



Universidad de las Ciencias Informáticas

Facultad 4

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

Título: Migración de Objetos de Aprendizaje entre  
versiones de XAUCE CRODA

Autor:

Hugo Cartaya Pardo

Tutores:

Ing. Sandy Nuñez Padrón

Ing. Yaillet Martínez Pérez

La Habana, 2016 “Año del 58 de la Revolución”

## Declaración de autoría

Declaro que soy el único autor del trabajo “Migración de Objetos de Aprendizaje entre versiones de XAUCE CRODA” y autorizo a la Facultad 4 de la Universidad de las Ciencias Informáticas a hacer uso del mismo en su beneficio.

Para que así conste firmo la presente a los \_\_\_\_ días del mes de \_\_\_\_\_ del año \_\_\_\_\_.

---

**Autor**

Hugo Cartaya Pardo

---

**Tutor**

Ing. Sandy Nuñez Padrón

---

**Tutor**

Ing. Yaillet Martínez Pérez

## Agradecimientos

Primero que todo agradecer al jurado, a mi oponente, a mis tutores, de forma especial a Sandy, que más que mi tutor ha sido mi compañero de tesis todo este tiempo. A todos mis profesores a lo largo de la carrera, pero quiero resaltar a Leonardo Boza, Irán Pantaleón, Yudanis Gago, Arlenis Palmero, Yuleisy González, Rigoberto David García, Iván Pérez Mallea, Arcel Labrada y Sandra de la Caridad Rodríguez.

A mis compañeros de aula, a los que están y los que estuvieron, a mis amigos y amistades de la facultad, a mis amigos de fiestas, travesuras y estudio de estos últimos años: Julio, Monedero, Ricardo, Gustavo y Yisenia.

Agradecer también a mis amigos y amistades de la Universidad. Especialmente a Danailys, Johana, Lien, Yunet a mi gente del voly de la 1 y a mi gente de Amor a Segunda Vista.

Agradecer a mis amigos y amistades de fuera de la universidad. A Oscar, a Juan Carlos por su ayuda incondicional con mi tesis, a mis viejos amigos de México, que a pesar de la distancia están cerca.

Quiero agradecer a mi novia, porque le ha tocado la etapa más difícil de estos 5 años y ha respondido con mucho amor y paciencia, a mi suegro, por aceptarme como uno más en la familia y a mi suegra que más que mi suegra, es mi amiga.

A mi otra familia, donde tengo una segunda mamá Yanny, un segundo papá Faby, otra abuela, tres hermanos (Ulises, Ian Carlos y Jessica) y como 500 primos.

A mi familia, mi extensa familia, mis tías y tíos (Dalia, Fide, Mundy, Dania, Hanais, Agosto e Iran), mis primos y primas (Dailin, Yaquelin, Daine, Michel, Sintia, Adrian y Claudia), A mi hermana linda y mi cuñado, a mi padre que siempre ha sido la meta de mi viaje y a mi madre que siempre ha estado ahí cuando la he necesitado. Además, quiero agradecer a 4 familiares que por desgracia no se encuentran junto a nosotros en este mundo, a mi tío Héctor, que fue como un padre para mí, en mi niñez, a mi abuela Iliá, por sus cuidados y preocupación todo el tiempo que estuvo con nosotros, a mi abuelo macho (como le decíamos de cariño), por todo lo que me enseñó y los muy buenos momentos que pasamos juntos, a mi abuela chole (como le decíamos cariñosamente), por su amor y por su paciencia con todos nosotros, a todos **MUCHAS GRACIA**, esta tesis es por y para ustedes.

## Dedicatoria

Le dedico mi tesis a mi familia, a mis amigos y a mi novia, espero se sientan orgullosos.



“Los sueños parecen al principio imposible, luego improbables, y luego, cuando nos comprometemos, se vuelven inevitables.”

Mahatma Gandni

## Resumen

Las Tecnologías de la Información y las Comunicaciones avanzan de manera acelerada y con esto la información se ha convertido en un factor muy importante para toda la sociedad. La Universidad de las Ciencias Informáticas tiene en su estructura organizativa al Centro de Tecnologías para la Formación, el cual ha desarrollado la herramienta de autor XAUCE CRODA, actualmente en explotación su versión 2.0. Dicha versión cuenta con una estructura de almacenamiento, la cual contiene una base de datos en XML nativo (eXist), otra en PostgreSQL y una estructura de directorios. La relación entre las bases de datos causa demoras en el *software* al dar respuestas a solicitudes realizadas por los usuarios y eXist no es compatible con los sistemas operativos de 64 bit. Teniendo en cuenta lo anterior, se determinó desarrollar la versión 3.0 la cual cuenta con una base de datos en PostgreSQL y una estructura de directorios, pero el proyecto desea tener en dicha versión los objetos de aprendizaje almacenados en la versión anterior. Por esta razón se hace imprescindible migrar los datos necesarios para cumplir con el nuevo objetivo del proyecto. Esto provocó que fuese necesario realizar un estudio de las herramientas y procedimientos de migración existentes, la creación de una estrategia de migración, así como verificar la calidad de la solución propuesta.

## Tabla de contenido

Introducción .....	1
Capítulo 1: Fundamentación Teórica .....	6
1.1 Introducción.....	6
1.2 Características de XAUCE CRODA.....	6
1.3 Conceptos y términos asociados al problema de investigación .....	6
1.3.1 Objeto de aprendizaje .....	6
1.3.2 Herramienta de autor .....	7
1.3.3 Base de datos .....	7
1.3.4 Sistema Gestor de Bases de Datos .....	8
1.3.5 XML .....	9
1.3.6 Migración de datos.....	9
1.3.7Estrategia .....	11
1.3.8 Calidad de datos.....	11
1.4 Estudio de procedimientos de migración de datos.....	12
1.4.1 Migración integrada de datos para el Sistema Nacional de Salud.....	12
1.4.2 Migración de la información del Sistema de Gestión Académica, Akademos, al Sistema de Gestión Académica de Pregrado .....	13
1.4.3 Migración de datos de los sistemas corporativos de la Universidad de Chile .....	14
1.4.4 Aportes obtenidos de los procedimientos estudiados.....	14
1.5 Herramientas informáticas utilizadas en los procesos de migración de datos....	15
1.5.1 Pentaho Data Integration. ....	15
1.5.2 Softek Transparent Data Migration Facility UNIX (IP) .....	16
1.5.3 Herramienta para la migración de datos del Sistema de Inventario Participativo al ASSETS .....	16
1.5.4 AdventNetSwisSQL .....	17
1.5.5 Ora2Pg .....	17

1.5.6 Resultados del estudio de las herramientas de migración de datos .....	17
1.6 Herramientas y tecnologías a utilizar .....	18
1.6.1 Sistema Gestor de bases de datos PostgreSQL .....	18
1.6.2 Administrador de bases de datos PgAdmin.....	19
1.6.3 Herramienta para el modelado Visual Paradigm .....	19
1.6.4 Sistema operativo Ubuntu.....	20
1.7 Lenguajes utilizados .....	21
1.8 Conclusiones del capítulo .....	22
Capítulo 2. Descripción y análisis de la solución propuesta .....	23
2.1 Introducción.....	23
2.2 Propuesta de arquitectura para la migración de los datos .....	23
2.3 Migración de datos .....	24
2.3.1 Etapas definidas a utilizar en la estrategia de migración de datos .....	24
2.3.2. Técnicas de migración de datos.....	26
2.4 Proceso de extracción, transformación y carga (ETL) .....	27
2.4.1 Extracción .....	27
2.4.2 Transformación .....	28
2.4.3 Carga de los datos.....	30
2.5 Estrategia de migración propuesta. ....	30
2.6 Descripción de la solución propuesta .....	32
2.6.1 Desarrollo de la migración de la tabla tb_metadata de la base de datos en PostgreSQL del AF hacia las tablas oa y oauser de la base de datos en PostgreSQL del AD. ....	33
2.7 Conclusiones parciales.....	38
Capítulo 3. Validación de la solución propuesta.....	40
3.1 Introducción.....	40
3.2 Estrategia de validación y prueba.....	40
3.2.1 Pruebas realizadas a la simulación de carga .....	41
3.2.2 Pruebas de aceptación .....	43

3.3 Conclusiones parciales.....	44
Conclusiones .....	45
Recomendaciones .....	46
Referencias Bibliográficas .....	47

## Introducción

El cuidado de la información es y será considerado el principal activo de las empresas. Si las empresas pierden datos, pierden dinero, afectando su competitividad y participación en el mercado. Por esta razón, resulta de vital importancia que al momento de reemplazar los sistemas obsoletos por nuevas tecnologías, también se formalice el cuidado de los datos procesados en el sistema anterior. (Argudo, 2014)

En la actualidad, las Tecnologías de la Información y las Comunicaciones (TIC) representan un eslabón vital en el desarrollo y progreso de la sociedad, permitiendo la interactividad, interconexión y digitalización de la información. En este proceso ha influido de manera considerable el campo de la informática, la cual surgió por la necesidad social creciente de desarrollar nuevos métodos y medios para gestionar de forma eficiente la información, donde esta última juega un papel fundamental para el análisis en la toma de decisiones. (Gerard, 1994)

Con el objetivo de preservar la información digital, surgieron las estructuras de almacenamiento. Entre las cuales se encuentran, entre otras, los directorios y las bases de datos, término utilizado por primera vez en un simposio celebrado en California en 1963. Las bases de datos proveen de varias ventajas a las empresas, entre las más importantes se encuentran la globalización de la información y la compartición de información entre los distintos niveles de la organización. (Cohen & Asín, 2000)

Para gestionar las bases de datos se utilizan los Sistemas Gestores de Bases de Datos (SGBD), cuyo principal objetivo es "manejar de manera clara, sencilla y ordenada un conjunto de datos y servir de interfaz entre las bases de datos, el usuario y las aplicaciones. Los SGBD además tienen el fin de mejorar las respuestas a los clientes y obtener de manera organizada y rápida la información, logrando la independencia e integridad de los datos." (García, 1999)

Los SGBD brindan soporte para la evolución de los datos y permiten su manipulación de forma interactiva, pero no ofrecen un claro soporte a la migración de datos, siendo un problema frecuente al cual se enfrentan hoy en día disímiles empresas y organizaciones en todo el mundo. (López, 2004)

La literatura formal no ha identificado un proceso consecuente para la migración de datos. Reconocidos académicos y notables implementadores de sistemas, no han establecido un acercamiento formal sobre este tema, mucho más cuando la cantidad

de información, las exigencias del negocio y la tecnología, avanzan tan aceleradamente. (Coleman K., 2005)

La Universidad de las Ciencias Informáticas (UCI) tiene en su estructura productiva el Centro de Tecnologías para la Formación (FORTES), el cual desarrolla aplicaciones y herramientas vinculadas al proceso de enseñanza-aprendizaje. Tiene como uno de sus productos a la herramienta de autor XAUCE CRODA. Dicha aplicación permite la creación de objetos de aprendizaje (OA) siguiendo el estándar SCORM.

Entre sus principales funcionalidades se encuentran:

- Creación de plantilla.
- Creación de OA.
- Revisión de plantillas.
- Revisión de OA.
- Descripción de los OA según el estándar LOM.

Actualmente dicha herramienta está en explotación en su versión 2.0. Dicha versión cuenta con tres estructuras para el almacenamiento de datos, las cuales son: una base de datos en XML nativo (eXist), otra en PostgreSQL y una estructura de directorios.

El modelo interno de eXist usa documentos XML como la unidad elemental de almacenamiento. La base de datos en PostgreSQL utiliza el lenguaje SQL para realizar consultas a las tablas. Estas dos estructuras se encargan de almacenar los OA y sus metadatos, los datos de los usuarios y la información necesaria para que la aplicación funcione correctamente y en la estructura de los directorios se almacenan los objetos multimedia contenidos en los OA.

La relación entre las bases de datos causa demoras en el *software* para dar respuestas a solicitudes realizadas por los usuarios y la base de datos eXist no es compatible con sistemas operativos de 64 bit. Lo anteriormente expuesto reduce la posibilidad de actualización de la base tecnológica del proyecto y trae como consecuencia que el *software* quede obsoleto al no poder adaptarse a las nuevas tecnologías.

Es un hecho destacado que los sistemas informáticos deben evolucionar para adecuarse a los siempre cambiantes requisitos del entorno. Por esta razón se dedica mucho más esfuerzo y dinero a la evolución del *software* que la que se le destina a la creación y desarrollo del mismo. (Pressman, 2015)

Idealmente, a la hora de introducir una modificación en una aplicación informática se debe seguir un proceso en el cual sea identificada la razón del cambio, se analice cómo afecta el cambio a introducir en la aplicación, para posteriormente realizar los cambios donde sean necesarios, finalizando con una exhaustiva fase de pruebas que demuestre que la modificación no ha introducido errores en el *software*. Pero, muchas empresas simplifican el proceso anterior, reduciéndolo en la mayoría de los casos a un análisis de la modificación para posteriormente aplicarla directamente al código sin excesivas pruebas y menor documentación. (Pressman, 2015)

Para darle solución a las problemáticas presentes en la versión 2.0, el proyecto productivo decidió desarrollar la versión 3.0, la cual cuenta con una estructura de almacenamiento compuesta por una base de datos en PostgreSQL para almacenar los metadatos de los OA, los datos referentes a los usuarios y los datos necesarios para que la herramienta funcione correctamente y una estructura de directorios donde se almacenan los OA y los recursos multimedia utilizados en la confección de los mismos.

El proyecto desea tener en la nueva versión todos los OA creados y almacenados en la versión anterior, así como los datos de los usuarios que los crearon, los recursos utilizados y los datos necesarios para que el nuevo sistema interactúe con dichos OA, por lo que surge el problema que da inicio a esta investigación:

### **¿Cómo migrar los OA generados y almacenados en la versión 2.0 hacia la versión 3.0 de XAUCE CRODA?**

Objeto de estudio:

**Proceso de migración de bases de datos y directorios.**

Campo de acción:

**Proceso de migración de bases de datos y directorios de la Versión 2.0 a la 3.0 de XAUCE CRODA.**

Trazándose como objetivo principal de la presente investigación:

**Desarrollar una estrategia que permita migrar los OA generados y almacenados en la Versión 2.0 hacia la versión 3.0 de XAUCE CRODA.**

Los objetivos específicos son:

- Elaborar el marco teórico de la investigación mediante el estudio del estado del arte acerca de las tendencias tecnológicas actuales y las estrategias utilizadas en la migración de datos.

- Crear y poner en práctica, en la herramienta XAUCE CRODA, una estrategia de migración de datos.
- Realizar pruebas que verifiquen la efectividad y calidad del proceso de migración y el cumplimiento de los objetivos propuestos.

Para darle cumplimiento a los mismos se definieron las siguientes tareas a cumplir:

- Análisis y selección de las diferentes tecnologías utilizadas para la migración de bases de datos y directorios.
- Análisis y selección de las estrategias de trabajo existentes para la migración de bases de datos y directorios.
- Creación y diseño de la arquitectura de la migración.
- Creación y diseño del diagrama de fases.
- Creación de una estrategia de migración que pueda ser utilizada en la migración entre las estructuras de almacenamiento de las versiones 2.0 y 3.0 de XAUCE CRODA.
- Uso de la estrategia de migración creada en la migración entre las estructuras de almacenamiento de las versiones 2.0 y 3.0 de XAUCE CRODA.
- Realización de pruebas al proceso de migración y a la versión 3.0 de XAUCE CRODA.

#### **Preguntas científicas:**

- ¿Cómo se presenta en la literatura científica los procesos de migración de datos?
- ¿Cómo aplicar en la herramienta XAUCE CRODA los procesos de migración de datos?
- ¿Cómo se asegura la calidad del proceso de migración de datos entre las versiones de XAUCE CRODA?

Resultado esperado: **estrategia de migración que permita migrar los objetos de aprendizaje desarrollados en la versión 2.0 hacia la versión 3.0 de XAUCE CRODA.**

Los Métodos científicos que se utilizaron durante la investigación fueron:

Métodos Teóricos

-Histórico-lógico: para realizar un estudio de la trayectoria histórica, evolución y desarrollo de estrategias para la migración de bases de datos existentes en el ámbito nacional e internacional.

-Analítico-sintético: para analizar las características de las diferentes estrategias y tecnologías usadas en la migración de datos, tomando los elementos más importantes que permitan lograr un correcto diseño de la nueva estrategia.

#### Métodos Empíricos

-Observación: permitió estudiar de cerca el objeto de la investigación, las acciones, causas y consecuencias logrando conocer la esencia del problema planteado, analizando desde varios puntos de vista la propuesta de solución y otras soluciones existentes e identificando qué está hecho y qué falta por hacer.

El presente documento está estructurado de la siguiente manera:

Capítulo 1 Fundamentación Teórica de la estrategia para la migración de datos de la versión 2.0 hacia la versión 3.0 de XAUCE CRODA: este capítulo presenta algunos elementos teóricos necesarios que ofrecen soporte para la realización de todo el trabajo en general, tales como: el estado del arte, y la descripción y selección de las herramientas a utilizar.

Capítulo 2 Descripción e implementación de la estrategia para la migración de datos de la versión 2.0 hacia la versión 3.0 de XAUCE CRODA: en este capítulo se presenta la propuesta de solución, se muestran además cada uno de los artefactos generados y se aplica la estrategia de migración.

Capítulo 3 Confección y realización de pruebas al proceso de migración y a la versión 3.0 de la herramienta XAUCE CRODA: en este se presenta un plan de pruebas para verificar que el proceso de migración se realice de forma correcta y no afecte el buen desempeño de la versión 3.0 de la herramienta XAUCE CRODA.

El presente trabajo contiene además Conclusiones, Recomendaciones, Referencias Bibliográficas, Glosario de Términos y Anexos.

## **Capítulo 1: Fundamentación Teórica**

### **1.1 Introducción**

En el presente capítulo se precisan un conjunto de conceptos y fundamentos que constituyen el marco teórico relacionado con el objeto de estudio definido en la investigación. Se destacan las principales características de los elementos asociados a la temática a investigar, así como el entorno nacional e internacional donde se desarrollan. Se exponen las principales herramientas que sirven de apoyo en la búsqueda de la solución planteada.

### **1.2 Características de XAUCE CRODA**

XAUCE CRODA es una herramienta de autor web desarrollada en la universidad de las Ciencias Informáticas, una aplicación libre que permite crear objetos y diseños de aprendizaje estandarizados de forma individual o mediante la participación activa y simultánea de varios docentes. Proporciona diversos tipos de recursos mediáticos, plantillas, tipologías de ejercicios y herramientas que promueven el trabajo colaborativo. (UCI, 2012)

CRODA permite la creación de objetos de aprendizaje reutilizables, accesibles, duraderos e interoperables empleando el estándar SCORM. Brinda un conjunto de plantillas que facilitan, a los autores, la actividad de estructurar sus OA. Con vista a la reutilización de los mismos, todos sus elementos pueden ser descritos con metadatos siguiendo el estándar *Learning Object Metadata* (LOM, por sus siglas en inglés). (Tabasco, 2012)

### **1.3 Conceptos y términos asociados al problema de investigación**

Para el mejor entendimiento de la situación existente alrededor del problema en cuestión y definir algunos de los conceptos más utilizados en la investigación, se propone plantearlos formalmente según algunos autores y la posición del autor de esta investigación con respecto a dichos conceptos.

#### **1.3.1 Objeto de aprendizaje**

El término OA fue introducido por Wayne Hodgins en 1992. A partir de esa fecha, han sido muchos los autores que han definido el concepto; de hecho, la falta de consenso

en su definición ha llevado a la utilización de múltiples términos sinónimos: *learning object*, objeto de aprendizaje reutilizable, objeto de conocimiento reutilizable, cápsula de conocimiento, entre otros. (Instituto de Ciencia de la Educación, 2006)

David Willey, en el año 2001 propone la siguiente definición: “cualquier recurso digital que puede ser usado como soporte para el aprendizaje”. (Instituto de Ciencia de la Educación, 2006)

Los recursos digitales constan de 2 partes: los contenidos y la información asociada a los contenidos denominada “Metadato”, que permiten la catalogación digital de la información que contiene un OA y su reutilización en diversos contextos. (Islas, 2008)

Según Aretio Garcia son: “recursos digitales autocontenidos, diseñados para utilizarse en procesos de enseñanza y aprendizaje, y se caracterizan por la capacidad de reuso que contienen, apoyándose fuertemente en cuestiones de programación orientada a objetos y clasificación bibliotecológica” (Garcia, 2005)

Partiendo de las definiciones anteriores, para la presente investigación el autor define que un OA es un recurso educativo de fácil construcción que contiene metadatos.

### **1.3.2 Herramienta de autor**

Las herramientas de autor son herramientas de desarrollo de *software* que posibilitan a diseñadores instructivos, educadores, maestros y aprendices diseñar un recurso multimedia interactivo, así como ambientes de aprendizaje en hipermedia sin el conocimiento de lenguajes de programación. (Dabbagh, 2001)

Las herramientas de autor tienen como objetivo la creación de aplicaciones educativas que contengan generalmente, ya sea implícitamente o explícitamente, un modelo particular de la tarea en el que el usuario final debe estar ocupado, así como un modelo del proceso editorial mismo. (Dabbagh, 2001)

De manera general, para la presente investigación, basándose en los conceptos anteriormente planteados, el autor define que una herramienta de autor es una aplicación que permite de forma sencilla la creación de OA.

### **1.3.3 Base de datos**

Una base de datos es un conjunto de datos persistentes que es utilizado por los sistemas de aplicación de alguna empresa dada. (Date, 2001)

Colección o depósito de datos, donde los datos están lógicamente relacionados entre sí, tienen una definición y descripción común y están estructurados de una forma particular. Una base de datos es también un modelo del mundo real y, como tal, debe poder servir para toda una gama de usos y aplicaciones. (Arias, 2007)

De manera general, para la presente investigación, basándose en los conceptos anteriormente planteados, el autor define que una base de datos es un conjunto de datos organizados, que se encuentran agrupados, estructurados e interrelacionados entre sí.

#### **1.3.4 Sistema Gestor de Bases de Datos**

El *software* que permite la utilización y/o la actualización de los datos almacenados en una (o varias) base(s) de datos por uno o varios usuarios desde diferentes puntos de vista y a la vez, se denomina sistema de gestión de bases de datos. (Matos, 1999)

Es una colección de programas cuyo objetivo es servir de interfaz entre la base de datos, el usuario y las aplicaciones. El objetivo fundamental de un SGBD consiste en suministrar al usuario las herramientas que le permitan manipular, en términos abstractos, los datos; o sea, de forma que no le sea necesario conocer el modo de almacenamiento de los datos en la computadora, ni el método de acceso empleado. (Matos, 1999)

SGBD libres:

- MariaDB
- PostgreSQL
- FireBird
- SQLite

SGBD propietarios:

- Oracle (Oracle Corp.)
- DB2 (IBM)
- InterBase (CodeGear)
- MaxDB (SAP)
- JavaDB (SUN Microsystems)

- SQL Server (Microsoft Corp.)
- Infomix (IBM)

De manera general, para la presente investigación, partiendo de los conceptos anteriormente planteados, el autor define un SGBD como una aplicación que permite a los usuarios definir, crear y mantener una base de datos y proporcionar un acceso controlado a la misma, asegurando su integridad, confidencialidad y seguridad.

### **1.3.5 XML**

Este estándar es un metalenguaje que puede ser utilizado para describir la estructura lógica y el contenido de una gran variedad de documentos, y puede ser adaptado para satisfacer una gran cantidad de aplicaciones. Además, XML no permite ninguna desviación de la sintaxis estándar. Esto asegura que todos los documentos codificados en XML puedan ser editados, almacenados y transmitidos sin importar el analizador de XML del receptor. (XML, 2015)

Un documento XML tiene dos estructuras, una lógica y otra física. Físicamente, el documento está compuesto por unidades llamadas entidades. Una entidad puede hacer referencia a otra entidad, causando que esta se incluya en el documento. Cada documento comienza con una entidad documento, también llamada "raíz". El documento está compuesto de declaraciones, elementos, comentarios, referencias a caracteres e instrucciones de procesamiento, todos los cuales están indicados por una marca explícita. (XML, 2015)

De manera general, para la presente investigación, basándose en los conceptos anteriormente planteados, el autor define que XML es un metalenguaje de sintaxis estándar, que contiene una raíz compuesta por declaraciones, elementos, comentarios, referencias e instrucciones.

### **1.3.6 Migración de datos**

La migración de datos constituye el proceso por el cual grandes volúmenes de datos son trasladados desde sistemas existentes hacia nuevos sistemas. (Lau, 2007)

Durante la migración se realizan procesos de extracción, transformación y carga, los cuales permiten obtener los datos desde su origen, modificarlos para cumplir con la integridad y la consistencia e insertarlos finalmente en la base de datos destino. (Lau, 2007)

Existen varias etapas a seguir en todo proceso de migración de datos entre las cuales se destacan las siguientes (Herrera, 2007):

- Levantamiento de información: contempla la definición de la información que requiere el nuevo sistema y se identifica en la fuente de origen.
- Análisis de Calidad: se verifica la calidad de los datos, con el propósito de establecer los tiempos reales de las etapas de transformación, limpieza y mapeo de datos.
- Transformación/Mapeo de datos: contempla la definición de los diccionarios con las reglas de transformación y el mapeo de campos entre las fuentes de origen y destino.
- Limpieza de datos: contempla la depuración de los datos en las fuentes de origen.
- Empleo del proceso de extracción, transformación y carga (ETL por sus siglas en inglés): se realiza el empleo de extracción, transformación y carga de los datos.
- Pruebas: se realizan las pruebas funcionales de la migración, determinando que la información sea traspasada con éxito al nuevo sistema.
- Cargas reales: se realiza la migración real de los datos desde el antiguo sistema al nuevo sistema.
- Revisión de resultados: se realiza la revisión final de la información cargada al nuevo sistema.

Para lograr una migración exitosa hay que tener en cuenta el uso de determinadas técnicas, entre las cuales se encuentran (Zaragoza, 2010):

- Planeación: este proceso de planeación y análisis del trabajo puede llevarse algún tiempo adicional, pero esta fase es la de mayor importancia para la migración de datos, repercutiendo en el desarrollo exitoso de la misma.
- Contador de registros: si la migración se realiza de forma manual mediante alguna consulta de inserción es recomendable inicializar un contador para cada registro insertado con éxito y otro para los no insertados, así obviamente la suma de ambos debe ser igual a los registros originales.

- Mapeado de tipos de datos: existen determinadas plataformas que no soportan algunos tipos de datos, así que es necesario planificar el mapeo de los campos en la nueva base de datos.
- Restricciones y disparadores: antes de iniciar la migración de la base de datos, es recomendable deshabilitar los *triggers* y restricciones que puedan generar error al momento que el SGBD ejecute el proceso de escritura de los datos.
- Codificación de caracteres: cuando el copiado se efectúa de forma automática, es preciso identificar la codificación de caracteres que la base de datos destino espera, pues así se impide la sustitución automática de caracteres o en su caso, pérdida de los mismos.

De manera general, para la presente investigación, basándose en los conceptos anteriormente planteados, el autor define que migración de bases de datos es un proceso organizado en el cual se trasladan datos de una o varias fuentes origen hacia una o más fuentes destino.

### **1.3.7 Estrategia**

Según Henry Mintzberg el término estrategia no es más que “el patrón de una serie de acciones que ocurren en el tiempo”. (Mintzberg, 1997)

La estrategia es un plan, una especie de curso de acción conscientemente determinado, una guía para abordar una situación específica. (Cruz, 2006)

Teniendo en cuenta los conceptos anteriormente expuestos el autor define para esta investigación que una estrategia es proceso seleccionado a través del cual se prevé alcanzar un cierto estado futuro.

### **1.3.8 Calidad de datos**

La mala calidad de los datos influye de manera muy significativa y profunda en la efectividad y eficiencia de las organizaciones, así como en todo el negocio, llevando en algunos casos a pérdidas multimillonarias. Cada día se hace más notoria la importancia y necesidad en distintos contextos de un nivel de calidad adecuado para los datos. (Marotta, 2009)

La calidad de los datos es fundamental para toda migración pues determina el nivel de exactitud y estandarización en que se encuentra la información. Para esto es preciso realizar un perfilamiento de datos, donde se estudia su contenido, estructura, calidad y

dependencias, y de ahí obtener estadísticas e información de los mismos. (Marotta, 2009)

Este proceso es muy importante pues permite conocer el tipo de información que contiene cada fuente y campo, además de detectar todos aquellos formatos, registros nulos, valores indebidos, datos ausentes o duplicados, y definir cómo tratar cada una de las casuísticas que van apareciendo. (Marotta, 2009)

### ***1.4 Estudio de procedimientos de migración de datos***

Realizar un estudio de procedimientos que apoyan la migración de datos tiene como objetivo principal tomar experiencias en cuanto al funcionamiento de los procesos que llevan a cabo los sistemas actuales en el mundo, y al mismo tiempo analizar en qué medida los procedimientos encontrados ayudan en la solución de la problemática planteada.

#### **1.4.1 Migración integrada de datos para el Sistema Nacional de Salud**

El Centro Nacional de Genética Médica (CNGM) tiene dentro de sus funciones las investigaciones de la genética médica. En la actualidad la Genética Médica enfrenta un grupo de dificultades asociado al crecimiento del volumen de información que generan los estudios genéticos y el descubrimiento vertiginoso del genoma humano, y en este sentido Cuba a través del CNGM ha desarrollado diversos estudios en consonancia con esta tendencia mundial. En este centro el volumen de datos histórico es valioso y no debe perderse por lo que se llevó a cabo una migración de los datos al Sistema Informático de Genética Médica (alasMEDIGEN), que resuelve entre otros problemas el intercambio de información entre los genetistas. Para esto se puso en práctica el desarrollo de una propuesta del proceso de migración de los datos que permitió que el sistema nuevo funcione reutilizando la información del sistema origen. (Dominguez, 2011)

A continuación, se describen algunos de los procesos empleados (Dominguez, 2011):

- Se revisó el diseño de la estructura de las tablas.
- Se revisó si el diseño de la estructura de las tablas origen se podían migrar al diseño de la estructura de las tablas nuevas, elaborándose un documento donde se detalló por cada tabla origen y campo si era posible migrarlo, y se detalló la tabla destino con los campos que contiene.

- Se revisó la calidad de la información de las tablas viejas, para desechar la información que no fuese valiosa, y no era necesario migrar.
- Se realizó un script para el traspaso de la información de una base de datos a otra.
- Se diseñaron nuevas tablas para poder realizar el traspaso de los datos de unas tablas a otras.

#### **1.4.2 Migración de la información del Sistema de Gestión Académica, Akademos, al Sistema de Gestión Académica de Pregrado**

Akademos contiene información de gran importancia que constituye el grosor histórico de la universidad, pues almacena todos los datos académicos de los estudiantes que han cursado en la misma. En la UCI se realizó una migración de toda la información contenida en la base de datos de Akademos para el nuevo modelo de datos del Sistema de Gestión Académica de Pregrado, debido a que con el surgimiento del mismo, era necesario que éste contara con todos los datos referente a los estudiantes que se encontraban almacenados en la base de datos del anterior Sistema de Gestión Académica, garantizando así la continuidad de la gestión de los procesos involucrados en la formación docente en la universidad. (Lamas, 2010)

Para llevar a cabo esta migración de Akademos al Sistema de Gestión Académica de Pregrado se pusieron en marcha una serie de fases de la migración de datos, para luego poder aplicar los procedimientos o estrategias establecidas para llegar al resultado final. (Lamas, 2010)

En la estrategia para la migración de los datos se realizó (Lamas, 2010):

- Un análisis de los sistemas.
- La descripción de los campos con los que contaba cada sistema.
- La migración haciendo uso de una herramienta de migración de datos.
- Una descripción de la cantidad de datos trasladada de un sistema a otro.
- Confección de reportes de los datos.

### **1.4.3 Migración de datos de los sistemas corporativos de la Universidad de Chile**

La migración de datos de aplicaciones y bases de datos antiguas a nuevas es un subproducto necesario del crecimiento organizacional y de la adaptación a las cambiantes necesidades de negocios. Pero el mezclar datos con diferentes formatos, estructuras y parámetros, continúa siendo una de las prácticas más riesgosas de las áreas de Tecnología de Información, resultando frecuentemente en grandes retrasos, costos elevados más de lo pronosticado e inclusive en fallas completas del proyecto. *Innovative* utiliza el sistema de perfilamiento (“*profiling*”) automático de datos y las soluciones de principio-a-fin para la integración y manejo de calidad de datos. (Universidad de Chile, 2012)

En este proceso se llevaron a cabo los siguientes pasos (Universidad de Chile, 2012):

- Se realiza una inspección a los datos para identificar formatos inconsistentes, estructuras de datos dañadas, información faltante y otras anomalías que corrompen la calidad de los mismos.
- Usando los resultados del perfilamiento de datos, se escriben especificaciones de mapeo para transformar los datos que ayudarán a asegurar una conversión exitosa.
- Conducir una prueba de conversión usando una muestra de datos estadísticamente significativa para confirmar que las especificaciones de mapeo están correctas antes de aplicarlas al universo de los datos.
- Limpiar, estandarizar, identificar clientes duplicados y enlazar clientes relacionados durante el proceso de transformación para maximizar la exactitud y eliminar duplicidades.
- Implementar procedimientos para monitorear y mejorar la calidad de datos en una forma continua.

### **1.4.4 Aportes obtenidos de los procedimientos estudiados**

Después de realizado un estudio de estos procedimientos, se arribó a la conclusión de que existe una tendencia común en la forma de trabajo para llevar a cabo una migración de datos, en la que se realizan una serie de pasos que guían dicho proceso. Cada uno de ellos fue exitoso en su objetivo, sin embargo, la problemática a la que dan solución no es la misma que la existente en la presente investigación. Debido a

que, en dichos procesos, la migración se realiza solo entre bases de datos relacionales y la migración a realizar, entre las versiones de XAUCE CRODA, se hace entre bases de datos relacionales (las bases de datos en PostgreSQL), bases de datos no relacionales (la base de datos eXist) y estructuras de directorios. Es por esto que sólo se tendrán en cuenta los aspectos positivos que ayuden a diseñar una estrategia de migración que permita resolver el problema en cuestión.

## ***1.5 Herramientas informáticas utilizadas en los procesos de migración de datos***

Existen muchas herramientas informáticas que sirven de apoyo a la migración de datos y posibilitan el diseño, creación e implementación del proceso ETL. A continuación, se describen algunas de las más usadas.

### **1.5.1 Pentaho Data Integration.**

Esta herramienta está compuesta de 4 herramientas principales (Kettle, 2007):

- *Spoon*, donde se diseñan las transformaciones de forma gráfica.
- *Pan*, donde se ejecutan las transformaciones diseñadas en *Spoon*. Esto se conoce como “*jobs*” (trabajos) y crea dependencia entre dichas transformaciones.
- *Chef*, el cual permite diseñar la carga de datos a través de una interfaz gráfica. Incluye el control de estado de los trabajos.
- *Kitchen*, el cual permite ejecutar los trabajos *batch* diseñados con *Chef*.

Es una de las herramientas libres más antiguas, tiene una gran cantidad de usuarios y una nueva dirección por parte del soporte técnico de Pentaho 4. Permite implementar los procesos de extracción, transformación y carga de datos (ETL). El uso de Pentaho *Data Integration* posibilita evitar grandes cargas de trabajo manual frecuentemente difícil de mantener y de desplegar. Con Pentaho se pueden realizar diversas tareas, entre ellas cabe resaltar (Kettle, 2007):

- Exportar de bases de datos a ficheros u otras bases de datos.
- Importar en bases de datos ficheros en formato Excel o texto.
- Migración de datos entre diferentes bases de datos.
- Explotación de los datos existentes en bases de datos (tablas, vistas).

- Limpieza de datos aplicando transformaciones de datos con condiciones complejas.

### **1.5.2 Softek Transparent Data Migration Facility UNIX (IP)**

*Softek Transparent Data Migration Facility* UNIX (IP) es un producto de migración de datos diseñado para trasladar datos sin interrupciones, en remoto con entornos UNIX. (IBM, 2008)

Características (IBM, 2008):

- Realiza migraciones de datos en segundo plano mientras mantiene la disponibilidad de las aplicaciones.
- Excelente para migraciones de datos en remoto a través de la red.
- Ayuda a garantizar la integridad de los datos y permite realizar copias puntuales de datos para que estén disponibles en cualquier momento tras la sincronización inicial.
- Ofrece monitorización centralizada, control y gestión de migraciones a través de una consola común que gestiona entornos *Softek TDMF Windows* (IP), *TDMF UNIX* (IP), *Softek Replicator UNIX*, y *Softek Replicator Windows*.
- Ofrece soporte para migraciones en múltiples sistemas operativos como *Red Hat Enterprise Linux*, *AIX*, *HP-UX* y *Solaris*.

### **1.5.3 Herramienta para la migración de datos del Sistema de Inventario Participativo al ASSETS**

Esta herramienta se desarrolló en la UCI, en un principio para automatizar la migración de datos al ASSETS, eliminando las actividades manuales e intelectuales asociadas a esta tarea, la cual está integrada al Sistema de Inventario Participativo y ha contribuido a elevar la eficiencia y la eficacia de la información del estado del patrimonio de la universidad. Se desarrolló bajo la metodología RUP, en la plataforma .NET, con lenguaje de programación C# y de modelado UML. (López, 2004)

Presenta las siguientes funcionalidades (López, 2004):

- Conectarse a las bases de datos origen y destino.
- Cargar información de las bases de datos.
- Mostrar información de las bases de datos origen y destino en forma de árbol.

- Relacionar tablas.
- Relacionar atributos.
- Visualizar relaciones establecidas entre las tablas y los atributos de las bases de datos origen y destino.
- Migrar datos.

#### **1.5.4 AdventNetSwisSQL**

AdventNetSwisSQL ofrece una herramienta que ayuda a la migración de las estructuras y de los datos de bases de datos como Oracle, PostgreSQL y SQL Server. (SwinSQL, 2009)

Características (SwinSQL, 2009):

- Migración de datos de forma rápida asegurando la integridad de los mismos.
- Realiza la migración de datos desde documentos Excel y CSV hacia una base de datos.
- Migra tablas, índices y restricciones junto con los datos.
- Soporta conectividad JDBC 6.
- Suspende o reanuda la migración en cualquier momento.
- Fácil de usar e intuitiva interfaz gráfica de usuario.
- Ofrece un excelente soporte a sus clientes desde la evaluación del producto hasta la finalización con éxito de su migración.

#### **1.5.5 Ora2Pg**

Ora2Pg es un módulo de Perl para exportar una base datos Oracle a un esquema de PostgreSQL. Esta aplicación se conecta a la base de datos Oracle, extrae su estructura, y genera un script SQL que puede cargar en la base de datos PostgreSQL. La misma hace un respaldo de todo el esquema (tablas, vistas, secuencias, índices, privilegios) con todas las llaves ya sean primarias, foráneas y únicas exportadas a código PostgreSQL sin necesidad de editar el SQL generado. Posee una extensa guía de migración muy intuitiva en su sitio oficial. (Ortíz, 2009)

#### **1.5.6 Resultados del estudio de las herramientas de migración de datos**

Luego del estudio de algunas de las herramientas usadas en la realización de procesos de migración de datos, se propone utilizar Pentaho *Data Integration* debido a

las facilidades de uso que brinda. Dicha herramienta permite migrar datos entre bases de datos relacionales y no relacionales, además de que los trabajos y transformaciones diseñados y confeccionados en ella se pueden reutilizar en un proceso de migración en varios momentos a lo largo del tiempo, lo cual se ajusta perfectamente a las necesidades de la investigación.

## ***1.6 Herramientas y tecnologías a utilizar***

En este apartado se hace un análisis de las herramientas y tecnologías que serán usadas de formas complementarias para la realización de la migración. Se valoran sus principales características y rasgos significativos, así como los elementos técnicos que permitieron su elección para desarrollar la propuesta de solución.

### **1.6.1 Sistema Gestor de bases de datos PostgreSQL**

El SGBD utilizado por el proyecto XAUCE CRODA es PostgreSQL, por lo que es de interés para el mismo que la solución propuesta en esta investigación se realice sobre este SGBD, el cual presenta todas las funciones y funcionalidades necesarias para el desarrollo de la solución propuesta.

PostgreSQL versión 9.3 es un sistema de gestión de bases de datos objeto-relacional, distribuido bajo licencia BSD 7 y con su código fuente disponible libremente. Utiliza un modelo cliente/servidor y usa multiprocesos en lugar de multi-hilos para garantizar la estabilidad del sistema. Un fallo en uno de los procesos no afectará el resto y el sistema continuará funcionando. Diseñado para ambientes de alto volumen, PostgreSQL incluye características de la orientación a objetos, como puede ser la herencia, tipos de datos, funciones, restricciones, *triggers* (disparadores), reglas, procedimientos almacenados, secuencias, relaciones y vistas. (González, 2012)

Ejemplos de su avanzada funcionalidad son consultas SQL declarativas, control de concurrencia multi-versión, soporte multi-usuario, transacciones, optimización de consultas y arreglos. PostgreSQL se caracteriza por ser un sistema estable, de alto rendimiento, gran flexibilidad, gran compatibilidad, además tiene características que permiten extender fácilmente el sistema. Además, cuenta con una documentación muy bien organizada, pública y libre. (González, 2012)

### 1.6.2 Administrador de bases de datos PgAdmin

El administrador de bases de datos utilizado por el proyecto XAUCE CRODA es PgAdmin III, por lo que es de interés para el mismo que la solución propuesta en esta investigación se realice sobre esta herramienta, el cual presenta todas las funciones y funcionalidades necesarias para el desarrollo de la solución propuesta.

PgAdmin III es una aplicación gráfica con licencia de código abierto para administrar bases de datos PostgreSQL. Está escrita en C++ usando la librería gráfica multiplataforma wxWidgets, lo que permite que se pueda usar en Linux, FreeBSD, Solaris, Mac OS X y Windows. Es capaz de gestionar versiones a partir del PostgreSQL 7.3 ejecutándose en varias plataformas, así como versiones comerciales de PostgreSQL como *Pervasive Postgres*, *Enterprise DB*, *Mammoth Replicatory SRA* *PowerGres*. (PgAdmin, 2010)

Es una herramienta diseñada para responder a las necesidades de todos los usuarios, desde escribir consultas SQL simples hasta desarrollar bases de datos complejas. La interfaz gráfica soporta todas las características de PostgreSQL y facilita la administración. La aplicación también incluye un editor SQL con resaltado de sintaxis, un editor de código de la parte del servidor y un agente para lanzar scripts programados. La conexión al servidor puede hacerse mediante conexión TCP/IP o Unix *Domain Sockets* (en plataformas \*nix), y puede encriptarse mediante SSL para mayor seguridad. (PgAdmin, 2010)

Es posible combinar el uso de esta herramienta con el *Visual Paradigm* sirviendo una de apoyo a la otra, además permite la creación y visualización de diagramas entidad-relación. PgAdmin está desarrollado por una comunidad de expertos en PostgreSQL y está disponible en varios idiomas. (PgAdmin, 2010)

### 1.6.3 Herramienta para el modelado Visual Paradigm

*Visual Paradigm* para UML 8.0 es una herramienta para desarrollo de aplicaciones utilizando modelado UML ideal para Ingenieros de *Software*, Analistas de Sistemas y Arquitectos de sistemas que están interesados en construcción de sistemas a gran escala y necesitan confiabilidad y estabilidad en el desarrollo orientado a objetos. (Visual Paradigm, 2009)

*Visual Paradigm* también ofrece (Visual Paradigm, 2009):

- Navegación intuitiva entre la escritura del código y su visualización.

- Potente generador de informes en formato PDF/HTML.
- Documentación automática Ad-hoc.
- Ambiente visualmente superior de modelado.
- Sofisticado diagramador automáticamente de *layout*.
- Sincronización de código fuente en tiempo real.

#### **1.6.4 Sistema operativo Ubuntu**

El uso del Software Libre es sustentable en Cuba a partir de las ventajas que tiene con respecto a los del tipo privativo. Por esto, su aplicación como plataforma informática de trabajo adquiere una relevante significación. (Universidad Central de las Villas, 2015)

Algunas de las ventajas del software libre (Universidad Central de las Villas, 2015):

- La gran mayoría de las distribuciones y aplicaciones son gratis.
- Gran variedad de programas, versiones y personalizaciones.
- Es mucho menos propenso a virus.

La UCI apuesta por el uso del software libre, por lo que toda su producción se desarrolla en computadoras con sistemas operativos libres. La herramienta de autor XAUCE CRODA y su proceso de desarrollo no quedaron exentos de esta directriz, por lo que su estructura de almacenamiento y el entorno de trabajo de esta investigación se encuentran en un entorno productivo con sistema operativo Ubuntu 16.4. Debido a esto se determinó confeccionar la estrategia de migración para que se usara sobre sistemas operativos libres. Todas las herramientas usadas en este proceso son compatibles con este entorno de trabajo, por lo que no se asegura que estas sean compatibles con otros sistemas operativos.

Ubuntu está basado en Debian GNU/Linux, Ubuntu concentra su objetivo en la facilidad de uso, la libertad en la restricción de uso, los lanzamientos regulares (cada 6 meses) y la facilidad en la instalación. (LinuxZone, 2010)

Características de Ubuntu (LinuxZone, 2010):

- Los desarrolladores de Ubuntu se basan en gran medida en el trabajo de las comunidades de Debian y GNOME.
- Las versiones estables se liberan cada 6 meses y se mantienen actualizadas en materia de seguridad hasta 18 meses después de su lanzamiento.

- El entorno de escritorio oficial es GNOME y se sincronizan con sus liberaciones.
- El navegador Web oficial es Mozilla Firefox.
- El sistema incluye funciones avanzadas de seguridad y entre sus políticas se encuentra el no activar, de forma predeterminada, procesos latentes al momento de instalarse. Por eso mismo, no hay un firewall predeterminado, ya que no existen servicios que puedan atentar a la seguridad del sistema.
- Para labores y tareas administrativas en terminal incluye una herramienta llamada sudo (similar al Mac OS X), con la que se evita el uso del usuario *root* (administrador).
- Mejora la accesibilidad y la internacionalización, de modo que el *software* está disponible para tanta gente como sea posible.

## ***1.7 Lenguajes utilizados***

### **Lenguaje de consulta estructurado SQL**

SQL (*Structured Query Language*) es un lenguaje de consulta estructurado de alto nivel, no procedural, normalizado, que permite la consulta y actualización de los datos de BD relacionales. (Scribd, 2011)

Una de sus características es el manejo del álgebra y el cálculo relacional permitiendo efectuar consultas con el fin de recuperar, de manera sencilla, información de interés de una base de datos, así como también hacer cambios sobre ella. Es utilizado por los programas de aplicaciones para realizar consultas a las bases de datos y ha sido adoptado por los principales proveedores y programadores de aplicaciones de clientes. (SwinSQL, 2009)

### **Lenguaje de programación de consola *Bourne Again Shell***

*Bourne Again Shell* (*bash*) no es únicamente un *shell*, sino que también es un lenguaje de scripting en sí mismo. El *shell* scripting permite utilizar las capacidades de la *shell* para automatizar multitud de tareas que, de otra forma, requerirían múltiples comandos introducidos de forma manual. *Bash* es la versión mejorada de *Bourne Shell*, fue escrita en 1987 por Brian Fox y es la consola por defecto en la mayoría de sistemas Linux. (*GNU Operating System*, 2002)

La sintaxis de *bash* cuenta con las extensiones de *shell Bourne*, pero también cuenta con otras (*GNU Operating System*, 2002):

- Matemáticas con enteros.
- Redirección de entrada/salida.
- Expresiones regulares.
- Escapado de caracteres con *Backslash*.

## ***1.8 Conclusiones del capítulo***

Luego de realizar el estudio de los procesos homólogos, se determinó que ninguno de los entornos en los que se realizaron, tiene las mismas características presentes en el entorno de la migración de datos en cuestión. Por lo que se determinó confeccionar una estrategia propia para la migración de datos a realizar. Para confeccionar dicha estrategia de migración se seleccionaron las siguientes herramientas y tecnologías: Pentaho *Data Integration* 5.1.0.0, para diseñar y construir las transformaciones y trabajos necesarios para realizar la migración entre las bases de datos y realizar el monitoreo en dicho proceso, PostgreSQL 9.4 para gestionar las bases de datos relacionales, *Visual Paradigm* 8.0 para la confección de diagramas, PgAdmin III para administrar las bases de datos en PostgreSQL y Ubuntu 16.4 como sistema operativo. Se seleccionaron, además, los siguientes lenguajes: SQL como lenguaje de consulta estructurado y *bash* como lenguaje de programación de consola.

## Capítulo 2. Descripción y análisis de la solución propuesta

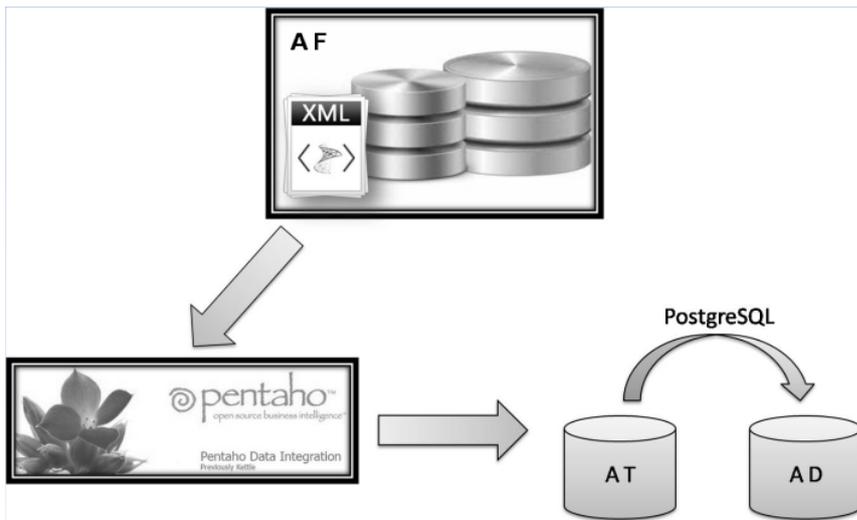
### 2.1 Introducción

Actualmente muchas organizaciones llevan a cabo la migración de datos para preservar información histórica actualizada en un sistema y utilizarla en la elaboración de estadísticas reales para su posterior uso en la toma de decisiones. Esta no es una tarea simple por lo que requiere un buen diseño de una estrategia que permita que se realice con la calidad requerida y sin pérdida de información. En el siguiente capítulo se explican los pasos que se llevarán a cabo en la estrategia de migración que se propone para dar solución a la problemática por la cual surge la presente investigación, así como las características de los sistemas involucrados.

### 2.2 Propuesta de arquitectura para la migración de los datos

Para la presente investigación se definió la siguiente arquitectura (Ilustración 1):

- **Área fuente (AF):** está conformada por la estructura de almacenamiento de la versión 2.0 de XAUCE CRODA. En esta área se realiza el levantamiento de la información deseada, el análisis de calidad de los datos y el mapeo de datos. Permite definir los datos que se desean migrar hacia el área destino.
- **Área temporal (AT):** está conformada por una estructura de almacenamiento clon de la estructura de almacenamiento de la versión 3.0 de XAUCE CRODA. Esta área es utilizada como punto medio entre el origen y el destino para poder realizar pruebas al proceso de migración antes de migrar los datos hacia la estructura de almacenamiento de XAUCE CRODA versión 3.0. Esto permite que la migración de datos hacia la estructura de almacenamiento destino se realice solo cuando la calidad del proceso sea la deseada.
- **Área de datos (AD):** está conformada por la estructura de almacenamiento de la versión 3.0 de XAUCE CRODA. En esta área se realizan las cargas reales y la revisión de resultados.



**Ilustración 1. Arquitectura de la migración**

## ***2.3 Migración de datos***

Existen diversos motivos para hacer una migración de datos, en el caso en cuestión, se determinó realizar este proceso debido a que la nueva versión de XAUCE CRODA necesitaba una actualización de contenidos, los cuales se encuentran almacenados en la versión 2.0 de la aplicación.

### **2.3.1 Etapas definidas a utilizar en la estrategia de migración de datos**

En la presente investigación se definieron las siguientes etapas o fases por las características propias del proceso de migración en cuestión y el entorno en que se desarrolla, a continuación, se presenta una descripción de cada fase y los resultados obtenidos en cada una de ellas:

- **Levantamiento y análisis de calidad:** se realiza el levantamiento de la información y se verifica la calidad de los datos en la fuente de origen (en el anexo 1 se muestra el diagrama de la base de datos en PostgreSQL de la fuente origen), con el propósito de definirlos datos deseados y los tiempos reales en el resto de las etapas del proceso. Al término de esta etapa se identificó la información deseada y se determinó que la misma tenía la calidad necesaria, en el caso de la base de datos eXist se determinó migrar 21 objetos de aprendizaje, en el caso de la base de datos en PostgreSQL se determinó migrar 20 tablas con un total de 1237 tuplas y en el caso de la estructura de directorios se determinó migrar 45 recursos multimedia.

- **Mapeo de datos:** se diseña el mapeo de datos entre las fuentes origen y destino. Al término de esta etapa se obtuvo como resultado 20 diagramas que especifican hacia donde se migrarán los datos seleccionados de la base de datos PostgreSQL (Anexo 3).
- **Construcción del ETL** (Extracción, transformación y carga): se realiza la construcción de los procesos de extracción, transformación y carga de los datos con la herramienta Pentaho *Data Integration*, teniendo en cuenta el análisis realizado en las fases anteriores. Al término de esta etapa se obtuvo como resultado el diseño de 37 transformaciones para migrar los datos de la base de datos en PostgreSQL y 1 transformación para migrar los OA almacenados en eXist.
- **Simulación de carga:** el objetivo de esta etapa es migrar los datos deseados, 20 tablas con un total de 1237 tuplas y 21 OA con los ETL antes diseñados y 45 recursos multimedia con el uso de un script.
- **Pruebas:** se realizan pruebas funcionales al proceso de migración, para determinar si los datos son traspasados con éxito al AT, con el uso del módulo de monitoreo de la herramienta Pentaho *Data Integration*. Fue necesario realizar dos iteraciones de pruebas debido a la ocurrencia de errores en la primera iteración, donde se detectaron 167 errores. Dichos errores se clasificaron en dos grupos, tuplas con datos nulos (65 errores) y tuplas con datos duplicados (102 errores). Antes de realizar la segunda iteración se modificó el diseño de algunas transformaciones del proceso ETL para eliminar los errores del tipo tuplas con datos duplicados y se levantaron las restricciones de la base de datos destino para eliminar los errores del tipo tuplas con datos nulos. En la realización de la segunda iteración se obtuvo como resultado la migración del 100% de los datos deseados.
- **Cargas Reales:** se realiza la migración de los datos desde el AT hacia el AD, con el proceso ETL creado en la última iteración de la fase construcción del ETL y levantando las restricciones de la base de datos destino. Logrando migrar 20 tablas con un total de 1237 tuplas y 21 OA con el uso del ETL y 45 recursos multimedia con el uso de un script.
- **Revisión de resultado:** se realizan pruebas de aceptación, por parte del cliente, las cuales garantizan no solo la correcta migración de los datos, sino que también garantiza que los datos migrados hacia el AD no afecten el correcto funcionamiento de la aplicación.

El cumplimiento de dichas etapas o fases en la puesta en práctica de la migración en el orden determinado en la investigación (Ilustración 2) garantiza que se migren los datos deseados de forma íntegra hacia el AD. En los sub-epígrafes 2.5 y 2.6, se realiza una descripción más detallada de la puesta en práctica de dicha solución.

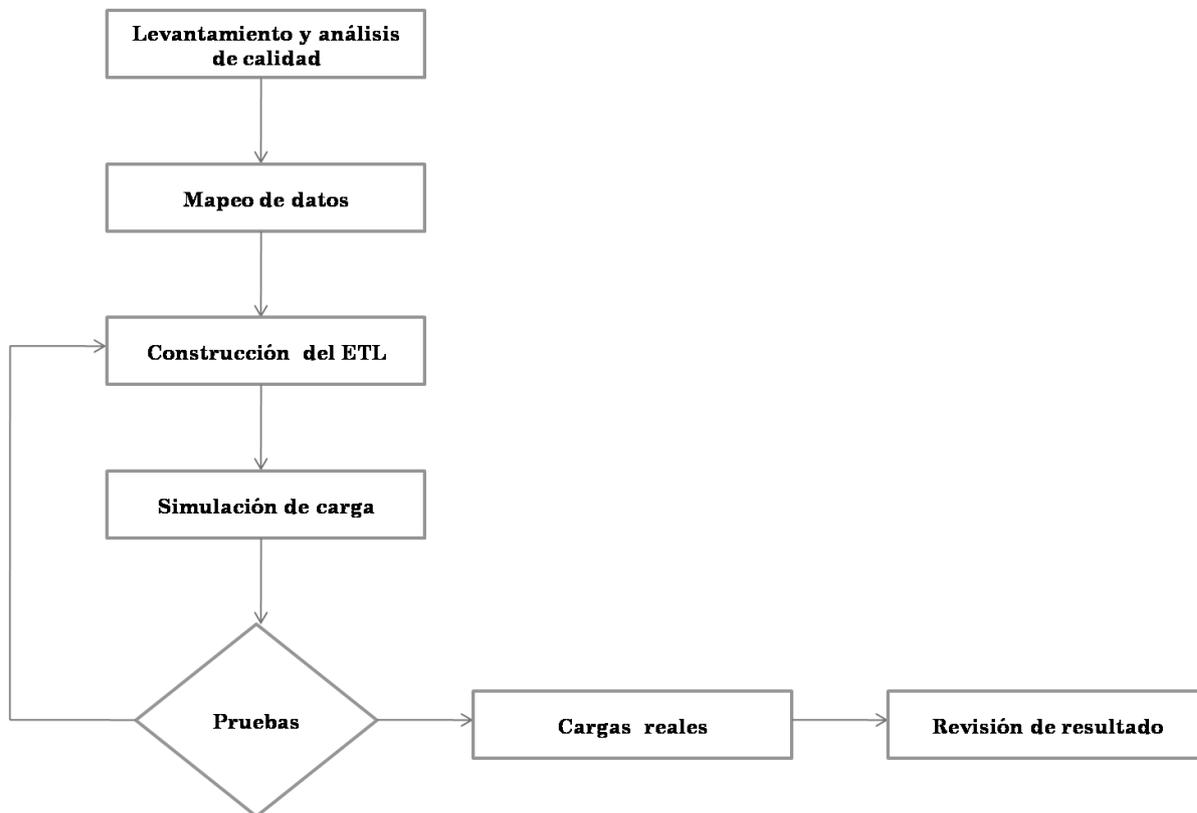


Ilustración 2. Diseño de las fases de la migración

### 2.3.2. Técnicas de migración de datos

Existen varias técnicas para ayudar a mejorar el proceso de migración de datos. Algunas de dichas técnicas se han utilizado en el proceso de migración en cuestión, ellas son:

- **Planeación:** esta técnica es utilizada luego de terminar la primera etapa del proceso, al tener la certeza de la cantidad de datos a migrar y la calidad de los mismos, para ordenar y planificar las acciones que se tomarán a cabo a lo largo de la migración. Esta técnica permitió determinar que el proceso de migración (la puesta en práctica de todas las etapas restantes) se realizaría en un intervalo de entre 8 y 10 semanas.

- **Mapeado de tipos de datos:** esta técnica es utilizada en la segunda y tercera fase del proceso de migración, garantiza que los datos que se van a migrar estén en el formato esperado por la estructura de almacenamiento destino y de esta forma eliminar incompatibilidades de formato en el proceso de migración.
- **Restricciones y disparadores:** antes de iniciar las fases de Simulación de cargas y Cargas Reales, es recomendable deshabilitar todos los *triggers* y restricciones que existan en la base de datos destino para disminuir la posibilidad de ocurrencia de errores en el proceso de migración.

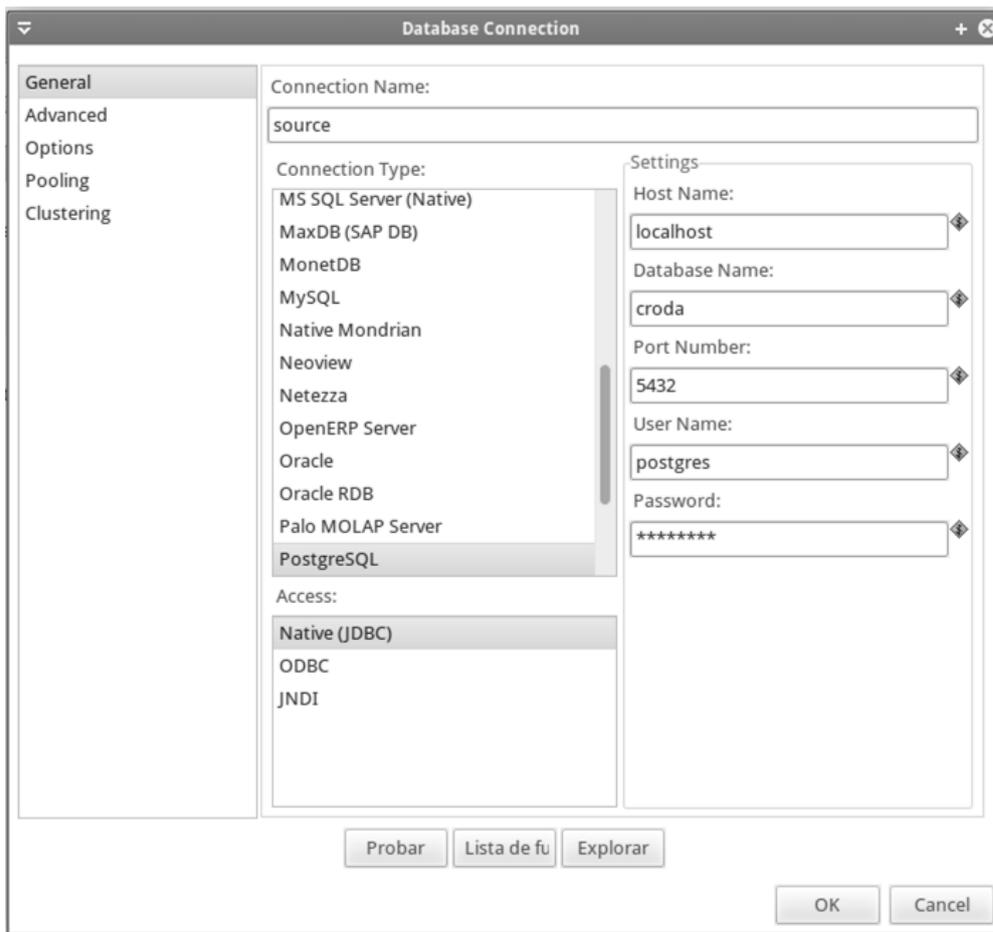
## ***2.4 Proceso de extracción, transformación y carga (ETL)***

El proceso ETL permite mover datos desde distintas fuentes, reformatearlos y cargarlos en bases de datos. Es utilizado para migrar datos de una o más fuentes a terceros y también para convertir bases de datos de un tipo o formato a otro. (Muñoz, 2011)

### **2.4.1 Extracción**

La primera parte del proceso ETL consiste en extraer los datos desde los sistemas de origen. En este subproceso se convierten los datos a un formato preparado para iniciar el proceso de transformación. Aquí se verifican los datos extraídos, donde se comprueba si los datos cumplen con lo que se espera y se adaptan al formato estándar diseñado, de lo contrario son rechazados. (Muñoz, 2011)

En el proceso ETL diseñado para la migración en cuestión, se utilizó para la extracción de los datos el componente Entrada tabla, el cual define la conexión con la fuente origen (en este caso la base de datos en PostgreSQL y eXist de la versión 2.0 de XAUCE CRODA). A continuación, en la ilustración 3 se muestra la configuración de dicho componente.



**Ilustración 3. Configuración del componente Entrada tabla**

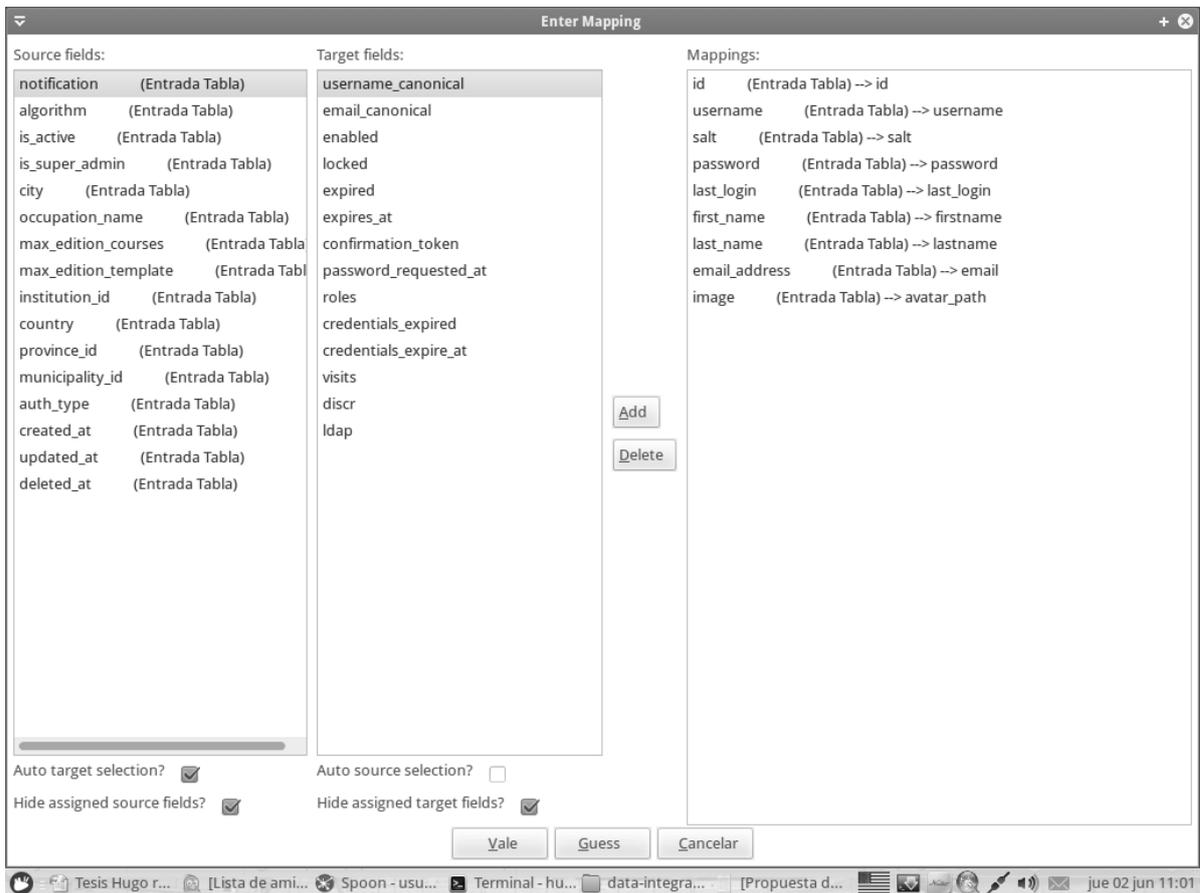
## 2.4.2 Transformación

La fase de transformación aplica una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados. (Muñoz, 2011)

En el proceso ETL diseñado para la migración de datos en cuestión, se utilizó para la transformación de los datos el componente Selecciona/Renombrar valores, el cual define la conexión entre los datos seleccionados por el componente de extracción y por el de carga, realizando el mapeo de datos diseñado anteriormente. Además, modifica los nombres de los atributos de las tablas a migrar para una mayor compatibilidad. A continuación, en la ilustración 4 y 5 se muestra la configuración de dicho componente.



**Ilustración 4. Configuración del componente Selecciona/Renombra valores para la modificación de los nombres de los atributos de las tablas a migrar.**



**Ilustración 5. Configuración del componente Selecciona/Renombra valores para la determinación del mapeo de datos.**

### 2.4.3 Carga de los datos

La fase de carga es el momento en que los datos son insertados en la base de datos del sistema destino, una vez terminada la transformación y limpieza de los mismos. El principal objetivo de esta fase es lograr que los datos estén listos para ser consultados. (Muñoz, 2011)

En el proceso ETL diseñado para la migración en cuestión, se utilizó para la Carga de los datos el componente Insertar/Actualizar, el cual define la conexión con la fuente destino (en este caso la base de datos en PostgreSQL de la versión 3.0 de XAUCE CRODA). A continuación, en la ilustración 6 se muestra la configuración de dicho componente.

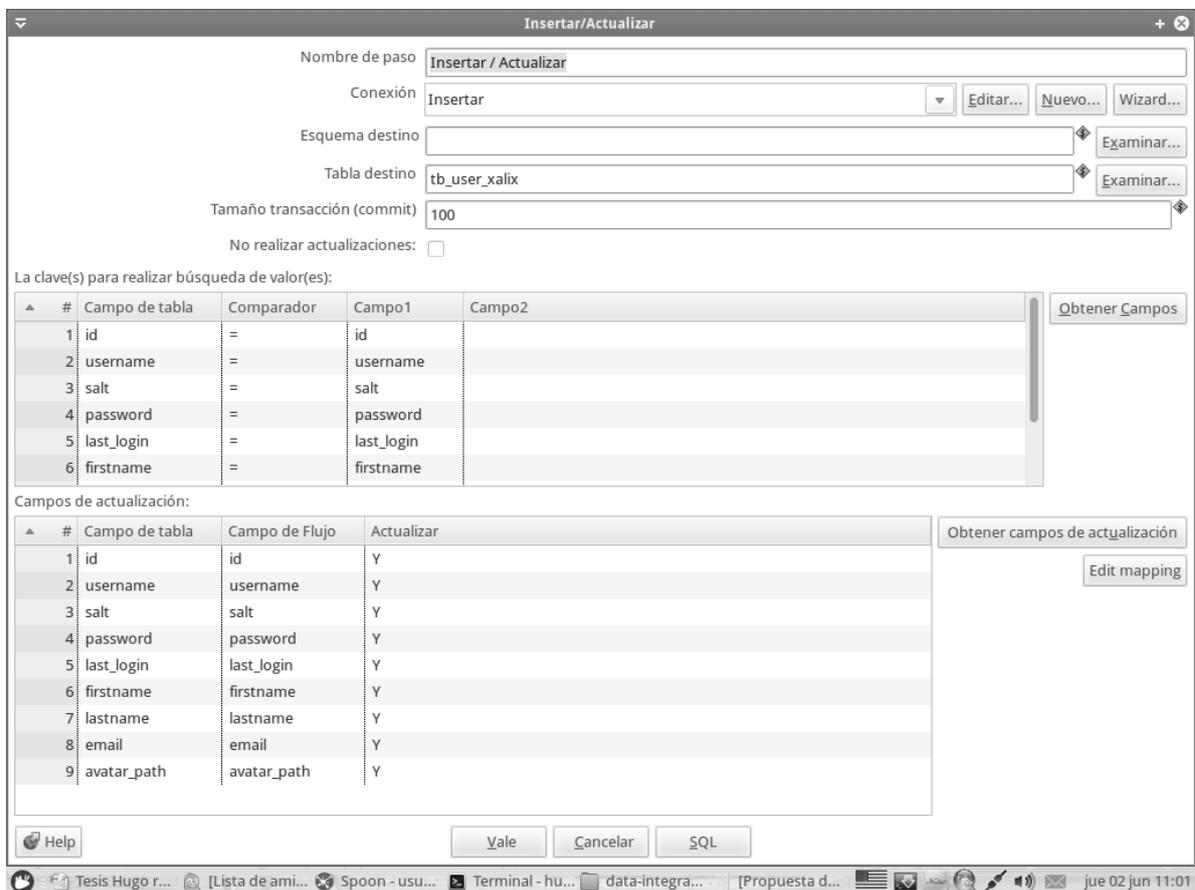


Ilustración 6. Configuración del componente Insertar/Actualizar

### 2.5 Estrategia de migración propuesta.

Siguiendo el diseño de las fases, la primera etapa es: levantamiento y análisis de calidad, en la cual, se recomienda realizar un análisis de las fuentes origen y las

fuentes destino, para determinar qué datos se desean migrar y la calidad de los mismos. Para llevar a cabo esta tarea se necesita tener:

- Las estructuras de almacenamientos (*Back-up* de las bases de datos y la estructura de los directorios)
- Los diagramas de base de datos de las fuentes origen.
- Los diagramas de base de datos de las fuentes destinos.

Para realizar dicho análisis, se recomienda analizar los datos de las tablas de las bases de datos y de la estructura de directorio de la fuente origen, en este proceso se debe determinar si los datos tienen la calidad necesaria para realizar la migración. Al término de esta fase se tienen identificados todos los datos a migrar, verificando la calidad de los mismos.

Luego de esta fase se realiza el mapeo de datos, para determinar hacia dónde se migrarán los datos seleccionados. En el caso de las bases de datos relacionales se recomienda usar diagramas de relación entre las tablas de las fuentes origen y destino usando la herramienta Visual Paradigm para UML 8.0.

Al término de esta fase se procede a la construcción del ETL, proceso en el que se establecen las reglas de migración que se ejecutarán para la realización de la extracción, transformación y carga, el cual se propone realizarlo con la herramienta *Pentaho Data Integration*.

Para la extracción, la herramienta brinda una variada cantidad de opciones, la elección de una u otra está en dependencia de las fuentes de origen de la información. Se escoge la que se necesite y se configura la conexión.

Para la transformación de los datos, la herramienta brinda varias opciones, la elección de una, varias o ninguna, radica en el tratamiento que sea necesario darle a los datos. Se escoge la que se necesite y se configura en dependencia de la transformación. En este paso se puede incorporar elementos para la limpieza de los datos; simplificarlos o eliminar datos repetidos, para evitar anomalías en inserciones.

Para la carga de los datos en la fuente destino la herramienta brinda una serie de opciones, la elección entre estas depende del tipo de la fuente destino. Se escoge la que se necesite y se configura la conexión.

Para garantizar una mayor calidad en la migración, se realiza una simulación de carga, usando el ETL y el *script* construido, hacia el AT. Esto reduce la incidencia de errores en la etapa cargas reales. Para su control se propone confeccionar tablas de Diseño del trabajo, con las cuales se tiene constancia de cuáles datos se migran hacia dónde y la ocurrencia de errores o no en la carga.

Durante el proceso de migración se propone utilizar la opción de *monitoring* de la herramienta Pentaho *Data Integration*, la cual está disponible a partir de la versión 3.1.0. Dicha opción permite realizar el seguimiento del rendimiento de operaciones sucesivas de una transformación. Esta es una herramienta importante que le permite ajustar con precisión el rendimiento de la transformación, ya que el rendimiento está determinado por la etapa más lenta en la transformación.

En dependencia del resultado de esta última fase, se determina si es necesario repetir las fases construcción del ETL, simulación de cargas y pruebas o si el proceso tiene la calidad deseada para realizar las cargas reales.

Asumiendo que no sea necesario repetir las fases, se procede entonces a migrar los datos hacia el AD utilizando los ETL utilizados en la última iteración del ciclo (construcción del ETL, simulación de carga y pruebas). Luego se cambia la fuente destino por AD y se realiza la migración de datos de la misma manera que en la fase simulación de carga. También se vuelve a usar la opción de *monitoring* para realizar el seguimiento del rendimiento en esta fase.

Por último, se realiza la comprobación del cumplimiento de los objetivos trazados con la migración, esto es posible a través de la realización de las pruebas de aceptación por parte del usuario o del cliente.

## ***2.6 Descripción de la solución propuesta***

Una vez diseñada la arquitectura y las etapas a seguir en la migración de los datos, se procede a la aplicación de la estrategia creada para dar solución al problema planteado. A continuación, se describen en un ejemplo ilustrativo, los pasos empleados en dicho proceso.

### 2.6.1 Desarrollo de la migración de la tabla tb\_metadata de la base de datos en PostgreSQL del AF hacia las tablas oa y oouser de la base de datos en PostgreSQL del AD.

Inicialmente se realizó un análisis de los datos contenidos en la estructura de almacenamiento de la fuente origen y destino, lo que dio como resultado los datos deseados de la base de datos en PostgreSQL de la versión 2.0 de XAUCE CRODA (ilustraciones 7 y 8). Entre estos datos deseados se encuentra la tabla tb\_metadata de la base de datos en PostgreSQL, la cual se tomará de ejemplo para explicar el proceso de migración. A dicha tabla se le realizó una revisión para determinar que sus datos presentan la calidad necesaria para realizar el proceso.

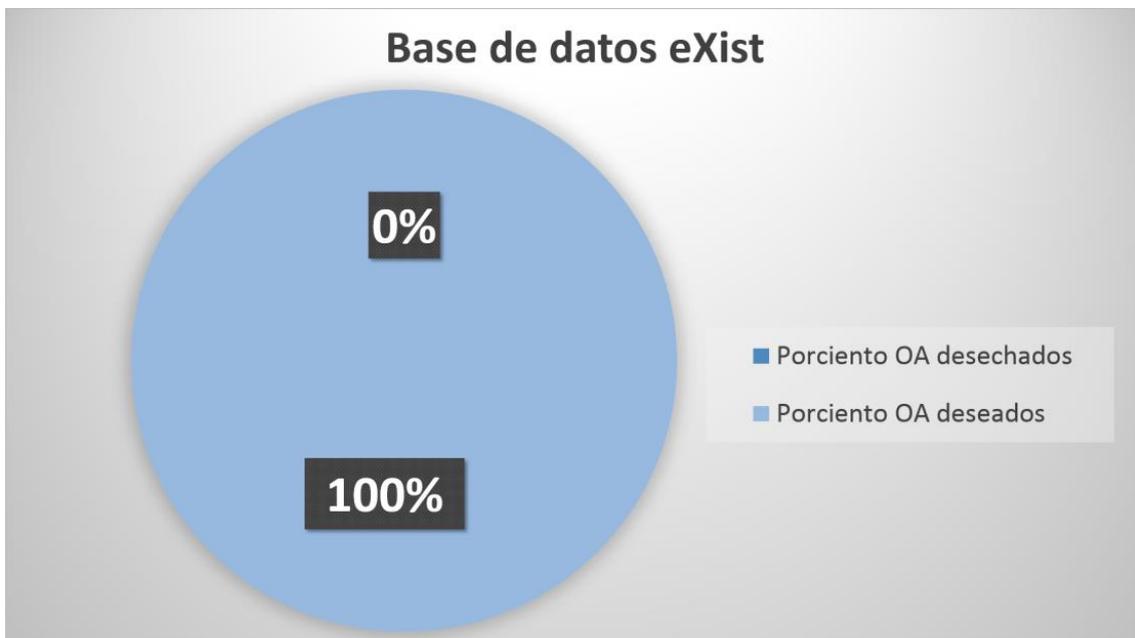
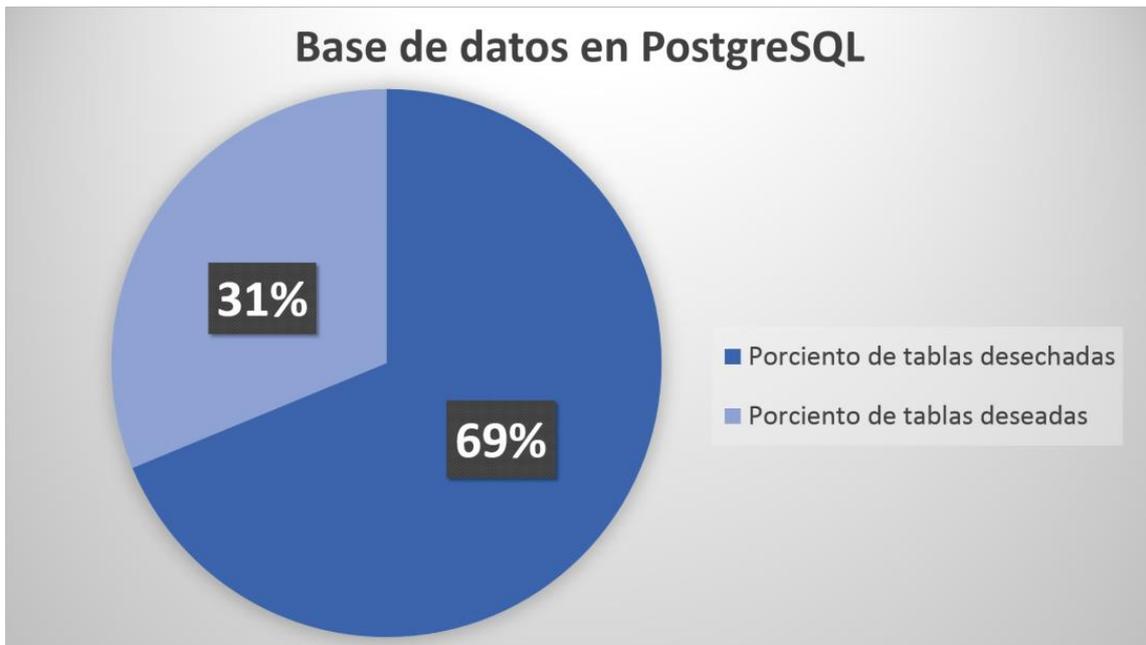
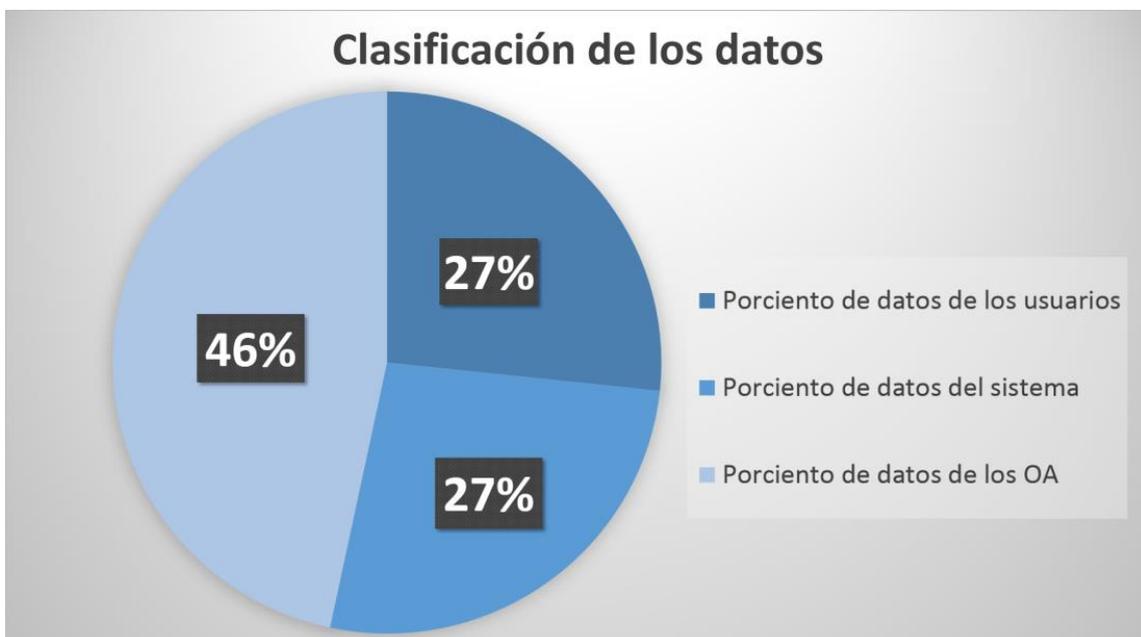


Ilustración 7. Datos de la base de datos eXist



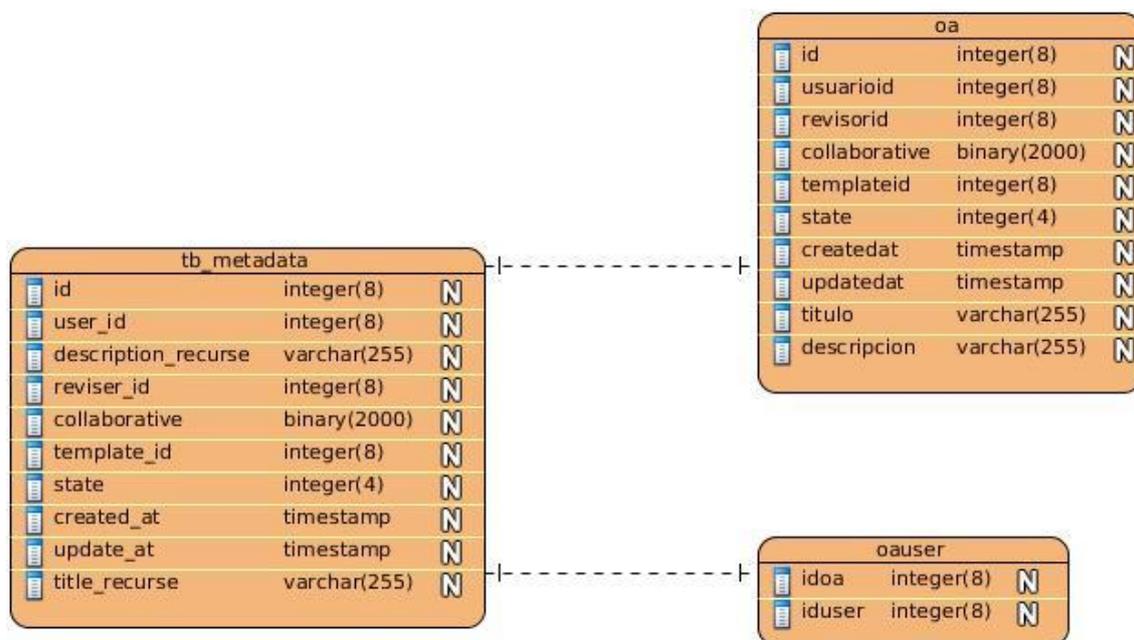
**Ilustración 8. Datos de la base de datos en postgresQL**

Los datos deseados se clasificaron en tres grupos (datos de usuarios, datos del sistema y datos de los OA) para determinar si existe o no la necesidad de realizar la migración dado que la prioridad es la migración de los OA creados en la versión 2.0 de XAUCE CRODA, lo cual dio como resultado que casi el 50% de los datos a migrar pertenecen a los OA como se muestra en la ilustración 9.



**Ilustración 9. Clasificación de los datos deseados**

Terminada la primera etapa se realiza el mapeo de datos, definiendo qué datos de la fuente origen se guardarán en qué campos de la fuente destino y verificando que los formatos de los mismos sean compatibles. Para ello se confeccionaron 12 diagramas (Ver anexo 3) incluido el de la tabla `tb_metadata` (ilustración10) para esbozar el mapeo de datos en PostgreSQL que se realizará posteriormente en el proceso ETL.



**Ilustración 10. Esbozo del mapeo de datos de la tabla `tb_metadata`**

Haciendo uso del Pentaho *Data Integration* se creó un diseño de trabajo para cargar los datos contenidos en la tabla `tb_metadata`, hacia las tablas `oa` y `oauser`. El mismo, en este caso, consta de tres componentes:

**Entrada Tabla:** para la conexión a la fuente origen y la carga de los datos contenidos en ella.

**Seleccionar/Renombrar valores:** para renombrar los nombres de los atributos que no coincidan en el origen y el destino, lo cual permite tener una mayor limpieza en el proceso y realiza el mapeo de datos definido anteriormente, enlazando los campos de la fuente de origen con los correspondientes en el destino.

**Insertar / Actualizar:** para la conexión a la fuente destino y la inserción de los datos deseados provenientes de la fuente origen.

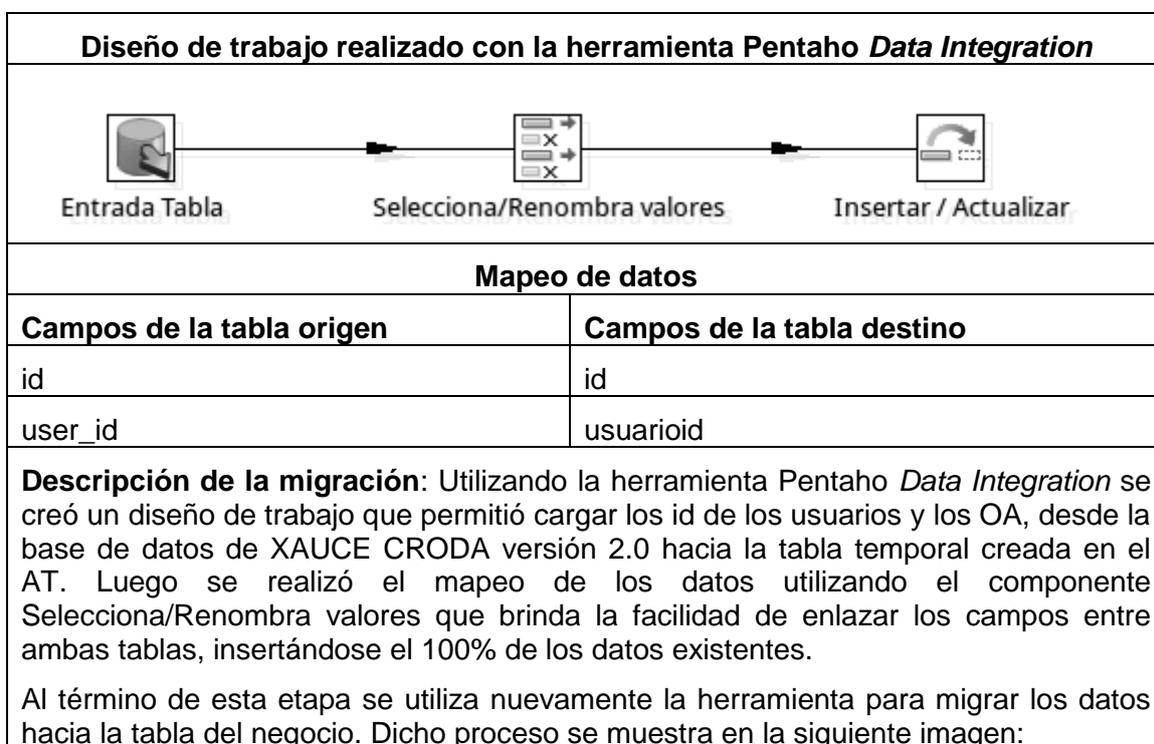
Luego de terminada la construcción del ETL se realiza la simulación de carga hacia el AT. Para su control se confeccionaron tablas de Diseño del trabajo (Anexo 4), dichas tablas se realizan por cada transformación, en el caso de la tabla `tb_metadata` se

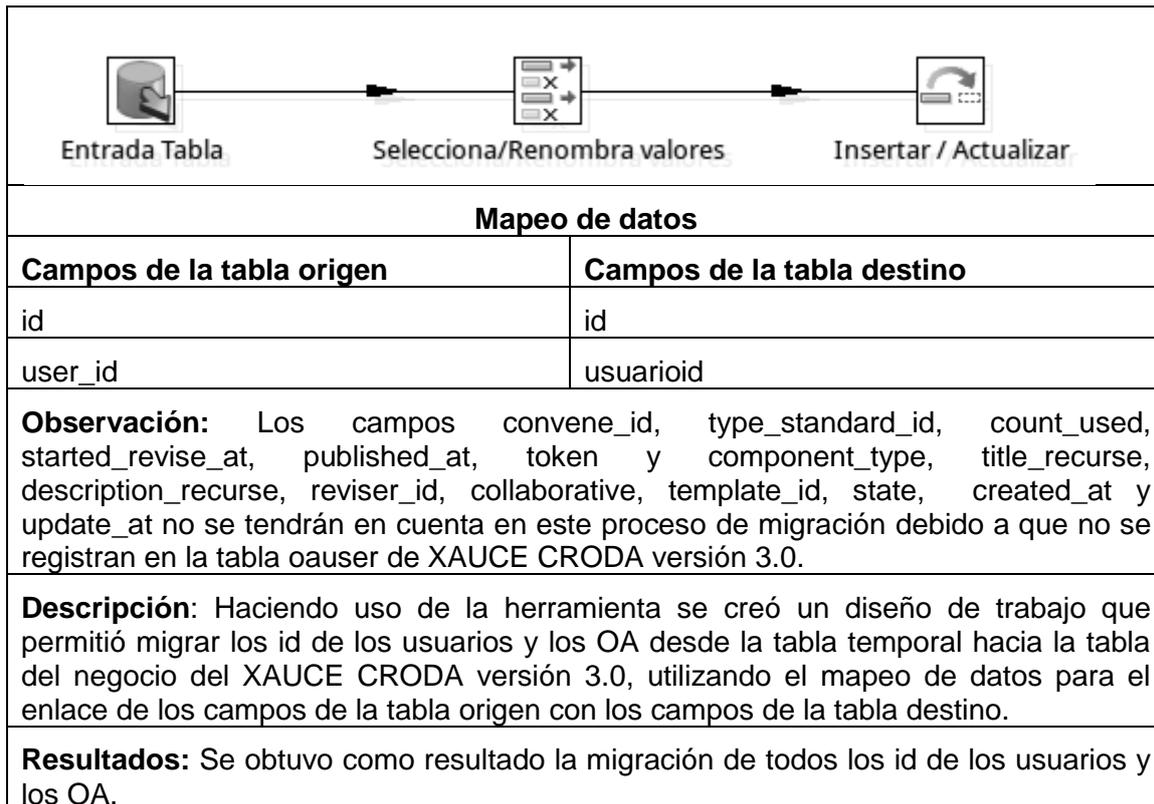
realizaron dos transformaciones las cuales se documentaron en dos tablas de diseño de trabajo (tabla 1 y 2) con las que se tiene constancia de cuáles datos se migran hacia dónde y la ocurrencia de errores o no en la carga. Para migrar los datos deseados de la base de datos no relacional y de los directorios se confeccionaron dos *scripts*.



title_recurse	titulo
description_recurse	descripcion
reviser_id	revisorid
collaborative	collaborative
template_id	templateid
state	state
created_at	createdat
update_at	updateat
<p><b>Observación:</b> Los campos convene_id, type_standard_id, count_used, started_revise_at, published_at, token y component_type no se tendrán en cuenta en este proceso de migración debido a que no se registran en la tabla tb_permission de XAUCE CRODA versión 3.0.</p>	
<p><b>Descripción:</b> Haciendo uso de la herramienta se creó un diseño de trabajo que permitió migrar los metadatos de los OA desde la tabla temporal del AT hacia la tabla del negocio del XAUCE CRODA versión 3.0, utilizando el mapeo de datos para el enlace de los campos de la tabla origen con los campos de la tabla destino.</p>	
<p><b>Resultados:</b> Se obtuvo como resultado la migración de todos los metadatos de los OA.</p>	

**Tabla 1. Descripción de la solución para la migración de la tabla tb\_metadata hacia la tabla oa**





**Tabla 2. Descripción de la solución para la migración de la tabla `tb_metadata` hacia la tabla `oauser`**

Con el uso de la herramienta Pentaho *Data Integration* se realizaron pruebas al proceso de migración (Etapa Pruebas) para verificar el nivel de calidad de mismo durante la migración entre el AF y el AT, luego de lograr la calidad necesaria en el proceso se realizó la migración entre el AT y el AD (Etapa Cargas reales) con el diseño de los ETL de la última iteración de las fases construcción del ETL, simulación de carga y pruebas y como última etapa se realizó la Revisión de resultados para garantizar el cumplimiento de los objetivos trazados y el buen funcionamiento de la herramienta de autor XAUCE CRODA 3.0.

## 2.7 Conclusiones parciales

A través de todo el transcurso de este capítulo se llevaron a cabo una serie de tareas y actividades, las cuales permitieron llegar a las siguientes conclusiones: la definición de una arquitectura permitió definir una estructura en la cual se pudiera realizar la migración de datos, la definición de las etapas de la migración permitió definir por fases las acciones que se realizarán de forma ordenada en la migración de datos, el uso de la herramienta Pentaho *Data Integration* permitió el diseño y confección del

proceso ETL. Con los elementos anteriormente descritos se pudo diseñar una estrategia de migración de datos.

## Capítulo 3. Validación de la solución propuesta

### 3.1 Introducción

Para realizar una buena toma de decisiones, se requiere cierta certeza sobre la confiabilidad de la información que se tiene, la cual está directamente ligada a la calidad de los datos almacenados, cómo son obtenidos y las manipulaciones por las que pasan antes de ser presentados a los usuarios finales. Las pruebas constituyen un paso fundamental en toda migración, ya que permiten operar sobre datos limpios, correctos y útiles; verificando la exactitud y el significado del dato en el sistema. El objetivo de las pruebas es descubrir el mayor número de errores con la mínima cantidad de esfuerzo y tiempo posible. (Pressman, 2015)

Teniendo en cuenta lo planteado anteriormente, se puede afirmar que validar la solución es un proceso que asegura la correcta migración de los datos a un sistema deseado. En el presente capítulo se exponen las pruebas que se realizaron en la investigación, dejando así constancia de la exitosa migración de datos provenientes de la versión 2.0 hacia la versión 3.0 de XAUCE CRODA.

### 3.2 Estrategia de validación y prueba

En la Ingeniería de Software, el proceso de pruebas permite verificar la calidad del producto. Es la fase final por donde pasa un sistema o proceso antes de ser entregado para su explotación y validación por parte del cliente. (Pressman, 2015)

La estrategia de prueba describe el enfoque y los objetivos generales de las actividades de prueba. Por tanto, después de realizar el proceso de migración de datos es preciso validar que la solución esté correcta, verificando que el 100% de los datos tomados de la fuente origen fueron insertados correctamente en la fuente destino. Esto tiene un gran nivel de importancia debido a que da la medida de que la migración se realizó satisfactoriamente. Todos los datos migrados hacia el AD tienen que cumplir con dos reglas básicas, que son:

- El proceso de migración de datos no genera ningún error.
- Los datos migrados contienen información completa (no presentan campos nulos).

Esto es posible en la etapa de Cargas reales debido a que las etapas construcción ETL, simulación de carga y pruebas, se realizan de forma cíclicas hasta lograr un alto nivel de calidad en el proceso de migración.

### 3.2.1 Pruebas realizadas a la simulación de carga

Durante el proceso de migración se utilizó la opción de *monitoring* de la herramienta Pentaho *Data Integration*. Dicha opción permite realizar el seguimiento del rendimiento de operaciones sucesivas de una transformación.

Con la opción *monitoring* se generaron gráficos estadísticos para comprobar qué porcentaje de los datos de la fuente origen se migraron hacia la fuente destino. Teniendo en cuenta los siguientes indicadores:

- **Null Count** (contar nulos): Cantidad de valores nulos en la fuente destino
- **Duplicate Count** (contar duplicados): Cantidad de valores duplicados en una columna

A continuación, se muestran algunos de los gráficos realizados durante el desarrollo de la migración. Los cuales muestran (ilustraciones 11 y 12) el porcentaje de datos migrados correctamente, el porcentaje de datos no migrados por ocurrencia de errores y la clasificación de dichos errores y el porcentaje de ocurrencia respectivamente

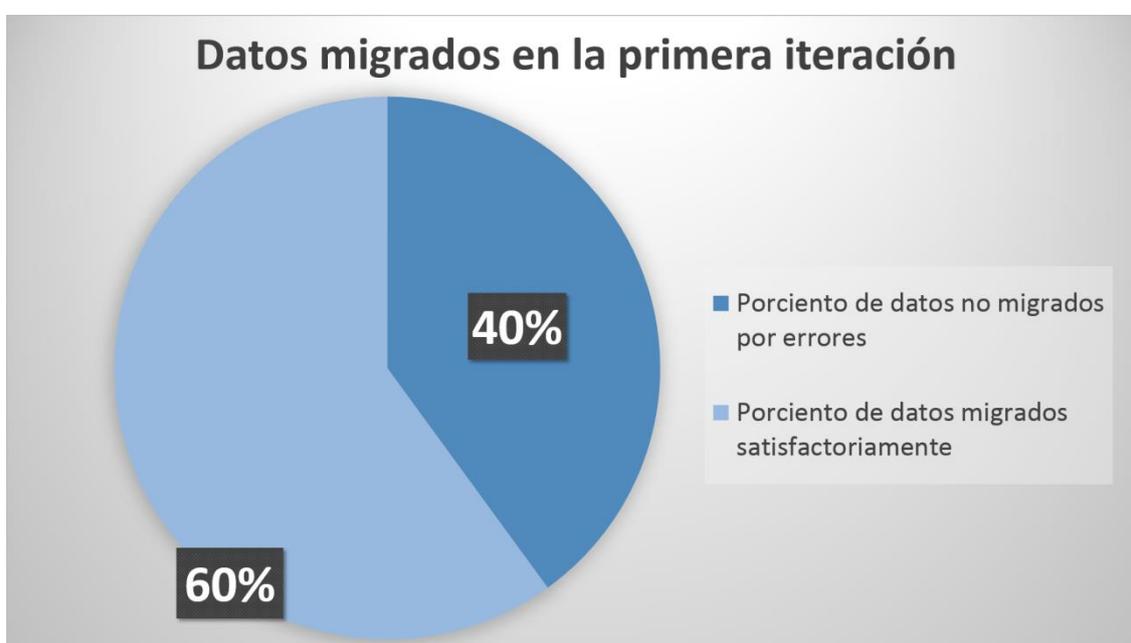
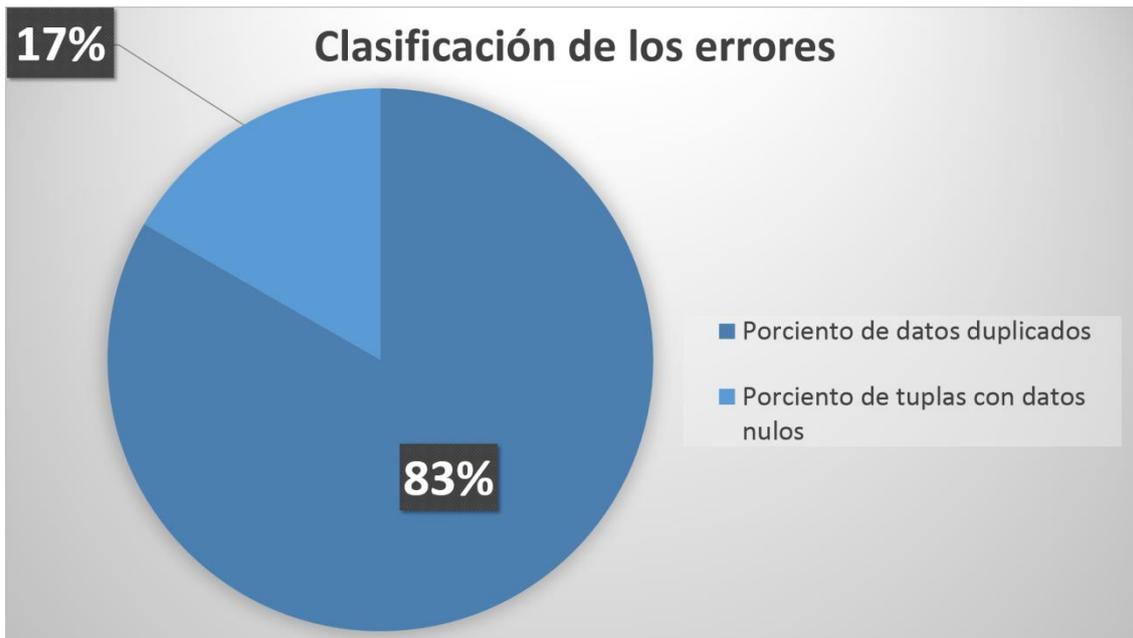


Ilustración 11. Resultado de la primera iteración de la migración

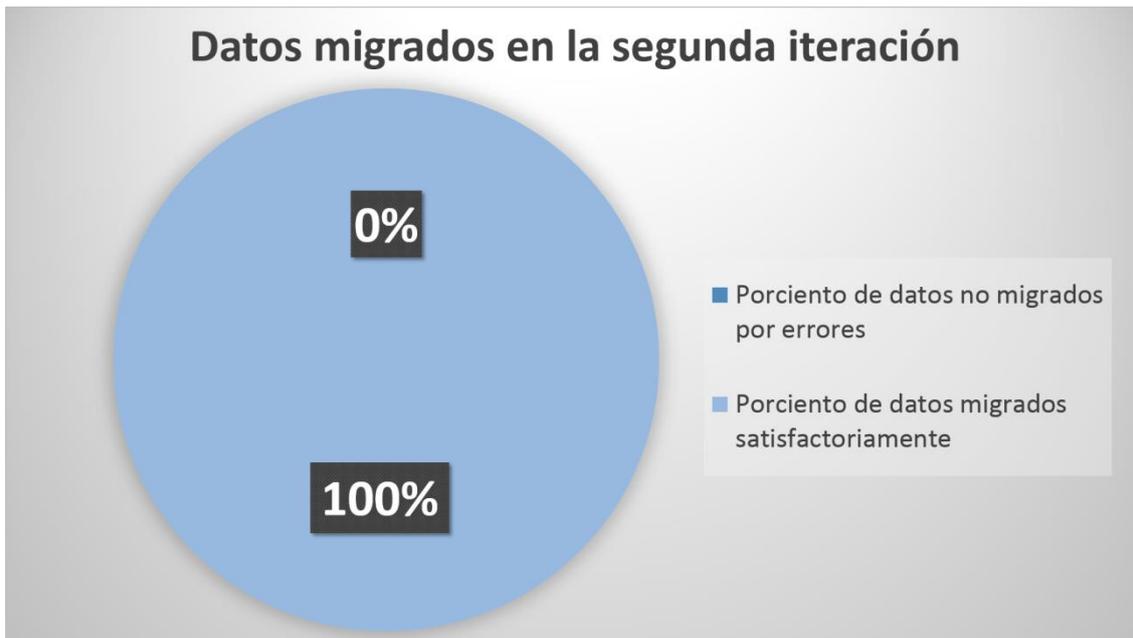


**Ilustración 12. Clasificación de los errores**

Los errores encontrados en esta etapa se clasificaron en dos grupos y la solución que se les dio fue:

- Datos duplicados (en las tablas `tb_userxalix` y `usuario`): se analizó el proceso de ETL y se reestructuró el mismo, para que no insertara datos ya migrados.
- Tuplas con datos nulos (en las tablas `role_user` y `role_permission`): se analizaron dichos datos y se detectó que se tenía toda la información, pero la misma procedía de distintas fuentes, por lo que para que se pudiera realizar de forma correcta la migración se determinó utilizar un *script* para levantar las restricciones que impedían que se insertaran y luego de terminada la migración ejecutar el mismo *script* para restablecerlas.

Dado que no se logró migrar satisfactoriamente todos los datos en la primera iteración, se corrigieron los errores y se repitieron las etapas construcción ETL, simulación de carga y pruebas, como estaba definido en el diagrama de etapas de la migración de datos presentado en la ilustración 1. Se obtuvo como resultado en la segunda iteración el 100% de los datos migrados (ilustración 13)



**Ilustración 13. Resultado de la segunda iteración de la migración**

Con las pruebas realizadas a la migración de datos, se pudieron examinar los errores que presentaba el proceso, lo cual brindó la posibilidad de corregirlos para que fuesen cargados con la calidad requerida hacia la versión 3.0 de la herramienta de autor XAUCE CRODA, y así lograr una migración satisfactoria.

### **3.2.2 Pruebas de aceptación**

Dentro del contexto de las pruebas, las más importantes son las de aceptación. Son aquellas pruebas que son diseñadas por el propio equipo de trabajo en base a los requisitos funcionales especificados en la fase de análisis para cubrir todo ese espectro, y ejecutadas por el propio usuario. Las mismas se utilizan para verificar que los datos en el sistema destino, cumplan con las especificaciones dadas, y confirmar que los usuarios principales del sistema no encuentren problemas en el uso de la aplicación. Con esto se puede determinar si los datos pueden ser usados en un ambiente de producción. (Ponce, 2014)

Para este tipo de prueba es importante mencionar que el sistema ya está en despliegue y está siendo utilizado sin dificultades por el cliente, quien pudo comprobar que los grupos de datos que existían en la estructura de almacenamiento de la versión 2.0 de XAUCE CRODA, fueron trasladados hacia su versión 3.0, por tanto, el desarrollo de esta prueba fue satisfactorio, quedando como constancia el acta de aceptación del cliente. (Anexo 3)

### ***3.3 Conclusiones parciales***

Para validar la propuesta de solución se realizó un monitoreo al proceso de migración dando como resultado la detección de errores en el proceso en la primera iteración, dichos errores se corrigieron y en la segunda iteración no se detectó la ocurrencia de ningún error, y se realizaron las pruebas de aceptación dando como resultado el cumplimiento de los objetivos trazados por el cliente.

## Conclusiones

A partir de los objetivos planteados y el trabajo realizado en esta investigación, se arribó a las siguientes conclusiones:

- Los procesos homólogos estudiados no satisfacen las necesidades de la migración de datos entre las versiones 2.0 y 3.0 de la herramienta XAUCE CRODA.
- La arquitectura, las fases de la migración y el proceso ETL son elementos necesarios para la confección de la estrategia de migración de datos entre las versiones 2.0 y 3.0 de XAUCE CRODA.
- La aplicación del monitoreo y las pruebas de aceptación al proceso de migración permiten que la migración de datos entre las versiones 2.0 y 3.0 de XAUCE CRODA se lleve a cabo con calidad.

## Recomendaciones

- Utilizar la estrategia diseñada en esta investigación para las posteriores migraciones realizadas entre las versiones 2.0 y 3.0 de XAUCE CRODA.
- Utilizar la estrategia diseñada en esta investigación para la migración de datos entre estructuras de almacenamiento con características iguales o similares a las definidas en la investigación.

## Referencias Bibliográficas

Argudo, W. O. (2014). *Propuesta de una metodología para el proceso de Migración de Datos en*. Sangolquí: Universidad de las Fuerzas Armadas ESPE.

Arias. (2007). *Sistema Integrado de Gestión Estadísticas*. La Habana.

Ciberaula. (2010). *Introducción a Apache*. Recuperado el 30 de Enero de 2016, de [http://www.ecured.cu/index.php/Servidor\\_web](http://www.ecured.cu/index.php/Servidor_web)

Cohen & Asín. (2000). *Sistemas de Información Para los Negocios: Un enfoque de toma de decisiones*. México: McGraw-Hill.

Coleman K., L. P. (2005). *Software Process Improvement in Small Organizations: A Case Study*. IEEE Software.

Cruz, F. N. (2006). *Estrategia del mundo Contemporáneo*. México: Hitek Patrick.

Dabbagh, N. H. (2001). *Authoring tools and Learning System*.

Date. (2001). *Introducción a los sistemas de bases de datos*. Mexico DF: Editorial Mexicana.

Gámez, Y. G. (2012). *Migración de los datos manejados por la Dirección de Cooperación Internacional hacia el*. Habana: UCI.

García, R. M. (1999). *Diseño de base de datos*.

Gerard, R. W. (1994). *Shaping the mind: Computer in education*. Sciences, Applied Science and technological Progress.

González, C. (2012). *Curso Base de Datos PostgreSQL*. Recuperado el 1 de Febrero de 2016, de [Curso Base de Datos PostgreSQL: http://www.usabilidadweb.com.ar/postgre.php](http://www.usabilidadweb.com.ar/postgre.php)

Henry Mintzberg, J. B. (1997). *El proceso estratégico: Conceptos, contextos y casos*. México: Prentice Hall Hispanoamericana.

Herrera, Á. A. (2007). *Migración de datos*. Recuperado el 4 de Febrero de 2016, de [Migración de datos: http://www.mailxmail.com/curso-migracion-datos/migracion-datos](http://www.mailxmail.com/curso-migracion-datos/migracion-datos)

IBM. (2008). *Softek TDMF Unix*. Recuperado el 29 de Enero de 2016, de [Softek TDMF Unix: http://www-05.ibm.com/services/es/sds/0905050200.html](http://www-05.ibm.com/services/es/sds/0905050200.html)

Lamas Collejo, Javier (2010). *Migración de la información del sistema de Gestión Académica, Akademos, al Sistema de Gestión Académica de Pregrado*. La Habana: Universidad de las Ciencias Informáticas.

Ponce González, José (2014). *Pruebas de aceptación orientadas al usuario: contexto ágil para un proyecto de gestión documental*. Sevilla: Universidad de Sevilla.

Kenneth, K. &. (2002). *Análisis y Diseño de Sistema*. USA: Editorial Prentice-Hall.

Kettle. (2007). *Xperimentos*. Recuperado el 3 de Febrero de 2016, de Xperimentos: <http://www.Xperimentos.com/kettle-240/>

Muñoz, J. N. (2011). *ETL Process Modeling Conceptual for Data Warehouses: A Systematic Mapping Study*. IEEE LATIN AMERICA TRANSACTIONS.

Lau, I. C. (2007). *Estado del arte de pruebas de base de datos para la migración y validación de datos*. Medellín: Universidad EAFIT.

LinuxZone. (2010). *Ubuntu: descripción de Ubuntu, descarga, características de Ubuntu*. . Recuperado el 1 de Febrero de 2016, de <http://www.linuxzone.es/distribuciones-principales/ubuntu/>

López, M. D. (2004). *Herramienta para la migración de datos del Sistema de Inventario Participativo al ASSETS*. Habana.

Dominguez Abreu, Lucía E. (2011). *Migración integrada de datos para el Sistema Nacional de Salud*. Habana: UCIENCIA.

Ortiz Valmaseda, Marcos (2009). *O2PMigration: Herramienta para la migración de base de datos Oracle a PostgreSQL*. Habana: Universidad de las Ciencias Informáticas.

Marotta, A. (2009). *Un caso de estudio en Calidad de Datos para*. Montevideo: FACULTAD DE INGENIERÍA UNIVERSIDAD DE LA REPÚBLICA.

Matos, R. M. (1999). *Diseño de base de datos*.

Martínez Tabasco, Milene Mayra (2012). *Desarrollo de un módulo para la construcción colaborativa de objetos de aprendizaje en la herramienta de autor CRODA 2.0*. Habana: Universidad de las Ciencias Informáticas.

Montero, J. H. (2008). Las herramientas de autor en el proceso de producción de cursos en formato digital. *Revista de Medios y Educación* , 59-72.

PgAdmin. (2010). *PgAdmin PostgreSQL*. Recuperado el 4 de Febrero de 2016, de PgAdmin PostgreSQL: <http://www.pgadmin.org/>

Pressman, R. S. (2015). *Ingeniería del Software* (7ma edición ed.). México: McGraw-Hill.

Scribd. (2011). *PostgreSQL*. Recuperado el 1 de Febrero de 2016, de <http://es.scribd.com/doc/36570462/postgreSQL-investigacion>

SwinSQL. (2009). *SwinSQL-Data Migration Tool*. Recuperado el 29 de Enero de 2016, de SwinSQL-Data Migration Tool: <http://www.swissql.com/products/datamigration/data-migration.html>.

Tamargo, M. L. (2011). *Primeros pasos para adentrarnos en el mundo de los Data Warehouse*. Recuperado el 3 de Febrero de 2016, de [http://www.betsime.disaic.cu/secciones/tec\\_ja\\_04.htm](http://www.betsime.disaic.cu/secciones/tec_ja_04.htm).

UCI. (2012). *uci*. Recuperado el 12 de Enero de 2016, de uci: <http://www.uci.cu/>

Universidad Central de las Villas. (2015). *Grupos de Redes*. Recuperado el 2016, de <http://net.uclv.edu.cu/software-libre/guia-cubana-de-migracion-al-software-libre/view>

Universidad Politécnica de Valencia. (2012). *Los objetos de aprendizaje como recurso para la docencia universitaria: criterios para su elaboración*. Valencia.

Visual Paradigm. (2009). *Visual Paradigm*. Recuperado el 29 de Enero de 2016, de <http://www.visual-paradigm.com>

XML. (2015). *Consortium. Extensible Markup Language*. Recuperado el 2 de Febrero de 2016, de Consortium. Extensible Markup Language: <http://www.w3.org/XML/>

Zaragoza, A. (2010). *Migración de datos. Técnicas de migración de datos*. Recuperado el 5 de Febrero de 2016, de Migración de datos. Técnicas de migración de datos: <http://es.scribd.com/doc/51334306/Migracion-de-datos>.