

Universidad de las Ciencias Informáticas
Facultad 6



Título: Mercado de Datos para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA.

Trabajo de Diploma para optar por el título de
Ingeniero en Ciencias Informáticas

Autores:

Andry Javier Vichot Chavez
Frank Delgado Rodríguez

Tutores:

Ing. Roberto Tellez Ibarra
Ing. Heriberto Llera Hernández

La Habana, Junio de 2016



La ciencia está hecha de datos, como una casa de piedras. Pero un montón de datos no es ciencia más de lo que un montón de piedras es una casa.

Henri Poincaré

Declaración de autoría

Declaramos ser autores de la presente tesis que tiene por título: Mercado de Datos para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo. Para que así conste firmamos la presente a los ____ días del mes de _____ del año _____.

Andry Javier Vichot Chavez

Firma del Autor

Frank Delgado Rodríguez

Firma del Autor

Ing. Roberto Tellez Ibarra

Firma del Tutor

Ing. Heriberto Llera Hernández

Firma del Tutor

Datos de contacto

Ing. Roberto Tellez Ibarra

Especialidad de graduación: Ingeniería en Ciencias Informáticas

Categoría docente: Asistente

Categoría científica: Ninguna

Años de experiencia en el tema: 7

Años de graduado: 7

Email: rtibarra@uci.cu

Ing. Heriberto Llera Hernández

Especialidad de graduación: Ingeniería en Telecomunicaciones

Categoría docente: Ninguna

Categoría científica: Ninguna

Años de experiencia en el tema: 8

Años de graduado: 25

Email: heriberto.llera@etecsa.cu

Agradecimientos

A Dios, Señor de la historia y de nuestros destinos, por haberme permitido llegar hasta este punto de mi vida personal y profesional, sin Él, nada hubiera sido posible.

A mis padres, por educarme bajo los principios de la responsabilidad, el respeto, la constancia y el sacrificio por aquello que se desea alcanzar.

A esa pequeña parte de mi familia, que a cada paso de mi vida estudiantil sufrió, se agotó, se alegró, en fin, estudió conmigo.

A Maricela, mi gran amiga y ángel de la guarda, que como en tantos momentos difíciles de mi vida, mi carrera universitaria fue también para ella prioridad de preocupación e interés desmedido. De corazón muchas gracias a ti Mari y también a tu hijo Juan por el libro tan útil que me envió y que conservo, así como a tu prima Yoelys que jugó un rol importantísimo en este episodio de mi vida.

A esos amigos míos: Basi, Migue, Pompi, Carlos, Rosa, Odalís, René, Nilmar y otros que en este tiempo la memoria puede ausentar sus nombres, pero no el corazón. En cada momento que simplemente me preguntaban por mis estudios, quizás hasta sin darse cuenta, me animaban y me comprometían a seguir adelante.

A Yanelis, que cuando en el tercer año de la carrera se me cerraba la puerta de la universidad, ella tuvo el gesto maravilloso de aprovecharse de que alguien, no sé aún quien, la dejaba entreabierta, y llamarme en el momento más inesperado para mí, logrando así sorprendentemente, que todo resurgiera como ave Fénix.

A Marisel, la primera que se comprometió de manera oficial, cuando en esa búsqueda desesperada de un tutor me tendió la mano que necesitaba para dar mis primeros pasos en esta investigación.

A mis dos tutores oficiales, Heriberto Llera y Roberto Téllez, que con toda profesionalidad se involucraron con su conocimiento y experiencia en este proyecto para que se hiciera realidad.

A la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA, por brindarme todo el apoyo que necesitaba para lograr este resultado, en especial a Raisa, por ofrecerme la palabra exacta que necesitaba aquel día en que todo parecía derrumbarse.

A cada uno de mis profesores desde mi primaria: Juana Resino, Margarita Novo, Ada y Barbarita, mi profe de Matemática Melba Cuétara en la secundaria y otros tantos hasta la universidad como: Permuy, Santos, Sixto, Neisi, Eddy, Yanio, Dayana, Dany, Eridnier, Yosvanis, Ernesto, José Carlos, Serrano, Garnache, Yadira, Dagmaris, Ariadna, Esley y muchos más.

Y por supuesto a cada uno de mis compañeros de aula, que fueron mis compañeros de viaje en esta desgastante carrera y que en algún momento fueron tan importantes cuando me prestaron su libreta de notas para copiar aquello que no pude escribir o cuando me aclararon una duda minutos antes de un examen y más aún, cuando les confesaba que no entendía algo después de recibir una kilométrica conferencia y ellos sencillamente me respondían: no te preocupes Frank, nosotros tampoco hemos entendido nada.

A todos, Muchas gracias.

Frank,

Agradecimientos

Son muchas las personas a las que tengo que agradecer por la ayuda que me han brindado a lo largo de esta carrera. Primeramente, a mi familia por su apoyo incondicional durante este tiempo, a mi abuela, mi mamá, mi papá y demás miembros. A Grisela, Daili, Cintia, Lacho, Danay, Jeralthine, Ismary y muchas otras personas, quienes fueron mis compañeros de trabajo en los primeros años en la universidad y a aquellos que, no por no mencionarlos signifique que no los tengo presente, sino porque necesitaría un millar de hojas para escribir sus nombres.

Quiero agradecer a Yanelis por enseñarnos a sustituir valores numéricos por “cuadritos, triangulitos, circulitos u otra figura geométrica” en caso de que se nos olvidara un resultado en alguno de sus exámenes con nombre y apellidos impresos. También a Sixto, Armenteros, Yanio, Neisi, Oscar, Eridnier, Eddy, Humberto, Serrano, Garnache, Dagmaris, Ariadna, Javier, Yosvany, Yadira, Ernesto y Carlos, entre otros por haber tenido el placer de tenerlos como profesores.

A Marisel por comprometerse con nosotros como tutora de la tesis al principio cuando el camino aún era angosto y turbio.

A Esley y Leonel por ser de los grandes consultores cada vez que había una duda. A nuestro tutor Tellez por dedicarnos gran parte de su tiempo en momentos difíciles.

A Nilmar por ser el grillo de la consciencia que constantemente me decía en cada uno de los turnos de trabajo “que piensas adelantar el día de hoy, eso está mal, cambia esto, estúdiate el libro tal” y otra serie de regaños constantes que permitieron que las cosas estuviesen listas para cada momento que era necesario.

A Juan Miguel por preocuparse por el curso de la tesis y brindar su ayuda.

A mis compañeros de aula por regalarme seis divertidos años y muchos momentos en los que nos reuníamos para estudiar para un examen y todos estábamos tan perdidos que inventábamos nuevos teoremas.

A mi novia, que no por ser la última es la menos importante, sino todo lo contrario, me daba fuerzas cuando me encontraba exhausto y en otras ocasiones cuando me veía estudiando muchas horas decía “si sigues así se te van a freír los cerebritos como si fueran rositas en un microondas”, algo a lo que nunca pude contener una carcajada.

En fin, a todas aquellas personas que de una forma u otra estuvieron involucradas en la realización de este trabajo.

Muchas gracias a todos.

Andry

Dedicatoria

Es una obligación y gozo dedicar enteramente mi trabajo en esta tesis a mi padre que fue, es y siempre será para mí, el inspirador y responsable de toda mi superación en el orden intelectual, cultural y científico.

Espero que desde el cielo esté satisfecho con lo poco, pero mucho que he logrado hacer.

Frank,

Este trabajo está dedicado a las personas más especiales en mi vida: mis padres. Por su amor, sacrificio, apoyo y confianza que me han sido indispensables para llegar a convertirme en lo que soy. A mi abuela por ser otra madre más y enseñarme tantas cosas de la vida que me sirvieron para lograr mi sueño.

Andry

Resumen

En el mundo moderno, las telecomunicaciones juegan un papel importante en el progreso de la sociedad. En nuestro país, la Empresa de Telecomunicaciones de Cuba (ETECSA), se responsabiliza en todo lo referente a este campo de la ciencia. Un área de gran importancia dentro de esta entidad es la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos, donde se realizan una serie de actividades para la elaboración del presupuesto. Sin embargo, debido al proceso manual de la gestión de la información que se efectuaba en la misma y las herramientas que se utilizaban (paquete de Microsoft Office), el trabajo se hacía muy agotador, lento y propenso a errores. En la presente investigación se describe la solución a esta problemática, teniendo en cuenta las fuentes de información de las que parte este proceso manual, para lo cual se implementó un mercado de datos haciendo uso de la suite *Pentaho* y como gestor de base de datos *PostgreSQL*. La aplicación informática además de mostrar vistas de análisis generales, brinda la posibilidad de generar reportes personalizados, que son utilizados por los especialistas de la Dirección para realizar investigación de mercado y análisis de factibilidad desde distintas perspectivas, teniendo como resultado final la elaboración del presupuesto anual de la división, todo con mayor rapidez, precisión y disminuyendo el desgaste de los recursos humanos.

Palabras clave

Mercadotecnia operativa, mercado de datos, investigación de mercado, análisis de factibilidad, elaboración de presupuesto.

Abstract

In the modern world, telecommunications play an important role in the progress of society. In our country the telecommunications company of Cuba (ETECSA), take responsibility in relation to this field of science. An area of great importance within this entity is the Direction of Operational Marketing Division of Fixed Services, where a series of activities for the elaboration of the budget takes place. However, due to the manual process of the management of the information that was carried out in it and tools that were used (Microsoft Office package) the work was done very exhausting, slow and prone to errors. This research describes the solution to this problem from the information sources where this manual process begin, for which a data mart using Pentaho suite was implemented and as database manager PostgreSQL. The computer application in addition shows views of general analysis, provides the ability to generate custom reports, which are used by specialists of the Direction to carry out market research and analysis of feasibility from different perspectives, resulting in final preparation of the annual budget of the division, all more quickly, precisely and reducing the use of human resources.

Keywords

Operational marketing, data mart, preparation of budget, market research, analysis of feasibility

Tabla de contenido

Introducción	1
Capítulo 1: Fundamentos teóricos sobre el desarrollo de un almacén de datos	4
1.1. Almacenes de datos y Mercados de datos.....	4
1.2. Almacenamiento de la información	6
1.2.1. Topología de esquemas	8
1.3. Metodología para el desarrollo de almacenes de datos	9
1.3.1. Ciclo de vida de la metodología de Kimball.....	11
1.3.2. Metodología de Desarrollo para Proyectos de Almacenes de Datos.....	12
1.4. Técnicas de captura de requisitos.....	14
1.5. Extracción, Transformación y Carga	15
1.6. Inteligencia de Negocio.....	16
1.7. Herramienta de modelado	16
1.8. Sistema Gestor de Base de Datos	17
1.9. Herramientas informáticas para el proceso de Extracción, Transformación y Carga e Inteligencia de Negocio.....	19
Capítulo 2: Análisis y diseño del Mercado de Datos para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA	23
2.1. Caracterización de la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos....	23
2.2. Necesidades de los usuarios	24
2.3. Reglas del negocio	25
2.4. Especificación de requerimientos.....	26
2.4.1. Requerimientos de información	26
2.4.2. Requerimientos funcionales.....	29
2.4.3. Requerimientos no funcionales.....	31
2.5. Casos de Uso del Sistema, Actores del Sistema y Diagrama de Casos de Uso del Sistema	32
2.5.1. Especificación de Casos de Uso del Sistema	34
2.6. Definición de la arquitectura del Mercado de Datos	36
2.7. Diseño de la solución.....	37
2.7.1. Diseño del subsistema de almacenamiento	37
2.7.2. Diseño del subsistema de integración.....	40
2.7.3. Diseño del subsistema de visualización	42
2.8. Política de respaldo	45

2.9. Esquema de seguridad	45
Capítulo 3: Implementación y pruebas del Mercado de Datos para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA.....	47
3.1. Implementación del subsistema de almacenamiento	47
3.1.1. Configuración de la seguridad de la base de datos.....	48
3.2. Implementación del subsistema de integración de datos	49
3.2.1. Limpieza y transformación de datos.....	49
3.2.2. Transformaciones y trabajos.....	49
3.2.3. Carga de datos	52
3.2.4. Gestión del cambio en las dimensiones.....	52
3.2.5. Gestión de los metadatos del proceso de integración.....	53
3.3. Implementación del subsistema de visualización de datos	54
3.3.1. Configurar la seguridad de los usuarios y roles.....	55
3.4. Pruebas	55
3.4.1. Pruebas de software.....	55
3.4.2. Diseño de los Casos de Prueba.....	57
3.5. Resultado de las pruebas	58
Conclusiones generales.....	61
Recomendaciones	62
Referencias bibliográficas	63
Glosario de términos.....	66
Anexos.....	67

Índice de figuras

Fig 1. Ciclo de vida de la metodología de Kimball.	11
Fig 2. Ciclo de vida de la metodología.	13
Fig 3. Diagrama de Casos de Uso del Sistema.	33
Fig 4. Arquitectura del MD Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA.	36
Fig 5. Dimensión dim_sector.	37
Fig 6. Tabla de hecho hech_SISREP.	38
Fig 7. Modelo dimensional.	39
Fig 8. Perfilado de datos del fichero Nauta_2015.xlsx.	41
Fig 9. Diseño general de la implementación de las transformaciones.	42
Fig 10. Estructura de navegación del MD para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos.	43
Fig 11. Diseño del cubo OLAP SISREP con la herramienta Schema Workbench.	44
Fig 12. Estructura física del MD para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA.	48
Fig 13. Adicionar rol Administrador de BD.	49
Fig 14. Transformación “hech_SINBAPLUS_2” para cargar la tabla hech_sinbapplus_2.	50
Fig 15. Trabajo “cargar datamart”.	51
Fig 16. Trabajo “cargar hechos”.	51
Fig 17. Reporte "Cumplimiento de los Ingresos de la División en MLC."	55
Fig 18. Modelo V.	57
Fig 19. Caso de prueba basado en el CUI “Presentar información relacionada con ingresos”.	57
Fig 20. Resultados de las pruebas del sistema.	59
Fig 21. Acta de aceptación del cliente.	67

Índice de tablas

Tabla 1. Transformación de ID_FILIAL en la ETL.	25
Tabla 2. Actores del sistema y sus responsabilidades.	32
Tabla 3. Especificación del Caso de Uso Presentar información relacionada con los ingresos.	34
Tabla 4. Hechos y dimensiones	38
Tabla 5. Matriz de bus o dimensional.....	38
Tabla 6. Descripción del reporte Comportamiento de los Ingresos acumulados en MLC por territorio.	44
Tabla 7. Roles y permisos del sistema.....	46

Introducción

En el mundo moderno, las telecomunicaciones juegan un papel importante en el progreso de la sociedad en general y representan sin duda alguna el desarrollo actual y futuro de la humanidad en sus diferentes esferas: económica, política y social. ETECSA es la entidad responsable en nuestro país en todo lo referente a las telecomunicaciones a nivel nacional (telefonía básica, móvil, datos e internet y otros servicios). Posee una estructura empresarial compleja, que jerárquicamente empezando por su estrato superior, contiene una dirección central (Presidente, Vice-presidente, Directores Centrales), varias divisiones (División de Servicios Móviles, División de Servicios Internacionales, División de Servicios Fijos, División de Telefonía Pública y otras Direcciones Centrales: Logística y Servicio, Negociación y Compras, entre otras) y las direcciones territoriales (una por cada provincia del país) a las cuales se les subordinan los distintos centros de telecomunicaciones. La presente investigación se centra en la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos, la que persigue asegurar, día a día, que el actual producto o servicio es comunicado y llega a los clientes de los segmentos objetivos a través de los canales de distribución escogidos, en condiciones competitivas ventajosas para el comprador (1). Esta dirección está constituida por su director, un especialista principal y otros cuatro especialistas que se subordinan directamente al principal. Entre sus funciones se encuentran: realizar una exhaustiva investigación de mercado y análisis de factibilidad, a partir de un número de indicadores, para crear el presupuesto de la División de Servicios Fijos de la empresa con carácter nacional. En la actualidad las herramientas con las que cuentan este colectivo de trabajo para desempeñar sus tareas, es el paquete de Microsoft Office (específicamente *Microsoft Office Excel*, *Microsoft Office Word*, *Microsoft Office PowerPoint*, *Microsoft Office Outlook* y *Microsoft Office Acces*). Estas herramientas procesan y muestran datos que ellos introducen manualmente a partir de la información que les brindan diferentes sistemas informáticos existentes en la entidad para otros fines específicos (SISREP, SINBAPLUS, SIPREC, CANAL COMERCIAL y NAUTA).

Debido a la utilización de estos sistemas, la información a la que tienen que acceder los especialistas no se encuentra en una única base de datos y se almacena en diferentes formatos, como “Excel”, “PDF” y otros. Estos inconvenientes traen consigo un trabajo agotador y lento debido al tiempo que tienen que emplear en la obtención y análisis de las fuentes, lo que conlleva a un coeficiente de error elevado por la elaboración manual de cada reporte. Teniendo en cuenta la repercusión de estas deficiencias en la misión final de la empresa, pueden provocar un mal funcionamiento de esta, así como pérdida económica significativa para la institución y el país. Se hace necesario por consiguiente, agrupar toda la información generada por estos sistemas permitiendo un sencillo análisis de la misma, que pertenece al área de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA. De esta forma se logrará una mayor rapidez en el proceso de obtención y análisis de información por parte de los especialistas, para que puedan acometer las otras

funciones que tienen asignadas en su contenido de trabajo, las que actualmente se ven limitadas en tiempo y calidad.

Por todo lo antes expuesto, se define como **problema de investigación**: ¿Cómo integrar la información que manejan los especialistas de la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de la empresa ETECSA, para lograr un análisis de dicha información?

La presente investigación tiene como **objeto de estudio**: los almacenes de datos, delimitando el **campo de acción**: a los mercados de datos para Mercadotecnia Operativa.

Con la finalidad de dar solución al problema de la investigación, se identifica como **objetivo general**: desarrollar un Mercado de Datos para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA para lograr un análisis de la información. En correspondencia con el mismo, se plantean los siguientes **objetivos específicos**:

1. Seleccionar la metodología, las herramientas y tecnologías a utilizar en el desarrollo del Mercado de Datos para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA.
2. Realizar el análisis y diseño del Mercado de Datos para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA.
3. Implementar el Mercado de Datos para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA.
4. Realizar pruebas al Mercado de Datos para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA.

Para cumplir con los objetivos planteados se definen las siguientes **tareas de investigación**:

1. Caracterización de las metodologías, herramientas y tecnologías a utilizar en el desarrollo de almacenes de datos.
2. Realización del levantamiento de los requisitos.
3. Descripción de los casos de uso del Mercado de Datos para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos.
4. Definición de la arquitectura del Mercado de Datos para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos.
5. Realización del diseño del Mercado de Datos.
6. Implementación del Mercado de Datos.
7. Aplicación de los casos de prueba.

Para respaldar el cumplimiento de todos los elementos planteados anteriormente, el presente trabajo de diploma estará compuesto por tres capítulos:

Capítulo 1. Fundamentos teóricos sobre el desarrollo de un Almacén de Datos

En el capítulo, se abordan los principales conceptos relacionados con los Almacenes de Datos y los Mercados de Datos, sus principales características y las ventajas que proporciona su utilización en el mundo empresarial. Se exponen además, aspectos fundamentales tales como la caracterización de la metodología de desarrollo a utilizar y las herramientas empleadas en los procesos de integración de datos e Inteligencia de Negocio a nivel internacional para el desarrollo de la solución.

Capítulo 2. Análisis y diseño del Mercado de Datos para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA

En este capítulo se realiza un estudio preliminar del negocio y de la organización, con el fin de obtener las reglas del negocio, identificar los requerimientos y los casos de uso con sus relaciones, describiendo brevemente los actores que van a interactuar con el sistema. Se define la arquitectura del Mercado de Datos y se diseñan los subsistemas de almacenamiento, de integración y visualización.

Capítulo 3. Implementación y pruebas del Mercado de Datos para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA

En este capítulo se hace referencia a la implementación de la solución, abordando específicamente cómo se realiza la implementación del subsistema de almacenamiento, de integración y de visualización, teniendo en cuenta los requerimientos y necesidades del negocio. De igual forma, hace referencia a las pruebas, para determinar que los procesos de Extracción, Transformación y Carga tengan la calidad requerida, así como los procesos de Inteligencia de Negocio.

Capítulo 1: Fundamentos teóricos sobre el desarrollo de un almacén de datos

Introducción

En el capítulo, se abordan los principales conceptos relacionados con los Almacenes de Datos y los Mercados de Datos, sus principales características y las ventajas que proporciona su utilización en el mundo empresarial. Se presentan también, aspectos fundamentales tales como la caracterización de la metodología de desarrollo a utilizar y las herramientas empleadas en los procesos de integración de datos e Inteligencia de Negocio a nivel internacional para el desarrollo de la solución.

1.1. Almacenes de datos y Mercados de datos

Resulta de gran interés analizar e interpretar la información que se genera en una empresa, con el fin de contar con elementos que le permitan tomar decisiones para maximizar sus ganancias o mejorar su rendimiento. Con el surgimiento y desarrollo de las tecnologías, el volumen de información generada ha ido aumentando considerablemente, dando lugar a grandes cantidades de datos históricos, cuyos análisis resultan cada vez más complejos.

Los Almacenes de Datos (AD), surgen como una solución a esta problemática, con el fin de apoyar el proceso de toma de decisiones en las empresas. La definición universalmente aceptada de un AD es la expresada por William H. Inmon, quien plantea que: "...un Almacén de Datos consiste en una colección de datos orientada al negocio, integrada, no volátil y variante en el tiempo, para el apoyo a la toma de decisiones administrativas (2)". Por su parte, Ralph Kimball quien es una de las personalidades más influyentes en el área, propone otra definición al catalogarlo como: "...una copia de datos transaccionales, específicamente estructurados para la consulta y el análisis (3)".

A pesar de plantear conceptos un poco diferentes, los dos convergen en un mismo punto y es que el AD tiene como tarea fundamental, organizar la información recogida de diferentes fuentes para facilitar su análisis y apoyar el proceso de toma de decisiones. Luego de un estudio de la bibliografía existente sobre el tema, se puede concluir que un AD es una colección de datos, que sirve de apoyo para la toma de decisiones administrativas y que debe cumplir las siguientes características:

- **Orientada:** debe estar orientado hacia los aspectos que resulten de interés para la organización. Por consiguiente, la información es organizada por temas, garantizando un acceso rápido y fácil a la misma.
- **Integrada:** un almacén se nutre de fuentes heterogéneas, por lo que resulta muy común encontrar la misma información almacenada de diferentes formas. Por tal motivo, los datos antes de ser cargados, deben pasar por un proceso de integración donde se corrigen, validan y estandarizan, para lograr llevarlos a un formato único que facilite su consulta en el AD.

- **Variable en el tiempo:** los cambios producidos con el paso del tiempo en los datos almacenados también se van registrando en el almacén, lo que permite brindar una perspectiva histórica de la información, que abarca un espacio de tiempo determinado.
- **No volátil:** los datos recogidos en el almacén van a poder ser leídos por los usuarios, pero no podrán ser modificados. Solamente se podrán introducir nuevos datos en el almacén o consultar los que ya se encuentran almacenados (2).

Ventajas y desventajas del uso de un AD

A continuación, se destacan algunas ventajas importantes de los AD:

1. Proporciona una herramienta para la toma de decisiones en cualquier área funcional, basándose en información integrada y global del negocio.
2. Facilita la integración de sistemas de aplicación no integrados.
3. Organiza y almacena los datos que se necesitan para el procesamiento de la información sobre una amplia perspectiva de tiempo.
4. Hace más fácil el acceso a una gran variedad de datos a los usuarios finales.
5. Permite la limpieza de los datos, es decir, eliminan los datos duplicados, erróneos o incompletos que suelen existir en las bases de datos operacionales.
6. Posibilita el ajuste de los datos para posibles combinaciones que se necesiten realizar.
7. Permite el almacenamiento de los datos históricos.
8. Posee mayor rendimiento, pues se tarda mucho menos en acceder a los datos del repositorio del AD que en hacer una consulta a bases de datos distintas.
9. Proporciona la capacidad de aprender de los datos del pasado y de predecir situaciones futuras en diversos escenarios.
10. Permite que los usuarios accedan a la información en línea, contribuyendo a la efectividad para operar en las tareas.

En los AD, también existen inconvenientes o desventajas. Algunas de estas son:

1. A lo largo de su vida se elevan los costos de mantenimiento.
2. Incremento continuo de los requerimientos de los usuarios (4).

Mercados de datos

Un Mercado de Datos (MD) es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica (3). A pesar de que un MD puede ser visto como un AD por poseer las mismas características que estos, los MD se centran solamente en los requerimientos de usuarios asociados a un departamento determinado, no contienen datos operacionales detallados y contienen menos información permitiendo que estos tengan mejor entendimiento y navegabilidad.

1.2. Almacenamiento de la información

Cuando se toma la decisión de crear un MD de un área funcional de la empresa es preciso encontrar la estructura adecuada para el análisis de su información, estructura que puede estar montada sobre una base de datos de Procesamiento Transaccional en Línea (OLTP), como el propio AD, o sobre una base de datos de Procesamiento Analítico en Línea (OLAP). La designación de una u otra dependerá de los datos, los requisitos y las características específicas de cada departamento.

Características de la base de datos OLTP y OLAP

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

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

Los sistemas **OLAP** son bases de datos orientadas al procesamiento analítico. Este análisis suele implicar, generalmente, la lectura de grandes cantidades de datos para llegar a extraer algún tipo de información útil: tendencias de ventas, patrones de comportamiento de los consumidores, elaboración de informes complejos u otro elemento. Este sistema es típico de los MD:

- El acceso a los datos suele ser de solo lectura. La acción más común es la consulta, con muy pocas inserciones, actualizaciones o eliminaciones.
- Los datos se estructuran según las áreas de negocio, y los formatos de los datos están integrados de manera uniforme en toda la organización.
- El historial de datos es a largo plazo, normalmente se realiza de dos a cinco años.
- Las bases de datos OLAP se suelen alimentar de información procedente de los sistemas operacionales existentes, mediante un proceso de Extracción, Transformación y Carga (5).

Los sistemas OLAP se subdividen en tres tipos de arquitectura:

- ✓ Sistemas MOLAP
- ✓ Sistemas ROLAP
- ✓ Sistemas HOLAP

Sistemas MOLAP

La arquitectura MOLAP usa unas bases de datos multidimensionales propietarias para proporcionar el análisis, su principal premisa es que el OLAP está mejor implementado almacenando los datos multidimensionalmente. Esto permite que los sistemas MOLAP tengan mejor tiempo de respuesta, una mejor estructura para maximizar las consultas, lo que constituye en sistema apropiado para cubos de rápida respuesta.

Sistemas ROLAP

La arquitectura ROLAP, accede a los datos almacenados en un AD para proporcionar los análisis OLAP con la premisa de que las capacidades OLAP se soportan mejor contra las bases de datos relacionales. El sistema ROLAP utiliza una arquitectura de tres niveles. La base de datos relacional maneja los requerimientos de almacenamiento de datos, y el motor ROLAP proporciona la funcionalidad analítica. El nivel de base de datos usa bases de datos relacionales para el manejo, acceso y obtención del dato. El nivel de aplicación es el motor que ejecuta las consultas multidimensionales de los usuarios.

El motor ROLAP se integra con niveles de presentación, a través de los cuales los usuarios realizan los análisis OLAP. Después de que el modelo de datos para el AD se ha definido, los datos se cargan desde el sistema operacional y se ejecutan rutinas de bases de datos para agregar el dato, si así es requerido por el modelo de datos. Se crean entonces los índices para optimizar los tiempos de acceso a las consultas.

Los usuarios finales ejecutan sus análisis multidimensionales, a través del motor ROLAP, que transforma dinámicamente sus consultas a consultas SQL. Se ejecutan estas consultas SQL en las bases de datos relacionales, y sus resultados se relacionan mediante tablas cruzadas y conjuntos multidimensionales para devolver los resultados a los usuarios.

La arquitectura ROLAP es capaz de usar datos precalculados si estos están disponibles, o generar dinámicamente los resultados desde los datos elementales si es preciso. Esta arquitectura accede directamente a los datos del AD, y soporta técnicas de optimización de accesos para acelerar las consultas.

Sistemas HOLAP

Un desarrollo un poco más reciente ha sido la solución OLAP híbrida (HOLAP), la cual combina las arquitecturas ROLAP y MOLAP para brindar una solución con las mejores características de ambas: desempeño superior y gran escalabilidad. Un tipo de HOLAP mantiene los registros de detalle (los volúmenes más grandes) en la base de datos relacional, mientras que mantiene las agregaciones en un almacén MOLAP separado (6).

Teniendo en cuenta las potencialidades de los sistemas ROLAP, a los cuales se ajustan los recursos disponibles, se seleccionó esta arquitectura.

1.2.1. Topología de esquemas

Existen tres tipos de esquemas para realizar el modelo de datos físico, los cuales se deben analizar para seleccionar el más adecuado según los requerimientos y necesidades del cliente, ellos son:

- ✓ Esquema en estrella
- ✓ Esquema copo de nieve
- ✓ Esquema constelación de hechos

Esquema en estrella

El esquema en estrella, consta de una tabla de hechos central y de varias tablas de dimensiones relacionadas a esta, a través de sus respectivas claves. El esquema en estrella es el más simple de interpretar y optimiza los tiempos de respuesta ante las consultas de los usuarios. Este modelo es soportado por casi todas las herramientas de consulta y análisis, y los metadatos son fáciles de documentar y mantener, sin embargo, es el menos robusto para la carga y es el más lento de construir.

Esquema copo de nieve

Este esquema representa una extensión del modelo en estrella cuando las tablas de dimensiones se organizan en jerarquías de dimensiones. Para dicho esquema existe una tabla de hechos central que está relacionada con una o más tablas de dimensiones, quienes a su vez pueden estar relacionadas o no con una o más tablas de dimensiones. Este modelo es más cercano a un modelo de entidad relación, que al modelo en estrella, debido a que sus tablas de dimensiones están normalizadas.

Uno de los motivos principales de utilizar este tipo de modelo, es la posibilidad de segregar los datos de las tablas de dimensiones y proveer un esquema que sustente los requerimientos de diseño. Otra razón es que es muy flexible y puede implementarse después de que se haya desarrollado un esquema en estrella.

Esquema constelación de hechos

Este modelo está compuesto por una serie de esquemas en estrella, donde existe más de una tabla de hechos. Dichas tablas yacen en el centro del modelo y están relacionadas con sus respectivas tablas de dimensiones, en ocasiones dos o más hechos comparten una misma dimensión, aunque no es necesario que las diferentes tablas de hechos compartan las mismas tablas de dimensiones (7). Dada las características de nuestra problemática la topología a utilizar es constelación de hechos.

1.3. Metodología para el desarrollo de almacenes de datos

Una metodología es el conjunto de métodos por los cuales se regirá una investigación científica. No existe una única metodología en la cual basarse para la construcción de un AD, sino que dependiendo del contexto en que se encuentre la empresa y los objetivos que persiga, se puede emplear una u otra. Estas metodologías se engloban dentro de dos grandes enfoques: descendente (*top-down*) y ascendente (*bottom-up*) que se corresponden con las metodologías propuestas por Bill Inmon y Ralph Kimball respectivamente. El enfoque descendente se adapta a la visión de Bill Inmon y considera que los pequeños almacenes departamentales o MD, se nutrirán del AD. Propone la creación de un repositorio de datos corporativo como fuente de información consolidada, consistente e histórica. Al ser construido descendentemente, los MD se nutren del AD corporativo, convirtiéndose en un complejo empresarial de bases de datos relacionales. Es un enfoque sistémico, que minimiza los problemas de integración, pero es costoso, debido a la gran cantidad de datos y su poca flexibilidad.

Al contrario de Inmon, Kimball plantea que el AD se compone por el conglomerado de todos los MD generados en una empresa y la información siempre se almacena en un modelo dimensional. Propone

construir MD independientes para evaluar las ventajas del nuevo sistema a medida que se avanza en su construcción. En él, las partes individuales se diseñan con detalle y luego se enlazan para formar componentes más grandes, que a su vez se acoplan hasta que se forma el sistema completo. Esta característica lo hace más flexible y sencillo de implementar, además, este enfoque parte de los requisitos del negocio, mientras que el enfoque descendente propone la validación de los requisitos una vez que se tiene el sistema.

En ciertos casos, las metodologías de desarrollo de AD son extensiones de las metodologías clásicas para bases de datos, en otros casos se ha adoptado un enfoque completamente nuevo, por ello existen otras metodologías que no siguen específicamente ninguno de los enfoques anteriormente mencionados, sino que realizan una selección de lo mejor de cada uno y definen su propia metodología.

A continuación, se describen brevemente algunas de ellas:

- **Metodología SQLBI:** avalada por Microsoft y orientada totalmente a sus herramientas: Microsoft SQL Server, SQL Server Analysis Services y su oferta más completa en este campo que es Microsoft Suite for Business Intelligence (8).
- **Hefesto:** entre sus principales directrices plantea que la construcción e implementación de un AD puede adaptarse muy bien a cualquier ciclo de vida de desarrollo de software, con la salvedad de que, para algunas fases en particular, las acciones que se han de realizar serán muy diferentes. Lo que se busca, es entregar una primera implementación que satisfaga una parte de las necesidades, para demostrar las ventajas del AD y motivar a los usuarios (9).
- **Metodología para el Diseño Conceptual de Almacenes de Datos:** esta última es presentada en la Tesis de Doctorado de Leopoldo Zenaido Zepeda Sánchez. Aporta como aspecto novedoso con respecto a las anteriores, la incorporación de una serie de transformaciones para llevar un diagrama relacional a uno dimensional y así obtener las estructuras candidatas que conformarán el Repositorio de Datos (10).

Como se observa, en las metodologías existentes, los autores las orientan a la optimización del rendimiento y a su visión de los principales procesos que se deben tener en cuenta para construir un AD flexible y dinámico intentando dar un acercamiento a una propuesta ideal para su desarrollo.

1.3.1. Ciclo de vida de la metodología de Kimball

El ciclo de vida Kimball comienza con una planificación de proyecto, donde se define el alcance, se identifican y programan las tareas, se planifica el uso de los recursos, conformando con todo esto, el plan de proyecto. En la segunda etapa se definen los requerimientos del negocio. Luego, el proyecto se enfoca en tres líneas concurrentes: tecnología, datos y aplicaciones de Inteligencia de Negocio. El ciclo de vida culmina con el despliegue y mantenimiento del producto. La Figura 1 muestra cada una de estas etapas.

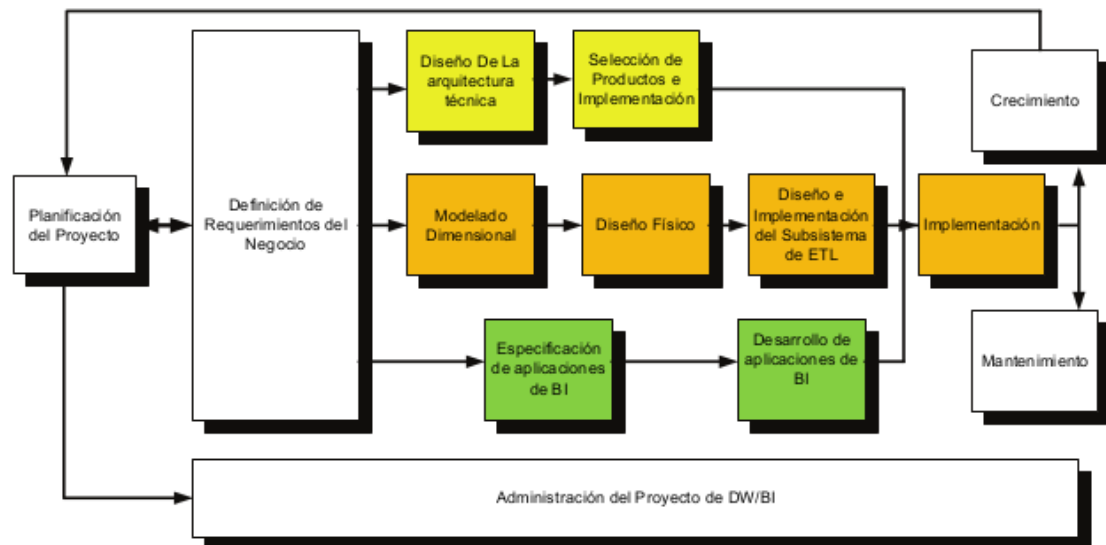


Fig 1. Ciclo de vida de la metodología de Kimball.

Para determinar la metodología de desarrollo a utilizar, se tomó como base la Metodología de Kimball por los siguientes elementos:

- Utiliza el Modelo Dimensional, muy eficaz en el proceso de la toma de decisiones que proporciona mayor agilidad en el desarrollo de un AD.
- Propone ir elaborando el AD a través de la construcción de los MD, lo que constituye una buena estrategia.
- Existe abundante documentación sobre la misma y se puede consultar vía web a través de los servicios que brindan el grupo creador de la metodología.
- Es una metodología madura y reconocida internacionalmente (11).

A pesar de las ventajas que ofrece la utilización de la Metodología de Kimball, esta no satisface las necesidades por las siguientes razones:

- No tiene definido un criterio que permita estimar los costos de desarrollo de un AD, basándose en las características de la construcción del mismo.
- Presenta un grupo de roles, pero no explica claramente cuáles son las competencias y responsabilidades de cada uno dentro del proyecto. Por la cantidad de roles que propone se necesita de grupos grandes para su desarrollo.
- Propone un gran número de actividades y artefactos que pueden extender los tiempos de desarrollo si se cuenta con pocos recursos humanos, además, no se especifica cómo deben realizarse estos artefactos.
- Está estructurada para el desarrollo de proyectos – productos, donde un proyecto desarrolla un producto determinado.
- No establece el análisis de diferentes criterios de diseño en el levantamiento de requisitos que permita la construcción más adecuada del almacén, teniendo en cuenta las metas de la organización, las necesidades de los usuarios y la disponibilidad de las fuentes operaciones (12).

1.3.2. Metodología de Desarrollo para Proyectos de Almacenes de Datos

Debido a las razones anteriores, se necesita una metodología que tenga como base la de Kimball y corrija las deficiencias antes expuestas, por ello se optó por la Metodología de Desarrollo para Proyectos de Almacenes de Datos, de Yanisbel González Hernández (11), la cual se basa en el ciclo de vida Kimball en la que incluye los casos de uso para guiar el proceso de desarrollo. Está adaptada a las necesidades de la Universidad de las Ciencias Informáticas (UCI), y cubre las fases por las que pasa la construcción de un AD.

Las características que presenta esta metodología es la identificación de requerimientos de información y a su vez, la trazabilidad que tienen estos en todo el ciclo de desarrollo del AD. También se adiciona una fase de pruebas que fortalece en gran medida la calidad con que se despliegue la solución propuesta. Además, se ajusta a las fases, actividades y artefactos del Programa de Mejora del Centro Nacional de Calidad del Software (CALISOFT), al nivel 2 del Modelo de Integración de Capacidades de Madurez (del inglés *Capability Maturity Model Integration*, CMMI) en la UCI. La Figura 2 muestra el ciclo de vida de la metodología.

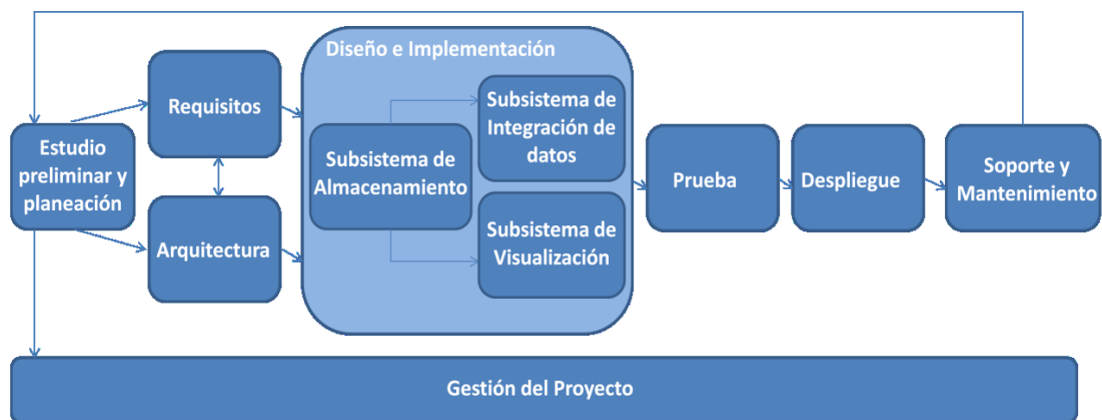


Fig 2. Ciclo de vida de la metodología.

El ciclo de vida de la metodología está organizado por fases, las cuales podrán ser implementadas de forma paralela según el componente que se está desarrollando para integrarse al final de la solución. A continuación se describen las fases del ciclo de vida de la metodología:

- **Estudio preliminar y planeación:** se realiza el estudio de la entidad cliente para determinar lo que se desea construir y las condiciones que existen para el desarrollo de la misma. En la planeación del proyecto se definen los objetivos, el alcance preliminar, los costos estimados y otras actividades.
- **Levantamiento de requisitos:** se realiza en tres direcciones; una, identificando las metas y objetivos de la organización; dos, identificando las necesidades de información y reglas del negocio; y tres, realizando un levantamiento detallado de las fuentes de datos a integrar. Es aquí donde se definen los requerimientos a través de la comparación de las necesidades y las reglas del negocio.
- **Arquitectura:** se define la arquitectura de la solución según los requisitos no funcionales obtenidos. Se definen aspectos como: la seguridad del sistema, la comunicación entre los subsistemas, la tecnología a utilizar, hardware y software, entre otros aspectos de gran importancia.
- **Diseño e implementación:** se define el diseño de las estructuras de almacenamiento, se diseñan los procesos de integración de datos como, las reglas de ETL, se diseñan los cubos para la presentación de los datos, así como el diseño visual de la aplicación definido por el cliente. Después se implementan cada uno de los subsistemas (almacenamiento, integración y visualización). Se lleva a cabo el diseño físico del Repositorio de Datos, se crean las estructuras de almacenamiento con las particiones y agregaciones correspondientes. Se crea el área temporal de almacenamiento, se ejecutan las reglas de ETL, haciendo los ajustes para

integrar la información. Se configuran e implementan las herramientas de Inteligencia de Negocio para obtener los reportes, gráficos, mapas y otros que cubran los requerimientos firmados con el cliente final.

- **Prueba:** se realizan las pruebas de unidad, luego las pruebas de integración y sistema, hasta llegar a las pruebas de aceptación con el cliente final.
- **Despliegue:** consta de dos etapas, despliegue piloto en el cual se configuran los servidores, se instalan las herramientas según la arquitectura definida y se carga una muestra de los datos para demostrar al cliente que el sistema funciona. Posterior a la aceptación del cliente, se realiza la carga histórica de los datos, la capacitación y la transferencia tecnológica. El resultado fundamental es la solución desplegada en el entorno real y en correcto funcionamiento.
- **Soporte y mantenimiento:** después de haber implantado la solución, se brindan los servicios de soporte en línea, vía telefónica, web u otros, hasta el acompañamiento junto al cliente según el contrato firmado y las condiciones de soporte establecidas.
- **Gestión y administración del proyecto:** se lleva durante todo el ciclo de vida, es aquí donde se controla, gestiona y chequea todo el desarrollo, los gastos, las utilidades, los recursos, las adquisiciones, y demás actividades relacionadas con la gestión del proyecto (12).

Por las características antes descritas y beneficios de la metodología, se considera que es la adecuada para el desarrollo del sistema a implementar como solución. En la presente investigación se ejecuta el sistema hasta la fase Prueba.

1.4. Técnicas de captura de requisitos

La obtención de requisitos en el desarrollo de los sistemas informáticos es importante para verificar si los resultados obtenidos logran satisfacer las necesidades del cliente, desde sus inicios los ingenieros han presentado dificultades para obtener los requisitos debido a que la información a veces proviene de diferentes personas y los datos pueden variar. Con el objetivo de darle solución a dichas dificultades se han desarrollado técnicas que permiten realizar el proceso de manera más segura. Algunas de las más utilizadas son: entrevistas, tormenta de ideas, cuestionarios, observaciones, discusiones, análisis de protocolo y casos de uso (13). En la presente investigación se utilizaron la observación y las entrevistas, permitiendo tener una mejor comprensión del problema y los objetivos de la solución, posibilitando tener una amplia visión del negocio y de las necesidades de los clientes. Otras de las técnicas que se utilizaron fueron las discusiones con los especialistas, para aclarar dudas y poder realizar una correcta captura de los requisitos.

1.5. Extracción, Transformación y Carga

Para el desarrollo de un AD es necesario realizar la Extracción, Transformación y Carga (del inglés *Extract-Transform-Load*, ETL) de los datos pertenecientes a una empresa. ETL es el proceso que organiza el flujo de los datos entre diferentes sistemas en una organización y aporta los métodos y herramientas necesarias para mover datos desde múltiples fuentes a un AD, reformatearlos, limpiarlos y cargarlos en otra base de datos o MD. ETL forma parte de la Inteligencia de Negocio (14).

Extracción: consiste en localizar y extraer los datos, ya que en la mayoría de los proyectos de almacenamiento se fusionan datos provenientes de diferentes sistemas de origen tales como ERP, CRM, bases de datos, planillas Excel, entre otros tipos. Los formatos de las fuentes normalmente se encuentran en bases de datos relacionales o ficheros planos, pero pueden incluir también bases de datos no relacionales y otras estructuras o formatos diferentes.

Transformación: la fase de transformación de un proceso de ETL aplica una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados. Algunas fuentes de datos requerirán alguna pequeña manipulación de los datos.

Carga: la fase de carga es el momento en el cual los datos de la fase anterior (transformación) son cargados en el sistema de destino. Dependiendo de los requerimientos de la organización, este proceso puede abarcar una amplia variedad de acciones diferentes. En algunas bases de datos se sobrescribe la información antigua con nuevos datos. Dado que los AD mantienen un historial de los registros, se puede hacer una auditoría de los mismos y disponer de un rastro de toda la historia de un valor a lo largo del tiempo.

Existen dos formas básicas de desarrollar el proceso de carga:

- **Acumulación simple:** la acumulación simple es la más sencilla y común, y consiste en realizar un resumen de todas las transacciones comprendidas en el período de tiempo seleccionado y transportar el resultado como una única transacción hacia el AD, almacenando un valor calculado que consistirá típicamente en un sumatorio o un promedio de la magnitud considerada.
- **Rolling:** el proceso de Rolling por su parte, se aplica en los casos en que se opta por mantener varios niveles de granularidad. Para ello se almacena información resumida a distintos niveles, correspondientes a distintas agrupaciones de la unidad de tiempo o diferentes niveles jerárquicos en

alguna o varias de las dimensiones de la magnitud almacenada (por ejemplo, totales diarios, totales semanales, totales mensuales, entre otros) (15).

1.6. Inteligencia de Negocio

La Inteligencia de Negocio es la habilidad para transformar los datos en información, y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios. Desde un punto de vista más pragmático, y asociándolo directamente con las tecnologías de la información, podemos definir Inteligencia de Negocio como el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada (interna y externa a la compañía) en información estructurada, para su explotación directa (reporte, análisis OLTP / OLAP, alertas) o para su análisis y conversión en conocimiento, dando así soporte a la toma de decisiones sobre el negocio.

La Inteligencia de Negocio actúa como un factor estratégico para una empresa u organización, generando una potencial ventaja competitiva, que no es otra que proporcionar información privilegiada para responder a los problemas del negocio: entrada a nuevos mercados, promociones u ofertas de productos, eliminación de islas de información, control financiero, optimización de costes, planificación de la producción, análisis de perfiles de clientes, rentabilidad de un producto concreto, y otras actividades (16).

1.7. Herramienta de modelado

Día a día la tecnología avanza, surgen nuevas y mejores formas de hacer las cosas, siempre buscando métodos más efectivos, confiables, con mayor calidad y menos riesgos. Las herramientas CASE (Computer Aided Software Engineering, Ingeniería de Software Asistida por Computadora) son un conjunto de herramientas y métodos asociados que proporcionan asistencia automatizada en el proceso de desarrollo del software a lo largo de su ciclo de vida. Fueron desarrolladas para automatizar esos procesos y facilitar las tareas de coordinación de los eventos que necesitan ser mejorados en el ciclo de desarrollo de software (17).

El Lenguaje Unificado de Modelado (Unified Modeling Lenguaje UML), es un lenguaje estándar para escribir planos de software, UML se puede utilizar para visualizar, especificar, construir y documentar los artefactos de un sistema que involucra una gran cantidad de software. UML prescribe un conjunto de notaciones y diagramas estándar para modelar sistemas orientados a objetos, y describe la semántica esencial de lo que estos diagramas y símbolos significan (18).

Para la presente investigación se decidió utilizar Visual Paradigm para UML en su versión 8.0 por las siguientes razones:

- Maximiza y acelera al equipo de desarrollo en conjunto a los programadores.
- Provee el modelado de procesos de negocios.
- Generador de código de base de datos.
- Proporciona el código y compatibilidad hasta con 10 lenguajes.

1.8. Sistema Gestor de Base de Datos

El Sistema de Gestión de Base de Datos (SGBD), es un tipo de software muy específico, dedicado a servir de interfaz entre la base de datos, los usuarios y las aplicaciones que lo utilizan. Se compone de lenguajes de definición, manipulación, consulta y seguridad de datos. El propósito general de los SGBD es el de manejar de manera clara, sencilla y ordenada un conjunto de datos.

Existen diferentes objetivos que deben cumplir los SGBD de los cuales se han enumerado los siguientes:

- Hacer transparente a los usuarios los detalles del almacenamiento físico de los datos, mediante varios niveles de abstracción de la información.
- Permitir la realización de cambios a la estructura de la base de datos, sin tener que modificar la aplicación que la emplea.
- Proveer a los usuarios la seguridad de que sus datos no podrán ser accedidos, ni manipulados por quien no tenga permiso para ello. Debido a esto, debe poseer un complejo sistema que maneje grupos, usuarios y permisos para las diferentes actividades que se pueden realizar dentro del mismo.
- Mantener la integridad de los datos.
- Proporcionar una manera eficiente de realizar copias de seguridad de la información almacenada en ellos, y permitir a partir de estas copias restaurar los datos.
- Controlar el acceso concurrente de los usuarios.
- Facilitar el manejo de grandes volúmenes de información.

Existen dos tipos de SGBD:

- **SGBD Multidimensionales:** estos aportan mucho rendimiento al AD en cuanto a la velocidad de respuesta, ya que los datos son almacenados en forma multidimensional, sin embargo, son difíciles de gestionar y de mantener.
- **SGBD Relacionales:** estos son cada vez más potentes y poseen una interfaz gráfica más avanzada (19).

PostgreSQL

Es un SGBD objeto-relacional, con su código fuente disponible libremente. Utiliza una arquitectura cliente/servidor y usa multiprocesos para garantizar la estabilidad del sistema, es decir, un fallo en uno de los procesos no afectará el resto y el sistema continuará funcionando. Además, es un sistema que funciona de manera excelente con grandes cantidades de datos y una gran concurrencia de usuarios. Se decidió utilizar PostgreSQL 9.2.

A continuación, se enumeran algunas de las características más importantes y soportadas por PostgreSQL:

- Juegos de caracteres internacionales
- Múltiples métodos de autenticación
- Completa documentación
- Licencia BSD
- Disponible para Linux y UNIX en todas sus variantes (AIX, BSD, HP-UX, SGI IRIX, Mac OS X, Solaris, Tru64) y Windows 32/64bit.
- El máximo de bases de datos que posee es ilimitado.
- Posee un máximo de tamaño de tabla de 32 TB.
- El máximo de tamaño de registro es de 1.6 TB.
- El máximo de tamaño de campo es de 1GB.
- El máximo de registros por tabla es ilimitado.
- El máximo de campos por tabla es de 250 a 1600 (depende de los tipos usados).
- El máximo de índices por tabla es ilimitado (20).

PgAdmin

PgAdmin es una plataforma de desarrollo y administración de código abierto para el SGBD PostgreSQL. Es multiplataforma debido a que puede ser usada en Linux, FreeBSD, Solaris, Mac OSX y Windows para gestionar dicho SGBD, así como las versiones derivadas y comerciales como Postgres Plus Advanced

Server y Greenplum database. Entre algunas de las funcionalidades que incluye la aplicación se encuentra un editor de sobresaltado de sintaxis SQL (syntax high lighting SQL editor) y un editor de código seguro del lado del servidor (server-side code editor). También, puede realizarse la conexión mediante TCP/IP o Unix Domain Sockets, y puede ser encriptada mediante SSL para mayor seguridad.

La aplicación es desarrollada por una comunidad de expertos de PostgreSQL de todo el mundo, por lo que es un software libre lanzado bajo la licencia de PostgreSQL y está disponible en más de una docena de idiomas. Además, tiene acceso a todos los objetos de PostgreSQL, los cuales son mostrados con su definición SQL y una lista de propiedades. También se pueden explorar los objetos dependientes, las dependencias y las estadísticas de los objetos consultados en cada caso (21).

En la presente investigación se escogió PgAdmin III en su versión 1.16 como herramienta de administración, pues está diseñada para responder, desde simples consultas SQL hasta el desarrollo de complejas bases de datos; y la interfaz gráfica que posee es compatible con todas las características de PostgreSQL, ya que es el cliente de base de datos oficial.

1.9. Herramientas informáticas para el proceso de Extracción, Transformación y Carga e Inteligencia de Negocio

En el desarrollo de un MD es importante realizar una adecuada selección de las herramientas informáticas para los procesos de ETL e Inteligencia de Negocio, que garanticen una solución con calidad. Existen diversas herramientas en el mercado para estos fines que permiten cumplir los objetivos de la investigación, con mayor o menor potencialidad, algunas son privativas y otras libres.

Pentaho Data Integration

En la presente investigación se determinó utilizar Pentaho Data Integration en su versión 5.4.0.1 para el desarrollo de los procesos de ETL, ya que es una herramienta libre, muy potente y una de las más utilizadas por los usuarios, considerándola la más completa por la gran cantidad de conectores que posee y la posibilidad de crear flujos de trabajo integrados con transformaciones de datos de manera muy sencilla y funcional. Dentro de las principales características que posee dicha herramienta se encuentran:

Entorno gráfico de desarrollo

- Uso de tecnologías estándar: Java, XML, JavaScript
- Fácil de instalar y configurar
- Multiplataforma: Windows, Macintosh y Linux

- Basado en dos tipos de objetos: Transformaciones (colección de pasos en un proceso ETL) y Trabajos (colección de transformaciones)

Incluye cuatro herramientas:

- **Spoon:** para diseñar transformaciones ETL usando el entorno gráfico.
- **PAN:** para ejecutar transformaciones diseñadas con spoon.
- **CHEF:** para crear trabajos.
- **Kitchen:** para ejecutar trabajos (22).

Mondrian OLAP Server

Mondrian es un servidor OLAP de código abierto que gestiona la comunicación entre una aplicación OLAP (escrita en Java) y la base de datos con los datos fuente. Además, permite crear cubos OLAP para el análisis multidimensional, y combinado con Jpivot, el servidor OLAP Mondrian, permite realizar consultas al AD, lo que posibilita que los resultados sean presentados a través del navegador (23). En la presente investigación se utilizó la versión 3.10.0.1 de Mondrian. Entre sus ventajas se encuentran:

- Hace posible la agilización de consultas de grandes cantidades de datos.
- Posee una alta velocidad de respuesta.
- Permite realizar consultas al MD.
- Es un motor ROLAP con caché.

Schema Workbench

Pentaho Schema Workbench es una herramienta de análisis caracterizada por su potencia gráfica y su capacidad multitarea. Puede utilizarse para ingeniería inversa a una base de datos, a un modelo para visualizarlo mejor o realizar su mantenimiento. El esquema de Mondrian Workbench es una interfaz de diseño que permite crear y probar los cubos OLAP visualmente.

El motor de Mondrian procesa las solicitudes de MDX con el ROLAP (Relational OLAP). Estos archivos son los modelos de esquemas de metadatos XML creados en una estructura específica que utiliza el motor de Mondrian. Estos modelos XML pueden ser considerados como cubos. Con esta aplicación, se puede configurar una conexión JDBC con el modelo físico, para luego elaborar el esquema lógico de manera simple y efectiva (24). Se seleccionó la versión 3.10.0.1 para la investigación.

Report Designer

Pentaho Report Designer es un potente generador de informes con prestaciones profesionales, con capacidad de personalización de informes a las necesidades de los negocios, destinado a desarrolladores que permite la distribución de los resultados de los análisis en múltiples formatos. Está estructurada de forma que los desarrolladores puedan acceder a sus prestaciones de forma rápida. Incluye un editor de consultas para facilitar la confección de los datos que serán utilizados en un informe (25). Se utilizó la versión 5.4.0.1.

Pentaho BI Server

Es una plataforma de Inteligencia de Negocio “orientada a la solución” y “centrada en procesos”. Para la solución se escogió la versión 5.4.0.1. Pentaho consiste en una Suite completa de Inteligencia de Negocio que incluye todos los principales componentes requeridos para implementar soluciones de Inteligencia de Negocio, tales como:

- Informes
- Dashboards
- Cubos OLAP
- Procesos ETL
- Data integration
- Suscripciones
- Data Mining

Características básicas de la aplicación Pentaho:

- Plataforma 100% J2EE, asegurando la escalabilidad, integración y portabilidad.
- Servidor: puede correr en servidores compatibles con J2EE como JBOSS AS, WebSphere, Tomcat, WebLogic y Oracle AS.
- Base de datos: vía JDBC, IBM DB2, Microsoft SQL Server, MySQL, Oracle, PostgreSQL, NCR Teradata, Firebird.
- Sistema operativo: no hay dependencia. Lenguaje interpretado.
- Lenguaje de programación: Java, Javascript, JSP, XSL (XSLT/XPath/XSL-FO (26)).

Conclusiones del capítulo

El capítulo expone una panorámica general del proceso de desarrollo de los AD, así como el estudio de las tendencias actuales en cuanto a la metodología de desarrollo, las tecnologías y herramientas que se utilizarán en este Trabajo de Diploma. Luego de esta investigación se arribaron a las siguientes conclusiones:

- La Metodología de Desarrollo para proyectos de Almacenes de Datos, cubre las fases por las que pasa la construcción de un AD y brinda diversas ventajas que facilitan su desarrollo; además está adaptada a las necesidades de la Universidad, teniendo como base el ciclo de vida Kimball.
- Las técnicas de captura de requisitos seleccionadas (entrevistas, discusiones y observaciones) permitieron la obtención de las necesidades del cliente, así como la definición de los requisitos del MD.
- El Lenguaje de Modelado Unificado en su versión 2.0 y el Visual Paradigm en su versión 8.0 permitirán la elaboración de los principales diagramas que formarán parte de la solución.
- El uso de PostgreSQL en su versión 9.2 como Sistema Gestor de Bases de Datos, así como el PgAdmin III en su versión 1.16, como herramienta de interfaz gráfica para la administración de los datos, posibilitarán la disponibilidad de las estructuras físicas para un correcto almacenamiento de la información.
- Las herramientas seleccionadas para los procesos de ETL, Pentaho Data Integration en su versión 5.4.0.1, contribuirán a que los datos tengan la calidad requerida para ser cargados al MD.
- Las herramientas de Inteligencia de Negocio seleccionadas, Pentaho BI Server, Mondrian, Pentaho Schema Workbench y Pentaho Report Designer en sus versiones 5.4.0.1, 3.10.0.1, 3.10.0.1 y 5.4.0.1 respectivamente, permitirán la implementación de una capa de visualización al MD para mostrar los reportes que servirán para el análisis de la información.

Capítulo 2: Análisis y diseño del Mercado de Datos para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA

Introducción

En este capítulo se realiza un estudio preliminar del negocio y de la organización, se definen las reglas del negocio, se identifican los requerimientos de información y los funcionales, los casos de uso de información, así como una breve descripción de los actores que van a interactuar con el sistema. Se define la arquitectura del MD y se diseñan los subsistemas de almacenamiento, de integración y visualización. Además, se establece el esquema de seguridad para interactuar con la Base de Datos y la aplicación.

2.1. Caracterización de la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos

ETECSA es la empresa encargada de brindar y desarrollar todo tipo de telecomunicación a nivel nacional, tanto en el interior, como desde y hacia el exterior del país. Para alcanzar esta misión satisfactoriamente, ella se encuentra estructurada en divisiones (Servicios Móviles, Servicios Internacionales, Servicios Fijos, Telefonía Pública, entre otras), cada una de ellas se componen de varias direcciones (Capital Humano, Contabilidad, Mercadotecnia Estratégica, Mercadotecnia Operativa, Logística y otras). Tal es el caso de la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos, la cual es la responsable de realizar el presupuesto anual de dicha empresa para esta división, esta se sustenta de un análisis de factibilidad e investigación mercado que se logra gracias a la información dispersa en diversos sistemas informáticos que posee la institución para otros fines. La mencionada información contiene un número de datos preliminares, a partir de los cuales los especialistas de la Dirección antes señalada, construyen los indicadores que le permiten arribar a conclusiones para elaborar el determinante presupuesto que necesita la empresa. A continuación, se brinda algunos de los 26 indicadores identificados:

- Ingreso promedio por línea: es el cociente de los ingresos percibidos por la empresa en relación con la cantidad de líneas, a nivel de segmentos y de forma global, en sus dos tipos de monedas. Es un indicador propio de la modalidad telefonía básica.
- Tráfico promedio por línea: es el cociente del tráfico telefónico por minutos en relación con la cantidad de líneas, a nivel de segmentos y de forma global, en sus dos tipos de monedas. Es un indicador propio de la modalidad telefonía básica.
- Tarifa promedio por línea: es el cociente de los ingresos percibidos por la empresa en relación con el tráfico telefónico por minutos, a nivel de segmentos y de forma global, en sus dos tipos de monedas. Es un indicador propio de la modalidad telefonía básica.

- Crecimiento de nuevos servicios: muestra el comportamiento creciente o decreciente de los servicios en un mes deseado para un año determinado, por territorios (dpa), en sus distintos tipos de agrupamiento, a partir de sus cifras de plan y real. Esto es solo en la modalidad de telefonía básica, en sus dos tipos de moneda.
- Desviación de ingreso para tráfico local, larga distancia nacional (LDN) y larga distancia internacional (LDI): muestra el comportamiento de los ingresos entre un año determinado y su año precedente a nivel de segmentos. Este indicador se halla para los dos tipos de monedas (excepto LDI que es solo en MLC), con sus segmentos correspondientes.
- Peso de los ingresos para tráfico local, LDN y LDI: muestra el comportamiento de los ingresos en forma porcentual, desglosado por segmentos, a partir de un mes seleccionado en un año determinado y sus precedentes (el mismo mes en años anteriores). Este indicador se realiza en los dos tipos de monedas (excepto para LDI que es solo en MLC).
- Comportamiento de servicios de acceso a Internet: refleja el comportamiento en un mes seleccionado de un año determinado, la cantidad de recargas (físicos) y los ingresos de Datos e Internet, desglosado en sus distintos tipos de servicios de acceso.

2.2. Necesidades de los usuarios

Puesto que existe una proporcionalidad directa entre el éxito de la solución y las necesidades de los clientes, se realizaron entrevistas particulares y reuniones colectivas con los especialistas del área en cuestión, con el fin de obtener toda la información necesaria para cumplir correctamente con los indicadores y objetivos que persigue el equipo de trabajo de la Mercadotecnia Operativa Fija. Para esto se analizó gran volumen de información referente al año actual, así como de cinco años precedentes en la que se observó diversos problemas que repercuten en el resultado final. Las entrevistas arrojaron que los especialistas eran muy dependientes de terceras personas y sistemas existentes en la empresa con otros fines, para el desarrollo de sus actividades laborales.

2.3. Reglas del negocio

Las Reglas del Negocio (RN) tienen como objetivo principal, definir las políticas o condiciones no violables durante todo el desarrollo de la solución, para lograr el correcto funcionamiento del sistema. Debido a que controlan aspectos de dicho sistema, la correcta identificación de estas es de gran importancia. Seguidamente se muestran tres de las 29 RN descritas en el artefacto Reglas del Negocio del Expediente de Proyecto:

- **Reglas de visualización**

RN1. Los valores correspondientes a ingresos, tanto para el plan como para el real en sus dos tipos de monedas, se mostrarán con dos lugares decimales.

- **Reglas de variables**

RN6. El indicador desviación, utilizado en el análisis de los ingresos a nivel de dirección territorial y nacional para los servicios de telefonía básica, las ventas de telefonía básica, servicios de datos e internet, en el mes seleccionado y sus respectivos acumulados (sumatoria de los meses precedentes incluyendo el seleccionado en un año determinado), es una diferencia que se calcula de la forma: ingreso real – ingreso plan.

- **Reglas de transformación**

RN28. Para la variable ID_FILIAL contenida en la fuente de dato SINBAPLUS, se define que cada número de la columna ID_FILIAL representa una provincia o área de la DPA. A continuación, se muestra la correspondencia entre número de fila y provincia o área.

Tabla 1. Transformación de ID_FILIAL en la ETL.

ID_Filial (SINBAPLUS)	Correspondencia Provincia o área
1	Pinar del Rio
3	Norte
4	Sur
5	Este
6	Oeste
50	Área 5
51	Área 6
35	La Habana (general)
8	Matanzas
9	Villa Clara
10	Cienfuegos

11	Sancti Spíritus
12	Ciego de Ávila
13	Camagüey
14	Las Tunas
15	Holguín
16	Granma
17	Santiago de Cuba
18	Guantánamo
19	Isla de la Juventud
56	Mayabeque
57	Artemisa

2.4. Especificación de requerimientos

El análisis detallado de requisitos contribuye a una correcta elaboración de un MD, además de guiar el proceso de desarrollo hacia la construcción de un sistema correcto. Una buena descripción de los requisitos posibilita llegar a acuerdos y resultados de las especificaciones que debe cumplir el sistema. En dicha fase se definen los requerimientos de información, funcionales y no funcionales del sistema, partiendo de las necesidades de los clientes. Se trabaja conjuntamente con los usuarios finales, permitiendo que estas especificaciones de la aplicación sean descritas por el personal de mayor conocimiento de los procesos que se ejecutan en la organización. De todo el correcto cumplimiento de estos procesos de análisis depende la exitosa implementación del sistema.

2.4.1. Requerimientos de información

El usuario final debe tener a su disposición los Requerimientos de Información (RI) en el momento que este necesite realizar consultas de análisis con los datos, lo cual le conducirá a realizar conclusiones y posible toma de decisiones. A continuación, se enumeran los RI que fueron identificados, tras una exhaustiva y detallada investigación referente a la información que procesa Mercadotecnia Operativa de Servicios Fijos:

- RI1.** Obtener los ingresos acumulados de ventas de telefonía básica, servicios de telefonía básica y datos e internet para un mes determinado por: DPA, tiempo y moneda.
- RI2.** Obtener los ingresos de ventas de telefonía básica, servicios de telefonía básica y datos e internet para un mes determinado por: DPA, tiempo y moneda.
- RI3.** Obtener los ingresos nacionales de ventas de mercancías para un mes determinado y el acumulado por: concepto, tiempo y moneda.

- RI4.** Obtener los ingresos nacionales de servicios para un mes determinado y el acumulado por: concepto, tiempo y moneda.
- RI5.** Obtener los ingresos nacionales de ventas y servicios de la empresa para un mes determinado y el acumulado por: concepto, tiempo y moneda.
- RI6.** Obtener el total de ingresos promedio por línea para un mes determinado por: concepto, tiempo y moneda.
- RI7.** Obtener el total de tráfico promedio por línea para un mes determinado por: concepto, tiempo y moneda.
- RI8.** Obtener el total de tarifas promedio por línea para un mes determinado por: concepto, tiempo y moneda.
- RI9.** Obtener los ingresos promedio por línea para un mes determinado por: concepto, tiempo, moneda y segmento.
- RI10.** Obtener el tráfico promedio por línea para un mes determinado por: concepto, tiempo, moneda y segmento.
- RI11.** Obtener las tarifas promedio por línea para un mes determinado por: concepto, tiempo, moneda y segmento.
- RI12.** Obtener el comportamiento del estado de los servicios de las líneas de la telefonía básica para un mes determinado por: DPA, agrupamiento, modalidad, concepto y tiempo.
- RI13.** Obtener el comportamiento del estado de los servicios de las líneas de la telefonía básica acumulado para un mes determinado por: DPA, agrupamiento, modalidad, concepto y tiempo.
- RI14.** Obtener el comportamiento del plan de servicios de datos e internet para un mes determinado por: DPA, tiempo y concepto.
- RI15.** Obtener comportamiento para un mes determinado de las 10 llamadas con destino al exterior con mayor tiempo de tráfico telefónico por: segmento, tiempo, país destino y moneda.
- RI16.** Obtener el comportamiento de los incrementos de los servicios tradicionales de Transmisión por Datos (TxD) e Internet para un mes determinado por: tiempo y moneda.
- RI17.** Obtener el comportamiento de los servicios instalados de TxD e Internet y Cuentas Permanentes para un mes determinado por: tiempo y moneda.
- RI18.** Obtener el comportamiento de los minutos de la tarifa mixta para un mes determinado por: tiempo y moneda.

- RI19.** Obtener el comportamiento de los minutos promedio por líneas para un mes determinado por: tiempo y moneda.
- RI20.** Obtener el comportamiento de minutos para un mes determinado por: DPA, tiempo y moneda.
- RI21.** Obtener el comportamiento de los ingresos de la tarifa mixta para un mes determinado por: tiempo y moneda.
- RI22.** Obtener el comportamiento de los ingresos promedio por líneas para un mes determinado por: tiempo y moneda.
- RI23.** Obtener el comportamiento promedio de ingresos para un mes determinado por: DPA, tiempo y moneda.
- RI24.** Obtener el comportamiento de físicos (cantidad de líneas) para un mes determinado por: DPA y tiempo.
- RI25.** Obtener el comportamiento de los ingresos de tráfico local para un mes determinado por: tiempo y moneda.
- RI26.** Obtener el comportamiento de los minutos de tráfico local para un mes determinado por: tiempo y moneda.
- RI27.** Obtener el comportamiento de los ingresos de metrado local y su desviación para un mes determinado por: segmento, tiempo y moneda.
- RI28.** Obtener el ingreso promedio por línea de metrado local para un mes determinado por: DPA, tiempo y moneda.
- RI29.** Obtener el comportamiento de los ingresos de LDN para un mes determinado por: tiempo y moneda.
- RI30.** Obtener el comportamiento de los minutos de LDN para un mes determinado por: tiempo y moneda.
- RI31.** Obtener el comportamiento de los ingresos de LDN y su desviación para un mes determinado por: segmento, tiempo y moneda.
- RI32.** Obtener el ingreso promedio por línea de LDN para un mes determinado por: DPA, tiempo y moneda.
- RI33.** Obtener el comportamiento de los ingresos de LDI para un mes determinado por: tiempo y moneda.
- RI34.** Obtener el comportamiento de los minutos de LDI para un mes determinado por: tiempo y moneda.

- RI35.** Obtener el comportamiento de los ingresos de LDI y su desviación para un mes determinado por: segmento, tiempo y moneda.
- RI36.** Obtener el ingreso promedio por línea de LDI para un mes determinado por: DPA, tiempo y moneda.
- RI37.** Obtener los ingresos de servicios TxD para un mes determinado por: DPA, tiempo, concepto y moneda.
- RI38.** Obtener comportamiento de los ingresos de datos para un mes determinado por: tiempo y moneda.
- RI39.** Obtener cantidad de cuentas de servicio de acceso a internet creadas para un mes determinado por: tiempo, moneda y sai.
- RI40.** Obtener cantidad de tarjetas de servicio de acceso a internet vendidas para un mes determinado por: tiempo, moneda y sai.
- RI41.** Obtener cantidad de recargas de servicio de acceso a internet realizadas para un mes determinado por: tiempo, moneda y sai.
- RI42.** Obtener ingresos acumulados de cuentas de servicio de acceso a internet para un mes determinado por: tiempo, moneda y sai.
- RI43.** Obtener ingresos acumulados de tarjetas de servicio de acceso a internet para un mes determinado por: tiempo, moneda y sai.
- RI44.** Obtener ingresos acumulados de recargas de servicio de acceso a internet para un mes determinado por: tiempo, moneda y sai.

2.4.2. Requerimientos funcionales

Los Requerimientos Funcionales (RF) son aquellos que representan las capacidades y condiciones que el sistema debe cumplir para dar respuesta a los RI. A continuación, se muestran los RF que fueron identificados:

RF1. Autenticar usuario.

Consiste en introducir un usuario y contraseña válidos en el sistema e ingresar al mismo con los permisos concedidos.

RF2. Adicionar roles.

Consiste en asignar distintos roles con diferentes niveles de permisos a los usuarios del sistema.

RF3. Eliminar roles.

Consiste en quitar roles pertenecientes a usuarios del sistema.

RF4. Adicionar usuarios.

Consiste en agregar nuevos usuarios al sistema.

RF5. Eliminar usuarios.

Consiste en quitar usuarios del sistema.

RF6. Crear reportes.

Consiste en crear los diferentes reportes OLAP y PDF que estarán disponibles en el sistema.

RF7. Modificar reportes.

Consiste en realizar cambios a los diferentes reportes OLAP y PDF que están disponibles en el sistema.

RF8. Eliminar reportes.

Consiste en quitar los diferentes reportes OLAP y PDF que no tienen que estar disponibles en el sistema.

RF9. Extraer información.

Consiste en efectuar la extracción de los datos de los ficheros fuentes.

RF10. Realizar transformación y carga.

Consiste en efectuar los procesos de transformación y carga de los datos extraídos.

RF11. Abrir navegador OLAP.

Permite al usuario seleccionar las columnas y filas, y poner los filtros necesarios para realizar el análisis.

RF12. Mostrar editor MDX.

Permite al usuario editar las consultas MDX del reporte sobre el cual quiere realizar los cambios.

RF13. Mostrar padres.

Permite al especialista mostrar el nivel superior en la jerarquía de cada miembro del reporte que se está visualizando.

RF14. Ocultar repeticiones.

Permite al usuario no mostrar los miembros con datos repetidos en el reporte.

RF15. Intercambiar ejes.

Permite al usuario invertir los ejes (filas, columnas) del reporte sobre el cual se quiere cambiar la forma de visualizarse (rotar las caras del cubo).

RF16. Mostrar gráfico.

Muestra un gráfico determinado vinculado a los datos de un reporte seleccionado.

RF17. Configurar gráfico.

Permite al usuario configurar las propiedades de un gráfico vinculado al reporte seleccionado.

RF18. Configurar impresión.

Permite al usuario configurar los parámetros de impresión del reporte que se está visualizando.

RF19. Exportar a PDF.

Permite al usuario exportar a formato PDF el reporte OLAP que se está visualizando.

RF20. Exportar a Excel.

Permite al usuario exportar a formato XLS el reporte OLAP que se está visualizando.

RF21. Mostrar propiedades.

Permite al usuario mostrar descripciones adicionales que fueron agregadas previamente en el diseño de los cubos OLAP, del reporte que se está visualizando.

RF22. Suprimir filas/columnas.

Permite al usuario no mostrar las filas y columnas que contengan valores nulos en el reporte que se está visualizando.

RF23. Detallar miembros.

Permite al usuario desplegar los miembros de los niveles de la jerarquía del reporte que se está visualizando.

RF24. Entrar en detalles.

Permite al usuario mostrar solamente los miembros de un nivel seleccionado de la jerarquía del reporte que está visualizando (taladrar).

RF25. Mostrar datos de origen.

Permite al usuario mostrar una tabla con información detallada sobre la procedencia exacta en la base de datos del valor seleccionado de una medida en el reporte que se está visualizando.

2.4.3. Requerimientos no funcionales

Los Requerimientos No Funcionales (RNF) son las propiedades, cualidades o características que el sistema debe cumplir, las cuales le reportan al cliente alguna utilidad, como el rendimiento, la eficiencia y la fiabilidad, dándole más confianza y seguridad a la aplicación. Luego de un análisis minucioso, se definieron 24 RNF para el MD, descritos en el artefacto Especificación de Requisitos de Software del Expediente de Proyecto. A continuación, se muestran algunos de ellos:

- **Requisitos de fiabilidad**

RNF1. Disponibilidad

El sistema debe tener disponibilidad al 100% entre las 8:00 am y las 5:00 pm en los días laborables (lunes a viernes).

- **Requisitos de eficiencia**

RNF4. Tiempo de respuesta

El sistema debe tener un tiempo de respuesta aproximado de 1 minuto, no debe excederse de 2 minutos.

RNF6. Capacidad

El sistema deberá permitir varios usuarios (entre dos y diez personas) conectados sin que esto afecte la respuesta de las consultas.

RNF7. Restricciones de diseño

El SGBD que se utilizará será PostgreSQL versión 9.2 y como interfaz de administración de dicho gestor PgAdmin III versión 1.16.

RNF13. Interfaces de usuario

El sistema debe permitir la visualización de un reporte determinado sin exceder los tres clics.

RNF15. Interfaces de hardware

Para el proceso de transformación es necesaria como mínimo, una memoria RAM de 4 GB, un microprocesador Core i3 y 100 Gb de disco duro.

2.5. Casos de Uso del Sistema, Actores del Sistema y Diagrama de Casos de Uso del Sistema

Después de definidos los RI y los RF, estos se agruparon en los Casos de Uso (CU) correspondientes, identificando además los actores encargados de inicializar tales CU. Los Casos de Uso del Sistema (CUS), se especifican en la fase de análisis y diseño de un AD. La función de ellos es describir la relación que existe entre la aplicación y los usuarios. La Tabla 2 brinda la información concerniente a los actores del sistema, con una concisa descripción de sus responsabilidades, seguidamente se muestra el agrupamiento de los RI que conforman los cinco Casos de Uso de Información (CUI), así como el de los RF, que componen los siete Casos de Uso Funcionales (CUF). Con estos tres importantes componentes (CUI, CUF y actores del sistema), se genera el Diagrama de Casos de Uso del Sistema (DCUS) para la solución (Ver Figura 3).

Tabla 2. Actores del sistema y sus responsabilidades.

Actor	Responsabilidad
Administrador	Inserta y elimina los usuarios del sistema. Inserta y elimina los roles asignados a los usuarios del sistema. Inserta, modifica y elimina los reportes disponibles en el sistema. Consulta información de los reportes.

Administrador de ETL	Realiza la extracción, transformación y carga de los datos de los ficheros fuentes.
Especialista	Consulta la información de los reportes.

Casos de Uso de Información

Los CUI tienen como objetivo agrupar un conjunto de requisitos de información teniendo en cuenta los conceptos del negocio que manejan y sus variables de entrada y salida.

- CUI 1.** Presentar información relacionada con ingresos.
- CUI 2.** Presentar información relacionada con tráfico telefónico.
- CUI 3.** Presentar información relacionada con tarifas telefónicas.
- CUI 4.** Presentar información relacionada con líneas telefónicas.
- CUI 5.** Presentar información relacionada con servicios de datos e Internet.

Casos de Uso Funcionales

Los CUF tienen como meta agrupar los RF definidos para los subsistemas que componen la solución.

- CUF 1.** Autenticar usuarios.
- CUF 2.** Gestionar usuarios.
- CUF 3.** Gestionar reportes.
- CUF 4.** Gestionar roles.
- CUF 5.** Extraer datos.
- CUF 6.** Realizar transformación y carga.
- CUF 7.** Realizar opciones sobre el reporte.

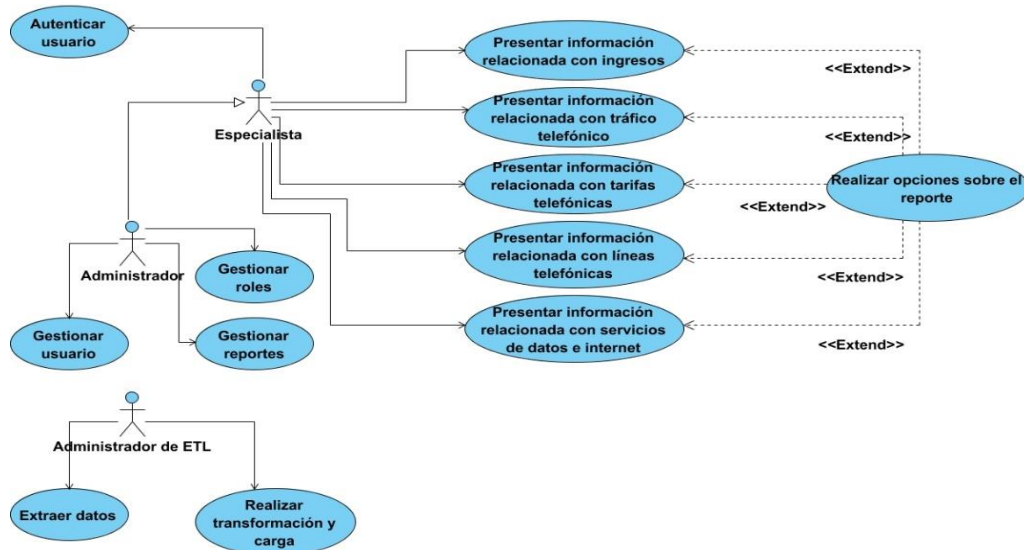


Fig 3. Diagrama de Casos de Uso del Sistema.

2.5.1. Especificación de Casos de Uso del Sistema

Seguidamente se muestra la Especificación de Casos de Uso del Sistema correspondiente al CU 8. “Presentar información relacionada con los ingresos.” El resto de las descripciones de los CU se encuentran en el artefacto Especificación de Casos de Uso del Expediente de Proyecto.

Tabla 3. Especificación del Caso de Uso Presentar información relacionada con los ingresos.

Objetivo	Presentar información relacionada con los ingresos.	
Actores	Especialista.	
Resumen	El CU inicia cuando el especialista desea hacer un análisis de la información relacionada con los ingresos desde diferentes perspectivas. El especialista selecciona el reporte que desea ver, el sistema muestra la información contenida en él y las opciones de los posibles cambios que le puede hacer al reporte. El CU finaliza cuando el especialista termina el análisis de la información relacionada con los ingresos.	
Complejidad	Media	
Prioridad	Crítica	
Precondiciones	El especialista se autenticó correctamente. Los datos correspondientes fueron cargados en el MD. Los reportes relacionados con los ingresos fueron creados.	
Postcondiciones	Los reportes correspondientes fueron consultados por el especialista.	
Flujo de eventos		
Flujo básico Presentar información relacionada con los ingresos		
	Actor	Sistema
1.	El especialista se autentica en el sistema.	
2.		Muestra la interfaz principal con su área de análisis (A.A) SAMKOSF y los libros de trabajos contenidos en esta.
3.	El especialista selecciona el libro de trabajo (L.T) Ingresos, contenido en el A.A. SAMKOSF	
4.		Muestra los reportes contenidos dentro del L.T Ingresos.
5.	El especialista selecciona el reporte deseado del L.T Ingresos.	
6.		Muestra la información contenida en el reporte seleccionado y brinda la posibilidad al

		especialista de hacerle cambios al reporte para su análisis. Ir al CU: Realizar opciones sobre el reporte. Finaliza el CU.
Flujos alternos		
2a Los datos son incorrectos		
	Actor	Sistema
		Muestra mensaje "Los datos son incorrectos". Vuelve al paso 1 del Flujo Básico de Eventos.
Opciones de Reportes		
Perspectivas de análisis	Posibles resultados	
	Medidas	Periodicidad
Variables de entrada relacionadas con el CU: <ul style="list-style-type: none"> • Tiempo • Moneda • Concepto • DPA • Segmento • SAI 	Variables de salida disponibles: <ul style="list-style-type: none"> • Plan • Real • Desviación • Por ciento • Ingreso • Suma_cuota_básica • Suma_ing_trlocal_tot • Suma_ing_ldnm_tot • Suma_ing_ldimet 	Rango de tiempo en que se solicitan las variables de salida: <ul style="list-style-type: none"> • Mensual
Relaciones	CU Incluidos	No aplica
	CU Extendidos	Realizar opciones sobre el reporte: Paso 6 del Flujo básico. Realizar opciones sobre el reporte en el CU: Presentar información relacionada con los ingresos.
Requisitos no funcionales	Sección: "3.2 Requisitos no funcionales" del documento: "Especificación de requisitos de software".	
Asuntos pendientes	Posibles mejoras al CU.	

2.6. Definición de la arquitectura del Mercado de Datos

La arquitectura general del Mercado de Datos para Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA, está constituida por las fuentes de datos y por tres subsistemas fundamentales: el subsistema de integración, el subsistema de almacenamiento y el subsistema de visualización como se muestra en la Figura 4:

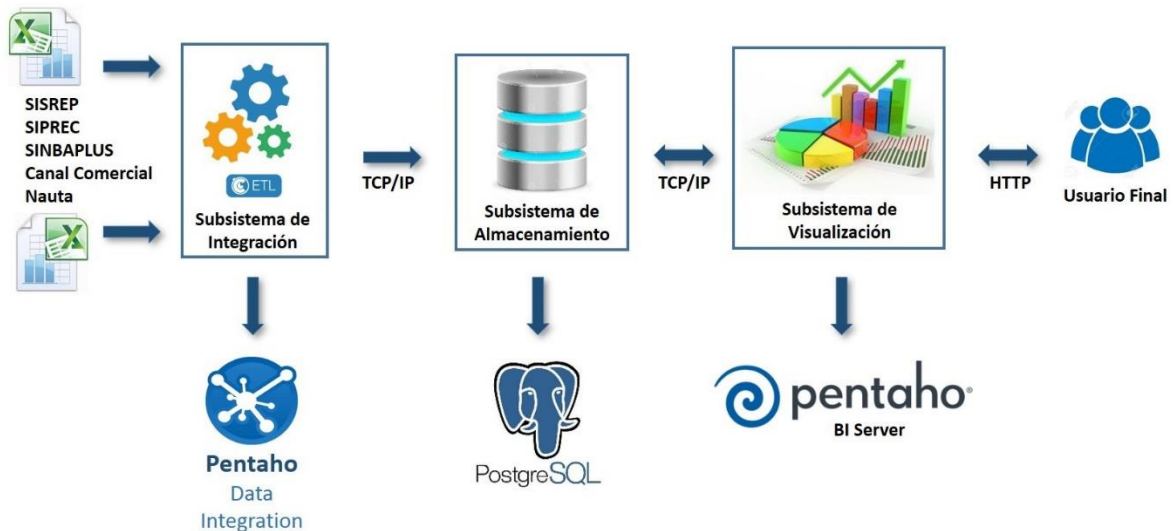


Fig 4. Arquitectura del MD Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA.

De inmediato se explicarán cada uno de los subsistemas que definen la arquitectura del MD:

- **Subsistema de integración:** es el encargado de integrar, estandarizar y limpiar la información de las fuentes de datos, con el fin de cargarla hacia el almacén, logrando así los procesos de ETL de los datos.
- **Subsistema de almacenamiento:** es el que se ocupa de almacenar toda la información que el subsistema de integración procesó, en el área específica del MD. El almacén estará compuesto por tablas de hechos y dimensiones.
- **Subsistema de visualización:** tiene como objetivo principal consultar la información almacenada en el AD, permitiendo mostrar los reportes que necesitan los clientes, bajo distintos criterios de análisis.

2.7. Diseño de la solución

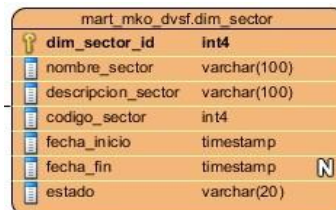
Para lograr un resultado exitoso del MD es determinante realizar un eficiente diseño de cada uno de los subsistemas que lo componen, producto de la repercusión que tienen en el momento de integrarse como un único sistema para la solución.

2.7.1. Diseño del subsistema de almacenamiento

Para el correcto desarrollo y funcionamiento del MD, en esta etapa se realiza el diseño del modelo dimensional, el cual contiene las tablas de hechos relacionadas con sus correspondientes dimensiones, identificadas en el negocio, para la solución.

- **Dimensiones**

Las dimensiones representan las características de un hecho, que permite su análisis posterior en el proceso de toma de decisiones y brinda una perspectiva adicional a un hecho dado. Una de las características principales que poseen las dimensiones es la definición de jerarquías entre sus atributos, las cuales tienen como fin plasmar la forma en que se puede consolidar la realización del proceso de análisis, ya sea mediante el uso de sumas, porcentos, máximos, mínimos, entre otros. Seguidamente, la Figura 5 muestra un ejemplo de una de las dimensiones que contiene la solución.



mart_mko_dvsf.dim_sector	
dim_sector_id	int4
nombre_sector	varchar(100)
descripcion_sector	varchar(100)
codigo_sector	int4
fecha_inicio	timestamp
fecha_fin	timestamp N
estado	varchar(20)

Fig 5. Dimensión dim_sector.

- **Tablas de hechos**

Las tablas de hechos pueden representar un objeto o evento del negocio que es utilizado por los analistas de la información, para analizar el comportamiento de los datos de la empresa. Además, estas tablas contendrán las medidas numéricas las cuales simbolizan las variables de salida del almacén. A continuación, la Figura 6 muestra un ejemplo de una de las tablas de hechos que contiene la solución.

hech_SISREP		
dim_tiempo_id	int4	
dim_concepto_id	int4	
dim_moneda_id	int4	
dim_dpa_id	int4	
plan	numeric(12, 2)	N
real	numeric(12, 2)	N

Fig 6. Tabla de hecho hech_SISREP.

Matriz bus o matriz dimensional

La Matriz bus es una herramienta potente para la planificación y la comunicación, pues permite definir la arquitectura general de los datos para el MD, así como determinar el impacto que provocaría durante el desarrollo del sistema un cambio a la solución. En ella se describen las relaciones que existen entre las tablas de hechos y las dimensiones. Las columnas de la matriz representan las dimensiones utilizadas en el MD y las filas los hechos identificados. Las celdas sombreadas con una X, indican que la columna de dimensión está relacionada con la fila del proceso de negocio (hecho). Esto posibilita ver de inmediato cuáles son las dimensiones que merecen una atención especial debido a su participación o relación con múltiples hechos. En las tablas 4 y 5, se muestran los hechos y dimensiones, así como la Matriz bus o dimensional para el MD de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA:

Tabla 4. Hechos y dimensiones

Hechos		Dimensiones	
H1: SISREP	H4: SINBAPLUS 2	D1: dim_concepto	D6: dim_moneda
H2: SIPREC	H5: canal comercial	D2: dim_dpa	D7: dim_segmento
H3: SINBAPLUS 1	H6: nauta	D3: dim_tiempo	D8: dim_sector
		D4: dim_agrupamiento	D9: dim_modalidad
		D5: dim_pais_destino	D10: dim_sai

Tabla 5. Matriz de bus o dimensional.

Hechos/Dimensiones	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
H1	X	X	X			X				
H2		X	X			X				
H3			X			X	X			

H4		X	X		X	X		X		
H5	X	X	X	X					X	
H6			X			X				X

Modelo de datos

El modelo dimensional tiene como objetivo primordial mostrar los datos del negocio en una estructura lógica, presentando las relaciones existentes entre las tablas de hechos y las dimensiones. A continuación, se muestra en la Figura 7 el diseño del modelo dimensional propuesto, para el desarrollo del MD de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA, en el que se observa la relación antes mencionada, entre las tablas de hechos propuestas y las dimensiones definidas para la solución, evidenciándose como topología de esquema constelación de hechos.



Fig 7. Modelo dimensional.

2.7.2. Diseño del subsistema de integración

En esta etapa se detallan las fuentes de datos, se efectúa un perfilado de datos, se elabora el diccionario de datos, se realiza el diseño general de las transformaciones y se establece un mapa lógico de datos, para asegurar que los procesos de ETL se ejecuten correctamente.

Registro de sistemas fuentes

El Expediente de Proyecto contiene un artefacto llamado Registro de Sistema Fuente, el cual tiene como finalidad detallar las fuentes de datos, logrando así un mejor entendimiento de la información contenida en las mismas. El sentido que tiene este documento es contribuir a la documentación de los sistemas fuentes, concentrándose en lo referente al estado físico de los datos y de los responsables de los sistemas fuentes en cuestión. La información utilizada por Mercadotecnia Operativa de Servicios Fijos se encuentra en 255 ficheros Excel, con un tamaño bruto de 174 MB, la cual es controlada mediante políticas de control de acceso, de manera que cada especialista puede acceder solo a la parte de la información que le corresponde por su contenido de trabajo, mientras que el especialista principal y el director de la dirección tienen acceso pleno a dicha información.

Diccionario de datos

El diccionario de datos sirve de apoyo para los procesos de ETL, debido a que este contiene la información necesaria para un correcto entendimiento de los sistemas fuentes, enfocado principalmente en la documentación de las variables con las que interactúa la fuente de datos. En él se describen cada una de estas variables especificando el significado que tienen en el negocio y los posibles valores que pueden tomar. Todo esto es recogido en el artefacto Diccionario de Datos que se encuentra en el Expediente de Proyecto. De forma general se identificaron 10 variables con las que interactúa la fuente de datos, que corresponden a cada una de las dimensiones del MD.

Perfilado de datos

El perfilado de los datos es un proceso que se caracteriza por analizar las fuentes de datos, con el fin de comprender a profundidad todos sus detalles en cuanto a estructura y contenido, para determinar entonces las transformaciones que se le deben hacer a estos. Esto se refleja en el artefacto Perfilado de Datos del Expediente de Proyecto, donde se detalla el nombre de la tabla, el campo, el tipo de dato, la longitud de la cadena, si hay nulos, valores máximos, entre otros. Es también aquí donde se reflejan los problemas detectados y su solución. En la Figura 8 se muestra el resultado del perfilado de los datos realizado al fichero Excel Nauta_2015.xlsx.

Porcentaje de valores nulos para Análisis de los servicios de nauta



Fig 8. Perfilado de datos del fichero Nauta_2015.xlsx.

Diseño general de las transformaciones

En los procesos de ETL, las transformaciones son una serie de pasos, que cuando se ejecutan posibilitan que los datos se carguen correctamente en sus correspondientes tablas del MD. Posterior al perfilado de datos, se determinaron las transformaciones que se realizaron para poblar el MD Mercadotecnia Operativa de Servicios Fijos. Los pasos que se hicieron para implementar cada una de ellas fueron los siguientes:

1. Se efectuó la extracción de los datos de una determinada fuente (Excel), con lo que se obtuvo los diferentes indicadores con sus valores numéricos.
2. Se continuó con la limpieza y transformación de los datos, donde se aplicó las necesarias reglas de transformación.
3. Se localizó el identificador correspondiente en las tablas de dimensiones.
4. Se validó los tipos de datos; en caso de existir problemas en este sentido, se envían hacia un Excel de errores, la columna que fue validada y una descripción del error que se produjo; de lo contrario, se insertan o actualizan estos en la tabla de hecho que le corresponde.
5. Se obtuvo la información necesaria del sistema, para crear los metadatos del proceso, tales como: nombre del fichero fuente, nombre de la transformación y fecha de ejecución de la transformación.

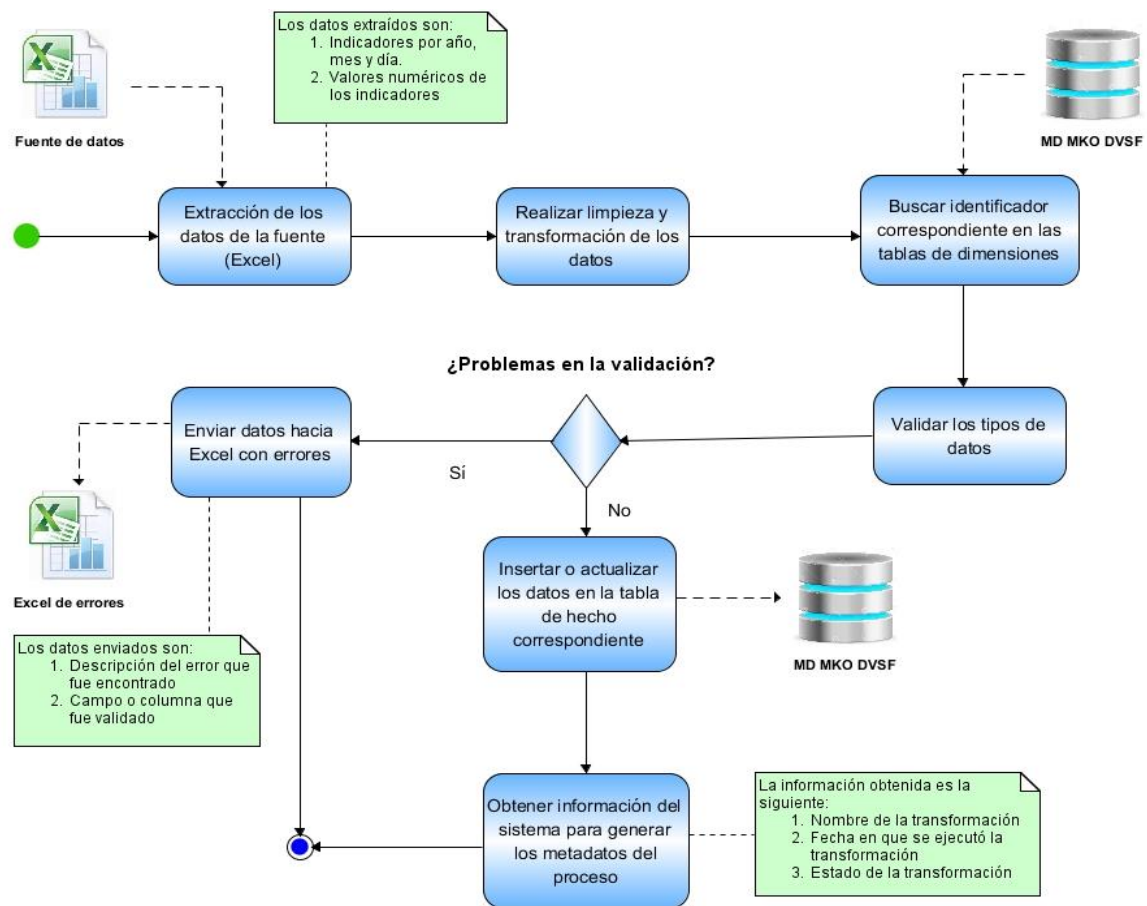


Fig 9. Diseño general de la implementación de las transformaciones.

Mapa lógico de datos

El artefacto Mapa Lógico de Datos del Expediente de Proyecto, tiene como función describir los datos contenidos en las tablas de los hechos y las dimensiones identificados en el MD Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA, o sea, dicho artefacto informa sobre cada dato: el nombre de la base de datos al que pertenece, nombre de la tabla, nombre de la columna, tipo de dato, entre otras descripciones, con el fin de lograr un correcto entendimiento del contenido, estructura y dependencia de los datos.

2.7.3. Diseño del subsistema de visualización

La capa de visualización es una de las más importantes, por ello se define la arquitectura de la información que se va a mostrar, se diseñan los cubos OLAP y los reportes candidatos, los que garantizarán que la información presentada sea la correcta.

Arquitectura de información

Para tener una mejor idea con respecto a cada elemento y a la estructura que forman parte del sistema que serán mostrados en el MD en su capa de visualización, se genera un artefacto en el Expediente de Proyecto llamado Vista de Presentación, el cual brinda un entorno de monitoreo, análisis y control de la información que guarda el MD para la Dirección de Mercadotecnia Operativa de Servicios Fijos, para el análisis de la información por parte de los especialistas. En este documento se ha identificado un A.A, que contiene cinco L.T y 44 reportes agrupados en ellos. En la Figura 10 se muestran parte de estos elementos, los cuales componen la estructura de navegación de la información presentada en la capa de visualización del MD:



Fig 10. Estructura de navegación del MD para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos.

Diseño de los cubos OLAP:

El diseño de los cubos OLAP se realizó mediante la herramienta Pentaho Schema Workbench, en la cual se definen los hechos identificados, las dimensiones que corresponden a cada uno de estos hechos con sus niveles de jerarquía, así como la conexión a la base de datos que contiene los datos para el cubo multidimensional. En la presente investigación se modelaron 6 cubos y 10 dimensiones, correspondientes a cada una de las tablas de hechos y dimensiones respectivamente. A continuación, se muestra en la Figura 11, un ejemplo del diseño de uno de los cubos modelados (SISREP) con sus dimensiones (dim_concepto, dim_dpa, dim_moneda y dim_tiempo) y sus medidas (plan, real, desviación y porciento).



Fig 11. Diseño del cubo OLAP SISREP con la herramienta Schema Workbench.

Diseño de los reportes candidatos

Para tener una mejor comprensión de los elementos que componen los reportes utilizados por los usuarios del sistema existe el artefacto Reportes Candidatos que forma parte del Expediente de Proyecto, en el cual se describen los reportes candidatos que fueron identificados en el desarrollo del MD Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA. A continuación, se muestra la composición de uno de ellos:

Tabla 6. Descripción del reporte Comportamiento de los Ingresos acumulados en MLC por territorio.

Área de análisis (A.A)	SAMKOSF.
Libro de Trabajo (L.T)	LT1- Ingresos.
Reporte (Tabla de Salida - T.S)	TS1- Comportamiento de los Ingresos acumulados en MLC por territorio.
Descripción	El reporte muestra los ingresos (Plan, Real, Desviación y Porcentaje) de ventas de telefonía básica, de servicios de telefonía básica y de servicios de datos e internet, percibidos desde el mes de enero hasta el mes seleccionado para el análisis, por cada provincia o área, en un tiempo determinado, en MLC.

Elementos del reporte	<ul style="list-style-type: none"> • DPA • Tiempo • Moneda
Frecuencia de emisión	Mensual.
Funciones	Deviación: Ingreso Real - Ingreso Plan Porciento: $(\text{Ingreso Real} / \text{Ingreso Plan} - 1) * 100$
Gráfico	

2.8. Política de respaldo

En función de definir una política de respaldo consistente en el MD Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA, se decide evaluar las salvadas, que serán conservadas por un período de cinco años, a partir de dos aspectos esenciales:

- Periodicidad de las salvadas: las salvadas de toda la información contenida en la base de datos se realizan mensualmente, verificando en todo momento que exista una copia escrita de la información almacenada en el servidor. Las tablas que se involucran en este proceso son las seis tablas de hechos y las diez tablas de dimensiones identificadas en el proceso de análisis.
- Salvadas existentes: actualmente no existen salvadas en esta área, por lo cual se realizará una salvada referente a la carga histórica del MD.

2.9. Esquema de seguridad

El MD Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA posee una seguridad que está regida por los niveles de acceso al sistema, los cuales están limitados específicamente por los permisos y roles que a los usuarios se les ha sido otorgado para interactuar con la base de datos y el sistema de forma general como bien se muestra en la Tabla 6. Para garantizar una buena seguridad en la base de datos se definió el rol Administrador de BD quien posee acceso total a la base de datos del sistema teniendo la responsabilidad de efectuar la política de respaldo definida con anterioridad. Además, se definió una persona que tiene permisos de lectura y escritura sobre la base de datos encargada de realizar los procesos de integración, el cual debe de ser informático, quien desempeñará el rol de Administrador de ETL. Por otra parte, para lograr una sólida seguridad en la aplicación se definieron dos roles más, Administrador y Analista, el primero tiene acceso total en lo que concierne a: administrar usuarios, roles y reportes; el otro solo tiene acceso de lectura al A.A. SAMKOSF.

Tabla 7. Roles y permisos del sistema.

Roles	Permisos
Administrador de BD	Acceso total a la base de datos. Responsable de ejecutar la política de respaldo.
Administrador de ETL	Permisos de lectura y escritura sobre la base de datos.
Administrador	Acceso total al A.A. SAMKOSF y a la base de datos. Gestiona los usuarios, roles y reportes.
Analista	Tiene acceso restringido de sólo lectura al A.A. SAMKOSF y a la base de datos.

Conclusiones del capítulo

Después de haber realizado el análisis y diseño del Mercado de Datos para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA se arribaron a las siguientes conclusiones:

- El levantamiento de requisitos arrojó como principal resultado la identificación de 44 requerimientos de información, 25 requisitos funcionales y 24 no funcionales, sirviendo de base para elaborar el DCUS. De igual manera, se identificaron 29 reglas del negocio, permitiendo definir reglas de transformación que serán utilizadas en los procesos de integración. Por último, se realizó la descripción de los CU del MD para especificar cada una de las funcionalidades del sistema.
- La arquitectura base del MD definida, permitió identificar los elementos y subsistemas que están implicados en el desarrollo de la solución.
- El modelo dimensional diseñado, representa las relaciones entre las seis tablas de hechos y las diez dimensiones identificadas para el MD.
- Durante el diseño de los subsistemas de integración y visualización, quedó definido el diseño general para las transformaciones, la arquitectura de información del MD, el diseño de los reportes candidatos y de los cubos OLAP, todo esto servirá de guía para la implementación de dichos subsistemas.
- A través de las políticas de respaldo definidas, se definió que la periodicidad que van a tener las salvadas es mensual.
- Los roles y permisos definidos en el MD, contribuirán a la seguridad de la aplicación.

Capítulo 3: Implementación y pruebas del Mercado de Datos para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA

Introducción

Este capítulo se refiere a la implementación de la solución, abordando específicamente cómo se realiza la implementación de los subsistemas de almacenamiento, integración y visualización del MD para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA, teniendo en cuenta los requerimientos y necesidades del negocio. También, hace referencia a las pruebas para determinar que los procesos de ETL tengan la calidad requerida, así como los casos de pruebas, basados en CU y las reglas de transformación, para validar los reportes del MD y el cumplimiento de las RN respectivamente.

3.1. Implementación del subsistema de almacenamiento

En la base de datos, los datos se encuentran organizados en estructuras lógicas que facilitan la correcta manipulación de los mismos. Estas estructuras son denominadas esquemas y tablas.

Esquemas

Los esquemas en una base de datos representan una forma de organizar la información contenida en la misma. Dentro de los esquemas se pueden encontrar funciones, operadores y tipos de datos que facilitarán su implementación. En el presente trabajo se definieron tres esquemas, los cuales serán explicados a continuación:

- **Esquema dimensiones:** contiene las tablas de las dimensiones generales del posible AD de la empresa, las que se utilizaron para implementar el MD.
- **Esquema mart_mko_dvsf:** contiene las tablas de hechos y las tablas de dimensiones propias del MD para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA.
- **Esquema metadatos:** contiene las tablas para la captura de los metadatos de los procesos de ETL.

Tablas

Con el diseño del modelo físico se genera el script de la base de datos, y a partir de este se concluye que la solución propuesta tiene 18 tablas en total, de ellas 13 contenidas en el esquema mart_mko_dvsf, tres en el esquema dimensiones y dos en el esquema metadatos. En la Figura 12 se muestra la estructura física del MD.

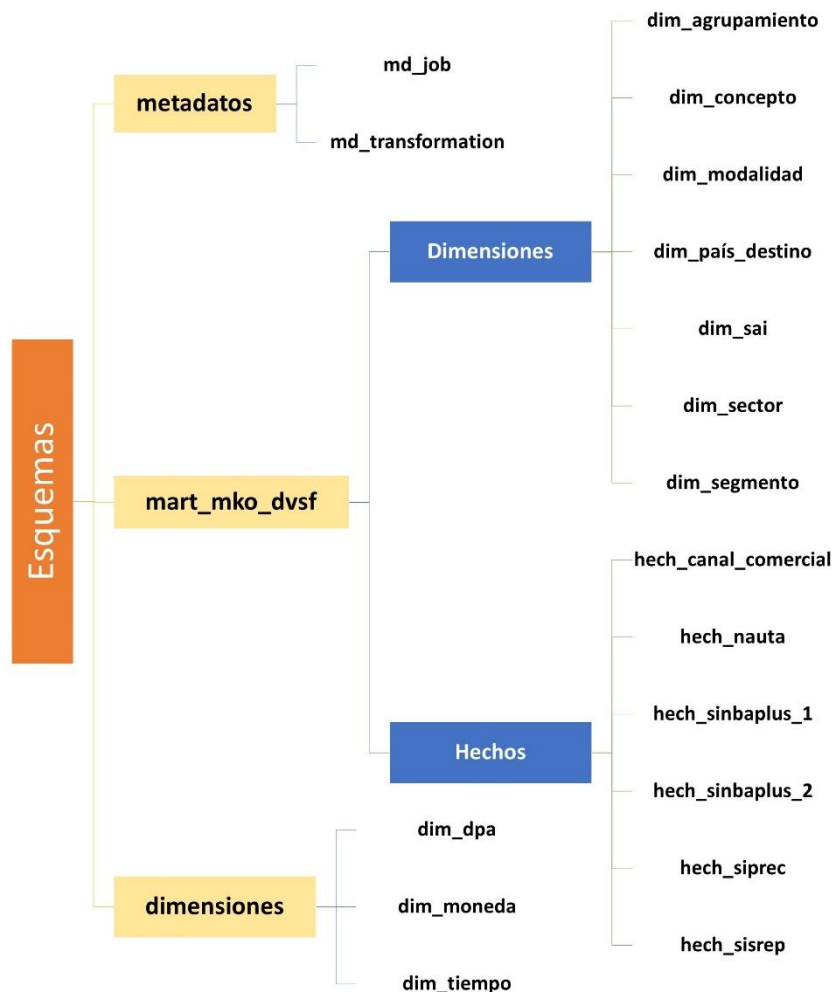


Fig 12. Estructura física del MD para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA.

3.1.1. Configuración de la seguridad de la base de datos

Con el objetivo de restringir el acceso a la base de datos y de este modo tener control sobre la seguridad de la misma, fue definido el rol Administrador de la BD mediante la herramienta de interfaz gráfica PgAdmin. Este rol posee todos los permisos sobre la base de datos y tiene definido una contraseña para acceder a ella. A continuación, se muestra en la Figura 13, cómo se adicionó dicho rol.

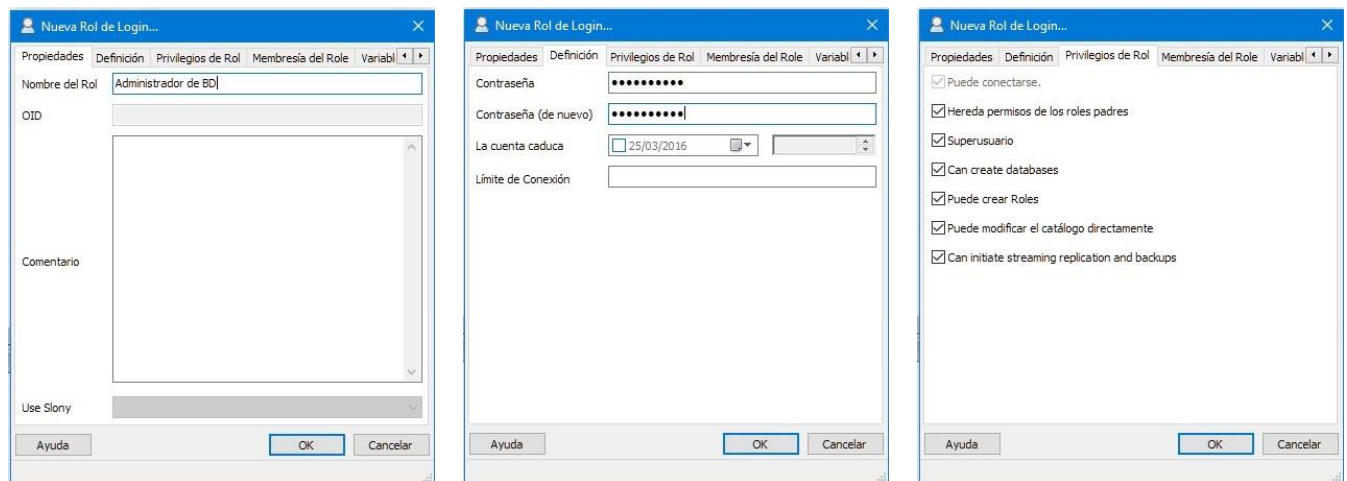


Fig 13. Adicionar rol Administrador de BD.

3.2. Implementación del subsistema de integración de datos

En el proceso de extracción se obtienen los datos que serán almacenados en el MD para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA, todo esto mediante las fuentes de información proporcionadas por el cliente. Los datos fueron extraídos de 255 ficheros Excel con el objetivo de adaptarlos al modelo relacional establecido con anterioridad. Cada uno de los ficheros contiene la información histórica referente a los diferentes sistemas que proporcionan la información manipulada por los especialistas.

3.2.1. Limpieza y transformación de datos

Seguidamente a la extracción de los datos, se realizó la limpieza de estos. En el proceso se detectaron posibles errores e incoherencias que existían en la fuente de datos, dándole el tratamiento adecuado para corregirlos, con el fin de que los datos que se cargasen a la base de datos fuesen fieles, completos y sin errores, para que se pudieran interpretar de forma correcta por el cliente. Luego de la limpieza, se realizaron transformaciones a los datos aplicando las RN. Con esto se logró que la información tuviese la calidad requerida por el usuario final, pues se eliminaron datos duplicados, se detectaron y corrigieron errores ortográficos y se trataron los posibles valores nulos que no tuviesen significado en el negocio, resultando en datos listos para la carga del MD.

3.2.2. Transformaciones y trabajos

En el MD para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA se implementaron un total de 24 transformaciones, diez para la carga de las dimensiones y 14 para

los hechos. En la Figura 14 que se muestra más adelante, se evidencian los pasos lógicos que se siguieron para la carga de la tabla hech_sinbaplus_2, los que se describen a continuación:

1. Se realizó la extracción de los datos a partir de los ficheros Excel correspondientes al sistema SINBAPLUS, el cual recoge la información relacionada con la cantidad de llamadas de larga distancia internacional que se hicieron.
2. Haciendo uso del paso Select values, se eliminaron campos que eran innecesarios.
3. Con el componente Filter rows se garantizó que en el flujo no hubiese tuplas vacías.
4. A través de los componentes Value Mapper y Replace in string, se aplicó la RN 28 y se corrigieron errores en los formatos de la fecha y el país respectivamente.
5. Con el paso Split Fields y Set field value to a constant, se dividió en tres campos la fecha y se estableció la moneda utilizada.
6. Se realizó la búsqueda en las tablas de dimensiones dim_dpa, dim_tiempo, dim_sector, dim_pais y dim_moneda, para obtener los identificadores de los valores de cada una de ellas.
7. Una vez que el flujo tuviese la estructura correspondiente a la tabla de hecho en la base de datos, se validaron los tipos de datos, y se guardaron en un fichero Excel los errores de validación en caso de haber existido para su posterior análisis.
8. Luego de comprobar que los datos eran correctos, se procedió a realizar la carga de los mismos hacia la tabla hech_sinbaplus_2.

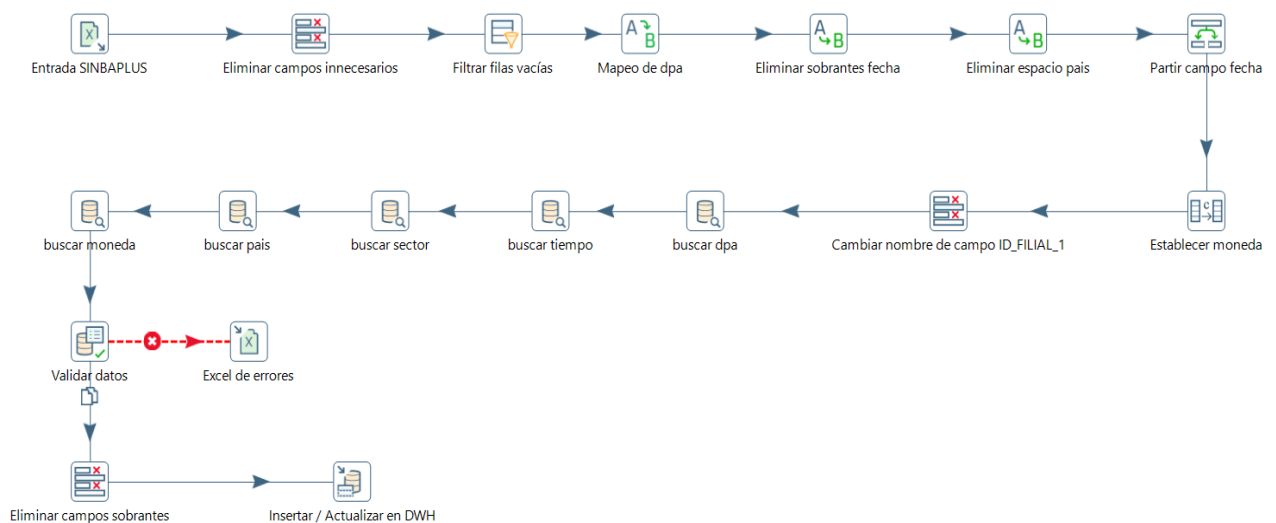


FIG 14. Transformación “hech_SINBAPLUS_2” para cargar la tabla hech_sinbaplus_2.

Con la finalidad de conseguir una estandarización en el desarrollo de las transformaciones y los trabajos, se utilizó una nomenclatura específica. Cuando se tratase de una transformación en el proceso de integración de datos que cargara los hechos, se designó de la siguiente manera:

- hech_<nombre>. Ejemplo: hech_SINBAPLUS_2.ktr

Cuando la transformación fuese para cargar las dimensiones se nombró de la siguiente manera:

- dim_<nombre>. Ejemplo: dim_tiempo.ktr

Cuando se refiriera a un trabajo, se denominó de la siguiente forma:

- cargar_<nombre>. Ejemplo: cargar_hechos.kjb

Finalmente se implementaron tres trabajos, uno que ejecuta todas las transformaciones que hacen posible la carga de los hechos (cargar_hechos), otro que permite cargar todas las dimensiones (cargar_dimensiones) y un tercero que ejecuta los dos trabajos explicados anteriormente (cargar_datamart). En este último, debido a la dependencia que tienen las tablas de hechos con las tablas de dimensiones, fue necesario ejecutar primero el trabajo cargar_dimensiones y seguidamente el trabajo cargar_hechos. A continuación, en las Figuras 15 y 16, se muestran el trabajo que puebla el MD y el trabajo que ejecuta todas las transformaciones que cargan los hechos, respectivamente.



Fig 15. Trabajo “cargar datamart”.

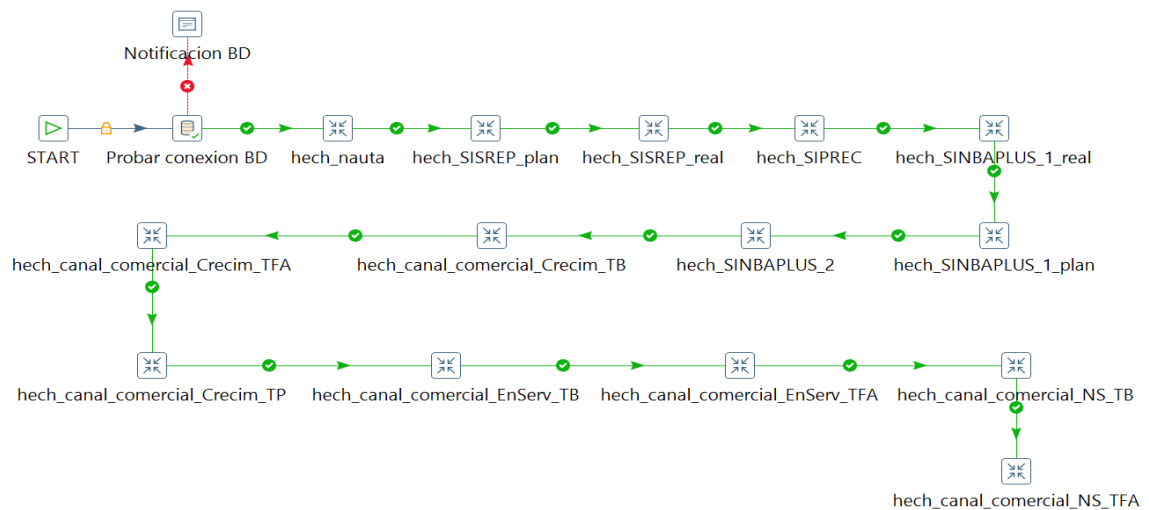


Fig 16. Trabajo “cargar hechos”.

3.2.3. Carga de datos

La carga de todos los datos fue el paso final de la implementación del subsistema de integración, donde los datos que fueron limpiados y transformados en los procesos anteriores, se cargaron hacia las tablas de hechos y dimensiones correspondientes en el modelo dimensional del MD para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA, facilitando la consulta de los mismos por parte de los especialistas.

3.2.4. Gestión del cambio en las dimensiones

Las dimensiones lentamente cambiantes o SCD (*Slowly Changing Dimensions*) son dimensiones en las cuales sus datos tienden a modificarse a través del tiempo, ya sea de forma ocasional o constante, que implique a un solo registro o la tabla completa. Cuando ocurren estos cambios, se puede optar por seguir alguna de estas dos opciones:

- Registrar el historial de cambios.
- Reemplazar los valores que sean necesarios.

Inicialmente Ralph Kimball planteó tres estrategias a seguir cuando se tratan las SCD: tipo 1, tipo 2 y tipo 3; pero a través de los años la comunidad de personas que se encargan de modelar base de datos profundizaron en las definiciones iniciales e incluyeron varios tipos de SCD más como son el tipo 0, tipo 4 y tipo 6. A continuación se detallará cada tipo de estrategia SCD:

- **Tipo 0:** este es un enfoque pasivo, es decir, no se hace nada al respecto. Los valores permanecen como estaba la dimensión cuando los registros fueron creados.
- **Tipo 1:** este tipo es el más básico y sencillo de implementar. En este caso cuando un registro presenta un cambio en alguno de los valores de sus campos, se procede simplemente a actualizar el dato en cuestión, sobrescribiendo el antiguo. Usualmente este tipo es utilizado en casos donde la información histórica no sea importante de mantener, tal como sucede, por ejemplo, cuando se debe modificar el valor de un registro porque tiene errores ortográficos.
- **Tipo 2:** en este enfoque se inserta un nuevo registro cada vez que existe un cambio en la dimensión. Se agrega un campo de versión u opcionalmente se agregan dos columnas para capturar la fecha de inicio y final de ese valor. Con este método se puede relacionar fácilmente el período de tiempo para el cual es válido cierto dato en la dimensión. Esta técnica permite guardar ilimitada información de cambios.
- **Tipo 3:** este método da seguimiento al cambio agregando nuevas columnas. Una columna mantendría el dato actual, y otra el dato nuevo por el que se quiere cambiar el actual, así como

una columna de fecha efectiva del cambio. Este enfoque sólo puede mantener un cambio histórico, a diferencia del Tipo 2 que puede mantener cambios ilimitados en la historia.

- **Tipo 4:** esta técnica mantiene una tabla histórica para todos los cambios y una tabla con el valor actual de la dimensión. Esta tabla histórica indicará, por ejemplo, qué tipo de operación se ha realizado (Insertar, Modificar, Eliminar), sobre qué campo y en qué fecha. El objetivo de mantener esta tabla es el de contar con un detalle de todos los cambios, para luego analizarlos y poder tomar decisiones acerca de cuál técnica SCD podría aplicarse mejor.
- **Tipo 6:** este procedimiento, conocido también como Híbrido, es una combinación de los Tipos 1, 2 y 3 ($1 + 2 + 3 = 6$). El enfoque es usar una Dimensión Tipo 1 (escribiendo el dato actual), pero agregar un par adicional de columnas con las fechas de validez (Tipo 2) (27).
- Se decidió con previo consentimiento del cliente que, para la gestión del cambio en las dimensiones en el MD para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA, la estrategia de SCD a utilizar fuese de Tipo 2, donde los valores de las dimensiones pueden cambiar en el tiempo.

3.2.5. *Gestión de los metadatos del proceso de integración*

Los metadatos son datos sobre los datos, o información descriptiva sobre los datos y otras estructuras, como objetos, RN y procesos que manipulan los datos. Con esta información se puede saber el momento en que se creó un archivo, quién lo creó, cuándo fue actualizado la última vez, su tamaño y su extensión, entre otros.

Según Ralph Kimball, los metadatos se pueden dividir en tres categorías (28):

Metadatos técnicos: se usan a menudo por un personal más técnico, tal como los desarrolladores. Incluye temas como las definiciones de tablas y tipos de datos. Estos objetos son utilizados frecuentemente durante el diseño de la aplicación y el proceso de desarrollo.

- Ejemplos: la definición de la fuente y el destino, sus estructuras de tabla, campos y atributos, la documentación para las derivaciones de auditoría y dependencias.

Metadatos del negocio: ayudan a definir los términos en el lenguaje cotidiano, sin reparos a la implementación técnica. Por ejemplo, el lenguaje utilizado para describir un cliente y la forma en que se categoriza, a menudo es específico del negocio, y podría diferir entre las divisiones de la compañía.

- Ejemplos: las reglas comerciales, gestión, definiciones comerciales, la terminología de auditoría, glosarios, algoritmos y linaje que utilizan el lenguaje comercial.

Metadatos de proceso: se refiere a los metadatos generados y capturados cuando se ejecuta un proceso. Permite que los administradores gestionen su sistema y asegura que los procesos funcionen

sin problemas. Si hay un problema con alguno de ellos, los metadatos operacionales también ayudan a los administradores a identificar y localizar los problemas.

- Ejemplos: información acerca de la ejecución de las aplicaciones, incluyendo la frecuencia, conteos de registro, un análisis de componente por componente y otras estadísticas con fines de auditoría.

Para la investigación que se realiza se decidió utilizar metadatos de procesos, para los cuales se utilizó la herramienta Pentaho Data Integration 5.4.0.1, que permite gestionar y almacenar en la base de datos las informaciones referentes a las transformaciones y los trabajos como, por ejemplo, el nombre, la fecha de inicio y fin de la ejecución, los errores que se produjeron, y otra serie de características que fueron almacenadas. Todos estos datos guiaron los procesos de ETL, permitiendo que los mismos tuviesen la calidad requerida.

3.3. Implementación del subsistema de visualización de datos

Haciendo uso del mapa de navegación se puede obtener una visualización más detallada de cómo se brinda la información, así como su organización. El MD para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA está estructurado por un A.A, que contiene cinco L.T dentro de los cuales se encuentran las 44 tablas de salidas (TS) o reportes que fueron implementados. Seguidamente se muestra la estructura que posee la capa de visualización:

Descripción del Área de Análisis (A.A)

- A.A SAMKOSF: congrega toda la información concerniente a la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA.

Descripción de los Libros de Trabajo (L.T)

- L.T 01- Ingresos: contiene 24 reportes relacionados con los ingresos.
- L.T 02- Tráfico telefónico: contiene nueve reportes relacionados con el tráfico telefónico.
- L.T 03- Tarifas telefónicas: contiene dos reportes relacionados con las tarifas telefónicas.
- L.T 04- Líneas telefónicas: contiene tres reportes relacionados con las líneas telefónicas
- L.T 05- Datos e Internet: contiene seis reportes relacionados con el comportamiento de datos e Internet.

A continuación, se muestra la TS o reporte “Cumplimiento de los Ingresos de la División en MLC”, el cual se encuentra dentro del L.T 01-Ingresos:

concepto	2014		2015							
	Marzo		Marzo							
	Medidas		Medidas							
	• Real	• Plan	• Real	• Desv.	• %	• Plan Ac.	• Real Ac.	• Desv. Ac.	• % Ac.	
INGRESOS TOTALES	79,556.94	77,848	90,581.47	12,734	16	220,448	239,648.60	19,200	8.71	
VENTAS DE MERCANCIAS	695.19	1,929	2,349.53	421	22	4,793	6,670.52	1,877	39.17	
VENTA DE MERCANCIAS TELEFONÍA BÁSICA	136	149	215.49	67	45	533	726.98	194	36.52	
APARATOS TELEFÓNICOS ALÁMBRICOS (INSTALACIÓN)		0		0	-100	0		0	-100	
APARATOS TELEFÓNICOS ALÁMBRICOS (NUEVOS PRODUCTOS)	39.14	40	58.22	18	46	137	157.26	20	14.49	
APARATOS TELEFÓNICOS INALÁMBRICOS (NUEVOS PRODUCTOS)	70	76	134.17	58	76	289	493.42	205	71.02	
VENTA DE OTROS EQUIPOS	30	36	28	-8	-23	122	91	-31	-26	
DEVOLUCIONES EN VENTAS	3	3	3	0	-3	12	11	-1	-8	
REBAJAS Y BONIFICACIONES EN VENTAS	0	1	1.68	1	225	3	3.07	0	-3.35	
VENTA PBX COMERCIAL	270.12	508	1,105.68	597	118	614	1,633.79	1,019	165.95	
INGRESOS NACIONALES	29,116.47	31,325	44,439.77	13,115	42	86,642	117,748.36	31,107	35.90	
SERVICIOS INTERNACIONALES DE TELEFONIA BASICA	0	25	0	-25	-99	60	0	-60	-100	
SERVICIOS DE DATOS E INTERNET	6,374.26	6,931	9,024.22	2,094	30	20,062	26,274.35	6,213	30.97	
SERVICIOS DE TELEFONIA BASICA	6,188.37	6,051	6,315.53	265	4	17,983	18,381.39	399	2.22	

Fig 17. Reporte "Cumplimiento de los Ingresos de la División en MLC."

3.3.1. Configurar la seguridad de los usuarios y roles

Con la finalidad de robustecer la seguridad del sistema, en el transcurso de la implementación del subsistema de visualización del MD para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA, se establecieron dos roles y cinco usuarios que se distinguen en cuanto a los permisos para acceder a la información. Seguidamente se puntualiza cada uno de ellos:

- **Rol de administrador:** posee todos los permisos sobre la aplicación y lo ocupa el usuario administrador del sistema.
- **Rol de analista:** posee permiso de sólo lectura y lo ocupa el usuario especialista del sistema.

3.4. Pruebas

En este epígrafe se realizarán las validaciones al sistema y las pruebas pertinentes que comprueben la calidad del producto, como son los Casos de Prueba, con el fin de comprobar si los resultados obtenidos a partir de la entrada de datos satisfacen los resultados esperados.

3.4.1. Pruebas de software

En la construcción de un software se corre el riesgo de cometer errores, puesto que el mismo transcurre por muchas etapas complejas. Para evitar esto, durante todo el proceso de desarrollo es

importante que el mismo esté acompañado de una actividad que permita detectar posibles fallos en la implementación, calidad o usabilidad de este. Durante las diferentes fases del ciclo de vida del software se integran las pruebas de software, a través de las cuales se podrá constatar la calidad del producto y así garantizar que este cumpla con las necesidades del cliente. A continuación, se muestran algunas de las pruebas que pueden ser utilizadas para la validación de un software (29).

- ✓ **Pruebas unitarias:** esta prueba centra el proceso de verificación en la menor unidad del diseño del software (el componente de software o módulo).
- ✓ **Pruebas de integración:** consiste en construir el sistema a partir de los distintos componentes y probarlo con todos integrados. Estas pruebas deben realizarse progresivamente.
- ✓ **Pruebas de regresión:** consiste en volver a ejecutar un subconjunto de pruebas que han sido llevadas a cabo anteriormente, para asegurarse que los cambios que se hayan realizado no introduzcan un comportamiento no deseado o errores adicionales.
- ✓ **Pruebas del sistema:** se refiere al comportamiento del sistema integrado. Estas se aplican generalmente para probar los requerimientos de la solución.
- ✓ **Pruebas de aceptación:** se realizan para probar que el sistema cumpla con los requerimientos y expectativas del cliente. Estas se pueden distinguir entre dos pruebas:
 - **Pruebas alfa:** las realiza el usuario en presencia del personal de desarrollo del proyecto haciendo uso de una máquina preparada para tal fin.
 - **Pruebas beta:** las realiza el usuario después de que el equipo de desarrollo les entregue una versión casi definitiva del producto.

CALISOFT propone el Modelo V para garantizar la calidad del producto final, en el presente trabajo se determinó usar dicho modelo para guiar todas las pruebas. A continuación, se muestra un esquema del ciclo de vida del software propuesto en el Modelo V, donde a la izquierda se encuentran las diferentes etapas de desarrollo y a la derecha las pruebas correspondientes a cada una de ellas, como se muestra en la Figura 18:

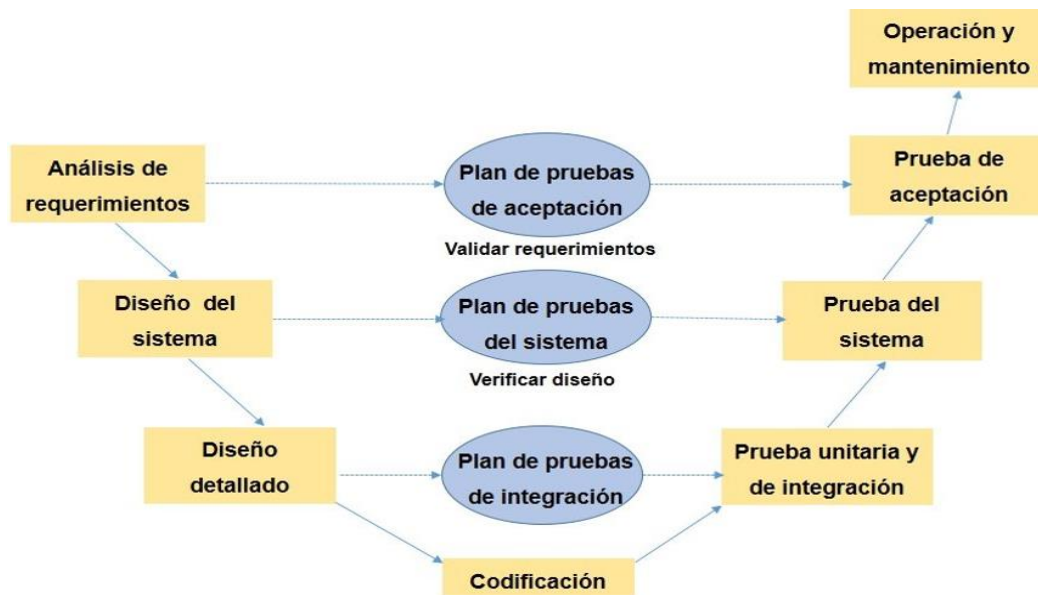


Fig 18. Modelo V.

3.4.2. Diseño de los Casos de Prueba

Los Casos de Prueba (CP) utilizados por el analista, consisten en un grupo de variables o condiciones mediante las cuales se podrá determinar si los requisitos de una aplicación son parciales o completamente satisfactorios y para identificar los posibles fallos de implementación. Con el objetivo de verificar los RI identificados en la etapa de análisis, se diseñaron cinco CP basados en los cinco CUI y dos CP basados en reglas de transformación, los cuales están recogidos en el Expediente de Proyecto. Seguidamente se muestra en la Figura 19 el CP correspondiente al CUI “Presentar información relacionada con ingresos”, donde se aprecia el reporte “Cumplimiento de los Ingresos de la División en MLC.”, contenido en el L.T 01- “Ingresos”.

Escenario	Descripción	Variables de entrada	Variables de salida	Respuesta del sistema	Flujo central
EC 1.1 L.T 01- Ingresos_Análisis del cumplimiento de los ingresos nacionales MLC en el mes.	Muestra los ingresos a nivel nacional de las ventas de mercancías y servicios en el mes seleccionado y desde el mes de enero hasta el dicho mes desglosado por sus distintos conceptos, en un tiempo determinado, en MLC.	Concepto Tiempo Moneda	Real del periodo anterior. Plan. Real. Desviación. Porcentaje. Plan acumulado. Real acumulado. Desviación acumulada. Porcentaje acumulado.	Se muestra la información correspondiente al escenario.	1. Se selecciona el área de análisis A.A SAMKOSF. 2. Se selecciona el L.T 01- Ingresos. 3. Se selecciona el reporte que corresponde al escenario especificado. 4. Se visualiza el reporte en el área de trabajo.

Fig 19. Caso de prueba basado en el CUI “Presentar información relacionada con ingresos”.

3.5. Resultado de las pruebas

Las pruebas realizadas al MD para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA arrojaron los siguientes resultados:

- **Pruebas unitarias y de integración:** Las pruebas unitarias y de integración realizadas al subsistema de integración de datos y a los diferentes componentes relacionados con la capa de visualización, detectaron en la primera iteración cuatro No Conformidades (NC). Una vez resueltas las NC detectadas se realizó una segunda iteración donde fue verificado el cumplimiento de la solución a dichas NC. Las pruebas consistieron en una serie de consultas SQL aplicadas a la base de datos que comprobaron la coincidencia de los datos de las fuentes con los almacenados en el subsistema de almacenamiento. La siguiente consulta SQL corresponde al RI4:

SELECT

```
dim_concepto.descripcion, dim_dpa.nombre_filial, dim_tiempo.anno_nombre,  
dim_tiempo.mes_nombre, hech_sisrep.plan, hech_sisrep."real"
```

FROM

```
mart_mko_dvsf.hech_sisrep, dimensiones.dim_moneda,  
dimensiones.dim_tiempo, dimensiones.dim_dpa,  
mart_mko_dvsf.dim_concepto
```

WHERE

```
dim_moneda.dim_moneda_id = hech_sisrep.dim_moneda_id AND  
dim_tiempo.dim_tiempo_id = hech_sisrep.dim_tiempo_id AND  
dim_concepto.dim_concepto_id = hech_sisrep.dim_concepto_id AND  
dim_dpa.dim_dpa_id = hech_sisrep.dim_dpa_id AND  
dim_moneda.dim_moneda_id = 1 AND  
dim_tiempo.anno_nombre = '2015' AND  
dim_tiempo.mes_nombre = 'Marzo' AND  
dim_dpa.nombre_filial = 'ETECSA NACIONAL' AND  
dim_concepto.descripcion IN ('INGRESOS TOTALES', 'VENTAS DE MERCANCIAS',  
'VENTA DE MERCANCIAS TELEFONÍA BÁSICA',  
'APARATOS TELEFÓNICOS ALÁMBRICOS (INSTALACIÓN)',  
'APARATOS TELEFÓNICOS ALÁMBRICOS (NUEVOS PRODUCTOS)', 'APARATOS  
TELEFÓNICOS INALÁMBRICOS (NUEVOS PRODUCTOS)', 'VENTA DE OTROS EQUIPOS',  
'DEVOLUCIONES EN VENTAS', 'REBAJAS Y BONIFICACIONES EN VENTAS',
```

'VENTA PBX COMERCIAL', 'INGRESOS NACIONALES', 'SERVICIOS INTERNACIONALES DE TELEFONIA BASICA', 'SERVICIOS DE DATOS E INTERNET', 'SERVICIOS DE TELEFONIA BASICA');

- **Pruebas del sistema:** Las pruebas del sistema ejecutadas mediante los CP diseñados para cada uno de los CUI identificados, arrojaron 24 NC en la primera iteración. Después de corregidas las NC detectadas, se ejecutó una segunda iteración donde se comprobó la eliminación de dichas NC y se detectaron otras cinco, que posterior a su erradicación, se verificó en una tercera iteración que el sistema quedó libre de NC. Los resultados obtenidos se muestran a través de un gráfico de barras (Ver Figura 20) que contiene las NC y su distribución por iteraciones.

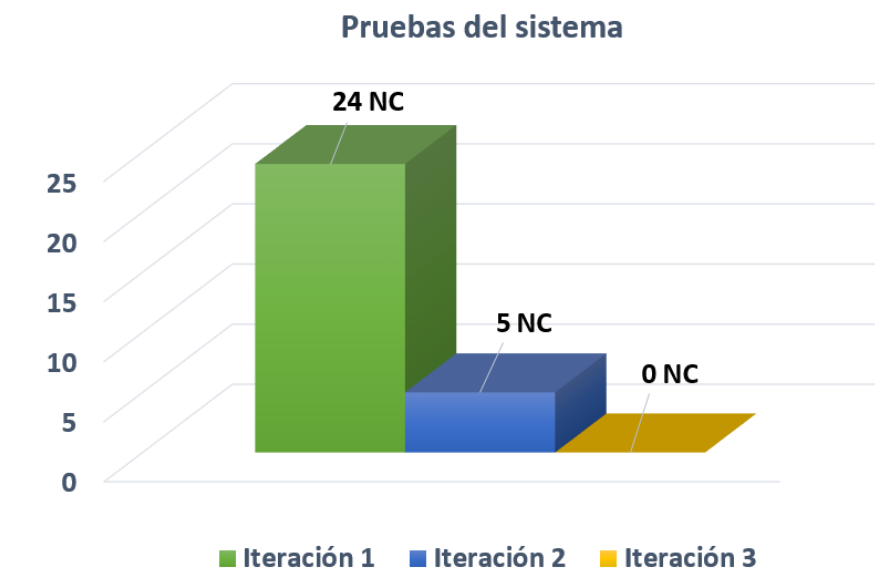


Fig 20. Resultados de las pruebas del sistema.

- **Pruebas de aceptación:** En conjunto con el cliente se realizaron las pruebas de aceptación de la solución, las cuales arrojaron resultados satisfactorios, quedando comprobado que el sistema cumple con sus necesidades y que están satisfechos con el producto elaborado.

Conclusiones del capítulo

Después de haber realizado la implementación y pruebas del Mercado de Datos para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA, se arribaron a las siguientes conclusiones:

- Durante la implementación del subsistema de almacenamiento quedaron definidas las estructuras físicas del MD con 18 tablas y tres esquemas.

- En la implementación del subsistema de integración, se ejecutaron un total de 24 transformaciones: diez para la carga de las dimensiones y 14 para los hechos, de esta manera se logró poblar el MD satisfactoriamente. Además, se definió que el tipo de SCD a utilizar fuese de Tipo 2 para la gestión del cambio en las dimensiones; y que los metadatos a utilizar fuesen los de procesos, permitiendo que se gestionaran y capturaran los datos cada vez que se ejecute un proceso, y asegurando que los mismos funcionaran sin problemas.
- A través de la implementación del subsistema de visualización quedó definida un Área de Análisis, cinco Libros de Trabajos y 44 reportes, permitiendo la visualización de la información. Además, se establecieron los roles y permisos de los usuarios que accederán a dicha información, con el objetivo de proporcionar seguridad al sistema.
- Con la utilización de los casos de pruebas realizados por reglas de transformación y por CUI para validar los procesos de ETL y de Inteligencia de Negocio respectivamente, se logró probar el MD, y así determinar que el sistema cumpliera con los requisitos del cliente.

Conclusiones generales

En la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA, se detectaron una serie de deficiencias con respecto al manejo de la información, lo cual dificultaba el análisis de la misma. Para aminorar estos inconvenientes, se desarrolló el Mercado de Datos para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA, el cual expuso resultados satisfactorios y cumple con los objetivos propuestos en la investigación:

1. A raíz de la investigación sobre los fundamentos teóricos de los almacenes de datos, se seleccionó la metodología, las herramientas y tecnologías a utilizar en el desarrollo del MD, quedando la Metodología de Desarrollo para Proyectos de Almacenes de Datos como metodología utilizada, así como diferentes herramientas que permitieron la correcta implementación del MD.
2. A partir del estudio preliminar del negocio, la fase de análisis y diseño del Mercado de Datos para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA, permitió identificar los requisitos de información, los funcionales y no funcionales. Además, se diseñaron los tres subsistemas fundamentales (almacenamiento, integración y visualización), arrojando como elementos esenciales el modelo dimensional, el perfilado de los datos, el diseño general de las transformaciones, la arquitectura de información del sistema y los reportes candidatos, los cuales sirvieron como base para la implementación de la solución.
3. Como resultado final, se implementó y probó el Mercado de Datos para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA, a través de la implementación de los tres subsistemas, lo que contribuyó a que los datos estuviesen organizados en una estructura física (esquemas y tablas), fuesen cargados satisfactoriamente hacia dicha estructura y se visualizara la información correcta para su posterior análisis. Además, se comprobó utilizando los casos de pruebas, pruebas unitarias, de integración, de sistema y de aceptación, que el sistema cumple con las necesidades del cliente.

Recomendaciones

- Mantener el soporte a las tecnologías utilizadas en el desarrollo del MD para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA.
- Utilizar la presente investigación como base de referencia para la construcción de los otros MD de cada dirección, que conformarán el Almacén de Datos Corporativo de la Empresa.

Referencias bibliográficas

1. SAINZ, Ana. Marketing Estratégico y Operativo. Universidad Politécnica de Madrid. 2009, [citado el: 8 de Octubre de 2015], pp. 1-115. Disponible en: http://ocw.upm.es/teoria-de-la-senal-y-comunicaciones-1/sistemas-de-telecomunicacion-2011/Contenidos/Material_de_clase/t.8/consulta/Marketing_Estrategico_y_Operativo.pdf
2. INMON, William H. *Building the Data Warehouse*. Fourth ed. Indianapolis (USA): Wiley Publishing, Inc, 2005. 576 p. ISBN 0-7645-9944-5.
3. KIMBALL, Ralph AND ROSS, Margy. *The Data Warehouse Toolkit, the Complete Guide to Dimensional Model*. Second ed. New York (USA): Wiley Publishing, Inc, 2002. 447 p. ISBN 0-471-20024-7.
4. LÓPEZ GARCÍA, David. Análisis de las posibilidades de uso de Big Data en las organizaciones. Universidad de Cantabria. 2013, [citado el: 23 de Septiembre de 2015], pp. 1-75. Disponible en: <http://repositorio.unican.es/xmlui/bitstream/handle/10902/4528/TFM%20-%20David%20L%C3%B3pez%20Garc%C3%ADa.pdf?sequence=1>
5. PAREKH, Arpit. Introduction on Data Warehouse with OLTP and OLAP. *International Journal Of Engineering And Computer Science*. 2013, vol. Volumen 2, no. 8 [citado el: 24 de Septiembre de 2015], pp. 2569-2573. ISSN 2319-7242. Disponible en: <http://www.ijecs.in/issue/v2-i8/44%20ijecs.pdf>
6. HERNÁNDEZ, Orallo José, RAMÍREZ, Quintana M AND FERRI, Ramírez Cèsar. *Introducción a la Minería de Datos*. Madrid (España): Pearson Prentice Hall, 2005. 656 p. ISBN 84-205-4091-9.
7. IBARRA, Maria de los Angeles. Procesamiento Analítico en Línea (OLAP). Universidad Nacional del Nordeste. 2006, [citado el: 28 de Septiembre de 2015], pp. 1-46. Disponible en: <http://exa.unne.edu.ar/informatica/SO/OLAPMonog.pdf>
8. FERRARI, Alberto AND RUSSO, Marco. Introduction to the SQLBI Methodology. 2008, [citado el: 13 de Diciembre de 2015], pp. 44. Disponible en: <https://www.sqlbi.com/wp-content/uploads/Introduction-to-SQLBI-Methodology-draft-1.0.pdf>
9. LEONARD BRIZUELA, Eric Ismael AND CASTRO BLANCO, Yudi. Metodologías para desarrollar Almacén de Datos. *Revista de Arquitectura e Ingeniería*. 2013, vol. 7, no. 3 [citado el: 24 de Octubre de 2015], pp. 1-12. ISSN 1990-8830. Disponible en: <http://www.redalyc.org/articulo.oa?id=193930080003>
10. ZEPEDA SÁNCHEZ, Leopoldo Zenaido. *Metodología para el Diseño Conceptual de Almacenes de Datos*. Tesis Doctoral. Universidad Politécnica de Valencia, 2008. pp. 211. Disponible en:
11. GONZÁLEZ HERNÁNDEZ, Yanisbel. *Metodología de Desarrollo para Proyectos de Almacenes de Datos*. Tesis de Maestría. Universidad de las Ciencias Informáticas, La Habana (Cuba), 2013. Disponible en: http://repositorio_institucional.uci.cu/jspui/handle/ident/8094

12. GONZÁLEZ HERNÁNDEZ, Yanisbel. Propuesta de Metodología para el Desarrollo de Almacenes de Datos en DATEC. *UCIENCIA*. 2012, pp. 1-15. Disponible en: http://repositorio_institucional.uci.cu/jspui/handle/ident/4323
13. MÉNDEZ, Gonzalo. Ingeniería de Requisitos. Universidad Complutense de Madrid. 2009, [citado el: 18 de Febrero de 2016], pp. 63. Disponible en: <http://www.fdi.ucm.es/profesor/gmendez/docs/is0809/03-requisitos.pdf>
14. LEÓN GUZMÁN, Elizabeth. Inteligencia de Negocios. Universidad Nacional de Colombia. 2015, pp. 1-39. Disponible en: <http://disi.unal.edu.co/~eleonguz/cursos/bda/presentaciones/Sesion2-IntroduccionBI.pdf>
15. ESPINOSA, Roberto. 2010. *Herramientas ETL. ¿Que son, para que valen?. Productos mas conocidos. ETL's Open Source*. Dataprix. [citado el: 6 de Octubre de 2015]. Disponible en: <http://www.dataprix.com/blogs/respinosamilla/herramientas-etl-que-son-para-que-valen-productos-mas-conocidos-etl-s-open-sour>.
16. KUMARI, Navita. Business Intelligence in a Nutshell. *International Journal Of Engineering And Computer Science*. 2013, vol. 1, no. 4 [citado el: 11 de Octubre de 2015], pp. 969-975. ISSN 2320 – 9801. Disponible en: http://www.ijrcce.com/upload/2013/june/26_BUSINESS.pdf
17. Introducción a Herramientas CASE y System Architect. Universidad Politécnica de Valencia. [citado el: 11 de Octubre de 2015], pp. 1-13. Disponible en: http://users.dsic.upv.es/assignaturas/eui/mtp/doc-practicas/intro_case_SA.pdf
18. GUPTA, Meenu, YADAV, Rajeev AND MEMORIA, Minakshi. UML Modelling: The Rational E-Development. *International Journal of Electronics and Computer Science Engineering*. vol. 2, no. 3 [citado el: 2 de Octubre de 2015], pp. 854-860. ISSN 2277-1956. Disponible en: <http://www.ijecse.org/wp-content/uploads/2013/06/Volume-2Number-3PP-854-860.pdf>
19. BERNABEU, Darío. 2009. *SGB*. Dataprix. [citado el: 28 de Septiembre de 2015]. Disponible en: <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/i-data-warehousing-investigacion-y-sistematizacion-concepto-14>.
20. MARTINEZ GUERRERO, Rafael. 2010. *Sobre PostgreSQL*. PostgreSQL-es. [citado el: 28 de Septiembre de 2015]. Disponible en: http://www.postgresql.org.es/sobre_postgresql.
21. OBE, Regina AND HSU, Leo. *PostgreSQL Up and Running*. Sebastopol, California (USA): O'Reilly Media, Inc., 2012. 164 p. ISBN 978-1-449-32633-3.
22. HATLE, Shrey, SAYEED, Ashiqa AND GUPTA, Deepti. Pentaho Data Integration Tool. Universidad de Maryland. 2013, [citado el: 11 de Octubre de 2015], pp. 1-19. Disponible en: <http://casci.umd.edu/wp-content/uploads/2013/12/FinalReport-Pentaho.pdf>

23. HYDE, Julian, WOOD, Sherman AND EMBERSON, Richard. Developing OLAP solutions with Mondrian/JasperAnalysis. *Mondrian 3.0.4 Technical Guide*. 2009, [citado el: 15 de Octubre de 2015], pp. 1-254. Disponible en: <http://mondrian.pentaho.org>, <https://sourceforge.net/projects/mondrian> and https://sourceforge.net/project/showfiles.php?group_id=35302
24. CADAHÍA FERNÁNDEZ, Claudia Daniela. *Mercado de datos Series históricas de turismo para el Sistema de Información de Gobierno*. Trabajo de Diploma. Universidad de las Ciencias Informáticas, La Habana (Cuba), 2012. Disponible en: http://repositorio_institucional.uci.cu/jspui/handle/ident/TD_05906_12
25. GARCÍA MATTÍO, Mariano AND BERNABEU, Dario R. *Pentaho 5.0 Reporting By Example Beginner's Guide*. Birmingham (Mumbai): Packt Publishing Ltd, 2013. ISBN 978-1-78216-224-7.
26. PADIERNA BEDOYA, Edgar David 2012. *Pentaho - Business Intelligence*. Prezi. [citado el: 11 de Octubre de 2015]. Disponible en: <https://prezi.com/zqx29l0rwmhh/pentaho-business-intelligence/>.
27. NUNO SANTOS, Vasco AND BELO, Orlando. No Need to Type Slowly Changing Dimensions. En: *Information Systems*. Avila (Spain). 2011. pp. 129-136. [citado el: 3 de Mayo de 2016]. Disponible en: https://www.researchgate.net/publication/259293002_No_Need_to_Type_Slowly_Changing_Dimensions
28. KIMBALL, Ralph AND CASERTA, Joe. *The Data Warehouse ETL Toolkit*. Indianapolis (USA): Wiley Publishing, Inc, 2004. 491 p. ISBN 0-764-57923-1.
29. SOMMERVILLE, Ian. *Ingeniería del Software*. Séptima ed. Madrid (España): Prentice Hall, 2005. 687 p. ISBN 8478290745.

Glosario de términos

- **Almacén de Datos:** es una Base de Datos corporativa caracterizada por integrar y depurar información de una o más fuentes distintas, permitiendo su análisis desde diversas perspectivas.
- **Área de Análisis:** agrupación de información según un propósito, aunque el criterio depende de las necesidades de la institución o empresa donde se aplica el sistema. Permite restringir el número de usuarios que acceden a los datos.
- **BI:** conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada (interna y externa a la organización) en información estructurada, para su explotación directa o para su análisis y conversión en conocimiento, dando así soporte a la toma de decisiones sobre el negocio.
- **Cubo:** colección de dimensiones y medidas en un área temática particular.
- **Casos de Uso del Sistema (CUS):** proceso dentro del negocio que se estudia, por lo que se corresponde con una secuencia de acciones con un orden lógico, y que producen un resultado observable para ciertos actores del negocio.
- **Dimensión:** característica de un hecho que permite su análisis posterior en el proceso de toma de decisiones y brinda una perspectiva adicional a un hecho dado.
- **ETL:** proceso a través del cual se gestionan datos obtenidos de múltiples fuentes, con el fin de extraerlos, transformarlos y cargarlos en bases de datos especializadas, denominadas Mercado de Datos, para analizar y apoyar una determinada línea de producto o unidad de negocios.
- **Hecho:** operación que se realiza en el negocio la cual está estrechamente relacionada con el tiempo y es objeto de análisis para la toma de decisiones.
- **Jerarquía:** implica una organización de niveles dentro de una dimensión, donde cada nivel representa el total agregado de los datos del nivel inferior.
- **Libro de Trabajo:** estructura organizativa que agrupa los reportes generados dentro de las Áreas de Análisis. Puede ser creado teniendo en cuenta criterios que permitan organizar la información: emisor de los reportes, receptor del reporte, contenido, entre otros.
- **Mercado de Datos:** es una base de datos departamental, que se especializa en el almacenamiento de los datos de un área específica, brindando una estructura óptima para analizar los procesos que tienen lugar dentro del departamento. Están orientados a temas específicos y contienen datos de solo una línea del negocio.
- **RN:** regla del negocio.
- **SGBD (Sistema Gestor de Bases de Datos):** conjunto de programas que permiten crear y mantener una base de datos, asegurando su integridad, confidencialidad y seguridad.

Anexos



**Mercado de Datos para la Dirección de Mercadotecnia
Operativa de la División de Servicios Fijos**

ACTA DE ACEPTACIÓN

Entre la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos de ETECSA, representado en este acto por el ciudadano cubano Heriberto Llera Hernández, mayor de edad, portadora de carné de identidad 69122001502 en su condición de Especialista de la Dirección de Mercadotecnia, facultado para este acto, que en lo sucesivo se denominará la "PARTE CLIENTE", por una parte; y por la otra el equipo de desarrollo del **Mercado de Datos para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos**, representado en este acto por los ciudadanos Frank Delgado Rodríguez y Andry Javier Vichot Chavez, ciudadanos cubanos, mayores de edad, portadores de carné de identidad 77012802302 89071217706 respectivamente, quienes actúan en su condición de TESISTAS, suficientemente facultados para este acto, que en lo sucesivo se denominará la "PARTE PROVEEDORA"; acuerdan expresamente que:

La Parte Cliente, luego de concluir el proyecto **Mercado de Datos para la Dirección de Mercadotecnia Operativa de la División de Servicios Fijos**, determina que el mismo se efectuó satisfactoriamente.

Comentarios:

Y para que así conste se suscribe la presente Acta en La Habana a los 9 días del mes de JUNIO de 2016.

<p>Entrega: </p> <p>Nombre y apellidos: <u>FRANK DELGADO RODRIGUEZ</u></p> <p>Cargo: _____</p>	<p>Recibe: </p> <p>Nombre y apellidos: <u>Heriberto Llera Hernández</u></p> <p>Cargo: <u>Especialista DVSF</u></p>
<p>Entrega: </p> <p>Nombre y apellidos: <u>Andry Javier Vichot Chavez</u></p> <p>Cargo: _____</p>	<div data-bbox="1209 1207 1461 1480" style="text-align: center;">  </div>

Fig 21. Acta de aceptación del cliente