

Universidad de las Ciencias Informáticas

FACULTAD 6



Mercado de datos Unión de Jóvenes Comunistas para la Facultad 6 de la Universidad de las Ciencias Informáticas

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

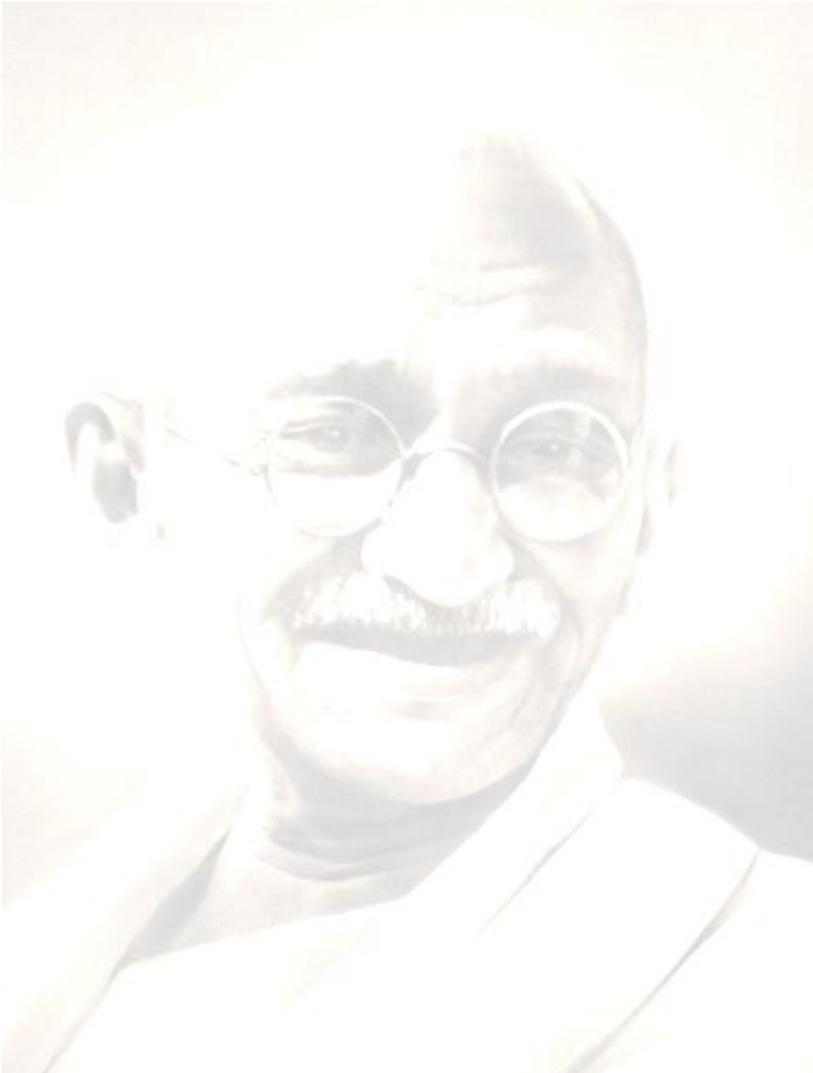
AUTORES:

Olivia Anet Silva Puente
Lianfranco Pinillo Vélez

TUTORES:

Ing. Fabian López García
Ing. Leonel Pérez Nieblas

La Habana, Junio de 2012
“Año 54 de la Revolución”



"Nuestra recompensa se encuentra en el esfuerzo y no en el resultado. Un esfuerzo total es una victoria completa."

Mahatma Gandhi

(1869-1948)

DECLARACIÓN DE AUTORÍA

Declaramos ser autores de la presente tesis “Mercado de datos Unión de Jóvenes Comunistas para la Facultad 6 de la Universidad de las Ciencias Informáticas” y reconocemos a la Universidad de las Ciencias Informáticas (UCI) los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste firmamos la presente a los ____ días del mes de _____ del año 2011.

Olivia Anet Silva Puente

Firma de la autora

Leonel Pérez Nieblas

Firma del tutor

Lianfranco Pinillo Vélez

Firma del autor

Fabian López García

Firma del tutor

DATOS DE CONTACTO

Ing. Leonel Pérez Nieblas.

Instructor recién graduado.

Graduado en el 2011 de Ingeniero en Ciencias Informáticas de la Universidad de las Ciencias Informáticas (UCI).

e-mail: lnieblas@uci.cu

Ing. Fabian López García

Instructor recién graduado.

Graduado en el 2011 de Ingeniero en Ciencias Informáticas de la Universidad de las Ciencias Informáticas (UCI).

e-mail: flgarcia@uci.cu

AGRADECIMIENTOS

De Olivia:

A mis abuelos por enseñarme el camino a seguir durante estos años.

A mis padres por brindarme su apoyo todo este tiempo.

A mi tía mami por enseñarme a confiar.

A todo el resto de la familia, en especial a mis hermanas y mis abuelos Daisy y Roberto.

A Ramiro, por soportar mis majaderías.

A Teresa y a Masiel, por brindarme su confianza.

A mis viejos amigos Laura y Popi, por poder contar con su ayuda.

A mi amiga Karla, por los momentos compartidos juntos.

A Yuri, Yaimé y demás muchachitas del apartamento.

A mis queridos tutores, por su ayuda.

A mi compañero de tesis, porque sin él este resultado no hubiese sido posible.

En fin, a todos los que de una forma u otra han tenido que ver con este trabajo, les agradezco de todo corazón.

De Lianfranco:

Si comparamos el desarrollo de una persona con una planta, veremos que existen puntos muy similares, este humilde ser, quien un día fue semilla, quiere agradecer a todos los que pusieron un granito de abono en su crecimiento y evolución como persona y profesional durante todo este tiempo. Como rayo de sol tan necesario para crecer siempre estuvo ahí mi familia, a mi madre querida por tantos años de amor y apoyo incondicional, ésta va para ti mulata. Papá espero que estés orgulloso y gracias por enseñarme que las barreras solo existen para quienes las quieren ver, al menos en esta vida. Hoy tanto como sonrío, también lamento no poder compartir mi alegría con esos ángeles que siempre he sentido en mi guarda, mis abuelos ausentes Ramón, Rafael y Cándida, gracias a Dios me queda una testigo que siempre ha estado ahí desde la cuna: mi abuela Julia, besos para ti. Espero también que algunos más jóvenes se sientan igual de orgullosos este día, mi hermano te quiero. Sin dejar a nadie fuera mencionaré a otra gran familia, de los cuales aprendo siempre algo cada día: mis amigos, primos, compañeros de clase, por mencionar algunos: Luis, Noslen, Ariel, en fin gracias a todos, pues de seguro les digo que este momento es la suma de todo su amor y apoyo.

Por último y no menos importante agradecer a mi dúo por ser como es, una chica extraordinariamente preocupada y dedicada, nunca cambies. Te espera un buen futuro. A mi UCI, si así es, siempre te sentiré mía. A la Revolución Cubana, que permite que un muchacho de Párraga, sin computadora y sin grandes recursos pueda formarse como todo un profesional de este país. A los tutores de la tesis, el tribunal y la oponente por sus observaciones que tributaron a la calidad de este trabajo.

DEDICATORIA

A mis abuelos que han sido los merecedores de este título.

Olivia

Este Trabajo de Diploma va dedicado a todos los que creyeron en mí y me brindaron su apoyo, sobre todo mi madre, te quiero vieja.

Lianfranco

RESUMEN

La Unión de Jóvenes Comunistas es una organización política que agrupa de forma selectiva a la vanguardia juvenil cubana. La misma cuenta con comités de base en todos los centros estudiantiles, laborales y unidades militares del país. En la facultad 6 de la Universidad de las Ciencias Informáticas la información referente a las reuniones de sus comités de base está siendo gestionada por un sistema desarrollado en el Centro de Tecnologías de Gestión de Datos, conocido con el nombre de Sistema de Recogida de Información. Este sistema no permite analizar la información existente, lo que dificulta la toma de decisiones para los directivos de esta organización y resta utilidad a los datos almacenados. El objetivo de esta investigación es desarrollar un mercado de datos que permita centralizar toda la información referente a las reuniones de esta organización, para facilitar el análisis estadístico y apoyar el proceso de toma de decisiones.

PALABRAS CLAVES: mercado de datos, toma de decisiones, Unión de Jóvenes Comunistas.

TABLA DE CONTENIDOS

INTRODUCCIÓN 1

CAPÍTULO 1: FUNDAMENTOS TEÓRICOS DE LOS ALMACENES DE DATOS..... 5

1.1 Introducción 5

1.2 Almacén de datos 5

1.2.1 Características de un almacén de datos 6

1.2.2 Componentes de un almacén de datos 6

1.3 Mercado de datos 8

1.3.1 Ventajas de los mercados de datos 9

1.4 Etapas de desarrollo de un almacén de datos..... 9

1.4.1 Análisis y diseño 9

1.4.2 Extracción, transformación y carga 9

1.4.3 Inteligencia de Negocio 10

1.5 Modelo de datos 11

1.5.1 Modelo Entidad-Relación 11

1.5.2 Modelo Dimensional..... 11

1.6 Tipos de modelado de un almacén de datos..... 11

1.6.1 Componentes de los esquemas de modelado..... 13

1.7 Modos de almacenamiento de datos..... 14

1.8 Metodologías para el desarrollo de un almacén de datos 15

1.8.1 Metodología seleccionada para el desarrollo del mercado de datos 16

1.9 Aplicaciones existentes 17

1.9.1 Aplicaciones existentes en el mundo que utilizan almacenes de datos 17

1.9.2 Aplicaciones existentes en Cuba que utilizan almacenes de datos 18

1.10 Herramientas 18

1.10.1 Herramientas de modelado 18

1.10.2 Sistema Gestor de Base de datos 19

1.10.3 Administrador de Base de datos 19

1.10.4 Servidor de Aplicaciones..... 20

1.10.5 Herramientas para el proceso de ETL..... 20

1.10.6 Herramientas para la Inteligencia de Negocios 21

1.11 Conclusiones parciales 22

CAPÍTULO 2: ANÁLISIS Y DISEÑO 23

2.1 Introducción 23

2.2 Análisis del mercado de datos 23

2.2.1 Descripción del negocio	23
2.2.2 Reglas del negocio.....	23
2.2.3 Requisitos de información	24
2.2.4 Requisitos funcionales	25
2.2.5 Requisitos no funcionales	26
2.2.6 Modelo de casos de uso del sistema.....	28
2.3 Arquitectura del mercado de datos.....	30
2.4 Diseño del mercado de datos.....	31
2.4.1 Diseño del subsistema de almacenamiento	31
2.4.2 Diseño del subsistema de integración.....	33
2.4.3 Diseño del subsistema de visualización	34
2.5 Seguridad del mercado de datos.....	36
2.6 Conclusiones parciales	36
CAPÍTULO 3: IMPLEMENTACIÓN DEL MERCADO DE DATOS	38
3.1 Implementación del subsistema de almacenamiento	38
3.2 Implementación del subsistema de integración	39
3.3 Implementación del subsistema de visualización	41
3.4 Conclusiones parciales	43
CAPÍTULO 4: PRUEBAS DEL MERCADO DE DATOS	45
4.1 Introducción	45
4.2 Pruebas de software	45
4.3 Herramientas para la aplicación de las pruebas.....	46
4.4 Resultado de las pruebas	50
4.5 Conclusiones parciales	51
CONCLUSIONES	53
RECOMENDACIONES	54
REFERENCIAS BIBLIOGRÁFICAS	55
BIBLIOGRAFÍA.....	58
ANEXOS.....	61
GLOSARIO	67

ÍNDICE DE FIGURAS

Figura 1. Componentes de un almacén de datos.....	8
Figura 2. Esquema estrella	12
Figura 3. Esquema copo de nieve.....	12

Figura 4. Esquema constelación de hechos.....	13
Figura 5. Diagrama de casos de uso del sistema.....	30
Figura 6. Modelo de datos	33
Figura 7. Procesos de integración.....	34
Figura 8. Arquitectura de información	35
Figura 9. Diseño del cubo reunión	35
Figura 10. Transformación para la carga del hecho reunión	40
Figura 11. Trabajo para la carga de las dimensiones.....	41
Figura 12. Trabajo para la carga de los hechos	41
Figura 13. Trabajo general.....	41
Figura 14. Mapa de navegación.....	42
Figura 15. Vista de análisis Cantidad de reuniones visitadas por carácter y estructura de la reunión ..	43
Figura 16. Modelo V.....	46

ÍNDICE DE TABLAS

Tabla 1. Convenciones de nombrado.....	27
Tabla 2. Actores del sistema.....	28
Tabla 3. Matriz bus	32
Tabla 4. Roles y permisos para acceder a la información	36
Tabla 5. Esquemas y tablas.....	38
Tabla 6. Caso de prueba Presentar información relacionada con las reuniones.	46
Tabla 7. Lista de chequeo aplicada al artefacto Registro del Sistema Fuente.....	47

INTRODUCCIÓN

Desde el comienzo de la humanidad el intercambio de información ha sido de gran importancia para la evolución de la sociedad. Su creciente desarrollo trae como resultado que la cantidad de datos a manejar aumente continuamente. Como consecuencia de este cúmulo de información, se hace necesario la creación de sistemas para el control y almacenamiento de la misma, capaces de brindar herramientas confiables y de calidad, que estén acorde a las exigencias demandadas por los usuarios, lo cual no sería posible sin el actual desarrollo de las tecnologías de la información.

Actualmente los procesos de dirección de la sociedad, en la mayoría de las actividades en que estos tengan lugar, requieren un manejo adecuado de grandes volúmenes de información y de la disponibilidad de una fuente de datos estadísticos, lo cual en el pasado resultaba engorroso, pero hoy con los avances alcanzados en los sistemas de información informáticos se obtienen en un menor tiempo y con una mayor exactitud en sus resultados.

Los sistemas de información se han convertido en un factor relevante para las organizaciones a nivel mundial, mejorando la forma en que éstas operan y contribuyendo a su proceso de toma de decisiones. Esto se debe a la posibilidad que brindan dichos sistemas de ejecutar una correcta gestión de la información.

Cuba, para no quedarse al margen en el empleo de los sistemas de información informáticos, ha puesto empeño en su uso, sobre todo en organizaciones que necesitan recopilar, describir y analizar un conjunto de datos estadísticos para mantener la integridad del país. En el desarrollo de estas aplicaciones informáticas para la gestión de la información, la Universidad de las Ciencias Informáticas (UCI) ha representado un papel fundamental, dentro de ella principalmente el Centro de Tecnologías de Gestión de Datos (DATEC), cuya especialidad es el desarrollo de soluciones, productos y servicios relacionados con la tecnología de gestión de datos.

En el centro DATEC se implementó un sistema de información, conocido como Sistema de Recogida de Información, que permite almacenar las actas que son generadas mensualmente por las reuniones de los comités de base de la Unión de Jóvenes Comunistas (UJC) dentro de la Facultad 6 de la UCI. Las actas de los comités de base se encontraban almacenadas en formato word, por lo que resultaba muy complejo para el comité primario la revisión de este cúmulo de datos. Actualmente dicho sistema

no permite obtener reportes estadísticos de los principales indicadores que necesitan conocer y manejar los directivos de esta organización en la facultad.

El análisis de la información referente a la asistencia a las reuniones, la cantidad de reuniones realizadas, las opiniones emitidas por los militantes y los acuerdos tomados en cada una de estas reuniones, resulta necesario para contribuir a la correcta toma de decisiones de los principales directivos de la UJC en la Facultad 6.

Por lo anteriormente expuesto se plantea como **problema de la investigación**: ¿Cómo contribuir a la toma de decisiones de los dirigentes de la Unión de Jóvenes Comunistas en la Facultad 6 de la Universidad de las Ciencias Informáticas?

Definiendo como **objeto de estudio**: los almacenes de datos, enmarcado en el **campo de acción** mercado de datos Unión de Jóvenes Comunistas para la Facultad 6 de la Universidad de las Ciencias Informáticas.

Para solucionar el problema planteado, se identifica como **objetivo general** de la investigación: Desarrollar el mercado de datos Unión de Jóvenes Comunistas para la Facultad 6 de la Universidad de las Ciencias Informáticas, que contribuya a la toma de decisiones.

En correspondencia con el mismo, se definen los siguientes **objetivos específicos**:

- Fundamentar la selección de las metodologías, herramientas y tecnologías a utilizar en el desarrollo de almacenes de datos.
- Realizar el análisis y diseño del mercado de datos Unión de Jóvenes Comunistas.
- Implementar el mercado de datos Unión de Jóvenes Comunistas.
- Realizar pruebas al mercado de datos Unión de Jóvenes Comunistas.

Para darle cumplimiento a los objetivos específicos planteados, se propone la realización de las siguientes **tareas de la investigación**:

- Caracterización de las metodologías, herramientas y tecnologías a utilizar en el desarrollo de almacenes de datos.
- Levantamiento de requisitos.
- Descripción de los casos de uso del mercado de datos.

- Definición de los hechos, las medidas y las dimensiones del mercado de datos.
- Diseño del modelo de datos.
- Definición de la arquitectura del mercado de datos.
- Diseño del subsistema de integración.
- Diseño del subsistema de visualización.
- Diseño de los casos de pruebas.
- Implementación del subsistema de integración.
- Implementación del subsistema de visualización.
- Aplicación de las listas de chequeo.
- Aplicación de los casos de pruebas.

Estructura de la Tesis

El presente trabajo está compuesto por cuatro capítulos estructurados de la siguiente manera:

Capítulo 1: Fundamentos teóricos de los almacenes de datos.

En este capítulo se abordan los principales conceptos relacionados con los almacenes de datos, sus principales características y las ventajas que proporciona su uso en el mundo empresarial. Se exponen además, otros aspectos, tales como: la metodología y herramientas a utilizar para el desarrollo de un almacén de datos.

Capítulo 2: Análisis y diseño del mercado de datos Unión de Jóvenes Comunistas para la Facultad 6 de la Universidad de las Ciencias Informáticas.

Este capítulo contiene la descripción de los pasos a seguir durante el análisis y el diseño de la solución. Se definen los requisitos que debe cumplir el sistema, así como el modelo dimensional propuesto para el desarrollo del mercado de datos a partir de los indicadores que se seleccionaron.

Capítulo 3: Implementación del mercado de datos Unión de Jóvenes Comunistas para la Facultad 6 de la Universidad de las Ciencias Informáticas.

En este capítulo se implementan todos los procesos de extracción, transformación y carga de los datos, donde se realizan los flujos de integración y los trabajos para enlazar todas las

transformaciones. Además, se llevan a cabo los procesos de Inteligencia de Negocio, donde se implementan el modelo de datos y las vistas de análisis.

Capítulo 4: Pruebas del mercado de datos Unión de Jóvenes Comunistas para la Facultad 6 de la Universidad de las Ciencias Informáticas.

En este capítulo se aplicarán los casos de prueba y las listas de chequeo para probar el mercado de datos. Estas pruebas son realizadas con el objetivo de garantizar que el sistema cumple con las necesidades del cliente y con la calidad requerida.

CAPÍTULO 1: FUNDAMENTOS TEÓRICOS DE LOS ALMACENES DE DATOS

1.1 Introducción

En este capítulo se reflejan los principales conceptos relacionados con los almacenes de datos, sus características, componentes, así como las ventajas y desventajas que trae consigo el uso de los mismos. Se abordan además, otros aspectos fundamentales como la metodología y las herramientas a utilizar en su implementación.

1.2 Almacén de datos

El término almacén de datos, en inglés “Data Warehouse” fue introducido por William H. Inmon¹, más conocido por Bill Inmon a principios de la década de los ´90, quien lo definió como: “...es un conjunto de datos orientados hacia una materia, integrados, no transitorios y que varían con el tiempo, los cuales apoyan el proceso de toma de decisiones de una administración...” [1].

Por su parte, Ralph Kimball² define un almacén de datos de una forma más sencilla y práctica, pero igual de importante: “...es una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis...” [2].

Otra definición en la cual coinciden varios autores expone: “...es una gran colección de datos que recoge información de múltiples sistemas fuentes u operacionales dispersos, y cuya actividad se centra en la Toma de Decisiones...” [3].

Estos conceptos a pesar de ser distintos coinciden en el señalamiento de que los almacenes de datos tienen como tarea fundamental: organizar la información recogida de diferentes fuentes para facilitar su análisis y apoyar el proceso de toma de decisiones.

¹William H. Inmon: licenciado en Ciencias Matemáticas y máster en Ciencias de la Computación, es considerado el padre del Data Warehouse y ha escrito una serie de libros junto a otros colaboradores, donde abordan ampliamente los tópicos y niveles de la arquitectura Data Warehouse.

²Ralph Kimball: autor ampliamente publicado sobre el tema de almacén de datos. Es conocido como el “Gurú del Data Warehousing”.

Después de un estudio detallado de la bibliografía existente sobre este tema, se puede concluir que un almacén de datos es una colección de datos, que sirve de apoyo para la toma de decisiones administrativas.

1.2.1 Características de un almacén de datos

A continuación se explican cuatro características fundamentales de los almacenes de datos que fueron mencionadas en la definición de Bill Inmon, las cuales han sido una gran ventaja para la toma de decisiones en aquellas empresas u organizaciones en que son utilizadas [4]:

Orientado a temas: la información recogida en el almacén se clasifica en dependencia a los aspectos que son de mayor importancia para la empresa, garantizando un acceso fácil y rápido a la misma.

Integrado: un almacén de datos está conformado por distintas fuentes heterogéneas, por lo que resulta muy común encontrar la misma información almacenada de diferentes formas. Por este motivo, los datos antes de ser cargados, deben pasar por un proceso de integración donde se corrigen, validan y estandarizan, para lograr llevarlos a un formato único que facilite su consulta en el almacén de datos.

Variable en el tiempo: los cambios producidos con el paso del tiempo en los datos almacenados también se van registrando en el almacén, lo que permite brindar una perspectiva histórica de la información, que abarca un espacio de tiempo de cinco a diez años.

No volátil: los datos recogidos en el almacén van a poder ser leídos por los usuarios, pero no podrán ser modificados. Solamente se podrán introducir nuevos datos en el almacén o consultar los que ya se encuentran almacenados.

1.2.2 Componentes de un almacén de datos

Los almacenes de datos están compuestos por un conjunto de elementos los cuales se explican a continuación [2]:

Sistemas fuentes operacionales: son los sistemas utilizados en las empresas para gestionar sus transacciones, información que es almacenada en diferentes formatos de acuerdo a las necesidades del negocio. Estos sistemas, conservan pocos datos históricos, pues generalmente realizan salvadas de

la información para trabajar con los datos generados en un corto período de tiempo y de esta forma hacer las recuperaciones más fácilmente.

Área de procesamiento (staging area): es un área de almacenamiento donde se realizan un conjunto de procesos comúnmente conocidos como extracción, transformación y carga (ETL), en los cuales se invierten la mayor cantidad de tiempo y esfuerzo durante la construcción de un almacén de datos. Primeramente, se realiza la extracción de los datos necesarios para el almacén de las diferentes fuentes, para luego pasar por un proceso de transformación donde se eliminan errores e inconsistencias que dificulten su posterior análisis. Finalmente, una vez que los datos están listos para ser almacenados, son cargados en el área de presentación del almacén de datos.

Área de presentación: en esta área los datos son almacenados, organizados y puestos a disposición de los usuarios para ser consultados, analizados o realizar reportes sobre ellos. En ella se almacena toda la información que puede ser de utilidad para el proceso de toma de decisiones en la empresa, diseñada mediante esquemas dimensionales. Generalmente, es referenciada como un conjunto de mercados de datos integrados, donde cada uno representa un proceso específico del negocio.

Herramientas de acceso a los datos: en este componente, se utiliza la palabra herramienta para referirse a la variedad de habilidades que pueden ser provistas a los usuarios del negocio, para soportar el proceso de toma de decisiones. Por definición, su actividad fundamental consiste en consultar la información que se encuentra en el área de presentación, lo que constituye el objetivo principal de los almacenes de datos.



Figura 1. Componentes de un almacén de datos

1.2.3 Ventajas y desventajas de un almacén de datos

A continuación se mencionan las ventajas y desventajas de los almacenes de datos [5]:

Ventajas:

- Integrar datos históricos sobre la actividad de la organización o negocio en un único repositorio.
- Analizar los datos del negocio desde la perspectiva de su evolución en el tiempo.
- Prever tendencias de evolución del negocio.
- Identificar nuevas oportunidades de negocio y tomar decisiones estratégicas.
- Reducir los costes materiales y humanos en la toma de decisiones.

Desventajas:

- Riesgo de fracaso en la construcción del sistema por cambios continuos en los requisitos de los usuarios.
- Problemas con la privacidad de los datos.

1.3 Mercado de datos

Un mercado de datos, en inglés “Data Marts” es un subconjunto de datos de un almacén relativos a los requisitos de un departamento o área de negocio concretos. Este subconjunto de datos puede funcionar de forma autónoma, o bien enlazado al almacén de datos [3].

Los mercados de datos son utilizados para facilitar la construcción y utilización del almacén debido a que éste puede alcanzar un gran volumen. A continuación se mencionan las características de los mercados de datos que los diferencian de los almacenes de datos [3]:

- Se centran en los requisitos de los usuarios asociados a un departamento o área de negocio concretos.
- Son más sencillos a la hora de utilizarlos y comprender sus datos, debido a que la cantidad de información que contienen es mucho menor que en los almacenes de datos.

1.3.1 Ventajas de los mercados de datos

Dentro de las ventajas de aplicar un mercado de datos a un negocio, se han seleccionado las siguientes [6]:

- Son simples de implementar.
- Conllevan poco tiempo de construcción y puesta en marcha.
- Permiten manejar información confidencial.
- Reflejan rápidamente sus beneficios y cualidades.
- Reducen la demanda del depósito de datos.

1.4 Etapas de desarrollo de un almacén de datos

1.4.1 Análisis y diseño

La etapa de análisis es la base para el desarrollo del mercado de datos, debido a que ésta es el punto de partida para los posteriores procesos de diseño e implementación.

Durante esta etapa se realiza un estudio detallado del negocio para conocer cuáles son las necesidades reales del cliente, se realiza de igual manera un levantamiento de los requisitos, lo cual será una guía para su desarrollo en la fase de implementación del almacén de datos.

Con la culminación de esta fase inicial quedan generados un conjunto de artefactos, entre los que se encuentran: el diagrama de diseño de la base de datos y los diagramas de casos de uso.

Finalizada la fase de análisis se da paso a la fase de diseño, donde es transformado el modelo lógico obtenido en la etapa anterior en modelo físico. Entre los artefactos generados al concluir esta etapa se encuentran: la matriz bus y el modelo de datos.

1.4.2 Extracción, transformación y carga

La etapa de extracción, transformación y carga ETL (por sus siglas en inglés) cobra gran importancia dentro de la realización del almacén de datos, en la misma queda definida la forma en que se almacenarán los datos en dicho almacén. Antes de realizar el proceso de ETL es necesario llevar a cabo el perfilado de los datos, que permite obtener un conocimiento de la calidad que presenta la

información con la que se va a trabajar. Una vez concluido el mismo, se ejecutan las operaciones del proceso de ETL, las cuales son detalladas a continuación.

1. Extracción

La operación de extracción consiste en extraer los datos de los sistemas de origen. Cada uno de estos sistemas puede utilizar distintos formatos, por lo que se hace necesario convertir los datos en un formato común, preparado para iniciar la operación de transformación.

Una parte fundamental dentro de la operación de extracción es analizar los datos extraídos, de lo cual resulta un chequeo que verifica que los datos cumplan con la estructura esperada y en caso de no ser así son rechazados.

2. Transformación

La operación de transformación consiste en la aplicación de un conjunto de reglas o funciones a los datos extraídos. Estas transformaciones se realizan debido a las necesidades técnicas del almacén. Entre las actividades que se llevan a cabo durante esta operación se encuentran: la limpieza, la normalización y la selección de los datos.

3. Carga

La operación de carga consiste en añadir la información al almacén de datos, una vez que la misma ha sido extraída de las fuentes.

Si no se realiza un correcto proceso de extracción transformación y carga (ETL) se pudieran obtener datos incorrectos, lo que afectaría el proceso de toma de decisiones, es por eso que este proceso constituye aproximadamente un 70% del trabajo de la construcción de un AD³ [7].

1.4.3 Inteligencia de Negocio

La Inteligencia de Negocio BI (por sus siglas en inglés) es un conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento, mediante el análisis de datos existentes en una organización o empresa [8].

³ Siglas usadas para hacer referencia a un almacén de datos.

Este conjunto de estrategias y herramientas tienen en común las siguientes características [8]:

- **Accesibilidad a la información:** los datos son la fuente principal de este concepto. Lo primero que deben garantizar este tipo de herramientas y técnicas será el acceso de los usuarios a los datos, con independencia de la procedencia de estos.
- **Apoyo en la toma de decisiones:** se busca ir más allá en la presentación de la información, de manera que los usuarios tengan acceso a herramientas de análisis que les permitan seleccionar y manipular sólo aquellos datos que les interesen.
- **Orientación al usuario final:** se busca independencia entre los conocimientos técnicos de los usuarios y su capacidad para utilizar estas herramientas.

1.5 Modelo de datos

Un modelo de datos es un lenguaje utilizado para describir una base de datos y refleja cómo quedarán relacionados los datos dentro de la misma y las condiciones que éstos deben cumplir para mostrar lo que se desea obtener.

1.5.1 Modelo Entidad-Relación

Un diagrama o modelo entidad-relación es un lenguaje para el modelado de datos de un sistema de información. Estos modelos expresan entidades más relevantes para el sistema, sus inter-relaciones y propiedades. Trabajan dividiendo los datos en muchas entidades discretas donde cada una se convierte en una tabla física en la base de datos operacional [9]. Los modelos entidad-relación no son recomendables para el diseño de los almacenes de datos, ya que no garantizan una recuperación de toda la información.

1.5.2 Modelo Dimensional

Para el diseño de un almacén de datos es utilizado el modelo dimensional. En este modelo la información es organizada de forma tal que se garantice la velocidad y eficiencia en la organización de la misma. Una de sus características principales es que no necesita una predefinición de los reportes, debido a que se diseñan de forma tal que cubra el universo de variantes que los usuarios necesiten al consultar la información almacenada [9].

1.6 Tipos de modelado de un almacén de datos

Para el modelado de un almacén de datos son utilizados tres tipos de esquemas los cuales son explicados a continuación [10]:

Esquema estrella: el esquema en estrella consta de una tabla de hechos central y de varias tablas de dimensiones relacionadas a ésta, a través de sus respectivas claves.

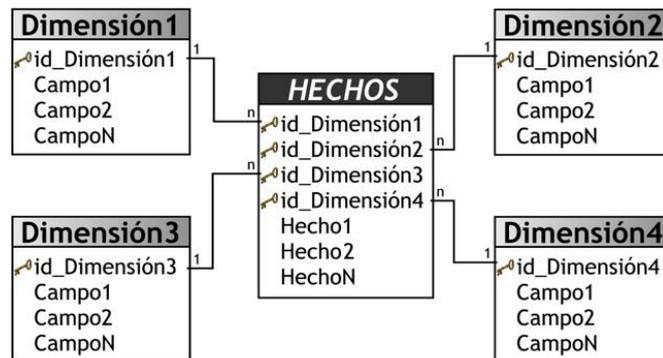


Figura 2. Esquema estrella

Esquema copo de nieve: este esquema representa una extensión del modelo en estrella, cuando las tablas de dimensiones se organizan en jerarquías de dimensiones.

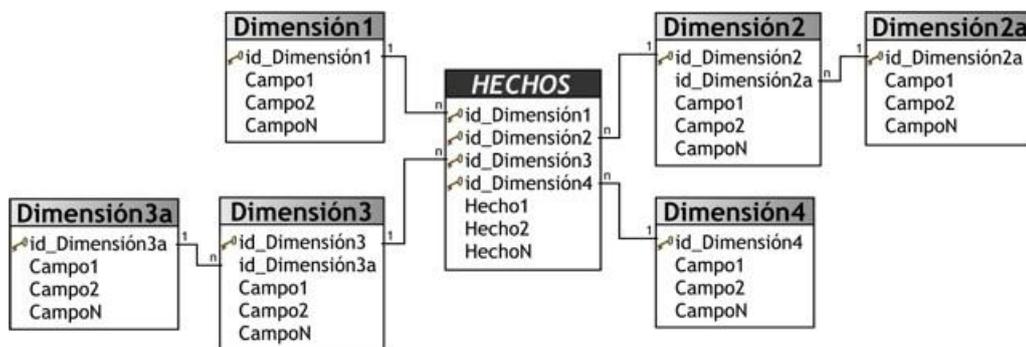


Figura 3. Esquema copo de nieve

Esquema constelación de hechos: los esquemas en estrella y copo de nieve pueden generalizarse con la inclusión de distintas tablas de hechos que comparten todas o algunas de las dimensiones. Dentro de las ventajas de utilizar este tipo de esquema se pueden mencionar las siguientes: permite tener más de una tabla de hechos, por lo cual se podrán analizar más aspectos claves del negocio con

un mínimo esfuerzo adicional de diseño; además contribuye a la reutilización de las tablas de dimensiones, porque una misma tabla de dimensión puede utilizarse para varias tablas de hechos.

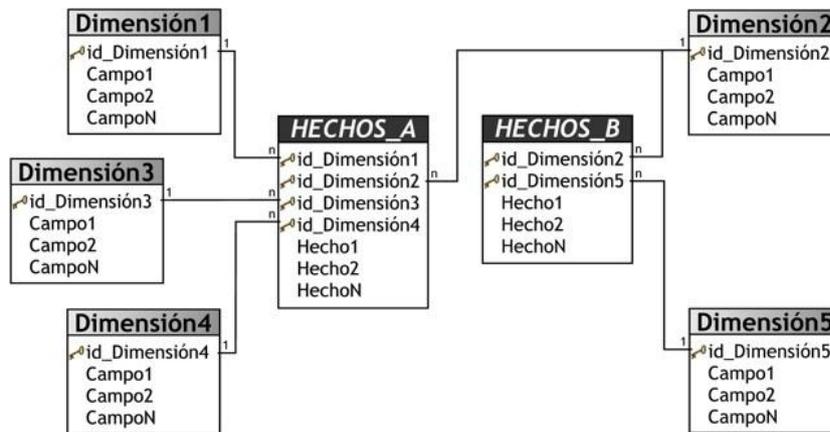


Figura 4. Esquema constelación de hechos

Para el desarrollo del modelo de datos del mercado de datos Unión de Jóvenes Comunistas se decide utilizar el esquema constelación de hechos, ya que el mismo contará con más de una tabla de hechos que comparten las mismas dimensiones. Otra de las cuestiones a tomar en cuenta para la utilización del mismo son los beneficios que trae consigo, los cuales fueron mencionados anteriormente.

1.6.1 Componentes de los esquemas de modelado

Los esquemas de modelado cuentan con varios componentes, como son: las tablas de hechos y las tablas de dimensiones, la granularidad y la jerarquía. Los mismos son explicados a continuación:

Tablas

Hechos: los hechos son operaciones que se realizan en el negocio en un tiempo determinado. Son las variables sobre las que se van a realizar operaciones de agregación que conduzcan a conclusiones sobre la evolución del área o departamento que se estudie. Son objetos de análisis para la toma de decisiones [11]. La tabla de hechos es la tabla primaria en el modelo dimensional.

Dimensión: la dimensión es una característica de un hecho que permite su análisis posterior, o sea, el eje de análisis por el que se desea categorizar la información [11]. La tabla de dimensión contiene la descripción textual del negocio y son las compañeras integrales de las tablas de hechos.

Granularidad

Los datos contenidos en las tablas de hechos y dimensiones están conformados por niveles de información. La granularidad es el nivel más bajo de información que será almacenado en la tabla de hechos. El primer paso al diseñar una tabla de hechos es determinar la granularidad [12].

Jerarquía

Es el nombre que toma la organización dentro de los campos de una dimensión. La jerarquía es tenida en cuenta a la hora de añadir información.

1.7 Modos de almacenamiento de datos

El Procesamiento Analítico en Línea OLAP (por sus siglas en inglés) es la técnica que permite ver y manipular los datos por dimensiones, proveyendo a los gerentes y analistas fácil acceso a la información, con el fin de soportar el proceso de toma de decisiones. En esta técnica de análisis, en lugar de ejecutar múltiples consultas, los datos son estructurados para permitir un acceso rápido y fácil a las respuestas de las preguntas que son típicamente formuladas. De esta manera, OLAP, brinda flexibilidad en la visualización de la información [13].

Las herramientas OLAP son utilizadas para la construcción de sistemas de apoyo a la toma de decisiones, permitiendo realizar sofisticados análisis multidimensionales en tiempo real, lo que posibilita entrelazar información de diferentes fuentes de datos. Las mismas están dirigidas a los usuarios finales, por lo que requieren de una interfaz simple y de una buena integración de los sistemas que la componen.

Existen tres modelos OLAP, los cuales son: ROLAP, MOLAP y HOLAP. El proceso de análisis se realiza de igual forma en los tres modelos, lo que varía en uno y otro caso es la metodología de almacenamiento.

ROLAP (Procesamiento Analítico Relacional en Línea): en este modelo la información es almacenada en filas y columnas de forma relacional. El mismo presenta los datos a los usuarios en forma de dimensiones de negocio. Para lograr esto se utilizan las etiquetas de los metadatos, lo cual es usado también, para ocultar las estructuras de almacenamiento. Estos metadatos son almacenados en las tablas relacionales, dentro de un esquema nombrado metadatos. Entre las ventajas que

presenta este modelo destacan: brinda una gran seguridad e integridad a la información que se encuentra almacenada en las bases de datos y permite que la misma pueda ser compartida con aplicaciones SQL.

MOLAP (Procesamiento Analítico Multidimensional en Línea): en este modelo los datos son almacenados dimensionalmente. Las estructuras de datos están fijas para que la lógica al procesar la información pueda estar basada en métodos bien definidos, para establecer las coordenadas del almacenamiento de los datos. La arquitectura MOLAP utiliza bases de datos multidimensionales para proporcionar el análisis.

HOLAP (Procesamiento Analítico Híbrido en Línea): este modelo, como su nombre lo indica, es un híbrido entre los métodos ROLAP y MOLAP. Permite almacenar una parte de los datos como en un sistema MOLAP y el resto como en uno ROLAP.

De estos modelos mencionados se ha decidido utilizar ROLAP para el desarrollo del mercado de datos por las ventajas que trae consigo su uso, las cuales fueron mencionadas anteriormente.

1.8 Metodologías para el desarrollo de un almacén de datos

Las metodologías de desarrollo definen un conjunto de pasos y procesos a seguir, que permiten estructurar, controlar y planificar el proceso de desarrollo de software, con el objetivo de entregar al cliente un producto con calidad y en el tiempo debido.

El diseño de un almacén de datos es un proceso que ha ido madurando desde sus inicios en la década del 90. Este desarrollo ha traído consigo disímiles enfoques por parte de varias personalidades influyentes en el tema, de los cuales destacan: la Metodología Kimball⁴ y la Metodología Inmon⁵. Ambas metodologías han servido de guía sobre este tema a la comunidad mundial. La principal diferencia entre ambos estilos metodológicos radica en la forma de enfrentar el problema.

Metodología Kimball

⁴ Metodología Kimball: es nombrada así en honor a su creador Ralph Kimball.

⁵ Metodología Inmon: nombrada de esta forma en honor a su creador William H. Inmon.

Esta metodología defiende una arquitectura ascendente (bottom-up). La idea es construir mercados de datos independientes para evaluar las ventajas del nuevo sistema a medida que avanzamos [14]. Es decir, esta metodología está basada en crear por cada departamento un conjunto de mercados de datos independientes, orientados a los temas que estén relacionados con él.

Otra de las características que aporta gran valor a esta metodología, es la propuesta de Kimball de dividir la Inteligencia de Negocio entre el hecho y las dimensiones, lo cual conduce a una solución completa en una menor cantidad de tiempo.

Una de las principales ventajas de este enfoque es que posee una amplia documentación, donde se pueden encontrar respuestas a disímiles interrogantes.

Metodología Inmon

Esta metodología defiende una arquitectura descendente (top-down), la cual propone construir los mercados de datos cuando haya sido terminado el almacén de datos completo de la organización. Esta solución, según Inmon, considerará mejor todos los datos corporativos.

Este enfoque plantea además, la creación de un repositorio de datos corporativo como fuente de información consolidada, persistente, histórica y de calidad. Como el almacén de datos se construye descendentemente los mercados de datos se nutren del almacén corporativo, convirtiéndose en un complejo empresarial de bases de datos relacionales [15].

1.8.1 Metodología seleccionada para el desarrollo del mercado de datos

Para el desarrollo del mercado de datos Unión de Jóvenes Comunistas se decide utilizar la Propuesta de Metodología para el desarrollo de almacenes de datos en DATEC, la cual tiene como base el enfoque planteado por Kimball, pero se adapta a las características de trabajo en la UCI y en el centro DATEC. A continuación se exponen los elementos que influyeron en la decisión de tomar como base el enfoque planteado por Kimball para la confección de esta metodología:

- Facilita el proceso de toma de decisiones, debido a que identifica las tablas de hechos y dimensiones, lo cual agiliza considerablemente el proceso de desarrollo.

- Propone ir construyendo los mercados de datos departamentales y después el almacén de datos, lo que trae como ventaja que la puesta en marcha de los mercados se realice en un corto período de tiempo y luego se valore si se construye o no el almacén.
- Cuenta con una abundante documentación, lo que facilita en gran medida el trabajo de los desarrolladores, que pueden dar respuestas a sus interrogantes consultando la misma.

Como complemento de esta metodología y por las particularidades de trabajo en la UCI, se tomó lo planteado por Leopoldo Zenaido Zepeda en su tesis de doctorado, que propone incluir los casos de uso para guiar el proceso de desarrollo y de esta forma estar más alineados a las tendencias y normas de la universidad.

En el ciclo de vida de esta metodología tienen lugar los siguientes flujos de trabajo:

- Estudio preliminar o planeación.
- Levantamiento de requisitos.
- Arquitectura.
- Diseño e implantación.
- Prueba.
- Despliegue.
- Soporte y mantenimiento.
- Gestión y administración del proyecto.

1.9 Aplicaciones existentes

En la actualidad se evidencia una gran competencia entre las empresas a nivel mundial, razón que ha obligado a los principales directivos a encontrar soluciones en vías de mejorar su proceso de gestión empresarial. Como vía de solución han decidido incorporar a sus organizaciones tecnologías que permitan realizar estudios de mercado y tendencias, así como facilitar al análisis de la información generada en cada una de ellas, con el objetivo de apoyar el proceso de toma de decisiones.

1.9.1 Aplicaciones existentes en el mundo que utilizan almacenes de datos

La utilización de los almacenes de datos ha tenido un gran auge en el mundo empresarial como solución a los problemas de su gestión. Muchas empresas han decidido utilizar esta tecnología de

almacenamiento con el fin de analizar su información con mayor rapidez. Después de un estudio bibliográfico del tema se pueden mencionar algunas empresas: el Corte Inglés, los Supermercados Casino en Francia, Canadian Tyre en Canadá, Procter & Gamble en México. También grandes trasnacionales como Coca Cola, Walt Disney, Nike, Maybelline, Adidas, 3M, y Bosh Siemens se han incorporado a la utilización de los almacenes de datos para la realización de estudios de mercado.

1.9.2 Aplicaciones existentes en Cuba que utilizan almacenes de datos

En nuestro país existen varias empresas que han logrado aplicar esta tecnología a sus negocios. Luego de un estudio referente al tema se puede mencionar: la Corporación CIMEX, la cual se dedica fundamentalmente a la exportación e importación de mercancías. La función principal del mercado de datos de esta corporación es la gestión de inventario, permitiendo así una mejor gestión de compra-venta. Otras empresas cubanas que utilizan esta tecnología de almacenamiento son Cubacel, Unión Cupet y Copextel.

1.10 Herramientas

En la actualidad son utilizadas diversas herramientas con el propósito de mejorar los procesos de diseño, construcción, implementación y mantenimiento de los almacenes de datos.

1.10.1 Herramientas de modelado

Las herramientas de modelado son utilizadas por los diseñadores y analistas durante el proceso de desarrollo del software. Las mismas son de gran ayuda, ya que proveen métodos, técnicas y utilidades que facilitan y agilizan el trabajo.

Visual Paradigm for UML v8.0: esta herramienta soporta el ciclo de vida completo de desarrollo del software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue. Es muy fácil de usar, ya que posee una interfaz gráfica amigable y posibilita diseñar todos los tipos de diagramas de clases necesarios, generar código desde cualquiera de estos diagramas y generar documentación. Además, es fácil de instalar y actualizar y sus diferentes ediciones son compatibles entre ellas [16].

Esta es la herramienta escogida para el modelado durante el ciclo de vida de desarrollo del mercado de datos por todas las características mencionadas anteriormente.

1.10.2 Sistema Gestor de Base de datos

Un Sistema de Gestión de Base de datos es una aplicación informática que permite a los usuarios definir, crear, mantener y consultar una base de datos, así como proporcionar acceso controlado a la misma [17].

PostgreSQL v9.1: es un sistema de gestión de bases de datos objeto-relacional, distribuido bajo licencia BSD y con su código fuente disponible libremente. PostgreSQL utiliza un modelo cliente/servidor y usa multiprocesos en vez de multihilos para garantizar la estabilidad del sistema. Un fallo en uno de los procesos no afectará el resto y el sistema continuará funcionando [18].

Ventajas de PostgreSQL

- Soporta distintos tipos de datos: además del soporte para los tipos base, también soporta datos de tipo fecha, monetarios, elementos gráficos, datos sobre redes y cadenas de bits. También permite la creación de tipos propios.
- Incorpora una estructura de datos “array”.
- Incorpora funciones de diversa índole: manejo de fechas, geométricas, orientadas a operaciones con redes, etc.
- Permite la declaración de funciones propias, así como la definición de disparadores.
- Soporta el uso de índices, reglas y vistas.
- Incluye herencia entre tablas, por lo que a este gestor de bases de datos se le incluye entre los gestores objeto-relacionales.
- Permite la gestión de diferentes usuarios, como también los permisos asignados a cada uno de ellos [19].

Debido a las ventajas antes mencionadas, a su propuesta de código abierto y a su adaptabilidad al problema en cuestión se ha decidido utilizar esta herramienta para el desarrollo del mercado de datos.

1.10.3 Administrador de Base de datos

PgAdmin v1.14: es una herramienta de código abierto para la administración de bases de datos PostgreSQL.

Características:

- Interfaz administrativa gráfica.
- Herramienta de consulta SQL (con un explain gráfico).
- Editor de código procedural.

PgAdmin se diseña para responder a las necesidades de la mayoría de los usuarios, desde escribir simples consultas SQL hasta desarrollar bases de datos complejas. La interfaz gráfica soporta todas las características de PostgreSQL y hace simple la administración. Está disponible en más de una docena de lenguajes y para varios sistemas operativos, incluyendo Microsoft Windows, Linux, FreeBSD, Mac OSX y Solaris [20].

1.10.4 Servidor de Aplicaciones

Apache Tomcat v6: es un servidor web desarrollado en un entorno abierto y participativo y publicado bajo la licencia Apache versión 2. Apache Tomcat es una implementación de software de código abierto de la especificación Java Servlet y tecnologías Java Server Pages [21].

1.10.5 Herramientas para el proceso de ETL

DataCleaner v1.5.3: es una aplicación de código abierto utilizada para el análisis, perfilado, y limpieza de datos. Estas actividades ayudan a administrar y controlar la calidad de los datos.

DataCleaner es la alternativa gratuita al software de gestión de datos maestros MDM (por sus siglas en inglés), metodologías, proyectos de almacenamiento de datos, la investigación estadística, la preparación para la extracción, transformación y carga (ETL) y más actividades [22].

Pentaho Data Integration v4.2.1: esta aplicación también conocida como Kettle, es una de las soluciones más extendidas y mejor valoradas por los desarrolladores de almacenes de datos. Permite realizar transformaciones y trabajos de una forma muy sencilla e intuitiva, por lo que es considerada una herramienta muy poderosa para realizar el proceso ETL. El uso de Kettle permite evitar grandes cargas de trabajo manual, frecuentemente difíciles de mantener y de desplegar [23].

Características

A parte de ser una herramienta de código abierto y sin costes de licencia, las características básicas de esta herramienta son:

- Posee un entorno gráfico de desarrollo.
- Uso de tecnologías estándar: Java, XML, Java Script.
- Fácil de instalar y configurar.
- Multiplataforma: Windows, Macintosh, Linux.
- Basado en dos tipos de objetos: transformaciones (colección de pasos en un proceso ETL) y trabajos (colección de transformaciones) [23].

1.10.6 Herramientas para la Inteligencia de Negocios

Mondrian v3.0.4: es un servidor OLAP de código abierto utilizado para gestionar la comunicación entre una aplicación OLAP y la base de datos con los datos fuente. Entre sus características se encuentra la facilidad que brinda para el análisis de grandes volúmenes de información. Mondrian soporta el lenguaje Microsoft's Multidimensional Expressions (DMX). También soporta los APIs: Java OLAP (JOLAP) y XML para el análisis de aplicaciones programadas. Es una de las aplicaciones más importantes de la plataforma Pentaho BI.

Schema Workbench v3.2.1: es una herramienta gráfica que permite desarrollar y probar cubos OLAP, además permite publicarlos al servidor BI para que puedan ser utilizados en los análisis por los usuarios de la plataforma. Es una aplicación multiplataforma desarrollada en Java que posee dos áreas: la zona en la que se muestra la estructura jerárquica del esquema OLAP y la zona de edición de las propiedades de cada elemento.

Pentaho BI Server v3.8: es una aplicación desarrollada en Java que permite la gestión de los recursos BI. Posee una interfaz donde se muestran los informes y las vistas OLAP. Cuenta además, con una consola que permite administrar la aplicación y conocer el rendimiento de la misma.

Ventajas:

- Se integra con la mayoría de los entornos y se puede comunicar con otras aplicaciones vía webservices.
- Integra todos los recursos informacionales en una única plataforma de explotación.
- Proporciona mucha libertad al usuario y a los desarrolladores para crear contenidos nuevos [24].

1.11 Conclusiones parciales

Luego de realizar un estudio bibliográfico concerniente al tema objeto de estudio se arriba a las siguientes conclusiones:

- Fueron reflejados los principales conceptos relacionados con los almacenes de datos, permitiendo un correcto desarrollo de la investigación.
- Se seleccionaron las herramientas y la metodología a utilizar, logrando planificar adecuadamente el proceso de desarrollo del mercado de datos.

CAPÍTULO 2: ANÁLISIS Y DISEÑO

2.1 Introducción

En este capítulo se realiza un estudio preliminar del negocio, donde son identificadas las necesidades de información del cliente. Quedarán expuestos los requisitos funcionales y no funcionales, las reglas del negocio, así como los casos de uso del sistema y la interacción con los actores. Otros elementos que serán reflejados en este capítulo son los hechos, las dimensiones y las medidas del mercado de datos, el modelo de datos dimensional y la matriz bus.

2.2 Análisis del mercado de datos

2.2.1 Descripción del negocio

La UJC en la Facultad 6 agrupa a sus militantes por comités de base en cada uno de los grupos, departamentos y áreas de servicios que conforman la facultad. La autoridad superior al comité de base es el comité primario, que está integrado por los dirigentes de la organización en la facultad, que a su vez se subordinan al comité UJC-UCI, conformado por los cuadros a nivel de universidad. Para recoger la información generada en cada una de las reuniones llevadas a cabo por estas estructuras (comité de base, comité primario y comité UJC-UCI) se implementó el Sistema de Recogida de Información. El mismo almacena los datos referentes a las actas de las reuniones, como son: la asistencia, el tipo de estructura en que se desarrolla la reunión, las opiniones emitidas y los acuerdos tomados. Sin embargo, no permite obtener reportes estadísticos sobre estos datos, ni analizar la información desde la perspectiva del tiempo, condiciones que resultan importantes para apoyar la toma de decisiones de los principales directivos de la organización en la facultad.

2.2.2 Reglas del negocio

Las reglas del negocio (RN) describen las políticas, normas, operaciones, definiciones y restricciones presentes en una organización y son de vital importancia para que la misma cumpla sus objetivos. Deben ser explícitas, escritas y expresadas en términos sencillos [25]. A continuación se mencionan las reglas del negocio identificadas durante la fase de análisis del negocio.

RN1-Los identificadores de las dimensiones no pueden ser nulos.

RN2-En la dimensión comité de base en el campo nombre del comité de base de aparecer alguna irregularidad se cambiará por el nombre del comité de base correcto.

- De aparecer *Dpto o Departamento* seguido del nombre de éste, se tomará solamente el nombre del mismo.
- Al presentarse siglas, tales como: *PP, CB, CB2, H, SD* se sustituirá por *Práctica Profesional, Ciencias Básicas, Ciencias Básicas 2, Humanidades y Sistemas Digitales* respectivamente.
- De comenzar con las palabras *Acta del o Análisis* seguido del nombre del comité de base, se dejará solamente el nombre.
- De iniciar con las siglas *C/B:* seguido del nombre se dejará solamente el nombre.

RN3-En la dimensión clasificación en el campo clasificación de aparecer el valor *cconstructiva*, se sustituirá por la palabra *constructiva*.

RN4-En la dimensión área_ujc en el campo área de obtenerse el valor *OrdenDía* se modificará por el valor *Orden del día*.

RN5-En la dimensión carácter en el campo carácter en caso de encontrarse el valor *extraorinario* se sustituirá por la palabra *extraordinario*.

2.2.3 Requisitos de información

Los requisitos de información (RI) describen qué información debe almacenar el sistema para satisfacer las necesidades de clientes y usuarios [26]. Luego de realizado un estudio del negocio se identificaron los siguientes requisitos de información:

RI1-Obtener la cantidad de reuniones por estructura, carácter y fecha de la reunión.

RI2-Obtener la asistencia a las reuniones por estructura y carácter de la reunión.

RI3-Obtener la cantidad de reuniones visitadas por estructura y carácter de la reunión.

RI4-Obtener la cantidad de opiniones por área y clasificación.

RI5-Obtener la cantidad de acuerdos por fecha de cumplimiento.

2.2.4 Requisitos funcionales

Los requisitos funcionales (RF) son capacidades o condiciones que debe cumplir el sistema, son de gran importancia para satisfacer las necesidades del cliente. A continuación se exponen los RF definidos:

RF1- Autenticar usuario.

RF2- Gestionar rol.

RF2.1- Adicionar rol.

RF2.2- Eliminar rol.

RF2.3- Modificar rol.

RF2.4- Visualizar rol.

RF3- Gestionar usuario.

RF3.1- Adicionar usuario.

RF3.2- Eliminar usuario.

RF3.3- Modificar usuario.

RF3.4- Visualizar usuario.

RF4- Gestionar reporte.

RF4.1- Adicionar reporte.

RF4.2- Modificar reporte.

RF4.3- Eliminar reporte.

RF4.4- Visualizar reporte.

RF5- Extraer información.

RF6- Realizar transformación y carga.

RF7- Abrir navegador OLAP.

RF8- Mostrar editor MDX.

RF9- Mostrar Padres.

RF10- Ocultar repeticiones.

RF11- Intercambiar ejes.

RF12- Mostrar gráfico.

RF13- Configurar gráfico.

RF14- Configurar impresión.

RF15- Exportar a PDF.

RF16-Exportar a Excel.

RF17-Mostrar propiedades.

RF18-Suprimir filas.

RF19-Detallar miembro.

RF20-Abrir detalle.

RF21-Mostrar datos de origen.

2.2.5 Requisitos no funcionales

Los requisitos no funcionales (RNF) son propiedades o cualidades que el producto debe tener. Se identificaron diez RNF dentro del mercado de datos Unión de Jóvenes Comunistas, los cuales se encuentran explicados a continuación:

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.

Disponibilidad

RNF3-Garantizar la persistencia de la información: se debe realizar un respaldo total de los datos del mercado de datos con una frecuencia mensual y anual. La información estará almacenada en la oficina del Secretario General de la UJC en la Facultad 6. Éste será el responsable de la información.

Soporte

RNF4-Lograr la homogeneidad de la estructura de los elementos definidos en el almacén: las estructuras del almacén de datos deben tener un nombre estándar, teniendo en cuenta el tipo de

estructura que sea. En la siguiente tabla se definen convenciones de nombrado con el objetivo de manejar un vocabulario común en todo el almacén de datos, permitiendo un entendimiento claro y conciso por parte de los desarrolladores.

Tabla 1. Convenciones de nombrado

Estructura	Descripción	Ejemplo
Tablas de hechos	Todas las tablas de hechos tendrán una cadena que demuestra que son hechos y el concepto que describen.	hech_<concepto>
Tablas de dimensiones	Todas las tablas de dimensiones tendrán una cadena que demuestra que son dimensiones y el concepto que describen.	dim_<concepto>

Interfaz

RNF5-Acceder al sistema: el usuario deberá acceder a la aplicación mediante el protocolo HTTP, utilizando un navegador web.

Restricciones de diseño

RNF6-Utilizar el Sistema Gestor de Base de datos definido durante la investigación: el gestor de base de datos que se utilizará es PostgreSQL v9.1 y como interfaz de administración de dicho gestor PgAdmin v1.14.

RNF7-Utilizar las herramientas para la implementación de la capa de Inteligencia de Negocio definidas durante la investigación: de la suite Pentaho, se utilizarán los siguientes componentes: Schema Workbench v3.2.1 y Pentaho BI Server v3.8. Para la utilización de los mismos se requiere la instalación de la máquina virtual de java: Java Virtual Machine v6 o superior.

Requisitos de Software

RNF8-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, además el plugin Adobe Flash Player. El servidor de aplicaciones, por su parte, debe contar con las siguientes herramientas: Pentaho BI Server v3.8, Java Virtual Machine v6 o superior y Apache Tomcat v6.0, además debe poseer un navegador web y el plugin Adobe Flash Player. Por último, el servidor de base de datos debe tener instalado: el PostgreSQL v9.1, como Sistema Gestor de Base de datos y el PgAdmin v1.14, como Administrador de Base de datos.

Requisitos de Hardware

RNF9-Proporcionar características mínimas de hardware a las estaciones de trabajo clientes: las estaciones de trabajo clientes deben contar como mínimo con una memoria RAM de 512 MB.

RNF10-Proporcionar características mínimas de hardware a los servidores: Para lograr una explotación aceptable del sistema los servidores deben contar con los siguientes requerimientos mínimos de hardware: 1 GB de memoria RAM y 80 GB de capacidad de almacenamiento de disco duro.

2.2.6 Modelo de casos de uso del sistema

Actores del sistema

A continuación se relacionan los actores del sistema con la descripción de cada uno de ellos.

Tabla 2. Actores del sistema

Actor	Descripción
Administrador	Responsable de administrar los usuarios del sistema, asignarles sus respectivos roles, a la vez que debe administrar dichos roles y los reportes.
Administrador ETL	Responsable de llevar a cabo los procesos de extracción, transformación y carga de los datos.
Analista	Responsable de inicializar los casos de uso relacionados con la visualización de los reportes del mercado de datos.

Diagrama de casos de uso del sistema

Para la confección del diagrama de casos de uso del sistema fueron identificados casos de uso de información y funcionales, los cuales son descritos en los Anexos 1 y 2 de este documento.

Los casos de uso informativos se agrupan por el tipo de información que es manejada en la UJC en la Facultad 6 y en dependencia de las necesidades de información de los usuarios. Los casos de uso de información que se definieron son los siguientes:

- Presentar información relacionada con las reuniones.
- Presentar información relacionada con las opiniones.
- Presentar información relacionada con los acuerdos.

Los casos de uso funcionales agrupan requerimientos funcionales que describen de manera detallada el comportamiento del sistema con los distintos actores que interactúan con él [27]. Los casos de uso funcionales identificados son los siguientes:

- Realizar la extracción de los datos.
- Realizar la transformación y carga de los datos.
- Autenticar usuario.
- Gestionar usuario.
- Gestionar rol.

- Gestionar reporte.
- Realizar operaciones sobre reportes.

A continuación se muestra el diagrama de casos de uso del sistema en el cual quedan reflejadas las interacciones de los actores con los casos de uso.

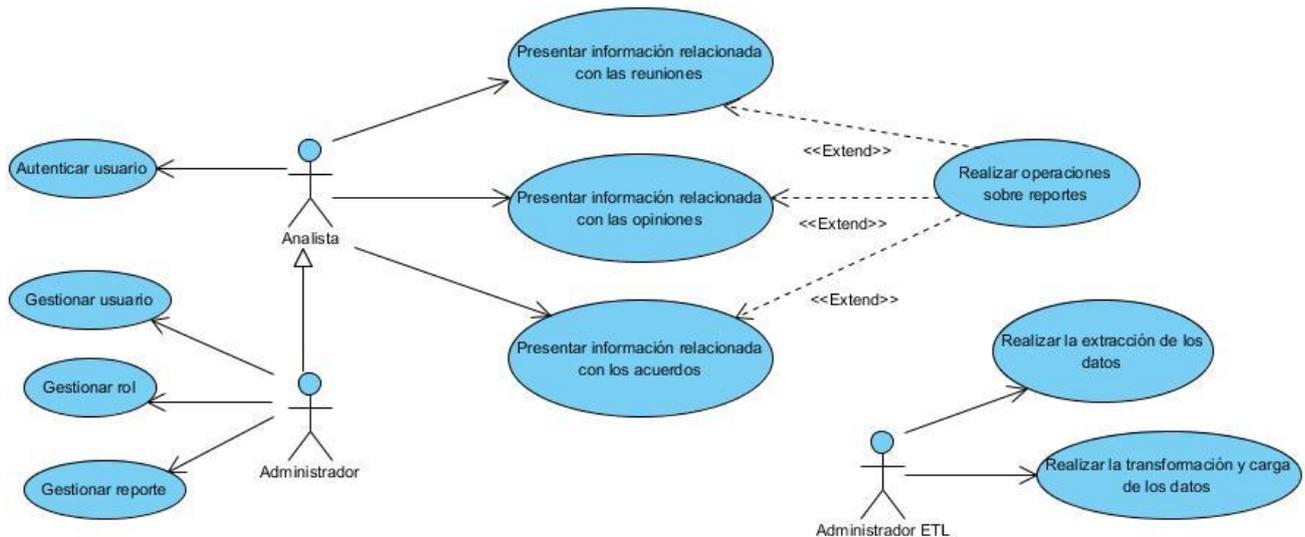


Figura 5. Diagrama de casos de uso del sistema

2.3 Arquitectura del mercado de datos

Para dar inicio a la etapa de diseño del mercado de datos es necesario definir la arquitectura que se va a utilizar. El mismo estará compuesto por la fuente de datos y tres subsistemas, los cuales se explican seguidamente:

Fuente de datos: es la fuente de información con la que se va a trabajar, en este caso es la base de datos del Sistema de Recogida de Información.

Subsistema de integración: en este subsistema serán extraídos y transformados los datos provenientes de la base de datos fuente.

Subsistema de almacenamiento: en este subsistema serán cargados los datos luego de ser transformados por el subsistema de integración.

Subsistema de visualización: en este subsistema se consulta la información que está almacenada en el subsistema de almacenamiento utilizando la herramienta de acceso a datos (Pentaho BI Server).

2.4 Diseño del mercado de datos

En la etapa de diseño de un mercado de datos quedan identificados los aspectos lógicos con que éste debe contar, como son: el diseño de los subsistemas de almacenamiento, integración y visualización. Igualmente en esta etapa se deben definir las políticas de seguridad con las que debe cumplir el sistema.

2.4.1 Diseño del subsistema de almacenamiento

Para diseñar el subsistema de almacenamiento fueron identificados los hechos y las dimensiones con las que debe contar el mercado de datos, igualmente fue realizada la matriz bus y quedó confeccionado el modelo de datos dimensional.

Hechos

- hech_reunion: en este hecho se recoge la información relacionada con las reuniones.
- hech_opinion: en este hecho se recoge la información relacionada con las opiniones.
- hech_acuerdos: en este hecho se recoge la información relacionada con los acuerdos.

Dimensiones

- dim_estructura: esta dimensión recoge la estructura que puede tener una reunión.
- dim_tiempo_dia: esta dimensión recoge la información relacionada con la fecha
- dim_caracter: esta dimensión recoge el carácter que puede tener una reunión.
- dim_clasificacion: esta dimensión recoge la clasificación que puede tener una opinión.
- dim_area: esta dimensión recoge el área a la que pertenece una opinión.
- dim_comite_base: esta dimensión recoge la información relacionada con el comité de base.

Matriz bus

La matriz bus representa las relaciones que existen entre los hechos y las dimensiones, su objetivo es evitar el solapamiento entre los hechos. A continuación se muestra la matriz bus, donde son relacionados los hechos y las dimensiones identificados para la solución.

Tabla 3. Matriz bus

Dimensiones	Hechos		
	hech_reunion	hech_opinion	hech_acuerdos
dim_estructura	X	X	X
dim_tiempo_dia	X	X	X
dim_caracter	X	X	X
dim_clasificacion		X	
dim_area_ujc		X	
dim_comite_base	X	X	X
dim_tiempo_dia_cumplimiento			X

Modelo de datos dimensional

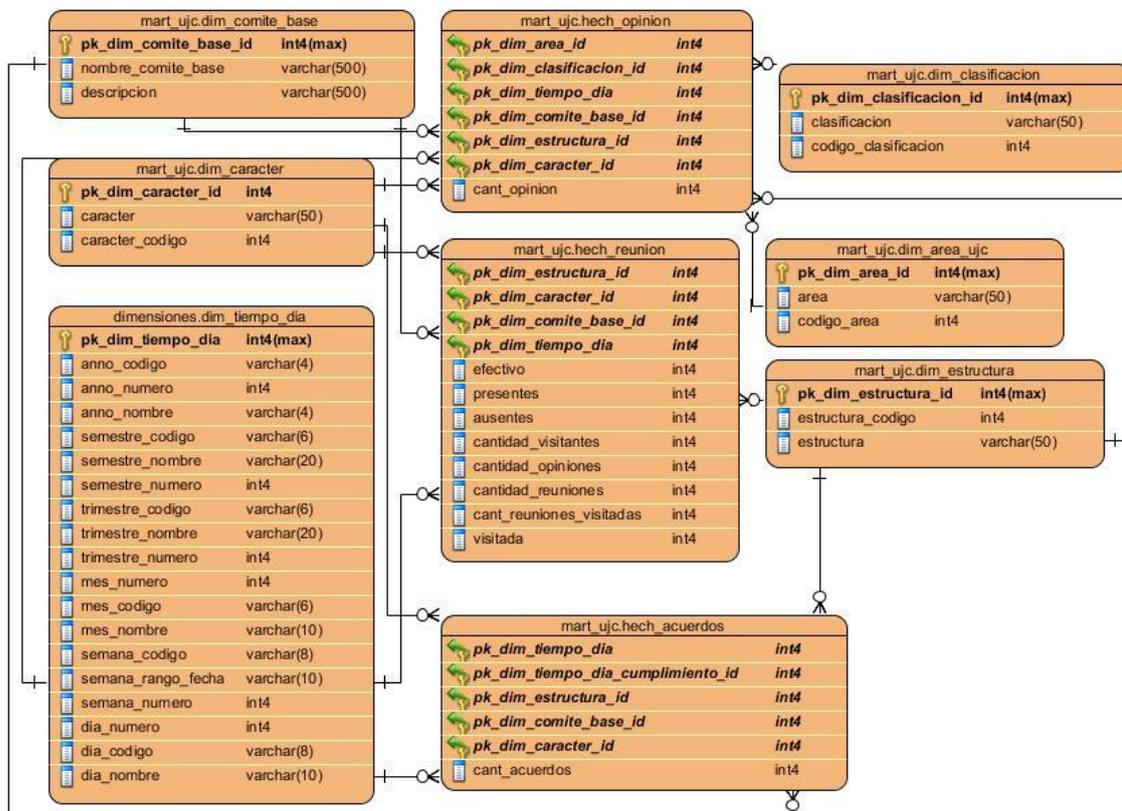


Figura 6. Modelo de datos

2.4.2 Diseño del subsistema de integración

El diseño del subsistema de integración del mercado de datos comprende dos procesos fundamentales, tales como: el perfilado de los datos, que permite conocer el estado en el que éstos se encuentran y el diseño de las transformaciones, que es realizado una vez que los datos son extraídos desde los sistemas fuentes y de esta forma la información queda integrada y estandarizada.

Perfilado de los datos

El perfilado de los datos consiste en analizar la información que se encuentra almacenada en los sistemas fuentes. Este proceso permite conocer cuál es la estructura, el formato y la calidad que tienen los datos con los que se va a trabajar. Es válido aclarar que luego de cargados los datos en el mercado de datos se debe realizar nuevamente el perfilado para valorar la calidad de la información final.

Diseño de las transformaciones

El diseño de las transformaciones abarca de manera general como se llevará a cabo el proceso de ETL. En primer lugar, se extraen los datos de la base de datos fuente. Luego de extraídos los datos son realizadas las validaciones definidas por el analista, en caso de encontrarse valores incorrectos estos datos serán almacenados en un fichero excel para darle tratamiento. Si todos los datos cumplen con la calidad requerida se le realizan un conjunto de transformaciones y son cargados en la base de datos destino.

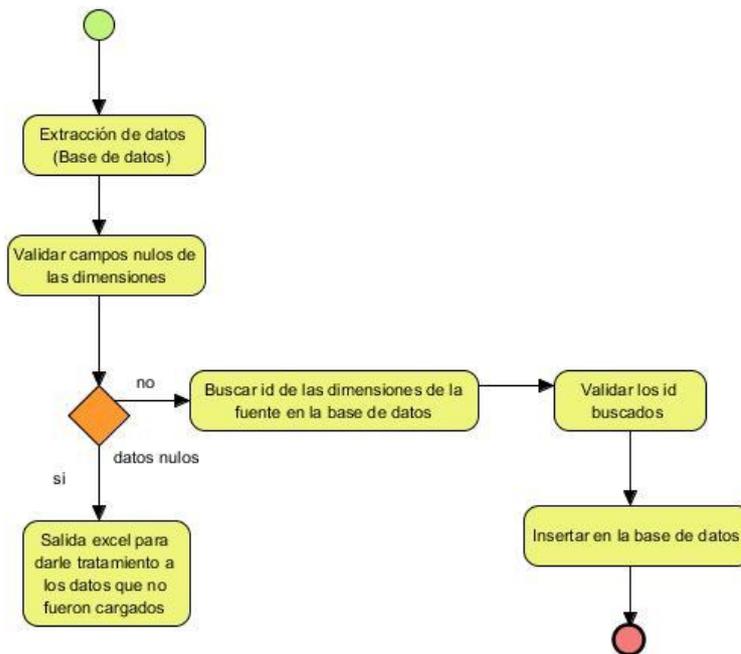


Figura 7. Procesos de integración

2.4.3 Diseño del subsistema de visualización

Dentro de las tareas que son realizadas durante el diseño del subsistema de visualización se encuentran: el diseño de la arquitectura de información y de los cubos OLAP.

Diseño de la arquitectura de información

La capa de visualización del mercado de datos está formada por el área de análisis (A.A) "UJCF6", la cual está compuesta por tres libros de trabajo (L.T). A continuación se explica detalladamente cómo queda conformada la arquitectura de información del mercado de datos Unión de Jóvenes Comunistas.

A.A UJCF6: agrupa toda la información referente a las reuniones, los acuerdos tomados y las opiniones emitidas por los militantes de la UJC de la Facultad 6. Dicha información se muestra mediante vistas de análisis, que permiten apoyar el proceso de toma de decisiones en esta organización.

L.T Acuerdos: libro de trabajo contenido dentro del A.A UJCF6, conformado por una vista de análisis que permite realizar un análisis de datos correspondiente a los acuerdos tomados en las reuniones de la UJC en la Facultad 6.

L.T Opinión: libro de trabajo contenido dentro del A.A UJCF6, conformado por tres vistas de análisis que permiten realizar un análisis de datos correspondiente a las opiniones emitidas en las reuniones de los comités de base por los militantes de la UJC en la Facultad 6.

L.T Reunión: libro de trabajo contenido dentro del A.A UJCF6, conformado por seis vistas de análisis que permiten realizar un análisis de datos correspondiente a las reuniones de la UJC en la Facultad 6.

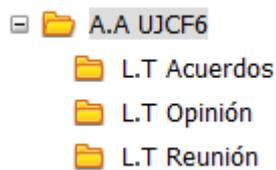


Figura 8. Arquitectura de información

Diseño de los cubos OLAP

El diseño de los cubos OLAP es fundamental para organizar la información de forma multidimensional. Para la solución del mercado de datos fueron definidos tres cubos multidimensionales, seis dimensiones y nueve medidas. A continuación se muestra cómo queda diseñado el cubo reunión.



Figura 9. Diseño del cubo reunión

2.5 Seguridad del mercado de datos

La seguridad es un aspecto fundamental con la que deben contar todos los sistemas de información, ya que la misma provee mecanismos de precaución que evitan la pérdida de información o el acceso a la misma por personas no autorizadas. La seguridad en el mercado de datos estará garantizada a través de la autenticación por parte de los usuarios, a los cuales se les asigna un rol y sus respectivos permisos para interactuar con la aplicación.

Tabla 4. Roles y permisos para acceder a la información

Roles	Permisos
Administrador	Posee acceso total para gestionar los permisos, roles y usuarios. También posee acceso total al A.A UJCF6 del mercado de datos.
Analista	Posee acceso de sólo lectura al A.A UJCF6, pero puede visualizar todas las vistas de análisis de esta área.

Otra de las cuestiones a tener en cuenta para la seguridad del mercado de datos es garantizar la persistencia de la información. Para lograr este objetivo se realizará un respaldo total de la información del mercado de datos mensualmente y anualmente.

2.6 Conclusiones parciales

Luego de finalizar la etapa de análisis y diseño del mercado de datos Unión de Jóvenes Comunistas se arriba a las siguientes conclusiones:

- Fueron identificadas las reglas del negocio y los requisitos necesarios para dar cumplimiento a las necesidades del cliente.
- Se confeccionó el diagrama de casos de uso del sistema, logrando perfeccionar la estructura del proceso de negocio.
- Se diseñó el modelo de datos y la matriz bus, reflejando las relaciones entre los hechos y las dimensiones y permitiendo que no existiera solapamiento entre los hechos.

- Fueron diseñados los subsistemas (almacenamiento, integración y visualización), permitiendo una arquitectura adecuada para la posterior implementación del sistema.
- Se elaboró el esquema de seguridad para lograr un producto seguro.

CAPÍTULO 3: IMPLEMENTACIÓN DEL MERCADO DE DATOS

Este capítulo está dirigido a desarrollar la implementación de los subsistemas de almacenamiento, integración y visualización del mercado de datos Unión de Jóvenes Comunistas.

3.1 Implementación del subsistema de almacenamiento

Para el desarrollo del mercado de datos se definieron tres esquemas: dimensiones, mart_ujc y metadatos, los cuales se explican a continuación.

- Esquema dimensiones: en el mismo se encuentra la dimensión compartida dim_tiempo_dia.
- Esquema mart_ujc: en el mismo se encuentran los hechos y las dimensiones restantes con las que cuenta el mercado de datos
- Esquema META-DATOS: en el mismo se encuentran tres tablas de metadatos⁶, cuyo objetivo consiste en registrar información acerca de las ejecuciones de los trabajos, las transformaciones y las dimensiones lentamente cambiantes⁷.

La solución cuenta con 12 tablas: tres tablas de hechos, seis tablas de dimensiones y tres tablas de metadatos. Seguidamente se muestran cada uno de los esquemas con las tablas que los componen.

Tabla 5. Esquemas y tablas

⁶ Metadatos: se refieren a la información generada cuando se ejecuta un proceso. Permite que los administradores gestionen su sistema y aseguren que las cosas funcionen sin problemas. Si hay un problema con algún proceso, los metadatos ayudan a los administradores a identificar y localizar los problemas.

⁷ Dimensiones lentamente cambiantes: son aquellas dimensiones en las que puede variar la información en cualquier instante de tiempo.

Esquemas	Tablas
dimensiones	dim_tiempo_dia
mart_ujc	dim_estructura dim_caracter dim_clasificación dim_area_ujc dim_comite_base hech_reunion hech_opinion hech_acuerdos
META-DATOS	META-G_CAMBIOS META-TRABAJOS META-TRANSFORMACION

3.2 Implementación del subsistema de integración

Para comenzar con la implementación del subsistema de integración se extraen los datos de los sistemas fuentes, posteriormente se da inicio a la etapa de transformaciones, a las cuales le fueron aplicadas las reglas del negocio que arrojaron los resultados del perfilado de datos.

Después de aplicarse un análisis a los datos de la fuente, se pudo detectar que el campo que más dificultades presentaba era comité de base, pues el usuario al entrar la información correspondiente al mismo cometió múltiples errores, por ejemplo: el uso indebido de espacios innecesarios o caracteres extraños, la utilización de abreviaturas no autorizadas o la combinación de ambos. El perfilado de datos arrojó además, que no existían campos nulos en la fuente, sin embargo se observaban varios espacios en blanco y errores ortográficos.

Implementación de las transformaciones

Durante la realización del proceso de ETL del mercado de datos se realizaron tres transformaciones para la carga de los hechos y seis transformaciones para la carga de las dimensiones. El

procedimiento general para la carga del hecho reunión se describe de la siguiente manera: se obtuvieron los identificadores necesarios para la carga del hecho, detectada la existencia de irregularidades en el flujo de datos se ejecutaron las transformaciones necesarias para garantizar la integridad de los mismos. Los componentes utilizados para la limpieza de la información fueron: replace in string, seleccionar/renombrar, valores de javascript y mapeo de valores. A continuación se muestra una imagen con los pasos utilizados para la carga del hecho antes descrito.

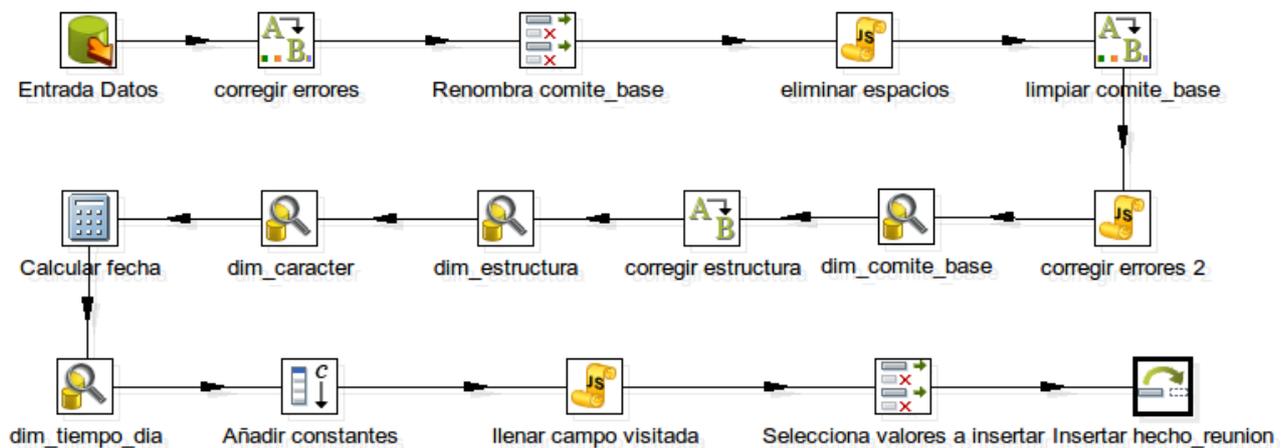


Figura 10. Transformación para la carga del hecho reunión

Implementación del trabajo general

Luego de realizadas todas las transformaciones necesarias para cargar los hechos y las dimensiones se realizó la implementación del trabajo general, cuyo objetivo es que las transformaciones se realicen en un orden lógico. Primeramente se realiza el trabajo job_transf para la carga de las dimensiones (ver figura 11) y posteriormente se realiza el trabajo job_hechos para la carga de los hechos (ver figura 12). Una vez concluidos se implementa el trabajo general (ver figura 13).

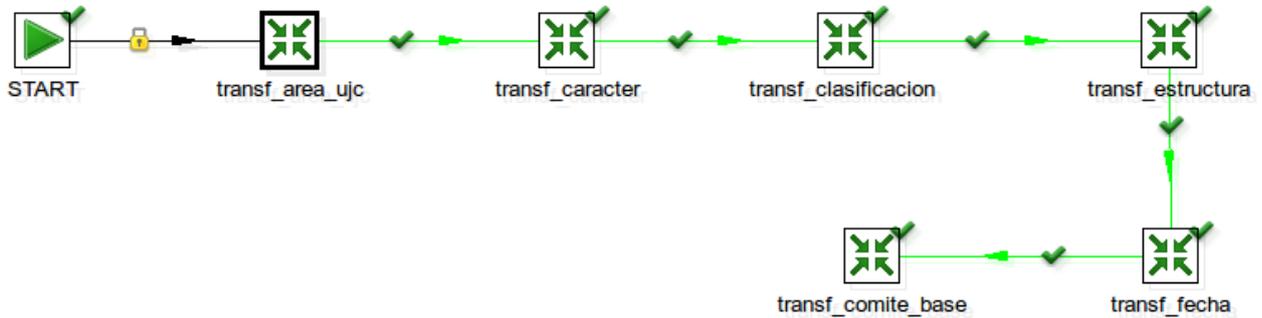


Figura 11. Trabajo para la carga de las dimensiones



Figura 12. Trabajo para la carga de los hechos

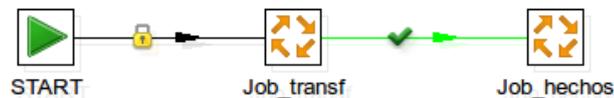


Figura 13. Trabajo general

3.3 Implementación del subsistema de visualización

Mapa de navegación

Antes de comenzar con la implementación del subsistema de visualización se muestra cómo queda conformado el mapa de navegación del mercado de datos Unión de Jóvenes Comunistas. El mismo manifiesta de forma general la manera en que los usuarios podrán acceder a la información.



Figura 14. Mapa de navegación

Implementación de las vistas de análisis

Luego de realizado el diseño del subsistema de visualización se deben implementar las vistas de análisis para completar la capa de visualización. El mercado de datos contará con diez vistas de análisis, las cuales están contenidas en sus tres L.T. Seguidamente se explica cómo quedan conformados los mismos y la información que brinda cada una de sus vistas de análisis.

Libro de trabajo Acuerdos

- Cantidad de acuerdos por fecha de cumplimiento del acuerdo: esta vista de análisis muestra la cantidad de acuerdos que fueron tomados en las reuniones de la organización, según la fecha de cumplimiento propuesta para este acuerdo.

Libro de trabajo Opinión

- Cantidad de opiniones por clasificación: esta vista de análisis muestra la cantidad de opiniones que fueron emitidas en las reuniones según la clasificación de la opinión (constructiva, no constructiva, debate, inquietud u otras).
- Cantidad de opiniones por clasificación de la opinión: esta vista de análisis muestra un gráfico de barras que representa la misma información que se muestra en la vista anterior, pero de una forma más dinámica para la vista del usuario.
- Cantidad de opiniones por área de opinión: esta vista de análisis muestra la cantidad de opiniones que fueron emitidas en las reuniones según el área de opinión (gastronomía, comedores, docencia, residencia, servicios, orden del día u otras).

Libro de trabajo Reunión

- Asistencia a las reuniones por carácter de la reunión: esta vista de análisis muestra la asistencia a las reuniones según el carácter de la reunión (ordinario y extraordinario).
- Asistencia a las reuniones por estructura de la reunión: esta vista de análisis muestra la asistencia a las reuniones según la estructura de la reunión (comité de base, comité primario y comité UJC-UCI).

- Cantidad de reuniones por carácter de la reunión: esta vista de análisis muestra la cantidad de reuniones que fueron realizadas según el carácter de la reunión.
- Cantidad de reuniones por estructura de la reunión: esta vista de análisis muestra la cantidad de reuniones que fueron realizadas según la estructura de la reunión.
- Cantidad de reuniones por fecha de la reunión: esta vista de análisis muestra la cantidad de reuniones según la fecha en la que fue realizada la reunión.
- Cantidad de reuniones visitadas por carácter y estructura de la reunión: esta vista de análisis muestra la cantidad de reuniones que fueron visitadas según el carácter y la estructura de la reunión.

A continuación se muestra la vista de análisis Cantidad de reuniones visitadas por carácter y estructura de la reunión perteneciente al L.T Reunión.

		Medidas
Carácter	Estructura	● Cantidad de reuniones visitadas
☐ Todos	☐ Todas	275
	Comité de base	270
	Comité UJC	1
	Comité primario	4
ordinario	☐ Todas	251
	Comité de base	248
	Comité primario	3
extraordinario	☐ Todas	24
	Comité de base	22
	Comité UJC	1
	Comité primario	1

Figura 15. Vista de análisis Cantidad de reuniones visitadas por carácter y estructura de la reunión

3.4 Conclusiones parciales

Una vez terminada la fase de implementación del mercado de datos se arriba a las siguientes conclusiones:

- Se llevó a cabo la implementación del subsistema de almacenamiento, lográndose de esta forma una mejor estructuración de la información.
- Fue implementado el subsistema de integración, quedando conformadas las transformaciones para la carga de los hechos y las dimensiones, así como el trabajo general.
- Se realizó la implementación del subsistema de visualización, permitiendo mostrar la información final contenida en el mercado de datos y de esta forma dar cumplimiento a las necesidades del cliente.

CAPÍTULO 4: PRUEBAS DEL MERCADO DE DATOS

4.1 Introducción

En este capítulo se exponen las pruebas realizadas al mercado de datos, así como las listas de chequeo y sus resultados. Estas pruebas son realizadas con el objetivo de garantizar que el sistema cumpla con las necesidades del cliente y con la calidad requerida.

4.2 Pruebas de software

Los productos de software se han convertido en objetivos estratégicos de las organizaciones, los procesos más importantes que se desarrollan en ellas dependen en gran medida del buen funcionamiento de estos productos. Por este motivo se hace necesario que los productos de software cuenten con una alta calidad y para ello deben aplicárseles pruebas antes de su salida al mercado.

Para obtener un producto con calidad, que cumpliera con los requerimientos establecidos, se realizaron las pruebas de software de acuerdo al Modelo V utilizado en el centro DATEC para evaluar y determinar la calidad del producto. Este modelo está determinado por la secuencia de un conjunto de actividades, la cuales se muestran en la figura 16.

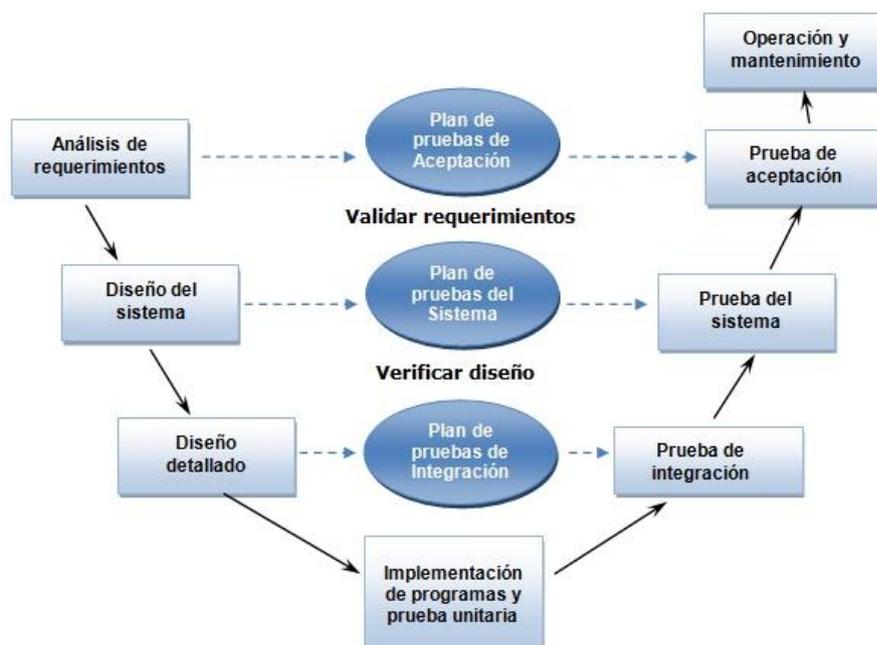


Figura 16. Modelo V

Las pruebas aplicadas al mercado de datos según el modelo empleado por el centro son explicadas a continuación:

- **Prueba unitaria:** la prueba unitaria o de unidad centra el proceso de verificación en la menor unidad del diseño del software: el módulo [28].
- **Prueba de integración:** la prueba de integración es una técnica sistemática para construir la estructura del programa, mientras que al mismo tiempo, se llevan a cabo pruebas para detectar errores asociados con la integración [28].
- **Prueba de sistema:** son usualmente conducidas para asegurar que todos los módulos trabajan como sistema, sin error. Es similar a la prueba de integración pero con un alcance mucho más amplio. Las pruebas del sistema examinan qué tan bien el sistema cumple con los requerimientos de la organización y su utilidad, seguridad y desempeño. También prueba documentación del sistema [29].
- **Prueba de aceptación:** son realizadas principalmente por los usuarios con el apoyo del equipo del proyecto. El propósito es confirmar que el sistema está terminado, que desarrolla puntualmente las necesidades de la organización y que es aceptado por los usuarios finales [29].

4.3 Herramientas para la aplicación de las pruebas

Para probar el mercado de datos fueron diseñados tres casos de prueba, con el objetivo de verificar los tres casos de uso de información identificados para la solución. A continuación se muestra un fragmento del caso de prueba correspondiente al caso de uso de información Presentar información relacionada con las reuniones, donde se aprecia la primera vista de análisis correspondiente al L.T Reunión con una breve descripción, además de las variables de entrada y salida, la respuesta que debe dar el sistema y el flujo central de ejecución del caso de prueba.

Tabla 6. Caso de prueba Presentar información relacionada con las reuniones.

Escenario	Descripción	Variables de entrada	Variables de salida	Respuesta del sistema	Flujo central
EC 1.1 Asistencia a las reuniones por carácter de la reunión	Permite visualizar el reporte con las variables presentes en el mismo.	Carácter	Efectivos Presentes Ausentes	El sistema muestra todas las variables disponibles para el reporte Asistencia a las reuniones por carácter de la reunión.	Se abre la aplicación. Se autentica. Se entra al sistema. Se selecciona en el área superior izquierda el área de análisis A.A UJCF6. Se selecciona el libro de trabajo L.T Reunión. Se selecciona en el área inferior izquierda el reporte: "Asistencia a las reuniones por carácter de la reunión"

Otra de las herramientas utilizadas para la aplicación de las pruebas fue la lista de chequeo de ETL, cuyo objetivo es validar los artefactos desarrollados (Diccionario de datos, Mapa lógico de datos, Perfil de los datos y Registro del Sistema fuente). La misma se divide en 3 secciones:

- **Estructura del documento:** se comprueban todos los aspectos definidos por el expediente de proyecto.
- **Elementos definidos por el modelo de desarrollo:** abarca todos los indicadores a evaluar durante la etapa de desarrollo del mercado según el modelo de desarrollo.
- **Semántica del documento:** se verifica la ortografía, redacción y demás indicadores en el documento.

Seguidamente se mencionan los elementos que forman parte de la lista de chequeo de ETL y se muestra como quedó conformada la lista de chequeo aplicada al artefacto Registro del Sistema Fuente (ver tabla 7).

- **Peso:** define si el indicador a evaluar es crítico o no.
- **Indicadores a evaluar:** son los indicadores a evaluar en las secciones Estructura del documento, Semántica del documento y Elementos definidos por el modelo de desarrollo.
- **Evaluación (Eval):** es la forma de evaluar el indicador en cuestión. El mismo se evalúa de 1 en caso de que exista alguna dificultad sobre el indicador y 0 en caso de que el indicador revisado no presente problemas.
- **N.P. (No Procede):** se usa para especificar que el indicador no es necesario evaluarlo en ese caso.
- **Cantidad de elementos afectados:** especifica la cantidad de errores encontrados sobre el mismo indicador.
- **Comentario:** especifica los señalamientos o sugerencias que quiera incluir la persona que aplica la lista de chequeo. Pueden o no existir señalamientos o sugerencias.

Tabla 7. Lista de chequeo aplicada al artefacto Registro del Sistema Fuente.

Estructura del documento					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de	Comentarios

				elementos afectados	
crítico	¿El entregable contiene las secciones obligatorias de la plantilla estándar definida para el expediente de proyecto?	0		0	
crítico	¿El alcance del proyecto describe correctamente los datos de las dimensiones y hechos del mercado de datos?	0		0	
crítico	¿El objetivo expresa correctamente el propósito del documento?	0		0	
	¿Se hace un uso adecuado del control del documento?	0		0	
	¿En la sección de acrónimos se definen todos los acrónimos utilizados en el documento?	1		1	No se definieron en la sección Definiciones, Acrónimos y Abreviaturas las siglas de CP: comité primario.
	¿Queda establecido en el entregable el sistema fuente de los datos?	0		0	

	¿Queda definido en el entregable los datos de contacto dentro de la empresa con la que se trabaja?	0		0	
	¿Se especifica el sistema gestor en el que se encuentra disponible la fuente de datos?	0		0	
	¿Se especifica en el entregable las propiedades de la fuente de datos (tamaño, cantidad de usuarios, cantidad de transacciones diarias, accesibilidad y disponibilidad)?	0		0	
Indicadores definidos en el desarrollo					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
	¿Se utilizó un lenguaje cuyas sentencias son expresables mediante una sintaxis bien definida?	0		0	
Semántica del documento					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
crítico	¿Se han identificado errores ortográficos en los entregables?	0		0	

crítico	¿Se entiende claramente lo que se ha especificado en el documento?	0		0	
	¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?	1		Todos	El número de páginas de los elementos que aparecen en el índice no coincide con el contenido de la misma.

4.4 Resultado de las pruebas

A continuación se detallan los resultados de cada una de las pruebas aplicadas de acuerdo a la propuesta del modelo V:

Pruebas unitarias

Una vez culminada la implementación se realizaron pruebas unitarias a los flujos de integración de datos y a los diferentes componentes relacionados con la capa de visualización, a través de casos de pruebas que le fueron aplicados a los casos de uso de información y casos de prueba de ETL. Con las mismas fueron detectadas dos no conformidades relacionadas con la estructura de los casos de prueba y las vistas de análisis, que fueron resueltas satisfactoriamente.

Pruebas de integración

Las pruebas de integración fueron llevadas a cabo a medida que se fue implementando cada módulo del mercado de datos. La primera prueba de integración se realizó a los procesos de ETL, donde fueron encontradas tres no conformidades las cuales se mencionan a continuación:

- No se había validado que los campos nulos fueran enviados a un excel de error para darle posterior tratamiento.

- Las variables de entrada y salida eran locales, lo que no permitía que las transformaciones se llevaran a cabo desde cualquier máquina.
- No existía la estructura adecuada de las carpetas que contenían las transformaciones.

La segunda prueba de integración se llevó a cabo una vez culminada la implementación de la capa de visualización, en la cual se detectaron dos no conformidades que son mencionadas seguidamente:

- No se había personalizado el sistema para el cliente.
- Las vistas de análisis no se habían guardado correctamente, lo que impedía que la información fuera mostrada completamente.

Pruebas de sistema

Luego de concluido el sistema fueron llevadas a cabo varias pruebas, donde fue encontrada una no conformidad relacionada con una de las vistas de análisis, la cual no brindaba la información que era necesaria para el cliente, debido a un error en el diseño de los cubos OLAP.

Pruebas ejecutadas aplicando listas de chequeo

Una vez aplicada la lista de chequeo a los artefactos de ETL se detectaron 3 no conformidades, las mismas son mencionadas a continuación:

- El número de página que aparecía en el índice de uno de los artefactos no coincidía con el contenido que se reflejaba en dicha página.
- En dos de los artefactos no se habían definido todos los acrónimos y abreviaturas que se utilizaban.

Todas las no conformidades detectadas en las pruebas antes mencionadas fueron corregidas en el tiempo establecido para ello.

4.5 Conclusiones parciales

Luego de finalizadas las pruebas al mercado de datos, se puede concluir:

- Fueron realizadas pruebas unitarias, de integración y de sistema, las cuales permitieron detectar ocho no conformidades y darle solución a las mismas rápidamente.
- Se aplicaron casos de prueba a cada uno de los casos de uso de información y casos de uso de ETL y listas de chequeo para los artefactos de ETL, logrando identificar cinco no conformidades.
- Se realizó el tratamiento y seguimiento de todas las no conformidades encontradas para prevenir posibles errores.

CONCLUSIONES

Al concluir la solución, se puede plantear que fueron cumplidos los objetivos trazados y las tareas de la investigación propuestas:

- Se seleccionaron las herramientas necesarias para la implementación del mercado de datos Unión de Jóvenes Comunistas; fueron aplicadas las técnicas OLAP, permitiendo el análisis multidimensional de los datos y fue escogida la metodología que guió todo el proceso para el desarrollo del mismo.
- Se realizó el análisis y diseño del mercado de datos, lográndose identificar los requisitos con los que debía contar el sistema. Fueron diseñados los subsistemas de almacenamiento, integración y visualización, sirviendo éste de base para la posterior implementación del mercado de datos.
- Se implementaron los subsistemas del mercado de datos, permitiendo: organizar los datos a través de la utilización de una estructura física, compuesta por esquemas y tablas; cargar los datos en dicha estructura y por último visualizar la información requerida, para apoyar la toma de decisiones.
- Se probó el mercado de datos, logrando evaluar los procesos de integración de datos e Inteligencia de Negocio y dar solución satisfactoria a las no conformidades encontradas.

RECOMENDACIONES

Se recomienda:

- Que se ponga en explotación el mercado de datos en la organización UJC para lograr que sean probadas las funcionalidades del mismo y detectar nuevas funcionalidades.
- Que se haga extensible esta solución para el resto de las facultades de la UCI.

REFERENCIAS BIBLIOGRÁFICAS

1. Berríos, Ericka Graciela Sevilla. *GUÍA METODOLÓGICA PARA LA DEFINICIÓN Y DESARROLLO DE UN DATA WAREHOUSE*. Managua, Nicaragua : s.n., 2003.
2. Kimball, Ralph y Ross, Margy. *The Data Warehouse Toolkit Second Edition The Complete Guide to Dimensional Modeling*. Estados Unidos de América : Wiley Computer Publishing, 2002. ISBN 0-471-20024-7.
3. Cabrera, Maria Evelia Casales. Facultad de Ciencias. *Facultad de Ciencias*. [En línea] [Citado el: 2 de noviembre de 2011.] <http://hp.fciencias.unam.mx/~alg/bd/dwh.pdf>.
4. Alarcos. *Alarcos*. [En línea] [Citado el: 11 de Noviembre de 2011.] <http://alarcos.inf-cr.uclm.es/doc/bbddavanzadas/08-09/FUNCIONALIDAD%204.pdf>.
5. García, Gerardo Clemente. *Un Sistema para el Mantenimiento de Almacenes de Datos*. Valencia, España : s.n., 2008.
6. Dataprix. *Dataprix*. [En línea] [Citado el: 3 de noviembre de 2011.] <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/i-data-warehousing-investigacion-y-sistematizacion-concepto-13>.
7. Sinnexus. *Sinnexus*. [En línea] [Citado el: 29 de octubre de 2011.] www.sinnexus.com.
8. DeciSop Todo sobre toma de decisiones. *DeciSop Todo sobre toma de decisiones*. [En línea] [Citado el: 3 de Noviembre de 2011.] <http://decisop.wordpress.com/category/toma-de-decisiones/>.
9. Sierra, Julio Ernesto Ortiz. *Diseño e Implementación de un Mercado de Datos para la Oficina Nacional de Estadísticas*. Ciudad de La Habana : s.n., 2009.
10. Data Prix. *Data Prix*. [En línea] [Citado el: 4 de Noviembre de 2011.] <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/arquitectura-del-data-warehouse/34-datawarehouse-manager>.
11. Rubia, J. M. *Introducción a los almacenes de datos*. 2009.

12. GestioPolis. *GestioPolis*. [En línea] [Citado el: 10 de Noviembre de 2011.] <http://www.gestipolis.com/administracion-estrategia/almacenes-de-datos-y-microsoft-sql-server.htm>.
13. Rivera, Juan David Peña, Jesús Armando y Suárez, Daza. *Utilización de Información Histórica para Decisiones Empresariales*. Santafé de Bogotá : s.n., 2005.
14. Hernández, Yanisbel González. *Propuesta de Metodología para el Desarrollo de Almacenes de Datos en DATEC*. La Habana : s.n., 2012.
15. autores, Colectivo de. *Metodología de Desarrollo de Soluciones BI y Warehousing (BI & W)*. La Habana : s.n.
16. Visual Paradigm. *Visual Paradigm*. [En línea] [Citado el: 11 de noviembre de 2011.] <http://www.visual-paradigm.com>.
17. A. Riscos Núñez. [En línea] [Citado el: 13 de Noviembre de 2011.] <http://www.cs.us.es/cursos/bd/tema1.pdf>.
18. PostgreSQL-es. *PostgreSQL-es*. [En línea] [Citado el: 11 de noviembre de 2011.] http://www.postgresql.org.es/sobre_postgresql.
19. EncuRed. *EncuRed*. [En línea] [Citado el: 13 de noviembre de 2011.] <http://www.ecured.cu/index.php/PostgreSQL>.
20. ArPUG Grupo de Usuarios PostgreSQL de Argentina. *ArPUG Grupo de Usuarios PostgreSQL de Argentina*. [En línea] [Citado el: 2 de noviembre de 2011.] <http://www.arpug.com.ar/trac/wiki/PgAdmin>.
21. Apache Tomcat. *Apache Tomcat*. [En línea] [Citado el: 4 de noviembre de 2011.] <http://tomcat.apache.org/>.
22. DataCleaner. *DataCleaner*. [En línea] [Citado el: 4 de noviembre de 2011.] <http://datacleaner.eobjects.org/>.
23. Gravatar. *Gravatar*. [En línea] [Citado el: 5 de noviembre de 2011.] <http://www.gravatar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/>.

24. stratebi. *stratebi*. [En línea] [Citado el: 15 de noviembre de 2011.] http://www.stratebi.es/todobi/jun10/Comparativa_OSBI.pdf.
25. Mario Saffiro. Tecnologías de Información y Gestión de Procesos de Negocios (BPM). *Tecnologías de Información y Gestión de Procesos de Negocios (BPM)*. [En línea] [Citado el: 20 de enero de 2012.] <http://msaffirio.wordpress.com/2011/08/20/reglas-de-negocio-business-rules/>.
26. autores, Colectivo de. puc-rio.br. *puc-rio.br*. [En línea] [Citado el: 22 de enero de 2012.] wer.inf.puc-rio.br/WERpapers/artigos/artigos_wer00/toro.pdf.
27. Vilalta Consultores. vico.org. *vico.org*. [En línea] [Citado el: 25 de marzo de 2012.] www.vico.org/aRecursos/TRAD_Cards/.../TRAD_Req_CUs_esp.pdf.
28. Scribd. *Scribd*. [En línea] [Citado el: 25 de marzo de 2012.] <http://es.scribd.com/doc/32165558/CAPITULO-17-ESTRATEGIAS-DE-PRUEBA-DEL-SOFT>.
29. delfin. *delfin*. [En línea] [Citado el: 25 de marzo de 2012.] delfin.mxl.uabc.mx/~angelica/Pruebas.pdf.

BIBLIOGRAFÍA

1. Berríos, Ericka Graciela Sevilla. *GUÍA METODOLÓGICA PARA LA DEFINICIÓN Y DESARROLLO DE UN DATA WAREHOUSE*. Managua, Nicaragua : s.n., 2003.
2. Kimball, Ralph y Ross, Margy. *The Data Warehouse Toolkit Second Edition The Complete Guide to Dimensional Modeling*. Estados Unidos de América : Wiley Computer Publishing, 2002. ISBN 0-471-20024-7.
3. Cabrera, Maria Evelia Casales. Facultad de Ciencias. *Facultad de Ciencias*. [En línea] [Citado el: 2 de noviembre de 2011.] <http://hp.fciencias.unam.mx/~alg/bd/dwh.pdf>.
4. Alarcos. *Alarcos*. [En línea] [Citado el: 11 de Noviembre de 2011.] <http://alarcos.inf-cr.uclm.es/doc/bbddavanzadas/08-09/FUNCIONALIDAD%204.pdf>.
5. García, Gerardo Clemente. *Un Sistema para el Mantenimiento de Almacenes de Datos*. Valencia, España : s.n., 2008.
6. Dataprix. *Dataprix*. [En línea] [Citado el: 3 de noviembre de 2011.] <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/i-data-warehousing-investigacion-y-sistematizacion-concepto-13>.
7. Sinnexus. *Sinnexus*. [En línea] [Citado el: 29 de octubre de 2011.] www.sinnexus.com.
8. DeciSop Todo sobre toma de decisiones. *DeciSop Todo sobre toma de decisiones*. [En línea] [Citado el: 3 de Noviembre de 2011.] <http://decisop.wordpress.com/category/toma-de-decisiones/>.
9. Sierra, Julio Ernesto Ortiz. *Diseño e Implementación de un Mercado de Datos para la Oficina Nacional de Estadísticas*. Ciudad de La Habana : s.n., 2009.
10. Data Prix. *Data Prix*. [En línea] [Citado el: 4 de Noviembre de 2011.] <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/arquitectura-del-data-warehouse/34-datawarehouse-manager>.
11. Rubia, J. M. *Introducción a los almacenes de datos*. 2009.

12. GestioPolis. *GestioPolis*. [En línea] [Citado el: 10 de Noviembre de 2011.] <http://www.gestipolis.com/administracion-estrategia/almacenes-de-datos-y-microsoft-sql-server.htm>.
13. Rivera, Juan David Peña, Jesús Armando y Suárez, Daza. *Utilización de Información Histórica para Decisiones Empresariales*. Santafé de Bogotá : s.n., 2005.
14. Hernández, Yanisbel González. *Propuesta de Metodología para el Desarrollo de Almacenes de Datos en DATEC*. La Habana : s.n., 2012.
15. autores, Colectivo de. *Metodología de Desarrollo de Soluciones BI y Warehousing (BI & W)*. La Habana : s.n.
16. Visual Paradigm. *Visual Paradigm*. [En línea] [Citado el: 11 de noviembre de 2011.] <http://www.visual-paradigm.com>.
17. A. Riscos Núñez. [En línea] [Citado el: 13 de Noviembre de 2011.] <http://www.cs.us.es/cursos/bd/tema1.pdf>.
18. PostgreSQL-es. *PostgreSQL-es*. [En línea] [Citado el: 11 de noviembre de 2011.] http://www.postgresql.org.es/sobre_postgresql.
19. EncuRed. *EncuRed*. [En línea] [Citado el: 13 de noviembre de 2011.] <http://www.ecured.cu/index.php/PostgreSQL>.
20. ArPUG Grupo de Usuarios PostgreSQL de Argentina. *ArPUG Grupo de Usuarios PostgreSQL de Argentina*. [En línea] [Citado el: 2 de noviembre de 2011.] <http://www.arpug.com.ar/trac/wiki/PgAdmin>.
21. Apache Tomcat. *Apache Tomcat*. [En línea] [Citado el: 4 de noviembre de 2011.] <http://tomcat.apache.org/>.
22. DataCleaner. *DataCleaner*. [En línea] [Citado el: 4 de noviembre de 2011.] <http://datacleaner.eobjects.org/>.
23. Gravatar. *Gravatar*. [En línea] [Citado el: 5 de noviembre de 2011.] <http://www.gravatar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/>.

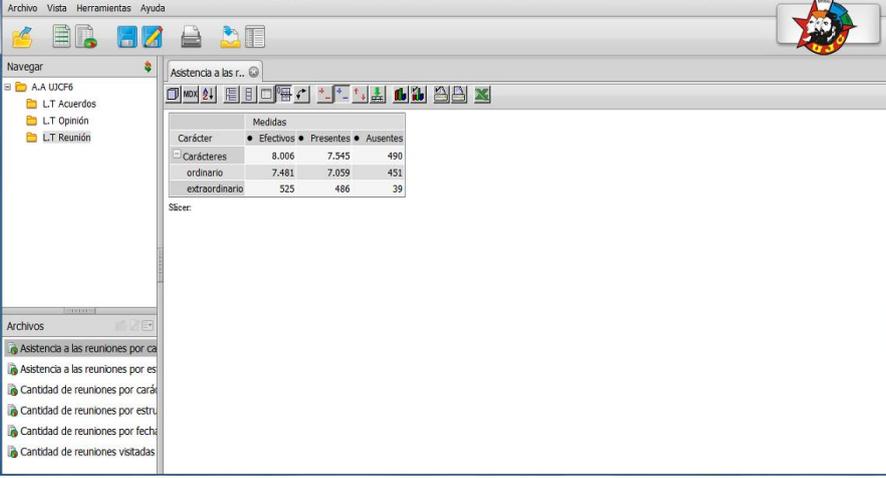
24. stratebi. *stratebi*. [En línea] [Citado el: 15 de noviembre de 2011.] http://www.stratebi.es/todobi/jun10/Comparativa_OSBI.pdf.
25. Mario Saffiro. Tecnologías de Información y Gestión de Procesos de Negocios (BPM). *Tecnologías de Información y Gestión de Procesos de Negocios (BPM)*. [En línea] [Citado el: 20 de enero de 2012.] <http://msaffirio.wordpress.com/2011/08/20/reglas-de-negocio-business-rules/>.
26. autores, Colectivo de. puc-rio.br. *puc-rio.br*. [En línea] [Citado el: 22 de enero de 2012.] wer.inf.puc-rio.br/WERpapers/artigos/artigos_wer00/toro.pdf.
27. Vilalta Consultores. vico.org. *vico.org*. [En línea] [Citado el: 25 de marzo de 2012.] www.vico.org/aRecursos/TRAD_Cards/.../TRAD_Req_CUs_esp.pdf.
28. Scribd. *Scribd*. [En línea] [Citado el: 25 de marzo de 2012.] <http://es.scribd.com/doc/32165558/CAPITULO-17-ESTRATEGIAS-DE-PRUEBA-DEL-SOFT>.
29. delfin. *delfin*. [En línea] [Citado el: 25 de marzo de 2012.] delfin.mxl.uabc.mx/~angelica/Pruebas.pdf.
30. Domínguez, Candelaria Yurena Ávila. *Almacén de datos (DATA WAREHOUSE)*.
31. Luján-Mora, Sergio. *Diseño de Almacenes de Datos con UML*.
32. Rojas, Richard y Boucchechter, Israel. *Ciclos de vida de ingeniería del software*. 2005.
33. Alarcos. *Alarcos*. [En línea] [Citado el: 26 de marzo de 2012.] <http://alarcos.inf-cr.uclm.es/doc/ISOFTWAREI/Tema09.pdf>.
34. Sánchez, Leopoldo Zenaido. *Metodología para el Diseño Conceptual de Almacenes de Datos*. Valencia, España : s.n., 2008.

ANEXOS

Anexo 1. Especificación de los casos de uso de información.**Caso de uso (CU) Presentar información relacionada con las reuniones.**

Objetivo	Presentar información solicitada por el actor que está relacionada con las reuniones de la UJC, permitiendo conocer al actor la información sobre: la asistencia a las reuniones, la cantidad de reuniones que fueron visitadas, así como la cantidad de reuniones según la estructura y carácter de la misma.	
Actores	Analista	
Resumen	El caso de uso inicia cuando el actor desea consultar la información relacionada con las reuniones desde diferentes perspectivas de análisis. Luego de seleccionar el reporte deseado, el sistema muestra la información contenida en él y las opciones de los posibles cambios que le puede hacer al reporte. El caso de uso finaliza cuando el especialista termina de analizar la información relacionada con las reuniones.	
Complejidad	Alta	
Prioridad	Media	
Precondiciones	El usuario se autenticó correctamente. Los datos correspondientes fueron cargados en el mercado de datos. Los reportes relacionados con las reuniones fueron creados.	
Postcondiciones	Los reportes correspondientes al caso de uso fueron consultados.	
Flujo de eventos		
Flujo básico Presentar información relacionada con las reuniones		
	Actor	Sistema
1.	Se autentica en el sistema	
2.		Valida que los datos sean correctos. Muestra la interfaz principal con el

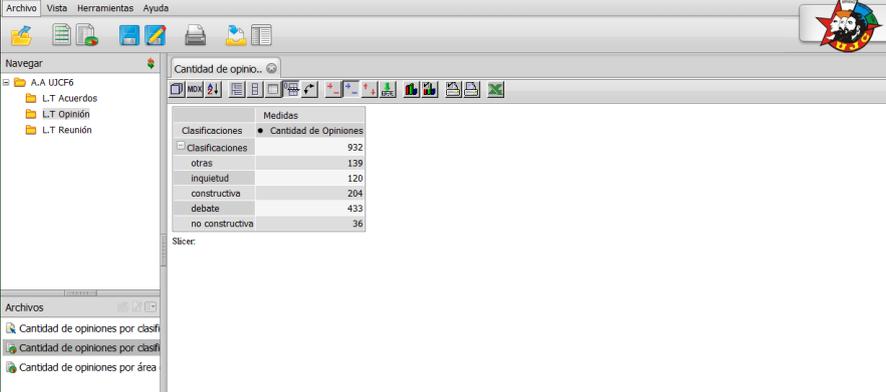
		área de análisis existente.
3.	Selecciona el área de análisis A.A.UJCF6	
4.		Muestra los libros de trabajo contenidos en el A.A.UJCF6
5.	Selecciona el libro de trabajo L.T Reuniones	
6.		Muestra los reportes contenidos en el L.T Reuniones
7.	Selecciona el reporte que desea analizar	
8.		Muestra la información contenida en el reporte seleccionado y brinda opciones al actor para hacer cambios al reporte durante su análisis. Ir al caso de uso Realizar operaciones sobre reportes. Finaliza el caso de uso.
Flujos alternos		
2a. Los datos son incorrectos.		
	Actor	Sistema
		Muestra un mensaje de error. Vuelve al paso 1 del flujo básico.
Opciones de reportes de Presentar información relacionada con las reuniones		
Perspectivas de análisis		Posibles resultados
		Medidas
		Periodicidad
Variables de entrada relacionadas con el caso de uso Presentar información relacionada con las reuniones:	Variables de salida disponibles en el hecho Reuniones:	Rango de tiempo en que se solicitan las variables de salida:
Estructura	-Efectivos	Mensual/Anual
Tiempo_día	-Presentes	

<p>Carácter</p> <p>Comité_base</p>	<p>-Ausentes</p> <p>-Cantidad de visitantes</p> <p>-Cantidad de opiniones</p> <p>-Cantidad de reuniones</p> <p>-Cantidad de reuniones visitadas.</p>																	
<p>Prototipo de interfaz de usuario:</p>	 <table border="1" data-bbox="715 862 957 952"> <thead> <tr> <th>Carácter</th> <th>Efectivos</th> <th>Presentes</th> <th>Ausentes</th> </tr> </thead> <tbody> <tr> <td>Caracteres</td> <td>8.006</td> <td>7.545</td> <td>490</td> </tr> <tr> <td>ordinario</td> <td>7.481</td> <td>7.059</td> <td>451</td> </tr> <tr> <td>extraordinario</td> <td>525</td> <td>486</td> <td>39</td> </tr> </tbody> </table>		Carácter	Efectivos	Presentes	Ausentes	Caracteres	8.006	7.545	490	ordinario	7.481	7.059	451	extraordinario	525	486	39
Carácter	Efectivos	Presentes	Ausentes															
Caracteres	8.006	7.545	490															
ordinario	7.481	7.059	451															
extraordinario	525	486	39															
<p>Relaciones</p>	<p>Casos de Uso Incluidos</p>	<p>No aplica.</p>																
	<p>Caso de Uso Extendidos</p>	<p>Realizar operaciones sobre reportes: Paso 8 del Flujo Básico.</p> <p>Realizar operaciones sobre reportes en el caso de uso Presentar información relacionada con las reuniones.</p>																
<p>Requisitos funcionales</p>	<p>no</p> <p>Sección: “3.2 Requisitos no funcionales” del documento: “0114_Especificación de requisitos de software”.</p>																	
<p>Asuntos pendientes</p>	<p>Posibles mejoras al caso de uso.</p>																	

Caso de uso Presentar información relacionada con las opiniones.

Objetivo	Presentar información solicitada por el actor que está relacionada con las opiniones emitidas en las reuniones de la UJC, permitiendo conocer al actor la cantidad de opiniones que fueron emitidas según la clasificación y el área de opinión.	
Actores	Analista	
Resumen	El caso de uso inicia cuando el especialista desea consultar la información relacionada con las opiniones desde diferentes perspectivas de análisis. Luego de seleccionar el reporte deseado, el sistema muestra la información contenida en él y las opciones de los posibles cambios que le puede hacer al reporte. El caso de uso finaliza cuando el especialista termina de analizar la información relacionada con las opiniones.	
Complejidad	Alta	
Prioridad	Media	
Precondiciones	El usuario se autenticó correctamente. Los datos correspondientes fueron cargados en el mercado de datos. Los reportes relacionados con las opiniones fueron creados.	
Postcondiciones	Los reportes correspondientes al caso de uso fueron consultados.	
Flujo de eventos		
Flujo básico Presentar información relacionada con las opiniones		
	Actor	Sistema
1.	Se autentica en el sistema	
2.		Valida que los datos sean correctos. Muestra la interfaz principal con el área de análisis existente.
3.	Selecciona el área de análisis A.A.UJCF6	
4		Muestra los libros de trabajo contenidos en el A.A.UJCF6
5.	Selecciona el libro de trabajo L.T Opiniones	
6.		Muestra los reportes contenidos en el L.T Opiniones

7.	Selecciona el reporte que desea analizar	
8.		Muestra la información contenida en el reporte seleccionado y brinda opciones al actor para hacer cambios al reporte durante su análisis. Ir al caso de uso Realizar operaciones sobre reportes. Finaliza el caso de uso.
Flujos alternos		
2a. Los datos son incorrectos.		
	Actor	Sistema
		Muestra un mensaje de error. Vuelve al paso 1 del flujo básico.
Opciones de reportes de Presentar información relacionada con las opiniones		
Perspectivas de análisis		Posibles resultados
		Medidas
		Periodicidad
Variables de entrada relacionadas con el caso de uso Presentar información relacionada con las opiniones:	Variables de salida disponibles en el hecho	Rango de tiempo en que se solicitan las variables de salida:
Clasificación	Opiniones:	
Área	-Cantidad de opiniones	Mensual/Anual
Tiempo_día		
Comité_base		
Estructura		
Carácter		

<p>Prototipo de interfaz de usuario:</p>		
<p>Relaciones</p>	<p>Casos de Uso Incluidos</p>	<p>No aplica.</p>
	<p>Casos de Uso Extendidos</p>	<p>Realizar operaciones sobre reportes: Paso 8 del Flujo Básico.</p> <p>Realizar operaciones sobre reportes en el caso de uso Presentar información relacionada con las opiniones.</p>
<p>Requisitos funcionales</p>	<p>no Sección: “3.2 Requisitos no funcionales” del documento: “0114_Especificación de requisitos de software”.</p>	
<p>Asuntos pendientes</p>	<p>Posibles mejoras al caso de uso.</p>	

GLOSARIO

UJC: Unión de Jóvenes Comunistas.

DATEC: Centro de Tecnologías de Gestión de Datos.

ETL: proceso de extracción, transformación y carga.

BI: Inteligencia de Negocio.

Licencia BSD: las licencias BSD son una familia de licencias de software libre permisiva. Poseen menos restricciones en comparación con otras licencias y además permiten el uso del código fuente en software no libre.

Multihilos: un hilo de ejecución, en sistemas operativos, es una característica que permite a una aplicación realizar varias tareas a la vez, es decir concurrentemente. Los distintos hilos de ejecución comparten una serie de recursos, tales como: el espacio de memoria y los archivos abiertos. Esta técnica permite simplificar el diseño de una aplicación que debe llevar a cabo distintas funciones simultáneamente.

Array: los arrays o matrices son un tipo de variable que permiten tener más de un elemento o valor en su interior, a los que se pueden acceder mediante un índice.

Java Servlet: es un programa que se ejecuta en un servidor. Por lo general son aplicaciones Java que corren en un entorno de servidor web.

Java Server Pages: es una tecnología Java que permite a los desarrolladores generar contenido dinámico para web, en forma de documentos HTML, XML u otros.

XML: es un formato universal para documentos y datos estructurados en Internet; este estándar permite el intercambio de información estructurada entre diferentes plataformas. Por lo que se puede usar en bases de datos, editores de texto y hojas de cálculo.

Java Script: es un lenguaje interpretado orientado a las páginas web, con una sintaxis semejante a la del lenguaje Java.

Webservices: es un sistema software diseñado para soportar la interoperabilidad máquina-máquina a través de una red.

HTTP: protocolo de transferencia de hipertexto, es el método más común de intercambio de información en la web (www), es utilizado para transferir las páginas web a un ordenador.

Firefox: es un navegador web libre y de código abierto.

TCP/IP: software de red basado en protocolos de red estándar del sector. Proporciona un marco de trabajo robusto, escalable, multiplataforma y cliente-servidor.

Java Virtual Machine: es la máquina virtual de Java, es decir funciona como intérprete de Java.

SQL: es un lenguaje de consulta estructurado y declarativo de acceso a bases de datos relacionales, que permite especificar diversos tipos de operaciones en ellas.

MDX: lenguaje de consulta multidimensional.

API: una API o Interfaz de Programación de Aplicaciones, es el conjunto de funciones y procedimientos (o métodos si se refiere a programación orientada a objetos), que ofrece cierta biblioteca para ser utilizado por otro software como una capa de abstracción.

No conformidad: defecto, error o sugerencia que se le hace al equipo de desarrollo una vez encontrada alguna dificultad en lo que se está evaluando.