

Universidad de las Ciencias Informáticas

Facultad 6



Título: *Sistema de Información de Gobierno. Mercado de Datos Balance económico financiero: extracción, transformación y carga de los datos*

*Trabajo de Diploma para optar por el título de
Ingeniero Informático*

Autoras: *Taimí González Reyes*

Yndira Inciarte Calderón

Tutoras: *Lic. Elena Leonila Fernández García*

Ing. Kenia Riverón Ovalle

La Habana, 2011

DECLARACIÓN DE AUTORÍA

Declaramos ser autoras 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 firmamos la presente a los ____ días del mes de _____ del año _____.

Firma del Autor

Taimí González Reyes

Firma del Autor

Yndira Inciarte Calderón

Firma del Tutor

Lic. Elena Leonila Fernández García

Firma del Tutor

Ing. Kenia Riverón Ovalle

DATOS DEL CONTACTO

Tutoras:

Tutora: Lic. Elena Leonila Fernández García

Especialidad de graduación: Matemática

Categoría docente: Instructor

Categoría Científica: Licenciada

Años de experiencia en el tema: 32

Años de graduado: 32

Correo Electrónico: efernandezg@uci.cu

Tutora: Ing. Kenia Riverón Ovalle

Especialidad de graduación: Ingeniero en Ciencias Informáticas

Categoría Científica: Ingeniera

Años de experiencia en el tema: 2

Años de graduado: 2

Correo Electrónico: kriveron@uci.cu

AGRADECIMIENTOS

Agradezco a mis padres por estar siempre a mi lado apoyándome incondicionalmente en cada momento.

A mis abuelas que no las tengo hoy físicamente pero colaboraron con mi cuidado y formación.

A mis hermanitas Tata y Jessy que las quiero mucho y espero servirles de ejemplo.

Al resto de la familia que me ha apoyado durante todo esta etapa. A Arle por soportarme en estos cinco años.

A mis compañeros y al claustro de profesores que han colaborado con mi preparación en la universidad.

En fin a todas las personas que me han demostrado su amistad y me han ayudado cuando más lo he necesitado, gracias.

Tay

Gracias ante todo a mi reina, mi guía y mi ejemplo, la razón por la cual he llegado hasta aquí, mi mamita.

Gracias a mi abuela por siempre darme su bendición y tener fe en mi.

Gracias a esa persona que se convirtió en mi razón de vivir, mi Cto, te agradezco la comprensión que tuviste en todo momento.

Gracias a mis amigas, Magnolia que estuvo durante estos cinco años y que sabes que este espacio es tuyo. A Milena por convertirte en más que una hermana para mi y a Yordi que no por encontrarse lejos dejó de apoyarme y aconsejarme.

Gracias a mi familia por creer en mi y apoyarme, en especial a Leo.

Gracias a todos los profesores y amistades que me dieron su apoyo cuando lo necesité.

Yndy

DEDICATORIA

Dedicada especialmente a mi mamita, abuela, Leo y familia.

Yndy

A alguien muy importante en mi vida, mi abuela Santa

Tay

RESUMEN

El presente trabajo surge a partir de la necesidad que tiene el país de recopilar los datos de cada una de las Unidades Presupuestadas(UP), para ello la Oficina Nacional de Estadísticas (ONE) se apoya en la red de Estadística distribuida a lo largo del territorio nacional con el fin de agilizar el proceso de toma de decisiones y la velocidad de respuesta del sistema a las consultas que se realicen sobre la información. Se tiene como tarea la creación de un Mercado de Datos (MD) que recoja los indicadores del Balance Económico Financiero (BEF) del país. El análisis, diseño e implementación de los procesos de integración de dicho MD constituye la esencia de este trabajo.

Abarca la metodología y herramientas que se emplean en el desarrollo de este tipo de soluciones. Como resultados se obtienen las estructuras dimensionales para el BEF que contienen las tablas de hechos y medidas asociadas, las dimensiones con niveles de jerarquías, que garantizan el análisis estadísticos.

Se realiza el proceso de extracción de los datos de las fuentes del Sistema de Integración de Gestión Estadística (SIGE), luego se realizan un conjunto de transformaciones para posteriormente cargar estos datos en el MD presentado. En la investigación se incluyen las estrategias de seguridad, respaldo y recuperación de datos y también se efectúan pruebas que validan la solución en cuestión.

PALABRAS CLAVES

BEF: Balance Económico Financiero.

MD: Mercado de Datos.

ONE: Oficina Nacional de Estadísticas.

SIGE: Sistema de Integración de Gestión Estadística.

UP: Unidad Presupuestada.

ÍNDICE DE CONTENIDO

INTRODUCCIÓN -----	1
CAPÍTULO 1 -----	4
MERCADO DE DATOS BALANCE ECONÓMICO FINANCIERO. EXTRACCIÓN, TRANSFORMACIÓN Y CARGA -----	4
Introducción-----	4
1.1 Almacén de datos-----	4
1.2 Actualidad de los almacenes de datos-----	7
1.3 Técnicas de almacenamiento de la información-----	9
1.4 Mercado de datos-----	10
1.5 Tendencias y tecnologías actuales-----	12
Conclusiones parciales-----	18
CAPÍTULO 2 -----	19
ANÁLISIS Y DISEÑO DE LOS PROCESOS DE INTEGRACIÓN DEL MERCADO DE DATOS BALANCE ECONÓMICO FINANCIERO. -----	19
Introducción-----	19
2.1 Necesidades del usuario-----	19
2.2 Definición del negocio-----	20
2.3 Reglas del negocio-----	21
2.4 Especificación de requerimientos-----	23
2.5 Modelo de casos de uso del sistema-----	26
2.6 Especificación del modelo dimensional-----	29
2.7 Seguridad-----	37
Conclusiones Parciales-----	38
CAPÍTULO 3 -----	39
IMPLEMENTACIÓN Y PRUEBA DE LOS PROCESOS DE INTEGRACIÓN DEL MERCADO DE DATOS BALANCE ECONÓMICO FINANCIERO. -----	39
Introducción-----	39
3.1 Modelo de datos físico-----	39
3.2 Perfilado de datos-----	40
3.3 Implementación de los procesos de extracción, transformación y carga-----	42
3.4 Estrategia de validación y prueba-----	45

Conclusiones Parciales	54
CONCLUSIONES	55
RECOMENDACIONES	56
REFERENCIAS BIBLIOGRÁFICAS	57
BIBLIOGRAFÍA	59
GLOSARIO DE TÉRMINOS	60
ANEXOS	61

ÍNDICE DE ILUSTRACIONES

Figura 1 Diagrama de casos de usos del sistema-----	26
Figura 2 Diseño lógico del mercado de datos -----	31
Figura 3 Carga de la tabla hech_estado_gastos-----	43
Figura 4 Trabajo para la carga de las tablas del esquema dimensiones-----	44
Figura 5 Trabajo para la carga de las tablas del esquema mart_balance -----	44
Figura 6 Ejecución del trabajo en Spoon -----	47
Figura 7 Posicionar al directorio de PDI -----	47
Figura 8 Ejecución del trabajo en Kitchen-----	48
Figura 9 Resultados de la ejecución del trabajo en Spoon-----	48
Figura 10 Resultados de la ejecución del trabajo en Kitchen-----	49
Figura 11 Archivos de errores-----	49
Figura 12 Reporte: Por División Política Administrativa y entidad los indicadores del estado de gastos ----	53
Figura 13 Carga de la tabla dim_dpa -----	61
Figura 14 Carga de la tabla dim_entidad -----	61
Figura 15 Carga de la tabla dim_indicadores_generales -----	61
Figura 16 Carga de la tabla dim_nae -----	61
Figura 17 Carga de la tabla dim_organismo -----	62
Figura 18 Carga de la tabla dim_subordinación -----	62
Figura 19 Carga de la tabla dim_temporal_mes -----	62
Figura 20 Carga de la tabla dim_forma_gasto -----	62
Figura 21 Carga de la tabla dim_indicador_balance -----	63
Figura 22 Carga de la tabla dim_nae_gastos -----	63

Figura 23 Carga de la tabla hech_estado_situacion -----	63
Figura 24 Carga de la tabla hech_estado_resultado-----	64
Figura 25 Carga de la tabla hech_inversiones_donaciones-----	64
Figura 26 Carga de la tabla hech_disponibilidad_fondos-----	65
Figura 27 Carga de la tabla hech_indicadores_generales -----	65
Figura 28 Carga de la tabla hech_indicadores_seleccionados_contabilidad -----	66
Figura 29 Carga de la tabla hech_cumplimiento_plan_economico-----	67

ÍNDICE DE TABLAS

Tabla 1 Tabla de roles y permisos -----	26
Tabla 2 Extraer información balance económico financiero-----	28
Tabla 3 Realizar transformación y carga balance económico financiero-----	29
Tabla 4 Matriz BUS -----	30
Tabla 5 Tabla de hecho Estado de resultado-----	33
Tabla 6 Tabla de hecho Indicadores generales -----	34
Tabla 7 Tabla de hecho Indicadores seleccionados de la contabilidad-----	34
Tabla 8 Tabla de hecho Estado de situación -----	35
Tabla 9 Tabla de hecho Cumplimiento del plan económico-----	35
Tabla 10 Tabla de hecho Disponibilidad de fondos -----	36
Tabla 11 Tabla de hecho Estado gastos-----	36
Tabla 12 Tabla de Inversiones y donaciones -----	37
Tabla 13 Tablas del modelo datos -----	40
Tabla 14 Perfilado de los estándares de medidas del modelo 5910 -----	41
Tabla 15 Perfilado del análisis de cadenas del modelo 5910-----	41
Tabla 16 Perfilado análisis numérico del modelo 5910-----	42
Tabla 17 Funciones a probar del caso 1-----	46
Tabla 18 Principales opciones de línea de comandos-----	47
Tabla 19 Funciones a probar del caso 2-----	50
Tabla 20 Cantidad de registros en BD fuente y BD destino-----	51
Tabla 21 Cantidad de registros por tabla en BD destino duplicados. -----	52
Tabla 22 Muestra de datos de la tabla hech_estado_gastos -----	53

INTRODUCCIÓN

INTRODUCCIÓN

Con el desarrollo de las tecnologías la vida de los habitantes del planeta se hace cada día más amena. El avance tecnológico, el crecimiento acelerado de programas y aplicaciones en el mercado internacional provocan competencia entre compañías dedicadas a la creación de las tecnologías. El uso de estos medios tecnológicos en varios sectores contribuye a mejorar la labor de los trabajadores en cada una de las empresas y favorecen el incremento económico de cada país. En estas entidades se generan datos que ayudan a la toma de decisiones de una nación, por lo que se hace necesario conservarlos.

De estas premisas surgieron las bases de datos (BD) convirtiéndose en una herramienta primordial para guardar información. Las BD al tener almacenado un cúmulo gigantesco de datos, provocaban una demora en la entrega de información a los clientes y en el análisis de los resultados económicos. Por lo que varios especialistas al surgir esta problemática se dieron a la tarea de buscar otra alternativa que facilitara la recopilación de la información sin tener que afectar la toma de decisiones en un momento determinado.

Cada país estima el desarrollo económico alcanzado mediante el análisis de la información que se obtiene a través de los datos estadísticos. Cuba a pesar de ser un país subdesarrollado no se encuentra ajena a este evento y aprovechando al máximo los beneficios que traen consigo las Tecnologías de la Información (TIC) es capaz de recoger los datos que se generan en cada una de las entidades a lo largo de todo el país.

En Cuba la ONE tiene entre sus funciones la de integrar el trabajo estadístico estatal del país en un sistema eficiente, destinado a brindar la información requerida para el análisis del comportamiento de la economía, lo cual realiza a través de la captación, el análisis y la difusión de los datos recogidos. La ONE capta estos datos en una serie de modelos que recogen información de todos los sectores de la economía y la sociedad directamente desde las entidades que producen la información. Estos provienen de distintas fuentes, reuniéndose en varios momentos y formatos, por lo que se hace muy difícil la consulta sobre los mismos, aumentando el costo del esfuerzo y el tiempo en el proceso de recuperación y elaboración de informes lo que ocasiona el no poder obtener cualquier conclusión sobre el tema deseado en el tiempo que se precisa.

Con el fin de solucionar este problema el país realizará por primera vez la captación de los Estados Financieros de todas las UP. Se necesita almacenar dicha información de tal manera que se pueda acceder a la misma tanto en el formato que poseen los modelos de la contabilidad, que son los

INTRODUCCIÓN

indicadores estadísticos, como en el formato de los estados financieros y viabilizar los trabajos de integración y análisis de los datos almacenados en distintas BD para que los Órganos del Estado puedan disponer de esta información en el proceso de toma de decisiones. Ante la presente situación surge como **problema de la investigación**:

¿Cómo lograr la integración de los datos en el área Balance Económico Financiero del Sistema de Información del Gobierno?

Teniendo como **objeto de estudio** la integración de los datos y específicamente como **campo de acción** la extracción, transformación y carga de datos para el área Balance Económico Financiero del Sistema de Información del Gobierno.

Para dar solución al problema de la investigación se ha planteado como **objetivo general**:

Implementar los procesos extracción, transformación y carga de los datos del área Balance Económico Financiero del Sistema de Información del Gobierno para contribuir a la calidad de los datos.

Para lograr el objetivo general anteriormente señalado se trazan los siguientes **objetivos específicos**:

1. Realizar el análisis y diseño del mercado de datos del área Balance Económico Financiero.
2. Implementar el mercado de datos (Extracción, Transformación y Carga de los datos) del área Balance Económico Financiero.
3. Validar el mercado de datos (Extracción, Transformación y Carga de los datos) del área Balance Económico Financiero.

Se planificaron una serie de **tareas** para dar cumplimiento a los objetivos específicos:

1. Caracterización de las metodologías, herramientas y tecnologías a utilizar en el desarrollo de almacenes de datos.
2. Levantamiento de requisitos.
3. Descripción de los casos de uso del mercado de datos.
4. Definición de los hechos, las medidas y las dimensiones del mercado de datos.
5. Diseño del modelo de datos.
6. Definición de la arquitectura del mercado de datos.
7. Diseño del subsistema de integración.

INTRODUCCIÓN

8. Diseño de los casos de pruebas.
9. Implementación del subsistema de integración.
10. Aplicación de las listas de chequeo.
11. Aplicación de los casos de pruebas.

El trabajo a presentar cuenta con tres capítulos:

Capítulo 1- Mercado de Datos Balance económico financiero. Extracción, transformación y carga: Hace referencia al estado del arte del tema tratado, se describen además el objeto de estudio, el campo de acción, así como las herramientas y la metodología con las que se llevará a cabo la propuesta.

Capítulo 2- Análisis y diseño de los procesos de integración del Mercado de Datos Balance económico financiero: En este capítulo se define el análisis del MD, se identifican los requisitos funcionales y no funcionales, se explican las necesidades de los usuarios, se describen los casos de uso (CU), se especifican los actores con su descripción. Se definen las dimensiones del almacén de datos y sus hechos, así como la matriz bus que los relaciona.

Capítulo 3- Implementación y prueba de los procesos de integración del Mercado de Datos Balance económico financiero: Se implementan los procesos de extracción, transformación y carga de los datos, detallándose los procesos de limpieza de los mismos. Se realizan pruebas para validar que la información fue cargada correctamente a la BD destino, comprobándose además la calidad de los datos.

CAPÍTULO 1

MERCADO DE DATOS BALANCE ECONÓMICO FINANCIERO. EXTRACCIÓN, TRANFORMACIÓN Y CARGA.

Introducción

En este capítulo se detallarán diferentes elementos teóricos y metodológicos sobre los que se fundamenta el MD, incluye un estado del arte a nivel internacional y nacional sobre el desarrollo alcanzado de los MD, se exponen las tendencias, tecnologías y software usados en la actualidad para la creación de los MD. Recoge además un estudio de los aspectos a tener en cuenta para un mejor entendimiento del campo de acción.

1.1 Almacén de Datos

Según Ralph Kimball¹ un AD “es una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis”[1]. Además Willian H. Inmon² lo define como “una colección de datos orientado a un determinado ámbito, que ayuda a la toma de decisiones en la entidad en la que se utiliza. Se trata, sobre todo, de un expediente completo de una organización, más allá de la información transaccional y operacional, almacenada en una BD diseñada para favorecer el análisis y la divulgación eficiente de datos” [2].

Características de los almacenes de datos

En términos de las características del repositorio de datos:

- **Orientado a temas** - Los datos en la BD están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí.
- **Variante 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 sólo lectura, y se mantiene para futuras consultas.

¹ Ralph Kimball: Considerado uno de los autores del tema de almacenamiento de datos.

² Willian H. Inmon: Reconocido como el padre de los almacenes de datos.

- **Integrado** - La BD contiene los datos de todos los sistemas operacionales de la organización, y dichos datos deben ser consistentes [2].

Ventajas y desventajas de los almacenes de datos

Los AD proporcionan un menor coste en la toma de decisiones, un mejor servicio al cliente y permite el rediseño de los procesos, además aporta facilidad en el manejo de la información por lo que su implantación puede beneficiar a una empresa en:

- Ventaja competitiva debido a una buena toma de decisiones.
- Los AD integran información proveniente de múltiples sistemas externos, permitiendo consolidar los datos y poder tener acceso a la información de todas las fuentes en una sola.
- Obtienen una BD clasificada por temas e histórica y se hace más fácil el acceso a una gran diversidad de datos.
- Permite el manejo de mayores volúmenes de información en archivos más pequeños, debido a que los AD mediante la agregación, reducen significativamente el tamaño de los archivos.
- El tiempo de respuesta en la generación de reportes y consultas se reduce considerablemente.

A la vez resulta muy difícil el mantenimiento de los mismos por lo que utilizarlos podría llevar a algunos problemas:

- Su construcción puede requerir de mucho tiempo.
- La integración de las herramientas es muy compleja.
- La subestimación del tiempo requerido para extraer, limpiar y cargar los datos puede ser larga.

Arquitectura de los almacenes de datos

Modelo multidimensional

Un AD representa los datos que contiene usando modelos multidimensionales lo cual facilita su análisis. Este tipo de modelamiento es una técnica para modelar BD 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 [3].

La estructura básica para este tipo de modelo está definida por esquemas y tablas.

El modelo multidimensional de datos puede estar fomentado por un esquema relacional, almacenando datos en tablas relacionales especializadas, llamadas tablas de hechos y de dimensiones. Este tipo de esquema provee una vista multidimensional de los datos usando un modelo relacional como soporte.

Implementar un AD a través del modelo multidimensional tiene un número importante de ventajas [4]:

- Un modelo multidimensional es una estructura homogénea y predecible. Reportes escritos, herramientas de consulta e interfaces de usuario, todas pueden hacer suposiciones acerca del modelo multidimensional para que las interfaces del usuario sean más entendibles y los procesos, más eficientes.
- Es una estructura predecible. Cada dimensión es equivalente. Todas las dimensiones pueden ser vistas como un conjunto de puntos igualmente simétricos dentro de una tabla de hechos.
- Es extensible para acomodar nuevos elementos de datos inesperados y nuevas decisiones de diseño. Esto se lleva a cabo añadiendo nuevos hechos de forma inesperada, añadiendo nuevas dimensiones, añadiendo nuevos atributos dimensionales y cambiando datos de una cierta granularidad para pasarlos a otra.
- Una última característica del modelo multidimensional es la creciente cantidad de herramientas administrativas y de software que manejan y usan la granularidad.

Esquemas de los almacenes de datos

Un AD adopta diferentes esquemas para mejorar el rendimiento de la consultas, su diseño es diferente al diseño de una BD operacional ya que en un AD los datos están desnormalizados para proporcionar acceso inmediato a los datos sin tener que realizar una gran cantidad de ensambles, mientras que en una BD los datos están altamente normalizados para soportar frecuentes actualizaciones y para mantener la integridad referencial. A continuación se describen un conjunto de esquemas arquitectónicos:

- **Estrella:** El esquema de la estrella es la arquitectura de AD más simple. En este diseño del AD la tabla de hechos está rodeada por dimensiones y juntos forman una estructura que permite implementar mecanismos básicos para poder utilizarla con una herramienta de consultas OLAP. Implementa un diseño lógico relacional de BD que resulta en que las tablas de hechos representan la Tercera Forma Normal (3FN) y las dimensiones representan la Segunda Forma Normal (2FN).

- **Copos de nieve:** El esquema en copo de nieve es una variedad más compleja del esquema estrella. El afinamiento está orientado a facilitar mantenimiento de dimensiones. Lo que distingue a la arquitectura en copo de nieve del esquema estrella, es que las tablas de dimensiones en este modelo representan relaciones normalizadas (3NF) y forman parte de un modelo relacional de BD.
- **Constelación de hechos:** Este esquema es más complejo que las otras arquitecturas debido al factor de que contiene múltiples tablas de hechos. Con esta solución las tablas de dimensiones pueden estar compartidas entre más que una tabla de los hechos. El esquema de constelación de hechos tiene mucha flexibilidad y este factor es su gran virtud. Sin embargo, el problema es que cuando el número de las tablas vinculadas aumenta, la arquitectura puede llegar a ser muy compleja y difícil para mantener [5].

Tablas

- **Hechos:** Los hechos son operaciones que se realizan en el negocio en un tiempo determinado. Son las variables sobre las que se van a realizar operaciones de agregación que conduzcan a conclusiones sobre la evolución del área o departamento que se estudie. Son objetos de análisis para la toma de decisiones [6].
- **Dimensión:** La dimensión es una característica de un hecho que permite su análisis posterior, o sea, el eje de análisis por el que se desea categorizar la información [6].

Granularidad

Las tablas de hechos y dimensiones contienen datos de interés, que presentan un nivel de granularidad. La granularidad es el nivel más bajo de información que será almacenado en estas tablas. El primer paso al diseñar una tabla de dimensión es determinar la granularidad.

Jerarquía

Se define como el orden determinado dentro de los campos de una dimensión. Estas jerarquías son utilizadas a la hora de agregar o desagregar la información [6].

1.2 Actualidad de los almacenes de datos

La información estadística, debido a la importancia que tiene para los estudios de tendencias, toma de decisiones o análisis de datos, se convierte en una aplicación perfecta para la tecnología de AD. En estos

*CAPÍTULO 1 MERCADO DE DATOS BALANCE ECONÓMICO FINANCIERO. EXTRACCIÓN.
TRANSFORMACIÓN Y CARGA.*

momentos a nivel mundial existen varios sistemas de AD estadísticos. A continuación se exponen ejemplos de ello:

- Almacén Central de Datos del Sistema Estadístico Nacional de República Dominicana: Es una herramienta electrónica que tiene como objetivo principal disponer de estadísticas precisas y actualizadas para mejorar los servicios al alcance de los tomadores de decisiones, los investigadores y la sociedad en general [10].
- Almacén de Datos para el Análisis y Difusión de la Información Estadística del Turismo en España (DATATUR) : Es un AD que se encarga del análisis y difusión de la información estadística del turismo en España. Se ha desarrollado para facilitar y agilizar el almacenamiento, recuperación, análisis y difusión de la información en esta área [11].
- AD para la prestación del servicio público de información estadística Instituto Nacional de Estadística, Geografía e Informática (INEGI) en México: Este proyecto consiste en aplicar las tecnologías de BD y AD en el desarrollo de un almacén integrado de datos definitivos con información estadística obtenida de los programas de censos nacionales, encuestas y registros administrativos para la elaboración de productos, la toma de decisiones y la planeación [12].

En el país se ha estado trabajando mucho para lograr almacenar información en múltiples sectores. Algunas organizaciones han dado los primeros pasos en la aplicación de estas nuevas tecnologías, tal es el caso de:

- AD para la Gestión Contable de la Empresa de Proyectos de Arquitectura e Ingeniería EMPAI: Permite disponer de la capacidad y la función de reunir y analizar datos para, de modo sistemático y organizado, monitorear, obtener y difundir información relevante sobre el ambiente externo y las condiciones internas de la organización ofreciendo una visión global de los aspectos económicos, financieros, históricos, regulatorios, políticos, sociales y tecnológicos [13].
- También en concursos de computación se han presentado almacenes para CUBASEL (Empresa que brinda servicio de telefonía celular), sobre la plataforma Oracle con grandes avances desde su implantación.
- Grandes avances ha dado la Universidad de Ciencias Informáticas (UCI) en estos temas, está desarrollando un AD para la ONE el cual se encuentra en construcción.

1.3 Técnicas de almacenamiento de la información

El análisis de un AD se refiere a la forma en que el usuario consultará la información en el sistema y a la parte a la que éste tendrá acceso. Mediante una consulta se selecciona la información a analizar, dicho análisis depende de las necesidades de los usuarios y del tipo de decisiones que quieran que soporte el AD construido.

Las herramientas OLAP (On-Line Analytical Processing por sus siglas en inglés) permiten construir sistemas de ayuda a la toma de decisiones permitiendo realizar sofisticados análisis multidimensionales en tiempo real lo que posibilita entrelazar información de diferentes fuentes de datos.

Las estructuras de datos sobre las que se basan las herramientas OLAP pueden dividirse principalmente en tres categorías [7]:

- **MOLAP** (Multidimensional OLAP): basada en cubos de varias dimensiones para ubicar los datos, refleja con mayor exactitud la realidad modelada y es más sencilla de comprender por el usuario final que los modelos relacionales tradicionales o los ficheros planos de datos.
- **ROLAP** (Relacional OLAP), preserva el enfoque relacional de las BD y tiene el propósito de mantener el mismo grado de eficiencia que las herramientas MOLAP.
- **HOLAP** (Híbrido OLAP), híbrido entre MOLAP y ROLAP que trata de rescatar y de fusionar las virtudes de ambos modelos. Los datos primitivos que están almacenados en una BD relacional y que son accedidos por métodos ROLAP conviven con los datos agregados a diferentes niveles almacenados en una BD multidimensional, utilizando múltiples cubos en dependencia del fabricante.

Para la construcción de la solución ROLAP ofrece mejores ventajas:

Puede manejar grandes cantidades de datos: La limitación de tamaño de los datos de la tecnología ROLAP es la limitación en el tamaño de los datos de la BD relacional subyacente. En otras palabras, ROLAP se pone ninguna limitación en la cantidad de datos.

Puede aprovechar las funciones inherentes a la BD relacional: A menudo, la BD relacional ya viene con una serie de funcionalidades [8].

1.4 Mercado de Datos

Los AD contienen información que se subdivide a veces en unidades lógicas más pequeñas llamadas MD. Un MD es un subconjunto de un AD. Los MD tienen mayor utilidad para grupos más pequeños que deseen acceder a datos detallados. Se usa un almacén para resúmenes de datos que puedan ser usados para el resto de la compañía. Aunque se puede usar cualquier software de BD para establecer un AD, algunos proveedores proporcionan software de BD para establecer un AD [1].

Características de los mercados de datos

- Según las necesidades de los usuarios el diseño del MD se realiza siguiendo una estructura consistente.
- La información histórica que posee es mínima.
- Contiene el grado de granularidad necesaria.
- Da costes adicionales en hardware, software y accesos de red.
- Debido a que hay grupos de usuarios que solo acceden a un subconjunto preciso de datos, se hace más fácil el acceso a las herramientas de consulta y divide los datos para controlar mejores accesos [1].

Etapas en el desarrollo del mercado de datos

Análisis

El análisis es un proceso de exploración que permite obtener los elementos a tener en cuenta en la construcción del MD, donde a través de los requerimientos definidos se logra entender de una forma más fácil lo que el usuario necesita.

Con el desarrollo del análisis del MD se obtiene una visión más general de lo que se quiere. En el análisis se determinan los requisitos de información, funcionales y no funcionales definiendo así el alcance del MD, se definen además las reglas del negocio y se identifican las fuentes de datos operacionales y externas del negocio.

Diseño

En el diseño se define el cómo hacer la aplicación. Durante la fase de análisis se identifican requerimientos, el diseño responde a la forma en la que el sistema cumplirá estos requisitos. Se define el modelo de datos, el esquema arquitectónico, los hechos y las dimensiones.

Extracción, transformación y carga

ETL (Extract, Transform and Load por sus siglas en inglés) es un proceso de tres etapas en el uso de BD y almacenamiento de datos. Permite la integración y el análisis de los datos almacenados en distintas BD y formatos heterogéneos. Vale destacar que el éxito del AD está dado por el amplio estudio del negocio partiendo del conocimiento obtenido de las fuentes de datos externas y de los responsables de su mantenimiento, no del simple copiado de la información [9].

ETL es un proceso que implica las siguientes tareas:

- **Extraer datos** de sistemas operacionales o de archivos de fuentes que son la fuente principal para el almacenamiento de datos.
- **Transformación de los datos** que puede implicar la limpieza, filtrado, validación y aplicación de reglas de negocio.
- **Cargar los datos** en un AD o cualquier otra BD o aplicación que aloja los datos garantizando la representación de las tablas de dimensiones y de hechos.

Los términos estrechamente relacionados y gestionados por los procesos ETL son: la migración de datos, gestión de datos, limpieza de datos, sincronización de datos y consolidación de datos [5].

Fuentes de datos de origen

Los datos archivados presentan pocos problemas, estos en general no se cargan pues no son provechosos. Los datos del entorno operacional se deben analizar según los diferentes entornos, ya que se deben identificar los atributos de interés, unificar el significado de los estados, identificar la mejor fuente de estado para cada registro y validar que los datos sean correctos en origen; además se deben definir las tareas de procesamiento y transformación de los datos, para poder definir el proceso de carga inicial y de sincronización posteriormente y en ocasiones se deben realizar modificaciones en el origen para facilitar la tarea de sincronización.

Integración de los datos

Al integrarse múltiples BD aparece la redundancia, los atributos u objetos pueden mostrarse con diferentes nombres en cada BD hasta el dominio puede cambiar y puede hasta aparecer como dato derivado en otra tabla. Sirviendo la integración para evitar las redundancias y las inconsistencias de los datos dentro de la BD.

Limpieza de Datos

La limpieza de datos es uno de los principales problemas en el almacenamiento de datos. Identificar y corregir valores atípicos y errores, corregir datos inconsistentes y rellenar valores perdidos son unas de las tareas que se deben realizar en este proceso.

Inteligencia de Negocio

La Inteligencia de Negocio (BI por sus siglas en inglés) es una actividad, proceso o herramienta que se utiliza para obtener una mejor información que puede apoyar el proceso de toma de decisiones, o sea, buscar en los datos de la empresa y seleccionar la información útil que pueda proporcionar a los usuarios información valiosa. Si es bien utilizada trae beneficios a la empresa y mejora su rendimiento ya que hace que la recolección y difusión de información sea más rápido y eficiente.

1.5 Tendencias y Tecnologías Actuales

Metodologías

“La metodología es la filosofía que sustenta todo el proceso investigativo y a los métodos en particular” [14]. Es un conjunto de pasos y procedimientos que deben seguirse que favorecen a los desarrolladores a dividir las etapas de un proyecto, qué tareas se llevan a cabo en cada etapa, qué restricciones deben aplicarse, qué técnicas y herramientas se emplean y cómo se controla y gestiona un proyecto.

Las tendencias actuales de las metodologías para el desarrollo de los AD van enfocadas hacia la metodología que propone Ralph Kimball, el cual propone un modelo ascendente. Otra tendencia es la propuesta por Willian H. Inmon proponiendo de manera contraria a Kimball una metodología de formas descendente.

Para el desarrollo del MD se utilizará una adaptación de la metodología de Ralph Kimball ya que crea los conceptos de hechos y dimensiones, lo que indudablemente es muy eficaz en el proceso de la

toma de decisiones y proporciona mayor agilidad en el proceso de desarrollo. Propone ir construyendo el AD a través de la construcción de los MD departamentales, lo que constituye una estrategia buena y coincide con la división lógica de las empresas, entidades u organismos. Es reconocida por el resto de la comunidad que se dedica al tema. Tiene bien definidas las etapas, actividades, artefactos y roles.

Como complemento a la misma y fortaleciendo la etapa del levantamiento de requisitos; se toma lo planteado por Leopoldo Zenaido Zepeda Sánchez en su tesis de doctorado, orientando así el trabajo a los casos de uso y se logra estar más alineado con las tendencias y normas de la universidad. Esta metodología consiste en tres fases: La primera fase, está dedicada a examinar el esquema entidad relación (ER) de la DB operacional, generando los esquemas multidimensionales candidatos para el AD. En la segunda fase, los requisitos de usuario son recogidos por medio de entrevistas. El propósito de las entrevistas es obtener información acerca de las necesidades de análisis de los usuarios. La tercera fase, contrasta la información obtenida en la segunda fase, con los esquemas multidimensional candidatos formados en la primera fase generando así, la mejor solución que reflejan los requisitos de usuario [15].

Herramientas para el desarrollo del MD

Control de versiones

Subversion 1.5.4

Subversión (SVN) es un sistema de control de versiones que permite la manipulación de datos y ficheros en un repositorio central a través del tiempo. Puede ser usado en varios ordenadores ya que se puede acceder al mismo a través de la red permitiendo recuperar versiones antiguas de ficheros o examinar el historial de cambios.

Características principales de SVN:

- Versionado de directorios: ficheros y directorios se encuentran bajo control de versiones.
- Verdadero historial de versiones: se puede añadir, borrar, copiar, y renombrar ficheros y directorios donde cada fichero que se añade comienza con un nuevo historial.
- Manipulación consistente de datos: expresa las diferencias del fichero usando un algoritmo de diferenciación binario y son transmitidas a través de la red en ambas direcciones.

Gestor de Base de Datos

Un Sistema Gestor de Bases de Datos (SGBD) es un software que proporciona servicios para la creación, el almacenamiento, el procesamiento y la consulta de la información almacenada en BD de forma segura y eficiente. Un SGBD actúa como un intermediario entre las aplicaciones y los datos [16]. El objetivo principal es proporcionar un entorno eficiente a la hora de almacenar y recuperar la información de las BD. Este software facilita el proceso de definir, construir y manipular BD para diversas aplicaciones.

PostgreSQL 8.4

Se selecciona como SGBD PostgreSQL ya que se considera como la mejor opción de código abierto para BD, es un sistema de BD utilizado por miles de organizaciones alrededor del mundo, puede ser ejecutado sobre las plataformas más usadas en todo el mundo: Windows, Linux y GNU e implementan una gran cantidad de interfaces para propiciar el acceso a datos, de aplicaciones desarrolladas sobre una larga lista de lenguajes de programación.

PostgreSQL es un sistema de gestión de BD objeto-relacional basado en Postgres versión 4.21. Soporta una gran parte del estándar SQL y ofrece muchas características modernas: consultas complejas: claves externas, desencadenantes, puntos de vista y la integridad transaccional.

Además, PostgreSQL puede ser ampliado por el usuario en muchos aspectos, por ejemplo mediante la adición de nuevos tipos de datos, las funciones, los operadores de funciones de agregado, los métodos de índice, lenguas de procedimiento [17].

La versión 8.4 continúa con el rápido desarrollo de las BD de código abierto más avanzada. Esta versión contiene una gran cantidad de mejoras para la administración, consulta y programación de BD PostgreSQL.

Muchos de los cambios en PostgreSQL 8.4 son herramientas de administraciones nuevas o mejoradas y el control y comandos. También hay mejoras en el rendimiento incluidos en esta versión.

Ventajas que ofrece PostgreSQL 8.4 [18]:

- Instalación ilimitada.
- Estabilidad y confiabilidad legendarias.
- Ahorros considerables en costos de operación.

- Extensible.
- Multiplataforma.
- Diseñado para ambientes de alto volumen.

pgAdmin III 1.10

pgAdmin es una herramienta de código abierto para la administración de BD PostgreSQL y sus derivados. Está diseñado para responder a las necesidades de todos los usuarios, desde escribir simples consultas SQL para el desarrollo de BD complejas. Posee una interfaz gráfica compatible con todas las características de PostgreSQL y facilita la administración. La aplicación también incluye un editor de resaltado de sintaxis SQL y un editor de código del lado del servidor. La conexión al servidor se puede hacer a través de TCP / IP y puede ser encriptado con SSL (Secure Sockets Layer por sus siglas en inglés) para la seguridad. No se requieren manejadores adicionales para comunicarse con el servidor de BD. Es un software libre publicado bajo la licencia de PostgreSQL [19].

Herramienta de modelado

Se puede definir a las Herramientas CASE (Computer Aided Software Engineering por sus siglas en inglés) como un conjunto de programas y ayudas que dan asistencia a los analistas, ingenieros de software y desarrolladores, durante todos los pasos del ciclo de vida de desarrollo de un software [20]. La utilización de herramientas de modelado permite a los arquitectos de datos y administradores de BD junto con los desarrolladores gestionar y mantener aplicaciones que trabajan con un volumen grande de datos.

Visual Paradigm 3.4

Como herramienta de modelado se seleccionó Visual Paradigm (VP por sus siglas en inglés), por ser un modelador UML (Unified Modeling Language por sus siglas en inglés) que permite el diseño de sistemas con todo tipo de tipos de diagramas UML. También es compatible con una explotación extensiva en caso de uso, diagrama de SysML (Systems Modeling Language por sus siglas en inglés) requerimiento y diseño de BD. Con VP-UML, el equipo de desarrollo de software puede realizar el análisis y diseño de sistemas con eficacia [21].

Características:

- Como característica principal se le confiere la de ser multiplataforma ya que tiene la capacidad de ejecutarse sobre diferentes sistemas.

- Fácil de usar y presenta un ambiente gráfico agradable para el usuario. Ofrece soluciones de software que permiten a las organizaciones desarrollar las aplicaciones de calidad más rápido, bien y más barato ya que usa como lenguaje de modelado UML.
- Permite la integración con herramientas de desarrollo, la generación de código base para diferentes lenguajes de programación, configurar las líneas de redacción, el modelado de BD, el modelado de requerimientos, el modelado del proceso de negocio, la interoperabilidad, la generación de documentación además su notación es muy parecida a la estándar.

Herramientas para el trabajo en ETL

Pentaho Data Integration 4.0

Para la realización de los procesos de ETL se seleccionó Pentaho Data Integration (PDI) 4.0, esta herramienta pone en práctica la metodología de ágil desarrollo y las buenas prácticas. La misma posibilita la integración de datos, dentro del cual hay un grupo de herramientas para llevar a cabo el proceso de ETL. Se puede realizar al mismo tiempo diseño de las ETL, modelamiento y visualización de datos, que a su vez ayuda a reducir costos, mejorar la productividad y acorta el tiempo necesario para obtener resultados concretos [22].

Programas que forman la herramienta:

- **Spoon:** Herramienta gráfica que posibilita el diseño de las transformaciones y trabajos. Posee opciones para previsualizar y testear los elementos desarrollados. Es la principal herramienta de trabajo de PDI y con la que se construirán y validarán los procesos ETL.
- **Pan:** Es un motor de transformación de datos que realiza muchas funciones tales como lectura, manipulación, y escritura de datos hacia y desde varias fuentes de datos [23]. Herramienta que permite la ejecución de las transformaciones diseñadas en spoon (bien desde un fichero o desde el repositorio). Permite desde la línea de comandos preparar la ejecución mediante scripts.
- **Kitchen:** Similar a Pan, pero para ejecutar los trabajos diseñados por Spoon en xml o en un catálogo de BD [23].
- **Carte:** es un pequeño servidor web que permite la ejecución remota de transformaciones y trabajos.

Cuando se trabaja con Spoon existen dos vías para guardar lo que se ha diseñado:

- **Repositorio:** se dispone de una BD, con una estructura especial, donde son guardadas las transformaciones y trabajos construidos.
- **Ficheros:** las transformaciones y trabajos son guardados a nivel del sistema de ficheros, en archivos xml (con extensión .ktr para las transformaciones y .kjb para los trabajos).

DataCleaner 1.5.3

Es una aplicación de código abierto para el perfil, la validación y comparación de datos. Permite la evaluación del nivel de calidad de los datos contenidos en el sistema de información [24]. Es una aplicación muy fácil de usar, genera sofisticados informes y gráficos que permiten a los usuarios determinar de un vistazo el nivel de calidad de los datos, identificar y analizar la estructura del origen de datos y combinar resultados y gráficos, creando vistas fáciles de interpretar para evaluar la calidad de los datos.

Las características incluyen:

- Los perfiles de datos se utilizan para calcular y analizar diversas medidas importantes basadas en los valores de los datos.
- Validación de los datos: el validador le dará un resultado que puede ser interpretado como bueno o malo, ya que el validador valida los datos.
- Los datos de comparación.
- Diccionario de gestión.
- Análisis del modelo.
- Soporta acceso de lectura a muchos tipos de AD.
- DataCleaner 1.5.3 puede acceder y analizar prácticamente cualquier AD, incluyendo: Bases de datos como Oracle, Microsoft SQL Server, PostgreSQL, MySQL, OpenOffice (ODB) y más.
- Archivos separados por comas y separados por tabuladores (.csv / .tsv).
- Hojas de cálculo Excel (.xls).
- Archivos xml.

Conclusiones Parciales

Actualmente los AD tienen un impacto positivo en los procesos empresariales, hecho que quedó evidenciado en el transcurso del estudio del arte realizado en este capítulo.

En el desarrollo del MD se utilizarán las siguientes herramientas y metodología:

- Para el modelado Visual Paradigm 3.4.
- PostgreSQL 8.4 como gestor de BD.
- Para el proceso de ETL, Pentaho Data Integration 4.0 y Data Cleaner 1.5.3.
- Metodología adaptada de Kimball propuesta por el grupo de AD (DATEC).

CAPÍTULO 2

ANÁLISIS Y DISEÑO DE LOS PROCESOS DE INTEGRACIÓN DEL MERCADO DE DATOS BALANCE ECONÓMICO FINANCIERO.

Introducción

En este capítulo se realiza el análisis y diseño del MD, donde se define el negocio, se identifican los requisitos funcionales y no funcionales, se explican las necesidades de los usuarios, se especifican los actores, se describen los CU. En el diseño se obtiene el modelo de datos. Además se definen las dimensiones del MD y sus hechos, así como la matriz bus que los relaciona.

Tema de Análisis

Para la construcción de la propuesta se definió como tema de análisis: balance económico financiero de las UP.

2.1 Necesidades del usuario

Constituye la forma de acceder y analizar la información por parte del usuario. A continuación se detallan las necesidades de usuarios encontradas:

1. El usuario necesita analizar la información del formulario 5910 “Estado de Situación” por alguna de las siguientes dimensiones: división política administrativa, organismo, entidad, subordinación, nomenclador de actividades económicas de todos los indicadores que se solicitan en el formulario en el tiempo por: apertura y en esta fecha.
2. El usuario necesita analizar la información del formulario 5911 “Estado de Resultado” por alguna de las siguientes dimensiones: división política administrativa, organismo, entidad, subordinación, nomenclador de actividades económicas de todos los indicadores que se solicitan en el formulario en el tiempo por: hasta esta fecha año anterior, plan anual y hasta la fecha.
3. El usuario necesita analizar la información del formulario 5912 “Inversiones y Donaciones” por alguna de las siguientes dimensiones: división política administrativa, organismo, entidad, subordinación, nomenclador de actividades económicas de todos los indicadores que se solicitan en el formulario en el tiempo por: plan, apertura y en esta la fecha.
4. El usuario necesita analizar la información del formulario 5913 “Disponibilidad de Fondos” por alguna de las siguientes dimensiones: división política administrativa, organismo, entidad,

subordinación, nomenclador de actividades económicas de todos los indicadores que se solicitan en el formulario en el tiempo por: importe.

5. El usuario necesita analizar la información del formulario 5914 “Estado de Gastos” por alguna de las siguientes dimensiones: división política administrativa, organismo, entidad, subordinación, nomenclador de actividades económicas, forma de gastos y nomenclador de actividades económicas por gastos de todos los indicadores que se solicitan en el formulario en el tiempo por: plan y real.
6. El usuario necesita analizar la información del modelo 0005-11 “Indicadores Generales” por alguna de las siguientes dimensiones: división política administrativa, organismo, entidad, subordinación, nomenclador de actividades económicas de todos los indicadores que se solicitan en el formulario en el tiempo por: real del año actual.
7. El usuario necesita analizar la información del modelo 5901-08 “Indicadores Seleccionados de la Contabilidad” por alguna de las siguientes dimensiones: división política administrativa, organismo, entidad, subordinación, nomenclador de actividades económicas de todos los indicadores que se solicitan en el formulario en el tiempo por: saldo inicial y saldo final.
8. El usuario necesita analizar la información del modelo 5903-04 “Cumplimiento del Plan Económico” por alguna de las siguientes dimensiones: división política administrativa, organismo, entidad, subordinación, nomenclador de actividades económicas de todos los indicadores que se solicitan en el formulario en el tiempo por: real acumulado del período del año actual.

2.2 Definición del negocio

El Ministerio de Finanzas y Precios (MFP) pone en vigor las Normas Cubanas de Información Financiera como base para el registro de los hechos económicos en las entidades del país, asegurando la integridad y el control financiero de los intereses del Estado cubano en entidades públicas, privadas y asociaciones con capital extranjero, incluyendo los principios, normas y procedimientos de contabilidad y costos necesarios. Establece las proformas y las indicaciones metodológicas de los Estados Financieros para las UP con el propósito de homogeneizar la información y posibilitar los procesos de agregación de la información a diferentes niveles, es decir, los cinco formularios estadísticos del balance financiero de las UP: 5910 “Estado de situación”, 5911 “Estado de resultados”, 5913 “Estado de disponibilidad de fondos”, 5914 y 5915 “Estado de gastos”.

Hasta el momento la ONE captaba los indicadores estadísticos, pero mediante un grupo de actividades se ha verificado que en muchos casos estos indicadores no correspondían con el balance financiero de las UP; por estas razones se deciden captar los registros primarios, es decir los cinco formularios definidos por el MFP y a través de las posibilidades que nos brinda el AD, facilitar los indicadores contables de los modelos estadísticos: 0005 "Indicadores Generales", 5901 "Indicadores Seleccionados de la Contabilidad" y 5903 "Cumplimiento del plan económico"; correspondientes a las UP. Toda esta información estadística de las UP es transferida a través de la red de gobierno para el Sistema de Información posibilitando la toma de decisiones por cada uno de los especialistas en las diferentes esferas de la economía.

2.3 Reglas del negocio

Las reglas de negocios definen y controlan la estructura, el funcionamiento y la estrategia de una organización. Pueden estar formalmente definidas en manuales de procedimiento, contratos o acuerdos. Son dinámicas, están sujetas a cambios en el tiempo y pueden encontrarse en todo tipo de aplicaciones [25]. Para la realización del MD se identificaron 97 reglas del negocio, a continuación se listan algunas de ellas, el resto se encuentra en el artefacto Reglas del negocio y transformación del Expediente de Proyecto.

MODELO 5910

RN1. Para obtener el activo circulante es necesario realizar la adición de los siguientes indicadores: Efectivo en Caja, Efectivo en Banco, Efectos por Cobrar, Cuentas por Cobrar a Corto Plazo, Cuentas por Cobrar Compra de Moneda, Pagos Anticipados a Suministradores, Pagos Anticipados del Proceso Inversionista, Anticipos a Justificar, Adeudos del Presupuesto del Estado, Adeudos del Organismo Superior, Reparaciones Generales en Proceso, Inventarios y Créditos Documentarios.

RN2. Para obtener el inventario es necesario realizar la adición de los siguientes indicadores: Materias Primas y Materiales, Combustibles y Lubricantes, Partes y Piezas de Repuesto, Envases y Embalajes, Útiles y Herramientas y Producción Terminada menos la adición de: Descuentos Comerciales e Impuesto, Mercancía para la Venta, Medicamentos, Base Material de Estudio, Vestuario y Lencería, Alimentos, Otros Inventarios, Producciones en Proceso, Producción Propia para Insumo, Inventarios Ociosos e Inventarios de Lento Movimiento.

RN3. Los Activos a Largo Plazo se obtienen a través de las Inversiones a Largo Plazo o Permanente.

RN4. El Activo Fijo se obtiene de la adición de los siguientes indicadores: Activos Fijos Tangibles Netos, Activos Fijos Intangibles Netos, Fondos Bibliotecarios, Inversiones Materiales, Equipos por Instalar y Materiales del Proceso Inversionista.

RN5. Para obtener los Activos Fijos Tangibles Netos es necesario realizar la adición de los siguientes indicadores: Depreciación de Activos Fijos Tangibles, Compra de Activos Fijos Tangibles y Activos Fijos Tangibles.

RN6. Para obtener los Activos Fijos Intangibles Netos es necesario realizar la adición de los siguientes indicadores: Activos Fijos Intangibles, Compra de Activos Fijos Intangibles y Amortización de Activos Fijos Intangibles.

RN7. Para obtener Otros Activos es necesario realizar la adición de los siguientes indicadores: Pérdidas en Investigación, Faltantes de Bienes en Investigación, Cuentas por Cobrar Diversas - Operaciones Corrientes, Cuentas por Cobrar Diversas del Proceso Inversionista, Efectos por Cobrar en Litigio, Cuentas por Cobrar en Litigio, Efectos por Cobrar Protestado, Cuentas por Cobrar en Proceso Judicial, Operaciones entre Dependencias – Activos y Depósitos y Finanzas.

RN8. Para obtener el Total del Activo es necesario realizar la adición de los siguientes indicadores: Otros Activos, Activo Fijo, Activos a Largo Plazo y Activo Circulante.

RN9. Para obtener el Pasivo Circulante es necesario realizar la adición de los siguientes indicadores: Cuentas por Pagar a Corto Plazo, Cuentas por Pagar Compra de Monedas, Cuentas por Pagar - Activos Fijos Tangibles, Cuentas por Pagar del Proceso Inversionista, Cobros Anticipados, Depósitos Recibidos, Obligaciones con el Presupuesto del Estado, Obligaciones con el Órgano u Organismo, Nóminas por Pagar, Retenciones por Pagar, Préstamos Recibidos, Gastos Acumulados por Pagar, Provisión para Vacaciones y Otras Provisiones Operacionales.

RN10. Para obtener el Pasivo a Largo Plazo es necesario realizar la adición de los siguientes indicadores: Cuentas por Pagar a Largo Plazo y Obligaciones a Largo Plazo.

RN11. Para obtener Otros Pasivos es necesario realizar la adición de los siguientes indicadores: Sobrantes en Investigación, Cuentas por Pagar Diversas y Operaciones entre Dependencias – Pasivo.

RN12. Para obtener el Total de Pasivo es necesario realizar la adición de los siguientes indicadores: Otros Pasivos, Pasivo a Largo Plazo y Pasivo Circulante.

RN13. Para obtener Total de Patrimonio Neto es necesario realizar la adición de los siguientes indicadores: Inversión Estatal, Recursos Recibidos Unidad Presupuestada, Recursos Entregados Unidad Presupuestada, Donaciones Recibidas, Donaciones Entregadas, Efectivo Depositado al Presupuesto del Estado y Resultado.

RN14. Para obtener el Total del Pasivo y del Patrimonio Neto es necesario realizar la adición de los siguientes indicadores: Total de Pasivo y Total de Patrimonio Neto.

RN15. La Diferencia se obtiene mediante la resta de En esta Fecha y Apertura.

2.4 Especificación de requerimientos

Requisitos Funcionales

El proceso de obtención de los requisitos es importante en el desarrollo del MD pues de la calidad de estos depende el éxito del producto. Los requisitos funcionales describen la función que el sistema debe hacer. Para el desarrollo de los procesos de integración se identificaron los siguientes requisitos funcionales:

RF1. Extraer datos del balance económico financiero

RF 1.1. Extraer la información de la base de datos SIGE, del estado de situación.

RF 1.2. Extraer la información de la base de datos SIGE, del estado de resultado.

RF 1.3. Extraer la información de la base de datos SIGE, de las inversiones y donaciones.

RF 1.4. Extraer la información de la base de datos SIGE, de la disponibilidad de fondos.

RF 1.5. Extraer la información de la base de datos SIGE, del estado de gastos.

RF2. Realizar transformación y carga del balance económico financiero

RF 2.1. Transformar y cargar la información de la base de datos SIGE, del estado de situación.

RF 2.2. Transformar y cargar la información de la base de datos SIGE, del estado de resultado.

RF 2.3 Transformar y cargar la información de la base de datos SIGE, de las inversiones y donaciones.

RF 2.4. Transformar y cargar la información de la base de datos SIGE, de la disponibilidad de fondos.

RF 2.5. Transformar y cargar la información de la base de datos SIGE, del estado de gastos.

RF 2.6. Transformar y cargar la información del MD Balance económico financiero, de los indicadores generales.

RF 2.7. Transformar y cargar la información del MD Balance económico financiero, de los indicadores seleccionados de la contabilidad.

RF 2.8. Transformar y cargar la información del MD Balance económico financiero, del cumplimiento del plan económico.

Requisitos no funcionales

Los requisitos no funcionales definen características que restringen el sistema como el rendimiento, la protección, la disponibilidad y otras propiedades. A continuación se muestran los requisitos no funcionales seleccionados para los procesos de ETL, el resto de los requisitos identificados para la construcción del MD se encuentran en el artefacto Especificación de Requerimientos del Expediente de Proyecto:

Fiabilidad

RNF 1. Asegurar recuperación ante fallos.

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

RNF 2. Garantizar la persistencia de la información.

Para garantizar la persistencia de la información se realizará un respaldo total de los datos del MD con una frecuencia mensual y anual. Toda esta información se almacenará en el área de la dirección de informática en un banco de datos especial. Además cada 6 meses, antes de los períodos vacacionales de julio y diciembre, se realizará una salva general de los datos en tres versiones (abuelo, padre e hijo) y se guardará en un DVD. Esta información se almacenará en el edificio correspondiente a la oficina de estadísticas de La Habana y será responsabilidad del grupo de administración de redes de la ONE.

RNF 3. Garantizar el cumplimiento de actualización de los datos en el almacén.

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

Usabilidad

RNF 4. Lograr que los elementos definidos en el mercado tengan una estructura homogénea.

Las estructuras del MD se nombrarán de una manera estándar teniendo en cuenta el tipo de estructura que se maneje.

Requisitos de seguridad

RNF 5. 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.

Requisitos de software

RNF 6. Instalar en las estaciones de trabajo el software necesario

- Pentaho Data Integration 4.0 o posterior para la ejecución de las transformaciones y trabajos diseñados.
- Máquina Virtual de Java para la ejecución del Pentaho Data Integration 4.0.

Requisitos de hardware

RNF 7. 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, Pentium III.

RNF 8. 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 permite hasta 9 conexiones.
- 1 Microprocesador Dual Core / 2.2 GHz.
- Disco duro de 160 GB.

2.5 Modelo de casos de uso del sistema

Roles y Permisos

Los roles y permisos están dados por el tipo de función que realice cada una de las personas en el sistema.

Roles	Permisos	
	Lectura	Escritura
Administrador_ETL	X	X

Tabla 1 Tabla de roles y permisos

El administrador de ETL es el encargado de realizar todos los procesos correspondientes a la extracción, transformación y carga de los datos.

Casos de uso del sistema

Los casos de uso del sistema (CUS) ayudan a describir qué es lo que el sistema debe hacer desde el punto de vista del usuario, es decir, describen la interacción entre el sistema y el usuario.

CUS-1 Extraer información del balance económico financiero.

CUS-2 Realizar transformación y carga del balance económico financiero.

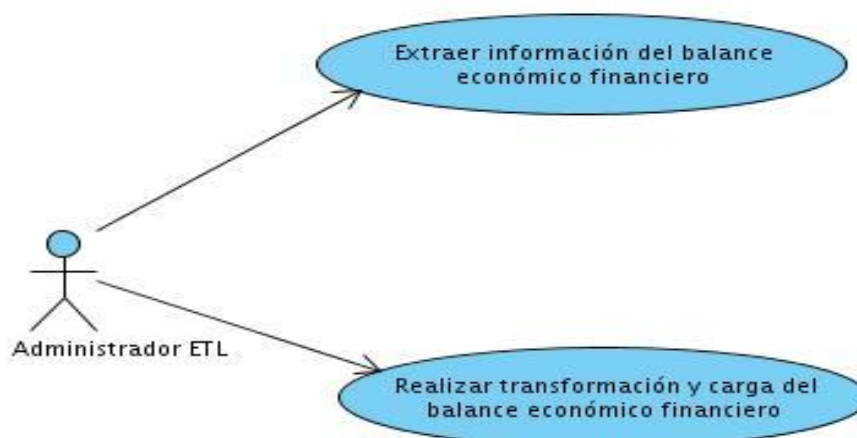


Figura 1 Diagrama de casos de usos del sistema

Especificación de los casos de uso del sistema.

- CU Extraer información balance económico financiero.

Caso de Uso:	Extraer información del balance económico financiero.
Tipo:	Funcional
Actores:	Administrador ETL
Resumen:	El CU inicia cuando el administrador ETL desea extraer los datos de la fuente. Se selecciona la fuente de información correspondiente y extrae los datos contenidos en ella. El CU termina cuando el administrador extrae los datos.
Precondiciones:	Disponibilidad de la fuente.
Referencias	RF1
Prioridad	Crítico

Flujo Normal de Eventos

Acción del Actor	Respuesta del Sistema.
<p>1. El administrador ETL selecciona el modelo al cual desea realizar la extracción:</p> <ul style="list-style-type: none"> • Estado de Resultado. • Estado de Situación. • Disponibilidad de Fondos. • Inversiones y Donaciones. • Estado de Gastos. • Indicadores Generales. • Indicadores Seleccionados de la Contabilidad. • Cumplimiento del Plan Económico. 	

CAPÍTULO 2 ANÁLISIS Y DISEÑO DE LOS PROCESOS DE INTEGRACIÓN.

2. El administrador ETL realiza la conexión a la BD SIGE.	3. El sistema responde a la solicitud de conexión.
4. El administrador ETL selecciona estructuras o archivos a extraer.	
5. El administrador ETL realiza la extracción.	6. Ejecuta la extracción de los datos. Finaliza el caso de uso.
Prototipo de Interfaz	
Flujos Alternos	
Acción del Actor	Respuesta del Sistema
	2.1 El sistema notifica el error al administrador ETL.
Poscondiciones	Datos disponibles para transformar.

Tabla 2 Extraer información balance económico financiero

- CU Realizar transformación y carga balance económico financiero.

Caso de Uso:	Realizar transformación y carga del balance económico financiero.
Tipo:	Funcional
Actores:	Administrador ETL
Resumen:	El CU se inicia cuando el administrador ETL realiza la transformación y carga de los datos. El administrador ETL realiza las transformaciones convenientes y carga la información resultante en el MD. El caso de uso finaliza cuando los cambios se encuentran registrados en la BD.

Precondiciones:	Extracción completada. BD disponible.	
Referencias	RF2	
Prioridad	Crítico	
Flujo Normal de Eventos		
Acción del Actor	Respuesta del Sistema.	
2. El administrador ETL selecciona estructuras a transformar.	1. El sistema muestra las estructuras a transformar.	
3. El administrador ETL carga los datos seleccionados en memoria.		
4. El administrador ETL aplica las transformaciones pertinentes y genera datos de auditoría.		
5. El administrador ETL carga datos en el MD.	6. Ejecuta la consulta. Finaliza el caso de uso.	
Prototipo de Interfaz		
Poscondiciones	Los datos de la BD son transformados y cargados.	

Tabla 3 Realizar transformación y carga balance económico financiero

2.6 Especificación del modelo dimensional

Matriz BUS

La matriz BUS muestra la relación que existe entre los hechos y las dimensiones de la solución. Las filas contienen todos los hechos que fueron identificados y en las columnas las dimensiones. En la misma quedan evidenciadas las relaciones existentes entre las tablas demostrando

TH/DIM	ORG	SUB	DPA	NAE	Temporal mes	Entidad	Indicador general	Indicador balance	NAE por gastos	Forma de gastos
Estado de situación	x	x	x	x	x	x		x		
Estado de resultados	x	x	x	x	x	x		x		
Disponibilidad de fondos	x	x	x	x	x	x		x		
Inversiones y donaciones	x	x	x	x	x	x		x		
Estado de gastos	x	x	x	x	x	x		x	x	x
Indicadores generales	x	x	x	x	x	x	x			
Indicadores de la contabilidad	x	x	x	x	x	x	x			
Cumplimiento del plan económico	x	x	x	x	x	x	x			

Tabla 4 Matriz BUS

Modelo de Datos

El modelo de datos describe todo lo relacionado con la BD, mostrando las condiciones que deben cumplir los datos para ser mostrados correctamente. Refleja la relación de cada una de las tablas de hechos y dimensiones, lo que define el esquema arquitectónico del MD.

Por las características que refleja el modelo de datos diseñado para la solución, se define como esquema arquitectónico constelación de hechos, ya que las tablas de dimensiones están relacionadas con más de una tabla de hechos.

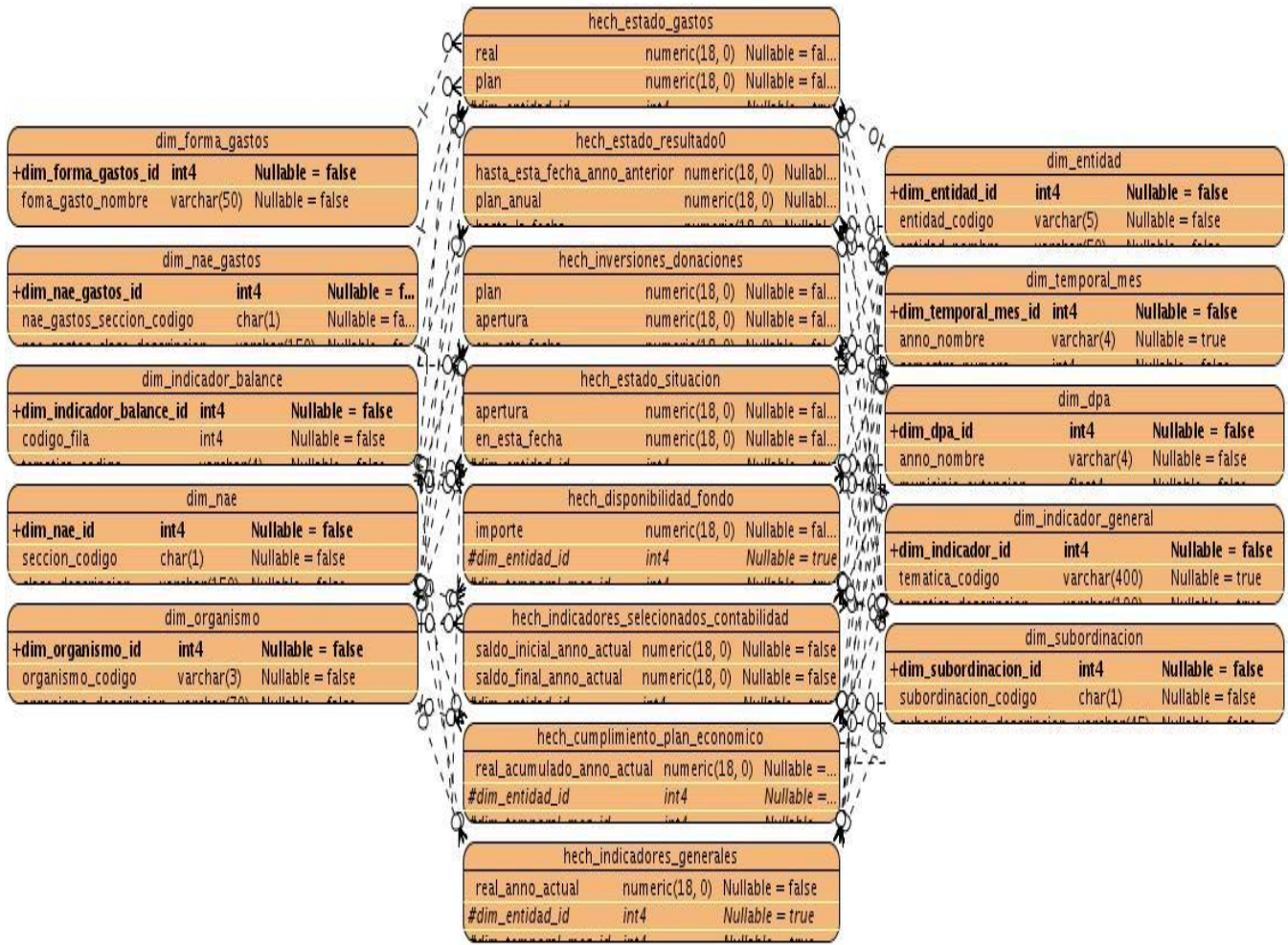


Figura 2 Diseño lógico del mercado de datos

Dimensiones

1. Entidad (dim_entidad)

Jerarquía:

- Centro informante

2. Nomenclador de Actividades Económicas (dim_nae)

Jerarquía:

- Sección
- División

- Clase

3. División Político-Administrativa (dim_dpa)

Jerarquía:

- Provincia
- Municipio

4. Organismos (dim_organismo)

Jerarquía:

- Organismo

5. Subordinación (dim_subordinacion)

Jerarquía:

- Subordinación

6. Temporal mes (dim_temporal_mes)

Jerarquía:

- Año
- Mes

7. Indicador balance (dim_indicador_balance)

Jerarquía:

- Temática
- Subtemática
- Indicador

8. Forma de gastos (dim_forma_gatos)

Jerarquía:

- Forma de gastos

9. Nomenclador de actividades económicas por gastos (dim_nae_gastos)

Jerarquía:

- Sección
- División
- Clase

10. Indicador general (dim_indicador_general)

Jerarquía:

- Temática
- Subtemática
- Indicador

Tablas de hechos y medidas asociadas

- Tabla de hecho Estado de resultado

Atributos	ID	Tipo de dato
hech_estado_resultado	hech_estado_resultado_id	integer
dim_subordinacion	dim_subordinacion_id	integer
dim_organismo	dim_organismo_id	integer
dim_dpa	dim_dpa_id	integer
dim_centro_informante	dim_centro_informante_id	integer
dim_nae	dim_nae_id	integer
dim_temporal_mes	dim_temporal_mes_id	integer
dim_indicador_balance	dim_indicador_balance_id	integer
Medidas		
hasta_esta_fecha_anno_anterior	numeric	
plan_anual	numeric	
hasta_la_fecha	numeric	

Tabla 5 Tabla de hecho Estado de resultado

➤ Tabla de hecho Indicadores generales

Atributos	ID	Tipo de dato
hec_indicadores_generales	hec_indicadores_generales_id	integer
dim_subordinacion	dim_subordinacion_id	integer
dim_organismo	dim_organismo_id	integer
dim_dpa	dim_dpa_id	integer
dim_entidad	dim_entidad_id	integer
dim_nae	dim_nae_id	integer
dim_temporal_mes	dim_temporal_mes_id	integer
dim_indicador_general	dim_indicador_general_id	integer
Medidas		
real_anno_actual		numeric

Tabla 6 Tabla de hecho Indicadores generales

➤ Tabla de hecho Indicadores seleccionados de la contabilidad

Atributos	ID	Tipo de dato
hech_indicadores_seleccionados_contabilidad	hech_indicadores_seleccionados_conta	integer
dim_subordinacion	dim_subordinacion_id	integer
dim_organismo	dim_organismo_id	integer
dim_dpa	dim_dpa_id	integer
dim_entidad	dim_entidad_id	integer
dim_nae	dim_nae_id	integer
dim_temporal_mes	dim_temporal_mes_id	integer
dim_indicador_general	dim_indicador_general_id	integer
Medidas		
saldo_inicial_anno_actual		numeric
saldo_final_anno_actual		numeric

Tabla 7 Tabla de hecho Indicadores seleccionados de la contabilidad

➤ Tabla de hecho Estado de situación

Atributos	ID	Tipo de dato
hech_estado_situacion	hech_estado_situacion_id	integer
dim_subordinacion	dim_subordinacion_id	integer
dim_organismo	dim_organismo_id	integer
dim_dpa	dim_dpa_id	integer
dim_entidad	dim_entidad_id	integer
dim_nae	dim_nae_id	integer
dim_temporal_mes	dim_temporal_mes_id	integer
dim_indicador_balance	dim_indicador_balance_id	integer
Medidas		
apertura	numeric	
en_esta_fecha	numeric	

Tabla 8 Tabla de hecho Estado de situación

➤ Tabla de hecho Cumplimiento del plan económico

Atributos	ID	Tipo de dato
hech_cumplimiento_plan_economico	hech_cumplimiento_plan_economico_id	integer
dim_subordinacion	dim_subordinacion_id	integer
dim_organismo	dim_organismo_id	integer
dim_dpa	dim_dpa_id	integer
dim_entidad	dim_entidad_id	integer
dim_nae	dim_nae_id	integer
dim_temporal_mes	dim_temporal_mes_id	integer
dim_indicador_general	dim_indicador_general_id	integer
Medidas		
real_acumulado_anno_actual	numeric	

Tabla 9 Tabla de hecho Cumplimiento del plan económico

➤ Tabla de hecho Disponibilidad de fondo

Atributos	ID	Tipo de dato
hech_disponibilidad_fondo	hech_disponibilidad_fondo_id	integer
dim_subordinacion	dim_subordinacion_id	integer
dim_organismo	dim_organismo_id	integer
dim_dpa	dim_dpa_id	integer
dim_entidad	dim_entidad_id	integer
dim_nae	dim_nae_id	integer
dim_temporal_mes	dim_temporal_mes_id	integer
dim_indicador_balance	dim_indicador_balance_id	integer
Medidas		
importe		numeric

Tabla 10 Tabla de hecho Disponibilidad de fondos

➤ Tabla de hecho Estado gastos

Atributos	ID	Tipo de dato
hech_estado_gastos	hech_estado_gastos_id	integer
dim_subordinacion	dim_subordinacion_id	integer
dim_organismo	dim_organismo_id	integer
dim_dpa	dim_dpa_id	integer
dim_entidad	dim_entidad_id	integer
dim_nae	dim_nae_id	integer
dim_temporal_mes	dim_temporal_mes_id	integer
dim_indicador_balance	dim_indicador_balance_id	integer
Medidas		
real		numeric
plan		numeric

Tabla 11 Tabla de hecho Estado gastos

➤ Tabla de Inversiones y donaciones

Atributos	ID	Tipo de dato
hech_inversiones_donaciones	hech_inversiones_donaciones_id	integer
dim_subordinacion	dim_subordinacion_id	integer
dim_organismo	dim_organismo_id	integer
dim_dpa	dim_dpa_id	integer
dim_entidad	dim_entidad_id	integer
dim_nae	dim_nae_id	integer
dim_temporal_mes	dim_temporal_mes_id	integer
dim_indicador_balance	dim_indicador_balance_id	integer
Medidas		
plan	numeric	
apertura	numeric	
en_esta_fecha	numeric	

Tabla 12 Tabla de Inversiones y donaciones

2.7 Seguridad

Esquema de Seguridad

El esquema de seguridad estará respaldado por los niveles de acceso, específicamente por los roles definidos.

La arquitectura del sistema de seguridad y alta disponibilidad debe ser de 3 capas:

- Funcionamiento: Dispositivos de seguridad (firewall, inspectores de contenido, sensores).
- Servidores: Servidor de gestión de administración y de BD.
- Presentación: Consolas de administración.

La configuración y prueba de todos los equipos deberá efectuarse en las instalaciones de la ONE, conjuntamente con personal técnico especializado, de acuerdo a las especificaciones de configuración y prueba que serán elaboradas.

El mantenimiento preventivo deberá efectuarse mediante una visita semestral, durante el período de vigencia de la garantía técnica.

El mantenimiento correctivo deberá efectuarse anualmente, a requerimiento de la ONE, con tiempos de respuesta por soporte técnico de 4 horas y de solución de hasta 48 horas, los 365 días del año, de fácil ubicación vía e-mail o telefónica.

Política de respaldo y recuperación

La política de respaldo y recuperación que utiliza la solución es sencilla, pero a la vez sólida, para ello se miden 3 puntos esenciales:

- Periodicidad de las salvas: Mensualmente se realizan las salvas de toda la información contenida en la BD. El Banco de Datos es el encargado de recoger esta información antes salvada. Cada seis meses preferentemente antes el periodo vacacional se hace una salva total y se extrae fuera de la organización en DVD-RW con tres versiones (abuelo, padre e hijo).
- Tablas involucradas: Las tablas que se involucran en la realización son las tablas de hecho y las tablas de dimensiones.
- Backups existentes: Actualmente existen backups en esta área.

Conclusiones Parciales

Durante el desarrollo de este capítulo quedaron definidos varios aspectos importantes para la implementación de los procesos de integración del MD:

- Mediante la definición del negocio se lograron detectar 97 reglas del negocio, las cuales constituyen un hito importante para la construcción del MD.
- Se obtuvieron dos CU y se describieron para darle solución a las 8 necesidades de los usuarios encontradas.
- Se definieron 10 dimensiones, 8 tablas de hechos y las medidas asociadas, posibilitando la creación del modelo lógico y físico, necesario para la obtención del script de la BD que posteriormente se utilizará para poblar el MD.

CAPÍTULO 3

IMPLEMENTACIÓN Y PRUEBA DE LOS PROCESOS DE INTEGRACIÓN DEL MERCADO DE DATOS BALANCE ECONÓMICO FINANCIERO.

Introducción

En el presente capítulo se implementa y se prueba el MD. Se especifica la estructura del modelo de datos definido. También se describen los procesos de perfilado de datos y ETL. Por último se llegará a la realización de las pruebas para la validación de los procesos de ETL, donde se podrá probar la solución con el fin de detectar errores en ella y mitigarlos antes de la entrega.

3.1 Modelo de datos físico

El modelo de datos se usa para organizar todo lo relacionado con la BD, permitiendo mostrar de una manera gráfica la interacción entre las distintas tablas de la solución lo que facilita la creación del script de la BD. El modelo físico se usa para describir los datos en el nivel más bajo.

Estructuras de datos

Una estructura de datos es una forma de organizar un conjunto de datos. Se caracteriza por las funciones que se usan para almacenar y acceder a los elementos individuales de los datos. Para un mejor entendimiento de la solución se estructuraron los datos por esquemas y tablas.

Esquemas y tablas

Los esquemas representan una forma para tener organizada la información dentro de una BD, agrupan tablas, campos en cada tabla y las relaciones entre cada campo y cada tabla. Cada tabla tiene una o más columnas y filas. Las columnas guardan parte de la información y los elementos que se quieren recoger en la columna y cada fila de la tabla conforman un registro.

Para el desarrollo de la solución se cuenta con dos esquemas:

- esquema dimensiones: contiene todas las dimensiones que se relacionan con todos los hechos del MD.
- esquema mart_balance: agrupa todos los hechos y las dimensiones que son únicas para cada hecho.

Para la solución fueron modeladas 18 tablas que representan cada una de las entidades modeladas en el análisis.

En la siguiente tabla se describe dicha estructura:

Esquemas	Tablas
dimensiones	dim_nae
dimensiones	dim_entidad
dimensiones	dim_dpa
dimensiones	dim_organismo
dimensiones	dim_subordinación
dimensiones	dim_temporal_mes
dimensiones	dim_indicador_general
mart_balance	dim_indicador_balance
mart_balance	dim_nae_gastos
mart_balance	dim_forma_gastos
mart_balance	hech_cumplimiento_plan_economico
mart_balance	hech_indicadores_seleccionados_contabilidad
mart_balance	hech_indicadores_generales
mart_balance	hech_estado_situacion
mart_balance	hech_estado_resultado
mart_balance	hech_inversiones_donaciones
mart_balance	hech_disponibilidad_fondo
mart_balance	hech_estado_gastos

Tabla 13 Tablas del modelo datos

3.2 Perfilado de datos

El perfilado de datos se utiliza para analizar los datos de la fuente con el objetivo de obtener estadísticas e información de los mismos. El análisis de la información realizado en la presente investigación permitió identificar que los datos de la fuente no presentan los siguientes problemas: valores indebidos, escritos incorrectamente, ausentes o duplicados ya que son provenientes de una fuente confiable.

CAPÍTULO 3 IMPLEMENTACIÓN Y PRUEBA

La realización del perfilado de datos permitió un efectivo análisis de los datos extraídos. A continuación se muestran algunos de los reportes generados, el resto del proceso del perfilado de datos se puede encontrar en el artefacto Perfil de los datos del Expediente de Proyecto.

- En el perfilado de los estándares de medidas se obtiene el resultado de todas las filas, valores nulos y vacíos, así como los valores más altos y más bajos la tabla analizada.

	codcentroinformante	m5910enestafec	codfila	idfechadelinformeacumulado	m5910diferencia	idcodvariante	m5910apertura
Row count	33	33	33	33	33	33	33
Null values	0	0	0	0	0	0	0
Empty va...	0	0	0	0	0	0	0
Highest ...	09260	92770819	92	2011-01-30 00:00:00.0	13769474	0	93219328
Lowest v...	09260	-16815200	10	2011-01-30 00:00:00.0	-3121175	0	0

Tabla 14 Perfilado de los estándares de medidas del modelo 5910

- Con el reporte del análisis de cadenas se obtiene toda la información de la cantidad, mínima, máxima y promedio de caracteres, los espacios en blanco, los caracteres no escritos, entre otros parámetros.

	codcentroinformante	codfila
Char count	165	62
Max chars	5	2
Min chars	5	1
Avg chars	5	1,88
Max white spaces	0	0
Min white spaces	0	0
Avg white spaces	0	0
Uppercase chars	0%	0%
Lowercase chars	0%	0%
Non-letter chars	100%	100%
Word count	33	33
Max words	1	1
Min words	1	1

Tabla 15 Perfilado del análisis de cadenas del modelo 5910

- El perfilado análisis numérico permite conocer el menor y el mayor valor, la suma de los valores, la medida geométrica, el promedio y la desviación estándar, así como la diferencia.

	m5910enestafec	m5910diferencia	idcodvariante	m5910apertura
Highest value	92770819	13769474	0	93219328
Lowest value	-16815200	-3121175	0	0
Sum	631978243	32136415	0	650287428
Mean	19150855,85	973830,76	0	19705679,64
Geometric mean			0	0
Standard deviation	27861690,88	4152654,1	0	28187215,94
Variance	776273818784822,5	17244536100923,69	0	794519142344066,9

Tabla 16 Perfilado análisis numérico del modelo 5910

3.3 Implementación de los procesos de extracción, transformación y carga

Extracción de los datos

El primer proceso de ETL consiste en extraer los datos desde los sistemas de origen, que pueden ser provenientes de diferentes fuentes. A través de las fuentes se establece desde dónde se extraerán los datos para analizarlos.

Cada sistema puede usar formatos distintos o una organización diferente de los datos. El formato de la fuente de esta investigación se encuentra en una BD relacional y en archivos xls para el caso de las dimensiones. Este proceso convierte los datos a un formato preparado para iniciar el proceso de transformación.

Para este proceso el sistema realiza la conexión con el servidor SIGE (Sistema Integrado de Gestión Estadística) específicamente con la BD PADTSI_ENCUESTA (Paquete de Ayuda a la Toma de Decisiones y Soluciones Integrales) y crea una salva (backup) de forma tal que se eviten los problemas de acceso a dicha BD.

Implementación de las transformaciones

La transformación es el proceso básico de ETL, se compone de pasos que están enlazados a través de saltos. Los pasos son los elementos más pequeños dentro de las transformaciones. Los saltos son el medio por donde fluye la información entre los diferentes pasos.

Después de realizada la extracción de los datos el sistema se encuentra listo para la etapa de transformación. Durante el proceso se llevaron a cabo tareas tales como: unión por clave para que a través de la entidad a la que pertenece la información de la fuente se obtengan la división política

administrativa (dpa), organimo (org), subordinación (surb) y el nomenclador de actividades económicas (nae) a la cual pertenece cada entidad; ordenamiento de filas para organizar los campos de las tablas; agrupamiento y cálculo de datos para calcular las medidas de algunos indicadores, así como asignación de llaves para relacionar la información de los hechos con las dimensiones.

Carga de datos

La carga es el último subproceso dentro de los procesos de ETL, el cual consiste en cargar todos los datos que ya han sido transformados satisfactoriamente.

Para una mayor organización en la carga de los datos se realizó primero la carga de las tablas que se encuentran en el esquema dimensiones, o sea, las que no dependen de la información de otras tablas. Luego se realizó la carga de las tablas del esquema mart_balance cargando en primer lugar las dimensiones, los hechos que tributan a los modelos de las UP y por último se cargaron los hechos que tienen alguna dependencia de otros, que son los correspondientes a los modelos de contabilidad.

En la ilustración 3 se muestra la transformación para la carga de la tabla hech_estado_gastos, el resto de las transformaciones realizadas podrán ser vistas en el Anexo 1.

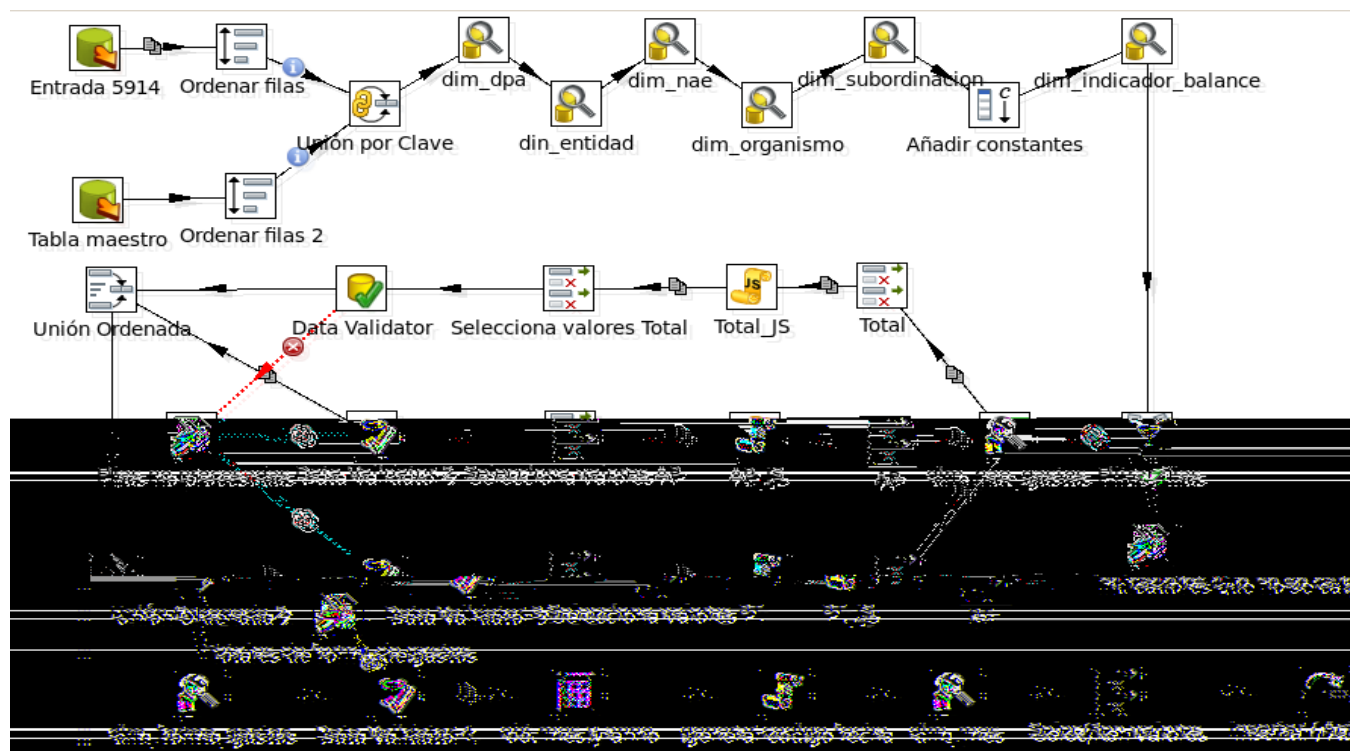


Figura 3 Carga de la tabla hech_estado_gastos

Implementación de los trabajos

Un trabajo es un conjunto sencillo o complejo de tareas con el objetivo de realizar una acción determinada. En los trabajos se utilizan pasos específicos que son diferentes a los disponibles en las transformaciones. Permite ejecutar una o varias transformaciones de las diseñadas siguiendo una secuencia de ejecución.

Después que se realizaron las transformaciones a los datos se organizó la carga de las tablas estableciendo el orden explicado anteriormente. La carga en este orden es importante ya que se evita cargar llaves nulas que podrían pertenecer a otras tablas que no han sido cargadas.

En la siguiente figura se muestran los trabajos diseñados para el proceso de ETL anteriormente expuesto:

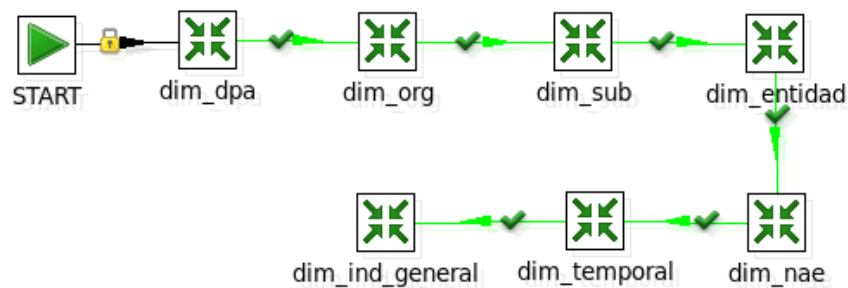


Figura 4 Trabajo para la carga de las tablas del esquema dimensiones

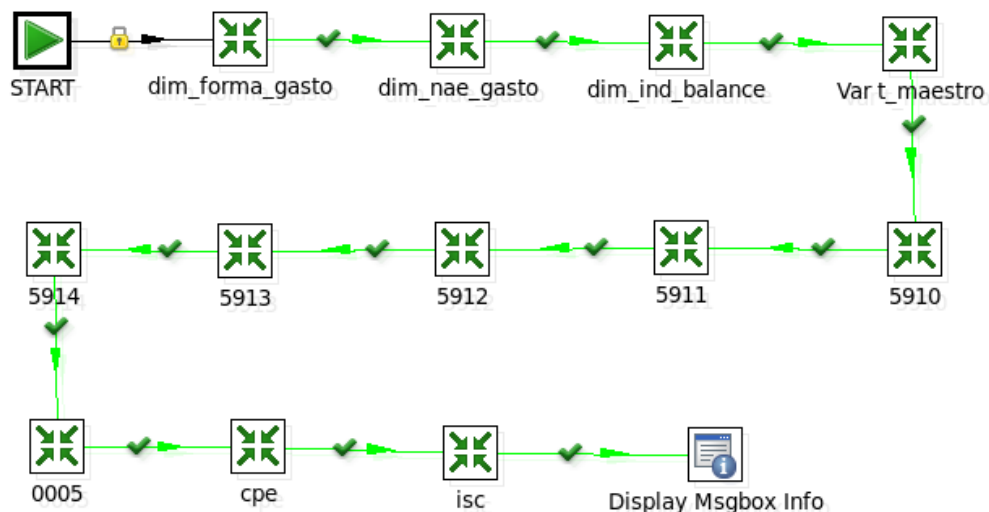


Figura 5 Trabajo para la carga de las tablas del esquema mart_balance

3.4 Estrategia de validación y prueba

Luego de realizado los procesos de ETL se procede a validar el resultado obtenido mediante casos de pruebas y listas de chequeo, con el fin de comprobar que los datos fueron correctamente cargados a la BD destino.

Listas de chequeo

Las listas de chequeo se realizan con el objetivo de evaluar y verificar el grado de cumplimiento de determinadas reglas establecidas mediante un listado de preguntas.

En esta investigación se aplicaron las siguientes listas de chequeo a los artefactos de los procesos ETL con el fin de evaluar y verificar el potencial de cada uno de ellos midiendo la confiabilidad y seguridad de los datos cargados:

- Lista de chequeo del Mapa Lógico de Datos.
- Lista de chequeo del Diccionario de Datos.
- Lista de chequeo de Registro de Sistemas Fuentes.

Casos de prueba

Los casos de pruebas constituyen un conjunto de escenarios mediante los cuales se determina si el requisito de una aplicación es total o parcial. Se pueden utilizar varios de estos para determinar si el requisito es completamente satisfactorio.

Para la validación de las reglas del negocio y de los procesos de integración de datos se realizaron casos de pruebas a los procesos de ETL, seleccionando los escenarios a probar que no son mas que las tablas de hechos y dimensiones del MD y los distintos campos de dichas tablas, los cuales pueden tomar los valores de válidos o inválidos en caso de que estén correctos o incorrectos. Los casos de prueba diseñados para la solución están en la carpeta Casos de prueba ETL del Expediente de Proyecto.

Pruebas de desarrollo

Las pruebas de desarrollo se realizan con el objetivo de asegurar la calidad de los datos en el MD, específicamente para los procesos de ETL. Consiste en realizar comparaciones entre los datos provenientes del sistema origen y los datos obtenidos en la BD destino.

Especificación de las pruebas

Para controlar la calidad se siguió el criterio de que los datos extraídos de las BD fuentes deben ser los mismos a los cargados en la BD destino siguiendo la siguiente estrategia:

Caso 1.

- Activación de los procesos de ETL.
- Realización de la transferencia de datos.
- Revisión de archivos de errores en la transferencia.

Caso 2.

- Comparar la cantidad de registros por tabla entre BD fuente y la BD destino.
- Verificar duplicación de datos en las tablas de hechos y dimensiones

Especificación del procedimiento de la prueba

A continuación se describen los casos de prueba definidos anteriormente y el resultado obtenido.

Para el caso1:

No. Paso	Función a probar	Acción	Resultados esperados
1	Activación de los procesos de ETL.	Ejecución de los procesos ETL.	Inicio del proceso ETL.
2	Realización de la transferencia de datos.		Integración de los datos.
3	Revisión de los archivos de errores.		Se crean archivos en caso de ocurrir errores en la transferencia.

Tabla 17 Funciones a probar del caso 1

Paso 1:

Para la activación de los procesos ETL a través de la interfaz gráfica de PDI Spoon solo es necesario ejecutar el trabajo programado.

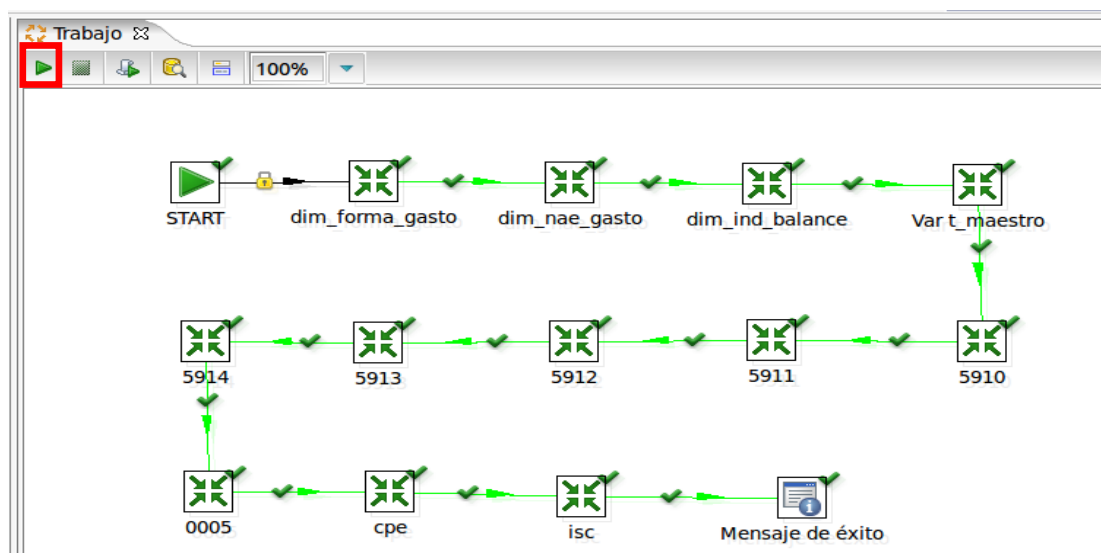


Figura 6 Ejecución del trabajo en Spoon

Para el caso de no presentar esta interfaz, Kitchen, otras de las herramientas que ofrece PDI, posibilita la ejecución de los trabajos diseñados a través de la consola.

Descripción del comando	Comandos Linux	Comandos Windows
Ejecutar Kitchen	Kitchen.sh	Kitchen.bat
Ejecutar archivo XML (.kjb)	-file=nombredearchivo	/file:nombredearchivo
Establecer el archivo de log	-log=nombredearchivo	/log:nombredearchivo
Establecer el nivel de log	-level=niveldelog	/level:niveldelog

Tabla 18 Principales opciones de línea de comandos

Para realizar alguna ejecución es necesario estar posicionado en el directorio de PDI.

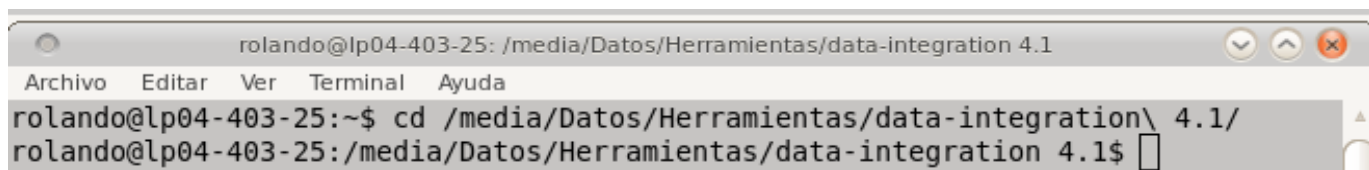


Figura 7 Posicionar al directorio de PDI

Luego se utiliza el comando `-file=nombredearchivo` para la ejecución del trabajo diseñado.

```

rolando@lp04-403-25: /media/Datos/Herramientas/data-integration 4.1
rolando@lp04-403-25:~$ cd /media/Datos/Herramientas/data-integration\ 4.1/
rolando@lp04-403-25:/media/Datos/Herramientas/data-integration 4.1$ sh kitchen.sh
-file="/home/rolando/Escritorio/Balance/Trabajo.kib"
    
```

Figura 8 Ejecución del trabajo en Kitchen

Paso 2:

La transferencia de datos se realizó correctamente logrando la integración de los datos en el MD.

Execution results		
Trabajo / Entrada de Trabajo	Comentario	Resultado
dim_ind_balance	Start of job execution	
dim_ind_balance	Job execution finished	Exito
Var t_maestro	Start of job execution	
Var t_maestro	Job execution finished	Exito
5910	Start of job execution	
5910	Job execution finished	Exito
5911	Start of job execution	
5911	Job execution finished	Exito
5912	Start of job execution	
5912	Job execution finished	Exito
5913	Start of job execution	
5913	Job execution finished	Exito
5914	Start of job execution	
5914	Job execution finished	Exito
0005	Start of job execution	
0005	Job execution finished	Exito
cpe	Start of job execution	
cpe	Job execution finished	Exito
isc	Start of job execution	
isc	Job execution finished	Exito
Trabajo: Trabajo	Job execution finished	Exito

Figura 9 Resultados de la ejecución del trabajo en Spoon

CAPÍTULO 3 IMPLEMENTACIÓN Y PRUEBA

```

rolando@ip04-403-25: /media/Datos/Herramientas/data-integration 4.1
Archivo  Editar  Ver  Terminal  Ayuda
INFO 09-06 09:25:40,045 - Ordenar filas 4 - Finished processing (I=0, O=0, R=50, W=50, U=0, E=0)
INFO 09-06 09:25:40,051 - Agrupar por 2 3 - Finished processing (I=0, O=0, R=381, W=378, U=0, E=0)
INFO 09-06 09:25:40,057 - Ordenar filas 5 - Finished processing (I=0, O=0, R=379, W=379, U=0, E=0)
INFO 09-06 09:25:40,061 - Ordenar filas 6 - Finished processing (I=0, O=0, R=812, W=812, U=0, E=0)
INFO 09-06 09:25:40,067 - Ordenar filas 7 - Finished processing (I=0, O=0, R=2237, W=2237, U=0, E=0)
INFO 09-06 09:25:40,388 - Filas No Procesadas 5901 - Finished processing (I=0, O=33096, R=33095, W=33095, U=0, E=0)
INFO 09-06 09:25:40,632 - dim_ind_general - Finished processing (I=2236, O=0, R=2237, W=2236, U=0, E=0)
INFO 09-06 09:25:40,635 - Renombrar 7 - Finished processing (I=0, O=0, R=2236, W=2236, U=0, E=0)
INFO 09-06 09:25:42,637 - Insertar / Actualizar hech_ISC - Finished processing (I=2236, O=53, R=2236, W=2236, U=0, E=0)
INFO 09-06 09:25:42,648 - Trabajo - Starting entry [Mensaje de éxito]
INFO 09-06 09:25:42,860 - Trabajo - Finished job entry [Mensaje de éxito] (result=[true])
INFO 09-06 09:25:42,860 - Trabajo - Finished job entry [isc] (result=[true])
INFO 09-06 09:25:42,860 - Trabajo - Finished job entry [cpe] (result=[true])
INFO 09-06 09:25:42,860 - Trabajo - Finished job entry [0005] (result=[true])
INFO 09-06 09:25:42,860 - Trabajo - Finished job entry [5914] (result=[true])
INFO 09-06 09:25:42,860 - Trabajo - Finished job entry [5913] (result=[true])
INFO 09-06 09:25:42,861 - Trabajo - Finished job entry [5912] (result=[true])
INFO 09-06 09:25:42,861 - Trabajo - Finished job entry [5911] (result=[true])
INFO 09-06 09:25:42,861 - Trabajo - Finished job entry [5910] (result=[true])
INFO 09-06 09:25:42,861 - Trabajo - Finished job entry [Var t_maestro] (result=[true])
INFO 09-06 09:25:42,861 - Trabajo - Finished job entry [dim_ind_balance] (result=[true])
INFO 09-06 09:25:42,861 - Trabajo - Finished job entry [dim_nae_gasto] (result=[true])
INFO 09-06 09:25:42,861 - Trabajo - Finished job entry [dim_forma_gasto] (result=[true])
INFO 09-06 09:25:42,861 - Trabajo - Job execution finished
INFO 09-06 09:25:42,862 - Kitchen - Finished!
INFO 09-06 09:25:42,862 - Kitchen - Start=2011/06/09 09:20:48.752, Stop=2011/06/09 09:25:42.862
INFO 09-06 09:25:42,862 - Kitchen - Processing ended after 4 minutes and 54 seconds (294 seconds total).
    
```

Figura 10 Resultados de la ejecución del trabajo en Kitchen

Paso 3:

Los archivos de log de errores no recibieron información ya que la integración se realizó exitosamente.

Nombre	Tamaño	Tipo
Filas no procesadas 0005.xls	133,5 KiB	Archivo de copia de seguridad de Microsoft Office Excel
Filas no procesadas 5901.xls	161,0 KiB	Archivo de copia de seguridad de Microsoft Office Excel
Filas no procesadas 5903.xls	125,0 KiB	Archivo de copia de seguridad de Microsoft Office Excel
Indicadores que no se cargan 5914.xls	16,5 KiB	Archivo de copia de seguridad de Microsoft Office Excel
log_dim_forma_gasto_20110526.txt	0 bytes	documento de texto sencillo
log_dim_ind_balance_20110526.txt	0 bytes	documento de texto sencillo
log_dim_nae_gasto_20110526.txt	0 bytes	documento de texto sencillo
log_hech_0005_20110526.txt	0 bytes	documento de texto sencillo
log_hech_5910_20110526.txt	0 bytes	documento de texto sencillo
log_hech_5911_20110526.txt	0 bytes	documento de texto sencillo
log_hech_5912_20110526.txt	0 bytes	documento de texto sencillo
log_hech_5913_20110526.txt	0 bytes	documento de texto sencillo
log_hech_5914_20110526.txt	0 bytes	documento de texto sencillo
log_hech_cpe_20110526.txt	0 bytes	documento de texto sencillo
log_hech_isc_20110526.txt	0 bytes	documento de texto sencillo
log_var_maestro_20110526.txt	0 bytes	documento de texto sencillo
Totales de forma de gastos.xls	60,5 KiB	Archivo de copia de seguridad de Microsoft Office Excel
Validación CUI Ind_BEf.xls	13,5 KiB	Archivo de copia de seguridad de Microsoft Office Excel

Figura 11 Archivos de errores

Para el caso 2:

No. Paso	Función a probar	Acción	Resultados esperados
1	Cantidad de registros por tabla de la BD fuente y la BD destino.	Contar la cantidad de registros en BD fuente y BD destino y comparar.	Igual cantidad de registros.
2	Verificar duplicación de datos en las tablas de hechos y dimensiones de la BD destino.	Ejecución de los procesos ETL.	Igual cantidad de registros.

Tabla 19 Funciones a probar del caso 2

Paso 1:

Cantidad de registros por tabla de la BD fuente y la BD destino:

Tablas fuentes	SQL	Cantidad de filas	Tablas destino	SQL	Cantidad de filas
tbm591000p1	select count(*) from mod_modelos .tbm591000p1	1310	hech_estado_situacion	select count(*) from mart_balance. hech_estado_situacion	1310
tbm591100p1	select count(*) from mod_modelos .tbm591100p1	316	hech_estado_resultado	select count(*) from mart_balance. hech_estado_resultado	316
tbm591200p1	select count(*) from mod_modelos .tbm59120	100	hech_inversiones_donaciones	select count(*) from mart_balance. hech_inversiones_donacion	100

				es	
tbm5913 00p1	select count(*) from mod_modelos .tbm59130	570	hech_disponibilidad_ fondo	select count(*) from mart_balance. hech_disponib ilidad_fondo	570
tbm5915 00p1	select count(*) from mod_modelos .tbm00150	259	hech_estado_gastos	select count(*) from mart_balance. hech_estado_ gastos	250

Tabla 20 Cantidad de registros en BD fuente y BD destino

La cantidad de filas de la tabla destino hech_estado_gastos no coincide con la cantidad de filas de la tabla fuente tbm591500p1, ya que esta última contiene dos indicadores que no son necesarios integrarlos al MD. En este caso estos indicadores pertenecen a diferentes entidades por lo que la diferencia de registros entre las tablas es de nueve.

Paso 2:

Tablas destino	Cantidad de registros (antes)	SQL	Cantidad de registros (actual)	Duplicados
dim_forma_gastos	2	select count(*) from mart_balance.dim_forma_gastos	2	0
dim_indicador_balan ce	236	select count(*) from mart_balance.dim_indicador_balan ce	236	0
dim_nae_gastos	711	select count(*) from mart_balance.dim_nae_gastos	711	0
hech_estado_situaci on	1310	select count(*) from mart_balance.hech_estado_situaci	1310	0

		on		
hech_estado_resultado	316	select count(*) from mart_balance.hech_estado_resultado	316	0
hech_inversiones_donaciones	100	select count(*) from mart_balance.hech_inversiones_donaciones	100	0
hech_disponibilidad_fondo	570	select count(*) from mart_balance.hech_disponibilidad_fondo	570	0
hech_estado_gastos	259	select count(*) from mart_balance.hech_estado_gastos	250	0

Tabla 21 Cantidad de registros por tabla en BD destino duplicados.

Pruebas de integración

Las pruebas de integración comprueban el correcto funcionamiento entre los distintos componentes de un software una vez que han sido probados por separados, con el propósito de comprobar que interactúan correctamente.

Para validar que los datos cargados dentro del MD corresponden con los que se visualizan al cliente, se realizaron pruebas de integración entre los procesos de ETL realizados y la capa de visualización del MD Balance Económico Financiero. Se tomó una muestra de los datos cargados en la tabla hech_estado_gastos para compararlos con los datos que se muestran en el reporte: Por División Política Administrativa y entidad los indicadores del estado de gastos. (Figura 12)

real	plan	dim_entidad_id	dim_dpa_id	dim_indicador_balance_id
1713.71	11560	1721	1949	758
428.72	1320	1721	1949	763
74.05	840	1721	1949	764
0	640	1721	1949	765

0	1200	1721	1949	766
---	------	------	------	-----

Tabla 22 Muestra de datos de la tabla hech_estado_gastos

1721: Ministerio de la Agricultura.

1949: La Habana, Plaza de la Revolución.

758: Alimento.

763: Materiales y Artículos de Consumo.

764: Libros y Revistas.

765: Útiles y Herramientas.

766: Partes y Piezas de Repuestos.

Plaza de La Revolución	Total		
De la Asistencia Social			
Materias Primas y Materiales	15.560	2.216,48	
Alimento	11.560	1.713,71	
Libros y Revistas	840	74,05	
Materiales de Cosntrucción			
Materiales para la Enseñanza			
Materiales y Artículos de Consumo	1.320	428,72	
Medicamentos y Materiales afines			
Partes y Piezas de Repuestos	1.200	0	
Vestuario y Lencería			
Útiles y herramientas	640	0	
Combustibles y Lubricantes	3.094,08	449,63	
Energía	15.000	1.848,35	
Gastos de Personal	687.812,51	111.110,99	
Otros Gastos de la Fuerzas de Trabajo	77.035	12.919,34	
Depreciación y Amortización	13.000	2.139,06	
Otros Gastos Monetarios	243.796,49	41.423,52	
Gastos Financieros			
De la Asistencia Social	1.055.298,08	172.107,37	

Figura 12 Reporte: Por División Política Administrativa y entidad los indicadores del estado de gastos

Conclusiones Parciales

Después de analizar los resultados obtenidos en la etapa de implementación y validación se definen como conclusiones:

- Quedó definida la estructura de los datos a partir del modelo de datos físico contando con dos esquemas: dimensiones y mart_balance, asegurando una mejor organización en la BD.
- Se desarrolló el proceso ETL, extrayendo los datos de la BD SIGE, transformados y cargados en el MD Balance económico financiero, logrando un almacenamiento organizado de los datos.
- Obtención de los casos de pruebas y listas de chequeos asegurando la calidad y veracidad de los datos en el MD.

CONCLUSIONES

Al culminar el trabajo de diploma, se le dio cumplimiento a los objetivos específicos planteados dando así solución al problema identificado.

- El estudio del estado del arte de los procesos de integración de datos del Mercado de Datos Balance Económico Financiero permitieron sentar las bases para su posterior desarrollo.
- La modelación e implementación de las estructuras dimensionales requeridas permitieron el manejo de la información de los modelos estadísticos y contables de las Unidades Presupuestadas.
- Los procesos de extracción, transformación y carga de los datos aseguraron la integración de los mismos en el Mercado de Datos Balance Económico Financiero.
- La aplicación de casos de prueba y listas de chequeos permitió validar la solución lográndose resultados satisfactorios.

RECOMENDACIONES

Con el propósito de mejorar la propuesta realizada en este trabajo de diploma, se recomienda:

- Realizar la integración de toda la información del balance financiero de las Unidades Presupuestadas del país.
- Probar los procesos de extracción, transformación y carga implementados en la solución con toda la información de los modelos estadísticos de las Unidades Presupuestadas, para encontrar no conformidades que no se detectaron durante la etapa de prueba realizada.
- Realizar un estudio de técnicas de optimización, que puedan ser aplicadas al proceso de extracción, transformación y carga desarrollado.

REFERENCIAS BIBLIOGRÁFICAS

1. Kimball, R., *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. 2nd Edition. ed. 2002.
2. Inmon, W.H., *Building the Data Warehouse*. 2ª ed. ed. 1996., New York
3. Wolff, C.G. *Modelamiento Multidimensional* 2002 miércoles 28 [cited; Available from: <http://www.inf.udec.cl/revista/edicion4/cwolff.htm>].
4. Chaudhuri, D., *An Overview of Data Warehousing and OLAP Technology*.
5. ETL, A.d.d.-[cited 2010 diciembre 16]; Available from: http://etl-tools.info/en/bi/etl_process.htm.
6. Rubia, J.M., *Introducción a los almacenes de datos* Marzo 2009.
7. [cited 2010 5 diciembre]; Available from: <http://www.1keydata.com/datawarehousing/molap-rolap.html>.
8. *ETL Tools. ETL and Data Warehousing portal*. [cited 2010 diciembre 13]; Available from: <http://www.etltools.org/>.
9. Tactuk, D.P. *Portal de las Estadísticas Dominicanas*. [cited 2011 febrero 25]; Available from: <http://www.one.gob.do/index.php?module=articles&func=display&aid=1377>.
10. Jorge Rubio Navarro, J.M.S. **Volume**,
11. Salazar, R.L. [cited 2011 febrero 25]; Available from: <http://www.inegi.org.mx>.
12. MSc. Ing. Julio Alfredo García Caraballo, D.I.J.A.M.M., *El Proceso de Inteligencia Empresarial en las Empresas del Grupo de Diseño e Ingeniería de la Construcción*. REVISTA DE ARQUITECTURA E INGENIERÍA, agosto 2010. **Vol. No. 4**.
13. Autores, C.d., . *Metodología de la investigación*. . 2003.
14. Sánchez, L.Z.Z., *Metodología para el Diseño Conceptual de Almacenes de Datos*. . 2008, Valencia
15. Angel, M. septiembre 23, 2008, Sicilia.
16. *TortoiseSVN el mejor interfaz para el control de versiones* [cited 2010 Diciembre 7]; Available from: <http://tortoisesvn.net/node/337>.
17. [cited 2010 noviembre 3]; Available from: http://soporte.tiendainux.com/portal/Portfolio/postgresql_ventajas.html.
18. Forero, E.S. *Knol. A unit of knowledge*. Julio 2008 [cited 2011 Enero 21]; Available from: <http://knol.google.com/k/an%C3%A1lisis-y-dise%C3%B1o-de-software#>.
19. Autores, C.d., . *Herramientas Case*.

REFERENCIAS BIBLIOGRÁFICAS

20. Paradigm., V. 2010.
21. [cited 2010 noviembre 15]; Available from: <http://kettle.pentaho.com>.
22. community, p.[cited 2010 diciembre 8]; Available from: <http://wiki.pentaho.com/pages/viewpage.action?pageId=14844841>.
23. [cited 2010 diciembre 10]; Available from: <http://datacleaner.eobjects.org>.
24. [cited 2011 febrero 25]; Available from: <http://msdn.microsoft.com/es-es/library/aa577691%28v=bts.10%29.aspx>.
25. *Tigris.org Open Source Software Engineering Tools* [cited 2010 Diciembre 7]; Available from: <http://tortoisesvn.tigris.org/servlets/NewsItemView?newsItemId=2168>.

BIBLIOGRAFÍA

BIBLIOGRAFÍA

[cited 2011 febrero 25]; Available from: <http://msdn.microsoft.com/es-es/library/aa577691%28v=bts.10%29.aspx>.

[cited 2010 diciembre 10]; Available from: <http://datacleaner.eobjects.org>.

[cited 2010 noviembre 15]; Available from: <http://kettle.pentaho.com>.

community, p.[cited 2010 diciembre 8]; Available from: <http://wiki.pentaho.com/pages/viewpage.action?pageId=14844841>.

ETL Tools. ETL and Data Warehousing portal. [cited 2010 diciembre 13]; Available from: <http://www.etltools.org/>.

ETL, A.d.d.- [cited 2010 diciembre 16]; Available from: http://etl-tools.info/en/bi/etl_process.htm.

Inmon, W.H., *Building the Data Warehouse*. 2ª ed. ed. 1996., New York.

Kimball, R., *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. 2nd Edition. ed. 2002.

MSc. Ing. Julio Alfredo García Caraballo, D.I.J.A.M.M., *El Proceso de Inteligencia Empresarial en las Empresas del Grupo de Diseño e Ingeniería de la Construcción*. REVISTA DE ARQUITECTURA E INGENIERÍA, agosto 2010. **Vol. No. 4**.

pgAdmin PostgreSQL Tools [cited 2010 diciembre 7]; Available from: <http://www.pgadmin.org>.

PostgreSQL Global Development Group. [cited 2010 noviembre 17]; Available from: <http://www.postgresql.org>.

Tamargo, L.L.C., *Diseño Físico del Data Warehouse*, in *La revista del empresario cubano*. 2011.

GLOSARIO DE TÉRMINOS

sql: es un lenguaje declarativo de acceso a bd relacionales que permite especificar diversos tipos de operaciones en éstas.

tcp/ip: modelo de descripción de protocolos de red.

indicadores: magnitud utilizada para medir o comparar los resultados efectivamente obtenidos, en la ejecución de un proyecto, programa o actividad. <http://www.definicion.org/indicador>

modelo: están conformados por un encabezado, pie de firma, indicadores y aspectos que recogen información de las entidades.

estadísticas: ciencia que estudia la recolección, análisis e interpretación de datos.

bases de datos: conjunto de datos pertenecientes a un mismo contexto y almacenados sistemáticamente para su posterior uso.

centros informantes: empresas u organismos que suministran información a las oficinas de estadísticas en sus diferentes niveles.

TH: tabla de hecho.

DIM: dimensión.

ANEXOS

Anexo 1

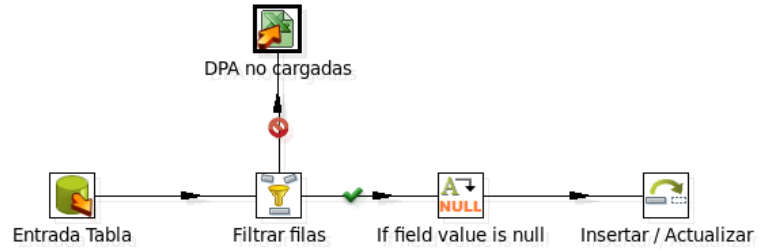


Figura 13 Carga de la tabla dim_dpa

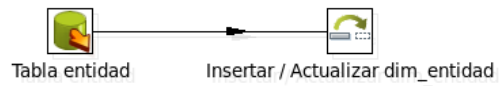


Figura 14 Carga de la tabla dim_entidad

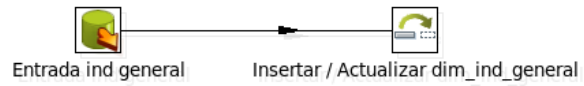


Figura 15 Carga de la tabla dim_indicadores_generales

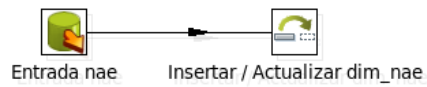


Figura 16 Carga de la tabla dim_nae



Figura 17 Carga de la tabla dim_organismo

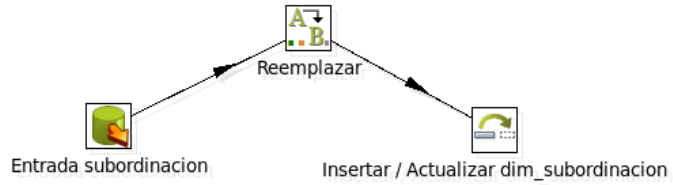


Figura 18 Carga de la tabla dim_subordinación



Figura 19 Carga de la tabla dim_temporal_mes

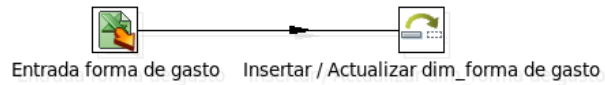


Figura 20 Carga de la tabla dim_forma_gasto

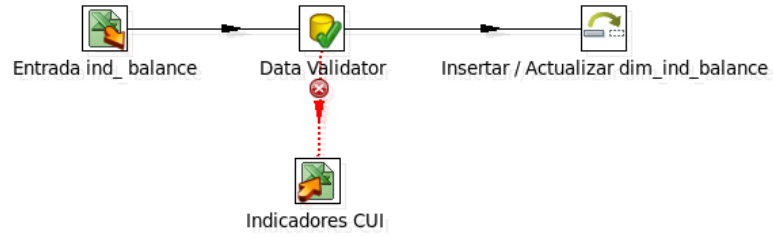


Figura 21 Carga de la tabla dim_indicador_balance

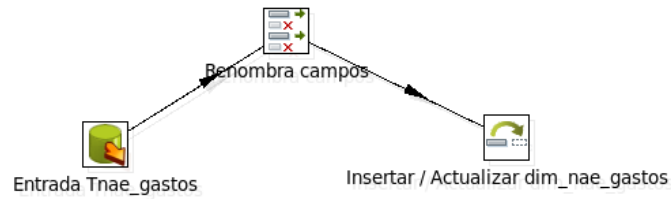


Figura 22 Carga de la tabla dim_nae_gastos

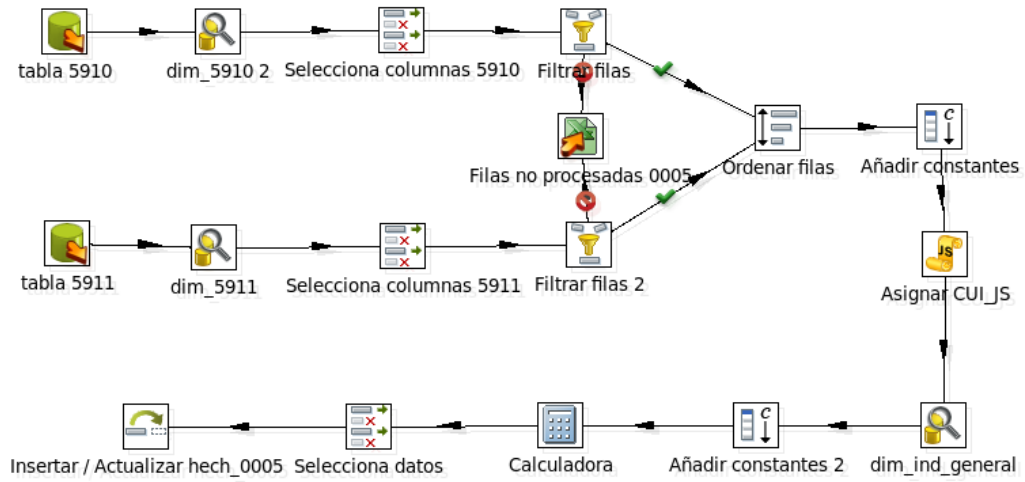


Figura 23 Carga de la tabla hech_estado_situacion

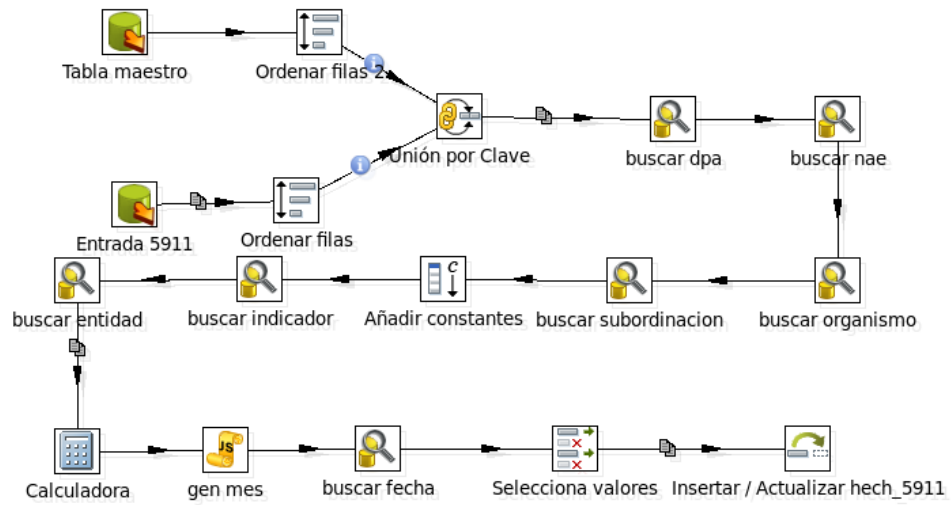


Figura 24 Carga de la tabla hech_estado_resultado

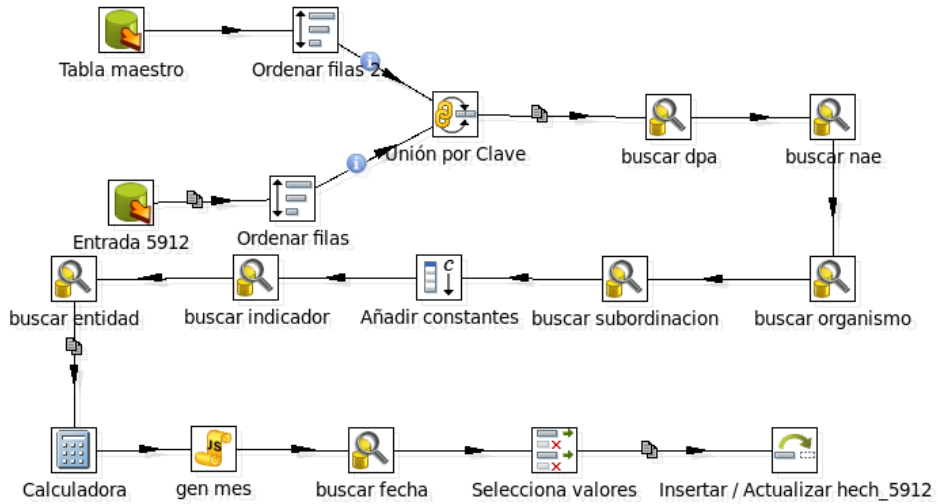


Figura 25 Carga de la tabla hech_inversiones_donaciones

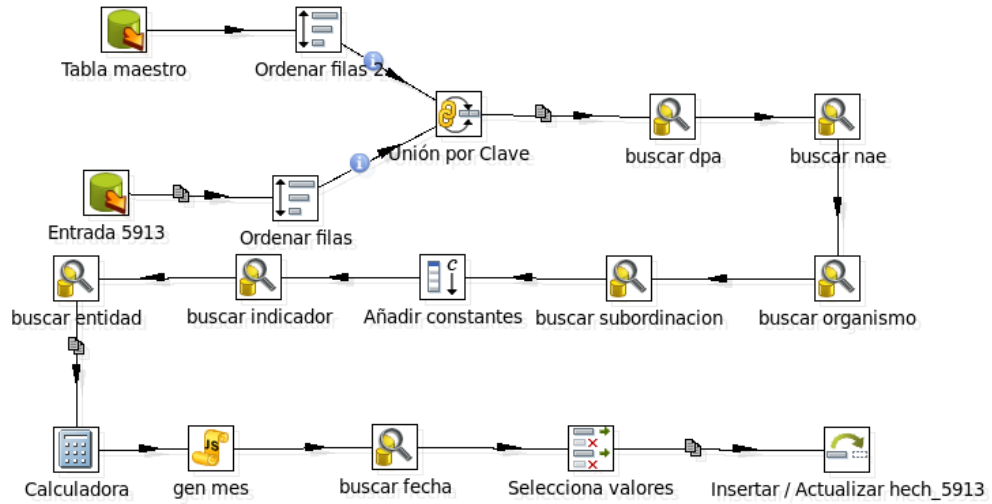


Figura 26 Carga de la tabla hech_disponibilidad_fondos

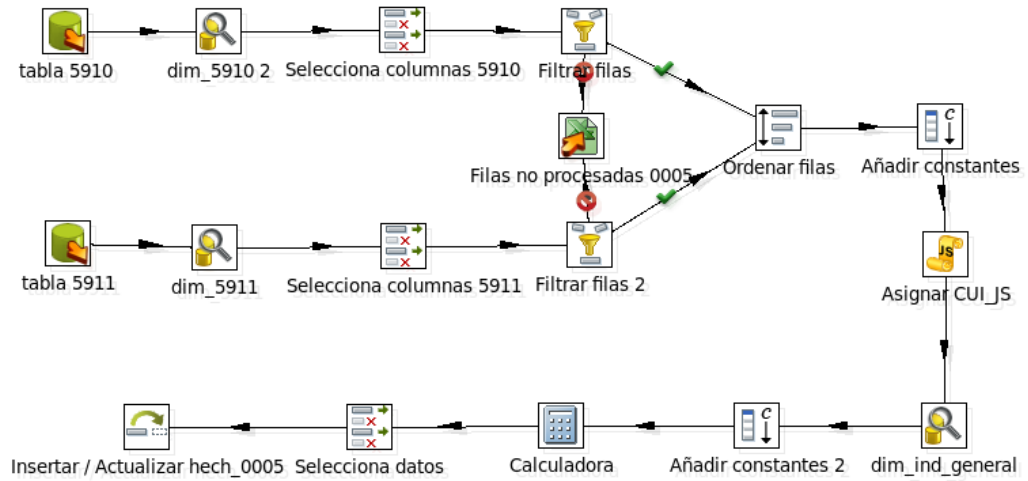


Figura 27 Carga de la tabla hech_indicadores_generales

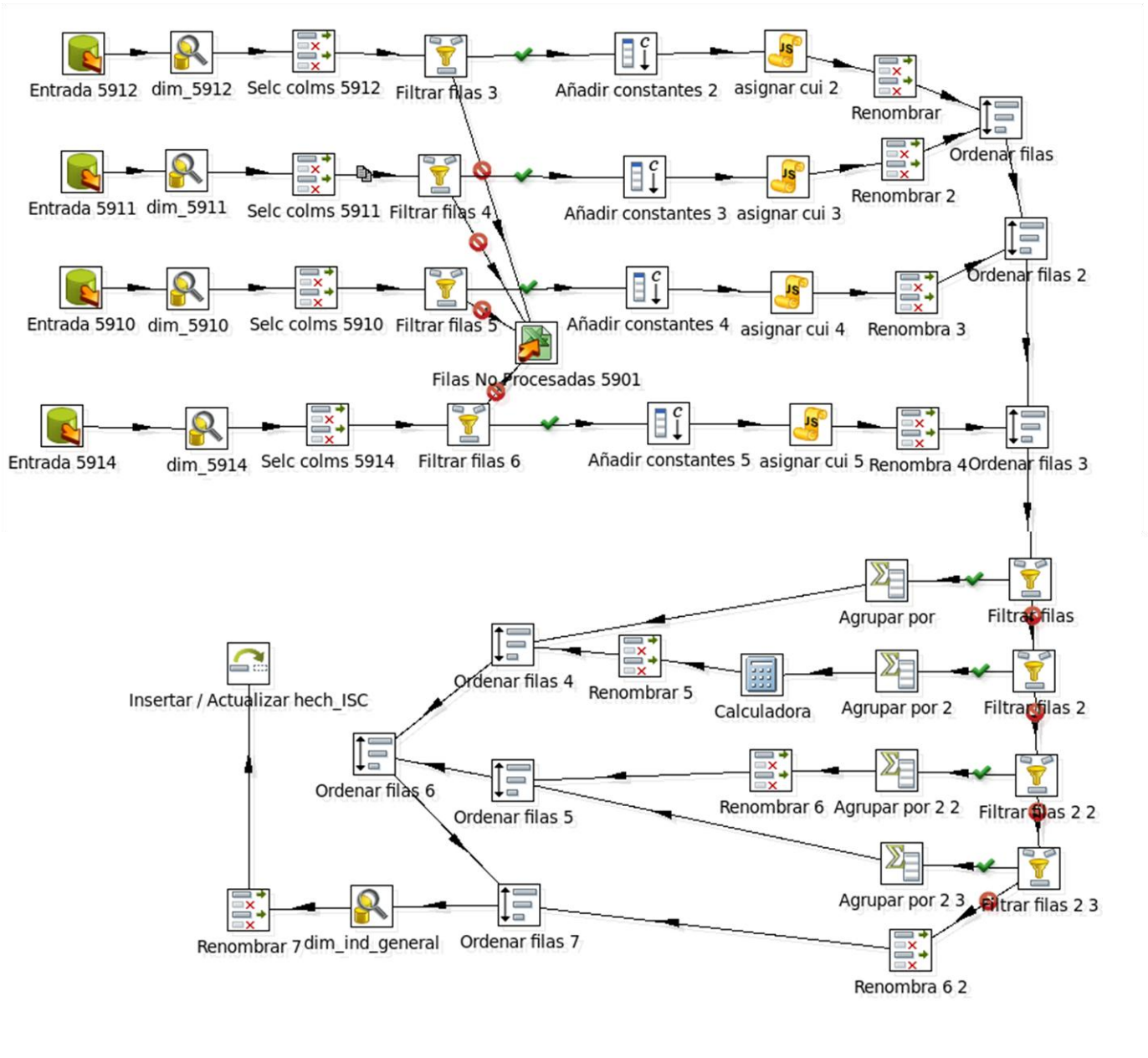


Figura 28 Carga de la tabla hech_indicadores_seleccionados_contabilidad

ANEXOS

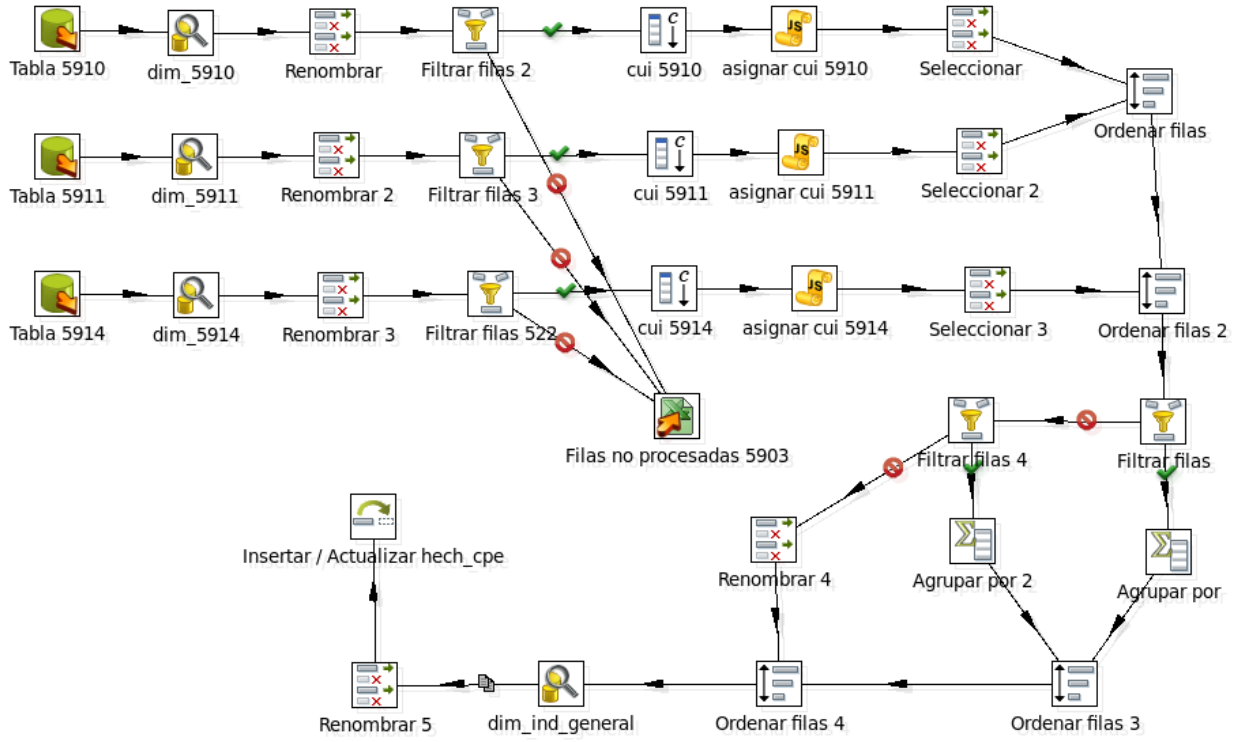


Figura 29 Carga de la tabla hech_cumplimiento_plan_economico