



**ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO**

**FACULTAD DE INFORMÁTICA Y ELECTRÓNICA**

**ESCUELA INGENIERÍA EN SISTEMAS**

**“PROPUESTA METODOLÓGICA PARA APLICAR BUSINESS  
INTELLIGENCE CASO PRÁCTICO “COHERVI S.A.””**

**TESIS DE GRADO PREVIA OBTENCIÓN DEL TÍTULO DE  
INGENIERO EN SISTEMAS INFORMÁTICOS**

**PRESENTADO POR:**

Elsi Elizabeth Ilbay Yupa

**RIOBAMBA-OCTUBRE**

**2009**

## **AGRADECIMIENTO**

*Mi más profundo agradecimiento a Dios, por ser mi guía y compañero, quien me ha dado fortaleza y sabiduría para sobrellevar todos los obstáculos que se presentaron en la ejecución de esta tesis.*

*Al mismo tiempo a la gente de la comunidad open source, en especial a dos personas: Mariano y Darío, quienes han aportado constantemente con sus conocimientos y experiencias.*

*A mi mejor amigo Omar quien me ha apoyado incondicionalmente de forma personal y profesional.*

*A mi Directora de Tesis, ing. Ivonne Rodríguez por su experiencia científica en un marco de confianza, afecto y amistad, fundamentales para el progreso de este trabajo.*

*Al Ing. Fernando Proaño, por su comprensión, apoyo y sugerencias durante el desarrollo de la misma.*

## **DEDICATORIA**

*A mis Padres y hermanos por creer y confiar siempre en mí, por su apoyo y motivación que me brindaron a lo largo de la carrera.*

*A mis amigos por su confianza, lealtad y por enseñarme que no hay límites, que lo que me proponga lo puedo lograr y que solo depende de mi.*

## FIRMAS RESPONSABLES Y NOTAS

NOMBRES	FIRMAS	FECHA
Dr. Romeo Rodríguez <b>DECANO DE LA FACULTAD</b> <b>INFORMATICA Y ELECTRÓNICA</b>	_____	_____
Ing. Iván Menes <b>DIRECTOR DE ESCUELA</b> <b>INGENIRIA EN SISTEMAS</b>	_____	_____
Ing. Ivonne Rodríguez <b>DIRECTOR DE TESIS</b>	_____	_____
Ing. Fernando Proaño <b>MIEMBRO DE TESIS</b>	_____	_____
Tlgo. Carlos Rodríguez <b>DIRECTOR DEL</b> <b>CENTRO DE</b> <b>DOCUMENTACIÓN</b>	_____	_____

**Nota:** \_\_\_\_\_

## **RESPONSABILIDAD DEL AUTOR**

“Yo, Elsi Elizabeth Ilbay Yupa, soy responsable de las ideas, doctrinas y resultados expuestos en esta Tesis, y el patrimonio intelectual de la misma pertenecen a la Escuela Superior Politécnica de Chimborazo”.

---

FIRMA

## ÍNDICE DE ABREVIATURAS

<b>SAG</b>	Sistema de Análisis Gerencial
<b>CASE</b>	Ingeniería de Software Asistida por Computadores
<b>DBMS</b>	Sistema Administrador de Base de Datos
<b>ESPOCH</b>	Escuela Superior Politécnica de Chimborazo
<b>XP</b>	Programación Extrema
<b>HW</b>	Hardware
<b>JDBC</b>	Open Databases Connectivity
<b>ODBC</b>	On-Line Analytical Processing (Proceso analítico en línea)
<b>RDBMS</b>	Sistema Administrador de Base de Datos Relacionales
<b>SW</b>	Software
<b>OLAP</b>	Procesamiento analítico en línea
<b>OLTP</b>	Procesamiento de transacciones en línea
<b>BI</b>	Business Intelligence
<b>MOLAP</b>	Procesamiento analítico en línea multidimensional
<b>ROLAP</b>	Procesamiento analítico en línea relacional
<b>HOLAP</b>	Procesamiento analítico en línea híbrido
<b>CMI</b>	Cuadros de Mando Integrales
<b>DSS</b>	Sistemas de Soporte a la Decisión
<b>EIS</b>	Sistemas de Información Ejecutiva
<b>CMO</b>	Cuadro de mando Operativo
<b>MIS</b>	Sistemas de información gerencial
<b>AIS</b>	Sistemas de información Administrativa
<b>SSEE</b>	Sistemas Expertos basados en Inteligencia Artificial
<b>GDSS</b>	Sistemas de apoyo a decisiones de grupo
<b>ROI</b>	Retorno de inversión
<b>NPV</b>	Valor presente neto
<b>IRR</b>	Índice de retorno interno
<b>BDM</b>	Modelado de Base de Datos
<b>GA</b>	General Available
<b>MEDEPROBI</b>	Metodología para desarrollar proyectos de Business Intelligence.

# ÍNDICE GENERAL

## ÍNDICE DE ABREVIATURAS

## ÍNDICE GENERAL

## ÍNDICE DE FIGURAS

## ÍNDICE DE TABLAS

## CAPÍTULO I

<b>MARCO REFERENCIAL</b> .....	<b>20</b>
<b>1.1. ANTECEDENTES</b> .....	<b>20</b>
<b>1.2. JUSTIFICACIÓN</b> .....	<b>21</b>
<b>1.3. OBJETIVOS</b> .....	<b>22</b>
1.3.1 OBJETIVO GENERAL .....	22
1.3.2 OBJETIVOS ESPECÍFICOS.....	22
<b>1.4. HIPÓTESIS</b> .....	<b>22</b>
<b>MARCO TEORICO</b> .....	<b>23</b>
<b>2.1 INTRODUCCIÓN</b> .....	<b>23</b>
<b>2.2 BUSINESS INTELLIGENCE</b> .....	<b>24</b>
2.2.1 VENTAJAS DE BUSINESS INTELLIGENCE.....	25
2.2.2 DIFERENCIA ENTRE DATOS, INFORMACIÓN, CONOCIMIENTO .....	26
<b>2.3 ÁREAS APLICATIVAS DE BUSINESS INTELLIGENCE</b> .....	<b>28</b>
<b>2.4 ARQUITECTURA TÍPICA DE SOLUCIONES DE BUSINESS INTELLIGENCE</b> .....	<b>28</b>
2.4.1 FUNCIONES DE LA BASE DE DATOS RELACIONAL: .....	28
2.4.2 FUNCIONES DE LA BASE DE DATOS MULTIDIMENSIONAL. ....	29
2.4.3 FUNCIONES DEL VISUALIZADOR:.....	29
<b>2.5 COMPONENTES DE BUSINESS INTELLIGENCE</b> .....	<b>29</b>
2.5.1 DATAWAREHOUSE.....	29
2.5.1.1 <i>Arquitectura del datawarehouse</i> .....	31
2.5.2 DATA MARTS.....	32
2.5.2.1 <i>Datamarts Dependientes</i> .....	33
2.5.2.2 <i>Datamarts Independientes</i> .....	34
2.5.3 PROCESAMIENTO ANALÍTICO EN LÍNEA (OLAP) .....	34
2.5.4 PROCESAMIENTO TRANSACCIONAL EN LÍNEA (OTLP) .....	35
2.5.5 CUBO MULTIDIMENSIONAL.....	37
2.5.5.1 <i>Estructura del Cubo</i> .....	38
2.5.5.2 <i>Esquema del Cubo</i> .....	39
2.5.5.3 <i>Almacenamiento de Cubo</i> .....	40
<b>2.6 PRODUCTOS DE BUSINESS INTELLIGENCE</b> .....	<b>42</b>
2.6.1 CUADROS DE MANDO INTEGRALES (CMI).....	42
2.6.1.1 <i>Diferencia con otras herramientas de Business Intelligence</i> .....	42
2.6.1.2 <i>Tipos de Cuadros de Mando</i> .....	43
2.6.1.3 <i>Beneficios de la implantación de un Cuadro de Mando Integral</i> .....	45
2.6.1.4 <i>Riesgos de la implantación de un Cuadro de Mando Integral</i> .....	45
2.6.2 SISTEMAS DE SOPORTE A LA DECISIÓN (DSS) .....	45
2.6.2.1 <i>Diferencia con otras herramientas de Business Intelligence</i> .....	47
2.6.2.2 <i>Tipos de Sistemas de Soporte a Decisiones</i> .....	47
2.6.3 SISTEMAS DE INFORMACIÓN EJECUTIVA (EIS).....	48

<b>ANÁLISIS DE LAS METODOLOGÍAS DE BUSINESS INTELLIGENCE.....</b>	<b>50</b>
<b>3.1 INTRODUCCIÓN.....</b>	<b>50</b>
<b>3.2 METODOLOGÍA RALPH KIMBALL.....</b>	<b>52</b>
3.2.1. PLANIFICACIÓN DEL PROYECTO.....	54
3.2.2 DEFINICIÓN DE LOS REQUERIMIENTOS DEL NEGOCIO .....	55
3.2.3 MODELADO DIMENSIONAL .....	57
3.2.4 DISEÑO FÍSICO .....	58
3.2.5 DISEÑO Y DESARROLLO DE PRESENTACIÓN DE DATOS .....	59
3.2.6 DISEÑO DE LA ARQUITECTURA TÉCNICA .....	60
3.2.7 SELECCIÓN DE PRODUCTOS E INSTALACIÓN .....	62
3.2.8 ESPECIFICACIÓN DE APLICACIONES PARA USUARIOS FINALES .....	63
3.2.9 DESARROLLO DE APLICACIONES PARA USUARIOS FINALES .....	65
3.2.10 IMPLEMENTACIÓN.....	67
3.2.11 MANTENIMIENTO Y CRECIMIENTO .....	70
3.2.12 GERENCIAMIENTO DEL PROYECTO.....	72
<b>3.3 METODOLOGÍA CRISP-DM.....</b>	<b>73</b>
3.3.1 COMPRENSIÓN DEL NEGOCIO: .....	73
3.3.2 COMPRENSIÓN DE LOS DATOS .....	74
3.3.3 PREPARACIÓN DE LOS DATOS .....	74
3.3.4 MODELADO DEL ALMACÉN .....	76
3.3.4.1 Métodos OLAP .....	76
3.3.4.2 Análisis y modelado multidimensional.....	78
3.3.4.3 Esquema multidimensional.....	79
3.3.5 EVALUACIÓN .....	84
3.3.5.1 Evaluación del modelo .....	85
Óptica de negocio .....	85
Óptica económica.....	86
Óptica del usuario .....	86
Óptica técnica .....	87
3.3.6 DESPLIEGUE: .....	88
<b>3.4 METODOLOGÍA MÉTRICA VERSIÓN 3 INTEGRÁNDOSE CON UNA METODOLOGÍA DE DESARROLLO Y EXPLOTACIÓN DE DATAWAREHOUSES .....</b>	<b>89</b>
3.4.1 PLANIFICACIÓN DEL SISTEMA.....	90
3.4.1.1 Definición del plan de trabajo.....	90
3.4.2 GESTIÓN DE LA CONFIGURACIÓN DEL SISTEMA .....	90
3.4.2.1 Especificar la política de gestión de la configuración para los componentes del sistema... 90	
3.4.3 GESTIÓN DE LA CALIDAD DEL SISTEMA .....	90
3.4.3.1 Especificar la política de gestión de calidad para el sistema .....	90
3.4.4 ESTUDIO DE VIABILIDAD DEL SISTEMA.....	90
3.4.4.1 Establecimiento del Alcance del Sistema .....	91
3.4.4.2 Estudio de la Situación Actual .....	91
3.4.4.3 Definición de Requisitos del Sistema.....	92
3.4.4.4 Selección de la Solución.....	93
3.4.5. ANÁLISIS DEL SISTEMA DE INFORMACIÓN.....	93
3.4.5.1 Definición del sistema.....	93
3.4.5.2 Establecimientos de Requisitos.....	94
3.4.5.3 Identificación de los subsistemas de análisis.....	95
3.4.5.4 Elaboración del modelo de datos.....	95
3.4.5.5 Definir las interfaces de usuario.....	96
3.4.5.6 Análisis de consistencia y especificación de requisitos.....	97
3.4.5.7 Aprobación del análisis de información.....	97
3.4.6 DISEÑO DEL SISTEMA DE INFORMACIÓN.....	97
3.4.6.1 Definir la arquitectura del sistema.....	97
3.4.6.2 Diseño de los casos de uso real.....	98
3.4.6.3 Diseño físico de los datos.....	98

3.4.6.4 Aprobación del diseño del sistema de información.....	99
3.4.7 CONSTRUCCIÓN DEL SISTEMA DE INFORMACIÓN.....	99
3.4.7.1 Ejecución de las pruebas unitarias.....	100
3.4.7.2 Aprobación del Sistema de Información.....	101
<b>DESARROLLO DE LA METODOLOGÍA PROPUESTA.....</b>	<b>102</b>
<b>4.1 INTRODUCCIÓN.....</b>	<b>102</b>
<b>4.2 ETAPAS DE LA METODOLOGÍA PARA DESARROLLAR PROYECTOS DE BUSINESS INTELLIGENCE.....</b>	<b>104</b>
4.2.1 FASE I: PLANIFICACION DEL PROYECTO.....	106
4.2.1.1 PROPOSITO:.....	106
4.2.1.2 DESCRIPCION:.....	107
4.2.1.3 RESULTADOS:.....	107
4.2.1.4 ACTIVIDADES.....	107
4.2.1.4.1 Definir el ámbito del negocio.....	107
4.2.1.4.2 Analizar los escenarios problemas.....	108
4.2.1.4.3 Determinar alternativas de Solución.....	108
4.2.1.4.4 Analizar unidades organizativas afectadas.....	108
4.2.1.4.5 Definir los costos.....	109
4.2.1.4.6 Realizar el cronograma de trabajo.....	109
4.2.1.4.7 Analizar los riesgos.....	109
4.2.1.4.8 Realizar estudio de factibilidad.....	110
4.2.2 FASE II: DEFINICION DE REQUERIMIENTOS.....	112
4.2.2.1 PROPOSITO:.....	112
4.2.2.2 DESCRIPCIÓN:.....	112
4.2.2.3 RESULTADOS:.....	112
4.2.2.4 ACTIVIDADES:.....	113
4.2.2.4.1 Mencionar técnicas utilizadas en la recopilación de requerimientos.....	113
4.2.2.4.2 Describir los requerimientos.....	113
4.2.3 FASE III: DISEÑO DE LA ARQUITECTURA.....	113
4.2.3.1 PROPOSITO:.....	113
4.2.3.2 DESCRIPCION:.....	114
4.2.3.3 RESULTADOS:.....	114
4.2.3.4 ACTIVIDADES:.....	114
4.2.3.4.1 Diseñar la arquitectura.....	115
4.2.3.4.2 Identificar casos de uso.....	115
4.2.3.4.3 Desarrollar Casos de Uso y Diagramas de Casos de Uso del sistema Propuesto.....	116
4.2.3.4.4 Determinar los datos fuentes.....	116
4.2.3.4.5 Definir Base de Datos Intermedia de la bodega de datos.....	117
4.2.3.4.6 Realizar el modelo dimensional.....	118
4.2.3.4.7 Elaborar el modelo físico.....	121
4.2.3.4.8 Diseñar cubos multidimensionales.....	124
4.2.3.4.9 Especificar características de aplicaciones para usuarios finales.....	125
4.2.3.4.10 Desarrollar diagramas de secuencia de los procesos de carga.....	127
4.2.3.4.11 Realizar Diagrama de componentes.....	129
4.2.3.4.12 Definir diagrama de nodos.....	129
4.2.4 FASE IV: CONSTRUCCION Y DESARROLLO.....	130
4.2.4.1 PROPOSITO:.....	130
4.2.4.2 DESCRIPCION:.....	130
4.2.4.3 RESULTADOS:.....	130
4.2.4.4 ACTIVIDADES:.....	131
4.2.4.4.1 Seleccionar herramientas.....	131
4.2.4.4.2 Especificar permisos y accesos de los usuarios.....	132
4.2.4.4.3 Desarrollar el diseño.....	133
4.2.5 FASE V: PRUEBAS E IMPLANTACION.....	133
4.2.5.1 PROPOSITO:.....	133
4.2.5.2 DESCRIPCION:.....	134
4.2.5.3 RESULTADOS:.....	134
4.2.5.4 ACTIVIDADES:.....	134

4.2.5.4.1 Planificar las prueba.....	135
4.2.5.4.2 Generar casos de pruebas.....	135
4.2.5.4.3 Realizar procedimientos de ejecución y reportes.....	137
4.2.5.4.4 Formalizar procedimientos de corrección de errores.....	138
4.2.5.4.5 Implantar el sistema.....	140
4.2.6 FASE VI: ADMINISTRACION Y MANTENIMIENTO DEL SISTEMA.....	141
4.2.6.1 PROPOSITO:.....	142
4.2.6.2 DESCRIPCION:.....	142
4.2.6.3 RESULTADOS:.....	142
4.2.6.4 ACTIVIDADES:.....	142
4.2.6.4.1 Capacitar y dar soporte técnico.....	143
4.2.6.4.2 Realizar la gestión de los módulos.....	143
<b>IMPLEMENTAR BUSINESS INTELLIGENCE EN COHERVI S.A.....</b>	<b>144</b>
<b>5.1 INTRODUCCIÓN.....</b>	<b>144</b>
<b>5.2 ETAPAS DE LA METODOLOGÍA PARA DESARROLLAR PROYECTOS DE BUSINESS INTELLIGENCE.....</b>	<b>145</b>
5.2.1 FASE I: PLANIFICACION DEL PROYECTO.....	145
5.2.1.1 PROPOSITO:.....	145
5.2.1.2 DESCRIPCION:.....	146
5.2.1.3 RESULTADOS:.....	146
5.2.1.4 ACTIVIDADES.....	146
5.2.1.4.1 Definir el ámbito del negocio.....	146
5.2.1.4.1.1 Misión.....	146
5.2.1.4.1.2 Visión.....	146
5.2.1.4.1.3 Antecedente Tecnológico.....	147
5.2.1.4.2 Analizar los escenarios problemas.....	148
5.2.1.4.3 Determinar alternativas de Solución.....	150
5.2.1.4.4 Analizar unidades organizativas afectadas.....	152
- Ejecutar pruebas del sistema.....	153
5.2.1.4.5 Definir los costos.....	153
5.2.1.4.6 Realizar el cronograma de trabajo.....	154
5.2.1.4.7 Analizar los riesgos.....	155
5.2.1.4.8 Realizar un estudio de factibilidad.....	157
5.2.2 FASE II: DEFINICION DE REQUERIMIENTOS.....	158
5.2.2.1 PROPOSITO:.....	158
5.2.2.2 DESCRIPCIÓN:.....	158
5.2.2.3 RESULTADOS:.....	159
5.2.2.4 ACTIVIDADES:.....	159
5.2.2.4.1 Mencionar técnicas utilizadas en la recopilación de requerimientos.....	159
5.2.2.4.2 Describir los requerimientos.....	159
5.2.3 FASE III: DISEÑO.....	162
5.2.3.1 PROPOSITO:.....	162
5.2.3.2 DESCRIPCION:.....	162
5.2.3.3 RESULTADOS:.....	162
5.2.3.4 ACTIVIDADES:.....	162
5.2.3.4.1 Diseñar la arquitectura.....	162
5.2.3.4.2 Identificar Casos de Uso.....	163
5.2.3.4.3 Desarrollar Casos de uso y Diagramas de Casos de uso del sistema Propuesto.....	164
5.2.3.4.4 Determinar los datos fuentes.....	175
5.2.3.4.5 Definir base de datos intermedia de la bodega de datos.....	182
5.2.3.4.7 Elaborar el modelo físico.....	191
5.2.3.4.8 Diseñar cubos multidimensionales.....	200
5.2.3.4.9 Especificar características de aplicaciones para usuarios finales.....	204
5.2.3.4.10 Desarrollar Diagrama Secuencia de los procesos de carga.....	208
5.2.3.4.11 Realizar Diagrama de componentes.....	211
5.2.3.4.12 Definir diagrama de nodos.....	211
5.2.4 FASE IV: CONSTRUCCION Y DESARROLLO.....	212
5.2.4.1 PROPOSITO:.....	212
5.2.4.2 DESCRIPCION:.....	212

5.2.4.3 RESULTADOS:	212
5.2.4.4 ACTIVIDADES:	213
5.2.4.4.1 Seleccionar herramientas:	213
5.2.4.4.2 Especificar permisos y accesos de los usuarios:	214
5.2.4.4.3 Desarrollar el Diseño:	215
5.2.4.4.3.1 Construir el Datawarehouse en Mysql:	216
5.2.4.4.3.2 Instalación y configuración de la Suite de Pentaho:	217
5.2.4.4.3.3 Realizar procesos de carga usando Pentaho Data Integration:	223
5.2.4.4.3.4 Desarrollar cubos multidimensionales con Schema Workbench y publicarlos usando Pentaho Design Studio:	229
5.2.4.4.3.5 Elaborar Reportes Estáticos – Ad Hoc usando Pentaho Report Designer y Metadata Editor:	235
5.2.4.4.3.6 Ejecutar consola de usuario:	242
5.2.5 FASE V: PRUEBAS E IMPLANTACION:	242
5.2.5.1 PROPOSITO:	242
5.2.5.2 DESCRIPCION:	243
5.2.5.3 RESULTADOS:	243
5.2.5.4 ACTIVIDADES:	243
5.2.5.4.1 Planificar las prueba:	243
5.2.5.4.2 Generar casos de pruebas:	244
5.2.5.4.3 Realizar procedimientos de ejecución y reportes:	251
5.2.5.4.4 Formalizar procedimientos de corrección de errores:	255
5.2.5.4.5 Implantar el sistema:	256
5.2.6 FASE VI: ADMINISTRACION Y MANTENIMIENTO DEL SISTEMA:	257
5.2.6.1 PROPOSITO:	257
5.2.6.2 DESCRIPCION:	258
5.2.6.3 RESULTADOS:	258
5.2.6.4 ACTIVIDADES:	258
5.2.6.4.1 Capacitar y dar soporte técnico:	258
5.2.6.4.2 Realizar la gestión de los módulos:	259
<b>5.3 RESULTADOS OBTENIDOS DE LA METODOLOGÍA PROPUESTA</b>	<b>262</b>
<b>CONCLUSIONES</b>	
<b>RECOMENDACIONES</b>	
<b>RESUMEN</b>	
<b>SUMMARY</b>	
<b>GLOSARIO</b>	
<b>BIBLIOGRAFÍA</b>	
<b>ANEXOS</b>	

## ÍNDICE DE FIGURAS

Figura II.1 – Concepto de Business Intelligence .....	24
Figura II.2 - Arquitectura típica de soluciones de Business Intelligence.....	28
Figura II.3 - Arquitectura para manejo de información .....	29
Figura II.4 - Arquitectura de un Datawarehouse centralizado .....	32
Figura II.5 - Arquitectura Datamarts Dependientes .....	33
Figura II.6 - Arquitectura de Datamarts Independientes.....	34
Figura II.7 - Esquema en Estrella .....	39
Figura II.8 – Esquema en copo de nieve.....	40
Figura II.9 - Diferencia con otras herramientas de Business Intelligence .....	43
Figura II.10 - Cuadro de Mando Integral .....	44
Figura II.11 - Sistema de Soporte a la Decisión.....	46
Figura II.12 - Sistemas de Información Ejecutiva.....	49
Figura III.13 - ¿Qué es un Data Warehouse?.....	52
Figura III.14 - Business Dimensional Lifecycle propuesto por Ralph Kimball.....	53
Figura III.15 - Iteración entre Planificación y Requerimientos.....	54
Figura III.16 - Iteración entre Requerimientos del Negocio y etapas subsiguientes.....	56
Figura III.17 - Diagramas para Modelado Dimensional.....	57
Figura III.18 - Proceso de diseño físico a alto nivel .....	58
Figura III.19 - Proceso ETL.....	59
Figura III.20 - Modelo de Arquitectura Técnica a alto nivel .....	61
Figura III.21 - Ejemplo de matriz de evaluación de productos .....	62
Figura III.22 - Variedad de interfases por perfil .....	63
Figura III.23 - Configuración de Meta data y construcción de reportes .....	65
Figura III.24 - Implementación como convergencia de la tecnología, los datos y las aplicaciones .....	67
Figura III.25 - Data Warehousing como proceso de naturaleza espiral.....	70
Figura III.26 - Gerenciamiento del proyecto.....	72
Figura III.27 - Metodología de Crisp-Dm .....	73
Figura III.28 - Esquema multidimensional .....	79
Figura III.29 - Esquema en estrella .....	80
Figura IV.30 - Etapas de la Metodología para Desarrollar Proyectos de Business Intelligence	104
Figura IV.31 - Esquema de cada fase. ....	105
Figura IV.32 - Descripción de la fase.....	107
Figura IV.33 - Actividades de la planificación del proyecto. ....	107
Figura IV.34 - Descripción de la fase.....	112
Figura IV.35 - Actividades de Definición de requerimientos.....	113
Figura IV.36 - Descripción de la fase.....	114
Figura IV.37 - Actividades del Diseño de la arquitectura. ....	114
Figura IV.38 - Diseño de la Arquitectura. ....	115
Figura IV.39 - Esquema de una vista multidimensional de las tablas de hecho. ....	119
Figura IV.40 - Representación de un diagrama en estrella. ....	121
Figura IV.41 - Representación de un diagrama copo de nieve.....	121
Figura IV.42 - Modelo lógico y modelo físico.....	122
Figura IV.43 – Conversión del modelo lógico al físico.....	123
Figura IV.44 - Proceso de carga.....	128
Figura IV.45 - Descripción de la fase.....	130
Figura IV.46 - Actividades de Construcción y Desarrollo.....	131
Figura IV.47 - Descripción de la fase.....	134
Figura IV.48 - Actividades de Pruebas e Implantación. ....	134
Figura IV.49 - Arquitectura centralizada.....	141
Figura IV.50 - Arquitectura distribuida. ....	141
Figura IV.51 - Descripción de la fase.....	142
Figura IV.52 - Actividades de Administración y mantenimiento del Sistema. ....	142
Figura V.53 - Descripción de la fase.....	146
Figura V.54 - Caso de uso sistema actual .....	150

Figura V.55 - Caso de uso sistema propuesto .....	151
Figura V.56 - Descripción de la fase.....	158
Figura V.57 - Descripción de la fase.....	162
Figura V.58 - Diseño de la Arquitectura.....	163
Figura V.59 - Caso de uso # 01 .....	165
Figura V.60 - Caso de uso # 02 .....	167
Figura V.61 - Caso de uso # 03 .....	168
Figura V.62 - Caso de uso # 04 .....	169
Figura V.63 - Caso de uso # 05 .....	170
Figura V.64 - Caso de uso # 06 .....	171
Figura V.65 - Caso de uso # 07 .....	172
Figura V.66 - Caso de uso # 08 .....	173
Figura V.67 - Caso de uso # 09 .....	174
Figura V.68 – Datamart Compras.....	189
Figura V.69 – Datamart Inventario (Fact_inventario) .....	190
Figura V.70 - Datamart Inventario (Fact_caducados) .....	190
Figura V.71 - Datamart ventas.....	191
Figura V.72 - Conversión del modelo lógico al físico data mart compras .....	193
Figura V.73 - Conversión del modelo lógico al físico data mart inventario (Fact_inventario) ...	194
Figura V.74 - Conversión del modelo lógico al físico data mart inventario (Fact_caducados) .	194
Figura V.75 - Conversión del modelo lógico al físico data mart ventas.....	195
Figura V.76 - Diseño del ETL desde el sistema fuente a dwh_cohervi .....	208
Figura V.77 - Diseño del ETL desde dwh_cohervi a ddm_cohervi.....	209
Figura V.78 – Diagrama de componentes.....	211
Figura V.79 – Diagramas de nodos.....	211
Figura V.80 - Descripción de la fase.....	212
Figura V.81 – Arquitectura y Herramientas .....	215
Figura V.82 – Arquitectura Pentaho Open BI Suite .....	217
Figura V.83 – Ejecutar el archivo start-pentaho.bat .....	219
Figura V.84 – Pantalla de servicios levantados de la Suite de Pentaho .....	219
Figura V.85 – Ventana inicial de la Suite de Pentaho .....	220
Figura V.86 - Ejecutar el archivo startup.bat.....	220
Figura V.87 – Pantalla de servicios levantados de la Consola de Administrador .....	220
Figura V.88 – Ventana inicial de la Consola de Administrador .....	221
Figura V.89 – Crear un DataSource.....	221
Figura V.90 - Parámetros de conexión del DataSource.....	222
Figura V.91 - Resultado de Conexión.....	222
Figura V.92 - Diseño de ETL.....	224
Figura V.93 – Ejecutar el archivo Spoon.bat .....	224
Figura V.94 - Ventana inicial de la herramienta ETL.....	225
Figura V.95 - Área de trabajo de la herramienta ETL.....	225
Figura V.96 - Parámetros de conexión Pentaho Data Integration.....	227
Figura V.97 - Poblamiento de la base de datos dwh_cohervi.....	227
Figura V.98 - Poblamiento de la base de datos ddm_cohervi .....	228
Figura V.99 - Carga periódica .....	228
Figura V.100 - Ejecutar archivo schema-worbench.bat .....	229
Figura V.101 - Pantalla inicial del schema workbench.....	230
Figura V.102 - Parámetros de conexión.....	230
Figura V.103 – Área para diseñar cubos.....	231
Figura V.104 - Despliegue del esquema desarrollado .....	231
Figura V.105 - Parámetros de Publicación .....	232
Figura V.106 - Repositorio de publicación .....	232
Figura V.107 - Ejecutar el archivo startup.bat .....	233
Figura V.108 - Ventana principal del Design Studio .....	233
Figura V.109 - Pantalla de Creación Design Studio .....	234
Figura V.110 - Pantalla de Diseño del Design Studio .....	234
Figura V.111 - Pantalla para ejecutar el archivo ReportDesigner.exe.....	235
Figura V.112 - Pantalla inicial de Report Designer .....	236

Figura V.113 - Parámetros de conexión.....	236
Figura V.114 - Selección de Datasource y query.....	237
Figura V.115 - Ventana de Diseño del reporte.....	237
Figura V.116 - Parámetros de publicación.....	238
Figura V.117 - Repositorio de publicación.....	238
Figura V.118 - Ejecutar la herramienta MetadataEditor.....	239
Figura V.119 - Ventana inicial MetadataEditor.....	239
Figura V.120 - Parámetros de conexión Metadata Editor.....	240
Figura V.121 - Creación de Nuevo modelo de Negocios.....	240
Figura V.122 - Selección de tablas físicas.....	241
Figura V.123 - Modelo de Negocios.....	241
Figura V.124 - Parámetros de publicación de Metadata Editor.....	242
Figura V.125 - Descripción de la fase.....	243
Figura V.126 - Sistema implementado.....	257
Figura V.127 - Descripción de la fase.....	258

## ÍNDICE DE TABLAS

Tabla II.1 - Diferencia entre sistemas OLTP y OLAP .....	36
Tabla IV-2. Técnicas y Herramientas propuestas .....	105
Tabla IV-3. Caso de uso .....	108
Tabla IV.4 - Unidades organizativas .....	108
Tabla IV.5 – Cronograma de trabajo .....	109
Tabla IV.6 - Descripción de riesgos: .....	109
Tabla IV.7 - Categorización de riesgos:.....	110
Tabla IV.8 - Plan de contingencia.....	110
Tabla IV.9 - Factibilidad técnica.....	110
Tabla IV.10 - Factibilidad Operativa .....	111
Tabla IV.11 - Factibilidad Económica. ....	111
Tabla IV.12 - Identificar casos de uso.....	115
Tabla IV.13 - Tablas fuentes .....	116
Tabla IV.14 - Modelo de datos fuente .....	116
Tabla IV.15 - Descripción Bases de Datos Fuente y Base de Datos Intermedia .....	117
Tabla IV.16 – Detalle de los campos de la BD Fuente - Destino.....	117
Tabla IV.17 – Granulidad.....	120
Tabla IV.18 – Tablas de Hecho y medidas .....	120
Tabla IV.19 - Matriz de verificación del modelo dimensional .....	122
Tabla IV.20 - Definición de los atributos de una tabla de hechos. ....	123
Tabla IV.21 - Diseño de cubos multidimensionales .....	124
Tabla IV.22 - Definición de plantillas de reportes.....	126
Tabla IV.23 - Formato de planilla de reporte .....	127
Tabla IV.24 - Especificación de los procesos ETL.....	129
Tabla IV.25 - Áreas a evaluar para escoger las herramientas.....	132
Tabla IV.26 - Permiso de base de Datos.....	132
Tabla IV.27 - Acceso a aplicaciones .....	133
Tabla IV.28 - Pruebas unitarias. ....	135
Tabla IV.29 - Caso de prueba .....	136
Tabla IV.30 - Informe de problema.....	137
Tabla IV.31 - Informe de problemas identificados. ....	138
Tabla IV.32 - Funciones del Proceso de carga .....	143
Tabla V.33 - Infraestructura de Servidores .....	147
Tabla V.34 - Caso de uso Sistema Actual.....	149
Tabla V.35 - Caso de uso .....	150
Tabla V.36 -Unidades organizativas.....	152
Tabla V.37 – Cronograma de Trabajo .....	154
Tabla V.38 - Descripción de riesgos: .....	155
Tabla V.39 - Categorización de riesgos: .....	155
Tabla V.40 - Plan de contingencia. ....	156
Tabla V.41 - Factibilidad técnica.....	157
Tabla V.42 - Factibilidad Operativa .....	157
Tabla V.43 – Hardware del Servidor .....	160
Tabla V.44 – Hardware del Cliente .....	161
Tabla V.45 – Software del Servidor.....	161
Tabla V.46 – Software del Cliente .....	161
Tabla V.47 – Casos del uso del Sistema de Análisis Gerencial .....	164
Tabla V.48 - CU_AUT # 01 .....	164
Tabla V.49 - CU_VAOLAP # 02.....	166
Tabla V.50 - CU_VREP_EST_ADH # 03 .....	167
Tabla V.51 - CU_DVENTAS # 04.....	168
Tabla V.52 - CU_DCOMPRAS # 05.....	170
Tabla V.53 - CU_DINVENTARIO #06.....	171
Tabla V.54 - CU_GVOLAP #07 .....	172
Tabla V.55 - CU_GREP_ADHOC #08 .....	173

Tabla V.56 - CU_AGER #07.....	174
Tabla V.57 – Tablas de la Base de Datos Fuente .....	175
Tabla V.58 - Modelo de datos fuente.....	176
Tabla V.59 – Descripción de la BD Fuente y BD Intermedia .....	182
Tabla V.60 - Detalle de los campos de la BD Fuente - Destino.....	183
Tabla V.61 - Dimensiones y Fact Tables.....	187
Tabla V.62 - Granulidad.....	188
Tabla V.63 - Tabla de Hechos y medidas .....	188
Tabla V.64 - Matriz de verificación del modelo dimensional .....	192
Tabla V.65 - Atributos de la tabla de hechos Compras.....	196
Tabla V.66 - Atributos de la tabla de hechos ventas .....	196
Tabla V.67 - Atributos de la tabla de hechos inventario (Fact_inventario).....	196
Tabla V.68 - Atributos de la tabla de hechos inventario (Fact_caducados).....	197
Tabla V.69 - Atributos de la dimensión tipo de venta .....	197
Tabla V.70 - Atributos de la dimensión tiempo.....	197
Tabla V.71 - Atributos de la dimensión vendedor .....	198
Tabla V.72 - Atributos de la dimensión cliente .....	198
Tabla V.73 - Atributos de la dimensión zona .....	198
Tabla V.74 - Atributos de la dimensión producto .....	199
Tabla V.75 - Atributos de la dimensión proveedor.....	199
Tabla V.76 - Cubo multidimensional de Compras.....	200
Tabla V.77 - Cubo multidimensional de ventas .....	201
Tabla V.78 - Cubo multidimensional de inventario (Fact_inventario).....	202
Tabla V.79 - Cubo multidimensional de inventario (Fact_caducados).....	203
Tabla V.80 - Definición de plantillas de reportes.....	204
Tabla V.81 - Planilla de reporte para Departamento de Compras .....	205
Tabla V.82 - Plantilla de reporte para Departamento de ventas.....	206
Tabla V.83 - Planilla de reporte para Departamento de Inventario .....	207
Tabla V.84 – Especificación de ETL desde Foxprox a BD dwh_cohervi.....	208
Tabla V.85 - Especificación de ETL desde dwh_cohervi a ddm_cohervi.....	210
Tabla V.86 - Herramientas seleccionadas .....	213
Tabla V.87 - Permiso de base de Datos.....	214
Tabla V.88 - Acceso a las herramientas.....	215
Tabla V.89 - Caso de prueba 1 .....	244
Tabla V.90 - Caso de prueba 2.....	245
Tabla V.91 - Caso de prueba 3.....	246
Tabla V.92 - Caso de prueba 4.....	247
Tabla V.93 - Caso de prueba 5.....	248
Tabla V.94 - Caso de prueba 6.....	249
Tabla V.95 - Caso de prueba 7.....	250
Tabla V.96 - Caso de prueba 1 paso 1 .....	251
Tabla V.97 - Caso de prueba 1 paso 2.....	252
Tabla V.98 - Caso de prueba 2.....	252
Tabla V.99 - Caso de prueba 3.....	253
Tabla V.100 - Caso de prueba 4.....	253
Tabla V.101 - Caso de prueba 5.....	254
Tabla V.102 - Informe de problema.....	255
Tabla V.103 - Informe de problemas identificados estado abierto .....	255
Tabla V.104 – Requerimientos el Servidor .....	256
Tabla V.105 – Requerimientos del Cliente.....	256
Tabla V.106 - Extracción, transformación y carga .....	259
Tabla V.107 - Datawarehouse. ....	260
Tabla V.108 - Acceso a datos o consultas.....	261

# **CAPÍTULO I**

## **MARCO REFERENCIAL**

### **1.1. ANTECEDENTES**

Vivimos en una época en que la información es la clave para obtener una ventaja competitiva en el mundo de los negocios. Para mantenerse competitiva una empresa, los gerentes y tomadores de decisiones requieren de un acceso rápido y fácil a información útil y valiosa de la empresa. Una forma de solucionar este problema es por medio del uso de Business Intelligence o Inteligencia de Negocios.

Business Intelligence suele definirse como la transformación de los datos de la compañía en conocimiento para obtener una ventaja competitiva. Considerada también como un conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales en información desestructurada (interna y externa a la compañía) en información estructurada, para su explotación directa (reporting, análisis OLAP) o para su análisis y conversión en conocimiento soporte a la toma de decisiones sobre el negocio.

Los datos constituyen un elemento primordial para la empresa y sobre todo para el desarrollo de la metodología de Business Intelligence y cuyo objetivo central es la información relevante sobre la situación de la empresa la misma que será de gran apoyo a los tomadores de decisiones.

La implementación de soluciones de Business Intelligence garantiza aspectos como: Accesibilidad, Integración tecnológica y toma de decisiones.

## **1.2. JUSTIFICACIÓN**

Las compañías de la actualidad son juzgadas no únicamente por la calidad de sus productos o servicios, sino también por el grado en el que comparten información con sus clientes, empleados y socios. Sin embargo, la gran mayoría de las organizaciones tienen una abundancia de datos, pero una penuria de conocimiento. Es por ello que se desarrollará la metodología Business Intelligence, la misma que trata de englobar todos los sistemas de información de una organización para obtener de ellos no solo información, si no una verdadera inteligencia que le confiera a la organización una ventaja competitiva.

En la actualidad el sector comercial ha evolucionado notablemente, siendo esta una causa para que los administradores pierdan el control de los resultados del negocio, es decir no tienen los informes oportunos en el momento que ellos desean y sobre todo no disponen de información que les ayude a la toma de decisiones.

En nuestra facultad en los años anteriores se han desarrollado las siguientes tesis que tienen relación con el tema de estudio:

- ✚ Estudio y uso de los métodos y técnicas de Datawarehouse y Datamining aplicado en el ámbito académico de la Escuela Superior Politécnica de Chimborazo
- ✚ Estudio de la tecnología OLAP e implementación de un sistema de soporte decisional para fundación marco.
- ✚ Estudio de la tecnología Business Intelligence y su aplicación en un modelo de sistema de información gerencial en Petroproducción.

Todas estas tesis han enfocado a ciertos componentes de un Business Intelligence, además la ingeniería de Software no es suficiente para implementar proyectos que involucren BI, es por ello que es necesario crear una guía metodológica del BI que se adapte a los requerimientos de las empresas, es decir el BI se basa en el análisis de la fuentes y proporciona a los usuarios del sistema una manera universal de acceder, ver y utilizar información que se guarda en diferentes fuentes de gestión empresarial a través de una sola aplicación.

Esta metodología ayudará a los distintos departamentos de la empresa, a obtener resultados benéficos tales como: calcular la rentabilidad de la empresa, generar reportes globales o por secciones, crear escenarios con respecto a una decisión, hacer pronósticos de ventas y devoluciones, compartir información entre departamentos , análisis multidimensionales, generar y procesar datos, cambiar la estructura de toma de decisiones, mejorar el servicio al cliente .

### **1.3. OBJETIVOS**

#### **1.3.1 Objetivo general**

- ✚ Desarrollar una metodología para aplicar la Tecnología Business Intelligence en las empresas.

#### **1.3.2 Objetivos específicos**

- ✚ Revisar las tecnologías Business Intelligence que actualmente se están aplicando en el país.
- ✚ Análisis de las metodologías de Business Intelligence.
- ✚ Desarrollar una metodología que permita aplicar Business Intelligence correctamente en las empresas.
- ✚ Emplear la metodología propuesta para aplicar Business Intelligence en la empresa "COHERVI" S.A.

### **1.4. HIPÓTESIS**

La metodología propuesta ayudará a aplicar Business Intelligence correctamente en las áreas estratégicas de una empresa.

## **CAPÍTULO II**

### **MARCO TEORICO**

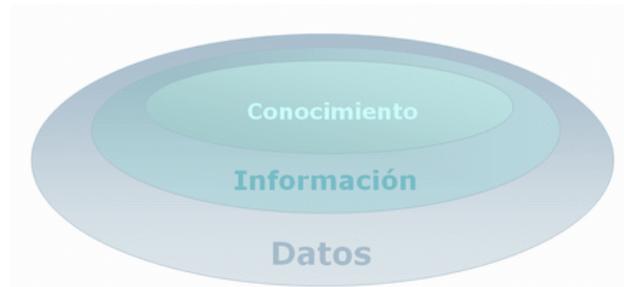
#### **2.1 INTRODUCCIÓN**

Business Intelligence es un término ideado por Gartner Group en los años 80 para escribir la capacidad de una organización para acceder y explotar la información residente en una base de datos, de manera que los usuarios puedan analizar esa información y desarrollar y adquirir, con base en ella, teorías y conocimientos que apoyen la toma de decisiones del negocio.

El cambiante entorno económico y la problemática de sistemas impulsan al surgimiento del llamado Business Intelligence, el cual es un concepto que trata de englobar todos los Sistemas de Información de una organización para obtener de ellos no solo información o conocimiento, si no una verdadera inteligencia que le confiera a la organización una ventaja competitiva por sobre sus competidores.

## 2.2 BUSINESS INTELLIGENCE

Business Intelligence es la habilidad para transformar los datos en información, y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios.



**Figura II.1 – Concepto de Business Intelligence**

Desde un punto de vista más pragmático, y asociándolo directamente con las tecnologías de la información, podemos definir Business Intelligence como el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada (interna y externa a la compañía) en información estructurada, para su explotación directa (reporting, análisis OLTP / OLAP, alertas, etc) o para su análisis y conversión en conocimiento, dando así soporte a la toma de decisiones sobre el negocio.

La inteligencia de negocio actúa como un factor estratégico para una empresa u organización, generando una potencial ventaja competitiva, que no es otra que proporcionar información privilegiada para responder a los problemas de negocio: entrada a nuevos mercados, promociones u ofertas de productos, eliminación de islas de información, control financiero, optimización de costes, planificación de la producción, análisis de perfiles de clientes, rentabilidad de un producto concreto, etc...

Las compañías actualmente usan una amplia gama de tecnologías y productos para saber qué es lo que está pasando en la organización. Las herramientas más comunes (simple consulta y reporte de datos, procesamiento analítico en línea, análisis estadístico, predicciones y minería de datos) pueden ser usadas de una gran variedad de formas. El objetivo de todo esto es transformar las montañas de datos en información útil para la empresa.

Los principales objetivos de la implementación de soluciones de BI son:

- ✚ **Accesibilidad:** se refiere a la posibilidad de los usuarios de consultar la información y manipular las diferentes variables por medio de una herramienta de fácil manejo y administración.
- ✚ **Integración tecnológica:** se refiere a la posibilidad del usuario final de consultar diferentes fuentes de datos sobre una interfaz que le permita manipular las variables de consulta para elaborar reportes e indicadores.
- ✚ **Toma de decisión:** las soluciones de BI brindan la posibilidad al usuario final de visualizar mediante herramientas gráficas indicadores que ayudan en la toma de decisiones.

### 2.2.1 Ventajas de Business Intelligence

Las ventajas derivadas de la implantación de este tipo de soluciones son:

- ✚ Disponer de más y mejor información de gestión. Ello mejora la toma de decisiones.
- ✚ Agilizar los plazos de elaboración y presentación de resultados, trasladando los esfuerzos hacia labores de análisis y reduciendo los costes de dichos procesos.
- ✚ Unificar las fuentes de información de gestión y homogenizar los conceptos, asegurando datos únicos y compartidos.
- ✚ Mejorar las estrategias comerciales, lo cual tendrá un impacto directo en la rentabilidad de la organización.

### Qué puede hacer Business Intelligence?

Con BI se puede:

- ✚ Generar reportes globales o por secciones
- ✚ Crear una base de datos de clientes
- ✚ Crear escenarios con respecto a una decisión.
- ✚ Hacer pronósticos de ventas y devoluciones.
- ✚ Compartir información entre departamentos.
- ✚ Análisis multidimensionales.

- ✚ Generar y procesar datos.
- ✚ Cambiar la estructura de toma de decisiones.
- ✚ Mejorar el servicio al cliente.

### **2.2.2 Diferencia entre Datos, información, conocimiento**

¿En qué se diferencia el conocimiento de los datos y de la información? En una conversación informal, los tres términos suelen utilizarse indistintamente y esto puede llevar a una interpretación libre del concepto de conocimiento. Quizás la forma más sencilla de diferenciar los términos sea pensar que los datos están localizados en el mundo y el conocimiento está localizado en agentes de cualquier tipo (personas, empresas, máquinas, mientras que la información adopta un papel mediador entre ambos.

#### **Datos**

Los datos son la mínima unidad semántica, y se corresponden con elementos primarios de información que por sí solos son irrelevantes como apoyo a la toma de decisiones. También se pueden ver como un conjunto discreto de valores, que no dicen nada sobre el por qué de las cosas y no son orientativos para la acción.

Un número telefónico o un nombre de una persona, por ejemplo, son datos que, sin un propósito, una utilidad o un contexto no sirven como base para apoyar la toma de una decisión.

Como cabe suponer, los datos pueden provenir de fuentes externas o internas a la organización, pudiendo ser de carácter objetivo o subjetivo, o de tipo cualitativo o cuantitativo, etc.

#### **Información**

La información se puede definir como un conjunto de datos procesados y que tienen un significado (relevancia, propósito y contexto), y que por lo tanto son de utilidad para quién debe tomar decisiones, al disminuir su incertidumbre. Los datos se pueden transformar en información añadiéndoles valor:

- ✚ **Contextualizando:** se sabe en qué contexto y para qué propósito se generaron.
- ✚ **Categorizando:** se conocen las unidades de medida que ayudan a interpretarlos.
- ✚ **Calculando:** los datos pueden haber sido procesados matemática o estadísticamente.
- ✚ **Corrigiendo:** se han eliminado errores e inconsistencias de los datos.
- ✚ **Condensando:** los datos se han podido resumir de forma más concisa (agregación).

Por tanto, la información es la comunicación de conocimientos o inteligencia, y es capaz de cambiar la forma en que el receptor percibe algo, impactando sobre sus juicios de valor y sus comportamientos.

*Información = Datos + Contexto (añadir valor) + Utilidad (disminuir la incertidumbre)*

## **Conocimiento**

El conocimiento es una mezcla de experiencia, valores, información y *know-how* que sirve como marco para la incorporación de nuevas experiencias e información, y es útil para la acción. Se origina y aplica en la mente de los conocedores. En las organizaciones con frecuencia no sólo se encuentra dentro de documentos o almacenes de datos, sino que también está en rutinas organizativas, procesos, prácticas, y normas.

El conocimiento se deriva de la información, así como la información se deriva de los datos. Para que la información se convierta en conocimiento es necesario realizar acciones como:

- ✚ Comparación con otros elementos.
- ✚ Predicción de consecuencias.
- ✚ Búsqueda de conexiones.
- ✚ Conversación con otros portadores de conocimiento.

## 2.3 ÁREAS APLICATIVAS DE BUSINESS INTELLIGENCE

- ✚ Ventas: Análisis de ventas; Detección de clientes importantes; Análisis de productos, líneas, mercados; Pronósticos y proyecciones.
- ✚ Marketing: Segmentación y análisis de clientes; Seguimiento a nuevos productos.
- ✚ Finanzas: Análisis de gastos; Rotación de cartera; Razones Financieras.
- ✚ Manufactura: Productividad en líneas; Análisis de desperdicios; Análisis de calidad; Rotación de inventarios y partes críticas.
- ✚ Embarques: Seguimiento de embarques; Motivos por los cuales se pierden pedidos.

## 2.4 ARQUITECTURA TÍPICA DE SOLUCIONES DE BUSINESS INTELLIGENCE

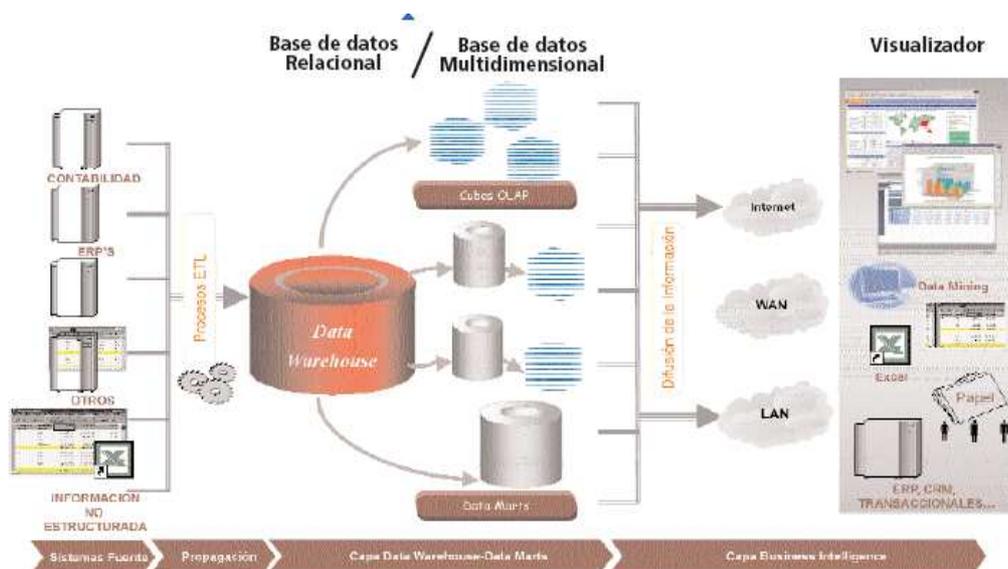


Figura II.2 - Arquitectura típica de soluciones de Business Intelligence

En general, la arquitectura típica de estas soluciones está basada en tres capas: BBDD Relacional, BBDD Multidimensional y Herramienta de visualización

### 2.4.1 Funciones de la Base de Datos Relacional:

- ✚ Depuración y homogenización.
- ✚ Almacenamiento de datos (hasta el nivel de detalle máximo y recogiendo datos actuales e históricos).
- ✚ Motor de cálculo, si fuese necesario.

### 2.4.2 Funciones de la Base de Datos Multidimensional.

- ✚ Velocidad de acceso y consulta.
- ✚ Capacidad de análisis desde varios puntos de vista (dimensiones de análisis).

### 2.4.3 Funciones del visualizador:

- ✚ Análisis de la información (drill-down, drill-through y rotación de filas por columnas).

## 2.5 COMPONENTES DE BUSINESS INTELLIGENCE

### 2.5.1 DataWarehouse

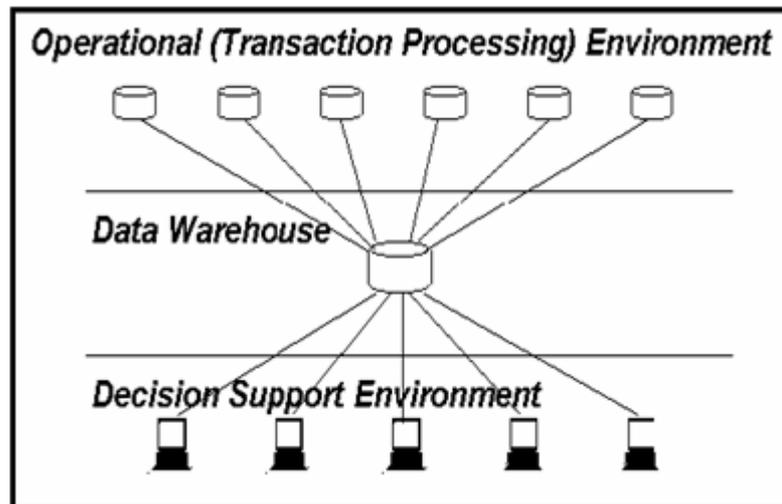


Figura II.3 - Arquitectura para manejo de información

Un datawarehouse es una base de datos diferente a la de los sistemas operacionales en cuanto a que:

- ✚ Es orientado a sujetos
- ✚ Maneja grandes cantidades de información
- ✚ Comprende múltiples versiones de un esquema de bases de datos
- ✚ Condensa y agrega información

Integra y asocia información de muchas fuentes de datos. Un Data Warehouse organiza y orienta los datos desde la perspectiva del usuario final. Muchos sistemas operativos organizan sus datos desde la perspectiva del negocio, para mejorar la

rapidez de acceso y actualización de los datos. Este tipo de organización, no es la más indicada para operaciones de consulta, típicas en el ambiente decisional.

La mayoría de los DataWarehouses contienen información histórica que se retira con frecuencia de los sistemas operativos porque ya no es necesaria para las aplicaciones operacionales y de producción. Por la necesidad de administrar tanto la información histórica como las actuales, un DataWarehouse es mayor que las bases de datos operacionales.

Un Datawarehouse guarda información histórica y la administra. Como la información histórica muchas veces es manejada por diferentes versiones de esquemas de bases de datos, el datawarehouse debe controlar la información originada por bases de datos diferentes.

Un datawarehouse condensa y agrega la información para presentarla en forma comprensible a las personas. La condensación y adición es esencial para retroceder y entender la imagen global.

Debido a que las organizaciones han administrado históricamente sus operaciones utilizando numerosas aplicaciones de software y múltiples bases de datos, se requiere de un datawarehouse para recopilar y organizar en un solo lugar la información que estas aplicaciones han acumulado al paso de los años.

El objetivo del Datawarehouse es el de satisfacer los requerimientos de información interna de la empresa para una mejor gestión. El contenido de los datos, la organización y estructura son dirigidos a satisfacer las necesidades de información de los analistas y usuarios tomadores de decisiones. El Datawarehouse es el lugar donde la gente puede acceder a sus datos.

Los Datawarehouses generan bases de datos tangibles con una perspectiva histórica, utilizando datos de múltiples fuentes que se fusionan en forma congruente. Estos datos se mantienen actualizados, pero no cambian al ritmo de los sistemas transaccionales.

Muchos datawarehouses se diseñan para contener un nivel de detalle hasta el nivel de transacción, con la intención de hacer disponible todo tipo de datos y características, para reportar y analizar.

Así un datawarehouse resulta ser un recipiente de datos transaccionales para proporcionar consultas operativas, y la información para poder llevar a cabo análisis multidimensional. Dentro de un datawarehouse existen dos tecnologías una relacional para consultas y otra multidimensional para realizar el análisis.

El modelo relacional en el cual se basa OLTP (Procesamiento Transaccional en Línea), tiene como objetivo mantener la integridad de la información (relaciones entre los datos) necesaria para operar un negocio de la manera más eficiente. Sin embargo, este modelo no corresponde a la forma como el usuario percibe la operación de un negocio.

El Datawarehouse está basado en un procesamiento distinto al utilizado por los sistemas operacionales, es decir, este se basa en OLAP (Procesos de Análisis en Línea), usado en el análisis de negocios y otras aplicaciones que requieren una visión flexible del negocio.

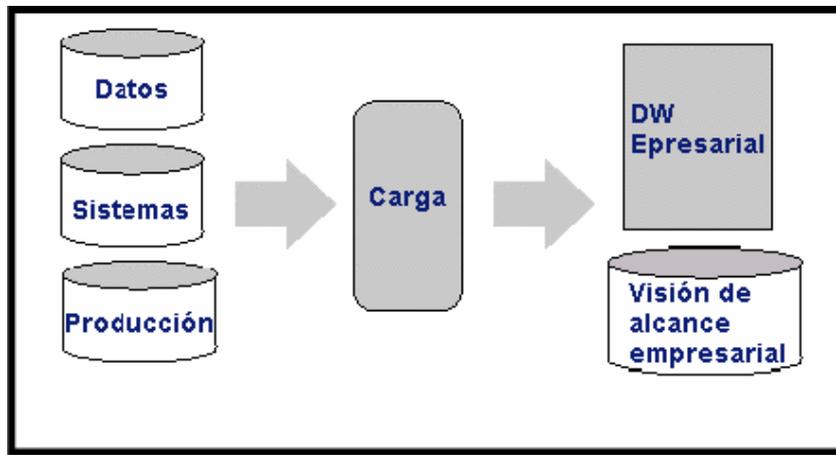
#### **2.5.1.1 Arquitectura del datawarehouse**

El éxito de la tecnología de DW ha sido muy grande en los últimos años y ha permitido evolucionar mucho la materia. Las distintas experiencias demostraron, ideas que en un principio parecían correctas, no lo eran tanto. Nos centraremos en los cambios que ha sufrido el DW con respecto a la arquitectura del sistema.

#### **Datawarehouse Central**

El concepto inicial detrás del Datawarehouse es el de crear un repositorio de alcance empresarial que homogeneizara y uniera todos los datos de la organización en una única estructura, desde donde todos los departamentos pudieran obtener una visión coherente de la organización.

Este concepto implicaba que todos los sistemas de producción provean de información al datawarehouse y que todas las extracciones y transformaciones dentro de toda la organización estuvieran bajo el control de un único proceso.



**Figura II.4 - Arquitectura de un Datawarehouse centralizado**

Los problemas en proyectos que aplicaron este concepto, fueron apareciendo.

Construir un datawarehouse global centralizado es complicado por varias razones que incluyen temas políticos empresariales, temas financieros, temas de performance y temas de estrategia. Los temas políticos principalmente se relacionan con pertenencia de datos y provincialismo, algunos departamentos no están de acuerdo con compartir sus datos.

Financiar estos proyectos implica invertir mucho dinero y esperar varios meses para obtener algo de utilidad. Surgen problemas también de performance al momento que usuarios de distintos departamentos quieren consultar su información al mismo tiempo dentro del datawarehouse. Por último, dado el tiempo que se necesita para desarrollar un sistema de este tamaño desalienta a los usuarios dado que al momento que se complete los requerimientos ya habrán cambiado.

### **2.5.2 Data Marts**

Los Data Marts, son almacenes de información específica que apuntan a un área del negocio en particular. El concepto en este caso deriva de la certeza que cualquier usuario tiene necesidades de información limitada, y aunque típicamente existen requerimientos para análisis funcionales cruzados, el tamaño de los requerimientos es reducido materialmente si limitamos el tamaño del datawarehouse en sí mismo.

Dos estrategias distintas se desarrollan a partir del concepto de data marts, la de data marts dependientes y la de data marts independientes.

Los principales beneficios de utilizar Datamarts son:

- ✚ Acelerar las consultas reduciendo el volumen de datos a recorrer
- ✚ Estructurar los datos para su adecuado acceso por una herramienta
- ✚ Dividir los datos para imponer estrategias de control de acceso
- ✚ Segmentar los datos en diferentes plataformas hardware
- ✚ Permite el acceso a los datos por medio de un gran número de herramientas del mercado, logrando independencia de estas.

### 2.5.2.1 Datamarts Dependientes

En esta arquitectura los datos son cargados desde los sistemas de producción hacia el datawarehouse empresarial y entonces subdivididos en data marts. Se llaman data marts dependientes porque utilizan los datos y metadatos del datawarehouse en lugar de obtenerlos de los sistemas de producción.

Esta solución resuelve los problemas de performance, estrategia, finanzas e incluso algunos de los problemas políticos. Aunque tiene esos puntos a favor, sigue teniendo que construir el datawarehouse global antes que los data marts sean implementados.

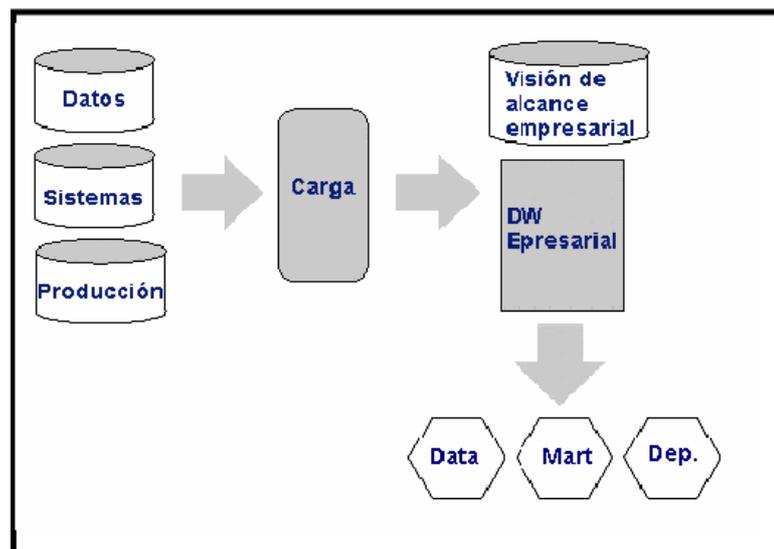


Figura II.5 - Arquitectura Datamarts Dependientes

### 2.5.2.2 Datamarts Independientes

Esta óptica es considerada por muchos como una alternativa del datawarehouse central. Con ella es posible comenzar con un sistema pequeño, invirtiendo menos dinero y obteniendo resultados limitados entre tres a seis meses. Los que proponen esta arquitectura, argumentan que luego de comenzar con un datamarts pequeño, otros marts pueden proliferar en otras líneas de negocio o departamentos que tengan necesidades, y que satisfaciendo las distintas necesidades divisionales, una organización puede construir su camino para el datawarehouse completo. En este caso de data marts múltiples también se tienen procesos de carga múltiples donde los datos son extraídos desde sistemas de producción quizás en forma redundante.

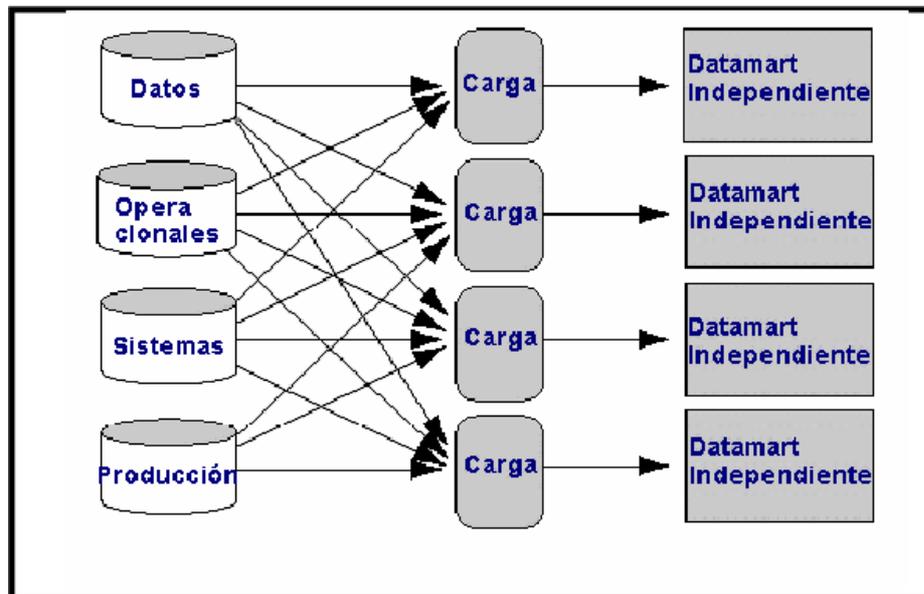


Figura II.6 - Arquitectura de Datamarts Independientes.

### 2.5.3 Procesamiento Analítico en Línea (OLAP)

La tecnología de Procesamiento Analítico en Línea, OLAP (*Online Analytical Processing*) permite un uso más eficaz de los datawarehouses para el análisis de datos en línea, lo que proporciona respuestas rápidas a consultas analíticas complejas e iterativas utilizada generalmente para sistemas de ayuda para la toma de decisiones.

1. OLAP presenta los datos a los usuarios a través de un modelo de datos intuitivo y natural. Con este estilo de navegación, los usuarios finales pueden

ver y entender más efectivamente la información de sus bases de datos, permitiendo así a las organizaciones reconocer mejor el valor de sus datos.

2. OLAP acelera la entrega de información a los usuarios finales que ven estas estructuras de datos como cubos denominadas multidimensionales debido a que la información es vista en varias dimensiones. Esta entrega es optimizada ya que se prepararan algunos valores calculados en los datos por adelantado, en vez de de realizar el cálculo al momento de la solicitud. La combinación de navegación fácil y rápida le permite a los usuarios ver y analizar información más rápida y eficientemente que lo que es posible con tecnologías de bases de datos relacionales solamente.
3. El resultado final: se pasa más tiempo analizando los datos y menos tiempo analizando las bases de datos.

#### **2.5.4 Procesamiento Transaccional en Línea (OTLP)**

Los sistemas OLTP son bases de datos orientadas al procesamiento de transacciones. Una transacción genera un proceso atómico (que debe ser validado con un *commit*, o invalidado con un *rollback*), y que puede involucrar operaciones de inserción, modificación y borrado de datos. El proceso transaccional es típico de las bases de datos operacionales.

- ✚ El acceso a los datos está optimizado para tareas frecuentes de lectura y escritura. (Por ejemplo, la enorme cantidad de transacciones que tienen que soportar las BD de bancos o hipermercados diariamente).
- ✚ Los datos se estructuran según el nivel aplicación (programa de gestión a medida, ERP o CRM implantado, sistema de información departamental...).
- ✚ Los formatos de los datos no son necesariamente uniformes en los diferentes departamentos (es común la falta de compatibilidad y la existencia de islas de datos).
- ✚ El historial de datos suele limitarse a los datos actuales o recientes.

## Las características principales del OLAP

- ✚ **Rápido:** proporciona la información al usuario a una velocidad constante. La mayoría de las peticiones se deben de responder al usuario en cinco segundos o menos.
- ✚ **Análisis:** realiza análisis estadísticos y numéricos básicos de los datos, predefinidos por el desarrollador de la aplicación o definido “ad hoc” por el usuario.
- ✚ **Compartida:** implementa los requerimientos de seguridad necesarios para compartir datos potencialmente confidenciales a través de una gran población de usuarios.

El OLAP es un componente clave en el proceso de almacenamiento de datos (data warehousing) y los servicios OLAP proporcionan la funcionalidad esencial para una gran variedad de aplicaciones que van desde reportes corporativos hasta soporte avanzado de decisiones.

- ✚ **Multidimensional:** llena la característica esencial del OLAP, que es ver la información en determinadas vistas o dimensiones.
- ✚ **Información:** acceden a todos los datos y a la información necesaria y relevante para la aplicación, donde sea que ésta resida y no esté limitada por el volumen.

## Diferencias entre Procesamiento Transaccional en Línea (OLTP) Y Procesamiento Analítico en Línea (OLAP).

La siguiente tabla muestra la diferencia que existe entre OLTP y OLAP

**Tabla II.1 - Diferencia entre sistemas OLTP y OLAP**

PROCESAMIENTO TRANSACCIONAL EN LÍNEA(OLTP)	PROCESOS DE ANÁLISIS EN LÍNEA(OLAP)
Admiten el acceso simultáneo de muchos usuarios -miles- que agregan y modifican datos.	Admiten el acceso simultáneo de muchos Usuarios -cientos- que consultan y no modifican datos.
Representan el estado, en cambio constante, de una organización, pero no guardan su historial.	Guardan el historial de una organización.
Contienen grandes cantidades de datos,	Contienen grandes cantidades de datos,

incluidos los datos extensivos utilizados para comprobar transacciones.	Sumarizados, consolidados y transformados. También de detalle pero solo los necesarios para el análisis.
Tienen estructuras de base de datos complejas.	Tienen estructuras de Base de datos simples.
Se ajustan para dar respuesta a la actividad transaccional.	Se ajustan para dar respuesta a la actividad de consultas.
Proporcionan la infraestructura tecnológica necesaria para admitir las operaciones diarias de la empresa.	Proporcionan la infraestructura tecnológica necesaria para admitir análisis de los datos de la empresa.
Los analistas carecen de la experiencia técnica necesaria para crear consultas "ad hoc" contra la compleja estructura de datos.	Pueden combinar datos de orígenes heterogéneos en una única estructura homogénea y simple, facilitando la creación de informes y consultas
Las consultas analíticas que resumen grandes volúmenes de datos afectan negativamente a la capacidad del sistema para responder a las Transacciones en línea.	Organizan los datos en estructuras simplificadas buscando la eficiencia de las consultas analíticas más que del proceso de transacciones.
El rendimiento del sistema cuando está respondiendo a consultas analíticas complejas puede ser lento o impredecible, lo que causa un servicio poco eficiente a los usuarios del proceso analítico en línea	Contienen datos transformados que son válidos, coherentes, consolidados y con el formato adecuado para realizar el análisis sin interferir en la operatoria transaccional diaria.
Los datos que se modifican con frecuencia interfieren en la coherencia de la información analítica.	Proporcionan datos estables que representan el historial de la empresa. Se actualizan periódicamente con datos adicionales, no con transacciones frecuentes.
La seguridad se complica cuando se combina el análisis en línea con el proceso de transacciones en línea.	Simplifican los requisitos de seguridad.

### 2.5.5 Cubo Multidimensional

Es una estructura de almacenamiento que permite realizar las diferentes y posibles combinaciones para visualizar los resultados de una organización hasta un determinado grado de detalle. Esta estructura es independiente del sistema

transaccional de la organización y facilita consultar información histórica de manera rápida y eficiente; ofreciendo la posibilidad de navegar y analizar los datos requeridos.

### 2.5.5.1 Estructura del Cubo

Está determinado por sus medidas y dimensiones:

#### **Medida:**

La medida o hecho es un dato numérico que representa una actividad específica de un negocio, estas medidas representan el patrón de interés o el evento dentro de una empresa que necesita ser analizado. Los hechos son implícitamente definidos por la combinación de valores de las dimensiones, existen tres tipos de hechos:

- ✚ **Evento:** Con la granularidad más fina, típicamente modela eventos del mundo real.
- ✚ **Fotos Fijas (snapshots):** Modelan entidades en un punto dado en el tiempo.
- ✚ **Fotos Fijas acumulativas:** Modelan actividades en un punto dado en el tiempo.

Una medida contiene una propiedad numérica y una fórmula. Existen tres clases de medidas:

- ✚ **Medidas aditivas:** Pueden ser combinadas a lo largo de cualquier dimensión.
- ✚ **Medidas semi-aditivas:** No pueden ser combinadas a lo largo de una o más dimensiones.
- ✚ **Medidas no aditivas:** No pueden ser combinadas a lo largo de ninguna dimensión.

#### **Dimensión:**

Las dimensiones son usadas para seleccionar y agregar datos a un cierto nivel deseado de detalle. Cada instancia de una dimensión o valor de una dimensión, corresponde a un nivel particular. Los datos dentro de un cubo tienen la ventaja de que se puede manejar cualquier número de dimensiones. Sin embargo usualmente un cubo tiene entre 4 y 12 dimensiones. Generalmente un cubo soporta una vista de dos o tres dimensiones simultáneamente.

### 2.5.5.2 Esquema del Cubo

Existen básicamente dos tipos de esquema de cubo: en estrella y en copo de nieve.

#### Esquema en Estrella (*star schema*)

Consiste en estructurar la información en procesos, vistas y métricas recordando a una estrella. Es decir, tendremos una visión multidimensional de un proceso que medimos a través de unas métricas. A nivel de diseño, consiste en una tabla de hechos (*fact table*) en el centro para el hecho objeto de análisis y una o varias tablas de dimensión (*dimension table*) por cada dimensión de análisis que participa de la descripción de ese hecho. En la tabla de hecho encontramos los atributos destinados a medir (cuantificar) el hecho: sus métricas. Mientras, en las tablas de dimensión, los atributos se destinan a elementos de nivel (que representan los distintos niveles de las jerarquías de dimensión) y a atributos de dimensión (encargados de la descripción de estos elementos de nivel). En el esquema en estrella la tabla de hechos es la única tabla del esquema que tiene múltiples joins que la conectan con otras tablas (*foreign keys* hacia otras tablas). El resto de tablas del esquema (tablas de dimensión) únicamente hacen join con esta tabla de hechos. Las tablas de dimensión se encuentran además totalmente denormalizadas, es decir, toda la información referente a una dimensión se almacena en la misma tabla.

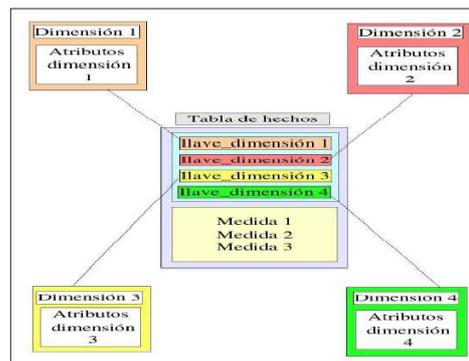


Figura II.7 - Esquema en Estrella

### Esquema en Copo de Nieve (*snowflake schema*)

Es un esquema de representación derivado del esquema en estrella, en el que las tablas de dimensión se normalizan en múltiples tablas. Por esta razón, la tabla de hechos deja de ser la única tabla del esquema que se relaciona con otras tablas, y aparecen nuevas joins gracias a que las dimensiones de análisis se representan ahora en tablas de dimensión normalizadas. En la estructura dimensional normalizada, la tabla que representa el nivel base de la dimensión es la que hace join directamente con la tabla de hechos. La diferencia entre ambos esquemas (star y snowflake) reside entonces en la estructura de las tablas de dimensión. Para conseguir un esquema en copo de nieve se ha de tomar un esquema en estrella y conservar la tabla de hechos, centrándose únicamente en el modelado de las tablas de dimensión, que si bien en el esquema en estrella se encontraban totalmente denormalizadas, ahora se dividen en subtablas tras un proceso de normalización. Es posible distinguir dos tipos de esquemas en copo de nieve, un snowflake completo (en el que todas las tablas de dimensión en el esquema en estrella aparecen ahora normalizadas en el snowflake) o un snowflake parcial (sólo se lleva a cabo la normalización de algunas de ellas).

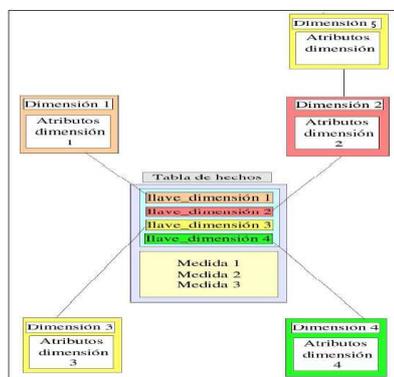


Figura II.8 – Esquema en copo de nieve

#### 2.5.5.3 Almacenamiento de Cubo

Todas ellas se basan en la tecnología OLAP (Procesamiento analítico en línea), que permite trabajar sobre los datos al pensarlos como hipercubos con información de la empresa. Existen diversas opciones, en las que se transa menor tiempo de acceso a cambio de mayor espacio utilizado en disco, viceversa.

Existen tres modos posibles:

**MOLAP:**

Los datos se almacenan en forma multidimensional, permitiendo operaciones rápidas de búsqueda y resúmenes, gracias a los datos precalculados existentes. Por otra parte posee grandes tiempos al agregar datos nuevos, así como un alto requerimiento de espacio en disco. Consta de dos niveles: la base de datos multidimensional (MDDDB), que se preocupa de manejar, acceder y obtener los datos, y el motor analítico, que ejecuta las consultas de los usuarios.

Existe una integración entre la presentación y la aplicación. Presenta dos características distintivas: la consolidación previa de datos (unir resultados parciales ya calculados para encontrar el resultado final), y la dispersión de los datos (cuando algunas celdas del cubo se encuentran vacías).

**ROLAP:**

Los datos se encuentran almacenados en tablas de bases de datos relacionales. Es el OLAP más rápido en la inserción de datos, y el de menor uso de espacio en disco. Por otra parte, resultan comparativamente más lentos que los otros modos.

La implementación se realiza mediante un esquema de tres niveles, la base de datos relacional almacena los datos, el motor OLAP permite el funcionamiento analítico (ejecuta las consultas de los usuarios), y una tercera aplicación presenta los informes.

**HOLAP:**

Corresponde a la implementación híbrida, que trata de obtener lo mejor de los dos modos anteriores. El sistema se comporta como un MOLAP para la información resumida (y constantemente utilizada), y como un ROLAP al solicitar un detalle de ésta. Los recursos promedio utilizados se ubican entre las dos opciones anteriores, al igual que el rendimiento.

## 2.6 PRODUCTOS DE BUSINESS INTELLIGENCE

Los principales productos de Business Intelligence que existen hoy en día son:

- 📊 Cuadros de Mando Integrales (CMI)
- 📊 Sistemas de Soporte a la Decisión (DSS)
- 📊 Sistemas de Información Ejecutiva (EIS)

### 2.6.1 Cuadros de Mando Integrales (CMI)

El Cuadro de Mando Integral (CMI), también conocido como Balanced Scorecard (BSC) o dashboard, es una herramienta de control empresarial que permite establecer y monitorizar los objetivos de una empresa y de sus diferentes áreas o unidades.

También se puede considerar como una aplicación que ayuda a una compañía a expresar los objetivos e iniciativas necesarias para cumplir con su estrategia, mostrando de forma continuada cuándo la empresa y los empleados alcanzan los resultados definidos en su plan estratégico.

#### 2.6.1.1 Diferencia con otras herramientas de Business Intelligence

El Cuadro de Mando Integral se diferencia de otras herramientas de Business Intelligence, como los Sistemas de Soporte a la Decisión (DSS) o los Sistemas de Información Ejecutiva (EIS), en que está más orientado al seguimiento de indicadores que al análisis minucioso de información.

Por otro lado, es muy común que un CMI sea controlado por la dirección general de una compañía, frente a otras herramientas de Business Intelligence más enfocadas a la dirección departamental.

El CMI requiere, por tanto, que los directivos analicen el mercado y la estrategia para construir un modelo de negocio que refleje las interrelaciones entre los diferentes componentes de la empresa.

Una vez que lo han construido, los responsables de la organización utilizan este modelo como mapa para seleccionar los indicadores del CMI



Figura II.9 - Diferencia con otras herramientas de Business Intelligence

### 2.6.1.2 Tipos de Cuadros de Mando

El **Cuadro de Mando Operativo (CMO)**, es una herramienta de control enfocada al seguimiento de variables operativas, es decir, variables pertenecientes a áreas o departamentos específicos de la empresa. La periodicidad de los CMO puede ser diaria, semanal o mensual, y está centrada en indicadores que generalmente representan procesos, por lo que su implantación y puesta en marcha es más sencilla y rápida. Un CMO debería estar siempre ligado a un DSS (Sistema de Soporte a Decisiones) para indagar en profundidad sobre los datos.

El **Cuadro de Mando Integral (CMI)**, por el contrario, representa la ejecución de la estrategia de una compañía desde el punto de vista de la Dirección General (lo que hace que ésta deba estar plenamente involucrada en todas sus fases, desde la definición a la implantación). Existen diferentes tipos de cuadros de mando integral que son utilizados tanto indicadores financieros como no financieros, y que los objetivos estratégicos se organizan en cuatro áreas o perspectivas: financiera, cliente, interna y aprendizaje/crecimiento.

- ✚ La **perspectiva financiera** incorpora la visión de los accionistas y mide la creación de valor de la empresa. Responde a la pregunta: ¿Qué indicadores tienen que ir bien para que los esfuerzos de la empresa realmente se transformen en valor? Esta perspectiva valora uno de los objetivos más relevantes de organizaciones con ánimo de lucro, que es, precisamente, crear valor para la sociedad.

- ✚ La **perspectiva del cliente** refleja el posicionamiento de la empresa en el mercado o, más concretamente, en los segmentos de mercado donde quiere competir. Por ejemplo, si una empresa sigue una estrategia de costes es muy posible que la clave de su éxito dependa de una cuota de mercado alta y unos precios más bajos que la competencia. Dos indicadores que reflejan este posicionamiento son la cuota de mercado y un índice que compare los precios de la empresa con los de la competencia.
- ✚ La **perspectiva interna** recoge indicadores de procesos internos que son críticos para el posicionamiento en el mercado y para llevar la estrategia a buen puerto. En el caso de la empresa que compite en coste, posiblemente los indicadores de productividad, calidad e innovación de procesos sean importantes. El éxito en estas dimensiones no sólo afecta a la perspectiva interna, sino también a la financiera, por el impacto que tienen sobre las rúbricas de gasto.
- ✚ La perspectiva de **aprendizaje y crecimiento** es la última que se plantea en este modelo de CMI. Para cualquier estrategia, los recursos materiales y las personas son la clave del éxito. Pero sin un modelo de negocio apropiado, muchas veces es difícil apreciar la importancia de invertir, y en épocas de crisis lo primero que se recorta es precisamente la fuente primaria de creación de valor: se recortan inversiones en la mejora y el desarrollo de los recursos



Figura II.10 - Cuadro de Mando Integral

### **2.6.1.3 Beneficios de la implantación de un Cuadro de Mando Integral**

- ✚ La fuerza de explicitar un modelo de negocio y traducirlo en indicadores facilita el consenso en toda la empresa, no sólo de la dirección, sino también de cómo alcanzarlo.
- ✚ Clarifica cómo las acciones del día a día afectan no sólo al corto plazo, sino también al largo plazo.
- ✚ Una vez el CMI está en marcha, se puede utilizar para comunicar los planes de la empresa, a unir los esfuerzos en una sola dirección y evitar la dispersión. En este caso, el CMI actúa como un sistema de control por excepción.
- ✚ Permita detectar de forma automática desviaciones en el plan estratégico u operativo, e incluso indagar en los datos operativos de la compañía hasta descubrir la causa original que dio lugar a esas desviaciones.

### **2.6.1.4 Riesgos de la implantación de un Cuadro de Mando Integral**

- ✚ Un modelo poco elaborado y sin la colaboración de la dirección es papel mojado, y el esfuerzo será en vano.
- ✚ Si los indicadores no se escogen con cuidado, el CMI pierde una buena parte de sus virtudes, porque no comunica el mensaje que se quiere transmitir.
- ✚ Cuando la estrategia de la empresa está todavía en evolución, es contraproducente que el CMI se utilice como un sistema de control clásico y por excepción, en lugar de usarlo como una herramienta de aprendizaje.
- ✚ Existe el riesgo de que lo mejor sea enemigo de lo bueno, de que el CMI sea perfecto, pero desfasado e inútil.

### **2.6.2 Sistemas de soporte a la Decisión (DSS)**

Un Sistema de Soporte a la Decisión (DSS) es una herramienta de Business Intelligence enfocada al análisis de los datos de una organización.

En principio, puede parecer que el análisis de datos es un proceso sencillo, y fácil de conseguir mediante una aplicación hecha a medida o un ERP sofisticado.

Sin embargo, no es así: estas aplicaciones suelen disponer de una serie de informes predefinidos en los que presentan la información de manera estática, pero no permiten profundizar en los datos, navegar entre ellos, manejarlos desde distintas perspectivas.



Figura II.11 - Sistema de Soporte a la Decisión

El DSS es una de las herramientas más emblemáticas del Business Intelligence ya que, entre otras propiedades, permiten resolver gran parte de las limitaciones de los programas de gestión. Estas son algunas de sus características principales:

- ✚ **Informes dinámicos, flexibles e interactivos**, de manera que el usuario no tenga que ceñirse a los listados predefinidos que se configuraron en el momento de la implantación, y que no siempre responden a sus dudas reales.
- ✚ **No requiere conocimientos técnicos**. Un usuario no técnico puede crear nuevos gráficos e informes y navegar entre ellos, haciendo *drag&drop* o *drill through*. Por tanto, para examinar la información disponible o crear nuevas métricas no es imprescindible buscar auxilio en el departamento de informática.
- ✚ **Rapidez en el tiempo de respuesta**, ya que la base de datos subyacente suele ser un datawarehouse corporativo o un datamart, con modelos de datos en estrella o copo de nieve. Este tipo de bases de datos están optimizadas para el análisis de grandes volúmenes de información (análisis OLTP-OLAP).
- ✚ **Integración entre todos los sistemas/departamentos de la compañía**. El proceso de ETL previo a la implantación de un Sistema de Soporte a la Decisión garantiza la calidad y la integración de los datos entre las diferentes unidades de la empresa. Existe lo que se llama: *integridad referencial absoluta*.
- ✚ **Cada usuario dispone de información adecuada a su perfil**. No se trata de que todo el mundo tenga acceso a toda la información, sino de que tenga

acceso a la información que necesita para que su trabajo sea lo más eficiente posible.

- ✚ **Disponibilidad de información histórica.** En estos sistemas está a la orden del día comparar los datos actuales con información de otros períodos históricos de la compañía, con el fin de analizar tendencias, fijar la evolución de parámetros de negocio.

### 2.6.2.1 Diferencia con otras herramientas de Business Intelligence

El principal objetivo de los Sistemas de Soporte a Decisiones es, a diferencia de otras herramientas como los Cuadros de Mando (CMI) o los Sistemas de Información Ejecutiva (EIS), explotar al máximo la información residente en una base de datos corporativa (datawarehouse o datamart), mostrando informes muy dinámicos y con gran potencial de navegación, pero siempre con una interfaz gráfica amigable, vistosa y sencilla.

Otra diferencia fundamental radica en los usuarios a los que están destinadas las plataformas DSS: cualquier nivel gerencial dentro de una organización, tanto para situaciones estructuradas como no estructuradas. (En este sentido, por ejemplo, los CMI están más orientados a la alta dirección).

Por último, destacar que los DSS suelen requerir (aunque no es imprescindible) un motor OLAP subyacente, que facilite el análisis casi ilimitado de los datos para hallar las causas raíces de los problemas/pormenores de la compañía.

### 2.6.2.2 Tipos de Sistemas de Soporte a Decisiones

#### ✚ **Sistemas de información gerencial (MIS)**

Los sistemas de información gerencial (MIS, *Management Information Systems*), también llamados Sistemas de Información Administrativa (AIS) dan soporte a un espectro más amplio de tareas organizacionales, encontrándose a medio camino entre un DSS tradicional y una aplicación CRM/ERP implantada en la misma compañía.

### **Sistemas de información ejecutiva (EIS)**

Los sistemas de información ejecutiva (EIS, *Executive Information System*) son el tipo de DSS que más se suele emplear en Business Intelligence, ya que proveen a los gerentes de un acceso sencillo a información interna y externa de su compañía, y que es relevante para sus factores clave de éxito.

### **Sistemas expertos basados en inteligencia artificial (SSEE)**

Los sistemas expertos, también llamados sistemas basados en conocimiento, utilizan redes neuronales para simular el conocimiento de un experto y utilizarlo de forma efectiva para resolver un problema concreto. Este concepto está muy relacionado con el datamining.

### **Sistemas de apoyo a decisiones de grupo (GDSS)**

Un sistema de apoyo a decisiones en grupos (GDSS, *Group Decision Support Systems*) es "un sistema basado en computadoras que apoya a grupos de personas que tienen una tarea (u objetivo) común, y que sirve como interfaz con un entorno compartido".

El supuesto en que se basa el GDSS es que si se mejoran las comunicaciones se pueden mejorar las decisiones.

## **2.6.3 Sistemas de Información Ejecutiva (EIS)**

Un Sistema de Información para Ejecutivos o Sistema de Información Ejecutiva es una herramienta software, basada en un DSS, que provee a los gerentes de un acceso sencillo a información interna y externa de su compañía, y que es relevante para sus factores clave de éxito.

La finalidad principal es que el ejecutivo tenga a su disposición un panorama completo del estado de los indicadores de negocio que le afectan al instante, manteniendo también la posibilidad de analizar con detalle aquellos que no estén cumpliendo con las expectativas establecidas, para determinar el plan de acción más adecuado.



Figura II.12 - Sistemas de Información Ejecutiva

De forma más pragmática, se puede definir un EIS como una aplicación informática que muestra informes y listados (*query & reporting*) de las diferentes áreas de negocio, de forma consolidada, para facilitar la monitorización de la empresa o de una unidad de la misma.

El EIS se caracteriza por ofrecer al ejecutivo un acceso rápido y efectivo a la información compartida, utilizando interfaces gráficas visuales e intuitivas. Suele incluir alertas e informes basados en excepción, así como históricos y análisis de tendencias. También es frecuente que permita la domiciliación por correo de los informes más relevantes.

A través de esta solución se puede contar con un resumen del comportamiento de una organización o área específica, y poder compararla a través del tiempo.

## CAPÍTULO III

### ANÁLISIS DE LAS METODOLOGÍAS DE BUSINESS INTELLIGENCE.

#### 3.1 INTRODUCCIÓN

El término Data Warehouse fue acuñado por Bill Inmon a principios de la década de los 90 y lo definió de la siguiente manera (dada la popularidad e importancia de esta definición preferimos no traducirla completamente): “Un warehouse es una colección de datos **subject-oriented**, **integrated**, **time-variant** y **non-volatile** para ayudar al proceso de toma de decisiones gerenciales<sup>1</sup>.

Analicemos cada uno de estos términos al igual que lo hace el autor.

- ✚ **Subject-Oriented:** datos que brindan información sobre un “sujeto” del negocio en particular, en lugar de concentrarse en la dinámica de las transacciones de la organización.
- ✚ **Integrated:** los datos con los que se nutre el data warehouse vienen de diferentes fuentes y son integrados para dar una visión de un “todo” coherente.

---

<sup>1</sup> William H. Inmon, 1992. “Building the Data Warehouse”, Wiley-QED John Wiley & Sons, Inc...

- ✚ **Time-variant:** todos los datos en el data warehouse son asociados con un período de tiempo específico.
- ✚ **Non-Volatile:** los datos son estables en el data warehouse. Mas datos son agregados pero los datos existentes no son removidos.

Obviamente que esta definición, ya clásica, debe tomarse como la definición “pura” sobre data warehouse. Después de diez años, sin embargo, algunos términos han sido manejados según las necesidades y capacidades del mercado, dando origen a conceptos como el de Data Mart (para referirse a data warehouse sobre áreas específicas en lugar del warehouse corporativo) o data warehouse volátiles, que ante la imposibilidad de almacenar toda la información histórica, almacenan una foto sobre determinado períodos, etc.

Ralph Kimball define data warehouse de una forma más sencilla y práctica pero igual de importante, un data warehouse es “una copia de los datos transaccionales específicamente estructurados para consultas y análisis”<sup>2</sup>.

Por nuestro lado, podemos decir que un Data Warehouse es una base de datos orientada al análisis de la información histórica contenida en ella. Dependiendo las necesidades de análisis de la organización puede almacenarse desde unos meses hasta varios años de información.

El modelo que soporta la información que contiene se encuentra diseñado, estructurado e implementado con la finalidad y propósito del análisis y navegación de los datos.

Se entiende por navegación o drilling de los datos, la posibilidad de ver información correspondiente a diferentes contextos o entornos, por ejemplo, analizar las ventas anuales y poder “abrir las” por sucursal, después analizar en más en detalle una sucursal para ver cómo se discriminan las ventas por cada producto, etc.

---

<sup>2</sup> Ralph Kimball, 1992. The Data Warehouse Toolkit, Wiley Computer Publishing

Según la implementación seleccionada, los datos son almacenados en forma relacional (RDBMS, respetando ciertos estándares a nivel de definición y consulta de datos) o en formato multidimensional (las bases de datos multidimensionales son arquitecturas propietarias definidas por cada proveedor que son frecuentemente actualizadas desde base de datos relacionales).

Típicamente los usuarios tienen sólo permisos de lectura sobre el warehouse (read-only).

Comúnmente se dice que los data warehouse son fuentes secundarias de información pues no generan información por si mismos, sino que son actualizados desde sistemas fuentes existentes internamente en la organización (sistema de ventas, sistema presupuestario, etc.) o sistemas externos de información (datos meteorológicos, información de la competencia, cotizaciones de la bolsa, etc.).

Por todo esto los Data Warehouses son identificados como ambientes OLAP, On-Line Analytical Processing, en contraposición a los ambientes transaccionales clásicos (OLTP, On-Line Transaction Processing).

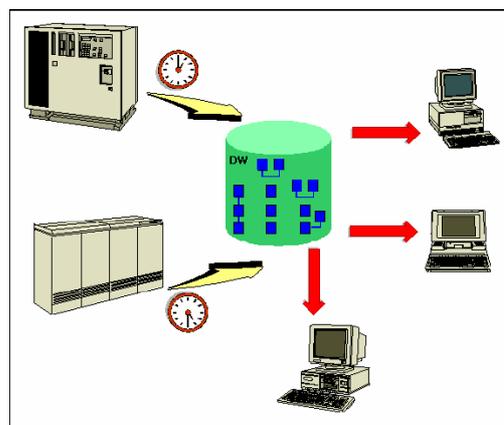
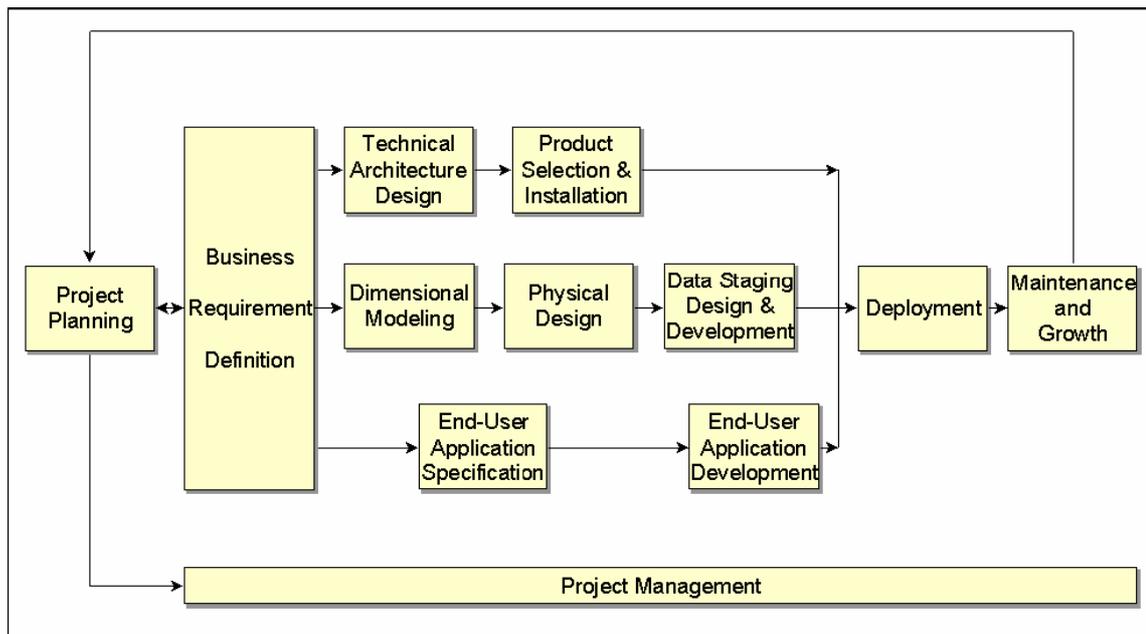


Figura III.13 - ¿Qué es un Data Warehouse?

### 3.2 METODOLOGÍA RALPH KIMBALL

Tomaremos como base el ciclo de vida de los Data Warehouses definido por Ralph Kimball. El marco presentado por Ralph Kimball con el nombre de Business Dimensional Lifecycle (El Ciclo de Vida Dimensional del Negocio (BDL))

(BDL<sup>3</sup>) ilustra las diferentes etapas por las que debe pasar todo proceso de Data Warehousing. Este enfoque de implementación de data warehouses es ilustrado en la siguiente figura. Este diagrama ilustra la secuencialidad de tareas de alto nivel requeridas para el efectivo diseño, desarrollo e implementación de data warehouses. El diagrama muestra una vista general del mapa de ruta de un proyecto en el cual cada rectángulo es un mojón que nos indica dónde estamos parados, por dónde pasamos y hacia dónde debemos dirigirnos.



**Figura III.14 - Business Dimensional Lifecycle propuesto por Ralph Kimball**

Es importante aclarar, como lo hacen los autores, que el BDL no intenta reflejar un proyecto en término de tiempos y plazos. Como se puede notar cada rectángulo del diagrama tiene el mismo ancho, con la excepción del gerenciamiento del proyecto. Cualquiera que haya pasado por algún proyecto de Data Warehousing sabe que la magnitud de recursos y tiempo requerido para cada rectángulo del ciclo de vida no es igual. El BDL se focaliza en secuencialidad y concurrencia no en tiempos y plazos. A continuación pasaremos de describir cada una de las etapas del BDL.

<sup>3</sup> BDL: Ciclo de vida dimensional del negocio

### 3.2.1. Planificación del Proyecto

La planificación busca identificar la definición y el alcance del proyecto de data warehouse, incluyendo justificaciones del negocio y evaluaciones de factibilidad.

La planificación del proyecto se focaliza sobre recursos, perfiles, tareas, duraciones y secuencialidad. El plan de proyecto resultante identifica todas las tareas asociadas con el BDL e identifica las partes involucradas.

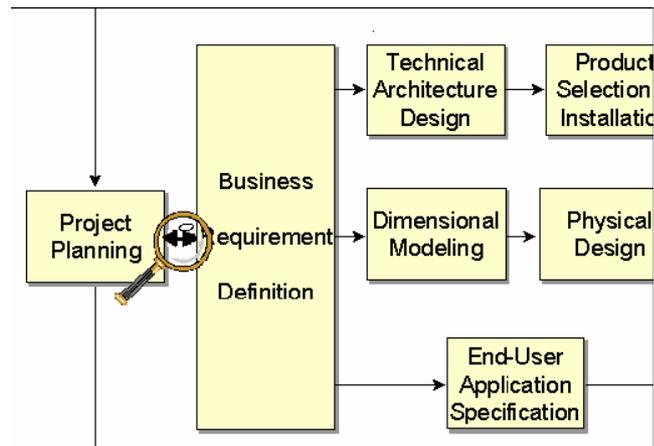


Figura III.15 - Iteración entre Planificación y Requerimientos

La planificación del proyecto es dependiente de los requerimientos del negocio, como es denotado en el diagrama del BDL, ya que los requerimientos del negocio determinan el alcance del proyecto, definen los recursos necesarios, etc y la planificación acotará los requerimientos ya sea por cuestiones de recursos y/o tiempo.

Esta etapa se concentra sobre la definición del proyecto (identificación del escenario del proyecto para saber de dónde surge la necesidad del data warehouse).

Según sentencia Kimball, antes de comenzar un proyecto de data warehouse o data mart, hay que estar seguro si existe la demanda y de dónde proviene. Si no se tiene un sólido usuario sponsor y no hay usuarios entusiasmados, posponga el proyecto.

Factores asociados con estas etapas incluyen: identificación de los usuarios sponsors, convincentes motivaciones del negocio, cooperación entre áreas de sistemas y negocios, cultura analítica de la organización y análisis de factibilidad (tanto tecnológica como de disponibilidad de datos). Para medir estos factores propone un test de buena disposición del proyecto dónde describe diferentes escenarios posibles.

Adicionalmente propone técnicas (relevamientos de alto nivel, priorización de requerimientos y pruebas de concepto) para mitigar las deficiencias que el proyecto pudiera tener en algunos de los factores mencionados anteriormente.

Cómo metodología de estas etapas propone identificar el alcance preliminar basándose en los requerimientos del negocio y no en fechas límites (deadlines) construyendo la justificación del proyecto en términos del negocio con indicadores como el **ROI** (retorno de inversión), **NPV** (valor presente neto) y el **IRR** (índice de retorno interno).

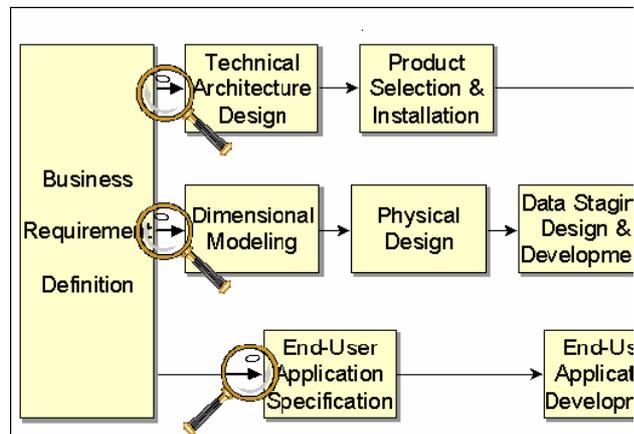
A nivel de planificación del proyecto, establece la identidad del mismo, el personal (staff): los usuarios sponsors, líderes, gerentes del proyecto (tanto de sistemas como del sector usuarios), equipo “corazón” del proyecto (analistas, arquitectos, DBAs, diseñadores, responsables de extracción, desarrolladores, instructores, etc.), equipo “especial” del proyecto (soporte, seguridad informática, programadores, analistas QA), el desarrollo del plan del proyecto, el seguimiento y monitoreo.

### **3.2.2 Definición de los Requerimientos del Negocio**

Un factor determinante en el éxito de un proceso de Data Warehousing es la interpretación correcta de los diferentes niveles de requerimientos expresados por los diferentes niveles de usuarios.

La técnica utilizada para relevar los requerimientos de los analistas del negocio difiere de los enfoques tradicionales guiados por los datos.

Los diseñadores de los data warehouses deben entender los factores claves que guían al negocio para determinar efectivamente los requerimientos y traducirlos en consideraciones de diseño apropiadas.



**Figura III.16 - Iteración entre Requerimientos del Negocio y etapas subsiguientes**

La definición de los requerimientos del negocio establece la base para las tres etapas paralelas subsiguientes focalizadas en la tecnología, los datos y las aplicaciones por lo cual es altamente crítica y es el centro de atención del BDL.

Los usuarios finales y sus requerimientos impactan siempre en las implementaciones realizadas de un data warehouse. Según la perspectiva de Kimball, los requerimientos del negocio se posicionan en el centro del “universo del data warehouse”. Como destaca siempre el autor, los requerimientos del negocio deben determinar el alcance del data warehouse (qué datos debe contener, cómo debe estar organizado, cada cuánto debe actualizarse, quiénes y desde dónde accederán, etc).

Kimball da consejos y técnicas para descubrir eficazmente los requerimientos del negocio.

Estas tácticas y estrategias se focalizan sobre las entrevistas de relevamiento (diferentes tipos, preparación de la entrevista, roles a cubrir, búsqueda de información pre- entrevista, selección de entrevistados, desarrollo de los cuestionarios, planificación, preparación de los entrevistados, conducción de la entrevista, contenido, cierre, revisión de resultados, etc.).

### 3.2.3 Modelado Dimensional

La definición de los requerimientos del negocio determina los datos necesarios para cumplir los requerimientos analíticos de los usuarios. Diseñar los modelos de datos para soportar estos análisis requiere un enfoque diferente al usado en los sistemas operacionales. Básicamente se comienza con una matriz donde se determina la dimensionalidad de cada indicador y luego se especifican los diferentes grados de detalle (atributos) dentro de cada concepto del negocio (dimensión), como así también la granularidad de cada indicador (variable o métrica) y las diferentes jerarquías que dan forma al modelo dimensional del negocio (BDM)<sup>4</sup> o mapa dimensional.

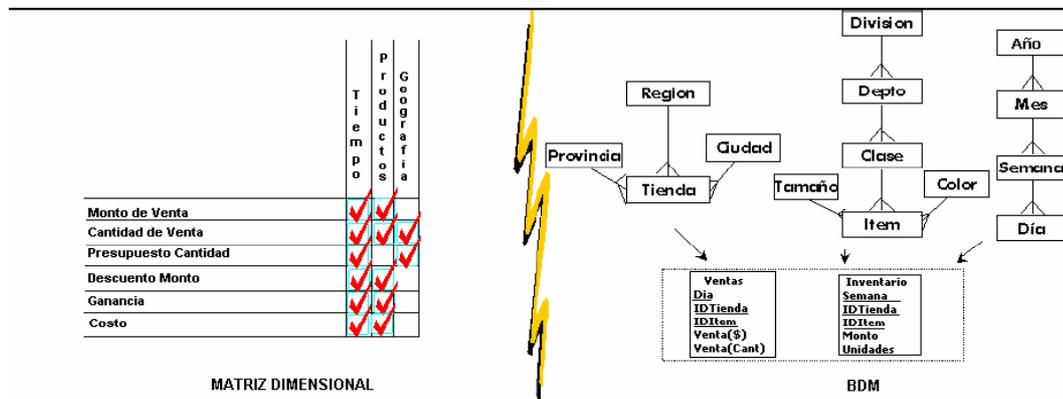


Figura III.17 - Diagramas para Modelado Dimensional

“Matriz Dimensional que determina la dimensionalidad de cada indicador y modelo dimensional del negocio (BDM) que ilustra las diferentes jerarquías dentro de cada dimensión.”

Ralph Kimball es realmente un referente en el tema de modelado dimensional como lo demuestran sus numerosos papers y libros publicados al respecto. Gran parte de las técnicas básicas en este tema (identificación de dimensiones, atributos, star schemas, snowflake schemas, soluciones verticales, etc.) son tratadas e introducidas constituyendo la base de la teoría sobre modelado dimensional.

<sup>4</sup> BDM: Modelo dimensional del negocio

Comienza fijando los pre-requisitos para realizar el **BDM**, después identifica los correspondientes Data Marts (como conjunto de indicadores), las correspondientes dimensiones, luego sigue con la construcción de una matriz dimensional de alto nivel (Data Warehouse Bus Architecture Matrix), donde las filas son los Data Marts y las columnas son todas las dimensiones marcando la intersección de aquellas dimensiones que están en juego en determinados Data Marts. La metodología continúa con el diseño de cada Data Mart, teniendo como primer paso la elección del Data Mart, segundo la declaración de la granularidad, tercero la elección de las dimensiones y cuarto la elección de las facts o indicadores. Además, Kimball recomienda un conjunto de gráficos y diagramas que ayudarán a esta etapa del proyecto (Data Warehouse Bus Architecture Matrix, diagrama de Fact Table, detalle de Fact Table y diagrama de Dimensiones). Finalmente, Kimball propone una serie de pasos y pautas para el equipo de trabajo asignado al modelado dimensional que ayuda a identificar y resolver futuros problemas de forma temprana, como así también el análisis de alto nivel de los sistemas fuentes como para mejorar las estimaciones y alcances del modelo realizado.

### 3.2.4 Diseño Físico

El diseño físico de las base de datos se focaliza sobre la selección de las estructuras necesarias para soportar el diseño lógico. Algunos de los elementos principales de este proceso son la definición de convenciones estándares de nombres y seteos específicos del ambiente de la base de datos. La indexación y las estrategias de particionamiento son también determinadas en esta etapa.

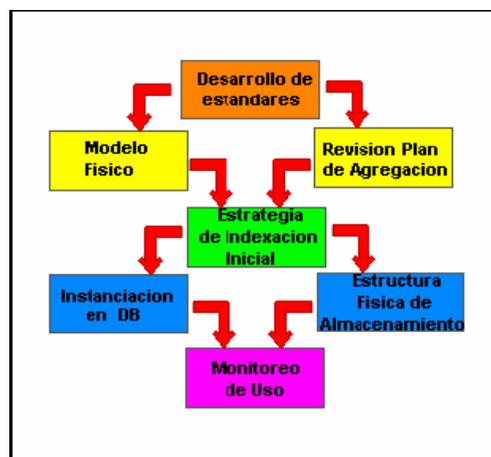


Figura III.18 - Proceso de diseño físico a alto nivel

### 3.2.5 Diseño y Desarrollo de Presentación de Datos

Esta etapa es típicamente la más subestimada de las tareas en un proyecto de data warehouse.

Las principales subetapas de esta zona del ciclo de vida son: la extracción, la transformación y la carga (ETL process). Se definen como procesos de extracción a aquellos requeridos para obtener los datos que permitirán efectuar la carga del Modelo Físico acordado. Así mismo, se definen como procesos de transformación los procesos para convertir o recodificar los datos fuente a fin poder efectuar la carga efectiva del Modelo Físico. Por otra parte, los procesos de carga de datos son los procesos requeridos para poblar el Data Warehouse.

Todas estas tareas son altamente críticas pues tienen que ver con la materia prima de la data warehouse: los datos. La desconfianza y pérdida de credibilidad del data warehouse serán resultados inmediatos e inevitables si el usuario choca con información inconsistente. Es por ello que la calidad de los datos es un factor determinante en el éxito de un proyecto de data warehousing. Es en esta etapa donde deben sanearse todos los inconvenientes relacionados con la calidad de los datos fuente.

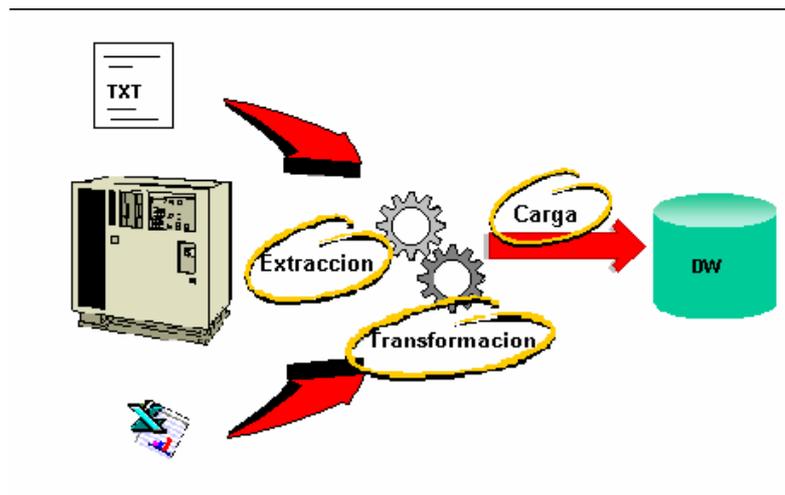


Figura III.19 - Proceso ETL

Como advierte Kimball, el proceso de Data Staging es el iceberg de un proyecto de Datawarehousing. Son muchos los desafíos que deben enfrentarse para lograr datos de alta calidad de los sistemas fuentes. Como comentamos anteriormente, en general es una de las etapas más subestimadas, que siempre termina tomando más tiempo del previsto. Ralph Kimball propone entonces un plan de 10 items que ayudarán a guiar esta etapa del BDL.

**Plan:**

1. Crear un diagrama de flujo fuente-destino esquemática, de una página y de muy alto nivel.
2. Probar, elegir e implementar una herramienta de data staging.
3. Profundizar en detalle por tabla destino, gráficamente describir las reestructuraciones transformaciones complejas. Gráficamente ilustrar la generación de las claves subrogadas.

Desarrollo preliminar de la secuencialidad de los trabajos.

**Carga de dimensiones:**

4. Construir y probar la carga de una tabla dimensional estática. La principal meta de este paso es resolver los problemas de infraestructura que pudieran surgir (conectividad, transferencia, seguridad, etc.)
5. Construir y probar los procesos de actualización de una dimensión.
6. Construir y probar las cargas de las restantes dimensiones.

**Fact Tables y automatización:**

7. Construir y probar la carga histórica de las fact tables (carga masiva de datos). Incluyendo búsqueda y sustitución de claves.
8. Construir y probar los procesos de cargas incrementales.
9. Construir y probar la generación de agregaciones.
10. Diseñar, construir y probar la automatización de los procesos.

### **3.2.6 Diseño de la Arquitectura Técnica**

Los ambientes de data warehousing requieren la integración de numerosas tecnologías. Se debe tener en cuenta tres factores:

los requerimientos del negocio, los actuales ambientes técnicos y las directrices técnicas estratégicas futuras planificadas para de esta forma poder establecer el diseño de la arquitectura técnica del ambiente de data warehousing.

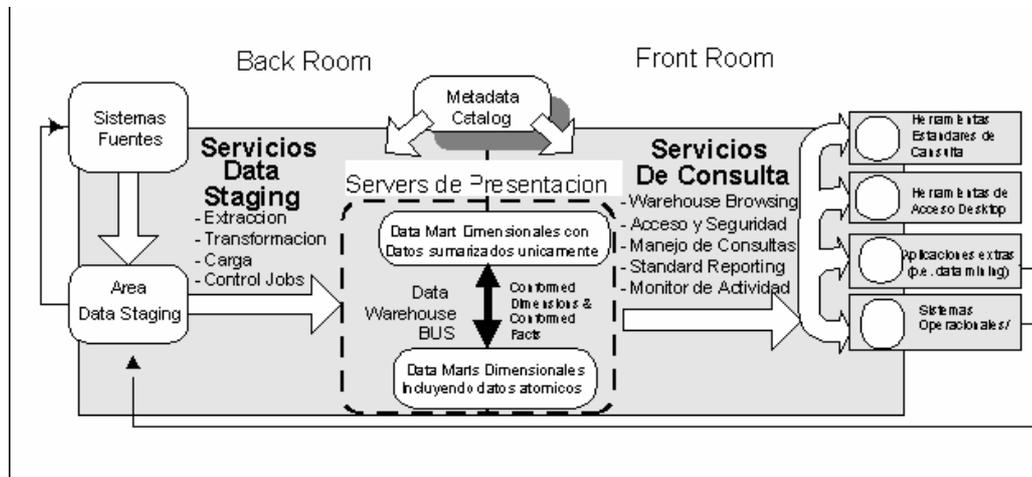


Figura III.20 - Modelo de Arquitectura Técnica a alto nivel

Ralph Kimball hace una analogía entre los planos arquitectónicos de una casa y la arquitectura de un warehouse, nadie comenzaría diciendo, "Hey, tengo algo de madera y algo de concreto, vamos a construir una casa". Hay que tener un **plan** antes de comenzar, no es simplemente reordenar y explotar la información. Al igual que en una construcción, los planos sirven para comunicar los deseos entre los clientes y el arquitecto, como así también para medir esfuerzos y materiales necesarios para la obra (comunicación, planificación, flexibilidad y mantenimiento, documentación, productividad y reuso). Finalmente, argumenta Kimball, "un buen conjunto de planos, como cualquier buena documentación, nos ayudará más tarde cuando sea tiempo de remodelar o hacer incorporaciones"

Una de las ventajas que trae la utilización de una arquitectura BUS, como la propuesta por el autor, es la fácil incorporación de nuevas instancias a la arquitectura (plug and play). Entre los muchos ensayos y enfoques existentes para la definición y desarrollo de arquitecturas de sistemas, Kimball se basa en el Framework propuesto por John A. Zachman de Zachman International y lo adapta a un ambiente de datawarehousing, simplificándolo considerablemente ya que en este tipo de ambientes, en teoría, no debe ser atacado los problemas de infraestructura de los procesos transaccionales. El Data Warehouse Architecture Framework propuesto por Kimball incluye tres áreas

(columnas): el área de arquitectura de datos (qué), el área de arquitectura técnica (cómo) y el área de infraestructura (dónde). A su vez, cada una de estas áreas tienen diferentes niveles de detalles (filas): Nivel de Requerimientos del Negocio, Nivel de Modelo Arquitectónico, Nivel de Modelo Detallado, Nivel de Implementación.

Como podemos inferir de las identificaciones de los diferentes niveles, Kimball recomienda un enfoque top-down, comenzando con una visión global y dividiendo la arquitectura en pequeñas piezas hasta llegar al grado donde las piezas pueden ser realmente implementadas.

Como siempre el énfasis está puesto en los requerimientos del negocio, son nuestra guía primaria para desarrollar nuestra arquitectura y para priorizar. La arquitectura técnica se divide en dos partes, el back room (la parte interna del warehouse) y el front room (la cara pública del warehouse), interactuando constantemente. Mientras los requerimientos del negocio nos dicen qué necesitamos hacer, la arquitectura técnica nos responde el interrogante de cómo lo haremos.

### 3.2.7 Selección de Productos e Instalación

Utilizando el diseño de arquitectura técnica como marco, es necesario evaluar y seleccionar componentes específicos de la arquitectura como ser la plataforma de hardware, el motor de base de datos, la herramienta de ETL o el desarrollo pertinente, herramientas de acceso, etc.

Una vez evaluados y seleccionados los componentes determinados se procede con la instalación y prueba de los mismos en un ambiente integrado de data warehousing.

Característica	Peso	Producto A	Producto B	Producto C	...
<b>Capacidades Basicas de ETL</b>					
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	85				
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	70				
<b>Job Control &amp; Scheduling</b>					
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	90				
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	85				
<b>Metada &amp; Estandares</b>					
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx					
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx					
<b>Info de Vendedor</b>	70				
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	85				
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx					
<b>Puntaje Total</b>					
<b>Ranking</b>					

Figura III.21 - Ejemplo de matriz de evaluación de productos

### 3.2.8 Especificación de Aplicaciones para Usuarios Finales

No todos los usuarios del warehouse necesitan el mismo nivel de análisis. Es por ello que en esta etapa se identifican los diferentes roles o perfiles de usuarios para determinar los diferentes tipos de aplicaciones necesarias en base al alcance de los diferentes perfiles (gerencial, analista del negocio, vendedor, etc.)

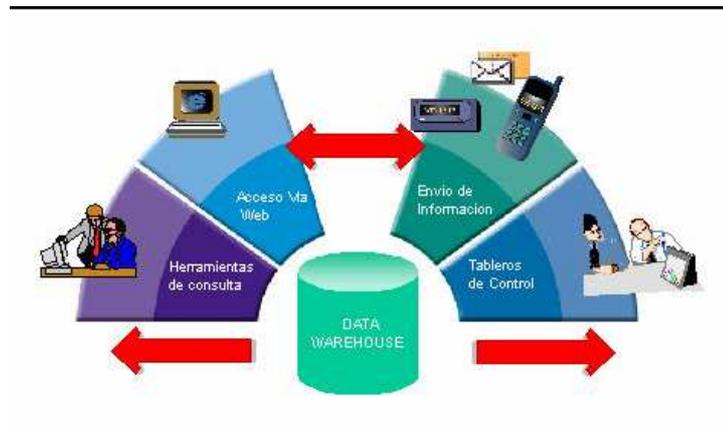


Figura III.22 - Variedad de interfaces por perfil

Los diferentes roles o perfiles de usuarios determinan la interfase o ventana al warehouse. Herramientas de diseño de reportes y consultas avanzadas para analistas, tableros de control para gerentes, acceso mediante inter/intra net para usuarios internos/externos remotos, envío de información por dispositivos no estándares para usuarios internos/externos, etc.

Kimball se concentra sobre el proceso de creación de aplicaciones “templates”. Comienza definiendo el concepto de la aplicación para usuario final y su rol en el acceso a la información del negocio. Brinda un marco metodológico bastante estándar en lo que ha desarrollo de aplicaciones (como piezas de software) se refiere. Divide el proceso de creación de las aplicaciones para usuarios finales en dos grandes fases: especificación y desarrollo. Clasifica a los usuarios según su perfil de consulta, desde usuarios con un perfil más estratégico y menos predecibles (power users) hasta usuarios netamente operacionales que consumen una serie de reportes

estándares (final users) pasando por los usuarios gerenciales con uso de interfases push-button (EIS users).

Kimball destaca, como uno de los requisitos a cumplir por las aplicaciones, la posibilidad de hacer análisis AD HOC. Análisis ad hoc es simplemente la habilidad para los usuarios de cambiar los parámetros sobre un reporte para crear sus propias versiones personalizadas de ese reporte. De esta forma se dará respuesta a necesidades como “Quiero ver este reporte, pero por mes en lugar de trimestre”, “Puedo ver este reporte a nivel provincia en lugar de región”, “Puedo ver esta evolución de ventas mensual pero sólo de mi equipo de ventas?”. Más importante aún es que las respuestas a estos interrogantes lo resuelven los propios usuarios sin la necesidad de la intervención del departamento sistemas y maximizando el tiempo de análisis por sobre el tiempo de construcción e integración de la información.

Kimball define entonces que las aplicaciones “templates” para usuarios finales proveen el marco (layout) y la estructura de un reporte para ser especializado por un conjunto de parámetros. El usuario selecciona los parámetros de una pick listo o aceptando los valores por defecto cuando ejecuta el template. Este enfoque orientado a parámetros permite a los usuarios generar docenas o potencialmente cientos de reportes de estructura similar desde un mismo template. Todo esto de forma amigable y con interfases gráficas que hacen uso de todo el trabajo realizado en la construcción del warehouse en etapas previas del BDL. Una advertencia que hace Kimball es el tiempo que existe entre el relevamiento y especificación de las aplicaciones para usuarios finales y el momento del desarrollo e implementación de la misma. Como pudimos ver, durante las primeras semanas del proyecto se realiza el relevamiento de los requerimientos de los usuarios y recién una vez que existen datos en el warehouse (aunque más no sea datos de prueba) puede comenzarse con la construcción de la aplicación final (al menos es lo recomendable), es decir luego del diseño lógico y físico. Kimball insta entonces a manejar con mucho cuidado el tema de la documentación explícita (tratar que lo menos posible quede en la propiedad intelectual de los que hicieron el relevamiento). Sino se deberá realizar un esfuerzo significativo en el redescubrimiento de los puntos especificados en etapas tempranas del proyecto.

Kimball destaca cuatro pasos principales (siempre enfatizando el hecho de involucrar a los usuarios en cada uno de estos pasos).

- ✚ Determinación del conjunto de templates iniciales (identificar reportes candidatos, clasificarlos y priorizarlos)
- ✚ Diseño de la estrategia de navegación dentro de la aplicación (esquema de pantallas, esquema de carpetas, directorios, criterios de agrupamiento -por datos, por dueño, por regla del negocio, etc.)
- ✚ Determinación de estándares (nombre de objetos, ubicación de objetos, formato de las salidas)
- ✚ Detalle de las especificaciones (definición: nombre, descripción o propósito, frecuencia, parámetros, restricciones, layout, etc.).

### 3.2.9 Desarrollo de Aplicaciones para Usuarios Finales

Siguiendo a la especificación de las aplicaciones para usuarios finales, el desarrollo de las aplicaciones de los usuarios finales involucra configuraciones de la meta data y construcción de reportes específicos.

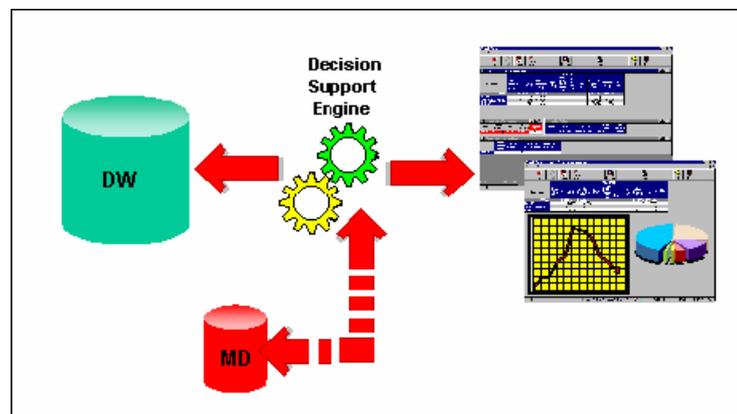


Figura III.23 - Configuración de Meta data y construcción de reportes

Una vez que se ha cumplido con todos los pasos de la especificación y se tiene la posibilidad de trabajar con algunos datos de prueba, comienza el desarrollo de la aplicación.

## Selección de un enfoque de implementación

- ✚ Basado en Web
- ✚ Inter/Intra net
- ✚ Usuarios altamente distribuidos
- ✚ Manejo centralizado de nuevas versiones
- ✚ Herramienta propietaria
- ✚ Mayor complejidad de uso
- ✚ Para usuarios más capacitados
- ✚ Instalación local
- ✚ EIS
- ✚ Acceso estructurado
- ✚ Secuencialidad de pantallas
- ✚ Push-Button
- ✚ Interface personalizada
- ✚ API (Application Programming Interface)
- ✚ Desarrollos propios sobre la base de un conjunto de funcionalidades

## Desarrollo de la aplicación

- ✚ Definición de herramienta de acceso al Meta data
- ✚ Desarrollo de plantillas y esquema de navegación de la aplicación
- ✚ Selección de reportes para pre-ejecución

## Prueba y verificación de datos

- ✚ Descripciones
- ✚ Información duplicada
- ✚ Relaciones entre atributos
- ✚ Consistencia e integridad de datos con sistemas fuentes

## Documentación y Roll Out

- ✚ Retroalimentación con los resultados de la puesta en producción

## Mantenimiento

- ✚ Nuevos plantillas
- ✚ Incorporación de nuevos sistemas fuentes
- ✚ Monitoreo de performance
- ✚ Eliminación de plantillas en desuso

### 3.2.10 Implementación

La implementación representa la convergencia de la tecnología, los datos y las aplicaciones de usuarios finales accesible desde el escritorio del usuario del negocio. Hay varios factores extras que aseguran el correcto funcionamiento de todas estas piezas, entre ellos se encuentran la capacitación, el soporte técnico, la comunicación, las estrategias de feedback. Todas estas tareas deben ser tenidas en cuenta antes de que cualquier usuario pueda tener acceso al data warehouse.

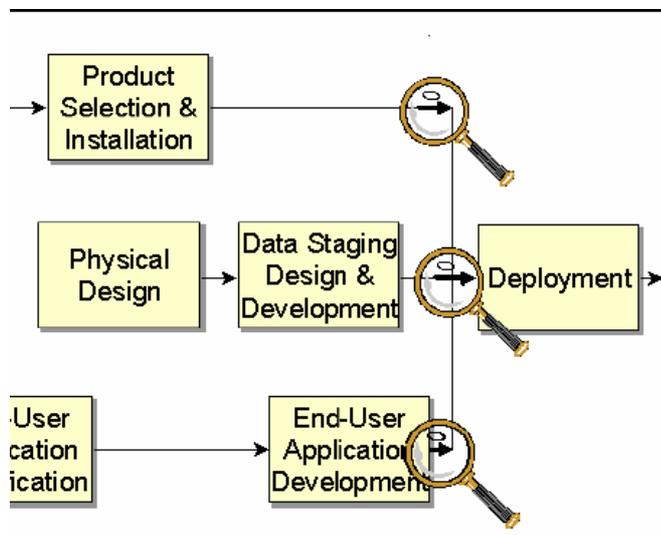


Figura III.24 - Implementación como convergencia de la tecnología, los datos y las aplicaciones

Kimball presenta una serie de tareas que deben cumplirse para garantizar un fin de proyecto y un producto de calidad.

La tecnología que reside en el escritorio del usuario es la última pieza que debe ser ubicada antes de la salida a producción (roll out o deployment). Desafortunadamente, afirma Kimball, las organizaciones frecuentemente subestiman el esfuerzo y el tiempo requerido para esta etapa. Kimball propone entonces una checklist sobre actividades que deberían ocurrir antes de la implantación para asegurar que la infraestructura correspondiente al ambiente del usuario este correcta. Esta checklist incluye: configuración de hardware, conexión a las bases, acceso a intranet o internet, direcciones LAN (si no son dinámicamente asignadas), auditorias de tecnología sobre las configuraciones en las que se encontraban las PCs, preveer actualizaciones de hardware y software (determinando responsables, proyecto o área de usuario),

verificaciones de seguridad (logon de red y base de datos), prueba de procedimientos de instalación en una variedad de máquinas, planificación de instalación con la correspondiente educación a los usuarios, etc.

Ralph Kimball dedica gran parte de este capítulo a la educación de los usuarios finales afirmando que “una robusta estrategia de educación para usuarios de negocios es un prerrequisito para el éxito del proyecto”. Entre sus recomendaciones para cumplir exitosamente con esta tarea se encuentran dos aspectos, por un lado el alcance de la educación, y por otro, la metodología de presentación. Para lo primero, debe instruirse al usuario en tres aspectos claves: contenido del warehouse, aplicación y herramientas de acceso. En cuanto a lo segundo, para los usuarios finales no es fácil entender los límites entre las aplicaciones, los contenidos y las herramientas del warehouse.

Para los usuarios el data warehouse es un todo, no la suma de componentes discretos, por lo tanto la educación también debe reflejar la misma perspectiva. No debe interpretarse educación de los usuarios finales como educación sólo sobre las herramientas de acceso. La educación sobre herramientas es inútil al menos que se complemente con educación sobre el contenido del warehouse (cuáles datos están disponibles, qué significan, cómo se usan y para qué usarlos). Otro factor clave es la correcta entrega de niveles de educación según el perfil del usuario (usuario final, power user, etc.)

El armado de un ambiente de training (como subconjunto de datos de producción) es siempre recomendable, dado que los tiempos de respuesta en producción, si bien pueden ser apropiados para el análisis de la información, tal vez no sean los mejores para un curso de un día de duración.

Además, el contenido del warehouse puede variar entre curso y curso haciendo más complicada la administración y mantenimiento del material educativo.

Kimball establece dos premisas para regular las relaciones entre la educación y el warehouse.

Realizar educación de usuarios finales sólo si el warehouse está listo (en tiempo y forma) y establecer la política de “Sin Educación, entonces Sin Acceso”. Cumpliendo estas dos premisas se puede garantizar el éxito del proyecto en esta etapa final del ciclo de vida.

El soporte es la otra columna que ayudará a implementar con éxito el proyecto. Se deberá definir claramente los niveles de soporte (a nivel aplicación, modelo, calidad de datos) de forma transparente para el usuario. La documentación juega un papel importantísimo en la ayuda a usuarios finales, dado que el datawarehouse evolucionará rápidamente, la documentación escrita puede quedar obsoleta también de forma prematura. La documentación en línea empieza entonces a tener vida propia, ya sea dentro del warehouse en sí mismo o como herramientas anexas, como el data warehouse web site propuesto por Kimball.

Finalmente, la propuesta de Kimball se orienta a la metodología de implementación. Para ello propone un esquema de versionado (release framework). Primero se pasa por la versión Alpha, primera oportunidad para el grupo de trabajo de conducir una prueba del sistema de principio a fin.

Todos los componentes del sistema deben ser testeados (infraestructura técnica, extracción, transformación, carga, procedimientos de calidad, performance, templates, etc.). Típicamente es de naturaleza iterativa y sólo puede decirse que se termina esta versión cuando el grupo de trabajo confía plenamente en la calidad del warehouse como un todo (por lo tanto a la hora de estimar los tiempos debe tenerse en cuenta que este será el objetivo final de la versión Alpha por lo cual debe asignarse un tiempo prudencial). Durante la etapa Alpha un conjunto limitado de usuarios finales son provistos con acceso al warehouse (sólo aquellos que pertenecen al grupo de trabajo). Luego viene la versión Beta.

El objetivo de esta versión es conducir una prueba a nivel usuario de principio a fin. El grupo Beta está formado por los power users, la determinación de la cantidad de este grupo es también importante, pues si son pocos la aplicación no se probará adecuadamente, y si son muchos se estará saltado la etapa de testing cayendo directamente en una salida a producción.

Una vez superadas estas dos versiones, al igual que en un proceso de diseño e implementación de software, llegamos a un estado GA (general availability). Lo que sugiere Kimball es que, todo cambio y/o modificación que se realice posteriormente al warehouse, pase internamente por un estado Alpha y Beta aunque externamente sea una nueva versión, release o pack desde el último GA.

### 3.2.11 Mantenimiento y crecimiento

Como se remarca siempre, Data Warehousing es un proceso (de etapas bien definidas, con comienzo y fin, pero de naturaleza espiral) pues acompaña a la evolución de la organización durante toda su historia. Se necesita continuar con los relevamientos de forma constante para poder seguir la evolución de las metas por conseguir. Según afirma Kimball, “si se ha utilizado el BDL el data warehouse esta preparado para evolucionar y crecer”.

Al contrario de los sistemas tradicionales, los cambios en el desarrollo deben ser vistos como signos de éxito y no de falla. Es importante establecer las prioridades para poder manejar los nuevos requerimientos de los usuarios y de esa forma poder evolucionar y crecer.

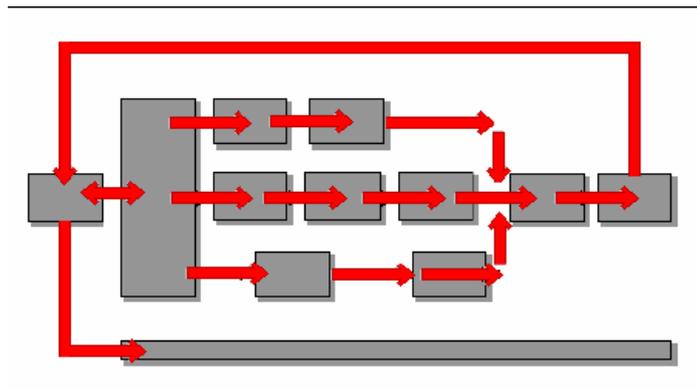


Figura III.25 - Data Warehousing como proceso de naturaleza espiral

El sentido de las flechas fija claramente el concepto evolutivo y espiral del BDL.

Una vez que se ha construido e implantado el data warehouse no hay tiempo para el descanso, rápidamente debemos estar preparados para administrar el mantenimiento y crecimiento del mismo. Si bien las tareas pueden llegar a parecer similares a las tratadas en otras etapas del BDL, existe una diferencia clave: los usuarios están ahora accediendo al warehouse.

Kimball brinda entonces una serie de puntos a tener en cuenta para mantener exitosamente el warehouse. Entre ellos se destacan: el continuo soporte y la constante

capacitación a usuarios de negocios, el manejo de la infraestructura (monitoreo de base de datos, tráfico, etc.), tuning de rendimiento sobre las consultas, mantenimiento del meta data y procesos ETLs. Otros aspectos involucran el monitoreo regular del cumplimiento de las expectativas sobre el warehouse (variables de medición del éxito del proyecto que se habían fijado en la etapa de relevamiento), relevamiento de casos de estudio (situaciones reales donde una decisión basada en información del warehouse tuvo impacto sobre el negocio ROI), constante publicidad interna del uso del warehouse (permitiendo acceso siempre y cuando se tenga la capacitación correspondiente) y constante comunicación con los sectores de negocios y sistemas para asegurar la buena salud del datawarehouse.

En cuanto a cómo prepararse para el crecimiento del warehouse, Kimball recomienda la creación de un comité conformado por analistas del negocio y sistemas que establezca prioridades y procedimientos.

El manejo de prioridades lo discrimina según sean mejoras menores o mayores (clasificadas según el impacto en el warehouse y excluyendo los errores, los cuales deben ser resueltos inmediatamente). Como mejoras menores incluye incorporación de datos sencillos, nuevas agregaciones, cambios en niveles superiores del modelo (atributos de alto nivel, lejanos a las bases tables), tareas que en definitiva no lleven más de días o semanas.

Advierte sobre el razonamiento simplista de decidir no aplicar una mejora pues el cambio es muy menor, "1000 pequeñas mejoras es una ventaja estratégica". Dentro de las mejoras de mayor impacto se encuentran las que modifican y/o crean bases tables o atributos asociados a las mismas, tareas que pueden disparar nuevos proyectos de meses de duración.

Finalmente, Kimball concluye su libro garantizando que el seguimiento y cumplimiento de todas las etapas BDL ayudará a obtener un warehouse mantenible y de controlado crecimiento.

Como es ilustrado en el diagrama del BDL, se deberá volver a iterar sobre el mismo pasando nuevamente por cada una de las etapas para de esta forma administrar correctamente los cambios a warehouses existentes, como así también para respetar y aprovechar definiciones realizadas para otros proyectos.

### 3.2.12 Gerenciamiento del Proyecto

El gerenciamiento del proyecto asegura que las actividades del BDL se lleven en forma y sincronizadas.

Como lo indica el diagrama, el gerenciamiento acompaña todo el ciclo de vida. Entre sus actividades principales se encuentra el monitoreo del estado del proyecto y la comunicación entre los requerimientos del negocio y las restricciones de información para poder manejar correctamente las expectativas en ambos sentidos.

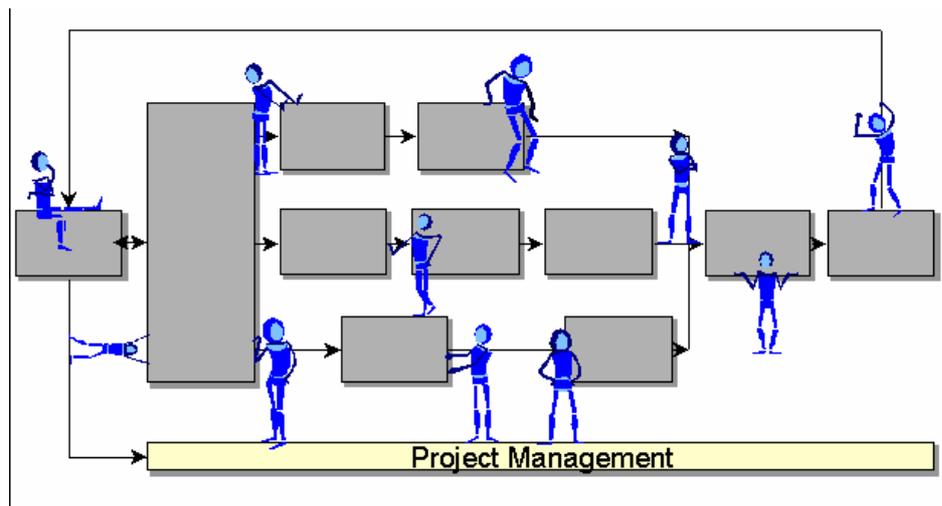


Figura III.26 - Gerenciamiento del proyecto

El gerenciamiento del proyecto debe asegurar que las actividades del BDL se lleven en forma y sincronizadas.

### 3.3 METODOLOGÍA CRISP-DM

Cross-Industry Standard Process for Data Mining) es un consorcio de empresas (inicialmente bajo una subvención inicial de la Comisión Europea), incluyendo SPSS, NCR y DaimlerChrysler.

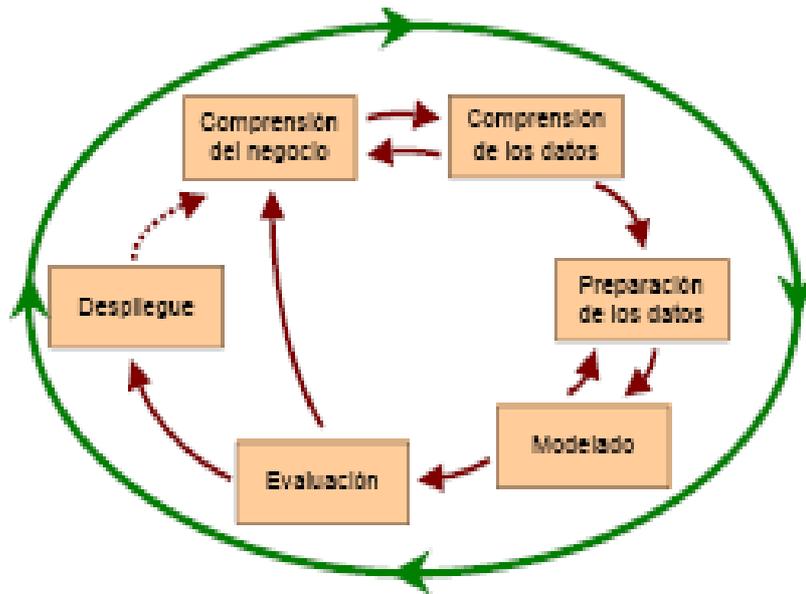


Figura III.27 - Metodología de Crisp-Dm

En esta metodología se escoge las técnicas OLAP (Online Analytical Processing) las fases cubiertas en el presente documento son:

- ✚ Comprensión del negocio
- ✚ Comprensión de los datos
- ✚ Preparación de los datos.
- ✚ Modelado.
- ✚ Evaluación.
- ✚ Despliegue

#### 3.3.1 Comprensión del negocio:

En esta fase consiste en entender los objetivos y requerimientos del proyecto desde una perspectiva de negocio.

Subfases:

- ✚ Establecimiento de los objetivos de negocio (contexto inicial, objetivos y criterios de éxito),
- ✚ Evaluación de la situación (inventario de recursos, requerimientos, suposiciones y restricciones, riesgos y contingencias, terminología y costes y beneficios),
- ✚ Establecimiento de los objetivos de minería de datos (objetivos de minería de datos y criterios de éxito) y
- ✚ Generación del plan del proyecto (plan del proyecto y evaluación inicial de herramientas y técnicas).

### 3.3.2 Comprensión de los datos

Consisten en recopilar y familiarizarse con los datos, identificar los problemas de calidad de datos y ver las primeras potencialidades o subconjuntos de datos que pueden ser interesantes analizar (según los objetivos de negocio en la fase anterior).

Subfases:

- ✚ recopilación inicial de datos (informe de recopilación),
- ✚ descripción de datos (informe de descripción),
- ✚ exploración de datos (informe de exploración) y
- ✚ verificación de calidad de datos (información de calidad).

### 3.3.3 Preparación de los datos

La staging área ó preparación de los datos, es una colección de procesos que limpian, transforman, combinan, y preparan los datos originales para su utilización en el Datawarehouse. En la staging área, los datos originales son transformados a formatos comunes, comprobada su consistencia y su integridad referencial, y preparados para cargar en la base de datos del Datawarehouse.

Una vez localizadas las fuentes de datos, estos se han de preparar para que se les puedan aplicar los métodos o herramientas que construirán el modelo deseado.

Esta fase aunque parezca sencilla conlleva aproximadamente el 70% del esfuerzo en los proyectos de data mining de nueva implantación.

En este punto hay que asegurarse de unas cuantas cosas:

- ✚ Que los datos tengan la calidad suficiente: es decir, que no contengan errores, redundancias o que presenten otro tipo de problemas.
- ✚ Que los datos sean los necesarios, quizás haya que no nos harán falta y quizás tendremos que añadir.
- ✚ Que estén en la forma adecuada: muchos métodos de construcción de modelos requieren que los datos estén en un formato determinado que no ha de coincidir necesariamente con el que están almacenados. Las técnicas utilizadas para asegurar los tres aspectos comentados son la limpieza de datos, la transformación de los datos y la reducción de la dimensionalidad.
- ✚ La limpieza de datos, consiste en procesar los datos eliminando los atributos que sean erróneos o redundantes, siendo los factores de distorsión más importantes:
  - Datos incompletos, puede pasar especialmente en aquellos atributos en que cuando se diseñó el proceso correspondiente a la entrada de datos se decidió que no eran obligatorios o que tenían formato libre.
  - Datos redundantes, repetición de tuplas.
  - Datos incorrectos o inconsistentes, muy común cuando el tipo de valores que puede recibir un atributo no está controlado porque está declarado como "texto libre".
- ✚ Transformación de datos, no siempre los datos están en la forma más adecuada para poder aplicar los métodos que hacen falta para la tarea que se ha de llevar a cabo y el modelo que se quiere obtener.
- ✚ Reducción de la dimensionalidad, una de las justificaciones más frecuentes para la utilización de técnicas de data mining es su capacidad para trabajar con grandes conjuntos de datos. Ahora bien el tamaño de un conjunto de datos, o de un problema de data mining, la da tanto la cantidad de registros que tiene como el número de atributos que se manejan.

### 3.3.4 Modelado del almacén

En esta fase, varias técnicas de modelado son seleccionadas y aplicadas y sus parámetros son calibrados a valores óptimos.

Existen varias técnicas para los mismos problemas, algunas de estas técnicas tienen requerimientos específicos en el formato de los datos, por lo que puede ser necesario el paso atrás a la fase de preparación de los datos.

Como primer paso se selecciona en función del problema la técnica a utilizar, en el caso que nos ocupa sobre el almacén construido se aplicarán técnicas de análisis en línea de la información u OLAP.

Antes de construir el modelo, necesitamos generar un procedimiento o mecanismo para probar la calidad y validez del modelo.

Una vez construido el modelo el diseñador interpreta el modelo de acuerdo a sus conocimientos del dominio y al criterio de éxito del proyecto.

Esta tarea interfiere con la fase de evaluación siguiente, el diseñador contacto con los analistas del negocio y expertos del dominio para discutir los resultados en el contexto del negocio.

#### 3.3.4.1 Métodos OLAP

Los métodos OLAP<sup>5</sup> surgieron para analizar los datos de ventas y marketing, así como para procesar datos administrativos y consolidar datos procedentes de diversas fuentes de cara a efectuar un análisis de rentabilidad, mantenimiento de la calidad y otros tipos de aplicaciones que se caracterizan porque redefinen de manera continua y flexible el tipo de información que hay que extraer, analizar y sintetizar (en comparación con las bases de datos tradicionales, dirigidas a responder consultas muy prefijadas y rutinarias).

Los sistemas OLAP se alimentan de los datos generados por los sistemas transaccionales (facturación, ventas, producción,...). Herramientas típicas de OLAP son las que permiten un análisis multidimensional de los datos en contra de las típicas facilidades de creación de resúmenes e informes propios de los sistemas de bases de datos tradicionales.

---

<sup>5</sup> E.F. Codd, S.B. Codd, C.T. Salley. *Beyond decisión support*, *Computerworld*, 27, July 1993.

Desde un punto de vista conceptual, OLAP tiene las cuatro siguientes funcionalidades.<sup>6</sup>

- ✚ Querying: posibilidad de generar consultas potentes a través de una interfase simple y declarativa.
- ✚ Restructuring: capacidad de reestructurar información en una base de datos multidimensional explotando la dimensionalidad de los datos y mostrando diferentes perspectivas de los datos.
- ✚ Classification: capacidad de clasificar ó agrupar conjuntos de datos en una manera apropiada para la consiguiente sumarización.
- ✚ Summarization/consolidation: esta es una generalización de los operadores de agregación en SQL estándar. En resumen, la sumarización mapea conjuntos de valores de tipo numérico a un único valor “consolidado”.

La unidad de datos de OLAP es el cubo, que es una representación de los datos de interés para el análisis. Posteriormente, hablaremos en mayor detalle de cada una de las operaciones que permiten “cortar” y mirar los cubos desde la perspectiva de muchos grupos diferentes de usuarios.

La característica principal de los cubos es que optimizan las consultas.

Normalmente se guardan en forma de tabla relacional especial que facilita ciertos tipos de consultas. Por ejemplo, hay columnas de las tablas que se llaman columnas de dimensión que facilitan y proveen datos para resúmenes e informes.

Las columnas llamadas columnas agregadas permiten precalcular cantidades como conteos, sumas y medias.

Construir un cubo requiere un análisis detallado de las necesidades de datos del grupo de usuarios a los cuales va dirigido y puede requerir mucho tiempo, tanto de diseño como de instalación por primera vez. Compensa por el hecho que facilita extraordinariamente las tareas de análisis de datos de los diversos grupos de usuarios y, una vez establecido, resulta más sencillo de modificar que las tablas relacionales tradicionales.

---

<sup>6</sup> Marc Gyssens (LUC), Laks V.S. Lakshmanan (CUM). *A foundation for multidimensional databases, In Proc. Of the 22nd VLDB Conference, Mumbai (Bombay), India, 1996.*

### 3.3.4.2 Análisis y modelado multidimensional

El objetivo del análisis multidimensional es ganar comprensión en el conocimiento contenido en las bases de datos. Su principal ventaja es que facilitan los análisis complejos (al estar muy próximos a la manera de pensar del analista) y la visualización de los datos en el datamart para procesos de toma de decisiones, reduciendo la confusión y disminuyendo las interpretaciones erróneas.

Además, ya que los datos están almacenados físicamente en una estructura multidimensional ó base de datos n-dimensional, la velocidad de estas operaciones es varias veces superior y más consistente de lo que es posible en otras estructuras de bases de datos. La combinación de simplicidad y velocidad es uno de los principales beneficios del análisis multidimensional.

El modelo está basado en la noción de dimensión que permite especificar diferentes maneras de estudiar la información, de acuerdo con las perspectivas del negocio bajo las cuales el análisis puede ser realizado. Cada dimensión se organiza en una jerarquía de niveles, correspondiendo a dominios de datos en diferentes niveles ó granularidades. Un esquema multidimensional consiste en un conjunto de tablas de hechos (también llamadas f-tables) que se definen respecto a combinaciones particulares de niveles.

Una instancia multidimensional asocia medidas, que corresponden a los datos a ser estudiados, con coordenadas simbólicas a las tablas de hechos.

Finalmente, en una dimensión, los valores con un gran nivel de detalle pueden hacer roll-up (agruparse) a valores más generales.<sup>7</sup>

---

<sup>7</sup> Alberto Abelló (UPC), José Samos (UGR), Félix Saltor (UPC). *A DataWarehouse multidimensional data models classification*, 2001.

### 3.3.4.3 Esquema multidimensional

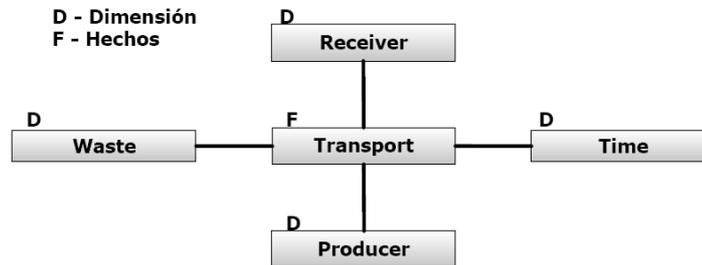


Figura III.28 - Esquema multidimensional

#### Estructura del cubo

La estructura del cubo se define por medio de sus medidas y dimensiones. Derivan de tablas del origen de datos del cubo.

El conjunto de tablas del que derivan las dimensiones y medidas de un cubo se denomina esquema del cubo. Las medidas del cubo derivan de las columnas de la tabla de hechos. Las dimensiones del cubo derivan de columnas de las tablas de dimensiones.

Hay dos tipos comunes de esquemas de cubo: en estrella y en copo de nieve<sup>8</sup>. En un esquema en estrella, cada tabla de dimensión se combina con la tabla de hechos.

En un esquema en copo de nieve, una o más tablas de dimensiones que no se combinan con la tabla de hechos son para dimensiones con varias tablas de dimensiones.

#### Esquema en estrella

Para facilitar el acceso a los datos y el análisis, un datamart organiza físicamente en estructuras llamadas esquemas en estrella.

---

<sup>8</sup> Microsoft Training and Certification. Designing and implementing OLAP solutions with MS SQL Server 2000. Microsoft Technical Report, 2001.

Un datawarehouse construido sobre datamarts utiliza uno o más esquemas en estrella para representar eventos o procesos específicos del negocio.

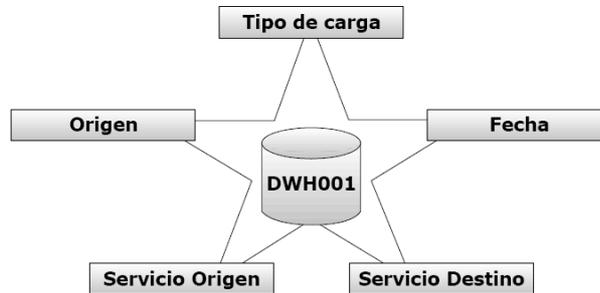


Figura III.29 - Esquema en estrella

Un esquema en estrella se caracteriza por tener una tabla central de hechos rodeada por tablas de dimensiones que contienen información desnormalizada de los hechos<sup>9</sup>. Los siguientes elementos son característicos en un esquema en estrella:

- ✚ El centro del esquema es la tabla de hechos, contiene las métricas o medidas del negocio.
- ✚ Las puntas de la estrella son las tablas de dimensiones. Estas puntas son utilizadas para describir la información existente en un proceso específico del negocio y proveen el contexto a los datos numéricos.

### Tabla de hechos

La tabla de hechos es la tabla central del esquema en estrella que representa datos numéricos en el contexto de las entidades del negocio. La tabla de hechos está constituida por medidas y por foreign keys.

### Medidas

Una medida, es una columna numérica de la tabla de hechos. Las medidas representan los valores que se van a analizar, tales como las unidades vendidas o el número de empleados.

---

<sup>9</sup> Microsoft Training and Certification. *Designing and implementing OLAP solutions with MS SQL Server 2000*. Microsoft Technical Report, 2001.

## Foreign keys

Una foreign key es la representación de la primary key de una dimensión en la tabla de hechos. Las foreign keys son tomadas de las primary keys para cada dimensión de la tabla. La combinación de estas claves es el identificador para cada registro de la tabla de hechos.

## Tablas de dimensiones

Una tabla de dimensión representa una entidad del negocio, dotando de contexto a los datos numéricos de la tabla de hechos. El diseño de las tablas de dimensiones apunta a las necesidades analíticas del usuario presentando información descriptiva que es fácil de utilizar para los usuarios.

Una tabla de dimensiones:

- ✚ Describe las entidades del negocio, en los esquemas en estrella representan una única entidad del negocio, tales como un producto o un cliente.
- ✚ Contienen atributos que proveen contexto a los datos numéricos.
- ✚ Presentan los datos organizados en jerarquías, en cada dimensión, se pueden organizar los datos en una o varias jerarquías.

## Dimensiones del cubo

Cuando se diseña un datawarehouse es una buena práctica crear una tabla de dimensión separada para las fechas más que utilizar una columna de fecha/hora en la tabla de hechos.<sup>10</sup>

Como con otras tablas de dimensiones, se utiliza una clave entera para ligar la tabla de dimensión de fechas con la tabla de hechos. Crear una tabla de dimensión de fechas separada tiene ciertas ventajas sobre almacenar una columna de fecha/hora en la tabla de hechos:

---

<sup>10</sup> *Microsoft Training and Certification. Designing and implementing OLAP solutions with MS SQL Server 2000. Microsoft Technical Report, 2001.*

- ✚ Contiene propiedades adicionales de la fecha, tales como la estación para un mes ó una marca de festivo para un día.
- ✚ Reduce el espacio de almacenamiento y diseñar y procesar una dimensión basada en una tabla con sólo unos pocos registros es mucho más rápido que extraer los valores de la dimensión de una gran tabla de hechos.
- ✚ Se utiliza en múltiples tablas de hechos, esto permite crear una dimensión única y compartida en múltiples cubos.

### **Almacenamiento de cubos**

Una vez decidido emplear un entorno de consulta es necesario decidir su modo de almacenamiento.

Las opciones de almacenamiento físico afectan al rendimiento, los cubos pueden necesitar una cantidad considerable de espacio para contener los datos y las agregaciones (resúmenes de datos precalculados que mejoran el tiempo de respuesta de consultas al tener preparadas las respuestas antes de que se planteen las preguntas) en estructuras multidimensionales.

Los cubos pueden dividirse en particiones y cada una de ellas se puede almacenar mediante un modo de almacenamiento distinto.

Los modos de almacenamiento son:

- ✚ OLAP multidimensional (MOLAP).
- ✚ OLAP relacional (ROLAP).
- ✚ OLAP híbrido (HOLAP).

ROLAP es la arquitectura de bases de datos multidimensional en la que los datos se encuentran almacenados en una base de datos relacional, la cual tiene forma de estrella (también llamada copo de nieve ó araña). En ROLAP, en principio la base de datos sólo almacena información relativa a los datos en detalle, evitando acumulados (evitando redundancias).

Con el modo de almacenamiento ROLAP, la respuesta de consultas suele ser más lenta que la de los otros dos modos de almacenamiento. ROLAP se usa normalmente

para el acceso a grandes conjuntos de datos que se consultan con poca frecuencia, tales como datos históricos de años no recientes.

En un sistema MOLAP, en cambio, los datos se encuentran almacenados en ficheros con estructura multidimensional, los cuales reservan espacio para todas las combinaciones de todos los posibles valores de dimensión que representan acumulados.

Es decir, un sistema MOLAP contiene precalculados (almacenados) los resultados de todas las posibles consultas a la base de datos. MOLAP consigue consultas muy rápidas a costa de mayores necesidades de almacenamiento, y retardos en las modificaciones (que no deberían producirse salvo excepcionalmente), y largos procesos batch de carga y cálculo de acumulados. En ROLAP, al contener sólo las combinaciones de valores de dimensión que representan detalle, es decir, al no haber redundancia, el fichero de base de datos es pequeño. Los procesos batch de carga son rápidos (ya que no se requiere agregación), y sin embargo, las consultas pueden ser muy lentas, por lo que se aplica la solución de tener al menos alguna consultas precalculadas.

En general, el modo MOLAP es más apropiado para las particiones de cubos que se utilizan con frecuencia y donde se necesita una rápida respuesta de consultas.

HOLAP, esta es la solución recomendada, combina atributos de los modos MOLAP y ROLAP. La idea que yace detrás de HOLAP es que obtengamos los beneficios de una arquitectura MOLAP para la información resumida, pero necesitamos el detalle, el sistema “retrocederá” a una base de datos relacional. Por lo tanto el rendimiento aumenta aunque solo a nivel de información resumida, y la escalabilidad puede ser resuelta porque todo el detalle está almacenado en tablas relacionales convencionales.

Las particiones almacenadas como HOLAP son más pequeñas que sus equivalentes MOLAP y responden más rápidamente que las particiones ROLAP a las consultas que implican datos de resumen. El modo de almacenamiento HOLAP suele ser más adecuado para particiones en cubos que requieren una respuesta de consultas rápida basada en una gran cantidad de datos de origen.

### 3.3.5 Evaluación

Es importante evaluar el modelo y revisar la construcción con el fin de comprobar que se cumplen los objetivos del negocio. Aquí es crítico determinar si partes importantes del negocio no han sido lo suficientemente consideradas. Al final de esta fase, el líder del proyecto debe decidir exactamente cómo utilizar los resultados del proceso.

Es en esta fase y con la ayuda del analista de los datos que gracias a la naturaleza iterativa de este tipo de proyectos, pueden surgir nuevas preguntas a responder que hagan que el proyecto retorne a la fase de conocimiento del negocio a fin de poder responderlas.

Una buena manera de definir los resultados de un proyecto de este tipo es utilizar la ecuación:

$$\text{RESULTADOS} = \text{MODELOS} + \text{DESCUBRIMIENTOS}$$

En esta ecuación estamos definiendo que el resultado del proyecto no son sólo los modelos (aunque, por supuesto, son importantes), también los descubrimientos, que se definen como cualquier cosa (aparte del modelo) que son importantes para alcanzar los objetivos del negocio (o importantes para conducir a nuevas preguntas o a los efectos secundarios, por ejemplo problemas en la calidad de los datos no cubiertos en el ejercicio de data mining).

Las fases de este paso son: la evaluación de los resultados, la revisión del procesos y determinar los siguientes pasos a seguir.

- ✚ Evaluación de los resultados, pasos de evaluación previos trabajan con factores tales como la precisión y la generalidad del modelo. Este paso evalúa el nivel en que el modelo se encuentra respecto a los requerimientos del negocio e intenta determina si existen requerimientos que no cumple el modelo.
- ✚ Revisión del proceso, en este punto el modelo parece ser satisfactorio y cubrir los requerimientos del negocio. Es ahora cuando se realiza una revisión del proceso para determinar si hay algún factor o tarea importante que de alguna manera no se ha tenido en cuenta. Esta revisión cubre además cuestiones de calidad, por ejemplo, ¿hemos construido correctamente el modelo?, ¿hemos

utilizado atributos que podíamos utilizar y estarán disponibles para análisis futuros?

- ✚ Determinar los siguientes pasos, de acuerdo con los resultados de la revisión, se decide como proceder al siguiente paso. Se decide si finalizar el proyecto y pasarlo a desplegar o iniciar nuevas iteraciones ó nuevos proyectos. Esta tarea incluye evaluar los recursos existentes y presupuestarlos para influir en la decisión.

### 3.3.5.1 Evaluación del modelo

Resume las principales ventajas que supone la implantación de la solución en la empresa, algunos comentarios son resultado de entrevistas realizadas con un grupo de los potenciales usuarios del modelo.

Si consideramos la suma de ventajas, podemos ver que adoptar esta solución comporta incrementar el potencial global de análisis de la información y mejorar la capacidad y control de gestión de la empresa, características cada vez más necesarias para seguir siendo competitivos.

Las ventajas se analizan desde 4 ópticas diferentes para ver su equilibrio global:

- ✚ Óptica de negocio, es decir, los beneficios directos al negocio que aporta la solución.
- ✚ Óptica económica, el factor económico de la solución: ahorro de costes y rentabilidad de la inversión.
- ✚ Óptica del usuario, desde este punto de vista se analizan las ventajas y el impacto tecnológico de la solución para el usuario.
- ✚ Óptica técnica, factores técnicos que hacen también muy coherente y ventajosa la solución presentada.

#### Óptica de negocio

Los beneficios más importantes de la solución son el impacto que causa en el negocio. Los listados de papel se sustituyen por hojas de cálculo formateadas y enriquecidas con todas las posibilidades de Microsoft Excel. Las siguientes funciones permiten analizar mejor los datos para una toma de decisiones más efectiva y certera:

- ✚ Análisis dinámico e interactivo, el análisis multidimensional de la información permite analizarla desde múltiples visiones, con el fin de ver más información en los mismos datos.
- ✚ Resaltados y gráficos, los datos se pueden resaltar de manera que los incrementos se muestren en verde y los descensos en rojo. Así se puede centrar la atención en los puntos que la merecen. Los gráficos también cumplen su función de resumir la información a alto nivel.
- ✚ Datos enriquecidos, los datos se pueden insertar dentro de hojas de cálculo ya definidas que pueden contener fórmulas para calcular más datos, como previsiones e indicadores estadísticos.

Otro punto fuerte del modelo es su velocidad, ya que para este caso se ha decidido la utilización de la técnica de almacenamiento MOLAP que consigue consultas muy rápidas ya que las agregaciones están precalculadas. Los datos serán refrescados de las fuentes operacionales una vez a la semana, no es necesario hacerlo más a menudo pero en función de cómo evolucione el modelo y su explotación, esta técnica de almacenamiento puede ser cambiada sin problemas.

### **Óptica económica**

No hay un coste adicional por licencia, la solución se apoya en la infraestructura ofimática ya implementada en la empresa.

El trabajo derivado de la creación de informes puede verse como un coste variable dependiente de la cantidad de informes que se quieran desarrollar que a posteriori repercute en un ahorro del tiempo de diseño de informes, un único informe satisface decenas de informes convencionales.

Se ahorra en papel y en tiempo de impresión.

### **Óptica del usuario**

Desde este punto de vista analizamos cómo verá el usuario la implantación de esta solución y cómo le afectará.

Observamos que con el datamart y la aplicación de técnicas OLAP sobre él se facilita tremendamente lo que antes era muy costoso y dependía en gran medida de la habilidad técnica de los usuarios, la creación de informes para los directivos.

El usuario tiene cierta familiaridad con Excel, luego no hay un “nuevo sistema” que implantar ya que todo el reporting lo recibe en la herramienta que más conoce:

Microsoft Excel. No es necesaria formación (o es muy mínima).

No es necesaria ninguna instalación en el PC, no hay que molestar al usuario para instalar algo y el coste de formación es bajísimo debido a que la herramienta de consulta es Microsoft Excel.

El usuario tiene más potencia de análisis. Ahora tiene una herramienta flexible que le permite manipular fácilmente la información.

### **Óptica técnica**

Existen factores técnicos que hacen también muy coherente y ventajosa la solución presentada. Esta se integra con los sistemas informáticos de la organización, con SQL Server y se reaprovecha la inversión ofimática, la solución se apoya en la infraestructura que ya tiene actualmente la empresa.

### **Revisión del proceso**

Se decide revisar el proceso de carga y limpieza de datos, incidiendo especialmente en la calidad de los datos de la tabla de clientes. A veces se introduce un mismo cliente con diferentes códigos y a menudo el personal encargado de introducir los datos en las aplicaciones da de alta un cliente y utilizan siempre ese código de cliente, cuando posiblemente ya esté dado de alta en el sistema.

### **Siguientes pasos**

Los datos consensuados en una única fuente abren la posibilidad de nuevos análisis, por ejemplo se ha observado que sería interesante ampliar el modelo incluyendo datos procedentes de las operativas de carga y descarga de los buques, lo que podría desembocar en una optimización de estas operativas tan costosas y cruciales para la naviera. Se desea además incluir información procedente de las unidades organizativas de transporte terrestre y distribución a fin de poder brindar al cliente un mejor servicio, pudiendo informar con facilidad de cuáles son los tiempos de entrega previstos de su mercancía en función de datos históricos. A fin de mejorar este punto se propone tener los metadatos dentro de la estructura del datamart a fin de observar fácilmente cómo los datos se integran y transforman antes de ser almacenados.

Otras preguntas que han surgido es si sería posible realizar clasificaciones de los clientes y de la carga por puertos, preguntas para las que el modelo está preparado para responder aplicando sobre él técnicas de clustering y árboles de decisión.

Se ha conseguido una recomendación a seguir en los proyectos de datawarehousing y datamining, y es que en las siguientes iteraciones el proyecto pase de ser una iniciativa de IT a ser una iniciativa del negocio<sup>11</sup>, permitiendo el desarrollo incremental y manejable para asegurar la flexibilidad del modelo y haciendo que los cambios en los procesos del negocio se reflejen en el diseño del modelo.

Adicionalmente se plantea la posibilidad de la adquisición de un software generador y distribuidor de informes por correo electrónico para que la información circule apropiadamente desde los sistemas hasta los usuarios, permitiendo además la gestión por excepción avisando de la ocurrencia de un evento importante en el negocio como la superación de umbrales y ratios y la comparación de ventas con previsiones y gastos con presupuestos. De esta forma los directivos y mandos pueden centrarse en sus tareas productivas sabiendo que el sistema de reporting les alertará ante situaciones alarmantes.

### 3.3.6 Despliegue:

Se trata de explotar la potencialidad de los modelos, integrarlos en los procesos de toma de decisión de la organización, difundir informes sobre el conocimiento extraído, etc.

Subfases:

- ✚ planificación del despliegue (plan del despliegue),
- ✚ planificación de la monitorización y del mantenimiento (plan de la monitorización y del despliegue),
- ✚ generación del informe final (informe final, presentación final) y
- ✚ revisión del proyecto (documentación de la experiencia).

---

<sup>11</sup> David Sammon, Pat Finnegan. *The ten commandments of data warehousing. The DATABASE for Advances in Information Systems, volume 32, number 4, fall 2000.*

### 3.4 METODOLOGÍA MÉTRICA VERSIÓN 3 INTEGRÁNDOSE CON UNA METODOLOGÍA DE DESARROLLO Y EXPLOTACIÓN DE DATAWAREHOUSES

Pasos Principales de Construcción utilizando esta metodología son los siguientes:

- ✚ Determinar el área del negocio
- ✚ Determinar que sistemas transaccionales están asociados al área de negocio seleccionada.
- ✚ Definir el origen de los datos operacionales. El modelo de datos fuente.
- ✚ Documentar las bases de datos de los sistemas transaccionales.
- ✚ Determinar las superposiciones y relaciones entre los sistemas transaccionales.
- ✚ Especificar los requerimientos para el soporte a las decisiones (DSS y EIS).
- ✚ Estudiar y analizar diferentes herramientas para OLAP y DataMining. Herramientas de Acceso al componente de almacenamiento físico datawarehouse.
- ✚ Definir las diferentes customizaciones de las herramientas seleccionadas.
- ✚ Especificar los mapeos de datos entre el datawarehouse y los sistemas transaccionales.
- ✚ Definir el modelo de datos del datawarehouse. El modelo de datos conceptual datawarehouse.
- ✚ Diseñar el datawarehouse.
- ✚ Definir la Arquitectura tecnológica datawarehouse.
- ✚ Especificar la extracción, transformación y carga de datos entre los sistemas transaccionales y el Datawarehouse. En donde la extracción es la selección sistemática de datos operacionales usados para poblar el componente de almacenamiento físico datawarehouse; la transformación es el proceso para transformar, sumarizar y realizar otros cambios en los datos operacionales para reunir los objetivos de orientación a los temas a explotar y finalmente la carga es la inserción sistemática de datos en el componente de almacenamiento físico datawarehouse
- ✚ Definir las interfaces de Usuario para el acceso y explotación de los datos.

### **3.4.1 Planificación del sistema**

El objetivo de esta fase es obtener un marco de referencia para el desarrollo del sistema de información.

#### **3.4.1.1 Definición del plan de trabajo**

Se determinan los productos finales y las fechas previstas de obtención y entrega de los mismos. Se detallan las actividades, asignando participantes, tiempos y responsables de cada una de ellas.

### **3.4.2 Gestión de la Configuración del Sistema**

#### **3.4.2.1 Especificar la política de gestión de la configuración para los componentes del sistema.**

Se realiza la estrategia de gestión de configuración que regirá durante todo el proyecto. Se decide que circuito se sigue para los cambios como así también que componentes serán objeto de control de configuración.

### **3.4.3 Gestión de la Calidad del Sistema**

#### **3.4.3.1 Especificar la política de gestión de calidad para el sistema**

Se realiza un plan de control de calidad donde se especifica en términos generales la estrategia a seguir para evaluar la calidad en el sistema. Esta estrategia apunta fundamentalmente a las pruebas que se realizan, cómo se realizan, quiénes son los participantes y responsables.

### **3.4.4 Estudio de Viabilidad del Sistema.**

El objetivo del Estudio de Viabilidad del Sistema es el análisis de un conjunto concreto de necesidades para proponer una solución a corto plazo, que tenga en cuenta restricciones económicas, técnicas, legales y operativas.

#### **3.4.4.1 Establecimiento del Alcance del Sistema**

En esta actividad se estudia el alcance de la necesidad planteada por el cliente o usuario, o como consecuencia de la realización de un Plan, realizando una descripción general de la misma. Se determinan los objetivos, se inicia el estudio de los requisitos y se identifican las unidades organizativas afectadas estableciendo su estructura.

Se analizan las posibles restricciones, tanto generales como específicas, que puedan condicionar el estudio y la planificación de las alternativas de solución que se propongan

##### **Estudio de la Solicitud**

Se realiza una descripción general de la necesidad planteada por el usuario.

Antes de iniciar el estudio de los requisitos del sistema se establecen los objetivos generales del Estudio de Viabilidad.

##### **Identificación del Alcance del Sistema**

Se analiza el alcance de la necesidad planteada y se identifican las posibles restricciones.

##### **Especificación del Alcance**

En función del alcance del sistema y los objetivos del Estudio de viabilidad del Sistema, se determinan las actividades y tareas a realizar. En particular, hay que decidir si se realiza o no el estudio de la situación actual y, en el caso de considerarlo necesario, con qué objetivo. Se identifican los usuarios participantes de las distintas unidades organizativas afectadas para la realización del Estudio de Viabilidad del Sistema.

#### **3.4.4.2 Estudio de la Situación Actual**

La situación actual es el estado en el que se encuentran los sistemas de información existentes en el momento en el que se inicia su estudio.

Teniendo en cuenta el objetivo del estudio de la situación actual, se realiza una valoración de la información existente acerca de los sistemas de información afectados. En función de dicha valoración, se especifica el nivel de detalle con que se debe llevar a cabo el estudio. Como resultado de esta actividad se genera un diagnóstico, estimando la eficiencia de los sistemas de información existentes e identificando los posibles problemas y las mejoras.

 **Determinar el área del negocio a ser soportadas en la primera fase del desarrollo.**

En esta tarea se debe determinar el área de negocio donde se comienza a construir el sistema.

 **Descripción de los Sistemas de Información Existentes**

Determinar que sistemas transaccionales están asociados a las áreas de negocio seleccionadas.)

En esta tarea se describen los sistemas de información existentes afectados, según el alcance y nivel de detalle establecido en la tarea.

Valoración del Estudio de la Situación Actual.

 **Realización del Diagnóstico de la Situación Actual.**

Con el fin de elaborar el diagnóstico de la situación actual se analiza la información de los sistemas de información existentes, obtenida en la tarea anterior y se identifican problemas, deficiencias y mejoras.

Estas últimas deben tenerse en cuenta en la definición de los requisitos.

### **3.4.4.3 Definición de Requisitos del Sistema**

Esta actividad incluye la determinación de los requisitos generales, mediante un conjunto de sesiones de trabajo con los usuarios participantes. Una vez finalizadas, se analiza la información obtenida definiendo los requisitos y sus prioridades, que se añaden al catálogo de requisitos que sirve para el estudio y valoración de las distintas alternativas de solución.

 **Identificación de Requisitos.**

En esta actividad se enumeran los requisitos generales, los cuáles se deberán refinar en tareas posteriores.

 **Catalogación de Requisitos**

Se analiza la información obtenida en las sesiones de trabajo para la Identificación de Requisitos, definiendo y catalogando los requisitos (funcionales y no funcionales) que debe satisfacer el sistema.

 **Estudio de Alternativas de Solución.**

Este estudio se centra en proponer diversas alternativas que respondan satisfactoriamente a los requisitos planteados, considerando también los resultados obtenidos en el Estudio de la Situación Actual.

 **Preselección de Alternativas de Solución.**

Una vez definidos los requisitos a cubrir por el sistema, se estudian las diferentes opciones que hay para configurar la solución.

 **Descripción de alternativas de Solución.**

Estudiar y analizar diferentes herramientas para OLAP y datamining, herramientas de Acceso a Datos.

#### **3.4.4.4 Selección de la Solución.**

Antes de finalizar el Estudio de Viabilidad del Sistema, se debaten las ventajas de cada una de las soluciones descritas, con el fin de seleccionar la más adecuada.

Finalmente, se aprueba la solución o se determina su inviabilidad.

 **Evaluación de las Alternativas y Selección.**

Se exponen las mejores alternativas con su valoración y se selecciona una.

 **Aprobación de la Solución.**

Se aprueba o rechaza la solución seleccionada.

#### **3.4.5. Análisis del sistema de información.**

El objetivo de este proceso es la obtención de una especificación detallada del sistema de información que satisfaga las necesidades de información de los usuarios y sirva de base para el posterior diseño del sistema.

##### **3.4.5.1 Definición del sistema.**

Esta actividad tiene como objetivo efectuar una descripción del sistema, delimitando su alcance, estableciendo las interfaces con otros sistemas e identificando a los usuarios representativos. Algunas de las tareas de esta actividad se iniciaron ya en parte en el proceso de Estudio de Viabilidad del Sistema.

 **Determinación del Alcance del Sistema.**

En esta tarea se delimita el sistema de información, utilizando como punto de partida la descripción de la solución del proceso en el Estudio de Viabilidad del Sistema

#### **Identificación de los Usuarios Participantes y Finales.**

En esta tarea se identifican los usuarios participantes y finales, interlocutores tanto en la obtención de requisitos como en la validación de los distintos productos y la aceptación final del sistema.

#### **3.4.5.2 Establecimientos de Requisitos.**

El objetivo de esta actividad es obtener un catálogo detallado de los requisitos, a partir del cual se pueda comprobar que los productos generados en las actividades de modelización se ajustan a los requisitos de usuario.

#### **Obtención de Requisitos**

En esta tarea comienza la obtención detallada de información mediante sesiones de trabajo con los usuarios.

#### **Especificación de Casos de Uso.**

El objetivo de esta tarea es especificar cada caso de uso identificado en la tarea anterior.

##### ▪ **Análisis de requisitos**

En esta tarea se estudia la información capturada previamente en esta actividad, para detectar inconsistencias, ambigüedades, duplicidad o escasez de información, etc.

El análisis de los requisitos y de los casos de uso asociados permite identificar funcionalidades o comportamientos comunes, reestructurando la información de los casos de uso a través de las generalizaciones y relaciones entre ellos.

#### **Validación de Requisitos.**

Mediante esta tarea, los usuarios confirman que los requisitos especificados en el catálogo de requisitos, así como los casos de uso, son válidos, consistentes y completos.

### 3.4.5.3 Identificación de los subsistemas de análisis.

El objetivo de esta actividad es facilitar el análisis del sistema de información llevando a cabo la descomposición del sistema en subsistemas.

#### **Determinación de Subsistemas de Análisis.**

La descomposición del sistema en subsistemas debe estar, principalmente, orientada a los procesos de negocio, aunque también es posible adoptar otros criterios lógicos.

#### **Integración de Subsistemas de Análisis.**

Objetivo de esta tarea es la coordinación en la elaboración de los distintos modelos de análisis de cada subsistema, asegurando la ausencia de duplicidad de elementos y la precisión en la utilización de los términos del glosario.

### 3.4.5.4 Elaboración del modelo de datos.

El objetivo de esta actividad es identificar las necesidades de información de cada uno de los procesos que conforman el sistema de información, con el fin de obtener un modelo de datos que contemple.

Todas las entidades, relaciones, atributos y reglas de negocio necesarias para dar respuesta a dichas necesidades.

#### **Definir el origen de los datos operacionales.**

El objetivo de esta tarea es determinar el origen de los datos de los sistemas transaccionales. Se determina la o las bases de datos fuentes.

#### **Documentar las bases de datos de los sistemas transaccionales.**

En esta tarea se debe obtener el modelo de datos de las bases de datos fuente. Este modelo de datos se debe estudiar y analizar a efectos de tener conocimientos de las estructuras y objetos de BD intervinientes en los sistemas transaccionales.

#### **Determinar las superposiciones y relaciones entre los sistemas transaccionales.**

Se debe determinar las superposiciones o duplicidad de datos en las bases de datos fuentes, en caso de haber varias bases de datos como así también en caso de una sola base fuente.

**✚ Definir el modelo de datos del DW. El modelo de datos conceptual DW.**

A partir de los modelos de datos fuentes y según requisitos se define el modelo de datos del Datawarehouse y Datamarts.

**✚ Especificar los mapeos de datos entre el DW y los sistemas transaccionales.**

Se realizan los mapeos entre los modelos de datos fuentes y el datawarehouse/datamart. Esta tarea básicamente es de refinamiento del modelo de datos del Datawarehouse/Datamarts contemplando los procesos de extracción, transformación y carga entre las diferentes bases de datos.

### **3.4.5.5 Definir las interfaces de usuario.**

El objetivo es realizar un análisis de los procesos del sistema de información en los que se requiere una interacción del usuario, con el fin de crear una interfaz que satisfaga todos los requisitos establecidos, teniendo en cuenta los diferentes perfiles a quienes va dirigido.

**✚ Especificación de Principios Generales de la Interfaz.**

El objetivo de esta tarea es especificar los estándares, directrices y elementos generales a tener en cuenta en la definición de la interfaz de usuario, tanto para la interfaz interactiva.

**✚ Definir las diferentes personalizaciones de las herramientas seleccionadas.**

Se comienza a especificar las personalizaciones de las herramientas en términos de funcionalidad.

**✚ Especificar la extracción, transformación y carga de datos entre los sistemas.**

Se especifican los procesos etls de extracción, transformación y carga entre las bases de datos fuentes y bases de datos destino. También se definen, si es necesario procesos etls entre el Datawarehouse y Datamarts.

#### **3.4.5.6 Análisis de consistencia y especificación de requisitos.**

El objetivo de esta actividad es garantizar la calidad de los distintos modelos generados en el proceso de Análisis del Sistema de Información, y asegurar que los usuarios y los Analistas tienen el mismo concepto del sistema.

##### **🚦 Verificación de los modelos.**

El objetivo de esta tarea es asegurar la calidad formal de los distintos modelos, conforme a la técnica seguida para la elaboración de cada producto.

##### **🚦 Validación de los Modelos.**

El objetivo de esta tarea es validar los distintos modelos con los requisitos especificados para el sistema de información, tanto a través del catálogo de requisitos, mediante la traza de requisitos, como a través de la validación directa del usuario.

#### **3.4.5.7 Aprobación del análisis de información.**

En esta tarea se aprueba el Análisis del Sistema.

##### **🚦 Presentación y Aprobación del Análisis del Sistema de Información.**

En esta tarea se realiza la presentación del análisis del sistema de información para la aprobación final del mismo.

#### **3.4.6 Diseño del sistema de información.**

El objetivo del proceso de Diseño del Sistema de Información es la definición de la arquitectura del sistema y del entorno tecnológico que le va a dar soporte, junto con la especificación detallada de los componentes del sistema de información.

##### **3.4.6.1 Definir la arquitectura del sistema.**

En esta actividad se define la arquitectura general del sistema de información, especificando las distintas particiones físicas del mismo, la descomposición lógica en subsistemas de diseño y la ubicación de cada subsistema en cada partición, así como la especificación detallada de la infraestructura tecnológica necesaria para dar soporte al sistema de información.

#### **Identificación de Subsistemas de Diseño**

En esta tarea se divide de forma lógica el sistema de información en subsistemas de diseño, con el fin de reducir la complejidad y facilitar el mantenimiento. Hay que tomar como referencia inicial los subsistemas de análisis especificados en el proceso de Análisis del Sistema de Información.

#### **3.4.6.2 Diseño de los casos de uso real.**

Esta actividad tiene como propósito especificar el comportamiento del sistema de información para un caso de uso, mediante objetos o subsistemas de diseño que interactúan.

#### **Diseño de la Realización de los Casos de Uso.**

El objetivo de esta tarea es definir cómo interactúan entre sí los objetos identificados en la tarea anterior para realizar, desde un punto de vista técnico, un caso de uso del sistema de información. Para ello, se parte de los escenarios especificados en el análisis, y se detallan teniendo en cuenta que se deben llevar cabo sobre un entorno tecnológico concreto.

#### **3.4.6.3 Diseño físico de los datos.**

En esta actividad se define la estructura física de datos que utilizará el sistema, a partir del modelo lógico de datos.

#### **Diseño del Modelo Físico de Datos.**

El objetivo de esta tarea es realizar el diseño del modelo físico de datos a partir del modelo lógico de datos o de clases/entidades.

#### **Definir las interfaces de Usuario para el acceso y explotación de los datos.**

En esta tarea se especifica las interfaces de usuario para el acceso a los datos.

#### **Diseñar la extracción, transformación y carga de datos entre los sistemas transaccionales y el Datawarehouse.**

Se especifica con detalle todos los componentes y parametrizaciones necesarias para realizar los procesos etls.

 **Verificación y Aceptación de la Arquitectura del Sistema.**

El objetivo de esta actividad es garantizar la calidad de las especificaciones del diseño del sistema de información y la viabilidad del mismo.

 **Verificación de las Especificaciones de Diseño.**

El objetivo de esta tarea es asegurar la calidad formal de los modelos utilizados.

#### **3.4.6.4 Aprobación del diseño del sistema de información.**

En esta actividad se aprueba el diseño para seguir avanzando en la construcción del sistema.

 **Presentación y Aprobación del Diseño del Sistema de Información.**

En esta tarea se realiza la presentación del diseño del sistema de información para la aprobación final del mismo.

#### **3.4.7 Construcción del Sistema de Información.**

En este proceso se genera el código de los componentes del Sistema de Información y las personalizaciones de las herramientas.

Para conseguir dicho objetivo, en este proceso se realizan las pruebas unitarias, las pruebas de integración de los subsistemas y componentes de acuerdo al plan de pruebas establecido.

 **Preparación del Entorno de Generación y Construcción.**

El objetivo de esta actividad es asegurar la disponibilidad de todos los medios y facilidades para que se pueda llevar a cabo la construcción del sistema de información. Entre estos medios, cabe destacar la preparación de los puestos de trabajo, equipos físicos y lógicos, gestores de bases de datos, herramientas de generación de código, bases de datos o archivos De prueba, entre otros.

 **Implantación de la Base de Datos Física.**

En esta tarea se debe implementar las base de datos y su gestor

 **Preparación del Entorno de Construcción.**

En esta tarea se prepara el entorno en el que se construirán los componentes del sistema de información.

 **Generar el código de los componentes y personalización de herramientas.**

En esta tarea se desarrollan y parametrizan en las herramientas correspondientes todas las especificaciones de diseño.

### **3.4.7.1 Ejecución de las pruebas unitarias.**

En esta actividad se realizan las pruebas unitarias de cada uno de los componentes del sistema de información, una vez codificados, con el objeto de comprobar que su estructura es correcta y que se ajustan a la funcionalidad establecida.

 **Preparación del entorno de las Pruebas Unitarias.**

En esta tarea se preparan todos los recursos necesarios para realizar las pruebas unitarias de cada uno de los componentes del sistema de información.

 **Realización y Evaluación de las Pruebas Unitarias.**

El objetivo de esta tarea es comprobar el correcto funcionamiento de los componentes del sistema de información codificados o personalizaciones de las herramientas. Para cada verificación establecida, se realizan las pruebas con los casos de pruebas asociados, efectuando el correspondiente análisis y evaluación de los resultados, y generando un registro conforme a los criterios establecidos en el plan de pruebas. Seguidamente, se analizan los resultados de las pruebas unitarias, evaluándose las mismas para comprobar que los resultados son los esperados. Si los resultados no son los esperados hay que proceder a realizar las correcciones pertinentes.

 **Ejecutar las pruebas de integración.**

El objetivo de las pruebas de integración es verificar si los componentes o subsistemas interactúan correctamente a través de sus interfaces, tanto internas como externas, cubren la funcionalidad establecida, y se ajustan a los requisitos especificados en las verificaciones correspondientes.

 **Preparación de las Pruebas de Integración.**

En esta tarea se disponen todos los recursos necesarios para realizar las pruebas de integración de los componentes y subsistemas que conforman el sistema de información.

 **Realización de las Pruebas de Integración.**

El objetivo de esta tarea es verificar el correcto funcionamiento de las interfaces existentes entre los distintos componentes y subsistemas, conforme a las verificaciones establecidas para el nivel de pruebas de integración.

 **Evaluación del Resultado de las Pruebas de Integración.**

El objetivo de esta tarea es analizar los resultados de las pruebas de integración y efectuar su evaluación. Dicha evaluación recoge el grado de cumplimiento de las pruebas.

### **3.4.7.2 Aprobación del Sistema de Información.**

En esta tarea se recopilan los productos del sistema de información y se presentan para su aprobación.

 **Presentación y Aprobación del Sistema de Información.**

Se realiza una presentación del sistema para su aprobación.

## CAPÍTULO IV

### DESARROLLO DE LA METODOLOGÍA PROPUESTA

#### 4.1 INTRODUCCIÓN

En la actualidad existen algunas metodologías para desarrollar proyectos de Business intelligence, se realizó un estudio de las metodologías existentes pero ninguna se ajusta a las necesidades de los usuarios dando una guía práctica para desarrollar un Data Warehouse.

En este proyecto se propone la Metodología para Desarrollar Proyectos de Business Intelligence “MEDEPROBI” se basa en el modelo en espiral ya que las fases son de carácter evolutivo e incremental, donde cada una de ellas trabaja con técnicas, herramientas, recursos, entradas y salidas.

Cada fase tiene su propósito, actividades, resultados y descripción de las actividades. Las fases de la metodología son seis: Planificación del Proyecto, definición de requerimientos, Diseño de la Arquitectura, Construcción y Desarrollo, Pruebas e Implantación, Administración y Mantenimiento del Sistema.

Una solución de Business Intelligence parte de los sistemas de origen de una organización (bases de datos, ERPs, ficheros de texto...), sobre los que suele ser necesario aplicar una transformación estructural para optimizar su proceso analítico.

Para ello se realiza una fase de extracción, transformación y carga (ETL) de datos. Esta etapa suele apoyarse en un almacén intermedio (datawarehouse), que actúa como pasarela entre los sistemas fuente y los sistemas destino (generalmente un datamart), y cuyo principal objetivo consiste en evitar la saturación de los servidores funcionales de la organización.

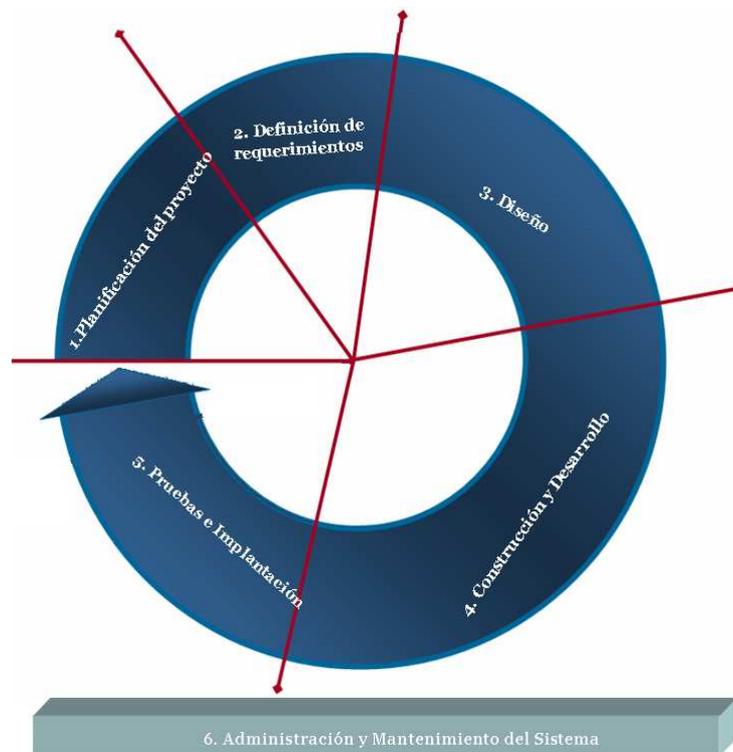
La información resultante, ya unificada, depurada y consolidada, se almacena en un datawarehouse corporativo, que puede servir como base para la construcción de distintos datamarts departamentales.

Estos datamarts se caracterizan por poseer la estructura óptima para el análisis de los datos de esa área de la empresa, ya sea mediante bases de datos transaccionales (OLTP) o mediante bases de datos analíticas (OLAP)

## 4.2 ETAPAS DE LA METODOLOGÍA PARA DESARROLLAR PROYECTOS DE BUSINESS INTELLIGENCE.

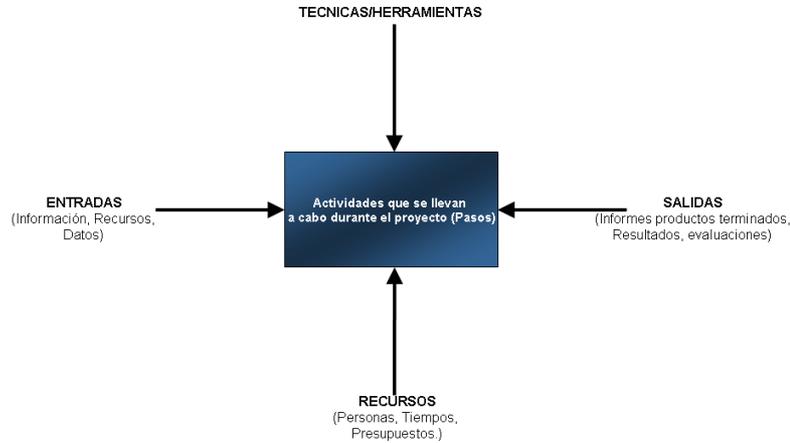
En la figura IV.30 se observa el ciclo de vida propuesto, Metodología para Desarrollar Proyectos de Business intelligence “MEDEPROBI”, esta figura muestra el proceso general donde cada cuadrante representa una fase que el desarrollador debe seguir, determinará donde empieza y hacia dónde se dirige, y en algún momento determinará en qué lugar se encuentra.

La metodología consta de las siguientes fases: Planificación del proyecto, Definición de Requerimientos, Diseño de la Arquitectura, Construcción y Desarrollo, Pruebas e Implantación y Administración y Mantenimiento.



**Figura IV.30 - Etapas de la Metodología para Desarrollar Proyectos de Business Intelligence**

Cada fase tiene un esquema que contempla Entradas, Salidas, Recursos y Técnicas/herramientas. De la siguiente manera:



**Figura IV.31 - Esquema de cada fase.**

En cada fase se aplicará múltiples técnicas y herramientas como las que propone el siguiente cuadro.

**Tabla IV-2. Técnicas y Herramientas propuestas**

FASE	TECNICAS	HERRAMIENTAS
Planificación del proyecto	<ul style="list-style-type: none"> <li>- Caso de uso</li> <li>- UML (Diagramas de Casos de uso)</li> <li>- Diagramas de Pert</li> </ul>	<ul style="list-style-type: none"> <li>- Rational Rose 2000</li> <li>- Erwin</li> <li>- Microsoft Office Visio</li> <li>- Microsoft Office Word</li> </ul>
Definición de Requerimientos	Encuestas, entrevistas, observación, cuestionarios	Microsoft Office Word
Diseño	<ul style="list-style-type: none"> <li>- Casos de uso</li> <li>- UML(Diagramas de caso de uso, diagramas de estados, diagrama de componentes, diagrama de nodos)</li> <li>- Modelado de datos</li> </ul>	<ul style="list-style-type: none"> <li>- Microsoft Office Visio</li> <li>- Microsoft Office Word</li> <li>- Racional Rose 2000</li> <li>- Power Designer 11</li> </ul>
Construcción y desarrollo	Diccionarios de Datos	<ul style="list-style-type: none"> <li>- Herramientas de Inteligencia de Negocios Comercial u Open Source</li> <li>- Sistemas Manejadores de Bases de Datos</li> </ul>

**Tabla IV-2. Técnicas y Herramientas propuestas(continuación)**

Pruebas e implantación	Pruebas Unitarias, pruebas de Integración.	- Sistemas Manejadores de Base de Datos - Microsoft Office Visio
Administración y mantenimiento del sistema	Exposición, preguntas y respuestas, demostración	Microsoft Office Word

#### **4.2.1 FASE I: PLANIFICACIÓN DEL PROYECTO**

La planificación del proyecto es la primera fase de la Metodología para Desarrollar Proyectos de Business Intelligence “MEDEPROBI”, por ende como en todo proyecto es necesario realizar una planificación, la misma que determinará el éxito o el fracaso del proyecto, es necesario poner énfasis en esta fase además nos permite determinar si es factible o no el proyecto.

##### **4.2.1.1 PROPOSITO:**

El propósito de esta fase es realizar un estudio de la situación actual del negocio, es decir conocer los antecedentes históricos y tecnológicos, recursos con los que como recursos humanos, técnicos, realizar una previa planificación del proyecto, determinar si es o no factible y sobre todo determinar los posibles riesgos y alternativas de solución, es importante saber que contamos con el apoyo de los ejecutivos para llevar a cabo este proyecto caso contrario no podemos avanzar con el mismo.

#### 4.2 1.2 DESCRIPCIÓN:

Una descripción grafica de la fase como muestra la siguiente figura.

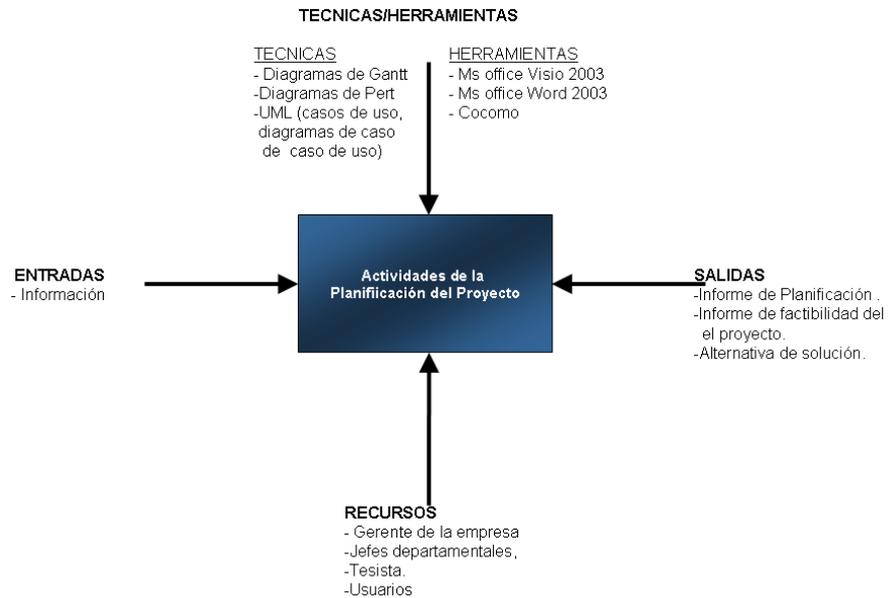


Figura IV.32 - Descripción de la fase.

#### 4.2.1.3 RESULTADOS:

- ✚ Valoración general del negocio y alternativa de solución.

#### 4.2.1.4 ACTIVIDADES.

Este gráfico nos muestra las actividades de esta fase.



Figura IV.33 - Actividades de la planificación del proyecto.

##### 4.2.1.4.1 Definir el ámbito del negocio.

En esta fase se realizará un estudio detallado de la empresa donde se conocerá todo lo relacionado a los antecedentes históricos y tecnológicos, es decir misión, visión, estructura funcional y tecnológica.

#### 4.2.1.4.2 Analizar los escenarios problemas.

Para determinar el escenario o los escenarios problema se utilizará casos de uso, donde un caso de uso es una técnica narrativa para definir requerimientos y escenarios del sistema en un lenguaje natural.

**Tabla IV-3. Caso de uso**

Caso de uso:	Nombre del caso de uso.
Actores:	Quienes participan en el caso de uso mencionado.
Descripción:	Narrar en lenguaje natural en que consiste el caso de uso.

#### 4.2.1.4.3 Determinar alternativas de Solución.

Luego de identificar el escenario problema en esta actividad realizamos una descripción general de la solución a los problemas encontrados en lenguaje natural, de igual forma utilizaremos casos de usos y diagramas de casos de uso.

#### 4.2.1.4.4 Analizar unidades organizativas afectadas.

En este apartado se detalla las unidades organizativas afectadas y sus responsabilidades.

**Tabla IV.4 - Unidades organizativas.**

<b>UNIDADES ORGANIZATIVAS</b>	<b>RESPONSABILIDADES</b>
Autoridades de la Empresa	Avalan u promueven el proyecto y donde a demás son los principales usuarios.
Usuarios	El área de negocio designada para comenzar el proyecto, el Departamento de Ingeniería. El usuario tomador de decisiones es el Director del Departamento y sus colaboradores.
Directora y miembro de la tesis	Guían al tesista y tienen funciones de revisión y auditoria.

**Tabla IV.4 - Unidades organizativas (continuación)**

Personal de la dirección de Sistemas	Dar soporte informático, los participantes son el director de la dirección y el responsable del mantenimiento del Sistema.
Tesista	Principal responsable de llevar adelante el proyecto. Sus funciones van desde la planificación del proyecto, definición de requerimientos, diseño de la arquitectura, construcción y desarrollo, pruebas e implantación, administración y mantenimiento y la documentación de la tesis para su presentación.

**4.2.1.4.5 Definir los costos.**

En este apartado realizar una estimación del esfuerzo de desarrollo a través de los puntos de función o COCOMO, costos del entorno de desarrollo para este costo tomar en cuenta el hardware, software y recursos humanos.

**4.2.1.4.6 Realizar el cronograma de trabajo.**

En esta actividad desarrollar el cronograma de trabajo, haciendo uso de los diagramas de Gantt.

**Tabla IV.5 – Cronograma de trabajo**

Id	Tarea

**4.2.1.4.7 Analizar los riesgos.**

En este apartado se describen, se categorizan los riesgos y realizar un plan de contingencia.

**Tabla IV.6 - Descripción de riesgos:**

RIESGO	DESCRIPCION

**Tabla IV.7 - Categorización de riesgos:**

RIESGOS	PROBABILIDAD		
	ALTA	MEDIA	BAJA

**Tabla IV.8 - Plan de contingencia.**

RIESGO	PLAN DE GESTION
R <sub>n</sub>	Problema:
	Actividades a desarrollar:
	Responsable:

**4.2.1.4.8 Realizar estudio de factibilidad.**

Se realizara un estudio de viabilidad para determinar si es factible o no realizar el proyecto.

**Factibilidad Técnica:**

Determinar los recursos hardware y software para implantar la solución.

**Tabla IV.9 - Factibilidad técnica.**

<b>HARDWARE</b>	<b>COSTOS</b>
<b>SOFTWARE</b>	<b>COSTOS</b>
<b>HUMANOS</b>	<b>COSTOS</b>
<b>TOTAL</b>	

**Factibilidad Operativa.**

Demostrar que los tiempos de respuesta del sistema propuesto son mucho mejores que el sistema actual.

**Tabla IV.10 - Factibilidad Operativa**

SISTEMA ACTUAL		SISTEMA PROPUESTO	
ACTIVIDADES	TIEMPOS	ACTIVIDADES	TIEMPOS
<b>TOTAL</b>			

**Factibilidad Económica.**

Es relación entre costos y beneficios al desarrollar las operaciones, cuando hablamos de beneficios nos referimos a cero errores, seguridad, efectividad y eficiencia. Costos implica todos los gastos directos e indirectos en la ejecución y mantenimiento del sistema.

En este apartado se representará la curva del sistema actual y el sistema propuesto.

**Tabla IV.11 - Factibilidad Económica.**

SISTEMA PROPUESTO			
COSTO	VALOR	BENEFICIO	VALOR
<b>TOTAL</b>			

**Factibilidad Legal.**

Garantizar que todo esté dentro de la ley o constitución de la empresa y organismos que rigen a los mismos.

#### 4.2.2 FASE II: DEFINICIÓN DE REQUERIMIENTOS.

La fase de definición de requerimientos es el cerebro del proyecto, se deberá realizar un informe detallado de los requerimientos de los usuarios ejecutivos de esta fase depende el enfoque del proyecto.

##### 4.2.2.1 PROPÓSITO:

El propósito de esta fase es determinar los requerimientos que guiarán el proyecto, además se conocerá el ámbito del sistema, perspectivas del producto, funciones y limitaciones generales.

##### 4.2.2.2 DESCRIPCIÓN:

Una descripción grafica de la fase como muestra la siguiente figura.

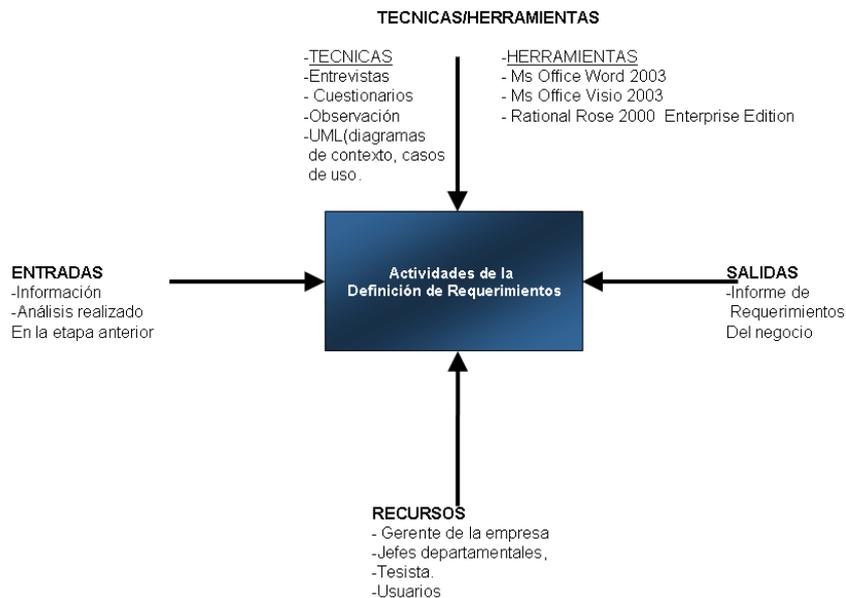


Figura IV.34 - Descripción de la fase.

##### 4.2.2.3 RESULTADOS:

🚩 Informe de los requerimientos del proyecto.

#### 4.2.2.4 ACTIVIDADES:

Es importante realizar cada una de las actividades como lo muestra la siguiente figura.



Figura IV.35 - Actividades de Definición de requerimientos.

##### 4.2.2.4.1 Mencionar técnicas utilizadas en la recopilación de requerimientos.

Al inicio de todo proyecto es necesario involucrarse en el proceso de funcionamiento de la empresa para ello hacer uso de ciertas técnicas de campo tales como: entrevistas, encuestas, observaciones, filmaciones, prototipos, etc.

##### 4.2.2.4.2 Describir los requerimientos.

Esta actividad es el pilar para el desarrollo del proyecto de Business Intelligence, engloba todas las necesidades de los distintos departamentos de la empresa, por lo tanto se clasificará a los requerimientos como: requerimientos funcionales, tecnológicos y humanos.

Los requerimientos se describirán en un lenguaje natural.

#### 4.2.3 FASE III: DISEÑO DE LA ARQUITECTURA.

Para realizar el diseño de la arquitectura se debe comenzar analizando los sistemas actuales, estos deben ser consistentes y manejar de forma correcta sus transacciones, pues para el desarrollo del datawarehouse se toma como hecho que estos sistemas son confiables.

##### 4.2.3.1 PROPÓSITO:

El propósito de esta fase es diseñar la arquitectura con la que se desarrollará el proyecto, es decir identificar los principales componentes que serán necesarias utilizar permitiendo de esta manera hacer un desarrollo más confiable y eficiente del proyecto.

#### 4.2.3.2 DESCRIPCIÓN:

Una descripción gráfica de la fase como muestra la siguiente figura.

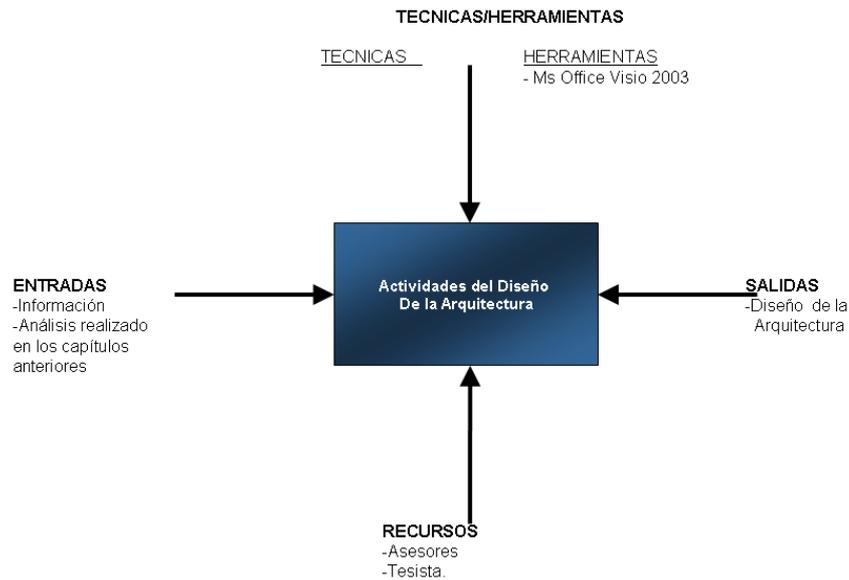


Figura IV.36 - Descripción de la fase.

#### 4.2.3.3 RESULTADOS:

🚩 Diseño de la arquitectura del proyecto.

#### 4.2.3.4 ACTIVIDADES:

A continuación se dará a conocer las principales actividades a desarrollar en esta fase como muestra la siguiente figura.



Figura IV.37 - Actividades del Diseño de la arquitectura.

#### 4.2.3.4.1 Diseñar la arquitectura.

En la actualidad existen diferentes arquitecturas para el diseño de un Datawarehouse, sin embargo la arquitectura que presentamos hace un enfoque global a todas y cada una de las capas de Business Intelligence.

Construir el sistema completo puede verse con facilidad como lo muestra la figura IV.38 Las flechas a la derecha indican “Sistema DW: Carga y control de calidad” y “Sistema DW: OLAP”. La primera se ocupa de construir el Data warehouse y los procesos de carga. La segunda se ocupa de construir las bases de datos multidimensionales y configurar las herramientas de usuario final.

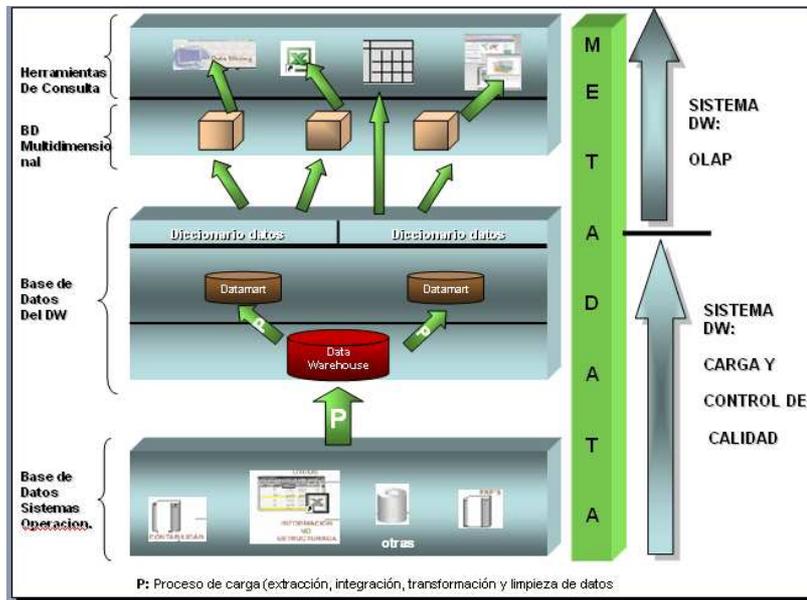


Figura IV.38 - Diseño de la Arquitectura.

#### 4.2.3.4.2 Identificar casos de uso.

En esta actividad será necesario identificar los casos de uso que intervendrán en la ejecución del proyecto, se utilizará una tabla con los siguientes campos: número y nombre del caso de uso, el evento y los actores.

Tabla IV.12 - Identificar casos de uso

No	Nombre del caso uso	Eventos	Actores

#### 4.2.3.4.3 Desarrollar Casos de Uso y Diagramas de Casos de Uso del sistema Propuesto.

En esta actividad a través de los casos de uso y los respectivos diagramas describiremos como se va llevar a cabo el sistema propuesto.

#### 4.2.3.4.4 Determinar los datos fuentes.

Se mencionará todos los sistemas fuentes que serán necesarios para realizar el proyecto de Inteligencia de Negocios.

En esta actividad es importante familiarizarse con los datos de los Sistemas Transaccionales u otros archivos que están relacionados a las actividades de la empresa, por tal razón hacer referencia al Diagrama Entidad Relación.

En esta actividad se realizará dos tablas, la primera que constará de tres columnas donde se indicará el número, nombre y una descripción de la entidad como lo indica la siguiente tabla.

**Tabla IV.13 - Tablas fuentes**

No	Nombre de la entidad	Descripción

La segunda tabla consta de las siguientes columnas: nombre de la entidad con sus respectivos atributos y tipo de datos, en la posterior columna hacer una breve descripción de la entidad fuente.

**Tabla IV.14 - Modelo de datos fuente**

Nombre de la entidad	Atributos	Tipo de Dato	Descripción

#### 4.2.3.4.5 Definir Base de Datos Intermedia de la bodega de datos.

La elección de una estructura de Datawarehouse responde a diferentes necesidades: En el caso de necesitar cruzar datos lo natural es crear un **Datawarehouse** que soporte los requerimientos de los usuarios.

Si cada departamento necesita información, es posible desarrollar los **Datamarts**, donde cada datamart satisfaga las distintas necesidades departamentales.

En esta actividad es importante realizar una base de datos intermedia, ya que la misma tendrá datos limpios, esta base de datos la obtendremos luego de realizar los proceso de Extracción, Transformación y Carga (ETL), que nos servirá para realizar el Datawarehouse o Datamarts.

En esta actividad se utilizarán dos tablas, en la primera se definirán las entidades que se tomarán en cuenta en la creación de la Base, la misma que consta de los siguientes campos: nombre de la entidad de la Base de Datos fuente, nombre del Datawarehouse y una descripción de cada una de las entidades.

**Tabla IV.15 - Descripción Bases de Datos Fuente y Base de Datos Intermedia**

Nombre de la entidad BD fuente	Nombre de la entidad BD intermedia	Descripción de las entidades de la BD intermedia

En la segunda tabla se registrarán las entidades de la Base de Datos Fuente y Base de Datos Intermedia.

En la primera columna se anotará el nombre de la entidad fuente "Bd\_fuente", seguido de sus respectivos campos, en la segunda columna se escribirá el nombre de la nueva entidad "Bd\_intermedia", seguida de los campos seleccionados para construir el Datawarehouse o Datamarts.

**Tabla IV.16 – Detalle de los campos de la BD Fuente - Destino**

BASE DE DATOS SISTEMA	
FUENTE	DESTINO
<b>Bd_fuente</b> Campo1 Campo2	<b>Bd_intermdia</b> Campo1

#### 4.2.3.4.6 Realizar el modelo dimensional.

El modelo dimensional es una técnica de diseño lógico que busca presentar los datos de una forma intuitiva y que proporcione acceso de alto desempeño, es decir en esta actividad se elabora el modelo lógico, que soportará las consultas, mediante las cuales se obtendrá la información requerida por los niveles directivos como apoyo a sus procesos de gestión y de toma de decisiones.

En esta actividad realizamos los siguientes pasos:

##### Definir las tablas de hechos

En un proyecto se puede definir una o más tablas de hechos, la tabla de hechos o “fact table”, debe tener una estructura que se divide en dos partes:

La primera de ellas debe contener las llaves foráneas que van a relacionar a las dimensiones. Por cada dimensión que exista asociada a una tabla de hechos, debe aparecer una llave foránea a dicha dimensión. Esto permite que al momento de utilizar los hipercubos se involucre información contenida en los atributos de las dimensiones que no estén almacenados en la tabla de hechos.

La segunda son los campos que almacenan la información concreta del registro y particular al problema y que permiten hacer las sumalizaciones que se requieran en la aplicación, este atributo se conoce como las Medidas.

Las medidas son las que miden una determinada variable de estudio y por lo general son cantidades de cosas (cantidades de ventas, de dinero, etc). Estas medidas pueden ser: *Medida regular* toma su dato directamente de una fuente disponible. Es un compendio de información que ya se tiene, tal como el número de unidades vendidas, ingresos, gastos, niveles de inventario. Y la *medida calculada* obtiene como resultado un nuevo dato numérico para medidas que no están en una fuente directa disponible. Es derivada de otras medidas.

Ejemplo de este tipo de medidas son: ganancia (ingresos - costos), margen de ganancia (ingreso – costo/ingreso), tiempo promedio de espera (fecha de entrega- echa de la orden), etc.

- ✚ **Identificar para cada tabla de hechos, las dimensiones que las referencian.**

Para cada tabla de hechos se identifican las dimensiones que se va utilizar en el desarrollo del proyecto, es importante tomar en cuenta los requerimientos de la fase II, en este paso, se visualizará cada tabla de hechos(eje X) con sus respectivas dimensiones(eje Y), , tal como se esquematiza en la Figura 4-13.

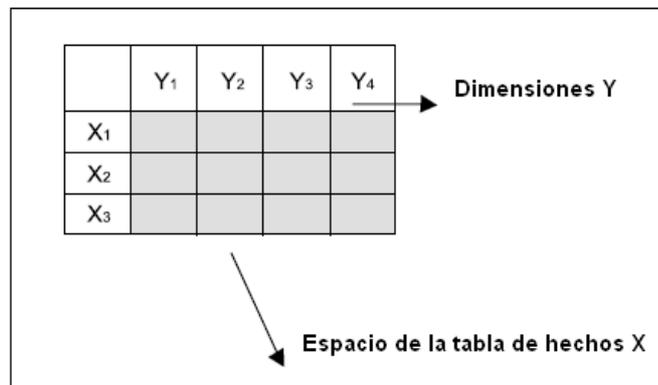


Figura IV.39 - Esquema de una vista multidimensional de las tablas de hecho.

- ✚ **Establecer el nivel de granulación de cada dimensión y los niveles de agregación de cada tabla de hechos.**

Una vez que las dimensiones han sido identificadas se debe establecer, para cada una de ellas, el menor nivel de granulación, el cual corresponde al conjunto de atributos que referencian el mayor nivel de detalle deseado para la dimensión o tabla de hechos. La dimensión tiempo provee un claro ejemplo, el menor nivel de granulación requerido sería el día.

Una vez que se han definido los menores niveles de granulación para cada dimensión, se identifican los niveles de agregación requeridos para los valores almacenados en la tabla de hechos, por cada dimensión. Estos niveles de agregación representan la jerarquía de cada dimensión. La dimensión tiempo sería un ejemplo del uso de jerarquías, donde el año es un nivel superior, luego sigue el semestre, trimestre, mes y por último el nivel mas inferior sería el día.

En esta actividad se utilizará dos tablas:

En la primera se indicará como llenar la granulación de las dimensiones, en la columna inicial se escribirá el nombre de la dimensión y en la segunda los campos que componen la dimensión, como lo indica la siguiente tabla.

**Tabla IV.17 – Granulidad**

<b>Dimensión</b>	<b>Granulación</b>

En la segunda tablas se anotará las Fact Tables con sus respectivas agregaciones, en la primera columna se escribirá el nombre de la Fact Tables y en la segunda se registrará los nombres de las medidas calculadas y no calculadas, como lo indica la siguiente tabla.

**Tabla IV.18 – Tablas de Hecho y medidas**

<b>Tablas de Hecho</b>	<b>Agregación</b>

**✚ Elaborar diagrama en estrella o copo de nieve que represente el datamart.**

Luego de identificar los elementos que conforman la estructura de la vista multidimensional, de la información gerencial requerida por la organización, se pasa a la elaboración de una representación gráfica, en forma de estrella o copo de nieve; para ello se puede utilizar la notación simplificada de los diagramas E-R.

En esta notación, el diagrama en estrella está conformado por una entidad central asociativa, que corresponde a la tabla de hechos, y por un conjunto de trayectorias de entidades y relaciones de uno a muchos, que corresponde a las dimensiones y a sus jerarquías. En la Figura IV.40 se presenta un diagrama en estrella con esta notación. Y en la figura IV.41, se presenta un diagrama en copo de nieve.

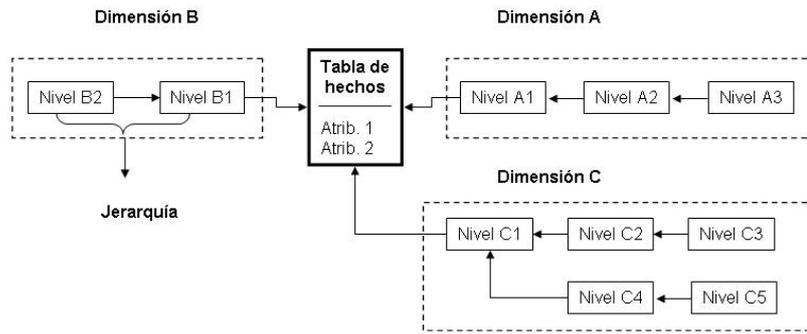


Figura IV.40 - Representación de un diagrama en estrella.

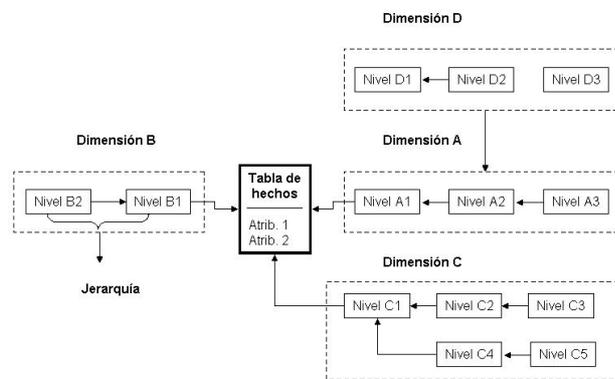


Figura IV.41 - Representación de un diagrama copo de nieve.

#### 4.2.3.4.7 Elaborar el modelo físico.

Durante esta actividad, se realizará la transformación del modelo lógico conceptual en la estructura física, que posteriormente será implementada. Este proceso de transformación se realiza mediante los siguientes pasos:

##### Verificación del modelo dimensional.

Realizar la verificación del modelo lógico, obtenido en la actividad anterior, para garantizar que el modelo soporte todas las consultas requeridas por los niveles ejecutivos, retornando siempre información confiable.

Para iniciar este proceso de verificación se debe elaborar una matriz de cruce, entre los requerimientos de información gerencial, definidos en la fase II y las estructuras (estrellas o copo de nieve), definidas en la actividad anterior. En la matriz de cruce se confirma si el requerimiento está completamente soportado. Si esta verificación no es correcta, se debe retornar a la actividad anterior, para

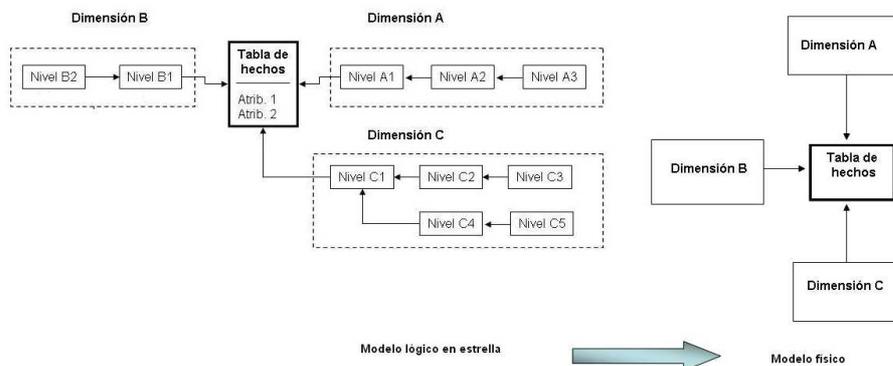
incorporar las estructuras que soporten los requerimientos faltantes de información.

**Tabla IV.19 - Matriz de verificación del modelo dimensional**

	Tablas de hechos	Dimensiones
	H_1,H_2, H_n	D_1,D_1,D_n
Requerimiento <sub>1</sub>		
Requerimieno <sub>2</sub>		
Requerimiento <sub>n</sub>		

 **Definición del esquema físico de las dimensiones y sus jerarquías.**

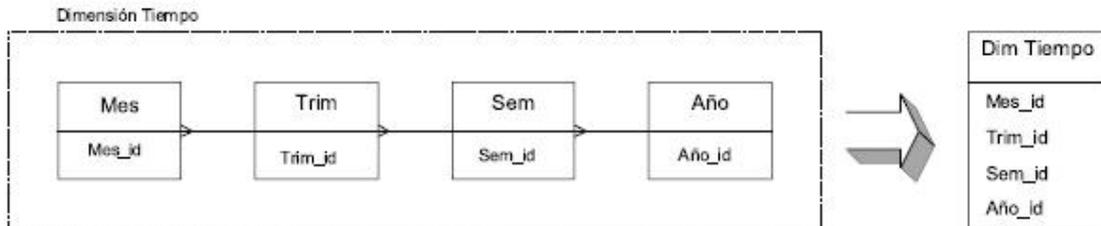
El modelo en estrella o copo de nieve que conforma la estructura lógica, debe ser convertido en una estructura totalmente desnormalizada, tal como se presenta en la Figura IV.42. Este modelo físico está conformado por una tabla de hechos y por las entidades en las cuales se almacenarán los dominios de las dimensiones con sus correspondientes niveles jerárquicos.



**Figura IV.42 - Modelo lógico y modelo físico**

En la conversión del modelo lógico al modelo físico la llave primaria de la dimensión se conforma como una llave compuesta por las llaves de cada una de las entidades que conforman la trayectoria de la dimensión.

En la Figura IV.43, se presenta, de manera gráfica, este esquema de conversión. Tomando en cuenta la dimensión tiempo con sus jerarquías.



**Figura IV.43 – Conversión del modelo lógico al físico**

**🚧 Definición de los atributos que conforman las tablas de hecho.**

En este paso se identifican para cada tabla de hechos y cada dimensión las características de los atributos que conforman cada estructura.

Una vez asignados todos los atributos, se realiza un análisis cruzado entre la tabla de hechos y las dimensiones, para establecer los tipos de cálculo matemático que pueden ser realizados, sobre la tabla de hechos.

En la siguiente tabla se hará constar en el nombre del Datamart, nombre de la tabla de hechos o dimensión, nombre del atributo, tipo, clave, y una descripción del atributo. Vea la siguiente tabla.

**Tabla IV.20 - Definición de los atributos de una tabla de hechos.**

Nombre del Datamart			
Tabla de hechos / Dimensión			
Atributo	Tipo	Pk	Descripción

#### 4.2.3.4.8 Diseñar cubos multidimensionales.

Los cubos multidimensionales trabajan con las tablas de hechos y las dimensiones, se basan en las representaciones en estrella o copo de nieve, para el diseño de los cubos multidimensionales nos basamos en el siguiente cuadro.

**Tabla IV.21 - Diseño de cubos multidimensionales**

Nombre del cubo:			
Requerimiento de usuario:			
Data source:			
Tabla fuente:			
Dimensiones	Nombre	Tabla fuente	
Medidas	Nombre	Campo fuente	Función
Miembros calculados	Nombre	Función de cálculo	
Tipo de almacenamiento			
Lectura/Escritura			
Roles			

Se usará herramientas de open source para desarrollar estos cubos, además el lenguaje utilizado para ejecutar cubos es MDX, que no es más que un lenguaje para realizar queries en base de datos multidimensionales, de la misma forma análoga al SQL en Base de datos relaciones.

Para conocer más detalles de los cubos multidimensionales y el lenguaje MDX ver Anexo 1.

### **Seleccionar los operadores de las Herramientas OLAP.**

Una vez diseñados los cubos multidimensionales se trabajará con los sistemas OLAP los mismos que se caracterizan por soportar análisis de los usuarios y unas posibilidades de navegación, seleccionando la información a obtener.

#### **Operaciones en cubos OLAP**

-  **Roll - Up (drill-up):** resumir los datos, esta operación permite subir en la jerarquía o reducir las dimensiones
-  **Drill-Down (roll-down):** contrario al anterior, esta operación permite bajar la jerarquía o introducir nuevas dimensiones.
-  **Slice and dice:** esta operación permite selección y proyección.
-  **Pívor (rotar):** permite reorientar el cubo.
-  **Drill:** utiliza las coordenadas dimensionales especificadas por un usuario para una celda en un cubo para moverse a otro cubo a ver información relacionada.
  - **Drill across:** implica más de una tabla de hechos.
  - **Drill through:** ir desde el nivel de máximo detalle del cubo a sus tablas relacionales.

#### **4.2.3.4.9 Especificar características de aplicaciones para usuarios finales.**

Una aplicación de usuario final, provee un diseño y estructura a los reportes tomando como base los datos del Datawarehouse.

#### **Especificación de aplicaciones para usuarios finales.**

Hay algunos pasos importantes en el proceso de especificación de las aplicaciones de usuario final.

-  Determinar el conjunto inicial de plantillas de reportes.
-  Determinar la navegación en los reportes.
-  Determinar el estándar de plantillas de reportes.
-  Determinar la especificación de estas plantillas.

### **Determinar el conjunto inicial de plantillas de reportes.**

Realizar una lista de reportes que se va a desarrollar en la aplicación, para esto realizar tres tareas:

- ✚ *Identificar los posibles reportes:* mirando la documentación del proyecto y los requerimientos del negocio. Con esta información se hace una lista de reportes a planificar. Por ejemplo tener una lista de los departamentos que serán involucrados en los reportes y análisis.
- ✚ *Consolidar la lista de candidatos:* una vez que se tiene la lista de reportes, se debe categorizar dicha lista. Las categorías se reconocen rápidamente debido a la forma como están estructuradas las compañías. Para analizar como categorizar los reportes se puede tener en cuenta las siguientes preguntas: ¿Cómo es el negocio?, ¿Cuál es la tendencia?, entre otra.
- ✚ *Priorizar la lista de reportes:* una vez establecida la lista de reportes, se debe definir con el usuario final para establecer prioridad en los reportes que ellos van a utilizar.

**Tabla IV.22 - Definición de plantillas de reportes**

No	Departamento involucrado	Lista de reportes

### **Definir la navegación para los reportes.**

En esta parte se establecen estándares en cuanto a los elementos de la Base de Datos. Esta estandarización es muy importante porque ayuda a entender la naturaleza de los reportes. La estandarización se enfoca principalmente a los elementos de datos del Datawarehouse que son utilizados para alimentar los reportes.

### **Detalles de especificación de las plantillas de reportes.**

Cada organización tiene un diferente grado de detalle en la documentación del proyecto. También se debe determinar si el equipo que trabaja en el proyecto hace formalmente la documentación de las aplicaciones. Si la respuesta es negativa se

debe presionar para que la tenga al día y a disposición para cualquier usuario que la quiera mirar.

*Formato de especificación:* la especificación de la aplicación para el usuario final tiene dos partes fundamentales.

- ✚ La definición: provee información básica acerca de la plantilla de reporte.
- ✚ El diseño: provee una representación visual de lo que debería mostrar el reporte.

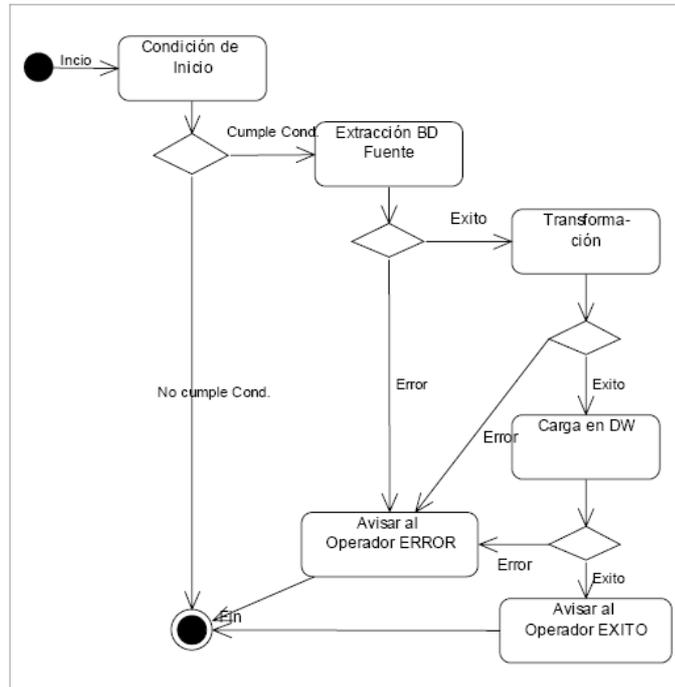
La definición de las plantillas de reportes incluye el nombre, la descripción o propósito, frecuencia con que debe ser actualizado, parámetros, entrada, entre otros.

**Tabla IV.23 - Formato de planilla de reporte**

Nombre de la planilla	
Descripción o propósito	
Requerimiento.	
Frecuencia de actualización	
Entradas	
Diseño de la plantilla	

#### **4.2.3.4.10 Desarrollar diagramas de secuencia de los procesos de carga.**

Se mostrará un diagrama de secuencia donde se detallan los diferentes procesos de la extracción, transformación y carga de los datos desde la base de datos del sistema fuente a la base de datos del Datawarehouse y a los Datamarts.



**Figura IV.44 - Proceso de carga.**

**Condición de Inicio:** El proceso comienza cuando el servicio detecta que se ha cumplido la condición de inicio de los procesos. La condición de inicio se programa indicando con qué frecuencia se debe ejecutar. Diariamente se verifica que la condición se cumpla de acuerdo a su programación temporal, en caso de no cumplirse no se ejecuta el proceso siguiente.

**Extracción BD Fuente:** Una vez que Condición de Inicio se cumple se ejecuta la extracción de los datos de la base de datos fuente. La extracción de los datos puede ser desde la base de datos fuente o desde la base de detalle dentro del datawarehouse. En caso que en el proceso de extracción surja un error el proceso lo detecta e informa al operador. Este informe es vía mensaje a la consola de principal del servidor. En caso de error el proceso se detiene, en caso de éxito se ejecuta el proceso Transformación.

**Transformación:** Este proceso realiza las transformaciones necesarias de los datos (en caso que sea necesario) antes de introducirlos a la base de datos destino. En caso de error, se informa al operador de la misma manera que en el proceso anterior y se detiene la ejecución. En caso de éxito se ejecuta el proceso Carga en DW.

**Carga en Data Warehouse:** Una vez extraídos y transformados, los datos son introducidos a la base de datos destino. Esta es la base de datos de Detalle y el Datamart. El manejo del error es igual que en los dos procesos anteriores. En caso de ejecución satisfactoria se informa al operador que el proceso terminó correctamente.

**Avisar al Operador ÉXITO:** Este proceso informa al operador por medio de un mensaje a la consola que el proceso completo terminó satisfactoriamente.

**Avisar al Operador ERROR:** Este proceso informa al operador por medio de un mensaje a la consola que en el proceso se produjo un error. Se informa el texto de error y el proceso en cuestión se detiene para que el operador resuelva el error.

La siguiente tabla muestra la especificación de los procesos ETL.

**Tabla IV.24 - Especificación de los procesos ETL**

Nombre del proceso ETL		
Fuente		
Destino		
Transferencias	Características de las transferencias	
Éxito		
Fracaso		
Actualización		
Tipo almacenamiento script ETL (archivos)		
Observaciones		

#### **4.2.3.4.11 Realizar Diagrama de componentes.**

A través de un diagrama de componentes demostrar la organización de los componentes, interfaces y las dependencias entre ellas.

#### **4.2.3.4.12 Definir diagrama de nodos.**

En esta actividad se diseñará la estructura física del sistema que se llevará a cabo. El objetivo es mostrar la disposición de las particiones físicas del sistema y la asignación de los componentes software a estas particiones. Es decir, las relaciones físicas entre los componentes software y hardware en el sistema a entregar.

#### 4.2.4 FASE IV: CONSTRUCCIÓN Y DESARROLLO

Esta fase constituye la ejecución del proyecto como se analizó en las fases anteriores empezamos con una planificación del proyecto, definición de requerimientos y el diseño de la arquitectura todo esto nos servirá de guía.

Para construir el Data Warehouse se trabajará con una base de datos auxiliar y datamart porque datamart por dos razones: una por consultas más rápidas y menos usuarios, dos por tiempo de desarrollo más rápido.

##### 4.2.4.1 PROPÓSITO:

Definir y desarrollar técnicas y herramientas que permitan desarrollar las actividades de esta fase para construir el proyecto.

##### 4.2.4.2 DESCRIPCIÓN:

Una descripción gráfica de la fase como muestra la siguiente figura.

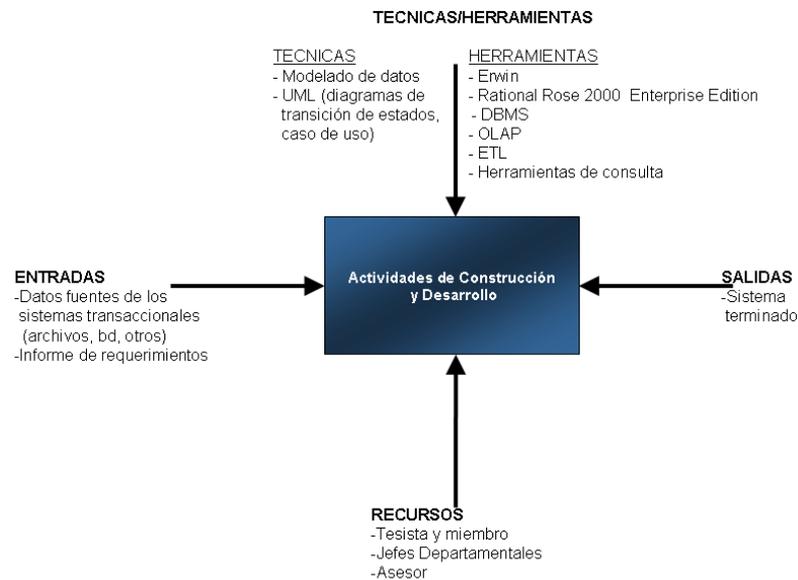


Figura IV.45 - Descripción de la fase.

##### 4.2.4.3 RESULTADOS:

- 🚩 Sistema terminado listo para realizar pruebas.

#### 4.2.4.4 ACTIVIDADES:

Actividades a desarrollar en esta fase son las siguientes como ilustra la siguiente figura.



Figura IV.46 - Actividades de Construcción y Desarrollo.

##### 4.2.4.4.1 Seleccionar herramientas.

Para seleccionar herramientas ETL, Base de datos y herramientas de visualización, se debe tener en cuenta dos componentes fundamentales para la selección de productos tales como los Requerimientos técnicos y Requerimientos de negocio.

Para la selección de la herramienta se debe hacer un riguroso estudio de cada una de ellas, mirar sus ventajas y desventajas y saber el alcance que se quiere con cada una para saber las cosas que se pueden o no se pueden hacer.

Otro punto importante en escoger la herramienta son los requerimientos de los usuarios: los requerimientos del negocio, mirar el análisis y el grado de procesamiento que se puede obtener con cada una ya sea un motor de base, una herramienta de ETL o una herramienta OLAP o de reportes. También se deben mirar las limitaciones de los usuarios en lo referente al hardware, pues no tiene sentido contratar una herramienta que consume muchos recursos cuando no se tiene la plataforma para soportarla.

Además hay que tomar en cuenta el plan del proyecto, la arquitectura para mirar los alcances técnicos y la documentación de los requerimientos que ayudan a la mejor opción.

**Áreas a evaluar para escoger las herramientas.**

Se tiene cuatro áreas fundamentales para evaluar y escoger las herramientas:

**Tabla IV.25 - Áreas a evaluar para escoger las herramientas.**

AREA	DESCRIPCIÓN
Plataforma de hardware	Se deben evaluar las diferentes plataformas para mirar su capacidad y escoger las herramientas teniendo en cuenta las limitaciones que podrían presentar. Las plataformas de servidores donde se ejecutarán las herramientas dimensiona, la plataforma donde tenemos los Datamarts y los demás servidores de aplicaciones deben ser analizadas cuidadosamente.
Plataforma de Base de Datos	Para proyectos pequeños se pueden utilizar un motor de BD sencillo como mysql que no es muy pesado y que podrán soportar las plataformas. Para proyectos grandes se debe considerar un motor de BD con más características y seguridad que permite realizar mejores aplicaciones por ejemplo PostgreSQL, Oracle, SQL Server, internase entre otras.
Herramientas de Almacenamiento	Esta es el área de trabajo del DWH donde la información se extrae de las fuentes, se realiza limpieza a los datos, se hacen las transformaciones y se envía a los datamarts. Estas son las herramientas más costosas y complejas del proyecto.
Herramientas de acceso a los datos	Esta elección es difícil porque en el mercado no existe un líder, y además se terminan usando varias de estas herramientas por la gran variedad de requerimientos que desea cumplir.

**4.2.4.4.2 Especificar permisos y accesos de los usuarios.**

Administrar los permisos y accesos de los usuarios a las distintas bases de datos y aplicaciones del sistema. La tabla IV-26. Muestra los permisos que tienen los distintos tipos de usuarios y estos permisos son lectura y escritura sobre las tablas o bases de datos.

**Tabla IV.26 - Permiso de base de Datos**

Tipo de usuarios	Permisos		
	Bases de Datos		

La tabla IV-27. Indica los accesos de los tipos de usuarios sobre las distintas aplicaciones tales como: DBMS, ETL, aplicaciones OLAP e interfaz de usuarios.

**Tabla IV.27 - Acceso a aplicaciones**

Tipo de usuarios	Acceso		
	Aplicaciones		

#### **4.2.4.4.3 Desarrollar el diseño**

En esta actividad será necesario ejecutar el diseño desarrollado en la fase anterior, es decir comenzar a desarrollar los procesos ETL, Datawarehouse y OLAP donde en cada etapa se realizará lo siguiente:

- Instalación y configuración de las herramientas seleccionadas.
- Ejecutar los procesos.

#### **4.2.5 FASE V: PRUEBAS E IMPLANTACIÓN.**

Esta fase es la encargada de realizar todas las pruebas posibles de tal manera que se demuestre la ejecución de los requerimientos establecidos en la fase II.

##### **4.2.5.1 PROPÓSITO:**

Lograr el correcto funcionamiento de los componentes del sistema, detectar errores a través de los cuales realizar las respectivas correcciones y conseguir que el sistema cumpla con el funcionamiento esperado y permita al usuario del sistema que determine su aceptación, desde el punto de vista de su funcionalidad y rendimiento.

#### 4.2.5.2 DESCRIPCIÓN:

Una descripción gráfica de la fase como muestra la siguiente figura.

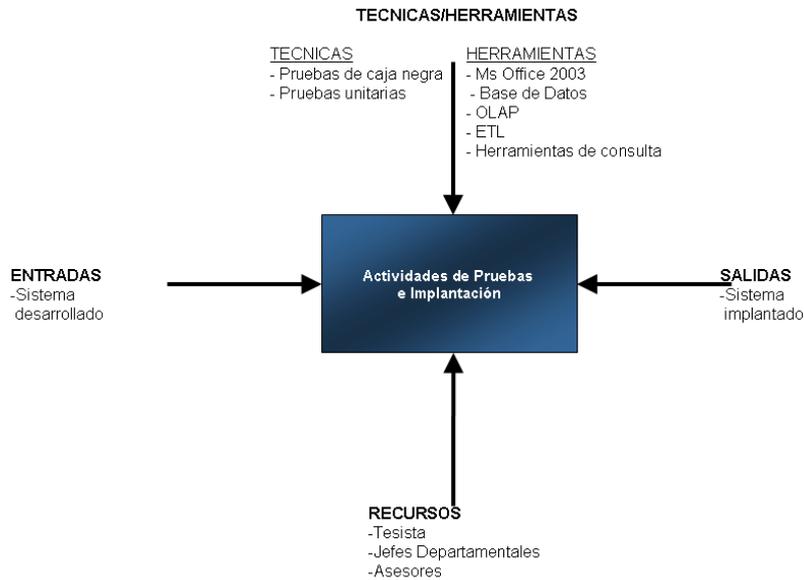


Figura IV.47 - Descripción de la fase.

#### 4.2.5.3 RESULTADOS:

✚ Sistema Implantado

#### 4.2.5.4 ACTIVIDADES:

Para llevar a cabo esta fase es necesario realizar una serie de actividades, como muestra la figura.



Figura IV.48 - Actividades de Pruebas e Implantación.

#### 4.2.5.4.1 Planificar las prueba.

En esta actividad se definen los objetivos de la prueba, los objetos probados, los alcances de la prueba, el método utilizado, los recursos empleados, el plan de tiempos y finalmente las responsabilidades para llevar adelante la prueba.

#### 4.2.5.4.2 Generar casos de pruebas

Los casos de prueba (para las pruebas funcionales o de caja negra) están basados en lotes de datos reales, los cuáles se han obtenido de la base de datos fuente. Para el diseño de los casos, se toma como base la lista de funciones, llamando lista de funciones a los requerimientos.

Las pruebas unitarias de cálculo se realizan en base a diversas consultas (instrucciones SQL) realizadas manualmente en la base de datos fuentes. De manera de poder conseguir resultados y valores para cotejarlos con los resultados que se obtienen utilizando business intelligence. Ver anexo 1.

En la fase de pruebas se muestran los casos de prueba completos y la integración con los formularios que se describen a continuación.

**Tabla IV.28 - Pruebas unitarias.**

No. De Caso de prueba	
Nombre del caso de prueba	
Instrucción SQL	
Resultado	

**Tabla IV.29 - Caso de prueba**

CASOS DE PRUEBA			
Caso número:	Proyecto:	Documento Soporte:	
	Sistema:		
Pre-requisitos del caso de prueba			
Entradas requeridas:			
Descripción de la prueba:			
Paso No.	No. De función a probar	Acción o instrucción	Resultado Esperado

**Guía para llenado**

**Propósito del formulario**

Describir en forma detallada cada caso de prueba, su objetivo y los pasos a seguir por parte del “probador” para ejecutarlo.

**Generado por:**

Desarrollador de casos de prueba.

**Para uso de:**

Probador.

**Partes de que se compone:**

 Encabezado

Esta página contiene la información general que se describe el caso así como sus Pre requisitos, nombre de documentos ampliatorios, entradas necesarias y descripción de la prueba o caso por lo que deberá ser leída y entendida por el probador antes de ejecutarla prueba.

 Detalle:

Estas páginas contienen los pasos específicos que el probador deberá seguir para ejecutar el caso. Incluye para cada paso, el número de identificación de la función o funciones que están siendo probadas, las acciones a realizar por parte del

probador, y los resultados que se esperan como respuesta del sistema a cada acción.

#### 4.2.5.4.3 Realizar procedimientos de ejecución y reportes.

El probador analiza el caso de pruebas, siguiendo las instrucciones paso a paso especificadas en el descriptivo del caso de prueba. Se puede utilizar el formulario **“Casos de Prueba”**.

Si la ejecución de un caso de prueba produce resultados incorrectos o inesperados, el probador completa un **“Informe de Problema”**.

Tabla IV.30 - Informe de problema.

INFORME DE PROBLEMA		
Caso número:	Proyecto:	
Fecha:	Sistema:	
Probador:		
Descripción del problema:		
Otros documentos:		

#### Guía para llenado.

##### Propósito del formulario

Describir en forma detallada el error detectado en determinada prueba.

##### Para uso de:

Probador.

##### Partes de que se compone:

###### Encabezado:

Contiene la información del sistema a probar, el nombre del probador y del número de caso a cual pertenece la prueba.

###### Detalle:

Descripción del problema: Explicación detallada del error encontrado.

Otros Documentos: Son documentos respaldatorios que aclaran el problema detectado y que servirán para corregir los errores.

**4.2.5.4.4 Formalizar procedimientos de corrección de errores**

Con la documentación recopilada de las pruebas se procede a la solución de los errores y observaciones detectadas.

Una vez que el o los problemas han sido solucionados y se han entregado las correcciones, se volverán a ejecutar los casos que prueban los módulos corregidos. Este ciclo se repetirá hasta que los casos produzcan los resultados esperados. En este proceso se irá actualizando el status de cada problema en el documento **“Registro de problemas identificados”**.

Este proceso asegura la entrega final de un producto de alta calidad y finalizará con la consecuente aceptación del mismo. Durante la ejecución de las pruebas, se tendrán reuniones de revisión en forma regular y continua, con el fin de revisar el estado de la ejecución del plan, así como el estado de los problemas detectados.

Se considera que las pruebas han terminado, cuando todos los casos de prueba hayan sido ejecutados exitosamente y no queden problemas a resolver.

**Tabla IV.31 - Informe de problemas identificados.**

INFORME DE PROBLEMAS IDENTIFICADOS							
Proyecto:				Registro de problemas:			
Sistema							
Elaborado por:				Pagina.....de.....			
				Fecha			
Prueba No.	Problema	Código de severidad	Probador	Responsable	Estado	Fecha de corrección requerida	Comentarios
Número de problemas identificados							
.....							
Severidad	1.....	2.....	3.....	4.....	5.....		

### **Guía para llenado.**

#### **Propósito del formulario:**

Registrar todos los problemas registrados durante la ejecución de los casos de prueba, llevando registro que facilite su control y seguimiento.

#### **Generado por:**

Responsable de pruebas, tomando como entrada los informes de prueba generados por los probadores.

#### **Para uso de:**

Responsable de pruebas.

#### **Frecuencia de generación:**

Se actualiza diariamente. Cada día se inicia una página nueva con un número consecutivo.

#### **Descripción de los campos:**

- ✚ Elaborado por: Nombre del autor, normalmente el responsable.
- ✚ Fecha: Fecha de identificación de cuando fueron reportados los problemas de esa página.
- ✚ Página: Número de página del registro o LOG.
- ✚ Caso Número: Número del caso en donde se identificó el problema.
- ✚ Número de problema Número Identificador del problema asignado por el coordinador de pruebas.
- ✚ Código de severidad: Código de severidad del problema.

Las siguientes definiciones describen los códigos de severidad que serán usados para clasificar los problemas:

- **Severidad 1** -Este 'detiene todo'. Este tipo de error detiene la prueba que se esta corriendo y las posteriores.
- **Severidad 2** -Este código de severidad esta normalmente asociado con un error en una rutina lógica o en una función específica. Se interrumpe la prueba para esa función o rutina lógica o en una función o rutina particular, pero otras pruebas pueden continuar.
- **Severidad 3** -Este código de severidad indica un error menor en donde una rutina lógica o función no se interrumpe, pero los resultados obtenidos pueden no ser los esperados.
- **Severidad 4**-Este código de severidad indica un error menor o problema en la documentación. Las pruebas no se interrumen. También puede agregarse el estado de **Observación 5**. Este estado indica que es una propuesta de mejora,

no un error. Si la mejora es leve se analizará y se realizará. Si en el análisis de la observación se determina que es de cierta envergadura se deberá proponerla como un cambio siguiendo el procedimiento establecido para estos casos.

- ✚ Probador: Nombre del probador que identificó el problema.
- ✚ Responsable: Nombre de la persona a la que se asignó el problema para su solución. Generalmente será el tesista.  
Estado: estado del problema Puede ser “abierto” cuando aun está en proceso de solución;”CERRADO”: cuando ya se solucionó y la re-prueba fue exitosa.
- ✚ Fecha de solución: Si se trata de un problema ABIERTO, se pone la fecha PLANEADA de solución. Si se trata de un problema CERRADO, se actualiza este campo con la fecha real, de solución.
- ✚ Comentario: Este campo se puede usar para hacer anotaciones adicionales.
- ✚ Total de problemas identificados: Es el número total de problemas que se han identificado hasta esa página del LOG.
- ✚ Severidad “1”, “2”, “3”, “4”: En cada uno de estos campos se pone el total de problemas identificados con esa severidad hasta esa página del LOG.

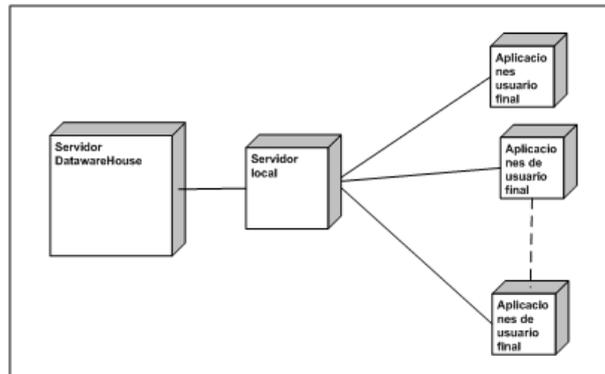
#### **4.2.5.4.5 Implantar el sistema.**

Es importante dar soporte técnico a los usuarios, configurar el hardware y software, establecer conexión a las bases de datos, acceso a Internet o intranet.

Una vez terminadas las actividades de pruebas se determinará que el sistema esta listo para implantar, para ello se hará la disposición de las particiones físicas del sistema y la asignación de los componentes software a estas particiones. Es decir, las relaciones físicas entre los componentes software y hardware en el sistema. Para una mejor comprensión se utilizará diagrama de despliegue para la distribución de las particiones físicas.

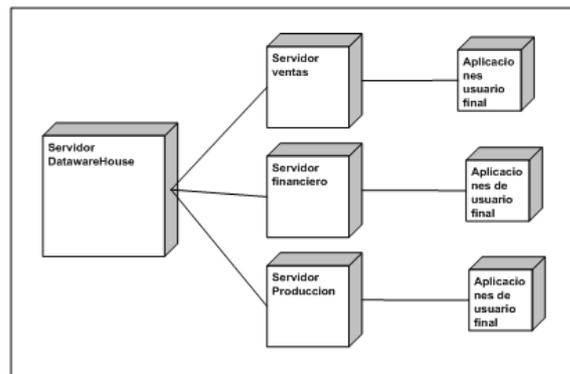
En el diseño físico la plataforma puede centralizarse en una sola ubicación o distribuirse regional, nacional o internacionalmente a continuación se describirá las siguientes alternativas de arquitectura:

1. Para almacenar los datos, que podría obtenerse desde fuentes múltiples internas y externas, es consolidar la base de datos en un data warehouse integrado. El enfoque consolidado proporciona eficiencia tanto en la potencia de procesamiento como en los costos de soporte. Ver figura IV.49



**Figura IV.49 - Arquitectura centralizada.**

2. Los datos son consolidados lógicamente pero se almacenan por separado sin las bases de datos físicas relacionadas, en los mismos sitios físicos o en diferentes. Ver la figura IV.50.



**Figura IV.50 - Arquitectura distribuida.**

#### **4.2.6 FASE VI: ADMINISTRACIÓN Y MANTENIMIENTO DEL SISTEMA.**

Una vez implantado el sistema, no significa que se termina al contrario como se dijo en los capítulos anteriores el desarrollo de un proyecto de business intelligence es un proceso de naturaleza espiral, es por ello que se necesita continuar con los relevamientos de forma constante para poder seguir con la evolución.

#### 4.2.6.1 PROPÓSITO:

El propósito de esta fase es administrar y dar mantenimiento al datawarehouse, datamart y su explotación.

#### 4.2.6.2 DESCRIPCIÓN:

Una descripción gráfica de la fase como muestra la siguiente figura.

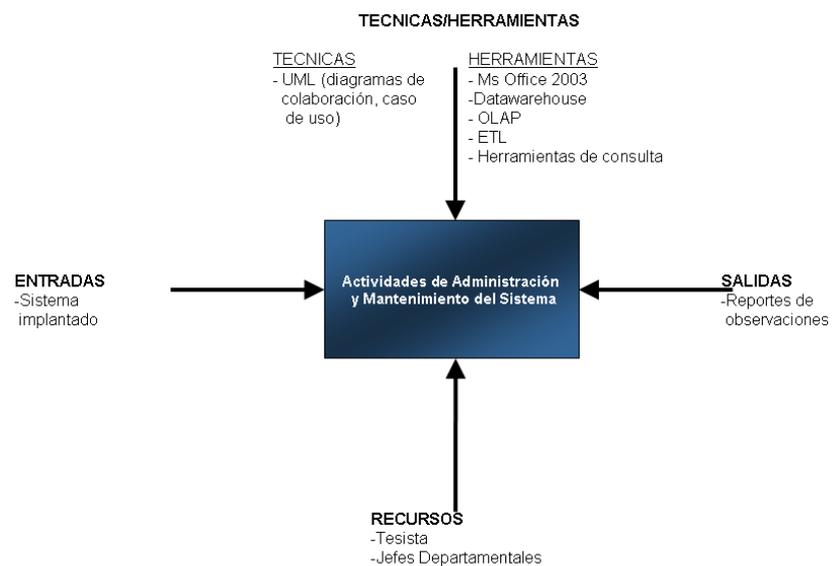


Figura IV.51 - Descripción de la fase.

#### 4.2.6.3 RESULTADOS:

🚩 Reporte de observaciones.

#### 4.2.6.4 ACTIVIDADES:

En estas actividades se mostrarán la interacción entre actores y componentes.



Figura IV.52 - Actividades de Administración y mantenimiento del Sistema.

#### 4.2.6.4.1 Capacitar y dar soporte técnico.

Los encargados del área de sistema capacitarán y darán soporte técnico a los usuarios del proyecto constantemente.

#### 4.2.6.4.2 Realizar la gestión de los módulos.

Se realizará la gestión de los siguientes módulos:

- ✚ Extracción, transformación y carga
- ✚ Datawarehouse.
- ✚ Acceso a datos o consultas.

En un caso de uso se describirá en qué consiste la gestión de cada módulo cómo lo muestra la siguiente tabla.

**Tabla IV.32 - Funciones del Proceso de carga**

<b>Caso de uso</b>	
Requerimiento que lo implementa	
Descripción	
Figuras asociadas	
Observaciones	

#### **Guia para llenado**

*Caso de uso:* nombre del modulo.

*Requerimiento que lo implementa:* Módulo necesario para el funcionamiento del sistema.

*Descripción:* Describir las funciones del modulo, herramientas que dan soporte usuarios responsables de utilizar este módulo.

*Figuras asociadas:* Enumerar las figuras relacionas con este modulo.

*Observaciones:* describir todo tipo de observaciones que tiene el administrador.

## **CAPÍTULO V**

### **IMPLEMENTAR BUSINESS INTELLIGENCE EN COHERVI S.A**

#### **5.1 Introducción**

La mayoría de las organizaciones cuentan con datos de los sistemas de ingreso de transacciones, vitales para registrar las operaciones que sostienen a una empresa. A pesar de la riqueza de estos datos, no se puede recurrir a ellos con facilidad cuando necesitamos encontrar respuestas sobre el funcionamiento de la empresa.

Si bien con los sistemas tradicionales se pueden preparar reportes ad-hoc para encontrar las respuestas a algunas de estas preguntas, se necesita mucho tiempo y recursos del departamento de sistemas para poder responderlas.

Además interfiere en el procesamiento de los sistemas transaccionales, aumentando los tiempos de respuestas de los mismos.

Con sistemas del tipo “Inteligencia de Negocios” (*Business Intelligence*), se puede formular y responder las preguntas claves sobre el funcionamiento de la universidad accediendo directamente a los indicadores de éxito, señalar cuáles son los factores que realmente inciden en el buen o mal funcionamiento, detectar situaciones fuera de lo normal, encontrar los factores que maximicen el beneficio y predecir el comportamiento futuro con un alto porcentaje de certeza.

Para esto es necesario implementar un Sistema de Apoyo Gerencial. El sistema debe estar orientado a brindar información interrelacionada para quienes tienen responsabilidades en el ámbito estratégico y táctico de la organización.

Los usuarios de este sistema no son usuarios comunes, sino usuarios que toman decisiones y planifican día a día, a mediano plazo o a largo plazo, en donde la calidad de la información juega un papel categórico.

Usuarios que necesitan disponer de información tanto consolidada como detallada de cómo marchan las actividades ya cumplidas, predecir tendencias y comportamientos para tomar decisiones proactivas.

Los usuarios de este sistema pertenecen a todos los departamentos y áreas de la empresa, ya que en cada área es necesario algún tipo de información como la mencionada. Inicialmente, los usuarios potenciales de este sistema pueden pertenecer a:

- Gerencia
- Jefes departamentales.

## **5.2 ETAPAS DE LA METODOLOGÍA PARA DESARROLLAR PROYECTOS DE BUSINESS INTELLIGENCE.**

En este capítulo se desarrolla cada una de las fases de la MEDEPROBI, a continuación se detalla cada una de las fases.

### **5.2.1 FASE I: PLANIFICACIÓN DEL PROYECTO**

#### **5.2.1.1 PROPÓSITO:**

A través de un estudio en la empresa COHERVI S.A lograr conocer la situación actual, es decir conocer el movimiento y la lógica de negocio de la misma.

### 5.2 1.2 DESCRIPCIÓN:

Una descripción grafica de la fase como muestra la siguiente figura.

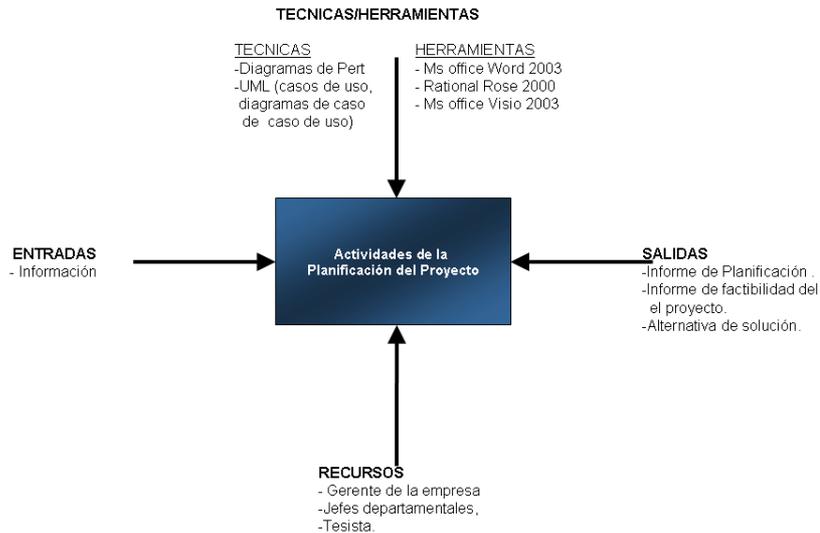


Figura V.53 - Descripción de la fase.

### 5.2.1.3 RESULTADOS:

- 🚩 Valoración general del negocio y alternativa de solución.

### 5.2.1.4 ACTIVIDADES.

Las actividades desarrolladas en esta fase son las siguientes:

#### 5.2.1.4.1 Definir el ámbito del negocio.

##### 5.2.1.4.1.1 Misión

El compromiso central de COHERVI S.A es la satisfacción de las necesidades y expectativas del cliente, suministrándole productos a bajos precios, garantizando sus productos de calidad para mantenernos como líderes en el mercado.

##### 5.2.1.4.1.2 Visión

Nuestra visión es la de ser una empresa líder en el mercado con los más bajos precios y de la más alta calidad para cubrir las necesidades locales, cantonales y provinciales.

### 5.2.1.4.1.3 Antecedente Tecnológico

Para la implantación del sistema se requiere de una buena infraestructura que soporte que garantice el acceso a la aplicación desde donde se requiera. COHERVI S.A. cuenta actualmente con una infraestructura aceptable y que servirá de base para la implantación de Sistema de Análisis Gerencial (SAG)

Entre los principales recursos tecnológicos que posee la COHERVI S.A. se encuentran:

#### Red empresarial

- ✚ La red utiliza principalmente un cableado de cobre para cubrir la red empresarial.
- ✚ Se cuenta con equipos de conmutación que aceptan medios de cobre.
- ✚ LAN, para acceso a la Intranet.
- ✚ Acceso a Internet en ciertos nodos conectados a la red.

**Tabla V.33 - Infraestructura de Servidores**

HARDWARE	PLATAFORMA
Un servidor de Desarrollo - DD: 40GB - RAM: 512Mb - Procesador: Intel Pentium III - Modelo: Compaq	Windows Server 2003
Un servidor de Producción - DD: 100Gb - RAM: 1Gb - Procesador: Intel Xeon - Modelo: HP	Windows Server 2003
Un servidor de Internet - DD: 80Gb - RAM: 512Mb - Procesador: Intel PIV - Modelo: Genérico	Linux

**Tabla V.33 - Infraestructura de Servidores (continuación)**

Un servidor de Intranet - DD:80Gb - RAM: 1Gb - Procesador: Intel PIV - Modelo: Genérico	Linux
Un servidor de Video - DD: 500Gb - RAM: 4 GB - Procesador: Intel PIV - Modelo: Genérico	Windows XP Service Pack 2

### **Clientes**

- Existen equipos con características heterogéneas en cuanto a hardware, a través de los cuales se brinda el acceso al sistema y a Internet o que se utilizan como estaciones de trabajo aisladas. Todos tienen sistemas operativos compatibles con Win32.

### **Sistema Integrado**

COHERVI S.A cuenta con un sistema integrado para el manejo de facturación, clientes, proveedores e inventario, pero no cuenta con una herramienta o sistema que le ayude a tener conocimiento de la situación actual de la empresa.

El recurso tecnológico con el que cuenta la empresa brinda las facilidades necesarias para la construcción e implantación del Sistema de Análisis Gerencial (SAG).

#### **5.2.1.4.2 Analizar los escenarios problemas.**

Con los sistemas tradicionales se preparan reportes para encontrar las respuestas a algunas preguntas, pero se necesita dedicar aproximadamente un 80% del tiempo asignado al análisis de localización y presentación de los datos, como también asignación de recursos humanos y de procesamiento del departamento de sistemas

para poder responderlas, sin tener en cuenta la degradación de los sistemas transaccionales.

Esta problemática se debe a que dichos sistemas transaccionales no fueron contruidos con el fin de brindar síntesis, análisis, consolidación, búsquedas y proyecciones.

En todos los casos se observa la necesidad de considerar como punto de partida la información existente en las bases de datos de la empresa. La empresa cuenta con un sistema, esta información es la plataforma para diseñar y construir un Sistema de Análisis Gerencial. Los sistemas que sustentan las principales actividades de la empresa son:

✚ Sistema Administrativo.

- Contabilidad

✚ Sistema de producción

- Inventarios
- Facturación
- Proveedores
- Cuentas por Cobrar

Luego de un análisis exhaustivo en COHERVI S.A y empleando las técnicas de casos de uso describiremos de una mejor manera las actividades que se desarrollan en la actualidad.

**Tabla V.34 - Caso de uso Sistema Actual**

Caso de uso:	Sistema Actual
Actores:	Gerente, jefe de sistemas.
Descripción:	En el sistema actual el proceso de consultas de reportes es el siguiente: El gerente solicita reportes de situación de ventas, el departamento de sistemas recepta el pedido, lo procesa y entrega al gerente, pero los reportes no son actualizados.

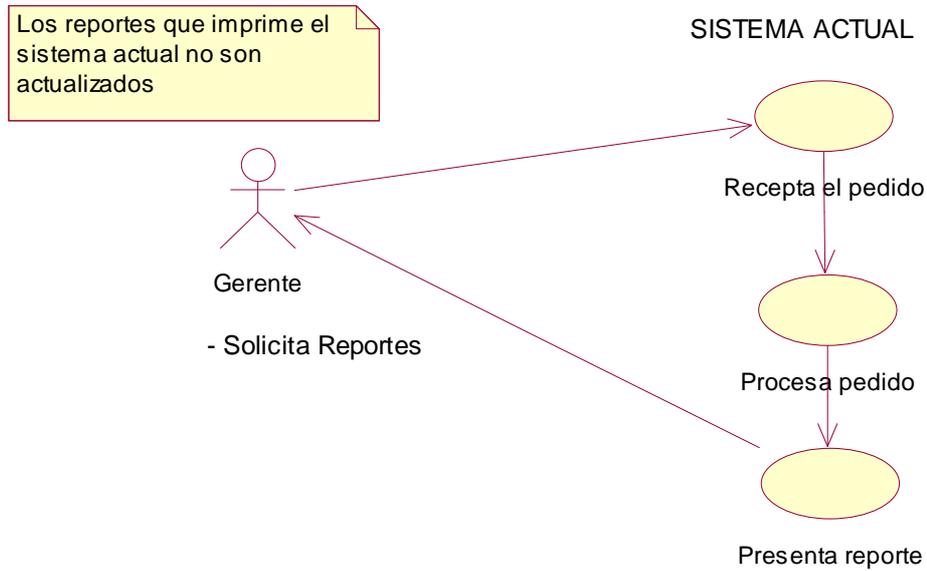


Figura V.54 - Caso de uso sistema actual

#### 5.2.1.4.3 Determinar alternativas de Solución.

##### 5.2.1.4.3.1 Descripción general del Sistema de Análisis Gerencial (SAG).

A través de un caso de uso y diagramas se demostrará la alternativa de solución.

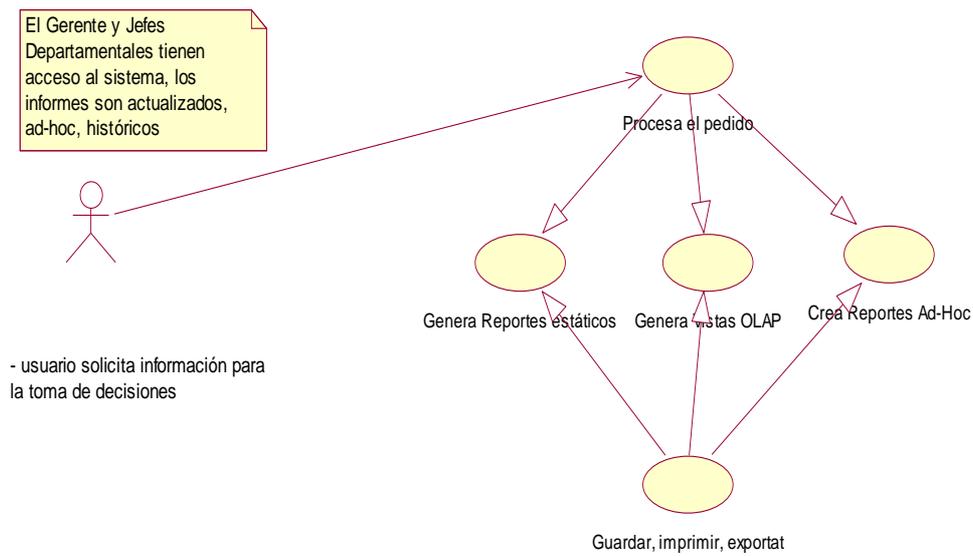
Tabla V.35 - Caso de uso

Caso de uso:	Sistema Propuesto
Actores:	Gerente, jefes departamentales.
Descripción:	Se desarrollará un Sistema de Análisis Gerencial (SAG) que genere una variedad de productos de información, con el fin de satisfacer las necesidades cambiantes de las personas que toman decisiones a través de toda la empresa. Las personas encargadas de tomar decisiones a nivel estratégico requieren informes, pronósticos e información más resumida, ad hoc, para respaldar sus responsabilidades de elaboración de políticas y de planeación no estructuradas

**Tabla V.35 - Caso de uso(continuación)**

Descripción:	<p>El SAG permitirá:</p> <ul style="list-style-type: none"><li>▪ Suministrar información y técnicas de apoyo a las decisiones para analizar problemas u oportunidades específicas.</li><li>▪ Consultas y respuestas interactivas</li><li>▪ Formato ad hoc, flexible y adaptable.</li><li>▪ Información generada mediante modelación analítica de los datos empresariales.</li><li>▪ Informes OLAP</li><li>▪ Reportes estáticos</li></ul>
--------------	--

**Diagrama de Caso de uso**



**Figura V.55 - Caso de uso sistema propuesto**

#### 5.2.1.4.4 Analizar unidades organizativas afectadas.

Las unidades que participan en el proyecto son:

**Tabla V.36 -Unidades organizativas.**

UNIDADES ORGANIZATIVAS	RESPONSABILIDADES
- Gerencia(Gerente)	<ul style="list-style-type: none"> <li>- Determinar el área del negocio a ser soportadas en el proyecto de desarrollo.</li> <li>- Identificar a los usuarios participantes y finales.</li> <li>- Aprobar el sistema.</li> </ul>
- Departamentos (jefes departamentales)	<ul style="list-style-type: none"> <li>- Establecer del alcance del sistema</li> <li>- Planificación del sistema</li> <li>- Establecer requisitos</li> <li>- Ejecutar las pruebas del sistema</li> <li>- Aceptar el sistema.</li> </ul>
ESPOCH (Directora y miembro de la tesis)	<ul style="list-style-type: none"> <li>- Establecer del alcance del sistema</li> <li>- Evaluar el desarrollo del sistema</li> <li>- Aprobar el sistema dentro de los alcances de la tesis</li> <li>- Revisar y corregir el trabajo de tesis.</li> </ul>
Departamento de Sistemas (Jefe y colaborador del área)	<ul style="list-style-type: none"> <li>- Determinar que sistemas transaccionales están asociados a las áreas de negocio seleccionadas</li> <li>- Analizar diferentes herramientas para OLAP que se adecuen a la empresa. Herramientas de Acceso al componente de almacenamiento físico DataWarehouse.</li> <li>- Definir el origen de los datos operacionales. El modelo de datos fuente.</li> </ul>
ESPOCH(Tesista)	<ul style="list-style-type: none"> <li>- Establecer del alcance del sistema</li> <li>- Planificación del sistema</li> <li>- Estudiar la situación actual</li> <li>- Obtener y definir los requisitos del sistema</li> <li>- Estudiar y analizar diferentes herramientas para ETL, Datawarehouse y OLAP,</li> <li>- Seleccionar alternativas para la solución</li> <li>- Definir el modelo de datos del DW. El modelo de datos conceptual DW</li> <li>- Definir las interfaces de usuario para el accesos a los datos</li> </ul>

**Tabla V.36 -Unidades organizativas(continuación)**

ESPOCH(Tesista)	<ul style="list-style-type: none"><li>- Definir las diferentes customizaciones de las herramientas seleccionadas</li><li>- Especificar la extracción, transformación y carga de datos entre los sistemas transaccionales y el DW</li><li>- Definir la arquitectura del sistema</li><li>- Definir las interfaces de Usuario para el acceso y explotación de los datos.</li><li>- Generar el código de los componentes</li><li>- <b>Ejecutar pruebas del sistema.</b></li></ul>
-----------------	---

**5.2.1.4.5 Definir los costos.**

El costo del proyecto no se definirá por cuanto que es un trabajo de tesis, es decir COHERVI S.A no pagará invertirá en el sistema a desarrollar.

**5.2.1.4.6 Realizar el cronograma de trabajo.**

En esta actividad desarrollar el cronograma de trabajo, haciendo uso de los diagramas de Gantt.

**Tabla V.37 – Cronograma de Trabajo**

Id.	Nombre de tarea	Comienzo	Fin	Duración	nov 2008		dic 2008				ene 2009				feb 2009				mar 2009				abr 2009				may 2009				jun 2009				jul 2009				ago 2009				sep 2009					
					2/11	9/11	16/11	23/11	30/11	7/12	14/12	21/12	28/12	4/1	11/1	18/1	25/1	1/2	8/2	15/2	22/2	1/3	8/3	15/3	22/3	29/3	5/4	12/4	19/4	26/4	3/5	10/5	17/5	24/5	31/5	7/6	14/6	21/6	28/6	5/7	12/7	19/7	26/7	2/8	9/8	16/8	23/8	30/8
1	Investigación	03/11/2008	31/03/2009	107d																																												
2	Planificación del Proyecto	01/04/2009	15/04/2009	11d																																												
3	Definición de Requerimientos	16/04/2009	30/04/2009	11d																																												
4	Diseño	01/05/2009	29/05/2009	21d																																												
5	Construcción y Desarrollo	01/06/2009	31/08/2009	66d																																												
6	Pruebas e Implantación	01/09/2009	25/09/2009	19d																																												
7	Administración y Mantenimiento del Sistema	28/09/2009	30/09/2009	3d																																												

### 5.2.1.4.7 Analizar los riesgos.

En el siguiente cuadro se realiza un resumen de los posibles riesgos que se pueden dar en la ejecución del proyecto.

**Tabla V.38 - Descripción de riesgos:**

<b>RIESGO</b>	<b>DESCRIPCION</b>
R1	Incumplimiento de los objetivos
R2	Incumplimiento del cronograma de trabajo
R3	Falta de comunicación entre el tesista y los altos dirigentes de la empresa.
R4	Problemas de acceso a las bases de datos.
R5	Insuficiente información para el desarrollo del proyecto.
R6	Insatisfacción de los usuarios.

Una vez identificados los riesgos es necesario categorizarlos correctamente, el mismo que nos ayudará a identificar los riesgos de mayor prioridad.

**Tabla V.39 - Categorización de riesgos:**

<b>RIESGOS</b>	<b>PROBABILIDAD</b>		
	<b>ALTA</b>	<b>MEDIA</b>	<b>BAJA</b>
R1			X
R2		X	
R3		X	
R4	X		
R5		X	
R6			X

Luego de haber determinado los posibles riesgos y categorizarlos es necesario establecer un plan de contingencia.

**Tabla V.40 - Plan de contingencia.**

<b>RIESGO</b>	<b>PLAN DE GESTION</b>
R <sub>1</sub>	Problema: El proyecto no cumple con los objetivos planteados.
	Actividades a desarrollar: <ul style="list-style-type: none"> <li>- Seguir y evaluar continuamente el avance del proyecto.</li> </ul> Interactuar continuamente con los jefes departamentales.
	Responsable: Elsi Ilbay
R <sub>2</sub>	Problema: Retrazo en la entrega del proyecto final.
	Actividades a desarrollar: <ul style="list-style-type: none"> <li>- Planificar entregas parciales del trabajo desarrollado.</li> <li>- Controlar continuamente el avance del proyecto en base a la planificación planteada inicialmente.</li> </ul>
	Responsable: Elsi Ilbay
R <sub>3</sub>	Problema: Retraso en las actividades del proyecto
	Actividades a desarrollar: <ul style="list-style-type: none"> <li>- Buscar al jefe inmediato superior para nos ayude a coordinar las actividades que se llevaran a cabo en el desarrollo del proyecto.</li> </ul>
	Responsable: Elsi Ilbay
R <sub>4</sub>	Problema: Debido a que no podemos acceder a la base de datos de la empresa no se podrá interactuar con la misma.
	Actividades a desarrollar: <ul style="list-style-type: none"> <li>- Gestionar los permisos necesarios para acceder a la base de datos o a una réplica.</li> </ul>
	Responsable: Elsi Ilbay
R <sub>5</sub>	Problema: Mal enfoque de la solución del problema.
	Actividades a desarrollar: <ul style="list-style-type: none"> <li>- Planificar reuniones de trabajo para abstraer los verdaderos requerimientos del cliente.</li> </ul>
	Responsable: Elsi Ilbay
R <sub>6</sub>	Problema: Desperdicios de los recursos empleados en una aplicación no utilizada.
	Actividades a desarrollar: <ul style="list-style-type: none"> <li>- Identificar y evaluar permanentemente el cumplimiento de los requerimientos de los usuarios.</li> </ul>
	Responsable: Elsi Ilbay

**5.2.1.4.8 Realizar un estudio de factibilidad.**

Para determinar si el proyecto es factible o no se analiza lo siguiente: **Factibilidad Técnica:**

COHERVI S.A en la actualidad cuenta con la tecnología necesaria, la misma que permitirá viabilizar y hacer posible el desarrollo e implantación del sistema propuesto que cuenta con:

**Tabla V.41 - Factibilidad técnica.**

<b>HARDWARE</b>	
<ul style="list-style-type: none"> <li>- 5 Servidores (desarrollo, Internet, aplicaciones, video, producción).</li> <li>- 3 PC en el Departamento de Sistemas.</li> <li>- Acceso a Internet.</li> <li>- 1 lapto del tesista.</li> </ul>	
<b>SOFTWARE</b>	
<ul style="list-style-type: none"> <li>▪ Contamos con las Bases de Datos del Sistema integrado</li> <li>▪ Herramientas de Documentación (Microsoft Office 2003).</li> <li>▪ Herramientas de Modelado de Datos (Power Designer v8.0).</li> <li>▪ Herramientas CASE: Rational Rose 2000.</li> <li>▪ Herramientas libres para el desarrollo de la aplicación.</li> </ul>	
<b>HUMANOS</b>	
<ul style="list-style-type: none"> <li>▪ Tesista.</li> </ul>	

**Factibilidad Operativa.**

Demostrar que los tiempos de respuesta del sistema propuesto son mucho mejores que el sistema actual.

**Tabla V.42 - Factibilidad Operativa**

<b>SISTEMA ACTUAL</b>		<b>SISTEMA PROPUESTO</b>	
<b>ACTIVIDADES</b>	<b>TIEMPOS</b>	<b>ACTIVIDADES</b>	<b>TIEMPOS</b>
<ul style="list-style-type: none"> <li>▪ Reportes</li> </ul>	15 -30 días	<ul style="list-style-type: none"> <li>▪ Reportes OLAP.</li> </ul>	2 días
<b>TOTAL</b>	<b>30 días</b>		<b>10 min</b>

**Factibilidad Económica.**

La factibilidad económica no contempla en el desarrollo de la tesis, ya que COHERVI S.A no tiene una inversión inicial.

**Factibilidad Legal.**

El SAG cumple con todo lo estipulado en la ley de la empresa.

**5.2.2 FASE II: DEFINICIÓN DE REQUERIMIENTOS.**

La fase de definición de requerimientos es el cerebro del proyecto por ende se ha asignado el mayor esfuerzo en comprender la lógica de negocios.

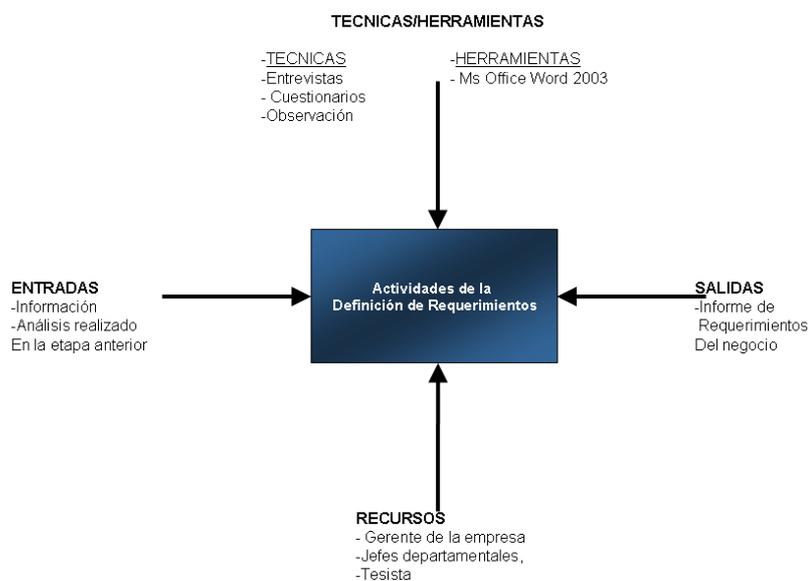
Se ha realizado un informe detallado de los requerimientos de los usuarios ejecutivos.

**5.2.2.1 PROPÓSITO:**

Definir correctamente los requerimientos de los usuarios de área de ventas, inventario y compras.

**5.2.2.2 DESCRIPCIÓN:**

Una descripción grafica de la fase como muestra la siguiente figura.



**Figura V.56 - Descripción de la fase.**

### **5.2.2.3 RESULTADOS:**

- ✚ Lista de requerimientos de los departamentos de compras, ventas e inventario.

### **5.2.2.4 ACTIVIDADES:**

Las actividades a realizar en esta fase son las siguientes:

#### **5.2.2.4.1 Mencionar técnicas utilizadas en la recopilación de requerimientos.**

Las técnicas utilizadas en la recopilación de información en el proyecto de Sistema de Análisis Gerencial son: entrevistas, encuestas y la observación.

#### **5.2.2.4.2 Describir los requerimientos.**

Luego de varias sesiones de trabajo con los usuarios participantes, se ha analizado la información recolecta, donde se define los siguientes requerimientos.

#### **Requerimientos funcionales:**

Planteados por los usuarios de la empresa COHERVI S.A, como necesidad de información para su departamento.

### **DEPARTAMENTO DE VENTAS**

1. Realizar un análisis de las ventas, de los últimos años analizados por los tipos de venta existentes.
2. Observar las variaciones en las ventas por las distintas zonas de la localidad a lo largo del tiempo.
3. Analizar montos de ventas por productos en los últimos años.
4. Comparar las ventas por zonas, realizadas por cada vendedor durante los últimos tres años.
5. Analizar el desempeño de los distintos vendedores y la evolución de tal desempeño en los últimos años.
6. Estudiar el comportamiento de las ventas realizadas a los clientes y un detalle de los productos adquiridos por ellos en el último semestre.

### **DEPARTAMENTO DE COMPRAS.**

1. Analizar la compra de mercadería a los distintos proveedores estudiándola por producto a un nivel de detalle por mes y año.
2. Comparar el monto de compras realizadas durante los últimos años.
3. Considerar que productos han sido adquiridos en los últimos meses por la empresa.

### **DEPARTAMENTO DE INVENTARIO.**

1. Realizar un análisis del stock de bodega por producto en el último trimestre.
2. Analizar cantidad de productos caducados en el último trimestre del año en curso.
3. Estudiar la rotación de los productos en el último trimestre de acuerdo a los productos vendidos.

### **Requerimientos tecnológicos:**

Dentro de los requerimientos tecnológicos se debe disponer del siguiente hardware y software, tanto en los equipos cliente como en el servidor.

COHERVI S.A cuenta con la tecnología necesaria para llevar a cabo este proyecto sin embargo haremos mención de los requerimiento hardware necesarios.

### **Hardware de Servidor**

Para nuestro sistema utilizaremos uso de un equipo, donde se encontrará el repositorio "Datawarehouse" y el servidor OLAP.

**Tabla V.43 – Hardware del Servidor**

<b>Computadora/Procesador:</b>	Intel PIV 3.2 Ghz
<b>Memoria:</b>	4GB de RAM
<b>Disco Duro:</b>	500Gb
<b>Hardware de red:</b>	Para accesos a las bases de datos fuentes, para la distribución y servicios a las estaciones clientes.

## Hardware del Cliente

**Tabla V.44 – Hardware del Cliente**

<b>Computadora/Procesador:</b>	Pentium (166 Mhz o superior)
<b>Memoria:</b>	128 Mb de RAM
<b>Disco Duro:</b>	40Gb o superior
<b>Hardware de Red:</b>	Para acceso al Datawarehouse.

## Software de Servidor:

**Tabla V.45 – Software del Servidor**

<b>Sistema Operativo</b>	Sistema operativo Microsoft Windows XP Service Pack II
Java Virtual Machine	Sun 1.6

## Software del Cliente:

**Tabla V.46 – Software del Cliente**

<b>Sistema Operativo:</b>	Sistema operativo Microsoft Windows, Linux
<b>Explorador Web</b>	Internet Explorer, Mozilla Firefox, etc

## Recursos humanos:

Para la ejecución del proyecto será necesario contar con dos tipos de recursos humanos, los mismos que son los siguientes

- ✚ **Recursos humanos para construir el sistema:** son los jefes de las distintas áreas de la empresa de los departamentos de: Ventas, Compras e Inventario.
- ✚ **Recursos humanos para el soporte técnico y mantenimiento del sistema:** El personal del Departamento de Sistema será el encargado de administrar el Sistema de Análisis Gerencial.

### 5.2.3 FASE III: DISEÑO

#### 5.2.3.1 PROPÓSITO:

Definir correctamente los componentes hardware software que se utilizarán para la ejecución del sistema.

#### 5.2.3.2 DESCRIPCIÓN:

Una descripción gráfica de la fase como muestra la siguiente figura.

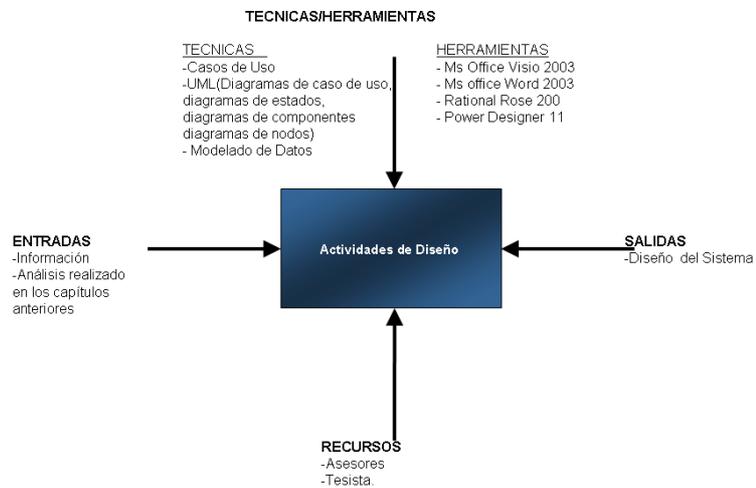


Figura V.57 - Descripción de la fase.

#### 5.2.3.3 RESULTADOS:

- 🚩 Diseño de la arquitectura del proyecto.

#### 5.2.3.4 ACTIVIDADES:

A continuación se dará a conocer las principales actividades a desarrollar en esta fase:

##### 5.2.3.4.1 Diseñar la arquitectura.

La arquitectura que se utilizará para el desarrollo del Sistema de Análisis Gerencia (SAG), está representa por la Figura V.58.

En la primera capa tenemos las bases de datos del sistema Integrado y ciertos archivos de Excel, que se convierten en los sistemas fuentes.

La capa dos que corresponde a la capa del DataWarehouse donde se encuentran los datos limpios y preparados para formar los cubos OLAP, reportes Ad Hoc y otros reportes.

La capa tres pertenece a la interfaz gráfica, a través de la cual los usuarios accederán a los datos, realizarán el análisis de la información.

Las primera flecha representan a los procesos de Extracción, Transformación y Carga (ETL) de datos, que es uno de los proceso claves para el funcionamiento de los sistemas de Inteligencia de Negocios.

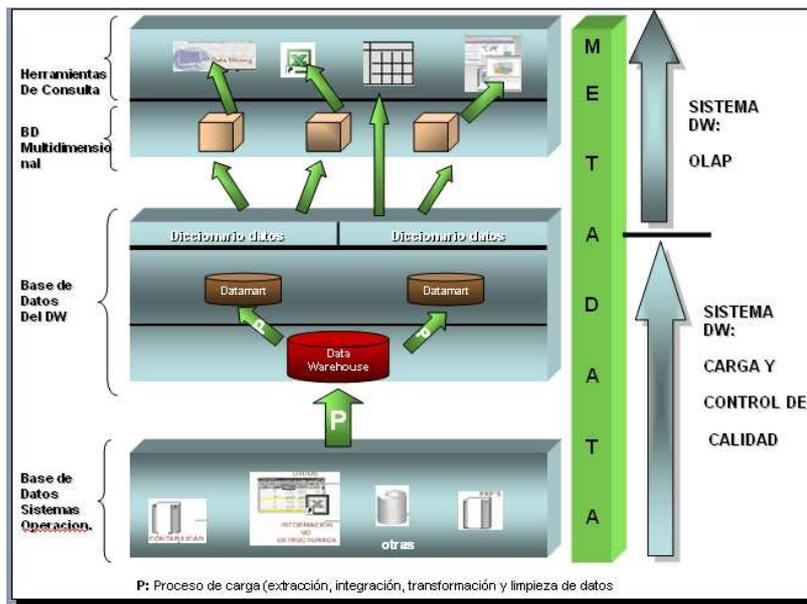


Figura V.58 - Diseño de la Arquitectura.

#### 5.2.3.4.2 Identificar Casos de Uso.

La siguiente tabla nos muestra los distintos casos de uso que han sido identificados para el desarrollo del Sistema de Análisis Gerencial.

**Tabla V.47 – Casos del uso del Sistema de Análisis Gerencial**

No	Nombre de caso de uso	Eventos	Actores
1	Autenticación del usuario	Ingresar a la página Web	Administrador Jefes departamentales
2	Visualización de análisis OLAP	Mostrar análisis OLAP	Jefes departamentales: ventas, compras e inventario.
3	Visualización de Reportes estáticos y Ad Hoc	Mostrar Reportes	Jefes departamentales: ventas, compras e inventario.
4	Ventas	Análisis	Jefe de ventas
5	Compras	Análisis	Jefe de compras
6	Inventario	Análisis	Jefe de inventario
7	Generar OLAP	Crear Cubos	Jefes departamentales
8	Generar reportes Ad Hoc	Crear reportes Ad Hoc	Jefes departamentales
9	Análisis Gerencia	Análisis y reportes	Gerente

**5.2.3.4.3 Desarrollar Casos de uso y Diagramas de Casos de uso del sistema Propuesto.**

**Tabla V.48 - CU\_AUT # 01**

<b>CASO DE USO</b>	Autenticación del usuario
<b>ACTORES</b>	Administrador, Jefes departamentales.
<b>TIPO</b>	Primario esencial
<b>PROPOSITO</b>	Ingresar a la página Web
<b>VISION GENERAL</b>	El usuario accede al sitio Web para observar la información.
<b>REFERENCIAS</b>	Ninguna

Tabla V.48 - CU\_AUT # 01(continuación)

CURSO TIPICO DE EVENTOS	
<p><b>ACTOR</b></p> <ol style="list-style-type: none"> <li>1. usuario solicita iniciar sesión</li> <li>3. Ingresa datos</li> <li>5. Visualiza opciones</li> </ol>	<p><b>SISTEMA</b></p> <ol style="list-style-type: none"> <li>2. Solicita usuario y clave</li> <li>4. Verifica y valida datos                             <ol style="list-style-type: none"> <li>4.1 Satisfactorio (accede a la pagina web)</li> <li>4.2 Insatisfactorio (mensaje usuario no encontrado)</li> </ol> </li> </ol>

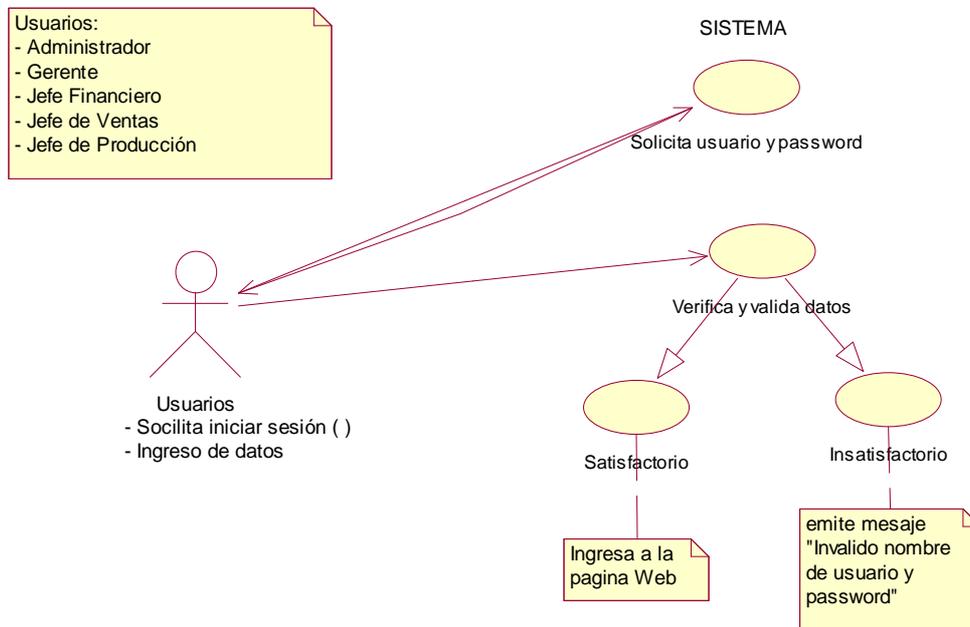


Figura V.59 - Caso de uso # 01

**Tabla V.49 - CU\_VAOLAP # 02**

<b>CASO DE USO</b>	Visualización de análisis OLAP
<b>ACTORES</b>	Jefes departamentales
<b>TIPO</b>	Primario esencial
<b>PROPOSITO</b>	Analizar áreas departamentales
<b>VISION GENERAL</b>	El usuario accede al sitio Web para analizar la situación del departamento que le corresponde.
<b>REFERENCIAS</b>	CU_AUT, requerimientos departamento de ventas, compras e inventario
<b>CURSO TIPICO DE EVENTOS</b>	
<p><b>ACTOR</b></p> <ol style="list-style-type: none"> <li>1. Selecciona carpeta análisis OLAP</li> <li>3. Selecciona una vista</li> <li>5. Selecciona opciones             <ol style="list-style-type: none"> <li>4.1 Datos</li> <li>4.2 Gráficos</li> <li>4.3 Exportar</li> </ol> </li> <li>7. Usuario sale del sistema.</li> </ol>	<p><b>SISTEMA</b></p> <ol style="list-style-type: none"> <li>2. Muestra lista de vistas</li> <li>4. Muestra la vistas de análisis</li> <li>6. Ejecuta proceso (datos, grafico, exportar)</li> </ol>
<b>CURSOS ALTERNATIVOS</b>	
<p><b>Línea 4.1 (Datos):</b> El usuario podrá navegar y visualizar los datos haciendo uso de las operaciones OLAP como son el Drill up y drill down.</p> <p><b>LINEA 4.2 (gráficos):</b> Si seleccionar gráficos tendrá una lista de gráficos que el usuario podrá escoger para ver los datos como gráficos de barra, circulares, en 3D, etc.</p> <p><b>Línea 4.3 (Exportar):</b> el usuario tiene dos opciones de exportar: a Excel o Adobe Reader.</p>	

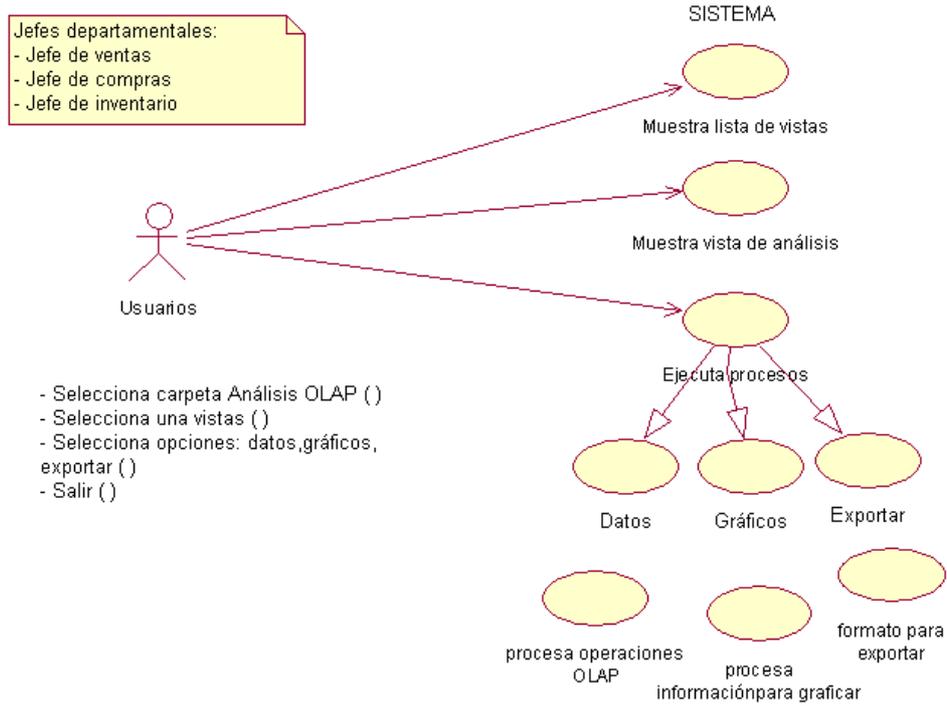
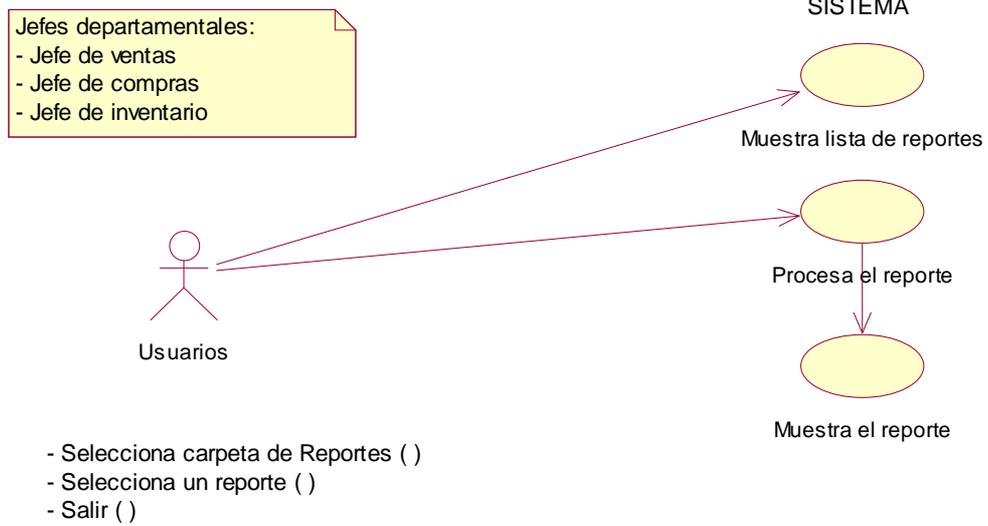


Figura V.60 - Caso de uso # 02

Tabla V.50 - CU\_VREP\_EST\_ADH # 03

<b>CASO DE USO</b>	Visualización de reportes estáticos y Ad Hoc
<b>ACTORES</b>	Jefes departamentales
<b>TIPO</b>	Primario esencial
<b>PROPOSITO</b>	Analizar áreas departamentales
<b>VISION GENERAL</b>	El usuario accede al sitio Web para analizar la situación del departamento que le corresponde.
<b>REFERENCIAS</b>	CU_AUT, requerimientos departamento de ventas, compras e inventario.
<b>CURSO TIPICO DE EVENTOS</b>	
<b>ACTOR</b>	<b>SISTEMA</b>
1. Selecciona carpeta de Reportes	2. Muestra lista de reportes
3. Selecciona reporte	4. Procesa el reporte
6. Usuario sale del sistema.	5. Muestra el reporte



**Figura V.61 - Caso de uso # 03**

**Tabla V.51 - CU\_DVENTAS # 04**

<b>CASO DE USO</b>	Análisis ventas
<b>ACTORES</b>	Jefe de ventas
<b>TIPO</b>	Primario esencial
<b>PROPOSITO</b>	Analizar el departamento de ventas
<b>VISION GENERAL</b>	El usuario accede al sitio web para analizar la información.
<b>REFERENCIAS</b>	CU_AUT, CU_VAOLAP, CU_VREP_EST_ADH, requerimientos departamento de ventas.
<b>CURSO TIPICO DE EVENTOS</b>	
<b>ACTOR</b> 1. Solicita información	<b>SISTEMA</b> 2. Análisis de ventas de los productos. 3. Análisis de ventas de crédito y contado. 4. Análisis de ventas según zonas. 5. Análisis de ventas según zona de cada vendedor. 6. Análisis del desempeño de los vendedores. 7. Análisis de rotación de los productos según los productos vendidos.

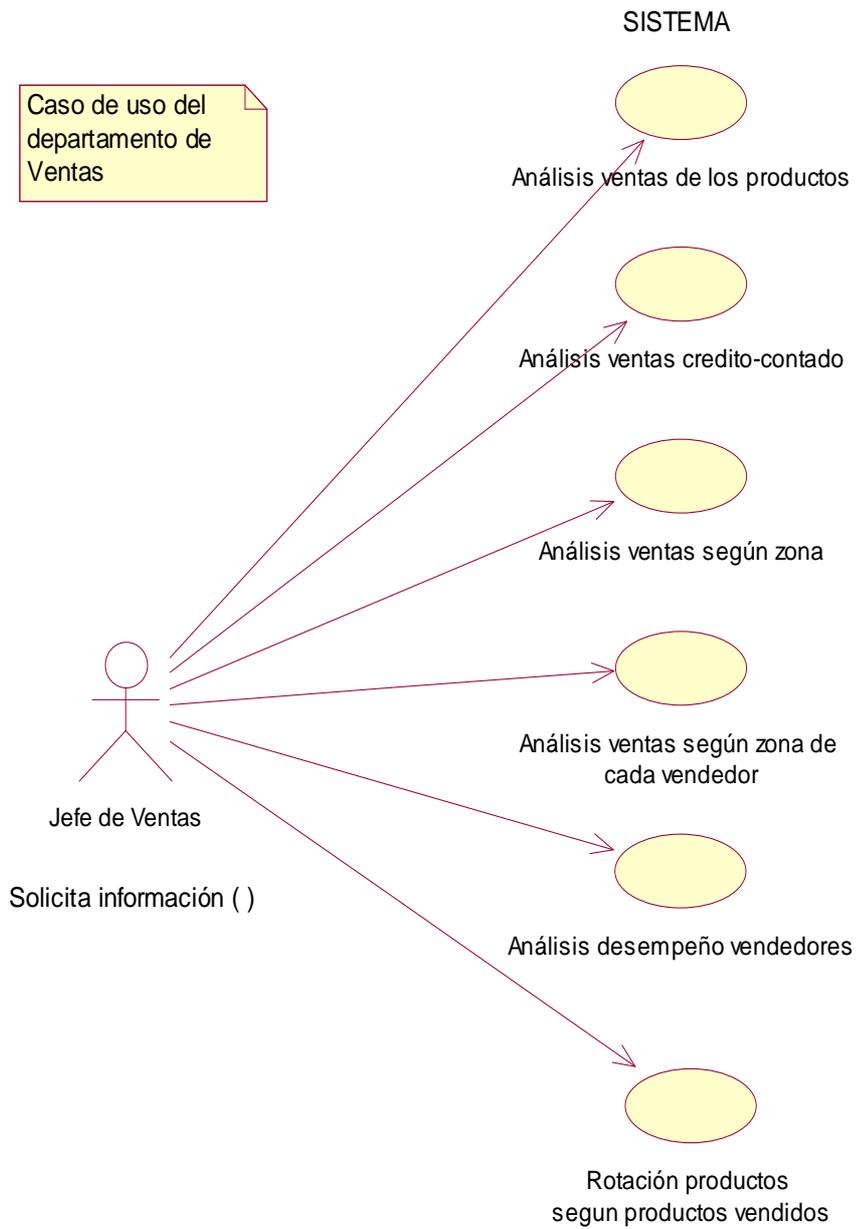
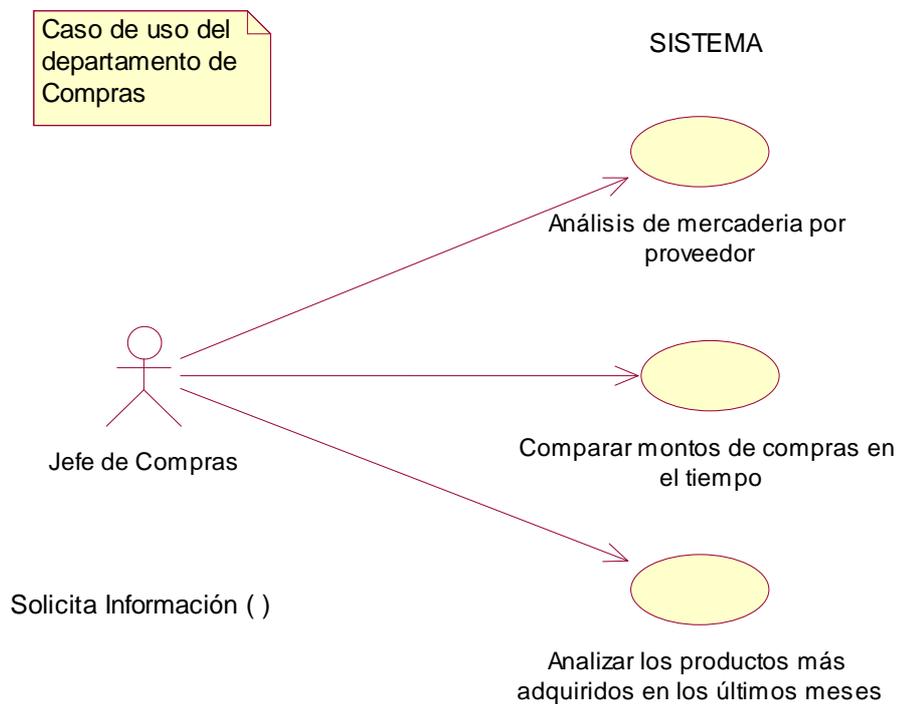


Figura V.62 - Caso de uso # 04

**Tabla V.52 - CU\_DCOMPRAS # 05**

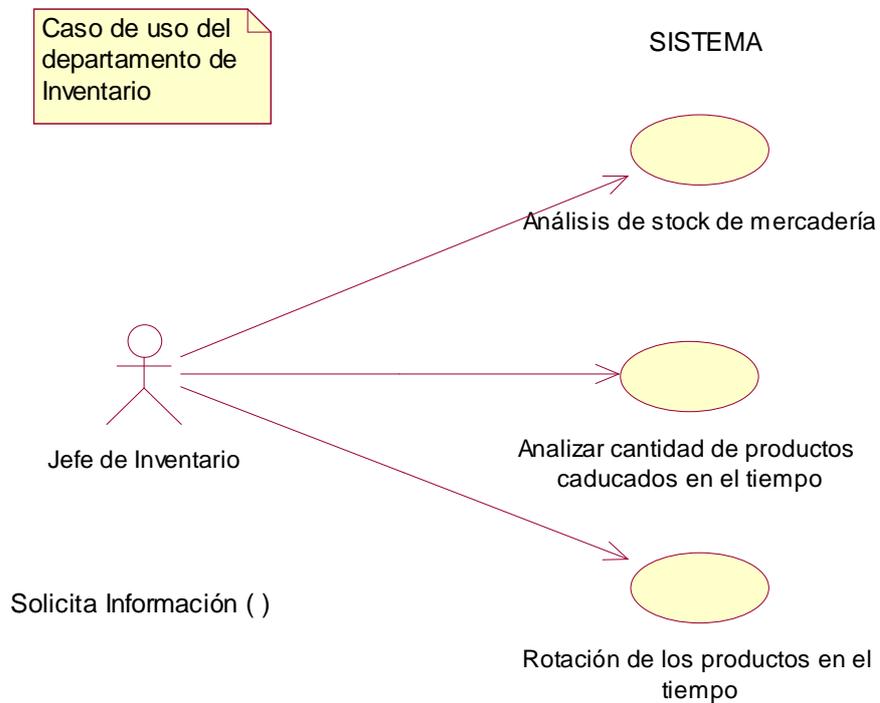
<b>CASO DE USO</b>	Análisis compras
<b>ACTORES</b>	Jefe de compras
<b>TIPO</b>	Primario esencial
<b>PROPOSITO</b>	Analizar el departamento de compras
<b>VISION GENERAL</b>	El usuario accede al sitio web para analizar la información.
<b>REFERENCIAS</b>	CU_AUT, CU_VAOLAP, CU_VREP_EST_ADH, requerimientos departamento de compras.
<b>CURSO TIPICO DE EVENTOS</b>	
<b>ACTOR</b> 1. Solicita información	<b>SISTEMA</b> 2. Análisis de mercadería por proveedor. 3. Comparar montos de compra en los últimos años. 4. Analizar los productos mas adquiridos en los últimos meses.



**Figura V.63 - Caso de uso # 05**

**Tabla V.53 - CU\_DINVENTARIO #06**

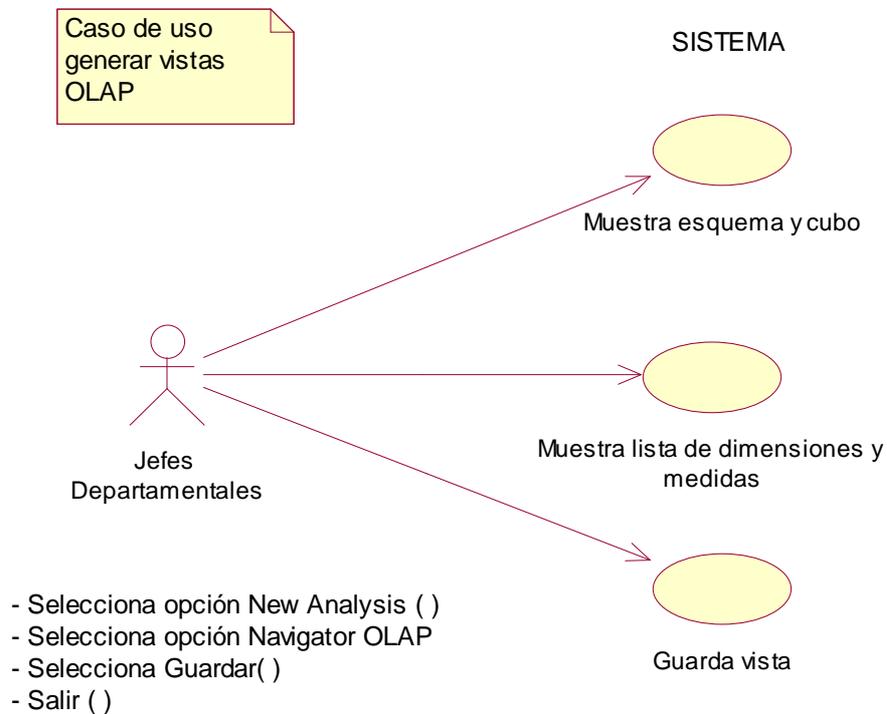
<b>CASO DE USO</b>	Análisis inventario
<b>ACTORES</b>	Jefe de inventario
<b>TIPO</b>	Primario esencial
<b>PROPOSITO</b>	Analizar el departamento de inventario
<b>VISION GENERAL</b>	El usuario accede al sitio web para analizar la información.
<b>REFERENCIAS</b>	CU_AUT, CU_VAOLAP, CU_VREP_EST_ADH, requerimientos departamento de inventario.
<b>CURSO TIPICO DE EVENTOS</b>	
<b>ACTOR</b> 1. Solicita información	<b>SISTEMA</b> 2. Análisis stock de mercadería 3. Analizar cantidad de productos caducados en el trimestre. 4. Rotación de los productos en los últimos meses.



**Figura V.64 - Caso de uso # 06**

**Tabla V.54 - CU\_GVOLAP #07**

<b>CASO DE USO</b>	Generar vistas OLAP
<b>ACTORES</b>	Administrador
<b>TIPO</b>	Primario esencial
<b>PROPOSITO</b>	Generar vistas OLAP
<b>VISION GENERAL</b>	Los jefes departamentales pueden generar los reportes para el análisis, es decir crear cubos OLAP, guardar y visualizarlos.
<b>REFERENCIAS</b>	CU_AUT, CU_VAOLAP,
<b>CURSO TIPICO DE EVENTOS</b>	
<b>ACTOR</b>	<b>SISTEMA</b>
<ol style="list-style-type: none"> <li>1. Selecciona opción New Análisis</li> <li>3. Selecciona la opción Navigator OLAP</li> <li>4. Selecciona guardar</li> <li>6. Salir del sistema.</li> </ol>	<ol style="list-style-type: none"> <li>2. Muestra esquema y cubo</li> <li>4. Muestra lista de dimensiones y medidas.</li> <li>5. Guarda la vista</li> </ol>



**Figura V.65 - Caso de uso # 07**

Tabla V.55 - CU\_GREP\_ADHOC #08

<b>CASO DE USO</b>	Generar reportes Ad Hoc
<b>ACTORES</b>	Jefes departamentales
<b>TIPO</b>	Primario esencial
<b>PROPOSITO</b>	Generar reportes Ad Hoc
<b>VISION GENERAL</b>	Los jefes departamentales pueden generar los reportes guardar y visualizarlos.
<b>REFERENCIAS</b>	CU_AUT, CU_VREP_EST_ADH
<b>CURSO TIPICO DE EVENTOS</b>	
<b>ACTOR</b>	<b>SISTEMA</b>
<ol style="list-style-type: none"> <li>1. Selecciona opción New Reports</li> <li>3. Selecciona campos para generar el reporte</li> <li>4. Selecciona guardar</li> <li>6. Salir del sistema.</li> </ol>	<ol style="list-style-type: none"> <li>2. Muestra un Wizard</li> <li>4. Procesa el reporte</li> <li>5. Guarda el reporte</li> </ol>

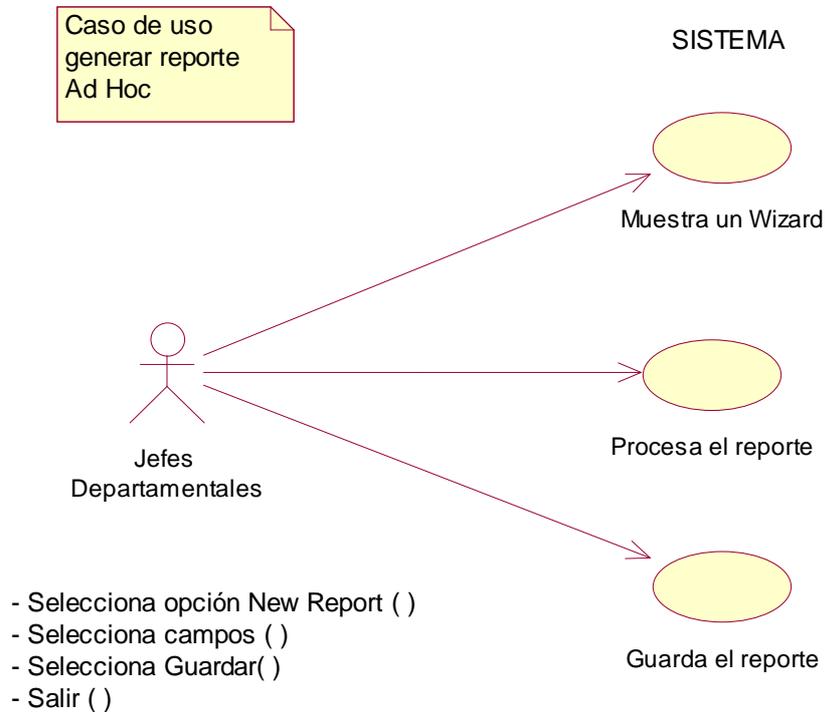
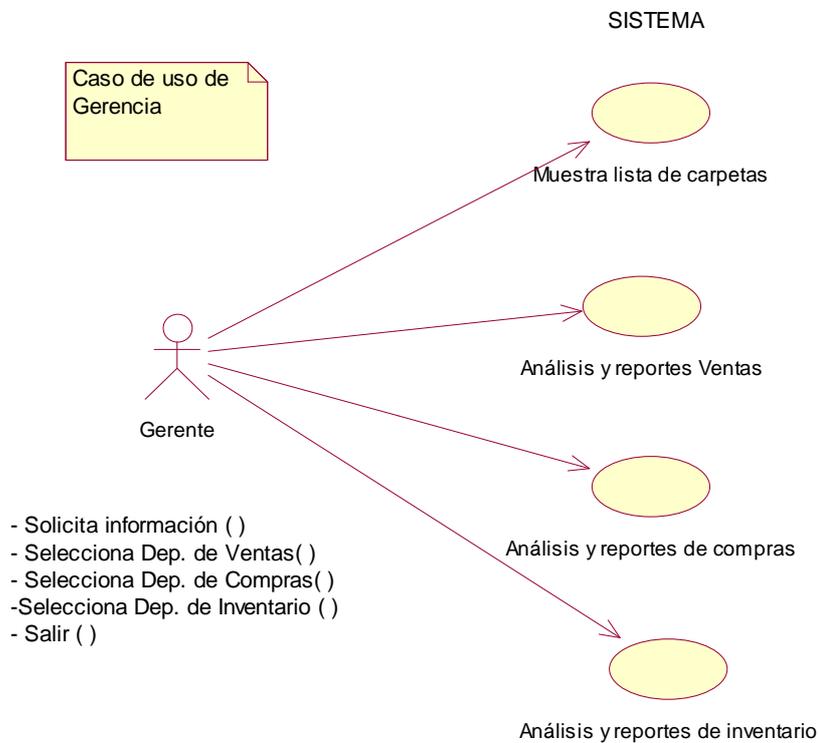


Figura V.66 - Caso de uso # 08

**Tabla V.56 - CU\_AGER #07**

<b>CASO DE USO</b>	Análisis Gerencial
<b>ACTORES</b>	Gerente
<b>TIPO</b>	Primario esencial
<b>PROPOSITO</b>	Informar la situación de la empresa.
<b>VISION GENERAL</b>	Proporcionar al gerente la información necesaria para la toma de decisiones.
<b>REFERENCIAS</b>	CU_AUT, CU_VAOLAP , CU_VREP_EST_ADH, CU_DVENTAS, CU_DCOMPRAS, CU_DINVENTARIO
<b>CURSO TIPICO DE EVENTOS</b>	
<b>ACTOR</b>	<b>SISTEMA</b>
<ol style="list-style-type: none"> <li>1. Solicita información</li> <li>3. Selecciona Departamento de Ventas</li> <li>5. Selecciona Departamento de Compras</li> <li>7. Selecciona Departamento de Inventario</li> <li>9. Salir del sistema.</li> </ol>	<ol style="list-style-type: none"> <li>2. Muestra lista de carpetas de los departamentos</li> <li>4. Análisis OLAP y reportes de Ventas</li> <li>6. Análisis OLAP y reportes de Compras</li> <li>8. Análisis OLAP y reportes de Inventario</li> </ol>



**Figura V.67 - Caso de uso # 09**

#### 5.2.3.4.4 Determinar los datos fuentes.

El modelo de datos que se describe a continuación pertenece al Sistema Integrado, el cual es la fuente de datos para el Sistema de Análisis Gerencial, a partir de este modelo se empieza a construir el Datawarehouse.

##### **Diagrama Entidad-Relación**

No existe un modelo Entidad relación, la base de datos está formada por archivos dbfs, donde el diseño está constituido por tablas sueltas.

##### **Descripción del modelo de datos fuente.**

Los datos fuentes para el proyecto serán la Base de Datos del Sistema Integrado y ciertos archivos de Excel.

En la siguiente tabla describiremos cada una de las entidades de la Base de Datos Fuente que nos servirá para desarrollar el SAG.

**Tabla V.57 – Tablas de la Base de Datos Fuente**

No.	Nombre de la entidad	Descripción
1	Ccdebito	Tiene el movimiento de ventas
2	Vendedor	Tabla de vendedores
3	Zona	Tabla de zonas
4	Incabinv	Tiene cabecera del documento de ingreso-egreso a bodega
5	Indetinv	Tiene el detalle del documento de ingreso-egreso a bodega
6	Icmarti0	Tiene precio de los productos
7	Icmarexis0	Tiene la existencia de los productos.
8	Cppmpvd0	Tabla de proveedores
9	Clientes	Tabla de clientes
10	Icrtico0	Tabla de tipos de documentos de inventario(1A, 11, etc)
11	Indetcst	Tabla de compras
No.	Nombre del archivo	Descripción
1	ProductosCaducador.xls	Archivo que contiene datos de los productos

En la siguiente tabla describiremos cada una de las entidades mencionadas en la tabla anterior con sus respectivos atributos, tipo de datos y descripción de la entidad.

**Tabla V.58 - Modelo de datos fuente**

Nombre de la tabla	Atributos	Tipo de Dato	Descripción
CCDEBITO	Codcia	Carácter (2)	Código de la empresa
	Tipdoc*	Carácter (2)	Tipo de documento (21,22,23,2P)
	Numdoc*	Carácter (6)	Número del documento
	Codbod	Carácter (2)	Código de la bodega
	Numped	Carácter (6)	Número de pedido
	Fectrn*	Date	Fecha
	Pzodia	Numérico (3)	Plazo en días
	Codcli*	Numérico (5)	Código del cliente
	Seccli	Numérico (6)	Sección cliente
	Valdoc*	Numérico (12)	Valor de la factura
	Ptgiva	Numérico (5)	Porcentaje de IVA grabado
	Codusr	Carácter (3)	Código de usuario del sistema
	Hordig	Carácter (5)	Hora digitada la operación
	Usrmod	Carácter (3)	Usuario que modifica
	Hormod	Carácter (5)	Hora que se modifica la factura
	Canmod	Numérico (1)	Cantidad modificada
	Fecvct	Date	Fecha de vencimiento
	Codven*	Numérico (5)	Código del vendedor
	Codzon*	Carácter (5)	Código de zona
	Pagado	Carácter (1)	Documento pagado.
VENDEDOR	Codven*	Numérico (5)	Código del vendedor.
	Nomven*	Carácter (40)	Nombre del vendedor.
	Cupvt	Numérico (10)	Cupo de venta

**Tabla V.58 - Modelo de datos fuente(continuación)**

ZONAS	Codzon*	Carácter (5)	Código de zona
	Codven	Numérico (5)	Código del vendedor
	Nomzon*	Carácter (40)	Nombre de la zona
	Limzon	Carácter (60)	Límite de la zona
	Supzon	Numérico (5)	Supervisor de la zona
	Estzon	Carácter (5)	Estado de la zona
INCABINV	Codcli	Carácter (2)	Código de la empresa
	Tipdoc	Carácter (2)	Tipo de documento
	Numdoc	Carácter (6)	Número del documento
	Codbod	Carácter (2)	Código de bodega
	Fectrn	Date	Fecha
	Pzodia	Numérico (3)	Plazo en días
	Codpvd	Numérico (5)	Código de proveedor
	Facpvd	Carácter (15)	Número de factura
	Ptgiva	Numérico (5)	Porcentaje de IVA
	Observación	Memo (10)	Memo
	Hordig	Carácter (5)	Hora digitado el documento
	Codusr	Carácter (3)	Código de usuario
	Codcia	Carácter (2)	Código de la empresa
	Numdoc*	Carácter (6)	Número de documento
	Tipdoc*	Carácter (2)	Tipo de documento
	Codbod	Carácter (2)	Código de bodega
	Codprd*	Numérico (5)	Código del proveedor
	Grpfam	Carácter (6)	Grupo o familia de productos
	Carcaj	Numérico (5)	Cantidad en cajas
	Canund	Numérico (4)	Cantidad en unidades
	Canprd*	Numérico (7)	Cantidad de productos

**Tabla V.58 - Modelo de datos fuente(continuación)**

INCABINV	Prcunt	Numérico (10)	Precio de venta al público (no esta en todos los productos)
	Pagiva	Lógico (1)	Producto gravado con IVA 12% o 0%
	Dscto	Numérico (5)	Descuento
	Fectrn	Date	Fecha
	Cstunt*	Numérico (14)	Precio de costo(no esta en todos los productos)
	Ingegr	Carácter (1)	Si es ingreso o egreso.
ICMARTIO	Cod_artic*	Numérico (5)	Código del artículo
	Nom_artic	Carácter (40)	Nombre del artículo
	Codi_equiv	Carácter (15)	Código de barra
	Unidad_medid	Carácter (4)	Unidad de medida
	Pres_artic	Carácter (4)	Presentación de artículos
	Marc_artic	Carácter (4)	Marca del artículo
	Mode_artic	Carácter (4)	Modelo del artículo
	Proc_artic	Carácter (4)	Procedencia del artículo
	Tope_fact	Numérico (4)	Cantidad en cajas
	Codi_prove*	Numérico (5)	Código del proveedor
	Cost_provi*	Numerico	Costo del producto
	Cost_pro98*	Numérico (15)	Costo del producto en cajas
	Prec_punid*	Numérico (12)	Pvp unidad
	Prec_pdoce*	Numérico (12)	Pvp docena
	Prec_pcjbt*	Numérico (12)	Pvp caja, jaba, bulto,
	Cod_iva	Numérico (1)	Código del IVA 1,0
	Porc_dscto	Numérico (6)	Porcentaje de descuento
	Porc_prbod	Numérico (5)	Porcentaje utilidad en cjbt
	Porc_prdoc	Numérico (5)	Porcentaje utilidad en docena
	Porc_prund	Numérico (5)	Porcentaje de utilidad en unidad
	Nomb_gener*	Carácter (25)	Descripción que se mostrará para listar productos en la ventana de facturación.

**Tabla V.58 - Modelo de datos fuente(continuación)**

ICMARTIO	Apli_artic	Carácter (20)	Cantidad de producto que viene en caja.
	Clas_conce	Carácter (6)	
	Ubic_fisic	Carácter (6)	Ubicación de la bodega
	Stat_regis	Carácter (1)	Estado del producto
	Cant_docen	Numérico (2)	Cantidad que determina la docena
	Fecha_ullis	Date	Fecha de la última lista
	Ident_user	Carácter (3)	Identificación del usuario que registra precios
ICMEXISO	Codi_empre	caracter (2)	Código de la empresa
	Codi_bodeg	caracter (2)	Código de bodega
	Codi_artic*	Numerico	Código del artículo
	Nomb_artic*	caracter (14)	Nombre del artículo
	Cant_exist*	Numerico	Existencia de productos
	Ulti_fingr	Date	Ultima fecha de ingreso
CPPMPVD0	Código	Numérico (5)*	Código del proveedor
	Cod_conta	Carácter (14)	Código contable
	Nombre	Carácter (40)*	Nombre del proveedor
	Ruc	Carácter (15)	Registro único de contribuyentes
	Diascred	Numérico (3)	Días de crédito
	Direc1	Carácter (35)	Dirección
	Vended1	Carácter (35)	Vendedor
	Telefax	Numérico (8)	Número de telefax
	Telef1	Numérico (8)	Teléfono
	Telef2	Numérico (8)	
	Ciudad	Carácter (15)	
	CONSEP	Lógico (1)	
	Pernat	Lógico (1)	

**Tabla V.58 - Modelo de datos fuente(continuación)**

CLIENTES	Codclt	Numérico (6)	Código del cliente
	Codven	Numérico (5)	Código del vendedor
	Zonacl	Carácter (5)	Zona del cliente
	Rutaclt	Numérico (3)	Ruta del cliente
	Locali	Carácter (15)	Localidad
	Nomclt	Carácter (40)	Nombre del cliente
	Ruc_cedula	Carácter (15)	Ruc o cédula del cliente
	Dirclt	Carácter (40)	Dirección del cliente
	Tlfclt	Carácter (11)	Teléfono
	Rznsoc	Carácter (30)	Razón social
	Tele_local	Carácter (11)	Teléfono del local
	Forma_pago	Carácter (40)	Forma de pago
	Días	Carácter (10)	Días
	Montar.	Numérico (10)	Monto de crédito
	Observ	Carácter (80)	Observación
	Estatus	Carácter (1)	Estado del cliente
	Catego	Carácter (4)	Categoría
	Ultfac	Numérico (5)	Última factura
	Saldo_p	Numérico (15)	Saldo del cliente
	Plzodia	Numérico (3)	Plazo en días
Usrmod	Carácter (3)	Nombre de usuario	
Hormod	Carácter (5)	Hora modificada	
Fecmod	Date	Fecha modificada	
Icrtico0	codi_empre	Carácter (2)	Código de la empresa
	Codi_compr	Carácter (2)*	Código de la compra

**Tabla V.58 - Modelo de datos fuente(continuación)**

Icrtico0	Tipo_compr	Carácter (3)	Tipo de comprobante
	Nomb_docum	Carácter (30)*	Nombre del documento
	Nomb_abrev	Carácter (10)	Nombre abreviado
	Tipo_conta	Numérico	Contabilidad tipo
	Form_impre	Numérico	Forma Impresa
	Afec_exist	Carácter (1)	Afecta a la existencia
	Afec_prec1	Carácter (1)	Afecta al precio
	Ulti_docum	Numérico	Último documento
	Fech_ultim	Date	Última fecha
	Firm_apro1	Carácter (16)	Firma de aprobación
	Firm_apro2	Carácter (16)	Firma de recepción
	Stat_regis	Carácter (1)	Registro de estado
	User_habil	Carácter (80)	Usuarios habilitados
Indetcst	Codcia	Carácter (2)	Código de la empresa
	Tipdoc*	Carácter (2)	Tipo de documento (11, 1D)
	Numdoc	Carácter (6)	Numero de ingreso a inventario
	Codbod	Carácter (2)	Código de bodega
	Codpvd*	Numerico	Código de proveedor
	Codprd*	Numerico	Código de producto
	Fectrn*	Date	Fecha
	Canprd*	Numerico	Cantidad de productos
	Canped	Numerico	Cantidad pedida
	Undxcj	Carácter (1)	Unidades q vienen en cajas
	Undcaj	Carácter (1)	Unidades o cajas
	Prc1st	Numerico	Precio en lista
	Cstprm	Numerico	Costo promoción
	Cstunt *		Costo unitario
	Estados.xls		

**5.2.3.4.5 Definir base de datos intermedia de la bodega de datos.**

Tomando en cuenta a la Fase II del proyecto, la empresa Cohervi S.A, necesita información para cada departamento:

El Departamento de Ventas, Inventario y Producción los requerimientos de los distintos departamentos, e construirá una base de datos intermedia llamada dwh\_cohervi. En la siguiente tabla se indicará las entidades que serán tomadas en cuenta para desarrollar el Datawarehouse del SAG.

**Tabla V.59 – Descripción de la BD Fuente y BD Intermedia**

<b>Nombre de la entidad BD fuente</b>	<b>Nombre de la entidad del dwh</b>	<b>Descripción de la entidad intermedia</b>
Ccdebito	dwh_movimientos	Movimientos de ventas
Vendedor	dwh_vendedor	Vendedores
Zona	dwh_zona	Zonas
Indetinv	Dwh_kardex	Detalle de inventarios
Indetcst	Dwh_compras	Ingreso de compra de productos
Icmarti0	Dwh_producto	Descripción, Costos y pvp de los productos
Cientes	Dwh_cliente	Datos de los clientes
Cppmpvd0	Dwh_proveedor	Datos de los proveedores
Icrtico0	dwh_tipodoc	Tipo de documentos de inventario (1N, 11,10, etc)
Estados	dwh_plancuentas	Datos del plan de cuentas.

A continuación se detalla las tablas fuente y tablas de la base de datos intermedia utilizadas en la aplicación con sus respectivos campos:

**Tabla V.60 - Detalle de los campos de la BD Fuente - Destino**

<b>BASE DE DATOS SISTEMA</b>	
<b>FUENTE</b>	<b>DESTINO</b>
<b>Ccdebito</b> Codcia Tipdoc Numdoc Codbod Numped Fectrn Pzodia Codcli Seccli Valdoc Ptgiva Codusr Hordig Usrmod Hormod Canmod Fecvct Codven Codzon Pagado	<b>Dwh_movimientos</b> tipdoc numdoc fectrn codcli valdoc codven codzon
<b>Vendedor</b> Codven Nomven Cupvt	<b>Dwh_vendedor*</b> codven nomven
<b>Zona</b> Codzon Codven Nomzon Limzon Supzon	<b>Dwh_zona*</b> codzon codven nomzon

**Tabla V.60 - Detalle de los campos de la BD Fuente – Destino (continuación)**

<p><b>Indetinv</b></p> <p>Codcia          Numdoc*          Tipdoc*          Codbod          Codprd*          Grpfam          Carcaj          Canund          Canprd*          Prcunt          Pagiva          Dscto          Fectrn*          Cstunt          Ingegr</p>	<p><b>Dwh_kardex</b></p> <p>numdoc          tipdoc          codprd          ingegr          fectrn          canprd</p>
<p><b>BASE DE DATOS INVENTPT</b></p>	
<p><b>Icmarti0</b></p> <p>Cod_artic*          Nom_artic          Codi_equiv          Unidad_medid          Pres_artic          Marc_artic          Mode_artic          Proc_artic          Tope_fact          Codi_prove*          Cost_provi*          Cost_pro98*          Prec_punid*          Prec_pdoce*          Prec_pcjbt*          Cod_iva          Porc_dscto          Porc_prbod</p>	<p><b>Dwh_producto</b></p> <p>Cod_artic          Codi_prove          Tope_fact1 (cantidad en cajas)          Cost_provi(precio costo unidad)          Cost_pro98 (precio costo caja)          Prec_unid          Prec_pdoce          Prec_pcjbt          Nomb_gener</p>

**Tabla V.60 - Detalle de los campos de la BD Fuente – Destino (continuación)**

Porc_prdoc Porc_prund Nomb_gener* Apli_artic Clas_conce Ubic_fisic Stat_regis Cant_docen Fecha_ullis Ident_user <b>Cppmpvd0</b> Codigo Cod_conta Nombre RUC Diascred Direc1 Vended1 Telefax Telef1 Telef2 Ciudad CONSEP Pernat	<b>Dwh_proveedor</b>  codigo nombre diascred
<b>Icrtico0</b> codi_empre Codi_compr Tipo_compr Nomb_docum Nomb_abrev Tipo_conta Form_impre Afec_exist Afec_prec1 Ulti_docum Fech_ultim	<b>dwh_tipovta</b>

**Tabla V.60 - Detalle de los campos de la BD Fuente – Destino (continuación)**

BASE DE DATOS COBRANZAS		
Firm_apro1	<b>Dwh_cliente</b>	
Firm_apro2		
Stat_regis		
User_habil		
<b>Clientes</b>		
Codclt		
Codven		
Zonacl		
Rutaclt		
Locali		
Nomclt		
Ruc_cedula		
Dirclt		
Tifclt		
Rznsoc		
Tele_local		
Forma_pago		
Dias		
Montar.		
Observ		
Estatus		
Catego		
Utlfac		
Saldo_p		
Plzodia		
Usrmod		
Hormod		
Fecmod		
FUENTES EXTERNAS (ProductosCaducados.xls)		

### 5.2.3.4.6 Realizar el modelo dimensional.

Para desarrollar el modelo dimensional del Sistema de Análisis Gerencial se desarrollará Datamarts por cada departamento de la empresa.

A continuación definimos las tablas de hechos y las dimensiones que utilizaremos en el desarrollo del sistema:

#### **Tablas de Hechos**

- Fact\_compras
- Fact\_inventario
- Fact\_ventas
- Fact\_caducados

#### **Dimensiones**

De acuerdo a los requerimientos de la fase II, en la siguiente tabla se detallará las dimensiones que se utilizará en el desarrollo del Sistema de Análisis Gerencial.

**Tabla V.61 - Dimensiones y Fact Tables**

	<b>Cliente</b>	<b>TipoVenta</b>	<b>Vendedor</b>	<b>Producto</b>	<b>Tiempo</b>	<b>Zona</b>	<b>Proveedor</b>
Fact_compras				X	X		X
Fact_inventario		X		X	X		
Fact_ventas	X	X	X	X		X	
Fact_caducados				X			

- ✚ **Establecer el nivel de granulación de cada dimensión y los niveles de agregación de cada tabla de hechos.**

La siguiente tabla indica el nivel de granulación de las dimensiones que se utilizará en la construcción de los Datamarts.

**Tabla V.62 - Granulidad**

<b>Dimensión</b>	<b>Granulación</b>
Dim_vendedores	Id, nombre
Dim_clientes	Id, codigo, nombres, estado, dirección, montocredito, plazodias, saldopendiente,
Dim_proveedores	Id,nombre
Dim_productos	Id, codigo, nombre, p_unidad, p_docena, p_cajas, c_compra, estado
Dim_tiempo	Idtiempo, Fecha, anio, semestre, trimestre, mesnumero, mesnombre, día
Dim_tipocomprobantes	Id, codigo, nombre
Dim_zona	Id, descripción

En esta tabla detallaremos las medidas que se utilizaran en la elaboración de los DataMarts. Veamos la tabla.

**Tabla V.63 - Tabla de Hechos y medidas**

<b>Tablas de Hecho</b>	<b>Agregación</b>
Fact_compras	Cantidad, montocompras
Fact_inventario	Ingreso, egreso
Fact_ventas	Importe
Fact_caducados	Diasvence, diascaducidad

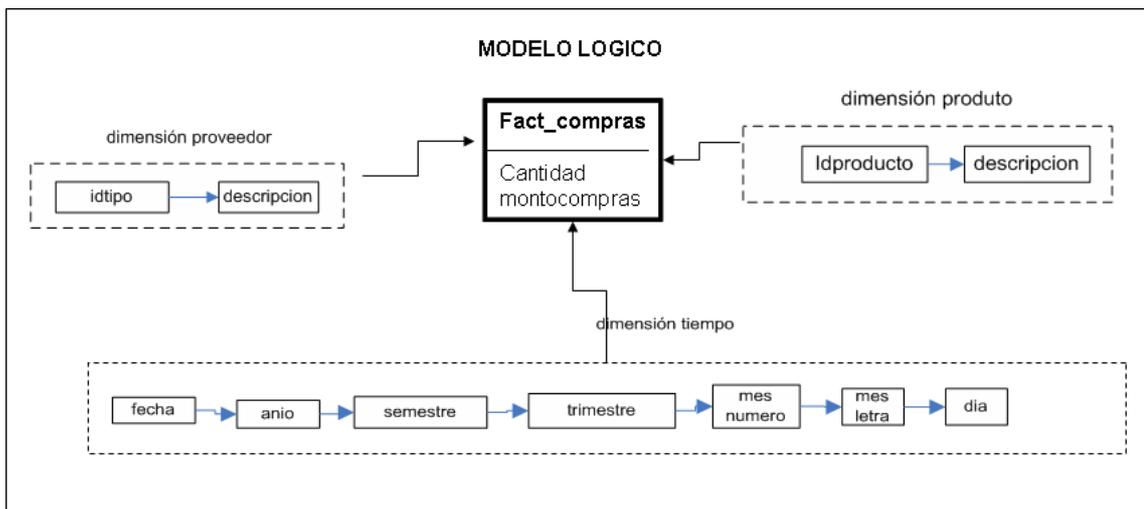
✚ **Elaborar diagrama en estrella que represente el datamart.**

Para desarrollar el sistema SAG, se ha considerado a tres departamentos como son Compras, Inventario y Ventas, por tal razón el sistema consta de cuatro tres Datamats siendo los siguientes:

1. Compras
2. Inventario
3. Ventas

**Datamart Compras**

Este Datamarts trabajará con tres dimensiones, las mismas que son Dimensión proveedor, producto y tiempo.



**Figura V.68 – Datamart Compras**

## Datamart Inventario

Este Datamart estará formado por dos Fact Tables: inventario y caducados, las dimensiones son: tipcomprobantes, productos, tiempo.

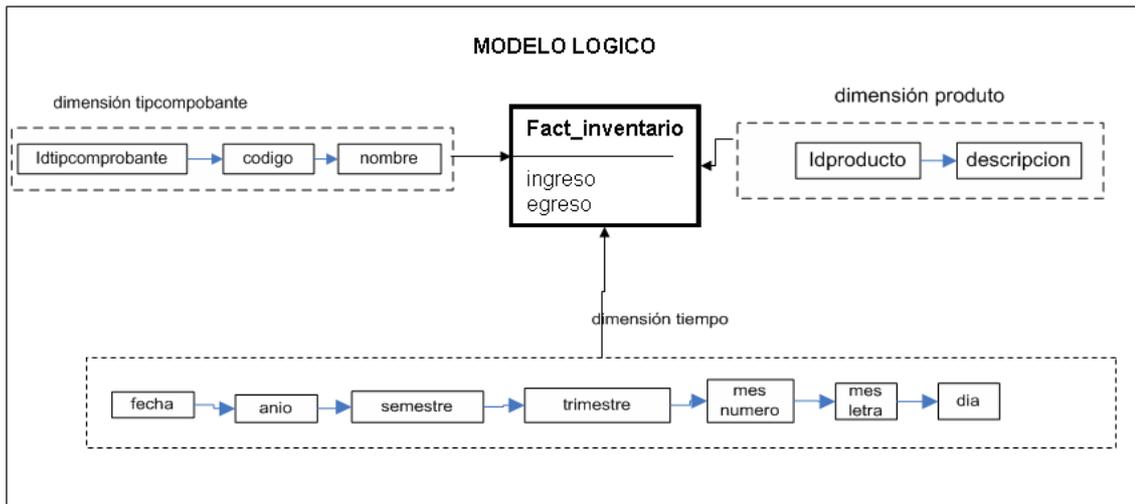


Figura V.69 – Datamart Inventario (Fact\_inventario)

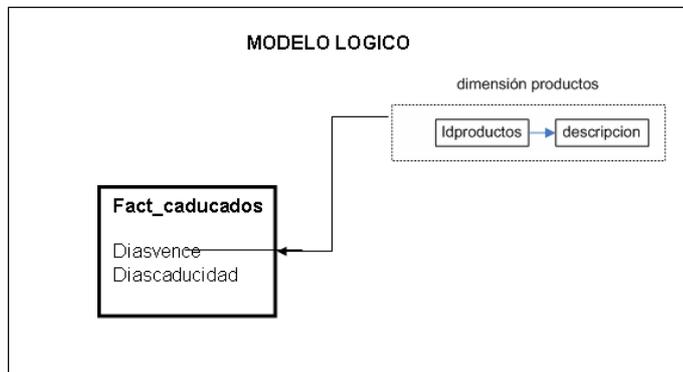


Figura V.70 - Datamart Inventario (Fact\_caducados)

## Datamart Ventas

Este Datamart constará de tres dimensiones tales como: dimensión tiempo, dimensión producto y dimensión cliente.

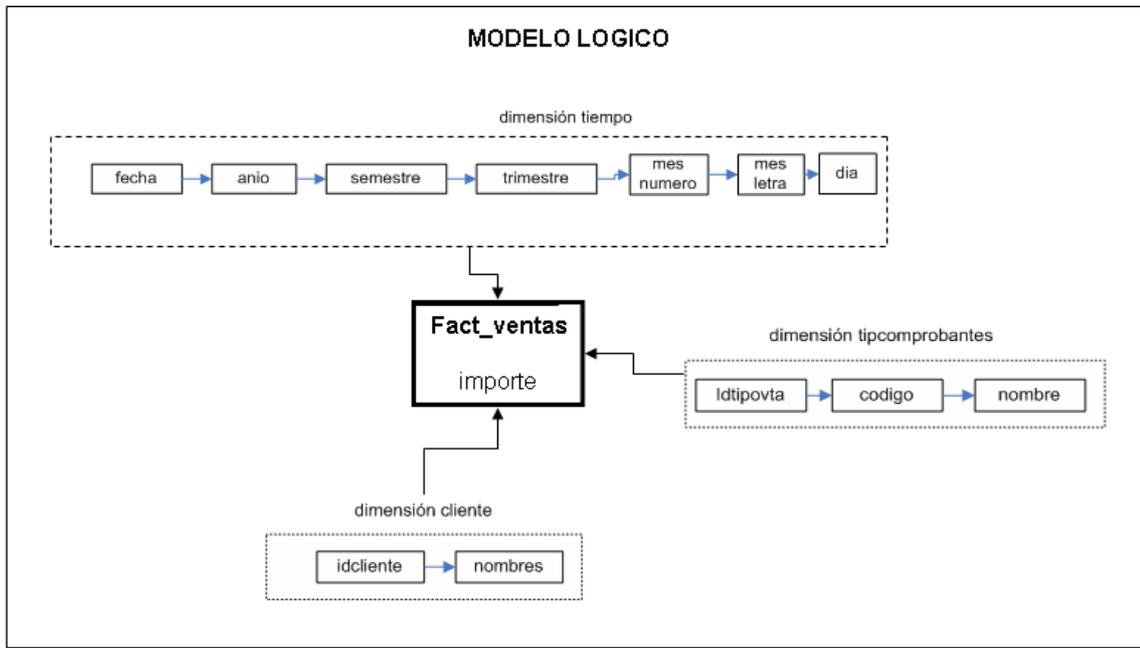


Figura V.71 - Datamart ventas.

### 5.2.3.4.7 Elaborar el modelo físico.

Se realizará la transformar el modelo lógico conceptual en la estructura física.

#### 🚦 Verificación del modelo dimensional.

A través de la siguiente matriz verificaremos el cumplimiento de los requerimientos de los distintos departamentos de la empresa Cohervi S.A.

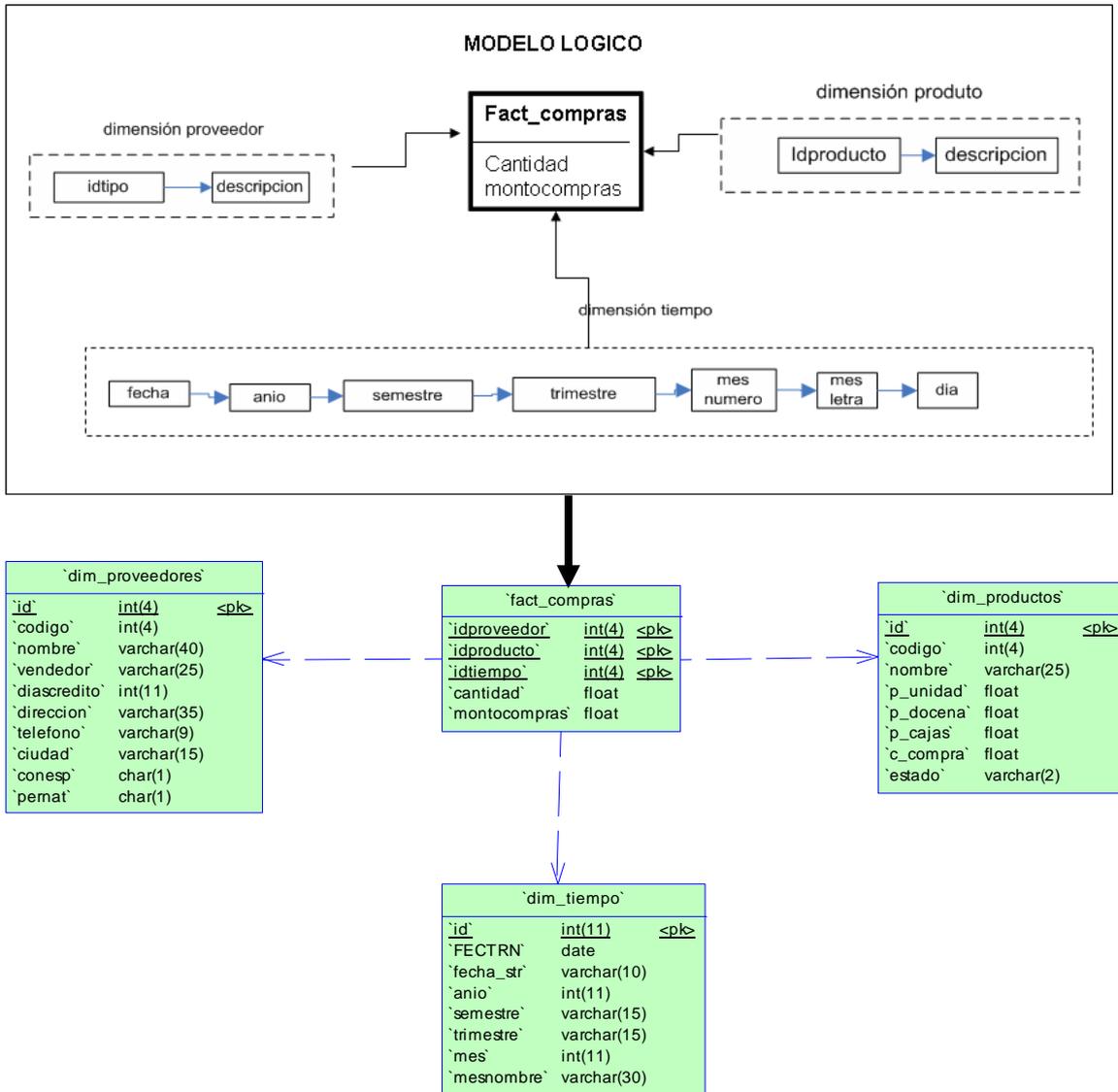
**Tabla V.64 - Matriz de verificación del modelo dimensional**

	compras	caducados	ventas	Inventario	Tipocompr obantes	Zona	vendedor	Cliente	proveedor	Producto	tiempo
<b>VENTAS</b>											
R1			X		X						X
R2			X			X					X
R3			X								X
R4			X			X	X				X
R5			X				X				X
R6			X					X			X
<b>COMPRAS</b>											
R1	X									X	X
R2	X										X
R3	X									X	X
<b>INVENTARIO</b>											
R1				X						X	X
R2		X								X	X
R3				X						X	X

**Definición del esquema físico de las dimensiones y sus jerarquías.**

Una vez definidas las tablas de hecho y las dimensiones precedemos a realizar la conversión del modelo lógico al modelo físico con el que trabajaremos para la elaboración de los cubos multidimensionales.

**DATAMART COMPRAS**



**Figura V.72 - Conversión del modelo lógico al físico data mart compras**

## DATAMART INVENTARIO

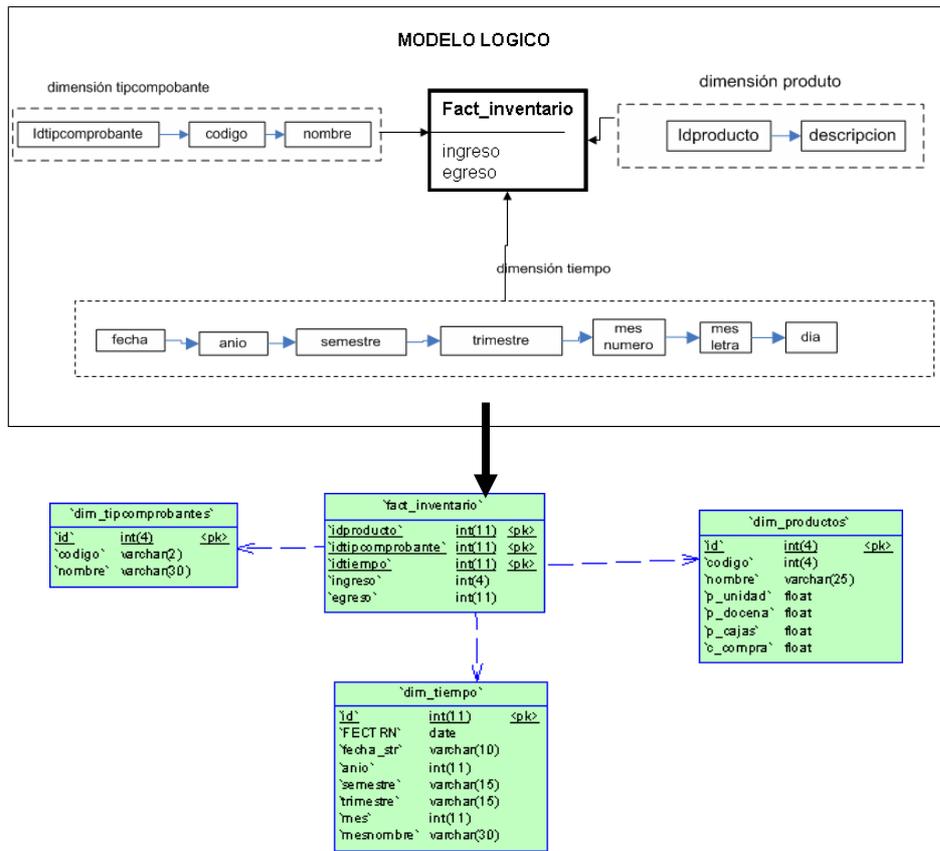


Figura V.73 - Conversión del modelo lógico al físico data mart inventario (Fact\_inventario)

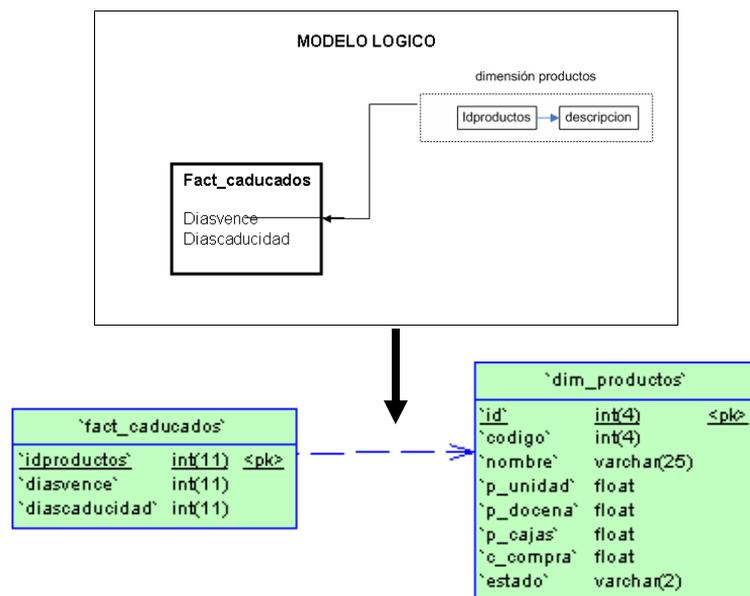


Figura V.74 - Conversión del modelo lógico al físico data mart inventario (Fact\_caducados)

## DATAMART VENTAS

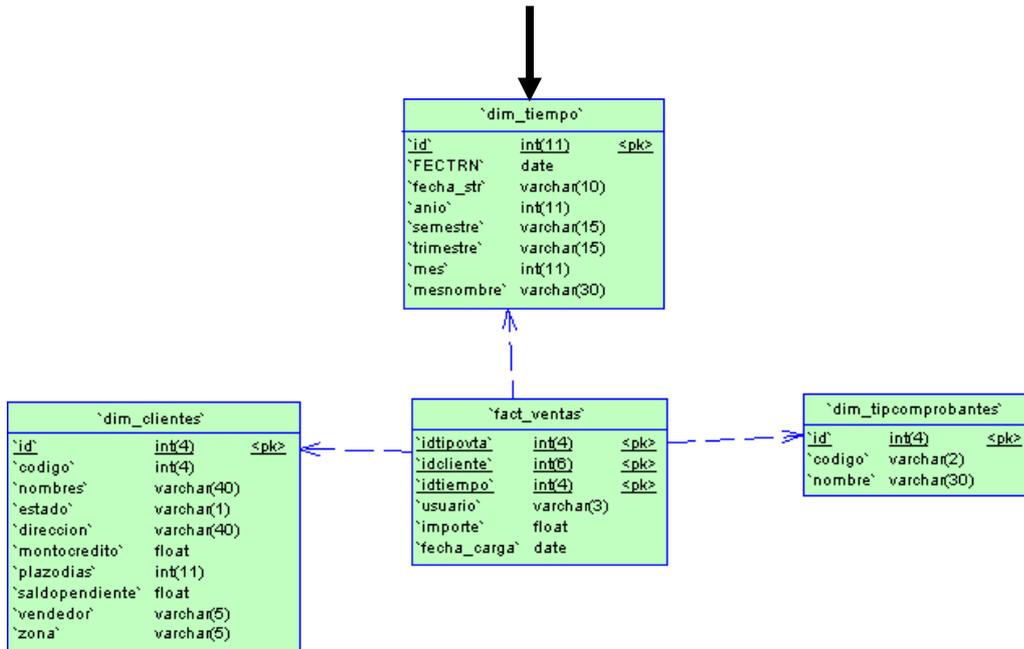
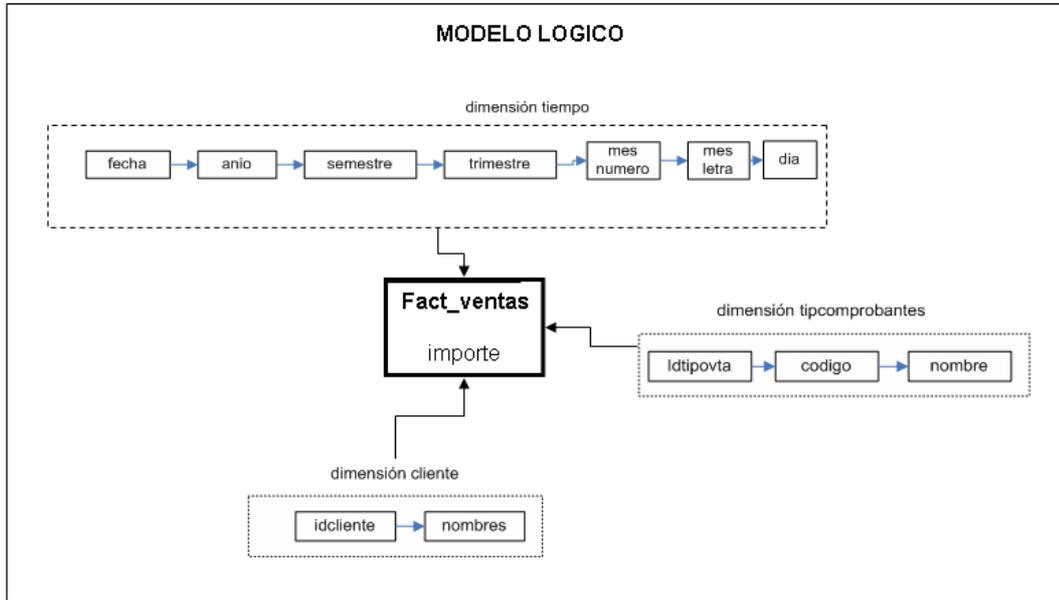


Figura V.75 - Conversión del modelo lógico al físico data mart ventas

### Definición de los atributos que conforman las tablas de hecho.

La siguiente tabla muestra los atributos de cada tabla de hechos y cada dimensión que conforman cada estructura.

**TABLAS DE HECHOS**

**Tabla V.65 - Atributos de la tabla de hechos Compras**

<b>Nombre del Datamart</b>	<b>Compras</b>		
<b>Tabla de hechos</b>	<b>Fact_Compras</b>		
<b>Atributo</b>	<b>Tipo</b>	<b>Pk</b>	<b>Descripción</b>
Idproveedor	Integer	Si	Identif. De la dimensión proveedor
Idproducto	Integer	Si	Identif. De la dimensión producto
Idtiempo	Integer	Si	Identif. De la dimensión tiempo
Cantidad	Float		Valor 1 de la tabla de hechos
Montocompras	Float		Valor 2 de la tabla de hechos

**Tabla V.66 - Atributos de la tabla de hechos ventas**

<b>Nombre del Datamart</b>	<b>Ventas</b>		
<b>Tabla de hechos</b>	<b>Fact_ventas</b>		
<b>Atributo</b>	<b>Tipo</b>	<b>Pk</b>	<b>Descripción</b>
Idtipovta	Integer	Si	Identif. De la dimensión tipocomprobantes
Idcliente	Integer	Si	Identif. De la dimensión cliente
Idtiempo	Integer	Si	Identif. De la dimensión tiempo
Importe	Float		Valor de la tabla de hechos

**Tabla V.67 - Atributos de la tabla de hechos inventario (Fact\_inventario)**

<b>Nombre del Datamart</b>	<b>Inventario</b>		
<b>Tabla de hechos</b>	<b>Fact_inventario</b>		
<b>Atributo</b>	<b>Tipo</b>	<b>Pk</b>	<b>Descripción</b>
Idproductos	Integer	Si	Identif. De la dimensión producto
Idtipocomprobantes	Integer	Si	Identif. De la dimensión tipocomprobante
Idtiempo	Date	Si	Identif. De la dimensión tiempo
Ingreso	Real		Valor 1 de la tabla de hechos
Egreso	Real		Valor 2 de la tabla de hechos

**Tabla V.68 - Atributos de la tabla de hechos inventario (Fact\_caducados)**

<b>Nombre del Datamart</b>	<b>Inventario</b>		
<b>Tabla de hechos</b>	<b>Fact_caducados</b>		
<b>Atributo</b>	<b>Tipo</b>	<b>Pk</b>	<b>Descripción</b>
Idproductos	Integer	Si	Identif. De la dimensión producto
Diasvence	Integer		Valor 1 de la tabla de hechos
Diascaducidad	Integer		Valor 2 de la tabla de hechos

## DIMENSIONES

Los atributos de cada una de las dimensiones son las siguientes:

**Tabla V.69 - Atributos de la dimensión tipo de venta**

<b>Nombre del Datamart</b>			
<b>Dimensión</b>	<b>Tipoventa</b>		
<b>Atributo</b>	<b>Tipo</b>	<b>Pk</b>	<b>Descripción</b>
Id	Integer	Si	Id de la dimensión
Codigo	Varchar(2)		Codigo del tipo de comprobante
Nombre	Varchar(30)		Nombre del comprobante

**Tabla V.70 - Atributos de la dimensión tiempo**

<b>Nombre del Datamart</b>			
<b>Dimensión</b>	<b>Tiempo</b>		
<b>Atributo</b>	<b>Tipo</b>	<b>Pk</b>	<b>Descripción</b>
Id	Integer	Si	Pk de la dimensión
Fecha	Date		La fecha
Anio	Varchar(15)		Anio en letras
semestre	Varchar(15)		Semestre del año
Trimestre	Varchar(15)		Trimestre del año
Mesnumero	Integer		Mes en número
Mesletra	Varchar(15)		Mes en letra

**Tabla V.71 - Atributos de la dimensión vendedor**

<b>Nombre del Datamart</b>			
<b>Dimensión</b>	<b>Vendedor</b>		
<b>Atributo</b>	<b>Tipo</b>	<b>Pk</b>	<b>Descripción</b>
Idvendedor	Integer	Si	Id de la dimensión
Nombre	Varchar(50)		Nombre del vendedor

**Tabla V.72 - Atributos de la dimensión cliente**

<b>Nombre del Datamart</b>			
<b>Dimensión</b>	<b>Cliente</b>		
<b>Atributo</b>	<b>Tipo</b>	<b>Pk</b>	<b>Descripción</b>
Id	Integer	Si	Id de la dimensión
Código	Integer		Código del cliente
Nombres	Varchar(40)		Nombre del cliente
Estado	Varchar(1)		Estado del cliente
Dirección	Varchar(40)		Dirección del cliente
Montocredito	Float		Monto de crédito del cliente
Plazodias	Integer		Plazo en días del cliente
Saldopendiente	Float		Saldo del cliente

**Tabla V.73 - Atributos de la dimensión zona**

<b>Nombre del Datamart</b>	<b>Ventas</b>		
<b>Dimensión</b>	<b>Zona</b>		
<b>Atributo</b>	<b>Tipo</b>	<b>Pk</b>	<b>Descripción</b>
Idzona	Integer	Si	Id de la dimensión
Descripción	Varchar(50)		

**Tabla V.74 - Atributos de la dimensión producto**

Nombre del Datamart	Ventas		
Dimensión	Producto		
Atributo	Tipo	Pk	Descripción
Id	Integer	Si	Id de la dimensión
Codigo	Integer		Código del producto
Nombre	Varchar(25)		Nombre del producto
P_unidad	Float		Precio unidad
P_docena	Float		Precio docena
P_cajas	Float		Precio cajas
C_compra	Float		Costo de compra
Estado	Varchar(2)		Estado del producto

**Tabla V.75 - Atributos de la dimensión proveedor**

Nombre del Datamart			
Dimensión	Proveedor		
Atributo	Tipo	Pk	Descripción
Id	Integer	Si	Id de la dimensión
Codigo	Integer		
Nombre	Varchar(50)		Nombre del proveedor
Vendedor	Varchar(25)		Nombre del vendedor
Diascredito	Integer		Días de crédito
Direccion	Varchar(50)		Dirección del proveedor
Telefono	Varchar(9)		Teléfono del proveedor
Ciudad	Varchar(15)		Ciudad del proveedor
Conesp	Varchar(1)		Contribuyente especial
Pernat	Varchar(1)		Persona natural

### 5.2.3.4.8 Diseñar cubos multidimensionales.

Para el SAG trabajará con cuatro cubos llamados compras, ventas, inventario y caducados.

**Tabla V.76 - Cubo multidimensional de Compras**

<b>Nombre del cubo:</b>		Compras	
<b>Requerimiento de usuario:</b>		Requerimientos del departamento de compras	
<b>Data source:</b>		ddm_cohervi	
<b>Tabla fuente:</b>		Fact_compras	
<b>Dimensiones</b>	<b>Nombre</b>	<b>Tabla fuente</b>	
	Proveedores	Dim_proveedores	
	Productos	Dim_productos	
	Tiempo	Dim_tiempo	
<b>Medidas</b>	<b>Nombre</b>	<b>Campo fuente</b>	<b>Función</b>
	Cantidad	Cantidad	Suma
	Montocompras	Montocompras	Suma
<b>Miembros calculados</b>	<b>Nombre</b>	<b>Función de cálculo</b>	
	n/a	n/a	
<b>Tipo de almacenamiento</b>		ROLAP	
<b>Lectura/Escritura</b>		Lectura	
<b>Roles</b>		Administrador, compras	

**Tabla V.77 - Cubo multidimensional de ventas**

<b>Nombre del cubo:</b>		Ventas	
<b>Requerimiento de usuario:</b>		Requerimientos del departamento de ventas	
<b>Data source:</b>		Ddm_cohervi	
<b>Tabla fuente:</b>		Fact_ventas	
<b>Dimensiones</b>	<b>Nombre</b>	<b>Tabla fuente</b>	
	Tipo venta	dim_tipcomprobantes	
	Cliente	Dim_clientes	
	Zona	Dim_clientes	
	Vendedores	Dim_clientes	
	Tiempo	Dim_tiempo	
<b>Medidas</b>	<b>Nombre</b>	<b>Campo fuente</b>	<b>Función</b>
	Importe	importe	Suma
<b>Miembros calculados</b>	<b>Nombre</b>	<b>Función de cálculo</b>	
	n/a	n/a	
Tipo de almacenamiento		ROLAP	
Lectura/Escritura		Lectura	
Roles		Administrador, ventas	

**Tabla V.78 - Cubo multidimensional de inventario (Fact\_inventario)**

<b>Nombre del cubo:</b>		Inventario	
<b>Requerimiento de usuario:</b>		Requerimientos del departamento de inventario	
<b>Data source:</b>		Ddm_cohervi	
<b>Tabla fuente:</b>		Fact_inventario	
<b>Dimensiones</b>	<b>Nombre</b>	<b>Tabla fuente</b>	
	Tipos	Dim_tipcomprobantes	
	Comprobantes		
	Productos	Dim_productos	
	Tiempo	Dim_tiempo	
<b>Medidas</b>	<b>Nombre</b>	<b>Campo fuente</b>	<b>Función</b>
	Ingreso	Ingreso	Suma
	Egreso	Egreso	suma
<b>Miembros calculados</b>	<b>Nombre</b>	<b>Función de cálculo</b>	
	Stock	Ingreso-egreso	
Tipo de almacenamiento		ROLAP	
Lectura/Escritura		Lectura	
Roles		Administrador, inventario	

**Tabla V.79 - Cubo multidimensional de inventario (Fact\_caducados)**

<b>Nombre del cubo:</b>		Caucados	
<b>Requerimiento de usuario:</b>		Requerimientos del departamento de inventario	
<b>Data source:</b>		Ddm_cohervi	
<b>Tabla fuente:</b>		Fact_caducados	
<b>Dimensiones</b>	<b>Nombre</b>	<b>Tabla fuente</b>	
	Productos	Dim_productos	
<b>Medidas</b>	<b>Nombre</b>	<b>Campo fuente</b>	<b>Función</b>
	Diasvence	Diasvence	Suma
	Diascaducidad	Diascaducidad	suma
<b>Miembros calculados</b>	<b>Nombre</b>	<b>Función de cálculo</b>	
	Stock	Stock	
Tipo de almacenamiento		ROLAP	
Lectura/Escritura		Lectura	
Roles		Administrador, inventario	

 **Seleccionar los operadores de las Herramientas OLAP.**

Los operadores con los que trabajará el SAG serán con los mas conocidos, los mismos que les permitirán a los usuarios a navegar fácilmente. Estos operadores son:

- **Roll - Up (drill-up):** el usuario puede subir en jerarquía o reducir las dimensiones
- **Drill-Down (roll-down):** el usuario podrá bajar la jerarquía o introducir nuevas dimensiones.
- **Slice and dice:** esta operación permite selección y proyección.
- **Pívor (rotar):** permite reorientar el cubo.
- **Drill:** utiliza las coordenadas dimensionales especificadas por un usuario para una celda en un cubo para moverse a otro cubo a ver información relacionada.

#### 5.2.3.4.9 Especificar características de aplicaciones para usuarios finales.

Las características que se han tomado en cuenta para el SAG son las siguientes:

##### **Determinar el conjunto inicial de plantillas de reportes.**

Cohervi S.A cuenta con distintos departamento pero como vimos en la fase II, los departamentos involucrados en los reportes y análisis serán: departamento financiero, ventas y producción.

A través de la siguiente tabla definiremos reportes que tendrá la aplicación SAG.

**Tabla V.80 - Definición de plantillas de reportes**

No	Departamento involucrado	Lista de reportes
1	Departamento de ventas	<ul style="list-style-type: none"><li>- Análisis de venta de los productos en los últimos años</li><li>- Análisis de ventas de contado y a crédito</li><li>- Análisis de ventas de acuerdo a las zonas</li><li>- Análisis de ventas según la zona de cada vendedor</li><li>- Análisis del desempeño de los vendedores</li><li>- Análisis de venta de productos a los clientes</li><li>- Rotación de los productos de acuerdo a los productos vendidos.</li></ul>
3	Departamento de producción	<ul style="list-style-type: none"><li>- Análisis de compra de mercadería</li><li>- Análisis de stock de mercadería</li><li>- Análisis de productos caducados</li></ul>

##### **Definir la navegación para los reportes.**

Los usuarios podrán navegar en los reportes correspondientes a cada departamento, los mismos que son estáticos y reportes ad hoc

**Detalles de especificación de las plantillas de reportes.**

**Tabla V.81 - Planilla de reporte para Departamento de Compras**

Nombre de la planilla	Compras
Descripción o propósito	Proporcionar información al gerente de compras.
Requerimiento.	Departamento: Compras
Frecuencia de actualización	Semanal
Entradas	n/a

Diseño de la plantilla explorar producción: datos

Explorar

- A. montos de compra
- A. compras en el tiempo
- Cantidad de productos, etc

Datos
Grafico

Drill: jerarquía

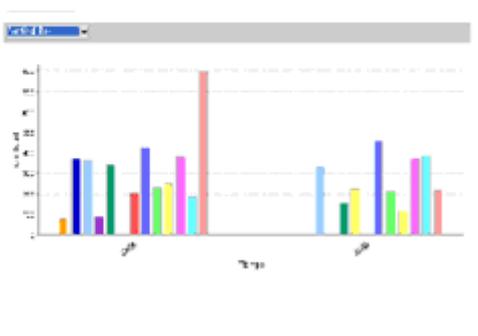
	Medidas
Dimensiones	Dimensiones
Datos	Datos
Datos	Datos

Diseño de la plantilla explorar producción: datos

Explorar

- A. montos de compra
- A. compras en el tiempo
- Cantidad de productos, etc

Datos
Grafico



The chart displays a bar graph with a vertical axis ranging from 0 to 80. The horizontal axis is labeled 'Tiempo' and shows three groups of bars. Each group contains five bars of different colors (blue, green, yellow, red, purple). The second group has a significantly taller red bar, reaching approximately 75 on the scale. Above the chart is a dropdown menu with the text 'Visualización'.

**Tabla V.82 - Plantilla de reporte para Departamento de ventas**

Nombre de la planilla	Ventas
Descripción o propósito	Proporcionar información al usuario sobre la situación del área de ventas
Requerimiento.	R1,R2,R3,R4,R5,R6,R7 del Departamento de ventas
Frecuencia de actualización	Semanal
Entradas	n/a

Diseño de la plantilla explorar ventas: datos

**Explorar**

VENTAS:

- A. Ventas por producto
- A. Ventas de contado y a crédito
- A. Ventas según las zonas.
- A. Ventas según zona de cada vendedor.
- A. Desempeño de los vendedores.
- A. Venta de los productos a los clientes.
- A. Rotación de productos según productos vendidos.

Datos
Grafico

Drill: jerarquía

Medidas	
Dimensiones	Dimensiones
Datos	Datos
Datos	Datos

Diseño de la plantilla explorar ventas: gráfico

**Explorar**

VENTAS:

- A. Ventas por producto
- A. Ventas de contado y a crédito
- A. Ventas según las zonas.
- A. Ventas según zona de cada vendedor.
- A. Desempeño de los vendedores.
- A. Venta de los productos a los clientes.
- A. Rotación de productos según productos vendidos.

Datos
Grafico

**Tabla V.83 - Planilla de reporte para Departamento de Inventario**

Nombre de la planilla	Inventario
Descripción o propósito	Proporcionar información al usuario sobre la situación del área de producción.
Requerimiento.	R1,R2,R3 del Departamento de producción
Frecuencia de actualización	Semanal
Entradas	n/a

Diseño de la plantilla explorar producción: datos

**Explorar**

PRODUCCION:

- A. Compra de mercadería.
- A. Stock de mercadería.
- A. Productos caducados.

**Datos** **Grafico**

Drill: jerarquía

	Medidas
Dimensiones	Dimensiones
Datos	Datos
Datos	Datos

Diseño de la plantilla explorar producción: datos

**Explorar**

PRODUCCION:

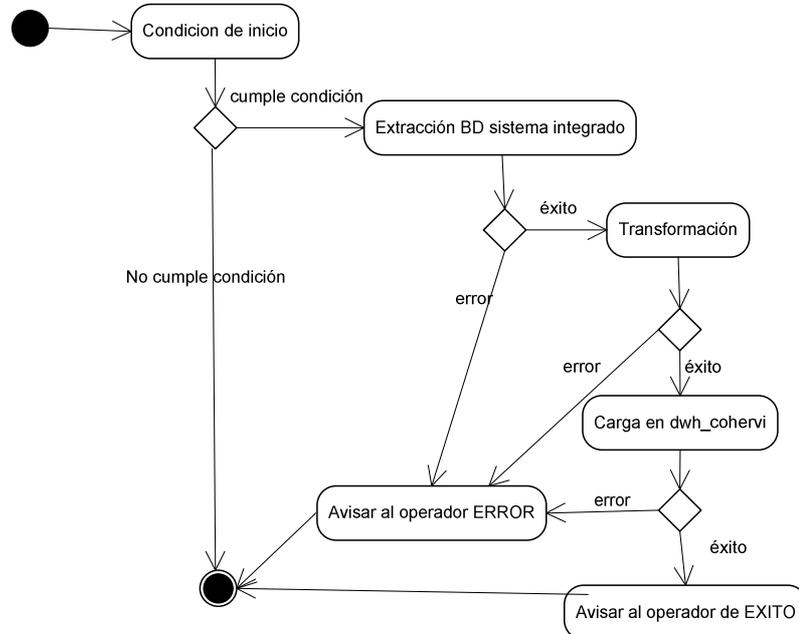
- A. Compra de mercadería.
- A. Stock de mercadería.
- A. Productos caducados.

**Datos** **Grafico**

**5.2.3.4.10 Desarrollar Diagrama Secuencia de los procesos de carga.**

El SAG tiene dos procesos de carga:

1. Extracción, Transformación y Carga (ETL) desde el sistema fuente llamado sistema integrado al destino dwh\_cohervi.



**Figura V.76 - Diseño del ETL desde el sistema fuente a dwh\_cohervi**

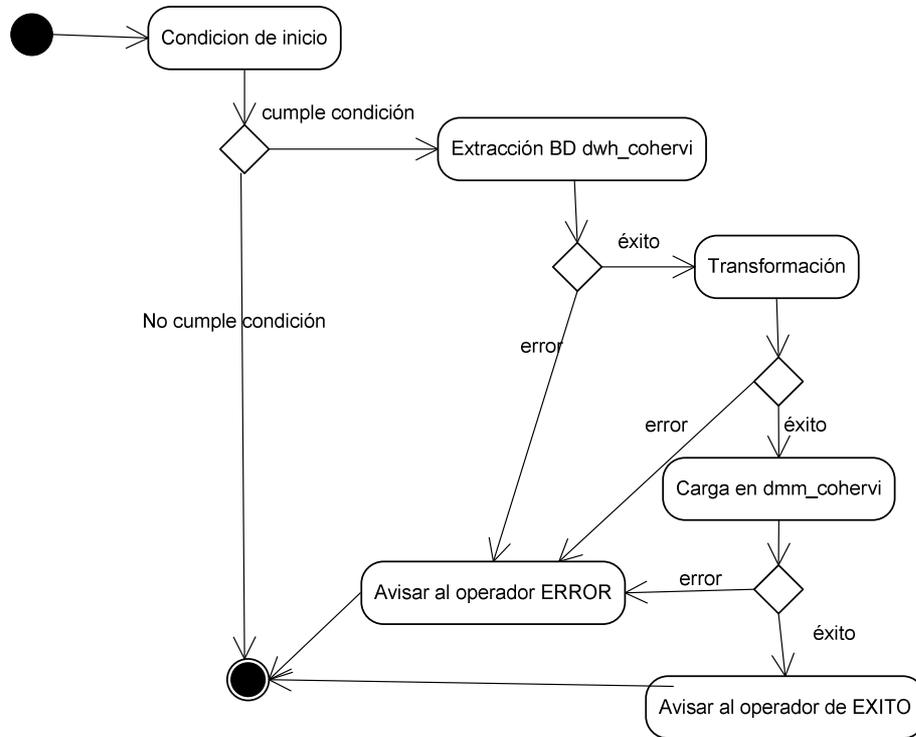
**Tabla V.84 – Especificación de ETL desde Foxprox a BD dwh\_cohervi**

Nombre del proceso ETL	ETL_BdSistemaIntegrado_BDdwh_cohervi	
Fuente	BD sistema Integrado(FoxPro), Excel	
Destino	BD dwh_cohervi	
Transferencias	Características de las transferencias	Transformación del tipo de dato char de las claves primarias de BD fuente a claves primarias integer de la BD destino
		Transferir todos los campos seleccionados
Éxito	Enviar mail al Operador con copia al operador alternativo. Con Texto: "Proceso ETL_DBsistemaIntegrado_DBdwh_cohervi realizado con EXITO".	

**Tabla V.84 – Especificación de ETL desde Foxprox a BD dwh\_cohervi (continuación)**

<p>Fracaso</p>	<p>Enviar mail al Operador con copia al operador alternativo.                  Con Texto:                  “Proceso ETL_DBsistemaIntegrado_DBdwh_cohervi realizado con ERRORES”.                  Setear opción de escribir archivo de LOG ante errores.</p>
<p>Actualización</p>	<p>Semanal</p>
<p>Tipo almacenamiento script ETL (archivos)</p>	<p>Archivo</p>
<p>Observaciones</p>	<p>Todas las tablas se transfieren una vez a la semana ya que la información que necesitan se analizará cada semana.                  Además, el volumen actual de datos no implica un costo en tiempo determinante en la transferencia. De manera que se mantienen los datos del Sistema Integrado en la base Intermedia actualizados con una frecuencia razonable para satisfacer los requerimientos actuales y futuros.</p>

2. Extracción, Transformación y Carga (ETL) desde base de datos fuente dwh\_cohervi al destino ddm\_cohervi.



**Figura V.77 - Diseño del ETL desde dwh\_cohervi a ddm\_cohervi**

**Tabla V.85 - Especificación de ETL desde dwh\_cohervi a ddm\_cohervi**

Nombre del proceso ETL	ETL_Bddwh_cohervi_BDddm_cohervi	
Fuente	BD dwh_cohervi	
Destino	BD ddm_cohervi	
	BD dwh_cohervi	BD ddm_cohervi
Transferencias	dwh_movimientos	Fact_ventas
	dwh_vendedor	dim_vendedor
	dwh_zona	dim_zona
	dwh_compras,dwh_existencia	Fact_inventario
	dwh_cliente	dim_cliente
	dwh_proveedor	dim_proveedor
Característica de la Transferencia	Reemplazar datos existentes Transferir todas las columnas Usar opciones de transferencia	
Éxito	Enviar mail al Operador con copia al operador alternativo. Con Texto: "Proceso ETL_DBdwh_cohervi_DBddm_cohervi realizado con EXITO".	
Fracaso	Enviar mail al Operador con copia al operador alternativo. Con Texto: "Proceso ETL_DBdwh_cohervi_DBddm_cohervi realizado con ERRORES". Setear opción de escribir archivo de LOG ante errores.	
Actualización	Último día de cada semanal	
Tipo almacenamiento script ETL (archivos)	Archivo	
Observaciones		

### 5.2.3.4.11 Realizar Diagrama de componentes.

Este diagrama nos permitirá tener una visión general de la organización física del Sistema de Análisis Gerencial. Muestra la organización de componentes software y dependencias entre ellas.

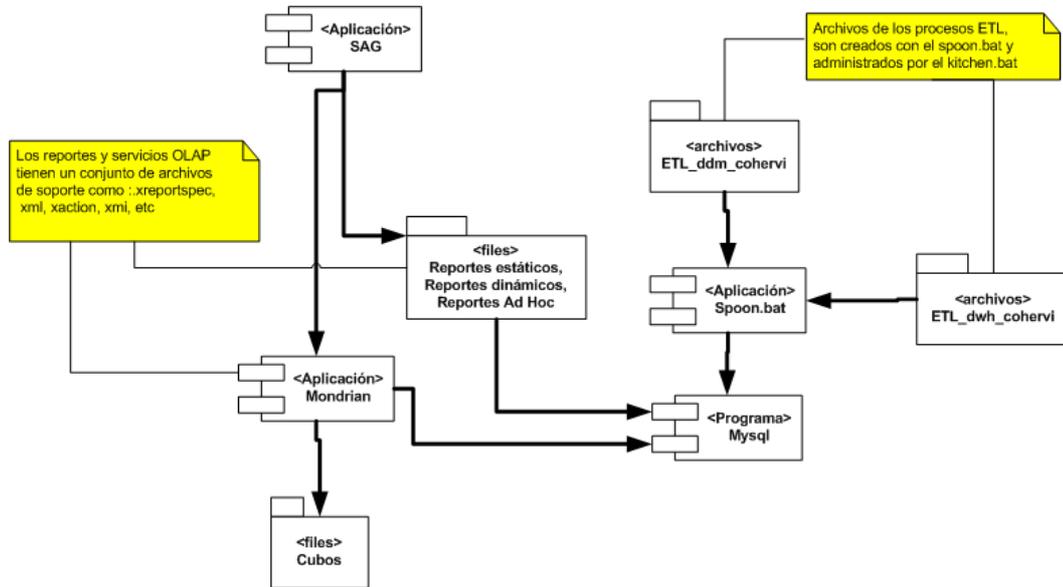


Figura V.78 – Diagrama de componentes

### 5.2.3.4.12 Definir diagrama de nodos.

El SAG, físicamente esta estructurado de la siguiente manera: un servidor donde se encuentra el Sistema Manejador de Base de Datos, el servidor OLAP y el servidor de aplicaciones, a su vez tenemos cuatro PC que se conectarán a la aplicación, estas pcs son del Departamento de ventas, Inventario, compras y sistemas(Administrador).

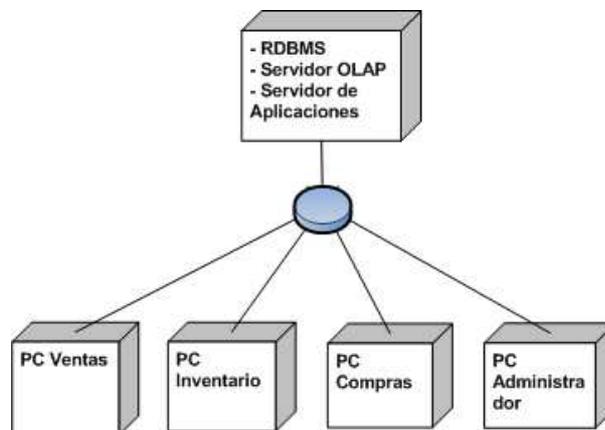


Figura V.79 – Diagramas de nodos

#### 5.2.4 FASE IV: CONSTRUCCIÓN Y DESARROLLO

El Sistema de Análisis Gerencial, está formado por un datawarehouse, que será el repositorio donde se almacenarán los datos limpios, tras realizar el proceso ETL. A continuación tenemos una base de datos llamada ddm\_cohervi donde se crearán los tres datamarts:

- Ventas
- Compras
- Inventario

##### 5.2.4.1 PROPÓSITO:

Usar las técnicas y herramientas definidas en los capítulos anteriores.

##### 5.2.4.2 DESCRIPCIÓN:

Una descripción gráfica de la fase como muestra la siguiente figura.

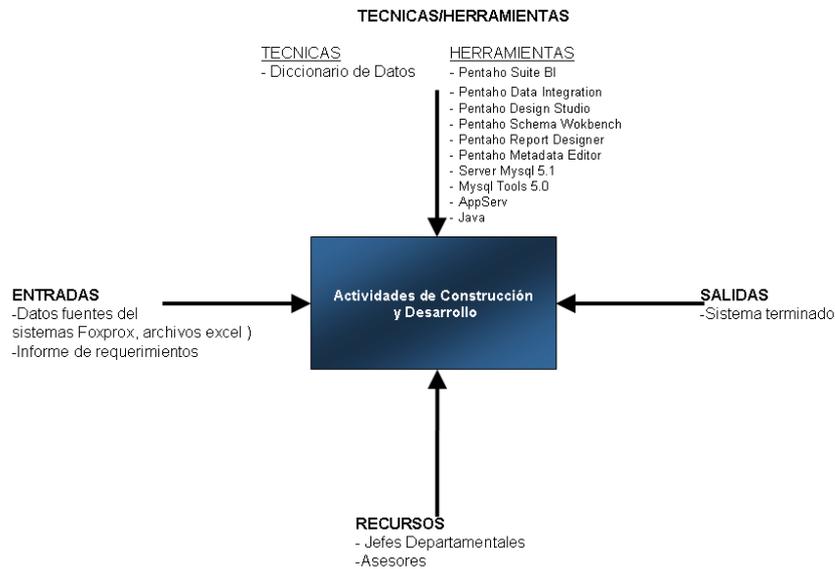


Figura V.80 - Descripción de la fase.

##### 5.2.4.3 RESULTADOS:

- 🚩 Sistema terminado listo para realizar pruebas.

#### 5.2.4.4 ACTIVIDADES:

Actividades a desarrollar a desarrollar son las siguientes:

##### 5.2.4.4.1 Seleccionar herramientas.

Luego de un estudio minucioso de las distintas herramientas para desarrollar proyectos de Business Intelligence, en nuestro proyecto trabajamos con herramientas open source "PENTAHO", ya que Cohervi S.A, no cuenta con el presupuesto para la compra de software comercial.

El siguiente cuadro hace un resumen de las áreas evaluadas y la selección de herramientas para la ejecución de nuestro proyecto:

**Tabla V.86 - Herramientas seleccionadas**

AREA	DESCRIPCIÓN
Plataforma de hardware	Windows XP Service Pack II
Plataforma Business Intelligence	PENTAHO 3.0
ETL	Pentaho Data Integrator (Kettle)
DBMS	Mysql Server 5.1 y las Tools
OLAP	Servidor OLAP Mondrian
Herramientas de acceso a los datos	Web de Pentaho

**5.2.4.4.2 Especificar permisos y accesos de los usuarios.**

La siguiente tabla nos indica los permisos que son lectura y escritura sobre las tablas o bases de datos.

**Tabla V.87 - Permiso de base de Datos**

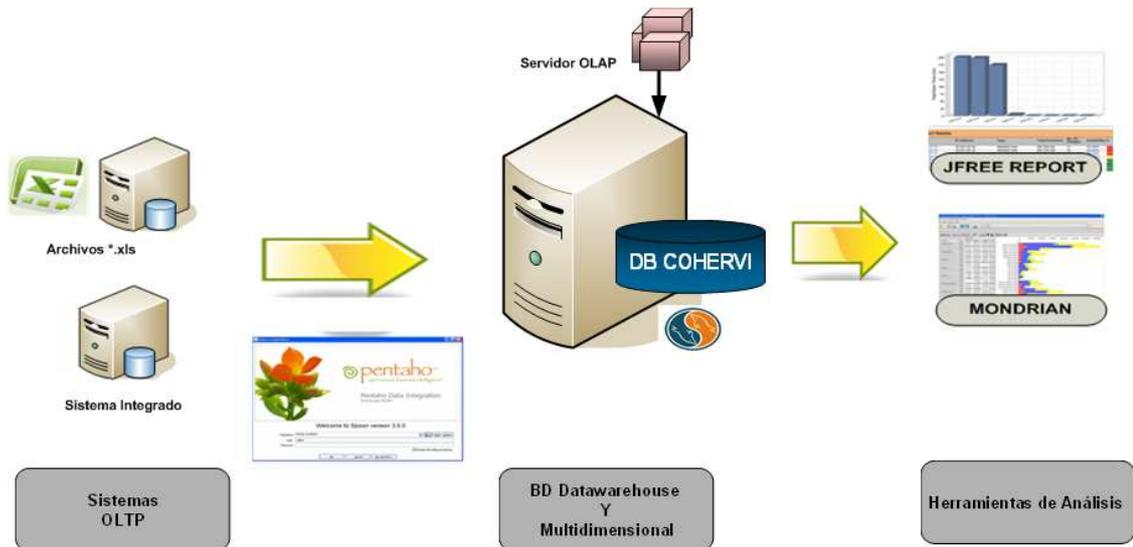
Tipo de usuarios		Permisos		
		Bases de Datos		
		BD Sistema Integrado, Excel	BD Intermedia	BD Datamart
Tomador de Decisiones	Departamento de Ventas	N/A	N/A	Lectura a todas las tablas
	Departamento de Compras	N/A	N/A	Lectura a todas las tablas
	Departamento de inventario	N/A	Lectura a todos los registros	Lectura a todas las tablas
Departamento de Computo	Administración Datawarehouse	N/A	Acceso Completo de Administración del Archivo	Acceso completo de Administración de BD.
	Administración de ETL	Lectura a todas las tablas	Lectura/Escritura a a todos los registros	Lectura/Escritura en todas las tablas
	Administración de de la plataforma Pentaho	N/A	Lectura a todos los registros	Lectura/Escritura en todas las tablas.

La tabla V-88. Indica los accesos de los tipos de usuarios sobre las distintas herramientas tales como: DBMS, ETL, aplicaciones OLAP e interfaz de usuarios.

**Tabla V.88 - Acceso a las herramientas**

Tipo de usuario	Permisos			
	Herramientas			
	DBMS (Mysql)	DTS (Spoon)	Servicios OLAP (Mondrian)	Interfaz de Usuario Pentaho (User Console Login)
Jefe de Ventas	NO	NO	NO	SI
Jefe de Inventario o Bodega	NO	NO	NO	SI
Jefe de Compras	NO	NO	NO	SI
Asistente y Jefe de Sistemas	SI	SI	SI	SI

**5.2.4.4.3 Desarrollar el Diseño.**



**Figura V.81 – Arquitectura y Herramientas**

El proyecto será desarrollado en la plataforma Windows XP Service Pack II y herramientas de Business Intelligence PENTAHO.

Para iniciar el proceso de desarrollo tomar en cuenta los siguientes pasos:

1. Descargar las herramientas tales como:

JAVA

 [jdk-6u10-windows-i586-p](#)

PENTAHO

 [pentaho-bi-suite-3.0-windows-installer](#)

 [pdi-ce-3.2.0-stable.zip](#)

 [psw-ce-3.0.4.stable.zip](#)

MYSQL

 [mysql-connector-java-5.1.6.tar](#)

 [mysql-5.1.29-rc-win32](#)

 [mysql-gui-tools-5.0-r10-win32](#)

 [appserv-win32-2.5.9](#)

2. Pre requisito para la instalación de la Suite de Pentaho y sus las herramientas será necesario instalar java (jdk-6u10-windows-i586-p)

#### **5.2.4.4.3.1 Construir el Datawarehouse en Mysql**

Para construir el Datawarehouse usamos el Sistema Manejador de Base de Datos MYSQL, por ser el gestor más usado en el mundo del software libre, por su diseño multihilo que le permite soportar una gran carga de forma muy eficiente.

En la actualidad mysql tiene paquete completo de Tools que permiten administrar esta base de datos, pero para el presente proyecto también usaremos AppServ que es un administrador de base de datos web por facilidad del desarrollador.

### **Server Mysql 5.1, Mysql Tools 5.0 y AppServ**

#### **Instalación**

La instalación no se indicará paso a paso, debido a que es muy sencilla con solo dar siguiente, siguiente esta base de datos está instalada.

Las Tools de Mysql y AppServ, se instala de la misma manera que el Server Mysql.

## Ejecución

Para ejecutar Mysql ir a Inicio/Archivos de Programa/Mysql.

## Creación del Datawarehouse y Datamart usando AppServ.

Primero debemos crear las bases de datos (dwh\_cohervi – ddm\_cohervi), luego crear cada una de las tablas basándonos en el diseño de la fase anterior, para un mayor detalle de cada una de las tablas de las dos bases de datos ver Anexos. Diccionario de datos.

### 5.2.4.4.3.2 Instalación y configuración de la Suite de Pentaho.

## Pentaho

Pentaho se define a sí mismo como una plataforma de BI “orientada a la solución” y “centrada en procesos” que incluye todos los principales componentes requeridos para implementar soluciones basados en procesos y ha sido concebido desde el principio para estar basada en procesos. Véase la siguiente figura.

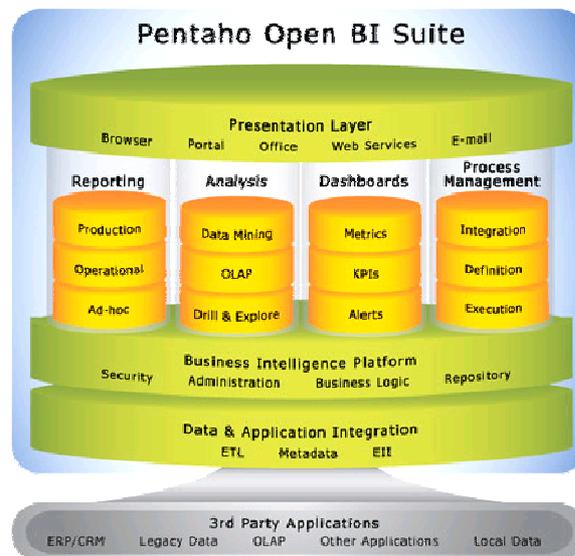


Figura V.82 – Arquitectura Pentaho Open BI Suite

Las herramientas con las que cuenta Pentaho son las siguientes:

-  Data Integration(Spoon)
-  Report designer:
-  Schema workbench
-  Design studio
-  Metadata editor
-  Weka

## Instalación

La herramienta no necesita instalación, porque viene en una carpeta comprimida, se descomprime en cualquier ruta en nuestro caso lo vamos hacer en: D:\COHERVI\pentaho

## Configuración

En la configuración se debe realizar lo siguiente:

### Acceso remoto

Se debe cambiar el nombre de **localhost** que viene por defecto, al nombre o IP de la máquina donde está la herramienta de Pentaho, en los siguientes archivos:

- Web.xml ubicado en la ruta:\biserver-ce\tomcat\webapps\pentaho\WEB-INF.
- Welcome.html en la ruta:\biserver-ce\tomcat\webapps\pentaho\welcome

### Password de Publicación

Esta configuración es necesaria para la publicación de los reportes, cubos, metadata y otros archivos, creados con las distintas herramientas de pentaho.

El nombre del archivo es Publisher\_config.xml ubicado en la ruta: \pentaho\biserver-ce\pentaho-solutions\system, en este archivo se registrará el password de publicación, quedando de la siguiente manera:

```
<publisher-password>admin</publisher-password>
```

## Ejecución

Para ejecutar la Suite de Pentaho abrir la consola de Windows y acceder a la ruta que indica la siguiente figura y ejecutar el archivo **start-pentaho.bat**



Figura V.83 – Ejecutar el archivo start-pentaho.bat

La siguiente pantalla nos indica que el servicio ha sido iniciado satisfactoriamente

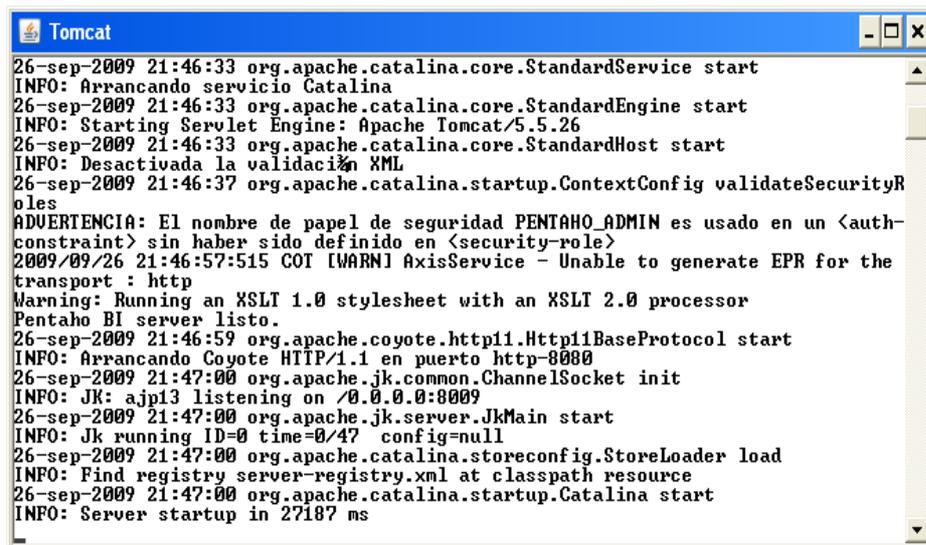


Figura V.84 – Pantalla de servicios levantados de la Suite de Pentaho

Abrir un navegador de internet (Firefox Mozilla, internet Explorer y otros), poner la siguiente URL <http://zeus:8080/pentaho/>, donde se desplegará la siguiente pantalla.



**Figura V.85 – Ventana inicial de la Suite de Pentaho**

A continuación levantar la consola de administración, ir a la siguiente ruta: `\pentaho\administration-console` y ejecutar el archivo `startup.bat`.



**Figura V.86 - Ejecutar el archivo startup.bat**

La siguiente pantalla nos indica que el servicio ha sido iniciado correctamente.



**Figura V.87 – Pantalla de servicios levantados de la Consola de Administrador**

Abrir un navegador de internet (Firefox Mozilla, internet Explorer y otros), poner la siguiente URL <http://zeus:8099/>, donde se desplegará la siguiente pantalla.

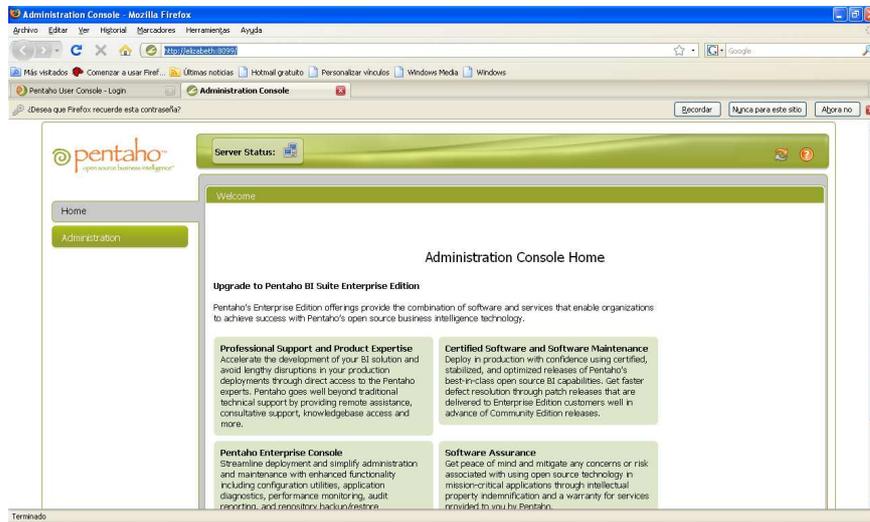


Figura V.88 – Ventana inicial de la Consola de Administrador

## Configuración adicional

### + Creación de Datasources

La creación del Datasources se debe realizar en la consola de administración, en donde se escogerá la opción **Administration**, seleccionar la opción **Datasources** y elegir el signo mas(+) para añadir un nuevo datasources, como indica en la siguiente figura.

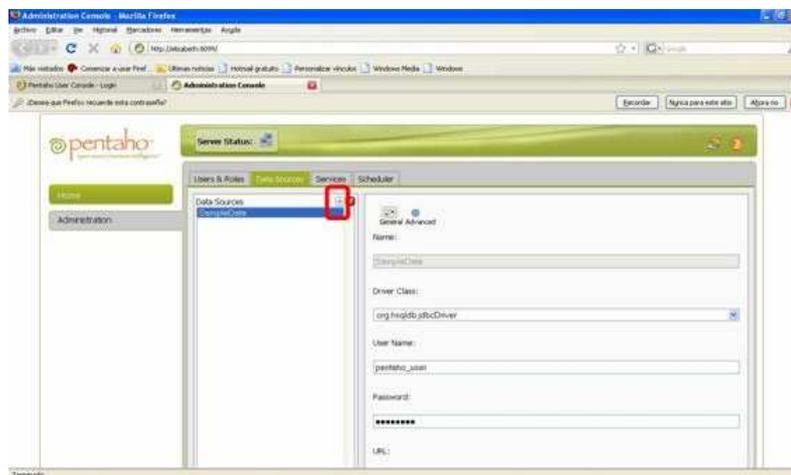
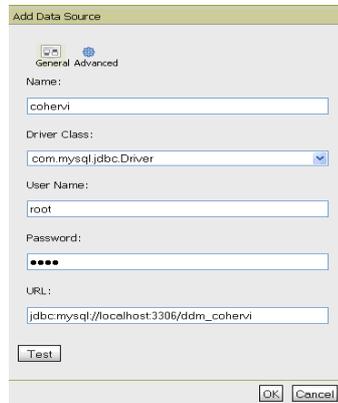


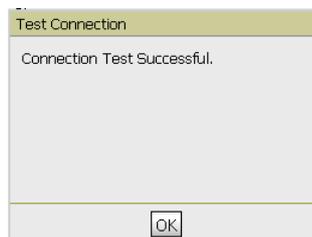
Figura V.89 – Crear un DataSource

En esta pantalla se ingresarán los parámetros de conexión a nuestra base de datos tales como nombre del datasources, el driver, el usuario y la URL de la base de datos.



**Figura V.90 - Parámetros de conexión del DataSource**

A continuación escoger la opción Test, donde se desplegará la siguiente pantalla indicando que la conexión ha sido satisfactoria.



**Figura V.91 - Resultado de Conexión**

### Creación del repositorio de soluciones

El repositorio se llamará **sistema**, es una carpeta que contendrá todos los archivos de Business Intelligence como reportes estáticos, dinámicos, Ad-Hoc, vistas OLAP y otros.

Se debe ir al repositorio general de soluciones de la herramienta que se encuentra en la siguiente ruta: `\pentaho\biserver-ce\pentaho-solutions`, crear una carpeta llamada sistema que debe contener el archivo `index.xml`, con el siguiente contenido:

```
<?xml version='1.0' encoding='UTF-8'?>
<index>
  <name>%name</name>
  <description>%description</description>
  <icon>solutions.png</icon>
  <visible>>true</visible>
  <display-type>icons</display-type>
</index>
```

### **Personalizar la interfaz**

Para personalizar la interfaz se debe modificar los archivos:

PUCLogin.jsp y Home.jsp, ubicado en la ruta: \pentaho\biserver-  
ce\tomcat\webapps\pentaho\jsp

#### **5.2.4.4.3 Realizar procesos de carga usando Pentaho Data Integration**

##### **Pentaho Data Integration**

Es un proyecto que incluye un conjunto de herramientas para realizar ETL. Uno de sus objetivos es que el proyecto ETL sea fácil de generar, mantener y desplegar.

Se compone de 4 herramientas:

-  SPOON: permite diseñar de forma gráfica la transformación ETL.
-  PAN: ejecuta las transformaciones diseñadas con SPOON.
-  CARTE: ejecuta transformaciones y trabajos de forma remota utilizando un servlet pequeño, al servidor que ejecuta Carte se denomina servidor esclavo.
-  KITCHEN permite ejecutar los trabajos batch diseñados con Spoon.

Los archivos generados por el sistema Integrado Foxprox son ficheros de base de datos (DBFS) y archivos externos de Excel estos archivos son extraídos, transformados y cargados al datawarehouse por medio de la herramienta Pentaho Data Integration.

En la figura V.92 se muestra el diseño de las ETL

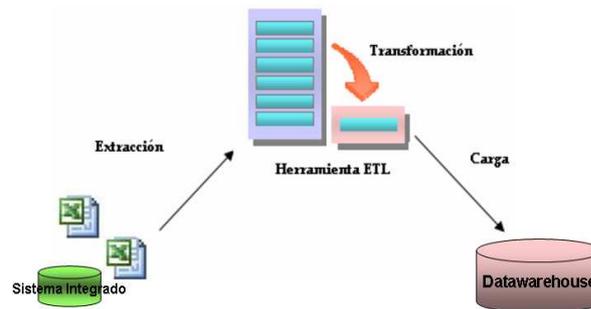


Figura V.92 - Diseño de ETL.

### Instalación:

La herramienta ETL Pentaho Data Integration seleccionada es la versión 3.2.0, esta herramienta viene en una carpeta comprimida, se descomprime en cualquier ruta en nuestro caso lo vamos hacer en la siguiente dirección: D:\COHERVI\Pentaho Data Integration.

### Ejecución:

Para ejecutar ingresar a la consola de Windows y ejecutar el archivo spoon.bat como se indica en la siguiente pantalla.



Figura V.93 – Ejecutar el archivo Spoon.bat

Posteriormente nos indica una ventana como muestra la siguiente figura.



Figura V.94 - Ventana inicial de la herramienta ETL

Para ingresar a la ventana de trabajo de la herramienta se debe seleccionar la opción **Sin Catálogo** como se muestra en la Figura v.95

A continuación, tenemos esta pantalla donde podremos crear conexiones a las distintas bases de datos, realizar transformaciones, trabajos, hacer uso de los distintos componentes que ofrece Spoon, realizar tareas programadas, entre otras.



Figura V.95 - Área de trabajo de la herramienta ETL

## **- Generar conexiones a las diversas bases de datos**

Al crear una transformación o trabajo en Pentaho Data Integration generalmente se establecen conexiones a las diversas fuentes y destino, a continuación se determinan los pasos que se deben seguir para realizarlos.

### **Fuentes y Destinos de datos soportados**

Al realizar la población de un destino un punto central es el acceso a varias fuentes posibles, es así, que ésta herramienta posee una compatibilidad amplia con fuentes y destinos, los mismos que son: Mysql, SQL Server, Microsoft Access, Oracle, SyBase, Informix, IBM Db2, PostgreSQL, Sqlite, Hypersonic, DBase II,IV or 5, Gupta SQL Base, entre otros.

### **Tipos de conexión utilizados**

Al momento de establecer conexiones a fuentes y destinos la herramienta permite acceder a través de Native (JDBC), ODBC y JNDI, en el sistema usamos una conexión ODBC para acceder a los archivos fuentes y la conexión JDBC para acceder data Warehouse y al Datamarts.

### **Crear conexión**

Se debe crear un ODBC para los archivos fuentes, para acceder a una conexión crear una transformación.

Posteriormente, Se selecciona la opción **Conexiones a Base de datos**, elegir **Nuevo**, se obtendrá una pantalla donde debe ingresar los datos de la base de datos fuente tales como:

- Nombre de la conexión, IP de la maquina, Nombre de la base de datos, Puerto, Usuario, Password.

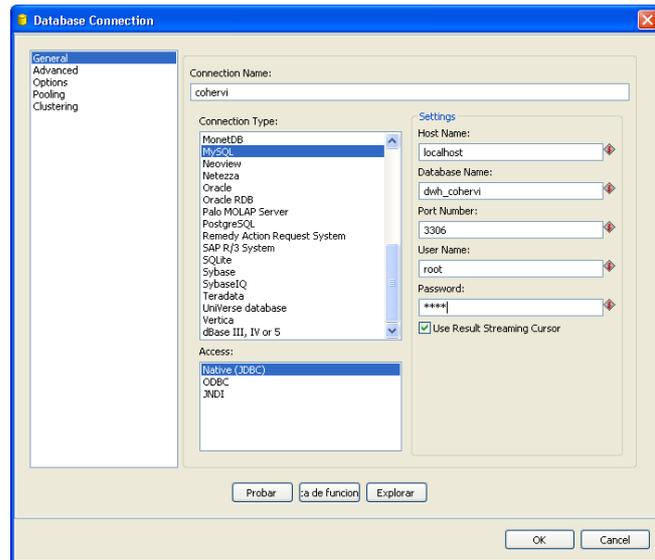


Figura V.96 - Parámetros de conexión Pentaho Data Integration

### - Generar transformaciones para poblamiento individual de información

La figura V.97 muestra el Job para la carga de la base de datos dwh\_cohervi

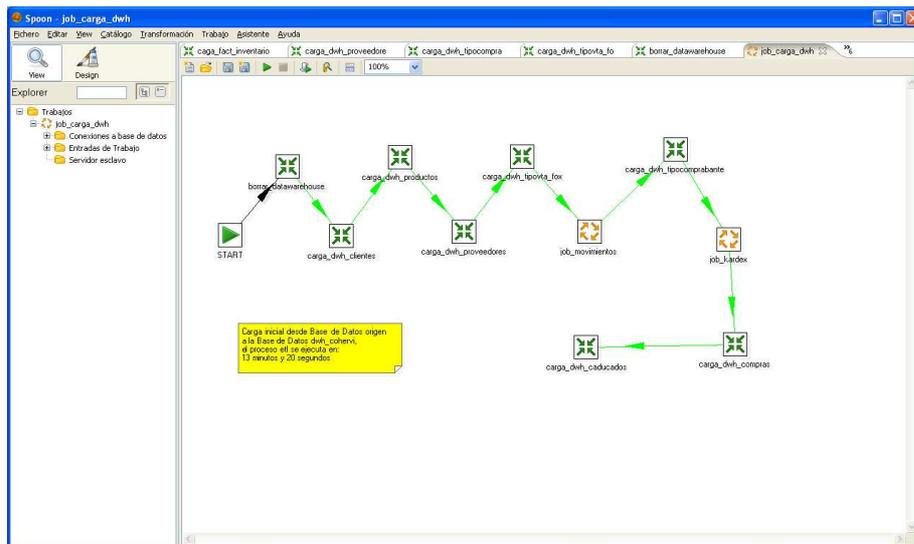


Figura V.97 - Poblamiento de la base de datos dwh\_cohervi

Una vez que tenemos poblada la base de datos dwh\_cohervi, se realiza el proceso de poblamiento de la base de datos ddm\_cohervi, cuyo Job es el siguiente.

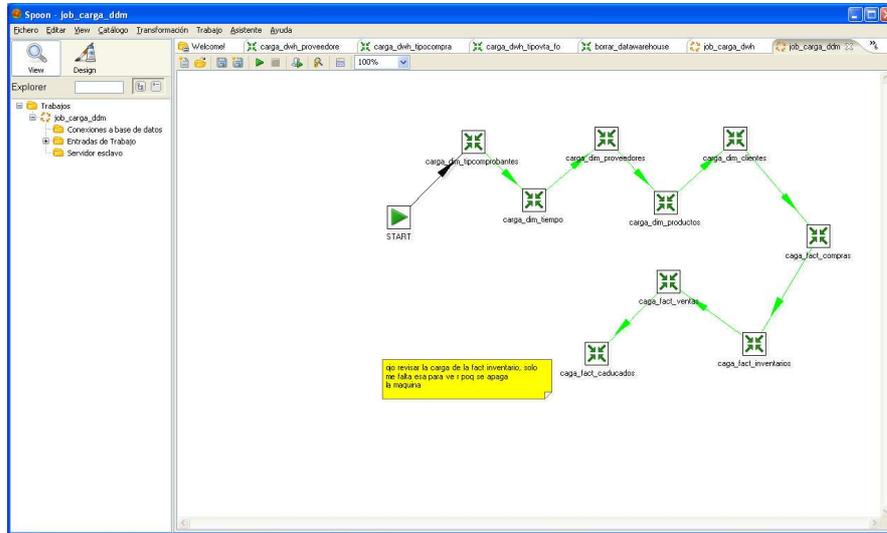


Figura V.98 - Poblamiento de la base de datos ddm\_cohervi

Finalmente tenemos el Job que ejecuta tanto el poblamiento de la base de datos dwh\_cohervi y ddm\_cohervi, cuya carga es programada semanalmente a las 6:00 PM

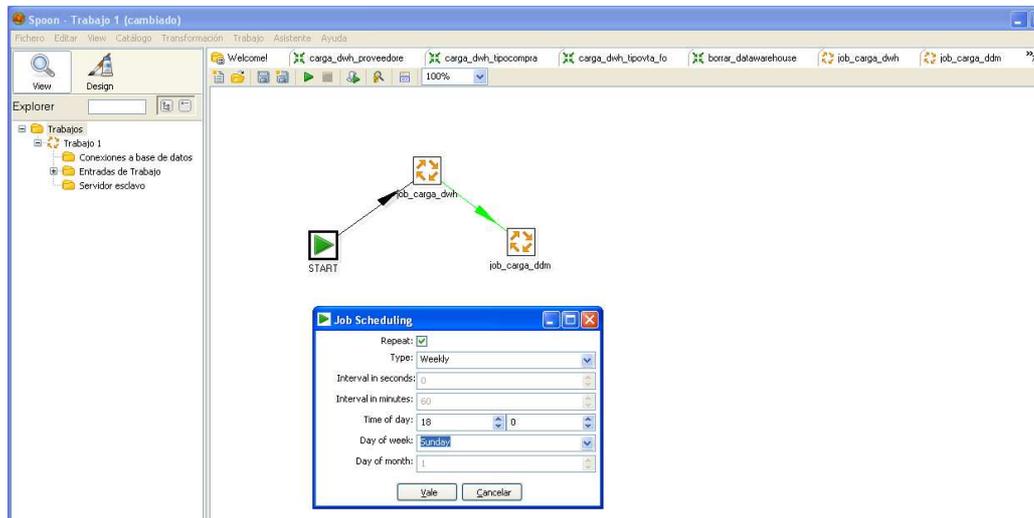


Figura V.99 - Carga periódica

#### 5.2.4.4.3.4 Desarrollar cubos multidimensionales con Schema Workbench y publicarlos usando Pentaho Design Studio.

##### SCHEMA WOKBENCH:

Es una herramienta que permite crear cubos para realizar análisis de información con el motor OLAP Mondrian, trabaja conjuntamente con Jpivot y utiliza el lenguaje de MDX para editar las consultas.

##### Instalación

Esta herramienta viene en una carpeta comprimida, se descomprime en cualquier ruta en nuestro caso lo vamos hacer en la siguiente dirección: D:\COHERVI\SAG\pentaho\Schema Workbench.

##### Ejecución

Para ejecutar Schema Workbench ir a la dirección donde se descomprimió la herramienta, abrir la consola de windows y ejecutar el archivo **workbench.bat**.



Figura V.100 - Ejecutar archivo schema-worbench.bat

Tras ejecutar el comando se desplegara la pantalla inicial del Schema Workbench.

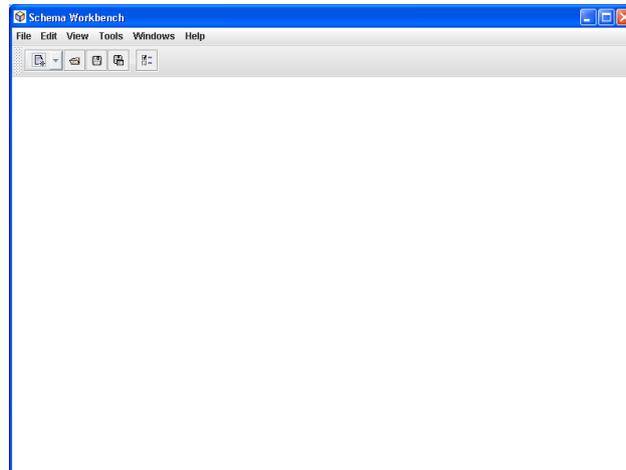


Figura V.101 - Pantalla inicial del schema workbench.

### Generar la conexión a la base de datos

Ir a la barra de menú **Tools**, escoger la opción **Preferences** y llenar los datos tales como driver de la base de datos, la URL de conexión, el usuario y password de la base de datos.

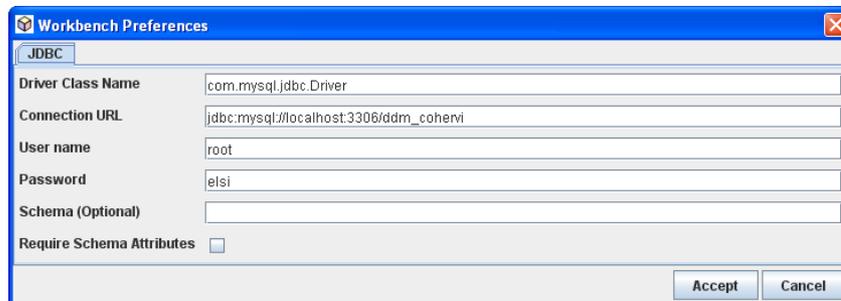


Figura V.102 - Parámetros de conexión

### Generar cubos

El Sistema de análisis Gerencial cuenta con tres cubos:

1. Cubo de Compras
2. Cubo de Inventario
3. Cubo de Ventas

A continuación se describirá la forma en la que se debe crear los cubos.

Crear el esquema, para ello ir a la barra de menú escoger **File**, luego **New** y finalmente **Schema**, asignamos un nombre para la aplicación nuestro esquema se llamará COHERVI, además en esta pantalla tenemos una propia barra de herramientas, donde se podrá añadir un cubo, dimensiones, medidas, roles, etc.

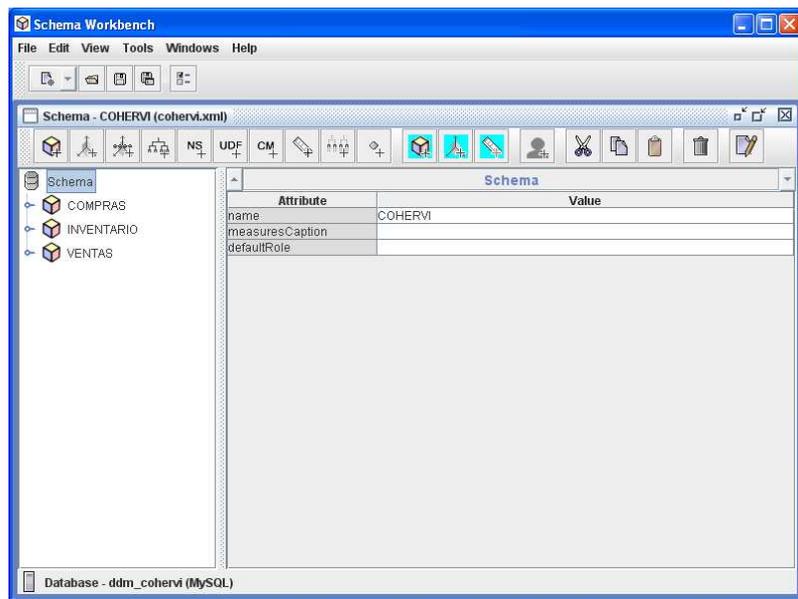


Figura V.103 – Área para diseñar cubos

Seleccionar el componente **CUBO** de la barra de herramientas del esquema, asignar un nombre, posteriormente añadir el componente dimensión, luego las medidas, quedando el cubo Compras, Inventario y Ventas de la siguiente manera:

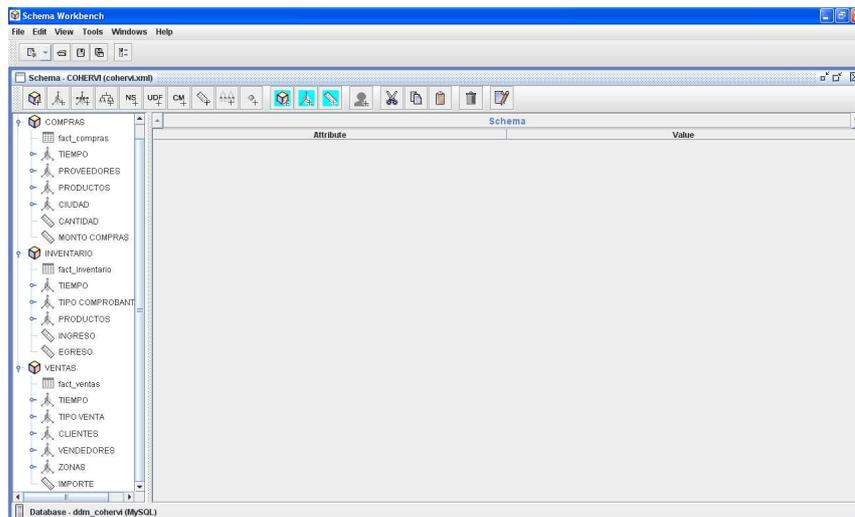


Figura V.104 - Despliegue del esquema desarrollado

Para realizar la publicación a la Suite de Pentaho, ir a la barra de herramientas del esquema seleccionar **File**, escoger la opción **Publish**, donde se obtendrá la siguiente pantalla, en la que se ingresará la URL, el password de publicación (admin), el usuario y password (password). A continuación elegir la opción **OK**.



Figura V.105 - Parámetros de Publicación

En esta pantalla ingresar la ruta del repositorio donde se almacenará el esquema, en nuestra aplicación lo guardaremos en la siguiente ruta: /sistema/Ventas/Vistas OLAP, escoger la opción **Publish**.

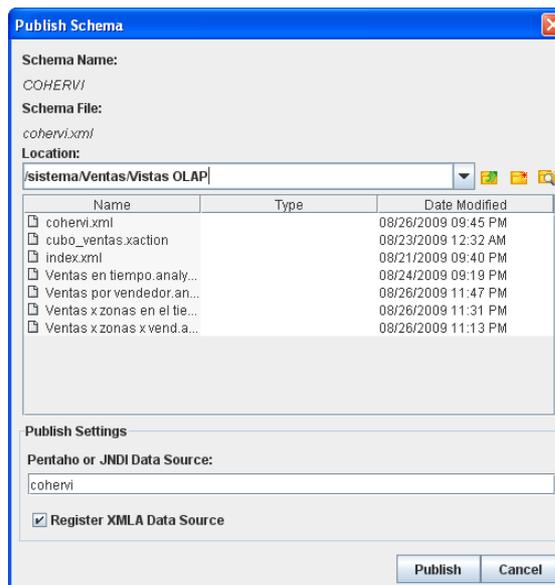


Figura V.106 - Repositorio de publicación

## DESIGN STUDIO

Es una colección de editores, vistas y módulos de administración integrada en una sola aplicación que proporcionan un entorno gráfico para editar cubos y reportes realizados con Schema Workbench, BIRT y otros.

### Instalación

No necesita instalación, se descomprime el archivo zip dentro de la carpeta pentaho

### Ejecución

Abrir la consola de Windows y ejecutar el archivo startup.jar, como indica en la figura



Figura V.107 - Ejecutar el archivo startup.bat

Luego de ejecutar el archivo startup.jar tenemos la ventana principal de Design Studio

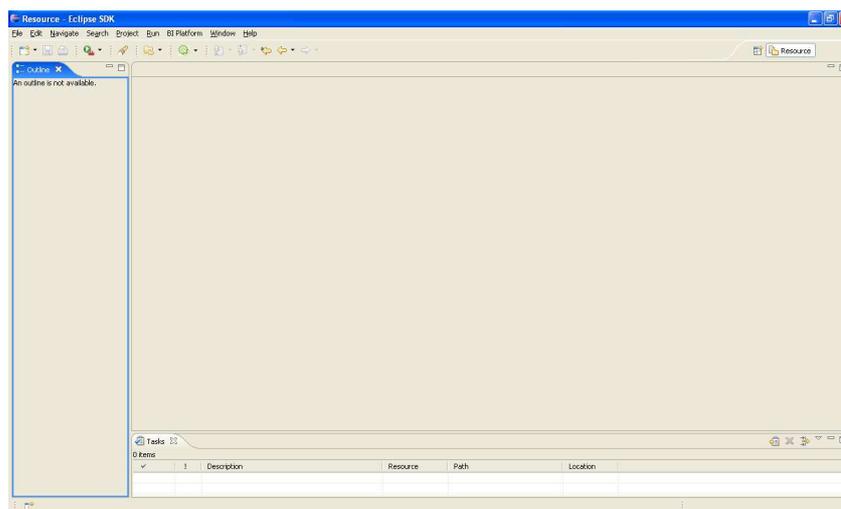


Figura V.108 - Ventana principal del Design Studio

## Generar archivos xaction

En la venta principal de Design Studio, en la barra de menú escoger la opción **File**, seleccionar **New**, elegir **Project** y seguir los pasos del Wizard, con el proyecto creado, ir a la barra de menú escoger la opción **BI Plataforma** y elegir **New Xaction Secuence** y llenar los datos como nombre del archivo xaction y escoger el template.

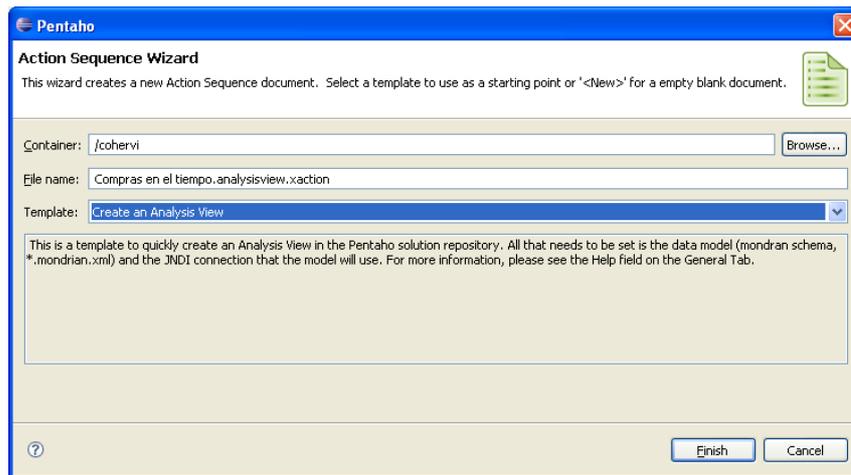


Figura V.109 - Pantalla de Creación Design Studio

En esta ventana se debe definir el repositorio donde se encuentra el cubo, nombre del data sources, nombre del rol (creado en el esquema workbench) y llenar la sentencia MDX en la parte que dice Query.

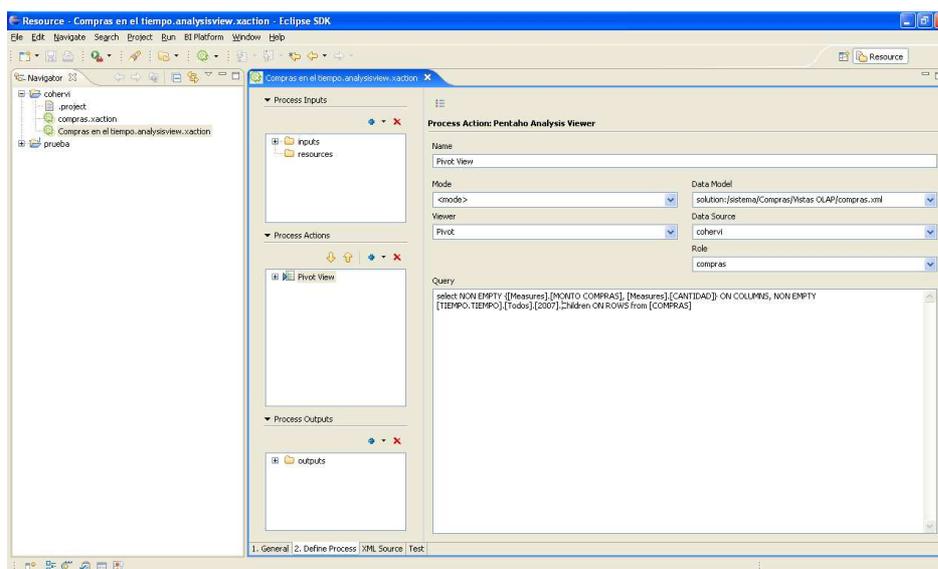


Figura V.110 - Pantalla de Diseño del Design Studio

Finalmente copiar el archivo a nuestro repositorio ubicado en la siguiente ruta.  
\\pentaho\biserver-ce\pentaho-solutions\sistema\Compras\Vistas OLAP.

Se repite este procedimiento para cada uno de los cubos creados en esta tesis.

#### **5.2.4.4.3.5 Elaborar Reportes Estáticos – Ad Hoc usando Pentaho Report Designer y Metadata Editor**

##### **Report Designer**

Pentaho Report Designer es un front-end que utiliza JFreeReport, para generar reportes utilizando la técnica de arrastrar y soltar. Los reportes terminados pueden ser exportados como archivos de PDF, rtf, html y otros.

##### **Intalación**

La herramienta Pentaho Design no realizamos intalación, se descomprime el arhivo zip dentro de la ruta: \\pentaho

##### **Ejecución**

Para ejecutar abrir la consola de windows y ejecutar el archivo ReportDesigner.exe



**Figura V.111 - Pantalla para ejecutar el archivo ReportDesigner.exe**

Luego de ejecutar el archivo tenemos la pantalla inicial del Report Designer.

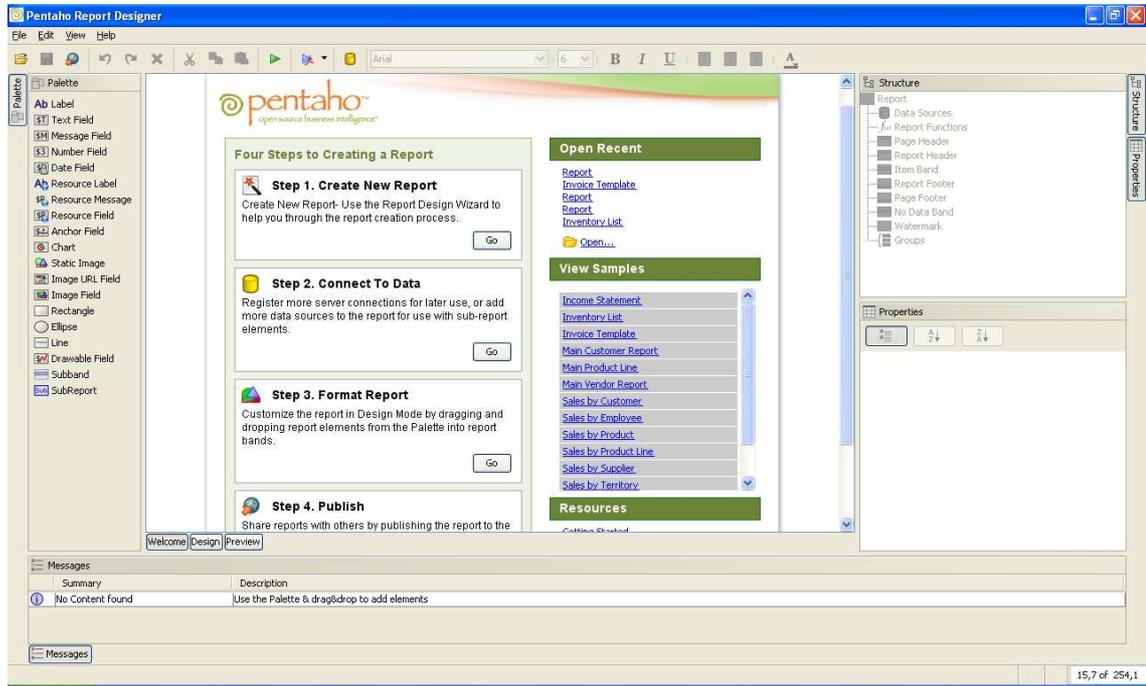


Figura V.112 - Pantalla inicial de Report Designer

### Generar la conexión

Para generar la conexión se debe crear un nuevo reporte, donde deberá ir a **File**, luego escoger **New** y en la parte derecha seleccionar **Add Data Source**, donde aparecerá la siguiente pantalla en la que debe crear una nueva conexión eligiendo el botón **Add**, En esta pantalla ingresar los parámetros de conexión tales como nombre del JNDI, el driver, el string de conexión, usuario y password del datawarehouse, finalmente escoger el botón **Test**.

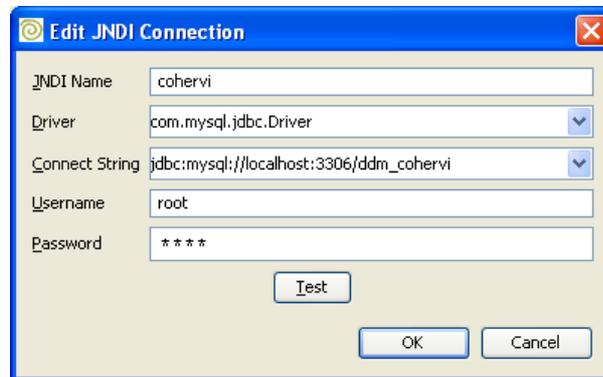


Figura V.113 - Parámetros de conexión

## Desarrollar reportes

Seleccionar el JNDI, luego en la parte derecha en **Query Details**, registrar el SQL que desea cargar en el reporte quedando de la siguiente manera:

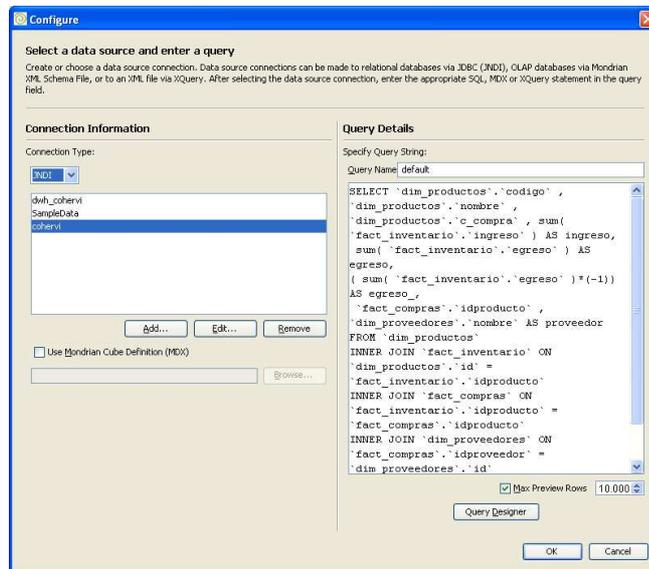


Figura V.114 - Selección de Datasource y query

En esta pantalla en la parte de **Design** usar los componentes de la parte izquierda y crear el reporte, donde el diseño de uno de los reportes será el que muestra en la figura, a continuación seleccionar **Preview**, para visualizar el reporte

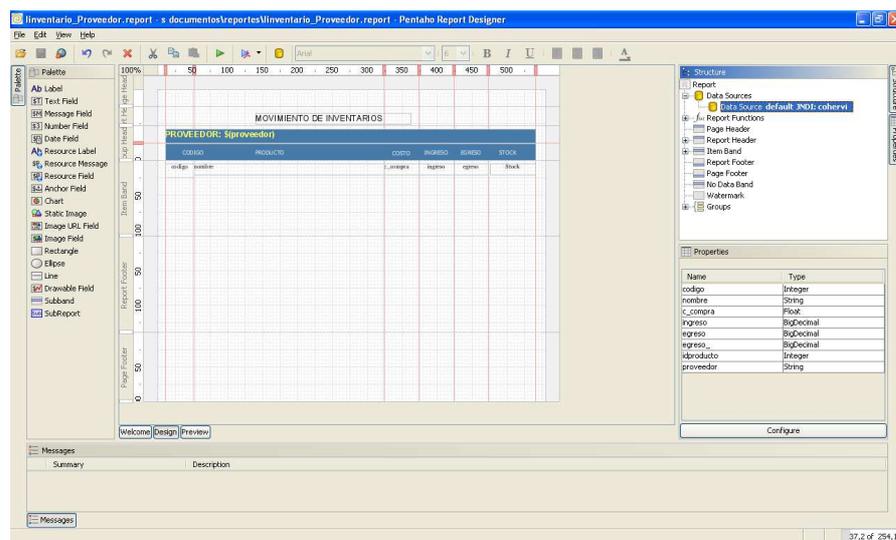


Figura V.115 - Ventana de Diseño del reporte

Una vez terminado el reporte se publicará, ir a la barra de herramientas y escoger **Publish** to Server, donde se ingresará la URL, el password de publicación, usuario y contraseña, finalmente seleccionar el botón **OK**



Figura V.116 - Parámetros de publicación

En esta pantalla se debe registrar datos del reporte y el repositorio de Pentaho, seguidamente pulsar el botón **Publish**.

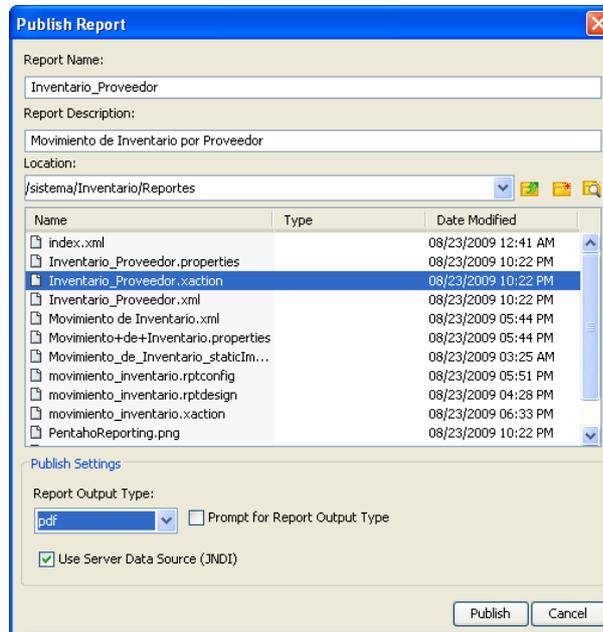


Figura V.117 - Repositorio de publicación

## METADATA EDITOR

Es una herramienta que permite crear una capa lógica intermedia para configurar los modelos de metadatos que servirán para realizar los reportes Ad Hoc.

### Instalación

La herramienta Metadata Editor no realizamos instalación, se descomprime el archivo zip dentro de la ruta: \pentaho

### Ejecución

Abrir la consola de Windows y ejecutar el archivo MetadaEditor.bat



Figura V.118 - Ejecutar la herramienta MetadataEditor

Luego de ejecutar el archivo bat tenemos la siguiente pantalla

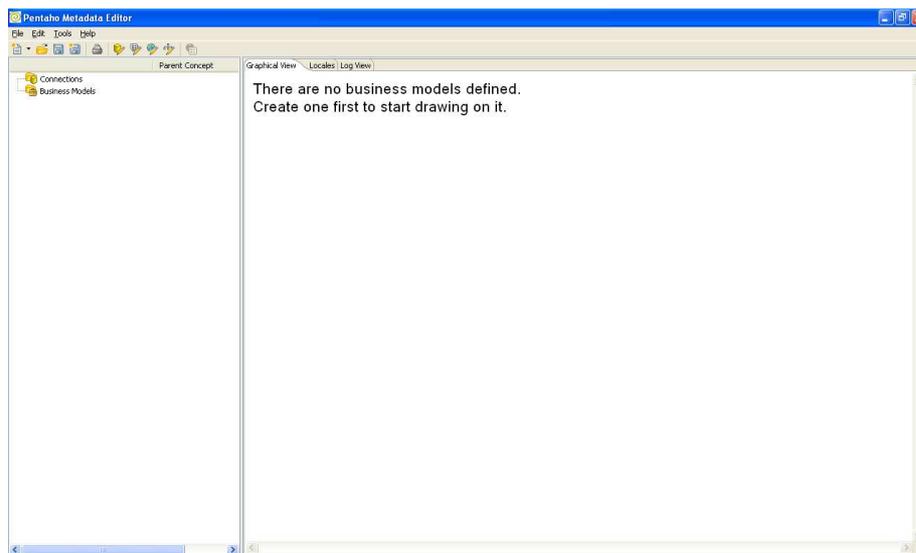


Figura V.119 - Ventana inicial MetadataEditor

## Generar la conexión

En la pantalla inicial seleccionar **File**, luego **New** y finalmente **New Connection**. En esta pantalla ingresar el nombre de la conexión y el JNDI, seleccionar **Probar**

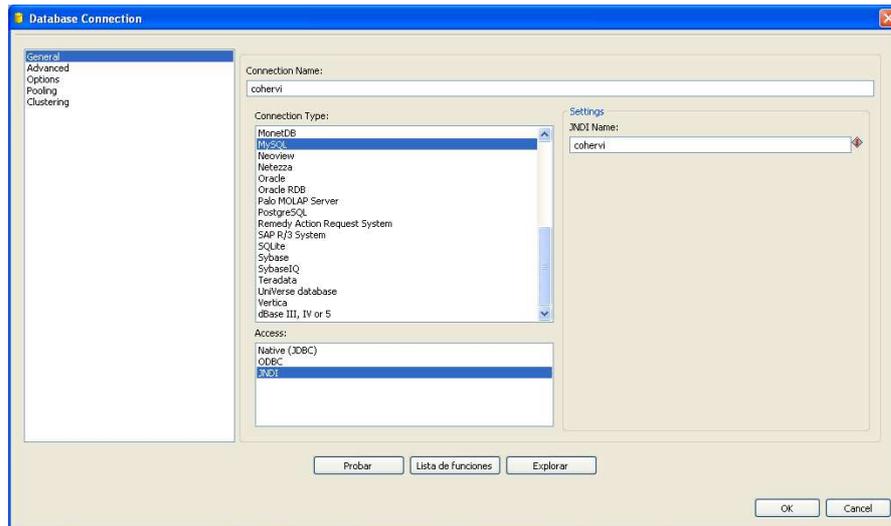


Figura V.120 - Parámetros de conexión Metadata Editor.

## Desarrollar un modelo de negocios

Seleccionar **New Business Model**, asignar un nombre y escoger el botón **OK**.

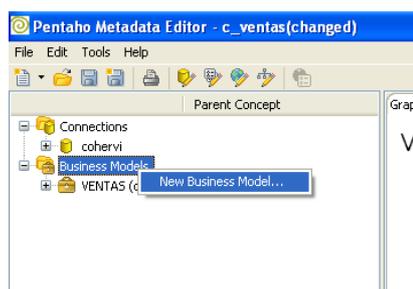


Figura V.121 - Creación de Nuevo modelo de Negocios

A continuación seleccionar **Business Tables**, donde se obtendrá la siguiente pantalla, en donde se elegirá las tablas que deseamos añadir al modelo de negocios.



Figura V.122 - Selección de tablas físicas

Seleccionar la opción **Relationships**, donde se definirá los keys de las tabla origen – destino y la el tipo de relación, elegir la opción **Business View**, en esta pantalla añadir las categorías que deseamos mostrar en nuestro modelo, para la aplicación seleccionaremos todas las categorías y obtenemos nuestro modelo de negocio.

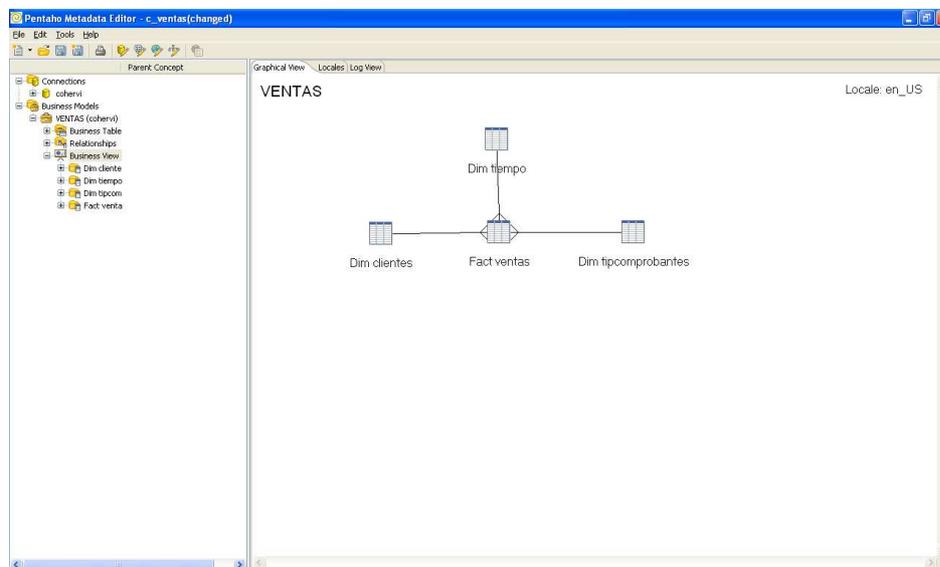
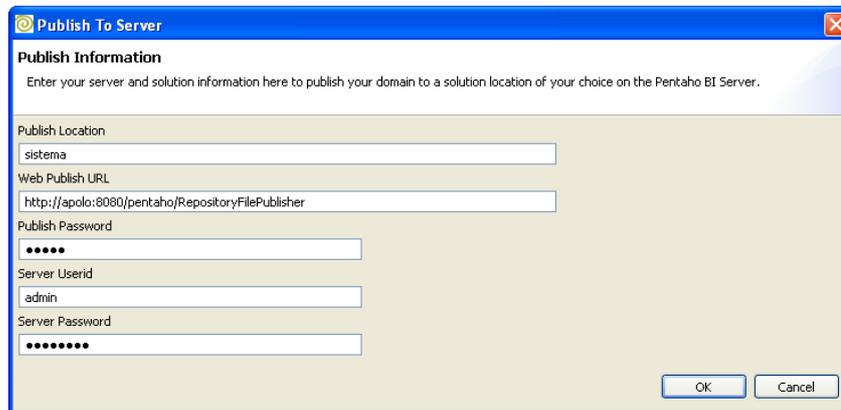


Figura V.123 - Modelo de Negocios

Finalmente seleccionamos **File**, luego **Publish**, donde se obtendrá la siguiente pantalla, en la que se ingresará el repositorio, la URL, el password de publicación(admin), el usuario y password del usuario de pentaho y seleccionar el botón **OK**.



**Figura V.124 - Parámetros de publicación de Metadata Editor**

#### **5.2.4.4.3.6 Ejecutar consola de usuario**

Para ingresar a la consola de usuario, deber seleccionar un navegador de internet e ingresar la siguiente URL **http://zeus:8080/pentaho/**, donde el usuario deberá autenticarse ingresando su login y password.

Para ver la funcionalidad de Pentaho revisar manual de usuario.

### **5.2.5 FASE V: PRUEBAS E IMPLANTACIÓN.**

#### **5.2.5.1 PROPÓSITO:**

Detectar y realizar el seguimiento de errores e implantar el sistema.

### 5.2.5.2 DESCRIPCIÓN:

Una descripción gráfica de la fase como muestra la siguiente figura.

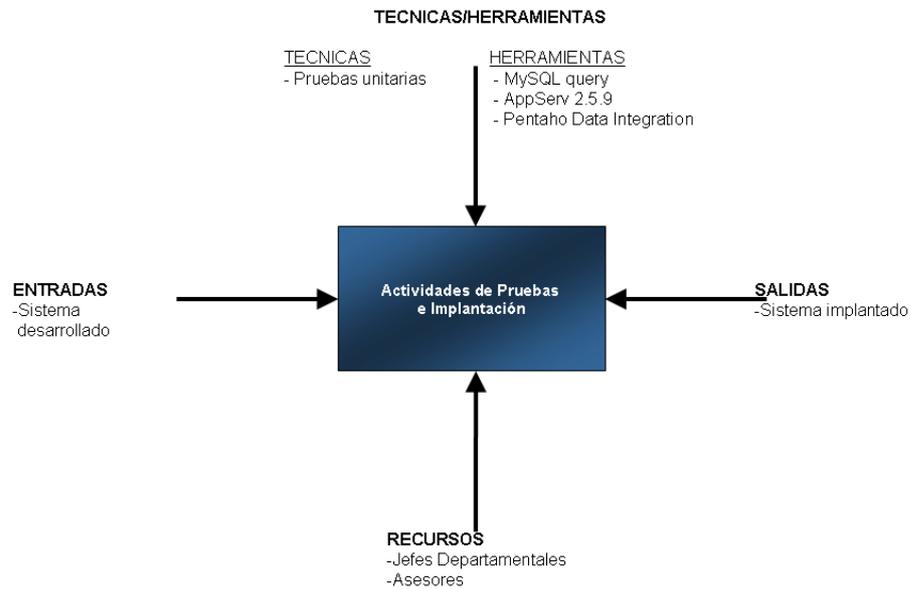


Figura V.125 - Descripción de la fase.

### 5.2.5.3 RESULTADOS:

🚩 Sistema Implantado

### 5.2.5.4 ACTIVIDADES:

Las actividades a realizar en esta fase son:

#### 5.2.5.4.1 Planificar las prueba.

En esta actividad se definen los objetivos de la prueba, los objetos probados, los alcances de la prueba, el método utilizado, los recursos empleados, el plan de tiempos y finalmente las responsabilidades para llevar adelante la prueba.

### 5.2.5.4.2 Generar casos de pruebas

Los casos de prueba (para las pruebas funcionales o de caja negra) están basados en lotes de datos reales, los cuáles se han obtenido de la base de datos fuente. Para el diseño de los casos, se toma como base la lista de funciones, llamando lista de funciones a los requerimientos.

Las pruebas unitarias de cálculo se realizan en base a diversas consultas (instrucciones SQL) realizadas manualmente en la base de datos fuentes. De manera de poder conseguir resultados y valores para cotejarlos con los resultados que se obtienen utilizando business intelligence. Ver anexo 1.

En la fase de pruebas se muestran los casos de prueba completos y la integración con los formularios que se describen a continuación.

**Tabla V.89 - Caso de prueba 1**

<b>CASO DE PRUEBA</b>			
<b>Caso número:</b> 1	<b>Proyecto:</b> SAG	<b>Documento Soporte:</b>	Documento actual Capítulo de análisis y diseño
<b>Versión:</b> 1.0	<b>Módulo:</b> Base de Datos dwh_cohervi		
<b>Pre-requisitos del caso de prueba</b>	<ul style="list-style-type: none"> <li>- Disponibilidad de la Base de Datos Cohervi</li> <li>- Base de Datos dwh_cohervi creada</li> <li>- Proceso ETL Base de Datos Cohervi a Base de Datos dwh_cohervi creado y ejecutado.</li> </ul>		
<b>Entradas requeridas:</b>	N/A		
<b>Descripción de la prueba:</b>	Verificar los modelos de dato y datos		
<b>Paso No.</b>	<b>No. De función a probar</b>	<b>Acción o Instrucción</b>	<b>Resultado esperado</b>
1	Cantidad de Tablas en DBFuente vs. Cantidad de Tablas en DBdwh	Comparar cantidad de tablas.	Cantidad Tablas iguales

**Tabla V.89 - Caso de prueba 1 (continuación)**

2	Cantidad de reg x tabla en DBFuente Vs. Cantidad de reg x tabla en DBdwh	Contar cantidad de Registros por cada tabla de DBFuente y DBdwh y comparar.	Cantidad de registros iguales
---	--	---	-------------------------------

**Tabla V.90 - Caso de prueba 2**

<b>CASO DE PRUEBA</b>			
<b>Caso número:</b> 2	<b>Proyecto:</b> SAG	<b>Documento Soporte:</b>	Documento actual Capítulo Diseño
<b>Versión:</b> 1.0	<b>Módulo:</b> Extracción, Transformación y Carga		
<b>Pre-requisitos del caso de prueba</b>	<ul style="list-style-type: none"> <li>- Disponibilidad de la Base de Datos Cohervi</li> <li>- Base de Datos dwh_cohervi lista para su uso</li> <li>- Creados y listos los Proceso ETL Base de Datos Cohervi a Base de Datos dwh_cohervi.</li> </ul>		
<b>Entradas requeridas:</b>	N/A		
<b>Descripción de la prueba:</b>	Prueba de Extracción, Transformación y Carga EXITOSA desde la base de datos Fuente a la base de datos DBdwh_cohervi.		
<b>Paso No.</b>	<b>No. De función a probar</b>	<b>Acción o Instrucción</b>	<b>Resultado esperado</b>
1	Activación Automática	Se deberá forzar la ejecución definiendo el día hora en el job_carga_dwh para que se active el ETL en el momento deseado.	Inicio del proceso ETL

**Tabla V.90 - Caso de prueba 2(continuación)**

2	Realización de la transferencia de datos	N/A.	Nuevos datos en la Base de Datos
3	Mensaje EXITOSO en la consola	N/A	Mensaje exitoso en la consola tras ejecutar el proceso batch

**Tabla V.91 - Caso de prueba 3**

<b>CASO DE PRUEBA</b>			
<b>Caso número:</b> 3	<b>Proyecto:</b> SAG	<b>Documento Soporte:</b>	Documento actual Capítulo Diseño
<b>Versión:</b> 1.0	<b>Módulo:</b> Extracción, Transformación y Carga		
<b>Pre-requisitos del caso de prueba</b>	<ul style="list-style-type: none"> <li>- Disponibilidad de la Base de Datos Cohervi</li> <li>- Base de Datos dwh_cohervi lista para su uso</li> <li>- Creados y listos los Proceso ETL Base de Datos Cohervi a Base de Datos dwh_cohervi.</li> </ul>		
<b>Entradas requeridas:</b>	N/A		
<b>Descripción de la prueba:</b>	Prueba de Extracción, Transformación y Carga ERRONEA desde la base de datos Fuente a la base de datos DBdwh_cohervi.		
<b>Paso No.</b>	<b>No. De función a probar</b>	<b>Acción o Instrucción</b>	<b>Resultado esperado</b>
1	Activación Automática	Se deberá forzar la ejecución definiendo el día hora en el job_carga_dwh para	Inicio del proceso ETL

**Tabla V.91 - Caso de prueba 3 (continuación)**

1	Activación Automática	que se active el ETL en el momento deseado.	
2	Mensaje ERRONEO en la consola	N/A	Mensaje de error en la consola tras ejecutar el proceso batch

**Tabla V.92 - Caso de prueba 4**

<b>CASO DE PRUEBA</b>			
<b>Caso número:</b> 4	<b>Proyecto:</b> SAG	<b>Documento Soporte:</b>	Documento actual Capítulo Diseño
<b>Versión:</b> 1.0	<b>Módulo:</b> Extracción, Transformación y Carga		
<b>Pre-requisitos del caso de prueba</b>	<ul style="list-style-type: none"> <li>- Disponibilidad de la Base de Datos Cohervi</li> <li>- Base de Datos ddm_cohervi lista para su uso</li> <li>- Creados y listos los Proceso ETL de la Base de Datos dwh_cohervi a Base de Datos ddm_cohervi.</li> </ul>		
<b>Entradas requeridas:</b>	N/A		
<b>Descripción de la prueba:</b>	Prueba de Extracción, Transformación y Carga EXITOSA desde la Base de Datos dwh_cohervi a la Base de Datos ddm_cohervi.		
<b>Paso No.</b>	<b>No. De función a probar</b>	<b>Acción o Instrucción</b>	<b>Resultado esperado</b>
1	Activación Automática	Se deberá forzar la ejecución definiendo el día hora	Inicio del proceso ETL

**Tabla V.92 - Caso de prueba 4(continuación)**

		en el job_carga_ddm para que se active el ETL en el momento deseado.	
2	Realización de la transferencia de datos	N/A.	Nuevos datos en la Base de Datos
3	Mensaje EXITOSO en la consola	N/A	Mensaje EXITOSO en la consola tras ejecutar el proceso batch

**Tabla V.93 - Caso de prueba 5**

<b>CASO DE PRUEBA</b>			
<b>Caso número:</b> 5	<b>Proyecto:</b> SAG	<b>Documento Soporte:</b>	Documento actual Capítulo Diseño
<b>Versión:</b> 1.0	<b>Módulo:</b> Extracción, Transformación y Carga		
<b>Pre-requisitos del caso de prueba</b>	- Disponibilidad de la Base de Datos Cohervi - Base de Datos ddm_cohervi lista para su uso - Creados y listos los Proceso ETL de la Base de Datos dwh_cohervi a Base de Datos ddm_cohervi.		
<b>Entradas requeridas:</b>	N/A		
<b>Descripción de la prueba:</b>	Prueba de Extracción, Transformación y Carga ERRONEA desde la Base de Datos dwh_cohervi a la Base de Datos ddm_cohervi.		
<b>Paso No.</b>	<b>No. De función a probar</b>	<b>Acción o Instrucción</b>	<b>Resultado esperado</b>
1	Activación Automática	Se deberá forzar la ejecución definiendo el día hora	Inicio del proceso ETL

**Tabla V.93 - Caso de prueba 5 (continuación)**

		en el job_carga_ddm para que se active el ETL en el momento deseado.	
2	Mensaje ERRONEO en la consola	N/A	Mensaje ERRONEO en la consola tras ejecutar el proceso batch

**Tabla V.94 - Caso de prueba 6**

<b>CASO DE PRUEBA</b>			
<b>Caso número:</b> 6	<b>Proyecto:</b> SAG	<b>Documento Soporte:</b>	<ul style="list-style-type: none"> <li>- Documento actual</li> <li>- Capítulo Análisis donde se detalla los requerimientos del sistema.</li> <li>- Ver Anexos para más detalle de las operaciones con Cubos</li> </ul>
<b>Versión:</b> 1.0	<b>Módulo:</b> Servicios OLAP		
<b>Pre-requisitos del caso de prueba</b>	Las tablas de dimensiones y hechos deben estar cargadas en el datamart. Los cubos multidimensionales deben estar procesados y listos para su uso.		
<b>Entradas requeridas:</b>	N/A		
<b>Descripción de la prueba:</b>	Son Pruebas de Funcionalidad en Servicios OLAP. Se probará que la información en cada uno de los cubos se corresponda a la obtenida de la base de datos del sistema fuente, los datos de la base fuente se obtendrán por medio de instrucciones SQL.		

**Tabla V.94 - Caso de prueba 6 (continuación)**

<b>Paso No.</b>	<b>No. De función a probar</b>	<b>Acción o Instrucción</b>	<b>Resultado esperado</b>
1		Utilizar las funciones de slice, drill-down, drill-up, swap y filtrado del manejador de Cubos.	La cantidad de ... cargos docentes según cargo y dedicación debe ser igual a la obtenida directamente desde la base del Sistema Académico.

**Tabla V.95 - Caso de prueba 7**

<b>CASO DE PRUEBA</b>			
<b>Caso número:</b> 7	<b>Proyecto:</b> SAG	<b>Documento Soporte:</b>	<ul style="list-style-type: none"> <li>- Documento actual</li> <li>- Capítulo Análisis donde se detalla los requerimientos del sistema.</li> <li>- Ver Anexos para mas detalle de las operaciones con Cubos</li> </ul>
<b>Versión:</b> 1.0	<b>Módulo:</b> Reportes Estáticos y Reportes Ad Hoc		
<b>Pre-requisitos del caso de prueba</b>	<p>Las tablas de dimensiones y hechos deben estar cargadas en el datamart.</p> <p>Los cubos multidimensionales deben estar procesados y listos para su uso.</p>		
<b>Entradas requeridas:</b>	N/A		
<b>Descripción de la prueba:</b>	<p>Son Pruebas de Funcionalidad de los reportes. Se probará que la información en cada uno de los reportes se corresponda a la obtenida de la base de datos del sistema fuente, los datos de la base fuente se obtendrán por medio de instrucciones SQL.</p>		

**Tabla V.95 - Caso de prueba 7 (continuación)**

Paso No.	No. De función a probar	Acción o Instrucción	Resultado esperado
1		Utilizar las sentencias SQL previa la realización de los reportes para ver los registros	La cantidad de ... cargos docentes según cargo y dedicación debe ser igual a la obtenida directamente desde la base del Sistema Académico. L

**5.2.5.4.3 Realizar procedimientos de ejecución y reportes.**

La siguiente tabla nos indica el procedimiento que se realizará para ejecutar el caso de prueba planificado en la actividad anterior:

**CASO DE PRUEBA 1. PASO 1:**

La cantidad de tablas de la base de datos fuente es igual a la cantidad de tablas en la Base de Datos dwh\_Cohervi, como lo indica la siguiente tabla.

**Tabla V.96 - Caso de prueba 1 paso 1**

CANTIDAD DE TABLAS	
BDFUENTE	BD dwh_cohervi
9	9

### CASO DE PRUEBA 1. PASO 2:

En este paso se cuenta la cantidad de registros de cada una de las tablas fuente y dwh\_cohervi.

**Tabla V.97 - Caso de prueba 1 paso 2**

TABLAS DE LA BDFUENTE	CANTIDAD DE FILAS	TABLAS DE DWH_COHERVI	CANTIDAD DE FILAS
ProductosCaducados.xls	17301	Dwh_cohervi	17301
Clientes	1053	dwh_clientes	1053
INDETCST	34852	Dwh_compras	34852
INDETINV	1863771	Dwh_kardex	1863771
CCDEBITO	752868	Dwh_movimientos	752868
ICMARTIO	17301	dwh_productos	17301
CPPMPVD0	117	dwh_proveedores	117
ICRTICO0	18	dwh_tipcomprobantes	18
FAPTOVTA	12	dwh_tipovtas	12

### CASO DE PRUEBA 2:

Los resultados del caso de prueba 2 son los siguientes:

**Tabla V.98 - Caso de prueba 2**

PASO NO.	NO. DE FUNCIÓN A PROBAR	RESULTADO ESPERADO	RESULTADO OBTENIDO
1	Activación Automática	Inicio del proceso ETL	Inicio OK del proceso ETL
2	Realización de la transferencia de datos	Nuevos datos en la BD	Nuevos datos en la BD
3	Mensaje de ÉXITO en la consola	Mensaje EXITOSO en la consola tras ejecutar el proceso batch	Mensaje de finalización con ÉXITO del proceso en la consola

**CASO DE PRUEBA 3:**

Los resultados de los casos de prueba son los siguientes:

**Tabla V.99 - Caso de prueba 3**

<b>PASO NO.</b>	<b>NO. DE FUNCIÓN A PROBAR</b>	<b>RESULTADO ESPERADO</b>	<b>RESULTADO OBTENIDO</b>
1	Activación Automática	Inicio del proceso ETL	Inicio OK del proceso ETL
2	Mensaje ERRONEO en la consola	Mensaje ERRONEO en la consola tras ejecutar el proceso batch	Mensaje de finalización con ERRONEO del proceso en la consola

**CASO DE PRUEBA 4:**

Los resultados del caso de prueba 2 son los siguientes:

**Tabla V.100 - Caso de prueba 4**

<b>PASO NO.</b>	<b>NO. DE FUNCIÓN A PROBAR</b>	<b>RESULTADO ESPERADO</b>	<b>RESULTADO OBTENIDO</b>
1	Activación Automática	Inicio del proceso ETL	Inicio OK del proceso ETL
2	Realización de la transferencia de datos	Nuevos datos en la BD	Nuevos datos en la BD
3	Mensaje de ÉXITO en la consola	Mensaje EXITOSO en la consola tras ejecutar el proceso batch	Mensaje de finalización con ÉXITO del proceso en la consola

**Tabla V.100 - Caso de prueba 4 (continuación)**

DB dwh_cohervi		DB ddm_cohervi	
Tabla	Cantidad de Filas	Tabla	Cantidad de filas
Dwh_clientes	1054	Dim_clientes	1054
Dwh_productos	17302	Dim_productos	17302
Dwh_proveedores	118	Dim_proveedores	118
		Dim_tiempo	9349
Dwh_tipcomprobantes,	18	Dim_tipcomprobantes	31
dwh_tipovta	12		
dwh_caducados	17301	Fact_caducados	15075

**CASO DE PRUEBA 5:**

Los resultados de los casos de prueba son los siguientes:

**Tabla V.101 - Caso de prueba 5**

PASO NO.	NO. DE FUNCIÓN A PROBAR	RESULTADO ESPERADO	RESULTADO OBTENIDO
1	Activación Automática	Inicio del proceso ETL	Inicio OK del proceso ETL
2	Mensaje ERRONEO en la consola	Mensaje ERRONEO en la consola tras ejecutar el proceso batch	Mensaje de finalización con ERRONEO del proceso en la consola

Si la ejecución de un caso de prueba produce resultados incorrectos o inesperados, el probador completa un **“Informe de Problema”**.

**Tabla V.102 - Informe de problema.**

<b>INFORME DE PROBLEMA</b>	
Caso número: 4	Proyecto: SAG
Fecha:	Sistema:
Probador: Elizabeth Ilbay	
Descripción del problema: al realizar la consulta no carga los registros que debería mostrar la tabla dwh_kardex al llenar la Fact._inventario	
Otros documentos:	Consulta SQL, SQL adjuntado a anexo.

#### 5.2.5.4.4 Formalizar procedimientos de corrección de errores

El problema que surgió en el desarrollo del SAG, fue en una transformación para la carga de la fact\_table de inventarios.

**Tabla V.103 - Informe de problemas identificados estado abierto**

<b>INFORME DE PROBLEMAS IDENTIFICADOS</b>							
Proyecto: SAG				Registro de problemas:			
Sistema: ddm_cohervi							
Elaborado por: Elsi Ilbay				Página 1			
				Fecha :			
<b>N</b>	<b>Problema</b>	<b>Código de severidad</b>	<b>Probador</b>	<b>Responsable</b>	<b>Estado</b>	<b>Fecha de corrección requerida</b>	<b>Comentarios</b>
4	No carga los registros el proceso ETL	2	Elizabeth Ilbay	Elizabeth Ilbay	Abierto	04/09/09	El proceso ETL se suspende

**Tabla V.103 - Informe de problemas identificados estado abierto (continuación)**

Número de problemas identificados					
Los problemas encontrados en la carga es en una Fact._table, por cuanto el computador se cae al realizar el proceso ETL.					
Severidad	1 X	2.....	3.....	4.....	5.....
ad		.....	.	.....	

#### 5.2.5.4.5 Implantar el sistema.

Para la implantación del sistema se necesitan de ciertos requerimientos hardware y software en el Servidor y Clientes.

### REQUERIMIENTOS SERVIDOR

**Tabla V.104 – Requerimientos el Servidor**

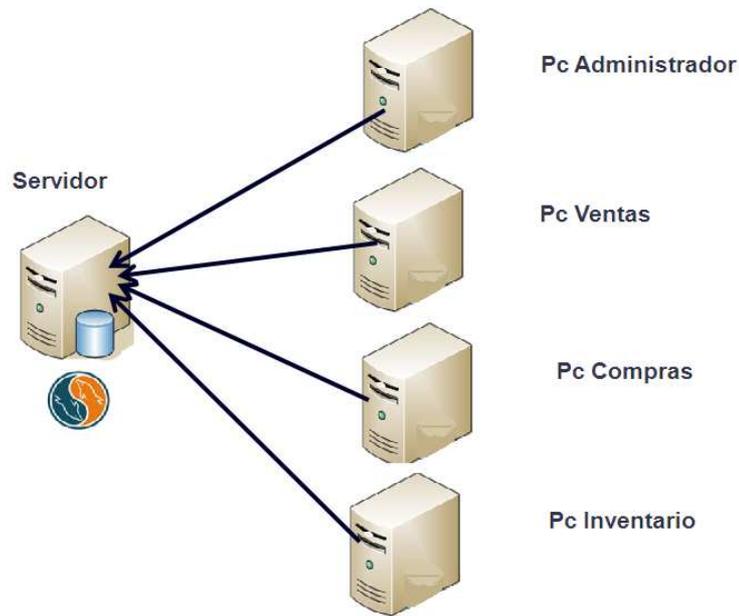
HARDWARE	
Discos Duro:	500Gb
Memoria:	4 GB
Procesador:	Intel PIV 3.2 Ghz
SOFTWARE	
Java Virtual Machine:	Sun 1.6

### REQUERIMIENTOS CLIENTE

**Tabla V.105 – Requerimientos del Cliente**

HARDWARE	
Discos Duro:	40Gb
Memoria:	1 GB
Procesador:	Mínimo Intel PII
SOFTWARE	
Explorador Web:	Internet Explorer, Mozilla Firefox, etc
Sistema Operativo:	Windows, Linux

Una vez implantado el sistema queda de la siguiente manera:



**Figura V.126 - Sistema implementado**

### **5.2.6 FASE VI: ADMINISTRACIÓN Y MANTENIMIENTO DEL SISTEMA.**

Una vez implantado el sistema, no significa que se termina al contrario como se dijo en los capítulos anteriores el desarrollo de un proyecto de business intelligence es un proceso de naturaleza espiral, es por ello que se necesita continuar con los relevamientos de forma constante para poder seguir con la evolución.

#### **5.2.6.1 PROPÓSITO:**

Administrar las bases de datos dwh\_cohervi y ddm\_cohervi y la Suite de Pentaho.

### 5.2.6.2 DESCRIPCIÓN:

Una descripción gráfica de la fase como muestra la siguiente figura.

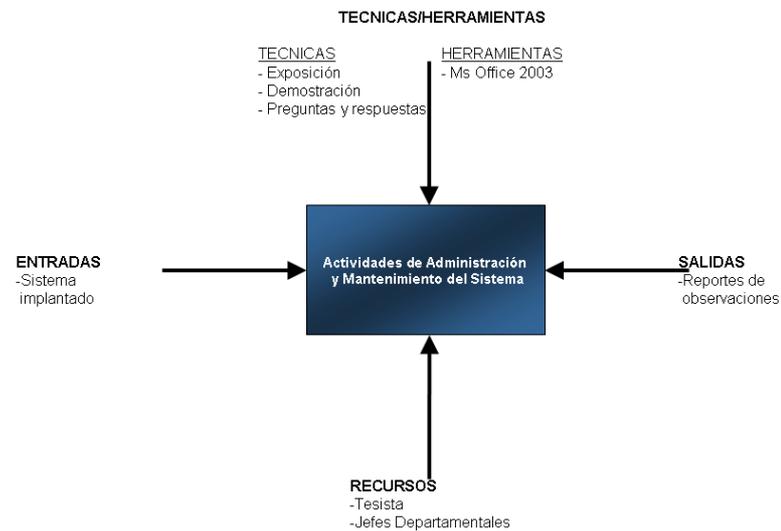


Figura V.127 - Descripción de la fase.

### 5.2.6.3 RESULTADOS:

- 🚩 Reporte de observaciones.

### 5.2.6.4 ACTIVIDADES:

Las actividades a desarrollar en esta fase son las siguientes:

#### 5.2.6.4.1 Capacitar y dar soporte técnico.

Se realizó la capacitación a los jefes departamentales durante 40 horas, los temas tratados fueron:

- 🚩 Manipulación de la consola de usuario
- 🚩 Manejo de la consola de administrador
- 🚩 Creación y manipulación de Vista OLAP
- 🚩 Creación y manipulación de Reportes estáticos y Ad-Hoc

**5.2.6.4.2 Realizar la gestión de los módulos.**

Se realizará la gestión de los siguientes módulos:

**Tabla V.106 - Extracción, transformación y carga**

Caso de uso	Gestión de Extracción y Carga
Requerimiento que lo implementa	Módulo necesario para el funcionamiento del sistema.
Descripción	<ul style="list-style-type: none"> <li>• Es el módulo que contiene las facilidades necesarias para construcción, y mantenimiento de la extracción y carga de datos en el dwh_cohervi, ddm_cohervi</li> </ul> <p>El módulo se implementa por medio de la herramienta Pentaho Data Integration.</p> <p>Con los ETL se importa, exporta y transforman los datos entre la base de datos fuente y el datawarehouse. También desde la Bd dwh_cohervi y ddm_cohervi.</p> <p>Estos procesos se automatizan de manera que se ejecutan sin intervención de los usuarios.</p> <p>Los procesos que se hacen referencia son:</p> <p>Condición de Inicio, Extracción de BD Fuente, Transformación, Carga En el DW y datamart, Aviso al Operador de Error y Aviso al Operador de Éxito.</p> <ul style="list-style-type: none"> <li>• Los usuarios responsables de utilizar este módulo son los usuarios “Administrador” y su principal tarea es la Administración y Mantenimiento de los procesos de extracción, transformación y carga. Los procesos de extracción, transformación y carga.</li> </ul> <p>Los usuarios del módulo son “Administrador” y tienen permisos de Administrador (permisos totales sobre de las bases de datos de dwh_cohervi y ddm_cohervi)</p>
Figuras asociadas	Figura de Procesos de Extracción, Transformación y Carga
Observaciones	n/a

**Tabla V.107 - Datawarehouse.**

<b>Caso de uso</b>	<b>Gestión del Warehouse</b>
Requerimiento que lo implementa	Módulo necesario para el funcionamiento del sistema.
Descripción	<ul style="list-style-type: none"> <li>• Este modulo tiene funciones de administración y mantenimiento de las bases de datos que componen el Datawarehouse y el Datamart. Permite controlar el funcionamiento diario de estas bases de datos. Entre sus tareas se encuentran las siguientes: <ul style="list-style-type: none"> <li>– Mantenimiento de tablas, vistas e índices.</li> <li>– Mantenimiento de la integridad entre los objetos de las bases de datos.</li> <li>– Mantenimiento de las agregaciones y actualización de las existentes.</li> <li>– Acumulación de los datos del datawarehouse y datamart para su salvaguarda.</li> <li>– Procesos automáticos de backups.</li> <li>– Administración de la seguridad y perfiles de usuarios.</li> <li>– Mantener el Metadatos del datawarehouse y del datamart.</li> </ul> </li> <li>• La Gestión del warehouse se realiza por medio de las herramientas que provee MYSQL para realizar las tareas diarias administración y mantenimiento.</li> <li>• Los usuarios del este módulo son “Administrador” y tienen permisos de Administrador (permisos totales sobre de las bases de datos dwh_cohervi y ddm_cohervi).</li> </ul>
Figuras asociadas	n/a
Observaciones	n/a

**Tabla V.108 - Acceso a datos o consultas.**

<b>Caso de uso</b>	<b>Gestión de Consultas - Acceso a Datos</b>
Requerimiento que lo implementa	Módulo necesario para el funcionamiento del sistema.
Descripción	<ul style="list-style-type: none"><li>• Permite llevar toda la lógica necesaria para apoyar el proceso de gestión de consultas y análisis de información. Implementa por medio de las herramientas OLAP Mondrian, Metadata Editor, BIRT.</li></ul> Tareas más significativas son: <ul style="list-style-type: none"><li>– Dirigir las consultas a las tablas adecuadas.</li><li>– Creación y mantenimiento de Cubos multidimensionales.</li><li>– Realizar los reportes y gráficos requeridos por los usuarios.</li><li>– Planificación de las consultas junto con los usuarios.</li><li>– Apoyo a los usuarios en modelos de análisis de la información.</li></ul> <ul style="list-style-type: none"><li>• Los usuarios de este módulo son los denominados Administrador, Ventas, Inventario, Compras</li></ul> Accesos a Datos, y sus permisos de accesos en las bases de datos están restringidos a solo consultas en la base de datos ddm_cohervi.
Figuras asociadas	n/a
Observaciones	n/a

### 5.3 RESULTADOS OBTENIDOS DE LA METODOLOGÍA PROPUESTA

Una vez desarrollada y aplicada la metodología en COHERVI S.A se realiza un análisis de los resultados obtenidos en base a tres variables, que son:

- A. Tiempo de elaboración de informes
- B. Complejidad de elaboración de los informes
- C. Satisfacción del usuario estratégico

#### MUESTRA

La muestra tomada en consideración para analizar esta variable es una persona, se realizó la entrevista al jefe del departamento de sistemas Ing. Moisés López, quien clasifica a los informes en: baja, media y alta complejidad y bajo, medio y alto tiempo de elaboración de los informes.

#### INFORME vs REPORTE

INFORME = Reporte de Análisis (OLAP) (dirigido → usuario de análisis)

REPORTE = Reporte Transaccional (dirigido → usuario final)

#### a) TIEMPO DE ELABORACIÓN DE INFORMES

Los informes se construye en base a requerimientos, los puntos a evaluar en este criterio son:

1. **Análisis de requerimientos:** El tiempo que se tomó en analizar los requerimientos de cada uno de los departamentos.
2. **Construcción:** Se consideró el tiempo de elaboración del informe, el mismo que dependerá si es un informe de baja, media o alta complejidad.
3. **Pruebas:** Tiempo que se utilizó en comprobar el informe elaborado con los usuarios de las áreas departamentales.

#### PUNTAJES: (El puntaje es de acuerdo a un rango de tiempo en horas) SISTEMA TRANSACCIONAL

BAJA: 0,008 h (30 segundos) - 8 horas

MEDIA: 8,0003h (8 horas + 1 segundo) - 24 horas

ALTA: mayor a 24 h

**PUNTAJES: (El puntaje es de acuerdo a un rango de tiempo en horas)) SISTEMA DE ANALISIS GERENCIAL**

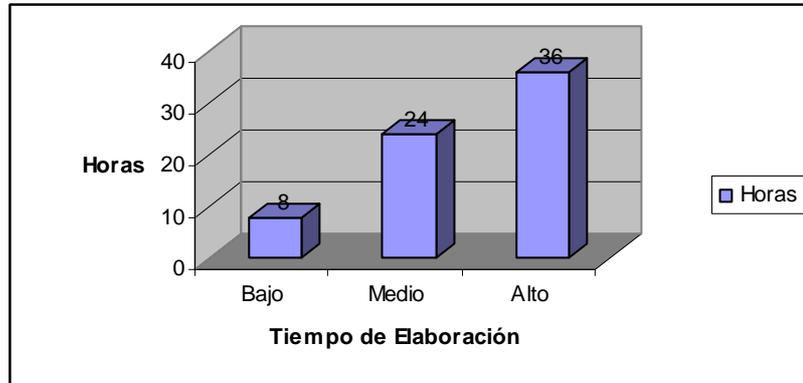
BAJA: 0,005h (30 segundos) - 3 horas  
 MEDIA: 3,0003h (3horas +1segundo) - 10 horas  
 ALTA: mayor a 10 h

**Tabla V.109 - Comparación del tiempo de elaboración de informes utilizando el Sistema Transaccional VS SAG**

CRITERIOS ELABORACION DE INFORMES/REPORTES	TIEMPO DE ELABORACION DE INFORMES (Tiempos/Horas)					
	SISTEMA TRANSACCIONAL ACTUAL			SAG (Sistema de Análisis Gerencial)		
	BAJA	MEDIA	ALTA	BAJA	MEDIA	ALTA
	1. Análisis de Requerimientos	2	4	4	2	4
2. Construcción	4	16	24	0,5	4	8
3. Pruebas	2	4	8	0,5	2	4
<b>Total</b>	<b>8</b>	<b>24</b>	<b>36</b>	<b>3</b>	<b>10</b>	<b>16</b>

La siguiente tabla muestra los criterios de elaboración de informes/reportes correspondientes a las filas, las columnas representan los tiempos de elaboración de informes del Sistema Transaccional vs. Sistema de Análisis Gerencial. Los mismos que tienen un tiempo de elaboración de bajo, medio y alto, los valores de las celdas constituyen los puntajes asignados a cada uno de las sistemas (Transaccional y SAG) de acuerdo a los criterios de los desarrolladores (jefe de sistemas y tesista).

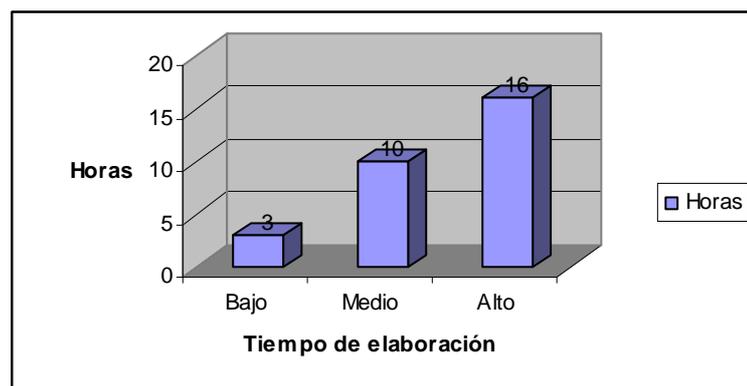
### TIEMPO DE ELABORACIÓN DE INFORMES UTILIZANDO EL SISTEMA TRANSACCIONAL - (Tiempo/Horas)



#### ANÁLISIS:

Los tiempos que se maneja en elaborar informes de análisis utilizando el Sistema Transaccional se lo clasifica en tres tipos: Bajo, Medio y Alto. Estos informes son desarrollados en 8, 24 y 36 horas respectivamente, tiempo que es considerado muy extenso para emitir informes a las áreas departamentales.

### TIEMPO DE ELABORACIÓN DE INFORMES UTILIZANDO EL SISTEMA DE ANÁLISIS GERENCIAL (SAG) - (Tiempo/Horas)



#### ANÁLISIS:

Esta gráfica indica, que construir informes utilizando el Sistema de Análisis Gerencial (SAG) se lo realiza en menor tiempo, los informes de bajo, medio y alto tiempo son

desarrollados en 3,10 y 16 horas respectivamente, tiempo reducido a la mitad del tiempo con respecto a los informes desarrollados con el Sistema Transaccional.

Estos informes permiten a los jefes departamentales contar con información actualizada e histórica permitiéndoles realizar el análisis de las Ventas, informe de compras, stock de inventario, monto de compras, etc.

**b) COMPLEJIDAD DE ELABORACIÓN DE LOS INFORMES**

En esta variable para ponderar los valores de alta, media y baja complejidad se toman en cuenta los siguientes parámetros.

Reporte estático → sencilla consulta SQL

Reporte ad Hoc → reporte dinámico personalizado

Vista de análisis → cubos OLAP

PUNTAJES: (El puntaje asignado es de acuerdo a la siguiente ponderación de: 1,2,3)

Alta = 3 (Vistas de análisis + gráfico+cálculos a nivel de medidas + operaciones OLAP)

Media = 2 (reporte Ad Hoc + cálculos a nivel de medidas)

Baja = 1 (reporte estático – sin gráfica ni cálculos)

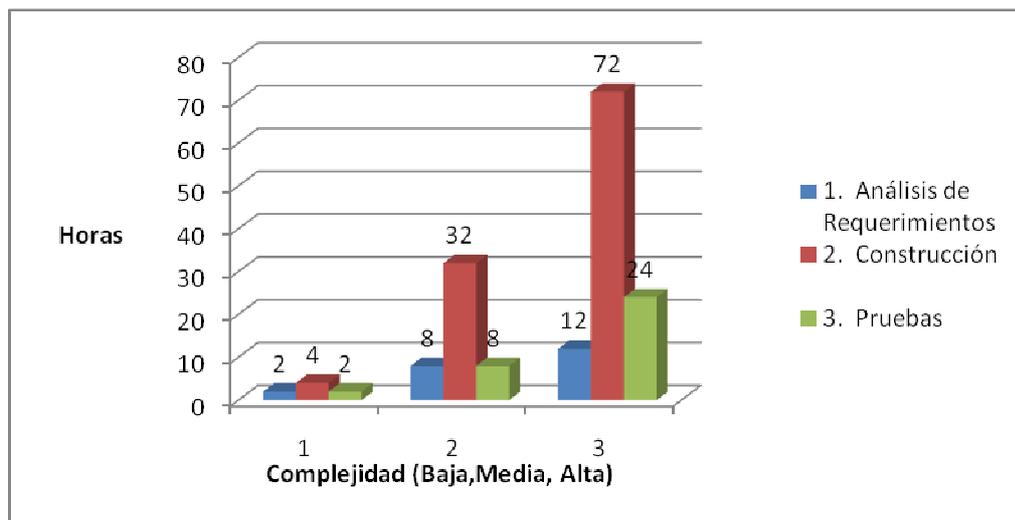
**Tabla V.110 - Tiempo en relación a la complejidad**

CRITERIOS ELABORACION DE INFORMES/REPORTES	COMPLEJIDAD DE INFORMES (Tiempos/Horas)					
	SISTEMA TRANSACCIONAL ACTUAL			SAG (Sistema de Análisis Gerencial)		
	BAJA	MEDIA	ALTA	BAJA	MEDIA	ALTA
	1. Análisis de Requerimientos	2	8	12	2	8
2. Construcción	4	32	72	0,5	8	24
3. Pruebas	2	8	24	0,5	4	12
<b>Total</b>	<b>8</b>	<b>48</b>	<b>108</b>	<b>3</b>	<b>20</b>	<b>48</b>

La siguiente tabla muestra los criterios de elaboración de informes/reportes correspondientes a las filas, las columnas representan la complejidad de los informes

del Sistema Transaccional vs. Sistema de Análisis Gerencial. Los mismos que tienen baja, media y alta complejidad, los valores de las celdas representan los puntajes que se ha asignado a cada uno de los sistemas (Transaccional y SAG) de acuerdo a una ponderación de 1, 2, 3.

### COMPLEJIDAD DE INFORMES UTILIZANDO EL SISTEMA TRANSACCIONAL (Tiempo/Horas)

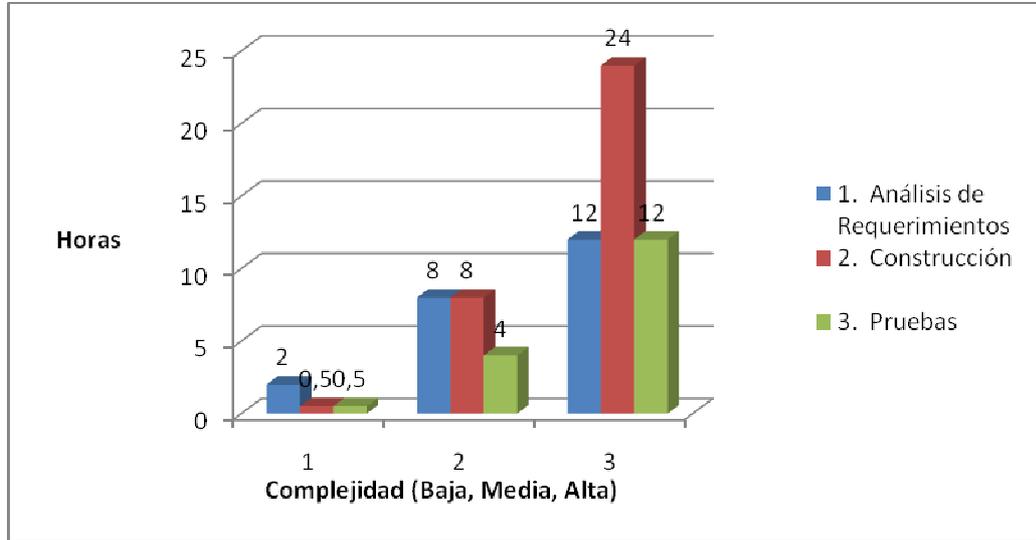


#### ANÁLISIS

La siguiente gráfica da a conocer los criterios de elaboración tales como: Análisis de requerimientos, Construcción y Pruebas, cada criterio es representativo de acuerdo a la complejidad (baja, media y alta) de los informes utilizando el Sistema Transaccional.

En el desarrollo de los informes de **Baja complejidad** en el Análisis de Requerimientos tenemos un tiempo de 2 horas, en la Construcción 4 horas y en las pruebas 2 horas. En los informes de **Complejidad Media** el Análisis de Requerimientos el tiempo de 8 horas, en la Construcción 32 horas y en las pruebas 8 horas. En los **Complejidad Alta** el Análisis de Requerimientos es de 12 horas, la Construcción es en 72 horas y las Pruebas en 3 horas.

**COMPLEJIDAD DE INFORMES UTILIZANDO EL SISTEMA DE ANALISIS GERENCIAL (SAG)- Tiempo/Horas**



**ANÁLISIS**

La gráfica representa el tiempo que se requiere para desarrollar los informes de complejidad baja, media y alta de acuerdo a los criterios de elaboración como son: Análisis de Requerimientos, Construcción y Pruebas utilizando el Sistema de Análisis Gerencial (SAG).

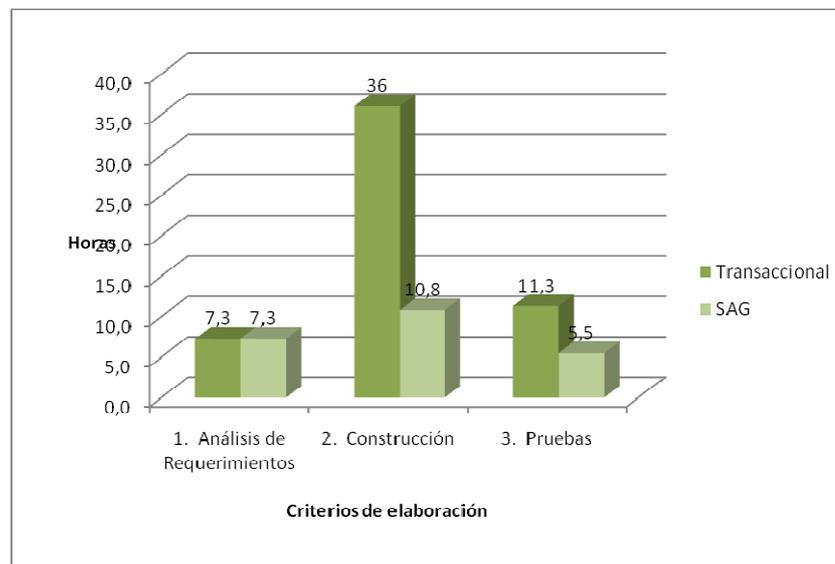
El tiempo empleado en un informe de **Baja Complejidad** en el Análisis de Requerimientos es de 2 horas, en la Construcción 0,50 horas y Pruebas en 0,50 horas. En los informes de Complejidad Media en el Análisis de Requerimientos es de 8 horas, en la Construcción 8 horas y Pruebas en 4 horas. Informes de **Alta Complejidad** en el Análisis de Requerimientos es de 12 horas, en la Construcción 24 horas y Pruebas en 12 horas.

**Tabla V.111 – Promedio de Tiempos/Horas de acuerdo a los criterios de elaboración de los Sistemas Transaccional vs. SAG**

CRITERIOS DE ELABORACION DE INFORMES	SISTEMAS (Horas/Promedio)	
	TRANSACCIONAL	SAG
1. Análisis de Requerimientos	7,3	7,3
2. Construcción	36	10,8
3. Pruebas	11,3	5,5

La siguiente tabla nos indica, en las filas los criterios de elaboración de informes y las columnas representan a los sistemas: Transaccional y Sistema de Análisis Gerencial (SAG), los valores de las celdas corresponden a los tiempos promedios en desarrollar cada uno de los criterios de elaboración de informes.

### Promedio de Tiempos/Horas de acuerdo a los criterios de elaboración de los Sistemas Transaccional vs. SAG



### ANÁLISIS

La gráfica da a conocer el tiempo de elaboración de informes de acuerdo a su complejidad promedio en los Sistemas de Transaccional vs. Sistema de Análisis Gerencial (SAG).

En el Análisis de requerimientos el tiempo promedio es de 7,3 horas tanto para el Sistema Transaccional y el Sistema de Análisis Gerencial (SAG), en la Construcción de los informes usando el Sistema Transaccional es de 36 horas mientras que el SAG es de 10,8 horas, en las Pruebas en el Sistema Transaccional es de 11,3 horas mientras que en el SAG es de 5,5 horas.

Concluimos que la elaboración de los informes de análisis de acuerdo a la complejidad, es mejor desarrollarlos utilizando el Sistema de Análisis Gerencial (SAG) mejor, porque los tiempos de construcción y pruebas son menores que los que se emplean en los Sistemas Transaccionales.

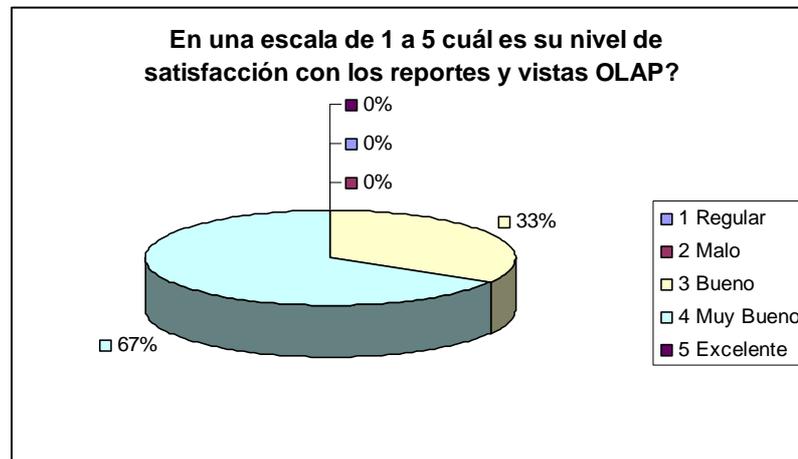
### c) Satisfacción del usuario estratégico

El universo tomado en consideración para analizar y obtener los resultados son los jefes departamentos de: Ventas, Compras e inventario, universo que ha sido considerado para demostrar el grado de satisfacción de los usuarios.

Las preguntas realizadas en la encuesta son las siguientes:

1. En una escala de 1 a 5 cuál es su nivel de satisfacción con los reportes y vistas OLAP?

Escala		Ni	Fr	%fr	Na	Fra
1	Regular	0,00	0,00	0,00	0,00	0,00
2	Malo	0,00	0,00	0,00	0,00	0,00
3	Bueno	1,00	0,33	33,33	1,00	0,33
4	Muy Bueno	2,00	0,67	66,67	3,00	1,00
5	Excelente	0,00	0,00	0,00	3,00	1,00
<b>Total</b>		<b>3</b>	<b>1,00</b>	<b>100,00</b>		

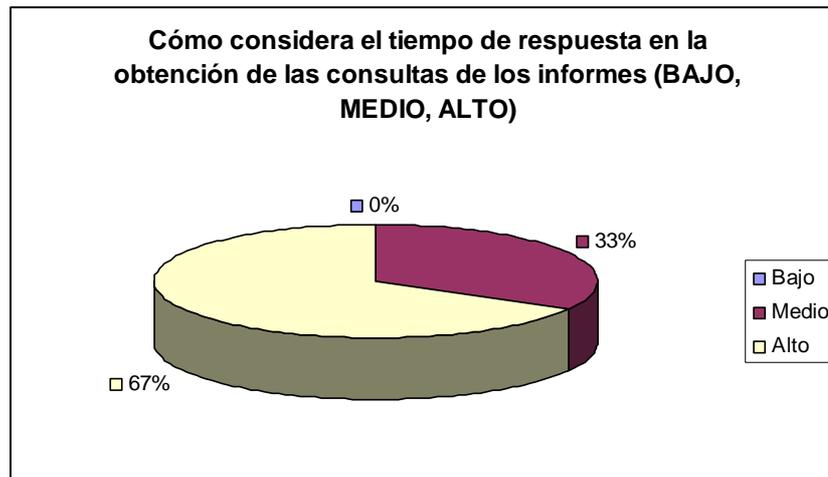


### ANÁLISIS

Dos personas de las tres a las que se realizó la entrevista informan que su nivel de satisfacción con los reportes y vistas OLAP es 4, correspondiente a Muy bueno, representando a un 67% de la población.

**2. Cómo considera el tiempo de respuesta en la obtención de las consultas de los informes (BAJO, MEDIO, ALTO)?**

Tiempo de respuesta	Ni	fr	%fr	na	Fra
Bajo	0,00	0,00	0,00	0,00	0,00
Medio	1,00	0,33	33,33	1,00	0,33
Alto	2,00	0,67	66,67	3,00	1,00
<b>Total</b>	<b>3</b>	<b>1,00</b>	<b>100,00</b>		

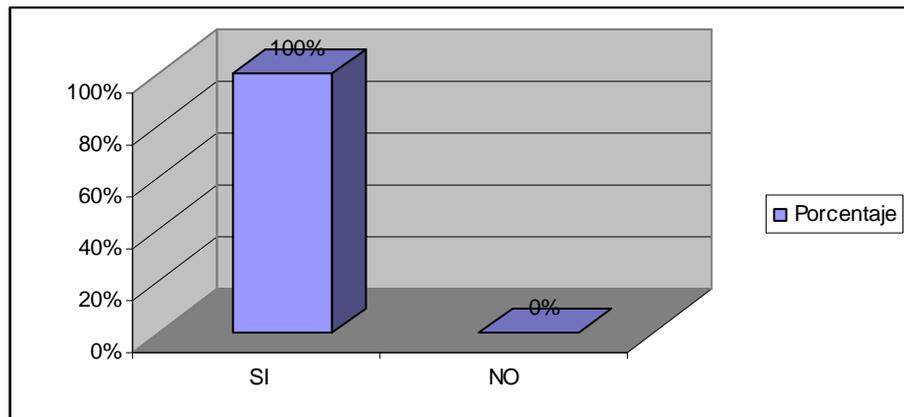


**ANÁLISIS**

Dos jefes departamentales responden que el tiempo de respuesta en la obtención de las consultas es Alto, es decir las consultas son ejecutadas rápidamente. Porcentaje que representa al 67%.

### 3. La información que obtiene al consultar los informes son los correctos?

DATOS	Ni	fr	%fr	na	Fra
SI	3	1	100%	3	100
NO	0	0	0%	3	100
<b>Total</b>	3	1	100%		

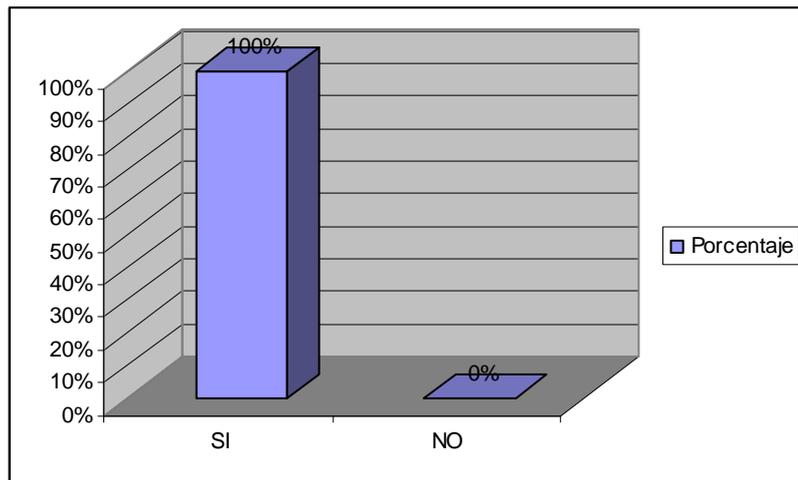


#### ANÁLISIS

Los tres jefes departamentales responden que la data que obtienen en los informes es la correcta, es decir la información concuerda con los registros contables. El porcentaje corresponde al 100%.

**4. Considera fácil el uso del Sistema de Análisis Gerencial (SAG)?**

DATOS	ni	fr	%fr	Na	Fra
SI	3	1	100	3	100
NO	0	0	0	3	100
<b>Total</b>	<b>3</b>	<b>1</b>	<b>100</b>		

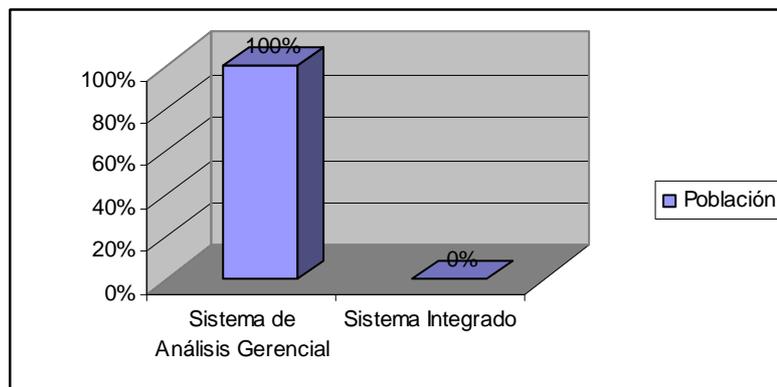


**ANÁLISIS**

Los jefes departamentales responden que el uso del Sistema de Análisis Gerencial es fácil, porcentaje que representa el 100%.

**5. Para obtener los informes ud prefiere el sistema Integrado o el Sistema de Análisis Gerencial(SAG)**

DATOS	Ni	fr	%fr	Na	Fra
Sistema de Análisis Gerencial	3	1	100	3	100
Sistema Integrado	0	0	0	3	100
<b>Total</b>	<b>3</b>	<b>1</b>	<b>100</b>		



**ANÁLISIS**

Los jefes departamentales responden que para obtener los informes, prefieren el Sistema de Análisis Gerencial (SAG), ya que les permite obtener sus propios informes de acuerdo a sus puntos de vista a analizar, porcentaje que representa el 100%.

## CONCLUSIONES

- ✚ La metodología planteada para implementar una solución Business Intelligence se ha aplicado en la empresa COHERVI S.A, obteniéndose una mejora en la elaboración de informes de análisis y satisfacción de las áreas departamentales, considerando tres variables: el Tiempo de elaboración de informes, Complejidad de elaboración de los informes y Satisfacción del usuario estratégico. Esto a su vez, permite ejecutar una planificación más adecuada con respecto, a: la de cantidad de compra de productos, proyecciones de ventas, analizar el movimiento de los distintos clientes y buscar nuevos mercados
- ✚ El tiempo promedio en la elaboración de informes de análisis de baja, media y alta complejidad utilizando el Sistema Transaccional es 7,3 – 36 y 11,3 horas respectivamente, este tiempo es mayor que el utilizado en la elaboración de informes en el Sistema de Análisis Gerencial, cuyo tiempo promedio es a su vez de 7,3 para los informes de alta complejidad, 10,8 para los de media y 5,5 horas para los de baja.
- ✚ Una vez que los usuarios han utilizado los informes analíticos de la Solución Bi implementada, se han pronunciado en un 100% de preferencia en relación a los reportes del sistema tradicional que han venido trabajando.
- ✚ El utilizar la metodología propuesta se redujeron los tiempos de elaboración del proyecto de inteligencia de negocio, por ser una metodología que utiliza técnicas y herramientas de fácil manejo en cada una de las fases, y cuyo repositorio trabaja con Data marts que cubren las necesidades de un grupo de trabajo o determinado departamento dentro de la organización.
- ✚ Los proyectos de Business Intelligence están orientados principalmente a usuarios de tipo gerencial, los que podrán acceder a información de forma fácil y actualizada, de esta manera lograrán contar la información fiable para la toma de decisiones oportunas en bienestar de la empresa.
- ✚ Luego de realizar un análisis de las herramientas tanto de tipo comercial y Open Source, se ha determinado en virtud de la necesidades específicas tanto técnicas y económicas, que la mejor herramienta para las necesidades de la empresa "COHERVI .S.A" es la orientada a open source.

## RECOMENDACIONES

- ✚ La Metodología para Desarrollo de Proyectos de Business Intelligence “MEDEPROBI” puede ser utilizada en cualquier proyecto de Business Intelligence, por lo que se recomienda su uso, trabaja en fases de tipo evolutivo e incremental, donde cada una utiliza técnicas, herramientas, recursos humanos, entradas y salidas.
- ✚ Se recomienda administrar y realizar un seguimiento de los proyectos de Business Intelligence porque los datos de los sistemas transaccionales pueden cambiar día a día y afectará a los Datamarts.
- ✚ La Suite de Pentaho es un proyecto desarrollado en JSP, y necesita establecer la conexión a la Base de Datos, por ello se recomienda antes de iniciar el desarrollo de reportes, cubos OLAP, vistas de negocio, definir el/los Datasources.
- ✚ Es fundamental involucrar a todas las personas a las cuales va enfocada el proyecto de inteligencia de negocios, además indicarles cuales son los objetivos que se puede lograr con este tipo de proyectos para no crear en ningún momento falsa expectativas.

## RESUMEN

La metodología propuesta "MEDEPROBI" permite construir proyectos de Business Intelligence de forma rápida y práctica, y se basa en las metodologías de Ralph Kimball, Crispdm y Métrica versión 3.

La metodología propuesta se ha utilizado para implementar Business Intelligence en la empresa "COHERVI S.A" ubicada en el cantón Milagro, provincia del Guayas, obteniéndose información íntegra, veraz y confiable de compra-venta de productos de consumo masivo a nivel local cantonal y provincial, que faciliten a las áreas departamentales y gerencia la toma de decisiones.

MEDEPROBI trabaja en fases de tipo evolutivo e incremental, donde cada una utiliza técnicas (UML, Diagramas de Gantt, entrevistas, modelado de datos, pruebas de caja negra, etc), herramientas (MS Microsoft Office, Visio, Rational Rose 2000, DBMS, Pentaho Community), recursos humanos, entradas(información, datos, entrevistas, etc) y salidas(informes, requerimientos, diseño,etc).

La solución Business Intelligence implementada a través de la metodología propuesta en la empresa "COHERVI S.A", denominada Sistema de Análisis Gerencial (SAG), utiliza el gestor de base de datos MySQL aplicando Datamarts dependientes, y el sistema de soporte de decisiones PENTAHO COMMUNITY.

## SUMMARY

The suggested "MEDEPROBI" methodology which is based on Ralph Kimball, Crispdm and Métrica v3 allows to construct BI projects in a fast and easy way.

The suggested "MEDEPROBI" methodology which is based on Ralph Kimball, Crispdm and Métrica v3 allows to construct BI projects in a fast and easy way.

MEDEPROBI suggested methodology has been used in implementing BI at "COHERVI S.A", an enterprise located in El Cantón Milagro, in Guayas province, resulting in successful integration of credible and reliable information about the buying and selling process of mass consumer oriented products at a local scale in Cantón and province that will lead to an easier decision making among the directors and other business areas.

MEDEPROBI has evolutive and incremental stages in which every one uses techniques (like UML, Gantt Charts, interviews, data modelers, black box tests, and others), set of tools (like MS Microsoft Office, Visio, Rational Rose 2000, DBMS, Pentaho Community), human resources, inputs (like information, data, interviews, etc) and outputs (reports, requirements, design, etc.)

Manager Analisis System (Sistema de Análisis Gerencial, SAG, in Spanish) was the solution implemented at "COHERVI S.A" which uses the suggested methodology and tools like MySQL database management systems through the construction of linked Datamarts and the PENTAHO COMMUNITY decision support system.

## GLOSARIO

**OLTP:** OLTP es la sigla en inglés de Procesamiento de Transacciones En Línea (OnLine Transaction Processing) es un tipo de sistemas que facilitan y administran aplicaciones transaccionales, usualmente para entrada de datos y recuperación y procesamiento de transacciones (gestor transaccional).

**OLAP:** OLAP es el acrónimo en inglés de procesamiento analítico en línea (*On-Line Analytical Processing*). Es una solución utilizada en el campo de la llamada Inteligencia empresarial (o *Business Intelligence*) cuyo objetivo es agilizar la consulta de grandes cantidades de datos.

**ETL:** ETL son las siglas en inglés de Extraer, Transformar y Cargar (*Extract, Transform and Load*). Es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, data mart, o data warehouse para analizar, o en otro sistema operacional para apoyar un proceso de negocio.

**RDBMS:** (Relational Database Management System o RDBMS - Sistema de Gestión de Base de Datos Relacional o SGBDR). Tipo de SGBD (o DBMS en inglés) para bases de datos relacionales (que emplea el modelo de datos); o sea, soporte de tablas relacionadas.

**AD HOC:** Ad hoc es una locución latina que significa literalmente «para esto». Generalmente se refiere a una solución elaborada específicamente para un problema o fin preciso y, por tanto, no es generalizable ni utilizable para otros propósitos. Se usa pues para referirse a algo que es adecuado sólo para un determinado fin. En sentido amplio, *ad hoc* puede traducirse como

**UML:** Lenguaje Unificado de Modelado (UML, por sus siglas en inglés, *Unified Modeling Language*) es el lenguaje de modelado de sistemas de software más conocido y utilizado en la actualidad; está respaldado por el OMG (Object Management Group). Es un lenguaje gráfico para visualizar, especificar, construir y documentar un sistema. UML ofrece un estándar para describir un "plano" del sistema (modelo), incluyendo aspectos conceptuales tales como procesos de negocio y funciones del sistema

**DATAWAREHOUSE:** Repositorio completo de datos de la empresa, donde se almacenan datos estratégicos, tácticos y operativos, el objetivo de obtener información estratégica y táctica.

**DATAMARTS:** Repositorio parcial de datos de la empresa, donde se almacenan datos tácticos y operativos, al objeto de obtener información táctica.

**JAVA:** Java es un lenguaje de programación orientado a objetos desarrollado por Sun Microsystems a principios de los años 90. El lenguaje en sí mismo toma mucha de su sintaxis de C y C++, pero tiene un modelo de objetos más simple y elimina herramientas de bajo nivel, que suelen inducir a muchos errores, como la manipulación directa de punteros o memoria

**METADATA:** la metadata o Meta Tags, como "Datos acerca de los datos" o "información acerca de la información" . A través de esta metodología es posible describir el contenido de un recurso de aprendizaje. Otra definición es: información sobre objetos web, comprensible por máquinas.

## BIBLIOGRAFÍA

### BIBLIOGRAFÍA GENERAL

1. Abukari, K. y Job, V. Business Intelligence in action. España: Proquest Editorial, 2003
2. Goodwin, C. Business Intelligence – Assault on the data mountain. New York: Proquest Editorial, 2003
3. Laudon Keneth C. y Laudon Jane P. *Sistemas de Información Gerencial*. México: Editorial Pearson, 2004
4. Martín, J. The Journal of Business Strategy. New York: ProQuest Editorial, 2006

### BIBLIOGRAFÍA EN INTERNET

#### **CONCEPTO GENERALES DE BUSINESS INTELLIGENCE**

-  [http://es.wikipedia.org/wiki/Inteligencia\\_empresarial](http://es.wikipedia.org/wiki/Inteligencia_empresarial)
-  <http://todobi.blogspot.com/2005/06/empresas-business-intelligence.html>
-  <http://www.monografias.com/trabajos14/bi/bi.shtml>
-  <http://www.monografias.com/trabajos29/sistema-business-intelligence/sistema-business-intelligence.shtml>  
2008-08-15

#### **DATAMARTS**

-  [http://es.wikipedia.org/wiki/Data\\_mart](http://es.wikipedia.org/wiki/Data_mart)
-  <http://www.pcanete.com.ar/data-mart/>  
2008-09-08

#### **DATAWARE HOUSE**

-  [http://es.wikipedia.org/wiki/Almac%C3%A9n\\_de\\_datos](http://es.wikipedia.org/wiki/Almac%C3%A9n_de_datos)
-  [http://books.google.com.ec/books?id=2OCbq8Azdm8C&dq=DATAWAREHOUSE&printsec=frontcover&source=bl&ots=T6IM9qU9EL&sig=aGliW06F\\_U73jtT561-glpA\\_ggg&hl=es&ei=9VeuSsSZNlu\\_IAfe5dXhBg&sa=X&oi=book\\_result&ct=result&resnum=4#v=onepage&q=&f=false](http://books.google.com.ec/books?id=2OCbq8Azdm8C&dq=DATAWAREHOUSE&printsec=frontcover&source=bl&ots=T6IM9qU9EL&sig=aGliW06F_U73jtT561-glpA_ggg&hl=es&ei=9VeuSsSZNlu_IAfe5dXhBg&sa=X&oi=book_result&ct=result&resnum=4#v=onepage&q=&f=false)
-  <http://www.atinternet.com/es/Recursos/Desarrollostecnologicos/DatawarehouseyCubeOLAP.aspx>

 <http://haciendocubos.com/2007/08/01/el-modelo-estrella/>

 [http://geeks.ms/blogs/gelexgaray/archive/2006/08/23/\\_BF00\\_Estrella-o-copo-de-nieve\\_3F00\\_.aspx](http://geeks.ms/blogs/gelexgaray/archive/2006/08/23/_BF00_Estrella-o-copo-de-nieve_3F00_.aspx)

2008-10-15

## **DESCARGAS HERRAMIENTAS**

### **BIRT**

 <http://download.eclipse.org/birt/downloads/>

2008-12-05

### **MYSQL**

 <http://dev.mysql.com/downloads/>

 <http://dev.mysql.com/downloads/gui-tools/5.0.html>

 <http://dev.mysql.com/downloads/connector/>

2008-12-05

### **JAVA**

 <http://java.sun.com/javase/downloads/index.jsp>

2008-12-05

### **PENTAHO**

 <http://sourceforge.net/projects/pentaho/files/>

2008-12-05

### **FOROS DE PENTAHO**

 [www.redopenbi.com](http://www.redopenbi.com)

 <http://forums.pentaho.org/>

2009-01-01

### **TECNOLOGÍA DE PENTAHO**

 [www.pentaho.com](http://www.pentaho.com)

 [www.wiki.pentaho.com](http://www.wiki.pentaho.com)

2009-01-01

# **ANEXOS**

## **ANEXO 1**

### **MDX**

Hasta ahora se ha usado una notación para referirse a un miembro en una dimensión. Se había dicho que primero se colocaba la dimensión, entre otros. Todo esto pertenece a un lenguaje llamado MDX.

Las siglas MDX significa Multi Dimensional Expressions. Es un lenguaje que sirve para hacer consultas en una base de datos multidimensional, análogamente a como lo haría el lenguaje SQL en las bases de datos comunes. Es la forma de expresar el análisis de un cubo OLAP.

Desde que la empresa Microsoft entro al mundo de la inteligencia de negocios, en el año de 1998, con una versión de servicios OLAP en la aplicación SQL Server 7.0, el lenguaje MDX se ha venido convirtiendo en una lengua estándar para este tipo de consultas. Es verdad que aunque fue creado por compañías propietarias, el lenguaje MDX ha sido adoptado también por el mundo del Software libre, para las aplicaciones OLAP existentes.

Este lenguaje es útil para muchas cosas, desde aspectos de seguridad en el cubo, hasta el diseño de la navegación hacia adentro y fuera del cubo. En definitiva este lenguaje fue creado específicamente para el manejo de cubos OLAP. Una de las grandes características de este lenguaje es el de permitir el manejo de las agregaciones y los resultados precalculados, sin necesidad de que aparezcan físicamente establecidos en las bases de datos.

Siguiendo con la notación del lenguaje, la forma natural de llamar aun miembro con un jerarquía suficiente podría ser como este: [lugar].[Riobamba], [lugar].[quito], [lugar].[cuenca]. Esta sería la forma completa para referirse al barrio Licán de la ciudad de Riobamba. Aunque muchas veces es posible usar expresiones más cortas,

como [ciudad].[Licán], esto se puede hacer, pero solo bajo ciertas condiciones de nombrado y ciertos datos.

Por ejemplo, para el caso de la dimensión tiempo, es posible tener una expresión como “[tiempo].[octubre]”. Pero como se podría saber si el octubre al que se hace referencia es al del año 2001 o 2002 o 2003?. Una forma de solucionar este inconveniente es el de hacer unas convenciones de nombramiento. Por ejemplo, si el mimbro no se llamara octubre sino “octubre 2001”, ya se podría identificar con una corta expresión, es decir que se puede definir como: [tiempo].[octubre 2001], en vez de : [tiempo].[all]. [2001].[octubre].

La sintaxis MDX no es solo para obtener datos de las bases multidimensionales, también provee una funcionalidad extra para el manejo de sets, manejo de miembros calculados y la escritura de información acerca de dimensiones y celdas. Para finalizar esta definición de MDX, se examina la siguiente consulta MDX, definida de a siguiente manera.

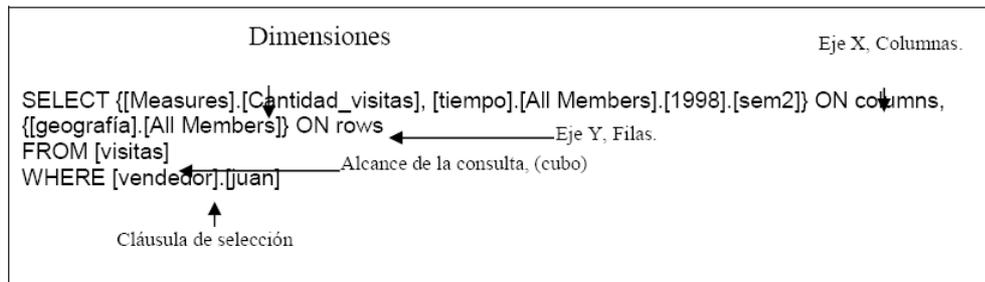
```
SELECT {[Measures]. [Cantidad_visitas], [tiempo]. [All Members]. [1998]. [sem2]} ON  
columns, {[geografia].[All Members]} ON rows  
FROM [visitas]  
WHERE [vendedor]. [juan]
```

Esta sencilla consulta mdx presenta la misma estructura que una consulta SQL es decir, tiene un select, un from y un where. Es importante notar que en el lenguaje mdx las medidas también son interpretadas como dimensiones. En este ejemplo se está consultando el número de visitas hechas en el segundo semestre de 1998 en todas las zonas, veredas y municipios por el vendedor Juan. Como se identifica en la sintaxis se da la opción de escoger como comparar las dimensiones, en filas o en columnas. [visitas] es el nombre del cubo.

## Conceptos avanzados con mdx

En la explicación anterior se revisaron los conceptos básicos de cómo construir una MDX, en esta parte se explicarán conceptos mas avanzados de MDX.

La estructura detallada de una consulta MDX se expone en el siguiente diagrama.



**Figura. Estructura de una consulta MDX**

Existe una similitud entre las consultas MDX Y SQL, debido a que comparten sintaxis como lo son las cláusulas SELECT, FROM Y WHERE. Sin embargo existen ciertos elementos que se distancian de SQL para dar soporte a las necesidades multidimensionales.

En MDX se identifican varios miembros, que cumplen un rol determinado por su posición en la jerarquía.

### **Children.**

La primera función es Children la cual, identifica el miembro o los miembros de un nivel por debajo del nivel inicial. Para usar la función Children se debe especificar el miembro, para que la función retorne todos los niveles por debajo de ella.

### **Count**

La función Count retorna el número de ítems de un cubo, por ejemplo las dimensiones, las tuplas, las celdas, niveles en una dimensión o una jerarquía.

### **Avg**

La función Avg retorna el promedio de un número total de miembros. En la siguiente consulta de ejemplo se utiliza esta función.

### **Max**

La función max retorna el máximo valor de un número total de miembros. Max (-Set- [Numeric Expression]).

### **Min**

La función min retorna un mínimo valor de un número total de miembros. Min ({Set}, [Numeric Expression]).

### **Crossjoin**

Frecuentemente en MDX, los resultados de una consulta deberán necesitar más miembros de más de una dimensión.

Crossjoin ({Set1},{Set2})

Esta función retorna el producto cartesiano de distintos miembros de dos sets, que retorna todas las posibles combinaciones de esos miembros.

## ANEXO 2

### DICCIONARIO DE DATOS

#### base de datos dwh\_cohervi

Servidor: localhost Base de datos: dwh\_cohervi

#### dwh\_caducados

Campo	Tipo	Nulo	Predeterminado
CODI_ARTIC	int(11)	No	0
NOMB_GENER	varchar(50)	No	
STAT_REGIS	varchar(1)	No	
F_EMISION	date	No	
F_VENCE	date	No	

#### dwh\_clientes

Campo	Tipo	Nulo	Predeterminado
ID	int(11)	No	
CODCLT	int(11)	No	0
CODVEN	int(11)	No	
ZONACL	varchar(5)	No	
NOMCLT	varchar(40)	No	
STATUS	varchar(1)	No	
DIRCLT	varchar(40)	No	
MONTCR	float	No	
PLZODIA	int(11)	No	
SALDO_P	float	No	

#### dwh\_compras

Campo	Tipo	Nulo	Predeterminado
ID	int(11)	No	
CODPVD	int(11)	No	
CODPRD	int(11)	No	
FECTRN	date	No	
CANPRD	float	No	
CSTUNT	float	No	

#### dwh\_existencia

Campo	Tipo	Nulo	Predeterminado
CODI_ARTIC	int(11)	No	
CANT_EXIST	int(11)	No	

#### dwh\_kardex

Campo	Tipo	Nulo	Predeterminado
ID	int(11)	No	
TIPDOC	varchar(2)	No	
NUMDOC	int(11)	No	
CODPRD	int(11)	No	
FECTRN	date	No	
CANPRD	int(11)	Sí	0
INGEGR	varchar(1)	No	

#### dwh\_movimientos

Campo	Tipo	Nulo	Predeterminado
ID	int(11)	No	
TIPDOC	varchar(2)	Sí	ND
NUMDOC	int(10)	No	0
FECTRN	date	No	
CODCLI	int(10)	No	
VALDOC	float	Sí	0
CODUSR	varchar(3)	No	

### dwh\_productos

Campo	Tipo	Nulo	Predeterminado	
ID	int(11)	No		
CODL_ARTIC	int(11)	No		
NOMB_GENER	varchar(25)	No		
CODL_PROVE	int(11)	No		
PREC_PUNID	float	No		
PREC_PDOCE	float	No		
PREC_PCJBT	float	No		
COST_PROD	float	No		
STAT_REGIS	varchar(2)	No		

### dwh\_proveedores

Campo	Tipo	Nulo	Predeterminado	
ID	int(11)	No		
CODIGO	int(11)	No		
NOMBRE	varchar(40)	No		
DIASCRED	int(11)	No		
DIREC1	varchar(35)	No		
VENDED1	varchar(35)	No		
TELEF1	varchar(9)	No		
CIUDAD	varchar(15)	No		
CONESP	varchar(1)	No		
PERNAT	varchar(1)	No		

### dwh\_tipcomprobantes

Campo	Tipo	Nulo	Predeterminado	
ID	int(4)	No		
CODL_COMPR	varchar(2)	No		
NOMB_DOCUM	varchar(30)	No		

### dwh\_tipovtas

Campo	Tipo	Nulo	Predeterminado	
ID	int(11)	No		
PTOVTA	varchar(2)	No		
NOMVTA	varchar(20)	No		

### dwh\_vendedores

Campo	Tipo	Nulo	Predeterminado	
ID	int(10)	No		
CODVEN	int(10)	No		
NOMVEN	varchar(40)	No		

### dwh\_zonas

Campo	Tipo	Nulo	Predeterminado	
IDZONA	int(11)	No		
CODZON	varchar(5)	No		
NOMZON	varchar(40)	No		

## Base de Datos ddm\_cohervi

Servidor: localhost ▶ Base de datos: ddm\_cohervi

### authorities

Campo	Tipo	Nulo	Predeterminado
AUTHORITY	varchar(50)	No	
DESCRIPTION	varchar(100)	Sí	NULL

### dim\_clientes

Campo	Tipo	Nulo	Predeterminado
id	int(4)	No	
codigo	int(4)	No	0
nombres	varchar(40)	No	ND
estado	varchar(1)	No	-
direccion	varchar(40)	No	ND
montocredito	float	No	0
plazodias	int(11)	No	0
saldopendiente	float	No	0
version	int(2)	No	
f_inicio	date	No	2000-01-01
f_fin	date	No	2050-01-01
vendedor	varchar(5)	No	ND
zona	varchar(5)	No	ND

### dim\_productos

Campo	Tipo	Nulo	Predeterminado
id	int(4)	No	
codigo	int(4)	No	0
nombre	varchar(25)	No	ND
p_unidad	float	No	0
p_docena	float	No	0
p_cajas	float	No	0
c_compra	float	No	0
estado	varchar(2)	No	ND
version	int(11)	No	
f_inicio	date	No	2000-01-01
f_fin	date	No	2050-12-31

### dim\_proveedores

Campo	Tipo	Nulo	Predeterminado
id	int(4)	No	
codigo	int(4)	No	0
nombre	varchar(40)	No	ND
vendedor	varchar(25)	No	ND
diascredito	int(11)	No	0
direccion	varchar(35)	No	ND
telefono	varchar(9)	No	0
ciudad	varchar(15)	No	ND
conesp	char(1)	No	-
permat	char(1)	No	-
version	int(11)	No	0
f_inicio	date	No	2000-01-01
f_fin	date	No	2050-12-31

### dim\_tiempo

Campo	Tipo	Nulo	Predeterminado	
id	int(11)	No		
FECTRAN	date	No	2000-01-01	
fecha_str	varchar(10)	No	2000-01-01	
año	int(11)	No	0	
semestre	varchar(15)	No	ND	
trimestre	varchar(15)	No	ND	
mes	int(11)	No	0	
mesnombre	varchar(30)	No	ND	
version	int(2)	No		
f_inicio	date	No	2000-01-01	
f_fin	date	No	2050-12-31	

### dim\_tipcomprobantes

Campo	Tipo	Nulo	Predeterminado	
id	int(4)	No		
codigo	varchar(2)	No	ND	
nombre	varchar(30)	No	ND	
version	int(4)	No		
f_inicio	date	No	2000-01-01	
f_fin	date	No	2050-12-31	

### fact\_caducados

Campo	Tipo	Nulo	Predeterminado	
idproductos	int(11)	No		
diasvence	int(11)	No		
diascaducidad	int(11)	No		
fecha_carga	datetime	No		

### fact\_compras

Campo	Tipo	Nulo	Predeterminado	
idproveedor	int(4)	No		
idproducto	int(4)	No		
idtiempo	int(4)	No		
cantidad	float	No		
montocompras	float	No		
fecha_carga	date	No		

### fact\_ventas

Campo	Tipo	Nulo	Predeterminado	
idtiopvta	int(4)	No		
idcliente	int(6)	No		
idtiempo	int(4)	No		
usuario	varchar(3)	No		
importe	float	No		
fecha_carga	date	No	2000-01-01	

### granted\_authorities

Campo	Tipo	Nulo	Predeterminado	
USERNAME	varchar(50)	No		
AUTHORITY	varchar(50)	No		

### users

Campo	Tipo	Nulo	Predeterminado	
USERNAME	varchar(50)	No		
PASSWORD	varchar(50)	No		
DESCRIPTION	varchar(100)	Sí	NULL	
ENABLED	bit(1)	No		



 El negocio para su Negocio	<b>SAG</b>	COHERVI S.A
	Sistema de Análisis Gerencial	

# MANUAL DE USUARIO

## SAG

Sistema de Análisis Gerencial

	<b>SAG</b>	COHERVI S.A
	Sistema de Análisis Gerencial	

## 1. INTRODUCCIÓN

Este manual es desarrollado para los jefes departamentales y administrador del área de sistemas de la empresa COHERVI S.A, usando la herramienta Open Source Pentaho.

El Sistema de Análisis Gerencial (SAG), permitirá el acceso de los usuarios del departamento de: Ventas, Inventario y Compras. Para ingresar al sistema, el usuario deberá autenticarse ingresando su login y password, una vez validado su autenticación podrá acceder a las carpetas de Reportes y Vistas OLAP.

Dentro de la carpeta de Reportes tiene acceso a reportes estáticos y Ad Hoc, dentro de la carpeta “Vistas OLAP” podrá manipular cubos que le permitirán tener varios enfoques del negocio y mayor flexibilidad de manipularlos en tiempo real a través de los navegadores de Internet, según las necesidades específicas de cada usuario, ya sea a nivel de dimensiones (Por ejemplo: tiempo, proveedores, productos, ciudad, zonas, vendedores) o medidas (valores o cálculos, por ejemplo: cantidad, importe), obteniendo así la información relevante que se desea para la toma de decisiones.

Las dimensiones de las vistas OLAP permiten agrupar o desplegar de diferentes maneras la información requerida gracias a los botones del menú, además permiten exportar a MS Excel o PDF.

Cada una de las Vistas OLAP, tiene la posibilidad de representar en distintos tipos de gráficos la información obtenida, los cuales tienen la opción de personalizar su diseño para un mejor entendimiento.

A través de una de las opciones que proporciona Pentaho, el usuario podrá crear sus propios reportes en tiempo real (Reportes Ad Hoc), seleccionar sus plantillas del reporte, definir el tipo de formato en el que desea visualizar y guardarlos.

Pentaho posee una sección de administración para controlar el adecuado uso del sistema, usuarios, grupos y data sources, a la cual tendrá acceso el usuario que posea permisos de “Administrador”.

	<b>SAG</b>	COHERVI S.A
	Sistema de Análisis Gerencial	

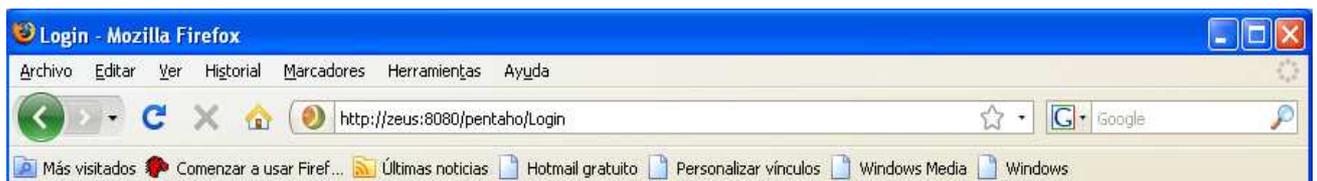
## 2. INSTRUCCIONES

### 1.1. 2.1 Conexión al Sistema.

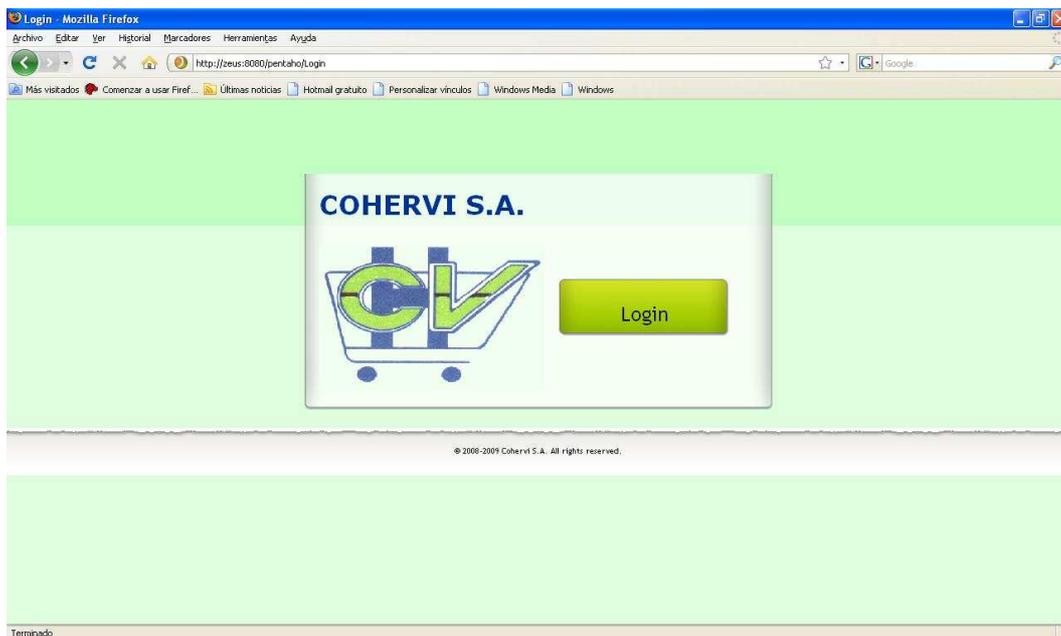
Para conectarse al Sistema de Análisis Gerencial (SAG) se debe seguir los siguientes pasos:

En el navegador de Internet (por ejemplo: Internet Explorer, Mozilla Firefox, Opera) ingrese la siguiente dirección URL:

<http://zeus:8080/pentaho>



A continuación se despliega la siguiente página de inicio, donde deberá seleccionar el botón **Login**.



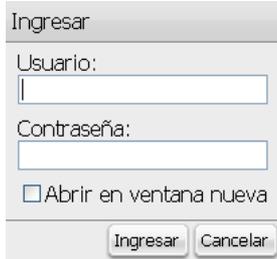
	<b>SAG</b>	COHERVI S.A
	Sistema de Análisis Gerencial	

## 1.2. 2.2 Ingreso al Sistema.

En esta ventana el usuario deberá ingresar su login y password para manipular e interactuar con los reportes estáticos, Ad-Hoc y Vistas OLAP para su respectivo análisis, ingresando en los casilleros sus datos de usuario asignados.

Por ejemplo **Usuario:** v\_olga

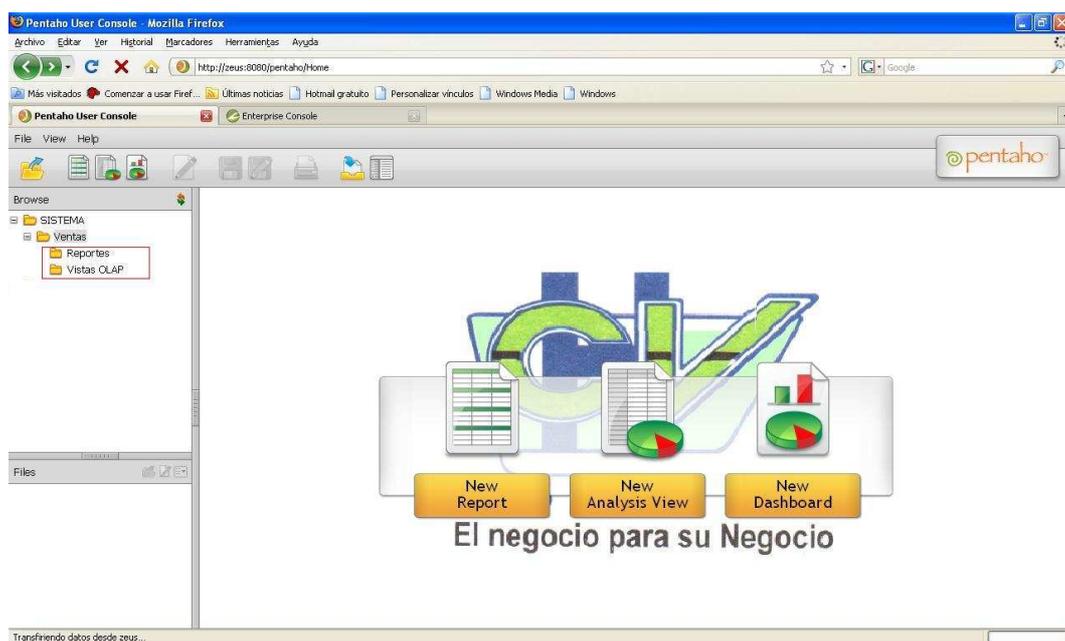
**Contraseña:** v\_olga



## 1.3.

## 1.4. 2.3 Página Principal

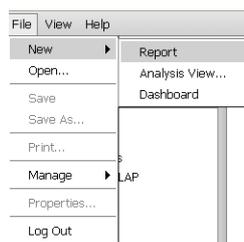
En esta pantalla, el usuario v\_olga que pertenece al departamento de Ventas, tendrá acceso a las carpetas Reportes y Vistas OLAP, además dispones de una barra de menú con las opciones de File, View y Help.



A continuación se describirá los componentes de la barra de menú



 **FILE:** Tiene varias opciones tales como:

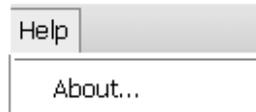


- **New:** Se utiliza para crear un Nuevo Reporte (Report) o Vista OLAP (Análisis View).
- **Open:** Permite abrir el repositorio de soluciones del sistema
- **Save:** Se utiliza para guardar el reporte o la vista OLAP
- **Save As:** esta opción permite asignar un nombre al archivo creado y seleccionar el repositorio donde deseo guardar.
- **Print:** Permite imprimir el reporte o Vista OLAP
- **Manage:** se utiliza para editar
- **Properties:** Permite conocer las propiedades de los documentos (reportes y análisis OLAP).
- **Log Out:** Se utiliza para salir del sistema

 **VIEW:** Esta opción permite personalizar las propiedades de la ventana, si deseamos ver en forma de Browser, Wokspace, Use Descriptions for Tooltips y Refresh.



🚧 **HELP:** tiene la opción de acceder al **About**

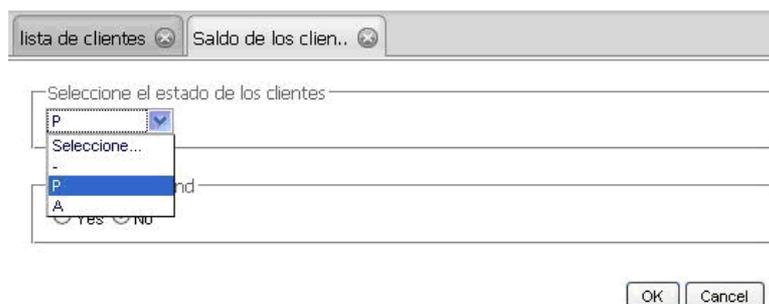


### 1.5. 2.4 Manejo de Reportes Estáticos y Ad-Hoc

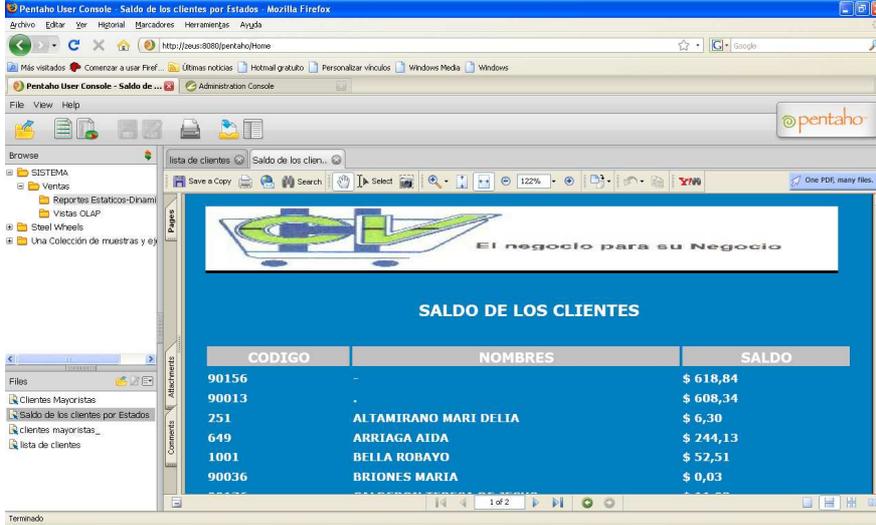
Una vez que el usuario este en la pantalla principal, tiene acceso a las Carpetas ReportesEstaticos – Dinamicos y Vistas OLAP, seleccione la Carpeta Reportes Estaticos-Dinamicos y fíjese en la parte inferior nos indica una lista de reportes.



Seleccionar uno de los reportes ejemplo Saldo de Clientes por estado, se presentará la siguiente pantalla donde deberá escoger el estado del cliente que puede ser Activo o Pasivo como indica en la siguiente figura, a continuación elegir el botón OK.



En esta pantalla tenemos el resultado de filtrar los clientes por el estado Pasivo



CODIGO	NOMBRES	SALDO
90156	-	\$ 618,84
90013	-	\$ 608,34
251	ALTAMIRANO MARI DELIA	\$ 6,30
649	ARRIAGA AIDA	\$ 244,13
1001	BELLA ROBAYO	\$ 57,51
90036	BRIONES MARIA	\$ 0,03

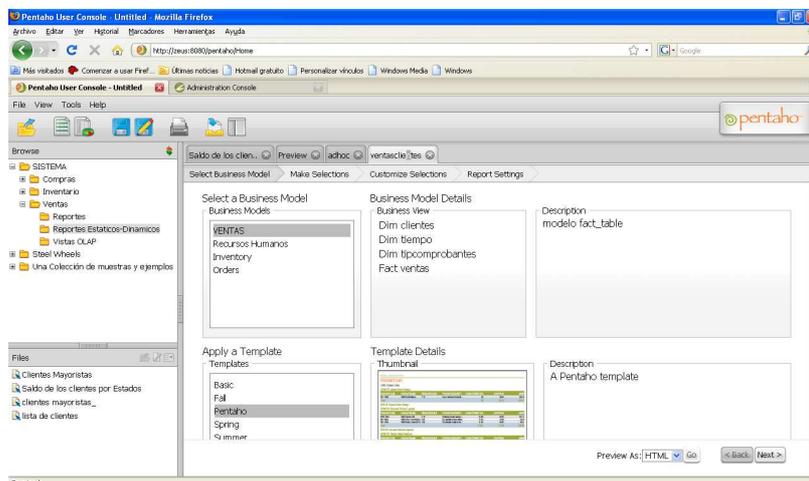
El usuario podrá crear sus propios reportes Ad –Hoc de la siguiente manera:

En la barra de herramientas seleccionar el icono que está encerrado en círculo.

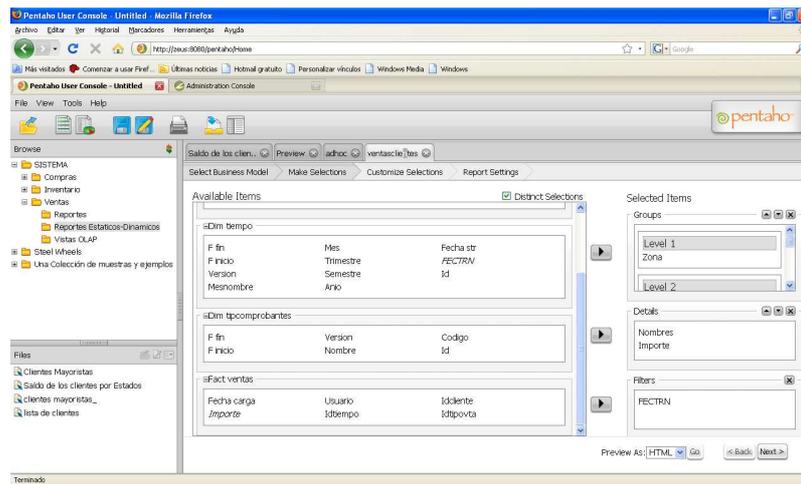


Para crear un reporte Ad Hoc deberá realizar cuatro pasos como indica el siguiente Wizard.

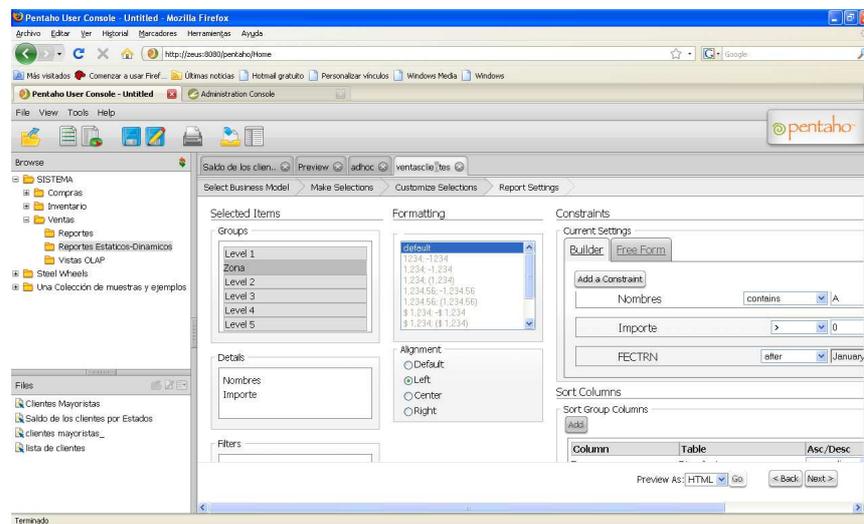
1. **Select a Business Model:** en este paso debe escoger el modelo de negocios, nuestro modelo se llama VENTAS, elegir el tipo de plantilla para el reporte, el formato para vitalizarlo, a continuación seleccione el botón Next.



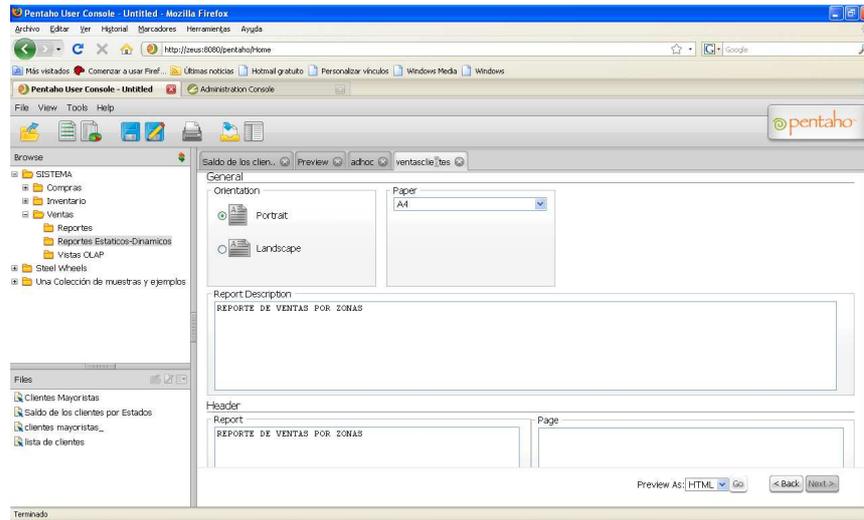
2. **Make selector:** en esta paso visualizamos las tablas con sus respectivos campos del modelo de negocio seleccionado en el paso anterior, escoger los campos que se utilizarán para: agrupar, mostrar en el reporte y campos por los que se realizarán los filtros, seguidamente escoger el botón Next.



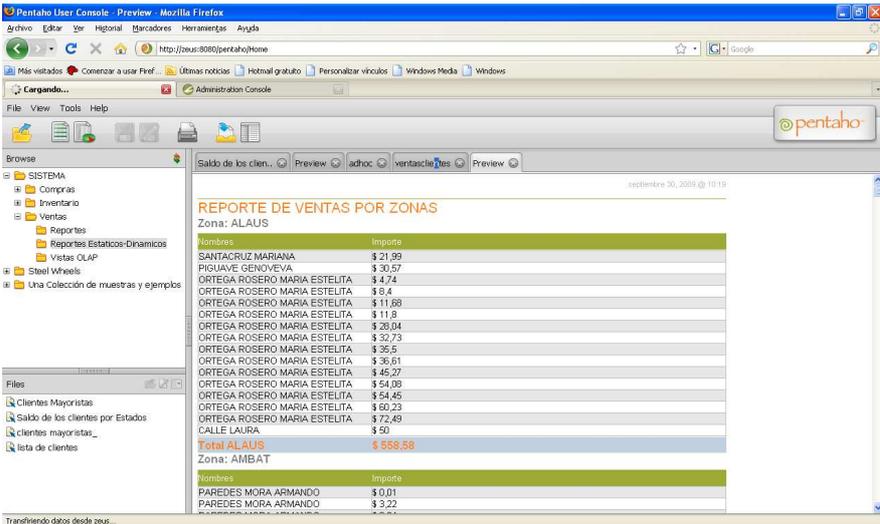
3. **Customize selection:** En este paso se define las condiciones para ciertos campos por ejemplos, el importe debe ser mayor a 0, los datos deben ser mostrados a partir del 01 de enero del 2008, además permite establecer sumalizaciones, promedios, contar y otras funciones, dar formato de número, ordenar las columnas de forma ascendente y descendente, alinear los campos, etc, seguidamente seleccionar el botón Next.



4. **Reports Setting:** En esta paso se debe configurar la página, escoger tipo de papel, dar una pequeña descripción del reporte, asignar un título y pie de página y número de página.



Se debe escoger el botón Preview, para visualizar el reporte creado y constatar que esté, de la forma que lo queremos obtener, en caso de no ser así, regrese al reporte y pulse el botón Back para retroceder y modificar uno de los pasos.

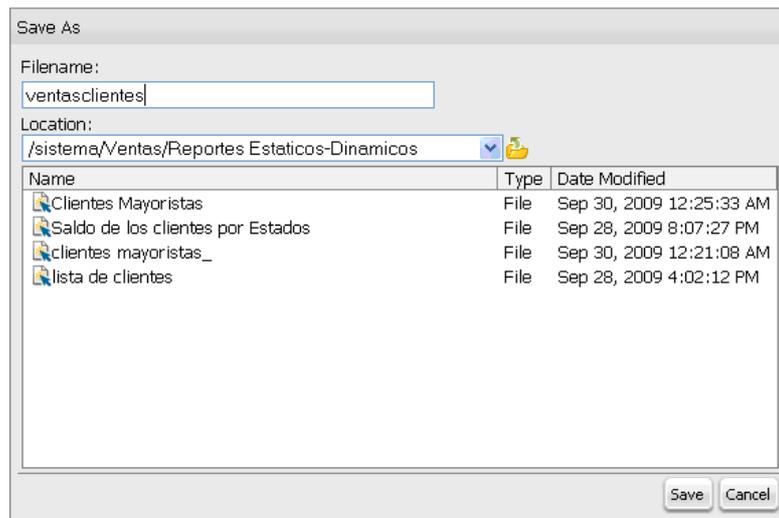


REPORTE DE VENTAS POR ZONAS	
Zona: ALAUS	
Nombres	Importe
SANTACRUZ MARIANA	\$ 21,99
PIGUAVE GENOVEVA	\$ 30,57
ORTEGA ROSERO MARIA ESTELITA	\$ 4,74
ORTEGA ROSERO MARIA ESTELITA	\$ 6,4
ORTEGA ROSERO MARIA ESTELITA	\$ 11,68
ORTEGA ROSERO MARIA ESTELITA	\$ 11,8
ORTEGA ROSERO MARIA ESTELITA	\$ 26,04
ORTEGA ROSERO MARIA ESTELITA	\$ 32,73
ORTEGA ROSERO MARIA ESTELITA	\$ 36,5
ORTEGA ROSERO MARIA ESTELITA	\$ 36,61
ORTEGA ROSERO MARIA ESTELITA	\$ 46,27
ORTEGA ROSERO MARIA ESTELITA	\$ 54,08
ORTEGA ROSERO MARIA ESTELITA	\$ 54,45
ORTEGA ROSERO MARIA ESTELITA	\$ 60,23
ORTEGA ROSERO MARIA ESTELITA	\$ 72,49
CALLE LAURA	\$ 50
<b>Total ALAUS</b>	<b>\$ 558,58</b>
Zona: AMBAT	
Nombres	Importe
PAREDES MORA ARMANDO	\$ 0,01
PAREDES MORA ARMANDO	\$ 3,22

A continuación escoger de la barra de menú el icono que está encerrado en círculo



Donde tenemos la siguiente pantalla en la que se asignará un nombre al reporte ejemplo ventasclientes, seguidamente escoger el botón Save.



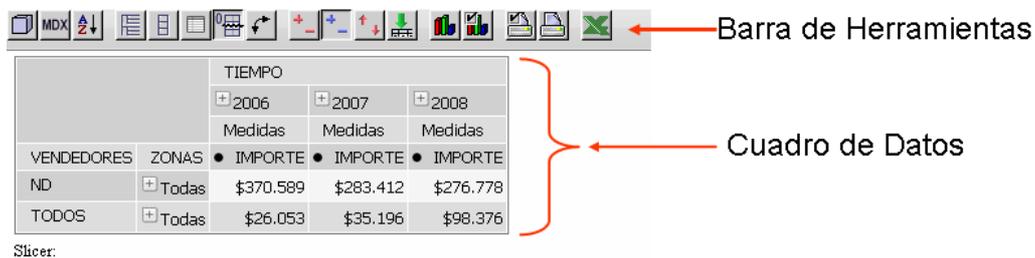
Para volverlo a utilizar ir al panel de lista de reportes, escoger el reporte y visualizarlo.

1.6.

1.7. **2.4 Manejo de Vistas OLAP**

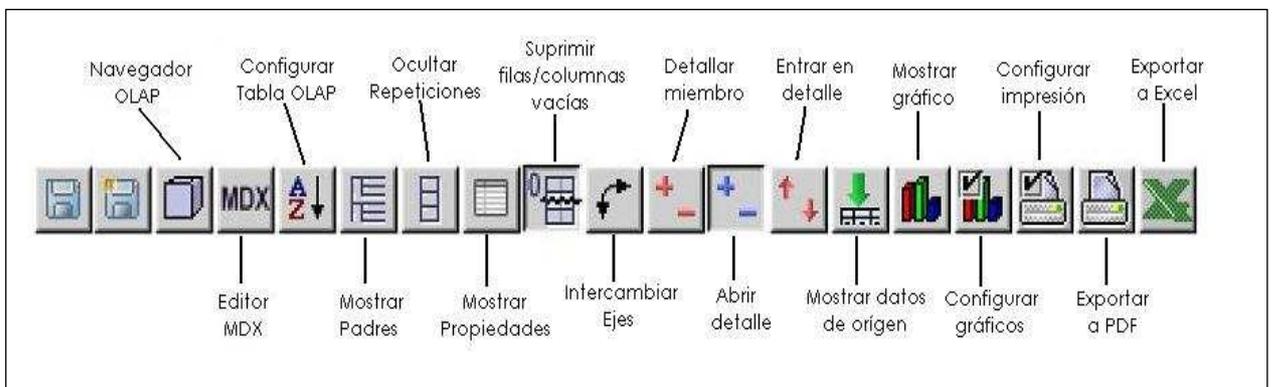
En el siguiente ejemplo se ilustra el uso y navegación de las vistas, se tomará como ejemplo el cubo de Ventas.

Una vez que nos encontramos dentro de la ventana de Home, escogemos la carpeta Ventas, luego Vistas OLAP, pulsamos la vista Ventas x zona x vend, a continuación se despliega la siguiente pantalla, en donde se podrá visualizar el reporte.



**2.4.1. Elementos de la Barra de Herramientas.**

Se puede construir una vista personalizada en la cual la vista muestre diferentes enfoques de análisis según las necesidades específicas, en este caso queremos analizar únicamente el importe de las ventas, para lo cual a continuación se realizará una explicación de las funcionalidades disponibles en la barra de herramientas para lograr dicho objetivo.



	<b>SAG</b>	COHERVI S.A
	Sistema de Análisis Gerencial	

La siguiente tabla describe cada uno de las opciones:

Icono	Nombre	Descripción
	Navegador OLAP	Se utiliza para definir el esquema general de su consulta, seleccionando las dimensiones para mostrar como filas o columnas y los miembros que se utilizarán como filtros, así mismo funciona con las medidas.
	Editor MDX	El análisis de Pentaho (Mondrian) utiliza el lenguaje MDX para la definición de consultas multidimensionales de datos. El editor MDX despliega un cuadro de diálogo que le permite visualizar y/o editar la consulta actual MDX del reporte.
	Configurar tabla OLAP	Este botón ofrece la posibilidad de fijar el tipo de visualización y el tipo de orden de los datos del reporte, además permite aplicar filtros hacia arriba o hacia abajo basados en rangos, y mostrar u ocultar las propiedades de miembro.
	Mostrar padres	Esta opción le proporciona la capacidad de mostrar el o los padres de cada miembro (Jerarquía de la dimensión) de una consulta específica en su propia fila o columna.
	Ocultar repeticiones	Se utiliza para impedir la visualización de cabeceras repetidas cuando hay múltiples dimensiones anidadas en un acceso común.
	Mostrar propiedades	Permite desplegar las propiedades del miembro en el reporte.
	Suprimir filas/columnas vacías	Esta opción oculta o muestra las filas que poseen datos con valor cero en el reporte.
	Intercambiar ejes	Esta es una de las más potentes características de la vista de análisis de Pentaho. Esta provee la capacidad de girar o transportar (pívor) los datos del reporte entre sus ejes, intercambiando dimensiones/miembros de las filas con las dimensiones/miembros de las columnas y manteniendo el mismo enfoque de análisis.
<p>A continuación se presentan 3 botones de detalle, con los cuales se puede obtener diferentes modos para detallar la información del reporte en la vista de análisis y así obtener diferentes comportamientos en nuestras consultas.</p>		
	Detallar miembro	En este modo, cualquier instancia de miembros que sean pulsadas serán expandidas (varias instancias de un miembro se muestran cuando

	<b>SAG</b>	COHERVI S.A
	Sistema de Análisis Gerencial	

		usted tiene dimensiones anidadas en un eje).
	Abrir detalle	En este modo, sólo la instancia del miembro que es pulsada será expandida, aunque haya otras instancias del mismo miembro.
	Entrar en detalle	En este modo, el miembro pulsado será reemplazado por sus hijos, aislando el análisis a esa sección específica.
	Mostrar datos de origen	En el análisis OLAP, los usuarios generalmente observan un nivel de información resumida que se ha agregado de los hechos detallados. Esta opción permite describir un valor específico que sea pulsado, es decir muestra el detalle de los hechos o registros individuales que resultan en el indicador resumido de dicha celda.
	Mostrar gráfico	Se utiliza para generar un gráfico que represente la consulta actual del reporte.
	Configurar gráficos	Este botón abre el cuadro de diálogo "Propiedades de Gráfico", el cual proporciona la capacidad de configurar diversos gráficos con propiedades como: tipo de gráfico, formato, tamaño, color, títulos, etc.
	Configurar impresión	Este botón abre el cuadro de diálogo "Propiedades de Impresión", el cual proporciona la capacidad para definir ajustes relacionados con la impresión de un reporte como el título, orientación de la página, tamaño de papel, y si desea o no imprimir el gráfico en una página independiente. Los ajustes se almacenan para la sesión actual del usuario.
	Exportar a PDF	Se utiliza para generar una versión en formato PDF (exportar a PDF) de la vista de análisis del reporte actual, basado en los ajustes de impresión.
	Exportar a Excel	Se utiliza para iniciar la herramienta MS Excel y generar una versión en formato .xls (exportar a MS Excel) de la vista de análisis del reporte actual, basado en los ajustes de impresión.

Una vez que conocemos la funcionalidad y uso de cada una de las opciones de la barra de herramientas, vamos a continuar con el ejemplo del manejo de vista, en el cual queremos analizar únicamente el importe en el tiempo en las distintas zonas y por los diferentes proveedores.

Para esto pulsar el botón “Navegador OLAP” en donde seleccionaremos la dimensión “TIEMPO Y VENDEDORES” que se encuentra filtrada y la situamos en las filas pulsando el botón respectivo  , después de eso pulsamos la sección de Medidas y dejamos seleccionado solamente la medida “IMPORTE” tal como se muestra en la imagen.



Pulsar el botón aceptar y a continuación el botón aplicar, y los cambios a la consulta del reporte han sido aplicados correctamente y se desplegará la siguiente vista de análisis.



		Medidas
		● IMPORTE
+ 2006	+ Todas	\$396.642
+ 2007	+ Todas	\$318.608
+ 2008	+ Todas	\$375.154

Slicer:

En este punto vamos a empezar a navegar dentro del reporte hasta encontrar la información que necesitamos utilizando las opciones de la barra de herramientas, por ejemplo en la siguiente imagen se puede ver como se puede explotar la dimensión ZONAS, tenemos seleccionadas las opciones:  para evitar filas o columnas con valores cero y  para detallar una instancia de miembro al pulsar el signo que se encuentra junto al nombre de dicho miembro (Se utiliza “+” para expandir y “-” para contraer).

En muchos casos queremos ver nuestra información no solamente en base a los códigos o identificadores sino también sus descripciones, para esto podemos utilizar la opción , en este caso se puede ver las propiedades de las dimensiones

En otras ocasiones será interesante ver la jerarquía de una dimensión para lo cual utilizamos la opción , aquí podemos observar los padres de cada miembro dentro de la vista de análisis, por ejemplo se muestra la estructura de la dimensión TIEMPO (Año, Semestre, Trimestre, Mes, Fecha) hasta el miembro actual de la dimensión, en este caso Fecha.



		Medidas
TIEMPO	ZONAS	● IMPORTE
+ 2006	+ Todas	\$396.642
+ 2007	+ Todas	\$318.608
- 2008	+ Todas	\$375.154
+ SEMESTRE I	+ Todas	\$201.164
- SEMESTRE II	+ Todas	\$173.990
+ TRIMESTRE III	+ Todas	\$98.135
- TRIMESTRE IV	+ Todas	\$75.855
+ octubre	+ Todas	\$32.942
+ noviembre	+ Todas	\$30.876
- diciembre	+ Todas	\$12.037
2008-12-01	+ Todas	\$3.146
2008-12-02	+ Todas	\$986

La opción  es muy importante ya que me permite intercambiar la información de los ejes de la vista de análisis obteniendo una forma de visualización distinta pero muy útil, a continuación se muestra un ejemplo claro:



		Medidas
TIEMPO	ZONAS	● IMPORTE
+ 2006	+ Todas	\$396.642
+ 2007	+ Todas	\$318.608
+ 2008	+ Todas	\$375.154

Slicer:

Luego de usar la opción , la tabla queda de la siguiente manera:



		TIEMPO		
		+ 2006	+ 2007	+ 2008
	ZONAS	ZONAS	ZONAS	ZONAS
Medidas	+ Todas	+ Todas	+ Todas	+ Todas
IMPORTE	\$396.642	\$318.608	\$375.154	

Slicer:

A continuación se mostrará el funcionamiento de los modos de detalle faltantes, comenzando por la opción , la cual detalla una instancia de un miembro que sea pulsada, pero con la diferencia que expande o contrae dicha instancia para todos los miembros de la dimensión anidada.

La opción  permite aislar una sección específica que deseamos analizar, en este caso llegamos al nivel REGION dentro de la dimensión UBICACIONES GEOGRAFICAS ocultando a los padres del miembro (Se utiliza “↓” para expandir y “↑” para contraer).



		Medidas
↑ TIEMPO	ZONAS	● IMPORTE
↓ 2006	↓ Todas	\$396.642
↓ 2007	↓ Todas	\$318.608
↓ 2008	↓ Todas	\$375.154

Slicer:

Si se desea ver todos los registros que formaron un valor específico, en este caso todos los servidores por institución que suman la cantidad total del año “2009”, se utiliza la opción de datos de origen  la cual presentará un cuadro con la descripción de ese dato.



		Medidas
TIEMPO	ZONAS	● IMPORTE
+ 2006	+ Todas	↓\$396.642
+ 2007	+ Todas	↓\$318.608
+ 2008	+ Todas	↓\$375.154

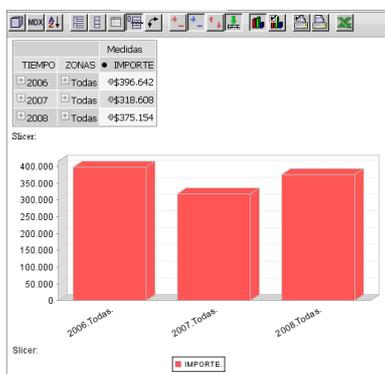
Slicer:

Drill Through Table for IMPORTE									
▲ AÑ±o	● Semestre	● Trimestre	● Mes	● Fecha	● Nombre	● Clientes	● Vendedores	● Zona	● IMPORTE
2008	SEMESTRE I	TRIMESTRE I	enero	2008-01-02	CONTADO CAJAS	ACOSTA MARIO	ND	BUCAY	0,10
2008	SEMESTRE I	TRIMESTRE I	enero	2008-01-02	CONTADO CAJAS	BUSCIA ZAMORA SONIA	ND	BABA	37,99
2008	SEMESTRE I	TRIMESTRE I	enero	2008-01-02	CONTADO CAJAS	CABEZAS BELGICA	ND	CUMAN	16,89
2008	SEMESTRE I	TRIMESTRE I	enero	2008-01-02	CONTADO CAJAS	CONSUMIDOR FINAL	TODOS	BUCAY	29,00
2008	SEMESTRE I	TRIMESTRE I	enero	2008-01-02	CONTADO CAJAS	GUZMAN CORSINO	ND	CUMAN	39,28
2008	SEMESTRE I	TRIMESTRE I	enero	2008-01-02	CONTADO CAJAS	PASFREI	ND	BUCAY	3,50
2008	SEMESTRE I	TRIMESTRE I	enero	2008-01-02	CONTADO CAJAS	YANES OLGA	ND	CUMAN	28,50
2008	SEMESTRE I	TRIMESTRE I	enero	2008-01-02	CONTADO UNIDAD	BUSCIA ZAMORA SONIA	ND	BABA	5,44
2008	SEMESTRE I	TRIMESTRE I	enero	2008-01-02	CONTADO UNIDAD	CABEZAS BELGICA	ND	CUMAN	64,22
2008	SEMESTRE I	TRIMESTRE I	enero	2008-01-02	CONTADO UNIDAD	CONSUMIDOR FINAL	TODOS	BUCAY	1,00

Página 1/742 Ir a página 1 Filas/página 10

En la mayoría de los casos un gráfico puede resultar más útil que una tabla de valores al momento de realizar análisis, debido a su facilidad de comprensión mostrando la misma información pero de una manera fácil, entendible e intuitiva; para esto existe la opción , la cual genera un gráfico con la información de la vista de análisis.

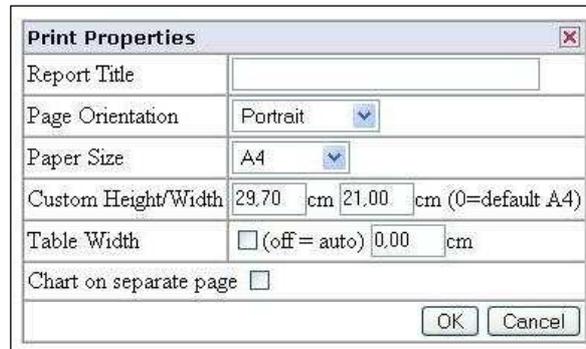
Es importante mencionar que no todos los tipos de gráficos son buenas alternativas al momento de analizar información, así mismo que no todos los reportes permiten generar gráficos entendibles debido a la cantidad de información y cruces de datos existentes, por lo cual es recomendable probar varias alternativas y generar gráficos de reportes resumidos que no desplieguen demasiada información, más bien sean consolidados.



Entonces para obtener un buen gráfico que represente la información deseada, es adecuado manejar la opción , que permite configurar las propiedades de diseño y presentación del mismo, por ejemplo **etiquetas**, colores, tipo de gráfico, etc.

Chart Properties	
Chart Type	Vertical Bar
Enable Drill Through	<input type="checkbox"/>
Chart Title	
Chart Title Font	SansSerif Bold 18
Horizontal axis label	
Vertical axis label	
Axes Label Font	SansSerif Plain 12
Axes Tick Label font	SansSerif Plain 12 30°
Show Legend	<input checked="" type="checkbox"/> Bottom
Legend Font	SansSerif Plain 10
Show Slicer	<input checked="" type="checkbox"/> Bottom Left
Slicer Font	SansSerif Plain 12
Chart Height	300
Chart Width	500
Background (R, G, B)	255 255 255
OK Cancel	

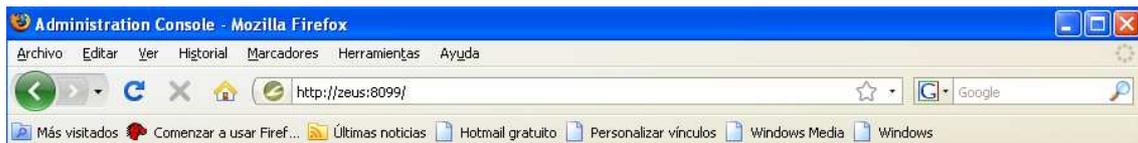
Así mismo podemos configurar las propiedades de impresión (o exportación) para la vista de análisis del reporte.



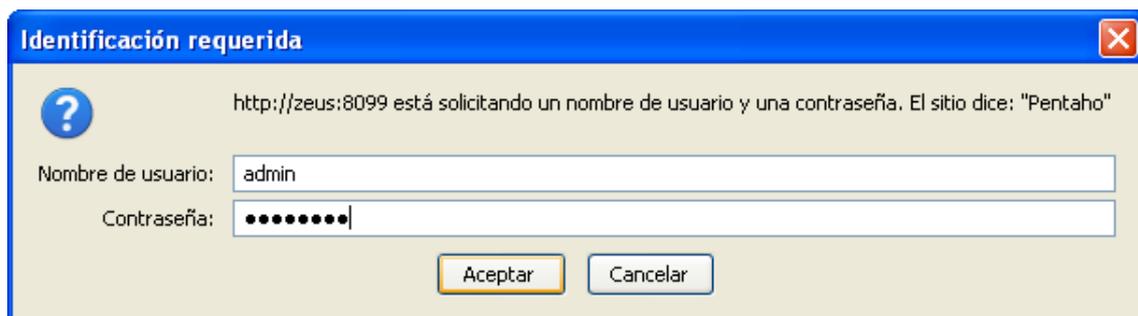
### 1.8. 2.5 Administración del Sistema.

Esta sección permite al usuario(s) que posea el perfil de “Administrador del Sistema” dar mantenimiento y control del mismo para su adecuado funcionamiento y del respectivo repositorio de soluciones BI.

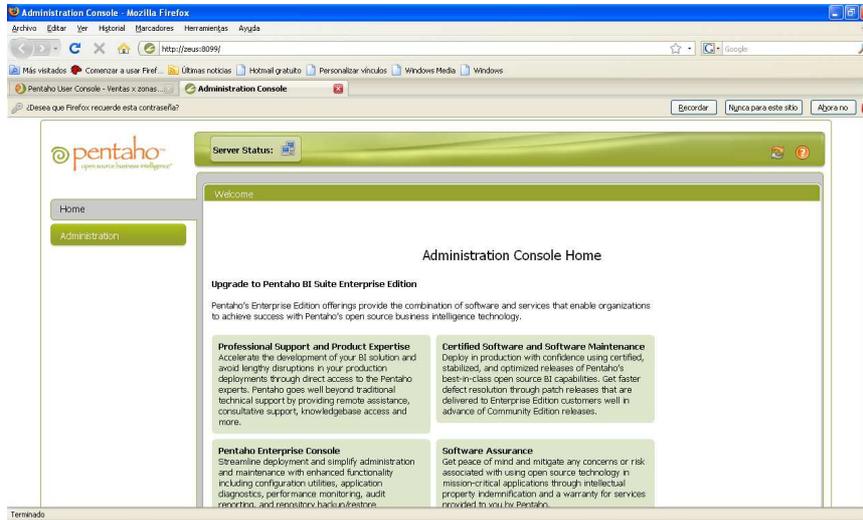
Para ingresar a la consola de administración abrir un navegador de internet y poner la siguiente URL: <http://zeus:8099/> .



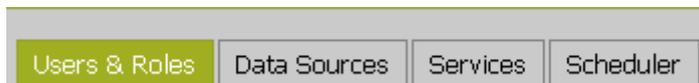
En la siguiente pantalla ingresar el nombre de usuario (admin) y contraseña (password).



Una vez validado su autenticación, deberá pulsar el botón Administration

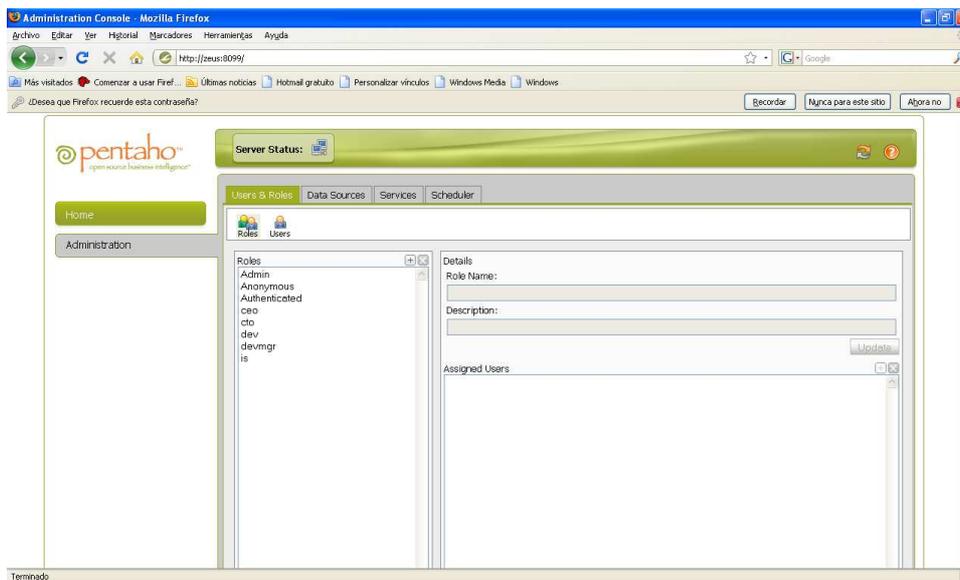


La consola de administración permite las siguientes opciones:



, a continuación se detallan cada una de ellas.

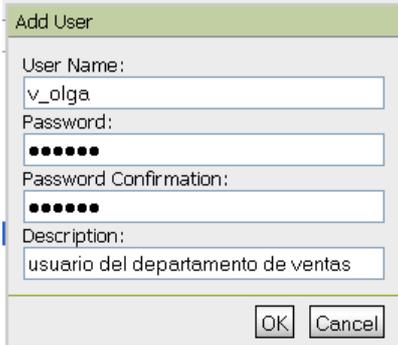
- 1. Users & Roles:** Esta opción permite crear usuarios y grupos de Pentaho.



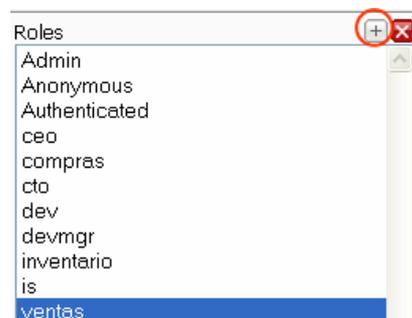
Para crear un usuario debe elegir el botón  y seleccionar el signo + que se encuentra encerrado en el círculo.



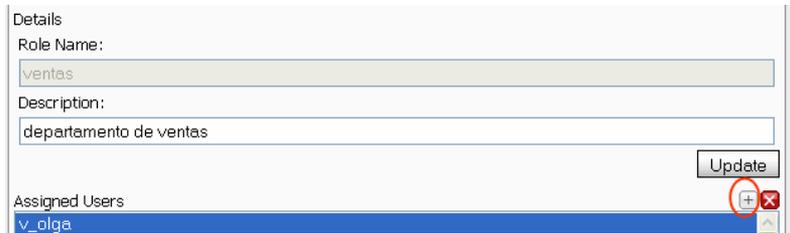
Donde se desplegará una ventana, en la que debe ingresar el nombre y password del usuario, por ejemplo usuario v\_olga y password v\_olga, seguidamente pulsar el botón OK.



Para crear un grupo, debe elegir el icono , y seleccionar el signo + que se encuentra encerrado en el círculo.



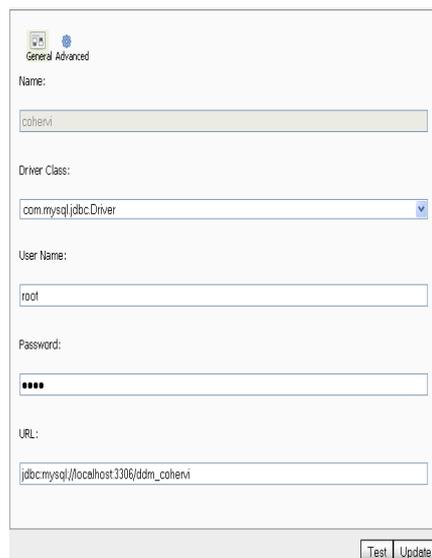
Para asignar un usuario a un grupo se realiza en la sección de grupos, seleccionando el signo +, que se encuentra encerrado en el círculo.



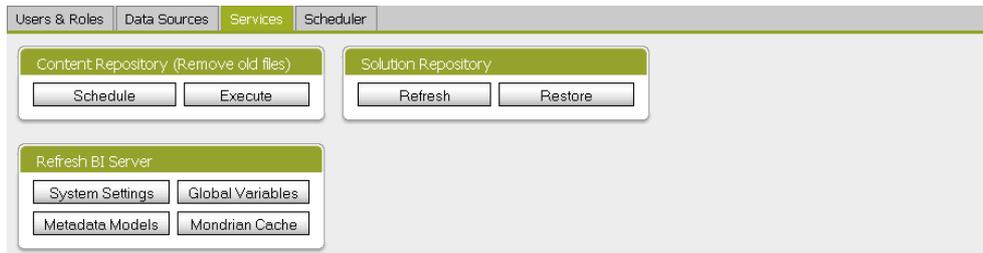
- Data Sources:** Permite crear los string de conexión a la base de datos, para crear un nuevo data sources, debe seleccionar el signo + que está encerrado en el círculo.



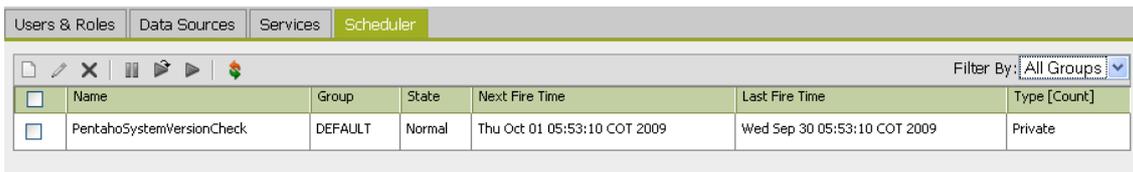
En la siguiente pantalla debe ingresar un nombre del data source, el driver de base de datos, el usuario, el password y la URL, seguidamente pulsar el botón Test.



3. **Services:** Esta opción permite administrar el repositorio de contenido, repositorio de soluciones y Refrescar la Suite de Pentaho, con solo pulsar una de los botones se ejecuta automáticamente la acción.



#### 4. Scheduler:



	Name	Group	State	Next Fire Time	Last Fire Time	Type [Count]
<input type="checkbox"/>	PentahoSystemVersionCheck	DEFAULT	Normal	Thu Oct 01 05:53:10 COT 2009	Wed Sep 30 05:53:10 COT 2009	Private