UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS

FACULTAD 3

Desarrollo de un Almacén de Datos para el Banco Nacional de Cuba

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

Autor: Raúl José Pacheco Rivero

Tutor:
Ing. Yoan Asdrúbal Quintana Ramírez

La Habana, Junio de 2013
"Año 55 de la Revolución"

"El software es como la entropía: difícil de atrapar, no pesa, y cumple la Segunda Ley de la Termodinámica, es decir, tiende a incrementarse"

Norman Augustine

Declaración de Autoría

Declaro ser autor de la presente tesis y reconozco a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo. Para que así conste firmo la presente a los ____ días del mes de _____ del año 2013.

Raúl José Pacheco Rivero

_____

Firma del Autor:

Ing. Yoan Asdrúbal Quintana Ramírez

_____

Firma del Tutor:

Datos de Contacto

Tutor:

Ing. Yoan Asdrúbal Quintana Ramírez

Graduado en la Universidad de las Ciencias Informáticas en el año 2011. Jefe del equipo de desarrollo de Base de Datos del sistema Quarxo.

E-mail: yaquintana@uci.cu

Dedicatoria

A mis abuelos Edita y Papile.

A mis padres.

A mi Chinita linda.

Agradecimientos

A mi abuela Edita por su educación, cariño, amor, apoyo y porque siempre la voy a llevar en mi corazón.

A mi abuelo Papile, que yo sé que me cuida y guía desde el cielo.

A mis padres por todo su apoyo y por creer en mí.

A mi novia Yusleidy (La China) por estar siempre a mi lado y darme todo su amor y apoyo incondicional, por todos los momentos lindos que he pasado junto a ella.

A mis hermanos Iliria, Ernesto y Yanetsi.

A toda mi familia por confiar en mí y apoyarme.

A mi suegra por todo el apoyo y cariño que me ha brindado y por aceptarme en la familia.

A mi tutor por ayudarme en todo momento y contribuir en este logro.

A todos mis compañeros que han estado conmigo estos 5 años.

A todos los profesores que contribuyeron con mi educación profesional.

A Fidel y la Revolución por brindarme la oportunidad de realizar mi sueño.

A todos los que de una forma u otra me han ayudado y apoyado.

Resumen

La presente investigación surge como parte de la colaboración que existe entre la Universidad de las Ciencias Informáticas y el Banco Nacional de Cuba. El principal objetivo de este último es perfeccionar el sistema monetario, normalizar las relaciones financieras externas del país y apoyar las gestiones de créditos de las empresas cubanas y de los bancos integrantes del sistema financiero cubano. En el presente Trabajo de Diploma se realiza el análisis, diseño e implementación de un Almacén de Datos para las operaciones contables que se gestionan en el Banco Nacional de Cuba, con el propósito de almacenar gran cantidad de información y con ello viabilizar la integración de los datos de la institución.

Su construcción está basada en la metodología Hefesto, la estructura lógica propuesta, el diseño y la implementación son consecuentes con ésta. Igualmente se definen e implementan los mecanismos de extracción, transformación y carga de los datos correspondientes al modelo propuesto.

Se realizan pruebas de rendimiento para determinar la velocidad de respuesta al ejercer consultas y obtener resultados y se verifica la calidad de los datos mediante las pruebas de perfilado de datos.

Palabras claves:

Almacén de Datos, Hefesto, metodología, operaciones contables.

ÍNDICE DE CONTENIDOS

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

Introducción

Con el creciente desarrollo tecnológico actual, las tecnologías de la información y las comunicaciones se han incorporado ágilmente a la mayoría de las actividades del ser humano, tornándose casi imprescindibles si se quiere estar acorde al acelerado desarrollo de todos los sectores de la sociedad, del cual no se encuentra exento el sector empresarial.

El avance de las tecnologías informáticas provee a las empresas de sistemas que agilizan y hacen más eficientes los procesos administrativos. Tareas que eran realizadas engorrosamente por ser repetitivas y de un grado de complejidad elevado, son ejecutadas de forma rápida teniendo en cuenta el nivel de seguridad y precisión al manejar grandes fuentes de información.

El Banco Nacional de Cuba es el encargado de perfeccionar el sistema monetario, normalizar las relaciones financieras externas del país y apoyar las gestiones de créditos de las empresas cubanas y de los bancos integrantes del sistema financiero cubano (1).

En el BNC[1] se encuentra en explotación el sistema Quarxo, sistema contable desarrollado por la Universidad de las Ciencias Informáticas, el cual posee una base de datos en la que son almacenadas todas las operaciones contables realizadas en la entidad para tener constancia de las mismas. El volumen de información en la base de datos aumenta considerablemente, lo que trae como consecuencia una sobrecarga en la misma y un aumento en el tiempo de respuesta de la aplicación. En la instalación inicial del sistema Quarxo se cargó la información almacenada a partir del año 2007, el resto de los datos aún se encuentra en el antiguo sistema contable usado por la entidad: SABIC[2], debido a que los mismos son consultados ocasionalmente por lo que se consideró que sería saturar la Base de Datos con información que aunque es muy importante, no es indispensable para correcto funcionamiento de Quarxo, garantizando el rendimiento del servidor de Base de Datos.

El SABIC fue desarrollado por la Dirección de Sistemas Automatizados del BCC[3] para satisfacer las necesidades de procesamiento de datos de bancos e instituciones financieras de nuestro país. Para el desarrollo de este sistema fueron utilizadas tecnologías que han ido quedando atrás con la aparición de herramientas computacionales en la actualidad (2). Lo que trae como consecuencia diferencias en cuanto al formato en que se maneja la información. Además al estar desarrollado sobre sistema operativo *MS-DOS* dificulta el acceso a la información pues los especialistas deben reiniciar el sistema

[1] Banco Nacional de Cuba
[2] Sistema Automatizado para la Banca Internacional de Comercio
[3] Banco Central de Cuba

operativo *Windows XP* para acceder por *MS-DOS*, consultar la información deseada y luego volver a reiniciar para seguir trabajando con el sistema Quarxo.

Como consecuencia de este proceso de actualización del sistema contable del BNC, los datos de las operaciones contables que se manejan en la entidad han quedado distribuidos en dos bases de datos: la base de datos en *SQL Server* 2005 perteneciente al sistema Quarxo, que contiene la información de las operaciones a partir del año 2007 hasta la fecha, y la base de datos en *Visual FoxPro* del sistema SABIC, que contiene la información histórica de las operaciones desde el año 1980 hasta el 2007.

Según la información ofrecida por Ileana Torres Sánchez[4], la carga de los datos de las operaciones almacenadas en la Base de Datos del sistema SABIC tiene como premisa la disponibilidad de la información como material de consulta para garantizar las relaciones financieras con empresas radicadas en Cuba o en el exterior. Su principal objetivo es mantener el registro y control de la deuda que el Estado y el BNC tienen contraída con acreedores extranjeros, las cuales están sujetas frecuentemente a procesos de renegociación. Además proporciona la información relacionada con las operaciones contables asociadas a las cartas de créditos, negociaciones, otros financiamientos, garantías y préstamos realizados por la institución en ese período.

Problema científico: ¿Cómo lograr la integración de los datos de las operaciones contables que se gestionan en el Banco Nacional de Cuba?

Objeto de estudio: Proceso de desarrollo de un Almacén de Datos.

Campo de acción: Proceso de desarrollo de un Almacén de Datos para las operaciones contables que se gestionan en el Banco Nacional de Cuba.

Objetivo general: Desarrollar un Almacén de Datos para la integración de la información de las operaciones contables del Banco Nacional de Cuba.

Objetivos específicos:

- ✓ Fundamentar la investigación, mediante la elaboración del Marco Teórico.
- ✓ Analizar, diseñar e implementar un almacén de datos para el almacenamiento de los datos históricos del BNC.

[4] Técnico medio en Contabilidad, Directora del área de Contabilidad del BNC

- ✓ Validar la solución propuesta mediante la aplicación de pruebas de volumen, carga y perfilado de datos.

Tareas de la investigación:

- ✓ Análisis de los requerimientos para el diseño del Almacén de Datos.
- ✓ Análisis de los sistemas de almacenamiento de datos utilizados por la entidad.
- ✓ Elaboración del modelo lógico de la estructura del Almacén de Datos.
- ✓ Diseño e implementación del Almacén de Datos del histórico pasivo.
- ✓ Ejecución de los procesos de extracción, transformación, carga y limpieza de los datos que se almacenarán en el Almacén de Datos.
- ✓ Validación mediante la aplicación de pruebas de volumen, carga y perfilado de los datos a la solución propuesta.

Posibles resultados: Un Almacén de Datos que permita la integración de los datos históricos en esta institución.

El presente trabajo está estructurado por **cuatro capítulos**, distribuidos de la siguiente manera:

Capítulo 1: Fundamentación Teórica

Se estudiarán los conceptos, tecnologías y metodologías que son utilizadas para el desarrollo de los Almacenes de Datos, así como sus características, arquitectura, herramientas y ventajas.

Capítulo 2: Análisis y diseño de un Almacén de Datos para el BNC

Se mostrarán los pasos a seguir para la realización del análisis y diseño del Almacén de Datos según los describe la metodología a utilizar. Con la aplicación de esta metodología se lleva a cabo la construcción de un modelo lógico a partir del modelo conceptual propuesto.

Capítulo 3: Implementación de un Almacén de Datos para el BNC

Se realiza la implementación de los procesos de Extracción, Transformación y Carga mediante la utilización de las herramientas definidas en el Capítulo 1.

Capítulo 4: Validación de la solución

Se hace énfasis en aspectos tales como las pruebas de volumen y carga, análisis de los tiempos de respuesta, así como hacer pruebas para verificar la calidad de los datos almacenados.

Capítulo 1: Fundamentación Teórica

En el presente capítulo se exponen los conceptos asociados a los Almacenes de Datos. Se estudian las herramientas, técnicas y tecnologías para el desarrollo de un Almacén de Datos. Además, se realiza un estudio de la metodología que guiará el análisis, diseño y las herramientas a utilizar.

1. Almacén de Datos

   - Según Inmon un Almacén de Datos es una colección de datos orientada a un determinado ámbito (empresa, organización), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza (3).
   - Según Kimball los Almacenes de Datos son una copia de los datos transaccionales estructurados específicamente para consultas y análisis (4).
   - Según Claudia Imhoff los Almacenes de Datos son un conjunto de datos almacenados en forma de repositorio que ofrecen una vista común de los datos de la empresa, independientemente de cómo pueden más tarde ser utilizados por los consumidores y que son consultados por otro conjunto de Almacenes de Datos: los Mercados de Datos (5).
   - A modo de conclusión un Almacén de Datos es una colección de datos donde se almacena información histórica durante un amplio período de tiempo, la cual puede ser consultada con el objetivo de ayudar en la toma de decisiones de la empresa.

1.1. Características de los Almacenes de Datos

Los Almacenes de Datos poseen las siguientes características (6).

- Orientada al negocio

La primera característica de los Almacenes de Datos, es que la información se clasifica en base a los aspectos que son de interés para la empresa, excluye la información que no será utilizada exclusivamente en el proceso de toma de decisiones. Esta clasificación afecta el diseño y la implementación de los datos encontrados en el Almacén de Datos, debido a que la estructura del mismo difiere considerablemente a la de los clásicos procesos operacionales orientados a las aplicaciones.

- Integrada

La integración implica que todos los datos de diversas fuentes que son producidos por distintos departamentos, secciones y aplicaciones, tanto internas como externas, deben ser consolidados en una

instancia antes de ser agregados al Almacén de Datos. A este proceso se lo conoce como Extracción, Transformación y Carga de Datos (*ETL*, por sus siglas en inglés). La integración de datos resuelve diferentes tipos de problemas relacionados con las convenciones de nombres, unidades de medidas, codificaciones, fuentes múltiples, entre otras.

- Variante en el tiempo

Debido al volumen de información que se manejará en el Almacén de Datos, cuando se le realiza una consulta, los resultados deseados demorarán en originarse. Este espacio de tiempo que se produce desde la búsqueda de datos hasta su consecución es del todo normal en este ambiente y es, precisamente por ello, que la información que se encuentra dentro del depósito de datos se denomina de tiempo variable. Esta característica básica, es muy diferente de la información encontrada en el ambiente operacional, en el cual, los datos se requieren en el momento de acceder, es decir, que se espera que los valores procurados se obtengan a partir del momento mismo de acceso.

Esto contribuye a una de las principales ventajas del Almacén de Datos: los datos son almacenados junto a sus respectivos históricos. Esta cualidad que no se encuentra en fuentes de datos operacionales, garantiza poder desarrollar análisis de la dinámica de la información. Es decir, que gracias a la dimensión tiempo se podrá tener acceso a diferentes versiones de la misma información.

- No volátil

La información es útil para el análisis y la toma de decisiones solo cuando es estable. Los datos operacionales varían momento a momento, en cambio, los datos una vez que entran en el Almacén de Datos no cambian. La actualización, o sea, insertar, eliminar y modificar, se hace de forma muy habitual en el ambiente operacional sobre una base, registro por registro. En los depósitos de datos la manipulación básica de los datos es mucho más simple, debido a que solo existen dos tipos de operaciones: la carga de datos y el acceso a los mismos.

1.2. Objetivos de un Almacén de Datos

Un Almacén de Datos tiene como objetivos principales: (7)

- Hacer que la información de una organización sea fácilmente accesible: El contenido del Almacén de Datos debe ser comprensible. Los datos deben ser intuitivos y evidentes para el usuario de negocios, no solamente para el desarrollador.

- Presentar la información de la organización de forma coherente: Los datos en el almacén deben ser creíbles.
- Protege la información: El Almacén de Datos de forma efectiva debe controlar el acceso a la información confidencial de la organización.
- Servir como base para la toma de decisiones: El Almacén de Datos debe tener los datos correctos en el mismo para apoyar la toma de decisiones.

1.3. Estructura de un Almacén de Datos

Están compuestos por diversos tipos de datos, que se organizan y dividen de acuerdo con el nivel de detalle que posean (6).

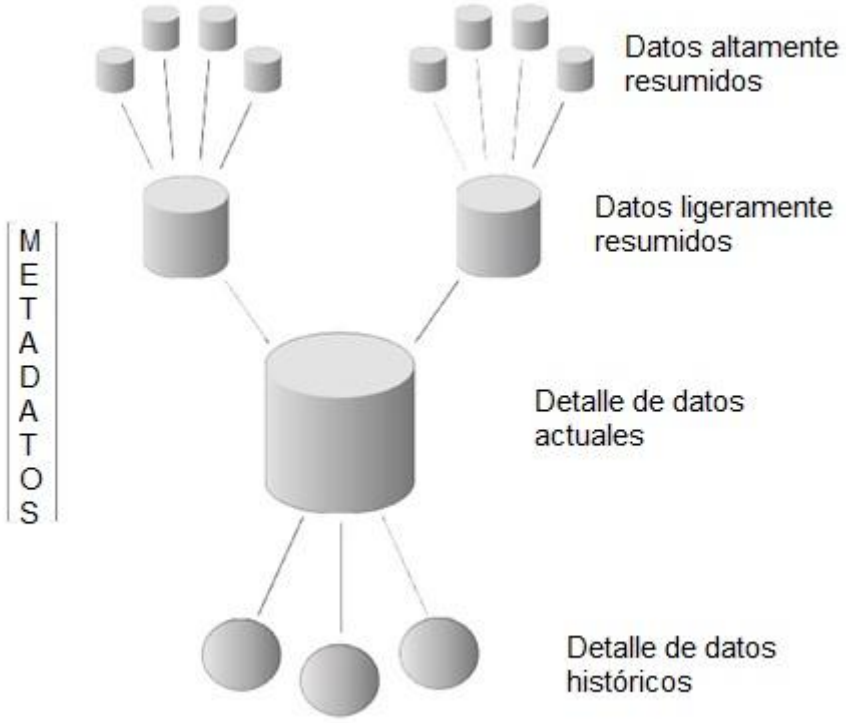

Figura 1 Estructura de los Almacenes de Datos.

Detalle de datos actuales: Son aquellos que reflejan las ocurrencias más recientes. Generalmente se almacenan en disco, aunque su administración sea costosa y compleja, con el fin de conseguir que el acceso a la información sea sencillo y veloz, ya que son bastante voluminosos. Su gran tamaño se debe a que los datos residentes poseen el más bajo nivel de granularidad, o sea, se almacenan a nivel de detalle.

Detalle de datos históricos: Representan aquellos datos antiguos, que no son frecuentemente consultados. También se almacenan a nivel de detalle, normalmente, sobre alguna forma de almacenamiento externa, ya que son muy pesados y en adición a esto, no son requeridos con mucha periodicidad. Estos tipos de datos son consistentes con los de Detalle de datos actuales.

Datos ligeramente resumidos: Son los que provienen desde un bajo nivel de detalle y sumarizan o agrupan los datos bajo algún criterio o condición de análisis. Habitualmente son almacenados en disco.

Datos altamente resumidos: Son aquellos que compactan aún más a los datos ligeramente resumidos. Se guardan en disco y son muy fáciles de acceder.

Metadatos: Representan la información acerca de los datos. De muchas maneras se sitúa en una dimensión diferente al de otros datos del Almacén de Datos, ya que su contenido no es tomado directamente desde el ambiente operacional.

2. Metodologías para el diseño del Almacén de Datos

2.1. Metodología Hefesto

Hefesto es una metodología para el desarrollo de Almacenes de Datos, cuya propuesta está fundamentada en una amplia investigación, comparación de metodologías existentes y experiencias en procesos de confección de Almacenes de Datos (6).

Esta metodología permite la construcción de un Almacén de Datos de forma sencilla, ordenada e intuitiva. Hefesto es una metodología bien fundamentada y explícita que facilita la construcción de un Almacén de Datos de manera metódica y sencilla, guiándose por pasos lógicos relacionados sólidamente durante todas las etapas del proceso de confección (6).

Hefesto, comienza recolectando las necesidades de información de los usuarios y se obtienen las preguntas claves del negocio. Luego, se identifican los indicadores resultantes de las interrogantes y sus respectivas perspectivas de análisis; mediante estos indicadores se construye el modelo conceptual del Almacén de Datos. Después, se analizan los *OLTP*[5] para señalar las correspondencias con los datos fuentes y seleccionar los campos de estudio de cada perspectiva. Seguidamente se construye el modelo lógico, explicitando las jerarquías que deben intervenir. Por último, se definirán los procesos de extracción, transformación, carga y limpieza de los datos fuente.

[5] Procesamiento de Transacciones En Línea (OnLine Transaction Processing)

La metodología HEFESTO puede resumirse a través del siguiente gráfico:

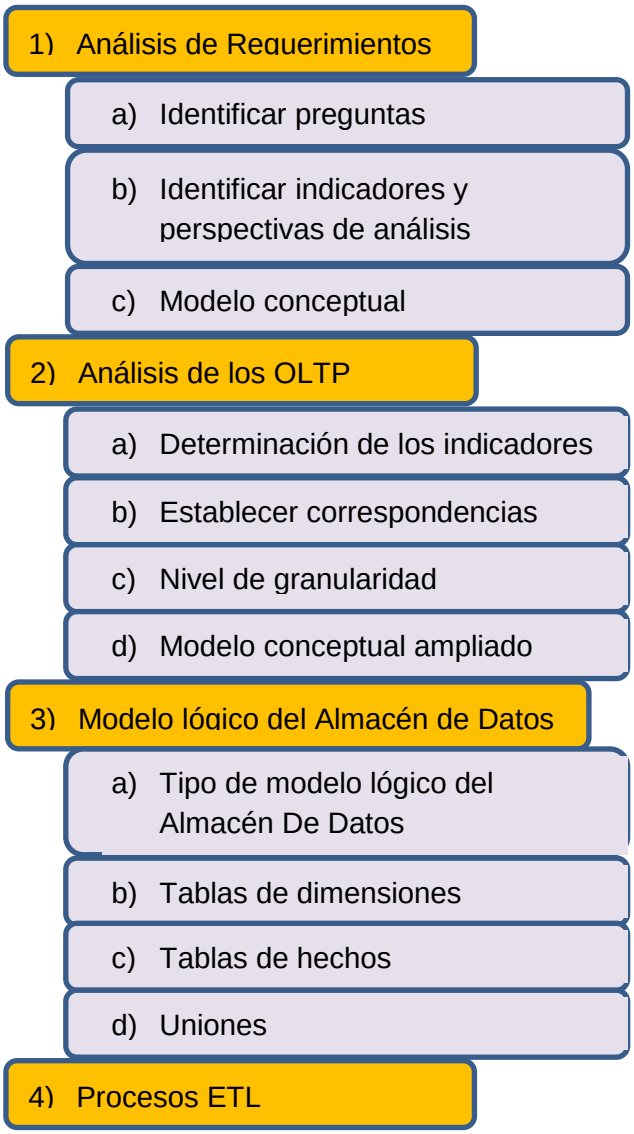

Figura 2 Pasos de la metodología Hefesto.

La metodología está orientada a la construcción de Almacenes de Datos para *OLAP*[6] y comprende las siguientes fases: (8)

- Análisis de requerimientos:

Identificar preguntas para las que se quieren tener respuesta y los objetivos que se quieren conseguir con el nuevo sistema.

[6] Procesamiento Analítico en Línea (On-Line Analytical Processing)

Analizar las preguntas para determinar las perspectivas de análisis y los indicadores de negocio.

Diseñar el modelo conceptual, que incluirá las perspectivas e indicadores identificados. A través del modelo se podrán alcanzar claramente cuáles son los alcances del proyecto, y será un punto de partida con alto nivel de definición para su exposición a los usuarios y responsables.

- Análisis de los sistemas transaccionales:

Determinación de indicadores: identificar el origen de los indicadores en los sistemas transaccionales y determinar la forma de su cálculo.

Correspondencias: establecer correspondencias entre los elementos definidos en el modelo conceptual y las fuentes de datos existentes en los *OLTP* (sistemas transaccionales).

Definición del nivel de granularidad: nivel de detalle de los datos a obtener para cada dimensión de análisis.

Modelo conceptual ampliado con los campos identificados para cada perspectiva.

- Modelo lógico del ETL:

Tipo de modelo lógico del Almacén de Datos: selección del tipo de esquema que se utilizará (estrella, copo de nieve o constelación).

Tabla de dimensiones: Construcción de las tablas de dimensiones para cada una de las perspectivas de análisis considerada.

Tablas de hechos: definición de las tablas de hechos que contendrán la información a partir de los cuales se construirán los indicadores de análisis.

Uniones: relaciones entre las tablas de dimensiones y las tablas de hechos.

- Procesos ETL:

Análisis, definición y desarrollo de todos aquellos procesos necesarios para la extracción, transformación y carga de datos desde los sistemas origen para llenar el Almacén de Datos.

Revisión y mantenimiento del Almacén de Datos:

Ajustes en el diseño del Almacén de Datos y mantenimiento en el tiempo.

2.2. Metodología *Rapid Warehousing*

Esta metodología es iterativa, y está basada en el desarrollo incremental de proyectos de Almacenes de Datos dividido en cinco fases (8):

- Definición de los objetivos: En esta fase se especificará el equipo de proyecto, el alcance del sistema y cuáles son las funciones que el Almacén de Datos realizará como suministrador de información de negocio estratégica para la empresa. Se definirán así mismo, los parámetros que permitan evaluar el éxito del proyecto.
- Definición de los requerimientos de información: Tal como sucede en todo tipo de proyectos, sobre todo si se involucran técnicas novedosas como son las relativas al Almacén de Datos, es importante analizar las necesidades y hacer comprender las ventajas que este sistema puede reportar.
- Diseño y modelado: Los requerimientos de información identificados durante la fase anterior proporcionarán las bases para realizar el diseño y el modelado del Almacén de Datos. En esta fase se identificarán las fuentes de los datos (sistema operacional, fuentes externas) y las transformaciones necesarias para, a partir de dichas fuentes, obtener el modelo lógico del Almacén de Datos. Este modelo estará formado por entidades y relaciones que permitirán resolver las necesidades del negocio de la organización.
- Implementación: La implementación de un Almacén de Datos lleva implícitos los siguientes pasos:
  - Extracción de los datos del sistema operacional y transformación de los mismos.
  - Carga de los datos validados en el Almacén de Datos. Esta carga debe ser planificada con una periodicidad que se adaptará a las actualizaciones detectadas durante las fases de diseño del nuevo sistema.
  - Explotación del Almacén de Datos mediante diversas técnicas dependiendo del tipo de aplicación que se le dé a los datos.
- Revisión: La construcción del Almacén de Datos no finaliza con la implantación del mismo, sino que es una tarea iterativa en la que se trata de incrementar su alcance aprendiendo de las experiencias anteriores. Después de implantarse, se debería revisar, planteando preguntas que permitan, después de los seis o nueve meses posteriores a su puesta en marcha, definir cuáles serían los aspectos a mejorar o potenciar en función de la utilización que se haga del nuevo sistema.

2.3. Metodología Kimball

La metodología de Kimball[7] se enfoca principalmente en el diseño de la base de datos que almacenará la información para la toma de decisiones. El diseño se basa en la creación de tablas de hechos que son tablas que contienen la información numérica de los indicadores a analizar, es decir, la parte cuantitativa de la información (9).

Está compuesta por 4 fases las cuales son:

- Requerimientos y gestión de proyectos.
- Arquitectura técnica.
- Implementación.
- Implantación y crecimiento.

En la figura se muestra el ciclo de vida de la metodología Kimball para el desarrollo de un Almacén de Datos:

[7] Ralph Kimball

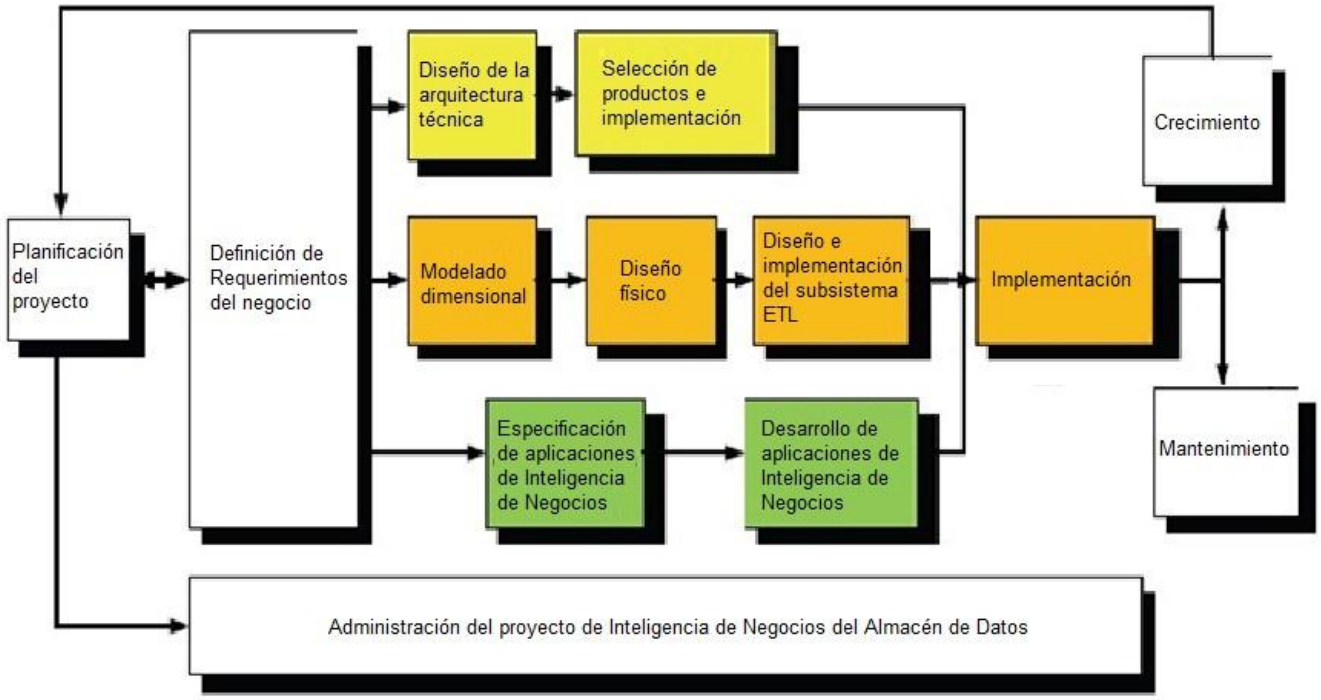

Figura 3 Ciclo de vida de la metodología Kimball.

Como se evidencia en la figura anterior, esta metodología propone dentro de cada una de sus fases la realización de un conjunto de actividades, lo que la convierte en una metodología madura y robusta, por lo que demanda más recursos, tiempo y documentación.

2.4. Justificación de la metodología a utilizar

El Banco Nacional de Cuba necesita un Almacén de Datos, el cual debe desarrollarse en un corto periodo de tiempo, por lo que se necesita una metodología sencilla y ágil pero a su vez madura, que garantice la integración de la información que actualmente se maneja en el sistema implantado.

Por tales razones se escoge la metodología Hefesto la cual permite la construcción de un Almacén de Datos de forma sencilla, ordenada e intuitiva y está fundamentada en una amplia investigación, comparación de metodologías existentes y experiencias en procesos de confección de Almacenes de Datos.

Además, esta metodología brinda las siguientes ventajas:

- 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 Almacén de Datos.
- Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.
- Es independiente del tipo de ciclo de vida que se emplee para contener la metodología.
- Es independiente de las herramientas que se utilicen para su implementación.
- Es independiente de las estructuras físicas que contengan el Almacén de Datos 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.

3. Herramientas para el desarrollo de un Almacén de Datos.

Para el desarrollo de un Almacén de Datos es necesario tener en cuenta las herramientas para realizar el diseño y la implementación del mismo. Algunas de las herramientas son: herramientas *CASE*, Sistemas Gestores de Bases de Datos (SGBD) y herramientas de Extracción, Transformación y Carga de datos.

3.1. Herramienta *CASE*

La herramienta *CASE* se utilizará para el diseño del Almacén de Datos. Entre ellas se encuentra:

3.1.1. ER/Studio

ER/Studio es una herramienta para el diseño de bases de datos, que brinda productividad en su diseño, generación y mantenimiento de aplicaciones. Desde un modelo lógico de los requerimientos de información, hasta el modelo físico perfeccionado para las características específicas de la base de datos diseñada, permite visualizar la estructura, los elementos importantes, y optimizar el diseño de la base de datos. *ER/Studio* soporta principalmente bases de datos relacionales SQL y bases de datos que incluyen *Oracle*, *Microsoft* SQL *Server*, *Sybase* (10).

3.1.2. Justificación de la herramienta CASE a utilizar

Se utilizará *ER/Studio* 7.5 ya que los especialistas del BNC especificaron su uso debido a que la Base de Datos del producto Quarxo se encuentra diseñada con esta herramienta, por lo que facilitará un

mayor y más fácil entendimiento por parte del área Informática en la entidad respecto a la solución propuesta.

3.2. Sistema Gestor de Bases de Datos

Los gestores de bases de datos dan soporte a los sistemas *OLTP* para registrar las transacciones diarias de las empresas. Estos gestores se han perfeccionado para dar soporte a los sistemas *OLAP* para la exploración de datos y para la extracción de conocimiento.

3.2.1. Microsoft SQL Server 2005

Está construido sobre las fortalezas de *SQL Server* 2000, aumenta el rendimiento, confiabilidad, disponibilidad, capacidad de programación y facilidad de uso del *SQL Server* 2000, ofrece nuevas oportunidades de diseño para los desarrolladores de cubos y proporciona un enfoque nuevo para el uso de los sistemas *OLAP*. Brinda la herramienta *Microsoft SQL Server* 2005 *Analysis Services* (*SSAS*) a la cual se le han agregado nuevas características, además de las ya existentes en *Analysis Services*. Esta herramienta incluye mejoras como métricas empresariales personalizables en los cubos, denominadas indicadores clave de rendimiento (*KPI*, por sus siglas en inglés); la creación de varias tablas de hechos en un único cubo; medidas de suma parcial para agregar medidas a una dimensión y no a otras; mejoras en las dimensiones como atributos, pues en versiones anteriores las dimensiones se basaban directamente en los niveles de una jerarquía y ahora se basan en atributos que corresponden a las columnas de las tablas de dimensión, separando así las características estructurales de una dimensión de su característica de exploración; relaciones de dimensiones de referencia, admitiéndose las dimensiones de referencia mediante el uso de relaciones entre las dimensiones de referencia y un grupo de medida de otra dimensión lo cual posibilita asociar estas dimensiones a un cubo sin crear un esquema de copo de nieve; además brinda un tamaño de dimensiones prácticamente ilimitado puesto que ya no se depende del almacenamiento residente en memoria (11).

Microsoft SQL Server 2005 brinda una nueva plataforma para crear soluciones de integración de datos de alto rendimiento que se llama *Microsoft SQL Server* 2005 *Integration Services* (*SSIS*) que sustituye a los Servicios de transformación de datos (*DTS*, por sus siglas en inglés) del *SQL Server* 2000. Incluye paquetes de extracción, transformación y carga (*ETL*) para el almacenamiento de datos; resuelve muchas dificultades y limitaciones de los *DTS,* incluyendo mejoras como una nueva arquitectura extensible, un nuevo diseñador de paquetes, estructuras de bucle y transformaciones, así como mejoras en la implementación, administración y el rendimiento de los paquetes; nuevas herramientas gráficas y asistentes para crear y depurar paquetes; tareas para realizar funciones de flujo de trabajo como la

ejecución de comandos *SQL* y mensajería por correo electrónico; transformaciones para borrar, agregar, mezclar y copiar datos; servicio de administración e interfaces de programación de aplicaciones (*API*, por sus siglas en inglés) para programar el modelo de objetos de *Integration Services* (11).

3.2.2. Justificación del Gestor de Bases de Datos a utilizar

Se usará *SQL Server* 2005 porque es el sistema gestor de bases de datos que utiliza el Banco Nacional de Cuba siendo una especificación por parte de los especialistas de la entidad. Además, este gestor de bases de datos ofrece oportunidades de diseño para el desarrollo de Almacenes de datos con herramientas como *SQL Server* 2005 *Analysis Services* (*SSAS*) para el desarrollo de cubos. También proporciona una plataforma para crear soluciones de integración de datos de alto rendimiento llamada *SQL Server* 2005 *Integration Services* (*SSIS*) la cual incluye paquetes de extracción, transformación y carga para el almacenamiento de datos.

3.3. Herramientas de Extracción, Transformación y Carga de Datos

Las herramientas de Extracción, Transformación y Carga de Datos proporcionan funcionalidades para:

- Obtener la información necesaria a partir de datos almacenados en fuentes externas.
- Realizar operaciones sobre los datos para que puedan ser cargados en el Almacén de Datos.
- Almacenar los datos en el Almacén de Datos final.
- Control de la extracción de los datos y su automatización.
- Proporcionar la gestión integrada del Almacén de Datos y los Mercados de Datos existentes.

3.3.1. Pentaho Data Integration (PDI, también conocido como Kettle)

Pentaho Data Integration es una herramienta para diseñar y ejecutar transformaciones y trabajos[8] *ETL* usando el entorno gráfico. Provee consistencia y una sola versión de todos los recursos de información, que es uno de los más grandes desafíos para las organizaciones hoy en día (12).

Dentro de sus principales características se encuentran:

- Código abierto y sin costes de licencia.
- Entorno gráfico de desarrollo.

[8] Conjunto sencillo o complejo de tareas cuyo objetivo consiste en realizar una acción determinada, definiendo una secuencia lógica para la ejecución de las transformaciones.

- Uso de tecnologías estándar: *Java*, *XML*, *JavaScript*.
- Fácil de instalar y configurar.
- Multiplataforma: *Windows*, *Macintosh*, *Linux*.
- Basado en dos tipos de objetos: Transformaciones (colección de pasos en un proceso *ETL*) y trabajos (colección de transformaciones).

3.3.2. SQL Server Integration Services

Es el instrumento en el marco de *Microsoft* para la extracción, transformación y carga de datos en *SQL Server* 2005 (13).

Principales ventajas:

- Facilita el movimiento de los datos necesarios para el éxito de almacenamiento de datos: los trabajos de movimiento de datos son programados y pueden ser organizados en los flujos complejos no lineales como se requiere.
- Su interfaz gráfica de usuario es muy fácil de usar.
- El flujo de datos y la secuencia se controla mediante tareas de flujo de datos: las tareas de flujo de datos pueden acceder a datos de una fuente, procesarla, y guardarlas en un objetivo.
- Transformaciones avanzadas tales como transformaciones de minería de texto y búsquedas difusa: búsqueda aproximada y puntuaciones de confianza.

3.3.3. Justificación de la Herramienta ETL a utilizar

Luego de un estudio de las herramientas *Kettle* y *SSIS*, se ha seleccionado la versión 4.2.1 de *Kettle* para el desarrollo del proceso de *ETL*, dado que es una herramienta libre multiplataforma. Además, posee gran soporte técnico y los usuarios comparten muchos consejos y trucos en los foros. También, posee gran cantidad de conectores y la posibilidad de crear flujos de trabajo integrados con transformaciones de datos de manera muy sencilla y funcional.

3.4. Herramienta de Conversión de Bases de Datos

ESF Database Convert es una herramienta que permite realizar conversiones entre múltiples bases de datos utilizadas en la actualidad, ofrece la posibilidad de seleccionar las tablas y campos a convertir e incluye soporte para una amplia gama de formatos (14).

ESF Database Convert incorpora un asistente que guía al usuario paso a paso mediante el proceso de conversión. *ESF Database Convert* incluye soporte para las bases de datos más populares: *SQL Server*, *Oracle*, *PostgreSQL*, *MySQL*, *InterBase*, *Access*, *Visual FoxPro*, *Lotus* y *dBase*.

Principales características:

- Asistente integrado para realizar la conversión.
- Convierte correctamente llaves primarias, indexaciones, Auto-ID, tablas.
- Facilidad de uso.

3.5. Herramientas para la Validación de los resultados

3.5.1. DataCleaner

Es una aplicación de calidad de los datos, diseñada para ayudar al usuario a perfilar, comparar, validar y supervisar. *DataCleaner* consiste de una interfaz gráfica de usuario independiente para la creación de perfiles, comparación y validación y una aplicación *web* para la supervisión.

Esta utilidad fue desarrollada como una alternativa al *software* de las metodologías de gestión de los datos maestros, proyectos de almacenamiento de datos, investigación estadística y preparación para actividades de extracto de transformación de carga (15).

Características claves de *DataCleaner*:

- Perfilado de la base de datos en cuestión de minutos.
- Acceso a Almacén de Datos soportados en *Oracle*, *MySQL*, MS *Access*, archivos *Excel.*

3.5.2. DataGenerator para MS SQL Server

Es una pieza de *software* que se puede utilizar para obtener datos de pruebas realistas necesarias para verificar el comportamiento de las aplicaciones de bases de datos. Los datos pueden ser insertados en la base de datos directamente o se puede generar una secuencia de comandos *SQL* con instrucciones de inserción. Para cada columna, la herramienta elige un generador de datos apropiado que produce un determinado tipo de datos (16).

Características claves de *DataGenerator* para *MS SQL Server*:

- Versiones de *MS SQL* compatibles: 2000, 2005, 2008 y 2012.
- Más de 40 generadores de datos incorporados.
- Generación de datos de múltiples tablas.

- Carga datos de fuentes externas.

Conclusiones del capítulo

Actualmente se le dificulta a los especialistas del BNC consultar toda la información contable de la entidad debido a que se encuentra almacenada en dos bases de datos diferentes en cuanto a su estructura y definición, por lo que afecta la toma de decisiones a los ejecutivos de la entidad, haciéndose necesario que se integre de manera eficiente la información almacenada en estas dos fuentes.

La utilización de las herramientas caracterizadas facilitará el desarrollo de un Almacén de Datos que permitirá integrar de una forma estandarizada la información almacenada tanto en el sistema que actualmente se encuentra en explotación como en el antiguo sistema SABIC, garantizando la obtención de una solución con la calidad requerida en el menor tiempo y coste posible, apto para utilizarse en el BNC.

Capítulo 2: Análisis y Diseño de un Almacén de Datos para el BNC

En este capítulo se mostrarán las fases a seguir para la realización del análisis y el diseño del Almacén de Datos utilizando la metodología Hefesto y se explicarán los pasos de cada una de ellas. Se definirá un modelo conceptual que luego será ampliado a partir del establecimiento de los indicadores y perspectivas identificados de los requerimientos del cliente. Se establecerán las correspondencias entre el modelo conceptual y los *OLTP*. Finalmente, se diseñará el modelo lógico del Almacén de Datos.

1. Análisis de requerimientos

El análisis de los requerimientos de los diferentes usuarios, es el punto de partida de esta metodología, ya que ellos son los que deben, en cierto modo, guiar la investigación hacia un desarrollo que refleje claramente lo que se espera del depósito de datos, en relación a sus funciones y cualidades.

El objetivo principal de esta fase, es obtener e identificar las principales necesidades de información, que son esenciales para llevar a cabo las metas y estrategias de la empresa, facilitando la toma de decisiones.

1.1. Identificar preguntas**:**

En la entrevista con el Ing. Yulier Matías León[9] se investigó cuáles eran las necesidades y la información clave que considerasen más importante en relación a las actividades diarias del BNC y que estuviese de alguna manera soportada por alguna base de datos convencional.

A continuación, se analizó qué era lo que les interesaba conocer sobre dichas actividades.

A partir de la entrevista realizada se definieron las siguientes preguntas que conformarán los requerimientos para el desarrollo del Almacén de Datos.

- Se desea conocer el importe del asiento contable de cada Carta de Crédito almacenada en la tabla Histórico en una fecha determinada, así como los datos correspondientes a cada Carta de Crédito.

[9] Ing. Yulier Matías León: Jefe de la línea Soluciones Financieras y que anteriormente se desempeñaba como analista principal del proyecto Quarxo, por lo que tiene amplio conocimiento de todos los procesos contables que se realizan en el BNC

- Se desea conocer el importe del asiento contable de cada Financiamiento almacenado en la tabla Histórico en una fecha determinada, así como los datos correspondientes a cada Financiamiento.
- Se desea conocer el importe del asiento contable de cada Negociación almacenada en la tabla Histórico en una fecha determinada, así como los datos correspondientes a cada Negociación.
- Se desea conocer el importe del asiento contable de cada Renegociación almacenada en la tabla Histórico en una fecha determinada, así como los datos correspondientes a cada Renegociación.
- Se desea conocer el importe del asiento contable de cada Préstamo almacenado en la tabla Histórico en una fecha determinada, así como los datos correspondientes a cada Préstamo.
- Se desea conocer el importe del asiento contable de cada Garantía almacenada en la tabla Histórico en una fecha determinada, así como los datos correspondientes a cada Garantía.

Debido a que la dimensión Tiempo es un elemento fundamental en el Almacén de Datos, se hace hincapié en ella.

Como se puede apreciar, las necesidades de información expuestas están acorde a los objetivos y estrategias del BNC, ya que es precisamente esta información la que apoya la toma de decisiones.

1.2. Identificar indicadores y perspectivas de análisis

Una vez establecidas las preguntas claves, se procede a su descomposición para descubrir los indicadores que se utilizarán y las perspectivas de análisis que intervendrán.

Para ello, se tuvo en cuenta que los **indicadores**, para que sean realmente efectivos son, en general, valores numéricos y representan lo que se desea analizar concretamente, por ejemplo: saldos o importes, promedios, cantidades, sumatorias, fórmulas (6).

En cambio, las **perspectivas** se refieren a los objetos mediante los cuales se quiere examinar los indicadores, con el fin de responder a las preguntas planteadas, por ejemplo: clientes, proveedores, sucursales, países, productos, rubros. Cabe destacar, que el Tiempo es comúnmente una perspectiva (6).

A continuación se identifican los indicadores y perspectivas por cada pregunta:

- Importe del asiento contable de cada Carta de Crédito en el Histórico en un tiempo determinado.
    - Indicadores: Importe del asiento contable.

    - Perspectivas: Carta de Crédito, Histórico, Tiempo.
- Importe del asiento contable de cada Financiamiento en el Histórico en un tiempo determinado.
    - Indicadores: Importe del asiento contable.
    - Perspectivas: Financiamiento, Histórico, Tiempo.
- Importe del asiento contable de cada Negociación en el Histórico en un tiempo determinado.
    - Indicadores: Importe del asiento contable.
    - Perspectivas: Negociación, Histórico, Tiempo.
- Importe del asiento contable de cada Renegociación en el Histórico en un tiempo determinado.
    - Indicadores: Importe del asiento contable.
    - Perspectivas: Renegociación, Histórico, Tiempo.
- Importe del asiento contable de cada Préstamo en el Histórico en un tiempo determinado.
    - Indicadores: Importe del asiento contable.
    - Perspectivas: Préstamo, Histórico, Tiempo.
- Importe del asiento contable de cada Garantía en el Histórico en un tiempo determinado.
    - Indicadores: Importe del asiento contable.
    - Perspectivas: Garantía, Histórico, Tiempo.

A continuación se relacionan los indicadores y perspectivas que se identificaron en los requerimientos anteriores:

Perspectivas:

- Histórico.
- Tiempo.
- Carta de Crédito.
- Financiamiento.
- Renegociación.
- Negociación.
- Préstamo.
- Garantía.

Indicadores:

- Importe del asiento contable.

1.3. Modelo conceptual

En esta etapa, se construyeron los modelos conceptuales (descripción de alto nivel de la estructura de la base de datos, en la cual la información es representada a través de objetos y relaciones) a partir de los indicadores y perspectivas obtenidas en el paso anterior.

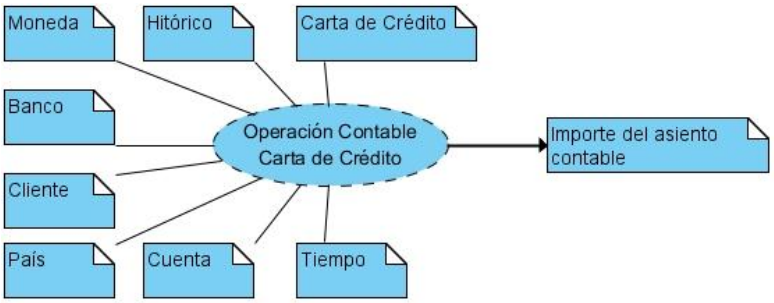

Figura 4 Modelo conceptual para Operación Contable de Carta de Crédito.

Además, se definieron los modelos conceptuales para las Operaciones Contables de Financiamiento, Garantía, Negociación, Préstamo y Renegociación (Ver anexos 1 al 5).

2. Análisis de los OLTP

En esta fase se analizan la base de datos del sistema Quarxo, la cual contiene toda la información transaccional que se maneja diariamente en el BNC y la base de datos del sistema SABIC, que contiene los datos históricos hasta el año 2007.

2.1. Establecer correspondencias con los requerimientos

Cálculo de los Indicadores:

- Importe del asiento contable: se obtiene de la columna “IMP_ASIENT” de la tabla “H_HISTOR”

Las relaciones identificadas fueron las siguientes:

- La tabla “H_HISTOR” se relaciona con la perspectiva “Histórico”.
- La tabla “o_carta_credito” se relaciona con la perspectiva “Carta de Crédito”.
- La tabla “o_otros_financiamientos” se relaciona con la perspectiva “Financiamientos”.
- Las tablas “o_negociacion” y “o_negociacion_union” se relaciona con la perspectiva “Negociaciones”.
- Las tablas “o_reneg” y “o_renegociacion_bancaria” se relaciona con la perspectiva “Renegociaciones”.
- La tabla “o_prestamo” se relaciona con la perspectiva “Préstamos”.

- La tabla “o_garant” se relaciona con la perspectiva “Garantía”.
- La tabla “C_CLIENT” se relaciona con la perspectiva “Cliente”.
- La tabla “C_BANCOS” se relaciona con la perspectiva “Banco”.
- La tabla “C_MONEDA” se relaciona con la perspectiva “Moneda”.
- La tabla “C_PAISES” se relaciona con la perspectiva “País”.
- Los campos “CUE_SUBCUE”, “TIP_CONTRA” y “COD_CONTRA” de la tabla “H_HISTOR” conforman la perspectiva “Cuenta”.
- El campo “IMP_ASIENT” de la tabla “H_HISTOR” se relaciona con el indicador “Importe del asiento contable”

2.2. Seleccionar los campos que integrarán cada perspectiva. Nivel de granularidad

Una vez establecidas las relaciones con los *OLTP*, se examinan y seleccionan los campos que conformarán cada perspectiva, ya que será a través de estos por los que se manipularán y filtrarán los indicadores.

Una vez que se recolecta toda la información pertinente y se consultó con los usuarios cuales eran los datos que consideraban de interés para analizar los indicadores ya expuestos, los resultados obtenidos fueron los siguientes:

referencia_original
referencia_correspondencia
tolerancia
please_open
descripcion_mercancia
referencia_externa
transferible
nombre_respaldo
tipo_cambio
tasa_interes
stand_by
financiado
porciento_financiado
vista
porciento_a_la_vista
colateral
porciento_colateral
tipo_deuda
asegurada
facilidad_linea_credito
plaza_proveedor
por_cuenta
referencia_autorizacion
pago_adelantado
estado
id_acuerdo
codigo_proveedor
numero_cheque
pago_vista
crc

Figura 5 Datos perspectiva Carta de Crédito.

Figura 6 Datos perspectiva histórico.

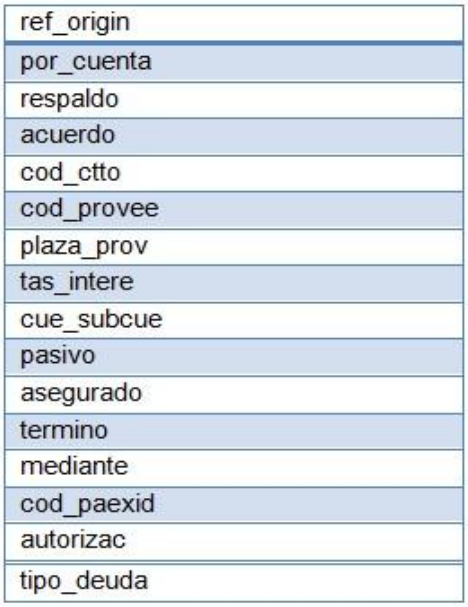

Figura 7 Datos perspectiva Financiamiento.

Figura 8 Datos perspectiva Garantía.

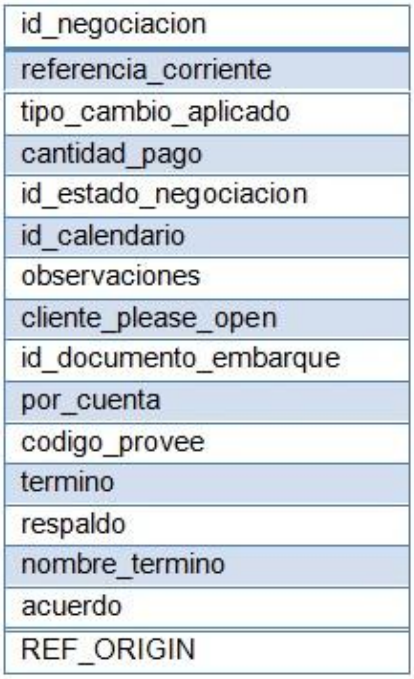

Figura 9 Datos perspectiva Negociación.

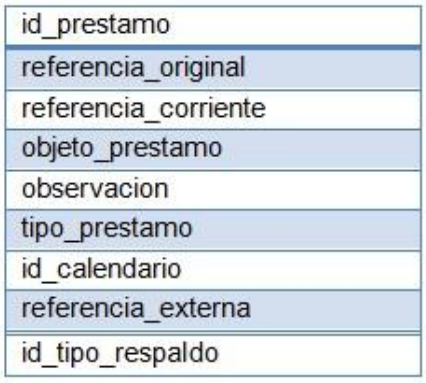

Figura 10 Datos perspectiva Préstamo.

Figura 11 Datos perspectiva Renegociación.

Después de describir los campos, se procede a ampliar el modelo conceptual, ubicando debajo de cada perspectiva los campos que la componen y de cada indicador la forma mediante la cual será calculado. Con este modelo culmina el análisis del Almacén de Datos. Como puede apreciarse, el modelo conceptual permite comprender cuáles serán los resultados que se obtendrán, las variables que se utilizarán para analizarlos y la relación que existe entre ellos.

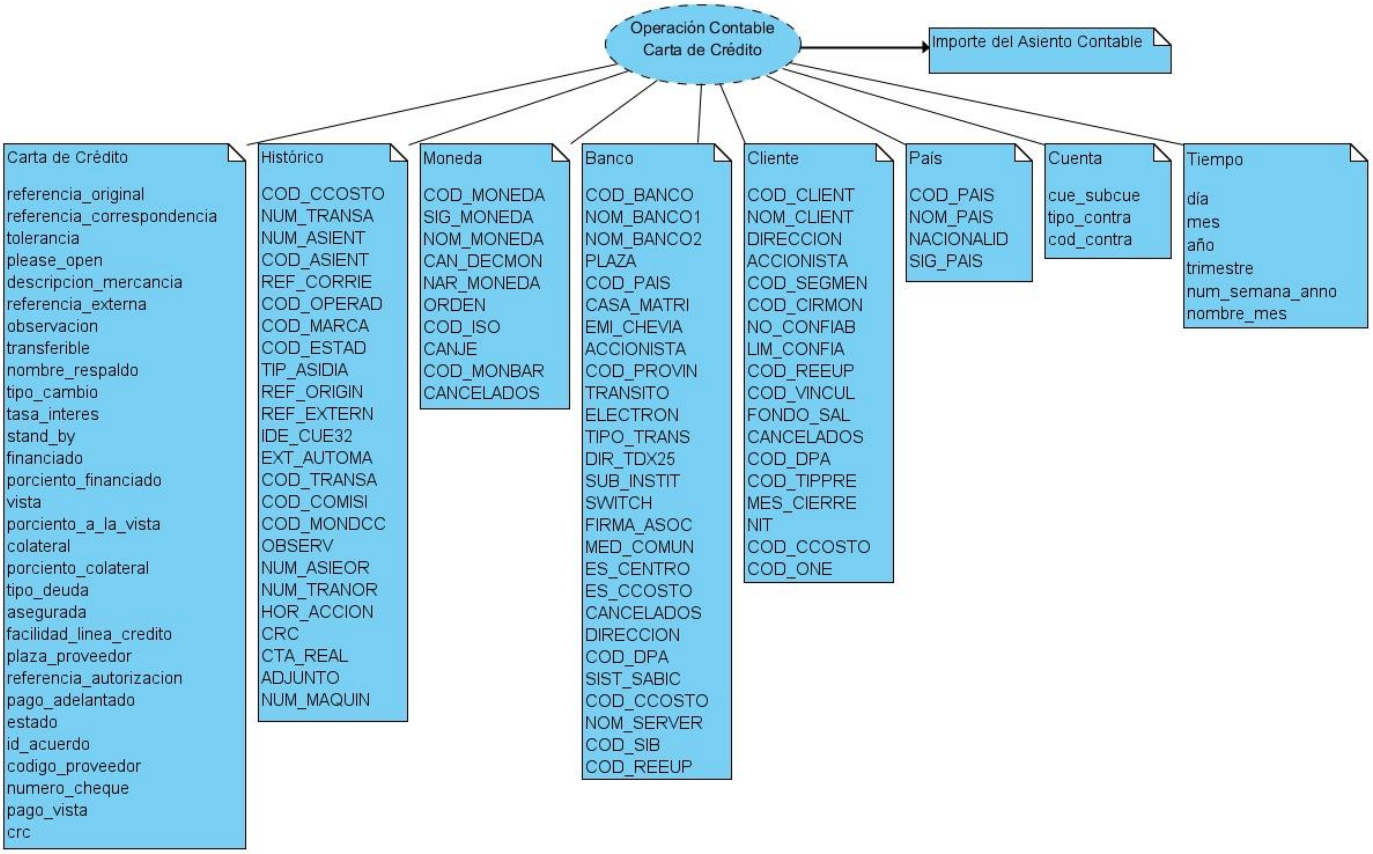

Figura 12 Modelo conceptual ampliado para Operación Contable de Carta de Crédito.

Además, se definieron los modelos conceptuales ampliados para las Operaciones Contables de Financiamiento, Garantía, Negociación, Préstamo y Renegociación (Ver anexos 6 al 10).

3. Elaboración del modelo lógico de la estructura del Almacén de Datos

En esta etapa se diseñó la estructura del Almacén de Datos, teniendo como base el modelo conceptual anterior. Para ello, primero se define el tipo de modelo que se utilizará y luego se llevaron a cabo las acciones propias al caso, para diseñar las tablas de dimensiones y de hechos. Finalmente, se realizan las uniones pertinentes entre estas tablas y se determinaron las jerarquías.

3.1. Modelo Lógico del Almacén de Datos

Las bases de datos multidimensionales implican tres variantes posibles de modelado, que permiten realizar consultas de soporte de decisión:

- Esquema en estrella (*Star Scheme*): Consta de una tabla de hechos central y de varias tablas de dimensiones relacionadas a esta, a través de sus respectivas claves. Este modelo debe estar totalmente desnormalizado, es decir, que no puede presentarse en tercera forma normal (3ra FN) (6).

- Esquema copo de nieve (*Snowflake Scheme*): Este esquema representa una extensión del modelo en estrella cuando las dimensiones se organizan en jerarquías de dimensiones (6).
- Esquema constelación o copo de estrellas (*Starflake Scheme*): Este modelo está compuesto por una serie de esquemas en estrella. Está formado por una tabla de hechos principal y por una o más tablas de hechos auxiliares, las cuales pueden ser sumarizaciones de la principal. Dichas tablas yacen en el centro del modelo y están relacionadas con sus respectivas tablas de dimensiones (6).

Luego de analizar las características de los diferentes esquemas se opta por utilizar un modelo híbrido el cual contemplará las características de los esquemas de Copo de Nieve y Constelación, debido a la complejidad del diseño del Almacén de Datos.

Estos esquemas, a su vez proporcionaran las siguientes ventajas al diseño:

Copo de nieve:

- Mejor utilización del espacio.
- Muy útil en tablas de dimensiones de muchas tuplas.
- Las dimensiones estarán normalizadas, por lo que requiere menos esfuerzo de diseño.

Constelación:

- Permitirá 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.
- Contribuirá a la reutilización de dimensiones, ya que una misma dimensión puede utilizarse para varias tablas de hechos.

3.2. Diseñar tablas de dimensiones

En este paso se diseñan las tablas de dimensiones que formarán parte del Almacén de Datos y que contienen la información clave de la entidad. Cada perspectiva definida en el modelo conceptual constituye una tabla de dimensión. Se toman cada una de las perspectivas con sus campos relacionados y se realiza el siguiente proceso:

- Se eligen los nombres para identificar cada tabla de dimensión.
- Se añaden los campos que representan su clave principal.
- Se añaden los campos definidos en cada perspectiva.

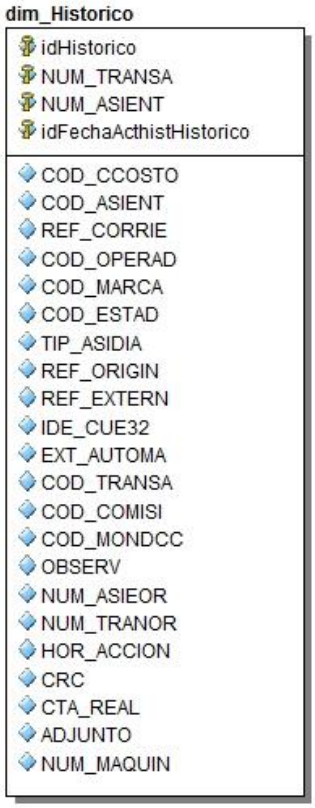

Figura 13 Dimensión Histórico.

Se definieron las dimensiones: dim_Banco, dim_Carta_Credito, dim_Cliente, dim_Cuenta, dim_Fecha, dim_Financiamiento, dim_Garantia, dim_Moneda, dim_Negociacion, dim_Pais, dim_Prestamo y dim_Renegociacion (Ver Anexo 11).

3.3. Diseñar tablas de hechos

En este paso, se definen las tablas de hechos, que son las que contendrán los hechos a través de los cuales se construirán los indicadores de estudio. Para cada hecho se realiza el siguiente proceso:

- Se define su clave primaria, que se compone de la combinación de las claves primarias de cada tabla de dimensión relacionada.
- Se crean tantos campos de hechos como indicadores definidos en el modelo conceptual.

Figura 14 Hecho Operación_Histórico_Préstamo.

Además, se definieron los hechos: Operación_Histórico_Carta_Crédito, Operación_Histórico_Financiamiento, Operación_Historico_Garantia, Operación_Historico_Negociacion y Operación_Historico_Renegociacion (Ver Anexo 12).

3.4. Realizar uniones

Luego de haber diseñado las tablas de dimensiones y las tablas de hechos a partir de los indicadores y las perspectivas identificadas en el modelo conceptual, se realiza el modelo lógico previo, el cual expresa las correspondencias que existen entre dichas tablas.

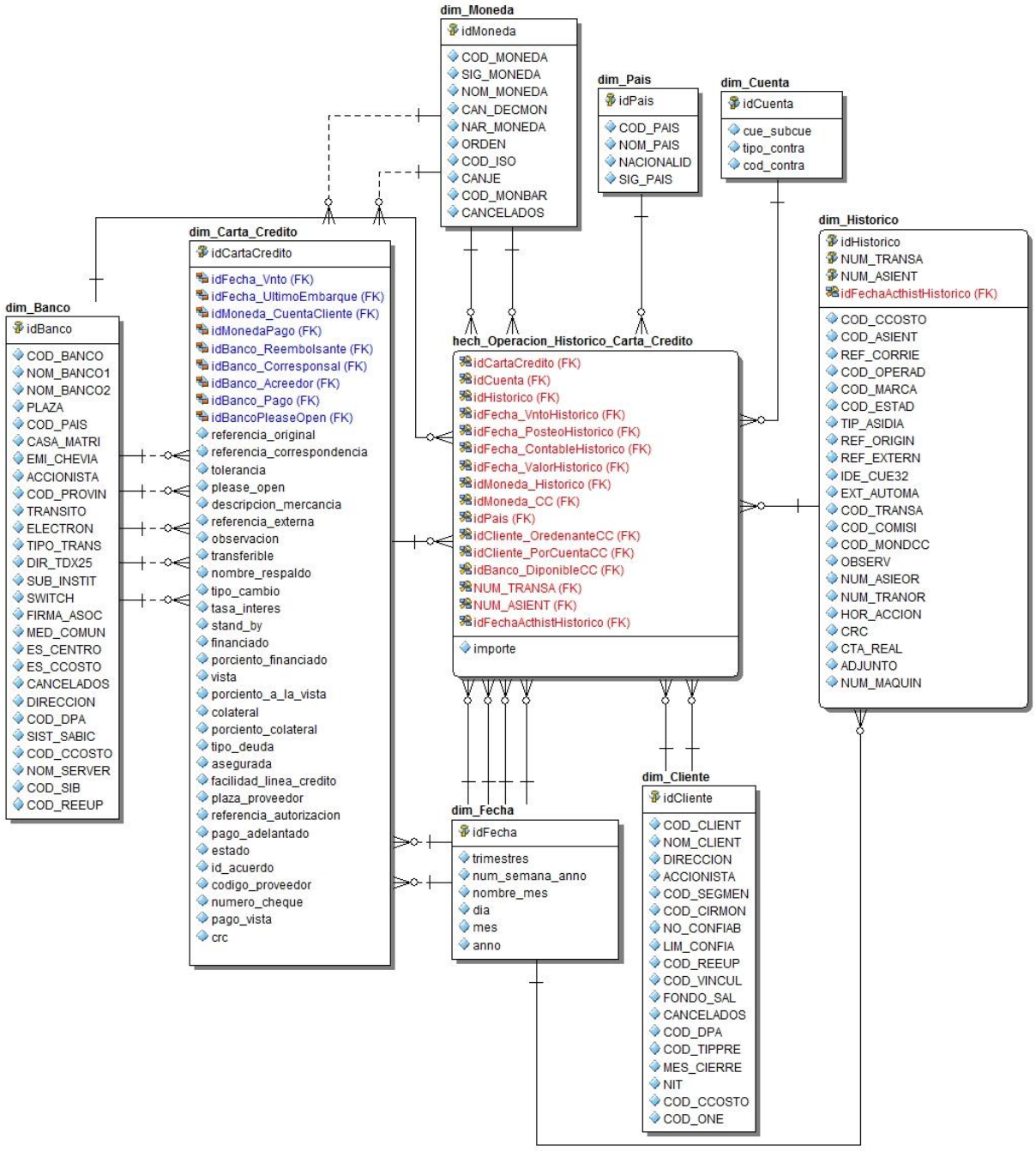

Figura 15 Modelo lógico para Carta de Crédito.

Además, se definieron los modelos lógicos para Financiamiento, Garantía, Negociación, Préstamo y Renegociación (Ver Anexos del 13 al 18).

3.5. Determinar jerarquías

No se determina ninguna jerarquía pues no se quiere analizar ningún dato desde un nivel general a otro más detallado y viceversa.

Conclusiones del capítulo

El presente capítulo fue desarrollado sobre los pasos definidos por la metodología Hefesto para el análisis y diseño de un Almacén de Datos, permitiendo obtener las tablas de dimensiones y las tablas de hecho que definirán la estructura del propio almacén. Además, permite un entendimiento del diseño realizado a partir de los indicadores definidos en la primera fase de la metodología empleada, proporcionando una entrada apropiada como punto de partida para la implementación del Almacén de Datos.

Capítulo 3: Implementación de un Almacén de Datos para el BNC

En este capítulo se define la implementación del Almacén de Datos para el Banco Nacional de Cuba, describiendo los procesos de Extracción, Transformación y Carga mediante la utilización de las herramientas descritas con anterioridad.

1. Propuesta de la Arquitectura de Integración de Datos

Una arquitectura en el ámbito computacional es un conjunto de estructuras o reglas que proveen un esqueleto para el diseño general de un producto o sistemas. No es recomendable iniciar el desarrollo de una solución sin haberla planificado, identificado sus fuentes, su esquema, el movimiento de los datos y determinado su enfoque de almacenamiento de datos. Para comenzar a describir la arquitectura de este proceso, es necesario recordar algunos elementos referentes a la implementación de este sistema como son:

- Fuente de datos.
- Área temporal.
- Almacén de Datos Operacional.

Fuente de datos:

Son los datos en bruto, que se encuentran almacenados en las bases de datos del sistema Quarxo y del sistema SABIC. Estos sufrirán un proceso de extracción hacia un servidor local, para facilitar el trabajo con la transformación y la homogeneidad de los tipos de datos, la información y los campos de las tablas.

Área temporal:

Constituye el área de preparación de datos para facilitar los procesos y técnicas de integración para luego ser cargados al destino. En este caso se utilizará el gestor *Microsoft SQL Server* 2005, creándose las tablas y atributos necesarios para dicho montaje.

Almacén de Datos Operacional:

El Almacén de Datos Operacional constituye el destino hacia donde se integrarán los datos a cargar, el cual responderá a las necesidades del negocio y de integración, en correspondencia con los procesos *ETL* que se implementarán. El mismo estará soportado físicamente en el gestor de base de datos *Microsoft SQL Server* 2005.

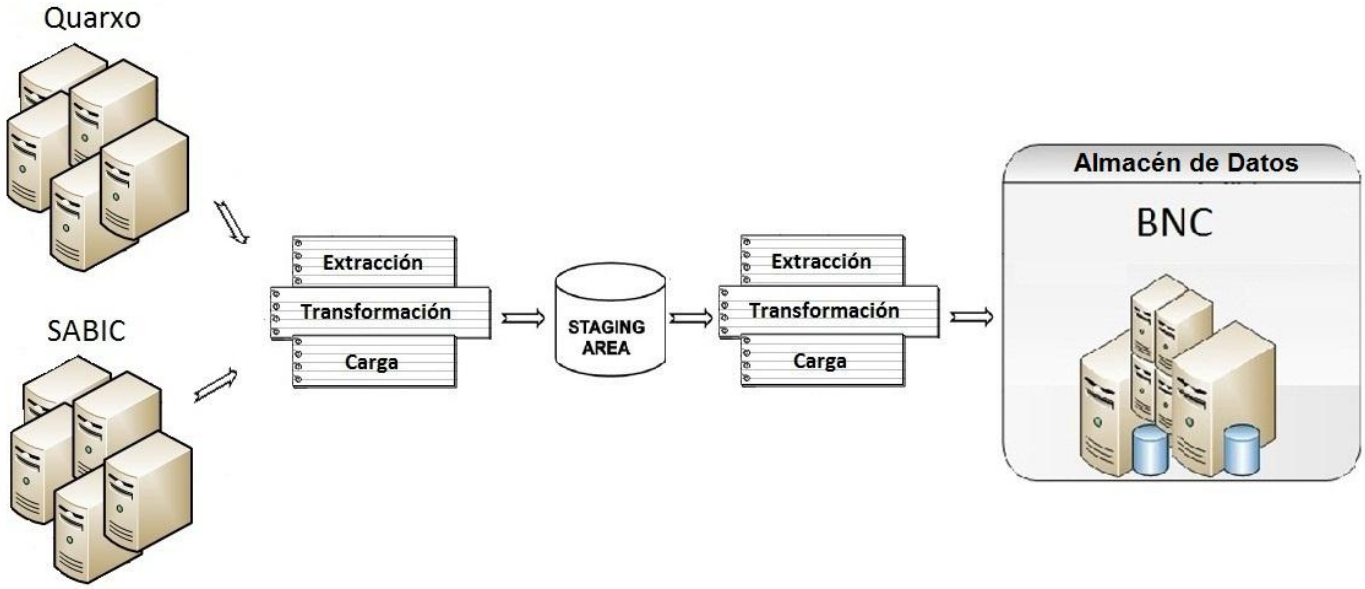

Figura 16 Arquitectura de integración de datos.

2. Aspectos generales de los sistemas fuentes.

Los sistemas externos que constituyen la fuente de integración son el sistema Quarxo y el sistema SABIC.

2.1. Sistema Quarxo

El sistema Quarxo contiene toda la información de las operaciones contables del BNC a partir del año 2007 hasta la fecha. La base de datos del mismo utiliza el gestor *Microsoft SQL Server* 2005.

2.2. Sistema SABIC

El sistema SABIC almacena la información de las operaciones contables del BNC hasta el año 2007. El mismo utiliza el sistema operativo *MS-DOS* y base de datos en *Visual FoxPro*, lo cual dificulta el acceso a los datos que posee la aplicación.

3. Aspectos generales de la base de datos intermedia

La estructura de las tablas de la base de datos intermedia se corresponde con la estructura de las tablas de la base de datos del sistema Quarxo. Para su diseño solo se utilizó la estructura de las 13 tablas que se utilizarán en el proceso de carga del Almacén de Datos.

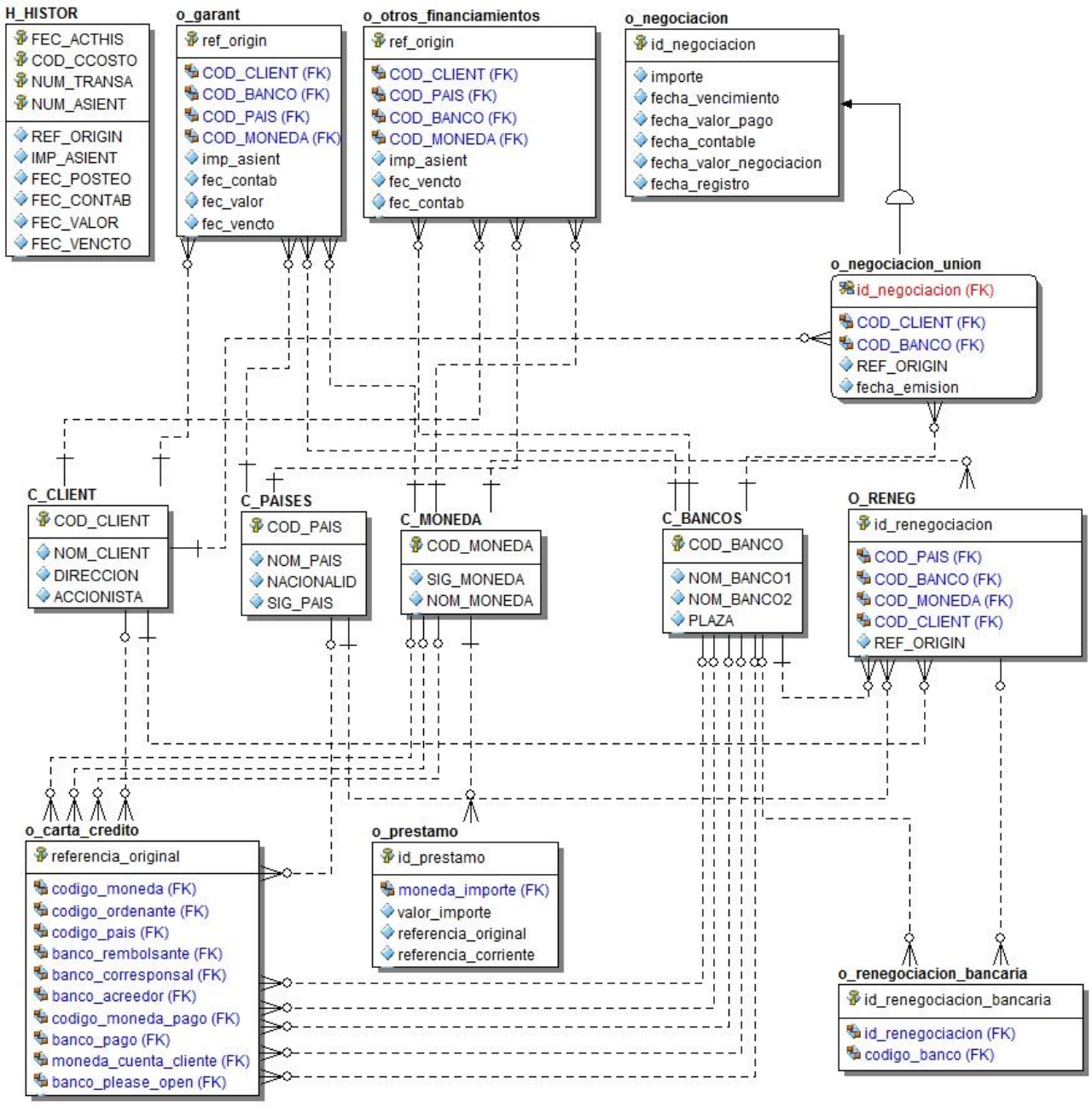

Figura 17 Estructura de la base de datos intermedia.

Como se aprecia en el diagrama, la base de datos intermedia está formada por las tablas de clasificadores, las cuáles contienen los datos de los clientes, países, monedas y bancos; las tablas de operativos que contiene la información de las cartas de crédito, préstamos, garantías, otros financiamientos, negociaciones y renegociaciones, además posee la tabla H_HISTOR que contiene los datos de la contabilidad.

Esta base de datos será utilizada con la misma frecuencia con la que se carguen datos al Almacén de Datos, para esto se le deberá borrar toda la información que contienen sus tablas e insertarle los nuevos datos que serán transformados y cargados en el propio almacén.

4. Procesos de Extracción, Transformación y Carga de Datos

El proceso de *ETL* se basa en controlar la fuente, la transformación correspondiente y el destino de los datos en todo el proceso. La transformación de los datos se hará de acuerdo a las reglas que se definieron en el negocio. Se definen transformaciones tales como: cambios de formato que aseguran la unicidad y estandarización de los tipos de datos. Para estos procesos *ETL* se utilizarán las ventajas que brinda la herramienta *Pentaho Data Integration*, como se explicó en el Capítulo 1.

4.1. Procesos de Extracción, Transformación y Carga de Datos hacia la base de datos intermedia

4.1.1. Procesos ETL para el sistema Quarxo

En la configuración de los procesos para este sistema inicialmente se definieron las tablas fuentes con sus respectivos atributos.

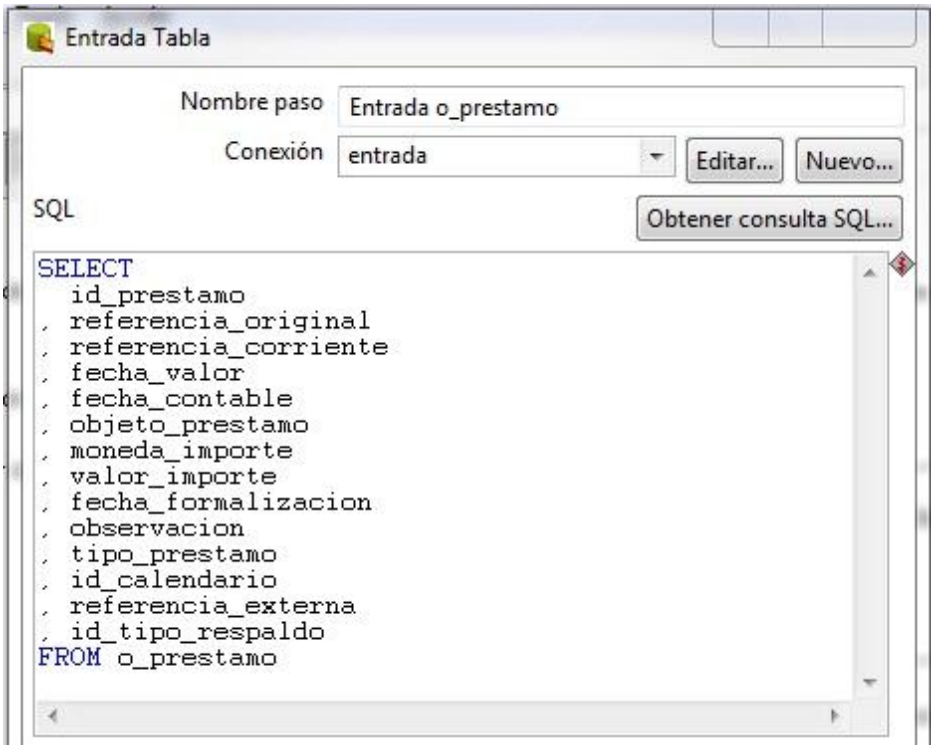

Figura 18 Entrada tabla o_prestamo.

Luego se realizaron las validaciones de datos nulos.

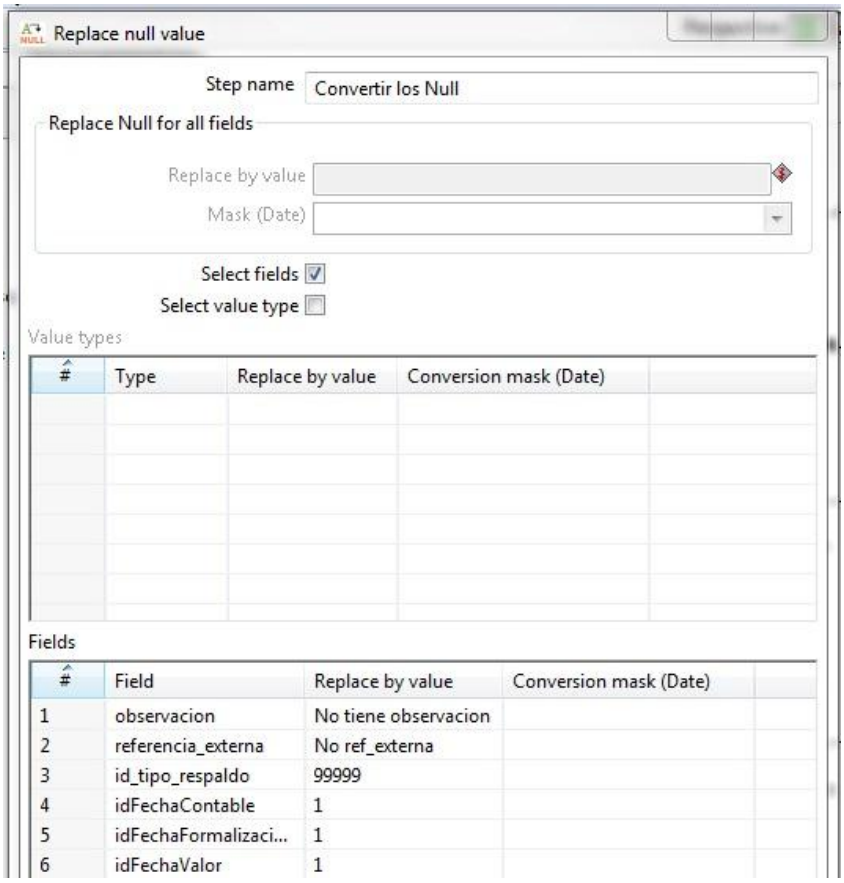

Figura 19 Conversión de los valores nulos.

Por último se definieron las tablas destino.

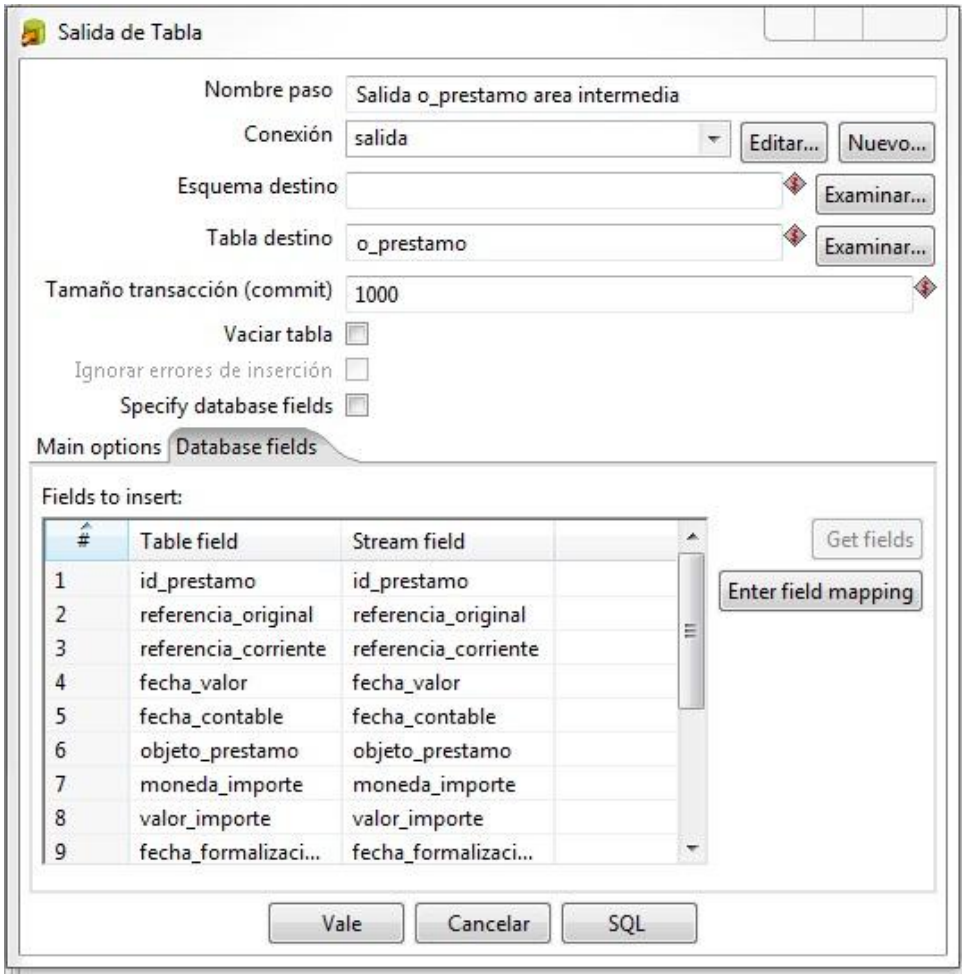

Figura 20 Salida tabla o_prestamo en la base de datos intermedia.

En la siguiente figura se muestra cómo se configuró el proceso para la tabla o_prestamo.

Figura 21 Proceso de carga de la tabla o_prestamo de Quarxo.

Las tablas restantes se cargaron de igual forma que el ejemplo anterior. Además, se implementó un trabajo para el proceso de carga en general en el cual se agrupan las cargas de todas las tablas; en este trabajo se cargaron inicialmente las tablas de clasificadores, ya que las tablas de operativos tienen

relaciones de llaves foráneas con dichos clasificadores, seguido se cargaron las tablas correspondientes a los operativos y por último la tabla que almacena la información de la contabilidad.

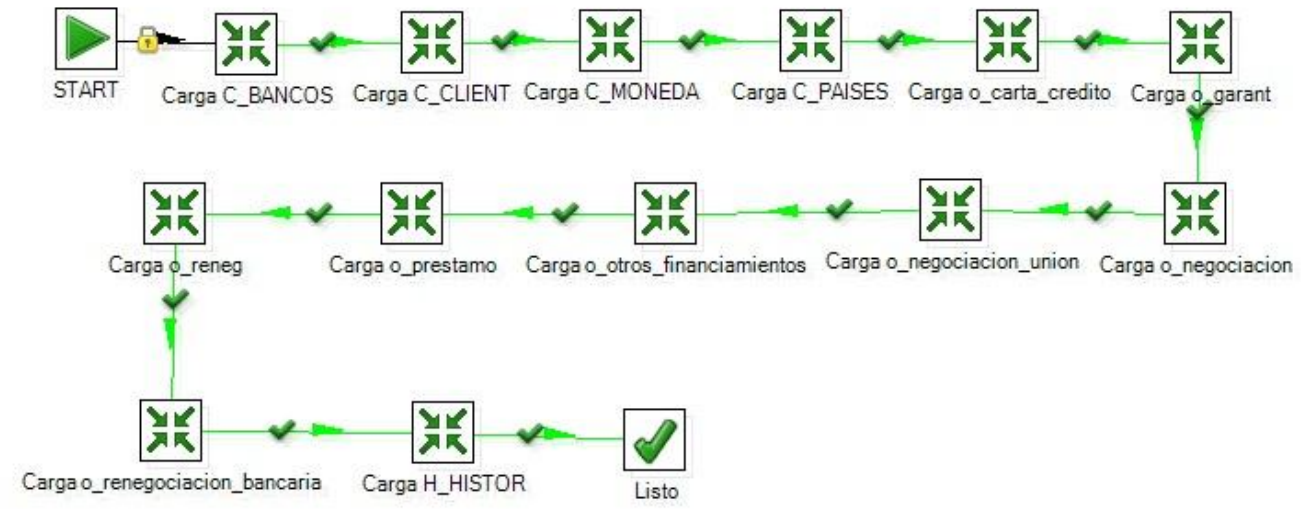

Figura 22 Proceso para unificar la carga del sistema Quarxo.

4.1.2. Procesos ETL para el sistema SABIC

Para la carga de los datos del sistema SABIC, inicialmente se cargó la base de datos de *Visual FoxPro* en *SQL Server* 2005 con la ayuda de la herramienta *ESF Database Convert* con el objetivo de estandarizar las bases de datos en un mismo gestor. A continuación se describe brevemente este proceso.

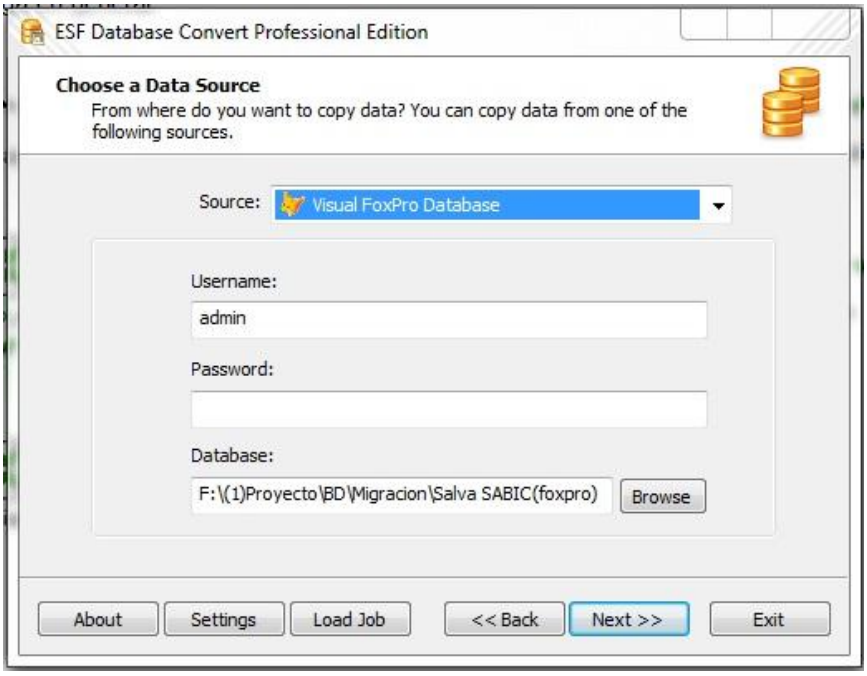

Figura 23 Selección de la base de datos origen en *Visual FoxPro.*

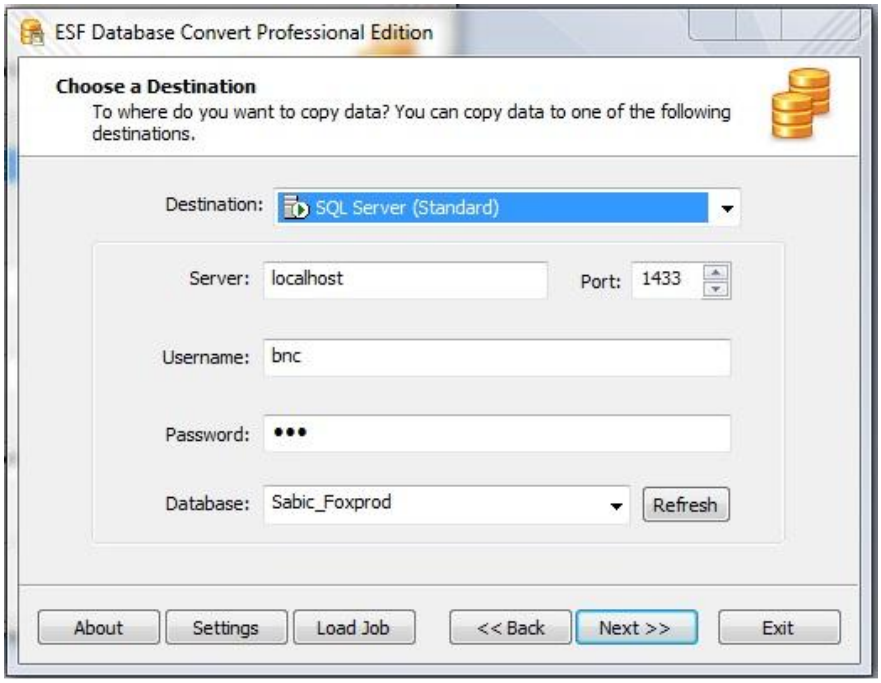

Figura 24 Selección de la base de datos destino en *SQL Server* 2005.

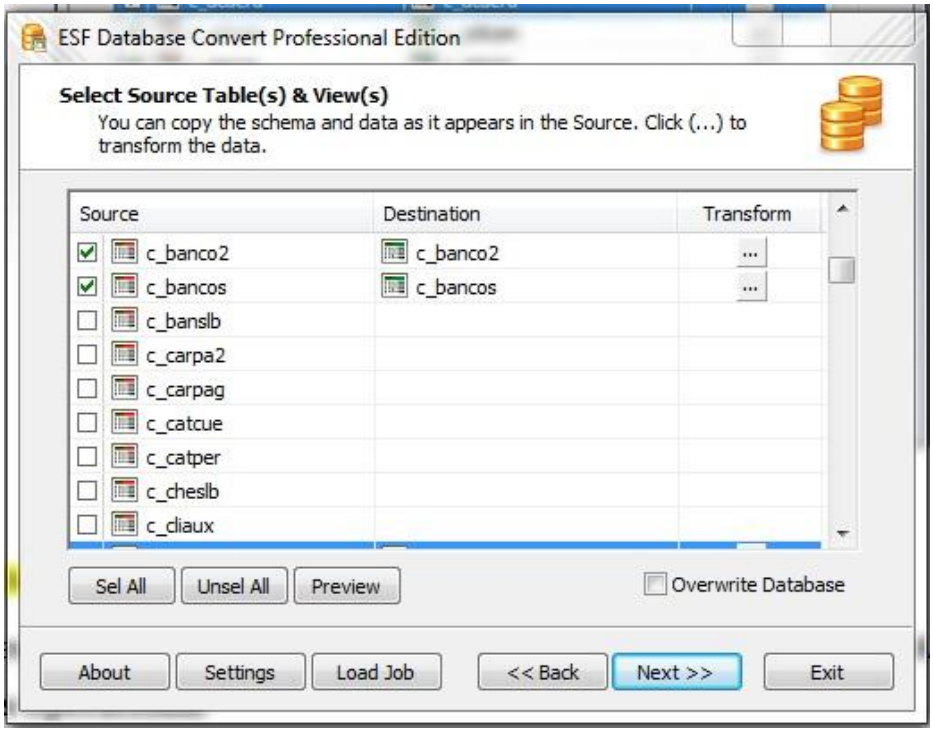

Figura 25 Selección de las tablas a cargar.

Como se aprecia en las imágenes anteriores el proceso comienza seleccionando la fuente de datos que se quiere transformar, luego se especifica la base de datos en la que se almacenara la información después de transformada y se procede a seleccionar las tablas a las que se realizara este proceso, una vez seleccionada las tablas se selecciona "*Next* (Siguiente)" hasta llegar al final del proceso.

4.1.2.1. Carga de las Cartas de Crédito

El proceso de carga de las cartas de crédito consta de 2 transformaciones las cuales se describen a continuación.

Para realizar la carga de las cartas de créditos se extraen los datos de la tabla O_CCRED de la base de datos, luego se renombraron los valores seleccionados en concordancia con los nombres de la tabla destino, seguidamente se modificaron los valores estado, por_cuenta y codigo_cliente, se hace una búsqueda de las cartas de crédito de apertura, se modifica el valor de la referencia original, se buscan los códigos de las monedas de la carta de crédito y la moneda de pago, se seleccionan los valores y se insertan en la tabla o_carta_credito de la base de datos intermedia.

En la siguiente figura se muestra el proceso descrito anteriormente.

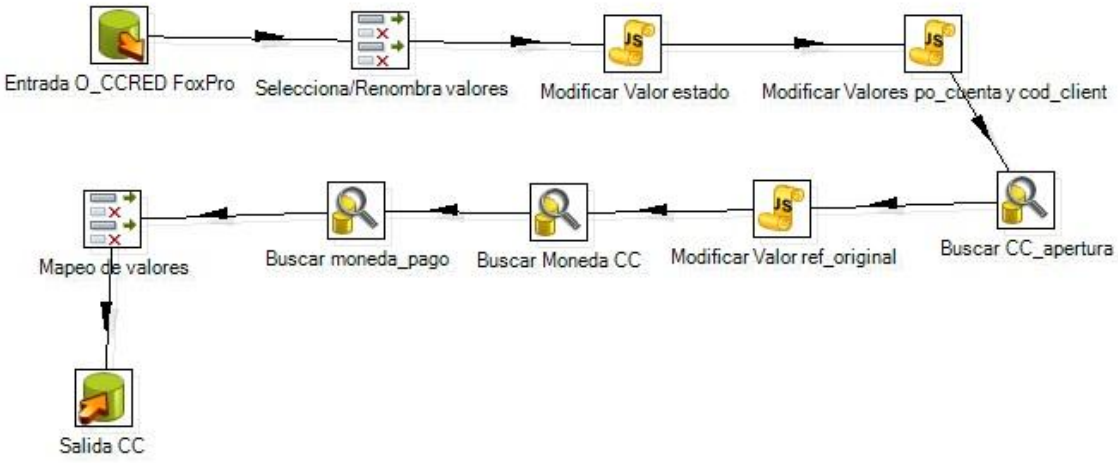

Figura 26 Primera transformación Carta de Crédito.

Para realizar la carga de las cartas de créditos de apertura se extraen los datos de la tabla O_CCRED_APERT de la base de datos, luego se busca el código de moneda, se renombran los valores y se insertan en la tabla o_carta_credito de la base de datos intermedia.

En la siguiente figura se muestra el proceso descrito anteriormente.

Figura 27 Transformación Carta de Crédito Apertura.

Además, se definió un trabajo para agrupar los procedimientos anteriores. El mismo inicia cargando los datos existentes en el operativo del sistema SABIC y estableciendo como valor predeterminado ***-1*** para los acuerdos que se encuentren en ***null***, luego borra la restricción que existe en la tabla de carta de crédito con la tabla acuerdo para permitir insertar en la misma el valor por defecto definido anteriormente, seguidamente ejecuta la primera transformación de carta de crédito y a continuación modifica a ***null*** los identificadores de acuerdos que tengan valor ***-1*** y además por petición del BNC los que tengan valor ***7;*** al ejecutarse este último paso se procede a establecer la relación existente entre carta de crédito y acuerdo para garantizar que se mantenga la restricción establecida en el futuro trabajo con la Base de Datos. Finalmente, se modifican los proveedores en las cartas de crédito almacenando

en este campo los códigos de proveedores correspondientes a los identificadores que poseían con anterioridad la carta de crédito.

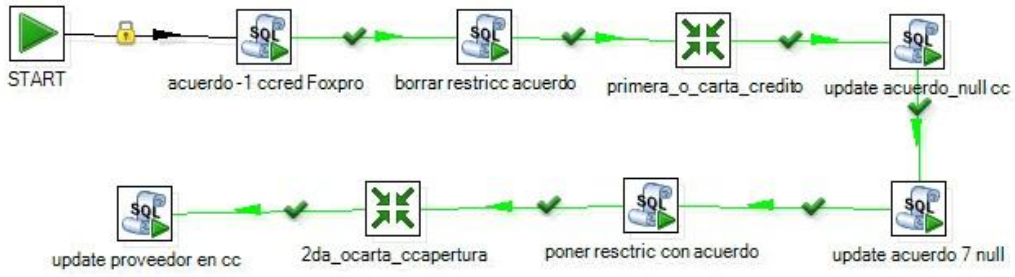

Figura 28 Trabajo para la carga de Carta de Crédito.

4.1.2.2. Carga de las Negociaciones

El proceso de carga de las negociaciones consta de 4 transformaciones, a continuación se describen algunas de ellas.

Para realizar la primera carga de las negociaciones se extraen los datos de la tabla M_DIARIO de la base de datos, luego se seleccionan los valores a utilizar, seguidamente se modificaron los valores importe, id_estado e id_documento, se añade una secuencia para autoincrementar los valores del campo id_negociacion, se renombren los valores y se insertan en la tabla o_negociacion de la base de datos intermedia.

En la siguiente figura se muestra el proceso descrito anteriormente.

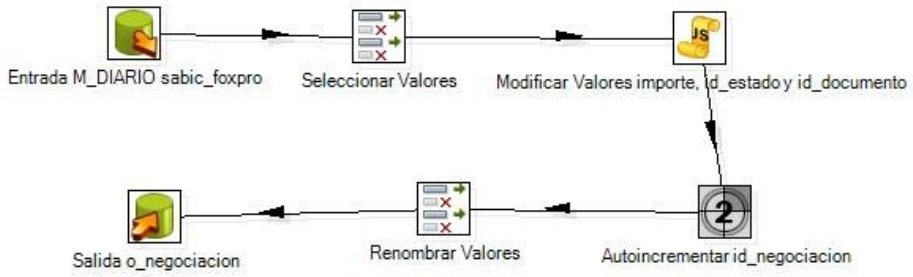

Figura 29 Primera carga Negociación.

Para realizar la carga de negociación_union se extraen los datos de la tabla o_negociacion cargada anteriormente, seguidamente se modificaron valor del campo nombre_termino, se renombran los valores y se insertan en la tabla o_negociacion_union de la base de datos intermedia.

En la siguiente figura se muestra el proceso descrito anteriormente.

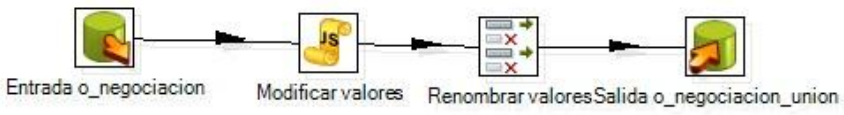

Figura 30 Carga o_negociacion_union.

Además, se definió un trabajo para agrupar las transformaciones definidas para esta área. En la primera transformación de manera general se cargan las negociaciones a partir de la tabla M_DIARIO, en la segunda se asignan valores a las fechas que se encuentran en ***null;*** en la tercera se cargan las negociaciones a partir de los datos obtenidos en la tabla union_negociacion_aux; finalmente se cargan la tabla o_negociacion_union a partir de los datos obtenidos de la tabla o_negociacion cargada al comienzo del proceso.

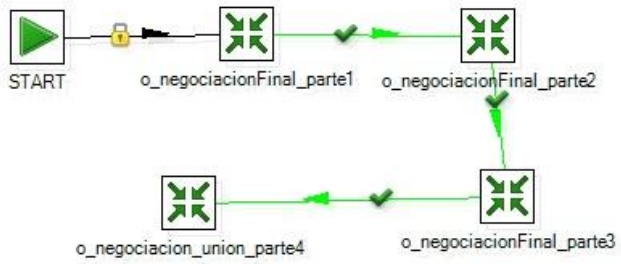

Figura 31 Trabajo para la carga de Negociación.

4.1.2.3. Carga de los Préstamos

El proceso de carga de los préstamos consta de una transformación que se explicará a continuación.

Para la carga de los préstamos se extraen los datos de la tabla O_PRESTA de la base de datos, luego se seleccionan los valores de los campos a utilizar; luego se modifican algunos valores utilizando el componente de *java script*, tales como tipo de operación, importe y fecha valor; se añade una secuencia para el identificador del préstamo y el identificador del calendario; luego se renombran los valores según

el formato de la tabla de salida; finalmente se adiciona en la tabla o_prestamo de la base de datos intermedia.

Figura 32 Carga de los préstamos del sistema SABIC.

Además se definió un trabajo para ejecutar la transformación definida anteriormente.

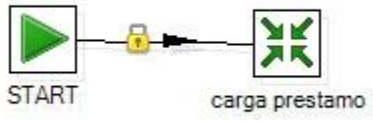

Figura 33 Trabajo para la carga de Préstamo.

4.1.2.4. Carga general

Para la carga general se definió un trabajo en el cual se agrupan los trabajos definidos anteriormente para cada una de las áreas.

Figura 34 Trabajo general sistema SABIC.

4.2. Procesos de Extracción, Transformación y Carga de Datos hacia el Almacén de Datos Operacional

4.2.1. Implementación de las transformaciones

Las transformaciones están compuestas por pasos, que constituyen el elemento más pequeño de la misma y además se encuentran unidos a través de saltos. Una vez que se toma la decisión de qué

reglas de transformación serán establecidas, deben crearse e incluirse las definiciones en las rutinas de transformación.

Para realizar la extracción de los datos correspondientes a cada una de las tablas de dimensiones, se accede a la base de datos fuente, de donde son extraídos los campos necesarios y se procede a realizar las transformaciones pertinentes.

4.2.1.1. Carga de la dimensión dim_Prestamo

La dimensión préstamo contiene todos los datos de los préstamos otorgados por el BNC, contiene propiedades como referencia original, fecha contable, fecha de formalización, y otras características que lo describen aún más.

Para realizar esta trasformación, se extraen inicialmente los datos de la tabla o_prestamo, luego se hacen las transformaciones necesarias para obtener las fechas valor, fecha contable y fecha formalización, se validan los datos y se asignan valores constantes a los campos que contengan valores nulos y por último se insertan los datos en la dimensión préstamo.

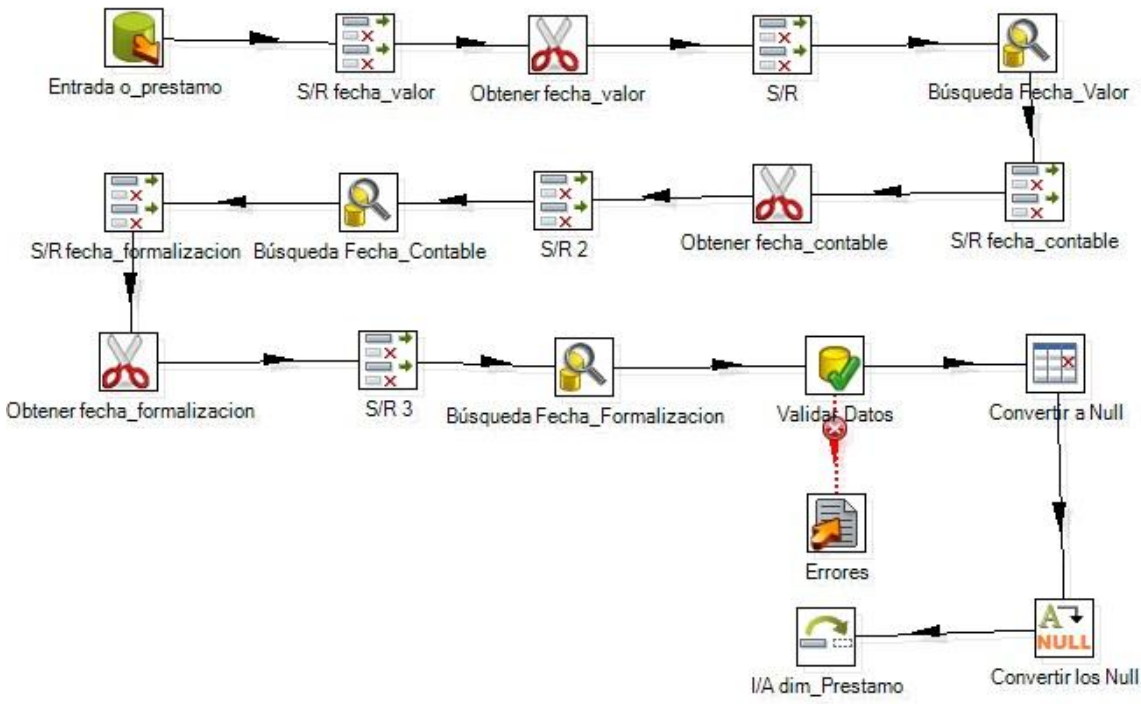

Figura 35 Carga de la dimensión Préstamo.

Para realizar la extracción de los datos correspondientes a cada una de las tablas de hechos, se accede a la base de datos fuente, de donde son extraídos los campos necesarios atendiendo a las dimensiones con que se relaciona cada hecho y se procede a realizar las transformaciones pertinentes.

4.2.1.2. Carga del hecho hech_Operacion_Historico_Prestamo

Este hecho contiene toda la información referente a las operaciones contables de los préstamos, tales como importe del asiento contable, así como los identificadores de las dimensiones que se relacionan con dicha tabla de hechos.

Para realizar esta trasformación, se extraen inicialmente los datos de la tabla H_HISTOR; luego se realiza un proceso de filtrado para comprobar que corresponda con un préstamo haciendo una búsqueda en la tabla de dimensiones de los préstamos cargados, si el resultado devuelto por la búsqueda es nulo no se realiza el proceso para esa tupla, ya que no corresponde con un préstamo; en caso contrario se hacen las transformaciones necesarias para obtener la fecha de actualización del histórico, de forma similar se transforman las fechas de vencimiento, de posteo, contable y de valor; seguidamente se buscan los identificadores de las diferentes dimensiones relacionada con la tabla de hechos; se asignan valores constantes a los campos que contengan valores nulos y por último se insertan los datos en la tabla de hechos hech_Operacion_Historico_Prestamo.

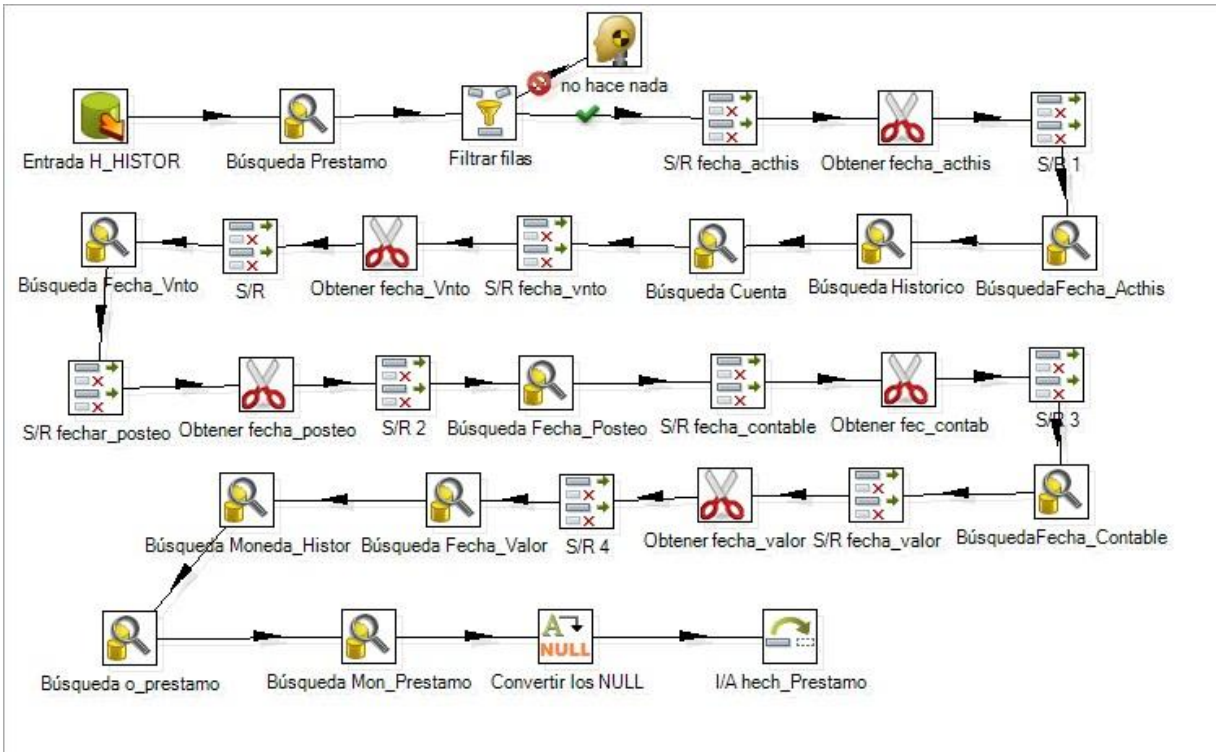

Figura 36 Carga del hecho Operación_Histórico_Préstamo.

4.2.2. Implementación de los trabajos

En el contexto de integración de datos, el término trabajo o *Job* se entiende como un conjunto sencillo o complejo de tareas cuyo objetivo consiste en realizar una acción determinada. La implementación de un trabajo define una secuencia lógica para la ejecución de las transformaciones, mediante el uso de pasos definidos, los cuales son diferentes a los disponibles en las transformaciones. Además, es posible ejecutar una o varias transformaciones de las que se hayan diseñado y orquestar una secuencia de ejecución para ellas. Los trabajos se encuentran en un nivel superior a las transformaciones.

Para definir el orden de la carga de los datos se crea un trabajo donde inicialmente se verifica que el servidor donde se encuentra el Almacén de Datos y la base de datos intermedia esté funcionando, si no está funcionando se manda un mensaje al usuario y se aborta la ejecución del trabajo. Si el servidor está en funcionamiento entonces se ejecutan las transformaciones en el orden determinado, comenzando con la carga de las tablas de dimensiones y luego con la carga de las tablas de hechos. Cuando todas las transformaciones se hayan ejecutado se le comunica al usuario.

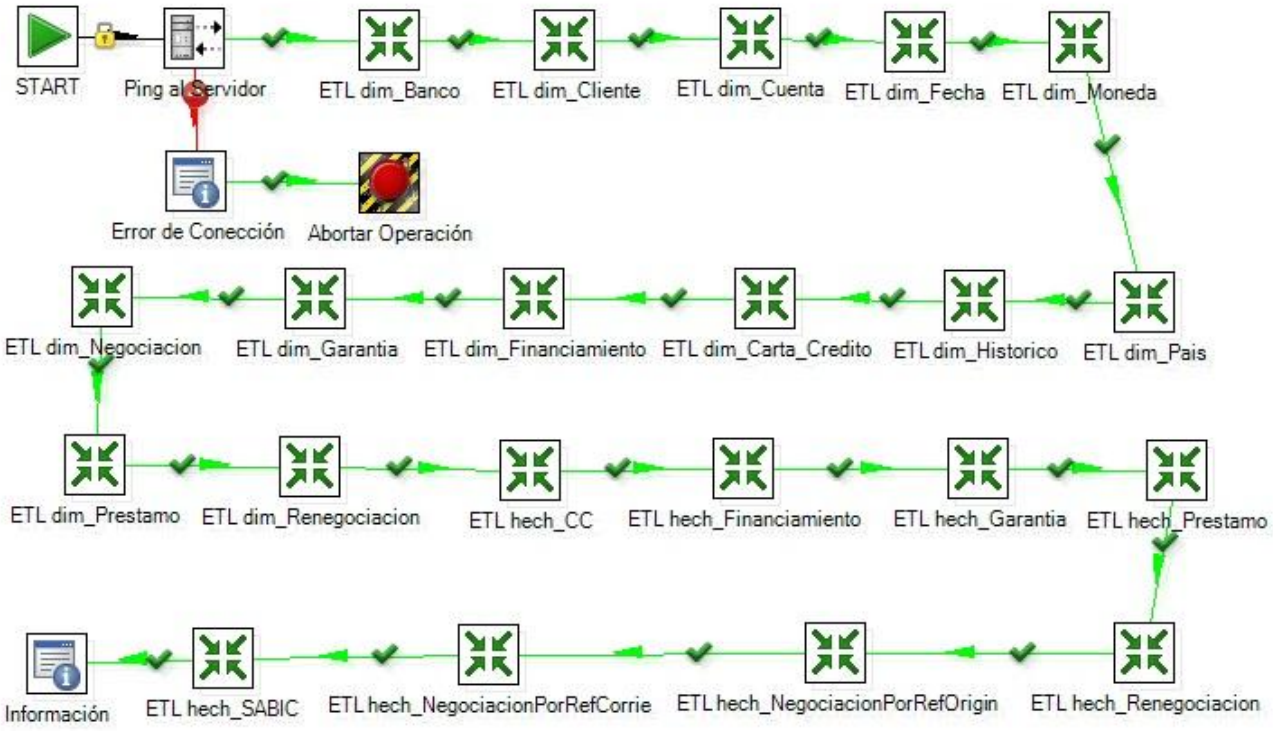

Figura 37 Trabajo o *JOB* para el Proceso ETL.

Conclusiones de capítulo

En este capítulo mostraron los principales conceptos que componen los procesos de *ETL*. Se definió la arquitectura que se utilizó en la solución y las características a tener en cuenta para seleccionar las fuentes de datos. Adicionalmente se fundamentó la necesidad de utilizar un área de almacenamiento intermedio para sustentar la solución informática. Además, se dio a conocer como quedó físicamente el Almacén de Datos, logrando una estructura robusta entre las relaciones de las tablas de hechos y dimensiones.

Capítulo 4: Validación de la Solución.

En este capítulo se llevan a cabo las pruebas para valorar el rendimiento del Almacén de Datos, así como la calidad de los datos almacenados, ya que esta etapa resulta tan importante como el diseño y la implementación del propio Almacén de Datos.

1. Calidad de los Datos

El término "calidad de datos" es asociado a los sistemas de información, teniendo en cuenta que dentro de los procesos que manejan grandes volúmenes de datos, puede encontrarse información incompleta e inconsistente (17).

En un Almacén de Datos son más frecuentes estos problemas ya que se utilizan varias fuentes de datos para su almacenamiento.

1.1. Perfilado de Datos

El perfilado de datos es el proceso que permite recopilar información sobre los datos existentes. Fue utilizada la herramienta *DataCleaner* en su versión 1.5.3 para realizar el perfilado de datos a la base de datos relacional que almacena la información del Almacén de Datos. Los reportes arrojados por este proceso indican que las cargas de los datos correspondientes a las tablas de hechos se realizaron correctamente. No fueron almacenados valores vacíos ni nulos. Este proceso se realizó para todas las tablas de hechos (Ver Anexos del 37 al 43).

El siguiente gráfico muestra los resultados correspondientes al perfilado de datos realizado a la tabla hech_Operacion_Historico_SABIC.

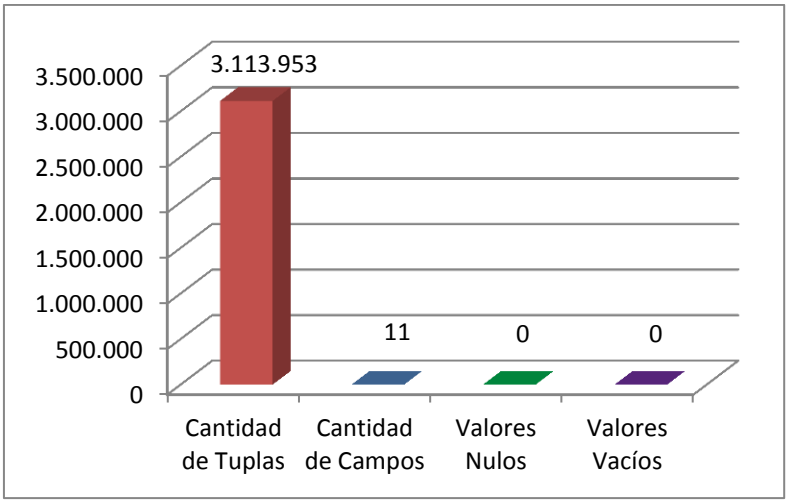

Figura 38 Resultados perfilado de datos hech_Operacion_Historico_SABIC.

2. Pruebas de Volumen y Carga

Las pruebas de volumen son pruebas típicas de entornos que utilicen bases de datos. Las mismas se realizan con el objetivo de analizar el comportamiento del sistema o base de datos con volúmenes de datos almacenados lo más similar posible a los esperados en la explotación real del sistema (18).

Las pruebas se realizaron bajo las siguientes condiciones:

Fuente de datos a reportar: Almacén de Datos en *SQL Server* 2005.

Tipo de consulta a realizar: Consultas *SQL*.

Características del *Hardware* del Servidor:

- *Hardware*: 1Gb de memoria RAM, 160 Gb de capacidad de disco duro *HDD*, procesador *Intel Core* 2 *Duo* a 2.2 GHz de velocidad.
- *Software*: SO *Microsoft Windows* 7 *Ultimate*, *SQL Server* 2005.

Para realizar estos tipos de pruebas se utilizaron los beneficios que brinda la herramienta *DataGenerator para SQL Server*. Este es un generador de carga diseñado para la realización de pruebas de este tipo. Genera carga por diversos protocolos, ya sea, *HTTP*, *SQL*. Al aplicarse este tipo de prueba sobre el almacén se observó que:

Al realizarse a la tabla hech_Operacion_Historico_SABIC la siguiente consulta, SELECT * FROM hech_Operacion_Historico_SABIC, se obtuvo:

Cantidad de Datos Cargados	750.000	1.500.000	3.000.000
Tiempo de demora (seg) en Realizar la Consulta	14	28	65

Tabla 1 Resultados prueba volumen y carga.

En el siguiente gráfico se muestra el comportamiento de este resultado:

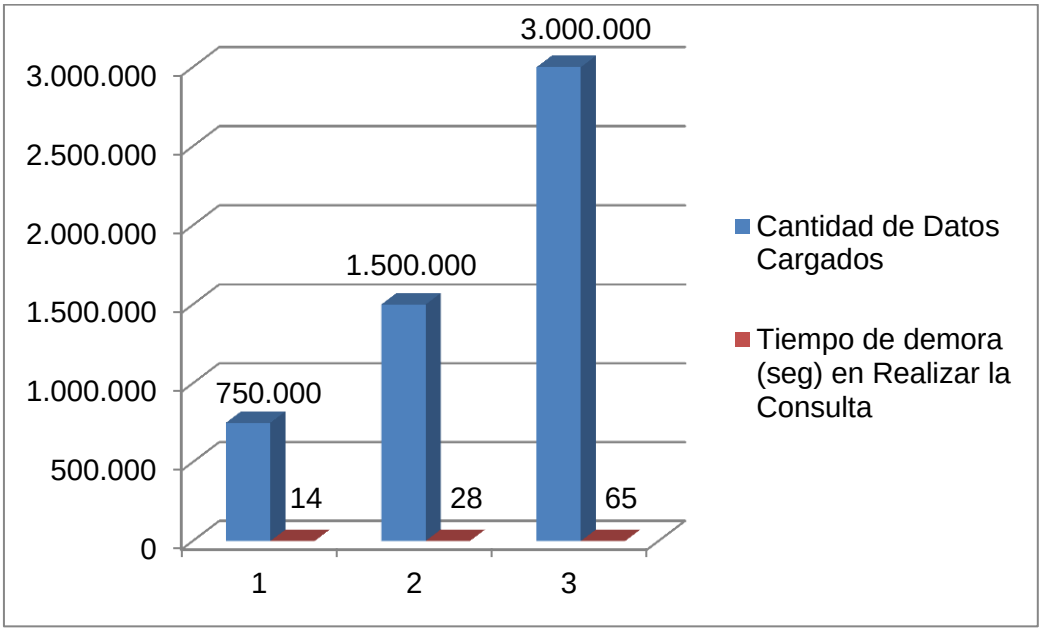

Figura 39 Resultado prueba volumen y carga hech_Operacion_Historico_SABIC.

Para interpretar los resultados de estas pruebas hay que tener en cuenta los siguientes aspectos:

- **El tiempo de respuesta aumenta demasiado cuando la base de datos se llena mucho:** El tiempo de respuesta no debería aumentar demasiado si se pasa de una base de datos con 100 filas en sus tablas a una con 50 000 filas. La tecnología de indexado de bases de datos hace que hallar una fila en una tabla tarde unos cuantos milisegundos, incluso si hay cientos de miles de filas. Por lo tanto, si el tiempo de respuesta aumenta mucho después de pasar de una base de datos de tamaño moderado a una de tamaño extremo, entonces probablemente aún no se han indexado las columnas apropiadas (18).

Los resultados arrojados se enmarcan en 2 variables básicas, la cantidad de datos cargados y el tiempo de respuesta de la base de datos ante una consulta realizada. Esta prueba se organizó en tres fases, en las que se fueron doblando las cantidades de datos a consultar con respecto a la anterior. Los resultados obtenidos fueron satisfactorios porque a pesar de que las cantidades de datos cargados eran

grandes, los tiempos de respuesta de la base de datos fueron aceptables dado que es un almacén de datos y la cantidad de información que se maneja es elevada.

Además, al hacer la carga del almacén no se presentaron problemas de límite de capacidad, ni de volumen. Las llaves autogeneradas no se salieron del rango especificado, ni se detectaron problemas con los tipos de datos definidos en el paso de diseño. Todo esto verifica que *SQL Server* 2005 como gestor utilizado y el diseño de las estructuras de la base de datos implementadas soportan completamente el almacenamiento de los niveles de información requeridos para la puesta en producción del Almacén de Datos.

Conclusiones del capítulo

En este capítulo se realizaron las pruebas de volumen y carga las cuáles validaron la infraestructura de *hardware* y *software* propuestas, garantizando la capacidad de gestión de los datos almacenados. Las pruebas de carga resultaron un elemento fundamental en el proceso de optimización y demostraron que los tiempos de respuestas fueron aceptables. El proceso de perfilado de datos realizado permitió examinar los datos existentes y obtener estadísticas e información sobre los mismos.

Conclusiones Generales

Con el desarrollo del trabajo de diploma se cumplió con los objetivos propuestos:

- La caracterización de la metodología, tecnologías y herramientas definidas para el desarrollo del Almacén de Datos establecieron las bases teóricas para darle solución al problema planteado.
- La ejecución de los procesos definidos por la metodología utilizada en las etapas de Análisis de Requerimientos, Análisis de los sistemas transaccionales y Diseño del Almacén de datos, en función del problema planteado, permitió obtener el diseño de una solución capaz de mantener integrada toda la información relacionada con la contabilidad en el BNC.
- La implementación de los procesos *ETL* permitió obtener un Almacén de Datos Operacional poblado de la información histórica de las operaciones contables del BNC garantizando la consulta de los mismos por parte de los especialistas de la institución.
- La validación de la solución a través de pruebas realizadas al Almacén de Datos demostró que el mismo cumple con los requerimientos necesarios para ser usado por el BNC.

Recomendaciones

Con el objetivo de mejorar el uso del Almacén de Datos en el BNC se recomienda:

- Mantener actualizado el contenido del Almacén de Datos, incorporando una mayor cantidad de datos.
- Implementar minerías de datos sobre la información almacenada en el Almacén de Datos, debido a que este tipo de análisis prepara, sondea y explora los datos para sacar la información oculta en ellos contribuyendo al proceso de toma de decisiones del BNC.

Trabajos Citados

1. **Martín Correa, Jorge Luis y Leandro Sosa, Alejandro.** *Módulo de seguridad para el sistema informático del Banco Nacional de Cuba.* La Habana : s.n., 2010.

2. **Quintana Ramírez, Yoan Asdrubal .** *Diseño e Implementación de los módulos Chequera y Transferencias del sistema Quarxo.* La Habana : s.n., 2011.

3. **Inmon, William H.** *Building the Data Warehouse Third Edition.* s.l. : John Wiley & Sons, Inc., 2002.

4. **Kimball.** *The Data Warehouse Lifecycle Toolkit.* New York : Wiley, 1998.

5. **Imhoff, Claudia, Galemmo, Nicholas y G. Geiger, Jonathan.** *Mastering Data Warehouse Design Relational and Dimensional Techniques.* Indianapolis : Wiley Publishing, Inc., 2003. ISBN: 0-471-32421-3.

6. **Bernabeu, Ricardo Darío.** *DATA WAREHOUSING:Investigación y Sistematización de Conceptos - HEFESTO: Metodología propia para la Construcción de un Data Warehouse.* Córdoba, Argentina : s.n., 2007.

7. **Kimball, Ralph y Ross, Margy .** *The Data Warehouse Toolkit Second Edition The Complete Guide to Dimensional Modeling.* s.l. : John Wiley & Sons, Inc., 2002.

8. **Espinosa, Roberto.** El Rincon del BI. *Fases en la implantación de un sistema DW. Metodologia para la construcción de un DW.* [En línea] 5 de Diciembre de 2009. [Citado el: 4 de Diciembre de 2012.] http://churriwifi.wordpress.com/2009/12/05/5-fases-en-la-implantacion-de-un-sistema-dw-metodologia-para-la-construccion-de-un-dw/.

9. **Huamantumba, Rayner.** *Manual para diseño y desarrollo de Datamart.* 2007.

10. ER/Studio Portal | Web Based Collaborative Data Model Reporting. [En línea] Embarcadero Technologies: Database Tools and Developer Software. [Citado el: 10 de Diciembre de 2012.] www.embarcadero.com/products/er-studio-portal.

11. **TANG, Z y J., MACLENNAN.** *Data Mining with SQL Server 2005.* Indianápolis : Wiley Publishing, Inc, 2005.

12. Gravitar. Información sin Límites. [En línea] [Citado el: 7 de Diciembre de 2012.] http://www.gravitar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/.

13. **McKnight, William; , SVP Data Warehousing; , Conversion.** *Choosing Microsoft SQL Server 2005 for Data Warehousing.* s.l. : Microsoft Corporation, 2006.

14. PHPNuke. [En línea] PHPNuke, Inc., 2013 . [Citado el: 10 de mayo de 2013.] http://downloads.phpnuke.org/es/download-item-view-g-m-m-l-z/ESF_Database_Convert.htm.

15. SOFTPEDIA. [En línea] Softpedia, 2013. [Citado el: 10 de mayo de 2013.] http://www.softpedia.es/programa-DataCleaner-160026.html.

16. SOFTPEDIA. [En línea] Softpedia, 2013. [Citado el: 10 de mayo de 2013.] http://www.softpedia.es/programa-Datanamic-Data-Generator-for-MS-SQL-Server-197943.html.

17. **Hernández Monteagudo, Yohana y Buchillón Soris, Adalennis .** *Mercado de datos Gestión académica para la Sala Situacional de la Universidad de las Ciencias Informáticas.* Ciudad de la Habana : s.n., 2012.

18. **Yoemny González Almaquer .** Ilustrados. Una Comunidad Educativa Mundial. [En línea] Ilustrados, 2011. [Citado el: 10 de mayo de 2013.] http://www.ilustrados.com/tema/11762/Evaluacion-desempeno-aplicaciones.html.

19. **Fernández Álvarez, Imias y Hernández Armas, Yulienni.** *ANÁLISIS DE UN ALMACÉN DE DATOS PARA LA RED NACIONAL DE GENÉTICA MÉDICA.* CIUDAD DE LA HABANA : s.n., 2010.

20. **Rivadera, Gustavo R.** *La metodología de Kimball para el diseño de almacenes de datos (Data warehouses).* 2010.

21. **Kimball, Ralph y Caserta, Joe .** *The Data Warehouse ETL Toolkit.Practical Techniques for Extracting, Cleaning, Conforming, andDelivering Data.* Indianapolis, : Wiley Publishing, Inc., 2004. IN 46256.

22. **Inmon, William H.** *Building the Data Warehouse, Fourth Edition.* Indianapolis : Wiley Publishing, Inc., 2005. IN 46256.

23. —. *THE DATA WAREHOUSE ENVIRONMENT:QUANTIFYING COST JUSTIFICATIONAND RETURN ON INVESTMENT.* 2000.

24. **Mundy, Joy, Thornthwaite, Warren y Kimball, Ralph.** *The Microsoft Data Warehouse Toolkit With SQL Server™ 2005 and the Microsoft® Business Intelligence Toolset.* Indianapolis : Wiley Publishing, Inc., 2006. IN 46256.

25. blogspot. [En línea] 4 de Febrero de 2008. [Citado el: 30 de Noviembre de 2012.] http://rimenri.blogspot.com/2008/02/inteligencia-de-negocios-business.html.

26. **Sommerville, Ian.** *Software Engineering.* Vol. 8va edición.

27. Sitio Oficial Visual Paradigm. [En línea] [Citado el: 19 de Enero de 2013.] http://www.visual-paradigm.com.

28. Sitio Oficial PostgresSQL. [En línea] [Citado el: 19 de Enero de 2013.] http://www.postgresql.org/.

29. Pentaho Open Source Bussiness Intelligence. [En línea] [Citado el: 19 de Enero de 2013.] http://www.pentaho.com.

30. **Microsoft.** Microsoft SQL Server. [En línea] [Citado el: 3 de Diciembre de 2012.] http://www.microsoft.com/en-us/sqlserver/default.aspx.

31. **Peña Torres, Hector y García Díaz, Efraín.** *Desarrollo del subsistema Reportes de Quarxo.* La Habna : s.n., 2012.

32. **Microsoft Partner.** SolidQ. [En línea] Microsoft , 2002. [Citado el: 4 de Abril de 2013.] http://blogs.solidq.com/BICorner/Post.aspx?ID=171&title=Curso+MS+Business+Intelligence-SSAS%3a+Agregaciones+%2836%29.

33. **Microsoft.** MSDN. [En línea] [Citado el: 4 de Abril de 2013.] http://msdn.microsoft.com/es-es/library/ms170208%28v=SQL.90%29.aspx.

34. How-to Geek. [En línea] HowToGeek, 2006. [Citado el: 20 de Febrero de 2013.] http://www.howtogeek.com/howto/database/reset-identity-column-value-in-sql-server/.

35. **Escobar Domínguez, René Raydel .** *Desarrollo de un Almacén de Datos para el Control del Consumo de Portadores Energéticos en la Oficina Nacional de Estadística.* Ciudad de La Habana : s.n., 2010.

36. Fabulous Force Database Tools. [En línea] fabFORCE.net, 2003. [Citado el: 10 de Diciembre de 2012.] http://www.fabforce.net/dbdesigner4/.

37. Oracle Database. [En línea] Oracle. [Citado el: 12 de Diciembre de 2012.] http://www.oracle.com/us/products/database/overview/index.html.

Anexos