

Facultad de Ciencias y Tecnologías Computacionales



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

Título: Mercado de Datos para el almacén de libros en la Facultad de Ciencias y Tecnologías Computacionales.

Autores:

Claudia Sánchez López

Julio César Cano Gutiérrez

Tutores:

Ing. Yamila Mateu Romero

Ing. Yanitza de la Caridad Gutiérrez Acea

La Habana, mayo de 2017

“Año 59 de la Revolución”



**"No se vive celebrando victorias,
sino superando derrotas."**

*Mercado de datos para el almacén de libros de la facultad
CITEC.*

Declaración de Autoría

Declaramos ser autores del presente trabajo de diplomas y reconocemos a la Universidad de las Ciencias Informáticas (UCI) los derechos patrimoniales de la misma, con carácter exclusivo.

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

Autores:

Claudia Sánchez López

Julio César Cano Gutiérrez

Tutores:

Ing. Yamila Mateu Romero

Ing. Yanitza de la Caridad Gutiérrez Acea

Mercado de datos para el almacén de libros de la facultad CITEC.

Datos de Autor

Tutora: Ing. Yamila Mateu Romero

Universidad de las Ciencias Informáticas, La Habana, Cuba

Especialidad de graduación: Ingeniería Informática

Años de graduado: 9

Área de investigación: Gestión de proyectos

Correo Electrónico: ymateu@uci.cu

Tutora: Ing. Yanitza de la Caridad Gutiérrez Acea

Universidad de las Ciencias Informáticas, La Habana, Cuba

Especialidad de graduación: Ingeniería Informática

Años de graduado:1

Área de investigación: Almacenes de Datos

Correo Electrónico: ycgutierrez@uci.cu

Autor: Claudia Sánchez López

Universidad de las Ciencias Informáticas, La Habana, Cuba

Correo Electrónico: csanchez@estudiantes.uci.cu

Autor: Julio César Cano Gutiérrez

Universidad de las Ciencias Informáticas, La Habana, Cuba

Correo Electrónico: jccano@estudiantes.uci.cu

Mercado de datos para el almacén de libros de la facultad CITEC.

Agradecimientos

Agradezco a mi Vane por ser el motor que me impulsa a superarme cada vez más en la vida.

A mi madre, gracias por todo tu amor, cariño y paciencia espero que sepas que todos mis logros serán siempre también los tuyos porque no existe mayor bendición que la dicha de que seas mi mamá.

A mi abuela Elvia y mi abuelo Aristides por todas las enseñanzas y los consejos.

A mi hermano Joel por sus ocurrencias y su amor.

A mi papá Julio por su amor y sus sacrificios, mi principal objetivo siempre va a ser que se sientan orgullosos de su hijo, hasta ahora creo que voy bien.

A Aracelis por ser la mejor madrastra del mundo, gracias por la ayuda y las comidas, quiero que sepas que te considero mi segunda mamá

Al Julio el integrante más nuevo que tiene la familia por la felicidad que le has dado a mi mamá y sobre todo por el padre que te has convertido para mi hermano.

A mi tía Yamy y mi tío Pablo por el apoyo y la ayuda incondicional.

A Elizabeth María Mesa Lago por ser una de las personas que más me apoyo y ayudo durante la vida universitaria.

A mis psicólogas Claudia Beatriz, Norys y Deborah por los cafés, las malas noches y los buenos consejos.

A todos los amigos que hice durante todo este tiempo de estudio: a Teudys y Saídy los mejores profesores de toda la universidad a Karel e Iván Presidente y Vicepresidente de la República Emo, al Rolo, Rodolfo, Osmin, Rubén y Víctor por los días y las noches de dotarnos de conocimientos, al piquete del WoW por el estrés, la emoción y las noches sin dormir.

A Enmanuel, Yuniór, Roly, Lenia, Lizardra, Suinny, Milagros, Yanelis, Midalis, Chabelis, René, Asiel, El padrino el médico de la timba, Raciél, Raimel, el Migue, Henry, Leo, Leosdany, El loquiño, Juan Pablo, Darayne, Flavia, Julito con todos ustedes mi gratitud por hacer de la UCI un lugar más divertido.

Mercado de datos para el almacén de libros de la facultad CITÉC.

A todas las personas que aportaron su granito de arena en la realización de este trabajo en especial a los mejores tutores que se pueden pedir: Yamila, Yanitza y el Barra por los conocimientos y la ayuda brindada, a los miembros del tribunal por todas las críticas constructivas, gracias por el apoyo.

A mi dúo de tesis por el apoyo, la confianza y la dedicación en la realización de esta investigación.

A Migdalia la mejor tía de la UCI.

A mi familia en general porque son lo mejor que tengo.

A todos los profesores que de una forma u otra me ayudaron y fueron parte de mi formación durante toda la carrera.

A todos muchas gracias.

Julio César Cano Gutiérrez

Agradezco a todas aquellas personas que de una forma u otra estuvieron involucradas en la realización de este trabajo, a mis tutores, por su apoyo incondicional, por el conocimiento que me supieron brindar y por su gran ayuda, a los miembros del tribunal por sus críticas constructivas en todo lo largo de este proceso, a todos muchas gracias por su apoyo.

A mi dúo de tesis por haberse comportado como lo hizo, por siempre estar cuando había que estar y por haber asumido este reto de realizar la tesis en tan poco tiempo.

A mi mamá Sonia, por ser la gran mujer que es, por estar en todo momento a mi lado, por no dejarme rendirme nunca cuando yo sentía que la única salida era esa, por todos sus consejos, por su apoyo incondicional sin importar el momento ni el lugar, por todos los sacrificios que haces y sé que seguirás haciendo cada día que pasa por mí, dicen que los hijos son los que escogen a sus padres, si la vida me diera la oportunidad de nacer otra vez no dudaría ni un segundo en volver a escogerte como mi mamá, TEAMO.

A mi papa Gustavo, que por cosas de la vida no pudo estar este día aquí presente, pero sé que en estos momentos diera lo que no tiene por estar aquí y ver que su Pipo, como me

Mercado de datos para el almacén de libros de la facultad CITEC.

decía el cuando era niña está cumpliendo su sueño y el de él, porque esa siempre fue mi idea, hacer que mi papá se sintiera orgulloso de tener una hija ingeniera, por todos sus sacrificios, por querer siempre hacer hasta lo imposible por complacerme, por siempre estar pendiente de mí y de mis cosas, por sus consejos y su comprensión, por todo eso y por mucho más yo puedo decir que tengo al mejor papá del mundo, gracias por ser mi papá, TE AMO.

A mis abuelos Mirta y Tavo, por no querer perderse ni un segundo de las cosas que me pasan tanto buenas como malas, siempre están ahí para apoyarme y darme los consejos que necesite, y aunque hoy no pudieron estar aquí por cosas de la edad, sé que en estos momentos están ansiosos por que los llame y les diga como salió todo.

A mis tíos Damaris y Adel por siempre apoyarme y estar pendientes de las cosas que me pasan, por formar parte de mi vida y siempre querer lo mejor para mí, por considerarme como su hija y siempre querer complacerme en todo, por aconsejarme y alarme las orejas en los momentos que ha hecho falta, los quiero.

A mi hermano Gustavito, que, aunque siempre estamos fajados, yo lo quiero mucho, y me gustaría que siguiera mis pasos y se convirtiera en un profesional.

Al calvo de la familia: mi tío Reinerio, a mi prima Dalgis que más que prima es mi hermana y quien lo iba a decir porque mira que nos alabamos los pelos cuando chicas, a mi tía Andrea, a mis tías del valle, Bárbara y Yanira, a mi primo Michel por soportarme desde que era niña.

A fausto y a Victoria, por ser los mejores suegros del mundo, por sus consejos, por estar pendiente de mí, gracias por ser como son conmigo y dejarme formar parte de su linda familia, los quiero.

A Luis Miguel por haberse convertido en una de las personas más importantes de mi vida, por ayudarme cuando más lo necesite, por sus consejos, por su comprensión, su dedicación. Pero lo más importante es darle las gracias por regalarme lo mejor que me han regalado en mi vida y espero con ansias y desvelo el día que lo pueda tener en mis brazos, sé que vamos a hacer los mejores padres del mundo para Bruno.

Mercado de datos para el almacén de libros de la facultad CITEC.

A mi familia en general por siempre estar pendientes de mí cada vez que llegaba la etapa de pruebas y siempre llamaban cada cinco minutos a ver cómo había salido. A todos mil gracias.

A los amigos que hice durante todo este tiempo de estudio: a Yanelis por estar siempre que la necesito y por ser tan especial, a mi negrita Suinny por ser diferente a los demás, pero siempre está lista para dar los mejores consejos, a mi amigo Julio por ser incondicional conmigo y demostrarme siempre que es una de las mejores personas que he conocido, a Milagro, a Lyszandra, a Lijandy, a Rene, a Asiel, a Midalis por nunca olvidarse de buscarme para ir a almorzar, al negro de mi life: Magdiel por esas noches de desvelo y por siempre estar cuando lo necesite, aun cuando ya no estaba en la escuela. A Yanitza porque a pesar de ser una de mis tutoras se convirtió en mi amiga desde el primer día que entre a esta universidad, muchas gracias por todo mi negra. A mis compañeros de mi primer grupo en la UCI, a Leo, al Nicher, a Miguel Ángel.

A todos los profesores que de una manera u otra me ayudaron y fueron partícipes de mi formación durante toda la carrera.

A todos muchas gracias.

Claudia Sánchez López

Mercado de datos para el almacén de libros de la facultad CITEC.

Dedicatoria

Dedico esta tesis especialmente a mi niña Vanessa Cano Hidalgo y a mi hermanito Joel Rizo Gutiérrez, para que este logro les sirva de ejemplo para superarse en la vida.

A mis padres y mis abuelos por todo su esfuerzo y apoyo incondicional y estar cuando más los he necesitado a mi Mita y mi Mamí que son las que más me han peleado para que estudie.

No se puede pedir una mejor familia.

Julio César Cano Gutiérrez

Dedico esta tesis especialmente a mi abuela Caridad, que, aunque la vida me la quito de mi lado muy temprano, sé que en estos momentos estaría muy orgullosa de verme cumplir mi sueño.

A mis padres por siempre apoyarme incondicionalmente y darme fuerzas cuando más la necesito para no rendirme nunca y siempre mirar hacia adelante, gracias por ser los mejores padres del mundo.

Claudia Sánchez López

Mercado de datos para el almacén de libros de la facultad CITEC.

Resumen

La siguiente investigación surge teniendo en cuenta las necesidades de extraer información valiosa del almacén de libros de la Facultad de Ciencias y Tecnologías Computacionales y para facilitar la toma de decisiones en el área del Vicedecanato de Administración. Teniendo en cuenta la cantidad de libros que existen y la cantidad de procesos que se realizan en el almacén, cada vez se generan más información, lo que dificulta el análisis de los mismos por parte de la encargada del local. Teniendo como objetivo el desarrollo de un Mercado de Datos para contribuir al análisis de los datos históricos que de estos procesos se derivan y permitiendo su análisis desde diferentes indicadores claves de seguimiento por el área, para su implementación se utiliza la Metodología para el desarrollo de proyectos de Almacenes de Datos utilizada por el Centro de Tecnologías de Gestión de Datos. Se realiza el análisis, diseño e implementación de sus tres subsistemas: almacenamiento, integración y visualización, haciendo uso de las herramientas como Visual Paradigm 8.0, PostgreSQL 9.4, la Suite de Pentaho 6.0. A través de Casos de Prueba y Listas de Chequeo se contribuyó a que la solución desarrollada garantizara la integridad de todos los requisitos planteados por el cliente. Al concluir exitosamente las pruebas se obtuvieron como resultado de la investigación un Mercado de Datos funcional y poblado con una pequeña muestra de datos que será incrementada a lo largo de tiempo, brindando información valiosa a través de vistas de análisis y reportes operacionales que contribuyen a la toma de decisiones sobre el uso de la bibliografía impresa.

Palabras claves: Almacén de libros, Mercado de Datos, Toma de Decisiones, Vicedecanato de Administración.

Mercado de datos para el almacén de libros de la facultad CITEC.

Abstract

The following investigation happens taking into account the needs to extract valuable information of the store of books of the Faculty of Computational Science and Technologies and to facilitate the decision making in the area of the Vice-Dean's Office of Administration. Taking into account the quantity of books that exist and the quantity of processes that come true in the store, each time generate themselves more data, what makes difficult the analysis of the same on behalf of the forewoman of the site. Aiming at the development of a Data Mart to contribute to the analysis of the historic data that of these processes derive and permitting your analysis from different key follow-up indicators for the area, for your implementation the developmental Methodology of projects of Data Warehouse used by the Center of Tecnologías of Data Management is used. The analysis, design and implementation of your three subsystems are accomplished: Storage, integration and visualization, making use of the tools like Visual Paradigm 8.0, PostgreSQL 9.4, Pentaho's Suite 6.0. Through trial Cases and the Checklists contributed himself to that the developed solution guaranteed the integrity of all the requirements put forward by the customer. When concluding proofs successfully they got as a result of investigation a functional Data Mart and town with a little sample of data that will be incremented in the course of time, offering valuable information through sights of analysis and operational reports that contribute to the photo decisions on the use of the printed bibliography.

Passwords: Store of books, Data Marts, Decision making, Vice-Dean's Office of Administration.

Mercado de datos para el almacén de libros de la facultad CITEC.

Índice

Introducción	1
Capítulo 1: Fundamentación teórica de los Almacenes de Datos.....	6
1.1. Almacenes de Datos.....	6
1.1.1. Definición de Almacén de Datos.....	6
1.1.2. Definición de Mercado de Datos	7
1.1.3. Características de los Almacenes de Datos	7
1.1.4. Características de los mercados de datos.....	8
1.1.5. Modelo Multidimensional	9
1.1.6. Topologías de esquemas.....	10
1.2. Inteligencia de Negocio	13
1.3. Extracción, Transformación y Carga	13
1.4. Sistemas de almacenamiento de datos	14
1.5. Herramientas para el desarrollo de Mercados de Datos	18
1.6. Conclusiones del capítulo	23
Capítulo 2: Análisis y diseño del Mercado de Datos.....	25
2.1. Análisis	25
2.1.1. Descripción del negocio	25
2.1.2. Tema de análisis.....	25
2.2. Necesidades de usuarios.....	26
2.3. Requisitos de información.....	26
2.4. Requisitos funcionales	26
2.5. Requisitos no funcionales	27
2.6. Reglas del negocio	29
2.7. Modelo de casos de uso del sistema.....	30
2.9. Diseño de los subsistemas de almacenamiento, integración y visualización.....	34
2.9.1. Subsistema de Almacenamiento	34
2.9.2. Arquitectura del Almacén de Datos	34
2.9.3. Subsistema de Integración.....	37
2.9.4. Subsistema de Visualización	39
2.10. Políticas de respaldo y recuperación	41
2.11. Conclusiones del capítulo	42
Capítulo 3: Implementación y prueba del Mercado de Datos	43
3.1. Implementación del Subsistema de Almacenamiento	43

Mercado de datos para el almacén de libros de la facultad CITEC.

3.1.1. Estándar de codificación	43
3.2. Implementación del Subsistema de Integración de datos	44
3.2.1. Transformaciones y trabajos	45
3.3. Implementación del Subsistema de Visualización	47
3.3.1. Implementación de la capa de visualización	48
3.3.2 Implementación de los reportes candidatos	48
3.4 Pruebas	50
3.4.1. Herramientas de prueba	53
3.4.2. Resultados de las pruebas	58
Conclusiones del capítulo	59
Conclusiones	60
Recomendaciones	61
Referencias	62
Bibliografía	64

Mercado de datos para el almacén de libros de la facultad CITEC.

Índice de Figuras

Figura 1: Esquema Estrella.....	11
Figura 2: Esquema Copo de Nieve.....	12
Figura 3:Esquema Constelación de Hechos.	13
Figura 4: Ciclo de vida de la Metodología.	17
Figura 5: Diagrama de casos de uso del sistema.	32
Figura 6: Arquitectura del almacén de datos.	35
Figura 7: Modelo de Datos.....	37
Figura 8: Distribución por tipo de datos.....	38
Figura 9: Diseño de las transformaciones para la carga de dimensiones.	38
Figura 10: Diseño de las transformaciones para la carga de hechos.....	39
Figura 11: Diseño del mapa de navegación.....	40
Figura 12: Proceso de integración de datos para el hecho reintegro.....	45
Figura 13: Proceso de integración de datos para la dimensión libro.	46
Figura 14: Trabajo para las dimensiones.	46
Figura 15: Trabajo para los hechos.	47
Figura 16: Implementación de los cubos OLAP.....	47
Figura 17: Arquitectura de información del MD para el almacén de la facultad CITEC.....	48
Figura 18: Reporte reintegro	49
Figura 19: Cantidad por fecha, encargado del almacén y persona.....	50
Figura 20: Modelo v.....	51
Figura 21: Comportamiento de los indicadores por secciones.	58

Mercado de datos para el almacén de libros de la facultad CITEC.

Índice de Tablas

Tabla 1: Descripción de los requisitos funcionales del sistema.....	27
Tabla 2: Descripción de los actores del sistema.....	31
Tabla 3: CU Mostrar información sobre el almacén.....	32
Tabla 4: Descripción de las tablas de dimensiones.....	35
Tabla 5: Descripción de las tablas de hechos.....	35
Tabla 6: Matriz Bus.....	36
Tabla 7: Reporte candidato "Cantidad de préstamos " del A.A Préstamo.	40
Tabla 8: Reporte candidato "Cantidad de reintegros" del A.A Reintegro.	40
Tabla 9: Reporte candidato "Cantidad de saldo" del A.A Almacén.	41
Tabla 10: Roles y permisos.	41
Tabla 11 :Diseño de CP para el CU "Mostrar información sobre el almacén"	53
Tabla 12: Lista de chequeo aplicada.....	55

Mercado de datos para el almacén de libros de la facultad CITEC.

Introducción

Las tecnologías en el mundo de hoy presentan un desarrollo científico técnico, es por ello que la informatización de la sociedad ha crecido a nivel mundial y esto trae consigo que aumente la capacidad de generación y almacenamiento de la información

. Por lo antes señalado las empresas y organizaciones mundiales se han visto en la necesidad de encontrar una nueva solución que almacene esta gran cantidad de datos y a su vez, poder extraer información realmente útil para las mismas, todo esto conlleva al surgimiento de las nuevas Tecnologías de la Información y las Comunicaciones (TIC). Las ventajas y aplicaciones de las TIC son diversas, las mismas brindan soluciones a los problemas que se presentan a diario en diferentes sectores de la sociedad (Almenara, 2005) .

Las TIC proporcionan mecanismos y procesos que permiten escoger la alternativa que sea más conveniente para el éxito de la empresa. Dentro de estos procesos se encuentra la Inteligencia de Negocios (BI - por sus siglas en inglés), que contribuye a la generación del conocimiento necesario para que diferentes empresas puedan aplicar la toma de decisiones basadas en información precisa y oportuna.

La Universidad de las Ciencias Informáticas (UCI), tiene como principal misión formar profesionales comprometidos con su Patria, altamente calificados en la rama de la Informática y servir de soporte a la industria cubana de la informática. La misma cuenta con el Centro de Tecnologías de Gestión de Datos (DATEC), integrado por cuatro líneas de desarrollo, teniendo como principal misión crear bienes y servicios informáticos relacionados con la gestión de datos, área del conocimiento que agrupa tanto a los sistemas de información, como a los denominados sistemas de inteligencia empresarial o de negocios, cuyo propósito fundamental es apoyar el proceso de toma de decisiones. Dentro de estas líneas se encuentran los Almacenes de Datos, la cual realiza soluciones informáticas con el objetivo de apoyar diferentes áreas dentro y fuera de la UCI.

La UCI está conformada por varias facultades, en cada facultad existe un almacén de libros donde los estudiantes adquieren los volúmenes de las asignaturas que cursan como apoyo al proceso docente-educativo. Dentro de estas facultades se encuentra La Facultad de Ciencias y Tecnologías Computacionales (CITEC), la cual cuenta con un depósito de libros donde los estudiantes adquieren los volúmenes de las asignaturas que cursan por año académico. Sin embargo, existen un grupo de datos con los que se puede hacer más efectivo el control de este proceso, como lo son:

- Cantidad de volúmenes total por año académico
- Cantidad de reintegros por fecha.
- Estado de los libros por un número que lo caracterice.

Mercado de datos para el almacén de libros de la facultad CITEC.

- Cantidad de libros en manos de estudiantes y profesores
- Cantidad de libros entregados hasta la fecha

Estos datos se encuentran almacenados en un sistema orientado a la gestión de la información donde se realiza la informatización del proceso, y no orientada al análisis de los datos. Para los especialistas que en este local laboran se hace difícil obtener este tipo de informaciones a lo largo del tiempo en los cursos académicos por cada año y muchas veces estos informes no salen en tiempo o requieren de horas extras para que puedan ser presentados a la administración de la Facultad.

Por todo lo anteriormente planteado surge como **Problema de la investigación:** ¿Cómo contribuir al análisis de datos en el almacén de libros de la Facultad de Ciencias y Tecnologías computacionales?

Dicha investigación tiene como **objeto de estudio:** los Almacenes de Datos. Delimitándose como **campo de acción:** Mercado de Datos para el análisis de la información en el almacén de libros.

Para darle solución al problema planteado, se determina como **objetivo general:** desarrollar un Mercado de Datos para el almacén de libros de la Facultad de Ciencias y Tecnologías Computacionales que permita el almacenamiento de toda la información y la toma de decisiones.

Para darle solución al objetivo general se identificaron los siguientes **objetivos específicos:**

1. Fundamentar la metodología, herramientas y tecnologías a utilizar en el desarrollo del Mercado de Datos.
2. Realizar el análisis y diseño del Mercado de Datos para el almacén de libros de la facultad CITEC.
3. Realizar la implementación del Mercado de Datos para el almacén de libros de la facultad CITEC.
4. Realizar las pruebas al Mercado de Datos para el almacén de libros de la facultad CITEC.

Con el propósito de cumplir con el objetivo general planteado fueron formuladas las siguientes **preguntas científicas:**

1. ¿Qué metodología, herramientas y tecnologías utilizar para llevar a cabo la implementación del Mercado de datos para el almacén de libros de la facultad CITEC?
2. ¿Qué características debe cumplir el Mercado de datos para el almacén de libros de la facultad CITEC para que cumpla con las necesidades de los clientes?
3. ¿Cómo desarrollar el Mercado de datos para el almacén de libros de la facultad CITEC de manera que contribuya al proceso de toma de decisiones?
4. ¿Cómo probar el Mercado de datos para el almacén de libros de la facultad CITEC con el fin de garantizar el correcto funcionamiento del mismo?

Mercado de datos para el almacén de libros de la facultad CITEC.

Se propone la realización de las siguientes **tareas de la investigación** para dar cumplimiento a los objetivos específicos planteados:

1. Estudio, definición y análisis de la metodología y herramientas a utilizar para guiar el desarrollo del Mercado de Datos.
2. Levantamiento de requisitos para identificar las necesidades del cliente.
3. Documentación de los casos de uso para describir las acciones que debe realizar el actor y las respuestas que debe dar el sistema.
4. Definición de los hechos, dimensiones y medidas del mercado de datos para el diseño del modelo de datos.
5. Diseño del modelo de datos para identificar la relación entre las tablas de hechos y dimensiones candidatas en la solución.
6. Definición de la arquitectura del mercado de datos para definir los subsistemas que forman parte de la solución.
7. Diseño del subsistema de almacenamiento para tener almacenada toda la información correspondiente al mercado de datos para la facultad CITEC.
8. Diseño del subsistema de integración para diseñar los procesos de carga de las dimensiones y hechos.
9. Diseño del subsistema de visualización para realizar el diseño de los cubos OLAP y de los reportes candidatos.
10. Diseño de los casos de prueba para aplicarlos luego durante las pruebas del sistema.
11. Implementación del modelo de datos para cumplir con el diseño de la estructura de la base de datos.
12. Implementación del subsistema de integración para poblar la base de datos para el mercado de datos.
13. Implementación del subsistema de visualización para mostrar los reportes contenidos dentro de los libros de trabajo según el área de análisis identificada.
14. Aplicación de las listas de chequeo para verificar la estructura del documento, los indicadores definidos en el desarrollo y la semántica del documento.
15. Aplicación de los casos de prueba para garantizar el correcto funcionamiento del sistema.

Para el desarrollo de la investigación fueron utilizados métodos científicos, dentro de los que se encuentran los métodos teóricos que permiten estudiar las características que no son observables del modelo de investigación, y los métodos empíricos que facilitan la descripción de las características fenomenológicas del objeto.

Mercado de datos para el almacén de libros de la facultad CITEC.

Métodos teóricos: Permiten conocer las relaciones que fluyen alrededor del objeto de estudio.

- **Análisis-Síntesis:** se llevó a cabo para analizar definiciones sobre sistemas existentes, que realicen un trabajo similar al propuesto, con el propósito de arribar a conclusiones que sustenten la necesidad de la investigación. Así como para la apropiación de conocimientos necesarios para el desarrollo del presente trabajo.
- **Inductivo-Deductivo:** la utilización de este método permitió establecer conclusiones y obtener una idea del funcionamiento del sistema, partiendo de la información consultada para el mismo.
- **Histórico-lógico:** es utilizado para analizar la evolución cronológica y funcional de las soluciones existentes vinculadas al campo de acción, para de esta forma alcanzar el objetivo definido.
- **Modelación:** se logra una relación entre el modelo y el objeto, para ello se definen las dimensiones del mercado de datos, el hecho asociado a las dimensiones definidas además se estructura el modelo dimensional y se transforma al diseño físico.

Métodos empíricos: Permiten la observación y el análisis inicial de la información.

- **Entrevista:** se entrevista al vicedecano de administración de la facultad y a la encargada del almacén de libros para el levantamiento de los requisitos, a partir de ellos se definen los requisitos informacionales.
- **Consulta de la información en todo tipo de base:** se utilizó para la elaboración del marco teórico de la investigación.
- **Pruebas informáticas:** permitirán verificar la funcionabilidad de la base de datos.

El presente documento está estructurado como se muestra a continuación: resumen, introducción, capítulo uno, capítulo dos, capítulo tres y conclusiones.

Capítulo 1: Fundamentación teórica de los Almacenes de Datos

En este capítulo se realizará un estudio de los principales conceptos, tecnologías y metodologías que son utilizadas para el desarrollo del almacén de datos, así como sus características, arquitectura, herramientas, ventajas y desventajas.

Capítulo 2: Análisis y diseño del Mercado de Datos

En este capítulo se mostrarán las estructuras multidimensionales modeladas para el almacén de datos, basándose en el análisis realizado anteriormente según lo describe la metodología utilizada. Con la aplicación de esta metodología se lleva a cabo la construcción de un modelo lógico y su implementación a partir del modelo conceptual propuesto.

Capítulo 3: Implementación y prueba del Mercado de Datos

En este capítulo se realizará la implementación del Mercado de datos para el almacén de libros de la facultad CITEC. Se lleva a cabo la implementación del modelo de datos, así como de los reportes

Mercado de datos para el almacén de libros de la facultad CITEC.

candidatos. Se llevarán a cabo las pruebas al Mercado de datos para el almacén de libros de la facultad CITEC. Se realizará la configuración de la seguridad de los usuarios y las pruebas necesarias.

Mercado de datos para el almacén de libros de la facultad CITEC.

Capítulo 1: Fundamentación teórica de los Almacenes de Datos.

En este capítulo se realiza el análisis de los principales elementos para el desarrollo de la solución propuesta. Se abordarán los principales elementos teóricos relacionados con los Almacenes de Datos (AD) y Mercados de Datos (MD); reflejándose las principales características, ventajas y desventajas que ayudan a su desarrollo. Además, se realiza un estudio y selección de las herramientas a utilizar en la solución para las diferentes etapas por las que transita, la metodología adoptada para guiar el proceso de desarrollo y a su vez reconocer las diferentes actividades, roles, procesos y características que posee la misma. También se describen las diferentes fases para la construcción de los AD y MD para lograr un mejor entendimiento de los procesos que son implementados.

1.1. Almacenes de Datos

Actualmente las empresas dedican una parte importante de su tiempo y de sus recursos económicos y humanos a la obtención, procesamiento y proyección de la información para el desarrollo de sus actividades cotidianas. La información es un recurso vital para toda organización, y el buen manejo de esta puede significar la diferencia entre el éxito o el fracaso para todos los proyectos que se emprendan dentro del negocio. Muchas de estas entidades presentan un gran reto, debido al excesivo volumen de información que maneja, la aparición de diversos formatos, influyendo esto en la demora de respuesta hacia los usuarios que dependan de la misma. Con el pasar de los años la información es considerada un fenómeno que crece a raíz de las actividades generadas por las entidades, donde cada una depende de la otra y apoyan diferentes áreas dentro de la institución. Esto trae consigo un aumento considerable de tiempo y acumulación de datos a procesar, por lo que se hace necesaria la gestión de toda esta información para tener un control más amplio sobre el funcionamiento de todas las actividades de la organización. Con el propósito de apoyar el proceso de toma de decisiones, los AD (también conocidos como Data Warehouse), permiten consultar, analizar y presentar la información que sea requerida por los especialistas, para darle respuesta a peticiones realizadas por el usuario y contribuir así a la toma de decisiones por parte de la gerencia de la entidad o la empresa. (Imon B. , 2002)

1.1.1. Definición de Almacén de Datos

Existen diversas tendencias y formas de conceptualizar a los AD y aunque se diferencian en algunos aspectos, todos estos conceptos giran sobre un mismo eje central. A continuación, se enuncian los conceptos definidos por Bill Inmon, Ralph Kimball:

Definición de Bill Inmon

Bill Inmon a principios de los noventa definió el tema de los Almacenes de Datos (Data Warehouse, DW): *“Un almacén de datos es una colección de datos orientados por temas, integrados, variables en el tiempo y no volátiles para el apoyo de la toma de decisiones”* (Imon B, 2002). Orientados por temas porque los

Mercado de datos para el almacén de libros de la facultad CITEC.

datos son estructurados por temas de interés para facilitar su análisis por parte de los usuarios; integrado debido a que los datos que se introducen en el almacén de datos se obtienen de una variedad de fuentes de datos y el proceso de Extracción, Transformación y Carga (ETL) es el encargado de su integración; variables en el tiempo porque es importante conocer la variación de los datos en el tiempo, por esto los datos siempre tienen que estar ligados a un instante de tiempo y no volátil porque en el almacén los datos pueden ser consultados pero no modificados, por tanto la información es permanente y la actualización consiste exclusivamente en la incorporación de nuevos datos.

Definición de Ralph Kimball

Ralph Kimball es reconocido a nivel mundial en el diseño de los almacenes de datos como creador del enfoque multidimensional. Se ha dedicado al desarrollo de su metodología para que este concepto sea bien aplicado en las organizaciones y se asegure la calidad en el desarrollo de estos proyectos. Define almacén de datos como *"una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis"*. También fue Kimball quien determinó que un almacén de datos no era más que: *"la unión de todos los Data Mart de una entidad"* (Kimball, 2002).

Siguiendo y tomando de base las dos definiciones enunciadas por Bill Inmon y Ralph Kimball se decidió tomar como concepto para usar durante el desarrollo del mercado:

"Una colección de datos estructurados, orientado a temas, integrados, no volátiles y variables en el tiempo, donde la información es almacenada para el apoyo a la toma de decisiones y puede ser agrupada en unidades lógicas llamadas mercado de datos"

1.1.2. Definición de Mercado de Datos

"Un MD es un base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento" (Kimball, 2002).

1.1.3. Características de los Almacenes de Datos

Los AD reúnen características especiales, las cuales se mencionan a continuación:

- **Integrado:** como los datos almacenados provienen de fuentes diferentes deben integrarse en una estructura consistente que elimine las inconsistencias existentes en los mismos.
- **Temático:** los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales.
- **Histórico:** los datos se organizan y almacenan en jerarquías en el tiempo, lo que permite análisis comparativos de estados actuales y de períodos anteriores.

Mercado de datos para el almacén de libros de la facultad CITEC.

- **No volátil:** el almacén de datos puede ser leído, pero no modificado. Es decir, se incorporan los últimos valores que tomarán las distintas variables contenidas en él, sin ningún tipo de acción sobre los valores que ya existían. (Imon, 2005)

1.1.4. Características de los mercados de datos

- Según las necesidades de los usuarios el diseño del mercado de datos se realiza siguiendo una estructura consistente.
- La información histórica que posee es mínima comparada con la información histórica que guardan los almacenes de datos.
- Presentan mayor nivel de detalle, por lo que contienen el grado de granularidad necesaria.
- A la hora de construirlo presenta costes adicionales en *hardware*, *software* y accesos de red.
- Debido a que hay grupos de usuarios que solo acceden a un subconjunto preciso de datos, se hace más fácil el acceso a las herramientas de consulta y divide los datos para controlar mejores accesos. (Kimbal, 2002)

Ventajas y desventajas de los AD y MD según (Orallo, 2010)

Entre las ventajas de un almacén de datos se encuentran:

- Proporciona una herramienta para la toma de decisiones en cualquier área funcional, basándose en información integrada y global del negocio.
- Facilita la aplicación de técnicas y estadísticas de análisis para encontrar relaciones ocultas entre los datos del almacén; obteniendo un valor añadido para el negocio de dicha información.
- Mejora la entrega de información, es decir, información completa, correcta, consistente, oportuna y accesible; información que los usuarios necesitan, en el momento adecuado y en el formato apropiado.
- Proporciona la capacidad de aprender de los datos del pasado y de predecir situaciones futuras en diversos escenarios.
- Simplifica dentro de la empresa la implantación de sistemas de gestión integral de la relación con el cliente.
- Permite reaccionar rápidamente a los cambios del mercado.
- Aumenta la competitividad en el mercado.

Mercado de datos para el almacén de libros de la facultad CITEC.

- Se consolida información de diferentes sistemas de origen, sin importar si estos provienen de la misma o varias fuentes.
- Existe consistencia de la información ya que se logra consolidar varios departamentos en uno solo. Es más fácil la toma de decisiones con la información consolidada que separada.
- El hecho de tener información ya almacenada y consolidada hace más fácil realizar el análisis de la misma.
- Realizar un almacén de datos provee las ventajas de utilizar información de múltiples fuentes de información sin importar la compatibilidad de ambas.
- Un almacén de datos ayuda a tener mejores tiempos de respuestas y mejora el proceso de producción.

Entre las desventajas de un AD Y MD:

- Requiere una gran inversión, debido a que su correcta construcción no es tarea sencilla y consume muchos recursos, además, su implementación implica desde la adquisición de herramientas de consulta y análisis, hasta la capacitación de los usuarios.
- Existe resistencia al cambio por parte de los usuarios.
- No todos los usuarios confiarán en el almacén de datos en una primera instancia, pero sí lo harán una vez que comprueben su efectividad y ventajas. Además, su correcta utilización surge de la propia experiencia.

1.1.5. Modelo Multidimensional

El modelo multidimensional está diseñado para permitir el almacenamiento y la recuperación eficiente de grandes volúmenes de datos íntimamente relacionados y almacenados, vistos y analizados desde diferentes perspectivas llamadas dimensiones. El mismo busca presentar la información de una manera estándar, sencilla y sobre todo intuitiva para los usuarios, además de que permite el acceso a la información de manera más rápida por parte de los manejadores de bases de datos. En un esquema multidimensional se representa una actividad que es el objeto de análisis (hecho) y las dimensiones que caracterizan la actividad (dimensiones).

El modelo dimensional divide el mundo de los datos en dos conjuntos: las tablas de hecho y de dimensiones.

A continuación, se explican cada una de ellas:

Mercado de datos para el almacén de libros de la facultad CITEC.

Tabla de hechos: Las tablas de hechos contienen los hechos, medidas o indicadores que serán utilizados por los analistas de negocio para apoyar el proceso de toma de decisiones. Las propias tablas de hechos incluyen atributos que podemos necesitar como Dimensiones en los Cubos, unas dimensiones un poco particulares, por su gran cantidad de miembros y por no necesitar de su propia tabla de dimensión, son las conocidas **Dimensiones degeneradas:** este término se refiere al campo que será utilizado como criterio de análisis y que es almacenado en una tabla de hechos, en vez de ser definido como una dimensión. Esto sucede cuando un campo posee el mismo nivel de granularidad que los datos almacenados en una tabla de hechos, y por tal motivo no se pueden realizar agrupaciones o sumalizaciones a través del mismo. Por tanto, se podría plantear la opción de simplemente incluir estos campos en una tabla de dimensión, pero en este caso se estaría manteniendo una fila de esta dimensión por cada fila en la tabla de hechos, por consiguiente, se tendría la duplicación de información y complejidad, que precisamente es lo que se pretende evitar. Por lo que se utiliza este tipo de dimensiones, incluyendo los campos en las tablas de hechos con el objetivo de eliminar la duplicación de los datos y simplificar las consultas.

Tabla de dimensiones: Las tablas de dimensiones definen cómo están los datos lógicamente organizados y proveen el medio para analizar el contexto del negocio, contienen datos cualitativos. Además, representan los aspectos de interés, mediante los cuales los usuarios podrán filtrar y manipular la información almacenada en la tabla de hechos. Los datos dentro de estas tablas, que proveen información del negocio o que describen alguna de sus características, son llamados datos de referencia (Robelo, 2010).

Medidas: atributo numérico de un hecho que representan el comportamiento del negocio relativo a las dimensiones.

1.1.6. Topologías de esquemas

La topología de esquema no es más que la forma en la cual se va a estructurar el depósito de datos. Es importante definir que tipología se empleará, ya que esta decisión afecta considerablemente la elaboración de los modelos dimensional y físico. Generalmente se utiliza la que se adapte mejor a los requerimientos y necesidades del cliente. Para el modelado de los almacenes de datos existen tres topologías. (Bernabeu, 2010)

Esquema Estrella: su estructura es una tabla central y un conjunto de tablas que la atienden radialmente, de este se deriva su nombre, del hecho que su diagrama forma una estrella, con puntos radiales desde el centro. El esquema en estrella es el más simple de interpretar y optimiza los tiempos de respuesta ante las consultas de los usuarios. Este modelo es soportado por casi todas las herramientas de consulta y análisis. Es necesario destacar que este es un esquema totalmente desnormalizado.

Mercado de datos para el almacén de libros de la facultad CITEC.

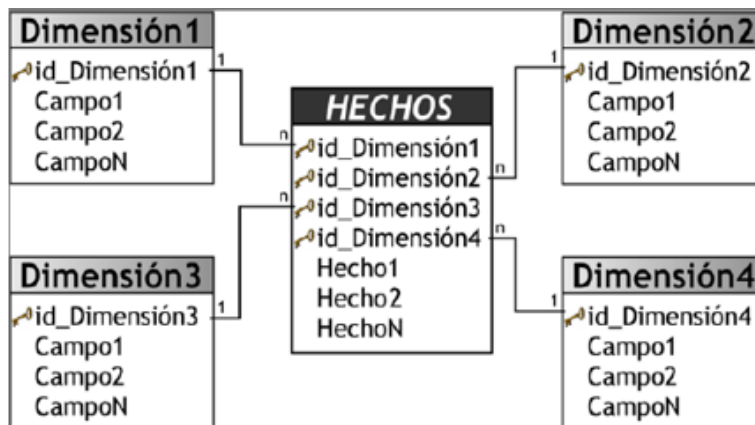


Figura 1: Esquema Estrella.

A continuación, se muestran algunas características de este modelo, que ayudarán a comprender mejor sus ventajas:

- Posee los mejores tiempos de respuesta.
- Su diseño es fácilmente modificable.
- Existe paralelismo entre su diseño y la forma en que los usuarios visualizan y manipulan los datos.
- Simplifica el análisis.
- Facilita la interacción con herramientas de consulta y análisis.

Esquema Copo de Nieve (Snowflake): es un esquema derivado del esquema de estrella, las tablas de dimensiones se ramifican en más puntas. Este modelo es más cercano a un modelo de entidad relación, que, al modelo en estrella, debido a que sus tablas de dimensiones están normalizadas. Uno de los motivos principales de utilizar este tipo de modelo, es la posibilidad de segregar los datos de las tablas de dimensiones y proveer un esquema que sustente los requisitos de diseño. Otra razón es que es muy flexible y puede implementarse después de que se haya desarrollado un esquema en estrella.

Mercado de datos para el almacén de libros de la facultad CITEC.

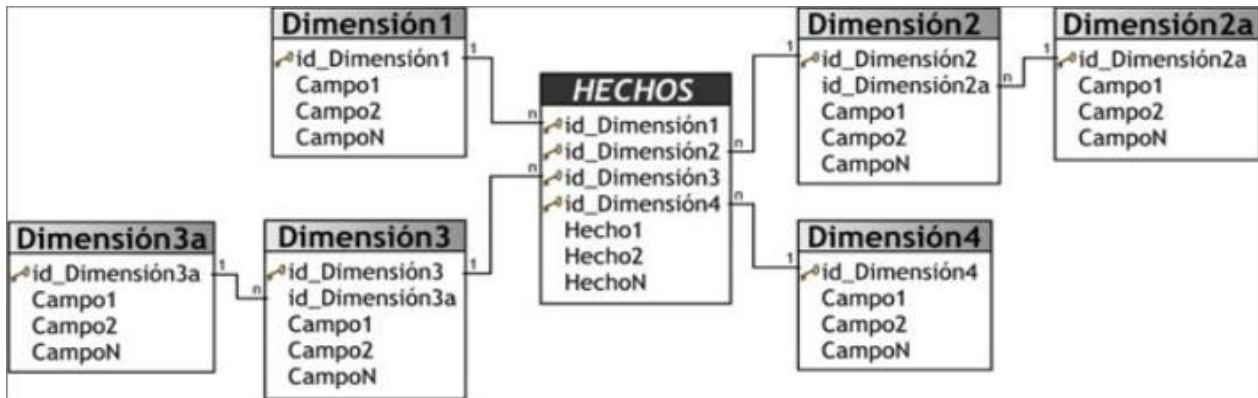


Figura 2: Esquema Copo de Nieve.

Se pueden definir las siguientes características de este tipo de modelo:

- Posee mayor complejidad en su estructura.
- Hace una mejor utilización del espacio.
- Es muy útil en tablas de dimensiones de muchas tuplas.
- Las tablas de dimensiones están normalizadas, por lo que requiere menos esfuerzo de diseño.
- Puede desarrollar clases de jerarquías fuera de las tablas de dimensiones, que permiten realizar análisis de lo general a lo detallado y viceversa.

A pesar de todas las características y ventajas que trae aparejada la implementación del esquema copo de nieve, existen dos grandes inconvenientes:

- Si se poseen múltiples tablas de dimensiones, cada una de ellas con varias jerarquías, se creará un número de tablas bastante considerable, que pueden llegar al punto de ser inmanejables.
- Al existir muchas uniones y relaciones entre tablas, el desempeño puede verse reducido.

La existencia de las diferentes jerarquías de dimensiones debe estar bien fundamentada, ya que de otro modo las consultas demorarán más tiempo en devolver los resultados, debido a que se deben realizar las uniones entre las tablas.

Esquema Constelación de Hechos: la constelación de hechos es un conjunto de tablas de hechos que comparten algunas tablas de dimensiones. Se puede decir que está compuesto por una serie de esquemas en estrella. Su diseño y cualidades son muy similares a las del esquema en estrella, pero posee una serie de diferencias con el mismo, que son precisamente las que lo destacan y caracterizan.

Mercado de datos para el almacén de libros de la facultad CITEC.

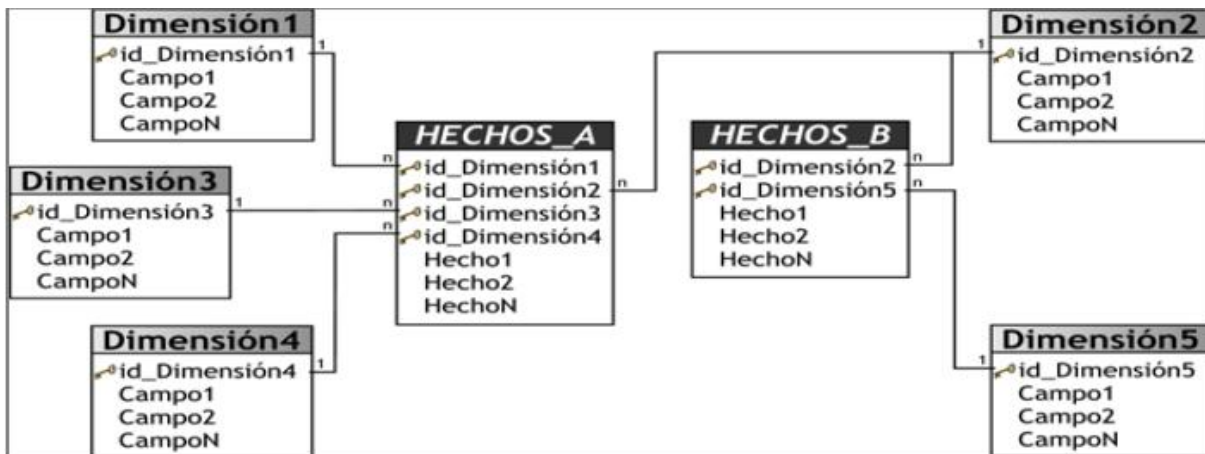


Figura 3:Esquema Constelación de Hechos.

1.2. Inteligencia de Negocio

Inteligencia de negocios o BI, por sus siglas en inglés, según el Data Warehouse Institute, lo define como la combinación de tecnología, herramientas y procesos que permiten transformar los datos almacenados en información, esta información en conocimiento y este conocimiento dirigido a un plan o una estrategia comercial. La inteligencia de negocios debe ser parte de la estrategia empresarial, esta le permite optimizar la utilización de recursos, monitorear el cumplimiento de los objetivos de la empresa y la capacidad de tomar buenas decisiones para así obtener mejores resultados (Sinnexus, 2011).

Algunas de sus principales características son:

- **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.

1.3. Extracción, Transformación y Carga

El término ETL se deriva de sus siglas en inglés Extract-Transform-Load que significan Extraer, Transformar y Cargar. ETL es el proceso que organiza el flujo de los datos entre diferentes sistemas en una organización y aporta los métodos y herramientas necesarias para mover datos desde múltiples fuentes a un almacén de datos, reformatearlos, limpiarlos y cargarlos en otra base de datos. ETL forma parte de la Inteligencia Empresarial (Business Intelligence), también llamado "Gestión de los Datos" (Data Management). La idea es que un proceso ETL lea los datos primarios de sistemas principales, realice

Mercado de datos para el almacén de libros de la facultad CITEC.

transformación, validación, el proceso cualitativo, filtración y al final escriba datos en el almacén y en este momento los datos se encuentran disponibles para analizar por los usuarios (Sinnexus, 2011).

1.4. Sistemas de almacenamiento de datos

Existen dos tipos de almacenamiento, uno es el Procesamiento de Transacciones en Línea (*On-Line Transactional Processing*, OLTP por sus siglas en inglés) y el otro Procesamiento Analítico en Línea (*On-Line Analytical Processing*, OLAP por sus siglas en inglés).

Los sistemas OLTP son bases de datos orientadas al procesamiento de transacciones, al que se le pueden realizar operaciones de inserción, modificación y borrado de datos. Este proceso es típico en bases de datos operacionales que se encargan de la informatización de un proceso donde se realiza toda la gestión del mismo (Sinnexus, 2011).

Por su parte los sistemas OLAP son bases de datos orientadas al procesamiento analítico, en el que el acceso a los datos es de solo lectura y la operación más común es la consulta, inserciones, actualizaciones o eliminaciones. Este sistema permite acceder a grandes volúmenes de datos, de los que se puede extraer información útil. Proporciona análisis estadísticos y permite ver la información en determinadas vistas y dimensiones. Dentro de este sistema se pueden encontrar tres tipos de almacenamiento: el relacional, el multidimensional y el híbrido.

Procesamiento Analítico en Línea Relacional (ROLAP)

El tipo de almacenamiento de datos ROLAP guarda los datos en una base de datos relacional. Este utiliza una arquitectura de tres niveles. La base de datos relacional maneja el almacenamiento de los datos, el motor OLAP proporciona la funcionalidad analítica, y una herramienta especializada es empleada para el nivel de presentación. Como parte de su función se encuentran los usuarios finales que ejecutan su análisis multidimensional, a través del motor OLAP, el cual transforma los datos del lenguaje de consulta estructurado (*Structured Query Language*, SQL por sus siglas en inglés) ejecutadas en las bases de datos relacionales y los resultados son devueltos a los usuarios (Tamayo, 2006).

La arquitectura ROLAP es capaz de usar datos pre-calculados (si estos están disponibles), o de generar dinámicamente los resultados desde la información elemental. Esta arquitectura accede directamente a los datos del almacén y soporta técnicas de optimización para acelerar las consultas como tablas particionadas, soporte a la desnormalización e índices.

Procesamiento Analítico en Línea Multidimensional (MOLAP)

El tipo de almacenamiento de datos MOLAP usa una base de datos multidimensional. Este utiliza una arquitectura de dos niveles: la base de datos multidimensional y el motor analítico. El primer nivel es el encargado del manejo, acceso y obtención de los datos. El nivel de aplicación es el responsable de la

Mercado de datos para el almacén de libros de la facultad CITEC.

ejecución de las consultas OLAP. El nivel de presentación se integra con el de aplicación y proporciona una interfaz a través de la cual los usuarios finales visualizan los análisis OLAP (Tamayo, 2006).

La información procedente de los sistemas transaccionales se carga en el sistema MOLAP. Una vez cargados los datos en la base de datos multidimensional, se realiza una serie de cálculos para obtener datos agregados a través de las dimensiones del negocio, poblando la estructura de la base de datos multidimensional.

Luego de completar esta estructura, se generan índices y se emplean algoritmos de tablas *hash*¹ para mejorar los tiempos de acceso de las consultas. Una vez que el proceso de poblado ha finalizado, la base de datos multidimensional está lista para su uso. Los usuarios solicitan informes a través de la interfaz y la lógica de aplicación de la base de datos multidimensional obtiene los datos.

Procesamiento Analítico en Línea Híbrido (HOLAP)

Se han desarrollado soluciones de OLAP híbridas que combinan el uso de las arquitecturas ROLAP y MOLAP. En una solución con HOLAP, los registros detallados (los volúmenes mayores) se mantienen en la base de datos relacional, mientras que los agregados lo hacen en un almacén MOLAP independiente. En los últimos años se han producido debates alrededor de los tipos de almacenamiento MOLAP y ROLAP. Por lo general, las implementaciones de MOLAP presentan mejor rendimiento que la tecnología relacional; sin embargo, tienen problemas de escalabilidad, por ejemplo, la adición de dimensiones a un esquema ya existente. Por otra parte, las implementaciones de ROLAP son más escalables y atractivas debido a que aprovechan las inversiones efectuadas en tecnología de base de datos relacionales. Una vez analizados los diferentes tipos de almacenamiento se selecciona ROLAP ya que ahorra espacio de almacenamiento y es útil cuando se trabaja con amplios conjuntos de datos. Su ventaja principal reside en la posibilidad de utilizar una tecnología ampliamente extendida y utilizada para la gestión de datos, los sistemas relacionales (Ibarzabal, 2003). Además, se tuvo en cuenta que el Sistema Gestor de Base de Datos (SGBD) PostgreSQL soporta el almacenamiento relacional, no así el multidimensional. En la actualidad los SGBD que dan soporte al almacenamiento multidimensional son propietarios, por lo que no están en correspondencia con las políticas de desarrollo de la UCI y el país.

Una vez analizados los diferentes tipos de almacenamiento se selecciona ROLAP ya que ahorra espacio de almacenamiento y es útil cuando se trabaja con amplios conjuntos de datos. Su ventaja principal reside en la posibilidad de utilizar una tecnología ampliamente extendida y utilizada para la gestión de datos, los sistemas relacionales.

Metodologías de desarrollo para los Almacenes de Datos

¹ Hash es una función computable mediante un algoritmo, que tiene como entrada un conjunto de elementos y los mapea en un rango de salida finito.

Mercado de datos para el almacén de libros de la facultad CITEC.

En el diseño de un almacén de datos, se destacan un conjunto de metodologías que definen y guían todo el ciclo de vida del desarrollo de dicho almacén. Estas diferentes metodologías se pueden englobar dentro de dos grandes bloques: descendente o top-down y ascendente o bottom-up, que se corresponden con las metodologías propuestas por Bill Inmon y Ralph Kimball respectivamente. Estos autores merecen una especial atención porque, en muchos aspectos, se consideran los precursores del Almacén de Datos y sus opiniones son muy valoradas en la industria.

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

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

Basada en el enfoque planteado por Kimball el centro DATEC propone una metodología para este tipo de soluciones: "Metodología para el desarrollo de proyectos de Almacenes de datos", la cual se adapta a las características particulares de la UCI y del centro DATEC.

Metodología para el desarrollo del MD para el almacén de libros de la facultad CITEC

Para el desarrollo de la presente investigación se decidió utilizar la "Metodología para el desarrollo de proyectos de Almacenes de datos" definida por Yanisbel González Hernández en su tesis de maestría (González Hernández, 2010.). El centro definió trabajar con una adaptación del ciclo de vida de la metodología de Kimball por los siguientes elementos:

- Identifica la tabla de hechos y la tabla de dimensiones, lo cual agiliza el proceso de desarrollo y con ello la toma de decisiones.
- Propone la construcción de Mercados de Datos departamentales y después el AD, esto trae como ventaja, que la creación y la puesta en marcha de los Mercados de Datos se produce en un lapso de tiempo corto y después se valora si se construye o no el Almacén de Datos.
- Existe amplia documentación de la misma, así cualquier duda que exista puede ser atendida rápidamente.

Es una metodología madura y reconocida por los usuarios dedicados al tema, además de tener bien definidas sus etapas, actividades, roles y artefactos. Como complemento de este modelo y por las características de trabajo de la UCI se tomó lo planteado por Leopoldo Zenaido Zepeda en su tesis de doctorado en conjunto con un colectivo de autores, que proponen 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 (Zepeda Sánchez, 2008).

Mercado de datos para el almacén de libros de la facultad CITEC.

Esta metodología define que el diseño y construcción de los AD, consta de siete fases y el flujo de trabajo Gestión de Proyecto, el cual se ejecuta durante todo el ciclo de vida del proyecto. De las siete fases, algunas se ejecutan de forma paralela, tal es el caso de las fases Requisitos y Arquitectura, y la fase Diseño e Implementación, donde también se ejecutan componentes en paralelo, lo que facilita que el desarrollo del AD sea más ágil. La Figura 4 muestra el ciclo de vida de la metodología.

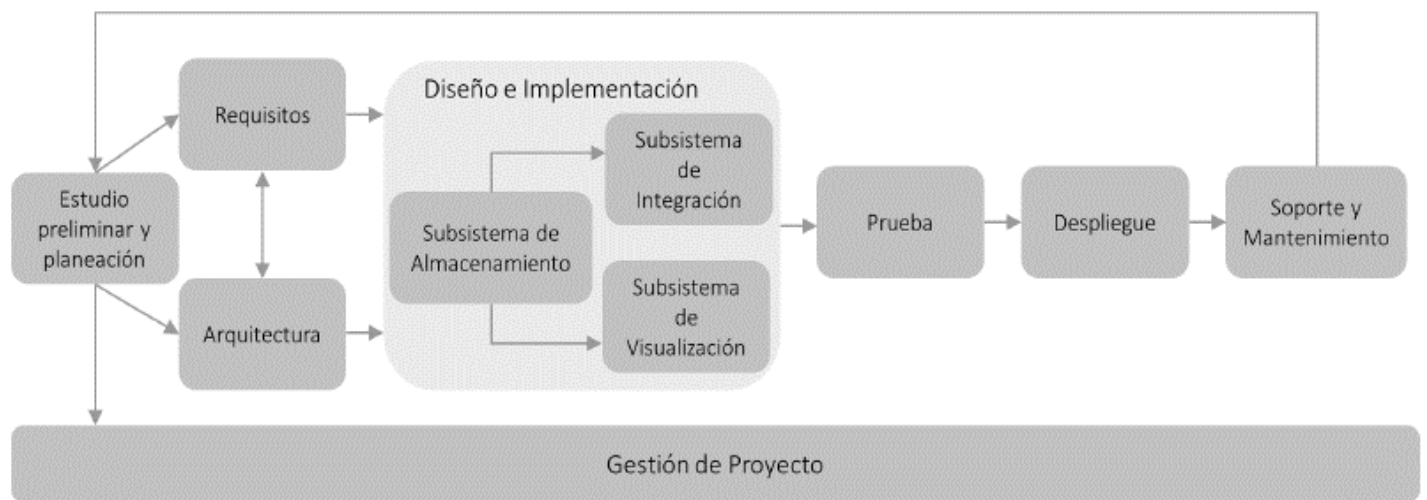


Figura 4: Ciclo de vida de la Metodología.

Fases de la metodología

Estudio preliminar y planeación: En esta fase se realiza un estudio integral de la organización, que se basa en el diagnóstico del negocio, los datos y la infraestructura tecnológica. Además, se definen una serie de aspectos importantes en la gestión de proyecto, tales como el alcance del mismo, los riesgos, la calidad, cronograma, presupuesto, costo, etc.

Requisitos: En esta fase se realizan entrevistas al cliente para identificar las necesidades del mismo, y a partir de estas realizar el levantamiento de los requisitos tanto de información, como funcionales y no funcionales. También se realiza el análisis de los requisitos definidos, lo que permitirá identificar las estructuras bases del modelo lógico dimensional.

Arquitectura: Se definen las vistas arquitectónicas de la solución y aspectos como la seguridad, la tecnología, los subsistemas y la comunicación entre ellos, etc.

Diseño e implementación: En esta fase se diseñan e implementan los tres subsistemas que conforman un AD, definiéndose el diseño de las estructuras de almacenamiento de datos, los procesos de integración, como el mapa lógico, los cubos OLAP, y se implementa el repositorio de datos, la integración de datos y la presentación de los mismos.

Mercado de datos para el almacén de libros de la facultad CITEC.

Prueba: En esta fase se realizan pruebas con el objetivo de validar la calidad del producto. Estas pruebas son de tipo Unitarias que permiten comprobar el funcionamiento de cada subsistema o componente, pruebas de integración, para verificar la correcta integración de los componentes y subsistemas, y pruebas de sistema, donde se valida si fueron cumplidos los requisitos de información y funcionales definidos por el cliente.

Despliegue: Esta fase se divide en dos etapas. Una primera etapa, que consiste en un despliegue piloto, donde se configuran los servidores, se instalan las herramientas en base a la arquitectura definida y se carga una muestra de los datos en un ambiente controlado que le demostrará al cliente que la solución funciona, para luego realizar la carga histórica de los datos. La segunda etapa consiste en la capacitación y transferencia tecnológica de la solución al cliente. Esta etapa tiene como finalidad, una solución desplegada en un entorno real y con un correcto funcionamiento.

Soporte y mantenimiento: Esta fase comienza cuando la solución está implementada y en explotación. Tiene como objetivo evitar que el sistema quede fuera de servicio debido a fallos en su funcionamiento u obsoleto.

Gestión de proyecto: constituye la columna vertebral del mismo, pues es aquí donde se controla, gestiona y chequea el desarrollo del AD, el cronograma y otras actividades de gestión y administración de proyecto.

En la realización del Mercado de Datos para el almacén de libros de la facultad CITEC, de las siete fases existentes de la metodología escogida para su desarrollo solo se utilizarán para el mismo las cinco primeras fases, siendo la fase de prueba la última para la implementación del mercado y el flujo de trabajo gestión de proyecto que se ejecuta durante todo el ciclo de vida del proyecto.

1.5. Herramientas para el desarrollo de Mercados de Datos

Actualmente existe un incremento de las tecnologías de información, creándose una constante competitividad por parte de los productores de estas herramientas, lo cual genera grandes exigencias en cuanto a productividad y calidad de los productos.

✓ **Herramienta de modelado**

Las herramientas CASE (Ingeniería de Software Asistida por Computadora), hoy en día son de gran importancia para la modelación de proyectos, flujos de información, entidades de datos, así como requerimientos del sistema, facilitando el proceso de Planeación. Mucha de la información que se captura durante esta fase se usará también durante el desarrollo y el mantenimiento del ciclo de vida del sistema.

Visual Paradigm For UML 8.0

Herramienta profesional de modelado CASE, que soporta el ciclo de vida completo del desarrollo del software, comprendido en: análisis y diseño orientado a objetos, construcción, pruebas y despliegue. Este

Mercado de datos para el almacén de libros de la facultad CITEC.

software de modelado UML (Lenguaje Unificado de Modelado) ayuda a una rápida construcción de aplicaciones de calidad, sin un elevado costo. Permite elaborar diagramas de clases, así como documentación.

Es una herramienta propietaria que posee varias versiones, como son: Enterprise, Professional, Standard, Modeler, Personal y Community (que es gratuita), además sus distintas ediciones tienen compatibilidad. Soporta aplicaciones Web, aunque las imágenes y reportes generados no son de muy buena calidad. Está disponible en varios idiomas. Posee generación de código para Java en Lenguaje de Marcado de Hipertexto (HTML) y es muy fácil de instalar y actualizar (Technologies, s.f.).

Para el modelado de este sistema se utilizará esta herramienta, debido a las grandes posibilidades que este brinda; en la UCI existe una licencia para la utilización del mismo, lo cual posibilita su utilización (Technologies, s.f.).

✓ **Sistema Gestor de Bases de Datos**

Un Sistema Gestor de Bases de Datos (SGBD) es una colección de programas cuyo objetivo es servir de interfaz entre la base de datos, el usuario y las aplicaciones. Se compone de un lenguaje de definición de datos, de un lenguaje de manipulación de datos y de un lenguaje de consulta. Un SGBD permite definir los datos a distintos niveles de abstracción y manipular dichos datos, garantizando la seguridad e integridad de los mismos (CAVSI, 2016).

PostgreSQL 9.4.1

PostgreSQL es un sistema de gestión de base de datos relacional orientada a objetos y libre, publicado bajo la licencia BSD, es la licencia de software otorgada principalmente para los sistemas BSD (Berkeley Software Distribution). Es más completo que MySQL ya que permite métodos almacenados, restricciones de integridad y vistas.

Principales características de este gestor de bases de datos:

- Implementación del estándar SQL92/SQL99.
- 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 de arreglo (array).
- Incorpora funciones de diversa índole: manejo de fechas, geométricas, orientadas a operaciones con redes, etcétera.

Mercado de datos para el almacén de libros de la facultad CITEC.

- Permite la declaración de funciones propias, así como la definición de disparadores.
- Soporta el uso de índices, reglas y vistas.
- Incluye herencia entre tablas.
- Permite la gestión de diferentes usuarios, como también los permisos asignados a cada uno de ellos.

Ventajas:

- Por su arquitectura de diseño, posee una gran escalabilidad. Es capaz de ajustarse al número de CPU y a la cantidad de memoria que posee el sistema de forma óptima, haciéndole soportar una mayor cantidad de peticiones simultáneas de manera correcta.
- Implementa el uso de reversiones (*rollback's*), subconsultas y transacciones, haciendo su funcionamiento mucho más eficaz.
- Tiene la capacidad de comprobar la integridad referencial, así como también la de almacenar procedimientos en la propia base de datos.
- Tiene mejor soporte para los disparadores (*triggers*) y procedimientos en el servidor (SCRIBD).

Esta es la herramienta a utilizar para el desarrollo del mercado de datos, ya que permite independencia tecnológica, por ser una propuesta de Código Abierto (*Open Source*) que sobrepasa a muchas propietarias y por su gran potencialidad y adaptabilidad al problema en cuestión.

✓ **Administrador de bases de datos**

Existen múltiples herramientas para la administración de bases de datos, para la realización de este trabajo se utilizó PgAdmin III.

PgAdmin III 1.16.1: es una de las herramientas más populares para administrar las bases de datos en PostgreSQL y presenta las siguientes características:

- Es un software libre.
- Accede a todos los objetos del PostgreSQL y responde las necesidades de todos los usuarios, desde escribir consultas SQL simples hasta desarrollar bases de datos complejas.
- Permite que se pueda usar en Linux, FreeBSD, Solaris, Mac OS X y Windows (EASY, 2010).

✓ **Herramienta de Perfilado**

Mercado de datos para el almacén de libros de la facultad CITEC.

DataCleaner 3.1: es una aplicación de código abierto para el perfilado, la validación y comparación de datos. Estas actividades ayudan a administrar y supervisar la calidad de los datos con el fin de garantizar que la información es útil y aplicable a su situación de negocio. Es una de las aplicaciones más fáciles de usar para la calidad de los datos. Normalmente se utiliza antes, durante y después del proceso de Extracción, Transformación y Carga (ETL).

- Antes, para profundizar en los orígenes de los datos que serán usados en el trabajo.
- Durante, en caso de existir cualquier desajuste inesperado durante el proceso ETL.
- Después de asegurar la coherencia y la calidad en la fuente de datos que han poblado.

DataCleaner puede acceder y analizar prácticamente cualquier almacén de datos, incluyendo:

- Base de datos como Oracle, Microsoft SQL Server, PostgreSQL, MySQL, OpenOffice (ODB), entre otras.
- Hojas de cálculo Excel.
- Archivos XML (DATA CLEANER., 2012).
- ✓ **Herramienta de desarrollo**

Suite Pentaho 6.0: La Open BI Suite de Pentaho, provee un completo espectro de funcionalidades de Business Intelligence (BI, Inteligencia de Negocios), incluyendo reportes, análisis, tableros de control, minería de datos, integración de datos y una plataforma de BI que la han convertido en la suite de código abierto más popular del mundo (CORAIL, 2012).

Algunas de las funcionalidades integradas con la plataforma de Inteligencia de Negocios (BI) de Pentaho son:

- Pentaho Data Integration.
- Pentaho Schema Workbench.
- Pentaho BI Server.
- Pentaho Report Designer.

Pentaho Data Integration 6.0: es una de las herramientas utilizadas en el proceso de Extracción, Transformación y Carga (ETL) de código abierto que reúne un conjunto de componentes que permiten modelar y ejecutar transformaciones sobre flujos de datos.

Presenta las siguientes características:

Mercado de datos para el almacén de libros de la facultad CITEC.

- Puede funcionar sobre varias plataformas a través de un sistema que soporte Java 1.4 o una versión superior.
- Exige solamente alrededor de 128 MB de RAM.
- Provee un Lenguaje de programación Java para la Conectividad de Base de Datos (JDBC)₂ que permite la conexión con cualquier base de datos sin tener que instalar un cliente adicional.
- Se integra con ficheros de Microsoft Office, Servicios Web y cubos MOLAP.
- Esta herramienta incluye procesamiento optimizado de los ficheros planos.
- Brinda soporte para metadatos e incorpora operaciones de transformación.
- Permite operar con los campos en el flujo de datos, renombrando, calculando campos en función de otros y realizando búsquedas auxiliares en bases de datos.
- Corrección de errores (Gravitar, 2012).

Pentaho Schema Workbench 3.10.0.1: es un entorno visual para el desarrollo y prueba de cubos OLAP Mondrian. Provee un mecanismo para buscar datos con rapidez y tiempo de respuesta uniforme independientemente de la cantidad de datos en el cubo o la complejidad del procedimiento de búsqueda. Permite la ejecución de consultas de Expresiones Multidimensionales (MDX) contra el esquema y la base de datos y la navegación por la base de datos subyacente (Gravitar, 2012).

Pentaho BI Server 6.0: La aplicación más conocida de la Plataforma de BI es el servidor de BI de Pentaho, que funciona como una red basada en sistema de gestión de informes, el servidor de integración de aplicación es un motor de flujo de trabajo ligero (secuencias de acción.) Está diseñado para integrarse fácilmente en cualquier proceso de negocio (EASY, 2010).

Pentaho Report Designer 6.0: es una herramienta de diseño de informes basados en múltiples fuentes de origen. Dichos informes pueden ser consultados a través del servidor de Pentaho o, de ser necesario, distribuidos por correo a múltiples usuarios.

Algunas de sus características son:

- Generales: previsualizador, editor gráfico, creación basada en bandas, drag & drop.
- Elementos que se pueden integrar en los informes: Chart, List, Table, Dynamic CrossTabs, Text, Dynamic Text, Image, Label, subreports, barcodes/sparklines, drill-back, crosstabs y elementos flash.
- Formato de salida de gráficos: PNG, JPG, SVG, EPS, PDF.
- Propiedades de los informes: Paginación, Templates, subreports, javascript, scripting, hyperlinks...

Mercado de datos para el almacén de libros de la facultad CITEC.

- Tipos de informes: estáticos, paramétricos, ad-hoc.
 - Formato de salida: PDF, HTML, EXCEL y RTF.
 - Distribución: bursting, email, web service.
 - Fuentes de datos: JDBC, ODBC, Reflection, Hibernate, Kettle, Mondrian, OLAP4J, Pentaho Metadata, Scripted data.
 - Existencia de un API de programación.
 - Funcionalidades avanzadas de creación de mensajes de correo electrónico.
 - Integración de un wizard para la creación de informes.
- ✓ **Servidor de Aplicaciones**

Apache Tomcat 5.5: herramienta utilizada como servidor web.

Algunas de sus características son:

- Es el servidor web más utilizado para trabajar en entornos web.
- Es una implementación completamente funcional de los estándares de JSP y servlets.
- Puede especificarse como el manejador de las peticiones de JSP y servlets recibidas por servidores Web populares.
- Está integrado en la implementación de referencia Java 2 Enterprise Edition (J2EE) de Sun Microsystems (Cristian, 2006).

1.6. Conclusiones del capítulo

Una vez concluido el estudio de las diferentes tecnologías para dar solución al problema planteado, se puede aseverar que los almacenes de datos constituyen un elemento primordial en la gestión de grandes volúmenes de información, auxiliándose en los mercados de datos para informaciones clasificadas específicamente. Se abordaron aspectos relacionados con esta tecnología y su relación con los subsistemas de almacenamiento, integración y visualización, así mismo se opta por esta tecnología ya que constituye una solución viable en cuanto al almacenamiento y estructura de los libros en el almacén de la facultad CITEC. Para la implementación de la solución se concluye que: La Propuesta de Metodología para el Desarrollo de Almacenes de Datos permite guiar el proceso de desarrollo de los subsistemas de almacenamiento, integración y visualización del MD por las diferentes etapas del ciclo de vida. Las herramientas que ayudarán a construir el subsistema de almacenamiento son: Visual Paradigm en su versión 8.0 que posibilita la generación de los diagramas necesarios para modelar el funcionamiento del sistema propuesto. Así como el SGBD PostgreSQL 9.4.1 y el administrador de base de datos PgAdmin III 1.16.1 para los procesos de integración, se decide

Mercado de datos para el almacén de libros de la facultad CITEC.

utilizar el DataCleaner 3.1 para el perfilado de los datos, el Pentaho Data Integration 6.0 para extraer, transformar y cargar los datos y el Report Designer 6.0 para los reportes operacionales.

Mercado de datos para el almacén de libros de la facultad CITEC.

Capítulo 2: Análisis y diseño del Mercado de Datos.

En el presente capítulo se realiza una descripción detallada del negocio y los temas de análisis. Se definen las reglas del negocio, los requisitos de información, los funcionales, no funcionales y los casos de uso del sistema, Se identificarán los hechos, medidas y dimensiones quedando conformado el modelo de datos de la solución y la matriz bus. Además, serán definidos y diseñados los Subsistemas de Almacenamiento, Integración y Visualización de los datos.

2.1. Análisis

2.1.1. Descripción del negocio

La facultad CITEC cuenta con un almacén de libros, en el cual la encargada del mismo conduce el proceso de almacenamiento de información de estos en su mayoría de forma manual, es por eso que la gestión de los procesos como: préstamo de libros, despacho, entrega y recepción de materiales o reportes administrativos resultan lentos y en ocasiones ineficientes si se tiene en cuenta la posible ocurrencia de errores de naturaleza humana. Al estar el proceso semi-informatizado, existe un grupo de informaciones con las que se puede hacer más efectivo el control de los libros como puede ser la cantidad de reintegros de libros por fecha, cantidad de préstamos de libros por fecha, almacén y persona, obtener la cantidad de salidas de libros por fecha y almacén, la cantidad de libros en manos de estudiantes y profesores, obtener el saldo total de los libros por fecha y almacén, entre otros; que para su obtención se necesitaría de un enorme esfuerzo del personal que labora en el almacén pues habría que consultar cada unidad y repetir una y otra vez los procedimientos.

El Vicedecanato de Administración de la facultad CITEC es el departamento encargado de atender el almacén de libros a nivel de facultad por lo que también necesita este tipo de información para realizar controles sistemáticos, obtener estadísticas de consumo y para realizar estos análisis existen en formato duro un gran cúmulo de datos que es necesario buscar en distintos documentos, los cuales deben almacenarse durante cierto tiempo para realizar estudios y tomar decisiones a partir del análisis de la evolución de los indicadores mencionados anteriormente, lo que atenta contra la veracidad de la información. Estos datos se encuentran almacenados en un sistema informático que facilita la gestión de los recursos del almacén de libros para apoyar cada uno de los procesos del mismo.

2.1.2. Tema de análisis

Un tema de análisis no es más que la división o categorización de la información de una organización según las diferentes temáticas que contenga. Se elige un tema de análisis basándose en los objetivos que se persigan. La identificación de los temas de análisis es de suma importancia para el desarrollo del almacén de datos, estos permiten la factibilidad, utilidad, y éxito de las estructuras que se están diseñando.

Se define como tema de análisis:

Mercado de datos para el almacén de libros de la facultad CITEC.

- La información de los libros almacenados en el Sistema de Gestión para el almacén de libros de la facultad CITEC.

2.2. Necesidades de usuarios

Después del análisis del negocio se define como necesidad de los usuarios el análisis de la información relacionada con el almacén de libros de la facultad CITEC, que se recopila a través de la base de datos del Sistema de Gestión para el almacén de libros de la facultad CITEC.

La situación actual muestra que no es posible obtener por parte de los directivos del vicedecanato de administración ni de la encargada del almacén de libros la información sobre aspectos asociados al uso de los mismo. Partiendo de esto se describen los requisitos de información, funcionales y no funcionales del sistema.

2.3. Requisitos de información

Los requisitos de información (RI) describen la información y los datos que el sistema debe proveer o debe acceder. Estos se definen a partir de las necesidades de información identificadas en el negocio, que permitan el análisis del comportamiento de los indicadores a medir según los objetivos y metas de la organización.

A continuación, se muestran los requisitos de información definidos:

- **RI 1:** Obtener la cantidad de reintegros de libros por fecha, encargado del almacén y persona.
- **RI 2:** Obtener la cantidad de reintegros de libros por fecha, encargado del almacén y libro.
- **RI 3:** Obtener por ciento de reintegros por fecha y encargado del almacén.
- **RI 4:** Obtener la cantidad de entradas de libros por fecha y libro.
- **RI 5:** Obtener porcentaje de entradas de libros por fecha y libro.
- **RI 6:** Obtener la cantidad de salidas de libros por fecha y libro.
- **RI 7:** Obtener por ciento de salidas de libros por fecha y libro.
- **RI 8:** Obtener el saldo total de los libros por fecha y libro.
- **RI 9:** Obtener por ciento del saldo total de los libros por fecha y libro.
- **RI 10:** Obtener cantidad de préstamos de libros por fecha, encargado del almacén y persona.
- **RI 11:** Obtener la cantidad de préstamos de libros por área y categoría.
- **RI 12:** Obtener por ciento de préstamos de libros por fecha, encargado del almacén y persona.

2.4. Requisitos funcionales

Los requisitos funcionales de un sistema describen el comportamiento del sistema. Estos requisitos dependen del tipo de software que se comience a desarrollar, lo que esperan los usuarios del software y el enfoque general tomado por la organización al escribir los requisitos (Sommerville, 2011).

Mercado de datos para el almacén de libros de la facultad CITEC.

Tabla 1: Descripción de los requisitos funcionales del sistema.

No	Requisitos Funcionales	Descripción
1	Extraer los datos de la fuente	Permite extraer los datos de la fuente.
2	Realizar transformación y carga	Realiza los procesos de transformación y carga de los datos extraídos.

2.5. Requisitos no funcionales

Pueden estar relacionados con las propiedades del sistema como la fiabilidad, tiempo de respuesta y la capacidad de almacenamiento. Alternativamente, pueden definir restricciones en la implementación del sistema, como las capacidades de los dispositivos de Entrada/Salida (E/S) o de las representaciones de datos utilizadas en las interfaces con otros sistemas. Los requisitos no funcionales, como el rendimiento, la seguridad y la disponibilidad, suelen especificar o restringir características del sistema en su conjunto (Sommerville, 2011).

Con el propósito de satisfacer al máximo las necesidades del cliente, así como la calidad de la herramienta se definió un listado de las propiedades o cualidades que el producto debe tener, las cuales se describen a continuación:

✓ **Disponibilidad**

RNF 1: asegurar la disponibilidad del sistema.

- El sistema debe estar disponible durante el horario de trabajo, de lunes a viernes, de 8:00am a 5:00pm.

RNF 2: asegurar la recuperación ante un fallo.

- El sistema debe ser capaz de recuperarse ante un fallo, teniendo en cuenta la complejidad y naturaleza de este.

RNF 3: garantizar la persistencia de la información.

- Se debe realizar un respaldo total de los datos del almacén de datos al concluir cada semestre. Esta información se almacenará en el vicedecanato de administración y será responsabilidad del vicedecano de administración de la facultad CITEC. Este requisito también contribuye a la seguridad del sistema.

Mercado de datos para el almacén de libros de la facultad CITEC.

✓ **Restricciones de diseño**

RNF 4: la estructura del mercado de datos debe tener un nombre estándar teniendo en cuenta el tipo de estructura que sea: dim_[nombre] para las dimensiones y hech_[nombre] para los hechos.

RNF 5: utilizar el Sistema Gestor de Base de Datos definido durante la investigación.

- El gestor de base de datos que se utilizará es PostgreSQL 9.4.1 y como interfaz de administración de dicho gestor PgAdmin III 1.16.1.

RNF 6: utilizar las herramientas para la implementación de la capa de inteligencia de negocios definidas durante la investigación.

De la suite Pentaho, se usarán los siguientes componentes:

- MV Java 7.51
- Schema Workbench 3.10.0.1
- Pentaho BI Server 6.0
- Pentaho Data Integration 6.0
- Data Cleaner 3.1
- Pentaho Report Designer 6.0

✓ **Usabilidad:**

RNF 7: el sistema debe contar con un diseño del modelo físico sencillo, con una estructura y distribución que permita trabajar con rapidez y eficiencia.

✓ **Fiabilidad:**

RNF 8: el acceso a la información debe estar disponible el tiempo especificado y se tendrán en cuenta los permisos establecidos.

✓ **Seguridad:**

RNF 9: los permisos correspondientes al usuario autenticado se activarán una vez que éste se autentique y en caso de cambiar, tendrá acceso solo a la información que le compete de acuerdo con sus privilegios.

✓ **Acceso al sistema:**

RNF 10: el usuario deberá acceder a la aplicación mediante el protocolo HTTP, usando preferiblemente el navegador web Firefox 52.0 en adelante.

RNF 11: proporcionar características mínimas de hardware a las estaciones de trabajo.

Características de un cliente ligero.

- 1 GB RAM.

Mercado de datos para el almacén de libros de la facultad CITEC.

- 60 GB de disco duro.
- Intel Pentium 4 de procesador.

RNF 12: proporcionar características mínimas de hardware a los servidores.

Para lograr una explotación aceptable del sistema los servidores deben contar con los siguientes requerimientos de hardware:

- 8 GB RAM.
- 320 GB de disco duro.
- Intel Core i5 4690k de procesador.

✓ Interfaces de usuario

RNF 13: Mostrar todos los textos que aparezcan en la interfaz del sistema en idioma español. Los mensajes de procesamiento y error, así como los nombres de las vistas y la arquitectura de información deben ser en idioma español y tener una estándar en todo el sistema.

✓ Interfaces de software

RNF 14: Instalar en las estaciones de trabajo el software necesario para el correcto funcionamiento del sistema.

Las configuraciones de software de las máquinas clientes deben contar al menos con:

- Firefox 52.0 o superior.

Las configuraciones de software del servidor deben contar al menos con:

- PostgreSQL 9.4.1, como servidor de base datos.
- PgAdmin III 1.18.1, como administrador de base datos
- Pentaho Data Integration 6.0, donde se utiliza el PAN para ejecutar las transformaciones y el Kitchen para ejecutar los trabajos.
- Pentaho BI Server 6.0, como servidor de aplicaciones para la visualización de las vistas de análisis y los reportes operacionales.
- MV Java 7.51.

2.6. Reglas del negocio

Las reglas del negocio describen las políticas, normas, operaciones, definiciones y restricciones presentes en una organización y son de vital importancia para alcanzar los objetivos propuestos. Las reglas del negocio deben ser expresadas en lenguaje natural y orientadas al negocio, a continuación, se mencionan las reglas del negocio asociadas al almacén de datos operacional para el almacén de libros de la facultad CITEC.

Mercado de datos para el almacén de libros de la facultad CITEC.

Reglas de variables

- **RN-1:** no pueden existir campos nulos.
- **RN-2:** los identificadores deben ser únicos.
- **RN-3:** el control de toda la información de los libros es realizado por la encargada del almacén y el vicedecano de administración de la facultad.
- **RN-4:** $\text{cant_prestamos} = \text{sumatoria de todos los préstamos realizados por el almacén.}$
- **RN-5:** $\text{cant_entradas} = \text{sumatoria de todos los libros que entran nuevos al almacén.}$
- **RN-6:** $\text{cant_salidas} = \text{sumatoria de todos los libros que salen del almacén.}$
- **RN-7:** $\text{saldo} = \text{cant_entradas} - \text{cant_salidas.}$
- **RN-8:** $\text{cant_reintegros} = \text{sumatoria de todos los reintegros realizados al almacén.}$

Reglas de transformación

- **RN-9:** cuando el nombre del autor no aparece se le asigna el valor: Autor no Registrado.
- **RN-10:** cuando la fecha de devolución aparece nula se le asigna la fecha: 9999/12/31.
- **RN-11:** cuando la credencial aparece null se le asigna el valor: Desconocido.
- **RN-12:** cuando la persona aparece null se le asigna el valor: Persona desconocida.

2.7. Modelo de casos de uso del sistema

El modelo de casos de uso del sistema es una forma de representar gráficamente la relación existente entre los actores y casos de uso del sistema (CUS), de manera que sirva de guía para reflejar las metas y funcionalidades que persigue el negocio (Pressman, 2005).

Los CUS fueron identificados a partir del agrupamiento de los requisitos de información y los requisitos funcionales, para un total de 6 CUS y dos actores que se relacionan con ellos, según sus responsabilidades.

Casos de uso de información:

- Mostrar información sobre los préstamos (Obtener la cantidad de préstamos de libros por fecha, almacén y persona; obtener la cantidad de préstamos de los libros por año académico; obtener el por ciento de préstamos de libros por fecha y almacén).

Mercado de datos para el almacén de libros de la facultad CITEC.

- Mostrar información sobre los registros (Obtener cantidad de reintegros de los libros por fecha, encargado y persona; obtener la cantidad de reintegros de los libros por fecha y organismo; obtener el por ciento de reintegros por fecha, encargado del almacén y organismo).
- Mostrar información sobre el almacén (Obtener la cantidad de entradas de libros por fecha y almacén; obtener el por ciento de entradas de libros por fecha y almacén; obtener la cantidad de salidas de libros por fecha y almacén; obtener el por ciento de salidas de libros por fecha y almacén; obtener el saldo total de los libros por fecha y almacén; obtener el por ciento del saldo total de los libros por fecha y almacén).

Casos de uso funcionales:

- Extraer datos de la fuente.
- Realizar transformación y carga de los datos.
- Realizar opciones sobre reportes

A continuación, se presenta la descripción de las responsabilidades de cada uno de los actores en el sistema (Tabla 2) y el diagrama de CUS (Figura 6), donde para la confección del mismo se utilizó el patrón concordancia en adición ya que existen tres casos de usos relacionados, dos bases y la concordancia.

Descripción de los actores del sistema

Tabla 2: Descripción de los actores del sistema.

Actor	Objetivo
Administrador ETL	Responsable de llevar a cabo los procesos de extracción, transformación y carga de los datos.
Especialista	Responsable de inicializar los casos de uso relacionados con la visualización de los reportes del mercado de datos y realiza las opciones sobre los reportes.

Diagrama de CUS

Mercado de datos para el almacén de libros de la facultad CITEC.

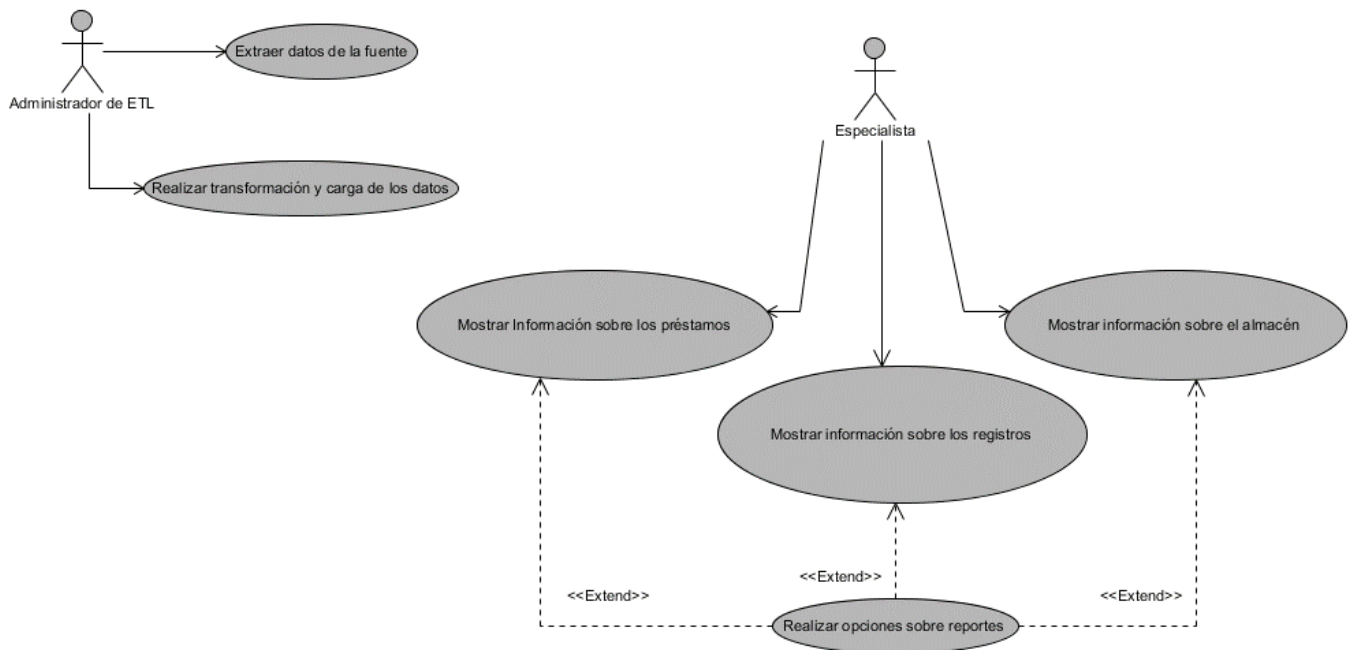


Figura 5: Diagrama de casos de uso del sistema.

Especificación de casos de uso

CU Mostrar información sobre el almacén.

Tabla 3: CU Mostrar información sobre el almacén.

Objetivo	Mostrar información sobre el almacén .
Actores	Especialista: (Inicia) Mostrar información sobre el almacén.
Resumen	El caso de uso inicia cuando el actor desea hacer un análisis de la información relacionada con el almacén. El actor selecciona el reporte que desea ver, el sistema muestra la información contenida en él. El caso de uso finaliza cuando el actor termina el análisis de la información relacionada con el almacén.
Complejidad	Media.
Prioridad	Alta.
Precondiciones	El especialista tiene que estar autenticado. Los datos correspondientes fueron cargados en el mercado de datos. Los reportes relacionados con la información de la población fueron creados

Mercado de datos para el almacén de libros de la facultad CITEC.

Postcondiciones	Los reportes correspondientes al caso de uso fueron consultados.		
Flujo de eventos			
Flujo básico Mostrar información sobre tendencias de las NC de revisiones.			
	Actor	Sistema	
1	Selecciona el área de información del almacén.		
2		Muestra las áreas de análisis correspondientes.	
3	Selecciona la información que desea consultar.		
4		Muestra los reportes contenidos en él.	
5	Selecciona el reporte que desea analizar.		
6		Muestra la información contenida en el reporte seleccionado y brinda opciones al actor para visualizar los reportes durante su análisis. Ir al CU Visualizar reportes. Finaliza el CU.	
Flujos alternos			
2a Los datos son incorrectos			
	Actor	Sistema	
		Muestra un mensaje de error. Vuelve al paso 1 del flujo básico.	
Opciones de Reportes			
Perspectivas de análisis		Posibles resultados	
		Medidas	Periodicidad
Variables de entrada relacionadas con el CU: <ul style="list-style-type: none"> • tiempo. • tipo de revisión. • almacén. 		Variables de salida disponibles: <ul style="list-style-type: none"> • existencia de los libros en el almacén. 	Rango de tiempo en que se solicitan las variables de salida: <ul style="list-style-type: none"> • Semestral.
Relaciones	CU incluidos	No aplica.	

Mercado de datos para el almacén de libros de la facultad CITEC.

	CU extendidos	No aplica.
Requisitos no funcionales	Sección: "2.5 Requisitos no funcionales" del documento: "Especificación de requisitos de software".	
Asuntos pendientes		

2.9. Diseño de los subsistemas de almacenamiento, integración y visualización

El proceso de diseño de un almacén define el modelo conceptual, lógico, físico y de visualización (BI) del mismo. En el modelo conceptual se especifican los requerimientos del usuario a través de las dimensiones, hechos y medidas que componen el almacén; en el modelo lógico se genera el modelo de datos; en el modelo físico se diseñan las transformaciones, y en el de visualización se diseñan reportes candidatos por cada uno de los libros de trabajo definidos en el área de análisis.

2.9.1. Subsistema de Almacenamiento

El subsistema de almacenamiento contiene toda la información correspondiente al MD, es por ello que, luego del estudio realizado sobre la información del almacén y la identificación de los RI a través de las necesidades de los clientes, fueron creadas 3 tablas de hechos, 5 medidas y 4 tablas de dimensiones.

2.9.2 Arquitectura del Almacén de Datos

La arquitectura del almacén de datos quedó estructurada de manera tal que se compone por la fuente de datos (base de datos del Sistema de gestión para el almacén de libros de la facultad CITEC) y tres subsistemas bases, los cuales son:

- Subsistema de integración: encargado de la extracción e integración de la información para su posterior carga al almacén de datos.
- Subsistema de almacenamiento: encargado de almacenar toda la información del almacén de datos en las diferentes tablas de hechos y dimensiones definidas.
- Subsistema de visualización: encargado de consultar los datos guardados en el almacén, con el objetivo de mostrarlos a los usuarios finales en los distintos reportes, contribuyendo a la toma de decisiones.

Mercado de datos para el almacén de libros de la facultad CITEC.

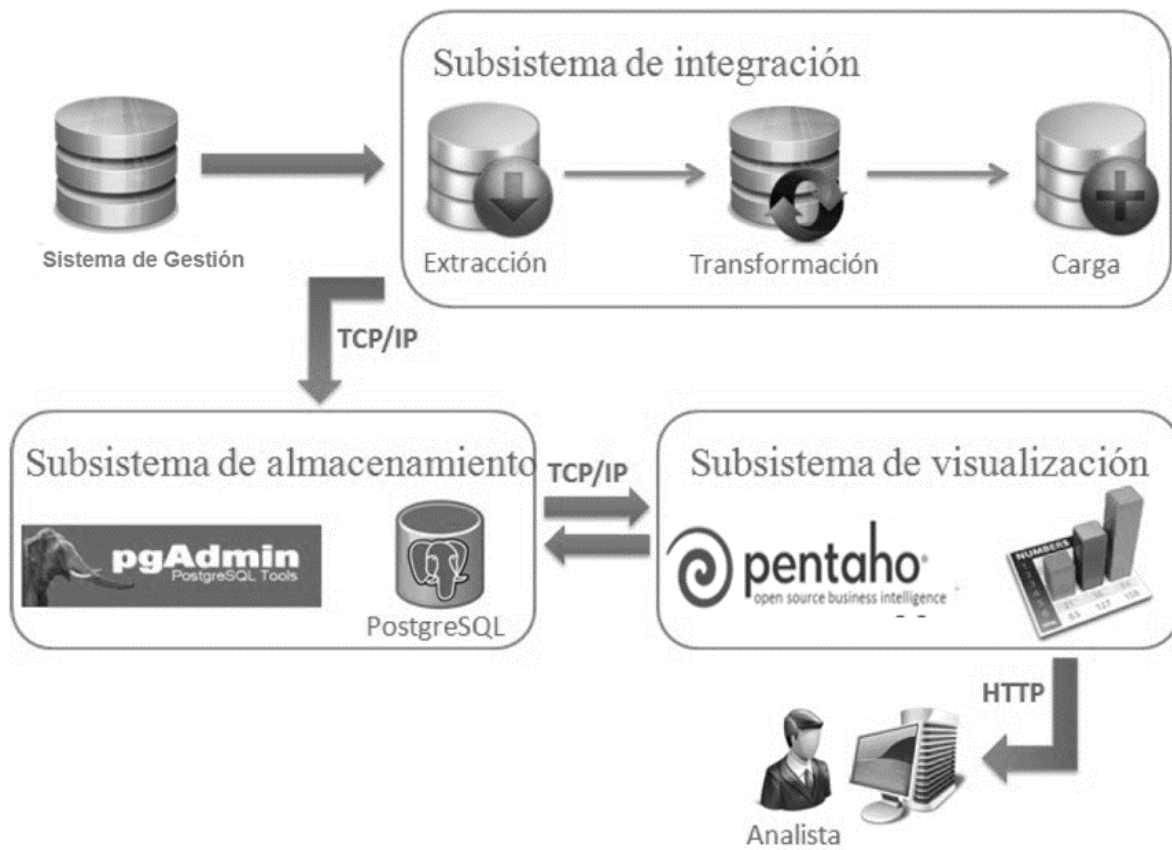


Figura 6: Arquitectura del almacén de datos.

A continuación, se presenta la descripción de las tablas de dimensiones (Tabla 4) y hechos (Tabla 5).

Tabla 4: Descripción de las tablas de dimensiones.

No.	Dimensiones	Descripción
D1	dim_fecha	Define una línea de tiempo específica donde el nivel más bajo es el día.
D2	dim_libro	Representa el autor y el título de los libros.
D3	dim_persona	Representa los datos de una persona.
D4	dim_almacen	Representa los datos del almacén de libros.

Tabla 5: Descripción de las tablas de hechos.

NO.	Hechos	Descripción
H1	hech_prestamo	Contiene la medida correspondiente a la cantidad de préstamos.

Mercado de datos para el almacén de libros de la facultad CITEC.

H2	hech_almacen	Contiene la medida correspondiente a la cantidad de entradas y salidas que se realizan en el almacén.
H3	hech_reintegro	Contiene la medida correspondiente a la cantidad de reintegros.

Matriz bus

La Matriz Bus es la representación de la relación que existe entre los hechos y las dimensiones. Se define como la habilidad para describir y seguir la vida tanto de una dimensión como de un hecho, la cual permite determinar el impacto que provocaría un cambio durante el desarrollo del sistema.

Tabla 6: Matriz Bus.

Matriz Bus					
Hechos	Dimensiones				
	D1	D2	D3	D4	D5
H1	XX	X	X	X	
H2	X	X			
H3	X	X	X	X	X

Modelo de datos

El modelo de datos representa la relación entre las tablas de hechos y dimensiones identificadas mediante el proceso de análisis. De esta manera los datos del negocio quedan reflejados en forma de cubos de datos. El modelo propuesto para la solución del MD para el almacén de libros de la facultad CITEC se basa en una topología Constelación de Hechos, ya que existen dimensiones compartidas para más de una tabla de hechos. A continuación, se muestra el modelo diseñado para la solución (Figura 8).

Mercado de datos para el almacén de libros de la facultad CITEC.

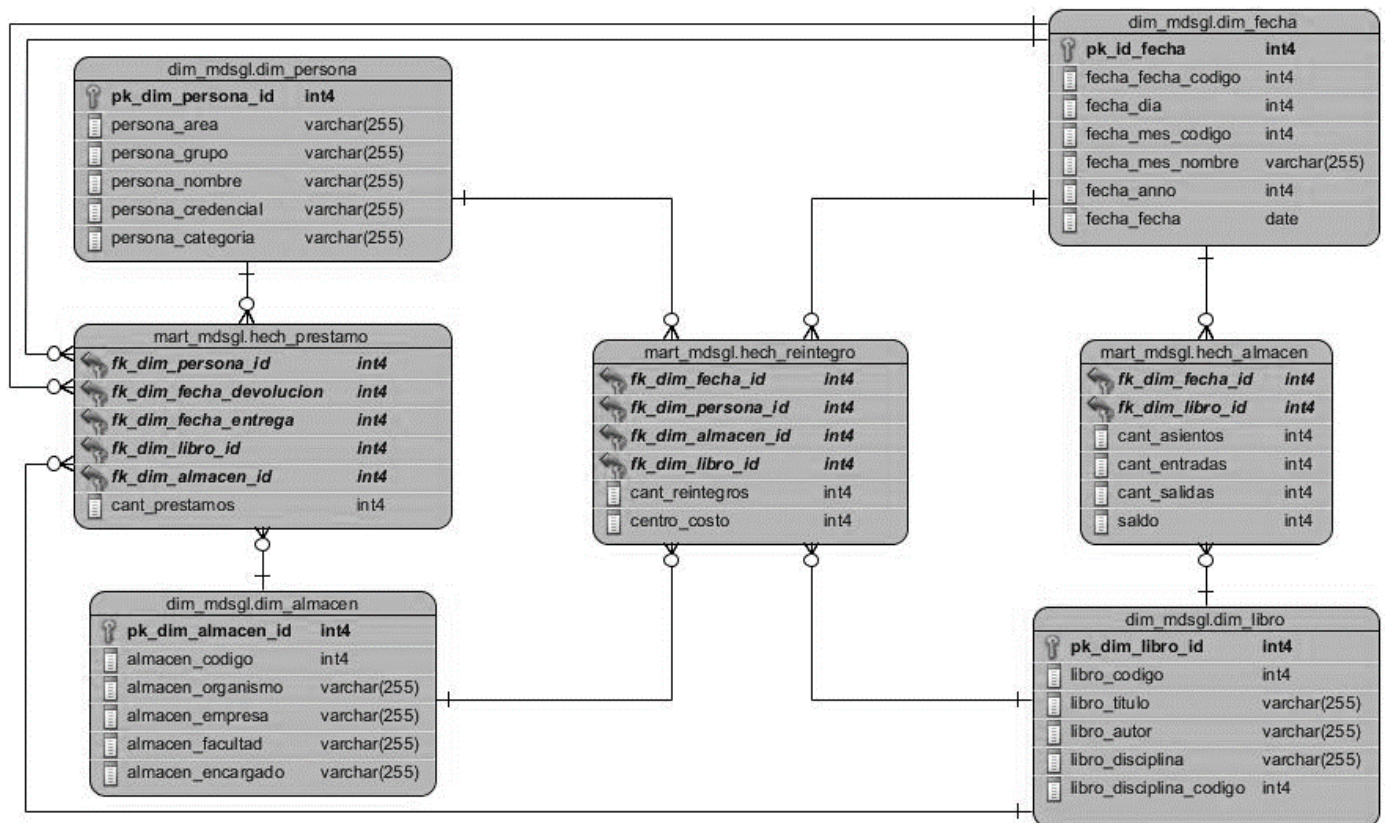


Figura 7: Modelo de Datos.

2.9.3. Subsistema de Integración

El subsistema de integración comprende el perfilado de los datos y la extracción de los mismos desde los sistemas fuentes, los cuales sufren un conjunto de transformaciones. Conformando estos de manera que fuentes separadas puedan ser aprovechadas conjuntamente, y finalmente hace su entrega en un formato listo para el almacenamiento, de forma tal, que permita a los desarrolladores construir la capa de presentación de la aplicación y los especialistas puedan tomar decisiones o realizarle técnicas de minería de datos a los que están almacenados. El perfilado, el diccionario de datos, el diseño de las transformaciones, el registro de sistemas fuentes y el mapa lógico de datos constituyen elementos esenciales para lograr el diseño del subsistema de integración.

Perfilado de datos

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

Mercado de datos para el almacén de libros de la facultad CITEC.

Al realizar el perfilado de los datos a la base de datos del sistema de gestión se identificó que los tipos de datos son en su mayoría enteros y cadenas, fueron detectados 52% de datos de tipo textos (String), 32% de números enteros (Integer) y 16% de tipo fecha (Date), así como se evidencia en la Figura 9. También fueron analizadas las tablas de la fuente para detectar la cantidad de filas nulas en cada una de ellas, por lo que no fueron almacenados valores vacíos ni nulos.

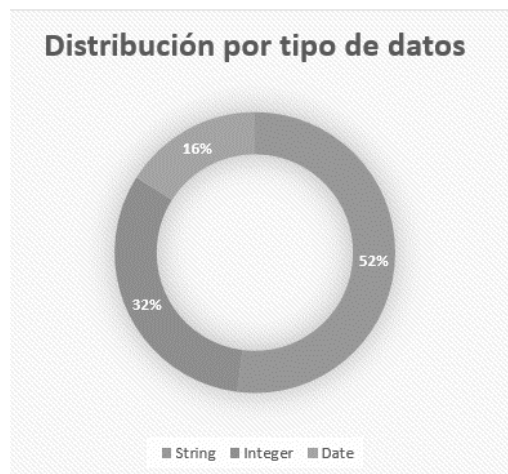


Figura 8: Distribución por tipo de datos.

Diseño de las transformaciones

Una vez que se conoce la estructura, contenido y fiabilidad de los datos, se procede a realizar el diseño de las transformaciones. Estos diseños pueden variar a la hora de la implementación de las mismas, porque durante el proceso de desarrollo suelen surgir situaciones con los datos y se llevan a cabo diversas estrategias para resolverlas.

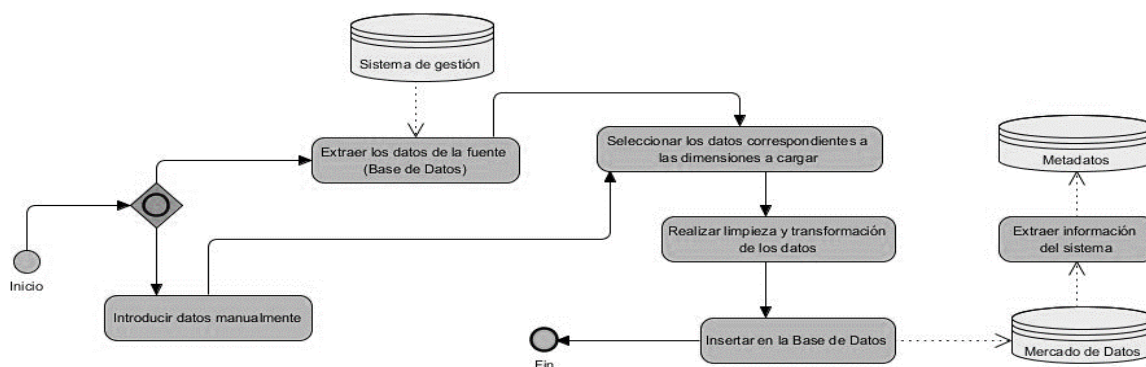


Figura 9: Diseño de las transformaciones para la carga de dimensiones.

Mercado de datos para el almacén de libros de la facultad CITEC.

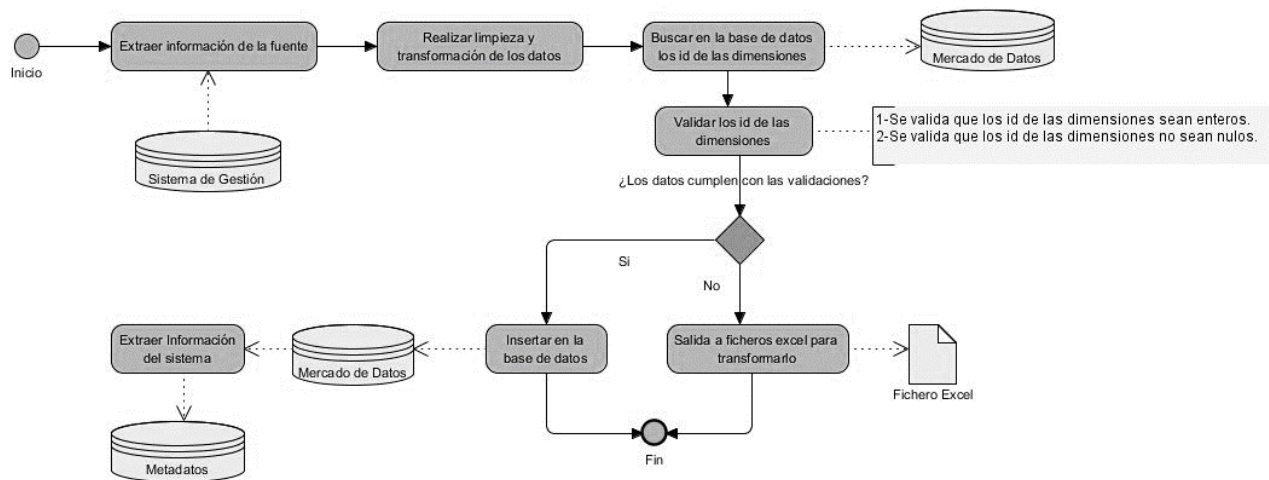


Figura 10: Diseño de las transformaciones para la carga de hechos.

2.9.4. Subsistema de Visualización

El diseño del subsistema de visualización comprende la realización de los cubos OLAP, además de los distintos reportes que contribuyan a la toma de decisiones de los usuarios finales del MD. En estos cubos se definen las dimensiones, medidas y las distintas jerarquías que posean las dimensiones para su posterior publicación en el Pentaho BI Server, el cual proporciona el servidor y plataforma web al usuario final. Este podrá interactuar con la solución BI previamente creada.

Arquitectura de información

Para el Mercado de datos para el almacén de libros de la facultad CITEC se definió como Área de Análisis General (A.A.G.) Almacén de Libros. En la investigación actual se identificaron 3 áreas de análisis que se divide en 3 libros de trabajo los que incluyen las tablas de salidas definidas por el cliente.

- **Área de análisis (AA):** Son las secciones en las que se divide el AD, las que contienen los diferentes libros de trabajo y las vistas de análisis.
- **Libro de Trabajo (LT):** Son las diferentes categorías en las que se pueden agrupar cada una de las vistas de análisis.
- **Vistas de análisis:** Informe que expresa mediante tablas los valores de los indicadores analizados.

A continuación, se muestra la estructura de navegación definida para el MD:

Mercado de datos para el almacén de libros de la facultad CITEC.

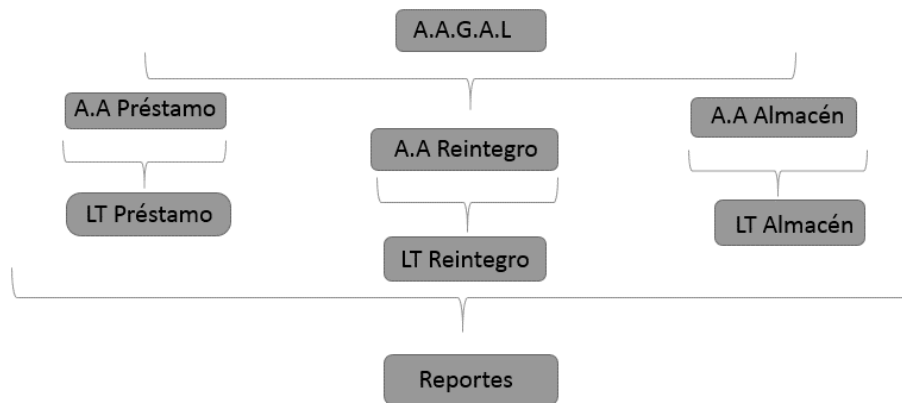


Figura 11: Diseño del mapa de navegación.

Diseño de los reportes candidatos

Los reportes candidatos responden a las necesidades de los usuarios y se corresponden con las salidas que debe mostrar el sistema. A continuación, se muestran ejemplo de reportes candidatos de cada una de las áreas de análisis definidas:

Tabla 7: Reporte candidato "Cantidad de préstamos " del A.A Préstamo.

Área de análisis (AA)	Préstamo.
Libro de Trabajo (LT)	Préstamo.
Reporte (Tabla de Salida – TS)	Cantidad de préstamos.
Descripción	Representa la cantidad de préstamos de los libros.
Elementos del reporte	Persona, fecha de entrega, fecha de devolución, almacén, libro.
Frecuencia de emisión	Semestral.
Funciones	Suma.
Gráfico	Línea.

Tabla 8: Reporte candidato "Cantidad de reintegros" del A.A Reintegro.

Área de análisis (AA)	Reintegro.
Libro de Trabajo (LT)	Reintegro.
Reporte (Tabla de Salida – TS)	Cantidad de reintegros.
Descripción	Representa la cantidad de reintegro de los libros.
Elementos del reporte	Fecha, persona, libro, almacén.
Frecuencia de emisión	Semestral.
Funciones	Suma.

Mercado de datos para el almacén de libros de la facultad CITEC.

Gráfico	Línea.
---------	--------

Tabla 9: Reporte candidato "Cantidad de saldo" del A.A Almacén.

Área de análisis (AA)	Almacén.
Libro de Trabajo (LT)	Saldo.
Reporte (Tabla de Salida – TS)	Cantidad de Saldo.
Descripción	Representa la cantidad de libros que quedan en el almacén.
Elementos del reporte	Asientos, entradas, salidas.
Frecuencia de emisión	Semestral
Funciones	Resta
Gráfico	Línea

2.10. Políticas de respaldo y recuperación

Las políticas de respaldo y recuperación de datos garantizan la integridad de la información, de manera que se encuentre almacenada en un lugar seguro en caso de fallos en el sistema. Para el MD en desarrollo se realizarán salvadas a la información contenida en la base de datos diariamente a través de tareas programadas en el sistema operativo y al finalizar cada semestre, ya que ese es el período en que se generan más préstamos y devoluciones en el almacén, estas salvadas se guardarán durante seis años y el responsable de las mismas es el vicedecano de administración de la facultad. En cuanto a las tablas involucradas, se les realizarán las salvadas a las tres tablas de hechos identificadas y sus dimensiones asociadas.

Roles y permisos

Es de gran importancia para un sistema de información contar con un mecanismo de protección contra aquellas acciones que puedan afectar la integridad de los datos almacenados. Por tal motivo, para el acceso a los Subsistemas de almacenamiento e integración es necesario definir los roles que tendrán acceso a la base de datos.

Tabla 10: Roles y permisos.

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

Mercado de datos para el almacén de libros de la facultad CITEC.

2.11. Conclusiones del capítulo

Una vez abordados los principales elementos relacionados con los artefactos generados durante la etapa de análisis y diseño de los subsistemas de almacenamiento, integración y visualización, se puede concluir que: Se realizó un estudio de las necesidades de información del cliente que permitió la identificación de 12 requisitos de información, 14 no funcionales, 2 funcionales y 12 reglas del negocio, las cuales han sido empleadas durante el diseño de los subsistemas. Mediante el diseño del modelo de datos fueron identificadas 4 tablas de dimensiones y 3 tablas de hechos, que garantizan el correcto funcionamiento del sistema. El diseño de las transformaciones para la carga de las dimensiones y los hechos constituye una aproximación a los pasos que se deben realizar para lograr la estandarización de la información y su almacenamiento.

Mercado de datos para el almacén de libros de la facultad CITEC.

Capítulo 3: Implementación y prueba del Mercado de Datos

En este capítulo se realizará la implementación de los Subsistemas de Almacenamiento, Integración y Visualización. Este proceso se efectúa sobre las tablas del modelo de datos definido en el capítulo anterior. La implementación del subsistema de almacenamiento comprende la definición de los estándares de codificación y la construcción del modelo físico. Por su parte, en el subsistema de integración de datos se implementan las transformaciones y los trabajos, así como los metadatos. Se exponen las pruebas realizadas a los subsistemas y los resultados obtenidos en cada una de ellas. Dichas pruebas son realizadas para garantizar el cumplimiento de las exigencias del cliente y la calidad del producto.

3.1. Implementación del Subsistema de Almacenamiento

Una vez diseñado el modelo dimensional siguiendo una estandarización de los nombres, se dio lugar al modelo físico, permitiendo describir el almacenamiento de los datos y la relación entre las tablas. Además, fueron creados los esquemas, así como las tablas correspondientes a cada uno de ellos.

Esquemas

- dim_mdsgl: contiene las 4 dimensiones identificadas para el MD.
- mart_mdsgl: contiene las 3 tablas de hechos identificadas para el MD.
- metadatos_mdsgl: contiene una tabla con la información de las actualizaciones para el MD.

3.1.1. Estándar de codificación

Los estándares de codificación pretenden organizar la forma en que se denominan las estructuras con el objetivo de lograr un patrón que tribute a la correcta normalización de los términos utilizados. Esta codificación está más bien dirigida a los desarrolladores, para que exista un vocabulario común en todo el almacén de datos, que permita un entendimiento claro y conciso.

En la solución propuesta se mantiene la misma nomenclatura atendiendo a la clasificación de las estructuras, teniendo en cuenta si la misma es una dimensión, un hecho, una transformación, un metadato o un trabajo. Si la tabla es una dimensión, el nombre estaría compuesto por las letras “dim” separadas del nombre de la misma por el carácter “_”, ejemplo dim_fecha. En caso de ser una tabla de hecho, se le antepone las letras “hech” e igualmente se separa del nombre del hecho por el carácter “_”, ejemplo hech_prestamo.

En el caso de los atributos de las dimensiones se siguió la misma estrategia para cada una de ellas. Las llaves primarias de las dimensiones fueron denominadas de la forma “pk_dim_dimension_id”. Las medidas contienen las letras “cant”, el carácter “_” y luego se especifica lo que se va a contar, ejemplo

Mercado de datos para el almacén de libros de la facultad CITEC.

cant_entradas. Por su parte los metadatos están conformados por la palabra “metadatos” y el nombre del mismo seguido del caracter “_”, ejemplo metadatos_mdsgl.

Luego de finalizar el proceso de estandarización de los nombres, queda organizada la nomenclatura utilizada para la denominación de las tablas, atributos y medidas dentro de la base de datos, así como de las transformaciones y trabajos.

3.2. Implementación del Subsistema de Integración de datos

La implementación del subsistema de integración implica que se lleven a cabo los procesos de ETL. Para el desarrollo del presente trabajo se utiliza como estrategia de integración de datos: Extracción, Transformación y Carga. Se decide la utilización de esta estrategia pues da soporte a distintos orígenes de datos desde base de datos relacionales y no relaciones hasta archivos XML, archivos en formatos Excel y DBF, entre otro gran conjunto de ficheros.

El proceso de ETL brinda la posibilidad de enviar datos a otros *softwares* en tiempo real, tiene una amplia capacidad de transformación, desde una sencilla conversión de un tipo de dato, cálculos simples, hasta transformaciones complejas como agregaciones y sumalizaciones. Posee una amplia visualización de flujos, ya que se puede observar gráficamente cual es el recorrido del flujo de información desde que los mismos son leídos de los sistemas fuentes hasta que son escritos en las distintas tablas de hechos y dimensiones y posee funcionalidades avanzadas que permiten limpiar los datos de origen evitando cargar información errónea en el almacén de datos.

Según (Kimball 2004), ETL consiste en:

- **Extracción:** La primera parte del proceso de ETL consiste en extraer los datos desde los sistemas de origen. La extracción convierte los datos a un formato preparado para iniciar el proceso de transformación.
- **Transformación:** Cualquier operación realizada sobre los datos para que puedan ser cargados en el almacén de datos o se puedan migrar de éste a otra base de datos. La fase de transformación aplica una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados. Algunas fuentes de datos requerirán alguna pequeña manipulación de los datos.
- **Carga:** Consiste en almacenar los datos en la base de datos final. La fase de carga es el momento en el cual los datos de la fase anterior (transformación) son cargados en el sistema de destino.

Ralph Kimball define 34 sistemas ETL clasificados en cuatro grupos, donde para llevar a cabo el proceso de integración de datos del presente trabajo fueron utilizados los siguientes subsistemas (Díaz 2011):

Mercado de datos para el almacén de libros de la facultad CITEC.

Extracción

- Sistema de extracción (subsistema 3): permite la extracción de los datos desde la fuente origen hasta la fuente destino.

Limpieza y conformación

- Limpieza de datos (subsistema 4): implementa los procesos de calidad de los datos que permiten detectar las incoherencias de calidad.
- Rastreo de eventos de errores (subsistema 5): captura todos los errores que proporcionan información valiosa sobre la calidad de los datos y permite la mejora de los mismos.

Entrega

- Claves subrogadas (subsistema 10): permite crear claves subrogadas independientes para cada tabla.

Gestión

- Programador de trabajos (subsistema 22): permite gestionar ETL de la categoría de trabajos.
- Repositorio de metadatos (subsistema 34): captura los metadatos de los procesos de ETL, de los datos del negocio y de los aspectos técnicos.

3.2.1. Transformaciones y trabajos

El elemento básico de diseño de los procesos ETL es la transformación, la cual está compuesta por pasos enlazados entre sí a través de los saltos. Los pasos constituyen el elemento más pequeño dentro de las transformaciones, y a través de los saltos fluye la información entre los diferentes pasos. A continuación, se muestra un ejemplo de transformación para el hecho reintegro y la dimensión libro:

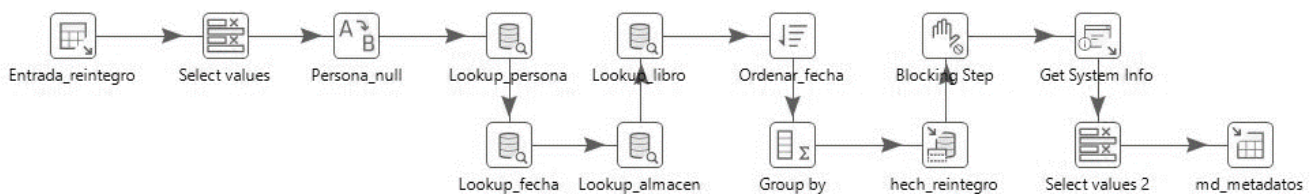


Figura 12: Proceso de integración de datos para el hecho reintegro.

Mercado de datos para el almacén de libros de la facultad CITEC.

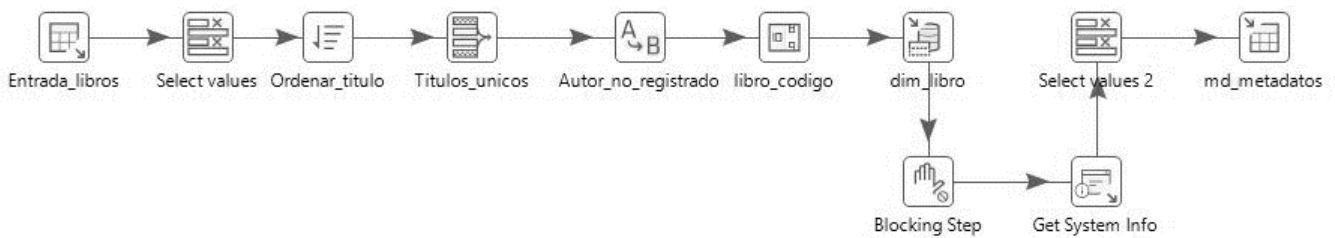


Figura 13: Proceso de integración de datos para la dimensión libro.

Un trabajo o job es similar a un proceso: conjunto de tareas con el objetivo de realizar una acción determinada. Estos permiten ejecutar varias transformaciones o trabajos previamente diseñados y organizar una secuencia de ejecución de estos. Los trabajos se encuentran en un nivel superior de las transformaciones. Para el MD para el almacén de libros de la facultad CITEC tres trabajos, uno para los hechos (Figura 15), otro para las dimensiones (Figura 16) y uno general para la ejecución de ambos trabajos.

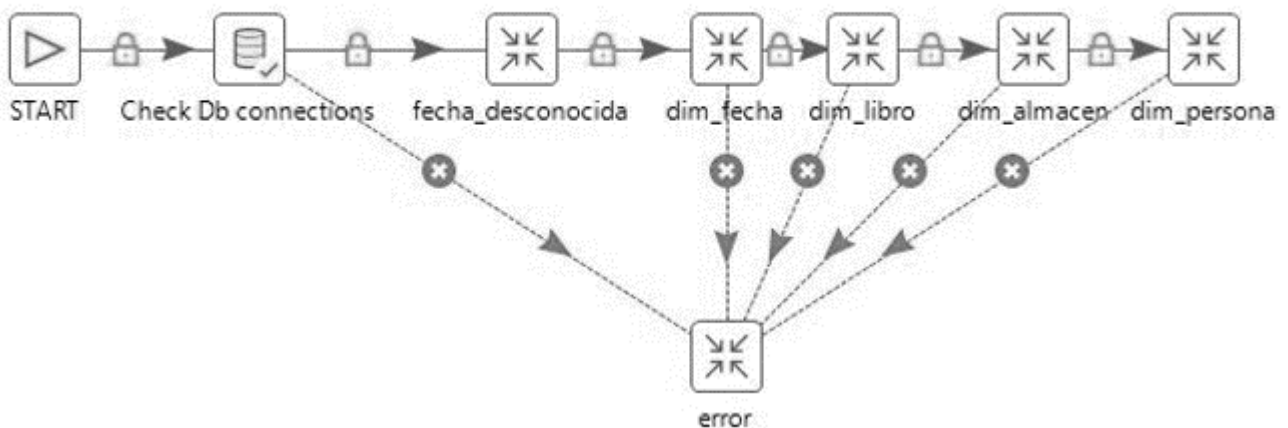


Figura 14: Trabajo para las dimensiones.

Mercado de datos para el almacén de libros de la facultad CITEC.

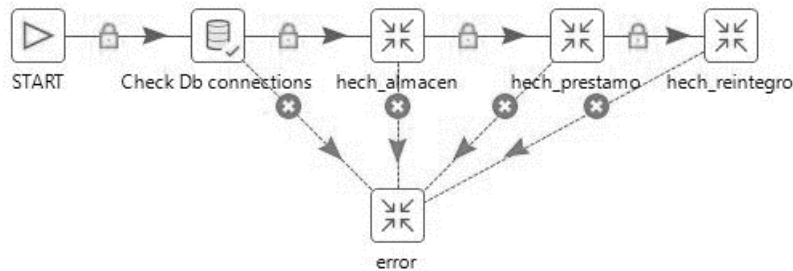


Figura 15: Trabajo para los hechos.

3.3. Implementación del Subsistema de Visualización

Luego de realizada la carga de los datos se procede a la implementación de los cubos OLAP y la creación de los reportes candidatos, acciones correspondientes a la implementación del subsistema de visualización.

Implementación de los cubos OLAP

En la implementación de los cubos OLAP se definieron todas las dimensiones, medidas y jerarquías que conforman el esquema Almacén. También se creó un cubo por cada una de las tablas de hechos (Figura 16).

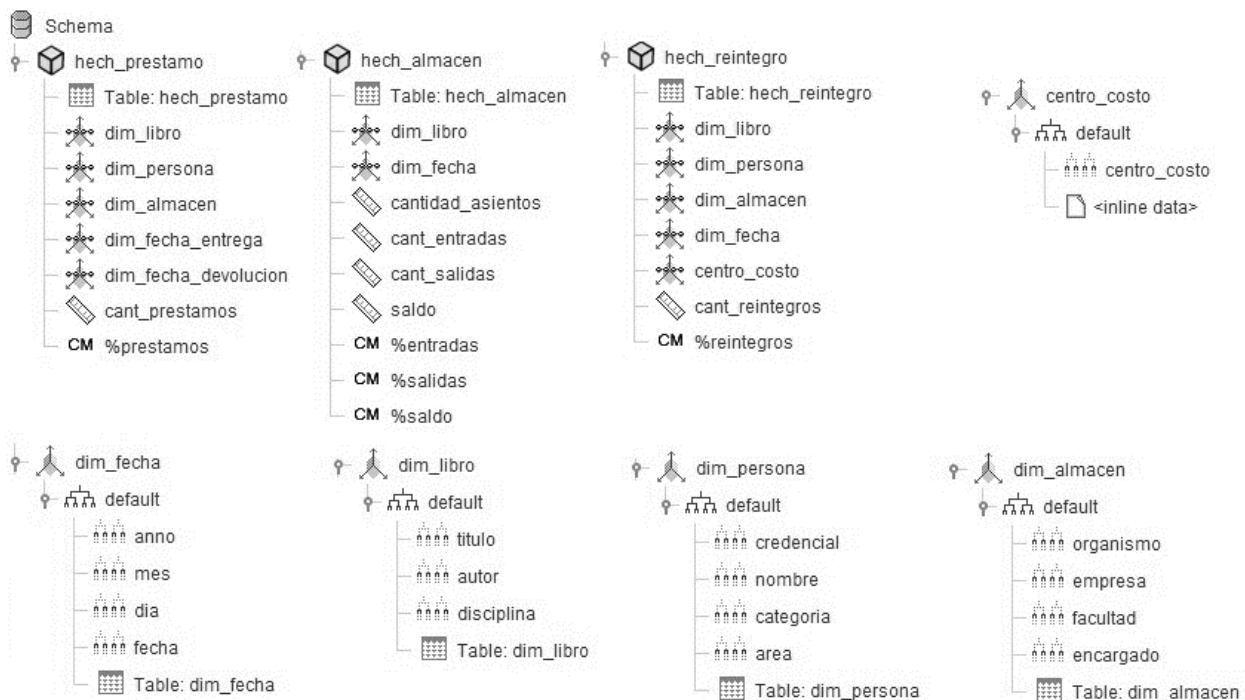


Figura 16: Implementación de los cubos OLAP.

Mercado de datos para el almacén de libros de la facultad CITEC.

3.3.1. Implementación de la capa de visualización.

El mapa de navegación o arquitectura de información está diseñado de manera que permita tener una mejor visualización de cómo estará estructurada la información en el MD para el almacén de libros de la facultad CITEC, la cual cuenta con un Área de Análisis General (A.A.G), tres Áreas de Análisis (A.A) y tres Libros de Trabajo (L.T) con sus reportes asociados (Figura 17).

- **A.A.G. MDSGL:** agrupa la información referente a todos los libros que existen en el almacén.
- **A.A. Prestamo:** contiene la información referente a todos los préstamos que se realizan en el almacén.
- **A.A. Reintegro:** contiene la información referente a los reintegros de los libros que se le hacen al almacén.
- **A.A. Almacen:** contiene toda la información referente a las entradas, salidas y existencia de los libros que hay dentro del almacén.
- **L.T. Prestamo:** contiene la información relacionada a la cantidad de préstamos que se realizan en el almacén.
- **L.T. Reintegro:** contiene la información relacionada a la cantidad de reintegros que se le hacen al almacén.
- **L.T. Almacen:** contiene la información relacionada a la cantidad de entradas, salidas y existencia de los libros dentro del almacén.

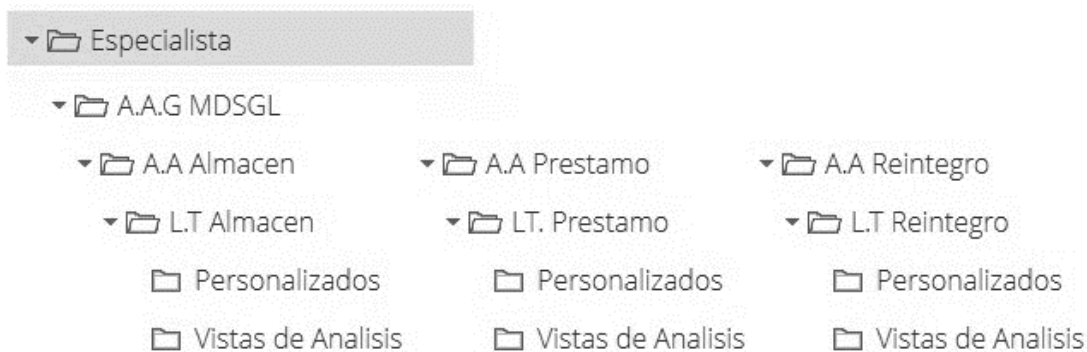


Figura 17: Arquitectura de información del MD para el almacén de la facultad CITEC.

3.3.2 Implementación de los reportes candidatos

Los reportes candidatos necesarios para satisfacer las necesidades del cliente fueron implementados en la herramienta Pentaho BI Server, mediante consultas MDX, las cuales realizan las consultas utilizando los términos hechos, dimensiones, medidas y varían su complejidad en dependencia de los tipos de reportes.

Mercado de datos para el almacén de libros de la facultad CITEC.

18 de May de 2017.
3:18:25 PM

Almacén de libros

Cantidad de Elementos: 27

Reporte Reintegros

Fecha	Encargado del Almacén	Reintegros	Código Libro	Persona
10-04-2017	Juanita Pérez Pérez	1	20	Asley Arias Sainz
10-04-2017	Juanita Pérez Pérez	1	20	Mailin Trueba Pérez
10-04-2017	Juanita Pérez Pérez	1	20	María Teresa Sardañas Rodríguez
10-04-2017	Juanita Pérez Pérez	1	20	Carlos Leandro Milán González
10-04-2017	Juanita Pérez Pérez	1	20	Marlon Ruiz Plaza
10-04-2017	Juanita Pérez Pérez	1	20	Chabelys Marlen Lestapier Lao
10-04-2017	Juanita Pérez Pérez	1	20	Claudia Sánchez López
10-04-2017	Juanita Pérez Pérez	1	20	Enmanuel Adalberto Naranjo Nápoles
10-04-2017	Juanita Pérez Pérez	1	20	Gisell Carreño Cesé
10-04-2017	Juanita Pérez Pérez	1	20	Idel González Sierra
10-04-2017	Juanita Pérez Pérez	1	20	Jorge Alfonso Torres
10-04-2017	Juanita Pérez Pérez	1	20	José Rubén Machado Vega
10-04-2017	Juanita Pérez Pérez	1	20	Alejandro De León Abreu
10-04-2017	Juanita Pérez Pérez	1	20	Alejandro Miguel Medina Pichs
10-04-2017	Juanita Pérez Pérez	1	20	Ariam Lázaro Alvarez Rodríguez
10-04-2017	Juanita Pérez Pérez	1	20	Ariel Enrique Parra López

Figura 18: Reporte reintegro

Mercado de datos para el almacén de libros de la facultad CITEC.

Fecha	Almacen	Persona	Medidas
+ 2017	Yanet Pérez Pérez	+ Aideé Mora Urdaneta	• Cantidad de reintegros 1
		+ Adiel Alfonso Cordovi	1
		+ Adriana González Ortega	1
	Juan Pérez Pérez	+ Adrián Avilés Avila	2
		+ Ailed Nuñez Ramos	2
	Julio Pérez Pérez	+ Adrian Arencibia Herrera	1
		+ Adrián Céspedes Calzado	1
		+ Abel Machado Montes	1
	Juanita Pérez Pérez	+ Mailin Trueba Pérez	1
		+ Enmanuel Adalberto Naranjo Nápoles	1
		+ Alejandro Miguel Medina Pichs	1
		+ Gisell Carreño Cesé	1
		+ Lizardsandra Díaz Grecesqui	1
		+ Luis Miguel Puris Hernández	1
		+ Asiel Riverón Hernández	1
+ Idel González Sierra		1	

Figura 19: Cantidad por fecha, encargado del almacén y persona.

3.4 Pruebas

Las pruebas constituyen un elemento crítico para la calidad del software. De acuerdo a la IEEE una prueba se define como: *Actividad en la cual un sistema o componente es ejecutado bajo condiciones específicas, se observan o almacenan los resultados y se realiza una evaluación de algún aspecto del sistema o componente.*

Una prueba se considera exitosa si encuentra alguna deficiencia en el *software*. Para obtener diferentes tipos de errores en el sistema se hace necesario aplicar un amplio conjunto de pruebas al sistema, con la finalidad de medir el grado en que el producto cumple con los requerimientos definidos. El proceso de pruebas comienza con la planificación de las mismas, seguidamente la ejecución, control y como paso final la evaluación. Para llevar a cabo las pruebas fue utilizado el Modelo V, método empleado por el centro DATEC para garantizar el buen funcionamiento y la calidad de los productos de almacenes de datos.

El modelo V es un método para comprobar el buen funcionamiento de los almacenes de datos, proviene del principio que establece que los procedimientos utilizados para probar si la aplicación cumple las especificaciones del cliente, ya deben haberse creado en la fase de diseño. Es una representación de dos cascadas enfrentadas y relacionadas, con su vértice en la codificación como punto en común. La cascada

Mercado de datos para el almacén de libros de la facultad CITEC.

a la izquierda muestra las actividades relacionadas con el análisis y diseño del MD y la cascada derecha muestra las actividades relacionadas con el aseguramiento de la calidad, haciendo uso de las herramientas Casos de Prueba y Listas de Chequeo. La siguiente figura muestra la relación entre los componentes del modelo.

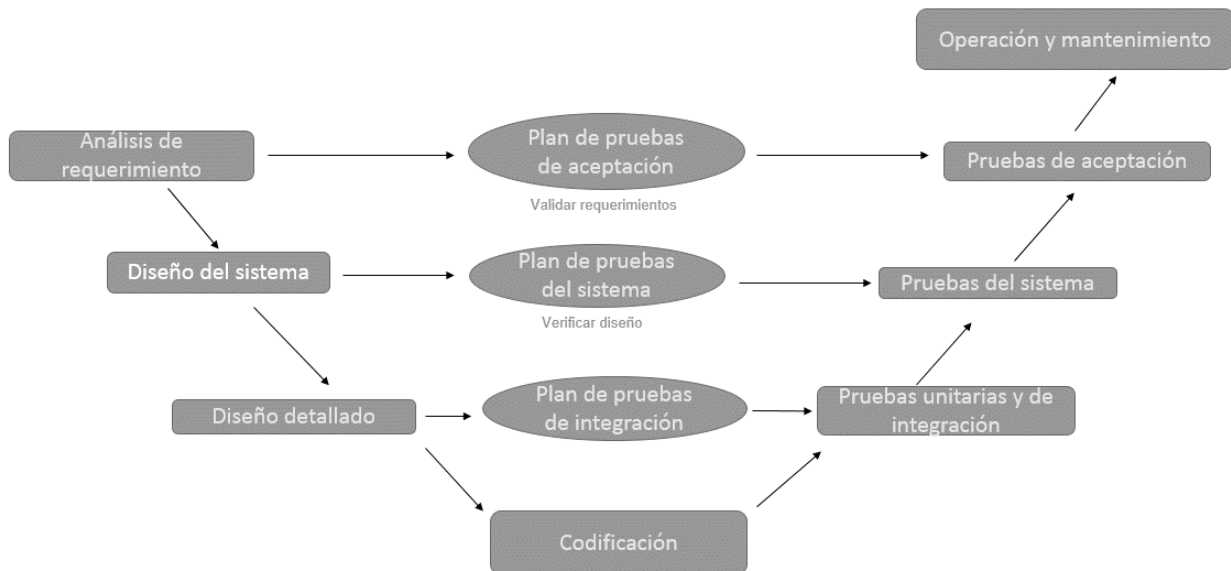


Figura 20: Modelo V.

Según (Lantares, 2003), estas pruebas consisten en:

Pruebas unitarias

La prueba de unidad es la primera fase de las pruebas dinámicas y se realizan sobre cada módulo del software de manera independiente.

Objetivo: comprobar que el módulo, entendido como una unidad funcional de un programa independiente, está correctamente codificado. En estas pruebas cada módulo será probado por separado y lo hará, generalmente, la persona que lo creó. En general, un módulo se entiende como un componente software que cumple las siguientes características:

- Debe ser un bloque básico de construcción de programas.
- Debe implementar una función independiente simple.
- Podrá ser probado al cien por cien por separado.
- No deberá tener más de 500 líneas de código.

Pruebas de integración

Aun cuando los módulos de un programa funcionen bien por separado es necesario probarlos conjuntamente: un módulo puede tener un efecto adverso o inadvertido sobre otro módulo; las

Mercado de datos para el almacén de libros de la facultad CITEC.

subfunciones, cuando se combinan, pueden no producir la función principal deseada; la imprecisión aceptada individualmente puede crecer hasta niveles inaceptables al combinar los módulos; los datos pueden perderse o malinterpretarse entre interfaces, etc. Por lo tanto, es necesario probar el software ensamblando todos los módulos probados previamente. Este es el objetivo de las pruebas de integración. A menudo hay una tendencia a intentar una integración no incremental; es decir, a combinar todos los módulos y probar todo el programa en su conjunto. El resultado puede ser un poco caótico con un gran conjunto de fallos y la consiguiente dificultad para identificar el módulo (o módulos) que los provocó. En contra, se puede aplicar la integración incremental en la que el programa se prueba en pequeñas porciones en las que los fallos son más fáciles de detectar. Un tipo de integración incremental es la denominada ascendente. Veamos los pasos a seguir para la misma:

Integración incremental ascendente:

- Se combinan los módulos de bajo nivel en grupos que realicen una subfunción específica.
- Se escribe un controlador (un programa de control de la prueba) para coordinar la entrada y salida de los casos de prueba.
- Se prueba el grup.
- Se eliminan los controladores y se combinan los grupos moviéndose hacia arriba por la estructura del programa.

Pruebas del sistema

Este tipo de pruebas tiene como propósito ejercitar profundamente el sistema para verificar que se han integrado adecuadamente todos los elementos del sistema (hardware, otro software, etc.) y que realizan las funciones adecuadas. Concretamente se debe comprobar que:

- Se cumplen los requisitos funcionales establecidos.
- El funcionamiento y rendimiento de las interfaces hardware, software y de usuario.
- La adecuación de la documentación de usuario.
- Rendimiento y respuesta en condiciones límite y de sobrecarga.

Para la generación de casos de prueba del sistema se utilizan técnicas de caja negra. Este tipo de pruebas se suelen hacer inicialmente en el entorno del desarrollador, denominadas Pruebas Alfa, y seguidamente en el entorno del cliente denominadas Pruebas Beta.

Pruebas de aceptación

A la hora de realizar estas pruebas, el producto está listo para implantarse en el entorno del cliente. El usuario debe ser el que realice las pruebas, ayudado por personas del equipo de pruebas, siendo deseable, que sea el mismo usuario quien aporte los casos de prueba. Estas pruebas se caracterizan por:

Mercado de datos para el almacén de libros de la facultad CITEC.

- Participación activa del usuario, que debe ejecutar los casos de prueba ayudado por miembros del equipo de pruebas.
- Están enfocadas a probar los requisitos de usuario, o mejor dicho a demostrar que no se cumplen los requisitos, los criterios de aceptación o el contrato. Si no se consigue demostrar esto el cliente deberá aceptar el producto.
- Corresponden a la fase final del proceso de desarrollo de software.

Es muy recomendable que las pruebas de aceptación se realicen en el entorno en que se va a explotar el sistema (incluido el personal que lo maneje). En caso de un producto de interés general, se realizan pruebas con varios usuarios que reportarán sus valoraciones sobre el producto. Para la generación de casos de prueba de aceptación se utilizan técnicas de caja negra haciendo uso del método de partición equivalente.

3.4.1. Herramientas de prueba

Para la aplicación de los tipos de pruebas definidos anteriormente se utilizarán las herramientas siguientes:

Casos de prueba

Tabla 11: Diseño de CP para el CU "Mostrar información sobre el almacén"

Escenario	Descripción	Variables de entrada	Variables de salida	Respuesta del sistema	Flujo central
EC 1: Muestra información sobre libros por fecha y almacén.	Muestra la cantidad de libros que hay en el almacén según la fecha y el almacén.	<ul style="list-style-type: none"> • Fecha • Almacén 	Cantidad de libros.	El sistema muestra todas las variables disponibles para el análisis, ubicadas en las filas y las columnas que pueden ser visualizadas en el reporte.	El usuario se autentica, selecciona el A.A.G Almacén de libros / A.A Almacén / L.T Entrada /Se selecciona el reporte correspondiente.
EC 2: Muestra información sobre el saldo total de los libros por fecha y almacén.	Muestra la cantidad de saldo de libros que hay en el almacén según la fecha y el almacén.	<ul style="list-style-type: none"> • Fecha • Almacén 	Cantidad de saldo	El sistema muestra todas las variables disponibles para el análisis, ubicadas en las filas y las columnas que pueden ser visualizadas en el reporte.	El usuario se autentica, selecciona el A.A.G Almacén de libros / A.A Almacén / L.T Saldo /Se selecciona el reporte correspondiente

Mercado de datos para el almacén de libros de la facultad CITEC.

El propósito de los casos de prueba (CP) es la comprobación de la calidad que posee el software desarrollado, en ellos se identifican los fallos en la implementación de la solución, los resultados esperados y el cumplimiento de las especificaciones del sistema.

Estos son esenciales para todas las actividades de pruebas porque son la base para diseñar y ejecutar los procedimientos de pruebas. Reflejan trazabilidad con los CU, ya que estos muestran una secuencia ordenada de eventos, al describir flujos básicos, flujos alternos, precondiciones y postcondiciones. Si los casos de prueba no son correctos, la calidad del sistema se pone en duda y las pruebas dejan de ser confiables.

Listas de Chequeo

La lista de chequeo consta de una serie de preguntas, en forma de cuestionario, mediante el cual se verifica el grado de cumplimiento de determinadas reglas establecidas para los procesos de desarrollo del sistema, además de medir la calidad de los artefactos de los procesos de ETL generados durante la realización del producto. Esta evaluación se desarrolla a través del análisis de un grupo de indicadores, distribuidos en tres secciones fundamentales:

- **Estructura del documento:** abarca todos los aspectos definidos por el expediente de proyecto o el formato establecido por el proyecto.
- **Indicadores definidos:** abarca todos los indicadores a evaluar durante la etapa de desarrollo.
- **Semántica del documento:** contempla todos los indicadores a evaluar respecto a la ortografía, redacción y demás.

La estructura de la lista de chequeo está formada por los siguientes elementos:

- **Peso:** define si el indicador a evaluar es crítico o no. El mismo se describe con una C si es crítico.
- **Indicadores a evaluar:** constituyen los indicadores a evaluar en las secciones Estructura del documento, Semántica del documento e Indicadores definidos para el artefacto a evaluar.
- **Evaluación:** es la forma de evaluar el indicador en cuestión. El mismo se evalúa de uno en caso de que exista alguna dificultad sobre el indicador y de cero, en caso de que el indicador revisado no presente problemas.
- **N.P. (No Procede):** se usa para especificar que no es necesario evaluar el indicador en ese caso.
- **Cantidad de elementos afectados (CEA):** 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.

A continuación, se muestra la lista de chequeo aplicada en esta investigación:

Mercado de datos para el almacén de libros de la facultad CITEC.

Tabla 12: Lista de chequeo aplicada

Estructura del documento					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
C	1. ¿Los entregables contienen las secciones obligatorias de la plantilla estándar definidas para un expediente de proyecto? (portada, control de versiones, reglas de confidencialidad, tabla de contenidos y contenido) (ver expediente de proyecto)	0		0	
Indicadores definidos en el desarrollo					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
	1. ¿Se utilizó un lenguaje cuyas sentencias son expresables mediante una sintaxis bien definida?	0		0	
C	2. ¿Los reportes pueden configurarse a través de la interfaz del sistema?	0		0	
	3. ¿La interfaz está orientada a facilitar el uso de las funciones	0		0	

*Mercado de datos para el almacén de libros de la facultad
CITEC.*

	del sistema por parte de los usuarios?				
C	4. ¿No existen restricciones para construir cubos OLAP con dimensiones y niveles de agregación ilimitados?	0		0	
C	5. ¿Los usuarios son capaces de manipular los resultados de manera que se ajusten a sus necesidades, conformando nuevos reportes?	0		0	
	6. ¿El sistema responde de una forma rápida a la información solicitada por el usuario?	0		0	
	7. ¿El sistema refleja cualquier lógica del negocio para poder responder a preguntas específicas?	0		0	
C	8. ¿El sistema garantiza la confidencialidad y seguridad de acceso a los datos por roles de usuarios?	0		0	
	9. ¿Los datos e información derivados del proceso de análisis realizado mediante la aplicación, apoyan la toma de decisiones?	0		0	
C	10. ¿Los cambios en los datos se reflejan automáticamente en	0		0	

Mercado de datos para el almacén de libros de la facultad CITEC.

	los reportes de forma instantánea?				
Semántica del documento					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
C	1. ¿Se han identificado errores ortográficos en los entregables?	0		0	
C	2. ¿Se entiende claramente lo que se ha especificado en el documento?	0		0	
	3. ¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?	1		2	

Aplicación de la lista de chequeo

Durante la revisión del desarrollo del mercado de datos, a través de la aplicación de la lista de chequeo, se identificó un indicador con dificultades del cual se generó una NC. Esta fue tratada al corregir los dos elementos afectados, haciendo coincidir el número de página que aparecía en el índice con el contenido que se reflejaba en dicha página. De forma general la implementación fue evaluada de **Bien**, ya que no hubo ningún indicador crítico afectado, no existieron problemas con los formatos de las plantillas, no se encontraron errores ortográficos en los documentos revisados y fue solucionada la NC generada. La siguiente figura, representa el comportamiento de los indicadores en las diferentes secciones de la lista de chequeo, luego de evaluar la implementación del mercado de datos:

Mercado de datos para el almacén de libros de la facultad CITEC.

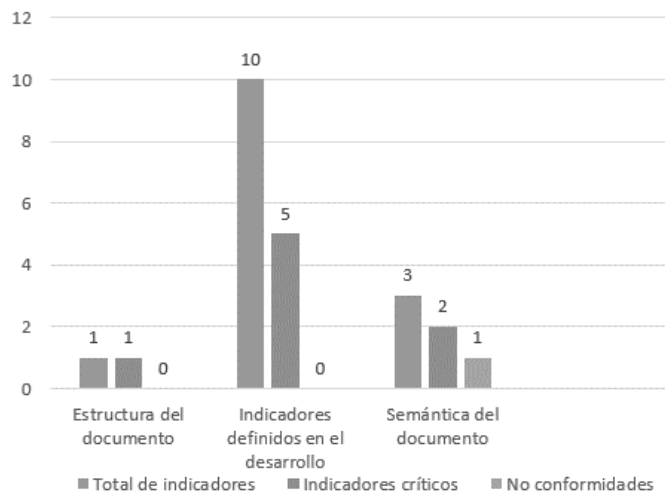


Figura 21: Comportamiento de los indicadores por secciones.

3.4.2. Resultados de las pruebas

Las pruebas realizadas al MD para el almacén de libros de la facultad CITEC arrojaron los siguientes resultados:

Pruebas unitarias y de integración

Una vez terminada la implementación se realizaron un conjunto de pruebas de unidad e integración a los distintos componentes de los subsistemas arrojando un total de 5 no conformidades (NC), las cuales se muestran a continuación:

- La estructura de carpetas del repositorio de ETL no está correcta.
- Los metadatos relacionados con la gestión del cambio en la fuente no han sido implementados.
- Debe arreglar el nombre de las medidas.
- Errores lógicos en la implementación de las transformaciones.
- Debe arreglar el nombre de los hechos.

Las mismas fueron resueltas satisfactoriamente en el tiempo establecido.

Pruebas del sistema

En las pruebas de sistema realizadas fueron identificadas un total de 42 NC, estas se basaban principalmente en errores relacionados con los nombres de las columnas de la fuente de datos y abuso del uso de mayúsculas. Una vez resueltas las NC detectadas se realizó una segunda iteración donde fue verificado el cumplimiento de la solución a dichas NC, arrojando como resultado que las mismas fueron resueltas satisfactoriamente.

Mercado de datos para el almacén de libros de la facultad CITEC.

Conclusiones del capítulo

En el presente capítulo se abordó sobre la implementación y las pruebas realizadas a los Subsistemas de almacenamiento, integración y visualización del MD para el almacén de libros de la facultad CITEC, las cuales concluyen el proceso de construcción y validación, arrojando las siguientes conclusiones: Se implementaron los tres subsistemas que componen la aplicación: almacenamiento, integración y visualización, teniendo como resultado la disponibilidad de la información. La estructura física del Subsistemas de almacenamiento del MD está conformada por tres esquemas: dim_mdsgl, mart_mdsgl y metadatos_mdsgl, posibilitando la correcta integración de los datos a la base de datos. Se realizaron las transformaciones para la carga de los hechos, para las dimensiones y para los trabajos, que posibilitaron la carga de los datos de la fuente. Con la aplicación de las listas de chequeo a los artefactos de ETL, los casos de prueba de integración basados en los casos de uso y los aplicados a las reglas de transformación, se logró probar que los Subsistemas de almacenamiento, integración y visualización cumplen con los requisitos identificados.

Mercado de datos para el almacén de libros de la facultad CITEC.

Conclusiones

El estudio de los distintos temas relacionados con el desarrollo de los almacenes de datos, proporcionó la elaboración del presente trabajo, el cual arrojó como resultado el “Mercado de datos para el almacén de libros de la facultad de Ciencias y Tecnologías Computacionales”. Los siguientes resultados demuestran el cumplimiento de los objetivos propuestos en la investigación:

- El estudio de las metodologías y herramientas realizado, garantizó que la metodología seleccionada guiara el proceso de desarrollo de los Subsistemas de almacenamiento, integración y visualización del MD para el almacén de libros de la CITEC a través de cada etapa del ciclo de vida y la selección de las herramientas y tecnologías para la construcción de la solución propuesta cumplen con la política de migración a software libre, dando soporte a las necesidades del equipo de desarrollo.
- En el análisis y diseño de los Subsistemas de almacenamiento e integración del MD se identificaron 12 requisitos de información, 2 funcionales, 14 no funcionales y 12 reglas del negocio, logrando cumplir con las necesidades del cliente.
- La implementación de los subsistemas de almacenamiento e integración posibilitaron la obtención de un mercado de datos correctamente poblado, con información disponible para ser consultada por parte de los usuarios, brindando apoyo al proceso de toma de decisiones.
- Las pruebas efectuadas durante las distintas etapas de desarrollo permitieron comprobar la funcionalidad del sistema a partir de los requisitos establecidos. Los resultados obtenidos durante las últimas pruebas realizadas fueron satisfactorios, validando el cumplimiento de los objetivos propuestos.

Mercado de datos para el almacén de libros de la facultad CITEC.

Recomendaciones

Para el presente trabajo de diploma se recomienda:

- Desplegar la solución y garantizar que la carga al mercado de datos se realice semestralmente para que se puedan realizar análisis certeros en el Vicedecanato de Administración.
- Integrar el análisis realizado en este mercado de datos, con la promoción para obtener una relación entre resultado docente y aprovechamiento de la bibliografía impresa por disciplina en la facultad CITEC.

Mercado de datos para el almacén de libros de la facultad CITEC.

Referencias

- Almenara. (2005). *Las nuevas Tecnologías de la Información y las Comunicaciones*.
- Bernabeu, D. R. (2010). *Hefesto: Metodología para la construcción de un Data Warehouse*.
- BibliotecaUCI.(s.f.)*. (s.f.). Obtenido de <http://biblioteca.uci.cu>.
- CAVSI. (2016). *¿Que es un Sistema Gestor de Bases de Datos o SGBD?* Obtenido de <http://www.cavsi.com/preguntasrespuestas/que-es-un-sistema-gestor-de-base-de-datos-o-sgbd/>.
- CORAIL, M. C. (2012). *Pentaho Open BI*. Obtenido de <http://www.cognus.cl/content/view/271452/Pentaho-Open-BI.html>.
- Cristian. (2006). *El servidor Apache Tomcat*. Colombia.
- DATALEANER. (2012). *DataCleaner news*. Obtenido de <http://datacleaner.eobjects.org/>
- EASY. (2010). *EASY*.
- González Hernández, Y. (2010.). *Metodología para el Desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocios*. Ciudad de La Habana.
- Gravitar. (2012). *Gravitar*.
- Ibarzabal, J. (2003). *Estrategia de reporting*.
- Imon. (2005). *Building the DataWarehouse*. Indianapolis.
- Imon, B. (2002). *Almacenes de Datos*.
- Kimbal. (2002). *The Data Warehouse Toolkit*. New York.
- Lantares, S. (2003). *"Data warehouse: la prueba de la calidad."*. Obtenido de Solutions Lantares: <http://www.lantares.com/blog/test-de-calidad-para-data-warehouse-y-bases-de-datos>.
- Orallo, J. H. (2010). *"Data Warehouse, o Almacén de Datos, y Cubo OLAP."*. Obtenido de El Mejor Proyecto: <http://elmejorproyecto2010.blogspot.com/2010/04/data-warehouse-o-almacen-de-datos-y.html>.
- POSTGRESQL-ES*. (s.f.). Obtenido de http://www.postgresql.org.es/sobre_postgresql.
- Pressman, R. S. (2005). *Ingeniería del Software. Un enfoque práctico*. Quinta Edición.
- Robelo, B. (2010). *Conceptos sobre Datawarehouse*. Obtenido de <http://www.bernardorobelo.com/conceptos-sobre-datawarehouse.html>.
- Sinnexus. (2011). *Business Intelligence*. Obtenido de http://sinnexus.es/business_intelligence/olap_vs_oltp.aspx.
- Sommerville. (2011). *Un Enfoque Práctico*.
- Tamayo, M. y. (2006). *Análisis del modelo de almacenamiento MOLAP frente al modelo de almacenamiento ROLAP*. Colombia: Ingeniería e Investigación, Vol. 26. ISBN: 0120-5609.

Mercado de datos para el almacén de libros de la facultad CITEC.

Technologies. (s.f.). *Visual Paradigm. Boost Productivity with Innovative and Intuitive Technologies*. Obtenido de <http://www.visualparadigm.com/product/vpuml/>.

TOMCAT, A. (s.f.). *Apache Tomcat*. Obtenido de <http://tomcat.apache.org/>.

Zepeda Sánchez, L. Z. (2008). *Metodología para el Diseño Conceptual de Almacenes de Datos*. Universidad Politécnica de Valencia. España.

Mercado de datos para el almacén de libros de la facultad CITEC.

Bibliografía

- Almenara. (2005). *Las nuevas Tecnologías de la Información y las Comunicaciones*.
- Bernabeu, D. R. (2010). *Hefesto: Metodología para la construcción de un Data Warehouse*.
- BibliotecaUCI.(s.f.). (s.f.). Obtenido de <http://biblioteca.uci.cu>.
- CAVSI. (2016). *¿Que es un Sistema Gestor de Bases de Datos o SGBD?* Obtenido de <http://www.cavsi.com/preguntasrespuestas/que-es-un-sistema-gestor-de-base-de-datos-o-sgbd/>.
- CORAIL, M. C. (2012). *Pentaho Open BI*. Obtenido de <http://www.cognus.cl/content/view/271452/Pentaho-Open-BI.html>.
- Cristian. (2006). *El servidor Apache Tomcat*. Colombia.
- DATALEANER. (2012). *DataCleaner news*. Obtenido de <http://datacleaner.eobjects.org/>
- EASY. (2010). *EASY*.
- González Hernández, Y. (2010.). *Metodología para el Desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocios*. Ciudad de La Habana.
- Gravitar. (2012). *Gravitar*.
- Ibarzabal, J. (2003). *Estrategia de reporting*.
- Imon. (2005). *Building the DataWarehouse*. Indianapolis.
- Imon, B. (2002). *Almacenes de Datos*.
- Kimbal. (2002). *The Data Warehouse Toolkit*. New York.
- Lantares, S. (2003). *"Data warehouse: la prueba de la calidad."*. Obtenido de Solutions Lantares: <http://www.lantares.com/blog/test-de-calidad-para-data-warehouse-y-bases-de-datos>.
- Orallo, J. H. (2010). *"Data Warehouse, o Almacén de Datos, y Cubo OLAP."*. Obtenido de El Mejor Proyecto: <http://elmejorproyecto2010.blogspot.com/2010/04/data-warehouse-o-almacen-de-datos-y.html>.
- POSTGRESQL-ES. (s.f.). Obtenido de http://www.postgresql.org.es/sobre_postgresql.
- Pressman, R. S. (2005). *Ingeniería del Software. Un enfoque práctico*. Quinta Edición.
- Robelo, B. (2010). *Conceptos sobre Datawarehouse*. Obtenido de <http://www.bernardorobelo.com/conceptos-sobre-datawarehouse.html>.
- Sinnexus. (2011). *Business Intelligence*. Obtenido de http://sinnexus.es/business_intelligence/olap_vs_oltp.aspx.
- Sommerville. (2011). *Un Enfoque Práctico*.
- Tamayo, M. y. (2006). *Análisis del modelo de almacenamiento MOLAP frente al modelo de almacenamiento ROLAP*. Colombia: Ingeniería e Investigación, Vol. 26. ISBN: 0120-5609.

Mercado de datos para el almacén de libros de la facultad CITEC.

Technologies. (s.f.). *Visual Paradigm. Boost Productivity with Innovative and Intuitive Technologies*. Obtenido de <http://www.visualparadigm.com/product/vpuml/>.

TOMCAT, A. (s.f.). *Apache Tomcat*. Obtenido de <http://tomcat.apache.org/>.

Zepeda Sánchez, L. Z. (2008). *Metodología para el Diseño Conceptual de Almacenes de Datos*. Universidad Politécnica de Valencia. España.