



UNIVERSIDAD TÉCNICA DEL NORTE
FACULTAD DE INGENIERÍA EN CIENCIAS APLICADAS
CARRERA DE INGENIERÍA EN SISTEMAS
COMPUTACIONALES

TEMA:

“Estudio de metodologías de Data Warehouse para la implementación de repositorios de información para la toma de decisiones gerenciales.”

AUTOR:

Roberto Xavier Villarreal Rosero

DIRECTOR:

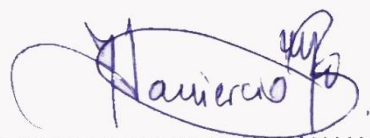
Ing. Mauricio Rea

Ibarra – Ecuador

2013

CERTIFICACIÓN DIRECTOR

Certifico que la tesis “**Estudio de metodologías de Data Warehouse para la implementación de repositorios de información para la toma de decisiones gerenciales.**”, ha sido realizada en su totalidad por el egresado: Roberto Xavier Villarreal Rosero, portador de la cédula de identidad número: 1716275795.

A handwritten signature in blue ink, appearing to read "Mauricio Rea", is written over a horizontal dotted line.

Ing. Mauricio Rea
Director de la Tesis



COOPERATIVA DE AHORRO Y CRÉDITO
"SAN ANTONIO LTDA"

Juntos hacia el progreso

CERTIFICACIÓN

Fundada el 25 de Abril de 1960
Acuerdo Ministerial 1596.
Orden Registro Cooperativas 613

Ibarra, 23 de abril del 2013

Señores
UNIVERSIDAD TÉCNICA DEL NORTE
Presente

De mis consideraciones.-

Siendo auspiciantes del proyecto de tesis del egresado ROBERTO XAVIER VILLARREAL ROSERO con CI: 1716275795, quien desarrolló su trabajo con el tema "ESTUDIO DE METODOLOGÍAS DE DATA WAREHOUSE PARA LA IMPLEMENTACIÓN DE REPOSITORIOS DE INFORMACIÓN PARA LA TOMA DE DECISIONES GERENCIALES.", me es grato informar que se han superado con satisfacción las pruebas técnicas y la revisión de cumplimiento de los requerimientos funcionales, por lo que se recibe el proyecto como culminado y realizado por parte del egresado ROBERTO XAVIER VILLARREAL ROSERO. Una vez que hemos recibido la capacitación y documentación respectiva, nos comprometemos a continuar utilizando el mencionado aplicativo en beneficio de nuestra institución.

El egresado ROBERTO XAVIER VILLARREAL ROSERO puede hacer uso de este documento para los fines pertinentes en la Universidad Técnica del Norte.

Atentamente,

Ing. Norma Vásquez
Gerente General
COAC "SAN ANTONIO LTDA."



iii

MATRIZ SAN ANTONIO DE IBARRA
Hnos Mideros 6-33 y 27 de Noviembre
Telfs: (06) 2932 063 - 2932 169 - 2932 365

AGENCIA ATUNTAQUI
Av. Salinas y Atahualpa 14-07
Telf: (06) 2910 230

AGENCIA IBARRA
Velasco 633 (entre Bolívar y Olmedo)
Telfs.: (06) 2950305 - 2600270 - 2600277

Email: coacsanantonio@hotmail.es



UNIVERSIDAD TÉCNICA DEL NORTE
CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE INVESTIGACIÓN
A FAVOR DE LA UNIVERSIDAD TÉCNICA DEL NORTE

Yo, ROBERTO XAVIER VILLARREAL ROSERO, con cédula de identidad Nro. 1716275795, manifiesto mi voluntad de ceder a la Universidad Técnica del Norte los derechos patrimoniales consagrados en la ley de propiedad intelectual del Ecuador, artículo 4, 5 y 6, en calidad de autor del trabajo de grado denominado: **“ESTUDIO DE METODOLOGÍAS DE DATA WAREHOUSE PARA LA IMPLEMENTACIÓN DE REPOSITORIOS DE INFORMACIÓN PARA LA TOMA DE DECISIONES GERENCIALES.”**, que ha sido desarrollado para optar por el título de Ingeniería en Sistemas Computacionales, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En mi condición de autor me reservo los derechos morales de la obra antes mencionada, aclarando que el trabajo aquí descrito es de mi autoría y que no ha sido previamente presentado para ningún grado o calificación profesional.

En concordancia suscribo este documento en el momento que hago entrega del trabajo final en formato impreso y digital a la biblioteca de la Universidad Técnica del Norte

Nombre: ROBERTO XAVIER VILLARREAL ROSERO

Cédula: 1716275795

Ibarra a los 07 días del mes de octubre del 2013



UNIVERSIDAD TÉCNICA DEL NORTE

BIBLIOTECA UNIVERSITARIA

AUTORIZACIÓN DE USO Y PUBLICACIÓN A FAVOR DE LA UNIVERSIDAD TÉCNICA DEL NORTE

1. IDENTIFICACIÓN DE LA OBRA

La UNIVERSIDAD TÉCNICA DEL NORTE dentro del proyecto Repositorio Digital institucional determina la necesidad de disponer los textos completos de forma digital con la finalidad de apoyar los procesos de investigación, docencia y extensión de la universidad.

Por medio del presente documento dejo sentada mi voluntad de participar en este proyecto, para lo cual ponemos a disposición la siguiente investigación:

DATOS DE CONTACTO	
CÉDULA DE IDENTIDAD	1716275795
APELLIDOS Y NOMBRES	ROBERTO XAVIER VILLARREAL ROSERO
DIRECCIÓN	Av. El Retorno, Conjunto Villasol del Retorno Casa 6
EMAIL	otreborex@yahoo.es
TELÉFONO FIJO	
TELÉFONO MOVIL	0992516956
DATOS DEL TRABAJO DE GRADO	
TÍTULO	“ESTUDIO DE METODOLOGÍAS DE DATA WAREHOUSE PARA LA IMPLEMENTACIÓN DE REPOSITORIOS DE INFORMACIÓN PARA LA TOMA DE DECISIONES GERENCIALES.”
AUTOR	ROBERTO XAVIER VILLARREAL ROSERO
FECHA	07 de octubre de 2013

PROGRAMA	PREGRADO
TÍTULO	INGENIERÍA EN SISTEMAS COMPUTACIONALES
DIRECTOR	ING. MAURICIO REA

2. AUTORIZACIÓN DE USO A FAVOR DE LA UNIVERSIDAD

Yo, ROBERTO XAVIER VILLARREAL ROSERO, con cédula de identidad Nro. 1716275795, en calidad de autor y titular de los derechos patrimoniales del trabajo de grado descrito anteriormente, hago entrega del ejemplar respectivo en forma digital y autorizo a la Universidad Técnica del Norte, la publicación de la obra en el Repositorio Digital Institucional y el uso del archivo digital en la biblioteca de la universidad con fines académicos, para ampliar la disponibilidad del material y como apoyo a la educación, investigación y extensión, en concordancia con la Ley de Educación Superior Artículo 144.



.....

Nombre: ROBERTO XAVIER VILLARREAL ROSERO

Cédula: 1716275795

Ibarra a los 07 días del mes de octubre del 2013

CONSTANCIAS

El autor manifiesta que la obra objeto de la presente autorización es original y se la desarrollo sin violar derechos de autor de terceros, por lo tanto la obra es original y que es el titular de los derecho patrimoniales, por lo que asume la responsabilidad sobre el contenido de la misma y saldrá en defensa de la Universidad en caso de reclamación por parte de terceros.

Ibarra a los 07 días del mes de octubre de 2013



.....

Nombre: ROBERTO XAVIER VILLARREAL ROSERO

Cédula: 1716275795

DEDICATORIA

Este trabajo de tesis va dedicado a mis padres por ser el pilar fundamental en todo lo que soy, en toda mi educación, tanto académica, como de la vida, por su incondicional apoyo perfectamente mantenido a través del tiempo.

Dedicado también a mi hijo y esposa por todo su amor. Cualquier sacrificio es nada comparado a la recompensa de tenerles a mi lado.

Todo este trabajo ha sido posible gracias a ellos.

AGRADECIMIENTO

A Dios

Por haberme permitido llegar a este punto y haberme dado salud, amor y paciencia para culminar este objetivo en mi vida.

A mi padre y madre.

Por apoyarme siempre en las decisiones que he tomado y ser ese pilar necesario para caminar día a día, siempre con ese amor incondicional que solo los padres pueden ofrecer a sus hijos.

A mi esposa Jesse y mi pequeño hijo Paquito.

Por ser esa fuente de inspiración que me han permitido terminar este objetivo. Hijo mío quiero que te sientas orgulloso de tu padre y madre. Les amo a los dos.

Al Ing. Mauricio Rea.

Por ser mi amigo y asesor en este tesis. Muchas gracias Ingeniero.

A la UTN

Por brindarme sus conocimientos y poder conocer a tan buenos catedráticos y amigos.

ÍNDICE DE CONTENIDOS

CERTIFICACIÓN DIRECTOR.....	II
CERTIFICACIÓN.....	III
CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE INVESTIGACIÓN.....	IV
AUTORIZACIÓN DE USO Y PUBLICACIÓN	V
CONSTANCIAS	VII
DEDICATORIA	VIII
AGRADECIMIENTO	IX
ÍNDICE DE CONTENIDOS.....	X
ÍNDICE DE TABLAS	XVI
RESUMEN	XVII
SUMMARY	XVIII
CAPÍTULO I	19
1.1. ANTECEDENTES.....	20
1.2. PROBLEMA	21
1.3. OBJETIVOS.....	22
1.4. JUSTIFICACIÓN	23
1.5. ALCANCE	24
CAPÍTULO II	26
2.1. HISTORIA DE BUSINESS INTELLIGENCE.....	27
2.2. DIFERENCIA ENTRE DATOS, INFORMACIÓN Y CONOCIMIENTO	28
2.3. DEFINICIÓN DE BUSINESS INTELLIGENCE.....	30
2.4. VENTAJAS Y DESVENTAJAS DE UTILIZAR UN PROCESO DE INTELIGENCIA DE NEGOCIOS.....	32
2.4.1 VENTAJAS.....	32
2.4.2 DESVENTAJAS.....	34
2.5. ARQUITECTURA TÍPICA DE SOLUCIONES BUSINESS INTELLIGENCE	36

2.5.1	FUNCIONES DE LA BASE DE DATOS RELACIONAL	37
2.5.2	FUNCIONES DE LA BASE DE DATOS MULTIDIMENSIONAL	37
2.5.3	FUNCIONES DEL VISUALIZADOR	37
2.6	COMPONENTES DE BUSINESS INTELLIGENCE	37
2.6.1	DATA WAREHOUSE	37
2.6.1.1	CARACTERÍSTICAS DE UN DATA WAREHOUSE	38
2.6.2	DATA MARTS.....	41
2.6.2.1	DATA MARTS DEPENDIENTES.....	42
2.6.2.2	DATA MARTS INDEPENDIENTES	43
2.6.3	PROCESAMIENTO ANALÍTICO EN LÍNEA (OLAP)	44
2.6.3.1	CARACTERÍSTICAS PRINCIPALES DEL OLAP	45
2.6.4	PROCESAMIENTO TRANSACCIONAL EN LÍNEA (OLTP)	46
2.6.4.1	DIFERENCIAS ENTRE PROCESAMIENTO TRANSACCIONAL EN LÍNEA (OLPT) Y PROCESAMIENTO ANALÍTICO EN LÍNEA (OLAP).....	46
2.7	BASE DE DATOS MULTIDIMENSIONAL.....	48
2.7.1	TABLAS DE DIMENSIONES	49
2.7.2	TABLA DIMENSIÓN TIEMPO.....	51
2.7.3	TABLAS DE HECHOS	51
2.7.4	TIPOS DE MODELAMIENTO DE UN DATA WAREHOUSE	53
2.7.4.1	ESQUEMA EN ESTRELLA (STAR SCHEMA)	53
2.7.4.2	ESQUEMA DE COPO DE NIEVE	54
2.7.4.3	ESQUEMA CONSTELACIÓN	55
2.7.5	TIPOS DE IMPLEMENTACIÓN DE UN DATA WAREHOUSE	56
2.7.5.1	ROLAP	56
2.7.5.2	MOLAP	57
2.7.5.3	HOLAP.....	57
2.7.6	CUBO MULTIDIMENSIONAL.....	57
2.7.6.1	INDICADORES.....	59
2.7.6.2	ATRIBUTOS	60
2.7.6.3	JERARQUÍAS	60
CAPÍTULO III		65
3.1	CUADROS DE MANDO INTEGRALES (CMI)	66
3.1.1	DIFERENCIAS CON OTRAS HERRAMIENTAS DE BUSSINES INTELLIGENCE	66
3.2	TIPOS DE CUADRO DE MANDO	67

3.3	SISTEMAS DE SOPORTE A LA DECISIÓN	69
3.3.1	TIPOS DE PROBLEMAS.....	70
3.3.2	FUNCIONES Y CARACTERÍSTICAS.....	70
3.3.3	OBJETIVOS DE LOS DSS.....	71
3.3.4	VENTAJAS DE LOS DSS.....	72
3.3.6	ARQUITECTURAS.....	73
3.4	SISTEMAS DE INFORMACIÓN EJECUTIVA	75
3.4.1	CARACTERÍSTICAS	75
3.4.2	DESARROLLO	76
3.4.3	FACTORES DE ÉXITO.....	77
3.4.4	FACTORES DE FRACASO	78
3.4.5	TENDENCIAS EN EL FUTURO.....	79
CAPÍTULO IV		80
4.1	ANÁLISIS DE LAS METODOLOGÍAS	81
4.1.1	METODOLOGÍA HEFESTO	81
4.1.2	METODOLOGÍA BILL INNMON.	92
4.1.3	METODOLOGÍA RALPH KIMBALL	94
4.1.4	CRITERIOS DE SELECCIÓN.	99
4.1.5	COMPARATIVA DE LAS METODOLOGÍAS.	100
4.1.6	SELECCIÓN DE LA METODOLOGÍA.....	102
CAPÍTULO V		104
5.1	INTRODUCCIÓN.....	105
5.2	MÓDULOS DEL SISTEMA	106
5.3	FUNCIONALIDADES PRINCIPALES	107
CAPÍTULO VI		109
6.	METODOLOGÍA PARA DESARROLLO DEL PROYECTO.....	110
6.1	PLANEACIÓN Y ADMINISTRACIÓN DEL PROYECTO	110
6.1.1	OBJETIVO DEL PROYECTO	110
6.1.2	DEFINICIÓN DEL PROYECTO.....	110
6.1.3	ALCANCE DEL PROYECTO	111
6.1.4	JUSTIFICACIÓN DEL PROYECTO EN EL NEGOCIO	112
6.2	ANÁLISIS DE REQUERIMIENTOS.....	112
6.2.1	LEVANTAMIENTO DE REQUERIMIENTOS.....	113

6.2.1.1	DOCUMENTACIÓN DE REQUERIMIENTOS	114
6.3	MODELAMIENTO DIMENSIONAL	117
6.3.1	EL DATA MART	117
6.3.2	DEFINICIÓN DE LA GRANULARIDAD	118
6.3.3	DIMENSIONES	118
6.3.4	TABLAS DE HECHOS.....	121
6.3.5	DISEÑO DEL MODELO DIMENSIONAL	124
6.4	DISEÑO TÉCNICO DE LA ARQUITECTURA	126
6.4.1	DATOS.....	127
6.4.2	MAPEO DE LOS DATOS EN LOS MODELOS DIMENSIONALES.....	127
6.4.3	INFRAESTRUCTURA.....	131
6.4.4	CONSTRUCCIÓN DE LOS CUBOS.....	132
6.5	REPORTES IMPLEMENTADOS	135
6.6	MANTENIMIENTO Y CRECIMIENTO.....	141
CAPÍTULO VII		143
CONCLUSIONES		144
ANÁLISIS DE IMPACTO		145
RECOMENDACIONES		147
BIBLIOGRAFÍA.....		148
ANEXOS		149

ÍNDICE DE ILUSTRACIONES

Ilustración 1 Ejemplo tabla reporte trimestral	25
Ilustración 2 Ejemplo gráfico reporte trimestral	25
Ilustración 3 Transformación de Datos a Conocimiento	31
Ilustración 4 Arquitectura de una solución BI	36
Ilustración 5 Características de un Data Warehouse	38
Ilustración 6 Data Warehouse no volátil.....	40
Ilustración 7 Arquitectura Data Marts Dependientes	43
Ilustración 8 Arquitectura Data Marts Independientes.....	44
Ilustración 9 Ejemplo de tablas de dimensiones	50
Ilustración 10 Tabla de hechos ventas.....	52
Ilustración 11 Esquema en estrella	53
Ilustración 12 Esquema copo de nieve	54
Ilustración 13 Esquema constelación.....	55
Ilustración 14 Cubo Multidimensional.....	58
Ilustración 15 Jerarquía Fecha.....	61
Ilustración 16 Esquema estrella de ejemplo.....	62
Ilustración 17 Cubo de ventas.....	62
Ilustración 18 Cubo de ventas con atributos	63
Ilustración 19 Cubo de ventas con indicador.....	63
Ilustración 20 Representación matricial Cubo de ventas.....	64
Ilustración 21 Diferencias con otras herramientas BI	67
Ilustración 22 Cuadro de Mando Integral	69
Ilustración 23 Vista de un DSS	74
Ilustración 24 Logotipo Hefesto.....	81
Ilustración 25 Etapas de la metodología Hefesto	83
Ilustración 26 Metodología Hefesto.....	83
Ilustración 27 Modelo conceptual-Metodología Hefesto	86
Ilustración 28 Modelo conceptual ampliado – Hefesto	89
Ilustración 29 Tabla de dimensiones-Hefesto	90
Ilustración 30 Metodología Innmon- DW Corporativo	94

Ilustración 31 Metodología Kimball- Arquitectura Bus del DW.....	96
Ilustración 32 Ciclo de vida Ralph Kimball	97
Ilustración 33 Modelo de las versiones de Pentaho	105
Ilustración 34 Arquitectura simplificada de Pentaho	108
Ilustración 35 Modelo dimensional.....	125
Ilustración 36 Diseño de la arquitectura	126
Ilustración 37 Ejemplo de un proceso ETL.....	128
Ilustración 38 Ejemplo de entrada tabla cnxinversiones.....	129
Ilustración 39 Ejemplo salida a tabla del data warehouse	129
Ilustración 40 Ejemplo de proceso ETL en una tabla de hechos	130
Ilustración 41 Job o Tarea para ejecutar ordenadamente todas las transformaciones	131
Ilustración 42 Estructura de un cubo en XML.....	135
Ilustración 43 Resultados gráficos del reporte: Socios por oficinas	136
Ilustración 44 Resultados del reporte: Socios por género	137
Ilustración 45 Resultados del reporte: Socios retirados por períodos.....	137
Ilustración 46 Resultado de reporte Inversiones por oficina	138
Ilustración 47 Cartera por estado.....	138
Ilustración 48 Consolidado de cartera de crédito	139
Ilustración 49 Cartera por líneas de crédito.....	139
Ilustración 50 Cartera por origen de recursos	140
Ilustración 51 Cartera de crédito por períodos	141

ÍNDICE DE TABLAS

Tabla 1 Diferencias entre OLAP y OLTP	48
Tabla 2 Comparativa de las metodologías estudiadas	100
Tabla 3 Tabla comparativa de Metodología Innmon vs Kimball	101
Tabla 4 Puntuación a los criterios de evaluación	102
Tabla 5 Calificaciones a las metodologías	103
Tabla 6 Funcionalidades de versiones	106
Tabla 7 Dimensión clientes	119
Tabla 8 Dimensión estados clientes	119
Tabla 9 Dimensión Oficina	120
Tabla 10 Dimensión Tipo de clientes	120
Tabla 11 Tabla de hechos clientes	121
Tabla 12 Tabla de Hechos clientes	122
Tabla 13 Tabla de hechos inversiones	122
Tabla 14 Tabla de hechos Inversiones	122
Tabla 15 Tabla de hechos créditos	123
Tabla 16 Tabla de hechos Cartera de crédito	124
Tabla 17 Herramientas software para BD y Análisis	132
Tabla 18 Cubos diseñados con Schema Workbench	134

RESUMEN

El presente trabajo de investigación está orientado a conocer los conceptos y las estrategias de Inteligencia de Negocios, además, estudiar y hacer un análisis de tres metodologías de Data Warehouse que actualmente son las más utilizadas, destacar las ventajas de la utilización de un proceso metodológico para la obtención de un sistema de calidad de Inteligencia de Negocios.

Además dentro de los objetivos de este estudio se plantea implementar un repositorio de información el cual permita a la empresa donde se está implementando la aplicación, obtener los reportes gerenciales que permitan realizar el análisis y la toma de decisiones que beneficien a la empresa y que le conlleven a ser más competitiva en el mercado del sector financiero.

A manera de introducción en el Capítulo I se da una corta descripción sobre los procesos de BI en las empresas, el aporte de la tecnología a las mismas.

En el capítulo II se profundiza los conceptos de Inteligencia de negocios, sus ventajas y desventajas.

El capítulo III habla sobre productos de BI, las características y diferencias entre cada uno de ellos.

En el capítulo IV se hace el análisis de las metodologías de construcción de Data Warehouse planteadas para el estudio y se desarrolla el proceso de selección para determinar cuál metodología es la más adecuada para la construcción del DW.

En el capítulo V se muestra las herramientas con las cuales se va a construir el DW y posterior explotación para obtener conocimiento.

En el capítulo VI siguiendo la metodología que se seleccione en el capítulo IV se desarrolla fase a fase la creación del repositorio.

En el capítulo VII se realiza las respectivas conclusiones, el análisis de impacto en la Cooperativa y las recomendaciones.

SUMMARY

The following research paper focuses on understanding Business Intelligence concepts and strategies, presents an overview of three common Data Warehouse methodologies, and highlights the advantages of utilizing a methodological process to obtain a quality Business Intelligence system.

Additionally included in the objectives of this study is a proposal for an information repository allowing a business implementing the application to generate managerial reports for the purpose of analyzing and making business decisions that would enhance its competitiveness in the financial sector market.

As an introduction, chapter I provides a short description of Business Intelligence (BI) processes used in firms and the contribution of technology in these businesses.

Chapter II examines in more depth BI concepts, their advantages, as well as their disadvantages.

Chapter III discusses BI products, and outlines their characteristics and differences.

Chapter IV analyzes Data Warehouse (DW) construction methodologies proposed in the study and elaborates on the selection process required to identify which methodology is most adequate for building the DW.

Chapter V reviews the tools with which the DW will be implemented and how to profit from them to gain knowledge.

Chapter VI develops a step-by-step process for creating the repository following the methodology selected in chapter IV.

Finally, chapter VII offers conclusions, as well as an impact analysis in the Cooperativa and recommendations.

CAPÍTULO I

INTRODUCCIÓN



1.1. Antecedentes

La Cooperativa de Ahorro y Crédito “San Antonio Ltda.”, es una institución financiera con 52 años de servicio en el norte del país, brinda sus servicios en la oficina matriz ubicada en San Antonio de Ibarra y en sus agencias ubicadas en las ciudades de Atuntaqui e Ibarra.

Su misión principal es apoyar el desarrollo integral de sus asociados, mediante la prestación de soluciones financieras y servicios complementarios oportunos, a través de un equipo humano capacitado e innovador, con la aplicación de tecnología adecuada y alianzas estratégicas con entidades afines.

En la actualidad la Cooperativa de Ahorro y Crédito “San Antonio Ltda.” cuenta con un sistema administrativo financiero adquirido a principios de la década pasada y del cual proviene toda la información que le es útil, pero que no le es suficiente para responder con todos los requerimientos y necesidades de información. Por parte de los proveedores se cuenta con soporte técnico, los cuales dependiendo del tiempo que disponen cumplen con los requerimientos que la Cooperativa demanda. Este sistema cuenta con un módulo de reportes gerenciales que muestran información plana y que por su simplicidad no es utilizado por gerencia para la toma de decisiones.

Con el crecimiento acelerado que ha presentado la Cooperativa en estos últimos 5 años, los continuos cambios y requerimientos de información interna y externa, hace suponer que se necesita de algún mecanismo tecnológico o proceso de inteligencia de negocios que coadyuve a la obtención de conocimiento y a la mejora toma de decisiones. Y por otra parte y aunque no menos importante deslinda la necesidad directa de esperar que los proveedores del sistema solucionen esta escasez de información.

1.2 Problema

La información, el conocimiento y la toma de decisiones son pieza clave para obtener una ventaja competitiva en el mundo de los negocios. Para mantenerse competitiva una empresa, los gerentes y quienes toman las decisiones requieren de un acceso fácil y rápido a la 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 datos de la empresa 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 estructurada para su análisis.

Actualmente el conformar un reporte de análisis gerencial demanda mucho tiempo y esfuerzo en encontrar y unir la información dispersa que se debe consolidar utilizando herramientas como Microsoft Excel la cual demanda experticia y conocimiento específico en la manipulación de esta herramienta.

Además los reportes que se obtienen en el sistema administrativo financiero no cubren las expectativas de quienes toman las decisiones y en ocasiones los archivos planos de información no son fáciles de comprender rápidamente, inclusive con los conocimientos respectivos. Para eso es necesario apoyarse en herramientas que faciliten la comprensión y que a través de técnicas de visualización y métodos, permitan a los administradores y a cualquier usuario o persona comprender rápidamente los resultados del análisis.

1.3 Objetivos

1.3.1 Objetivo General

- Investigar las metodologías de Data Warehouse propuestas por Kimball, Innmon y Hefesto para la implementación de repositorios de información para la toma de decisiones gerenciales.

1.3.2 Objetivos Específicos

- Determinar los criterios para seleccionar la metodología de Data Warehouse más conveniente de entre las mencionadas.
- Documentar el proceso del estudio comparativo de las metodologías mencionadas.
- Implementar un repositorio de información para la obtención de reportes de cartera de crédito para la Cooperativa de Ahorro y Crédito “San Antonio Ltda.” con la herramienta de BI Pentaho.
- Determinar las ventajas y desventajas que se obtienen al utilizar un proceso de BI.
- Realizar pruebas del repositorio a implementarse.

1.4 Justificación

Gestionar la información en las empresas es, hoy en día, una herramienta clave para poder sobrevivir en un mercado cambiante, dinámico y global. Aprender a competir con esta información es fundamental para la toma de decisiones, el crecimiento y la gestión de nuestra empresa.

La disciplina denominada como Business Intelligence nos acerca a los sistemas de información que nos ayudan a la toma de decisiones en nuestra organización. Las pequeñas y medianas empresas disponen, como todas las empresas de sistemas de información sofisticados y que es conveniente analizar y optimizar

El estudio a realizarse nos ayudará, a través de la aplicación de la metodología de DW seleccionada y la generación del repositorio de información, mejorar la orientación al cliente interno, nuestros procesos, nuestra gestión económica/financiera.

Esta investigación que a través de capítulos teóricos y del caso práctico plasmado en el repositorio de información, desde el rigor técnico y la visión particularizada para una pequeña o mediana empresa pretende ser una solución adaptada a las necesidades reales de la Cooperativa.

Los usuarios de este repositorio serán:

- Gerencia General
- Jefes departamentales y de agencias

Estos usuarios se valdrán de esta herramienta para revisar y analizar el estado de colocación de créditos por líneas, consolidados de colocaciones por fechas/tiempo, lo cual facilitará el análisis y la toma de decisiones necesarias para buscar estrategias importantes para mejorar el mercado.

1.5 Alcance

Se estudiará las metodologías de Data Warehouse (Ralph Kimball, Innmon y Hefesto) para comprender el proceso de BI, realizar la extracción de los datos y hacer que estos se transformen en información útil que facilite el análisis y la toma de decisiones gerenciales e inclusive se pueda prever lo que va a pasar.

Por ser ampliamente utilizadas en los proyectos de BI se realizará la comparación de las metodologías Ralph Kimball (enfoque multidimensional), Bill Innmon (enfoque Empresarial Warehouse) y la metodología Hefesto que mantiene un enfoque realista y práctico.

Existen muchas metodologías de Data Warehouse que utilizan aproximaciones de las dos primeras metodologías y que no entrarán en el estudio de este proyecto.

La creación del repositorio se basa principalmente en realizar los reportes que se presentan a la asamblea de socios y que están relacionados con cartera de crédito y socios.

- Socios ingresados por períodos
- Socios retirados por períodos
- Operaciones de crédito por tipo y oficina
- Consolidado de créditos totalizado
- Créditos por estado

En el siguiente ejemplo se muestra la presentación de los reportes realizados en una hoja de cálculo de Excel y que se presenta en la reunión semestral con los representantes de los socios de la Cooperativa.

CONSOLIDADO		
TIPO DE CARTERA	Nº DE OPERACIONES	VALOR
MICROCREDITO	122	1,390,996
CONSUMO	16	158,210
VIVIENDA	6	103,493
TOTAL	144	1,652,699

Ilustración 1 Ejemplo tabla reporte trimestral

Fuente: *Coopetiva de Ahorro y Crédito "San Antonio Ltda."*

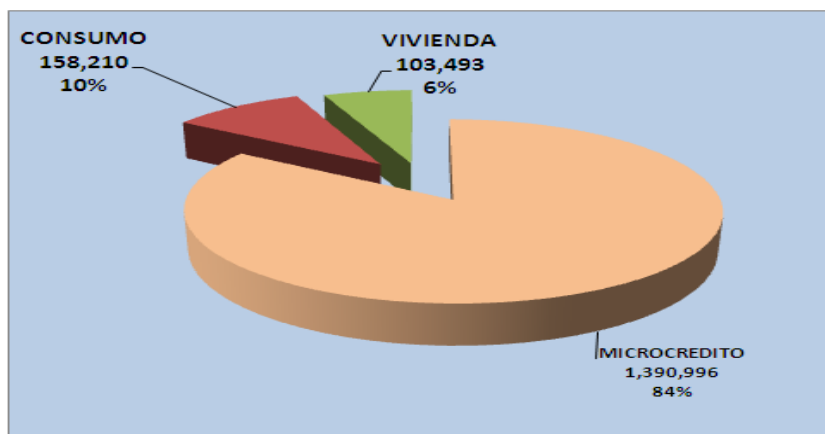


Ilustración 2 Ejemplo gráfico reporte trimestral

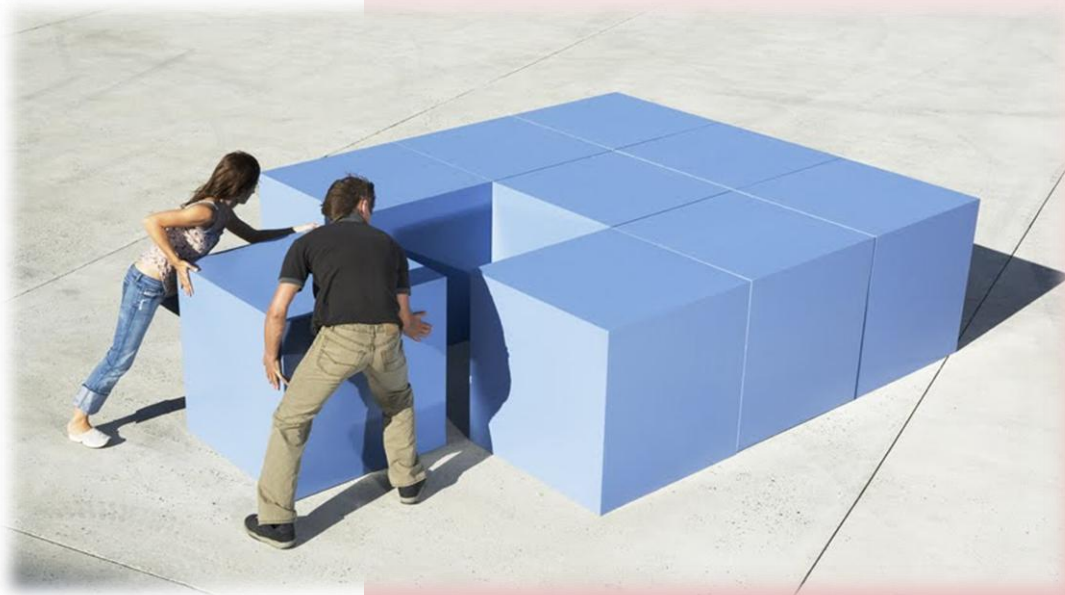
Fuente: *Coopetiva de Ahorro y Crédito "San Antonio Ltda."*

En el gráfico anterior por ejemplo se refleja el total de créditos y el porcentaje que representa según tipo de cartera que fueron aprobados por la Comisión de Crédito en el segundo semestre del año 2011.

La preparación y elaboración de estos reportes demanda tiempo y esfuerzo del personal encargado en realizar esta tarea. Con la implementación de este sistema se pretende facilitar la creación de este tipo de reportes que permitirá la toma de mejores decisiones con el análisis de los mismos.

CAPÍTULO II

BUSINESS INTELLIGENCE



Introducción

En el presente capítulo se realizará un análisis de la historia y una breve descripción del concepto de Business Intelligence y todos los componentes que se ven englobados dentro de él, así como cuáles son sus principales características y cuáles han sido las claves del éxito del mismo

2.1. Historia de Business Intelligence

A continuación, se detallan los que se podrían denominar hitos más importantes en la historia de Business Intelligence, abarcando hechos anteriores a la creación de este concepto, pero que son antecedentes que han ayudado a la definición del mismo:

- 1969: Creación del concepto de base de datos (Codd¹)
- 1970-1979: Desarrollo de las primeras bases de datos y las primeras aplicaciones empresariales (SAP, JD Edwards, Siebel, PeopleSoft). Estas ofrecieron la posibilidad de realizar “data entry”² en los sistemas, aumentando la información disponible, pero no fueron capaces de ofrecer un acceso rápido y fácil a la misma.
- 1980-1988: Creación del concepto de Almacén de Datos (Ralph Kimball, Bill Inmon), y aparición de los primeros sistemas de reporting. A pesar de esto, la funcionalidad ofrecida era escasa, además de ser complicado para el usuario. En esta época ya existían sistemas de bases de datos relativamente potentes, pero no había aplicaciones para facilitar su explotación.
- 1989: Introducción del término Business Intelligence (Howard Dresner³)
- 1990-1999: Business Intelligence 1.0. Proliferación de múltiples aplicaciones BI. Estos proveedores resultaban caros, pero facilitaron el acceso a la información. Hay

¹ Codd: Científico informático conocido por sus aportes a la teoría de bases de datos relacionales

² Data entry: proceso de ingresar datos a una computadora para que éstos sean procesados

³ Howard Dresner: Investigador de Gartner Group que popularizó el acrónimo BI para referirse a Business Intelligence o Inteligencia de Negocios

que decir en su contra, que en algunos casos, agravaron el problema que pretendían resolver, ya que los sistemas de información carecían de integridad.

- 2000-2009: Business Intelligence 2.0. Se consolidan las aplicaciones de Business Intelligence en algunas plataformas (Oracle, SAP, IBM, Microsoft, etc.). A parte de la información estructurada, se empieza a considerar otro tipo de información y documentos no estructurados.

2.2. Diferencia entre datos, información y 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. (Davenport & Prusak, 1999).

- **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 porqué 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. Los datos pueden ser una colección de hechos almacenados en algún lugar físico como un papel, un dispositivo electrónico (CD, DVD, disco duro...), o la mente de una persona. En este sentido las tecnologías de la información han aportado mucho a recopilación de datos.

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*⁴ 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 Definición de Business Intelligence

Business Intelligence puede ser descrito como un concepto que integra por una parte el almacenamiento de grandes cantidades de datos, y por otra el procesamiento de los mismos, siendo el principal objetivo la transformación de estos en conocimiento y en decisiones en tiempo real a través del análisis y la exploración.

Otra definición nos dice:

*“BI es un **proceso interactivo** para **explorar y analizar información estructurada** sobre un **área** (normalmente almacenada en un **Data Warehouse**), para descubrir tendencias o patrones, a partir de los cuales derivar ideas y extraer conclusiones.*

*El proceso de Business Intelligence incluye la **comunicación** de los descubrimientos y **efectuar** los cambios. (Cano. 2007)*

⁴ La palabra Know How proviene del vocablo anglosajón Know How to do it, es decir saber cómo hacer algo o saber cómo hacerlo.

El concepto de Business Intelligence parte de la premisa de que, si se cuenta con un conjunto de datos lo suficientemente grande y se aplica un análisis a los mismos, se puede inferir conocimiento.

Una frase muy popular en el mundo de Business Intelligence enuncia: “Business Intelligence es el proceso de convertir datos en conocimiento, y el conocimiento en acción, para la toma de decisiones”. (Bernabeu, 2009)

Dicha afirmación se ve ilustrada perfectamente en la Ilustración 3.

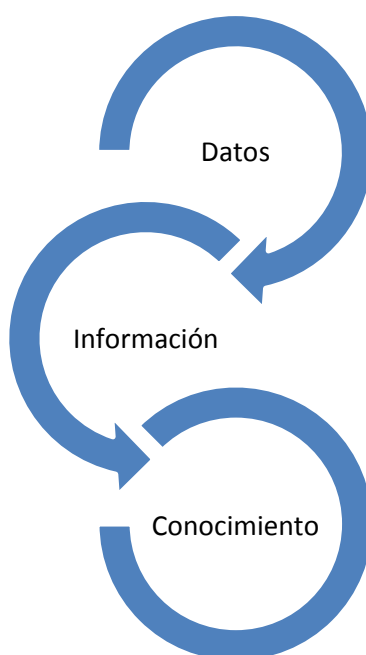


Ilustración 3 Transformación de Datos a Conocimiento

Fuente: Propia

Otro enunciado acerca de BI, que dice (Dario, 21): “Inteligencia de Negocios es el proceso de convertir datos en conocimiento y el conocimiento en acción, para la toma de decisiones”.

Business Intelligence se centra principalmente en los procesos de recolección y utilización efectiva de la información, al objeto de mejorar la forma de operar de una organización, ofreciendo a sus usuarios el acceso a la información que es clave a la

hora de llevar a cabo sus tareas habituales, con el fin de poder tomar decisiones oportunas basadas en datos certeros y correctos.

Al contar con información correcta y exacta en tiempo real, es posible identificar y corregir situaciones que, en un futuro, se pueden convertir en problemas para la organización, ofreciendo también la posibilidad de identificar nuevas oportunidades o readaptarse ante la ocurrencia de sucesos inesperados.

Cuanto más relevante y útil sea la inteligencia que posea una organización sobre un negocio (clientes, proveedores, socios, operaciones), mayor será su ventaja competitiva sobre sus competidores, pudiendo tomar mejores decisiones.

2.4 Ventajas y desventajas de utilizar un proceso de Inteligencia de Negocios

2.4.1 Ventajas

Uno de los objetivos básicos de los sistemas de información es que nos ayuden a la toma de decisiones. Cuando un responsable tiene que tomar una decisión pide o busca información, que le servirá para reducir la incertidumbre. Sin embargo, aunque todos la utilicen, no todos los responsables recogen la misma información: depende de muchos factores, como pueden ser su experiencia, formación, disponibilidad, etc. Del mismo modo, los responsables pueden necesitar recoger más o menos información dependiendo que su mayor o menor aversión al riesgo.

A partir de los datos que nos proporciona el sistema de *Business Intelligence* podemos descubrir conocimiento. Por ejemplo, en un concesionario de coches descubrimos la relación entre el número de visitas al concesionario y el número de vehículos vendidos en el mes siguiente.

Parece claro que el número de visitas al concesionario parece un indicador clave, pero ¿todos los concesionarios lo recogen?

Como hemos visto, *Business Intelligence* nos servirá como ayuda para la toma de decisiones y, posteriormente, para descubrir cosas que hasta ahora desconocíamos.

Los beneficios que se pueden obtener a través del uso de BI pueden ser de distintos tipos (Cano, 2007):

- **Beneficios tangibles:** por ejemplo: reducción de costes, generación de ingresos, reducción de tiempos para las distintas actividades del negocio.
- **Beneficios intangibles:** el hecho de que tengamos disponible la información para la toma de decisiones hará que más usuarios utilicen dicha información para tomar decisiones y mejorar la nuestra posición competitiva.
- **Beneficios estratégicos:** Todos aquellos que nos facilitan la formulación de la estrategia, es decir, a qué clientes, mercados o con qué productos dirigirnos.

(Dario, 21), cita algunos de los beneficios o ventajas de los cuales podemos mencionar algunos como por ejemplo:

- Reduce el tiempo mínimo que se requiere para recoger toda la información relevante de un tema en particular, ya que la misma se encontrará integrada en una fuente única de fácil acceso.
- Automatiza la asimilación de la información, debido a que la extracción y carga de los datos necesarios se realizará a través de procesos predefinidos.
- Proporciona herramientas de análisis para establecer comparaciones y tomar decisiones.
- Cierra el círculo que hace pasar de la decisión a la acción.
- Permite a los usuarios no depender de reportes o informes programados, porque los mismos serán generados de manera dinámica.
- Posibilita la formulación y respuesta de preguntas que son claves para el desempeño de la empresa.
- Permite acceder y analizar directamente los indicadores de éxito.

- Se pueden identificar cuáles son los factores que inciden en el buen o mal funcionamiento de la empresa.
- Se podrán detectar situaciones fuera de lo normal.
- Permitirá predecir el comportamiento futuro con un alto porcentaje de certeza, basado en el entendimiento del pasado.
- El usuario podrá consultar y analizar los datos de manera sencilla e intuitiva.

2.4.2 Desventajas

En la fuente (Students, 2011) nombre algunas de las desventajas más comunes que se pueden presentar en la implementación de un Data Warehouse y son:

- Requiere una gran inversión, debido a que su correcta construcción no es tarea sencilla y consume muchos recursos, de los 3 niveles fundamentales de la organización que son: Económicos, Humanos, y Materiales.
- No solo es capacitar a los recursos humanos para que hagan uso adecuado de esta herramienta, sino que también se ven afectados los recursos económicos porque si no se quiere estar de alguna forma “REZAGADO”, se debe contar con lo más novedoso y por ende los más costoso, además los recursos tecnológicos en ocasiones no permiten alcanzar los resultados esperados ya que habrán de ser sometidos al máximo rendimiento de sus funciones.
- Ya que se considera una tecnología con una evolución cambiante demasiado acelerada, esto debido a que los fabricantes de software y hardware ven un amplio mercado para sus “NOVEDADES”, por lo que no existe un regulación en precios, y en ocasiones el más caro no siempre es el mejor ni asegura el éxito a una empresa que lo implementa en sus TICS.
- Además, su misma implementación implica desde la adquisición de herramientas de consulta y análisis, hasta la capacitación de los, usuarios, y de aquellos operarios encargados de utilizarlas, por lo que las compañías no quieren correr el riesgo de que un operario experimentado lleve lo aprendido en la organización a otra, y que el conocimiento gestado pierda calidad y privacidad.

- Existe resistencia al cambio por parte de los usuarios esto por que constantemente se tienen dudas, la más común es: ¿Por qué cambiar si este modelo está funcionando de forma adecuada, cómoda y correcta?
- No existe una “Gestión” de cuando debemos cambiar nuestros sistemas de comunicación e información, por lo que los cambios aparentemente rápidos suelen ser tediosos para clientes y usuarios.
- Los beneficios del almacén de datos son apreciados en el mediano y largo plazo. Este punto deriva del anterior, y básicamente se refiere a que no todos los usuarios confiarán en el DW en una primera instancia, pero sí lo harán una vez que comprueben su efectividad y ventajas, si no se le da un seguimiento o servicio post uso, (post venta que permita verter sus opiniones) los clientes y usuarios no ven la denominada mejora continua.
- Además, su correcta utilización surge de la propia experiencia, por lo que las personas constantemente se sienten amenazadas cuando algún otro miembro es introducido al núcleo de trabajo, ya sea este de mayor o menor edad.
- Si se incluyen datos propios y confidenciales de clientes, proveedores, etc. El depósito de datos atentará contra la privacidad de los mismos, ya que cualquier usuario podrá tener acceso a ellos.
- El uso o acceso desmedido de los datos resulta ser molesto para los clientes quienes son “BOMBARDEADOS”, con propaganda indeseable o bien se sienten violentados hacia su intimidad al no tener poder de decisión en que si permiten o no se conozca su información particular.
- Infravaloración de los recursos necesarios para la captura, carga y almacenamiento de los datos.
- Recordar que la principal “MISIÓN”, del Ingeniero es el “HACER MÁS CON MENOS”, sacará el máximo rendimiento a los equipos y dejar de lado la denominada obsolescencia.
- Infravaloración del esfuerzo necesario para su diseño y creación.

- Incremento continuo de los requerimientos de los usuarios.
- Subestimación de las capacidades que puede brindar la correcta utilización del Data Warehouse y de las herramientas de BI en general, al final no todos quedan conformes.

2.5 Arquitectura típica de soluciones Business Intelligence



Ilustración 4 Arquitectura de una solución BI

Fuente: Propia

En general, la arquitectura típica de estas soluciones está basada en tres capas: BDD Relacional, BDD Multidimensional y Herramientas de visualización.

2.5.1 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.

2.5.2 Funciones de la Base de datos Multidimensional

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

2.5.3 Funciones del visualizador

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

2.6 Componentes de Business Intelligence

2.6.1 Data Warehouse

Una de las definiciones más famosas sobre DW, es la de Inmon, quien define: “Un Data Warehouse es una colección de datos orientada al negocio, integrada, variante en el tiempo y no volátil para el soporte del proceso de toma de decisiones de la gerencia”.

A primera vista y de acuerdo con la definición anterior se pensaría en una base de datos como un concepto similar a un Data Warehouse, pero en verdad no lo es, a continuación se nombra algunas diferencias.

Un Data Warehouse es una base de datos diferente a los 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

Debido a que Inmon, es reconocido mundialmente como el padre del DW, la explicación de las características más sobresalientes de este concepto se basó en su definición.

2.6.1.1 Características de un Data Warehouse

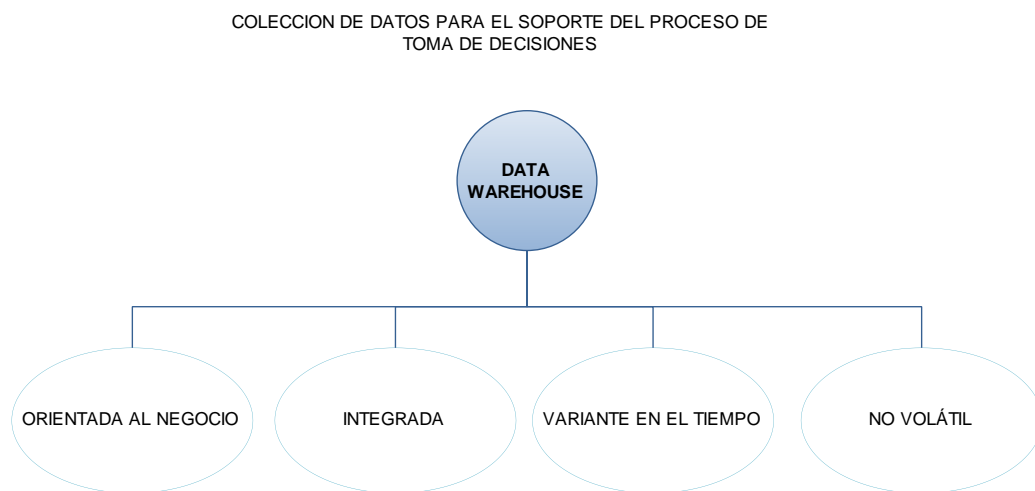


Ilustración 5 Características de un Data Warehouse

Fuente: Propia

1. Orientada al negocio.- La primera característica del DW, es que la información se clasifica en base a los aspectos que son de interés para la organización. Esta clasificación afecta el diseño y la implementación de los datos encontrados en el

almacén de datos, debido a que la estructura del mismo difiere considerablemente a la de los clásicos procesos operacionales orientados a las aplicaciones.

2. Integrada.- La integración implica que todos los datos de diversas fuentes que son producidos por distintos departamentos, secciones y aplicaciones, tanto internos como externos, deben ser consolidados en una instancia antes de ser agregados al DW, y deben por lo tanto ser analizados para asegurar su calidad y limpieza, entre otras cosas. A este proceso se lo conoce como Integración de Datos, y cuenta con diversas técnicas y subprocesos para llevar a cabo sus tareas. Una de estas técnicas son los procesos ETL ⁵ que más tarde se explicará en detalle.

3. Variante en el tiempo. - Debido al gran volumen de información que se manejará en el DW, cuando se le realiza una consulta, los resultados deseados demorarán en originarse. Este espacio de tiempo que se produce desde la búsqueda de datos hasta su consecución es del todo normal en este ambiente y es, precisamente por ello, que la información que se encuentra dentro del depósito de datos se denomina de tiempo variable.

Es importante tener en cuenta la granularidad⁶ de los datos, así como también la intensidad de cambio natural del comportamiento de los fenómenos de la actividad que se desarrolle, para evitar crecimientos incontrolables y desbordamientos de la base de datos.

El intervalo de tiempo y periodicidad de los datos debe definirse de acuerdo con la necesidad y requisitos de los usuarios.

4. No volátil.- La información es útil para el análisis y la toma de decisiones solo cuando es estable. Los datos operacionales varían momento a momento, en cambio, los datos una vez que entran en el DW no se modifican.

5 ETL.- 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.

6 Granularidad.- La granularidad representa el nivel de detalle al que se desea almacenar la información sobre el negocio que se esté analizando.

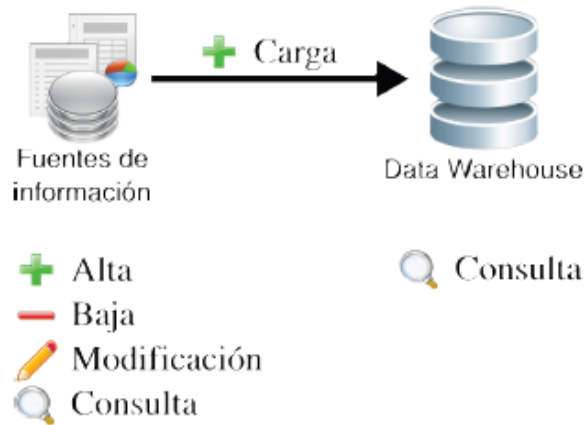


Ilustración 6 Data Warehouse no volátil

Fuente: (Dario, 21)

La actualización, o sea, insertar, eliminar y modificar, se hace de forma muy habitual en el ambiente operacional sobre una base, registro por registro, en cambio en el depósito de datos la manipulación básica de los datos es mucho más simple, debido a que solo existen dos tipos de operaciones: la carga de datos y el acceso a los mismos.

Por esta razón es que en el DW no se requiere mecanismos de control de concurrencia y recuperación.

2.6.1.2 Arquitectura del Data Warehouse

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 centramos en los cambios que ha sufrido el DW con respecto a la arquitectura del sistema.

El concepto inicial detrás del DW es el 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 DW y que todas las extracciones y transformaciones dentro de toda la organización estuvieran bajo el control de un único proceso.

Los problemas en proyectos que aplicaron este concepto, fueron apareciendo. Construir un DW global y 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 pertenencias 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 requieren consultar su información al mismo tiempo dentro del DW. 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.

2.6.2 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 tienen necesidades de información limitada, y aunque típicamente existen requerimientos para análisis funcionales cruzados, el tamaño de los mismos es reducido materialmente si limitamos el tamaño del DW en sí mismo.

Los principales beneficios de utilizar Data Marts 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 de hardware
- Permite el acceso a los datos por medio de un gran número de herramientas del mercado, logrando independencia de estas.

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

2.6.2.1 Data Marts Dependientes

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

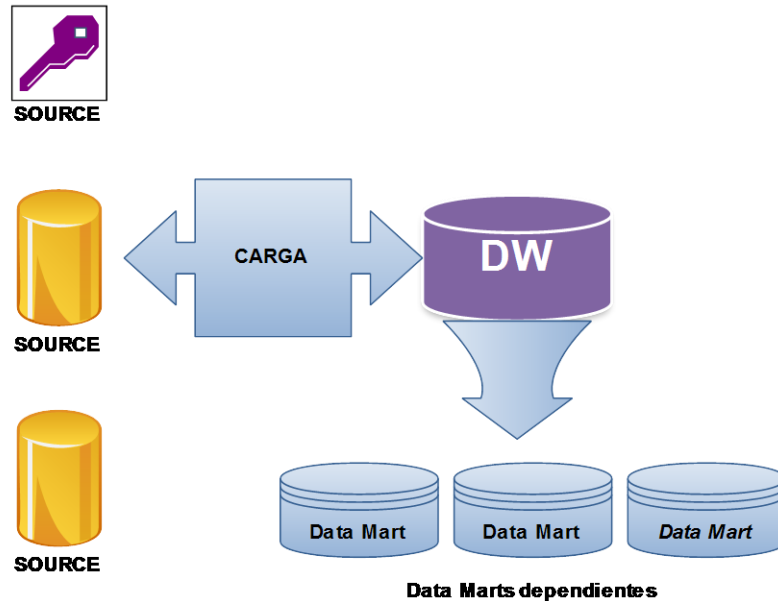


Ilustración 7 Arquitectura Data Marts Dependientes

Fuente: Propia

Esta solución resuelve los problemas de performance, estrategia, finanzas e incluso algunos de los problemas políticos. Aunque tienen estos puntos a favor, siguen teniendo que construir el DW global antes que los data marts sean implementados.

2.6.2.2 Data Marts Independientes

Esta óptica es considerada por muchos como una alternativa del DW 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 data marts pequeños, otros data 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 DW 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.

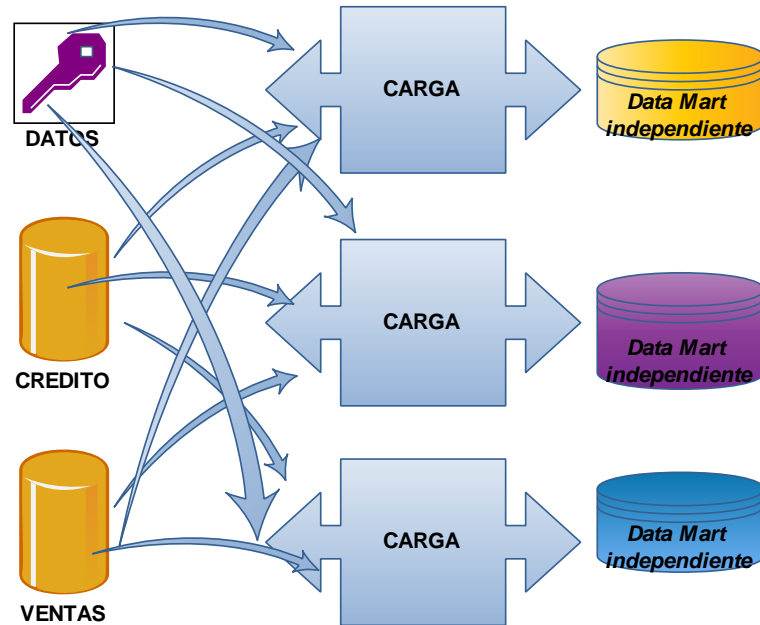


Ilustración 8 Arquitectura Data Marts Independientes

Fuente: Propia

2.6.3 Procesamiento Analítico en Línea (OLAP)

La tecnología de OLAP permite un uso más eficaz de los DW para el análisis de datos en línea, lo que proporciona respuestas rápidas a consultas analíticas complejas e iterativas utilizadas 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 la organización 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 realizar el cálculo al momento de la solicitud. La combinación de navegación fácil y rápida le permita 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.

2.6.3.1 Características principales del OLAP

- **Rápido:** proporciona la información al usuario a una velocidad constante. La mayoría de las peticiones de deben 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 en “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 proporciona 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 esta resida y no esté limitada por el volumen.

2.6.4 Procesamiento Transaccional en Línea (OLTP)

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 operaciones.

- El acceso a los datos está optimizado por tareas frecuentes de lectura y escritura. (Por ejemplo, la enorme cantidad de transacciones que tienen que soportar las BD de bancos o supermercados 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.

2.6.4.1 Diferencias entre Procesamiento Transaccional en Línea (OLPT) y Procesamiento Analítico en Línea (OLAP)

La tabla que a continuación se muestra indica las diferencias que existe entre OLTP y OLAP.

OLTP	OLAP
Admite el acceso simultáneo de muchos –miles- de usuarios que agregan y modifican datos	Admite el acceso simultáneo de muchos –cientos- de usuarios 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, incluidos los datos extensivos utilizados para comprobar transacciones	Contienen grandes cantidades de datos, sumariados, consolidados y transformados. También en detalle pero solo los necesarios para el análisis.
Tienen estructuras de bases de datos complejas.	Tienen estructuras de bases de datos simples.
Se ajustan para dar respuesta a la actividad transaccional.	Se ajustan para dar respuesta a la actividad de consulta.
Proporcionan la infraestructura tecnológica necesaria para admitir las operaciones diarias de la empresa.	Proporcionan la infraestructura tecnológica necesaria para admitir análisis de datos de la empresa.
Los analistas carecen de la experiencia técnica necesaria para crear consultas “ad-hoc” contra la compleja estructura de los 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	Contienen datos transformados que son válidos, coherentes, consolidados

analíticas complejas puede ser lento o impredecible, lo que causa un servicio poco eficiente a los usuarios del proceso analítico en línea.	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.

Tabla 1 Diferencias entre OLAP y OLTP

Fuente: Propia

2.7 Base de datos Multidimensional

Una base de datos multidimensional es una base de datos en donde su información se almacena en forma multidimensional, es decir, en varias dimensiones a través de tablas de hechos y tablas de dimensiones.

Proveen una estructura que permite, a través de la creación y consulta a una estructura de datos determinada generalmente cubos multidimensionales, tener acceso flexible a los datos, para explorar y analizar sus relaciones, y consiguientes resultados.

Las bases de datos multidimensionales implican tres variantes posibles de modelamiento, que permiten realizar consultas de soporte de decisión:

- Esquema en estrella (Star Scheme).
- Esquema copo de nieve (Snowflake Scheme).
- Esquema constelación o copo de estrellas (Starflake Scheme).

Los mencionados esquemas pueden ser implementados de diversas maneras, que, independientemente al tipo de arquitectura, requieren que toda la estructura de datos este desnormalizada o semi desnormalizada, para evitar desarrollar uniones (Join) complejas para acceder a la información, con el fin de agilizar la ejecución de consultas. Los diferentes tipos de implementación son los siguientes:

- Relacional – ROLAP
- Multidimensional – MOLAP
- Híbrido – HOLAP

2.7.1 Tablas de dimensiones

Las tablas de dimensiones definen como están organizados lógicamente los datos y proveen el medio para analizar el contexto del negocio. Contienen datos cualitativos. Representan los aspectos de interés, mediante los cuales los usuarios podrán filtrar y manipular la información almacenada en la tabla de hechos.

En la siguiente figura se pueden apreciar algunos ejemplos:

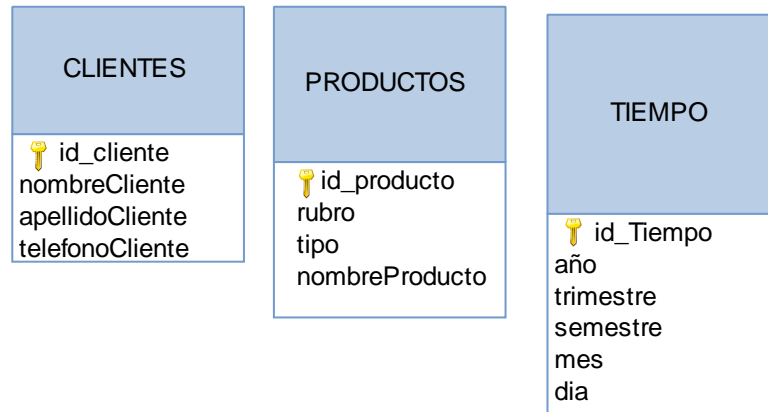


Ilustración 9 Ejemplo de tablas de dimensiones

Fuente: Propia

Como se puede ver en el gráfico anterior, cada tabla posee un identificador único (primary key) y al menos un campo o atributo de referencia que describe los criterios de análisis relevantes para la organización, estos son por lo general de tipo texto.

Usualmente la cantidad de tablas de dimensiones, aplicadas a un tema de interés en particular, varían entre tres y quince.

Es recomendable manejar un sistema de claves en el DW (Claves Subrogadas) totalmente diferente al de los OLTP, ya que si estos últimos son recodificados, el almacén quedaría inconsistente y debería ser poblado nuevamente en su totalidad.

2.7.2 Tabla dimensión Tiempo

En un DW, la creación y el mantenimiento de una tabla de dimensión Tiempo es obligatoria, y la definición de granularidad y estructuración de la misma depende de la dinámica del negocio que se esté analizando. Toda la información dentro del depósito, como ya se ha explicado, posee su propio sello de tiempo que determina la ocurrencia de un hecho específico, representando de esta manera diferentes versiones de una misma situación.

Es importante tener en cuenta que la dimensión tiempo no es solo una secuencia cronológica representada de forma numérica, sino que mantiene niveles jerárquicos especiales que inciden notablemente en las actividades de la organización. Esto se debe a que los usuarios podrán por ejemplo analizar los socios ingresados en determinado período año, trimestre, mes, semana, día, etc.

2.7.3 Tablas de hechos

Las tablas de hechos contienen, precisamente, los hechos que serán utilizados por los analistas de negocio para apoyar el proceso de toma de decisiones. Contienen datos cuantitativos.

Los hechos son datos instantáneos en el tiempo, que son filtrados, agrupados y explorados a través de condiciones definidas en las tablas de dimensiones.

Una tabla de hecho posee una clave primaria que está compuesta por las claves primarias de las tablas de dimensiones relacionadas a este.

La ilustración siguiente puede aclarar el concepto de una tabla de hechos.

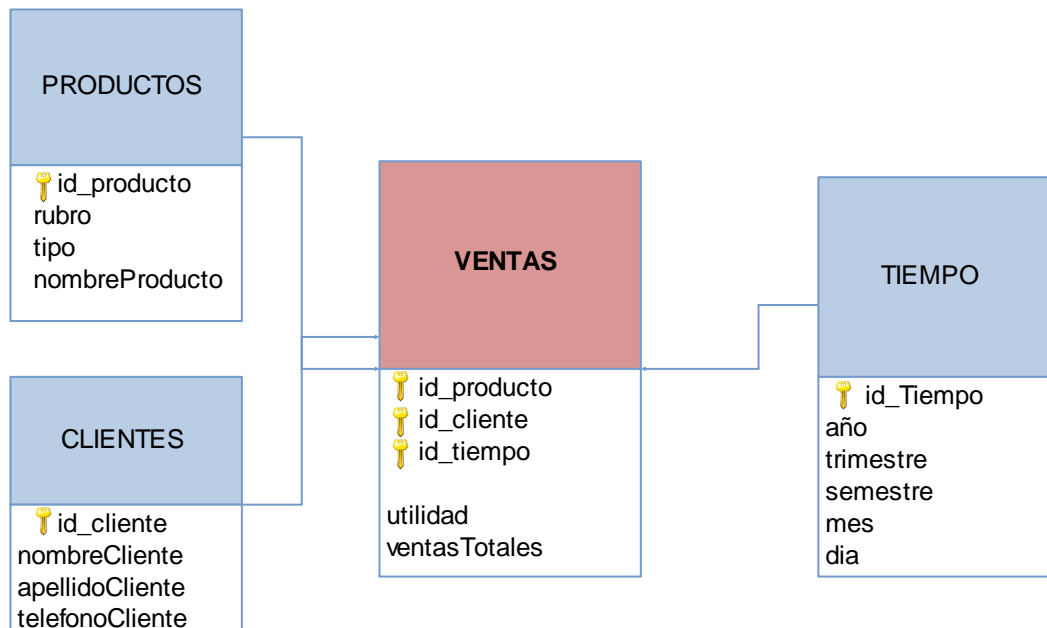


Ilustración 10 Tabla de hechos ventas

Fuente: Propia

Como se muestra en la ilustración la tabla de hechos VENTAS y a los costados de esta tabla se encuentran las tablas de dimensiones CLIENTES, TIEMPO Y PRODUCTOS que se encuentran conectadas a la tabla VENTAS por las claves principales de las tablas de dimensiones. Los hechos en este ejemplo son utilidad y ventas Totales.

Los hechos son aquellos datos que residen en una tabla de hechos y que son utilizados para crear indicadores, a través de sumalizaciones ⁷ preestablecidas al momento de crear un cubo multidimensional.

⁷ Las sumalizaciones no están referidas solo a sumas, sino también a promedios, mínimos, máximos, totales por sector, porcentajes, fórmulas predefinidas, etc, dependiendo de los requerimientos de información del negocio.

2.7.4 Tipos de modelamiento de un Data Warehouse

2.7.4.1 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 (dimensión table) para 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 las otras tablas). El resto de las 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.

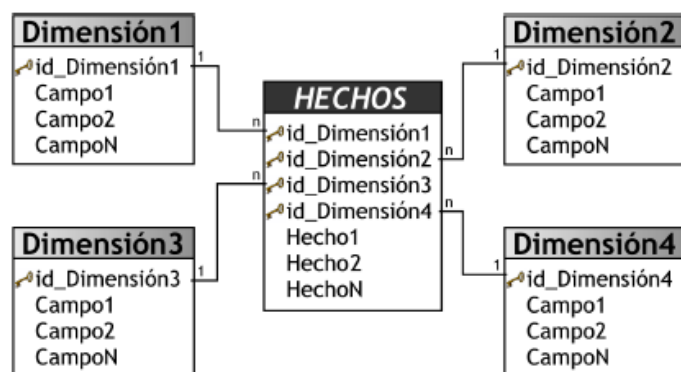


Ilustración 11 Esquema en estrella

Fuente: (Dario, 21)

La mayoría de los data Warehouse están diseñados en base al esquema en estrella para representar el sistema de datos multidimensional. El sistema en estrella se caracteriza por tener una o más tablas de hechos que contienen la información principal del Data Warehouse, y un número indeterminado de tablas de dimensión.

Cada una de las tablas de dimensión contiene información sobre las entradas (tuplas) de un determinado atributo en la tabla de hechos. Cada tabla de dimensión está relacionada con la tabla de hechos mediante el sistema clave primaria – clave ajena. Las dimensiones no se relacionan entre sí. Una tabla de hechos contiene claves y medidas.

2.7.4.2 Esquema de copo de nieve

Este esquema representa una extensión del modelo en estrella cuando las tablas de dimensiones se organizan en jerarquías de dimensiones.

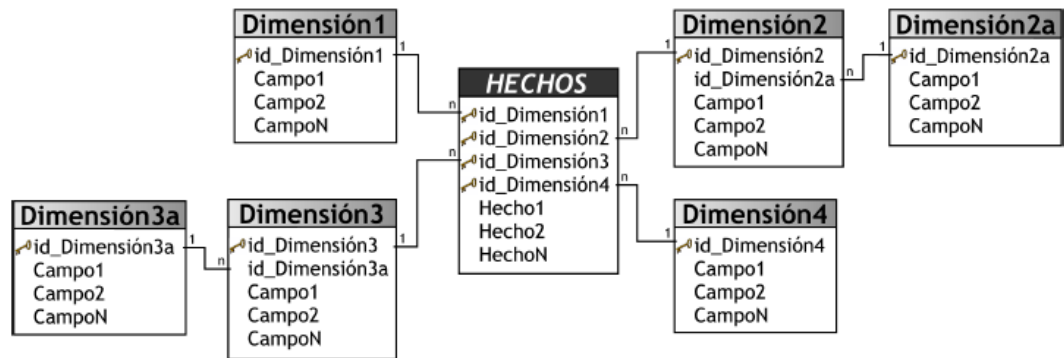


Ilustración 12 Esquema copo de nieve

Fuente: (Dario, 21)

En este esquema se puede decir que es el más parecido al modelo entidad relación ya que las tablas de dimensiones están normalizadas.

La finalidad de normalizar las tablas es reducir el espacio de almacenamiento al eliminar la redundancia de datos; pero tiene la contrapartida de generar peores rendimientos al tener que crear más tablas de dimensiones y más relaciones entre las tablas (*JOINS*) lo que tiene un impacto directo sobre el rendimiento.

2.7.4.3 Esquema constelación

Este esquema es más complejo que las otras arquitecturas debido a que contiene múltiples tablas de hechos. Con esta solución las tablas de dimensiones pueden estar compartidas entre varias tablas de hechos.

El esquema de constelación de hechos tiene mucha flexibilidad y esta característica es su gran virtud.

Sin embargo, el problema es que cuando el número de las tablas vinculadas aumenta, la arquitectura puede llegar a ser muy compleja y difícil para mantener.

En la siguiente ilustración se indica la conformación de un esquema tipo constelación.

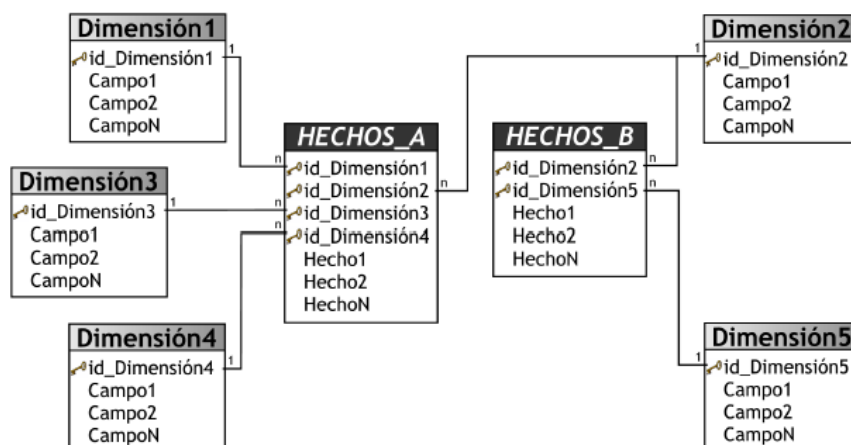


Ilustración 13 Esquema constelación

Fuente: (Dario, 21)

2.7.5 Tipos de implementación de un Data Warehouse

La implementación de un Data Warehouse basa principalmente en la tecnología OLAP que permite trabajar sobre los datos al pensarlos como cubos multidimensionales o hipercubos con información de la empresa. Existen diferentes maneras en las que se puede implementar un Data Warehouse con la única diferencia que se transa menor tiempo de acceso a cambio de mayor espacio de utilización de disco y viceversa.

Existen tres modelos posibles:

- ROLAP
- MOLAP
- HOLAP

2.7.5.1 ROLAP

Significa Procesamiento Analítico OnLine Relacional, es decir, se trata de sistemas y herramientas OLAP (*Procesamiento Analítico OnLine*) construidos sobre una base de datos relacional. Es una alternativa a la tecnología MOLAP (Multidimensional OLAP) que se construye sobre bases de datos multidimensionales. Ambos tipos de herramientas, tanto ROLAP como MOLAP, están diseñadas para realizar análisis de datos a través del uso de modelos de datos multidimensionales, aunque en el caso de ROLAP estos modelos no se implementan sobre un sistema multidimensional, sino sobre un sistema relacional clásico.

2.7.5.2 MOLAP

Significa procesamiento analítico multidimensional en línea. En esta implementación los datos se almacenan en forma multidimensional permitiendo operaciones rápidas de búsquedas y resúmenes gracias a los datos precalculados existentes. Como limitante, ocupa más espacio en disco para almacenamiento, pero como se sabe el costo de espacio en disco es cada vez más barato.

Consta de dos niveles, el nivel de la base de datos multidimensional (MDDDB) que se encarga de manejar, acceder y obtener los datos y del motor analítico que ejecuta la consulta de los usuarios.

2.7.5.3 HOLAP

Significa procesamiento analítico en línea híbrido y es una combinación de ROLAP y MOLAP. HOLAP permite almacenar una parte de datos como ROLAP y otra como MOLAP. Utiliza las características de MOLAP para mejorar el acceso a las consultas y ROLAP para optimizar el tiempo en que se procesa el cubo.

2.7.6 Cubo Multidimensional.

Un cubo multidimensional, representa o convierte los datos planos que se encuentran en filas y columnas, en una matriz de N dimensiones.

Los objetos más importantes que se pueden incluir en un cubo multidimensional, son los siguientes:

- **Indicadores:** sumalizaciones que se efectúan sobre algún hecho o expresiones basadas en sumalizaciones, pertenecientes a una tabla de hechos.
- **Atributos:** campos o criterios de análisis, pertenecientes a tablas de dimensiones.
- **Jerarquías:** representa una relación lógica entre dos o más atributos.

De esta manera en un cubo multidimensional, los atributos existen a lo largo de varios ejes o dimensiones, y la intersección de las mismas representa el valor que tomará el indicador que se está evaluando.

En la siguiente representación matricial se puede ver más claramente lo que se acaba de decir.

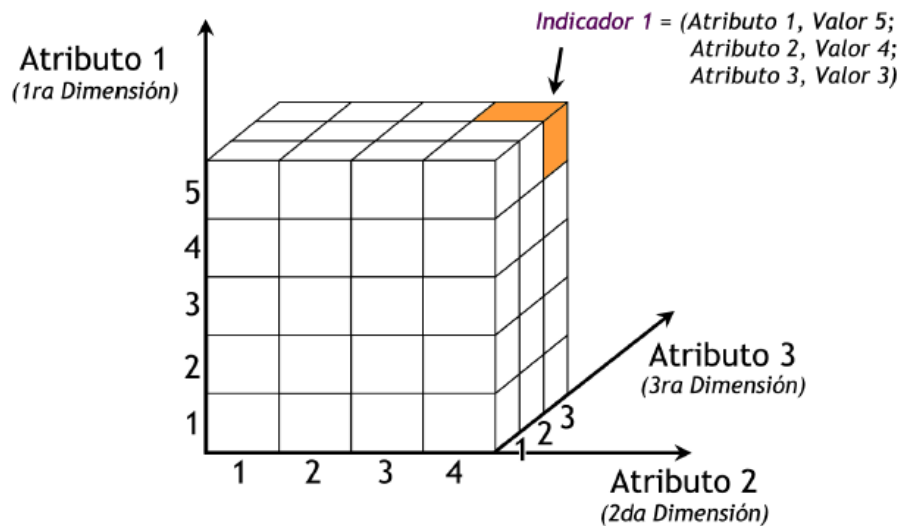


Ilustración 14 Cubo Multidimensional

Fuente: (Dario, 21)

Para la creación del cubo de la figura anterior, se definieron tres Atributos (“Atributo 1”, “Atributo 2” y “Atributo 3”) y se definió un Indicador (“Indicador 1”).

Entonces el cubo quedó compuesto por 3 dimensiones o ejes (una por cada Atributo), cada una con sus respectivos valores asociados. También, se ha seleccionado una intersección al azar para demostrar la correspondencia con los valores de las Atributos. En este caso, el indicador “Indicador 1”, representa el cruce del Valor “5” de “Atributo 1”, con el Valor “4” de “Atributo 2” y con el Valor “3” de “Atributo 3”.

Se puede observar, que el resultado del análisis está dado por los cruces matriciales de acuerdo a los valores de las dimensiones seleccionadas.

Más específicamente, para acceder a los datos del DW, se pueden ejecutar consultas sobre algún cubo multidimensional previamente definido. Dicho cubo debe incluir entre otros objetos: indicadores, atributos, jerarquías, etc., basados en los campos de las tablas de dimensiones y de hechos, que se deseen analizar. De esta manera, las consultas son respondidas con gran performance, minimizando al máximo el tiempo que se hubiese incurrido en realizar dicha consulta sobre una base de datos transaccional.

2.7.6.1 Indicadores

Los indicadores son sumalizaciones efectuadas sobre algún hecho o expresiones basadas en sumalizaciones, que serán incluidos en algún cubo multidimensional, con el fin de analizar los datos almacenados en el DW. El valor que estos adopten estará condicionado por los atributos/jerarquías que se utilicen para analizarlos

Los indicadores, además de hechos, pueden estar compuestos por otros indicadores, pero no ambos simultáneamente. Pueden utilizarse para su creación funciones de sumalización (suma, conteo, promedio, etc.), funciones matemáticas, estadísticas, operadores matemáticos y lógicos.

2.7.6.2 Atributos

Los atributos constituyen los criterios de análisis que se utilizarán para analizar los indicadores dentro de un cubo multidimensional. Los mismos se basan, en su gran mayoría, en los campos de las tablas de dimensiones y/o expresiones.

Dentro de un cubo multidimensional, los atributos son los ejes del mismo.

2.7.6.3 Jerarquías

Una jerarquía representa una relación lógica entre dos o más atributos pertenecientes a un cubo multidimensional; siempre y cuando posean su correspondiente relación “padre-hijo”.

Las jerarquías poseen las siguientes características:

- Pueden existir varias en un mismo cubo.
- Están compuestas por dos o más niveles.
- Se tiene una relación “1-n” o “padre-hijo” entre atributos consecutivos de un nivel superior y uno inferior.

Por lo general, las jerarquías pueden identificarse fácilmente, debido a que existen relaciones “1-n” o “padre-hijo” entre los propios atributos de un cubo.

La principal ventaja de manejar jerarquías, reside en poder analizar los datos desde su nivel más general al más detallado y viceversa, al desplazarse por los diferentes niveles.

La siguiente figura muestra un pequeño ejemplo de lo recién expuesto:

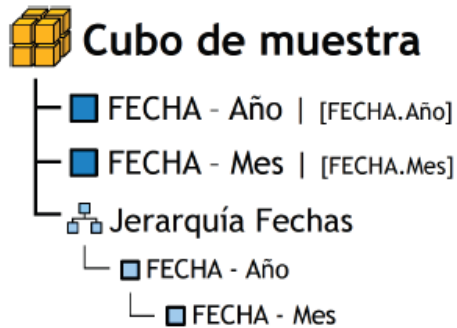
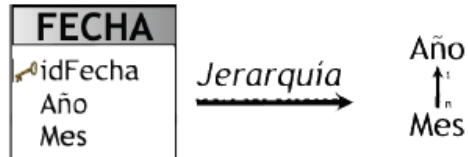


Ilustración 15 Jerarquía Fecha

Fuente: (Dario, 21)

Un ejemplo muy práctico y que permite visualizar los conceptos que hasta el momento se ha revisado es el siguiente.

Se tomará como ejemplo el esquema de estrella anteriormente visto.

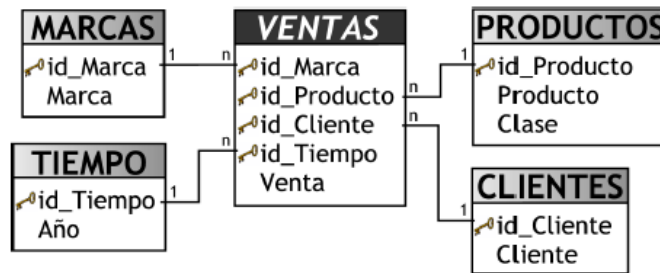


Ilustración 16 Esquema estrella de ejemplo

Fuente: (Dario, 21)

Para comenzar se creará el cubo multidimensional llamado “Cubo de Ventas”. Gráficamente lo veremos así.



Ilustración 17 Cubo de ventas

Fuente: (Dario, 21)

Luego se crearán dos atributos:

De la tabla de dimensión “PRODUCTOS”, se tomará el campo “Producto” para la creación del atributo denominado:

- “PRODUCTOS - Producto”.

De la tabla dimensión “MARCAS”, se tomará el campo “Marca” para la creación del atributo denominado:

- “MARCAS - Marca”.

Gráficamente:

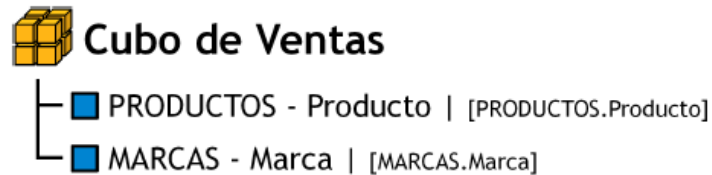


Ilustración 18 Cubo de ventas con atributos

Fuente: (Dario, 21)

También se creará un indicador.

De la tabla de hechos "VENTAS", se sumará el hecho "Venta" para crear el indicador denominado:

- "VENTAS - Venta".

La fórmula utilizada para crear este indicador es la siguiente:

- "VENTAS - Venta" = SUM(VENTAS.Venta).

Gráficamente:

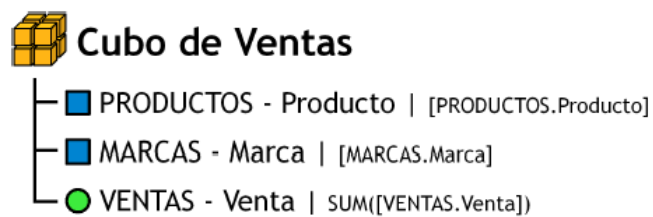


Ilustración 19 Cubo de ventas con indicador

Fuente: (Dario, 21)

En este instante tenemos un cubo de ventas con dos dimensiones, PRODUCTOS y MARCAS, cuya representación matricial es la siguiente:

PRODUCTOS - Producto	P1	40	25	60
	P2	21	55	45
	P3	13	32	43
		M1	M2	M3
		MARCAS - Marca		

Ilustración 20 Representación matricial Cubo de ventas

Fuente: (Dario, 21)

La intersección de los ejes representa las ventas de cada producto con su respectiva marca (indicador “VENTAS - Venta”).

Por ejemplo:

- Las ventas asociadas al producto “P1” y a la marca “M1” son 40.
- Las ventas asociadas al producto “P1” y a la marca “M2” son 25.
- Las ventas asociadas al producto “P1” y a la marca “M3” son 60.

CAPÍTULO III

PRODUCTOS DE BUSINESS INTELLIGENCE



Los principales productos de Business Intelligence que existen en la actualidad son:

- Cuadros de mando Integrales (CMI)
- Sistema de soporte a la decisión (DDS)
- Sistema de información ejecutiva (EIS)

3.1 Cuadros de mando integrales (CMI)

El cuadro de mando integral o más conocido como Balance Scorecard 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.

3.1.1 Diferencias con otras herramientas de Bussines Intelligence

El cuadro de mando integral se diferencia de otras herramientas de BI, como los Sistemas de Soporte a la Decisión (DSS) o los sistemas de información Ejecutiva, en que está orientado al seguimiento de indicadores que al análisis minucioso de información.

Por otro lado, es muy común que un CMI sea controlador por la dirección general de una empresa, frente a otras herramientas de BI 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 selecciona los indicadores del CMI.

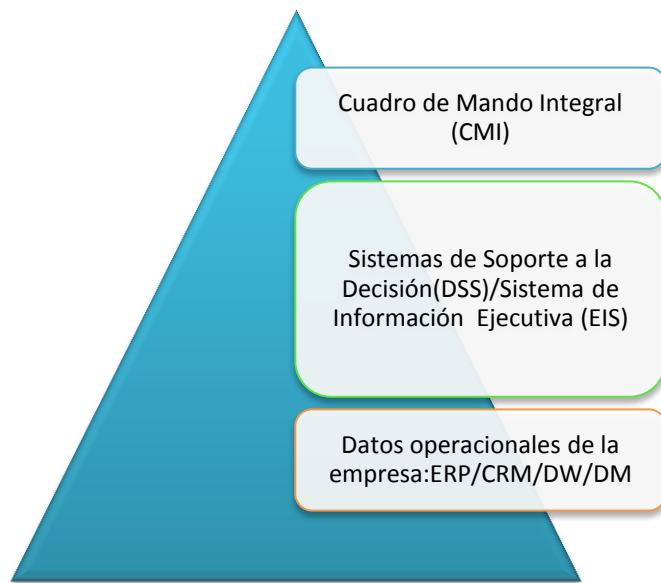


Ilustración 21 Diferencias con otras herramientas BI

Fuente: Propia

3.2 Tipos de cuadro 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 implementación y puesta en marcha es más sencilla y rápida. Un CMO debería estar siempre ligado a un DSS para indagar en profundidad sobre sus datos.

El **Cuadro de Mando Integral (CMI)**, por el contrario, representa la ejecución de la estrategia de la 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 CMI 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 tiene 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 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.
- Las perspectivas de aprendizaje y crecimiento son las últimas que se plantean 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

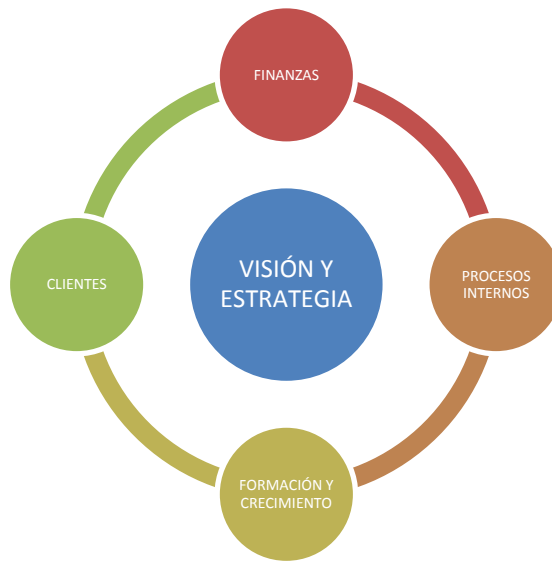


Ilustración 22 Cuadro de Mando Integral

Fuente: Propia

3.3 Sistemas de Soporte a la Decisión

Un Sistema de Soporte a la Decisión (DSS) es una herramienta de Business Intelligence enfocada al análisis de los datos de una organización. Conjunto de programas y herramientas que ayudan a la toma de decisiones de los administradores al combinar datos, modelos analíticos sofisticados y software amigable en un solo sistema que puede dar soporte a la toma de decisiones semi-estructuradas y no estructuradas.

3.3.1 Tipos de Problemas

- Problemas no estructurados

Estas decisiones no cuentan con un procedimiento definido para tomarlas.

- Problemas estructurados.

Son repetitivos, rutinarios y cuentan con un procedimiento definido para tomarlas, de esta forma cada vez que se presentan no se manejan como si fueran nuevos.

- Problemas semi-estructurados.

En este tipo de decisiones sólo parte del problema tiene una respuesta ya definida proporcionada por un procedimiento que es aceptado.

3.3.2 Funciones y Características

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:

- **Interactividad.**- Sistema computacional con la posibilidad de interactuar en forma amigable y con respuestas a tiempo real con el encargado de tomar decisiones.
- **Tipo de Decisiones.**- Apoya el proceso de toma de decisiones estructuradas y no estructuradas.
- **Frecuencia de Uso.**-Tiene una utilización frecuente por parte de la administración media y alta para el desempeño de su función.

- **Variedad de Usuarios.**- Puede emplearse por usuarios de diferentes áreas funcionales como ventas, producción, administración, finanzas y recursos humanos.
- **Flexibilidad.**-Permite acoplarse a una variedad determinada de estilos administrativos: Autocráticos, Participativos, etc.
- **Desarrollo.**-Permite que el usuario desarrolle de manera directa modelos de decisión sin la participación operativa de profesionales en informática.
- **Interacción Ambiental.**- Permite la posibilidad de interactuar con información externa como parte de los modelos de decisión.
- **Comunicación Inter-Organizacional.**- Facilita la comunicación de información relevante de los niveles altos a los niveles operativos y viceversa, a través de gráficas.
- **Acceso a Bases de Datos.**-Tiene la capacidad de acceder información de las bases de datos corporativos.
- **Simplicidad.**-Simple y fácil de aprender y utilizar por el usuario final.

3.3.3 Objetivos de los DSS

- Ayudar a los gerentes a tomar decisiones para resolver problemas semi-estructuradas
- Apoyar el juicio del gerente en lugar de tratar de remplazarlo.
- Mejorar la eficacia del gerente en la toma de decisiones, más que su eficiencia.

3.3.4 Ventajas de los DSS

- Menores Costos
- Disponibilidad de gran variedad de herramientas en el mercado que operan en el ambiente de microcomputadoras
- Muy baja dependencia de personas que se encuentran fuera de control del encargado de la toma de decisiones
- Mejor uso de los datos.
- Ahorro de costos.
- Mejores decisiones.
- Respuesta rápida a situaciones inesperadas.
- Equipo de trabajo más efectivo.
- Ahorro de tiempo.
- Mejora la comunicación al interior de la organización.
- Incrementa la creatividad en la toma de decisiones
- Análisis temporales.

3.3.5 Desventajas de los DSS

- Necesidad de mayor conocimiento de uso por parte del encargado de la toma de decisiones.
- Herramientas más costosas.

3.3.6 Arquitecturas

Una vez más, diferentes autores identifican diferentes componentes para un DSS; los cuales se muestran a continuación:

- **El sistema de gestión de base de datos.**
Almacena información de diversos orígenes, puede proceder de los repositorios de datos de una organización tradicional, de fuentes externas (como Internet), o del personal (de ideas y experiencias de los usuarios individuales).
- **El sistema gestor de modelos.**
Se ocupa de las representaciones de los acontecimientos, hechos o situaciones utilizando varios tipos de modelos (dos ejemplos serían modelos de optimización y modelos de búsqueda-objetivo).
- **El sistema gestor y generador de diálogos.**
Se trata de la interfaz de usuario; es, por supuesto, el componente que permite a un usuario interactuar con el sistema.

3.3.7 Factores para el Éxito de un DSS

- Capacitación
- Involucramiento
- Experiencia de los usuarios
- Apoyo de la alta dirección
- Nivel de utilización
- Novedad de la aplicación

3.3.8 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 (data warehouse o data mart), mostrando informes muy dinámicos y con gran potencial de navegación, pero siempre con una interfaz gráfica amigable, vistosa y sencilla.



Ilustración 23 Vista de un DSS

Fuente: (Dario, 21)

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.

3.4 Sistemas de Información Ejecutiva

Un sistema de Información Ejecutiva (Executive information system, EIS) es una herramienta de Inteligencia empresarial (Business Intelligence), orientada a usuarios de nivel gerencial, que permite monitorizar el estado de las variables de un área o unidad de la empresa a partir de información interna y externa a la misma, y 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.

“Un EIS es un tipo de un DSS (Sistema de Soporte a las Decisiones).”

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.

3.4.1 Características

Un buen sistema de información para ejecutivos presenta información en forma de gráficos, columnas y textos. La capacidad para hacer gráficos se necesita para facilitar en el análisis rápido de las condiciones y tendencias corrientes; las tablas presentan mayor detalle y permiten el análisis de variaciones; la información de textos añade interpretaciones y detalles de los datos. Las principales características de los sistemas de información para ejecutivos (EIS) son las siguientes:

- Están diseñados para cubrir las necesidades específicas y particulares de la alta administración de la empresa.

- Extraen, filtran, comprimen y dan seguimiento a información crítica del negocio.
- Implica que los ejecutivos puedan interactuar en forma directa con el sistema sin el apoyo o auxilio de intermediarios.
- Es un sistema desarrollado con altos estándares en sus interfaces hombre-máquina, caracterizado por gráficas de alta calidad, información tabular y en forma de texto.
- Capacidades de análisis de datos, tales como hoja electrónica de cálculo, lenguajes especializados de consulta que utilicen comandos de SQL.
- Herramientas para la organización personal del ejecutivo, tales como calendario, agenda y tarjetero electrónico.

3.4.2 Desarrollo

El proceso de desarrollo de un EIS tiene características que lo hacen único. En primera instancia, porque es el primer sistema que se desarrolla en la empresa dirigido al ejecutivo; quien es el usuario de este sistema.

En segundo lugar, las técnicas utilizadas para el análisis y desarrollo de los tradicionales Sistemas Transaccionales no necesariamente funcionan en un 100% de manera similar durante el desarrollo de un EIS. Es importante tomar en cuenta los tres pasos a planificar para construir un sistema EIS, los cuales son:

- Adquisición de datos
- Modelización
- Presentación

3.4.3 Factores de éxito

Se supone que los EIS están hechos a la medida para ser usados por ejecutivos de alto nivel; sin embargo, pocos son los que los usan. Usualmente los sistemas que apoyan decisiones son difíciles de justificar usando métodos económicos estándar de evaluación. Un EIS es comúnmente desarrollado con altas expectativas de éxito, sin embargo terminan fracasando, y las razones principales son, según algunos investigadores basados en los siguientes criterios de evaluación:

- **Acceso.**

Es obvio que si un sistema se encuentra disponible significa que los usuarios tendrán acceso a él, de una manera no complicada y desde puntos clave. Si esto no sucede, el sistema no servirá de nada.

- **Uso.**

Un indicador importante es su frecuencia de uso. Si un sistema no es usado, o simplemente, los usuarios potenciales no lo emplean, esto se reflejará en el éxito del sistema.

- **Satisfacción.**

Si el sistema no puede satisfacer a sus usuarios, éstos no usarán el sistema

- **Impacto Positivo.**

Un sistema es exitoso si tiene un impacto benéfico en los ejecutivos y la organización, al hacer los primeros, mejores decisiones la organización recibe más beneficios.

- **Difusión.**

Otro punto que indica el éxito es la propagación del sistema. El número de personas que usa el sistema aumenta después de que los usuarios iniciales lo han probado. Para que un EIS tenga éxito es necesario que cumpla con los siguientes factores:

- Que se vea bien.
- Que sea relevante.
- Que sea rápido.
- Que la información esté disponible y actualizada.

3.4.4 Factores de fracaso

Los siguientes son los factores considerados como " factores de fracaso" en los sistemas de Información para Ejecutivos:

- **Tecnológicos.**

El EIS no es mejor que el sistema original. La tecnología no es adecuada o es inapropiada. Las interfaces son complicadas o los menús son extensos.

- **Soporte**

Los requerimientos de información de los usuarios fueron ignorados. Los cambios en las necesidades de información de los usuarios no fueron llevados a cabo. No hay mejoras en los reportes electrónicos. Inadecuado conocimiento del negocio entre los miembros del equipo de soporte. No ataca los problemas significativos de la empresa. Falta de disponibilidad de los datos. Entrega tardía de datos operativos.

- **Relacionados con el Usuario**

- a. Limitante de enfoque del EIS a un usuario.
- b. Falta de compromiso de los usuarios.
- c. Falta de claridad del propósito del EIS en el ejecutivo patrocinador.

- d. No proveer los medios a los ejecutivos para comunicar ideas.
- e. Usuarios no capaces de comunicar decisiones
- f. Resistencia organizacional

3.4.5 Tendencias en el futuro

En el pasado, la falta de aplicaciones adecuadas para apoyar el concepto era otra de las grandes limitantes, pero actualmente las soluciones presentes en el mercado tienen la capacidad de integrarse fácilmente con otras aplicaciones. Existen sistemas que pueden manejar información tanto interna como externa a la organización. Las nuevas soluciones son más flexibles a las necesidades de los usuarios y generalmente giran en torno de estructuras basadas en los hechos y no solamente basada en los libros contables.

Debido a que estos sistemas financieros dependen del poder de la computación para manejar los elementos individuales de la información y procesar las transacciones, permiten que los ejecutivos desempeñen un rol más estratégico en la organización. La tecnología de los EIS está evolucionando muy rápido, y es muy probable que en el futuro los sistemas sean diferentes de los que se usan actualmente. Algunos puntos que pueden anticiparse son:

- Una mejor integración con otras aplicaciones.
- Mejor software comercial para el desarrollo de EIS.
- Mejores interfaces sistema-ejecutivo. Puede decirse que los EIS, además de tablas y gráficas, incluirán voz, video e imágenes, es decir, integrarán la tecnología de multimedia en sus desarrollo.
- Durante la toma de decisiones y solución de problemas a veces no se reconoce que también interviene el punto de vista de la persona que ha detectado el problema, por lo tanto hay situaciones en las que hay que decidir o resolver algo, siempre y cuando a la persona le "importen " y por lo cual hace juicios sobre esta

CAPÍTULO IV

ANÁLISIS Y SELECCIÓN DE LA METODOLOGÍA DE DATA WAREHOUSE



En este capítulo se realizará una comparación entre las tres metodologías Ralph Kimball (y su enfoque dimensional), Bill Inmon (y su enfoque Empresarial Warehouse) y Hefesto (enfoque práctico) para la construcción de Data Warehouse, en el caso de particular de esta tesis se analizará desde el punto de vista de la construcción de un Data Warehouse para la Cooperativa de Ahorro y Crédito “San Antonio Ltda.”, para los procesos de cartera de crédito, inversiones (DPF) y Socios/Clientes.

Se debe tomar en cuenta que a la hora de escoger una metodología debemos revisar una que se adapte a al tipo de proyecto y organización dentro de la cual nos encontramos, tomando en cuenta particularidades como tipos de datos que se manejan, la complejidad de los mismos y del modelo relacional, fuentes de datos, número de transacciones, la perspectiva de crecimiento, además si los requerimientos de análisis de información son similares en los diferentes departamentos o no.

4.1 Análisis de las metodologías

4.1.1 Metodología Hefesto

Se presenta en esta parte la metodología HEFESTO, que permite la construcción de Data Warehouse de forma sencilla, ordenada e intuitiva. Su nombre fue inspirado en el dios griego de la construcción y el fuego, y su logotipo es el siguiente:



Ilustración 24 Logotipo Hefesto

Fuente: (Dario, 21)

La Metodología Hefesto, creada por Bernabeu Ricardo Dario (disponible con licencia GNU Free Documentation License). Esta metodología es muy realista y pragmática (no se enfoca en un excesivo número de documentos). Hefesto, se enfoca en el análisis de los requerimientos de la empresa, la identificación de las carencias de información que se tienen, los indicadores y "perspectivas" del negocio, y acto seguido se procede al análisis de las fuentes de datos.

HEFESTO es una metodología propia, cuya propuesta está fundamentada en una muy amplia investigación, comparación de metodologías existentes, experiencias propias en procesos de confección de almacenes de datos.

Cabe destacar que HEFESTO está en continua evolución, y se han tenido en cuenta, como gran valor agregado, todos los feedbacks que han aportado quienes han utilizado esta metodología en diversos países y con diversos fines.

Según (Dario, 21) (Bernabeu, 2009) la idea principal de esta metodología, es comprender cada paso que se realizará, para no caer en el tedio de tener que seguir un método al pie de la letra sin saber exactamente qué se está haciendo, ni por qué.

La construcción e implementación de un DW puede adaptarse muy bien a cualquier ciclo de vida de desarrollo de software, con la salvedad de que para algunas fases en particular, las acciones que se han de realizar serán muy diferentes. Lo que se debe tener muy en cuenta, es no entrar en la utilización de metodologías que requieran fases extensas de reunión de requerimientos y análisis, fases de desarrollo monolítico que conlleve demasiado tiempo y fases de despliegue muy largas. Lo que se busca, es entregar una primera implementación que satisfaga una parte de las necesidades, para demostrar las ventajas del DW y motivar a los usuarios.

La metodología HEFESTO, puede ser embebida en cualquier ciclo de vida que cumpla con la condición antes declarada.

- 1) ANÁLISIS DE REQUERIMIENTOS**
 - ↳ a) Identificar preguntas
 - ↳ b) Identificar indicadores y perspectivas de análisis
 - ↳ c) Modelo Conceptual
- 2) ANÁLISIS DE LOS OLTP**
 - ↳ a) Determinación de Indicadores
 - ↳ b) Establecer correspondencias
 - ↳ c) Nivel de granularidad
 - ↳ d) Modelo Conceptual ampliado
- 3) MODELO LÓGICO DEL DW**
 - ↳ a) Tipo de Modelo Lógico del DW
 - ↳ b) Tablas de dimensiones
 - ↳ c) Tablas de hechos
 - ↳ d) Uniones
- 4) PROCESOS ETL**

Ilustración 25 Etapas de la metodología Hefesto

Fuente: (Dario, 21)

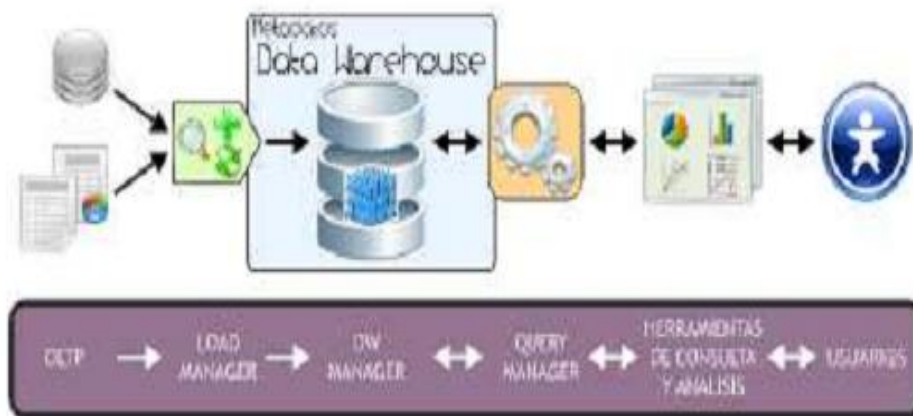


Ilustración 26 Metodología Hefesto

Fuente: (Dario, 21)

Etapas de la metodología

1) Análisis de requerimientos

Lo primero que se hará será identificar los requerimientos de los usuarios a través de preguntas que expliciten los objetivos de su organización. Luego, se analizarán estas preguntas a fin de identificar cuáles serán los indicadores y perspectivas que serán tomadas en cuenta para la construcción del DW. Finalmente se confeccionará un modelo conceptual en donde se podrá visualizar el resultado obtenido en este primer paso.

Es muy importante tener en cuenta que HEFESTO se puede utilizar para construir un Data Warehouse o un Data Mart a la vez, es decir, si se requiere construir por ejemplo dos Data Marts, se deberá aplicar la metodología dos veces, una por cada Data Mart. Del mismo modo, si se analizan dos áreas de interés de negocio, como el área de “Ventas” y “Compras”, se deberá aplicar la metodología dos veces.

a) Identificar preguntas

El primer paso comienza con el acopio de las necesidades de información, el cual puede llevarse a cabo a través de muy variadas y diferentes técnicas, cada una de las cuales poseen características inherentes y específicas, como por ejemplo entrevistas, cuestionarios, observaciones, etc.

El análisis de los requerimientos de los diferentes usuarios, es el punto de partida de esta metodología, ya que ellos son los que deben, en cierto modo, guiar la investigación hacia un desarrollo que refleje claramente lo que se espera del depósito de datos, en relación a sus funciones y cualidades.

El objetivo principal de esta fase, es la de obtener e identificar las necesidades de información clave de alto nivel, que es esencial para llevar a cabo las metas y estrategias de la empresa, y que facilitará una eficaz y eficiente toma de decisiones.

Debe tenerse en cuenta que dicha información, es la que proveerá el soporte para desarrollar los pasos sucesivos, por lo cual, es muy importante que se preste especial atención al relevar los datos.

Una forma de asegurarse de que se ha realizado un buen análisis, es corroborar que el resultado del mismo haga explícitos los objetivos estratégicos planteados por la empresa que se está estudiando.

Otra forma de encaminar el relevamiento, es enfocar las necesidades de información en los procesos principales que desarrolle la empresa en cuestión.

La idea central es, que se formulen preguntas complejas sobre el negocio, que incluyan variables de análisis que se consideren relevantes, ya que son estas las que permitirán estudiar la información desde diferentes perspectivas.

Un punto importante que debe tenerse muy en cuenta, es que la información debe estar soportada de alguna manera por algún OLTP, ya que de otra forma, no se podrá elaborar el DW.

b) Identificar indicadores y perspectivas

Una vez que se han establecido las preguntas de negocio, se debe proceder a su descomposición para descubrir los indicadores que se utilizarán y las perspectivas de análisis que intervendrán.

Para ello, se debe tener en cuenta que los indicadores, para que sean realmente efectivos son, en general, valores numéricos y representan lo que se desea analizar concretamente, por ejemplo: saldos, promedios, cantidades, sumatorias, fórmulas, etc.

En cambio, las perspectivas se refieren a los objetos mediante los cuales se quiere examinar los indicadores, con el fin de responder a las preguntas planteadas, por ejemplo: clientes, proveedores, sucursales, países, productos, rubros, etc. Cabe destacar, que el Tiempo es muy comúnmente una perspectiva.

c) Modelo Conceptual

En esta etapa, se construirá un modelo conceptual a partir de los indicadores y perspectivas obtenidas en el paso anterior. Modelo Conceptual: descripción de alto nivel de la estructura de la base de datos, en la cual la información es representada a través de objetos, relaciones y atributos.

A través de este modelo, se podrá observar con claridad cuáles son los alcances del proyecto, para luego poder trabajar sobre ellos, además al poseer un alto nivel de definición de los datos, permite que pueda ser presentado ante los usuarios y explicado con facilidad.

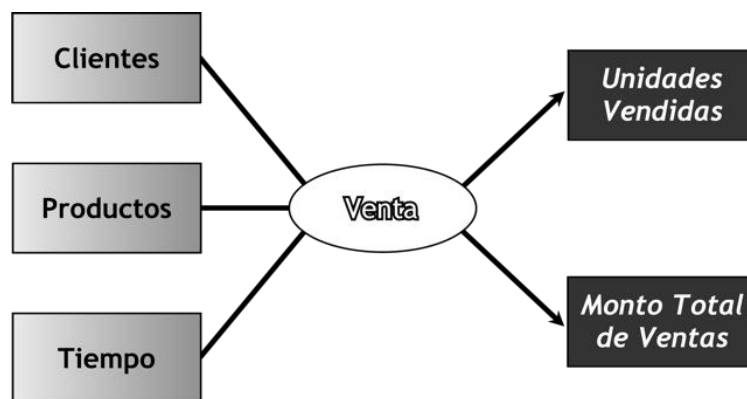


Ilustración 27 Modelo conceptual-Metodología Hefesto

Fuente: (Dario, 21)

2) Análisis de los OLTP

Seguidamente, se analizarán las fuentes OLTP para determinar cómo serán calculados los indicadores y para establecer las respectivas correspondencias entre el modelo conceptual creado en el paso anterior y las fuentes de datos. Luego, se definirán qué campos se incluirán en cada perspectiva. Finalmente, se ampliará el modelo conceptual con la información obtenida en este paso.

a) Conformar Indicadores

En este paso se deberán explicitar como se calcularán los indicadores, definiendo los siguientes conceptos para cada uno de ellos:

- Hecho/s que lo componen, con su respectiva fórmula de cálculo. Por ejemplo: Hecho1 + Hecho2.
- Función de sumariazación que se utilizará para su agregación. Por ejemplo: SUM, AVG, COUNT, etc.

b) Establecer correspondencias

El objetivo de este paso, es el de examinar los OLTP disponibles que contengan la información requerida, como así también sus características, para poder identificar las correspondencias entre el modelo conceptual y las fuentes de datos.

La idea es, que todos los elementos del modelo conceptual estén correspondidos en los OLTP.

c) Nivel de granularidad

Una vez que se han establecido las relaciones con los OLTP, se deben seleccionar los campos que contendrá cada perspectiva, ya que será a través de estos por los que se examinarán y filtrarán los indicadores.

Para ello, basándose en las correspondencias establecidas en el paso anterior, se debe presentar a los usuarios los datos de análisis disponibles para cada perspectiva. Es muy importante conocer en detalle que significa cada campo y/o valor de los datos encontrados en los OLTP, por lo cual, es conveniente investigar su sentido, ya sea a través de diccionarios de datos, reuniones con los encargados del sistema, análisis de los datos propiamente dichos, etc.

Luego de exponer frente a los usuarios los datos existentes, explicando su significado, valores posibles y características, estos deben decidir cuáles son los que consideran relevantes para consultar los indicadores y cuáles no.

Con respecto a la perspectiva “Tiempo”, es muy importante definir el ámbito mediante el cual se agruparán o sumarán los datos. Sus campos posibles pueden ser: día de la semana, quincena, mes, trimestres, semestre, año, etc.

Al momento de seleccionar los campos que integrarán cada perspectiva, debe prestarse mucha atención, ya que esta acción determinará la granularidad de la información encontrada en el DW

d) Modelo Conceptual ampliado

En este paso, y con el fin de graficar los resultados obtenidos en los pasos anteriores, se ampliará el modelo conceptual, colocando bajo cada perspectiva los campos seleccionados y bajo cada indicador su respectiva fórmula de cálculo. Gráficamente:

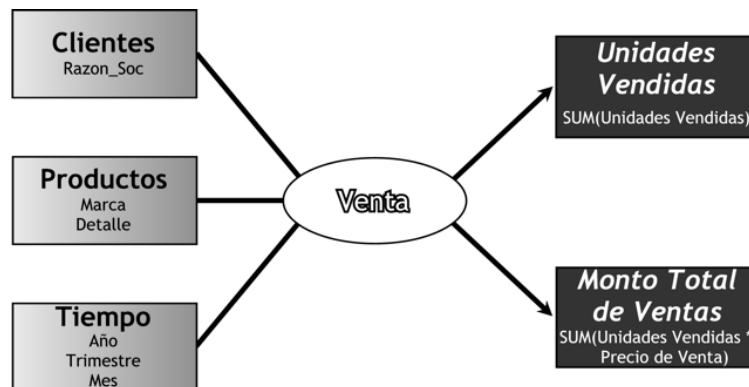


Ilustración 28 Modelo conceptual ampliado – Hefesto

Fuente: (Dario, 21)

3) Modelo lógico del Data Warehouse

A continuación, se confeccionará el modelo lógico de la estructura del DW, teniendo como base el modelo conceptual que ya ha sido creado. Para ello, primero se definirá el tipo de modelo que se utilizará y luego se llevarán a cabo las acciones propias al caso, para diseñar las tablas de dimensiones y de hechos. Finalmente, se realizarán las uniones pertinentes entre estas tablas. Modelo Lógico: representación de una estructura de datos, que puede procesarse y almacenarse en algún SGBD⁸.

a) Tipo de Modelo Lógico del DW

Se debe seleccionar cuál será el tipo de esquema que se utilizará para contener la estructura del depósito de datos, que se adapte mejor a los requerimientos y necesidades de los usuarios. Es muy importante definir objetivamente si se empleará un esquema en estrella, constelación o copo de nieve, ya que esta decisión afectará considerablemente la elaboración del modelo lógico.

b) Tablas de dimensiones

⁸ SGBD.- es una agrupación de programas que sirven para definir, construir y manipular una base de datos.

En este paso se deben diseñar las tablas de dimensiones que formaran parte del DW.

Para los tres tipos de esquemas, cada perspectiva definida en el modelo conceptual constituirá una tabla de dimensión. Para ello deberá tomarse cada perspectiva con sus campos relacionados y realizarse el siguiente proceso:

- Se elegirá un nombre que identifique la tabla de dimensión.
- Se añadirá un campo que represente su clave principal.
- Se redefinirán los nombres de los campos si es que no son lo suficientemente intuitivos.

Gráficamente:

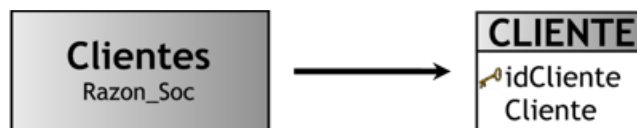


Ilustración 29 Tabla de dimensiones-Hefesto

Fuente: (Dario, 21)

c) Tablas de hechos

En este paso, se definirán las tablas de hechos, que son las que contendrán los hechos a través de los cuales se construirán los indicadores de estudio.

d) Uniones

Para los tres tipos de esquemas, se realizarán las uniones correspondientes entre sus tablas de dimensiones y sus tablas de hechos.

4) Integración de datos

Una vez construido el modelo lógico, se deberá proceder a poblarlo con datos, utilizando técnicas de limpieza y calidad de datos, procesos ETL, etc.; luego se definirán las reglas y políticas para su respectiva actualización, así como también los procesos que la llevarán a cabo.

a) Carga Inicial

Debemos en este paso realizar la Carga Inicial al DW, poblando el modelo de datos que hemos construido anteriormente. Para lo cual debemos llevar adelante una serie de tareas básicas, tales como limpieza de datos, calidad de datos, procesos ETL, etc.

La realización de estas tareas puede contener una lógica realmente compleja en algunos casos. Afortunadamente, en la actualidad existen muchos software que se pueden emplear a tal fin, y que nos facilitarán el trabajo.

Se debe evitar que el DW sea cargado con valores faltantes o anómalos, así como también se deben establecer condiciones y restricciones para asegurar que solo se utilicen los datos de interés.

Cuando se trabaja con un esquema constelación, hay que tener presente que varias tablas de dimensiones serán compartidas con diferentes tablas de hechos, ya que puede darse el caso de que algunas restricciones aplicadas sobre una tabla de dimensión en particular para analizar una tabla de hechos, se puedan contraponer con otras restricciones o condiciones de análisis de otras tablas de hechos.

Primero se cargarán los datos de las dimensiones y luego los de las tablas de hechos, teniendo en cuenta siempre, la correcta correspondencia entre cada elemento. En el caso en que se esté utilizando un esquema copo de nieve, cada vez que existan jerarquías de dimensiones, se comenzarán cargando las tablas de dimensiones del nivel más general al más detallado.

Concretamente, en este paso se deberá registrar en detalle las acciones llevadas a cabo con diferente software. Por ejemplo, es muy común que sistemas ETL trabajen con "pasos" y "relaciones", en donde cada "paso" realiza una tarea en particular del proceso ETL y cada "relación" indica hacia donde

debe dirigirse el flujo de datos. En este caso lo que se debe hacer es explicar que hace el proceso en general y luego que hace cada "paso" y/o "relación". Es decir, se partirá de lo más general y se irá a lo más específico, para obtener de esta manera una visión general y detallada de todo el proceso.

Es importante tener presente, que al cargar los datos en las tablas de hechos pueden utilizarse pre-agregaciones, ya sea al nivel de granularidad de la misma o a otros niveles diferentes.

b) Actualización

Cuando se haya cargado en su totalidad el DW, se deben establecer sus políticas y estrategias de actualización o refresco de datos.

Una vez realizado esto, se tendrán que llevar a cabo las siguientes acciones:

- Especificar las tareas de limpieza de datos, calidad de datos, procesos ETL, etc., que deberán realizarse para actualizar los datos del DW.
- Especificar de forma general y detallada las acciones que deberá realizar cada software.

4.1.2 Metodología Bill Inmon.



Esta metodología la definió su autor en el año de 1992 en el libro "Building the Data Warehouse", en el que se proponía mecanismos necesarios para llevar a cabo la correcta realización de un Data Warehouse.

Innmon ve la necesidad de transferir la información de los diferentes OLTP (Sistemas Transaccionales) de las organizaciones a un lugar centralizado donde los datos puedan ser utilizados para el análisis a la Fábrica de Información Corporativa (CIF o Corporate Information Factory). Insiste además en que ha de tener las siguientes características:

- **Orientado a temas.**- Los datos en la base de datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí.
- **Integrado.**- La base de datos contiene los datos de todos los sistemas operacionales de la organización, y dichos datos deben ser consistentes.
- **No volátil.**- La información no se modifica ni se elimina, una vez almacenado un dato, éste se convierte en información de sólo lectura, y se mantiene para futuras consultas.
- **Variante en el tiempo.**- Los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones.

La información debe estar a los máximos niveles de detalle. Los Data Warehouse departamentales o data marts son tratados como subconjuntos de este Data Warehouse corporativo, son construidos para cubrir las necesidades individuales de análisis de cada departamento, y siempre a partir del Data Warehouse central

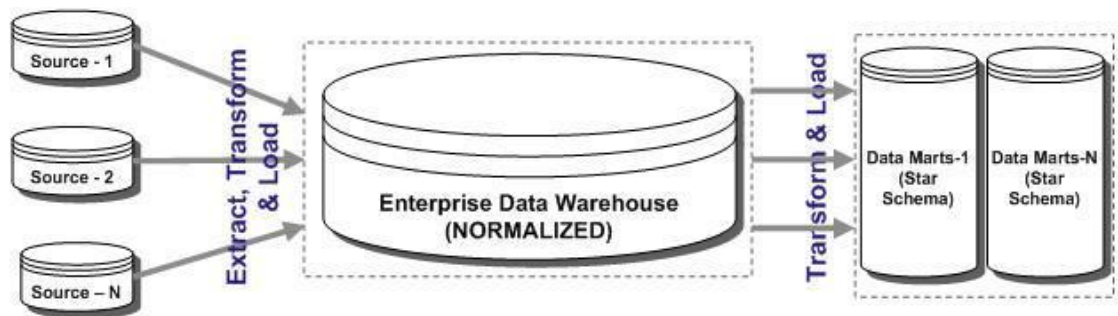


Ilustración 30 Metodología Inmon- DW Corporativo

Fuente: (Dario, 21)

La metodología Inmon también se referencia normalmente como Top-down. Los datos son extraídos de los sistemas operacionales por los procesos ETL y cargados en las “area stage”, donde son validados y consolidados en el DW corporativo, donde además existen los llamados metadatos que documentan de una forma clara y precisa el contenido del DW. Una vez realizado este proceso, los procesos de refresco de los Data mart departamentales obtienen la información de este, y con las consiguientes transformaciones, organizan los datos en las estructuras particulares requeridas por cada uno de ellos, refrescando su contenido.

La metodología para la construcción de un sistema de este tipo es la habitual para construir un sistema de información, utilizando las herramientas habituales (esquema Entidad Relación, DIS (Data Item Sets, etc.). Para el tratamiento de los cambios en los datos, usa la Gestión de las dimensiones continuas y discretas (inserta fechas en los datos para determinar su validez para la dimensión continua o bien mediante el concepto de snapshot o foto para la dimensión discreta).

Al tener este enfoque global, es más difícil de desarrollar en un proyecto sencillo (pues se intentará abordar el “todo”, a partir del cual luego se irá al “detalle”).

4.1.3 Metodología Ralph Kimball



El Data Warehouse es un conglomerado de todos los Data marts dentro de una empresa, siendo una copia de los datos transaccionales estructurados de una forma especial para el análisis, de acuerdo al Modelo Dimensional (no normalizado), que incluye, como se explicó, las dimensiones de análisis y sus atributos, su organización jerárquica, así como los diferentes hechos de negocio que se quieren analizar. Por un lado se tiene las tablas para las representar las dimensiones y por otro lado tablas para los hechos. Los diferentes Data marts están conectados entre sí por la llamada estructura de bus, que contiene los elementos anteriormente citados a través de las dimensiones conformadas (que permiten que los usuarios puedan realizar consultas conjuntas sobre los diferentes data marts, pues este bus contiene los elementos en común que los comunican). Una dimensión conformada puede ser, por ejemplo, la dimensión cliente, que incluye todos los atributos o elementos de análisis referentes a los clientes y que puede ser compartida por diferentes data marts (crédito, clientes, inversiones, etc).

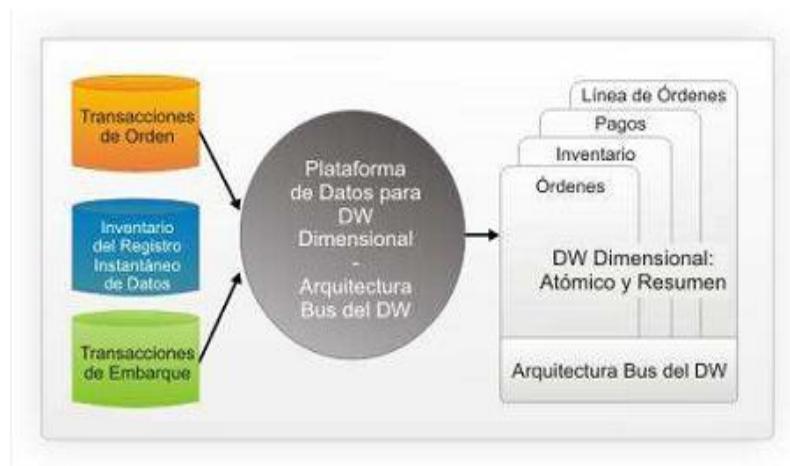


Ilustración 31 Metodología Kimball- Arquitectura Bus del DW

Fuente: (Open Business Intelligence-La red del business intelligence)

Esta metodología también se referencia como Bottom-up, pues al final el Data Warehouse Corporativo no es más que la unión de los diferentes data marts, que están estructurados de una forma común a través de la estructura de bus. Esta característica le hace más flexible y sencilla de implementar, pues se puede construir un Data mart como primer elemento del sistema de análisis, y luego ir añadiendo otros que comparten las dimensiones ya definidas o incluyen otras nuevas. En este sistema, los procesos ETL extraen la información de los sistemas operacionales y los procesan igualmente en las “area stage”, realizando posteriormente el llenado de cada uno de los Data mart de una forma individual, aunque siempre respetando la estandarización de las dimensiones (dimensiones conformadas).

La metodología para la construcción del Data Warehouse incluye las 4 fases que son:

- Selección del proceso de negocio.
- Definición de la granularidad de la información.
- Elección de las dimensiones de análisis.

- Identificación de los hechos o métricas. Tratamiento de los cambios, Dimensiones Lentamente Cambiantes (SCD).

Etapas de la Metodología de Kimball

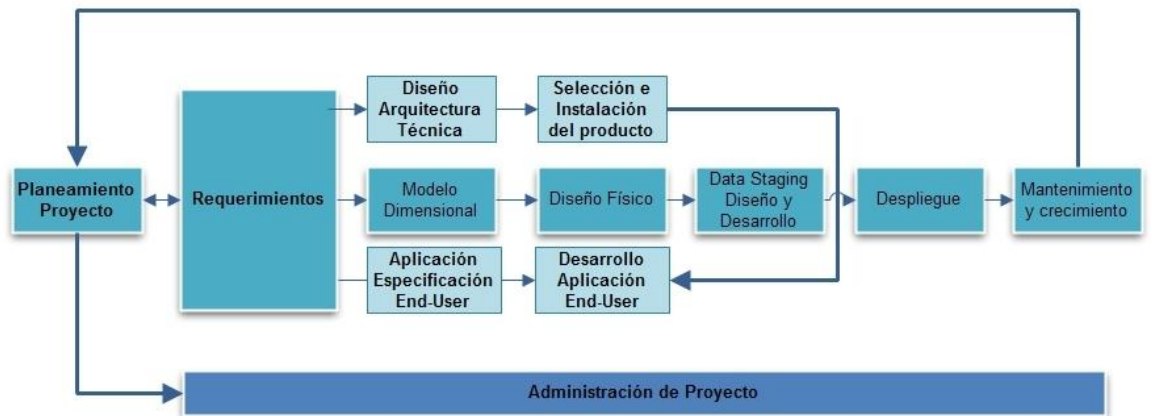


Ilustración 32 Ciclo de vida Ralph Kimball

Fuente: Propia

1. Planeamiento del proyecto

En esta etapa inicial se busca identificar el escenario del proyecto para determinar el alcance, definir el proyecto, incluyendo justificaciones del negocio, generando la información suficiente para poder dar seguimiento al progreso del proyecto.

2. Análisis de Requerimientos

Es un factor determinante para el éxito del proyecto, ya que se necesita de la interpretación correcta de los diferentes niveles de requerimientos expresados por los diferentes usuarios y establece la base de las tres etapas paralelas siguientes.

3. Selección e instalación de productos

En base al diseño de la arquitectura se evalúa y selecciona los componentes específicos de la arquitectura, como la plataforma, el motor de base de datos, herramientas ETL y herramientas de acceso.

4. Modelamiento dimensional

El diseño del modelo dimensional busca presentar los datos de una forma intuitiva y que proporcione acceso de alto desempeño.

5. Diseño Físico

Esta etapa se focaliza sobre la selección de las estructuras que soporta el diseño lógico, que incluye los nombres de columnas, tipos de datos, declaraciones de claves.

6. Diseño y desarrollo de la presentación de los datos

En esta etapa se ejecutan los procesos de extracción, transformación y carga (ETL), este proceso comprende varios aspectos que son determinantes en el proyecto de inteligencia de negocios, por lo que para su desarrollo se debe seguir un plan para su correcto desarrollo.

7. Especificación de aplicaciones para usuarios finales.

En esta etapa se centra más en el front room, ya que se proporcionará la interfaz que se mostrará al usuario.

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

8. Mantenimiento y crecimiento

Cuando se desarrolla un proyecto Data Warehouse se debe pensar en el mantenimiento posterior, pues estas aplicaciones tienden a crecer a medida que crecen los datos de la organización.

4.1.4 Criterios de selección.

La tabla siguiente muestra cualitativamente las características de cada una de las metodologías descritas anteriormente tomando como valoraciones las siguientes:

- **M** (Muy Bueno).
- **B** (Bueno).
- **R** (Regular).
- **N** (No disponible).

4.1.5 Comparativa de las metodologías.

	Parámetro de evaluación	Metodología Ralph Kimball	Metodología Bill Inmon	Metodología Hefesto
1	Facilidad de seguimiento	M	B	M
2	Semejanza con fases de desarrollo de sistemas	M	R	B
3	Detalle de pasos o etapas a seguir	M	B	M
4	Rapidez de implementación	M	B	M
5	Cantidad de fases	B	N	B
6	Antigüedad	M	M	R
7	Difusión	B	B	B
8	Presentación de casos de estudio	M	N	M

Tabla 2 Comparativa de las metodologías estudiadas

Fuente: Propia

Metodología	Kimball	Innmon
Definición Data Mart	Un data mart mantiene los datos al menor nivel de detalle, los cuales se refieren a un proceso de negocio	Un data mart mantiene los datos agregados que se relacionan a la unidad de negocio
	Un data mart se construye mediante la extracción de datos directamente desde los sistemas operacionales	Un data mart se construye mediante la extracción de datos del data warehouse de la empresa (también llamados data marts dependientes).
	Los data marts están vinculados entre sí.	Los data marts no están vinculados entre sí.
Enfoque de desarrollo	Se basan en procesos específicos del negocio y están vinculadas a las dimensiones, que forman la arquitectura de bus data warehouse.	El diseño de un Data Warehouse para toda la empresa se basa en su modelo de datos. Es una aplicación progresiva de Las áreas temáticas, de acuerdo con las prioridades establecidas

Tabla 3 Tabla comparativa de Metodología Innmon vs Kimball

Fuente: Propia

Para la selección de la metodología se ha tomado en consideración las siguientes puntuaciones para los criterios de evaluación presentados en la tabla 2.

Parámetros	Justificación
1, 2, 3, 6 y 7	Se dan las puntuaciones de M=3, B=2, R=1 y N=0 debido a que estos parámetros se relacionan con el ciclo propuesto por la metodología y su facilidad de seguimiento, que si bien es un factor muy importante en la selección, no es determinante
4 y 5	Se dan puntuaciones de M=4, B=3, R=2 y N=1 que son más altas que en el caso anterior debido a que reflejan la rapidez que ofrece la metodología para la obtener el producto final que en este caso es un factor determinante en la selección
8	Se dan puntuaciones de N=0 que en este caso significa que carece de esta característica y M=1 en el caso contrario.

Tabla 4 Puntuación a los criterios de evaluación

Fuente: Propia

4.1.6 Selección de la metodología

Por lo estudiado de la Metodología de Innmon vemos que es más apropiada para sistemas complejos, donde se quiere asegurar la perdurabilidad y consistencia de la información aunque cambien los procesos de negocio de la organización.

Para proyectos pequeños donde se quiere asegurar la usabilidad de los usuarios que permita un desarrollo rápido e incremental de la solución donde no se tiene claro el panorama global, el enfoque de Kimball es el más apropiado.

La metodología Hefesto por su parte aunque se enmarca bien en los procesos de desarrollo de software sigue siendo una metodología muy joven y que pese a todas las contribuciones o feedbacks que se hace a la metodología, esta se encuentra embebida dentro de los 2 metodologías analizadas anteriormente y más particularmente en la metodología de Kimball.

Con las puntuaciones descritas anteriormente se obtienen los siguientes resultados:

Metodología	Calificación
Ralph Kimball	23
Bill Innmon	15
Hefesto	20

Tabla 5 Calificaciones a las metodologías

Fuente: Propia

Por todo lo expuesto anteriormente en la comparativa de la metodología más las conclusiones realizadas en esta sección se concluye que para el desarrollo de este tipo de proyectos donde se necesita resultados breves es conveniente utilizar el enfoque de Kimball para el diseño del DW.

El DW se va a desarrollar intentando que las dimensiones estén conformadas (dentro del concepto de Data Warehouse bus), con lo que se dejará la posibilidad a una ampliación posterior dentro el ámbito de la organización, añadiendo nuevos cubos que utilizarán las dimensiones conformadas ya definidas y permitirán el análisis posterior de otros hechos que también son importantes analizarlos.

CAPÍTULO V

ESTUDIO DE LA HERRAMIENTA PENTAHO



5.1 Introducción

Pentaho Corp., empresa dueña de Pentaho, fue fundada en el año 2004 por pioneros en Business Intelligence Open Source comerciales provenientes de empresas como: Business Objects, Cognos, Hyperion, Jboss, Oracle, Red Hat y SAS

Pentaho, provee una alternativa de soluciones de BI en distintas áreas como en la Arquitectura, Soporte, Funcionalidad e Implantación. Estas soluciones al igual que su ambiente de implantación están basados en JAVA, haciéndolo flexible en cubrir amplias necesidades empresariales. A través de la integración funcional de diversos proyectos de Open Source permite ofrecer soluciones en áreas como: Análisis de información, Reportes, Tableros de mando conocido como “DashBoards”, Flujos de Trabajo y Minería de Datos.

Pentaho ofrece dos versiones de su solución en base al modelo anteriormente descrito: una versión comunitaria gratuita orientada principalmente al mundo académico, y una versión Enterprise comercial orientada a la implementación profesional tanto en empresas privadas como en instituciones gubernamentales u otras sin fines de lucro que pretendan potenciar sus capacidades analíticas para mejorar su gestión. El siguiente esquema resume el modelo:

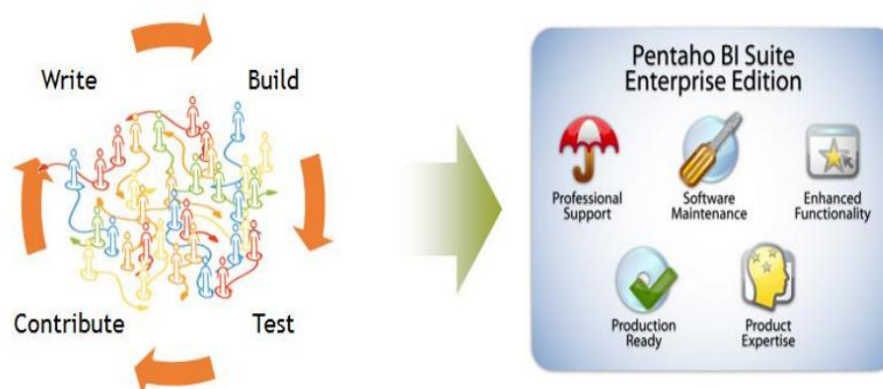


Ilustración 33 Modelo de las versiones de Pentaho

Fuente: (Bouman & Van Dongen, 2009)

Open Source Comunnity	Enterprise Edition
Pentaho provee Lead developers y lineamientos técnicos para el desarrollo del producto.	Pentaho certifica estos productos como listos para una implementación empresarial.
La comunidad contribuye con nuevas soluciones y se utiliza como un “test bed” para estas innovaciones.	Agrega funcionalidades mejoradas y una mayor facilidad de uso e implementación.
Mejora eficazmente la calidad del producto mediante una rápida identificación de defectos u oportunidades de mejora.	Provee soporte técnico profesional para la herramienta con altos niveles de servicio a un bajo costo de suscripción.

Tabla 6 Funcionalidades de versiones

Fuente: Propia

5.2 Módulos del sistema

En esta parte de la tesis se describen los módulos con los que cuenta Pentaho y se realiza una descripción breve de su utilización dentro de la Suite.

Cabe destacar que todos los módulos aquí descritos sirven para crear una solución de Inteligencia de negocios completa e integral sin requerir de otros complementos.



Pentaho Data Integration: Limpia e integra datos desde las diferentes fuentes u orígenes de datos



Pentaho Reporting: Accede a los datos y provee información de alto valor a la organización.



Pentaho Analysis: Permite explorar y analizar los datos de manera interactiva, rápida y ágil.



Pentaho Dashboards: Entrega resúmenes de información y métricas de alto valor a través de una interfaz interactiva y fácil de comprender



Pentaho Data Mining: Descubre patrones escondidos en los datos, y permite predecir eventos futuros.

5.3 Funcionalidades Principales

La suite de herramientas de Pentaho permite elegir entre el uso completo de las herramientas o la utilización específica de ciertas herramientas para resolver los problemas de negocio.

En un proyecto de Inteligencia de negocios no es necesaria la utilización de todos los módulos con que cuenta Pentaho sino únicamente de los que conlleven a la solución de los requerimientos que se plantee.

Una arquitectura simplificada de la solución puede ser representada mediante el siguiente esquema:



Ilustración 34 Arquitectura simplificada de Pentaho

Fuente: (Bouman & Van Dongen, 2009)

Como se puede ver en la ilustración anterior las herramientas de Pentaho nos presenta todas las funcionalidades necesarias para que desde el comienzo del proceso de desarrollo, iniciando en las fuentes de datos como puede ser una base de datos relacional o fuentes no estructuras como hojas de Excel, texto plano, pasando por herramientas de extracción , transformación y carga de datos (ETL), el manejo de grandes volúmenes de datos con hadoop, herramientas para optimizar el análisis de los datos con Pentaho Analysis y herramientas para realizar reportes , cuadros de mando y una serie de herramientas para el análisis de los cubos de información nos permiten presentar al usuario de negocios un método fácil por el cual se pueda tomar decisiones que beneficien a las empresas.

CAPÍTULO VI

IMPLEMENTACIÓN, PRUEBAS Y VERIFICACIÓN DEL FUNCIONAMIENTO MEDIANTE LA METODOLOGÍA SELECCIONADA



6. Metodología para desarrollo del proyecto

Para el desarrollo de la construcción de los Data mart, cubos OLAP y reportes, se sigue la metodología de Ralph Kimball, dado que establece claros procesos para todo el ciclo del desarrollo del proyecto y garantiza la calidad y eficiencia de la solución de inteligencia de negocios.

Esta metodología fue desarrollada desde el inicio del proceso de construcción, hasta llegar a las etapas de interacción con el usuario y documentación del proyecto.

En las siguientes secciones se describen los procesos realizados para cada fase del proyecto que garantizan su calidad y cumplimiento.

6.1 Planeación y Administración del Proyecto

6.1.1 Objetivo del Proyecto

Construir los Data mart, cubos OLAP y reportes de Business Intelligence, necesarios para el análisis de la información de los módulos de cartera de crédito, clientes e Inversiones.

6.1.2 Definición del Proyecto

Para el proyecto desarrollado se ha identificado un alto interés por parte de los directivos de la Cooperativa de Ahorro y Crédito "San Antonio Ltda.", para el éxito de su implementación.

La demanda del proyecto se da debido a la necesidad de obtener mejor información del sistema financiero, para tomar mejores decisiones a nivel gerencial y así mejorar su competitividad y rendimiento. Esta demanda se

satisface al implementar para cada sistema una solución de inteligencia de negocios.

Dicha Solución BI está compuesta por varios Data marts, desarrollados por medio de tablas , en donde se encuentran ya no los simples datos operacionales de la base de datos sino la información útil lista para ser reportada o como en nuestro caso información lista para el respectivo análisis multidimensional por medio de cubos OLAP.

Posteriormente el debido análisis multidimensional para por último definir el reporte de la información desarrollando con herramientas Open Source como Mondrian 3.0.4 su navegador OLAP Jpivot y Java 6.0 como plataforma, además del servidor Web también de código libre denominado Apache Tomcat 6.0.16.

6.1.3 Alcance del Proyecto

El presente proyecto de tesis tiene todos los componentes de una solución BI con análisis multidimensional en herramientas Open Source.

Como fuente de datos principal se tiene la base de datos de producción del sistema financiero CONEXUS llamada Conexus de la cual se obtendrá solo la información importante y necesaria para el análisis de la misma.

A partir de esto se define mediante la implementación de procesos ETL cada uno de los Data mart que se utilizan para el análisis multidimensional en cubos OLAP y su posterior explotación. El proceso ETL está realizado mediante la herramienta Pentaho Data Integration (Spoon) el cual carga la información útil en una base de datos Mysql la cual contiene las dimensiones y tablas de hechos.

Posteriormente haciendo uso de la herramienta Mondrian Schema Workbench se realizará la creación de los cubos multidimensionales para su posterior explotación a través del servidor OLAP y Jpivot.

Todas estas herramientas soportan varios sistemas operativos, incluyendo Microsoft Windows y GNU/Linux lo que permite una gran portabilidad y la

posibilidad de realizar el análisis OLAP vía web, gracias a que el análisis OLAP se realiza utilizando un navegador de Internet.

6.1.4 Justificación del Proyecto en el Negocio

Además de ser un proyecto académico, el proyecto busca mejorar la productividad y rendimiento en cuanto al reporte de información se refiere en cada uno de los módulos del sistema citados anteriormente en los objetivos del proyecto en el sistema Conexus.

Como justificación para el negocio se prevén los siguientes beneficios:

1. Solución Business Intelligence económicas

Ya que con la realización del denominado análisis multidimensional con herramientas Open Source, estamos dando una solución BI libre en cuanto a la capa BI se refiere.

2. Mejora de proceso de toma de decisiones.

Es fundamental considerar este beneficio, puesto que los usuarios del sistema podrán obtener un mayor nivel en cuánto a tomar decisiones importantes, al tener reportes BI con información actualizada, clara y precisa cuyos reportes además ofrece un alto nivel estadístico y de interacción con el usuario final.

6.2 Análisis de Requerimientos

Para el levantamiento de requerimientos se realizaron reuniones con los Jefes de área de Crédito, Inversiones y Atención al cliente. Además y como se mencionó en la justificación del proyecto de tesis la creación del repositorio se basa principalmente en realizar los reportes que se presentan trimestralmente a la asamblea de socios y que están relacionados con cartera de crédito y socios e inversiones.

6.2.1 Levantamiento de Requerimientos

Los requerimientos obtenidos en las reuniones fueron.

1. Módulo de Clientes

- a) Ver socios ingresados por períodos
- b) Ver los socios por sexo
- c) Ver socios retirados por períodos.
- d) Ver el total de socios activos e inactivos en determinada fecha
- e) Ver el ingreso promedio de los socios

2. Módulo de inversiones

- a) Ver el número de inversiones por oficinas
- b) Ver inversiones por tipo de pago de interés
- c) Ver inversiones por estados
- d) Ver tasa de interés promedio por oficina
- e) Ver el monto promedio de depósito por oficina
- f) Ver inversiones por captadas por períodos.
- g) Ver inversiones por clientes

3. Módulo de Cartera de crédito (saldos de cartera).

- a) Ver cartera de crédito por estado (VIGENTE, RECLASIFICADA, VENCIDA)
- b) Consolidado de cartera de crédito
- c) Ver cartera por líneas de crédito
- d) Ver créditos por origen de recursos
- e) Ver créditos por períodos.
- f) Ver créditos por oficiales de crédito
- g) Ver créditos por socios
- h) Ver monto promedio de créditos por oficina

Estos requerimientos fueron los acordados para implementar en reportes estadísticos con la empresa, todos ellos soportados con datos encontrados en la base de datos respectivamente que funcionan como fuente.

6.2.1.1 Documentación de Requerimientos

1. Módulo de Clientes

a) Ver socios por oficinas

Con este reporte se podrá revisar los socios que tiene cada oficina

b) Ver los socios por género

Con este reporte se mostrará el número de socios por género.

c) Ver socios retirados por períodos.

Con este reporte se podrá visualizar el número de socios que se retiran de la Cooperativa durante un período

d) Ver el total de socios activos e inactivos en determinada fecha

Con este reporte se podrá ver la cantidad de socios activos e inactivos de la Cooperativa

e) Ver el ingreso económico promedio de los socios

Este reporte permitirá analizar la situación económica de cada socio y será utilizado por la unidad de cumplimiento y riesgos.

f) Ver socios por estado civil.

Se podrá ver el número de socios por estado civil

g) Ver socios por oficina

Se visualizará el número de socios por oficina.

2. Módulo de inversiones

- a) Ver el número de inversiones por oficinas.

Con este reporte se podrá observar el número de inversiones que tenga cada oficina.

- b) Ver inversiones por tipo de pago de interés

Con este reporte se podrá ver el detalle de inversiones que existe por tipo de pago de interés.

- c) Ver inversiones por estados

Con este reporte se podrá visualizar el detalle de inversiones por estados. (VIGENTE, VENCIDA)

- d) Ver tasa de interés promedio por oficina

Con este reporte se podrá ver la tasa de interés promedio de inversiones en detalle o por oficina.

- e) Ver el monto promedio de depósito por oficina

Con este reporte se podrá observar el monto promedio de las inversiones por oficina.

- f) Ver plazo promedio de inversiones en días

Con este reporte se podrá observar el plazo promedio en día de las inversiones que mantiene la Cooperativa.

- g) Ver monto en inversiones

Con este reporte se podrá visualizar el monto en inversiones que mantiene la cooperativa.

- h) Ver inversiones por captadas por períodos.

Con este reporte se podrá visualizar por períodos las inversiones que mantiene la Cooperativa.

- i) Ver inversiones por clientes

Con este reporte se podrá ver el número de inversiones que mantiene cada cliente por oficina.

3. Módulo de Cartera de crédito (saldos de cartera).

- a) Ver cartera de crédito por estado (VIGENTE, RECLASIFICADA, VENCIDA).

Con este reporte se podrá observar en reporte de cartera de crédito por estado.

- b) Consolidado de cartera de crédito

Con este reporte se podrá ver el consolidado de cartera de las tres oficinas.

- c) Ver cartera por líneas de crédito

Con este reporte se podrá la cartera de créditos por línea de crédito (Microcrédito, Consumo, Vivienda).

- d) Ver créditos por origen de recursos.

Con este reporte se podrá ver el reporte de cartera por origen de recursos o fuente de fondeo).

- e) Ver créditos por períodos.

Con este reporte podemos observar por períodos la cartera de crédito

- f) Ver créditos por oficiales de crédito.

Con este reporte se podrá ver la cartera de crédito que mantiene cada oficial.

- g) Ver créditos por socios

Con este reporte se podrá ver una vista general de la cartera que mantiene cada socio en la Cooperativa.

- h) Ver monto promedio de créditos por oficina

Con este reporte podemos observar el monto promedio de la cartera por oficina.

6.3 Modelamiento Dimensional

Para iniciar el Modelamiento dimensional se debe tener en cuenta el principal objetivo de cualquier data mart es el análisis de la información. Este análisis es realizado por medio de reportes, por lo tanto al modelar el data mart se debe tener como objetivo la información deseada en los reportes.

Además se tomará en cuenta que los data mart realizados para cada uno de los módulos han sido realizados por medio de la creación de procesos ETL y creación de procedimientos almacenados. En las siguientes secciones se detalla cada componente del modelo dimensional.

6.3.1 El Data mart

El modelo diseñado e implementado es de varios data mart en modelo constelación lo cual hace que las dimensiones de un data mart puedan ser utilizadas en otro data mart para el análisis de otro tipo de requerimiento y/o módulo a ser analizado.

Los data mart implementados en este proyecto tienen cada uno de ellos como fuente la base de datos conexus, por esta razón y por la facilidad se utilizará procesos ETL para la extracción, transformación y carga.

Es importante mencionar que en cada una de las dimensiones y tablas de hechos de nuestros data mart se agregó información adicional la misma que sirve para permitir al usuario final el poder necesario para personalizar sus reportes, de manera que pueda habilitar y/o ocultar en tiempo de ejecución las dimensiones y medidas adicionales en cada uno de los reportes estadísticos Business Intelligence que se ha desarrollado.

6.3.2 Definición de la granularidad

Se definió la granularidad de las tablas de hechos como las más bajas o granulares posibles. Por ejemplo se puede consultar a detalle el número de socios retirados en un período específico es así que la granularidad para este ejemplo es el día que se deriva directamente desde año->trimestre->mes->semana->día

De esta forma será posible llegar al grado de detalle que se desee y consultar registros de manera específica, aunque este no sea el objetivo de un data mart. Las medidas, que son los campos de valor de las tablas de hechos son los valores con la granularidad establecida.

6.3.3 Dimensiones

Se definen las dimensiones que soportan los requerimientos definidos, cumpliendo con la granularidad de la tabla de hechos. Las siguientes secciones relacionan las tablas diseñadas para la base de datos con su dimensión correspondiente.

1. Módulo de Clientes

a. Dimensión Clientes

Tabla: clientes

Contiene la información necesaria acerca de los Clientes en la Cooperativa, tales como código de cliente, nombres completos, tipo de personas, edad, sexo, estado civil, instrucción, profesión y más información adicional como el período, las fechas de entrada , salida, estado, tipo de cliente.

Campo	Tipo
<u>clien_cod_clien</u>	int(11)
clien_nom_clien	varchar(61)
tpers_des_tpers	varchar(40)
clien_num_anios	decimal(32,0)
sexos_des_sexos	varchar(40)
ecivi_des_ecivi	varchar(40)
instr_des_instr	varchar(40)
profe_des_profe	varchar(40)
clien_fec_entra	datetime
clien_fec_salid	datetime
clien_ctr_estad	int(11)
clien_cod_usuar	int(11)
clien_ctr_socio	int(11)

Tabla 7 Dimensión clientes

Fuente: Propia

b. Dimensión Estados clientes

Tabla: estados_clientes

Contiene la información necesaria para identificar o filtrar el estado de socios, como código de estado y descripción (Activos, Inactivo, retirado, fallecido)

Campo	Tipo
<u>estad_cod_estad</u>	int(11)
estad_des_estad	varchar(45)

Tabla 8 Dimensión estados clientes

Fuente: Propia

c. Dimensión Oficina
Tabla: oficinas

Contiene la información necesaria acerca de una oficina o agencia de la Cooperativa, como código de oficina y nombre de oficina.

	Campo	Tipo
<input type="checkbox"/>	<u>ofici_cod_ofici</u>	int(11)
<input type="checkbox"/>	ofici_nom_ofici	char(50)

Tabla 9 Dimensión Oficina

Fuente: Propia

c. Dimensión tipo_cliente
Tabla: tipo_cliente

Contiene la información necesaria acerca de los tipos de clientes y realizar filtros (socios, clientes, inversionistas y garantes)

Campo	Tipo
<u>tipo_cod_clien</u>	int(11)
tipo_des_clien	varchar(45)

Tabla 10 Dimensión Tipo de clientes

Fuente: Propia

6.3.4 Tablas de Hechos

A continuación se detallan las tablas de hechos para cada uno de los módulos de igual forma como anteriormente se lo hizo con las dimensiones:

1. Módulo de Clientes

a. Tabla de hechos clientes

Tabla: fact_clientes

Campo	Tipo
<u>clien_cod_ofici</u>	int(11)
<u>clien_cod_clien</u>	int(11)
<u>clien_fec_entra</u>	datetime
<u>clien_fec_salid</u>	datetime
<u>clien_ctr_socio</u>	int(11)
<u>clien_cod_estad</u>	int(11)
<u>clien_val_ingre</u>	decimal(32,2)

Tabla 11 Tabla de hechos clientes

Fuente: Propia

Contiene la información necesaria acerca de los valores numéricos que son las medidas para los clientes, como son: clien_val_ingre que representa el valor de los ingresos de los socios.

Nombre de la columna	Descripción	Tipo	Tamaño
<u>clien_cod_ofici</u>	Código de la oficina	int	11
<u>clien_cod_clien</u>	Código del cliente	int	12
<u>clien_fec_entra</u>	Fecha de entrada	datetime	11
<u>clien_fec_salid</u>	Fecha de salida	datetime	11
<u>clien_ctr_socio</u>	Código de control de tipo de cliente	Int	11
<u>clien_cod_estad</u>	Código de control de estado del cliente	datetime	-
<u>clien_val_ingre</u>	Valor de los ingresos económicos del clientes	decimal	(32,2)

Tabla 12 Tabla de Hechos clientes

Fuente: Propia

2. Módulo de Inversiones

a. Tabla de hechos inversiones

Tabla: fact_inversiones

	Campo	Tipo
<input type="checkbox"/>	Field1	int(11)
<input type="checkbox"/>	Field2	varchar(12)
<input type="checkbox"/>	Field3	int(11)
<input type="checkbox"/>	Field4	int(11)
<input type="checkbox"/>	Field5	int(11)
<input type="checkbox"/>	Field6	datetime
<input type="checkbox"/>	Field7	decimal(16,4)
<input type="checkbox"/>	Field8	decimal(16,4)

Tabla 13 Tabla de hechos inversiones

Fuente: Propia

Contiene la información necesaria acerca de los valores numéricos que son las medidas para las inversiones, como son:

Field7.- Tasa de la inversión

Field8.- Monto de inversión

La medida para el número de inversiones se la obtiene del conteo de las inversiones con el campo Field1 correspondiente a la oficina.

Nombre de la columna	Descripción	Tipo	Tamaño
Field1	Código de la oficina	int	11
Field2	Código de la inversión	varchar	12
Field3	Código de cliente	int	11
Field4	Código de estado	int	11
Field5	Código de tipo de inversión	int	11
Field6	Fecha de inicio de la inversión	datetime	-
Field7	Tasas de la inversión	decimal	16,4
Field8	Monto de la inversión	decimal	16,4

Tabla 14 Tabla de hechos Inversiones

Fuente: Propia

2. Módulo de Cartera

a. Tabla de hechos cartera de crédito

Tabla: fact_liquidaciones

	Campo	Tipo
<input type="checkbox"/>	Field1	int(12)
<input type="checkbox"/>	Field2	char(12)
<input type="checkbox"/>	Field3	int(11)
<input type="checkbox"/>	Field4	int(11)
<input type="checkbox"/>	Field5	int(13)
<input type="checkbox"/>	Field6	datetime
<input type="checkbox"/>	Field7	decimal(16,2)
<input type="checkbox"/>	Field8	decimal(16,2)

Tabla 15 Tabla de hechos créditos

Fuente: Propia

Contiene la información necesaria acerca de los valores numéricos que son las medidas para las inversiones, como son:

Field7.- Saldo del crédito

Field8.- Tasa de crédito

La medida para el número de créditos se la obtiene del conteo de los créditos con el campo Field1 correspondiente a la oficina.

Nombre de la columna	Descripción	Tipo	Tamaño	
Field1	Código de la oficina	int	11	
Field2	Código de la crédito	char	12	
Field3	Código de cliente	int	11	
Field4	Código de usuario	int	11	
Field5	Código de estado del crédito	int	13	
Field6	Fecha de inicio crédito	datetime	-	aaaa-mm-dd
Field7	Valor del crédito(Saldo)	decimal	16,2	

Field8	Tasa del crédito	decimal	16,2	
--------	------------------	---------	------	--

Tabla 16 Tabla de hechos Cartera de crédito

Fuente: Propia

6.3.5 Diseño del modelo dimensional

Luego de haber determinado las tablas que funcionarán como dimensiones y tablas de hechos para los diferentes módulos y de haber presentado una solución de varios data marts se presenta ahora el diseño multidimensional de cada data mart creado.

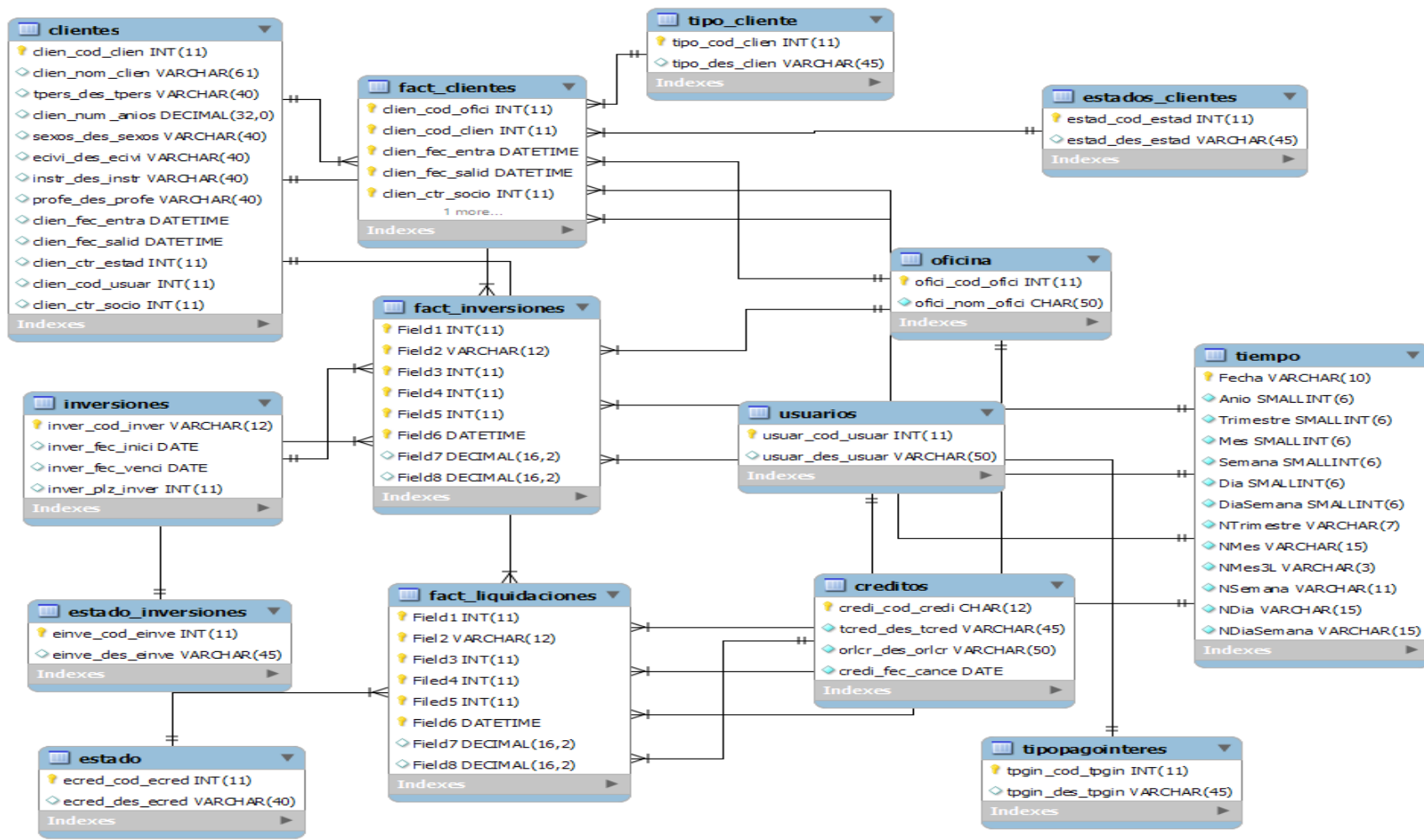


Ilustración 35 Modelo dimensional

6.4 Diseño Técnico de la Arquitectura

La solución es un sistema de información que se conforma de varias tecnologías utilizadas para implementar la solución orientada al usuario final, con la capacidad de integrar los datos que produce cada Programa Social y transformarlos en información activa y productiva para la toma de decisiones.

Por lo expuesto, el sistema de información se enmarca en la categoría de un sistema del tipo “Inteligencia de Negocio” (Business Intelligence) apoyada en la Bases de producción del sistema financiero conexus del cual se definen las estructuras de los Data marts por cada Módulo para luego formar los cubo con Mondrian y poder observar y navegar en el Cubo mediante el Visor Web JPivot.

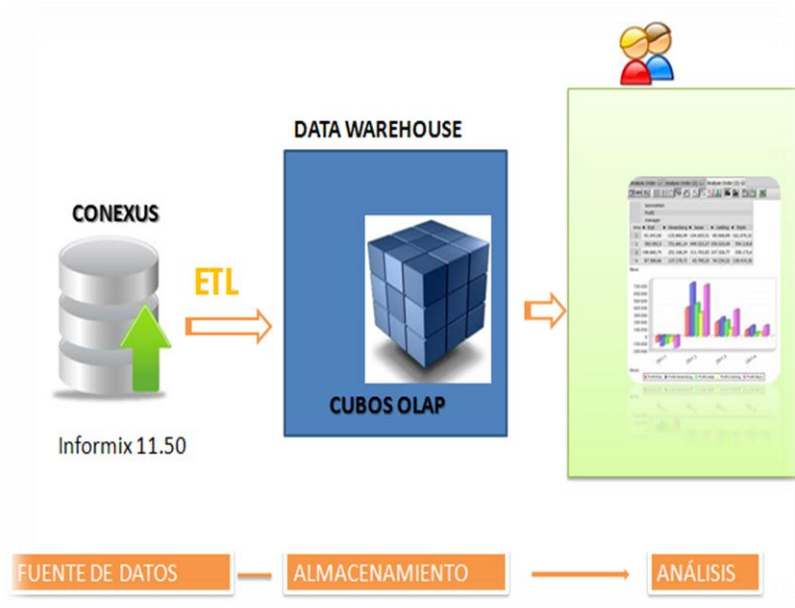


Ilustración 36 Diseño de la arquitectura

Fuente: Propia

6.4.1 Datos

Los datos constituyen la información de las bases de datos operacionales que son las fuentes, estos datos son los que alimentan las dimensiones creadas las mismas que son siempre son actualizadas a través de una tarea programada que es ejecutada periódicamente.

La tarea programada ejecuta un archivo .bat que hace la llamada al programa kitchen.bat o kitchen.sh encargado de correr trabajos realizados en Pentaho Data Integration el cual actualiza el repositorio de información.

El contenido de archivo .bat es el siguiente:

```
C:\Pentaho\data-integration\Kitchen.bat /rep:"repositorio"  
/job:JobCargaTransformaciones /dir:/cooperativa /user:admin /pass:admin  
/param:org=22 /level:Basic >> C:\log\log.log
```

Esta tarea se ejecuta cada hora para mantener la base de datos de información actualizada y que se pueda ver con claridad los cambios que se realizan en los módulos del sistema. Además guarda en un archivo .log la información de la ejecución de la tarea (job)

6.4.2 Mapeo de los datos en los modelos dimensionales

Para consolidar los datos que formarán parte de los data mart, se ha mapeado los datos en transformaciones, las mismas que a través de procedimientos almacenados o consultas sql realizan la extracción, transformación y carga de los datos a las dimensiones y tablas de hechos de los data marts.

Se detalla un ejemplo de mapeo de datos de las tablas del sistema OLTP al Data Warehouse.

Para un mejor entendimiento se ha denotado a las transformaciones con las letras TR seguidas de un nombre descriptivo en este caso la transformación del ejemplo se llama TRInversiones

Para denotar e identificar a las tablas de hechos se ha utilizado la palabra en inglés FACT (significa hecho) seguido de un guión medio y una palabra descriptiva de la tabla de hecho por ejemplo FACT_Inversiones

En este ejemplo se muestra la dimensión Inversiones donde se carga en una tabla los datos que serán posteriormente los filtros para las búsquedas. Para más información sobre cómo realizar la carga de dimensiones y tablas de hechos consultar el ANEXO 4

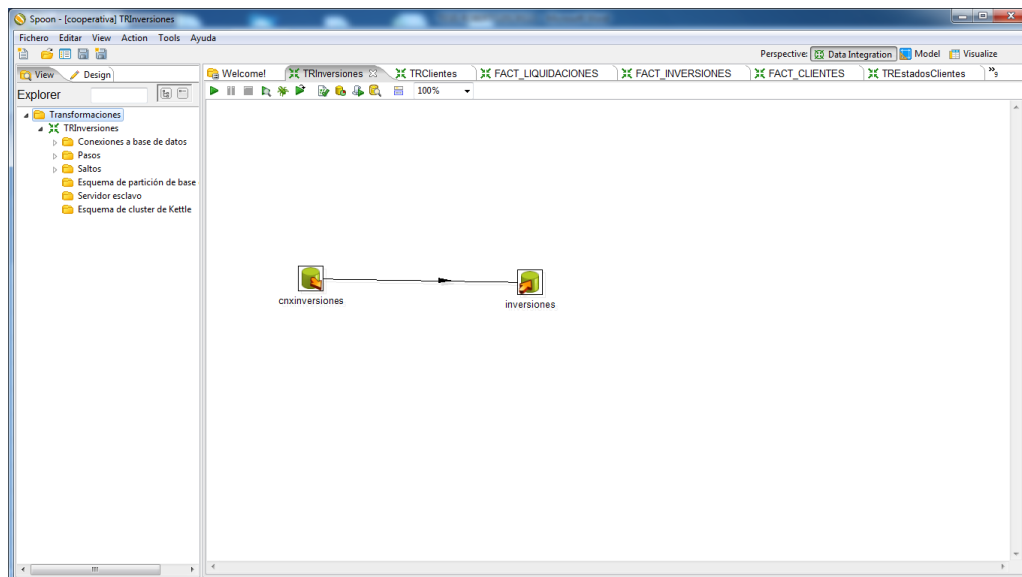


Ilustración 37 Ejemplo de un proceso ETL

Fuente: Propia

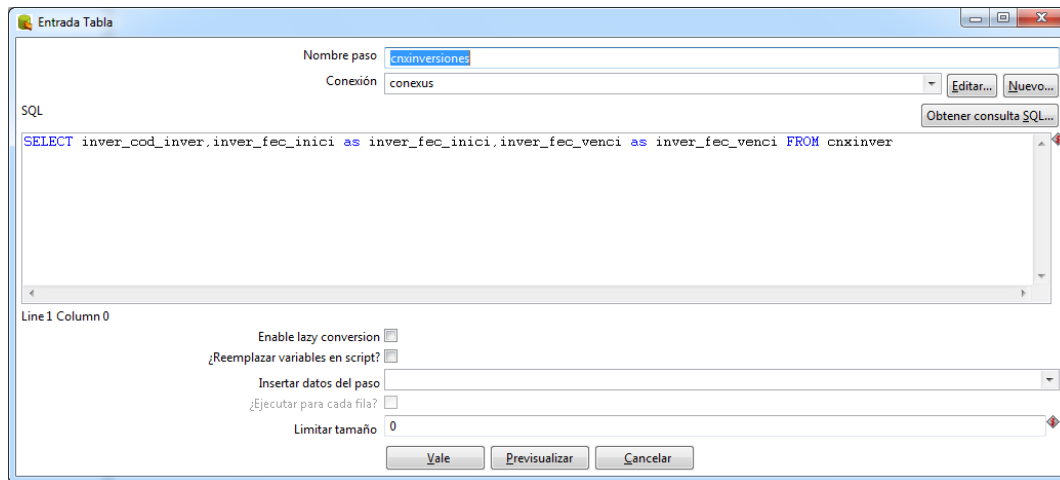


Ilustración 38 Ejemplo de entrada tabla cnxinversiones

Fuente: Propia

En esta parte en la configuración de la salida de Tabla se realiza el mapeo de los campos de la Entrada de la tabla cuyo origen es la base de datos del sistema conexus hacia los campos del repositorio o data warehouse llamado cooperativa_dw.

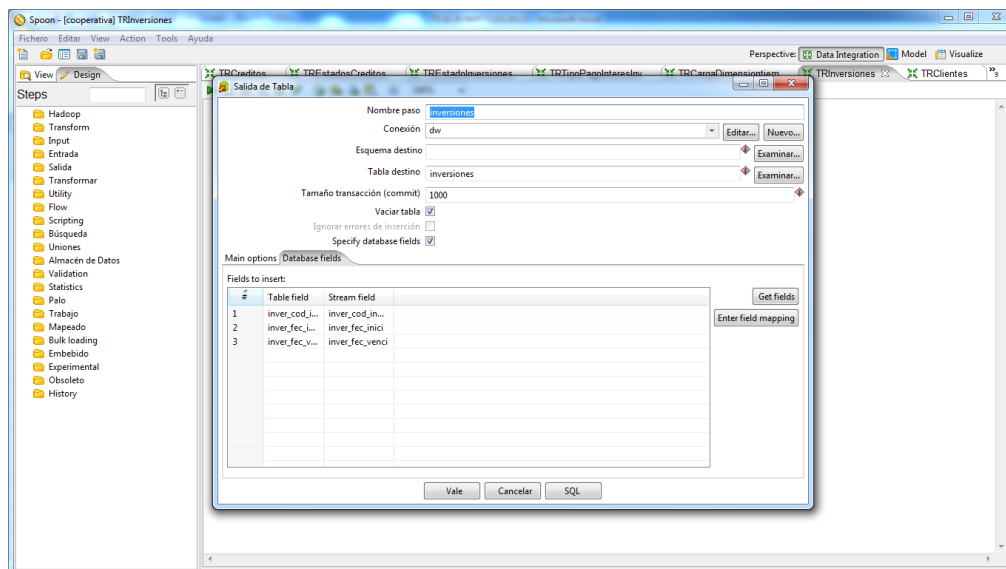


Ilustración 39 Ejemplo salida a tabla del data warehouse

Fuente: Propia

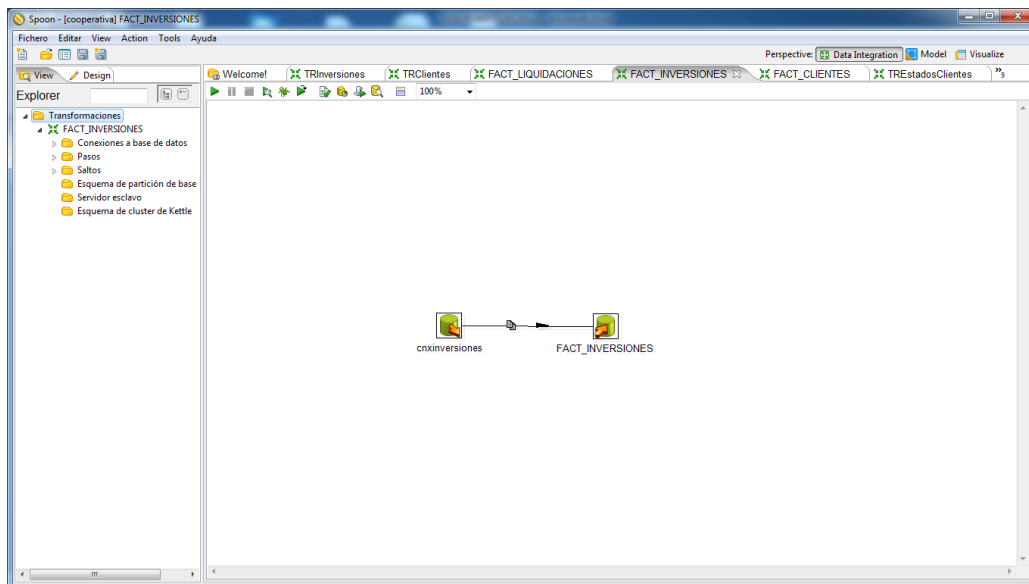


Ilustración 40 Ejemplo de proceso ETL en una tabla de hechos

Fuente: Propia

En la ilustración anterior podemos observar la tabla de hecho llamada FACT_INVERSIONES la cual carga los datos para el análisis de la información.

En la siguiente ilustración se puede observar el Job o Trabajo completo donde se realiza la carga ordenada de todas las transformaciones seguidas finalmente por la carga de todas las tablas de hechos las cuales corresponden a los módulos que van a ser analizados como son Inversiones, Clientes y Créditos (Liquidaciones)

Las herramientas software que se utilizan son las siguientes:

DESCRIPCIÓN	LICENCIA	TIPO	CANTIDAD
IDS Informix 11.50	Enterprise	x64	Propietario
Mysql Server 5.0.51b-community	Free	X32	Propietario
Jdk-6u7-windows-i586-p	GNU/GPL		
Mondrian-Jpivot	GNU/GPL		
The Apache Tomcat 6.0 Servlet/JSP Container	GNU/GPL		
Pentaho Data Integration	GNU/GPL		
Pentaho Server	GNU/GPL		
Schema-workbench	GNU/GPL		

Tabla 17 Herramientas software para BD y Análisis

Fuente: Propia

6.4.4 Construcción de los Cubos

El análisis OLAP de la solución es realizado con la herramienta Schema WorkBench. Por lo tanto, los cubos que reflejan el diseño del modelo dimensional de los data marts serán construidos de acuerdo a los requerimientos de Mondrian para tal fin.

Se deben crear dos archivos para utilizar un cubo en Mondrian, en las siguientes secciones se describen las estructuras de estos archivos.

Luego de haber establecido las estructuras de estos archivos, un usuario puede navegar por las estructuras (dimensiones, atributos y medidas) de forma intuitiva e interactiva, utilizando únicamente el Mouse desde la interfaz gráfica ofrecida por Mondrian-Jpivot.

Mondrian maneja dimensiones, niveles, categorías y medidas. Estas determinan la ruta posible de drill down y drill up en la herramienta OLAP Mondrian. Los valores que toma cada nivel son conocidos como las categorías. Las Medidas

representan los indicadores de gestión del negocio para analizar (datos numéricos).

Estructuras XML

Mondrian para entenderse con los data marts desarrollados, diseñada con el modelo dimensional, usa un archivo XML. Este archivo es en el cual se describen las dimensiones, medidas y cubos que se usan en el proceso de análisis multidimensional de los data marts.

En estas descripciones se asocian los nombres de los campos que se utilizaron en los data marts de las bases de datos operacionales.

El archivo XML que define las dimensiones y las medidas usadas en el cubo que soporta el data mart.

Los esquemas XML que se han desarrollado uno por cada requerimiento de reportes citados al inicio de este capítulo es decir 3 cubos OLAP, uno para análisis de créditos otro para análisis de inversiones y otro para análisis de los clientes.

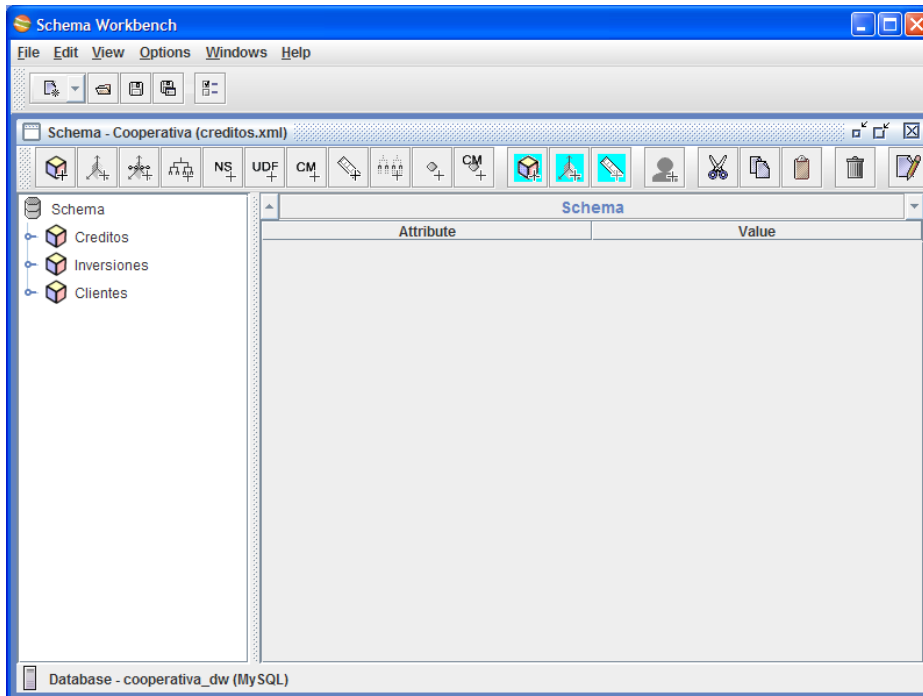


Tabla 18 Cubos diseñados con Schema Workbench

Fuente: Propia

Estructura de los cubos

Para aclarar la implementación de cada cubo OLAP en primer lugar luego de tener listos nuestros data marts, procedemos a desarrollar el esquema XML del cubo.

```

1  <Cube name="Inversiones" visible="true" cache="true" enabled="true">
2  <Table name="fact_inversiones">
3  </Table>
4  <Dimension type="StandardDimension" visible="true" foreignKey="Field1" highCardinality="false" name="Oficina">
12 <Dimension type="StandardDimension" visible="true" foreignKey="Field3" highCardinality="false" name="Clientes">
20 <Dimension type="StandardDimension" visible="true" foreignKey="Field2" highCardinality="false" name="Inversiones">
34 <Dimension type="StandardDimension" visible="true" foreignKey="Field4" highCardinality="false" name="Estado de inversiones">
42 <Dimension type="StandardDimension" visible="true" foreignKey="Field5" highCardinality="false" name="Tipo pago intereses">
50 <Dimension type="StandardDimension" visible="true" foreignKey="Field6" name="Fecha inicio">
66 <Measure name="Total en inversiones" column="Field8" aggregator="sum" visible="true">
67 </Measure>
68 <Measure name="Nº de inversiones" column="Field2" aggregator="count" visible="true">
69 </Measure>
70 <Measure name="Tasa de interes promedio (%)" column="Field7" aggregator="avg" visible="true">
71 </Measure>
72 <Measure name="Monto promedio de depositos" column="Field8" aggregator="avg" visible="true">
73 </Measure>
74 </Cube>
75

```

Ilustración 42 Estructura de un cubo en XML

Fuente: Propia

Donde en primer lugar se define el esquema del cubo, teniendo en cuenta que un esquema puede tener uno o varios cubos y de igual manera el cubo tiene una o varias dimensiones las cuales tienen una o varias jerarquías con sus niveles, y sin olvidarnos por cierto de las medidas y también las medidas calculadas del cubo.

6.5 Reportes Implementados

Los reportes estadísticos business intelligence implementados cumplen con los requerimientos. Su implementación está hecha sobre Mondrian el servidor OLAP de Pentaho, y con el uso de su interfaz gráfica es posible ver estos reportes y analizar la información con tablas de datos o gráficas de diversos tipos (barras, pies, líneas, etc.).

Para este análisis de información se dan diferentes posibilidades de navegación, como el drill up y drill down. Además es posible analizar solo algunas dimensiones, o incluir a varias para un reporte. También es posible restringir en un análisis a determinados miembros de una dimensión que sean del interés del análisis, en vez de incluir a todos.

En las siguientes secciones se documentan estos reportes en forma gráfica, aunque en la herramienta también se muestra la tabla de resultados de los datos.

Es importante notar que no es necesario escribir un query MDX para navegar por el cubo OLAP.

Esta navegación puede ser hecha de forma sencilla e interactiva con el usuario usando solo el mouse, seleccionando las opciones que se quieren. Además de los reportes acordados es posible realizar muchos más, haciendo uso de las

características de Mondrian y de las dimensiones implementadas. Se mostrará un solo reporte por módulo para dar claridad al documento.

A continuación se muestra una lista de varios ejemplos que son el resultado de los reportes implementados

1. Modulo Clientes

a. **Reporte:** Ver socios por oficinas

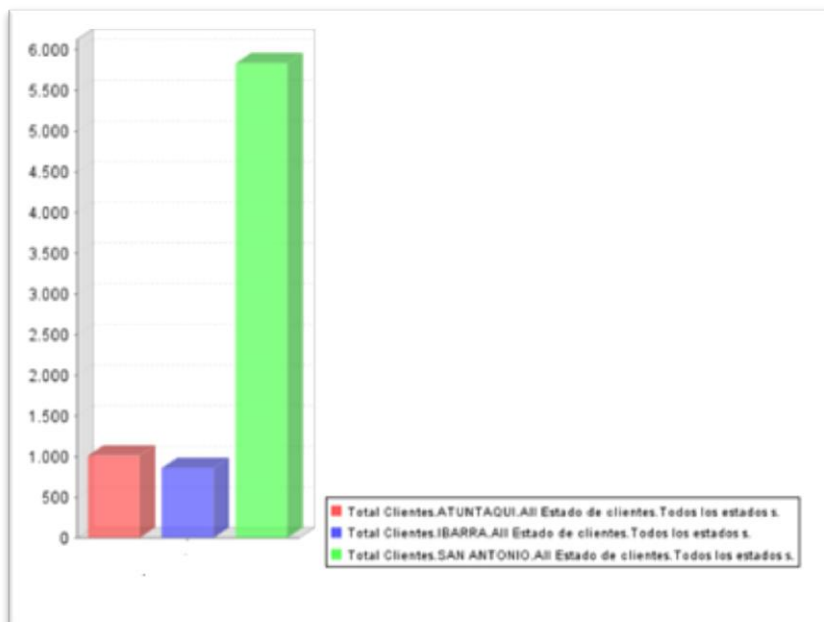


Ilustración 43 Resultados gráficos del reporte: Socios por oficinas

Fuente: Propia

b. **Reporte:** Ver socios por género

				Medidas					
				Total Clientes					
				Genero					
				FEMENINO		MASCULINO		NO DEFINIDO	
				Todos los estados		Todos los estados		Todos los estados	
Oficinas	Todos los tipos de clientes	Fecha de entrada	Fecha de Salida	● ACTIVOS	● INACTIVOS	● ACTIVOS	● INACTIVOS	● ACTIVOS	● INACTIVOS
☐ Oficinas	SOCIO	☐ All Tiempo(Fec.Entrada),Fecha de entradas	☐ All Tiempo (Fec.Salida),Fecha de Salidas	2.739	18	2.869	29	60	1

Ilustración 44 Resultados del reporte: Socios por género

Fuente: Propia

c. Reporte: Socios retirados por período

				Medidas			
				Total Clientes			
				Oficinas			
Genero	Todos los tipos de clientes	Todos los estados	Fecha de entrada	Fecha de Salida	● ATUNTAQUI	● SAN ANTONIO	● IBARRA
☐ Genero	SOCIO	RETIRADO	☐ All Tiempo(Fec.Entrada),Fecha de entradas	☐ 2012	27	98	58
			☐ T1/2012		5	13	6
			☐ T2/2012		10	24	10
			☐ T3/2012		7	28	28
			☐ T4/2012		5	33	14

Ilustración 45 Resultados del reporte: Socios retirados por períodos

Fuente: Propia

2.- Modulo de inversiones.

a. Reporte: Inversiones por oficina

					Medidas			
					Número de Inversiones			
					Oficinas			
Todos los clientes	Inversiones	Estados de inversiones	Tipo intereses	Fecha de inicio	Oficinas	ATUNTAQUI	IBARRA	SAN ANTONIO
Al Clientes.Todos los clientes	Al Inversiones	Al Estado de inversiones.Estados de inversiones	Al Tipo pago Intereses.Tipo Interesses	Fecha de inicio	457	60	39	358

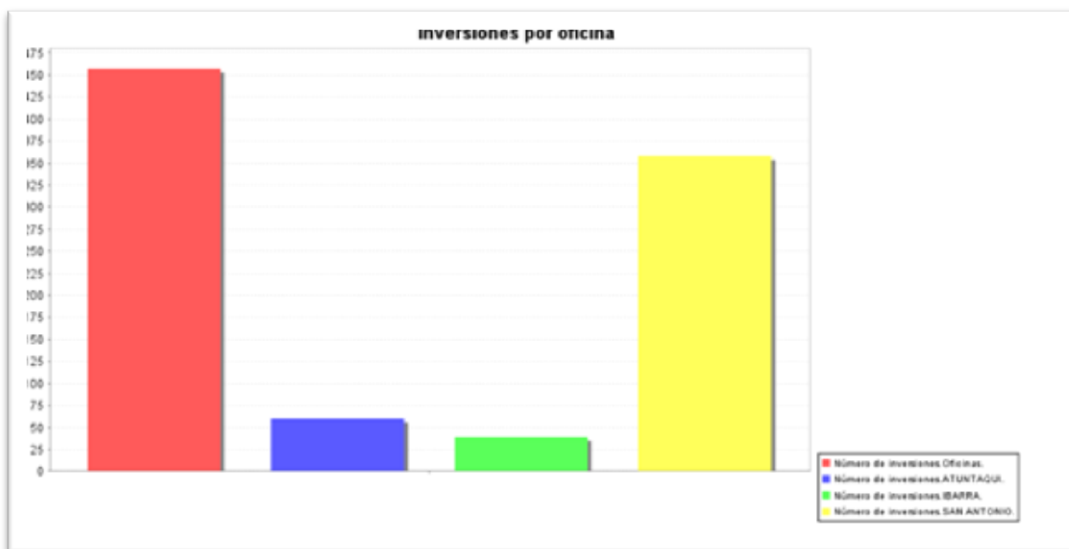


Ilustración 46 Resultado de reporte Inversiones por oficina

Fuente: Propia

3. Módulo de cartera de crédito

a. Reporte: Ver cartera de crédito por estado

						Medidas	
Oficinas	Linea de credits-Origen de recursos	Usuarios	Clientes	Estados	Fecha inicio del credito	Número de créditos	
Oficina	Linea de credito-Origen de recurso	Usuario	Cliente	Estado	Fecha de inicio de credito	2.786	
				ACTIVO	Fecha de inicio de credito	2.397	
				RECLASIFICADO	Fecha de inicio de credito	326	
				VENCIDO	Fecha de inicio de credito	63	

Ilustración 47 Cartera por estado

Fuente: Propia

b. Reporte: Ver consolidado de cartera de crédito

						Medidas
Oficinas	Linea de creditos-Origen de recursos	Usuarios	Cientes	Estados	Fecha inicio del credito	• Número de créditos
Oficina	Linea de credito-Origen de recurso	Usuario	Cliente	Estado	Fecha de inicio de credito	2.786
ATUNTAQUI	Linea de credito-Origen de recurso	Usuario	Cliente	Estado	Fecha de inicio de credito	604
IBARRA	Linea de credito-Origen de recurso	Usuario	Cliente	Estado	Fecha de inicio de credito	667
SAN ANTONIO	Linea de credito-Origen de recurso	Usuario	Cliente	Estado	Fecha de inicio de credito	1.515

Ilustración 48 Consolidado de cartera de crédito

Fuente: Propia

c. Reporte: Ver cartera por líneas de crédito

						Medidas
Oficinas	Linea de creditos-Origen de recursos	Usuarios	Cientes	Estados	Fecha inicio del credito	• Número de créditos
Oficina	Linea de credito-Origen de recurso	Usuario	Cliente	Estado	Fecha de inicio de credito	2.786
	CONSUMO	Usuario	Cliente	Estado	Fecha de inicio de credito	382
	MICROCREDITO	Usuario	Cliente	Estado	Fecha de inicio de credito	1.459
	MICROCREDITO PSNM	Usuario	Cliente	Estado	Fecha de inicio de credito	578
	MICROCREDITO RABOBANK-UACNOR	Usuario	Cliente	Estado	Fecha de inicio de credito	2
	MICROCREDITOS COLAC	Usuario	Cliente	Estado	Fecha de inicio de credito	184
	MICROCREDITOS GRUPALES	Usuario	Cliente	Estado	Fecha de inicio de credito	13
	MIGRANTES	Usuario	Cliente	Estado	Fecha de inicio de credito	128
	VIVIENDA	Usuario	Cliente	Estado	Fecha de inicio de credito	17
	VIVIENDA FINANZAS POPULARES	Usuario	Cliente	Estado	Fecha de inicio de credito	23

Ilustración 49 Cartera por líneas de crédito

Fuente: Propia

d. Reporte: Ver cartera por origen de recursos

Oficinas	Linea de creditos-Origen de recursos	Usuarios	Cientes	Estados	Fecha inicio del credito	Medidas
Oficina	Linea de credito-Origen de recurso	Usuario	Ciente	Estado	Fecha de inicio de credito	• Número de créditos 2.786
	CONSUMO	Usuario	Ciente	Estado	Fecha de inicio de credito	382
	MICROCREDITO	Usuario	Ciente	Estado	Fecha de inicio de credito	1.459
	COLAC	Usuario	Ciente	Estado	Fecha de inicio de credito	1
	CON RECURSOS PROPIOS	Usuario	Ciente	Estado	Fecha de inicio de credito	1.432
	MICROCREDITOS ASOCIATIVOS RFR	Usuario	Ciente	Estado	Fecha de inicio de credito	21
	PSMN	Usuario	Ciente	Estado	Fecha de inicio de credito	2
	RABOBANK-UACACNOR	Usuario	Ciente	Estado	Fecha de inicio de credito	3

Ilustración 50 Cartera por origen de recursos

Fuente: Propia

e. Reporte: Ver cartera por períodos

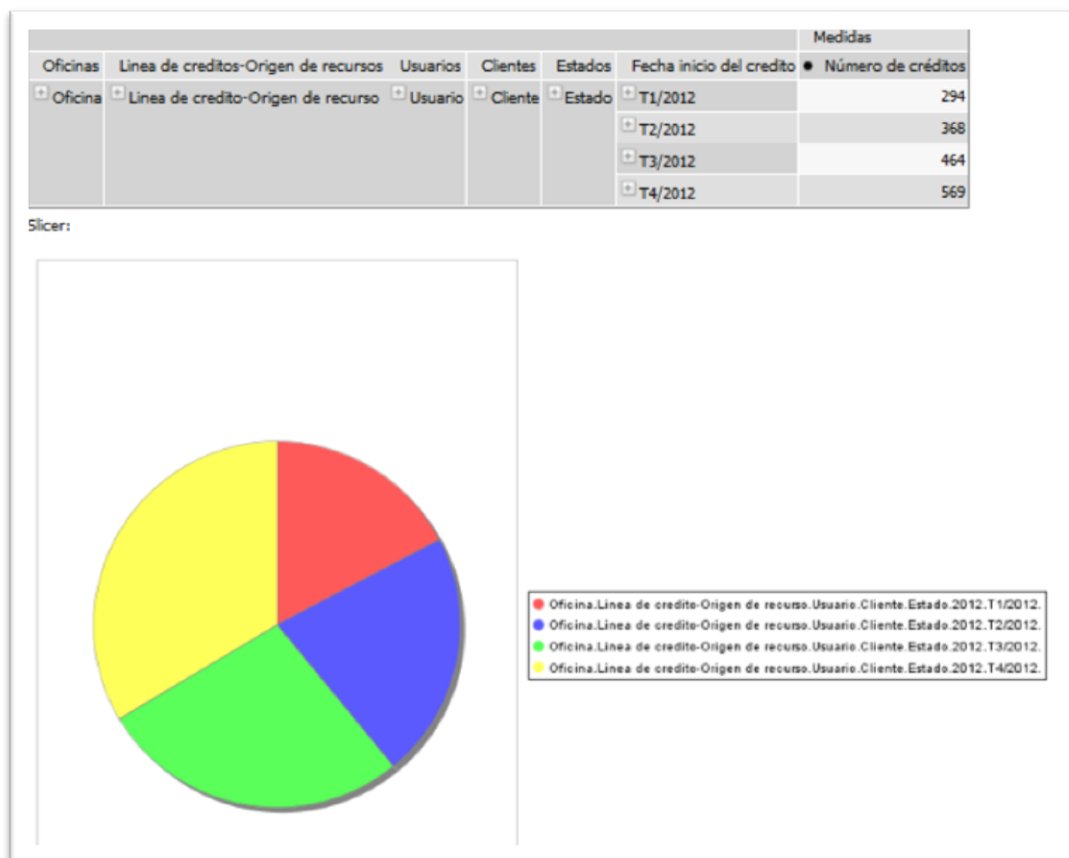


Ilustración 51 Cartera de crédito por períodos

Fuente: Propia

6.6 Mantenimiento y crecimiento

Luego de pasar exitosamente la etapa de pruebas de todo el sistema y ser aceptado por la gerencia general, el Data Warehouse está actualmente en producción. Los directivos de la Cooperativa recibieron el sistema con mucho entusiasmo, satisfacción y expectativa. Tanto la Cooperativa como para mi persona estamos convencidos de haber hecho un producto de excelente calidad, pues se basó en una metodología de desarrollo especializada en DW, que se siguió con prudencia, lo que garantizó el éxito del proyecto.

Se realizó la capacitación en el uso del producto, confirmando la aceptación y acogida que tuvo el proyecto. La capacitación fue rápida ya que la herramienta es fácil y orientada al usuario.

En cuanto al mantenimiento y crecimiento del repositorio de información, se puede decir que se hará actualizaciones diarias cada hora, por el momento para analizar el funcionamiento y mantener la información diaria y oportuna.

En cuanto al crecimiento, la tabla tendrá un mayor crecimiento es la tabla de hechos de liquidaciones (FACT_LIQUIDACIONES), ya que la liquidación de créditos es un proceso que se realiza diariamente. No existe riesgo alguno de que los datos se pierdan o entren en conflicto con la aplicación del sistema, ya que ambas bases de datos tanto de origen (fuente) como destino de información son bases de datos que garantizan y soportan la cantidad de información que puede ser generada por la Cooperativa.

La carga actual de datos al repositorio se la realiza en un tiempo aproximado de 15 segundos, tal como lo muestra el histórico log de carga y análisis generado por el PDI.

CAPÍTULO VII

CONCLUSIONES, ANÁLISIS DE IMPACTO Y RECOMENDACIONES



CONCLUSIONES

Como resultado del desarrollo del presente documento de tesis en donde se realizó el estudio de metodologías de Data Warehouse para la implementación de repositorios de información para la toma de decisiones gerenciales, y luego la aplicación de la metodología seleccionada en la implementación de una aplicación utilizando la herramienta de Inteligencia de negocios Pentaho para crear reportes que permitan conseguir los objetivos planteados, se muestran las siguientes conclusiones:

- La implementación de herramientas de Inteligencia de negocios en las empresas colabora al mejoramiento de la administración y gestión de los datos, mostrando una mejor visión del estado actual e histórico de las empresas o negocios a través de la toma de decisiones oportunas.
- La utilización de una metodología de desarrollo tanto para implementaciones de software como de desarrollo de Data Warehouse permiten obtener productos de calidad y en tiempos relativamente cortos ya que se conoce los pasos a seguir y las posibles complicaciones que se puede tener en el transcurso.
- El uso de la metodología Ralph Kimball representa un proceso eficaz en tiempo y recursos debido a que se obtiene la solución al problema en corto plazo, acoplándose a la metodología tradicional de desarrollo de software.
- Existen excelentes herramientas de software libre para el desarrollo de sistemas de inteligencia de negocios. La versión community de Pentaho por ahora está siendo mantenido por la comunidad, pero actualmente existe alternativas a jpivot como visor OLAP.
- Cuando se trata de reportes que utilizan una gran cantidad de información el usuario debe tener conocer las dimensiones y medidas que van a interactuar en el reporte BI, para mejorar el rendimiento.

ANÁLISIS DE IMPACTO

Existen cinco tipos generales de beneficios que otorga la DW y BI : ahorros de tiempo para los usuarios y para los proveedores de data, mayor y mejor información, mejores decisiones, mejora de los procesos de negocios, y apoyo para la obtención de objetivos estratégicos

El éxito del Data Warehouse no está en la construcción sino en utilizarlo para mejorar los procesos empresariales, operacionales y de toma de decisiones, para que esto suceda se deben tener en cuenta los impactos producidos.

Impacto en la gente

La construcción requiere de la participación activa de quienes utilizarán el Data Warehouse, depende tanto de la realidad de la empresa como de las condiciones que existan en ese momento, las cuales determinarán cuál será su contenido.

El Data Warehouse provee los datos que posibilitará a los usuarios a acceder a su propia información en el momento que la necesitan. Esta posibilidad para entregar información presenta varias implicancias:

- Los usuarios deberán adquirir nuevas destrezas.
- Se eliminará los largos tiempos de análisis y programación para obtener información. Como la información estará lista para ser utilizada, probablemente, aumenten las expectativas.
- Se reducirá hasta casi eliminarse la gran cantidad de reportes en papel.
- La madurez del Data Warehouse dependerá del uso activo y retroalimentación de sus usuarios.

Impactos en los procesos empresariales y de toma de decisiones

- Mejora del proceso de toma de decisiones por medio de la disponibilidad de la información.
- Las decisiones se toman más rápidamente por gente más informada.
- Los procesos empresariales pueden ser optimizados, se elimina el tiempo de espera de información que, generalmente, es incorrecta o no se encuentra.
- Se reducen los costos de los procesos y muchas veces se aclaran sus conexiones y dependencias, aumentando así la eficiencia en dichos procesos.
- El Data Warehouse permite que los datos de los sistemas operaciones sean utilizados y examinados, cuando estos datos se organizan para tener significado para la empresa la gente comienza a observar posibles defectos de las aplicaciones actuales.
- Aumenta la confianza de las decisiones tomadas en base a la información del Data Warehouse, debido a que tanto los responsables de la toma de decisiones como los afectados conocen que están basadas en información de buena calidad.

RECOMENDACIONES

A partir del desarrollo del presente trabajo se dará algunas conclusiones respecto del mismo, que los usuarios deberán tomar en cuenta para la correcta utilización de los software OLAP que han sido objeto de este estudio:

Se recomienda también que si usted utiliza bases de datos en cualquier DBMS difundido en el mercado, y necesita realizar análisis multidimensional sobre esos datos que además tienen estructuras complejas, lo haga con Mondrian el mismo que permite realizar conexiones JDBC.

Implementar Mondrian como herramienta OLAP en una empresa, requiere de un equipo que como mínimo tenga 8 GB de memoria RAM y un buen procesador para que ofrezca un mejor rendimiento en su reporte dinámico, además de configurar la máquina virtual de java en el archivo de inicio del servidor sea este Tomcat el cual se utilizó para la implementación de esta solución o cualquiera para que el procesamiento y presentación de resultados sea eficiente.

Al trabajar con Mondrian y si no se conoce tan a fondo el desarrollo de un análisis OLAP se recomienda utilizar la herramienta Schema Workbench de Pentaho que es totalmente Open Source y es una gran ayuda para el desarrollo de cubos OLAP con Mondrian.

Se recomienda además como posible tema de tesis para otros estudiantes profundizar el estudio de las herramientas de Pentaho que pueden ser utilizadas para procesos de migración de datos.

BIBLIOGRAFÍA

- Bernabeu, D. (7 de 05 de 2009). *Dataprix Knowledge is the goal*. Obtenido de <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/-metodologia-hefesto/51-introduccion>: <http://www.dataprix.com>
- Bouman, R., & Van Dongen, J. (2009). *Pentaho Solutions Business Intelligence and Data Warehousing with Pentaho and Mysql*. Indianapolis: Willey.
- Cano, J. L. (2007). *Business Intelligence: Competir con información*. Madrid: Fundación Cultural Banesto.
- Dario, I. B. (2009 de abril de 21). DATA WAREHOUSING: Investigación y sistematización de conceptos. Córdoba, Argentina.
- Davenport, T. H., & Prusak, L. (1999). *Working Knowledge: How Organizations Manage What They Know*. USA.
- El Rincon de BI- Descubriendo el Business Intelligence*. (s.f.). Obtenido de <http://churriwifi.wordpress.com/category/pentaho/>
- Kimball, R., & Ross, M. (2002). *The Data Warehouse Toolkit* (Segunda edición ed.). Willey.
- Open Business Intelligence-La red del business intelligence*. (s.f.). Obtenido de <http://www.redopenbi.com/>
- stratebi-open business intelligence*. (s.f.). Obtenido de <http://www.stratebi.com/>
- Students, B. I. (23 de enero de 2011). *www.blogspot.com*. Obtenido de <http://businessintelligencemustudents.blogspot.com/2011/01/retos-y-desventajas-de-bussines.html>
- Todo Bi- Business Intelligence, Data Warehouse, CRM y mucho más*. (s.f.). Obtenido de <http://todobi.blogspot.com/>

ANEXOS

ANEXO 1

INSTALACIÓN DE SUITE PENTAHO

1) Requisitos previos: máquina virtual Java y la base de datos MySQL (u otra de las soportadas).

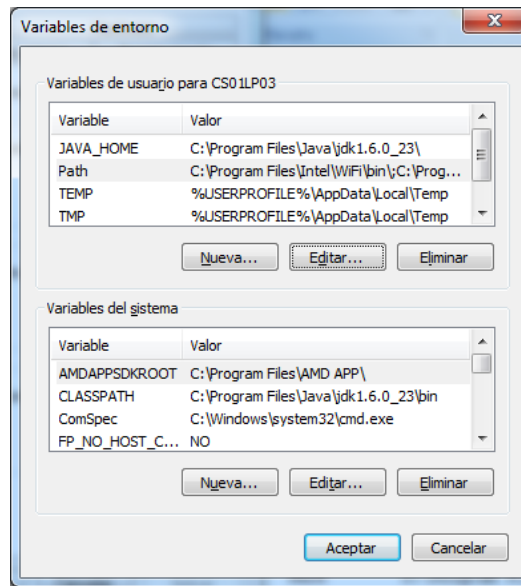
Para poder ejecutar la plataforma de BI de Pentaho es necesario disponer de una máquina virtual Java instalada en el equipo donde se va a trabajar. Pentaho recomienda la versión 1.5 de Sun JRE aunque en este proyecto se usó la versión 1.6.

Para ver la versión instalada, se ejecuta el comando: `java -version`. En el caso de no disponer de la máquina.

A continuación comprobar que la variable de entorno **JAVA_HOME** apunte al directorio donde tenemos instalado Java. Igualmente, la variable **PATH** también deberá apuntar al directorio de ejecutables de la instalación de Java. En este, el valor de las variables será el siguiente:

```
JAVA_HOME    c:\Archivos de Programa\Java\jdk1.6.0_23
PATH         c:\Archivos de Programa \Java\jdk1.6.0_23\bin
```

Para configurar las variables, lo realizaremos desde Propiedades del Sistema, Variables de Entorno.



Configuración de variables de entorno

Con respecto a MySQL, en el caso de que no se tenga instalado, se puede descargar el software de la web y realizar la instalación según las instrucciones que se proporcionan en el portal de documentación.

2) Descomprimir los ficheros de la plataforma.

Se selecciona una carpeta (por ejemplo `c:\pentaho`), y en ella vamos a descomprimir el fichero Zip descargado de la web. Tras el proceso, se mostrará dos carpetas diferenciadas, llamadas **administration-console** y **biserver-ce**. La primera carpeta alberga los ficheros de la plataforma de administración, donde se puede configurar y administrar el servidor BI. La segunda, es la plataforma de BI propiamente dicha (la que utilizarán los usuarios), que utiliza tomcat.

En este momento, ya se podría arrancar la plataforma desde los correspondientes scripts que se encuentran en la carpeta `c:\pentaho\biserver-ce` (`start-pentaho.bat` para iniciar el servidor y `stop-pentaho.bat` para pararlo). Este Script arranca en primer lugar la base de datos HSQLDB de ejemplo (donde residen los datos necesarios para el funcionamiento de la plataforma, junto con

datos de pruebas para los ejemplos precargados). A continuación, arranca la plataforma de BI, a través del tomcat. Como no se desea trabajar con la base que viene por defecto sino con MySQL, se procederá a realizar una serie de ajustes antes de arrancar la plataforma.

3) Creación de catálogos en base de datos necesarios para la plataforma.

La plataforma Pentaho necesita dos bases de datos para su funcionamiento (además de la base de datos de test para poder trabajar con el set de ejemplos). Las bases de datos y su cometido son las siguientes:

- **hibernate:** esta base de datos almacena la autenticación de usuarios y los datos de autorizaciones, el contenido BI (solution repository) y los orígenes de datos disponibles en la plataforma.
- **quartz:** es el repositorio para el scheduler Quartz, que es uno de los componentes que forma la plataforma, que nos permite la planificación de procesos dentro del servidor BI.
- **sampledate:** contiene las tablas para ilustrar y hacer posible la ejecución de todos los ejemplos por defecto que proporciona la plataforma, para poder hacernos una idea de sus funcionalidades y sus posibilidades de análisis.
- Por defecto, los catálogos de estas bases de datos estarán creados en la base de datos HSQLDB que se puede arrancar en la configuración del servidor por defecto. Para crearlos en MySQL, como es nuestro caso, ejecutaremos los scripts que se encuentran en la carpeta `c:\pentaho\biserver-ce\data`. El orden de ejecución será el siguiente:

```
mysql> source 1_create_repository_mysql.sql;
...output
mysql> source 2_create_quartz_mysql.sql;
...output
mysql> source 3_create_sample_datasource_mysql.sql;
...output
mysql> source 4_load_sample_users_mysql.sql;
...output
mysql> source 5_sample_data_mysql.sql;
...output
```

La ejecución de los scripts sql se la puede realizar desde MySQL Query Browser (la herramienta gráfica para ejecución de sentencias SQL) o bien desde línea de comandos con la utilidad MySQL que llevar incluido el servidor MySQL para ejecutar scripts.

4) Configuración JDBC, Hibernate and Quartz.

Todas las aplicaciones de Pentaho, incluyendo el Pentaho Server, utilizan la conectividad JDBC (Java Database Connectivity) para la comunicación con las bases de datos. Por tanto, será necesario disponer de los correspondientes conectores según la base de datos que vayamos a utilizar. En este caso, se deja el conector para MySQL (donde ira las bases de datos de Hibernate y Quartz y del data warehouse cooperativa_dw). Para la conexión a la base de datos Informix se realiza una conexión mediante ODBC. Las carpetas donde vamos a copiar serán las siguientes:

- **C:\Pentaho\biserver-ce\tomcat\common\lib:** ubicación de los drivers JDBC para poder utilizar en el servidor Pentaho la base de datos para la que el conector proporciona conectividad.
- **C:\Pentaho\administration-console\jdbc:** es necesario colocar aquí también los drivers para poder definir correctamente las conexiones a base de datos en la consola de administración.
- A continuación, se configurará los ficheros de parametrización del sistema para que Hibernate y Quartz lean de los catálogos de base de datos en Mysql que se ha creado en el punto 3, en lugar de la base de datos HSQLDB proporcionada por defecto.
- **Configuración de Hibernate (I):** en el fichero ***applicationContext-spring-security-jdbc.xml*** (ubicado en la carpeta C:\Pentaho\biserver-ce\pentaho-solutions\system), modificamos la parte que esta subrayada a continuación, con los valores referidos para utilizar MySQL.

```
<!-- This is only for Hypersonic. Please update this section
for any other database you are using -->
<bean id="dataSource"
```



```

class="org.springframework.jdbc.datasource.DriverManagerDataSo
urce">
<property name="driverClassName" value="com.mysql.jdbc.Driver"
/>
<property name="url"
value="jdbc:mysql://localhost:3306/hibernate" />
<property name="username" value="hibuser" />
<property name="password" value="password" />
</bean>

```

- **Configuración de Hibernate (II):** en el fichero *applicationContext-spring-security-hibernate.xml* (ubicado en la carpeta C:\Pentaho\biserver-ce\pentaho-solutions\system), modificaremos la parte que se ve subrayada a continuación, con los valores referidos para utilizar MySQL.

```

jdbc.driver=com.mysql.jdbc.Driver
jdbc.url=jdbc:mysql://localhost:3306/hibernate
jdbc.username=hibuser
jdbc.password=password
hibernate.dialect=org.hibernate.dialect.MySQLDialect

```

- **Configuración de Hibernate (y III):** en el fichero *hibernate-settings.xml* (ubicado en la carpeta C:\Pentaho\biserver-ce\pentaho-solutions\system\hibernate), modificamos la parte que se ve subrayada a continuación.

```

<config-file>
    system/hibernate/mysql5.hibernate.cfg.xml
</config-file>

```

Con lo anterior, se ha configurado la seguridad JDBC de la plataforma.

5) Configuración Servidor Apache-Tomcat.

La plataforma Pentaho utiliza Apache-Tomcat como servidor de aplicaciones para desplegar los servicios que la componen. El servidor lleva una configuración por defecto que podemos modificar (por ejemplo, para variar el puerto donde nos conectamos, para el caso de que haya conflicto con otras aplicaciones instaladas en el servidor), la ubicación HTML, el lenguaje, etc. Para ello, se debe modificar el fichero **web.xml** que se encuentra en la carpeta C:\Pentaho\biserver-ce\tomcat\webapps\pentaho\WEB-INF.

A continuación se muestra algunas configuraciones que se puede modificar.

- **solution-path**

Con este parámetro, se le indica a la plataforma BI la ubicación de la carpeta pentaho-solutions. Por defecto, tiene el valor `c:\biserver-ce\`.

En este caso, vamos a cambiar el valor para que apunte a la carpeta donde hemos instalado:

```
<context-param><param-name>solution-path</param-name>
  <param-value>C:\Pentaho\biserver-ce\pentaho-solutions</param-
value>
</context-param>
```

- **base-url**

Al instalar, la ruta URL por defecto para acceder a la plataforma será la siguiente:

`http://localhost:8080/pentaho`

Para cambiar la ruta se debe modificar el parámetro `base_url` dentro del mismo archivo. En este caso, como se cambiará el puerto por defecto, modificando su valor indicando lo siguiente:

```
<context-param>
  <param-name>base-url</param-name>
  <param-value>http://localhost:8081/pentaho/</param-value>
</context-param>
```

- **port**

En la ruta `C:\Pentaho\biserver-ce\tomcat\conf`, tenemos el fichero **server.xml**, donde podemos modificar el puerto por defecto de nuestro servidor BI (que es el 8080).

```
<!-- Define a non-SSL HTTP/1.1 Connector on port 8080 -->
<Connector port="8081" maxHttpHeaderSize="8192"
  maxThreads="150" minSpareThreads="25" maxSpareThreads="75"
  enableLookups="false" redirectPort="8443" acceptCount="100"
  connectionTimeout="20000" disableUploadTimeout="true" />
```

Ahora pararemos el servidor tomcat y al arrancar la nueva URL de acceso será la siguiente: <http://localhost:8081/pentaho>

- **Publicación de contenido:** Para habilitar la publicación, se modifica el fichero **publisher_config.xml**, que se encuentra en la carpeta `C:\Pentaho\biserver-ce\pentaho-solutions\system`. Ahí se debe indicar la contraseña de publicación. Por defecto, no tiene ninguna contraseña, y por tanto, no está habilitada la publicación.

```
<publisher-config>
  <publisher-password>
    passpublic
  </publisher-password>
</publisher-config>
```

ANEXO 2

UTILIZACIÓN DE LA BARRA DE NAVEGACIÓN OLAP


















Los botones que se mostraron en la figura son los siguientes, en su respectivo orden:

- Navegador OLAP
- Mostrar editor MDX
- Configurar tabla OLAP
- Ocultar Repeticiones
- Suprimir filas/columnas vacías
- Detallar miembro
- Abrir detalle
- Entrar en detalle
- Mostrar datos origen
- Mostrar Gráfico
- Configurar Gráfico
- Configurar impresión
- Exportar a PDF
- Exportar a Excel

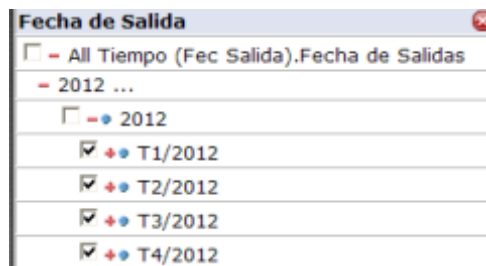
La tabla a continuación descrita, muestra una lista de la simbología empleada en la barra de navegación interactiva del reporte estadístico implementado en esta aplicación.

Esta simbología está acompañada de una descripción para cada uno de los botones y/o símbolos que componen la misma, de la siguiente manera:

Barra de Navegación de los Reportes Estadísticos	
SIMBOLOGÍA	DESCRIPCIÓN
	Navegador OLAP , le da varias opciones al usuario para personalizar su reporte estadístico, habilitando y/o deshabilitando las dimensiones o filas informativas y las medidas o columnas de valores numéricos. Además ayuda al usuario para filtrar sus resultados estadísticos por dimensiones y/o medidas.
	Mostrar Editor MDX , muestra un editor avanzado para realizar una consulta mdx, para ello el usuario debe tener un buen conocimiento en construcción de cubos utilizando el lenguaje mdx.
	Configurar tabla OLAP , muestra un cuadro de diálogo que permite configurar la tabla de forma que se pueda ordenar los registros de las dimensiones y también las medidas, por grupos y por jerarquía.
	Ocultar Repeticiones , sirve para ocultar o habilitar los nombres repetidos en el reporte estadístico.
	Suprimir Filas/Columnas vacías , permite ocultar o mostrar los registros de las dimensiones con medidas vacías.
	Intercambiar ejes , permite intercambiar los ejes de las dimensiones mostradas en el reporte estadístico.
	Detallar miembro , muestra el detalle de todos los miembros del reporte estadístico
	Abrir detalle , muestra el detalle por cada miembro del reporte estadístico seleccionado
	Entrar en detalle , muestra en detalle el reporte estadístico, sin repeticiones
	Mostrar datos origen , permite mostrar los datos origen para una medida o columna de valor numérico dada

	<p>Mostrar gráficos, muestra un gráfico estadístico de acuerdo al tipo de gráfico y especificaciones configuradas por el usuario</p>
	<p>Configurar gráficos, muestra un cuadro de diálogo de configuración de gráficos estadísticos, con varias opciones como:</p> <ul style="list-style-type: none"> • Tipo de gráfico, que son barras, barras apiladas, áreas y pasteles en dos y tres dimensiones. • Título del gráfico. • Dimensiones del gráfico • Etiquetas, para los ejes horizontal y vertical
	<p>Configurar impresión, permite configurar la impresión que se exportará a PDF, y se puede ingresar el título para el reporte, sus dimensiones y damos la facilidad al usuario de escoger si desea el reporte de forma horizontal o vertical.</p>
	<p>Exportar a PDF, exporta el resultado del reporte estadístico business Intelligence en formato PDF con las especificaciones de impresión configuradas por medio del botón anterior</p>
	<p>Exportar a Excel, exporta el resultado del reporte estadístico business Intelligence en formato EXCEL</p>

De esta forma, vamos configurando cómo será la visualización en la tabla. Además, podremos modificar la jerarquía visible en cada dimensión (en principio se verá la jerarquía por defecto, para el caso de tener varias jerarquías en la misma dimensión). También podemos seleccionar el nivel dentro de la jerarquía o los elementos a visualizar de todos los disponibles.



Jpivot - Selección miembros

Una vez realizadas las consiguientes modificaciones en el layout, pulsaremos el botón Aplicar y los cambios serán visibles en la tabla pivot.

					Medidas		
					Total Clientes		
					Oficinas		
Genero	Todos los tipos de clientes	Todos los estados	Fecha de entrada	Fecha de Salida	● ATUNTAQUI	● SAN ANTONIO	● IBARRA
⊕ Genero	SOCIO	RETIRADO	⊕ All Tiempo(Fec Entrada).Fecha de entradas	2012	23	52	48
				⊕ T1/2012	5	9	6
				⊕ T2/2012	9	15	10
				⊕ T3/2012	8	21	27
				⊕ T4/2012	1	7	5

Jpivot - Informe Modificado

Editor MDX.

Como hemos indicado antes, conforme vamos seleccionando las dimensiones e indicadores en el navegador OLAP, al actualizar la consulta, internamente se traduce a lenguaje MDX que es el que utiliza Mondrian para luego construir las sentencias SQL que se ejecutarán contra la base de datos relacional. Con este control, podemos visualizar la sentencia MDX que se ha construido de forma automática, e incluso, si dominamos este lenguaje, escribir directamente las consultas que darán como resultado la tabla pivot correspondiente.



Jpivot - Editor MDX

Habr  que tener en cuenta en la sintaxis de las sentencias **MDX** como hayamos llamado a las dimensiones, jerarqu as e indicadores.

Propiedades de visualizaci3n.

El siguiente grupo de iconos nos permite establecer propiedades de visualizaci3n de la tabla de pivoteo sobre los datos.

- **Mostrar padres:** podemos forzar la visualizaci3n de los elementos padre conforme vayamos profundizando en los datos.

Time		Medidas	
Time	Markets	• Quantity	
<input type="checkbox"/> All Years	<input type="checkbox"/> All Markets	105.331	
<input type="checkbox"/> 2003	<input type="checkbox"/> All Markets	36.439	
<input type="checkbox"/> QTR1	<input type="checkbox"/> All Markets	4.561	
<input type="checkbox"/> QTR2	<input type="checkbox"/> All Markets	5.695	
<input type="checkbox"/> QTR3	<input type="checkbox"/> All Markets	6.629	
<input type="checkbox"/> QTR4	<input type="checkbox"/> All Markets	19.554	
<input type="checkbox"/> 2004	<input type="checkbox"/> All Markets	49.417	
<input type="checkbox"/> 2005	<input type="checkbox"/> All Markets	19.475	

MOSTRAR PADRES DESACTIVADO

Time			Markets		Medidas	
(All)	Years	Quarters	(All)	• Quantity		
<input type="checkbox"/> All Years			<input type="checkbox"/> All Markets	105.331		
All Years	2003		<input type="checkbox"/> All Markets	36.439		
		<input type="checkbox"/> QTR1	<input type="checkbox"/> All Markets	4.561		
		<input type="checkbox"/> QTR2	<input type="checkbox"/> All Markets	5.695		
		<input type="checkbox"/> QTR3	<input type="checkbox"/> All Markets	6.629		
	<input type="checkbox"/> QTR4	<input type="checkbox"/> All Markets	19.554			
	2004		<input type="checkbox"/> All Markets	49.417		
	2005		<input type="checkbox"/> All Markets	19.475		

MOSTRAR PADRES ACTIVADO

Jpivot - Mostrar Padres

- **Ocultar repeticiones:** con esta opción podemos ocultar repeticiones de los valores de los miembros de una jerarquía y así hacer más comprensible el análisis de los resultados.

Ocultar repeticiones			Medidas		
Time	Markets	● Sales	Time	Markets	● Sales
+ 2003	- All Markets	3.677.384	+ 2003	- All Markets	3.677.384
	+ APAC	343.082	+ 2003	+ APAC	343.082
	+ EMEA	1.681.987	+ 2003	+ EMEA	1.681.987
	+ Japan	292.558	+ 2003	+ Japan	292.558
	+ NA	1.359.757	+ 2003	+ NA	1.359.757
+ 2004	- All Markets	4.987.740	+ 2004	- All Markets	4.987.740
	+ APAC	601.606	+ 2004	+ APAC	601.606
	+ EMEA	2.396.408	+ 2004	+ EMEA	2.396.408
	+ Japan	168.479	+ 2004	+ Japan	168.479
	+ NA	1.821.247	+ 2004	+ NA	1.821.247
+ 2005	- All Markets	1.980.825	+ 2005	- All Markets	1.980.825
	+ APAC	337.018	+ 2005	+ APAC	337.018
	+ EMEA	929.829	+ 2005	+ EMEA	929.829
	+ Japan	42.921	+ 2005	+ Japan	42.921
	+ NA	671.057	+ 2005	+ NA	671.057

Jpivot - Ocultar repeticiones

- **Suprimir filas/columnas vacías:** oculta las filas o columnas que no tuviesen valores.
- **Intercambiar ejes:** nos permite de una forma rápida pasar las filas a columnas y viceversa, y así cambiar la forma de ver la información.

Opciones de navegación.

Una opción muy interesante es determinar la forma en que se realiza la navegación por la tabla. Para ello, tenemos 3 posibles opciones.

- **Detallar miembro:** cuando navegamos por un miembro de una dimensión (por ejemplo, el mercado EMEA en la imagen anterior), independientemente de que estemos en un año u otro, se abre el desglose del miembro en todas las ocurrencias que tuviera en la tabla (en los diferentes años en el caso de nuestro ejemplo).

- **Abrir detalle:** en contraposición a Detallar miembro, con la opción Abrir detalle sobre se abre el nivel del miembro que hayamos seleccionado, no todas las ocurrencias.
- **Entrar en detalle:** cambiamos la forma de navegación, sustituyendo el icono + por una flecha, que nos permite ir bajando y subiendo por la información sin que se vaya realizando un desglose. Es una forma de navegación mucho más útil para ir analizando aspectos concretos.

Las tres opciones de navegación descritas son excluyentes entre sí (solo podremos tener seleccionada una de ellas a la vez).

					Medidas		
					Total Clientes		
					Oficinas		
Genero	Todos los tipos de clientes	Todos los estados	Fecha de entrada	Fecha de Salida	ATUNTAQUI	SAN ANTONIO	IBARRA
<input type="checkbox"/> Genero	SOCIO	RETIRADO	<input type="checkbox"/> All Tiempo(Fec Entrada).Fecha de entradas	2012	023	052	048
				T1/2012	05	09	06
				T2/2012	09	015	010
				T3/2012	08	021	027
				T4/2012	01	07	05

Jpivot - Entrar en detalle

- **Mostrar datos Origen:** con esta opción mostramos en la parte inferior de la tabla pivote una tabla adicional donde se muestran los datos originales que dan lugar a los resultados mostrados en la tabla principal. Puede ser útil para buscar determinados valores en registros individuales de datos cuando se produzca una alarma visualizando los resultados de la tabla principal.

Modo gráfico y exportación PDF/Excel.

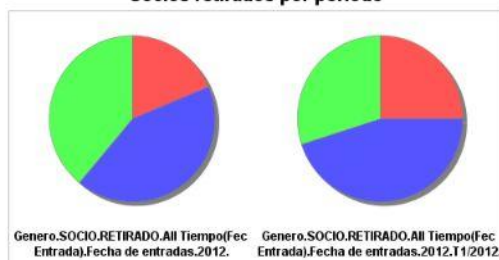
Finalmente, el último set de botones de la barra de herramientas nos permite configurar el gráfico que se muestra como complemento de la tabla pivot o realizar la exportación de los resultados en formato PDF o Excel.

- **Mostrar gráficos:** al seleccionar la opción, se visualiza adicionalmente un gráfico con los resultados de la tabla. Los tipos de gráfico que se pueden utilizar son de barras, de línea, de área o de pastel.

				Medidas			
				Total Clientes			
				Oficinas			
Genero	Todos los tipos de clientes	Todos los estados	Fecha de entrada	Fecha de Salida	● ATUNTAQUI	● SAN ANTONIO	● IBARRA
☐ Genero	SOCIO	RETIRADO	☐ All Tiempo(Fec Entrada).Fecha de entradas	☐ 2012	⊖23	⊖52	⊖48
				☐ T1/2012	⊖5	⊖9	⊖6
				☐ T2/2012	⊖9	⊖15	⊖10
				☐ T3/2012	⊖8	⊖21	⊖27
				☐ T4/2012	⊖1	⊖7	⊖5

Slicers:

Socios retirados por periodo



Jpivot - Tabla y grafico

- **Configurar gráficos:** en esta opción configuramos las propiedades del gráfico (tipo de gráfico, fuentes, títulos, color de fondo, etc.).
- **Configurar impresión:** configuramos alguna de las propiedades que tendrá el PDF que se genere en la opción Exportar a PDF (título, tamaño de tabla, orientación del papel, etc.).
- **Exportar a PDF:** genera un documento PDF con los resultados de la tabla pivot según la configuración indicada.
- **Exportar a Excel:** nos permite exportar la tabla de resultados visibles a un fichero con formato Excel.

ANEXO 4

MANUAL DE PENTAHO DATA INTEGRATION PDI

- **REQUISITOS PREVIOS A LA INSTALACIÓN DEL PDI**

Requisitos mínimos de hardware

- Procesador de arquitectura Pentium de 2.0 GHZ
- 768 MB de memoria RAM
- Disco Duro con al menos 2 GB libres

Requisitos de software

- Java run time environment 5 o posteriores
- MySQL versión 5 o posteriores

- **PASOS PARA LA INSTALACIÓN**

Los siguientes son los pasos para instalar el PDI en un computador:

- Descargar el archivo .zip del sitio web de Pentaho que contiene el PDI: pdi-ce-4.2.0-stable.zip
- Descomprimir el archivo en la ubicación C:\Pentaho donde están albergados los directorios de bi-server, consola de administración y schema-workbench.

- **PASOS PARA LA IMPLEMENTACIÓN**

De manera resumida los pasos para la implementación son:

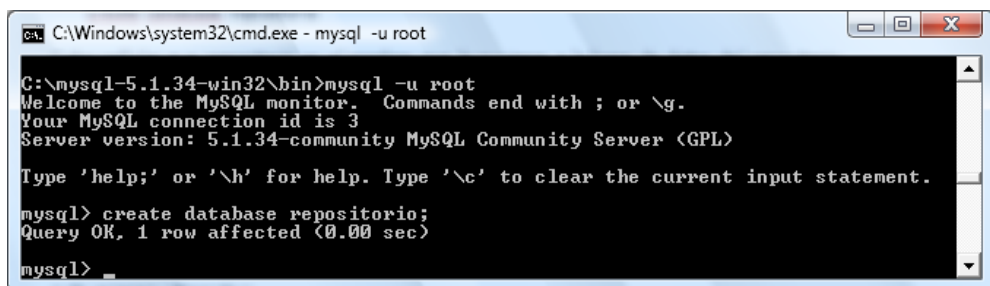
- Creación de la base de datos que contendrá el repositorio
- Creación del repositorio
- Creación de las conexiones a los orígenes de datos y al Data Warehouse
- Extracción de los datos
- Creación de una tabla INPUT
- Definición de las tablas de dimensiones
- Definición de la tabla de hecho
- Carga o exportación de la tabla de hecho

Vamos a detallar cada uno de pasos mencionados.

- **CREACIÓN DE LA BASE DE DATOS DEL REPOSITORIO**

Como era de suponer, antes de crear el repositorio debemos crear la base de datos que lo contenga. Para esto ejecutamos MySQL, ingresando a la carpeta bin de MySQL y ejecutamos:

- `mysqld --console`
- `mysql -u root`
- `create database repositorio;`



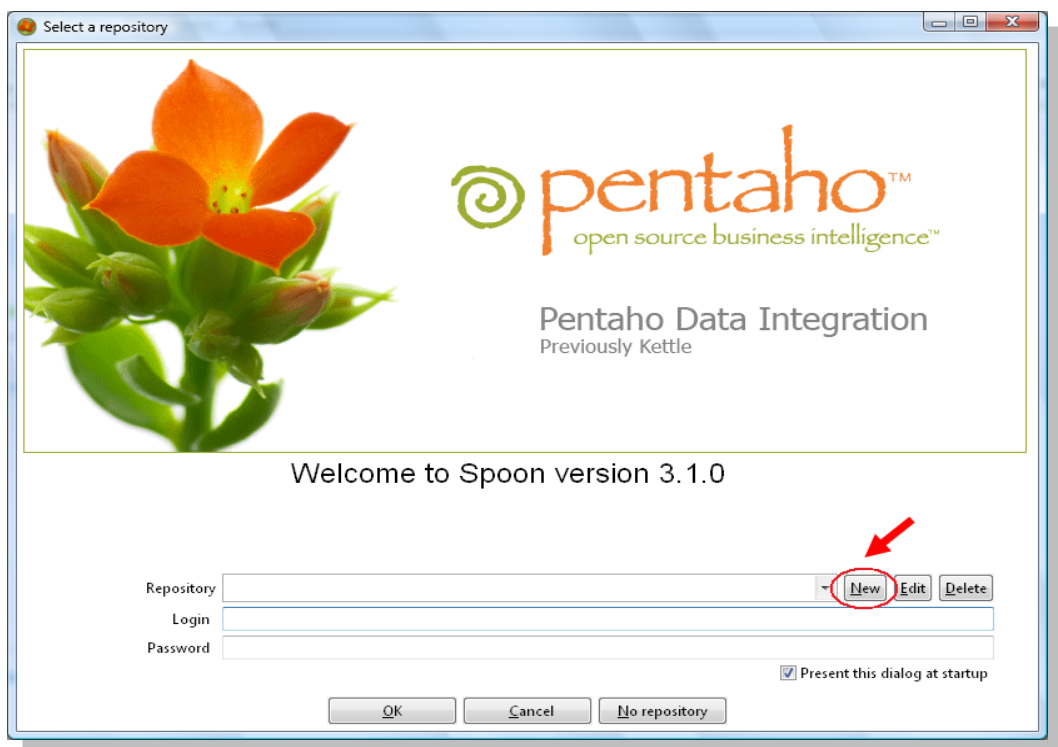
```
C:\Windows\system32\cmd.exe - mysql -u root
C:\mysql-5.1.34-win32\bin>mysql -u root
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 3
Server version: 5.1.34-community MySQL Community Server <GPL>

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.
mysql> create database repositorio;
Query OK, 1 row affected (0.00 sec)
mysql> _
```

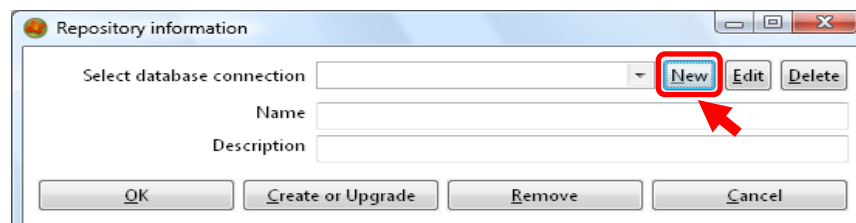
Cabe señalar que “repositorio” es el nombre que le pusimos a la base de datos del repositorio.

- **CREACION DEL REPOSITORIO**

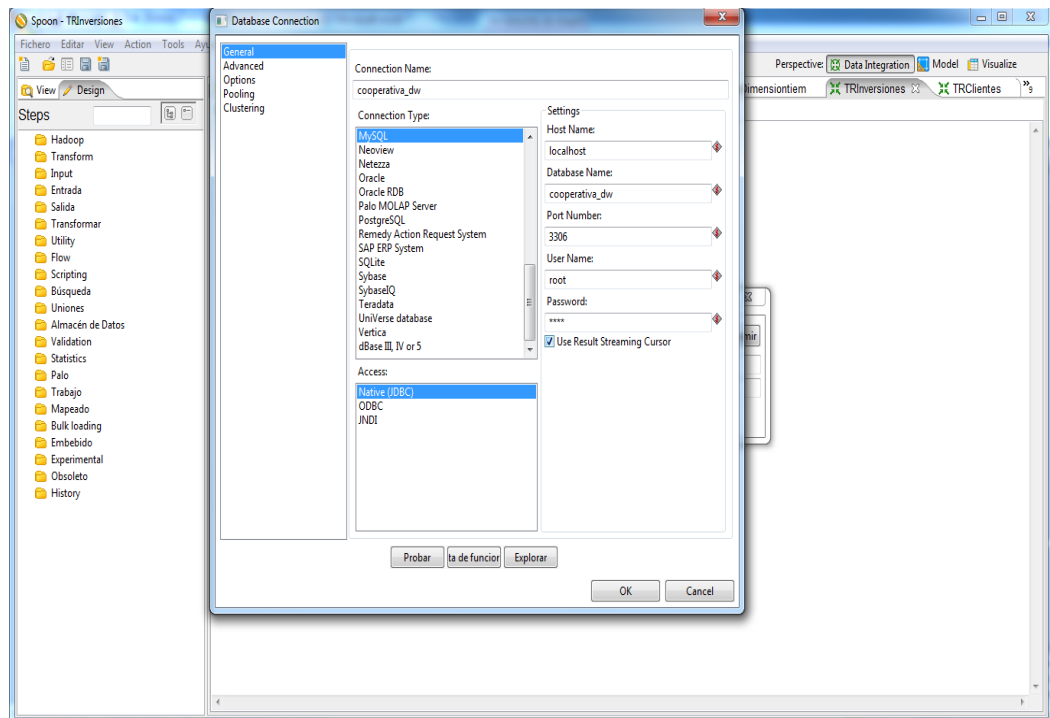
- Para crear el repositorio entramos a la carpeta donde descomprimimos el archivo del PDI y ejecutamos el archivo spoon.bat
- Se presentará un cuadro de diálogo “Selecciona un catálogo”, donde presionaremos el botón “New”.



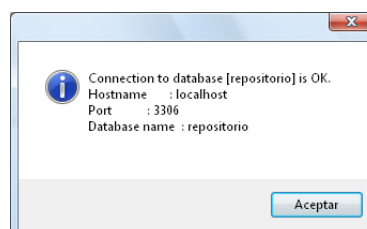
- Se nos presentará una nueva ventana donde presionaremos el botón “New”.



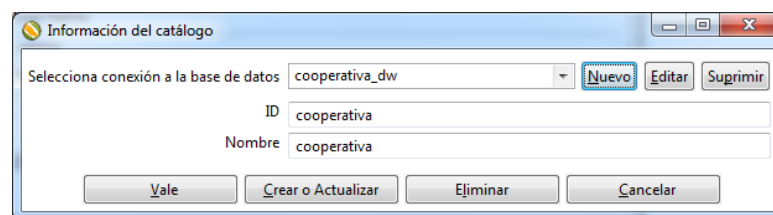
- Se nos presentará una nueva ventana, la cual llenamos con los datos encerrados en cuadrados. Luego de llenar los datos realizamos un test de la conexión con el botón Probar.



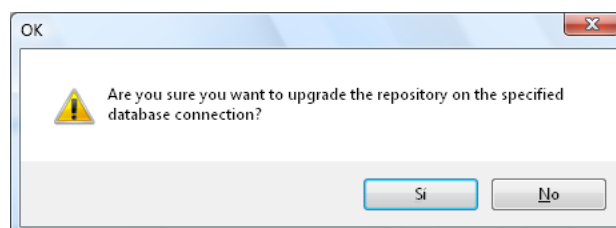
- Si la conexión está bien definida y la base de datos a la cual nos conectamos existe entonces luego de presionar el botón de Test debemos obtener:



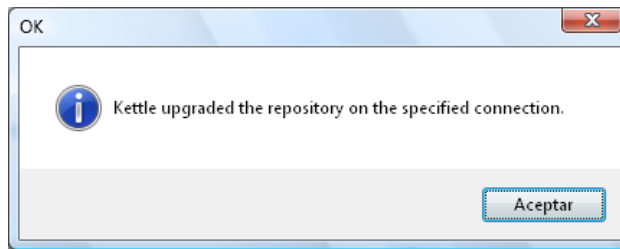
- Presionamos el botón de Aceptar y en la parte inferior de la ventana presionamos el botón de OK adicional. Se ha creado de esta manera la conexión.
- Se nos presenta nuevamente la ventana anterior y debemos presionar **“Create or Upgrade”** para de esa manera terminar de definir nuestro repositorio.



- Además en la parte anterior tenemos que especificar un ID y un Nombre para el catálogo
- Al presionar este botón, el PDI crea en nuestra base “cooperativa_dw” una tablas que el usara para poder almacenar la metadata de las transformaciones.



- Luego de presionar este botón debemos obtener la siguiente ventana de anuncio:

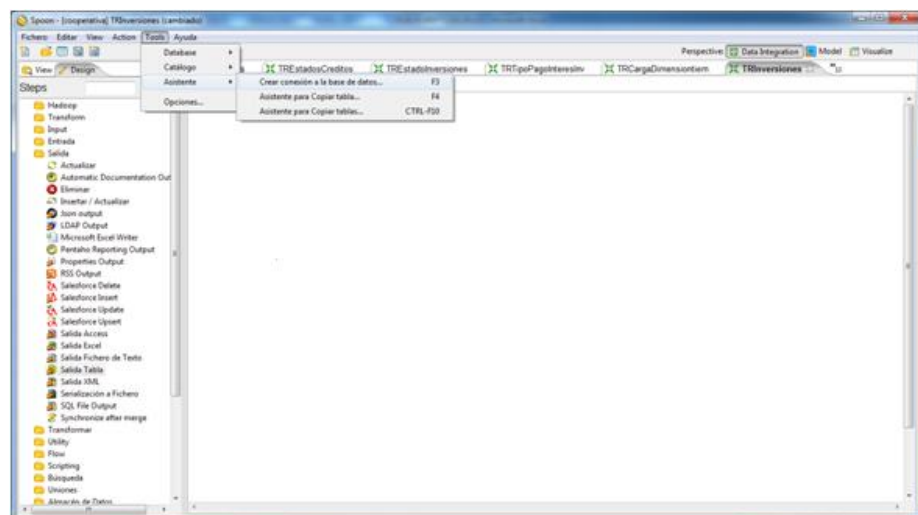


- Luego presionamos OK, se nos presentará la ventana inicial donde seleccionamos nuestro repositorio y nos logeamos con las credenciales:

Login: admin
Password: admin

• CREACIÓN DE LAS CONEXIONES A LOS ORIGENES DE DATOS Y AL DATA WAREHOUSE

- En el menú Tools->Asistente seleccionamos la opción “Crear conexión a la base de datos”



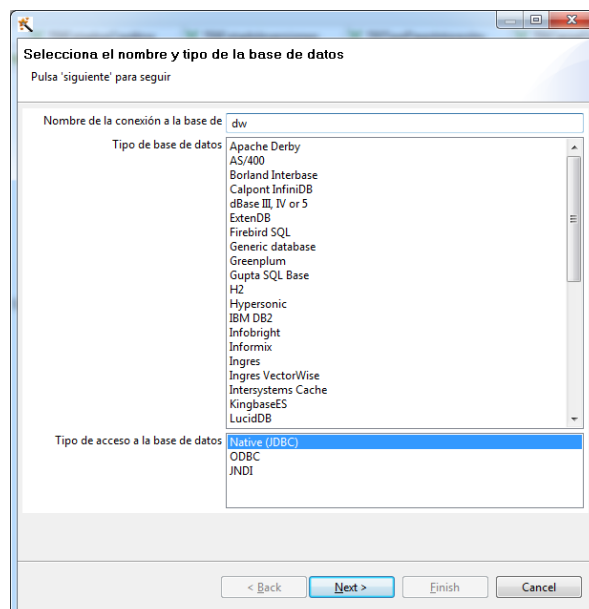
- Se presenta la siguiente ventana donde especificamos el nombre de la conexión a la base de datos, se selecciona el tipo de base de datos a la que nos queremos conectar y se especifica el tipo de acceso mediante el cual nos conectaremos a la base de datos.

Para este caso se colocó y escogió los siguientes datos.

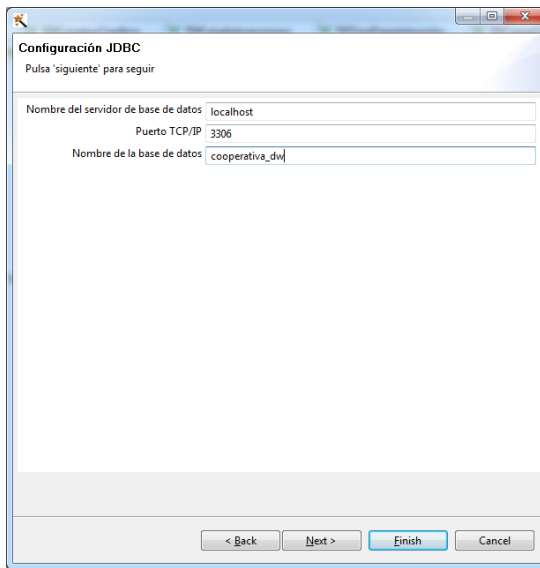
Nombre de la conexión a la base: dw

Tipo de base de datos: MYSQL

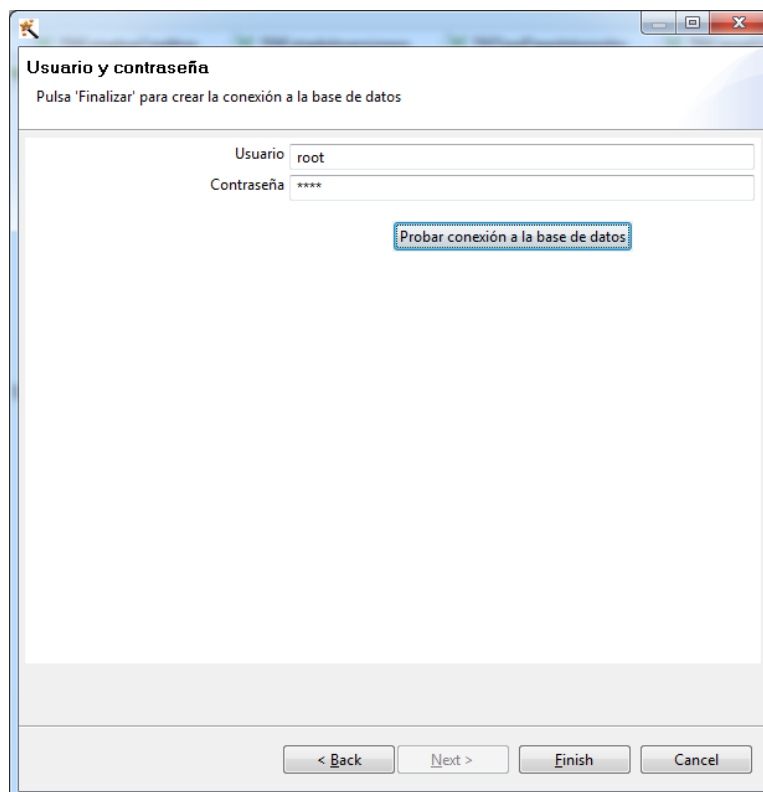
Tipo de acceso a la base de datos: Native (JDBC)



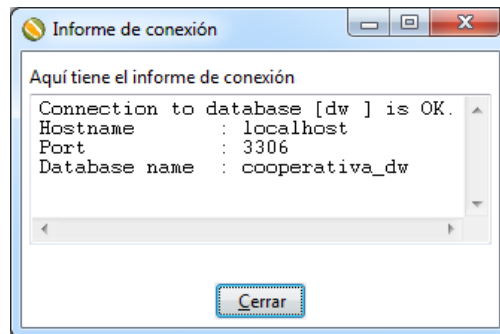
- Ingresar el nombre del servidor de la base de datos, el puerto de conexión y el nombre de la base de datos a la que nos queremos conectar.



- Ingresamos el usuario y contraseña para acceder a mysql



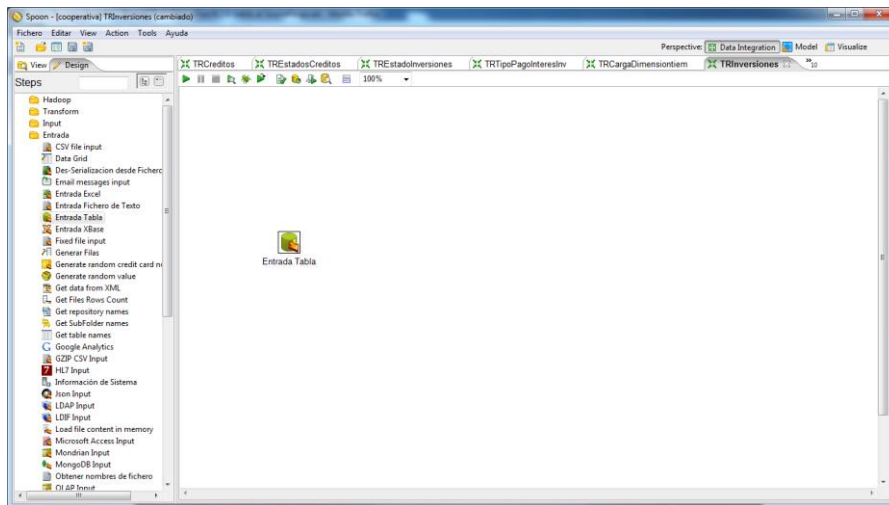
- Hacer clic en el botón “Probar conexión a la base de datos” y verificar que la conexión a la base de datos es correcta.



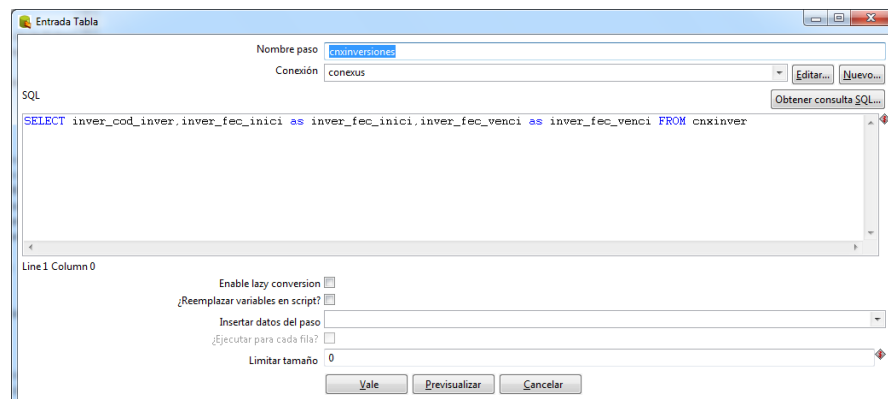
Clic en el botón “Cerrar” y hemos finalizado la creación de la conexión a la base de datos de destino.

- **EXTRACCIÓN DE LOS DATOS DESDE LA BASE DE DATOS INFORMIX**
 - Luego de terminar el paso anterior inmediatamente se nos abrirá la siguiente ventana.

Del costado del lado izquierdo seleccionamos “**Entrada Tabla**” y arrastramos a la parte derecha.

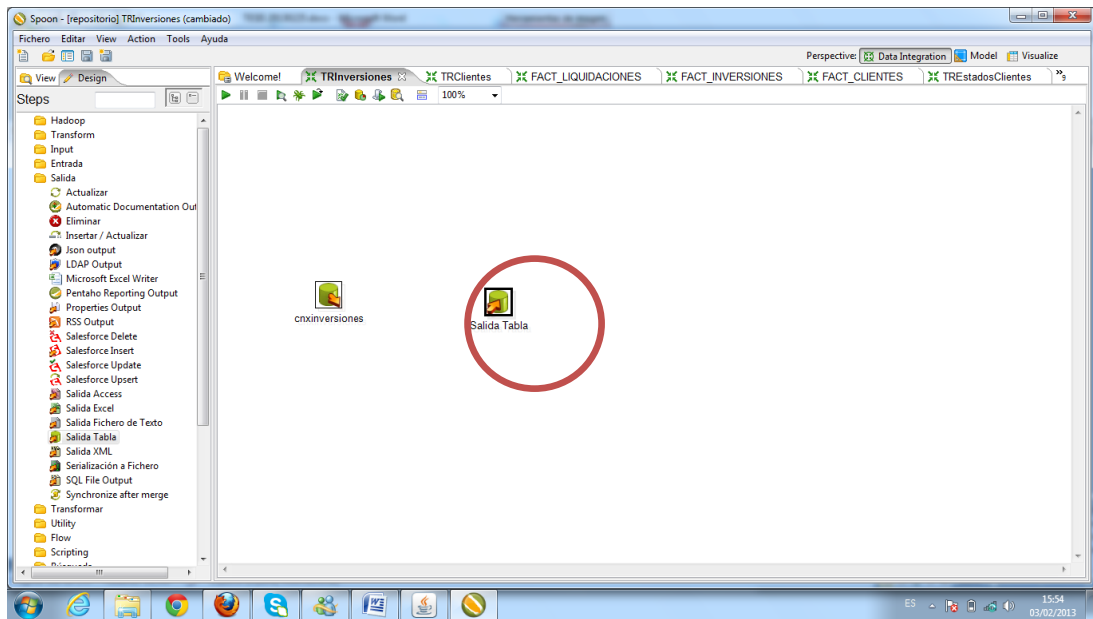


- Hacemos doble clic en el gráfico de **“Entrada Tabla”** y configuramos la ventana como se explica en el gráfico siguiente.

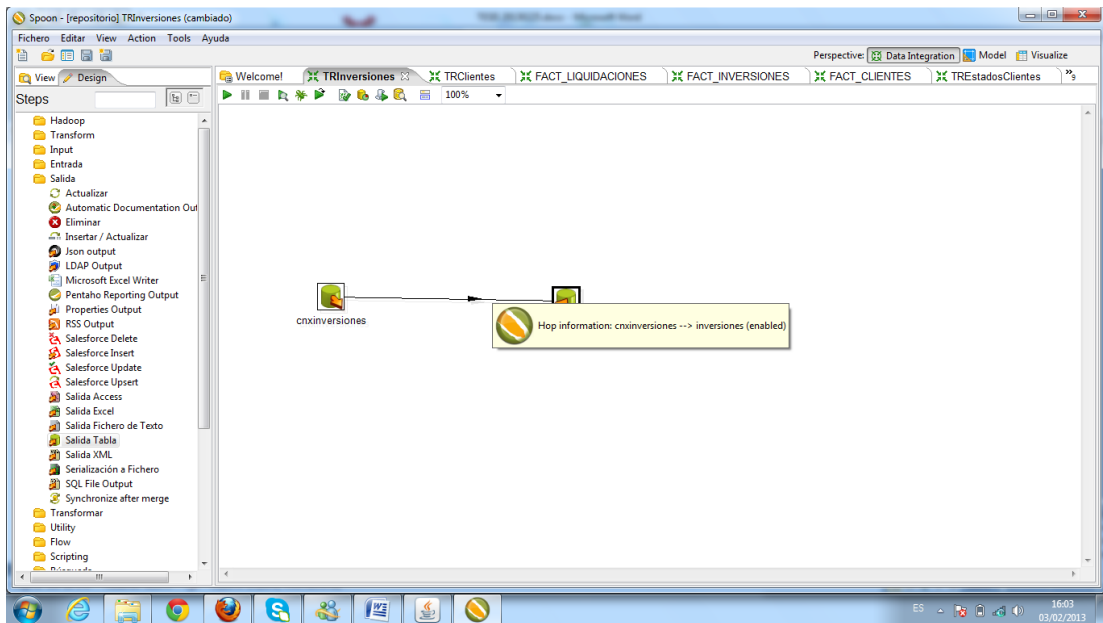


Colocamos un nombre para el paso, escogemos la conexión a la base de datos origen y escribimos la consulta SQL la cual obtendrá los datos que posteriormente se cargarán en la dimensión del data warehouse.

De la barra izquierda Steps abrimos el árbol Salida y arrastramos **“Salida Tabla”** a la zona de diseño:

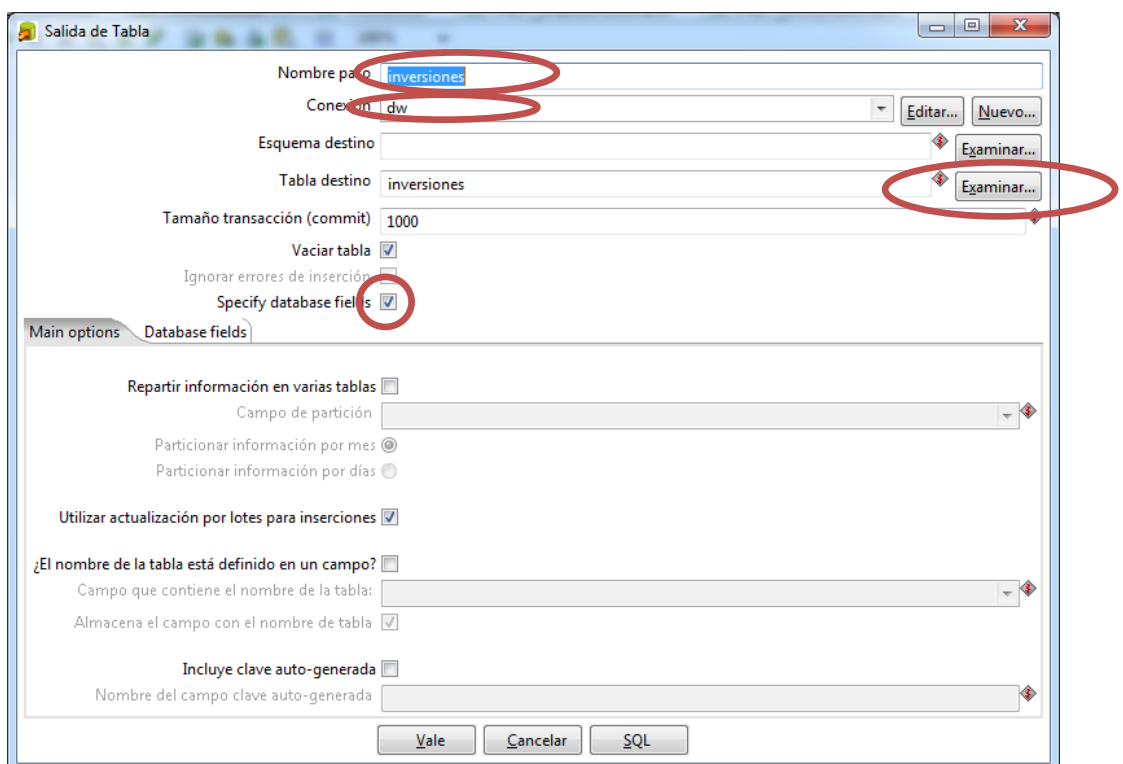


Realizar el Hop o salto que permitirá tener el enlace entre el origen y destino, para esto manteniendo presionada la tecla Shift arrastramos desde la Entrada de tabla llamada cnxinversiones hasta la tabla Salida Tabla. Se formará la flecha tal como se muestra en la ilustración inferior.

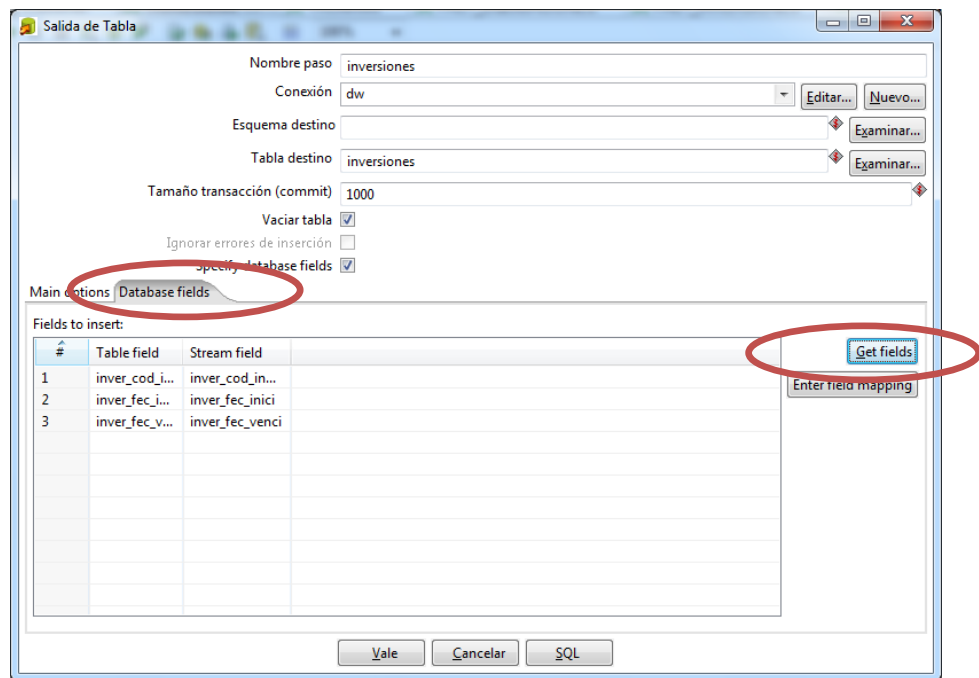


Ahora como siguiente paso hacemos doble clic sobre el elemento de Salida arrastrado anteriormente y realizamos las configuraciones como lo indica el gráfico de la parte inferior.

- En esta parte colocamos un Nombre paso que es el nombre con el que se distinguirá al paso.
- Escogemos la conexión del Data Warehouse.
- Escogemos la base de destino de donde deben colocarse los datos extraídos.
- Y marcamos el check que dice **“Specify database fields”**, para posteriormente hacer el mapeo de los campos de origen con los campos de destino del data warehouse.



En la misma pantalla de la configuración de la “Salida de tabla hacemos clic en la pestaña Database field y se mostrará la siguiente vista donde, haciendo clic en Get fields automáticamente se mapean los campos de origen y destino.



Con estos pasos ya tenemos realizado el mapeo de datos para la extracción desde la fuente de datos hasta el destino o data warehouse.

Luego de realizar los procesos de extracción para cada una de las tablas del repositorio de información procederemos a realizar el JOB o trabajo para que se carguen todas las dimensiones y tablas de hecho en el data warehouse.

En esta parte es importante indicar que las dimensiones son los primeros trabajos que se deben realizar la carga, finalizando siempre con las tablas de hechos. Esto nos evitará problemas en el trabajo de carga.

