

UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS

FACULTAD 3



Histórico de precios para el sistema Xedro-Apside

*Trabajo de diploma para optar por el título de
Ingeniero en Ciencias Informáticas*

Autor: Pablo Miguel Choy Oliva

Tutoras:

Ing. Olga Yarisbel Rojas Grass

Ing. Virtudes Milagro Figueredo Lara

*Ciudad de La Habana, Junio 2015
"Año 57 de la Revolución"*



*"Las inteligencias poco capaces se interesan
en lo extraordinario; las inteligencias poderosas
en las cosas ordinarias"*

Victor Hugo

Declaración de autoría

Declaración de autoría

Declaro ser el autor de la presente tesis y reconozco 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 2015.

Firma del tesista

Pablo Miguel Choy Oliva

Firma del tutor

Ing. Olga Yarisbel Rojas Grass

Firma del tutor

Ing. Virtudes Milagro Figueredo Lara

Dedicatoria

A mi papá y mi mamá por ser el motivo más grande de estar aquí y por el apoyo, confianza y amor que siempre me han brindado.

A mi hermano por ser guía y ejemplo para mí y estar siempre ahí cuando más lo necesité.

A mis abuelos por su amor y su confianza.

Agradecimientos

Con esta tesis termina una etapa en mi vida de esfuerzo y estudio, por este motivo quiero expresar mis agradecimientos:

Quiero dar un agradecimiento especial a mi papá Leonardo Choy Peña y mi mamá Gisela Rodríguez Escalona por ser incondicionales, por apoyarme siempre y enseñarme a ser una mejor persona. Gracias por enseñarme a luchar por alcanzar mis metas. A ellos le debo todo lo que soy y mi gran sueño es que nunca me falten.

A mi hermano Julio Cesar Brito Rodríguez por estar siempre a mi lado y por la confianza que siempre ha depositado en mí.

A mis abuelos por el cariño y la confianza que han depositado en mí durante todo los momentos importantes de mi vida y a mi familia en general por esperar lo mejor de mí en todo momento.

A mis tutoras Olga y Virtudes por ayudarme siempre a corregir los mínimos detalles de este trabajo, guiándome durante todo el proceso de desarrollo. A Yanet por asumir el rol de tutora y ayudarme siempre que la necesité. A mi tribunal y oponente por corregir cada detalle y estar dispuestos a ayudarme.

A todos mis amigos que estuvieron conmigo en estos cinco años de carrera en los buenos y los malos momentos en especial quiero mencionar a: Yasiel, Dito, Rolando, Abel Alexis, Yuleidis, Dachel, Dayana, Rayco, Abelito, Samper, Leo, Ivis, Alexander y dentro de esta familia quiero agradecer a Elizabeth por ser durante todos estos años mi amiga y haber entrado en mi vida para convertirse en algo especial para mí.

Por último agradecer a todos los que han contribuido de una forma u otra a lo largo de estos cinco años a mi preparación profesional y como una mejor persona, a mis compañeros, profesores, todo han puesto su granito de arena. A todos y cada uno de ustedes...

MUCHISIMAS GRACIAS.

Resumen

La solución propuesta es un mercado de datos cuyo objetivo fundamental es contribuir al análisis estadístico del histórico de precios en el área de mercadotecnia para la Empresa BK Import/Export. Con el fin de dar cumplimiento al objetivo propuesto, fue centralizada la información que se encuentra almacenada en los documentos Excel, cuyo análisis de manera independiente se torna difícil y complejo. Dicho mercado de datos brinda un espacio único donde la información se encuentra centralizada, consistente y disponible. Todo el conjunto de transformaciones que sufren los datos para ser estandarizados y posteriormente visualizados, se llevan a cabo a través de procesos de integración de datos y técnicas de Inteligencia de negocios, haciendo uso del conjunto de herramientas de la Suite Pentaho. Una vez concluidos dichos procesos, se obtiene como resultado un mercado de datos poblado y una aplicación web para la visualización de diferentes reportes que incluyen el análisis de los principales indicadores del área mercadotecnia.

Palabras Claves: centralizada, consistente, disponible, mercado de datos, transformaciones

INTRODUCCIÓN.....	1
CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA.....	4
Introducción.....	4
1.1 Estado actual de los almacenes y mercado de datos	4
1.2 Almacenes de Datos	6
1.2.1 Características del almacén de datos	6
1.3 Mercado de datos	9
1.3.1 Características de los mercados de datos	10
1.4 Metodologías existentes para el desarrollo de un Mercado de Datos.....	10
1.4.1 Propuesta de metodología para el desarrollo de Almacén de Datos en DATEC	11
1.5 Sistemas gestores de bases de datos	12
1.5.1 Sistema gestor de bases de datos PostgreSQL 9.1	13
1.5.1.1 Administrador de base de datos PgAdmin III 1.14	14
1.6 Sistemas de almacenamiento de datos	14
1.6.1 OLAP (Procesamiento Analítico en Línea).....	15
1.6.1.1 ROLAP (Procesamiento Analítico en Línea Relacional)	16
1.6.1.2 MOLAP (Procesamiento Analítico en Línea Multidimensional).....	17
1.6.1.3 HOLAP (Procesamiento Analítico en Línea Híbrido)	17
1.6.2 Valoración general.....	17
1.7 Modelo Entidad-relación y Modelo dimensional	18
1.7.1 Modelo Entidad-relación.....	18
1.7.2 Modelo dimensional.....	19
1.8 Herramientas de modelado	21
1.8.1 Herramienta de modelado Visual Paradigm 8.0.....	21
1.9 Herramientas para la integración de datos.....	22
1.9.1 Pentaho Data Integration 4.2.0.....	22
1.9.2 DataCleaner 1.5.4	22
1.10 Herramientas para la inteligencia de negocios	23
1.10.1 Pentaho Schema Workbench 3.2.1	23
1.10.2 Pentaho BI Server 4.5.0	23
CAPÍTULO 2: ANÁLISIS Y DISEÑO	25
Introducción.....	25
2.1 Necesidades del negocio.....	25
2.2 Especificación de los requisitos.....	25
2.2.1 Requisitos de información	25
2.2.2 Requisitos funcionales.....	26
2.2.3 Requisitos no funcionales.....	26
2.3 Reglas del negocio	27
2.4 Diagrama de casos de uso del sistema	28

2.4.1	Actores del sistema	28
2.4.2	Caso de uso de información	29
2.4.3	Caso de uso funcionales	30
2.5	Definición de la arquitectura de un mercado de datos	31
2.6	Diseño de la solución.....	32
2.6.1	Diseño del subsistema de integración.....	35
2.6.2	Diseño del subsistema de visualización.....	37
2.7	Política de respaldo y recuperación	38
2.8	Esquema de seguridad	38
	Conclusiones del capítulo	39
	CAPÍTULO 3 IMPLEMENTACIÓN Y VALIDACIÓN	40
	Introducción	40
3.1	Implementación del subsistema de almacenamiento	40
3.1.1	Esquemas y tablas	40
3.1.2	Restricciones y secuencias	41
3.2	Implementación del subsistema de integración	42
3.2.1	Implementación de las transformaciones.....	42
3.2.2	Implementación del trabajo	43
3.3	Implementación del subsistema de visualización	43
3.3.1	Cubos de datos	43
3.4	Pruebas.....	45
3.4.1	Pruebas funcionales	46
3.4.2	Pruebas de aceptación.....	47
3.4.3	Aplicación de la técnica de ladov	48
	Conclusiones del capítulo	50
	CONCLUSIONES GENERALES	51
	RECOMENDACIONES	52
	REFERENCIAS BIBLIOGRÁFICAS	53
	ANEXOS	55

Tabla 1. Comparación de los sistemas similares	5
Tabla 2. Diferencia entre AD y bases de datos.....	7
Tabla 3. Criterios comparativos de tipos de almacenamiento.....	18
Tabla 4. Descripción del CUI Obtener información de compra.....	29
Tabla 5 Descripción del CU Autenticar usuario	30
Tabla 6. Matriz BUS	34
Tabla 7. Permisos de los usuarios en la base de datos	38
Tabla 8. Permisos de los usuarios en la aplicación.....	39
Tabla 9. Niveles de satisfacción.....	49
Tabla 10. Resultados de las escalas de satisfacción	50

Figura 1. Caso de uso del sistema	28
Figura 2. Arquitectura del mercado de datos	32
Figura 3. Modelo de datos.....	35
Figura 4. Diseño de la transformación para la carga de los indicadores.....	36
Figura 5. Diseño de la transformación para la carga de los datos del hecho compra	36
Figura 6. Cubo compra	37
Figura 7. Estructura de datos	41
Figura 8. Transformación para la carga de la dimensión dim_país.....	42
Figura 9. Transformación para la carga de la dimensión dim_oferta	43
Figura 10. Transformación del trabajo general del mercado de datos Histórico de precios	43
Figura 11. Interfaz de la vista de análisis.....	44
Figura 12. Reporte obtener el precio por proveedor, día, mes y año dado un producto.....	45
Figura 13. Caso de prueba realizado al RI obtener el precio por proveedor, día, mes y año dado un producto	47
Figura 14. Descripción de las variables	47
Figura 15. Cuadro Lógico de ladov	48
Figura 16. Nivel de satisfacción	49
Figura 17. Página principal del documento de liberación del MD	55
Figura 18. Resultados de los elementos probados y su estado final	56
Figura 19. Firmas de los involucrados en el acta de liberación.....	56
Figura 20. Acta de aceptación	57
Figura 21. Cuestionario de los especialistas.....	58
Figura 22. Vista principal de la aplicación	59

Introducción

La informática es una ciencia que propone el tratamiento automatizado de la información utilizando dispositivos electrónicos y sistemas computacionales. La información se ha convertido en uno de los recursos más valiosos y la materia prima indispensable para el desarrollo, equilibrio y adaptación de las empresas, sectores y países; además es un recurso estratégico determinante de la competitividad y un sustento a los procesos de toma de decisiones. No basta con lograr la persistencia de los datos en el tiempo, es necesario su almacenamiento y procesamiento. En este sentido, se han alcanzado avances significativos en los repositorios de información, evolucionando desde un fichero de texto hasta grandes base de datos.

Con el fin de obtener la mayor información posible, las principales esferas de la sociedad en el mundo han aplicado la tecnología de los Almacenes de Datos (AD). Los mismos constituyen un instrumento eficaz para la dirección de las empresas, pues permiten realizar preguntas, consultas y análisis de los datos según sus necesidades, sin tener que acudir al personal informático de la institución. Particularmente los AD orientados al análisis de históricos de precios se han convertido en herramientas que ofrecen beneficios para llevar a cabo los diferentes procesos dentro de las organizaciones, constituyendo uno de los principales eslabones para la toma de decisiones.

En Cuba en materia de Inteligencia de negocios y AD, no se ha avanzado lo suficiente para explotar las funcionalidades que estas tecnologías pueden proporcionar a las organizaciones. A pesar de esto con el paso del tiempo las entidades se van sintiendo cada vez más atraídas por los beneficios asociados a las soluciones de Inteligencia de negocios. Las organizaciones cubanas precisan de una mayor automatización para mejorar el desempeño de sus tareas operativas pues en algunos casos existe poca homogeneización de la información disponible a este nivel.

Como parte de la estrategia del estado cubano surge la Universidad de las Ciencias Informáticas (UCI), institución que se caracteriza por su modelo de enseñanza docente productiva. La misma cuenta con varios centros de desarrollo de software entre ellos está el Centro de Informatización de Entidades (CEIGE) ubicado en la facultad 3.

El CEIGE informatiza los procesos referentes a la gestión empresarial por lo que como tarea recibe la informatización de los procesos de la Empresa BK Import/Export. La misma se dedica a la prestación de servicios de importación, exportación y comercialización con una estructura que permite asimilar los cambios que se originen como resultado de su acción y del entorno. Reconocida en el mercado mundial como importador de excelencia. En ella se maneja un gran número de productos y servicios de diversos proveedores que son almacenados en documentos Excel. Esto dificultan la integración de los principales reportes e históricos de precios y la selección de proveedores al iniciar el proceso de importación porque la mayor parte de la

información se encuentra en formato Excel, trayendo consigo lentitud en este proceso a partir de que la información es inconsistente, está descentralizada y no está disponible, afectando la toma de decisiones.

En correspondencia con lo planteado anteriormente se genera el siguiente **problema a resolver**: ¿Cómo contribuir a la centralización, consistencia y disponibilidad de la información para el análisis estadístico del histórico de precios para el Sistema Xedro-Apside?

Se define como **objeto de estudio** los AD, enmarcado en el **campo de acción** los Mercados de Datos (MD) en sistemas comerciales.

Objetivo general: Desarrollar un MD que permita la centralización, consistencia y disponibilidad de la información del histórico de precios en el Sistema Xedro-Apside.

Objetivos específicos:

OE1. Definir el marco teórico de la investigación mediante el estudio y el análisis de los principales referentes teóricos acerca de los MD para el desarrollo de la solución.

OE2. Realizar el análisis y diseño del MD para el histórico de precios del Sistema Xedro-Apside.

OE3. Realizar la implementación del MD para el histórico de precios del Sistema Xedro-Apside.

OE4. Validar la propuesta de solución mediante la aplicación de técnicas y pruebas.

Como **resultado esperado** se obtendrá un MD poblado, así como la capa de visualización de los datos que permitirá a la organización y análisis de precios sobre los productos para el proceso de toma de decisiones en el momento de realizar las importaciones.

Por el razonamiento antes realizado se arriba a la siguiente **idea a defender**: El desarrollo de un MD para el análisis estadístico del histórico de precio en el sistema Xedro-Apside contribuirá a la centralización, consistencia y disponibilidad de la información.

La presente investigación está guiada por **métodos científicos** los cuales son:

Métodos teóricos

Análisis histórico-lógico: Mediante el estudio del estado del arte de los Almacenes de Datos y los Mercados de Datos, se podrá conocer la evolución histórica y el desarrollo actual de los mismos, enfocado en los históricos de precios de la Empresa BK Import/Export.

Analítico-Sintético: Se utilizará para distinguir, extraer y unificar los elementos que forman parte del proceso de construcción del mercado de datos e indicadores relacionados con los históricos de precios de la Empresa BK Import/Export.

Método empírico

Entrevista: Se conocerá las funcionalidades que tiene el mercado de datos a desarrollar, a partir de las necesidades del cliente.

El trabajo está compuesto por 3 capítulos:

Capítulo 1: Fundamentación teórica

En este capítulo se definen conceptos fundamentales sobre los AD y los MD, con sus principales características, metas y elementos que los componen, se hacen referencia al desarrollo de los procesos de integración de datos. Se fundamenta la selección de la metodología, las herramientas y tecnologías que serán utilizadas.

Capítulo 2: Análisis y diseño

En este capítulo se hará un estudio preliminar del negocio con el objetivo de definir: reglas del negocio, casos de uso y su descripción, identificación de dimensiones, hechos y medidas. Se realiza el análisis del sistema a desarrollar, con el propósito de refinar y estructurar los requisitos obtenidos con anterioridad para facilitar su comprensión, preparación, modificación y mantenimiento.

Capítulo 3: Implementación y validación

En este capítulo se hace referencia a la implementación de la solución, abordando específicamente cómo se realiza la misma en el subsistema de almacenamiento, integración y visualización, teniendo en cuenta los requisitos y necesidades del negocio. Una vez terminado de implementar los subsistemas, se dará paso a la validación de la propuesta de solución mediante la aplicación de pruebas, para determinar que la solución a desarrollar cuente con la calidad requerida.

CAPÍTULO 1: Fundamentación teórica

Introducción

Durante el desarrollo del presente capítulo se detallan diferentes elementos teóricos sobre las tecnologías de AD y MD. Se analizan las diferentes metodologías existentes a nivel mundial, además de la definida en la universidad, haciendo la selección de la que regirá el proceso de desarrollo. Se define las herramientas que se utilizan para el proceso de desarrollo del producto, dejando plasmado los elementos relacionados con la elección de cada una de ellas. De igual forma se exponen las principales características del gestor de bases de datos seleccionado.

1.1 Estado actual de los almacenes y mercado de datos

Con el transcurso del tiempo y con el crecimiento de la información en distintas fuentes de datos, los directivos de las empresas se dieron cuenta de que era inevitable la utilización de los AD y las herramientas de Inteligencia de negocio (BI por sus siglas en inglés). Actualmente existen muchas corporaciones que manejan grandes cantidades de datos como son el caso de Adidas, Hewlett Packard y Sun, dichas empresas construyen sus sistemas operacionales con herramientas propietarias [1].

El uso de los AD se ha extendido a muchas esferas, ya no están concebidos solo en la economía. La salud, las telecomunicaciones y los sistemas bancarios son ejemplos fehacientes que demuestran que los AD han ido tomando cada día más fuerza y ya no pasarán desapercibidos como antes. Prestigiosos bancos del mundo utilizan AD para gestionar su información, algunos de ellos son: El Banco París de Francia, *European Central Bank de la Unión Europea*, BBVA considerado el segundo banco más grande de España [2].

En Cuba el uso de estos sistemas de almacenamiento de información, aún se consideran muy incipientes. No obstante, el estado cubano y el Ministerio de la Informática y las Comunicaciones, siguen haciendo esfuerzos y demostrando su interés en dar pasos sólidos hacia una futura utilización de esta tecnología por sus beneficios. Las entidades cubanas que por sus características y ventajas, emplean los MD para analizar el gran cúmulo de información, se destacan:

- ✓ La Dirección de Colaboración Económica del Ministerio de Comercio Exterior y la Inversión Extranjera utiliza un MD para almacenar la información referente a las inversiones extranjeras y colaboraciones económicas, que le permite garantizar la accesibilidad, calidad y el análisis de la información en apoyo a la toma de decisiones [3].
- ✓ El MD de la Dirección de Cuadros de la Administración Provincial de Artemisa almacena los datos manejados por parte de los directivos de la provincia, garantizando la disponibilidad de la información y facilitando el análisis de los datos para mejorar la toma de decisiones [4].

Capítulo 1: Fundamentación teórica

- ✓ La Oficina Nacional de Estadísticas (ONE), creada para garantizar la producción de estadísticas de calidad, utiliza un MD para llevar el registro de todas las instituciones del país y así almacenar, recuperar y presentar la información proveniente de un modelo estadístico, garantizando el proceso de toma de decisiones [5].

A partir del estudio realizado de las diferentes instituciones que utilizan MD se realizó un análisis comparativo, definiendo las principales características que contribuirán en el desarrollo de la propuesta de solución. A continuación se muestra una tabla comparativa que detalla las características antes mencionadas:

Tabla 1. Comparación de los sistemas similares

Mercados	Características			
	Modelo	Esquema	Metodología	Almacenamiento
MD1	Dimensional	Constelación de hechos	Kimball	OLAP
MD2	Dimensional	Estrella	Kimball	ROLAP
MD3	Dimensional	Estrella	Kimball	ROLAP

Leyenda:

MD1: La Dirección de Colaboración Económica del Ministerio de Comercio Exterior y la Inversión Extranjera.

MD2: El MD de la Dirección de Cuadros de la Administración Provincial de Artemisa.

MD3: La Oficina Nacional de Estadísticas (ONE).

Luego del análisis realizado se concluye que los sistemas no cuentan con las mismas características para tomarlas como base para el desarrollo de la propuesta de solución. El estudio permitió llegar a las siguientes conclusiones:

- ✓ El 100% de los MD fueron desarrollados utilizando el Modelo Dimensional.
- ✓ El 33.3% de los MD utilizaron como forma de representación de las tablas el esquema Constelación de hechos, mientras que 66.6% utilizó el esquema Estrella.
- ✓ El 100% de los MD fueron desarrollados empleando la metodología de desarrollo implementada por Kimball.
- ✓ El 33.3% de los MD utilizaron el modo de almacenamiento de datos Procesamiento Analítico en Línea (OLAP) y el 66.6% emplearon el Procesamiento Analítico en Línea Relacional (ROLAP).

Partiendo del análisis antes expuesto se propone desarrollar un MD que permita la centralización, consistencia y disponibilidad de la información para el análisis estadístico de los históricos de precios. Los MD antes mencionados no están en función del análisis de históricos de precios, pero algunas de las características evidenciadas se pueden tomar como base para el desarrollo de la propuesta de solución.

1.2 Almacenes de Datos

Debido a la gran cantidad de información que se maneja actualmente en los diferentes organismos y empresas del mundo, ha surgido la necesidad de digitalizarla en función de la manipulación eficiente de la misma. A pesar del esfuerzo realizado por los especialistas en la materia el volumen de datos sigue siendo cada vez mayor. Con el objetivo de darle solución a este problema ha surgido el término almacén de datos. Un almacén de datos es una gran colección de datos orientados a temas, integrados, no volátiles e históricos, que recoge información de múltiples sistemas fuentes u operacionales dispersos, y cuya actividad se centra en la toma de decisiones [6].

Un AD es un sistema computarizado para guardar registros cuya principal finalidad es almacenar información de manera que los usuarios puedan actualizarla y recuperarla. Una vez reunidos los datos de los sistemas fuentes se guardan durante mucho tiempo, permitiendo el acceso a datos históricos y proporcionando una mayor facilidad a la hora de realizar consultas en función de la toma de decisiones [7]. Otra definición de AD es: "una colección de información corporativa derivada directamente de los sistemas operacionales y de algunos orígenes de datos externos. Su propósito específico es soportar la toma de decisiones en un negocio, no las operaciones de un negocio" [8].

Un AD es "una colección de datos, orientados a hechos relevantes del negocio, integrados, que incluyen el tiempo como característica importante de referencia y no volátiles para el proceso de toma de decisiones" [9]. También se puede citar lo planteado por otro reconocido autor como Ralph Kimball (2002), quien define a un Almacén de Datos como: "los AD son una copia de los datos de la transacción estructurados específicamente para la pregunta y el análisis". Como se ha podido ver, existen disímiles autores y estudiosos del tema que definen de forma muy particular lo que para ellos representa un almacén de datos. Sin embargo, no deja de ser un hecho que un AD es un sistema para la colección de datos con respecto a temas específicos, en los que la información va a prevalecer de una manera efectiva a lo largo del tiempo, y su principal razón de existir es la de brindar ayuda a una organización o empresa con respecto a la toma de decisiones.

1.2.1 Características del almacén de datos

El AD posee un grupo de características que los distinguen y que están estrechamente relacionadas con su estructura y funcionamiento. Estas características son [10]:

✓ Temático

Porque los datos están almacenados por materias o temas. Estos se organizan desde la perspectiva del usuario final, mientras que en las base de datos operacionales se organizan desde la perspectiva de la aplicación, con vistas a lograr una mayor eficiencia en el acceso a los datos.

Capítulo 1: Fundamentación teórica

✓ Integrado

Porque todos los datos almacenados en el Almacén de Datos están integrados. Las base de datos operacionales orientadas hacia las aplicaciones fueron creadas sin pensar en su integración, por lo que un mismo tipo de dato puede ser expresado de distintas maneras en dos base de datos operacionales diferentes (Por ejemplo, para representar el sexo: 'Femenino' y 'Masculino', 'F' y 'M' o '0' y '1').

✓ No volátil

Porque únicamente hay dos tipos de operaciones en el Almacén de Datos: la carga de los datos procedentes de los entornos operacionales (carga inicial y carga periódica) y la consulta de los mismos. La actualización de datos no forma parte de la operativa normal de un AD, ya que no son muy frecuentes los cambios en ellos y por ende se puede mantener por largos períodos de tiempo la información.

✓ Histórico

El tiempo debe estar presente en todos los registros contenidos en un Almacén de Datos. Las base de datos operacionales contienen los valores actuales de los datos. Un AD es una serie de instantáneas en el tiempo tomadas periódicamente, que permiten mantener y referenciar información.

Diferencia entre AD y Base de datos

A partir de lo anterior se pueden establecer diferencias sustanciales entre las base de datos operacionales y los AD, las cuales apuntan hacia la utilización de los AD como solución a muchos de los problemas que no se resuelven con las base de datos. La siguiente tabla muestra de manera sintetizada estas diferencias [11].

Tabla 2. Diferencia entre AD y base de datos

Base de Datos Operacional	Almacén de Datos
Datos operacionales	Datos del negocio para la información
Orientado a la aplicación	Orientado a sujeto
Actual	Actual e histórico
Detallada	Detallada y resumida
Cambia continuamente	Estable

Componentes de un Almacén de Datos

La composición de un Almacén de Datos está definida por una serie de elementos que muestran de manera general el ambiente de estos. Es irrefutable que la construcción de un AD está justificada por la necesidad en específico de cada empresa u organización que lo requiera. A continuación se hace referencia a los componentes de un AD.

Sistema de fuentes operacionales

Con el objetivo de gestionar las transacciones que se hacen diariamente en las empresas aparecen los sistemas de fuentes operacionales. Dichas transacciones pueden ser almacenadas con distintos formatos que suelen ser vistos de distintas maneras, desde una base de datos relacional hasta ficheros como excel, xml, dbf, texto plano.

Por no tenerse control o casi ningún control sobre el volumen y formato de los datos de estas fuentes, estos sistemas están localizados fuera del depósito central. Estos componentes tienen una importancia vital a la hora de evaluar el rendimiento y la disponibilidad de la información y funcionan con el objetivo de ir realizando salvadas a la información que se encargan de gestionar.

Área de procesamiento

El sistema que se encuentra entre las fuentes de datos y el AD es llamado área de procesamiento. Tiene entre sus objetivos facilitar la extracción de los datos desde las fuentes orígenes. El área de procesamiento realiza lo que se conoce como limpieza de datos con el fin de mejorarlos y se utiliza para acceder con un nivel detallado a la información que no está contenida en el AD.

El área de procesamiento es el “área donde los datos extraídos por los procesos de Extracción Transformación y Carga (ETL) son transformados y limpiados antes de ser cargados dentro del Almacén de Datos”. Finalmente, es el componente de un almacén de datos donde los datos son almacenados temporalmente y que realiza un conjunto de procesos como los de Extracción, Transformación y Carga (ETL) [12].

Extracción, Transformación y Carga de los datos

Al proceso de analizar la información consolidada en un AD y que implica actividades de extracción de diversas fuentes de datos, transformación de la información necesaria y finalmente su carga, se le denomina proceso ETL. Otro aspecto importante y que provee los procesos ETL es la limpieza de los datos, que es la habilidad de chequear, filtrar y corregir los errores que puedan ser encontrados en los datos. Entre los principales productores actuales de herramientas ETL se destacan Oracle con *OracleWarehouseBuilder*, Microsoft con *MicrosoftwithDataTransformation Services* e IBM con *DataWareHouse Center*. A continuación se describe más detalladamente dicho proceso [12]:

Capítulo 1: Fundamentación teórica

- ✓ Extracción: Acción de obtener la información deseada a partir de los datos almacenados en fuentes externas.
- ✓ Transformación: Cualquier operación realizada sobre los datos para que puedan ser cargados en el Almacén de Datos o se puedan migrar de este a otra base de datos.
- ✓ Carga: Consiste en almacenar los datos en la base de datos final.

El flujo de trabajo que representa a los procesos de ETL está compuesto por varias funcionalidades que a continuación se mencionan [12]:

- ✓ Identificación de la información relevante a los recursos externos.
- ✓ Extracción de la información identificada.
- ✓ Transportación de la información al área de procesamiento.
- ✓ Transformación de la información proveniente de múltiples recursos a un formato común.
- ✓ Limpieza de los datos entrados a la base de datos.
- ✓ Propagación de los datos al Almacén de Datos y actualización del MD.

Área de presentación

Se ha definido como área de presentación al componente de un AD donde los datos significativos para la toma de decisiones están organizados, almacenados y disponibles para la consulta o reporte que deseen hacer los usuarios. Según Kimball en esta área las estructuras aparecen como esquemas dimensionales llamados esquemas en estrella [13].

Herramientas de acceso a datos

Las herramientas de acceso a datos son vistas como la variedad de capacidades que se les proveen a los usuarios del negocio para la toma de decisiones. Básicamente son herramientas que permiten la consulta del área de presentación de un AD. Pueden ser herramientas de consultas muy simples o tan complejas y sofisticadas como una aplicación de modelado o de Minería de Datos [14].

1.3 Mercado de datos

A través del análisis de diversas fuentes existentes sobre la tecnología de AD, muchos autores hacen referencia al término de MD, cuando se refieren al proceso de almacenar información. Josep Curto un estudioso en el ámbito de los AD y MD, define a un MD como: “un subconjunto de datos del Almacén de Dato, con el objetivo de responder a un determinado análisis, función o necesidad y con una población de usuarios específica” [15]. Otros autores definen un MD como: “bases de datos multidimensionales orientadas a una materia específica” [8].

Los MD son un subconjunto de datos de un AD donde se almacenan la mayoría de las actividades de análisis que en el entorno de Inteligencia de negocios se llevará a cabo [16] . Además son un

subconjunto de datos que pueden funcionar de forma autónoma, o bien enlazado al Almacén de Datos. El motivo por el cual se crean MD es el crecimiento que tiene el almacén y así facilitar su construcción y utilización [17].

Por lo anteriormente mencionado se puede llegar a la conclusión de que un MD es un componente de un Almacén de Datos pero dedicado a un área más específica, que contienen información de datos operacionales y ayudan a decidir sobre estrategias del negocio teniendo en cuenta experiencias pasadas.

Ventajas de utilizar MD [17]

- ✓ Poco volumen de datos.
- ✓ Mayor rapidez de consulta.
- ✓ Consultas SQL o MDX sencillas.
- ✓ Facilidad para conservar los datos a través del tiempo.
- ✓ Fácil acceso a los datos que se necesitan frecuentemente.
- ✓ Crea vista colectiva para grupo de usuarios.
- ✓ Facilidad de creación.
- ✓ Costo inferior de instalación que la de un almacén de datos.

1.3.1 Características de los mercados de datos

Algunas características a tener en cuenta de los MD son las que se relación a continuación [13]:

- ✓ Según las necesidades de los usuarios el diseño del mercado de datos se realiza siguiendo una estructura consistente.
- ✓ La información histórica que posee es mínima.
- ✓ Contiene el grado de granularidad necesaria.
- ✓ Da costes adicionales en hardware, software y accesos de red.
- ✓ Debido a que hay grupos de usuarios que solo acceden a un subconjunto preciso de datos, se hace más fácil el acceso a las herramientas de consulta y divide los datos para controlar mejores accesos.

1.4 Metodologías existentes para el desarrollo de un Mercado de Datos

En la producción de software es primordial el uso de las metodologías, con el fin de organizar, planificar y controlar el proceso de desarrollo de estos, sirviendo estas de guía para mejorar la productividad en el desarrollo y calidad del software.

Desde el surgimiento de las soluciones de AD e Inteligencia de negocios se han venido desarrollando simultáneamente las metodologías para el desarrollo y la implementación de este tipo de soluciones. A nivel internacional existen dos enfoques bien definidos para la creación de un AD: la metodología de Kimball y la metodología de Inmon.

Inmon en su metodología se basa en un enfoque descendente proponiendo primero la creación del AD y a partir de este los MD que se nutrirán de él. Inmon propone la construcción de un repositorio de datos corporativo como fuente de información consistente, consolidada, histórica y de calidad. Como el almacén de datos se construye descendientemente los MD se nutren del almacén corporativo, convirtiéndose en un complejo empresarial de BD relacionales [18].

Por otra parte Kimball plantea la creación del almacén siguiendo un enfoque ascendente, creando primero los MD independientes por cada área departamental y el almacén de datos sería la unión de todos los MD.

1.4.1 Propuesta de metodología para el desarrollo de Almacén de Datos en DATEC

Para el desarrollo del MD del Sistema Xedro-Apside se define como metodología a utilizar: Propuesta de metodología para el desarrollo de AD en DATEC, que toma como base el enfoque de Kimball y se ajusta a las condiciones y características de la producción del centro y la Universidad.

Se tomó como base la Metodología de Kimball por los siguientes elementos:

- ✓ Crea los conceptos de Hechos y Dimensiones, lo que indudablemente es muy eficaz en el proceso de la toma de decisiones y proporciona mayor agilidad en el proceso de desarrollo.
- ✓ Propone ir construyendo el AD a través de la construcción de los MD departamentales, lo que constituye una estrategia buena y coincide con la división lógica de las empresas, entidades, organismos, entre otros.
- ✓ Es una metodología madura y reconocida por el resto de la comunidad dedicada al tema. Tiene bien definidas las etapas, actividades, artefactos y roles [18].

Durante el ciclo de vida de esta metodología se destacan las siguientes fases de desarrollo:

- ✓ Estudio preliminar o planeación: se realiza el estudio de la entidad cliente, la planeación del proyecto, se definen los objetivos, el alcance preliminar, los costos estimados y otras actividades.
- ✓ Requisitos: se realiza en dos direcciones, una, mediante la identificación de las necesidades de información y reglas del negocio; y la otra con un levantamiento detallado de las fuentes de datos a integrar. Después se procede a la definición de los requerimientos.
- ✓ Arquitectura y diseño: se definen las estructuras de almacenamiento, se diseñan las reglas de extracción, transformación y carga, definiéndose la arquitectura de información que regirá el desarrollo de la solución.
- ✓ Implementación: se diseña físicamente el repositorio de datos, se crean las estructuras de almacenamiento, el área temporal de almacenamiento, se ejecutan las reglas de ETL y

se configuran e implementan las herramientas de inteligencia de negocios para la obtención de los elementos que se acordaron con el cliente final.

- ✓ Prueba: se realizan las pruebas al sistema desde las Pruebas de Unidad hasta las de aceptación con el cliente final.
- ✓ Despliegue: se realiza un despliegue piloto en el cual se configuran los servidores y se instalan las herramientas y se carga una muestra de los datos para demostrar que el sistema funciona. Posterior a la aceptación del cliente se realiza la carga de los datos así como la capacitación y transferencia tecnológica.
- ✓ Soporte y mantenimiento: tras la implantación de la solución se brindan los servicios de soporte en línea, vía telefónica, web u otras según el contrato firmado y las condiciones de soporte establecidas.
- ✓ Gestión y administración del proyecto: a lo largo del ciclo de vida se realizan actividades de control, gestión y chequeo del desarrollo, los gastos, las utilidades, los recursos y demás actividades por parte del grupo de dirección del proyecto [18].

El MD realizado en la investigación no transitará por todas las fases del ciclo de vida de la metodología a utilizar, solo llegará hasta la fase de pruebas. Atendiendo a que las dos fases posteriores a esta son ejecutadas por el personal del departamento, y la gestión y administración del proyecto es desarrollada por los especialistas que pertenecen al grupo de dirección.

1.5 Sistemas gestores de bases de datos

Se considera un SGBD a un conjunto de programas que ayudan a administrar la información que está contenida en una base de datos. Estos programas se encargan de la integridad y la seguridad de los datos; y garantizan la interacción con el sistema operativo. Entre los tipos de gestores de bases de datos que existen actualmente se pueden mencionar a los relacionales, jerárquico y de red; pero el más utilizado es el relacional (SGBDR) [19].

Los sistemas gestores de bases de datos definen además lenguajes para permitir a los administradores de la base de datos especificar los datos que componen la base de datos. Estos lenguajes se pueden clasificar en Lenguaje de definición de datos (LDD o DDL), y en Lenguaje de manipulación de datos (MDL o DML). Algunas características que se pueden mencionar de los sistemas gestores de bases de datos son las siguientes:

- ✓ Son capaces de controlar la concurrencia y las operaciones que implican la recuperación de fallos.
- ✓ Definen usuarios y sus restricciones de acceso.
- ✓ Respetan la integridad y seguridad de los datos.
- ✓ Toleran definiciones de esquemas y vistas.

1.5.1 Sistema gestor de base de datos PostgreSQL 9.1

Creado por el proyecto POSTGRES de la universidad de Berkeley, PostgreSQL es un sistema objeto-relacional, ya que incluye características de la orientación a objetos, como puede ser la herencia, tipos de datos, funciones, restricciones, disparadores, reglas e integridad transaccional [19].

Ventajas

- ✓ Cliente/Servidor: utiliza una arquitectura de procesos por usuario cliente/servidor. Hay un proceso maestro que se ramifica para proporcionar conexiones adicionales para cada cliente que intente conectarse a PostgreSQL.
- ✓ Altamente Extensible: soporta los tipos de datos base, así como: tipo, fecha, monetarios, elementos gráficos, datos sobre redes (MAC, IP), cadenas de bits, etc. Además, operadores, funciones, métodos de acceso y tipos de datos definidos por el usuario.
- ✓ Soporte SQL Compresivo: soporta la especificación SQL99 e incluye características avanzadas tales como las uniones SQL92.
- ✓ Es un gestor bajo licencia Berkeley Software Distribution (BSD), que posee una gran escalabilidad, haciéndolo idóneo para su uso en sitios web. Además, por su arquitectura de diseño, escala muy bien al aumentar el número de CPUs y la cantidad de RAM.
- ✓ Sus tablas pueden llegar a 32 TB, sus tuplas 1.6 TB y los campos a 1GB de tamaño respectivamente.
- ✓ El tamaño de la base de datos es ilimitada.

Desventajas

- ✓ Consume bastantes recursos y carga con mucha facilidad el sistema.
- ✓ Velocidad de respuesta un poco deficiente al gestionar bases de datos relativamente pequeñas, aunque esta misma velocidad la mantiene al gestionar bases de datos realmente grandes.
- ✓ No soporta la multidimensionalidad.

Justificación del sistema gestor de base de datos a utilizar

La universidad se encuentra inmersa en un proceso de migración, vinculada principalmente a las nuevas tendencias tecnológicas, que consiste en la utilización de herramientas netamente libres, por lo que se seleccionó PostgreSQL como gestor de base de datos, por el hecho de que existe una amplia comunidad de desarrollo, en la cual se trabaja de forma desinteresada, altruista y libre. Las características antes mencionadas de este gestor hacen que sea flexible, potente y profesional.

1.5.1.1 Administrador de base de datos PgAdmin III 1.14

Por otra parte, la robustez de la combinación PostgreSQL-PgAdmin permite obtener resultados con marcadas características de rapidez y flexibilidad, al permitir al usuario la modificación de código de acuerdo a sus necesidades. A partir de lo antes expuesto se define que para el desarrollo de este trabajo se utilizará como herramienta de administración PgAdmin III 1.14.

Es una aplicación gráfica para administrar el gestor de base de datos PostgreSQL, siendo el más completo y popular con licencia de código abierto. Soporta todas las características de PostgreSQL y facilita enormemente su administración. Permite escribir consultas SQL simples, a través de un editor SQL con resaltado de sintaxis. Esta herramienta incluye también un editor de código de la parte del servidor y un agente para lanzar *scripts* programados [20]. Algunas de sus principales características son [25]:

- ✓ Multiplataforma
- ✓ Diseñado para múltiples versiones de PostgreSQL y derivados
- ✓ Amplia documentación
- ✓ Acceso a todos los objetos de PostgreSQL
- ✓ Interfaz multilingüe

1.6 Sistemas de almacenamiento de datos

Existen dos tipos de almacenamiento, uno es el Procesamiento de Transacciones en Línea (*On-Line Transactional Processing*, OLTP por sus siglas en inglés) y el otro Procesamiento Analítico en Línea (*On-Line Analytical Processing*, OLAP por sus siglas en inglés). Los sistemas OLTP son bases de datos orientadas al procesamiento de transacciones, al que se le pueden realizar operaciones de inserción, modificación y borrado de datos. Este proceso es típico en bases de datos operacionales [21].

Por su parte los sistemas OLAP son bases de datos orientadas al procesamiento analítico, en el que el acceso a los datos es de solo lectura y la operación más común es la consulta, con pocas

inserciones, actualizaciones o eliminaciones. Este sistema permite acceder a grandes volúmenes de datos, de los que se puede extraer información útil. Este proceso es típico en los MD.

Para el desarrollo de la investigación se utilizará OLAP porque ofrece gran ventaja sobre los sistemas OLTP ya que los datos se estructuran según las áreas del negocio y los formatos de los mismos están integrados de manera uniforme en toda la organización, lo que favorece en la rapidez del tiempo de respuesta de las consultas. Proporciona análisis estadísticos y permite ver la información en determinadas vistas y dimensiones. Dentro de este sistema se pueden encontrar tres tipos de almacenamiento: el relacional, el multidimensional y el híbrido.

1.6.1 OLAP (Procesamiento Analítico en Línea)

El sistema Proceso Analítico en Línea (OLAP) es una proyección multidimensional redundante de una relación. Según la definición que le dio Codd, "es un tipo de procesamiento de datos que se caracteriza, entre otras cosas, por permitir el análisis multidimensional de datos" [21].

OLAP es una tecnología que permite a los usuarios mejorar la visión que tienen de sus datos de una manera rápida, interactiva y fácil de usar. Al computar todas las agrupaciones, realiza una agregación de sus resultados en un espacio N-dimensional para responder consultas [12, 22].

Dicho análisis se basa en modelar la información en medidas, dimensiones y hechos. Las medidas son los valores de un dato en particular, las dimensiones son las descripciones de las características que definen dicho dato y los hechos son la definición de una o más medidas para una combinación particular de dimensiones.

En otras palabras una aplicación OLAP permite ver los datos en función de muchas dimensiones, posibilitando a los usuarios expresar los datos de la misma forma que deseen. Entre las funcionalidades que puede ofrecer OLAP, se incluye la declaración de dimensiones, jerarquías, óptima indexación de los datos, definición de operaciones predefinidas de navegación en las dimensiones y de agrupación de medidas, por tanto entre los parámetros que definen estas funcionalidades se destacan [7, 12]:

- ✓ Seleccionar solo la información de un miembro en particular de una dimensión, o lo que es lo mismo, se trabaja con un subconjunto del total de los datos.
- ✓ Pasar la información del nivel anterior de la dimensión actual a una jerarquía definida, consolidando los datos del nivel actual y mostrando el valor consolidado, correspondiente al nivel inmediatamente superior.
- ✓ Realizar la operación inversa del *Roll-up*. Permite ver la información del nivel siguiente de la dimensión actual en una jerarquía definida.
- ✓ Cambiar la dimensión que está caracterizando los datos actualmente considerados.

- ✓ Visualizar la información de otro miembro del mismo nivel de la dimensión que se está evaluando. No detalla ni consolida la información, sino que cambia el miembro para el cual se están presentando los datos.
- ✓ Permitir la consulta de información del nivel inferior en la dimensión actual y la navegación por fuera del modelo multidimensional. La ejecución de esta operación depende de poder establecer el acceso al sistema fuente desde el OLAP.

Características de OLAP [23]

- ✓ El acceso a los datos suele ser de sólo lectura. La acción más común es la consulta, con muy pocas inserciones, actualizaciones o eliminaciones.
- ✓ Los datos se estructuran según las áreas de negocio, y los formatos de los datos están integrados de manera uniforme en toda la organización.
- ✓ El historial de datos es a largo plazo, normalmente, de dos a cinco años.
- ✓ Provee análisis multidimensional dinámico, permitiendo a los usuarios finales realizar actividades analíticas y de navegación, que incluyen cálculo de dimensiones, análisis en períodos de tiempo, visualización de subconjuntos de datos, subir o bajar niveles, comparaciones de varias dimensiones en el área de visualización, entre otros.
- ✓ Está basado en una modalidad cliente/servidor multiusuario, que ofrece respuestas rápidas, independientemente del tamaño y la complejidad de la base de datos.
- ✓ Solución de manipulación de datos multiusuario de alta capacidad diseñada para soportar y operar en una estructura de datos multidimensional.
- ✓ Preparado físicamente para responder rápida y consistentemente a los usuarios finales o cargar datos en tiempo real desde las bases de datos.

1.6.1.1 ROLAP (Procesamiento Analítico en Línea Relacional)

En ROLAP se utiliza una arquitectura de tres niveles. La base de datos relacional maneja el almacenamiento de datos, el motor OLAP proporciona la funcionalidad analítica, y alguna herramienta especializada es empleada para el nivel de presentación. El nivel de aplicación es el motor OLAP, que ejecuta las consultas de los usuarios. La arquitectura ROLAP es capaz de usar datos precalculados (si estos están disponibles), o de generar dinámicamente los resultados desde la información elemental (menos resumida).

Esta arquitectura accede directamente a los datos del Almacén de Datos y soporta técnicas de optimización para acelerar las consultas como tablas particionadas, soporte a la desnormalización, precalculado de datos, índices, entre otros. Se pueden mencionar como algunos productos que basan sus implementaciones en ROLAP a [22]:

- ✓ *BusinessObjects* [BOS]
- ✓ *Microstrategy's DSS Agent* [MIC]
- ✓ *Redbrick* [RED]
- ✓ *Oracle Warehouse* [ORA]
- ✓ *DB2 Data Warehouse* [DB2]

1.6.1.2 MOLAP (Procesamiento Analítico en Línea Multidimensional)

Un sistema MOLAP usa una Base de Datos Multidimensional (BDMD), en la que la información se almacena multidimensionalmente. MOLAP utiliza una arquitectura de dos niveles: la BDMD y el motor analítico. La BDMD es la encargada del manejo, acceso y obtención de los datos.

La información procedente de los sistemas transaccionales se carga en el sistema MOLAP. Una vez cargados los datos en la BDMD, se realiza una serie de cálculos para obtener datos agregados a través de las dimensiones del negocio, poblando la estructura de la BDMD; luego de llenar esta estructura, se generan índices y se emplean algoritmos de tablas de dispersión para mejorar los tiempos de accesos de las consultas [22].

1.6.1.3 HOLAP (Procesamiento Analítico en Línea Híbrido)

Solución OLAP híbrida que combina el uso de las arquitecturas ROLAP y MOLAP. En una solución con HOLAP, los registros detallados (los volúmenes más grandes) se mantienen en la base de datos relacional, mientras que los agregados lo hacen en un almacén MOLAP independiente. Un sistema HOLAP resuelve el problema de dispersión, dejando los datos más granulares (menos agregados) en la base de datos relacional, pero almacena los agregados en un formato multidimensional, minimizando así la presencia de celdas vacías [22].

1.6.2 Valoración general

Una vez analizados los diferentes tipos de almacenamiento se selecciona ROLAP ya que ahorra espacio de almacenamiento y es útil cuando se trabaja con amplios conjuntos de datos. Su ventaja principal reside en la posibilidad de utilizar una tecnología ampliamente extendida y utilizada para la gestión de datos, los sistemas relacionales [24]. Además se tuvo en cuenta que el Sistema Gestor de Base de Datos (SGBD) PostgreSQL soporta el almacenamiento relacional, no así el multidimensional. En la actualidad los SGBD que dan soporte al almacenamiento multidimensional son propietarios, por lo que no están en correspondencia con las políticas de desarrollo de la UCI y el país. A continuación se presenta una tabla comparativa de los tipos de almacenamiento antes mencionados:

Tabla 3. Criterios comparativos de tipos de almacenamiento

Criterios	Tipo de almacenamiento		
	MOLAP	ROLAP	HOLAP
Almacenamiento de las agregaciones	Modelo multidimensional	Base de datos relacional	Modelo multidimensional
Almacenamiento de los datos	Modelo multidimensional	Base de datos relacional	Base de datos relacional
Facilidad de creación	Sencillo	Muy sencillo	Sencillo
Escalabilidad	Problemas de escalabilidad	Son más escalables	Escalable
Recomendados para	Cubos con uso frecuente	Datos que no son frecuentemente usados	Si el cubo requiere una rápida respuesta
Velocidad de respuesta	Buena	Regular o Baja	Buena para consultas que poseen agregaciones, Regular para datos de bajo nivel

1.7 Modelo Entidad-relación y Modelo dimensional

1.7.1 Modelo Entidad-relación

El modelo Entidad-relación es un lenguaje para el modelado de los datos dentro de un sistema de información. Permite la creación de esquemas conceptuales de bases de datos. En ellos los datos son fraccionados en muchas entidades donde estas posteriormente se convierten en una tabla física dentro de la base de datos [25].

Dentro de este modelo hay una gran simetría por lo parecido de sus tablas, provocando que no se pueda identificar que tabla es más significativa que otra. Otra desventaja que poseen es que no existe una manera eficiente de diferenciar cuales son las tablas que contienen medidas numéricas con respecto a las que incluyen información estática.

De manera general los modelos Entidad-relación no son recomendables para el diseño de un AD, dado que no garantizan una óptima recuperación de la información almacenada. Las consultas se tornan muy complejas, dado que abarcan una gran cantidad de registros y tablas, que hacen de este modelo, algo muy difícil de entender por los usuarios.

1.7.2 Modelo dimensional

Cuando se hace referencia a los AD es imposible obviar el Modelo dimensional. A diferencia de los sistemas de bases de datos más convencionales, un Almacén de Datos requiere de un Modelo dimensional para su diseño. Estos modelos poseen en su estructura la misma información que maneja el modelo Entidad-Relación, pero la diferencia entre ellos es la manera de organizarla. El modelado dimensional se puede adaptar a dos entornos principales: el relacional y el dimensional (ROLAP y MOLAP) y es una técnica que hace más entendible para el usuario la base de datos. Existen dos tipos de tablas fundamentales en este modelo [26]:

Tabla de dimensiones

Estas acompañan a la tabla de hechos y determinan los parámetros (dimensiones) de los que dependen los hechos registrados en la tabla de hechos. Son elementos que contienen atributos (o campos) que se utilizan para restringir y agrupar los datos almacenados en una tabla de hechos cuando se realizan consultas sobre dichos datos en un entorno de AD o MD. Los atributos de las dimensiones sirven como fuente primaria de las restricciones de las consultas, agrupaciones y las etiquetas de los reportes [12].

Tabla de hechos

Todo AD tiene incluido una o varias tablas de hechos, eso depende en gran medida de su complejidad. Estas tablas básicamente capturan los datos encargados de medir las operaciones de un equipo. Contienen datos numéricos (hechos) que proporcionan información sobre el historial de la organización.

Están compuestas por un gran número de filas en dependencia de los años de trabajo de un proyecto, y su capacidad de almacenar información está a un nivel elevadísimo de detalle, esto sobresale como una de sus principales características. No deben incluir datos que no procedan de los campos numéricos y los campos de índice, guardan relación entre los hechos con las entradas en las tablas de dimensiones correspondientes. Este tipo de tabla incluye atributos para calcular las medidas y parámetros para determinar la descomposición de los datos, estos se describen como [12]:

- ✓ **Medidas:** Una medida es un atributo (campo) de una tabla que se desea analizar, agrupando sus datos, usando los criterios de corte conocidos como dimensiones[27]. Las medidas más útiles para incluir en una tabla de hechos son aquellas medidas numéricas que pueden calcularse con la suma de varias cantidades de la tabla. Es decir, las medidas candidatas son los datos numéricos, pero no cada atributo numérico es una medida candidata. En consecuencia, por lo general los hechos a almacenar en una tabla de hechos van a ser casi siempre valores numéricos, enteros o reales.

- ✓ **Granularidad:** La granularidad significa especificar el nivel de detalle. La elección de la granularidad depende de los requerimientos del negocio y lo que es posible a partir de los datos actuales [27]. La granularidad de la tabla de hechos representa el nivel más atómico por el cual se definen los datos.

Atributos e indicadores

Los atributos son criterios utilizados para analizar los indicadores. Se basan, en los datos de referencia de las tablas de dimensiones, además son campos o criterios de análisis, pertenecientes a las tablas de dimensiones [28].

La calidad de todo almacén de dato, se mide por la definición de los atributos de las dimensiones. Su poder es directamente proporcional a la calidad y profundidad de estos atributos[13]; los indicadores son variables que pueden tomar un valor de una determinada unidad de medida y de un determinado tipo de dato.

Las tablas de hechos y dimensiones poseen distintas formas de representación, basado en esquemas, estos se definen como [26]:

- ✓ **Esquema en estrella**

En el esquema en estrella la tabla de hechos es la única tabla que tiene múltiples uniones que la conectan con otras tablas. El resto de las tablas del esquema (tablas de dimensión) únicamente hacen uniones con esta tabla de hechos. Toda la información referente a una dimensión, se almacena en la misma tabla. En el esquema copo de nieve las tablas de dimensiones están normalizadas, con respecto al esquema estrella.

- ✓ **Esquema copo de nieve**

En cada dimensión se almacenan jerarquías de atributos o bien simplemente se separan atributos en otra entidad por razones de desempeño y mejor utilización del espacio. El uso más común de este esquema, es cuando las tablas de dimensiones son muy grandes o complejas y es muy difícil representar los datos en un esquema estrella.

- ✓ **Constelación de hechos**

Para cada esquema estrella o esquema copo de nieve de un almacén de dato es posible construir un esquema de constelación de hechos. Este esquema es más complejo que las otras arquitecturas, debido a que contiene múltiples tablas de hechos. Con esta solución las tablas de dimensiones pueden estar compartidas entre más de una tabla de hechos.

Justificación del esquema de almacenamiento de datos a utilizar

Para la realización del MD se utilizará el esquema en estrella porque posee mejor rendimiento y velocidad que los demás esquemas de almacenamientos, debido a que no utiliza un gran

número de uniones para consultar la información. El esquema estrella permite estructurar la información en procesos, vistas y métricas. Mientras, en las tablas de las dimensiones, los atributos se destinan a elementos de nivel (que representan los distintos niveles de las jerarquías de dimensión) y a atributos de dimensión (encargados de la descripción de estos elementos de nivel).

1.8 Herramientas de modelado

Se puede definir a las herramientas CASE¹ como un conjunto de programas y ayudas que dan asistencia a los analistas, ingenieros de software y desarrolladores, durante todos los pasos del ciclo de vida de desarrollo de un software. Las herramientas CASE permiten la aplicación práctica de metodologías estructuradas, las cuales al ser realizadas con una herramienta agilizan el trabajo. Facilitan la realización de prototipos y el desarrollo conjunto de aplicaciones, simplifican el mantenimiento de los programas, mejoran y estandarizan la documentación. Además, aumentan la portabilidad de las aplicaciones, facilitan la reutilización de componentes, posibilitan el desarrollo y refinamiento visual de las aplicaciones, mediante la utilización de gráficos [29].

1.8.1 Herramienta de modelado Visual Paradigm 8.0

Es una herramienta CASE que utiliza lenguaje de modelado UML y soporta las etapas del ciclo de vida de desarrollo de software. Permite la ingeniería inversa, generación de código, importación desde Rational Rose, exportación/importación XML, generador de informes, editor de figuras y otros elementos particulares [30].

Visual Paradigm es un entorno de creación de diagramas para UML. El mismo utiliza el diseño centrado en casos de uso y enfocado al negocio que genera un software de mayor calidad. Emplea un lenguaje estándar común a todo el equipo de desarrollo que facilita la comunicación. También mantiene sincronizado el modelo con el código en todo el ciclo de desarrollo. Presenta disponibilidad de múltiples versiones, disponibilidad para integrarse en los principales IDEs que existen en la actualidad y disponibilidad en múltiples plataformas.

Justificación de la herramienta de modelado a utilizar

Se decidió utilizar Visual Paradigm porque es una herramienta CASE profesional que soporta todo el ciclo de vida del desarrollo de software. Además, brinda la posibilidad de modelar todos los tipos de diagramas de clases, permite la compatibilidad entre ediciones, la generación de código desde diagramas y la documentación asociada a cada etapa del proceso de desarrollo. Esta herramienta presenta licencia gratuita y comercial. La UCI cuenta con la licencia para su uso.

¹ CASE: Ingeniería de Software Asistida por Computadoras

1.9 Herramientas para la integración de datos

El proceso de ETL se utiliza para la construcción de los MD, ya que es necesario que la información contenida represente la realidad del negocio, sea confiable y esté disponible en el momento que los usuarios y la organización la necesiten. Para los procesos ETL, es preciso contar con herramientas que permitan reducir tiempo y costos. El desarrollo y la diversificación de las herramientas para los procesos de ETL, actualmente son crecientes, y se refleja en la amplia variedad de herramientas tanto comerciales como de código abierto [31]. Las herramientas a utilizar para la integración de los datos es Pentaho Data Integration para realizar la ETL y el Datacleaner para la limpieza de los datos.

1.9.1 Pentaho Data Integration 4.2.0

Es una herramienta multiplataforma que permite extraer la información de las diferentes fuentes, transformar la información a través de un modelo dimensional y cargar los resultados de la transformación en una base de datos tipo AD, a partir del entorno gráfico de desarrollo [31]. Pentaho es una herramienta libre muy poderosa y una de las más antiguas y utilizadas por los usuarios. Sus principales ventajas son [32]:

- ✓ Multiplataforma.
- ✓ Uso de tecnologías estándar: java, XML, *javascript*.
- ✓ Incluye cuatro herramientas:
 - ✓ *Spoon*: para diseñar transformaciones de ETL usando el entorno gráfico.
 - ✓ *PAN*: para ejecutar transformaciones diseñadas con *spoon*.
 - ✓ *CHEF*: para crear trabajos.
 - ✓ *Kitchen*: para ejecutar trabajos.
- ✓ Basada en dos tipos de objetos: transformaciones (colección de pasos en los procesos de ETL) y trabajos (colección de transformaciones).
- ✓ Soporta diferentes lenguajes de base de datos como por ejemplo: MySQL y Postgres.

1.9.2 DataCleaner 1.5.4

Permite la evaluación del nivel de calidad de los datos contenidos en el sistema de información. Es una aplicación muy fácil de usar, genera sofisticados informes y gráficos que permiten a los usuarios determinar de un vistazo el nivel de calidad de los datos, identificar y analizar la estructura del origen de datos y combinar resultados y gráficos, creando vistas fáciles de interpretar para evaluar la calidad de los datos [33].

1.10 Herramientas para la inteligencia de negocios

Las herramientas de inteligencia de negocios han sido creadas para ayudar a la toma de decisiones entre las diferentes empresas e instituciones de un estado. Además, permiten mostrar una visión general de todos los procesos de la entidad a sus directivos, facilitando un mejor entendimiento en el análisis y en la presentación de los datos.

La plataforma *Open Source Pentaho Business Intelligence* cubre muy amplias necesidades de análisis de los datos y de los informes empresariales. Las soluciones de Pentaho están escritas en java y tienen un ambiente de implementación también basado en java. Eso hace que Pentaho sea una solución muy flexible para cubrir una amplia gama de necesidades empresariales.

Las tecnologías de BI son herramientas de soporte de decisiones que permiten en tiempo real, acceso interactivo, análisis y manipulación de información crítica para la empresa. Intentan ayudar a las personas a entender los datos más rápido, a fin de que puedan tomar mejores y más rápidas decisiones y finalmente, mejorar sus movimientos hacia la consecución de objetivos de negocios. Los impulsores claves detrás de los objetivos de inteligencia de negocio son incrementar la eficiencia organizacional y la efectividad. A continuación se mencionan las herramientas definidas para realizar la inteligencia de negocio.

1.10.1 Pentaho Schema Workbench 3.2.1

Es una interfaz de diseño que permite crear y probar esquemas de cubos Mondrian OLAP visualmente. La Plataforma de BI de Pentaho incrusta el motor de consulta Mondrian, como parte de su arquitectura. Además permite la ejecución de consultas MDX. Está distribuida bajo la licencia EPL [34].

1.10.2 Pentaho BI Server 4.5.0

Es una herramienta de reporte que permite crear informes, ya sea para ejecutarlos directamente o para publicarlos en la plataforma de Inteligencia de negocios y que desde allí puedan ser utilizados por los usuarios. Incluye asistentes para facilitar la configuración de propiedades, cuenta también con un editor de consultas para facilitar la confección de los datos que serán utilizados en un informe [35].

Capítulo 1: Fundamentación teórica

Conclusiones del capítulo

La construcción del MD Histórico de precios constituye una solución viable a la problemática planteada, teniendo en cuenta que permitirá el almacenamiento de la información y el análisis de los históricos de precios para los especialistas de mercadotecnia de la Empresa BK Import/Export. Para guiar el proceso de desarrollo de la solución se definió la Metodología para el Desarrollo de AD de DATEC. El diseño será realizado según el esquema estrella, como modo de almacenamiento se utilizará ROLAP y la suite de Pentaho para el proceso de inteligencia de negocio y extracción, transformación y carga de los datos.

CAPÍTULO 2: Análisis y diseño

Introducción

En este capítulo se tiene en cuenta las necesidades del negocio, así como las reglas que tiene que cumplir. Se especifican los requisitos de información (RI), los funcionales y los no funcionales. Se realiza el diagrama de caso de uso del sistema. También se define la arquitectura de información que regirá el desarrollo de la solución propuesta. Se construye la matriz bus, el modelo de datos donde se determinan: las dimensiones, los hechos y las medidas.

2.1 Necesidades del negocio

Las funcionalidades del negocio están en correspondencia con las necesidades del cliente. Para identificar la situación que afrontaban los especialistas y los problemas del departamento de mercadotecnia fue utilizado como método de recopilación de información la entrevista. Como resultado de su aplicación se obtuvo que:

- ✓ Es difícil encontrar el precio de determinado producto en los catálogos debido a que el nombre puede variar.
- ✓ Es difícil realizar búsquedas de los productos en los pliegos de concurrencia debido al gran cúmulo de información que se almacena.

Lo planteado anteriormente provoca que se realice un incorrecto análisis estadístico de los históricos de precios de los productos y que el proceso se retrase.

2.2 Especificación de los requisitos

El análisis de requisitos constituye una de las fases principales en la construcción de un MD. El levantamiento de los requisitos consiste en identificar las necesidades de información de la organización, las características y las cualidades que debe poseer el sistema. En las soluciones de MD se identifican tres tipos de requisitos: de información, funcionales y no funcionales.

2.2.1 Requisitos de información

Los requisitos de información describen la información y los datos que el sistema debe almacenar para satisfacer a los clientes. Esto se define a partir de las necesidades de información identificadas en el negocio, que permite el análisis del comportamiento de los indicadores a medir según los objetivos y las metas de la organización. A continuación se muestran los requisitos de información identificados durante el proceso de análisis.

RI1 Obtener el precio por proveedor dado un producto.

RI2 Obtener la cantidad de producto por unidad de medida, número de oferta, país, moneda, proveedor y fecha.

RI3 Obtener el precio por proveedor, día, mes y año dado un producto.

2.2.2 Requisitos funcionales

Los requisitos funcionales definen las capacidades o condiciones que el sistema debe cumplir. Permiten expresar específicamente las responsabilidades del sistema que se propone con la intención de lograr la satisfacción plena del cliente. A continuación se relacionan:

RF1 Autenticar usuario

RF2 Administrar roles

RF3 Administrar reporte

RF4 Administrar usuario

RF5 Extraer información

RF6 Realizar transformación y carga de los datos

2.2.3 Requisitos no funcionales

Los requisitos no funcionales (RNF) describen las propiedades y cualidades que deben tener la solución, para que sea rápido, confiable y seguro. A continuación se mencionan:

Usabilidad

RNF1 Mostrar los mensajes, títulos y demás textos que aparezcan en la interfaz del sistema en idioma español: Los títulos de los componentes de la interfaz, los mensajes para interactuar con los usuarios y los mensajes de error, deben ser en idioma español y tener una apariencia uniforme en todo el sistema.

RNF2 Agilizar el acceso a los reportes del almacén de datos mediante la distribución de la información por áreas de análisis y libros de trabajo: El usuario podrá acceder de manera rápida a la información que solicita, la cual se encontrará ubicada dentro del libro de trabajo y área correspondiente, de acuerdo al objetivo de su solicitud.

Confiabilidad

RNF3 Garantizar la persistencia de la información: Se debe realizar un respaldo total de los datos del MD con una frecuencia mensual. La información estará almacenada en el departamento de mercadotecnia y el responsable de la misma será el informático con que cuenta el departamento.

Interfaz

RNF4 Acceder al sistema: El usuario deberá acceder a la aplicación mediante el protocolo de transferencia de hipertexto, utilizando un navegador web.

Restricciones de diseño

RNF5 Utilizar el lenguaje de programación definido durante la investigación: Para la programación en el MD se utilizará PostgreSQL/pgAdmin como lenguaje dentro del Sistema Gestor de base de datos, para la implementación de los procesos de integración de datos el lenguaje JavaScript y el lenguaje MDX para realizar las consultas.

RNF6 De la suite Pentaho, se utilizarán las herramientas Schema Workbench 3.2.1 y Pentaho BI Server 4.5.0. Para el uso de las herramientas anteriores se requiere la instalación de la máquina virtual de java Java Virtual Machine 7.0 o una versión superior.

Requisitos de software

RNF7 Instalar en las estaciones de trabajo el software necesario para el correcto funcionamiento del sistema: Las estaciones de trabajo clientes deben contar con un navegador web (en caso de utilizar el navegador Firefox, una versión posterior a la 2.0). El servidor de aplicaciones debe contar con las herramientas: Pentaho BI Server 4.5.0, Java Virtual Machine 7.0 o superior, Apache Tomcat 6.0 y Un navegador web. El servidor de base de datos debe contar con las herramientas: PostgreSQL 9.1 como Sistema Gestor de Base de Datos y PgAdmin 1.14 como Administrador de Base de Datos.

Requisitos de hardware

RNF8 Proporcionar características mínimas de hardware a las estaciones de trabajo clientes y servidores: las estaciones de trabajo deben contar con un mínimo de 512 Mb de memoria RAM. Los servidores para lograr una explotación aceptable del sistema deben contar con los requerimientos mínimos de hardware: 1 Gb RAM y 80 Gb de disco duro.

2.3 Reglas del negocio

Las reglas de negocios definen y controlan la estructura, funcionamiento y estrategia de una organización mediante políticas, medidas, procedimientos y restricciones que son de vital importancia para alcanzar sus objetivos. Las reglas de negocios pueden estar formalmente definidas en manuales, contratos o bien pueden existir como conocimiento con que cuentan los miembros de la institución. El proceso de especificación implica reconocerlas dentro del negocio y aplicarlas a la solución propuesta.

Durante el análisis fueron identificadas las siguientes reglas del negocio:

RN1 Los identificadores de las dimensiones no pueden tomar valor nulo y no pueden existir llaves primarias repetidas en la dimensiones.

RN2 Los valores que identifiquen cantidad tienen que ser mayor que cero.

2.4 Diagrama de casos de uso del sistema

Un caso de uso (CU) constituye una secuencia de interacciones que se desarrollan entre los actores y un sistema, en respuesta a un evento que inicia un actor sobre el propio sistema. Los requisitos de información y los requisitos funcionales son agrupados en casos de uso de información y casos de uso funcionales respectivamente, los cuales aportan una idea enfocada al correcto funcionamiento del sistema. En la figura que se muestra a continuación queda evidenciado el diagrama de casos de uso del sistema en el que está representado los CU de información y funcionales.

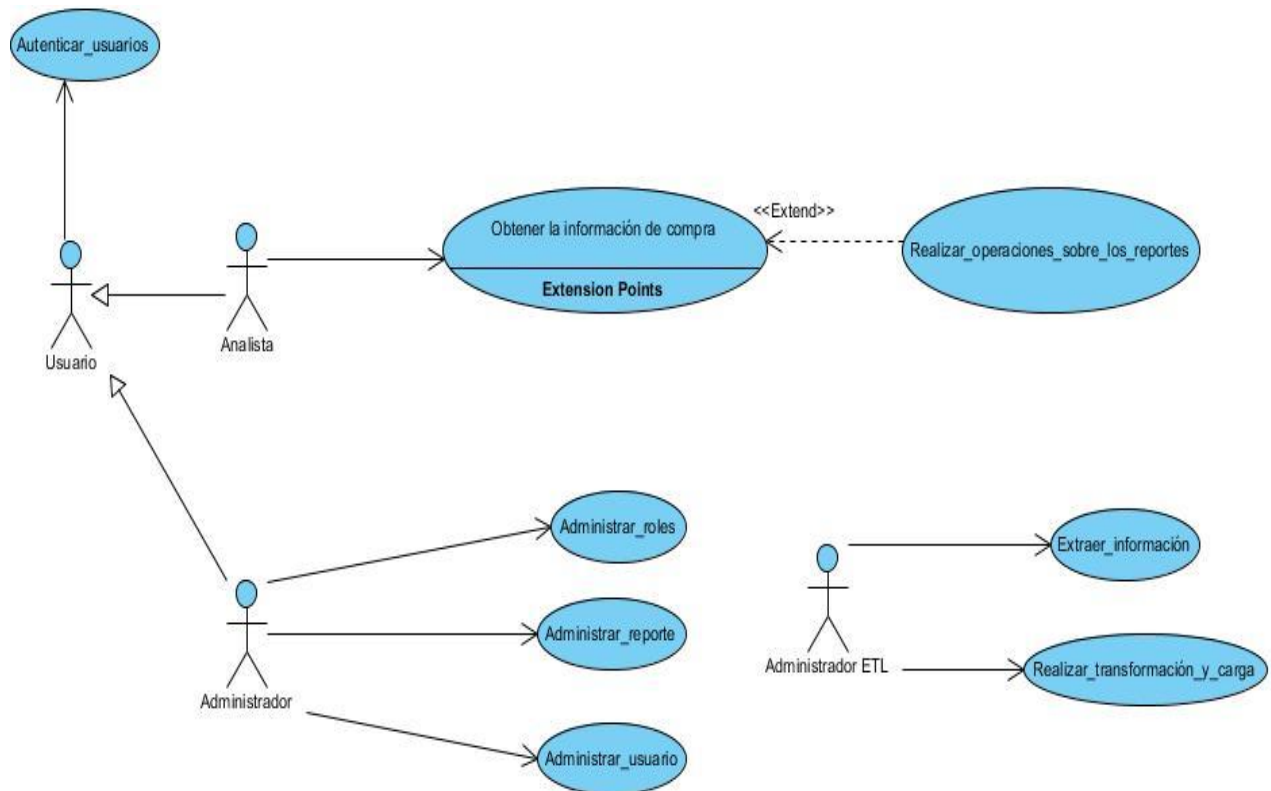


Figura 1. Caso de uso del sistema

2.4.1 Actores del sistema

Administrador: encargado de administrar los usuarios del sistema, asignarles sus respectivos roles, además de administrar dichos roles y los reportes.

Analista: encargado de los casos de uso relacionados con la presentación y análisis de la información del MD.

Administrador ETL: encargado de llevar a cabo los procesos de extracción, transformación y carga de los datos del sistema fuente.

2.4.2 Caso de uso de información

Los casos de uso de información (CUI) agrupan un conjunto de requisitos de información teniendo en cuenta los conceptos del negocio que manejan, fundamentalmente por tema de análisis. En la presente investigación se identificó un tema de análisis, generando un grupo de CUI

CUI1 Obtener la información de compra: visualiza los reportes de los requisitos de información de la compra.

Tabla 4. Descripción del CUI Obtener información de compra

Objetivo	Mostrar toda la información referente a los indicadores compra.	
Actores	Analista: (Inicia) seleccionando el Área de Análisis (A.A) Mercadotecnia	
Resumen	El caso de uso se inicia cuando el actor desea hacer un análisis de la información relacionada con los productos desde diferentes perspectivas. El actor selecciona el reporte que desea ver, el sistema muestra la información contenida y a partir de la información mostrada el actor realiza los cambios deseados al reporte. Luego de esto finaliza el caso de uso.	
Complejidad	Alta	
Prioridad	Media	
Precondiciones	El usuario se autenticó correctamente. Los datos correspondientes fueron cargados en el MD. Los reportes relacionados con los asuntos del producto fueron creados.	
Postcondiciones	Los reportes correspondientes al caso de uso fueron consultados.	
Flujo de eventos		
Flujo básico Presentar información de indicadores de productos		
	Actor	Sistema
1.	Selecciona el A.A de Mercadotecnia.	Muestra el Área de Análisis (A.A) de Mercadotecnia.
2.	Selecciona el Libro de Trabajo (L.T) compra.	Muestra el contenido en el L.T de compra.
3.	Selecciona el reporte que desea analizar.	Muestra la información contenida en el reporte seleccionado y brinda opciones al actor para hacer cambios al reporte durante su análisis. Finaliza el CU.

Prototipo de Interfaz de usuario								
							Medidas	
Moneda.Moneda	Fecha.Fecha	Oferta.Oferta	Pais.Pais	Producto.Producto	Proveedor.Proveedor	Unidad de medida	Cantidad de Productos	precio_producto min
Todas	Fecha	Todas	Todos	Todos	Todos	Todas	13.369	0,21
Relaciones					CU Incluidos	N/A		
					CU Extendidos	Realizar operaciones sobre los reportes		
Requisitos no funcionales					Ver el artefacto: CEIGE_Xedro-Apside_Histórico de precio_010113_Especificación_de_requisitos_de_software, el epígrafe 3.4			
Asuntos pendientes					N/A			

2.4.3 Caso de uso funcionales

Los casos de usos funcionales (CUF) agrupan los requisitos funcionales para los diferentes subsistemas que componen la solución. La investigación cuenta con siete CUF.

CUF1 Autenticar usuario: se realiza la autenticación de los usuarios en el sistema.

CUF2 Administrar usuario: se insertan, modifican o eliminan los usuarios que interactúan con el sistema.

CUF3 Administrar rol: se insertan, modifican o eliminan en el sistema.

CUF4 Administrar reporte: se eliminan, insertan o modifican los reportes que se visualizan.

CUF5 Realizar_operaciones_sobre_los_reportes: se realizan las operaciones o funciones deseadas sobre los reportes.

CUF6 Extraer información: se realiza la extracción de los datos necesarios de las BD fuentes.

CUF7 Realizar_transformación_y_carga: se realiza la transformación y carga de los datos necesarios para la construcción del MD.

A continuación en la tabla se describe el caso de uso Autenticar usuario. Las demás descripciones de los CU se pueden encontrar en el artefacto CEIGE_Xedro-Apside_Histórico de precio Especificación de casos de uso.

Tabla 5 Descripción del CU Autenticar usuario

Objetivo:	Permitir que la persona se autentique y acceda al sistema si es un usuario del mismo.
Actores:	Analista y Administrador.

Resumen:	Inicia cuando el actor se conecta al sistema para acceder a las funcionalidades del mismo. Para ello debe proporcionar un usuario y contraseña. El caso de uso termina una vez validado el usuario y contraseña cuando el usuario logra acceder al sistema.	
Complejidad:	Media	
Prioridad:	Baja	
Precondiciones:	El sistema debe estar disponible. El usuario debe existir en la BD.	
Postcondiciones:	El usuario ha sido autenticado en el sistema.	
Flujo Normal de Eventos		
1. El usuario interactúa con el sistema.	2. El sistema muestra un formulario para introducir los datos.	
3. El usuario introduce los datos.	4. El sistema valida los datos.	
	5. El sistema le da los permisos para entrar a la aplicación. Finaliza el caso de uso.	
Flujo Alternativo		
Acción del Actor	Respuesta del Sistema	
	5.1 Si los datos no son válidos, muestra un mensaje de error y regresa al punto 2 del Flujo Normal de Eventos . Finaliza el caso de uso.	

2.5 Definición de la arquitectura de un mercado de datos

La arquitectura base de un MD define los aspectos arquitectónicos de la solución. Esta presenta cuatro niveles:

- ✓ **Fuente de datos:** se refiere al origen de los datos.
- ✓ **Subsistema de integración:** incluye los procesos que permiten que los datos de las fuentes sean extraídos, transformados y cargados hacia las fuentes destino.

- ✓ **Subsistema de almacenamiento:** base de datos relacional que contiene las tablas de dimensiones y hechos cargadas a través de los procesos de ETL.
- ✓ **Subsistema de visualización:** comprende las interfaces orientadas a usuario que extraen información y le facilitan la toma de decisiones.

El subsistema de integración se nutre de las diferentes fuentes de datos y se encarga de los procesos que integran y transforman la información para su carga. Los usuarios que acceden a este subsistema son los encargados de la administración de los procesos de ETL.

El subsistema de almacenamiento se abastece de los datos cargados en los procesos de ETL. La información es almacenada en una base de datos en el SGBD PostgreSQL y se administra por los usuarios autorizados a través de la herramienta PgAdmin III.

El subsistema de visualización permite mostrar la información estandarizada en forma de reporte a través de la herramienta *Pentaho Business Intelligence*; los usuarios que acceden están autorizados para obtener los reportes.



Figura 2. Arquitectura del mercado de datos

2.6 Diseño de la solución

El diseño es un refinamiento del análisis que tiene en cuenta los requisitos no funcionales y cómo cumple el sistema sus objetivos. El diseño debe ser suficientemente claro para que el sistema pueda ser implementado sin ambigüedades. Se modela el sistema incluyendo la arquitectura, para que soporte los requisitos funcionales y no funcionales, así como las restricciones que este propone.

2.6.1 Diseño del subsistema de almacenamiento

Para el diseño del subsistema de almacenamiento es necesario identificar las dimensiones y los hechos con sus medidas asociadas, así como las relaciones existentes entre hechos y dimensiones; definiendo una política de respaldo y recuperación que garantice la integridad de los datos almacenados.

Dimensiones

Las dimensiones representan cada uno de los ejes en un espacio multidimensional. Suministra el contexto en el que se obtienen las medidas de un hecho. También se utilizan para seleccionar y agrupar los datos en un nivel de detalle deseado.

A continuación se describen las dimensiones y la jerarquía que están relacionadas con el repositorio principal donde se va a almacenar la información atómicamente.

1. **Dimensión producto** (dim_producto): almacena la descripción del producto.
2. **Dimensión proveedor** (dim_proveedor): describe la información de los proveedores atendiendo al nombre y sus siglas.
3. **Dimensión fecha** (dim_fecha): esta dimensión es una de las comunes e importantes en el diseño de MD, debido a que define el tiempo para enmarcar la información almacenada y organizarla atendiendo al momento en que fue captada.
4. **Dimensión país** (dim_pais): almacena los valores de los distintos países.
5. **Dimensión oferta** (dim_oferta): almacena la información de las ofertas atendiendo al número de oferta y a la moneda que se maneja la oferta.
6. **Dimensión unidad de medida** (dim_um): describe los valores bajos los cuales puede clasificarse la cantidad de productos.
7. **Dimensión moneda** (dim_moneda): describe los valores de tipo de moneda bajos los cuales se hace el pago de los productos.

Tabla de hechos

Teniendo en cuenta que en el caso del MD en cuestión no fue identificado ningún campo para el cual sea necesario aplicar el concepto de dimensión degenerada, las tablas de hechos diseñadas solo almacenan las medidas numéricas y las llaves asociadas a cada una de las dimensiones con que se relacionan. Para el desarrollo del MD Histórico de precios se identificó un hecho.

1. **Hecho compra** (hech_compra): se define como medida numérica cantidad y precio del producto.

Matriz bus o matriz dimensional

La matriz bus es esencialmente la arquitectura dimensional de los datos de la organización para cada proceso del negocio identificado, en este caso como es un solo proceso de negocio todas las dimensiones tienen relación con el único hecho identificado.

Tabla 6. Matriz BUS

Dimensiones	Hechos
	Compra
Producto	X
Proveedor	X
Fecha	X
País	X
Oferta	X
Unidad de medida	X
Moneda	X

Modelo de datos

Una vez definido dentro del negocio las dimensiones y medidas se procede a la estructuración del modelo dimensional. A continuación se muestra un fragmento del modelo de datos diseñado para el desarrollo del Mercado de datos Histórico de precios, donde se evidencia la topología estrella, atendiendo a que existe un solo hecho y varias dimensiones.



Figura 3. Modelo de datos

2.6.1 Diseño del subsistema de integración

El subsistema de integración comprende el perfilado de los datos y la extracción de los mismos desde los sistemas fuentes, los cuales sufren un conjunto de transformaciones. Conforma los datos de manera que fuentes separadas puedan ser aprovechadas conjuntamente, y finalmente hace su entrega en un formato listo para el almacenamiento, de forma tal, que permita a los desarrolladores construir la capa de presentación de la aplicación y los usuarios finales puedan tomar decisiones. El perfilado de los datos y el diseño de las transformaciones constituyen elementos esenciales para lograr el diseño del subsistema de integración.

Perfilado de los datos

Por perfilado de datos se entiende el análisis de los datos que permite comprender su contenido, estructura, calidad y dependencias. Se realiza con el objetivo de conocer el estado en que se encuentran los datos que próximamente se extraerán de las fuentes, así como administrar y supervisar la calidad de los mismos. Se verifica la existencia de valores nulos, distintos y duplicados, permitiendo definir nuevas reglas del negocio que pasarían a ser las reglas de transformación aplicadas durante la implementación del subsistema de integración. De esta manera se garantiza que la información es útil y aplicable a la situación del negocio.

Diseño de los procesos de integración

Una vez que se conoce la estructura, contenido y fiabilidad de los datos, se procede a realizar el diseño de las transformaciones. Estos diseños pueden variar a la hora de la implementación de las transformaciones, porque durante el proceso de desarrollo de las mismas suelen surgir situaciones con los datos y se llevan a cabo disímiles estrategias para resolverlas. Para cargar los indicadores del Mercado de datos Histórico de precios se diseña la siguiente transformación.

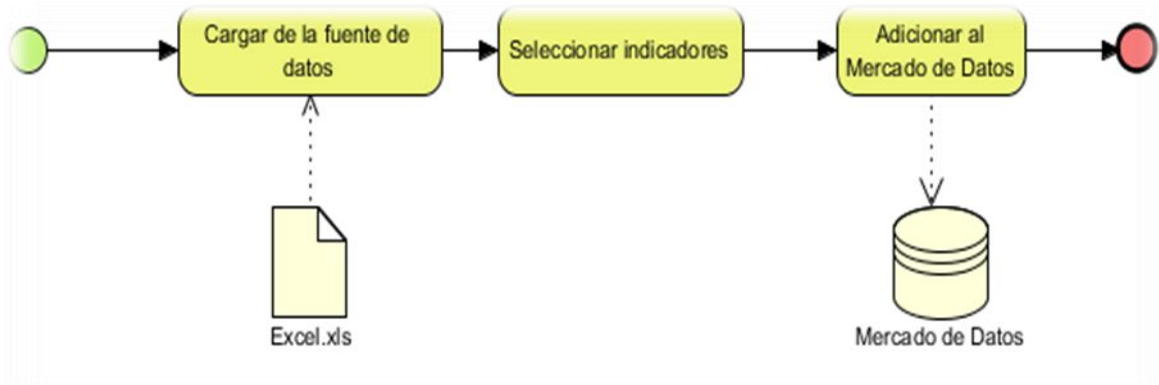


Figura 4. Diseño de la transformación para la carga de los indicadores

En el caso de las tablas de hechos, se diseña la siguiente transformación que permitirá cargar los datos hacia el Mercado de datos Histórico de precios:

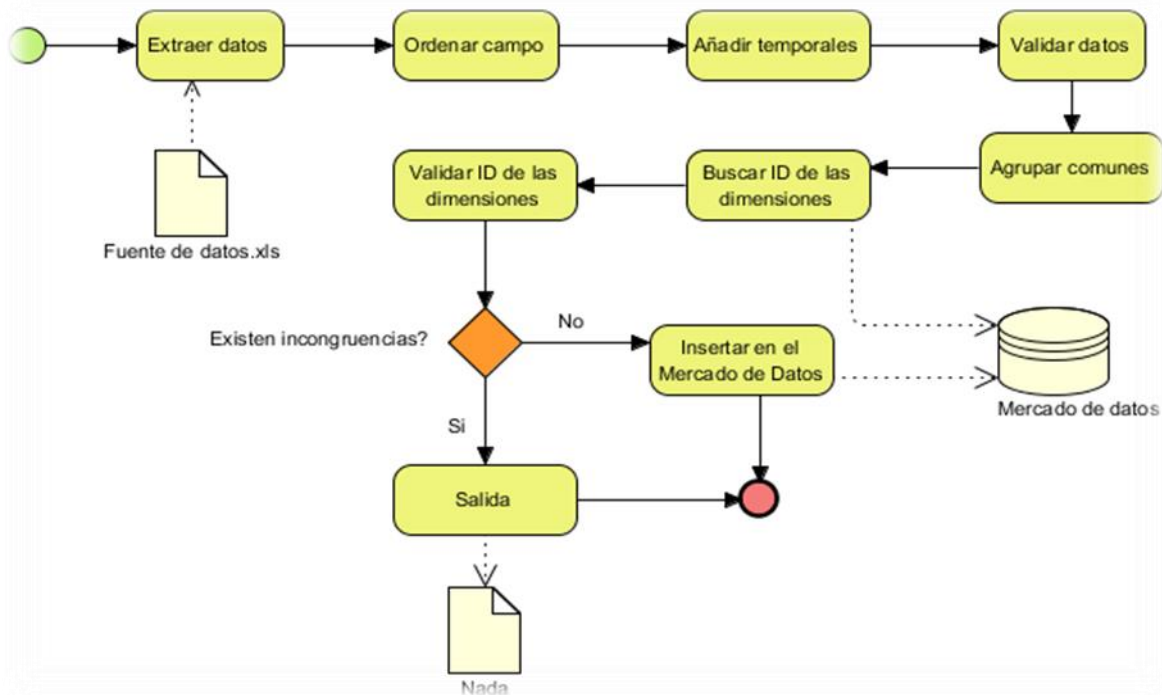


Figura 5. Diseño de la transformación para la carga de los datos del hecho compra

2.6.2 Diseño del subsistema de visualización

Para realizar el diseño del subsistema de visualización es necesario diseñar los reportes que responden a las necesidades de información del cliente y los cubos OLAP. Un reporte OLAP permite diseñar y emular un esquema de cubos, mediante el mapeo de los datos dimensionales. Además cualquier base de datos dimensional permite crear reportes en línea haciendo uso de dimensiones y métricas.

Diseño de los cubos OLAP

Existen diversas estructuras de datos a través de las cuales se puede representar la información de un MD, pero los cubos multidimensionales constituyen una de las más utilizadas. Un cubo multidimensional o hipercubo representa o convierte los datos planos que se encuentran en filas y columnas de una tabla. En el Mercado de datos Histórico de precios se tienen un cubo multidimensional, es decir, un cubo por cada tabla de hechos diseñada, siete dimensiones y dos medidas físicas, distribuidos de la siguiente manera:

El cubo **Compra** contiene siete dimensiones utilizables y dos medidas físicas. La figura 6 muestra el cubo compra con sus dimensiones y sus medidas asociadas.

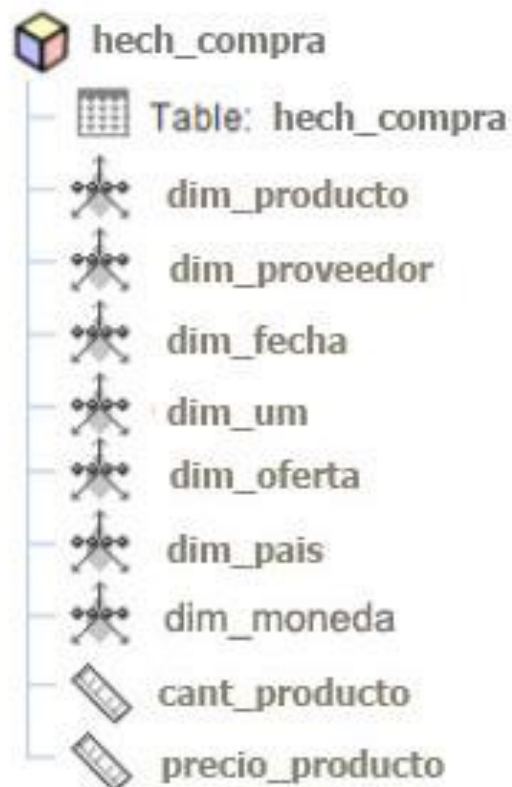


Figura 6. Cubo compra

2.7 Política de respaldo y recuperación

Con el objetivo de garantizar la persistencia de la información, se establece una política de respaldo y de recuperación de los datos.

Periodicidad de las salvas: las salvas de toda la información contenida en la base de datos se realizarán con una periodicidad anual. Se debe verificar la existencia de una copia de toda la información almacenada. La tabla involucrada en este proceso es la tabla compra identificada en el proceso de análisis con las tablas de dimensiones asociadas.

Backups existentes: cada año se realizarán varias salvas, una mensual y otra a fin de año. Se cuenta, además, con un servidor de respaldo con todos los datos en caso de ocurrencia de incidentes que atenten contra la seguridad de la entidad. Se debe realizar una copia de los datos en otros medios de almacenamiento, dígame memoria o DVD.

2.8 Esquema de seguridad

Para evitar el daño de los datos y de las funcionalidades de una aplicación informática es necesario el desarrollo de un esquema de seguridad estable que garantice la seguridad y protección de los datos. Esta seguridad se rige principalmente por los roles y permisos que los usuarios poseen para la interacción con el sistema.

Seguridad en la base de datos

Como complemento del esquema de seguridad mencionado anteriormente se definieron los roles mencionados a continuación para el acceso a la base de datos.

Tabla 7. Permisos de los usuarios en la base de datos

Usuarios	Permisos
Administrador	Tiene total acceso a todas las funciones de la base de datos
Administrador ETL	Tiene solamente permiso para realizar las funciones de ETL

Seguridad en la aplicación

Debido al sucesivo incremento de las aplicaciones desplegadas en el servidor BI de la Suite del Pentaho, así como también los usuarios que tienen acceso a las mismas se definen los siguientes roles para lograr una mayor organización a la hora de acceder a la aplicación.

Tabla 8. Permisos de los usuarios en la aplicación

Usuarios	Permisos
Administrador	Tiene acceso total al sistema
Analista	Tiene acceso de solo lectura al MD para el área de mercadotecnia

Después de analizar los usuarios y permisos sobre la base de datos y la aplicación no se puede dejar de mencionar que la herramienta Pentaho BI Server ofrece su propia seguridad dividida en tres aspectos fundamentales:

- ✓ **Seguridad de acceso a datos de objetos:** Contiene los usuarios, contraseñas, autorizaciones permitidas, recursos web y protección a datos.
- ✓ **Autenticación:** Está relacionado con el procesamiento de información interactiva de inicio de sección, es decir el nombre de usuario y contraseña, comparándola con la información recuperada del almacén de datos de seguridad.
- ✓ **Autorización de recursos web (URL):** Brinda protección a las URL para determinar si algún usuario determinado tiene acceso o no a alguna página. De esta operación se encarga el administrador de recursos web brindándole a cada usuario un permiso de seguridad delimitando el acceso o no a las páginas.

Conclusiones del capítulo

Se obtuvo el análisis y diseño del MD a partir de los requisitos funcionales, no funciones y de información definidos a partir de las necesidades del cliente, se diseñaron los subsistemas almacenamiento, integración y visualización que componen la solución permitiendo generar el modelo de datos, las transformaciones generales y el cubo OLAP. Para garantizar la seguridad del sistema se definieron los parámetros a partir de los roles que corresponden a los usuarios que interactuarán con el MD.

Capítulo 3 Implementación y validación

Introducción

Una vez realizado el análisis y diseño del Mercado de datos Histórico de precios y teniendo como guía la metodología utilizada, se procede a realizar la implementación física del sistema y el proceso de validación del mismo. Esta nueva etapa comprende la implementación de cada uno de los subsistemas que conforman el MD: subsistemas de almacenamiento, integración y visualización. Una vez concluido el proceso de construcción del Mercado de datos Histórico de precios, se hace necesaria la validación de la solución mediante la aplicación de técnicas y pruebas. Dichos aspectos, unidos a una correcta preparación de los usuarios finales para la explotación del MD, tributan al éxito en la toma de decisiones.

3.1 Implementación del subsistema de almacenamiento

Para la implementación del subsistema de almacenamiento fue necesaria la utilización de un gestor de base de datos. El SGBD utilizado fue PostgreSQL que permitió representar el diseño de la solución, mediante la creación de las tablas con sus respectivos atributos y relaciones.

El desarrollo exitoso de un modelo físico es un aspecto importante dentro de la construcción de un MD. El punto de partida para llevar a cabo este modelo es el modelo lógico, el modelo dimensional elaborado en el capítulo anterior. A continuación se describe lo relacionado con la implementación del modelo de datos físico en el Mercado de datos Histórico de precios describiendo la estructura de la base de datos conformada por los esquemas, tablas, secuencias.

3.1.1 Esquemas y tablas

Los esquemas y las tablas permiten brindar una idea general de la base de datos, así como analizar y profundizar los contenidos básicos de la misma. Para la solución propuesta se cuenta con siete tablas identificadas. Para una buena organización de las mismas se agruparon por esquemas, las dimensiones están contenidas dentro del esquema “dimensiones” y la tabla del hecho está contenida en el esquema “hecho”. A continuación se muestra la estructura de datos en el PostgreSQL.

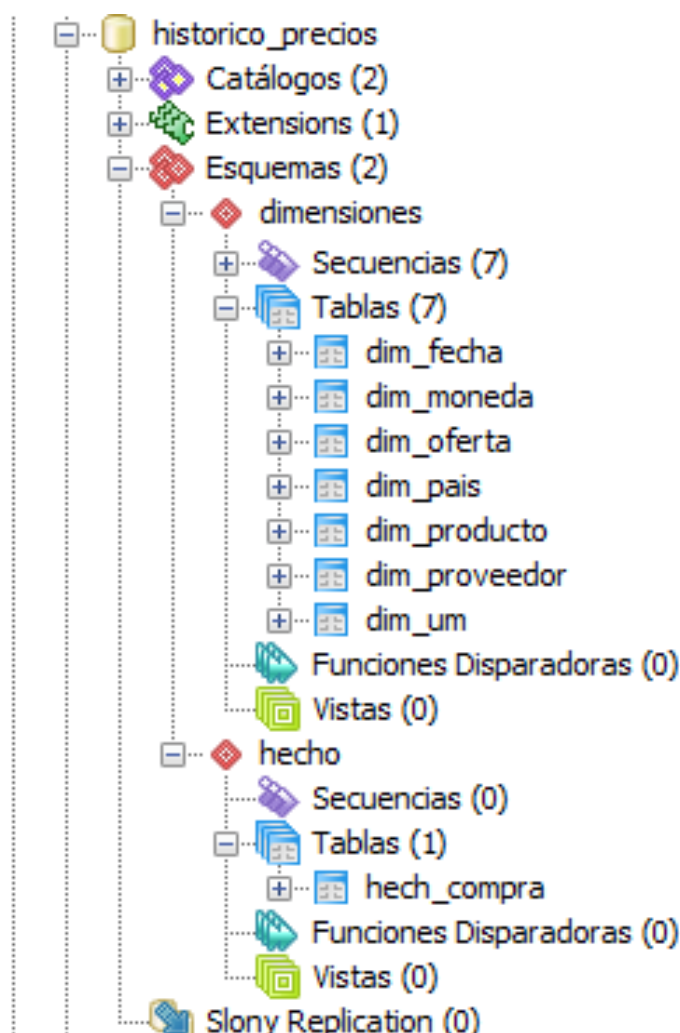


Figura 7. Estructura de datos

3.1.2 Restricciones y secuencias

Cuando se diseña una base de datos se debe reflejar fielmente el universo del discurso que se está tratando, o mejor, las restricciones existentes en el mundo real. Los componentes de una restricción son los siguientes:

- ✓ La operación de actualización (inserción, borrado o eliminación) cuya ejecución ha de dar lugar a la comprobación del cumplimiento de la restricción.
- ✓ La condición que debe cumplirse, la cual es en general una proposición lógica, definida sobre uno o varios elementos del esquema, que puede tomar uno de los valores de verdad (cierto o falso).
- ✓ La acción que debe llevarse a cabo dependiendo del resultado de la condición.
- ✓ Las restricciones son condiciones que se le aplican a una BD para que cumpla con ciertos parámetros. Las mismas pueden ser creadas automáticamente al definir una tabla (en el caso de las llaves primarias) o ser introducidas por el programador de la BD cuando se busca algo específico.

Existen 4 tipos de restricciones que se relacionan a continuación:

- ✓ **Clave foránea:** son las que referencian una clave de otra tabla, se usan para relacionar tablas diferentes.
- ✓ **Duplicidad:** implica que no debe haber dos valores iguales en la misma columna.
- ✓ **Clave primaria:** son valores que deben cumplir con un conjunto de restricciones: no tener valores nulos, ser únicos para cada tupla y ser necesarios.
- ✓ **Valor no nulo:** no debe existir ninguna casilla de la columna que esté vacía.

En el Mercado de datos Histórico de precios en cada tabla de dimensiones el tipo de clave es primaria y para las tablas de hechos son foráneas.

Las secuencias son atributos que se incrementan secuencialmente durante el ingreso de los datos. En el presente trabajo se definieron 7 llaves primarias y 7 secuencias.

3.2 Implementación del subsistema de integración

El proceso de integración de los datos consta de tres etapas fundamentales relacionadas entre sí: extracción, transformación y carga de los datos. Para identificar y corregir los problemas se realiza la limpieza de los datos que permite llenar valores ausentes y corregir errores. Una vez que los datos son transformados se cargan, poblando las dimensiones y el hecho que conforma la estructura del subsistema de almacenamiento del Mercado de datos Histórico de precios.

3.2.1 Implementación de las transformaciones

Las transformaciones constituyen un elemento básico dentro de la implementación del proceso de ETL. Una transformación está compuesta por pasos, que constituyen el elemento más pequeño de la transformación y se encuentran unidos a través de saltos. El principal componente utilizado en cada una de las transformaciones fue: entrada Excel con el paso entrada a tabla para leer la información de la fuente de datos. A continuación se muestran algunas de las transformaciones realizadas para desarrollar la solución:

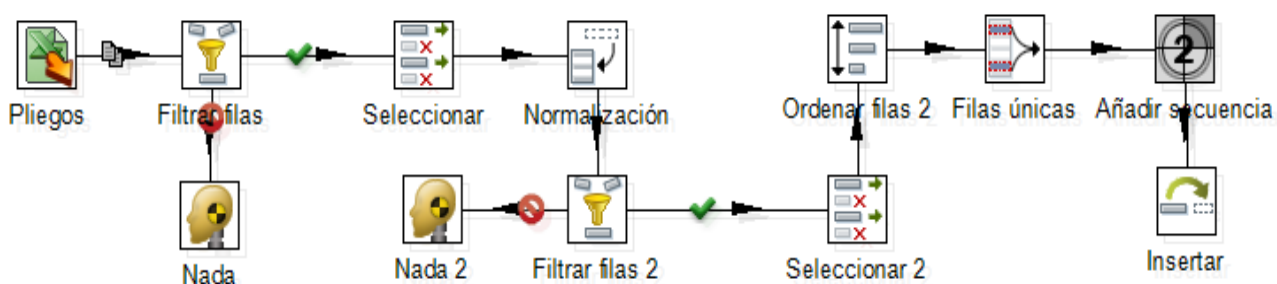


Figura 8. Transformación para la carga de la dimensión dim_país



Figura 9. Transformación para la carga de la dimensión dim_oferta

3.2.2 Implementación del trabajo

Al concluir la realización de todas las transformaciones necesarias para la carga de los datos, se realiza la implementación del trabajo, el cual se encarga de ejecutar todas las transformaciones en un orden lógico definido, primero las dimensiones y después el hecho. A continuación se muestra el trabajo realizado para desarrollar la solución:

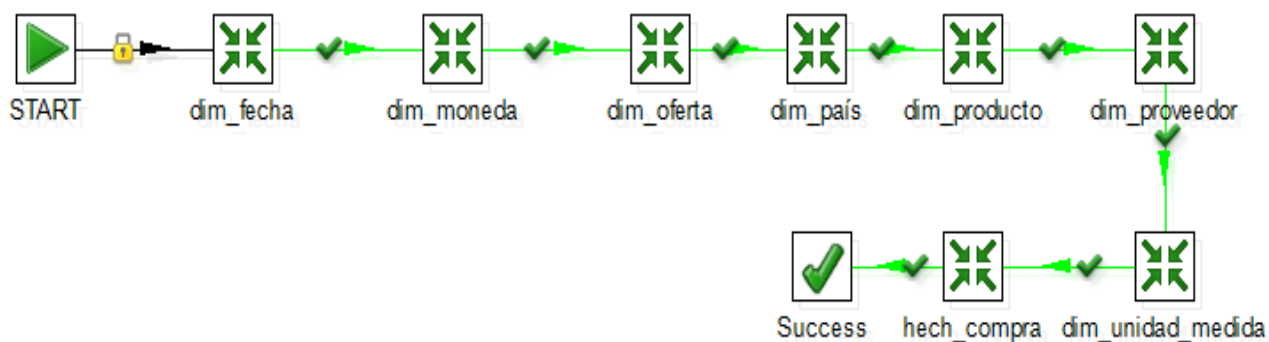


Figura 10. Transformación del trabajo general del mercado de datos Histórico de precios

3.3 Implementación del subsistema de visualización

Una vez realizada la carga de los datos y como parte de la implementación del tercer y último subsistema del Mercado de datos Histórico de precios, se realizó el cubo OLAP y los reportes candidatos, así como la configuración del control de acceso al sistema.

3.3.1 Cubos de datos

En el sistema se definió un cubo OLAP, que abarcó todas las dimensiones descritas anteriormente. Dicho cubo se nombrará de acuerdo al proceso de negocio con el cual está relacionado y contendrán las medidas necesarias para regir el hecho numérico en esta estructura y según los objetivos estratégicos a cumplir.

Capítulo 3: Implementación y validación

El cubo de datos se desarrolló mediante el uso de la herramienta *Schema Workbench*. El tipo de almacenamiento definido para este cubo es el ROLAP. El primer paso para la creación del cubo es la definición del mismo y donde se deciden cuáles son los atributos que son necesarios para realizar el análisis. Como segundo paso se especifica la estructura de las dimensiones diseñando los campos calculables, las medidas, junto con la creación de todas las tablas relacionales requeridas. Para poder realizar el análisis del cubo OLAP, debe publicarse dentro de la plataforma Pentaho. A continuación se muestra la interfaz al publicar el cubo en el Pentaho BI server.

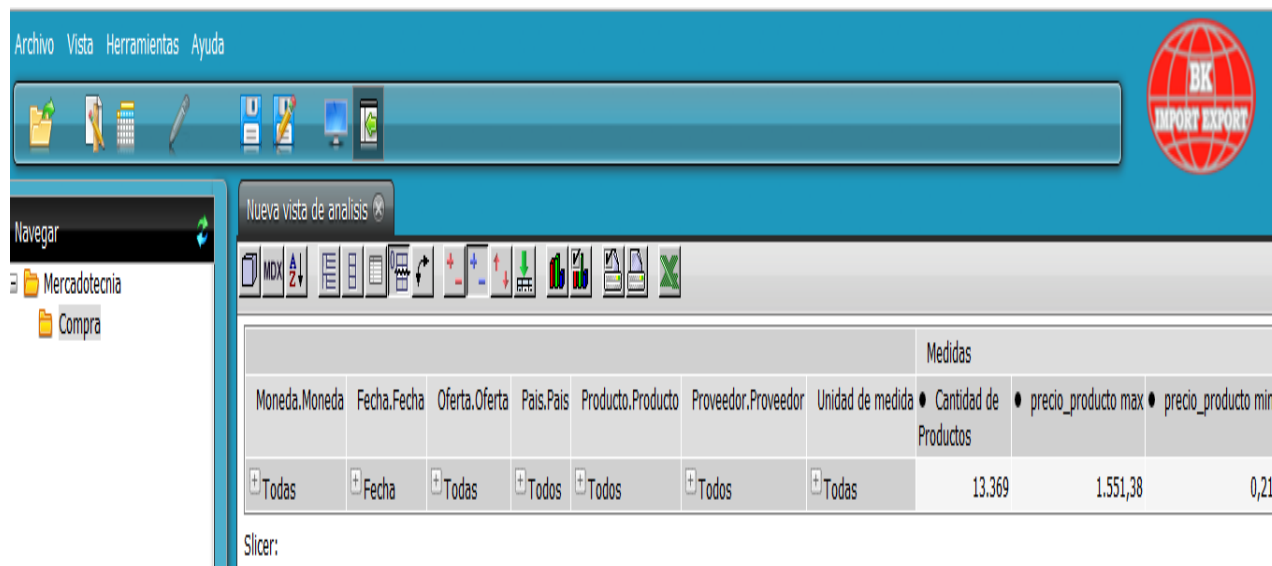


Figura 11. Interfaz de la vista de análisis

Vistas de análisis

La solución propuesta cuenta con 3 vistas de análisis, con el objetivo de satisfacer las necesidades de información de la Empresa BK Import/Export. El Pentaho BI-Server permite mostrar el resultado del análisis realizado, pues a través de la misma es posible ver los reportes y analizar la información mediante tablas de datos; permitiendo además, desplegar el cubo de información, y modificar las vistas de análisis o crear unas totalmente nuevas. Por otra parte, esta interfaz brinda la posibilidad de imprimir el reporte o salvarlo en un archivo de formato PDF o XLS.

Reportes

En la implementación de la solución se generaron 3 reportes candidatos útiles para el departamento de mercadotecnia, a partir de los datos registrados.

Para la definición de los reportes se utilizó la herramienta *Pentaho BI Server*. Esta herramienta permite consultar los datos de distintas fuentes y ponerlos a disposición de los usuarios en diferentes formatos (HTML, PDF, Microsoft Excel, y texto plano). Las herramientas de Pentaho permiten distribuir los reportes a los usuarios interesados, así como también publicarlos para que los usuarios puedan observar la información que necesitan. A continuación se muestra un reporte que responde a las necesidades del cliente.

Capítulo 3: Implementación y validación

Producto: ABANICO LIJADOR MANGO 6MM 60X30 G120					
Precio del producto	Proveedor	Día	Mes	Año	
3,12	NSV	23	enero	2015	
3,6	Apolo	14	enero	2015	
3,6	Apolo	17	febrero	2015	
3,6	Apolo	27	enero	2015	
4,1	Fersume S.L	17	diciembre	2014	
4,1	Fersume S.L	22	enero	2015	
9,22	Cronaser S.L	26	diciembre	2014	
Producto: ABANICO LIJADOR MANGO 6MM 60X30 G60					
Precio del producto	Proveedor	Día	Mes	Año	
3,08	Apolo	14	enero	2015	
3,08	Apolo	17	febrero	2015	
3,08	Apolo	27	enero	2015	
3,12	NSV	23	enero	2015	
4,51	Fersume S.L	17	diciembre	2014	
4,51	Fersume S.L	22	enero	2015	
9,22	Cronaser S.L	26	diciembre	2014	
Producto: ALICATE FEGEMU UNIVERSAL VDE-1000V 200MM					
Precio del producto	Proveedor	Día	Mes	Año	
9,27	Apolo	14	enero	2015	
9,27	Apolo	17	febrero	2015	
9,27	Apolo	27	enero	2015	
11,66	Fersume S.L	17	diciembre	2014	
11,66	Fersume S.L	22	enero	2015	
16,37	BSP	19	enero	2015	
18,42	Cronaser S.L	26	diciembre	2014	
21,43	NSV	23	enero	2015	
Producto: ALLGAS Mobile Pro Soldador Portatil, incluye botellas					
Precio del producto	Proveedor	Día	Mes	Año	
36,22	Rodabilisa	19	febrero	2015	
63,98	Guarconsa	18	febrero	2015	
148,26	NSV SL	9	diciembre	2014	
223,08	Apolo	14	enero	2015	
223,08	Apolo	17	febrero	2015	
274,75	Fersume S.L	17	diciembre	2014	

Figura 12. Reporte obtener el precio por proveedor, día, mes y año dado un producto

3.4 Pruebas

Las pruebas se centran principalmente en la evaluación o la valoración de la calidad del producto y representan un elemento crítico para la garantía del mismo. Es una actividad en la cual un sistema o uno de sus componentes se ejecutan en circunstancias previamente especificadas, los resultados se observan, se registran y se realiza una evaluación de algún aspecto.

El objetivo de la etapa de pruebas es garantizar la calidad del producto desarrollado. Además, esta etapa implica:

1. Verificar la interacción de los componentes.
2. Verificar la integración adecuada de los componentes.
3. Verificar que todos los requisitos se han implementado correctamente.
4. Identificar y asegurar que los defectos encontrados se han corregido antes de entregar el software al cliente.

A continuación se exponen algunas de las pruebas que pueden ser utilizadas para la validación de un producto de software:

Capítulo 3: Implementación y validación

- ✓ **Prueba unitaria:** Es el proceso de probar los componentes individuales de la solución. El propósito es identificar diferencias entre la especificación de los artefactos y el comportamiento real de cada módulo.
- ✓ **Prueba de integración:** Es el proceso en el cual los componentes son agregados para crear componentes más grandes. Es la prueba realizada para mostrar que aunque los componentes hayan pasado satisfactoriamente las pruebas de unidad, la integración de los componentes es incorrecta.
- ✓ **Prueba de sistema:** Se refiere al comportamiento del sistema integrado. La prueba de sistema se aplica generalmente para probar los requisitos no funcionales de la solución.
- ✓ **Pruebas funcionales:** Aseguran el trabajo apropiado de los requisitos funcionales, incluyendo la navegación, entrada de datos, procesamiento y obtención de resultados.
- ✓ **Pruebas de aceptación:** El objetivo de las pruebas de aceptación es validar que un sistema cumpla con el funcionamiento esperado y permitir 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 Pruebas funcionales

Las pruebas funcionales se realizaron a través de los casos de prueba basados en CU con el objetivo de validar si el comportamiento observado del software cumple o no con sus especificaciones.

El propósito de un caso de prueba es especificar una forma de probar el sistema, incluyendo las entradas para la validación, los resultados esperados y las condiciones bajo las que ha de probarse.

Para validar los requerimientos del sistema de este trabajo se le realizan casos de prueba a cada CUI, con el objetivo de comprobar la disponibilidad de los perfiles de análisis y los indicadores a medir, así como también verificar el cumplimiento de los requisitos de información a través de los reportes candidatos. Se realizaron tres casos de pruebas evidenciándolo en el documento *Diseno_de_casos_de_prueba.xls* almacenado en el expediente de proyecto. A continuación se describe el caso de prueba asociado al CUI Obtener el precio por proveedor, día, mes y año dado un producto.

Capítulo 3: Implementación y validación

Escenario	Descripción	Variables de Entrada	Variables de Salida	Respuesta del sistema	Flujo central
EC 1: Obtener el precio por proveedor, día, mes y año dado un producto.	Muestra la información de los productos y proveedores.	Proveedor Día Mes Año Producto Precio	Precio	El sistema muestra todas las variables disponibles para los análisis, ubicados en las filas y las columnas que pueden ser visualizadas para cada reporte.	Se autentica. Se despliega hacia la derecha el componente ubicado en el lateral izquierdo que contiene el navegador. Se selecciona Mercadotecnia. Se selecciona Compra. En la parte inferior izquierda se selecciona el reporte deseado. En el área de trabajo se visualiza la tabla correspondiente al reporte.

Figura 13. Caso de prueba realizado al RI obtener el precio por proveedor, día, mes y año dado un producto

No	Nombre de campo	Clasificación	Valor Nulo	Descripción
1	Proveedor	Lista desplegable	No	Describe el universo de valores bajo los cuales puede clasificarse la información atendiendo a los Proveedores
2	Día	Lista desplegable	No	Describe el universo de pregunta bajo los cuales puede clasificarse la información atendiendo al Día
3	Mes	Lista desplegable	No	Describe el universo de pregunta bajo los cuales puede clasificarse la información atendiendo al Mes
4	Año	Lista desplegable	No	Describe el universo de pregunta bajo los cuales puede clasificarse la información atendiendo al Año
5	Producto	Lista desplegable	No	Describe el universo de pregunta bajo los cuales puede clasificarse la información atendiendo al Producto
6	Precio	Valor fijo	No	Valores estadísticos que se cargan calculados.

Figura 14. Descripción de las variables

Resultado de las pruebas

El MD fue probado en una iteración a través de pruebas de caja negra. En esta iteración no se detectaron no conformidades, lo que evidencia el correcto funcionamiento del sistema desarrollado, cumpliendo satisfactoriamente las restricciones y necesidades dadas por el cliente. Los resultados antes descritos lo constituye el acta de liberación, que puede ser consultada en el Anexo #1.

3.4.2 Pruebas de aceptación

El MD desarrollado fue presentado al cliente, luego de haber interactuado con la aplicación y comprobado su funcionalidad se mostró satisfecho con el producto entregado. Prueba de lo planteado anteriormente constituye el acta de aceptación emitida (ver Anexo #4), en la cual consta que el MD realizado cumple con las condiciones de calidad y satisface las necesidades plasmadas por el cliente, además logra la centralización, consistencia y disponibilidad que contribuye al análisis estadístico del histórico de precio en el sistema Xedro-Apside.

3.4.3 Aplicación de la técnica de ladov

Para conocer el nivel de satisfacción del cliente al interactuar con la aplicación fue utilizada la técnica de ladov. Fue necesario formular tres preguntas cerradas en un cuestionario (ver Anexo #5), que se relacionan a través del Cuadro Lógico de ladov [36].

La técnica de ladov en su versión original fue creada por su autor para el estudio de la satisfacción por la profesión en carreras pedagógicas. Posteriormente se ha generalizado su uso en otras áreas.

4. Luego de haber visto el sistema, le gusta la solución desarrollada	1. ¿Cree usted que la forma en la que se almacena la información en la empresa facilita el análisis de los precios de los productos?								
	No			No sé			Sí		
	2. ¿Usted considera que el Mercado de datos desarrollado garantizan la centralización, consistencia y disponibilidad de la información permitiendo el análisis estadístico del Histórico de precios?								
	Sí	No sé	No	Sí	No sé	No	Sí	No sé	No
Me gusta mucho	1	2	6	2	2	6	6	6	6
Me gusta más de lo que me disgusta	2	2	3	2	3	3	6	3	6
Me da lo mismo	3	3	3	3	3	3	3	3	3
Me disgusta más de lo que me gusta	6	3	6	3	4	4	3	3	4
No me gusta nada	6	6	6	6	4	4	6	6	5
No sé decir	2	3	6	3	3	3	6	6	4

Figura 15. Cuadro Lógico de ladov

El número resultante de la interrelación de las tres preguntas indica la posición de cada encuestado en la escala de satisfacción siguiente:

1. Clara satisfacción.
2. Más satisfecho que insatisfecho.
3. No definida.
4. Más insatisfecho que satisfecho.
5. Clara insatisfacción.
6. Contradictoria.

Capítulo 3: Implementación y validación

Si un encuestado responde a la pregunta uno "No", se va a la zona izquierda del cuadro, debajo de la pregunta uno, donde aparece No. Si a la pregunta dos responde "No sé" se busca el "No sé" aparece debajo del No anterior. Si a la pregunta tres responde: "Me disgusta más de lo que me gusta" entonces se busca en las filas, a la izquierda, la casilla donde aparece esa respuesta y se busca el punto donde se interceptan la fila "Me disgusta más de lo que me gusta" con la columna "No sé". El resultado de dicho encuestado es "3", que equivale a "satisfacción no definida". Así se procede con cada usuario de la muestra, en dependencia de sus respuestas. De esta forma se van clasificando en las seis categorías antes mencionadas.

Para obtener el índice de satisfacción grupal (ISG) se trabaja con los diferentes niveles de satisfacción que se expresan en una escala numérica que oscila entre +1 y -1 de la siguiente forma [36]:

Tabla 9. Niveles de satisfacción

+1	Máximo de satisfacción	A
0,5	Más satisfecho que insatisfecho	B
0	No definido y contradictorio	C
-0,5	Más insatisfecho que satisfecho	D
-1	Máxima insatisfacción	E

La satisfacción grupal se calcula por la siguiente fórmula:

$$\text{ISG} = \frac{A(+1) + B(+0,5) + C(0) + D(-0,5) + E(-1)}{N}$$

En esta fórmula A, B, C, D, E, representan el número de encuestados con índice individual (1; 2; 3; 4; 5 ó 6) y N representa el número total de la muestra. El índice grupal arroja valores entre + 1 y -1. Los valores que se encuentran comprendidos entre -1 y -0,5 indican insatisfacción; los comprendidos entre -0,49 y + 0,49 evidencian contradicción y los que caen entre 0,5 y 1 indican que existe satisfacción.

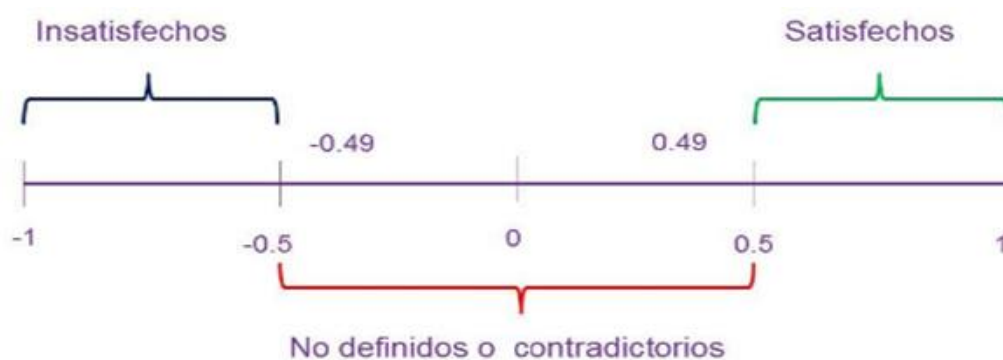


Figura 16. Nivel de satisfacción

Capítulo 3: Implementación y validación

Después del análisis del cuestionario aplicado a los especialistas de mercadotecnia de la Empresa BK Import/Export se obtuvo el resultado siguiente:

Tabla 10. Resultados de las escalas de satisfacción

Total de usuarios de la muestra (N)	4
Clara satisfacción	2
Más satisfecho que insatisfecho	0
No definida	0
Más insatisfecho que satisfecho	0
Clara insatisfacción	0
Contradictoria	2

A partir de estos resultados para calcular el ISG, las variables de la fórmula tomarían los siguientes valores: A=2; C=2.

Calculando el ISG quedaría de la siguiente forma:

$$\text{ISG} = \frac{2(+1)+2(0)}{4} = 0.5$$

El resultado está dentro del intervalo [0.5, 1] por tanto el nivel de satisfacción de la muestra es de "Satisfecho". Por los resultados antes evidenciados se puede afirmar que se cumplió el problema de investigación planteado. La solución desarrollada garantiza la consistencia, centralización y disponibilidad de la información para el análisis estadístico en el sistema Xedro-Apside.

Conclusiones del capítulo

A partir del diseño propuesto se obtuvo la implementación del Mercado de Datos Histórico de precios mediante el desarrollo de los subsistemas que componen la solución permitiendo el análisis de los históricos de precios y la generación de reportes. La calidad del MD implementado se comprobó mediante la aplicación de pruebas internas garantizando su correcto funcionamiento, además fueron realizadas pruebas de aceptación y fue aplicada la técnica de ladov, con la que se obtuvo un índice de satisfacción por parte del cliente.

Conclusiones generales

La investigación realizada cumple los objetivos planteados inicialmente mediante el desarrollo del Mercado de datos Histórico de precios, arribando a las siguientes conclusiones:

- ✓ Se evidenció la necesidad de desarrollar un MD para el análisis de histórico de precios que permita mantener la información centralizada, consistente y disponible utilizando la metodología de DATEC, el modo de almacenamiento ROLAP y el modo de representación en esquema Estrella.
- ✓ Se obtuvo el modelo de datos, el diseño del cubo OLAP y de las transformaciones a partir de los requisitos de información, funcionales y no funcionales definidos con el cliente.
- ✓ La solución implementó los subsistemas de integración, almacenamiento y visualización permitiendo el análisis del histórico de precios y la generación de reportes.
- ✓ Se validó el MD a partir de la aplicación de pruebas internas y de aceptación con el cliente garantizando su correcto funcionamiento y se demostró su satisfacción aplicando la técnica de ladov.

Recomendaciones

Con el propósito de mejorar la propuesta realizada en este trabajo, se sugiere:

- ✓ Poblar el mercado de datos con toda la información existente en el departamento de mercadotecnia de la Empresa BK Import/Export.
- ✓ Realizar la integración del Mercado de Datos Histórico de precios con el Sistema Xedro-Apside.

Referencias bibliográficas

1. Ganapati, P. *Sun Microsystems Data Warehouse in 2009 Guinness Book*. Available from: <http://www.wired.com/2008/09/sun-microsystem/>.
2. Natixis *Natixis basa su mensajería SWIFT e infraestructura de datos en la solución integrada de Informatica PowerCenter y B2B Data Transformation*. 2010.
3. Torres Ilanes, J., Alfonso Collado, Andy Carlos, *Mercado de Datos para la Dirección de Colaboración Económica del Ministerio de Comercio Exterior y la Inversión Extranjera*. 2012, Universidad de las Ciencias Informáticas: Universidad de las Ciencias Informáticas.
4. Valdés Rodríguez, Y., *Mercado de datos que contribuya a la toma de decisiones en la Dirección de Cuadros de la Administración Provincial de Artemisa*. 2012, Universidad de las Ciencias Informáticas: Universidad de las Ciencias Informáticas.
5. Rodríguez Cabrera, M., Vilaró González, Yaillet de la Caridad *Análisis, diseño e implementación del mercado de datos para los indicadores específicos de la ocupación para la Oficina Nacional de Estadísticas*. 2010, Universidad de las Ciencias Informática.
6. Competencias, U.E.O.a. *Inteligencia de Negocios (Business Intelligence)*. 2015; Available from: <http://erangel.foroactivo.mx/t352p15-5-inteligencia-de-negocios-business-intelligence>.
7. CASTILLO, D.O.A.F.Y.J.N.P. *Estado actual de las tecnologías data warehousing y OLAP aplicadas a bases de datos espaciales* Luís Joyanes Aguilar. [cited 20 de Enero 2015]; Available from: <http://novella.mhhe.com/sites/dl/free/8448118952/540197/ActasVol2SISOFT2006.pdf#page=113>.
8. ABBEY, M.J.C.Y.M., *Oracle Data Warehousing*. 2002.
9. Inmon, W.H., *Using the DataWareHouse*. 1997.
10. Buigues, E.B.d.A. *Data Warehouse y Data Warehousing*. 2011.
11. VELASCO, R.H. *Almacenes de datos (Datawarehouse)*. [cited 23 de Enero 2015]; Available from: <http://www.rhernando.net/modules/tutorials/doc/bd/dw.pdf>.
12. KONCILIA, R.W.Y.C., *DATA WAREHOUSES AND OLAP*. 2007.
13. Kimball, R., *The Data Warehouse Lifecycle Toolkit*. 2008.
14. *TopicosAdmonDatOrientWeb*. Available from: <http://topiavandebasedatos.wikispaces.com/3.3+Mercados+de+datos+%28Data+Mart%29>.
15. Bill Inmon, D.m., Data warehouse, Datamartist Tool, Gartner, Ralph Kimball. *Data warehouse or Data Mart?* [cited 25 de Enero 2015]; Available from: <http://www.datamartist.com/data-warehouse-vs-data-mart>.
16. IMHOFF, C., *Mastering Data Warehouse Design, Relational and Dimensional Techniques*. 2003.
17. SINNEXUS. *Datamart*. [cited 26 de Enero 2015]; Available from: http://www.sinnexus.com/business_intelligence/datamart.aspx.

Referencias bibliográficas

18. Hernández, I.Y., *PROPUESTA DE METODOLOGIA PARA EL DAsARROLLO DE AD EN DATEC*. 2010: La Habana.
19. M, R. *Sobre PostgreSQL*. [cited 31 de Enero 2015]; Available from: http://www.postgresql.org/es/sobre_postgresql.
20. pgAdmin. *PostgreSQL Tools*. 2012; Available from: <http://www.pgadmin.org>.
21. asociados., C.y., *Providing OLAP to user-analysts: An IT mandate*. 1993.
22. VÁZQUEZ, F.P.Y.G. *Relevamiento: Diseño Físico de Sistemas OLAP*. [cited 28 de enero 2015]; Available from: http://www.fing.edu.uy/~fpiedrab/downloads/Physical_OLAP_Design.pdf.
23. V., P.F.P. *Sistemas de Soporte a la toma de Decisiones*. [cited 30 de Enero 2015]; Available from: <http://palomo.usach.cl/bdnc/2005-02/Presentaciones/U3-1-OLAP.pdf>.
24. Becker, S., *Data Warehousing and Web Engineering*. 2003: idea Group publishing.
25. Chen, P.P.-S., *The entity-relationship model—toward a unified view of data*. 1976.
26. Dataprix. *OLAP, MOLAP y ROLAP*. [cited 31 de Enero 2015]; Available from: <http://www.dataprix.com/olap-rolap-molap>.
27. Rivadera, G.R., *La metodología de Kimball para el diseño de almacenes de*. 2010.
28. González, Y.P., *Sistema de información de gobierno Mercado de datos Inmigración y extranjería*. Ciudad de la Habana. 2011.
29. Computación., M.y.T.d.I.P.D.d.S.I.y. and E.-F. . *Introducción a Herramientas CASE y System Architect* [cited 31 de Enero 2015]; Available from: http://users.dsic.upv.es/asignaturas/eui/mtp/doc-practicas/intro_case_SA.pdf.
30. López, P. *Herramienta CASE Visual Paradigm*. [cited 2 de febrero 2015]; Available from: <http://ocw.unican.es/enseñanzas-tecnicas/ingenieria-del-software-i/practicas-1/is1-p01-trans.pdf>.
31. Corporation., P. *Pentaho. Powerful Analytics Made Easy*. 2014; Available from: <http://www.pentaho.com/explore/pentaho-data-integration/>.
32. Madrid, U.p.d. *Comparativa de Herramientas UML de libre distribución*. [cited 2 de Junio 2015]; Available from: <http://www.visual-paradigm.com/product/vpuml>.
33. *DataCleaner The premier data quality solution*. 2014.
34. *Schema_workbench*. 2014.
35. *Pentaho Reporting*. 2014; Available from: <https://github.com/pentaho/pentaho-reporting>.
36. López Rodríguez, A., González Maura, Viviana *La técnica de ladov. Una aplicación para el estudio de la satisfacción de los alumnos por las clases de educación física*. 2002.

Anexos

Anexo #1. Acta de liberación del Mercado de datos Histórico de precios

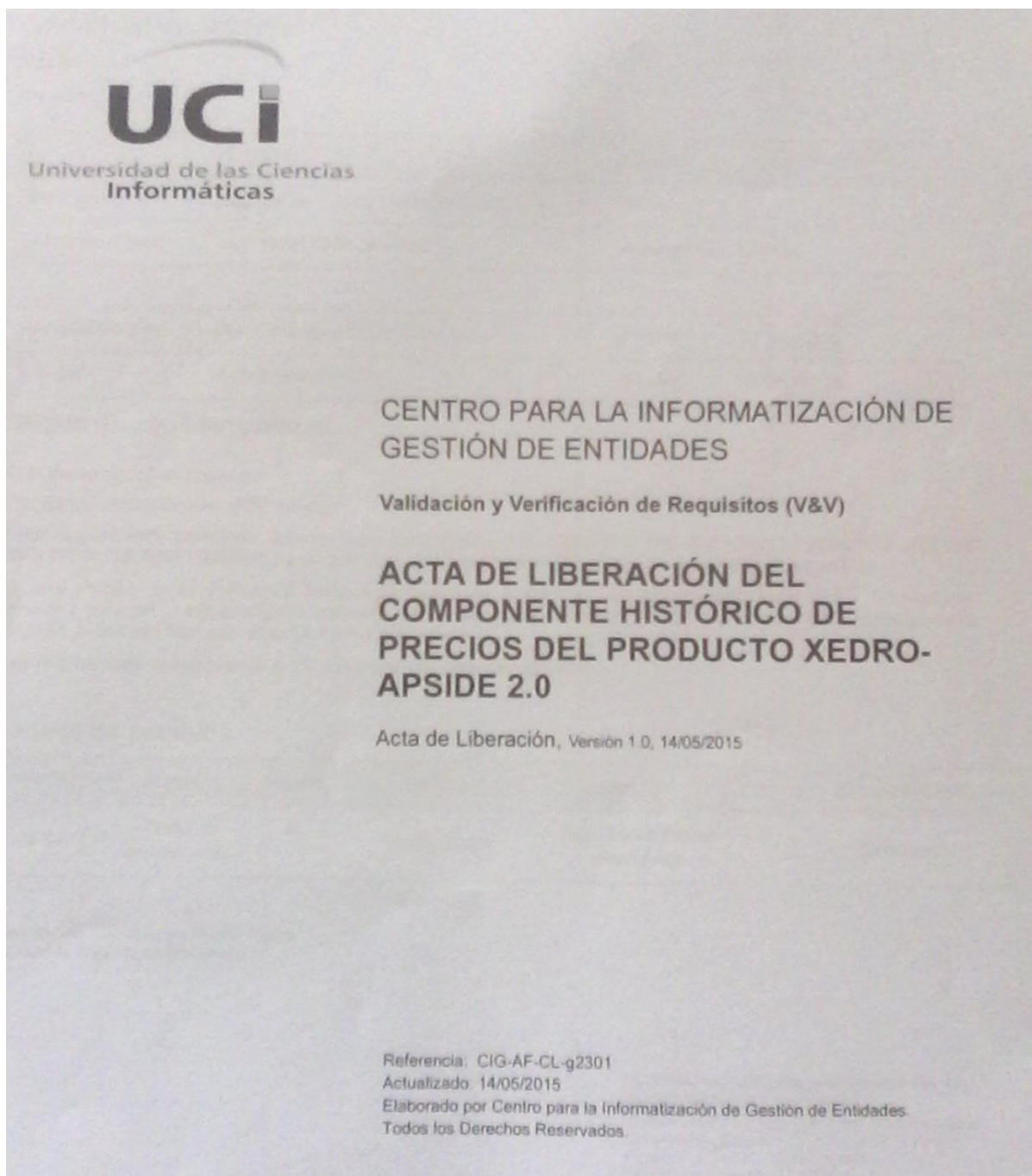


Figura 17. Página principal del documento de liberación del MD

Anexo #2. Acta de liberación del Mercado de datos Histórico de precios

UCI | CIG-AF-CL09001 : Acta de liberación

2 Datos del producto

Emitida a favor de: Histórico de Precios del Sistema de Gestión de Importación y Exportación BK Import/Export
 Fecha de emisión del acta: 14/05/2015
 Responsable: Pablo M. Choy Oliva
 Cargo: Estudiante

2.1 Clasificado como:

- Aplicación

2.2 Detalle de los elementos probados y su estado final:

Artefacto	Estado Final
Aplicación	0 No Conformidades

2.3 Cantidad de iteraciones:

Para la revisión se emplearon un total de 1 iteración de trabajo para lograr el resultado de 0 (cero) No Conformidad.

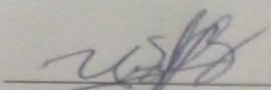
Artefacto	Versión	Estado final	Cantidad Iteraciones	Tipos de pruebas realizadas	Fecha de liberación
Aplicación	Versión 1.0	No Conformidades	1 iteración total	Pruebas Funcionales	14/05/2015

CENTRO DE INFORMATIZACIÓN DE LA GESTIÓN DE ENTIDADES
 Universidad de las Ciencias Informáticas
 Carretera a San Antonio Km 2 1/2, Tomera,
 Boyeros, Ciudad de La Habana, Cuba
 Teléfono: + 53 (7) 537 3680
 E-mail: software.gestion@uci.cu

Figura 18. Resultados de los elementos probados y su estado final

Anexo #3. Acta de liberación del Mercado de datos Histórico de precios

UCI | CIG-AF-CL09001 : Acta de liberación


 Ing. Yisel Niño Benítez
 Asesor de Calidad

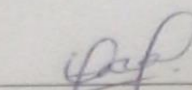

 Ing. Virtudes M. Figueredo Lara
 Jefe de Proyecto

Figura 19. Firmas de los involucrados en el acta de liberación

Anexo #4. Acta de aceptación del Mercado de datos Histórico de precios

UCI Universidad de las Ciencias Informáticas

ACTA DE ACEPTACIÓN

En cumplimiento del **Convenio de colaboración** de la Empresa Comercial BK Import/Export y en función de la ejecución del proyecto: Sistema de Gestión Comercial de Importación, se han efectuado las actividades siguientes:

- Revisión y Aprobación de los Requisitos de Información por el Cliente.
- Realización de pruebas exploratorias al Mercado de Datos desarrollado para realizar el análisis del Histórico de Precios.

La Parte Cliente, luego de haber revisado los productos de trabajo determina que aceptan la solución realizada.

Entrega		Recibe	
Nombre y apellidos:	Pablo Miguel Choy Oliva	Nombre y apellidos:	<i>Elon Estrella</i>
Cargo:	Estudiante de la Universidad de las Ciencias Informáticas	Cargo:	<i>Director de Mercadería</i> Empresa Comercial BK Import/Export



Ing. Virtudes Milagro Figueredo Lara
Jefe de Proyecto 

Fecha: 20/05/2015



Figura 20. Acta de aceptación

Anexo #5. Cuestionario de los especialistas

Cuestionario de los Especialistas

Estimado especialista:

Lea con cuidado cada pregunta antes de responder. En este cuestionario NO DEBE PONER SU NOMBRE. Le agradecemos su participación y franqueza al decir honestamente lo que piensa sobre lo que le preguntamos.

Preguntas:

1. ¿Cree usted que la forma en la que se almacena la información en la empresa facilita el análisis de los precios de los productos?
Sí__ No__ No sé__
2. ¿Usted considera que el Mercado de datos desarrollado garantizan la centralización, consistencia y disponibilidad de la información permitiendo el análisis estadístico del Histórico de precios? Sí__ No__ No sé__
3. ¿Cómo el departamento garantiza que los cambios realizados en la información por un especialista, sea consultada por el resto del personal?
No se garantiza__ Se copian información entre todos__ No es necesario ver los cambios__ Otra vía __
4. Luego de haber visto el sistema, le gusta la solución desarrollada.
__ Me gusta mucho
__ Me gusta más de lo que me disgusta
__ Me da lo mismo
__ Me disgusta más de lo que me gusta
__ No me gusta nada
__ No sé decir
5. ¿Cree usted que sería beneficioso para el departamento contar con este sistema que permite mantener la información disponible en todo momento y que los cambios realizados por una persona sean vistos por el resto de los especialistas?

Figura 21. Cuestionario de los especialistas

Anexo #6. Vista principal de la aplicación



Figura 22. Vista principal de la aplicación