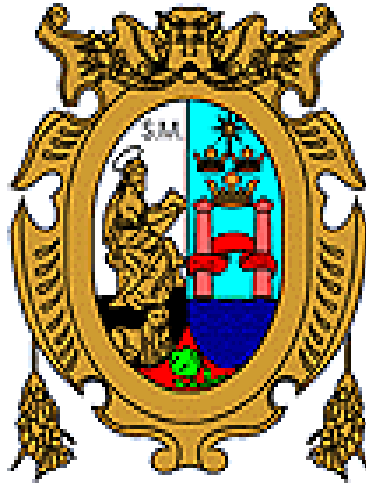


UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS

FACULTAD DE INGENIERIA DE SISTEMAS E INFORMATICA

E.A.P. DE INGENIERÍA DE SISTEMAS



**IMPLEMENTACION DE UNA HERRAMIENTA DE
INTELIGENCIA DE NEGOCIOS PARA LA
ADMINISTRACION DE JUSTICIA SOBRE UNA
METODOLOGIA AD-HOC**

**TESIS
para optar el Titulo Profesional de Ingeniero de
Sistemas**

AUTORES
Arturo Eguila Canales
Alex Alsey Parco Iquiapaza

Lima – Peru

2007

INDICE

RESUMEN

ABSTRACT

INDICE.....	1
INDICE DE FIGURAS.....	3
INDICE DE TABLAS.....	5
INTRODUCCIÓN.....	8

CAPITULO I

1. DEFINICIÓN DEL PROBLEMA.....	9
1.1. Descripción de la realidad.....	11
1.2. Antecedentes del problema	19
1.3. Justificación e importancia de la investigación.....	31
1.4. Delimitación del problema.....	34

CAPITULO II

2. OBJETIVOS

2.1. Objetivos generales.....	36
2.2. Objetivos específicos.....	37

CAPITULO III

3. MARCO TEÓRICO CONCEPTUAL

3.1. Antecedentes de la investigación	39
3.2. Bases teóricas.....	40
3.3. Definición de términos básicos.....	141

CAPITULO IV

4. METODOLOGÍA DE LA INVESTIGACIÓN

4.1. Metodología de Desarrollo.....	144
4.1.1. Planificación del Proyecto.....	147
4.1.2. Definición de Requerimientos de Negocios.....	149
4.1.3. Modelado Dimensional.....	150
4.1.4. Diseño Físico.....	168

4.1.5. Diseño y Desarrollo de Presentación de Datos.....	170
4.1.6. Diseño de La Arquitectura Técnica.....	190
4.1.7. Selección de Productos de Instalación.....	191
4.1.8. Especificación de Aplicaciones para usuarios Finales.....	191
4.1.9. Desarrollo de aplicaciones para usuarios Finales.....	192
4.1.10. Implementación.....	213
4.1.11. Gerenciamiento.....	214
CONCLUSIONES	215
RECOMENDACIONES.....	216
REFERENCIAS BIBLIOGRÁFICAS.....	217

INDICE DE FIGURAS

Fig. 1.1. Causas nuevas por año, Juzgados de Familia.....	13
Fig. 1.2. Acciones en las causas en seguimientos por año, Juzgados Familia..	15
Fig. 1.3. Procesos iniciados por año, Juzgados penales.....	16
Fig. 1.4. Acciones en procesos en seguimiento por año, Juzgados Penales.....	16
Fig. 1.5. Procesos iniciados por año, Salas Penales.....	17
Fig. 1.6. Acciones en procesos en seguimiento por año, Salas Penales.....	18
Fig. 3.1. Niveles de la Organización.....	46
Fig. 3.2. Clasificación de los Gerentes en las Organizaciones.....	49
Fig. 3.3. Data Warehouse Integrado.....	52
Fig. 3.4. Data Warehouse Temático.....	53
Fig. 3.5. Data Warehouse No Volátil.....	54
Fig. 3.6. Data Warehouse variante en el Tiempo.....	55
Fig. 3.7. Arquitectura real de un Data Warehouse.....	61
Fig. 3.8. Esquema Estrella.....	69
Fig. 3.9. Esquema Copo de Nieve.....	70
Fig. 3.10. Constelación de Hechos.....	70
Fig. 3.11. Ejemplo de un Cubo con tres dimensiones.....	71
Fig. 3.12. Dimensión con Niveles.....	72
Fig. 3.13. Slice & Dice.....	73
Fig. 3.14. Rotación.....	74
Fig. 3.15. Drill-Down y Drill-Up.....	75
Fig. 3.16. Roll-Up.....	76
Fig. 3.17. Drill-Across.....	76
Fig. 3.18. Drill-Trough.....	77
Fig 3.19. Ventas.....	81
Fig. 3.20. Identificación de Dimensión.....	82
Fig. 3.21. Datamart de Ventas.....	83
Fig. 3.22. Calificación de Productos.....	85
Fig. 3.23. ETL Datamart.....	86
Fig. 3.24. MOLAP.....	95

Fig. 3.25. ROLAP.....	96
Fig. 3.26. HOLAP.....	98
Fig. 3.27. Arquitectura Básica para Mining	107
Fig. 3.28. Etapas de Construcción de un Mining	110
Fig. 3.29. Fases de Metodología Kimball.....	113
Fig. 3.30. Iteración entre Planificación y Requerimientos.....	114
Fig. 3.31. Iteración entre Requerimientos del Negocio y etapas subsiguientes	117
Fig. 3.32. Diagramas para Modelado Dimensional.....	118
Fig. 3.33. Proceso de Diseño Físico a Alto Nivel.....	120
Fig. 3.34. Proceso ETL.....	121
Fig. 3.35. Modelo de Arquitectura Técnica a Alto Nivel.....	123
Fig. 3.36. Ejemplo de Matriz de Evaluación de Productos.....	125
Fig. 3.37. Variedad de Interfases por perfil.....	125
Fig. 3.38. Configuración de Metadata y construcción de reportes.....	129
Fig. 3.39. Implementación como convergencia de la Tecnología, los Datos y las Aplicaciones.....	132
Fig. 3.40 Data Warehousing como proceso de naturaleza.....	137
Fig. 3.41. Gerenciamiento del Proyecto.....	139
Fig. 4.1. Metodología Kimball.....	146

INDICE DE TABLAS

Tabla 1.1: Causas nuevas del año 2006.....	14
Tabla.3.1: Diferencias entre Sistema Tradicional versus Data Warehouse.....	58
Tabla 3.2: Modelo Dimensional Complejo.....	84
Tabla 3.3.: MOLAP, ROLAP, HOLAP.....	99

RESUMEN

En el presente trabajo se desarrolla el Análisis y Diseño para el desarrollo e implementación futura de una herramienta de Inteligencia de Negocios para la toma de decisiones en el Área de Defensoría de Oficio del Ministerio de Justicia, el propósito de la implementación de dicha herramienta es tener un mejor control y gestión de la información de el sistema de defensores de oficio de forma que, ayude a mejorar la calidad del servicio que presta dicha entidad, por la toma de decisiones eficientes conseguidas a partir de la información que este sistema proporcione a los directivos de dicha entidad.

Este trabajo muestra la metodología utilizada para el desarrollo del Datamart, el cual plantearemos cómo desarrollarlo y mostraremos mediante prototipos su funcionamiento.

Palabras Claves: Datamart, Minjus, Proyecto, OLAP, Transacción, Metodología, Defensoría.

ABSTRACT

The present work develops the Analysis and Design for the development and future implementation of a tool of Business Intelligence for the taking of decisions in the Area of Defensoría de Oficio of the Justice Ministry, the intention of the implementation of this tool is to have a better control and management of the information of the office defenders system so, it will help to improve the quality of the service that gives this organization, by the efficient decision taking from the information that this system provides to the managers of this organization.

This work shows the methodology used for the development of the Datamart, which we will show as to develop it and we will show by means of prototypes its operation.

Key words: Datamart, Minjus, Project, OLAP, Transaction, Methodology, Defensoría.

INTRODUCCIÓN

En la actualidad el Ministerio de Justicia cumple un papel importante en la defensa pública asistiendo en la asesoría y defensa legal gratuita.

Muchas veces el no contar con información actualizada y oportuna deriva en una toma de decisiones fuera de tiempo o en una decisión incorrecta, ante esta situación surge la necesidad de contar con una herramienta de Inteligencia de Negocios que ayude a la toma de decisiones basada en la información confiable y oportuna, de manera que al mostrarle a la Alta Dirección la información lo que ellos necesitan, puedan entenderla y usarla en forma eficiente y que esté actualizada al momento que ellos lo soliciten o requieran.

CAPITULO I

1. DEFINICIÓN DEL PROBLEMA

La Dirección de Defensoría de Oficio toma decisiones y planifican a corto, mediano y largo plazo, actualmente cuenta con un sistema transaccional, este sistema permite a los encargados de la institución controlar todo lo relativo al registro, asignación, seguimiento de casos, así como el control y supervisión en tiempo real, a nivel nacional, de los procesos, que abarca a los Defensores de Oficio, abogados de Consultorios Jurídicos Populares, Centro de Asistencia Legal Gratuita – ALEGRA y la implementación del Nuevo Código Procesal Penal, pudiendo en todo momento conocer los avances de los procesos a cargo de cada uno de ellos; así como de registrar información del Patrocinado a fin de agilizar sus procesos.

Pero la Dirección de Defensoría de Oficio necesita disponer de información tanto consolidada como detallada de cómo marchan las actividades ya cumplidas, para tomar decisiones proactivas que ayuden a una mejor administración del área de Defensoría. El Sistema de Defensores de Oficio proporciona reportes para encontrar las respuestas a algunas de las preguntas, pero se necesita una herramienta especializada por que el sistema de Defensores de Oficio

no fue construido con el fin de brindar síntesis, análisis, consolidación, búsquedas y proyecciones sino para ayudar a controlar sus procesos.

Por lo tanto el problema encontrado es la falta de información actualizada y consolidada de los indicadores que permita a esta área del Ministerio de Justicia una adecuada toma de decisiones para el control y seguimiento de los casos asignados a esta área.

Los análisis y requerimientos de informes acerca del manejo y seguimiento de los casos toman cierto tiempo para ser elaborados debido a que para el análisis de las diversas fuentes de información se debe invertir tanto en tiempo como en personal, el cual se dedicará a realizar estas funciones y de esta forma obtener los resultados requeridos.

Además, la Implementación de una herramienta de Inteligencia de Negocios eliminará la dependencia obligatoria de las áreas afectadas hacia el Área de sistemas, la cual en algunos casos se encuentra ocupada.

1.1 Descripción de la realidad:

La Dirección de Defensoría de Oficio y Servicios Jurídicos Populares, presta sus servicios a través de las sedes de cortes superiores, juzgados de paz letrados, establecimientos penitenciarios, centros de asistencia legal gratuitos-alegra y módulos básicos de justicia.

Esta institución fue creada para proporcionar defensa legal a los imputados o acusados por un crimen, simple delito o falta que carezcan de abogado, asegurando de esta manera el derecho a defensa por un defensor de oficio y el debido proceso en el juicio.

Actualmente La Dirección de Defensoría de Oficio no cuenta con una herramienta de Inteligencia Negocios (Datamart) y sólo cuentan con sistemas transaccionales con las cuales se llevan a cabo el control, supervisión y seguimiento de los casos.

La Dirección de Defensoría de Oficio y Servicios Jurídicos Populares es el órgano de línea de la Dirección Nacional de Justicia, encargado de normar, promover, ejecutar y supervisar el asesoramiento legal gratuito que prestan los Consultorios Jurídicos Populares a nivel nacional; promueve y coordina el Sistema de Defensoría de Oficio, garantizando el derecho de defensa, a las personas de escasos recursos económicos.

Existen diferentes puntos del país donde se presta el servicio de defensa pública. La prestación de los servicios se hace a través de abogados que forman parte de la institución, mediante el registro y seguimiento de casos llevados por los abogados. El Defensor debe registrar todas las actividades relacionadas con su patrocinado haciendo uso de un Sistema de Seguimiento de casos distribuido, esta es una solución eficiente y flexible, que aumenta la eficiencia en el control de los procesos llevados por los defensores, al tiempo que reduce los tiempos de respuesta y actualización de la información.

Los análisis y requerimientos de informes acerca del seguimiento de casos y control del Área de Defensores de Oficio toman cierto tiempo en ser elaborados ya que para el análisis de las diversas fuentes de información se debe invertir en tiempo y personal que se dedique a estas funciones para obtener los resultados requeridos.

Al mes de diciembre del año pasado (2006), se contó con 483 Defensores de Oficio a nivel nacional, de los cuales 141 asumieron los procesos de salas penales, 125 en los juzgados penales, 70 en los juzgados de familia, 65 en los juzgados mixtos, 24 en las áreas policiales, 46 en los establecimientos penitenciarios y 12 brindaron atención en los Centros de Asistencia Legal Gratuita - ALEGRA.

Durante el año 2006, los Defensores de Oficio asignados en los juzgados de familia y ALEGRA asumieron 28 155 causas nuevas, cifra que muestra un incremento de 3 476 (14,08 %) con relación al año anterior que registró 24 679, correspondiendo 22 427 (79,66 %) a procesos civiles y 5 728 (20,34 %) a infracciones penales.

Estos incrementos de nuevos casos requieren un número mayor de personal de la Dirección de Defensoría de oficio.



Figura 1.1: Causas nuevas por año, juzgados de Familia

Causas nuevas, asumidas por los Defensores de Oficio asignados en los juzgados de familia y ALEGRA, según Distrito Judicial, 2006

Distrito Judicial	Causas nuevas							
	Total		Sedes Judiciales				ALEGRA	
	Procesos civiles	Infracciones penales	Juzgados de familia		Juzgados de familia (MBJ)		Procesos civiles	Infracciones penales
			Procesos civiles	Infracciones penales	Procesos civiles	Infracciones penales		
Total	22 427	5 728	17 726	5 078	4 076	526	625	124
Amazonas	304	24	240	17	64	7	-	-
Ancash	1 336	55	1 162	48	133	4	41	3
Apurímac	377	30	316	26	61	4	-	-
Arequipa	1 626	320	838	242	754	76	34	2
Ayacucho	302	144	207	54	10	11	85	79
Cajamarca	719	19	591	18	128	1	-	-
Callao	750	679	650	664	27	14	73	1
Cañete	683	102	683	102	-	-	-	-
Cono Norte	1 828	731	1 464	665	269	64	95	2
Cusco	869	254	692	215	167	32	10	7
Huancavelica	697	31	697	31	-	-	-	-
Huanuco	596	73	408	63	188	10	-	-
Huaura	322	60	322	60	-	-	-	-
Ica	536	158	358	134	160	23	18	1
Junín	1 008	192	813	161	195	31	-	-
La Libertad	479	81	345	76	133	5	1	-
Lambayeque	732	95	634	90	98	5	-	-
Lima	3 666	2 077	2 600	1 884	862	178	204	15
Loreto	737	50	659	43	78	7	-	-
Madre de Dios	64	14	-	-	-	-	64	14
Moquegua	226	18	226	18	-	-	-	-
Pasco	157	19	157	19	-	-	-	-
Piura	1 430	135	1 000	101	430	34	-	-
Puno	837	124	677	115	160	9	-	-
San Martín	489	67	489	67	-	-	-	-
Santa	709	32	608	30	101	2	-	-
Tacna	421	63	364	56	57	7	-	-
Tumbes	289	36	289	36	-	-	-	-
Ucayali	238	45	237	43	1	2	-	-

Fuente : Dirección Nacional de Justicia

Tabla 1.1: Causas nuevas del año 2006

En el 2006, los Defensores de Oficio asignados en los juzgados de familia, realizaron 39 230 acciones relacionadas con las causas que se encuentran en seguimiento, cifra superior en 7 780 (24,74 %) con relación al año anterior que registró 31 450, correspondiendo 35 528 (90,56 %) a procesos civiles y 3 702 (9,44 %) a infracciones penales.



Figura 1.2: Acciones en las causas en seguimientos por año, Juzgados Familia

Procesos Iniciados

Durante el año 2006, los Defensores de Oficio asignados en los juzgados penales iniciaron un total de 64 626 procesos a nivel nacional, cifra inferior en 348 (0,54 %) respecto al año anterior que registró 64 974 procesos.

En la Actualidad no se pueden tomar decisiones acertadas en cuanto a la cantidad de personal adecuado para el Área de Defensores de oficio por que no se tiene muchas referencias en cuanto a la exactitud del volumen de trabajo ni de la calidad de los servicios prestado por los Defensores de Oficio.



Figura 1.3: Procesos iniciados por año, Juzgados penales

Acciones en los Procesos en Seguimiento

Asimismo, los Defensores de Oficio realizaron 34 749 acciones en los procesos que se encuentran en seguimiento, cifra superior en 6 201 (21,72 %) con relación a lo registrado el año anterior que alcanzó 28 548.



Figura 1.4: Acciones en los procesos en seguimiento por año, Juzgados

Penales

Durante el año 2006, los Defensores de Oficio asignados en salas penales iniciaron 21 887 procesos a nivel nacional, cifra superior en 2 247 (11,44 %) con relación a los 19 640 registrados en el año 2005 .



Figura 1.5: Procesos iniciados por año, Salas Penales

Durante el año 2006, los Defensores de Oficio asignados en salas penales realizaron 61 709 acciones en los procesos que se encuentran en seguimiento, cifra que muestra un incremento de 3 568 (6,14 %) con relación a las 58 141 registradas en el año anterior. Actualmente el Ministerio designa una cantidad de Defensores de Oficio para llevar los casos.

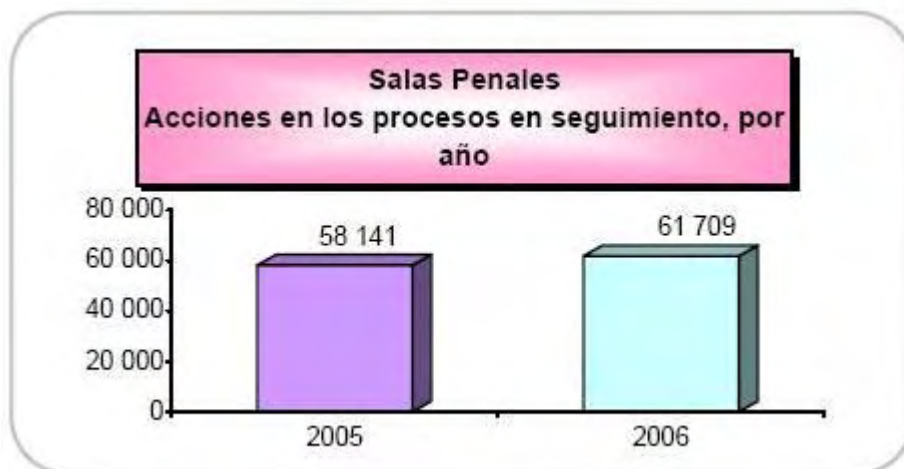


Figura 1.6: Acciones en procesos en seguimiento por año, Salas Penales

El Sistema actual permite realizar las siguientes tareas a los defensores:

- El Defensor puede acceder al Sistema desde cualquier punto del país utilizando una conexión Internet.
- El Defensor puede acceder a los datos históricos en cualquier momento, simplemente ingresando con su propio usuario y clave (Personal autorizado por al Área).
- Estado del caso en cualquier momento, pudiendo revisar acciones del pasado, y acciones actuales.
- Permite realizar un seguimiento completo, desde el momento en que se está procesando la investigación del inculpado, hasta el fin del proceso en sí .

- Es posible actualizar en cualquier momento los datos del patrocinado, ya sea modificando la información existente o adicionando nueva información .
- En caso de que un Defensor tome un caso existente, tenga la historia previa con su respectivo detalle.
- Se puede tener la información detallada de cada Defensor y de manera actualizada.

1.2 Antecedentes del problema:

La mayoría de empresas del Perú y del extranjero ya cuentan de una u otra manera con diferentes herramientas de Inteligencia de Negocios como un Datamart ya que las empresas actualmente tienen la necesidad constante de consumir información para poder sobrevivir.

Sin embargo muchos de estos Datamarts fueron creados enfocados en los datos y no en las necesidades de información de los usuarios, por lo que algunas empresas aceptan el reto de implementar un Datamart y muchas de ellas lo implementaron con éxito, por ejemplo:

MUNDI HOGAR DE MEXICO

Mundihogar es una empresa comercializadora líder en el ramo mueblero. Fundada en 1951, actualmente cuenta con 22 tiendas, distribuidas en el área metropolitana de Guadalajara, Aguascalientes, San Francisco del Rincón, Tepatlán, León, Mazatlán y Tijuana.

El Reto

En el año de 2002 se detectó la necesidad de contar con un Sistema de Información que le ofreciera, tanto a la dirección como a los diferentes responsables de las áreas un ambiente gráfico y de fácil manejo con capacidad de análisis dinámico para servir de apoyo en la toma de decisiones. Las áreas involucradas fueron Crédito, Cobranza, Comercialización, Tesorería, Nóminas y Contabilidad.

Este sistema de información ejecutivo se diseñaría en base a las necesidades de información de la Dirección y las Gerencias y se apoyaría en un Datawarehouse para almacenar información actual e histórica. Todo esto manteniendo el equilibrio entre Costo-Beneficio.

La Propuesta.

Al iniciar el proyecto de Inteligencia de Negocios de Mundihogar el primer Módulo a desarrollar fue Crédito y Cobranza donde se incluyó información de la cartera, la cobranza y la tendencia de cobranza lo que le permitió a Mundihogar analizar el comportamiento pasado de sus cuentas por cobrar así como estimar la recuperación probable en base a la tendencia de cobranza del mes en curso. El siguiente Modelo que se desarrolló fue la parte Contable en donde se consolidó la información de las diferentes empresas que conforman el grupo (tanto de comercialización como de servicios), para poder obtener sus reportes Financieros consolidados, dentro de estos reportes Financieros–Contables se incluyeron: Estado de Resultados, Balance General, Estado de Origen y Aplicación y Gastos tanto Consolidado como por Sucursal.

El siguiente Módulo fue el de Comercialización que fue diseñado para obtener todos los indicadores tanto del área de Ventas, de Compras como del área de Inventarios. Este modelo permite a la dirección analizar y manipular la información de ventas desde el punto de vista general de la empresa, y con la capacidad de profundizar (hacer “drill down”) por Sucursal, Línea de Producto, Vendedor e incluso llegar al detalle del producto mismo que se vendió permitiendo

identificar oportunidades de negocio, así como visualizar la tendencia de ventas y el estado que guardan las ventas contra periodos anteriores. Esto mismo aplica a la información de Compras y de Inventarios.

Posteriormente se elaboraron los modelos de Tesorería y Recursos Humanos que permiten identificar y conocer los recursos tanto económicos como humanos con que cuenta la empresa.

El sistema de información fue creado utilizando MS-SQL como Warehouse de datos y Artus como herramienta OLAP (Análisis dinámico en línea).

El Cambio.

El gran avance registrado por Mundihogar ha hecho que la dirección general de la compañía cuente con la información clara, oportuna, confiable y automática para la toma de decisiones, “es la eficiencia total de la empresa”, como comentó el Director General. Se han reducido sustancialmente los recursos destinados a elaborar los reportes en hojas de cálculo que se presentaban a la Dirección, destinando este ahorro a mejorar las áreas de influencia, así como adelantarse oportunamente a los hechos, tomando las decisiones adecuadas y oportunas.

Actualmente los gerentes de área realizan consultas y fundamentan sus decisiones en base a la información reportada por el modelo de Inteligencia de Negocios de Mundihogar. Dedicando más tiempo al análisis de información y a la implementación de planes de acción.

Resultados.

El sistema de Inteligencia de Negocios (BI por sus siglas en inglés) de Mundihogar ofrece a sus directivos información confiable y oportuna en un ambiente gráfico, fácil de usar y con capacidad de análisis dinámico; esto ha permitido a los gerentes y directores de Mundihogar dedicar más su tiempo a analizar información y tomar decisiones.

Hasta hace unos meses Mundihogar manejaba en sus bodegas un inventario con un índice de 45 días, lo cual según el Director General ya era sano financieramente hablando, hoy en día, gracias al sistema de Inteligencia de Negocios desarrollado para Mundihogar, el índice de su inventario es de sólo 15 días; esto ha generado grandes beneficios financieros a la empresa lo cual ha representado con este sólo hecho un retorno de la inversión en este proyecto bastante alto y en un plazo menor a 1 año.

Los objetivos principales del proyecto de Inteligencia de Negocios en Mundihogar que se plantearon inicialmente era tener acceso oportuno a información confiable sobre las áreas de mayor impacto sobre el

negocio así como mantener una historia del desempeño pasado para poder analizarlo y compararlo con el actual.

El resultado obtenido en un periodo de 6 meses es un Panel de control con los principales indicadores de la empresa y 6 modelos de información que se actualizan diariamente, de manera automática y que son la base del proceso de gestión del negocio y análisis de información para tomar decisiones.

CPINGREDIENTE OBTIENE EL PODER DE LA INFORMACIÓN

CPIngrediente, Empresa Mexicana (antes Arancia Corn Products), es el líder nacional en el ramo de la manufactura y comercialización de productos derivados del maíz, cuentan con 4 plantas en México y con 1200 empleados aproximadamente, CPIngredientes es subsidiaria de Corn Products Internacional, Inc.

Corn Products International, Inc. es una compañía líder a nivel mundial, tiene presencia en 22 países con 41 plantas, es la única compañía en el negocio de refinación de maíz con operaciones de manufactura y distribución en los 3 países que conforman el Tratado de Libre Comercio de Norteamérica.

El Reto.

En el año de 1998 la Empresa con el apoyo de el Departamento de Sistemas implementó un sistema de operación ERP “Enterprise Resource Planning”, en este se lleva la operación de toda la empresa, pero surgió la necesidad de contar con una herramienta alterna, el sistema transaccional satisfacía la operación pero no así la extracción y la explotación de la información, por lo que se comenzó a buscar alternativas con las cuales se pudiera crear un Data Warehouse. La opción elegida tendría que ser de manejo sencillo y de precio razonable, para lograr un parámetro alto de costo-beneficio.

La Propuesta.

Cuando se inicia el proyecto de BI en CPIIngredientes se toma la decisión de arrancar con el departamento de ventas teniendo como principal objetivo el conocer su mercado, es decir, hacia donde iba su mercancía, cómo estaban vendiendo, cuánto estaban vendiendo y dónde estaba el potencial de crecimiento de la empresa.

Conforme avanzó el proyecto este fue creciendo hacia otras áreas por lo que se creó un Datamart para el departamento de Logística (Transportes).

Ambos Datamarts fueron creados sobre una plataforma híbrida UNIX-NT, utilizando Red Brick como warehouse Server y Metacube como herramienta OLAP para consultas Ad-hoc al modelo construido.

El Cambio.

Dados los nuevos requerimientos de expandir la tecnología de BI a otros modelos dentro de las áreas de la Empresa y el alto costo de la solución con la que se contaba en el momento, el equipo de Sistemas de CPIngredientes calificó las diferentes opciones que había en el mercado para la creación del Data Warehouse.

Se desarrollaron los Datamarts para el departamento de Compras y para el área de Costos. Actualmente se tiene el objetivo a corto plazo de crear uno para el Área de Finanzas y la redefinición del Datamart de Transportes, “El objetivo es que en la información donde el dinero juegue un papel importante (mercados, costos, inventarios, gastos, clientes, etc.) se tenga una herramienta con la cuál se pueda evaluar el desarrollo y crecimiento de la empresa y no sólo los datos arrojados por la operación” (Ing. José Luis Vázquez Blasnich, Jefe de Soporte Operativo y Desarrollo).

Resultados.

En el área de compras han logrado unificar toda la información que se generaba por separado como el uso de reportes del ERP, la

extracción de listados, la captura manual en hojas de Excel, etc. Toda esta información ahora se encuentra concentrada en el Datamart, lo que le permite a la gerencia de este departamento evaluar mes con mes los diferentes parámetros como precios, volúmenes comprados, historiales de precios y compras con algún determinado proveedor, con lo cuál obtienen herramientas para realizar mejores negocios.

Otro de los beneficios que se tuvo es que los sistemas ERP al ser transaccionales, están orientados a un período de tiempo, en contraste con el Datamart que crea un histórico con lo cual se tiene un panorama más amplio de la situación de la empresa en un área determinada y no sólo una foto del momento en curso.

Un objetivo primordial que se estableció cuando comenzó el proyecto de Business Intelligence en CPIngredientes era cambiar el modelo de negocios, es decir darle la responsabilidad de la toma de decisiones a quien debe tenerla realmente, por lo tanto, quién tuviera esa responsabilidad debería de tener la información como principal herramienta.

Proyectos Futuros.

En cuanto a proyectos futuros se ha planteado la creación del Datamart Financiero el cual será un proyecto que involucrará a mucha gente y a muchas áreas de la Empresa, por lo que éste es hasta ahora

el proyecto más ambicioso de Business Intelligence desarrollado por CPIngredientes. Este proyecto juega un papel protagónico al prestar los servicios de soporte y consultoría al personal de CPIngredientes.

Empresa AVON

Aquí se llevaron a cabo los siguientes procesos: Análisis, diseño y Desarrollo de un Datamart Comercial para la explotación de información de campañas comerciales por las áreas de Marketing y Ventas.

Implementación de herramienta E.T.L (Extracción, Transformación y carga). Herramientas BI, desarrollo del proceso de migración de data del AS 400 a Oracle 9i. (Data Clearing). Extracción, Transformación.

El Proyecto de migración de Data llevado acabo el año 2005 Servicio de Tunning para la optimización de BD mejoró la calidad de la información obtenida ofreciendo información consolidada y a tiempo para el personal de la empresa.

La implementación del Datamart actualmente les permite ahorrar en su presupuesto y obtener información consolidada para la toma de decisiones, reducir los tiempos de respuesta de información y obtener reportes actuales de su información más importante.

BANCO DE CRÉDITO DEL PERÚ

Se implementó Datamart en todas sus fases como: Análisis, Diseño, Desarrollo e implementación de una aplicación de Business Intelligence para el área de marketing de este banco Peruano:

La implementación del Datamart les permite:

- ✓ Realizar un correcto seguimiento de la rentabilidad por tipo de transacciones que se realizan en cada canal de atención: ATM, POS, Ventanilla, Plataforma, etc. Esta aplicación permite a la alta dirección tomar decisiones como rentabilizar, clausurar o mantener un canal operativo
- ✓ Realizar un correcto seguimiento del tiempo de atención en los distintos canales de atención y mejorar la calidad del servicio al cliente reduciendo el tiempo de atención por cada tipo de cliente.
- ✓ Realizar un correcto seguimiento de las transacciones que se realizan en los distintos canales de atención.

Laboratorios Hersil

En este Laboratorio se llevo a cabo la Implementación de la Herramienta E.T.L. BI Tool para la automatización de los procesos de carga de la información de ventas proveniente de Farmacias, de Prescripciones Médicas (información comprada en el mercado) para ser entregada a las Gerencias de marketing y Comercial, Este implementación del Datamart permitió que la información llegue en forma oportuna y eficiente. Reduciendo los tiempos de procesamiento de esta información de semanas a horas, lo cual traducido en cuestiones de dinero les permitió un ahorro en su presupuesto asignado en todos sus proyectos.

1.3 Justificación e Importancia de la investigación.

1.3.1. Justificación de la Investigación.

La Dirección de Defensoría de Oficio necesita disponer de información tanto consolidada como detallada de cómo marchan las actividades ya cumplidas, para tomar decisiones proactivas que ayuden a una mejor administración del área de defensoría. por lo tanto se necesita una herramienta especializada con el fin de brindar síntesis, análisis, consolidación, búsquedas y proyecciones.

Durante el año 2006, los Defensores de Oficio asignados en los juzgados de familia y ALEGRA asumieron 28 155 causas nuevas, cifra que muestra un incremento de 3 476 (14,08 %) con relación al año anterior que registró 28 155, correspondiendo 22 427 (79,66 %) a procesos civiles y 5 728 (20,34 %) a infracciones penales. Debido a este constante aumento año a año por lo tanto es necesario una herramienta de Inteligencia de negocio en este caso Datamart para poder predecir el número de defensores necesarios por Defensoría para los siguientes periodos y los volúmenes de trabajo que serán asignados a ellos y de ese modo estar dentro de los límites de presupuesto asignado al Ministerio, además poder medir la eficiencia de los defensores para mejorar la calidad de sus servicios prestados y tener información actualizada de los casos iniciales y un seguimiento de los mismos llevados por ellos.

De este modo explotar toda la información y conservar la historia transaccional de la institución para llevar a cabo un análisis de las variables en el tiempo y proceder a la toma de decisiones.

Actualmente en la Defensoría de Oficio existen los siguientes problemas:

- ❖ Se toma más del tiempo previsto recolectando y preparando la información para el área de defensores de oficio necesaria para concluir los casos de forma satisfactoria, y una vez implementado el Datamart para esta área permitirá analizar la información. requerida de forma correcta y a tiempo.
- ❖ En ocasiones se nos es difícil encontrar información que existe dentro de la Institución, debido a que para ello participan diversas áreas dentro de la institución.
- ❖ Los reportes de información necesaria toman cierto tiempo en ser elaborados ya que para el análisis de las diversas fuentes de información se debe invertir en tiempo y personal que se dedique a estas funciones para obtener los resultados requeridos
- ❖ Se trabaja horas extras el fin de mes para procesar documentos o reportes de los casos iniciales, concluidos o .en proceso, lo cual requiere energía extra por parte del personal, mas tiempo y mas gasto

de dinero para la institución, lo cual influye negativamente en el presupuesto asignado

- ❖ No se sabe con certeza si los defensores están alcanzando los objetivos planeados, debido a que muchas veces no se cuenta con mucha información del seguimiento de los casos llevados por ellos.

Finalidad e importancia de la investigación.-

Finalidad

La finalidad del presente trabajo de investigación es proponer una herramienta de Inteligencia de negocios que contribuya a la mejorar la administración del área de defensoría, lo cual nos favorecerá a todos los peruanos para mejorar la calidad del servicio prestado.

Importancia

El trabajo es importante en primer lugar, porque contribuirá a orientar a la Dirección de Defensoría de Oficio y contará con una herramienta que juegue el papel de soporte para la toma de decisiones, de respuesta ágil y rápida, con información precisa.

La información consolidada permitirá realizar un mejor control con la aplicación de tecnologías de información, en este caso, el uso de Datamart, esto contribuirá en beneficio de todos los peruanos con lo cual

permitirá disponer de herramientas de gestión y a su vez de control para el Sector de justicia.

Limitaciones de la investigación

1.4.1. Delimitación Temporal

La parte descriptiva de la investigación se realizará en el período comprendido entre los meses de Julio y Agosto del 2007. En este periodo se implementará los prototipos de la herramienta de Inteligencia de Negocios sobre una metodología Ad Hoc debido al periodo de tiempo corto para su respectivo desarrollo.

1.4.2. Delimitación Espacial

El ámbito físico geográfico es el Perú específicamente Lima, ya que se aplica al sistema de Justicia Nacional (Ministerio de Justicia).

1.4.3. Delimitación Conceptual

Está delimitado por el área de Inteligencia de Negocios, sus herramientas respectivas, la Administración Estratégica de la Dirección del Ministerio de Justicia y estrategias administrativas .

1.4.4. Delimitación Social

Esta investigación circunscribe su estudio al personal directivo de la Dirección Nacional de Justicia, coordinadores y administrativo, instituciones públicas y organizaciones sociales.

CAPITULO II

2.1. OBJETIVOS

Se busca la implementación futura de una herramienta de Inteligencia de Negocios para la Administración del área de Defensoría de Oficio, que permita determinar los factores estratégicos que la afectan.

2.1.1. Objetivos generales:

Proporcionar una herramienta que contribuya a mejorar el control y la gestión de la Información que el sistema de defensores de oficio proporciona a las diferentes sedes judiciales del país, se desarrollará una herramienta que ayude a la toma de decisiones por parte de la Dirección Nacional de Justicia.

Explotar la información existente mediante reportes que nos muestre el estado actual de los servicios prestados por el Área de la Dirección de Defensoría de Oficio, de este modo se mejorará el control de los procesos llevados por cada defensor.

2.1.2. Objetivos Específicos:

Los Objetivos Específicos de este Proyecto de Investigación son:

- ✓ Proveer el diseño de una herramienta integral que ayude a controlar de manera efectiva la asignación y seguimiento de todos los casos registrados como nuevos, en seguimiento y concluidos por los defensores de tal manera de contar con estadísticas y métricas que permitan observar los volúmenes de trabajo de los defensores, así como tomar acciones correctivas para garantizar la calidad de los servicios prestados, optimizando el uso de los recursos de manera progresiva.
- ✓ Proporcionar información detallada de los datos personales, ubicación y asignación exacta de cada uno de los defensores de oficio.
- ✓ Generar consultas, reportes, estadísticas y métricas que permita evaluar, en todo momento, los niveles de calidad de trabajo del Defensor y la estrategia en relación al trabajo desarrollado por cada uno de ellos.
- ✓ Proveer a los niveles directivos del Ministerio de Justicia, un mecanismo capaz de otorgar información acerca del avance global de la gestión de los Defensores para cada una de las áreas de

interés, pudiendo establecer métricas comparativas entre el avance real de todas las actividades desarrolladas bajo un determinado periodo de tiempo.

CAPITULO III

3. MARCO TEÓRICO CONCEPTUAL

3.1. Antecedentes de la investigación:

Como resultado de búsquedas realizadas se han encontrado conceptos sobre la presente investigación que nos servirán para la elaboración de nuestro proyecto y que aparecerán en un marco conceptual.

En la mayoría de empresas del Perú y del extranjero se puede apreciar que muchas ya cuentan de una u otra manera con diferentes Datamart, lo cual está muy bien, ya que las empresas de hoy se han vuelto "Inforveras", es decir, tienen la necesidad constante de consumir información para poder sobrevivir.

Sin embargo muchos de estos Datamart fueron creados enfocados en los datos y no en las necesidades de información de los usuarios.

En la actualidad hemos encontrado pocas implementaciones en cuanto a desarrollos de trabajos sobre Datamart, Datamining en el ámbito de Ministerios del estado, específicamente, en el Ministerio de Justicia el cual no cuenta con un Datamart hasta la actualidad. Asimismo, con relación a las variables del tema, no hemos encontrado investigaciones que hayan abordado este tema aplicados a esta problemática por lo que optamos por la implementación para un área del Ministerio de Justicia.

3.2. Fundamentación o Bases Teóricas:

Algo peor que no tener información disponible es tener mucha información y no saber qué hacer con ella. La Inteligencia de Negocios o **Business Intelligence** (BI) es la solución a ese problema, pues por medio de dicha información puede generar escenarios, pronósticos y reportes que apoyen a la toma de decisiones, lo que se traduce en una ventaja competitiva. La clave para BI es la información y uno de sus mayores beneficios es la posibilidad de utilizarla en la toma de decisiones. En la actualidad hay una gran variedad de software de BI con aplicaciones similares que pueden ser utilizados en las diferentes áreas de la empresa, tales como, ventas, marketing, finanzas, etc. Son muchas las empresas que se han beneficiado por la implementación de una sistema de BI, además se pronostica que con el tiempo se convertirá en una necesidad de toda empresa.

3.2.1. Business Intelligence

Business Intelligence es un término creado por Gartner Group en 1993 con el concepto de Datawarehouse. Inteligencia de Negocios (Business Intelligence BI) es el proceso de identificar, recolectar, entregar y analizar los datos para transformarlos en conocimiento y acciones poderosas de negocio.

Es una amplia categoría de soluciones y software para recolectar, consolidar, analizar y proveer acceso a los datos en una forma que permita a los usuarios de la empresa tomar mejores decisiones de negocios.

BI apoya a los que toman decisiones con la información correcta, en el momento y lugar correcto, lo que les permite tomar mejores decisiones de negocios. La información adecuada en el lugar y momento adecuado incrementa la efectividad de cualquier empresa.

“La tecnología de BI no es nueva, ha estado presente de varias formas por lo menos en los últimos 20 años, comenzando por generadores de reportes y sistemas de información ejecutiva en los 80's ” Candice Goodwin [1]. Entiéndase como sinónimos de tecnología de BI los términos aplicaciones, soluciones o software de inteligencia de negocios.

¿Qué puede hacer Business Intelligence? [2]

Con BI se puede:

- ✓ Generar reportes globales o por secciones.
- ✓ Crear una base de datos de clientes.
- ✓ Crear escenarios con respecto a una decisión hacer pronósticos de ventas y devoluciones.

- ✓ Compartir información entre áreas o departamentos de una organización.
- ✓ Análisis multidimensionales.
- ✓ Generar y procesar datos.
- ✓ Cambiar la estructura de toma de decisiones.
- ✓ Mejorar el servicio al cliente.

Según Kobana Abukari y Vigía Job [3], “BI es una de las iniciativas administrativas más robustas que los administradores inteligentes pueden emplear para ayudar a sus organizaciones a crear más valor para los accionistas”.

BI ha tenido mucho éxito ya que le da una ventaja a las empresas sobre sus competidores al juntar a las personas y a la tecnología para resolver problemas.

¿Quién necesita soluciones de Business Intelligence?

Si usted puede contestar afirmativamente por lo menos a una de las siguientes preguntas, entonces usted es candidato a beneficiarse de las soluciones de BI.

- a) ¿Pasa más tiempo recolectando y preparando información que analizándola?
- b) ¿En ocasiones le frustra el no poder encontrar información que usted está seguro que existe dentro de la empresa?
- c) ¿Pasa mucho tiempo tratando de hacer que los reportes en Excel luzcan bien?
- d) ¿Quisiera tener una guía sobre las cosas que han sucedido cuando los administradores anteriores implementaban determinada estrategia?
- e) ¿No sabe qué hacer con tanta información que tiene disponible en la empresa?
- f) ¿Quiere saber qué productos fueron los más rentables durante un periodo determinado?
- g) ¿No sabe cuáles son los patrones de compra de sus clientes dependiendo de las zonas?
- h) ¿Ha perdido oportunidades de negocio por recibir información retrasada?
- i) ¿Trabaja horas extras el fin de mes para procesar documentos o reportes?
- j) ¿No sabe con certeza si su gente está alcanzando los objetivos planeados?
- k) ¿No sabe si mantiene una comunicación estrecha entre las diversas áreas de su empresa hacia una estrategia común?
- l) ¿No tiene idea de por qué sus clientes le regresan mercancía?

CONCEPTOS DE INTELIGENCIA DE NEGOCIOS

3.2.1.1. DATAWAREHOUSE

Conceptos Previos

Actualmente las empresas están inmersas en entornos cada vez más competitivos, en los cuales es fundamental el disponer de información valiosa con la que se puedan adoptar estrategias empresariales que lleven a situar el negocio por delante de sus competidores. Para este propósito es indispensable disponer de información que permita a los directivos de las empresas a tomar decisiones a nivel estratégico de su negocio.

Para esto, las empresas, independiente de su tamaño, cuentan con un conjunto de aplicaciones de procesamiento transaccional que mecanizan los procesos operativos, muy estructurados y repetitivos, que constituyen las funcionalidades básicas de la entidad. En este conjunto de aplicaciones se procesan, de manera automática, grandes volúmenes de datos referentes a las actividades rutinarias, que se almacenan en bases de datos operativas. De ellas se puede extraer información, fundamentalmente válidas para transacciones del día a día, es decir, sirven para apoyar y ejecutar las decisiones operativas que conducen a las actividades básicas,

pero no sirven para realizar análisis mas avanzados, incluso de tipo estratégico, ya que no están diseñadas para apoyar este tipo de tareas.

A la hora de atender las necesidades de información para la toma de decisiones a nivel estratégico, los sistemas operacionales no son de mucha utilidad. Esto es debido a que estas necesidades se basan en el análisis de una gran cantidad de datos, en el cual es importante un valor muy detallado del negocio como valor totalizado.

Como solución a los problemas de información de las empresas, a partir de los datos almacenados en las bases de datos operacionales, es posible extraer un cúmulo de datos que aporten un valor añadido a la gestión de la empresa, lo que constituirán los Data Warehouses.

El ingreso de datos en el Data Warehouse viene desde el ambiente operacional en casi todos los casos. El Data Warehouse es siempre un almacén de datos transformados y separados físicamente de la aplicación donde se encontraron los datos en el ambiente operacional. Como apoyo al entendimiento del origen de los datos para la transformación, es necesario hacer una referencia a la estructuración de las organizaciones con respecto a sus sistemas de información.

SISTEMAS DE INFORMACIÓN

Los sistemas de información se han dividido de acuerdo al siguiente esquema:

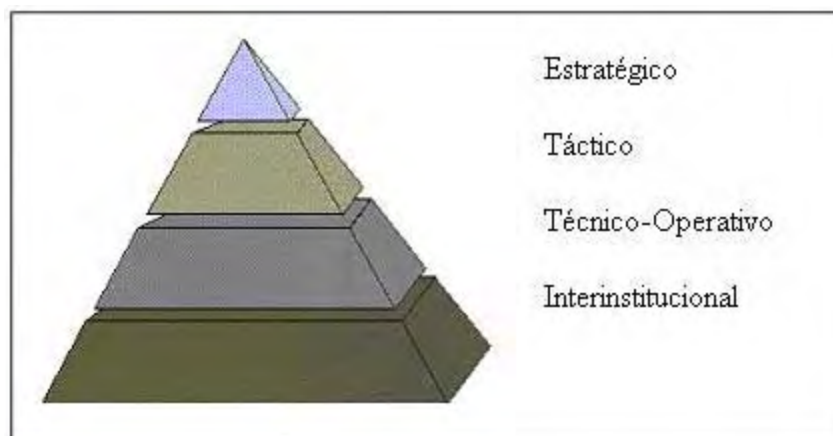


Figura 3.1: Niveles de la Organización

a. Sistemas Estratégicos

Están orientados a soportar la toma de decisiones, facilitan la labor de la dirección, proporcionando un soporte básico, en forma de mejor información, para la toma de decisiones. Se caracterizan porque son sistemas sin carga periódica de trabajo, es decir, su utilización no es predecible, al contrario de los casos posteriores, cuya utilización es periódica.

Destacan entre estos sistemas: los Sistemas de Información Gerencial (MIS), Sistemas de Información Ejecutivos (EIS), Sistemas de Información Georeferencial (GIS), Sistemas de Simulación de Negocios (BIS y que en la práctica son sistemas expertos o de Inteligencia Artificial - AI).

b. Sistemas Tácticos

Diseñados para soportar las actividades de coordinación y manejo de documentación, definidos para facilitar consultas sobre información almacenada en el sistema, proporcionar informes y, en resumen, facilitar la gestión independiente de la información por parte de los niveles intermedios de la organización.

Destacan entre ellos: los Sistemas Ofimáticos (OA), Sistemas de Transmisión de Mensajería (Correo electrónico y Servidor de fax), coordinación y control de tareas (Work Flow) y tratamiento de documentos (Imagen, Trámite y Bases de Datos Documentales).

c. Sistemas Técnico – Operativos

Cubren el núcleo de operaciones tradicionales de captura masiva de datos (Data Entry) y servicios básicos de tratamiento de datos, con tareas predefinidas (contabilidad, facturación, almacén, presupuesto, personal y otros sistemas administrativos). Estos sistemas están evolucionando con la irrupción de sensores, autómatas, sistemas multimedia, bases de datos relacionales más avanzadas y data warehousing.

d. Sistemas Interinstitucionales

Este último nivel de sistemas de información recién está surgiendo, es consecuencia del desarrollo organizacional orientado a un mercado de carácter global, el cual obliga a pensar e implementar estructuras de comunicación más estrechas entre la organización y el mercado (Empresa Extendida, Organización Inteligente e Integración Organizacional), todo esto a partir de la generalización de las redes informáticas de alcance nacional y global (INTERNET), que se convierten en vehículo de comunicación entre la organización y el mercado, no importa dónde esté la organización (INTRANET), el mercado de la institución (EXTRANET) y el mercado (Red Global).

Sin embargo, la tecnología data warehousing basa sus conceptos y diferencias entre dos tipos fundamentales de sistemas de información en todas las organizaciones: los sistemas técnico - operacionales y los sistemas de soporte de decisiones. Este último es la base de un Data Warehouse.

Asimismo, podemos clasificar dentro de una organización, a los gerentes o los cargos de alto mando, encargados de tomar decisiones.

La forma de clasificar a los gerentes en las organizaciones estructuradas tradicionalmente, es decir, las que cuentan con estructuras de trabajo deliberadas o estructuradas configuradas como una pirámide, se muestra en la siguiente figura.

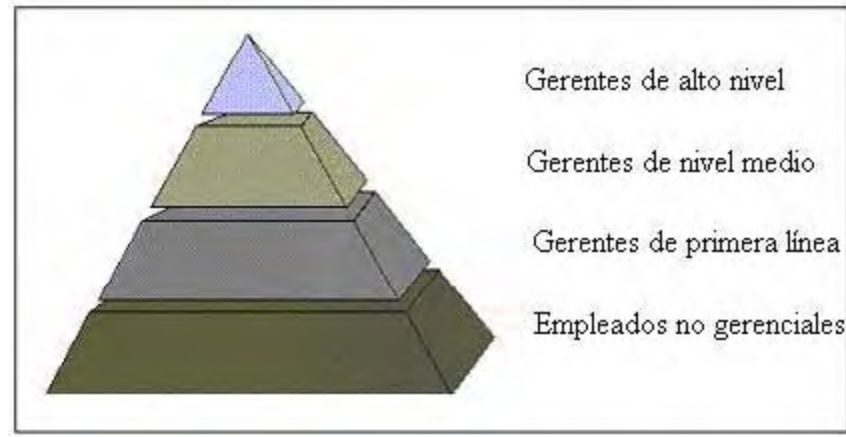


Figura 3.2: Clasificación de los Gerentes en las Organizaciones

Aquí se distinguen los gerentes de primera línea, que ocupan el nivel más bajo de la gerencia, y con frecuencia cumplen con labores como la de supervisión.

Los gerentes de nivel medio, pueden tener diversos títulos, como jefe de departamento o agencia, líder de proyecto, gerente de planta, jefe de unidad, decano o gerente de división.

En la cumbre de la organización, o cerca de ella, están los gerentes de alto nivel, responsables de tomar las decisiones y establecer las políticas y estrategias que afectan a toda la organización. Es precisamente a estos individuos, y a los gerentes del nivel medio, a los cuales está dirigido el propósito de este proyecto: facilitarles el proceso de toma de decisiones.

DATA WAREHOUSE

Los Data Warehousing, nacen debido a la necesidad de contar con información de apoyo a la toma de decisiones, dado que los datos operacionales no cumplen eficientemente con este objetivo. Esto conlleva a desarrollar Data Warehouse para poner el uso estratégico de la información como generador de ventajas competitivas.

Para entender el concepto Data Warehouse, daremos tres definiciones de autores diferentes:

“Un Data Warehouse es una copia de los datos de las transacciones de una organización, estructurados específicamente para la pregunta y el análisis”. (KIMBALL, RALPH) [4].

“Un Data Warehouse es algo que provee dos beneficios empresariales reales: Integración y Acceso de Datos. El Data Warehouse elimina una gran cantidad de datos inútiles y no deseados, como también el procesamiento desde el ambiente operacional clásico” (SUSAN OSTTERFELDT) [5].

“Un Data Warehouse es una colección de datos integrados, temáticos, no volátiles y variantes en el tiempo, organizados para soportar necesidades empresariales orientadas a la toma de decisiones”. (INMON, W.H.) [6].

La primera definición, no deja muy claro el concepto de Data Warehouse, ya que da una idea no muy clara ni muy global del concepto.

La segunda definición refleja claramente el principal beneficio que el Data Warehouse aporta a una organización, eliminar aquellos datos que obstaculizan la labor de análisis y entregar la información que se requiere en la forma más apropiada, facilitando así el proceso de gestión.

Para este proyecto, profundizaremos la tercera definición dada por Inmon, considerado el padre del Data Warehouse, por su sencillez, claridad y fácil comprensión.

Podemos concluir, que un Data Warehouse, es el proceso de extraer y filtrar datos de las operaciones comunes de las empresas, procedentes de los distintos subsistemas operacionales, para transformarlos, integrarlos, resumirlos y almacenarlos en un depósito o almacén de datos, para poder acceder a ellos cada vez que se necesite mediante mecanismos flexibles para el usuario.

CARACTERÍSTICAS

Siguiendo con la definición de Inmon, un Data Warehouse se caracteriza por ser:

➤ Integrado

El aspecto más importante en los Data Warehouses es que la información encontrada siempre esta integrada. La integración de los datos

se muestra de muchas maneras: en conversiones de nombres consistentes, en la medida uniforme de las variables, en la conversión de estructuras consistentes, en atributos físicos de los datos consistentes, fuentes múltiples, entre otros.

Los datos almacenados en el Data Warehouse, deben integrarse en una estructura consistente, por lo que las inconsistencias existentes en los diversos sistemas operacionales, deben ser eliminadas. La información suele estructurarse también en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.

Como se muestra en la siguiente Figura un Data Warehouse integra datos recogidos de diferentes sistemas operacionales de la organización y/o fuentes externas.

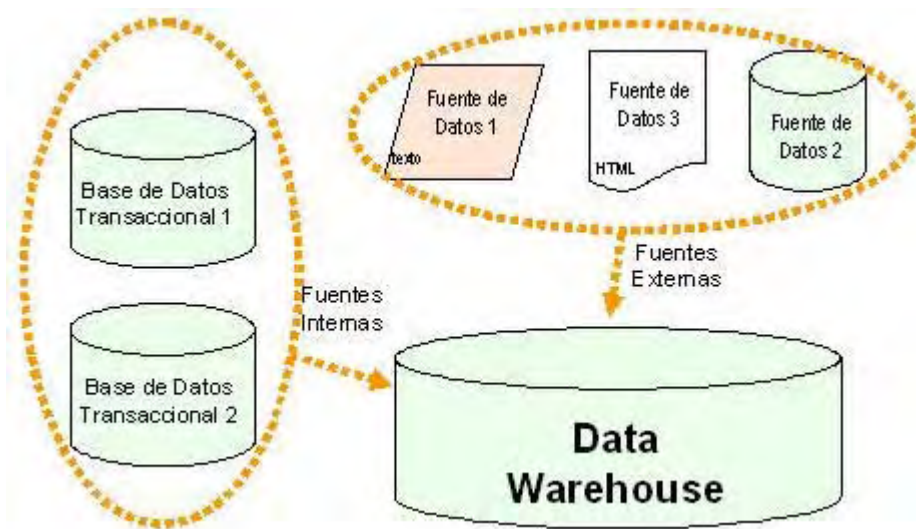


Figura 3.3: “Data Warehouse Integrado”

➤ Temáticos

Sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional. Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. Por ejemplo, como se aprecia en la siguiente figura, todos los datos sobre vendedores pueden ser consolidados en una única tabla del Data Warehouse. Así, las peticiones de información sobre vendedores serán más fáciles de responder dado a que toda la información está en el mismo lugar.

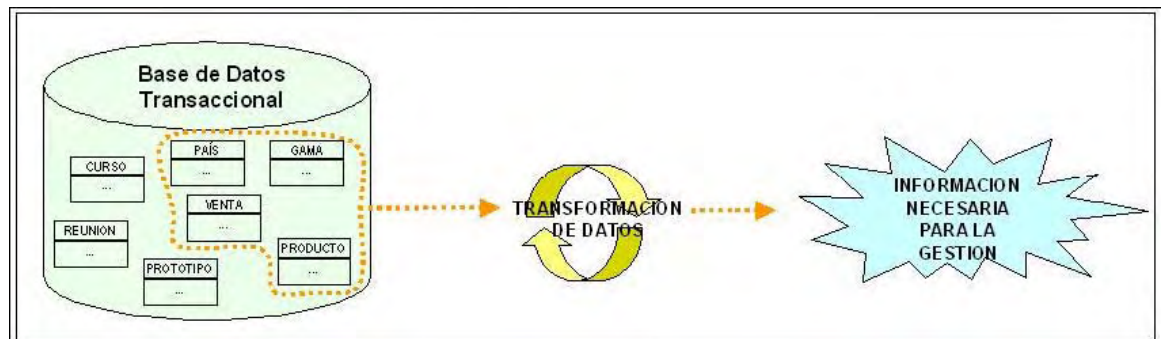


Figura 3.4: “Data Warehouse Temático”

➤ **No Volátil**

La información es útil solo cuando es estable. Los datos operacionales cambian sobre una base momento a momento. La perspectiva más grande, esencial para el análisis y la toma de decisiones, requiere una base de datos estable.

En la siguiente figura, se puede ver que la actualización (insertar, actualizar, borrar), se hace en el ambiente operacional. Pero la manipulación básica de los datos que ocurre en el Data Warehouse son: la carga inicial de los datos y el acceso a los mismos. Por lo tanto, la información de un Data Warehouse existe para ser leída, y no modificada.

Los datos almacenados no son actualizados, sólo son incrementados

Es importante mencionar que el periodo de tiempo cubierto por un Data Warehouse varía entre 2 y 10 años.

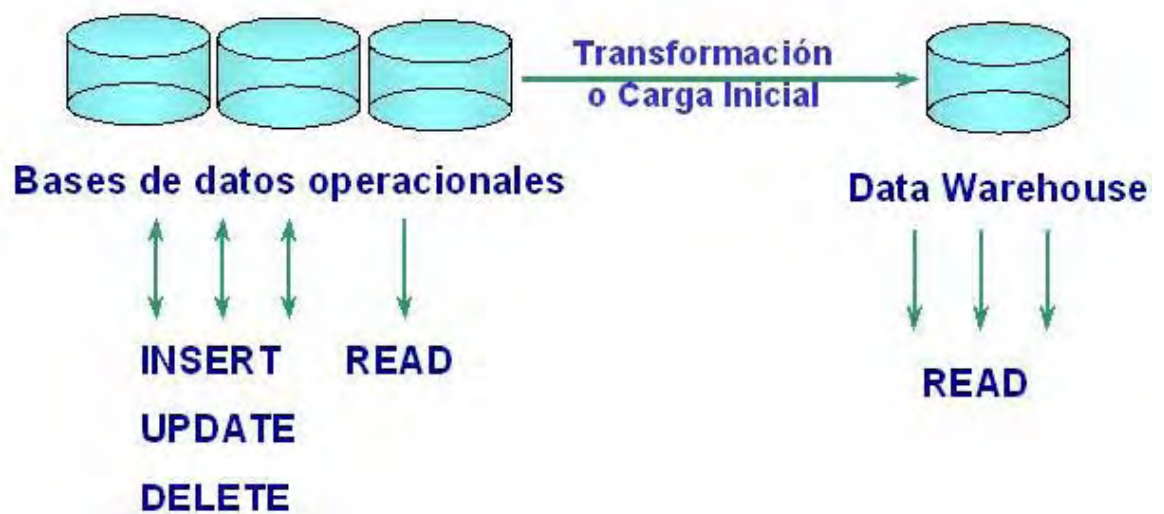


Figura 3.5: “Data Warehouse No Volátil”

➤ Variante en el Tiempo

El tiempo es la parte implícita de la información en un Data Warehouse. En los sistemas operacionales, los datos siempre reflejan el estado de la actividad del negocio en el momento presente. Por el contrario, la información contenida en el Data Warehouse sirve, entre otras cosas, para realizar análisis de tendencias. Por lo tanto, el Data Warehouse se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.

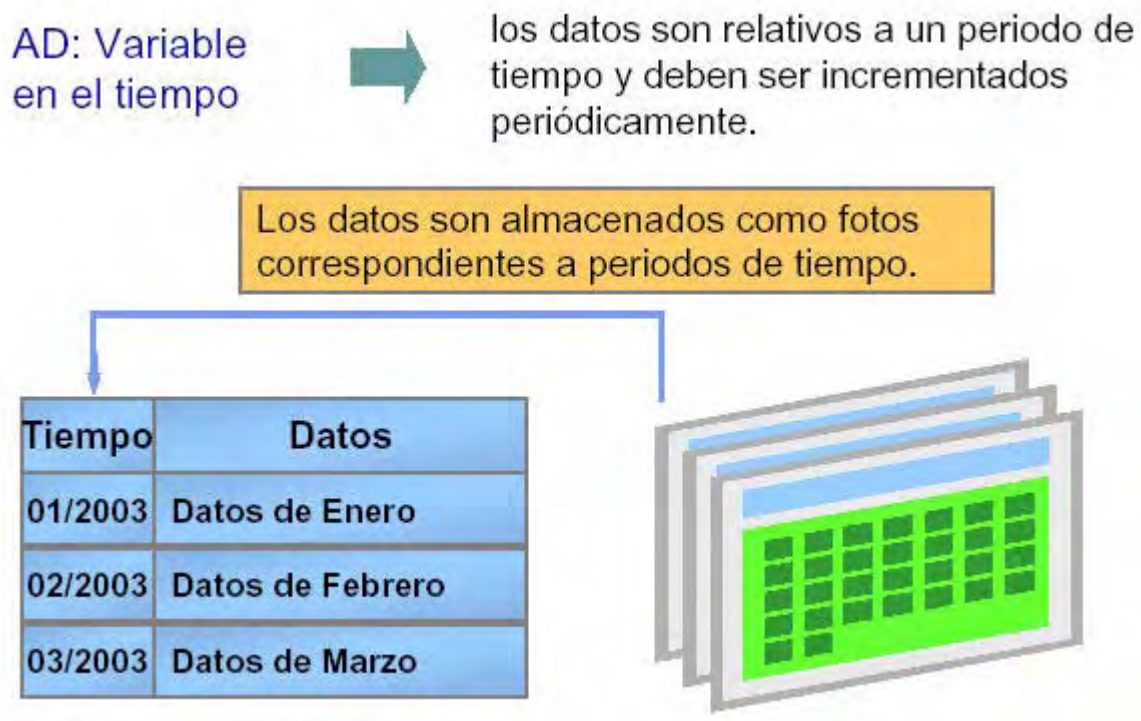


Figura 3.6: Datawarehouse -Variante en el Tiempo

PROBLEMAS Y VENTAJAS DEL DATA WAREHOUSE

Los Data Warehouse convierten los datos operacionales de una organización en una herramienta competitiva, que permite a los usuarios finales (gerentes de alto nivel, o nivel medio) examinar los datos de modo más estratégico, realizar análisis y detección de tendencias, seguimiento de medidas críticas, producir informes con mayor rapidez, un acceso más fácil, más flexible y más intuitivo a la información que se necesite en cada momento.

Frecuentemente, datos que son difíciles de interpretar, desde varias fuentes, se convierten en información lista para el usuario final.

Estas ventajas competitivas para el negocio, se pueden traducir en:

- Proporciona una herramienta para la toma de decisiones en cualquier área funcional, basándose en información integrada y global del negocio.
- Facilita la aplicación de técnicas estadísticas de análisis y modelización para encontrar relaciones ocultas entre los datos del almacén; obteniendo un valor añadido para el negocio de dicha información.
- Proporciona la capacidad de aprender de los datos del pasado y de predecir situaciones futuras en diversos escenarios.
- Simplifica dentro de la empresa la implantación de sistemas de gestión integral de la relación con el cliente.

- Supone una optimización tecnológica y económica en entornos de Centro de Información, estadística o de generación de informes con retornos de la inversión espectaculares.

Aun con los beneficios que se incorporan a la organización, junto con la implantación de un Data Warehouse, también se pueden presentar ciertos problemas, como:

- Infravaloración del esfuerzo necesario para su diseño y creación.
- Infravaloración de los recursos necesarios para la captura, carga y almacenamiento de los datos.
- Incremento continuo de los requisitos de los usuarios.
- Privacidad de los datos.

DIFERENCIAS ENTRE SISTEMA TRADICIONAL (OLTP) Y DATA WAREHOUSE

En cuanto a los requerimientos de diseño y características de implementación, los Sistemas Tradicionales (OLTP, Online Transactional Processing) y los Data Warehouse poseen muchas diferencias, las que podemos resumir en la siguiente tabla:

SISTEMA TRADICIONAL	DATA WAREHOUSE
Predomina la actualización	Predomina la consulta
La actividad más importante es del tipo operativo(día a día)	La actividad más importante es el análisis y la decisión estratégica
Predomina el proceso puntual	Predomina el proceso masivo
Mayor importancia a la estabilidad	Mayor importancia al dinamismo
Datos en general desagregados	Datos en distintos niveles de detalle y agregación
Importancia del dato actual	Importancia del dato histórico
Importancia del tiempo de respuesta de la transacción instantánea	Importancia de la respuesta masiva
Estructura relacional	Visión multidimensional
Usuarios de perfiles medios o bajos	Usuarios de perfiles altos
Explotación de la información relacionada con la operativa de cada aplicación	Explotación de toda la información interna y externa del negocio

Tabla.3.1: “Diferencias entre Sistema Tradicional versus Data Warehouse”

ARQUITECTURA DE UN DATA WAREHOUSE

Una de las razones por las que el desarrollo de un Data Warehouse crece rápidamente, es que es una tecnología muy entendible. Un Data Warehouse puede representar de mejor forma la amplia estructura de una empresa, para administrar los datos que serán útiles para la toma de decisiones dentro de la organización. A fin de comprender cómo se relacionan todos los componentes involucrados en un Data Warehouse, es esencial analizar su arquitectura.

- ✓ **Arquitectura real:** Es generalmente la arquitectura elegida para los sistemas de decisión. El almacenamiento de los Data Warehouses se realiza en un SGBD (Sistema Gestor de Base de Datos) separado del sistema de producción. Este SGBD se alimenta por extracciones periódicas. Antes de la carga, los datos sufren importantes procesos de integración, limpieza y transformación. La ventaja de esta solución es que se dispone de datos preparados para las necesidades de la decisión que responden bien a los objetivos de Data Warehouse. La principal razón para justificar la arquitectura real es la inadaptación de los datos de producción a las necesidades de los sistemas de decisión. Las estructuras de datos en un sistema de producción son complejas en cuanto a almacenamiento e implican

herramientas de programación para acceder a ellas. Los datos también están codificados. En un contexto de ayuda a la decisión, el dato debe ser comprensible por el usuario. Es necesario transformar todos los códigos legibles y comprensibles. Las cargas relacionadas con los accesos son también incompatibles con los objetivos de rendimiento del sistema de producción. Los datos están dispersos. El Data Warehouse debe integrar los datos unos con los otros a fin de asegurar una coherencia global.

- ✓ **Arquitectura virtual:** En una arquitectura virtual, los datos del Data Warehouse residen en el sistema tradicional. Se hacen visibles por productos middleware o pasarelas. En esta arquitectura no hay coste de almacenamiento suplementario y el acceso se hace en tiempo real. Sin embargo, las numerosas desventajas de este tipo de arquitectura impiden frecuentemente su elección. Los datos no son preparados. El párrafo anterior muestra la dificultad real que presenta la utilización de los datos de la empresa. Los accesos de decisión pueden perturbar el rendimiento del sistema de la empresa, tanto más cuanto que los procesos de transformación y de integración están aquí relacionados forzosamente con los procesos de acceso. Según los antecedentes recopilados en diferentes fuentes, no existe ninguna implementación de Data Warehouse virtual.

- ✓ **Arquitectura remota:** La arquitectura remota es una combinación de los dos tipos de arquitectura descritas anteriormente. El objetivo es implementar físicamente los niveles agregados (los niveles de datos utilizados mas a menudo) a fin de facilitar el acceso y conservar el nivel de detalle en los sistemas de producción dando acceso por medio de Middleware o de Gateways. Esta arquitectura se utiliza también muy raramente.

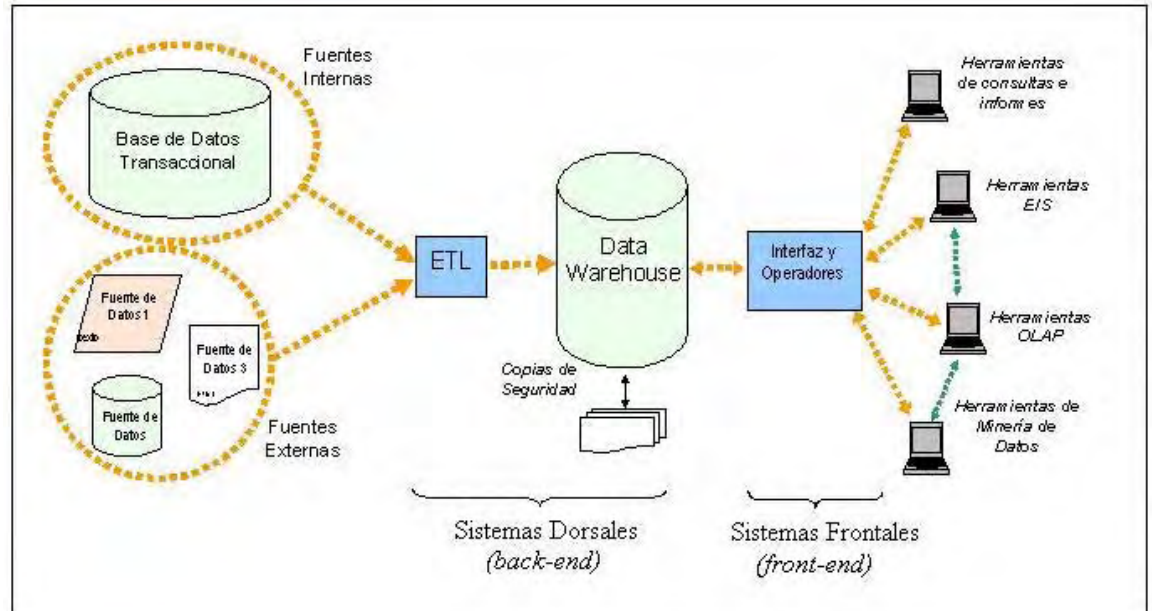


Figura 3.7: "Arquitectura real de un Data Warehouse"

En la figura, se muestra la arquitectura de un Data Warehouse, que viene determinada por su situación central como fuente de información para las herramientas de análisis. En la figura 3.7 se puede distinguir dos tipos

de sistemas: Sistemas Dorsales (conocidos como back-end) y los Sistemas Frontales (o front-end tools).

❖ **Sistemas Dorsales (back-end)**

Los Sistemas Dorsales, que son herramientas de adquisición de datos de fuentes externas, se encargan principalmente de extracción, limpieza, transformación, integración, carga y refresco de los datos.

Dentro de la clasificación de sistemas dorsales, podemos mencionar los sistemas ETL y el Data Warehouse mismo, que se describen a continuación.

Los Sistemas ETL (Extraction, Transformation, Load), se encargan de las funciones de:

- La extracción de las distintas fuentes de datos, sean estas transaccionales o externas.
- Transformación: filtrado de los datos, limpieza, consolidación de los datos, etc.
- La carga inicial del Data Warehouse: ordenación, agregaciones, etc.
- Refresco del Data Warehouse: operación periódica que propaga los cambios de las fuentes externas al Data Warehouse mismo Repositorio Propio de Datos (Data Warehouse), que posee información relevante de los datos o metadatos. Los metadatos contienen información relativa a

los datos. Comúnmente, los metadatos permiten mantener información sobre:

- La procedencia de los datos
- La periodicidad de refresco de los datos
- La fiabilidad y forma de cálculo de los datos
- Semántica de los datos y su localización en el DW.
- Localización de los datos en los sistemas de producción y reglas de transformación.
- Especificación de fórmulas de cálculo de agregados.
- Información sobre frecuencias de carga, mecanismo de historización, etc.

❖ **Sistemas Frontales (front-end)**

Los sistemas frontales, o herramientas front-end, ofrecen al usuario final mecanismos eficaces de acceso a la información mediante mecanismos simples y potentes, y así poder tener una conexión eficaz con el Data Warehouse. Cumplen con las tareas de realizar consultas, informes, análisis y extracción del conocimiento (data mining). Dentro de esta clasificación podemos mencionar las Interfaces y Gestores de Consultas y los Sistemas de Integridad y Seguridad.

Las Interfaces y Gestores de Consulta permiten acceder a los datos y sobre ellos conectar herramientas más sofisticadas como OLAP (Online Analytical Processing, EIS (Executive Information System) es un sistema de información y un conjunto de herramientas asociadas, proporciona a los directivos acceso a la información de estado y sus actividades de gestión, Está especializado en analizar el estado diario de la organización (mediante indicadores clave) para informar rápidamente sobre cambios a los directivos, La información solicitada suele ser, en gran medida, numérica (ventas semanales, nivel de stocks, balances parciales, etc.) y representada de forma gráfica al estilo de las hojas de cálculo.

Es un software diseñado para apoyar a los ejecutivos de la empresa al entregarle datos en forma gráfica, apoyando la toma de las grandes decisiones

EIS es un sistema de información para directivos que permite automatizar la labor de obtener los datos más importantes de una organización, resumirlos y presentarlos de la forma más comprensible posible, provee al ejecutivo acceso fácil a información interna y externa al negocio con el fin de dar seguimiento a los factores críticos del éxito.

EIS se enfoca primordialmente a proporcionar información de la situación actual de la compañía y dejan en un plano secundario la visualización o proyección de esta información en escenarios futuros.

MODELO DE DATOS

Un modelo de datos es un sistema formal y abstracto que permite describir los datos de acuerdo con reglas y convenios predefinidos. Es formal pues los objetos del sistema se manipulan siguiendo reglas perfectamente definidas y utilizando exclusivamente los operadores definidos en el sistema, independientemente de lo que estos objetos y operadores puedan significar.

Para entender una de las partes principales de la arquitectura del Data Warehouse, como es el modelo de datos, es necesario conocer primero la diferencia que existe entre los modelos existentes: Modelo Relacional y Modelo multidimensional.

✓ Modelo Relacional

El modelo relacional puede considerarse como un lenguaje de programación más bien abstracto, orientado de manera específica hacia las aplicaciones de bases de datos.

En términos tradicionales una relación se asemeja a un archivo, una tupla a un registro, y un atributo a un campo. Pero estas correspondencias son aproximadas, en el mejor de los casos.

Una relación no debe considerarse como “solo un archivo”, sino más bien como un archivo disciplinado, siendo el resultado de esta disciplina una simplificación considerable de las estructuras de datos

con las cuales debe interactuar el usuario, lo cual a su vez simplifica los operadores requeridos para manejar esas estructuras.

El modelo relacional es el pilar fundamental para el diseño de la mayoría de las bases de datos que existen hoy en las grandes y pequeñas empresas. La composición de estas bases de datos son decenas de tablas relacionadas a través de una compleja tela de araña de uniones.

La implementación de bases de datos con un modelo relacional da lugar a escenarios como los siguientes:

- Legibilidad limitada: Los usuarios finales no son capaces de entender el modelo entidad relación. Evidentemente no pueden navegar por dicho modelo en busca de información.
- Dificultad para las herramientas de consulta en el acceso a un modelo entidad relación general: Los optimizadores de las bases de datos no siempre realizan la selección correcta y a menudo adolecen de prestaciones mediocres o inaceptables cuando trabajan en entornos de grandes volúmenes de información.
- La utilización de la técnica del modelado relacional frustra el principal atractivo de Data Warehouse, que es, la recuperación de información intuitiva y con alto rendimiento.

✓ **Modelo Multidimensional**

El modelo multidimensional es una técnica para modelar bases de datos simples y entendibles al usuario final, ya sea, para presentar la información en un marco estándar e intuitivo que permitan un acceso de alto rendimiento. Los objetivos del modelo multidimensional son:

- Representar los datos en forma cercana a la intuición del usuario.
- Resolver problemas planteados en sistemas relacionales.

Este tipo de modelado es independiente de las tecnologías y permite el empleo de cualquier base de datos, ya sea multidimensional, MOLAP, ésta almacena y manipula los datos en estructuras especiales (matrices multidimensionales); Base Datos relacional, ROLAP, la que almacena los datos en un SGBD relacional extendido y transforma operaciones sobre datos multidimensionales en operaciones relacionales en SQL; Base de datos objetos, etc.

Componentes del Modelo Multidimensional

El Modelo Multidimensional está compuesto principalmente por dos elementos: esquemas y tablas.

Las tablas en el Modelo Multidimensional se dividen en dos tipos:

- **Tablas Hechos:** Es el objeto a analizar, además posee atributos llamados atributos de hechos o síntesis, estos atributos son de tipo cuantitativo cuyos valores (cantidades) se obtienen generalmente por aplicación de una función estadística que resume un conjunto de valores en un único valor. Además contienen funciones resumen, funciones de tipo estadístico que se aplican a los atributos de hecho.

- **Tablas Dimensiones:** Representan cada uno de los ejes en un espacio multidimensional. Como todas las tablas también poseen atributos llamados de dimensión o de clasificación, estos son de tipo cualitativo (sus valores son modalidades) que suministran el contexto en el que se obtienen las medidas en un esquema de hecho. Las dimensiones poseen jerarquías, que son varios atributos unidos mediante una relación de tipo jerárquico.

Un esquema multidimensional se puede representar como un esquema estrella (star join), copo de nieve (snowflake) o Constelación de hechos.

Esquema estrella: Formado por una tabla de hechos con una única tabla para cada dimensión,

El esquema estrella deriva su nombre del hecho que su diagrama forma una estrella, con puntos radiales desde el centro. El centro de la estrella consiste de una o más tablas fact, y las puntas de la estrella son las tablas lock_up.

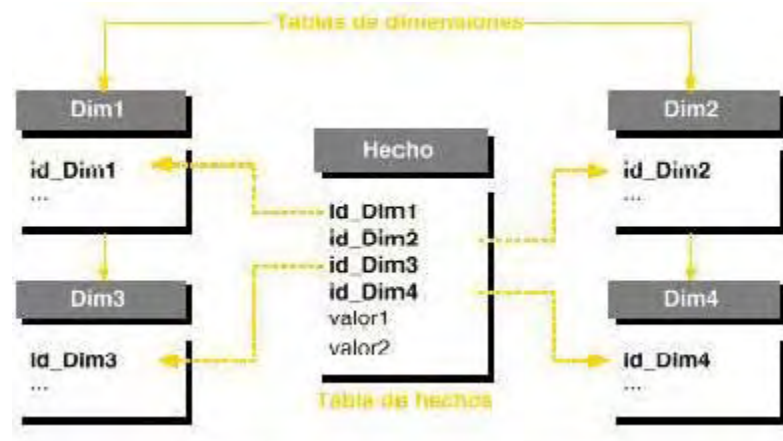


Figura 3.8: “Esquema Estrella”

Esquema copo de nieve: La diferencia del esquema snowflake comparado con el esquema estrella, está en la estructura de las tablas lock_up: las tablas lock_up en el esquema snowflake están normalizadas. Cada tabla lock_up contiene sólo el nivel que es clave primaria en la tabla y la foreign key de su parentesco del nivel más cercano del diagrama como se puede observar en la siguiente figura:

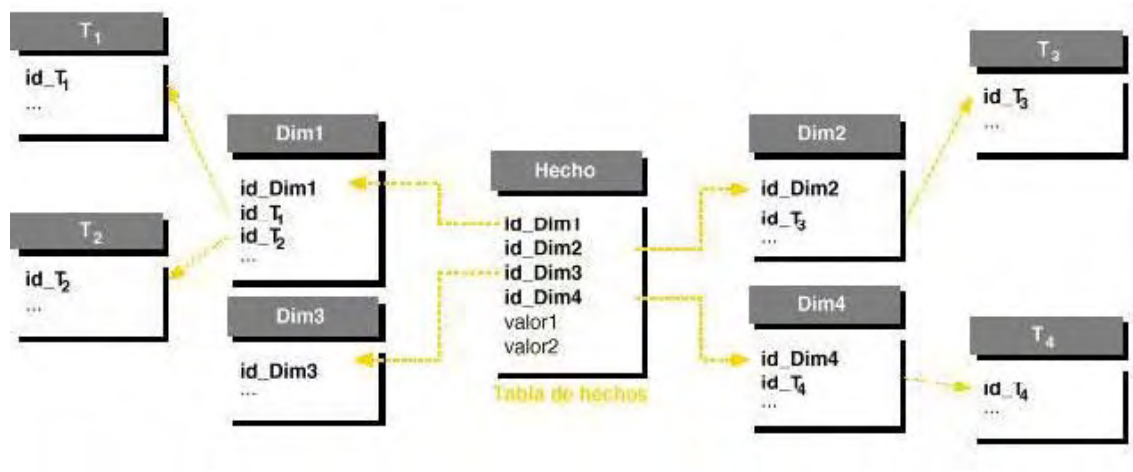


Figura 3.9: “Esquema Copo de Nieve”.

Constelación de Hechos: es un conjunto de tablas de hechos que comparten algunas tablas de dimensiones, como lo muestra la siguiente figura:

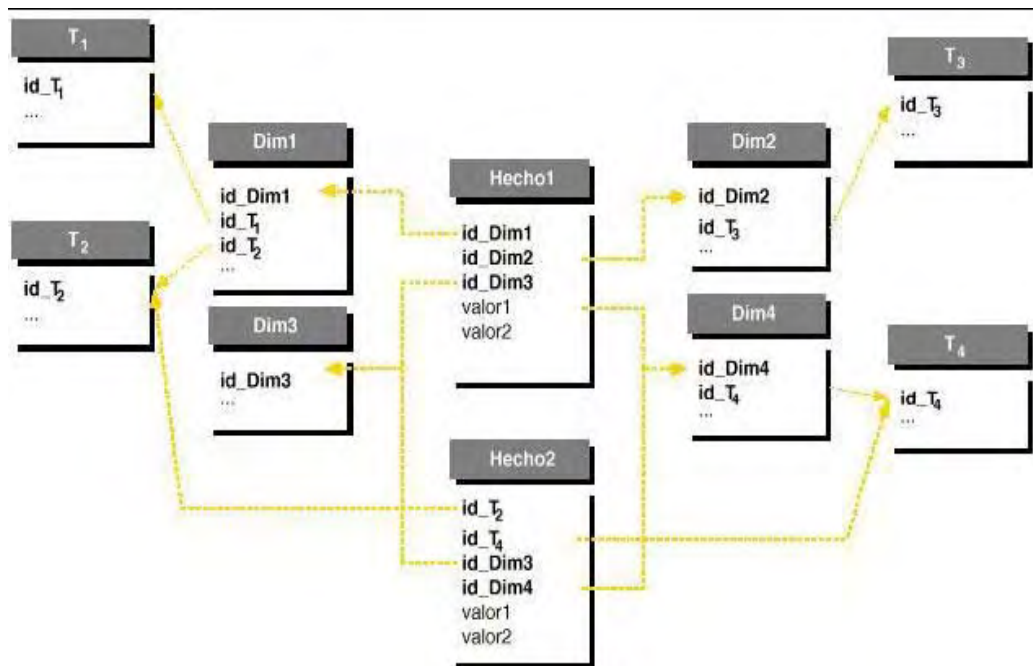


Figura 3.10: “Constelación de Hechos”

Estructura Básica de las Bases de Datos Multidimensionales.

Un modelo multidimensional soporta el manejo de una basta cantidad de datos empresariales y temporales. De esta forma surge la instancia del modelo multidimensional, también conocido como **Cubo** o **Hipercubo**, en la cual podemos encontrar en interacción los conceptos mencionados en estos párrafos.

En una base de datos multidimensional la información se representa como una matriz multidimensional, en los ejes de esta matriz se encuentran los criterios de análisis, en los cruces están los valores a analizar. Como ejemplo de Cubo encontramos el que nos muestra la siguiente figura, la cual contiene tres dimensiones: modelo, vendedor y color; donde cada dimensión tiene tres diferentes niveles o hechos, para finalmente encontrar la medida al intersectar los valores.

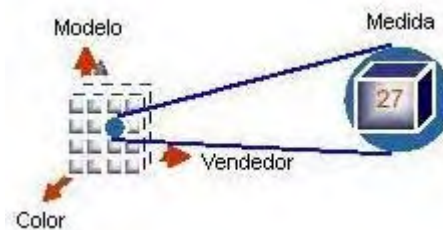


Fig. 3.11: “Ejemplo de un Cubo con tres dimensiones”.

Los Cubos o Hipercubos están formados por:

- Dimensiones: Las dimensiones son un concepto esencial de bases de datos multidimensionales. Las dimensiones representan los criterios de análisis de los datos, macro-objetos del problema y variables independientes. Si una dimensión tiene más de un nivel entonces los miembros de la dimensión pueden ser organizados en una o más jerarquías, como lo muestra la siguiente figura:

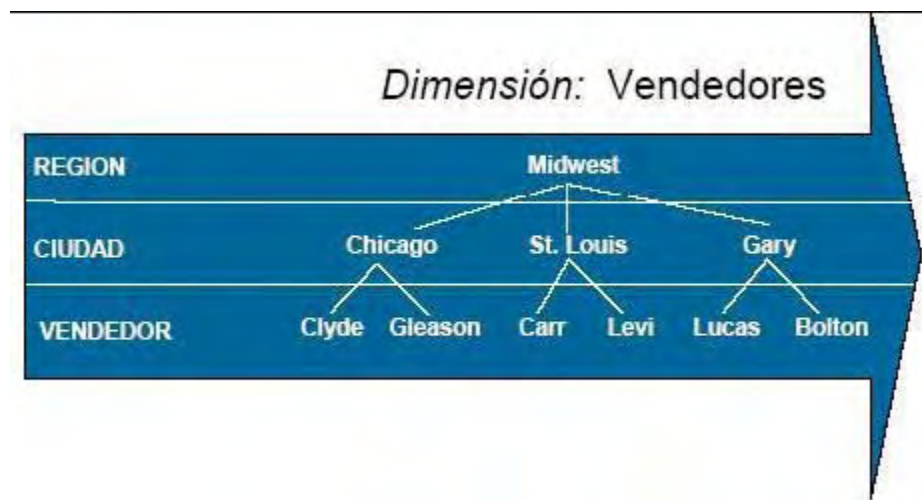


Figura 3.12: "Dimensión con Niveles"

- Medida: Las medidas son un dato numérico que representa una actividad específica de un negocio, mientras que una dimensión representa una perspectiva de los datos. Cada dimensión está descrita por un conjunto de atributos (datos agregados). A su vez se pueden intersectar estas dimensiones para obtener un valor, llamado medida.

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

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

Operaciones en Modelos Multidimensionales

Las principales operaciones en los modelos multidimensionales, son:

- ❖ Slice & Dice: Éste nos permite hacer una selección de valores de las dimensiones que uno requiere. En la siguiente figura se puede observar un ejemplo de este tipo en el cual el usuario solo necesita regiones y productos en un tiempo específico.

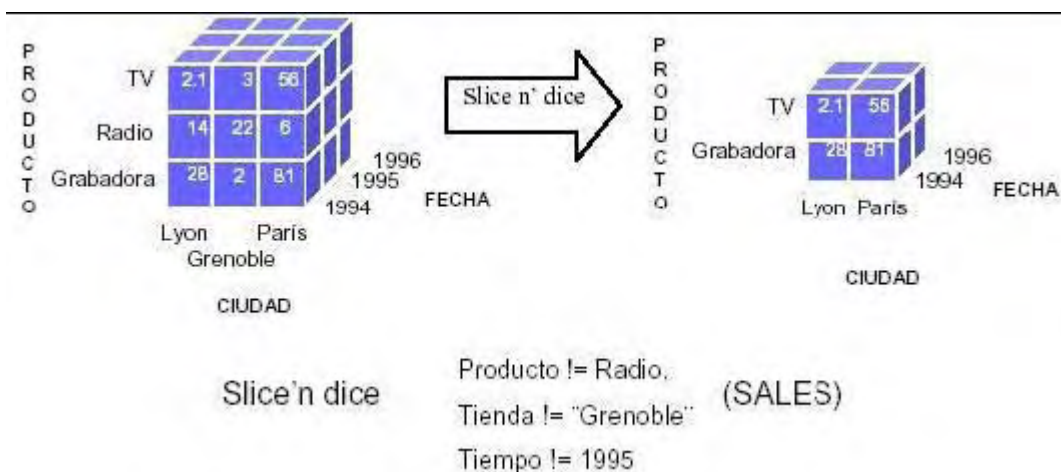


Figura 3.13: "Slice & Dice".

- ❖ Rotación: Este operador selecciona el orden de visualización de las dimensiones. En la siguiente figura se muestra un ejemplo de ello.



Figura 3.14: “Rotación”

- ❖ Movimientos en la Jerarquía de una dimensión (Drill Down, Drill Up): Este operador nos permite bajar o subir a los niveles más atómicos de nuestro esquema multidimensional. En la figura siguiente se puede ver cómo se ajustan las escalas de los ejes, produciendo movimientos en la jerarquía.

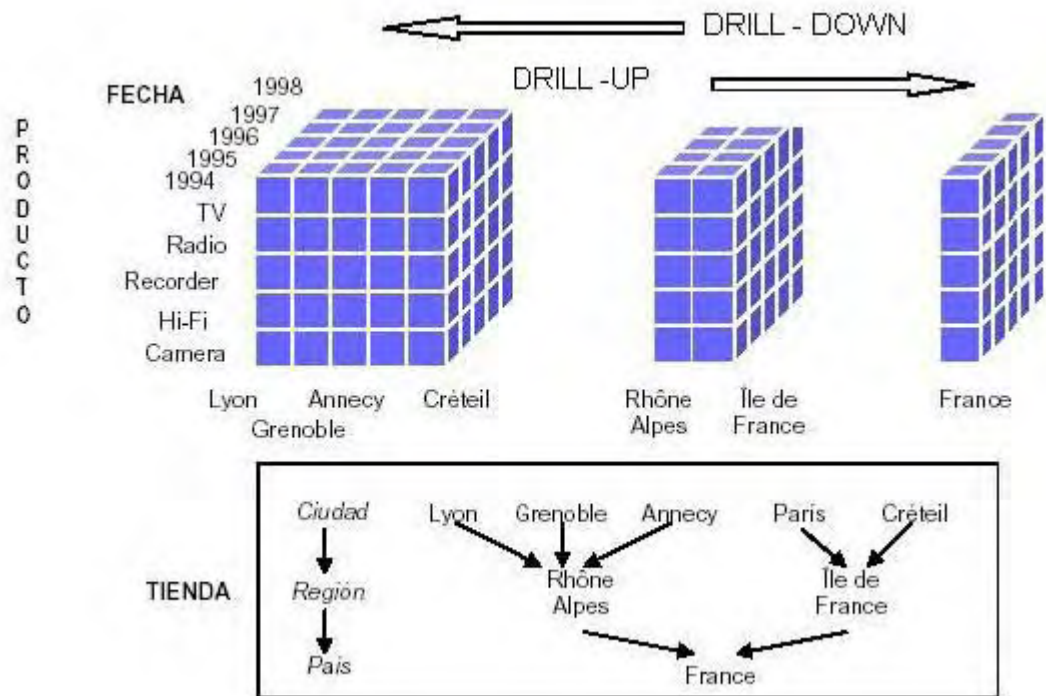


Figura 3.15: “Drill-Down y Drill-Up”.

- ❖ **Consolidación (Roll-Up):** Este operador nos permite hacer cálculos a las medidas en función de agrupamientos. Realiza el recálculo de la medida de acuerdo a los ajustes de escala, además se debe especificar cuál es la operación que calcula el nuevo valor de la medida. Esta operación pueden ser: suma, promedio, etc. También pueden haber medidas con comportamientos diferentes. Por ejemplo en la figura siguiente se realiza un Roll –Up suma sobre la dimensión Tienda.

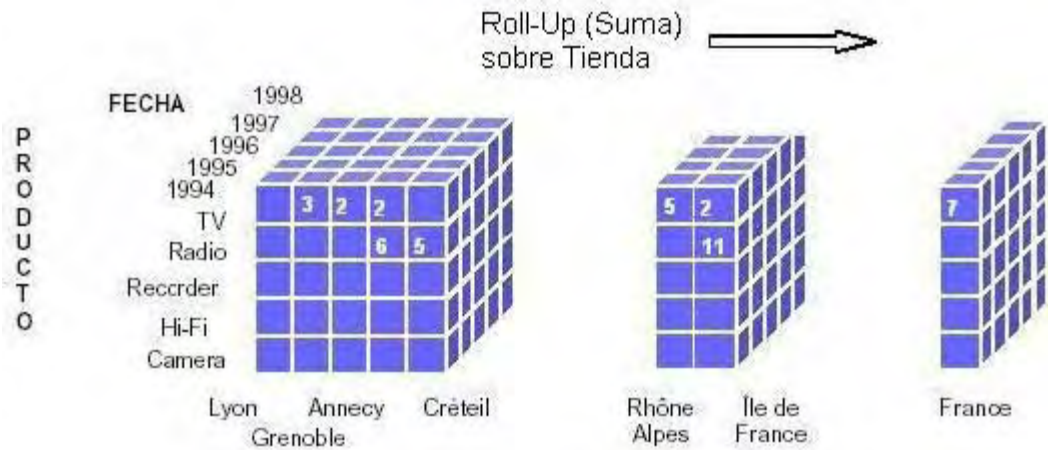


Figura 3.16: “Roll-Up”

- ❖ Drill-Across: Operador que relaciona dos cubos, como lo muestra la figura siguiente:

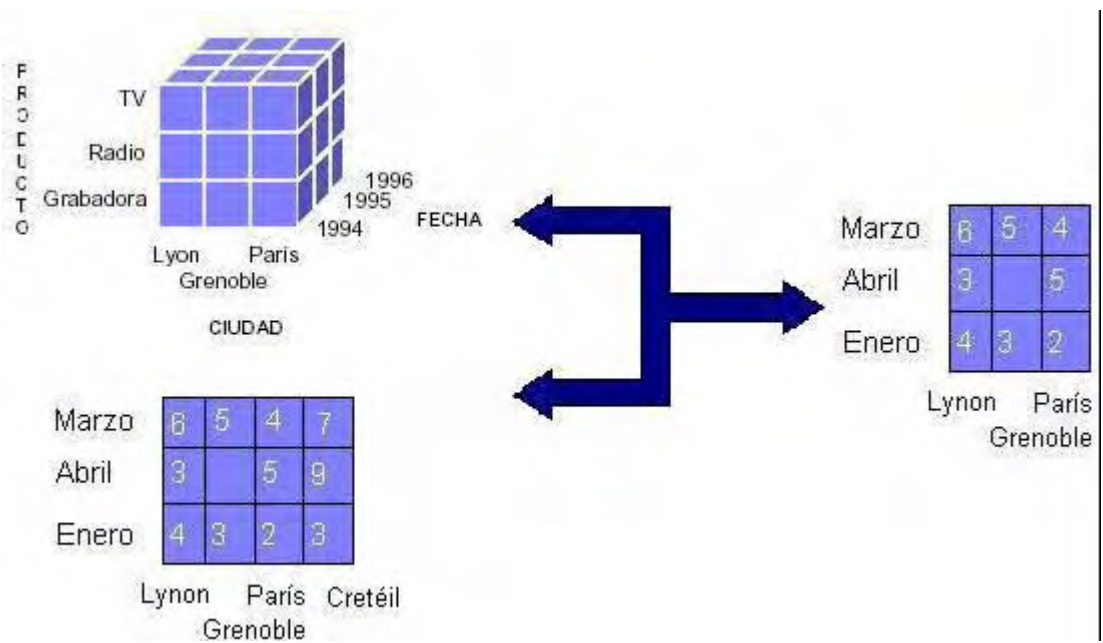


Figura 3.17. “Drill-Across”.

- ❖ Drill-Through: Operador que accede a datos descriptivos. Un ejemplo se muestra en la siguiente figura:

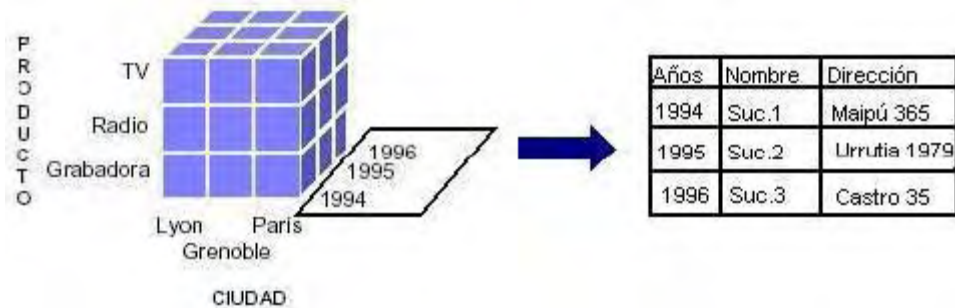


Figura 3.18: “Drill-Trough”

Ventajas del Modelo Multidimensional

Implementar un Data Warehouse a través del modelo multidimensional tiene un número importante de ventajas:

- Primero, es un modelo, una estructura homogénea y predecible. Reportes escritos, herramientas de consulta e interfaces de usuario, todas pueden hacer suposiciones acerca del modelo multidimensional para que las interfaces del usuario sean más entendibles y los procesos, más eficientes.
- Una segunda ventaja del modelo multidimensional reside en que es una estructura predecible. Cada dimensión es equivalente. Todas las dimensiones pueden ser vistas como un conjunto de puntos igualmente simétricos dentro de una tabla de hechos.

- Una tercera ventaja del modelo multidimensional, es que es extensible para acomodar nuevos elementos de datos inesperados y nuevas decisiones de diseño. Esto se lleva a cabo añadiendo nuevas dimensiones, nuevos atributos dimensionales y cambiando datos de una cierta granularidad para pasarlos a otra.
- Una última ventaja del modelo multidimensional es la creciente cantidad de herramientas administrativas y de software que manejan y usan la granularidad.

Ventajas DW:

- ✓ Da apoyo al CRM.
- ✓ Cumple muy bien funciones de marketing.
- ✓ Se anticipa a las necesidades de los clientes.
- ✓ Administra mucha información.

Desventajas DW:

- ✓ Es muy costoso.
- ✓ Se invierte mucho tiempo en implementarlo.
- ✓ Se invierte mucho tiempo en poblarlo.

3.2.1.2. DATAMART

Las corporaciones de hoy se esfuerzan por conducir sus negocios hacia una base internacional. Vemos compañías que surgieron en Estados Unidos y se expandieron a Europa, Asia y África. La expansión del negocio crea la necesidad de acceder a datos corporativos que están ubicados en diferentes puntos geográficos. Por ejemplo, un ejecutivo de ventas de una compañía con origen en Perú que está situado en Brasil puede necesitar acceso a la base de datos de la empresa para identificar los clientes potenciales que residen solo en Brasil.

Este problema se soluciona creando versiones más pequeñas del DataWarehouse, los datamarts. Estas versiones se crean usando algún criterio particular, como por ejemplo el lugar geográfico. En el ejemplo anterior los datos de los clientes que residen en Brasil se deben almacenar en el datamart de la sucursal en ese país.

La existencia de los datamarts crea nuevas formas de pensar cuando se diseñan los repositorios corporativos de datos.

Algunas corporaciones reemplazan completamente el concepto de tener un DataWarehouse central, por varios datamarts más pequeños que se alimenten directamente de los sistemas operacionales.

Otras compañías usan datamarts para complementar sus DataWarehouses, mueven datos desde el DataWarehouse hacia varios datamarts con el fin de permitir un análisis más eficiente. La separación de los datos se determina según criterios como departamentos, áreas geográficas, periodos de tiempo, etc.

Finalmente, algunas organizaciones usan sus datamarts como el primer paso de almacenamiento de datos operacionales. Luego los datos de todos los datamarts se replican en un DataWarehouse corporativo central.

Los Datamart se implementan antes que el Data Warehouse, como un plan piloto, siendo el conjunto de Datamart específicos orientados al Data Warehouse. Se debe restringir el acceso al Datamart y a las bases de datos con las que trabaja a sólo personal autorizado.

En la mayoría de empresas del Perú y del extranjero se puede apreciar que muchas ya cuentan de una u otra manera con diferentes datamart, lo cual está muy bien, ya que las empresas de hoy se han vuelto "Inforveras", es decir, tienen la necesidad constante de consumir información para poder sobrevivir...

Sin embargo muchos de estos datamart fueron creados enfocados en los datos y no en las necesidades de información de los usuarios.

Para reconocer si el datamart ha sido creado enfocado en los datos y no en las necesidades de información, se deben identificar alguno de estos síntomas:

- Los usuarios solo utilizan reportes similares a los reportes de su sistema transaccional.
- Los reportes toman mucho tiempo para procesar.
- Los usuarios no utilizan el Datamart.

- Necesitan que el personal de sistemas hagan los reportes porque ellos no saben como hacerlo.
- Los usuarios exportan la información a una hoja de cálculo para agruparla o enlazarla con otra información.

Si su empresa tiene alguno de estos síntomas es momento de implementar un Datamart Enfocado al Usuario.

Los pasos para implementar un Datamart enfocado en el usuario [7] son:

Paso 1: Identificar los temas de análisis.

Esta tarea consiste en definir los temas y sus respectivos indicadores que serán analizados y explotados por los usuarios, por ejemplo:

Tema: Ventas

Indicadores: Cantidad Vendida, Precio Unitario, Total, Descuento, IGV, etc.



Figura 3.19: Ventas

Paso 2: Identificar las dimensiones de Información

Las dimensiones de Información es la forma cómo el usuario podrá agrupar la Información.

Las dimensiones siempre deben responder una de estas 6 preguntas: A Quién, Dónde, Cuándo, Qué, Cómo y A quien.

Recuerde que el usuario siempre necesitará que el datamart le responda cualquiera de estas preguntas con la finalidad de poder tomar dediciones mas acertadas.

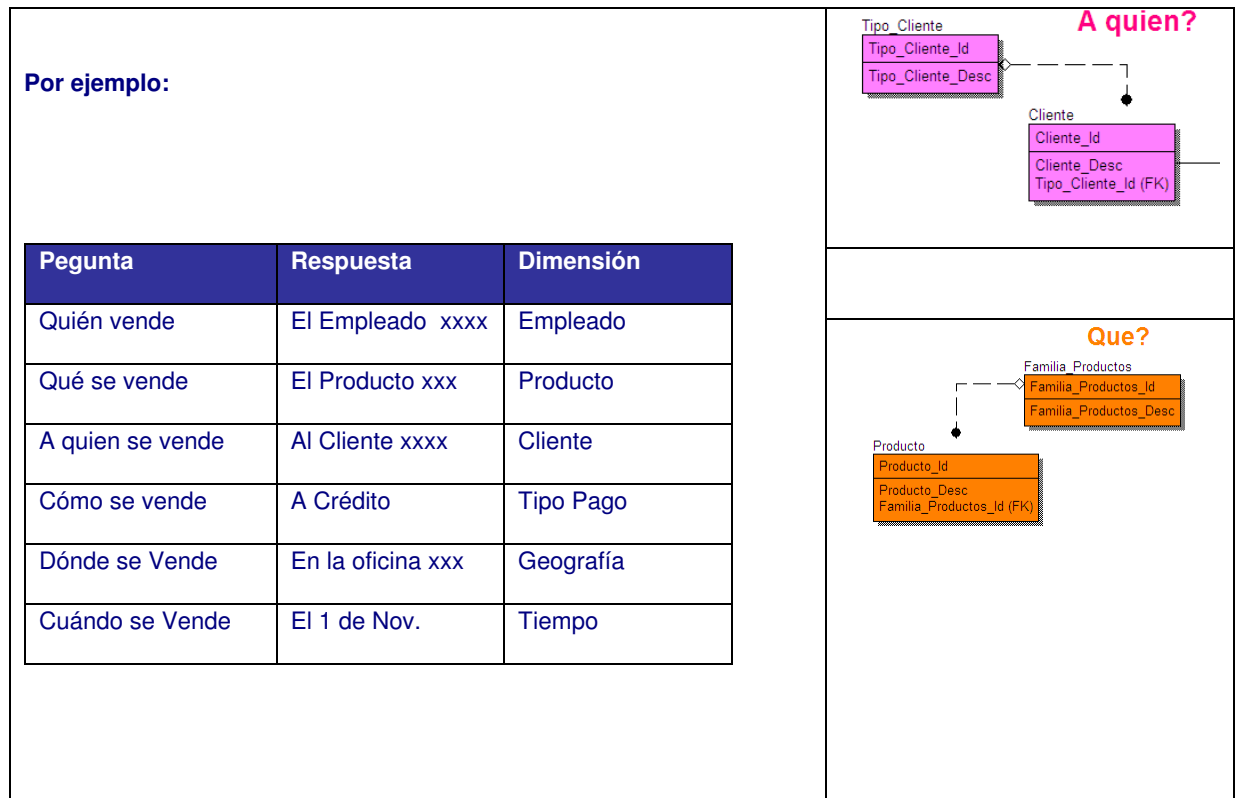


Figura 3.20: Identificación de Dimensiones

Paso 3: Elaboración del Modelo Multidimensional Básico

Con ayuda de alguna herramienta CASE, deberá diseñar un modelo multidimensional capaz de soportar cualquiera de las consultas que los usuarios puedan hacer en un futuro al datamart.

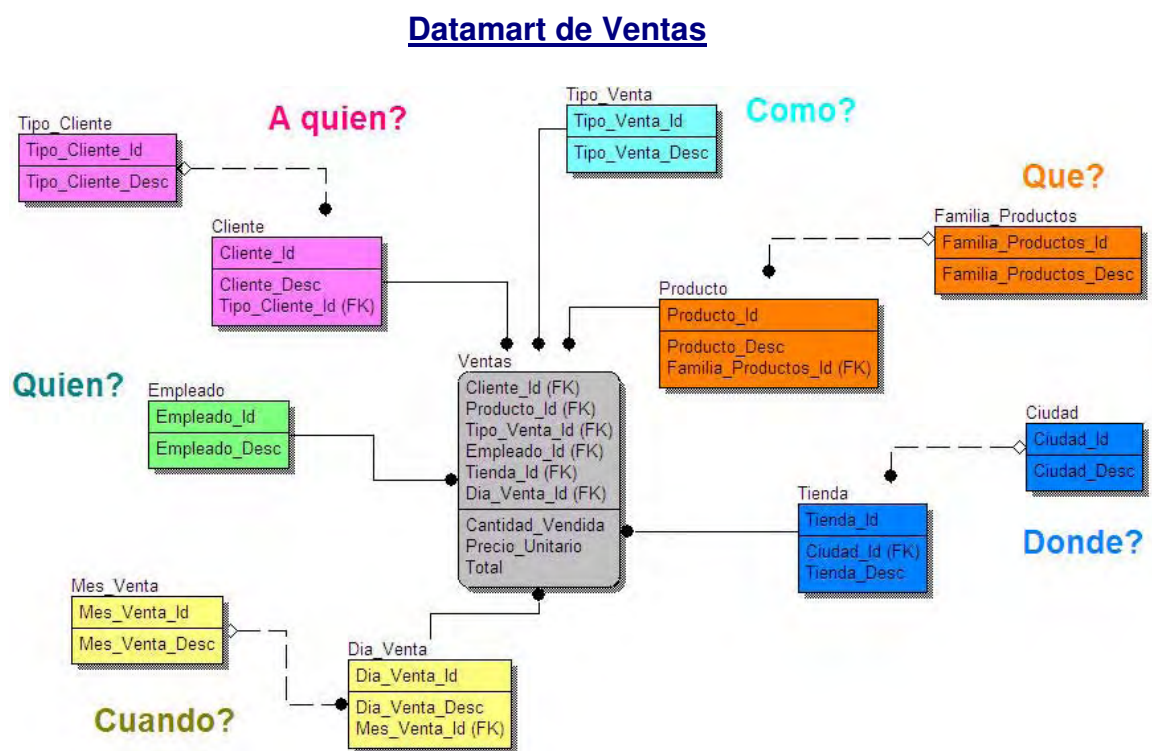


Figura 3.21: Datamart de Ventas

El esquema empleado como Copo de Nieve, Estrella o cualquiera de los derivados como constelación de estrellas, tormenta de nieve, etc dependerá de la herramienta de explotación que estén utilizando.

Pueden existir más de una dimensión que respondan a una de las preguntas, por ejemplo:

Quién Vende: El Empleado o El Distribuidor

Paso 4: Elaboración del Modelo Multidimensional Complejo

En esta etapa se realiza el proceso de calificación a los indicadores y a los atributos.

Tipo	Descripción	Ejemplo
Bandas, Rangos	Califica a un indicador	[0 – 500> [500 – 1000>
Grupos Personalizados y Consolidados	Califica a un atributo	Tipo Producto

Tabla 3.2: Modelo Multidimensional Complejo

En un modelo multidimensional complejo los usuarios podrán segmentar a los clientes, productos o cualquier otro atributo en forma fácil y sencilla, pudiendo diferenciarlos según cuanto consumen y/o como consumen.

Por ejemplo:

Si queremos segmentar a productos según la rotación de los últimos 6 meses, se puede crear un grupo personalizado llamado: Calificación de productos en el que se especifica si tiene alta, mediana o baja rotación.

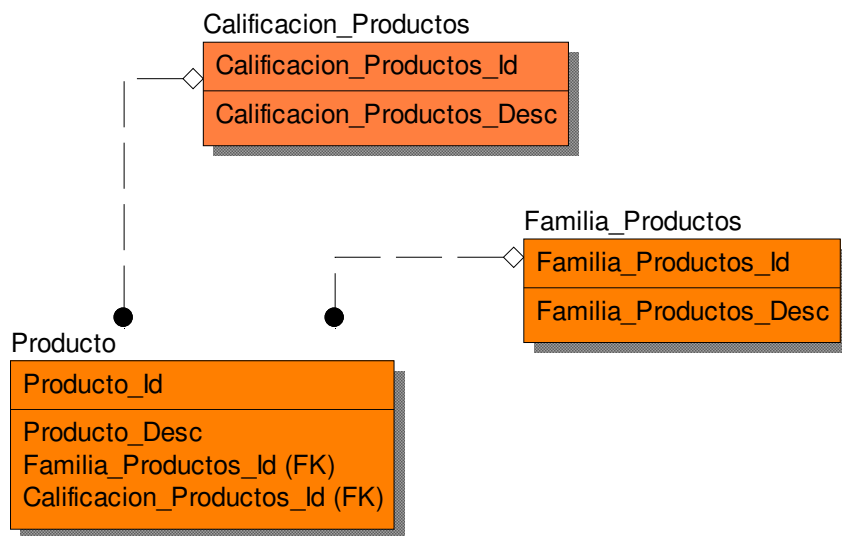


Figura 3.22: Calificación de Productos

Paso 5: Elaboración de las especificaciones de carga de datos.

Este es el momento donde se trata de buscar la información en las diferentes fuentes de datos de la organización para cargar el modelo multidimensional creado.

Paso 6: Creación de Base de Datos.

Se debe crear una base de datos que contendrá la información del datamart que será explotada.

En el caso que no exista una metadata se deberá crear una base de datos en blanco con las características y especificaciones técnicas de la herramienta que será utilizada para la explotación de los datos.

Paso 7: Construcción de la Arquitectura del Datamart.

Este es el momento donde se construyen la arquitectura del datamart, que en el caso de las herramientas MOLAP serian los cubos de información.

Paso 8: ETL

Consiste en extraer, transformar y cargar los datos de los sistemas fuentes hacia la base de datos del **datamart**.

Los programas de ETL deben cumplir con las especificaciones que se desarrollaron en el paso 5, con la finalidad de llevar una correcta documentación del proyecto.

Los programas de cargas deben verificar los errores de integridad referencial y limpiarla en el caso que se detecte alguna ocurrencia.

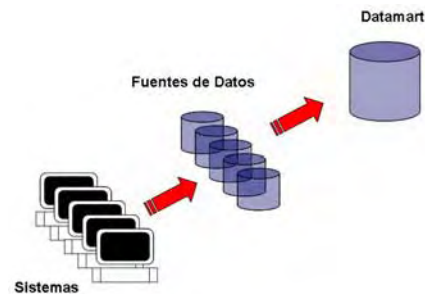


Figura 3.23: ETL - Datamart

Una mala planificación o construcción de los programas de ETL pueden ocasionar que los usuarios dejen de usar el Datamart, por ejemplo:

- ❖ La información está inconsistente.
- ❖ Sólo se tiene cargado pocos periodos debido a la falta de espacio en el disco.

- ❖ La carga se paralizó debido a que uno de los programas identificó que existe datos inconsistentes.
- ❖ Los programas no se ejecutaron en el momento que se requerían.
- ❖ No se puede reprocesar la información y se mantienen datos no certeros.

Paso 9: Implementación

un motivo relevante por el cual los usuarios no utilizan los datamart es por falta de capacitación

Paso 10: Capacitación al Usuario

Otro de los principales motivos por el cual los usuarios no utilizan los datamart es por falta de capacitación.

La capacitación que debe recibir el usuario debe estar enfocada en:

- a. **El Modelo Multidimensional.**- Es importante que los usuarios conozcan los diferentes modelo multidimensionales de la empresa.
- b. **La Herramienta de explotación.**- Se dice que los usuarios solo utilizan el 20% de las opciones que cuentan las herramientas de explotación por falta de capacitación.

- c. **Las herramientas de gestión.**- Los usuarios deben ser constantemente capacitados en herramientas de gestión como creación de dashboard, scorecard, etc

Ventajas:

- ✓ Se puede contar con él en menor tiempo.
- ✓ Permite probar en un área específica de la empresa, para determinar si es factible o no implementar un Data warehouse.
- ✓ Permite diferir los altos costos del Data Warehouse durante el tiempo en que se vayan implementando los DataMart.

Desventajas:

- ✓ ·De esta forma, posterga aún más la implementación del Data Warehouse.
- ✓ ·Al enfocarse a un área específica, no logra cumplir un papel de anticipador de necesidades.

3.2.1.3. OLAP (Online Analytical Processing)

OLAP (Proceso Analítico en Línea) desarrolla un análisis multidimensional de los datos, que consiste en organizar la información según los parámetros que

la alta dirección considere oportuno para darles un sentido y llevar a cabo el análisis de los mismos. Para un concepto más formal de OLAP, citamos a dos autores, quienes definen OLAP como:

“Es una categoría de tecnología de software que permite a los ejecutivos y analistas, una vista rápida, consistente, con un acceso interactivo a una completa variedad de vistas posibles de la información que ha sido transformada de los registros de una empresa”.

“Se trata de un término inventado para describir una aproximación dimensional interactiva al soporte de toma de decisiones (análisis desde la perspectiva de sus componentes o dimensiones, contemplando también los distintos niveles o jerarquías que éstas poseen): Proceso, Dimensiones, Interacción, Análisis, Toma de decisiones. Este tipo de Análisis, se soporta mediante una visión dimensional: Modelo Multidimensional o Cubo”.

Como se describió comúnmente los sistemas OLAP ofrecen a los empleados calificados una estructura de datos conocida como un “cubo”, que virtualmente contiene niveles de reportes resumidos actualizados. Este cubo es multidimensional y ofrece a los usuarios finales diferentes alternativas y perspectivas de ver los datos, a partir del mismo cubo.

Por ejemplo, un cubo que contiene datos de ventas puede producir reportes tan diversos como: ventas totales por trimestre, porcentaje de ventas por vendedor, ventas promedio por región geográfica, etc.

Las herramientas OLAP se caracterizan por una rápida recuperación de los datos, ya que la recuperación se puede realizar en una sola entrada y salida, los

índices son pequeños y los datos se encuentran modelados en un esquema multidimensional. Además, el almacenamiento es muy eficiente, ya que los bloques solo contienen datos y los índices son simples.

HERRAMIENTAS OLAP

Las herramientas de OLAP presentan al usuario una visión multidimensional de los datos (esquema multidimensional) para cada actividad que es objeto de análisis. El usuario formula consultas a la herramienta OLAP seleccionando atributos de este esquema multidimensional sin conocer la estructura interna (esquema físico) del almacén de datos. La herramienta OLAP genera la correspondiente consulta y la envía al gestor de consultas del sistema. Una consulta a un Data Warehouse consiste generalmente en la obtención de medidas sobre los hechos parametrizados por atributos de las dimensiones y restringidas por condiciones impuestas sobre las dimensiones.

Utilizando herramientas OLAP, los usuarios pueden acceder al DW, brindando a los responsables de las tomas de decisiones de las organizaciones, el potencial de mejorar su comprensión del negocio y los cambios que lo afectan, de incrementar su habilidad para identificar o generar soluciones posibles a problemas de decisión, y de efectuar oportunamente formulaciones tácticas o estratégicas alineadas con los objetivos de la organización. Esta característica de analizar y sintetizar la información a partir de OLAP, surge de la elaboración

de múltiples escenarios que contestan preguntas tales como “qué pasaría si...” o “por qué...”.

En estos verdaderos modelos de simulación, y a partir de la modificación de ciertas variables clave, se analiza el comportamiento del resto de las variables.

Lo realmente interesante de las herramientas OLAP son sus operadores de refinamiento o manipulación de consultas:

- Drill
- Roll
- Slice & Dice
- Pívor

El carácter agregado de las consultas en el Análisis de Datos, aconseja la definición de nuevos operadores que faciliten la agregación (consolidación) y la disgregación (división) de los datos:

Roll (Agregación): permite eliminar un criterio de agrupación en el análisis, agregando los grupos actuales.

Drill (Disgregación): permite introducir un nuevo criterio de agrupación en el análisis, disgregando los grupos actuales.

Las operaciones de agregación (roll) y disgregación (drill) se pueden hacer sobre atributos de una dimensión sobre los que se ha definido una jerarquía:

(drill-down, roll-up) como por ejemplo: Departamento - Categoría - Producto (Producto), Año - Trimestre - Mes – Día (Tiempo). O sobre dimensiones independientes (drill-across, roll-across). Por ejemplo: Producto - Almacén – Tiempo.

Slice & Dice: permite seleccionar y proyectar datos en el informe.

Pivot: permite la reorientación de las dimensiones en el informe.

CARACTERÍSTICAS DE LAS HERRAMIENTAS OLAP

En 1994, Codd y Codd introdujeron reglas sobre el modelo OLAP, pensadas para eliminar interpretaciones erróneas sobre las características de esta tecnología. En forma resumida, las doce reglas de Codd para los sistemas OLAP son las siguientes:

1. Vistas Multidimensionales. Manejo y organización conceptual y física de la información en forma multidimensional.
2. Transparencia. Capacidad para acceder a datos de otras fuentes de manera sencilla y transparente.
3. Accesibilidad. Habilidad para obtener información completa y estructurada de fuentes externas de datos, tales como bases de datos relacionales, archivos planos, etc.

4. Desempeño y consistencia. El número de dimensiones utilizadas en el sistema no debe degradar el desempeño del sistema, ni tampoco afectar la consistencia de la información.

5. Cliente/servidor. Las herramientas deben poder operar en ambientes cliente/servidor.

6. Dimensionalidad genérica. Cada dimensión deberá ser tratada de igual manera.

7. Uso eficiente del almacenamiento. Manejo eficiente de la porosidad (sparseness) de la base multidimensional, para ocupar la misma cantidad de espacio. Por porosidad se entiende la manera en la que herramienta maneja el espacio requerido para almacenar la información multidimensional; este punto es muy importante ya que, debido a la estructura de los datos en las bases multidimensionales se cuenta con muchas celdas o campos vacíos. Un buen manejo de la porosidad implica que la herramienta es capaz de detectar las celdas vacías, y hacer eficiente el espacio que éstos requieren.

8. Soporte o múltiples usuarios. Permitir el acceso de múltiples usuarios al mismo tiempo al modelo.

9. Operaciones entre dimensiones sin límite. Capacidad para realizar operaciones entre varias dimensiones sin ningún tipo de restricción.

10. Manipulación intuitiva de datos. Capacidad de navegación a través de los datos, dimensiones y jerarquías de la base mediante una interfaz de usuario fácil de usar.

11. Producción flexible de reportes. Utilitarios para la creación rápida de reportes, consultas y gráficos.

12. Capacidad ilimitada para dimensiones y relaciones (jerarquías). Capacidad para manejar un número ilimitado de jerarquías, relaciones y dimensiones de datos.

Lo interesante no es poder realizar consultas que, en cierto modo, se pueden hacer con selecciones, proyecciones, concatenaciones y agrupamientos tradicionales.

ESTRATEGIAS DE ALMACENAMIENTO.

OLAP soporta tres tipos de almacenamiento:

MOLAP

En el modo de almacenamiento MOLAP (**OLAP Multidimensional**) una copia de los datos de origen del cubo, junto con sus agregaciones, es almacenada en una estructura multidimensional.

Debemos tener en cuenta que mientras los datos de origen cambian directamente con las operaciones, los objetos con almacenamiento MOLAP deben ser procesados para incorporar estos cambios.

El tiempo comprendido entre un procesamiento y el siguiente, crea un periodo de **latencia** durante el que puede que la información OLAP no coincida con los datos de origen actuales.

Como característica del almacenamiento MOLAP podemos destacar:

- Provee excelente rendimiento y compresión de datos.
- Tiene mejor tiempo de respuesta, dependiendo sólo del porcentaje de las agregaciones del cubo.
- La estructura está muy optimizada para maximizar el rendimiento de las consultas.
- En general este método, es muy apropiado para cubos con uso frecuente por su rápida respuesta.

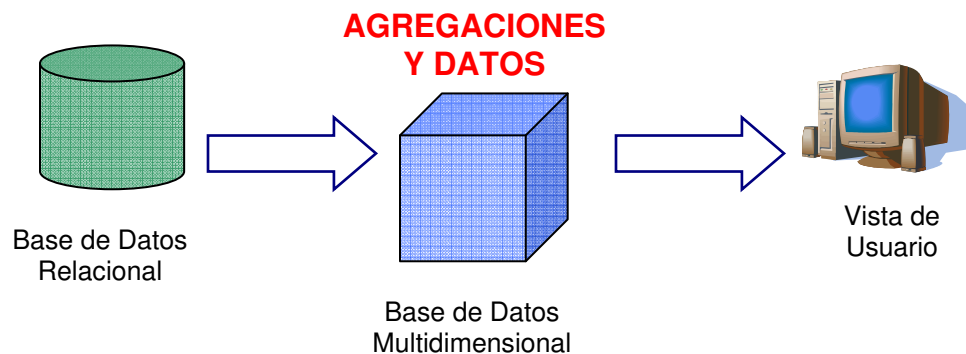


Figura 3.24: MOLAP

ROLAP

En un modelo ROLAP (**OLAP Relacional**) toda la información del cubo, sus datos, su agregación, sumas, etc., son almacenados en una base de datos relacional.

A diferencia del modo de almacenamiento MOLAP, ROLAP no almacena copia de la base de datos, accede a las tablas originales cuando necesita responder a las consultas, generalmente es mucho más lenta que las otras estrategias de almacenamiento (MOLAP o HOLAP).

ROLAP se utiliza para ahorrar espacio de almacenamiento cuando se trabaja con grandes conjuntos de datos que se consultan con poca frecuencia; por ejemplo, datos exclusivamente históricos.

Los usos comunes de este esquema son:

- Cuando los clientes desean ver los cambios inmediatamente.
- Cuando contamos con grandes conjuntos de datos que **no** son frecuentemente buscados

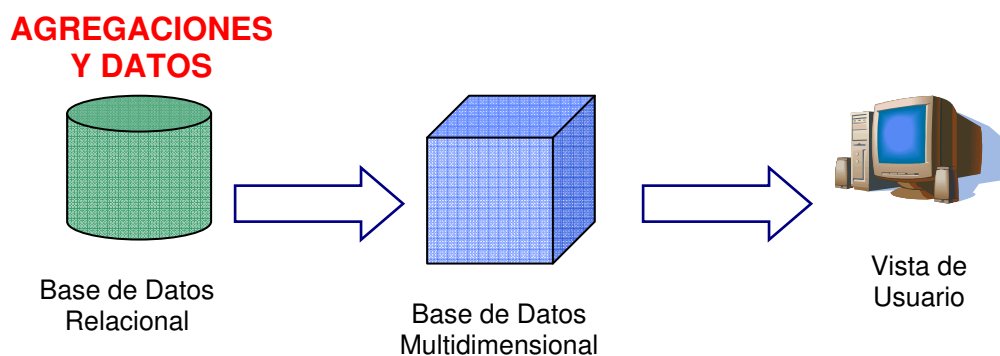


Figura 3.25: ROLAP

HOLAP

HOLAP (**OLAP híbrido**) combina atributos de MOLAP y ROLAP.

Al igual que MOLAP, HOLAP hace que las agregaciones se almacenen en una estructura multidimensional, y los datos a nivel de detalle, en una base de datos relacional como lo hace el almacenamiento ROLAP.

Para procedimientos de búsqueda que accedan datos sumariados, HOLAP es equivalente a MOLAP. Por el contrario, si los procesos de consultas accedieran a los máximos niveles de detalle, deberían recuperar los datos de la base de datos relacional y esto no sería tan rápido comparado con una estructura MOLAP.

Los cubos almacenados como HOLAP, son más pequeños que los MOLAP y responden más rápidos que los ROLAP.

Usos comunes de HOLAP

- ↗ Cubos que requieren rápida respuesta
- ↗ Cuando existen sumalizaciones basadas en una gran cantidad de datos de origen.
- ↗ Solución de compromiso para bajar el espacio ocupado sin perjudicar totalmente el rendimiento de las consultas.

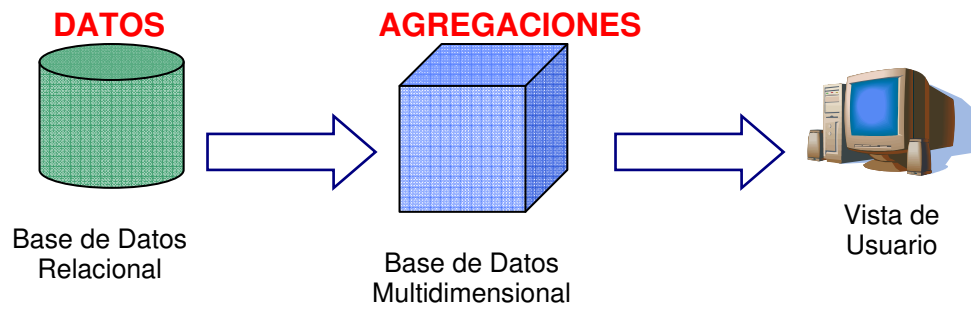


Figura 3.26: HOLAP

Debemos tener en cuenta que si los usuarios generan consultas que deben utilizar los datos del nivel mas bajo HOLAP no suele ser la mejor opción

	MOLAP	ROLAP	HOLAP
Almacenamiento de las Agregaciones	Modelo Multidimensional	Base de datos relacional	Modelo Multidimensional

Almacenamiento de los datos	Modelo Multidimensional	Base de datos relacional	Base de datos relacional
Facilidad de Creación	Sencillo	Muy Sencillo	Sencillo
Velocidad de respuesta	Buena	Regular o Baja	Buena para consultas que posean agregaciones, Regular para datos de bajo nivel
Escalabilidad	Problemas de escalabilidad	Son más escalables	
Recomendados para	Cubos con uso frecuente	Datos que no son frecuentemente usados	Si el cubo requiere una rápida respuesta

Tabla 3.3. MOLAP; ROLAP y HOLAP

➡ MOLAP es un OLAP basado en el acceso a una base de datos multidimensional

➡ ROLAP es un OLAP basado en el acceso a una base de datos relacional

HOLAP es un OLAP situado entre ROLAP y MOLAP, accede a la Multidimensional y a la Relacional.

3.2.1.4. DATAMINING.

Data Mining, la extracción de información oculta y predecible de grandes bases de datos, es una tecnología para ayudar a las compañías a descubrir información relevante en sus bases de información (datawarehouses). Las herramientas de Data Mining predicen futuras tendencias y comportamientos. Los análisis prospectivos automatizados ofrecidos por la automatización del Data Mining van más allá de los eventos pasados provistos por las herramientas usuales de sistemas de soporte de decisión.

Las herramientas de Data Mining pueden responder a preguntas de negocios que tradicionalmente consumen demasiado tiempo para poder ser resueltas. Estas herramientas exploran las bases de datos en busca de patrones ocultos, encontrando información predecible que un experto no puede llegar a encontrar.

Muchas compañías ya coleccionan y refinan cantidades masivas de datos. Las técnicas de Data Mining pueden ser implementadas rápidamente en plataformas ya existentes de software y hardware para acrecentar el valor de las fuentes de información existentes y pueden ser integradas con nuevos productos y sistemas pues son traídas en línea (on-line). Una vez que las herramientas de Data Mining fueron implementadas en computadoras cliente servidor de alta performance o de procesamiento paralelo, pueden analizar bases de datos masivas para brindar respuesta a preguntas tales como, "¿Cuáles clientes tienen más probabilidad de responder al próximo mailing promocional, y por qué? y

presentar los resultados en formas de tablas, con gráficos, reportes, texto o hipertexto.

LOS FUNDAMENTOS DEL DATA MINING

Las técnicas de Data Mining son el resultado de un largo proceso de investigación y desarrollo de productos. Esta evolución comenzó cuando los datos de negocios fueron almacenados por primera vez en computadoras, continuó con mejoras en el acceso a los datos, y más recientemente con tecnologías generadas para permitir a los usuarios navegar a través de los datos en tiempo real.

Data Mining toma este proceso de evolución más allá del acceso y navegación retrospectiva de los datos, hacia la entrega de información prospectiva y proactiva. Data Mining está listo para su aplicación en la comunidad de negocios porque está soportado por tres tecnologías que ya están suficientemente maduras:

- Recolección masiva de datos
- Potentes computadoras (algunas con multiprocesadores)
- Algoritmos de Data Mining

Los algoritmos de Data Mining utilizan técnicas que han existido por lo menos desde hace 10 años, pero que sólo han sido implementadas recientemente como herramientas maduras y confiables.

En la evolución desde los datos de negocios a información de negocios, cada nuevo paso se basa en el previo. Por ejemplo, el acceso a datos dinámicos es crítico para las aplicaciones de navegación de datos (OLAP), y la habilidad para almacenar grandes bases de datos (Datawarehouse) es crítica para Data Mining. Los componentes esenciales de la tecnología de Data Mining han estado bajo desarrollo por décadas, en áreas de investigación como estadísticas, inteligencia artificial y aprendizaje de máquinas. Hoy, la madurez de estas técnicas, junto con los motores de bases de datos relacionales de alta performance, hicieron que estas tecnologías fueran prácticas para los entornos de datawarehouse actuales.

EL ALCANCE DE DATA MINING

El nombre de Data Mining deriva de las similitudes entre buscar información de negocios en grandes bases de datos, encontrar información de la venta de un producto entre grandes montos de Gigabytes almacenados y minar una montaña para encontrar una veta de metales valiosos. Ambos procesos requieren examinar una inmensa cantidad de material, o investigar inteligentemente hasta encontrar exactamente donde residen los valores. Dadas bases de datos de suficiente tamaño y calidad, la tecnología de Data Mining puede generar nueva información al proveer las siguientes capacidades:

- **Predicción automatizada de tendencias y comportamientos.** Data Mining automatiza el proceso de encontrar información predecible en grandes bases de datos. Preguntas que tradicionalmente requerían un intenso análisis manual, ahora pueden ser contestadas directa y rápidamente desde los datos.

- **Descubrimiento automatizado de modelos previamente desconocidos.**

Las herramientas de Data Mining barren las bases de datos e identifican modelos previamente escondidos en un sólo paso. Otros problemas de descubrimiento de modelos incluye detectar transacciones fraudulentas de tarjetas de créditos e identificar datos anormales que pueden representar errores de tipeado en la carga de datos.

Cuando las herramientas de Data Mining son implementadas en sistemas de procesamiento paralelo de alta performance, pueden analizar bases de datos masivas en minutos. Procesamiento más rápido significa que los usuarios pueden automáticamente experimentar con más modelos para entender datos complejos. La alta velocidad de procesamiento junto a las técnicas de Data Mining hace que sea práctico para los usuarios analizar inmensas cantidades de datos. Grandes bases de datos, a su vez, producen mejores predicciones.

Las bases de datos pueden ser grandes tanto en profundidad como en ancho:

Más columnas: los analistas muchas veces deben limitar el número de variables a examinar cuando realizan análisis manuales debido a limitaciones de tiempo. Sin embargo, variables que son descartadas porque parecen sin importancia pueden proveer información acerca de modelos desconocidos. Un Data Mining de alto rendimiento permite a los usuarios explorar toda la base de datos, sin preseleccionar un subconjunto de variables.

Más filas: muestras mayores producen menos errores de estimación y desvíos, y permite a los usuarios hacer inferencias acerca de pequeños pero importantes segmentos de población.

Las técnicas más comúnmente usadas en Data Mining son:

- **Redes neuronales artificiales:** modelos predecibles no-lineales que aprenden a través del entrenamiento y semejan la estructura de una red Neuronal biológica.
- **Árboles de decisión:** estructuras de forma de árbol que representan conjuntos de decisiones. Estas decisiones generan reglas para la clasificación de un conjunto de datos. Métodos específicos de árboles de decisión incluyen Árboles de Clasificación y Regresión (CART:Classification And Regression Tree) y Detección de Interacción

Automática de Chi Cuadrado (CHAI: Chi Square Automatic Interaction Detection)

- **Algoritmos genéticos:** técnicas de optimización que usan procesos tales como combinaciones genéticas, mutaciones y selección natural en un diseño basado en los conceptos de evolución.

- **Método del vecino más cercano:** una técnica que clasifica cada registro en un conjunto de datos basado en una combinación de las clases del/de los k registro (s) más similar/es a él en un conjunto de datos históricos (donde $k > 1$). Algunas veces se llama la técnica del vecino k-más cercano.

- **Regla de inducción:** la extracción de reglas if-then de datos basados en significado estadístico.

ARQUITECTURA DATA MINING

Para aplicar mejor estas técnicas, deben estar totalmente integradas con el data warehouse así como con herramientas flexibles e interactivas para el análisis de negocios (herramientas OLAP). Varias herramientas de Data Mining actualmente operan fuera del warehouse, requiriendo pasos extra para extraer,

importar y analizar los datos. Además, cuando nuevos conceptos requieren implementación operacional, la integración con el warehouse simplifica la aplicación de los resultados desde Data Mining.

El punto de inicio ideal es un datawarehouse. Este datawarehouse puede ser implementado en una variedad de sistemas de bases relacionales y debe ser optimizado para un acceso a los datos flexible y rápido.

Un server OLAP permite que el usuario analice los datos de acuerdo a como quiera mirar el negocio, resumido por línea de producto, u otras perspectivas claves para su negocio. El server de Data Mining debe estar integrado con el data warehouse y el server OLAP para insertar el análisis de negocios directamente en esta infraestructura. A medida que el data warehouse crece, la organización puede aplicar extraer la información oculta y aplicarla en futuras decisiones.

IMPLEMENTACIÓN DE DATA MINING

¿Cuán exactamente es capaz Data Mining de decir cosas importantes que se desconocen o que van a pasar? La técnica usada para realizar estas predicciones en Data Mining se llama Modelado. Modelado en Data Mining es simplemente el acto de construir un modelo en una situación en donde se conoce la respuesta y luego se aplica en otra situación de la cual se desconoce la respuesta.

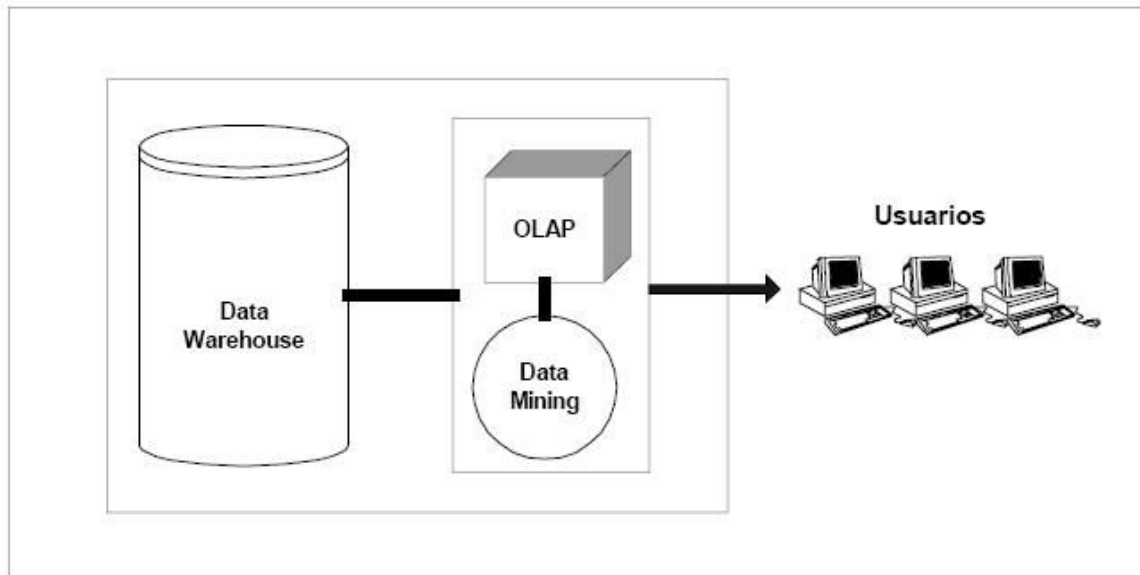


Figura 3.27: Arquitectura básica para Data Mining.

Este acto de construcción de un modelo es algo que la gente ha estado haciendo desde hace mucho tiempo, seguramente desde antes del auge de las computadoras y de la tecnología de Data Mining. Lo que ocurre en las computadoras, no es muy diferente de la manera en que la gente construye modelos. Las computadoras son cargadas con mucha información acerca de una variedad de situaciones donde una respuesta es conocida y luego el software de Data Mining en la computadora debe correr a través de los datos y distinguir las características de los datos que llevarán al modelo. Una vez que el modelo se construyó, puede ser usado en situaciones similares donde no conoce la respuesta. Si alguien que tiene un modelo que puede predecir el comportamiento de los clientes, ¿cómo puede saber si es realmente un buen modelo? La primera cosa que puede probar es que aplique el modelo a su base de clientes conocidos y usuales donde ya se conoce la respuesta. Con Data

Mining, la mejor manera para realizar esto es dejando de lado ciertos datos para aislarlos del proceso de Data Mining. Una vez que el proceso está completo, los resultados pueden ser testeados contra los datos excluidos para confirmar la validez del modelo. Si el modelo funciona, las observaciones deben mantenerse para los datos excluidos.

Entonces, los pasos típicos para realizar Data Mining son los siguientes:

- **Definición del problema:** de la misma manera que en un análisis tradicional, antes de iniciar un proceso de data mining debemos tener muy claro el problema que queremos resolver.
- **Recopilación y preparación de datos:** los datos originales de las BD transaccionales no están preparados para el análisis y, a veces, es necesario aplicar modificaciones, crear agregados y diseñar estructuras nuevas. Además, muchos métodos de data mining necesitan los datos en un formato específico.
- **Data mining:** consiste en construir un modelo sobre los datos con capacidad predictiva y/o descriptiva, de manera que pueda utilizarse para resolver el problema planteado. Para ello, se emplean técnicas estadísticas o de Inteligencia Artificial.

- **Validación:** después de construir el modelo, éste se debe validar antes de utilizarse. La validación puede ser de carácter técnico (utilizar muestras adicionales de datos para comprobar la capacidad predictiva o descriptiva del modelo) o conceptual (ver si la interpretación es satisfactoria, si el resultado es aplicable). Si el modelo no puede validarse, es posible que sea necesario aplicar de nuevo el método de data mining o modificar los datos.

- **Aplicación:** una vez validado, el modelo debe implementarse en el proceso que se desea mejorar. Dependiendo del proceso, esta implementación puede ser más o menos directa y requerir más o menos tiempo.

- **Monitorización:** debe existir un seguimiento de la implementación del modelo en el proceso que se desea mejorar para comprobar sus resultados reales. Si el resultado no es bueno, es posible que haya que redefinir los objetivos. Y puede que, aún siendo óptimo, sugiera nuevos objetivos que se pueden alcanzar.

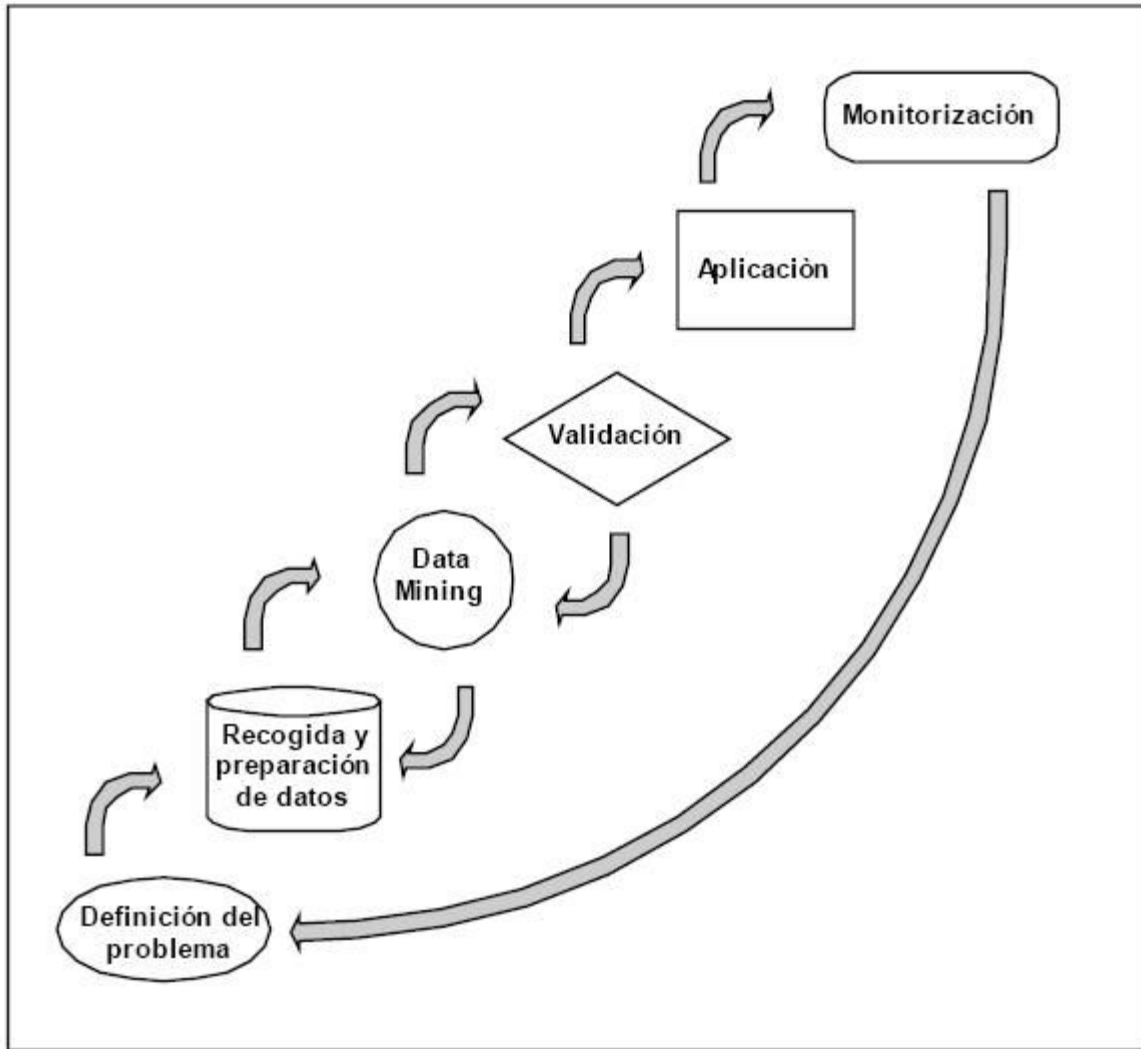


Figura 3.28: Etapas de construcción del Data Mining.

Ventajas:

- ✓ Permite obtener una imagen detallada de la información solicitada.
- ✓ Se puede llegar al conjunto de transacciones que generaron la información.

Desventajas:

- ✓ Puede colapsar el ancho de banda.
- ✓ Puede disminuir la capacidad de procesamiento.
- ✓ Puede empeorar el tiempo de respuesta.

3.2.1.5. METODOLOGÍA PARA EL DESARROLLO DE DATAMART**METODOLOGIA DE RALPH KIMBALL**

Ralph Kimball, es reconocido como uno de los padres del concepto de Datawarehouse, quien se ha dedicado desde hace más de 10 años al desarrollo de su metodología para que éste concepto sea bien aplicado en las organizaciones y se asegure la calidad en el desarrollo de estos proyectos.

La metodología de Ralph Kimball se enfoca principalmente en el diseño de la base de datos que almacenará la información para la toma de decisiones.

El diseño se basa en la creación de tablas de hechos, es decir, tablas que contengan la información numérica de los indicadores a analizar, o sea la parte cuantitativa de la información para la toma de decisiones.

Las tablas anteriores se relacionan con tablas de dimensiones, las cuales contienen la información cualitativa, de los indicadores, es decir, toda aquella información que clasifique la información requerida.

A éste modelo de datos se le conoce como "diseño estrella", existen variaciones de éste, llamados "copo de nieve" y "diseño flat". Todos estos diseños tienen la característica de preparar la información de acuerdo a la necesidad de tomar decisiones y no a los argumentos técnicos de espacio de almacenamiento.

Por lo tanto tomaremos como base el ciclo de vida de los Data Warehouses definido por Ralph Kimball. El marco presentado por Ralph Kimball con el nombre de Business Dimensional Lifecycle (BDL) ilustra las diferentes etapas por las que debe pasar todo proceso de Data Warehousing. Este enfoque de implementación de data warehouses es ilustrado en la siguiente figura. Este diagrama ilustra la secuencialidad de tareas de alto nivel requeridas para el efectivo diseño, desarrollo e implementación de data warehouses. El diagrama muestra una vista general del mapa de ruta de un proyecto en el cual cada rectángulo nos indica en que etapa estamos, por dónde pasamos y hacia dónde debemos dirigirnos.

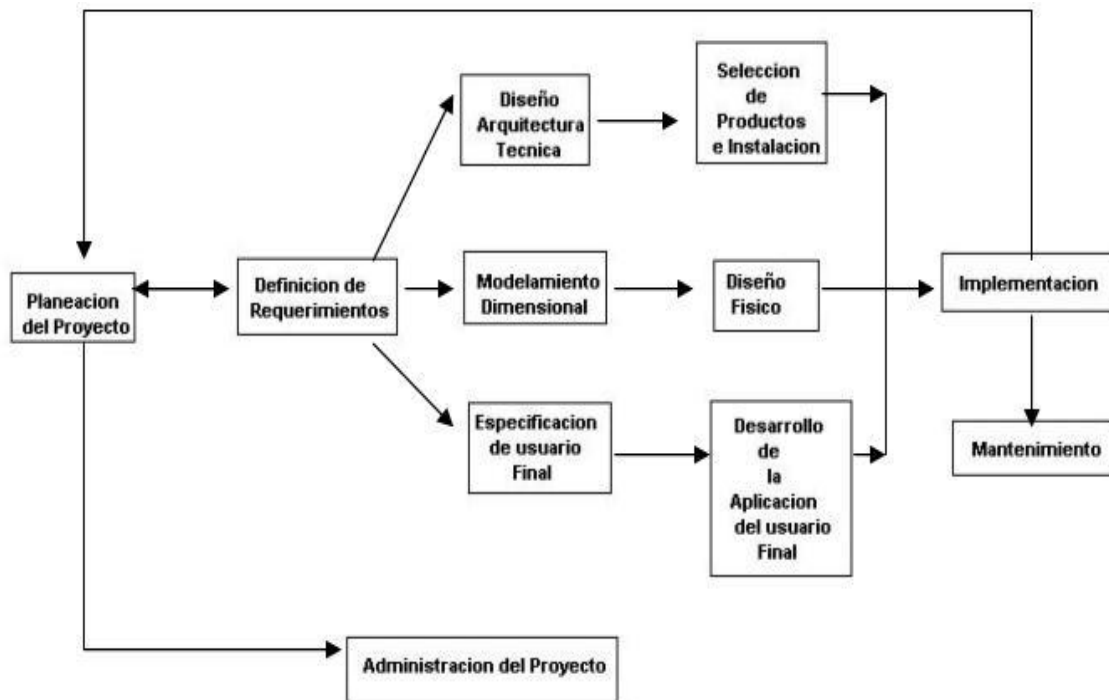


Figura 3.29 Fases de Metodología Kimball

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

A continuación pasaremos a describir cada una de las etapas del BDL.

PLANIFICACIÓN DEL PROYECTO

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

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

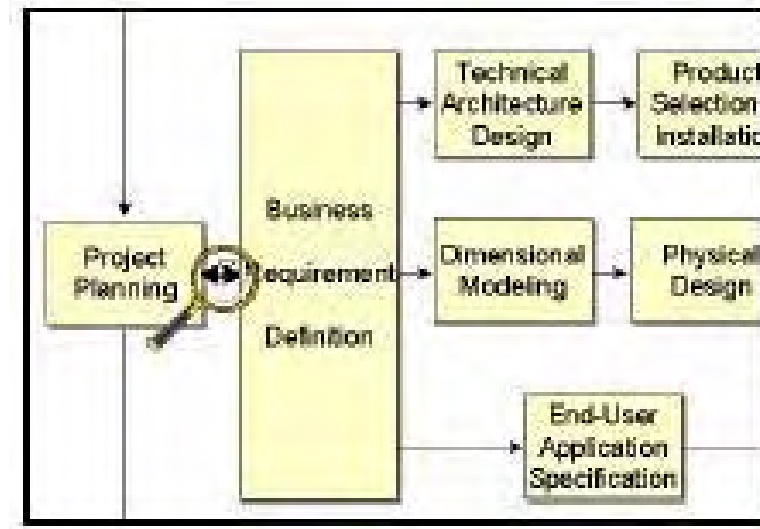


Figura 3.30: Iteración entre Planificación y requerimientos

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

Esta etapa se concentra sobre la definición del proyecto (identificación del escenario del proyecto para saber de dónde surge la necesidad del data warehouse). Según sentencia Kimball, “Antes de comenzar un proyecto de data warehouse o data mart, hay que estar seguro si existe la demanda y de dónde proviene. Si no se tiene un sólido usuario sponsor y no hay usuarios entusiasmados, posponga el proyecto”. Factores asociados con estas etapas incluyen: identificación de los usuarios sponsors, convincentes motivaciones del negocio, cooperación entre áreas de sistemas y negocios, cultura analítica de la organización y análisis de factibilidad (tanto tecnológica como de disponibilidad de datos). Para medir estos factores propone un test de buena disposición del proyecto donde describe diferentes escenarios posibles [Kim98]. Adicionalmente propone técnicas (relevamientos de alto nivel, priorización de requerimientos y pruebas de concepto) para mitigar las deficiencias que el proyecto pudiera tener en algunos de los factores mencionados anteriormente.

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

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

DEFINICIÓN DE LOS REQUERIMIENTOS DEL NEGOCIO

Un factor determinante en el éxito de un proceso de Data Warehousing es la interpretación correcta de los diferentes niveles de requerimientos expresados por los diferentes niveles de usuarios. La técnica utilizada para relevar los requerimientos de los analistas del negocio difiere de los enfoques tradicionales guiados por los datos. Los diseñadores de los data warehouses deben entender los factores claves que guían al negocio para determinar efectivamente los requerimientos y traducirlos en consideraciones de diseño apropiadas.

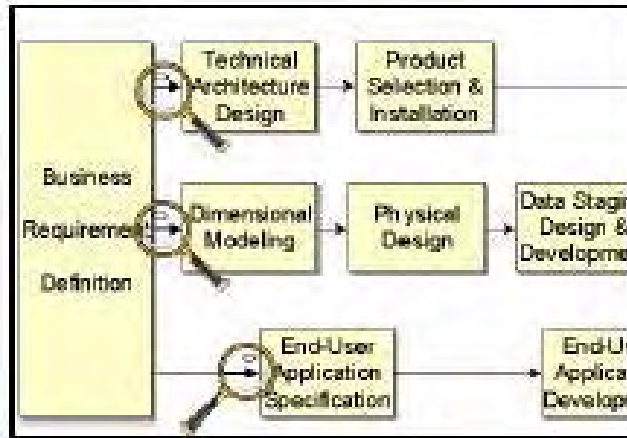


Figura 3.31: Iteración entre Requerimientos del Negocio y etapas subsiguientes

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

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

Kimball da consejos y técnicas para descubrir eficazmente los requerimientos del negocio. Estas tácticas y estrategias se focalizan sobre

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

MODELADO DIMENSIONAL

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

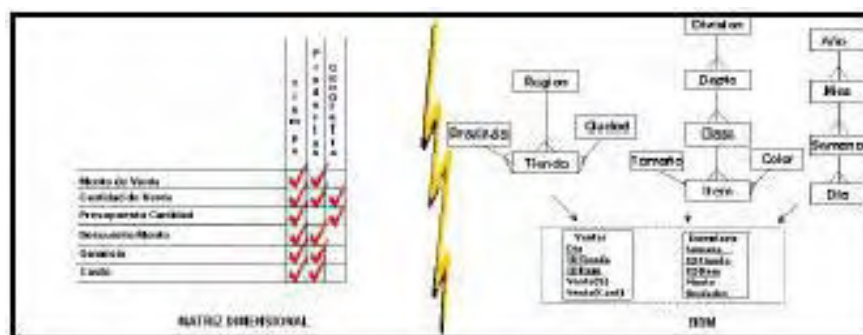


Figura 3.32: Diagramas para Modelado Dimensional

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

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

DISEÑO FÍSICO

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



Figura 3.33: Proceso de diseño físico a alto nivel

DISEÑO Y DESARROLLO DE PRESENTACIÓN DE DATOS

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

Todas estas tareas son altamente críticas pues tienen que ver con la materia prima del data warehouse: los datos. La desconfianza y pérdida

de credibilidad del data warehouse serán resultados inmediatos e inevitables si el usuario choca con información inconsistente. Es por ello que la calidad de los datos es un factor determinante en el éxito de un proyecto de data warehousing. Es en esta etapa donde deben sanearse todos los inconvenientes relacionados con la calidad de los datos fuente.

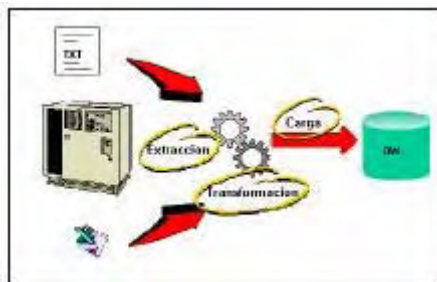


Figura 3.34: Proceso ETL

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

Plan:

1. Crear un diagrama de flujo fuente-destino esquemático, de una página y de muy alto nivel.
2. Probar, elegir e implementar una herramienta de data staging.

3. Profundizar en detalle por tabla destino, gráficamente describir las reestructuraciones o transformaciones complejas. Gráficamente ilustrar la generación de las claves surrogadas. Desarrollo preliminar de la secuencialidad de los trabajos.

Carga de dimensiones:

4. Construir y probar la carga de una tabla dimensional estática. La principal meta de este paso es resolver los problemas de infraestructura que pudieran surgir (conectividad, transferencia, seguridad, etc.)

5. Construir y probar los procesos de actualización de una dimensión.

6. Construir y probar las cargas de las restantes dimensiones.

Fact Tables y automatización:

7. Construir y probar la carga histórica de las fact tables (carga masiva de datos). Incluyendo búsqueda y sustitución de claves.

8. Construir y probar los procesos de cargas incrementales.

9. Construir y probar la generación de agregaciones.

10. Diseñar, construir y probar la automatización de los procesos.

planos sirven para comunicar los deseos entre los clientes y el arquitecto, como así también para medir esfuerzos y materiales necesarios para la obra (comunicación, planificación, flexibilidad y mantenimiento, documentación, productividad y reuso). Finalmente, argumenta Kimball, “un buen conjunto de planos, como cualquier buena documentación, nos ayudará más tarde cuando sea tiempo de remodelar o hacer incorporaciones”.

SELECCIÓN DE PRODUCTOS E INSTALACIÓN

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

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

Característica	Producto A	Producto B	Producto C	...
Capacidades Básicas de ETL				
Extracción de datos	85			
Transformación de datos	78			
Job Control & Scheduling				
Control de ejecución	80			
Planificación de jobs	85			
Metadatos & Excepciones				
Extracción de metadatos				
Transformación de metadatos				
Interfaz de Usuario				
Interfaz de usuario	70			
Reportes	85			
Puntuación Total				
Puntuación				

Figura 3.36: Ejemplo de matriz de evaluación de productos

ESPECIFICACIÓN DE APLICACIONES PARA USUARIOS FINALES

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

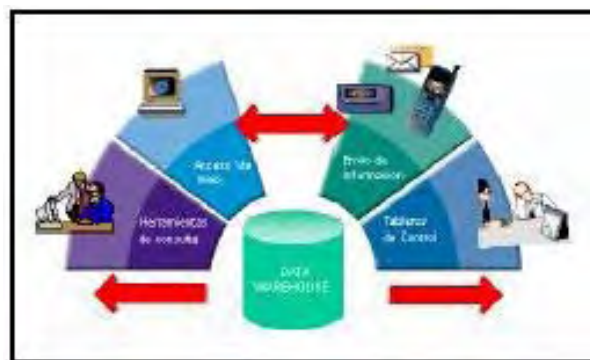


Figura 3.37: Variedad de interfases por perfil

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

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

Kimball destaca, como uno de los requisitos a cumplir por las aplicaciones, la posibilidad de hacer análisis AD HOC. Análisis ad hoc es simplemente la habilidad para los usuarios de cambiar los parámetros sobre un reporte para crear sus propias versiones personalizadas de ese reporte. De esta forma se dará respuesta a necesidades como “Quiero ver

este reporte, pero por mes en lugar de trimestre”, “Puedo ver este reporte a nivel provincia en lugar de región?”, “Puedo ver esta evolución de ventas mensual pero sólo de mi equipo de ventas?”. Más importante aún es que las respuestas a estos interrogantes lo resuelven los propios usuarios sin la necesidad de la intervención del departamento sistemas y maximizando el tiempo de análisis por sobre el tiempo de construcción e integración de la información.

Kimball define entonces que las aplicaciones “Templates” para usuarios finales proveen el marco (layout) y la estructura de un reporte para ser especializado por un conjunto de parámetros. El usuario selecciona los parámetros de una pick list o aceptando los valores por defecto cuando ejecuta el Template. Este enfoque orientado a parámetros permite a los usuarios generar docenas o potencialmente cientos de reportes de estructura similar desde un mismo Template. Todo esto de forma amigable y con interfases gráficas que hacen uso de todo el trabajo realizado en la construcción del warehouse en etapas previas del BDL.

Una advertencia que hace Kimball es el tiempo que existe entre el relevamiento y especificación de las aplicaciones para usuarios finales y el momento del desarrollo e implementación de la misma. Como pudimos ver, durante las primeras semanas del proyecto se realiza el relevamiento de los requerimientos de los usuarios y recién una vez que existen datos en el warehouse (aunque más no sea datos de prueba) puede comenzarse con la construcción de la aplicación final (al menos es lo

recomendable), es decir luego del diseño lógico y físico. Kimball insta entonces a manejar con mucho cuidado el tema de la documentación explícita (tratar que lo menos posible quede en la propiedad intelectual de los que hicieron el relevamiento). Sino se deberá realizar un esfuerzo significativo en el redescubrimiento de los puntos especificados en etapas tempranas del proyecto.

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

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

DESARROLLO DE APLICACIONES PARA USUARIOS FINALES

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

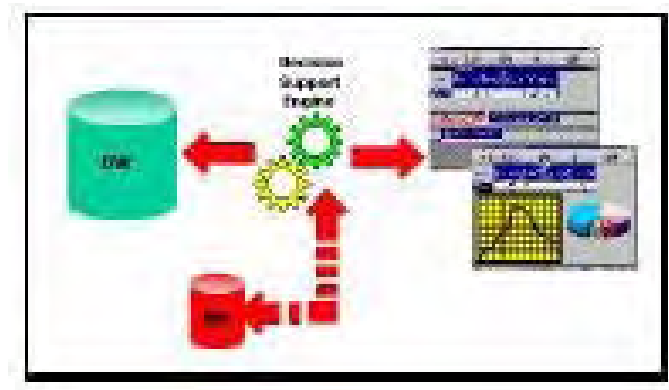


Figura 3.38: Configuración de Metadata y construcción de reportes

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

- Selección de un enfoque de implementación

Basado en Web

- _ Inter/Intra net
- _ Usuarios altamente distribuidos
- _ Manejo centralizado de nuevas versiones.

Herramienta propietaria

- _ Mayor complejidad de uso
- _ Para usuarios más capacitados
- _ Instalación local

EIS

- _ Acceso estructurado
- _ Secuencialidad de pantallas
- _ Push-Button

Interfase personalizada

- _ API (Application Programming Interface)
- _ Desarrollos propios sobre la base de un conjunto de funcionalidades

• Desarrollo de la aplicación

- _ Definición de herramienta de acceso al MetaData
- _ Desarrollo de templates y esquema de navegación de la aplicación
- _ Selección de reportes para pre-ejecución

• Prueba y verificación de datos

- _ Descripciones

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

- Documentación y Roll Out

- _ Retroalimentación con los resultados de la puesta en producción

- Mantenimiento

- _ Nuevos templates
 - _ Incorporación de nuevos sistemas fuentes
 - _ Monitoreo de performance
 - _ Eliminación de templates en desuso

IMPLEMENTACIÓN

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

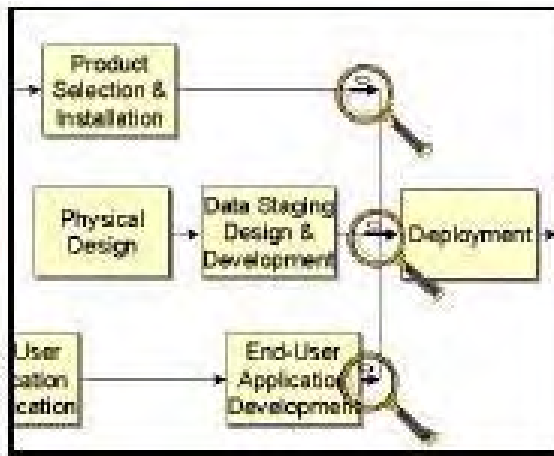


Figura 3.39: Implementación como convergencia de la tecnología, los datos y las aplicaciones

En esta etapa, mucho menos técnico y más orientado al gerenciamiento del proyecto, Kimball presenta una serie de tareas que deben cumplirse para garantizar un fin de proyecto y un producto de calidad.

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

asignadas), auditorías de tecnología sobre las configuraciones en las que se encontraban las PCs, proveer actualizaciones de hardware y software (determinando responsables, proyecto o área de usuario), verificaciones de seguridad (logon de red y base de datos), prueba de procedimientos de instalación en una variedad de máquinas, planificación de instalación con la correspondiente educación a los usuarios, etc.

Ralph Kimball dedica gran parte de esta etapa a la educación de los usuarios finales afirmando que “una robusta estrategia de educación para usuarios de negocios es un prerrequisito para el éxito del proyecto”.

Entre sus recomendaciones para cumplir exitosamente con esta tarea se encuentran dos aspectos, por una lado el alcance de la educación, y por otro, la metodología de presentación. Para lo primero, debe instruirse al usuario en tres aspectos claves: contenido del warehouse, aplicación y herramientas de acceso. En cuanto a lo segundo, para los usuarios finales no es fácil entender los límites entre las aplicaciones, los contenidos y las herramientas del warehouse.

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

clave es la correcta entrega de niveles de educación según el perfil del usuario (usuario final, power user, etc.)

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

Además, el contenido del warehouse puede variar entre curso y curso haciendo más complicada la administración y mantenimiento del material educativo. Kimball establece dos premisas para regular las relaciones entre la educación y el warehouse.

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

El soporte es la otra columna que ayudará a implementar con éxito el proyecto. Se deberá definir claramente los niveles de soporte (a nivel aplicación, modelo, calidad de datos) de forma transparente para el usuario. La documentación juega un papel importantísimo en la ayuda a usuarios finales, dado que el datawarehouse evolucionará rápidamente, la documentación escrita puede quedar obsoleta también de forma prematura. La documentación en línea empieza entonces a tener vida

propia, ya sea dentro del warehouse en sí mismo o como herramientas anexas, como el data warehouse web site propuesto por Kimball.

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

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

El objetivo de esta versión es conducir una prueba a nivel usuario de principio a fin. El grupo Beta está formado por los power users, la determinación de la cantidad de este grupo es también importante, pues

si son pocos la aplicación no se probará adecuadamente, y si son muchos se estará salteado la etapa de testing cayendo directamente en una salida a producción.

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

MANTENIMIENTO Y CRECIMIENTO

Como se remarca siempre, Data Warehousing es un proceso (de etapas bien definidas, con comienzo y fin, pero de naturaleza espiral) pues acompaña a la evolución de la organización durante toda su historia. Se necesita continuar con los relevamientos de forma constante para poder seguir la evolución de las metas por conseguir. Según afirma Kimball, “si se ha utilizado el BDL el data warehouse esta preparado para evolucionar y crecer”. Al contrario de los sistemas tradicionales, los cambios en el desarrollo deben ser vistos como signos de éxito y no de falla. Es importante establecer las prioridades para poder manejar los nuevos requerimientos de los usuarios y de esa forma poder evolucionar y crecer.

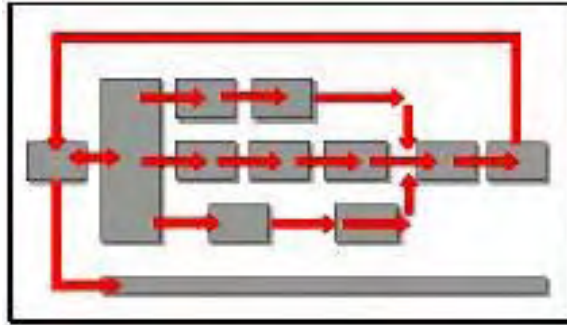


Figura 3.40: Data Warehousing como proceso de naturaleza espiral

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

Una vez que se ha construido e implantado el data warehouse no hay tiempo para el descanso, rápidamente debemos estar preparados para administrar el mantenimiento y crecimiento del mismo. Si bien las tareas pueden llegar a parecer similares a las tratadas en otras etapas del BDL, existe una diferencia clave: los usuarios están ahora accediendo al warehouse Kimball brinda entonces una serie de puntos a tener en cuenta para mantener exitosamente el warehouse. Entre ellos se destacan: el continuo soporte y la constante capacitación a usuarios de negocios, el manejo de la infraestructura (monitoreo de base de datos, tráfico, etc.), tuning de rendimiento sobre las consultas, mantenimiento del metadata y procesos ETLs. Otros aspectos involucran el monitoreo regular del cumplimiento de las expectativas sobre el warehouse (variables de medición del éxito del proyecto que se habían fijado en la etapa de

relevamiento), relevamiento de casos de estudio (situaciones reales donde una decisión basada en información del warehouse tuvo impacto sobre el negocio –ROI-), constante publicidad interna del uso del warehouse (permitiendo acceso siempre y cuando se tenga la capacitación correspondiente) y constante comunicación con los sectores de negocios y sistemas para asegurar la buena salud del datawarehouse.

En cuanto a cómo prepararse para el crecimiento del warehouse, Kimball recomienda la creación de un comité conformado por analistas del negocio y sistemas que establezca prioridades y procedimientos. El manejo de prioridades lo discrimina según sean mejoras menores o mayores (clasificadas según el impacto en el warehouse y excluyendo los errores, los cuales deben ser resueltos inmediatamente). Como mejoras menores incluye incorporación de datos sencillos, nuevas agregaciones, cambios en niveles superiores del modelo (atributos de alto nivel, lejanos a las bases tables), tareas que en definitiva no lleven mas días o semanas. Advierte sobre el razonamiento simplista de decidir no aplicar una mejora pues el cambio es muy menor, “1000 pequeñas mejoras es una ventaja estratégica”. Dentro de las mejoras de mayor impacto se encuentran las que modifican y/o crean bases tables o atributos asociados a las mismas, tareas que pueden disparar nuevos proyectos de meses de duración.

Finalmente, Kimball concluye su libro de su metodología garantizando que el seguimiento y cumplimiento de todas las etapas BDL ayudará a

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

GERENCIAMIENTO DEL PROYECTO

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

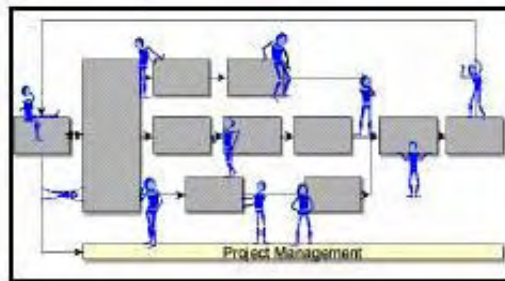


Figura 3.41: Gerenciamiento del Proyecto

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

PUNTOS PROBLEMÁTICOS DE UN DW

Según identifica Jeff Bedell, los principales puntos de atención que pueden llegar a complicar un proyecto de data warehousing se discriminan en según las siguientes tres áreas:

- **Rutinas de Carga.** Incluye programas de extracción y limpieza de datos. Surgen problemas en este punto dada la falta de integración y estructura consistente (alineada) entre los sistemas fuentes.
- **Mantenimiento.** Dados los diferentes períodos de almacenamiento para OLTP y OLAP y el hecho de que los DW son sistemas secundarios de información, otro problema surge para sincronizar los datos entre los sistemas operacionales fuentes y los warehouses.
- **Tuning.** Dado los patrones de uso y los métodos de acceso típicos de los sistemas OLAP, diseñadores y administradores deben realizar cambios significativos a los implementados en el tuning de sistemas OLTPs.

3.3. DEFINICIÓN DE TÉRMINOS BÁSICOS:

- ❖ **Agregación** : Actividad de combinar datos desde múltiples tablas para formar una unidad de información más compleja, necesitada frecuentemente para responder consultas del DataWarehouse en forma más rápida y fácil.
- ❖ **Datawarehouse** : Base de datos que almacena una gran cantidad de datos transaccionales integrados para ser usados para análisis gestionales por usuarios especializados (tomadores de decisión de la empresa).
- ❖ **DataMart** : Conjunto de hechos y datos organizados para soporte decisional basados en la necesidad de un área o departamento específico. Los datos son orientados a satisfacer las necesidades particulares de un departamento dado teniendo sólo sentido para el personal de ese departamento y sus datos no tienen porque tener las mismas fuentes que los de otro DataMart.
- ❖ **Datamining** : Análisis de los datos para descubrir relaciones, patrones, o asociaciones desconocidas.
- ❖ **Diccionario de Datos** : Un compendio de definiciones y especificaciones para las categorías de datos y sus relaciones.

- ❖ **Dimensión** : Entidad independiente dentro del modelo multidimensional de una organización, que sirve como llave de búsqueda (actuando como índice), o como mecanismo de selección de datos.
- ❖ **Drill Down** : Exponer progresivamente más detalle (dentro de un reporte o consulta), mediante selecciones de ítems sucesivamente.
- ❖ **Drill-Up** : Es el efecto contrario a drill -down. Significa ver menos nivel de detalle, sobre la jerarquía significa generalizar o sumarizar, es decir, subir en el árbol jerárquico.
- ❖ **DSS** : Sistema de Soporte de Decisiones. Sistema de aplicaciones automatizadas que asiste a la organización en la toma de decisiones mediante un análisis estratégico de la información histórica.
- ❖ **ETL (Extracción, Transformación y Transporte de datos)** : Pasos por los que atraviesan los datos para ir desde el sistema OLTP (o la fuente de datos utilizada) a la bodega dimensional. Extracción, se refiere al mecanismo por medio del cual los datos son leídos desde su fuente original. Transformación (también conocida como limpieza) es la etapa por la que puede atravesar una base de datos para estandarizar los datos de las distintas fuentes, normalizando y fijando una estructura para los datos. Finalmente está el Transporte, que consiste básicamente en llevar los datos leídos y estandarizados a la bodega dimensional (puede ser remota o localmente). Generalmente, para un DataMart no es necesario atravesar por todos estos pasos, pues al ser información

localizada, sus datos suelen estar naturalmente estandarizados (hay una sola fuente).

- ❖ **Jerarquía** : Es un conjunto de atributos descriptivos que permite que a medida que se tenga una relación de muchos a uno, se ascienda en la jerarquía.
- ❖ **OLAP (On-line Analytical Processing)** : Conjunto de principios que proveen un ambiente de trabajo dimensional para soporte decisional.
- ❖ **OLTP (On-line Transaction Processing)** : Sistema transaccional diario (o en detalle) que mantiene los datos operacionales del negocio.
- ❖ **Rollup** : Comando propio del lenguaje Oracle Express, que simboliza las sumas agregadas de una variable a través de los niveles jerárquicos de las dimensiones que la sustentan.
- ❖ **Snapshot** : Imagen instantánea de los datos en un tiempo dado.
- ❖ **Sumarización** : Actividad de incremento de la granularidad de la información en una base de datos. La sumarización reduce el nivel de detalle, y es muy útil para presentar los datos para apoyar al proceso de Toma de Decisiones.
- ❖ **Tabla Dimensional** : Dentro del esquema estrella, corresponde a las tablas que están unidas a la tabla central a través de sus respectivas llaves. La cantidad de estas tablas le otorgan la característica de multidimensionalidad a esta estrategia.

CAPITULO IV

4. METODOLOGÍA DE LA INVESTIGACIÓN

Para el desarrollo del DataMart se empleó la metodología Ralph Kimball, dado que es una metodología que ha sido usada con éxito en diferentes proyectos y también esta metodología tiene claramente definido sus procesos para todo el ciclo del proyecto.

4.1. METODOLOGÍA DE DESARROLLO DE RALPH KIMBALL

Ralph Kimball es reconocido como uno de los padres del concepto de Datawarehouse, se ha dedicado desde hace más de 10 años al desarrollo de su metodología para que éste concepto sea bien aplicado en las organizaciones y se asegure la calidad en el desarrollo de estos proyectos.

La metodología de Ralph Kimball se enfoca principalmente en el diseño de la base de datos que almacenará la información para la toma de decisiones.

El diseño se basa en la creación de tablas de hechos, es decir, tablas que contengan la información numérica de los indicadores a analizar, o sea la parte cuantitativa de la información para la toma de decisiones.

Las tablas anteriores se relacionan con tablas de dimensiones, las cuales contienen la información cualitativa, de los indicadores, es decir, toda aquella información que clasifique la información requerida.

A éste modelo de datos se le conoce como "diseño estrella", existen variaciones de éste, llamados "copo de nieve" y "diseño flat". Todos estos diseños tienen la característica de preparar la información de acuerdo a la necesidad de tomar decisiones y no a los argumentos técnicos de espacio de almacenamiento.

Por lo tanto Tomaremos como base el ciclo de vida de los Data Warehouses definido por Ralph Kimball. El marco presentado por Ralph Kimball con el nombre de Business Dimensional Lifecycle (BDL) ilustra las diferentes etapas por las que debe pasar todo proceso de Data Warehousing. Este enfoque de implementación de data warehouses es ilustrado en la siguiente figura. Este diagrama ilustra la secuencialidad de tareas de alto nivel requeridas para el efectivo diseño, desarrollo e implementación de data warehouses. El diagrama muestra una vista general del mapa de ruta de un proyecto en el cual cada rectángulo nos indica en que etapa estamos, por dónde pasamos y hacia dónde debemos dirigirnos.

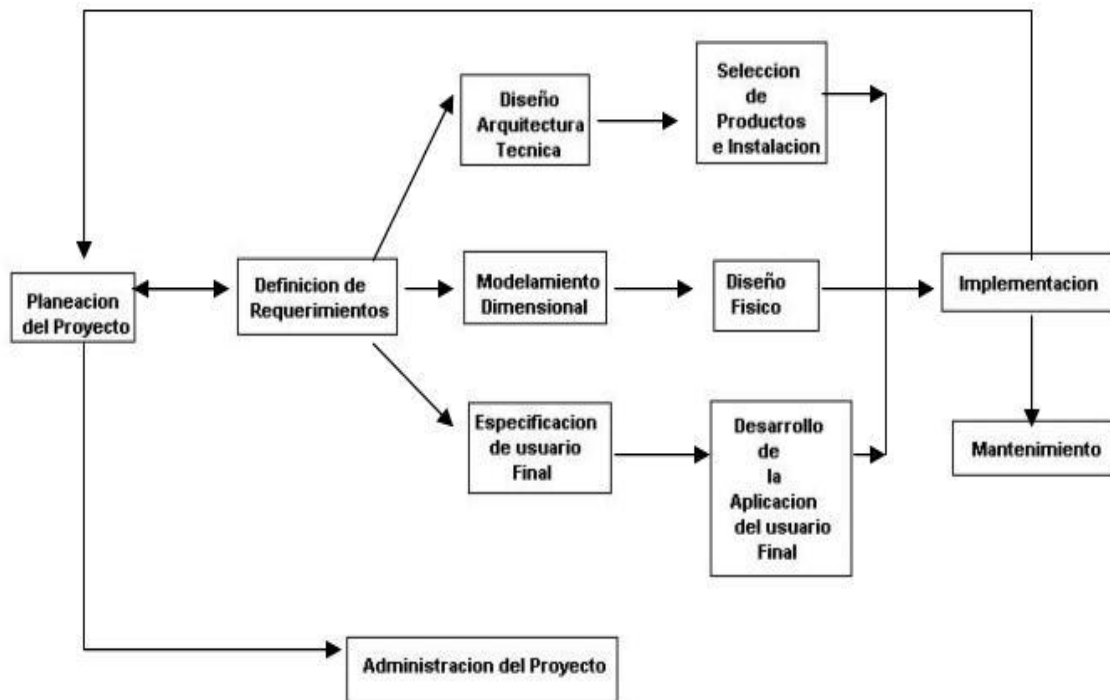


Figura 4.1. Metodología Kimball

4.1.1. Planificación del Proyecto

- **Objetivo del Proyecto**

El objetivo del proyecto es construir un Datamart para dar soporte a la toma de decisiones a las Dirección Nacional de Justicia y en la Dirección de Defensoría de Oficio.

- **Alcances del Proyecto**

Utilizando la metodología de Ralph Kimball, se diseñará un Datamart para el sector justicia.

El Datamart a diseñar será para el área de Defensoría de Oficio de la Dirección Nacional de Justicia.

La fuente de la que serán extraídos los datos será el sistema transaccional de defensores.

La tecnología de implementación será Microsoft SQL Server 2000 como Base de datos del Data Mart y DTS para la extracción, transformación y carga de la información, Power Designer 11 para el modelado.

El proyecto no incluye la implementación del Datamart , sólo se centrará en el análisis, diseño y los prototipos de Reportes.

- **Personal involucrado en el proyecto:**

Equipos de Requerimiento: Está conformado por los usuarios de los cuales se obtendrá información e identificarán las necesidades. Este equipo está conformado por: El director de la Dirección Nacional de Justicia, El director de Defensoría de Oficio, el Director de Informática y el especialista de Informática y Sistemas de la oficina de Defensoría de Oficio.

Equipo de Desarrolladores: Este equipo está conformado por el Líder Técnico, un consultor DW, un Programador de Extracción, un Programador de BI y un analista funcional.

Equipo de Usuarios de Prueba: Este equipo está conformado por el especialista de informática de la oficina de Defensoría de Oficio.

4.1.2. Definición de los Requerimientos del Negocio

Los Requerimientos del negocio son el conjunto de necesidades específicas que se han identificado durante las entrevistas al personal del área de defensoría. Cada uno de los requerimientos levantados pueden ser resuelto con uno o varios reportes.

Algunos de los requerimientos son :

- Número de Casos por Periodo

- Número de Casos por Distrito Judicial

- Delitos mas Frecuentes

- Número de Delitos por Periodo

- Número de Delitos por Sexo

- Número de Delitos por Edad

- Gestiones previas a un beneficio

4.1.3. Modelado Dimensional

En esta etapa se debe de seleccionar los procesos a modelar de acuerdo a las necesidades de los usuarios.

El análisis es realizado por medio de reportes por lo tanto al diseñar el DataMart se debe de tener en cuenta que información se desea obtener.

Diseñar los modelos de datos para soportar estos análisis requieren un enfoque diferente al usado en los sistemas transaccionales.

En esta fase se analiza el modelo de datos del sistema transaccional para construir el modelo Dimensional del Datamart.

Modelo del Datamart

A partir del modelo de datos del sistema transaccional se construye el modelo Dimensional del Datamart.

Dimensiones

Descripciones que no cambian (mucho) en el tiempo (delitos, dependencias, etc.)

Usados para vistas y profundización, Usualmente son Textos.

Las dimensiones que conforman el Datamart son:

N°	Dimensiones
----	-------------

-
- | | |
|----|-------------|
| 1. | Tiempo |
| 2. | Cliente |
| 3. | Defensor |
| 4. | Materia |
| 5. | Dependencia |
| 6. | Gestión |

Dimensión Tiempo

Descripción

Almacena las fechas para las que el Datamart contiene información relevante.

Jerarquías

Contiene la fecha (día, mes, año) y el número de semana en el año.

Nivel	Atributos
Nivel 1	Año
Nivel 2	Semestre
Nivel 3	Trimestre
Nivel 4	Mes
Nivel 5	Día

Atributos

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por Defecto
TiempoKey	Llave primaria de la Dimensión Tiempo	Cadena de caracteres	Ninguno
Año	Contiene el número del año.	Número entero	Ninguno

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por Defecto
Semestre	Contiene el número del Semestres.	Número entero	
Trimestre	Contiene el número del Trimestres.	Número entero	
Mes	Contiene el número del mes.	Número entero	Ninguno
Dia	Contiene el número de días en el mes.	Número Entero	Ninguno

Dimensión Cliente

Descripción

Contiene los datos de los clientes que solicitan los servicios de defensa legal gratuita a la defensoría de oficio.

Jerarquías

Datos generales del Cliente.

Nivel	Atributos
Nivel 1	IdCliente
Nivel 2	IdSexo
Nivel 2	IdDepartamento
Nivel 3	IdProvincia
Nivel 4	IdDistrito

Atributos

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por Defecto
ClienteKey	Llave Primaria de la Dimensión Cliente	Cadena de caracteres	Ninguno
IdCliente	Código del Cliente	Número entero	Ninguno

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por Defecto
IdDepartamento	Código del Departamento	Número entero	Ninguno
IdProvincia	Código de la Provincia	Número entero	Ninguno
IdDistrito	Código del Distrito	Número entero	Ninguno
IdSexo	Código del Sexo	Número entero	Ninguno
Nombre	Nombres del Cliente	Cadena de caracteres	Ninguno
DescripciónDepartamento	Descripción del Departamento	Cadena de caracteres	Ninguno
DescripciónProvincia	Descripción de la Provincia	Cadena de caracteres	Ninguno
DescripciónDistrito	Descripción del Distrito	Cadena de caracteres	Ninguno
DescripciónSexo	Descripción del Sexo	Cadena de caracteres	Ninguno
Edad	Edad	Numérico	Ninguno

Dimensión Defensor

Descripción

Contiene los datos de los defensores quienes son los abogados que prestan sus servicios a los clientes.

Jerarquías

Datos generales del defensor.

Nivel	Atributos
Nivel 1	IdDefensor
Nivel 2	IdSexo
Nivel 2	IdDepartamento
Nivel 3	IdProvincia
Nivel 4	IdDistrito

Atributos

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por Defecto
DefensorKey	Llave Primaria de la Dimensión Defensor	Cadena de caracteres	Ninguno
IdDefensor	Código del Defensor	Número entero	Ninguno

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por Defecto
IdDepartamento	Código del Departamento	Número entero	Ninguno
IdProvincia	Código de la Provincia	Número entero	Ninguno
IdDistrito	Código del Distrito	Número entero	Ninguno
IdSexo	Código del Sexo	Número entero	Ninguno
Nombre	Nombres del Defensor	Cadena de caracteres	Ninguno
DescripciónDepartamento	Descripción del Departamento	Cadena de caracteres	Ninguno
DescripciónProvincia	Descripción de la Provincia	Cadena de caracteres	Ninguno
DescripciónDistrito	Descripción del Distrito	Cadena de caracteres	Ninguno
DescripciónSexo	Descripción del Sexo	Cadena de caracteres	Ninguno

Dimensión Materia

Descripción

Contiene los datos de las materias o delitos en que incurre el cliente.

Jerarquía

Datos generales de la materia.

Nivel	Atributos
Nivel1	IdMateria

Atributos

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por Defecto
MateriaKey	Llave principal de la dimensión Materia	Cadena de caracteres	Ninguno
idMateria	Código de la materia	Número entero	Ninguno
Descripción	Descripción de la Materia	Cadena de caracteres (se utilizará mayúsculas)	Ninguno

Dimensión Dependencia

Descripción

Contiene los datos de las dependencias, tipo de dependencias y distritos judiciales.

Jerarquías

La Jerarquía representa la clasificación de la dependencia.

Nivel	Atributos
Nivel1	IdDependencia
Nivel2	IdDistritoJudicial
Nivel2	IdTipoDependencia

Atributos

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por Defecto
DependenciaKey	Llave principal de la dimensión dependencia	Cadena de caracteres	Ninguno
idDependencia	Código de la dependencia	Número entero	Ninguno
distritoJudicialId	Distrito Judicial al que pertenece la dependencia.	Número entero	Ninguno

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por Defecto
tipoDependenciald	Tipo de Dependencia al que pertenecen la dependencia.	Número entero	Ninguno
DependenciaDescripción	Descripción de la dependencia	Cadena de caracteres	Ninguno
DistritoDependencia	Descripción del Distrito Judicial	Cadena de caracteres	Ninguno
TipoDependenciaDependencia	Descripción del tipo de dependencia	Cadena de caracteres	Ninguno

Dimensión Gestión

Descripción

Contiene las Gestiones que son todas las acciones o diligencias que realiza el defensor.

Jerarquías

Datos generales de las Gestiones.

Nivel	Atributos
Nivel 1	IdGestión
Nivel 2	IdResultado
Nivel 3	IdApelación

Atributos

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por Defecto
GestiónKey	Llave principal de la dimensión Gestión	Cadena de caracteres	Ninguno
IdGestión	Código de la Gestión	Número entero	Ninguno
IdResultado	Código del Resultado	Número entero	Ninguno
IdApelación	Código de la Apelación	Número entero	Ninguno
GestiónDescripción	Descripción de la Gestión	Cadena de caracteres	Ninguno

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por Defecto
ResultadoDescripción	Descripción del Resultado	Cadena de caracteres	Ninguno
ApelaciónDescripción	Descripción de la Apelación	Cadena de caracteres	Ninguno

Facts Procesos

Los hechos son medidas numéricas del negocio.

- Número de Casos
- Cantidad de Delitos

Los hechos son agregados utilizando sumas, promedios, mínimos, máximos, etc.

Descripción

Contiene las llaves de las dimensiones y las medidas que fueron levantadas en las etapas anteriores del proyecto de acuerdo a las especificaciones del equipo de requerimiento.

Granularidad

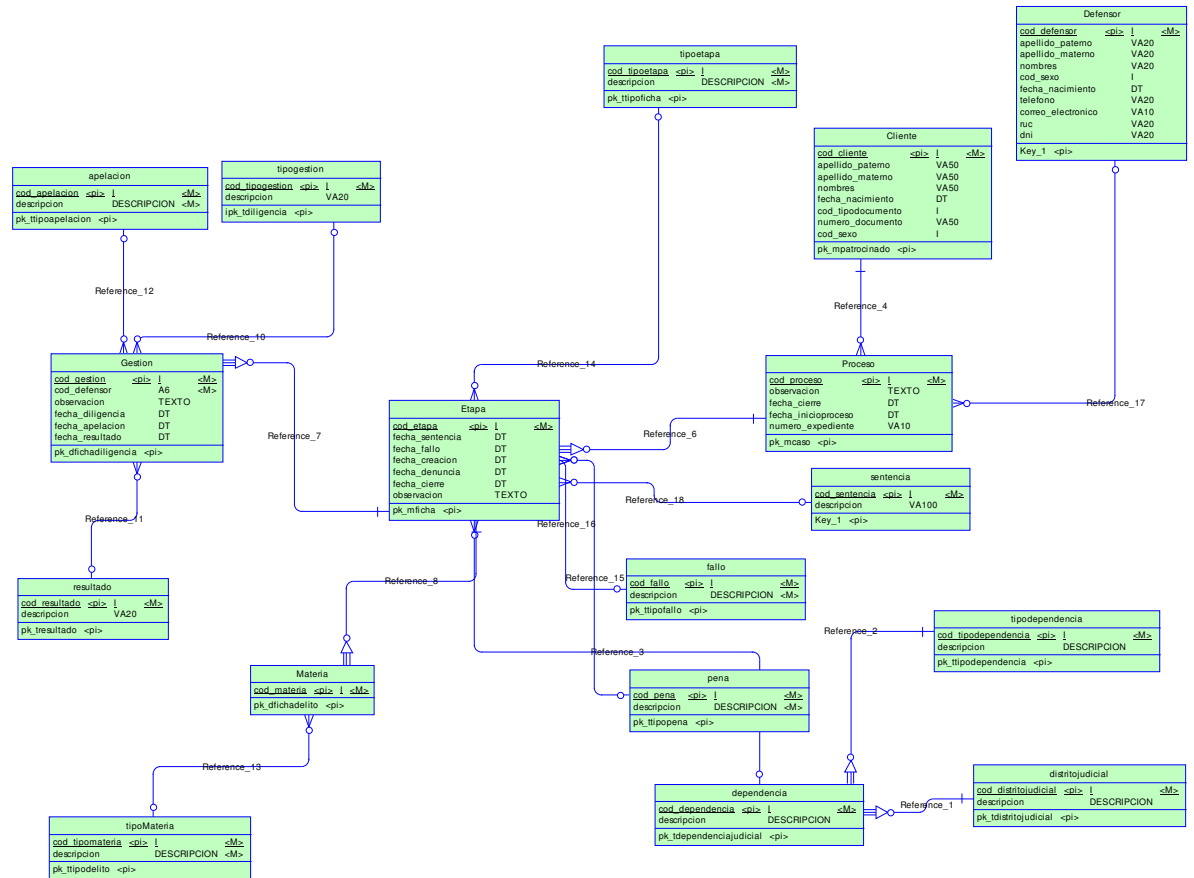
N°	Nombre de la Dimensión	Descripción	Llave Primaria
1.	Tiempo	Fecha en que se realizó la defensa.	Si
2.	Cliente	Cliente al que se le dio el servicio de defensoría	Si
3.	Defensor	Defensor que realizó la defensa.	Si
4.	Materia	Materias de los Procesos.	Si
5.	Dependencia	Dependencia donde se desarrolla el proceso	Si
6.	Gestión	Gestión desarrolladas por los defensores	Si

Medidas

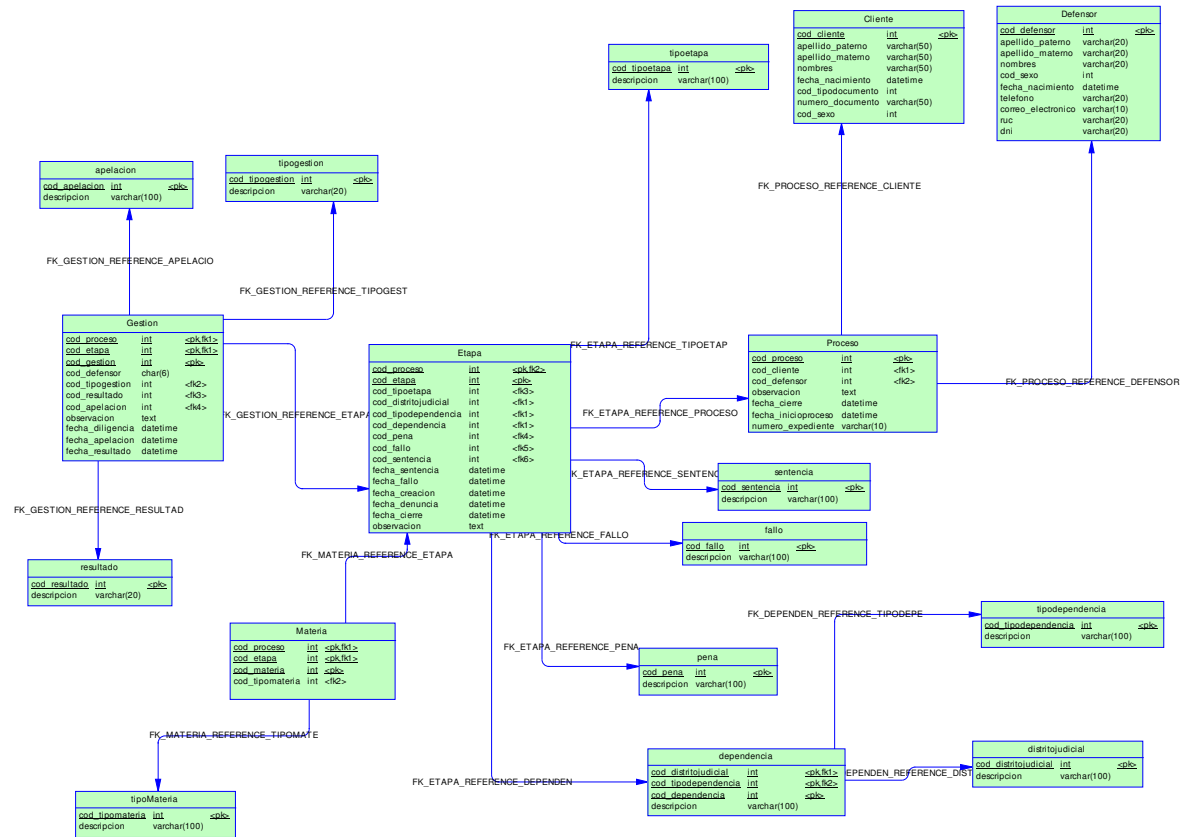
N°	Nombre	Descripción
1.	Num_SalAlternativas_Favorables	Nº de Salidas alternativas favorables.
2.	Num_SalAlternativas_Totales	Nº de Salidas alternativas totales.
3.	Num_Detenciones_Ilegales	Nº de Detenciones ilegales.
4.	Num_Absoluciones	Nº de Absoluciones.
5.	Num_Medidas_Cautelares	Nº de Medidas Cautelares.
6.	Num_Imputados_Ingresados	Nº de Imputados ingresados.
7.	Num_Imputados_Terminados	Nº de Imputados con procesos terminados.
8.	Num_Visitas_Carcel	Nº de Visitas a cárceles.
9.	Num_Casos_Ingresados	Nº de Casos Ingresados.
10.	Num_CasosxDefensor	Nº de Casos por Defensor.
11.	Num_Casos_Terminados	Nº de Casos Terminados.
12.	Num_Casos_Nuevos	Nº de Casos Nuevos.
13.	Num_Casos_Archivados	Nº de Casos Archivados.
14.	Num_Casos_Seguimiento	Nº de Casos en Seguimiento.
15.	Num_Delitos_Atendidos	Nº de Delitos Atendidos.
16.	Num_Delitos_Absolutorias	Nº de Delitos con Sentencia Absolutoria.
17.	Num_Delitos_Condenatorios	Nº de Delitos con Sentencia Condenatoria.

Modelo Lógico de Datos

En este proyecto contamos con una base de datos transaccional altamente normalizada para optimizar el ingreso de datos.

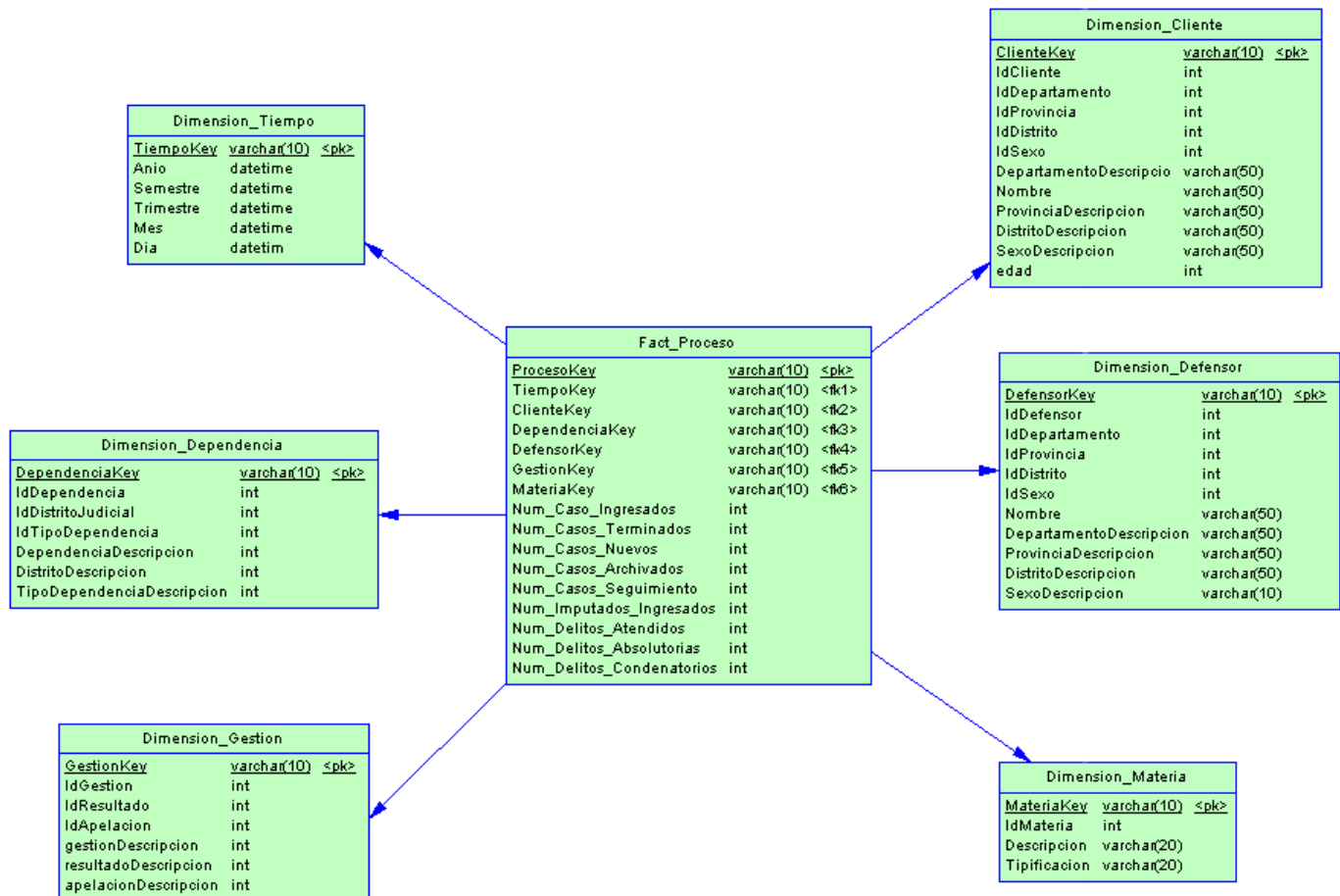


Modelo Físico de Datos



Modelo Dimensional

El esquema Estrella diseñado en el proyecto consiste de siete tablas, una tabla de hechos llamada Fact_Procesos y seis tablas de dimensiones llamadas Dimensión_Tiempo, Dimensión_Dependencia, Dimensión_Gestión, Dimensión_Cliente, Dimensión_Defensor, Dimensión_Materia.



4.1.4 Diseño Físico

En esta fase se procede a la creación física del DataMart, se realiza el diseño óptimo de los índices, el cual es basado en el esquema lógico y la carga de trabajo.

En nuestro caso el Sistema Administrador de Base de Datos es el Microsoft SQL Server 2000, el cual realiza automáticamente la selección óptima de los índices; además cuenta con un servidor de análisis multidimensional OLAP.

A continuación mencionamos algunas estrategias para mejorar el Diseño Físico del Datamart:

Particionamiento

El particionamiento permite que los datos de una tabla lógica se repartan en múltiples conjuntos de datos físicos. Este particionamiento se basa en una columna de la tabla de hechos, que comúnmente es la columna que indica la fecha. Dado que la columna por la cual se particione tiene que ser parte de la clave primaria, no puede ser una columna derivada ni contener valores nulos.

Clustering

Es una técnica muy útil para el acceso secuencial de grandes cantidades de datos. El clustering se obtiene definiendo un índice clustering para una tabla, el cual determina el orden secuencial físico en el que se almacenan las filas en los conjuntos de datos.

Esta técnica es importante porque mejora drásticamente la performance del acceso secuencial, y este tipo de acceso es el más usado en el procesamiento OLAP.

Indexado

Existen dos estrategias extremas de indexado: una es indexar todo, y la otra es no indexar nada, pero ninguna de las dos es conveniente. Las columnas que se elijan para indexar deben ser las que se usan mas frecuentemente para recuperar las filas, y las que tienen una alta distribución de valores.

4.1.5. Diseño y Desarrollo de Presentación de Datos

Paquetes DTS

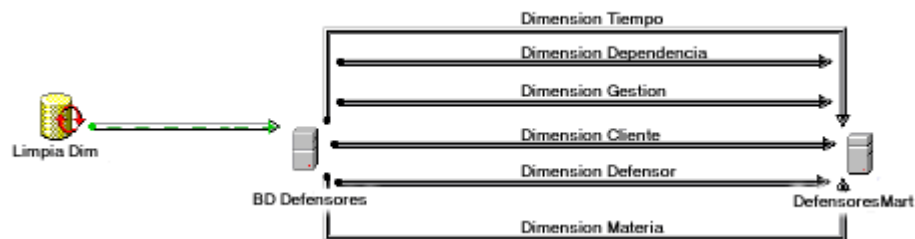
Es una herramienta para mover, copiar, modificar y trabajar con orígenes de datos iguales o diferentes, pero en este punto vamos a hacer la importación y exportación de datos entre dos orígenes de datos diferentes.

DTS tiene una arquitectura OLE DB por lo que puede copiar y transformar múltiples orígenes de datos.

Carga Original Data



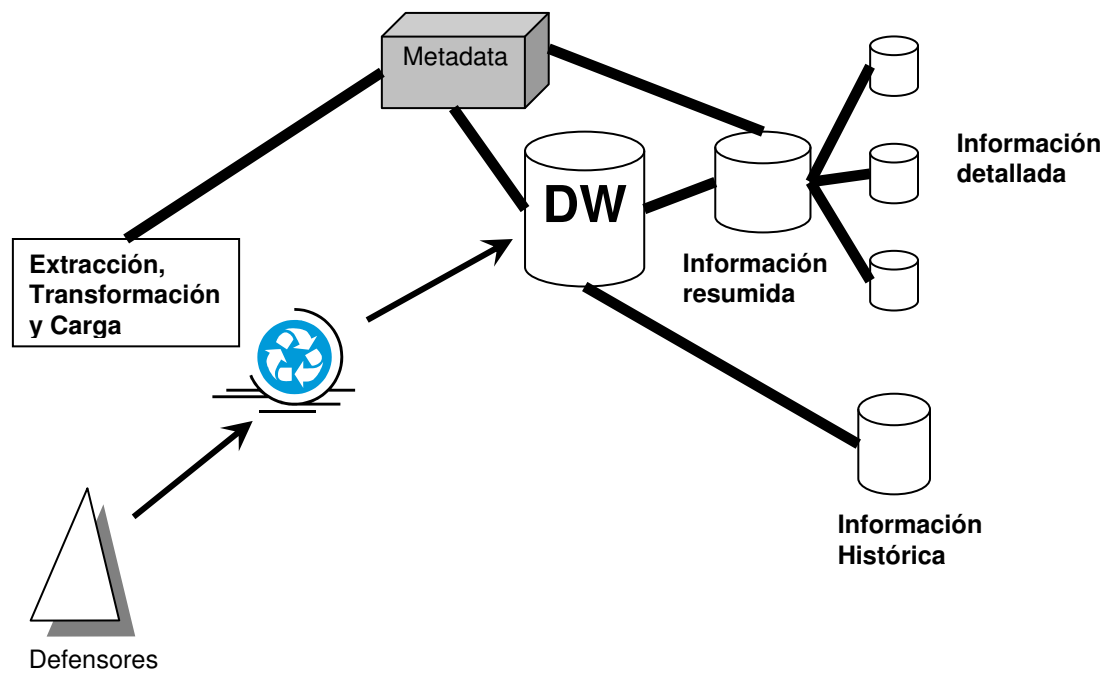
Paquetes DTS : Carga Dim —————> Carga de Dimensiones



Paquetes DTS : Carga Fact



Arquitectura de Extracción



Sistemas Transaccionales

Fuente	Tipo	Conexión
Fuente	Sybase ASE	ODBC

Procesos de Extracción y Transformación

Dimensión Cliente

Descripción

Representa la carga de la dimensión Cliente.

Descripción de Tablas Fuentes

Tipo de Fuente	Nombre de Tabla	Descripción
Cliente (ASE Sybase)	Cliente	Contiene los datos de los clientes (personas naturales)

Estandarización de Datos y Limpieza de Datos

Nombre	Llave	Tipo	Formato	Limpieza	Valor por Defecto
Cod_cliente	Si	Número (Entero Largo)	Numérico		0
Nombre	No	Texto	Texto (Mayúscula)	No debe ser nulo	No tiene

Fuentes de Datos

Tabla:	Cliente
--------	---------

Nombre	Llave	Tipo	Formato	Consideración Importante
Cod_cliente	Si	Numérico	Número	
nombres	No	Texto	Texto	
apellido_paterno	No	Texto	Texto	
apellido_materno	No	Texto	Texto	
fecha_nacimiento	No	Texto	Texto	
Cod_tipodocumento	No	Numérico	Número	
Número_documento	No	Texto	Texto	
Cod_sexo	No	Texto (1)	Texto	Puede ser "M" o "F".

Tabla Destino

Tabla:	tmpCliente	
Campo	Tipo	Mapeo
ClienteKey	TEXTO(20)	ClienteKey llenarCeros(cliente_Id,20)
IdCliente	Número (Entero Largo)	Cliente.cliente_id
Nombre	Texto(50)	Concatenar(cliente.nombres, cliente.apellido_paterno, cliente.apellido_materno)
Sexo	Texto(1)	cliente.sexo

Tabla:	tmpCliente	
Campo	Tipo	Mapeo
Edad	Número	Campo calculado

Proceso

1. Borrar Tablas Temporal

Eliminar la tabla temporal tmpCliente

2. Cargar registros de la tabla Cliente

Toma los datos de la tabla Cliente de acuerdo al mapeo, y lo lleva a la tabla temporal tmpCliente.

3. Carga de la Dimensión

Tomar los valores de la tabla temporal tmpCliente y llevarla a la dimensión Cliente. En caso que sean nuevos clientes insertarlos, en caso que sean clientes registrados, actualizar sólo: Nombre

Dimensión Defensor

Descripción

Representa la carga de la dimensión Defensor

Descripción de Tablas Fuentes

Tipo de Fuente	Nombre de Tabla	Descripción
Proveedor (ASE Sybase)	Defensor	Defensores que utiliza el sistema.
Ubicación_Geográfica(ASE Sybase)	Ubicación	Contiene Departamentos , Provincias y Distritos del Perú

Estandarización de Datos y Limpieza de Datos

Nombre	Llave	Tipo	Formato	Limpieza	Valor por Defecto
Cod_defensor	PK	Texto	Texto		NO TIENE
Departamento		Número	Numérico	No debe ser Nulo	NO TIENE
Provincia		Número	Numérico	No debe ser Nulo	NO TIENE
Distrito		Número	Numérico	No debe ser Nulo	NO TIENE

Nombre	Llave	Tipo	Formato	Limpieza	Valor por Defecto
Cod_sexo	No	Texto (1)	Texto	Puede ser "M" o "F".	sexo_id

Fuentes de Datos

Tabla: Defensor				
Nombre	Llave	Tipo	Formato	Consideración Importante
Cod_defensor		Número	Numérico	
Nombres		Texto(20)	Texto	
apellido_paterno		Texto(20)	Texto	
apellido_materno		Texto(20)	Texto	
idUbicación_Geográfica	FK	Texto(6)	Texto	
fecha_nacimiento		DateTime	dd/mm/yyyy	
Ruc		TEXT(11)	Texto	
Dni		TEXT(8)	Text	
Teléfono		TEXT(10)	Text	
correo_electrónico		TEXT(20)	Text	
Cod_sexo		Texto (1)	Texto	Puede ser "M" o "F".

Tabla: Ubicación_Geográfica				
Nombre	Llave	Tipo	Formato	Consideración Importante
idUbicación_Geográfica	PK	TEXTO(6)	TEXTO	
Departamento		Número	Numérico	
Provincia		Número	Numérico	
Distrito		Número	Numérico	

Tabla Destino

Tabla: TmpDefensor		
Campo	Tipo	Mapeo
DefensorKey	TEXTO(20)	DefensorKey llenarCeros(cod_defensor,20)
Cod_Defensor	Número	Defensor.cod_defensor

Tabla:	TmpDefensor	
Campo	Tipo	Mapeo
Nombres	TEXTO (30)	Defensor.Nombres
IdSexo	TEXTO(1)	Mayúscula(Defensor.sexo_id)
Departamento	Número	Mayúscula(Ubicación_Geográfica.Departamento)
Provincia	Número	Mayúscula(Ubicación_Geográfica.Provincia)
Distrito	Número	Mayúscula(Ubicación_Geográfica.Distrito)

Proceso

1. Borrar Tablas Temporal

Eliminar la tabla temporal tmpDefensor

2. Cargar registros de la tabla Defensores

Toma los datos de la tabla Defensores de acuerdo al mapeo indicado y carga los datos a la tabla tmpDefensor

3. Carga de la Dimensión

Tomar los valores de la tabla temporal tmpDefensor y llevarla a la dimensión Defensor. En caso que sean nuevos defensores insertarlos.

Dimensión Materia

Descripción

Representa la carga de la dimensión Materia.

Descripción de Tablas Fuentes

Tipo de Fuente	Nombre de Tabla	Descripción
Delito (ASE Sybase)	Materia	Contiene los datos de las materias.

Estandarización de Datos y Limpieza de Datos

Nombre	Llave	Tipo	Formato	Limpieza	Valor por Defecto
Cod_materia	Si	Número	Numérico		0
Descripción	No	Texto	Texto (Mayúscula)	No debe ser nulo	No tiene

Fuentes de Datos

Tabla: Materia				
Nombre	Llave	Tipo	Formato	Consideración Importante
Cod_materia	Si	Numérico	Número	
Descripción	No	Texto	Texto	

Tabla Destino

Tabla:	tmpMateria	
Campo	Tipo	Mapeo
MateriaKey	TEXTO(20)	MateriaKey llenarCeros(cod_materia,20)
IdMateria	Número	Materia.cod_materia
Descripción	Texto(50)	Materia.descripcion

Proceso

1. Borrar Tablas Temporal

Eliminar la tabla temporal tmpMateria

2. Cargar registros de la tabla Materia

Toma los datos de la tabla Materia de acuerdo al mapeo y lo lleva a la tabla temporal tmpMateria.

3. Carga de la Dimensión

Tomar los valores de la tabla temporal tmpMateria y llevarla a la dimensión Materia.

Dimensión Dependencia

Descripción

Representa la carga de la dimensión Dependencia.

Descripción de Tablas Fuentes

Tipo de Fuente	Nombre de Tabla	Descripción
Dependencia (ASE Sybase)	dependencia	Dependencias que utiliza el sistema
DistritoJudicial (ASE Sybase)	distritojudicial	Almacena los distritos judiciales de las dependencias
TipoDependencia(ASE Sybase)	tipodependencia	Almacena los tipos de dependencias

Estandarización de Datos y Limpieza de Datos

Nombre	Llave	Tipo	Formato	Limpieza	Valor por Defecto
Cod_dependencia	PK	Numérico	Número	No debe ser Nulo	NO TIENE
Cod_distritoJudicial	PK	Numérico	Número	No debe ser Nulo	NO TIENE

Nombre	Llave	Tipo	Formato	Limpieza	Valor por Defecto
Cod_tipoDependencia	PK	Numérico	Número	No debe ser Nulo	NO TIENE
dependenciaDescripción		Texto	Texto(Mayúscula)	No debe ser Nulo	NO TIENE
distritojudicialDescripción		Texto	Texto(Mayúscula)	No debe ser Nulo	NO TIENE
tipodependenciaDescripción		Texto	Texto(Mayúscula)	No debe ser Nulo	NO TIENE

Fuentes de Datos

Tabla: Dependencia				
Nombre	Llave	Tipo	Formato	Consideración Importante
Cod_dependencia	PK	Numérico	Número	
Cod_distritojudicial	PK	Numérico	Número	
Cod_tipodependencia	PK	Numérico	Número	
Descripción		TEXTO	TEXTO	

Tabla: DistritoJudicial

Nombre	Llave	Tipo	Formato	Consideración Importante
Cod_distritojudicial	PK	Numérico	Número	
Descripción		TEXTO	TEXTO	

Tabla: TipoDependencia				
Nombre	Llave	Tipo	Formato	Consideración Importante
Cod_tipodependencia	PK	Numéri co	Número	
Descripción		TEXTO	TEXTO	

Tabla Destino

Tabla:	TmpDependencia	
Campo	Tipo	Mapeo
DependenciaKey	TEXTO(20)	
idDependencia	Numérico	Dependencia. Cod_dependencia
idDistritoJudicial	Numérico	DistritoJudicial. Cod_distritoJudicial

Tabla:	TmpDependencia		
Campo	Tipo	Mapeo	
idTipoDependencia	Numérico	TipoDependencia.cod_tipoDependencia	
dependenciaDescripción	TEXTO(255)	Mayúscula(Dependencia.Descripción)	
distritoJudicialDescripción	TEXTO(255)	Mayúscula(DistritoJudicial.Descripción)	
tipoDependenciaDescripción	TEXTO(255)	Mayuscula(TipoDependencia.Descripción)	

Proceso

1. Borrar Tablas Temporal

Eliminar la tabla temporal tmpDependencia

2. Cargar registros de la tabla Dependencia

Toma los datos de la tabla Dependencia de acuerdo al mapeo indicado y carga los datos a la tabla tmpDependencia

3. Cargar registros de la tabla Distrito Judicial

Toma los datos de la tabla DistritoJudicial de acuerdo al mapeo indicado y carga los datos a la tabla tmpDependencia

4. Cargar registros de la tabla TipoDependencia

Toma los datos de la tabla TipoDependencia de acuerdo al mapeo indicado y carga los datos a la tabla tmpDependencia

5. Carga de la Dimensión

Tomar los valores de la tabla temporal tmpDependencia y llevarla a la dimensión Dependencia.

Dimensión Gestión

Descripción

Representa la carga de la dimensión Gestión.

Fuentes de Datos

Tabla: Gestión				
Nombre	Llave	Tipo	Formato	Consideración Importante
Cod_gestión	PK	Numérico	Número	
Cod_TipoGestión	FK	Numérico	Número	
Cod_resultado	FK	Numérico	Número	
Cod_apelación	FK	Numérico	Número	
Fecha_diligencia		DateTime	DateTime	
Fecha_resultado		DateTime	DateTime	
Fecha_apelación		DateTime	DateTime	

Tabla: TipoGestión				
Nombre	Llave	Tipo	Formato	Consideración Importante
TipoGestión_id	PK	Numérico	Número	
Descripción		TEXTO	TEXTO	

Tabla: resultado				
Nombre	Llave	Tipo	Formato	Consideración Importante
Tiporesultado_id	PK	Numérico	Número	
Descripción		TEXTO	TEXTO	

Tabla: apelacion				
Nombre	Llave	Tipo	Formato	Consideración Importante
Tipoapelación_id	PK	Numérico	Número	
Descripción		TEXTO	TEXTO	

Tabla Destino

Tabla:	TmpGestión	
Campo	Tipo	Mapeo
GestiónKey	TEXTO(20)	GestiónKey llenarCeros(cod_gestió n,20)
IdDigestión	Numérico	TipoDiligencia.diligenci ald
IdResultadp	Numérico	TipoResultado.resultad old

Tabla:	TmpGestión		
Campo		Tipo	Mapeo
IdApelación		Numérico	TipoApelación.apelaciónId
diligenciaDescripción		TEXTO(255)	Mayuscula(tipogestión.Descripción)
resultadoDescripción		TEXTO(255)	Mayúscula(resultado.Descripción)
apelacionDescripción		TEXTO(255)	Mayúscula(apelación.Descripción)

Proceso

1. Borrar Tablas Temporal

Eliminar la tabla temporal tmpGestión

2. Cargar registros de la tabla Gestión

Toma los datos de la tabla Gestión de acuerdo al mapeo indicado y carga los datos a la tabla tmpGestión

3. Cargar registros de la tabla TipoGestión

Toma los datos de la tabla Tipogestión de acuerdo al mapeo indicado y carga los datos a la tabla tmpGestión

4. Cargar registros de la tabla resultado

Toma los datos de la tabla resultado de acuerdo al mapeo indicado y carga los datos a la tabla tmpGestión

5. Cargar registros de la tabla apelación

Toma los datos de la tabla apelación de acuerdo al mapeo indicado y carga los datos a la tabla tmpGestión

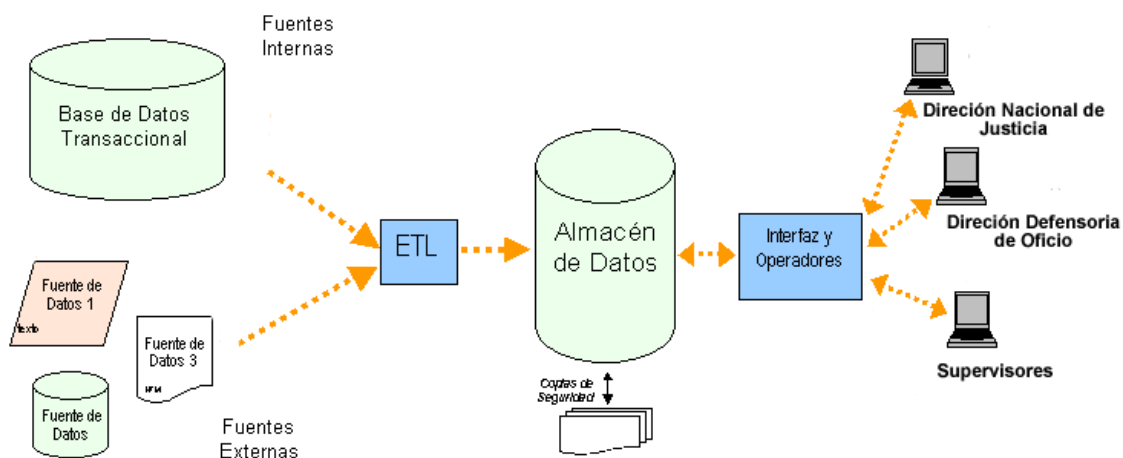
6. Carga de la Dimensión

Tomar los valores de la tabla temporal tmpGestión y llevarla a la dimensión Gestión.

4.1.6. Diseño de la Arquitectura Técnica

A partir de la información centralizada en la base de datos se desarrollará un DataMart que permitirá obtener información consolidada de la información que maneja la Dirección Nacional de Justicia generando así reportes gerenciales de manera rápida.

También la obtención y publicación de nuevos reportes se da de manera más ágil y eficiente, la publicación de estos reportes en un entorno web que permita a los usuarios acceder a la información de reportes desde cualquier navegador en la PC de su oficina mediante Internet de la institución.



4.1.7. Selección de Productos e Instalación

El Ministerio dispone de herramientas de software las cuales serán empleadas para el desarrollo del proyecto, por tal motivo no es necesario adquirir nuevos productos.

Las herramientas a utilizar son las siguientes:

Power Builder 10

Power Designer

SQL 2000 Server

Analisis Manager

Excel 2003

4.1.8. Especificación de Aplicaciones para Usuarios Finales

En esta fase se determinan las aplicaciones y qué usuarios tendrán acceso a estos.

Para este proyecto existen tres tipos de Usuarios finales :

El Director de la Dirección Nacional de Justicia: Este usuario tiene un perfil de consulta más estratégico y menos predecibles, para este usuario se desarrollará un tipo de reporte que le permita cambiar los parámetros de las consultas.

El Director de la Dirección de Defensoría de Oficio: Este usuario tiene un nivel de consulta táctico.

Usuarios Supervisores: Estos usuarios tienen un perfil de consulta más operacional para ellos se desarrollarán reportes mas estándares.

4.1.9. Desarrollo de Aplicaciones para Usuarios Finales

En esta fase se construye la vista de los reportes que serán usados por los usuarios finales, dichos reportes están basados en la información que a proporcionado el equipo de requerimientos.

Para la generación de reportes se han considerado desarrollar consultas agrupadas y dentro de cada grupo tener una serie de reportes.

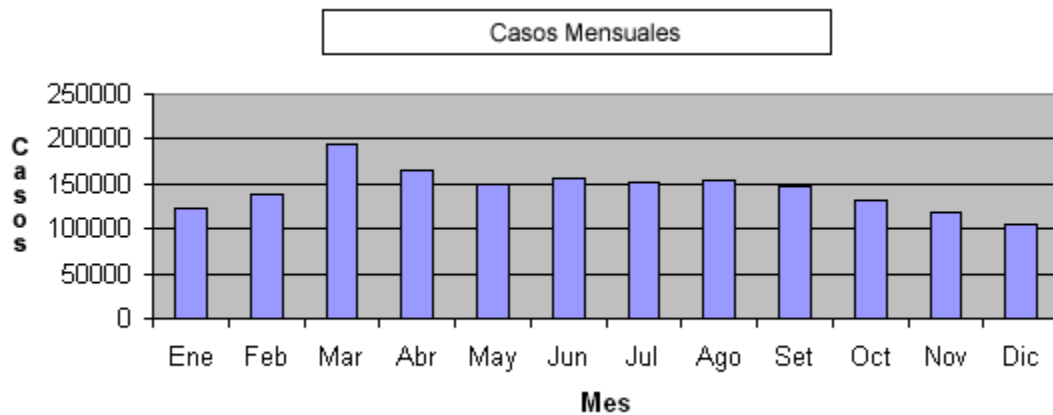
Vistas y Reportes

Tema	Reporte
Análisis de Casos	Casos por Periodo
	Casos por Distrito Judicial
Análisis de Delitos	Delitos mas Frecuentes
	Delitos por Periodo
	Delitos por Sexo
	Delitos por Edad
Análisis de Clientes	Imputados Ingresados por Distrito Judicial y periodo
	Variación Mensual del Ingreso de Imputados
	Número de Imputados Ingresados por Sexo y Periodo.
	Número de Imputados Ingresados por Rango de Edades.

ANÁLISIS DE CASOS

CASOS POR PERIODO

Diseño:



Tipo: Barras

Filas:

No.	Dimensión	Nivel
1	Caso	casos

Columnas:

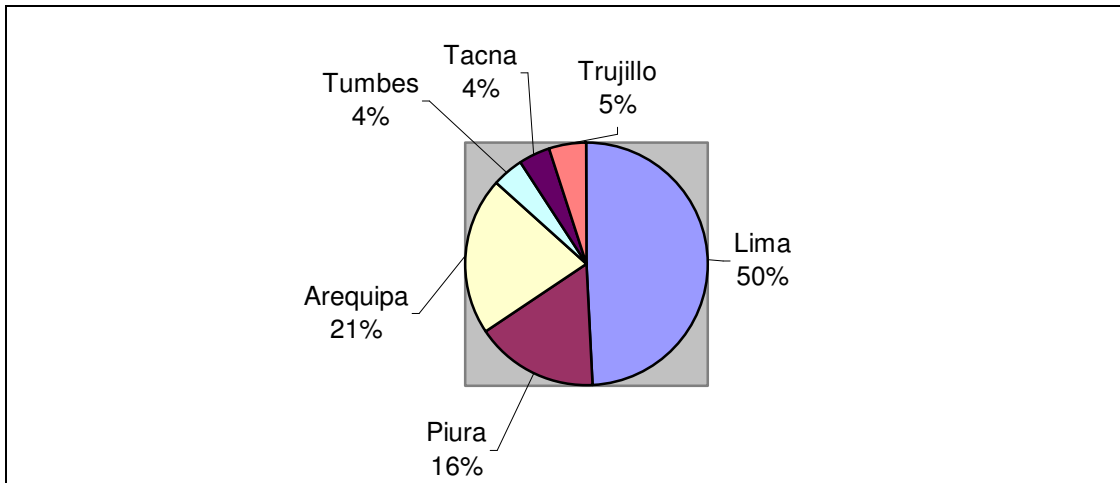
No.	Dimensión	Nivel
1	Tiempo	Fecha

Filtro:

No. Operación	
1	Ubicación = <Nombre de Provincia>
2	Frecuencia = <Frecuencia de Análisis a Ingresar>
3	Fecha Inicial = <Fecha Inicial a Ingresar>
4	Fecha Final = <Fecha Final a Ingresar>

CASOS POR DISTRITO JUDICIAL

Diseño:



0	Casos x Distrito	Distrito						
1								Total
2	Etapa	Lima	Trujillo	Tacna	Tumbes	Arequipa	Piura	
3	Investigacion Preparatoria	xxx	xxx	xxx	xxx	xxx	xxx	xxx
4	Etapa Intermedia	xxx	xxx	xxx	xxx	xxx	xxx	xxx
5	Juicio Oral	xxx	xxx	xxx	xxx	xxx	xxx	xxx
6	Total	xxx	xxx	xxx	xxx	xxx	xxx	xxx

Tipo: Pie y crosstab

Filas:

No.	Dimensión	Nivel
1	caso	etapa

Columnas:

No.	Dimensión	Nivel
1	Dependencia	distritoJudicial

Medida:

No.	Medida	Formato
1	Número de casos	Entero

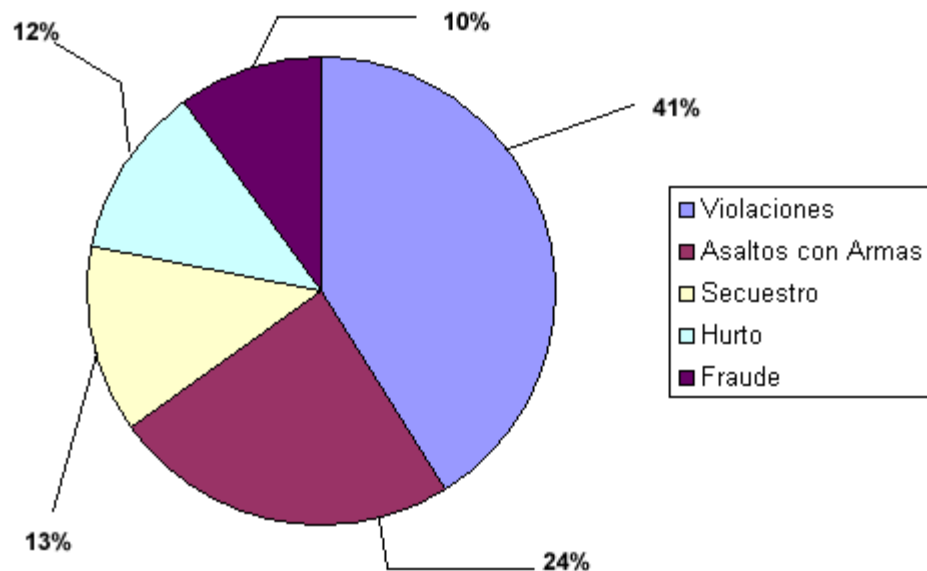
Filtro:

No.	Operación
1	Tipo de Ubicación = <Tipo de Ubicación a Ingresar>
2	Fecha >= <Fecha a Ingresar>
3	Frecuencia = <Frecuencia de Análisis a Ingresar>

ANÁLISIS DE DELITOS

DELITOS MÁS FRECUENTES

Diseño:



Delitos Distrito	Distrito					
Delitos	Lima	Trujillo	Tacna	Tumbes	Arequipa	Total
Violaciones	xxx	xxx	xxx	xxx	xxx	xxx
Asaltos con Armas	xxx	xxx	xxx	xxx	xxx	xxx
Secuestro	xxx	xxx	xxx	xxx	xxx	xxx
Hurto	xxx	xxx	xxx	xxx	xxx	xxx
Fraude	xxx	xxx	xxx	xxx	xxx	xxx
Total	xxx	xxx	xxx	xxx	xxx	xxx

Tipo: Pie y Crosstab

Filas:

No.	Dimensión	Nivel
1	Materia	materia

Columnas:

No.	Dimensión	Nivel
1	Cliente	Ubicación
2	Tiempo	Año, mes, semana, día

Medida:

No.	Medida	Formato
1	Cantidad	Entero

Filtro:

No.	Operación
1	Ubicación = <Ubicación a Ingresar>
2	Fecha >= <Fecha a Ingresar>
3	Frecuencia = <Frecuencia de Análisis >

CUADRO COMPARATIVO DE DELITOS POR PERIODO

Diseño:

Delitos Periodos		Meses												Total
Delitos	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Setiembre	Octubre	Noviembre	Diciembre		
Violaciones	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	
Asaltos con Armas	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	
Secuestro	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	
Hurto	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	
Fraude	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	
Violaciones	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	
Asalto con Armas	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	
Total	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	

Tipo: crosstab

Filas:

No.	Dimensión	Nivel / Categoría
1	Materia	Materia

Columnas:

No.	Dimensión	Nivel / Categoría
1	Tiempo	Meses

Medida:

No.	Medida	Formato
1	Delitos Promedio por Meses	Integer

Filtro:

No.	Operación
1	Delitos = <Delitos a analizar>

No. Operación	
2	periodo = <Año a Ingresar>
3	Periodo 1= <Mes>

CUADRO COMPARATIVO DE DELITOS POR SEXO

Diseño:

Delitos Sexo	Año 2007								
	Trimestre 1		Trimestre 2		Trimestre 3		Trimestre 4		Total
Delitos	Masculino	Femenino	Masculino	Femenino	Masculino	Femenino	Masculino	Femenino	
Violaciones	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
Asaltos con Armas	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
Secuestro	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
Hurto	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
Fraude	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
Total	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx

Tipo: crosstab

Filas:

No.	Dimensión	Nivel / Categoría
1	Delito	delito

Columnas:

No.	Dimensión	Nivel / Categoría
1	Tiempo	Trimestre

Medida:

No.	Medida	Formato
1	Delitos Promedio por Sexo	Integer

Filtro:

No.	Operación
1	Delitos = <Delitos a analizar>
2	periodo = <Año a Ingresar>

No. Operación	
3	Periodo 1= <Trimestre>

ANÁLISIS DE IMPUTADOS

NÚMERO DE IMPUTADOS INGRESADOS POR DISTRITO JUDICIAL Y PERIODO

Diseño:

Imputados	Año	Meses											
	Año 2007												Total
Distrito	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Setiembre	Octubre	Noviembre	Diciembre	
Lima	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
Trujillo	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
Tacna	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
Tumbes	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
Arequipa	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
Piura	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
Lambayeque	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
Total	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx

Tipo: Crosstab

Filas:

No.	Dimensión	Nivel
1	Dependencia	distritoJudicial

Columnas:

No.	Dimensión	Nivel
2	Tiempo	Año, mes

Medida:

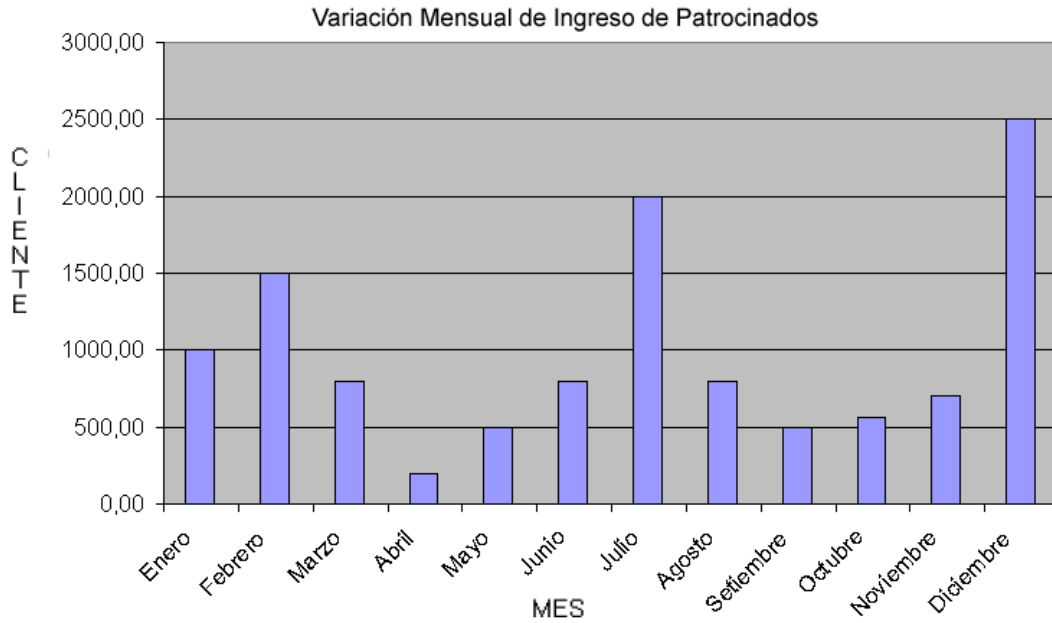
No.	Medida	Formato
1	Num_Imputados_Ingresados	Entero

Filtro:

No. Operación	
1	Ubicación = <Ubicación a Ingresar>
2	Fecha >= <Fecha a Ingresar>
3	Frecuencia = <Frecuencia de Análisis a Ingresar>

VARIACIÓN MENSUAL DEL INGRESO DE IMPUTADOS

Diseño:



Tipo: Barras

Filas:

No.	Dimensión	Nivel
1	Cliente	cliente

Columnas:

No.	Dimensión	Nivel
2	Tiempo	Año

Medida:

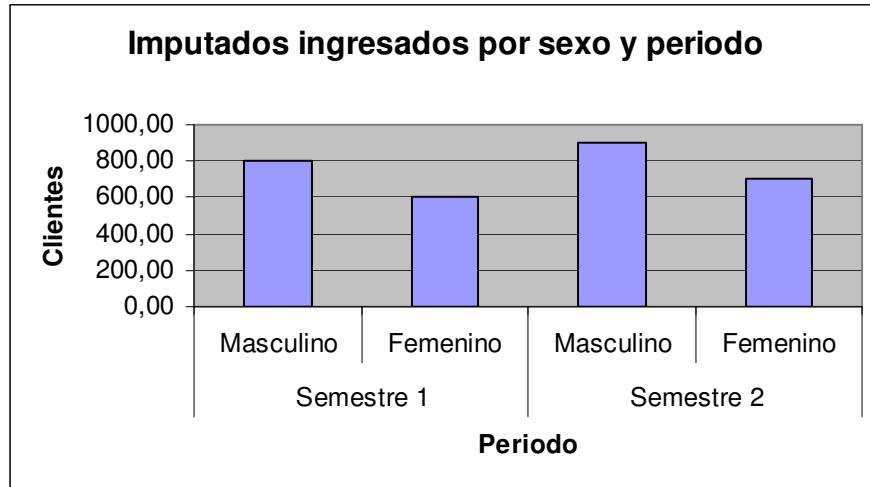
No.	Medida	Formato
1	Num_Imputados_Ingresados	Entero

Filtro:

No.	Operación
1	Ubicación = <Ubicación a Ingresar>
2	Fecha >= <Fecha a Ingresar>
3	Frecuencia = <Frecuencia de Análisis a Ingresar>

DETERMINAR EL NÚMERO DE IMPUTADOS INGRESADOS POR SEXO Y PERIODO.

Diseño:



Imputados	Año	Semestr				
	Año 2007					
	Semestre		Semestre 2		Total	
Distrito	Masculino	Femenino	Masculino	Femenino		
Trujillo	xxx	xxx	xxx	xxx	xxx	
Tacna	xxx	xxx	xxx	xxx	xxx	
Tumbes	xxx	xxx	xxx	xxx	xxx	
Arequipa	xxx	xxx	xxx	xxx	xxx	
Piura	xxx	xxx	xxx	xxx	xxx	
Lambayeque	xxx	xxx	xxx	xxx	xxx	
Total	xxx	xxx	xxx	xxx	xxx	

Tipo: Barras y Crosstab

Filas:

No.	Dimensión	Nivel
1	Dependencia	distritoJudicial

Columnas:

No.	Dimensión	Nivel
2	Tiempo	Año,Semestre

Medida:

No.	Medida	Formato
1	Num_Imputados_Ingresados	Entero

Filtro:

No.	Operación
1	Ubicación = <Ubicación a Ingresar>
2	Fecha >= <Fecha a Ingresar>
3	Frecuencia = <Frecuencia de Análisis a Ingresar>

DETERMINAR EL NÚMERO DE IMPUTADOS INGRESADOS POR RANGO DE EDADES.

Diseño:

Delitos	Año 2007										
	Semestre 1					Semestre 2					Total
Delitos	18 - 25	26 - 31	32 - 37	38 - 43	44 - mas	18 - 25	26 - 31	32 - 37	38 - 43	44 - mas	Total
Lima	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
Trujillo	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
Tacna	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
Tumbes	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
Arequipa	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
Piura	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
Lambayeque	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
Total	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx

Tipo: Crosstab

Filas:

No.	Dimensión	Nivel
1	Dependencia	distritoJudicial

Columnas:

No.	Dimensión	Nivel
2	Tiempo	Año,Semestre

Medida:

No.	Medida	Formato
1	Num_Imputados_Ingresados	Entero

Filtro:

No. Operación	
1	Ubicación = <Ubicación a Ingresar>
2	Fecha >= <Fecha a Ingresar>
3	Frecuencia = <Frecuencia de Análisis a Ingresar>

El siguiente cuadro muestra los objetivos y los reportes que pueden ayudar a alcanzarlos

Objetivos	Reporte
Proveer el diseño de una herramienta integral que ayude a controlar de manera efectiva la asignación y seguimiento de todos los casos registrados como nuevos, en seguimiento y concluidos por los defensores de tal manera de contar con estadísticas y métricas que permitan observar los volúmenes de trabajo de los defensores, así como tomar acciones correctivas para garantizar la calidad de los servicios prestados, optimizando el uso de los recursos de manera progresiva .	Casos por Periodo.
	Casos por Distrito Judicial.
	Relación entre imputados en trámite y terminados ingresados en los últimos doce meses por distrito judicial.
Generar consultas, reportes, estadísticas y métricas que permita evaluar, en todo momento, los niveles de calidad de trabajo del Defensor y la estrategia en relación al trabajo desarrollado por cada uno de ellos .	Delitos terminados con sentencias condenatorias y absolutorias en juicio oral por periodo.
	Relación entre prisión preventiva y otras medidas cautelares.
	Nº de salidas alternativas favorables por defensor.
	Formas de término agrupadas según categoría de delitos.
	El porcentaje de absoluciones en juicio oral de imputados terminados por esta vía procesal.
	Medidas cautelares decretadas por sexo.
Proveer a los niveles directivos del Ministerio de Justicia, un mecanismo capaz de otorgar información acerca del avance global de la gestión de los Defensores para cada una de las áreas de interés, pudiendo establecer métricas comparativas entre el avance real de todas las actividades desarrolladas bajo un determinado periodo de tiempo.	Delitos por Sexo.
	Delitos por Edad.
	Delitos por Periodo.
	Delitos más Frecuentes.
	Imputados Ingresados por Distrito Judicial y periodo.
	Variación Mensual del Ingreso de Imputados.
	Número de Imputados Ingresados por Sexo y Periodo.
	Número de Imputados Ingresados por Rango de Edades.
	Imputados Ingresados por Distrito Judicial y periodo.

4.1.10.Implementación

Para poder realizar la implementación del sistema previamente se debe haber realizado el análisis de los requerimientos de los usuarios del sistema, pudiendo definir así reglas del negocio, luego proceder a realizar el Diseño del Datamart y crear los procesos ETL, que son necesarios para crear el Datamart y crear las consultas.

Cuando se realiza la implementación se debe de tener en cuenta los siguientes puntos:

La capacitación, el soporte técnico, la comunicación, las estrategias de feedback. Todas estas tareas deben ser tenidas en cuenta antes de que cualquier usuario pueda tener acceso al Datamart

Plan de Implementación.- El plan de implementación considera las siguientes actividades:

1. Capacitación del sistema a los usuarios que disponga la Dirección Nacional de Justicia, los mismos que luego podrán capacitar a los usuarios de las diferentes defensorías.
2. Capacitación técnica de la herramienta al personal de Sistemas, con el propósito que puedan dar el soporte informático a los usuarios y el mantenimiento posterior del mismo.
3. Verificación Final de los datos antes de inicio de la operación.
4. Inicio de Operación y soporte de puesta en marcha.

4.1.11. Gerenciamiento del Proyecto

Esta fase de la metodología nos permitió administrar todo el ciclo de vida del proyecto y tomar las medidas preventivas y correctivas para afrontar los diferentes riesgos que se puedan presentar durante la ejecución del proyecto.

Se necesito gestionar los posibles riesgos tales como usuario no identificados con el proyecto. Así como también se realizaron actividades de seguimiento y control en la etapa de desarrollo y prueba.

CONCLUSIONES

Se ha presentado el diseño de un Datamart (prototipo) para la gestión del área de Defensoría del Ministerio de Justicia.

A partir de esta experiencia de diseño a un caso real como es la Defensoría de Oficio, se podrá implementar esta herramienta de inteligencia de negocios.

El diseño que se muestra facilitará la elaboración de reportes que satisfacen los requerimientos de los diferentes tipos de usuarios de nivel directivo.

La implementación de este Datamart ayudará a los Directores en las consultas de reportes diseñados para satisfacer sus requerimientos de forma sencilla a través de una herramienta de fácil aprendizaje.

Utilizar la Metodología de Ralph Kimball ayuda a definir mejor los procesos en el desarrollo del proyecto, pues sirve como guía en todas las etapas del proyecto.

RECOMENDACIONES

- Se requiere un entrenamiento al personal usuario del área de la Dirección Nacional de Justicia para el mejor manejo del Datamart.
- Al diseñar un Datamart se debe pensar en ser parte del Data Warehousing de la Institución.
- Utilizar Herramientas de Reporting para la construcción de consultas avanzadas y visualización de información orientadas al usuario final.
- Es importante utilizar una metodología de Datamart comprobada para garantizar el éxito del proyecto.
- Lo mas importante es enfocarse en las necesidades de la información y no en los datos.
- Siempre es importante incluir a usuarios sponsor en todo el desarrollo de un proyecto de Datamart.

REFERENCIAS BIBLIOGRÁFICAS

[1] Candice Goodwin, “To BI or not to BI”

<http://www.monografias.com/trabajos14/bi/bi.shtml>

2003

[2] ¿Que es Business Intelligence?

<http://www.logicareal.com/neobi.html>

[3] “To BI or not to BI”

<http://www.monografias.com/trabajos14/bi/bi.shtml>

2003

[4] Ralph Kimball, “Una definición de Data Warehousing”

<http://todobi.blogspot.com/2005/05/una-definicion-de-data-warehousing.html>

2005

[5] “La Tecnología Datawarehousing”

www.inf.udec.cl/revista/ediciones/edicion3/cwolff.PDF

2002

[6] "DW: definiciones de Inmon y Kimball"

<http://informationmanagement.wordpress.com/2006/11/28/dw-definiciones-de-inmon-y-kimball/>

[7] Datamart enfocado en el usuario

<http://www.bicase.com/datamarting.doc>

Datawarehouse y sus componentes"

<http://es.wikipedia.org/wiki/Datawarehouse>

"Intelligence y Business Solutions"

http://www.ibss.biz/casos_exito.htm

"Sistemas Decisionales"

<http://gobiernotic.blogspot.com/2007/06/8-claves-para-el-xito-de-un-cuadro-de.html>

"Tecnología Datawarehousing"

<http://html.rincondelvago.com/datawarehousing.html>

"Proceso de Datamining"

http://www.sonda.com/es/global/home/capacidades/business_intelligence/data_mining/

“Soluciones de Inteligencia de Negocios”

http://www.consolti.com/soluciones_de_inteligencia_de_negocios.php

“Generalidades y patrones de desarrollo de Datamarts”

<http://www.intersedes.ucr.ac.cr/index.html>

“Tareas en la implementación de Datwarehouse”

<http://www.virtual.unal.edu.co/cursos/sedes/manizales/4060029/lecciones/cap8-5.html>

www.technologyevaluation.com A definition of Data

Warehousing, Micheal Reed, August 24, 2000.

www.technologyevaluation.com The necessity of Data

Warehousing, Micheal Reed,

August 2, 2000.