

Universidad de las Ciencias Informáticas
Facultad de Ciencias y Tecnologías Computacionales



Título: Mercado de Datos para el Departamento de Mercadotecnia
perteneiente a la Dirección de Transferencia de Tecnología y Conocimiento

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

Autores:

Midalis Rodríguez Rosado
Jorge Alfonso Torres

Tutores:

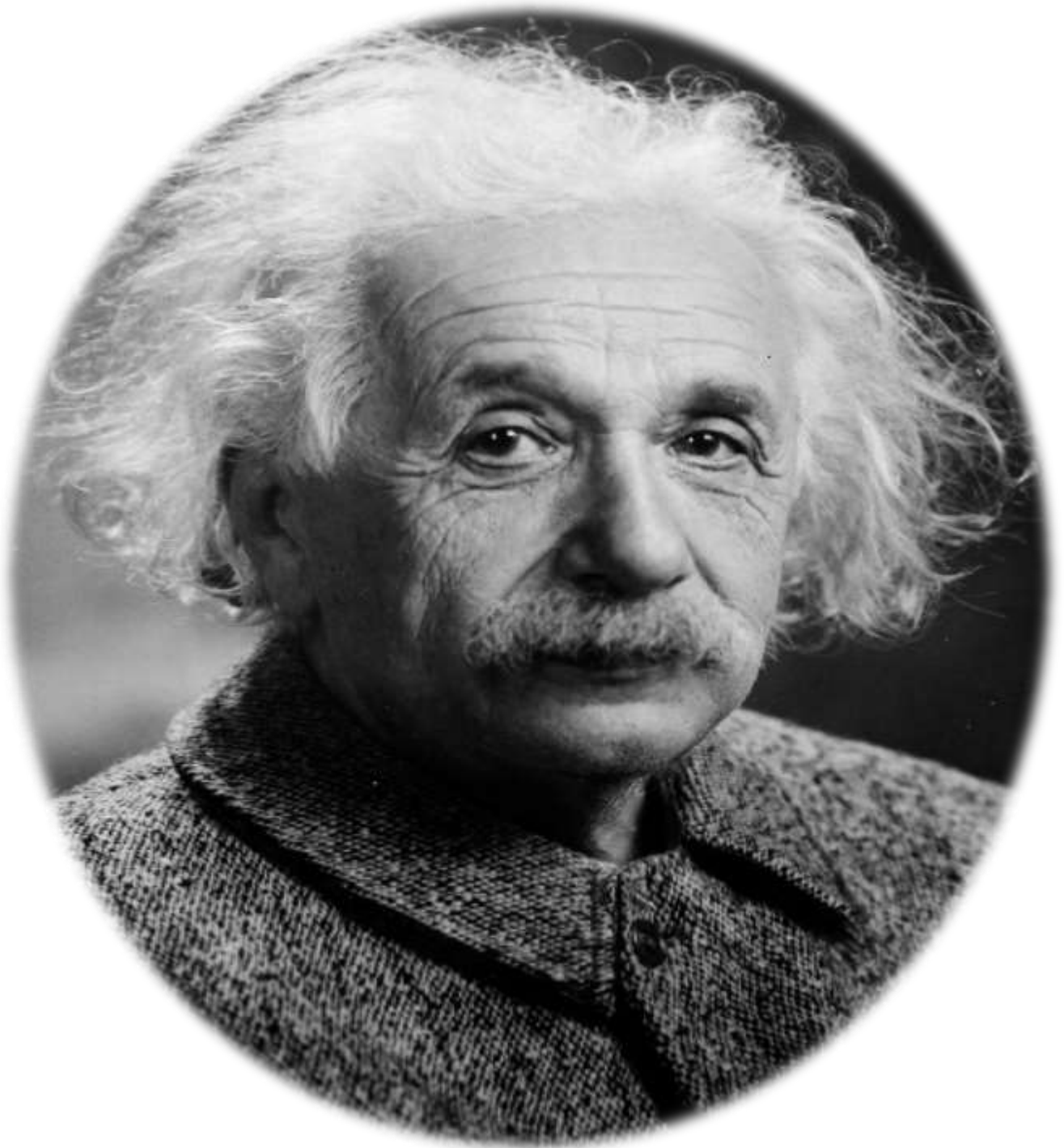
Ing. Aleida Eva Sáez Aldana
Ing. Elian Eduardo Castell Legrá

Consultantes:

Ing. Nolvis Pérez Franco
Ing. Yunier Vega Rodríguez

La Habana, junio de 2017

“Año 59 de la Revolución”



Temo el día en que la tecnología sobrepase nuestra humanidad. El mundo solo tendrá una generación de idiotas.

Albert Einstein

Declaración de Autoría

Declaración de Autoría

Declaramos ser autores de la presente tesis “Mercado de Datos para el Departamento de Mercadotecnia perteneciente a la Dirección de Transferencia de Tecnología y Conocimiento” 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 _____.

Autores: _____

Jorge Alfonso Torres

Midalis Rodríguez Rosado

Tutores: _____

Ing. Elian Eduardo Castell Legrá

Ing. Aleida Eva Sáez Aldana

Datos de contacto

Tutor: Ing. Elian Eduardo Castell Legrá ecastell@uci.cu

Especialidad de graduación: Ingeniería en Ciencias Informáticas

Años de graduado: 1

Área de investigación: Base de datos

Años de experiencia en el tema: 1

Tutor: Ing. Aleida Eva Sáez Aldana aleidaeva@uci.cu

Especialidad de graduación: Ingeniería en Ciencias Informáticas

Años de graduada: 13

Área de investigación: Base de datos

Años de experiencia en el tema: 9

Consultante: Ing. Nolvis Pérez Franco nfranco@uci.cu

Especialidad de graduación: Ingeniería en Ciencias Informáticas

Años de graduada: 4

Área de investigación: Mercadotecnia

Años de experiencia en el tema: 2

Consultante: Ing. Yunier Vega Rodríguez yvrodriguez@uci.cu

Especialidad de graduación: Ingeniería en Ciencias Informáticas

Años de graduado: 8

Área de investigación: Mercadotecnia

Años de experiencia en el tema: 6

Autor: Midalis Rodríguez Rosado

Correo electrónico: mrosado@estudianates.uci.cu

Autor: Jorge Alfonso Torres

Correo electrónico: jtorres@estudiantes.uci.cu

Agradecimientos

Resulta difícil agradecer, en tan solo pocas palabras, a todas aquellas personas que de una forma u otra han contribuido a que yo hoy, haga realidad mi sueño. Es por ello que quiero agradecer a:

Mi dúo de tesis, por confiar en mi para lograr este sueño ya cumplido. Hicimos un buen equipo.

A Elian por ser el mejor tutor del mundo, por no dudar ni un segundo cuando se lo propusimos, por apoyarnos siempre. Agradecer a Nolvis, fue con ella que dimos nuestros primeros pasos y nunca dudó en tendernos la mano en todo momento que estuvo. Muchas gracias.

Agradecer al tribunal por todos sus consejo y críticas constructivas. A Amarelys, que, a pesar de ser mi oponente, siempre estuvo allí para corregirme y darme buenos consejos. A Juan Pablo, el que tenía la solución a todo cuando no sabíamos salir del bache. Quiero agradecer de forma general a todos mis profesores de la carrera que contribuyeron a mi formación profesional en el día a día.

A ti papá por tu nobleza y entrega en cada una de las cosas que haces, por ser mi maestro y mi ejemplo. Espero te sientas orgulloso de mi. A mi mamá que a pesar de que no se encuentra físicamente conmigo siempre la llevo en mi corazón y cada una de las decisiones que tomo lo hago pensando en ella.

No me podía faltar esa personita que llegó a mi vida en el momento que más lo necesitaba y se ha convertido en una de las cosas más importantes que me han pasado. Gracias tata por abrirme las puertas de tu corazón, por hacerme reír, por cuidarme, por darme fuerza y aliento en todo momento.

Que sería de la vida en la universidad sin esas grandes amistades que un día conoces y sin darte cuenta al forman parte de tu familia. Es por ello que quiero agradecer a Yanelis, mi compañera de cuarto y estudio inseparable, Suinny, la más loca del grupo, pero especial en todo, René, amante de las fotos, Claudia, amiga incondicional, Lizzandra, dispuesta para peinarnos siempre, Milagrito con sus ocurrencias, Teudys y Saydi siempre que los necesite estuvieron ahí, Lenia mi compañera para las guaguas los viernes, Roly, Osmin, Cano, Asiel, Leandro, Rolando, Yazmin, amistades, en fin, a todos muchas gracias por hacer de mi estancia en la UCI un recuerdo inolvidable.

A gradecer de forma generar a todas aquellas personas que me brindaron de una forma u otra su apoyo incondicional para la realización de este trabajo.

Midalis Rodríguez Rosado

Agradecimientos

Agradezco infinitamente a mi Mamá por ser una inspiración y un ejemplo a seguir. Porque sé que todo lo que hace es por mi bien, por ser una luchadora incansable y darme más de lo que se puede dar.

Gracias a mi abuela Andrea, por tanto, amor y cariño que me ha dado, por ser una segunda madre para mí y ser gran responsable de la persona quien soy hoy en día.

A mi abuelo Gerardo, que a pesar de no estar físicamente sé que está muy orgulloso de mí.

A mis abuelos maternos, por desear que fuera un profesional y por ser importante para ellos, así como ellos para mí.

A mi papá y a mi hermano, por siempre estar ahí para mí y y brindarme tanto apoyo.

A mi tío Gerardo, por todo lo que ha hecho. A mis otros tíos, por ser un hijo más para ellos.

A Annareya, por ser la mejor novia que se puede tener en el mundo, y por dedicarme sus últimos 5 años.

A Midalis, por ser el mejor dúo de tesis de toda la UCI, y porque sin ella el resultado no fuese igual.

A Elián, un tutor incomparable y que gracias a él se pudo hacer la tesis.

A Barroso, que se comportó como un tutor más.

A La Mora, por convertirse en una tía más.

A Leonardo, por estar junto a mí desde 1er año. Al Migue, por entretenerme despertándolo en 1er año.

A los Carlos, por compartir tanto juntos (principalmente cosas del Madrid). Al Lija, por mantener mi imagen. A Elián nuevamente, Leosdany, Yanitza, Claudia, Loquiño y al resto de mis compañeros del grupo.

A Carlos Leandro, por compartir música, deporte y otros buenos momentos. A la Yane, por crecernos haciendo la tesis. A Suinny, Lizzandra, Milagros, Asiel y a los demás compañeros que conocí en el grupo.

A Nina, mi doctora y una hermana mayor. A Fabián, Arley, Junior, Luis Pablo, Aliesky, Marichal, Ale, Rafa y demás del piquete, demostrando que la amistad puede seguir sin importar el tiempo.

A Dayersi, por su antigua amistad y por regalarme mi bella ahijada. A Cilic y Alvarez por tantos momentos compartidos. A Pumukel, Rosi y mi otra hermanita Lyneth.

A mis otras amistades que no he mencionado, pero no me olvido de ellas.

A todos los profesores que han aportado con mi formación.

Jorge Alfonso Torres

Dedicatoria

Dedico esta tesis a mis padres, por haberme dado educación, un hogar donde crecer, equivocarme, desarrollarme, aprender y donde adquirí los valores que hoy definen mi vida. Este título es para ustedes.

Midalis Rodríguez Rosado

Dedico esta tesis a:

A mi madre, por ser el motor impulsor de mi vida.

A Mima, por dar tanto por mí.

A Pipo, porque sé que estaría muy orgulloso.

A Aba y Papi, por creer siempre en mí.

Jorge Alfonso Torres

Resumen

La Universidad de las Ciencias Informáticas, cuenta con la Dirección de Transferencia de Tecnología y Conocimiento, la que tiene asociado un Departamento de Mercadotecnia que se encarga de la comercialización y la negociación de las diferentes ofertas que brinda la universidad. Actualmente, en el departamento elaboran informes para la toma de decisiones que incluyen tablas resumen, gráficos, listados, cálculos de indicadores complejos y análisis estadísticos. Esta tarea es afectada por el manejo manual de los datos obtenidos desde diferentes fuentes, causando su pérdida y duplicación, además del consumo de tiempo. El presente trabajo de diploma tiene como objetivo general desarrollar un Mercado de Datos para dicho departamento que contribuya al proceso de toma de decisiones. El mismo permitirá obtener, de forma organizada y rápida, los datos a través de listados y gráficos. Para guiar el proceso de desarrollo de la solución se utilizó la Metodología para el desarrollo de proyectos de Almacenes de Datos, apoyándose en las herramientas Visual Paradigm 8.0, PostgreSQL 9.4 y PgAdmin III 1.20. Para la limpieza de datos se emplea el Data Cleaner 3.1 y Data Integration 6.0 para los procesos de extracción, transformación y carga. Para el diseño de cubos, implementación de reportes y presentación de la información se utiliza el Schema Workbench 3.11, Report Designer 6.0 y el BI Server 6.0. Después del análisis, diseño e implementación de la solución, se obtuvo un Mercado de Datos poblado al que se le aplicaron un conjunto de pruebas para verificar su correcto funcionamiento.

Palabras claves: Almacén de Datos, Mercado de Datos, toma de decisiones

Abstract

The University of Computer Science has the Technology and Knowledge Transfer Directorate, which has a Marketing Department associated with the commercialization and negotiation of the different offers offered by the university. Currently, the department specialists produce decision-making reports which include summary tables, charts, lists, complex indicator calculations and statistical analyzes. This task is affected by the manual handling of the data obtained from different sources, causing its loss and duplication, besides the time consumption. This diploma work is part of the development of a Data Market for this department that contributes to the decision-making process. It will allow to obtain, in an organized and fast, the data through lists and graphs. To guide the process of developing the solution, the Methodology for the development of Data Warehouse projects was used to guide the process, using Visual Paradigm 8.0, PostgreSQL 9.4, PgAdmin III 1.20, Data Cleaner 3.1, Data Integration 6.0, Schema Workbench 3.11, Report Designer 6.0, and BI Server 6.0. The analysis, design and implementation of the solution was performed, resulting in a populated Data Market, which meets the characteristics identified by the client. A set of tests were applied to verify the correct operation of the market, throwing several nonconformities resolved in their entirety.

Key words: *Data Mart, Data Warehouse, decision making*

Índice de Contenido

Introducción	1
Capítulo 1: Fundamentación teórica de la investigación.....	6
1.1 Tecnologías de almacenamiento de datos	6
1.1.1 Base de datos.....	6
1.1.2 Almacén de Datos.....	8
1.1.3 Mercado de Datos.....	10
1.2 Elementos de un Almacén de Datos.....	10
1.3 Base de datos multidimensional	11
1.3.1 Variantes de modelamiento.....	11
1.3.2 Variantes de implementación.....	13
1.4 Metodologías para el desarrollo de AD.....	14
1.5 Herramientas a utilizar en el desarrollo de la solución	17
1.6 Conclusiones del capítulo.....	20
Capítulo 2: Análisis y diseño de la solución.....	21
2.1 Estudio preliminar del negocio.....	21
2.2 Necesidades de usuario	22
2.3 Requisitos de información	22
2.4 Requisitos funcionales.....	24
2.5 Requisitos no funcionales.....	24
2.6 Reglas del negocio	26
2.7 Diagrama de Casos de Uso del Sistema	27
2.8 Arquitectura del Mercado de Datos	31
2.8.1 Subsistema de Almacenamiento.....	32
2.8.2 Subsistema de Integración.....	34
2.8.3 Subsistema de Visualización.....	38
2.9 Políticas de respaldo y recuperación	39
2.10 Conclusiones del capítulo.....	40
Capítulo 3: Implementación y Prueba de la solución.....	41
3.1 Implementación del Subsistema de Almacenamiento.....	41
3.1.1 Estándares de codificación	41
3.2 Implementación del Subsistema de Integración.....	42
3.2.1 Transformaciones y trabajos.....	43

3.3 Implementación del Subsistema de Visualización.....	45
3.3.1 Implementación de los cubos OLAP	45
3.3.2 Implementación de la capa de visualización	45
3.3.3 Implementación de las vistas OLAP.....	46
3.3.4 Implementación de los reportes candidatos	47
3.3.5 Implementación de los Dashboard.....	48
3.4 Pruebas.....	51
3.4.1 Herramientas para la aplicación de las pruebas.....	52
3.4.2 Resultados de las pruebas.....	54
3.5 Conclusiones del capítulo.....	56
Conclusiones Generales.....	57
Recomendaciones	58
Referencias Bibliográficas.....	59
Bibliografía.....	62
Anexos.....	65

Índice de Tablas

Tabla.1: Distribución de las áreas de análisis	22
Tabla.2: Descripción de los actores del sistema	28
Tabla.3: Descripción del caso de uso “Mostrar información sobre el PTTC”	29
Tabla.4: Descripción de hechos	32
Tabla.5: Descripción de dimensiones.....	33
Tabla.6: Matriz Bus correspondiente al Departamento de Mercadotecnia.....	34
Tabla.7: Reporte Cantidad de contratos del AA PTTC	39
Tabla.8: Reporte Cantidad de visitantes del AA Visitas al Salón de Exposiciones	39
Tabla.11: Roles y permisos.....	40
Tabla.12: Estándar de codificación para el MD	41
Tabla.13: Diseño del CP para el CU “Mostrar información sobre el PTTC”	53
Tabla.14: Resultados de las pruebas de sistema	55

Índice de Figuras

Fig.1: Representación del Esquema en estrella (Elaboración propia)	12
Fig.2: Representación del Esquema copo de nieve (Elaboración propia).....	12
Fig.3: Representación del Esquema constelación o copo de estrellas (Elaboración propia)	13
Fig.4: Metodología de desarrollo (Elaboración Propia).....	16
Fig.5: Diagrama de Caso de Uso del Sistema (Elaboración propia)	28
Fig.6: Arquitectura del Mercado de Datos (Elaboración propia)	32
Fig.7: Diagrama Entidad-Relación (Elaboración propia).....	34
Fig.8: Distribución por tipo de datos (Elaboración propia)	36
Fig.9: Cantidad de datos nulos en el excel PTTC (Elaboración propia).....	36
Fig.10: Diseño de las transformaciones para la carga de dimensiones (Elaboración propia)	37
Fig.11: Diseño de las transformaciones para la carga de hechos (Elaboración propia).....	37
Fig.12: Diseño del mapa de navegación (Elaboración propia).....	38
Fig.13: Proceso de integración de datos para la dimensión estado_legal (Elaboración propia).....	43
Fig.14: Proceso de integración de datos para el hecho PTTC (Elaboración propia).....	43
Fig.15: Trabajo para las dimensiones (Elaboración propia).....	44
Fig.16: Trabajo para los hechos (Elaboración propia)	44
Fig.17: Implementación de los cubos OLAP (Elaboración propia)	45
Fig.18: Arquitectura de información del MD (Elaboración propia).....	46
Fig.19: Vista de análisis referente al PTTC (Elaboración propia)	47
Fig.20: Reporte asociado al PTTC (Elaboración propia)	48
Fig.21: Dashboard correspondiente al PTTC (Elaboración propia)	50
Fig.22: Modelo V (Elaboración propia)	51

Introducción

Las Tecnologías de la Información y las Comunicaciones (TIC), han transformado el mundo que nos rodea, siendo piezas fundamentales para la superación y desarrollo de un país. Aportan herramientas específicas, que se incluyen de forma responsable en el día a día. Se puede considerar que la incorporación de las mismas es un proceso imparabile, por lo que hay que estar a la vanguardia para adaptarse a las nuevas tecnologías que vayan surgiendo. Su implantación permite evolucionar la manera de trabajar, agilizando los procesos existentes, de manera que contribuya a la informatización de la sociedad.

Con el uso de computadoras en los distintos sectores de la sociedad, se ha visto un mayor crecimiento en la capacidad de generar y almacenar información. Mientras mayor es la capacidad de los dispositivos de almacenamiento de datos, mayor es el interés de los humanos por guardar grandes volúmenes de datos. Esto trae consigo dificultades a la hora de extraer conocimiento, realmente útil, de la información almacenada. Las empresas e instituciones se han dado cuenta de la necesidad de procesar dicha información, que es realmente eficaz, si los datos están ordenados, analizados y transformados de modo que permitan resolver problemas específicos.

Cuba cuenta con varias empresas e instituciones encargadas de informatizar la sociedad y exportar productos y servicios a otros países. Una de estas instituciones es la Universidad de las Ciencias Informáticas (UCI), la que tiene como misión, formar profesionales comprometidos con su Patria y altamente calificados en la rama de la Informática, capaces de producir aplicaciones y servicios informáticos, a partir de la vinculación estudio-trabajo, como modelo de formación y servir de soporte a la industria cubana del software.

La UCI cuenta con una Dirección de Transferencia de Tecnología y Conocimiento (DTTC), la que tiene como misión, gestionar la transferencia de tecnologías y conocimientos de la institución hacia la sociedad cubana y al exterior del país, protegiendo la propiedad intelectual de los productos y servicios resultantes de su actividad y estableciendo la política de calidad institucional. Así mismo, coordina y organiza la elaboración y actualización de la planeación estratégica, los manuales de organización y los objetivos de trabajo, controlando su cumplimiento. Cuenta con tres departamentos: el Departamento de Desarrollo y Calidad, el Departamento de Aseguramiento a la Gestión y el Departamento de Mercadotecnia.

Este último es el encargado de la gestión de la comercialización y la negociación de los servicios académicos, editoriales e informáticos, así como de las aplicaciones que son desarrolladas por los centros de producción de la universidad. Para cumplir con dichos objetivos los especialistas que allí laboran son responsables de ejecutar los siguientes procesos:

1. Registro en el Plan de Transferencia y Tecnología del Conocimiento (PTTC).
2. Registro de visitas al salón de exposiciones.
3. Registro de expedientes comerciales.
4. Registro en el Centro Nacional de Derecho de Autor (CENDA).

La información que presenta el departamento data desde 2008 y es almacenada digitalmente en ficheros excel (.xls) con una periodicidad mensual y anual, indistintamente, para cada uno de los procesos. El primero, guarda todo lo referente a los contratos que se llevan a cabo en la institución; el segundo, lo relacionado con las visitas al salón de exposiciones; el tercero, los expedientes comerciales realizados, mientras que el cuarto, archiva los productos u obras literarias a inscribir en el CENDA para el derecho de autor.

A partir de la consulta y el análisis de los datos, los especialistas generan informes mensuales en formato PDF (.pdf) solicitados por el jefe de departamento para la toma de decisiones. Con la utilización de las herramientas de Microsoft Office, la cual posee una licencia privativa y su uso requiere de grandes costos para la institución, crean tablas resumen, gráficos o listados, de manera que la información sea mostrada desde diferentes perspectivas y aristas de análisis.

Realizar estos informes resulta engorroso por parte de los especialistas, ya que, para la creación de cada uno de ellos deben analizar y copiar los datos contenidos en los .xls para un documento .doc con sus tablas y gráficos correspondientes. Se hacen, además, cálculos de indicadores complejos y análisis estadísticos, comparativos y de tendencias de los mismos en diferentes periodos de tiempo. Como consecuencia se genera la pérdida y duplicación de los datos, además de la utilización de tiempo excesivo en la conformación y entrega de la información solicitada.

El jefe de departamento analiza la información enviada de cada uno de los procesos para realizar investigaciones de mercado y análisis de factibilidad; tarea que resulta compleja, al no existir una correspondencia entre los resultados que brindan los especialistas, pues esta no se encuentra centralizada e integrada. Las problemáticas, antes enunciadas, permiten a los autores declarar que no son pocas las dificultades a las que se enfrentan los diferentes especialistas para elaborar los informes, que, de forma mensual, deben rendir al jefe de departamento; lo que genera, afectaciones en el proceso de toma de decisiones en el Departamento de Mercadotecnia perteneciente a la DTTC.

Después de un análisis de la situación problemática, queda conformado el **problema de la investigación** mediante el cuestionamiento de: ¿Cómo contribuir al proceso de toma de decisiones en el Departamento de Mercadotecnia perteneciente a la DTTC?

Se define como **objeto de estudio**: los Almacenes de Datos, enmarcado en el **campo de acción**: Mercados de Datos.

Para darle solución al problema planteado, se define como **objetivo general**: desarrollar un Mercado de Datos para el Departamento de Mercadotecnia perteneciente a la DTTC que contribuya al proceso de toma de decisiones, del cual se desglosan los siguientes **objetivos específicos**:

1. Elaborar los fundamentos teóricos de la metodología, herramientas y tecnologías de almacenamiento a utilizar en el desarrollo del Mercado de Datos para el Departamento de Mercadotecnia.
2. Realizar el análisis y diseño del Mercado de Datos para el Departamento de Mercadotecnia.
3. Realizar la implementación y prueba del Mercado de Datos para el Departamento de Mercadotecnia.

Con el propósito de cumplir con los objetivos específicos fueron formuladas las **preguntas científicas**:

1. ¿Qué metodología, herramientas y tecnologías de almacenamiento utilizar, para llevar a cabo el desarrollo del Mercado de Datos para el Departamento de Mercadotecnia?
2. ¿Cómo realizar el análisis y diseño del Mercado de Datos para el Departamento de Mercadotecnia?
3. ¿Cómo implementar el Mercado de Datos para el Departamento de Mercadotecnia de manera que contribuya al proceso de toma de decisiones?
4. ¿Cómo probar el Mercado de Datos desarrollado para el Departamento de Mercadotecnia para garantizar el correcto funcionamiento del mismo?

A partir de las preguntas científicas planteadas anteriormente se trazaron las **tareas de la investigación** que a continuación se presentan:

1. Caracterización de la metodología, herramientas y tecnologías de almacenamiento a utilizar, en el desarrollo del Mercado de Datos para profundizar en el nivel de comprensión y dominio de las mismas.
2. Levantamiento de los requisitos del sistema para identificar las necesidades del cliente.
3. Descripción de los casos de uso para determinar cada una de las funcionalidades del Mercado de Datos.
4. Definición de la arquitectura del Mercado de Datos para identificar los subsistemas fundamentales que componen la solución.
5. Definición de los hechos, las medidas y las dimensiones para identificar los elementos que forman parte del modelo lógico del Mercado de Datos.
6. Realización del modelo lógico para determinar los elementos que conforman el modelo físico.

7. Realización del diseño del subsistema de almacenamiento para determinar las estructuras dimensionales del Mercado de Datos (hechos y dimensiones).
8. Realización del diseño del subsistema de integración para diseñar los procesos de carga de las dimensiones y hechos.
9. Realización del diseño del subsistema de visualización para realizar el diseño de los cubos OLAP.
10. Implementación del subsistema de almacenamiento para la construcción física del Mercado de Datos.
11. Implementación del subsistema de integración para poblar el Mercado de Datos con los datos proveniente de los sistemas fuentes.
12. Implementación del subsistema de visualización para mostrar la información mediante reportes contenidos dentro de los libros de trabajo según el área de análisis identificada.
13. Aplicación de las pruebas para garantizar el correcto funcionamiento del Mercado de Datos.

A continuación, se explican los métodos de la metodología de la investigación científica empleados para la obtención de información y el futuro desarrollo del Mercado de Datos.

Métodos Teóricos

- ❖ **Histórico – lógico:** para el desarrollo del presente trabajo de diploma se realizó un estudio de los fundamentos teóricos relacionados con los Mercados de Datos en cuanto a definiciones, ventajas, desventajas y características.
- ❖ **Analítico – sintético:** este método se utilizó para el análisis de documentos, materiales y temas relacionados con el desarrollo de los Mercado de Datos y de elementos claves para el Departamento de Mercadotecnia. Permitió definir los conceptos fundamentales del tema y se usó para especificar la metodología a utilizar, así como los elementos que conforman un Mercado de Datos. La síntesis permitió realizar un análisis del impacto de los resultados de la solución desarrollada como parte de la validación y cumplimiento de los objetivos de la presente investigación.

Métodos empíricos

- ❖ **Entrevistas:** se utilizó para la recolección de información mediante las entrevistas no estructuradas a los clientes, lo que permitió la recogida de datos durante el proceso de ingeniería de requisitos y la consulta de materiales y normativas referentes al departamento.

Estructura del trabajo de diploma

El presente trabajo de diploma está estructurado de la siguiente manera: resumen, introducción, tres capítulos, conclusiones, recomendaciones, referencias bibliográficas, bibliografía y anexos.

Capítulo 1: Fundamentación teórica de la investigación

En este capítulo se fundamentarán los elementos teóricos que sustentan el problema de la investigación y los objetivos del trabajo. Se analizará la metodología, herramientas y tecnologías de almacenamiento de datos que se ajustan al desarrollo de la investigación, justificando la selección y utilización de cada una de ellas.

Capítulo 2: Análisis y diseño de la solución

En este capítulo se realizará un estudio más detallado del negocio para determinar cuáles son las necesidades del cliente. Se describirán los requisitos funcionales y los requisitos de información para conformar el diagrama de caso de uso del sistema, además de las reglas del negocio y los requisitos no funcionales. Se definirá la arquitectura de implementación a utilizar para el desarrollo de la solución. Se identificarán los hechos, medidas y dimensiones dando paso a la creación del modelo lógico de datos. Serán diseñados los subsistemas de almacenamiento, integración y visualización para su posterior implementación. Se establecerán políticas de respaldo en caso de fallos y roles y permiso para el correcto manejo de los datos.

Capítulo 3: Implementación y prueba de la solución

En este capítulo se realizará la implementación de los subsistemas de almacenamiento, integración y visualización, teniendo en cuenta los requisitos de información, los requisitos funcionales, los requisitos no funcionales y las reglas del negocio. Además, se expondrán los resultados obtenidos al realizar las pruebas correspondientes a la solución.

Capítulo 1: Fundamentación teórica de la investigación

Capítulo 1: Fundamentación teórica de la investigación

En este capítulo se abarcan todos los elementos teóricos que sustentan el tema de la investigación, así como definiciones por varios autores para arribar a conclusiones, ventajas y desventajas. Se define la metodología a utilizar de forma que sirva de base para el desarrollo de la presente investigación. Además, se fundamentan las herramientas y tecnologías empleadas que permitirán un correcto desarrollo del Mercado de Datos (MD).

1.1 Tecnologías de almacenamiento de datos

El hombre, en el transcurso de la vida ha tenido la necesidad de almacenar información para obtener conocimiento de ella. Para realizar el proceso de almacenamiento de datos se han empleado diferentes métodos y dentro de estos, el uso de archiveros o almacenes físicos. En la actualidad, debido al creciente desarrollo de las tecnologías y al surgimiento de la computadora, mejora considerablemente, el almacenamiento y manejo de información, lo que facilita la vida al hombre.

Durante las décadas de los 60 y 70 surge el concepto de Base de Datos (BD); sin embargo, el objetivo principal siempre ha sido la administración óptima de la información y el uso que se le puede dar a la misma. Hoy, las necesidades de las empresas han cambiado, lo que trae consigo la imperiosa interacción con diversas fuentes de información que se traducen en desafío para las BD. Lo anterior ha provocado que los volúmenes de información sean mayores, su formato muy diverso, lo que incrementa así los tiempos de respuesta para analizar la información y tomar decisiones. (Anguiano, 2014)

Por lo antes mencionado, ha sido necesario desarrollar nuevas tecnologías para satisfacer las necesidades de los usuarios. Surgieron así, los Almacenes de Datos (AD) los que, además de contener una BD bien estructurada para el almacenamiento de la información, permiten que esta se muestre desde diferentes perspectivas, utilizando gráficos para el análisis estadístico, lo que facilita el proceso de toma de decisiones en las empresas.

1.1.1 Base de datos

Desde sus inicios, las BD fueron creadas para el almacenamiento de la información, permitiéndoles a sus usuarios recuperar y actualizar la misma. Estas se convirtieron, rápidamente, en una herramienta fundamental para el análisis de la información en las empresas. El término de base de datos apareció por primera vez en 1963, en un simposio celebrado en California, en el que se definió como un conjunto de información relacionada que se encuentra agrupada o estructurada. (Valdéz, 2007). A continuación, se exponen algunas de las definiciones dadas por diferentes autores.

Capítulo 1: Fundamentación teórica de la investigación

Para (Cabello, 2001), una base de datos es un conjunto de datos persistentes que es utilizado por los sistemas de aplicación de alguna empresa.

En otro término, (Navathe, 2015) la define como una colección de datos relacionados entre sí y que tienen un significado implícito.

Según (Anguiano, 2014), una base de datos es un contenedor que permite almacenar la información de forma ordenada con diferentes propósitos y usos, lo que facilita el almacenamiento de la información de diferentes departamentos. Este almacenamiento de la información por sí solo no tiene un valor, pero si combina o relaciona la información con diferentes departamentos puede dar valor.

Después de sistematizar las definiciones de diferentes estudiosos sobre el tema, los autores de la presente investigación la definen como una colección de datos relacionados entre sí, que se encuentran organizados en diferentes tablas, para que posteriormente puedan ser localizados, utilizados fácilmente, o sean explotados por un sistema de información de una empresa.

Las BD son de vital importancia en la dinámica del mundo actual, pues están detrás de toda la marejada que se mueve en la generación, recuperación y consumo de información. Para esto es de gran importancia conocer las ventajas y desventajas que las mismas ofrecen. Según (Group, 2015) estas son:

Ventajas de las BD

- ❖ Las bases de datos permiten obtener información estructurada de forma más sencilla.
- ❖ Facilita la estandarización de procesos y los nombres de registros.
- ❖ Permite controlar la duplicidad de datos (redundancia).
- ❖ Permite la centralización y unificación de datos.
- ❖ Flexibilidad y rapidez para obtener información.
- ❖ Las bases de datos no son instalables, sino que son portables. Basta con copiarlas e importarlas.

Desventajas de las BD

- ❖ Una BD suele requerir mucho espacio en disco, suelen volverse pesadas.
- ❖ Son un producto complejo, por lo que no toda persona será capaz de manipularla y/o ponerse al cargo de su mantenimiento.

Capítulo 1: Fundamentación teórica de la investigación

- ❖ Cuando la BD crece mucho puede llegar a ponerse lenta, lo que afecta las búsquedas y la recuperación de la información.
- ❖ Las BD son vulnerables a fallos que no están en las manos de los usuarios, como una descarga eléctrica.

1.1.2 Almacén de Datos

El poder de procesamiento y almacenamiento de los datos en las empresas, cada vez mayor, ha dado como resultado la creación de nuevas tecnologías que faciliten el proceso de toma de decisiones. Los AD surgieron como respuesta a lo antes mencionado, permitiendo a sus usuarios revisar, graficar y visualizar información multidimensionalmente para mejorar el acceso a los datos, mostrándolos desde diferentes perspectivas.

El AD, es actualmente, el centro de atención de las grandes instituciones, porque provee un ambiente para que las organizaciones hagan un mejor uso de la información que está siendo administrada por diversas aplicaciones operacionales. (Cuéllar M, 2014). Este es una BD corporativa que se caracteriza por integrar y depurar información de una o más fuentes distintas, para luego procesarla, permitiendo su análisis desde infinidad de perspectivas y con grandes velocidades de respuesta. (Sinnexus, 2016)

Bill Inmon¹ lo define como una recopilación de datos temáticos, integrados, no volátiles y con historial para la toma de decisiones. (Inmon, 2005)

Ralph Kimball² lo define como una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis. (Kimball, 2013)

En la presente investigación se definen los AD como una colección de datos específicamente estructurados, integrado, temático, no volátil y con historial para el proceso de apoyo a la toma de decisiones. Los AD extraen los datos de una o más fuentes distintas, en ocasiones con diferentes formatos, para luego procesarla y permitir su posterior interpretación. También se pueden separar en unidades lógicas más pequeñas llamadas Mercado de Datos.

Características de los AD (Inmon, 2005)

Según Bill Inmon en su definición de AD plantea cuatro características que a continuación se detallan:

¹ Creador del término Data Warehouse (DW) y considerado como el padre de la disciplina.

² Especialista reconocido a nivel mundial en el diseño de DW y creador del enfoque Multidimensional.

Capítulo 1: Fundamentación teórica de la investigación

Integrado: los datos almacenados en el AD deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre las diferentes fuentes deben ser eliminadas.

Temático: los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. De esta forma, las peticiones de información específicas serán más fáciles de responder dado que toda la información reside en el mismo lugar.

Histórico: el tiempo es parte implícito de la información contenida en un AD, sirve, entre otras cosas, para realizar análisis de tendencias. Por lo tanto, el AD se carga con los distintos valores que toma una variable en el tiempo de permitir comparaciones.

No volátil: los AD existen para ser leídos, pero no modificados. La información es, por tanto, permanente, significando la actualización como la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía.

Ventajas de los AD (Sinnexus, 2016)

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

Desventajas de los AD (Bernabeu, 2010)

- ❖ Requiere una gran inversión, debido a que su correcta construcción no es tarea sencilla y consume muchos recursos, además, su implementación implica desde la adquisición de herramientas de consulta y análisis, hasta la capacitación de los usuarios.
- ❖ Existe resistencia al cambio por parte de los usuarios.

Capítulo 1: Fundamentación teórica de la investigación

- ❖ Los beneficios del AD son apreciados en el mediano y largo plazo. No todos los usuarios confiarán en el AD en una primera instancia, pero sí lo harán una vez que comprueben su efectividad y ventajas. Además, su correcta utilización surge de la propia experiencia.
- ❖ Incremento continuo de los requerimientos del usuario.

1.1.3 Mercado de Datos

Un MD es un elemento sustantivo dentro de la última generación de bases de datos. Proporciona un esquema de almacenamiento que permite realizar consultas complejas, por medio de la interacción e interrelación de múltiples entidades, manteniendo solo un repositorio con la información agregada y ordenada, según las necesidades de los usuarios. Además, provee de una interfaz de consulta que permite al usuario la posibilidad de análisis de la información para la toma de decisiones. (Chinchilla, 2011)

Un MD es un AD departamental, especializado en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. (Sinnexus, 2016)

Los MD son un subconjunto de los datos del AD cuyo objetivo es responder a un determinado análisis, función o necesidad, con una población de usuarios específica. Pueden ser dependientes o independientes de un AD. (Díaz, 2010)

Los autores de la investigación lo definen como un AD para un área o departamento específico de una organización, permitiendo al usuario la posibilidad de análisis de la información para la toma de decisiones. Ofrece una mayor personalización de los datos permitiendo un manejo más eficiente de la información histórica, ejecución de procesamiento independiente del resto de los departamentos y un costo de almacenamiento y procesamiento inferior.

1.2 Elementos de un Almacén de Datos

Un **Esquema** es una colección de tablas de dimensiones y tablas de hecho. Este término se refiere a la representación gráfica de una estructura de base de datos (Diseño físico). Se encuentra dentro de una base de datos, que a su vez está dentro de un servidor. Estas entidades se acomodan como cajas anidadas. El servidor es la caja más externa y el esquema la más interna. (Díaz, 2010)

Las **tablas de dimensiones** definen como están organizados lógicamente los datos y proveen el medio para analizar el contexto del negocio. Representan los ejes del cubo, y los aspectos de interés, mediante los cuales el usuario podrá filtrar y manipular la información almacenada en la tabla de hechos. Cada tabla

Capítulo 1: Fundamentación teórica de la investigación

posee un identificador único y al menos un atributo que describe los criterios de análisis relevantes de la organización. (Bernabeu, 2010)

Las **tablas de hechos** son la representación en el AD de los procesos de negocio de la organización. Contienen los hechos, medidas o indicadores que serán utilizados por los analistas de negocio para apoyar el proceso de toma de decisiones. Los datos presentes en ellas constituyen el volumen de la bodega, y pueden estar compuestos por millones de registros dependiendo de su granularidad y de los intervalos de tiempo de los mismos. Estas tablas poseen una clave primaria que está compuesta por las claves primarias de las tablas de dimensiones relacionadas a este. (Díaz, 2010)

Las **medidas** o atributos de hecho representan la información relevante sobre el hecho. Estas medidas son de tipo calculable, su función está en dependencia de lo que se quiera obtener. (Rojas, 2009)

1.3 Base de datos multidimensional

Las BD multidimensionales, proveen una estructura que permite tener acceso flexible a los datos, para explorar y analizar sus relaciones, y resultados consiguientes. Estas se pueden visualizar como un cubo multidimensional, donde las variables asociadas existen a lo largo de varios ejes o dimensiones, y la intersección de las mismas representan la medida, indicador o el hecho que se está evaluando. Para la implementación de las mismas, según (Bernabeu, 2010) existen variantes de modelamiento y de implementación que a continuación se detallan

1.3.1 Variantes de modelamiento

Las bases de datos multidimensionales implican tres variantes posibles de modelamiento, que permiten realizar consultas de soporte a la toma de decisión. Según (Bernabeu, 2010) estas son:

- ❖ **Esquema en estrella (*Star Scheme*):** 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. En la figura 1 se puede apreciar un esquema en estrella estándar.

Capítulo 1: Fundamentación teórica de la investigación

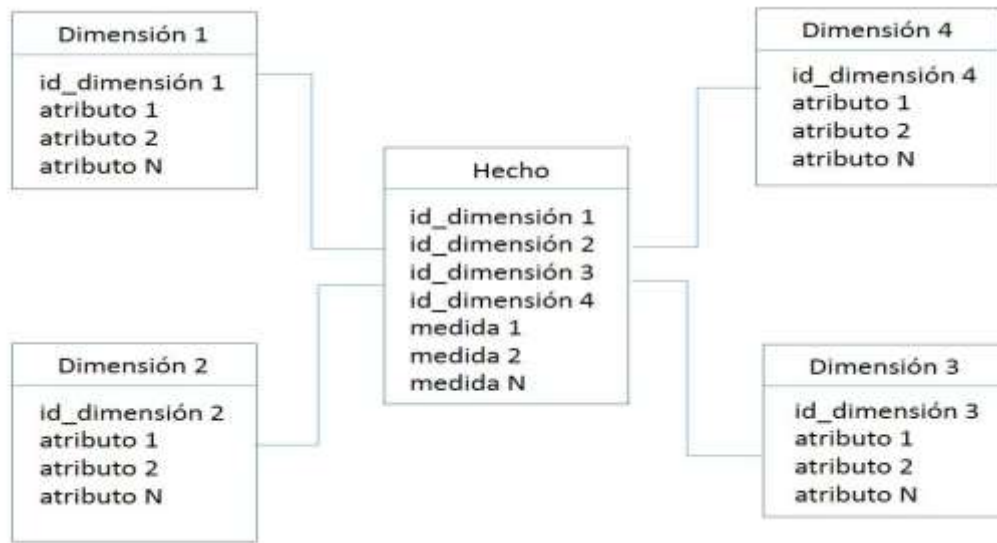


Fig.1: Representación del Esquema en estrella (Elaboración propia)

- ❖ **Esquema copo de nieve (Snowflake Scheme):** este esquema representa una extensión del modelo en estrella cuando las dimensiones se organizan en jerarquías de dimensiones. Consta de 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. En la figura 2 se puede apreciar un esquema copo de nieve estándar.

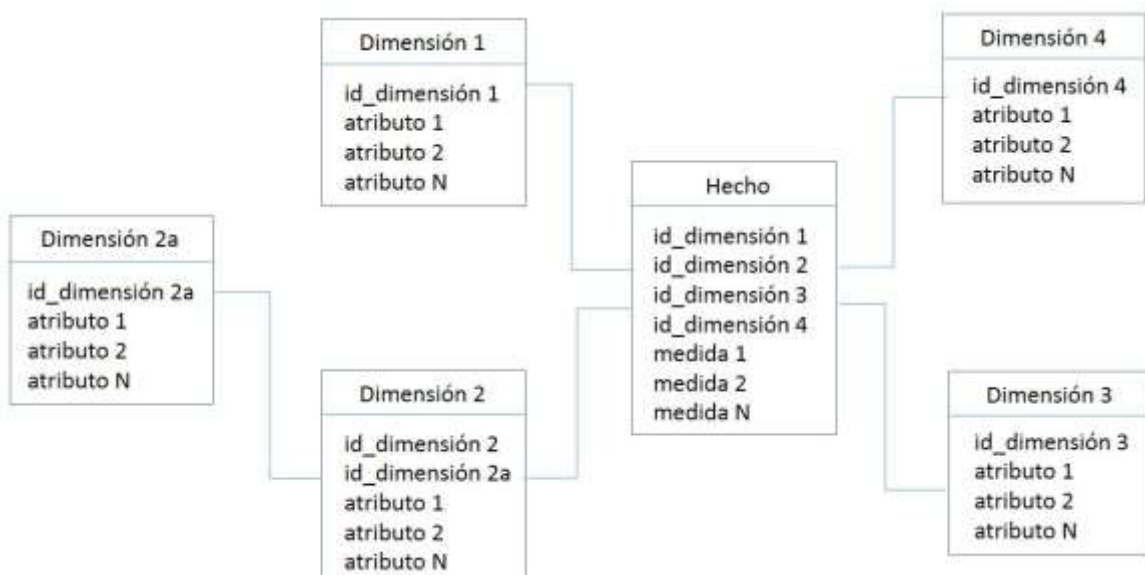


Fig.2: Representación del Esquema copo de nieve (Elaboración propia)

Capítulo 1: Fundamentación teórica de la investigación

- ❖ **Esquema constelación o copo de estrellas (Starflake Scheme):** un esquema de constelación está compuesto por una serie de esquemas en estrella. Consta de varias tablas de hechos que comparten dimensiones iguales. En la figura 3 se puede apreciar un esquema de constelación o copo de estrellas estándar.

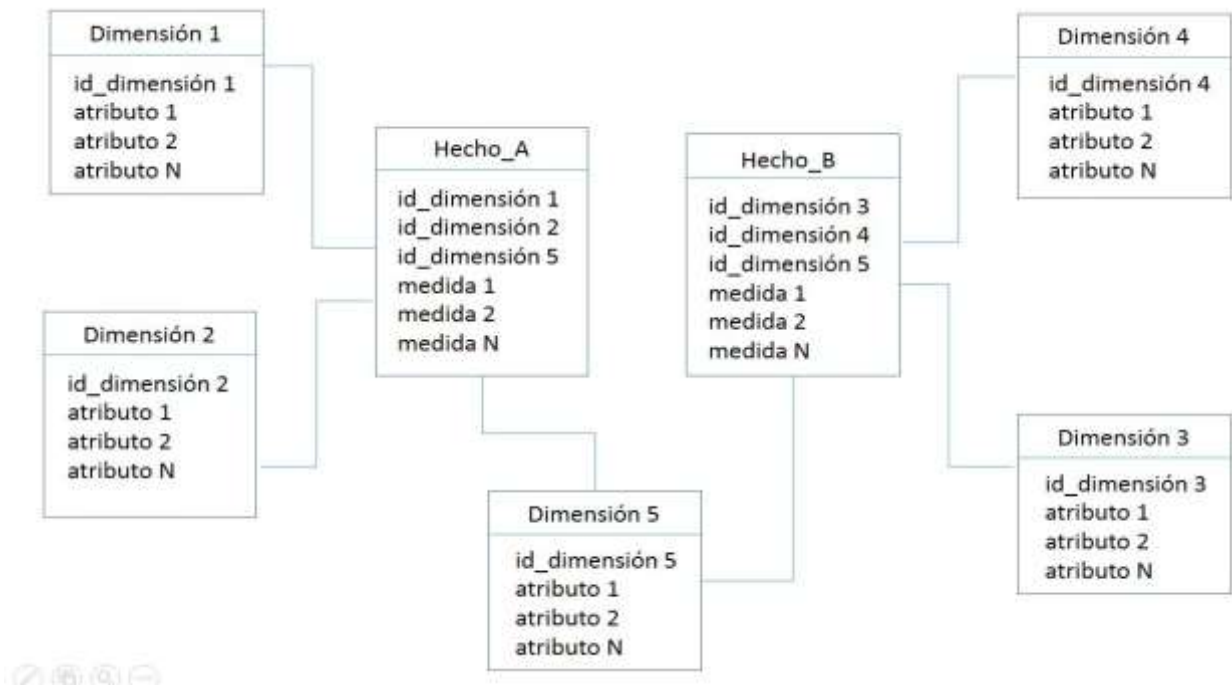


Fig.3: Representación del Esquema constelación o copo de estrellas (Elaboración propia)

1.3.2 Variantes de implementación

Los AD almacenan información para su posterior búsqueda y análisis, es por ello que se requiere de la tecnología OLAP por sus siglas en inglés (*On-Line Analytical Processing*) para implementar los esquemas antes mencionados.

OLAP incluye la incorporación de grandes cantidades de datos dispersos y puede abarcar millones de datos con relaciones complejas. Su objetivo es contribuir al análisis de estas relaciones mediante la facilitación de la búsqueda de patrones, tendencias y excepciones. OLAP logra proveer una visión multidimensional de los datos por medio de la consulta interactiva y el análisis de los mismos. Para esto, utiliza diferentes modos de implementación, ROLAP, MOLAP y HOLAP. Estos términos se refieren a las formas de almacenamiento de los datos. La decisión se debe tomar dependiendo de si se quiere que el almacenamiento refleje los análisis que se harán (multidimensional), o los datos se almacenarán sin tomar en cuenta para qué serán utilizados (de forma relacional). (Chinchilla, 2011). A continuación, se detallan las características de cada una de estas formas de almacenamiento según (Bernabeu, 2010)

Capítulo 1: Fundamentación teórica de la investigación

ROLAP (*Relational On Line Analytic Processing*): este tipo de organización física se implementa sobre tecnología relacional, pero disponen de algunas facilidades para mejorar el rendimiento. Es decir, ROLAP cuenta con todos los beneficios de una Sistema Gestor de Base de Datos (SGBD) relacional a los cuales se les provee extensiones y herramientas para poder utilizarlo como un Sistema Gestor de AD. Entre las características más importantes y sobresalientes de ROLAP, se encuentran las siguientes:

- ❖ Almacena la información en una BD relacional.
- ❖ Posee tres capas lógicas: de almacenamiento, de análisis y de presentación.
- ❖ Posee técnicas de particionamiento de datos.
- ❖ Posee optimizadores de consultas.
- ❖ Cuenta con extensiones del SQL, del inglés (*Structured Query Language*).

MOLAP (*Multidimensional On Line Analytic Processing*): el objetivo de los sistemas MOLAP es almacenar físicamente los datos en estructuras multidimensionales, de manera que la representación externa y la interna coincidan. Para ello, se dispone de estructuras de almacenamiento específicas (*Arrays*) y técnicas de compactación de datos que favorecen el rendimiento del depósito de datos. Las principales características de MOLAP son:

- ❖ Posee tecnología optimizada para consultas y análisis, basada en el modelo multidimensional.
- ❖ Cuenta con un motor especializado.
- ❖ Provee herramientas limitadas y propietarias.
- ❖ No es adecuada para muchas dimensiones.

HOLAP (*Hybrid On Line Analytic Processing*): constituye un sistema híbrido entre MOLAP y ROLAP, que combina estas dos implementaciones para almacenar algunos datos en un motor relacional y otros en una BD multidimensional.

1.4 Metodologías para el desarrollo de AD

Para la correcta implementación de una solución informática es aconsejable seguir una metodología que pueda guiar el proceso paulatinamente. Muchos son los autores que han diseñado o propuesto una metodología para el diseño y construcción de AD, las más conocidas son la de Bill Inmon, la de Hefesto y la de Kimball.

Capítulo 1: Fundamentación teórica de la investigación

Inmon se basa en los conceptos de diseño de bases de datos relacionales, utilizando un enfoque descendente (*Top-Down*), donde se implementa el AD y luego los MD que se desarrollen para cada área de la empresa. (Inmon, 2005)

Sin embargo, Kimball define una metodología utilizando un enfoque ascendente (*Bottom-Up*), en el que se implementan los MD para luego integrarlos en el AD. Este enfoque se centra en el modelo dimensional no normalizado. (Kimball, 2006)

La metodología Hefesto, por otra parte, define cuatro pasos fundamentales: análisis de requerimientos, análisis de los OLTP (*On Line Transaction Processing*), modelo lógico del AD e integración de datos los que pueden adaptarse muy bien a cualquier ciclo de vida de desarrollo de software. (Bernabeu, 2010)

Metodología a utilizar en la solución

Para el desarrollo de la presente investigación se decidió no utilizar ninguna de las metodologías antes propuestas, al considerar que estas se basan en el uso de herramientas propietarias y utilizan una terminología no estandarizadas para la identificación de requisitos de los usuarios, lo que hace difícil su utilización. Los temas relacionados con la gestión de proyectos se tratan de forma superficial. La mayoría de las tareas y fases que presentan se basan únicamente en el desarrollo del AD y no en el desarrollo integral del proyecto. Ninguna de las metodologías presentadas logra una alineación clara con algún estándar de calidad, esta función se le deja totalmente a la organización.

Por tanto, se decidió utilizar la “Metodología de Desarrollo de Proyectos de Almacenes de Datos”, definida por la MCs. Yanisbel González Hernández en su tesis de maestría. La misma es una combinación de las mencionadas anteriormente. Toma como base la metodología de Kimball para definir los aspectos específicos del desarrollo de AD. Para incorporar los principios básicos que permiten una adecuada gestión del proyecto, utiliza la Guía para los Fundamentos de la Dirección de Proyectos (Guía del PMBOK). Los temas asociados a CMMI (*Capability maturity model integration*) se incorporan a partir del Programa de Mejora, por lo tanto, hereda algunos de sus enfoques, artefactos y actividades. Se adapta a las características de trabajo específicas del centro DATEC, por lo que se escoge para todo desarrollo de AD en dicho centro. (González, 2013)

Esta metodología de desarrollo está dividida en siete fases y un flujo de trabajo. Algunas de estas fases podrán ser implementadas de forma paralela como es el caso de la fase de requisitos y arquitectura, además durante la fase de diseño e implementación, podrán desarrollarse varios componentes al mismo tiempo, lo que permite un desarrollo más ágil. Para el desarrollo de la presente investigación solo se va a tomar en cuenta las cinco primeras fases y el flujo de trabajo gestión de proyecto para guiar la solución de forma

Capítulo 1: Fundamentación teórica de la investigación

paulatina. A continuación, la figura 4 muestra las fases de la metodología con su flujo de trabajo, y posteriormente las definiciones que de ellas ofrece (González, 2013).

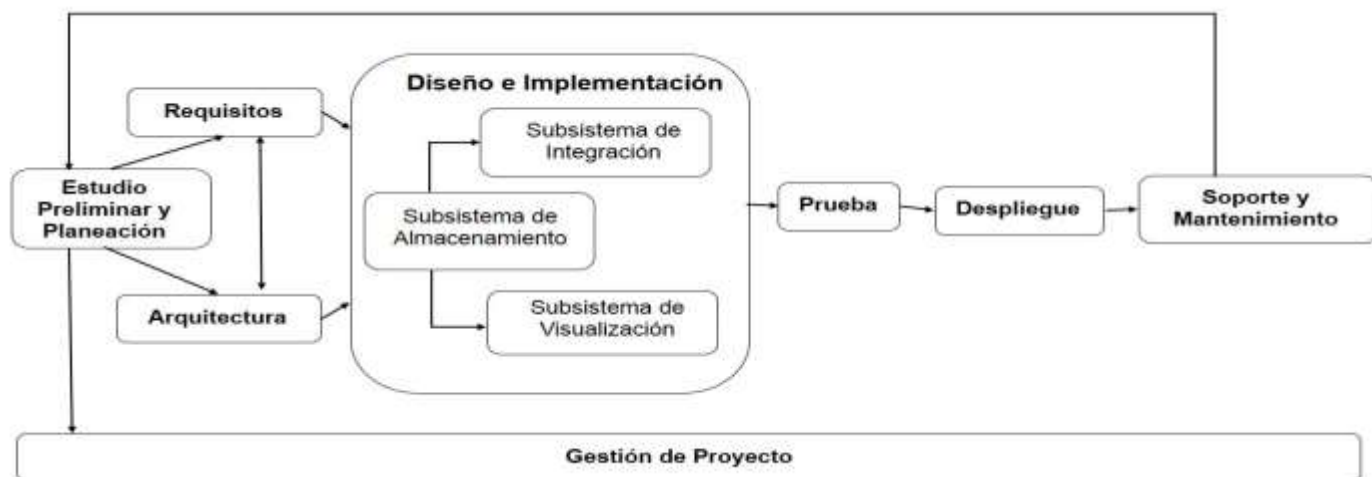


Fig.4: Metodología de desarrollo (Elaboración Propia)

- ❖ **Estudio preliminar y planeación:** se realiza un estudio minucioso en la entidad cliente. Esto incluye un diagnóstico integral de la organización, con el fin de determinar qué es lo que se desea construir y qué condiciones existen para el desarrollo y montaje de la misma. Además, se llevan a cabo las tareas de planeación del proyecto.
- ❖ **Requisitos:** se realiza el proceso entrevistas al cliente para determinar los requisitos de información. Se hace levantamiento detallado de las fuentes de datos para validar la disponibilidad de la información. Además, se definen los requisitos funcionales y no funcionales de la solución y se hace el análisis de los requisitos que dan paso al diseño e implementación.
- ❖ **Arquitectura:** se definen las vistas arquitectónicas de la solución, aspectos como, los subsistemas y componentes, la seguridad, la comunicación y la tecnología a utilizar.
- ❖ **Diseño e Implementación:** se define el diseño de las estructuras de almacenamiento de datos, se diseñan los procesos de integración de datos como, el mapa lógico de datos, los cubos OLAP para la presentación de la información, así como el diseño gráfico de la aplicación definido por el cliente. Después se implementan cada uno de los subsistemas (repositorio de datos, integración de datos, presentación de datos).
- ❖ **Prueba:** se realizan las pruebas que validan la calidad del producto, comenzando por las pruebas de unidad, seguidas las pruebas de integración y sistema, hasta llegar a las pruebas de aceptación

Capítulo 1: Fundamentación teórica de la investigación

con el cliente final. Esta fase no es la única en la que se realizan pruebas durante el desarrollo del proyecto, en todas las fases hay actividades de aseguramiento de la calidad.

- ❖ **Despliegue:** consta de dos etapas, despliegue piloto, donde se configuran los servidores necesarios y se instalan las herramientas según la arquitectura definida, se cargan una muestra de los datos en un ambiente controlado, con el fin de mostrarle al cliente final el sistema en funcionamiento. Una vez aceptada la solución por el cliente, se realiza la carga histórica de los datos, puede ser en el mismo entorno que el despliegue piloto u otro, todo depende de las condiciones que establezca el cliente. Además, se realiza la capacitación y transferencia tecnológica de la solución a los clientes. El resultado fundamental es la solución desplegada en el entorno real y en correcto funcionamiento.
- ❖ **Soporte y Mantenimiento:** comienza cuando la solución está implantada y en explotación, y se ejecuta según el contrato firmado y las condiciones de soporte establecidas. Puede realizarse a través de variados servicios, que pueden ser soporte en línea, vía telefónica, web, correo y el acompañamiento al cliente. Además, se realizan las tareas de manteniendo de la aplicación tan necesarias para este tipo de desarrollo y que garantizan el adecuado funcionamiento y crecimiento del almacén de datos.
- ❖ **Gestión del proyecto:** este flujo de trabajo se ejecuta a lo largo de todo el ciclo de vida del proyecto. Es aquí donde se controla, gestiona y chequea todo el desarrollo, los gastos, las utilidades, los recursos, las adquisiciones, los planes y cronogramas, entre otras actividades relacionadas con la gestión de proyectos. Esta fase es la columna vertebral del proyecto y si no se ejecuta de forma continua y correcta el proyecto puede fracasar.

1.5 Herramientas a utilizar en el desarrollo de la solución

Para la creación de la solución propuesta, es necesario la selección de herramientas que permitan su correcto desarrollo y funcionamiento. Hoy día, muchas son las empresas que se han dado a la tarea de crear y revitalizar herramientas que brindan una mejor compenetración con temas de base de datos, entre ellas, se pueden mencionar a *Microsoft* y *Oracle Corporation*, las que de una forma u otra se han apoderado del mercado y del mundo de las bases de datos. Desde el punto de vista económico, la utilización de las mismas, en empresas cubanas, tendrían un impacto negativo, dado los inaccesibles precios de sus licencias privativas. La tendencia hoy en el mercado del software indica una interesante propuesta, que convida a inclinarse por herramientas de software libre, que independientemente de los resultados que ya están ofreciendo, se libra de cuantiosas sumas de capital y da la posibilidad de desarrollar herramientas propias

Capítulo 1: Fundamentación teórica de la investigación

con interfaces amigables al usuario. A continuación, se detallan cada una de las herramientas a utilizar teniendo en cuenta que fueron seleccionadas por lo antes mencionado.

❖ Herramienta para el modelado de los datos

Visual Paradigm for UML Community Edition 8.0: es una herramienta para desarrollo de aplicaciones utilizando modelado UML (Lenguaje Unificado de Modelado) ideal para el desarrollo orientado a objetos. Visual Paradigm ofrece navegación intuitiva entre la escritura del código y su visualización. Potente generador de informes en formato PDF/HTML. Ambiente visualmente superior de modelado. Sofisticado diagramador automático de *layout* y sincronizador de código fuente en tiempo real. (Targetware, 2016)

❖ Sistema Gestor de Base de Datos

PostgreSQL 9.4: es un sistema de gestión de base de datos relacional orientado a objeto, distribuido bajo licencia BSD (*Berkeley Software Distribution*) y su código fuente está disponible libremente. Es el sistema de gestión de base de datos de código abierto más potente del mercado, sus últimas versiones son tan poderosas como las de los sistemas de gestión de base de datos comerciales. Utiliza un modelo cliente/servidor y usa multiprocesos en vez de multihilos para garantizar la estabilidad del sistema. Un fallo en uno de los procesos no afectará el resto y el sistema continuará funcionando. (Martínez, 2010)

❖ Administrador de Base de Datos

PgAdmin III 1.20: es una aplicación de diseño y gestión de base de datos *PostgreSQL*. Está escrito en C++, multiplataforma lo que permite que se ejecute en múltiples plataformas. La aplicación se puede utilizar para manejar *PostgreSQL* 7.3 y superiores y funciona sobre casi todas las plataformas. Este software fue diseñado para responder a las necesidades de todos los usuarios, desde la escritura de simples consultas SQL a la elaboración de base de datos complejas. (PostgreSQL, 2016)

❖ Herramienta para el perfilado de datos

Data Cleaner 3.1: esta herramienta permite descubrir y analizar la calidad de sus datos. Encuentra los patrones, valores faltantes, conjuntos de caracteres y otras características de los datos. Realiza comparaciones de tablas y columnas, lo que permite verificar la consistencia y veracidad de los datos. Esta herramienta permite analizar la calidad de la información que se encuentra almacenada en fuentes como: base de datos *Oracle*, *MySQL*, *PostgreSQL* y ficheros excel, XML. (Sorensen, 2012)

❖ Herramientas para la Extracción, Transformación y Carga (ETL)

Pentaho Data Integration (PDI) 6.0: herramienta de código abierto y multiplataforma, prepara y combina datos para crear una imagen completa del negocio y generar información útil. La plataforma proporciona

Capítulo 1: Fundamentación teórica de la investigación

datos precisos y preparados para el análisis de los usuarios finales desde cualquier fuente, permite la lectura y escritura de base de datos, el mapeo de valores y la normalización de los datos en distintas filas y viceversa con herramientas visuales para eliminar la codificación y la complejidad. (Pentaho, 2016)

❖ Herramientas de Inteligencia de Negocio (BI)

Las herramientas de BI son aplicaciones digitales diseñadas para colaborar con la inteligencia de negocio durante el análisis y la presentación de datos. Permite a las compañías contar con la información adecuada para una mejor toma de decisiones. Las compañías que implementan el BI logran sacar mayor provecho de las situaciones de crisis gracias a la posibilidad de contar con un análisis de mercado más acertado debido a que los datos pesados son transformados en importantes estrategias corporativas.

Pentaho Schema Workbench (PSW) 3.11.0.0: es una herramienta de desarrollo multiplataforma que permite crear, modificar y publicar un esquema de *Mondrian*³. Permite crear todos los objetos tales como, esquemas, cubos, dimensiones y métricas. Tiene dos áreas: la zona donde se encuentra la estructura jerárquica del esquema OLTP y la zona de edición de las propiedades de cada elemento. Presenta un menú superior para crear cubos, dimensiones, métricas, miembros calculados subconjuntos y roles, así como operaciones estándar como cortar, copiar y pegar. (Díaz, 2010)

Pentaho Report Designer (PRD) 6.0: es una herramienta de diseño desktop, que permite definir y construir informes, y luego publicarlos en el portal de BI para que puedan ser ejecutados por los usuarios. Trabaja con múltiples orígenes de datos (JDBC, Olap4J, Pentaho Análisis, Pentaho Data Integration, XML). PRD es un generador de informe que se divide en secciones o grupos de datos en los que los elementos del informe pueden ser posicionados. El resultado de los informes que se vayan diseñando se puede ver con las opciones de previsualización, y permite la salida de resultados en diferentes formatos como PDF, HTML, XLS, RTF y CSV. (Espinosa, 2010)

Pentaho BI Server 6.0: provee de servicios críticos incluyendo programación, seguridad, integración automatización y flujo de trabajo, proporcionando habilidades a los usuarios finales de Pentaho y proveniente un lugar central para administrar y mantener el despliegue de la empresa BI. Posee integración con procesos de negocio, administra y programa reportes, administra seguridad de usuarios asegurando la escalabilidad, integración y portabilidad. Puede trabajar sobre las BD: vía JDBC, IBM DB2, *Microsoft SQL Server*, *MySQL*, *Oracle*, *PostgresSQL* sin importar el sistema operativo. Repositorio de datos basado en XML. Todos los componentes están expuestos vía *Web Services* para facilitar la integración con Arquitecturas Orientadas a Servicios. (Gravitar, 2016)

³ Es un servidor/motor OLAP escrito en java que está licenciado bajo la Eclipse Public License y fue adquirido por Pentaho en 2005.

Capítulo 1: Fundamentación teórica de la investigación

1.6 Conclusiones del capítulo

La sistematización de los referentes teóricos que sustentan las diferentes tecnologías de almacenamiento, permitieron profundizar en el tema a investigar, lo que contribuyó a una mejor comprensión de los MD, recurso idóneo para el desarrollo de la propuesta de solución, debido a que la información que maneja responde a las necesidades de un departamento en específico. Para el desarrollo de la temática abordada se seleccionó la “Metodología de Desarrollo de Proyectos de Almacenes de Datos” empleada en el centro DATEC, de este modo se garantiza el correcto desarrollo del Mercado de Datos para el Departamento de Mercadotecnia, teniendo en cuenta cada una de las fases y flujo de trabajo que propone la misma. Por último, se realizó un estudio de las herramientas a utilizar teniendo en cuenta que las seleccionadas son todas de código abierto y sus características permiten el correcto modelamiento e implementación del Mercado de Datos para el Departamento de Mercadotecnia. De ahí que, se seleccionó, para el modelado de datos el Visual Paradigm for UML 8.0, como SGBD el PostgreSQL 9.4, como administrador de BD el PgAdmin III 1.20, para la limpieza de datos el Data Cleaner 3.1, se utilizó el PDI 6.0 para la extracción, transformación y carga de los datos, el PSW 3.11 para la creación de los cubos multidimensionales, para la creación de los reportes se utilizó el PRD 6.0 y como plataforma de visualización el Pentaho BI Server 6.0.

Capítulo 2: Análisis y diseño de la solución

Capítulo 2: Análisis y diseño de la solución

En este capítulo se realizará un estudio más detallado del negocio comprendiendo así las necesidades del cliente. Se definirán los requisitos de información, los requisitos funcionales y los requisitos no funcionales, además de las reglas del negocio de manera que cumpla con las especificaciones del cliente. Luego se realizará el diagrama de caso de uso del sistema con las especificaciones correspondientes a cada caso de uso. Se definirá la arquitectura a utilizar para el desarrollo del Mercado de Datos para el Departamento de Mercadotecnia definiéndose el diseño correspondiente a cada subsistema identificado, además de establecerse políticas de respaldo y recuperación en caso de ocurrencia de fallos.

2.1 Estudio preliminar del negocio

El Departamento de Mercadotecnia perteneciente a la DTTC se encarga de la gestión de la comercialización y la negociación de los servicios académicos, editoriales e informáticos, así como de las aplicaciones que son desarrolladas por los centros de producción de la universidad, de manera que permita a los directivos de dicha dirección, tener una visión objetiva de los deseos de sus clientes para diseñar productos y servicios atendiendo a dicho mercado. Para esto, los especialistas realizan una serie de actividades que se basan en el almacenamiento de la información de cada proceso que allí se realiza, para posteriormente, conformar un informe que se emite mensualmente. Dicho informe recoge información estadística de cada uno de los procesos, manteniendo el control sobre los mismos, de manera que apoye el proceso de toma de decisiones en la dirección. Estos procesos son:

- ❖ **Registro en el PTTC:** donde se almacena información referente a los contratos que se llevan a cabo en la institución, de estos se deben tener un control del estado legal en que se encuentran, los ingresos que genera cada uno de ellos y el centro que desarrolló el producto.
- ❖ **Registro de expedientes comerciales:** por cada producto comercializable se crea un expediente para ser mostrado a los clientes como proceso de marketing. Por cada expediente se debe conocer el tipo de producto, el centro que lo desarrolló, la marca, si es nuevo en el catálogo, si constituye una nueva versión y si tiene alcance en el mercado.
- ❖ **Registro de visitas al salón de exposiciones:** aquí se archivan todas las visitas ya sea de empresas, exposiciones de productos UCI u otras actividades realizadas en el salón; de estos se debe conocer el tipo de visitante, el tipo de organización, si la visita lleva presencia de asesores, si se planificó, se realizó y el interés.
- ❖ **Registro de productos en el CENDA:** aquí es donde a cada realizador de un producto u obra literaria se le otorga el derecho de ser reconocido legalmente como el propietario del producto, lo

Capítulo 2: Análisis y diseño de la solución

que se conoce como derecho de autor. De cada obra que se propone para ser registrada en el CENDA se debe conocer el centro y el estado legal en que se encuentra.

Toda la información con la que cuenta el departamento proviene de disímiles ficheros .xls que se encuentran en las PC de los especialistas, los que no son manejados por ningún sistema informático. Para la confección de los informes mensuales, los especialistas analizan y copian los datos que se encuentran en los ficheros excel para un documento word. Realizan, además, tablas resumen, comparaciones entre los indicadores y generan sus gráficos correspondientes, de manera que la información mostrada sirva de análisis para los directivos del departamento.

2.2 Necesidades de usuario

Para construir la base del Mercado de Datos para el Departamento de Mercadotecnia fue necesario identificar las necesidades de información de cada especialista para poder apoyar al proceso de toma de decisión en la dirección. Para esto se separó la información en áreas de análisis según la temática abordada, cada una de ellas, representa un proceso en específico ya que los cuatro son totalmente diferentes. A continuación, se muestran las áreas de análisis y su proceso correspondiente. Ver tabla 1.

Tabla.1: Distribución de las áreas de análisis

Áreas de análisis	Procesos
PTTC	Registro de contratos en el PTTC.
Visitas al Salón de Exposiciones	Registro de Visitas al Salón de Exposiciones.
Expedientes Comerciales	Registro de Expedientes Comerciales.
Productos en el CENDA	Registro de Productos en el CENDA.

2.3 Requisitos de información

Los Requisitos de información (RI) describen la información y los datos que el sistema debe proveer o debe acceder. Estos se definen a partir de las necesidades de información identificadas en el negocio, que permitan el análisis del comportamiento de los indicadores a medir según los objetivos y metas de la organización. (Schiefer, y otros, 2002). A continuación, se muestran los RI identificados por áreas de análisis.

Capítulo 2: Análisis y diseño de la solución

RI de Registro en el PTTC

- ❖ **RI-1** Obtener listado de contratos por centro, estado legal, encargado y tiempo dado el centro, el tiempo (año y mes), el estado legal y el encargado.
- ❖ **RI-2** Obtener el monto total en CUC por centro, estado legal, encargado y tiempo dado el centro, el tiempo (año y mes), el estado legal y el encargado.
- ❖ **RI-3** Obtener el monto total en CUP por centro, estado legal, encargado y tiempo dado el centro, el tiempo (año y mes), el estado legal y el encargado.
- ❖ **RI-4** Obtener el monto total en MT (moneda total) por centro, estado legal, encargado y tiempo dado el centro, el tiempo (año y mes), el estado legal y el encargado.
- ❖ **RI-5** Obtener la cantidad de contratos por centro, estado legal, encargado y tiempo dado el centro, el tiempo (año y mes), el estado legal y el encargado.
- ❖ **RI-6** Obtener el por ciento de contratos por centro, estado legal, encargado y tiempo dado el centro, el tiempo (año y mes), el estado legal y el encargado.

RI de Registro de Visitas al Salón de Exposiciones

- ❖ **RI-7** Obtener listado de visitantes por tipo de organización, tipo de visitante y tiempo dado el tipo de organización, el tipo de visitante, el tiempo (año y mes), el interés, la planificación, la realización y el uso de asesores.
- ❖ **RI-8** Obtener la cantidad de visitantes por tipo de organización, tipo de visitante y tiempo dado el tipo de organización, el tipo de visitante, el tiempo (año y mes), el interés, la planificación, la realización y el uso de asesores.
- ❖ **RI-9** Obtener la cantidad de participantes por tipo de organización, tipo de visitante y tiempo dado el tipo de organización, el tipo de visitante, el tiempo (año y mes), el interés, la planificación, la realización y el uso de asesores.
- ❖ **RI-10** Obtener el por ciento de visitantes por tipo de organización, tipo de visitante y tiempo dado el tipo de organización, el tipo de visitante, el tiempo (año y mes), el interés, la planificación, la realización y el uso de asesores.

RI de Registro de Expedientes Comerciales

- ❖ **RI-11** Obtener listado de expedientes comerciales por centro, marca, tipo de producto, su descripción y tiempo dado el centro, el tiempo (año), el alcance en el mercado, el tipo de producto, la marca sombrilla y novedad en el catálogo.

Capítulo 2: Análisis y diseño de la solución

- ❖ **RI-12** Obtener la cantidad de expedientes comerciales por centro, marca, tipo de producto, su descripción y tiempo dado el centro, el tiempo (año), el alcance en el mercado, el tipo de producto, la marca sombrilla y novedad en el catálogo.
- ❖ **RI-13** Obtener el por ciento de expedientes comerciales por centro, marca, tipo de producto, su descripción y tiempo dado el centro, el tiempo (año), el alcance en el mercado, el tipo de producto, la marca sombrilla y novedad en el catálogo.

RI de Registro de Productos en el CENDA

- ❖ **RI-14** Obtener listado de obras literarias por su tipo, centro, estado CENDA y tiempo dado el centro, el tiempo (año y mes), el estado CENDA y el tipo de obra literaria.
- ❖ **RI-15** Obtener la cantidad de obras literarias por su tipo, centro, estado CENDA y tiempo dado el centro, el tiempo (año y mes), el estado CENDA y el tipo de obra literaria.
- ❖ **RI-16** Obtener el por ciento de obras literarias por su tipo, centro, estado CENDA y tiempo dado el centro, el tiempo (año y mes), el estado CENDA y el tipo de obra literaria.

2.4 Requisitos funcionales

Para que los sistemas de software cumplan con las funcionalidades que el cliente necesita, hay que identificar los requisitos funcionales (RF) que representan las condiciones o funcionalidades con los que el sistema debe cumplir. (Kamber, y otros, 2006). A continuación, se enumeran los RF y sus descripciones se pueden encontrar en el artefacto “Especificación de requisitos de software”.

- ❖ **RF-1** Extraer datos: extrae los datos de los ficheros fuentes (.xls).
- ❖ **RF-2** Realizar transformación y carga: transforma los datos extraídos para poder cargarlos en la BD.

2.5 Requisitos no funcionales

Los requisitos no funcionales (RNF) representan características generales y restricciones de la aplicación o sistema que se esté desarrollando. Suelen presentar dificultades en su definición dado que su conformidad o no conformidad, podría ser sujeto de libre interpretación, por lo que es recomendable acompañar su definición con criterios de aceptación que se puedan medir como atractivo, usable, rápido o confiable. (PMOinformatica, 2015). A continuación, se detallan los RNF y sus descripciones.

Usabilidad

- ❖ **RNF-1** Facilitar la búsqueda y el acceso a los reportes, vistas de análisis y cuadros de mando del MD, mediante la distribución de la información por áreas de análisis. Se diseñará un mapa de

Capítulo 2: Análisis y diseño de la solución

navegación con el objetivo que el usuario pueda acceder de manera rápida a la información que solicita en el área correspondiente de acuerdo al objetivo de su solicitud.

Disponibilidad

- ❖ **RNF-2** Garantizar el cumplimiento de actualización de los datos en el MD. La información contenida en el MD tendrá un rigor y precisión mensual, ya que se realizará la carga de los datos con esta periodicidad a través de tareas programadas en el sistema operativo.

Seguridad

- ❖ **RNF-3** Garantizar la persistencia de la información. Se debe realizar un respaldo mensual al MD, asegurando que la información esté disponible para el usuario siempre que la necesite.
- ❖ **RNF-4** Garantizar que solo tengan acceso a la información del sistema los especialistas autorizados mediante la asignación de roles.

Restricciones de diseño

- ❖ **RNF-5** Lograr la homogeneidad de la estructura de los elementos definidos en el MD. La estructura del MD debe tener un nombre estándar teniendo en cuenta el tipo de estructura que sea: **dimension.dim_[nombre]** para las dimensiones y **mart_mercadotecnia.hech_[nombre]** para los hechos.
- ❖ **RNF-6** Utilizar los lenguajes de programación definidos durante la investigación. Para la realización del MD se utilizó como lenguaje de programación SQL y MDX.

Interfaz

- ❖ **RNF-7** Acceso al sistema. El usuario deberá acceder a la aplicación mediante el protocolo HTTP, usando el navegador web Firefox entre sus versiones 42.0 y 53.0.2.

Interfaces de usuario

- ❖ **RNF-8** Mostrar todos los textos que aparezcan en la interfaz del sistema en idioma español. Los mensajes de procesamiento y error, así como los nombres de las vistas y la arquitectura de información deben ser en idioma español y tener un estándar en todo el sistema.

Interfaces de hardware

- ❖ **RNF-9** El sistema interactuará con la impresora como interfaz de hardware. Esta interacción se ocasionará cuando se necesite imprimir un reporte en formato físico. El acceso a la impresora será mediante el protocolo TCP/IP a través de la interfaz que ofrece el hardware.
- ❖ **RNF-10** Proporcionar características mínimas de hardware a las estaciones de trabajo. Para lograr una explotación aceptable del sistema el hardware debe contar con dos GB de RAM.
- ❖ **RNF-11** Proporcionar características mínimas de hardware a los servidores. Para lograr una explotación aceptable del sistema, el servidor debe contar con cuatro GB de RAM y un TB de almacenamiento.

Capítulo 2: Análisis y diseño de la solución

Interfaces de software

- ❖ **RNF-12** Instalar en las estaciones de trabajo el software necesario para el correcto funcionamiento del sistema. Las configuraciones de software de las máquinas clientes deben contar con Firefox entres las versiones 42.0 y 53.0.2.
- ❖ **RNF-13** Instalar en el servidor el SGBD PostgreSQL 9.4, como administrador de BD PgAdmin III, de la Suite Pentaho 6.0 el PDI, el PRD, el PSW y el Pentaho BI server. Tener instalada la máquina virtual de Java versión 7.

2.6 Reglas del negocio

Una Regla de Negocio (RN) define o limita un aspecto del negocio con el objeto de establecer una estructura o un grado de influencia que condiciona el comportamiento de los actores del negocio. A menudo las RN están focalizadas en el control, en la forma de realizar los cálculos, otras permiten establecer las políticas, y así se tienen en cualquier actividad del negocio, que requiera que la gente actúe de una forma pre-establecida. (Saffirio, 2011)

Reglas de variables

- ❖ **RN-1** La cantidad de contratos va a ser igual a la sumatoria de todos los contratos archivados en el PTTC.
- ❖ **RN-2** El monto total en CUC va a ser igual a la sumatoria del monto en CUC para todos los contratos.
- ❖ **RN-3** El monto total en CUP va a ser igual a la sumatoria del monto en CUP para todos los contratos.
- ❖ **RN-4** El monto total en MT va a ser igual a la suma del monto total en CUC y el monto total en CUP para todos los contratos.
- ❖ **RN-5** El por ciento de los contratos va a ser igual a la $(\text{cantidad de contratos mostrados} * 100) / \text{total de contratos archivados}$.
- ❖ **RN-6** La cantidad de expedientes comerciales va a ser igual a la sumatoria de todos los expedientes comerciales registrados.
- ❖ **RN-7** El por ciento de los expedientes comerciales va a ser igual a la $(\text{cantidad de expedientes comerciales mostrados} * 100) / \text{total de expedientes comerciales}$.
- ❖ **RN-8** La cantidad de obras literarias va a ser igual a la sumatoria de todas las obras literarias registradas.
- ❖ **RN-9** El por ciento de las obras literarias va a ser igual a la $(\text{cantidad de obras literarias mostradas} * 100) / \text{total de obras literarias}$.
- ❖ **RN-10** La cantidad de visitantes va a ser igual a la sumatoria de todos los visitantes registrados.
- ❖ **RN-11** La cantidad de participantes va a ser igual a la sumatoria de todos los participantes registrados en cada una de las visitas.

Capítulo 2: Análisis y diseño de la solución

- ❖ **RN-12** El por ciento de los visitantes va a ser igual a la (cantidad de visitantes mostrados * 100) / total de visitantes.

Reglas de almacenamiento

- ❖ **RN-13** Los valores de tipo texto (String) tendrán una capacidad de almacenamiento de 50 caracteres.
- ❖ **RN-14** Los valores de tipo numérico (Integer) tendrán una capacidad de almacenamiento de cuatro caracteres.
- ❖ **RN-15** Los valores de tipo fecha (Data) tendrán una capacidad de almacenamiento de 10 caracteres.
- ❖ **RN-16** Los valores de tipo decimal (float) tendrán una capacidad de almacenamiento de ocho caracteres.

Reglas de transformación

- ❖ **RN-17** Los campos vacíos en las tablas significan que los indicadores no tienen valor.
- ❖ **RN-18** Cuando las columnas tienen campos vacíos se les debe especificar “SV”, lo que significa “Sin Valor”, en caso de ser String.
- ❖ **RN-19** Cuando las columnas tienen campos vacíos se les debe especificar “0” en caso de ser valores numéricos.
- ❖ **RN-20** Cuando el tipo de servicio es “no aplica” toma valor nulo.
- ❖ **RN-21** Cuando el tipo de servicio y el tipo de programa informático sean nulos toma valor “Programa informático” la columna tipo de programa informático.

Reglas de visualización

- ❖ **RN-22** El análisis estadístico es realizado por los jefes del Departamento de Mercadotecnia y los especialistas de cada proceso.
- ❖ **RN-23** Cada especialista tiene acceso solo a los reportes correspondiente a su área de análisis.
- ❖ **RN-24** El 100 % es representado en los reportes como 1.00.0 %.
- ❖ **RN-25** Los valores de tipo *float* se representarán con dos lugares después de la coma.
- ❖ **RN-26** Cuando los valores de tipo *float* pasan los tres dígitos estos son separados por un punto.

2.7 Diagrama de Casos de Uso del Sistema

El diagrama de Casos de Uso del Sistema (CUS) representa el conjunto de Casos de Uso (CU), actores y sus relaciones. En la presente investigación fueron identificados dos actores que representan a las personas que interactúan con el sistema, los que se describen en la tabla 2; además, fueron identificados siete casos de uso, de ellos cinco responden a los RI y dos a los RF conformando así el diagrama de CUS, que se muestra en la figura 5. En el mismo se ve reflejado el patrón Concordancia en Adición en el CU extendido “Realizar operaciones sobre reportes” ya que se utiliza para modelar una parte del CU que puede considerarse como un comportamiento opcional.

Capítulo 2: Análisis y diseño de la solución

Actores del sistema

Tabla.2: Descripción de los actores del sistema

Actor	Descripción
Administrador de ETL	Encargado de la realización de los procesos de extracción, transformación y carga de los datos.
Especialista	Encargado de mostrar la información referente a las áreas de análisis, la realización de los reportes y su visualización.

Diagrama de CUS

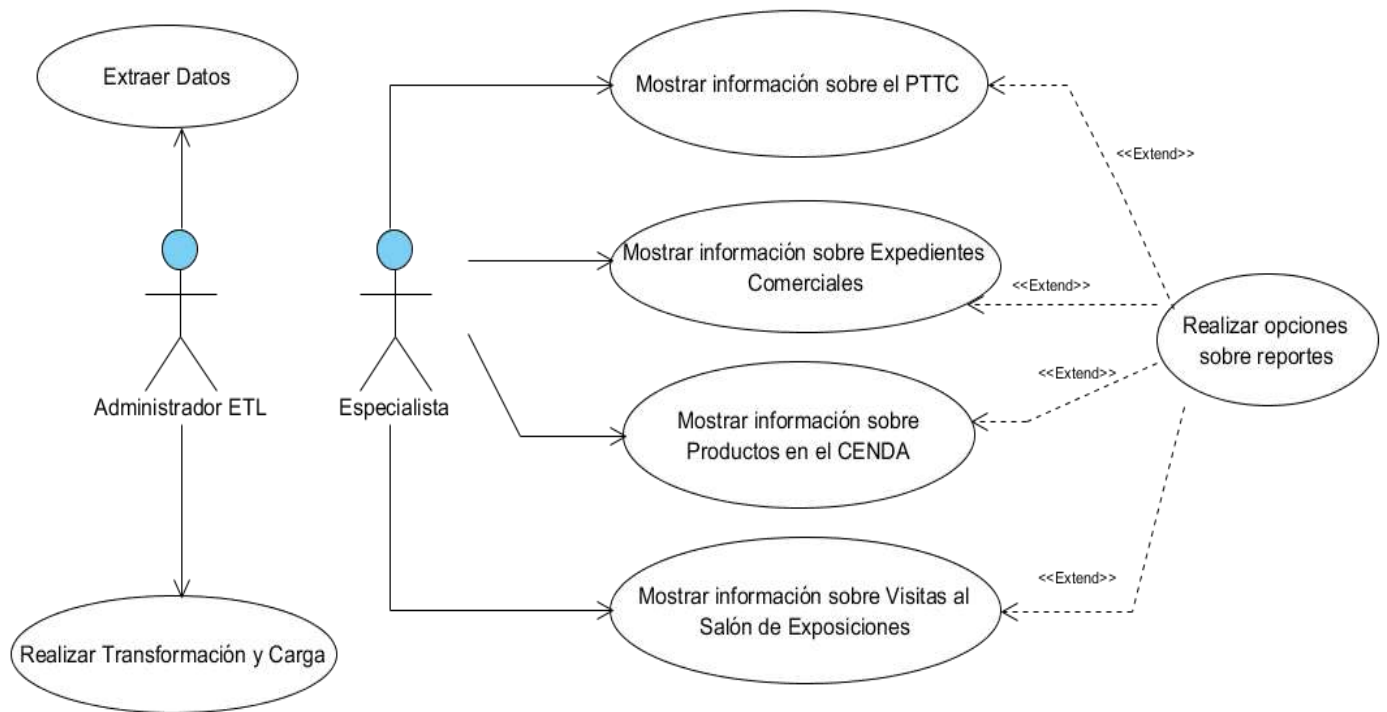


Fig.5: Diagrama de Caso de Uso del Sistema (Elaboración propia)

Especificación de casos de uso

La tabla siguiente muestra la explicación detallada del caso de uso "Mostrar información sobre el PTTC". El resto de las descripciones de los casos de uso se pueden encontrar en el artefacto "Especificación de casos de uso".

Capítulo 2: Análisis y diseño de la solución

Tabla.3: Descripción del caso de uso “Mostrar información sobre el PTTC”

Objetivo	Mostrar información sobre el PTTC.	
Actores	Especialista (Inicia).	
Resumen	El caso de uso inicia cuando el especialista desea consultar la información referente al PTTC. El especialista selecciona el reporte que desea ver y el sistema muestra la información contenida en él. El caso de uso finaliza cuando el especialista termina el análisis de la información relacionada con el PTTC.	
Complejidad	Media	
Prioridad	Alta	
Precondiciones	El especialista tiene que estar autenticado. El MD tiene que estar poblado. Los reportes relacionados con el PTTC fueron creados.	
Post condiciones	Los reportes correspondientes fueron consultados por el especialista.	
Flujo de eventos		
Flujo básico Mostrar información sobre el PTTC		
	Actor	Sistema
1	Selecciona el A.A PTTC.	
2		Muestra los Libros de Trabajos del área seleccionada.
3	Selecciona el Libro de Trabajo.	
4		Muestra las carpetas contenidas dentro del Libro de Trabajo.
5	Selecciona la carpeta deseada: 1 Vistas OLAP (Ver Evento 1) 2 Reportes personalizados (Ver Evento 2) 3 Cuadros de mando (Ver Evento 3)	
Evento 1: Vistas OLAP		
5.1.1	Selecciona la carpeta Vistas OLAP.	
5.1.2		Muestra las vistas contenidas dentro de la carpeta.

Capítulo 2: Análisis y diseño de la solución

5.1.3	Selecciona la vista que desea analizar.	
5.1.4		Muestra la información contenida en la vista seleccionada y brinda opciones al especialista para visualizar las vistas durante su análisis. Finaliza el CU.
Evento 2: Reportes personalizados		
5.2.1	Selecciona la carpeta Reportes personalizados.	
5.2.2		Muestra los reportes contenidos dentro de la carpeta.
5.2.3	Selecciona el reporte que desea analizar.	
5.2.4		Muestra el reporte sin datos.
5.2.5	Introduce los parámetros correspondientes.	
5.2.6		Muestra la información contenida en el reporte seleccionado y brinda opciones al especialista para visualizar los reportes durante su análisis. Finaliza el CU.
Evento 3: Cuadros de mando		
5.3.1	Selecciona la carpeta Cuadros de mando.	
5.3.2		Muestra los dashboard contenidos dentro de la carpeta.
5.3.3	Selecciona el dashboard que desea analizar.	
5.3.4		Muestra la información contenida en el dashboard seleccionado y brinda opciones al especialista para visualizar los dashboard durante su análisis. Finaliza el CU.
Opciones de Reportes		
Perspectivas de análisis		Posibles resultados
	Medidas	Periodicidad

Capítulo 2: Análisis y diseño de la solución

Variables de entrada relacionadas con el CU: <ul style="list-style-type: none"> • Estado legal • Encargado • Centro • tiempo (año y mes) 	Variables de salida disponibles: <ul style="list-style-type: none"> • Cantidad de contratos • Monto total en CUC • Monto total en CUP • Monto total en MT • Listado de contratos • Por ciento de contratos 	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 reportes”
Requisitos no funcionales	Sección: “3.2 Requisitos no funcionales” del documento: “Especificación de requisitos de software”.	
Asuntos pendientes		

2.8 Arquitectura del Mercado de Datos

Para el diseño de un MD es necesario definir una arquitectura que sirva para guiar el proceso de forma general. En la presente investigación la arquitectura diseñada consta de tres subsistemas fundamentales además de las fuentes de datos necesarias y los usuarios finales a los que está dirigido el trabajo.

Las fuentes de datos están compuestas por cuatro excel, estructurados de forma diferente, cada uno representa un proceso en específico. Dichas fuentes serán almacenadas en los esquemas definidos en el subsistema de almacenamiento con apoyo del gestor de BD PostgreSQL y el administrador de BD PgAdmin III. Para ello, es necesario realizar previamente los procesos de ETL definidos en el subsistema de integración mediante la utilización de las herramientas Pentaho Data Integration. Finalmente, estos datos serán mostrados a los jefes y especialistas del departamento, a través del subsistema de visualización que contendrá los pedidos de información definidos por los clientes mediante la herramienta Pentaho BI Server.

Por otra parte, se encuentran los usuarios finales, quienes, desde su puesto de trabajo, accederán al sistema instalado en un servidor mediante el protocolo HTTP del inglés (*Hypertext Transfer Protocol*) para realizar las operaciones correspondientes.

La figura 6, muestra cómo queda diseñada la arquitectura del MD para el Departamento de Mercadotecnia perteneciente a la DTTC.

Capítulo 2: Análisis y diseño de la solución



Fig.6: Arquitectura del Mercado de Datos (Elaboración propia)

2.8.1 Subsistema de Almacenamiento

Este subsistema es el encargado de almacenar toda la información recogida de las fuentes de datos correspondientes a los procesos que se llevan a cabo en el Departamento de Mercadotecnia. Después de los análisis correspondientes y la identificación de los RI según las necesidades de los clientes, fueron creadas cuatro tablas de hechos agrupadas en el esquema **mart_mercadotecnia**, y 19 tablas de dimensiones agrupadas en el esquema **dimension**. A continuación, se presenta la descripción de los hechos (ver tabla 4) y dimensiones (ver tabla 5).

Tabla.4: Descripción de hechos

No	Hechos	Descripción
H1	hecho_PTTC	Aquí se registran todos los contratos que se llevan a cabo en la institución.
H2	hecho_registro_expedientes_comerciales	Por cada producto comercializable se crea un expediente para ser mostrado a los clientes como proceso de marketing.
H3	hecho_registro_obras_CENDA	Aquí es donde a cada realizador de un producto u obra literaria se le otorga el derecho de ser reconocido legalmente como el propietario del producto a lo que se conoce como derecho de autor.
H4	hecho_registro_visitas_salón	Se archivan todas las visitas ya sea de empresas, exposiciones de productos UCI u otras actividades realizadas en el salón.

Capítulo 2: Análisis y diseño de la solución

Tabla.5: Descripción de dimensiones

No	Dimensiones	Descripción
D1	dim_centro	Representa a todos los centros de producción de la UCI.
D2	dim_fecha	Define una línea de tiempo específica donde el nivel más bajo es el mes.
D3	dim_estado_legal	Representa todos los posibles estados que puede tener un contrato.
D4	dim_encargado	Representa a los encargados de llevar a cabo cada contrato.
D5	dim_contrato	Representa todos los contratos que se tienen.
D6	dim_tipo_producto	Representa el tipo de producto de un expediente comercial.
D7	dim_expediente_com	Representa todos los expedientes comerciales que se tienen.
D8	dim_marca	Representa la marca sombrilla que tiene un expediente comercial.
D9	dim_novedad_catálogo	Representa si un producto es nuevo en el catálogo.
D10	dim_nuevaversión	Representa si un producto tiene una nueva versión.
D11	dim_alcance_mercado	Representa el alcance en el mercado de un producto.
D12	dim_obra_literaria	Representa las obras literarias que están en el CENDA y su tipo.
D13	dim_estado_CENDA	Representa el estado que puede tener una obra literaria en el CENDA.
D14	dim_asesor	Representa si una visita requirió la presencia de asesores.
D15	dim_clasificacion	Representa si las visitas al salón fueron planificadas y realizadas.
D16	dim_tipo_visitante	Representa el tipo de visitante que tiene una visita realizada al salón.
D17	dim_tipo_organización	Representa el tipo de organización que tiene una visita realizada al salón.
D18	dim_interés	Representa el tipo de interés que tiene una visita realizada al salón.
D19	dim_visitante	Contiene el nombre de todas las visitas hechas al salón.

Matriz bus

La matriz bus representa la relación que existe entre las tablas de hechos y las tablas de dimensiones, permite describir la trazabilidad de los mismos, evitando que ocurra el solapamiento entre las tablas de hechos.

A continuación, se muestra la matriz bus correspondiente al MD para el Departamento de Mercadotecnia, donde las filas contienen a los hechos y las columnas a las dimensiones. Se pudo identificar que la dimensión fecha es compartida entre todos los hechos y la dimensión centro la comparten los hechos H1 hecho_PTTC, H2 hecho_registro_expedientes_comerciales y H3 hecho_registro_obras_CENDA. Ver tabla 6.

Capítulo 2: Análisis y diseño de la solución

Tabla.6: Matriz Bus correspondiente al Departamento de Mercadotecnia

Hechos	Dimensiones																		
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16	D17	D18	D19
H1	x	x	x	x	x														
H2	x	x				x	x	x	x	x	x								
H3	x	x										x	x						
H4		x												x	x	x	x	x	x

Modelo de datos

El modelo de datos representa los tipos de datos que hay en una base de datos y la forma en que se relacionan entre sí. A continuación, se muestra un fragmento del modelo de datos diseñado para el Departamento de Mercadotecnia perteneciente a la DTTC (Ver figura 7). El diagrama ampliado se puede apreciar en los anexos (Ver [Anexo #1](#)).

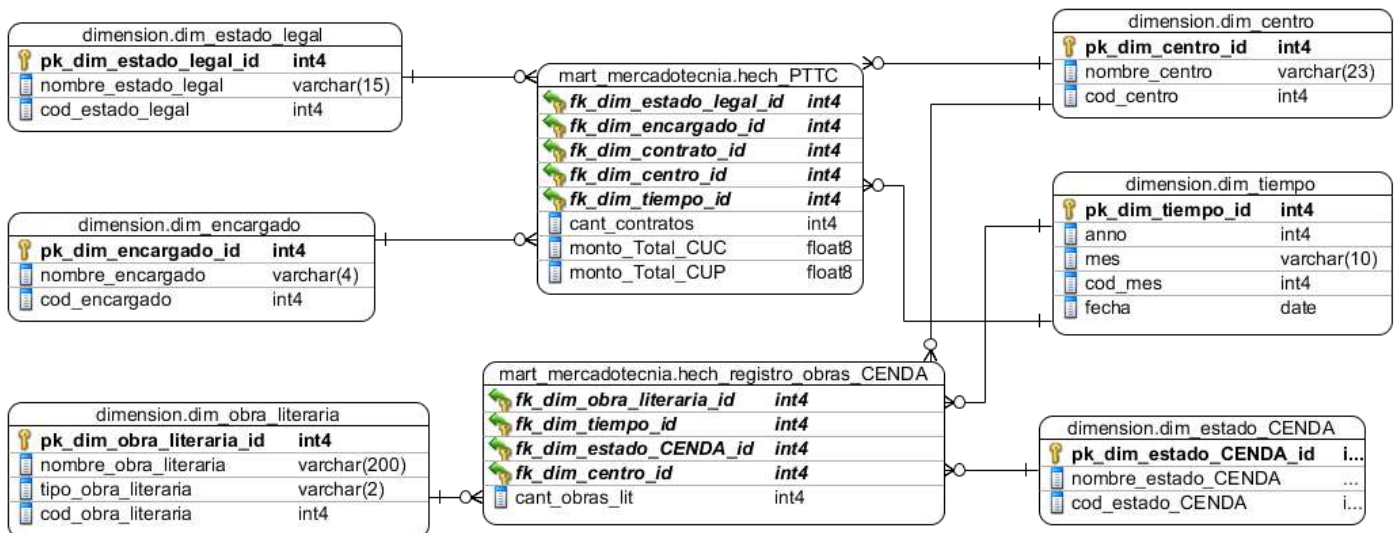


Fig.7: Diagrama Entidad-Relación (Elaboración propia)

En la presente investigación el modelo de datos antes realizado representa la variante de esquema constelación o copo de estrella ya que tiene varios hechos y estos comparten dimensiones.

2.8.2 Subsistema de Integración

En este subsistema es donde se realiza el diseño de las transformaciones para la extracción de los datos desde sus fuentes originales, se transforman y posteriormente son cargados en la base de datos destino mediante la herramienta PDI. La integración de datos se puede enfocar de varias formas, según (Azán Basallo, y otros, 2009) existen cuatro:

Capítulo 2: Análisis y diseño de la solución

- ❖ **Replicación de datos:** se basa en la creación y mantenimiento de múltiples copias de una misma BD. Un servidor mantiene la copia primaria de la BD y servidores adicionales mantienen las copias esclavas de la misma. La transferencia se realiza de BD a BD, este es el estilo más antiguo de integración que existe.
- ❖ **Integración de Información Empresarial (EII):** este tipo de solución consiste en crear un intermediario que contenga los directorios de la BD y que a su vez sirva de canal de consulta y representación de la información recuperada. La información es capturada en tiempo real, lo que implica que las fuentes de datos tengan una estructura tecnológica sólida y bien establecida.
- ❖ **Integración de Aplicaciones Empresariales (EAI):** es el proceso de integrar múltiples aplicaciones desarrolladas independientemente, que utilizan tecnología incompatible y que son gestionadas de forma independiente, permitiendo que se comuniquen e intercambien transacciones de negocio, mensajes y datos entre sí.
- ❖ **Extracción, Transformación y Carga de Datos (ETL):** como su nombre lo indica extrae información de un sistema fuente, transforma esos datos para satisfacer los requerimientos del negocio y carga el resultado en el sistema destino. Tanto la fuente como el destino son generalmente BD o archivos.

Atendiendo a las diferentes técnicas antes mencionadas se decidió utilizar ETL, ya que es la que más se ajusta a las características del negocio, teniendo en cuenta que la información se encuentra en ficheros .xls y se requiere de su transformación para la posterior carga.

Perfilado de datos

Analizar las fuentes existentes en la organización, es una de las tareas más importantes, consiste en realizar un primer análisis de los datos de origen, con el objetivo de conocer su estructura y nivel de calidad. Esto permite identificar los distintos tipos de datos con los que se cuenta y obtener estadísticas sobre ellos, además de identificar valores incorrectos, duplicados o nulos para poder corregirlos. Después de realizado el perfilado de datos a las fuentes existentes en el Departamento de Mercadotecnia fueron detectados 74.07 % de datos tipo texto (String), 3.70 % de tipo entero (Integer), 14.81 % de tipo fecha (Date) y 7.41 % de tipo decimal (Double) como se muestra en la figura 8, donde los valores están redondeados por exceso. También fueron analizadas la cantidad de filas nulas que presentaban cada uno de los ficheros. La figura 9 muestra la cantidad de datos nulos en el fichero PTTC por cada atributo de dicha tabla.

Capítulo 2: Análisis y diseño de la solución

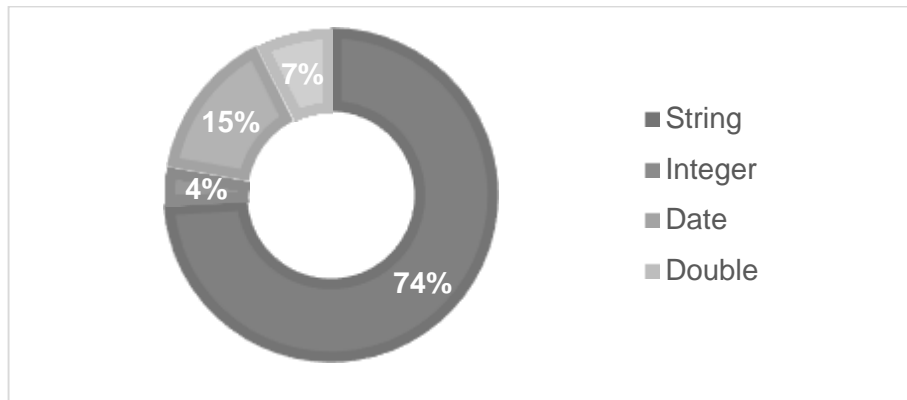


Fig.8: Distribución por tipo de datos (Elaboración propia)

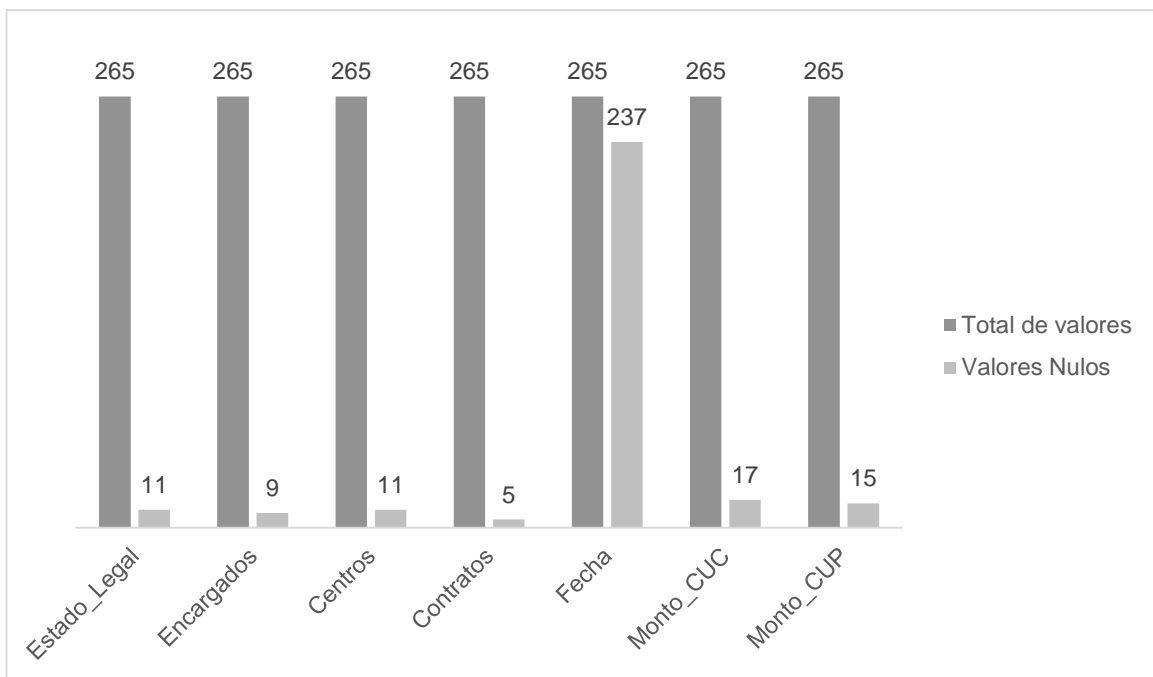


Fig.9: Cantidad de datos nulos en el excel PTTC (Elaboración propia)

Diseño de las transformaciones

Diseño de las transformaciones para la carga de dimensiones: para la carga de las dimensiones, lo primero que se realiza es la extracción de los datos de los ficheros fuentes (excel) o se introducen directamente. Luego se seleccionan los datos correspondientes a cargar en las dimensiones para realizarles la limpieza, transformación y carga en la base de dato destino. A continuación, se muestra en la figura 10 el diseño de las transformaciones para la carga de las dimensiones.

Capítulo 2: Análisis y diseño de la solución

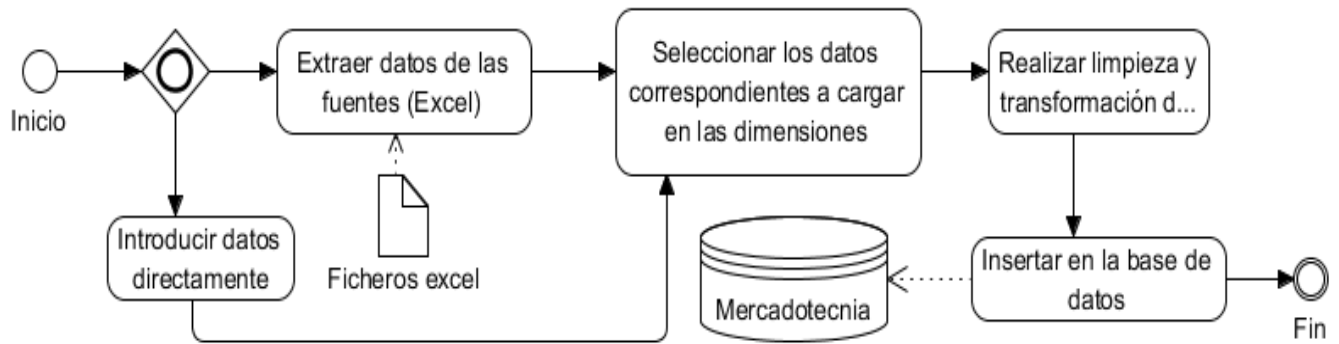


Fig.10: Diseño de las transformaciones para la carga de dimensiones (Elaboración propia)

Diseño para la carga de hechos: para la carga de los hechos se extraen los datos de los ficheros fuentes, en este caso excel y se les realiza la limpieza y transformación. Se buscan en la base de datos los id de las dimensiones y se validan que estos no sean nulos, en caso de serlo, pasan a un excel de errores para darle el tratamiento adecuado. Posteriormente, se obtienen las medidas deseadas y se insertan en la base de datos finalizando el proceso. A continuación, se muestra en la figura 11, el diseño de las transformaciones para la carga de los hechos.

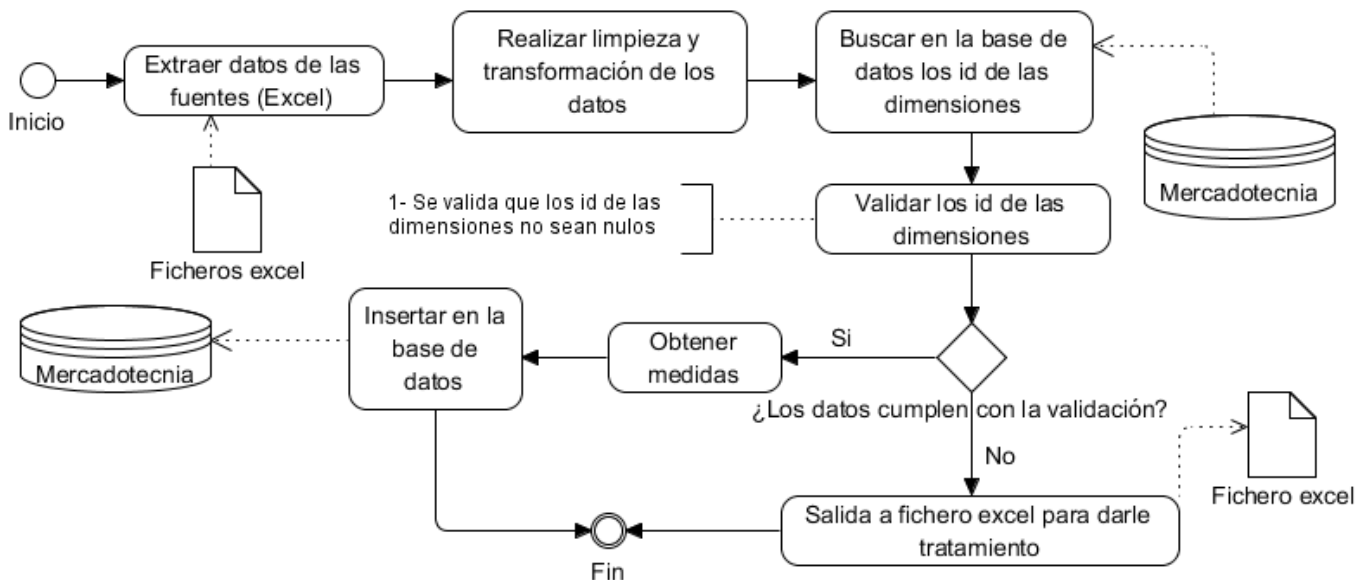


Fig.11: Diseño de las transformaciones para la carga de hechos (Elaboración propia)

Capítulo 2: Análisis y diseño de la solución

2.8.3 Subsistema de Visualización

En este subsistema se muestran los resultados finales a los usuarios a través de la herramienta Pentaho BI Server, además, se definirá la arquitectura de información y los reportes candidatos.

Arquitectura de información

Para lograr una organización referente a la información que posee el MD para el Departamento de Mercadotecnia se realizó la arquitectura de información. La misma cuenta con un Área de Análisis General (A.A.G) Mercadotecnia que agrupa cuatro Áreas de Análisis (A.A) específicas dependiendo de los procesos que se llevan a cabo en el negocio, estas son: Expedientes Comerciales, Productos en el CENDA, PTTC y Visitas al Salón de Exposiciones, cada una con su Libro de Trabajo (LT) asociado para un total de cuatro LT. Cada LT incluye tres carpetas relacionadas con las vistas OLAP, los reportes personalizados y los cuadros de mando. Cada carpeta contiene las Tablas de Salidas (TS) en correspondencia con los reportes candidatos asociados a las necesidades del cliente, con un total de 16 TS. A continuación, se muestra la figura 12 que representa la estructura de navegación definida para el MD.

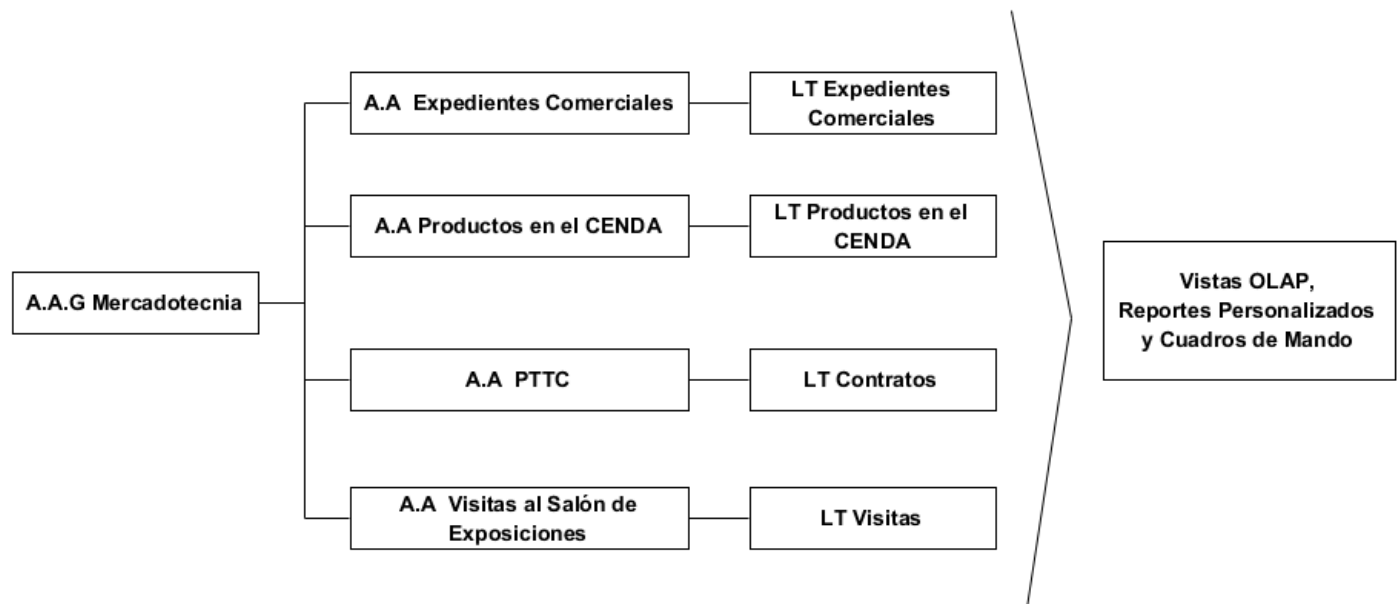


Fig.12: Diseño del mapa de navegación (Elaboración propia)

Diseño de los reportes candidatos

Después de identificadas las necesidades de información establecidas por el cliente, fueron realizados los reportes, que se corresponden con las salidas que debe mostrar el sistema. A continuación, se muestran dos ejemplos de Tablas de Salidas (TS) correspondientes a uno de los RI asociados a las A.A PTTC y Visitas al Salón de Exposiciones, tales como: **cantidad de contratos** y **cantidad de visitantes**, estos se pueden apreciar en las tablas 7 y 8, respectivamente. Las demás TS asociadas a cada uno de los requisitos se pueden encontrar en el artefacto "Salida del sistema".

Capítulo 2: Análisis y diseño de la solución

Tabla.7: Reporte Cantidad de contratos del AA PTTC

Área de análisis (AA)	PTTC
Libro de Trabajo (LT)	Contratos
Reporte (Tabla de Salida – TS)	Cantidad de contratos
Descripción	Muestra la cantidad de contratos archivados dado el centro, el tiempo (año y mes), el estado legal y el encargado.
Elementos del reporte	Centro, Estado legal, Encargado y Tiempo
Frecuencia de emisión	Mensual
Funciones	Suma
Gráfico	Barras verticales y de pastel

Tabla.8: Reporte Cantidad de visitantes del AA Visitas al Salón de Exposiciones

Área de análisis (AA)	Visitas al Salón de Exposiciones
Libro de Trabajo (LT)	LT Visitas
Reporte (Tabla de Salida – TS)	Cantidad de visitantes
Descripción	Muestra la cantidad de visitas hechas al salón dado el tipo de organización, el tipo de visitante, el tiempo (año y mes), el interés, la planificación, la realización y el uso de asesores.
Elementos del reporte	Tipo de organización, Tipo de visitante y Tiempo
Frecuencia de emisión	Mensual
Funciones	Suma
Gráfico	Barras verticales y de pastel

2.9 Políticas de respaldo y recuperación

En caso de ocurrir fallos en el sistema, ya sea por causas internas o externas, el mismo debe estar preparado para que no pierda la información que contiene. Para esto se implementan las políticas de respaldo y recuperación, las que permiten mantener la integridad de la información. El MD para el Departamento de Mercadotecnia realizará salvadas a la información contenida en la BD y las copias se realizarán de forma mensual a través de tareas programadas en el sistema operativo.

Roles y permisos

Para el acceso al MD se definieron roles para cada usuario existente en el sistema. La tabla 11 muestra la asignación de los roles con sus permisos asociados.

Capítulo 2: Análisis y diseño de la solución

Tabla.9: Roles y permisos

Roles	Base de Datos		Aplicación	
	Lectura	Escritura	Lectura	Escritura
Administrador de ETL	X	X		
Especialista			X	

2.10 Conclusiones del capítulo

En este capítulo se realizó un estudio detallado del negocio, lo que permitió identificar 13 RNF y 26 RN que detallan la propuesta de solución. Además, se identificaron 16 RI y dos RF, que permitieron la confección del diagrama de CUS. Se definió la arquitectura a utilizar en el MD identificándose tres subsistemas fundamentales: almacenamiento, integración y visualización, los que dieron paso a la estructura, funcionamiento e interacción de cada uno de ellos en el MD. Fueron identificados cuatro tablas de hechos y 19 tablas de dimensiones lo que posibilitó la creación del modelo lógico y la matriz bus correspondiente, para evitar el solapamiento entre los hechos. Como técnica de integración de datos se empleó ETL, la que facilitó la extracción de los datos de sistemas fuentes para luego transformarlos y cargarlos en la BD destino. Esta técnica permitió realizar 19 transformaciones correspondientes a las dimensiones, cuatro a los hechos y tres trabajos para la ejecución de las transformaciones de forma organizada. El perfilado de datos accedió a la identificación de los distintos tipos de datos que se encontraban en las fuentes y la cantidad de datos nulos presentes. Fue identificada la arquitectura de navegación para el acceso de forma rápida a la información solicitada. Por último, se asignaron los permisos necesarios para evitar la mala manipulación de los datos.

Capítulo 3: Implementación y Prueba de la solución

Capítulo 3: Implementación y Prueba de la solución

En este capítulo se realizará la implementación de la solución y sus pruebas correspondientes, para esto es necesario implementar los subsistemas de almacenamiento, integración y visualización, teniendo en cuenta los requisitos de información definidos con el cliente. Además, se expondrán los resultados de las pruebas realizadas al MD para el Departamento de Mercadotecnia para que cumpla con sus funcionalidades y características definidas.

3.1 Implementación del Subsistema de Almacenamiento

Una vez realizado el modelo lógico de los datos, se implementa el modelo físico teniendo en cuenta los esquemas, tablas de hechos y dimensiones. Los mismos están estructurados de manera que permitan tener la información organizada, usando las tablas de hechos y dimensiones creadas previamente, ello les permitirá acceder una vez otorgados los permisos. Se consideró que la variante de implementación a utilizar para el almacenamiento es ROLAP, ya que el MD para el Departamento de Mercadotecnia será implementado sobre el SGBD PostgreSQL que es de tipo relacional.

Esquemas

- ❖ **dimension:** contiene las 19 dimensiones identificadas en el MD para el Departamento de Mercadotecnia.
- ❖ **mart_mercadotecnia:** contiene los cuatro hechos identificados en el MD para el Departamento de Mercadotecnia.

3.1.1 Estándares de codificación

La tabla 12 muestra algunos de los estándares de codificación utilizados para el MD, con el objetivo de lograr una mayor comprensión entre las partes involucradas. Los demás se pueden apreciar en el artefacto “Estándares de Codificación”.

Tabla.10: Estándar de codificación para el MD

Tipo de Objeto	Función	Nomenclatura	Descripción
Esquemas	Hechos	mart_[nombre]	Esquema donde se encuentran las tablas de hechos.
	Dimensiones	dimension	Esquema donde se encuentran las tablas de dimensiones.
Tablas	Dimensiones	dim_[nombre]	Tablas de dimensiones utilizadas como perspectivas de análisis.

Capítulo 3: Implementación y Prueba de la solución

	Hechos	hech_[nombre]	Tablas de hechos que contienen los principales procesos con sus medidas asociadas.
Campos de las tablas	Llave de las dimensiones	pk_dim_[nombre]_id	Contiene el identificador de la llave primaria de las dimensiones.
	Llave de los hechos	fk_dim_[nombre]_id	Contiene el identificador de cada una de las llaves primaria de las dimensiones.

3.2 Implementación del Subsistema de Integración

Para poder extraer los datos desde los ficheros fuentes, para luego manipularlos, integrarlos, transformarlos y cargarlos, es necesario contar con algún proceso que se encargue de ello. Precisamente los ETL son los que cumplirán con tal fin. Tal y como sus siglas lo indican, los ETL, extraen datos de las diversas fuentes que se requieran, los transforman para resolver posibles problemas de inconsistencias entre los mismos y finalmente, después de haberlos depurado, se procede a su carga en el depósito de datos. (Bernabeu, 2010)

Bernabeu define tres funciones específicas de los ETL, estas son:

- ❖ **Extracción:** es aquí donde basándose en las necesidades y requisitos del usuario, se exploran las diversas fuentes OLTP que se tengan a disposición, y se extrae la información que se considere relevante al caso. En esta primera etapa los datos quedan preparados para el proceso de transformación.
- ❖ **Transformación:** esta función es la encargada de convertir aquellos datos inconsistentes en un conjunto de datos compatibles y congruentes, para que puedan ser cargados posteriormente. Estas acciones se llevan a cabo, debido a que pueden existir diferentes fuentes de información, y es vital conciliar un formato y forma única, definiendo estándares y convenciones de nombramientos, para que todos los datos estén integrados.
- ❖ **Carga:** este proceso es el encargado de cargar los datos provenientes de la fase anterior en la BD destino.

Los ETL, son los encargados de realizar dos tareas bien definidas:

- ❖ **Carga Inicial:** la carga inicial, se refiere precisamente a la primera carga de datos que se realizará.
- ❖ **Actualización, mantenimiento o refrescado periódico** (siempre teniendo en cuenta un intervalo de tiempo predefinido para tal operación): los mantenimientos periódicos mueven pequeños volúmenes de datos, y su frecuencia está dada en función de los requerimientos del usuario. El objetivo de esta tarea es añadir al depósito aquellos datos nuevos que se fueron generando desde el último refresco.

Capítulo 3: Implementación y Prueba de la solución

3.2.1 Transformaciones y trabajos

Para la extracción de los datos, como parte del proceso de ETL se realizaron las transformaciones correspondientes a los hechos y dimensiones identificados. A continuación, la figura 13 muestra un ejemplo de transformación para la dimensión estado_legal, la misma solo está presente en el hecho PTTC y la carga comienza con la extracción de los datos de los sistemas fuentes y con la inserción manual de una denominación (SV), posteriormente se unen las dos entradas, se filtran los datos nulos a un excel de errores para su posterior tratamiento, los datos no nulos pasan para el siguiente componente, se ordenan, se agrupan por estado legal, se añade un código y se finaliza con la carga de los datos en la BD destino.



Fig.13: Proceso de integración de datos para la dimensión estado_legal (Elaboración propia)

El hecho PTTC se inicia con la extracción de datos del fichero fuente PTTC, se continúa con la eliminación de los datos que no se utilizarán en la BD, se filtran las filas que poseen fecha nula enviándolas a un excel de errores para su posterior tratamiento, se sustituyen los campos nulos por los valores SV y 0, se obtienen los identificadores (id) de las dimensiones, posteriormente se eliminan todos los valores quedando solo los id, se ordenan y se agrupan las filas para luego cargar los datos en la BD destino. La figura 14 muestra la transformación antes descrita.

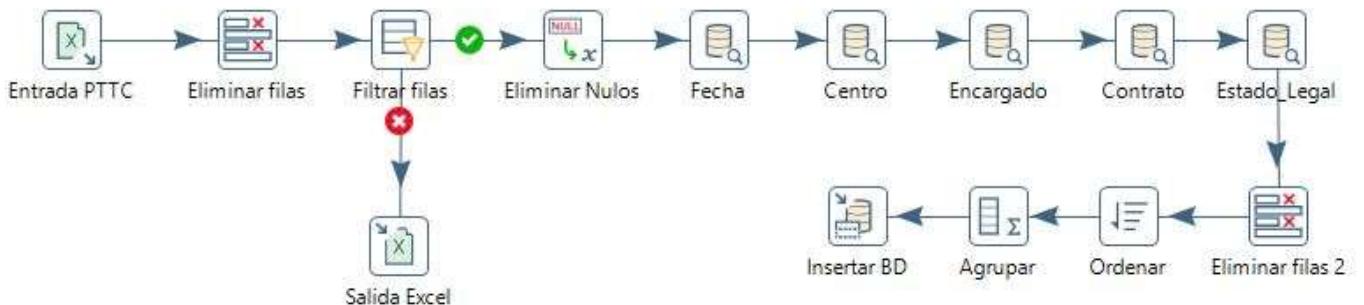


Fig.14: Proceso de integración de datos para el hecho PTTC (Elaboración propia)

Capítulo 3: Implementación y Prueba de la solución

Para realizar las transformaciones de forma automatizada se realizan los trabajos o Jobs. Para el MD del Departamento de Mercadotecnia se realizaron tres trabajos, cuyo objetivo consiste en evitar cargar individualmente las transformaciones.

La siguiente figura muestra la ejecución de todas las transformaciones correspondientes a las dimensiones anteriormente definidas (Ver figura 15). En caso de ocurrir algún error en la carga de una de las dimensiones, este se envía a una salida excel sin afectar el proceso.

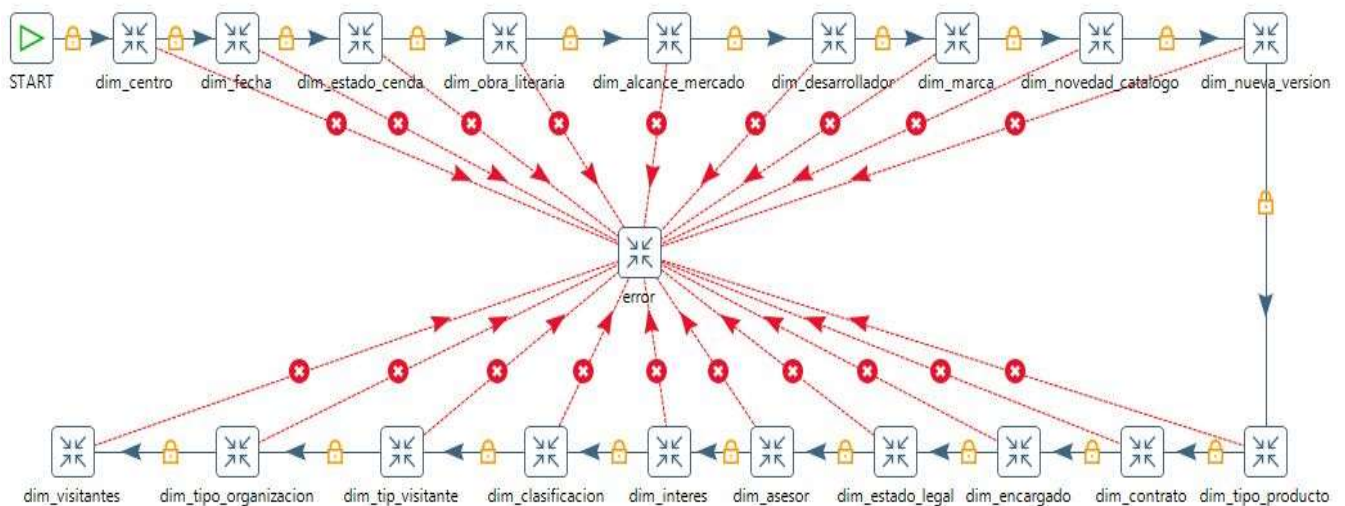


Fig.15: Trabajo para las dimensiones (Elaboración propia)

La figura 16 muestra la ejecución de todas las transformaciones correspondientes a los hechos definidos anteriormente. En caso de ocurrir algún error en la carga de uno de los hechos, este se envía a una salida excel sin afectar el proceso.

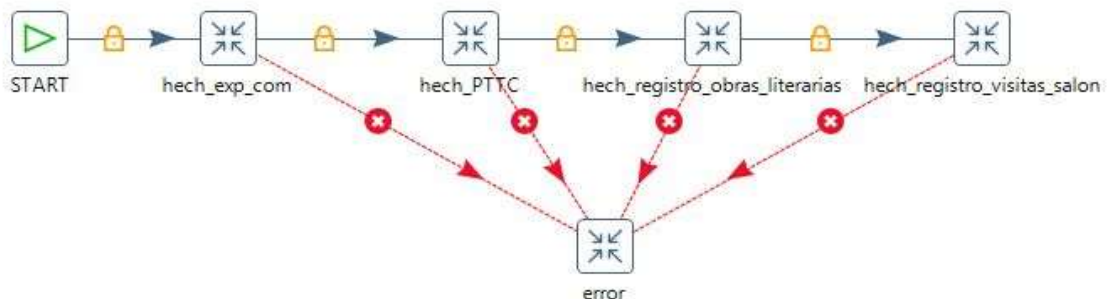


Fig.16: Trabajo para los hechos (Elaboración propia)

Fue realizado también un tercer trabajo, que ejecuta automáticamente la carga de los dos mencionados anteriormente.

Capítulo 3: Implementación y Prueba de la solución

3.3 Implementación del Subsistema de Visualización

Luego de haber concluido con los subsistemas anteriores, se procede a la implementación del subsistema de visualización, que es el responsable de mostrar el resultado final a los usuarios. Para esto se realizan los cubos OLAP, la capa de visualización y los reportes candidatos.

3.3.1 Implementación de los cubos OLAP

Para la implementación de los cubos OLAP del MD se definió un cubo por cada hecho identificado, así como las dimensiones, medidas y niveles de jerarquía que conforman el esquema para el MD. La figura 17 muestra la estructura de los cubos.



Fig.17: Implementación de los cubos OLAP (Elaboración propia)

3.3.2 Implementación de la capa de visualización

Tener estructurada y organizada la información para el entendimiento de los usuarios finales constituye un requisito fundamental. Para esto se crea la arquitectura de información o mapa de navegación. Esta, para el Departamento de Mercadotecnia, cuenta con un A.A.G, cuatro A.A y cuatro LT. Estos últimos contienen tres carpetas referentes a las vistas de análisis, reportes personalizados y cuadros de mando. A continuación, se detallan las características de los elementos que componen la arquitectura, su estructura física se puede apreciar en la figura 18.

- ❖ **AAG Mercadotecnia:** contiene la información referente al Departamento de Mercadotecnia perteneciente a la DTTC.

Capítulo 3: Implementación y Prueba de la solución

- ❖ **AA Expedientes Comerciales:** contiene la información referente a los expedientes comerciales llevados a cabo en la institución.
- ❖ **AA Productos en el CENDA:** contiene la información referente a los productos u obras literarias puestas a inscribirse en el CENDA.
- ❖ **AA PTTC:** contiene la información referente a los contratos llevados a cabo por la institución.
- ❖ **AA Visitas al Salón de Exposiciones:** contiene la información referente a las visitas realizadas en el salón de exposiciones.
- ❖ **LT Expedientes Comerciales:** contiene los reportes operacionales y dashboard referentes a la cantidad de expedientes comerciales.
- ❖ **LT Productos en el CENDA:** contiene los reportes operacionales y dashboard referentes a la cantidad de obras literarias y productos en el CENDA.
- ❖ **LT Contratos:** contiene los reportes operacionales y dashboard referentes a la cantidad de contratos llevados a cabo en la institución.
- ❖ **LT Visitas:** contiene los reportes operacionales y dashboard referentes a las visitas realizadas al salón.
- ❖ **Cuadros de Mando:** contiene los cuadros de mando referentes a cada LT.
- ❖ **Reportes Personalizados:** contiene los reportes personalizados referentes a cada LT.
- ❖ **Vistas OLAP:** contiene las vistas OLAP referentes a cada LT.

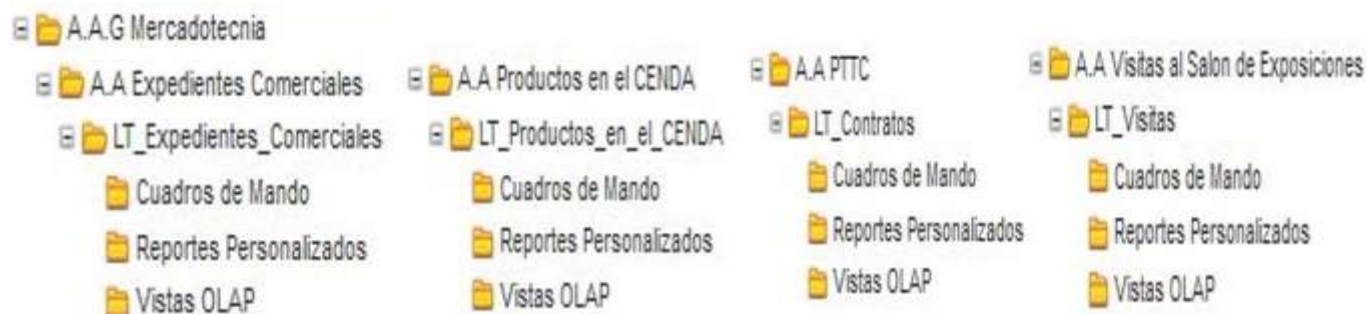


Fig.18: Arquitectura de información del MD (Elaboración propia)

3.3.3 Implementación de las vistas OLAP

Una vista es una tabla virtual cuyo contenido está definido por una consulta. Al igual que una tabla, una vista consta de un conjunto de columnas y filas de datos con un nombre. Actúa como filtro de las tablas subyacentes a las que se hace referencia en ella. Suelen usarse para centrar, simplificar y personalizar la percepción de la BD para cada usuario. Las vistas pueden emplearse como mecanismos de seguridad, que permiten a los usuarios obtener acceso a los datos por medio de la vista, pero no les conceden el permiso de obtener acceso directo a las tablas base subyacentes de la vista. (Microsoft, 2016)

Capítulo 3: Implementación y Prueba de la solución

La vista de análisis correspondiente al A.A PTTC (Ver figura 19) muestra la cantidad de contratos registrados en un determinado año por centro, estado legal y encargado. Con la opción de despliegue, según el criterio de búsqueda, se facilita el entendimiento por parte de los especialistas encargados de realizar análisis y dar respuestas al departamento.

		Fecha					
		* Todos					
		Estado Legal					
		Detenido			En Ejecución		Facturado
		Encargado					
		SCT3	SCT1	SCT3	SCT2	SCT1	
		Medidas	Medidas	Medidas	Medidas	Medidas	
Contrato	Centro	• Cant_contratos	• Cant_contratos	• Cant_contratos	• Cant_contratos	• Cant_contratos	
* Todos	* Todos	1	11	16	7	1	
	CEGEL					1	
	CEIGE	1		1	1		
	CESOL				1		
	CIDI				1		
	CIGED		2		1		
	CISED		1				
	FORTES			2			
	Soporte		8	13	3		

Fig.19: Vista de análisis referente al PTTC (Elaboración propia)

3.3.4 Implementación de los reportes candidatos

Después de analizar los pedidos de información realizados por el cliente y las fuentes de datos que recogían dicha información, fueron creados los reportes correspondientes al MD para el Departamento de Mercadotecnia. Estos fueron elaborados con la herramienta PRD, a través de consultas SQL. La figura 20 muestra un reporte correspondiente al A.A PTTC. En él se ilustra un listado de contratos con su centro, encargado, monto en CUP, monto en CUC, monto en MT y su fecha correspondiente. En la parte inferior se encuentra el total de reportes encontrados según el criterio de búsqueda seleccionado en la parte superior, el por ciento que estos representan y el total de efectivo correspondiente a las tres monedas.

Centro	Soporte
Año	2016
Mes	Mayo
Estado Legal	En Ejecución
Encargado	SCT3

Capítulo 3: Implementación y Prueba de la solución

09 junio, 2017, 12:28

Estado Legal: En Ejecución

DTTC

Dirección de Tecnología y Transferencia de Conocimiento

Contratos

Centro	Contrato	Encargado	Fecha	Monto CUP	Monto CUC	Monto Total
Soporte	Actualización, Transferencia y Soporte del Portal Web de LABIOFAM.	SCT3	09/05/2016	7.448,95	3.192,41	10.641,36
Soporte	Despliegue, acompañamiento y soporte al Sistema de Planificación de Actividades (SIPAC v2.1) para el MINAL	SCT3	16/05/2016	12.294,00	0,00	12.294,00
Soporte	Despliegue, acompañamiento y soporte al Sistema de Planificación de Actividades (SIPAC v2.1) para INDER	SCT3	30/05/2016	12.294,00	0,00	12.294,00
Soporte	Renovación Soporte a GINA (Suplemento al contrato específico 3 prorrogando vigencia de contrato)	SCT3	10/05/2016	4.724,40	0,00	4.724,40
Soporte	Servicio de soporte técnico a la Distribución Cubana de GNU/Linux, NOV para el Banco Central de Cuba (BCC).	SCT3	22/05/2016	12.153,96	0,00	12.153,96
Soporte	Servicio de soporte técnico al Sistema Ventanilla Única Aduanera (VUA) para AGR	SCT3	10/05/2016	11.010,60	0,00	11.010,60
Total	6			59.925,91	3.192,41	63.118,32

Por Ciento de Contratos: 1.00.00%

1 / 1

Fig.20: Reporte asociado al PTTC (Elaboración propia)

3.3.5 Implementación de los Dashboard

Un dashboard es una representación gráfica de las principales métricas que intervienen en el proceso de negocio de una empresa. Esta herramienta permite visualizar el problema y favorecer la toma de decisiones. El fin último es transformar los datos en información útil para orientar las estrategias hacia la consecución de los objetivos planteados. Debe ser un gráfico limpio y ordenado, que permita entender, de un solo vistazo, los datos que se plantean, con el fin de visualizar, contextualizar y comparar datos de forma tal que facilite el establecimiento de valoraciones útiles. (Elósegui, 2014)

Capítulo 3: Implementación y Prueba de la solución

Para presentar la información a los usuarios finales fueron creados, mediante gráficos, los dashboard, los que representan en alto nivel el contenido de BI. La figura 21 corresponde a un dashboard realizado para el A.A PTTC. Para que este muestre su información, primero se tiene que seleccionar el rango de fecha (por año) deseado. A continuación, se muestra un ejemplo que evidencia la cantidad de contratos en el año 2015, por centro, por encargado y según su estado legal.

Capítulo 3: Implementación y Prueba de la solución



Fig.21: Dashboard correspondiente al PTTC (Elaboración propia)

Capítulo 3: Implementación y Prueba de la solución

3.4 Pruebas

Las pruebas constituyen una de las actividades más importantes dentro del desarrollo de software, muestran la información sobre la calidad del producto diseñado a los responsables de esto. Pueden ser ejecutadas en diferentes momentos (al inicio, en el medio o al final), hasta que el producto cumpla con la calidad requerida y las expectativas del cliente.

La realización de las pruebas al MD para el Departamento de Mercadotecnia se sustentó en el Modelo V (figura 21), utilizado por el centro DATEC para garantizar el buen funcionamiento y la calidad de los productos de AD. Dicho modelo cuenta con un vértice común en la codificación, una rama izquierda donde se muestran las actividades relacionadas con el análisis y diseño del producto en cuestión y una rama derecha que muestra las actividades relacionadas con el aseguramiento de la calidad mediante los tipos de pruebas aplicadas.

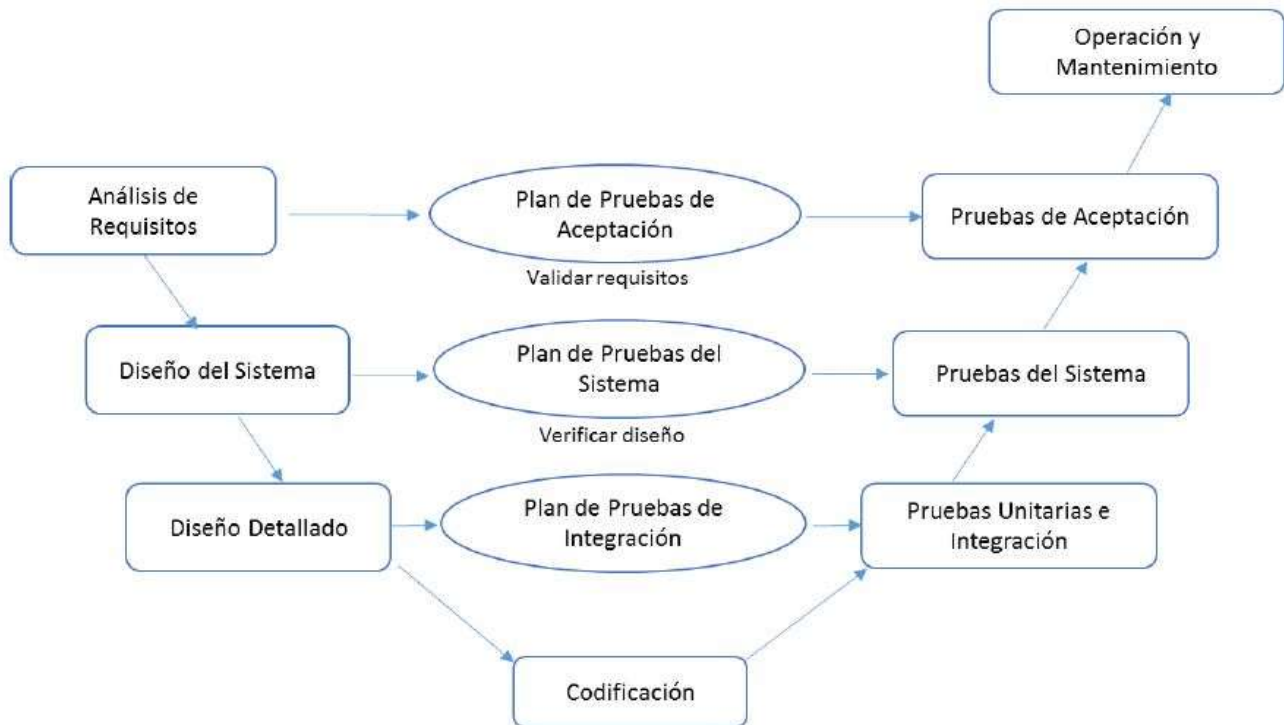


Fig.22: Modelo V (Elaboración propia)

Según (Herrera, 2012) estas pruebas consisten en:

Pruebas unitarias: tienen como objetivo verificar la funcionalidad y estructura de cada componente, de forma individual, una vez que ha sido codificado. Es decir, es mejor probar primero los bloques desarrollados más pequeños del programa, que inicialmente probar el software en su totalidad. Estas pruebas son realizadas por el programador a medida que implementa los componentes.

Capítulo 3: Implementación y Prueba de la solución

Pruebas de integración: verifica el correcto ensamblaje entre dos o más componentes una vez que han sido probados, unitariamente, con el fin de comprobar que interactúan correctamente a través de sus interfaces, tanto internas como externas. Este tipo de pruebas son dependientes del entorno en el que se ejecutan. Aunque el código sea correcto, pueden ser fallidas si existe un cambio en el entorno.

Pruebas de sistemas: las pruebas de sistema tienen un propósito particular, comparar todo el sistema de software completo e integrado con sus objetivos originales (normalmente desde el punto de vista de requisitos de la aplicación). Es decir, las pruebas de sistema buscan discrepancias entre el programa y sus objetivos o requerimientos, enfocándose en los errores hechos durante la transición del proceso al diseñar la especificación funcional.

Pruebas de aceptación: el objetivo de las pruebas de aceptación es comprobar si el software cumple con las expectativas del cliente. Permite al usuario de dicho sistema que determine su aceptación, desde el punto de vista de su funcionalidad y rendimiento. Las pruebas de aceptación son definidas por el usuario del sistema y preparadas por el equipo de desarrollo, aunque la ejecución y aprobación final corresponden al usuario.

3.4.1 Herramientas para la aplicación de las pruebas

A continuación, se detallan las herramientas aplicadas a los tipos de pruebas definidas anteriormente.

Casos de Prueba

Los casos de prueba (CP) especifican los requisitos de la aplicación, por lo que cada uno de ellos debe estar cubierto por un mínimo de un caso de prueba. Cada CP está compuesto por varios pasos a ejecutar, dependiendo de la complejidad del caso, y cada paso está compuesto por una acción y un resultado esperado. Para que un caso de prueba resulte exitoso, todos los pasos deben cumplir el resultado esperado. Si uno de los pasos no lo cumple, el CP resultará fallido. (Globe, 2012)

Para el MD correspondiente al Departamento de Mercadotecnia fueron diseñados cuatro CP, uno por cada caso de uso de información. La tabla 13 muestra dos de los escenarios correspondientes al CU “Mostrar información sobre el PTTC”.

Capítulo 3: Implementación y Prueba de la solución

Tabla.11: Diseño del CP para el CU “Mostrar información sobre el PTTC”

Escenarios	Descripción	Variables de entrada	Variable de salida	Respuesta del sistema	Flujo central
EC 1: Muestra información sobre los contratos archivados	Muestra un listado de contratos por centro, estado legal, encargado y tiempo dado el centro, el tiempo (año y mes), el estado legal y el encargado	Estado legal	Listado de contratos	El sistema muestra todas las variables por las cuales se puede filtrar la información para obtener el listado	El usuario se autentica, selecciona el A.A.G Mercadotecnia / A.A PTTC / L.T Contrato / Selecciona la carpeta deseada y el reporte, la vista o dashboard correspondiente
		Encargado			
		Centro			
		Tiempo (año y mes)			
EC 5: Muestra información sobre los contratos archivados	Muestra la cantidad de contratos por centro, estado legal, encargado y tiempo dado el centro, el tiempo (año y mes), el estado legal y el encargado	Estado legal	Cantidad de contratos	El sistema muestra todas las variables por las cuales se puede filtrar la información para obtener la cantidad contratos	El usuario se autentica, selecciona el A.A.G Mercadotecnia / A.A PTTC / L.T Contrato / Selecciona la carpeta deseada y el reporte, la vista o dashboard correspondiente
		Encargado			
		Centro			
		Fecha			

Listas de Chequeo

La lista de chequeo es la herramienta más fácil de implementar y una de las más efectivas para el control de los procesos. Su principal beneficio es asegurar el cumplimiento de las actividades o procesos, sin embargo, lo más importante es que se asegure el cumplimiento de los aspectos que afectan la calidad del resultado final del proceso o actividad. Deja trazabilidad de la actividad realizada para evaluaciones necesarias. Se concentra en los aspectos críticos del proceso que puedan generar resultados no esperados. Asegura la secuencia de ejecución de las tareas o actividades. (Alzate, 2015)

En el presente trabajo se aplicarán las listas de chequeo a los artefactos correspondientes a la obtención de requisitos, con el objetivo de comprobar el cumplimiento de las actividades, estos documentos son: Especificación de casos de uso, Especificación de requisitos de software, Reglas de negocio y transformación y Salidas del sistema. Las listas de chequeo responderán a los siguientes elementos:

- ❖ Indicador a evaluar: es el indicador a evaluar en las diferentes secciones

Capítulo 3: Implementación y Prueba de la solución

- ❖ Peso: describe la criticidad del indicador a evaluar
- ❖ Evaluación: forma de evaluar el indicador (1 si existe alguna dificultad, 0 en caso contrario)
- ❖ No procede: especifica la necesidad de no evaluar el indicador en ese caso
- ❖ Cantidad de elementos afectados: especifica la cantidad de no conformidades (NC) detectados en el indicador

Los indicadores a evaluar en las listas de chequeo estarán distribuidos en las siguientes secciones:

- ❖ Estructura del documento: contiene los aspectos definidos por el Expediente de proyecto
- ❖ Veracidad de la información: contiene los indicadores a evaluar respecto a la información mostrada
- ❖ Semántica del documento: contiene los indicadores a evaluar respecto a la ortografía y redacción

3.4.2 Resultados de las pruebas

A continuación, se muestran los resultados de las pruebas realizadas al MD para el Departamento de Mercadotecnia perteneciente a la DTTC.

Pruebas unitarias y de integración

Las pruebas unitarias y de integración fueron aplicadas por miembros del equipo de desarrollo y especialistas del centro DATEC. En las mismas se verificó el correcto diseño y funcionamiento de cada uno de los componentes por individual y luego su integración como un todo. Dichas pruebas arrojaron las siguientes NC, las que ya fueron resueltas:

- ❖ Redefinir los nombres de los id de las dimensiones de manera que cumpla con los estándares de codificación
- ❖ Redefinir el nombre de la BD por uno que se ajuste al departamento
- ❖ Colocarle una salida de errores a excel al componente filtrar de las ETL
- ❖ Colocarle una salida de errores a excel a los trabajos o jobs
- ❖ Redefinir el nombre del esquema hecho por mart_mercadotecnia

Pruebas del sistema

Las pruebas de sistema fueron realizadas por el equipo de trabajo con la utilización de la herramienta CP. Se verificó que la solución cumple con los objetivos propuestos, basándose en el resultado arrojado por cada CP diseñado, teniendo en cuenta los errores cometidos durante la implementación. La tabla 14 muestra la relación con la cantidad de NC detectadas en cada iteración, las que ya fueron resueltas, y la complejidad de las mismas.

Capítulo 3: Implementación y Prueba de la solución

Tabla.12: Resultados de las pruebas de sistema

Iteraciones	Cantidad de NC	Complejidad			NC Resueltas
		Alta	Media	Baja	
Iteración 1	7	3	2	2	7
Iteración 2	4	1	3	0	4
Iteración 3	0	0	0	0	0

Pruebas de aceptación

Después de concluida la implementación del MD para el Departamento de Mercadotecnia perteneciente a la DTTC, fue presentado al cliente, el Ing. Yunier Vega Rodríguez, especialista superior y Metodólogo del Departamento de Mercadotecnia, para su revisión; quedando satisfecho con los resultados obtenidos y dando su aprobación final. De esta manera se obtuvo la carta de aceptación, la que se puede apreciar en los anexos (Ver [Anexo #2](#)).

La figura 22 muestra un resumen de las NC detectadas en los niveles de pruebas: unidad e integración, sistema y aceptación, en las que se detectaron un total de, cinco, 11 y cero NC, respectivamente. Estas NC fueron resueltas satisfactoriamente.

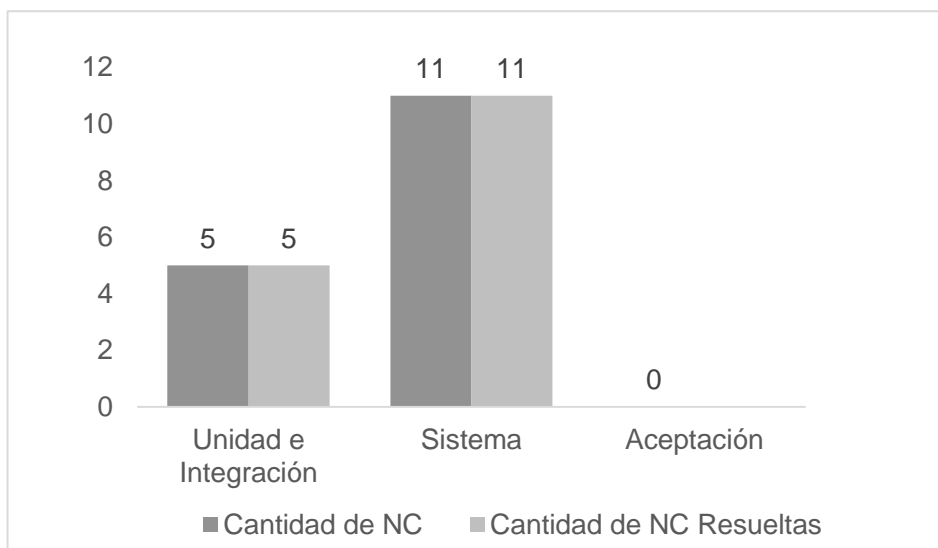


Fig.22: Resumen de NC detectadas por tipos de pruebas (Elaboración propia)

Capítulo 3: Implementación y Prueba de la solución

Listas de chequeo

Luego de haber aplicado las listas de chequeo, fueron detectadas nueve NC correspondientes al artefacto “Especificación de requisitos de software”. La figura 23 muestra el comportamiento de los indicadores identificados para dicho artefacto, en la que fueron definidos, para la Estructura del documento, un total de tres indicadores. De estos, dos fueron críticos y, además, se detectó una NC. En cuanto a la Veracidad de la información, se identificaron cuatro indicadores, de ellos, dos críticos y tres NC, mientras que, para la Semántica del documento, se identificaron un total de tres indicadores como críticos y 10 NC.

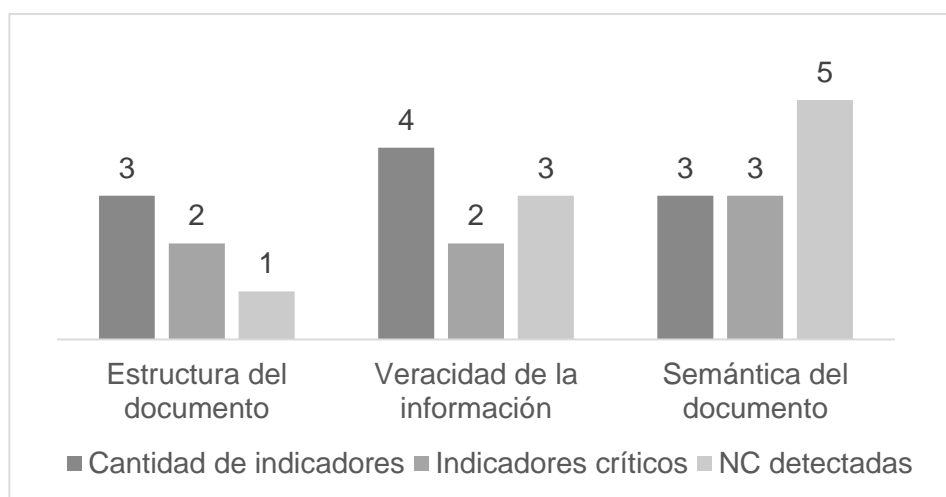


Fig.23: Comportamiento de los indicadores para la Especificación de requisitos (Elaboración propia)

3.5 Conclusiones del capítulo

La implantación del subsistema de almacenamiento permitió obtener una BD con cuatro tablas de hechos y 19 tablas de dimensiones que responden a los estándares de codificación definidos. El subsistema de integración permitió la realización de 23 transformaciones que garantizaron la limpieza y estandarización de los datos y tres trabajos que permiten la ejecución de las acciones de forma continua. Para la implementación del subsistema de visualización fueron creados los cubos OLAP, que permitieron el acceso multidimensional a los datos. Se realizó la arquitectura de información del MD para una mayor organización en cuanto a la estructura de las Áreas de Análisis. Fueron elaboradas las vistas de análisis, los reportes operacionales y dashboards, permitiendo mostrar la información a los usuarios finales mediante tablas, listas y gráficos, de forma amena e intuitiva. Después de realizada la implementación de los tres subsistemas, se obtuvo como resultado un MD poblado que posee todas las características definidas por el cliente. Este, fue sometido a un conjunto de pruebas que arrojaron cinco NC asociadas a unidad e integración y 11 NC a pruebas de sistema; todas resueltas. Posteriormente, se obtuvo la carta de aceptación que comprueba la conformidad con el producto.

Conclusiones Generales

Al finalizar el proceso de desarrollo del MD para el Departamento de Mercadotecnia perteneciente a la DTTC se puede afirmar que se le ha dado cumplimiento de forma satisfactoria al objetivo general del presente trabajo de diploma, obteniéndose los resultados que se esperaban. Es por ello que se llega a las siguientes conclusiones:

- ❖ El estudio de los fundamentos teóricos permitió comprender mejor el tema tratado. Se seleccionó para guiar el proceso de desarrollo, la Metodología para el desarrollo de proyectos de AD, la que se adapta a las características de trabajo específicas del centro DATEC, lo que permite un desarrollo más ágil. La selección de las herramientas basadas en software libre permitió la correcta implementación de la solución con interfaces amigables para el usuario.
- ❖ El proceso de entrevistas con el cliente permitió definir los requisitos de la solución, identificándose así 16 RI agrupados en cuatro CUI, dos RF agrupados en dos CUF, 13 RNF y 26 RN. Se definió la arquitectura de la solución identificándose tres subsistemas. El subsistema de almacenamiento permitió identificar cuatro tablas de hechos y 19 tablas de dimensiones dando paso al diseño del modelo lógico de datos que permitió la implementación del modelo físico. El subsistema de integración permitió definir las estrategias que deben seguirse para el desarrollo de los procesos de integración de datos, en el que fueron realizadas 23 transformaciones y tres trabajos. El subsistema de visualización permitió diseñar los cubos OLAP, las vistas de análisis, los reportes operacionales y dashboard que fueron implementados como parte de la solución.
- ❖ Una vez culminado el proceso de implementación se obtuvo como resultado un MD poblado. El producto fue sometido a un conjunto de pruebas, arrojando como resultado final un total de 16 NC distribuidas en sus diferentes niveles. Estas NC fueron resueltas, garantizando así, el correcto funcionamiento del mercado y la obtención de la carta de aceptación por parte del cliente.

Recomendaciones

Independientemente de que se hayan alcanzado los objetivos trazados al inicio de la investigación se recomienda:

- ❖ Adicionar nuevos requerimientos de información que pueda necesitar el cliente a fin de satisfacer sus necesidades.
- ❖ Desarrollar nuevos Mercados de Datos para los demás departamentos, con el objetivo de apoyar el proceso de toma de decisión en la Dirección de Transferencia de Tecnología y Conocimiento.

Referencias Bibliográficas

1. **Alzate, Fernando V. 2015.** CGE (Centro de Gestión Empresarial). *Cómo elaborar listas de chequeo*. [En línea] 2015. [Citado el: 18 de mayo de 2017.] <http://iso9001-calidad-total.com/como-elaborar-listas-de-chequeo/>.
2. **Anguiano, Jorge Daniel M. 2014.** IBM developerWorks. [En línea] 30 de 06 de 2014. [Citado el: 21 de 02 de 2017.] https://www.ibm.com/developerworks/ssa/data/library/tipos_bases_de_datos/.
3. **Azán Basallo, Yasser, Díaz Estrada, Anay y González Gómez, Salvador. 2009.** Una experiencia en integración de aplicaciones empresariales. [En línea] 2009. [Citado el: 17 de mayo de 2017.] <http://www.redalyc.org/pdf/3783/378343637001.pdf>.
4. **Bernabeu, Ricardo Dario. 2010.** Metodología propia para la construcción de un Data Warehouse. [En línea] 2010. [Citado el: 16 de marzo de 2017.] <http://www.dataprix.com/category/business-intelligence/business-intelligence/data-mart>.
5. **Cabello, Victoria Nevado. 2001.** *INTRODUCCION A LAS BASES DE DATOS RELACIONALES*. 2001.
6. **Chinchilla, Arley Ricardo. 2011.** 3, julio-septiembre de 2011, Vol. 24.
7. **Cuéllar M, Guillermo A. 2014.** DATA WAREHOUSE, ASPECTOS TÉCNICOS, CARACTERÍSTICAS, USOS, BENEFICIOS, COMPONENTES, HERRAMIENTAS OLAP. [En línea] 23 de enero de 2014. [Citado el: 17 de noviembre de 2016.] <http://fceca.unicauca.edu.co/old/datawarehouse.htm>.
8. **Díaz, Josep Curto. 2010.** *Introducción al Business Intelligence*. 2010.
9. **Elósegui, Tristán. 2014.** *¿Qué es y para qué sirve un dashboard?* [En línea] 27 de octubre de 2014. [Citado el: 18 de mayo de 2017.] <http://tristanelosegui.com/2014/10/27/que-es-y-para-que-sirve-un-dashboard/>.
10. **Espinosa, Roberto. 2010.** Descubriendo el Business Intelligence. [En línea] 15 de julio de 2010. [Citado el: 11 de abril de 2017.] <https://churriwifi.wordpress.com/2010/07/15/17-4-reporting-en-pentaho-con-pentaho-report-designer-otras-posibilidades-de-reporting-birt-y-jasperreports/>.
11. **Globe. 2012.** *Pruebas funcionales y casos de prueba*. [En línea] enero de 2012. [Citado el: 18 de mayo de 2017.] <https://www.globetesting.com/2012/01/pruebas-funcionales-y-casos-de-prueba/>.
12. **González, Hernández, Yanisbel. 2013.** Serie Científica de la Universidad de las Ciencias Informáticas. [En línea] 2013. <http://publicaciones.uci.cu>.
13. **Gravitar. 2016.** Gravitar Información sin límites. [En línea] 2016. [Citado el: 24 de noviembre de 2016.] <http://gravitar.biz/pentaho/>.

14. **Group, E-Cultura. 2015.** El pensante educación. *Ventajas y desventajas de las bases de datos.* [En línea] 18 de junio de 2015. [Citado el: 12 de mayo de 2017.] <https://educacion.elpensante.com/ventajas-y-desventajas-de-las-bases-de-datos/>.
15. **Herrera, Carlos Arturo. 2012.** SlideShare. [En línea] 15 de mayo de 2012. [Citado el: 24 de abril de 2017.]
16. **Inmon, W. H. 2005.** *Building the Data Warehouse.* Fourth Edition. s.l. : Wiley Publishing, Inc., 2005. págs. 29-130.
17. **Kamber, Micheline y Han, Jiawei. 2006.** *Data Mining Concepts and Techniques.* 2006.
18. **Kimball, Ralph. 2006.** *The Data warehouse Lifecycle Toolkit.* 2006.
19. **Kimball, Ralph, Ross Margy. 2013.** *The Data Warehouse Toolkit.* Third Edition. 2013.
20. **Martínez, Rafael. 2010.** PostgreSQL. [En línea] 2010. [Citado el: 23 de noviembre de 2016.] http://www.postgresql.org.es/sobre_postgresql.
21. **Microsoft. 2016.** SQL Server 2016 and later. *Vistas.* [En línea] 2016. [Citado el: 7 de junio de 2017.] <https://msdn.microsoft.com/es-es/library/ms190174.aspx>.
22. **Navathe, Elmasri. 2015.** *Fundamentals of Database Systems.* 2015.
23. **Pentaho. 2016.** Pentaho. [En línea] 2016. [Citado el: 24 de noviembre de 2016.] <http://www.pentaho.com/product/data-integration#data-integration->.
24. **PMOinformatica. 2015.** PMOinformatica.com. [En línea] 6 de 05 de 2015. [Citado el: 22 de 02 de 2017.] <http://www.pmoinformatica.com/2015/05/requerimientos-no-funcionales-ejemplos.html>.
25. **PostgreSQL. 2016.** [En línea] 2016. [Citado el: 23 de noviembre de 2016.] <https://www.postgresql.org/ftp/pgadmin3/>.
26. **Rojas, Pol Bartomeus. 2009.** *Estudio comparativo de bases de datos analíticas.* Barcelona : s.n., 2009.
27. **Saffirio, Mario. 2011.** Consultoría BPM y TI. [En línea] 20 de 08 de 2011. [Citado el: 22 de 02 de 2017.] <https://msaffirio.com/2011/08/20/reglas-de-negocio-business-rules/>.
28. **Schiefer, Josef, List, Beate y Bruckner, Robert. 2002.** *A HOLISTIC APPROACH FOR MANAGING REQUIREMENTS OF DATA WAREHOUSE SYSTEMS.* In *Eighth Americas Conference on Information Systems.* Vienna University of Technology. 2002.
29. **Sinnexus. 2016.** *Busines Intelligence Informatica estrategica.* [En línea] Copyright 2007 – 2016, 2016. [Citado el: 18 de octubre de 2016.] http://www.sinnexus.com/business_intelligence/datawarehouse.aspx.
30. **Sorensen, Kasper. 2012.** DataCleaner. [En línea] 2012. [Citado el: 23 de noviembre de 2016.] <https://datacleaner.org/news>.

31. **Targetware. 2016.** Software.com.mx. [En línea] 2016. [Citado el: 23 de noviembre de 2016.] <http://www.software.com.mx/p/visual-paradigm-para-uml>.
32. **Valdéz, Damián Pérez. 2007.** ¿Qué son las bases de datos? [En línea] 26 de octubre de 2007. [Citado el: 29 de marzo de 2017.] <http://www.maestrosdelweb.com/principiantes/%C2%BFque-son-las-bases-de-datos/>.

Bibliografía

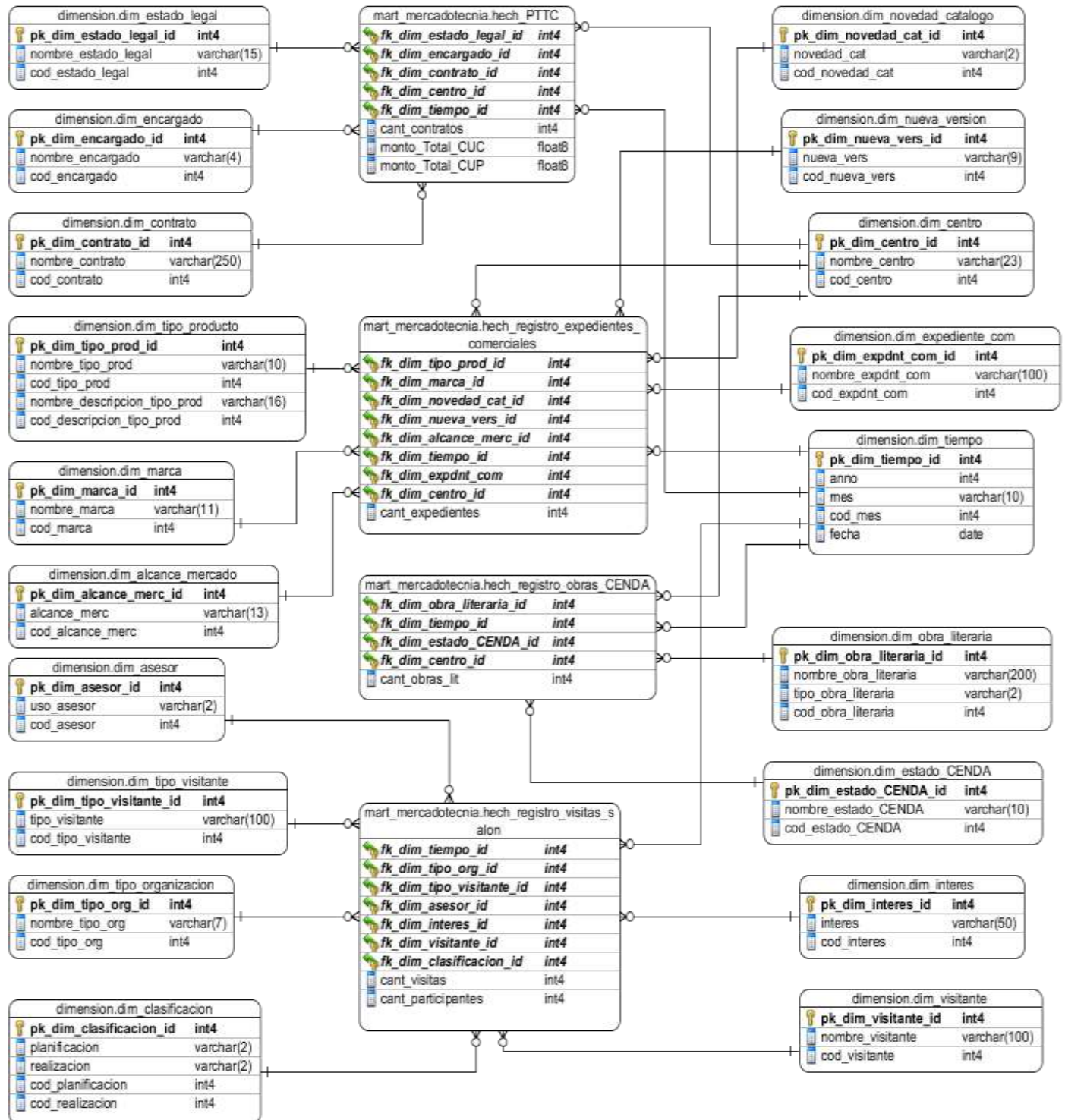
1. **Alzate, Fernando V. 2015.** CGE (Centro de Gestión Empresarial). *Cómo elaborar listas de chequeo*. [En línea] 2015. [Citado el: 18 de mayo de 2017.] <http://iso9001-calidad-total.com/como-elaborar-listas-de-chequeo/>.
2. **Anguiano, Jorge Daniel M. 2014.** IBM developerWorks. [En línea] 30 de 06 de 2014. [Citado el: 21 de 02 de 2017.] https://www.ibm.com/developerworks/ssa/data/library/tipos_bases_de_datos/ .
3. **Azán Basallo, Yasser, Díaz Estrada, Anay y González Gómez, Salvador. 2009.** Una experiencia en integración de aplicaciones empresariales. [En línea] 2009. [Citado el: 17 de mayo de 2017.] <http://www.redalyc.org/pdf/3783/378343637001.pdf>.
4. **Bernabeu, Ricardo Dario. 2010.** Metodología propia para la construcción de un Data Warehouse. [En línea] 2010. [Citado el: 16 de marzo de 2017.] <httpwww.dataprix.com/categorybusiness-intelligencebusiness-intelligencedata-mart>.
5. **Cabello, Victoria Nevado. 2001.** *INTRODUCCION A LAS BASES DE DATOS RELACIONALES*. 2001.
6. **Chinchilla, Arley Ricardo. 2011.** 3, julio-septiembre de 2011, Vol. 24.
7. **Cuéllar M, Guillermo A. 2014.** DATA WAREHOUSE, ASPECTOS TÉCNICOS, CARACTERÍSTICAS, USOS, BENEFICIOS, COMPONENTES, HERRAMIENTAS OLAP. [En línea] 23 de enero de 2014. [Citado el: 17 de noviembre de 2016.] <http://fccea.unicauca.edu.co/old/datawarehouse.htm>.
8. **Díaz, Josep Curto. 2010.** *Introducción al Business Intelligence*. 2010.
9. **Elósegui, Tristán. 2014.** *¿Qué es y para qué sirve un dashboard?* [En línea] 27 de octubre de 2014. [Citado el: 18 de mayo de 2017.] <http://tristanelosegui.com/2014/10/27/que-es-y-para-que-sirve-un-dashboard/>.
10. **Espinosa, Roberto. 2010.** Descubriendo el Business Intelligence. [En línea] 15 de julio de 2010. [Citado el: 11 de abril de 2017.] <https://churriwifi.wordpress.com/2010/07/15/17-4-reporting-en-pentaho-con-pentaho-report-designer-otras-posibilidades-de-reporting-birt-y-jasperreports/>.
11. **ETL-Tools.Info. 2016.** Business Intelligence - Almacenes de Datos - ETL. [En línea] 2016. [Citado el: 24 de noviembre de 2016.] http://etl-tools.info/es/bi/proceso_etl.htm.
12. **Globe. 2012.** *Pruebas funcionales y casos de prueba*. [En línea] enero de 2012. [Citado el: 18 de mayo de 2017.] <https://www.globetesting.com/2012/01/pruebas-funcionales-y-casos-de-prueba/>.
13. **González, Hernández , Yanisbel. 2013.** Serie Científica de la Universidad de las Ciencias Informáticas. [En línea] 2013. <http://publicaciones.uci.cu>.

14. **Gravitar. 2016.** Gravitar Información sin límites. [En línea] 2016. [Citado el: 24 de noviembre de 2016.] <http://gravitar.biz/pentaho/>.
15. **Group, E-Cultura. 2015.** El pensante educación. *Ventajas y desventajas de las bases de datos*. [En línea] 18 de junio de 2015. [Citado el: 12 de mayo de 2017.] <https://educacion.elpensante.com/ventajas-y-desventajas-de-las-bases-de-datos/>.
16. **Herrera, Carlos Arturo. 2012.** SlideShare. [En línea] 15 de mayo de 2012. [Citado el: 24 de abril de 2017.]
17. **Inmon, W. H. 2005.** *Building the Data Warehouse*. Fourth Edition. s.l. : Wiley Publishing, Inc., 2005. págs. 29-130.
18. **Inmon, William. 1999.** *Building the Operational Data Store*. 1999.
19. **Kamber, Micheline y Han, Jiawei. 2006.** *Data Mining Concepts and Techniques*. 2006.
20. **Kimball, Ralph. 2006.** *The Data warehouse Lifecycle Toolkit*. 2006.
21. **Kimball, Ralph, Ross Margy. 2013.** *The Data Warehouse Toolkit*. Third Edition. 2013.
22. **Martínez, Rafael. 2010.** PostgreSQL. [En línea] 2010. [Citado el: 23 de noviembre de 2016.] http://www.postgresql.org.es/sobre_postgresql.
23. **Microsoft. 2016.** SQL Server 2016 and later. *Vistas*. [En línea] 2016. [Citado el: 7 de junio de 2017.] <https://msdn.microsoft.com/es-es/library/ms190174.aspx>.
24. **Molina, Alonso de. 2015.** ESAN. [En línea] 2015. [Citado el: 24 de noviembre de 2016.] <http://www.esan.edu.pe/apuntes-empresariales/2015/07/20-herramientas-inteligencia-negocios-debes-conocer/>.
25. **Navathe, Elmasri. 2015.** *Fundamentals of Database Systems*. 2015.
26. **Pentaho. 2016.** Pentaho. [En línea] 2016. [Citado el: 24 de noviembre de 2016.] <http://www.pentaho.com/product/data-integration#data-integration->.
27. **Philip, Kotler y Armstrong, Gary. 1996.** *Mercadotecnia*. sexta. s.l. : Prentice Hall, 1996.
28. **PMOinformatica. 2015.** PMOinformatica.com. [En línea] 6 de 05 de 2015. [Citado el: 22 de 02 de 2017.] <http://www.pmoinformatica.com/2015/05/requerimientos-no-funcionales-ejemplos.html>.
29. **PostgreSQL. 2016.** [En línea] 2016. [Citado el: 23 de noviembre de 2016.] <https://www.postgresql.org/ftp/pgadmin3/>.
30. **Rojas, Pol Bartomeus. 2009.** *Estudio comparativo de bases de datos analíticas*. Barcelona : s.n., 2009.
31. **Rouse, Margaret. 2015.** Search Data Center. [En línea] 2015. [Citado el: 23 de noviembre de 2016.] <http://searchdatacenter.techtarget.com/es/definicion/Administrador-de-base-de-datos-DBA>.
32. **Saffirio, Mario. 2011.** Consultoría BPM y TI. [En línea] 20 de 08 de 2011. [Citado el: 22 de 02 de 2017.] <https://msaffirio.com/2011/08/20/reglas-de-negocio-business-rules/>.


33. **Sánchez, Jorge. 2016.** *Introducción a los Sistemas Gestores de Bases de Datos.* 2016.
34. **Schiefer, Josef, List, Beate y Bruckner, Robert. 2002.** *A HOLISTIC APPROACH FOR MANAGING REQUIREMENTS OF DATA WAREHOUSE SYSTEMS.* In *Eighth Americas Conference on Information Systems.* Vienna University of Technology. 2002.
35. **Sinnexus. 2016.** *Business Intelligence Informatica estrategica.* [En línea] Copyright 2007 – 2016, 2016. [Citado el: 18 de octubre de 2016.] http://www.sinnexus.com/business_intelligence/datawarehouse.aspx.
36. **Software, Testiando. 2013.** *Testiando Software Asegurando la calidad del software. Casos de Uso vs. Casos de Prueba.* [En línea] 9 de septiembre de 2013. [Citado el: 18 de mayo de 2017.] <https://testeandosoftware.com/casos-de-uso-vs-casos-de-prueba/>.
37. **Sorensen, Kasper. 2012.** *DataCleaner.* [En línea] 2012. [Citado el: 23 de noviembre de 2016.] <https://datacleaner.org/news>.
38. **Targetware. 2016.** *Software.com.mx.* [En línea] 2016. [Citado el: 23 de noviembre de 2016.] <http://www.software.com.mx/p/visual-paradigm-para-uml>.
39. **Urquizu, Ramel Vitier. 2013.** [En línea] 2013. [Citado el: 08 de noviembre de 2016.] <http://catalogoenlinea.uci.cu/cgi-bin/koha/opac-detail.pl?biblionumber=12991>.
40. **Valdéz, Damián Pérez. 2007.** *¿Qué son las bases de datos?* [En línea] 26 de octubre de 2007. [Citado el: 29 de marzo de 2017.] <http://www.maestrosdelweb.com/principiantes/%C2%BFque-son-las-bases-de-datos/>.

Anexos

Anexo #1: Modelo de datos ampliado



Anexo #2: Carta de Aceptación



Universidad de las Ciencias
Informáticas

**Dirección de Transferencia de Tecnologías y
Conocimientos (DTTC)**

La Habana, 7 de junio de 2017
"Año 59 de la Revolución"

**ACTA DE ACEPTACIÓN DE LA APLICACIÓN MERCADO DE DATO PARA EL DEPARTAMENTO
DE MERCADOTECNIA PERTENECIENTE A LA DTTC**

La Dirección de Transferencia de Tecnología y Conocimiento (DTTC), representado por Yunier Vega Rodríguez, portador de carné de identidad 85032512422 en su condición de Especialista Superior y Metodólogo del Departamento de Mercadotecnia, facultado para este acto, que en lo sucesivo se le denominará como **PARTE CLIENTE**, por una parte; y por la otra, el equipo desarrollador del Mercado de Datos para el Departamento de Mercadotecnia perteneciente a la DTTC, representado en este acto por los estudiantes Midalis Rodríguez Rosado y Jorge Alfonso Torres, portadores de carné de identidad 94071426138 y 91082430266 respectivamente, quienes actúan en su condición de TESISISTAS, suficientemente facultados para este acto, acuerdan expresamente que:

La **PARTE CIENTE**, luego de concluir el proyecto Mercado de Datos para el Departamento de Mercadotecnia perteneciente a la DTTC, determina que el mismo se efectuó satisfactoriamente.

Comentarios:

Y para que así conste se suscribe la presente **ACTA** en La Habana a los 7 días del mes de junio de 2017.

Entrega: <u>Midalis Rodríguez Rosado</u> Cargo: <u>Estudiante</u> Firma: <u>[Firma]</u>	Entrega: <u>Jorge Alfonso Torres</u> Cargo: <u>Estudiante</u> Firma: <u>[Firma]</u>
---	---

Recibe: Yunier Vega Rodríguez
 Cargo: Especialista Superior
 Firma: [Firma]

Universidad de las Ciencias Informáticas. Centro de Soporte
Carretera a San Antonio Km 2 1/2. Torrens, Boyeros. Ciudad de La Habana, Cuba