

*Universidad de las Ciencias Informáticas*

*Facultad 15*



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

***Título:*** *Diseño e implementación de una estrategia de carga inicial de datos legados del sistema de Banco Nacional de Cuba para el sistema SAGEB*

***Autora:*** *Madelyn Roa Vidal  
Renier Gutiérrez Prendes*

***Tutor:*** *Ing. Yanetsi Mestre Morales*

**FRASE**

*...la Informática se convertirá en una poderosísima fuerza científica, económica e incluso política del país...*

*Fidel Castro Ruz*



***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 2010

Madelyn Roa Vidal

Renier Gutiérrez Prendes

-----  
Firma del Autor

-----  
Firma del Autor

Yanetsi Mestre Morales

-----  
Firma del Tutor



## AGRADECIMIENTOS

De Madelyn:

*Agradezco a todas aquellas personas que de una forma u otra hicieron posible que mi sueño de ser ingeniera se cumpliera.*

*A toda mi familia, mis padres, mis abuelos, mis hermanos, mi novio, mis tíos, mi prima Ivett y a todas mis amistades.*

*Muy en especial a mi mejor amiga Yuseidy y a su familia.*

*A mi compañero de tesis Renier.*

*A mi tutora Yanetsi.*

*A todos muchas gracias por su apoyo incondicional.*

De Renier:

*A mis padres por los muchos sacrificios a lo largo de tantos años.*

*A mi hermano.*

*A mi novia por estar al lado mí todos estos años y darme fuerzas para continuar. A todos mis familiares por el apoyo y su preocupación constantes.*

*A mis amistades que me han apoyado y por su incondicionalidad.*

*A mi compañera de tesis Madelyn*

*A la tutora Yanetsi.*



**DEDICATORIA**

*De Madelyn:*

*A mi mamá y a mi papá.*

*A mis hermanos.*

*A mis abuelos.*

*A mi mejor amiga Yusteid.*

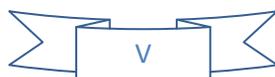
*A mi novio Yordano.*

*De Renier:*

*A mis padres, a mi hermano, a mi novia.*

## **RESUMEN**

El presente trabajo de diploma pretende implementar una estrategia para la carga de datos del sistema contable que utiliza en la actualidad el Banco Nacional de Cuba (BNC), para el nuevo sistema que está implementando el proyecto SAGEB (Sistema Automatizado para la Gestión Bancaria) en la UCI (Universidad de Ciencias Informáticas). Para ello se realizó el diseño de la estrategia a través de la metodología DWEP (Data Warehouse Engineering Process) y como herramienta de Extracción, Transformación y Carga (ETL) se hizo uso de Pentaho Data Integration (Kettle ETL) que es una de las herramientas ETL libres más usadas hoy en día, con su utilización se desarrollaron exitosamente toda las transformaciones y limpiezas necesarias, logrando migrar los datos sin pérdida de integridad de los mismos hacia una base de datos destino que presenta estructura diferente y en otro gestor de Bases de datos (SQL Server 2005).



**TABLA DE CONTENIDOS**

**FRASE** ..... I

**DECLARACIÓN DE AUTORÍA** ..... II

**AGRADECIMIENTOS**..... III

**DEDICATORIA** ..... IV

**RESUMEN**..... V

**TABLA DE CONTENIDOS** ..... VI

**ÍNDICE DE FIGURAS**..... VIII

**INTRODUCCIÓN** ..... 1

**CAPÍTULO I. FUNDAMENTACIÓN TEÓRICA** ..... 4

    1.1 Introducción..... 4

    1.2 ¿Qué es una base de datos? ..... 4

        1.2.2 Características de una base de datos ..... 5

    1.3 Migraciones de Datos ..... 5

        1.3.1 ¿Por qué la migración de Datos? ..... 6

        1.3.2 Aspectos a tener en cuenta durante la migración de los datos ..... 6

        1.3.3 Pasos para llevar a cabo una migración de datos ..... 7

        1.3.4 Resultado de la migración de datos ..... 7

    1.4 Herramientas ETL ..... 8

        1.4.1 Octopus ..... 9

        1.4.2 Talend..... 9

        1.4.3 Clovet ETL ..... 9

    1.4.4 Pentaho Data Integration (Kettle ETL)..... 10

    1.5 Metodologías y herramientas a utilizar ..... 11

        1.5.1 Rational Rose ..... 11

        1.5.2 UML (Lenguaje Unificado de Modelado) ..... 12

        1.5.3 DWEP (Data Warehouse Engineering Process)..... 12

    1.6 Conclusiones Parciales ..... 13

**CAPÍTULO II. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA** ..... 14

2.1 Introducción.....	14
2.2 Descripción de la solución .....	14
2.3 Aplicación del método DWEP .....	14
2.3.1 Mapeo de datos .....	14
2.4 Conclusiones Parciales .....	27
<b>CAPÍTULO III. DESARROLLO DE LA SOLUCIÓN .....</b>	<b>28</b>
3.1 Introducción.....	28
3.2 Funciones utilizadas de la herramienta.....	28
3.3 Desarrollo .....	29
3.3 Pruebas de integración de los datos.....	36
3.4 Resultado de la migración .....	43
3.5 Conclusiones Parciales .....	43
<b>CONCLUSIONES GENERALES .....</b>	<b>44</b>
<b>RECOMENDACIONES .....</b>	<b>45</b>
<b>REFERENCIAS BIBLIOGRÁFICAS .....</b>	<b>46</b>
<b>BIBLIOGRAFÍA .....</b>	<b>48</b>
<b>ANEXOS .....</b>	<b>50</b>
<b>GLOSARIO DE TÉRMINOS .....</b>	<b>67</b>

**ÍNDICE DE FIGURAS**

**Figura 1.** Niveles de mapeo de datos. .... 15

**Figura 2.** Mapeo de datos. Nivel 0. .... 16

**Figura 3.** Esquema conceptual de la fuente. .... 16

**Figura 4.** Esquema conceptual del destino. .... 17

**Figura 5.** Mapeo de datos. Nivel 1. .... 17

**Figura 6.** Clasificadores sin transformaciones. .... 18

**Figura 7.** Clasificadores con transformaciones. .... 18

**Figura 8.** Tablas de maestros. .... 18

**Figura 9.** Tablas de maestros críticos. .... 19

**Figura 10.** Mapeo de datos. Nivel 2. Tabla M\_Comins. .... 20

**Figura 11.** Mapeo de datos. Nivel 3. Tabla M\_Comins (Paso 1). .... 20

**Figura 12.** Mapeo de datos. Nivel 3. Tabla M\_Comins (Paso 2). .... 21

**Figura 13.** Mapeo de datos. Nivel 2. Tabla C\_Comisi y M\_Comisi. .... 26

**Figura 14.** Mapeo de datos. Nivel 3. Tabla C\_Comisi y M\_Comisi .... 27

**Figura 15.** Conexión a la Base de Datos. .... 30

**Figura 16.** Entrada Tabla. .... 31

**Figura 17.** Salida Tabla. .... 31

**Figura 18.** Filtrar campo. .... 32

**Figura 19.** Modificar valores. .... 33

**Figura 20.** Tabla resultante. .... 34

**Figura 21.** Esquema de la transformación M\_Comins. .... 34

**Figura 22.** Esquema de la transformación auxiliar C\_Cominsi. .... 35

**Figura 23.** Esquema de la transformación M\_Cominsi. .... 35

**Figura 24.** Esquema del trabajo para ejecutar las transformaciones. .... 36

**Figura 25.** Mapeo de datos. Nivel 2. Tabla C\_Client. .... 50

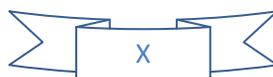
**Figura 26.** Mapeo de datos. Nivel 3. Tabla C\_Client. .... 50

**Figura 27.** Esquema de la transformación Tabla C\_Client. .... 51

**Figura 28.** Mapeo de datos. Nivel 2. Tabla C\_Bancos. .... ¡Error! Marcador no definido.

<b>Figura 29.</b> Mapeo de datos. Nivel 3. Tabla C_Bancos.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 30.</b> Esquema de la transformación Tabla C_Bancos.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 31.</b> Mapeo de datos. Nivel 2. Tabla C_Person. ....	51
<b>Figura 32.</b> Mapeo de datos. Nivel 3. Tabla C_Person. ....	52
<b>Figura 33.</b> Esquema de la transformación Tabla C_Person. ....	52
<b>Figura 34.</b> Mapeo de datos. Nivel 2. Tabla C_Paises.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 35.</b> Mapeo de datos. Nivel 3. Tabla C_Paises.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 36.</b> Esquema de la transformación Tabla C_Paises.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 37.</b> Mapeo de datos. Nivel 2. Tabla M_Cuemay.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 38.</b> Mapeo de datos. Nivel 3. Tabla M_Cuemay.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 39.</b> Esquema de la transformación Tabla M_Cuemay.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 40.</b> Mapeo de datos. Nivel 2. Tabla M_Compen.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 41.</b> Mapeo de datos. Nivel 3. Tabla M_Compen.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 42.</b> Esquema de la transformación Tabla C_Compen.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 43.</b> Mapeo de datos. Nivel 2. Tabla C_Areas.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 44.</b> Mapeo de datos. Nivel 3. Tabla C_Areas.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 45.</b> Esquema de la transformación Tabla C_Areas.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 46.</b> Mapeo de datos. Nivel 2. Tabla C_Cuenta.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 47.</b> Mapeo de datos. Nivel 3. Tabla C_Cuenta.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 48.</b> Esquema de la transformación Tabla C_Cuenta.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 49.</b> Mapeo de datos. Nivel 2. Tabla C_Asient.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 50.</b> Mapeo de datos. Nivel 3. Tabla C_Asient.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 51.</b> Esquema de la transformación Tabla C_Asient.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 52.</b> Mapeo de datos. Nivel 2. Tabla C_Concep.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 53.</b> Mapeo de datos. Nivel 3. Tabla C_Concep.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 54.</b> Esquema de la transformación Tabla C_Concep.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 55.</b> Mapeo de datos. Nivel 2. Tabla C_Tiptip.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 56.</b> Mapeo de datos. Nivel 3. Tabla C_Tiptip.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 57.</b> Esquema de la transformación Tabla C_Tiptip.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 58.</b> Mapeo de datos. Nivel 2. Tabla C_Otrcon.....	<b>¡Error! Marcador no definido.</b>

<b>Figura 59.</b> Mapeo de datos. Nivel 3. Tabla C_Otrcon.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 60.</b> Esquema de la transformación Tabla C_Otrcon.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 61.</b> Mapeo de datos. Nivel 2. Tabla C_Moneda.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 62.</b> Mapeo de datos. Nivel 3. Tabla C_Moneda.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 63.</b> Esquema de la transformación Tabla C_Moneda.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 64.</b> Mapeo de datos. Nivel 2. Tabla C_Provin.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 65.</b> Mapeo de datos. Nivel 3. Tabla C_Provin.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 66.</b> Esquema de la transformación Tabla C_Provin.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 67.</b> Mapeo de datos. Nivel 2. Tabla M_Perban.....	52
<b>Figura 68.</b> Mapeo de datos. Nivel 3. Tabla M_Perban.....	53
<b>Figura 69.</b> Esquema de la transformación Tabla M_Perban.....	53
<b>Figura 70.</b> Mapeo de datos. Nivel 2. Tabla M_Comper.....	54
<b>Figura 71.</b> Mapeo de datos. Nivel 3. Tabla M_Comper.....	54
<b>Figura 72.</b> Esquema de la transformación Tabla M_Comper.....	55
<b>Figura 73.</b> Mapeo de datos. Nivel 2. Tabla M_Percue.....	55
<b>Figura 74.</b> Mapeo de datos. Nivel 3. Tabla M_Percue (Paso1).....	56
<b>Figura 75.</b> Mapeo de datos. Nivel 3. Tabla M_Percue (Paso2).....	57
<b>Figura 76.</b> Esquema de la transformación Tabla M_Percue.....	57
<b>Figura 77.</b> Mapeo de datos. Nivel 2. Tabla M_Mayor.....	58
<b>Figura 78.</b> Mapeo de datos. Nivel 3. Tabla M_Mayor.....	59
<b>Figura 79.</b> Esquema de la transformación Tabla M_Mayor.....	60
<b>Figura 80.</b> Mapeo de datos. Nivel 2. Tabla M_Tasint.....	60
<b>Figura 81.</b> Mapeo de datos. Nivel 3. Tabla M_Tasint.....	61
<b>Figura 82.</b> Esquema de la transformación Tabla M_Tasint.....	62
<b>Figura 83.</b> Mapeo de datos. Nivel 2. Tabla M_Percli.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 84.</b> Mapeo de datos. Nivel 3. Tabla M_Percli.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 85.</b> Esquema de la transformación Tabla M_Percli.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 86.</b> Mapeo de datos. Nivel 2. Tabla M_Diario.....	63
<b>Figura 87.</b> Mapeo de datos. Nivel 3. Tabla M_Diario.....	63
<b>Figura 88.</b> Esquema de la transformación Tabla M_Diario.....	64



**Figura 89.** Mapeo de datos. Nivel 2. Tabla M\_Histor.....64  
**Figura 90.** Mapeo de datos. Nivel 3. Tabla M\_Histor.....65  
**Figura 91.** Esquema de la transformación Tabla M\_Histor.....66

## INTRODUCCIÓN

Desde el surgimiento de la informática hasta la actualidad se han desarrollado numerosos sistemas de gestión de base de datos (SGBD) con el objetivo de informatizar la sociedad y a su vez resolver los problemas existentes en cuanto a organización y almacenamiento de datos.

La cantidad de información que se maneja actualmente en las empresas provoca que su gestión y administración sea una tarea casi imposible de realizar. Con el transcurso del tiempo han surgido nuevas tecnologías y herramientas que han propiciado que los SGBD que existían hasta el momento se hayan visto obligados a migrar los datos a nuevas herramientas.

La migración de los datos consiste en mover, convertir, trasladar datos desde un SGBD a otro. La misma es un gran desafío porque involucra una multitud de problemas de conversión causados por las considerables incompatibilidades que pueden existir entre las bases de datos y la necesidad de preservar las propiedades de ejecución de las aplicaciones de base de datos existentes.

### **Situación Problemática**

El BNC, constituye una entidad en la cual la contabilidad es su actividad fundamental y para ejecutar sus funciones requiere del procesamiento de un gran volumen de datos. Además ha tenido que valerse de un sistema contable para controlar y agilizar sus operaciones.

Actualmente el BNC cuenta con una versión del sistema informático SABIC (Sistema Automatizado para la Banca Internacional de Comercio), el mismo se encuentra muy desactualizado. Su desarrollo sobre tecnología obsoleta lo ha convertido en un sistema decadente, propiciando que no pueda cubrir en su totalidad las exigencias de los procesos que se ejecutan dentro de esta entidad, el mismo actualmente trabaja con el sistema MS-DOS y los datos son almacenados en tablas en FoxPro, el cual no funciona adecuadamente con grandes volúmenes de información ya que no posee una gran capacidad de almacenamiento.

El BNC al igual que el resto de las entidades financieras del país, pretende usar una versión más ampliada

del SABIC, es decir, tendrá todas funcionalidades con la adición de otras, ya que el BNC tiene características diferentes al resto de los demás bancos existentes en el país.

El proyecto SAGEB de la UCI, se encuentra enfrascado en la realización de un software partiendo inicialmente del sistema SABIC. Una vez concluido el mismo, se deben migrar los datos de forma completa y consistente del sistema en FoxPro hacia el nuevo sistema, sin perder la información recopilada que tiene almacenado en dicha migración.

Dada la situación problemática antes vista se plantea el siguiente **problema a resolver**: ¿Cómo migrar los datos del sistema del BNC para la base de datos del sistema SAGEB?, teniendo como **objeto de estudio**: los mecanismos de transformación de datos y el **campo de acción** está centrado en el proceso de transformación de datos legados del sistema del BNC.

El **objetivo general** del presente trabajo se enmarca en: Diseñar e implementar una estrategia para llevar a cabo el proceso de migración de datos del sistema del BNC para la base de datos del sistema SAGEB. Para cumplir con el mismo se han propuesto las siguientes **tareas de investigación**:

- ✓ Efectuar un estudio sobre las herramientas, tecnologías, metodologías y tendencias actuales propuestas para la realización de transformaciones y migraciones de datos, con el objetivo de identificar y seleccionar las posibles a utilizar en la solución del problema planteado.
- ✓ Obtención de los artefactos que se generan en los flujos de trabajo: Análisis, Diseño e Implementación, para lograr una mejor comprensión de la estrategia a desarrollar.
- ✓ Realizar el proceso de extracción, transformación y carga de datos para filtrar, limpiar, homogenizar y agrupar la información proveniente de la fuente de datos.
- ✓ Realizar pruebas de integración de datos para comprobar la integridad y consistencia de los mismos luego de realizada la migración.

Se plantea como **resultado esperado**: la migración de los datos del sistema del BNC para la base de datos del sistema SAGEB.

## Estructura del contenido

El presente documento consta de 3 capítulos, en los cuales se desarrollan aspectos de importancia para lograr el objetivo que se persigue.

### Capítulo 1: Fundamentación teórica

- ✓ Conceptos fundamentales para el esclarecimiento del trabajo
- ✓ Tendencias actuales en el desarrollo de los sistemas de migración de datos.
- ✓ Técnicas para las transformaciones de datos.
- ✓ Herramientas utilizadas para la migración de datos.
- ✓ Metodologías y herramientas para llevar a cabo la propuesta de solución.

### Capítulo 2. Descripción de la solución propuesta

- ✓ Descripción de la propuesta.
- ✓ Aplicación de método DWEP.
- ✓ Diseño de la solución mediante los niveles del mapeo de datos.

### Capítulo 3. Desarrollo de la solución

- ✓ Desarrollo de la solución.
- ✓ Resultado de la migración.

## *CAPÍTULO I. FUNDAMENTACIÓN TEÓRICA*

### **1.1 Introducción**

En este capítulo se exponen una serie de conceptos fundamentales para el esclarecimiento del trabajo. Se hace un análisis de las herramientas que existen para la migración de datos, escogiendo la más adecuada. Además se exponen las metodologías y las herramientas que se utilizarán para llevar a cabo la solución.

### **1.2 ¿Qué es una base de datos?**

“El término de base de datos fue escuchado por primera vez en 1963, en un simposio celebrado en California, USA.” (2008). “Una base de datos se define como un fichero en el cual se almacena información en campos o delimitadores, teniendo acceso a ella posteriormente tanto de forma separada como de forma conjunta. Se utiliza normalmente para recoger grandes cantidades de información.” (Pérez, 2009).

“Los datos se recopilan, se estructuran, se centralizan y se almacenan convenientemente. El objetivo inmediato de este proceso es poder recuperar estos mismos datos u otros datos derivados de ellos en cualquier momento sin necesidad de volverlos a recopilar.”

“Desde el punto de vista informático, la base de datos es un sistema formado por un conjunto de datos almacenados en discos que permiten el acceso directo a ellos y un grupo de programas que manipulen esos datos.” (2008).

“Cada base de datos se compone de una o más tablas que guarda un conjunto de datos. Cada tabla tiene una o más columnas y filas. Las columnas guardan una parte de la información sobre cada elemento que queramos guardar en la tabla, cada fila de la tabla conforma un registro.” (Valdes, 2007)

Las bases de datos se han constituido como una de las herramientas más ampliamente difundidas en la actual sociedad de la información, utilizadas como fuentes secundarias en cuanto recuperación y almacenamiento de información en todos los campos a nivel científico, social, económico, político y cultural.

## **1.2.1 Importancia de una base de datos**

Las bases de datos, son uno de los recursos más importantes en el desarrollo de cualquier área del saber, el contexto histórico ha demostrado que son de una gran influencia en el progreso de los países a nivel mundial, permitido un almacenamiento y acceso confiable, eficiente y práctico en el uso de la información que se produce, éstas facilitan:

- ✓ El almacenamiento de grandes cantidades de información.
- ✓ La recuperación rápida y flexible de información.
- ✓ La organización y reorganización de la información.
- ✓ La impresión y distribución de información en varias formas.

## **1.2.2 Características de una base de datos**

“Entre las principales características de los sistemas de base de datos podemos mencionar:

- ✓ Independencia lógica y física de los datos.
- ✓ Redundancia mínima.
- ✓ Acceso concurrente por parte de múltiples usuarios.
- ✓ Integridad de los datos.
- ✓ Consultas complejas optimizadas.
- ✓ Seguridad de acceso y auditoria.
- ✓ Respaldo y recuperación.
- ✓ Acceso a través de lenguajes de programación estándar.” (Valdes, 2007)

## **1.3 Migraciones de Datos**

“La migración de datos es el proceso por el cual grandes volúmenes de datos son trasladados desde sistemas existentes hacia sistemas nuevos. Los cambios tecnológicos, los cambios de proveedores, las actualizaciones del software y los proyectos de almacenamiento de datos hacen que operaciones tan críticas y delicadas sean necesaria.”(2010)

En la actualidad, la migración de datos es una actividad regular del área de informática en las empresas.

“Abarcando además todos los pasos necesarios para limpiar, corregir y mover los datos al nuevo sistema.” (2010).

## 1.3.1 ¿Por qué la migración de Datos?

La necesidad de crear sistemas encaminados a garantizar el funcionamiento de la economía cubana en las nuevas circunstancias y en el marco de las transformaciones organizativas que se están desarrollando en la actualidad, constituye un reto. La informatización del BNC como parte del avance tecnológico que precisa nuestro país, demanda una elevada capacidad tecnológica y operativa, lo que trae consigo a gran escala la utilización de modernos medios de procesamiento de la información.

“Muchas circunstancias existen cuando una organización necesita migrar las aplicaciones o las bases de datos. Puede ser que sea tan simple como una mejora a una nueva versión del sistema. O puede ser que implique cambiar a una nueva base de datos o aplicación. Después de una fusión o de una adquisición, a menudo se retiran las aplicaciones redundantes, pero los datos tienen que ser preservados en el sistema de supervivencia. Después de una migración, el viejo sistema y el nuevo sistema tendrán que coexistir.” (2010).

## 1.3.2 Aspectos a tener en cuenta durante la migración de los datos

- ✓ Planeación: lo más importante al migrar una base de datos es llevar a cabo un proceso de planeación y análisis del trabajo, puesto que aunque pareciera tomarse algún tiempo adicional, éste será retribuido en el éxito de la operación y menos costos por errores de datos. Es importante que esto sea aplicado cuando la base de datos destino está en producción.
- ✓ Mapeador de Tipos de datos: algunas plataformas no soportan algunos tipos de datos, así que es necesario planificar el mapeo de los campos en la nueva base de datos.
- ✓ Restricciones y Triggers: antes de iniciar la migración de la base de datos, es recomendable deshabilitar los triggers y/o restricciones que puedan generar error al momento que el DBMS (Data Base Management System) ejecute el proceso de escritura de los datos.

- ✓ Codificación de Caracteres: cuando el copiado se realiza de forma automática, es necesario identificar la codificación de caracteres que la base de datos destino espera, pues así se evita el reemplazo automático de caracteres o pérdida de los mismos.

### 1.3.3 Pasos para llevar a cabo una migración de datos

- ✓ Limpieza de las tablas de la base de datos.
- ✓ Consolidación de las tablas de la base de datos.
- ✓ Mapeado de las tablas de Origen y Destino.
- ✓ Definición de Formato de Origen y Destino.
- ✓ Ruta de Saltos de Formato en la Migración.
- ✓ Definición del Set de Caracteres de Origen y Destino.
- ✓ Comprobación de los delimitadores de campos.
- ✓ Migración de Prueba.
- ✓ Evaluación y Comprobación de Errores.
- ✓ Depuración Final.

### 1.3.4 Resultado de la migración de datos

Las migraciones traen como resultado el mejor aprovechamiento de la tecnología que se posea y la mayor velocidad de acceso a la información. En muchos casos y dependiendo del tipo de migración, también se logra mayor seguridad y versatilidad.

Casos en los que es necesaria la migración de datos:

- ✓ Se ha vuelto muy lenta la carga y consulta de datos.
- ✓ Se debe mantener compatibilidad con otros productos.
- ✓ El volumen de información que se maneja necesita una mejor base de datos.
- ✓ Se desea mantener actualizados los productos.
- ✓ Es necesario cambiar hacia un producto de base de datos sin licencia.
- ✓ Se necesita bajar los gastos en tecnología.
- ✓ Se desea armar una red.
- ✓ Se han comprado nuevos equipos.

## 1.4 Herramientas ETL

“Cuando se implanta un sistema, normalmente se dispone de otro, por lo que es necesario hacer extracción de datos, transformación y carga de los mismos, esto es lo que se denomina: ETL. (Extract, Transform, Load)” (Mora, 2008). Un proceso ETL organiza el flujo de datos entre diferentes fuentes o sistemas y aporta los métodos y herramientas necesarias para mover datos hacia un formato único.

Entre las más populares herramientas ETL del mercado pueden ser citadas algunas como:

- ✓ Kettle.
- ✓ Talend Open Studio.
- ✓ Clover ETL.
- ✓ Octopus.

El proceso de **extracción**: La primera parte del proceso ETL consiste en extraer los datos desde los sistemas de origen. La extracción convierte los datos a un formato preparado para iniciar el proceso de transformación. Una parte intrínseca del proceso de extracción es la de analizar los datos extraídos, de lo que resulta un chequeo que verifica si los datos cumplen la pauta o estructura que se esperaba. De no ser así los datos son rechazados.

El proceso de **transformación**: Se aplican una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados. Algunas fuentes de datos requerirán alguna pequeña manipulación como la aplicación de cualquier forma, simple o compleja, de validación de datos, y la consiguiente aplicación de la acción que en cada caso se requiera.

La **carga** de los datos en la fuente destino: Este es el momento en el cual los datos de la fase anterior (*transformación*) son cargados en el sistema destino. Este proceso abarca una amplia variedad de acciones diferentes tales como mantener un historial de los registros de manera que se pueda hacer una auditoría de los mismos y disponer de un rastro de toda la historia de un valor a lo largo del tiempo almacenados en *logs*.

Beneficios principales de las herramientas ETL:

- ✓ Ganancias en términos de tiempo.
- ✓ Procesos automatizados.
- ✓ Total fiabilidad de los datos. (Collado)

## 1.4.1 Octopus

Octopus es una herramienta ETL para transformaciones de datos por conexión JDBC (Conectividad de la Base de Datos en Java). Octopus solo soporta fuentes de datos que vengan con el manejador JDBC, incluye también drivers especiales que permiten la conectividad con archivos CSV, XML, MS-SQL. La principal característica de esta herramienta es que, para acceder a las fuentes de datos que lo utilicen, debe existir un manejador JDBC disponible. Es una herramienta poderosa capaz de normalizar datos, crear llaves artificiales, tablas y llaves primarias. Todos los trabajos que corren en Octopus son independientes de los proveedores de base de datos.

## 1.4.2 Talend

Talend es una herramienta ETL que entre sus principales características se puede decir que es compatible con Microsoft SQL Server Integration Services (SSIS), funciona en Windows, Unix y Linux. Se puede conectar con: Oracle, DB2, MySQL, Sybase y Postgres. El código fuente Java/Eclipse está disponible para su descarga y personalización. La interfaz GUI para importar metadatos, configuraciones y uniones de componentes y generación de código proporciona ganancias en productividad para desarrolladores y acaba siendo más rápida que la programación estándar. Esta herramienta necesita de Java Data Base Connectivity (API que permite la realización de operaciones sobre bases de datos) para acceder a las fuentes. No tiene proceso automático de separación y redistribución de datos lo cual puede generar cuellos de botella. Es una herramienta poco intuitiva y difícil de entender, pero una vez superada esta dificultad inicial se observan las grandísimas posibilidades y la potencia de la aplicación.

## 1.4.3 Clovet ETL

Es un ambiente de transformación de datos de código abierto basado en Java, para datos estructurados, capaz de funcionar como aplicación independiente o estar incluida en otra aplicación. Maneja todas las bases de datos con conexión JDBC disponible para ellas, soporta valores nulos y 4 diferentes tipos de

datos: string, numeric, bytes y date. Realiza transformaciones gráficas basadas en XML para la descripción de los metadatos de los registros y puede ejecutarse en múltiples computadoras.

## 1.4.4 Pentaho Data Integration (Kettle ETL)

“La plataforma Open Source Pentaho Business Intelligence cubre amplias necesidades de Análisis de los Datos y de los Informes empresariales. Las soluciones de Pentaho están escritas en Java y tienen un ambiente de implementación también basado en Java. Eso hace que Pentaho es una solución muy flexible para cubrir una amplia gama de necesidades empresariales.” (2008).

Incluye herramientas para realizar consultas, generación de informes y reportes, análisis interactivo, tableros de mando, ETL/integración de datos, data mining (minería de datos), y un servidor para la plataforma de BI (Business Intelligence) que lo ha convertido en la suite BI de software libre más popular del mundo. Productos de Pentaho se utilizan en el Sistema de Mando Aéreo de US Army, Lifetime Networks, Terra Industries y Sun Microsystems.

Kettle es una herramienta de (Pentaho Data Integración) que permite implementar los procesos ETL, la misma es de código abierto compuesta por cuatro componentes fundamentales: SPOON para el diseño gráfico de las transformaciones, PAN para la ejecución de los trabajos y las transformaciones, CHEF para el diseño de la carga de datos y KITCHEN para la ejecución de los trabajos Batch diseñados con CHEF.

Características de Kettle:

- ✓ Plataforma: Windows, Unix y Linux.
- ✓ GUI: Interfaz con indicadores visuales de transformación. Informes disponibles de la capa de metadatos.
- ✓ Código: Aplicación 100% Java con transformaciones avanzadas en JavaScript mediante una interfaz empotrada. Diseño orientado a metadatos.
- ✓ Licencia: Mozilla Public License.
- ✓ Código fuente: El código fuente está disponible.
- ✓ Soporte: Existe un foro, un buscador de problemas y la comunidad Pentaho, con varios artículos técnicos que son mejores que algunos de los vendedores oficiales de productos para ETL.

# CAPÍTULO I. FUNDAMENTACIÓN TEÓRICA

---

- ✓ Conectividad: Soporta Oracle, DB2, SQL Server.
- ✓ Es una de las herramientas ETL libres más antiguas que tiene una gran cantidad de usuarios y una nueva dirección por parte del soporte técnico de Pentaho.

Luego de haber realizado un análisis de las herramientas ETL más usadas actualmente en el mercado, se pudo constatar que existen muchas posibilidades, pudo haberse seleccionado cualquiera de las mencionadas anteriormente pues todas son de excelente calidad. Sin embargo se escogió para la realización de las ETL del presente trabajo la herramienta Pentaho Data Integration (Kettle ETL), ya que la misma es una herramienta estable, a la cual se le han corregido sus errores, es capaz de realizar el proceso ETL eficientemente, tiene una amplia documentación, permite leer y escribir de cualquier bases de datos, operar eficientemente con los campos: renombrando, normalizando, calculando campos en función de otros, mapeando valores, realizando búsquedas auxiliares en bases de datos, etc. ,sin necesitar grandes recursos de hardware, ni un potente sistema operativo, además que la misma es la más usada en la UCI para la realización de las ETL en los almacenes de datos.

## 1.5 Metodologías y herramientas a utilizar

En cualquier ámbito de ingeniería hay una fractura entre los responsables de analizar y definir los problemas (necesidades), y los expertos en proveer soluciones (tecnología). Las metodologías nacen para intentar solucionar este conflicto. Su propósito es establecer un contrato social entre todos los participantes en un proyecto para conseguir la solución más eficaz con los recursos disponibles. Es un conjunto de procedimientos, técnicas, herramientas y un soporte documental que ayuda a los desarrolladores a producir nuevo software.

### 1.5.1 Rational Rose

Rational Rose, es una herramienta para realizar el modelado visual, forma parte de un conjunto más amplio de herramientas que todas juntas abarcan el ciclo de vida del desarrollo de un software. Permite completar diferentes disciplinas (flujos fundamentales) de RUP (Proceso Unificado de Desarrollo de Software); incluye, a su vez, herramientas de ingeniería inversa y generación de código que facilitan el tránsito hasta el producto final. Rational Rose, es considerado una de las mejores herramientas para traducir requisitos de alto nivel a una arquitectura basada en componentes. Se encuentra en la avanzada

en cuanto al desarrollo con UML por lo que se ha convertido en una de las mejores opciones por la notación estándar que provee para especificar, visualizar y construir productos de software.

## 1.5.2 UML (Lenguaje Unificado de Modelado)

“UML es un conjunto de herramientas, que permite modelar (analizar y diseñar) sistemas orientados a objetos.” (Gracia, 2005). Es el lenguaje de modelado de sistemas de software más conocido y utilizado en la actualidad. Ofrece un estándar para describir un "plano" del sistema (modelo), incluyendo aspectos conceptuales tales como procesos de negocio y funciones del sistema, y aspectos concretos como expresiones de lenguajes de programación, esquemas de bases de datos y componentes reutilizables.

Uno de los objetivos principales de la creación de UML fue posibilitar el intercambio de modelos entre las distintas herramientas CASE orientadas a objetos. Hay que tener en cuenta que el estándar UML no define un proceso de desarrollo específico, tan solo se trata de una notación. Es importante resaltar que UML es un "lenguaje" para especificar y no para describir métodos o procesos. Se utiliza para definir un sistema, para detallar los artefactos en el sistema y para documentar y construir. En otras palabras, es el lenguaje en el que está descrito el modelo. Se puede aplicar en el desarrollo de software entregando gran variedad de formas para dar soporte a una metodología de desarrollo de software pero no especifica en sí mismo qué metodología o proceso usar.

**UML es un método formal de modelado que aporta las siguientes ventajas:**

- ✓ Permite realizar una verificación y validación del modelo realizado.
- ✓ Se pueden automatizar determinados procesos y permite generar código a partir de los modelos y a la inversa.
- ✓ Mayor rigor en la especificación.

## 1.5.3 DWEP (Data Warehouse Engineering Process)

Este método propuesto en la tesis de doctorado de Sergio Luján Mora (TRUJILLO *et al.* 2002) fue creado para llevar a cabo el diseño de almacenes de datos, pues a pesar de existir una gran variedad de modelos para realizar esta tarea, ninguno permitía la representación de todas las etapas de diseño de los Data

# *CAPÍTULO I. FUNDAMENTACIÓN TEÓRICA*

---

Warehouse, como por ejemplo las transformaciones de datos. Este método está basado en UML y RUP, es orientado a objetos, independiente de cualquier implementación específica, ya sea relacional, multidimensional, etc. El mismo, a pesar de ser creado para diseñar almacenes de datos es el seleccionado para el diseño de las ETL de este trabajo, puesto que haciendo uso de él se podrán obtener diagramas bien detallados de las transformaciones de datos mostrando el nivel de detalle adecuado para su posterior implementación.

DWEP propone la estructuración del almacén de datos tres niveles:

- ✓ Conceptual: Define el almacén de datos desde un punto de vista conceptual, es decir, desde el mayor nivel de abstracción y contiene únicamente los objetos y relaciones más importantes.
- ✓ Lógico: Abarca aspectos lógicos del diseño del almacén de datos, como la definición de las tablas y claves, la definición de los procesos ETL, etc.
- ✓ Físico: Define los aspectos físicos del almacén de datos, como el almacenamiento de las estructuras lógicas en diferentes discos o la configuración de los servidores de bases de datos que mantienen el almacén de datos.

Este trabajo se centrará fundamentalmente en el nivel Lógico, específicamente en la definición de los procesos ETL.

## **1.6 Conclusiones Parciales**

En este capítulo se expusieron conceptos importantes para el entendimiento del trabajo, se presentaron algunas de las tecnologías, metodologías y herramientas existentes que se utilizarán en el diseño e implementación de la estrategia de migración de datos de FoxPro a SQL Server 2005.

## *CAPÍTULO II. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA*

### **2.1 Introducción**

En este capítulo se realizará una descripción detallada de cómo se realizará la migración de los datos utilizando la metodología DWEP, reflejada a través de varios diagramas.

### **2.2 Descripción de la solución**

Se desea diseñar e implementar una estrategia de carga inicial de los datos legados del sistema del BNC para el sistema SAGEB, la misma permitirá una rápida carga de datos y garantizará la consistencia de los mismos, realizando las limpiezas y transformaciones de datos necesarias ya que la base de datos origen y destino poseen estructuras diferentes. Para el diseño de la misma se utilizará el método DWEP.

### **2.3 Aplicación del método DWEP**

DWEP es un método global para llevar a cabo el diseño de todas las fases de los almacenes de datos, incluyendo las fuentes de datos operacionales, los procesos ETL y el propio esquema del almacén de datos. El mismo es un método muy útil para el diseño de las transformaciones de datos que se desea realizar, ya que describe de forma bien específica todo este proceso necesario para alcanzar el objetivo propuesto y propone la realización de diagramas para la especificación de los mapeos de datos en varios niveles de detalle.

Para determinar las tablas origen del BNC involucradas en la migración, se realizaron entrevistas con la especialista del mismo que atiende al proyecto SAGEB, determinándose además, las dependencias y el grado de prioridad de cada tabla, las cuales son especificadas más adelante.

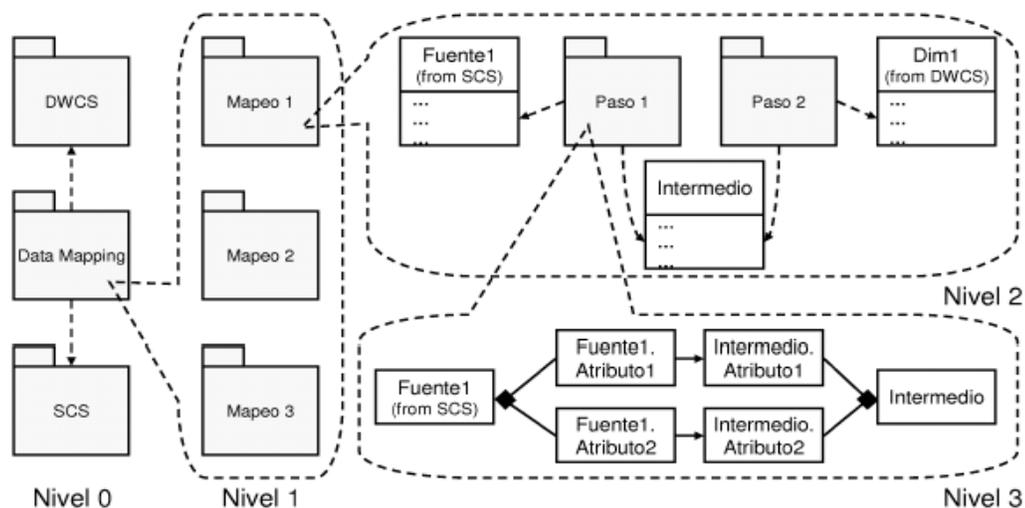
#### **2.3.1 Mapeo de datos**

El mapeo se utiliza para capturar las interconexiones entre los distintos elementos del diseño y se define mediante 3 elementos lógicos.

## CAPÍTULO II. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

- ✓ El proveedor: una entidad (esquema, tabla o atributo) responsable de generar los datos que posteriormente se propagan.
- ✓ El consumidor: que recibe los datos del proveedor.
- ✓ El emparejamiento: que define la forma en la cual el mapeo se realiza, incluyendo cualquier tipo de transformación o filtrado.

Debido a que este diagrama puede ser muy complejo se organiza en diferentes niveles gracias al uso de paquetes UML obteniéndose cuatro niveles como se muestra en la siguiente figura:



**Figura 1. Niveles de mapeo de datos.**

A continuación se pueden consultar los niveles del mapeo de datos, comenzando por el nivel 0, representado por un paquete llamado Data Mapping, donde se encuentran todas las operaciones de mapeos relacionado con el esquema conceptual de la fuente (ECF), este diagrama se puede ver en la figura 3 que representan las tablas que almacenan los datos del BNC y el esquema conceptual del destino (ECD) que representan las tablas destino del sistema SABIC, el cual se puede observar en la figura 4. En

# CAPÍTULO II. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

este nivel cada esquema se representa mediante un paquete. Los mapeos entre los diferentes esquemas se modelan en un único paquete de mapeo, que encapsula todos los detalles.

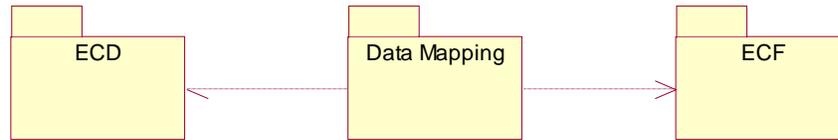


Figura 2. Mapeo de datos. Nivel 0.

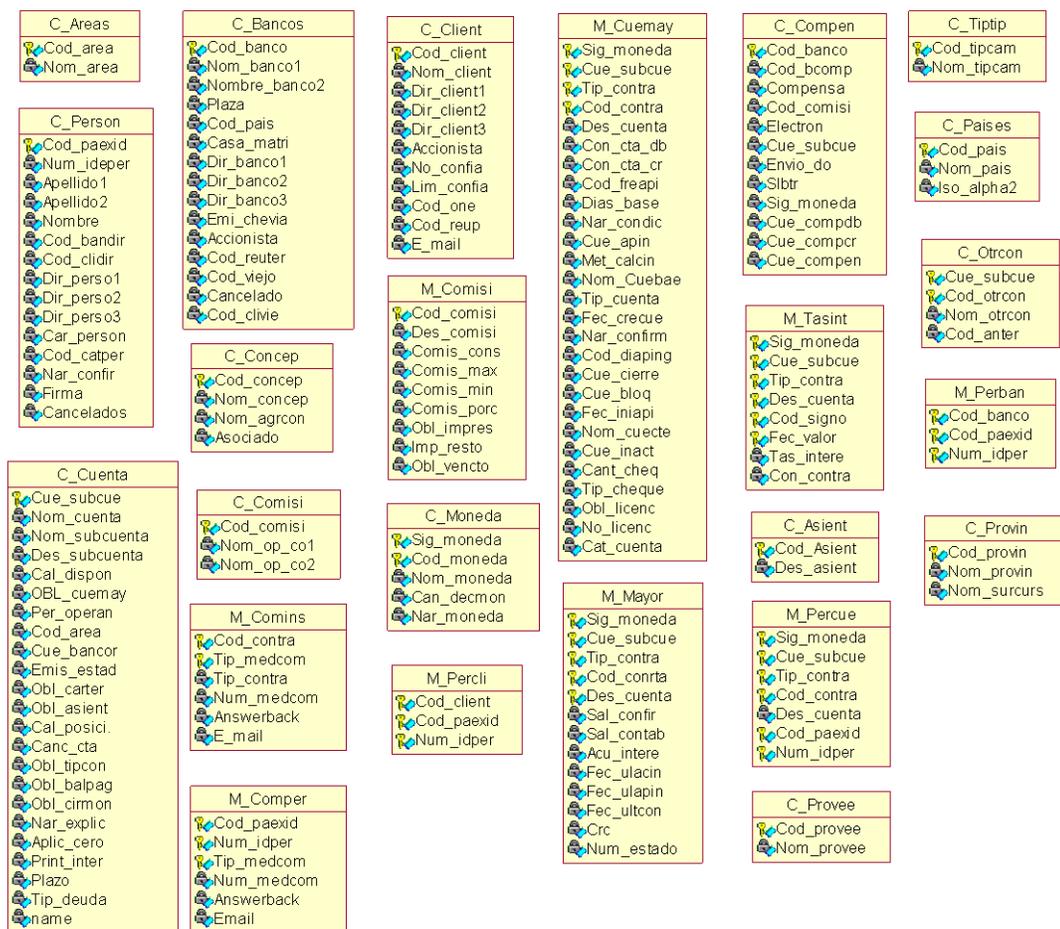


Figura 3. Esquema conceptual de la fuente.

# CAPÍTULO II. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

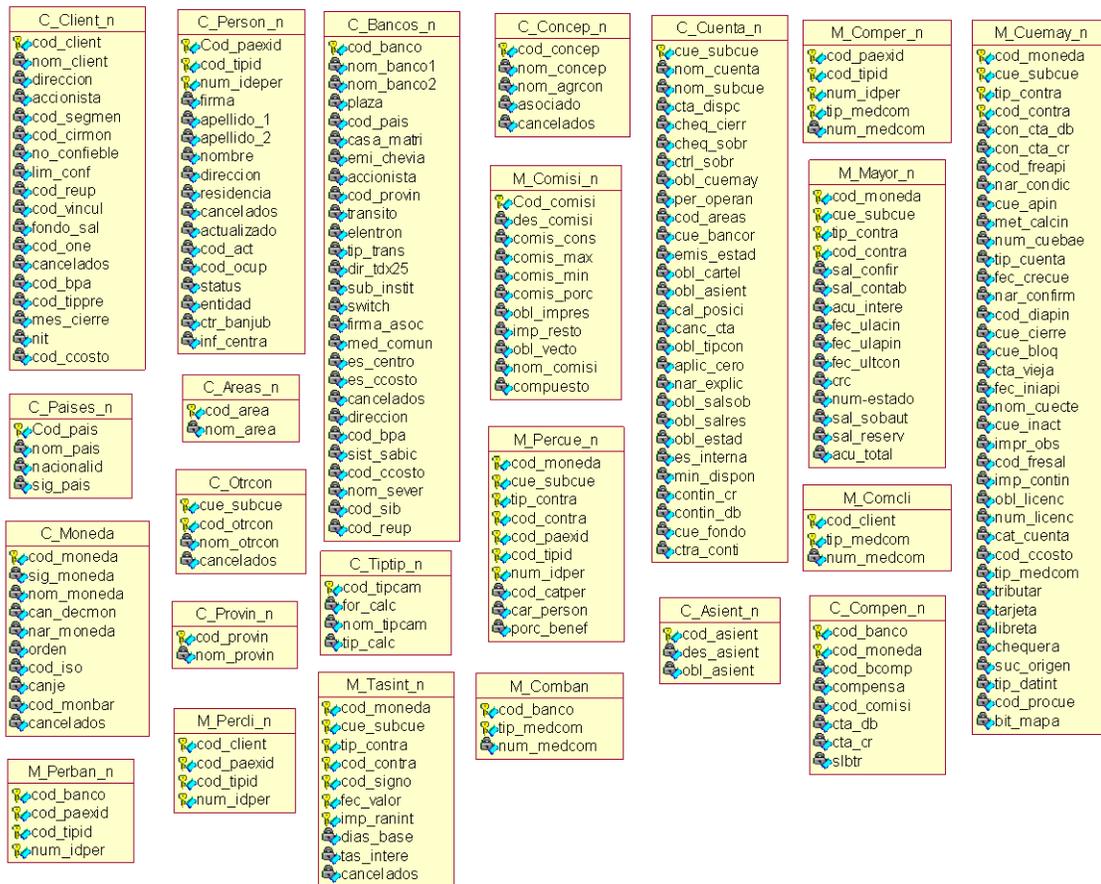


Figura 4. Esquema conceptual del destino.

El nivel 1, describe las relaciones de datos a nivel individual de las tablas fuentes hacia los respectivos destinos. Las mismas se agruparon según el grado de prioridad y dependencias entre ellas, para evitar problemas a la hora de realizar el traspaso de los datos, por ejemplo: una tabla que dependa de otra no puede migrarse antes de la cual depende.

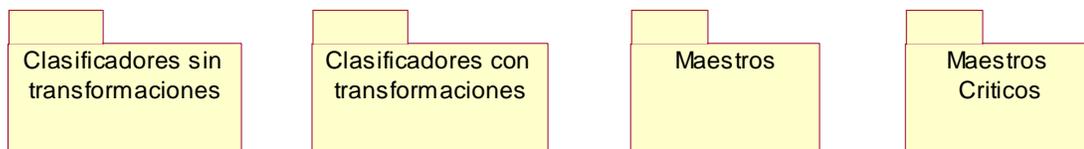
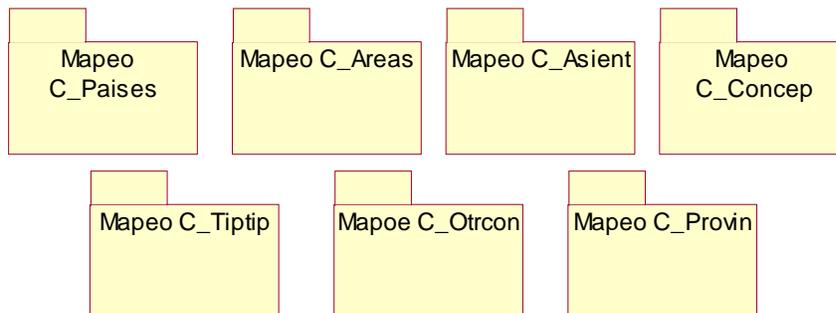


Figura 5. Mapeo de datos. Nivel 1.

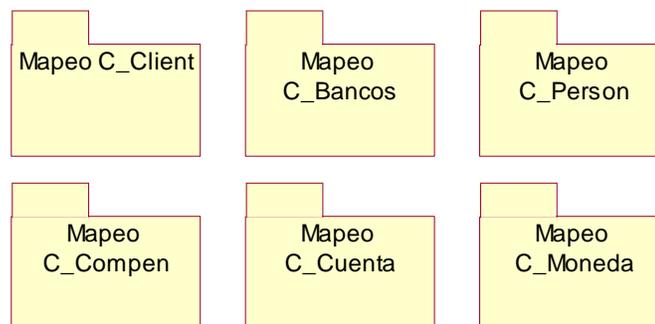
# CAPÍTULO II. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

---

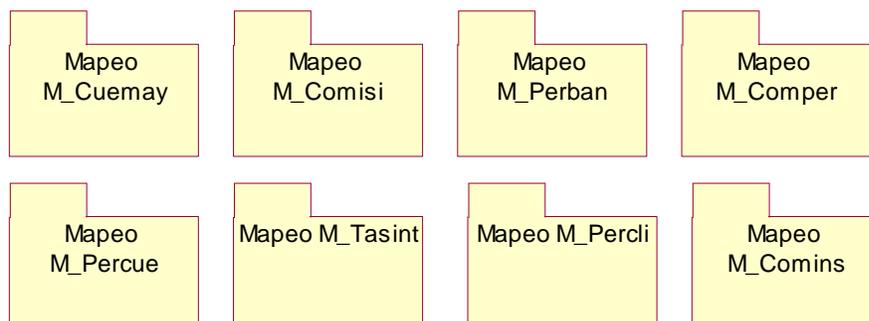
En los mismos según su descripción se encuentran los mapeos correspondientes.



**Figura 6. Clasificadores sin transformaciones.**



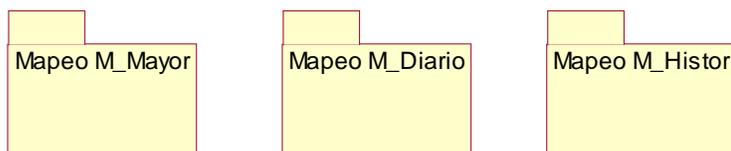
**Figura 7. Clasificadores con transformaciones.**



**Figura 8. Tablas de maestros.**

## CAPÍTULO II. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

---



**Figura 9. Tablas de maestros críticos.**

Mientras que el diagrama de mapeo en el nivel 1 describe las relaciones entre las fuentes y los destinos de datos mediante un único paquete, el diagrama de mapeo de datos en el nivel 2 detalla todas las transformaciones intermedias que tienen lugar durante ese flujo.

Los diagramas de mapeo de datos correspondientes al nivel 2 y al nivel 3 que a continuación se presentan, corresponden a las transformaciones de datos que se le realizarán a la tabla M\_Comins (Comunicación con las instituciones), en este caso la tabla M\_Comins se trasforma en dos tablas: la tabla M\_Comcli (Comunicación con los clientes) y M\_Comban (Comunicación con los bancos) lo que depende del valor del atributo tip\_contra, si es 1 entonces se mapean los datos hacia la tabla M\_Comcli de lo contrario pasan a la tabla M\_Comban. M\_Comcli está constituida por los atributos cod\_client, tip\_medcom, y el atributo num\_medcon, el cual surge de la unión de los atributos E\_mail y Num\_medcom y el cod\_client es el cod\_contra pero se le agrega un cero delante. La tabla M\_Comban, esta va a tener los mismos atributos que la tabla M\_Comcli, solo que el cod\_contra que pasa a ser el cod\_banco pasa como mismo está en la tabla M\_Comins .

# CAPÍTULO II. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

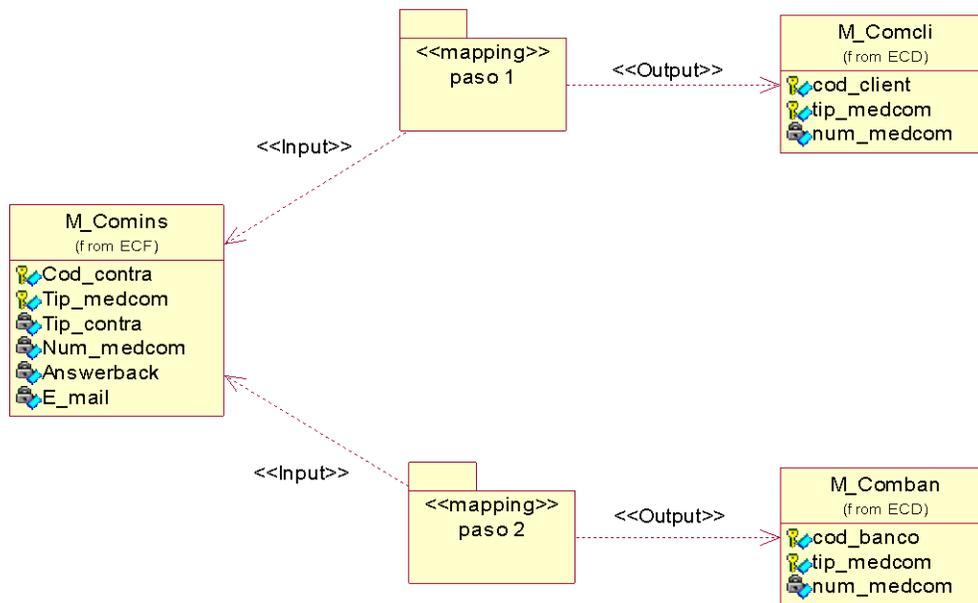


Figura 10. Mapeo de datos. Nivel 2. Tabla M\_Comins.

En el nivel 3 de diagrama de mapeo de datos se capturan los mapeos existentes a nivel de atributo.

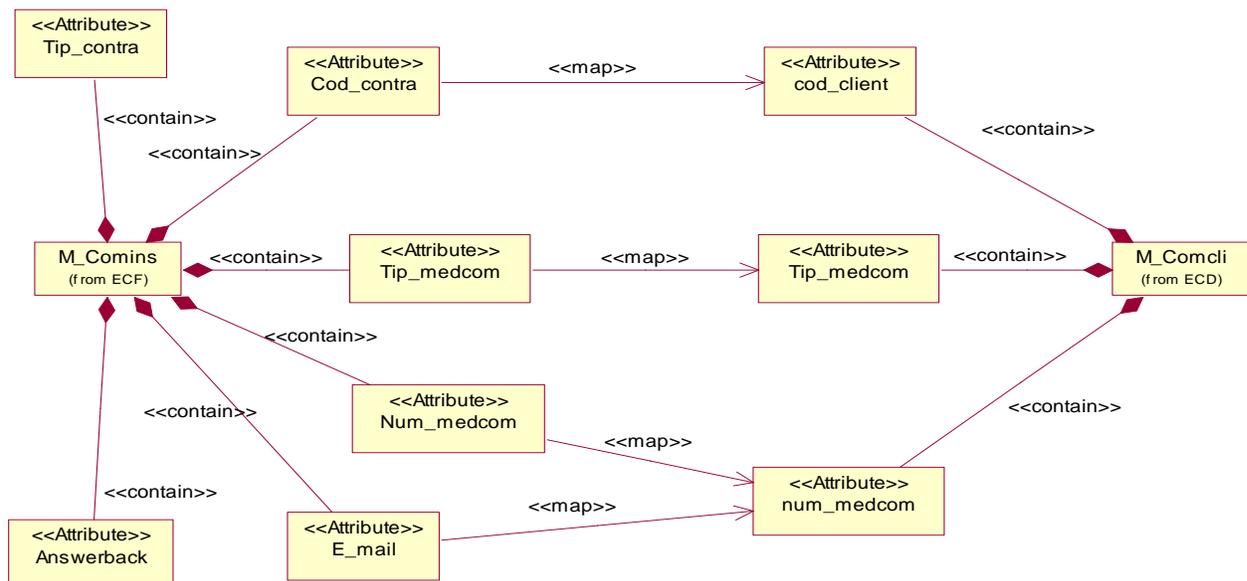


Figura 11. Mapeo de datos. Nivel 3. Tabla M\_Comins (Paso 1).

## CAPÍTULO II. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

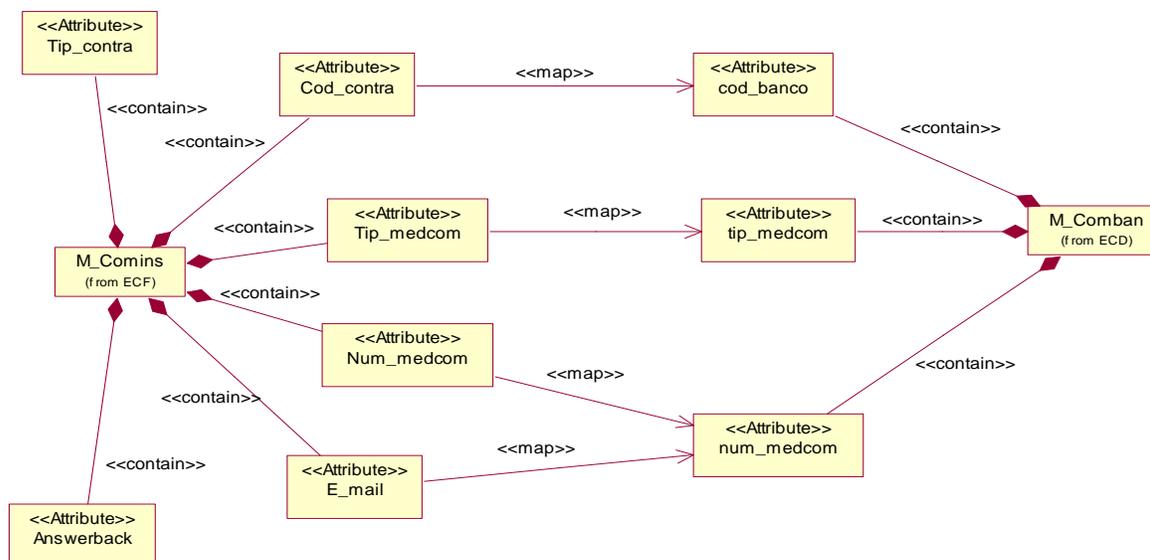


Figura 12. Mapeo de datos. Nivel 3. Tabla M\_Comins (Paso 2).

A continuación se presentan los diagramas correspondientes a las tablas M\_Comisi y C\_Comisi, estas son transformadas a una tabla nueva M\_Comisi\_n, la cual va a estar constituida por los atributos mapeados de estas dos tablas. Figuras 13 y 14.

Los diagramas de las demás tablas podrán encontrarse en los anexos.

La **tabla C\_Client** se transforma en la tabla C\_Client\_n, al realizar las transformaciones, los atributos Dir\_client1, Dir\_client2 y Dir\_client3 son transformados a un único atributo dirección, el atributo cod\_client es cargado a la tabla C\_Client\_n con la adición de un “0” al principio. El atributo E\_mail no es cargado a la nueva tabla C\_Client\_n. Los atributos cod\_segmen, cod\_cirmon, cod\_vincul, fondon\_sal, cancelados, cod\_bpa, cod\_tippre, mes\_cierre, nit y cod\_ccosto son adicionados en la tabla nueva C\_Client\_n. El atributo cod\_ccosto de la nueva tabla es una constante (001). Ver anexo 1.

La tabla **C\_Bancos** se transforma en la tabla C\_Bancos\_n, al realizar la transformación, los atributos Dir\_banco1, Dir\_banco2 y Dir\_banco3 son transformados a un único atributo dirección, los atributos

## *CAPÍTULO II. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA*

---

Cod\_reuter, Cod\_viejo y Cod\_clivie no son cargados a la nueva tabla C\_Bancos\_n. Los atributos cod\_provin, transito, cod\_ccosto, elatron, tip\_trans, sub\_instit, switch, dir\_tdx25, firma\_asoc, med\_comun, es\_centro, es\_ccosto, cod\_bpa, sist\_sabic, cod\_reup, cod\_sib y nom\_server, son adicionados en la tabla nueva C\_Bancos\_n. Ver anexo 2.

La tabla **C\_Person** se transforma en la tabla C\_Person\_n, al realizar la transformación, los atributos Dir\_perso1, Dir\_perso2 y Dir\_perso3 son transformados a un único atributo dirección, los atributos Cod\_bandir, Cod\_catper, Cod\_clidir, Car\_person y Nar\_confir, no son cargados la tabla C\_Person\_n. Los atributos: residencia, actualizados, cod\_act, cod\_ocup, status, entidad, ctr\_banjub e inf\_centra son adicionados en la tabla nueva C\_Person\_n y el cod\_tipid es una constante ("CI"). Ver anexo 3.

La tabla **C\_Paises** se transforma en la tabla C\_Paises\_n, en este caso no ocurre ninguna transformación, solamente el atributo nacionalidad es adicionado en la nueva tabla C\_Paises\_n y Iso\_alpha 2 cambia su nombre por sig\_pais. Ver anexo 4.

La tabla **M\_Cuemay** se transforma en la M\_Cuemay\_n. En la nueva tabla surge el atributo cod\_moneda que no es más que el código de la moneda al que corresponde la sigla de la moneda (sig\_moneda) de la tabla c\_moneda (clasificador moneda). El atributo cod\_ccosto es una constante (001) y el cod\_contra va a depender de 3 atributos: tip\_contra, cod\_contra y des\_cuenta. Si el tip\_contra es 1 se pone se rellena con 1 cero a la izquierda y se le concatena el des\_cuenta. Si el tip\_contra es 2 se concatena el cod\_contra con el des\_cuenta relleno con un cero a la izquierda. Los demás atributos son mapeados con la nueva tabla y pasados sin ninguna transformación (los que tienen el mismo nombre en la vieja y en la nueva los demás se desechan). Ver anexo 5.

La tabla **C\_Compen** se transforma en la tabla C\_Compen\_n, el atributo cod\_moneda de la nueva tabla es el código de la moneda correspondiente a la sigla de la moneda (sig\_moneda) de la tabla c\_moneda (clasificador moneda), y Cue\_comp\_db y Cue\_comp\_cr cambian su nombre por cta\_db y cta\_cr respectivamente. Los atributos cod\_banco, cod\_bcomp; compensa, slbtr y cod\_comisi son cargados sin ninguna transformación y electrón, cue\_subcue, envio\_do, y cue\_compen son desechados. Ver anexo 6.

## *CAPÍTULO II. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA*

---

La tabla **C\_Areas** se transforma en la tabla C\_Areas\_n, al realizar la transformación los datos son cargados de la tabla C\_Areas a la tabla C\_Areas\_n de la misma forma sin ocurrir ningún cambio. Ver anexo 7.

La tabla **C\_Cuenta** se transforma en la tabla C\_Cuenta\_n, al realizar la transformación, si el atributo obl\_cuemay tenía valor dos (2) ahora va a tomar el valor cero (0), en caso contrario sigue con el valor que tenía anteriormente. De igual forma cue\_subcue, nom\_cuenta, nom\_subcue, per\_operan, cod\_area, cue\_bancor, emis\_estado, obl\_asient, cal\_posici, canc\_cta, obl\_tipcon, nar\_explic, aplic\_cero, obl\_cartel son transferidos hacia la nueva tabla sin transformación. Los atributos Des\_subcuenta, Cal\_dispon, OBL\_balpag, OBL\_cimon, Print\_inter, Plazo, Tipo\_deuda y Name, no son cargados la tabla C\_Cuenta\_n. Los atributos cta\_dispo, cheq\_cierre, cheq\_sobr, ctr\_sobr, obl\_salsob, obl\_salres, obl\_estad, min\_dispon, contin\_cr, contin\_db, cue\_fondo, ctra\_conti, y es\_interna son adicionados en la tabla nueva C\_Cuenta\_n. Ver anexo 8.

La tabla **C\_Asient** se transforma en la tabla C\_Asient\_n, en este caso no ocurre ninguna transformación, solamente el atributo obl\_asient es adicionado en la nueva tabla C\_Asient\_n.

La tabla **C\_Concep** se transforma en la tabla C\_Concep\_n, en este caso no ocurre ninguna transformación, solamente el atributo cancelados es adicionado en la nueva tabla C\_Concep\_n. Ver anexo 9.

La tabla **C\_Tiptip** se transforma en la tabla C\_Tiptip\_n, en este caso no ocurre ninguna transformación, solamente los atributos for\_calc y tip\_calc son adicionados en la nueva tabla C\_Tiptip\_n. Ver anexo 10.

La tabla **C\_Otrcon** se transforma en la tabla C\_Otrcon\_n, el atributo Cod\_anter, no es cargado a la tabla C\_Otrcon\_n y el atributo cancelados es adicionado en la tabla nueva C\_Otrcon\_n. Ver anexo 11.

La tabla **C\_Moneda** se transforma en la tabla C\_Moneda\_n. El atributo Cod\_moneda de 3 caracteres pasa a ser de 2 tomando los 2 últimos valores. Los atributos Sig\_moneda, Nom\_moneda, Can\_decmon y

## *CAPÍTULO II. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA*

---

Nar\_moneda son transferidos hacia la nueva tabla sin transformación. Los atributos Orden, Cod\_iso, Canje, Cod\_monbar y Cancelados son adicionados en la tabla nueva C\_Cuenta\_n. Ver anexo 12.

La tabla **C\_Provinc** se transforma en la tabla C\_provinc\_n. Los atributos Cod\_provin, Nom\_provin son transferidos hacia la nueva tabla sin transformación. El atributo nom\_surcurs es desechado. Ver anexo 13.

La tabla **M\_Perban** se transforma en la tabla M\_Perban\_n. Los atributos Cod\_banco, Cod\_paexid, Num\_idper son trasferidos hacia la nueva tabla sin trasformación. El atributo Cod\_tipid es una constante con el valor "CI". Ver anexo 14.

La tabla **M\_Comper** se transforma en la tabla M\_Comper\_n. Los atributos Cod\_paexid, Num\_idper, Tip\_medcom, Num\_medcom son trasferidos hacia la nueva tabla sin trasformación. El atributo Cod\_tipid es una constante con el valor "CI". Ver anexo 15.

La tabla **M\_Percue** se transforma en la tabla M\_Percue\_n. En la nueva tabla surge el atributo cod\_moneda que no es más que el código de la moneda al que corresponde la sigla de la moneda (sig\_moneda) de la tabla c\_moneda (clasificador moneda). El atributo cod\_tipid es una constante ("CI") y el cod\_contra va a depender de 3 atributos: tip\_contra, cod\_contra y des\_cuenta. Si el tip\_contra es 1 se pone se rellena con 1 cero a la izquierda y se le concatena el des\_cuenta. Si el tip\_contra es 2 se concatena el cod\_contra con el des\_cuenta relleno con un cero a la izquierda. Los atributos cue\_subcue, tip\_contra, num\_idper son trasferidos hacia la nueva tabla sin trasformación. Los atributos cod\_catper, car\_person y porc\_benef son agregados a la nueva tabla. Ver anexo 16.

La tabla **M\_Mayor** se transforma en la tabla M\_Mayor\_n. En la nueva tabla surge el atributo cod\_moneda que no es más que el código de la moneda al que corresponde la sigla de la moneda (sig\_moneda) de la tabla c\_moneda (clasificador moneda). El atributo cod\_contra va a depender de 3 atributos: tip\_contra, cod\_contra y des\_cuenta. Si el tip\_contra es 1 se pone el cod\_contra, se rellena con 1 cero a la izquierda y se le concatena el des\_cuenta. Si el tip\_contra es 2 se concatena el cod\_contra con el des\_cuenta relleno con un cero a la izquierda. Los atributos tip\_contra, cue\_subcue y num\_estado son trasferidos

## *CAPÍTULO II. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA*

---

hacia la nueva tabla sin transformación. Los atributos sal\_sobaut, sal\_reserv y acu\_total son agregados a la nueva tabla. Ver anexo 17.

La tabla **M\_Tasint** se transforma en la tabla M\_Tasint\_n. En la nueva tabla surge el atributo cod\_moneda que no es más que el código de la moneda al que corresponde la sigla de la moneda (sig\_moneda) de la tabla c\_moneda (clasificador moneda). El atributo cod\_contra va a depender de 3 atributos: tip\_contra, cod\_contra y des\_cuenta. Si el tip\_contra es 1 se pone el cod\_contra, se rellena con 1 cero a la izquierda y se le concatena el des\_cuenta. Si el tip\_contra es 2 se concatena el cod\_contra con el des\_cuenta relleno con un cero a la izquierda. Los atributos tip\_contra, cue\_subcue y cod\_signo, fec\_valor, tas\_interere son transferidos hacia la nueva tabla sin transformacion. Los atributos imp\_ranint, dias\_base y cancelados son agregados a la nueva tabla. Ver anexo 18.

La tabla **M\_Percli** se transforma en la tabla M\_Percli\_n, El atributo cod\_tipid es una constante con el valor "CI". Los cod\_client, cod\_paexid y num\_idper son transferidos hacia la nueva tabla sin transformación. Ver anexo 19.

La tabla **M\_Diario** se transforma en la M\_Diario\_n. Los atributos num\_tranor, num\_asieor, num\_transa y num\_asient pasan a la nueva tabla y se le agregan 2 ceros a la izquierda y ide\_cue32 se le agrega un cero al final. En la nueva tabla surge el atributo cod\_moneda y cod\_mondcc que no son más que los códigos de la moneda al que corresponde la sigla de la moneda (sig\_moneda, sig\_mondcc) de la tabla c\_moneda (clasificador moneda). El cod\_contra va a depender de 3 atributos: tip\_contra, cod\_contra y des\_cuenta. Si el tip\_contra es 1 se pone se rellena con 1 cero a la izquierda y se le concatena el des\_cuenta. Si el tip\_contra es 2 se concatena el cod\_contra con el des\_cuenta relleno con un cero a la izquierda. La referencia corriente y la referencia original (ref\_corrie, ref\_origin) se transforma de la siguiente forma, se duplica el primer dígito de la referencia, se concatena el segundo, se rellena con cero a la izquierda lo siguiente hasta llegar a 7 y se le concatena el valor de la variable wcen\_costo que es una constante con el valor "000". El cod\_estad y cod\_marca son el cod\_superv y con\_balpag respectivamente y a los mismos se le adiciona 2 ceros a la izquierda. Los demás atributos son mapeados con la nueva tabla y pasados sin ninguna transformación (los que tienen el mismo nombre en la vieja y en la nueva). Ver anexo 20.

## CAPÍTULO II. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

La **tabla M\_Histor** se transforma en la M\_Histor\_n. El atributo `cod_ccosto` es una constante con valor "000". Los atributos `num_tranor`, `num_asieor`, `num_transa` y `num_asient` pasan a la nueva tabla y se le agregan 2 ceros a la izquierda y `ide_cue32` se le agrega un cero al final. En la nueva tabla surge el atributo `cod_moneda` y `cod_mondcc` que no son más que los códigos de la moneda al que corresponde la sigla de la moneda (`sig_moneda`, `sig_mondcc`) de la tabla `c_moneda` (clasificador moneda). El `cod_contra` va a depender de 3 atributos: `tip_contra`, `cod_contra` y `des_cuenta`. Si el `tip_contra` es 1 se pone se rellena con 1 cero a la izquierda y se le concatena el `des_cuenta`. Si el `tip_contra` es 2 se concatena el `cod_contra` con el `des_cuenta` relleno con un cero a la izquierda. La referencia corriente y la referencia original (`ref_corrie`, `ref_origin`) se transforma de la siguiente forma, se duplica el primer dígito de la referencia, se concatena el segundo, se rellena con cero a la izquierda lo siguiente hasta llegar a 7 y se le concatena el valor de la variable `wcen_costo` que es una constante con el valor "000". El `cod_estad` y `cod_marca` son el `cod_superv` y `con_balpag` respectivamente y a los mismos se le adiciona 2 ceros a la izquierda. Los demás atributos son mapeados con la nueva tabla y pasados sin ninguna transformación (los que tienen el mismo nombre en la vieja y en la nueva). Ver anexo 21.

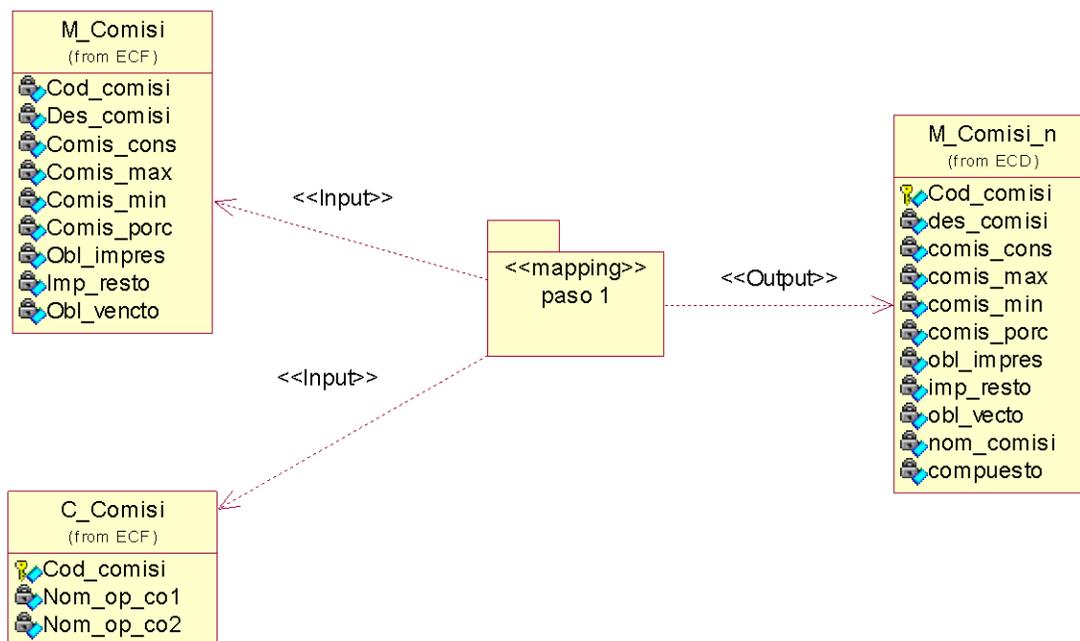


Figura 13. Mapeo de datos. Nivel 2. Tabla C\_Comisi y M\_Comisi.

# CAPÍTULO II. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

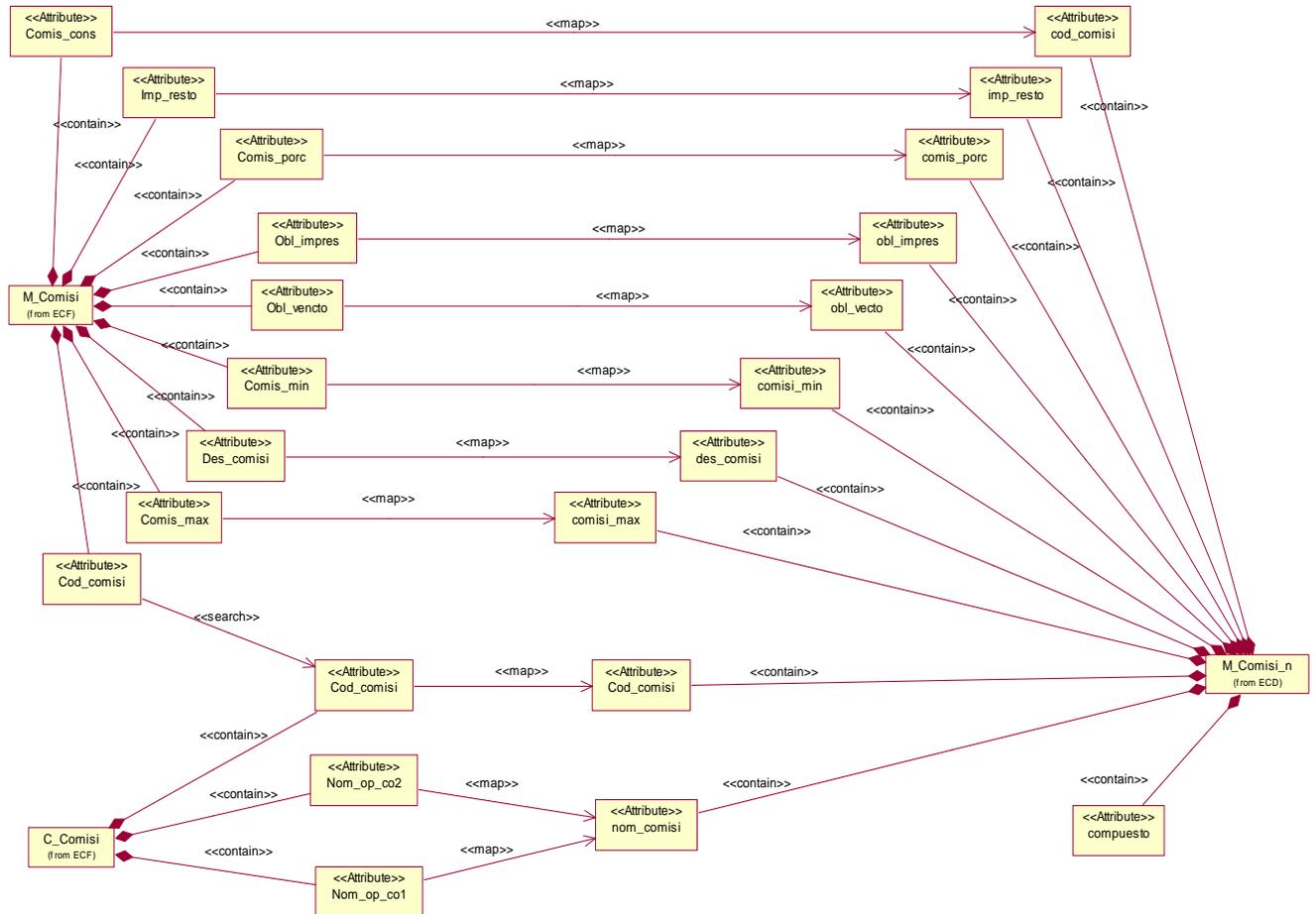


Figura 14. Mapeo de datos. Nivel 3. Tabla C\_Comisi y M\_Comisi

## 2.4 Conclusiones Parciales

En este capítulo se elaboraron los diagramas de mapeo de datos correspondientes a las tablas que se les harán las transformaciones de datos mediante los cuatro niveles de mapeo que propone el método DWEP, para así describir y ejemplificar mejor las acciones a realizar durante la implementación, lo cual se verá en el próximo capítulo.

## CAPÍTULO III. DESARROLLO DE LA SOLUCIÓN

### 3.1 Introducción

En este capítulo se explica cómo se realiza la implementación de la migración con el uso de la herramienta Pentaho Data Integration (Kettle ETL), comentando cada una de las funcionalidades que se utilizan para realizar la migración, además se exponen las pruebas realizadas a la migración de datos.

### 3.2 Funciones utilizadas de la herramienta

La herramienta Kettle ETL, está conformada por varias funcionalidades que permiten llevar a cabo las transformaciones de datos, a continuación se muestran algunas de ellas, las cuales fueron utilizadas en el desarrollo de la solución.

✓ **Función XBase input:**



XBase input

Esta función se utiliza para tomar los datos de los ficheros de de Fox Pro (.dbf).

✓ **Función output:**



Table output

Función para conectarse a la base de datos destino.

✓ **Función Add constants:**



Add constants

Función para añadir las constantes que sean necesarias.

✓ **Función If field value is null:**



If field value is null

Función para definir qué valores van a tomar los datos que sean nulos.

✓ **Función Filter rows:**



Filter rows

Función que permite filtrar por los valores necesarios (if-else).

✓ **Función Modified Java Script Value:**



Modified Java Script Value

Función que permite crear código Java Script.

✓ **Función Table output Mapping:**



Table output Mapping

Función para definir los mapeos de datos hacia la base de datos destino después de transformados los datos.

✓ **Función Database lookup:**



Database lookup

Función que permite hacer búsquedas en la base de datos.

## 3.3 Desarrollo

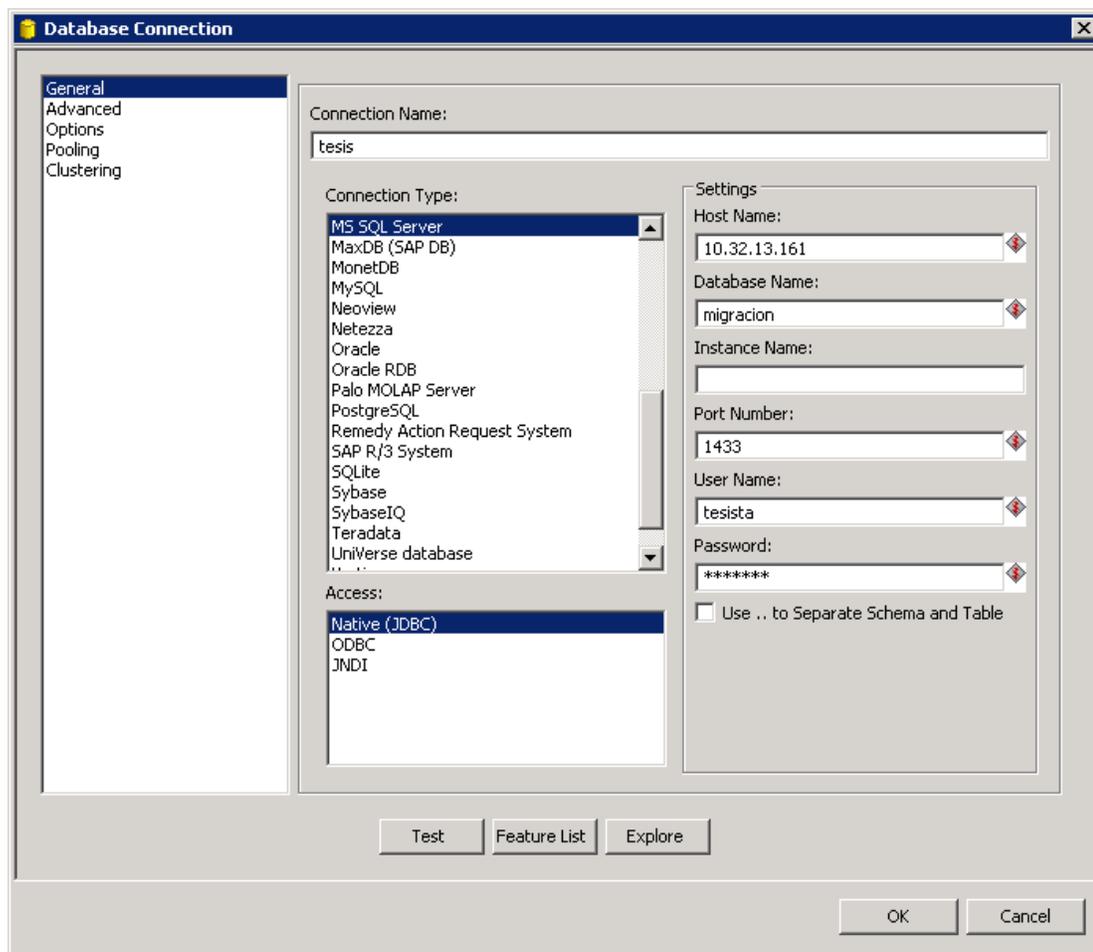
Para dar repuesta al objetivo propuesto, la migración se realizó en 2 fases:

# CAPÍTULO III. DESARROLLO DE LA SOLUCIÓN

**Fase 1: Desarrollo de las Transformaciones:** se realizó las trasformaciones necesarias para cada una de las tablas.

El siguiente diagrama corresponde a la tabla M\_Comins descrita en el capítulo anterior.

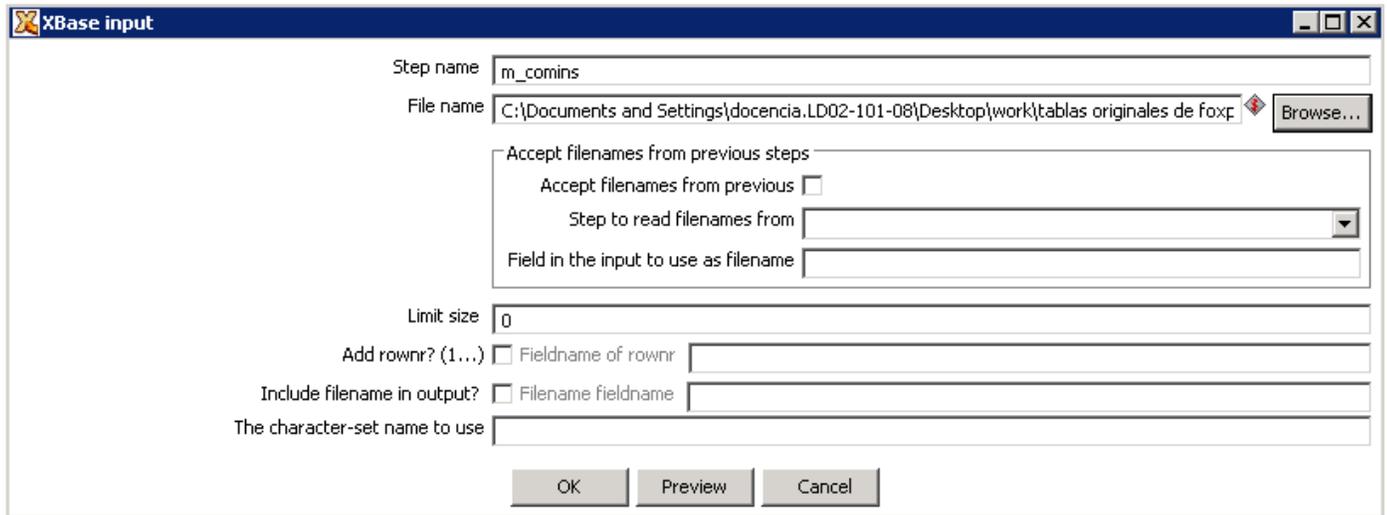
Estas trasformaciones comienzan creando la conexión a la base de datos destino, figura 15:



**Figura 15. Conexión a la Base de Datos.**

Posteriormente se especifica la entrada y salida a tabla para determinar la tabla que se utilizará y los campos a transformar, así como el destino de los mismos, figuras 16 y 17:

# CAPÍTULO III. DESARROLLO DE LA SOLUCIÓN

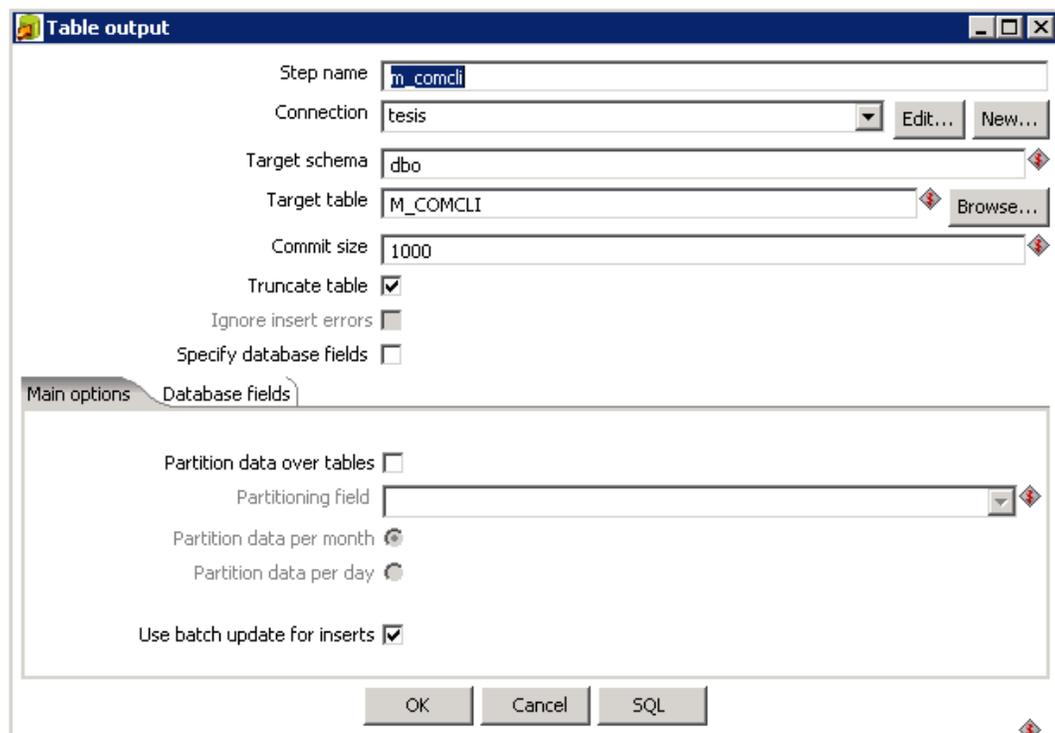


The screenshot shows the 'XBase input' dialog box. It has a title bar with a small icon and the text 'XBase input'. The main area contains several fields and options:

- Step name:** A text box containing 'm\_comins'.
- File name:** A text box containing 'C:\Documents and Settings\docencia.LD02-101-08\Desktop\work\tablas originales de foxp' with a 'Browse...' button to its right.
- Accept filenames from previous steps:** A section with a checkbox 'Accept filenames from previous' (unchecked), a dropdown menu 'Step to read filenames from', and a text box 'Field in the input to use as filename'.
- Limit size:** A text box containing '0'.
- Add rownr? (1...):** A checkbox (unchecked) followed by a text box 'Fieldname of rownr'.
- Include filename in output?:** A checkbox (unchecked) followed by a text box 'Filename fieldname'.
- The character-set name to use:** A text box.

At the bottom, there are three buttons: 'OK', 'Preview', and 'Cancel'.

Figura 16. Entrada Tabla.



The screenshot shows the 'Table output' dialog box. It has a title bar with a small icon and the text 'Table output'. The main area contains several fields and options:

- Step name:** A text box containing 'm\_comcli'.
- Connection:** A dropdown menu containing 'tesis' with 'Edit...' and 'New...' buttons to its right.
- Target schema:** A text box containing 'dbo' with a diamond icon to its right.
- Target table:** A text box containing 'M\_COMCLI' with a 'Browse...' button to its right.
- Commit size:** A text box containing '1000' with a diamond icon to its right.
- Truncate table:** A checked checkbox.
- Ignore insert errors:** An unchecked checkbox.
- Specify database fields:** An unchecked checkbox.

Below these options is a tabbed interface with two tabs: 'Main options' (selected) and 'Database fields'. Under 'Main options':

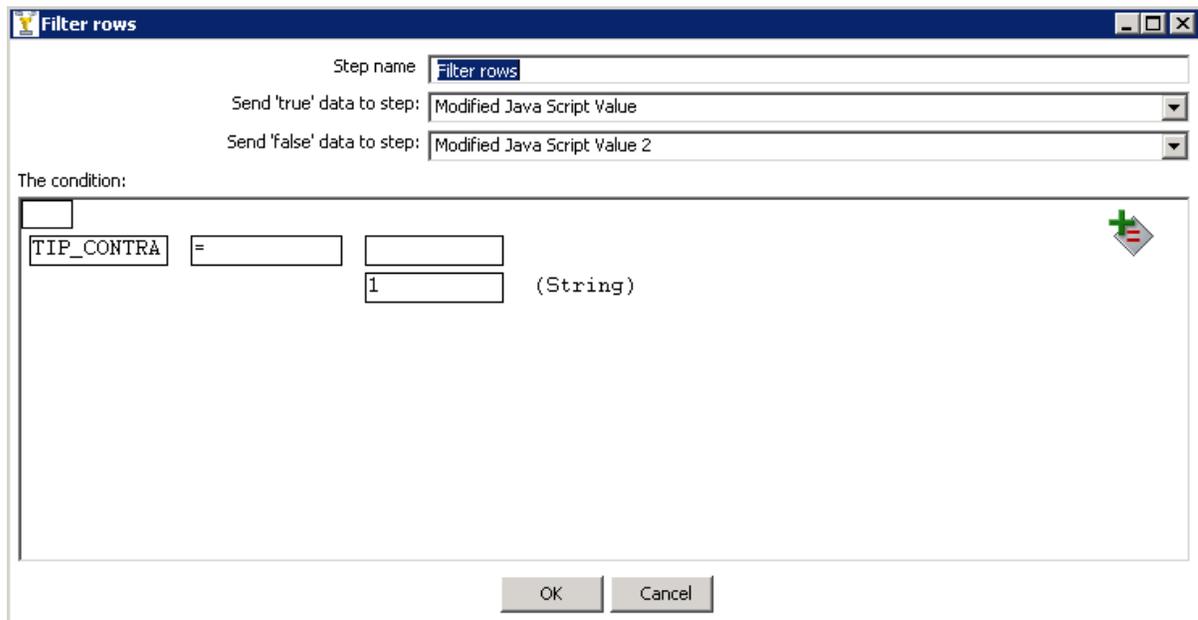
- Partition data over tables:** An unchecked checkbox.
- Partitioning field:** A dropdown menu with a diamond icon to its right.
- Partition data per month:** A radio button (selected).
- Partition data per day:** A radio button (unselected).
- Use batch update for inserts:** A checked checkbox.

At the bottom, there are three buttons: 'OK', 'Cancel', and 'SQL'.

Figura 17. Salida Tabla.

# CAPÍTULO III. DESARROLLO DE LA SOLUCIÓN

En el caso de la tabla M\_Comins (comunicación con la institución) se hace un filtro al campo tip\_contra, si este es 1 los datos pasarían a la tabla M\_Comcli (comunicación con cliente) y si es 2 a la tabla M\_Comban (comunicación con banco), figura 18:



**Figura 18. Filtrar campo.**

Una vez conocido el destino de los datos, se procede a realizar las transformaciones de los campos cod\_client y num\_medcom descritos en el capítulo anterior, figura19:

# CAPÍTULO III. DESARROLLO DE LA SOLUCIÓN

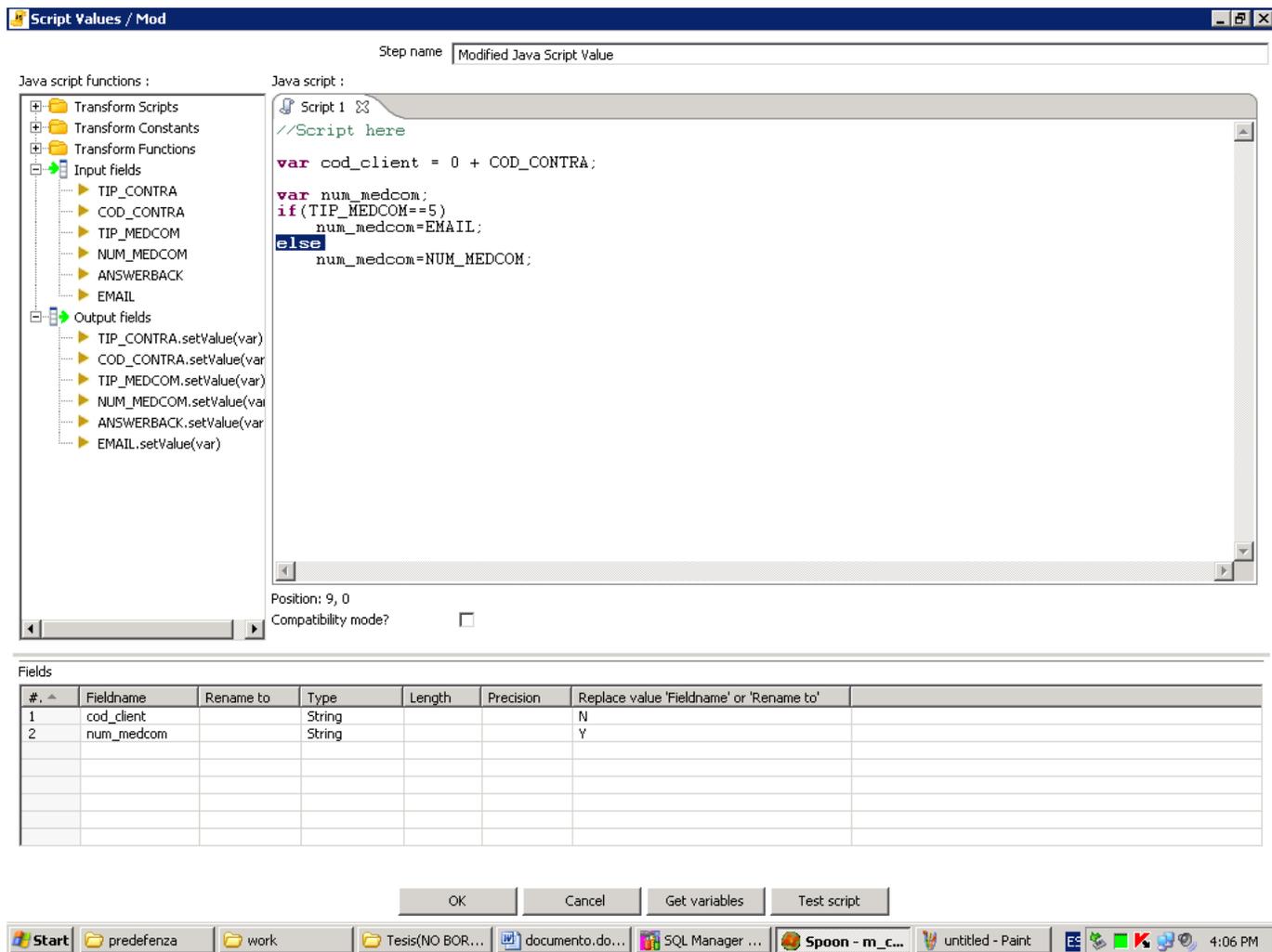


Figura 19. Modificar valores.

Al tener los datos ya transformados se mapean estos con la salida a la tabla, figura 20:

# CAPÍTULO III. DESARROLLO DE LA SOLUCIÓN

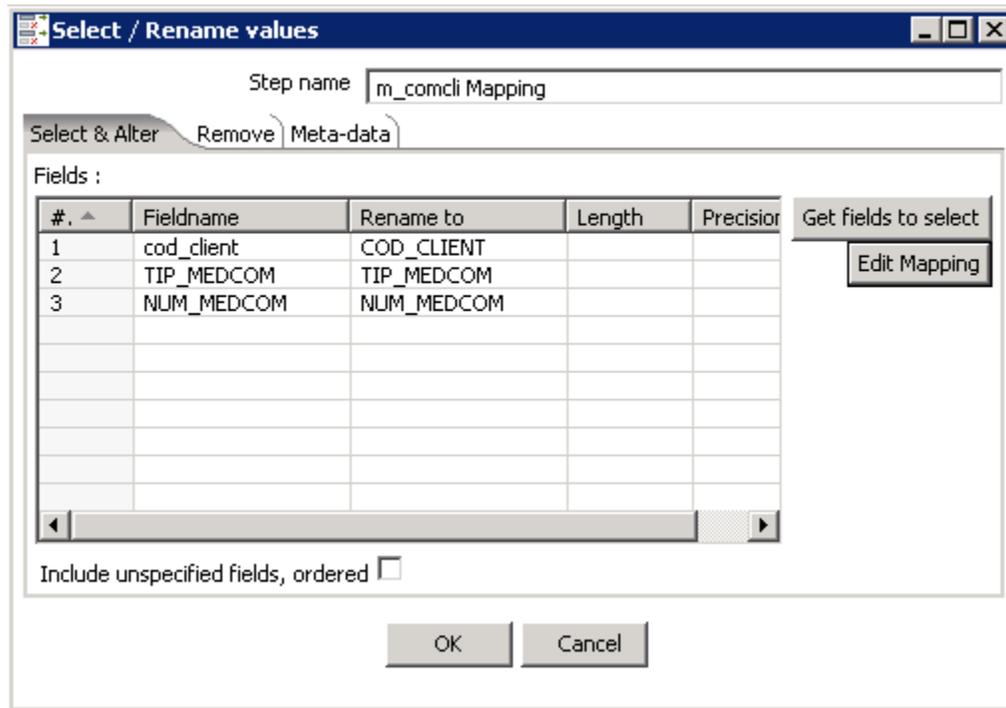


Figura 20. Tabla resultante.

El esquema resultante de la transformación es el que se muestra a continuación, figura 21:

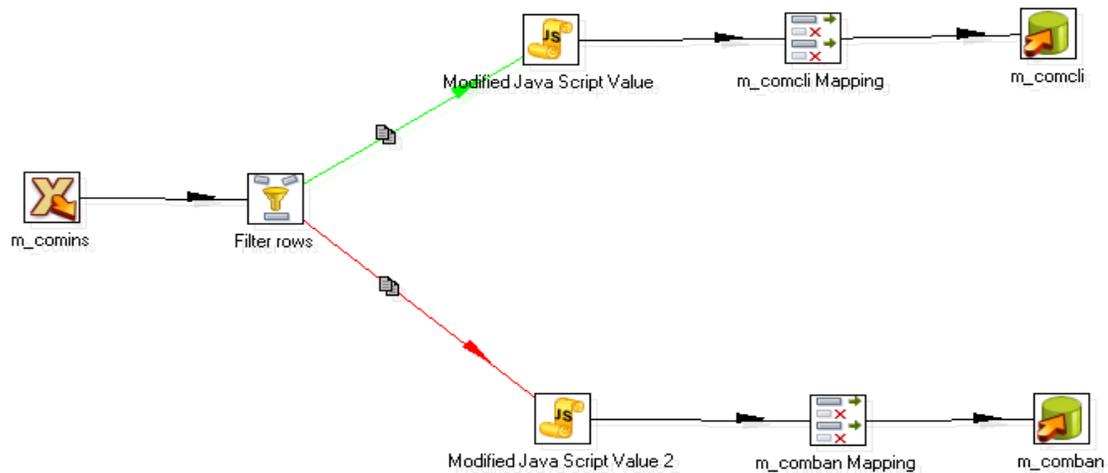


Figura 21. Esquema de la transformación M\_Comins.

# CAPÍTULO III. DESARROLLO DE LA SOLUCIÓN

Para la realización de la transformación para la tabla M\_Comisi también descrita en el capítulo 2 fue necesario realizar una transformación auxiliar para poder hacer la búsqueda necesaria en la tabla C\_Comisi ya que el Kettle no permite realizar búsquedas en ficheros de datos, estas solo se pueden realizar en base de datos. La misma se muestra a continuación.

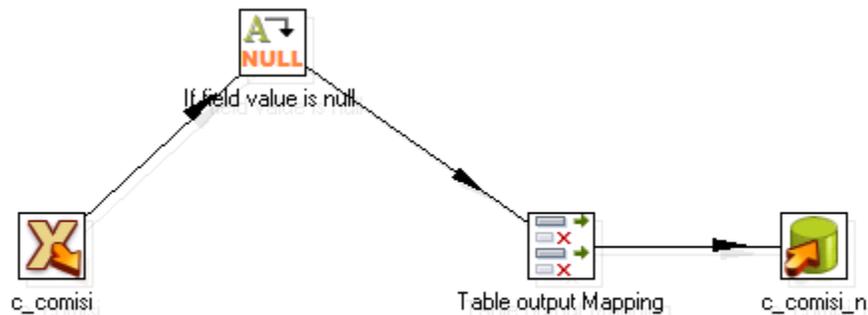


Figura 22. Esquema de la transformación auxiliar C\_Comisi.

Una vez insertados los datos en la tabla auxiliar se realiza la transformación para la nueva tabla M\_Comisi, Figura 23.

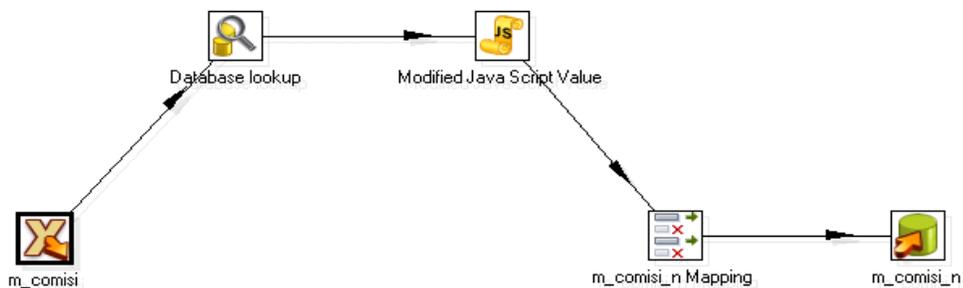


Figura 23. Esquema de la transformación M\_Comisi.

**Fase 2: Realización del Trabajo:** se construyó un esquema que contiene cada una de las transformaciones de la fase anterior y el orden en que se van a ejecutar cada una de ellas.

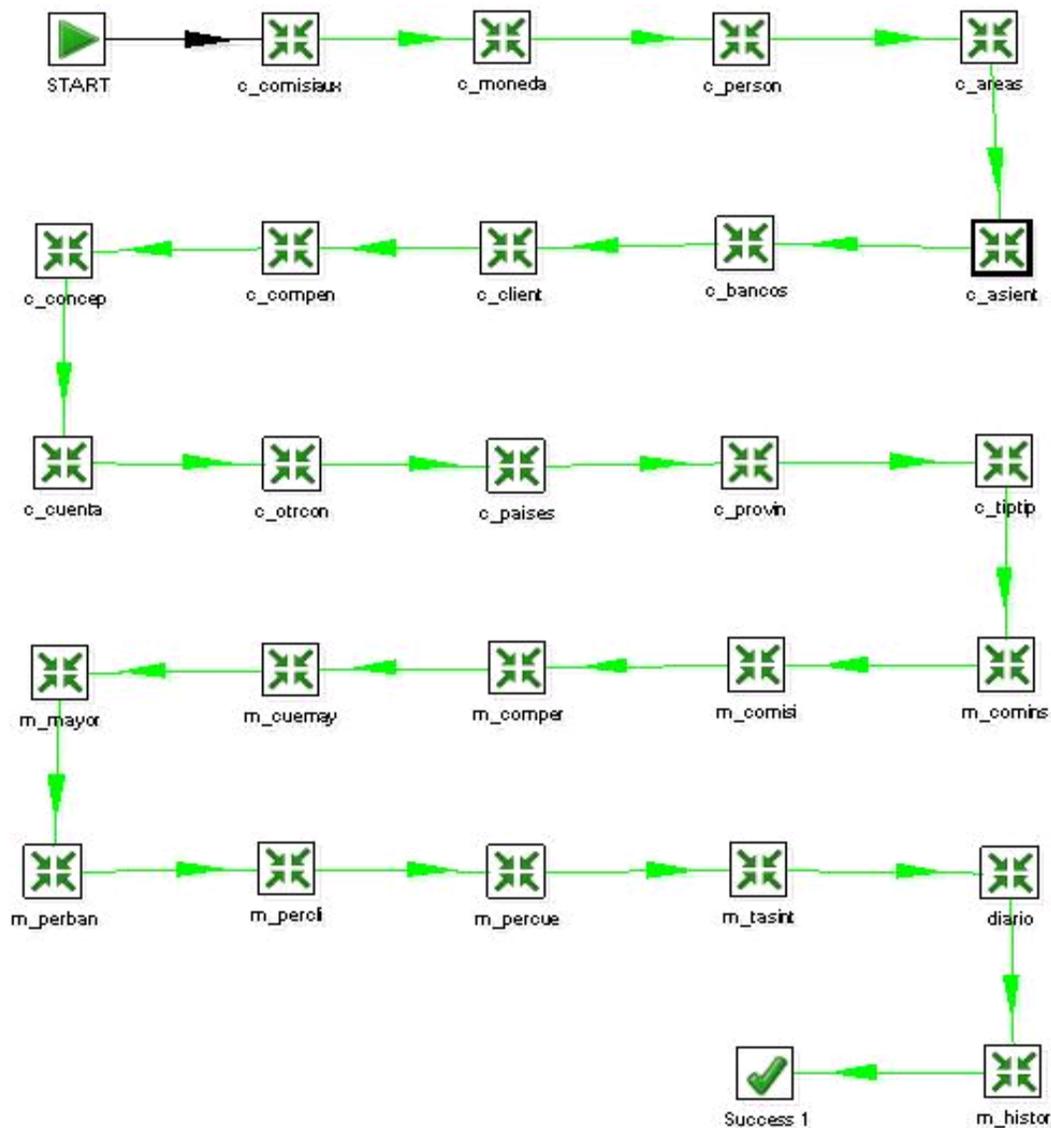


Figura 24. Esquema del trabajo para ejecutar las transformaciones.

### 3.3 Pruebas de integración de los datos

“Factor crítico para el éxito de la migración de la base de datos es la realización de pruebas, las cuales inicialmente, pueden ser a pequeña escala para validar o modificar la arquitectura final y el plan de migración, así como para comprobar que las aplicaciones que harán uso de la base de datos funcionan correctamente y optimizar los tiempos y recursos necesarios.” (Correa, 2008).

## CAPÍTULO III. DESARROLLO DE LA SOLUCIÓN

---

Para la realización de las pruebas de integración de datos se tomaron juegos de datos de las tablas origen y se determinaron los posibles resultados que se debían obtener. Se crearon tablas auxiliares por cada tabla origen y se le insertaron los datos que se deseaban lograr si la estrategia propuesta funcionaba correctamente.

Para finalmente comprobar si tuvo éxito la migración se realizaron consultas a las tablas finales después de la migración y a las tablas auxiliares, para verificar si los datos obtenidos eran iguales, garantizando así el éxito de la estrategia.

Las siguientes tablas muestran lo anteriormente explicado para las tablas: M\_Diario y M\_Histor.

**M\_Diario**

Nombre Atributo	Entrada	Transformación	Salida Esperada	Salida Real
Fec_posteo	1996/03/28		1996/03/28	3/28/1996
Num_tranor	00312	00+ Num_tranor	0000312	0000312
Num_asieor	872	00+ Num_asieor	00872	00872
Tip_asidia	02		02	02
Sig_moneda	CAD	Cod_moneda	03	03
Cue_subcue	7361		7361	7361
Tip_contra	2		2	2
Cod_contra	7307	Si el tip_contra es 1 se pone se rellena con 1 cero a la izquierda y se le concatena el des_cuenta. Si el tip_contra es 2 se	0730700	0730700

## *CAPÍTULO III. DESARROLLO DE LA SOLUCIÓN*

		concatena el cod_contra con el des_cuenta relleno con un cero a la izquierda.		
Des_cuenta	00	Se desecha		
Fec_valor	1993/03/08		1993/03/08	3/8/1993
Ref_corrie	H6311561	Se duplica el primer digito de la referencia, se concatena el segundo, se rellena con cero a la izquierda lo siguiente hasta llegar a 7 y se le concatena el valor de la variable wcn_costo (000).	0000000HH6000	0000000HH6000
Imp_asient	45942		45942	45942
Cod_asient	000		000	000
Fec_contab	1993/03/08		1993/03/08	3/8/1993
Cod_operad	999	00+ Cod_operad	00999	00999
Cod_superv	999	Cod_marca = 00+ Cod_superv	00999	00999
Fec_vecto	1996/03/23		1996/03/23	3/23/1996
Con_balpag		Cod_estad = 00+ Con_balpag		
Num_maquin	041		041	041

## *CAPÍTULO III. DESARROLLO DE LA SOLUCIÓN*

Ref_origin	H00055	El mismo tratamiento que la referencia corriente	0000000HH0000	0000000HH0000
Ref_extern				
Ext_automa	N		0	0
Cod_transa	99999		99999	99999
Sig_mondcc		Cod_mondcc		
Des_cuedcc		Se desecha		
Ide_cue32		Si no esta vacio idecue32 + 0		
Cod_comisi	000		000	000
Observ				
Num_transa	00312	00+ Num_transa	0000312	0000312
Num_asient	872	00+ Num_asient	00872	00872
CRC				
Cod_ccosto		Constante (000)	000	

A continuación se muestra la consulta realizada para comprobar la integridad de los datos.

```
SELECT * FROM dbo.M_DIARIO d, dbo.M_DIARIO_prueba p WHERE d.FEC_POSTEO =
p.FEC_POSTEO AND d.COD_CCOSTO = p.COD_CCOSTO AND d.NUM_TRANOR = p.NUM_TRANOR
AND d.NUM_ASIEOR = p.NUM_ASIEOR AND d.TIP_ASIDIA = p.TIP_ASIDIA AND d.COD_MONEDA =
p.COD_MONEDA AND d.CUE_SUBCUE = p.CUE_SUBCUE AND d.TIP_CONTRA = p.TIP_CONTRA
AND d.COD_CONTRA = p.COD_CONTRA AND d.FEC_VALOR = p.FEC_VALOR AND d.REF_CORRIE =
p.REF_CORRIE AND d.IMP_ASIENT = p.IMP_ASIENT AND d.COD_ASIENT = p.COD_ASIENT AND
d.FEC_CONTAB = p.FEC_CONTAB AND d.COD_OPERAD = p.COD_OPERAD AND d.COD_MARCA =
p.COD_OPERAD AND d.FEC_VENCTO = p.FEC_VENCTO AND d.COD_ESTAD = p.COD_ESTAD AND
d.NUM_MAQUIN = p.NUM_MAQUIN AND d.REF_ORIGIN = p.REF_ORIGIN AND d.REF_EXTERN =
p.REF_EXTERN AND d.EXT_AUTOMA = p.EXT_AUTOMA AND d.COD_TRANSA = p.COD_TRANSA
AND d.COD_MONDCC = p.COD_MONDCC AND d.IDE_CUE32 = p.IDE_CUE32 AND d.COD_COMISI =
```

## *CAPÍTULO III. DESARROLLO DE LA SOLUCIÓN*

p.COD\_COMISI AND d.OBSERV = p.OBSERV AND d.NUM\_TRANSA = p.NUM\_TRANSA AND d.NUM\_ASIENT = p.NUM\_ASIENT AND d.CRC = p.CRC AND d.CTA\_REAL = p.CTA\_REAL ;

### M\_Histor

Atributo	Entrada	Transformación	Salida Esperada	Salida Resultante
Fec_acthist	1996/03/28		1996/03/28	3/28/1996
Cod_ccosto		Constante(000)	000	000
Num_transa	00001	00+ Num_transa	0000001	0000001
Num_asient	012	00+ Num_asient	00012	00012
Fec_posteo	1996/03/28		1996/03/28	3/28/1996
Sig_moneda	DEM	Cod_moneda	08	08
Cue_subcue	4863		4863	4863
Tip_contra	2		2	2
Cod_contra	0850	Si el tip_contra es 1 se pone se rellena con 1 cero a la izquierda y se le concatena el des_cuenta. Si el tip_contra es 2 se concatena el cod_contra con el des_cuenta relleno con un cero a la izquierda.	0085000	0085000
Des_cuenta	00	Se desecha		
Cod_asient	000		000	000

## *CAPÍTULO III. DESARROLLO DE LA SOLUCIÓN*

Fec_contab	1990/06/14		1990/06/14	6/14/1990
Imp_asient	-185185.19		-185185.19	-185185.19
Fec_valor	1990/06/14		1990/06/14	6/14/1990
Ref_corrie	H6000012	Se duplica el primer dígito de la referencia, se concatena el segundo, se rellena con cero a la izquierda lo siguiente hasta llegar a 7 y se le concatena el valor de la variable wcen_costo (000).	0000000HH6000	0000000HH6000
Cod_operad	999	00+ Cod_operad	00999	00999
Cod_superv	999	Cod_marca =00+ Cod_superv	00999	00999
Fec_vecto	1993/02/05		1993/02/05	2/5/1993
Con_balpag		Cod_estado =00+ Con_balpag		
Tip_asidia	99		99	99
Ref_origin	H13684	Se duplica el primer dígito de la referencia, se concatena el segundo, se rellena con cero a la izquierda lo siguiente hasta	0000000HH1000	0000000HH1000

## *CAPÍTULO III. DESARROLLO DE LA SOLUCIÓN*

		llegar a 7 y se le concatena el valor de la variable wcen_costo (000).		
Ref_extern				
Ide_cue32		Si no está vacío idecue32 + 0		
Ext_automa	N		0	0
Cod_transa	99999		99999	99999
Cod_comisi	000		000	000
Sig_mondcc		Cod_moneda		
Des_cuedcc		Se desecha		
Observ				
Num_asieor	012	00+ Num_asieor	00012	00012
Num_tranor	00001	00+ Num_tranor	0000001	0000001
Hor_accion	08:01:10		08:01:10	08:01:10
CRC				

Se realizó la siguiente consulta para comprobar la integridad de los datos.

```
SELECT * FROM dbo.M_HISTOR h, dbo.M_HISTOR_prueba p WHERE h.FEC_ACTHIS = p.FEC_ACTHIS AND h.COD_CCOSTO = p.COD_CCOSTO AND h.NUM_TRANSA = p.NUM_TRANSA AND h.NUM_ASIENT = p.NUM_ASIENT AND h.FEC_POSTEO = p.FEC_POSTEO AND h.COD_MONEDA = p.COD_MONEDA AND h.CUE_SUBCUE = p.CUE_SUBCUE AND h.TIP_CONTRA = p.TIP_CONTRA AND h.COD_CONTRA = p.COD_CONTRA AND h.COD_ASIENT = p.COD_ASIENT AND h.FEC_CONTAB = p.FEC_CONTAB AND h.IMP_ASIENT = p.IMP_ASIENT AND h.FEC_VALOR = p.FEC_VALOR AND h.REF_CORRIE = p.REF_CORRIE AND h.COD_OPERAD = p.COD_OPERAD AND h.COD_MARCA = p.COD_OPERAD AND h.FEC_VENCTO = p.FEC_VENCTO AND h.COD_ESTAD = p.COD_ESTAD AND h.TIP_ASIDIA = p.TIP_ASIDIA AND h.REF_ORIGIN = p.REF_ORIGIN AND
```

# *CAPÍTULO III. DESARROLLO DE LA SOLUCIÓN*

---

h.REF\_EXTERN = p.REF\_EXTERN AND h.IDE\_CUE32 = p.IDE\_CUE32 AND h.EXT\_AUTOMA = p.EXT\_AUTOMA AND h.COD\_TRANSA = p.COD\_TRANSA AND h.COD\_COMISI = p.COD\_COMISI AND h.COD\_MONDCC = p.COD\_MONDCC AND h.OBSERV = p.OBSERV AND h.NUM\_ASIEOR = p.NUM\_ASIEOR AND h.NUM\_TRANOR = p.NUM\_TRANOR AND h.HOR\_ACCION = p.HOR\_ACCION AND h.CRC = p.CRC AND h.CTA\_REAL = p.CTA\_REAL AND h.ADJUNTO = p.ADJUNTO AND h.NUM\_MAQUIN = p.NUM\_MAQUIN;

Luego de haberse ejecutado las consultas especificadas anteriormente, se obtuvo que los datos esperados en el proceso de la migración eran correctos.

## **3.4 Resultado de la migración**

Con la realización de cada una de las transformaciones y la ejecución del trabajo que ejecutará todas las transformaciones de forma automática, los datos que estaban en Fox Pro son trasladados al Sistema Gestor de Bases de Datos SQL Server 2005, comprobando la integridad de los mismos a través de pruebas de integridad.

## **3.5 Conclusiones Parciales**

En este capítulo se describió detalladamente el proceso de Extracción, Transformación y Carga usando Pentaho Data Integration (Kettle ETL), así como las pruebas de integridad realizadas para verificar el éxito de la migración de datos.

### *CONCLUSIONES GENERALES*

Con este trabajo se diseñó y se implementó una estrategia de migración de datos desde el sistema del BNC en FoxPro a SQL Server 2005, que permite un rápido traspaso de los datos, haciendo las limpiezas y transformaciones necesarias para que la información mantenga su integridad y no se afecte su consistencia.

Para dar cumplimiento a esto:

- ✓ Se efectuó primeramente un estudio sobre las herramientas, tecnologías, metodologías y tendencias actuales propuestas para la realización de transformaciones y migraciones de datos, con el objetivo de identificar y seleccionar las posibles a utilizar en la solución del problema planteado.
- ✓ Se obtuvieron una serie de diagramas para lograr una mejor comprensión de la estrategia utilizando el método DWEP.
- ✓ Se realizó el proceso de extracción, transformación y carga de datos para filtrar, limpiar, homogenizar y agrupar la información proveniente de la fuente de datos.
- ✓ Por último se realizaron pruebas de integración de datos para comprobar la integridad y consistencia de los mismos luego de realizada la migración.

### ***RECOMENDACIONES***

Ya terminado este trabajo y teniendo en cuenta que se han cumplido todos los objetivos planteados, se recomienda:

- ✓ Motivar el estudio del método DWEP, para realizar migraciones de datos.
- ✓ Aplicar la metodología DWEP en otros proyectos de la universidad relacionados con migración de datos.

## *REFERENCIAS BIBLIOGRÁFICAS*

**2008.** [En línea] 2 de Febrero de 2008. [Citado el: 15 de Enero de 2010.]

<http://fundamentosinformaticosjl.wordpress.com/category/base-de-datos/>.

**2010.** [En línea] 2010. [Citado el: 1 de Febrero de 2010.]

[http://www.ar.atosorigin.com/esar/servicios/nuestra\\_oferta/integracion\\_de\\_sistemas/tecnologias\\_expertise/dases\\_de\\_datos\\_migracion\\_de\\_datos/default.htm](http://www.ar.atosorigin.com/esar/servicios/nuestra_oferta/integracion_de_sistemas/tecnologias_expertise/dases_de_datos_migracion_de_datos/default.htm).

**Pérez, Paqui Valle. 2009.** [En línea] 31 de Marzo de 2009. [Citado el: 25 de Enero de 2010.]

<http://www.mailxmail.com/curso-informatica-administracion-publica-3/bases-datos-concepto-caracteristicas-funcionalidades>.

**Valdes, Damián Perez. 2007.** [En línea] 6 de Octubre de 2007. [Citado el: 20 de Enero de 2010.]

<http://www.maestrosdelweb.com/principiantes/%C2%BFque-son-las-bases-de-datos/>.

**Collado, Albert.** [En línea] [Citado el: 5 de marzo de 2010.]

<http://dialnet.unirioja.es/servlet/articulo?codigo=2241561>.

TRUJILLO, J.; M. PALOMAR, et al. Extending UML for Multidimensional Modeling. Proceedings of the 5th International Conference on the Unified Modeling Language, Dresden, Germany, Springer- Verlag, 2002. 290-304 p.

**Mora, Roberto Canales. 2008.** [En línea] 3 de Diciembre de 2008. [Citado el: 10 de Marzo de 2010.]

<http://www.adictosaltrabajo.com/tutoriales/tutoriales.php?pagina=talend>.

**2010.** [En línea] 2010. [Citado el: 3 de Abril de 2010.] <http://es.talend.com/>.

**2008.** [En línea] 2008. [Citado el: 20 de Febrero de 2010.] <http://pentaho.almacen-datos.com/>.

## *REFERENCIAS BIBLIOGRÁFICAS*

---

**Gracia, Joaquin. 2005.** [En línea] 7 de Mayo de 2005. [Citado el: 21 de Marzo de 2010.]

<http://www.ingenierossoftware.com/analisisydiseno/uml.php>.

**2010.** [En línea] 2010. [Citado el: 22 de Febrero de 2010.] [http://es.talend.com/solutions-data-](http://es.talend.com/solutions-data-integration/data-migration.php)

[integration/data-migration.php](http://es.talend.com/solutions-data-integration/data-migration.php).

**Correa, Susana Corona. 2008.** [En línea] 27 de Noviembre de 2008. [Citado el: 23 de Abril de 2010.]

<http://www.enterate.unam.mx/Articulos/2006/agosto/migracion.htm>.

## BIBLIOGRAFÍA

**2008.** [En línea] 2 de Febrero de 2008. [Citado el: 15 de Enero de 2010.]

<http://fundamentosinformaticosjl.wordpress.com/category/base-de-datos/>.

**2010.** [En línea] 2010. [Citado el: 1 de Febrero de 2010.]

[http://www.ar.atosorigin.com/esar/servicios/nuestra\\_oferta/integracion\\_de\\_sistemas/tecnologias\\_expertise/dases\\_de\\_datos\\_migracion\\_de\\_datos/default.htm](http://www.ar.atosorigin.com/esar/servicios/nuestra_oferta/integracion_de_sistemas/tecnologias_expertise/dases_de_datos_migracion_de_datos/default.htm).

**Pérez, Paqui Valle. 2009.** [En línea] 31 de Marzo de 2009. [Citado el: 25 de Enero de 2010.]

<http://www.mailxmail.com/curso-informatica-administracion-publica-3/bases-datos-concepto-caracteristicas-funcionalidades>.

**Valdes, Damián Perez. 2007.** [En línea] 6 de Octubre de 2007. [Citado el: 20 de Enero de 2010.]

<http://www.maestrosdelweb.com/principiantes/%C2%BFque-son-las-bases-de-datos/>.

**Collado, Albert.** [En línea] [Citado el: 5 de marzo de 2010.]

<http://dialnet.unirioja.es/servlet/articulo?codigo=2241561>.

TRUJILLO, J.; M. PALOMAR, et al. Extending UML for Multidimensional Modeling. Proceedings of the 5th International Conference on the Unified Modeling Language, Dresden, Germany, Springer- Verlag, 2002. 290-304 p.

**Mora, Roberto Canales. 2008.** [En línea] 3 de Diciembre de 2008. [Citado el: 10 de Marzo de 2010.]

<http://www.adictosaltrabajo.com/tutoriales/tutoriales.php?pagina=talend>.

**2010.** [En línea] 2010. [Citado el: 3 de Abril de 2010.] <http://es.talend.com/>.

**2008.** [En línea] 2008. [Citado el: 20 de Febrero de 2010.] <http://pentaho.almacen-datos.com/>.

**Gracia, Joaquin. 2005.** [En línea] 7 de Mayo de 2005. [Citado el: 21 de Marzo de 2010.]

<http://www.ingenierosoftware.com/analisisydiseno/uml.php>.

**2010.** [En línea] 2010. [Citado el: 22 de Febrero de 2010.] <http://es.talend.com/solutions-data-integration/data-migration.php>.

**Espinosa, Roberto. 2010.** [En línea] 1 de Junio de 2010. [Citado el: 10 de Junio de 2010.]

<http://churriwifi.wordpress.com/2010/06/01/comparativa-talend-vs-kettle-pdi/>.

**Khirr. 2010.** [En línea] 19 de Febrero de 2010. [Citado el: 23 de Marzo de 2010.]

<http://www.actualidadinformatica.com/respaldar-base-de-datos-de-mysql/>.

**Roldan, Maria Carina. 2010.** [En línea] Abril de 2010. [Citado el: 11 de Mayo de 2010.]

<http://wiki.pentaho.com/display/COM/Pentaho+3.2+Data+Integration+-+Beginner%27s+Guide>.

**Correa, Susana Corona. 2008.** [En línea] 27 de Noviembre de 2008. [Citado el: 23 de Abril de 2010.]

<http://www.enterate.unam.mx/Articulos/2006/agosto/migracion.htm>.

**2009.** [En línea] 8 de Octubre de 2009. [Citado el: 27 de Abril de 2010.] <http://www.hosting-powerhost.com/2009/05/indicaciones-utiles-para-migrar-una-base-de-datos-a-mysql-es-decir-cuando-tenemos-que-subir-una-base-de-datos-local-en-cualquier-gestor-a-una-base-de-datos-remota-en-mysql/>.

**Oñate, Juan. 2006.** [En línea] 25 de Septiembre de 2006. [Citado el: 12 de Marzo de 2010.]

<http://www.techweek.es/software-servicios-ti/opinion/1000660004701/metodologia-exito-migracion-datos.1.html>.

Anexo 1: Mapeo de datos: Tabla C\_Client.

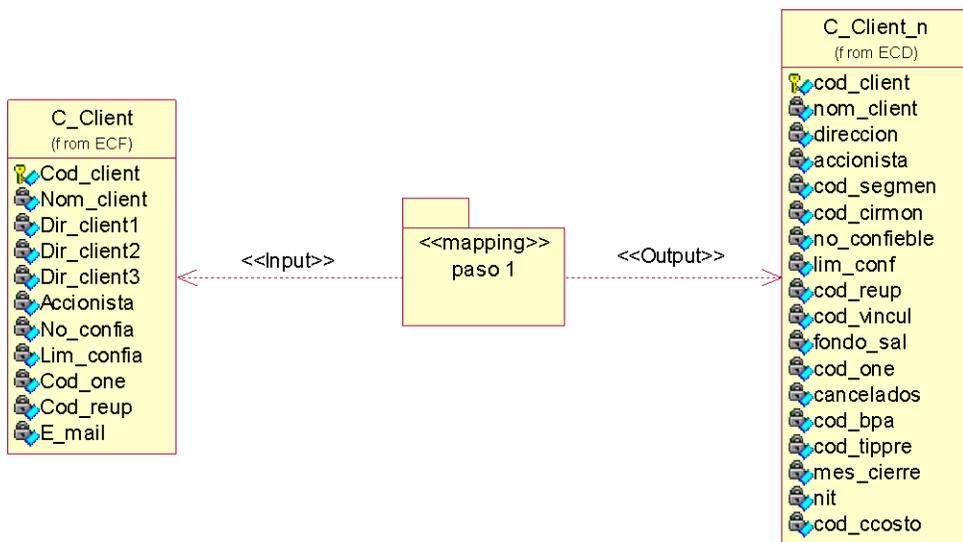


Figura 25. Mapeo de datos. Nivel 2. Tabla C\_Client.

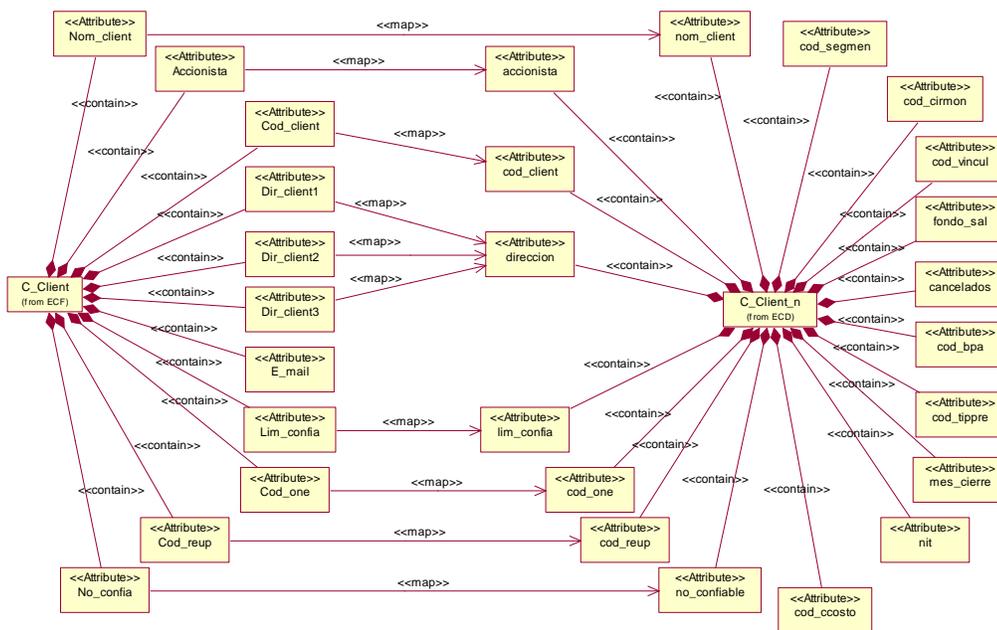


Figura 26. Mapeo de datos. Nivel 3. Tabla C\_Client.

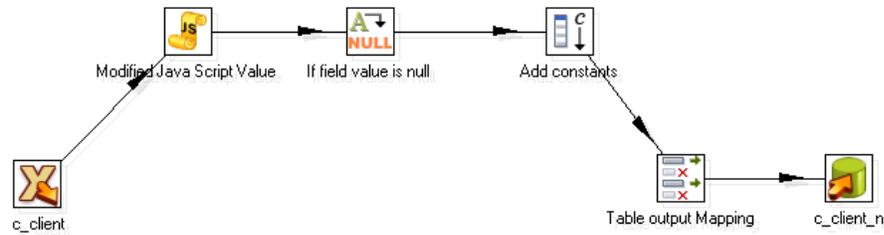


Figura 27. Esquema de la transformación Tabla C\_Client.

Anexo 2: Mapeo de datos: Tabla C\_Person.

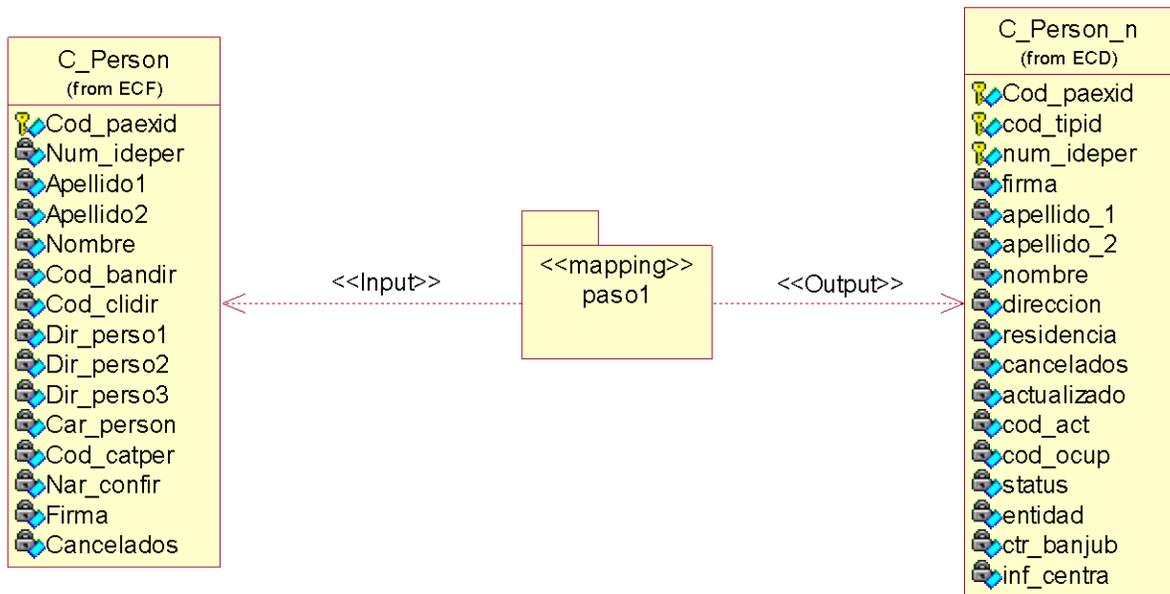


Figura 28. Mapeo de datos. Nivel 2. Tabla C\_Person.

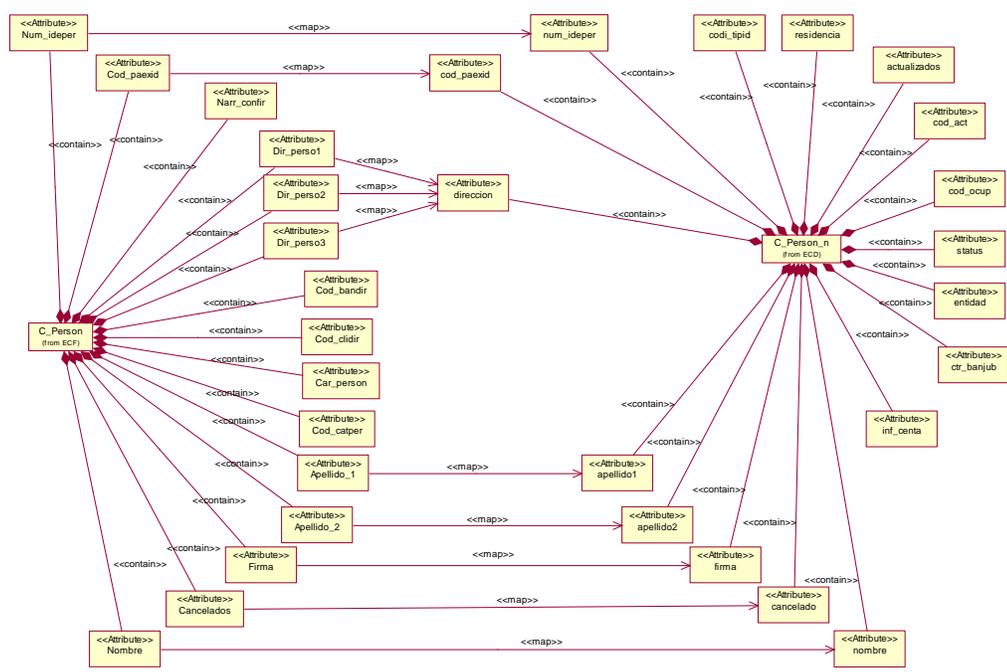


Figura 29. Mapeo de datos. Nivel 3. Tabla C\_Person.

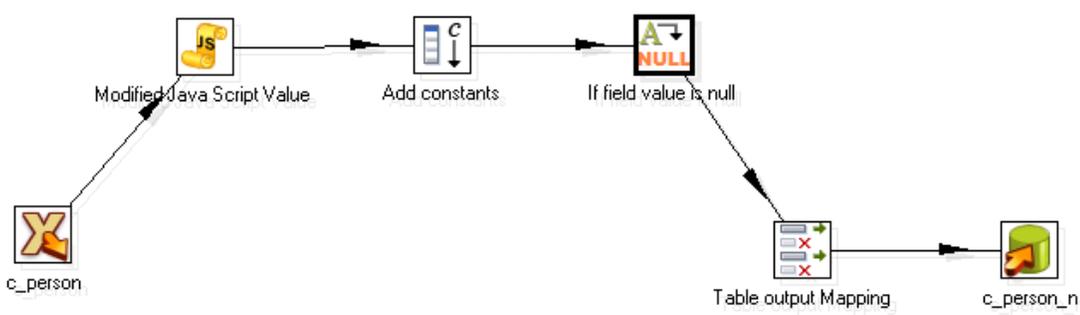


Figura 30. Esquema de la transformación Tabla C\_Person.

Anexo 3: Mapeo de datos: Tabla M\_Perban.

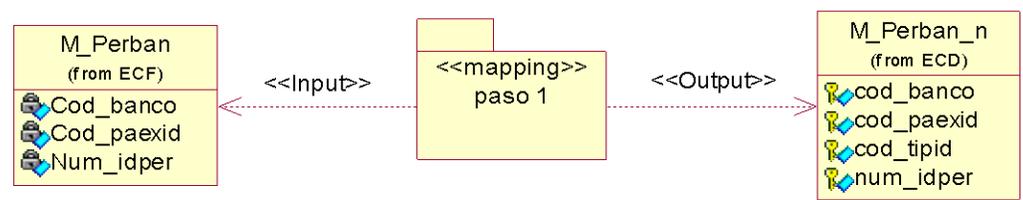


Figura 31. Mapeo de datos. Nivel 2. Tabla M\_Perban.

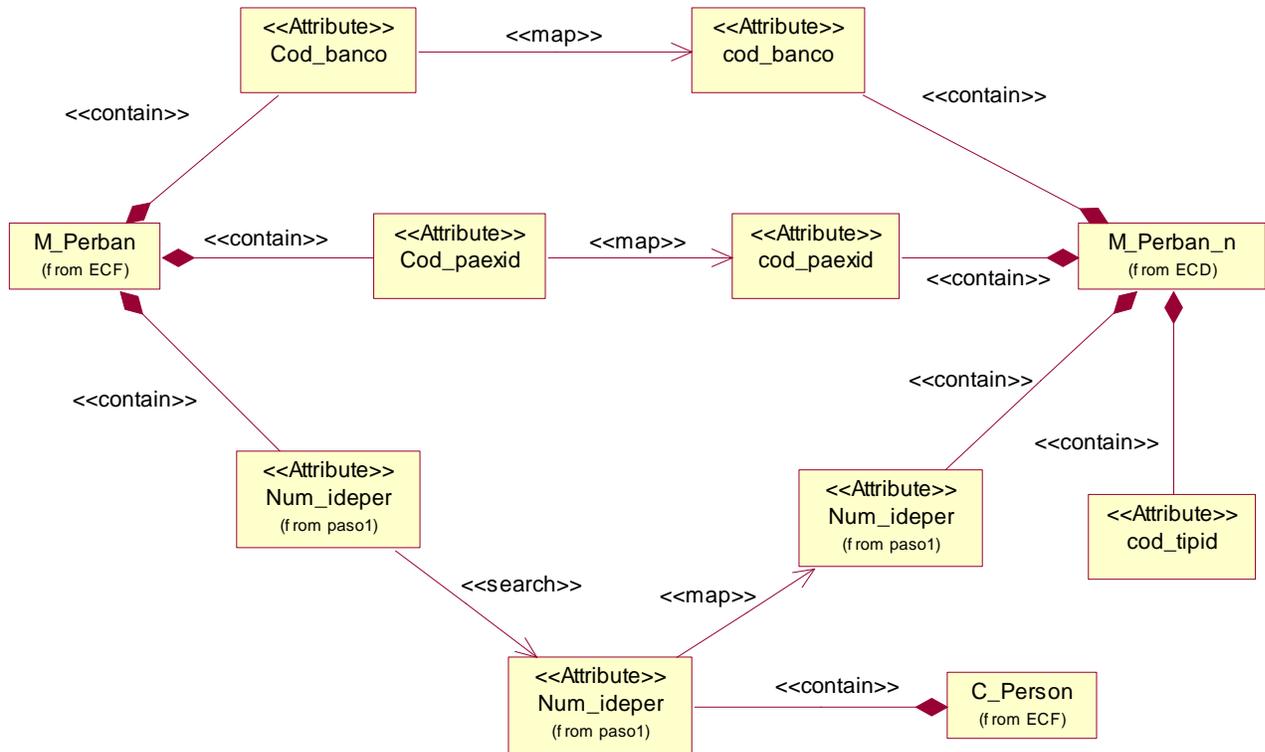


Figura 32. Mapeo de datos. Nivel 3. Tabla M\_Perban.

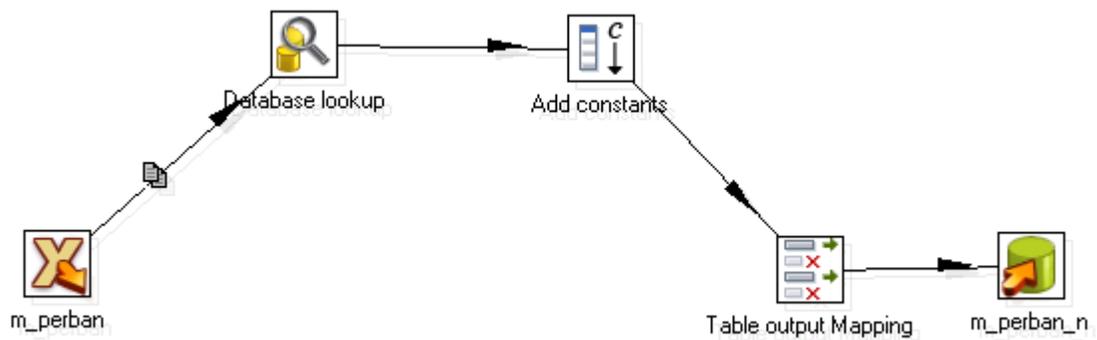


Figura 33. Esquema de la transformación Tabla M\_Perban.

Anexo 4: Mapeo de datos: Tabla M\_Comper.

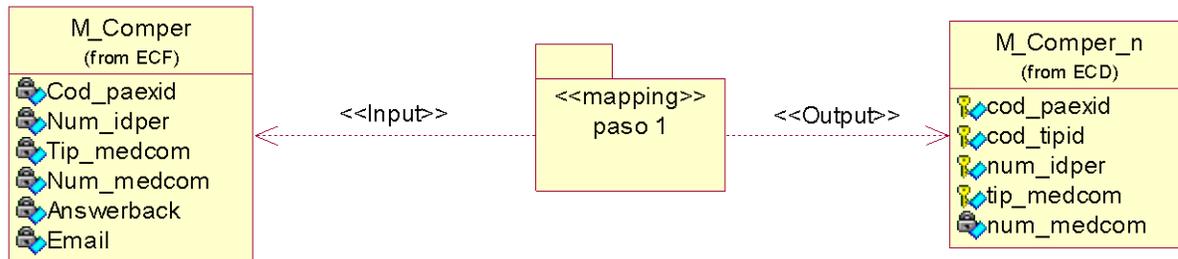


Figura 34. Mapeo de datos. Nivel 2. Tabla M\_Comper.

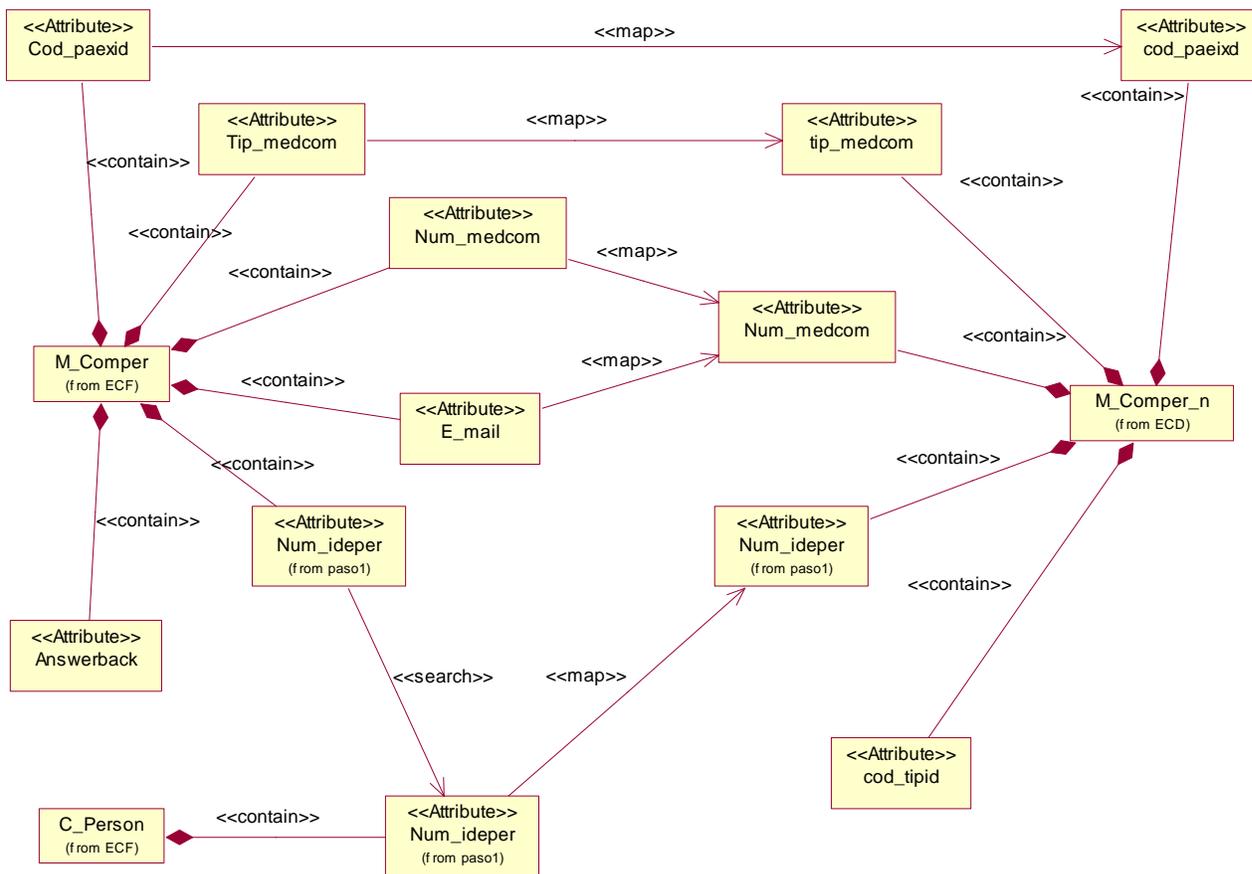


Figura 35. Mapeo de datos. Nivel 3. Tabla M\_Comper.

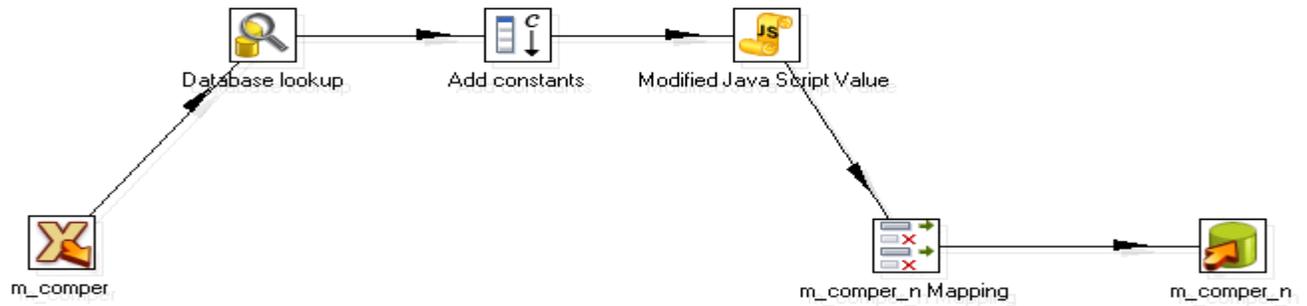


Figura 36. Esquema de la transformación Tabla M\_Comper.

Anexo 5: Mapeo de datos: Tabla M\_Percue.

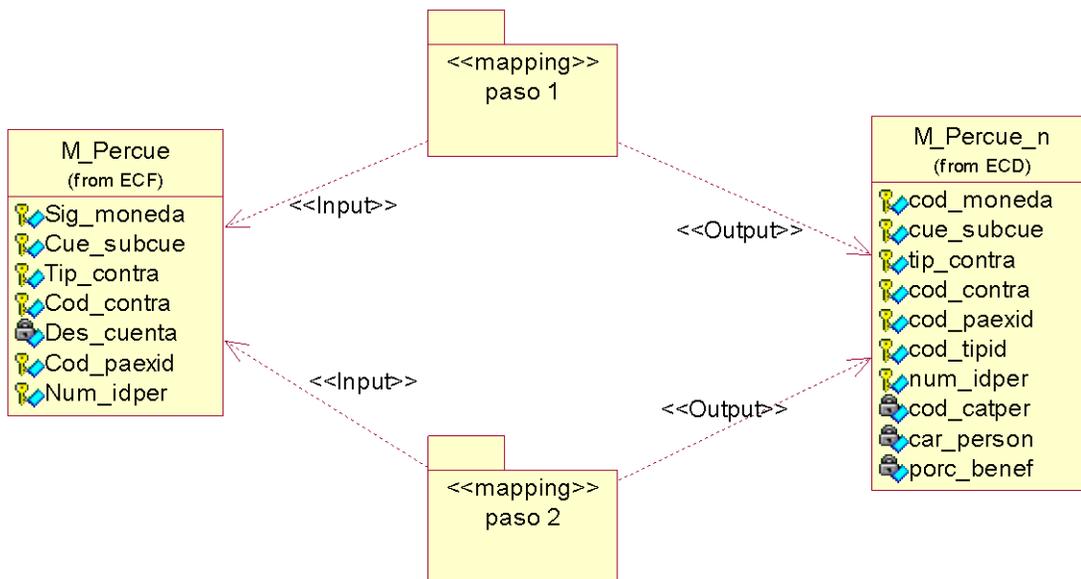


Figura 37. Mapeo de datos. Nivel 2. Tabla M\_Percue.

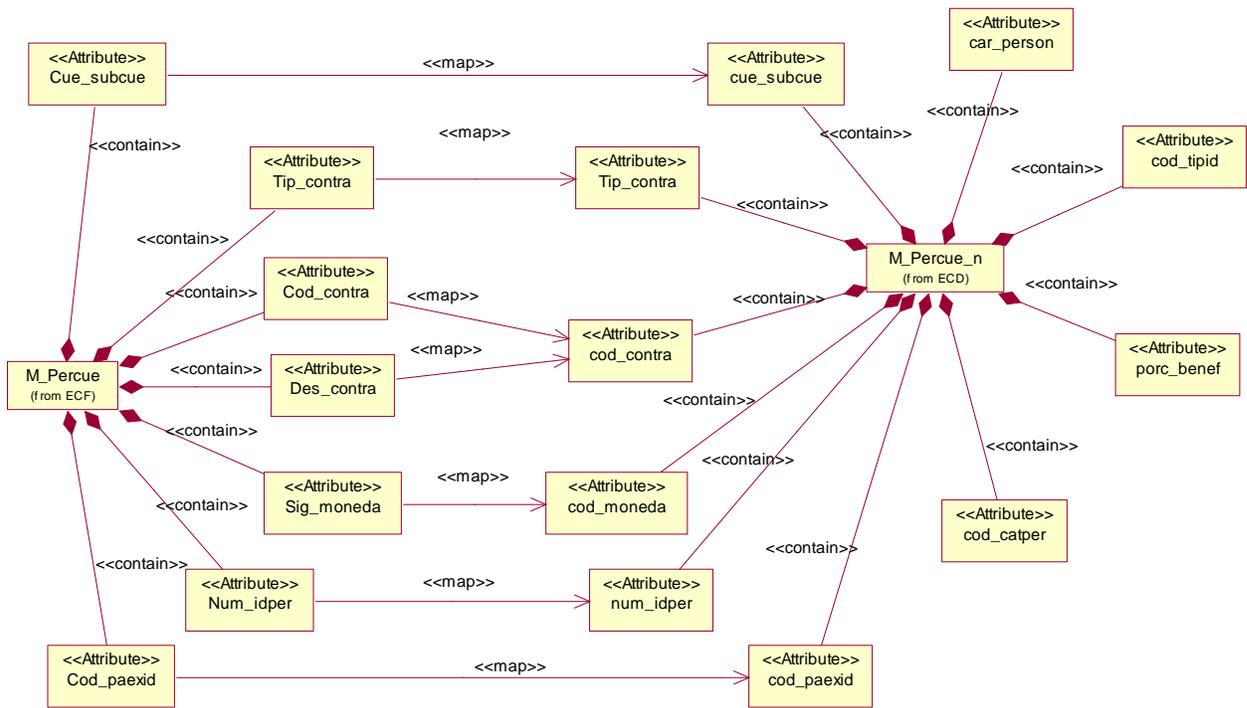


Figura 38. Mapeo de datos. Nivel 3. Tabla M\_Percue (Paso1).

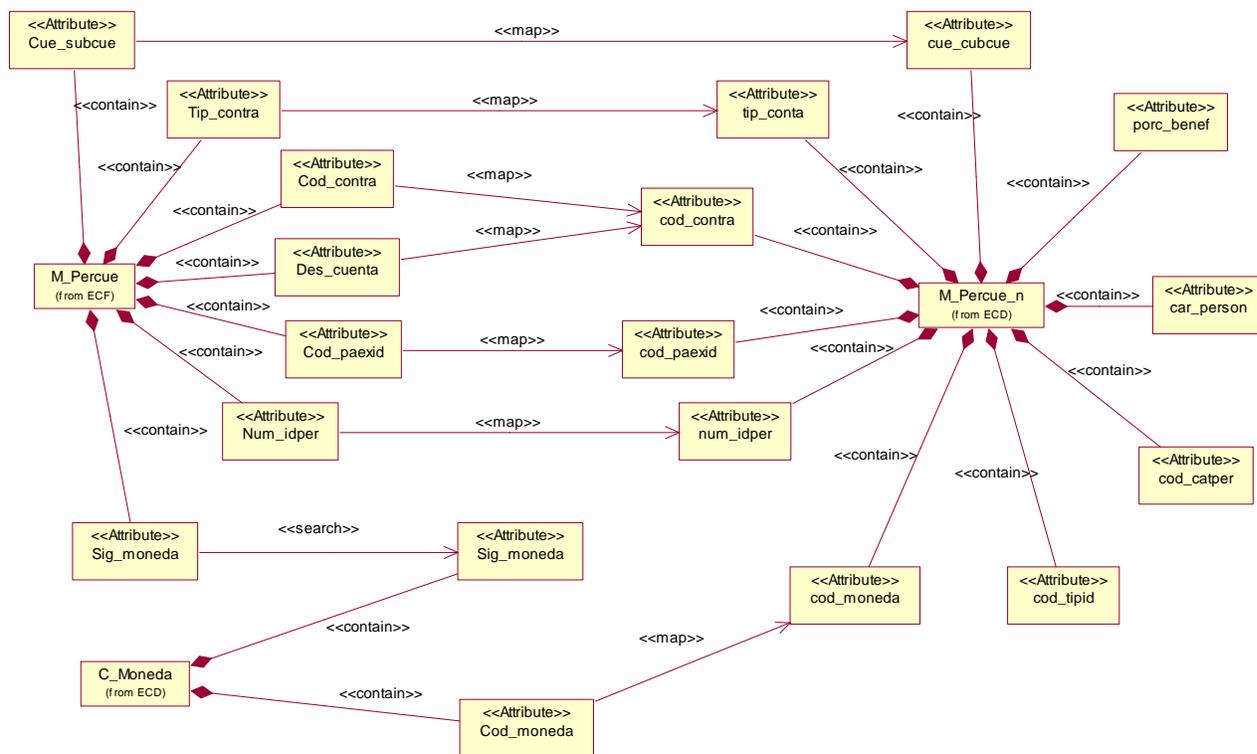


Figura 39. Mapeo de datos. Nivel 3. Tabla M\_Percue (Paso2).

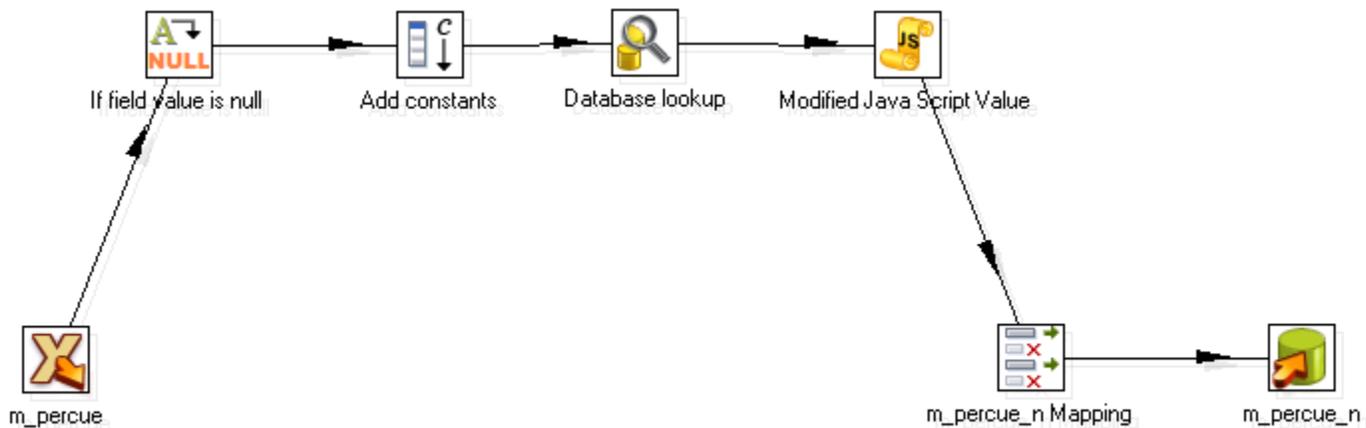


Figura 40. Esquema de la transformación Tabla M\_Percue.

## Anexo 6: Mapeo de datos: Tabla M\_Mayor.

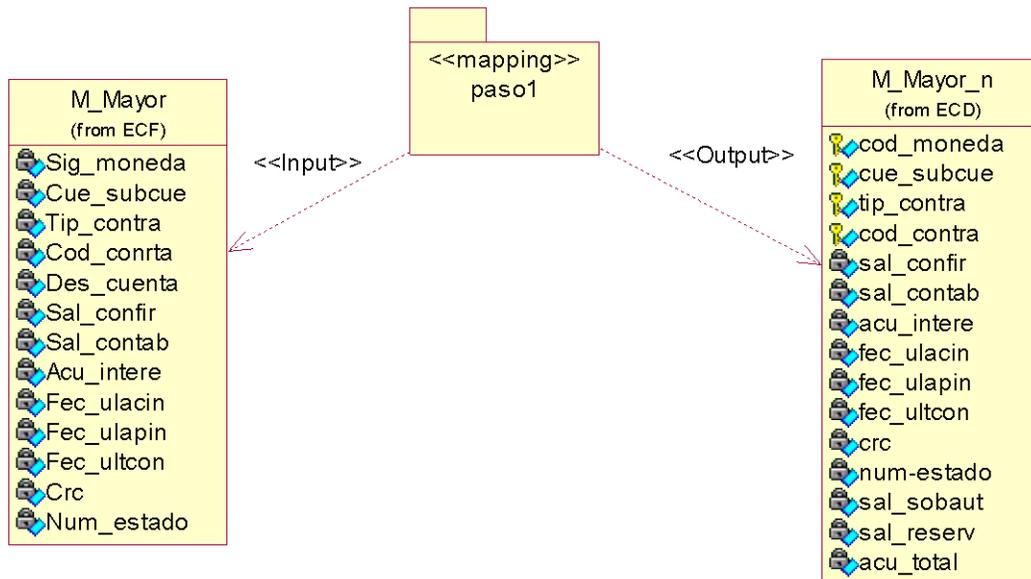


Figura 41. Mapeo de datos. Nivel 2. Tabla M\_Mayor.

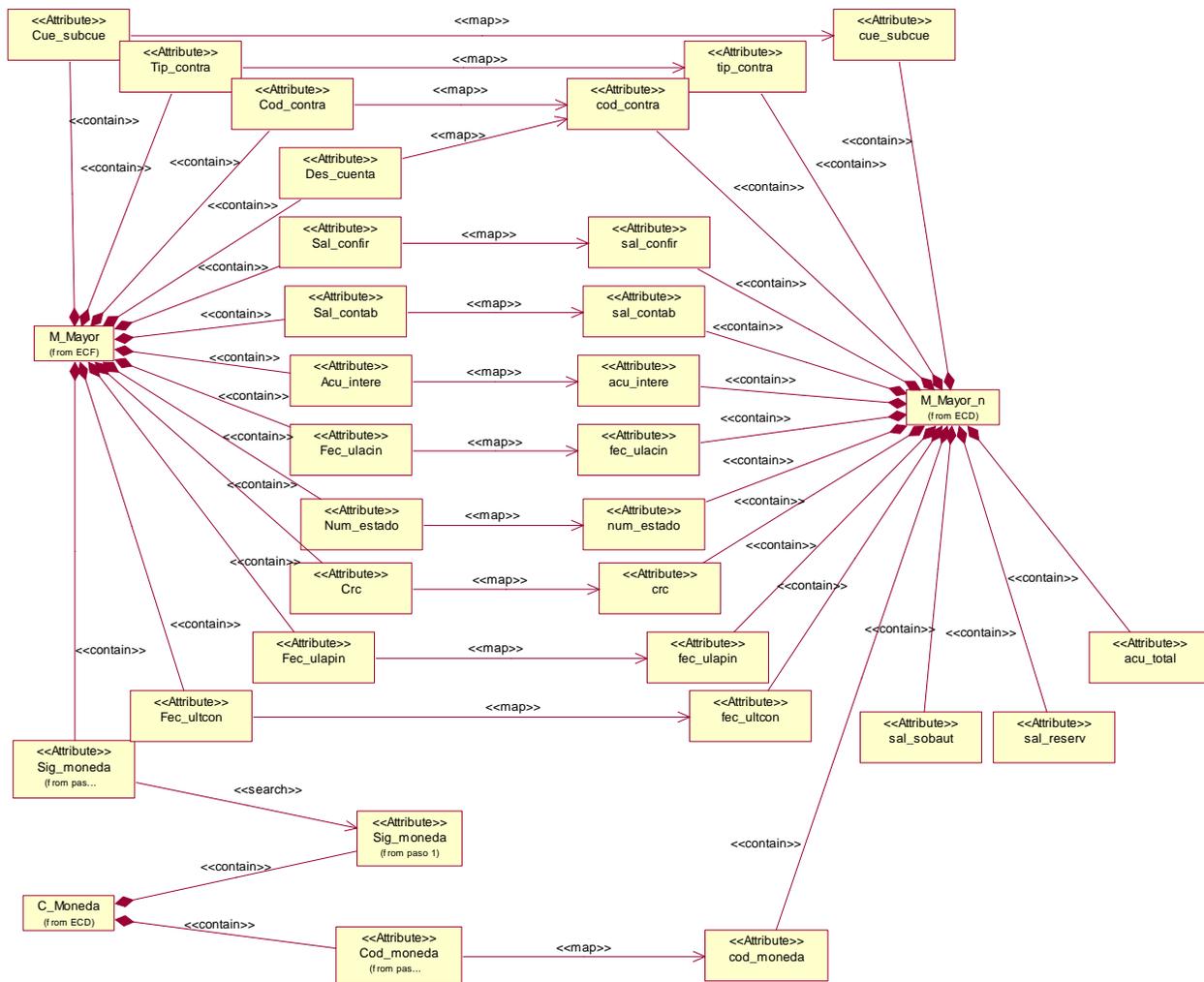


Figura 42. Mapeo de datos. Nivel 3. Tabla M\_Mayor.

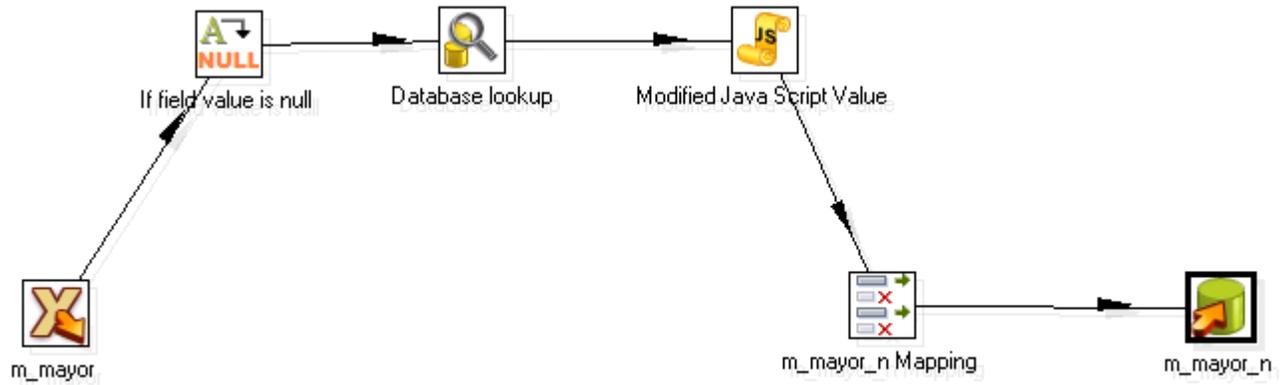


Figura 43. Esquema de la transformación Tabla M\_Mayor.

Anexo 7: Mapeo de datos: Tabla M\_Tasint.

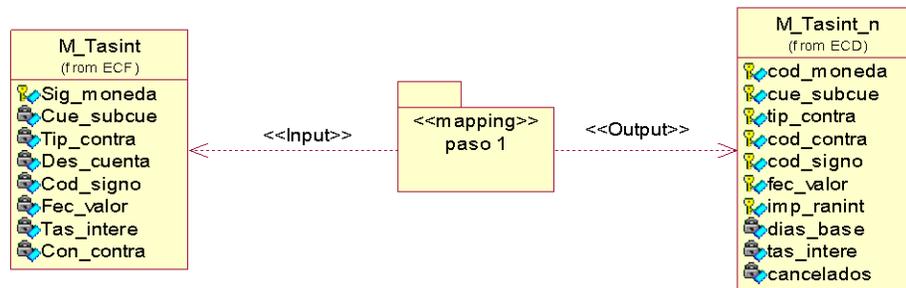


Figura 44. Mapeo de datos. Nivel 2. Tabla M\_Tasint.

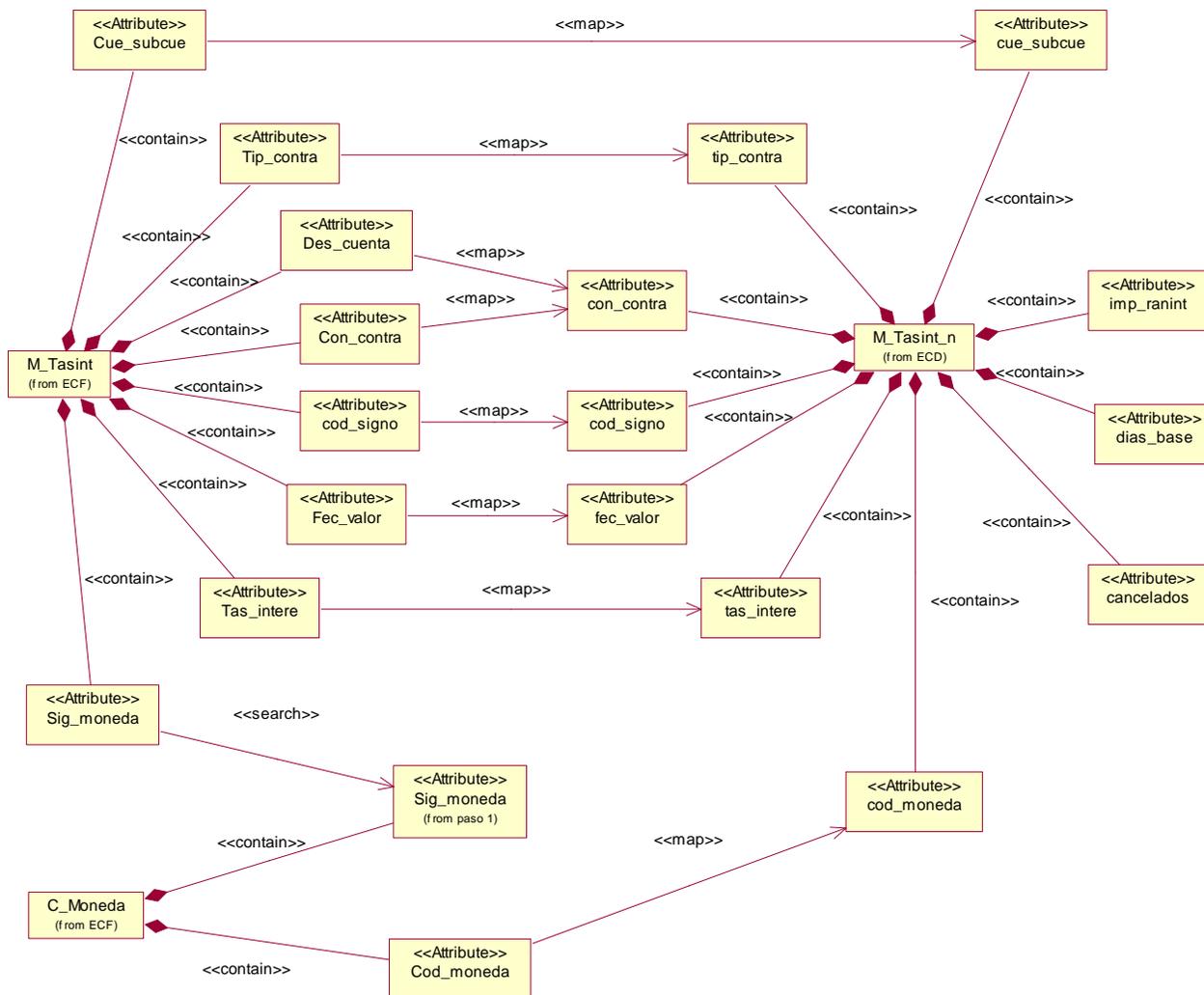


Figura 45. Mapeo de datos. Nivel 3. Tabla M\_Tasint.

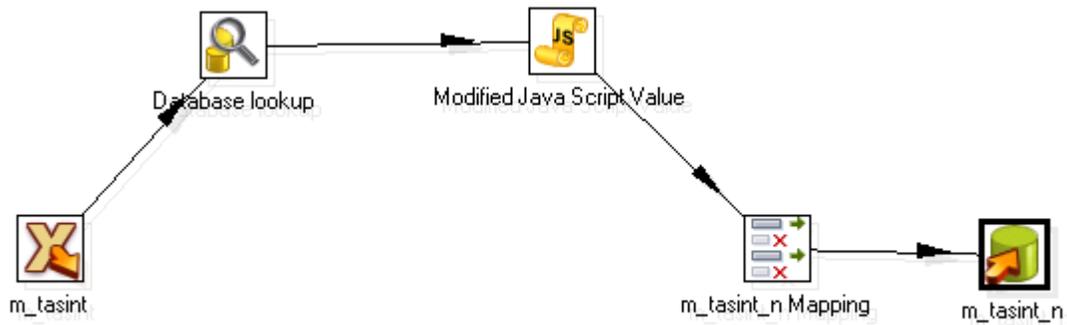


Figura 46. Esquema de la transformación Tabla M\_Tasint.

Anexo 8: Mapeo de datos: Tabla M\_Diario.

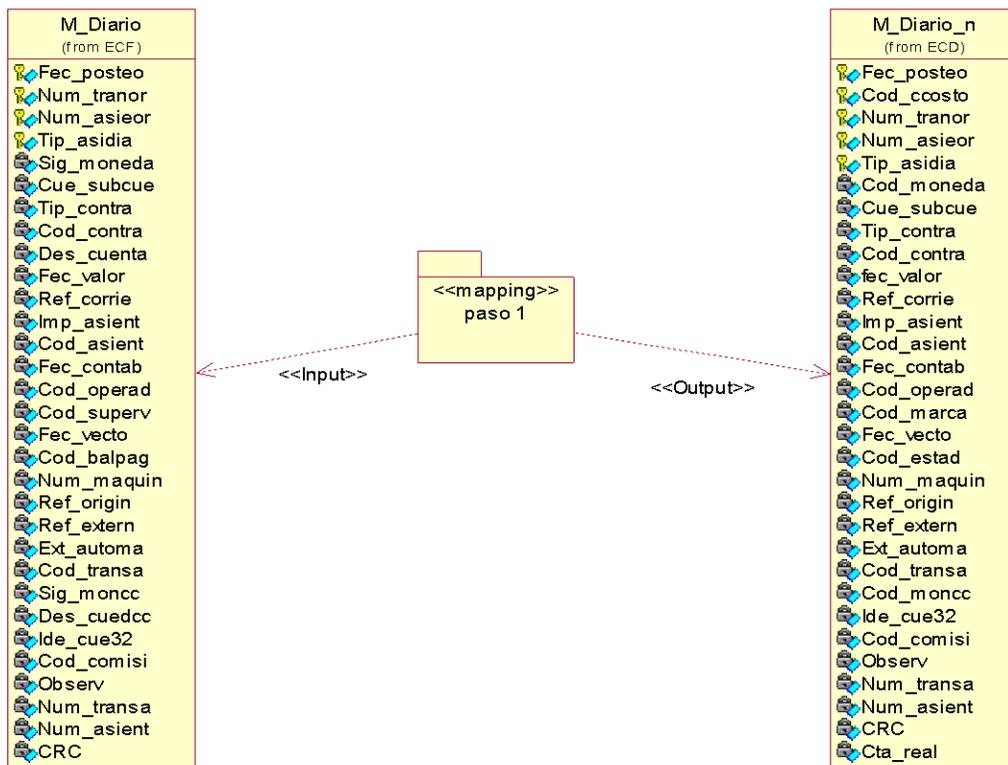


Figura 47. Mapeo de datos. Nivel 2. Tabla M\_Diario.

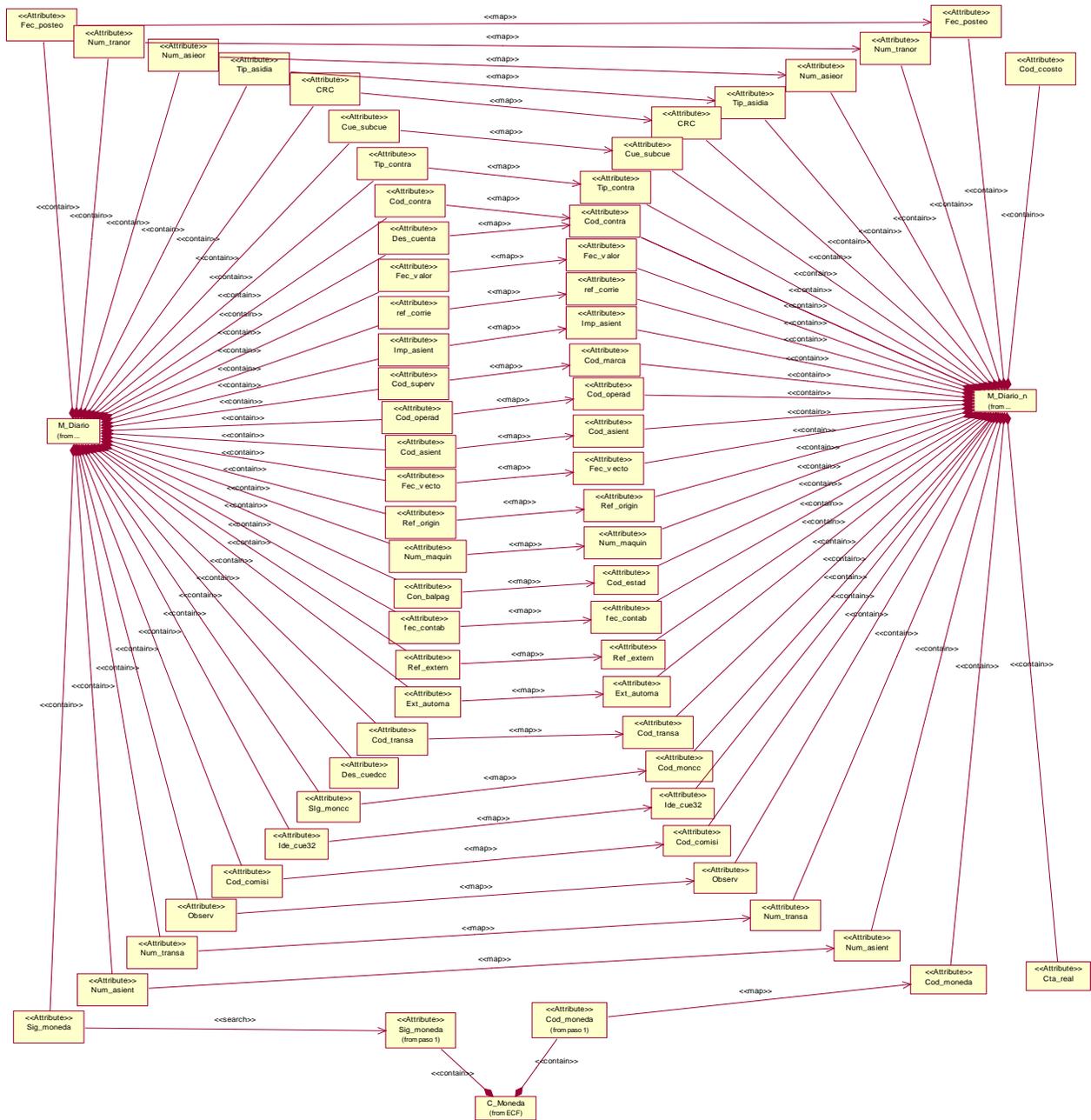


Figura 48. Mapeo de datos. Nivel 3. Tabla M\_Diario.

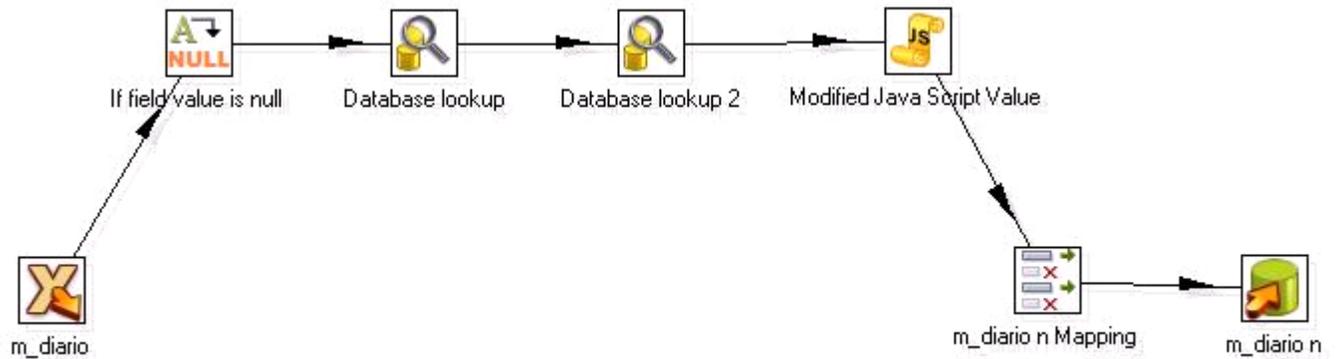


Figura 49. Esquema de la transformación Tabla M\_Diario.

Anexo 9: Mapeo de datos: Tabla M\_Histor.

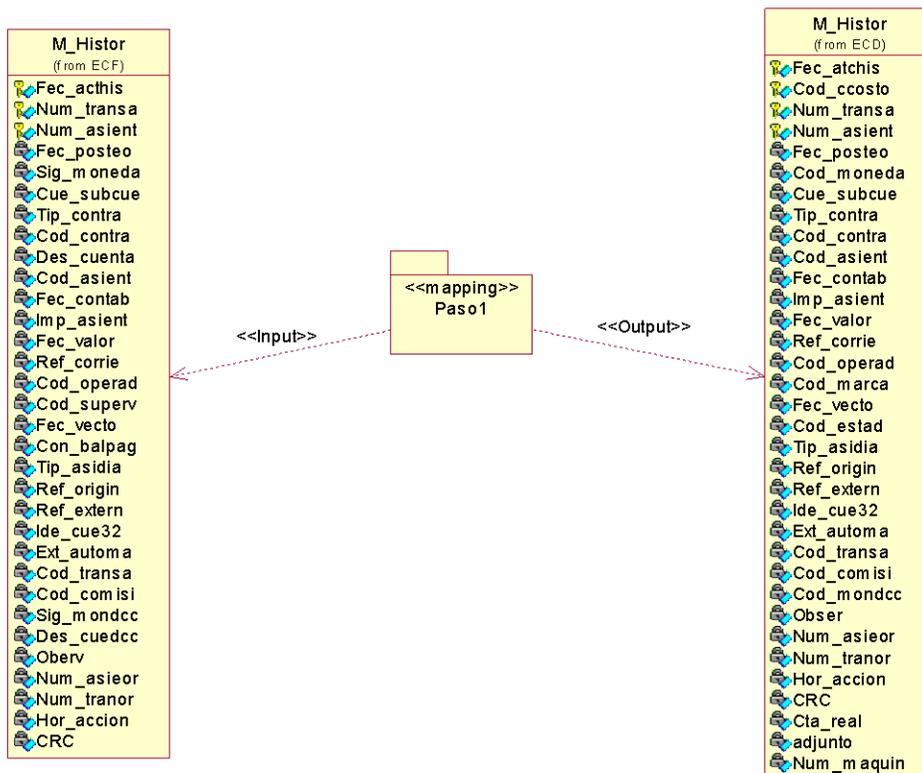


Figura 50. Mapeo de datos. Nivel 2. Tabla M\_Histor.



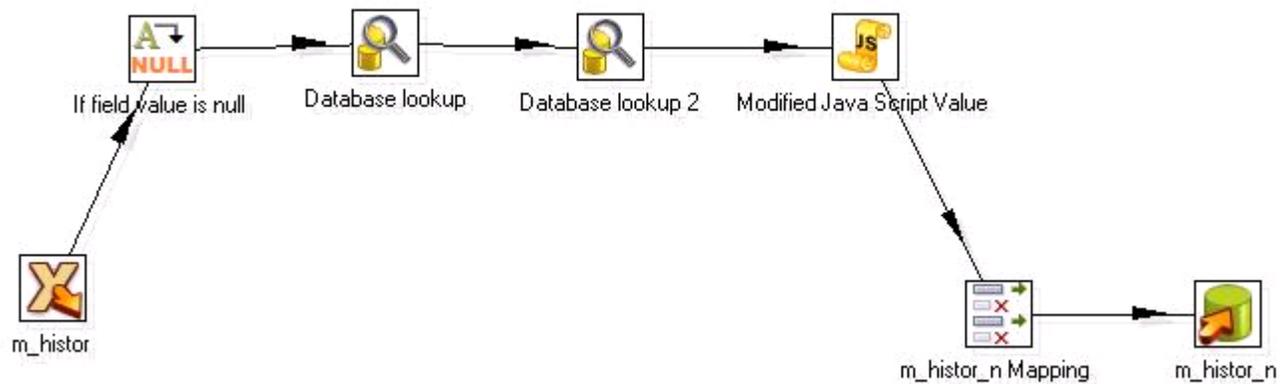


Figura 52. Esquema de la transformación Tabla M\_Histor.

# GLOSARIO DE TÉRMINOS

### ✓ Base de Datos

Base de Datos es un conjunto exhaustivo no redundante de datos estructurados organizados independientemente de su utilización y su implementación en máquina accesibles en tiempo real y compatibles con usuarios concurrentes con necesidad de información diferente y no predicable en tiempo.

### ✓ Entidad

Cualquier tipo de objeto o concepto sobre el que se recoge información: cosa, persona, concepto abstracto o suceso. Por ejemplo: coches, casas, empleados, clientes, empresas, oficios, diseños de productos, conciertos, excursiones, etc. Las entidades se representan gráficamente mediante rectángulos y su nombre aparece en el interior. Un nombre de entidad sólo puede aparecer una vez en el esquema conceptual. Hay dos tipos de entidades: fuertes y débiles. Una entidad débil es una entidad cuya existencia depende de la existencia de otra entidad. Una entidad fuerte es una entidad que no es débil.

### ✓ Atributo

Un atributo de una entidad es una característica interesante sobre ella, es decir, representa alguna propiedad que nos interesa almacenar.

### ✓ Tupla

Una ocurrencia de artículo o tupla consiste en un grupo de ocurrencias de campos relacionados, representando una asociación entre ellos.

### ✓ Cardinalidad

Las asociaciones son distinguidas por el número de posibles relaciones que una entidad determinada puede tener sobre esta. Este rango es llamado cardinalidad y es especificado por la cantidad mínima y máxima de instancias de la asociación. Típicamente la cota inferior es 0 o 1 y la superior 1 o n (muchos), esta forma de indicar la cardinalidad es llamada notación (mín., máx.) y es representada en las líneas que conectan los rombos y los rectángulos; que representan las asociaciones y los conjuntos de entidades respectivamente.

### ✓ Llave

Llave o clave: un atributo o conjunto de atributos de un artículo que define que cada ocurrencia de artículo de la base de datos sea único. En principio, cada artículo tiene una llave, ya que se tiene como hipótesis que cada elemento u ocurrencia del artículo es diferente de las demás. Por ejemplo, número de identidad

del trabajador.

✓ **Interrelaciones:**

Las interrelaciones son asociaciones o conexiones que existen entre dos o más entidades. Las entidades relacionadas pueden pertenecer al mismo o a distintos conjuntos de entidades. Una interrelación puede ser vista como una relación matemática entre  $n$  entidades cada una tomada de un conjunto de entidades, no necesariamente diferentes, donde cada  $e_i$  es una instancia del conjunto de entidades  $E_i$ . Cada tupla de la forma  $[e_1, e_2, \dots, e_n]$  es una instancia de la interrelación.

✓ **Fichero:**

Un fichero o archivo o conjunto de datos puede ser definido como un conjunto de ocurrencias de un mismo tipo de artículo. En la práctica, a menudo interesan las colecciones o conjuntos de objetos similares, necesitándose almacenar la información de las mismas propiedades para cada uno de ellos, por ejemplo, el conjunto de profesores de la Facultad.

✓ **Dominio:**

Para cada atributo existe un conjunto de valores posibles o permitidos, este conjunto de valores es llamado dominio o conjunto de valores de un atributo.

✓ **Trigger:**

Un trigger (o disparador) en una base de datos es un procedimiento que se ejecuta cuando se cumple una condición establecida al realizar una operación de inserción (INSERT), actualización (UPDATE) o borrado (DELETE).

✓ **Log:**

Un log es un registro de actividad de un sistema, que generalmente se guarda en un fichero de texto, al que se le van añadiendo líneas a medida que se realizan acciones sobre el sistema.