Universidad de las Ciencias Informáticas

Facultad 6

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

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

Autor:
Yuniel Ochoa Rodríguez

Tutores:
Ing. Yunier Santana Aldana
Ing. Roberto Tellez Ibarra

Ciudad de La Habana, junio 2011

"Lo fundamental es que seamos capaces de hacer cada día algo que perfeccione lo que hicimos el día anterior. "

Ernesto Guevara de la Serna

DECLARACIÓN DE AUTORÍA

Declaración de autoría

Declaro ser autor de la presente tesis y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

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

_____

Yuniel Ochoa Rodríguez

Firma del Autor

_____ _____

Firma del Tutor Firma del Tutor

Ing. Yunier Santana Aldana Ing. Roberto Tellez Ibarra

Tutores

Tutor: Ing. Yunier Santana Aldana
Especialidad de graduación: Ingeniería en Ciencias Informáticas
Categoría Científica: Ingeniero
Años de experiencia en el tema: 2
Años de graduado: 2
Correo electrónico: ysaldana@uci.cu

Tutor: Ing. Roberto Tellez Ibarra
Especialidad de graduación: Ingeniería en Ciencias Informáticas
Categoría Científica: Ingeniero
Años de experiencia en el tema: 2
Años de graduado: 2
Correo electrónico: rtibarra@uci.cu

A mis padres Idio y Orlidia, por ser mis ejemplos, mis amigos incondicionales, por estar en el momento en que más los necesito y brindarme su apoyo en difíciles ocasiones.

A mi hermano y mi sobrina querida, que son toda mi vida.

A mi novia, que me ha consolado cuando nadie lo ha podido hacer, que me ha dado las fuerzas para seguir adelante, que me ha levantado cuando me he caído y que me ha dado el amor que nunca había conocido.

A mis suegros y a mi cuñado por el apoyo que me han brindado durante estos años

A mis amistades de la UCI que han sido una familia para mi desde que entre a la universidad, hemos convivido por 5 años, si los menciono no cabrían en este documento, pero siempre hay algunos en especial, Ricardo Pico, Ariel Manresa, Yoendy, Leonel, en fin todos aquello que de alguna forma han compartido conmigo buenos y malos momentos. Gracias y éxitos para todos.

A mis tutores, Yunier y Roberto Tellez, que durante todo el tiempo que me han tutoreado, han sido capaz de ir moldeando mi forma de proyectarme y de ser mejor cada día, gracias.

A mi tribunal y mi oponente que con sus revisiones han logrado que esta investigación sea digna de un profesional por el que estoy luchado ser. En fin a todas aquellas personas que de alguna que de otra forma han contribuido con el éxito de esta investigación.

…A mis padres, en especial…

…A mi hermano…

… A mi novia…

Resumen

El presente trabajo de diploma enmarca su desarrollo en áreas de Mercados de Datos (MD), para el análisis, diseño e implementación de los indicadores relacionados con los temas de Cuentas nacionales. Para realizar este proceso de desarrollo se documentan las metodologías y las diferentes herramientas informáticas utilizadas, se realiza el análisis y el diseño, donde se hace el levantamiento de requisitos. De igual manera se definen e implementan los mecanismos de extracción, transformación y carga de los datos correspondientes a los excel propuestos, desarrollando además, la capa de Inteligencia del Negocio (BI) para la visualización y realización del análisis de la información contenida en el Almacén de Datos(AD) para el Control de las Cuentas nacionales. Incluyendo también, la aplicación de listas de chequeo y casos de pruebas, con el objetivo de obtener un MD que cumpla con todas las necesidades del cliente.

Palabras claves: Inteligencia del negocio, Cuentas nacionales, Almacén de Datos, Mercado de datos

Índice

INTRODUCCIÓN

La información es un elemento vital para cualquier empresa o institución, influye de manera directa en la forma en que estas operan. En la actualidad, manejar correctamente la información y realizar análisis sobre ésta, es una necesidad primordial para la sociedad. Los avances tecnológicos ocurridos durante las últimas décadas, han facilitado la manipulación y almacenamiento de modo eficiente de grandes volúmenes de datos.

Este avance tecnológico provocó una gran competencia en el mercado, lo que conllevó a que las empresas se esforzaran por elevar su nivel profesional en aras de ganar prestigio y demanda entre los clientes. Debido a esta alta competitividad, la información que se genera en las empresas crece en volumen, provocando que la toma de decisiones sea más difícil y la gestión de los datos de forma manual sea costosa en tiempo y esfuerzo; surgiendo la necesidad de informatizar los datos.

La estadística no está exenta de estos avances y transformaciones, el control estadístico permite recopilar, clasificar, resumir y analizar la información de las empresas e instituciones de gobierno. Cuba posee una larga historia en materia de estadística. La entidad rectora en el país es la Oficina Nacional de Estadísticas (ONE), que cuenta con el Sistema Estadístico Nacional (SEN) para organizar, dirigir, controlar y regular la estadística en el país. El SEN garantiza la captación de todas las informaciones estadísticas de la sociedad cubana a través de diferentes instrumentos de captación de datos como son los censos y encuestas económicas, sociales y demográficas que permiten, de manera sistemática, seguir el proceso de desarrollo a diferentes niveles. Para su mejor control la ONE tiene una estructura institucional distribuida territorialmente en las provincias y municipios del país. Existen 14 oficinas provinciales y 169 oficinas municipales de estadísticas. Las oficinas municipales, están subordinadas a las provinciales, las cuales son las encargadas de interactuar directamente con los Centros Informantes (CI), siendo estos, el último eslabón de la cadena de la actividad estadística.

Actualmente el Centro de Tecnologías de Gestión de Datos (DATEC) de la UCI trabaja en conjunto con la ONE para optimizar el uso que actualmente le dan a las tecnologías informáticas. Debido a que en estos momentos en la ONE, específicamente en el área de cuentas nacionales existen serias deficiencias al almacenar y gestionar los datos que se generan, pues la información es almacenada en formatos que dificultan el análisis, y sólo puede ser consultada por un especialista de la informática con alto conocimiento del negocio, debido a que algunas de las fuentes están codificadas de forma que si no se es conocedor de la terminología usada no se puede comprender el contenido. Además, se

generan reportes en tiempos prolongados de solicitud y los datos no se integran, lo que atenta contra la calidad de estos, pero principalmente no existe una aplicación informática que brinde reportes flexibles con información actualizada para apoyar el proceso de toma de decisiones.

Por la situación anteriormente descrita, se plantea como **problema de la investigación:**

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

La investigación tiene como **objeto de estudio,** Los Almacenes de Datos, enmarcado en el **campo de acción** Mercado de datos estadísticos para el área de Cuentas Nacionales del Sistema de Información de Gobierno.

El **objetivo general** de este trabajo es desarrollar el mercado de datos para el área de Cuentas nacionales del Sistema de Información de Gobierno.

En correspondencia con el objetivo general, se plantean como **objetivos específicos:**

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

Para el cumplimiento de los objetivos específicos se realizaron esencialmente las siguientes **tareas investigativas**:

- 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.

El trabajo de diploma está estructurado de la siguiente manera: introducción, cuatro capítulos, conclusiones, recomendaciones, referencias bibliográficas, bibliografía, anexos y glosario de términos.

Capítulo 1: Fundamentación teórica del mercado de datos Cuentas nacionales

En este capítulo se hace un análisis del estado del arte del objeto de estudio, la metodología a seguir en el diseño e implementación de un Mercado de Datos (MD), y la selección de herramientas útiles para llevar a cabo un almacenamiento de la información.

Capitulo 2: Análisis y diseño del mercado de datos Cuentas nacionales

Este capítulo contiene la descripción de los pasos a seguir durante el análisis y el diseño de la solución. Se identifican los casos de uso del sistema, requisitos funcionales y no funcionales, además de los requisitos de información y los multidimensionales. Se definen medidas, tablas de hechos y dimensiones para estructurar el modelo dimensional de la solución.

Capítulo 3: Implementación del mercado de datos Cuentas nacionales

Este capítulo contiene todos los elementos referentes a la implementación del sistema, las dimensiones y hechos que forman parte de la solución, la descripción de los roles, permisos creados y los pasos para la implantación del sistema.

Capítulo 4: Validación del mercado de datos Cuentas nacionales

En este capítulo se describe la manera en la que se realizó la validación de la solución propuesta, aplicando lista de chequeo y casos de pruebas, además de una carta de aceptación del cliente.

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA DEL MERCADO DE DATOS CUENTAS NACIONALES

1.1 Introducción

En este capítulo se abordan temas relacionados con las diferentes herramientas que existen en el mundo para la construcción de Mercados de Datos, sus características y aspectos más significativos, haciendo referencia al proceso de desarrollo de software, así como a las metodologías de desarrollo, tecnologías y tendencias actuales que pueden ser útiles en la propuesta de solución, argumentando el porqué de su uso**.**

1.2 Cuentas nacionales

1.2.1 ¿Qué son las Cuentas nacionales?

Las Cuentas nacionales son un registro contable de las transacciones realizadas por los distintos sectores de la economía en el cual se brinda una perspectiva global del sistema económico. Los esquemas contables sirven para organizar las nociones de la actividad económica con el fin de analizar y elaborar políticas y medir la actividad de un país en un período determinado. Por otro lado, el hecho de que diversos subtotales en las cuentas deban igualarse proporciona un mecanismo de control en cuanto a la consistencia recíproca que representan. Además, si es posible prever el comportamiento de algunas variables económicas claves, las identidades de las cuentas proveen una idea de cómo debe evolucionar la economía en su conjunto.

1.3 Indicadores económicos

Un indicador económico, como su nombre lo dice, sirve para indicar la situación de un aspecto económico particular en un momento determinado en el tiempo. Aspectos como los precios, el comercio exterior, las finanzas públicas, el sistema financiero y la producción son algunos de ellos. Para cada uno de estos aspectos existen diversos indicadores. Algunos de los más importantes son:

- **Los indicadores de proceso:** Se definen como el conjunto de datos obtenidos durante la ejecución del proceso que permiten conocer el comportamiento del mismo y, por tanto, predecir su comportamiento futuro en circunstancias similares [1].
- **Los indicadores de producto:** Son el conjunto de datos referidos al producto en sí (medidas obtenidas respecto a medidas previstas, por ejemplo) cuyo análisis indica hasta qué punto se ha conseguido el producto que se deseaba [1].
- **Los indicadores de servicio:** Igual que los indicadores de producto, son el conjunto de datos

referidos al servicio cuyo análisis indica el grado de cumplimiento de los niveles de servicio previamente establecidos [1].

1.3.1 Principales indicadores económicos que se miden sobre las cuentas en el mundo

El Producto Interno Bruto (PIB) puede calcularse a través de tres procedimientos fundamentales:

Método del gasto

En el método del gasto, el PIB se mide sumando todas las demandas finales de bienes y servicios en un período dado. En este caso se está cuantificando el destino de la producción. Existen cuatro grandes áreas de gasto: el consumo de las familias, la inversión en nuevo capital, el consumo del gobierno y los resultados netos del comercio exterior (exportaciones - importaciones).

Método de la distribución o del ingreso

Este método suma los ingresos de todos los factores que contribuyen al proceso productivo, como por ejemplo, sueldos y salarios, comisiones, alquileres, derechos de autor, honorarios, intereses y utilidades. El PIB es el resultado del cálculo por medio del pago a los factores de la producción.

Método de la oferta o del valor agregado

En términos generales, el valor agregado o valor añadido, es el valor de mercado del producto en cada etapa de su producción, menos el valor de mercado de los insumos utilizados para obtener dicho producto; es decir, que el PIB se cuantifica a través del aporte neto de cada sector de la economía.

Según el método del valor agregado, la suma de valor agregado en cada etapa de producción es igual al gasto en el bien final del proceso de producción.

1.3.2 Principales indicadores económicos que se miden sobre cuentas en Cuba

A continuación se dan a conocer algunos de los indicadores económicos que se miden sobre las cuentas en Cuba:

Producto interno bruto(PIB),consumo total o gasto total de consumo final, consumo final de los hogares, mercado estatal, mercado agropecuario, mercado de trabajadores por cuenta propia, formación bruta de capital, consumo final efectivo de los hogares entre otros.

1.4 Tecnologías de almacenamiento de datos

Desde el surgimiento de las tecnologías de almacenamiento de datos, estas se han ido convirtiendo en una herramienta fundamental para el control y manejo de operaciones, los documentos Excel, XML

entre muchos, que aunque hoy día aún se utilizan, ya sea como un simple modo de guardar información o unidas a otras tecnologías resolviendo problemas de almacenamiento y tratamiento de datos, como por ejemplo la utilización de los XML en multimedia y sitios web, que se usan como Bases de Datos, entre otros usos que se le dan a estas tecnologías de almacenamiento. Debido a que las empresas van acumulando grandes volúmenes de información a medida que pasa el tiempo, necesitando mantener la misma tal y como fue almacenada, para la realización de comparaciones entre informaciones históricas y análisis inteligentes sobre los datos, permitiendo a estas empresas ver el comportamiento de su desarrollo. Se hace imposible la utilización de estas estructuras como variantes de solución para la investigación, por la razón que sería difícil con el tiempo y con ese cúmulo de información realizar análisis sobre los datos. Debido a esto se le da paso a los grandes Almacenes de Datos (AD) por sus funcionalidades y características.

1.4.1 Almacén de Datos

En la actualidad las organizaciones, en sus bases de datos, almacenan datos tanto internos como externos de clientes, productos, servicios, estructura organizativa, canales de distribución, entre otros. Sin embargo, esta cantidad de datos no se suele corresponder con una mayor accesibilidad a la información de utilidad en la gestión comercial. Por lo que nace la necesidad de crear nuevas infraestructuras de comunicación con potentes y flexibles herramientas de tratamiento de información, que mejoren la calidad, cantidad y eficiencia de los datos comerciales, así como el análisis, procesamiento y comunicación de los mismos. Los almacenes AD, los cuales pueden aportar a las corporaciones la base tecnológica necesaria para afrontar los nuevos retos de la situación actual y las perspectivas de futuro de la gestión comercial, permitiendo además en primera instancia un almacenamiento adecuado de los datos obtenidos de las actividades habituales de la organización.

Hay que señalar que el diseño del AD no es un proceso trivial, se debe elegir, en base a la información que se desea explotar, los datos que se guardarán, la unidad mínima de éstos, la estructura de las entidades de información, las dimensiones que se estudiarán, estadísticos intermedios que se deben conservar y muchos aspectos más para que el diseño responda a las necesidades de información de distintos departamentos o áreas y niveles jerárquicos de la empresa, así como la eficiencia en la provisión operacional de dicha información [2].

1.4.1.1 Características de los Almacenes de Datos

- **Orientado por temas:** los datos en el almacén están organizados de manera que todos los elementos de datos relativos al mismo evento u objetivo queden unidos entre sí.

- **Variables en el tiempo:** los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones.
- **No volátil:** la información no se modifica ni se elimina, una vez almacenado un dato, éste se convierte en información de solo lectura, y se mantiene para futuras consultas.
- **Integrado:** la base de datos contiene la información de todos los sistemas operacionales de la organización, y dichos datos deben ser consistentes [3].

Hay muchas ventajas por las que se recomienda el uso de los AD. Algunas de ellas se mencionan a continuación:

- Integrar datos históricos sobre la actividad de la organización (o negocio) en un repositorio.
- Analizar los datos del negocio desde la perspectiva de su evolución en el tiempo.
- Prever tendencias de evolución del negocio.
- Identificar nuevas oportunidades de negocio y tomar decisiones estratégicas.
- Reducir los costes materiales y humanos en la toma de decisiones [4].

Sin embargo, a pesar de las ventajas que brindan presentan también algunas desventajas de las cuales es necesario tener conocimiento, ellas son:

- Pueden suponer altos gastos, además de los gastos de mantenimiento que son muy elevados.
- Pueden quedarse obsoletos relativamente pronto si los usuarios incrementan sus necesidades.
- Subestimación del tiempo requerido para extraer, limpiar y cargar los datos en el Almacén.
- Problemas con los sistemas de origen de los datos.
- La construcción de un Almacén de Datos puede requerir de mucho tiempo.
- La integración de las herramientas de Almacén de Datos, para conseguir un beneficio en la organización, es muy compleja [5].

1.4.2 Mercado de Datos

Los Mercados de Datos (MD), son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones, mientras que un DW entrega información a nivel corporativo [6]. El MD es un subconjunto de datos de un almacén relativo a los requisitos de un departamento o área de negocio concreto. Este subconjunto de datos puede funcionar de forma autónoma, o bien enlazado al AD. El motivo por el cual se crean MD es el crecimiento que tiene el almacén y así facilitar su construcción y utilización [5].

1.4.2.1 Características de los Mercado de Datos

- Los mercados de datos se enfocan a los requisitos de los usuarios que están asociados a un departamento específico de la empresa.
- Su utilización y comprensión es sencilla debido que contienen menor número de información que los almacenes de datos.

Beneficios de la utilización de MD:

- 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.
- Se realizan sobre ellos consultas MDX y/o SQL sencillas que facilitan el acceso a los datos que son utilizados con frecuencia.
- Facilidad para la historización de los datos.
- Validación directa de la información.

1.5 Metodologías

Una metodología es aquella guía que se sigue a fin de realizar las acciones propias de una investigación, es el conjunto de métodos que rigen una investigación científica o en una exposición doctrinal, indicando qué hacer y cómo actuar cuando se quiere obtener algún tipo de investigación. En el diseño de un AD, se ha destacado un conjunto de metodologías que definen y guían todo el ciclo de vida del desarrollo. Existen dos criterios bien identificados y que han marcado claramente su tendencia sirviéndole de guía a la comunidad mundial en cuanto a este tema.

Las dos vertientes fundamentales para el desarrollo de almacenes de datos son Ralph Kimball y Bill Inmon. Kimball (principal promotor del enfoque dimensional para el diseño de almacenes de datos), considera que un AD es una copia de los datos transaccionales específicamente estructurada para la consulta y el análisis. Bill Inmon (conocido por muchos como el padre del AD), plantea que un AD es un conjunto de datos orientados por temas, integrados, variantes en el tiempo y no volátiles, que tienen por objetivo dar soporte a la toma de decisiones.Para desarrollar proyectos de almacenes de datos se requiere de una metodología de desarrollo que ayude a su proceso de construcción. Existen en el mundo diferentes metodologías para el desarrollo de AD entre las que se el Ciclo de vida Kimball.

1.5.1 Ciclo de vida Kimball

El ciclo Kimball comienza con una planificación de proyecto, en la cual se define el alcance, se identifican y programan las tareas, se planifica el uso de los recursos, conformado con todo esto el plan de proyecto. En la segunda etapa de este ciclo se definen los requerimientos del negocio. Luego de definir los requerimientos del negocio, el proyecto se enfoca en tres líneas concurrentes: tecnología, datos y aplicaciones de la inteligencia de negocios.

La etapa de diseño del AD, está enmarcada en la línea de datos, donde se realiza el modelo dimensional y se analizan los datos del negocio para identificar la granularidad de las tablas de hechos, dimensiones y atributos asociados. Las construcciones primarias que se hacen en esta etapa son las tablas de dimensiones y de hechos. Las primeras contienen las métricas derivadas de los procesos de negocio o eventos. Hay que tener en cuenta que la granularidad debe ser lo más atómica posible, lo que permite una mayor flexibilidad y extensibilidad. Por su parte, las tablas de dimensiones contienen la descripción de atributos y características asociadas con medidas de eventos tangibles y específicos.

Para terminar el ciclo de vida se realiza el despliegue con el objetivo de dejar sentadas las bases de crecimiento y mantenimiento del AD [7].

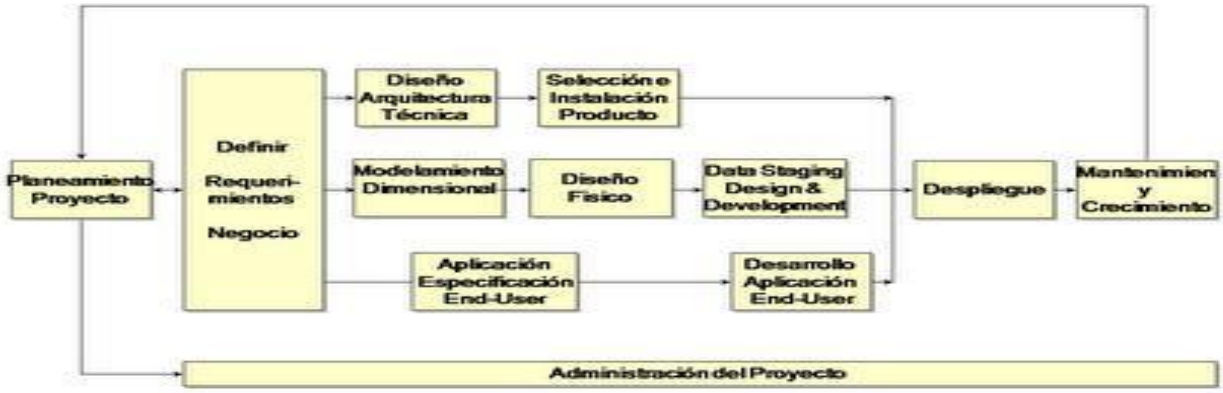


Figura 1: Ciclo de Vida Kimball

1.5.2 Metodología a usar para el desarrollo del almacén de datos

Para el desarrollo de la presente investigación se utiliza el modelo para el desarrollo de Soluciones de AD e Inteligencia de Negocio en DATEC (DW&BI), 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 de los productos.

La misma cubre todas las fases por las que pasa la construcción de un almacén de datos, desde el levantamiento de información inicial hasta la capa de visualización. Es una metodología mixta que reúne elementos de varias metodologías de desarrollo de proyectos de integración de datos, toma como base la metodología propuesta por Ralph Kimball. En una primera fase contempla el levantamiento de información a nivel de negocio para identificar los posibles indicadores y aspectos a medir en los análisis, que luego de algunas transformaciones se convierten en los requerimientos de información de entrada y de salida para la solución de integración.

Entre las principales características podemos encontrar que su 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 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.

DW&BI cuenta con 4 fases principales:

1. Requerimientos y gestión del proyecto.
2. Arquitectura.
3. Diseño e implementación.
4. Implantación y operaciones.

Durante su ciclo de vida 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. Luego se procede a la definición de los requerimientos.
- Arquitectura y diseño: se definen las estructuras de almacenamiento, se diseñan las reglas de ETL y se define la arquitectura de información que regirá el desarrollo de la solución.
- Implementación: se diseña físicamente el repositorio de datos, se crean las estructuras de almacenamiento, el área temporal de almacenamiento, se ejecutan las reglas de ETL y se

configuran e implementan las herramientas de BI para la obtención de los elementos 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, 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, 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.

Como se puede observar a continuación, en cada flujo intervienen grupos específicos, cada uno con actividades y responsabilidades concretas. La línea tiene una estructura conformada por cinco grupos fundamentales, cada uno de estos se especializa en un conjunto de actividades generales que contienen actividades específicas que tributan al desarrollo del proyecto, generan sus propios artefactos y se dividen por roles para darle cumplimiento a todas las actividades.

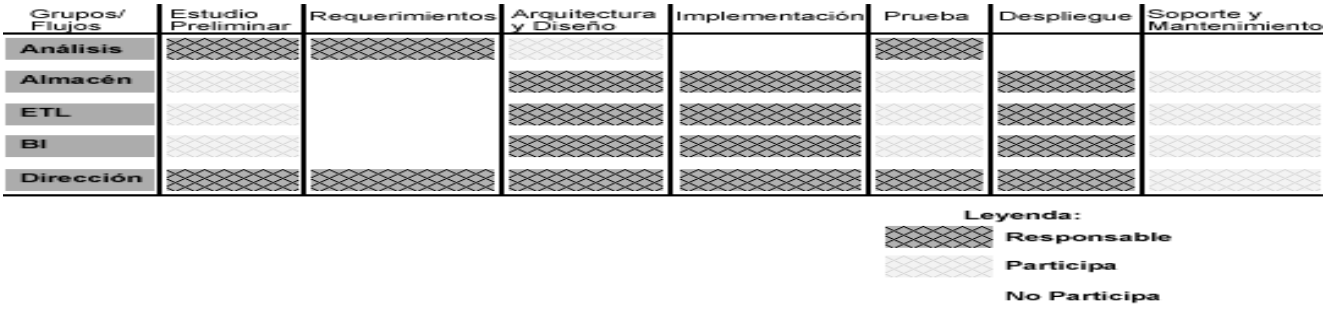


Figura 2: Relación de los grupos con los flujos.

1.6 Modelo de Datos

Un modelo de datos es “un conjunto de conceptos, reglas y convenciones que nos permiten describir y en ocasiones manipular los datos de un cierto mundo real que deseamos almacenar en la base de datos”. Al producto del modelo de datos se le llama esquema (descripción de la estructura de la base de datos) y a los datos en concreto almacenados en la base de datos en ese momento [8]. Existen varios modelos de BD entre los cuales se encuentra:

1.6.1 Modelo Entidad-Relación

Un diagrama o modelo entidad-relación es un lenguaje para el modelado de datos de un sistema de información. Estos modelos expresan entidades más relevantes para el sistema, sus inter-relaciones y propiedades. Trabajan dividiendo los datos en muchas entidades discretas donde cada una se convierte en una tabla física en la base de datos operacional. Los sistemas de información que se realizan bajo estas directrices comúnmente se denominan sistemas proceso de transacción en línea (OLTP). Su principal función es reflejar el estado y funcionamiento de las empresas mediante el registro de las operaciones que realizan diariamente. Los modelos entidad–relación no son recomendables para el diseño de los almacenes de datos debido a que no garantizan la recuperación óptima del gran cúmulo de información que se almacena. Además, estos diagramas tienden a resultar en un diseño normalizado mientras que en un almacén de datos este aspecto no es un requisito a tener muy en cuenta [9].

1.6.2 Modelo Multidimensional

La tecnología de los AD debido a su orientación analítica, impone un procesamiento y pensamiento distinto, la cual se sustenta por un modelamiento propio de bases de datos, conocido como Modelamiento Multidimensional, el cual busca ofrecer al usuario su visión respecto a la operación del negocio. Modelamiento Dimensional es una técnica para modelar bases de datos simples y entendibles al usuario final. La idea fundamental es que el usuario visualice fácilmente la relación que existe entre los distintos componentes del modelo [10].

A diferencia de los sistemas de bases de datos clásicos, la estructura de las tablas y sus relaciones son representadas mediante un modelo multidimensional, o sea, almacenan la misma información que el Diagrama Entidad Relación pero la organizan 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, por tanto, se ajusta bastante bien a las aspiraciones que se tienen, al disminuir la normalización, resultar sencillo e intuitivo a los usuarios y bastante adaptable.

Características básicas del modelo multidimensional

La estructura básica de los Almacenes de Datos está definida por dos elementos: esquemas y tablas.

Tablas: Como cualquier base de datos relacional, un DWH se compone de tablas. Hay dos tipos básicos de tablas en el modelo multidimensional:

- **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 dimensiones [11]. Contienen los valores de las medidas de negocios, por ejemplo: venta promedio en dólares, número de unidades vendidas.

- **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. Contienen el detalle de los valores que se encuentran asociados a la tabla hecho. 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 (AD) debido a que son las llaves que hacen el AD usable y entendible. Estos atributos son las llaves de entrada a los hechos o medidas almacenadas. 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 [12].

Esquemas AD: la colección de tablas en el almacén se conoce como Esquema. Los esquemas caen dentro de dos categorías básicas: esquemas estrellas y esquemas copo de nieve [10].

Esquema estrella

El modelo multidimensional es conocido también como esquema estrella, ya que su estructura es similar: una tabla central y un conjunto de tablas que la atienden radialmente. El nombre proviene debido a que el diagrama forma una estrella, con puntos radiales desde el centro. El centro de la estrella consiste de una o más tablas hechos y las puntas son las tablas dimensionales. Las tablas dimensionales sólo tienen conexión con la tabla hecho y ninguna más [10].

Esquemas copo de nieve

La diferencia del esquema copo de nieve comparado con el esquema estrella, está dada en la estructura de las tablas dimensionales: las tablas dimensionales en el esquema copo de nieve están normalizadas. Cada tabla dimensional contiene sólo el nivel que es clave primaria en la tabla y la llave foránea de su parentesco del nivel más cercano del diagrama [10].

Esquema constelación: está compuesto por una serie de esquemas de estrella, es decir, una tabla de hechos central con otras auxiliares y sus respectivas tablas de dimensiones.

1.7 Etapas de desarrollo de un Almacén de Datos

1.7.1 Análisis y Diseño

La primera etapa de desarrollo de un AD es el análisis, dicha etapa es una de las más importantes debido a que se realiza un estudio del negocio especificando aspectos importantes para el entendimiento de la necesidad de la organización y lograr un software que cumpla con las expectativas y los resultados esperados. El análisis tiene como función dar soporte a las actividades del negocio, es donde se determinan los requisitos iniciales definiendo el alcance del almacén mediante entrevistas con los usuarios finales, revisan informes existentes, definen las reglas del negocio que se aplicarán en la construcción del almacén, identifican las fuentes de datos operacionales y externas. Mientras, que en el diseño se construye el esquema conceptual, realizan la definición de los procesos de Extracción, Transformación y Carga (ETL, por sus siglas en inglés), además, de definir los mercados de datos e informes [13].

1.7.2 Extracción, Transformación y Carga

Es la tecnología enfocada a la integración de datos, tanto por lote como a tiempo real hacia almacenes de datos [12]. Estos procesos se combinan para extraer datos de bases de datos fuentes, archivos u

otros sistemas, y colocarlas en bases de datos destino. Los procesos ETL se utilizan para migrar datos de una o más bases de datos a terceros y también para convertir bases de datos de un tipo o formato a otro. Se utilizan además para sincronizar datos desde diversas aplicaciones.

Extracción

Consiste en adquirir los datos desde los sistemas de origen, estas fuentes pueden estar sobre sistemas incompatibles. En este subproceso se convierten los datos a un formato preparado para iniciar el proceso de transformación. Aquí se verifican los datos extraídos, donde se comprueba si los datos cumplen lo que se espera y se adaptan al formato estándar diseñado, de lo contrario los datos son rechazados. En el proceso de extracción es necesario causar un mínimo impacto en el sistema origen, pues si se necesita extraer muchos datos el sistema origen podría ralentizar o colapsar, provocando que no pueda implementarse con normalidad para su uso cotidiano.

Transformación

En esta fase se aplican una serie de Reglas de Negocios sobre los datos extraídos, con el objeto de convertirlos en datos aptos para ser cargados. Aquí se necesita obtener una buena calidad de los datos y para ello se hace necesario el control de los valores válidos, garantizar la coherencia entre los valores, la eliminación de duplicaciones y comprobar que las reglas del negocio no han sido forzadas [12]. En esta fase los datos deben ser limpiados, pues estos pueden estar sucios e incompletos. Por ello se realiza un proceso de limpieza que elimina errores e inconsistencias en los datos y resuelve el problema de identidad de los objetos. Luego que los datos han sido limpiados se procede a realizar las transformaciones mediante las reglas de transformación que pueden ser: combinar los datos de distintas fuentes, realizar búsqueda de valores en distintas tablas, darle tratamiento a valores nulos, entre otras.

Carga

La fase de carga es el momento cuando los datos, provenientes de la fase anterior, son incluidos en el sistema de destino, dependiendo de los requerimientos de la organización. El principal objetivo de esta fase es lograr que los datos estén listos para ser consultados. Este subproceso puede abarcar una amplia variedad de acciones diferentes. En algunas bases de datos se sobrescribe la información antigua con nuevos datos. En los almacenes de datos al mantener un historial de los registros se puede hacer una auditoria de los mismos y disponer de un rastro de toda la historia de un valor a lo largo del tiempo, independientemente de la acción a tomar para la carga, al realizar esta operación se aplicarán todas las restricciones que se hayan definido, los cuales contribuyen a que se garantice la calidad de los datos en el proceso ETL.

Si no se realiza un correcto proceso de ETL se pudieran obtener datos incorrectos lo que afectaría el proceso de toma de decisiones, es por eso que este proceso constituye aproximadamente un 70% del trabajo de la construcción de un AD. Los procesos ETL son los componentes más importantes de una infraestructura de inteligencia de negocio. Pueden ser invisibles por los usuarios, recuperan los datos de todos los sistemas operativos y los pre-elaboran para las herramientas de análisis y de reporte.

Algunas de las características que posee son:

1. Es un mecanismo de carga muy eficiente y efectivo orientado a los almacenes de datos.
2. Enfocado a migrar y mezclar datos.
3. Reduce la exposición a desarrollos manuales (codificación) producto de la existencia en el mercado de herramientas potenciales para la implementación visual, con manejo de excepciones, gestión y planificación de tareas.
4. Necesita pocos servicios de administración y mantenimiento.
5. Gran capacidad para llevar a cabo transformaciones.
6. Tecnología enfocada a la Integración de datos en bases de datos versátiles hacia los Almacenes de Datos.

1.7.3 Inteligencia de negocio

Se define por Inteligencia de Negocio o Business Intelligence (BI) a la transformación de los datos de la compañía en conocimiento para obtener una ventaja competitiva. Desde un punto de vista más pragmático, y asociándolo directamente a las tecnologías de la información, podemos definir Business Intelligence como el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada (interna y externa a la compañía) en información estructurada, para su explotación directa (reporting, análisis OLAP...) o para su análisis y conversión en soporte a la toma de decisiones sobre el negocio [14].

El análisis de datos es un proceso en el que, a través de las distintas técnicas del análisis, como: el procesamiento analítico en línea (*OnLine Analytical Processing*, OLAP por sus siglas en inglés).En general, estos sistemas OLAP deben:

- Soportar requerimientos complejos de análisis.
- Analizar datos desde diferentes perspectivas.
- Soportar análisis complejos contra un volumen ingente de datos.

Se decidió utilizar la tecnología de Procesamiento Analítico en Línea (OLAP) ya que permite que los datos sean clasificados en diferentes dimensiones y pueden ser vistos unos con otros en diferentes combinaciones para obtener diferentes análisis de los datos que contienen. En estos modelos los datos son vistos como cubos los cuales consisten en categorías descriptivas (dimensiones) y valores cuantitativos (medidas). Estos modelos les permiten formular consultas complejas, arreglar datos de un reporte, cambiar los datos resumidos o detallados etc. Las principales características de OLAP son:

- **Rápido:** la primera regla se refiere a que el sistema debe ser capaz de responder de una forma rápida y ágil a la información que le sea solicitada por el usuario, el cual no deberá esperar más de cinco segundos a la hora de resolver peticiones sencillas y no más de veinte segundos en las peticiones complejas [15].
- **Análisis:** significa que el sistema debe poder reflejar cualquier lógica del negocio para poder responder a las preguntas específicas y necesidades empresariales [15].
- **Compartido**: el sistema deberá proporcionar herramientas que garanticen la confidencialidad de los datos y la seguridad de acceso por perfiles de los usuarios [15].
- **Multidimensional*:*** la herramienta deberá proporcionar soporte a cada una de las múltiples jerarquías que puedan existir dentro de la organización de información. Son todos los datos e información derivada de este proceso de análisis, la cual permitirá la toma de decisiones [15].
- **Información:** son todos los datos e información derivada de este proceso de análisis, la cual permitirá la toma de decisiones [15].

Existen tres modelos de información, el Procesamiento Analítico Relacional en Línea (ROLAP, por sus siglas en inglés), el Procesamiento Analítico Multidimensional en Línea (MOLAP, por sus siglas en inglés) y el Procesamiento Analítico el Línea Híbrido (HOLAP, por sus siglas en inglés). En los cuales el proceso de análisis se realiza igual, lo que varía es la metodología de almacenamiento, influyendo en la velocidad de recuperación de la información, las zonas de ubicación y el procesamiento de los datos en general. Específicamente se decidió usar la arquitectura ROLAP ya que almacena los datos en una base de datos relacional, lo que implica que no es necesario que los datos se repliquen en un almacenamiento separado para el análisis. Los cálculos se realizan en una base de datos relacional, con grandes volúmenes de datos y tiempos de navegación no predecibles.

El sistema ROLAP utiliza una arquitectura de tres niveles:

1. El nivel de base de datos utiliza bases de datos relacionales para el manejo, acceso y obtención del dato.

2. El nivel de aplicación es el motor que ejecuta las consultas multidimensionales de los usuarios.

3. El motor ROLAP se integra con el nivel de presentación, a través de este nivel los usuarios realizan los análisis OLAP.

Figura 3: Arquitectura del sistema ROLAP

La arquitectura ROLAP es capaz de usar datos precalculados 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 AD, y soporta técnicas de optimización de accesos para acelerar las consultas [16].

1.8 Herramienta de modelado

1.8.1 Visual Paradigm versión 6.4

En la presente investigación se utiliza para el diseño el Visual Paradigm (VP) en su versión 6.4 (versión 3.4 de la suite del producto). VP es una herramienta UML (Lenguaje Unificado de Modelado) profesional que soporta el ciclo de vida completo del desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue. Los sistemas de modelado UML ayudan a una rápida construcción de aplicaciones de calidad y a un menor coste. Permite dibujar todos los tipos de diagramas de clases, ingeniería inversa, generar código a partir de diagramas, generación de objetos a partir de bases de datos y generación de bases de datos a partir de diagramas de entidad relación.

A continuación se brindan algunas características de esta herramienta:

- Es una potente herramienta CASE, para visualizar y diseñar elementos de software, para ello utiliza el Lenguaje Unificado de Modelado (UML).
- Proporciona a los desarrolladores una plataforma que les permite diseñar un producto rápidamente y con la calidad requerida.
- Facilita la interoperabilidad con otras herramientas CASE y se integra con múltiples herramientas de desarrollo, como Eclipse/IBMWebSphere, Jbuilder, NetBeans IDE, Oracle Jdeveloper.
- Genera código y realiza ingeniería inversa para diez lenguajes de programación, Java, C++, CORBA IDL, PHP, XML Schema y ADA.
- Genera código para C#, Visual Basic.net, Object Definition Lenguage(ODL), Flasch Action Script, Delphi,Perl y Phyton.
- Se integra con el Visio para importar imágenes del mismo para realizar los diagramas de despliegue.
- Genera documentación para el proyecto en HTML, MS Word y PDF.
- Además, exporta e importa los diagramas en el estándar XML y como imágenes (ya sea con extensiones jpg o png).
- Es gratis en su edición Community.

1.9 Sistema Gestor de Base de Datos

1.9.1 PostgreSQL versión 8.4

Se decidió utilizar PostgreSQL versión 8.4 ya que es un potente sistema de base de datos relacional libre (Open Source, su código fuente está disponible) liberado bajo licencia The PostgreSQL Licence (TPL) similar a la Berkeley software Distribución (BSD). 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 (savepoints), copia de seguridad en línea entre otras características.

Algunas de las mejoras que brinda son: [17]

1. Restauración de la base de datos usando procesos paralelos, acelerando la recuperación de un respaldo hasta en 8 veces respecto a la versión anterior.

2. Mejora el rendimiento de las aplicaciones.

3. Privilegios por columna, para poder controlar el acceso a un nivel de detalle mayor. **4.** Configuración de idioma y ordenamiento por base de datos. Para que se pueda seleccionar la configuración más adecuada dependiendo del idioma que se requiera. **5.** Actualización en el lugar de 8.3 a 8.4 con un mínimo de downtime. **6.** Nuevas herramientas para monitorear las consultas, entregando mayor información a los administradores para saber lo que está sucediendo en la base de datos.

1.10 Herramientas para el proceso de Extracción, Transformación y Carga

Pentaho Data Integration versión 4.0.1

Es una herramienta de código abierto adoptado por Pentaho BI. Proporciona la extracción de gran alcance, Transformación y Carga (ETL) utilizando un enfoque innovador, orientado a los metadatos. Presenta un enfoque orientado a los metadatos que simplemente es indicar qué quiere hacer, pero no cómo desea hacerlo. Ahora los desarrolladores ETL, BI y los administradores pueden crear complejas transformaciones y el empleo en un entorno gráfico, arrastrar y soltar sin tener que generar ningún código personalizado. Se utiliza para poblar el AD, importación de datos en bases de datos, que van desde archivos de texto a hojas de Excel, migración de datos entre aplicaciones de base de datos, exploración de datos en bases de datos existentes. (tablas, vistas, sinónimos), enriquecimiento de información por la búsqueda de datos en varios almacenes de información (bases de datos, archivos de texto, hojas de Excel), limpieza de datos mediante la aplicación de condiciones complejas en transformaciones de datos, integración de aplicaciones.

1.11 Herramienta para Inteligencia de Negocio

Pentaho Schema Workbench versión 3.2

Es un entorno visual para el desarrollo y prueba de cubos OLAP Mondrian. Si bien la definición del XML para esquemas Mondrian no es extremadamente compleja, en la práctica resulta engorroso recordar cada uno de los elementos junto a sus atributos y sub-elementos. Con esta aplicación, se puede configurar una conexión JDBC como el modelo físico, para luego elaborar el esquema lógico de manera simple y efectiva.

Funcionalidades que brinda:

- Editor de esquemas integrados con un origen de datos subyacente para su validación.
- Prueba de consultas MDX contra el esquema y la base de datos.
- Examinar la estructura subyacente de bases de datos [18].

Mondrian OLAP Server.

Mondrian es una de las aplicaciones más importantes de la plataforma Pentaho BI. Es un servidor OLAP open source, que gestiona comunicación entre una aplicación OLAP (escrita en Java) y la base de datos con los datos fuente. Mondrian utiliza MDX como lenguaje de consulta, que fue un lenguaje propuesto por Microsoft. Funciona sobre las bases de datos estándar del mercado: Oracle, DB2, SQL-Server, MySQL, etc., lo cual habilita y facilita el desarrollo de negocios basados en la plataforma Pentaho. Para obtener la funcionalidad de procesamiento analítico en línea (OLAP) se utilizan otras dos aplicaciones: el servidor OLAP Mondrian, que combinado con Jpivot, permiten realizar consultas al AD y permite que los resultados sean presentados mediante un navegador, de modo que el usuario pueda realizar las actividades típicas de navegación. Mondrian actúa como "JDBC para OLAP".

Funcionamiento de Mondrian OLAP Server:

1. El cliente manda una solicitud de consulta bajo la interfaz web Jpivot.
2. Mondrian recibe la solicitud y bajo el esquema de metadatos que definen sus conceptos multidimensionales, busca si ya tiene los datos en cache respondiendo rápidamente a la petición.
3. Si los datos no se encontraron en cache ejecuta las sentencias SQL para generar los datos.
4. Se almacenan los datos en cache para agilizar posteriores consultas.
5. Y finalmente se devuelve el resultado al usuario cliente a través de la interfaz.

Pentaho BI Server versión 3.6

La aplicación Pentaho BI Server en su versión 3.6 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. La plataforma Pentaho BI Server, provee el soporte y la infraestructura necesaria para crear soluciones de inteligencia empresarial a problemas de negocios. El marco proporciona los servicios básicos, incluidos autenticación, registro, auditoría, servicios web y motor de reglas. La plataforma también incluye un motor de soluciones que integra reportes, análisis, tableros de comandos y componentes de minería de datos. Permite integrar procesos de negocio, administrar y programar reportes y además permite administrar la seguridad de los usuarios.

El servidor Apache Tomcat en su versión 5.5

Tomcat es el servidor Web más utilizado a la hora de trabajar con Java en entornos web. Es una implementación completamente funcional de los estándares de JSP y Servlets. Tomcat también puede especificarse como el manejador de las peticiones JSP y Servlets recibidas por servidores Web

populares, como el servidor Apache HTTP de la Fundación de software de Apache o el servidor Microsoft Internet Information Server (IIS).Está integrado en la implementación de referencia Java 2 Enterprise Edition (J2EE) de Sun Microsystems[19].

El servidor Tomcat es gratis, es fácil de instalar, se ejecuta en máquinas más pequeñas y es compatible con las API más recientes de Java. Ocupa muy poco espacio, es muy fiable, pone a disposición de todo el mundo las últimas actualizaciones de Java.

1.12 Conclusiones

El estudio realizado en este capítulo permitió al autor de este trabajo obtener una serie de conocimientos sobre los diferentes tipos de Integración de Datos, centrándose en el proceso ETL y sus subprocesos correspondientes, así como las herramientas necesarias para el desarrollo de este trabajo, la metodología que se debe utilizar, las características y conceptos principales que fueron necesarias indagar para un mejor entendimiento del autor.

Después de este profundo estudio se decide que la metodología apropiada para satisfacer la problemática, es el Modelo para el Desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio en DATEC, pues cumple con los requerimientos necesarios para la implementación del Proceso ETL, el cual se diseñará mediante las herramientas: Pentaho Data Integration en su versión 4.0.1, Visual Paradigm su versión 6.4, PostgreSQL en su versión 8.4 y PgAdmin III en su versión 1.10.

CAPÍTULO 2: ANÁLISIS Y DISEÑO DEL MERCADO DE DATOS CUENTAS NACIONALES

En el presente capítulo se muestran los resultados y pasos a seguir para el análisis. Se determinan y justifican cada uno de los requisitos funcionales y no funcionales con los que debe contar el sistema: los primeros hacen referencia a las tareas o actividades que el sistema será capaz de hacer; y los segundos, a las cualidades que hacen al producto atractivo, usable, rápido y confiable. Además, se realiza una revisión de los casos de uso, los hechos, las medidas y las dimensiones del mercado de datos identificados en la tesis precedente, para realizar el refinamiento de los mismos.

2.1 Análisis del mercado de datos Cuentas nacionales

2.1.1 Descripción del negocio

La ONE es el organismo rector de la estadística en Cuba. Comprende la elaboración de estadísticas y análisis del Estado y del Gobierno a los efectos de conocer el comportamiento de los procesos económicos, demográficos, sociales y, especialmente para el control del plan de economía nacional, del presupuesto, los compromisos estadísticos internacionales, la población y otras instituciones. Por ser la institución rectora de las estadísticas en el país, tiene como principal objetivo el de actuar como un almacén donde se gestione y guarde toda la información estadística del país.

Ante la necesidad de tener centralizada toda la información existente en la ONE, surge el proyecto Sistema de Información del Gobierno con el objetivo de lograr un mejor monitoreo y control de los datos. Su objetivo fundamental es la creación de una herramienta que permita acceder a toda la información, que apoye la toma de decisiones en las diferentes áreas socioeconómicas. Una de estas áreas es la Cuentas Nacionales, en la que se gestionan todas las estadísticas referidas a las actividades de las empresas estatales, entidades presupuestadas, unidades individuales privadas, sistemas bancarios, es decir, del proceso económico en su conjunto. Cuentas nacionales recibe la información del Sistema de Información Estadística Nacional (SIE-N), del Sistema Nacional de Contabilidad (SCN), de las Unidades Básicas de Producción Cooperativa (UBPC), de las Cooperativas de Producción Agropecuaria (CPA), y de otras fuentes estadísticas.

2.1.2 Tema de análisis

Un tema de análisis no es más que la división o clasificación de la información de una organización de acuerdo las diferentes temáticas que contenga, este es elegido según los objetivos que persiguen las áreas de la organización. Con este se obtienen varias perspectivas que orientan el avance del cumplimiento de las tareas planteadas y garantizan la utilidad y el éxito del diseño de las estructuras que se desarrollan. La ONE maneja toda la información relacionada con Cuentas nacionales, es por

ello que se identifica como tema de análisis las cuentas nacionales.

2.1.3 Reglas del negocio

Las reglas del negocio describen las políticas, normas, operaciones, definiciones y restricciones presentes en una organización, son las transformaciones que se le deben realizar a ciertos datos para obtener otros datos o los segmentos de un dato que pueden generar otros datos sin necesidad de ser introducidos por una persona. A continuación se definen las reglas de negocios que son de gran importancia para el logro de los objetivos en la misma:

PIB = Bienes + Servicios básicos + otros servicios + derechos de importación.

Bienes = Agricultura, caza, silvicultura y pesca + Explotación de minas y canteras + Industrias manufactureras + Construcción.

Servicios Básicos =Electricidad, gas, agua + Transporte, almacenamiento, comunicaciones.

Otros servicios = Comercio, restaurante, hoteles + Establecimientos financieros, bienes, inmuebles, servicios a empresas + Servicios comunales, sociales, personales.

Oferta global = PIB + Importaciones de bienes y servicios.

Demanda global = Demanda interna+ Exportaciones de bienes y servicios.

Demanda interna = Formación bruta de capital + Consumo total.

Consumo total = Gobierno general + Hogares.

Consumo de gobierno = Servicios públicos generales, económicos y otros +Educación +Sanidad +Asistencia social +Viviendas y ordenamiento urbano y rural +Cultura, Deporte y Recreación.

Consumo final de los hogares = Mercado estatal +Mercado agropecuario + Mercado de trabajadores por cuenta propia +Otras fuentes.

Formación bruta de capital =Formación bruta de capital fijo + Variación de existencias.

Formación bruta de capital fijo = Construcción + Maquinarias y equipos + Otras inversiones + Reparaciones capitalizables.

Consumo total = Consumo final efectivo del gobierno + Consumo final efectivo de los hogares.

Estructura porcentual = año actual /PIB * 100.

- Oferta global = Oferta global / PIB * 100.
- Demanda global = Demanda global / PIB * 100.

- Consumo total = consumo total / PIB * 100.

Tasas de crecimiento = año actual /año anterior *100 – 100.

Dinámica = Año actual / Año anterior * 100.

2.1.4 Necesidades de los usuarios

En la ONE uno de los principales problemas presentes es la no automatización de varias de sus oficinas, lo que imposibilita tomar decisiones inmediatas y los principales cruces de variables. La propuesta de solución está enmarcada en resolver varios problemas que tiene actualmente la ONE por parte de sus trabajadores, estos serán los futuros usuarios de la herramienta a desarrollar. Por lo que se acordó la realización de un MD para las cuentas nacionales de esta institución. Las necesidades de los usuarios están enmarcadas en el siguiente tema.

- Indicadores sobre Cuentas nacionales.

2.1.5 Requisitos de Información

Los requisitos de información son las principales funcionalidades que el sistema debe tener disponible a la hora de realizar análisis sobre los datos. Constituyen una entrada fundamental para el proceso de inteligencia del negocio y para futuros reportes. Describen qué información debe almacenar el sistema para satisfacer las necesidades de clientes y usuarios. Identifican los conceptos relevantes sobre los que se debe almacenar información y los datos específicos que son de interés.

RI 1. Obtener serie del Producto Interno Bruto por tipo de precio en un tiempo dado.

RI 2. Obtener oferta y demanda global por tipo de precio en un tiempo dado.

RI 3. Obtener oferta y demanda global por tipo de precio en estructura porcentual en un tiempo dado.

RI 4. Obtener oferta y demanda global por tipo de precio en tasa de crecimiento en un tiempo dado.

RI 5. Obtener índices de precios (año anterior=100) en un tiempo dado.

RI 6. Obtener índices de precios por deflactor implícito en un tiempo dado.

RI 7. Obtener PIB por clase de actividad económica a precios de mercado de NAE de Cuba por tipo de actividad por tipo de precio en un tiempo dado.

RI 8. Obtener PIB por clase de actividad económica a precios de mercado de NAE de Cuba por tipo de actividad en estructura porcentual a precios corrientes en un tiempo dado.

RI 9. Obtener indicadores de tasas del Producto interno bruto por clase de actividad económica a precios de mercado

RI 10. Obtener índices de precios por clase de actividad económica en porciento (año anterior =100) un tiempo dado.

RI 11. Obtener deflactor implícito por clase de actividad económica en porciento (base=1997) en un tiempo dado.

RI 12. Selección de indicadores del Sistema de Cuentas Nacionales (SCN) por tipo de precio en un tiempo dado.

RI 13. Selección de indicadores del Sistema de Cuentas Nacionales (SCN) en dinámica en precio de 1997 en un tiempo dado.

RI 14. Obtener consumo final del gobierno por finalidades por tipo de precio en un tiempo dado.

RI 15. Obtener consumo final del gobierno por finalidades en tasa de crecimiento a precios constantes en un tiempo dado.

RI 16. Obtener consumo final de los hogares por fuentes de oferta por tipo de precio en un tiempo dado.

RI 17. Obtener consumo final de los hogares por fuentes de oferta en tasa de crecimientos a precios constantes de 1997 en un tiempo dado.

RI 18. Obtener gasto total de consumo final por tipo de precio en tiempo dado.

RI 19. Obtener gasto total de consumo final en tasas de crecimientos a precios constantes de 1997 en un tiempo dado.

RI 20. Obtener formación bruta de capital por tipo de precio en un tiempo dado.

RI 21. Obtener formación bruta de capital en tasas de crecimiento a precios constantes en un tiempo dado.

RI 22. Obtener saldo externo de bienes y servicios por tipo de precio en un tiempo dado.

RI 22. Obtener saldo externo de bienes y servicios en tasas de crecimiento a precios constantes en un tiempo dado.

RI 23. Obtener relaciones entre los principales agregados de Cuentas Nacionales a precios corrientes en un tiempo dado.

RI 24. Obtener financiamiento de la inversión a precios corrientes en un tiempo dado.

RI 25. Obtener financiamiento de la inversión en porcentajes del producto interno bruto en un tiempo dado.

RI 26. Obtener otros indicadores globales sobre población y los recursos laborales en la economía nacional a precios corrientes en un tiempo dado.

RI 27. Obtener otros indicadores globales sobre población y los recursos laborales en la economía nacional en dinámica en precios corrientes en un tiempo dado.

RI 29. Obtener indicadores económicos a precios corrientes en un tiempo dado.
RI 30. Obtener indicadores económicos a precios corrientes en estructura porcentual con respecto al PIB en un tiempo dado.

2.1.6 Requisitos Funcionales

Los requisitos funcionales son aquellos que definen las funciones que el sistema va a llevar a cabo. Deben estar orientados a las necesidades de los usuarios finales. A continuación se muestran los requisitos que han sido identificados para el desarrollo del mercado de datos:

RF1. Realizar extracción de los datos fuentes.

RF2. Realizar transformación y carga de los datos fuentes.

RF3. Autenticar usuario.

RF4. Adicionar usuario.

RF5. Eliminar usuario.

RF6. Adicionar rol.

RF7. Eliminar rol.

RF8. Adicionar reporte.

RF9. Eliminar reporte.

RF10. Modificar reporte.

RF11. Realizar cruce de variables.

RF12. Mostrar consulta MDX.

RF13. Suprimir filas y columnas vacías.

RF14. Mostrar gráfica.

RF15. Imprimir reporte.

RF16. Visualizar reporte.

RF17. Exportar reporte como Excel.

2.1.7 Requisitos no funcionales

Los requisitos no funcionales describen aquellas características no funcionales que los clientes y usuarios desean que tenga el sistema a desarrollar, ejemplos de estas son la seguridad, el

rendimiento, la fiabilidad entre otros. A continuación se presentan algunos de los requisitos no funcionales mediante los cuales se definen las propiedades o cualidades que el producto debe tener, para ver los restantes requisitos no funcionales remitirse al Anexo.

Fiabilidad

RNF 1. Asegurar la disponibilidad del sistema

El sistema debe permanecer disponible las 24 horas del día. En caso de fallo, restablecer el sistema lo más pronto posible.

Restricciones de diseño

RNF 2. Utilizar las herramientas y lenguaje de programación definido durante la investigación

Para la programación en el AD se utilizará PL/pgSQL como lenguaje dentro del SGBD el cual será PostgreSQL en su versión 8.4. Además, se utilizará el lenguaje MDX para realizar las consultas. Como Interfaz de Administración del Gestor de BD se usará el PgAdmin en su versión 1.10. Como herramienta de modelado el Visual Paradigm en su versión 6.4. Para el control de versiones del MD se utiliza el Subversion que permite manejar los archivos, carpetas y sus modificaciones en el transcurso del tiempo en su versión 1.5.4.

De la suite de Pentaho, Pentaho BI Suite Community Edition en su versión 3.5, se usarán los siguientes componentes.

- Schema Workbench 3.2.1: Es la herramienta gráfica que se utiliza para construir el esquema de metadatos multidimensional que soportará la creación de los reportes multidimensionales.
- Pentaho BI Server 3.5: Es el servidor de visualización de la suite, que se encarga de visualizar los reportes, tableros de control digital, controlar el acceso a la información y unificar en una solución de inteligencia de negocios el uso de las demás herramientas que componen la suite.
- Pentaho Administrator Console 3.5: Es la herramienta a usar para administrar el Pentaho BI Server, que permite la administración de las conexiones a las bases de datos, tareas programadas así como los roles y usuarios.
- Pentaho Desing Studio 3.5: Es la herramienta que proporciona el ambiente gráfico para construir y probar documentos de Secuencia de Acción e informes de reportes a través de su colección de editores, visores y módulos de administración integrados.

Se requiere el uso de la Java Virtual Machine para el uso de las herramientas anteriores.

Requisitos de software

RNF 3. Instalar en las estaciones de trabajo el software necesario para el correcto funcionamiento del sistema

Las configuraciones de software de las máquinas clientes deben contar al menos el siguiente elemento:

- Navegador web.
- Java Virtual Machine y Schema Workbench, Pentaho BI Server en caso de que un usuario capacitado requiera la construcción de esquemas multidimensionales para el diseño de nuevos reportes.
- Sistema Operativo debe ser Windows o Linux.

Requisitos de hardware

RNF 4. Proporcionar características mínimas de hardware a las estaciones de trabajo

- Características de un cliente ligero las cuales son: procesador Modelo Intel Celeron, velocidad de 1.33 GHz, 256 de memoria RAM.
- Ordenador 256 MB RAM, Pentium III.

RNF 5. Proporcionar características mínimas de hardware a los servidores

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

- 1 GB RAM.
- 1 Microprocesador Dual Core / 2.2 GHz.
- Disco duro de 160 GB.

Requisitos de seguridad

RNF 6. Restringir el acceso a los servicios alojados en el local de los servidores de información

En el local destinado para los servidores de información en la ONE se gestionarán los servicios correspondientes a la solución de software, los cuales son completamente externos y solo podrán ser invocados desde la sede.

RNF 7. Almacenar de manera cifrada las claves de los usuarios en la base de datos

Las contraseñas de los usuarios no deben ser almacenadas en texto plano, esta información debe ser privada y específica de cada uno, de manera que nadie pueda reemplazar la identidad de un usuario en el sistema.

RNF 8. Persistir la sesión entre componentes

La sesión del usuario activo persistirá en toda la solución de software. Los permisos correspondientes al usuario autenticado se activarán una vez que éste se autentique y en caso de cambiar, tendrá acceso sólo a la información que le compete de acuerdo con sus privilegios.

2.1.8 Caso de uso del sistema

Los casos de uso del sistema representan información de manera visual. Describen lo que debe hacer el sistema relacionado con el usuario, representando generalmente requisitos funcionales. Para una mejor comprensión de los mismos se hace una representación gráfica que contiene los casos de uso y la relación que mantiene con los usuarios que interactúan con él. Para la realización del MD fue necesario identificar casos de uso de información y funcionales, donde se realizó la especificación de cada uno de estos.

Casos de uso de información:

Los casos de uso de información se agrupan por el tipo de información que maneja la ONE y en dependencia de las necesidades de información de los usuarios. Estos se nombran como:

- Analizar indicadores del PIB.
- Analizar indicadores de oferta y demanda global.
- Analizar indicadores del PIB por clase de actividad económica.
- Analizar indicadores económicos.
- Analizar indicadores de precio.
- Analizar indicadores económicos de bienes y servicios
- Analizar indicadores de financiamiento de la inversión.
- Analizar indicadores globales de población.
- Analizar indicadores de los principales agregados.
- Analizar indicadores del SCN.

Casos de uso funcionales

Los casos de uso funcionales identificados están basados en la realización de las operaciones de ETL que se le realizarán a los clasificadores que contienen los datos estadísticos de Cuentas nacionales, además, se basan en la gestión de usuarios para la aplicación y en la realización de las consultas en la base de datos. Estos se nombran como:

- Extraer datos de Cuentas nacionales
- Cargar y Transformar los datos de Cuentas nacionales.

- Administrar usuario.
- Administrar roles.
- Administrar reportes.
- Autenticar usuario.
- Visualizar reportes.

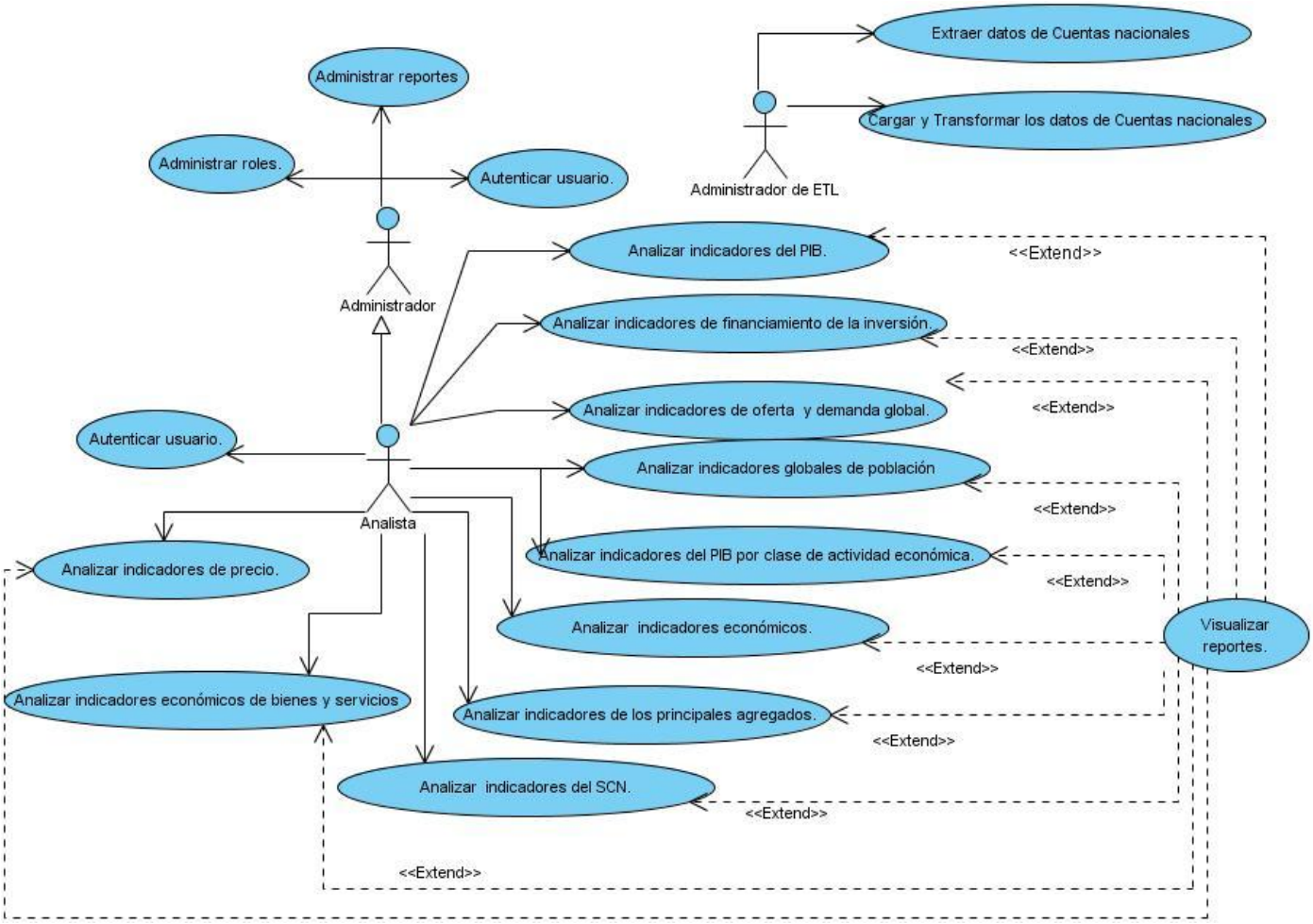


Figura 4: Diagrama de caso de uso del sistema.

Descripción de los casos de usos críticos:

Caso de Uso:	Realizar la extracción de los datos fuentes.
Tipo:	Funcional.
Actores:	Administrador ETL.

Resumen:	El caso de uso inicia cuando el administrador de ETL desea realizar la extracción de datos. Se selecciona la fuente de información correspondiente y extrae los datos contenidos en ella. El caso de uso finaliza cuando todos los datos de la fuente son extraídos.
Precondiciones:	Disponibilidad de las fuentes.
Referencias	RF 1
Prioridad	
Complejidad	Critico.

Flujo Normal de Eventos

Sección “”

Acción del Actor	Respuesta del Sistema.
1. El actor interactúa con la herramienta PDI para realizar la extracción de los datos	2. El sistema muestra los repositorios disponibles para acceder a las transformaciones.
3. El actor selecciona el repositorio con el cual va a trabajar.	**4.** El sistema muestra el área de trabajo de la herramienta.
5. El actor carga la transformación a ejecutar.	**6.** El sistema muestra la transformación seleccionada por el usuario.
7. El actor configura los parámetros de entrada de la transformación y presiona el botón previsualizar.	**8.** El sistema muestra los datos de los ficheros fuentes.
9. El actor ejecuta la transformación seleccionada.	

Flujos Alternos	
Acción del Actor	**Respuesta del Sistema**
	8.1 No responde a la solicitud de conexión.
	8.2 Notifica el error al administrador de ETL.
Prototipo de Interfaz	
	
Poscondiciones	Los datos de los ficheros excel son extraídos de la fuente y almacenados en un área temporal.

Tabla 1: Descripción del Caso de Uso Extraer datos de Cuentas nacionales

Caso de Uso:	Realizar la transformación y carga de los datos fuentes
Tipo:	Funcional.

Actores:	Administrador ETL.
Resumen:	El caso de uso se inicia cuando el administrador de ETL selecciona los datos que van a ser transformados, que previamente habían sido extraídos. El actor realiza las transformaciones pertinentes carga la información hacia el mercado de datos, finalizando así el caso de uso.
Precondiciones:	La información debe ser extraída correctamente hacia el área temporal y las estructuras del almacén deben estar disponibles para ser usadas.
Referencias	RF 2
Prioridad	Crítico.

Flujo Normal de Eventos	
Acción del Actor.	**Respuesta del sistema.**
1. El actor selecciona las estructuras del área temporal a transformar.	
2. El actor carga los datos seleccionados en memoria.	
3. El actor aplica las transformaciones pertinentes y genera datos de auditoría.	
4. El actor carga los datos en el almacén.	5. El sistema ejecuta la consulta para insertar los datos en el almacén.

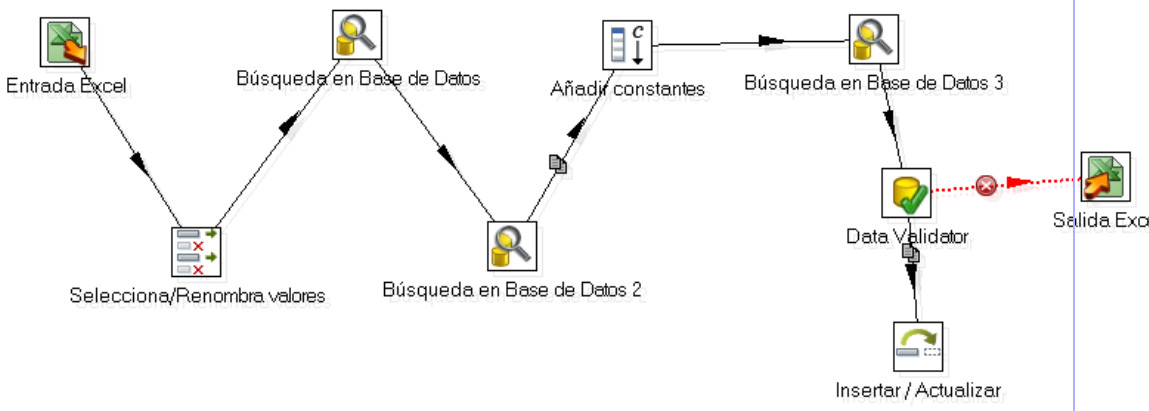
Prototipo de interfaz:	
Poscondiciones	Datos de la fuente transformados y cargados en el mercado de datos.

Tabla 2: Descripción del Caso de Uso Realizar la transformación y carga de Cuentas nacionales

2.2 Diseño del mercado de datos Cuentas nacionales

A continuación se explicara todo lo referente al diseño de propuesta de solución.

2.2.1 Matriz BUS o Matriz Dimensional

El propósito de la Matriz Dimensional es obtener un modelo lógico inicial, donde queda identificado las tablas hechos las cuales forman las áreas de análisis del AD y sus dimensiones relacionadas.

Hechos:

- TH-1: PIB actividad económica.
- TH-2: Serie del PIB.
- TH-3: Saldo externo de bienes y servicios.
- TH-4: Series de oferta y demanda global.
- TH-5: Indicadores de población.
- TH-6: Selección de indicadores del sistema de cuentas nacionales.
- TH-7: Financiamiento de la inversión.
- TH-8: Resumen de indicadores económicos.
- TH-9: Indicadores globales de la población.

- TH-10: Principales agregados de cuentas nacionales.

Dimensiones:

D-1: Temporal año.

D-2: Tipo de precio.

D-3: Tipo de actividad.

D-4: DPA

En la siguiente tabla se relacionan los hechos identificados con las dimensiones propuestas:

Hecho/DIM	temporal_ano	tipo_precio	tipo_actividad	dpa
financiamiento_inversiones	x	x		x
indicadores_globales_poblacion	x	x		x
indicadores_poblacion	x			x
oferta_y_demanda	x	x		x
pib_act_economica	x	x	x	x
principales_agregados_cuent_nac	x	x		x
resumen_indicadores_economicos	x			x
saldo_externo_bienes_servicio	x	x		x
selec_indicadores_scn	x	x		x
serie_pib	x	x		x

Tabla 3: Matriz Bus

2.2.2 Modelo de datos

Seguidamente se muestra el modelo de datos, correspondiente a las dimensiones y hechos seleccionados, así como las medidas especificadas según las series Cuentas nacionales.

.

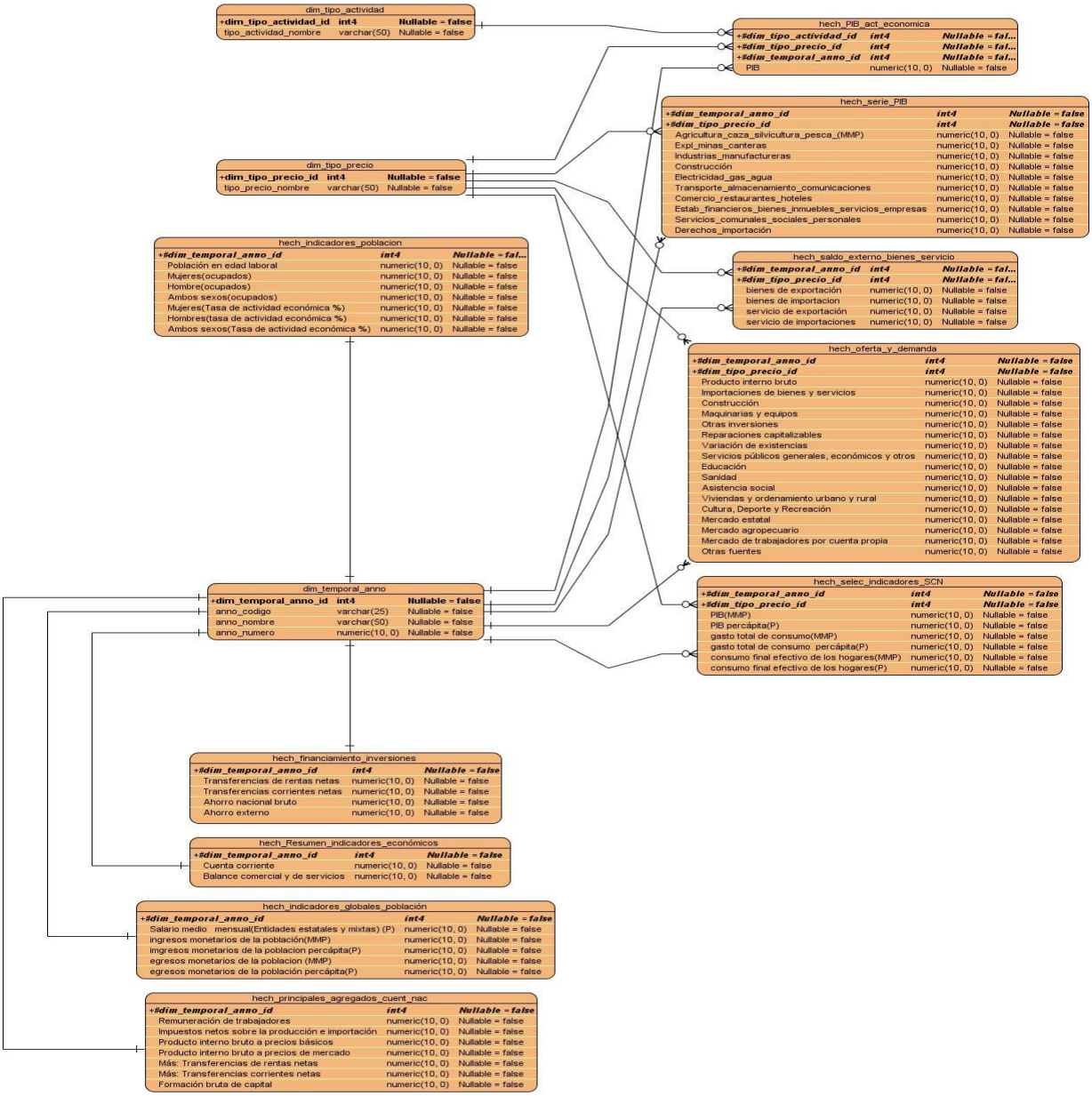


Figura 5: Modelo de datos

2.3 Conclusiones

En este capítulo se definió como tema de análisis Sistema de cuentas nacionales, se definieron los roles y permisos en el mercado de datos, se identificaron 30 requisitos de información, 17 requisitos funcionales y 18 requisitos no funcionales, 4 tablas de dimensiones y 10 tablas de hechos, se confeccionó la matriz BUS y el modelo de datos.

CAPÍTULO 3: IMPLEMENTACIÓN DEL MERCADO DE DATOS CUENTAS NACIONALES

Este capítulo tiene como objetivo principal desarrollar la implementación de los procesos ETL y BI para el área de Cuentas nacionales y darle solución a los requisitos del sistema.

3.1 Implementación de la base de datos

Después de realizado el diseño del modelo dimensional se procede a la transformación al modelo físico generando el script, que muestra la relación existente entre las tablas.

3.2 Estructura de los datos

Los esquemas son una forma más de tener organizada la información dentro de una base datos, éstos definen sus tablas, sus campos en cada tabla y las relaciones entre cada campo y cada tabla. Presentan varios niveles dentro de la base datos, uno de ellos es el esquema conceptual que se utiliza para que el diseñador transmita a la empresa lo que ha entendido sobre la información que ésta maneja. El esquema conceptual se construye utilizando la información que se encuentra en la especificación de los requisitos de usuario. Otro de los esquemas que se encuentran es el esquema lógico, éste fundamentalmente se utiliza para desarrollar el diseño físico ya que brinda una gran fuente de información. Además, desempeña un papel importante durante la etapa de mantenimiento del sistema [20].

En el presente trabajo se definieron 2 esquemas:

- dimensiones: contiene las tablas de las dimensiones propuestas que son comunes con la BD de la ONE.
- mart_cuentas_nacionales: contiene todas las tablas de hechos y las dimensiones propias del MD.

Después de haber diseñado el modelo de datos, la solución propuesta cuenta con 14 tablas en total, 4 dimensiones y 10 hechos, distribuidas en los 2 esquemas anteriormente planteados, como se muestra a continuación:

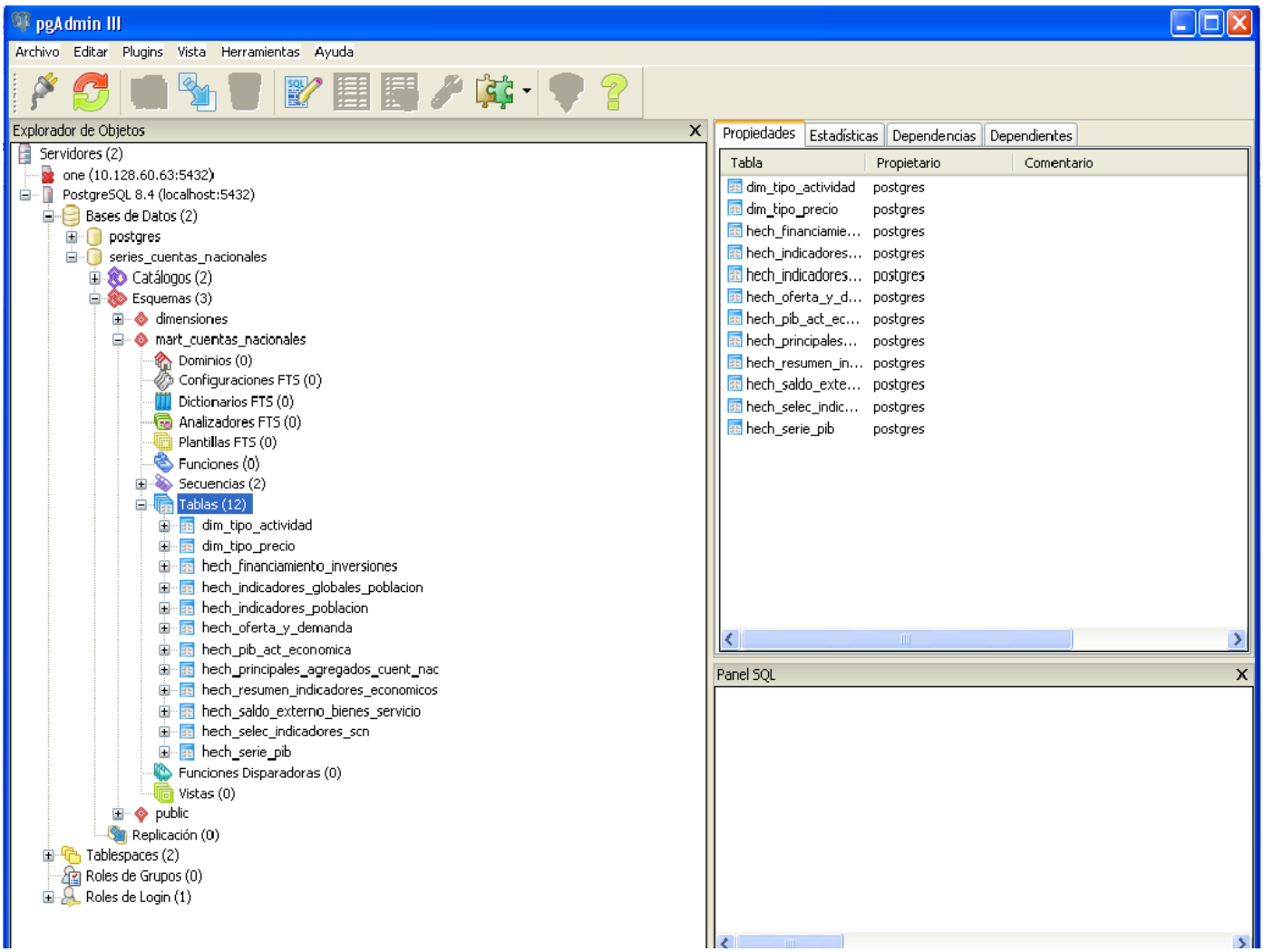


Figura 6: Estructura física de la Base de Datos

3.3 Implementación del subsistema de integración de datos

Para la integración de los datos, es recomendable analizar previamente la fuente de datos, que son aquellos datos que están almacenados en los sistemas fuentes que guardan la información histórica, que se encuentran en formato .xls y que sufrirán cambios para facilitar el trabajo con las transformaciones. Esta etapa de transformación y limpieza es fundamental, es el paso inicial para la carga de datos en la BD. Con la limpieza se detectan los datos erróneos, además de detectar entradas duplicadas y con las transformaciones se combinan y ordenan los datos.

Luego de realizada la extracción de los datos el sistema se encuentra listo para la etapa de transformación. Las transformaciones constituyen un elemento básico dentro de la implementación del

proceso ETL, está compuesta por pasos y estos a su vez se encuentran unidos a través de saltos, dichos pasos son los elementos más pequeños dentro de las transformaciones. Al concluir este proceso, los datos estarán listos para ser cargados [21].

El proceso de carga consiste en cargar todos los datos que han sido transformados anteriormente. Para este proceso se tiene que realizar inicialmente la carga de las tablas que no dependen de la información de otras tablas, con estas características se encuentran las tablas del esquema dimensiones y de las tablas del esquema mart_cuentas_nacionales las que representan dimensiones. Para la realización del proceso de ETL del MD Cuentas nacionales se realizaron 14 transformaciones para la carga de los hechos. A continuación se muestra un ejemplo de la carga correspondiente al hecho pib_act_economica:

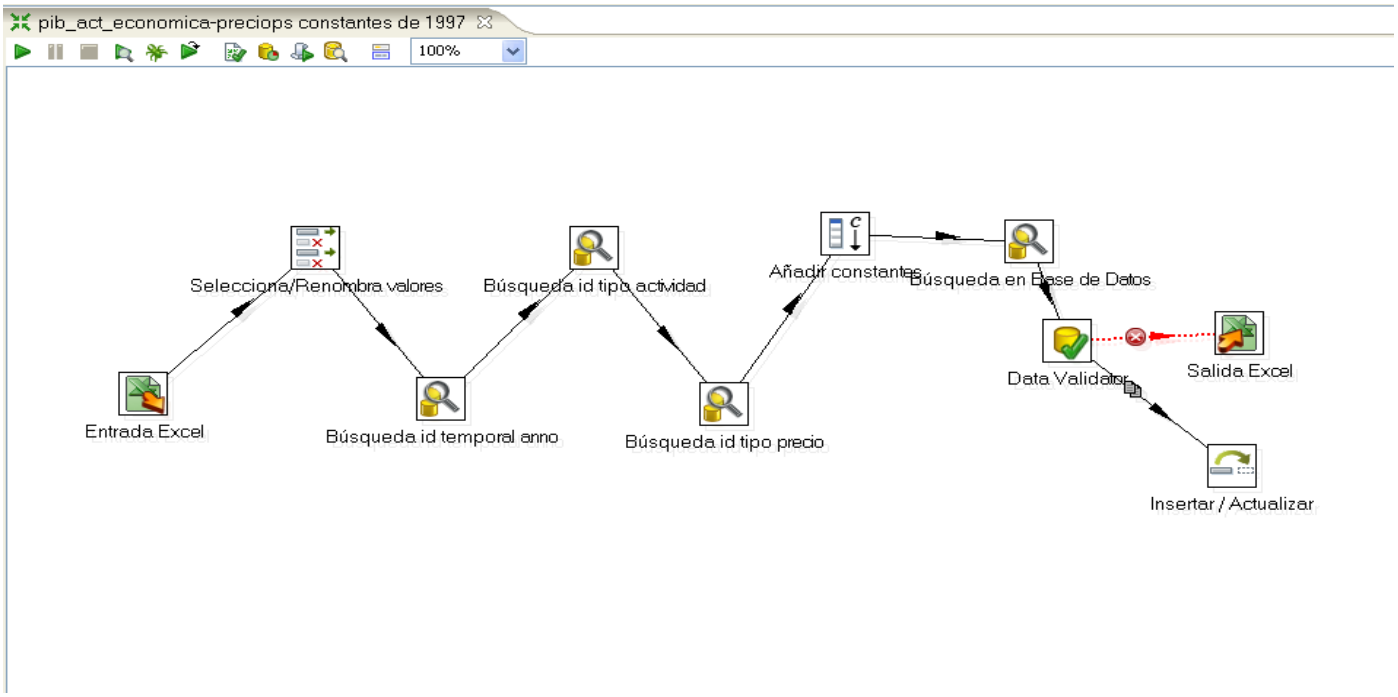


Figura 7: Transformación del hecho pib_act_economica

3.4 Implementación de los trabajos

Un trabajo o job es un conjunto de tareas que tienen como objetivo realizar una acción determinada. En los trabajos se utilizan pasos específicos que son distintos de los disponibles en las transformaciones. Permiten ejecutar una o varias transformaciones de las que han sido diseñadas, siguiendo una secuencia de ejecución para cada elemento que los conforman. Mediante los trabajos se definen el horario y frecuencia de la carga, así como el orden en que van a ser ejecutadas las

transformaciones para poder realizar exitosamente la carga de los datos. A continuación se muestra el trabajo realizado para la implementación del mercado de datos Cuentas nacionales.

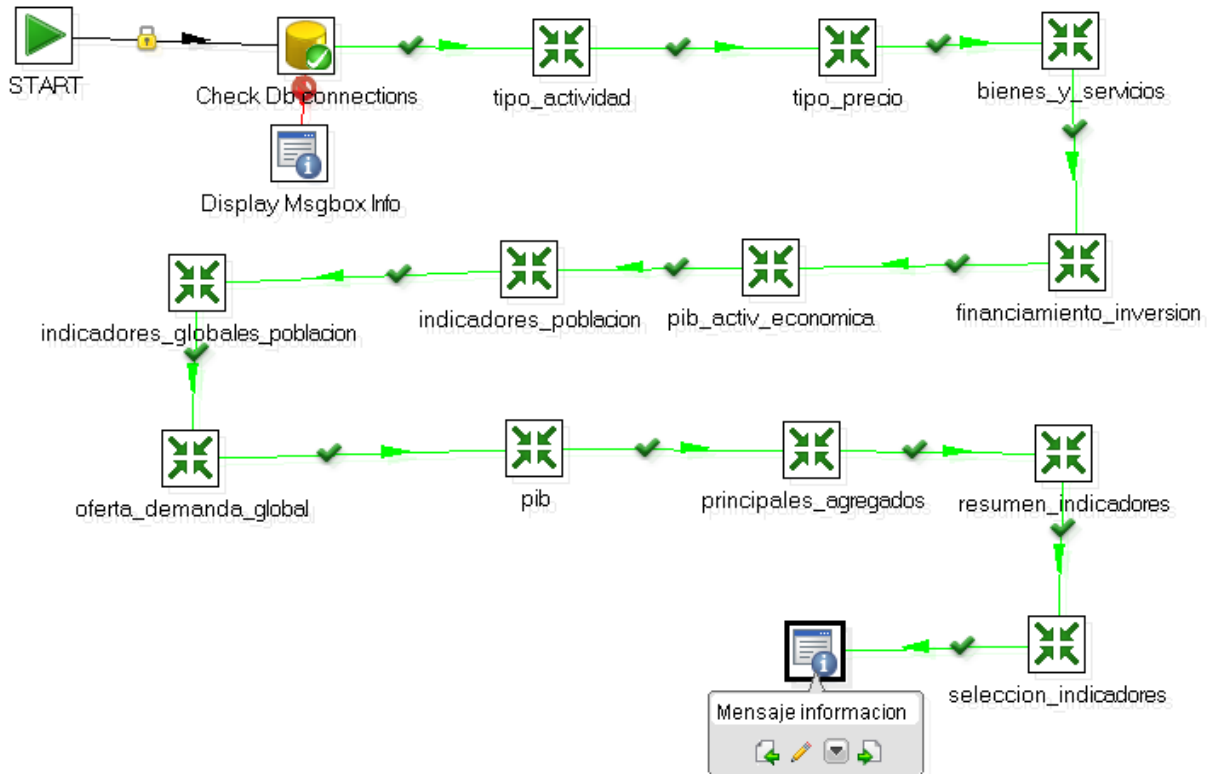


Figura 8: Trabajo del proceso de ETL

3.5 Implementación del subsistema de visualización de datos

3.5.1 Cubos OLAP

Para la visualización de los datos se utilizó Pentaho Schema Workbech que posibilita la creación de los cubos multidimensionales. Esta herramienta genera un fichero .xml que guarda los cubos, define las dimensiones, los niveles de jerarquías de las dimensiones y las medidas. Se modelaron 10 cubos multidimensionales, que se relacionaron con sus respectivas dimensiones y medidas. Para los hechos existe una dimensión que es común para todos, la dimensión temporal año, ya que los reportes se van a mostrar anualmente.

A continuación se muestran los cubos modelados y uno desglosado (oferta_y_demanda), el cual está relacionado con la dimensión tipo de precio ya que los reportes se van a mostrar a precios constantes y precios corrientes. También está relacionado con la dimensión temporal año debido a que los

reportes se muestran anualmente, y con la dimensión Nivel territorial porque la información que se muestra es nacional. Están, además, sus medidas que es donde se van a contener los valores de la oferta y demanda global en sus respectivos años.

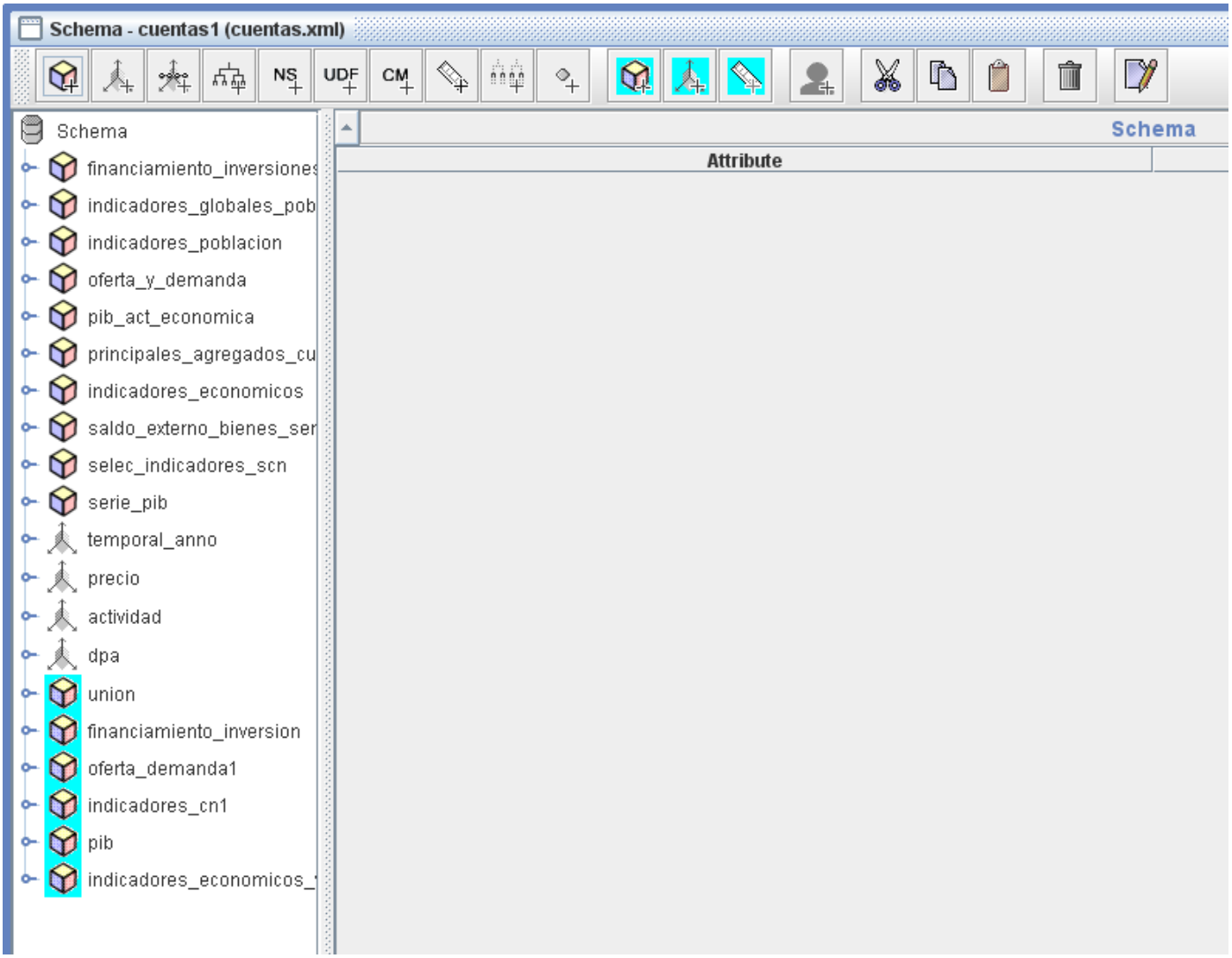


Figura 9: Diseño de los cubos utilizando Pentaho Schema Workbench.

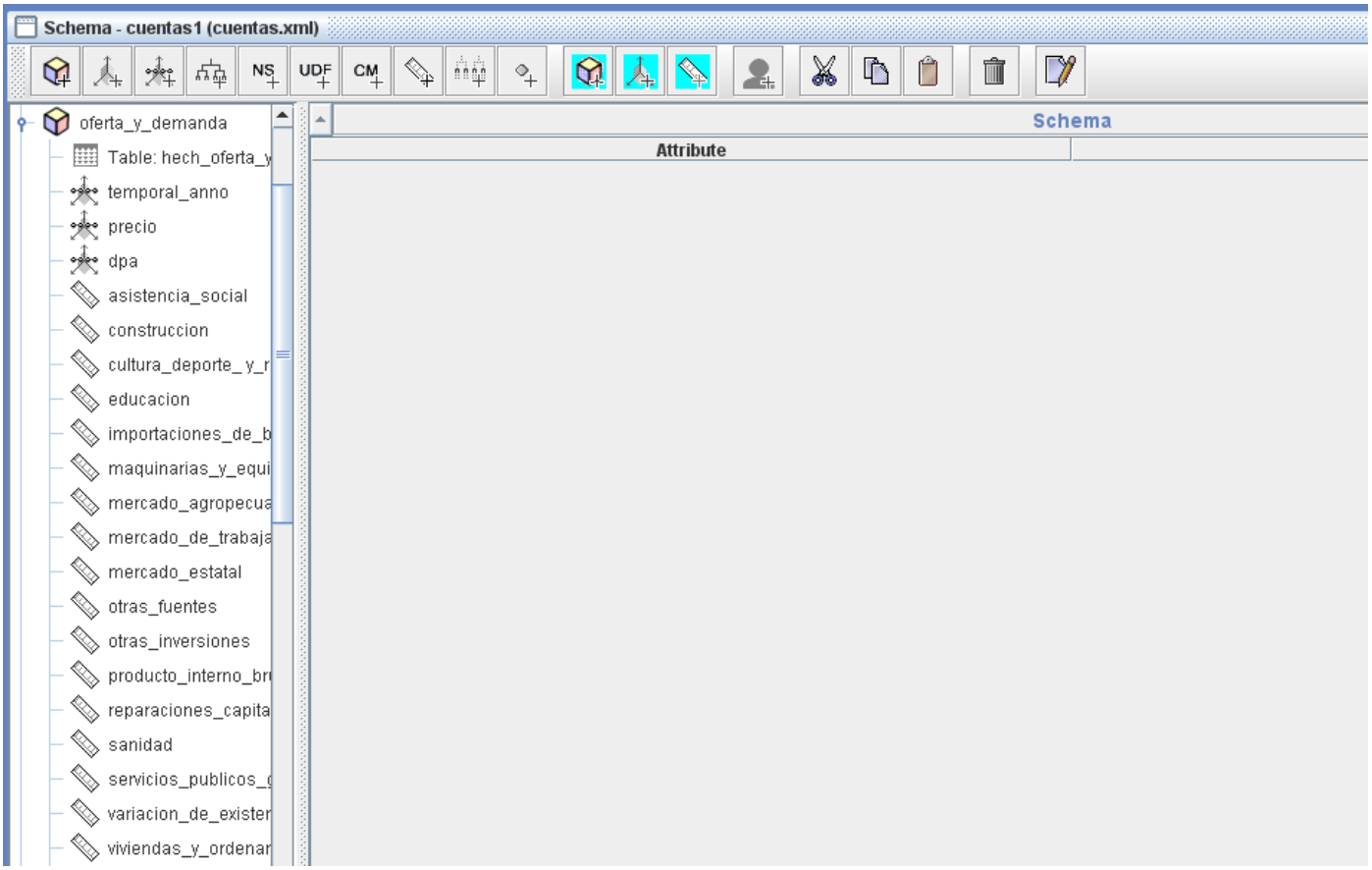


Figura 10: Elementos que componen al cubo oferta y demanda global

3.5.2 Navegación de la capa de visualización

La capa de visualización del MD Cuentas nacionales, contiene 1 áreas de análisis, 10 libros de trabajo, 30 reportes a continuación se detallan los elementos que la componen.

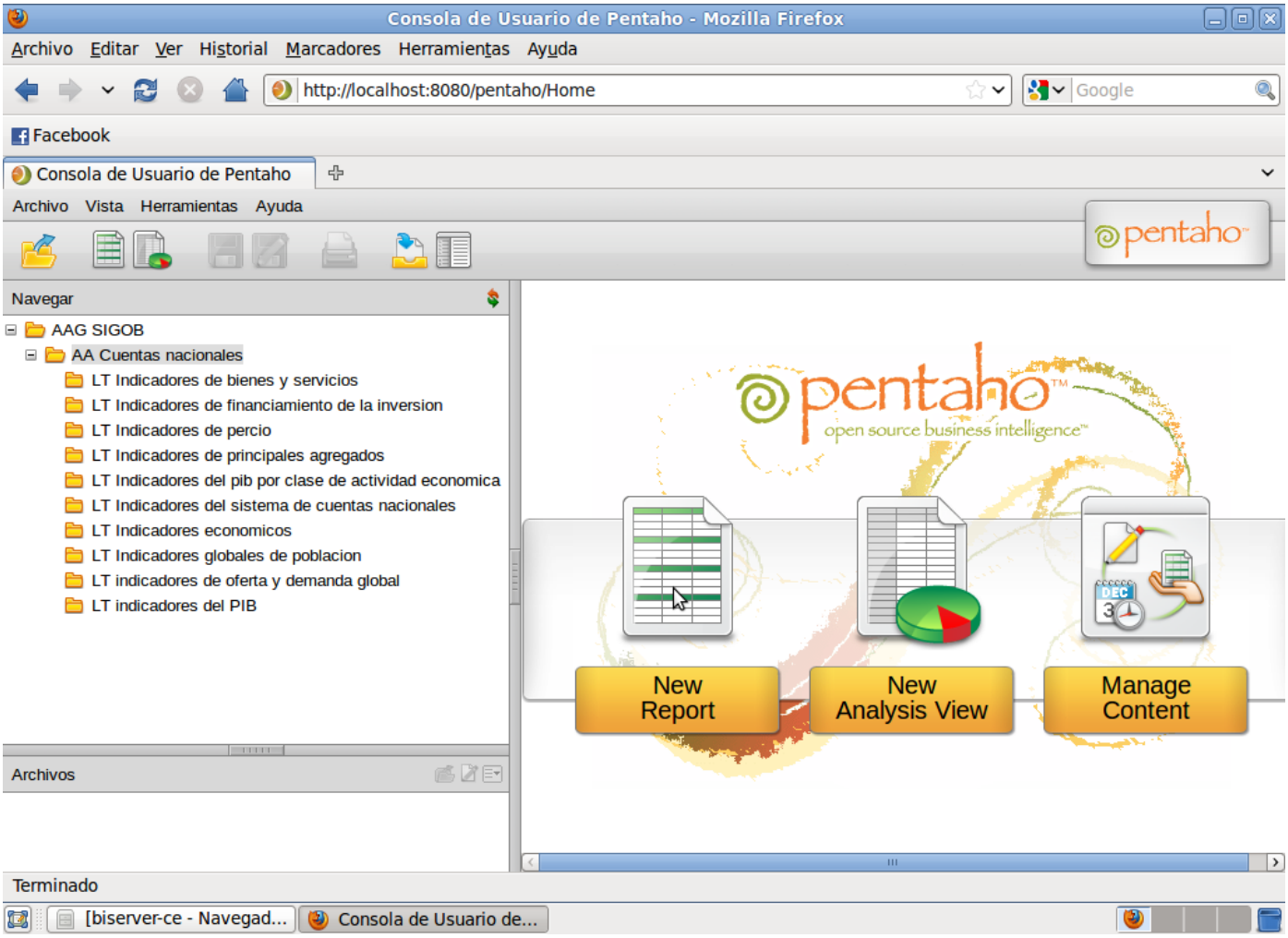


Figura 11: Arquitectura de información

A continuación se detallan los libros de trabajo correspondiente al área de análisis Cuentas nacionales.

AAG SIGOB.

AA Cuentas nacionales.

LT Indicadores de bienes y servicios: Libro de trabajo contenido dentro del área de análisis Cuentas nacionales. Contiene 2 reportes que permiten realizar un análisis de los datos correspondiente a los bienes y servicios.

LT Financiamiento de la inversión: Libro de trabajo contenido dentro del área de análisis Cuentas nacionales. Contiene 2 reportes que permiten realizar un análisis de los datos correspondiente al financiamiento de la inversión.

LT Indicadores de precio: Libro de trabajo contenido dentro del área de análisis Cuentas nacionales. Contiene 3 reportes que permiten realizar un análisis de los datos correspondiente a los índices de precios.

LT Indicadores de principales agregados: Libro de trabajo contenido dentro del área de análisis Cuentas nacionales. Contiene 1 reportes que permiten realizar un análisis de los datos correspondiente a los principales agregados.

LT Indicadores del PIB por clase de actividad económica: Libro de trabajo contenido dentro del área de análisis Cuentas nacionales. Contiene 4 reportes que permiten realizar un análisis de los datos correspondiente al PIB por clase de actividad económica.

LT Indicadores del sistema de cuentas nacionales: Libro de trabajo contenido dentro del área de análisis Cuentas nacionales. Contiene 2 reportes que permiten realizar un análisis de los datos correspondiente al sistema de Cuentas nacionales.

LT Indicadores económicos: Libro de trabajo contenido dentro del área de análisis Cuentas nacionales. Contiene 2 reportes que permiten realizar un análisis de los datos correspondiente a los resúmenes económicos.

LT Indicadores globales de población: Libro de trabajo contenido dentro del área de análisis Cuentas nacionales. Contiene 2 reportes que permiten realizar un análisis de los datos correspondiente a los indicadores globales de población.

LT Indicadores de oferta y demanda global: Libro de trabajo contenido dentro del área de análisis Cuentas nacionales. Contiene 11 reportes que permiten realizar un análisis de los datos correspondiente a la oferta y demanda global.

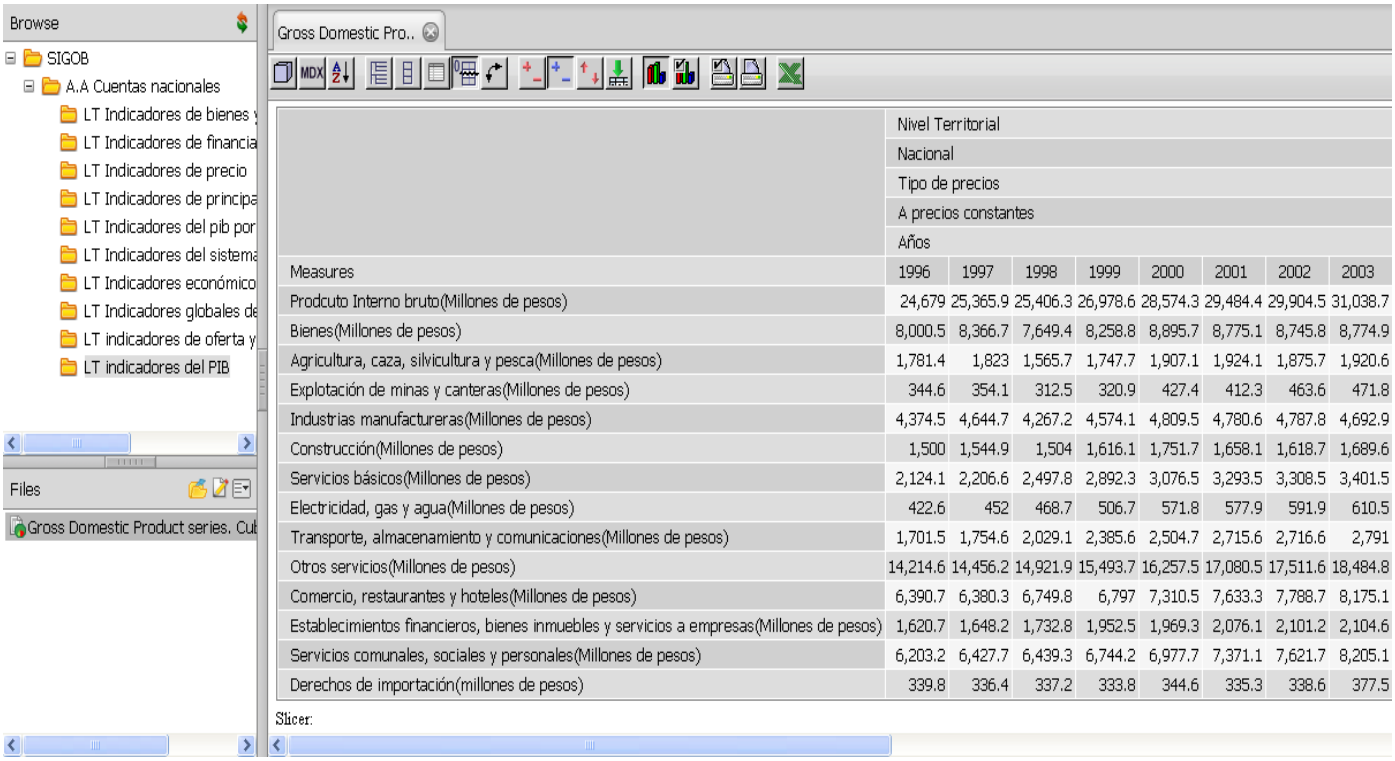
LT Indicadores del PIB: Libro de trabajo contenido dentro del área de análisis Cuentas nacionales. Contiene 1 reporte que permiten realizar un análisis de los datos correspondiente al PIB.

Los reportes se realizan a través de consultas .mdx, a continuación se muestra una consulta hecha para mostrar el reporte serie del Producto Interno Bruto por tipo de precios en un tiempo dado.

select NON EMPTY Crossjoin({[precio].[A precios constantes], [precio].[A precios corrientes]}, {[temporal_anno].[1996], [temporal_anno].[1997], [temporal_anno].[1998], [temporal_anno].[1999], [temporal_anno].[2000], [temporal_anno].[2001], [temporal_anno].[2002], [temporal_anno].[2003], [temporal_anno].[2004], [temporal_anno].[2005], [temporal_anno].[2006]}) ON COLUMNS,

NON EMPTY {[Measures].[prodcuto_interno_bruto], [Measures].[bienes], [Measures].[agricultura_caza_silvicultura_pesca_(mmp)], [Measures].[expl_minas_canteras], [Measures].[industrias_manufactureras], [Measures].[construccion], [Measures].[servicios], [Measures].[electricidad_gas_agua], [Measures].[transporte_almacenamiento_comunicaciones], [Measures].[otrosservicios], [Measures].[comercio_restaurantes_hoteles], [Measures].[estab_financieros_bienes_inmuebles_servicios_empresas], [Measures].[servicios_comunales_sociales_personales], [Measures].[derechos_importacion]} ON ROWS

from [serie_pib]



Measures	1996	1997	1998	1999	2000	2001	2002	2003
Prodcuto Interno bruto(Millones de pesos)	24,679	25,365.9	25,406.3	26,978.6	28,574.3	29,484.4	29,904.5	31,038.7
Bienes(Millones de pesos)	8,000.5	8,366.7	7,649.4	8,258.8	8,895.7	8,775.1	8,745.8	8,774.9
Agricultura, caza, silvicultura y pesca(Millones de pesos)	1,781.4	1,823	1,565.7	1,747.7	1,907.1	1,924.1	1,875.7	1,920.6
Explotación de minas y canteras(Millones de pesos)	344.6	354.1	312.5	320.9	427.4	412.3	463.6	471.8
Industrias manufactureras(Millones de pesos)	4,374.5	4,644.7	4,267.2	4,574.1	4,809.5	4,780.6	4,787.8	4,692.9
Construcción(Millones de pesos)	1,500	1,544.9	1,504	1,616.1	1,751.7	1,658.1	1,618.7	1,689.6
Servicios básicos(Millones de pesos)	2,124.1	2,206.6	2,497.8	2,892.3	3,076.5	3,293.5	3,308.5	3,401.5
Electricidad, gas y agua(Millones de pesos)	422.6	452	468.7	506.7	571.8	577.9	591.9	610.5
Transporte, almacenamiento y comunicaciones(Millones de pesos)	1,701.5	1,754.6	2,029.1	2,385.6	2,504.7	2,715.6	2,716.6	2,791
Otros servicios(Millones de pesos)	14,214.6	14,456.2	14,921.9	15,493.7	16,257.5	17,080.5	17,511.6	18,484.8
Comercio, restaurantes y hoteles(Millones de pesos)	6,390.7	6,380.3	6,749.8	6,797	7,310.5	7,633.3	7,788.7	8,175.1
Establecimientos financieros, bienes inmuebles y servicios a empresas(Millones de pesos)	1,620.7	1,648.2	1,732.8	1,952.5	1,969.3	2,076.1	2,101.2	2,104.6
Servicios comunales, sociales y personales(Millones de pesos)	6,203.2	6,427.7	6,439.3	6,744.2	6,977.7	7,371.1	7,621.7	8,205.1
Derechos de importación(millones de pesos)	339.8	336.4	337.2	333.8	344.6	335.3	338.6	377.5

Figura 12: Reporte Series del Producto interno bruto

3.6 Conclusiones

En este capítulo se realizó un análisis de la fuente de datos y los temas para la implementación del MD. Se mostró como quedó físicamente el almacén, a partir del cual se obtuvo resultados satisfactorios logrando poblarlo satisfactoriamente.

- Se implementó el proceso ETL, extrayendo los datos de las series de Cuentas nacionales los cuales fueron transformados y cargados.
- Quedó definida la estructura de los datos a partir del modelo de datos físico, contando con 2 esquemas: dimensiones y mart_cuentas_nacionales, compuesto por 4 dimensiones y 10 hechos.
- Se diseñaron e implementaron los cubos OLAP, quedando definidos 10 cubos, 4 dimensiones y 63 medidas.
- Se desarrollaron los subsistemas de visualización, determinándose 10 libros de trabajo, con un total de 31 vistas de análisis.

CAPITULO 4: VALIDACIÓN DEL MERCADO DE DATOS CUENTAS NACIONALES

4.1 Introducción

En el presente capitulo se realiza la configuración de la seguridad de los usuarios, se aplican los casos de pruebas y listas de chequeo para validar los requisitos y verificar si el producto final satisface o no, las necesidades de los clientes.

4.2 Configurar la seguridad de los usuarios

Pentaho BI Server brinda la posibilidad de agrupar a los usuarios según las necesidades de permisos y accesos que necesitará cada rol para realizar su función como trabajador del sistema.

Usuarios y Roles:

Los usuarios de una base de datos son una entidad de seguridad de la base de datos, a estos se les asignan los permisos pertinentes para velar por la seguridad de la BD. Se tendrán en cuenta la creación de usuarios para el uso de la herramienta en correspondencia de los roles que estos desempeñen a la hora de interactuar con esta.

Se definieron los usuarios y roles que trabajarán con el mercado de datos, para facilitar la organización del trabajo de los desarrolladores y garantizar la fiabilidad del sistema. A continuación se muestra cada uno de ellos:

- **Admin:** es el rol del administrador del sistema y tiene acceso a la aplicación en su totalidad, dígase administración y configuración tanto de los reportes, los usuarios como los roles.
- **Especialista:** su rol se basa en consultar la información y ver desde las diferentes perspectivas de análisis.
- **Authenticated:** es un rol por defecto que se le asigna a cada usuario cuando se adiciona y es el que le permite autenticarse en el sistema, es por eso que todos los usuarios tienen este rol además del específico definido por el trabajo que realiza.

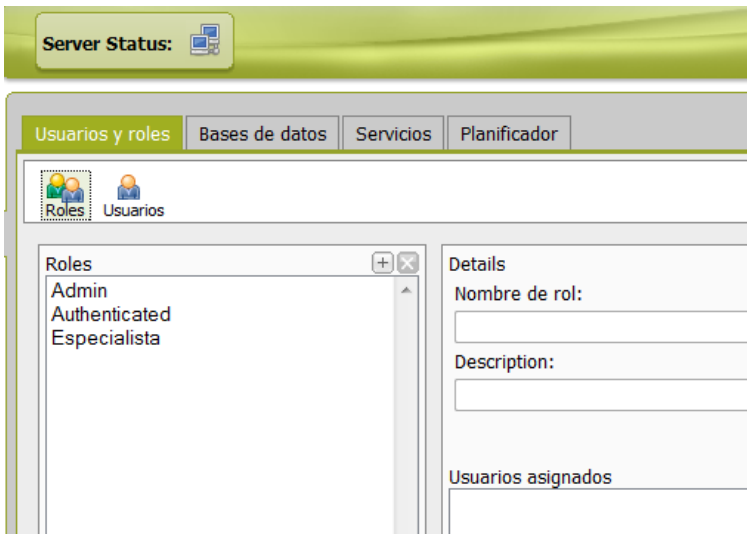


Figura 13: Usuarios y roles

Privilegios:

Los privilegios que se les asigna a los usuarios del sistema son basados en el rol que desempeñan. En caso del:

- **Administrador:** Es el usuario que responde al rol de Admin y se le asigna el permiso control total.
- **Especialista:** Responde al rol Especialista y al contrario del administrador este tipo de usuario solo tiene derechos de ejecutar los reportes y verlos desde las diferentes perspectivas de análisis pero sin poder guardar los cambios.

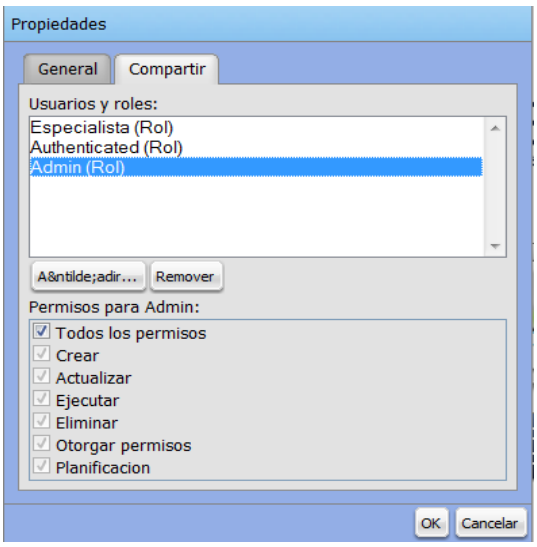


Figura 14: Roles y permisos

4.3 Pruebas

El desarrollo de sistemas de software implica la realización de una serie de actividades, que permiten verificar y revelar la calidad de un producto de software. Con el objetivo de validar el funcionamiento del mismo, es necesario incorporar pruebas que garanticen la calidad del software e identifiquen posibles fallos. Las pruebas de *software* son los procesos que permiten verificar y revelar la calidad de un producto. Son utilizadas para identificar posibles fallos de implementación, calidad y usabilidad en la solución. Para determinar el nivel de calidad se deben efectuar pruebas que permitan comprobar el grado de cumplimiento de las especificaciones iniciales del sistema. Para la validación del MD Cuentas nacionales se pueden aplicar diferentes tipos de pruebas a continuación se especifican algunas de ellas:

La prueba unitaria que no es más que una prueba que valida que las unidades individuales están trabajando correctamente [23].

Su objetivo es identificar defectos en las interfaces y las interacciones entre los componentes del sistema, ya que se unen para formar grandes subsistemas progresivamente. Dado que los componentes del sistema son identificados y diseñados durante el diseño físico, la mayoría de la información utilizada en la planificación de pruebas de integración deriva del diseño físico [24].

La prueba de sistema es una técnica para verificar si un sistema completo cumple con las especificaciones acordadas. El propósito de la prueba de sistema es detectar discrepancias entre el comportamiento del sistema construido y su especificación [25].

Pruebas de aceptación: son pruebas funcionales, pero vistas directamente desde el cliente. Digamos que son aquellas pruebas que demuestran al cliente que la funcionalidad está terminada y funciona correctamente. Las herramientas utilizadas para aplicar las pruebas de la solución definidas por el centro DATEC son los casos de pruebas y las listas de chequeo.

Casos de prueba: que no es más que un conjunto de entradas y condiciones que arrojan resultados esperados desarrollados con un objetivo particular [26].

4.3.1 Diseño de casos de pruebas

Los casos de prueba son un conjunto de condiciones o variables, bajo las cuales se determinará si los requisitos de una aplicación, son parcial o completamente satisfactorios. A continuación se muestra un ejemplo de un caso de prueba de interfaz, aplicado al caso de uso de información “Analizar indicadores del Producto interno bruto por clase de actividad económica”, para ver los restantes casos de prueba, remitirse al Anexo.

Escenario	Descripción	Variables		Descripción de la funcionalidad	Flujo central
		Perfiles de análisis	Indicadores a medir		
EC 1.1: Deflactor implícito del Producto interno bruto por clase de actividad económica	Reporte que se muestran el deflactor implícito del producto interno bruto por clase de actividad económica según el tipo de actividad por años**.**	Años Tipo de precio Tipo de actividad Nivel territorial	Deflactor(Base=1997)	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 e área de análisis de **A.A Cuenta**
EC 1.2: Estructura del Producto interno bruto por clase de actividad económica	Reporte que se muestran la estructura del producto interno bruto por clase de actividad económica según el tipo de precio	Años Tipo de precio Nivel territorial	Producto interno bruto Agricultura, ganadería y silvicultura Pesca Explotación de		

	por años		minas y canteras Industria azucarera Industrias manufactureras (excepto Industria azucarera) Construcción Suministro electricidad, gas y agua Transportes, almacenamiento y comunicaciones Comercio; reparación de efectos personales Hoteles y restaurantes Intermediación financiera Servicios empresariales, actv. inmobiliarias y de alquiler Administración pública, defensa; seguridad socia Ciencia e innovación tecnológica Educación Salud pública y asistencia socia Cultura y deporte Otras actv de serv. comunales, de asociaciones y personales Derechos de importación		**nacionales.** Se selecciona e libro de trabajo **L.T Indicadore del producto interno bruto por clase de actividad económica** En la parte inferior izquierd se selecciona e reporte desead En el área de trabajo se visualiza la tabl correspondient al reporte.
EC 1.3:	Reporte que se	Años	Cantidad(Millones		

Producto interno bruto por clase de actividad económica a precio de mercado según Nomenclador de actividad económica de Cuba	muestran el producto interno bruto por clase de actividad económica a precio de mercado según Nomenclador de actividad económica de cuba por años	Tipo de precios Tipo de actividad Nivel territorial	de pesos)
EC 1.4: Tasas del producto interno bruto por clase de actividad económica	Reporte que se muestran las tasas de crecimiento del producto interno bruto por clase de actividad económica según el tipo de precio por año	Año Tipo de precios Tipo de actividad Nivel territorial	Tasas de crecimiento

Tabla 5: Caso de prueba analizar indicadores del Producto interno bruto por clase de actividad económica

4.4 Elaboración, evaluación y aplicación de las listas de chequeo

Las listas de chequeo se realizan con el objetivo de evaluar la calidad de los artefactos que se generan en el análisis de la solución. Mediante estas listas se pueden cubrir todos los aspectos aplicables en los temas de análisis, particularmente para el tema de Cuentas nacionales en las oficinas de la ONE.

La lista de chequeo definida para la evaluación del MD, cuenta con varios puntos a medir, los mismos son: estructura del documento, Indicadores definidos en el desarrollo y semántica del documento. La lista recoge 14 indicadores a evaluar, de ellos ocho críticos (Ver Tabla 4).

Elementos que forman parte de la estructura de la Lista de Chequeo:

Peso: Define si el indicador a evaluar es crítico o no.

Evaluación (Eval): Es la forma de evaluar el indicador en cuestión. El mismo se evalúa de 1 en caso de mal y 0 en caso que elemento revisado no presente errores.

Cantidad de elementos afectados: Especifica la cantidad de errores encontrados sobre el mismo indicador.

Comentario: Especifica los señalamientos o sugerencias que quiera incluir la persona que aplica la lista de chequeo**.**

Estructura del Documento: Abarca todos los aspectos definidos por el expediente de proyecto o el formato establecido por el proyecto.

Indicadores definidos en el desarrollo: Abarca todos los indicadores a evaluar según el desarrollo.

Semántica del documento: Contempla todos los indicadores a evaluar respecto a la ortografía, redacción y demás.

N.P. (No Procede): Se usa para especificar que el indicador a evaluar no se puede aplicar en ese caso.

Estructura del documento					
Peso	**Indicadores a evaluar**	**Eval**	**(NP)**	**Cantidad de elementos afectados**	**Comentarios**
crítico	1. ¿Los entregables contienen las secciones obligatorias de la plantilla estándar definidas para un expediente de proyecto? (portada, control de versiones, reglas de confidencialidad, tabla de contenidos y contenido) (ver expediente de proyecto)				
Indicadores definidos en el desarrollo					
Peso	**Indicadores a evaluar**	**Eval**	**(NP)**	**Cantidad de elementos afectados**	**Comentarios**
	1. ¿Se utilizó un lenguaje cuyas sentencias son expresables mediante una				

	sintaxis bien definida?				
crítico	2. ¿Los reportes son configurables a través de la interfaz del sistema?				
	3. ¿La interfaz está orientada a facilitar el uso de las funciones del sistema por parte de los usuarios?				
crítico	4. ¿No existen restricciones para construir cubos OLAP con dimensiones y niveles de agregación ilimitados?				
crítico	5. ¿Los usuarios son capaces de manipular los resultados de manera que se ajusten a sus necesidades, conformando nuevos reportes?				
	6. ¿El sistema responde de una forma rápida a la información que le sea solicitada por el usuario?				
	7. ¿El sistema refleja cualquier lógica del negocio para poder responder a preguntas específicas?				
crítico	8. ¿El sistema garantiza la confidencialidad y seguridad de acceso a los datos por rol de los usuarios?				
	9. ¿Los datos e información derivados del proceso de análisis realizado mediante la aplicación, apoyan la toma de decisiones en la Institución?				
crítico	10. ¿Los cambios en los datos se reflejan automáticamente en los reportes de forma instantánea?				
Semántica del documento					
Peso	**Indicadores a evaluar**	**Eval**	**(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?				

Tabla 4: Lista de chequeo

Resultados y discusión de las pruebas

Para la validación de la solución se aplicó por parte de los especialistas del departamento los casos de prueba diseñados y la lista de chequeo elaborada. En una primera iteración se identificaron 26 no conformidades, las que posteriormente se corrigieron para lograr una correcta disponibilidad de la información y en estos momentos el MD está pasando por una segunda iteración en la que está siendo evaluado por los especialistas de Calisoft.

4.5 Conclusiones

En el presente capítulo se describió el desarrollo y la aplicación de una lista de chequeo, arrojando como resultado la identificación de forma general, de 14 indicadores, de ellos 8 indicadores con peso crítico. Se aplicaron las pruebas pertinentes para verificar la calidad, alcanzando el producto una evaluación de Bien. Además, se obtuvo la carta de aceptación por parte del cliente, quedando certificado que el sistema cumple con las necesidades del cliente.

Conclusiones generales.

Al concluir la solución se puede plantear que fueron cumplidos los objetivos trazados y las tareas de investigación propuestas. Por tanto, se arriban a las siguientes conclusiones:

- A partir del estudio realizado, se logra definir en el marco teórico de la investigación, que los mercados de datos son la solución más idónea para la situación problémica planteada.
- Mediante el proceso de análisis, diseño e implementación, se logra la estructura final del Mercado de Datos de Cuentas nacionales para la ONE.
- A través del proceso de validación mediante las listas de chequeo y los casos de prueba, se concluye que la solución desarrollada posee 26 no conformidades las cuales fueron resueltas, por lo que se estima que los resultados que se obtienen son satisfactorios.

- Proponer un mecanismo más efectivo para el tratamiento de errores de los ficheros fuentes, que le permita a los usuarios cargar los datos ya arreglados, sin necesidad de cargar toda la información nuevamente.
- Crear un repositorio para los metadatos, donde se guarde toda la información de la ejecución de las transformaciones de una manera más detallada y entendible para el administrador de ETL.
- Que se ponga en explotación el mercado de datos para probar las funcionalidades del mismo y detectar nuevas funcionalidades.

Bibliografía

1. Bernabeu, D. R. (2007). *Metodología propia para la Construcción de un Data Warehouse.* Córdoba, Argentina. Disponible en: http://www.dataprix.com/files/DWH_Metodologia_HEFESTO-V1.0.pdf
2. Chrysler, D., & Otros. (Agosto de 2000). *CRISP-DM 1.0 Guía paso a paso de Mineria de Datos.* Disponible en: http://www.dataprix.com/files/Metodologia_CRISP_DM.pdf
3. CICLO DE VIDA DEL SOFTWARE. [en línea] (2008). [Consulta 1 de junio del 2011]. Disponible en: < http://es.kioskea.net/contents/genie-logiciel/cycle-de-vie.php3_/>
4. *Cicpc.* (2009). Visitado el 21 de 10 de 2010, Disponible en: http://www.cicpc.gob.ve/index.php
5. Colectivo de autores Capri Software. S.L. [En línea] [Citado el: 16 de febrero de 2010.] http://www.capris.es/talend/Folleto_Talend_Open_Profiler.pdf
6. Colectivo de autores. 2010**.** *METODOLOGÍA PARA EL DESARROLLO DE SOLUCIONES DE ALMACENES DE DATOS E INTELIGENCIA DE NEGOCIO EN CENTALAD.* Habana, Cuba: s.n., 2010.
7. Decloix, S. (2008). *Les ETL Open Source "Une réelle alternative aux solutions propriétaires".* Disponible en: http://www.atolcd.com/fileadmin/Publications/Atol_CD_Livre_Blanc_ETL_Open_Source_01.pdf
8. Ferrari, A., & Russo, M. (October 1, 2008). *SQLBI METHODOLOGY AT WORK.* Disponible en: http://www.dataprix.com/files/SQLBI Methodology_At Work draft.0-1.pdf
9. gulsin [Online] [Cited: 03 24, 2011.] http://gulsin.org/2011/03/14/transformaciones-en-pentaho-kettle/.
10. Headquarters. (2009). *Visual Paradigm for UM*. Visitado el 17 de 12 de 2010, de http://www.visual-paradigm.com/product/vpuml
11. Hobbs y otros, Lilian. 2005. *Oracle Database 10g Data Warehousing.* EUA : ELSEVIER Digital Press, 2005.
12. Knight, B. (October 2008). *Professional Microsoft SQL Server 2008 Integration Services (Wrox Programmer to Programmer).* Disponible en: http://www.ebooks-space.com/ebook/1250/Professional-Microsoft-SQL-Server-2008-Integration-Services.html

13. Kristin Daniel.**INTEGRATION TESTING** [en línea] (2000). [Consulta 1 junio 2011] Disponible en: <http://www.sei.cmu.edu/intro/process/techngs/q_it.htm />.
14. Kristin Daniel.**FUNCTIONAL SYSTEM TESTING** [en línea] (2000). [Consulta 1 junio 2011] Disponible en: <http://www.sei.cmu.edu/intro/process/techngs/q_fst.htm />
15. **León, Eduardo. 2007.** Blog de Eduardo León. *Visual Paradigm, una herramienta de lo más útil.* [En línea] 2 de Abril de 2007. [Citado el: 3 de Diciembre de 2010.] http://slion2000.blogspot.com/2007/04/visual-paradigm-una-herramienta-de-lo.html.
16. Lsoto. (2008). *Mitecnologico*. Recuperado el 04 de 05 de 2010, de Mitecnologico:
17. Mustelier, Doris Medina. 2009. *Técnicas de ETL del SINSC en la República Bolivariana de las Américas.* Habana, Cuba: s.n., 2009
18. Pentaho. (2009). *Pentaho Open Source Business Intelligence: Kettle Project.* Visitado el 25 de 11 de 2010, de http://kettle.pentaho.org
19. **Portada sobre la plataforma Pentaho Open Source. 2010.** Portada sobre la plataforma Pentaho Open Source . *Business Intelligence.* [En línea] 2010. [Citado el: 2 de Diciembre de 2010.] http://pentaho.almacen-datos.com/mondrian.html.
20. PRUEBAS DE SOFTWARE [en línea] (2009). [Consulta 1 junio 2011] Disponible en: <http://www.slideshare.net/aracelij/pruebas-de-software/>.
21. **ralfm. 2007.** El rincón de Linux. *Introducción a PostgreSQL - Instalación e inicialización.* [En línea] Rafael Martinez, 08 de Junio de 2007. [Citado el: 29 de Noviembre de 2010.] http://www.linux-es.org/node/536.
22. **Ralph Kimball, Joe Caserta. 2004.** *The Data Wherehouse ETL toolkit.* s.l. : Wiley Publishing, Inc., 2004.
23. **Sánchez, Leopoldo Zenaido Zepeda. 2008.** *Metodología para el Diseño.* Valencia : s.n., 2008.
24. **Sherman Wood, JasperSoft. 2007.** pentaho. *Mondrian Documentation.* [En línea] Abril de 2007. [Citado el: 2 de Diciembre de 2010.] http://mondrian.pentaho.com/documentation/workbench.php.
25. **SINNEXUS. 2008.** SINNEXUS. *Business Intelligence+Informatica estrinta.* [En línea] 2008. [Citado el: 29 de Noviembre de 2010.] http://www.sinnexus.com/business_intelligence/index.aspx.
26. **Soluciones AG. 2009.** Soluciones AG. *Soluciones para alta gerencia AG, SA.* [En línea] 2009. [Citado el: 25 de Noviembre de 2010.] http://www.soluciones-ag.com/pdf_productos/Spanish_ER-Studio_Datasheet_2009.pdf.

27. **Summan. 2010.** Summan. *Manejo documental e infraestrutura informatica.* [En línea] 2010. [Citado el: 30 de Noviembre de 2010.] http://www.summan.com/index.php/productos/software/pentaho-.html

28. **UNIT TESTING EXAMPLE CONCEPTS AND FRAMEWORKL.** [en línea] (2008). [Consulta 1 junio 2011] Disponible en:< http://mauriziostorani.wordpress.com/>

29. Unión de Asociaciones, y Entidades de Atención al D. 2007. UNAD. *UNAD.* [En línea] 2007. [Citado el: 10 de 03 de 2010.] http://calidad.unad.org/asesoramiento/definicion_de_indicadores.html.

30. **Web, Elección del servidor de aplicaciones. 2003.** Información técnica y de gestión para IBM. [En línea] Febrero de 2003. [Citado el: 3 de Diciembre de 2010.] http://www.help400.es/asp/scripts/nwart.asp?Num=131&Pag=10&Tip=T.

31. **Wolff, Carmen Gloria. 2002.** Departamento Ingenieria y ciencia de la cumputación. [En línea] 28 de Agosto de 2002. [Citado el: 29 de Noviembre de 2010.] http://www.inf.udec.cl/~revista/ediciones/edicion4/modmulti.PDF.

Referencias Bibliográficas

[1] Unión de Asociaciones, y Entidades de Atención al D. 2007. UNAD. *UNAD.* [En línea] 2007. [Citado el: 10 de 03 de 2010.] Disponible en: http://calidad.unad.org/asesoramiento/definicion_de_indicadores.html/.

[2] Manjit, S. (2000).- ***Developing a Corporate Data Warehousing Strategy*** **Enterprise System Integration**, Chapter 33, CRC Press LLC, Boca Raton, Florida, 449-467.

[3] Imohoff, Claudia, Galemmo, Nicholas y Geiger, Jonathan G. 2003. *Mastering Data Warehouse Desing Relational and Dimensional Techniques.* s.l. : Wiley Publishing, Inc, 2003.

[4] Alonso, Roberto Abajo. 2006. *DATA WAREHOUSE.* Madrid : s.n., 2006.

[5] **Cabrera, María Evelia Casales. 2009.** Universidad autónoma de México. [En línea] 2009. [Citado el: 4 de noviembre de 2010.] Disponible en: http://hp.fciencias.unam.mx/~alg/bd/dwh.pdf.

[6] **Peñaloza, Lucía Victoria Hernández. 2008.** *Tesis para logra el título de Magíster: Diseño y Construcción de un Data Mart para la mantención de Indicadores de Sostenibilidad de la Industria del Salmón. Chile: s.n., 2008.*

[7] DIAZ MORALES, Themis Patricia y BERMUDEZ RODRIGUEZ, José Salvador. Diseño de un Datawarehouse para los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular. Tesis (Ingeniero en Ciencias Informáticas) Ciudad Habana. Universidad de las Ciencias Informáticas, 2010.

[8] Duque, Raúl González. 2008. Mundo Geek. *Mundo Geek.* [En línea] 2008. [Citado el: 23 de 02 de 2010.] Disponible en: http://mundogeek.net/archivos/2004/08/26/modelo-de-datos/.

[9] Hobbs y otros, Lilian. 2005. *Oracle Database 10g Data Warehousing.* EUA : ELSEVIER Digital Press, 2005.

[10] WOLFF, C. G. *Modelamiento Multidimensional* 2002.

[11] **Ponniah, Paulraj. 2001**. *Data Warehousing Fundamentals*. EUA : Wiley Publishing Inc, 2001.

[12] **Kimball, Ralph.** The Data Warehouse Lifecycle Toolkit. EUA: Wiley Publishing Inc.

[13] **Nader, Ing. Javier.***"Sistema de apoyo gerencial universitario" .* 2003. Tesis de Magister en Ingeniería de Software.

[14] [En línea] [Citado el: 21 de 11 de 2010.] Disponible en: http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/OLAP.pdf.

[15] 2010. Introducción a Pentaho Business Intelligence. [En línea] 2010. Disponible en: <http://pentaho.almacen-datos.com/>.

[16] **Armstrong, Smith.** *Oracle Discoverer 10g Handbook.* San Francisco, California : s.n., 2006.

[17] CASANOVA, J. *PostgreSQL 8.4 ha sido liberado*, 2009.

[18] **Sherman Wood, JasperSoft. 2007.** pentaho. *Mondrian Documentation.* [En línea] Abril de 2007. [Citado el: 2 de Diciembre de 2010.] Disponible en: <http://mondrian.pentaho.com/documentation/workbench.php/>.

[19] **Esteban, Eloy A. 2010.** Programación en castellano. *Artículo.* [En línea] 2010. [Citado el: 3 de Diciembre de 2010.] Disponible en: <http://www.programacion.com/articulo/tomcat_-_introduccion_134#tomcat1/>.

[20] Lsoto. (2008). *Mitecnologico*. Recuperado el 04 de 05 de 2010, de Mitecnologico:

[21] gulsin [Online] [Citado el: 24 de marzo del 2011.] Disponible en: <http://gulsin.org/2011/03/14/transformaciones-en-pentaho-kettle/>.

[22] **UNIT TESTING EXAMPLE CONCEPTS AND FRAMEWORKL.** [en línea] (2008). [Consulta 1 junio 2011] Disponible en:< http://mauriziostorani.wordpress.com/>

[23] Kristin Daniel**.INTEGRATION TESTING** [en línea] (2000). [Consulta 1 junio 2011] Disponible en: <http://www.sei.cmu.edu/intro/process/technqs/q_it.htm />.

[24] Kristin Daniel**.FUNCTIONAL SYSTEM TESTING** [en línea] (2000). [Consulta 1 junio 2011] Disponible en: <http://www.sei.cmu.edu/intro/process/technqs/q_fst.htm />

[25] PRUEBAS DE SOFTWARE [en línea] (2009). [Consulta 1 junio 2011] Disponible en: <http://www.slideshare.net/aracelij/pruebas-de-software/>.

GLOSARIO DE TÉRMINOS

A continuación se presentan los términos que podrían resultar de difícil comprensión, nuevos al lector o de diversos significados dependiendo del contexto que se analice. Esta sección tiene como objetivo facilitar la comprensión del contenido expuesto en el documento.

BI: Inteligencia de negocio

Código abierto: término con el que se conoce al software distribuido y desarrollado libremente.

Cubo: colección de dimensiones y medidas en un área temática particular.

Datamart: mercado de datos.

Datawarehouse: almacén de datos.

Dimensión: característica de un hecho que permite su análisis posterior en el proceso de toma de decisiones y brinda una perspectiva adicional a un hecho dado.

ETL: Extracción, transformación y carga

Hecho: operación que se realiza en el negocio la cual está estrechamente relacionada con el tiempo y es objeto de análisis para la toma de decisiones.

HOLAP: Procesamiento Analítico en Línea Híbrido.

Lista de chequeo: instrumento de medición y evaluación que consiste básicamente en un formulario de preguntas referentes al atributo de calidad que se está probando y de las características del documento en el caso de la documentación.

MOLAP: Procesamiento Analítico en Línea Multidimensional.

Open source: Código abierto.

PIB: Producto interno bruto

ROLAP: Procesamiento Analítico en Línea Relacional.

Tabla de hecho: contendrá el hecho a través del cual se construirá el indicador de estudio.