

Universidad de las Ciencias Informáticas
Centro de Informatización Universitaria
CENIA



Título: Migración de datos para el Sistema de Gestión de Residencia.

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

Autores: Yusniel Remos Díaz
Rodolfo Betancourt Hidalgo

Tutores: Ing. Laritza González Marrero
Ing. Tony Javier Sánchez Olivera
Ing. Yoan Carlos Machado Espinosa

La Habana, Cuba
Junio, 2012

DECLARACIÓN DE AUTORÍA

Declaramos que somos los únicos autores de este trabajo y autorizamos al Centro de Informatización Universitaria de la Universidad de las Ciencias Informáticas; así como a dicho centro para que hagan el uso que estimen pertinente con este trabajo.

Para que así conste firmamos la presente a los ____ días del mes de _____ del año_____.

Yusniel Remos Díaz

Autor

Rodolfo Betancourt Hidalgo

Autor

Ing. Laritza González Marrero

Tutor

Ing. Tony Javier Sánchez Olivera

Tutor

Ing. Yoan Carlos Machado Espinosa

Tutor

DATOS DEL CONTACTO

Ing. Laritza González Marrero:

Graduada de Ingeniera Informática en la Universidad de las Ciencias Informáticas.

Correo electrónico: lgmarrero@uci.cu

Ing. Tony Javier Sánchez Olivera:

Graduado de Ingeniero Informático en la Universidad de las Ciencias Informáticas.

Correo electrónico: tjsanchez@uci.cu

Ing. Yoan Carlos Machado Espinosa:

Graduado de Ingeniero Informático en la Universidad de las Ciencias Informáticas.

Correo electrónico: ycmachado@uci.cu

AGRADECIMIENTOS

RODOLFO BETANCOURT HIDALGO

Quiero agradecerle a toda mi familia, principalmente a mi mamá y a mi papá porque gracias a todo el apoyo que me han dado durante estos cinco años, es que he podido llegar hasta aquí.

Quiero agradecer a mis tutores Yoan Carlos, Laritza y Tony por toda la ayuda y apoyo que me han proporcionado, sin los cuales no habría sido posible realizar este trabajo.

Quiero agradecerles especialmente a mis amigos Rodaín, Yuni, Reynel, Zoe, Maryanís y Adriél, porque siempre me ayudaron cuando los necesité sin poner peros o evasiones.

AGRADECIMIENTOS

YUSNIEL REMOS DÍAZ

Quiero agradecer en primer lugar a mi mamá por haberme dado la vida y por estar siempre a mi lado, a mi papá por apoyarme en mis decisiones y por su confianza.

A mi prima por ese cariño tan grande que expresa en múltiples ocasiones, y que solo me hace quererla cada día más y a sus dos hijas por ser mi lucecita inspiradora.

A mis amigos Ramsés, Alex, Kike, Yordano y Oramis por ser amigos insustituibles, y compañeros de muchas alegrías y tristezas, no importa lo lejos que estemos, siempre seremos uno.

A mis compañeros del grupo 4 por su compañía en estos 5 años tan maravillosos de mi vida. A los del grupo 5 por haber compartido una parte de su vida conmigo.

A mis tutores porque sin su ayuda este trabajo no hubiese tenido un final, mucho menos feliz.

Al tribunal por darnos una oportunidad cuando todo parecía que se extendía.

A mis familiares y a todos los que han contribuido con mi formación en lo personal y en lo profesional.

A todos aquellos que me preguntaron alguna vez, ¿cómo va la tesis?

DEDICATORIA

Quiero dedicarle este trabajo a toda mi familia y especialmente a mi mamá y a mi papá.

Rodolfo

A mis padres por su dedicación, paciencia, confianza y por quererme tanto.

A mi prima la cual considero mi hermana y a sus dos hijas como sobrinas por su cariño insuperable.

Yusniel

RESUMEN

El presente trabajo lleva por título “Migración de datos para el Sistema de Gestión de Residencia” y surge con el objetivo de que el sistema cuente con el grosor histórico de los datos relacionados con el avituallamiento en la Universidad, brindando así la posibilidad de que este cuente con los datos referentes a los medios que poseen los estudiantes y profesores, los cuales se encuentran almacenados en archivos Excel en las lavanderías de la Universidad, por tales motivos el objetivo de este trabajo es desarrollar la migración de los datos relacionados con el avituallamiento en la Universidad, para la base de datos del Sistema de Gestión de Residencia. Para la realización de este trabajo se utilizó una estrategia elaborada a partir del estudio de las diferentes etapas de migración de datos ya existentes en el mundo, las cuales fueron redefinidas dependiendo de las características necesarias. También se hizo una plantilla para describir, recoger y organizar todo el proceso de migración. Se utilizó la herramienta Talend Open Studio para el apoyo de la migración de los datos, y se usó la herramienta Talend Open Profiler para la comprobación de los datos migrados.

PALABRAS CLAVES: migración, histórico, medios, datos, etapas, comprobación.

ÍNDICE DE CONTENIDO

Introducción	1
Capítulo 1. Fundamentación teórica.	6
1.1 Introducción	6
1.2 Base de Datos	6
1.3 Ventajas de una base de datos	8
1.3.1 Elementos de una base de datos relacional	11
1.4 Sistemas de Gestión de Base de Datos (SGBD)	11
1.5 Clasificación de los SGBD	13
1.6 Lenguaje de consulta SQL	14
1.7 Extracción, transformación y carga de datos (ETC)	15
1.8 Migración de datos	16
1.9 Etapas de la migración de datos	17
1.10 Aplicaciones para la migración de datos	18
1.11 Ambiente de desarrollo	21
1.12 Conclusiones	25
Capítulo 2. Descripción de la solución propuesta	26
2.1 Introducción	26
2.2 Descripción de la estrategia de migración	26
2.3 Análisis antes de la migración.	29
2.4 Desarrollo de la carga y migración de datos	39
2.5 Descripción de las funciones utilizadas en los scripts PHP	50
2.6 Conclusiones	51
Capítulo 3. Validación de la solución propuesta	52

ÍNDICE DE CONTENIDO

3.1 Introducción	52
3.2 Descripción del período de pruebas	52
3.3 Resumen de la cantidad de campos insertados en cada una de las tablas	63
3.4 Conclusiones.....	63
Conclusiones generales.....	64
Recomendaciones.....	65
Bibliografía.....	66
Glosario de términos	68

ÍNDICE DE FIGURAS

Figura 1. Archivo Excel donde se guardan los datos de la lavandería 1.	29
Figura 2. Archivo Excel donde se guardan los datos de la lavandería 2.	30
Figura 3. Archivo Excel donde se guardan los datos de la lavandería 4.	31
Figura 4. Archivo Excel donde se guardan los datos de la lavandería 3.	32
Figura 5. Homogeneización de la forma en que se guardaban los datos en los archivos Excel de las lavanderías.	35
Figura 6. Identificadores de los administradores de las lavanderías de la tabla tb_dadministradora_lavanderia.	53
Figura 7. Identificadores de las lavanderías de la tabla tb_dadministradora_lavanderia.....	54
Figura 8. Identificadores de las personas de la tabla lavanderias.	55
Figura 9. Identificadores de las personas de la tabla tb_dboleta_entrega.....	55
Figura 10. Identificadores de las lavanderías de la tabla lavanderias.	56
Figura 11. Identificadores de las lavanderías de la tabla tb_dboleta_entrega.	56
Figura 12. Identificadores de los administradores de las lavanderías de la tabla tb_dlavanderia...	57
Figura 13. Nombres de las lavanderías de la tabla tb_dlavanderia.	58
Figura 14. Nombres de las residencias de la tabla tb_dlavanderia.....	59
Figura 15. Identificadores de las boletas de entrega de la tabla tb_rmedios_boleta.....	59
Figura 16. Identificadores de los medios de la tabla tb_rmedios_boleta.....	60
Figura 17. Descripciones de los medios de la tabla tb_ntipo_medio.....	61
Figura 18. Nombres de los tipos de los medios de la tabla tb_ntipo_medio.	62

ÍNDICE DE TABLAS

Tabla 1. Análisis de los archivos Excel de las lavanderías antes de la carga.	29
Tabla 2. Análisis de las tablas del sistema antes de la migración.	32
Tabla 3. Análisis de las tablas intermedias creadas antes de la migración.....	35
Tabla 4. Carga de datos hacia la tabla lavanderia_1.	39
Tabla 5. Carga de datos hacia la tabla lavanderia_2.	40
Tabla 6. Carga de datos hacia la tabla lavanderia_3.	41
Tabla 7. Carga de datos hacia la tabla lavanderia_4.	43
Tabla 8 Transformación de la tabla lavanderia_1.	44
Tabla 9. Transformación de la tabla lavanderia_2.	45
Tabla 10. Transformación de la tabla lavanderia_3.	46
Tabla 11. Transformación de la tabla lavanderia_4.	48
Tabla 12. Migración de datos hacia la tabla tb_dboleta_entrega.	49
Tabla 13 Migración de datos hacia la tabla tb_rmedios_boleta.....	49
Tabla 14. Cantidad de campos insertados en cada tabla.....	63

INTRODUCCIÓN

Introducción

En la actualidad el mundo está inmerso en un acelerado desarrollo y especialmente lo están las Tecnologías de la Información y las Comunicaciones (TIC), alcanzando un elevado nivel de avance. Estos adelantos desempeñan un papel decisivo en las diversas actividades del ser humano, como son el comercio, la industria, la educación, las finanzas y la investigación; como consecuencia de esto el dominio y uso de la tecnología no es un privilegio, ya que se ha convertido en un requisito indispensable para que las personas y las empresas puedan ser más eficientes y competitivas.

La Universidad de las Ciencias Informáticas (UCI) fue creada con el objetivo de ser un centro de docencia, producción e investigación, lo cual ha posibilitado el desarrollo de varios proyectos, a través de los cuales, se están informatizando los principales procesos que se llevan a cabo en dicha entidad. En el Centro de Informatización Universitaria (CENIA) se desarrollan diversos proyectos productivos, tales como Postgrado, Pregrado, Cooperación Internacional, Investigaciones y Residencia, este último incluye los procesos vinculados con la convivencia y el bienestar de las personas que residen en la Universidad; este proyecto contiene los siguientes módulos: Alojamiento, Avituallamiento y Lavandería, Trabajo Educativo, Complejo Residencial (hotelito universitario) y Visitas a Profesores.

Como consecuencia de que este proyecto debe manipular una gran cantidad de información, que los datos tienen un elevado valor para la vida de los seres humanos y que por eso deben ser manejados con mucho cuidado, es necesario almacenarlos en una base de datos con lo cual se logra una mayor seguridad y rendimiento. Las bases de datos tienen una gran importancia porque permiten almacenar, administrar y recuperar información de manera eficiente y rápida. Son

INTRODUCCIÓN

usadas en muchas aplicaciones para impedir que el trabajo sea lento y engorroso. También de ellas se pueden obtener datos históricos acerca de los movimientos que se hayan hecho al manejar el sistema, con la meta de proporcionar buenos servicios a los clientes. Otras de las ventajas que brindan es que posibilitan el uso de restricciones de seguridad para mantener la integridad de la información, además permiten que se hagan reportes de forma rápida.

Los datos relacionados con el avituallamiento en la Universidad de las Ciencias Informáticas se encuentran guardados en varios archivos Excel; estos son de gran importancia porque incluyen los datos de los estudiantes y profesores que se encuentran en el centro.

El proceso de avituallamiento y lavandería se lleva a cabo en las lavanderías de la entidad. Este consiste en entregarle a los estudiantes o profesores los medios necesarios para su estancia en la residencia, como son: sábana, toalla, funda, colcha, almohada, mosquitero, sábana camera y cubre cama. El control de estos medios es llevado por cada uno de los administradores de las lavanderías en varios archivos Excel. Sin embargo no existe uniformidad en el manejo de esta información.

De manera que por un lado se tiene una aplicación web que permite la gestión de la información referente al avituallamiento, con su base de datos limpia, y por otro el conjunto de datos almacenados en archivos Excel; de lo cual emerge el siguiente **problema científico**: ¿Cómo tener en la base de datos del Sistema de Gestión de Residencia el grosor histórico de los datos relacionados con el avituallamiento en la Universidad?

Para darle solución a este problema se ha determinado como **objeto de estudio** la migración de datos, y el proceso de migración de la información hacia la base de datos del Sistema de Gestión de Residencia como **campo de acción**.

INTRODUCCIÓN

La investigación se sustenta sobre la siguiente **idea a defender**: El desarrollo de la migración de los datos hacia la base de datos del Sistema de Gestión de Residencia, permitirá al cliente contar con información histórica de gran utilidad para su trabajo, además con estadísticas más reales, gracias a los reportes que podrá realizar cuando la información ya esté en la base de datos.

Constituyendo el **objetivo general**: Desarrollar la migración de los datos relacionados con el avituallamiento en la Universidad, para la base de datos del Sistema de Gestión de Residencia.

Estableciendo además los siguientes **objetivos específicos**:

- Seleccionar las tecnologías que se van a usar para realizar la migración de los datos.
- Definir una estrategia para la migración de los datos.
- Migrar los datos para la base de datos del Sistema de Gestión de Residencia.
- Elaborar un plan de pruebas para verificar la correcta migración de los datos.

Para poder cumplir estos objetivos es necesario desarrollar el siguiente grupo de **tareas**:

- Estudio de la base de datos del Sistema de Gestión de Residencia.
- Estudio de alternativas de migración.
- Análisis de las herramientas de migración existentes.

INTRODUCCIÓN

- Migración de la información relacionada con el avituallamiento en la Universidad, para la base de datos del Sistema de Gestión de Residencia.
- Estudio de pruebas existentes para la comprobación de la integridad de los datos.
- Ejecución del plan de pruebas para la información migrada.
- Evaluación de los resultados.

Como **Métodos de Investigación Científica** se utilizaron:

Métodos Teóricos:

- **Histórico – lógico:** Este método permitió el análisis histórico del proceso de avituallamiento en la residencia, para conocer cuál es su situación antes de hacer la migración de los datos e identificar qué ventajas podría traer su ejecución.
- **Analítico sintético:** Este método posibilitó la recopilación de información sobre las migraciones de datos, el cual propició una mejor organización de la solución planteada, además de servir para comprender cómo funcionan los elementos que forman parte del proceso de migración.

Métodos empíricos:

Entrevista: Este método se empleó para comprender el significado de la información que se migró y para obtener los archivos en Excel donde se guardaba la misma, a través de reuniones con los administradores de las lavanderías de la Universidad.

Como **aportes prácticos esperados del trabajo** se han identificado:

- Recuperación bibliográfica sobre la migración de datos.

INTRODUCCIÓN

- Documentación asociada a la migración de los datos para la base de datos del Sistema de Gestión de Residencia.
- Datos migrados para la base de datos del Sistema de Gestión de Residencia.

Este trabajo consta esencialmente de 4 partes: Resumen, Introducción, Desarrollo y Conclusiones, el Desarrollo está dividido en 3 capítulos:

Capítulo 1: Fundamentación Teórica

En este capítulo se abordan de forma detallada los principales conceptos que se deben tener en cuenta para la realización del trabajo de diploma, las principales herramientas y plataformas utilizadas, así como los principales aspectos teóricos empleados para darle solución al problema.

Capítulo 2: Descripción de la solución propuesta

En dicho capítulo se lleva a cabo un análisis de la solución propuesta para realizar la migración de datos. Se describe la migración paso a paso teniendo en cuenta los problemas identificados previamente.

Capítulo 3: Validación de la solución propuesta

En este capítulo se incluye la validación del diseño de la migración y un resumen de la información que será trasladada hacia el Sistema de Gestión de Residencia.

Capítulo 1. Fundamentación teórica.

1.1 Introducción

En este capítulo se explican los conceptos asociados al dominio del problema, las herramientas, la tecnología y el lenguaje de programación empleado para resolverlo.

1.2 Base de Datos

Hay varias maneras de definir el concepto de base de datos, a continuación se plantean algunas de las que se consideran correctas:

- Es posible considerar a la propia base de datos como una especie de armario electrónico para archivar; es decir, es un depósito o contenedor de una colección de archivos de datos computarizados. (Date, 2001)
- Una base de datos es un conjunto de datos persistentes que es utilizado por los sistemas de aplicación de alguna empresa dada. (Date, 2001)
- Conjunto de datos interrelacionados entre sí, almacenados con carácter más o menos permanente en la computadora. O sea, que una base de datos puede considerarse una colección de datos variables en el tiempo. (García, 1999)

La clasificación de las bases de datos se realiza teniendo en cuenta las características de los datos que se van a guardar y los procedimientos a ejecutar con ellos:

- **Base de datos relacional:** Éste es el modelo más utilizado en la actualidad para modelar problemas reales y administrar datos dinámicamente. Tras

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

ser postulados sus fundamentos en 1970 por Edgar Frank Codd, de los laboratorios IBM en San José (California), no tardó en consolidarse como un nuevo paradigma en los modelos de bases de datos. Su idea fundamental es el uso de "relaciones". Estas relaciones podrían considerarse en forma lógica como conjuntos de datos llamados "tuplas". Pese a que ésta es la teoría de las bases de datos relacionales creadas por Edgar Frank Codd, la mayoría de las veces se conceptualiza de una manera más fácil de imaginar. Esto es pensando en cada relación como si fuese una tabla que está compuesta por registros (las filas de una tabla), que representarían las tuplas, y campos (las columnas de una tabla). En este modelo, el lugar y la forma en que se almacenen los datos no tienen relevancia (a diferencia de otros modelos como el jerárquico y el de red). Esto tiene la considerable ventaja de que es más fácil de entender y de utilizar para un usuario esporádico de la base de datos. La información puede ser recuperada o almacenada mediante consultas que ofrecen una amplia flexibilidad y poder para administrar la información. El lenguaje más habitual para construir las consultas en bases de datos relacionales es SQL, *Structured Query Language* o Lenguaje Estructurado de Consultas, un estándar implementado por los principales motores o sistemas de gestión de bases de datos relacionales. Durante su diseño, una base de datos relacional pasa por un proceso al que se le conoce como normalización de una base de datos. Durante los años 80 (1980-1989) la aparición de dBase produjo una revolución en los lenguajes de programación y sistemas de administración de datos. Aunque nunca debe olvidarse que dBase no utilizaba SQL como lenguaje base para su gestión.

- **Bases de datos Objeto-Relacionales:** Los modelos de datos relacionales orientados a objetos extienden el modelo de datos relacional

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

proporcionando un sistema de tipos más rico y que se adapte mejor a las nuevas aplicaciones, que requieren guardar tipos de datos más complejos. Permiten que los atributos de las tuplas tengan tipos complejos. Además, la base de datos es relacional por lo que conserva su rapidez y eficiencia, y permite hacer uso de nuevos elementos, como las relaciones de herencia, que modelan los objetos que serán guardados en dicha base de datos, por lo que el analista y el diseñador ven un modelo orientado a objetos. (Sánchez, 2007)

1.3 Ventajas de una base de datos

El empleo de bases de datos como plataforma para desarrollar sistemas de aplicación ha aumentado considerablemente con el paso del tiempo, lo cual ha sucedido por las variadas ventajas que brinda su uso, a las que se hace referencia a continuación:

- **Los datos pueden compartirse:** Compartir no sólo significa que las aplicaciones existentes puedan compartir la información de la base de datos, sino también que sea posible desarrollar nuevas aplicaciones para operar sobre los mismos datos. En otras palabras, es posible satisfacer los requerimientos de datos de aplicaciones nuevas sin tener que agregar información a la base de datos.
- **Es posible reducir la redundancia:** En sistemas que no son de bases de datos, cada aplicación tiene sus propios archivos exclusivos. A menudo este hecho puede conducir a una redundancia considerable de los datos almacenados, con el consecuente desperdicio de espacio de almacenamiento.

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

- **Es posible evitar la inconsistencia:** Resulta claro que una base de datos en un estado inconsistente es capaz de proporcionar a sus usuarios información incorrecta o contradictoria. Como alternativa; si no se elimina la redundancia pero se controla (haciéndola del conocimiento del SGBD), entonces el SGBD puede garantizar que la base de datos nunca será inconsistente a los ojos del usuario.
- **Es posible brindar un manejo de transacciones:** Una transacción es una unidad de trabajo lógica, que por lo regular comprende varias operaciones de la base de datos (en particular, varias operaciones de actualización).
- **Es posible mantener la integridad:** El problema de la integridad es el de asegurar que los datos de la base de datos estén correctos. La inconsistencia entre dos entradas que pretenden representar el mismo "hecho" es un ejemplo de la falta de integridad; desde luego, este problema en particular puede surgir sólo si existe redundancia en los datos almacenados. No obstante, aun cuando no exista redundancia, la base de datos podría seguir conteniendo información incorrecta. El control centralizado de la base de datos puede ayudar a evitar estos problemas (en la medida de lo posible) permitiendo que el administrador de bases de datos implemente las restricciones de integridad (también conocidas como reglas del negocio) que serán verificadas siempre que se realice una operación de actualización.
- **Es posible hacer cumplir la seguridad:** Al tener la completa jurisdicción sobre la base de datos, el administrador de bases de datos puede asegurar que el único medio de acceso a la base de datos sea a través de los canales adecuados y por lo tanto puede definir las reglas o restricciones de seguridad que serán verificadas siempre que se intente acceder a datos sensibles. Es posible establecer diferentes restricciones para cada tipo de

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

acceso (recuperación, inserción, eliminación, etcétera) para cada parte de la información de la base de datos. Sin embargo, sin dichas restricciones la seguridad de los datos podría de hecho estar en mayor riesgo que en un sistema de archivos tradicionales (dispersos); es decir, la naturaleza centralizada de un sistema de base de datos requiere, en cierto sentido, que también sea establecido un buen sistema de seguridad.

- **Es posible equilibrar los requerimientos en conflicto:** Al conocer los requerimientos generales de la empresa (a diferencia de los requerimientos de los usuarios individuales), el administrador de bases de datos puede estructurar los sistemas de manera que ofrezcan un servicio general que sea "el mejor para la empresa". Por ejemplo, es posible elegir una representación física de los datos almacenados que proporcione un acceso rápido para las aplicaciones más importantes (posiblemente a costa de un acceso más lento para otras aplicaciones).
- **Es posible hacer cumplir los estándares:** Con el control central de la base de datos, el administrador de bases de datos puede asegurar que todos los estándares aplicables en la representación de los datos sean observados. Estos estándares podrían incluir alguno o todos los siguientes: departamentales, de instalación, corporativos, de la industria, nacionales e internacionales. Es conveniente estandarizar la representación de datos, en particular como un auxiliar para el intercambio de datos o para el movimiento de datos entre sistemas (esta consideración se ha vuelto particularmente importante con el advenimiento de los sistemas distribuidos. En forma similar, los estándares en la asignación de nombres y en la documentación de los datos también son muy convenientes como una ayuda para compartir y entender los datos. (Date, 2001)

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

1.3.1 Elementos de una base de datos relacional

Las siguientes son definiciones necesarias para manejar con claridad las bases de datos relacionales:

- **Tupla:** Se trata de cada una de las filas de la tabla. Es importante señalar que no se pueden tener tuplas duplicadas en una tabla.
- **Atributo:** Se trata de cada una de las columnas de la tabla.
- **Dominio:** El dominio dentro de la estructura del modelo relacional es el conjunto de valores que puede tomar un atributo.
- **Relación:** La relación se representa mediante una tabla, la cual contiene los atributos (columnas) y las tuplas (filas).
- **Vista:** Se trata de una tabla ficticia la cual muestra atributos de otras tablas relacionadas.
- **Llave primaria:** Es el valor o conjunto de valores que identifican una fila dentro de una tabla.
- **Llave foránea:** Es el valor o valores de una tabla que corresponde con el valor de una clave primaria en otra tabla. (Álvarez, 2007)
- **Llave:** Es un atributo o conjunto de atributos cuyo valor es único y diferente para cada tupla.
- **Cardinalidad:** Número de tuplas que tiene una tabla.
- **Grado:** Es el número de campos en una tabla. (Uni12)

1.4 Sistemas de Gestión de Base de Datos (SGBD)

Existen diversas formas de definir el concepto de Sistemas de Gestión de Base de Datos (SGBD), de las cuales se muestran algunas a continuación:

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

- 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 (SGBD). (García, 1999)
- El SGBD (sistema de gestión de bases de datos) es el software que maneja todo acceso a la base de datos. (Date, 2001)
- Un sistema gestor de base de datos se define como el conjunto de programas que administran y gestionan la información contenida en una base de datos. (Alvarez, 2007)

Los Sistemas de Gestión de Base de Datos (SGBD) tienen entre sus objetivos principales los siguientes:

- **Independencia:** La independencia de los datos consiste en la capacidad de modificar el esquema (físico o lógico) de una base de datos sin tener que realizar cambios en las aplicaciones que se sirven de ella.
- **Consistencia:** En aquellos casos en los que no se ha logrado eliminar la redundancia, será necesario vigilar que aquella información que aparece repetida se actualice de forma coherente, es decir, que todos los datos repetidos se actualicen de forma simultánea. Por otra parte, la base de datos representa una realidad determinada que tiene determinadas condiciones.
- **Seguridad:** La información almacenada en una base de datos puede llegar a tener un gran valor. Los sistemas de gestión de bases de datos deben garantizar que esta información se encuentre segura.
- **Manejo de transacciones:** Una transacción es un programa que se ejecuta como una sola operación. Los sistemas de gestión de bases de datos proveen mecanismos para programar las modificaciones de los datos.

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

- **Tiempo de respuesta:** Lógicamente, es deseable minimizar el tiempo que el sistema de gestión de bases de datos tarda en dar la información solicitada y en almacenar los cambios realizados. (Date, 2001)

1.5 Clasificación de los SGBD

Los SGBD pueden ser libres, propietarios o con licenciamiento dual:

Ejemplos de SGBD Proprietarios:

- Oracle (Oracle Corp.).
- DB2 (IBM).
- InterBase (CodeGear).
- MaxDB (SAP).
- SQL Server (Microsoft Corp.).
- Informix (IBM).

Ejemplos de SGBD Libres:

- PostgreSQL.
- FireBird.
- SQLite.

Ejemplo de SGBD con licenciamiento dual:

- MySQL (Oracle Corp.).

Licenciamiento dual: Es el tipo de licenciamiento que por un lado brinda la licencia GNU GPL para la utilización compatible con la misma y por el otro proporciona la licencia propietaria, para las compañías interesadas en incluir el SGBD con esta forma de licenciamiento en productos privativos. Este tipo de

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

licencia es usada como herramienta de marketing para la adopción generalizada de productos. (Albert)

1.6 Lenguaje de consulta SQL.

Los Sistemas de Gestión de Base de Datos (SGBD) utilizan para el acceso a bases de datos el lenguaje de consulta estructurado o SQL (por sus siglas en inglés *Structured Query Language*), el cual es un lenguaje declarativo de acceso a bases de datos relacionales que permite especificar diversos tipos de operaciones en éstas. Una de sus características es el manejo del álgebra y el cálculo relacional permitiendo efectuar consultas con el fin de recuperar de una forma sencilla información de interés de una base de datos, así como también hacer cambios sobre ella. Como su nombre lo indica SQL permite realizar consultas a la base de datos, además realiza funciones de definición, control y gestión de la base de datos. Las sentencias SQL se clasifican según su finalidad dando origen a tres sub lenguajes:

- El **DDL** (*Data Definition Language*), lenguaje de definición de datos, incluye órdenes para definir, modificar o borrar las tablas que componen la base de datos, así como los índices, vistas, permisos, etc. que pudieran definirse sobre las mismas.
- El **DCL** (*Data Control Language*), lenguaje de control de datos, contiene elementos útiles para trabajar en un entorno multiusuario, en el que es importante la protección de los datos, la seguridad de las tablas y el establecimiento de restricciones en el acceso.
- El **DML** (*Data Manipulation Language*), lenguaje de manipulación de datos que permite recuperar los datos almacenados en la base de datos y también incluye órdenes para permitir al usuario actualizar la base de datos

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

añadiendo nuevos datos, suprimiendo datos antiguos o modificando datos previamente almacenados. (Date, 2001)

1.7 Extracción, transformación y carga de datos (ETC)

ETC son siglas que significan Extraer, Transformar y Cargar. Es el proceso que organiza el flujo de los datos entre diferentes sistemas en una organización y aporta los métodos y herramientas necesarias para mover datos desde múltiples fuentes a un almacén de datos, reformatearlos, limpiarlos y cargarlos en otra base de datos. La idea es que una aplicación ETC lea los datos primarios de unas bases de datos de sistemas principales, realice transformación, validación, el proceso cualitativo, filtración y al final escriba datos en el almacén y en este momento los datos son disponibles para analizar por los usuarios.

Extracción: Consiste en extraer los datos desde los sistemas de origen. Analiza los datos extraídos, lo que resulta un chequeo que verifica si estos datos cumplen la pauta o estructura que se esperaba y los convierte a un formato listo para iniciar el proceso de transformación. De no cumplirse esto, los datos son rechazados. Un requerimiento importante en esta fase es que cause un impacto mínimo en el sistema origen, lo que quiere decir que si los datos a extraer son muchos, el sistema de origen podría ralentizarse o colapsar provocando que no pueda utilizarse con normalidad.

Transformación: En esta fase se aplican una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados.

Carga: La fase de carga es el momento en el cual los datos de la fase anterior (transformación) son cargados en el sistema destino, sobrescribiendo la

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

información antigua por datos nuevos o manteniendo un historial de los registros, esto depende del tipo de base de dato. (Bermúdez, 2009)

1.8 Migración de datos

La migración de datos es el proceso a través del cual grandes magnitudes de datos que se encuentran almacenados, son trasladados hacia otro sitio de destino. Con mucha frecuencia se lleva a cabo la migración de datos debido a la necesidad de las entidades de cambiar el sistema de información que estaba usando, en el momento que determinan seleccionar otro sistema gestor de base de datos, o en el tiempo que se modernizan los sistemas y entonces los datos no son compatibles con el sistema nuevo, o cuando se le realiza una mejora a una nueva versión del sistema. Los tipos de migración de datos se pueden clasificar de dos formas atendiendo a su tiempo de uso o ejecución:

- **Única migración:** Es un proceso de migración de datos que se lleva a cabo una sola vez, sin importar la razón por la cual se esté realizando la migración (cambio de aplicación, cambio de base de datos, cambio de hardware, etc.). Solo se realizará una vez dentro de la organización, sin incluir las pruebas del proceso que se realizaron, previo a la migración del sistema.
- **Migración continua:** Procedimientos de migración de datos que se realizan continuamente dentro de la organización. Estas migraciones continuas se pueden presentar a través de interfaces entre sistemas que necesitan comunicar información continuamente entre ellos.

Con el objetivo de ejecutar correctamente la migración de datos es indispensable emplear varias técnicas como las que a continuación se mencionan:

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

- **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.
- **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:** Algunas plataformas no soportan algunos tipos de datos, así que es necesario planificar el mapeo de los campos en la nueva base de 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 evitaremos el reemplazo automático de caracteres o en su caso, pérdida de los mismos. (Bermúdez, 2009)

1.9 Etapas de la migración de datos

1. **Levantamiento de información:** Se realiza un análisis en la fuente de origen, de la información existente definiendo cuál es requerida por el Sistema de Gestión de Residencia.
2. **Mapeo de datos:** Se mapearon los campos entre las fuentes de origen y destino.

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

3. **Empleo del ETC (Extractor, Transformador y Cargador de datos):** Se utiliza la herramienta ETC, para cargar los datos de los archivos en Excel hacia las tablas de almacenamiento intermedio.
4. **Análisis de Calidad:** Se verifica la calidad de los datos.
5. **Transformación:** Se agregaron los campos necesarios en las tablas de almacenamiento intermedio en las que se insertaron valores que se obtuvieron de otras tablas del sistema mediante consultas y funciones.
6. **Limpieza de datos:** Contempla la depuración de los datos en las tablas intermedias.
7. **Migración de datos:** Se realiza la migración de los datos desde las tablas de almacenamiento intermedio hacia las del sistema.
8. **Pruebas a los datos migrados:** Se realizan las pruebas de la migración, determinando que la información sea traspasada con éxito hacia la fuente de destino.

1.10 Aplicaciones para la migración de datos

Para migrar datos se pueden utilizar varias aplicaciones, pero atendiendo a la cantidad de documentación que tienen publicada y a que se encuentran entre las más usadas, se decidió analizar las que se plantean a continuación, para llevar a cabo la migración de los datos.

Database Migration Toolkit Professional

Es un programa que permite convertir entre formatos de bases de datos diferentes. Ofrece un sencillo asistente paso a paso, que permite conectarse a la base de datos de origen, seleccionar las tablas, y luego convertir y copiar al destino. Permite optar por incluir todos los datos o simplemente copiar la

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

estructura de la base de datos, brinda todas las opciones de conversión necesarias, teniendo en cuenta las peculiaridades de los dos formatos de entrada y salida de base de datos. Esta aplicación tiene entre sus características las siguientes:

- Posee modo de conversión por lotes que puede mejorar la productividad, acelerando el proceso de conversión.
- Se destaca por ser el más intuitivo y una herramienta completa y eficaz en su categoría.
- Reduce el esfuerzo, coste y riesgo de migración.
- Soporta una serie de Sistemas de Base de Datos (SBD) como son Oracle, MySQL, SQL Server, PostgreSQL, IBM DB2, IBM Informix, InterSystems Caché, Teradata, Visual Foxpro, SQLite, FireBird, InterBase, Microsoft Access, Paradox, Lotus, dBase. (EAS12)

Softek Transparent Data Migration Windows UNIX (IP)

Es un producto de migración de datos diseñado para trasladar datos. Al proporcionar una metodología y plataforma de migración. Esta herramienta es una solución acreditada para entornos heterogéneos y resulta ideal para renovar tecnologías, consolidar servidores o dispositivos de almacenamiento, consolidar o reubicar centros de datos, y para implantar el almacenamiento por niveles. Este programa tiene entre sus características las siguientes:

- Realiza migraciones de datos en segundo plano mientras mantiene la disponibilidad de las aplicaciones.
- Garantiza la integridad de los datos y permite realizar copias puntuales de datos para que estén disponibles en cualquier momento tras la sincronización inicial.
- Es parte de la metodología de migración de Softek. (IBM12)

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

Talend Open Studio

Es una herramienta basada en código abierto de integración de datos, aprovecha el sistema código abierto para poner la integración de datos a disposición de todos los tipos de organizaciones de todos los tamaños, todos los niveles de pericia y de presupuesto. Ofrece soluciones de calidad de datos, completamente complementarias con sus soluciones de integración de datos. Las soluciones de integración de datos de Talend están optimizadas por la migración de datos de la empresa. Las características siguientes son especialmente críticas al diseño, al desarrollo y a la ejecución de los procesos de migración de datos:

- Proceso de modelado orientado a procesos de negocios que implica asegurar la coordinación apropiada durante la migración de los datos y de los procesos del negocio.
- El entorno de desarrollo completamente gráfico mejora y facilita la productividad, permite la reutilización de las asignaciones y de las transformaciones de datos para los procesos de sincronización.
- Plataforma altamente escalable y rápida de ejecución con un método que permite la elaboración de código de datos, usando los métodos de ETC (Extracción, Transformación y Carga de datos).

Características principales:

- Plataformas: Se ejecuta en Windows, Unix y Linux.
- Código fuente: El código fuente Java/Eclipse está disponible para su descarga y personalización.
- Soporte: Wiki y foro de Talend.
- Bases de datos: MySQL, MS SQL Server, DB2, Oracle, Ingres, PostgreSQL, Sybase, MS Access, Informix, Firebird. (José, 2009)

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

Se propone emplear la aplicación *Talend Open Studio* para migrar los datos hacia el Sistema de Gestión de Residencia, debido a la simplicidad que proporciona al usarse, a que es multiplataforma. También porque tiene una comunidad, la cual permite que se detecten errores en poco tiempo, que se les den solución a estos y que se hagan llegar dichas soluciones a toda la comunidad, a través de los foros utilizados para ello, de esta forma se ofrece un espacio a través del cual se puede recibir ayuda y soporte gratuitamente. Además es una aplicación libre que no cuenta con una licencia comercial, implicando esto un ahorro económico y la posibilidad de migrar datos correctamente para las entidades, sin afectar su presupuesto y por los abundantes conocimientos con que se cuenta en el centro acerca de su funcionamiento. Como consecuencia de los elementos planteados anteriormente se considera que el *Talend Open Studio* es la herramienta que se va a utilizar para realizar la migración de los datos hacia el Sistema de Gestión de Residencia.

1.11 Ambiente de desarrollo

En este epígrafe se elabora una exposición del gestor de base de datos, el lenguaje de programación y las herramientas utilizadas:

- **IDE (Entorno de Desarrollo Integrado) Netbeans 6.8 o superior:** Es una herramienta para programadores creada para escribir, compilar, depurar y ejecutar programas. Está escrito en Java, pero soporta otros lenguajes de programación entre los que se encuentran: Ruby, C/C++, PHP y Groovy. Existe además un número importante de módulos para extender el IDE NetBeans que es un producto libre y gratuito sin restricciones de uso.
- **Gestor de base de datos PostgreSQL 8.4.11:** Es un sistema del tipo objeto-relacional que está distribuido bajo la licencia BSD (*Berkeley*

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

Software Distribution) y con su código fuente disponible libremente. Es totalmente compatible con ACID (por sus siglas en inglés *Atomicity, Consistency, Isolation and Durability*) que en español significa Atomicidad, Consistencia, Aislamiento y Durabilidad, es la propiedad que debe cumplir una base de datos para realizar transacciones seguras. Este sistema funciona muy bien con grandes cantidades de datos y una alta concurrencia de usuarios accediendo al sistema. Su desarrollo comenzó hace más de 16 años, y durante este tiempo, estabilidad, potencia, robustez, facilidad de administración e implementación de estándares han sido las características que más se han tenido en cuenta durante su desarrollo. Esto hace posible contar con una amplia documentación pública, organizada, libre y en diferentes idiomas. Funciona en los principales sistemas operativos, incluyendo Linux, UNIX (AIX, BSD, HP-UX, SGI IRIX, Mac OS X, Solaris, Tru64) y Windows. A continuación se muestran algunas de las características más importantes y soportadas por PostgreSQL:

- ✓ Integridad referencial.
- ✓ Unicode.
- ✓ Múltiples métodos de autenticación.
- ✓ Uso de certificados SSL para autenticar a usuarios.
- ✓ Columnas auto-incrementales.
- ✓ Llaves primarias y foráneas.
- ✓ Joins.
- ✓ Vistas.
- ✓ Disparadores comunes, por columna, condicionales.
- ✓ Reglas.

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

Una de las mejoras más populares de esta versión es:

- ✓ Restauración de bases de datos en procesos paralelos, acelerando la recuperación de un respaldo hasta en ocho veces respecto a la versión anterior. (Casanova, 2009)
- **Talend Open Studio 4.1.0:** El Talend Open Studio es una herramienta de integración y gestión de datos, así como integración de aplicaciones empresariales, es decir una herramienta ETC; está basado en Java y puede ser ejecutado en Windows y Linux sin mayor dificultad.
- **PgAdmin III 1.12.2:** El PgAdmin III es una herramienta de código abierto para la administración de bases de datos PostgreSQL; fue diseñado para responder a las necesidades de la mayoría de los usuarios, desde escribir simples consultas SQL hasta desarrollar bases de datos complejas. La interfaz gráfica soporta todas las características de PostgreSQL y hace simple la administración; está disponible en varios sistemas operativos, incluyendo Microsoft Windows, Linux, FreeBSD, Mac OSX y Solaris.
- **Lenguaje de programación PHP 5.3.5:** El lenguaje de programación utilizado es PHP (acrónimo de PHP: *Hypertext Preprocessor*), es un lenguaje de código abierto muy popular especialmente adecuado para desarrollo web y que puede ser incrustado en HTML. PHP puede ser usado en los sistemas operativos que más se utilizan actualmente, como son Windows, Mac OS X y Linux, es decir, con PHP se tiene la posibilidad de seleccionar el sistema operativo deseado. Como la sintaxis de PHP es parecida a la del lenguaje C, cualquier persona que tenga experiencia trabajando con este tipo de lenguaje puede aprender PHP con más facilidad, lo cual es muy ventajoso porque se ahorra bastante tiempo al aprender un lenguaje similar a uno que ya se conoce, respecto a iniciarse en un lenguaje totalmente nuevo y desconocido. Una de sus características

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

más potentes es su soporte para gran cantidad de bases de datos. Entre su soporte pueden mencionarse InterBase, mSQL, MySQL, Oracle, Informix y PostgreSQL. PHP está escrito en C, por eso se ejecuta rápidamente haciendo uso de poca memoria. Como producto de código abierto y gracias a su amplia distribución en el mundo, PHP cuenta con la ayuda de una gran comunidad de programadores, esto permite que los fallos de funcionamiento se encuentren y se reparen de forma rápida; su código se pone al día muy a menudo con mejoras y extensiones del lenguaje para ampliar sus capacidades.

- **Talend Open Profiler 4.2.1:** Es una herramienta de código abierto que permite examinar las diferentes fuentes de datos existentes (por ejemplo bases de datos, aplicaciones, archivos.) y recoger las estadísticas y la información sobre estos datos para luego mostrar los resultados mediante informes o gráficos sofisticados. Esta herramienta permite la evaluación del nivel de calidad de los datos contenidos en el sistema de información.
- **Lenguaje de consulta SQL 2008:** Para el acceso a bases de datos los Sistemas de Gestión de Base de Datos (SGBD) utilizan el lenguaje de consulta estructurado o SQL (por sus siglas en inglés *Structured Query Language*), el cual es un lenguaje declarativo de acceso a bases de datos relacionales.

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

1.12 Conclusiones

En este capítulo se hizo referencia a los principales aspectos relacionados con la investigación, luego de realizar un estudio de las tecnologías, de haber plasmado las características fundamentales de cada una de ellas y teniendo en cuenta la estructura de la Universidad de las Ciencias Informáticas en cuanto a las tecnologías que se utilizan, se llegó a la conclusión de que entre las herramientas estudiadas para ejecutar la carga de datos, la más indicada para ser usada es el *Talend Open Studio*, como resultado de todas las facilidades que ofrece, lo cual influye directamente en la correcta migración de los datos y para el resto de las tareas las siguientes: el *Netbeans*, el *PgAdmin* y el *Talend Open Profiler*.

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

Capítulo 2. Descripción de la solución propuesta

2.1 Introducción

En este capítulo se plantea la propuesta de solución para la ejecución de la migración de la información hacia la base de datos del Sistema de Gestión de Residencia, así como la descripción de la forma en que se va a realizar.

2.2 Descripción de la estrategia de migración

Para el desarrollo de la migración de los datos se tomó en consideración la investigación realizada y se aplicó una estrategia que fue guiada por las etapas definidas en el capítulo anterior, donde se planteó como primera etapa el levantamiento de la información.

En esta etapa la labor realizada fue la búsqueda de la información, en donde se lleva a cabo el control de los medios que poseen los estudiantes y profesores de la Universidad, para ver cómo y dónde se encuentra recogida la misma. Además se analizaron los campos en la fuente de origen y en la fuente de destino, identificando los valores por defecto, nulos y las tablas del destino de los datos, llegando así a la conclusión de cuáles serían los campos críticos en las dos fuentes, necesarios para crear varias tablas de almacenamiento intermedio donde cargar toda la información de los archivos Excel, para su posterior migración hacia las tablas de la base de datos del sistema. Después se describieron los campos con los que cuenta el sistema y se analizaron los diferentes tipos de datos, garantizando el almacenamiento de la información. Luego se homogeneizó la forma en que se guardaban los datos en los archivos Excel, para realizar el mapeo

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

y la carga de datos, debido a que en todas las lavanderías esto se hacía de formas diferentes, donde se separaron los nombres y apellidos que se encontraban en la misma celda, dejando solo un atributo por cada una, logrando así una compatibilidad entre ambas fuentes a la hora de migrar los datos. Terminado el levantamiento de la información, conociendo la definición de los datos en la fuente de origen y en la fuente de destino, se prosiguió con el mapeo de datos, donde estos se enlazaron entre la fuente de origen y las tablas de almacenamiento intermedio, después se empleó la herramienta Talend Open Studio, para la carga de datos desde la fuente de origen hacia las tablas intermedias. Finalizada la importación se revisaron todas las tablas, controlando que la cantidad de datos insertados, fuese la misma que el número de datos en la fuente de origen.

Una vez realizada la carga se continuó con el análisis de la calidad de los datos, donde se ejecutaron consultas para verificar que no se encontraran personas repetidas, en una o en varias lavanderías, que no tuvieran medios y que no existieran filas vacías. Además se comprobó que la cantidad correspondiente a cada medio, no sobrepasara el número máximo permitido, como por ejemplo alguna persona con más de dos sábanas o más de una colcha.

Se agregaron los campos necesarios en las tablas de almacenamiento intermedio, en las que se insertaron valores que se obtuvieron de otras tablas del sistema mediante consultas y funciones. Cuando se hizo la búsqueda de los identificadores de persona mediante la función `fn_buscar_persona_avanzada`, se examinó este campo y se verificó que no se pudieron insertar todos, debido a que se detectaron nombres y/o apellidos mal escritos, incompletos o desordenados por lo que no coincidían con los de la tabla `tb_dpessoa_data_plus`. En estos casos se encontraron 1129 personas de un total de 8562, razón por la cual se llevó a cabo

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

la limpieza de los datos en las tablas intermedias. Algunas de las correcciones que se hicieron fueron las siguientes:

- La sustitución de datos como eliminar espacios entre caracteres.
- Adicionar algún carácter en alguna posición determinada.
- La corrección ortográfica de los nombres y apellidos.
- Rectificación del orden de los nombres y apellidos.

La corrección de esta información consumió bastante tiempo, como consecuencia de la cantidad de errores encontrados. Además se identificaron personas no registradas, que más tarde se registraron después de llegar a esa conclusión con los tutores.

Después de hacer estas actividades se obtuvieron todos los identificadores de las personas, que junto al resto de los datos ya estaban listos para ser migrados hacia la fuente de destino, lo cual se hizo a través de scripts PHP.

En la última fase con el objetivo de verificar que el número de valores migrados, se corresponden con los obtenidos se ejecutaron las pruebas, apoyados en la herramienta Talend Open Profiler. A continuación se describe el análisis antes de la migración.

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

2.3 Análisis antes de la migración.

Tabla 1. Análisis de los archivos Excel de las lavanderías antes de la carga.

Análisis de la fuente de origen antes de la carga	
Descripción de la fuente de origen:	
<p>En las lavanderías se recogen los nombres y apellidos de estudiantes y profesores de la Universidad. Así como los medios entregados a cada uno de ellos, dígame sábanas, colchas, almohadas, toallas, fundas, en el caso de los profesores se cuenta además con la cantidad de sábanas cameras, mosquiteros y cubre camas en archivos Excel, pero en todas se hace de manera diferente como se muestra en las siguientes figuras:</p>	

Figura 1. Archivo Excel donde se guardan los datos de la lavandería 1.



	A	B	C	D	E	F
	Nombres y apellidos	Sábana	Toalla	Funda	Colcha	Almohada
2	Aragón Vigoa, Dervis Marbel	2	1	1	1	1
3	Arbolález Pérez, Danailys	2	1	1	1	1
4	Alonso Tarafa, Thais	2	1	1	1	1
5	Chang Collado, Lienzt de la Caridad	2	1	1	1	1
6	Dussac Estupiñán, Laritza	2	1	1	1	
7	Hernández Romero, Selma	2	1	1	1	1
8	Hernández Peña, Rosmery	2	1	1	1	1
9	João Sáez, Claudia	2	1	1	1	1
10	Laurencio González, Elena	2	1	1	1	1
11	Muñoz Ruíz, Anabel					
12	Padilla Martínez, Johana	2	1	1	1	1
13	Rodríguez Sánchez, Grelenis Lirene	2	1	1	1	1
14	Rosell Aldana, Jeidy	2	1	1	1	1
15	Sepúlveda Delgado, Rossana	2	1	1	1	1
16	Valdés Cepero, Iria	2	1	1	1	1
17	Alvarez Carmona, Oslén	2	1	1	1	1
18	Alonso Benítez, Drymon	2	1	1	1	1
19	Alvarez Díaz, Diego Javier	2	1	1	1	1
20	Arrieta Romero, José Carlos	2	1	1	1	1
21	Alonso Peña, Carlos Andrés	2	1	1	1	1

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

Figura 2. Archivo Excel donde se guardan los datos de la lavandería 2.

QUINTO AÑO										
Facultad	1er apellido	2do apellido	1er nombre	2do nombre	sexo	Sabana	Fnda	Toalla	Colcha	Almohada
QUINTO AÑO										
Facultad 5	Antunez	Ojeda	Liset		Femenina	2	1	1	1	1
Facultad 5	Abraham	Ravelo	Jose	Carlos	Masculino	2	1	1	1	1
Facultad 5	Abreu	Fong	Patricia	Airela	Femenina	2	1	1	1	
Facultad 5	Ayala	Alonso	Juan	Carlos	Masculino	2	1	1	1	
Facultad 5	Arias	Moreno	Lisandra		Femenina					
Facultad 5	Armosto	Reyna	Jaqueline		Femenina	2	1	1	1	
Facultad 5	Areas	Arevalo	Juan	Carlos	Masculino					
Facultad 5	Aguilera	Diaz	Neurys	Eduardo	Masculino					
Facultad 5	Alvarez	Mesdell	Elieser		Masculino					
Facultad 5	Avila	Rodriguez	Ivan		Masculino	2	1	1	1	
Facultad 5	Alvarez	Yanes	Yissel		Femenina	2	1	1	1	
Facultad 5	Aleman	Siten	Ana	Laura	Femenina					
Facultad 5	Baltel	Caballero	Solangel		Masculino					
Facultad 5	Barrios	Gonzalez	Dianelys		Femenina					
Facultad 5	Baez	Garcia	Anibal		Masculino					
	Bolet	Sanchez	Arianna		Femenina	2	1	1	1	
Facultad 5	Bavastros	Montes De	Yamany		Masculino					
	Baeza	Orta	Yaimelys		Femenina	2	1	1	1	
Facultad 5	Blanco	Alvarez	Reldy		Masculino					
Facultad 5	Barrios	Matos	Yunior		Masculino					
Facultad 5	Barrios	Morales	Reinier	Felipe	Masculino					
	Barzaga	Chacon	Liliet		Femenina	2	1	1	1	
	Boteta	Vilva	Jaca	Manuel	Masculino	2	1	1	1	

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

Figura 3. Archivo Excel donde se guardan los datos de la lavandería 4.

	A	B	C	D	E	F	G	H	I
1	Nombre	2do nombre	Apellido	Apellido	S	T	F	C	A
2	Abel	José	Romero	Machado	2	1	1	1	1
3	Adolfo		Alonso	Corona	2	1	1	1	1
4	Adrián		Hernández	Barrios	2	1	1	1	0
5	Alberto	Daniel	Mesa	Sánchez	2	1	1	1	0
6	Alejandro		Mengana	Pérez	2	1	1	1	1
7	Alejandro		Vázquez	Chávez	2	1	1	1	1
8	Alexander		Lugones	Aguar	2	1	1	1	0
9	Alexei		Alayo	Rondón	2	1	1	1	1
10	Alisbet		Fernández	Rojas	2	1	1	1	1
11	Amed	Alcides	Cortina	Márquez	2	1	1	1	1
12	Anaeli		Romero	Saborit	2	1	1	1	1
13	Angel	Antonio	Sánchez	Lopez-Castro	2	1	1	1	0
14	Ariel		Fajardo	Delgado	2	1	1	1	1
15	Asiel		Almaguer	Trujillo	2	1	1	1	0
16	Beatriz		Maciá	Gonzalez	2	1	1	1	1
17	Carlos	Ivan	García	Delgado	2	1	1	1	0
18	Carlos	Manuel	Estrada	Fernández	2	1	1	1	1
19	Carlos	Manuel	Tamayo	Rodriguez	2	1	1	1	1
20	Carlos	Yusel	Carralero	Calzadilla	2	1	1	1	1
21	Dachel		Benitez	Abdo	2	1	1	1	1
22	Dailys		Torres	Martin	2	1	1	1	0
23	David		Fonseca	Martinez	2	1	1	1	1
24	David		Milanes	Vorobiov	2	1	1	1	1
25	David		Roman	Martinez	2	1	1	1	1
26	Dayán		Ruiz	Riera	2	1	1	1	0
27	Dayana		Pardo	Roque	2	1	1	1	0
28	Eduardo	Lazaro	Braña	Dominguez	2	1	1	1	1
29	Edwin		Ruiz	Llanes	2	1	1	1	0
30	Elaimy		Arias	Liemá	2	1	1	1	1

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

Figura 4. Archivo Excel donde se guardan los datos de la lavandería 3.

	A	B	C	D	E	F	G	H	I	J	K
1											
2			EDIFICIO 117		Residencia 4		Auditoria 05/06				
3	Apto	Nombre y Apellidos	Sab.	Sab c.	Toall.	Fund.	Fund.c	Colch.	Mosq.	Cubre C.	almo
4	101	JUAN PLACENCIAC		2	2	2		1		1	
5	102	YASEL COUSE SARDIÑAS	2		2	1		1		1	
6		AMANDA COUSE BETANCOURT	2		2	2		1			1
7		ODELKIS BETANCOUR OSVALDO	2		2	2		1			
8		KAREL GOMEZ	2		2	2		1		1	
9	103	MAIKEL YELANDI LEIVA	2		2	1		1			
10		KARINA PEREZ TERUEL	2		2	1		1			
11	104	WUILIAN AZCUY MORALES	2		2	2		1		1	1
12		YANEIKA MIRANDA NEGRIN	2		2	2		1		1	1
13	105	DISMEY SAAVEDRA LOPEZ			2	2		1			1
14	105	YUNIESKI ARMENTERO MORENO		2	2	2		1		1	1
15		YUDISLANDY AGUILA GLEZ	2		2	2		1		1	1
16	106	MAIKEL DAVID RUENES CORREA	2		1	2		1		1	
17	107	IYATNE MOMPIE	2		2	2				1	1
18		DACIEL OLIVERA CORTINA	2		2	1		1			
19	108	EDISEL NAVAS CONYEDO		2	2	1		1			
20		YAILEN COSTA MARRERO			2	1		1			
21		YANIRIS BLANCO ZAMORA	2		2	2		1		1	1
22		ERICK MARIO GOMEZ	2		2	2		1		1	

Tabla 2. Análisis de las tablas del sistema antes de la migración.

Análisis de la fuente de destino antes de la migración		
Nombre de la base de datos:		Sistema al que pertenece:
BD_sgu_residencia		Sistema de Gestión de Residencia
Descripción de la base de datos:		
<p>El Sistema de Gestión de Residencia, recoge todos los datos de los procesos residenciales que se llevan a cabo en la Universidad, ya sea por estudiantes, profesores o visitantes. Entre estos procesos se encuentran: alojamiento, avituallamiento y lavandería, complejo residencial, trabajo educativo y visitas a profesores. Los mismos contienen información esencial para la primera iteración de la migración.</p>		
Nombre de la tabla: tb_dboleta_entrega		
Atributo	Tipo de Dato	Descripción

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

id_boleta_entrega	int4	Identificador de la boleta entregada
fecha_entrega_boleta	timestamp(6)	Fecha en que se registró
id_persona	varchar(32)	Identificador de la persona
autorizado	varchar(50)	Persona autorizada a realizar la entrega
estado	int4	Estado de la entrega
id_lavanderia	int4	Identificador de la lavandería
activo	bool	Especifica si la boleta está activa
Cantidad de objetos a migrar:		8562
Nombre de la tabla: tb_rmedios_boleta		
Atributo	Tipo de Dato	Descripción
id_medio	int4	Identificador del medio entregado
id_boleta_entrega	int4	Identificador de la boleta entregada
cantidad	int4	Cantidad de un tipo de medio entregada
fecha	timestamp(6)	Fecha en que se registró
Cantidad de objetos a migrar:		39153
Nombre de la tabla: tb_dadministradora_lavanderia		
Atributo	Tipo de Dato	Descripción
id_persona	varchar(32)	Identificador de la persona
fecha	timestamp(6)	Fecha en que se registró
id_lavanderia	int4	Identificador de la lavandería
activo	bool	Especifica si la persona está activa
Cantidad de objetos a migrar:		4
Nombre de la tabla: tb_dlavanderia		
Atributo	Tipo de Dato	Descripción

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

id_persona	varchar(32)	Identificador de la persona
fecha	timestamp(6)	Fecha en que se registró
id_lavanderia	int4	Identificador de la lavandería
activo	bool	Especifica si la persona está activa
id_lavanderia	int4	Identificador de la lavandería
Cantidad de objetos a migrar:		4
Nombre de la tabla: tb_ntipo_medio		
Atributo	Tipo de Dato	Descripción
id_tipo_medio	int4	Identificador del tipo de medio
nombre_tipo_medio	varchar(30)	Nombre del tipo de medio
descripcion_medio	text	Descripción del medio
medio_activo	bool	Especifica si el medio está activo
fecha_registro_medio	timestamp(6)	Fecha en que se registró
fecha_modificado_medio	timestamp	Fecha en que se modificó
Cantidad de objetos a migrar:		8

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

Figura 5. Homogeneización de la forma en que se guardaban los datos en los archivos Excel de las lavanderías.

	A	B	C	D	E	F	G	H	I
1	1er nombre	2do nombre	1er apellido	2do apellido	S	T	F	C	A
2	Abel	José	Romero	Machado	2	1	1	1	1
3	Adolfo		Alonso	Corona	2	1	1	1	1
4	Adrián		Hernández	Barrios	2	1	1	1	0
5	Alberto	Daniel	Mesa	Sánchez	2	1	1	1	0
6	Alejandro		Mengana	Pérez	2	1	1	1	1
7	Alejandro		Vázquez	Chávez	2	1	1	1	1
8	Alexander		Lugones	Aguiar	2	1	1	1	0
9	Alexei		Alayo	Rondón	2	1	1	1	1
10	Alisbet		Fernández	Rojas	2	1	1	1	1
11	Amed	Alcides	Cortina	Márquez	2	1	1	1	1
12	Anaeli		Romero	Saborit	2	1	1	1	1
13	Angel	Antonio	Sánchez	Lopez-Castro	2	1	1	1	0
14	Ariel		Fajardo	Delgado	2	1	1	1	1
15	Asiel		Almaguer	Trujillo	2	1	1	1	0
16	Beatriz		Maciá	Gonzalez	2	1	1	1	1
17	Carlos	Ivan	García	Delgado	2	1	1	1	0

Tabla 3. Análisis de las tablas intermedias creadas antes de la migración.

Análisis de las tablas de almacenamiento intermedio antes de la migración		
Nombre de la base de datos:		Sistema al que pertenece:
BD_sgu_residencia		Sistema de Gestión de Residencia
Descripción de la base de datos:		
En las tablas de almacenamiento intermedio es donde se van a cargar los datos de las lavanderías.		
Nombre de la tabla intermedia: lavanderia_1		
Atributo	Tipo de Dato	Descripción
primer_apellido	varchar	Primer apellido
segundo_apellido	varchar	Segundo apellido

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

primer_nombre	varchar	Primer nombre
segundo_nombre	varchar	Segundo nombre
sabana	int4	Cantidad de sábanas recogidas
toalla	int4	Cantidad de toallas recogidas
funda	int4	Cantidad de fundas recogidas
colcha	int4	Cantidad de colchas recogidas
almohada	int4	Cantidad de almohadas recogidas
id_persona	varchar(32)	Identificador de la persona
id_s	int4	Identificador del medio sábana
id_t	int4	Identificador del medio toalla
id_f	int4	Identificador del medio funda
id_c	int4	Identificador del medio colcha
id_a	int4	Identificador del medio almohada
id_lavanderia	int4	Identificador de la lavandería
Cantidad de objetos a cargar:		1930
Nombre de la tabla intermedia: lavanderia_2		
Atributo	Tipo de Dato	Descripción
primer_apellido	varchar	Primer apellido
segundo_apellido	varchar	Segundo apellido
primer_nombre	varchar	Primer nombre
segundo_nombre	varchar	Segundo nombre
sabana	int4	Cantidad de sábanas recogidas
toalla	int4	Cantidad de toallas recogidas

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

funda	int4	Cantidad de fundas recogidas
colcha	int4	Cantidad de colchas recogidas
almohada	int4	Cantidad de almohadas recogidas
id_persona	varchar(32)	Identificador de la persona
id_s	int4	Identificador del medio sábana
id_t	int4	Identificador del medio toalla
id_f	int4	Identificador del medio funda
id_c	int4	Identificador del medio colcha
id_a	int4	Identificador del medio almohada
id_lavanderia	int4	Identificador de la lavandería
Cantidad de objetos a cargar:		2787
Nombre de la tabla intermedia: lavanderia_3		
Atributo	Tipo de Dato	Descripción
primer_apellido	varchar	Primer apellido
segundo_apellido	varchar	Segundo apellido
primer_nombre	varchar	Primer nombre
segundo_nombre	varchar	Segundo nombre
sabana	int4	Cantidad de sábanas recogidas
toalla	int4	Cantidad de toallas recogidas
funda	int4	Cantidad de fundas recogidas
colcha	int4	Cantidad de colchas recogidas
almohada	int4	Cantidad de almohadas recogidas
sabana_camera	int4	Cantidad de sábanas cameras recogidas

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

cubre_cama	int4	Cantidad de cubre camas recogidos
mosquitero	int4	Cantidad de mosquiteros recogidos
id_persona	varchar(32)	Identificador de la persona
id_s	int4	Identificador del medio sábana
id_t	int4	Identificador del medio toalla
id_f	int4	Identificador del medio funda
id_c	int4	Identificador del medio colcha
id_a	int4	Identificador del medio almohada
id_sc	int4	Identificador del medio sábana camera
id_cc	int4	Identificador del medio cubre cama
id_m	int4	Identificador del medio mosquitero
id_lavanderia	int4	Identificador de la lavandería
Cantidad de objetos a cargar:		2100
Nombre de la tabla intermedia: lavanderia_4		
Atributo	Tipo de Dato	Descripción
primer_apellido	varchar	Primer apellido
segundo_apellido	varchar	Segundo apellido
primer_nombre	varchar	Primer nombre
segundo_nombre	varchar	Segundo nombre
sabana	int4	Cantidad de sábanas recogidas
toalla	int4	Cantidad de toallas recogidas
funda	int4	Cantidad de fundas recogidas
colcha	int4	Cantidad de colchas recogidas

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

almohada	int4	Cantidad de almohadas recogidas
id_persona	varchar(32)	Identificador de la persona
id_s	int4	Identificador del medio sábana
id_t	int4	Identificador del medio toalla
id_f	int4	Identificador del medio funda
id_c	int4	Identificador del medio colcha
id_a	int4	Identificador del medio almohada
id_lavanderia	int4	Identificador de la lavandería
Cantidad de objetos a cargar:		1745

2.4 Desarrollo de la carga y migración de datos

Tabla 4. Carga de datos hacia la tabla lavanderia_1.

Carga de datos hacia la tabla lavanderia_1	
Descripción de la carga de datos:	
<p>Haciendo uso de la herramienta Talend Open Studio se importaron los datos del archivo Excel correspondiente a dicha lavandería hacia la tabla lavanderia_1. Después de esto se realizaron consultas para verificar que no se encontraran personas repetidas, que no tuvieran medios y que no existieran filas vacías.</p>	
Mapeo de datos	
Campos del archivo Excel	Campos de la tabla intermedia
1er nombre	primer_nombre
2do nombre	segundo_nombre
1er apellido	primer_apellido
2do apellido	segundo_apellido

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

S	sabana
T	toalla
F	funda
C	colcha
A	almohada
Consultas realizadas	
<p>SELECT primer_nombre, segundo_nombre, primer_apellido, segundo_apellido, COUNT (*) AS cantidad FROM lavanderia_1 GROUP BY primer_nombre, segundo_nombre, primer_apellido, segundo_apellido HAVING COUNT (*) > 1;</p> <p>SELECT primer_nombre, segundo_nombre, primer_apellido, segundo_apellido FROM lavanderia_1 WHERE sabana IS NULL, toalla IS NULL, funda IS NULL, colcha IS NULL, almohada IS NULL;</p> <p>SELECT primer_nombre, segundo_nombre, primer_apellido, segundo_apellido FROM lavanderia_1 WHERE primer_nombre IS NULL AND segundo_nombre IS NULL AND primer_apellido IS NULL AND segundo_apellido IS NULL AND sabana IS NULL AND toalla IS NULL AND funda IS NULL AND colcha IS NULL AND almohada IS NULL;</p>	
Cantidad de objetos cargados:	1930

Tabla 5. Carga de datos hacia la tabla lavanderia_2.

Carga de datos hacia la tabla lavanderia_2	
<p>Descripción de la carga de datos:</p> <p>Haciendo uso de la herramienta Talend Open Studio se importaron los datos del archivo Excel correspondiente a dicha lavandería hacia la tabla lavanderia_2. Después de esto se realizaron consultas para verificar que no se encontraran personas repetidas, que no tuvieran medios y que no existieran filas vacías.</p>	
Mapeo de datos	
Campos del archivo Excel	Campos de la tabla intermedia
1er nombre	primer_nombre

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

2do nombre	segundo_nombre
1er apellido	primer_apellido
2do apellido	segundo_apellido
S	sabana
T	toalla
F	funda
C	colcha
A	almohada
Consultas realizadas	
<p>SELECT primer_nombre, segundo_nombre, primer_apellido, segundo_apellido, COUNT (*) AS cantidad FROM lavanderia_2 GROUP BY primer_nombre, segundo_nombre, primer_apellido, segundo_apellido HAVING COUNT (*) > 1;</p> <p>SELECT primer_nombre, segundo_nombre, primer_apellido, segundo_apellido FROM lavanderia_2 WHERE sabana IS NULL, toalla IS NULL, funda IS NULL, colcha IS NULL, almohada IS NULL;</p> <p>SELECT primer_nombre, segundo_nombre, primer_apellido, segundo_apellido FROM lavanderia_2 WHERE primer_nombre IS NULL AND segundo_nombre IS NULL AND primer_apellido IS NULL AND segundo_apellido IS NULL AND sabana IS NULL AND toalla IS NULL AND funda IS NULL AND colcha IS NULL AND almohada IS NULL;</p>	
Cantidad de objetos cargados:	2787

Tabla 6. Carga de datos hacia la tabla lavanderia_3.

Carga de datos hacia la tabla lavanderia_3
<p>Descripción de la carga de datos:</p> <p>Haciendo uso de la herramienta Talend Open Studio se importaron los datos del archivo Excel correspondiente a dicha lavandería hacia la tabla lavanderia_3. Después de esto se realizaron consultas para verificar que no se encontraran personas repetidas, que no tuvieran medios y que no existieran filas vacías.</p>

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

Mapeo de datos	
Campos del archivo Excel	Campos de la tabla intermedia
1er nombre	primer_nombre
2do nombre	segundo_nombre
1er apellido	primer_apellido
2do apellido	segundo_apellido
S	sabana
T	toalla
F	funda
C	colcha
A	almohada
SC	sabana_camera
CC	cubre_cama
M	mosquitero
Consultas realizadas	
<p>SELECT primer_nombre, segundo_nombre, primer_apellido, segundo_apellido, COUNT (*) AS cantidad FROM lavanderia_3 GROUP BY primer_nombre, segundo_nombre, primer_apellido, segundo_apellido HAVING COUNT (*) > 1;</p> <p>SELECT primer_nombre, segundo_nombre, primer_apellido, segundo_apellido FROM lavanderia_3 WHERE sabana IS NULL, toalla IS NULL, funda IS NULL, colcha IS NULL, almohada IS NULL, sabana_camera IS NULL, cubre_cama IS NULL, mosquitero IS NULL;</p> <p>SELECT primer_nombre, segundo_nombre, primer_apellido, segundo_apellido FROM lavanderia_3 WHERE primer_nombre IS NULL AND segundo_nombre IS NULL AND primer_apellido IS NULL AND segundo_apellido IS NULL AND sabana IS NULL AND toalla IS NULL AND funda IS NULL AND colcha IS NULL AND almohada IS NULL, sabana_camera IS NULL, cubre_cama IS NULL, mosquitero IS NULL;</p>	
Cantidad de objetos cargados:	2100

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

Tabla 7. Carga de datos hacia la tabla lavanderia_4.

Carga de datos hacia la tabla lavanderia_4	
<p>Descripción de la carga de datos:</p> <p>Haciendo uso de la herramienta Talend Open Studio se importaron los datos del archivo Excel correspondiente a dicha lavandería hacia la tabla lavanderia_4. Después de esto se realizaron consultas para verificar que no se encontraran personas repetidas, que no tuvieran medios y que no existieran filas vacías.</p>	
Mapeo de datos	
Campos del archivo Excel	Campos de la tabla intermedia
1er nombre	primer_nombre
2do nombre	segundo_nombre
1er apellido	primer_apellido
2do apellido	segundo_apellido
S	sabana
T	toalla
F	funda
C	colcha
A	almohada
Consultas realizadas	
<pre>SELECT primer_nombre, segundo_nombre, primer_apellido, segundo_apellido, COUNT (*) AS cantidad FROM lavanderia_4 GROUP BY primer_nombre, segundo_nombre, primer_apellido, segundo_apellido HAVING COUNT (*) > 1; SELECT primer_nombre, segundo_nombre, primer_apellido, segundo_apellido FROM lavanderia_4 WHERE sabana IS NULL, toalla IS NULL, funda IS NULL, colcha IS NULL, almohada IS NULL; SELECT primer_nombre, segundo_nombre, primer_apellido, segundo_apellido FROM lavanderia_4 WHERE primer_nombre IS NULL AND segundo_nombre IS NULL AND primer_apellido IS NULL AND segundo_apellido IS NULL AND sabana IS NULL AND toalla IS NULL AND funda IS NULL AND colcha IS NULL AND</pre>	

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

almohada IS NULL;	
Cantidad de objetos cargados:	1745

Tabla 8 Transformación de la tabla lavanderia_1.

Transformación de la tabla lavanderia_1		
<p>Descripción de la transformación de datos:</p> <p>En esta tabla se adicionaron varios campos que no se encontraban en el archivo Excel y eran necesarios para realizar la correcta migración de datos. Luego mediante varias consultas y funciones se le insertaron valores revisando que no existieran identificadores de personas duplicados o nulos. En el caso de los nulos se verificaron que sus nombres y apellidos correspondieran con los de la tabla tb_dpersona_data_plus, en caso de no coincidir se registraron estas personas.</p>		
Atributo adicionado	Tipo de Dato	Descripción
id_persona	varchar(32)	Identificador de la persona
id_s	int4	Identificador del medio sábana
id_t	int4	Identificador del medio toalla
id_f	int4	Identificador del medio funda
id_c	int4	Identificador del medio colcha
id_a	int4	Identificador del medio almohada
id_lavanderia	int4	Identificador de la lavandería
Consultas realizadas		
<pre>UPDATE lavanderia_1 SET id_persona=(SELECT id_persona FROM sq_master_integracion.fn_buscar_persona_avanzada (primer_nombre ' ' segundo_nombre ' ' primer_apellido ' ' segundo_apellido,"",-1,-1) LIMIT 1);</pre> <pre>UPDATE lavanderia_1 SET id_s=21, id_t=22,id_f=26,id_c=20,id_a=27,</pre>		

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

```

id_lavanderia=28;

SELECT id_persona, COUNT (id_persona) AS cantidad FROM lavanderia_1
GROUP BY id_persona HAVING COUNT (id_persona) >1;

SELECT primer_nombre, segundo_nombre, primer_apellido, segundo_apellido
FROM lavanderia_1 WHERE id_persona IS NULL;

UPDATE lavanderia_1 SET id_persona=(SELECT
sq_master_integracion.fn_registrar_persona(primer_nombre,segundo_nombre,pri
mer_apellido,segundo_apellido,primer_nombre,',',e014d35b111b771c01848903d6
ef1aad')) WHERE id_persona IS NULL;
    
```

Tabla 9. Transformación de la tabla lavanderia_2.

Transformación de la tabla lavanderia_2		
Descripción de la transformación de datos: En esta tabla se adicionaron varios campos que no se encontraban en el archivo Excel y eran necesarios para realizar la correcta migración de datos. Luego mediante varias consultas y funciones se le insertaron valores revisando que no existieran identificadores de personas duplicados o nulos. En el caso de los nulos se verificaron que sus nombres y apellidos correspondieran con los de la tabla tb_dpersona_data_plus, en caso de no coincidir se registraron estas personas.		
Atributo adicionado	Tipo de Dato	Descripción
id_persona	varchar(32)	Identificador de la persona
id_s	int4	Identificador del medio sábana
id_t	int4	Identificador del medio toalla
id_f	int4	Identificador del medio funda
id_c	int4	Identificador del medio colcha
id_a	int4	Identificador del medio almohada
id_lavanderia	int4	Identificador de la lavandería

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

Consultas realizadas
<pre>UPDATE lavanderia_2 SET id_persona=(SELECT id_persona FROM sq_master_integracion.fn_buscar_persona_avanzada (primer_nombre ' ' segundo_nombre ' ' primer_apellido ' ' segundo_apellido,"",-1,-1) LIMIT 1);</pre>
<pre>UPDATE lavanderia_2 SET id_s=21, id_t=22,id_f=26,id_c=20,id_a=27, id_lavanderia=30;</pre>
<pre>SELECT id_persona, COUNT (id_persona) AS cantidad FROM lavanderia_2 GROUP BY id_persona HAVING COUNT (id_persona) >1;</pre>
<pre>SELECT primer_nombre, segundo_nombre, primer_apellido, segundo_apellido FROM lavanderia_2 WHERE id_persona IS NULL;</pre>
<pre>UPDATE lavanderia_2 SET id_persona=(SELECT sq_master_integracion.fn_registrar_persona(primer_nombre,segundo_nombre,pri mer_apellido,segundo_apellido,primer_nombre,',',e014d35b111b771c01848903d6 ef1aad')) WHERE id_persona IS NULL;</pre>

Tabla 10. Transformación de la tabla lavanderia_3.

Transformación de la tabla lavanderia_3		
<p>Descripción de la transformación de datos:</p> <p>En esta tabla se adicionaron varios campos que no se encontraban en el archivo Excel y eran necesarios para realizar la correcta migración de datos. Luego mediante varias consultas y funciones se le insertaron valores revisando que no existieran identificadores de personas duplicados o nulos. En el caso de los nulos se verificaron que sus nombres y apellidos correspondieran con los de la tabla tb_dpersona_data_plus, en caso de no coincidir se registraron estas personas.</p>		
Atributo adicionado	Tipo de Dato	Descripción
id_persona	varchar(32)	Identificador de la persona
id_s	int4	Identificador del medio sábana

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

id_t	int4	Identificador del medio toalla
id_f	int4	Identificador del medio funda
id_c	int4	Identificador del medio colcha
id_a	int4	Identificador del medio almohada
id_sc	int4	Identificador del medio sábana camera
id_cc	int4	Identificador del medio cubre cama
id_m	int4	Identificador del medio mosquitero
id_lavanderia	int4	Identificador de la lavandería
Consultas realizadas		
<pre> UPDATE lavanderia_3 SET id_persona=(SELECT id_persona FROM sq_master_integracion.fn_buscar_persona_avanzada (primer_nombre ' ' segundo_nombre ' ' primer_apellido ' ' segundo_apellido,',',',',-1,-1) LIMIT 1); UPDATE lavanderia_3 SET id_s=21, id_t=22, id_f=26, id_c=20, id_a=27, id_sc=25, id_cc=28 , id_m=29, id_lavanderia=27; SELECT id_persona, COUNT (id_persona) AS cantidad FROM lavanderia_3 GROUP BY id_persona HAVING COUNT (id_persona) >1; SELECT primer_nombre, segundo_nombre, primer_apellido, segundo_apellido FROM lavanderia_3 WHERE id_persona IS NULL; UPDATE lavanderia_3 SET id_persona=(SELECT sq_master_integracion.fn_registrar_persona(primer_nombre,segundo_nombre,pri mer_apellido,segundo_apellido,primer_nombre,',',e014d35b111b771c01848903d6 ef1aad')) WHERE id_persona IS NULL; </pre>		

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

Tabla 11. Transformación de la tabla lavanderia_4.

Transformación de la tabla lavanderia_4		
<p>Descripción de la transformación de datos:</p> <p>En esta tabla se adicionaron varios campos que no se encontraban en el archivo Excel y eran necesarios para realizar la correcta migración de datos. Luego mediante varias consultas y funciones se le insertaron valores revisando que no existieran identificadores de personas duplicados o nulos. En el caso de los nulos se verificaron que sus nombres y apellidos correspondieran con los de la tabla tb_dpersona_data_plus, en caso de no coincidir se registraron estas personas.</p>		
Atributo adicionado	Tipo de Dato	Descripción
id_persona	varchar(32)	Identificador de la persona
id_s	int4	Identificador del medio sábana
id_t	int4	Identificador del medio toalla
id_f	int4	Identificador del medio funda
id_c	int4	Identificador del medio colcha
id_a	int4	Identificador del medio almohada
id_lavanderia	int4	Identificador de la lavandería
Consultas realizadas		
<pre>UPDATE lavanderia_4 SET id_persona=(SELECT id_persona FROM sq_master_integracion.fn_buscar_persona_avanzada (primer_nombre ' ' segundo_nombre ' ' primer_apellido ' ' segundo_apellido,"",-1,-1) LIMIT 1); UPDATE lavanderia_4 SET id_s=21, id_t=22,id_f=26,id_c=20,id_a=27, id_lavanderia=47; SELECT id_persona, COUNT (id_persona) AS cantidad FROM lavanderia_4 GROUP BY id_persona HAVING COUNT (id_persona) >1; SELECT primer_nombre, segundo_nombre, primer_apellido, segundo_apellido FROM lavanderia_4 WHERE id_persona IS NULL;</pre>		

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

```
UPDATE lavanderia_4 SET id_persona=(SELECT
sq_master_integracion.fn_registrar_persona(primer_nombre,segundo_nombre,primer_apellido,segundo_apellido,primer_nombre,',','e014d35b111b771c01848903d6ef1aad')) WHERE id_persona IS NULL;
```

Tabla 12. Migración de datos hacia la tabla tb_dboleta_entrega.

Migración de datos hacia la tabla tb_dboleta_entrega	
<p>Descripción de la migración de datos:</p> <p>Para esta tabla se migraron los identificadores de personas e identificadores de lavanderías a través de scripts PHP.</p>	
Mapeo de datos	
Campos de las tablas intermedias	Campos de la tabla tb_dboleta_entrega
id_persona	id_persona
id_lavanderia	id_lavanderia
Cantidad de objetos a migrar:	8562

Tabla 13. Migración de datos hacia la tabla tb_rmedios_boleta.

Migración de datos hacia la tabla tb_rmedios_boleta	
<p>Descripción de la migración de datos:</p> <p>Para esta tabla se migraron los medios, sus identificadores y los identificadores de las boletas de entrega a través de scripts PHP.</p>	
Mapeo de datos	
Campos de las tablas intermedias	Campos de la tabla tb_rmedios_boleta
id_s	id_medio

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

id_t	id_medio
id_f	id_medio
id_c	id_medio
id_a	id_medio
id_sc	id_medio
id_cc	id_medio
id_m	id_medio
sabana	cantidad
toalla	cantidad
funda	cantidad
colcha	cantidad
almohada	cantidad
sabana_camera	cantidad
cubre_cama	cantidad
mosquitero	cantidad
Campos de las tabla tb_dboleta_entrega	Campos de la tabla tb_rmedios_boleta
id_boleta_entrega	id_boleta_entrega
Cantidad de objetos a migrar:	39153

2.5 Descripción de las funciones utilizadas en los scripts PHP

Teniendo los datos necesarios para la migración en las tablas intermedias, se crearon los scripts para ejecutar la migración de dichos datos hacia la fuente de destino. Lo primero que se hizo fue la función registrar_boleta_entrega, la cual recibe como parámetro el identificador de la persona y el identificador de la lavandería que son obtenidos de las tablas intermedias, para luego insertarlos en

CAPÍTULO 2. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

la tabla `tb_dboleta_entrega`. Este método retorna el identificador de la boleta de entrega, que se le pasa como parámetro a la función `registrar_medios_boleta` junto al identificador del medio y la cantidad de los medios adquiridos de las tablas de almacenamiento intermedio, para más tarde ser insertados en la tabla `tb_rmedios_boleta` con esta función. Este procedimiento se lleva a cabo en todas las tablas de almacenamiento intermedio, para realizar la migración completa de los datos. Una vez probado el proceso de migración de los datos en la base de datos de desarrollo, estos scripts PHP pueden ser reutilizados en la migración hacia el servidor de producción.

2.6 Conclusiones

En la elaboración del presente capítulo se explicaron las etapas que fueron atravesadas en el proceso de migración, a través de la estrategia empleada, el análisis antes de dicho proceso, el desarrollo de la carga de los datos usando la herramienta Talend Open Studio, la migración de la información y la descripción de las funciones utilizadas, estas fases quedaron enunciadas en el primer capítulo de la investigación.

CAPÍTULO 3. VALIDACIÓN DE LA SOLUCIÓN PROPUESTA

Capítulo 3. Validación de la solución propuesta

3.1 Introducción

En este capítulo se manifiestan las pruebas que se ejecutaron después de terminar la migración de los datos para el Sistema de Gestión de Residencia, las cuales son imprescindibles para migrar los datos correctamente, también en dicho capítulo se muestran los resultados obtenidos, es decir la cantidad de tablas y de campos que se migraron.

3.2 Descripción del período de pruebas

El período de pruebas se lleva a cabo con el objetivo de controlar que el número de datos obtenidos de la fuente de origen se corresponda con el número de los insertados en la fuente de destino; estas se ejecutaron cuando terminó la migración de los datos y poseen una vital importancia ya que son las que demuestran que la migración se llevó a cabo satisfactoriamente. Si mientras se realizan las pruebas aparecen errores, hay que regresar a la fase anterior para solucionar esos problemas y ejecutar nuevamente la etapa de pruebas; dicho período es fundamental para que se haga una migración adecuada, porque si se deja de hacer la comprobación de los datos después que llegue a su fin la migración, se corre el riesgo de que la misma no se haya realizado de una forma apropiada.

Esta etapa se realizó utilizando la herramienta *Talend Open Profiler*, la cual posibilita la generación de gráficos de las estadísticas, a partir de los campos almacenados en una base de datos, se crearon varios gráficos por cada uno de los campos que fueron migrados, para poder comprobar que todos los de la fuente

CAPÍTULO 3. VALIDACIÓN DE LA SOLUCIÓN PROPUESTA

de origen se corresponden con todos los insertados en la fuente de destino; entre las estadísticas que proporciona dicha herramienta se encuentran: el número total de campos, la cantidad de campos duplicados, los campos únicos, el número de campos en blanco y el total de campos insertados con valores por defecto. A continuación se muestran los gráficos realizados:

En la figura 6 se muestran los identificadores de los administradores de las lavanderías que fueron insertados en la tabla `tb_dadministradora_lavanderia` de la base de datos `BD_sgu_residencia`, los cuales son cuatro porque hay esa misma cantidad de lavanderías en la Universidad, por tanto se demuestra que se insertaron todos, además se evidencia que los cuatro identificadores son diferentes y únicos ya que pertenecen a personas distintas y que no hay elementos duplicados, en blanco o nulos.

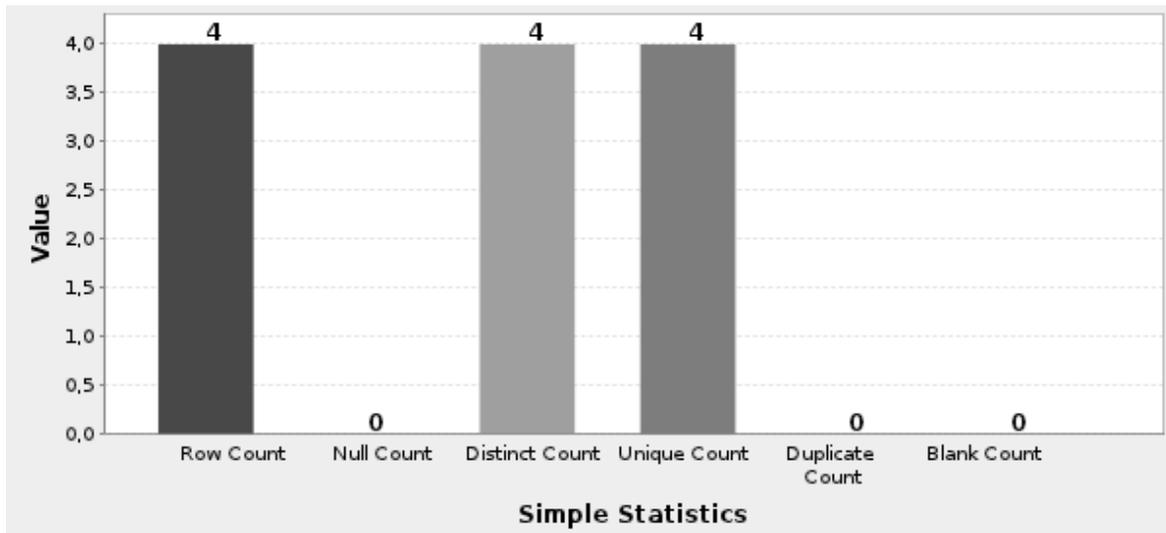


Figura 6. Identificadores de los administradores de las lavanderías de la tabla `tb_dadministradora_lavanderia`.

En la figura 7 se exponen los identificadores de las lavanderías que fueron insertados en la tabla `tb_dadministradora_lavanderia` de la base de datos `BD_sgu_residencia`, los cuales son cuatro porque existe esa misma cantidad de

CAPÍTULO 3. VALIDACIÓN DE LA SOLUCIÓN PROPUESTA

lavanderías en la Universidad y por tanto se demuestra que se insertaron todos, además se evidencia que los cuatro identificadores son diferentes y únicos ya que pertenecen a lavanderías distintas y que no hay elementos duplicados ni nulos.

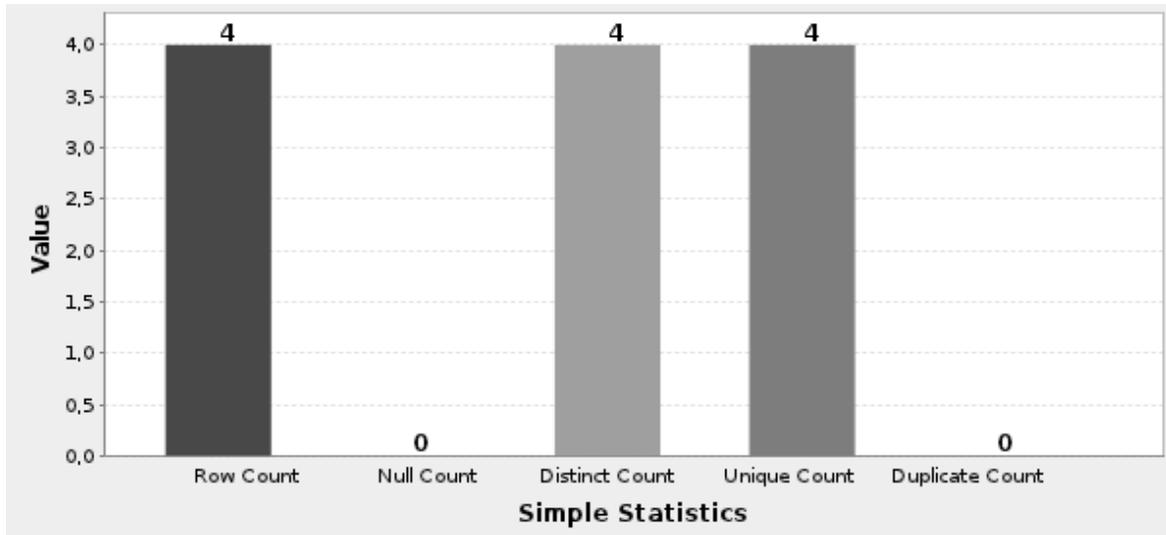


Figura 7. Identificadores de las lavanderías de la tabla tb_dadministradora_lavanderia. En las figuras 8 y 9 se muestran los indicadores de comparación de los identificadores de personas que fueron migrados hacia la tabla tb_dboleta_entrega de la base de datos BD_sgu_residencia desde la tabla lavanderias, donde se puede apreciar que los valores de los indicadores coinciden en su totalidad.

CAPÍTULO 3. VALIDACIÓN DE LA SOLUCIÓN PROPUESTA

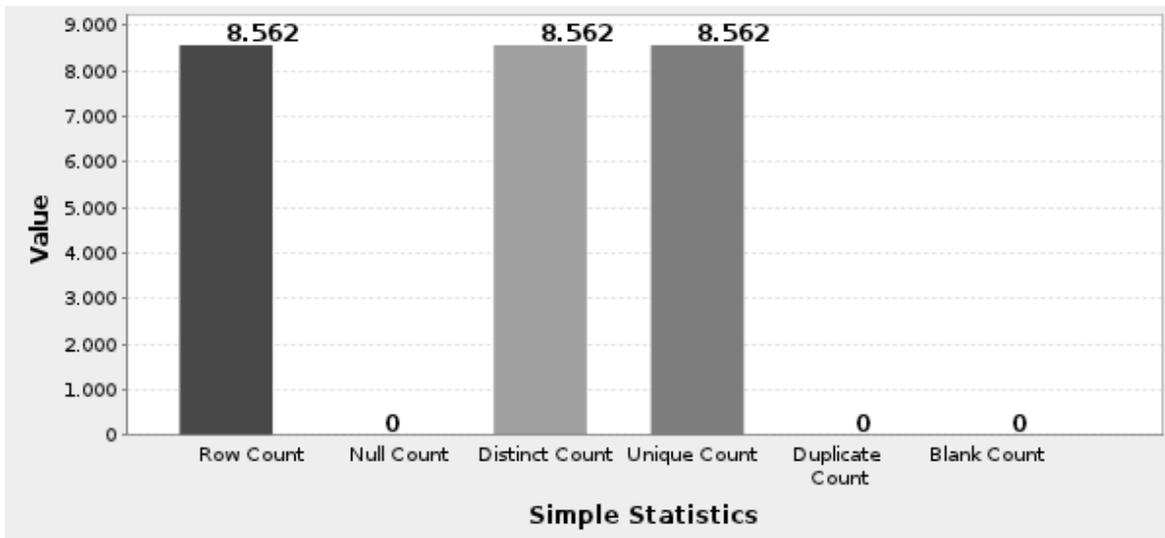


Figura 8. Identificadores de las personas de la tabla lavanderias.

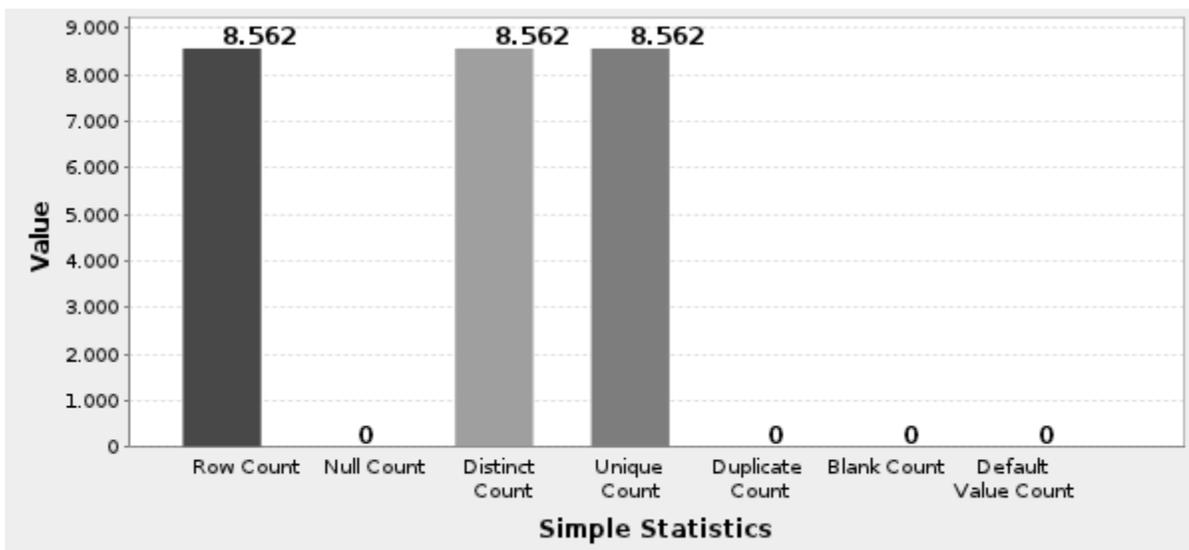


Figura 9. Identificadores de las personas de la tabla tb_dboleta_entrega.

En las figuras 10 y 11 se muestran los indicadores de comparación de los identificadores de las lavanderías que fueron migrados hacia la tabla tb_dboleta_entrega de la base de datos BD_sgu_residencia desde la tabla

CAPÍTULO 3. VALIDACIÓN DE LA SOLUCIÓN PROPUESTA

lavanderías, donde se puede apreciar que los valores de los indicadores coinciden en su totalidad.

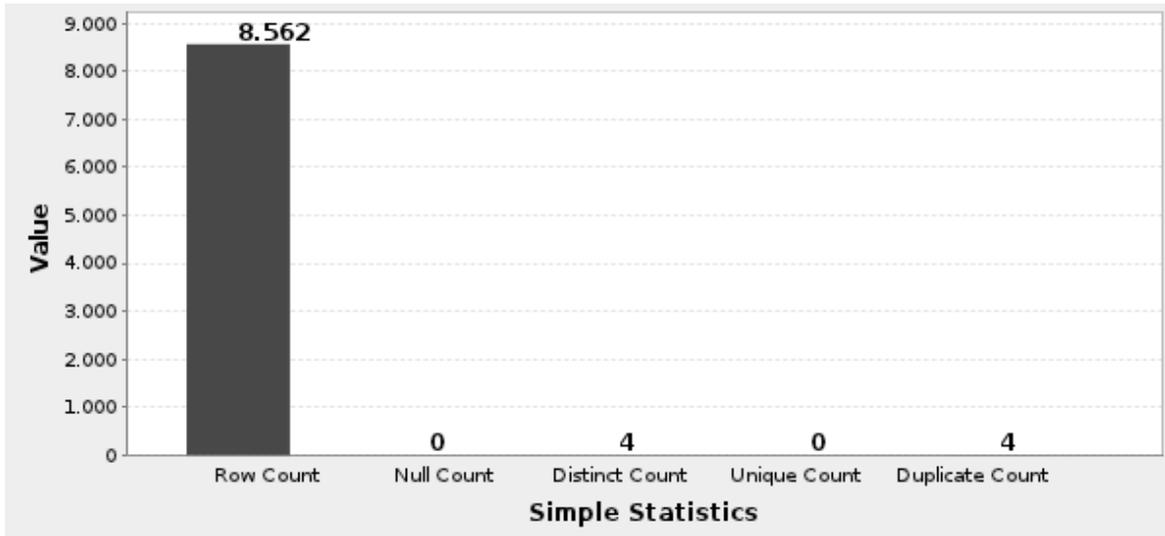


Figura 10. Identificadores de las lavanderías de la tabla lavanderías.

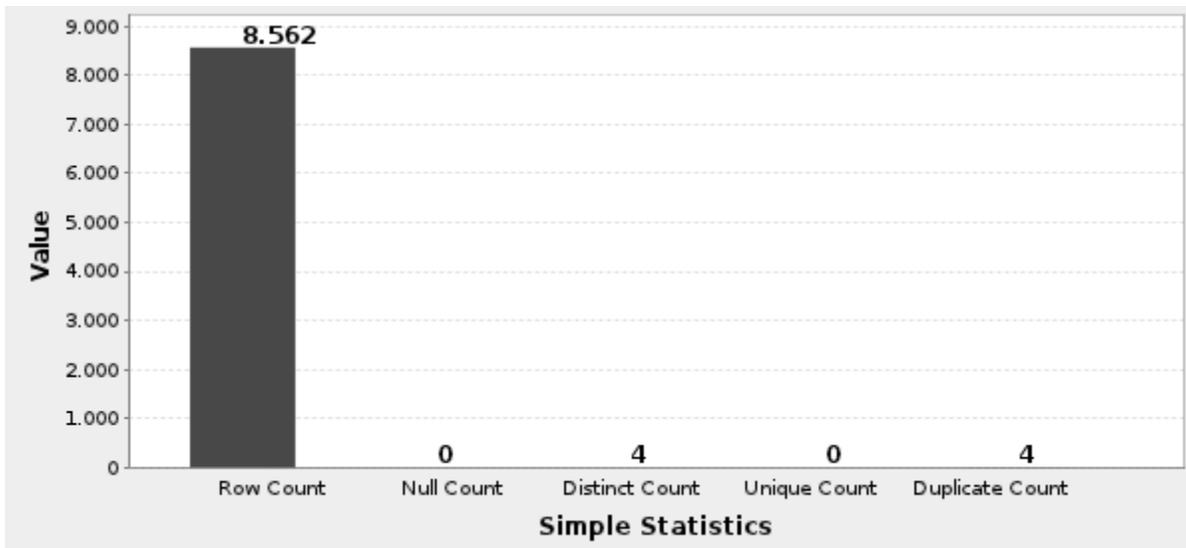


Figura 11. Identificadores de las lavanderías de la tabla tb_dboleta_entrega.

En la figura 12 se muestran los identificadores de los administradores de las lavanderías que fueron insertados en la tabla tb_dlavanderia de la base datos

CAPÍTULO 3. VALIDACIÓN DE LA SOLUCIÓN PROPUESTA

BD_sgu_residencia, los cuales son cuatro ya que la Universidad cuenta con esa misma cantidad de lavanderías, por tanto se demuestra que se insertaron todos, además se evidencia que los cuatro identificadores son diferentes y únicos ya que pertenecen a personas distintas y que no hay elementos duplicados, en blanco o nulos.

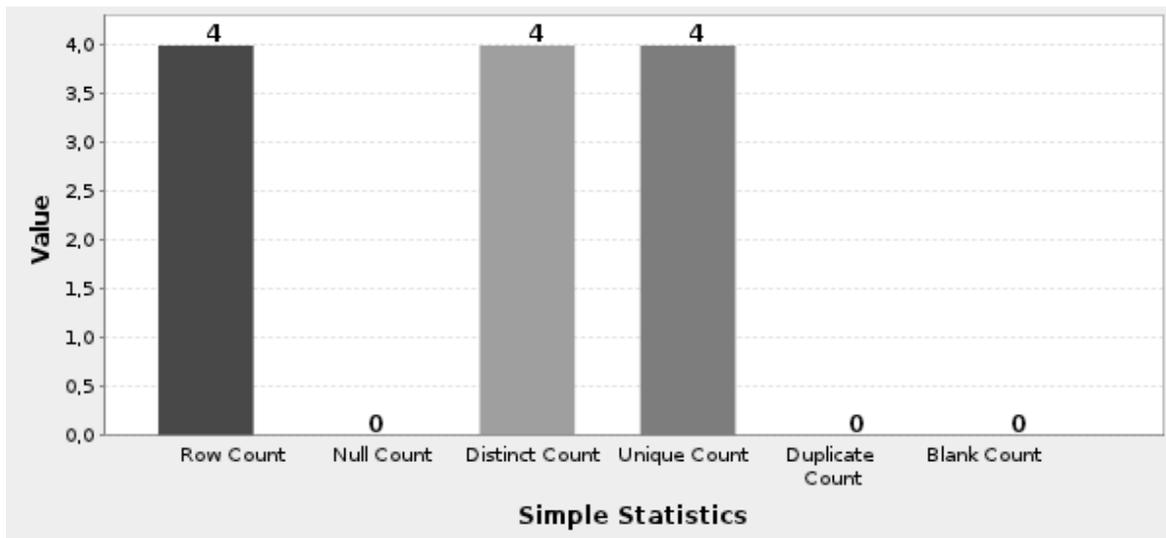


Figura 12. Identificadores de los administradores de las lavanderías de la tabla tb_dlavanderia.

En la figura 13 se muestran los nombres de las lavanderías que fueron insertados en la tabla tb_dlavanderia de la base de datos BD_sgu_residencia, los cuales son cuatro ya que la Universidad cuenta con esa misma cantidad de lavanderías, por tanto se demuestra que se insertaron todos, además se evidencia que los cuatro son diferentes y únicos ya que pertenecen a lavanderías distintas y que no hay elementos duplicados, en blanco o nulos.

CAPÍTULO 3. VALIDACIÓN DE LA SOLUCIÓN PROPUESTA

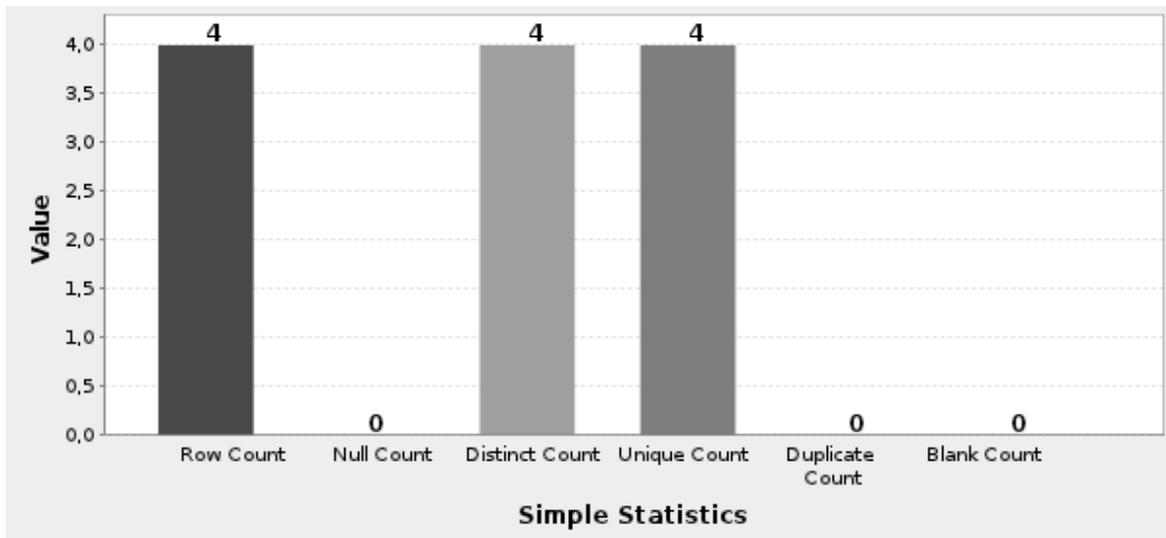


Figura 13. Nombres de las lavanderías de la tabla tb_dlavanderia.

En la figura 14 se muestran los nombres de las residencias que fueron insertados en la tabla tb_dlavanderia de la base datos BD_sgu_residencia, los cuales son cuatro ya que la Universidad cuenta con esa misma cantidad de residencias, por tanto se demuestra que se insertaron todos, además se evidencia que los cuatro son diferentes y únicos ya que pertenecen a residencias distintas y que no hay elementos duplicados, en blanco o nulos.

CAPÍTULO 3. VALIDACIÓN DE LA SOLUCIÓN PROPUESTA

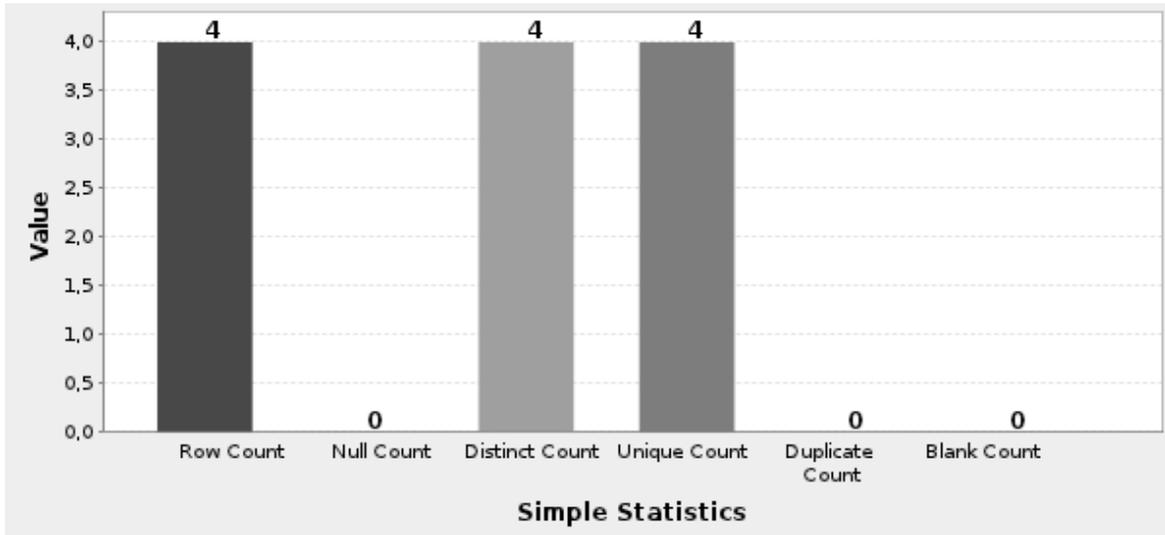


Figura 14. Nombres de las residencias de la tabla tb_dlavanderia.

En la figura 15 se muestran los identificadores de las boletas de entrega por cada medio, que fueron migrados para la tabla tb_rmedios_boleta de la base de datos BD_sgu_residencia, desde la tabla tb_dboleta_entrega.

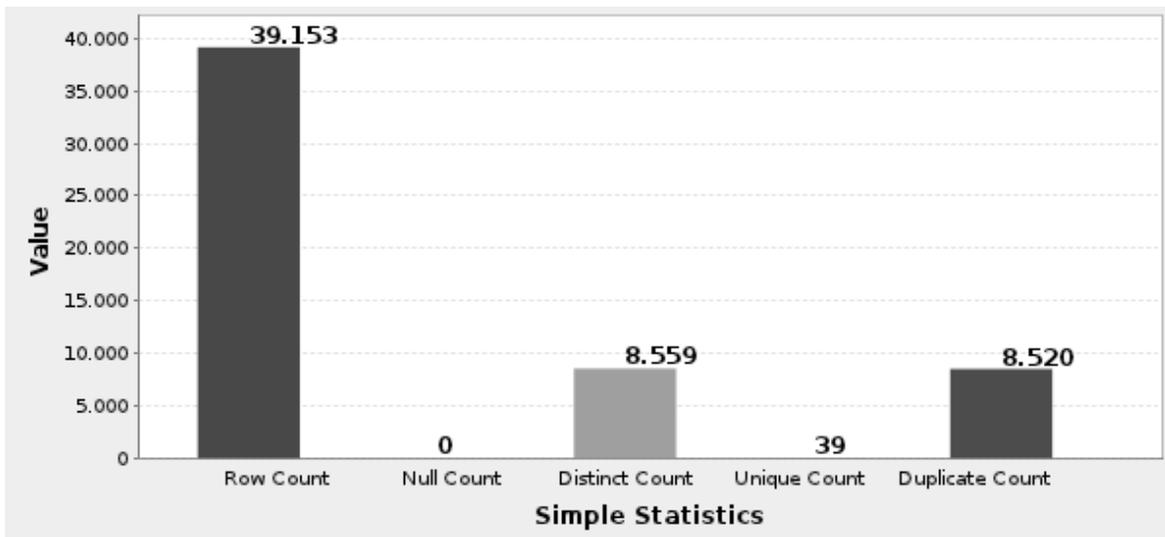


Figura 15. Identificadores de las boletas de entrega de la tabla tb_rmedios_boleta.

CAPÍTULO 3. VALIDACIÓN DE LA SOLUCIÓN PROPUESTA

En la figura 16 se muestran los identificadores de los medios que fueron migrados para la tabla `tb_rmedios_boleta` de la base de datos `BD_sgu_residencia`, desde la tabla `lavanderias`, donde se comprueba que no se insertó ningún elemento nulo, ni único, que existen ocho tipos diferentes de medios (sábana, toalla, funda, almohada, colcha, cubre_cama, sabana_camera, mosquitero) ya que se migró esa misma cantidad de elementos distintos y que también se encuentran duplicados ese mismo número de veces.

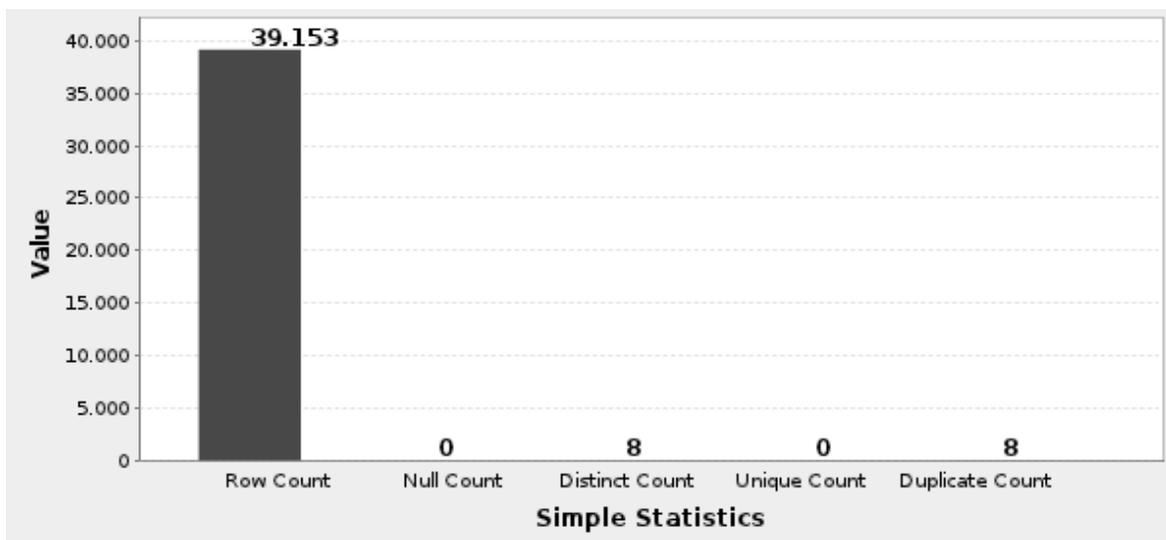


Figura 16. Identificadores de los medios de la tabla `tb_rmedios_boleta`.

En la figura 17 se muestran las descripciones de los medios que fueron insertados en la tabla `tb_ntipo_medio` de la base de datos `BD_sgu_residencia`, donde se demuestra que se insertaron ocho ya que ese es el número de medios que se entregan a las personas, que no hay valores nulos, únicos o en blanco, que se insertó un solo tipo de elemento y que hay esa misma cantidad de elementos duplicados.

CAPÍTULO 3. VALIDACIÓN DE LA SOLUCIÓN PROPUESTA

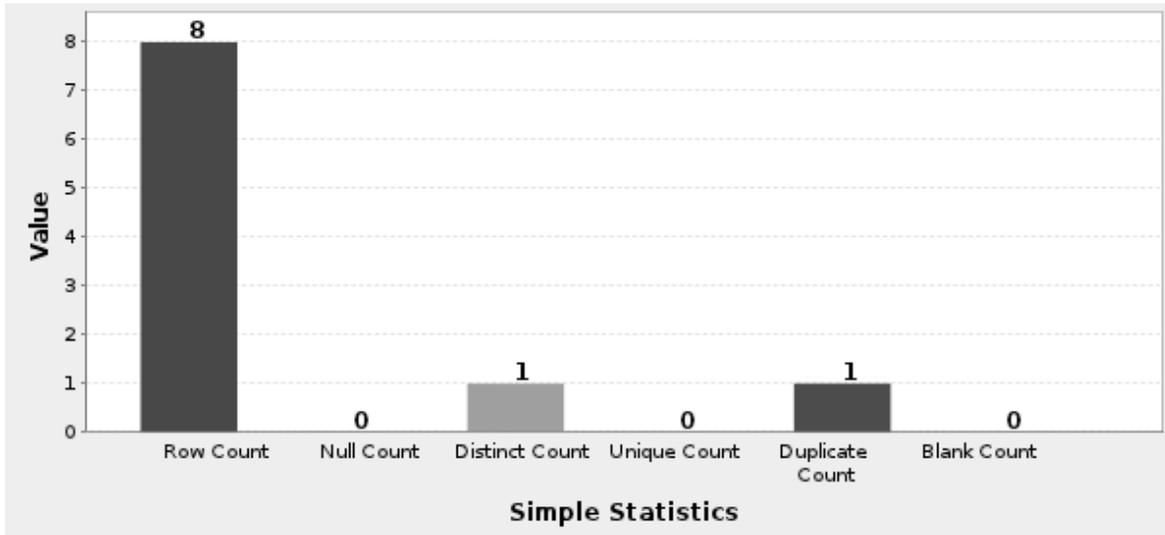


Figura 17. Descripciones de los medios de la tabla tb_ntipo_medio.

En la figura 18 se muestran los nombres de los tipos de los medios que fueron insertados en la tabla tb_ntipo_medio de la base de datos BD_sgu_residencia, donde se comprueba que son ocho ya que esa es la cantidad de medios que se entrega a las personas, que se insertaron esa misma cantidad de elementos distintos y únicos porque los nombres de los tipos de medios son diferentes y que no se insertó ningún elemento nulo, duplicado o en blanco.

CAPÍTULO 3. VALIDACIÓN DE LA SOLUCIÓN PROPUESTA

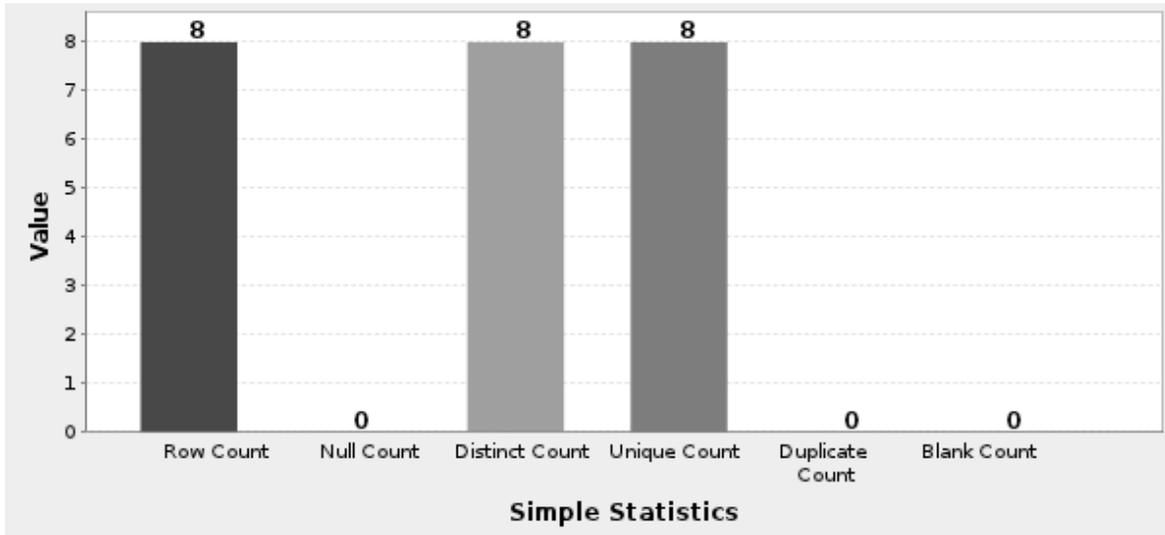


Figura 18. Nombres de los tipos de los medios de la tabla tb_ntipo_medio.

CAPÍTULO 3. VALIDACIÓN DE LA SOLUCIÓN PROPUESTA

3.3 Resumen de la cantidad de campos insertados en cada una de las tablas

Tabla 14. Cantidad de campos insertados en cada tabla.

Base de datos	Tabla	Campo	Total
BD_sgu_residencia	lavanderias	id_persona	8562
BD_sgu_residencia	tb_dboleta_entrega	id_persona	8562
BD_sgu_residencia	lavanderias	id_lavanderia	8562
BD_sgu_residencia	tb_dboleta_entrega	id_lavanderia	8562
BD_sgu_residencia	tb_dpersona_data_plus	id_persona	4
BD_sgu_residencia	tb_dadministradora_lavanderia	id_persona	4
BD_sgu_residencia	tb_dlavanderia	id_lavanderia	4
BD_sgu_residencia	tb_dadministradora_lavanderia	id_lavanderia	4
BD_sgu_residencia	tb_dpersona_data_plus	id_persona	4
BD_sgu_residencia	tb_dlavanderia	id_persona	4

3.4 Conclusiones

Durante la confección de este capítulo se verificó que los datos fueron migrados para el Sistema de Gestión de Residencia, asegurando que estén en la fuente de destino, usando la herramienta Talend Open Profiler, la cual permite confirmar la relación entre los datos que se encuentran en la fuente de origen y los que están en la fuente de destino, siendo útil de esta forma a la hora de ejecutar una fase fundamental para que los datos se migren correctamente, como es la comprobación de los datos.

CONCLUSIONES GENERALES

Conclusiones generales

Después de terminar este trabajo se le da cumplimiento a los objetivos que inicialmente se habían propuesto. Por tanto se concluye que:

- Se obtuvo la base teórica que apoya a la solución propuesta, a través de la investigación realizada.
- Se diseñó y desarrolló una estrategia para la migración de los datos que se guardaban en las lavanderías de la Universidad, hacia el Sistema de Gestión de Residencia.
- Se utilizó la herramienta Talend Open Studio para apoyar el proceso de migración de datos, la cual evidenció ser confiable para los procesos de migración de datos.
- Se utilizó la herramienta Talend Open Profiler para ayudar en la verificación del éxito de la migración y el análisis de la calidad de los datos después de ser migrados.

RECOMENDACIONES

Recomendaciones

- Crear puntos de comprobación en la planificación de la migración, para controlar en cada uno de ellos cómo se va desarrollando la misma.
- Realizar la verificación de la integridad de la información por parte del cliente desde el Sistema de Gestión de Residencia.

BIBLIOGRAFÍA

Bibliografía

1. *Estrategia de Migración hacia Sistemas de Gestores de Base de Datos Libres*. **Falcón, Antonio Mauri García y Abel Pérez**. La Habana : s.n., 2008.
2. **Date, Christopher J.** *Introducción a los Sistemas de Bases de Datos*. La Habana : Félix Varela, 2001.
3. *Base de Datos para la Residencia de la UCI*. **Sánchez, Leonel De Jesús Castillo Santana y Yaniel Alvarez**. La Habana : s.n., 2007.
4. **A. SILVERSCHATZ, H.F Korth y Sudarshan S.** *Fundamentos de Bases de Datos*. Madrid : s.n., 2006.
5. pgAdmin PostgreSQL Tools. *pgAdmin PostgreSQL Tools*. [En línea] [Citado el: 15 de diciembre de 2011.] <http://www.pgadmin.org/>.
6. Talend Open Integration Solutions. *Talend Open Integration Solutions*. [En línea] Talend Inc. [Citado el: 6 de enero de 2012.] www.talend.com.
7. **Jaque, Alvaro Alejandro Herrera**. mailxmail. *mailxmail*. [En línea] Grupo Intercom, 24 de 10 de 2007. [Citado el: 10 de enero de 2012.] <http://www.mailxmail.com/cursos-migracion-datos/migracion-datos>.
8. Red Costarricense de Software Libre. *Red Costarricense de Software Libre*. [En línea] [Citado el: 15 de enero de 2012.] http://www.softwarelibre.org/faq/software_libre.
9. **Guerrero, Rafael Martinez**. PostgreSQL-es. *PostgreSQL-es*. [En línea] [Citado el: 20 de enero de 2012.] http://www.postgresql.org.es/sobre_postgresql.
10. Infomática Hoy. *Infomática Hoy*. [En línea] [Citado el: 25 de enero de 2012.] <http://www.informatica-hoy.com.ar/software-linux/Beneficios-y-desventajas-de-Linux.php>.
11. **Mehdi Achour, Friedhelm Betz , Antony Dovgal, Nuno Lopes, Hannes Magnusson, Georg Richter, Damien Seguy, Jakub Vrana**. PHP.net. *PHP.net*. [En línea] [Citado el: 30 de enero de 2012.] <http://docs.php.net/manual/es/index.php>.
12. **Marley, Jimi**. Programación en castellano. *Programación en castellano*. [En línea] [Citado el: 5 de febrero de 2012.] http://www.programacion.com/articulo/por_que_elegir_php_143#porquephpventajas.

BIBLIOGRAFÍA

13. EASYFROM. *EASYFROM*. [En línea] EasyFrom Inc. [Citado el: 9 de febrero de 2012.] <http://www.easyfrom.net/es>.
14. **José, Juan.** islabit. *islabit*. [En línea] 1 de marzo de 2009. [Citado el: 15 de febrero de 2012.] <http://www.islabit.com/3119/talend-open-data-solutions.html>.
15. PostgreSQL Cuba. *PostgreSQL Cuba*. [En línea] [Citado el: 19 de febrero de 2012.] <http://postgresql.uci.cu/node/46>.
16. **García, Rosa María Mato.** *Diseño de bases de datos*. 1999.
17. mastermagazine. *mastermagazine*. [En línea] [Citado el: 24 de febrero de 2012.] <http://www.mastermagazine.info/termino/6051.php>.
18. **Albert, Phil.** linuxinsider. *linuxinsider*. [En línea] [Citado el: 6 de marzo de 2012.] <http://www.linuxinsider.com/story/38172.html>.
19. **Álvarez, Sara.** desarrolweb.com. *desarrolweb.com*. [En línea] 25 de septiembre de 2007. [Citado el: 10 de marzo de 2012.] <http://www.desarrolweb.com/articulos/estructura-modelo-relacional.html>.
20. Universidad del Azuay. *Universidad del Azuay*. [En línea] [Citado el: 15 de marzo de 2012.] <http://www.uazuay.edu.ec/analisis/EI%20modelo%20relacional.pdf>.
21. **Sancho, Antonia Chávez y Fernando.** temas de bases de datos. *temas de bases de datos*. [En línea] [Citado el: 20 de marzo de 2012.] <http://www.cs.us.es/cursos/bd-2001/temas>.
22. **Alvarez, Sara.** desarrolweb.com. *desarrolweb.com*. [En línea] 31 de julio de 2007. [Citado el: 25 de marzo de 2012.] <http://www.desarrolweb.com/articulos/sistemas-gestores-bases-datos.html>.
23. *Módulo de extracción, transformación y réplica de datos del Sistema para el Aseguramiento de la Actividad de Estructura y Composición de las FAR versión 3.* **Bermúdez, Julio Rodríguez.** La Habana : s.n., 2009.
24. IBM. *IBM*. [En línea] [Citado el: 6 de abril de 2012.] <http://www-05.ibm.com/services/es/sds/0905050300.html>.
25. **Casanova, Jaime.** PostgreSQL. *PostgreSQL*. [En línea] 1 de julio de 2009. [Citado el: 10 de abril de 2012.] <http://archives.postgresql.org/pgsql-es-fomento/2009-07/msg00000.php>.

GLOSARIO DE TÉRMINOS

Glosario de términos

1. **PHP:** Preprocesador de Hipertexto (Hypertext Preprocessor). Lenguaje interpretado de alto nivel casi siempre unido a etiquetas HTML. Comúnmente utilizado para el desarrollo de páginas dinámicas.
2. **IBM:** International Business Machines es una empresa multinacional estadounidense de tecnología y consultoría con sede en Armonk, Nueva York.
3. **SQL:** Structured Query Language o Lenguaje Estructurado de Consultas, es un lenguaje declarativo de acceso a bases de datos relacionales que permite especificar diversos tipos de operaciones en éstas.
4. **dBase:** Fue el primer sistema de gestión de base de datos usado ampliamente para microcomputadoras, publicado por Ashton-Tate.
5. **SGBD:** Sistemas de Gestión de Base de Datos.
6. **DDL:** Data Definition Language o lenguaje de definición de datos, incluye órdenes para definir, modificar o borrar las tablas que componen la base de datos, así como los índices, vistas, permisos, etc.
7. **DCL:** Data Control Language o lenguaje de control de datos, contiene elementos útiles para trabajar en un entorno multiusuario, en el que es importante la protección, la seguridad y el establecimiento de restricciones en el acceso.
8. **DML:** Data Manipulation Language o lenguaje de manipulación de datos, este permite recuperar los datos almacenados en la base de datos y también incluye órdenes para permitir al usuario actualizar la base de datos.
9. **Database Migration Toolkit Professional:** Es un programa que permite convertir entre formatos de bases de datos diferentes.
10. **Softek Transparent Data Migration Windows UNIX (IP):** Es un producto de migración de datos diseñado para trasladar datos.
11. **Java:** Lenguaje de programación originalmente desarrollado por la compañía Sun Microsystems.
12. **HTML:** Lenguaje de Marcado de Hipertexto (HyperText Markup Language). Utilizado para la elaboración de páginas Web describiendo la estructura y contenido de dicha página.