

Universidad de las Ciencias Informáticas

Facultad 6



Título: Sistema de Información de Gobierno. Mercado de datos para el área de Construcción.

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

Autor:

Leonel Pérez Nieblas

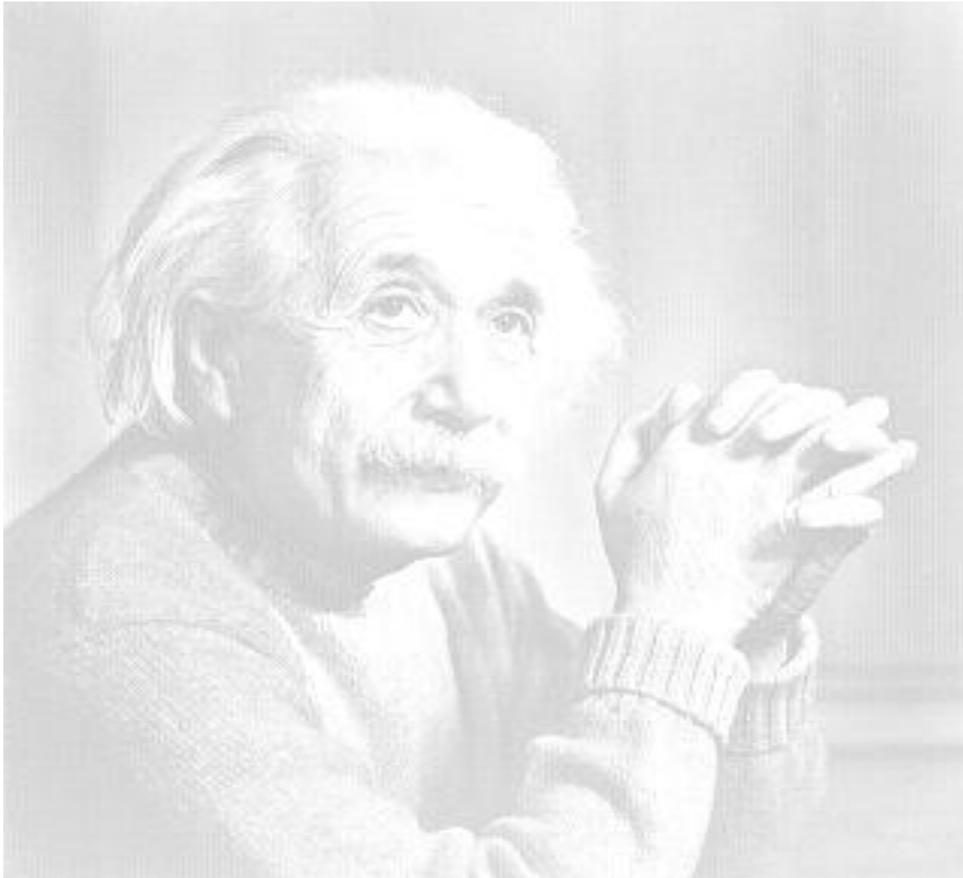
Tutora:

MSc. Yudelkis Abad Fuentes

Ciudad de la Habana, junio de 2011

“Año 53 de la Revolución”

“Nunca consideres el estudio como una obligación, sino como una oportunidad para penetrar en el bello y maravilloso mundo del saber”



Declaración de autoría

Declaro ser autor del presente trabajo “Sistema de Información de Gobierno. Mercado de datos para el área de Construcción” 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 2011.

Leonel Pérez Nieblas

Autor

Yudelkis Abad fuentes

Tutora

Tutora: Ms.C Yudelkis Abad Fuentes

Email: yabad@uci.cu

Graduada en la Universidad de las Ciencias Informáticas en 2007.

AGRADECIMIENTOS

A mi tutora Yudelkis Abad, por ser mi guía durante todo el proceso investigativo, por apoyarme en todo momento y trasmitirme parte de su preparación y experiencia.

A todos los profes que han contribuido en mi formación como profesional.

Al equipo de la FEU que me han enseñado a ser mejor persona.

A todos mis compañeros que me han acompañado durante estos 5 años.

A los profes y tesistas de almacenes de datos.

DEDICATORIA

A mi mamá Roxana, mi abuela Eugenia y mis hermanos Yasleidi y Reinier.

A mi padrastro Rolando y a mis tíos.

A mi novia Liniuska, por estar a mi lado en los momentos difíciles.

A todos mis primos, en especial a los que han sido como hermanos para mí.

RESUMEN

La Oficina Nacional de Estadística (ONE) es la empresa encargada del control estadístico en Cuba. Dicha empresa se divide en áreas socioeconómicas para organizar la información que recibe de los distintos centros informantes del país. Una de esas áreas es la de construcción, en donde se guarda información de las viviendas y obras terminadas, así como, de las presas en explotación. Toda la información que se recoge en esta área, se encontraba almacenada en ficheros Excel, volviendo engorroso el proceso de análisis, y por ende, el proceso de toma de decisiones. Como resultado se implementó el mercado de datos del área de construcción de la ONE, logrando un formato estándar, organizado y centralizado de los datos recogidos en todo el país a lo largo de los años. Los tiempos de respuestas del sistema –durante la visualización de cada uno de los reportes del mercado- son aceptables para la cantidad de usuarios que laboran en el área. El sistema implementado se validó a través de la lista de chequeo, los casos de pruebas diseñados y la carta de aceptación del cliente. En la construcción del mercado, se utilizó PostgreSQL como gestor de bases de datos, la suite Pentaho y el jakarta-jmeter en las pruebas de rendimiento.

Palabras claves: ONE, mercado de datos, toma de decisiones.

TABLA DE CONTENIDOS

AGRADECIMIENTOS.....	I
DEDICATORIA.....	II
RESUMEN	III
INTRODUCCIÓN.....	1
CAPÍTULO 1: Fundamentos teóricos	4
1.1 Control estadístico	4
1.2 Descripción del negocio.....	5
1.3 Aplicaciones existentes	5
1.3.1 Aplicaciones estadísticas en el mundo que utilizan almacenes de datos	5
1.3.2 Aplicaciones estadísticas en Cuba que utilizan almacenes de datos	5
1.4 Almacenes de datos	6
1.4.1 Definición de almacén de datos	6
1.4.2 Mercado de datos	7
1.4.3 Componentes de un almacén de datos	8
1.4.4 Metodologías para el desarrollo de un almacén de datos	9
1.4.5 Estructura de almacenamiento.....	12
1.4.6 Proceso OLAP	14
1.5 Herramientas	15
1.5.1 Herramientas de modelado.	15
1.5.2 Sistemas Gestores de Bases de Datos	16
1.5.3 Administrador de bases de datos	16
1.5.4 Servidor de Aplicaciones	17
1.5.5 Herramienta de desarrollo	17
1.6 Conclusiones	18
CAPÍTULO 2: Análisis y Diseño.	20
2.1 Análisis	20
2.1.1 Definición del negocio.....	20
2.1.2 Temas de análisis.....	20
2.1.3 Reglas del negocio.	21
2.1.4 Descripción de los actores del sistema.....	22
2.1.5 Necesidades de los usuarios.....	22
2.1.6 Requisitos de información.	22

2.1.7 Requisitos funcionales	23
2.1.8 Requisitos no funcionales.....	24
2.1.9 Casos de uso del sistema	27
2.1.10 Descripción de los casos de usos críticos	27
2.2 Diseño	31
2.2.1 Modelo conceptual.....	32
2.2.1.1 Dimensiones	32
2.2.1.2 Medidas.....	32
2.2.1.3 Tabla de hechos	33
2.2.2 Modelo de Datos	35
2.2.3 Diseño de las transformaciones	36
2.2.4 Diseño de los reportes candidatos	36
2.3 Políticas de seguridad	37
2.3.1 Roles y permisos	37
2.3.2 Salva de la base de datos	38
2.4 Conclusiones	38
CAPÍTULO 3: Implementación y pruebas	39
3.1 Implementación del modelo de datos.....	39
3.2 Implementación de los procesos de Extracción, Transformación y Carga	39
3.2.1 Implementación de las transformaciones.....	39
3.2.2 Implementación de los trabajos	40
3.3 Implementación de los procesos de Inteligencia de Negocio	41
3.3.1 Implementación de los cubos OLAP	41
3.3.2 Implementación de los reportes	42
3.3.3 Vistas de análisis	42
3.4 Pruebas.....	43
3.4.1 Prueba de Interfaz	44
3.4.2 Lista de chequeo.....	44
3.4.3 Pruebas de aceptación	45
3.4.4 Pruebas de rendimiento	45
3.6 Conclusiones	48
CONCLUSIONES	49
RECOMENDACIONES.....	50
REFERENCIAS BIBLIOGRÁFICAS	51
BIBLIOGRAFÍA	56

ANEXOS	60
GLOSARIO	67

ÍNDICE DE TABLAS

Tabla 1: Descripción de los Actores del sistema	22
Tabla 2: Ejemplos de convenciones de nombrado	26
Tabla 3: Descripción del caso de uso "Extraer datos de los archivos Excel"	28
Tabla 4: Descripción del caso de uso "Transformar y cargar datos"	30
Tabla 5: Descripción del caso de uso "Administrar Reportes"	31
Tabla 6: Tabla de hechos	34
Tabla 7: Matriz BUS	35
Tabla 9: Reporte candidato "Viviendas terminadas"	37
Tabla 8: Roles y permisos	38
Tabla 10: Esquemas y tablas del mercado	39

ÍNDICE DE FIGURAS

Figura 1: Relación entre los componentes de un Almacén de Datos	8
Figura 2: Metodología de DW&BI.....	11
Figura 3: Estructura de un Cubo OLAP	12
Figura 4: Diagrama de Casos de Uso	27
Figura 5: Proceso de diseño	32
Figura 6: Modelo de datos	35
Figura 7: Diseño de las transformaciones.....	36
Figura 8: Transformación para cargar el hecho viviendas terminadas	40
Figura 9: Trabajo para ejecutar las transformaciones.....	41
Figura 10: Diseño de los cubos OLAP	41
Figura 11: Vista del reporte Estructura de Obras	43
Figura 12: Resultado de las pruebas de interfaz	44
Figura 13: Comportamiento de los indicadores por secciones.....	45
Figura 14: Plan de pruebas	47
Figura 15: Resultados para 25 usuarios.....	47
Figura 16: Resultados para 50 usuarios.....	48
Figura 17: Resultados para 100 usuarios	48

INTRODUCCIÓN

En la actualidad, el centro del desarrollo de las aplicaciones comerciales está determinado por el volumen de información recogido a lo largo de vida del desarrollo de ese negocio (1) (2) (3) (4) así como, la conversión de esa información en conocimiento útil para la empresa. Sin embargo, un manejo eficiente de grandes volúmenes de información, no siempre es una tarea fácil, por lo que se requieren de técnicas innovadoras para su almacenamiento y procesamiento. En este sentido, se han dado avances significativos en los repositorios de información, los que han ido evolucionando desde un fichero de texto hasta grandes bases de datos.

Dentro de todas las técnicas de almacenamiento reportadas en la bibliografía, el uso de bases de datos, representan un salto cuantitativo en la automatización de procesos y almacenamiento de abundante información (5) (6). Estas técnicas, pueden clasificarse de dos maneras atendiendo a la información que manejan: bases de datos estáticas y bases de datos dinámicas. Las bases de datos dinámicas (7) son bases de datos donde la información almacenada se modifica con el tiempo, permitiendo operaciones como actualización, borrado y adición de datos, además de las operaciones fundamentales de consulta.

En cambio, las bases de datos estáticas (7) tienen por característica principal el uso de información invariante en el tiempo, permitiendo que se realicen extracciones de conocimiento a partir de la información recopilada, característica esta que le ha permitido ser la variante adoptada en proyectos que requieren de Inteligencia de Negocio.

La inteligencia de negocios es un término adoptado a finales de la década de los 80, mediante el cuál se pretendía definir el proceso de analizar los bienes o datos acumulados en la empresa y extraer una cierta inteligencia o conocimiento de ellos (8) (9).

Cuba posee una gran cantidad de información estadística que a lo largo de los años se ha ido recopilando en las distintas ramas de la economía como son: construcción, cultura y deporte, educación, comercio exterior, entre otras, provenientes del resultado de sus operaciones. Este proceso de recopilación, se encuentra dirigido y centralizado por la Oficina Nacional de Estadísticas (ONE), la que a pesar de tener toda esta información a disposición, se encuentra ante la dificultad de poder dar un significado que sea punto de apoyo para la toma de decisiones efectivas, debido a que la

información se encuentra almacenada en distintos ficheros de diferentes formatos, volviéndose engorrosa la centralización, manipulación e interacción con la misma.

En la ONE, específicamente en el área de construcción, se cuenta con gran cantidad de información sobre las construcciones que se desarrollan en el país, ya sean educacionales, industriales, hidráulicas, agropecuarias, marítimas, obras de salud, entre otras. La información se recopila anualmente por cada uno de los centros informantes del país en distintas hojas de cálculos, las que luego son integradas manualmente en un solo fichero al terminar el año. Dada esta situación, los reportes se generan en tiempos prolongados de solicitud, lo que frena el análisis y estudio de la información almacenada, y trae consigo, demoras en el proceso de toma de decisiones para los órganos del estado.

Por lo anteriormente planteado surge la **interrogante**: ¿Cómo contribuir al proceso de toma de decisiones en el área Construcción del Sistema de Información de Gobierno?

Se establece como **objeto de estudio**, Almacenes de Datos, y como **campo de acción** Mercado de datos estadístico para el área construcción del Sistema de Información de Gobierno.

Objetivo general: Desarrollar un mercado de datos para el área de construcción del Sistema de Información de Gobierno.

Objetivos específicos

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

Para lograr el objetivo propuesto se ha trazado una serie de **tareas de Investigación**:

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

- Definición de la arquitectura del mercado de datos
- Diseño del subsistema de integración.
- Diseño del subsistema de visualización.
- Diseño de los casos de pruebas.
- Implementación del subsistema de integración.
- Implementación del subsistema de visualización.
- Aplicación de las listas de chequeo.
- Aplicación de los casos de pruebas.

Estructura de la Tesis

El presente trabajo está compuesto por 3 capítulos estructurados de la siguiente manera:

Capítulo 1: Fundamentos teóricos

En este capítulo se abordan los principales conceptos tratados en la investigación realizada. La importancia que tiene para una organización un buen control estadístico, la descripción del proceso estadístico a modelar, la metodología a seguir en el diseño e implementación de un mercado de datos, y la selección de herramientas útiles para llevar a cabo un excelente almacenamiento de la información.

Capítulo 2: Análisis y Diseño del mercado de datos del área construcciones

Este capítulo contiene la descripción de los pasos a seguir durante el análisis y el diseño de la solución. Se abordan aspectos concernientes a la descripción de las fuentes a integrar. Se definen los requisitos que debe cumplir el sistema, así como el modelo dimensional propuesto para el desarrollo del Mercado de Datos a partir de los indicadores que se seleccionaron.

Capítulo 3: Implementación y pruebas del mercado de datos del área construcciones

En este capítulo se implementan todos los procesos de extracción, transformación y carga de los datos donde se realizan los flujos de integración y los trabajos para enlazar todas las transformaciones. Además, se realizan los procesos de inteligencia de negocio donde se implementa el modelo de datos, así como los cubos OLAP y los reportes candidatos. Se aplican los casos de pruebas diseñados y la lista de chequeo.

CAPÍTULO 1: Fundamentos teóricos

Desde su surgimiento, las bases de datos se convirtieron en la herramienta fundamental para almacenar gran cantidad de información. Debido a esto, en muy poco tiempo la información almacenada por las grandes empresas y negocios alcanzaron una dimensión considerablemente voluminosa. Con la acumulación de esta información se presentó la problemática de cómo darle un fin útil.

La solución sería unificar las diferentes fuentes de información de las cuales disponían, en un único lugar, al que sólo se le incorporaría información relevante, sobre la base de una estructura integrada. La respuesta a esto fueron los almacenes de datos como se conocen mundialmente.

En este capítulo se van a tratar aspectos sobre los almacenes de datos, conceptos, características, ventajas, componentes, metodologías y herramientas a utilizar en su implementación.

1.1 Control estadístico

La palabra estadística proviene del latín "*status*", que significa situación o estado y se reconoce como una ciencia que permite recopilar, organizar y procesar datos con la finalidad de ayudar en la toma de decisiones (10). En la actualidad, existe una amplia variedad de información referente a planos económicos, políticos y sociales, muy utilizados en pronósticos futuros de la rama en que se estudian, lo que implica que el uso de la estadística sea de vital importancia para lograr un control de datos exitoso.

Estudios demuestran, que las empresas que más han avanzado en su negocio, han sido aquellas que hicieron de las estadísticas una herramienta fundamental para tener una gestión eficiente, ya sea para negociar, tomar decisiones o aumentar la productividad. Toda decisión, está prácticamente en el aire si no se cuenta con datos estadísticos suficientes y fiables, una empresa sin información no puede verificar si un producto o una actividad pueden causarle pérdidas o generar utilidades. Tener datos estadísticos, permite a directivos tener claridad en lo que acontece en su empresa y tomar las decisiones más aceptadas (11).

El control estadístico se aplica para analizar la información de diversos temas de interés, entre ellos: comercio exterior, cultura y deporte, medio ambiente, inmigración y extranjería, inversiones,

construcciones, este último en particular controla la información referente a las construcciones que se realizan, ya sea el tipo, la cantidad y la fecha en que se realizan.

1.2 Descripción del negocio

El área de construcción de Cuba, recoge información acerca de las construcciones de viviendas y urbanizaciones que se construyen, ya sean agropecuarias, educacionales, obras para la salud, obras de redes eléctricas y construcciones industriales. La recogida de estos datos favorece en gran medida conocer las viviendas y obras terminadas por provincias, regiones, así como, sus estructuras, dinámicas y costo por año.

Con el devenir de los años, esta información recopilada se vuelve variable y abundante, pues existen datos de todos los municipios del país, lo que trae consigo que se haga muy difícil su manipulación y análisis.

1.3 Aplicaciones existentes

1.3.1 Aplicaciones estadísticas en el mundo que utilizan almacenes de datos

Como resultado de la experiencia a lo largo de la existencia de una empresa, estas acumulan gran cantidad de información (8) (9), considerada como un activo fundamental, en la toma de decisiones futuras. La conversión de esa información en conocimiento se ha convertido en la única fuente de competitividad sostenible, para incrementar su eficiencia, elevar la eficacia y su posición en el mercado.

Por ejemplo, grandes trasnacionales como Coca Cola, Walt Disney, Nike han incorporado la utilización de los almacenes de datos en la realización de estudios de mercado y de inteligencia de negocio, para realizar complejas consultas y engorrosos reportes en la toma de decisiones (12).

1.3.2 Aplicaciones estadísticas en Cuba que utilizan almacenes de datos

En Cuba existen diferentes entidades que han implementado almacenes de datos para gestionar su información. Ejemplo de ello constituye el almacén comercial de la Corporación CIMEX (13). En la Feria de Informática del 2002, se presentó un almacén de datos para CUBACEL (14). Todos estos almacenes recopilan información general para centrar su atención a la actividad del comercio, pero a pesar de dichos logros, no se ha encontrado referencia al desarrollo de un mercado de datos

estadísticos para llevar a cabo el control descriptivo de la información recogida en el sector de la construcción.

1.4 Almacenes de datos

La informática desde su surgimiento ha sido de ayuda para los científicos, dotándolos de métodos, modelos, técnicas y herramientas para el análisis de datos y el modelado robusto y coherente de los mismos. Los mecanismos para la recolección automática de estos datos y el desarrollo de la tecnología de recolección de datos, han generado que se puedan almacenar grandes cantidades de datos en distintos repositorios de información, los que según su estructura pueden ser ficheros de texto, hojas de cálculo y bases de datos.

De ellas, las bases de datos representan ventajas sobre otras técnicas de almacenamiento, porque en ellas se tienen un mejor control sobre la redundancia de los datos, estos son más consistentes y se tiene más información sobre la misma cantidad de datos (15). Según la variabilidad de los datos, estas se pueden clasificar en estáticas o dinámicas (16).

Las BD estáticas: Bases de datos de sólo lectura, utilizadas fundamentalmente para almacenar datos históricos que posteriormente puedan ser utilizados en el estudio del comportamiento de un fenómeno a través del tiempo, predecir patrones de comportamiento, realizar proyecciones futuras y tomar decisiones (16).

Las BD dinámicas: Bases de datos donde la información almacenada se modifica en tiempo de ejecución, permitiendo operaciones sobre los datos con comportamientos distintos, se pueden obtener varios resultados en una misma consulta (16).

En este trabajo se van a estudiar las bases de datos estáticas, fundamentalmente los almacenes de datos.

1.4.1 Definición de almacén de datos

Desde que surgió la tecnología de *warehousing* (década de los 90) hasta la fecha, ha venido posicionándose como la variante más acertada para analizar la información histórica. Un almacén es una base de datos, orientada al análisis de la información histórica contenida en ella, característica que lo ha convertido en una potente herramienta para la toma de decisiones.

En el libro “*Mastering Data Warehouse Design, Relational and Dimensional Techniques*” (17) se plantea que la definición universalmente aceptada es la desarrollada en los 90’s por William H. Inmon, más conocido como Bill Inmon, donde se plantea que “Son un conjunto de datos orientados a un tema, integrados, de tiempo variante y no volátiles usados en la estrategia de toma de decisiones administrativas” (18).

Entre las ventajas más importantes de su uso se encuentran:

- Hacen más fácil el acceso a una gran variedad de datos a los usuarios finales.
- Se obtiene una base de datos histórica y clasificada por temas.
- Integra información procedente de múltiples sistemas externos.
- Facilita la toma de decisiones estratégicas.

Normalmente, si el almacén crece muy rápido y los distintos departamentos contenidos en él, requieren sólo acceder a sus datos, se recomienda implementar almacenes más pequeños denominados mercado de datos, que representan una estrategia de "divide y vencerás".

1.4.2 Mercado de datos

Los *data mart*, conocidos como mercados de datos, son un subconjunto específico de datos con el objetivo de ayudar a que un área determinada dentro del negocio pueda tomar mejores decisiones. Un concepto más amplio sería: el *data mart* es un sistema orientado a la consulta, es consultado mediante herramientas OLAP que ofrecen una visión multidimensional de la información. En síntesis, se puede decir que los mercados de datos son pequeños almacenes de datos centrados en un tema o un área de negocio específico dentro de una organización (19).

Beneficios de la utilización de los mercados de datos:

- Acelera las consultas al reducir la cantidad de datos a recorrer.
- Estructura los datos para su adecuado acceso por una herramienta.
- Divide los datos para imponer estrategias de control de acceso.
- Segmenta los datos en diferentes plataformas de hardware.
- Permite el acceso a los datos mediante un gran número de herramientas del mercado, logrando así independencia de estas.

1.4.3 Componentes de un almacén de datos

Los almacenes de datos están compuestos por una serie de procesos que definen, en su conjunto, el ambiente que estos poseen. Aunque cada desarrollo de un almacén de datos es diferente debido a las características de las organizaciones, generalmente cumplen con los siguientes componentes descritos a continuación (20).

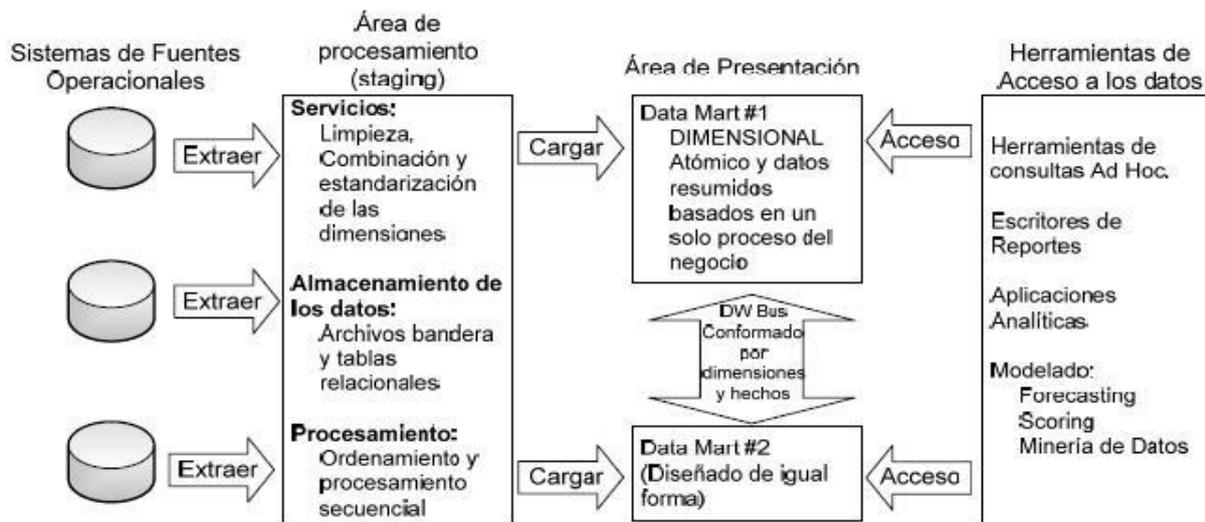


Figura 1: Relación entre los componentes de un Almacén de Datos

El primer componente hace referencia al **área de procesamiento**, encargada de los datos temporalmente y permitir la realización de un conjunto de procesos comúnmente llamados de Extracción, Transformación y Carga (ETL). Además, constituye la interfaz entre las fuentes operacionales y el área de presentación.

El segundo componente es el **Área de Presentación**. En este componente los datos se encuentran organizados, almacenados y disponibles para ser consultados, reportados o analizados por parte de los usuarios finales. Generalmente esta área es referenciada como una serie de mercados de datos integrados donde cada uno se encuentra representando a un proceso específico del negocio.

El tercer componente lo constituyen las **Herramientas de Acceso a Datos**. En este componente se usa la palabra herramientas para referirse a la variedad de capacidades que pueden ser provistos a los usuarios del negocio para el soporte a la toma de decisiones. Su actividad principal es la de consultar

el área de presentación del almacén de datos. El mismo puede abarcar desde una simple o personalizada herramienta de consulta hasta una compleja y sofisticada aplicación de modelado o de minería de datos¹.

El desarrollo de estos componentes, varían según la metodología a seguir, entre las más conocidas se encuentran: Kimball e Inmon (21).

1.4.4 Metodologías para el desarrollo de un almacén de datos

Metodología, es la ciencia que estudia los métodos del conocimiento. Se refiere a los métodos o procedimientos de investigación que se siguen para alcanzar una gama de objetivos en una ciencia (22).

En el diseño de un almacén de datos, se ha destacado un conjunto de metodologías que definen y guían todo el ciclo de vida del desarrollo. Dentro de los criterios que existen hay dos bien identificados y que han marcado claramente su tendencia sirviéndole de guía a los desarrolladores de este tema (23). Estas tendencias son las conocidas como: Metodología Kimball y Metodología de Inmon, este último creador del término *data warehouse*, así como, del CIF (*Corporate Information Factory*) conjuntamente con Claudia Imhoff. Por su parte, Ralph Kimball es un maestro del diseño de almacenes de datos.

Según Bill Inmon, el almacén de datos es una parte del todo que conforma un sistema de inteligencia. Una entidad tiene un almacén de datos y los mercados de datos tienen como fuente de información ese almacén. Esta tendencia es conocida como “*Top-Down*” (23).

Para Ralph Kimball el almacén se compone por el conjunto de mercado de datos que existan en la entidad donde esté implementado y la información siempre se almacena en un modelo dimensional. Esta tendencia es conocida como “*Bottom-Up*” (23).

¹ Proceso conocido como *Data Mining*, consiste en la extracción no trivial de información que reside de manera implícita en los datos. En otras palabras, la minería de datos prepara y explora los datos para sacar la información oculta en ellos.

1.4.4.1 Metodología a usar para el desarrollo del almacén de datos

En este trabajo se va a emplear la **Metodología para el desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio (DW&BI)** (24). La cual surge tomando como base la metodología propuesta por Ralph Kimball en 1992. Se le realizaron algunas modificaciones y adaptaciones en dependencia de las características del centro.

Algunas de las características principales y ventajas de esta metodología:

- La solución completa se puede implementar en poco tiempo.
- El desarrollo es iterativo e incremental donde se construye una pieza 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 su posterior integración en un almacén de datos de la organización. Esta estrategia coincide con la división lógica que se tiene en una organización.
- Cuenta con mayor velocidad de respuesta al cliente.
- Los productos son más comprensibles para los usuarios.
- Es resistente y tolerante ante los cambios.

La metodología cuenta con cuatro fases principales dentro de las cuales se encuentran 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, definiéndose la arquitectura de información que regirá el desarrollo de la solución.

- **Implementación:** Se diseña físicamente el repositorio de datos, se crean las estructuras de almacenamiento, el área temporal de almacenamiento, se ejecutan las reglas de ETL y se configuran e implementan las herramientas de BI para la obtención de los elementos que se acordaron con el cliente final.
- **Prueba:** Se realizan las pruebas al sistema desde las Pruebas de Unidad hasta las de Aceptación con el cliente final.
- **Despliegue:** Se realiza un despliegue piloto en el cual se configuran los servidores y se instalan las herramientas y se carga una muestra de los datos para demostrar que el sistema funciona. Posterior a la aceptación del cliente se realiza la carga de los datos así como la Capacitación y Transferencia Tecnológica.
- **Soporte y Mantenimiento:** Tras la implantación de la solución se brindan los servicios de soporte en línea, vía telefónica, web u otras según el contrato firmado y las condiciones de soporte establecidas.
- **Gestión y Administración del Proyecto:** A lo largo del ciclo de vida se realizan actividades de control, gestión y chequeo del desarrollo, los gastos, las utilidades, los recursos y demás actividades por parte del Grupo de Dirección del Proyecto.

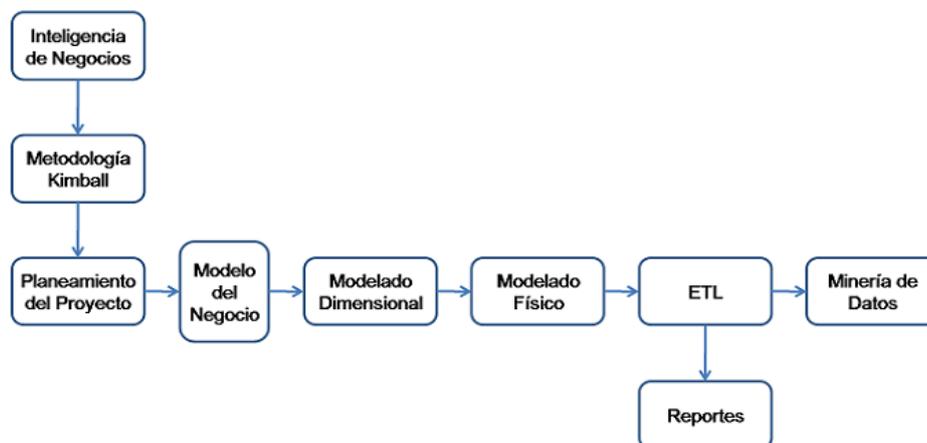


Figura 2: Metodología de DW&BI

Uno de los elementos fundamentales, luego de realizado el levantamiento de requisitos del almacén, lo constituye la estructura de almacenamiento como parte de la arquitectura y diseño, pues es ahí, donde se van a cargar todos los datos provenientes de las diversas fuentes. Esta estructura donde se representan las tablas y sus relaciones se denomina modelo multidimensional.

1.4.5 Estructura de almacenamiento

A diferencia de los sistemas de bases de datos clásicos, la estructura de las tablas y sus relaciones son representadas mediante un modelo multidimensional (25), o sea, almacenan la misma información que el diagrama entidad relación pero la organiza de forma diferente para garantizar la velocidad y eficiencia en la recuperación de la misma. Sus 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.
- Agregar nuevos atributos a las dimensiones. Esta característica de gran adaptabilidad es muy deseable, pues a medida que los analistas añaden nuevos requerimientos al sistema, se pueden ir incorporando las modificaciones sin que esto implique demasiados cambios.

El modelo multidimensional al disminuir la normalización, resulta sencillo a los usuarios y bastante adaptable. En la Figura 3 se muestra la estructura espacial que posee este tipo de diseño.

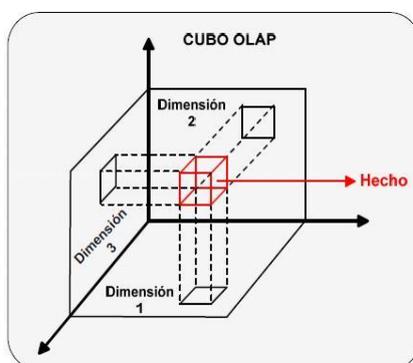


Figura 3: Estructura de un Cubo OLAP

En la figura se puede observar el modelo multidimensional de la información (cubo OLAP), el cual está dividido en dos grandes grupos: las dimensiones y las tablas de hechos. Las tablas de hechos son tablas primarias en el modelo multidimensional y contiene los valores del negocio. Dentro de ellas se almacenan las medidas, operaciones que generalmente son numéricas a diferencia de las

dimensiones que son textuales. Las dimensiones por su parte, van a ser las descripciones del entorno de los hechos.

Tablas de Hechos: Las tablas de hechos son las tablas primarias en el modelo dimensional, ya que almacenan los valores del negocio, cada tabla representa una relación de muchos a muchos, y contiene dos o más llaves extranjeras que se enlazan con sus respectivas tablas de dimensiones. Dada este tipo de información al almacenar, cada tabla de hecho generalmente define un mercado de datos determinado. La principal condición que debe cumplir las tablas de hechos es que el hecho debe almacenarse de tal forma que su valor sea numérico y a su vez sea aditivo para así poder realizar cálculos sobre él, ya sea por ciento, sumas o igualdades (26).

Tablas de Dimensiones: Las tablas de dimensiones son las compañeras integrales de las tablas de hechos, ellas contienen la descripción textual del negocio. En el modelo dimensional, las tablas de dimensiones poseen varios atributos que en su conjunto definen una fila en la tabla de dimensión. Los atributos de las dimensiones sirven como fuente primaria de las restricciones de las consultas, agrupaciones y las etiquetas de los reportes. Ellos desempeñan un rol de vital importancia dentro del almacén de datos debido a que son las llaves que hacen el almacén usable y entendible. Estos atributos son las llaves de entrada a los hechos o medidas almacenadas (27).

La calidad de todo almacén de datos se mide por la definición de los atributos de las dimensiones. Su poder es directamente proporcional a la calidad y profundidad de estos atributos (20).

Implementación del modelo multidimensional

El modelo multidimensional implica tres variantes de modelación según 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 de estrella, donde las tablas de dimensiones se ramifican en más puntas.

Esquema constelación: está compuesto por una serie de esquemas de estrella, es decir, varias tablas de hechos centrales y sus respectivas tablas de dimensiones.

1.4.6 Proceso OLAP

La tecnología de Procesamiento Analítico en Línea (OLAP) permite a los usuarios analizar diferentes dimensiones de datos multidimensionales. Su objetivo es agilizar la consulta de grandes cantidades de datos, lo que proporciona respuestas rápidas a consultas analíticas complejas e iterativas, utilizadas generalmente para sistemas de ayuda que contribuyan a la toma de decisiones. Las principales características de OLAP son:

- **Rápido:** proporciona la información al usuario a una velocidad constante (mayormente en cinco segundos o menos).
- **Análisis:** realiza análisis estadísticos y numéricos básicos de los datos.
- **Compartida:** permite compartir los datos potencialmente confidenciales a través de una gran cantidad de usuarios, implementando para esto los requerimientos de seguridad necesarios.
- **Multidimensional:** permite ver la información en determinadas vistas o dimensiones.
- **Información:** accede a todos los datos necesarios, donde quiera que estos residan, mientras no esté limitada por el volumen.

Tipos de OLAP

- **ROLAP:** En el Procesamiento Analítico Relacional en Línea (*Relational Online Analytical Process*, en inglés) los datos son almacenados en filas y columnas de forma relacional. Este modelo presenta los datos a los usuarios en forma de dimensiones de negocio.

La arquitectura ROLAP es capaz de usar datos pre-calculados si estos están disponibles, o de generar dinámicamente los resultados desde los datos elementales si es preciso. Esta arquitectura accede directamente a los datos del almacén y soporta técnicas de optimización de accesos para acelerar las consultas. Estas optimizaciones son, entre otras, particionado de los datos a nivel de aplicación, soporte a la desnormalización y joins múltiples (28).

- **MOLAP:** El Procesamiento Analítico Multidimensional en Línea (*Multidimensional Online Analytical Process*, en inglés) usa unas bases de datos multidimensionales para proporcionar el análisis, su principal premisa es que el OLAP está mejor implantado almacenando los datos multidimensionalmente.

Un sistema MOLAP usa una base de datos propietaria multidimensional, en la que la información se almacena y visualiza multidimensionalmente. Este sistema utiliza una arquitectura de dos niveles: Las bases de datos multidimensionales y el motor analítico. La base de datos multidimensional es la encargada del manejo, acceso y obtención del dato; y el motor analítico es el responsable de la ejecución de los requerimientos OLAP. El nivel de presentación se integra con el de aplicación y proporciona una interfaz a través de la cual los usuarios finales visualizan los análisis OLAP (28).

- **HOLAP:** Procesamiento analítico híbrido en línea. Permiten un análisis híbrido de la información, une lo mejor de las dos tecnologías anteriores. El análisis HOLAP ayuda a reducir costes de hardware ya que se necesita menos espacio en disco que en las bases de datos relacionales. Además, la respuesta de las consultas sobre las bases de datos multidimensionales son más rápidas que sobre las relacionales. Como aspecto negativo, los datos multidimensionales deben ser cargados antes de ser consultados y actualizados cuando se actualizan los datos de la organización.

1.5 Herramientas

En la actualidad se han desarrollado diversas herramientas con el fin de dar un acercamiento a la automatización del diseño, construcción, implementación y mantenimiento de los almacenes de datos.

1.5.1 Herramientas de modelado.

Las herramientas de modelado son de gran ayuda para el desarrollo de un software, estas herramientas son utilizadas por las personas que intervienen en el proceso de desarrollo de software como los diseñadores y analistas, durante las fases de desarrollo del producto. Estas herramientas se utilizan para agilizar y facilitar el trabajo, ya que proveen de métodos, técnicas y utilidades que ayudan al perfeccionamiento del desarrollo de software.

Visual Paradigm: es una herramienta UML profesional que soporta el ciclo de vida completo del desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue. El software de modelado UML ayuda a una más rápida construcción de aplicaciones de calidad, mejores y a un menor coste. Permite dibujar todos los tipos de diagramas de clases, código inverso, generar código desde diagramas y generar documentación. Genera código para un gran número de lenguajes

de programación entre los que se encuentra Java y permite integración con varias herramientas de Java. Además, brinda una versión libre para uso no comercial (29).

Esta es la herramienta propuesta y aprobada por el centro DATEC para el desarrollo de software, pues proporciona un entorno amigable e interactivo al usuario y permite la exportación de script de una manera íntegra y eficiente.

1.5.2 Sistemas Gestores de Bases de Datos

Se considera un Sistema Gestor de Base de Datos (SGBD) a un conjunto de programas que ayudan a administrar la información que está contenida en una base de datos. Estos programas se encargan de la integridad y la seguridad de los datos; y garantizan la interacción con el sistema operativo.

PostgreSQL: Es un potente Sistema de Base de Datos Relacional para grandes bases de Datos, liberado bajo licencia The PostgreSQL Licence (TPL) (30).

Funciona en la mayoría de los sistemas operativos actuales, soporta casi toda la sintaxis SQL, posee puntos de recuperación a un momento dado, *tablespaces*, replicación asincrónica, transacciones jerarquizadas y copia de seguridad en línea.

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

1.5.3 Administrador de bases de datos

Existen diversas herramientas para la administración de bases de datos entre ellas se tienen PgAdmin, MS manager for PostgreSQL

PgAdmin 1.10: Es una de las herramientas más populares para administrar las bases de datos en PostgreSQL, que presenta las siguientes características:

- Es un software libre.
- Accede a todos los objetos del PostgreSQL y responde las necesidades de todos los usuarios, desde escribir consultas SQL simples hasta desarrollar bases de datos complejas.

- Permite que se pueda usar en Linux, FreeBSD, Solaris, Mac OS X y Windows (31).

1.5.4 Servidor de Aplicaciones

Apache Tomcat: herramienta utilizada como servidor web que presenta diversas características entre las que se encuentran:

- 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 (32).

1.5.5 Herramienta de desarrollo

Suite Pentaho: La Open BI Suite de Pentaho, provee un completo espectro de funcionalidades de Inteligencia de Negocios, incluyendo reportes, análisis, minería de datos, integración de datos y una plataforma de BI que la han convertido en la suite de código abierto más popular del mundo (33).

Entre las funcionalidades integradas con la plataforma de BI de Pentaho se encuentran:

- Pentaho BI Server
- Pentaho Data Integration
- Pentaho Schema Workbench
- Mondrian OLAP Server

Pentaho BI Server: La aplicación más conocida de la Plataforma Pentaho BI es la BI Server, que 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.) Está diseñado para integrarse fácilmente en cualquier proceso de negocio (34).

Pentaho Data Integration: es una de las herramientas ETL de código abierto que reúne un conjunto de componentes que permiten modelar y ejecutar transformaciones sobre flujos de datos. Presenta las siguientes características:

- Puede funcionar sobre varias plataformas a través de un sistema que soporte Java 1.4 Runtime Environment o una versión superior.
- Exige solamente alrededor de 128 MB de RAM.
- Provee un JDBC² que permite la conexión con cualquier base de datos sin tener que instalar un cliente adicional.
- Se integra con ficheros de Microsoft Office, Web services y cubos MOLAP.
- Esta herramienta incluye procesamiento optimizado de los ficheros planos.
- Brinda soporte para metadatos e incorpora operaciones de transformación.
- Permite operar con los campos en el flujo de datos, renombrando, calculando campos en función de otros, correlacionando valores y realizando búsquedas auxiliares en bases de datos.
- Corrección de errores (35).

Pentaho Schema Workbench: es un entorno visual para el desarrollo y prueba de cubos OLAP Mondrian. Provee un mecanismo para buscar datos con rapidez y tiempo de respuesta uniforme independientemente de la cantidad de datos en el cubo o la complejidad del procedimiento de búsqueda. Permite la ejecución de consultas MDX contra el esquema y la base de datos y la navegación por la base de datos subyacente (36).

Mondrian OLAP Server: es un servidor OLAP de código abierto que gestiona la comunicación entre la aplicación OLAP y la base de datos. Permite crear cubos de información para análisis multidimensional. Este proporciona la conexión a la base de datos y ejecuta las sentencias SQL (37).

1.6 Conclusiones

La Oficina Nacional de Estadísticas entidad rectora de la estadística en Cuba amerita la utilización de una metodología robusta y madura que garantice el éxito de la integración de la información que actualmente dispone. A partir del estudio del estado del arte realizado se concluyó lo siguiente:

- La metodología de desarrollo adoptada es la Metodología para el desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio (DW&BI).
- Como Sistema Gestor de Base de Datos utilizar PostgreSQL.

² Java Database Connectivity (JDBC), permite la ejecución de operaciones sobre bases de datos desde el lenguaje de programación Java.

- Como Administrador de Base de Datos utilizar el PgAdmin
- Como herramienta CASE a utilizar, Visual Paradigm.
- Como herramientas para el desarrollo del mercado de datos:
 - Suite de Pentaho

CAPÍTULO 2: Análisis y Diseño.

2.1 Análisis

2.1.1 Definición del negocio.

La Oficina Nacional de Estadísticas es la institución rectora de los temas estadísticos en Cuba y uno de sus objetivos es funcionar como un repositorio central, donde se lleven a cabo un conjunto de procesos que ejecutan y supervisan la gestión estadística del país.

La información estadística está recogida en modelos de manera organizada por cada centro informante. Dicha información está constituida bajo los diferentes niveles de estructura nacional que posee la ONE y donde cada uno de estos modelos incluye un conjunto de indicadores. Los indicadores relacionados se ajustan a cada centro informante en dependencia del trabajo que realice cada uno, ya sea social, estatal, económicos o de otras índoles.

En el área de construcción se realiza un estudio de todas las construcciones que se realizan en el país, ya sean viviendas u obras terminadas y presas en explotación. La información que maneja esta área se encuentra dividida en los ficheros viviendas terminadas, obras terminadas y presas en explotación. En viviendas terminadas se guarda la información sobre el total de viviendas terminadas en el país, las estatales y no estatales, las UBPC, CPA y CSS. En el fichero obras terminadas se guardan las obras construidas en el país y el valor monetario para su construcción. En el archivo presas en explotación se guarda la cantidad de presas existentes en el país y su capacidad. Toda esta información es recogida anualmente.

2.1.2 Temas de análisis.

Un tema de análisis no es más que la división o clasificación de la información de una organización de acuerdo a las diferentes temáticas que contenga la organización. En el caso de la investigación centrada en los archivos Excel del área de construcción se define como temas de análisis: el control estadístico de las viviendas y obras terminadas en el país, así como las presas en explotación.

Dentro de este tema se recogen los siguientes datos:

- Viviendas terminadas.
- Estructura de viviendas terminadas
- Dinámica de viviendas terminadas.
- Obras terminadas.
- Estructura de obras terminadas.
- Dinámica de obras terminadas.
- Capacidad de las presas en explotación.
- Cantidad de las presas en explotación.

2.1.3 Reglas del negocio.

Las reglas del negocio son las transformaciones que se le deben realizar a ciertos datos para obtener a partir de ellos otros datos.

Estructura de viviendas terminadas

Información sobre las viviendas, representa el por ciento de las viviendas con respecto al total.

RN 1

Estructura = Cantidad Viviendas / Total Viviendas * 100

RN 2

Estructura: Valor mayor o igual que cero y menor que 100.

Dinámica de viviendas terminadas

Información sobre las viviendas, representa el por ciento de las viviendas construidas con respecto al año anterior.

RN 3

Dinámica = Año actual / Año anterior * 100

Estructura de obras terminadas

Información sobre las obras terminadas, representa el por ciento de las obras con respecto al total.

RN 4

Estructura = Cantidad Obras / Total Obras del año actual* 100

RN 5

Estructura: Valor mayor que cero y menor que 100.

Dinámica de obras terminadas

Información sobre las obras terminadas, representa el por ciento de las obras construidas con respecto al año anterior

RN 6

Dinámica = Año actual / Año anterior * 100

2.1.4 Descripción de los actores del sistema.

Actor	Descripción
Administrador	El administrador interactúa con el sistema para realizar todas las operaciones de administración de roles, usuarios y reportes.
Administrador ETL	El administrador ETL interactúa con la base de datos para monitorizar los procesos de extracción, transformación y carga de los datos.
Analista	El analista interactúa con el sistema para analizar y consultar la información.

Tabla 1: Descripción de los Actores del sistema

2.1.5 Necesidades de los usuarios.

Luego de un análisis del negocio que tiene lugar en la ONE se define como necesidad de los usuarios el análisis y la difusión de los indicadores relacionados con el área de construcción que se recopila anualmente a través de los siguientes archivos Excel:

- Viviendas terminadas por provincias.
- Obras terminadas.
- Cantidad y capacidad de presas en explotación por provincias.

2.1.6 Requisitos de información.

Ri1- Obtener información del total de viviendas terminadas por provincias, en el tiempo.

Ri2- Obtener información de la cantidad de viviendas estatales terminadas por provincias, en el tiempo.

Ri3- Obtener información de las viviendas no estatales terminadas por provincias, en el tiempo.

- Ri4-** Obtener información de las viviendas terminadas en las UBPC por provincias, en el tiempo.
- Ri5-** Obtener información de las viviendas terminadas en las CPA por provincias, en el tiempo.
- Ri6-** Obtener información de las viviendas terminadas en las CCS por provincias, en el tiempo.
- Ri7-** Obtener información de las viviendas terminadas de la población por provincias, en el tiempo.
- Ri8-** Obtener información de la estructura de las viviendas terminadas por provincias, en el tiempo.
- Ri9-** Obtener información de la dinámica de las viviendas terminadas por provincias, en el tiempo.
- Ri10-** Obtener información de la cantidad de obras terminadas por tipo de obra en el tiempo.
- Ri11-** Obtener información del valor de obras terminadas por tipo de obra en el tiempo.
- Ri12-** Obtener información de la estructura de la cantidad obras terminadas.
- Ri13-** Obtener información de la dinámica de la cantidad obras terminadas.
- Ri14-** Obtener información de la estructura del valor obras terminadas.
- Ri15-** Obtener información de la dinámica del valor obras terminadas.
- Ri16-** Obtener la cantidad de presas en explotación por provincias en el tiempo.
- Ri17-** Obtener la capacidad de presas en explotación por provincias en el tiempo.

2.1.7 Requisitos funcionales

- RF1-Extraer información de archivos Excel del área de construcción.
- RF2-Transformar y cargar información de archivos Excel del área de construcción.
- 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

2.1.8 Requisitos no funcionales.

Usabilidad

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

El sistema debe tener una interfaz gráfica uniforme a través del mismo 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 nomenclatura homogénea.

Las estructuras del mercado se nombrarán de una manera estándar teniendo en cuenta el tipo de estructura que se maneje. 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 el Almacén de Datos (Codificación y componentes de una solución de DWH).

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>

Tabla 2: Ejemplos de convenciones de nombrado

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 el siguiente elemento:

- 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.

2.1.9 Casos de uso del sistema

Para la confección del diagrama de caso de uso, se agruparon los 26 requisitos funcionales del sistema en casos de usos; y se definieron las relaciones existentes entre ellos y los actores del sistema (Ver figura 4).

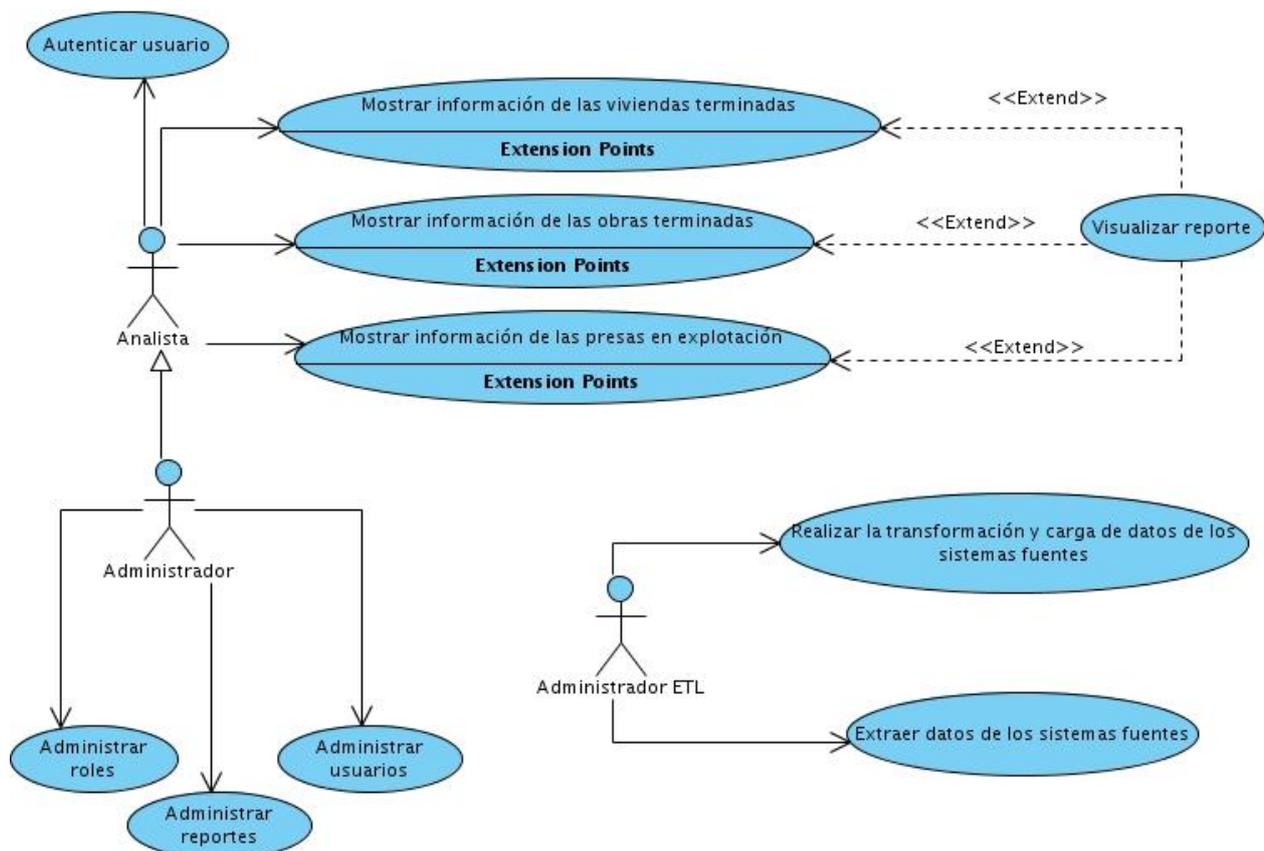


Figura 4: Diagrama de Casos de Uso

2.1.10 Descripción de los casos de usos críticos

Extraer datos de los sistemas fuentes

Caso de Uso:	Extraer datos de los sistemas fuentes
Tipo:	Funcional
Actores:	Administrador ETL
Resumen	Inicia cuando el Administrador ETL vaya a adicionar nuevos datos al mercado. Este caso de uso finaliza cuando los datos se encuentran en el área temporal.
Precondiciones:	Disponibilidad de las fuentes.
Referencias	RF 1
Prioridad	Crítico.
Flujo Normal de Eventos	
Acción del Actor	Respuesta del sistema
1. El usuario entra al PDI para realizar la extracción	1.1 El sistema levanta el área de trabajo y le da la posibilidad al usuario de cargar la transformación.
2. El usuario carga la transformación a ejecutar.	
3. El usuario configura los parámetros de entrada de la transformación	3.1 El sistema almacena los datos en el área temporal.
4. El usuario pre visualiza los datos	4.1 El sistema muestra los datos existentes en el área temporal
5. El usuario presiona el botón aceptar y cierra el área de trabajo	
Flujos Alternos	
Acción del Actor	Respuesta del sistema
	4.1 El sistema muestra un mensaje de error. Regresar al paso 3.
Pos condiciones	Datos disponibles para transformar.

Tabla 3: Descripción del caso de uso "Extraer datos de los archivos Excel"

Transformar y cargar datos de los sistemas fuentes

Caso de Uso:	Transformar y cargar datos de los sistemas fuentes
Tipo:	Funcional.

Actores:	Administrador ETL
Resumen	El caso de uso inicia cuando el actor desea realizar la transformación y carga de los datos almacenados en el área temporal, hacia el mercado de datos. El caso de uso finaliza cuando los datos son insertados satisfactoriamente en la Base de Datos.
Precondiciones:	<ul style="list-style-type: none"> • Los datos deben estar almacenados en el área temporal. • En la Base de Datos debe existir una estructura para almacenar la información.
Referencias	Ver requisito funcional 2
Prioridad	Crítico.
Flujo Normal de Eventos	
Acción del Actor.	Respuesta del sistema.
1. El administrador ETL interactúa con la herramienta PDI para realizar la extracción	1.1 El sistema levanta el área de trabajo y le da la posibilidad al usuario de cargar la transformación.
2. El administrador ETL carga la transformación a ejecutar.	
3. El administrador ETL configura las transformaciones definidas en las reglas del negocio.	3.1 El sistema realiza las transformaciones pertinentes
	3.2 El sistema almacena los datos en la Base de Datos
	3.3 El sistema muestra un mensaje de satisfacción en cada una de las operaciones realizadas
4. El administrador ETL cierra la ventana de trabajo	
Flujos Alternos	
Acción del Actor	Respuesta del Sistema
	3.3 El sistema muestra un mensaje de error y regresa al paso 3
Prototipo de Interfaz	
Poscondiciones	

Tabla 4: Descripción del caso de uso "Transformar y cargar datos"

Administrar Reportes

Caso de Uso:	Administrar Reportes	
Actores:	Administrador	
Resumen:	El caso de uso se inicia cuando el administrador interactúa con la aplicación para crear un nuevo reporte, actualizar o eliminar uno existente.	
Precondiciones:	El administrador debe estar autenticado.	
Referencia:	RF: 7, 8, 9	
Prioridad:	Crítica	
Flujo Normal de Eventos		
1. El administrador interactúa con la aplicación	1.1 El sistema le brinda la posibilidad de: a) Insertar reporte. b) Actualizar reporte c) Eliminar reporte.	
Sección Insertar Reportes		
Acción del Actor	Respuesta del Sistema	
2. El administrador selecciona la opción a)	2.1 El sistema muestra los esquemas y cubos disponibles en el almacén	
3. El administrador selecciona los datos deseados	3.1 El sistema carga los datos seleccionados	
	3.2 El sistema levanta un área de trabajo y le da la posibilidad al usuario de configurar el reporte (Ver caso de uso Visualizar Reporte)	
4. El administrador guarda el reporte		
Sección Modificar Reportes		
Acción del Actor	Respuesta del Sistema	

2. El administrador selecciona la opción b)	2.1 El sistema muestra los reportes existentes
3. El administrador selecciona el reporte deseado	3.1 El sistema carga el reporte en un área de trabajo
	3.2 El sistema le da la posibilidad al usuario de modificar las configuraciones del reporte (Ver caso de uso Visualizar Reporte)
4. El administrador modifica parámetros del reporte	4.1 El sistema visualiza los cambios en el reporte
5. El administrador guarda los cambios realizados	5.1 El sistema aplica los cambios realizados
6. El administrador cierra el área de trabajo	
Sección Eliminar Reportes	
Acción del Actor	Respuesta del Sistema
2. El administrador selecciona la opción c)	2.1 El sistema muestra los reportes existentes.
3. El administrador selecciona el reporte a eliminar	3.1 El sistema levanta una ventana de confirmación
4. El administrador acepta eliminar el reporte	4.1 El sistema actualiza el libro de trabajo.
Flujo Alternativo	
4. El administrador cancela la solicitud de eliminación.	
Poscondiciones	

Tabla 5: Descripción del caso de uso "Administrar Reportes"

2.2 Diseño

El proceso de diseño de un almacén define el modelo conceptual, lógico, físico y de visualización (BI) del mismo. En el modelo conceptual se especifica los requerimientos del usuario a través de las dimensiones, hechos y medidas que componen el almacén; en el modelo lógico se genera el modelo

de datos; en el modelo físico se diseñan las transformaciones, y en el de visualización se diseñan reportes candidatos por cada uno de los libros de trabajo definidos en el área de análisis (Ver figura 5).

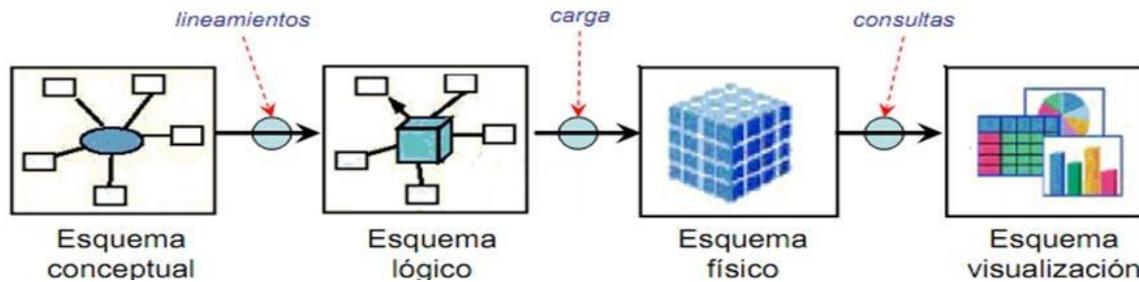


Figura 5: Proceso de diseño

2.2.1 Modelo conceptual

2.2.1.1 Dimensiones

Las dimensiones son las características de un hecho, que permiten su posterior análisis en el proceso de toma de decisiones; una entidad de negocio respecto a la cual se deben calcular las métricas.

Las dimensiones definidas son las siguientes:

- Dim_temporal_anno: dimensión que representa el año.
- Dim_provincia: dimensión que representa las provincias del país.
- Dim_tipo_obra: dimensión que representa los distintos tipos de obras que se pueden construir.
- Dim_dpa: dimensión que representa la a que región pertenecen los reportes de obras terminadas.

2.2.1.2 Medidas

Las medidas son los valores de datos numéricos que se analizan, son las variables de salida en el diseño, de ahí que representen lo contable que se necesita conocer. En este mercado se definieron las siguientes medidas:

- total_viviendas: total de viviendas construidas

- estatal: viviendas estatales construidas
- no_estatal: viviendas no estatales construidas
- ubpc: cantidad de UBPC construidas en el país
- cpa: cantidad de CPA construidas en el país
- ccs: cantidad de CCS construidas en el país
- población: viviendas de la población construidas
- cantidad_presas: cantidad de presas existentes en el país
- capacidad_presas: capacidad de las presas existentes en el país
- cantidad_obras: cantidad de obras construidas
- valor_obras: valor de las obras construidas

2.2.1.3 Tabla de hechos

Las tablas de hechos contienen las dimensiones y las medidas asociadas al mismo (38). La tabla de hechos tiene una clave primaria compuesta por las claves foráneas de las tablas dimensiones asociadas a él (Ver tabla 6).

En este mercado se definieron tres hechos:

- hech_viviendas_terminadas: en este hecho se guardan la información de las viviendas terminadas.
- hech_presas_en_explotacion: en este hecho se guarda la información de las presas en explotación existentes en el país.
- hech_obras_terminadas: en este hecho se guarda la información de las obras terminadas en el país.

Hechos	Dimensiones Asociadas	Medidas
		-total_viviendas

hech_viviendas_terminadas	-dim_provincia - dim_temporal_anno	-estatal -no_estatal -ubpc -cpa -ccs -poblacion
hech_presas_en_explotacion	-dim_provincia - dim_temporal_anno	-cantidad_presas -capacidad_presas
hech_obras_terminadas	-dim_provincia - dim_temporal_anno -dim_tipo_obra	-cantidad_obras -valor_obras

Tabla 6: Tabla de hechos

Matriz BUS

La matriz BUS representa la relación que existe entre los hechos y las dimensiones del Mercado (Ver tabla 7).

Hechos/Dimensiones	Dim_dpa	Dim_provincia	Dim_temporal_anno	Dim_tipo_obra
Hech_viviendas_terminadas		X	X	
Hech_presas_en_explotacion		X	X	
Hech_obras_terminadas	X		X	X

Tabla 7: Matriz BUS

2.2.2 Modelo de Datos

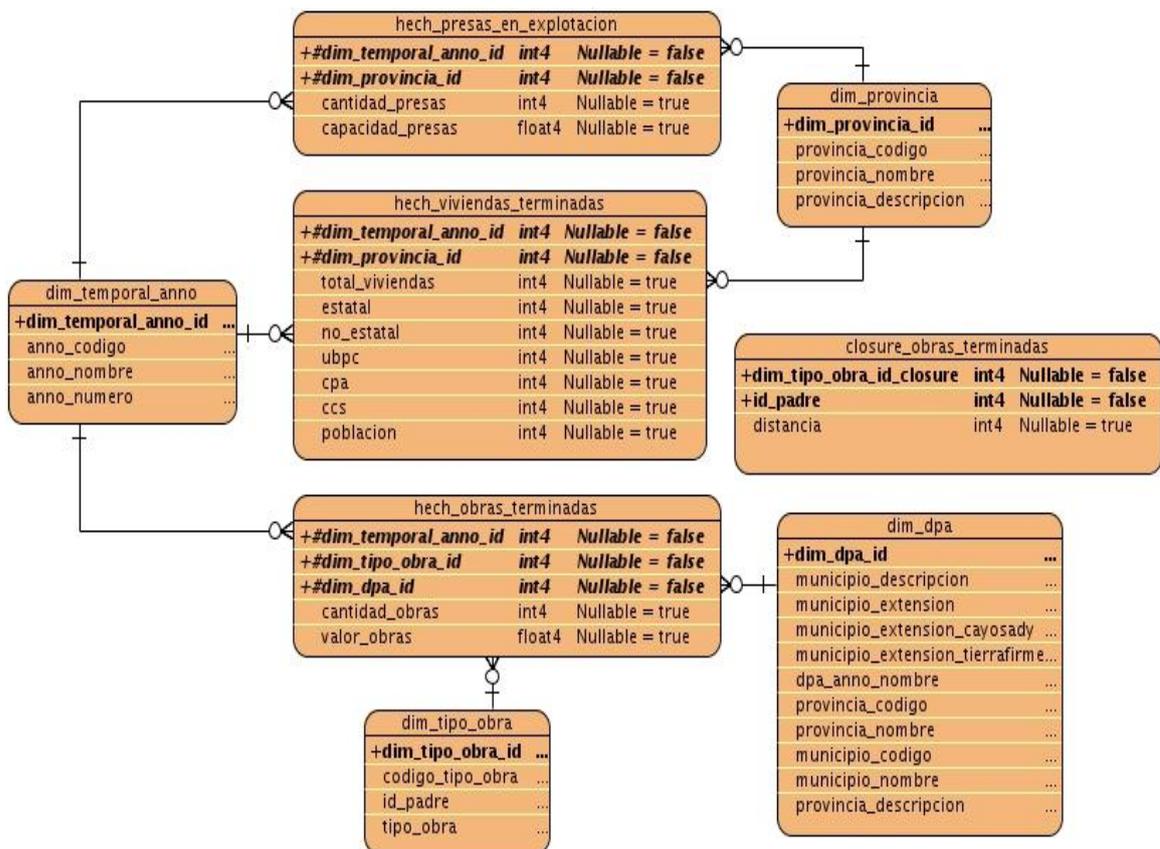


Figura 6: Modelo de datos

2.2.3 Diseño de las transformaciones

Las transformaciones son realizadas para el proceso de extracción, transformación y carga de los datos. Una vez ejecutada dicha transformación, se valida un grupo de restricciones definidas por el analista. En caso de no cumplirse una de estas restricciones, los datos se almacenan en un fichero Excel para darle tratamiento; en caso contrario, todos los datos se insertan en la tabla que tiene como salida dicha transformación (Ver figura 7).

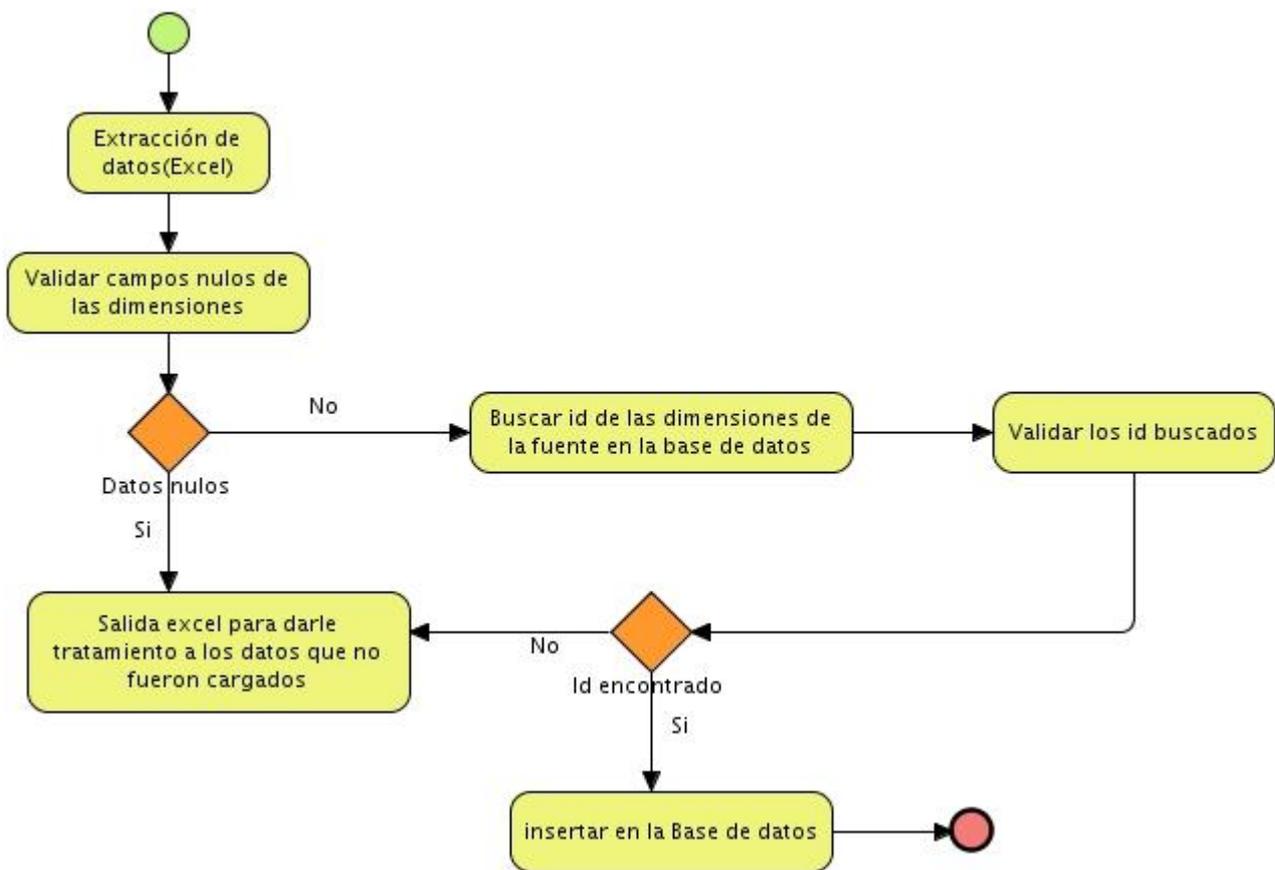


Figura 7: Diseño de las transformaciones

2.2.4 Diseño de los reportes candidatos

Los reportes candidatos responden a las necesidades de los usuarios, permitiendo identificar los reportes y demás salidas que debe producir el sistema. Estos son implementados en la Inteligencia de Negocio.

En el presente trabajo se diseñaron seis reportes candidatos agrupados en tres libros de trabajos (Ver anexo 5). A continuación se muestra un ejemplo del reporte candidato: “Viviendas terminadas”.

Area de análisis (AA)	Construcción
Libro de Trabajo (LT)	LT3 – Viviendas terminadas
Reporte (Tabla de Salida – TS)	TS2 – Viviendas terminadas
Descripción	Reporte que muestra la cantidad de viviendas terminadas en el país, ya sean estatales o no estatales.
Elementos del reporte	Provincia - dim_provincia Tiempo - dim_temporal_anno
Frecuencia de emisión	
Funciones	
Gráfico	Barras verticales 3D

Tabla 8: Reporte candidato "Viviendas terminadas"

2.3 Políticas de seguridad

2.3.1 Roles y permisos.

Para el acceso al mercado de datos se define un usuario por cada uno de los roles existentes en el sistema, con el objetivo de garantizar un control de acceso basado en roles (patrón RBAC), y así cada usuario, opera en el sistema según los permisos que se le definan al rol (Ver tabla 8).

Roles	Permisos	
	Lectura	Escritura
Sobre la Base de Datos		
Administrador	x	
Administrador ETL	x	x
Analista	x	
Sobre la aplicación		
Administrador	x	x
Analista	x	

Tabla 9: Roles y permisos

2.3.2 Salva de la base de datos

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 su seguridad y uso, será responsabilidad del grupo de administración de redes de la ONE.

2.4 Conclusiones

En este capítulo se crearon todos los artefactos necesarios para la implementación del mercado de datos:

- El negocio y sus reglas
- Los temas de análisis
- Requisitos de información, funcionales y no funcionales
- El proceso de diseño
- Y las políticas de seguridad definidas

CAPÍTULO 3: Implementación y pruebas

La implementación del mercado, está muy relacionada a los procesos de diseño del almacén. Ya que esta etapa del sistema, implementa el modelo físico y la capa de inteligencia de negocio: En el modelo físico se genera y puebla la base de datos, a través del script del mercado, y la implementación de las transformaciones; en análisis, se implementan y visualizan los reportes candidatos. En este capítulo se verán cada uno de los artefactos descritos anteriormente.

3.1 Implementación del modelo de datos

Es necesario tener una organización adecuada en el almacén de datos, por ello la implementación de las tablas del modelo de datos se dividieron en dos esquemas: “dimensiones” y “mart_construccion”, en los que se distribuyen las dimensiones y hechos del mercado (Ver tabla 10). En el esquema “dimensiones” se recogen las dimensiones que son comunes en el almacén y en el esquema “mart_construccion” los hechos y las dimensiones específicas del mercado.

Esquemas	Tablas
dimensiones	dim_provincia
dimensiones	dim_temporal_anno
dimensiones	dim_dpa
mart_construccion	closure_obras_terminadas
mart_construccion	dim_tipo_obra
mart_construccion	hech_viviendas_terminadas
mart_construccion	hech_obras_terminadas
mart_construccion	hech_presas_en_explotacion

Tabla 10: Esquemas y tablas del mercado

3.2 Implementación de los procesos de Extracción, Transformación y Carga

3.2.1 Implementación de las transformaciones

La transformación diseñada se aplicó a cada uno de los hechos del mercado (Ver anexo 6). Los componentes principales utilizados en cada una de ellas fueron: *data validator* para verificar la

existencia de campos nulos; y la búsqueda en base de datos para encontrar los identificadores de las dimensiones y poder realizar la carga de los hechos. En la figura 8 se muestra la transformación del hecho viviendas terminadas.

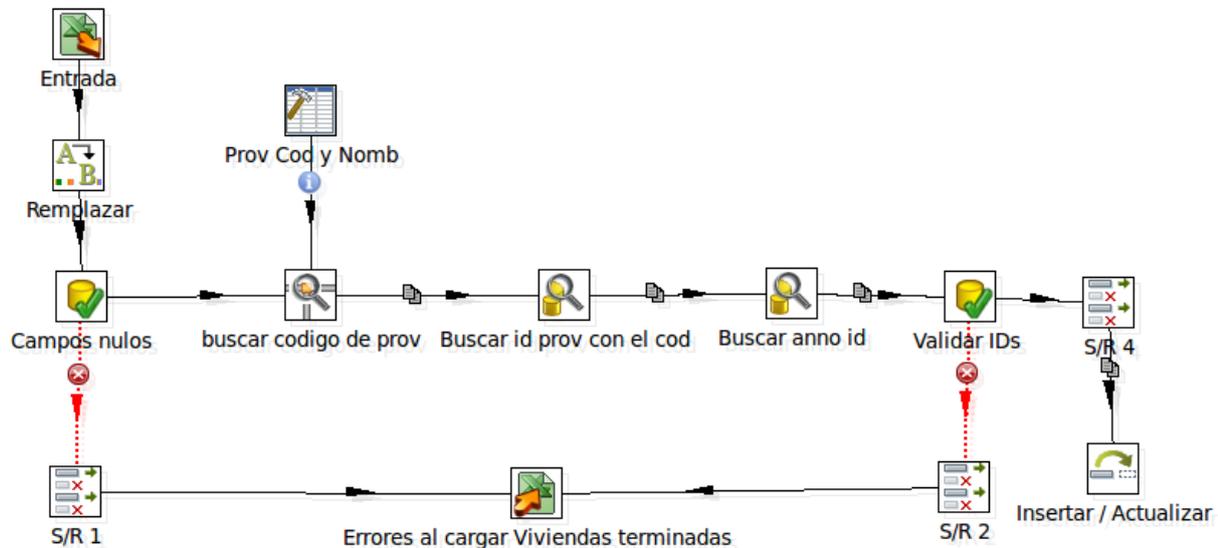


Figura 8: Transformación para cargar el hecho viviendas terminadas

3.2.2 Implementación de los trabajos

Al concluir la realización de todas las transformaciones necesarias para la carga de los datos, se realiza la implementación del trabajo, el cual se encarga de ejecutar todas las transformaciones en un orden lógico definido, primero las dimensiones y después los hechos (Ver figura 9). Al ejecutar el trabajo, este primeramente verifica la conexión con el servidor de base de datos, enviando un mensaje de error en caso de que no se pueda establecer la conexión y luego ejecuta cada una de las transformaciones definidas.

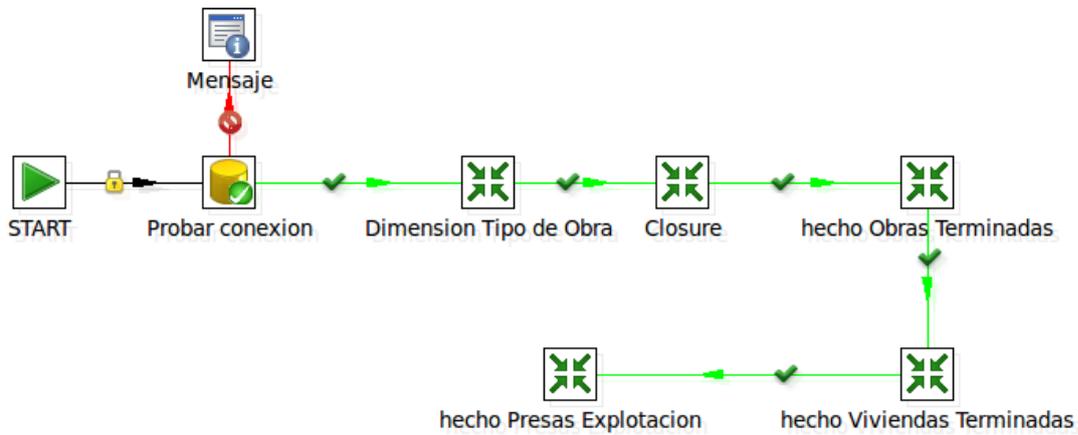


Figura 9: Trabajo para ejecutar las transformaciones

3.3 Implementación de los procesos de Inteligencia de Negocio

3.3.1 Implementación de los cubos OLAP

Para la implementación de los cubos, primero se definieron las dimensiones compartidas entre ellos, y luego se implementó un cubo por cada uno de los hechos del mercado (Ver figura 10).



Figura 10: Diseño de los cubos OLAP

3.3.2 Implementación de los reportes

En la implementación de los reportes se utilizaron consultas **mdx**. A continuación se muestra la consulta del reporte “Estructura de obras terminadas”, la cual devuelve la estructura de la cantidad y valor de obras por tipos de obras y año.

```
with set [dim_tipo_obra_set] as 'Union({[dim_tipo_obra].[Redes eléctricas]},
Union({[dim_tipo_obra].[Oleoductos y gasoductos]}, Union({[dim_tipo_obra].[Vías férreas]},
Union({[dim_tipo_obra].[Construcciones marítimas]}, Union({[dim_tipo_obra].[Redes de
comunicación]}, Union({[dim_tipo_obra].[Dragado]}, Union({[dim_tipo_obra].[Agropecuarias]},
Union([dim_tipo_obra].[Obras para el turismo].Children, {[dim_tipo_obra].[Obras para el
turismo]}))))))'
set [dim_tipo_obra_diferencia] as 'Except(Descendants([dim_tipo_obra].[Todos], 1.0),
[dim_tipo_obra_set])'
member [dim_tipo_obra].[Otras obras] as 'Sum([dim_tipo_obra_set])'
member [Measures].[Estructura de la cantidad de obras] as '((([dim_tipo_obra].CurrentMember,
[Measures].[cantidad_obras]) / ([dim_tipo_obra].[Todos], [Measures].[cantidad_obras]))',
FORMAT_STRING = "|0.00 %"
member [Measures].[Estructura del valor de obras] as '((([dim_tipo_obra].CurrentMember,
[Measures].[valor_obras]) / ([dim_tipo_obra].[Todos], [Measures].[valor_obras]))', FORMAT_STRING
= "|0.00 %"
select Crossjoin({[dim_dpa].[Nacional]}, {[Measures].[Estructura de la cantidad de obras],
[Measures].[Estructura del valor de obras]}) ON COLUMNS,
Crossjoin({[dim_temporal_anno].[2009]}, {[dim_tipo_obra].[Educativas], [dim_tipo_obra].[Obras
para la salud], [dim_tipo_obra].[Otras edificaciones], [dim_tipo_obra].[Construcciones Industriales],
[dim_tipo_obra].[Viales], [dim_tipo_obra].[Hidráulicas], [dim_tipo_obra].[Hidrológicas],
[dim_tipo_obra].[Otras obras]}) ON ROWS
from [Obras_Terminadas]
```

3.3.3 Vistas de análisis

La arquitectura de información del mercado de datos Construcción está definida por tres libros de trabajo, uno por cada hecho, en los cuales van a estar los reportes asociados a cada uno de ellos. A continuación se muestra la vista de análisis “Estructura de obras” perteneciente al libro de trabajo “LT Obras terminadas” (Ver figura 11).

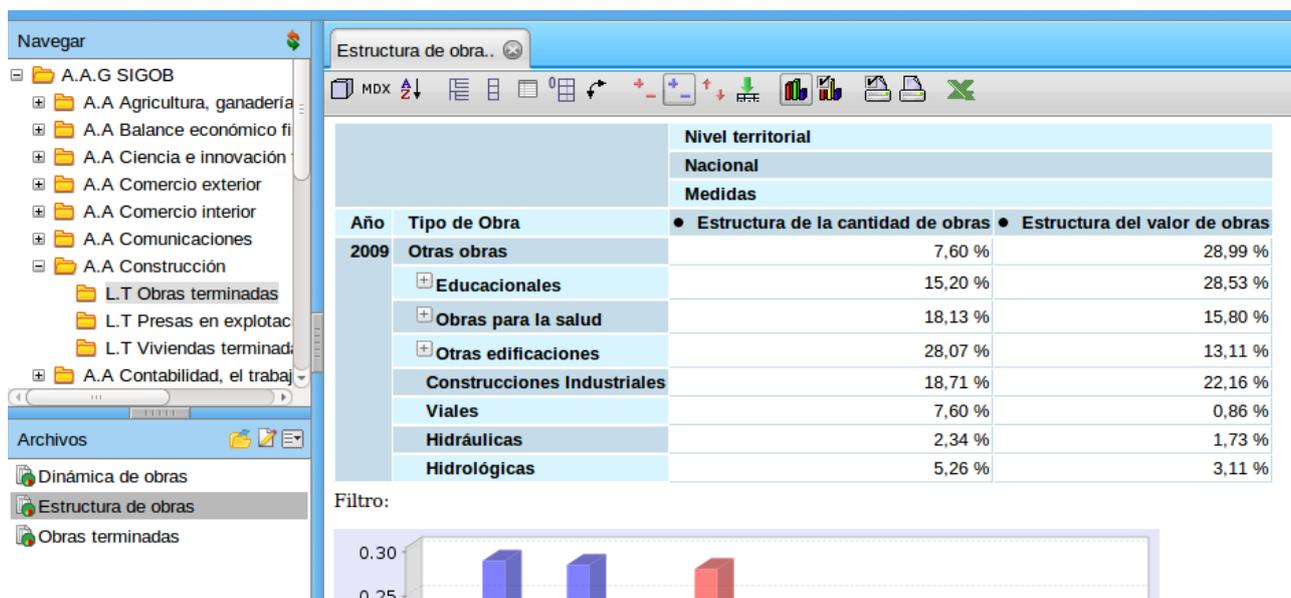


Figura 11: Vista del reporte Estructura de Obras

3.4 Pruebas

Las pruebas del software constituyen un elemento fundamental para garantizar la calidad del producto, de esta manera se puede verificar si el sistema cumple con lo esperado por el cliente. Para evaluar la funcionalidad, usabilidad, confiabilidad y desempeño del mercado (cuán rápido visualiza un reporte bajo condiciones particulares de trabajo), se diseñaron las siguientes pruebas:

- Funcionalidad: Se realizó una carta de aceptación por parte del cliente con el objetivo de validar que las funcionalidades del mercado, cumplieran con los requisitos del sistema.
- Calidad: Se diseñó una prueba de chequeo a los artefactos del expediente de proyecto con el objetivo de verificar la calidad del producto.
- Usabilidad: Se diseñó una prueba de Interfaz con el objetivo de garantizar que el sistema provee el acceso apropiado para interactuar con el mercado.
- Confiabilidad y desempeño: Se diseñaron pruebas de rendimiento para evaluar características de funcionamiento, tales como, tiempos de respuesta a un reporte, y otros elementos sensibles al tiempo.

3.4.1 Prueba de Interfaz

Para garantizar la usabilidad del sistema se realizaron las pruebas de interfaz aplicando los casos de pruebas (Ver anexo 1, 2 y 3). Se realizaron hasta el momento tres iteraciones de estos casos de prueba, a nivel departamental, a nivel de centro y Calidad UCI los cuales arrojaron los siguientes resultados:

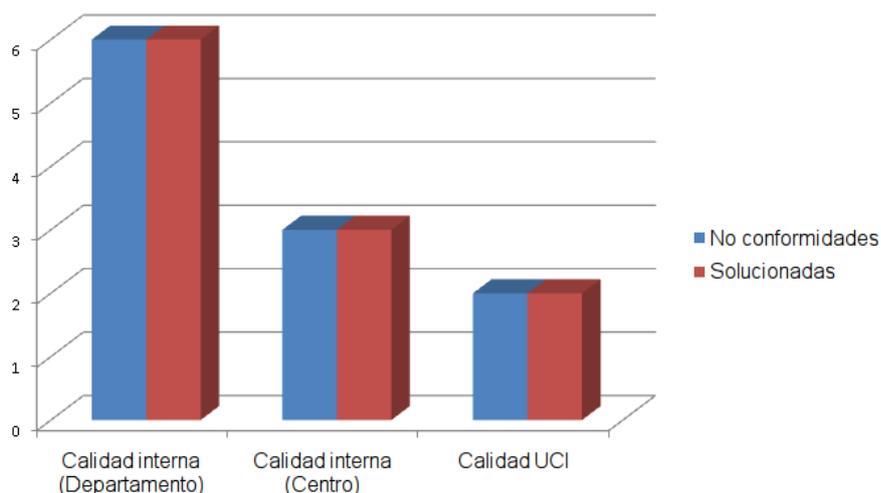


Figura 12: Resultado de las pruebas de interfaz

3.4.2 Lista de chequeo

La lista de chequeo definida para la evaluación de la calidad del mercado de datos está dividida en tres secciones: Estructura del documento, Indicadores definidos en el desarrollo y Semántica del documento (Ver anexo 4). De forma general se identificaron cinco indicadores, de ellos tres críticos. Al aplicarse la lista se generaron 4 no conformidades (ver expediente de proyecto), las cuales fueron solucionadas satisfactoriamente.

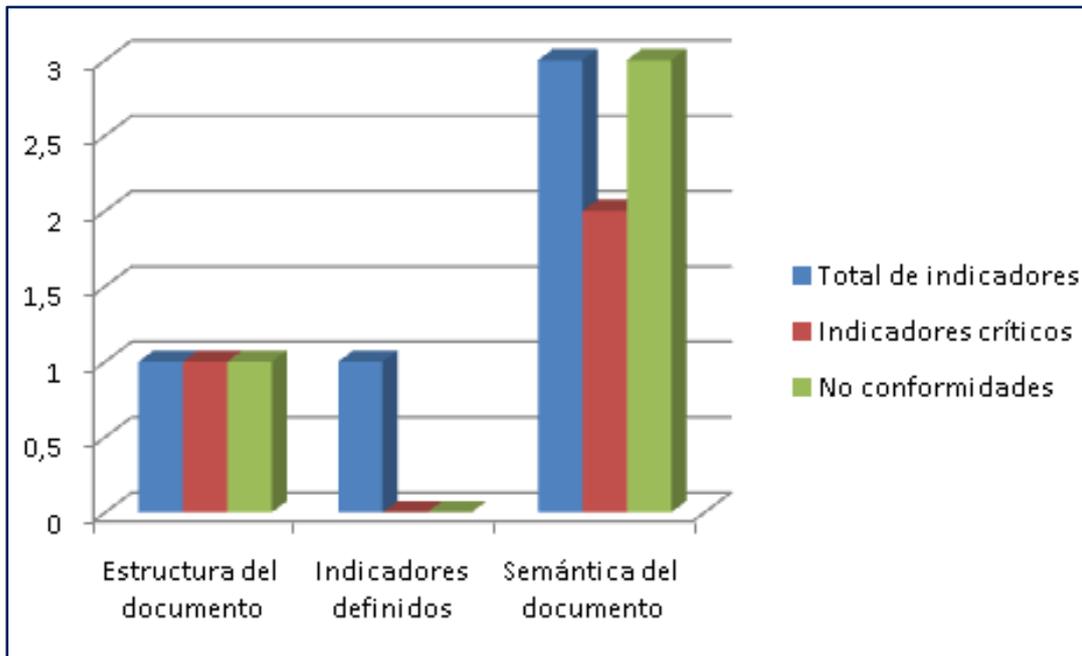


Figura 13: Comportamiento de los indicadores por secciones

3.4.3 Pruebas de aceptación

Una vez finalizado el desarrollo del mercado, se realizó las pruebas de aceptación por la especialista Elena Leonila Fernández García (representante de la Oficina Nacional de Estadística en la universidad), quien al interactuar con el sistema confirmó que el mercado satisface todas las necesidades de información identificadas con anterioridad. Dichos resultados quedaron oficializados con la carta de aceptación del cliente.

3.4.4 Pruebas de rendimiento

En las aplicaciones web, además de validar los requisitos funcionales, existen varios aspectos adicionales a analizar, como por ejemplo, el rendimiento del servidor, la facilidad de navegación y el tiempo de respuesta a eventos, con el objetivo de determinar no solo la velocidad, sino la eficiencia de los mismos.

Las limitaciones en los tiempos de respuesta de un sitio web están muy relacionadas a la velocidad del enlace donde se esté “navegando” (39). Según el autor Jakob Nielsen, en el libro “Usability Engineering” (40) existen tres límites importantes: (39)

- 0,1 segundo: es el límite en el cual el usuario siente que esta “manipulando” los objetos desde la interfaz de usuario.
- 1 segundo: es el límite en el cual el usuario siente que está navegando libremente sin esperar demasiado una respuesta del servidor.
- 10 segundos: es el límite en el cual se pierde la atención del usuario, si la respuesta tarda más de 10 segundos se deberá indicar algún mecanismo por el cual el usuario pueda interrumpir la operación.

Para corroborar el tiempo de visualización de un reporte, se utilizó la herramienta jakarta-jmeter en su versión 2.3.4 en una máquina Pentium III de 1GB de RAM y 160 GB de disco duro. La conexión al servidor de aplicaciones y al de bases de datos, se realizó desde una máquina remota.

El jmeter es programa de código abierto, implementada en Java que permite realizar test de comportamiento funcional y medir el rendimiento. Originalmente fue diseñado para probar las aplicaciones Web, pero desde entonces se ha expandido a otras funciones de prueba (41). También se puede utilizar para realizar pruebas de carga y estrés a los servidores (39).

Plan de Prueba

Se configuró un plan de prueba con la visualización de los 6 reportes implementados en el mercado. Para ello primero se creó un servidor proxy HTTP, se configuró un acceso de 25, 50 y 100 usuarios conectados concurrentemente al mercado y se definieron dos *listener* para visualizar los resultados y posibles errores detectados (Ver figura 14).

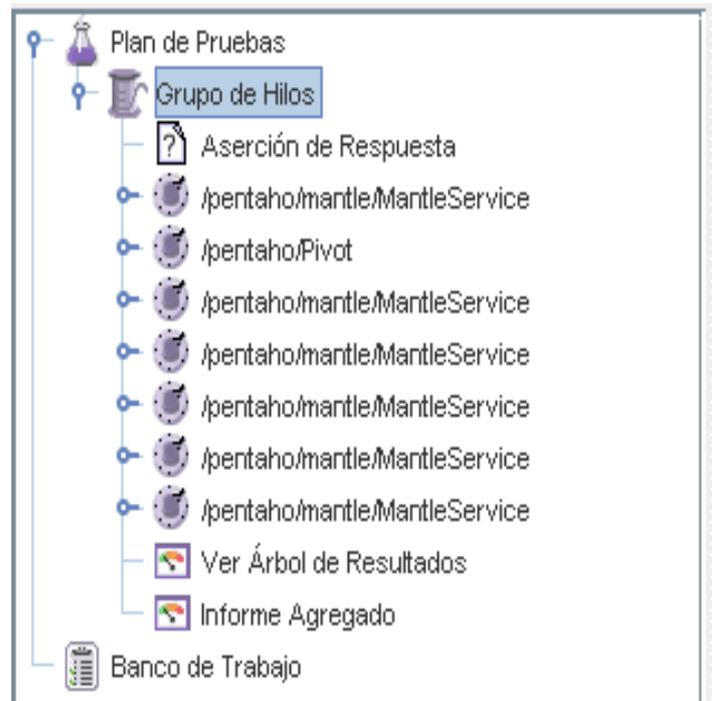


Figura 14: Plan de pruebas

Resultados de la prueba

Luego de concluidas las pruebas de rendimiento se obtuvo como tiempos promedios de respuesta del reporte de 24, 58 y 105 ms para 25, 50 y 100 accesos de usuarios respectivamente. A continuación se muestran los resúmenes correspondientes a dichas pruebas.

Label	# Muestras	Media	Mediana	Linea de 9...	Mín	Máx	% Error	Rendimie...	Kb/sec
/pentaho/mantle/MantleService	150	25	19	45	6	149	0,00%	59,8/sec	445,6
/pentaho/ViewAction	150	24	17	46	6	141	0,00%	60,1/sec	459,9
/pentaho/Pivot	150	23	17	42	6	151	0,00%	60,2/sec	459,6
/pentaho/js/webcontext.js	300	24	18	45	4	148	0,00%	118,9/sec	885,1
/pentaho/adhoc/js/common/ui/messages/nls/es/messa...	150	24	17	47	6	154	0,00%	60,3/sec	453,1
/pentaho/adhoc/js/common/ui/messages/nls/es-ar/mes...	150	22	18	42	6	144	0,00%	60,3/sec	454,0
/pentaho/adhoc/js/ui/messages/nls/es/message_string...	150	21	18	40	6	121	0,00%	60,3/sec	452,8
/pentaho/adhoc/js/ui/messages/nls/es-ar/message_stri...	150	26	19	48	5	158	0,00%	60,5/sec	454,5
/pentaho/jpivot/toolbar/non-empty-up.png	25	22	10	43	6	155	0,00%	16,1/sec	120,7
/pentaho/DisplayChart	150	24	18	44	5	139	0,00%	60,7/sec	456,6
TOTAL	1525	24	18	44	4	158	0,00%	592,9/sec	4459,1

Figura 15: Resultados para 25 usuarios

Label	# Muestras	Media	Mediana	Línea de 9...	Mín	Máx	% Error	Rendimie...	Kb/sec
/pentaho/mantle/MantleService	300	61	47	134	7	286	0,00%	59,5/sec	443,6
/pentaho/ViewAction	300	58	50	123	6	194	0,00%	59,6/sec	455,9
/pentaho/Pivot	300	57	43	129	6	211	0,00%	59,6/sec	455,2
/pentaho/js/webcontext.js	600	58	47	125	5	236	0,00%	118,4/sec	881,9
/pentaho/adhoc/js/common/ui/messages/nls/es/messa...	300	56	45	121	6	255	0,00%	59,6/sec	448,2
/pentaho/adhoc/js/common/ui/messages/nls/es-ar/mes...	300	55	45	124	6	195	0,00%	59,6/sec	448,5
/pentaho/adhoc/js/ui/messages/nls/es/message_string...	300	58	46	122	6	220	0,00%	59,7/sec	447,7
/pentaho/adhoc/js/ui/messages/nls/es-ar/message_stri...	300	56	45	124	5	278	0,00%	59,7/sec	448,3
/pentaho/jpivot/toolbar/non-empty-up.png	50	57	40	148	8	177	0,00%	17,9/sec	133,6
/pentaho/DisplayChart	300	59	47	131	6	282	0,00%	60,0/sec	451,6
TOTAL	3050	58	46	126	5	286	0,00%	59,7,3/sec	4492,3

Figura 16: Resultados para 50 usuarios

Label	# Muestras	Media	Mediana	Línea de 9...	Mín	Máx	% Error	Rendimie...	Kb/sec
/pentaho/mantle/MantleService	600	119	82	288	5	460	0,00%	59,7/sec	445,2
/pentaho/ViewAction	600	107	67	273	5	463	0,00%	59,8/sec	457,2
/pentaho/Pivot	600	108	69	278	5	415	0,00%	59,8/sec	456,3
/pentaho/js/webcontext.js	1200	105	65	270	5	474	0,00%	118,4/sec	881,3
/pentaho/adhoc/js/common/ui/messages/nls/es/messa...	600	109	71	278	5	459	0,00%	59,7/sec	448,9
/pentaho/adhoc/js/common/ui/messages/nls/es-ar/mes...	600	100	62	262	6	460	0,00%	59,6/sec	448,2
/pentaho/adhoc/js/ui/messages/nls/es/message_string...	600	104	69	266	5	378	0,00%	59,4/sec	445,9
/pentaho/adhoc/js/ui/messages/nls/es-ar/message_stri...	600	97	48	261	5	396	0,00%	59,5/sec	446,6
/pentaho/jpivot/toolbar/non-empty-up.png	100	123	97	283	6	394	0,00%	24,4/sec	182,2
/pentaho/DisplayChart	600	95	50	265	6	451	0,00%	59,8/sec	449,6
TOTAL	6100	105	64	271	5	474	0,00%	59,9,3/sec	4507,3

Figura 17: Resultados para 100 usuarios

3.6 Conclusiones

En el capítulo se describieron los elementos de implementación para la construcción del mercado de datos

- ✓ Como resultado de la implementación del proceso ETL, se logró poblar el mercado de manera satisfactoria.
- ✓ Se implementaron los cubos OLAP necesarios para satisfacer los requisitos de información del cliente.
- ✓ Se desarrolló el subsistema de visualización con los reportes básicos solicitados por el cliente.
- ✓ Se realizaron pruebas de funcionalidad, calidad, usabilidad y desempeño al mercado, con buenos resultados en cada una de ellas.
- ✓ Los tiempos de respuestas promedios de visualizar un reporte con 100 usuarios conectados concurrentes son de 105 ms.

CONCLUSIONES

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

- Se definieron los requisitos de información y funcionales necesarios para la obtención de un modelo lógico que sustentara los datos contenidos en las fuentes.
- Se implementó el mercado de datos del área de construcción de la ONE, logrando un formato estándar, organizado y centralizado de los datos recogidos en todo el país a lo largo de los años.
- Los tiempos de respuestas del sistema –durante la visualización de cada uno de los reportes del mercado - son aceptables para la cantidad de usuarios que laboran en el área.
- El sistema implementado se validó a través de la lista de chequeo, los casos de pruebas diseñados y la carta de aceptación del cliente.

RECOMENDACIONES

- Que se ponga en explotación el mercado de datos 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.

REFERENCIAS BIBLIOGRÁFICAS

1. Lefcovich, Mauricio. Estadística aplicada a los negocios. [Online] [Cited: 6 5, 2011.] <http://www.monografias.com/trabajos34/estadistica-negocios/estadistica-negocios.shtml>.
2. Ferrer, Alejandro Tiana. Tratamiento y Usos de la Información en Evaluación. [Online] [Cited: 05 30, 2011.] <http://www.campus-oei.org/calidad/tiana.htm>.
3. Vargas, Rafael Kauffmann. Estadística básica para supervisores aplicada al control de procesos. [Online] [Cited: 06 08, 2011.] <http://www.mailxmail.com/curso-estadistica-basica-supervisores-aplicada-control-procesos/estadistica-importancia-toma-decisiones>.
4. Esther López Vizcaíno, Pilar Romero Martínez, Rosa Díaz Doutón, Antonio Delgado Jul y José Méndez Martínez. ¿CÓMO POTENCIAR EL USO DE LA INFORMACIÓN ESTADÍSTICA? [Online] [Cited: 06 02, 2011.] http://www.google.com/url?sa=t&source=web&cd=7&ved=0CE8QFjAG&url=http%3A%2F%2Fwww.ican.es%2Fopencms%2Fopencms%2Fponencias%2FGALICIA%2FCOMO_POTENCIAR_EL_USO_DE_LA_INFORMACION_ESTADISTICA.doc&rct=j&q=como%20usar%20la%20informacion%20estadistica%20en%20las%.
5. Las técnicas del almacenamiento de datos han ido evolucionando a lo largo de los años. [Online] [Cited: 06 01, 2011.] <http://es.scribd.com/doc/55287696/Las-tecnicas-del-almacenamiento-de-datos-han-ido-evolucionando-a-lo-largo-de-los-anos>.
6. Rodríguez, Héctor Julio. Historia de las bases de datos en Ciencia de la Información. [Online] [Cited: 06 02, 2011.] http://recursostic.javeriana.edu.co/wiki/index.php/Historia_de_las_bases_de_datos_en_Ciencia_de_la_Informaci%C3%B3n.
7. Tipos De Base De Datos. Naturaleza Y Representación. *buenastareas.com*. [Online] [Cited: 06 04, 2011.] <http://www.buenastareas.com/ensayos/Tipos-De-Base-De-Datos-Naturaleza-y/12004.html>.

8. Ortiz, Marta Cecilia Ortiz. La inteligencia de negocios aplicada a las organizaciones en Latinoamérica. [Online] [Cited: 06 09, 2011.]
http://www.google.com/url?sa=t&source=web&cd=1&ved=0CBcQFjAA&url=http%3A%2F%2Fwww2.epm.com.co%2Fbibliotecaepm%2Fbiblioteca_virtual%2Fdocuments%2Fla_inteligencia_de_negocios_aplicada_a_las_organizaciones.pdf&rct=j&q=La%20inteligencia%20de%20negocios%20es%2.
9. Montoya, Lic. Ricardo Sánchez. Business intelligence... to bi or not to bi. [Online] [Cited: 05 30, 2011.] <http://www.gestiopolis.com/recursos/documentos/fulldocs/ger1/tobi.htm>.
10. Qué es estadística? [Online] [Cited: 10 10, 2010.] <http://www.misrespuestas.com/que-es-estadistica.html>.
11. Lefcovich, Mauricio León. Estadística aplicada a los negocio. [Online]
<http://manuelgross.bligoo.com/content/view/218974/La-estadistica-es-fundamental-para-la-gestion-eficiente.html>.
12. Chuc-Durán, Diana Graciela. Introducción a los Datawarehouses. [Online] [Cited: 05 25, 2011.]
http://www.publicaciones.ujat.mx/publicaciones/revista_dacb/Acervo/v6n1OL/v6n1a5-ol/index.html.
13. *Análisis, Diseño e Implementación del mercado de datos Indicadores relacionados con el Comercio Exterior*. Daimel Rubén González Alarcón, Reinier Cárdenas Ramírez. Habana : s.n., 2010.
14. *GIGA*. Habana : s.n. 1028-270x.
15. Ventajas e inconvenientes de los sistemas de bases de datos. [Online] [Cited: 01 10, 2011.]
<http://www3.uji.es/~mmarques/f47/apun/node7.html>.
16. ¿En que se clasifican las bases de datos? [Online] [Cited: 01 10, 2011.]
http://www.netronycs.com/clasificacion_de_base_datos.html.
17. Geiger, Claudia Imhoff. Nicholas Galemno. Jonathan G. *Mastering Data Warehouse Design, Relational and Dimentional Techniques*. 2003. 0-471-32421-3.
18. Herrera, Cristhian. Características del almacén de datos. [Online] 10 30, 2007.
<http://www.adictosaltrabajo.com/tutoriales/tutoriales.php?pagina=datawarehouse>.

19. Peñaloza, Lucía Victoria Hernández. Tesis para lograr el título de Magíster: Diseño y Construcción de un Data Mart para la mantención de Indicadores de Sostenibilidad de la Industria del Salmón. Chile. Chile : s.n., 2008.
20. Ross, Ralph Kimball y Margy. *The Data Warehouse Toolkit*. 2002.
21. Cavazos, Eliezer. Metodología para Business Intelligence. [Online] [Cited: 06 02, 2011.] <http://www.gravitar.biz/index.php/bi/metodologia-business-intelligence/>.
22. La metodología. [Online] [Cited: 01 10, 2011.] <http://www.ecured.cu/index.php/Metodolog%C3%ADa>.
23. Kimball vs Inmon. Ampliación de conceptos del Modelado Dimensional. [Online] [Cited: 01 10, 2011.] <http://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-del-modelado-dimencional/>.
24. Autores, Colectivo de. *"Metodología para el desarrollo de soluciones de Almacenes de Datos e Inteligencia de Negocio en CENTALAD"*. 2009.
25. Modelamiento multidimensional. [Online] [Cited: 05 30, 2011.] <http://www.google.com/url?sa=t&source=web&cd=1&ved=0CBsQFjAA&url=http%3A%2F%2Fwww.inf.udec.cl%2F~revista%2Fediciones%2Fedicion4%2Fmodmulti.PDF&rct=j&q=modelo%20multidimensional%20&ei=PnDuTaTdB6L00gGztszfAw&usg=AFQjCNGUM1nUN0UrK8yOYBuxFI3dXBnnnQ&sig2=iGxzY>.
26. Tablas de hechos. [Online] [Cited: 06 03, 2011.] <http://msdn.microsoft.com/es-es/library/ms244679%28v=vs.80%29.aspx>.
27. Tablas de dimensiones. [Online] [Cited: 06 03, 2011.] <http://msdn.microsoft.com/es-es/library/ms244714%28v=vs.80%29.aspx>.
28. gestiopolis. [Online] [Cited: 11 22, 2010.] <http://www.gestiopolis.com/canales8/ger/olap-online-analytic-processing.htm>.

29. Visual Paradigm for UML (ME). [Online] [Cited: 11 25, 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/).
30. Sitio de Postgres. [Online] [Cited: 11 23, 2010.] <http://www.postgresql.org/>.
31. Sitio de PgAdmin. [Online] [Cited: 11 25, 2010.] <http://www.pgadmin.org/>.
32. El servidor Apache Tomcat. [Online] [Cited: 11 20, 2010.] <http://casidiablo.net/el-servidor-apache-tomcat/>.
33. La Suite de Pentaho. [Online] [Cited: 11 30, 2010.] <http://demetole.blogspot.com/2010/05/la-suite-pentaho.html>.
34. Pentaho BI Plataform Server. [Online] [Cited: 12 01, 2010.]
<http://www.summan.com/index.php/productos/software/pentaho-.html>.
35. Características. [Online] [Cited: 11 22, 2010.] <http://www.gravitar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/>.
36. Ochoa, Darián González. Diseño e Implementación de un Almacén de Datos Operacionales para la corporación CIMEX. . Ciudad de La Habana. : Universidad de las Ciencias Informáticas.
37. PENTAHO: Creación de un cubo con Mondrian. . [Online] [Cited: 11 24, 2010.]
<http://www.gravitar.biz/index.php/bi/pentaho-ejemplo-cubo-mondrian/>.
38. Díaz., Mg. Samuel Oporto. [Online] 08 22, 2009. [Cited: 02 25, 2011.] www.wiphala.net. .
39. F. Javier Diaz, Claudia M. Tzancoff Banchoff, Anahí S. Rodríguez, Valeria Soria. Usando Jmeter para pruebas de rendimiento. [Online] [Cited: 05 30, 2011.]
http://www.google.com/url?sa=t&source=web&cd=1&ved=0CBoQFjAA&url=http%3A%2F%2Fwww.linti.unlp.edu.ar%2Fuploads%2Fdocs%2Fusando_jmeter_para_pruebas_de_rendimiento.pdf&rct=j&q=Usando%20Jmeter%20para%20pruebas%20de%20rendimiento&ei=5-zyTduoGci4tge2p-mTBw&usg=.
40. Nielsen, Jakob. Usability Engineering. [Online] [Cited: 05 29, 2011.]
<http://www.useit.com/jakob/useengbook.html>.

41. Apache Jakarta Project. [Online] [Cited: 04 28, 2011.] <http://jakarta.apache.org/jmeter/>.

BIBLIOGRAFÍA

- ¿En que se clasifican las bases de datos? [Online] [Cited: 01 10, 2011.] http://www.netronyos.com/clasificacion_de_base_datos.html.
- Análisis, Diseño e Implementación del mercado de datos Indicadores relacionados con el Comercio Exterior . Daimel Rubén González Alarcón, Reinier Cárdenas Ramírez. Habana : s.n., 2010.
- Apache Jakarta Project. [Online] [Cited: 04 28, 2011.] <http://jakarta.apache.org/jmeter/>.
- Autores, Colectivo de. "METODOLOGÍA PARA EL DESARROLLO DE SOLUCIONES DE ALMACENES DE DATOS E INTELIGENCIA DE NEGOCIO EN CENTALAD". 2009.
- Características. [Online] [Cited: 11 22, 2010.] <http://www.gravitar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/>.
- Cavazos, Eliezer. Metodología para Business Intelligence. [Online] [Cited: 06 02, 2011.] <http://www.gravitar.biz/index.php/bi/metodologia-business-intelligence/>.
- Chuc-Durán, Diana Graciela. Introducción a los Datawarehouses. [Online] [Cited: 05 25, 2011.] http://www.publicaciones.ujat.mx/publicaciones/revista_dacb/Acervo/v6n1OL/v6n1a5-ol/index.html.
- DataCleaner. [Online] [Cited: 11 24, 2010.] <http://datacleaner.eobjects.org/>.
- Díaz., Mg. Samuel Oporto. [Online] 08 22, 2009. [Cited: 02 25, 2011.] www.wiphala.net.
- El servidor Apache Tomcat. [Online] [Cited: 11 20, 2010.] <http://casidiablo.net/el-servidor-apache-tomcat/>.
- Esther López Vizcaíno, Pilar Romero Martínez, Rosa Díaz Doutón, Antonio Delgado Jul y José Méndez Martínez. ¿CÓMO POTENCIAR EL USO DE LA INFORMACIÓN ESTADÍSTICA? [Online] [Cited: 06 02, 2011.] [http://www.google.com/url?sa=t&source=web&cd=7&ved=0CE8QFjAG&url=http%3A%2F%2Fwww.icanes.es%2Fopencms%2Fopencms%2Fponencias%2FGALICIA%2FCOMO_POTENCIAR_EL_USO_DE_LA_INFORMACION_ESTADISTICA.doc&rct=j&q=como%20usar%20la%20informacion%20estadistica%20en%20las%](http://www.google.com/url?sa=t&source=web&cd=7&ved=0CE8QFjAG&url=http%3A%2F%2Fwww.icanes.es%2Fopencms%2Fopencms%2Fponencias%2FGALICIA%2FCOMO_POTENCIAR_EL_USO_DE_LA_INFORMACION_ESTADISTICA.doc&rct=j&q=como%20usar%20la%20informacion%20estadistica%20en%20las%20).
- ETL-Tools.Info. [Online] [Cited: 11 05, 2010.] http://etl-tools.info/es/bi/almacenedatos_arquitectura.htm.
- F. Javier Diaz, Claudia M. Tzancoff Banchoff, Anahí S. Rodríguez, Valeria Soria. Usando Jmeter para pruebas de rendimiento. [Online] [Cited: 05 30, 2011.]

http://www.google.com/url?sa=t&source=web&cd=1&ved=0CBoQFjAA&url=http%3A%2F%2Fwww.linti.unlp.edu.ar%2Fuploads%2Fdocs%2Fusando_jmeter_para_pruebas_de_rendimiento.pdf&rct=j&q=Usando%20Jmeter%20para%20pruebas%20de%20rendimiento&ei=5-zyTduoGci4tge2p-mTBw&usg=.

- Falcon Rodriguez, Yolanda y Leyva Osorio, Reynaldo. Mercado de Datos Estadístico de Inmigración y Extranjería para el Departamento de Turismo y Comercio de la Oficina Nacional de Estadísticas. . Habana : s.n., 2010.
- Ferrer, Alejandro Tiana. Tratamiento y Usos de la Información en Evaluación. [Online] [Cited: 05 30, 2011.] <http://www.campus-oei.org/calidad/tiana.htm>.
- Geiger, Claudia Imhoff. Nicholas Gallempo. Jonathan G. Mastering Data Warehouse Design, Relational and Dimensional Techniques. 2003. 0-471-32421-3.
- gestiopolis. [Online] [Cited: 11 22, 2010.] <http://www.gestiopolis.com/canales8/ger/olap-online-analytic-processing.htm>.
- GIGA. Habana : s.n. 1028-270x.
- Herrera, Cristhian. Características del almacén de datos. [Online] 10 30, 2007. <http://www.adictosaltrabajo.com/tutoriales/tutoriales.php?pagina=datawarehouse>.
- Kimball vs Inmon. Ampliación de conceptos del Modelado Dimensional. [Online] [Cited: 01 10, 2011.] <http://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-del-modelado-dimENSIONAL/>.
- La metodología. [Online] [Cited: 01 10, 2011.] <http://www.ecured.cu/index.php/Metodolog%C3%ADa>.
- La Suite de Pentaho. [Online] [Cited: 11 30, 2010.] <http://demetole.blogspot.com/2010/05/la-suite-pentaho.html>.
- Las técnicas del almacenamiento de datos han ido evolucionando a lo largo de los años. [Online] [Cited: 06 01, 2011.] <http://es.scribd.com/doc/55287696/Las-tecnicas-del-almacenamiento-de-datos-han-ido-evolucionando-a-lo-largo-de-los-anos>.
- Lefcovich, Mauricio León. Estadística aplicada a los negocios. [Online] <http://manuelgross.bligoo.com/content/view/218974/La-estadistica-es-fundamental-para-la-gestion-eficiente.html>.
- Lefcovich, Mauricio. Estadística aplicada a los negocios. [Online] [Cited: 6 5, 2011.] <http://www.monografias.com/trabajos34/estadistica-negocios/estadistica-negocios.shtml>.

- Modelamiento multidimensional. [Online] [Cited: 05 30, 2011.]
<http://www.google.com/url?sa=t&source=web&cd=1&ved=0CBsQFjAA&url=http%3A%2F%2Fwww.inf.udec.cl%2F~revista%2Fediciones%2Fedicion4%2Fmodmulti.PDF&rct=j&q=modelo%20multidimensional%20&ei=PnDuTaTdB6L00gGztszfAw&usg=AFQjCNGUM1nUN0UrK8yOYBuxF13dXBnnnQ&sig2=iGxzY>.
- Montoya, Lic. Ricardo Sánchez. Business intelligence... to bi or not to bi. [Online] [Cited: 05 30, 2011.] <http://www.gestiopolis.com/recursos/documentos/fulldocs/ger1/tobi.htm>.
- Nielsen, Jakob. Usability Engineering. [Online] [Cited: 05 29, 2011.]
<http://www.useit.com/jakob/useengbook.html>.
- Ochoa, Darián González. Diseño e Implementación de un Almacén de Datos Operacionales para la corporación CIMEX. . Ciudad de La Habana. : Universidad de las Ciencias Informáticas.
- Ortiz, Marta Cecilia Ortiz. La inteligencia de negocios aplicada a las organizaciones en Latinoamérica. [Online] [Cited: 06 09, 2011.]
http://www.google.com/url?sa=t&source=web&cd=1&ved=0CBcQFjAA&url=http%3A%2F%2Fwww2.epm.com.co%2Fbibliotecaepm%2Fbiblioteca_virtual%2Fdocuments%2Fla_inteligencia_de_negocios_aplicada_a_las_organizaciones.pdf&rct=j&q=La%20inteligencia%20de%20negocios%20es%2.
- Peñaloza, Lucía Victoria Hernández. Tesis para lograr el título de Magíster: Diseño y Construcción de un Data Mart para la mantención de Indicadores de Sostenibilidad de la Industria del Salmón. Chile. Chile : s.n., 2008.
- Pentaho BI Plataform Server. [Online] [Cited: 12 01, 2010.]
<http://www.summan.com/index.php/productos/software/pentaho-.html>.
- PENTAHO: Creación de un cubo con Mondrian. . [Online] [Cited: 11 24, 2010.]
<http://www.gravitar.biz/index.php/bi/pentaho-ejemplo-cubo-mondrian/>.
- Qué es estadística? [Online] [Cited: 10 10, 2010.] <http://www.misrespuestas.com/que-es-estadistica.html>.
- Rodríguez, Héctor Julio. Historia de las bases de datos en Ciencia de la Información. [Online] [Cited: 06 02, 2011.]
http://recursostic.javeriana.edu.co/wiki/index.php/Historia_de_las_bases_de_datos_en_Ciencia_de_la_Informaci%C3%B3n.
- Ross, Ralph Kimball y Margy. The Data Warehouse Toolkit. 2002.
- Sitio de PgAdmin. [Online] [Cited: 11 25, 2010.] <http://www.pgadmin.org/>.

- Sitio de Postgres. [Online] [Cited: 11 23, 2010.] <http://www.postgresql.org/>.
- Tablas de dimensiones. [Online] [Cited: 06 03, 2011.] <http://msdn.microsoft.com/es-es/library/ms244714%28v=vs.80%29.aspx>.
- Tablas de hechos. [Online] [Cited: 06 03, 2011.] <http://msdn.microsoft.com/es-es/library/ms244679%28v=vs.80%29.aspx>.
- Tipos De Base De Datos. Naturaleza Y Representación. buenastareas.com. [Online] [Cited: 06 04, 2011.] <http://www.buenastareas.com/ensayos/Tipos-De-Base-De-Datos-Naturaleza-y/12004.html>.
- Vargas, Rafael Kauffmann. Estadística básica para supervisores aplicada al control de procesos. [Online] [Cited: 06 08, 2011.] <http://www.mailxmail.com/curso-estadistica-basica-supervisores-aplicada-control-procesos/estadistica-importancia-toma-decisiones>.
- Ventajas e inconvenientes de los sistemas de bases de datos. [Online] [Cited: 01 10, 2011.] <http://www3.uji.es/~mmarques/f47/apun/node7.html>.
- Visual Paradigm for UML (ME). [Online] [Cited: 11 25, 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/).
- Warehouse, Data. Data Warehouse. [Online] [Cited: 01 10, 2011.] <http://www.buenastareas.com/ensayos/Data-Warehouse/22490.html>.

ANEXOS

Anexo 1: Caso de prueba viviendas terminadas

Escenario	Descripción	Variables		Respuesta del sistema	Flujo central
		Perfiles de análisis	Indicadores a medir		
Estructura y dinámica de viviendas terminadas	Muestra la estructura y dinámica de las viviendas terminadas por años y provincias	Año	-Estructura Total Viviendas	Se muestra la tabla con los valores correspondientes a cada escenario y su gráfico.	<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 Construcción.</p> <p>Se selecciona el libro de trabajo L.T Viviendas terminadas</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>
		Provincia	-Estructura Viviendas Estatales -Estructura Viviendas no Estatales -Estructura Viviendas de la Población -Dinámica Total Viviendas -Dinámica Estatal -Dinámica No Estatal -Dinámica Población		
Viviendas terminadas	Muestra las viviendas terminadas por años provincia.	Año	-Total Viviendas		
		Provincia	-Estatal -No_Estatal -UBPC -CPA -CCS -Población		

Anexo 2: Caso de prueba obras terminadas

Escenario	Descripción	Variables		Respuesta del sistema	Flujo central
		Perfiles de análisis	Indicadores a medir		
Dinámica de obras	Muestra la dinámica de las obras terminadas por años y tipo de obra y nivel territorial.	Año	-Dinámica Cantidad de Obras	Se muestra la tabla con los valores correspondientes a cada escenario y su gráfico.	Se abre la aplicación. Se autentifica. 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 AA Construcción. Se selecciona el libro de trabajo L.T Obras Terminadas En la parte inferior izquierda se selecciona el reporte deseado. En el área de trabajo se visualiza la tabla correspondiente al reporte.
		Tipo de obra	-Dinámica Valor de Obras		
		Nivel Territorial			
Estructura de obras	Muestra la estructura de las obras terminadas por años y tipo de obra y nivel territorial.	Año	-Estructura de la cantidad de obras		
		Tipo de obra	-Estructura del valor de obras		
		Nivel Territorial			
Obras terminadas	Muestra las obras terminadas por años y tipo de obra y nivel territorial.	Año	-Cantidad		
		Tipo de obra	-Valor		
		Nivel Territorial			

Anexo 3: Caso de prueba presas en explotación

Escenario	Descripción	Variables		Respuesta del sistema	Flujo central
		Perfiles de análisis	Indicadores a medir		
Presas en explotación	Muestra la capacidad y cantidad de las presas en explotación en el país por años y provincias	Año	-Cantidad -Capacidad	Se muestra la tabla con los valores correspondientes a cada escenario y su gráfico.	<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 AA Construcción.</p> <p>Se selecciona el libro de trabajo L.T Presas en explotación</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>
		Provincia			

Anexo 4: Lista de chequeo

Estructura del documento					
Peso	Indicadores a evaluar	Evaluación	(NP)	Cantidad de elementos afectados	Comentarios
crítico	1. ¿Los entregables contienen las secciones obligatorias de la plantilla estándar definidas para un expediente de proyecto? (ver expediente de proyecto)				
Indicadores definidos en el desarrollo					
Peso	Indicadores a evaluar	Evaluación	(NP)	Cantidad de elementos afectados	Comentarios
	1. ¿Se utilizó un lenguaje cuyas sentencias son expresables mediante una sintaxis bien definida?				
Semántica del documento					
Peso	Indicadores a evaluar	Evaluación	(NP)	Cantidad de elementos afectados	Comentarios
crítico	1. ¿Se han identificado errores ortográficos en los entregables?				
crítico	2. ¿Se entiende claramente lo que se ha especificado en el documento?				
	3. ¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?				

Anexo 5: Reportes candidatos

Area de análisis (AA)	Construcción
Libro de Trabajo (LT)	LT1 – Obras terminadas
Reporte (Tabla de Salida – TS)	TS1 – Dinámica de obras.
Descripción	Reporte que muestra la dinámica de las obras terminadas en el país.
Elementos del reporte	Tipo de obra - dim_tipo_obra Año - dim_temporal_año
Frecuencia de emisión	
Funciones	Dinámica: (tipo de obra del año actual / tipo de obra del año anterior) * 100
Gráfico	Barras verticales 3D

Area de análisis (AA)	Construcción
Libro de Trabajo (LT)	LT1 – Obras terminadas
Reporte (Tabla de Salida – TS)	TS2 – Estructura de obras
Descripción	Reporte que muestra la estructura de las obras terminadas en el país.
Elementos del reporte	Tipo de obra - dim_tipo_obra Año - dim_temporal_año
Frecuencia de emisión	
Funciones	Estructura: (tipo de obra / total de tipos de obras) * 100
Gráfico	Barras verticales 3D

Area de análisis (AA)	Construcción
Libro de Trabajo (LT)	LT1 – Obras terminadas
Reporte (Tabla de Salida – TS)	TS3 – Obras terminadas
Descripción	Reporte que muestra la cantidad y el valor de obras terminadas en el país por años y por tipo de obra.
Elementos del reporte	Tipo de obra - dim_tipo_obra Año - dim_temporal_año
Frecuencia de emisión	
Funciones	
Gráfico	Barras verticales 3D

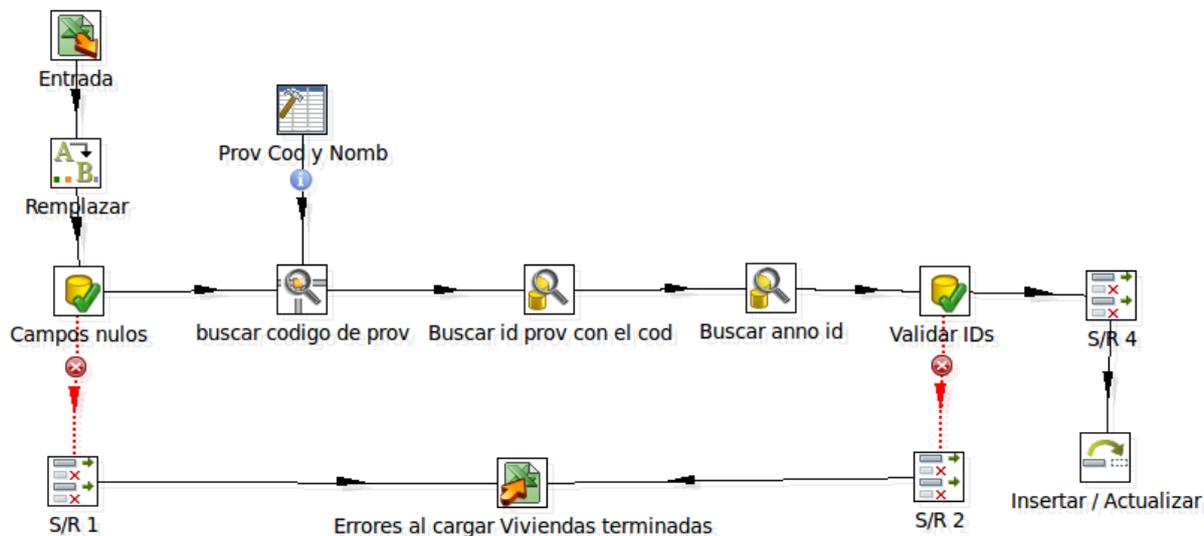
Area de análisis (AA)	Construcción
------------------------------	---------------------

Libro de Trabajo (LT)	LT2 – Presas en explotación
Reporte (Tabla de Salida – TS)	TS1 – Presas en explotación
Descripción	Reporte que muestra la cantidad y la capacidad de las presas en explotación en el país.
Elementos del reporte	Provincia - dim_provincia Año - dim_temporal_anno
Frecuencia de emisión	
Funciones	
Gráfico	Barras verticales 3D

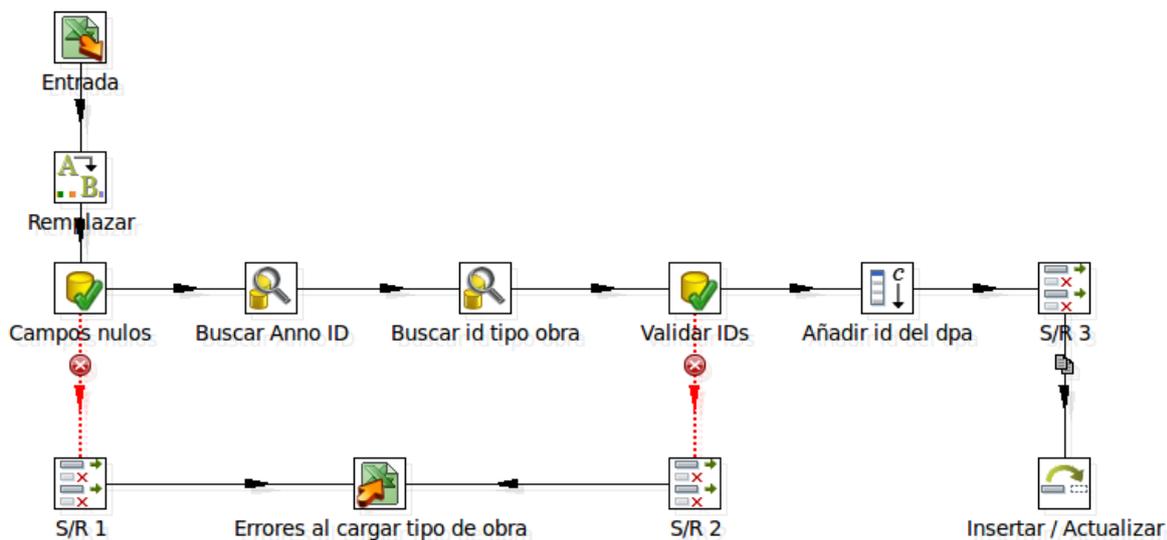
Area de análisis (AA)	Construcción
Libro de Trabajo (LT)	LT3 – Viviendas terminadas
Reporte (Tabla de Salida – TS)	TS1 – Estructura y dinámica de viviendas terminadas
Descripción	Reporte que muestra la estructura y dinámica de las viviendas terminadas en el país
Elementos del reporte	Año - dim_temporal_anno Viviendas- (total_viviendas, estatal, no_estatal, población)
Frecuencia de emisión	
Funciones	Estructura: viviendas/total de viviendas *100 Dinámica: (viviendas del año actual / viviendas del año anterior)*100
Gráfico	Barras verticales 3D

Area de análisis (AA)	Construcción
Libro de Trabajo (LT)	LT3 – Viviendas terminadas
Reporte (Tabla de Salida – TS)	TS2 – Viviendas terminadas
Descripción	Reporte que muestra la cantidad de viviendas terminadas en el país, ya sean estatales o no estatales.
Elementos del reporte	Provincia - dim_provincia Año - dim_temporal_anno
Frecuencia de emisión	
Funciones	
Gráfico	Barras verticales 3D

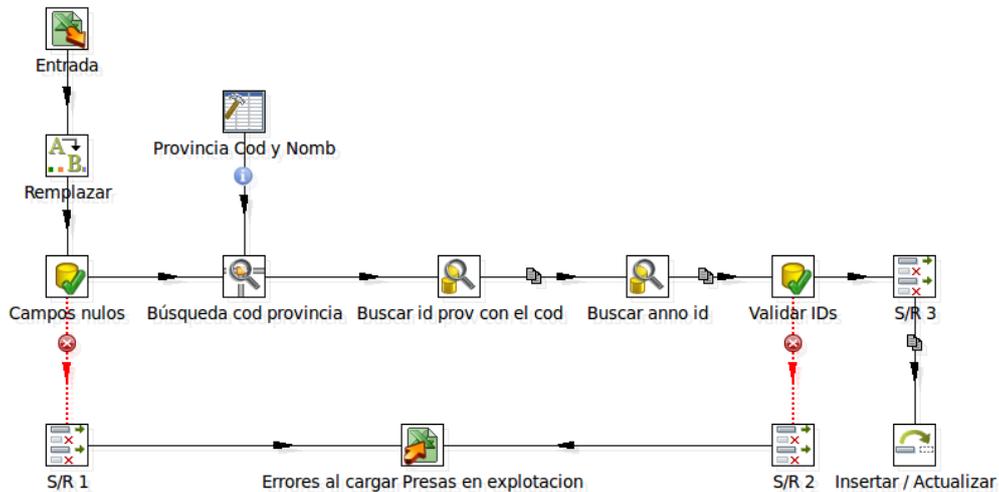
Viviendas terminadas



Obras terminadas



Presas en explotación



GLOSARIO

Centro informante: entidad de la Oficina Nacional de Estadísticas que le envía información.

HTTP(Hypertext Transfer Protocol): en español protocolo de transferencia de hipertexto, es el método más común de intercambio de información en la world wide web, el método mediante el cual se transfieren las páginas web a un ordenador.

Listener: es un componente que muestra los resultados de las muestras. Los resultados se pueden mostrar en un árbol, tablas, gráficos o simplemente escribir en un archivo de registro.