Universidad de las Ciencias Informáticas

Facultad 6

Título: Mercado de datos Laboral para el Sistema de Información Estadística Judicial para los Tribunales Populares.

TRABAJO DE DIPLOMA PARA OPTAR POR EL TÍTULO DE INGENIERO EN CIENCIAS INFORMÁTICAS

Autores: Martha Maricela Véliz Jaime
Félix Camilo Calderón Aguilera

Tutora: Ing. Yordanka Hechavarría Melo
Co-Tutora: Ing. Neysis Hernández Díaz

La Habana, junio del 2013
"Año 55 de la Revolución"

"El futuro de nuestra patria tiene que ser necesariamente un futuro de hombres de ciencia, tiene que ser un futuro de hombres de pensamiento, porque precisamente es lo que más estamos sembrando; lo que más estamos sembrando son oportunidades a la inteligencia (...)"

Fidel Castro Ruz

Declaración de autoría

Declaramos ser autores del presente trabajo “Mercado de datos Laboral para el Sistema de Información Estadística Judicial para los Tribunales Populares”, y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

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

Martha Maricela Véliz Jaime

_____

Firma de la Autora

Félix Camilo Calderón Aguilera

_____

Firma del Autor

Ing. Yordanka Hechavarría Melo

_____

Firma de la Tutora

Ing. Neysis Hernández Díaz

_____

Firma de la Co-Tutora

Datos de contacto

Ing. Yordanka Hechavarría Melo
Graduada en Ingeniería en Ciencias Informáticas en el 2009 en la Universidad de las Ciencias Informáticas. Actualmente trabaja en el Centro de Tecnologías de Gestión de Datos, específicamente en el departamento de Almacenes de Datos como especialista en la línea de Integración de Datos.
Correo electrónico: yordankah@uci.cu

Ing. Neysis Hernández Díaz
Graduada en la Universidad de las Ciencias Informáticas del curso 2006-2007. Profesora del departamento de Ciencias Básicas adjunta al Centro de Tecnología de Gestión de Datos.
Correo electrónico: nhernandez@uci.cu

Agradezco:

A mis padres por su gran amor y apoyo incondicional ante cada situación.

A mis hermanas por haber sido siempre un modelo a seguir, por todo el apoyo y amor brindado durante estos años de vida.

A mi novio por el apoyo y los consejos brindados durante los cinco años de carrera.

A Reinier Martínez por los consejos brindados con el objetivo de ver mis sueños hechos realidad.

A los amigos que me acompañaron durante estos cinco años de carrera, especialmente a Liniuska, Yoendy, Liuva y Maidelín.

A todos los profesores que han contribuido en la realización de este trabajo de diploma especialmente a Ramón por ser incondicional y estar dispuesto en cada momento a brindar su ayuda, a mi tutora Yordanka por su ayuda y preocupación constante durante el desarrollo del trabajo de diploma, a Manresa y Marisel por estar presentes cuando de ellos necesité.

Dedico esta tesis:

A mi familia, en especial a mis padres y hermanas por su amor, dedicación infinita y apoyo incondicional en los momentos más difíciles de mi vida.

A mi novio Fabian López García por su ayuda y apoyo ante las adversidades que transcurrieron durante el desarrollo de este trabajo de diploma.

Martha Maricela Véliz Jaime

Agradezco:

A mis padres, por todo el apoyo y amor que me han brindado durante toda mi vida.

A mis tutoras, por exigir siempre lo mejor de mí y tener la paciencia y dedicación necesaria para guiarme.

A Themis, Fabian, Liniuska, Marisel y a todos los profesores que de una forma u otra me dieron su apoyo cuando más lo necesitaba.

A mi dúo de tesis, que me ha guiado en los momentos más difíciles y que siempre estuvo presente cuando la necesitaba.

Dedico esta tesis:

A mi familia, en especial a mis padres, que si no fuera por ellos nunca sería la persona que soy en la actualidad.

Félix Camilo Calderón Aguilera

Resumen

La presente investigación surge como parte de la colaboración que existe entre la Universidad de las Ciencias Informáticas y el Departamento Independiente de Estadística Judicial del Tribunal Supremo Popular. El proceso de análisis de la información llevado a cabo por los especialistas de esta entidad presenta varios problemas, debido a que no poseen una herramienta que permita la integración y centralización de los datos. Por tal motivo se desarrolla el mercado de datos Laboral para el Sistema de Información Estadística Judicial para los Tribunales Populares que contribuirá al proceso de toma de decisiones que se desarrolla en el Tribunal Supremo Popular. En el desarrollo de este mercado de datos se tuvieron en cuenta los requisitos especificados por el cliente y se utilizó la propuesta de Metodología de Desarrollo de Almacenes de Datos empleada en DATEC, para guiar el proceso de desarrollo de la solución, así como, las herramientas: Visual Paradigm for UML 8.0, PostgreSQL 9.1, PgAdmin III 1.14, DataCleaner 1.5.4, Pentaho Data Integration 4.2.1, Mondrian Schema Workbench 3.2.1, Pentaho BI Server 3.10, Mondrian OLAP Server 3.0.4 y Apache Tomcat 6.0.29, con el objetivo de satisfacer las necesidades de los especialistas de dicha entidad.

Palabras claves:

Departamento Independiente de Estadística Judicial, Laboral, Mercado de datos, Tribunal Supremo Popular.

Índice

Índice VII
Introducción 1
Capítulo 1: Fundamentos teóricos de los almacenes de datos 5
Introducción 5
1.1 Almacenes de datos 5
1.1.1 Tendencias actuales del uso de almacenes de datos en el mundo 6
1.1.2 Arquitectura de un almacén de datos 7
1.2 Mercado de datos 7
1.3 Integración de datos 8
1.4 Inteligencia de negocio 9
1.5 Metodologías para el desarrollo de un almacén de datos 12
1.6 Herramientas para la creación de los almacenes de datos 14
1.6.1 Herramienta de modelado de datos 14
1.6.2 Herramientas de gestión y administración de bases de datos 16
1.6.3 Herramientas para el proceso de integración de datos 17
1.6.4 Herramientas para el proceso de inteligencia de negocio 19
1.7 Conclusiones del capítulo 20
Capítulo 2: Análisis y diseño del mercado de datos 21
Introducción 21
2.1 Descripción del negocio 21
2.2 Necesidades de los usuarios 22
2.3 Especificación de requerimientos 22
2.3.1 Requisitos de información 22
2.3.2 Requisitos funcionales 23
2.4.3 Requisitos no funcionales 23
2.4 Reglas del negocio 25
2.5 Casos de uso del sistema 26
2.5.1 Actores del sistema 26
2.5.2 Diagrama de casos de uso del sistema 26
2.5.3 Especificaciones de casos de uso 27

2.6 Arquitectura 30
2.7 Diseño del subsistema de almacenamiento 31
2.7.1 Estándares de codificación 31
2.7.2 Dimensiones 31
2.7.3 Hechos y medidas 32
2.7.4 Matriz bus 33
2.7.5 Modelo de datos 34
2.8 Diseño del subsistema de integración 35
2.9 Diseño del subsistema de visualización 39
2.10 Políticas de respaldo y recuperación 40
2.11 Conclusiones del capítulo 40
Capítulo 3: Implementación y pruebas del mercado de datos 41
Introducción 41
3.1 Implementación del subsistema de almacenamiento de datos 41
3.2 Implementación del subsistema de integración de datos 42
3.2.1 Transformaciones y trabajos 43
3.2.2 Gestión del cambio de las dimensiones 47
3.2.3 Gestión de los metadatos del proceso de integración 51
3.3 Implementación del subsistema de visualización de datos 51
3.3.1 Implementación de los cubos OLAP 51
3.3.2 Implementación de la capa de visualización 52
3.3.3 Implementación de la seguridad de los usuarios 54
3.4 Pruebas 55
3.4.1 Casos de prueba 55
3.4.2 Listas de chequeo 56
3.4.3 Resultados de las pruebas 57
3.5 Conclusiones del capítulo 59
Conclusiones generales 60
Recomendaciones 61
Referencias bibliográficas 62
Bibliografía 65

Anexos 68

Índice de figuras

Figura 1. Ciclo de vida de la propuesta de Metodología de Desarrollo de Almacenes de Datos 13
Figura 2. Diagrama de casos de uso 27
Figura 3. Arquitectura del MD Laboral 30
Figura 4. Modelo dimensional para el hecho seguridad social y el hecho disponibilidad laboral 35
Figura 5. Cantidad máxima de caracteres de la fuente Disponibilidad Laboral 37
Figura 6. Tipos de datos de las fuentes del MD Laboral 37
Figura 7. Diseño de las transformaciones para el hecho seguridad social 38
Figura 8. Diseño del mapa de navegación del MD Laboral 39
Figura 9. Trabajo para cargar el hecho seguridad social 43
Figura 10. Transformación correspondiente al hecho seguridad social 44
Figura 11. Transformación realizada al hecho seguridad social parte 1 45
Figura 12. Transformación realizada al hecho seguridad social parte 2 45
Figura 13. Transformación realizada al hecho seguridad social parte 3 45
Figura 14. Transformación realizada al hecho seguridad social parte 4 46
Figura 15. Transformación realizada al hecho seguridad social parte 5 46
Figura 16. Transformación realizada al hecho seguridad social parte 6 47
Figura 17. Transformación realizada al hecho seguridad social parte 7 47
Figura 18. Cubos OLAP correspondientes al MD Laboral 52
Figura 19. Mapa de navegación del MD Laboral 53
Figura 20. Vista de análisis Cantidad de pensiones a largo plazo para el mes de septiembre 54
Figura 21. Caso de prueba CU Presentar información sobre la disponibilidad laboral 56
Figura 22. Resultados de las pruebas aplicadas al MD Laboral 58
Figura 23. Comportamiento de los indicadores de la lista de chequeo del Diccionario de Datos 59
Anexo 1. Modelo dimensional correspondiente al MD Laboral 68

Índice de tablas

Tabla 1. Diferencias entre OLAP y OLTP 10
Tabla 2. Actores del sistema 26
Tabla 3. Descripción del CU Extraer información 27

Tabla 4. Descripción del CU Presentar información sobre la disponibilidad laboral 28
Tabla 5. Estándares de codificación 31
Tabla 6. Matriz bus 34
Tabla 7. Tablas ubicadas en los esquemas de la base de datos del MD Laboral 41

Introducción

El desarrollo de las Tecnologías de la Información y las Comunicaciones (TIC) ha traído consigo cambios significativos en la sociedad actual. Su uso contribuye a elevar los resultados y aumentar la competitividad de las empresas. En la actualidad se hace necesario disponer de un acceso rápido y sencillo a la información, razón por la cual numerosas instituciones han enfocado sus esfuerzos en llevar a cabo el control automatizado de los datos históricos que generan. Como consecuencia de estas necesidades se ha producido un amplio proceso de digitalización en diferentes órganos de trabajo, con el propósito de perfeccionar el proceso de análisis y manejo de la información. Todo esto facilita el control estadístico de los resultados obtenidos en una entidad a lo largo del tiempo, permitiendo conocer su estado actual, así como, establecer balances financieros.

Estos avances tecnológicos se han insertado en la sociedad cubana, proporcionando una vía factible para el manejo de grandes cúmulos de datos que se generan en los centros laborales, fundamentalmente en el campo estadístico. La Universidad de las Ciencias Informáticas (UCI), específicamente el departamento de Almacenes de Datos perteneciente al Centro de Tecnologías de Gestión de Datos (DATEC), trabaja en conjunto con el Tribunal Supremo Popular brindando servicios relacionados con bases de datos y análisis de información. El Tribunal Supremo Popular tiene la misión de ejercer la máxima autoridad judicial en nombre del pueblo de Cuba, de forma profesional, con transparencia y humanismo, para contribuir a la seguridad jurídica y al desarrollo de la sociedad socialista. Dentro de los departamentos que constituyen esta entidad se encuentra el de Estadística Judicial, cuya finalidad es presentar indicadores y cifras estadísticas que reflejen el comportamiento y desarrollo de la actividad jurídica en los municipios y provincias del país. Este departamento consta de cuatro áreas, entre las que se encuentra el área Laboral, donde está enmarcado el presente trabajo de diploma.

Las tecnologías y herramientas empleadas en Tribunales para consultar y almacenar la información traen consigo un difícil manejo de los datos correspondiente al área Laboral de esta institución, ocasionando la aparición de las siguientes deficiencias:

- El análisis estadístico se realiza a través de mecanismos no automatizados, poco confiables y en algunas ocasiones resulta tedioso, a causa de que la información es almacenada en herramientas basadas en aplicaciones de oficina como ficheros excel y dbf.
- Para el análisis de la información se necesita ser un especialista del tema con alto conocimiento del negocio.

- Se generan ficheros mensuales trayendo consigo que existan múltiples versiones y grandes cúmulos de datos, dificultándose el proceso de obtención de información estadística.
- El proceso de recuperación y elaboración de informes resulta costoso en tiempo y esfuerzo.

Todo esto obstaculiza la centralización, disponibilidad e integración de la información, dificultando así, el análisis estadístico de diferentes variables relacionadas con los datos que se procesan en el área Laboral.

La búsqueda de mejoras en las formas de almacenar la información y presentar los principales reportes, cruces de variables, indicadores, porcentajes y demás aspectos de interés, es una necesidad urgente para aumentar la disponibilidad de la información.

Por lo planteado anteriormente se define como **problema de la investigación**: ¿Cómo contribuir a la toma de decisiones en el área Laboral del Sistema de Información Estadística Judicial para los Tribunales Populares?

Teniendo la presente investigación como **objeto de estudio** los almacenes de datos, enmarcado en el **campo de acción** Mercado de datos Laboral para el Sistema de Información Estadística Judicial para los Tribunales Populares.

Definiéndose como **objetivo general**: desarrollar el Mercado de datos Laboral para el Sistema de Información Estadística Judicial para los Tribunales Populares que contribuya a la toma de decisiones.

Para dar solución al objetivo general se plantean los siguientes **objetivos específicos**:

- Fundamentar la selección de la metodología, herramientas y tecnologías a utilizar en el desarrollo del Mercado de datos Laboral.
- Realizar el análisis y diseño del Mercado de datos Laboral.
- Realizar la implementación y pruebas al Mercado de datos Laboral.

Para darle solución a los objetivos específicos se definieron las siguientes **tareas de la investigación**:

- Estudio y análisis de las metodologías, herramientas y tecnologías a utilizar en el desarrollo de almacenes de datos.
- Levantamiento de requisitos para definir las necesidades del cliente, así como, identificar los requerimientos de información, funcionales y no funcionales.
- Descripción de los casos de uso del mercado de datos para un mejor entendimiento de estos.
- Definición de la arquitectura del mercado de datos para establecer las bases del desarrollo del mercado de datos.
- Definición de los hechos, las medidas y las dimensiones del mercado de datos para el diseño del modelo de datos.

- Diseño del modelo de datos para recoger procesos fundamentales como: tablas de hechos y dimensiones candidatas en la solución del problema.
- Diseño del subsistema de integración para definir el flujo de datos desde los sistemas fuentes hacia el mercado de datos.
- Diseño del subsistema de visualización para definir el flujo de datos entre el sistema y el cliente.
- Diseño de los casos de pruebas para aplicarlos posteriormente durante la liberación del sistema.
- Implementación del modelo de datos para cumplir con el diseño de la estructura de la base de datos.
- Implementación del subsistema de integración para poblar el mercado de datos.
- Implementación del subsistema de visualización para gestionar los reportes candidatos necesarios y de esta forma satisfacer las necesidades del cliente.
- Aplicación de las listas de chequeo para comprobar el correcto funcionamiento de los subsistemas de integración y visualización.
- Aplicación de los casos de prueba para comprobar el correcto funcionamiento de los reportes candidatos definidos en el sistema.

El presente trabajo de diploma está estructurado de la siguiente manera: introducción, tres capítulos, conclusiones, recomendaciones, bibliografía, referencias bibliográficas y anexos.

Capítulo 1: Fundamentos teóricos de los almacenes de datos

Este capítulo está referido al estudio y análisis del estado del arte de los almacenes de datos y de los mercados de datos, así como, temas relacionados con la tecnología de los almacenes de datos respaldando el presente trabajo, abarcando definiciones, ventajas y desventajas de este tipo de soluciones, además de la metodología y herramientas a emplear.

Capítulo 2: Análisis y diseño del mercado de datos

En este capítulo se definirán los requisitos que permiten diseñar el diagrama de casos de uso del sistema, para lograr un mejor entendimiento de la solución a implementar. Se identificarán los hechos, medidas y dimensiones quedando conformado el modelo de datos de la solución. Además se diseñarán los subsistemas de almacenamiento, integración y visualización de los datos.

Capítulo 3: Implementación y pruebas del mercado de datos

En este capítulo se hace referencia a la implementación de la solución, abordando temas como la implementación de los subsistemas de almacenamiento, integración y visualización, teniendo en cuenta

los requisitos y necesidades del cliente. Además se realizarán pruebas al mercado de datos a través de la aplicación de las listas de chequeo y los casos de prueba para obtener un resultado de calidad.

Capítulo 1: Fundamentos teóricos de los almacenes de datos

Introducción

En este capítulo se abordarán los principales conceptos y definiciones relacionados a los Almacenes de Datos (AD) y Mercados de Datos (MD), así como, sus características, arquitectura, ventajas, desventajas y tendencias actuales de estos. Se realizará un estudio de la metodología y herramientas a utilizar para el desarrollo de un MD que permitirá al Tribunal Supremo Popular mejorar el proceso de toma de decisiones.

1.1 Almacenes de datos

En la actualidad el manejo de grandes volúmenes de información, así como, el análisis de datos que se pueden extraer de ellos constituye una prioridad para toda aquella empresa que pretenda tener éxito en su ámbito de acción. Con el creciente desarrollo de las tecnologías de información han aparecido soluciones a estos problemas, entre las que pueden mencionarse los AD.

Ralph Kimball, conocido autor en el tema de los AD o *Data Warehouse*, define un AD como "una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis" (1).

La definición más difundida y aceptada de un AD pertenece a William H. Inmon, licenciado en Ciencias Matemáticas y máster en Ciencias de la Computación quien plantea que un AD es "una colección de datos orientados a temas, integrados, no volátiles y variante en el tiempo, organizados para soportar necesidades empresariales" (2).

En la presente investigación se asumirá la definición de AD propuesta por Inmon, pues se considera el concepto más completo partiendo de las características de un AD.

Características de los almacenes de datos

Según la definición de Inmon, un almacén de datos se caracteriza por ser:

- Orientado a temas: los datos en la base de datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí. La información se clasifica en base a los aspectos que son de interés para la empresa.
- Variable en el tiempo: los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones. Los datos son relativos a un período de tiempo y estos deben ser integrados periódicamente.
- No volátil: una vez almacenado un dato éste se convierte en información de sólo lectura, y no puede ser modificado ni eliminado por los usuarios finales.
- Integrado: los datos provenientes de diferentes fuentes son integrados en un mismo repositorio con el objetivo de eliminar las inconsistencias presentes en la información (1).

Ventajas y desventajas de usar un almacén de datos

Algunas de las ventajas que trae consigo la utilización de un AD son:

- Decisiones soportadas por datos fiables, coherentes y homogéneos.
- Entornos amigables, pues los directivos o analistas de información acceden a los datos mediante interfaces sencillas de manejar.
- Identificar nuevas oportunidades de negocio y tomar decisiones estratégicas.
- Aumento de la competitividad en el mercado (3).

El empleo de los AD puede originar algunos inconvenientes, entre los que pueden mencionarse:

- La subestimación del tiempo requerido para extraer, limpiar y cargar los datos en el almacén.
- Los gastos de mantenimiento pueden ser muy elevados.
- La construcción de un almacén de datos puede requerir de mucho tiempo (4).

1.1.1 Tendencias actuales del uso de almacenes de datos en el mundo

Los AD se han convertido actualmente en una de las principales vías para soportar el proceso de toma de decisiones de las empresas. Existen a nivel mundial numerosas compañías que hacen uso de estas tecnologías, dentro de ellas se pueden mencionar la Twentieth Century Fox, que utiliza Inteligencia de Negocio o *Business Intelligence* (BI) para predecir el grado de popularidad de sus actores, argumentos y películas. La firma de entretenimiento estadounidense Walt Disney Studios Home Entertainment (WDSHE) ha hecho uso de esta importante solución en su departamento de ventas con resultados exitosos, planteándose actualmente extender su uso en todas las áreas de la organización. La compañía *Wal-Mart*, considerada como una de las más grandes de su tipo en el mundo, implantó la solución BI Neoview de HP, considerada como una plataforma de AD que integra productos y servicios que permiten realizar el análisis de datos de sus puntos de venta, con propósitos de capitalizar la información y optimizarla. En Cuba existen centros que utilizan AD como respuesta a la necesidad de obtener información valiosa y confiable para llevar a cabo el proceso de toma de decisiones. La Empresa de Proyectos de Arquitectura e Ingeniería (EMPAI) de Matanzas cuenta con un AD para gestionar, organizar y llevar el control de la información correspondiente a la gestión contable de la entidad. De igual forma la Oficina Nacional de Estadística e Información (ONEI) cuenta con un AD desarrollado en la UCI, el cual provee a la institución de una solución que le facilita tomar decisiones de manera efectiva.

1.1.2 Arquitectura de un almacén de datos

La arquitectura de los AD se define teniendo en cuenta las características y necesidades de la empresa en la cual se implanta. Sin embargo, las arquitecturas cuentan con aspectos comunes, dentro de los que se pueden mencionar:

- Repositorio de datos: el repositorio de datos operacionales es la fuente donde se encuentran los datos primarios, actuales e integrados, por lo tanto es el encargado de suministrar datos al sistema, estos datos operacionales pueden provenir de fuentes externas, estaciones de trabajo o servidores privados y principalmente de sistemas *mainframe*[1].
- Gestor de carga: también conocido como proceso de ETL (Extracción, Transformación y Carga, en inglés *Extraction, Transformation and Loading*), es el encargado de realizar las funciones de extracción, transformación y carga de las fuentes de datos del AD.
- Gestor del almacén de datos: realiza las operaciones relacionadas con la gestión de los datos dentro del AD utilizando herramientas específicas para la transformación de los datos, creación de copias de seguridad y archivado de datos, además de realizar el análisis de los datos para mantener su coherencia.
- Metadatos: describen la estructura de los datos dentro del almacén y pueden ser utilizados por los gestores de carga del AD y de consultas.
- Sistemas gestores de bases de datos (SGBD): proporciona métodos para mantener la integridad de los datos por medio del almacenamiento, modificación y extracción de la información en una BD.
- Gestor de consultas: encargado de gestionar las operaciones asociadas a las consultas.
- Herramientas de acceso para usuarios: tienen como objetivo proporcionar a los usuarios un medio de acceso a los datos, facilitando la toma decisiones estratégicas (5).

1.2 Mercado de datos

Un MD o *Data Mart*, como se conoce en inglés, es un subconjunto de datos de un almacén relativos a los requisitos de un departamento o área de negocio específico. Este subconjunto de datos puede funcionar de forma autónoma, o bien enlazado al AD (5).

[1]Mainframe: también conocido como computador central, son computadoras usadas principalmente por compañías para el procesamiento de grandes cantidades de datos.

Características de los mercados de datos

Dentro de las principales características de los MD se pueden encontrar:

- Se centran en los requisitos de los usuarios asociados a un área de negocio o departamento específico.
- Son más sencillos a la hora de utilizarlos y comprender sus datos, debido a que la cantidad de información que contienen es mucho menor que en los AD.
- Provee una interfaz de consulta que permite al usuario analizar la información para la toma de decisiones (5).

Ventajas de los mercados de datos

Algunas de las principales ventajas de aplicar un MD a un negocio son:

- Son simples de implementar.
- Conllevan poco tiempo de construcción y puesta en marcha.
- Permiten manejar información confidencial.
- Permiten satisfacer las necesidades de los especialistas de un área específica de una organización.
- Una misma organización puede tener varios MD, lo que permite tener la información dividida por áreas de análisis (6).

1.3 Integración de datos

La integración de datos permite combinar la información existente en diversos sistemas fuentes. ETL es un proceso de integración de datos que involucra extraer información de las fuentes, con el objetivo de eliminar sus inconsistencias e integrarla para ser cargada al AD.

Extracción: consiste en acceder a diversas fuentes y recuperar la información que será integrada en el AD. El propósito principal de la fase de extracción es capturar y copiar los datos requeridos de una o más fuentes de datos (7).

Transformación: en esta fase se realizan las operaciones de estandarización y limpieza de datos. La transformación se encarga de erradicar las inconsistencias en la codificación y los formatos de los datos que puedan existir dentro de la fuente de datos (7).

Carga: en esta fase se carga hacia la BD la información depurada en el proceso de transformación. Los datos cargados serán utilizados para la generación de reportes, los cuales se podrán analizar para contribuir a la toma de decisiones (7).

1.4 Inteligencia de negocio

BI es el proceso por el cual la información es sistemáticamente capturada, analizada y distribuida como conocimiento, para que los usuarios puedan tomar decisiones a partir de ella. En una empresa es necesario tomar decisiones día a día, basadas en la información existente en BD. La toma de decisiones implica riesgos que son necesarios minimizar, es ahí donde entran en juego las soluciones de BI (8):

- Sistemas de Soporte de Decisiones, en inglés *Decision Support Systems* (DSS).

Los DSS son sistemas de información basados en computadora, que combinan modelos y datos para intentar resolver problemas relacionados con la toma de decisiones, utilizando una interfaz amigable para el usuario. Estos generalmente son aplicaciones de computador, junto con un componente humano que puede filtrar a través de grandes cantidades de datos y escoger entre numerosas opciones (9). Dentro de los DSS se encuentran los AD, que tienen como objetivo responder a las necesidades de información de una organización, ofreciendo a los usuarios la posibilidad de consultar y analizar datos homogéneos y fiables.

- Sistemas de Información Ejecutiva, en inglés *Executive Information Systems* (EIS).

Un sistema de información ejecutiva es una herramienta de *software*, basada en un DSS, que provee a los ejecutivos de un acceso sencillo a información interna y externa de su empresa. La finalidad principal es que el ejecutivo tenga a su disposición un panorama completo del estado de los indicadores de negocio que le afectan al instante, manteniendo también la posibilidad de analizar con detalle aquellos que no estén cumpliendo con las expectativas establecidas por la compañía, para determinar el plan de acción más adecuado (10).

- Tecnologías OLAP (Procesamiento Analítico en Línea o por sus siglas en inglés, *On-Line Analytical Processing*)

Tecnología de análisis de datos que utiliza estructuras multidimensionales para proporcionar un acceso rápido a los datos, facilitando la consulta y análisis de grandes cantidades de información contenida en los AD (6). El principal elemento de estas tecnologías es el cubo OLAP, que organiza los datos mediante dimensiones, jerarquías y medidas en una estructura multidimensional, facilitando el análisis de los datos de una organización.

- Minería de Datos, en inglés *Data Mining*.

La minería de datos es el conjunto de técnicas encaminadas a la extracción de conocimiento procesable, implícito en las BD (11).

Procesamiento Analítico en Línea

Las tecnologías OLAP facilitan el acceso a la información, ofreciendo una respuesta rápida a las consultas de los usuarios para contribuir a la toma de decisiones de una organización.

A continuación se presentan algunas de las características del OLAP:

- Presenta una visión multidimensional lógica de los datos del AD, independiente de su forma de almacenamiento.
- Crea resúmenes, adiciones y jerarquías.
- Comprende consultas interactivas y análisis de los datos. Permite una profundización hacia niveles más detallados o un ascenso a niveles superiores de resumen y adición.
- Responde con rapidez a las consultas de modo que el proceso de análisis no se interrumpe (12).

Procesamiento de Transacciones en Línea o por sus siglas en inglés, *On-Line Transaction Processing* (OLTP)

A continuación se presentan algunas de las características del OLTP:

- Datos almacenados cambian continuamente.
- El historial de datos suele limitarse a los datos actuales o recientes.
- El acceso a los datos está optimizado para tareas frecuentes de lectura y escritura.
- Usa diagrama entidad relación (13).

La tecnología OLAP, basada en el modelo multidimensional de datos, facilita el análisis en línea, administración y ejecución de consultas, convirtiéndose en una ventaja frente al OLTP, donde los datos son almacenados para realizar control y operaciones sobre los mismos.

Tabla 1. Diferencias entre OLAP y OLTP (12)

	OLAP	**OLTP**
Función	Soporte a las decisiones	Operación diaria
Datos	Históricos, resumidos, multidimensionales, integrados, consolidados, opcionalmente detallados	Actuales, detallados, relacionales, aislados
Unidad de trabajo	Consultas complejas	Transacciones simples
Diseño de la BD	Orientado a una materia	Orientado a una aplicación

Modos de almacenamiento:

ROLAP (Procesamiento Analítico Relacional o en inglés *Relational On-line Analytical Process*): son los modelos en los cuales la organización física de los datos se implementa sobre tecnología relacional disponiendo de algunas facilidades para mejorar el rendimiento. Cuenta con todos los beneficios de un Sistema de Administración de Base de Datos Relacional o *Relational Database Management System* (RDBMS) a los cuales se les provee extensiones y herramientas para poder utilizarlo como un Sistema Gestor de AD (14).

Las ventajas que permite ROLAP son:

- Uso fundamental de la seguridad e integridad de la base de datos.
- Datos y estructura más dinámicos.
- Escalable para grandes volúmenes.
- Permite el análisis de una enorme cantidad de datos.
- Los datos pueden ser compartidos con aplicaciones SQL (Lenguaje de Consulta Estructurado o por sus siglas en inglés S*tructured Query Language*) (15).

MOLAP (Procesamiento Analítico Multidimensional en Línea o en inglés *Multidimensional On-line Analytical Process*): son los modelos en los cuales la organización física de los datos se realiza en estructuras multidimensionales de manera que la representación externa y la interna coincidan. Disponen de estructuras de almacenamiento específicas y técnicas de compactación de datos que favorecen el rendimiento del almacén de datos (14).

HOLAP (Procesamiento Analítico en Línea Híbrido o en inglés *Hybrid On-line Analytical Process*): son los modelos híbridos entre MOLAP y ROLAP, combinan estas dos implementaciones para almacenar algunos datos en un motor relacional y otros en una base de datos multidimensional (14).

En el desarrollo del MD Laboral se decide utilizar como modo de almacenamiento de datos ROLAP, pues el Sistema Gestor de Bases de Datos (SGBD) a utilizar en la solución es PostgreSQL 9.1 que permite el almacenamiento relacional.

Beneficios de implementar BI:

- Reducción de los costos de obtención de la información.
- Disponibilidad de un compendio descriptivo de toda la información contenida en los sistemas fuentes.
- Reducción de la brecha entre la información y las áreas que la requieren como soporte a la toma de decisiones.

- Disponibilidad de información consolidada e integrada de toda la organización (8).

1.5 Metodologías para el desarrollo de un almacén de datos

Una metodología de desarrollo de *software* es un conjunto de pasos y procedimientos que sirven como guía para desarrollar un producto. Estas surgen ante la necesidad de realizar una serie de procedimientos, técnicas, herramientas y soporte documental a la hora de desarrollar un *software* (16).

Existen diferentes metodologías que guían la construcción de los AD para hacer más fácil su desarrollo. Actualmente dos tendencias sobresalen frente a las demás, sirviendo de guías a la comunidad mundial en el tema de los AD. Estas tendencias se conocen como enfoque de Kimball y enfoque de Inmon, debido a sus creadores Ralph Kimball y William H. Inmon. La propuesta de Inmon se basa en un enfoque descendente (*Top_Down*), con la idea de implementar un AD, del cual se nutrirán posteriormente los MD que se desarrollen para las diferentes áreas de la empresa. Por otra parte la de Kimball proporciona un enfoque ascendente (*Bottom_Up*), permitiendo la implementación de MD que luego serán integrados en un gran AD. Este enfoque se centra en el modelado dimensional, haciendo énfasis en el diseño de los MD, asegurando en gran parte el éxito del proyecto. Además existen otras metodologías para el desarrollo de AD, entre las que se encuentran: Metodología Hefesto, Desarrollo de almacenes de datos dirigidos por modelos (Trujillo), Data Warehouse Engineering Process (DWEP) y Rapid Warehousing Methodology (RWM).

Metodología de desarrollo de almacenes de datos utilizada en DATEC

La propuesta de Metodología de Desarrollo de Almacenes de Datos utilizada en el centro DATEC toma como base la metodología de Kimball para definir los aspectos específicos del desarrollo de AD. En su concepción también se tiene en cuenta lo planteado en el Programa de Mejora llevado a cabo en la UCI para incorporar los temas asociados a CMMI (Modelo de Integración de Capacidades de Madurez o por sus siglas en inglés, *Capability Maturity Model Integrated*), así como, la incorporación de una etapa de prueba que fortalece en gran medida la calidad de la solución a desplegar.

Ciclo de vida de la propuesta de Metodología de Desarrollo de Almacenes de Datos

Para definir las fases del ciclo de vida de la metodología se tuvieron en cuenta las fases propuestas por la Metodología de Kimball y el Programa de Mejora llevado a cabo en la UCI.

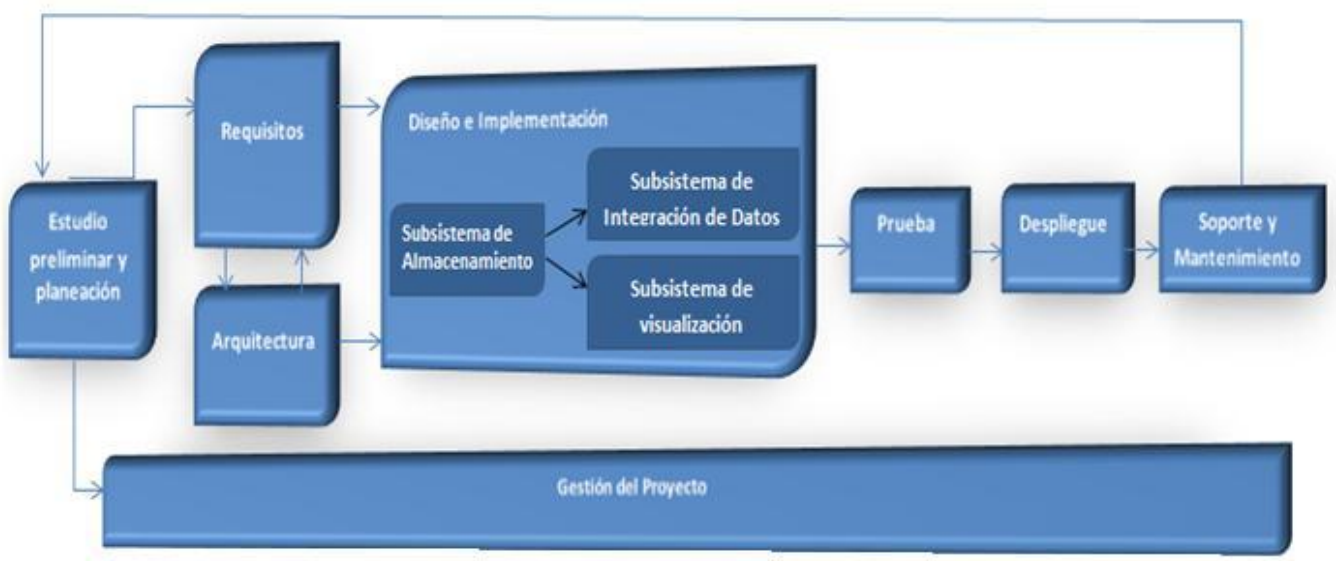


Figura 1. Ciclo de vida de la propuesta de Metodología de Desarrollo de Almacenes de Datos

- **Estudio preliminar y planeación**: se realiza un estudio minucioso en la entidad cliente. Esto incluye un diagnóstico integral de la organización, con el fin de determinar qué es lo que se desea construir y qué condiciones existen para el desarrollo y montaje de la misma. Además se llevan a cabo las tareas de planeación del proyecto.
- **Requisitos**: se realiza el proceso de entrevistas al cliente para determinar los requisitos de información. Se hace levantamiento detallado de las fuentes de datos para validar la disponibilidad de la información. Además se definen los requisitos funcionales y no funcionales de la solución y se hace el análisis de los requisitos que dan paso al diseño e implementación.
- **Arquitectura:** se definen las vistas arquitectónicas de la solución, aspectos como, los subsistemas y componentes, la seguridad, la comunicación y la tecnología a utilizar.
- **Diseño e Implementación:** se define el diseño de las estructuras de almacenamiento de datos, se diseñan los procesos de integración de datos como, el mapa lógico de datos, los cubos OLAP para la presentación de la información, así como el diseño gráfico de la aplicación definido por el cliente. Después se implementan cada uno de los subsistemas (repositorio de datos, integración de datos, presentación de datos).
- **Prueba:** se realizan las pruebas que validan la calidad del producto, comenzando por las Pruebas de Unidad, las Pruebas de Integración y Sistema, hasta llegar a las Pruebas de Aceptación con el

cliente final. Esta fase no es la única en la que se realizan pruebas durante el desarrollo del proyecto, en todas las fases hay actividades de aseguramiento de la calidad.

- **Despliegue:** consta de dos etapas, despliegue piloto, donde se configuran los servidores necesarios y se instalan las herramientas según la arquitectura definida, se cargan una muestra de los datos en un ambiente controlado, con el fin de mostrarle al cliente final el sistema en funcionamiento. Una vez aceptada la solución por el cliente, se realiza la carga histórica de los datos, puede ser en el mismo entorno que el despliegue piloto u otro, todo depende de las condiciones que establezca el cliente. Además se realiza la capacitación y transferencia tecnológica de la solución a los clientes. El resultado fundamental es la solución desplegada en el entorno real y en correcto funcionamiento.
- **Soporte y Mantenimiento:** comienza cuando la solución está implantada y en explotación, y se ejecuta según el contrato firmado y las condiciones de soporte establecidas. Puede realizarse a través de variados servicios, que pueden ser soporte en línea, vía telefónica, web, correo u otros y el acompañamiento al cliente. Además se realizan las tareas de manteniendo de la aplicación tan necesarias para este tipo de desarrollo y que garantiza el adecuado funcionamiento y crecimiento del AD.
- **Gestión del proyecto:** esta fase se ejecuta a lo largo de todo el ciclo de vida del proyecto. Es aquí donde se controla, gestiona y chequea todo el desarrollo, los gastos, las utilidades, los recursos, las adquisiciones, los planes y cronogramas entre otras actividades relacionadas con la gestión de proyectos. Esta fase es el pilar base del proyecto y si no se ejecuta de forma continua y correcta el proyecto puede fracasar (17).

Para el presente trabajo de diploma se emplea la propuesta de Metodología de Desarrollo de Almacenes de Datos utilizada en DATEC, cubriéndose en el desarrollo de la solución hasta la fase Prueba, debido a que las restantes fases serán llevadas a cabo por especialistas del centro, pues no se encuentran dentro del alcance de la investigación.

1.6 Herramientas para la creación de los almacenes de datos

La selección de las herramientas para la creación de los AD constituye un eslabón fundamental para llevar a cabo la implementación de una solución que cumpla con las expectativas del cliente.

1.6.1 Herramienta de modelado de datos

Existen en el mundo varias herramientas que brindan ayuda y asistencia a los analistas, ingenieros de *software* y desarrolladores durante el ciclo de vida de desarrollo de un producto. Dentro de estas se

encuentran las herramientas de ingeniería de software asistida por computadora o por sus siglas en inglés, *Computer Aided Software Engineering* (CASE) que proveen de apoyo en el modelado de soluciones.

A continuación se nombran algunas herramientas CASE orientadas al lenguaje unificado de modelado o por sus siglas en inglés, *Unified Modeling Language* (UML).

- Rational Rose
- Visual Paradigm for UML

La herramienta de modelado Rational Rose se caracteriza por (18):

- Disponible en múltiples plataformas.
- Permite desarrollo multiusuario.
- Genera documentación del sistema.
- Es software propietario.

La herramienta de modelado Visual Paradigm for UML posee (19):

- Soporte de UML versión 2.1.
- Modelado colaborativo con *Concurrent Versions System* (CVS) y Subversion.
- Disponibilidad en múltiples plataformas.
- Ingeniería inversa.
- Varios idiomas.
- Editor de detalles de casos de uso.
- Generación de código.
- Distribución automática de diagramas.
- Integración con Visio.
- Permite exportar los diagramas como imágenes.
- Editor de figuras.
- Es una herramienta privada.

En el presente trabajo de diploma se decidió utilizar como herramienta de modelado **Visual Paradigm for UML 8.0** pues propicia un conjunto de ayudas para el desarrollo de programas informáticos, desde la planificación, pasando por el análisis y el diseño, hasta la generación del código fuente de los programas y la documentación. Visual Paradigm ha sido concebido para soportar el ciclo de vida completo del proceso de desarrollo del *software* a través de la representación de todo tipo de diagramas (19).

Cabe destacar que en la UCI se paga una licencia que respalda el uso de esta herramienta. Además en el centro DATEC específicamente en el departamento de Almacenes de Datos se realizó la implementación del producto Vpmd para la herramienta de modelado anteriormente mencionada que consiste en una extensión que brinda la posibilidad de modelar las estructuras de datos dimensionales. La utilización de este producto posibilitará obtener el modelo físico de datos, del cual se obtendrá el script SQL para el gestor de base de datos PostgreSQL y permitirá generar una plantilla XML[2] para el trabajo con la herramienta Pentaho Schema Workbench empleado en la etapa de BI.

1.6.2 Herramientas de gestión y administración de bases de datos

Un SGBD es un tipo de *software* muy específico, dedicado a servir de interfaz entre la BD, los usuarios y las aplicaciones que lo utilizan. El propósito general de los SGBD es el de manejar de manera clara, sencilla y ordenada un conjunto de datos (20).

Entre los gestores de base de datos más utilizados se encuentran:

- MySQL
- Oracle
- PostgreSQL

Algunas de las características de MySQL son (21):

- Funciona en diferentes plataformas.
- Es *software* propietario y está patrocinado por una empresa privada, que posee el *copyright* de la mayor parte del código.
- Ofrece un sistema de privilegios y contraseñas seguro mediante el cifrado del tráfico de contraseñas al conectarse a un servidor.
- El servidor puede proporcionar mensajes de error a los clientes en muchos idiomas.
- Soporta gran cantidad de datos.

Entre las características de la herramienta Oracle se pueden mencionar que (22):

- Es un sistema multiplataforma, disponible en Windows, Linux y Unix.
- Brinda soporte a la mayoría de los lenguajes de programación.
- Posee un rico diccionario de datos
- Es un producto de elevado precio, por lo que generalmente se utiliza en empresas muy grandes y multinacionales.

[2]XML: Extensible Markup Language

- Los costos de soporte técnico y mantenimiento son elevados.

Entre las principales características que posee PostgreSQL se encuentran (23):

- Soporta distintos tipos de datos. Además del soporte para los tipos base, soporta datos de tipo fecha, monetarios, elementos gráficos, datos sobre redes y cadenas de bits, entre otros. Permite la creación de tipos propios.
- El código fuente se encuentra disponible para todos sin costo alguno.
- Incorpora arreglos como una estructura de datos.
- Soporta el uso de índices, reglas y vistas.
- Incluye herencia entre tablas.
- Permite la gestión de diferentes usuarios y los permisos asignados a cada uno de ellos.
- Funciona en los sistemas operativos Linux, UNIX y Windows.
- Debido a la liberación de la licencia, PostgreSQL se puede usar, modificar y distribuir de forma gratuita para cualquier fin.

La herramienta de gestión de BD escogida fue el **PostgreSQL 9.1** pues constituye un SGBD relacional orientado a objetos, que incluye características como la herencia, tipos de datos, funciones, restricciones, disparadores, reglas e integridad transaccional, liberado bajo la licencia BSD (*Berkeley Software Distribution*) (23). Cabe destacar que esta herramienta es libre constituyendo una ventaja frente a los SGBD MySQL y Oracle que son herramientas privadas que requieren de grandes sumas de dinero para lograr su obtención.

La herramienta de administración de BD escogida fue **PgAdmin III 1.14**. PgAdmin es una aplicación gráfica para administrar el gestor de BD PostgreSQL, siendo la más completa y popular con licencia *open source*. Está escrita en C y es diseñada para responder a las necesidades de todos los usuarios, desde escribir consultas SQL simples hasta desarrollar BD complejas. La interfaz gráfica soporta todas las características de PostgreSQL y facilita enormemente la administración. La aplicación puede utilizarse en plataformas como Linux, FreeBSD, Solaris, Mac OSX y Windows (24).

1.6.3 Herramientas para el proceso de integración de datos

La calidad de los datos es un tema frecuente en el campo de los AD, por lo que existen herramientas que ayudan a monitorear los datos para verificar que tengan la calidad necesaria. Para identificar inconsistencias en los datos y realizar el perfilado de estos se escoge la aplicación de código abierto **DataCleaner 1.5.4**. Estas actividades ayudan a administrar y supervisar la calidad de los datos, con el fin de garantizar que la información sea útil.

Sus características incluyen:

- Validación de los datos: el validador le dará un resultado que puede ser interpretado como bueno o malo.
- Compatibles con diferentes tipos de base de datos: Oracle, MySQL, PostgreSQL, Firebird, SQLite.
- Es multiplataforma y está desarrollado en *Java* (25).

Para realizar la integración de la información existen herramientas encargadas de la extracción, transformación y carga de los datos, entre las que se puede mencionar el **Pentaho Data Integration (PDI) 4.2.1**. PDI es una herramienta libre muy potente, así como, una de las más antiguas y utilizadas con gran soporte técnico. Posee gran cantidad de conectores y brinda la posibilidad de crear flujos de trabajo integrados con transformaciones de datos de manera muy sencilla y funcional (26).

Cuenta con las siguientes características (26):

- Cada proceso es creado con una herramienta gráfica donde se especifica qué se va hacer sin necesidad de escribir un código que indique cómo hacerlo.
- Admite una amplia gama de formatos de entrada y salida, incluyendo archivos de texto, hojas de datos, archivos XML, entre otros.
- Basado en repositorio, facilita la reutilización de componentes como transformación, colaboración y administración de modelos, conexiones.
- Depurador integrado.
- Librería de transformaciones completa con más de 100 objetos de mapeo.
- Fácil de instalar y configurar.
- Es multiplataforma: Windows, Macintosh, Linux.
- Sin costes de licencia.

Cuenta con cuatro componentes fundamentales:

- SPOON: permite diseñar transformaciones ETL usando el entorno gráfico.
- PAN: es un motor de transformación que realiza funciones tales como lectura, manipulación, y escritura de datos hacia y desde varias fuentes de datos. Permite la ejecución de los trabajos y las transformaciones.
- CHEF: para el diseño de la carga de datos.
- KITCHEN: para la ejecución de los trabajos Batch diseñados con CHEF.

1.6.4 Herramientas para el proceso de inteligencia de negocio

Las herramientas de inteligencia de negocio permiten manipular la información de las organizaciones, para lograr un mayor entendimiento de los datos que se manejen y que beneficie la toma de decisiones.

Mondrian Schema Workbench 3.2.1 es una interfaz de diseño que permite crear y probar esquemas de cubos OLAP visualmente. Los archivos de esquemas son modelos de metadatos XML que se crean en una estructura específica utilizada por el motor de Mondrian. La estructura de estos modelos se pueden considerar en forma de cubos, que utilizan hechos existentes y tablas de dimensiones que se encuentran en la BD (27).

Ofrece las siguientes funcionalidades:

- Editor de esquema integrado con el origen de datos subyacente para su validación.
- Probar consultas MDX (Expresiones multidimensionales o por sus siglas en inglés, *Multidimensional Expressions*) en contra del esquema de BD en pantalla.
- Examinar BD de estructura subyacente en pantalla.

Pentaho BI Server 3.10 provee el soporte y la infraestructura necesarios para crear soluciones de inteligencia empresarial a problemas de negocios. El marco proporciona los servicios básicos, incluidos autenticación, registro, auditoría, servicios web y motor de reglas. La plataforma cuenta con un motor de solución que integra reportes, análisis, tableros de comandos y componentes de minería de datos. Además esta herramienta incluye el servidor Mondrian OLAP Server para la consulta de la información y al Apache Tomcat como servidor web (28).

Algunas de sus ventajas son:

- Administra seguridad de usuarios.
- Integración con procesos de negocio.
- Administra y programa reportes.
- Está diseñado para integrarse fácilmente en cualquier proceso de negocio (28).

Mondrian OLAP Server 3.0.4 es un servidor OLAP *open source* escrito en *Java* que gestiona la comunicación entre una aplicación OLAP y la BD con los datos fuente. Utiliza MDX como lenguaje de consulta. Permite realizar consultas al AD y posibilita que los resultados sean presentados mediante un navegador, de modo que el usuario pueda realizar las actividades típicas de navegación. Entre sus principales características se encuentra la facilidad para el análisis de grandes volúmenes de información almacenados en BD que soporten Java Database Connectivity (JDBC) (27).

Apache Tomcat 6.0.29 es una implementación de *software* de código abierto de tecnologías Java Servlet y Java Server Pages (JSP). Se desarrolla en un entorno abierto y participativo y publicado bajo la licencia Apache versión 2. Apache Tomcat tiene la intención de ser una colaboración de los mejores desarrolladores de su clase en todo el mundo. Dado que Tomcat fue escrito en *Java*, funciona en cualquier sistema operativo que disponga de la máquina virtual de este lenguaje de programación y es actualmente un servidor web muy utilizado a la hora de trabajar con Java en entornos web (29).

Pentaho Report Designer 3.8.2 es una herramienta que forma parte de la unidad de reportes del Pentaho (Pentaho Reporting), que simplifica el proceso de generación de reportes, permitiendo a los diseñadores crear rápidamente informes sofisticados y ricos visualmente. Ofrece un entorno gráfico familiar y una estructura de reporte bastante acertada y flexible, para darle libertad al diseñador de generar reportes que se adapten a su gusto y necesidad (30).

Algunas de sus características son:

- Diseñador gráfico basado en “arrastrar y soltar” (*drag & drop*) que provee completo control de acceso a los datos, agrupaciones, cálculos, gráficas y formatos.
- Opciones de salida flexibles incluyendo los formatos Adobe PDF, HTML (Lenguaje de Marcado Hipertextual o por sus siglas en inglés, *HyperText Markup Language*) y Microsoft Excel (30).

1.7 Conclusiones del capítulo

En este capítulo se profundizó en la metodología, herramientas y tecnologías a utilizar en el desarrollo de AD, permitiendo obtener conocimientos que respaldan su desarrollo. La metodología de desarrollo de AD que se decidió utilizar, permitirá la implementación de un producto que cumpla con las expectativas del cliente. Se definió como herramienta a utilizar para la realización del diseño el Visual Paradigm 8.0 y para el proceso de ETL se seleccionó el Pentaho Data Integration (PDI) 4.2.1 por la facilidad de uso, mantenimiento y flexibilidad a la hora de realizar las transformaciones. Se seleccionó DataCleaner 1.5.4 para el depurado y limpieza de los datos, y en la realización de los procesos de BI se utilizarán las herramientas Mondrian OLAP Server 3.0.4, Mondrian Schema Workbench 3.2.1, Pentaho Report Designer 3.8.2 y el Pentaho BI Server 3.10 sobre el servidor Apache Tomcat 6.0.29.

Capítulo 2: Análisis y diseño del mercado de datos

Introducción

En este capítulo se describirá el negocio que servirá como punto de partida para el análisis y diseño del MD Laboral. Se definirán los requisitos funcionales, no funcionales y de información que permitirán diseñar el diagrama de casos de uso del sistema, para lograr un mejor entendimiento de la solución a implementar. Se identificarán los hechos, medidas y dimensiones que conformarán el modelo de datos del MD Laboral. Además, se diseñarán los subsistemas de almacenamiento, integración y visualización de los datos.

2.1 Descripción del negocio

El Tribunal Supremo Popular, órgano encargado de impartir justicia en nombre del pueblo de Cuba, recibe grandes cantidades de información de diferentes sectores del país, que se manejan y analizan en el Departamento Independiente de Estadística Judicial perteneciente a dicha institución. La recogida de información referente a las áreas que integran este departamento, se realiza mediante modelos estadísticos definidos que provienen de organismos a nivel provincial que se subordinan al Tribunal Supremo Popular. El área Laboral perteneciente al departamento previamente mencionado se encarga de llevar un control de los procesos de demandas, apelaciones, disponibilidad laboral y seguridad social. La información llega a la especialista que atiende el área Laboral por correo mensualmente y es almacenada en ficheros excel y dbf[3], provocando tardanza en la elaboración de informes y resultando costoso en tiempo y esfuerzo. Esta información es analizada atendiendo a diferentes indicadores, permitiendo llevar el control del estado de los procesos que se llevan a cabo en los centros laborales. El indicador3351 está compuesto por los indicadores de las decisiones prejudiciales del Órgano de Justicia Laboral de Base (O.J.L.B) asociadas a los temas de indisciplinas y derechos laborales. El indicador3352 contendrá los indicadores por los que se analizan las demandas realizadas por la administración y los trabajadores. Estas demandas son solucionadas por sentencias y autos[4]. El análisis de esta información es crucial para elaborar estadísticas destinadas a satisfacer los requerimientos informativos de los más altos niveles del gobierno.

[3]DBF: Database file.

[4]Autos: Son las demandas que se archivan o se anulan.

2.2 Necesidades de los usuarios

Debido a la importancia que tiene conocer las necesidades de información de los usuarios, se realizaron reuniones con la especialista del área Laboral perteneciente al Departamento Independiente de Estadística Judicial del Tribunal Supremo Popular, con el objetivo de analizar la información referente a dicha área. Luego de realizar un estudio de la organización y del negocio, las necesidades de los usuarios se enfocaron en el análisis de la información relacionada con las demandas, apelaciones, disponibilidad laboral y seguridad social.

2.3 Especificación de requerimientos

En la etapa de análisis del desarrollo de un MD se realiza un estudio del negocio para identificar y alcanzar un mayor conocimiento sobre las necesidades de la organización. Mediante el análisis se descubren aspectos importantes necesarios para el desarrollo del MD, tales como los requisitos de información, los funcionales y los no funcionales que luego serán archivados en la especificación de requisitos del negocio.

2.3.1 Requisitos de información

Los requisitos de información (RI) describen la información que debe estar disponible y almacenada en el sistema para satisfacer las necesidades de los usuarios. A continuación se especifican algunos de los 26 RI especificados en el documento DATEC_SIEJT_Laboral_0113_Especificación de Requisitos de Software, ubicado en el expediente del proyecto.

RI 1. Obtener la cantidad de procesos de disponibilidad laboral por tipo de causal, estado del proceso, período de tiempo, provincia y tribunal provincial.

RI 3. Obtener la cantidad de procesos apelados por período de tiempo, provincia y tribunal provincial.

RI 9. Obtener la cantidad de pensiones por período de tiempo, tipo de causal, provincia, estado del seguro social y tribunal provincial.

RI 11. Obtener la cantidad de expedientes por demandantes dado el período de tiempo, DPA[5], tipo de indicador3352 y tribunal municipal.

RI 13. Obtener la cantidad de demandantes por tipo de indicador3351, período de tiempo, DPA y tribunal municipal.

RI 14. Obtener la cantidad de demandas resueltas por período de tiempo, DPA, tipo de indicador3352 y tribunal municipal.

[5]DPA: División Político Administrativa.

RI 23. Obtener la cantidad de expedientes revisados y apelados por tipo de indicador3352, período de tiempo, DPA y tribunal municipal.

RI 24. Obtener la cantidad de medidas aplicadas por el O.J.L.B a causales radicadas por indisciplinas dado el período de tiempo, DPA, tipo de indicadores3351 y tribunal municipal.

RI 26. Obtener la cantidad de decisiones del O.J.L.B a causales radicadas por derechos laborales dado el período de tiempo, DPA, tipo de indicadores3351 y tribunal municipal.

2.3.2 Requisitos funcionales

Los requisitos funcionales (RF) son las funcionalidades que debe cumplir la solución a desarrollar, de acuerdo con las necesidades del cliente. A continuación se muestran algunos de los 18 RF identificados en el negocio y que son especificados en el documento DATEC_SIEJT_Laboral_0113_Especificación de Requisitos de Software ubicado en el expediente del proyecto.

RF 1. Autenticar usuario.

RF 2. Adicionar roles.

RF 3. Eliminar roles.

RF 13. Los datos del Sistema de Información Estadística Judicial para Tribunales se adquirirán a través de los excel y dbf que contienen información asociada a los procesos de disponibilidad laboral, seguridad social, apelaciones y demandas.

RF 15. Se cargará la información proveniente de los ficheros excel y dbf a partir del año 2003 hasta el 2012.

2.4.3 Requisitos no funcionales

Los requisitos no funcionales (RNF) son propiedades o características que la solución debe cumplir y que determinan como esta debe comportarse una vez finalizada. En la presente investigación se identificaron 27 RNF que se encuentran detallados en el artefacto DATEC_SIEJT_Laboral_0113_Especificación de Requisitos de Software, ubicado en el expediente del proyecto. A continuación se muestran algunos de los 27 RNF identificados:

- **Usabilidad**

RNF 3. Agilizar el acceso a los reportes del almacén de datos mediante la distribución de la información por áreas de análisis:

El usuario podrá acceder a la información de manera rápida y de acuerdo al objetivo de su solicitud en el área correspondiente.

- **Confiabilidad**

RNF 8. Asegurar la recuperación ante un fallo:

El sistema debe ser capaz de recuperarse ante un fallo, teniendo en cuenta la complejidad y naturaleza de éste. El tiempo para su correcta recuperación fluctúa entre 10 minutos y 24 horas. Este tiempo comprende la solución al problema, así como su validación y prueba.

- **Eficiencia**

RNF 10. Garantizar un tiempo de respuesta tolerable:

El sistema debe permitir que se muestre la información de las vistas de análisis y los reportes en un tiempo no mayor de 6 segundos.

- **Soporte**

RNF 12. Lograr la homogeneidad de la estructura de los elementos definidos en el almacén:

Las estructuras del almacén de datos deben tener un nombre estándar teniendo en cuenta el tipo de estructura que sea.

- **Restricciones de diseño**

RNF 15. Utilizar el sistema gestor de base de datos definido durante la investigación:

El gestor de base de datos que se utilizará es PostgreSQL 9.1 y como interfaz de administración el PgAdmin 1.14.

- **Requisitos para la documentación de usuarios en línea y ayuda del sistema.**

RNF 19. Confección de un manual de usuario:

El sistema debe estar acompañado de un documento el cual tiene como nombre Manual de Usuario, que explica cómo proceder en cada funcionalidad brindada por el sistema.

- **Interfaz**

RNF 20. Acceso al sistema:

El usuario deberá acceder a la aplicación mediante el protocolo HTTP, usando preferiblemente el navegador web Firefox 4.1 o superior.

- **Interfaces de usuario**

RNF 21. Garantizar una interfaz amigable al usuario:

El sistema debe tener una interfaz amigable, sencilla, con colores suaves y sin cúmulos de imágenes u objetos, teniendo en cuenta que los usuarios finales no son personas instruidas en el campo de la informática.

- **Interfaces de hardware**

RNF 23. Proporcionar características mínimas de hardware:

Para lograr un funcionamiento estable del sistema, los servidores y las estaciones de trabajo, se deben contar con los siguientes requisitos de hardware:

Servidor:

- RAM: al menos 2GB.
- Espacio en el disco: al menos 2GB.
- Procesador: dual-core o superior.

Estación de trabajo:

- RAM: al menos 512 GB.
- Espacio en el disco: al menos 1GB.
- Procesador: dual-core o superior.

- **Interfaz de software**

RNF 24. Instalar en las estaciones de trabajo el software necesario para el correcto funcionamiento del sistema:

Las configuraciones de software de las máquinas clientes deben cumplir los siguientes requisitos de software:

- Utilizar cualquier navegador, pero preferiblemente Firefox 4.1 o superior.
- Java Virtual Machine 1.6 o superior y Schema Workbench 3.2.1 en caso de que un usuario capacitado requiera la construcción de esquemas multidimensionales para el diseño de nuevas vistas de análisis y reportes.

- **Seguridad**

RNF 26. Garantizar el acceso a la información de acuerdo al usuario autenticado en el sistema:

El sistema muestra a cada usuario únicamente la información a la cual puede acceder, garantizando la autenticación como primera acción, en la cual se suministrará un nombre de usuario y una contraseña que deben ser de conocimiento exclusivo de la persona que se autentica, definiéndose previamente los roles y permisos de cada usuario.

2.4 Reglas del negocio

Las reglas del negocio (RN) se establecen con el objetivo de especificar las condiciones que deben tenerse en cuenta durante todo el proceso de desarrollo de la solución. A continuación se muestran algunas de las 18 RN identificadas para cumplir con las necesidades del cliente, especificadas en el

documento DATEC_SIEJT_Laboral_Reglas de negocio y transformación ubicado en el expediente del proyecto.

- Regla de variables

Las demandas pendientes al final (PF) se calculan:

PI (Pendientes al Inicio)+ Radicados=Total/Resolver-Resuelto=PF.

- Regla de almacenamiento

No deben aparecer las medidas con valores nulos, cuando exista un campo vacío debe ponerse cero (0).

- Regla de transformación

El valor de las medidas provenientes de las fuentes de disponibilidad laboral, apelaciones y seguridad social que tengan como tipo de dato varchar se cambiará por el tipo de dato integer.

- Regla de visualización

Los porcientos se visualizarán con solo un valor después de la coma.

2.5 Casos de uso del sistema

Los casos de uso del sistema (CUS) se definieron luego de agrupar los requisitos de información y funcionales, identificando además los actores que se relacionan con cada uno.

2.5.1 Actores del sistema

A continuación se describen las responsabilidades de los actores del sistema.

Tabla 2. Actores del sistema

Actor	Descripción
Especialista	El especialista se encarga de analizar y consultar la información de los diferentes indicadores.
Administrador	El administrador se encarga de gestionar los usuarios, roles, permisos, las vistas de análisis y los reportes.
Administrador ETL	El administrador de ETL se encarga de realizar los procesos de extracción, transformación y carga.

2.5.2 Diagrama de casos de uso del sistema

Los diagramas de casos de uso reflejan las interacciones entre los usuarios y el sistema. En el diseño del diagrama de casos de uso del sistema del MD Laboral se tuvieron en cuenta tres actores y 12 CUS, que

en conjunto conforman el diagrama. Además se utilizó el patrón de casos de uso CRUD completo presente en los casos de uso (CU) Gestionar_usuarios, Gestionar_roles y Gestionar_reportes.

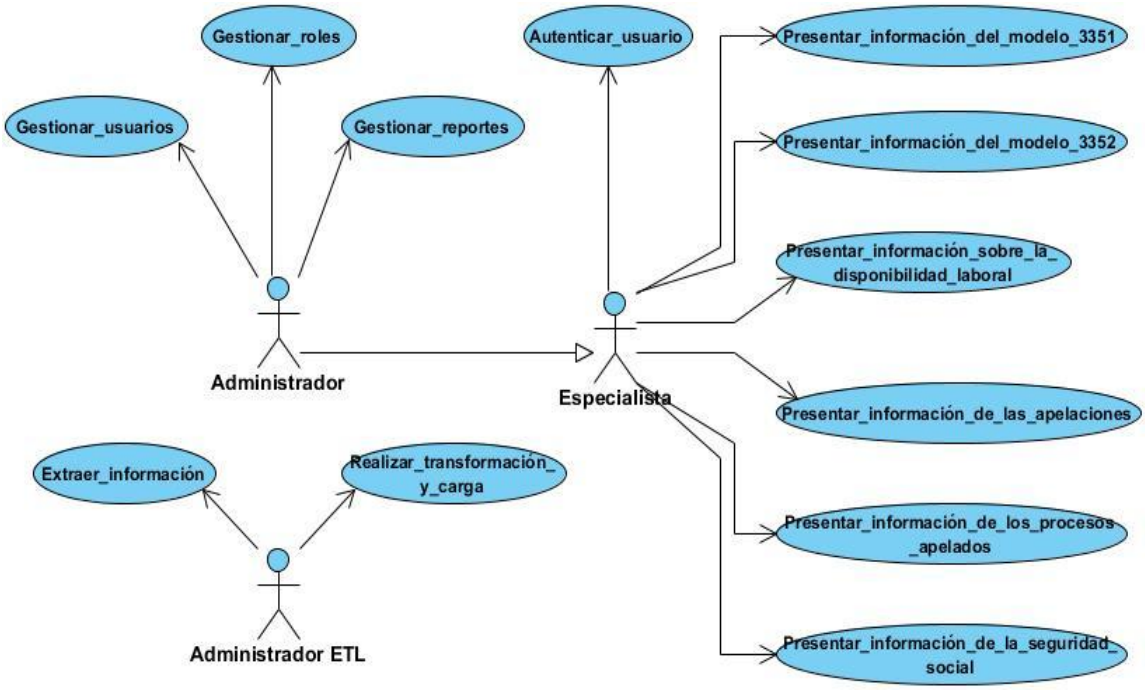


Figura 2. Diagrama de casos de uso

2.5.3 Especificaciones de casos de uso

Para consultar las descripciones de todos los casos de uso remitirse al artefacto DATEC_SIEJT_Laboral_0114_Especificación de Casos de Uso del Sistema, ubicado en el expediente del proyecto.

Tabla 3. Descripción del CU Extraer información

Objetivo	Extraer información.
Actores	Administrador ETL: (Inicia) Extraer información
Resumen	El CU inicia cuando el Administrador ETL desea realizar la extracción de los datos correspondientes a las fuentes de información. Se extraen los datos de la fuente. El CU finaliza una vez que los datos seleccionados por el Administrador ETL son extraídos.
Complejidad	Media
Prioridad	Crítica
Precondiciones	El Administrador ETL tiene que estar autenticado. Disponibilidad de las fuentes.
Postcondiciones	Los datos seleccionados de las fuentes de información quedan extraídos y disponibles para transformar.

Flujo de eventos		
Flujo básico Extraer datos		
	Actor	**Sistema**
1.	Ejecuta la transformación.	
2.		Realiza la conexión a la fuente de información correspondiente.
3.		Chequea la fecha de los datos.
4.		Verifica el control de las extracciones.
5.		Si no se extrajeron esos datos, procede a realizar la extracción. Finaliza el caso de uso.
Flujos alternos		
2ª. No responde a la solicitud de conexión.		
	Actor	**Sistema**
		Notifica el error al Administrador de ETL a través de un mensaje. Vuelve al paso 1 del flujo normal.
5ª. Se extrajeron los datos.		
		Aborta la ejecución del proceso. Finaliza el caso de uso.
Relaciones	**CU Incluidos**	No aplica.
	CU Extendidos	No aplica.
Requisitos no funcionales	Sección 3.2 “Requisitos no funcionales” del documento DATEC_SIEJT_Laboral_0113_Especificacion de Requisitos de Software	

Tabla 4. Descripción del CU Presentar información sobre la disponibilidad laboral

Objetivo	Mostrar información sobre los procesos de disponibilidad laboral.
Actores	Especialista: (Inicia) Presentar información sobre la disponibilidad laboral.
Resumen	El CU inicia cuando el actor desea consultar la información referente a los procesos de disponibilidad laboral. El CU finaliza una vez que los datos se muestran.
Complejidad	Media
Prioridad	Crítica
Precondiciones	Mercado de datos poblado.
Postcondiciones	Los reportes correspondientes fueron consultados por el Especialista.
Flujo de eventos	
Flujo básico Presentar información sobre la disponibilidad laboral.	
Actor	**Sistema**
1. Se autentica en el sistema.	

2.		Verifica autentificación. En caso de existir problemas con la autentificación remitirse al flujo alterno 1.
3.	Selecciona el A.A.G[6] SIEJT[7]	
4.		Muestra el A.A Laboral
5.	Selecciona el A.A[8] Laboral	
6.		Muestra las A.A: -DPA (1976-2010) -DPA (2011-Actualidad)
7.	Selecciona el A.A DPA (2011-Actualidad)	
8.		Muestra los LT[9] Materia laboral e Instrucción 203.
9.	Selecciona el LT Instrucción 203 y luego el LT 00-Disponibilidad laboral.	
10.		Muestra las vistas de análisis y los reportes existentes en el LT 00- Disponibilidad laboral.
11.	Selecciona la vista de análisis o reporte que desea consultar.	
12.		Muestra los datos de la vista de análisis o el reporte seleccionado. Finaliza el CU.
Flujos alternos		
1 Introduce los datos incorrectamente		
	Actor	**Sistema**
		Muestra mensaje ¨Los datos son incorrectos¨. Vuelve al paso 1 del flujo normal.
Relaciones	**CU Incluidos**	No aplica
	CU Extendidos	No aplica
Requisitos no funcionales	Sección 3.2 “Requisitos no funcionales” del documento DATEC_SIEJT_Laboral_0113_Especificación de Requisitos de Software	

[6]A.A.G: Área de Análisis General.

[7]SIEJT: Sistema de Información Estadística Judicial para los Tribunales Populares.

[8]A.A: Área de Análisis.

[9]LT: Libro de Trabajo.

2.6 Arquitectura

La arquitectura del MD Laboral está compuesta por la fuente de datos y los subsistemas de integración, almacenamiento y visualización, los cuales englobarán los aspectos comunes de la arquitectura de un AD. Las fuentes de datos contendrán la información que será utilizada durante el proceso de ETL, con el propósito de realizar su limpieza y transformación atendiendo a las reglas de transformación definidas durante la etapa de diseño. Culminado este proceso los datos transformados son almacenados en esquemas definidos en el subsistema de almacenamiento. Mediante el subsistema de visualización se consulta la información almacenada en la BD correspondiente al MD Laboral, con el objetivo de mostrar la información a los usuarios finales mediante vistas de análisis y reportes.

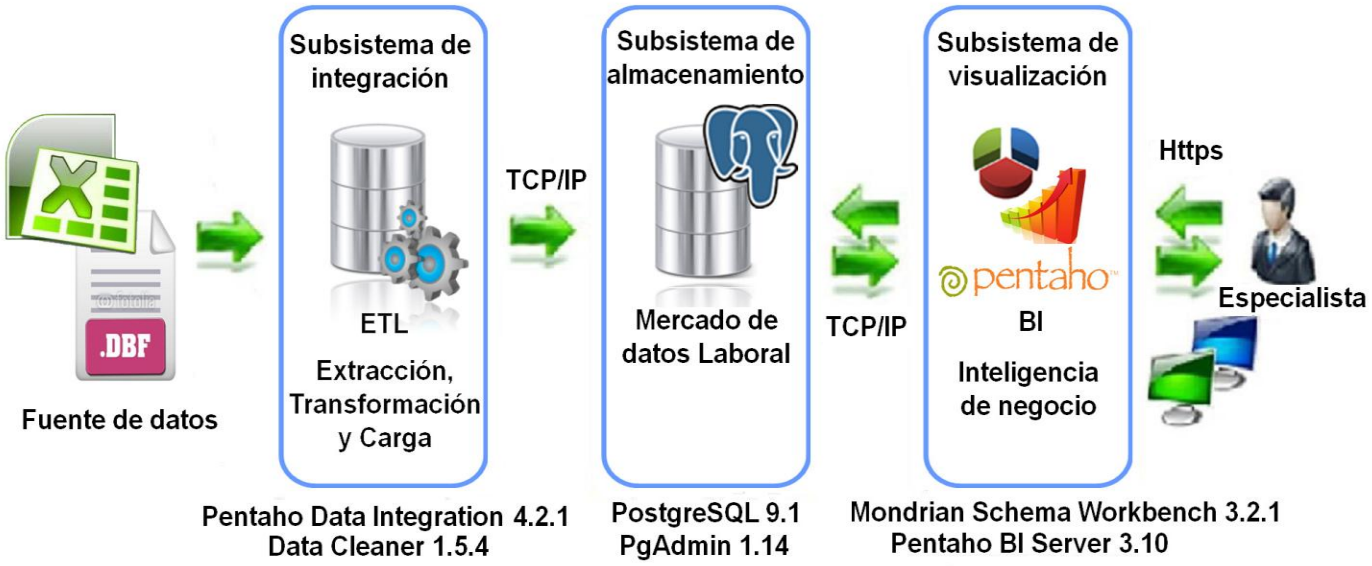


Figura 3. Arquitectura del MD Laboral

- Fuente de datos: incluye los ficheros excel y dbf que contienen la información referente a los procesos de demandas, disponibilidad laboral, apelaciones y seguridad social.
- Subsistema de integración: donde se realizará la extracción, transformación y carga de la información correspondiente al MD Laboral.
- Subsistema de almacenamiento: donde se almacenará en diferentes tablas la información correspondiente al MD Laboral.
- Subsistema de visualización: muestra la información almacenada en el MD, con el objetivo de presentarla al usuario final mediante vistas de análisis y reportes.

2.7 Diseño del subsistema de almacenamiento

Luego de haberse identificado los requisitos del negocio, es importante realizar un análisis de ellos, con el objetivo de tener una mejor comprensión antes de iniciar el diseño de la solución.

2.7.1 Estándares de codificación

Con el objetivo de lograr un mejor entendimiento entre las partes implicadas en el desarrollo de la solución, se definieron estándares de codificación para dejar definida la forma en que se deben hacer las codificaciones y que sirvan como punto de referencia para los desarrolladores del proyecto.

Tabla 5. Estándares de codificación

Tipo de objeto	Función	Nomenclatura
Esquema	Dimensiones compartidas	dimensiones
	Esquemas de datos	mart_[nombre del esquema]
Tabla	Dimensiones	dim_[nombre de la dimensión]
	Hechos	hech_[nombre del hecho]
Constraints	Llave primaria	[nombre de la tabla]_id
	Llave foránea	[nombre de la tabla]_id

2.7.2 Dimensiones

Las dimensiones buscan determinar un contexto para el análisis de los hechos. Se trata de grupos homogéneos de elementos, en muchas ocasiones, jerarquizados. Se utilizan para describir la información almacenada en la tabla de hechos (31). Las dimensiones reflejan características de un hecho que permiten su posterior análisis en el proceso de toma de decisiones. Las tablas descriptivas de las dimensiones se encuentran en el artefacto DATEC_SIEJT_Laboral_Especificación del modelo de datos, ubicado en el expediente del proyecto. Las dimensiones definidas para este trabajo son las siguientes:

- dim_dpa_municipio
- dim_dpa_provincia
- dim_temporal_mes
- dim_indicadores3351
- dim_indicadores3352
- dim_estado_proceso
- dim_causal
- dim_tipo_resolucion
- dim_estado_seguro_social

- dim_tipo_apelacion
- dim_tribunal_provincial
- dim_ tribunal_municipal

2.7.3 Hechos y medidas

Los hechos proporcionan una información cuantitativa sobre las características del negocio que se quieren analizar. Su finalidad es proporcionar información necesaria para la gestión, facilitando el conocimiento del negocio o proceso a modelar (31). Las tablas de hechos contienen las medidas y las claves subrogadas de aquellas dimensiones que definen su nivel de detalle. En el desarrollo del MD Laboral se definieron los siguientes hechos:

- hech_modelo3351
- hech_modelo3352
- hech_disponibilidad_laboral
- hech_procesos_apelados
- hech_apelaciones
- hech_seguridad_social

Las medidas son valores de datos numéricos que serán analizados por los usuarios, son las variables de salida en el diseño. Las medidas describen un proceso del mundo real que será objeto de un análisis (32). Las medidas pueden ser:

- aditivas: cuando pueden ser agregadas a través de cualquier dimensión y no pierde sentido su valor (33).
- semiaditivas: cuando pueden ser agregadas a través de algunas dimensiones y de otras no (33).
- no aditivas: cuando no pueden ser agregadas a través de ninguna dimensión (33).

Se identificaron las medidas aditivas:

- hech_modelo3351: cantidad de demandantes, cantidad de demandas radicadas, cantidad de medidas aplicadas por el O.J.L.B a causales radicadas por indisciplinas, cantidad de medidas aplicadas por la administración al trabajador y cantidad de decisiones del O.J.L.B a causales radicadas por derechos laborales.
- hech_modelo3352: cantidad de expedientes por demandantes, cantidad de demandas resueltas, cantidad de medidas aplicadas por indisciplina, cantidad de declaraciones de demanda, cantidad de demandas tramitadas en días hábiles y cantidad de expedientes revisados y apelados.
- hech_disponibilidad_laboral: cantidad de procesos de disponibilidad laboral.

- hech_procesos_apelados: cantidad de procesos apelados y cantidad de recursos apelados.
- hech_apelaciones: cantidad de recursos de apelación
- hech_seguridad_social: cantidad de pensiones.

Como medidas no aditivas se determinaron:

- hech_modelo3352: el porciento de expedientes por demandantes, el porciento de las demandas resueltas, el porciento de las medidas aplicadas por indisciplina, el porciento de las medidas aplicadas por indisciplina del total de demandas resueltas por sentencia y el porciento de declaraciones de demanda.
- hech_disponibilidad_laboral: el porciento de procesos de disponibilidad laboral resueltos.
- hech_apelaciones: el porciento de recursos de apelación resueltos por autos, el porciento de recursos de apelación resueltos por sentencia y el porciento de recursos de apelación resueltos.
- hech_seguridad_social: el porciento de las pensiones.

Una jerarquía representa una relación lógica entre dos o más atributos dentro de una misma dimensión (34). Las jerarquías poseen las siguientes características:

- Pueden existir varias en una misma dimensión.
- Están compuestas por niveles que describen el orden de una dimensión desde el nivel más resumido hasta el más detallado.

La principal ventaja de manejar jerarquías, reside en poder analizar los datos desde su nivel más general al más detallado y viceversa, al desplazarse por los diferentes niveles.

2.7.4 Matriz bus

La matriz bus representa la relación que existe entre los hechos y las dimensiones del MD Laboral. Las columnas de la matriz representan los hechos, mientras que las filas identifican las dimensiones definidas para la solución. Las celdas con una X indican que el hecho y la dimensión cuya columna y fila se interceptan en dicha celda, guardan relación entre sí. Mediante la matriz bus se identifican los hechos que comparten las mismas dimensiones, permitiendo apreciar la existencia o no de solapamiento de los hechos. A continuación se presenta dicha matriz:

Leyenda:

H1: hech_modelo3351

H2: hech_demandas3352

H3: hech_disponibilidad_laboral

H4: hech_procesos_apelados

H5: hech_apelaciones

H6: hech_seguridad_social

Tabla 6. Matriz bus

Dimensiones	Hechos					
	H1	**H2**	**H3**	**H4**	**H5**	**H6**
dim_estado_proceso			x			
dim_tipo_resolucion					x	
dim_tipo_apelacion					x	
dim_causal			x			x
dim_estado_seguro_social						x
dim_dpa_municipio	x	x				
dim_dpa_provincia			x	x	x	x
dim_temporal_mes	x	x	x	x	x	x
dim_indicadores3351	x					
dim_indicadores3352		x				
dim_tribunal_provincial			x	x	x	x
dim_tribunal_municipal	x	x				

2.7.5 Modelo de datos

El modelado de los datos permite identificar las relaciones entre los hechos y las dimensiones, así como las medidas pertenecientes a cada hecho. Entre los modelados de datos se encuentra el ER, que es utilizado para crear un único modelo de todos los procesos de la organización. Este enfoque resulta ser efectivo para crear sistemas eficientes basados en el OLTP (32). Además existe el modelado dimensional, que es una técnica de diseño lógico que busca presentar la información en un marco estándar e intuitivo que permita un acceso de alto rendimiento (32). Existen principalmente tres esquemas para el modelado dimensional: estrella, copo de nieve y constelación de hechos. En el esquema estrella las dimensiones no se normalizan, minimizándose el número de uniones y, por consiguiente, incrementando el rendimiento de las consultas. Lo que distingue al esquema copo de nieve del esquema previamente mencionado, es que las tablas de dimensiones en este modelo representan relaciones normalizadas y forman parte de un modelo relacional de BD, por su parte el esquema constelación de hechos es una generalización de los

esquemas en estrella y copo de nieve, que se obtiene con la inclusión de distintas tablas de hechos que comparten algunas de las dimensiones presentes en el modelo de datos.

En el diseño del modelo de datos perteneciente al MD Laboral se empleó la topología constelación de hechos, permitiendo que las tablas de dimensiones puedan estar compartidas entre más de una tabla de hechos. A continuación se muestra una porción del modelo de datos presente en los anexos para presentar los hech_seguridad_social y hech_disponibilidad_laboral, reflejando las dimensiones con las que interactúan.

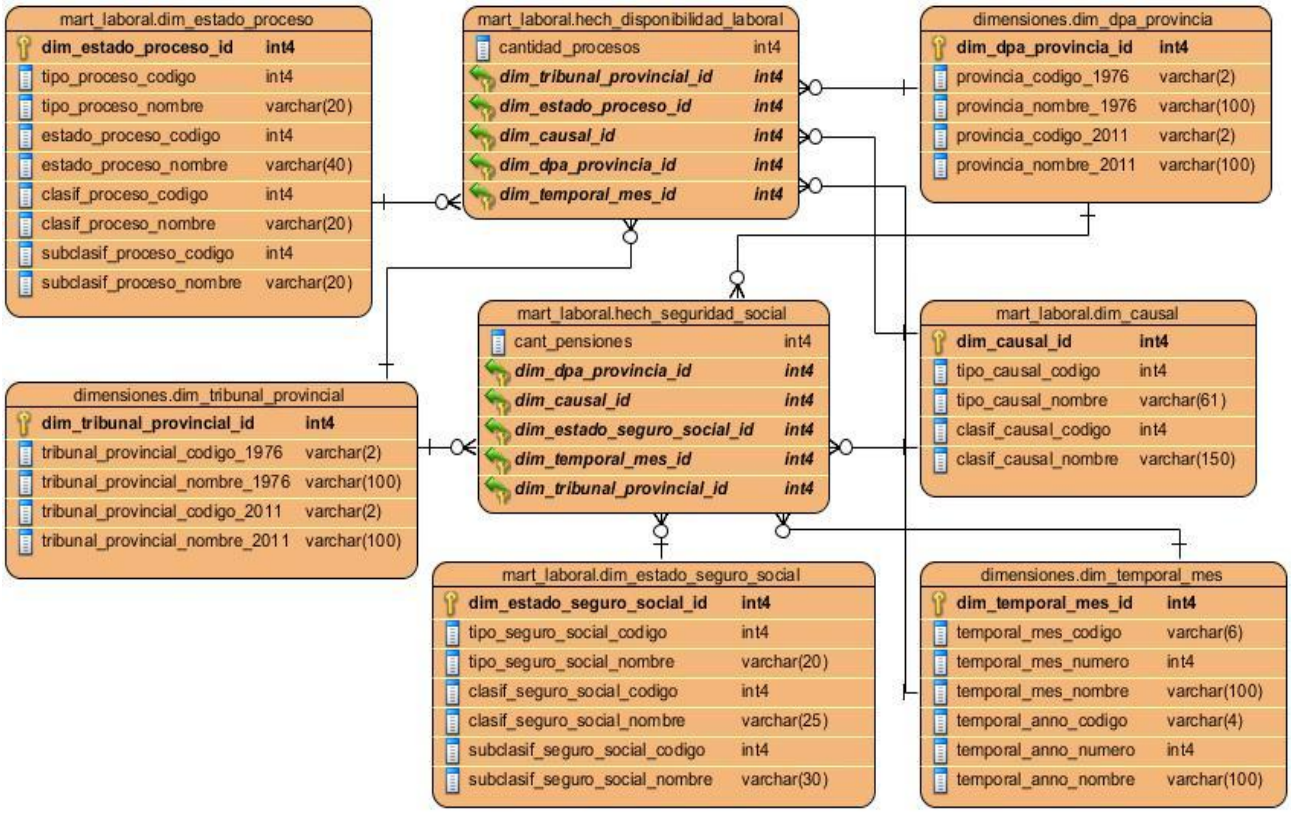


Figura 4. Modelo dimensional para el hecho seguridad social y el hecho disponibilidad laboral

2.8 Diseño del subsistema de integración

El diseño del subsistema de integración tiene el propósito de brindar información de cómo van a estar implementados los procesos de ETL en el mercado de datos a desarrollar.

Existen diferentes estrategias de integración de datos que posibilitan realizar la integración de la información procedente de diversas fuentes. Dentro de estas se encuentran:

- Replicación de datos: es una técnica de integración que se basa en la creación y mantenimiento de múltiples copias de una misma base de datos. En la mayoría de las implementaciones de replicación, un servidor mantiene la copia primaria de la base de datos y servidores adicionales mantienen las copias esclavas de la misma (35).
- ETL: extrae la información de un sistema fuente, transforma esos datos para satisfacer los requisitos del negocio y carga el resultado en el sistema destino. Tanto la fuente como el destino son generalmente base de datos y archivos. Esta técnica se encarga de la integración de datos, no de aplicaciones, y obtiene los datos directamente de la base de datos (35).
- Integración de Información Empresarial: es un mecanismo de transformación y acceso a datos transparentes y optimizados para suministrar una única interfaz a lo largo de los datos de las organizaciones. Este tipo de solución consiste en crear un intermediario que contenga los directorios de la base de datos y que a su vez sirva de canal de consulta y representación de la información recuperada. Esta estrategia no es factible para la integración de aplicaciones (35).
- Integración de Aplicaciones Empresariales: es el proceso de integrar múltiples aplicaciones que utilizan tecnología incompatible y que son gestionadas de forma independiente, permitiendo que se comuniquen e intercambien transacciones de negocio, mensajes, y datos entre sí. Las características más importantes de esta tecnología es que se utiliza para la integración de Aplicaciones a Aplicaciones y proporciona un enfoque de integración orientado a proceso basado en mensajes XML (35).

Para el desarrollo de la solución se utilizará ETL como estrategia de integración, siendo la más apropiada para integrar las fuentes de datos que serán utilizadas para poblar el MD Laboral. Esta estrategia permitirá extraer los datos procedentes de los excel y dbf asociados a los procesos de demandas, apelaciones, disponibilidad laboral y seguridad social, que luego serán transformados teniendo en cuenta las reglas del negocio, para finalmente cargar el resultado en la base de datos destino.

Perfilado de datos

El perfilado de datos se encarga de analizar las fuentes de datos con el objetivo de conocer su estructura, tipos de datos presentes y la calidad de estos, sirviendo de guía para las futuras transformaciones que se realizarán en el proceso de ETL. Permite obtener estadísticas e información sobre los datos, que dan la posibilidad de corregir problemas como valores escritos incorrectamente, duplicados o nulos.

En el análisis de las fuentes de datos que serán empleadas para poblar el MD Laboral se utilizó el DataCleaner para determinar la cantidad máxima de caracteres de tipo varchar que podía admitir cada

campo de la BD. Esta información fue utilizada como base durante el diseño del modelo de datos para establecer el tamaño máximo de los campos de tipo varchar. En la siguiente figura se muestra un ejemplo del resultado arrojado por el perfilado a la fuente de datos correspondiente a Disponibilidad laboral.

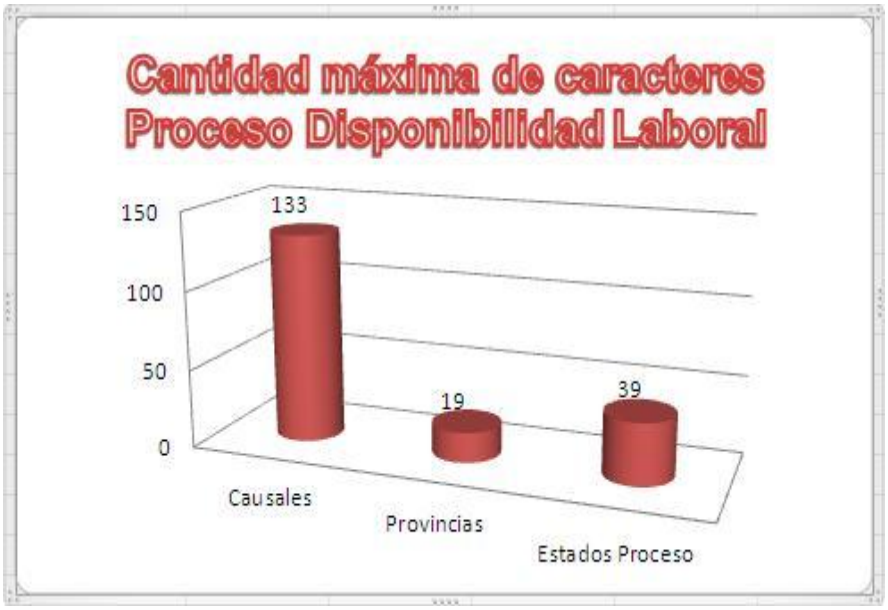


Figura 5. Cantidad máxima de caracteres de la fuente Disponibilidad Laboral

En la figura 6 se evidencia la distribución en porciento de los tipos de datos presentes en las fuentes del MD Laboral, permitiendo identificar como tipo de dato predominante el varchar. Además se detectó como problema predominante en las fuentes de datos correspondientes a los excel, la existencia del tipo de dato varchar en los valores asociados a las medidas, que serán cambiadas a integer en el proceso de ETL.

Figura 6. Tipos de datos de las fuentes del MD Laboral

Diseño general de las transformaciones

El diseño de las transformaciones sirve de guía para llevar a cabo los procesos de extracción, transformación y carga de los datos al MD Laboral. A continuación se muestra uno de los diseños de las transformaciones realizados para los hechos del MD Laboral.

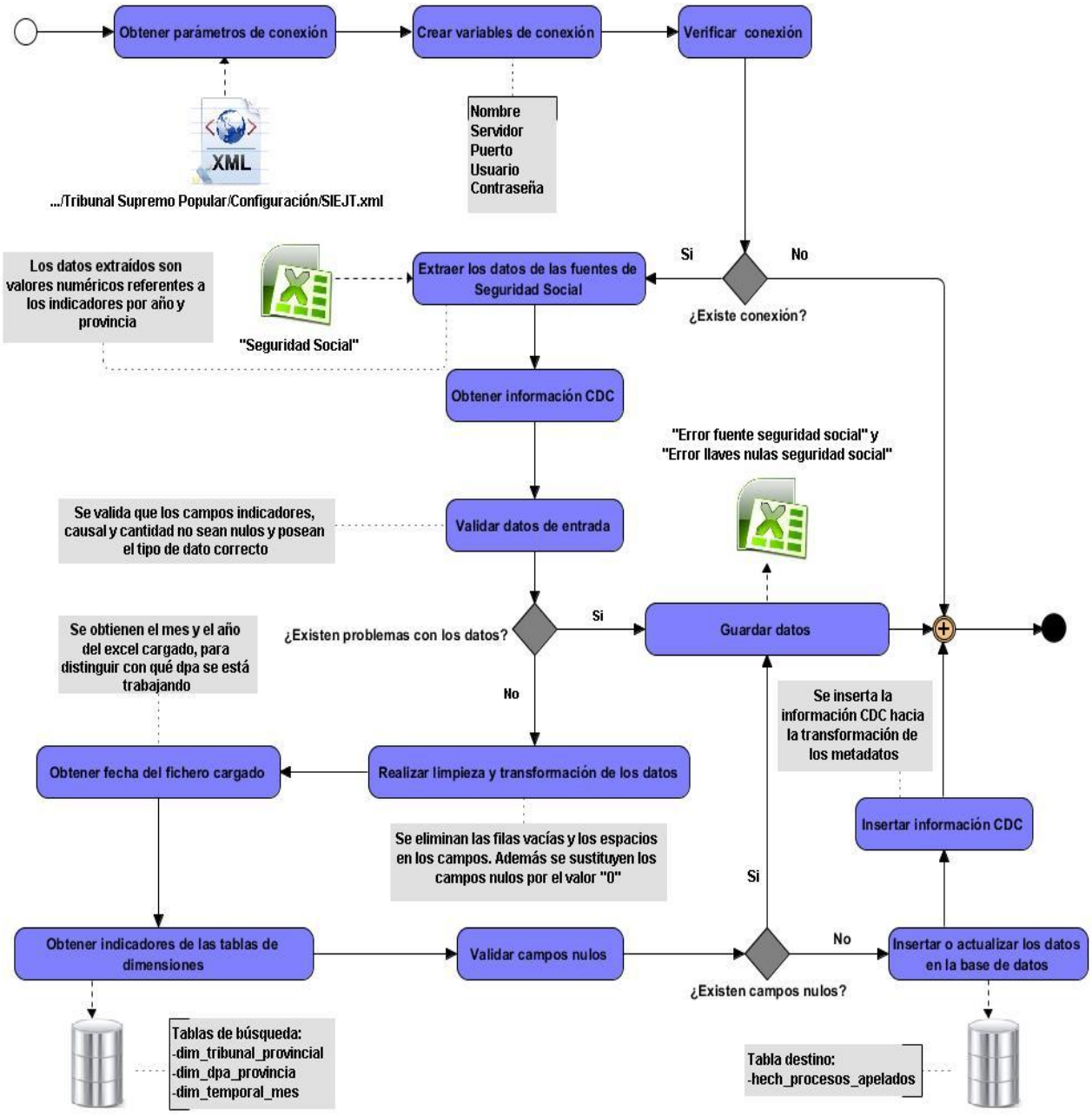


Figura 7. Diseño de las transformaciones para el hecho seguridad social

2.9 Diseño del subsistema de visualización

El diseño del subsistema de visualización se realiza con el objetivo de adquirir una perspectiva de cómo serán mostrados los datos a los usuarios finales.

Arquitectura de la información

La arquitectura de la información permite conocer cómo estarán estructurados los elementos que conforman el subsistema de visualización. Para el MD Laboral se identificó el área de análisis general Sistema de Información Estadística Judicial para Tribunales (SIEJT), el área de análisis Laboral que contiene cinco libros de trabajo y los reportes asociados a cada libro de trabajo, reflejando cada uno de los elementos en la figura 8.

Figura 8. Diseño del mapa de navegación del MD Laboral

Diseño de los cubos OLAP

Los cubos OLAP permiten presentar la información a diferentes niveles de agregación, mediante las operaciones *Drill Down* y *Roll Up*. *Drill Down* permite explorar los hechos hacia los niveles más detallados de la jerarquía de dimensiones, mientras que *Roll Up* explora los hechos iterativamente hacia el nivel más alto de agregación (36).

Entre las ventajas de los cubos OLAP se encuentran (36):

- Tiene acceso a grandes cantidades de información.
- Presentan los datos en diferentes perspectivas.
- Pueden responder con rapidez a consultas de usuarios.

En el diseño de los cubos OLAP se especifican las dimensiones, los niveles jerárquicos y las medidas correspondientes a cada cubo. En la presente investigación se modelarán 12 dimensiones y seis cubos multidimensionales, uno por cada tabla de hecho.

2.10 Políticas de respaldo y recuperación

Las políticas de respaldo y recuperación de datos brindan una guía para garantizar que la información esté almacenada en un lugar seguro en caso de fallos en el sistema. Las políticas de seguridad y respaldo que se utilizarán en el MD Laboral están divididas en dos puntos fundamentalmente.

- Periodicidad de las salvas: las salvas de la información contenida en la BD se realizarán mensualmente, confirmando que exista una copia de la información presente en el servidor.
- Tablas involucradas: las tablas que se involucran en la realización de las salvas son las seis tablas de hechos identificadas para el MD Laboral, con sus 12 dimensiones asociadas.

2.11 Conclusiones del capítulo

En este capítulo se realizó un análisis del negocio permitiendo alcanzar un mayor conocimiento sobre las necesidades de los usuarios y describir las reglas a las que tiene que ajustarse el negocio. Se definieron los requisitos funcionales, no funcionales y los de información que debe cumplir la solución para lograr el desarrollo de un MD que cumpla con las expectativas del cliente. El diseño de diagrama de casos de uso del sistema permitió reflejar las interacciones entre los actores y el sistema. La arquitectura propuesta para la construcción del MD Laboral permitió determinar la estructura, funcionamiento e interacción entre sus partes. La especificación de los hechos y las dimensiones que se utilizaron en la matriz bus posibilitaron evitar el solapamiento de los hechos. El diseño de las transformaciones, los cubos OLAP y la arquitectura de la información sentaron las bases para el comienzo de la implementación de la solución.

Capítulo 3: Implementación y pruebas del mercado de datos

Introducción

En este capítulo se realizará la implementación de los subsistemas que en conjunto forman la solución del problema de la investigación, teniendo en cuenta las necesidades de los usuarios y los requisitos del negocio. De igual manera, se aplicarán los diferentes tipos de pruebas, mediante los casos de prueba y las listas de chequeo, para asegurar que no existan fallos en la implementación del *software*, proporcionándole calidad a la solución.

3.1 Implementación del subsistema de almacenamiento de datos

Estructura de los datos

La herramienta de administración de bases de datos PgAdmin permite organizar la información de las BD en diferentes estructuras que facilitan su manipulación. Entre estas estructuras se encuentran los esquemas y las tablas. Los esquemas representan una forma de organizar la información contenida en una BD. En ellos se agrupan las tablas y los campos pertenecientes a cada tabla, posibilitando una adecuada organización de los datos. Dentro de los esquemas se pueden encontrar funciones, operadores y tipos de datos que facilitarán su implementación.

En el presente trabajo se definieron los siguientes esquemas: el esquema dimensiones, que contendrá la información correspondiente a las tablas de dimensiones que son comunes para el AD central, el esquema mart_laboral, que contendrá las tablas de hechos, dimensiones y una tabla closure pertenecientes específicamente al MD Laboral y el esquema metadatos que contiene las tablas para la captura de los metadatos de los procesos de ETL. A continuación se muestran los esquemas del MD Laboral con las 25 tablas correspondientes a cada uno, de las cuales 12 son de dimensiones, seis de hechos y seis para la captura de los metadatos. La tabla closure es utilizada con el objetivo de establecer una jerarquía que permita la obtención de la información de los procesos de demandas presente en el modelo 3352.

Tabla 7. Tablas ubicadas en los esquemas de la base de datos del MD Laboral

Esquemas	Tablas
dimensiones	dim_dpa_municipio
dimensiones	dim_dpa_provincia
dimensiones	dim_temporal_mes
dimensiones	dim_tribunal_municipal

dimensiones	dim_tribunal_provincial
mart_laboral	closure_dim_indicadores3352
mart_laboral	dim_causal
mart_laboral	dim_estado_proceso
mart_laboral	dim_estado_seguro_social
mart_laboral	dim_indicadores3351
mart_laboral	dim_indicadores3352
mart_laboral	dim_tipo_apelacion
mart_laboral	dim_tipo_resolucion
mart_laboral	hech_apelaciones
mart_laboral	hech_disponibilidad_laboral
mart_laboral	hech_modelo3351
mart_laboral	hech_modelo3352
mart_laboral	hech_procesos_apelados
mart_laboral	hech_seguridad_social
metadatos	md_carga_historica
metadatos	md_cdc
metadatos	md_mercado
metadatos	md_registro_hist_fichero
metadatos	md_temporal
metadatos	md_transformacion

3.2 Implementación del subsistema de integración de datos

El proceso de ETL inicia al extraer los datos que poblarán el MD Laboral, utilizando los ficheros excel y dbf proporcionados por el cliente. Los ficheros contienen la información referente a los procesos de demandas, apelaciones, seguridad social y disponibilidad laboral, que son analizados por los especialistas del área Laboral desde el año 2003 hasta el 2012, con el objetivo de presentar indicadores que reflejen el comportamiento de estos procesos. Culminado el proceso de extracción de los datos se procede a llevar a cabo la transformación de estos, con el objetivo de detectar los posibles errores presentes en las fuentes y corregirlos, asegurando la consistencia de la información, para luego realizar su carga hacia el MD Laboral.

3.2.1 Transformaciones y trabajos

Una vez concluida la extracción de los datos se procede a realizar su transformación, constituyendo el paso más importante en la implementación del proceso de ETL. En el presente trabajo se realizaron 24 transformaciones y 23 trabajos (*job*) con el propósito de realizar una carga de los datos exitosa. A continuación se muestra el trabajo diseñado para realizar la carga del hech_seguridad_social.

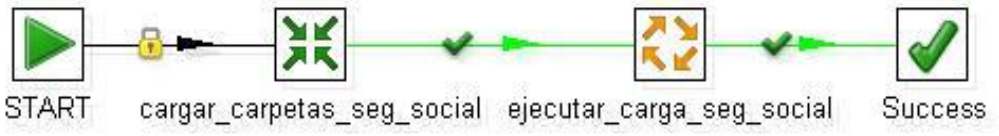


Figura 9. Trabajo para cargar el hecho seguridad social

Con el objetivo de cargar el hecho seguridad social se implementó el trabajo carga_seguridad_social, el cual se inicia empleando el componente STAR, luego utilizando el componente Transformación se realiza la llamada a la transformación cargar_excel_seg_social del cual se obtendrán los directorios, la fecha y las provincias que serán utilizados en la transformación hech_seguridad_social. Seguido de este paso se utiliza el componente Trabajo, mediante el cual se llama a la transformación del hech_seguridad_social. Una vez concluido este paso se utiliza el componente Succes para confirmar que el trabajo fue ejecutado correctamente.

En la siguiente figura se muestra la transformación correspondiente al hech_seguridad_social:

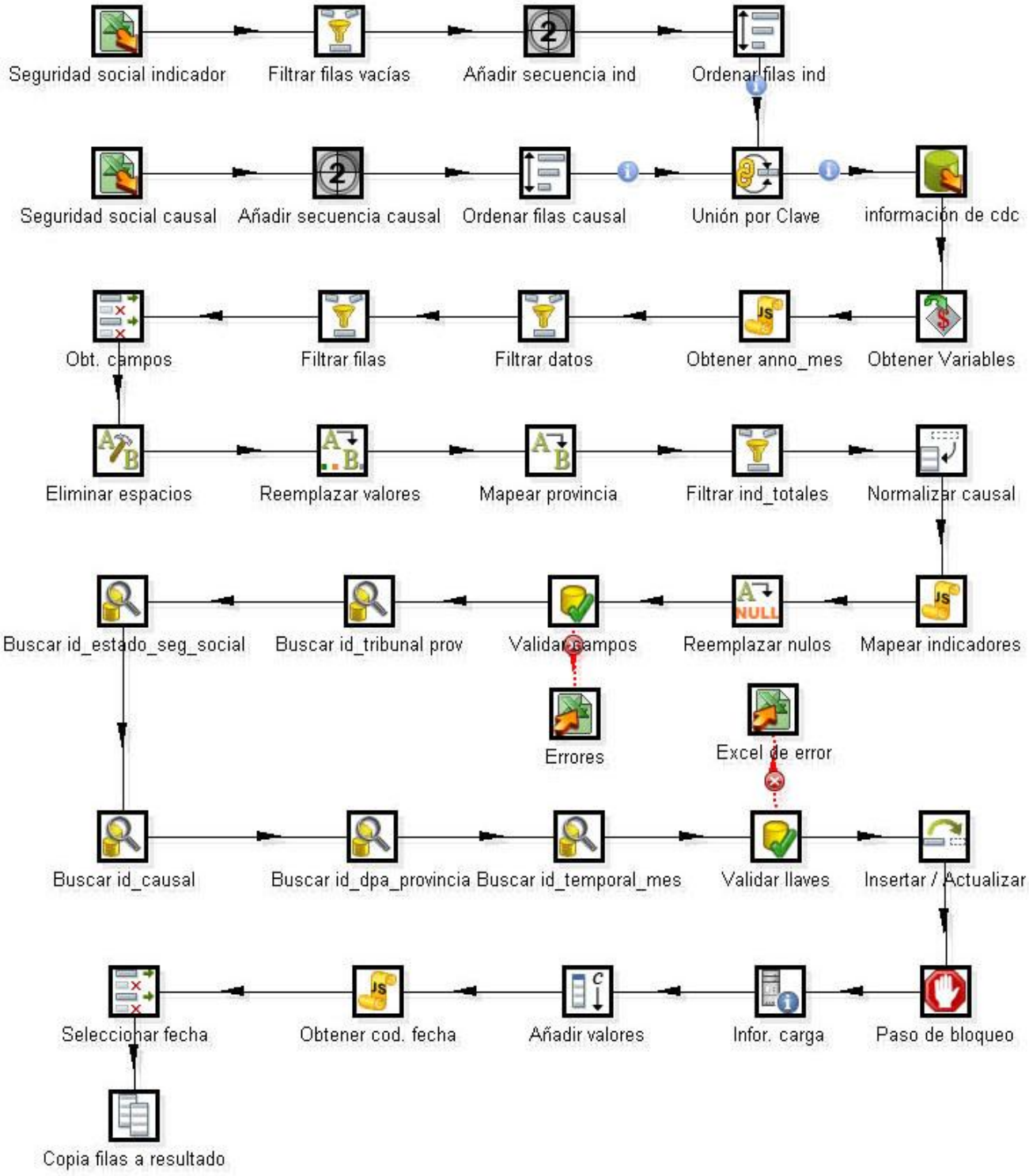


Figura 10. Transformación correspondiente al hecho seguridad social

Primeramente se realizó la extracción de los datos a partir de los ficheros con formato excel de seguridad social, declarando que los campos a extraer son los indicadores. Además se especifica que el nombre de la hoja a cargar es Seguridad Social y que se realizará la carga a partir de la fila seis y la columna dos. Luego utilizando el componente Filtrar filas se eliminan los indicadores que estén vacíos y posteriormente se les asigna un código que servirá para ordenarlos ascendentemente por medio del componente Ordenar filas.

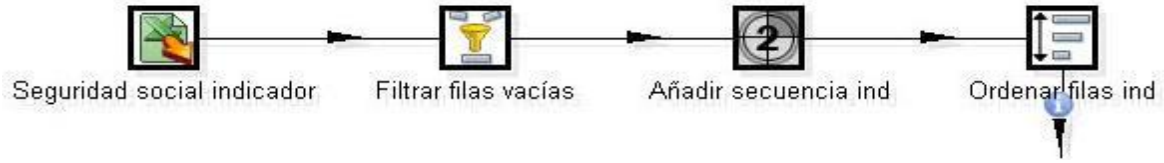


Figura 11. Transformación realizada al hecho seguridad social parte 1

Por otra parte, se realiza la misma operación que en el paso anterior pero definiendo que los campos a extraer son las causales, a partir de la hoja con nombre Seguridad Social y comenzando en la fila siete y la columna tres. Luego se le asigna un código a cada causal y se ordenan ascendentemente, para más tarde unir los dos flujos según el valor de la clave común dada y devolver un único flujo del conjunto. A continuación se obtiene la información que servirá para llevar un control de los cambios de los datos en la fuente.

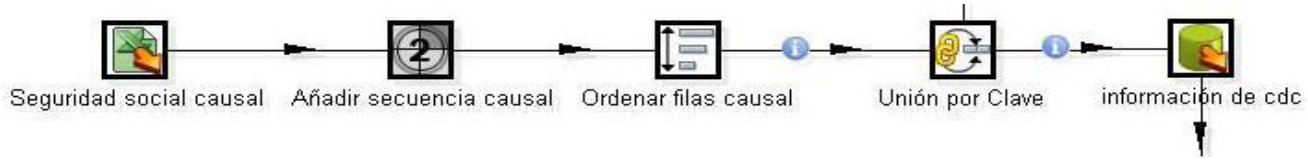


Figura 12. Transformación realizada al hecho seguridad social parte 2

Luego se obtienen las variables de entorno temporal y provincia que se guardarán en los campos tiempo y provincias respectivamente. A partir del campo tiempo se obtiene el mes y el año del excel cargado. Se filtra la información que tenga una fecha más actual que el valor de la última fecha actualizada y se eliminan los campos que no son necesarios. Seguido se establecen los tipos de datos de los campos indicadores y causales.

Figura 13. Transformación realizada al hecho seguridad social parte 3

En pasos posteriores se realiza una limpieza de los datos eliminando los espacios vacíos que presenten los indicadores, se mapean los valores de las provincias para renombrarlas con su valor escrito correctamente y se filtran los campos indicadores que se necesitan cargar. Además se normaliza la información de las causales guardándose los datos de las mismas en el nuevo campo cantidad.

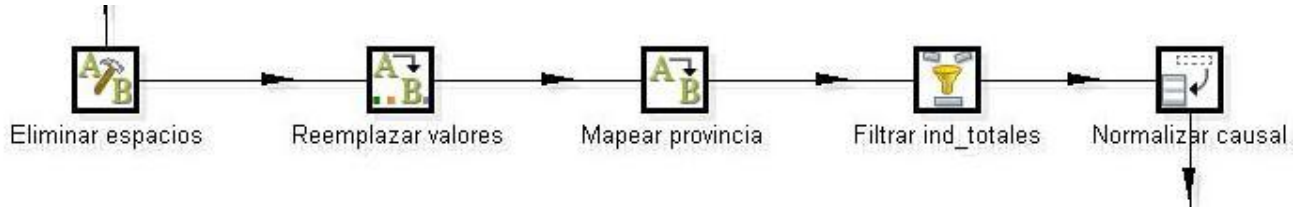


Figura 14. Transformación realizada al hecho seguridad social parte 4

Luego se mapean las causales y los indicadores mediante el componente Valores de Script para lograr que sus nombres comiencen con letra mayúscula. Seguidamente se sustituyen los valores del campo cantidad que vengan nulos por el valor “0” y se valida que los campos causal, indicadores y cantidad no sean nulos y contengan el tipo de dato que les corresponde, especificándose en un excel los errores encontrados. También se comparan los campos provincias e indicadores con sus valores correspondientes en la BD del MD Laboral para obtener los campos dim_tribunal_provincial_id y dim_estado_seguro_social_id.

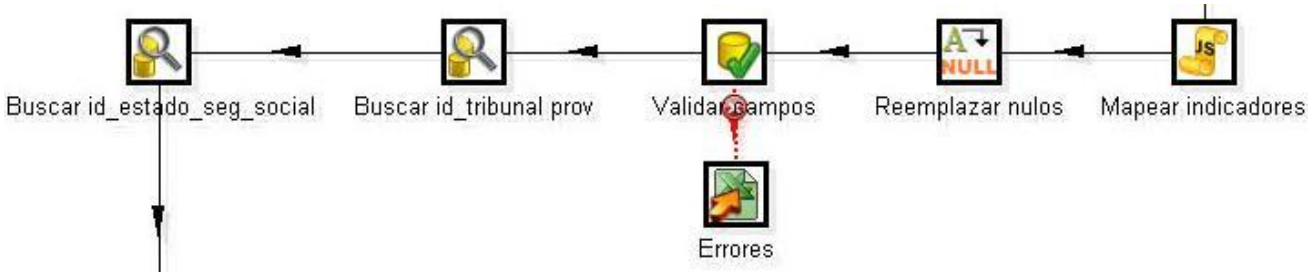


Figura 15. Transformación realizada al hecho seguridad social parte 5

Posteriormente se comparan los campos provincias, causal, mes y año con sus valores correspondientes en la BD del MD Laboral, para obtener los campos dim_causal_id, dim_dpa_provincia_id y dim_temporal_mes_id. Una vez obtenidos los identificadores se valida que ninguno pueda ser nulo y de encontrarse el caso se guardará el error en un excel, especificando los datos del problema. Al concluir el flujo se insertan los datos de seguridad social transformados e integrados al MD Laboral.

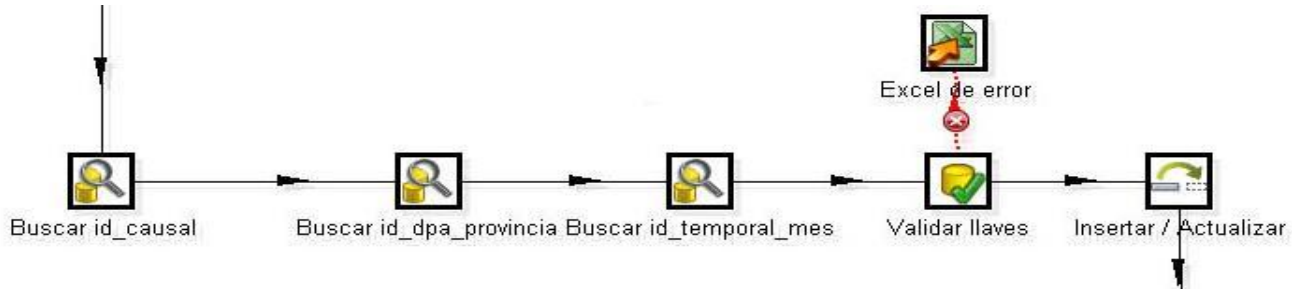


Figura 16. Transformación realizada al hecho seguridad social parte 6

Finalmente por medio del componente Paso de bloqueo se espera hasta que todas las filas de entrada hayan sido procesadas para seguir con la transformación. Se obtiene la información del sistema, se le añade un valor constante a los campos nombre_datamart, nombre_fuente y frecuencia_carga y se obtiene la fecha del sistema, para luego copiar las filas hacia la transformación de los metadatos.

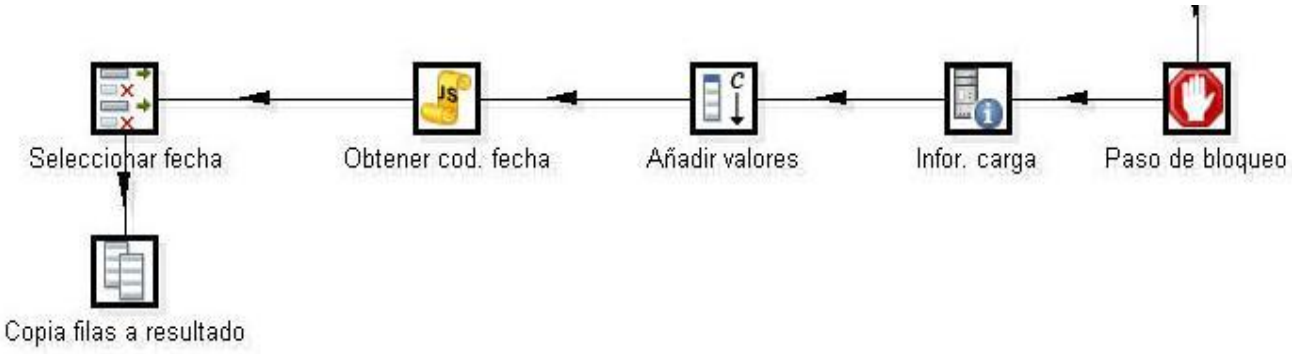


Figura 17. Transformación realizada al hecho seguridad social parte 7

Una vez terminado el proceso de ETL, las tablas que conforman la BD, contendrán los datos que fueron transformados y limpiados para poblar el MD Laboral.

3.2.2 Gestión del cambio de las dimensiones

Las dimensiones lentamente cambiantes o en inglés *Slowly Changing Dimensions* (SCD), son dimensiones en las cuales sus datos tienden a modificarse a través del tiempo, ya sea de forma ocasional o constante, o implique a un solo registro o la tabla completa (37). De acuerdo al tipo de cambio que sufren los datos se selecciona que tipo de SCD se utilizará para su manejo. A continuación se detallan los tipos de SCD:

SCD Tipo 0

Este es un enfoque pasivo, es decir no se hace nada al respecto. Los valores permanecen como estaba la dimensión cuando los registros fueron creados.

SCD Tipo 1: Sobrescribir

En este caso cuando un registro presente un cambio en alguno de los valores de sus campos, se debe proceder simplemente a actualizar el dato en cuestión, sobrescribiendo el antiguo (37). Para ejemplificar este caso, se tomará como referencia la siguiente tabla:

id_Producto	Rubro	Tipo	Producto
1	Rubro 1	Tipo 1	Producto 1

Ahora, se supondrá que este producto ha cambiado de Rubro, y ahora ha pasado a ser "Rubro 2", entonces se obtendrá lo siguiente:

id_Producto	Rubro	Tipo	Producto
1	Rubro 2	Tipo 1	Producto 1

SCD Tipo 2: Añadir fila

Esta estrategia requiere que se agreguen algunas columnas adicionales a la tabla de dimensión, para que almacenen el historial de cambios (37). Las columnas que suelen agregarse son:

- FechaInicio: fecha desde que entró en vigencia el registro actual. Por defecto suele utilizarse una fecha muy antigua, ejemplo: "01/01/1000".
- FechaFin: fecha en la cual el registro actual dejó de estar en vigencia. Por defecto suele utilizarse una fecha muy futurista, ejemplo: "01/01/9999".
- Versión: número secuencial que se incrementa cada nuevo cambio. Por defecto suele comenzar en "1".
- Versión actual: especifica si el campo actual es el vigente. Este valor puede ser en caso de ser verdadero: "true" o "1"; y en caso de ser falso: "false" o "0".

Una vez que ocurra algún cambio, se añadirá una nueva fila y se deberán completar los datos referidos al historial de cambios. Para ejemplificar este caso, se tomará como referencia la siguiente tabla:

id_Producto	Rubro	Tipo	Producto
1	Rubro 1	Tipo 1	Producto 1

A continuación se añadirán las columnas que almacenarán el historial:

id_Producto	Rubro	Tipo	Producto	FechaInicio	FechaFin	Versión	VersiónActual
1	Rubro 1	Tipo 1	Producto	01/01/1000	01/01/9999	1	True

			1				

Ahora, se supondrá que este producto ha cambiado de Rubro, y ahora ha pasado a ser "Rubro 2", entonces se obtendrá lo siguiente:

id_Producto	Rubro	Tipo	Producto	FechaInicio	FechaFin	Versión	VersiónActual
1	Rubro 1	Tipo 1	Producto 1	01/01/1000	06/11/2009	1	False
2	Rubro 2	Tipo 1	Producto 1	07/11/2009	01/01/9999	2	True

Como puede observarse, se lleva a cabo el siguiente proceso:

- Se añade una nueva fila con su correspondiente clave subrogada ("id_Producto").
- Se registra la modificación ("Rubro").
- Se actualizan los valores de "FechaInicio" y "FechaFin", tanto de la fila nueva, como la antigua (la que presentó el cambio).
- Se incrementa en uno el valor del campo "Versión" que posee la fila antigua.
- Se actualizan los valores de "VersionActual", tanto de la fila nueva como la antigua, dejando a la fila nueva como el registro vigente (True).

SCD Tipo 3: Añadir columna

Esta estrategia requiere que se agregue a la tabla de dimensión una columna adicional por cada columna cuyos valores se desea mantener un historial de cambios (37).

Se tomará como ejemplo la siguiente tabla:

id_Producto	Rubro	Tipo	Producto
1	Rubro 1	Tipo 1	Producto 1

A continuación se añadirá una columna para mantener el historial de cambios sobre los datos de la columna "Rubro":

id_Producto	Rubro	RubroAnterior	Tipo	Producto
1	Rubro 1	-	Tipo 1	Producto 1

Ahora, se supondrá que este producto ha cambiado de Rubro, y ahora ha pasado a ser "Rubro 2", entonces se obtendrá lo siguiente:

id_Producto	Rubro	RubroAnterior	Tipo	Producto
1	Rubro 2	Rubro 1	Tipo 1	Producto 1

Como puede observarse, se lleva a cabo el siguiente proceso:

- En la columna "RubroAnterior" se coloca el valor antiguo.
- En la columna "Rubro" se coloca el nuevo valor vigente.

SCD Tipo 4: Tabla de Historia separada

Esta técnica se utiliza en combinación con alguna otra y su función básica es almacenar en una tabla adicional los detalles de cambios históricos realizados en una tabla de dimensión (37). Esta tabla histórica indicará por ejemplo que tipo de operación se ha realizado (*Insert, Update, Delete*), sobre qué campo y en qué fecha. El objetivo de mantener esta tabla es el mantener un control de todos los cambios, para luego analizarlos y poder tomar decisiones acerca de cuál técnica SCD podría aplicarse mejor. Por ejemplo, la siguiente tabla histórica registra los cambios de la tabla de dimensión "Productos", la cual supondremos emplea el SCD Tipo 2:

id_Producto	Rubro_Cambio	Tipo_Cambio	Producto_Cambio	FechaDeCambio
1	Insert	-	-	05/06/2000
2	Insert	Insert	-	25/10/2002
3	-	Insert	-	17/01/2005
4	-	-	Insert	18/02/2009

Tomando como ejemplo el primer registro de esta tabla, la información allí guardada indica lo siguiente:

- El día "05/06/2000", el registro de la tabla de dimensión "Productos" con "id_Producto" igual a "1" sufrió un cambio de "Rubro", por lo cual se debió insertar ("Insert") una nueva fila con los valores vigentes.

SCD Tipo 6: Híbrido

Esta técnica combina las SCD Tipo 1, 2 y 3.

Se denomina SCD Tipo "6", simplemente porque: 6 = 1 + 2 +3.

Debido a que la información que se maneja en el área Laboral toma valores distintos de DPA para las fechas anteriores y posteriores al año 2011, se decidió utilizar para la gestión del cambio en las dimensiones dpa_provincia, dpa_municipio, tribunal_provincial y tribunal_municipal del MD Laboral, la estrategia de SCD Tipo 3 permitiendo darle seguimiento a los valores de sus columnas. Cabe destacar

que para el resto de las dimensiones pertenecientes al MD Laboral no fue necesario aplicar ningún tipo de SCD, ya que sus valores no tienden a variar con el paso del tiempo.

3.2.3 Gestión de los metadatos del proceso de integración

Los metadatos son "datos sobre los datos" que permiten gestionar, controlar, entender y preservar otra información. Los metadatos describen el contenido, la calidad y otras características de los datos, constituyendo un mecanismo para caracterizarlos (38). Existen diversos tipos de metadatos que son clasificados teniendo en cuenta diferentes parámetros. A continuación se muestran algunas de las clasificaciones de los metadatos:

- Metadatos administrativos: son utilizados para el manejo y administración de los recursos de información. Incluyen información sobre cuándo y cómo fue creado el recurso, quién es el responsable del acceso o de la actualización del contenido y también se incluye información técnica, como la versión de *software* o el *hardware* necesario para ejecutar dicho recurso.
- Metadatos descriptivos: tienen como propósito descubrir, identificar y seleccionar recursos de información.
- Metadatos técnicos: están relacionados con la función de un sistema o el modo en que interrelacionan sus componentes. Son utilizados para informar sobre los requisitos técnicos de *hardware* o *software.*
- Metadatos de proceso: permiten obtener información de los procesos en que se ejecutan.
- Metadatos de negocio: posibilita obtener los datos y la información referente a los aspectos del negocio, como son los datos provenientes de la fuente. Incluyen descripciones de datos que no están relacionadas a implementaciones de software, por ejemplo, el nombre del negocio y las reglas de negocio en relación a otros datos.

En la investigación se utilizaron los metadatos de proceso para obtener la información correspondiente a las transformaciones y los trabajos realizados para poblar el MD Laboral. Se definieron seis tablas de metadatos que almacenan la información correspondiente al nombre de la fuente, nombre del MD, nombre del hecho, líneas leídas, escritas, actualizadas, introducidas, salientes y rechazadas, así como, otros aspectos asociados a las transformaciones que permiten adquirir conocimiento de su estado.

3.3 Implementación del subsistema de visualización de datos

3.3.1 Implementación de los cubos OLAP

Una vez terminado el proceso de ETL se realizó la implementación de los cubos OLAP correspondientes al MD Laboral, mediante el uso de la herramienta Pentaho Schema Workbench. A través de los cubos

OLAP se accede a la información presente en las tablas de hechos con sus dimensiones, permitiendo el análisis y procesamiento rápido de los datos. En la presente investigación se implementaron seis cubos multidimensionales, uno por cada tabla de hechos, y 12 dimensiones correspondientes a cada una de las tablas de dimensiones. A continuación se muestra el diseño de los cubos OLAP para el MD Laboral.

Figura 18. Cubos OLAP correspondientes al MD Laboral

3.3.2 Implementación de la capa de visualización

El mapa de navegación permite tener una mejor visualización de cómo se va a estructurar la información. El MD Laboral está compuesto por un área de análisis general (A.A.G SIEJT), un área de análisis especifica del MD (A.A. Laboral), dos áreas de análisis para separar la información según la DPA que se esté utilizando, dos libros de trabajo (LT) generales dentro de los cuales se encuentran los LT 3351, 3352, 00-Disponibilidad laboral, 01-Apelaciones y 02-Seguridad social, que agrupan las vistas de análisis y los reportes asociados a cada LT. A continuación se muestra la estructura que posee la capa de visualización:

Figura 19. Mapa de navegación del MD Laboral

Descripciones del mapa de navegación:

A.A.G SIEJT: agrupa la información correspondiente a los MD diseñados para las diferentes áreas del departamento de Estadística judicial del Tribunal Supremo Popular.

A.A. Laboral: contiene toda la información específica del área Laboral, agrupándola en dos A.A para el DPA de los años anteriores y posteriores al 2011 respectivamente.

A.A. DPA (1976-2010): agrupa la información del área Laboral según la DPA anterior al año 2011.

A.A. DPA (2011-Actualidad): agrupa la información del área Laboral según la DPA posterior al año 2010.

LT Materia laboral: contiene los LT referentes a la materia laboral.

LT Instrucción 203: presenta los LT relacionados con la instrucción 203.

LT 3351: contiene las vistas de análisis y los reportes relacionados con las radicaciones de las demandas, los demandantes, las medidas aplicadas, las decisiones del O.J.L.B, las medidas aplicadas por el O.J.L.B a causales radicadas por indisciplinas y las medidas aplicadas por la administración al trabajador.

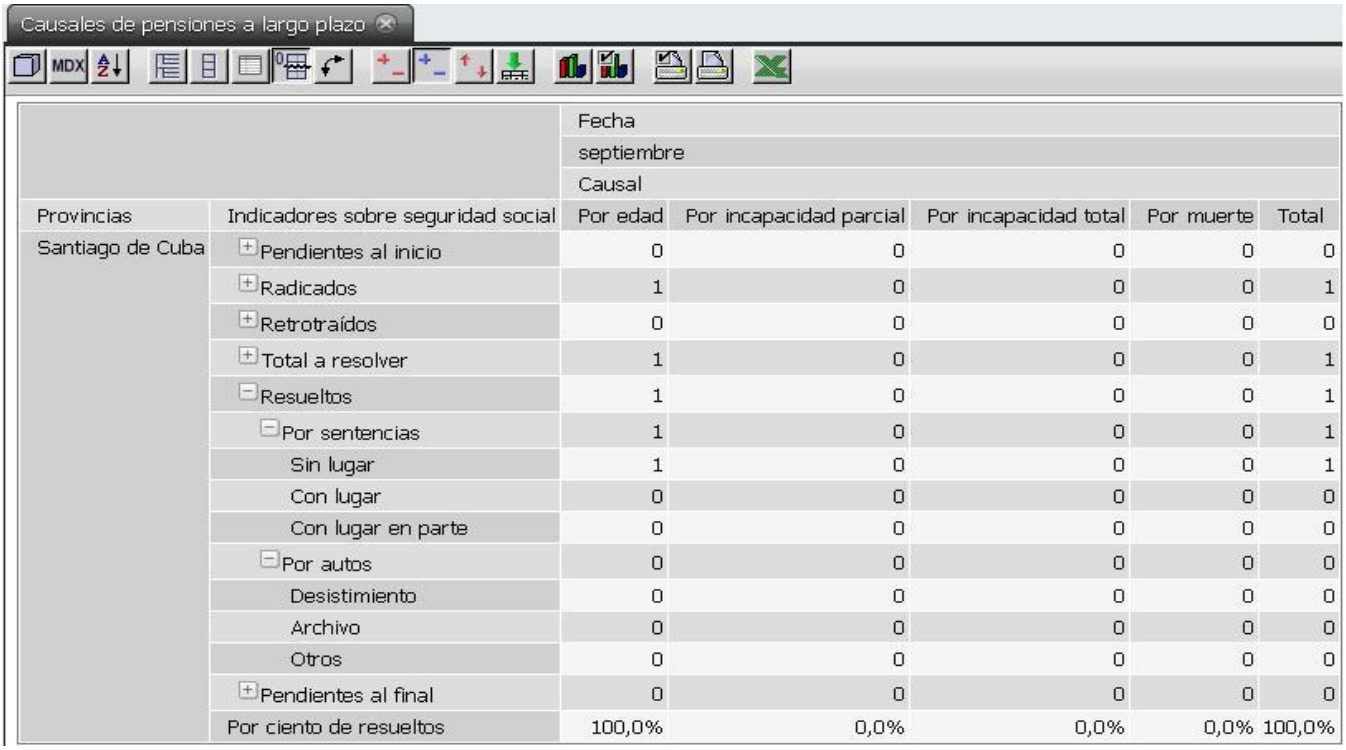
LT 3352: presenta las vistas de análisis y los reportes relacionados con los expedientes radicados-resueltos-pendientes, las medidas aplicadas, la declaración del total de las demandas establecidas y la declaración de las demandas por indisciplinas y derecho laboral establecidas por la administración y el trabajador.

LT 00-Disponibilidad laboral: contiene las vistas de análisis y los reportes asociados a la cantidad de procesos de disponibilidad laboral por causal y a los recursos de apelación presentados.

LT 01-Apelaciones: presenta las vistas de análisis y los reportes asociados al acumulado de las apelaciones y a las apelaciones resueltas por autos de no admisión, por otros autos y por sentencias.

LT 02-Seguridad social: contiene las vistas de análisis y los reportes relacionados con el acumulado de las pensiones a largo plazo, los recursos de apelación interpuestos y las causales de pensiones a largo plazo.

Las vistas de análisis y los reportes van a contener los valores asociados a los indicadores que son de interés para el cliente y que el mismo necesita consultar para su análisis. A continuación se refleja la vista de análisis "Cantidad de pensiones a largo plazo", ubicado en el LT 02-Seguridad social, en el cual se muestran las pensiones a largo plazo según las provincias, el mes, las causales y el estado del seguro social:



		Fecha				
		septiembre				
		Causal				
Provincias	Indicadores sobre seguridad social	Por edad	Por incapacidad parcial	Por incapacidad total	Por muerte	Total
Santiago de Cuba	Pendientes al inicio	0	0	0	0	0
	Radicados	1	0	0	0	1
	Retrotraídos	0	0	0	0	0
	Total a resolver	1	0	0	0	1
	Resueltos	1	0	0	0	1
	Por sentencias	1	0	0	0	1
	Sin lugar	1	0	0	0	1
	Con lugar	0	0	0	0	0
	Con lugar en parte	0	0	0	0	0
	Por autos	0	0	0	0	0
	Desistimiento	0	0	0	0	0
	Archivo	0	0	0	0	0
	Otros	0	0	0	0	0
	Pendientes al final	0	0	0	0	0
	Por ciento de resueltos	100,0%	0,0%	0,0%	0,0%	100,0%

Figura 20. Vista de análisis Cantidad de pensiones a largo plazo para el mes de septiembre

3.3.3 Implementación de la seguridad de los usuarios

Durante la implementación del subsistema de visualización del MD Laboral se crearon dos roles y dos usuarios, con el objetivo de proporcionar una mayor seguridad al sistema y definir los permisos de acceso a la información. A continuación se describen cada uno de ellos:

- Rol administrador de BD: tiene todos los permisos para administrar la BD, mediante el usuario administrador.

- Rol administrador de ETL: el usuario etl tendrá los permisos de lectura y escritura sobre las tablas de la BD.
- Rol analista: solo tiene permiso de lectura sobre la aplicación, mediante el usuario especialista_laboral.

3.4 Pruebas

Las pruebas de *software* tienen como propósito proporcionar información sobre la calidad de un software. Son una serie de actividades que se realizan para encontrar los posibles fallos de implementación, calidad o usabilidad de *software*, probando el comportamiento del mismo (39).

Las pruebas aplicadas al MD Laboral se especifican a continuación:

- **Pruebas unitarias:** permiten probar el correcto funcionamiento de un componente o subsistema específico. Esta prueba centra el proceso de verificación en la menor unidad del diseño del *software*, o sea, en algún componente del *software* o módulo (17).
- **Pruebas de integración:** permiten verificar la correcta integración de los componentes y subsistemas que conforman la solución. Estas pruebas son ejecutadas por los arquitectos de software (17).
- **Pruebas del sistema:** permiten validar el cumplimiento de los requisitos de información y funcionales definidos por los clientes. Son las pruebas más cercanas a la realidad del cliente, debido a que los probadores utilizan el sistema de la misma manera que será usado por los clientes. El propósito de las pruebas del sistema es detectar discrepancias entre el comportamiento del sistema construido y su especificación (17).
- **Pruebas de aceptación:** estas pruebas son realizadas por el cliente para verificar que se cumple con los requisitos planteados por el mismo y validar su conformidad con el producto. Son aquellas pruebas que demuestran al cliente que la funcionalidad está terminada y funciona correctamente (17).

Dentro de las herramientas utilizadas para que se apliquen los distintos tipos de pruebas se tienen los casos de prueba y las listas de chequeo.

3.4.1 Casos de prueba

Para lograr la calidad del producto de software es necesario realizar un conjunto de evaluaciones durante todo el proceso de desarrollo que implique al cliente y desarrollador. El diseño de los casos de prueba se realiza para comprobar que la documentación del producto se corresponde con lo establecido en la aplicación, o sea la aplicación se ejecuta sobre ciertas condiciones para descubrir errores antes de la

entrega del software al cliente. Para el MD Laboral se diseñaron cinco casos de prueba correspondientes a cinco CU de información. A continuación se muestra parte del diseño de caso de prueba correspondiente al CU Presentar información sobre la disponibilidad laboral, especificado en el documento DATEC-SIEJT_Laboral-CasoPrueba (CU Presentar información sobre la disponibilidad laboral) ubicado en el expediente del proyecto.

Escenario	Descripción	Variable entrada	Variable salida	Respuesta del sistema	Flujo central
EC 1.1 Cantidad de procesos de disponibilidad laboral por causal 1.	Muestra la cantidad de procesos de disponibilidad laboral por tipo de causal, estado del proceso, período de tiempo, provincia y tribunal provincial.	tribunal provincial	Obtener la cantidad de procesos de disponibilidad laboral por causal 1 dado el tribunal provincial.	El sistema muestra todas las variables disponibles para los análisis, ubicados en las filas y las columnas que pueden ser visualizadas en el reporte	Se abre la aplicación. Se autentica. Se entra al sistema. Se selecciona el área de análisis A.A.G SIEJT. Se selecciona el A.A Laboral y dentro, el A.A DPA (2011-Actualidad). Se selecciona el LT Instrucción 203 y dentro de este, el LT 00-Disponibilidad laboral. Se selecciona el reporte al que se le hace referencia en el escenario.
		causal	Obtener la cantidad de procesos de disponibilidad laboral por causal 1 dada la causal.		
		estado proceso	Obtener la cantidad de procesos de disponibilidad laboral por causal 1 dado el estado del proceso.		
		dpa provincia	Obtener la cantidad de procesos de disponibilidad laboral por causal 1 dado el dpa de la provincia.		
		temporal mes	Obtener la cantidad de procesos de disponibilidad laboral por causal 1 dado el mes y el año.		

Figura 21. Caso de prueba CU Presentar información sobre la disponibilidad laboral

3.4.2 Listas de chequeo

Las listas de chequeo van a contener indicadores a evaluar, los cuales estarán ubicados en tres secciones:

- Estructura del documento: contiene todos los aspectos definidos por el expediente del proyecto.
- Indicadores definidos en el desarrollo: contiene todos los indicadores a evaluar.
- Semántica del documento: abarca todos los indicadores a evaluar respecto a la ortografía, redacción y otros aspectos de forma y estilo.

Las listas de chequeo están especificadas por los siguientes elementos:

- Peso: define si el indicador a evaluar es crítico o no.
- Indicadores a evaluar: son los indicadores que servirán para evaluar las tres secciones fundamentales que componen las listas de chequeo.
- Evaluación: es el modo de evaluar el indicador, este obtiene evaluación de 1 en caso de que exista alguna dificultad sobre el indicador y de 0 en caso contrario.

- No procede: especifica que el indicador no presenta evaluación. Se usa para especificar que el indicador no es necesario evaluarlo en ese caso.
- Cantidad de elementos afectados: especifica la cantidad de errores que se identificaron en el indicador.
- Comentario: especifica los señalamientos o sugerencias que desee incluir la persona encargada de aplicar la lista de chequeo.

Evaluación del resultado de la lista de chequeo:

- Se aborta el proceso de aplicación de la lista en caso de:
  - ✓ Existan al menos dos indicadores críticos evaluados de mal.
  - ✓ Más del 50% de los indicadores a evaluar están evaluados de mal.
  - ✓ Se mantienen las no conformidades de una revisión a otra.
- Se evalúa de regular la calidad de los artefactos de ETL en caso de:
  - ✓ Incumple con los indicadores críticos a evaluar de las secciones Estructura del documento y Semántica del documento de la lista de chequeo.
  - ✓ Existe al menos un indicador crítico evaluado de mal.
  - ✓ Existen al menos cinco indicadores no críticos evaluados de mal.
- Los artefactos de los procesos de ETL son evaluados de bien cuando no cumple con ningún criterio de los dos puntos anteriores.

En esta investigación se aplicaron las siguientes listas de chequeo a los artefactos de los procesos de ETL con el fin de medir y evaluar la confiabilidad y seguridad de los datos cargados mediante preguntas bien elaboradas:

- Lista de chequeo del Mapa Lógico de Datos.
- Lista de chequeo del Diccionario de Datos.
- Lista de chequeo de Registro de Sistemas Fuentes.
- Listas de chequeo del Perfilado de Datos

3.4.3 Resultados de las pruebas

Una vez aplicadas las pruebas al MD Laboral se obtuvieron los siguientes resultados:

- Pruebas unitarias: fueron realizadas por especialistas del departamento de Almacenes de datos y durante su aplicación se detectaron siete no conformidades (NC), de las cuales dos tenían complejidad alta, cuatro complejidad media y una complejidad baja, que fueron corregidas durante el desarrollo de la solución. Entre las NC detectadas se encuentran:

- ✓ Abundar en los pasos del diseño de las transformaciones.
- ✓ Poner el cubo.xml dentro del área correspondiente al MD Laboral.

➢ Pruebas de integración: fueron aplicadas por miembros del equipo de desarrollo y especialistas del departamento de Almacenes de datos, arrojando cinco NC durante su aplicación que fueron solucionadas por el equipo de desarrollo.

- ✓ Corregir los datos duplicados de la provincia de Matanzas.
- ✓ Revisar los cálculos de los valores para administración y trabajadores.

➢ Pruebas del sistema: durante su aplicación se obtuvieron cuatro NC, una de complejidad alta, dos de complejidad media y una de complejidad baja, que fueron solucionadas satisfactoriamente. Entre las NC detectadas se pueden mencionar:

- ✓ El flujo central descrito en los casos de prueba no se corresponden con la aplicación.
- ✓ En las descripciones de las variables la cantidad es un valor fijo.

➢ Pruebas de aceptación: durante estas pruebas se realizó la auditoría de los datos a una porción de los modelos con formato excel y dbf, para comparar los reportes con que cuenta el cliente, existentes en estos modelos, con las vistas de análisis y los reportes realizados con el Pentaho BI Server. Las pruebas de aceptación arrojaron dos NC de complejidad alta y dos NC de complejidad media, que fueron solucionadas satisfactoriamente. Entre las NC identificadas se encuentran:

- ✓ No están implementados los reportes de los LT 3351 y 3352.
- ✓ Cambiar en los reportes el nombre de "A resolver" por Indicadores.

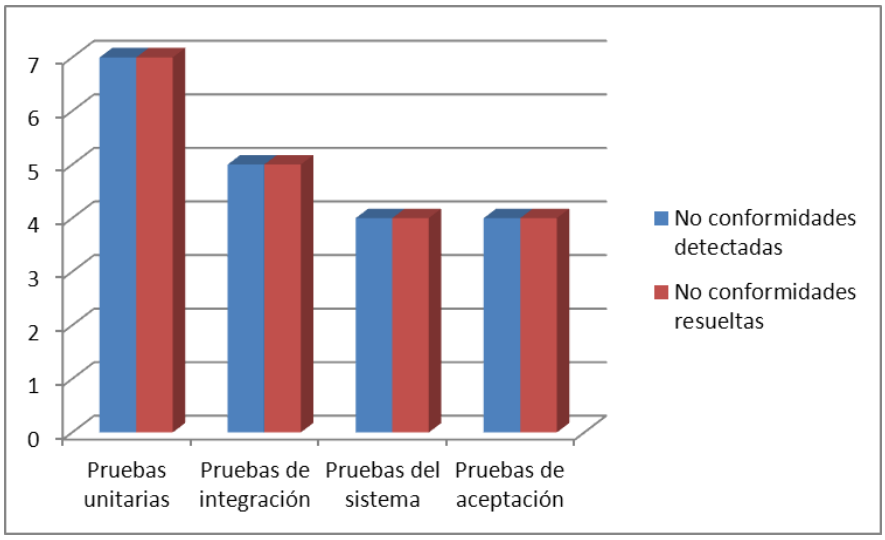


Figura 22. Resultados de las pruebas aplicadas al MD Laboral

En la siguiente figura se muestra el comportamiento de los indicadores evaluados en la lista de chequeo del Diccionario de Datos. En la lista de chequeo se identificaron 13 indicadores, de ellos cinco críticos, y luego de aplicada la herramienta no generó ninguna NC.

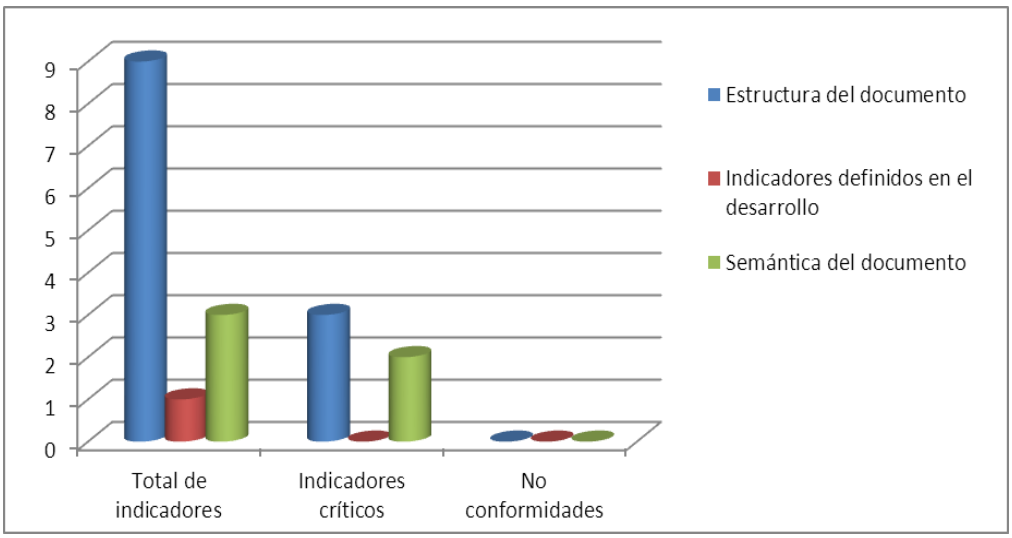


Figura 23. Comportamiento de los indicadores de la lista de chequeo del Diccionario de Datos

3.5 Conclusiones del capítulo

En este capítulo se realizó la implementación del MD Laboral teniendo en cuenta las necesidades de los usuarios, con el objetivo de cumplir con las expectativas del cliente. Se definieron los diferentes esquemas, tablas de hechos y de dimensiones que conforman el MD Laboral, permitiendo adquirir una mejor visión de la estructura de su BD. Se implementaron los diferentes trabajos y transformaciones que sirvieron para realizar la limpieza y estandarización de los datos, que luego se cargaron en la BD. Además se confeccionaron los cubos OLAP correspondientes a cada hecho, que permitieron el acceso a los datos presentes en las tablas de hechos. La implementación del mapa de navegación permitió adquirir una visualización de cómo se estructuraron las A.A, los LT y los reportes asociados a estos. Finalmente se aplicaron los casos de prueba y las listas de chequeo al MD Laboral, con el propósito de identificar posibles errores en la solución.

Conclusiones generales

Las siguientes conclusiones demuestran el cumplimiento de los objetivos específicos propuestos en la presente investigación:

- La selección de la metodología, herramientas y tecnologías a utilizar en el desarrollo de la solución, guiaron y facilitaron el proceso de construcción del mercado de datos Laboral.
- El análisis y diseño del mercado de datos Laboral permitió identificar los requisitos funcionales, no funcionales, de información y las reglas del negocio, que sirvieron como base para el diseño de la solución.
- El diseño de los subsistemas de almacenamiento, integración y visualización permitió obtener como elementos principales el modelo dimensional, el diseño de las transformaciones, la arquitectura de la información y el diseño de los cubos OLAP, que fueron la base para la implementación de la solución.
- Se implementaron los subsistemas de la solución, permitiendo obtener como resultado un mercado de datos poblado con información que va a estar disponible para ser consultada por los clientes a través de las vistas de análisis y los reportes, contribuyendo al proceso de toma de decisiones.
- Las pruebas realizadas permitieron validar la calidad del producto y obtener resultados satisfactorios.

Recomendaciones

- Integrar el MD Laboral al Sistema de Información de Gestión Estadística, para facilitar la carga incremental de los datos de la organización.
- Realizar mejoras en la capa de visualización de la información, aplicando otras técnicas de inteligencia de negocios, como los Cuadros de Mando Integral, que permitan obtener datos relevantes para la toma de decisiones en el negocio.

Referencias bibliográficas

1. EcuRed. *EcuRed.* [Online] http://www.ecured.cu/index.php/Almacén_de_Datos.
2. **Rizo Rizo, MSc. Emma R., et al.** Biblioteca Virtual de las ciencias en Cuba. *Biblioteca Virtual de las ciencias en Cuba.* [Online] http://www.bibliociencias.cu/gsdl/collect/libros/index/assoc/HASH0106/b6fac6b9.dir/doc.pdf
3. **Mazón López, José Norberto, Pardillo Vela, Jesús and Trujillo Mondéjar, Juan Carlos.** Google. *Google.* [Online] Enero 2011. http://books.google.com.cu/books/about/Diseño_y_explotación_de_almacenes_de_d.html?id=E7Aceg--o4oC&redir_esc=y.
4. **Torres Torrillas, Francisco José Lucas, et al.** UNIVERSIDAD DE CASTILLA-LA MANCHA ESCUELA SUPERIOR DE INFORMÁTICA. *UNIVERSIDAD DE CASTILLA-LA MANCHA ESCUELA SUPERIOR DE INFORMÁTICA.* [Online] http://alarcos.inf-cr.uclm.es/doc/bbddavanzadas/08-09/FUNCIONALIDAD 4.pdf.
5. **Casales Cabrera, María Evelia.** Scribd. *Scribd.* [Online] Julio 14, 2010. http://www.scribd.com/doc/34312997/Data-Warehouse.
6. **Ricardo Dario, Ing. Bernabeu.** DATAPRIX. *DATAPRIX.* [Online] Julio 19, 2010. http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/i-data-warehousing-investigacion-y-sistematizacion-concepto-13.
7. **Guevara Lenis, Jorge Eduardo and Valencia Arcos, Janeth Del Carmen.** DSpace. *DSpace.* [Online] http://bibdigital.epn.edu.ec/bitstream/15000/445/1/CD-0827.pdf.
8. EcuRed. *EcuRed.* [Online] 2012. http://www.ecured.cu/index.php/Inteligencia_en_el_Negocio.
9. Tecnologías de Información. *Tecnologías de Información.* [Online] http://www.tecnologias-informacion.com/soportedecisiones.html.
10. Trabajo en grupo y gestión del conocimiento. *Trabajo en grupo y gestión del conocimiento.* [Online] Enero 14, 2012. http://groupware-g9.blogspot.com/2012/01/sistemas-de-informacion-ejecutiva-eis.html.
11. **Marín Llanes, Luis A., Carro Cartaya, Juan C**. La Minería de Datos como herramienta de inteligencia competitiva. [Online] http://www.redciencia.cu/empres/Intempres2000/Sitio/Principal/Literatura/DATA-MINING.pdf.
12. **G. Bigatti, Ing. Cristian.** Scribd. *Scribd.* [Online] http://www.scribd.com/doc/48897874/16/Drill-Down-y-Roll-Up.

13. **Oporto díaz, Mg. Samuel**. Scribd. *Scribd*. [Online] http://www.scribd.com/doc/137601593/ Class-51-OLAP-ppt.
14. **Gerolami, Nicolás, Revello, Esteban and Venzal, Germain.** Facultad de Ingeniería. *Facultad de Ingeniería.* [Online] Diciembre 13, 2011. http://www.fing.edu.uy/~asabigue/prgrado/ 2010dw.pdf.
15. Scribd. *Scribd.* [Online] http://www.scribd.com/doc/96120219/Procesamiento-Y-Analisis-en-Linea-OLAP.
16. EcuRed. *EcuRed.* [Online] http://www.ecured.cu/index.php/Metodologías_de_desarrollo_de_software.
17. **González Hernández, Yanisbel.** *PROPUESTA DE METODOLOGIA DE DASARROLLO DE AlMACENES DE DATOS.* 2012.
18. Soluciones y propuestas Rational. *Soluciones y propuestas Rational*. [Online] http://www.rational.com.ar/herramientas.
19. EcuRed. *EcuRed.* [Online] http://www.ecured.cu/index.php/Visual_Paradigm.
20. EcuRed. *EcuRed.* [Online] http://www.ecured.cu/index.php/SGBD.
21. MySQL, The world´s most popular open source database. *MySQL, The world´s most popular open source database*. [Online] http://dev.mysql.com/doc/refman/5.0/es/features.html.
22. Oracle, Hardware and Software, Engineered to Work Together. *Oracle, Hardware and Software, Engineered to Work Together*. [Online] http://www.oracle.com.
23. PostgreSQL. *PostgreSQL.* [Online] http://www.postgresql.org.
24. PgAdmin. *PgAdmin.* [Online] http://www.pgadmin.org/.
25. Data Cleaner. *Data Cleaner.* [Online] http://datacleaner.eobjects.org.
26. Pentaho. *Pentaho.* [Online] http://kettle.pentaho.com/.
27. Pentaho. *Pentaho.* [Online] http://mondrian.pentaho.com/documentation.
28. Pentaho. *Pentaho.* [Online] http://community.pentaho.com/.
29. The Apache Software Foundation. *The Apache Software Foundation.* [Online] http://tomcat.apache.org/.
30. Manual Generador de Reportes de Pentaho. *Manual Generador de Reportes de Pentaho.* [Online] http://www.onuva.com/wp-content/uploads/2012/07/Manual_PRD.pdf.
31. **Chiciaza, Ing. Janneth Alexandra.** Universidad Técnica Particular de Loja. *Universidad Técnica Particular de Loja.* [Online] http://rsa.utpl.edu.ec/material/208/G181003.2.pdf.

32. **Rojas, Marianal Isabel and La Red Martinez, Mgter. David Luis.** [Online] 2009. http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MonoAdsDiseno.pdf.
33. HerramientaRecolAna_SIEJT_Laboral. HerramientaRecolAna_SIEJT_Laboral. [Expediente del proyecto].
34. **Kimball, Ralph.** *The Data Warehouse toolkit: Practical techniques for building Dimensional Data Warehouses.*
35. **Azán Basallo,Yasser,Díaz Estrada,Anay**. Una experiencia en integración de aplicaciones empresariales. Una experiencia en integración de aplicaciones empresariales. [Online] http://rcci.uci.cu/index.php/rcci/article/download/65/59%E2%80%8E.
36. EcuRed. *EcuRed*. [Online] http://www.ecured.cu/index.php/Cubos_OLAP.
37. DataPrix, Knowledge is the goal. *DataPrix, Knowledge is the goal*. [Online] http://www.dataprix.com/blogs/bernabeu-dario/dimensiones-lentamente-cambiantes.
38. EcuRed. *EcuRed*. [Online] http://www.ecured.cu/index.php/Metadatos.
39. EcuRed. *EcuRed*. [Online] http://www.ecured.cu/index.php/Pruebas_de_software.

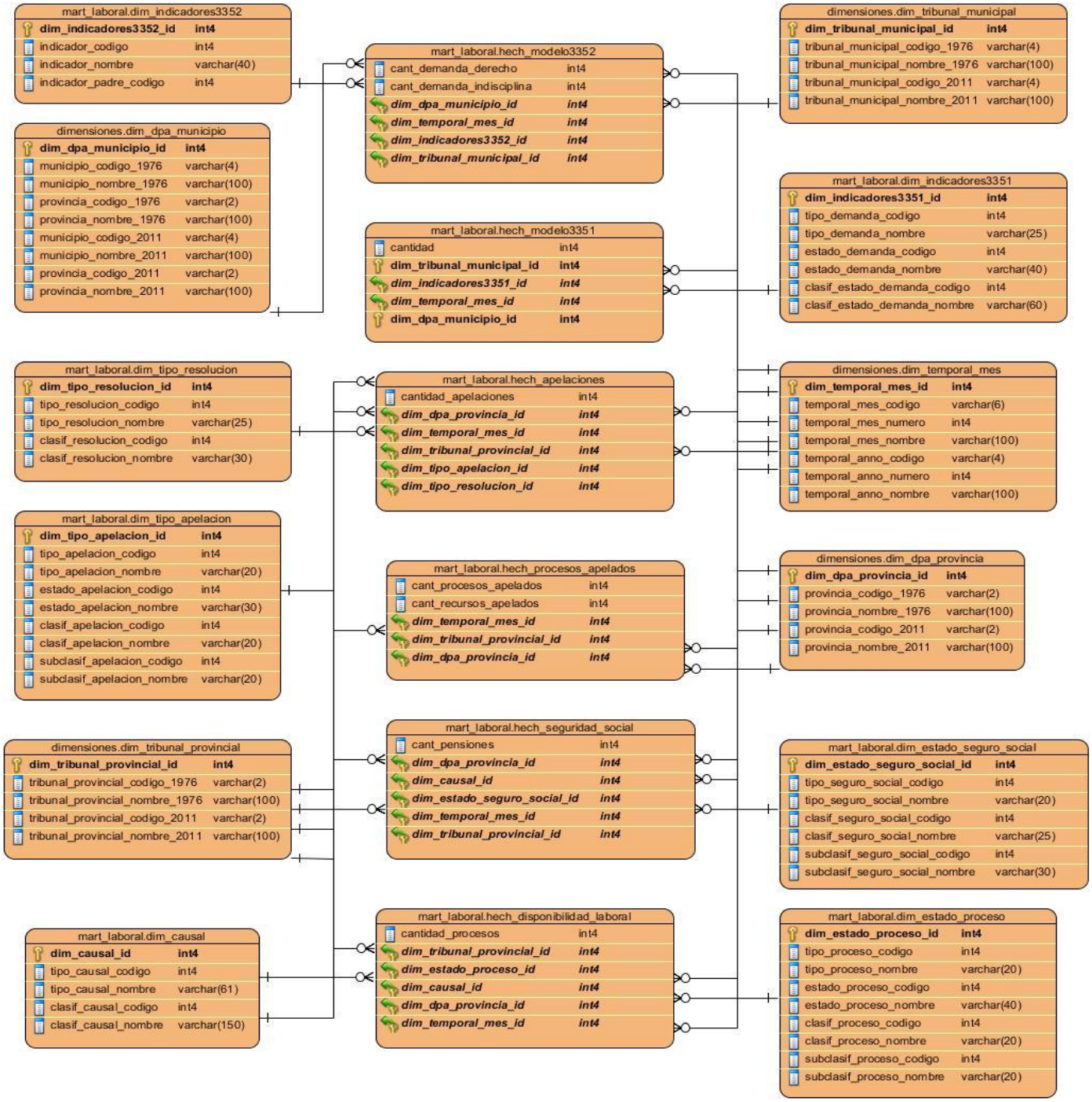
Bibliografía

1. EcuRed. *EcuRed.* [Online] http://www.ecured.cu/index.php/Almacén_de_Datos.
2. **Rizo Rizo, MSc. Emma R., et al.** Biblioteca Virtual de las ciencias en Cuba. *Biblioteca Virtual de las ciencias en Cuba.* [Online] http://www.bibliociencias.cu/gsdl/collect/libros/index/assoc/HASH0106/b6fac6b9.dir/doc.pdf
3. **Mazón López, José Norberto, Pardillo Vela, Jesús and Trujillo Mondéjar, Juan Carlos.** Google. *Google.* [Online] Enero 2011. http://books.google.com.cu/books/about/Diseño_y_explotación_de_almacenes_de_d.html?id=E7Aceg--o4oC&redir_esc=y.
4. **Torres Torrillas, Francisco José Lucas, et al.** UNIVERSIDAD DE CASTILLA-LA MANCHA ESCUELA SUPERIOR DE INFORMÁTICA. *UNIVERSIDAD DE CASTILLA-LA MANCHA ESCUELA SUPERIOR DE INFORMÁTICA.* [Online] http://alarcos.inf-cr.uclm.es/doc/bbddavanzadas/08-09/FUNCIONALIDAD 4.pdf.
5. **Casales Cabrera, María Evelia.** Scribd. *Scribd.* [Online] Julio 14, 2010. http://www.scribd.com/doc/34312997/Data-Warehouse.
6. **Ricardo Dario, Ing. Bernabeu.** DATAPRIX. *DATAPRIX.* [Online] Julio 19, 2010. http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/i-data-warehousing-investigacion-y-sistematizacion-concepto-13.
7. **Guevara Lenis, Jorge Eduardo and Valencia Arcos, Janeth Del Carmen.** DSpace. *DSpace.* [Online] http://bibdigital.epn.edu.ec/bitstream/15000/445/1/CD-0827.pdf.
8. EcuRed. *EcuRed.* [Online] 2012. http://www.ecured.cu/index.php/Inteligencia_en_el_Negocio.
9. Tecnologías de Información. *Tecnologías de Información.* [Online] http://www.tecnologias-informacion.com/soportedecisiones.html.
10. Trabajo en grupo y gestión del conocimiento. *Trabajo en grupo y gestión del conocimiento.* [Online] Enero 14, 2012. http://groupware-g9.blogspot.com/2012/01/sistemas-de-informacion-ejecutiva-eis.html.
11. **Marín Llanes, Luis A., Carro Cartaya, Juan C**. La Minería de Datos como herramienta de inteligencia competitiva. [Online] http://www.redciencia.cu/empres/Intempres2000/Sitio/Principal/Literatura/DATA-MINING.pdf.
12. **G. Bigatti, Ing. Cristian.** Scribd. *Scribd.* [Online] http://www.scribd.com/doc/48897874/16/Drill-Down-y-Roll-Up.

13. **Oporto díaz, Mg. Samuel**. Scribd. *Scribd*. [Online] http://www.scribd.com/doc/137601593/ Class-51-OLAP-ppt.
14. **Gerolami, Nicolás, Revello, Esteban and Venzal, Germain.** Facultad de Ingeniería. *Facultad de Ingeniería.* [Online] Diciembre 13, 2011. http://www.fing.edu.uy/~asabigue/prgrado/ 2010dw.pdf.
15. Scribd. *Scribd.* [Online] http://www.scribd.com/doc/96120219/Procesamiento-Y-Analisis-en-Linea-OLAP.
16. EcuRed. *EcuRed.* [Online] http://www.ecured.cu/index.php/Metodologías_de_desarrollo_de_software.
17. **González Hernández, Yanisbel.** *PROPUESTA DE METODOLOGIA DE DASARROLLO DE AlMACENES DE DATOS.* 2012.
18. Soluciones y propuestas Rational. *Soluciones y propuestas Rational*. [Online] http://www.rational.com.ar/herramientas.
19. EcuRed. *EcuRed.* [Online] http://www.ecured.cu/index.php/Visual_Paradigm.
20. EcuRed. *EcuRed.* [Online] http://www.ecured.cu/index.php/SGBD.
21. MySQL, The world´s most popular open source database. *MySQL, The world´s most popular open source database*. [Online] http://dev.mysql.com/doc/refman/5.0/es/features.html.
22. Oracle, Hardware and Software, Engineered to Work Together. *Oracle, Hardware and Software, Engineered to Work Together*. [Online] http://www.oracle.com.
23. PostgreSQL. *PostgreSQL.* [Online] http://www.postgresql.org.
24. PgAdmin. *PgAdmin.* [Online] http://www.pgadmin.org/.
25. Data Cleaner. *Data Cleaner.* [Online] http://datacleaner.eobjects.org.
26. Pentaho. *Pentaho.* [Online] http://kettle.pentaho.com/.
27. Pentaho. *Pentaho.* [Online] http://mondrian.pentaho.com/documentation.
28. Pentaho. *Pentaho.* [Online] http://community.pentaho.com/.
29. The Apache Software Foundation. *The Apache Software Foundation.* [Online] http://tomcat.apache.org/.
30. Manual Generador de Reportes de Pentaho. *Manual Generador de Reportes de Pentaho.* [Online] http://www.onuva.com/wp-content/uploads/2012/07/Manual_PRD.pdf.
31. **Chiciaza, Ing. Janneth Alexandra.** Universidad Técnica Particular de Loja. *Universidad Técnica Particular de Loja.* [Online] http://rsa.utpl.edu.ec/material/208/G181003.2.pdf.

32. **Rojas, Marianal Isabel and La Red Martinez, Mgter. David Luis.** [Online] 2009. http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MonoAdsDiseno.pdf.
33. HerramientaRecolAna_SIEJT_Laboral. HerramientaRecolAna_SIEJT_Laboral. [Expediente del proyecto].
34. **Kimball, Ralph.** *The Data Warehouse toolkit: Practical techniques for building Dimensional Data Warehouses.*
35. **Azán Basallo,Yasser,Díaz Estrada,Anay**. Una experiencia en integración de aplicaciones empresariales. Una experiencia en integración de aplicaciones empresariales. [Online] http://rcci.uci.cu/index.php/rcci/article/download/65/59%E2%80%8E.
36. EcuRed. *EcuRed*. [Online] http://www.ecured.cu/index.php/Cubos_OLAP.
37. DataPrix, Knowledge is the goal. *DataPrix, Knowledge is the goal*. [Online] http://www.dataprix.com/blogs/bernabeu-dario/dimensiones-lentamente-cambiantes.
38. EcuRed. *EcuRed*. [Online] http://www.ecured.cu/index.php/Metadatos.
39. EcuRed. *EcuRed*. [Online] http://www.ecured.cu/index.php/Pruebas_de_software.
40. Almacén de Datos para la Gestión Contable de la EMPAI. Data Warehouse Management Accounting Officer of the EMPAI. [Online] http:// redalyc.uaemex.mx/src/inicio/ArtPdfRed.jsp ?iCve=193915954004.
41. **Soni, Rajinder.** http://www.indiastudychannel.com/resources/84786-Business-Intelligence-The-Current-Need-Society.aspx(20th century fox).
42. BI Latino. *BI Latino*. [Online] http://www.bi-spain.com/articulo/57065/business-intelligence/ medios-de-comunicacion/losestudios-walt-disney-extienden-el-uso-de-herramientas-de-bi-en-sus-departamentos-de-finanzasmarketing-y-operaciones(walt disney).
43. BI Latino. *BI Latino*. [Online] http://www.bi-spain.com/articulo/44257/data-warehouse/los-almacenes-usa-wal-mart-implantan-lasolucion-bi-neoview-de-hp-para-el-analisis-de-datos-de-sus-puntos-de-venta(wallmart).

Anexos



Anexo 1. Modelo dimensional correspondiente al MD Laboral