

Universidad de las Ciencias Informáticas
Facultad 6



Título: Sistema de Información de Gobierno. Mercado de datos
Educación.

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

Autora:

Aylenis Rodríguez Pompa

Tutores:

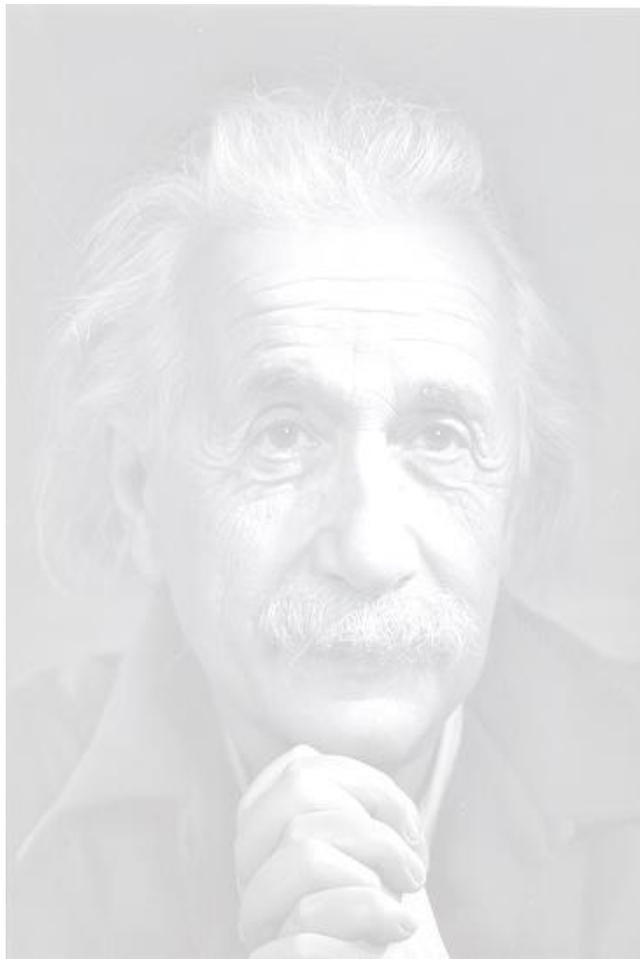
Ing. Rosayda Valiente Mesa

Ing. Lisdan Rodríguez Pérez

Ciudad de la Habana, junio de 2011
“Año 53 de la Revolución”

Los ideales que han iluminado mi camino, y una y otra vez me han infundido valor para enfrentarme a la vida con ánimo, han sido la bondad, la belleza y la verdad.

Albert Einstein



DECLARACIÓN DE AUTORÍA

Declaro ser autora del presente trabajo “Sistema de Información de Gobierno. Mercado de datos Educación” y reconozco a la Universidad de las Ciencias Informáticas (UCI) los derechos patrimoniales de la misma, con carácter exclusivo.

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

Autor: _____

Aylenis Rodríguez Pompa

Tutores: _____

Ing. Rosayda Valiente Mesa

Ing. Lisdan Rodríguez Pérez

DATOS DE CONTACTO

Ing. Rosayda Valiente Mesa

Universidad de las Ciencias Informáticas, La Habana, Cuba

E-mail: rvaliente@uci.cu

Ing. Lisdan Rodríguez López

Universidad de las Ciencias Informáticas, La Habana, Cuba

E-mail: lisdanrp@uci.cu

AGRADECIMIENTOS

A mis compañeros y profesores del proyecto por ayudarme cada vez que los necesité, al tribunal por sus críticas constructivas, a mi oponente por su preocupación, a mi cotutor Salvador. A Liniuska por haberme ayudado en el comienzo de mi tesis, a mis compañeras de apartamento por todos los momentos que hemos compartido juntas. A mi tutora Rosayda por su ayuda incondicional. A mi “dúo de tesis” Leonel, gracias por darme ánimo, apoyo y fuerzas cuando pensaba que todo estaba perdido y por haberme guiado en la realización de este trabajo.

A mis amigas del pre: Annia Tornes por estar pendiente de mi, por su eterna amistad y comprensión, Annia Leiva por sus repases de Matemática y en los momentos difíciles sacarme las mejores sonrisas, a mi negrita linda y a Rogsany.

A unas personitas que nunca olvidaré, que han estado a mi lado estos 5 años, a Betina por soportar mi inseguridad y mis cambios de humor, por tener siempre una respuesta a todas mis preguntas, por el cariño que se ha ganado y lo que significa hoy para mi, a Maray por brindarme su apoyo aunque yo siempre la esté molestando, eres especial para mí, a Edelis por tenerme presente y aceptarme como soy, a Beatriz Lara por haberme dado el privilegio de conocerla, a Leidiana por la paz que transmite. A mi primer compañero de mesa Alejandro, quiero que sepas que te quiero mucho.

Llega el momento más difícil de mi carrera, y es el de intentar resumir en nombres de personas lo agradecida que en este momento me siento. Pienso que esta sección, quizás, no sea suficiente para mencionar a todos aquellos que han puesto un poquito de empeño y su ayuda desinteresada para cumplir este, mi sueño, y si existe alguien que no sienta su presencia en mis palabras, reciba mis más sinceras disculpas.

Le dedico esta tesis a mis padres por confiar en mí y brindarme su apoyo incondicional, a mi papá por recordarme que yo soy más grande que las dificultades que se me presenten en la vida, por ser mi gran ejemplo a seguir, espero que hoy te sientas muy orgulloso de mí. A mi mamá linda por estar en cada prueba de programación, por apoyarme, por mimarme, por malcriarme, por encontrar siempre en ella todo el apoyo que necesité para lograr mis propósitos.

A mi hermanita Ailyn, por ser mi doctora, mi amiga y sobre todo por ser la estrella que me guía. En este día tan importante quisiera que estuvieras a mi lado y sé que te sentirías muy orgullosa de mí como lo estoy yo de ti.

A mi familia por siempre estar pendiente de cada paso que doy, a mis tías, en especial a mi segunda madre Isolidia por entregarme todo su amor sin pedir nada a cambio.

A Yiya y familia por haberme acogido en su casa y hacerme sentir como si estuviera en la mía. A mi abuela Lolí por confiar siempre en mí, a mis suegros por brindarme su apoyo y confianza.

A mi novio Yaser, que no por ser el último en mencionar es menos especial. Gracias por todo el apoyo que me diste cuando más lo necesité, si no fuera por ti no estaría aquí. Viste lo logramos.

La Oficina Nacional de Estadísticas (ONE), constituye el órgano rector de la estadística en Cuba y tiene como objetivo captar, analizar y difundir los datos de todo el país. En la Universidad de las Ciencias Informáticas (UCI), específicamente en el Centro de Tecnología de Gestión de Datos (DATEC), se está desarrollando el Sistema de Información de Gobierno (SIGOB), que es un sistema para la captación de información estadística a partir de formularios matriciales y encuestas. La misión fundamental de la ONE es garantizar la producción de estadísticas de calidad a través del Sistema Estadístico Nacional, ejerciendo una adecuada dirección, ejecución y control de la captación de las cifras económicas y sociales, así como su difusión de acuerdo con los requerimientos de la economía y las demás necesidades del país. Una de las áreas que presenta dificultad para almacenar y estudiar dicha información, es la correspondiente a la educación, que incluye los datos referentes a las distintas enseñanzas contenidas en el sistema educacional cubano. El presente trabajo aborda el desarrollo del mercado de datos Educación, contribuyendo a la toma de decisiones.

PALABRAS CLAVE: Oficina Nacional de Estadísticas, Educación y Mercado de Datos.

INTRODUCCIÓN	1
CAPÍTULO 1: FUNDAMENTOS TEÓRICOS	5
1.1 Almacenes de datos.....	5
1.1.1 Características del almacén de datos	5
1.1.2 Ventajas del uso de los almacenes de datos	6
1.1.3 Componentes del almacén de datos	7
1.1.4 Mercado de datos.....	8
1.1.5 Metodologías para el desarrollo de almacenes de datos	8
1.1.6 Implementación del Modelo multidimensional.....	11
1.2 Etapas de desarrollo de un Almacén de Datos	15
1.2.1 Análisis y diseño.....	15
1.2.2 Extracción, Transformación y Carga.....	15
1.2.3 Inteligencia de negocios	16
1.3 Herramientas.....	17
1.3.1 Herramientas de modelado	17
1.3.2 Sistema Gestor de Base de Datos.....	17
1.3.3 Herramienta informática para el proceso de extracción, transformación y carga	18
1.3.4 Herramientas informáticas para la inteligencia de negocios	19
Conclusiones	21
CAPÍTULO 2: ANÁLISIS Y DISEÑO	22
2.1 Necesidades de información	22
2.1.1 Requisitos de información	22
2.1.2 Roles y permisos.....	24
2.2 Especificación de requerimientos	24
2.2.1 Requisitos funcionales.....	24
2.2.2 Requisitos no funcionales.....	25
2.3 Modelo de Casos de Uso del Sistema.....	28
2.3.1 Actores del Sistema.....	28
2.3.2 Diagrama de Casos de Uso del Sistema	29
2.3.3 Especificación de casos de uso del sistema	30
2.4 Especificación del modelo dimensional	32
2.4.1 Tablas de dimensiones.....	32
2.4.2 Tablas de hechos	32
2.4.3 Matriz bus.....	33
2.4.4 Modelo físico	33
2.5 Estrategia de respaldo y recuperación	35
Conclusiones	35
CAPÍTULO 3: IMPLEMENTACIÓN Y PRUEBA	36
3.1 Implementación del modelo de datos	36

ÍNDICE

3.2 Implementación de los procesos de ETL.....	37
3.2.1 Extracción de los datos	37
3.2.2 Implementación de las transformaciones.....	37
3.2.3 Carga de datos.....	37
3.2.4 Implementación de los trabajos	38
3.3 Implementación del proceso de BI	39
3.3.1 Implementación de los cubos OLAP.....	39
3.3.2 Arquitectura de información.....	43
3.3.3 Implementación de los reportes candidatos.....	44
3.4 Pruebas.....	46
3.4.1 Validación y prueba	46
3.4.2 Modelo V	46
3.4.2 Lista de chequeo	50
3.4.3 Pruebas de aceptación.....	53
Conclusiones	53
CONCLUSIONES	54
RECOMENDACIONES	55
BIBLIOGRAFÍA.....	56
REFERENCIAS BIBLIOGRÁFICAS	58
ANEXOS.....	60
GLOSARIO	73

ÍNDICE DE TABLAS

Tabla 2.1 Ejemplos de convenciones de nombrado	27
Tabla 2.2 Actores del sistema.	29
Tabla 2.3 Matriz Dimensional.....	33
Tabla 3.1 Esquemas del mercado de datos Educación.....	36
Tabla 3.2 Diseño de caso de prueba "Analizar información de los círculos infantiles"	48
Tabla 3.3 Lista de chequeo.....	50

ÍNDICE DE FIGURAS

Figura 1.1 Relación entre los componentes de un almacén de datos.....	7
Figura 2.1 Diagrama de casos de uso del sistema.....	29
Figura 2.2 Modelo Físico.....	34
Figura 3.1 Carga de la tabla hech_indicadores_circuitos	38
Figura 3.2 Trabajo para el proceso de ETL.....	39
Figura 3.3 Diseño de los cubos utilizando Pentaho Schema Workbench.....	40
Figura 3.4 Elementos que componen el cubo Indicadores Circuitos.	40
Figura 3.5 Elementos que componen el cubo Postgrado.	41
Figura 3.6 Elementos que componen el cubo Presupuesto.	41
Figura 3.7 Elementos que componen el cubo Matrícula graduados por organismo.....	41
Figura 3.8 Elementos que componen el cubo Matrícula graduados técnica y profesional	42
Figura 3.9 Elementos que componen el cubo Matrícula grupo especialidad.	42
Figura 3.10 Elementos que componen el cubo de las escuelas por educación.....	42
Figura 3.11 Elementos que componen el cubo tasas brutas y netas.....	43
Figura 3.12 Elementos que componen el cubo de las escuelas por educaciones y provincias.....	43
Figura 3.13 Arquitectura de información	44
Figura 3.14 Reporte Indicadores Generales de los Circuitos Infantiles.	46
Figura 3.15 Modelo V.....	47
Figura 3.16 Resultados de los casos de prueba.....	50
Figura 3.17 Comportamiento de los indicadores por secciones.	53

INTRODUCCIÓN

Desde la antigüedad el hombre ha utilizado información de diversos tipos, y su uso le ha reportado beneficio propio. Con el vertiginoso desarrollo de la informática y las telecomunicaciones, la cantidad de datos manejados ha aumentado considerablemente a través de los años. Es por ello que en la actualidad se hace necesario la creación de sistemas para el control y almacenamiento de los mismos, capaces de brindar herramientas confiables y de gran calidad, que estén acorde a las exigencias demandadas por los usuarios. Como resultado del uso de las tecnologías de almacenamiento de datos, se han introducido en las empresas mejoras en cuanto a organización, recursos humanos, desarrollo tecnológico y otros aspectos de interés.

En diferentes países se han creado sitios web con el objetivo de centralizar información referente a las distintas esferas de la sociedad, como lo son: economía, salud, energía, educación, entre otras. Estos sitios permiten además consultar los datos, para ayudar a la toma de decisiones. En Cuba se hace necesario llevar un estricto control de la información sobre las distintas ramas sociales, es por ello que se crea en el año 1994 la ONE.

La ONE constituye el órgano rector de la estadística en Cuba, tiene como objetivo captar, analizar y difundir los datos de todo el país. La fuente fundamental de los datos que se exponen procede del Sistema de Información Estadístico Nacional (SIEN), a través del cual la ONE, con sus 15 delegaciones territoriales y 168 oficinas municipales en todo el país, centraliza la información de los centros fuentes. No obstante, es válido reconocer a los Organismos de la Administración Central del Estado y otras instituciones, que han prestado su colaboración suministrando valiosas informaciones (1).

En la Universidad de las Ciencias Informáticas se creó el Centro de Tecnología de Gestión de Datos (DATEC). El mismo está regido por un conjunto de políticas y procedimientos para la obtención de productos de software empaquetados, y se encarga además de brindar apoyo a la ONE. Actualmente está desarrollando el Sistema de Información de Gobierno (SIGOB), que es un sistema para la captación de información estadística a partir de formularios matriciales y encuestas.

La misión fundamental de la ONE es garantizar la producción de estadísticas de calidad a través del Sistema Estadístico Nacional, ejerciendo una adecuada dirección, ejecución y control de la captación

de las cifras económicas y sociales, así como su difusión de acuerdo con los requerimientos de la economía y las demás necesidades del país en información estadística. Una de las áreas que presenta dificultad para almacenar y estudiar dicha información, es la correspondiente a la educación, que incluye los datos referentes a las distintas enseñanzas contenidas en el sistema educacional cubano.

A partir del año 1959, se inician en la isla las primeras transformaciones en este sector, con la creación de un Sistema Nacional de la Educación, de acceso universal y gratuito a todos los ciudadanos cubanos. Desde entonces y hasta la actualidad se han venido desarrollando importantes cambios, de ahí la gran necesidad de poder tener al alcance información disponible y exacta de todo lo concerniente a esta área de importancia estratégica para el país, por constituir unas de sus prioridades.

La ONE cuenta con un grupo de series estadísticas para recoger la información de las distintas áreas, entre ellos el de Educación, las cuales son almacenados en formatos de difícil acceso para su consulta, lo que complejiza los procesos de análisis de datos. Esto dificulta la manera en que se realizan los principales reportes y cruces de variables, indicadores, tasas, porcentajes y otros, por lo que se afecta en gran medida la disponibilidad de la información de Educación para Órganos del Estado, obstaculizando así el proceso de toma de decisiones sobre dicha área. La principal misión de la ONE es disponer de esta información de forma oportuna y con alta calidad para apoyar el proceso de toma de decisiones a nivel nacional.

Es válido destacar la existencia de un trabajo de diploma que precede a la presente investigación, la misma solo tuvo alcance hasta el análisis y diseño del mercado de datos, pero se pudo constatar que este trabajo no cumple con las necesidades actuales del negocio, por lo que no da solución al problema aún existente.

Por lo anteriormente expuesto se plantea como **problema científico** a resolver:

¿Cómo contribuir a la toma de decisiones en el área de Educación del Sistema de Información de Gobierno?

Se define como **objeto de estudio** de la presente investigación: los almacenes de datos, delimitando el **campo de acción** al mercado de datos estadístico para el área de Educación del Sistema de Información de Gobierno.

Para solucionar el problema científico planteado, se identifica como **objetivo general** de la investigación:

Desarrollar el mercado de datos Educación del Sistema de Información de Gobierno que contribuya a la toma de decisiones.

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

- Realizar el análisis y diseño del mercado de datos del área Educación.
- Implementar el mercado de datos del área Educación.
- Validar el mercado de datos del área Educación.

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

1. Caracterización de las metodologías, herramientas y tecnologías a utilizar en el desarrollo de almacenes de datos.
2. Levantamiento de requisitos.
3. Descripción de los casos de uso del mercado de datos.
4. Definición de los hechos, las medidas y las dimensiones del mercado de datos.
5. Diseño del modelo de datos.
6. Definición de la arquitectura del mercado de datos.
7. Diseño del subsistema de integración.
8. Diseño del subsistema de visualización.
9. Diseño de los casos de pruebas.
10. Implementación del subsistema de integración.
11. Implementación del subsistema de visualización.
12. Aplicación de las listas de chequeo.
13. Aplicación de los casos de pruebas.

El presente trabajo está conformado por introducción, 3 capítulos, conclusiones, recomendaciones, bibliografías y anexos. La estructura de los capítulos es la siguiente:

Capítulo 1: Fundamentos teóricos

Se abordan los principales conceptos relacionados con los almacenes de datos. Se hace un estudio de las herramientas y metodología a utilizar para el desarrollo del mercado de datos perteneciente al área de Educación.

Capítulo 2: Análisis y diseño del mercado de datos Educación.

Se hace un estudio preliminar del negocio, con el fin de identificar las perspectivas e indicadores a tener en cuenta durante el desarrollo del mercado de datos. Se realiza el análisis y diseño de los datos de las series históricas del área de Educación de la ONE.

Capítulo 3: Implementación y prueba del mercado de datos Educación.

En este capítulo se define la implementación del mercado de datos Educación, describiendo los procesos de Extracción, Transformación y Carga (ETL) y de Inteligencia de Negocio (BI) mediante la utilización de las herramientas descritas con anterioridad. También se realizan las pruebas de validación al mercado de datos del área de Educación a través de las listas de chequeo, cartas de aceptación del cliente y los casos de pruebas.

CAPÍTULO 1: FUNDAMENTOS TEÓRICOS

En este capítulo se abordan los principales conceptos relacionados con los almacenes de datos. Se hace un estudio de las herramientas y metodología a utilizar para el desarrollo del mercado de datos perteneciente al área de Educación.

1.1 Almacenes de datos

Existen diversos conceptos sobre los almacenes de datos, a continuación se exponen algunos:

Bill Inmon define: “el almacén de datos es una colección de datos integrada en una base de datos, orientada según un tema, diseñadas para soportar un Sistema de Soporte a las Decisiones (DSS, por sus siglas en inglés), donde cada unidad de dato es relevante en algún momento del tiempo”.

Según Ralph Kimball, otro conocido autor en este tema, un almacén de datos es: “una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis”.

Otro de los conceptos es que: “un almacén de datos o *datawarehouse* es una gran colección de datos que recoge información de múltiples sistemas fuentes u operacionales dispersos, cuya actividad se centra en la toma de decisiones. Una vez reunidos los datos de los sistemas fuentes, se guardan durante mucho tiempo, lo que permite el acceso a datos históricos; así los almacenes de datos proporcionan al usuario una interfaz consolidada única para los datos” (2).

Por lo anteriormente planteado, se puede concluir que los almacenes de datos son considerados una potente herramienta para la consulta y el análisis de información. Le brindan un gran apoyo a las empresas en cuanto a comodidad del trabajo, facilitándoles el proceso de toma de decisiones.

1.1.1 Características del almacén de datos

Según Bill Inmon un almacén de datos debe tener las siguientes características:

Organizado en torno a temas: la información recogida en el almacén se clasifica en dependencia a los aspectos que son de mayor importancia para la empresa.

Integrado: constituye el aspecto más importante dentro de las características de un almacén. La integración de datos consiste en convenciones de nombres, codificaciones consistentes, medida uniforme de variables, entre otras.

Dependiente del tiempo: la dependencia se evidencia de diferentes formas:

- La información almacenada representa los datos sobre un largo período de tiempo.
- La estructura contiene (implícita o explícitamente) un elemento de tiempo (día, semana, mes y año).
- La información, una vez registrada correctamente, no puede ser actualizada.

No volátil: el almacén de datos sólo permite cargar datos nuevos y acceder a los que ya están almacenados, pero no permite eliminar ni modificar los ya existentes.

Histórico: en los sistemas operacionales, los datos reflejan el estado de la actividad del negocio en el momento actual de la empresa. El almacén de datos se carga con los diferentes valores que toma una variable en el tiempo para permitir que se puedan realizar comparaciones (3).

1.1.2 Ventajas del uso de los almacenes de datos

La inversión que realiza una organización para implantar un sistema de almacén de datos dentro de la empresa generalmente trae consigo un costo muy elevado, sin embargo, el resultado de dicha inversión es garantizado a gran escala. Se puede conseguir un beneficio competitivo debido a una buena toma de decisiones gracias al almacén de datos implantado (3). Entre las principales ventajas de los almacenes de datos se pueden citar:

- Contribuyen a la toma de decisiones tácticas y estratégicas, proporcionando un sentido automatizado para identificar información clave desde grandes volúmenes de datos generados por procesos tradicionales o elementos de software.
- Posibilitan medir las acciones y los resultados de una mejor forma.
- Los procesos empresariales pueden ser optimizados, permitiéndole a los usuarios dar prioridad a determinadas decisiones y acciones (3).

1.1.3 Componentes del almacén de datos

Los almacenes de datos se componen por varios procesos que definen, en su conjunto, el ambiente que estos poseen. Aunque en el desarrollo de un almacén de datos existen diferencias debido a las particularidades de las organizaciones, por lo general, incluyen la realización de los siguientes componentes, lo que se puede observar en la figura 1.1:

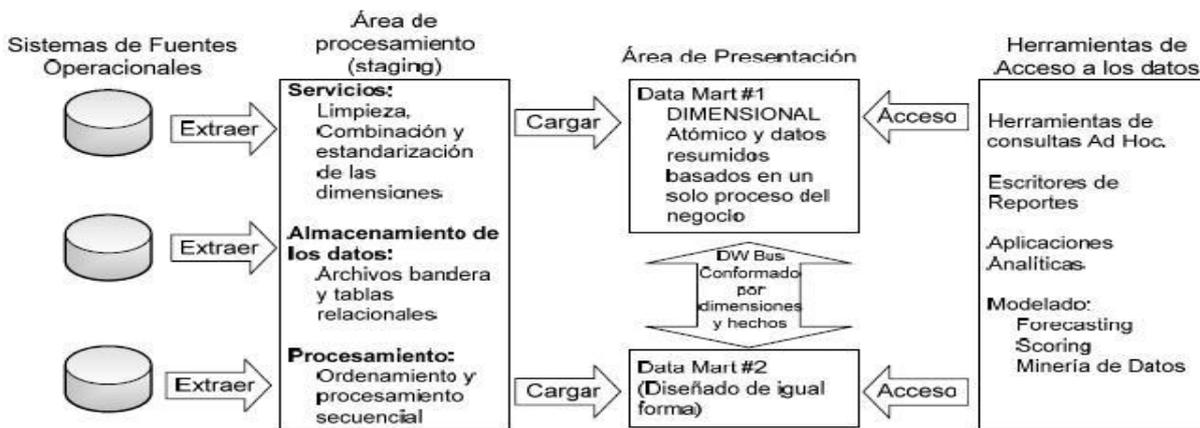


Figura 1.1 Relación entre los componentes de un almacén de datos.

El **área de procesamiento**, se encargará de los datos temporalmente, además de permitir la realización del proceso de ETL. Constituye la interfaz entre las fuentes operacionales y el área de presentación.

En el **área de presentación** los datos se encuentran organizados, almacenados y disponibles para poder consultarlos, realizar reportes o análisis por parte de los usuarios finales. Por lo general esta área es referenciada como una serie de mercados de datos integrados donde cada uno representa un proceso particular del negocio (4).

Las **herramientas de acceso a datos** son esenciales para consultar el área de presentación del almacén de datos, que puede abarcar desde una simple o personalizada herramienta de consulta, hasta una compleja y desarrollada aplicación de modelado o de minería de datos (5).

Estos sistemas lo poseen compañías o empresas para gestionar sus transacciones diarias. Las transacciones son almacenadas en diferentes formatos, incluyendo desde una base de datos relacional hasta cualquier tipo de ficheros, ya sea excel, XML, DBF, texto plano, entre otros.

1.1.4 Mercado de datos

Un *datamart* o mercado de datos es un subconjunto de un almacén de datos, que tiene como objetivo responder a un determinado análisis, función o necesidad y con una población de usuarios específica. En similitud con un almacén de datos, los datos se encuentran estructurados en modelos de estrella, copo de nieve o constelación de hechos. Un mercado de datos puede ser dependiente o independiente de un almacén de datos.

Ventajas de los mercados de datos

- Convertir los datos operacionales en información relacionada y estructurada llegando a generar el “conocimiento” necesario para la toma de decisiones.
- Centralizar y homogeneizar la información de gestión, evitando respuestas distintas a la misma pregunta.
- Permitir la visión global de la información en base a los conceptos de negocio que tratan los usuarios.
- Reducir costes evitando costosas extracciones manuales, permitiendo dedicar recursos a otras tareas.
- Mejorar la calidad de la gestión a partir de información relevante y con un significado homogéneo.
- Establecer una base única del modelo de información de las empresas y organizaciones (6).

1.1.5 Metodologías para el desarrollo de almacenes de datos

Existen diversas metodologías que definen y guían el ciclo de vida de desarrollo de un almacén. Las tendencias más comunes son las planteadas por Bill Inmon y Ralph Kimball, estas brindan un acercamiento a una propuesta ideal para el desarrollo de almacenes de datos. También existen otras variantes de estas metodologías como son: SQLBI, DM2, CRISP-DM y HEFESTO.

Cada una de estas metodologías, en dependencia de su autor, está orientada a la optimización del rendimiento y a la visión de los principales procesos que se deben tener en cuenta para construir un almacén de datos flexible y dinámico.

Bill Inmon y Ralph Kimball son considerados precursores en el estudio de los almacenes de datos. Ambos realizaron importantes aportes a las teorías de construcción de estos, es por ello que en este trabajo solo se realizará el estudio de lo planteado por ambos para definir la metodología a usar.

Inmon propone primeramente la construcción del almacén de datos y seguido de esta, los mercados de datos. Plantea además, la creación de un repositorio de datos corporativo como fuente de información consolidada, persistente, histórica y de calidad. Los mercados de datos se nutren del almacén de datos corporativo, debido a que son construidos descendentemente, convirtiéndose en un complejo empresarial de bases de datos relacionales.

Kimball a diferencia de Inmon, propone la elaboración del almacén de datos, pero partiendo de la construcción de los mercados de datos inicialmente. Entre sus principales aportes se encuentran la aparición de los conceptos de hechos y dimensiones, ambos agilizan el proceso de la toma de decisiones. Es considerada una metodología madura, donde están definidas etapas, actividades, artefactos y roles.

Para la realización del mercado de datos Educación se va a emplear la metodología para el desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio (DW&BI), basada en lo planteado por Kimball, pero adaptándola a necesidades y políticas de la Universidad de Ciencias Informáticas (UCI) (7), como se muestra en la figura 1.2.

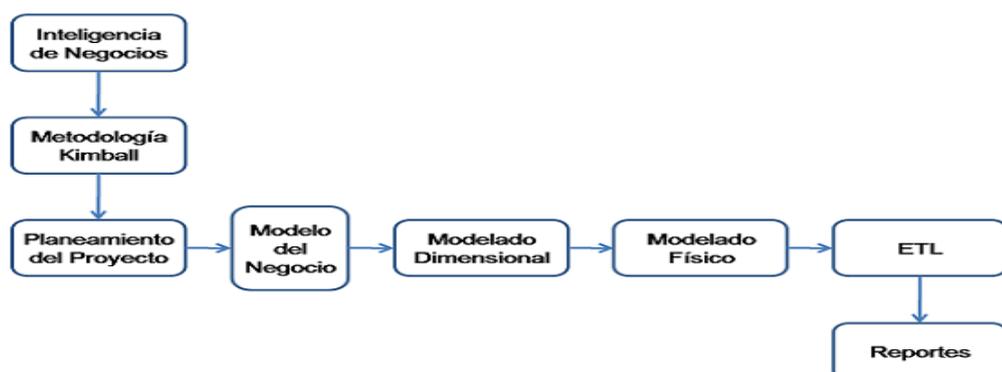


Figura 1.2 Metodología de DW & BI.

Las principales ventajas de esta metodología son:

- El desarrollo es iterativo e incremental, donde solo una pieza es construida a la vez (mercado de datos). Se basa en el paradigma de construcción adoptado por Kimball ("Bottom-Up"), que define la creación de los mercados de datos departamentales y luego su integración en un almacén de datos de la organización. Esta estrategia coincide con la estructura lógica que dispone la organización.
- Cuenta con mayor velocidad de respuesta al cliente.
- Los productos son más comprensibles para los usuarios.
- Es resistente y tolerante ante los cambios (7).

La metodología cuenta con cuatro fases principales, que incluyen las actividades a desarrollar, la descripción de cada una de ellas y los responsables. Las fases son: Requerimientos y Gestión de Proyectos, Arquitectura Técnica, Implementación e Implantación y Crecimiento. Dentro de dichas fases, se tienen los siguientes flujos de trabajo:

- **Estudio Preliminar o Planeación:** se realiza el estudio de la entidad cliente, la planeación del proyecto, se definen los objetivos, el alcance preliminar, los costos estimados y otras actividades.
- **Requerimientos:** se realiza en dos direcciones, una, mediante la identificación de las necesidades de información y reglas del negocio; y la otra con un levantamiento detallado de las fuentes de datos a integrar. Después se procede a la definición de los requerimientos.
- **Arquitectura y Diseño:** se definen las estructuras de almacenamiento, se diseñan las reglas de extracción, transformación y carga y se define la arquitectura de información que regirá el desarrollo de la solución.
- **Implementación:** se diseña físicamente el repositorio de datos, se crean las estructuras de almacenamiento, el área temporal de almacenamiento, se ejecutan las reglas de ETL y se configuran e implementan las herramientas de BI para la obtención de los elementos (reportes y gráficos) que se acordaron con el cliente final.
- **Prueba:** se realizan las pruebas al sistema, desde las pruebas de unidad hasta las de aceptación con el cliente final.
- **Despliegue:** se realiza un despliegue piloto en el cual se configuran los servidores y se instalan las herramientas, también se carga una muestra de los datos para demostrar que el sistema

funciona. Posterior a la aceptación del cliente se realiza la carga de los datos y la capacitación y transferencia tecnológica.

- **Soporte y Mantenimiento:** tras la implantación de la solución se brindan los servicios de soporte en línea, vía telefónica, web u otras según el contrato firmado y las condiciones de soporte establecidas.
- **Gestión y Administración del Proyecto:** a lo largo del ciclo de vida se realizan actividades de control, gestión y chequeo del desarrollo, los gastos, las utilidades, los recursos y demás actividades por parte del grupo de dirección del proyecto (8).

1.1.6 Implementación del Modelo multidimensional

A diferencia de los sistemas de bases de datos convencionales, la estructura de las tablas y sus relaciones son representadas mediante un modelo multidimensional, es decir, almacenan la misma información que el diagrama entidad relación pero organizada de manera diferente, de modo que se garantice velocidad y eficiencia en la recuperación de dicha información. Las principales ventajas están dadas por:

- Su enfoque al negocio y sus actividades.
- Búsquedas a gran velocidad.
- Adición de dimensiones y hechos que no se habían previsto, sin que esto implique volver a cargar los datos ya almacenados.
- Adición de nuevos atributos a las dimensiones (9).

El modelo multidimensional incluye tres variantes de modelación, que está determinado por la complejidad del sistema:

- **Esquema estrella:** es la técnica más común. En este esquema existe un único elemento central (tabla de hechos) conectado radialmente con las tablas de dimensiones.
- **Esquema copo de nieve (Snowflake):** es un esquema derivado del esquema de estrella, las tablas de dimensiones se ramifican en más puntas.
- **Esquema constelación:** está conformado por una serie de esquemas de estrella, o sea, una tabla de hechos central con otras auxiliares y sus respectivas tablas de dimensiones.

Para una adecuada comprensión de los distintos tipos de modelado existentes es necesario dominar algunos conceptos básicos referentes al tema, que a continuación se exponen:

Hecho

Es el objeto a analizar, posee atributos llamados de hechos o de síntesis, y son de tipo cuantitativo. Sus valores que se denominan medidas, se obtienen generalmente por la aplicación de una función estadística que resume un conjunto de valores en un único valor (10).

Dimensiones

Representan cada uno de los ejes en un espacio multidimensional. Suministran el contexto en el que se obtienen las medidas de un hecho. Las dimensiones se utilizan para seleccionar y agrupar los datos en un nivel de detalle deseado. Los componentes de una dimensión se denominan niveles y se organizan en jerarquías (10).

Es importante acotar que los hechos se guardan en tablas de hechos y las dimensiones en tablas de dimensiones.

Cubo

Consiste en un conjunto de celdas, cada una se identifica por la combinación de los miembros de las diferentes dimensiones y contiene el valor de la medida analizada para dicha combinación de dimensiones (11).

OLAP (Procesamiento analítico en línea)

OLAP es una clasificación de mercado de datos, basado en los cubos OLAP que se construyen, teniendo en cuenta los requisitos de cada área o departamento, las dimensiones y los indicadores necesarios de cada cubo relacional. El modo de creación, explotación y mantenimiento de los cubos OLAP es muy heterogéneo, en función de la herramienta final que se utilice, cumpliendo con las siguientes características:

- El acceso a los datos suele ser de sólo lectura. La acción más común es la consulta, con muy pocas inserciones, actualizaciones o eliminaciones.

- Los datos se estructuran según las áreas de negocio, y los formatos de los datos están integrados de manera uniforme en toda la organización.
- El historial de datos es a largo plazo, normalmente de dos a cinco años.
- Las bases de datos OLAP se suelen alimentar de información procedente de los sistemas operacionales existentes, mediante un proceso de extracción, transformación y carga (ETL) (12).

OLAP trabaja con tres tipos de almacenamiento:

ROLAP:

Hace posible que las agregaciones de la partición se almacenen en vistas de la base de datos relacional que se especificó en el origen de datos de la partición.

No permite almacenar una copia de los datos del origen en una estructura multidimensional en las carpetas de datos de Servicio de Análisis. En su lugar, cuando no se pueden derivar los resultados de la caché de consultas, se utilizan las vistas indexadas del origen de datos para responder a las consultas. La respuesta a las consultas suele ser más lenta que con los modos de almacenamiento MOLAP o HOLAP. El tiempo de procesamiento también suele ser más lento. No obstante, permite a los usuarios ver los datos y ahorrar espacio de almacenamiento al trabajar con conjuntos de datos grandes a los que no se suele consultar con frecuencia, como datos puramente históricos (13).

MOLAP (Multidimensional OLAP):

Da lugar a que las agregaciones de la partición y una copia de sus datos de origen se almacenen en una estructura multidimensional. Está optimizada para maximizar el rendimiento de las consultas. La ubicación de almacenamiento puede estar en el equipo en donde se define la partición. Dado que una copia de los datos de origen reside en la estructura multidimensional, las consultas se pueden resolver sin necesidad de obtener acceso a los datos de origen de la partición. Si se utilizan agregaciones, los tiempos de respuesta a las consultas pueden disminuir notablemente. Los datos de la estructura MOLAP de la partición están tan actualizados como el procesamiento más reciente de la misma (13).

HOLAP (Hybrid OLAP):

Combina atributos de los modos MOLAP y ROLAP. Al igual que MOLAP, hace que las agregaciones de la partición se almacenen en una estructura multidimensional. No almacena una copia de los datos de origen. Es el equivalente de MOLAP para las consultas que sólo tienen acceso a los datos de resumen de las agregaciones de una partición. Las consultas con acceso a los datos de origen como por ejemplo, la obtención de detalles para una celda de un cubo atómico sin datos de agregación, deben recuperar datos de la base de datos relacional y no serán tan rápidas como lo serían si los datos de origen se hubieran almacenado en la estructura MOLAP. Con el modo de almacenamiento HOLAP, los usuarios suelen experimentar notables diferencias en cuanto a los tiempos de las consultas, en dependencia de si esta puede resolver desde la caché o las agregaciones frente a los propios datos de origen (13).

Tipos de Almacenamiento OLAP

En el almacenamiento OLAP se trabajará específicamente con el ROLAP, el cual fue descrito con anterioridad, sus características lo convierte en el más indicado para la elaboración del Mercado de datos de Educación.

OLTP- Procesamiento transaccional en línea.

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

Estas bases de datos cumplen las siguientes características:

- El acceso a los datos está optimizado para tareas frecuentes de lectura y escritura. (Por ejemplo, la enorme cantidad de transacciones que tienen que soportar las BD de bancos o hipermercados diariamente).
- Los datos se estructuran según el nivel de aplicación (programa de gestión a medida, sistema de información departamental).
- Los formatos de los datos no son necesariamente uniformes en los diferentes departamentos (es común la falta de compatibilidad y la existencia de islas de datos).
- El historial de datos suele limitarse a los datos actuales o recientes (14).

1.2 Etapas de desarrollo de un Almacén de Datos

1.2.1 Análisis y diseño

Para la elaboración de un mercado de datos es necesaria la etapa de análisis y diseño, con el objetivo de tener un control acerca de las necesidades de los usuarios y poder obtener finalmente un sistema que responda a los intereses del negocio.

El análisis es la base fundamental para el desarrollo del mercado de datos, pues a partir de él se sientan las bases para los posteriores procesos de diseño e implementación.

Realizar el proceso de análisis es una tarea compleja, partiendo de que se necesita un estudio del proceso del negocio que se pretende informatizar para entender de manera clara y transparente lo que el usuario necesita. En la fase de análisis se generan un conjunto de artefactos que facilitan el desarrollo del sistema. La realización de dichos artefactos orienta el avance del cumplimiento de las tareas planteadas y garantiza la utilidad y el éxito del diseño de las estructuras.

Es necesario tener una descripción acerca de la organización que se desempeña como cliente, haciendo una definición del negocio en cuestión. Durante el período de análisis se tiene en cuenta el levantamiento de los requerimientos, creando una guía para los desarrolladores en la fase de implementación. Se elabora además el diagrama de diseño de la base de datos, donde se definen las relaciones entre los hechos y dimensiones, así como los diagramas de casos de uso. Se especifican los actores del negocio y del sistema, y su relación con los diferentes casos de uso.

En el diseño es donde se transforman los modelos lógicos conseguidos en la fase de análisis a modelos físicos. En esta fase se realiza el modelo de datos. Se construye también la matriz dimensional y se muestra el modelo de diseño realizado.

1.2.2 Extracción, Transformación y Carga

Extracción, transformación y carga (ETL), es un proceso de tres etapas en el uso de base de datos y almacenamiento de datos. Este permite la integración y el análisis de los datos almacenados en distintas bases de datos y de formatos heterogéneos. Una vez recopilados de diversas fuentes (extracción), los datos se modifican y se limpian para satisfacer las necesidades operacionales (transformación). Por último, se cargan en una base de datos de destino, de almacenamiento de datos o un mercado de datos para ser analizados. Excepto para el almacenamiento de datos e inteligencia empresarial, las herramientas ETL se pueden utilizar para mover datos de un sistema operativo a otro. El propósito del proceso de extracción es llegar a los sistemas de origen y recoger los datos necesarios para el almacén de datos.

Extracción

La primera parte de un proceso de ETL es extraer los datos de los sistemas de la fuente. Un elemento esencial de la extracción es el análisis de los datos extraídos, dando por resultado una evaluación donde se valida si los datos resuelven un patrón o una estructura prevista. Si no, los datos se rechazan o se transforman. La extracción convierte los datos en un formato común para el proceso de la transformación.

Transformación

La etapa de transformación de un proceso ETL consiste en la aplicación de una serie de reglas o funciones a los datos extraídos. Incluye la validación de los registros y su rechazo si no son aceptables. El importe de la manipulación necesaria para el proceso de transformación depende de los datos. Existen fuentes de datos que requieren poca transformación, mientras que otras pueden requerir una o más técnicas de transformación para cumplir con los requisitos empresariales y técnicos de la base de datos de destino o el almacén de datos. Los procesos más comunes utilizados para la transformación son la conversión, la limpieza, la normalización, el perfilado y selección.

Carga

La carga es la última etapa del proceso de ETL y se realiza en un repositorio de destino. Existen varias maneras en las que se pueden cargar los datos. Los datos son cargados al almacén durante esta fase. Dependiendo de los requisitos de la organización, este proceso se puede extender.

1.2.3 Inteligencia de negocios

Inteligencia de Negocio o *Business Intelligence*, es un agregado de aplicaciones y herramientas enfocadas al procesamiento de los datos en una empresa, con el fin de ayudar a la toma de decisiones y hacer varios análisis. Realiza gestión y consulta a los datos, generación de informes y reportes. En general, la inteligencia de negocio gestiona todo el procesamiento para traducir la lógica de negocios a la lógica de sistemas empresariales. Su mayor importancia radica en transformar datos primarios en información refinada y concentrada, para incrementar la eficiencia de las actividades de las áreas de una organización.

1.3 Herramientas

1.3.1 Herramientas de modelado

Para lograr la eficiencia hoy en día en el diseño de sistemas informáticos, se hace necesario el uso de herramientas que permitan comunicar con total exactitud y rapidez el objetivo del negocio durante todo el proceso de desarrollo.

El constante avance en el proceso de desarrollo de software dio lugar a la aparición de las denominadas herramientas CASE (Ingeniería de Software Asistida por Computadora), usadas para la modelación de proyectos, flujos de información, entidades de datos así como requerimientos del sistema. Estas facilitan en gran medida el proceso de planeación. Dentro de las herramientas de modelado se encuentran ER/Studio, Rational Rose, Visual Paradigm, entre otras.

La herramienta de modelado a utilizar es el Visual Paradigm 6.4: es una herramienta CASE que soporta las últimas versiones de UML (Lenguaje de Modelado Unificado). Provee el modelado de procesos de negocios. Es soportado por aplicaciones web, con opciones para varios idiomas. Permite además la generación de código para lenguajes como Java y exportación como HTML. Es muy fácil de instalar y actualizar. Es compatible con varias ediciones. Puede integrarse con distintas herramientas, como: Eclipse, NetBeans IDE, Oracle JDeveloper y otras (15).

1.3.2 Sistema Gestor de Base de Datos

Los sistemas gestores de base de datos son un conjunto de programas no visibles al usuario final que se encargan de la privacidad, la integridad, la seguridad de los datos y la interacción con el sistema operativo. Proporciona una interfaz entre los datos, los programas que los manejan y los usuarios finales. Cualquier operación que el usuario hace contra la base de datos está controlada por el gestor (16).

PostgreSQL 8.4

PostgreSQL es el sistema gestor de bases de datos, escogido para la implementación del sistema, por sus características, que se relacionan a continuación:

- Brinda confiabilidad e integridad a los datos. El PostgreSQL puede ser ejecutado en la mayoría de los Sistemas Operativos más utilizados en el mundo incluyendo, Linux, varias versiones de UNIX y Windows.
- Este robusto sistema gestor de bases de datos es una solución real, a los complejos problemas del mundo empresarial y a la vez mantiene la eficiencia, al consultar los datos.
- Es un producto sin costos de licencia, que se convierte en una alternativa extremadamente atractiva para las empresas, que buscan un ahorro significativo de costos en activos.
- Posee numerosas interfaces nativas de lenguajes como: C + +, Java, .Net, Perl, Python, Ruby, Tcl y ODBC.
- Pueden hacerse todas las modificaciones, mejoras o cambios que se estimen conveniente, pues el código fuente está disponible bajo los más liberales términos de licencia de código abierto (17).

PgAdmin 1.10

Es considerada una de las herramientas más usadas para administrar las bases de datos en PostgreSQL, dentro de sus principales características se encuentran:

- Posee licencia libre, disponible para todos los usuarios.
- Accede a todos los objetos del PostgreSQL y responde a 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 (18).

1.3.3 Herramienta informática para el proceso de extracción, transformación y carga

Pentaho Data Integration (PDI, también conocido como *Kettle*)

El PDI es el componente de Pentaho encargado de los procesos de ETL. Aunque las herramientas ETL son las más usadas en entornos de almacenes de datos, *Kettle* también puede ser utilizado para otros fines, como lo son:

- Migración de datos entre las aplicaciones o bases de datos.
- Exportación de datos desde bases de datos para archivos planos.
- Carga de datos de forma masiva en bases de datos.
- Limpieza de datos (19).

Dentro de sus principales características se encuentran:

- Cada proceso es creado con una herramienta gráfica donde se especifica qué se va hacer sin necesidad de escribir código que indique cómo hacerlo.
- Admite una amplia gama de formatos de entrada y salida, incluyendo archivos de texto, hojas de datos, archivos XML, propiedades de Java y los motores de base de datos open source y propietarios.
- Arquitectura extensible que permite desarrollar conectores y plugins fácilmente.
- Basado en repositorio, facilita la reutilización de componentes de transformación, colaboración y administración de modelos, conexiones, logs.
- En todos los pasos de base de datos existe un botón SQL que generará el código requerido para las tablas y/o índices. También está disponible a nivel de transformación y trabajo (19).

1.3.4 Herramientas informáticas para la inteligencia de negocios

Pentaho Schema Workbench

Es una interfaz de diseño que le permite crear y probar esquemas cubo OLAP Mondrian visualmente. El motor de Mondrian procesa las solicitudes de MDX con la ROLAP (OLAP Relacional) esquemas. Estos archivos son los modelos de esquemas XML de metadatos que se crean en una estructura específica utilizada por el motor de Mondrian. Ofrece como principales funcionalidades:

- Editor de esquema integrado con el origen de datos subyacente para su validación.
- Prueba de consultas MDX en contra del esquema de base de datos y pantalla.
- Examinar bases de datos de estructura subyacente de pantalla (20).

Mondrian OLAP Server

Para obtener la funcionalidad de Procesamiento Analítico en Línea se utilizan otras dos aplicaciones: el servidor OLAP Mondrian, que combinado con Jpivot, permite realizar consultas al almacén de datos y permite que los resultados sean presentados mediante un navegador de modo que el usuario pueda realizar las actividades típicas de navegación. Mondrian utiliza MDX como lenguaje de consulta, lenguaje propuesto por Microsoft. Funciona sobre las bases de datos estándar del mercado: Oracle, DB2, SQL-Server, MySQL y otras, lo cual facilita el desarrollo de negocio basado en la plataforma Pentaho (21).

Mondrian es una de las aplicaciones más importantes de la plataforma Pentaho BI. Es un servidor OLAP open source que gestiona comunicación entre una aplicación OLAP (escrita en Java) y la base de datos con los datos fuente.

Entre sus ventajas están:

- Agiliza la consulta de grandes volúmenes de datos.
- Alta velocidad de respuesta.
- Permite realizar queries a mercado de datos.
- Es un motor ROLAP con caché (21).

Pentaho BI Server 3.6

La plataforma Pentaho BI Server provee el soporte y la infraestructura necesaria para crear soluciones de inteligencia empresarial a problemas de negocios. El marco proporciona los servicios básicos, incluidos autenticación, registro, auditoría, servicios web y motor de reglas. La plataforma también incluye un motor de solución que integra reportes, análisis, tableros de comandos y componentes de minería de datos. La aplicación Pentaho BI Server funciona como un sistema basado en administración web de informes, el servidor de integración de aplicaciones y un motor de flujo de trabajo ligero (secuencias de acción.) Su diseño posibilita integrarse fácilmente en cualquier proceso de negocio (22).

Entre sus ventajas se encuentran:

- Integración con procesos de negocio.
- Administra y programa reportes.
- Administra seguridad de usuarios.

Apache Tomcat 5.5

Apache Tomcat es una implementación de software de código abierto de Java Servlet y tecnologías JavaServerPages. Las especificaciones Java Servlet y JavaServerPages se desarrollan bajo la Java CommunityProcess. Es desarrollado en un entorno abierto y participativo, publicado bajo la licencia Apache versión 2.

- 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 (23).

Conclusiones

En este capítulo se realizó un estudio sobre los principales conceptos relacionados a los términos de almacenes de datos, mercado de datos, así como una breve descripción acerca de las tecnologías de desarrollo de los almacenes de datos. Se decidió usar para el desarrollo del mercado de datos la metodología para el desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio (DW&BI). Para la realización de este trabajo se determinó que se usarán las herramientas que se mencionan a continuación, partiendo de las posibilidades y ventajas que estas ofrecen y teniendo en cuenta que están aprobada por el centro DATEC para el desarrollo de productos de software dentro de la UCI. Se propuso PostgreSQL en su versión 8.4 como Sistema Gestor de Base de Datos y Visual Paradigm 6.4 como herramienta CASE a utilizar. Como herramientas para el proceso de ETL se usará Pentaho Data Integration 4.1. Como herramientas para el proceso de inteligencia de negocio se determinó Apache Tomcat 5.5, Pentaho BI Server 3.6, Mondrian OLAP Server y Pentaho Schema Workbench.

CAPÍTULO 2: ANÁLISIS Y DISEÑO

En este capítulo se hace un estudio preliminar del negocio, con el fin de identificar las perspectivas e indicadores a tener en cuenta durante el desarrollo del mercado de datos. Se realiza el análisis y diseño de los datos de las series históricas del área de Educación de la ONE.

2.1 Necesidades de información

Luego de un previo análisis sobre el departamento de Educación de la ONE, donde se estudiaron las características y funcionamiento de este, se determinaron las necesidades de los usuarios. Todos estos requisitos informativos fueron derivados de los archivos excel o series que agrupan los indicadores de esta área.

2.1.1 Requisitos de información

Los requerimientos de información son aquellos que se corresponden con las necesidades de los usuarios y reflejan la información que estos desean. Se definen a partir de los elementos disponibles en las fuentes. A continuación se muestran todos los requisitos de información identificados para el departamento de Educación de la ONE:

1. Obtener la matrícula final por año de vida de los círculos infantiles.
2. Obtener la cantidad de círculos infantiles por año.
3. Obtener la capacidad final de los círculos infantiles por año.
4. Obtener la asistencia promedio por año de vida en los círculos infantiles al finalizar el año.
5. Obtener la cantidad de personal técnico de los círculos infantiles por año.
6. Obtener la cantidad de madres beneficiadas en círculos infantiles por año.
7. Obtener la cantidad de niños de 0-5 años de edad en los círculos infantiles.
8. Obtener la cantidad de escuelas por provincias, tipo de enseñanza y curso escolar.
9. Obtener la cantidad de personal docente frente al aula por provincias, tipo de enseñanza, sexo y curso escolar.
10. Obtener la matrícula inicial de las escuelas por provincias, tipo de enseñanza, sexo y curso escolar.
11. Obtener la cantidad de graduados por provincias, tipo de enseñanza, sexo y curso escolar.
12. Obtener el gasto público por año.
13. Obtener la per cápita por año.

14. Obtener matrícula inicial por diez mil de las escuelas por curso escolar, nivel de enseñanza y por sexo.
15. Obtener la cantidad de personal docente de las escuelas por curso escolar, nivel de enseñanza y por sexo.
16. Obtener la cantidad de escuelas por curso escolar y por nivel de enseñanza.
17. Obtener la matrícula inicial de las escuelas por curso escolar, nivel de enseñanza y por sexo.
18. Obtener la cantidad de personal docente frente al aula de las escuelas por curso escolar, nivel de enseñanza y por sexo.
19. Obtener la cantidad de becarios de las escuelas por curso escolar, nivel de enseñanza y por sexo.
20. Obtener la cantidad de seminternos de las escuelas por curso escolar, nivel de enseñanza y por sexo.
21. Obtener la cantidad de graduados de las escuelas por curso escolar, nivel de enseñanza y por sexo.
22. Obtener las tasas brutas por curso escolar y tipo de enseñanza.
23. Obtener las tasas netas por curso escolar y nivel de enseñanza.
24. Obtener el índice de paridad de las tasas brutas y netas por curso escolar y nivel de enseñanza.
25. Obtener la matrícula total de las tasas brutas y netas por curso escolar, sexo y tipo de enseñanza.
26. Obtener la matrícula de edad establecida de las tasas brutas y netas por curso escolar, sexo y tipo de enseñanza.
27. Obtener la población en edad escolar de las tasas brutas y netas por curso escolar, sexo y tipo de enseñanza.
28. Obtener la matrícula inicial de la enseñanza Técnica y Profesional por tipo de carrera y por curso escolar.
29. Obtener la cantidad de graduados de la enseñanza Técnica y Profesional por tipo de carrera y por curso escolar.
30. Obtener el total de nuevos ingresos por especialidades y por curso escolar.
31. Obtener la matrícula inicial por especialidades y por curso escolar.
32. Obtener la cantidad de graduados por especialidades y por curso escolar.
33. Obtener la matrícula inicial de graduados de los organismos por curso escolar y tipo de estudio.
34. Obtener el total de graduados de los organismos por curso escolar y tipo de estudio.

35. Obtener las actividades de postgrado por años y por figuras.

36. Obtener los participantes de postgrado por años y por figuras.

2.1.2 Roles y permisos.

La definición de los roles y permisos es de vital importancia para un sistema, mediante estos se garantiza la seguridad, pues cada persona tiene definido su función y los privilegios necesarios para poder ejecutarla. A continuación se especifican los roles y permisos:

Administrador de ETL: es el encargado de realizar los procesos de extracción, transformación y carga de datos hacia el mercado, tiene permisos de lectura y escritura.

Administrador de la Base de Datos: tiene la responsabilidad de mantener funcional la base de datos que conforma el mercado de datos. Tiene control total sobre el mercado, puede insertar, actualizar, crear objetos, eliminar, dar permisos, y demás tareas de administración.

Analista de información: es el encargado de trabajar con los reportes, tiene permisos de solo lectura sobre los mismos.

Administrador: administra el Pentaho BI Server, crea usuarios, reportes, define privilegios y otras tareas de administración sobre la herramienta de BI.

2.2 Especificación de requerimientos

2.2.1 Requisitos funcionales

Los requisitos funcionales son aquellas funcionalidades que el sistema debe cumplir. Son definidos a partir de las necesidades de los usuarios. A continuación se muestran los 26 requisitos funcionales identificados para el desarrollo del mercado de datos Educación:

RF1- Extraer información de los archivos excel del área de Educación.

RF2- Transformar y cargar información de los archivos excel del área de Educación.

RF3- Insertar usuario.

RF4- Eliminar usuario.

RF5- Insertar rol.

- RF6- Eliminar rol.
- RF7- Insertar reporte.
- RF8- Eliminar reporte.
- RF9- Modificar reporte.
- RF10- Configurar elementos del cubo.
- RF11- Ordenar resultados.
- RF12- Mostrar padres.
- RF13- Mostrar propiedades.
- RF14- Suprimir filas.
- RF15- Invertir eje.
- RF16- Detallar miembros.
- RF17- Entrar en detalles.
- RF18- Mostrar datos de origen.
- RF19- Mostrar gráfico.
- RF20- Configurar gráfico.
- RF21- Editar consulta MDX.
- RF22- Exportar reporte a “.pdf”.
- RF23- Exportar reporte a “.xls”.
- RF24- Imprimir reporte.
- RF25- Ver un reporte.
- RF26- Autenticar usuario.

2.2.2 Requisitos no funcionales

Los requisitos no funcionales son características que de una u otra forma pueden limitar el correcto funcionamiento del sistema, como por ejemplo la seguridad, el rendimiento, la fiabilidad, entre otros. También hacen posible que el producto pueda desplegarse al ser terminado y funcionar de manera

eficiente. Se seleccionaron 11 requisitos no funcionales, basados en las características del sistema que se va a realizar, los cuales se muestran a continuación.

Usabilidad

RNF 1. Cumplir con las pautas de diseño de las interfaces.

El sistema debe tener una interfaz gráfica uniforme, incluyendo pantallas, menús y opciones.

RNF 2. Mostrar los mensajes, títulos y demás textos que aparezcan en la interfaz del sistema en idioma español.

Los títulos de los componentes de la interfaz, los mensajes para interactuar con los usuarios y los mensajes de error, deben ser en idioma español y tener una apariencia uniforme en todo el sistema. Los mensajes de error deberán ser lo suficientemente informativos para dar a conocer la severidad del error.

RNF 3. Establecer tiempo de entrenamiento requerido para que usuarios normales sean productivos operando el sistema.

El tiempo de entrenamiento requerido para que usuarios normales sean productivos operando el sistema deberá ser entre siete y 14 días. Para aquellos usuarios con un nivel avanzado se define como valor máximo siete días. Para lograr el cumplimiento de los tiempos establecidos por parte de los usuarios es necesario un dominio del funcionamiento del negocio en correspondencia con el rol que ocupen.

RNF 4. Diseñar un reporte del almacén de datos de manera sencilla y ágil.

Un usuario con conocimientos básicos de sistema podrá diseñar un reporte del mercado de datos de manera ágil y sencilla sin necesidad de ser un experto en las herramientas requeridas para ello.

RNF 5. Navegar en los reportes del almacén de datos de manera ágil.

El usuario podrá realizar los cruces y agrupaciones de variables en los reportes del mercado de datos de manera dinámica en la misma pantalla de trabajo. Esto permite agilizar la navegabilidad de los usuarios en un reporte.

Fiabilidad

RNF 6. Garantizar la persistencia de la información.

Para garantizar la persistencia de la información se realizará un respaldo total de los datos del mercado, con una frecuencia anual. Toda esta información se almacenará en el área de la dirección de informática en un banco de datos especial. Esta información se almacenará en el edificio correspondiente a la oficina de estadísticas de La Habana y será responsabilidad del grupo de administración de redes de la ONE.

RNF 7. Garantizar el cumplimiento de actualización de los datos en el almacén.

La información contenida en el mercado tendrá una precisión y exactitud anual, en correspondencia con la periodicidad con que se recogen los datos.

Soporte

RNF 8. Lograr que los elementos definidos en el almacén tengan una estructura homogénea.

Se definen convenciones de nombrado con el objetivo de manejar un vocabulario común en el almacén de datos para un entendimiento claro y conciso de las estructuras por parte de los desarrolladores que interactúen con dicho almacén. En la tabla 2.2 se muestra la codificación y componentes de una solución de DWH.

Tabla 2.1 Ejemplos de convenciones de nombrado

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

Interfaz

RNF 9. Acceder al sistema

Se recomienda acceder a la aplicación a través del protocolo HTTP usando el navegador Firefox 2 o una versión superior.

Requisitos de software

RNF 10. Instalar en las estaciones de trabajo los requerimientos mínimos de software

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

- Navegador web (Preferentemente firefox 2.0 o superior)
- Java Virtual Machine 6.0
- Schema Workbench 3.2.1 en caso de que un usuario capacitado requiera la construcción de esquemas multidimensionales para el diseño de nuevos reportes.

Requisitos de hardware

RNF 11. Proporcionar características mínimas de hardware a los servidores

Para que el software funcione en óptimas condiciones los servidores deben contar con los siguientes requerimientos de hardware:

- 512 GB RAM
- 1 Microprocesador Core2Duo

2.3 Modelo de Casos de Uso del Sistema

Para la realización del mercado de datos fue necesario identificar una serie de casos de uso, como fueron los casos de uso funcionales y los de información.

2.3.1 Actores del Sistema

Los actores del sistema pueden representar el rol que juega una o varias personas, un equipo o un sistema automatizado. En la tabla 2.3 se muestra la relación de actores con sus descripciones.

Tabla 2.2 Actores del sistema.

Actor	Descripción
Administrador	Gestiona los reportes OLAP, permisos, roles y usuarios.
Administrador ETL	Realiza los procesos ETL.
Analista	Analiza y consulta la información de la educación.

2.3.2 Diagrama de Casos de Uso del Sistema

Un diagrama de casos de uso del sistema representa gráficamente a los procesos y su interacción con los actores. En la figura 2.1 se muestra el diagrama realizado para el mercado de datos del área Educación.

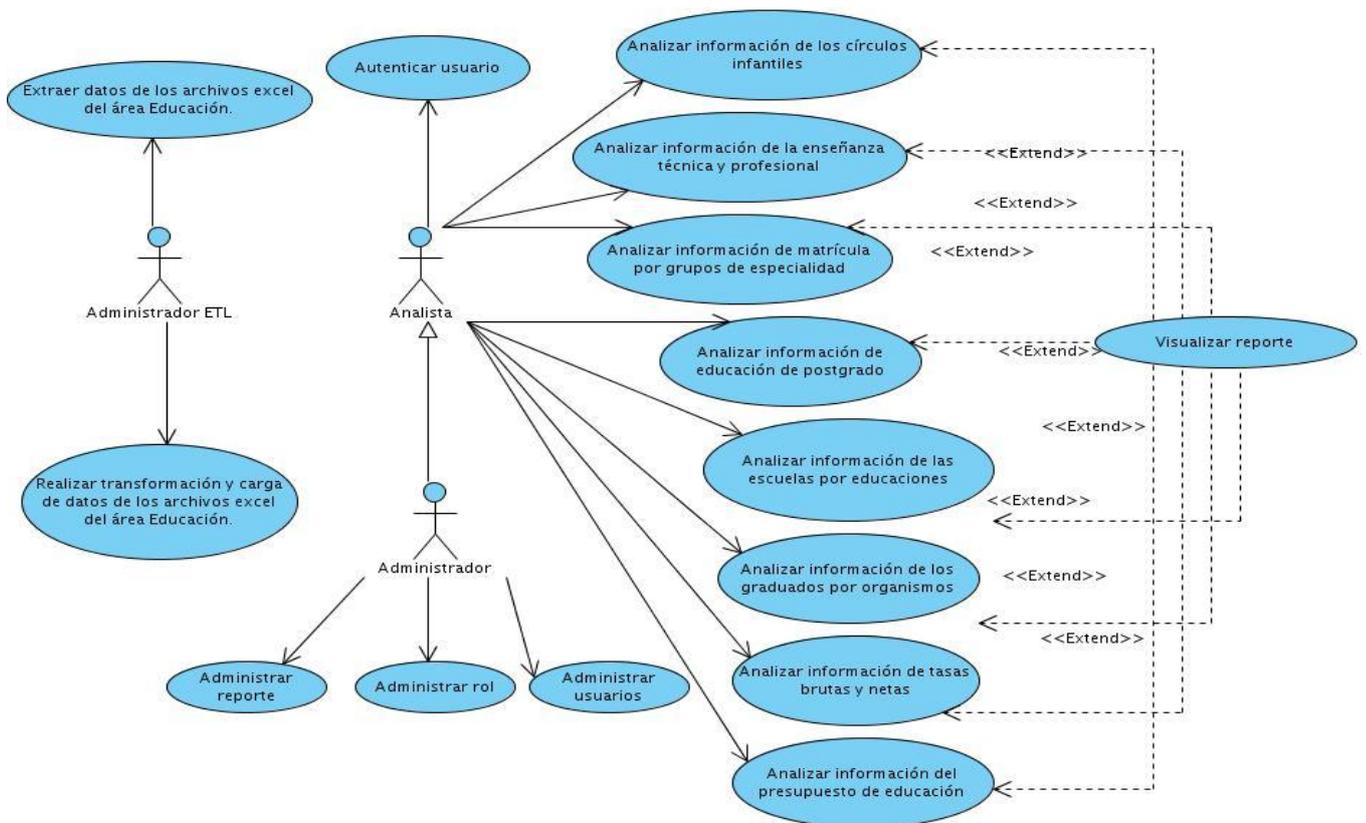


Figura 2.1 Diagrama de Casos de Uso del Sistema

2.3.3 Especificación de casos de uso del sistema

Los casos de uso del sistema son artefactos narrativos que describen, bajo la forma de acciones y reacciones, el comportamiento del sistema desde el punto de vista del usuario. Se definieron 15 casos de uso, de estos 8 de información y 7 funcionales, los que se mencionan a continuación:

- Analizar información de los círculos infantiles.
- Analizar información de las enseñanzas técnica y profesional.
- Analizar información de los graduados por organismos.
- Analizar información de educación de postgrado.
- Analizar información de las escuelas por educaciones.
- Analizar información de tasas y netas.
- Analizar información del presupuesto de educación.
- Analizar información de matrícula por grupos de especialidad.
- Administrar reportes.
- Administrar usuarios.
- Administrar rol.
- Visualizar reporte.
- Autenticar usuario.
- Realizar transformación y carga de datos de los archivos excel del área Educación.
- Extraer datos de los archivos excel del área Educación.

Se muestra la descripción del caso de uso Analizar información de las escuelas por educaciones. Para ver los demás remitirse al artefacto Descripción de Casos de Uso del Sistema.

Caso de Uso:	Analizar información de las escuelas por educaciones.
Tipo:	Información
Actores:	Analista, Administrador
Resumen:	El CU inicia cuando el actor entra al sistema, luego este selecciona el reporte de Escuelas, y cuando el sistema muestra el reporte termina el CU.
Precondiciones:	Compleitud del almacén. Carga de los datos.
Referencias	Ver RI8, RI9, RI10, RI11.

Prioridad	Crítico.	
Flujo Normal de Eventos		
Acción del Actor	Respuesta del Sistema	
1. El actor entra al sistema	2. El sistema muestra las áreas de análisis existentes.	
3. El actor selecciona el AA Educación.	4. El sistema muestra los libros de trabajos que contiene el AA Educación.	
5. El actor selecciona el LT Escuelas por educación.	6. El sistema muestra los reportes pertenecientes al LT Escuelas por educación.	
7. El actor selecciona el reporte deseado.	8. El sistema muestra la información contenida en el reporte seleccionado y brinda la posibilidad de analizar el reporte desde diferentes perspectivas. Ir al CU: Visualizar reporte.	
Opciones de reportes de Analizar información de las escuelas.		
Perspectivas de análisis	Posibles resultados	
	Medidas	Periodicidad
Variables de entrada relacionadas con el CU Analizar información de las escuelas: <ul style="list-style-type: none"> • Sexo. • Curso. • Nivel Enseñanza. • Provincia. • Tipo de Enseñanza. 	Variables de salida disponibles en el caso de uso Analizar información de las escuelas: <ul style="list-style-type: none"> • Escuelas. • Personal docente frente al aula. • Graduados. • Matrícula inicial de las escuelas. • Seminternos. • Becarios. • Personal docente. • Matrícula inicial por diez mil. 	Rango de tiempo en que se solicitan las variables de salida: <ul style="list-style-type: none"> • Anual.
Pos condiciones	Disponibilidad de opciones (cruces de variables) de reportes relacionados	

con el caso de uso **Analizar información de las escuelas.**

2.4 Especificación del modelo dimensional

2.4.1 Tablas de dimensiones

Las dimensiones definidas son las siguientes:

1. dim_sexo
2. dim_figura
3. dim_provincia
4. dim_tipo_carrera
5. dim_tipo_estudio
6. dim_especialidad
7. dim_curso_escolar
8. dim_temporal_anno
9. dim_tipo_ensennanza
10. dim_nivel_ensennanza
11. dim_dpa
12. dim_organismo

2.4.2 Tablas de hechos

Los hechos definidos son los siguientes:

1. hech_postgrado
2. hech_presupuesto
3. hech_tasas_brutas_netas
4. hech_indicadores_circulos
5. hech_escuelas_educaciones
6. hech_educacion_provincias
7. hech_matri_graduados_organismos
8. hech_matricula_grupos_especialidades
9. hech_matricula_graduados_tecnico_profesional

2.4.3 Matriz bus

En la matriz dimensional o matriz bus es donde se refleja la relación existente entre las tablas de hechos y sus dimensiones asociadas. A continuación se muestra la matriz bus definida a partir de las relaciones entre los hechos y las dimensiones identificadas en el mercado de datos Educación.

Matriz Dimensional												
Hechos	Dimensiones											
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
H1	X										X	
H2		X	X	X	X						X	
H3	X										X	
H4			X	X		X					X	
H5				X			X				X	
H6			X	X	X							
H7				X				X			X	X
H8				X						X	X	
H9	X								X		X	

Tabla 2.3 Matriz Dimensional

2.4.4 Modelo físico

Una vez definido dentro del negocio las dimensiones y las medidas, se procede a la estructuración del modelo que existirá. En tal sentido se puede destacar que por las necesidades actuales del negocio existen varias series históricas que unifican las dimensiones definidas y las medidas que se han especificado hasta el momento.

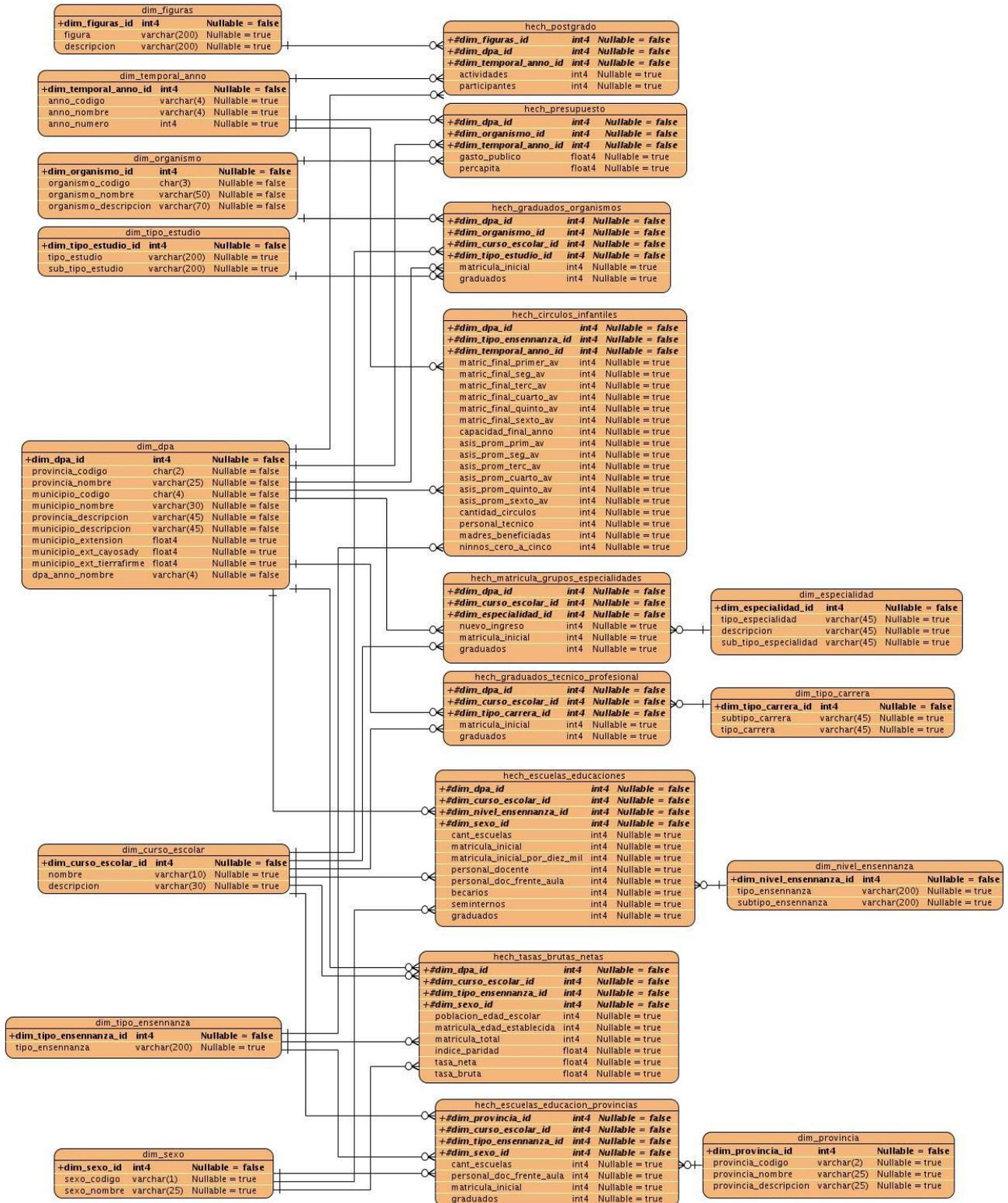


Figura 2.2 Modelo Físico

2.5 Estrategia de respaldo y recuperación

En un sistema informático juega un papel fundamental que el mismo esté preparado para recuperarse ante la pérdida de información. Los mercados de datos no están exentos de dicha problemática. Para contrarrestar la pérdida de información y poder poner en marcha nuevamente el mercado de datos se definieron las siguientes políticas:

- Hacer una copia total de la base de datos del mercado de datos cada vez que se realicen las cargas de los ficheros fuentes. En este caso anual.
- Se le realizará salvas a todas las tablas del mercado, tanto a las dimensiones como los hechos.
- Se guardarán las salvas en discos duros por un período de un año.
- Deben ejecutarse pruebas de restauración a las salvas guardadas para verificar que contienen la información correcta.

Conclusiones

En este capítulo se expusieron los resultados del proceso de análisis y diseño del mercado de datos Educación. Fueron identificados 26 requisitos funcionales y 11 no funcionales, así como 12 dimensiones, nueve hechos y 36 medidas. Se estructuró el modelo físico de la base de datos, se establecieron los roles y permisos para la aplicación, además de quedar definidos y descritos los 15 casos de usos del sistema y se elaboró la matriz bus.

CAPÍTULO 3: IMPLEMENTACIÓN Y PRUEBA

En este capítulo se define la implementación del mercado de datos Educación, describiendo los procesos de ETL y de BI mediante la utilización de las herramientas descritas con anterioridad. También se realiza la validación del mercado de datos.

3.1 Implementación del modelo de datos

Para el desarrollo del mercado de datos Educación se definieron dos esquemas: dimensiones, que contiene todas las dimensiones comunes del almacén y el esquema mart_educacion, que agrupa todos los hechos y dimensiones del mercado de datos Educación. La solución cuenta con 19 tablas, 12 dimensiones y nueve hechos. En la siguiente tabla se describe dicha estructura:

Tabla 3.1 Esquemas del mercado de datos Educación.

Esquemas	Tablas
dimensiones	dim_sexo
dimensiones	dim_figura
dimensiones	dim_provincia
dimensiones	dim_tipo_carrera
dimensiones	dim_tipo_estudio
dimensiones	dim_especialidad
dimensiones	dim_temporal_anno
dimensiones	dim_nivel_ensennanza
dimensiones	dim_tipo_ensennanza
dimensiones	dim_curso_escolar
dimensiones	dim_dpa
dimensiones	dim_organismo
mart_educacion	hech_presupuesto
mart_educacion	hech_postgrado
mart_educacion	hech_escuelas_educaciones
mart_educacion	hech_indicadores_circuitos
mart_educacion	hech_educacion_provincias

mart_educacion	hech_matri_graduados_organismos
mart_educacion	hech_tasas_brutas_netas
mart_educacion	hech_matricula_grupos_especialidades
mart_educacion	hech_matricula_graduados_tecnico_profesional

3.2 Implementación de los procesos de ETL

3.2.1 Extracción de los datos

El proceso de ETL se inicia al extraer los datos desde los sistemas de origen. El formato de los datos fuente correspondiente a las series del área de Educación se encuentra en ficheros de tipo excel, estos fueron estandarizados a un formato común para iniciar el proceso de transformación.

3.2.2 Implementación de las transformaciones

Después de realizada la extracción de los datos el sistema se encuentra listo para la etapa de transformación. Las transformaciones constituyen un elemento básico dentro de la implementación del proceso ETL, está compuesta por pasos y estos a su vez se encuentran unidos a través de saltos, dichos pasos son los elementos más pequeños dentro de las transformaciones. Al concluir este proceso, los datos estarán listos para ser cargados (24).

3.2.3 Carga de datos

La carga es el último subproceso dentro de los procesos de ETL, consiste en cargar todos los datos que han sido transformados anteriormente. Para este proceso se tiene que realizar inicialmente la carga de las tablas que no dependen de la información de otras tablas, con estas características se encuentran las tablas del esquema dimensiones y las del esquema mart_educacion que representan dimensiones. En la siguiente figura se muestra el proceso ETL descrito anteriormente para el hecho relacionado a los círculos infantiles.

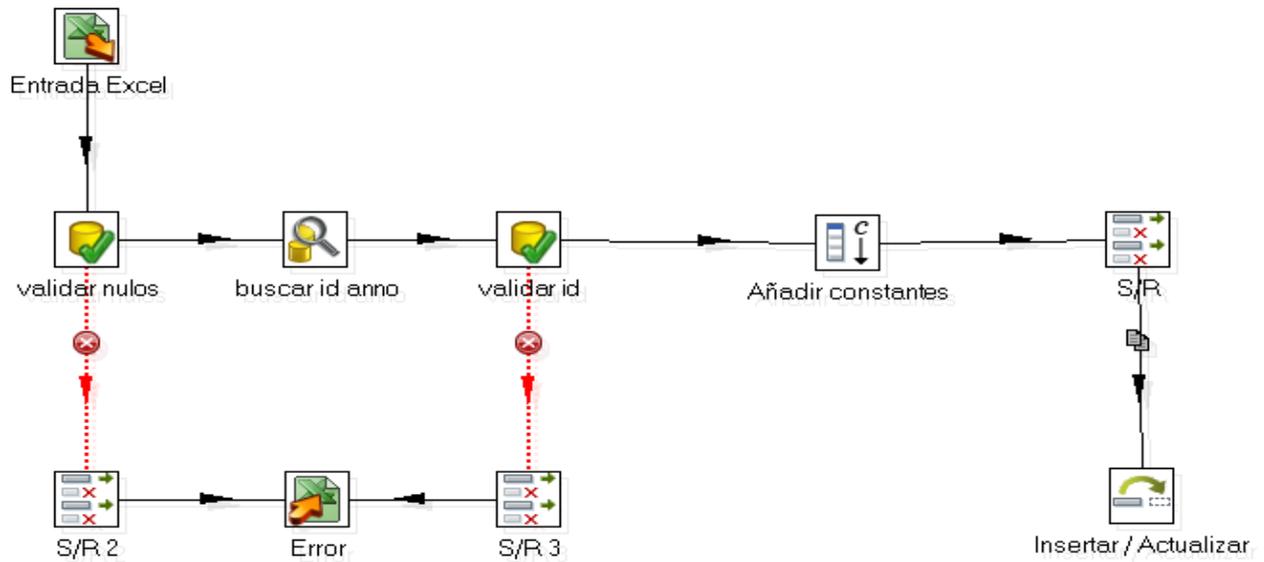


Figura 3.1 Carga de la tabla hech_indicadores_circulos

3.2.4 Implementación de los trabajos

Un trabajo o job es un conjunto de tareas que tienen como objetivo realizar una acción determinada. En los trabajos se utilizan pasos específicos que son distintos a los disponibles en las transformaciones. Permiten ejecutar una o varias transformaciones de las que han sido diseñadas, siguiendo una secuencia de ejecución para cada elemento que los conforman. Mediante los trabajos se definen el horario y frecuencia de la carga, así como el orden en que van a ser ejecutadas las transformaciones para poder realizar exitosamente la carga de los datos (25). En este caso el trabajo es programado para ejecutarse anualmente. Ver en la figura 3.2 el trabajo realizado al proceso de ETL para la implementación del mercado de datos Educación.

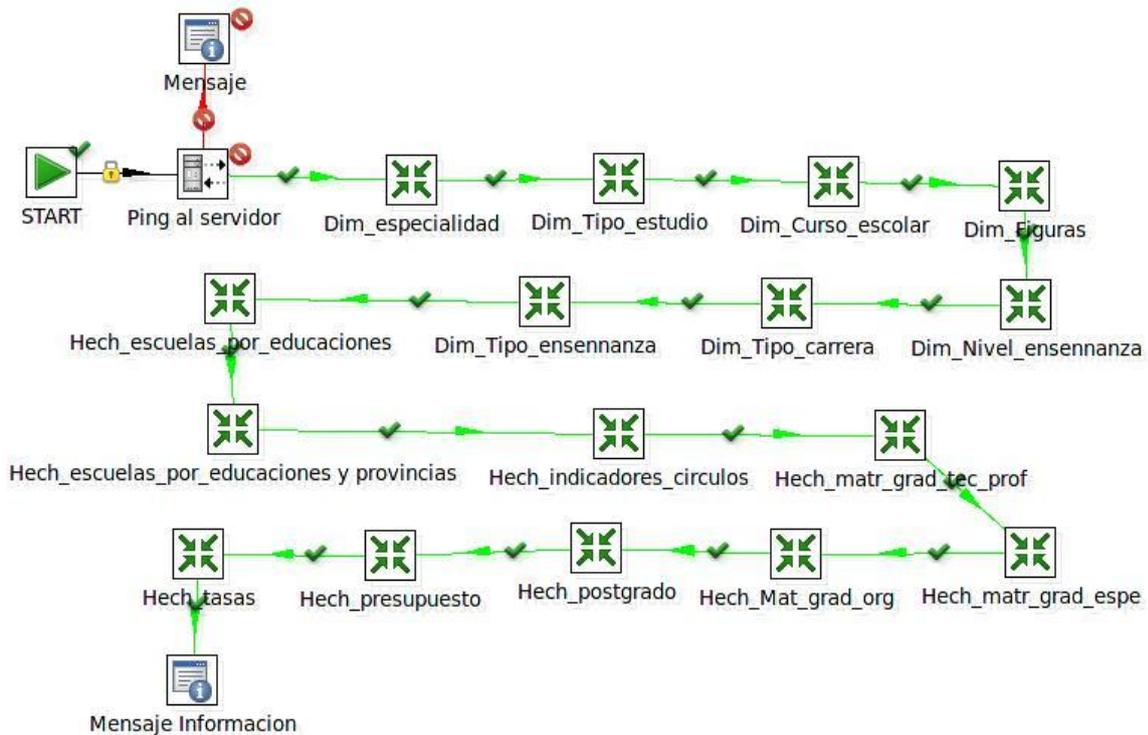


Figura 3.2 Trabajo para el proceso de ETL

3.3 Implementación del proceso de BI

3.3.1 Implementación de los cubos OLAP

Para la implementación de los cubos OLAP, es necesaria la creación de cubos multidimensionales, utilizando la herramienta Pentaho Schema Workbench. Esta permite generar un fichero xml donde se definen las dimensiones, los niveles de jerarquía, los hechos y la conexión con el almacén. En el mercado de datos del área de Educación se modelaron 9 cubos multidimensionales. (Ver en la figura 3.3)

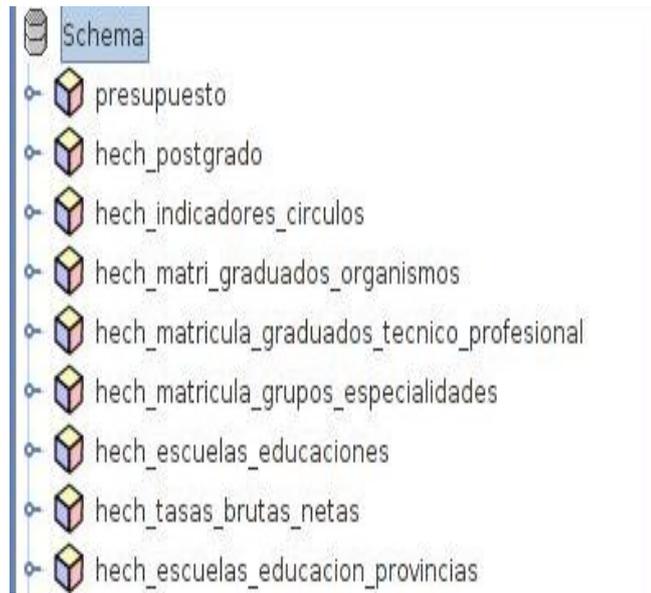


Figura 3.3 Diseño de los cubos utilizando Pentaho Schema Workbench.

En las figuras siguientes se muestran los niveles de jerarquía de las dimensiones.

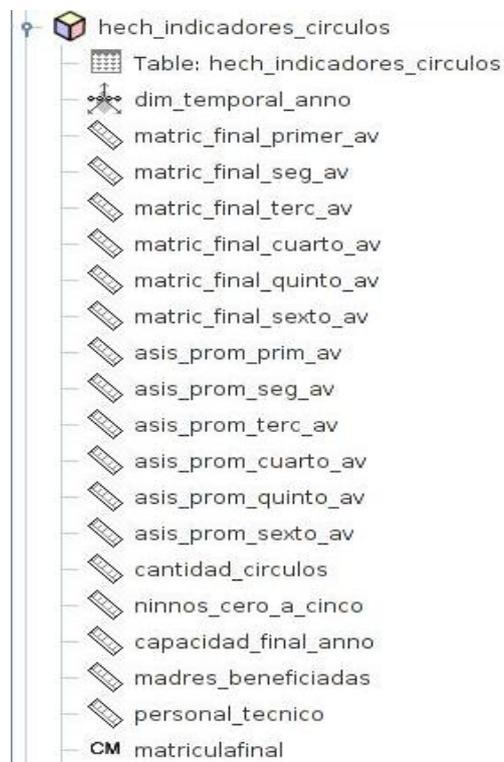


Figura 3.4 Elementos que componen el cubo Indicadores Círculos.

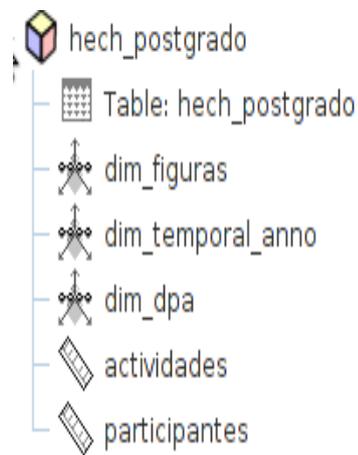


Figura 3.5 Elementos que componen el cubo Postgrado.



Figura 3.6 Elementos que componen el cubo Presupuesto.

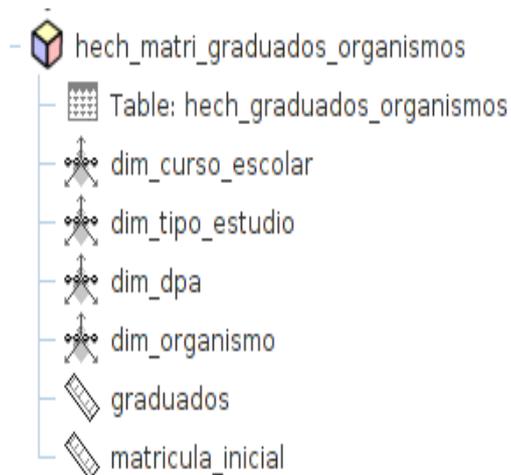


Figura 3.7 Elementos que componen el cubo Matrícula graduados por organismo.

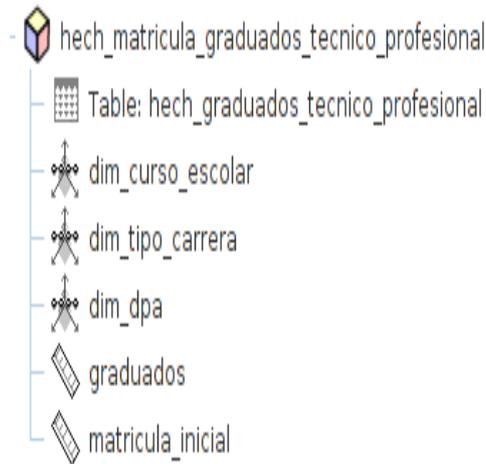


Figura 3.8 Elementos que componen el cubo Matrícula graduados técnica y profesional

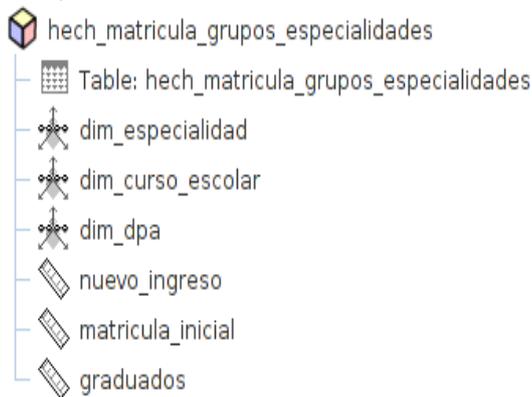


Figura 3.9 Elementos que componen el cubo Matrícula grupo especialidad.

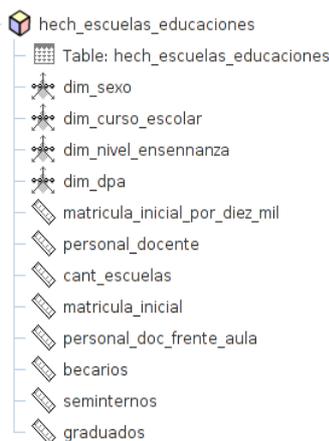


Figura 3.10 Elementos que componen el cubo de las escuelas por educación

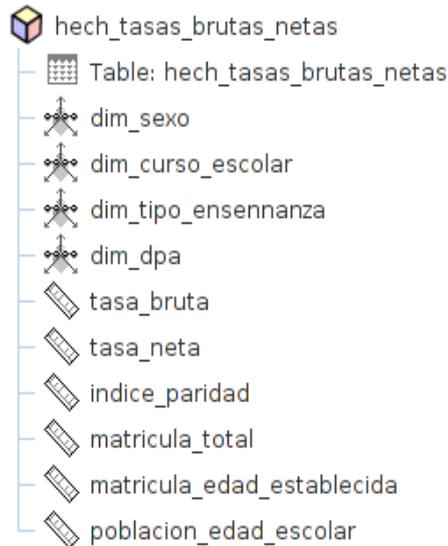


Figura 3.11 Elementos que componen el cubo tasas brutas y netas



Figura 3.12 Elementos que componen el cubo de las escuelas por educaciones y provincias

3.3.2 Arquitectura de información

A continuación se muestran los elementos que componen la estructura de navegación del mercado de datos Educación. Esta contiene un área de análisis, ocho libros de trabajos y 34 reportes asociados a estos. (Ver figura 3.13)

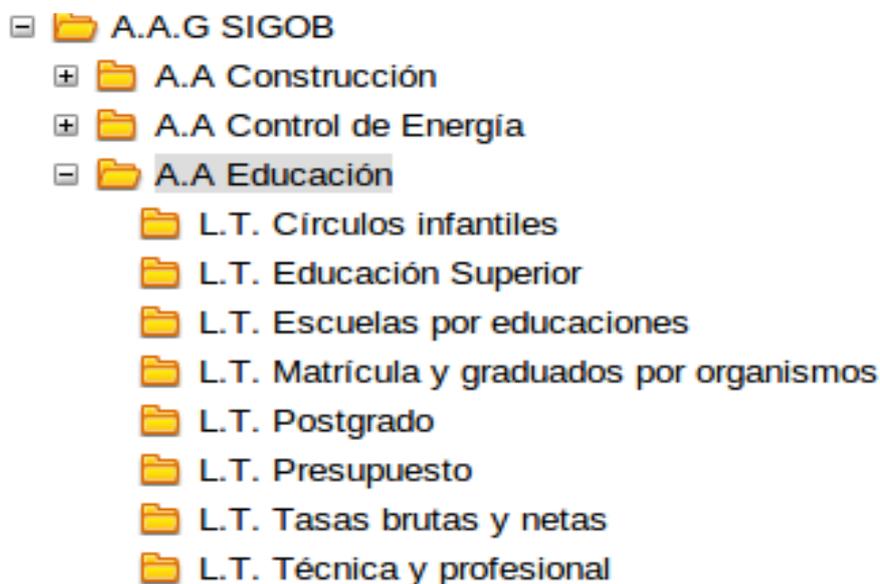


Figura 3.13 Arquitectura de información

3.3.3 Implementación de los reportes candidatos

Los libros de trabajo sirven para facilitarles el trabajo a los usuarios, pues contienen los reportes agrupados por temas afines.

Descripción de los Libros de Trabajo:

L.T Círculos infantiles: libro de trabajo contenido dentro del área de análisis Educación. Está compuesto por tres reportes que permiten realizar un análisis de datos correspondiente a los círculos infantiles.

L.T Escuelas por educaciones: libro de trabajo contenido dentro del área de análisis Educación. Está compuesto por 18 reportes que permiten realizar un análisis de datos correspondiente a las escuelas por educaciones.

L.T Técnica y profesional: libro de trabajo contenido dentro del área de análisis Educación. Está compuesto por dos reportes que permiten realizar un análisis de datos correspondiente a la enseñanza técnica y profesional.

L.T Matrícula grupo especialidades: libro de trabajo contenido dentro del área de análisis Educación. Está compuesto por cinco reportes que permiten realizar un análisis de datos correspondiente a la matrícula por grupo de especialidades.

L.T Tasas brutas y netas: libro de trabajo contenido dentro del área de análisis Educación. Está compuesto por un reporte que permite realizar un análisis de datos correspondiente a las tasas brutas y netas.

L.T Postgrado: libro de trabajo contenido dentro del área de análisis Educación. Está compuesto por un reporte que permite realizar un análisis de datos correspondiente a la educación de postgrado.

L.T Presupuesto: libro de trabajo contenido dentro del área de análisis Educación. Está compuesto por un reporte que permite realizar un análisis de datos correspondiente al presupuesto de educación.

L.T Graduados por organismos: libro de trabajo contenido dentro del área de análisis Educación. Está compuesto por dos reportes que permiten realizar un análisis de datos correspondiente a los graduados por organismos.

Reportes

La visualización de los reportes se realiza a través de consultas mdx, a continuación se muestra un ejemplo de estas para visualizar el reporte Indicadores generales de los círculos infantiles.

```
select NON EMPTY {[Measures].[cantidad_circulos], [Measures].[capacidad_final_anno],  
[Measures].[madres_beneficiadas], [Measures].[personal_tecnico], [Measures].[matriculafinal],  
[Measures].[asitenciapromedio]} ON COLUMNS, NON EMPTY  
Hierarchize({[dim_temporal_anno].[Todos]}) ON ROWS  
from [hech_indicadores_circulos]
```

Asistencia							
Nivel Territorial							
Nacional							
Tipo de enseñanza							
Círculos infantiles							
Medidas							
Año	● Asistencia promedio de primer año de vida (U)	● Asistencia promedio de segundo año de vida (U)	● Asistencia promedio de tercer año de vida (U)	● Asistencia promedio de cuarto año de vida (U)	● Asistencia promedio de quinto año de vida (U)	● Asistencia promedio de sexto año de vida (U)	● Asistencia promedio
1970	2.989	2.862	5.914	8.056	7.991	5.907	33.719
1971	2.737	2.621	5.431	7.366	7.250	5.274	30.679
1972	3.108	2.689	5.602	7.540	7.592	5.339	31.870
1973	3.018	3.087	6.450	8.652	7.900	5.523	34.630
1974	2.794	3.290	6.894	9.211	8.499	5.945	36.633
1975	2.901	3.616	7.581	10.114	9.538	6.611	40.361
1976	3.138	3.610	7.574	10.099	10.841	7.050	42.312
1977	2.781	5.800	7.730	9.662	10.937	8.047	44.957
1978	3.346	6.401	9.036	10.396	10.786	7.302	47.267

Figura 3.14 Reporte Indicadores Generales de los Círculos Infantiles.

3.4 Pruebas

3.4.1 Validación y prueba

Las pruebas son una medida de la calidad del software. Para determinar la calidad de un producto de software se deben desarrollar actividades que permitan comprobar el grado de cumplimiento de las especificaciones iniciales del sistema, y es aquí donde las pruebas de software desempeñan un papel fundamental. Estas verifican el desarrollo que va alcanzando el producto durante todas sus etapas, identificando posibles fallos de implementación, calidad, o usabilidad de un programa.

3.4.2 Modelo V

El Modelo V está relacionado con el modelo en cascada, ya que es una evolución del mismo, pero este modelo considera las actividades de prueba como un proceso que corre en paralelo con las actividades de análisis y diseño, en lugar de establecer una fase independiente al final del proyecto como lo hace el modelo en cascada (26). A continuación se muestra una imagen de como se aplica el Modelo V.

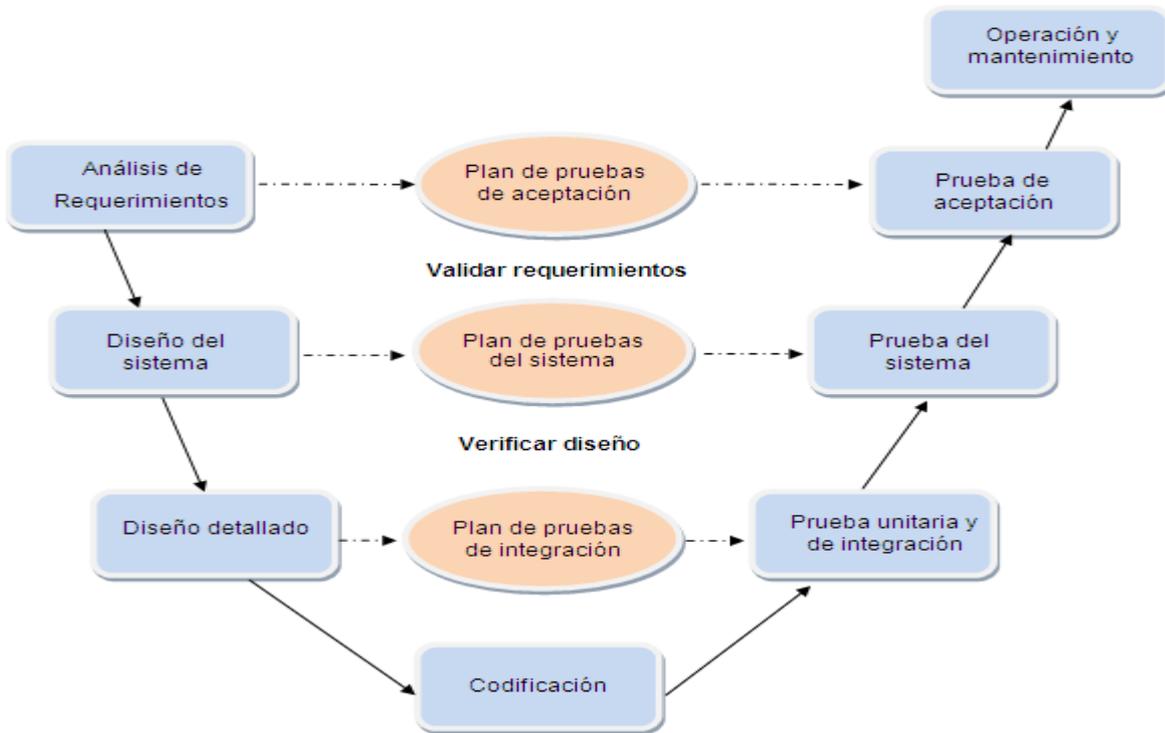


Figura 3.15 Modelo V

Para que un producto de software obtenga una buena calidad debe cumplir con las siguientes características:

- **Mantenibilidad:** capacidad del producto de software de ser modificado. Las modificaciones pueden incluir las correcciones, mejoras o adaptaciones del software a cambios en el ambiente, así como en los requisitos y las especificaciones funcionales.
- **Funcionalidad:** es la capacidad del software para proporcionar funciones que satisfacen las necesidades declaradas e implícitas cuándo el software se usa bajo las condiciones especificadas.
- **Portabilidad:** capacidad del producto de software de ser transferido de un ambiente a otro.
- **Confiabilidad:** la capacidad del producto de software para mantener un nivel de ejecución especificado cuando se usa bajo las condiciones especificadas.
- **Eficiencia:** capacidad del producto de software para proporcionar una ejecución o desempeño apropiado, en relación con la cantidad de recursos utilizados, bajo condiciones establecidas.

CAPÍTULO 3: IMPLEMENTACIÓN Y PRUEBA

- **Usabilidad:** capacidad del producto de software de ser comprendido, aprendido, utilizado y de ser atractivo para el usuario, cuando se utilice bajo las condiciones especificadas (27).

Existen diversas pruebas aplicables al mercado de datos como es el caso de las listas de chequeos, pruebas de interfaz y las pruebas de aceptación. En el mercado de datos del área de Educación se diseñaron los casos de pruebas de interfaz, a continuación se muestra un ejemplo para el caso de uso de información "Analizar información de los círculos infantiles". Para consultar los demás diseños de casos de prueba remitirse al Expediente de proyecto.

Tabla 3.2 Diseño de caso de prueba "Analizar información de los círculos infantiles".

Escenario	Descripción	Variables		Respuesta del sistema	Flujo central
		Perfiles de análisis	Indicadores a medir		
EC 1.1: Asistencia	Permite visualizar el reporte con las variables presentes en el mismo.	<ul style="list-style-type: none"> • Año. • Nivel Territorial • Tipo enseñanza 	<ul style="list-style-type: none"> • Asistencia promedio de primer año de vida. • Asistencia promedio de segundo año de vida. • Asistencia promedio de tercer año de vida. • Asistencia promedio de cuarto año de vida. • Asistencia promedio de quinto año de 	Se muestra la tabla con los valores correspondientes a cada escenario.	Se abre la aplicación. Se autentica. Se entra al sistema. Se despliega hacia la derecha el componente ubicado en el lateral izquierdo que contiene el navegador. Se selecciona el área de análisis de A.A Educación. Se selecciona el

CAPÍTULO 3: IMPLEMENTACIÓN Y PRUEBA

			<p>vida.</p> <ul style="list-style-type: none"> • Asistencia promedio de sexto año de vida. • Asistencia promedio 		<p>libro de trabajo</p> <p>L.T Círculos infantiles.</p> <p>En la parte inferior izquierda se selecciona el reporte deseado.</p> <p>En el área de trabajo se visualiza la tabla correspondiente al reporte.</p>
EC 1.2: Indicadores Generales	Permite visualizar el reporte con las variables presentes en el mismo.	<ul style="list-style-type: none"> • Año. • Nivel Territorial • Tipo enseñanza 	<ul style="list-style-type: none"> • Cantidad de círculos • Capacidad final del año • Madres beneficiadas • Personal técnico • Matrícula final • Asistencia promedio 		
EC 1.3: Matrícula	Permite visualizar el reporte con las variables presentes en el mismo	<ul style="list-style-type: none"> • Año • Nivel Territorial • Tipo de enseñanza 	<ul style="list-style-type: none"> • Matrícula final • Matrícula final de primer año de vida • Matrícula final de segundo año de vida • Matrícula final de tercer año de vida • Matrícula final de cuarto año de vida • Matrícula final de quinto año de vida 		

			<ul style="list-style-type: none"> • Matrícula final de sexto año de vida • Cantidad de niños de 0-5 años 		
--	--	--	---	--	--

Luego de aplicados los casos de pruebas a nivel interno con dos iteraciones (departamental y a nivel del centro) arrojaron los resultados que se muestran a continuación:

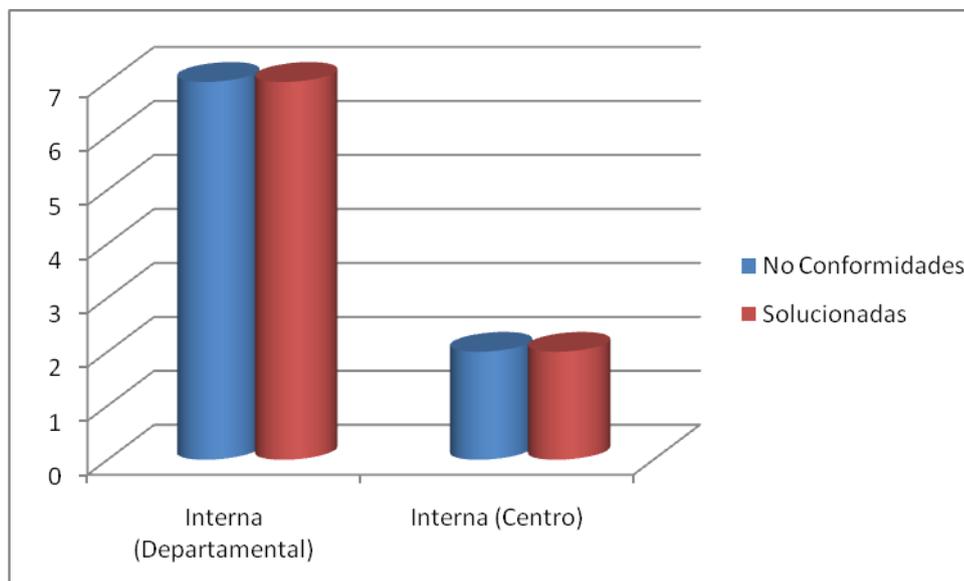


Figura 3.16 Resultados de los casos de prueba

3.4.2 Lista de chequeo

La lista de chequeo definida para la evaluación del mercado de datos está dividida en tres secciones: Estructura del documento, Indicadores definidos en el desarrollo y Semántica del documento; en los cuales se analizan 14 indicadores, de ellos ocho críticos (Ver Tabla 3.3).

Tabla 3.3 Lista de chequeo.

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

CAPÍTULO 3: IMPLEMENTACIÓN Y PRUEBA

				elementos afectados	
Crítico	1. ¿Los artefactos presentan las secciones obligatorias establecidas para el expediente de proyecto?	0		0	
Indicadores definidos en el desarrollo					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
	1. ¿El lenguaje utilizado contiene sentencias que son expresables mediante una sintaxis bien definida?	0		0	
Crítico	2. ¿Los reportes se configuran a través de la interfaz del sistema?	0		0	
	3. ¿La interfaz facilita el uso de las funciones del sistema por parte de los usuarios?	0		0	
Crítico	4. ¿No existen restricciones para construir cubos OLAP con dimensiones y niveles de agregación ilimitados?	0		0	
Crítico	5. ¿Los reportes pueden ser modificados dependiendo de las necesidades de los usuarios?	0		0	
	6. ¿Una vez que el usuario solicita la información el sistema responde con rapidez?	0		0	
	7. ¿El sistema refleja cualquier lógica del negocio para poder responder a preguntas específicas?	0		0	

CAPÍTULO 3: IMPLEMENTACIÓN Y PRUEBA

Crítico	8. ¿En dependencia del rol asignado al usuario, el sistema es capaz de garantizar la confidencialidad y seguridad de acceso a los datos?	0		0	
	9. ¿La toma de decisiones en la institución se apoya en los datos e información derivados del proceso de análisis?	0		0	
Crítico	10. ¿Los cambios en los datos se reflejan automáticamente en los reportes de forma instantánea?	0		0	
Semántica del documento					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
Crítico	1. ¿Se han identificado errores ortográficos en los entregables?	0		0	
Crítico	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?	0		0	

La imagen mostrada a continuación refleja un gráfico de barras en el cual aparece el comportamiento de los indicadores definidos para la lista de chequeo elaborada. De forma general se identificaron 14 indicadores, de ellos ocho críticos y luego de aplicada la lista de chequeo se generaron cero no conformidades.

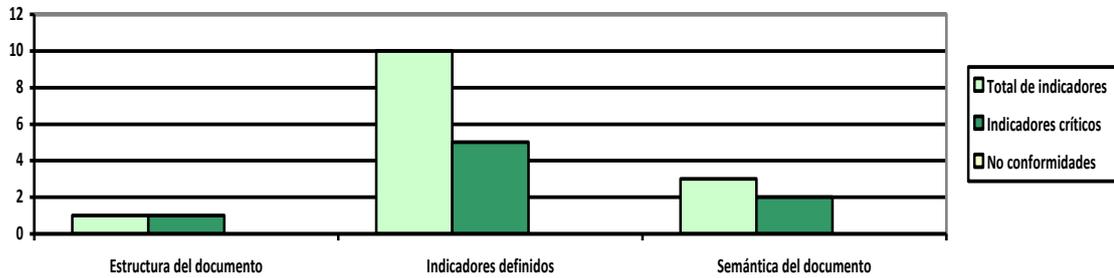


Figura 3.17 Comportamiento de los indicadores por secciones.

3.4.3 Pruebas de aceptación

Una vez finalizado el desarrollo del mercado de datos del área de Educación se realizaron las pruebas de aceptación; efectuadas por la especialista Elena Leonila Fernández García, quien es la representante de la ONE en la universidad. Los resultados arrojados luego de ejecutar dichas pruebas fueron satisfactorios, ya que se confirmó que el mercado cumple con todas las necesidades de información identificadas con anterioridad. Dichos resultados quedaron oficializados con la carta de aceptación del cliente donde especifica que el mercado cumple con los requisitos especificados por este.

Conclusiones

En este capítulo se describieron los elementos de implementación para la construcción del mercado de datos Educación. Se implementó el proceso ETL, extrayendo los datos de las series de Educación, estos fueron transformados y cargados. Quedó definida la estructura de los datos a partir del modelo de datos físico, contando con dos esquemas: “dimensiones” y “mart_educacion”, compuesto por 12 dimensiones y nueve hechos. Se diseñaron e implementaron los cubos OLAP, quedando definidos nueve cubos, 10 dimensiones y 36 medidas. Se desarrolló el subsistema de visualización, determinándose ocho libros de trabajo, con un total de 34 vistas de análisis. Se aplicó la lista de chequeo, generando cero no conformidades. También se realizaron los casos de pruebas, generándose siete no conformidades, las que quedaron resueltas. Además, se efectuó la prueba de aceptación del cliente, quedando el mercado de datos avalado.

CONCLUSIONES

Con la realización de este trabajo se cumplieron los objetivos propuestos:

- Se realizó el análisis y diseño del mercado de datos del área Educación, identificándose nueve requisitos funcionales y 30 no funcionales, así como 12 dimensiones, nueve hechos y 36 medidas. Además, se estructuró el modelo físico de la base de datos y quedaron definidos y descritos 13 casos de usos del sistema.
- Se implementó el mercado de datos del área Educación, definiéndose anteriormente la estructura de los datos a partir del modelo de datos físico, contando con dos esquemas: “dimensiones” y “mart_educacion”. Luego se llevó a cabo todo el proceso de ETL y BI, donde se diseñaron e implementaron los cubos OLAP, quedando definidos nueve cubos. Se desarrolló el subsistema de visualización, determinándose ocho libros de trabajo, con un total de 34 vistas de análisis.
- Se validó el mercado de datos del área Educación aplicando las listas de chequeos, casos de pruebas y pruebas de aceptación del cliente.

RECOMENDACIONES

- Que se ponga en explotación el mercado de datos en las oficinas de la ONE distribuidas por todo el país, para probar las funcionalidades del mismo y detectar nuevas funcionalidades.
- Elaborar una estrategia para cargar los datos que quedaron registrados en los ficheros de errores una vez que los especialistas los hayan revisado.

BIBLIOGRAFÍA

- Autores, Colectivo de. 2009. "METODOLOGÍA PARA EL DESARROLLO DE SOLUCIONES DE ALMACENES DE DATOS E INTELIGENCIA DE NEGOCIO EN CENTALAD". 2009.
- Características del Almacén de Datos. [En línea] [Citado el: 20 de 10 de 2010.] <http://hp.fciencias.unam.mx/~alg/bd/dwh.pdf>.
- **C, Casares.** Programación. [En línea] [Citado el: 10 de 11 de 2010.] http://www.programacion.com/bbdd/tutorial/warehouse/15/#warehousing_desarrollo_confi.
- Esquema Mondrian Workbench. [En línea] [Citado el: 15 de 11 de 2010.] http://mondrian.pentaho.com/documentation/schema_workbench.pdf.
- **Franco Capote, Rocío y Cedeño Pupo, Isabel.** *Análisis, Diseño e Implementación del mercado de datos Indicadores de la Educación en la Oficina Nacional de Estadísticas.* Ciudad Habana: s.n., 2010.
- gravitar [Online] [Citado el : 25 de 03 de , 2011.] <http://www.gravitar.biz/index.php/bi/introduccion-pentaho-parte-1/>
- gulsin [Online] [Citado el: 24 de 03 de , 2011.] <http://gulsin.org/2011/03/14/transformaciones-en-pentaho-kettle/>
- **H., Torres, y otros.** Bases de datos y data warehouse: Herramientas estratégicas para la eficacia comercial. [En línea] [Citado el: 29 de 11 de 2010.] http://www-lsi.ugr.es/~rosana/investigacion/bd_efsio4.pdf.
- **J., Martín, y otros.** Sistemas de soporte a la gestión del negocio. [Citado el: 27 de 11 de 2010.]
- **Kimball, Ralph y Ross, Margy.** 2002. *The Data Warehouse Toolkit. EUA . 2002.*
- La Suite de Pentaho. [En línea] [Citado el: 30 de 11 de 2010.] <http://demetole.blogspot.com/2010/05/la-suite-pentaho.html>.
- **Lefcovich, Mauricio León.** 2008. Estadística aplicada a los negocio. [En línea] 25 de 06 de 2008. [Citado el: 05 de 11 de 2010.] <http://manuelgross.bligoo.com/content/view/218974/La-estadistica-es-fundamental-para-la-gestion-eficiente.html>.
- **Leyva Osorio, Yolanda y Falcón Rodríguez, Reynaldo.** Mercado de Datos Estadístico de Inmigración y Extranjería para el Departamento de Turismo y Comercio de la Oficina Nacional de Estadísticas. 2010.

- One. [En línea] [Citado el: 15 de 10 de 2010.] <http://www.one.cu/queeslaone.htm>.
- Pentaho BI Plataform Server. [En línea] [Citado el: 01 de 12 de 2010.] <http://www.summan.com/index.php/productos/software/pentaho-.html>.
- rhernando. Almacenes de datos (Datawarehouse). [En línea] [Citado el: 25 de 10 de 2010.] <http://www.rhernando.net/modules/tutorials/doc/bd/dw.html>.
- **R, Kimball y Dirty Data.** [En línea] [Citado el: 20 de 11 de 2010.] <http://www.dbmsmag.com/9609d14.html>.
- **S., Hussain y J., Beg.** Data Quality: A Problem and an Approach. [En línea] [Citado el: 15 de 11 de 2010.] <http://doc.advisor.com/doc/13060>.
- Sitio de PgAdmin. [En línea] [Citado el: 25 de 11 de 2010.] <http://www.pgadmin.org/>.
- Sitio de Postgres. [En línea] [Citado el: 23 de 11 de 2010.] <http://www.postgresql.org/>.
- Visual Paradigm for UML (ME). [En línea] [Citado el: 25 de 11 de 2010.] [http://www.freedownloadmanager.org/es/downloads/Paradigma_Visual_para_UML_\(MI\)_1470_p/](http://www.freedownloadmanager.org/es/downloads/Paradigma_Visual_para_UML_(MI)_1470_p/).

REFERENCIAS BIBLIOGRÁFICAS

1. Oficina Nacional de Estadística. *Oficina Nacional de Estadística*. [Online] [Cited: octubre 2010, 20.] <http://www.one.cu/queeslaone.htm>.
2. www.rhernando.net . *www.rhernando.net* . [Online] 2001. [Cited: octubre 25, 2010.] <http://www.rhernando.net/modules/tutorials/viewexttutorial.php?tid=41>.
3. **Casales Cabrera, María Evelia**. Facultad de Ciencias. *Facultad de Ciencias*. [Online] 2009. [Cited: octubre 20, 2010.] <http://hp.fciencias.unam.mx/~alg/bd/dwh.pdf>.
4. ECURED. *ECURED*. [En línea] http://www.ecured.cu/index.php/Almac%C3%A9n_de_Datos.
5. **Kimball, Ralph and Data, Dirty**. [Online] [Cited: noviembre 20, 2010.] <http://www.dbmsmag.com/9609d14.html>.
6. Sinnexus. *Sinnexus*. [Online] http://www.sinnexus.com/business_intelligence/datamart.aspx.
7. *METODOLOGÍA PARA EL DESARROLLO DE SOLUCIONES DE ALMACENES DE DATOS E INTELIGENCIA DE NEGOCIO EN CENTALAD*. **autores, Colectivo de**. 2009.
8. Fases Construcción Mercados Datos. *Fases Construcción Mercados Datos*. [Online] [Cited: diciembre 1, 2010.] <http://www.mitecnologico.com/Main/FasesConstruccionMercadosDatos>.
9. **Fernando Berzal**. El modelo multidimensional. Data warehousing. *El modelo multidimensional. Data warehousing*. [Online] [Cited: noviembre 8, 2010.] http://www.google.com/url?sa=t&source=web&cd=2&ved=0CCAQFjAB&url=http%3A%2F%2Felix.ugr.es%2Fidbis%2Fdb%2Fdocs%2Fintro%2FF%2520Modelo%2520multidimensional.pdf&ei=39JrTd_NJ8mitgf9pIXnAg&usq=AFQjCNH2UCmEinC1Ce5JyNWlKfByAasFFw&sig2=w0DjF8Ow0llz_LPSmMD7Qw.
10. El Rincón del BI. *El Rincón del BI*. [Online] [Cited: noviembre 2010, 7.] <http://churriwifi.wordpress.com/2010/01/24/14-2-diseno-de-hechos-y-atributos-microstrategy/>.
11. Modelos Emergentes de Bases de Datos . *Modelos Emergentes de Bases de Datos* . [Online] [Cited: octubre 2010, 28.] <http://www.tutoriales.itsa.edu.mx/TopicosAvanzadosBD/index.php?mod=cubos&ban=0>.
12. **Ibarra, María de los Ángeles**. OLAP. *OLAP*. [Online] <http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/OLAPMonog.pdf>.
13. **Martínez Orol, Alfredo**. GestioPolis.com. *GestioPolis.com*. [Online] [Cited: octubre 2010, 25.] <http://www.gestiopolis.com/canales8/ger/olap-online-analytic-processing.htm>.
14. DataPrix. [Online] <http://www.dataprix.com/32-oltp>.
15. Sistema Gestor de Bases de Datos. *Sistema Gestor de Bases de Datos*. [Online] [Cited: noviembre 2010, 20.] <http://www.desarrolloweb.com/articulos/sistemas-gestores-bases-datos.html>.
16. Visual Paradigm for UML (ME). *Visual Paradigm for UML (ME)*. [Online] [Cited: noviembre 2010, 25.] [http://www.freedownloadmanager.org/es/downloads/Paradigma_Visual_para_UML_\(MI\)_1470_p/](http://www.freedownloadmanager.org/es/downloads/Paradigma_Visual_para_UML_(MI)_1470_p/).
17. Postgres. *Postgres*. [Online] [Cited: noviembre 2010, 23.] <http://www.postgresql.org/>.

REFERENCIAS BIBLIOGRÁFICAS

18. PgAdmin. *PgAdmin*. [Online] [Cited: noviembre 2010, 25.] <http://www.pgadmin.org/>.
19. PENTAHO. *PENTAHO*. [Online] [Cited: noviembre 2010, 5.] <http://www.gravitar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/>.
20. Esquema Mondrian Workbench. *Esquema Mondrian Workbench*. [Online] [Cited: noviembre 2010, 15.] http://mondrian.pentaho.com/documentation/schema_workbench.pdf.
21. PENTAHO: Creación de un cubo con Mondrian. *PENTAHO: Creación de un cubo con Mondrian*. [Online] [Cited: octubre 2010, 18.] <http://www.gravitar.biz/index.php/bi/pentaho-ejemplo-cubo-mondrian/>.
22. Pentaho BI Plataform Server. *Pentaho BI Plataform Server*. [Online] [Cited: diciembre 2010, 1.] <http://www.summan.com/index.php/productos/software/pentaho-.html>.
23. El servidor Apache Tomcat. *El servidor Apache Tomcat*. [Online] [Cited: noviembre 2010, 15.] <http://casidiablo.net/el-servidor-apache-tomcat/>.
24. Gulsin. *Gulsin*. [Online] [Cited: marzo 2011, 24.] <http://gulsin.org/2011/03/14/transformaciones-en-pentaho-kettle/>.
25. Gravatar. *Gravatar*. [Online] [Cited: marzo 2011, 25.] <http://www.gravitar.biz/index.php/bi/introduccion-pentaho-parte-1/>.
26. Pérez Lamancha, Beatriz. Proceso de Testing funcional independientemente. Tesis de Maestría en Informática. Proceso de Testing funcional independientemente. Tesis de Maestría en Informática. [Online] 2006. [Cited: Junio 05, 2011.] <http://www.fing.edu.uy/~bperez/.../Tesis%20-%20Beatriz%20Perez-%202006.pdf>.
27. Norma Cubana. Ingeniería de Software. Calidad del producto. *Norma Cubana. Ingeniería de Software. Calidad del producto*. [Online] abril 2005. <http://calisoft.uci.cu/tmp/documentos/normas/iso/NC-ISO-IEC%209126-1.pdf>.

ANEXOS

Descripción de los casos de uso de Información.

Caso de Uso:	Analizar información de las escuelas por educaciones.
Tipo:	Información
Actores:	Analista, Administrador
Resumen:	El CU inicia cuando el actor entra al sistema, luego este selecciona el reporte de Escuelas., y cuando el sistema muestra el reporte termina el CU.
Precondiciones:	Compleitud del almacén. Carga de los datos.
Referencias	Ver RI8, RI9, RI10, RI11.
Prioridad	Crítico.
Flujo Normal de Eventos	
Acción del Actor	Respuesta del Sistema
1. El actor entra al sistema	2 El sistema muestra las áreas de análisis existente.
3. El actor selecciona el AA Educación.	4 El sistema muestra los libros de trabajos que contiene el AA Educación.
5. El actor selecciona el LT Escuelas por educación.	6. El sistema muestra los reportes pertenecientes al LT Escuelas por educación.
7. El actor selecciona el reporte deseado.	8. El sistema muestra la información contenida en el reporte seleccionado y brinda la posibilidad de analizar el reporte desde diferentes perspectivas. Ir al CU: Visualizar reporte.
Opciones de reportes de Analizar información de las escuelas.	

Perspectivas de análisis	Posibles resultados	
	Medidas	Periodicidad
Variables de entrada relacionadas con el CU Analizar información de las escuelas: <ul style="list-style-type: none"> • Sexo. • Curso • Nivel Enseñanza. • Provincia • Tipo de Enseñanza. 	Variables de salida disponibles en el caso de uso Analizar información de las escuelas: <ul style="list-style-type: none"> • Escuelas. • Personal docente frente al aula. • Graduados. • Matrícula inicial de las escuelas. • Seminternos • Becarios • Personal docente • Matrícula inicial por diez mil. 	Rango de tiempo en que se solicitan las variables de salida: <ul style="list-style-type: none"> • Anual.
Pos condiciones	Disponibilidad de opciones (cruces de variables) de reportes relacionados con el caso de uso Analizar información de las escuelas.	

Caso de Uso:	Analizar información del presupuesto de educación
Tipo:	Información
Actores:	Analista, Administrador
Resumen:	El CU inicia cuando el actor entra al sistema, luego este selecciona el reporte del presupuesto de educación y cuando el sistema muestra el reporte termina el CU.
Precondiciones:	Compleitud del almacén. Carga de los datos.
Referencias	Ver RI12, RF13.

Prioridad	Crítico.	
Flujo Normal de Eventos		
Acción del Actor	Respuesta del Sistema	
1. El actor entra al sistema	2 El sistema muestra las áreas de análisis existente.	
3. El actor selecciona el AA Educación.	4 El sistema muestra los libros de trabajos que contiene el AA Educación.	
5. El actor selecciona el LT Presupuesto.	6. El sistema muestra los reportes pertenecientes al LT Presupuesto .	
7. El actor selecciona el reporte deseado.	8. El sistema muestra la información contenida en el reporte seleccionado y brinda la posibilidad de analizar el reporte desde diferentes perspectivas. Ir al CU: Visualizar reporte.	
Opciones de reportes de Analizar información del presupuesto de educación		
Perspectivas de análisis	Posibles resultados	
	Medidas	Periodicidad
Variables de entrada relacionadas con el hecho Analizar información del presupuesto de educación: <ul style="list-style-type: none"> • Año. • Nivel Territorial. • Organismos. 	Variables de salida disponibles en el caso de uso Analizar información del presupuesto de educación: <ul style="list-style-type: none"> • Gasto público • Percápita por año. 	Rango de tiempo en que se solicitan las variables de salida: <ul style="list-style-type: none"> • Anual.
Pos condiciones	Disponibilidad de opciones (cruces de variables) de reportes relacionados con el caso de uso Analizar información del presupuesto de educación.	

Caso de Uso:	Analizar información de las escuelas por educaciones.
Tipo:	Información
Actores:	Analista, Administrador

Resumen:	El CU inicia cuando el actor entra al sistema, luego este selecciona el reporte de Escuelas., y cuando el sistema muestra el reporte termina el CU.
Precondiciones:	Completitud del almacén. Carga de los datos.
Referencias	Ver RI8, RI9, RI10, RI11.
Prioridad	Crítico.

Flujo Normal de Eventos

Acción del Actor	Respuesta del Sistema
1. El actor entra al sistema	2 El sistema muestra las áreas de análisis existente.
3. El actor selecciona el AA Educación.	4 El sistema muestra los libros de trabajos que contiene el AA Educación.
5. El actor selecciona el LT Escuelas por educación.	6. El sistema muestra los reportes pertenecientes al LT Escuelas por educación.
7. El actor selecciona el reporte deseado.	8. El sistema muestra la información contenida en el reporte seleccionado y brinda la posibilidad de analizar el reporte desde diferentes perspectivas. Ir al CU: Visualizar reporte.

Opciones de reportes de Analizar información de las escuelas.

Perspectivas de análisis	Posibles resultados	
	Medidas	Periodicidad
Variables de entrada relacionadas con el CU Analizar información de las escuelas: <ul style="list-style-type: none"> • Sexo. • Curso • Nivel Enseñanza. • Provincia • Tipo de 	Variables de salida disponibles en el caso de uso Analizar información de las escuelas: <ul style="list-style-type: none"> • Escuelas. • Personal docente frente al aula. • Graduados. • Matrícula inicial de las escuelas. • Seminternos • Becarios • Personal docente 	Rango de tiempo en que se solicitan las variables de salida: <ul style="list-style-type: none"> • Anual.

Enseñanza.	<ul style="list-style-type: none"> • Matrícula inicial por diez mil. 	
Pos condiciones	Disponibilidad de opciones (cruces de variables) de reportes relacionados con el caso de uso Analizar información de las escuelas.	

Caso de Uso:	Analizar información de las escuelas por educaciones.
Tipo:	Información
Actores:	Analista, Administrador
Resumen:	El CU inicia cuando el actor entra al sistema, luego este selecciona el reporte de Escuelas., y cuando el sistema muestra el reporte termina el CU.
Precondiciones:	Compleitud del almacén. Carga de los datos.
Referencias	Ver RI8, RI9, RI10, RI11.
Prioridad	Crítico.

Flujo Normal de Eventos

Acción del Actor	Respuesta del Sistema
1. El actor entra al sistema	2 El sistema muestra las áreas de análisis existente.
3. El actor selecciona el AA Educación.	4 El sistema muestra los libros de trabajos que contiene el AA Educación.
5. El actor selecciona el LT Escuelas por educación.	6. El sistema muestra los reportes pertenecientes al LT Escuelas por educación.
7. El actor selecciona el reporte deseado.	8. El sistema muestra la información contenida en el reporte seleccionado y brinda la posibilidad de analizar el reporte desde diferentes perspectivas. Ir al CU: Visualizar reporte.

Opciones de reportes de Analizar información de las escuelas.

Perspectivas de análisis	Posibles resultados	
	Medidas	Periodicidad

<p>VARIABLES DE ENTRADA relacionadas con el CU</p> <p>Analizar información de las escuelas:</p> <ul style="list-style-type: none"> • Sexo. • Curso • Nivel Enseñanza. • Provincia • Tipo de Enseñanza. 	<p>VARIABLES DE SALIDA disponibles en el caso de uso</p> <p>Analizar información de las escuelas:</p> <ul style="list-style-type: none"> • Escuelas. • Personal docente frente al aula. • Graduados. • Matrícula inicial de las escuelas. • Seminternos • Becarios • Personal docente • Matrícula inicial por diez mil. 	<p>Rango de tiempo en que se solicitan las variables de salida:</p> <ul style="list-style-type: none"> • Anual.
Pos condiciones	Disponibilidad de opciones (cruces de variables) de reportes relacionados con el caso de uso Analizar información de las escuelas.	

Caso de Uso:	Analizar información de las escuelas por educación.
Tipo:	Información.
Actores:	Analista, Administrador.
Resumen:	El CU inicia cuando el actor entra al sistema, luego este selecciona el reporte la enseñanza primaria y media y cuando el sistema muestra el reporte termina el CU.
Precondiciones:	Complejidad del almacén. Carga de los datos.
Referencias	Ver RI14, RI15, RI16, RI 17, RI 18, RI 19, RI20, RI21
Prioridad	Crítico.

Flujo Normal de Eventos

Acción del Actor	Respuesta del Sistema
1. El actor entra al sistema	2 El sistema muestra las áreas de análisis existente.
3. El actor selecciona el AA Educación.	4 El sistema muestra los libros de trabajos que

	contiene el AA Educación.
5. El actor selecciona el LT escuelas por educación.	6. El sistema muestra los reportes pertenecientes al LT escuelas por educación .
7. El actor selecciona el reporte deseado.	8. El sistema muestra la información contenida en el reporte seleccionado y brinda la posibilidad de analizar el reporte desde diferentes perspectivas. Ir al CU: Visualizar reporte.

Opciones de reportes de Analizar información de las enseñanzas primaria y media

Perspectivas de análisis	Posibles resultados	
	Medidas	Periodicidad
Variables de entrada relacionadas con el CU Analizar información de la enseñanza primaria y media. <ul style="list-style-type: none"> • Sexo. • Curso. • Nivel Enseñanza. • Tipo carrera. • Nivel Territorial. 	Variables de salida disponibles en el caso de uso Analizar información de las enseñanzas primaria y media: <ul style="list-style-type: none"> • Matrícula inicial por diez mil. • Personal docente. • Escuelas. • Matrícula inicial de las escuelas. • Personal docente frente al aula. • Becarios. • Seminternos. • Graduados 	Rango de tiempo en que se solicitan las variables de salida: <ul style="list-style-type: none"> • Anual.
Pos condiciones	Disponibilidad de opciones (cruces de variables) de reportes relacionados con el caso de Analizar información de la enseñanza primaria y media.	

Caso de Uso:	Analizar información de tasas e índices
---------------------	---

Tipo:	Información
Actores:	Analista, Administrador
Resumen:	El CU inicia cuando el actor entra al sistema, luego este selecciona el reporte de tasas e índices y cuando el sistema muestra el reporte termina el CU.
Precondiciones:	Compleitud del almacén. Carga de los datos.
Referencias	Ver RI22, RI23, RI24, RI25.RI26, RI27.
Prioridad	Crítico.

Flujo Normal de Eventos

Acción del Actor	Respuesta del Sistema
1. El actor entra al sistema	2 El sistema muestra las áreas de análisis existente.
3. El actor selecciona el AA Educación.	4 El sistema muestra los libros de trabajos que contiene el AA Educación.
5. El actor selecciona el LT Tasa brutas y netas.	6. El sistema muestra los reportes pertenecientes al LT Tasa brutas y netas.
7. El actor selecciona el reporte deseado.	8. El sistema muestra la información contenida en el reporte seleccionado y brinda la posibilidad de analizar el reporte desde diferentes perspectivas. Ir al CU: Visualizar reporte.

Opciones de reportes de Analizar información de tasas e índices

Perspectivas de análisis	Posibles resultados	
	Medidas	Periodicidad
Variables de entrada relacionadas con el CU Analizar información de tasas e índices: <ul style="list-style-type: none"> • Sexo. • Curso. • Tipo enseñanza. • Nivel Territorial. 	Variables de salida disponibles en el caso de uso Analizar información de tasas e índices: <ul style="list-style-type: none"> • Tasas brutas. • Tasas netas. • índice de paridad. • Matrícula total de 	Rango de tiempo en que se solicitan las variables de salida: <ul style="list-style-type: none"> • Anual.

	<p>las tasas brutas y netas.</p> <ul style="list-style-type: none"> • Matrícula de edad establecida de las tasas brutas y netas. • Población en edad escolar de las tasas brutas y netas. 	
Pos condiciones	Disponibilidad de opciones (cruces de variables) de reportes relacionados con el caso de uso Analizar información de tasas e índices.	

Caso de Uso:	Analizar información de la enseñanza superior.
Tipo:	Información
Actores:	Analista, Administrador
Resumen:	El CU inicia cuando el actor entra al sistema, luego este selecciona el reporte de enseñanza superior y cuando el sistema muestra el reporte termina el CU.
Precondiciones:	Compleitud del almacén. Carga de los Datos.
Referencias	Ver RI28, RI29, RI30, RI31, RI32, RI33, RI34.
Prioridad	Crítico.

Flujo Normal de Eventos

Acción del Actor	Respuesta del Sistema
1. El actor entra al sistema	2 El sistema muestra las áreas de análisis existente.
3. El actor selecciona el AA Educación.	4 El sistema muestra los libros de trabajos que contiene el AA Educación.
5. El actor selecciona el LT Educación superior.	6. El sistema muestra los reportes pertenecientes al LT Educación superior.
7. El actor selecciona el reporte deseado.	8. El sistema muestra la información contenida en el reporte seleccionado y brinda la

	posibilidad de analizar el reporte desde diferentes perspectivas. Ir al CU: Visualizar reporte.	
Opciones de reportes de Analizar información de la enseñanza superior		
Perspectivas de análisis	Posibles resultados	
	Medidas	Periodicidad
Variables de entrada relacionadas con el CU Analizar información de la enseñanza superior : <ul style="list-style-type: none"> • Curso. • Especialidad. • Tipo estudio. • Nivel Territorial. 	Variables de salida disponibles en el caso de uso Analizar información de la enseñanza superior: <ul style="list-style-type: none"> • Matrícula inicial de la enseñanza Técnica y Profesional. • Graduados de la enseñanza Técnica y Profesional. • Nuevos ingresos. • Matrícula inicial. • Graduados. • Matrícula inicial de graduados. 	Rango de tiempo en que se solicitan las variables de salida: <ul style="list-style-type: none"> • Anual.
Pos condiciones	Disponibilidad de opciones (cruces de variables) de reportes relacionados con el caso de uso Analizar información de la enseñanza superior.	

Ejemplo de una caso de prueba

Escenario	Descripción	Variables	Respuest	Flujo central
-----------	-------------	-----------	----------	---------------

		Perfiles de análisis	Indicadores a medir	a del sistema	
EC1.1:Graduados por la enseñanza dirigida	Permite visualizar el reporte con las variables presentes en el mismo.	<ul style="list-style-type: none"> • Curso • Especialidad • Nivel Territorial 	<ul style="list-style-type: none"> • Graduados 	Se muestra la tabla con los valores correspondientes a cada escenario.	<p>Se abre la aplicación.</p> <p>Se autentifica.</p> <p>Se entra al sistema.</p> <p>Se despliega hacia la derecha el componente ubicado en el lateral izquierdo que contiene el navegador.</p> <p>Se selecciona el área de análisis de A.A Educación.</p> <p>Se selecciona el libro de trabajo L.T Educación superior.</p> <p>En la parte inferior izquierda se selecciona el reporte deseado.</p> <p>En el área de trabajo se visualiza la tabla correspondiente al</p>
EC1.2:Graduados por rama de la ciencia	Permite visualizar el reporte con las variables presentes en el mismo.	<ul style="list-style-type: none"> • Curso • Especialidad • Nivel Territorial 	<ul style="list-style-type: none"> • Graduados 		
EC1.3:Indicadores de la educación superior	Permite visualizar el reporte con las variables presentes en el mismo..	<ul style="list-style-type: none"> • Curso • Especialidad • Nivel Territorial 	<ul style="list-style-type: none"> • Becarios • Personal docente • Cantidad de escuelas • Matrícula inicial • graduados 		

EC1.4:Matrícula en la enseñanza	Permite visualizar el reporte con las variables presentes en el mismo..	<ul style="list-style-type: none"> • Curso • Especialidad • Nivel Territorial 	<ul style="list-style-type: none"> • Matrícula inicial 		reporte.
EC1.5:Matrícula inicial, Grupo de especialidades	Permite visualizar el reporte con las variables presentes en el mismo.	<ul style="list-style-type: none"> • Curso • Especialidad • Nivel Territorial 	<ul style="list-style-type: none"> • Matrícula inicial 		
EC1.6:Nuevo ingreso a la educación superior	Permite visualizar el reporte con las variables presentes en el mismo.	<ul style="list-style-type: none"> • Curso • Especialidad • Nivel Territorial 	<ul style="list-style-type: none"> • Nuevo ingreso 		

Escenario	Descripción	Variables		Respuesta del sistema	Flujo central
		Perfiles de análisis	Indicadores a medir		
EC1.1:Postgrado	Permite visualizar el reporte con las variables presentes en el mismo.	<ul style="list-style-type: none"> • Figuras • Año • Nivel • Territorial 	<ul style="list-style-type: none"> • Actividades. • Participantes. 	Se muestra la tabla con los valores correspondientes a cada escenario.	<p>Se abre la aplicación.</p> <p>Se autentifica.</p> <p>Se entra al sistema.</p> <p>Se despliega</p>

					<p>hacia la derecha el componente ubicado en el lateral izquierdo que contiene el navegador.</p> <p>Se selecciona el área de análisis de A.A Educación.</p> <p>Se selecciona el libro de trabajo L.T Postgrado.</p> <p>En la parte inferior izquierda se selecciona el reporte deseado.</p> <p>En el área de trabajo se visualiza la tabla correspondiente al reporte.</p>
--	--	--	--	--	--

GLOSARIO

Artefactos: Productos tangibles del proyecto que son producidos, modificados y usados por las actividades. Pueden ser modelos, elementos dentro del modelo, código fuente y ejecutables.

BI: Inteligencia de negocio.

Caso de Uso: Secuencias de acciones que el sistema puede llevar a cabo interactuando con sus actores, incluyendo alternativas dentro de las secuencias.

Data Mart: Mercado de datos.

Data Warehouse: Almacén de datos.

Esquema: Definición formal de todos los datos de objetos y atributos que se pueden almacenar.

ETL: Proceso de extracción, transformación y carga.

SIEN: Sistema de Información Estadístico Nacional.

SIGOB: Sistema de Información de Gobierno.

ONE: Oficina Nacional de Estadísticas.

OLAP: Procesamiento Analítico en Línea.

ROLAP: Procesamiento Analítico en Línea Relacional.

UML: Lenguaje visual para especificar, construir y documentar un sistema de software. Sus siglas vienen dadas por su nombre en inglés Unified Modeling Language.