



Universidad de las Ciencias Informáticas

Facultad 3

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

**Mercado de Datos para la gestión de la información
sobre operaciones con tarjetas bancarias en el Banco
Metropolitano.**

Autor: Yadini Pérez López

Tutor: Ing. Lissett Díaz Mesa

La Habana, junio de 2013

“Año 55 de la Revolución”

“En todos los asuntos humanos hay esfuerzos, y hay resultados, y la fortaleza del esfuerzo es la medida del resultado.”

James Allen

Declaración de Autoría

Declaro que soy la única autora de este trabajo y autorizo a la Universidad de las Ciencias Informáticas los derechos patrimoniales del mismo, con carácter exclusivo. Para que así conste, firmo la presente a los ____ días del mes de _____ del año 2013.

Yadini Pérez López
Autor

Ing. Lissett Díaz Mesa
Tutor

Datos de contacto

Autor

Yadini Pérez López

Correo electrónico: yperezl@estudiantes.uci.cu

Universidad de las Ciencias Informáticas, La Habana, Cuba

Tutor

Ing. Lissett Díaz Mesa

Correo electrónico: ldiazm@uci.cu

Universidad de las Ciencias Informáticas, La Habana, Cuba

Agradecimientos

A todas las personas que me han brindado su ayuda durante mis estudios.

A mis padres por su apoyo incondicional, dedicación, sacrificio... Por ser partícipes de cada momento de mi vida con la mayor entrega.

A mi madre por ser guía en mi camino, por ser la razón de mis esfuerzos, por ser el mayor de mis orgullos.

A Luis Miguel Rojas Aguilera por ser compañero, amigo, soporte en todo momento, por ser un gran regalo de la vida.

A todos mis compañeros por su afecto y preocupación, en especial a: Dairis Almaguer Pérez, Aylén Jomarrón Jomarrón, Diana Rodríguez Malagón, Karel Riverón Escobar y Jorge Fonseca Rodríguez.

A los profesores: Yanisbel González Hernández, Asnioby Hernández López, Joisel Pérez Pérez, José Antonio Sánchez Imbert y Annia Arencibia Morales, por la gran ayuda prestada en la realización de este trabajo.

A los miembros del tribunal que con sus revisiones han logrado que el presente trabajo tenga la calidad requerida.

A mi tutora Lissett Díaz Mesa que me brindó la oportunidad de crecer con la realización de este trabajo; y por estar presente en el desarrollo del mismo hasta el final.

Dedicatoria

El presente trabajo lo dedico a las personas más importantes en mi vida, ellas son el motivo de mis esfuerzos sin límites:

A mis queridos padres.

A Luis Miguel Rojas Aguilera.

Resumen

De los procesos que se llevan a cabo en el Banco Metropolitano de Cuba, las operaciones con tarjetas bancarias es uno de los que mayor atención requiere para evitar casos de fraude financiero mediante el empleo de las mismas. Actualmente en dicha institución se trabaja en la detección de este tipo de delitos mediante el estudio de las operaciones que realizan los clientes, sin embargo, ocasionalmente no es posible detectar a tiempo los presuntos casos de fraude, ya que no existe una estructura de datos que contenga la información a procesar de forma estandarizada y resumida, ni se aplican técnicas sobre los datos que automaticen y hagan más exacto el proceso de detección.

Con el objetivo de solucionar las deficiencias detectadas se decide construir un Mercado de Datos (MD) que contenga información preparada para facilitar la entrada de datos para el proceso de detección de fraude financiero con tarjetas bancarias. Para esto se realizó un estudio sobre Almacenes de Datos, se seleccionaron las herramientas para la construcción del MD, se siguieron los pasos que propone la metodología de la UCID en su primera versión, obteniendo el modelo conceptual y modelo de datos del MD, entre otros artefactos, además se realizaron pruebas para verificar la efectividad de la solución.

El MD desarrollado contiene información estructurada de forma tal que será posible llevar a cabo técnicas de Minería de Datos para la detección de posibles casos de fraude. Los especialistas obtendrán valores que les permitirán tomar decisiones en tiempo.

Palabras Claves:

Fraude financiero, mercado de datos, tarjetas bancarias.

Tabla de Contenidos

Introducción	1
Capítulo 1 Fundamentación Teórica	6
1. Introducción	6
2. Inteligencia de negocios.....	6
2.1 Almacén de Datos	7
2.2 Mercado de Datos	9
2.3 Elementos que integran un Mercado de Datos	9
2.4 Función de un Mercado de Datos.....	10
2.5 Arquitectura de un Mercado de Datos	12
2.5.1 Definiendo una arquitectura de flujo de datos para la solución.....	15
2.5.2 Modelado de una Base de Datos Multidimensional.....	15
2.5.3 Modos de almacenamiento de los datos	18
2.5.3.1 Definición de los modos de almacenamiento a utilizar.....	19
2.6 Ventajas e inconvenientes de los almacenes de datos	19
3. Almacenes de datos en el mundo bancario.....	20
3.1 Almacenes de datos en Cuba.....	21
4. Metodologías para el desarrollo de un Almacén de Datos.....	22
4.1 Metodología de Kimball	22
4.2 Metodología Hefesto	23
4.3 Metodología de la UCID versión 1.0 para el desarrollo de almacenes de datos	25
4.3.1 Metodología propuesta para la solución.....	26
5. Herramientas a utilizar en la construcción del Mercado de Datos	26
5.1 Herramienta de Modelado	26
5.2 Sistema de Gestión de bases de datos	27
5.3 Herramientas para el proceso ETL.....	27
5.4 Herramientas para la validación del MD	28
5.4.1 Herramienta para la realización de las pruebas de funcionalidad.....	28
5.4.2 Herramientas para la realización de pruebas de rendimiento y estrés.....	28

6. Conclusiones del capítulo	28
Capítulo 2 Análisis y diseño del Mercado de Datos	30
1. Introducción	30
2. Procedimiento para el desarrollo de la solución utilizando la metodología escogida	30
2.1 Fase # 1. Análisis de los requerimientos	30
2.1.1 Planificar las entrevistas	30
2.1.2 Identificar preguntas, perspectivas e indicadores	32
2.1.3 Modelo Conceptual	35
2.1.4 Aprobación del modelo conceptual por el cliente	37
2.2 Fase # 2. Análisis de los OLTP.....	37
2.2.1 Definir el estado general de los sistemas fuentes	38
2.2.2 Determinación de indicadores	38
2.2.3 Establecer correspondencias	39
2.2.4 Nivel de granularidad	41
2.2.5 Modelo conceptual ampliado.....	41
2.3 Fase # 3. Modelo lógico del MD	43
2.3.1 Definir el tipo de modelo del MD	43
2.3.2 Identificar dimensiones	43
2.3.3 Identificar hechos.....	45
2.3.4 Diseñar tablas y columnas físicas	46
3. Conclusiones del capítulo	47
Capítulo 3 Implementación y validación del Mercado de Datos.....	48
1. Introducción	48
2. Fase # 4 Procesos ETL.....	48
2.1 Mapeo de los datos fuente al destino	48
2.2 Establecer condiciones adicionales y restricciones.....	49
2.3 Cargas incrementales de datos	50
2.4 Automatización del proceso ETL	54
3. Validación del Mercado de Datos.....	54

3.1 Pruebas para validar la calidad del Mercado de Datos	55
3.1.1 Pruebas de funcionalidad.....	55
3.1.2 Pruebas de calidad de datos.....	56
3.1.3 Pruebas de confiabilidad de los datos	57
3.2 Pruebas para validar el rendimiento del Mercado de Datos	59
3.2.1 Pruebas de rendimiento	59
3.2.2 Pruebas de estrés.....	60
4. Conclusiones del capítulo	61
Conclusiones generales.....	62
Recomendaciones	63
Referencias Bibliográficas.....	64
Bibliografía.....	66
Anexos.....	68

Introducción

Es considerable el enorme cúmulo de datos que se genera en la actualidad, mayormente en el campo empresarial, donde las corporaciones han descubierto cuán importante y beneficioso es acumular información sobre los procesos que las mismas llevan a cabo diariamente; y luego analizarla e interpretarla, con el fin de mejorar el rendimiento de los negocios.

El proceso de almacenamiento de información relevante para su posterior análisis recibe el nombre de Inteligencia de Negocio (IN) o Inteligencia Empresarial (IE) (1). En los últimos años han surgido múltiples técnicas para llevar a cabo esta valiosa idea, pero el centro de atención, sin lugar a dudas, lo constituyen los Almacenes de Datos (del inglés Data Warehouse (DW)).

Los Almacenes de Datos (AD) brindan la vía más eficiente para estandarizar toda la información que genera una empresa y guardar la de mayor relevancia en una única fuente, para el posterior análisis e interpretación de los datos. Los AD están conformados por Mercados de Datos (del inglés Data Mart (DM)), que no son más que pequeños almacenes que procesan los datos de un área determinada en una entidad. (2)

El empleo de la IN proporciona un sin número de facilidades a las empresas en la actualidad, no solo para organizar y controlar todos sus procesos gerenciales o estrechar las relaciones con los clientes, sino que también constituye una forma de detectar actividades inusuales y dañinas para las entidades; siendo la más frecuente y peligrosa, lo que se conoce hoy como Fraude Financiero: Estafa destinada a realizar un beneficio financiero por medios ilegales en perjuicio de otros. (3)

Los servicios bancarios no están exentos al efecto de estos delitos, los fraudes bancarios ya no se realizan a punta de pistola, como en el viejo oeste norteamericano, sino de muchas otras maneras, frecuentemente con apoyo tecnológico; el uso fraudulento de tarjetas bancarias es una de las formas más frecuente. Con el fin de detectar y mitigar estos delitos, que pueden significar grandes pérdidas monetarias y de clientes para las empresas, se aplican técnicas y conceptos como los anteriormente abordados.

En Cuba múltiples organizaciones implementan con éxito el uso de la IN y los AD. El Banco Central de Cuba ha detectado también la necesidad de incorporar estas herramientas en las diferentes entidades bancarias existentes en el país, con el objetivo de llevar un control estricto de todos sus procesos de negocio, dada la relevancia y el nivel de confiabilidad que requieren dichos procesos, el Banco Metropolitano de Cuba es una de esas entidades; fue creado en 1997 con el objetivo fundamental de prestar servicios a las embajadas y entidades extranjeras radicadas en el

territorio nacional, así como a sus funcionarios, empleados y población en general. Los principales servicios que brinda son: (4)

- Cuentas corrientes y de ahorro en moneda nacional y en moneda libremente convertible.
- Certificado de depósitos, y depósitos a plazo fijo en moneda libremente convertible.
- Pago y venta de cheques de viajeros y tarjetas de créditos nacionales e internacionales.
- Servicio de cajeros automáticos las 24 horas en tres de sus oficinas.
- Pago directo de facturas domésticas de los clientes desde sus cuentas.
- Cartas de crédito, cobranzas y descuentos de letras, alquiler de cajas de seguridad y otras operaciones bancarias relacionadas con sus clientes que le permitan las leyes. (4)

Actualmente en el departamento de Solución y Prevención de Fraudes de esta institución, se realiza un trabajo de análisis encaminado a la detección de sucesos inusuales y comportamientos poco comunes en los usuarios al ejecutar operaciones que involucran tarjetas bancarias, generando como resultado reportes diarios de estas actividades. (5) Dichos reportes se obtienen a través de consultas al sistema contable SABIC¹, donde los datos relevantes para el estudio en cuestión, coexisten con un considerable número de datos que no son importantes en este sentido. Al no existir una estructura de datos resumida y preparada para el procesamiento de la información, el análisis se hace sumamente engorroso, debido también, a que el mismo se realiza manualmente por funcionarios del departamento, los cuales deciden, rigiéndose por estándares predefinidos, si se está o no, en presencia de un caso de fraude.

Todo esto conlleva a que el proceso de análisis de las operaciones que realizan los clientes, no contemple todos los elementos necesarios para detectar a tiempo, en muchas ocasiones, los presuntos casos de fraude. El proceso se realiza de forma superficial al no existir un lugar donde se concentre la información de importancia, ni técnicas que permitan explotar al máximo la misma. No existe precedente alguno en el banco del trabajo con herramientas que hagan más rápido, automatizado y fácil el proceso de gestión de la información como los AD.

Dada la situación presente, como parte de los objetivos de trabajo del Sistema Bancario Cubano para el año 2013, se determina la necesidad de poner en práctica técnicas y herramientas que contribuyan a la detección de fraude financiero con tarjetas bancarias en el Banco Metropolitano.

Teniendo en cuenta lo anteriormente expuesto, se identifica como **problema a resolver**: ¿Cómo estandarizar la información referente a las operaciones con tarjetas bancarias en el Banco

¹ **SABIC**: Sistema Automatizado para Banca Internacional de Comercio. Es el sistema contable en tiempo real que define el Banco Central de Cuba, como sistema base a utilizar para los diferentes bancos del país. (5)

Metropolitano, de manera tal que facilite la entrada de datos para el proceso de detección de fraude financiero?

El problema planteado delimita el **objeto de estudio**: Procesos de estandarización y almacenamiento de información.

Se identifica como **campo de acción**: Estandarización y almacenamiento de datos sobre las operaciones con tarjetas bancarias en el Banco Metropolitano.

Para solucionar el problema planteado, se identifica como **objetivo general** de la investigación: Desarrollar un Mercado de Datos (MD) que contenga la información relevante sobre las operaciones con tarjetas bancarias en el Banco Metropolitano, facilitando la entrada de datos para el proceso de detección de fraude financiero.

En correspondencia con el objetivo general se definen los siguientes **objetivos específicos**:

1. Analizar el estado del arte referente a los AD relacionados con el objeto de estudio de la investigación, así como las herramientas y metodologías posibles a utilizar en la construcción del MD.
2. Analizar los procesos del Banco Metropolitano que involucran tarjetas bancarias con el fin de identificar las necesidades de información del cliente.
3. Realizar el diseño e implementación del MD, validando que el mismo responda a las necesidades de información del cliente.
4. Realizar pruebas al MD para validar la solución desarrollada.

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

1. Análisis de los AD existentes a nivel internacional y nacional relacionados con la estandarización y almacenamiento de información bancaria, estableciendo similitudes con la investigación en curso.
2. Caracterización de las metodologías y herramientas posibles a utilizar en la construcción del MD, conociendo así las diferentes opciones que existen en este sentido.
3. Selección de la metodología y las herramientas necesarias para la construcción del MD.
4. Identificación de las preguntas que incluyan variables de análisis que se consideren relevantes, permitiendo el estudio de la información desde diferentes perspectivas.

5. Identificación de los indicadores y perspectivas de análisis, para una mejor comprensión de los aspectos a medir.
6. Realización del Modelo Conceptual del MD, obteniendo una representación de la relación existente entre perspectivas e indicadores.
7. Selección de los campos que integrarán cada perspectiva, definiendo así el nivel de granularidad del MD.
8. Establecimiento de las correspondencias entre los campos seleccionados y el sistema fuente, garantizando la existencia de la información que se quiere obtener en la fuente de datos.
9. Diseño de las tablas de dimensiones y hechos, logrando así una visión de los elementos físicos que formarán el MD.
10. Realización de las uniones entre tablas para la obtención del modelo físico del MD.
11. Realización de los procesos de: Extracción, Transformación y Carga, limpieza de datos y sentencias SQL para el llenado del MD.
12. Validación del MD para comprobar el correcto desempeño del mismo.

Se define como **idea a defender**: El desarrollo de un MD para el Banco Metropolitano, permitirá la estandarización de la información relevante sobre las operaciones con tarjetas bancarias; facilitando la entrada de datos para el proceso de detección de fraude financiero.

Entre los **métodos científicos** utilizados se destacan:

Como **métodos teóricos**: el **Analítico-Sintético** que se utilizó en la revisión de documentos, libros, artículos e informes para la extracción de elementos importantes que están relacionados con la investigación, con el objetivo de lograr una mayor visión sobre el tema, permitiendo tener un enfoque global del mismo. Se aplicó además el método **Inductivo – Deductivo** que se utilizó en la profundización de los diferentes aspectos relacionados con el tema de investigación; el método **Histórico-Lógico** que permitió conocer cómo se realiza actualmente el proceso de detección de fraude con tarjetas bancarias en el Banco Metropolitano, con el fin de capturar los requisitos de dicho proceso a tener en cuenta en la solución; y el método de **Modelación**, para realizar el diseño del modelo conceptual, modelo conceptual ampliado y modelo de datos del MD.

Se emplearon **métodos empíricos** como: **Entrevistas estructuradas** y **Observaciones** para la recopilación de información sobre las particularidades de los diferentes procesos del negocio a tener en cuenta en la solución.

Aportes de la investigación:

El **valor práctico** de la investigación está dado por:

- Obtención de la relación de indicadores y sus reglas de cálculo para definir, mediante estas medidas, comportamientos en los clientes del banco al operar con tarjetas bancarias.
- Obtención del diseño conceptual de los hechos y dimensiones del MD, pensado para la posterior aplicación de técnicas de minería de datos.
- Obtención de la implementación del MD.

Como resultado se espera contar con un MD que contenga, de forma estandarizada y con la calidad requerida, toda la información relevante referente a las operaciones que se realizan en el Banco Metropolitano con tarjetas bancarias; esto facilitará la entrada de los datos para el proceso de detección de fraude financiero para la posterior aplicación de técnicas de minería de datos.

El presente trabajo consta de 3 capítulos estructurados de la siguiente manera:

Capítulo 1: Fundamentación Teórica. Se abordan los principales conceptos relacionados con los AD, así como una descripción de todas las herramientas, tecnologías y metodologías a utilizar para dar solución al problema planteado.

Capítulo 2: Análisis y diseño del Mercado de Datos. Se describen los pasos de la metodología utilizada con el objetivo de facilitar la construcción del MD. En principio se identifican las necesidades de información del cliente para el diseño de un modelo conceptual del MD, luego se confecciona el modelo conceptual ampliado del MD y se concluye con la obtención del modelo de datos del MD.

Capítulo 3: Implementación y validación del Mercado de Datos. Se describe el proceso de implementación del MD, abordando los procesos de Extracción, Transformación y Carga (ETL) mediante la utilización de las herramientas descritas en el capítulo primero. También se realiza la validación del MD, a partir de la construcción de cubos de Procesamiento Analítico en Línea, que muestran el correcto funcionamiento de la solución y se realizan pruebas de rendimiento y estrés.

Capítulo 1 Fundamentación Teórica

1. Introducción

El creciente interés que experimentan las grandes empresas por el uso de sistemas que permitan el análisis de la información que las mismas generan y apoyen la toma de decisiones, ha dado paso al surgimiento de nuevas vías para lograr esta meta. Un MD constituye una de estas vías.

En este capítulo se realiza un estudio de los principales conceptos que giran alrededor de un MD, su definición, las ventajas que puede proporcionar su implantación y también los inconvenientes que pueden surgir al utilizarlo. Se expone la tendencia nacional e internacional del uso de los AD y MD en empresas con determinadas características. Se analizan las herramientas que pueden ser utilizadas para la construcción del MD, así como metodologías que garanticen el cumplimiento de los objetivos trazados.

2. Inteligencia de negocios

La IN no es más que el conjunto de estrategias y herramientas dirigidas a la administración y creación de conocimiento mediante el análisis de los datos de una organización, con el fin de obtener una ventaja competitiva. Este término abarca la comprensión del funcionamiento actual de la institución, así como la anticipación a eventos futuros, con el objetivo de apoyar la toma de decisiones. (1)

Se puede decir que la IN como concepto integra por un lado, el almacenamiento, y por el otro, el procesamiento de grandes cantidades de datos; con el principal objetivo de transformarlos en conocimiento y en decisiones que se toman en tiempo real a través de un proceso de análisis y exploración. (1)

La definición anteriormente expuesta puede representarse a través de la fórmula que propone la figura # 1:

$$\text{Datos} + \text{Análisis} = \text{Conocimiento}$$

Figura # 1 Inteligencia de Negocio (1)

Los datos son observaciones, hechos o imágenes que una vez formalizados, contextualizados, filtrados y resumidos, constituyen información. Dicha información, enriquecida por ideas,

procedimientos y reglas, se convierte en conocimiento, que permite realizar acciones y tomar decisiones.

Los datos se pueden encontrar en cualquier lugar en el mundo y el conocimiento está localizado en agentes (personas, organizaciones, etc.); mientras que la información adopta un papel mediador entre ambos conceptos.

La IN proporciona múltiples beneficios a las empresas:

- Reduce el tiempo mínimo que se requiere para recoger toda la información relevante del negocio, ya que la misma se encontrará integrada en una fuente única de fácil acceso.
- Permite a los usuarios no depender de reportes o informes programados, porque los mismos serán generados de manera dinámica.
- Posibilita la formulación y respuesta de preguntas que son claves para el desempeño de la empresa.
- Permite acceder y analizar directamente los indicadores de éxito.
- Se pueden detectar situaciones fuera de lo normal.
- Se pueden encontrar y/o descubrir cuáles son los factores que maximizarán el beneficio.
- Permite predecir el comportamiento futuro con un alto porcentaje de certeza, basado en el entendimiento del pasado.
- El usuario puede consultar y analizar los datos de manera sencilla. (1)

Para llevar a cabo la IN es necesario gestionar datos guardados, en ocasiones, en diversos formatos y fuentes, para luego depurarlos e integrarlos, y almacenarlos en un solo destino, depósito o base de datos, lo que permitirá su posterior análisis. Es de vital importancia contar con una herramienta que satisfaga todas estas necesidades.

Actualmente la herramienta más utilizada para lograr estos fines es un AD, que básicamente se encarga de consolidar, integrar y centralizar los datos que la empresa genera en todos los ámbitos de una actividad de negocios (Compras, Ventas, Producción, etc.), para luego ser almacenados mediante una estructura que permite el acceso y exploración de la información requerida, facilitando posteriormente, una amplia gama de posibilidades de análisis de múltiples variables, que permitirá la toma de decisiones estratégicas y tácticas.

2.1 Almacén de Datos

Definición de Bill Inmon

Una de las definiciones más famosas sobre AD, es la de W. H. Inmon, quien define:

“Un Data Warehouse es una colección de datos orientada al negocio, integrada, variante en el

tiempo y no volátil para el soporte del proceso de toma de decisiones de la gerencia.” (6)

- Orientado a temas: Los datos en la base de datos 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 solo lectura, y se mantiene para futuras consultas.
- Integrado: La base de datos contiene los datos de todos los sistemas operacionales² de la organización, y dichos datos deben ser consistentes. (6)

Inmon defiende una metodología descendente (top-down)³ a la hora de diseñar un AD, ya que de esta forma se considerarán mejor todos los datos corporativos. En esta metodología los MD se crearán después de haber terminado el AD completo de la organización. En el epígrafe 1.2.2 se explica qué es un MD.

Definición de Ralph Kimball

Ralph Kimball es otro conocido autor en el tema de los AD, él define un AD como:

"Una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis". (2)

También fue Kimball quien determinó que un AD no era más que:

"La unión de todos los Mercados de Datos de una entidad." (2)

Kimball defiende una metodología ascendente (bottom-up)⁴ a la hora de diseñar un AD, donde primero se construyen los diferentes MD y la integración de los mismos forman el AD de la empresa.

Una definición más amplia de almacén de datos

Las definiciones anteriores se centran en los datos en sí mismos, sin embargo, los medios para extraerlos, transformarlos y cargarlos, así como las diferentes formas para realizar la gestión de datos, incluyendo las herramientas para la IN, son componentes esenciales de un AD. Por lo tanto, para la realización de este trabajo, se tendrá en cuenta este criterio y se seguirá además, el

² **Sistemas operacionales:** Sistemas que realizan operaciones transaccionales de inserción, modificación y eliminación de datos. Los OLTP (Procesos de Transacciones en Línea) son sistemas operacionales. (1)

³ **Top-down:** Estrategia de procesamiento de información donde se formula un resumen del sistema, sin especificar detalles. Luego cada parte del sistema se refina diseñando con mayor detalle. (1)

⁴ **Bottom-up:** Estrategia de procesamiento de información donde las partes individuales se diseñan con detalle y luego se enlazan para formar componentes más grandes. (1)

enfoque bottom-up que propone Ralph Kimball para la construcción de un AD, ya que dicho trabajo consiste en la construcción de un MD que contenga información sobre las operaciones que involucran solo tarjetas bancarias en el Banco Metropolitano, y en futuros proyectos se continuará con el desarrollo de otros MD que contengan información relevante de otras áreas o procesos del banco, para luego integrar dichos mercados y obtener el AD de la institución.

2.2 Mercado de Datos

Un MD es una versión especial de un AD con el objetivo de responder a un determinado análisis, función o necesidad y con una población de usuarios específica, ayudando potencialmente a las empresas y entidades a la toma de decisiones. La principal diferencia entre un AD y un MD es su alcance. El MD está pensado para cubrir las necesidades de un grupo de trabajo o de un determinado departamento dentro de la organización, es el almacén natural para los datos departamentales. En cambio, el ámbito del AD es la organización en su conjunto. En síntesis, los MD son pequeños AD centrados en un tema o un área de negocio específico dentro de una organización. (1)

La idea de lo que es un AD o un MD difiere mayormente en la dimensión de la información que contengan. Ambos pueden compartir el mismo proceso de construcción, la arquitectura, las funciones que realizan. Por esta razón, a partir de este punto del documento, se hablará indistintamente de AD o MD.

2.3 Elementos que integran un Mercado de Datos

Dimensiones

Las dimensiones son utilizadas para indexar, ordenar, agrupar o abreviar el significado de un cubo. Las dimensiones poseen una granularidad⁵, tomando como valores un conjunto de elementos; ejemplos de dimensiones podrían ser: “productos”, “localidades”, “el tiempo” (medido en días, horas, semanas, etc.). Son las perspectivas de análisis en los cubos. Son catálogos de información complementaria necesaria para la presentación de los datos a los usuarios. Es decir, la información general complementaria a cada uno de los registros de la tabla de hechos. (6)

Hechos

Los hechos son todas aquellas sumalizaciones o acumulaciones preestablecidas que residen en una tabla de hechos para agilizar las consultas y permitir que los datos puedan ser accedidos por las diferentes dimensiones, y desde luego, explorados por ellas. Las sumalizaciones no están

⁵ **Granularidad:** Representa el nivel de detalle al que se desea almacenar la información sobre el negocio que se esté analizando. Mientras mayor sea el nivel de detalle de los datos, se tendrán mayores posibilidades de análisis. (1)

referidas solo a sumas, sino también a promedios, mínimos, máximos, totales por sector, porcentajes, fórmulas predefinidas, etc. dependiendo de los requerimientos de información del negocio. (7)

Cubos de Información

Los cubos de información o cubos OLAP⁶ funcionan como los cubos de rompecabezas en los juegos, en el juego se trata de armar los colores y en el MD se trata de organizar los datos por tablas o relaciones; los primeros (en el juego) tienen tres dimensiones, los cubos OLAP tienen un número indefinido de dimensiones, razón por la cual también reciben el nombre de hipercubos. Un cubo OLAP contendrá datos de una determinada variable que se desea analizar, proporcionando una vista lógica de los datos provistos por el sistema de información hacia el MD, esta vista estará dispuesta según unas dimensiones y podrá contener información calculada. El análisis de los datos está basado en las dimensiones del hipercubo, por lo tanto, se trata de un análisis multidimensional. (6)

A la información de un cubo puede acceder el especialista mediante "tablas dinámicas" en una hoja de cálculo o a través de programas personalizados. Las tablas dinámicas le permiten manipular las vistas (cruces, filtrados, organización, totales) de la información con mucha facilidad. Las diferentes operaciones que se pueden realizar con cubos de información se producen con mucha rapidez. Llevando estos conceptos a un AD, éste es una colección de datos que está formada por «dimensiones» y «variables», entendiendo como dimensiones a aquellos elementos que participan en el análisis y como variables a los valores que se desean analizar. (6)

2.4 Función de un Mercado de Datos

Un MD debe entregar la información correcta al personal correspondiente en el momento indicado y en el formato adecuado y brindar respuestas a las necesidades de los expertos. Los usuarios finales pueden hacer fácilmente consultas sobre sus MD sin tocar o afectar la operación del sistema.

En el funcionamiento de un MD son muy importantes las siguientes ideas:

- Integración de los datos provenientes de bases de datos en constante cambio distribuidas por las diferentes unidades de la organización (fuentes heterogéneas). Se debe facilitar una descripción global y un análisis comprensivo de toda la organización en el AD.
- Separación de los datos usados en operaciones diarias, de los usados en el MD para los

⁶ **OLAP (Procesamiento Analítico en Línea):** Es una solución utilizada en el campo de la IN, la cual consiste en consultas a estructuras multidimensionales que contienen datos resumidos de grandes Bases de Datos o Sistemas Transaccionales. (5)

propósitos de divulgación, de ayuda en la toma de decisiones, para el análisis o para operaciones de control. Ambos tipos de datos no deben coincidir en la misma base de datos, ya que obedecen a objetivos muy distintos y podrían entorpecerse entre sí.

Funciones ETL (extracción, transformación y carga)

El proceso de Extracción, Transformación y Carga (ETL por sus siglas en inglés) constituye el cimiento del AD. El sistema ETL hace o destruye el almacén. A pesar de ser algo oculto para el usuario final, consume el 70 % de los recursos necesarios para la implementación y mantenimiento de un AD. (8)

Como su nombre indica el proceso ETL consta de tres etapas principales:

Extracción

La extracción de datos es el proceso que se encarga de recuperar los datos de los sistemas fuentes, éstos pudieran no ser una bases de datos, usualmente se requiere extraer datos de ficheros, una cola, un servicio o del correo electrónico.

En ocasiones el proceso de extracción solo puede realizarse en determinados instantes de tiempo, ya que como se accede directamente a las fuentes, el proceso pudiera interferir con el trabajo diario sobre las mismas. Es por esto que se recomienda actualizar el AD en horarios no perjudiciales en este sentido. (8)

Transformación

Cuando se mueven los datos desde el sistema fuente hasta el AD se requiere hacer varias transformaciones en los datos para que éstos sean más prácticos a la hora de su utilización, esto brinda la posibilidad de estandarizar y reorganizar los datos según sea necesario. Las transformaciones más comunes son:

- Seleccionar solo ciertas columnas de una tabla para su carga (por ejemplo, que las columnas con valores nulos no se carguen).
- Traducir códigos (por ejemplo, si la fuente almacena una "H" para Hombre y "M" para Mujer pero el destino tiene que guardar "1" para Hombre y "2" para Mujer).
- Obtener nuevos valores calculados (por ejemplo, $total_venta = cantidad * precio$).
- Dividir una columna en varias (por ejemplo, columna "Nombre: García, Miguel"; pasar a dos columnas "Nombre: Miguel" y "Apellido: García"). (8)

Carga de datos

La fase de carga es el momento en el cual los datos de la fase anterior (transformación) son cargados en el destino.

La fase de carga interactúa directamente con la base de datos destino. Al realizar esta operación se aplicarán todas las restricciones y disparadores que se hayan definido. Los mismos contribuyen a que se garantice la calidad de los datos en el proceso ETL. (8)

El sistema ETL concede un valor añadido significativo a los datos. Va más allá de transportar los datos de una fuente hasta el MD. Concretamente el sistema ETL:

- Elimina los errores y corrige la ausencia de datos.
- Provee una medida documentada de la certidumbre de los datos.
- Ajusta datos de diversas fuentes para que puedan ser utilizados juntos.
- Estructura los datos para que puedan ser usados por las herramientas de cara al usuario final. (8)

En el capítulo tercero se abordará el proceso ETL llevado a cabo en la construcción del MD para el Banco Metropolitano.

2.5 Arquitectura de un Mercado de Datos

Una de las razones por las que el desarrollo de un AD crece rápidamente es que realmente es una tecnología muy entendible. Teniendo en cuenta las características generales de un AD se definen los componentes que intervienen en su arquitectura o ambiente, la misma está formada por diversos elementos que interactúan entre sí y que cumplen una función específica dentro del sistema. Básicamente la forma de operar se resume de la siguiente manera:

- Los datos son extraídos de las diferentes fuentes: aplicaciones, bases de datos, archivos, etc.
- Los datos son transformados, limpiados e integrados para luego ser cargados en el AD.
- Se visualiza la información contenida en el AD mediante diferentes técnicas.

La figura # 2 visualiza los procesos anteriormente descritos.

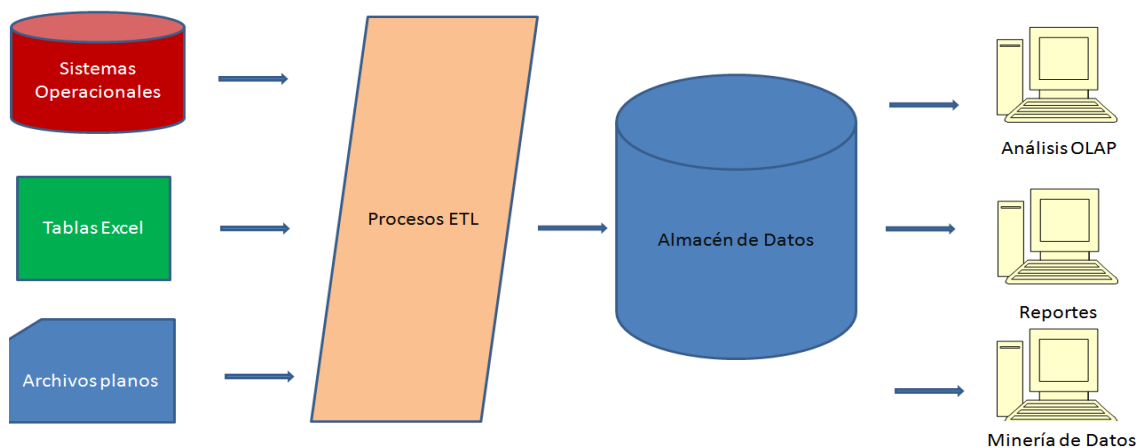


Figura # 2 Arquitectura de un AD

En la construcción de un AD es de suma importancia definir la arquitectura de flujo de datos que tendrá la solución, dicha arquitectura consiste en cómo ordenar los componentes que formarán el AD; las tiendas de datos⁷ son componentes de uso regular en este sentido.

Las tiendas de datos pueden clasificarse en tres tipos basándose en la accesibilidad:

Tienda de datos de cara al cliente: Es una tienda de datos que está disponible para el usuario final, siendo consultada por el mismo o por las aplicaciones del usuario final. (9)

Tienda de datos interna: Es una tienda de datos que es usada internamente por los componentes del AD, con el objetivo de integrar, limpiar o preparar los datos. No es accesible por el usuario final y/o aplicaciones de cara al mismo. (9)

Tienda de datos híbrida: Es una tienda de datos que es usada por mecanismos internos del AD y consultada por el usuario final y/o aplicaciones de cara al mismo. (9)

En cuanto al formato de los datos, las tiendas de datos pueden clasificarse en los siguientes tipos:

Área intermedia (del inglés Staging Area (SA)): Es una tienda de datos interna usada para la transformación y preparación de datos que se obtienen de las fuentes. Es aconsejable incluirlas en la arquitectura de un AD cuando los datos del origen se encuentran en disímiles tipos o formatos, o cuando no se tiene una fuente única, sino que la información proviene de diferentes orígenes. (1)

Tienda de datos normalizada (del inglés Normalized Data Store (NDS)): Es una tienda de datos interna en la forma de una o muchas bases de datos relacionales normalizadas. (9)

Tienda de datos operacional (del inglés Operational Data Store (ODS)): Es una tienda de datos en la forma de una o muchas bases de datos relacionales, que contiene los datos transaccionales más recientes, con el objetivo de apoyar las aplicaciones operacionales. (9)

Tienda de datos dimensional (del inglés Dimensional Data Store (DDS)): Es una tienda de datos de cara al cliente, en la forma de una o muchas bases de datos relacionales donde los datos están ordenados en un formato dimensional con el objetivo de dar apoyo a las consultas de análisis. (9)

Base de Datos Multidimensional: Las bases de datos multidimensionales, proveen una estructura que permite tener acceso flexible a los datos, para explorar y analizar sus relaciones, y resultados consiguientes. Éstas se pueden visualizar como un cubo multidimensional, en donde las variables asociadas existen a lo largo de varios ejes o dimensiones, y la intersección de las mismas representa la medida, indicador o el hecho que se está evaluando. (1)

⁷ **Tienda de Datos:** Una o muchas bases de datos o archivos contenidos dentro del AD, ubicados en una forma particular en dependencia de los objetivos que se quieran cumplir. (9)

La figura # 3 es una representación matricial donde se puede ver claramente lo que se explicó anteriormente.

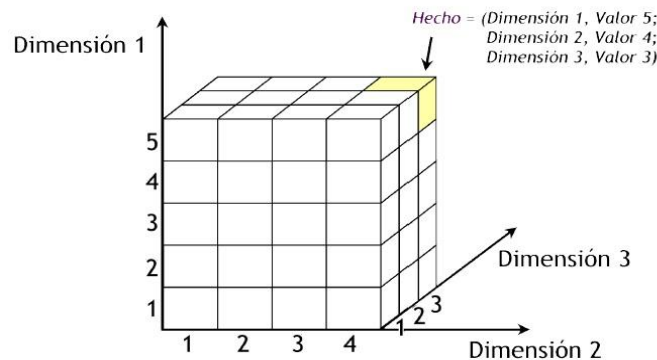


Figura # 3 Cubo Multidimensional (1)

En el cubo que se muestra en la figura # 3 existen tres dimensiones, “Dimensión 1”, “Dimensión 2” y “Dimensión 3”, cada una con sus respectivos valores asociados. También, se ha seleccionado un hecho al azar para demostrar su correspondencia con los valores de las dimensiones. En este caso, la medida a la que se hace referencia, representa el cruce del Valor “5” de “Dimensión 1”, con el Valor “4” de “Dimensión 2” y con el Valor “3” de “Dimensión 3”. Se puede observar que el resultado del análisis está dado por los cruces matriciales de acuerdo a los valores de las dimensiones seleccionadas. (1)

Arquitectura de flujo de datos

Una arquitectura de flujo de datos puede ser definida a partir del uso en conjunto de los diferentes tipos de tiendas de datos abordados, dándoles un orden determinado en dependencia de los objetivos que se quieran cumplir y de las peculiaridades de la fuente de datos. Algunas arquitecturas resultantes de aplicar lo anteriormente dicho son: Tienda de Datos Dimensional Simple, en la cual los componentes que integran el AD serán: un área intermedia, los procesos de control y auditoría, una tienda de datos dimensional y una base de datos multidimensional de cara al cliente; otro ejemplo es, una Tienda Datos Operacional + Tienda de Datos Dimensional, donde los componentes del AD serán: un área intermedia, los procesos de control y auditoría, una tienda de datos operacional, una tienda de datos dimensional y una base de datos multidimensional de cara al cliente. Es posible hacer disímiles combinaciones de los componentes que pueden integrar un AD, todo depende del fin o propósito del mismo. (9)

2.5.1 Definiendo una arquitectura de flujo de datos para la solución

Aplicando a la solución lo anteriormente abordado se decide incluir en la arquitectura de flujo de datos del MD, solamente una tienda de datos, la cual se puede clasificar según el formato de los datos que contendrá como una Base de Datos multidimensional y según la accesibilidad como una tienda de datos híbrida. No es necesario incluir un área intermedia en la arquitectura ya que los datos a procesar se encuentran en solo un sistema fuente, así como tampoco se requiere hacer grandes transformaciones en dichos datos.

Finalmente la arquitectura del MD quedará conformada por tres subsistemas: subsistema de integración de datos, el cual abarcará los procesos ETL; subsistema de almacenamiento, el cual contendrá el esquema de la Base de Datos Multidimensional y finalmente, subsistema de visualización, para el cual se construirán cubos OLAP, para la interacción por parte del cliente con los datos del MD. La arquitectura definida se muestra en la figura # 4.

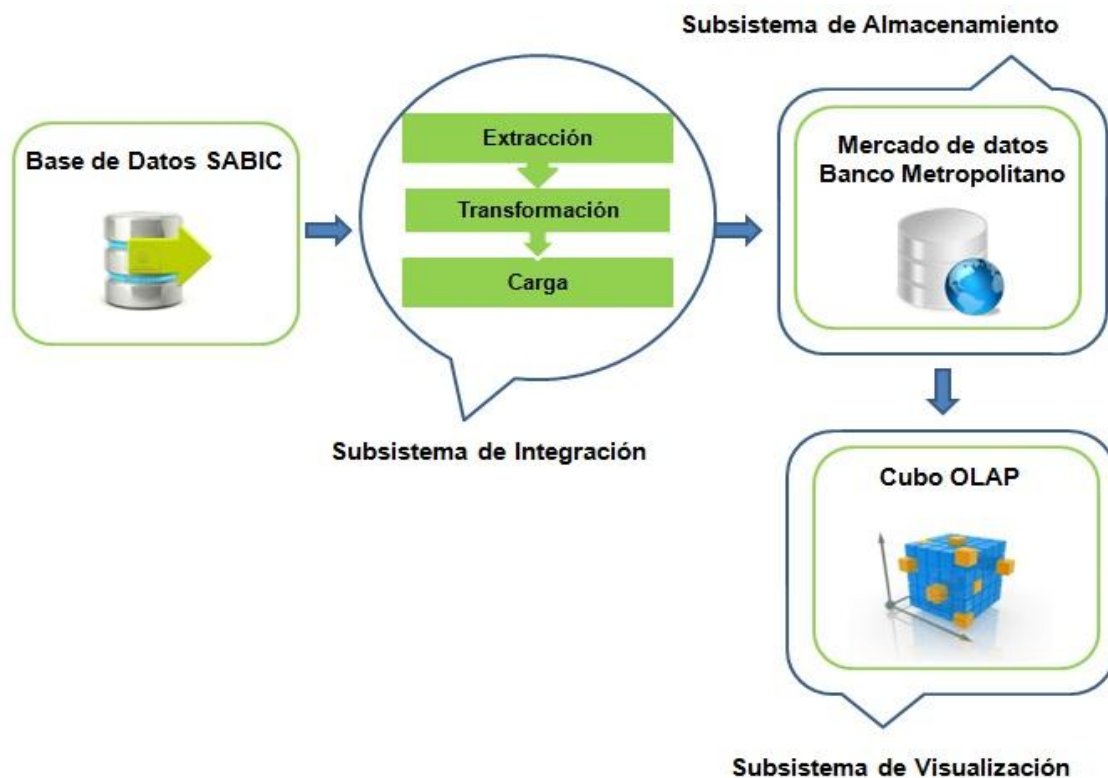


Figura # 4 Arquitectura de flujo de datos del MD

2.5.2 Modelado de una Base de Datos Multidimensional

Las bases de datos multidimensionales implican tres variantes posibles de modelamiento:

Esquema en estrella

El esquema en estrella consta de una tabla de hechos central y de varias tablas de dimensiones relacionadas a ésta a través de sus respectivas claves. En la figura # 5 se puede apreciar un esquema en estrella estándar: (1)

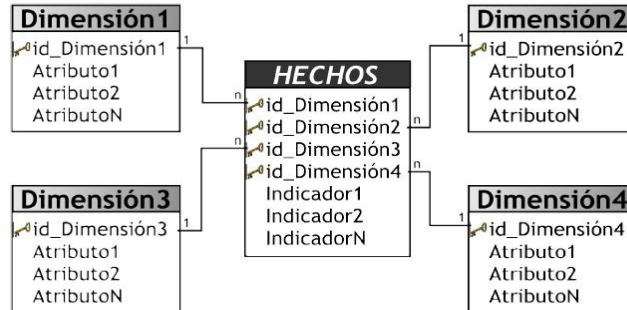


Figura # 5 Esquema en Estrella (1)

Este modelo debe estar totalmente desnormalizado. Las ventajas que trae aparejada la desnormalización, son las de obviar uniones (Join) entre las tablas cuando se realizan consultas, procurando así un mejor tiempo de respuesta y una mayor sencillez con respecto a su utilización. Un punto en contra, es que se genera un cierto grado de redundancia, pero el ahorro de espacio no es significativo. (1) Otro inconveniente de este modelo es que solo presenta una tabla de hechos, por lo tanto solo se miden aquellos aspectos relacionados con el hecho en cuestión. (1)

El esquema en estrella es el más simple de interpretar y para optimizar los tiempos de respuesta ante las consultas de los usuarios. Este modelo es soportado por casi todas las herramientas de consulta y análisis.

Esquema Copo de Nieve

Este esquema representa una extensión del modelo en estrella donde las dimensiones se organizan en jerarquías de dimensiones.

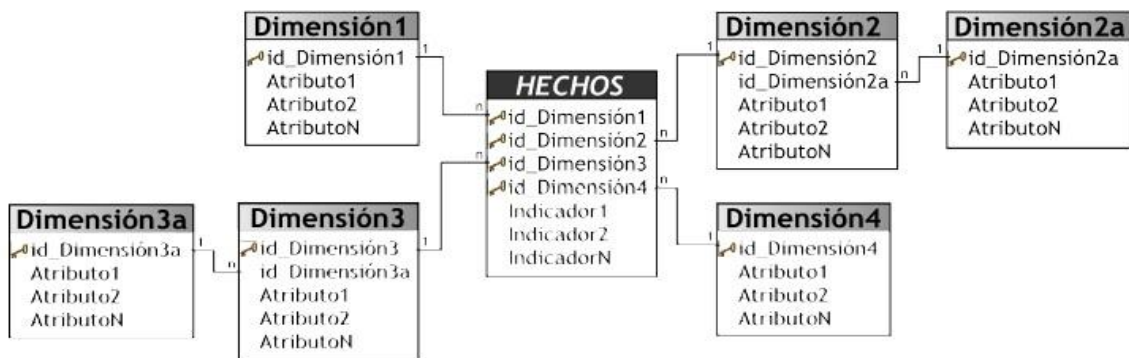


Figura # 6 Esquema Copo de Nieve (1)

Como se puede apreciar en la figura # 6, existe una tabla de hechos central que está relacionada con una o más tablas de dimensiones, quienes a su vez, pueden estar relacionadas o no, con una o más tablas de dimensiones. En este modelo las tablas de dimensiones están normalizadas.

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

- Posee mayor complejidad en su estructura que el esquema en estrella.
- Es muy útil en tablas de dimensiones de muchas tuplas.
- Puede desarrollar clases de jerarquías fuera de las dimensiones, que permiten realizar análisis de lo general a lo detallado y viceversa. (1)

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

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

Esquema Constelación

Este modelo está compuesto por una serie de esquemas en estrella, y tal como se puede apreciar en la figura # 7, está formado por una tabla de hechos principal ("HECHOS_A") y por una o más tablas de hechos auxiliares ("HECHOS_B"), las cuales pueden ser sumalizaciones de la principal. Dichas tablas yacen en el centro del modelo y están relacionadas con sus respectivas tablas de dimensiones. (1)

Diferentes tablas de hechos pueden estar relacionadas con las mismas tablas de dimensiones, éstas reciben el nombre de "dimensiones compartidas".

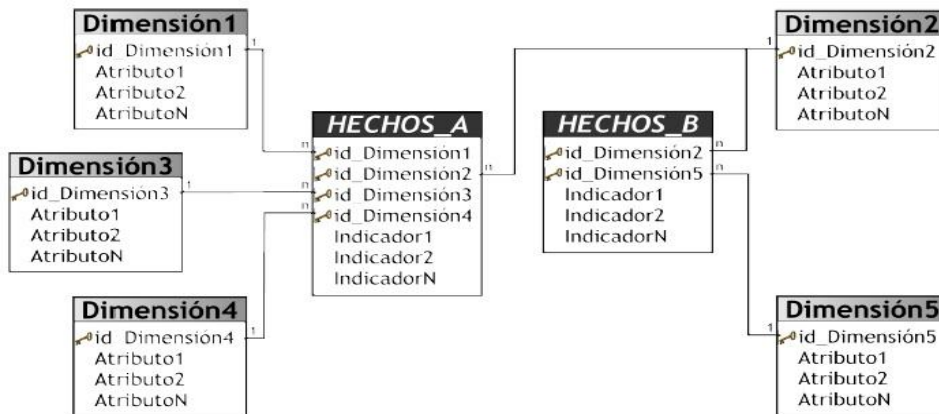


Figura # 7 Esquema Constelación (1)

Su diseño y cualidades son muy similares a las del esquema en estrella, pero posee una serie de diferencias con el mismo, que son precisamente las que lo caracterizan y destacan. Entre ellas se pueden mencionar:

- Permite tener más de una tabla de hechos, por lo cual se podrán analizar más aspectos claves del negocio con un mínimo esfuerzo adicional de diseño.
- Contribuye a la reutilización de dimensiones. (1)

El esquema constelación presenta algunos inconvenientes, donde el principal es que al existir dimensiones compartidas por diferentes hechos, si es necesario realizar cambios sobre determinada dimensión por requerimientos de alguna tabla de hechos, las restantes se verán afectadas también.

En el capítulo segundo se definirá el esquema para modelar el MD.

2.5.3 Modos de almacenamiento de los datos

Los esquemas mencionados pueden ser implementados de diversas maneras independientemente del tipo de arquitectura que se defina para el almacén. Los diferentes tipos de implementación o formas de almacenamiento de los datos a tener en cuenta para la construcción de un MD son los siguientes:

ROLAP

ROLAP (Procesamiento Analítico Relacional en Línea) los datos son almacenados en un formato relacional, lo que significa que existirá latencia⁸ cero entre la base de datos origen y el cubo. Una

⁸**Latencia:** En sistemas informáticos de datos se denomina latencia a la suma de retardos temporales dentro de una red. Un retardo es producido por la demora en la propagación y transmisión de paquetes dentro de la red.

desventaja de este modo de almacenamiento es que presenta un rendimiento de consulta bajo. Es un modo de almacenamiento lento, pero que ahorra espacio considerablemente. (10)

MOLAP

MOLAP (Procesamiento Analítico Multidimensional en Línea) es el modo de almacenamiento más usado. Está diseñado para ofrecer los mejores tiempos de respuesta a los usuarios. Los datos son almacenados en un formato optimizado en un cubo, los datos dentro del cubo se actualizan solo si el cubo está siendo procesado, por lo que la latencia, es alta.(10)

MOLAP brinda un mejor tiempo de respuesta que ROLAP, y posee una estructura optimizada para maximizar las consultas. Es un modo de almacenamiento apropiado para cubos de rápida respuesta. Por otra parte la capacidad de almacenamiento que requiere es mayor que ROLAP. (10)

HOLAP

HOLAP (Procesamiento Analítico Híbrido en Línea) constituye un sistema híbrido entre MOLAP y ROLAP, que combina estas dos implementaciones para almacenar algunos datos en un motor relacional y otros en una base de datos multidimensional, aprovechando las ventajas de ambos modos.(10)

2.5.3.1 Definición de los modos de almacenamiento a utilizar

El modo de almacenamiento que se seleccionó para poblar el MD es ROLAP, debido a que se maneja un considerable cúmulo de datos y es necesario tener presente el crecimiento del MD, por otra parte no es prioridad el tiempo de carga del MD, por lo tanto, no sería necesario implementar HOLAP o MOLAP en este sentido.

En la visualización de los datos del MD mediante la construcción de cubos OLAP, se utilizará el modo de almacenamiento MOLAP, con el objetivo de mejorar el tiempo de respuesta de las consultas al MD.

2.6 Ventajas e inconvenientes de los almacenes de datos

Ventajas

Hay muchas ventajas por las que es recomendable usar un AD. Algunas de ellas son:

- Los AD hacen más fácil el acceso a los usuarios finales a una gran variedad de datos.
- Facilitan el funcionamiento de los sistemas de apoyo a la toma de decisiones tales como: informes de tendencia, informes de excepción, informes que muestran los resultados reales

frente a los objetivos planteados a priori.

- Integra y consolida diferentes fuentes de datos y departamentos empresariales en una única plataforma sólida y centralizada.
- Provee la capacidad de analizar y explotar diferentes áreas de trabajo y de realizar un análisis inmediato de las mismas.
- Permite reaccionar rápidamente ante los cambios.
- Elimina la producción y el procesamiento de datos que no son utilizados ni necesarios, producto a aplicaciones mal diseñadas o no utilizadas.
- Mejora la entrega de la información que los usuarios necesitan, una información completa, correcta, consistente y oportuna. (1)

Inconvenientes

Utilizar AD también plantea algunos inconvenientes, algunos de ellos son:

- A lo largo de su vida los AD pueden suponer altos costos. El AD no suele ser estático. Los costos de mantenimiento son elevados.
- Los AD se pueden quedar obsoletos relativamente pronto.
- A veces ante una petición de información éstos devuelven información no del todo óptima, que también supone una pérdida para la organización.
- A menudo existe una delgada línea entre los AD y los sistemas operacionales. Hay que determinar qué funcionalidades de éstos se pueden aprovechar y cuáles se deben implementar en el AD, resultaría costoso implementar operaciones no necesarias o dejar de implementar alguna realmente importante. (1)

3. Almacenes de datos en el mundo bancario

Las transacciones bancarias generan un gran cúmulo de información que debe ser almacenada y protegida. Disímiles bancos en el mundo emplean los AD para llevar el control de los procesos que realizan diariamente, para garantizar la persistencia en el tiempo y la seguridad de la información que éstos generan. A continuación se expondrán algunos ejemplos de bancos que trabajan exitosamente con AD.

ING-DiBa es el mayor banco de Alemania. Hace pocos años mejoró su plataforma de análisis de datos, con el objetivo de hacer este proceso más rápido y profundo, y que a la vez, constituyera una solución fácil de manejar por los usuarios empresariales. Esto se logró con la construcción de un AD para el banco. El nuevo Sistema de Gestión de la Información, proporciona informes de gestión diaria para más de 300 personas dentro de la organización, produce informes periódicos y análisis de las transacciones que se realizan en el banco. Además del AD poseen sistemas para la

IN y para la aplicación de técnicas de minería de datos. La gerente de la división de AD en el banco, Gisela Hehn, expresó refiriéndose a la nueva tecnología: "Con nuestra amplia solución, hemos duplicado la velocidad del proceso, además del logro de una mayor profundidad de análisis, lo cual es de gran importancia estratégica para nosotros". (11)

El Banco Guayaquil es una institución financiera de banca internacional con base en la ciudad de Guayaquil, Ecuador. Este banco mantenía un sistema de información gerencial basado en la generación de reportes estáticos producidos por demanda. Estos reportes a su vez debían ser transformados con herramientas de análisis como hojas de cálculo, para producir la información mínima necesaria para la toma de decisiones gerenciales, demandando un gran esfuerzo técnico y analítico de diversas áreas de la institución. Ante ello la institución decidió implantar herramientas de IN y AD, las que le han permitido consolidar su información, controlando la duplicidad de datos, generando información para la entidad de control, cuyo destino de mayor importancia consiste en analizar el comportamiento de sus clientes, a partir de la información almacenada con el fin de detectar fraudes bancarios en tiempo real. (12)

El Banco Pichincha es el mayor banco del sector privado, por capitalización y número de depositantes, en el Ecuador. Ramón Larrea Dávila, vicepresidente ejecutivo de negocios del Banco Pichincha, planteó hace algunos años la necesidad de un mayor conocimiento de sus clientes para poder definir adecuadamente sus estrategias de negocio. Tras la implementación de herramientas de IN, AD y técnicas de minería de datos, el Banco Pichincha pudo analizar el comportamiento de sus clientes, segmentarlos, tomar decisiones estratégicas en función de este comportamiento, realizar predicciones y muy importante, detectar movimientos sospechosos en las cuentas de sus clientes, esto último brindó una vía de detección de fraudes bancarios. (13)

Dado el creciente éxito del uso de los AD en el mundo bancario se confirma la necesidad de llevar esta solución a los procesos diarios del Banco Metropolitano de Cuba. Dadas las exigencias del cliente, de construir un AD diseñado específicamente a la medida de sus procesos, no será posible tomar como base la estructura de AD que hayan sido diseñados para otros bancos, pero se pueden tomar como referencia y apoyo, los alentadores resultados obtenidos en el mundo con el uso de estos sistemas de gestión de la información.

3.1 Almacenes de datos en Cuba

Las empresas cubanas en la actualidad, necesitan obtener de una forma cada vez más rápida la información necesaria para tomar decisiones. Con la utilización de las tecnologías de la información y las comunicaciones se trata de alcanzar la mayor productividad posible para obtener ventajas competitivas.

Algunas instituciones han comenzado a hacer uso de las ventajas de un AD como es el caso de la corporación CIMEX, la Oficina Nacional de Estadísticas (ONE), la Red Nacional de Genética Médica, la Unidad Central de Cooperación Médica, el Sistema Penitenciario Cubano, la Universidad de las Ciencias Informáticas, en la cual se han desarrollados varios AD y MD con el fin de apoyar el sistema interno de estas organizaciones, lo cual ha constituido un verdadero éxito. (14)

En las entidades bancarias cubanas no existen precedentes del trabajo con la IN o los AD, sin embargo, la presente investigación es una muestra de que se comienza a trabajar en la inclusión de estas tecnologías para facilitar, fortalecer y hacer más seguros los servicios en los bancos de Cuba.

4. Metodologías para el desarrollo de un Almacén de Datos

4.1 Metodología de Kimball

La metodología de Ralph Kimball es una metodología madura y acabada, que genera abundante documentación y artefactos en cada una de las fases que propone, esto hace más claro y fácil el siguiente proceso. Está basada en un enfoque bottom-up, para la construcción de un AD. Permite presentar resultados parciales a los clientes en cortos plazos, para mantenerlos motivados e informados respecto a la evolución de la solución que esperan. (15)

La metodología Kimball propone un grupo de tareas como parte del ciclo de vida de desarrollo de un AD. Ver figura # 8.

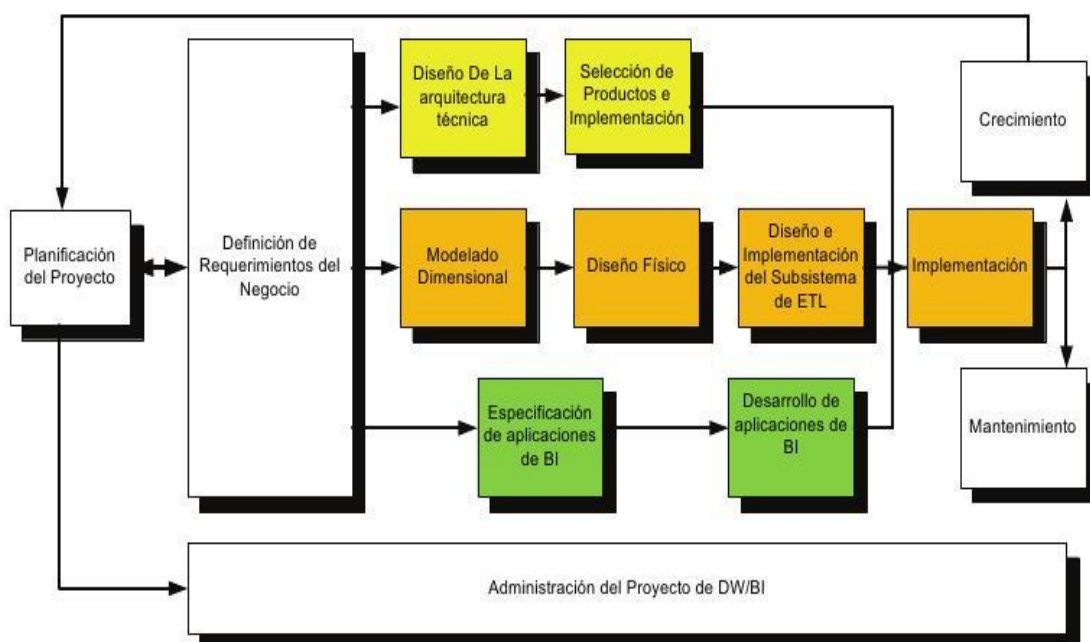


Figura # 8 Tareas que propone la metodología Kimball (15)

4.2 Metodología Hefesto

HEFESTO es una metodología cuya propuesta está fundamentada sobre la investigación, comparación de metodologías existentes y experiencias propias en procesos de confección de AD. (1)

El objetivo que se persigue es no llevar a cabo fases de desarrollo extensas que impliquen exceso de reuniones, análisis o documentación, que conlleven demasiado tiempo y fases de despliegue muy largas. Lo que se busca es entregar una primera implementación que satisfaga una parte de las necesidades, para demostrar las ventajas del AD y motivar al cliente. (1)

La metodología HEFESTO puede ser embebida en cualquier ciclo de vida que cumpla con la condición antes declarada.

Características de la metodología Hefesto:

- Los objetivos y resultados esperados en cada fase se distinguen fácilmente y son sencillos de comprender.
- Se basa en los requerimientos del usuario, por lo cual su estructura es capaz de adaptarse con facilidad y rapidez ante los cambios en el negocio.
- Reduce la resistencia al cambio, ya que involucra al usuario final en cada etapa para que tome decisiones respecto al comportamiento y funciones del AD.
- Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.
- Es independiente de las herramientas que se utilicen para su implementación.
- Es independiente de las estructuras físicas que contengan el AD y de su respectiva distribución.
- Cuando se culmina con una fase, los resultados obtenidos se convierten en el punto de partida para llevar a cabo el paso siguiente.
- Se aplica tanto para MD como para AD. (1)

Descripción:

Como se puede apreciar en la figura # 9, se comienza recolectando las necesidades de información de los usuarios y se obtienen las preguntas claves del negocio. Luego se deben identificar los indicadores y perspectivas de análisis, mediante las cuales se construirá el modelo conceptual de datos del AD. Después se analizarán los OLTP para señalar las correspondencias con los datos fuentes y seleccionar los campos de estudio de cada perspectiva. Una vez hecho

esto, se pasará a la construcción del modelo lógico del AD, explicitando las jerarquías que intervendrán. Por último se definirán los procesos ETL y limpieza de los datos fuente.

Hefesto presenta como inconvenientes de peso, que no explica ni define claramente cuáles artefactos se deben generar en sus respectivas fases, lo que podría ocasionar incongruencias en el proceso de construcción del AD, por otra parte tampoco genera documentación que sirva como constancia de los procesos que ya han sido realizados.

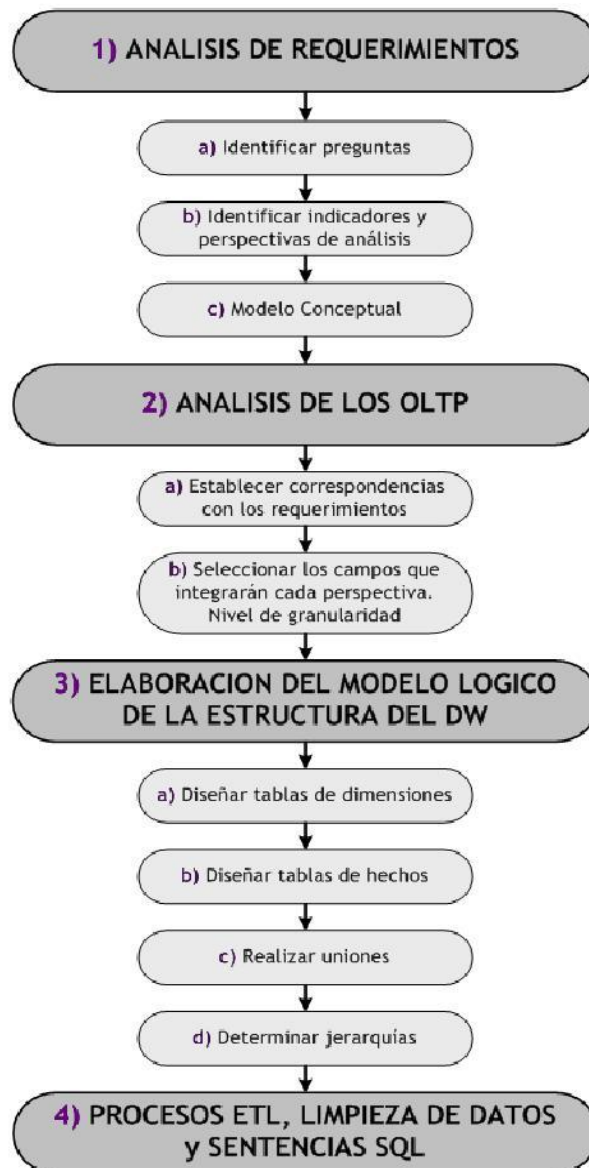


Figura # 9 Metodología Hefesto. Fases y tareas. (1)

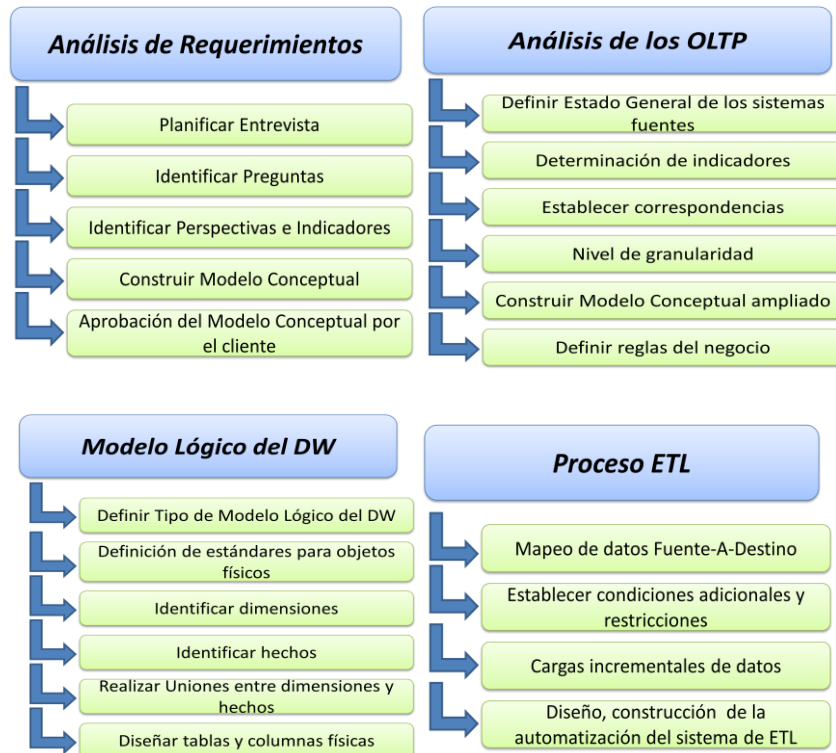
4.3 Metodología de la UCID versión 1.0 para el desarrollo de almacenes de datos

La metodología de la UCID⁹ es un híbrido entre la metodología de Hefesto y la de Ralph Kimball. Es el resultado de una adaptación a las necesidades propias de la UCID. (16)

La decisión de mezclar dichas metodologías fue determinada precisamente con el objetivo de aprovechar las características que posee Hefesto de ser una metodología sencilla y entendible, con la de Kimball de generar artefactos y documentación a la hora de construir el AD. Es importante destacar que algunos de los artefactos adoptados por la UCID fueron modificados, pues se determinó que no era necesario documentar todo lo que Kimball propone en su metodología.

La metodología rectora en el híbrido resultante es Hefesto, ya que se siguen fielmente todas sus fases y tareas, únicamente adicionando la generación de documentos y artefactos necesarios para hacer el proceso de construcción más legible y orientado. (16)

En la figura # 10 se pueden apreciar las cuatro fases que propone la metodología de la UCID en su versión 1.0 con sus respectivas tareas.



⁹ **UCID:** Unidad de Compatibilización e Integración de Software para la Defensa, es un centro de desarrollo de software de las Fuerzas Armadas Revolucionarias (FAR) que radica en la Universidad de las Ciencias Informáticas.

Figura # 10 Fases y tareas de la Metodología de la UCID versión 1.0 (16)

4.3.1 Metodología propuesta para la solución

Se decide utilizar la metodología de la UCID para la construcción del MD, ya que la misma se ajusta perfectamente a las características y necesidades que se presentan para la realización de este trabajo, donde todos los roles serán desempeñados por una sola persona, responsable de la construcción del MD; esta particularidad queda validada por el hecho de que esta metodología, adopta la sencillez y legibilidad de Hefesto. Por otra parte es necesario documentar los procesos más relevantes en la solución y los resultados generales y parciales de los mismos, esto queda validado en la metodología al generar algunos de los artefactos que propone Kimball. La utilización de la metodología de la UCID permitirá hacer el proceso de desarrollo del MD de forma sencilla, ordenada e intuitiva.

5. Herramientas a utilizar en la construcción del Mercado de Datos

Las herramientas para llevar a cabo procesos de IN han permitido hacer más fáciles estudios que antes implicaban trabajos engorrosos y extenuantes. En la actualidad existe un considerable número de herramientas para la creación y explotación de un AD.

5.1 Herramienta de Modelado

Visual Paradigm versión 8.0

Se caracteriza por:

- Disponibilidad en múltiples plataformas (Windows, Linux) y en varios idiomas.
- Permite realizar diseños centrados en casos de uso y enfocados al negocio que generan un software de mayor calidad.
- Uso de un lenguaje estándar común a todo el equipo de desarrollo que facilita la comunicación.
- Capacidades de ingeniería directa e inversa.
- Licencia: gratuita y comercial.
- Permite el diseño de diagramas de interacción, clases, modelado de base de datos etc.
- Generación de código, modelo a código, diagrama a código.
- Permite diseñar diagramas de flujo de datos.
- Tiene incluido editor de figuras. (17)

Se utilizará Visual Paradigm versión 8.0 como herramienta CASE para la construcción del MD, ya que sus características, anteriormente abordadas, constituyen elementos necesarios para obtener

la solución esperada. Siendo una herramienta bajo licencia GPL, se ajusta a las políticas de utilización de software libre de la UCI. Es además la herramienta de modelado definida a utilizar por el centro CEIGE¹⁰ al que pertenece el proyecto encargado de implementar la solución al problema planteado.

5.2 Sistema de Gestión de bases de datos

Microsoft SQL Server 2005

SQL Server 2005 está construido sobre las fortalezas de SQL Server 2000, aumenta el rendimiento, la confiabilidad, la disponibilidad, la capacidad de programación y la facilidad de uso del SQL Server 2000. Incluye herramientas para realizar IN y para la generación de reportes. (18)

Se utilizará en la construcción del MD como Gestor de Base de Datos: SQL Server 2005, cumpliendo con las preferencias del cliente, por otra parte, dado que la fuente de datos funciona sobre este gestor, existirá una mayor homogeneidad entre la fuente origen y el MD, por lo que el lenguaje de consultas no debe diferir.

5.3 Herramientas para el proceso ETL

SQL Server Integration Services 2005 (SSIS)

Es un conjunto de herramientas tanto para las exigencias de las operaciones ETL tradicionales, como para las necesidades de integración de datos para propósitos generales en continua evolución. Facilita la creación de soluciones complejas y sólidas para la extracción, transformación y carga de datos. Proporciona la capacidad de diseñar, crear, implementar y administrar paquetes que tratan los requisitos empresariales cotidianos. (19)

El proceso de llenado de la dimensión tiempo, no se realizará como en el resto de las dimensiones, para esto se utilizará, además de la herramienta SSIS, Microsoft SQL Server Analysis Services (SSAS), que forma parte de la suite de SQL Server 2005. La misma permite la creación de hechos, dimensiones y cubos OLAP, además brinda la posibilidad de crear la dimensión tiempo de forma física en el gestor de base de datos, facilitando en gran medida el proceso, pues al alcance de un clic se generan numerosos campos típicos de esta dimensión con un alto grado de detalle.

¹⁰ **CEIGE:** Centro de Informatización de la Gestión de Entidades, es un centro de desarrollo de software de la UCI especializado en aplicaciones de gestión empresarial.

5.4 Herramientas para la validación del MD

5.4.1 Herramienta para la realización de las pruebas de funcionalidad

Para verificar la funcionalidad del MD se construirán cubos OLAP, a partir de los cuales será posible navegar por las diferentes dimensiones y medidas de los mismos, esto servirá para comprobar que los datos persistentes en el MD, responden a las necesidades de información del cliente. Para la construcción del cubo se utilizará la herramienta SSAS.

SSAS proporciona una arquitectura cliente servidor que facilita un rápido acceso a los datos y una arquitectura para obtener acceso a los datos de la minería de datos. Organiza los datos en cubos, creando datos de agregación pre calculados para proporcionar respuestas rápidas a consultas analíticas complejas, permite la creación de modelos de minería de datos de orígenes multidimensionales (OLAP) y relacionales. Permite utilizar los modos de almacenamiento: MOLAP, ROLAP y HOLAP. (20)

5.4.2 Herramientas para la realización de pruebas de rendimiento y estrés

Apache JMeter

Para la realización de las pruebas de rendimiento se utilizará la herramienta Apache JMeter en su versión 2.4. Esta herramienta es utilizada para efectuar pruebas de carga, para analizar y medir el desempeño de una variedad de servicios: aplicaciones web, bases de datos, etc. Es utilizada comúnmente para hacer pruebas unitarias para conexiones de bases de datos y calcular el tiempo de respuesta de consultas que se realicen a las mismas.

Data Generator 2005 para SQL Server

Para la realización de las pruebas de estrés al MD se utilizará la herramienta Data Generator 2005 para SQL Server en su versión 2.3.0.1. Esta herramienta proporciona una vía de realizar pruebas de carga de bases de datos, especificando el número de tablas que se quieren llenar, así como la cantidad de tuplas que se insertarán por campo.

6. Conclusiones del capítulo

Con el estudio realizado sobre los MD, y los principales conceptos asociados a los mismos, se puede concluir que no existen precedentes de éstos en las diferentes sucursales del Banco Metropolitano. Por tanto, la construcción de un MD para esta institución, es una solución necesaria. El estudio realizado sobre las metodologías y herramientas existentes en la actualidad

para la construcción de MD, posibilitó la selección de la metodología de la UCID por ser una metodología sencilla que no genera grandes cúmulos de documentación o artefactos, pero que a su vez guía de forma ordenada todo el proceso de construcción de AD, lo cual se ajusta a las necesidades del desarrollador; por otra parte las herramientas seleccionadas en base al trabajo que se realizará, permitirán el correcto desarrollo de la solución. El estudio realizado permitirá seguir pasos concretos y predefinidos en las diferentes etapas de construcción del MD, enfocados a obtener la solución deseada en el tiempo dispuesto.

Capítulo 2 Análisis y diseño del Mercado de Datos

1. Introducción

La etapa de diseño de un MD es de vital importancia, puesto que el modelo resultante debe responder a las necesidades y peticiones de la empresa para la que se construye la solución, de lo contrario, los restantes pasos de desarrollo se darán con dificultad y sin el éxito esperado.

En este capítulo se realiza el análisis y diseño del MD para el Banco Metropolitano, en correspondencia con los pasos establecidos en cada una de las fases que propone la metodología de la UCID descrita en el capítulo primero, se transita desde el análisis de los requerimientos, análisis de los OLTP hasta la elaboración del modelo lógico del MD.

2. Procedimiento para el desarrollo de la solución utilizando la metodología escogida

La metodología de la UCID propone cuatro fases para la construcción de un MD, a continuación se abordará el procedimiento para el desarrollo de la solución enfocado en las tres primeras fases de dicha metodología.

2.1 Fase # 1. Análisis de los requerimientos

En esta fase se llevará a cabo la planificación de cada una de las entrevistas que se realizarán con el cliente, luego se identificarán los requerimientos del usuario a través de preguntas que expliciten los objetivos de su organización. Se analizarán dichas preguntas con el fin de identificar qué medidas y perspectivas serán tomadas en cuenta para la construcción del MD. Se confeccionará un modelo conceptual mediante el cual se tendrá una idea inicial de la información que se guardará posteriormente en el MD. Finalmente concluye esta primera fase con la aprobación del modelo conceptual por el cliente, donde se obtendrá un acta de validación para poder continuar con las demás fases.

2.1.1 Planificar las entrevistas

En un primer encuentro con los especialistas de diferentes departamentos de la Sede Central del Banco Metropolitano: Solución y Prevención de Fraudes, Sistemas de Información e Informática,

se llegó al acuerdo de realizar cinco entrevistas con el fin de construir el modelo conceptual del MD.

Entrevista 1

La entrevista fue realizada con el objetivo de conocer el funcionamiento del negocio en cuestión, profundizar en la estructura y contenido de la fuente de datos existente en el Banco Metropolitano. En el encuentro estuvieron presentes los especialistas del departamento de Sistemas de Información y de Solución y Prevención de Fraudes. Se logró analizar la estructura actual de la fuente de datos SABIC y se realizó un estudio de las tablas clasificadoras, maestras y operativas presentes en la misma.

Entrevista 2

La entrevista fue realizada con el objetivo de comenzar a identificar posibles indicadores y perspectivas a partir de preguntas elaboradas según las necesidades de información del cliente. Se logró identificar correctamente una parte de las preguntas, a la vez que se determinó qué tablas de la fuente de datos es necesario consultar con el fin de obtener los datos para dar respuesta a las mismas. En el encuentro estuvieron presentes los especialistas del departamento de Solución y Prevención de Fraudes. Una vez culminada la entrevista fue posible comenzar a trabajar en la construcción del modelo conceptual del MD (ver epígrafe 2.1.4), habiendo identificado los hechos con sus perspectivas e indicadores:

- hecho_capitalIlícito_cuentaCN
- hecho_capitalIlícito_cuentaCJ
- hecho_capitalIlícito_depósitoExtracciónCN
- hecho_capitalIlícito_depósitoExtracciónCJ

Entrevista 3

La entrevista fue realizada con el objetivo de continuar identificando posibles indicadores y perspectivas de análisis. Se logró identificar correctamente el resto de las preguntas, a la vez que se determinó qué tablas de la fuente de datos es necesario consultar con el fin de obtener los datos para dar respuesta a las mismas. En el encuentro estuvieron presentes los especialistas del departamento de Solución y Prevención de Fraudes. Una vez culminada la entrevista fue posible completar la construcción del modelo conceptual del MD (ver epígrafe 2.1.4), habiendo identificado nuevos hechos con sus perspectivas e indicadores:

- hecho_capitalIlícito_transferenciasCN
- hecho_capitalIlícito_transferenciasCJ
- hecho_capitalIlícito_transferenciasExtranjero

Entrevista 4

La entrevista fue realizada con el objetivo de definir el nivel de granularidad de cada una de las perspectivas identificadas para la solución. En el encuentro estuvieron presentes los especialistas del departamento de Solución y Prevención de Fraudes. Se logró seleccionar los campos que integrarán cada una de las perspectivas identificadas. Una vez culminada la entrevista, se pudo comenzar la construcción del Modelo Conceptual Ampliado de los hechos (ver epígrafe 2.2.5):

- hecho_capitalIlícito_cuentaCN
- hecho_capitalIlícito_cuentaCJ
- hecho_capitalIlícito_depósitoExtracciónCN
- hecho_capitalIlícito_depósitoExtracciónCJ
- hecho_capitalIlícito_transferenciaCN
- hecho_capitalIlícito_transferenciaCJ
- hecho_capitalIlícito_transferenciaExtranjero

Entrevista 5

La entrevista fue realizada con el objetivo de definir cómo calcular los indicadores de análisis del comportamiento de los clientes del banco con el empleo de tarjetas bancarias. En el encuentro estuvieron presentes los especialistas del departamento de Solución y Prevención de Fraudes. La entrevista dio cumplimiento al objetivo principal de la misma, ya que se obtuvieron fórmulas para calcular los indicadores de los hechos identificados (ver epígrafe 2.2.2).

2.1.2 Identificar preguntas, perspectivas e indicadores

El objetivo principal de este paso es identificar las necesidades de información claves del cliente. Es importante comprender las peticiones del mismo, ya que el MD debe estar en función de proporcionar la información necesaria para cumplir con las metas y estrategias del cliente. (16)

El proceso de identificación de las preguntas, perspectivas e indicadores de análisis se realizó mediante entrevistas estructuradas con el cliente. En las cuales se formularon interrogantes complejas sobre las diferentes operaciones que pueden realizar los usuarios del Banco Metropolitano involucrando tarjetas bancarias.

En las entrevistas realizadas se logró identificar un total de 26 preguntas, las cuales fueron descompuestas con el objetivo de descubrir las diferentes perspectivas e indicadores que intervienen en cada una de ellas.

Para mejor comprensión del lector se aclara que: las perspectivas identificadas se han subrayado de color azul y los indicadores de color rojo.

Pregunta 1

Cantidad de cuentas que posee un cliente natural de cierto país en un período de tiempo determinado.

Pregunta 2

Saldo total de las cuentas en USD¹¹ de un cliente natural de cierto país en un período de tiempo determinado.

Pregunta 3

Saldo total de las cuentas en CUC¹² de un cliente natural de cierto país en un período de tiempo determinado.

Pregunta 4

Saldo total de las cuentas en CUP¹³ de un cliente natural de cierto país en un período de tiempo determinado.

Pregunta 5

Cantidad de cuentas que posee un cliente jurídico por localización en un período de tiempo determinado.

Pregunta 6

Saldo total de las cuentas en USD de un cliente jurídico por localización en un período de tiempo determinado.

Pregunta 7

Saldo total de las cuentas en CUC de un cliente jurídico por localización en un período de tiempo determinado.

Pregunta 8

Saldo total de las cuentas en CUP de un cliente jurídico por localización en un período de tiempo determinado.

Pregunta 9

¹¹ **USD:** Dólar estadounidense, es la moneda oficial de Estados Unidos. En Cuba es una de las tres monedas base del sistema bancario.

¹² **CUC:** Peso cubano convertible, es una de las dos monedas oficiales de Cuba.

¹³ **CUP:** Peso cubano, es la moneda de curso legal de la República de Cuba.

Cantidad de depósitos de un cliente natural de cierto país por lugar de operación en un período de tiempo determinado.

Pregunta 10

Importe total de los depósitos de un cliente natural de cierto país por lugar operación en un período de tiempo determinado.

Pregunta 11

Cantidad de extracciones de un cliente natural de cierto país por lugar de operación en un período de tiempo determinado.

Pregunta 12

Importe total de las extracciones de un cliente natural de cierto país por lugar operación en un período de tiempo determinado.

Pregunta 13

Cantidad de depósitos de un cliente jurídico por localización y lugar de operación en un período de tiempo determinado.

Pregunta 14

Importe total de los depósitos de un cliente jurídico por localización y lugar operación en un período de tiempo determinado.

Pregunta 15

Cantidad de extracciones de un cliente jurídico por localización y lugar de operación en un período de tiempo determinado.

Pregunta 16

Importe total de las extracciones de un cliente jurídico por localización y lugar operación en un período de tiempo determinado.

Pregunta 17

Cantidad de transferencias de un cliente natural de cierto país por lugar de operación en un período de tiempo determinado.

Pregunta 18

Importe total de las transferencias de un cliente natural de cierto país por lugar de operación en un período de tiempo determinado.

Pregunta 19

Cantidad de beneficiarios de las transferencias de un cliente natural de cierto país por lugar de operación en un período de tiempo determinado.

Pregunta 20

Cantidad de transferencias de un cliente jurídico por localización y por lugar de operación en un período de tiempo determinado.

Pregunta 21

Importe total de las transferencias de un cliente jurídico por localización y por lugar de operación en un período de tiempo determinado.

Pregunta 22

Cantidad de beneficiarios de las transferencias de un cliente jurídico por localización y por lugar de operación en un período de tiempo determinado.

Pregunta 23

Cantidad de transferencias de un ordenante extranjero por país en un período de tiempo determinado.

Pregunta 24

Importe total de las transferencias de un ordenante extranjero por país en período de tiempo determinado.

Pregunta 25

Cantidad de beneficiarios de un ordenante extranjero por país en un período de tiempo determinado.

Pregunta 26

Cantidad de bancos remitentes por ordenante extranjero y por país en un período de tiempo determinado.

Para asegurar que se realizó un buen análisis, se comprobó que el resultado del mismo explicitara los objetivos estratégicos planteados por el Banco Metropolitano. Se verificó además, que todo el estudio realizado estuviera enfocado en las necesidades de información de los procesos principales que se llevan a cabo en dicha institución.

Una vez identificados los indicadores y perspectivas de análisis, se puede proceder a la construcción del modelo conceptual del MD.

2.1.3 Modelo Conceptual

A través del modelo conceptual se podrá observar con claridad el alcance de la propuesta de solución.; además, al poseer un nivel de definición de los datos comprensible para personas que no posean conocimientos informáticos, permite que pueda ser presentado y explicado con facilidad al cliente.

En diseño del modelo conceptual del MD se incluyeron tres áreas de análisis del negocio diferentes: “Cuenta”, “Depósito-Extracción” y “Transferencia”, por cada una de éstas se diseñó un modelo conceptual, ya que en las mismas intervienen diferentes perspectivas e indicadores.

En las figuras # 11, 12 y 13 se puede observar el modelo conceptual de cada una de las áreas de análisis identificadas.

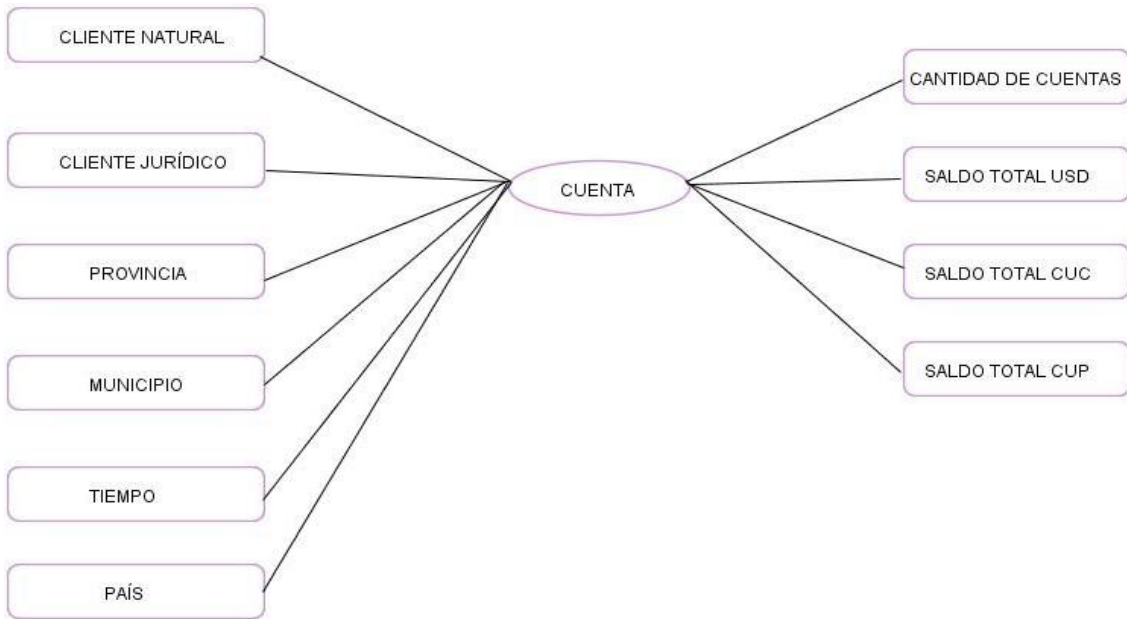


Figura # 11 Modelo Conceptual del área de análisis “Cuenta”

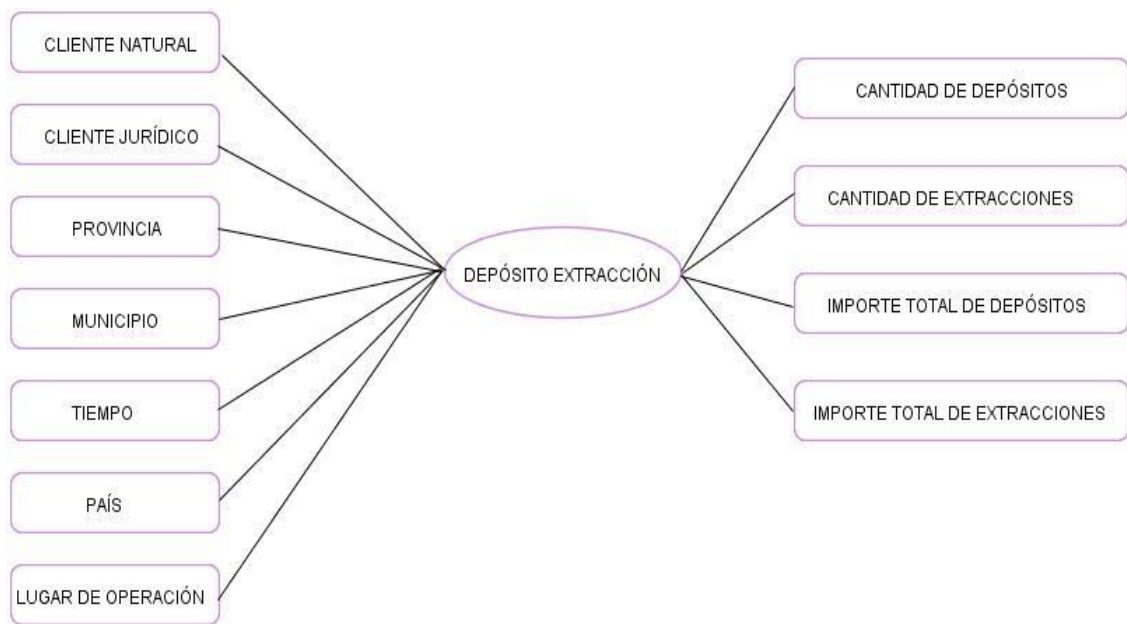


Figura # 12 Modelo Conceptual del área de análisis “Depósito - Extracción”

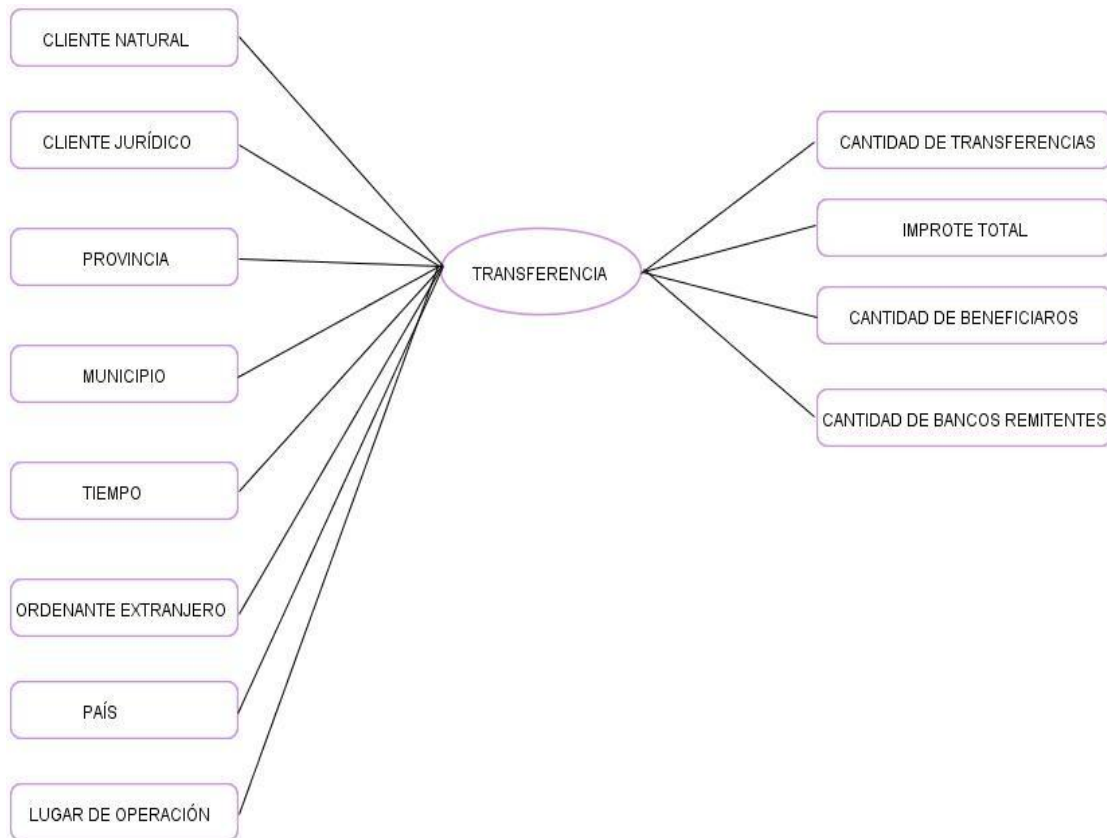


Figura # 13 Modelo Conceptual del área de análisis “Transferencia”

2.1.4 Aprobación del modelo conceptual por el cliente

Para asegurar que el modelo conceptual fue construido correctamente y que satisface las necesidades de información del cliente, deberá ser aprobado por el mismo en una reunión donde se le expone el modelo obtenido, utilizando siempre un lenguaje entendible para ambas partes. (16)

Con el objetivo de dar cumplimiento a esta última etapa de la fase primera, se realizó una reunión donde se le expuso al cliente el modelo conceptual diseñado. Al finalizar el cliente manifestó su aprobación con la firma del acta de validación del modelo. Ver anexo # 1.

2.2 Fase # 2. Análisis de los OLTP

En esta fase se analizará la fuente de datos SABIC con el objetivo de definir el estado general de la misma. Se determinará cómo serán calculados los indicadores y se establecerán las respectivas correspondencias entre el modelo conceptual creado en el paso anterior y las tablas presentes en SABIC. Luego se definirá qué campos se incluirán en cada perspectiva. Se ampliará

el modelo conceptual con la información obtenida en este paso, y se finalizará definiendo las reglas del negocio.

2.2.1 Definir el estado general de los sistemas fuentes

En esta etapa, con el fin de realizar un análisis de los OLTP, se comienza a definir el estado general de los sistemas fuentes en cuanto a la disponibilidad y accesibilidad. (16)

MetrodataW es el nombre del servidor físico donde se realiza una salva histórica de todos los datos que contiene el sistema contable SABIC del Banco Metropolitano. Es administrada en el gestor de base de datos Microsoft SQL Server 2005, sobre el sistema operativo Windows Server, por los especialistas: Francisco Ramos (Programador del Centro Informativo) y Alfredo Hernández (Subdirector del Centro Informativo). Actualmente se almacena un cúmulo de información que alcanza los 4.2 terabytes. Sobre esta fuente operan diariamente alrededor de 10 usuarios con los permisos pertinentes, que realizan aproximadamente 100 operaciones en cada jornada laboral. Está compuesta aproximadamente de 344 tablas y 91 procedimientos almacenados.

Dado lo anteriormente expuesto se puede decir que la fuente de datos es una gran base de datos que contiene un cúmulo de información grande y que sobre la misma operan aproximadamente 10 usuarios de la dirección.

2.2.2 Determinación de indicadores

Los indicadores identificados son valores numéricos que son calculados a partir de sumatorias o contadores de datos que se obtienen de la fuente. En la tabla # 1 se presentan las fórmulas correspondientes para calcular dichos indicadores:

Tabla # 1 Fórmulas para calcular los indicadores identificados

Indicador	Fórmula
Cantidad de cuentas de un cliente.	COUNT (cuenta)
Saldo total de las cuentas en CUC de un cliente.	SUM (saldo contable de cada cuenta en CUC)
Saldo total de las cuentas en CUP de un cliente.	SUM (saldo contable de cada cuenta en CUP)
Saldo total de las cuentas en USD de un cliente.	SUM (saldo contable de cada cuenta en USD)
Importe total de las transferencias por cliente.	SUM (importe de cada transferencia)
Cantidad de transferencias por cliente.	COUNT (transferencia)
Cantidad de transferencias por ordenante extranjero.	COUNT (transferencia)
Cantidad de beneficiarios de una transferencia por	COUNT (beneficiario)

ordenante.	
Cantidad de bancos remitentes desde los que transfirió un ordenante.	COUNT(banco remitente)
Cantidad de depósitos por cliente.	COUNT (crédito)
Importe total de los depósitos por cliente.	SUM (importe de cada depósito)
Cantidad de extracciones por cliente.	COUNT (débito)
Importe total de las extracciones por cliente.	SUM (importe de cada extracción)

2.2.3 Establecer correspondencias

El objetivo de este paso es verificar que la fuente de datos pueda proporcionar la información requerida, para ello es necesario establecer correspondencias entre la misma y el modelo conceptual elaborado. (16)

La idea es que los elementos del modelo conceptual (perspectivas e indicadores) estén correspondidos en el OLTP, más específicamente en las tablas o campos de las tablas desde donde se vaya a extraer la información que se necesitará posteriormente para completar las perspectivas e indicadores. (16)

En la tabla # 2 se pueden apreciar las correspondencias entre las perspectivas del modelo conceptual y las tablas presentes en la fuente de datos.

Tabla # 2 Correspondencia de las perspectivas con las tablas presentes en la fuente de datos

Perspectiva	Tabla de la Base de datos	Campo de la tabla
Cliente natural	C_PERSON	CÓDIGO (La unión de los campos: COD_PAEXID, COD_TIPID, NUM_IDEPER) DIRECCIÓN RESIDENCIA CANCELADOS STATUS ENTIDAD
	C_CATOCU	COD_OCUP NOM_OCUP
Cliente jurídico	M_CLIENTGEN	COD_CLIENT CANCELADOS TIP_CLIENT DIRECCIÓN
Lugar de operación	C_CAJERO	COD_BANCO COD_CAJERO NOM_CAJERO
	C_CCOSTO	COD_CCOSTO

		NOM_CCOSTO COD_BANCO
Provincia	M_CLIENTGEN C_DPA C_PROVIN	COD_DPA COD_PROVIN NOM_PROVIN
Municipio	C_DPA	NOM_MUNIC
Operante extranjero	M_HISREC	C_IMPRESO
País	C_PAISES	COD_PAIS NOM_PAIS
Tiempo	M_MAYOR H_HISTOR M_HISREC	FEC_ULTCON FEC_POSTEO C_IMPRESO

En la tabla # 3 se pueden apreciar las correspondencias entre los indicadores del modelo conceptual y las tablas presentes en la fuente de datos.

Tabla # 3 Correspondencia de los indicadores con las tablas presentes en la fuente de datos

Indicador	Tabla de la Base de datos	Campo de la tabla
Cantidad de cuentas por cliente	M_PERCUE M_CUECLIGEN	COD_MONEDA CUE_SUBCUE TIP_CONTRA COD_CONTRA COD_PAEXID COD_TIPID NUM_IDEPER COD_CLIENT COD_MONEDA CUE_SUBCUE TIP_CONTRA COD_CONTRA
Saldo total CUP	M_MAYOR	SAL_CONTAB COD_MONEDA
Saldo total CUC	M_MAYOR	SAL_CONTAB COD_MONEDA
Saldo total USD	M_MAYOR	SAL_CONTAB COD_MONEDA
Importe total de las transferencias por cliente	-	-
Cantidad de transferencias por cliente	-	-
Cantidad de transferencias por ordenante extranjero	M_HISREC	C_IMPRESO
Cantidad de beneficiarios por ordenante	M_HISREC	C_IMPRESO
Cantidad de bancos remitentes desde los que transfirió un ordenante extranjero.	M_HISREC	C_IMPRESO
Importe total de las extracciones	H_HISTOR	IMP_ASIENT

por cliente		
Cantidad de extracciones por cliente	H_HISTOR	IMP_ASIENT
Importe total de los depósitos por cliente	H_HISTOR	IMP_ASIENT
Cantidad de depósitos por cliente	H_HISTOR	IMP_ASIENT

Como se puede apreciar en la tabla # 3 en el caso de los identificadores del área de análisis “Transferencia”, no fue posible determinar fórmulas para el cálculo de los mismos, ya que actualmente no existe un camino en la fuente de datos para calcular dichos valores.

2.2.4 Nivel de granularidad

Una vez establecidas las relaciones con los OLTP, se examinaron y seleccionaron los campos que contendrá cada perspectiva, ya que a través de éstos, se manipularán y filtrarán los indicadores. De las tablas de la fuente se llevaron al MD la mayoría de los campos, por lo que se puede decir que se podrá tener un nivel bastante alto de detalle al analizar los datos.

En el epígrafe 2.2.5 se puede apreciar mediante el modelo conceptual ampliado del MD los campos seleccionados a integrar cada una de las perspectivas identificadas.

2.2.5 Modelo conceptual ampliado

Con la determinación del nivel de granularidad que tendrán las perspectivas identificadas se puede diseñar el modelo conceptual ampliado del MD, de esta forma se podrán ilustrar los resultados obtenidos en los pasos anteriores; colocando bajo cada perspectiva los campos elegidos y bajo cada indicador la fórmula para su cálculo. Las figuras # 14 y 15 visualizan el modelo conceptual ampliado del área de análisis “Cuenta”. Ver anexos 2, 3, 4 y 5 donde se muestran los modelos obtenidos en las restantes áreas de análisis: “Depósito-Extracción” y “Transferencia”.

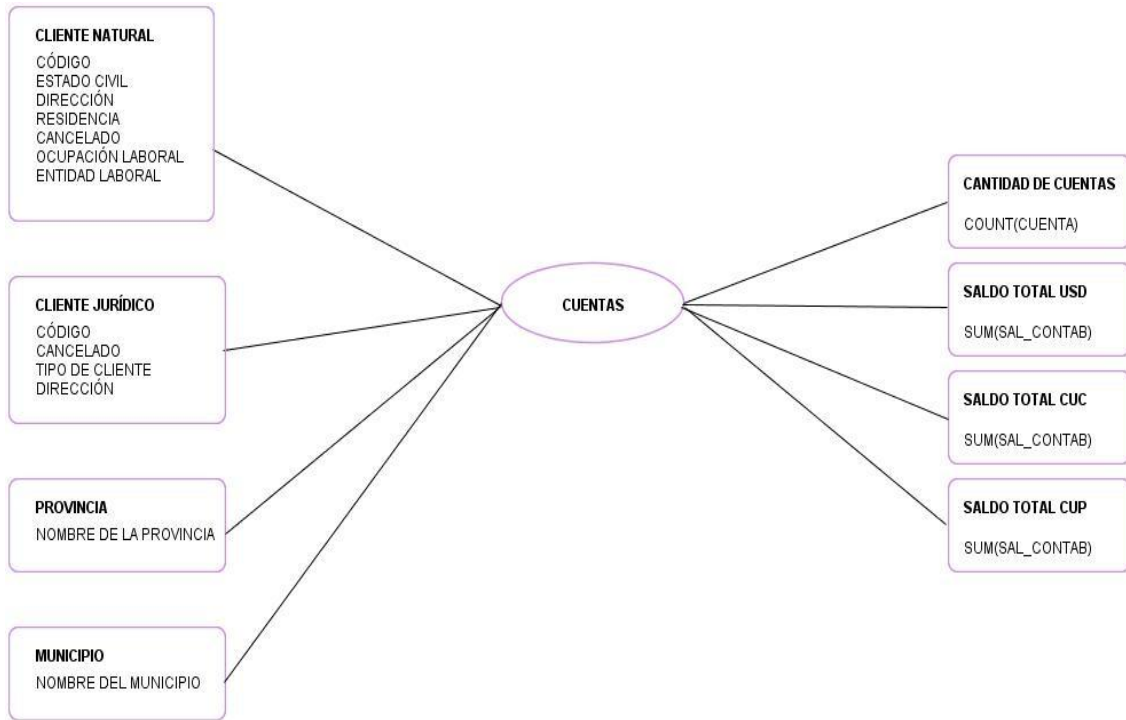


Figura # 14 Primera parte del Modelo Conceptual Ampliado del área de análisis “Cuenta”

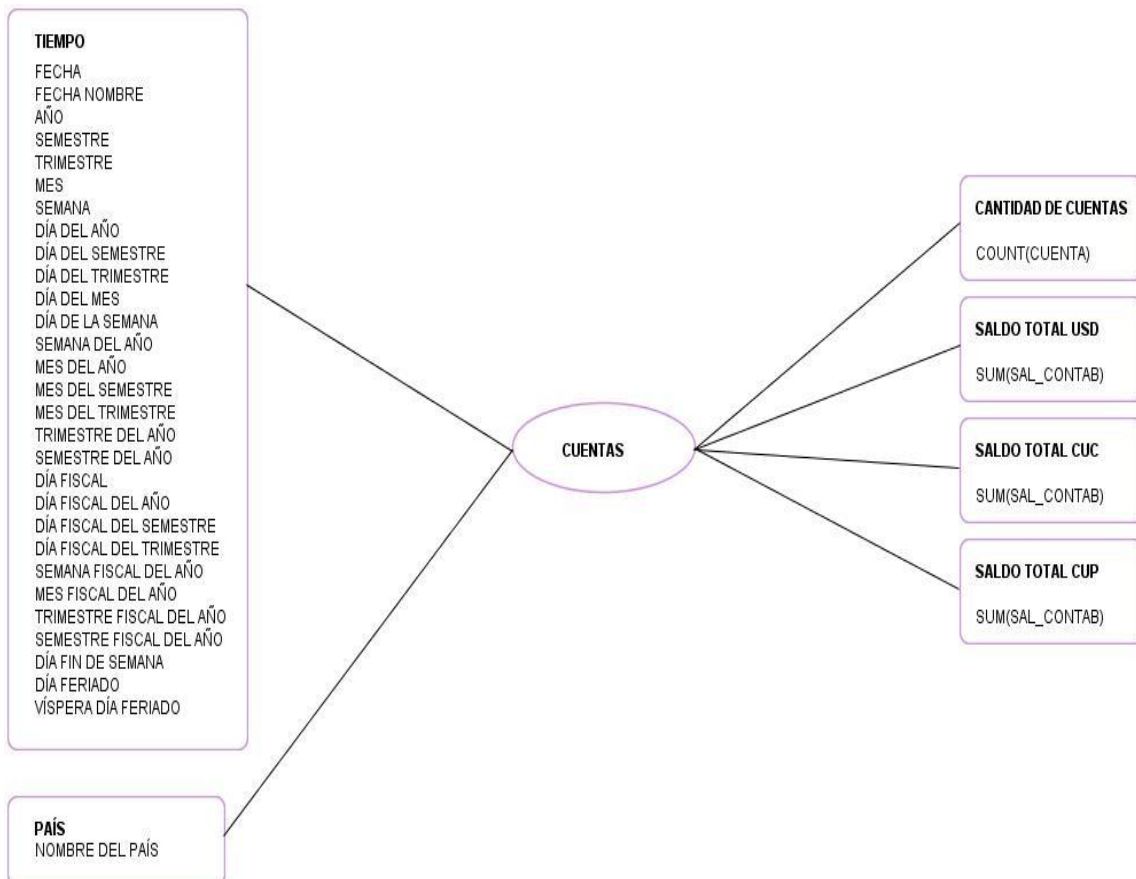


Figura # 15 Segunda parte del Modelo Conceptual Ampliado del área de análisis “Cuenta”

2.3 Fase # 3. Modelo lógico del MD

En esta fase se confeccionará el modelo lógico de la estructura del MD teniendo como base el modelo conceptual ampliado que ya ha sido creado. Para ello, primero se definirá el tipo de esquema que se utilizará; luego se llevarán a cabo las acciones propias al caso, para diseñar las tablas de dimensiones y de hechos. Se realizarán las uniones pertinentes entre estas tablas. Y finalmente se diseñarán las tablas y columnas físicas del AD. (16)

2.3.1 Definir el tipo de modelo del MD

Luego de haber realizado el análisis de los requerimientos del usuario y de los diferentes tipos de modelación que se pueden realizar para construir un MD, se decide, dadas las peculiaridades del negocio en cuestión, construir el modelo lógico como un esquema en constelación; debido a que se tienen tres áreas de análisis identificadas, las cuales serán representadas por varios hechos, según el tipo de cliente que se desee analizar: cliente jurídico o natural. Dadas las múltiples ventajas de este esquema, anteriormente expuestas en el capítulo primero, se obtendrá un modelo del MD que asegure el buen funcionamiento de la solución.

2.3.2 Identificar dimensiones

En este paso se diseñan las tablas de dimensiones que serán parte del MD; para esto cada perspectiva identificada en fases anteriores se toma como una dimensión para el modelo. Fue necesario cambiar algunos nombres de las tablas y campos de la fuente a la hora de incluirlos en el modelo con el objetivo de hacerlos lo más explícitos posible. En la tabla # 4 se pueden apreciar las dimensiones identificadas con sus respectivos campos.

Tabla # 4 Dimensiones y campos que integran las mismas

Dimensión	Campos
dim_clienteNatural	Id_dim_clienteNatural Ocupación laboral Entidad laboral Dirección Residencia Estado civil Cancelado

dim_clienteJurídico	Id_dim_clienteJurídico Código Dirección Cancelado TipoCliente
dim_provincia	Id_dim_provincia NombreProvincia CódigoProvincia
dim_municipio	Id_dim_municipio NombreMunicipio CódigoDPA
dim_lugarOperación	Id_dim_lugarOperación CentroCosto TipoLugar NombreLugar Terminal
dim_ordenanteExtranjero	Id_dim_ordenanteExtranjero nombreOrdenante
dim_país	Id_dim_país CódigoPaís NombrePaís
dim_tiempo	Fecha Fecha_nombre Año Semestre Trimestre Mes Semana Día_del_año Día_del_semestre Día_del_trimestre Día_del_mes Día_de_la_semana Semana_del_año Semana_del_semestre Semana_del_trimestre Trimestre_del_año Semestre_del_año

	Día_fiscal Día_fiscal_del_año Día_fiscal_del_semestre Día_fiscal_del_trimestre Semana_fiscal_del_año Mes_fiscal_del_año Trimestre_fiscal_del_año Semestre_fiscal_del_año Día_fin_de_semana Día_feriado Víspera_día_feriado
--	--

2.3.3 Identificar hechos

En este paso se definieron las tablas de hechos del MD, se debe aclarar que las tres áreas de análisis identificadas en fases anteriores, fueron particionadas en diferentes tablas de hechos según el tipo de cliente que se desee analizar: cliente jurídico o natural, fue necesario hacer esta especificación en el modelo ya que la información que reportan los clientes naturales no es la misma que la de los clientes jurídicos, así que las dimensiones asociadas a los hechos en este caso tampoco serían iguales, esto hace necesario crear hechos diferentes para cada tipo de cliente. Las tablas de hechos contienen los indicadores de estudio. En la tabla # 5 se pueden apreciar los hechos identificados con los indicadores correspondientes.

Tabla # 5 Hechos e indicadores identificados

Hecho	Indicadores
hecho_capitalilícito_cuentaClienteNatural	Cantidad Saldo total en CUP Saldo total en CUC Saldo total en USD
hecho_capitalilícito_cuentaClienteJurídico	Cantidad Saldo total en CUP Saldo total en CUC Saldo total en USD
hecho_capitalilícito_depósitoExtracciónClienteNatural	Cantidad de depósitos Saldo total de los depósitos Cantidad de extracciones Saldo total de las extracciones

hecho_capitallicito_depósitoExtracciónClienteJurídico	Cantidad de depósitos Saldo total de los depósitos Cantidad de extracciones Saldo total de las extracciones
hecho_capitallicito_transferencia_ClienteNatural	Cantidad de transferencias Saldo total de las transferencias Cantidad de beneficiarios
hecho_capitallicito_transferencia_ClienteJurídico	Cantidad de transferencias Saldo total de las transferencias Cantidad de beneficiarios
hecho_capitallicito_transferencia_Extranjero	Cantidad de transferencias Saldo total de las transferencias Cantidad de beneficiarios Cantidad de bancos remitentes

Es válido aclarar que no será posible efectuar el llenado de dos hechos del MD, debido a que en la fuente de datos no existe un camino para calcular los indicadores de estos hechos. Sin embargo se incluirán en el diseño del MD, ya que el cliente trabaja en una solución para este problema. Los hechos en cuestión son:

- hecho_capitallicito_depósitoExtracciónClienteJurídico
- hecho_capitallicito_depósitoExtracciónClienteNatural

2.3.4 Diseñar tablas y columnas físicas

Una vez definidos los hechos y las dimensiones que formarán parte del MD, se procede a la construcción del modelo físico del mismo. Donde se han realizado las uniones pertinentes entre los elementos anteriormente mencionados. En la figura # 16 se puede ver un fragmento del modelo de datos diseñado, en el cual no se han incluido los atributos de las tablas, pero se pueden apreciar las dimensiones y hechos incluidos en el diseño y las relaciones entre estos elementos. En el anexo # 6 se puede apreciar el modelo de datos del MD completo.

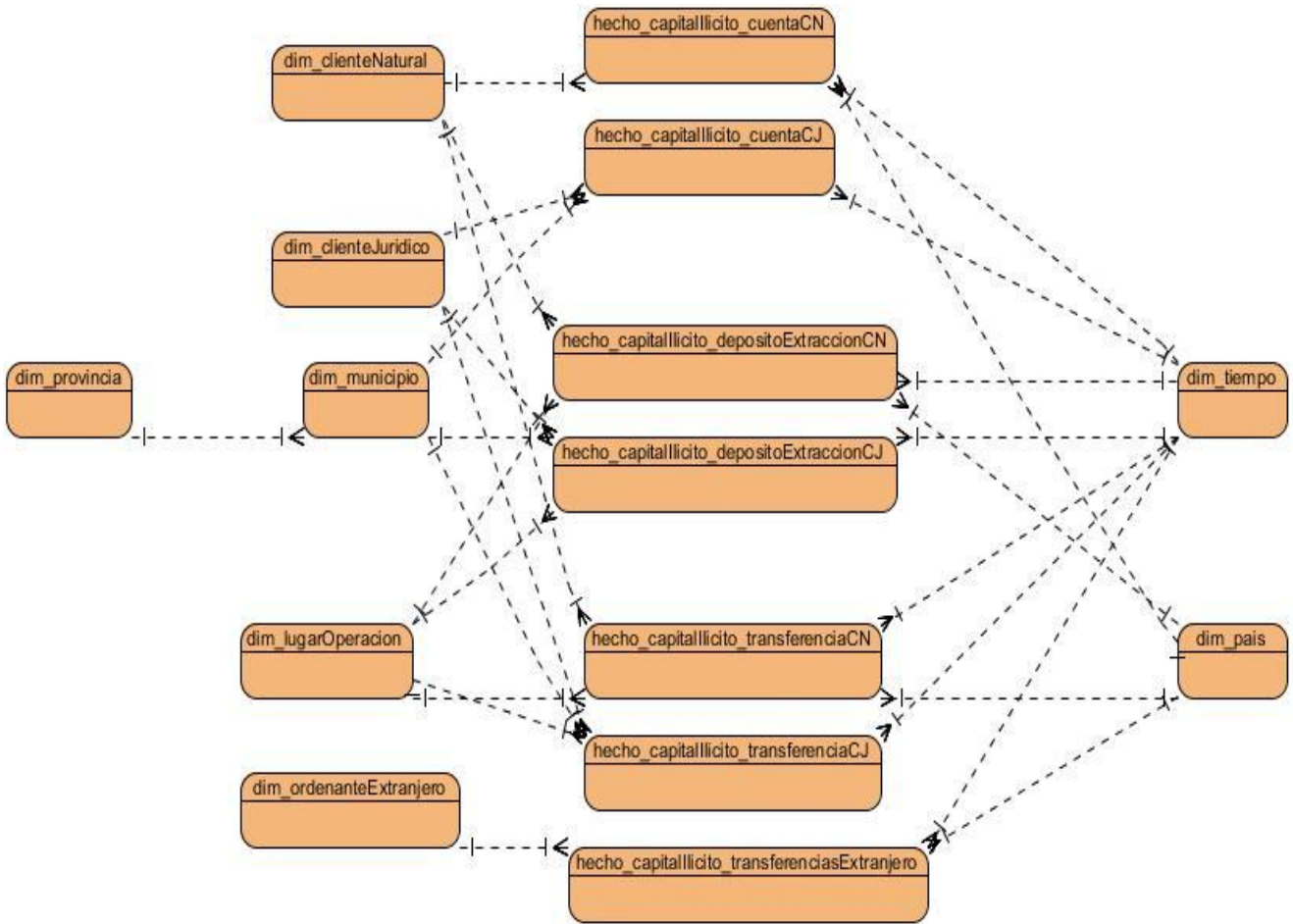


Figura # 16 Modelo de datos del MD

3. Conclusiones del capítulo

Siguiendo las tres primeras fases de la metodología de la UCID, se transitó por cada una de las tareas que proponen las mismas. La identificación de un total de 26 preguntas permitió conocer las necesidades de información del cliente así como definir los indicadores y perspectivas de análisis y el nivel de granularidad de cada perspectiva, lo que permitió conocer que el nivel de detalle al que se podrán realizar estudios a partir de los datos del MD es bastante específico. El diseño de las tablas de hechos y dimensiones y la unión entre las mismas, permitió la obtención del modelo físico del MD. Una vez generados los artefactos pertenecientes a estas fases es posible continuar con la siguiente etapa de la metodología, ya que el análisis y diseño realizado valida las necesidades de información y requerimientos del cliente de forma satisfactoria, lo que se corrobora con el acta de validación firmada por el mismo.

Capítulo 3 Implementación y validación del Mercado de Datos

1. Introducción

Este capítulo abarca la implementación y automatización de los procesos ETL como parte de la fase # 4 y última de la metodología de la UCID. Se abordan además, las pruebas de validación realizadas al MD, con el objetivo probar su correcto funcionamiento, la calidad de los datos que contiene, y verificar si el rendimiento actual del mismo se corresponde con las exigencias del cliente.

2. Fase # 4 Procesos ETL

Una vez creadas las tablas físicas del MD como resultado de las fases anteriores de la metodología aplicada, se puede proceder al llenado del MD mediante los procesos ETL. Para esto se realizará primeramente un mapeo de los datos fuentes hacia los destinos detallando claramente de dónde salen y hacia dónde se dirigen, luego se establecerán algunas restricciones y condiciones adicionales, con el fin de no obviar ningún aspecto relevante del negocio. Luego se procederá a la carga de los datos hacia las dimensiones y hechos según corresponda. Finalmente se diseñará y construirá la automatización de todo el proceso. (16)

2.1 Mapeo de los datos fuente al destino

Con el objetivo de evitar la pérdida de datos en el proceso ETL se hizo necesario llevar el control del mapeo de los mismos, documentando la relación existente entre los campos de la base de datos origen y su correspondiente en la base de datos destino. La tabla # 6 muestra un fragmento del mapeo de datos realizado. El resultado completo de este mapeo se puede visualizar en el anexo # 7.

Tabla # 6 Mapa lógico de los datos

Mapa lógico de datos					
Nombre BD Fuente: metro			Nombre BD Destino: TesisMD		
Nombre de la tabla	Nombre del campo	Tipo de dato	Nombre de la tabla	Nombre del campo	Tipo de dato
dim_clienteJurídico	código	varchar(50)	M_CLIENTGEN	COD_CLIENT	varchar(8)

dim_clienteJurídico	cancelado	bit	M_CLIENTGEN	CANCELADOS	bit
dim_clienteJurídico	tipoCliente	varchar(255)	M_CLIENTGEN C_SEGMEN	COD_SEGMEN NOM_SEGMEN	varchar(3) varchar(35)
dim_clienteJurídico	dirección	varchar(255)	M_CLIENTGEN	DIRECCION	varchar(254)
dim_clienteJurídico	procedencia	varchar(255)	-	-	-
dim_clienteNatural	código	varchar(620)	C_PERSON	COD_PAEXID+C OD_TIPID+NUM_ IDEPER	varchar(3) varchar(2) varchar(15)
dim_clienteNatural	estadoCivil	varchar(255)	C_PERSON	STATUS	varchar(1)
dim_clienteNatural	dirección	varchar(255)	C_PERSON	DIRECCION	varchar(254)
dim_clienteNatural	residencia	varchar(255)	C_PERSON	RESIDENCIA	varchar(254)
dim_clienteNatural	cancelado	bit	C_PERSON	CANCELADOS	bit

La tabla “dim_tiempo” del MD no se incluyó en el mapeo de datos realizado, ya que la misma no se construyó a partir de datos existentes en la fuente origen, sino que se generaron todos los campos que la integran, a partir de la definición de una fecha de inicio y fecha fin, que en el caso de esta primera solución serán: enero primero de 2004 y enero segundo de 2020 respectivamente, es necesario aclarar que la fecha de inicio se corresponde con la fecha a partir de la cual se mantienen salvas de las operaciones en el banco. Para la generación de este gran intervalo de fechas se utilizó la herramienta de la suite de SQL Server 2005: Analysis Services, la cual permite la creación de la dimensión tiempo de forma física en el MD, solo que genera los datos en idioma inglés, por lo que en el proceso de ETL de esta dimensión, fue necesario traducir todos los campos al idioma español.

2.2 Establecer condiciones adicionales y restricciones

Para la realización del proceso ETL es necesario tener en cuenta una serie de condiciones y restricciones adicionales, mediante las que se garantizará que los datos persistentes en el MD sean lo más confiables posible. Las restricciones establecidas son las siguientes:

1. Para efectuar el llenado de la tabla “hecho_capitalIlícito_cuentaCN” es necesario que los indicadores sean calculados por cliente natural, país y tiempo.
2. Para efectuar el llenado de la tabla “hecho_capitalIlícito_cuentaCJ” es necesario que los indicadores sean calculados por cliente jurídico, municipio y tiempo.
3. Para efectuar el llenado de la tabla “hecho_capitalIlícito_depósitoExtracciónCN” es necesario que los indicadores sean calculados por cliente natural, país, lugar de operación y tiempo.

4. Para efectuar el llenado de la tabla “hecho_capitalIlícito_depósitoExtracciónCJ” es necesario que los indicadores sean calculados por cliente jurídico, municipio, lugar de operación y tiempo.

5. Los saldos de las cuentas que serán guardados en las tablas “hecho_capitalIlícito_cuentaCN” y “hecho_capitalIlícito_cuentaCJ”, deberán ser calculados según el tipo de moneda de la cuenta, llevado a las tres monedas base del país: CUP, CUC y USD. En el caso de las cuentas que estén en otras monedas se les aplicará una tasa de cambio a la moneda base USD mediante la siguiente fórmula: saldo contable*0.20.

6. Para calcular el importe de las operaciones realizadas por un cliente en el día, se sumarán los importes de cada operación proveniente del origen, estos importes deben tener signo positivo si se ha efectuado un débito a la cuenta y signo negativo si se ha acreditado efectivo.

7. Para incluir el id de la dimensión tiempo en los hechos del área de análisis “Depósito-Extracción”, primero será necesario hacer corresponder la fecha en formato dd/mm/aaaa 12:00:00, que proviene del origen con la fecha en este mismo formato en la dimensión tiempo. En el caso de los hechos del área de análisis “Cuenta”, incluirán el id de la dimensión tiempo que se corresponda con la fecha actual del sistema.

8. Para realizar la carga de la dimensión municipio, será necesario efectuar la carga primeramente de la dimensión provincia, ya que entre ambas existe una jerarquía, donde municipio es el nivel más específico, por lo que contiene la llave primaria de la dimensión provincia.

9. En caso de que en la fuente de datos existan campos con valores nulos, vacíos o anómalos, éstos deberán ser transformados por el nuevo valor: “desconocido”.

2.3 Cargas incrementales de datos

El proceso de carga del MD se llevó a cabo a partir de la extracción de los datos de la fuente origen “metro”¹⁴, la transformación de los mismos para evitar valores faltantes o anómalos y finalmente la carga en las correspondientes tablas de dimensiones y hechos del MD. Para la implementación de los procesos ETL se utilizó la herramienta Microsoft SQL Server Integration Services 2005.

En la implementación de los procesos ETL se crearon dos paquetes de integración, el primero, “ETL MD carga inicial”, tiene como objetivo realizar la carga primera del MD, y el segundo “ETL MD carga diaria”, será el que se ejecute diariamente con el objetivo de realizar cargas

¹⁴ **metro**: Fragmento de prueba de la base de datos SABIC del Banco Metropolitano.

incrementales. Fue necesario implementar dos paquetes de integración, ya que en la carga inicial se incluyen tareas SQL para la inserción de tuplas destinadas al manejo de errores en las dimensiones: dim_país y dim_provincia, que no es prudente ejecutar por segunda vez, ya que se duplicarían estas tuplas sin sentido; por otra parte, la carga de la dimensión dim_tiempo, no se debe efectuar diariamente, solo será necesario incluirle nuevas fechas una vez que se aproxime la fecha fin establecida.

En los paquetes de integración para la carga inicial y final del MD, se realizó la implementación del escenario de flujo de datos de los procesos ETL de los hechos y dimensiones. En el caso del llenado de las dimensiones se extrajeron los datos de varias tablas del origen, los datos fueron sometidos a transformaciones, pues se hizo necesario validar que los datos que se insertaran en el MD no presentaran valores nulos, vacíos o anómalos. Estas transformaciones fueron de suma importancia, ya que en la fuente “metro”, existe un gran número de inconsistencias, valores faltantes y erróneos. Finalmente los datos transformados fueron insertados o actualizados en el destino (MD). La figura # 17 visualiza el escenario de flujo de datos del proceso ETL implementado para el llenado de la dimensión dim_clienteJurídico, el resto de los procesos para el llenado de las demás dimensiones se pueden encontrar en los anexos # 8, 9, 10, 11, 12 y 13.

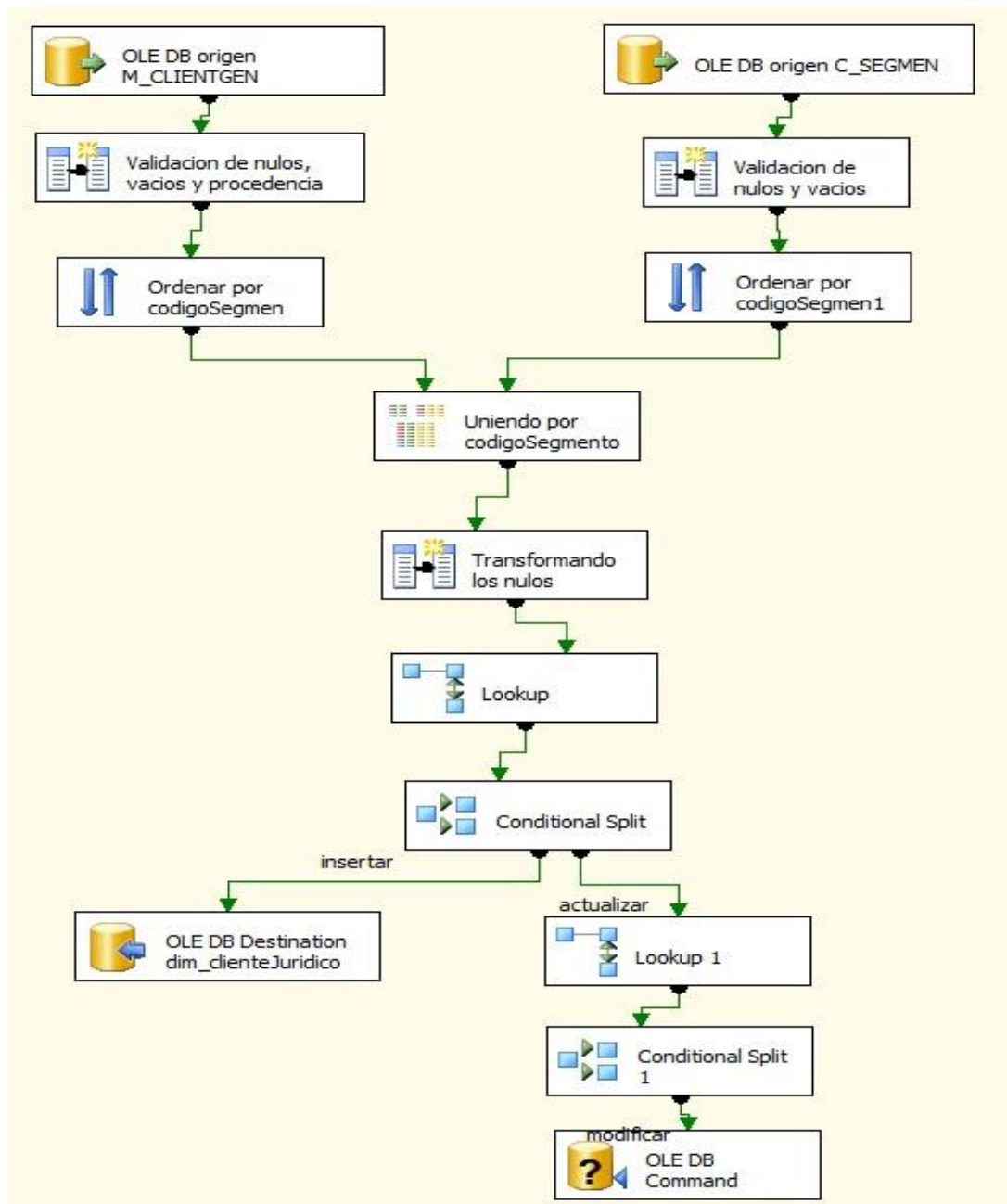


Figura # 17 Escenario de Flujo de Datos del proceso ETL de la tabla dim_clienteJurídico

Para la implementación de los procesos ETL en las tablas de hechos, se siguió básicamente el mismo proceder que en el caso de las tablas de dimensiones, solo que fue necesario además, realizar el cálculo de los indicadores o medidas, así como incorporar el id de las dimensiones que definen el hecho. EL anexo # 14 muestra el proceso ETL para el llenado de la tabla hecho_capitalIlicito_cuentaCJ.

En el caso de la implementación del escenario de control de flujo, se realizaron configuraciones similares en los paquetes de carga inicial y diaria, con la particularidad que en el primero se incluyeron tareas SQL para el manejo de tuplas erróneas en el flujo y se realiza la carga de la tabla dim_tiempo, mientras que en el paquete de carga diaria, no se incluye lo anterior. La figura # 18 muestra la configuración del escenario de control de flujo del paquete de carga inicial. En el anexo # 15 se puede visualizar la configuración para el paquete de carga diaria.

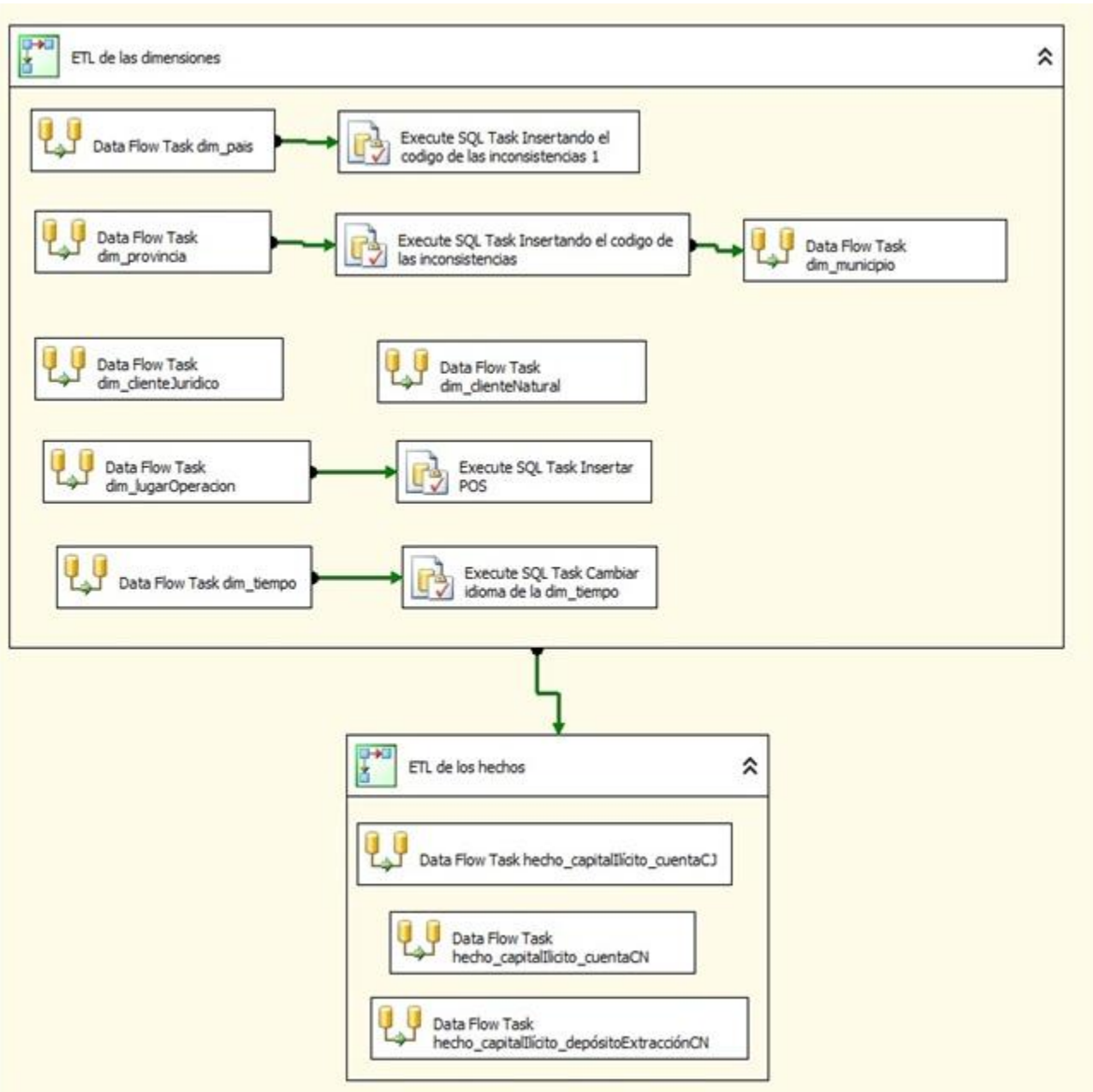


Figura # 18 Escenario de Control de flujo del proceso ETL para la carga inicial del MD

2.4 Automatización del proceso ETL

Una vez implementado el proceso ETL para cada una de las dimensiones y hechos, se pudo proceder a la automatización del mismo. Para esto se crearon dos trabajos independientes en el gestor de base de datos, uno destinado a la carga inicial del MD y otro a la carga diaria del mismo. La figura # 19 muestra los trabajos creados.

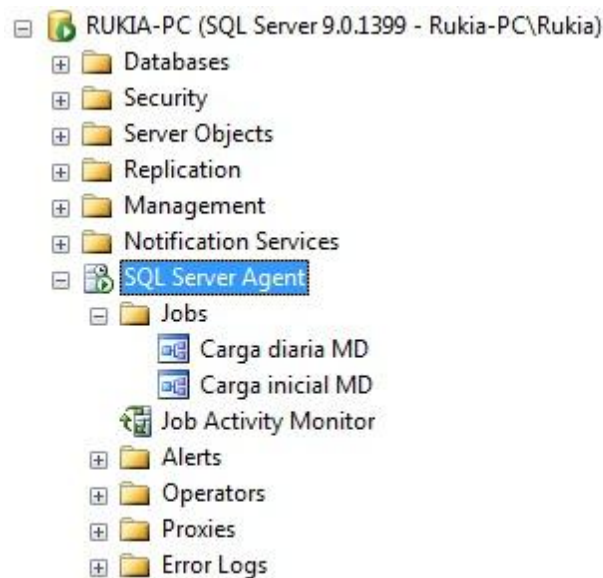


Figura # 19 Trabajos para la carga inicial y diaria del MD

En la configuración del trabajo creado para la carga diaria del MD se validó que el mismo se ejecutara diariamente a las 6.00 horas, esto responde a un requisito solicitado por el cliente. Por otra parte el trabajo creado para la carga inicial del MD se configuró a ejecutarse en una fecha determinada que será la fecha de carga histórica que definirá el cliente.

3. Validación del Mercado de Datos

La validación de todo sistema constituye el mecanismo mediante el cual se garantiza el correcto funcionamiento del mismo a partir de comprobar que se han cumplido con todas las metas y requisitos planteados por el cliente al inicio del ciclo de desarrollo, y verificando además, el rendimiento del sistema en un ambiente de explotación lo más similar posible al real. Para cumplir con lo anteriormente planteado se realizaron dos grupos de pruebas al MD, el primer grupo de pruebas se enfocó en validar el correcto funcionamiento de la solución, para esto se realizaron los siguientes tipos de pruebas: Pruebas de funcionalidad, pruebas de calidad de datos y pruebas de confiabilidad de los datos. El segundo grupo de pruebas se centró en verificar el correcto desempeño del MD, para esto se realizaron pruebas de rendimiento y estrés.

3.1 Pruebas para validar la calidad del Mercado de Datos

3.1.1 Pruebas de funcionalidad

Se realizaron pruebas de funcionalidad al MD con el objetivo de verificar la posibilidad de emplear los datos persistentes en el mismo para la realización de estudios, a partir de la selección indistinta de determinadas dimensiones y medidas. Para esto se construyeron cuatro cubos OLAP, mediante la utilización de la herramienta SQL Server Analysis Services 2005.

Primeramente se diseñaron los cubos: “CuboCuentaCJ”, “CuboCuentaCN”, “CuboDepósitoExtracciónCN” y “CuboDepósitoExtracciónCJ”, los cuales son una representación multidimensional de las tablas del MD: hecho_capitalIlicito_cuentaCJ, hecho_capitalIlicito_cuentaCN, hecho_capitalIlicito_depósitoExtracciónCN y hecho_capitalIlicito_depósitoExtracciónCJ respectivamente. La figura # 20 muestra el diseño del cubo “CuboCuentaCJ” con sus medidas y dimensiones asociadas, la representación de los cubos “CuboCuentaCN” y “CuboDepósitoExtracciónCN” se puede visualizar en los anexos # 16 y 17 respectivamente.

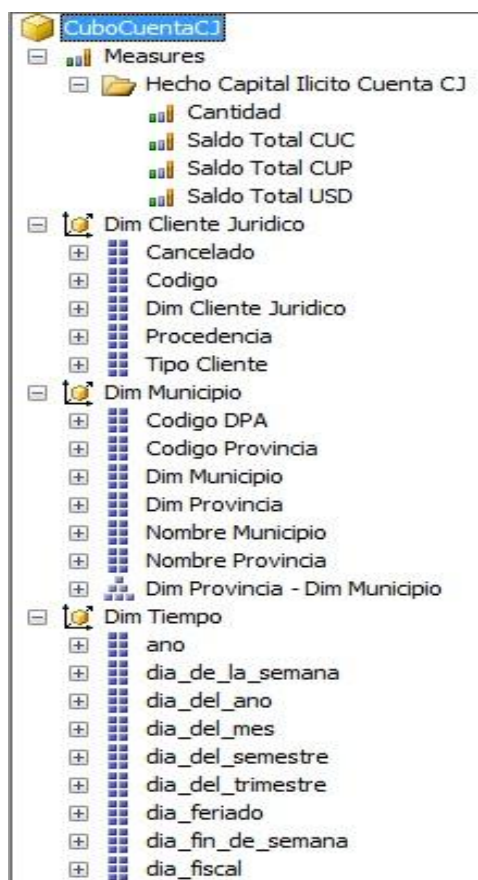


Figura # 20 Diseño del cubo OLAP “CuboCuentaCJ”

Luego de construidos los cubos OLAP fue posible visualizar la información contenida en el MD desde todos los niveles de detalle que se definió en el diseño. Actualmente es posible navegar por las diferentes estructuras según los datos que se deseen estudiar, éstos pueden ser visualizados, sin necesidad de realizar manualmente consultas SQL al MD. Los cubos “CuboCuentaCJ” y “CuboCuentaCN” se destinaron al análisis de la cantidad de cuentas y los saldos en las tres monedas base del país (CUC, CUP y USD) que poseen los clientes jurídicos y naturales respectivamente, mientras que los cubos “CuboDepósitoExtracciónCN” y “CuboDepósitoExtracciónCJ”, se destinaron al estudio de las operaciones de depósito y extracción diarias que realizan los clientes naturales y jurídicos desde determinados lugares de operación.

La figura # 21 muestra la navegación por los datos del MD mediante el cubo “CuboCuentaCJ”, visualizando la información referente a la cantidad de cuentas, saldos en CUC, CUP y USD según el tipo de cliente jurídico por municipio en el año 2013.

Drop Filter Fields Here										
		ano ▼				Grand Total				
		calendario 2013								
Nombre Municipio ▼	Tipo Cliente ▼	Cantidad	Saldo Total CUC	Saldo Total CUP	Saldo Total USD	Cantidad	Saldo Total CUC	Saldo Total CUP	Saldo Total USD	
<input type="checkbox"/> ARROYO NARANJO	ORGANISMO ESTATALES	3	-100	-500	0	3	-100	-500	0	
	Otras formas gestión no estat(OGE)	1	0	-500	0	1	0	-500	0	
	Trabajadores por cuenta propia(TCP)	19	-33	-67016.11	7	19	-33	-67016.11	7	
	Total	23	-133	-68016.11	7	23	-133	-68016.11	7	
<input type="checkbox"/> BOYEROS	Agricultores pequeños (APE)	3	-20.08	-198.35	4	3	-20.08	-198.35	4	
	Trabajadores por cuenta propia(TCP)	5	-71.25999	-5392.45	25	5	-71.25999	-5392.45	25	
	Total	8	-91.34	-5590.8	29	8	-91.34	-5590.8	29	
<input type="checkbox"/> CENTRO HABANA	Otras formas gestión no estat(OGE)	0	1	1	1	0	1	1	1	
	Trabajadores por cuenta propia(TCP)	11	15	-2435	15	11	15	-2435	15	
	Total	11	16	-2434	16	11	16	-2434	16	
<input type="checkbox"/> CERRO		64	-4713.23	-39626.07	12	64	-4713.23	-39626.07	12	
<input type="checkbox"/> COTORRO		1	9	-17670	9	1	9	-17670	9	
<input type="checkbox"/> DIEZ DE OCTUBRE		25	-26.67	-32734.66	27	25	-26.67	-32734.66	27	
<input type="checkbox"/> GUANABACOA		10	-17	-3702.09	3	10	-17	-3702.09	3	
<input type="checkbox"/> LA HABANA		2	0	0	0	2	0	0	0	
<input type="checkbox"/> LA HABANA DEL ESTE		15	14	-18783.34	14	15	14	-18783.34	14	
<input type="checkbox"/> LA HABANA VIEJA		0	12	12	12	0	12	12	12	
<input type="checkbox"/> LA LISA		28	10	-4804	20	28	10	-4804	20	
<input type="checkbox"/> MARIANAO		3	0	-2975	0	3	0	-2975	0	
<input type="checkbox"/> PLAYA		96	-12446.32	-34427.87	35	96	-12446.32	-34427.87	35	
<input type="checkbox"/> PLAZA DE LA REVOLUCION		15	-123	-7679.8	27	15	-123	-7679.8	27	
<input type="checkbox"/> REGLA		0	5	5	5	0	5	5	5	
<input type="checkbox"/> SAN MIGUEL DEL PADRON		0	14	14	14	0	14	14	14	
Grand Total		301	-17470.56	-238412.7	230	301	-17470.56	-238412.8	230	

Figura # 21 Visualización del cubo “CuboCuentaCJ”

3.1.2 Pruebas de calidad de datos

Dentro de los datos que se guardan en un MD no deben existir valores nulos o vacíos, ya que

dichos datos deben aportar información para realizar estudios posteriores. Con el objetivo de validar lo anterior, a pesar de que en el proceso ETL, se realizaron las transformaciones pertinentes en este sentido, se realizaron consultas SQL a cada una de las tablas de dimensiones y hechos del MD, con el fin de verificar que en las mismas no persistieran campos nulos o vacíos. Los resultados arrojados demuestran que no existe este tipo de valores en el MD, en comparación con el índice de valores nulos y vacíos presentes en la fuente de datos, el MD valida este aspecto en su totalidad. Ver figura # 22.

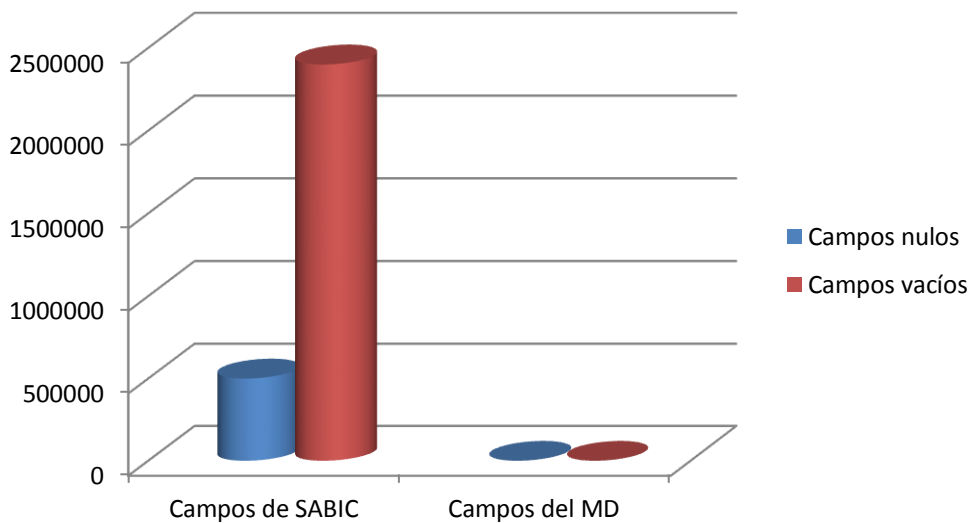


Figura # 22 Índices de valores nulos y vacíos presentes en la fuente de datos y en el MD

3.1.3 Pruebas de confiabilidad de los datos

Mediante los procesos ETL se pobló el MD, calculándose los indicadores según las dimensiones asociadas a los hechos. Sin embargo es necesario verificar cuán confiable son los datos que persisten en el MD, es importante verificar si los indicadores se han calculado correctamente y si la información que brinda el hecho en sí, es del todo confiable. Para esto se realizaron pruebas mediante la confrontación de los datos contenidos en la fuente origen y los del MD.

Con el objetivo de probar lo anteriormente planteado se consultaron los datos de la tabla hecho_capitalIlícito_cuentaCJ y se seleccionaron varias tuplas de la misma con el fin de confrontar este resultado con el que se obtendrá de la fuente origen. Una de las tuplas seleccionadas brinda la siguiente información:

Tupla: El cliente con id igual a “3099” en la dimensión dim_clienteJurídico, perteneciente al municipio con id igual a “1445” en la dimensión dim_municipio, en la fecha con id igual a “159 789” en la dimensión dim_tiempo, posee un total de dos cuentas bancarias, un saldo total en CUC y

CUP de 50.00 y 500.00 respectivamente.

La figura # 23 visualiza este resultado.

cantidad	saldoTotalUSD	saldoTotalCUC	saldoTotalCUP	dim_municipioid	dim_municipio	dim_tiempoid	dim_tiempo	dim_clienteJuridico	dim_clienteJuridico
2	0	-50	-500	1445		159789		3099	
0	0	0	0	1459		159789		3360	
0	0	0	0	1451		159789		3359	
0	0	0	0	1457		159789		3358	
2	0	0	0	1448		159789		3357	
0	0	0	0	1452		159789		3356	
0	0	0	0	1459		159789		3355	
0	0	0	0	1459		159789		3354	
0	0	0	0	1459		159789		3353	
1	0	0	-600	1455		159789		3352	
0	0	0	0	1459		159789		3361	
0	0	0	0	1459		159789		3351	
2	0	-46,26	-4717,45	1449		159789		3349	

Figura # 23 Tupla seleccionada de la tabla hecho_capitalilícito_cuentaCJ

Luego se consultó en la tabla dim_clienteJurídico en busca del código del cliente con id igual a “3099”. El resultado de esta consulta se puede visualizar en el anexo # 18.

El código obtenido se corresponde con el código de la tabla M_CLIENTGEN de la base de datos origen “metro”, con este código fue posible buscar en la tabla origen M_CUECLIGEN, que registra la relación cuenta-cliente, los códigos de las cuentas asociadas al cliente en cuestión. El resultado de esta búsqueda arrojó que el cliente posee dos cuentas bancarias. Ver anexo # 19.

Conociendo los códigos de las cuentas que posee el cliente en análisis solo resta buscar el saldo contable y la moneda que tienen estas cuentas en la tabla origen M_MAYOR, que precisamente guarda la información referente al estado de las cuentas de cada cliente del banco. El resultado de la consulta realizada arrojó que el cliente posee una cuenta en CUP con un saldo de 500.00 pesos y otra cuenta en CUC con un saldo de 50.00 CUC. La figura # 24 muestra el resultado de las consultas.

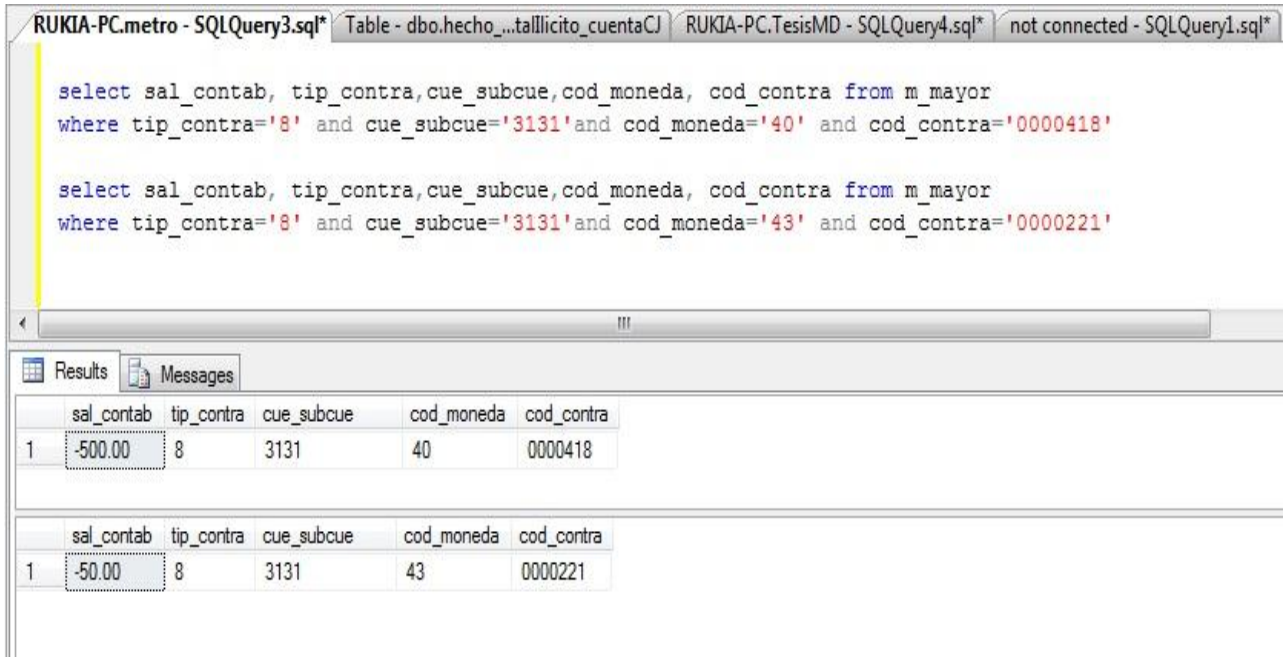


Figura # 24 Consultas y resultados de la búsqueda del saldo contable de las cuentas que posee el cliente en estudio

Como se pudo observar las consultas realizadas a la fuente de datos arrojaron los mismos resultados que los datos persistentes en el MD. Este proceso de verificación se realizó en varias tuplas en los restantes hechos arrojando excelentes resultados.

3.2 Pruebas para validar el rendimiento del Mercado de Datos

3.2.1 Pruebas de rendimiento

Dos de los requisitos que el cliente solicitó al inicio del ciclo de desarrollo del MD, fueron: la necesidad de que la solución permitiera la conexión y consulta paralela a diez usuarios, sin exceder, como tiempo de respuesta de las consultas, los 12 segundos. Con el objetivo de verificar el cumplimiento de estos requisitos se utilizó la herramienta Apache JMeter anteriormente descrita en el capítulo primero.

La prueba se realizó bajo las siguientes condiciones:

- Herramienta a probar: Mercado de datos para el Banco Metropolitano.
- Tipo de operación a realizar: Consulta SQL.
- Servicios de SQL Server 2005 y SQL Server Integration Services 2005 en ejecución.
- Hardware del Servidor de aplicaciones: 4 GB de memoria RAM, 720 GB como capacidad del disco duro.

- Software en el Servidor: SO Microsoft Windows Seven Ultimate, SQL Server 2005 como gestor de base de datos.
- La tabla seleccionada para la prueba se encontraba poblada con 1000 tuplas.

La prueba se configuró para un total de 10 conexiones realizando una consulta a la semejanza de las que se le efectuarán al MD en el entorno de trabajo real: “select * from hecho_capitalIlícito_cuentaCJ where saldoTotalCUC<=-1000 or saldoTotalUSD<=-1000 or saldoTotalCUP<=-5000”, esta prueba arrojó los resultados que se pueden apreciar en la figura # 25.

Label	# Muestras	Media	Mediana	Linea de 90%	Mín	Máx	% Error	Rendimiento
Petición de muestreo	10	0	1	1	0	1	100,00%	10,9/sec
TOTAL	10	0	1	1	0	1	100,00%	10,9/sec

Figura # 25 Tiempo de ejecución de una consulta al MD usando la herramienta JMeter

3.2.2 Pruebas de estrés

Otro aspecto de gran importancia a validar es cómo responde el MD ante cargas de grandes volúmenes de datos. Para verificar el correcto desempeño de la solución en este sentido, se utilizó la herramienta Data Generator 2005 para SQL Server. Se efectuó la carga de todas las tablas del MD, simulando la carga histórica del mismo. La tabla # 7 muestra la relación “Tabla del MD-tuplas insertadas”.

Tabla # 7 Relación “Tabla del MD-tuplas insertadas”

Tabla del MD	Tuplas insertadas
dim_país	300
dim_provincia	20
dim_municipio	180
dim_tiempo	11692
dim_lugarOperación	500
dim_clienteNatural	1000000
dim_clienteJurídico	1000
hecho_capitalIlícito_cuentaCJ	2000
hecho_capitalIlícito_cuentaCN	2000000

hecho_capitalIlicito_depósitoExtracciónCN	10000000
hecho_capitalIlicito_depósitoExtracciónCJ	10000000

La herramienta arrojó resultados alentadores, ya que se insertaron todas las tuplas ordenadas y no se generaron errores en el proceso. Visualizar figura # 26.



Figura # 26 Resultado de la prueba de estrés

4. Conclusiones del capítulo

La herramienta seleccionada para la realización de los procesos ETL permitió la correcta población de las tablas del MD. En la automatización de dichos procesos quedaron validados los requisitos solicitados por el cliente.

Las pruebas realizadas al MD arrojaron los resultados esperados, en las mismas se pudo comprobar el correcto funcionamiento de la solución, así como la inexistencia de valores nulos o vacíos y que los datos persistentes en el MD son del todo confiables. La verificación de los tiempos de respuestas de la solución y la capacidad de almacenar grandes cúmulos de datos, permitió comprobar que el MD cumple totalmente con los requisitos solicitados por el cliente.

Conclusiones generales

En aras de llevar a cabo la construcción de un MD para la gestión de la información sobre las operaciones con tarjetas bancarias en el Banco Metropolitano, facilitando la entrada de datos para el proceso de detección de fraude financiero, se cumplieron los objetivos trazados:

- El análisis del estado del arte referente a los AD, permitió la selección de las herramientas y metodologías utilizadas en la construcción del MD.
- EL análisis del negocio realizado posibilitó la identificación de un total de veintiséis preguntas facilitando la identificación de las necesidades de información del cliente.
- La elaboración de los modelos conceptuales y modelo de datos del MD permitió verificar que el diseño satisface los intereses y necesidades de información del cliente.
- La herramienta para los procesos ETL seleccionada posibilitó la exitosa población de las tablas del MD.
- Las pruebas realizadas permitieron verificar que se cumplieron con las necesidades de información y requisitos solicitados por el cliente.

Recomendaciones

- Finalizar la implementación del área de análisis “Transferencia” una vez que los funcionarios del Banco Metropolitano hayan creado un camino para obtener los datos pertenecientes a la dimensión y hechos de dicha área.
- Incorporar la implementación de reglas de seguridad al MD, a partir de la especificación de permisos de acceso y modificación sobre la solución.

Referencias Bibliográficas

1. **Ing. Bernabeu, Ricardo Dario.** *DATA WAREHOUSING: Investigación y Sistematización de Conceptos – HEFESTO: Metodología propia para la Construcción de un Data Warehouse.* Córdoba, Argentina : s.n., 2007.
2. **Kimball, Ralph.** *El juego de herramientas del Almacén de Datos.* 2002.
3. AGGNITIO. [En línea] [Citado el: 25 de noviembre de 2013.] <http://www.definicion-de.es/fraude-financiero>.
4. Banco Central de Cuba. [En línea] 2009. [Citado el: 10 de diciembre de 2012.] http://www.bc.gov.cu/espanol/bancos_comerciales.asp.
5. *Enriqueciendo el sistema contable en el Banco Central de Cuba.* **Sanabria, Lourdes Cerezal Tamargo y Jorge Torres.** 2002, págs. 24,25.
6. **Inmon, W.H.** *Building the Data Warehouse.* 1. 2002.
7. **Kimball, Ralph.** *El juego de herramientas ETL para almacenes de datos.* 2004.
8. **Departamento Central de Sistema de Gestión de Base de Datos, Universidad de las Ciencias Informáticas.** Entorno Virtual de Aprendizaje. [En línea] 2012. [Citado el: 10 de enero de 2013.] <http://eva.uci.cu/SGBD/Extracción, transformación y carga.pdf>.
9. **Rainardi, Vincent.** *Building a Data Warehouse.* New York : s.n., 2008.
10. **Departamento Central de Sistemas de Gestión de Base de Datos, Universidad de las Ciencias Informáticas.** Entorno Virtual de Aprendizaje. [En línea] 2011. [Citado el: 10 de enero de 2013.] <http://eva.uci.cu/SGBD/Almacenes de Datos.pdf>.
11. **Zhulkovsky, Vitali.** Bnak System & Technology. [En línea] [Citado el: 12 de enero de 2013.] www.banktech.com/business-intelligence/ing-diba-upgrades-to-sas-analytics-platf.
12. BI-LATINO.com. [En línea] 20 de noviembre de 2008. [Citado el: 10 de enero de 2013.] <http://www.bi-spain.com/articulo/64211/data-warehouse/banca-y-finanzas/el-banco-de-guayaquil-consolida-sus-fuentes-de-datos-gracias-al-datawarehouse-con-enfoque-financiero-de-finanware>.
13. BI-LATINO.com. [En línea] 20 de noviembre de 2008. [Citado el: 10 de enero de 2013.] <http://www.bi-spain.com/articulo/65165/calidad-de-datos/banca-y-finanzas/el-banco-pichincha>

optimiza-el-conocimiento-y-gestion-de-sus-clientes-con-herramientas-de-data-mining-y-data-warehouse.

14. **Hernández, Yanisbel González.** *Propuesta de metodología para la construcción de almacenes de datos en DATEC.* La Habana : s.n., 2012.
15. **Rivadera, Ing.Gustavo R.** *La metodología de Kimball para el diseño de almacenes de datos.* 2010.
16. **UCID, Colectivo de autores de.** *Metodología Híbrida para la construcción de un Data Warehouse.* La Habana : s.n., 2010.
17. Visual Paradigm. [En línea] [Citado el: 3 de enero de 2013.] <http://www.visual-paradigm.com/aboutus/newsreleases/vpuml80.jsp>.
18. **Microsoft.** Microsoft SQL Server 2005. [En línea] 2005. [Citado el: 7 de enero de 2013.] <http://msdn.microsoft.com/en-us/library/bb545450.aspx>.
19. —. Microsoft SQL Server 2005. [En línea] 2005. [Citado el: 7 de enero de 2013.] <http://msdn.microsoft.com/en-us/sqlserver/bb671084.aspx>.
20. —. Microsoft SQL Server. [En línea] 2005. [Citado el: 7 de enero de 2013.] <http://msdn.microsoft.com/en-us/library/ms141085.aspx>.

Bibliografía


1. **Ing. Bernabeu, Ricardo Dario.** *DATA WAREHOUSING: Investigación y Sistematización de Conceptos – HEFESTO: Metodología propia para la Construcción de un Data Warehouse.* Córdoba, Argentina : s.n., 2007.
2. **Kimball, Ralph.** *El juego de herramientas del Almacén de Datos.* 2002.
3. AGGNITIO. [En línea] [Citado el: 25 de noviembre de 2013.] <http://www.definicion-de.es/fraude-financiero>.
4. Banco Central de Cuba. [En línea] 2009. [Citado el: 10 de diciembre de 2012.] http://www.bc.gov.cu/espanol/bancos_comerciales.asp.
5. *Enriqueciendo el sistema contable en el Banco Central de Cuba.* **Sanabria, Lourdes Cereza Tamargo y Jorge Torres.** 2002, págs. 24,25.
6. **Inmon, W.H.** *Building the Data Warehouse.* 1. 2002.
7. **Kimball, Ralph.** *El juego de herramientas ETL para almacenes de datos.* 2004.
8. **Departamento Central de Sistema de Gestión de Base de Datos, Universidad de las Ciencias Informáticas.** Entorno Virtual de Aprendizaje. [En línea] 2012. [Citado el: 10 de enero de 2013.] <http://eva.uci.cu/SGBD/Extracción, transformación y carga.pdf>.
9. **Rainardi, Vincent.** *Building a Data Warehouse.* New York : s.n., 2008.
10. **Departamento Central de Sistemas de Gestión de Base de Datos, Universidad de las Ciencias Informáticas.** Entorno Virtual de Aprendizaje. [En línea] 2011. [Citado el: 10 de enero de 2013.] <http://eva.uci.cu/SGBD/Almacenes de Datos.pdf>.
11. **Zhulkovsky, Vitali.** Bnak System & Technology. [En línea] [Citado el: 12 de enero de 2013.] www.banktech.com/business-intelligence/ing-diba-upgrades-to-sas-analytics-platf.
12. BI-LATINO.com. [En línea] 20 de noviembre de 2008. [Citado el: 10 de enero de 2013.] <http://www.bi-spain.com/articulo/64211/data-warehouse/banca-y-finanzas/el-banco-de-guayaquil-consolida-sus-fuentes-de-datos-gracias-al-datawarehouse-con-enfoque-financiero-de-finanware>.
13. BI-LATINO.com. [En línea] 20 de noviembre de 2008. [Citado el: 10 de enero de 2013.] <http://www.bi-spain.com/articulo/65165/calidad-de-datos/banca-y-finanzas/el-banco-pichincha>.

optimiza-el-conocimiento-y-gestion-de-sus-clientes-con-herramientas-de-data-mining-y-data-warehouse.

14. **Hernández, Yanisbel González.** *Propuesta de metodología para la construcción de almacenes de datos en DATEC.* La Habana : s.n., 2012.
15. **Rivadera, Ing.Gustavo R.** *La metodología de Kimball para el diseño de almacenes de datos.* 2010.
16. **UCID, Colectivo de autores de.** *Metodología Híbrida para la construcción de un Data Warehouse.* La Habana : s.n., 2010.
17. Visual Paradigm. [En línea] [Citado el: 3 de enero de 2013.] <http://www.visual-paradigm.com/aboutus/newsreleases/vpuml80.jsp>.
18. **Microsoft.** Microsoft SQL Server 2005. [En línea] 2005. [Citado el: 7 de enero de 2013.] <http://msdn.microsoft.com/en-us/library/bb545450.aspx>.
19. —. Microsoft SQL Server 2005. [En línea] 2005. [Citado el: 7 de enero de 2013.] <http://msdn.microsoft.com/en-us/sqlserver/bb671084.aspx>.
20. —. Microsoft SQL Server. [En línea] 2005. [Citado el: 7 de enero de 2013.] <http://msdn.microsoft.com/en-us/library/ms141085.aspx>.
21. **Davis, G. y Olsón.** *Management Information Systems: Conceptual foundations, Structure and Development.* 2. Nueva York : s.n., 1985.
23. **Ashwani, Nanda.** *Hands-on Microsoft SQL Server 2008 Integration Services.* New York : s.n., 2011.
24. **Walsh Earp, Richard.** *Essential SQL on SQL Server 2008.* New York : s.n., 2011.

Anexos

Anexo # 1 Acta de validación del Modelo Conceptual del Mercado de Datos

 Acta de validación

Se valida que el proyecto: **SAGEB** ha concluido el proceso de **construcción del modelo conceptual del Mercado de Datos** como parte de la fase de **Análisis de los requerimientos**, de forma satisfactoria y obteniendo los artefactos establecidos.

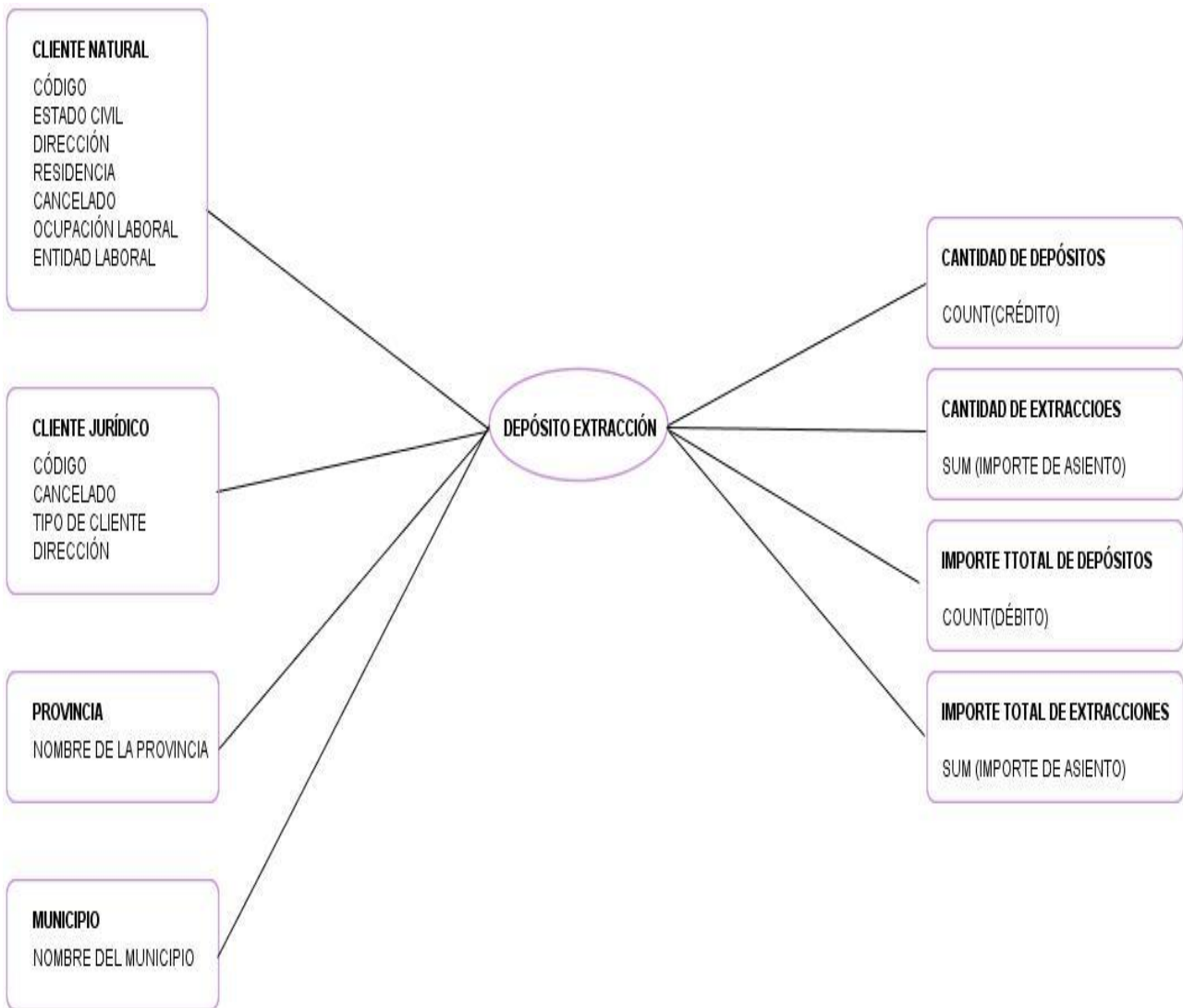
Para que así conste se emite este documento a los 14 días del mes de marzo del año 2013.

Yadira Pérez López
Firmado: (responsable de la actividad)

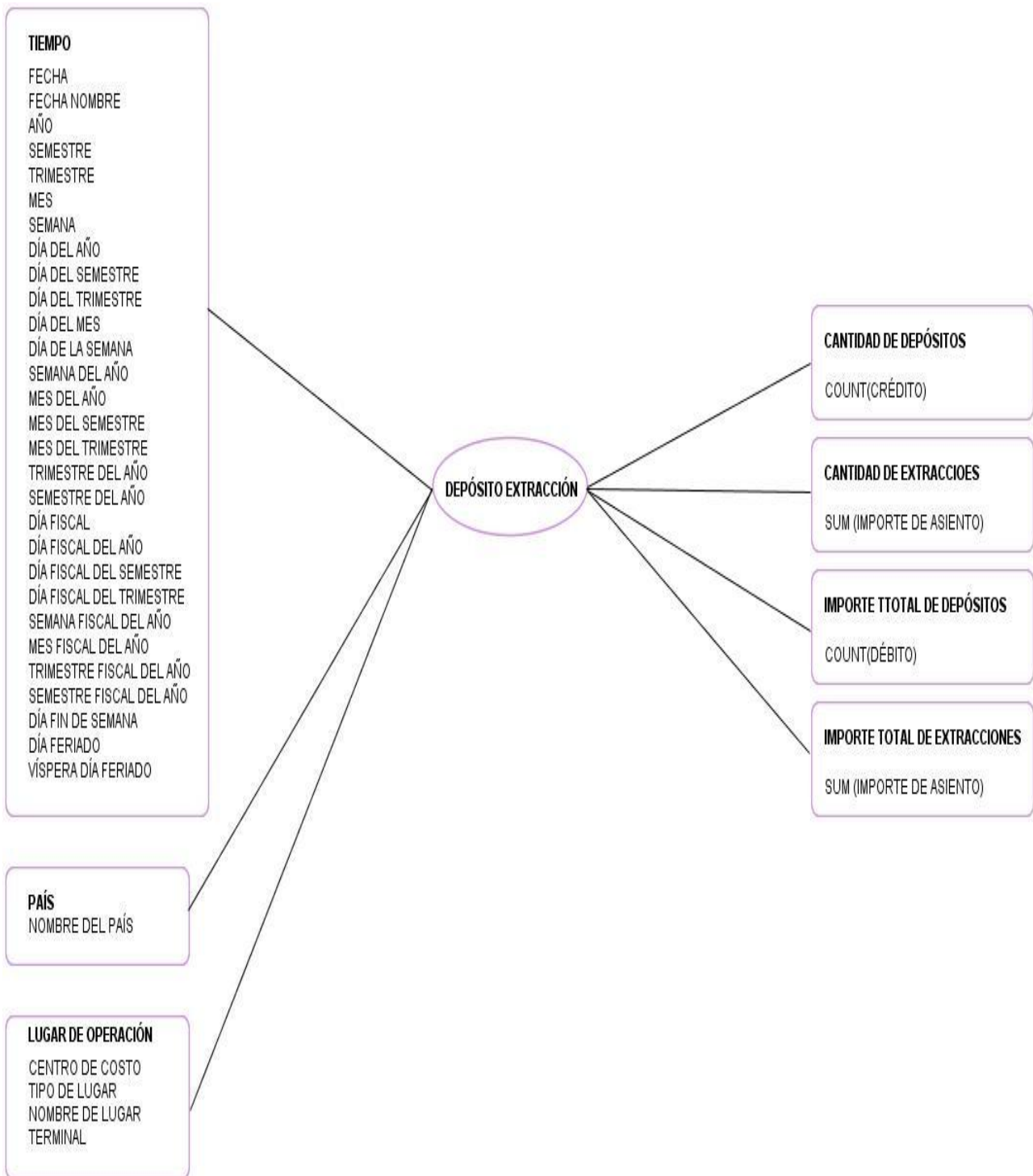
Firmado: (cliente)
ANÍE BLANCO PARDO
Antonio Arroyo Alfaro

Página 2 de 2

Anexo # 2 Primera parte del Modelo Conceptual Ampliado del área de análisis “Depósito - Extracción”



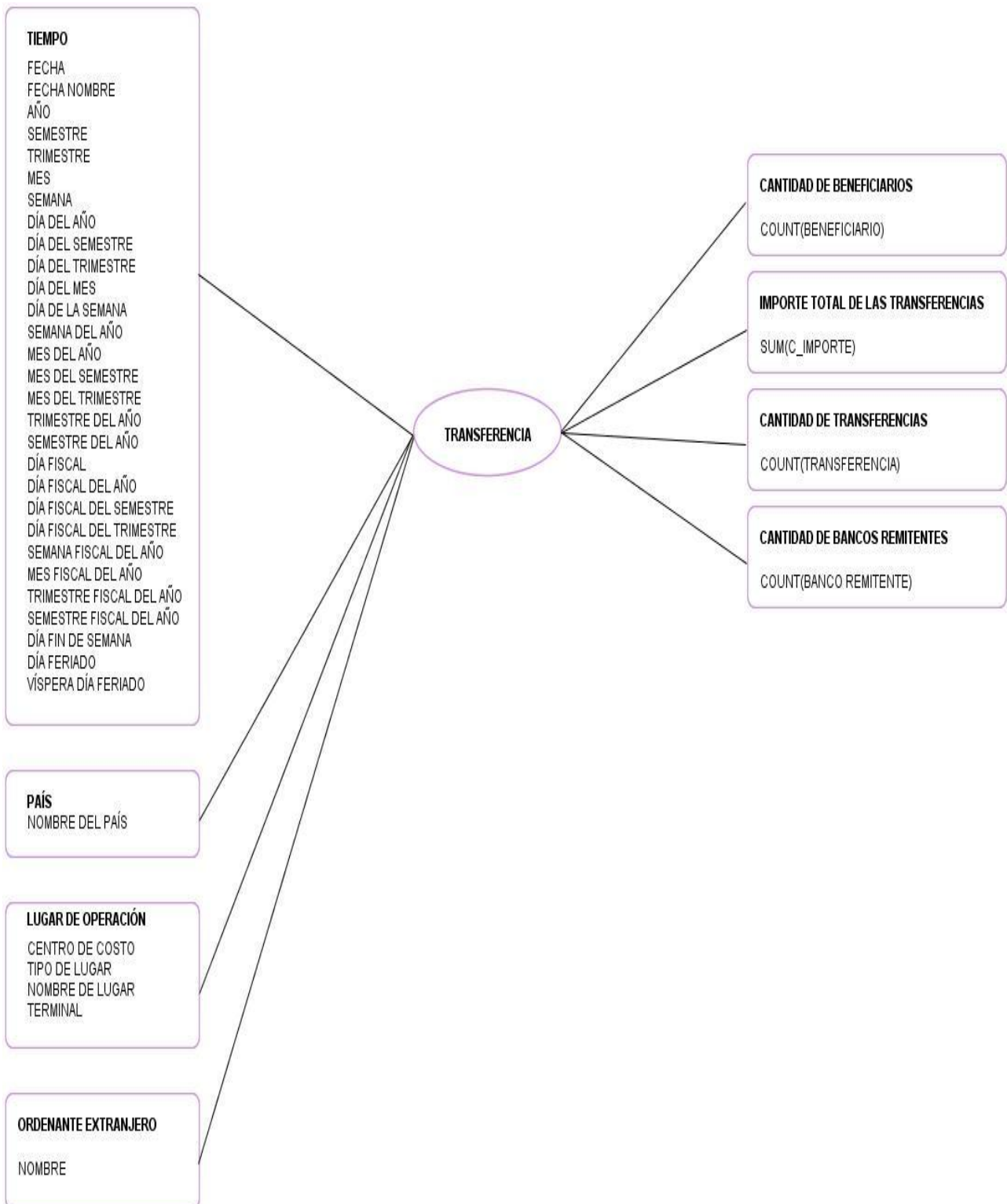
Anexo # 3 Segunda parte del Modelo Conceptual Ampliado del área de análisis “Depósito - Extracción”



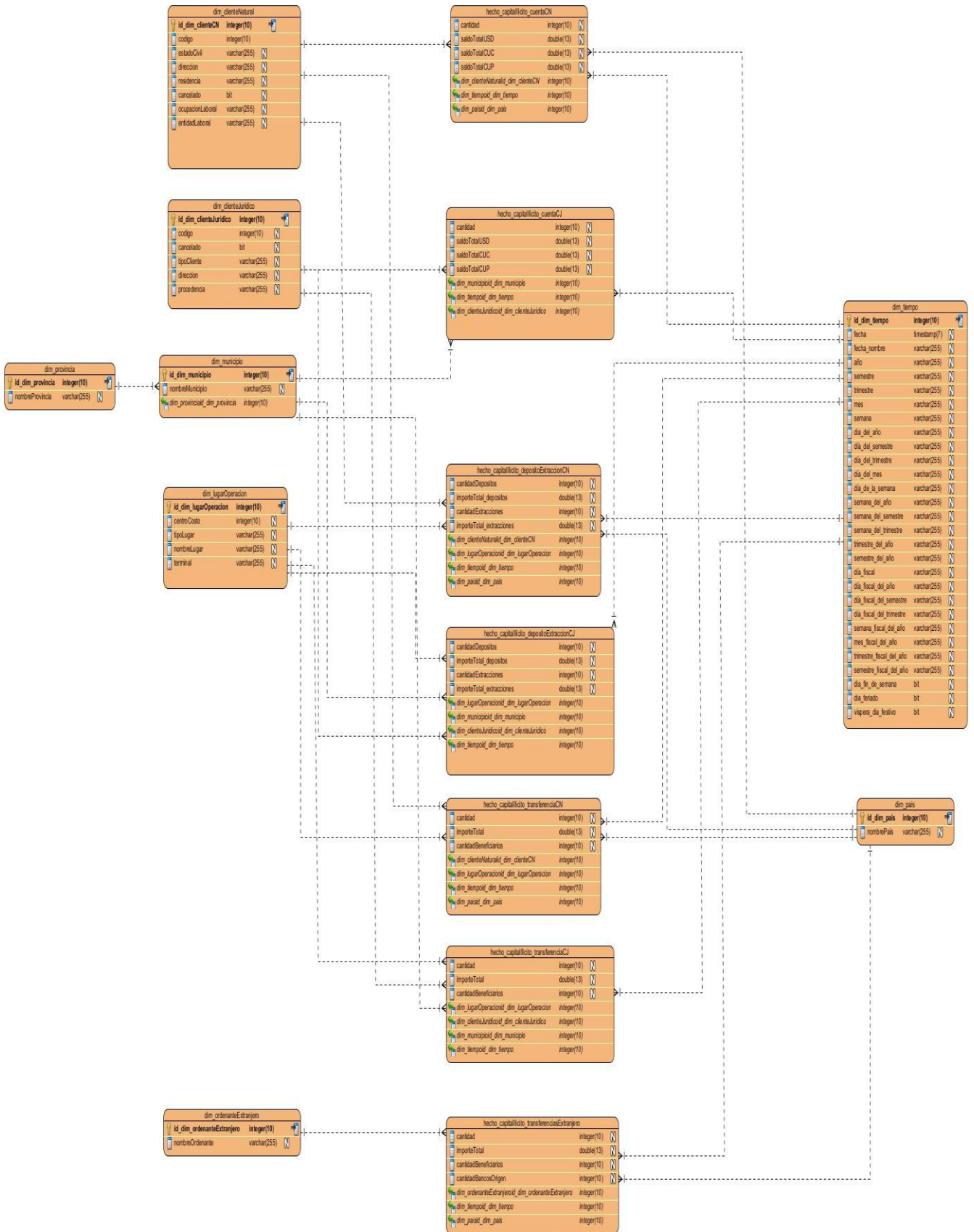
Anexo # 4 Primera parte del Modelo Conceptual Ampliado del área de análisis “Transferencia”



Anexo # 5 Segunda parte del Modelo Conceptual Ampliado del área de análisis “Transferencia”



Anexo # 6 Modelo de datos del Mercado de Datos

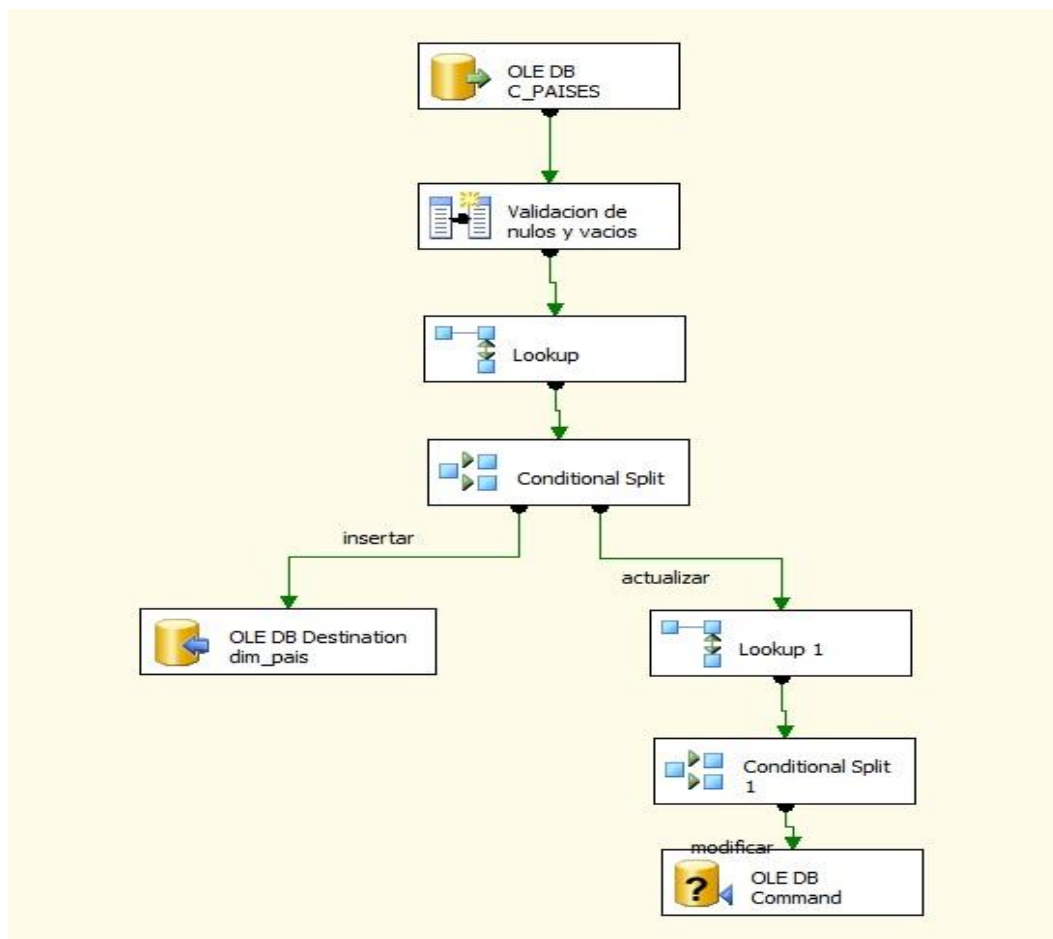


Anexo # 7 Mapa lógico de los datos

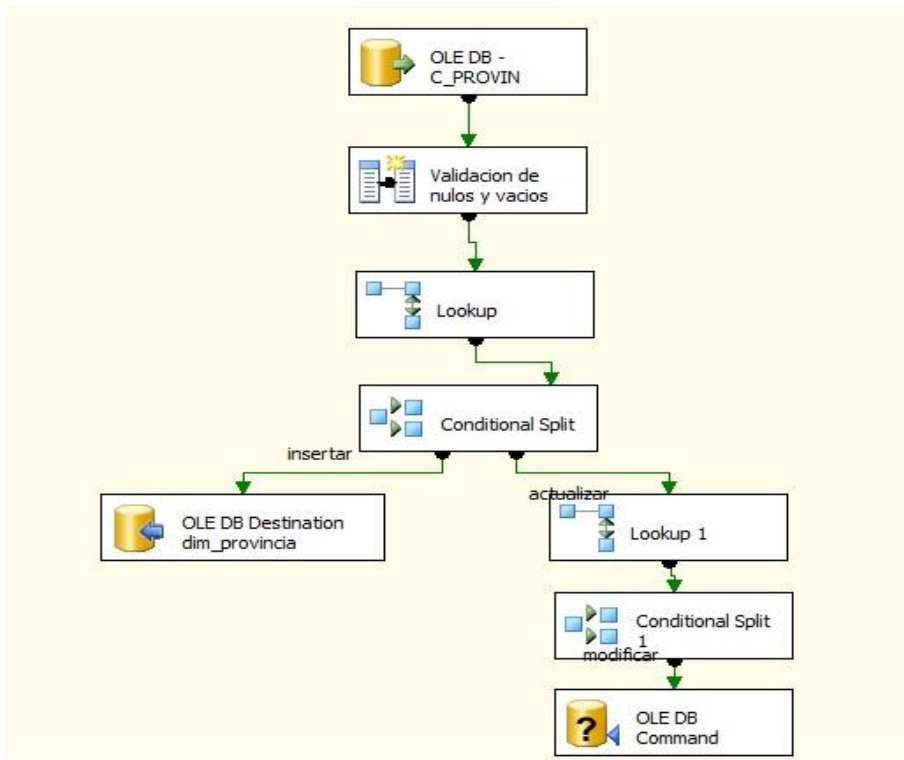
Mapa lógico de datos					
Nombre BD Fuente: metro			Nombre BD Destino: TesisMD		
Nombre de la tabla	Nombre del campo	Tipo de dato	Nombre de la tabla	Nombre del campo	Tipo de dato
dim_clienteJurídico	código	varchar(50)	M_CLIENTGEN	COD_CLIENT	varchar(8)
dim_clienteJurídico	Cancelado	bit	M_CLIENTGEN	CANCELADOS	bit
dim_clienteJurídico	tipoCliente	varchar(255)	M_CLIENTGEN C_SEGMEN	COD_SEGMEN NOM_SEGMEN	varchar(3) varchar(35)
dim_clienteJurídico	dirección	varchar(255)	M_CLIENTGEN	DIRECCION	varchar(254)
dim_clienteJurídico	procedencia	varchar(255)	-	-	-
dim_clienteNatural	código	varchar(620)	C_PERSON	COD_PAEXID+C OD_TIPID+NUM_ IDEPER	varchar(3) varchar(2) varchar(15)
dim_clienteNatural	estadoCivil	varchar(255)	C_PERSON	STATUS	Varchar(1)
dim_clienteNatural	dirección	varchar(255)	C_PERSON	DIRECCION	varchar(254)
dim_clienteNatural	residencia	varchar(255)	C_PERSON	RESIDENCIA	varchar(254)
dim_clienteNatural	cancelado	bit	C_PERSON	CANCELADOS	bit
dim_clienteNatural	ocupacionLa boral	varchar(255)	C_PERSON C_CATOCU	COD_OCUP NOM_OCUP	varchar(2) varchar(50)
dim_clienteNatural	entidadLabor al	varchar(255)	C_PERSON	ENTIDAD	varchar(30)
dim_país	nombrePaís	varchar(255)	C_PAISES	NOM_PAIS	varchar(35)
dim_país	códigoPaís	varchar(6)	C_PAISES	COD_PAIS	varchar(3)
dim_provincia	nombreProvi ncia	varchar(6)	C_PROVIN	NOM_PROVIN	varchar(35)
dim_provincia	códigoProvin cia	varchar(255)	C_PROVIN	COD_PROVIN	varchar(2)
dim_municipio	nombreMuni cipio	varchar(255)	C_DPA	NOM_MUNIC	varchar(25)
dim_municipio	códigoDPA	varchar(6)	C_DPA	COD_DPA	varchar(4)
dim_lugarOperación	centroCosto	varchar(50)	C_CCOSTO C_CAJERO	COD_CCOSTO COD_CCOSTO	varchar(3)
dim_lugarOperación	tipoLugar	varchar(255)	M_HISTOR	COD_TRANSA	varchar(255)
dim_lugarOperación	nombreLugar	varchar(255)	C_CCOSTO C_CAJERO	NOM_CCOSTO NOM_CAJERO	varchar(45) varchar(50)

dim_lugarOperación	terminal	varchar(255)	C_CCOSTO C_CAJERO	COD_CCOSTO COD_CAJERO	varchar(3) varchar(3)
hecho_capitalIlícito_cuentaCN/CJ	saldoTotalUSD	decimal(20, 2)	M_MAYOR	SAL_CONTAB COD_MONEDA	decimal(20, 2) varchar(2)
hecho_capitalIlícito_cuentaCN/CJ	saldoTotalCUC	decimal(20, 2)	M_MAYOR	SAL_CONTAB COD_MONEDA	decimal(20, 2) varchar(2)
hecho_capitalIlícito_cuentaCN/CJ	saldoTotalCUP	decimal(20, 2)	M_MAYOR	SAL_CONTAB COD_MONEDA	decimal(20, 2) varchar(2)
hecho_capitalIlícito_depósitoExtracciónCN/CJ	importeTotal_extracciones	decimal(16, 2)	M_HISTOR	IMP_ASIENT	decimal(16, 2)
hecho_capitalIlícito_depósitoExtracciónCN/CJ	importeTotal_depósitos	decimal(16, 20029)	M_HISTOR	IMP_ASIENT	decimal(16, 2)

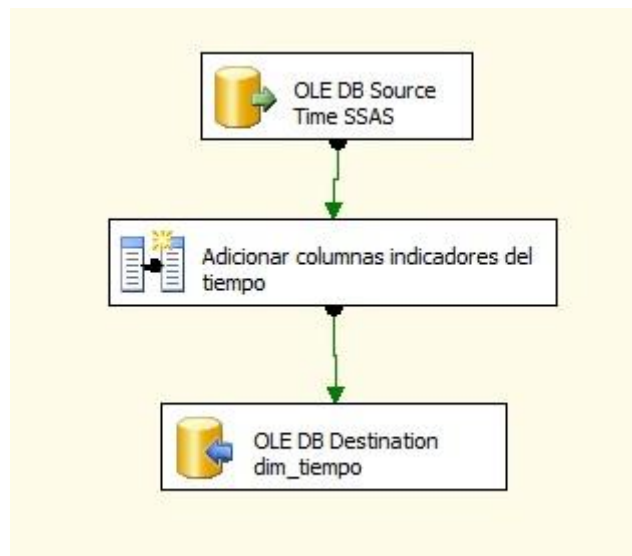
Anexo # 8 Escenario de Flujo de Datos del proceso ETL de la tabla dim_país



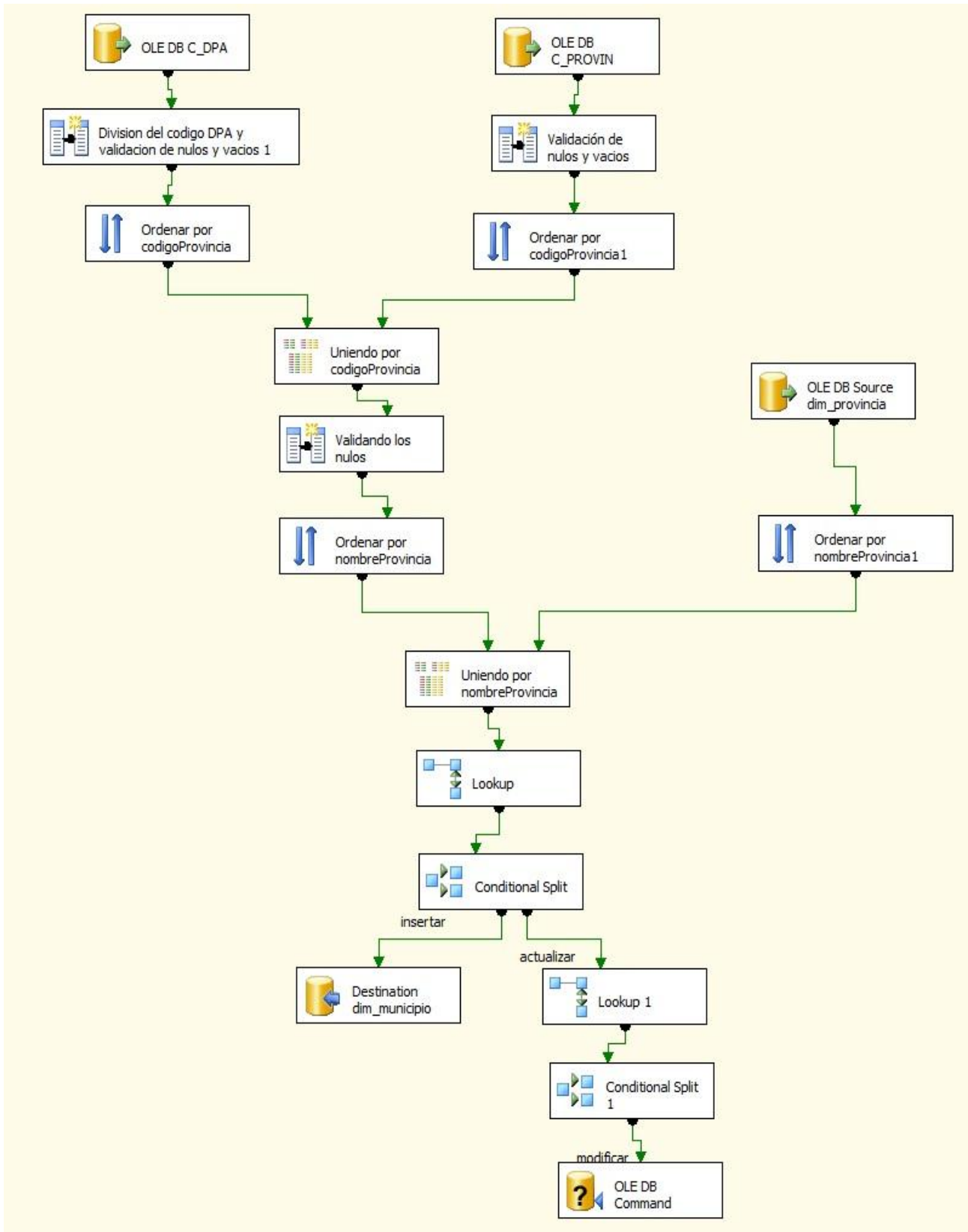
Anexo # 9 Escenario de Flujo de Datos del proceso ETL de la tabla dim_provincia



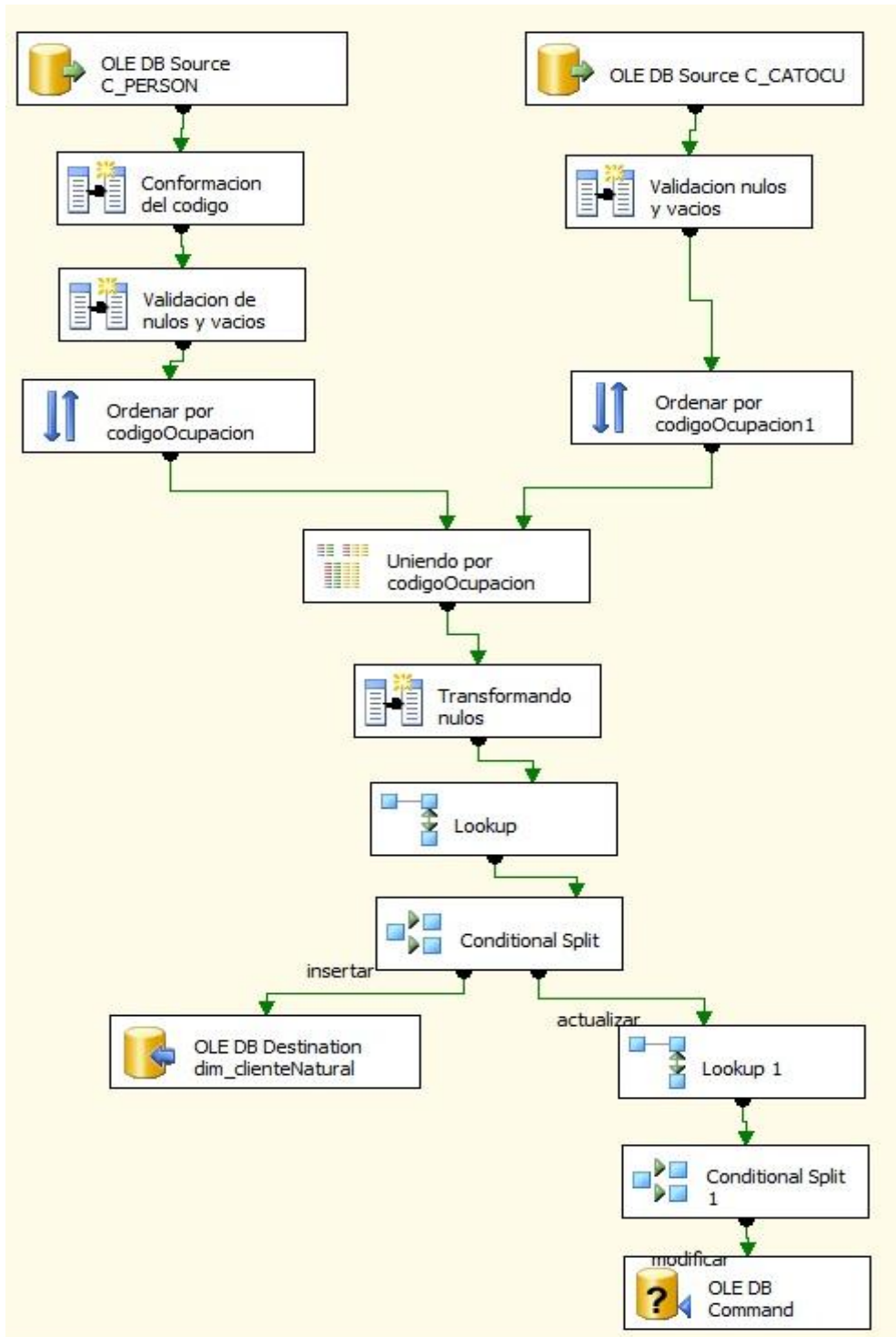
Anexo #10 Escenario de Flujo de Datos del proceso ETL de la tabla dim_tiempo



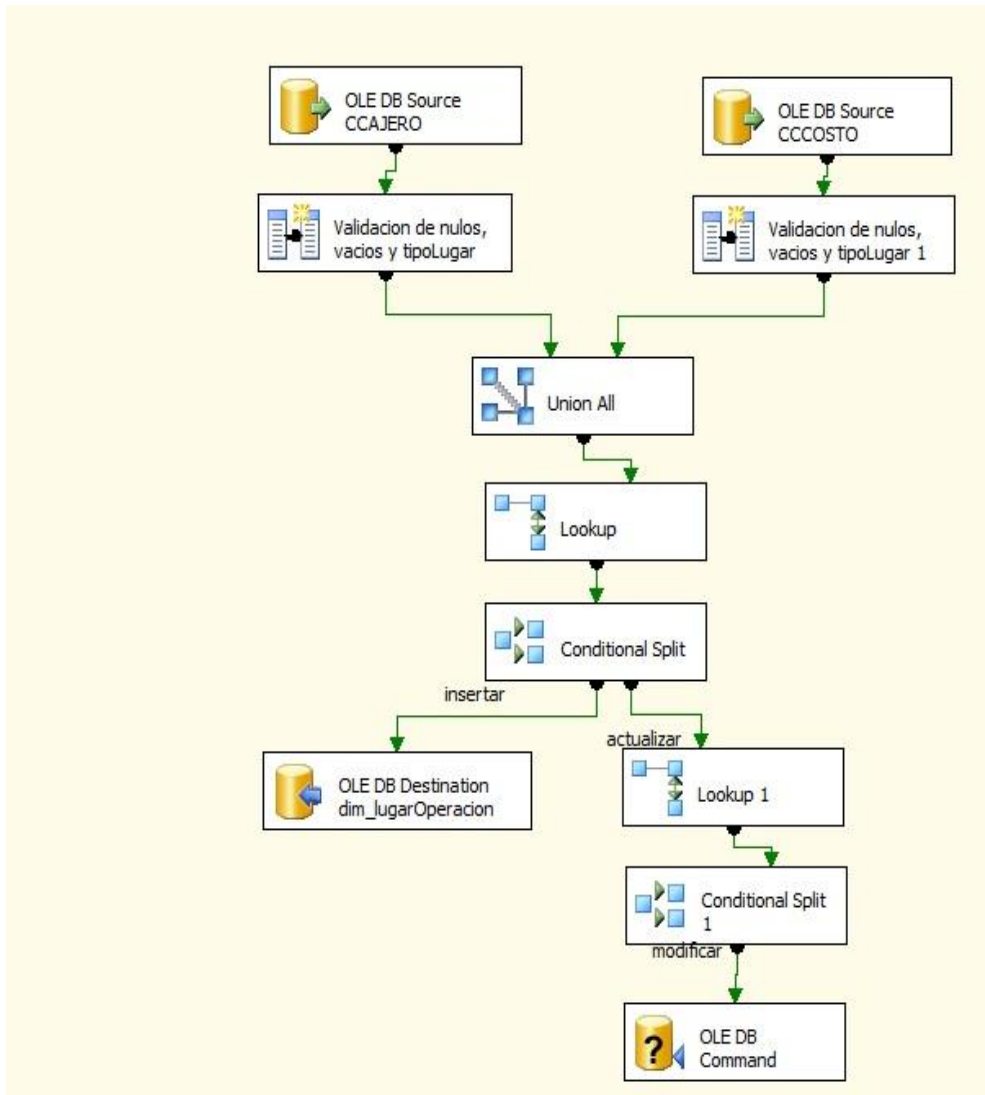
Anexo # 11 Escenario de Flujo de Datos del proceso ETL de la tabla dim_municipio



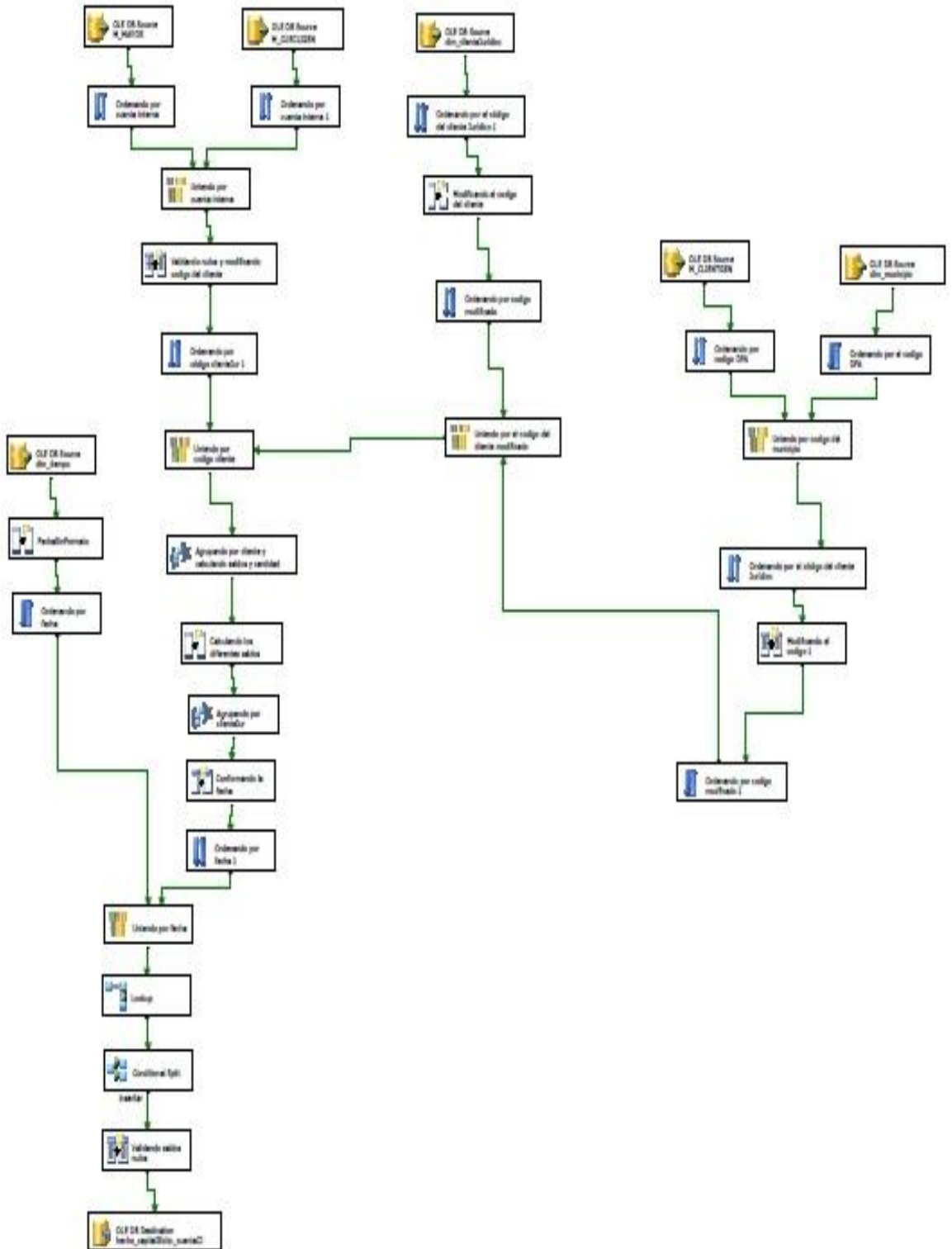
Anexo # 12 Escenario de Flujo de Datos del proceso ETL de la tabla dim_clienteNatural



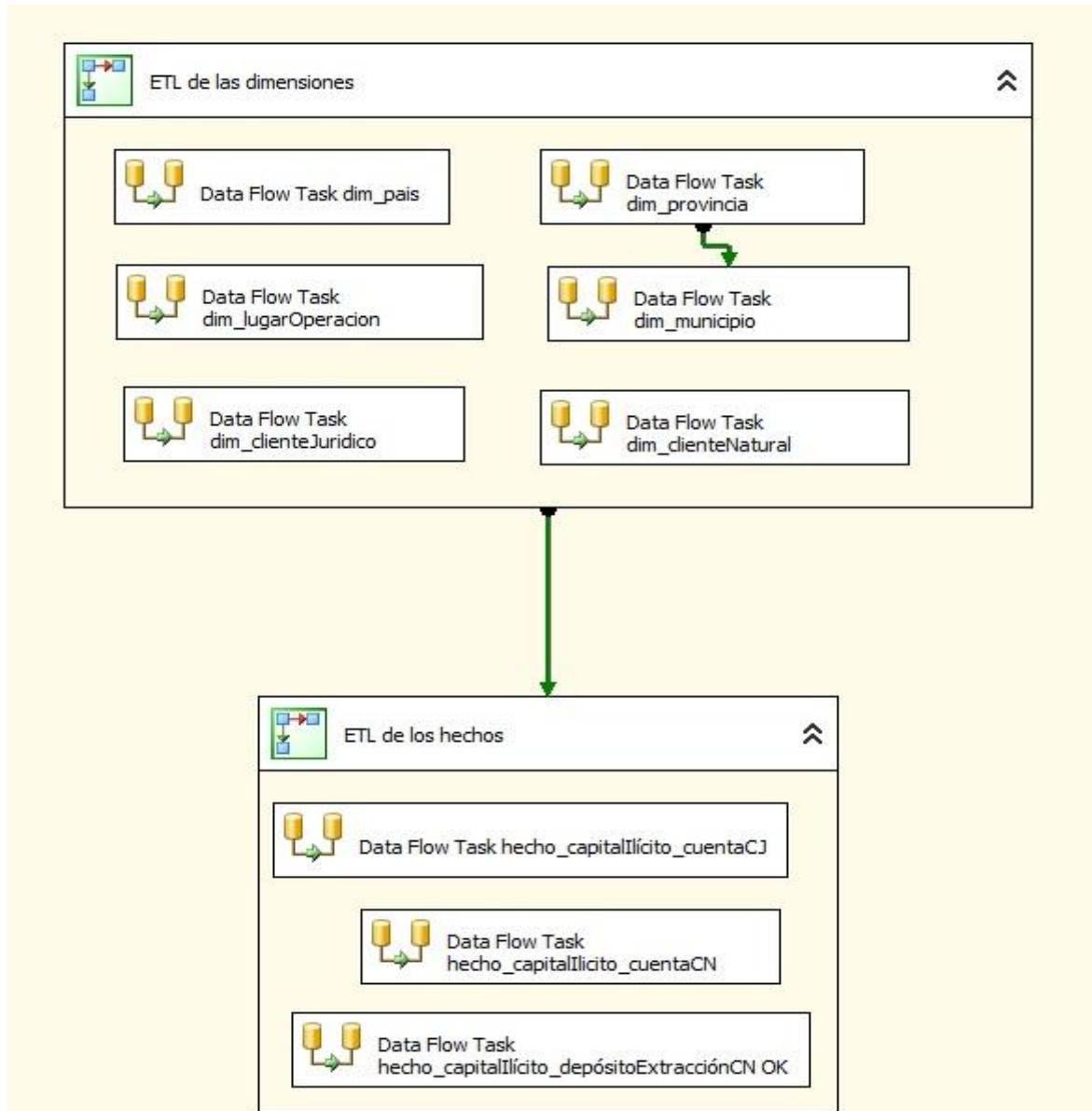
Anexo # 13 Escenario de Flujo de Datos del proceso ETL de la tabla dim_lugarOperación



Anexo # 14 Escenario de Flujo de Datos del proceso ETL de la tabla hecho_capitalIlicito_cuentaCJ



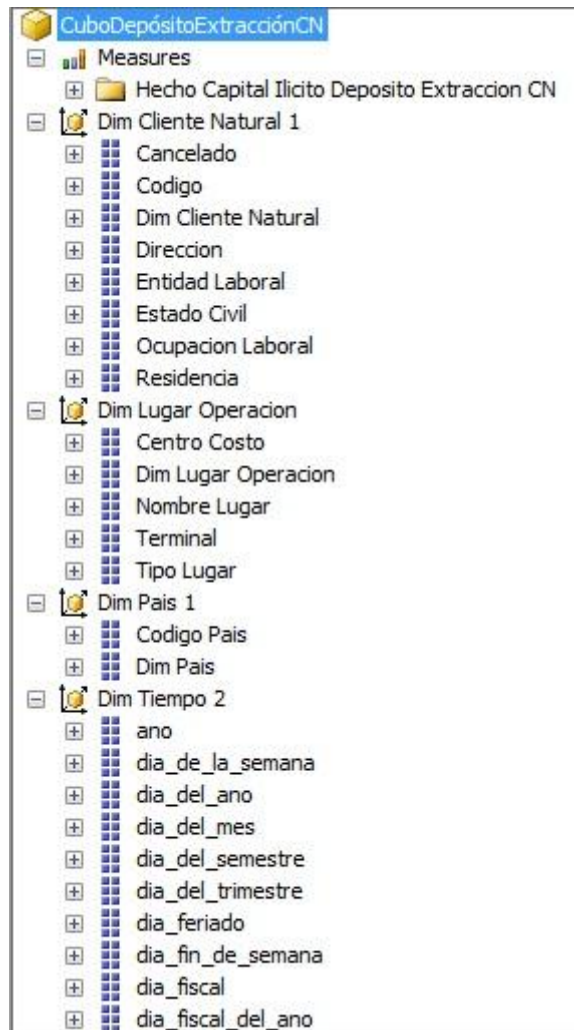
Anexo # 15 Control de flujo del proceso ETL para la carga diaria del Mercado de Datos



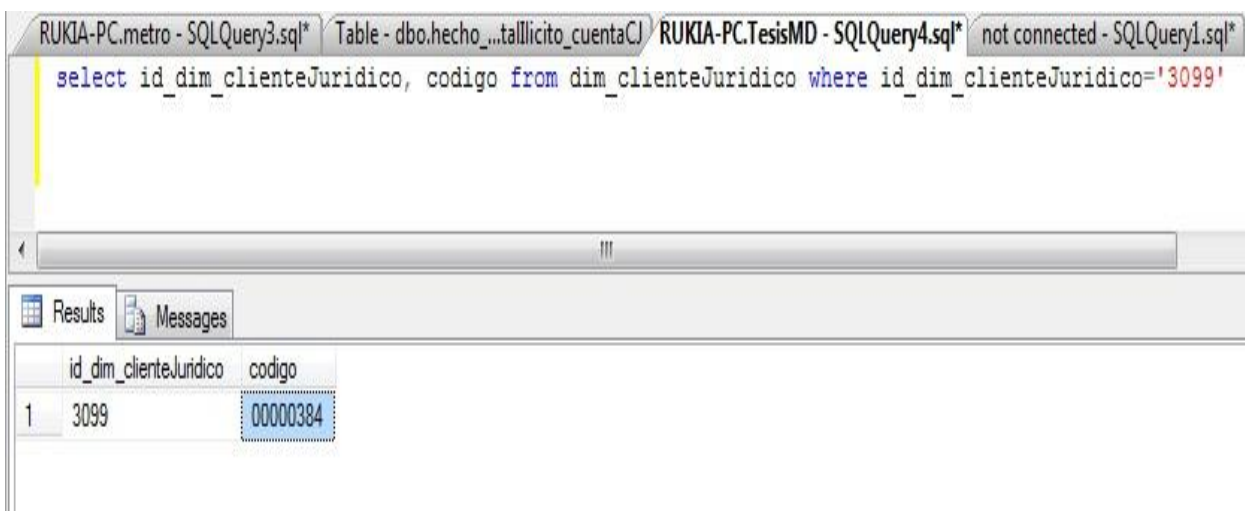
Anexo # 16 Diseño del cubo OLAP "CuboCuentaCN"



Anexo # 17 Diseño del cubo OLAP “CuboDepósitoExtracciónCN”



Anexo # 18 Consulta y resultado de la búsqueda del cliente con id en la dimensión dim_clienteJurídico igual a “3099”.



Anexo # 19 Consulta y resultado de la búsqueda de la cantidad de cuentas que posee el cliente con código igual a “00000384” en la tabla M_CUECLIGEN de la base de datos origen “metro”.

The screenshot shows a SQL query window with the following query:

```
select * from m_cuecligen where cod_client='00000384'
```

The results are displayed in a table with the following columns: COD_CLIENT, COD_MONEDA, CUE_SUBCUE, TIP_CONTRA, and COD_CONTRA. The results are as follows:

	COD_CLIENT	COD_MONEDA	CUE_SUBCUE	TIP_CONTRA	COD_CONTRA
1	00000384	40	3131	8	0000418
2	00000384	43	3131	8	0000221