

Universidad de las Ciencias Informáticas

FACULTAD 8



Análisis, diseño e implementación del mercado de datos para los indicadores específicos de la ocupación para la Oficina Nacional de Estadística.

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

Autoras:

Meylin Rodríguez Cabrera.

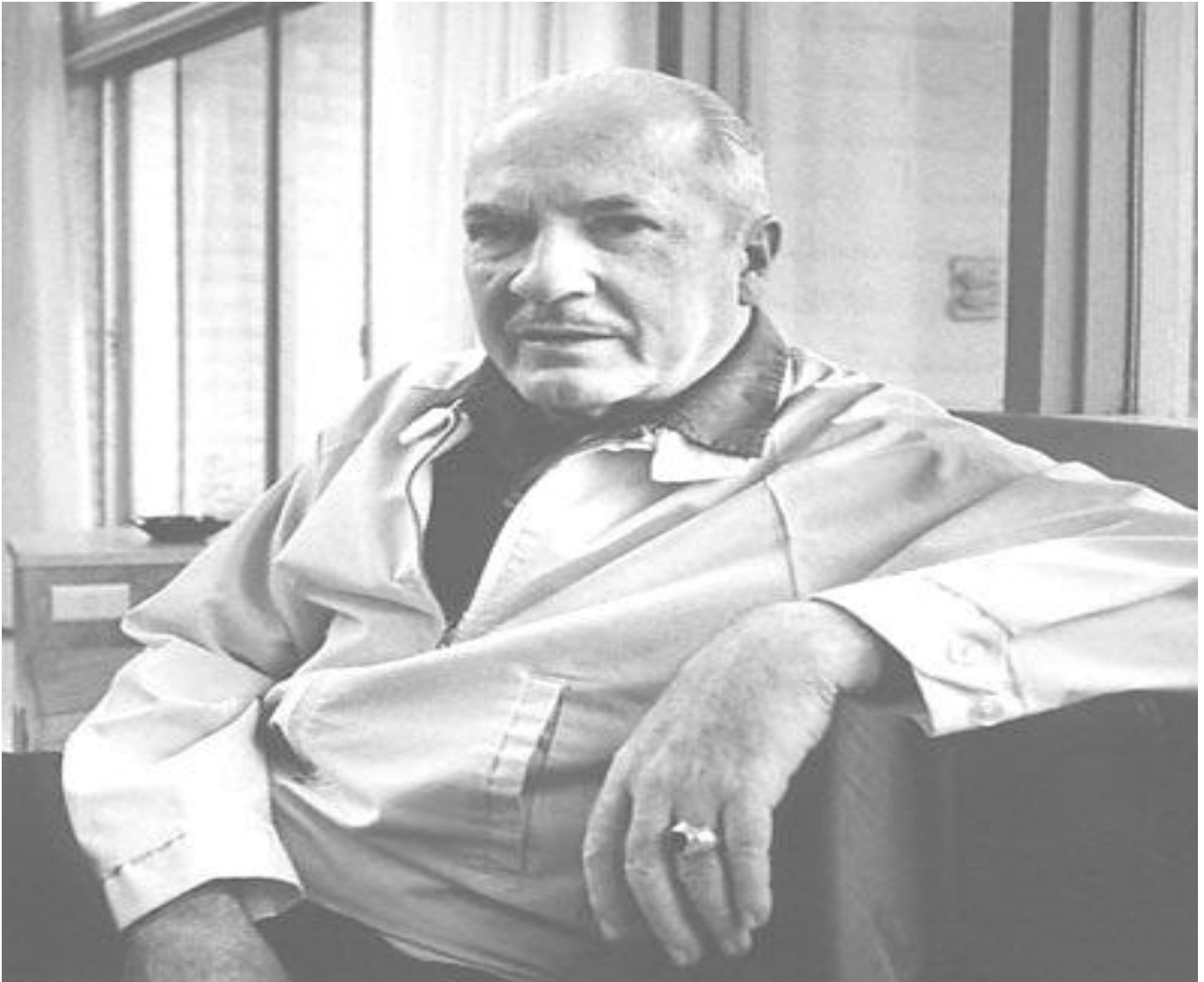
Yaillet de la Caridad Vilaró González.

Tutora:

Ing. Yisel Barrabía Legrá.

Ciudad de La Habana, junio 2010

“Año 52 de la Revolución”



No existe la suerte. Sólo hay preparación adecuada o inadecuada para hacer frente a una estadística.

Robert Heinlein

Declaración de autoría

Declaramos que somos las únicas autoras del trabajo “Análisis, diseño e implementación del mercado de datos para los indicadores específicos de la ocupación para la Oficina Nacional de Estadística” y autorizamos a la Facultad 8 de la Universidad de las Ciencias Informáticas a hacer uso del mismo en su beneficio. Para que así conste firmamos la presente a los ____ días del mes de _____ del año _____.

Meylin Rodríguez Cabrera
Autora

Yaillet de la Caridad Vilaró González
Autora

Yisel Barrabia Legrá
Tutora

Agradecimientos

Queremos agradecerle de forma especial a Yisel, nuestra tutora, por apoyarnos y por estar siempre a nuestro lado.

A mi mamá, a mi papá y a Agustín, pues me apoyaron en los momentos más difíciles de la carrera y depositaron toda su confianza en mí.

A mis abuelos y a María por creer que sería capaz de lograr mi sueño.

A mis tías y tíos especialmente a **Bety** y a **Yosvani** por su preocupación por mis estudios todo este tiempo.

De forma general a toda **mi linda familia**, a todos los que me ayudaron y depositaron su granito de arena de forma especial.

Meylin

A **mis padres**, que me dieron la vida, me educaron, nunca me dieron la espalda en cada momento difícil e hicieron todo lo que estaba a su alcance y más para que lo tuviera todo en estos cinco años. Por eso y por mucho más, a ustedes, dos gracias y felicidades, que este es el fruto de su esfuerzo.

A **mis abuelos** porque siempre confiaron en mí.

A **mis hermanos** que de una forma u otra me brindaron su ayuda en los momentos difíciles.

A todos los que directa o indirectamente hayan puesto un granito de arena en que hoy esté optando por este título.

Yaillet de la Caridad

Dedicatoria

Dedico este trabajo a quien ha sido un motor impulsor en mi vida y un ejemplo a seguir, a Gina de la Caridad, mi mamá, que me ha apoyado incondicionalmente todos estos años de mi carrera.

Meylin

Dedico este trabajo a quienes han sabido guiar mis pasos, mi familia, y a todos los que han tenido que ver con la realización de este sueño.

Yaillet de la Caridad

Resumen

En el presente trabajo de diploma se muestra el resultado del análisis, diseño e implementación del mercado de datos para los indicadores específicos de la ocupación para la Oficina Nacional de Estadística. El mismo se enmarca en el área de los almacenes de datos y mercados de datos para el análisis de la información estadística. Incluye una revisión detallada acerca de las herramientas de modelado, sistema gestor de bases de datos y metodología a utilizar para el desarrollo la solución, abordando temas de interés, como lo son; la definición y reglas del negocio, las necesidades de los usuarios, todos con el objetivo de lograr un mejor entendimiento del negocio.

Como resultado se exponen las estructuras dimensionales para el modelo estadístico “Número de trabajadores por categoría ocupacional y sexo”, que comprende las dimensiones, medidas y el hecho imprescindibles para soportar el análisis estadístico. La principal novedad consiste en realizar la implementación del mercado de datos utilizando herramientas libres.

Palabras claves: almacenes de datos, mercados de datos

Tabla de contenidos

Introducción	1
Capítulo 1	6
Fundamentación teórica.....	6
1.1 Introducción	6
1.2 Análisis de los trabajos relacionados con el tema en el ámbito internacional y nacional	6
1.2.1 INE en España	6
1.2.2 INE en Chile	7
1.3 Almacenes de datos.....	7
1.4 Mercados de datos.....	9
1.5 Modelos y modos de almacenamiento de los datos	9
1.5.1 Modelos de almacenamiento de datos	9
1.5.2 Modos de almacenamiento de datos	10
1.6 Sistemas gestores de bases de datos.....	11
1.6.1 Selección del gestor de bases de datos a utilizar	13
1.7 Herramientas CASE.....	14
1.7.1 Selección de las herramientas CASE a utilizar.....	15
1.8 Metodologías de desarrollo	15
1.8.1 Selección de la metodología a utilizar	17
1.9 Conclusiones del capítulo	18
Capítulo 2	19
Análisis y Diseño de la solución	19
2.1 Introducción	19

2.2 Análisis	19
2.2.1 Definición del negocio	19
2.2.2 Temas de análisis	19
2.2.3 Roles y permisos.....	19
2.2.4 Reglas del negocio.....	20
2.2.5 Necesidades de los usuarios.....	22
2.2.6 Requisitos	23
2.2.6.1 Requisitos funcionales.....	23
2.2.6.2 Requisitos no funcionales.....	23
2.2.6.3 Requisitos de información	24
2.2.6.4 Requisitos multidimensionales	26
2.2.7 Casos de uso del sistema	32
2.2.7.1 Casos de uso funcionales.....	32
2.2.7.2 Caso de uso informativo.....	32
2.2.8 Diagrama de casos de uso.....	33
2.2.9 Descripción textual de los casos de uso.....	33
2.3 Diseño.....	38
2.3.1 Matriz dimensional	38
2.3.2 Modelo de datos.....	40
2.3.3 Esquema de seguridad	44
2.3.4 Política de respaldo y recuperación.....	44
2.4 Conclusiones parciales	45
Capítulo 3	46

Implementación y prueba.....	46
3.1 Introducción	46
3.2 Implementación.....	46
3.2.1 Modelo de datos físico	46
3.2.2 Estructuras de datos	46
3.2.2.1 Esquemas y tablas	46
3.2.2.2 Restricciones y secuencias	49
3.2.2.3 Índices.....	51
3.2.2.4 Descripción del artefacto DDL	52
3.2.3 Usuarios y privilegios	52
3.2.3.1 Usuarios	52
3.2.3.2 Privilegios.....	53
3.2.3.3 Descripción del artefacto DCL	53
3.2.4 Carga de nomencladores	54
3.2.4.1 Nomenclador	54
3.2.4.2 Descripción del artefacto DML.....	55
3.3 Guía de Implantación	55
3.3.1 Requerimientos	55
3.3.2 Secuencia de pasos	56
3.4 Validación y pruebas.....	56
3.4.1 Listas de chequeo del análisis.....	56
3.4.2 Validación de requisitos por el cliente	57
3.4.3 Lista de chequeo del diseño.....	57

3.4.4 Pruebas de implantación	57
3.5 Conclusiones parciales	58
Conclusiones generales.....	59
Recomendaciones	60
Bibliografía.....	61
Referencias bibliográficas	63
Anexos.....	66
Glosario de términos.....	67

Índice de tablas

Tabla 1: Descripción del caso de uso Extraer información del archivo DBF del número de trabajadores por categoría ocupacional y sexo.....	33
Tabla 2: Descripción del caso de uso Transformar y cargar información del archivo DBF del número de trabajadores por categoría ocupacional y sexo	35
Tabla 3: Descripción del caso de uso Obtener información sobre el número de trabajadores ocupados por categoría ocupacional y sexo.....	37
Tabla 4: Matriz dimensional	39
Tabla 5: Descripción de la dimensión Sector ocupación	40
Tabla 6: Descripción de la dimensión Categoría ocupacional	41
Tabla 7: Descripción de la dimensión Contratados por tiempo determinado y por edad.....	41
Tabla 8: Descripción de los atributos de la tabla de hecho identificada.....	42
Tabla 9: Descripción de las medidas de la tabla de hecho identificada	43
Tabla 10: Descripción de esquemas y tablas	47
Tabla 11: Descripción de llaves foráneas.....	50
Tabla 12: Descripción de llaves primarias.....	50
Tabla 13: Descripción de los índices.....	51
Tabla 14: Descripción de los casos de prueba	58

Índice de figuras

Figura 1: Diagrama de casos de uso del sistema33

Figura 2: Modelo dimensional40

Figura 3: Hecho identificado42

Introducción

El mundo informático se encuentra inmerso en un acelerado desarrollo, pues es cada vez más dinámico e impredecible, el cual se manifiesta en diversas ramas de la sociedad, proporcionando formas de comunicación cada vez más eficaces y flexibles.

La informática a nivel mundial se conoce como la nueva revolución tecnológica, pues han sido numerosos los avances alcanzados en dicha materia, debido a la necesidad de información y de nuevos conocimientos que tiene el mundo actual, no únicamente para el desarrollo de los pueblos, sino también para el de las nuevas tecnologías.

Cuba va aparejada al creciente desarrollo de dicha tecnología, aportando resultados significativos en las diversas esferas.

Muchas de las empresas se han visto obligadas a informatizarse, digitalizando el gran cúmulo de información poseído para poder establecer relaciones con otras instituciones, facilitando los convenios y relaciones entre las mismas.

La Universidad de las Ciencias Informáticas (UCI) está a la vanguardia de la producción de software, es un motor impulsor para la fomentación del desarrollo de la Industria del Software en Cuba, son múltiples los proyectos productivos que se realizan en ella, y que contribuyen a la adquisición de nuevos conocimientos.

Entre los proyectos productivos enfrascados en dicha universidad se encuentra la realización del almacén de datos para la Oficina Nacional de Estadística (ONE), la cual almacena gran cantidad de información sobre estadísticas del país, siendo su objetivo fundamental la integración del trabajo estadístico del país en un sistema eficiente, que sea capaz de responder con calidad a las necesidades de información del país para enfrentar las metas del progreso del desarrollo, tanto social como económico de la isla.

Situación problemática

La ONE fue creada el 21 de abril de 1994, como resultado de la reorganización de los organismos de la Administración Central del Estado. Su misión es garantizar la producción de estadísticas de calidad a través del Sistema Estadístico Nacional, ejerciendo una adecuada dirección, ejecución y control de la

captación de las cifras económicas y sociales, así como su adecuada difusión de acuerdo con las necesidades de la economía y las demás insuficiencias del país en información estadística.

La ONE es el órgano rector de la estadística en Cuba, tiene como objetivo fundamental captar, analizar y difundir los datos recogidos a lo largo y ancho de todo el país. Para ello cuenta con una serie de modelos estadísticos como es el caso del modelo 5200-03, en el cual se tiene almacenado el número de trabajadores por categoría ocupacional y sexo, y se recoge información de todos los sectores de la economía y la sociedad. Estos datos son almacenados en formatos de difícil acceso para su consulta, por lo que se hace muy complejo el proceso de acceso y divulgación de dicha información.

La forma de almacenar, recuperar y presentar la información proveniente del modelo 5200-03 dificulta la realización de los principales reportes y cruces de variables, indicadores, tasas, porcentajes y demás aspectos de interés; obstaculizando así la disponibilidad de información para los Órganos del Estado y afectando el proceso de toma de decisiones.

Problema a resolver

¿Cómo agilizar los procesos de extracción y consulta de los datos almacenados en el modelo 5200-03?

Objetivo general

Desarrollar el análisis, diseño e implementación del mercado de datos para los indicadores específicos de la ocupación para la ONE.

Objeto de estudio

Los almacenes de datos.

Objetivos específicos

- ✓ Elaborar el marco teórico de la investigación acerca de las principales tendencias de implementación de los mercados de datos.
- ✓ Realizar el estudio del estado del arte para los indicadores específicos de la ocupación.
- ✓ Realizar el análisis del modelo 5200-03 de la ONE.
- ✓ Diseñar el mercado de datos para los indicadores específicos de la ocupación para el almacén de datos de la ONE.

- ✓ Implementar y cargar los clasificadores para el mercado de datos para los indicadores específicos de la ocupación para el almacén de datos de la ONE.
- ✓ Validar la solución desarrollada mediante la realización de pruebas.

Tareas de la investigación:

- ✓ Definición de los conceptos relacionados con el desarrollo de mercados de datos.
- ✓ Selección de la herramienta de modelado, sistema gestor de base de datos y la metodología a utilizar en el desarrollo.
- ✓ Planificación y realización de las entrevistas.
- ✓ Identificación de la estructura de usuarios y permisos.
- ✓ Definición de los temas de análisis.
- ✓ Identificación de las necesidades de información, requisitos funcionales y no funcionales.
- ✓ Modelación de los requerimientos.
- ✓ Validación de los requerimientos.
- ✓ Definición de los requisitos de entrada y de salida.
- ✓ Elección de la granularidad del proceso del negocio.
- ✓ Definición de las dimensiones del mercado de datos.
- ✓ Definición de los hechos asociados a las dimensiones definidas.
- ✓ Estructuración del modelo dimensional.
- ✓ Transformación del modelo dimensional al diseño físico.
- ✓ Implementación de la base de datos.
- ✓ Montaje de los clasificadores para el mercado de datos para los indicadores específicos de la ocupación para el almacén de datos de la ONE.
- ✓ Realización de las pruebas al mercado de datos.

Campo de acción

Mercados de datos para los indicadores específicos de la ocupación.

Resultados esperados

Análisis, diseño e implementación del mercado de datos para los indicadores específicos de la ocupación y toda la documentación correspondiente al tema.

Idea a defender

Si se realiza un mercado de datos para los indicadores específicos de la ocupación se contribuirá a mejorar la realización de los principales reportes y cruces de variables, indicadores, tasas, porcentajes y demás aspectos de interés, facilitando la disponibilidad de información para los Órganos del Estado y el proceso de toma de decisiones.

Métodos científicos de investigación:

Teóricos:

Analítico - sintético: para estudiar los temas relacionados con el desarrollo del mercado de datos para los indicadores específicos de la ocupación, así como su evolución en la ONE, además de la metodología a utilizar en el desarrollo.

Análisis histórico - lógico: para el análisis de documentos, materiales, y temas relacionados a las mejores prácticas en el desarrollo de mercados de datos; además para definir la metodología a utilizar y las dimensiones del mercado de datos.

Modelación: para modelar el mercado de datos.

Empíricos:

Entrevistas: para obtener la información acerca de las necesidades de información, requisitos funcionales y no funcionales con los que debe cumplir la solución.

Estructura capitular:

Capítulo #1: Fundamentación teórica.

Se definen los temas relacionados con el desarrollo de los almacenes de datos, mercados de datos, las herramientas para el modelado, los sistemas gestores de bases de datos y las metodologías a utilizar en el desarrollo del mercado.

Capítulo #2: Análisis y diseño de la solución.

Se identifican las necesidades de información, así como los requerimientos funcionales y no funcionales, la modelación y validación de los mismos; además se define el hecho y las dimensiones del mercado de datos.

Capítulo #3: Implementación y prueba.

Se transforma el modelo dimensional, se selecciona la granularidad del proceso de negocio, se identifican los clasificadores y finalmente se implementa la base de datos, concluyendo la misma con un proceso de pruebas.

Capítulo 1 Fundamentación teórica

1.1 Introducción

Este capítulo abarca temas de la investigación relacionados con algunos conceptos básicos implicados en el desarrollo de los almacenes de datos, mercados de datos, gestores de bases de datos, herramientas y la metodología a utilizar para el desarrollo del mercado de datos para los indicadores específicos de la ocupación.

1.2 Análisis de los trabajos relacionados con el tema en el ámbito internacional y nacional

Diversas son las empresas del mundo que están interesadas en integrar y centralizar sus datos con el objetivo de reducir sus costos y adquirir una visión más completa sobre sus clientes, productos, riesgos y finanzas. Muchas de ellas son grandes empresas, reconocidas en todo el mundo, que manejan gran cantidad de información, lo que provoca que el uso de los almacenes de datos sea indispensable para lograr su correcto funcionamiento. Para el desarrollo de este trabajo se realiza un estudio de algunas de ellas que tratan varios indicadores, entre los que se encuentra el de ocupación.

1.2.1 INE1 en España

La organización estadística en España se crea para la producción de estadísticas oficiales de este país, pues su misión fundamental es la elaboración y perfección de los datos recogidos de distintos indicadores. Presentan un apartado de oferta pública, en el cual se recoge cualquier convocatoria de empleo, ya sea por contratación temporal o fija. La mayor complejidad de esta organización es la descentralización de la producción estadística dentro de la Administración General del Estado. La información de este INE se publica en su página web, donde presenta información de diversos indicadores, como es el caso del indicador sociedad, donde almacena datos sobre el mercado laboral. Recoge operaciones estadísticas elaboradas por el INE y por otros organismos. Se estructura por temas y apartados para que la localización de la misma sea de forma sencilla. La información estadística está permanentemente actualizada. (INE, 2010)

¹ INE: Instituto Nacional de Estadística.

1.2.2 INE2 en Chile

El INE es uno de los organismos públicos de mayor trayectoria en ese país. En la actualidad ha ampliado las áreas de trabajo. Entrega al país más de 70 indicadores de calidad, en una diversidad temática: empleo, precios, población, cultura, seguridad ciudadana, economía, entre otros. Es el organismo rector del SEN³, que integra a los principales productores y usuarios de este tipo de información. Su finalidad es coordinar y mejorar la pertinencia de la producción estadística, intercambiar puntos de vista y normar estándares comunes. El INE en Chile presenta la información a sus usuarios en un sistema que contiene datos relevantes. En su sección Microdatos presenta un apéndice que aborda directamente el tema del indicador de ocupación, permitiendo acceder a la base de datos y realizar cruces de variables. (INE, 2008)

Cuba hasta el momento no tiene implementado ningún sistema de mercado de datos o almacenes de datos que se encargue de recoger información estadística de este indicador.

Después de haber revisado estas posibles soluciones se determinó que ninguna satisface las necesidades requeridas para la realización del mercado de datos para los indicadores específicos de la ocupación para la ONE.

1.3 Almacenes de datos

La aparición de los ordenadores para la automatización de la información introdujo un cambio considerable para la gestión de la misma, desplazando el centro de informática hacia la estructuración de los datos, convirtiéndose las bases de datos en una herramienta esencial para el control y el manejo de las operaciones del comercio, debido a la necesidad de las empresas de disponer de gran cúmulo de información almacenada en diferentes fuentes de datos. A su vez, esta acumulación de información evidencia que podría ser de gran ayuda tener almacenada en un mismo lugar la mayoría de las operaciones.

Debido a la idea de unir las distintas fuentes de información en un lugar único para la futura introducción de la documentación relevante, y como respuesta a la misma, es que surgen los almacenes de datos.

Varios conocedores del tema plantean:

² INE: Instituto Nacional de Estadísticas.

³ SEN: Sistema Estadístico Nacional.

Capítulo 1: Fundamentación teórica

Datawarehouse es un conjunto de datos orientados por temas, integrados, variantes en el tiempo y no volátiles, que tienen por objetivo dar soporte a la toma de decisiones. (Inmon, 2006)

Datawarehouse es una copia de los datos transaccionales específicamente estructurada para la consulta y el análisis. (Kimball, 2006)

De forma general se puede decir que un almacén de datos es una base de datos que se caracteriza por la purificación e integración de la información recogida de diferentes fuentes para después procesarla.

Entre las características de los almacenes de datos se pueden mencionar las siguientes:

- ✓ Orientado hacia la información relevante de la organización: se diseña para consultar eficientemente información relativa a las actividades de la organización, no para soportar los procesos que se realizan en ella.
- ✓ Integrado: integra datos recogidos de diferentes sistemas operacionales de la organización.
- ✓ Variable en el tiempo: los datos son relativos a un período de tiempo y deben ser incrementados periódicamente.
- ✓ No volátil: los datos no son actualizados, únicamente son incrementados. (Hernández, 2003)

Estos sistemas poseen las siguientes ventajas:

- ✓ Rentabilidad de las inversiones realizadas para su creación.
- ✓ Aumento de la competitividad en el mercado.
- ✓ Aumento de la productividad de los técnicos de dirección. (Hernández, 2003)

Algunas de las desventajas de estos almacenes son:

- ✓ Infravaloración del esfuerzo necesario para su diseño y creación.
- ✓ Infravaloración de los recursos necesarios para la captura, carga y almacenamiento de los datos.
- ✓ Incremento continuo de los requisitos de los usuarios.
- ✓ Privacidad de los datos. (Hernández, 2003)

1.4 Mercados de datos

Un datamart es un contenedor de datos que ayuda en el proceso de toma de decisiones sobre experiencias pasadas. Su creación se debe específicamente a datos seleccionados por una necesidad concreta.

Entre las principales características están:

- ✓ Se centran en los requisitos de los usuarios asociados a un departamento o área de negocio específico.
- ✓ No contienen datos operacionales detallados.
- ✓ 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 la de los almacenes de datos. (Casales, 2009)

1.5 Modelos y modos de almacenamiento de los datos

Un modelo de almacenamiento de datos es un lenguaje orientado a describir la estructura de una base de datos, pues permite representar el orden de los datos, las restricciones de integridad y las operaciones de manipulación de los datos y los modos de almacenamiento de los datos permiten organizar la información contenida en la base de datos.

1.5.1 Modelos de almacenamiento de datos

Los modelos de datos se utilizan para describir las características y las relaciones que existen entre los datos, para su posterior manipulación.

Con el desarrollo de las bases de datos se han empleado modelos para realizar el diseño de las mismas, entre los cuales se pueden mencionar:

Modelo relacional: el diagrama de entidad-relación es un lenguaje para realizar el modelado de los datos de un sistema de información, se basa en la separación de los datos en entidades para formar parte del diseño físico. (Castillo, 2008)

Modelo dimensional: es uno de los más reconocidos en el mundo de los almacenes de datos, contiene la misma información que un modelo de entidad-relación, pero la forma de empaquetar los datos en un formato simétrico tiene como objetivo garantizar una ejecución eficiente y rápida de las consultas, así

como lograr una mayor comprensión del usuario. Dicho modelo separa sus datos en dos grandes tipos: las medidas, que generalmente son valores numéricos que se almacenan en las tablas de hechos, las cuales son las tablas primarias de este modelo; y las descripciones de los entornos, que son textuales y se almacenan en las tablas de dimensiones. (Verástegui, 2007)

Los hechos se pueden definir como objetos o variables a analizar, que se utilizan para realizar operaciones de agregación y arribar a conclusiones sobre un área o departamento en específico. La acción de llevar todos los hechos que se relacionan en el modelo a un nivel de desagregación es a lo que se le denomina granularidad. (Rubia, 2007)

Las dimensiones son los ejes por los que se caracteriza la información. La forma de agrupar los elementos de las dimensiones en un orden determinado dentro de las variables es a lo que se le denomina jerarquía. (Rubia, 2007)

Las jerarquías contienen sus atributos en una tabla única, tomando las dimensiones como puntas y la tabla de hecho como centro de una estrella. A esta modelación se le conoce como: esquema de estrella. (Rubia, 2007)

Este modelo posee diversas ventajas, y es que permite adicionar:

- ✓ Nuevos hechos a la tabla de hechos, siempre y cuando no afecten el nivel de granularidad del mismo.
 - ✓ Nuevas dimensiones y hechos aunque no se hayan tenido en cuenta antes.
 - ✓ Dimensiones, siempre y cuando no generen varios valores para cada fila en la tabla de hechos.
- (Rubia, 2007)

1.5.2 Modos de almacenamiento de datos

OLAP⁴ es una tecnología que permite organizar grandes volúmenes de bases de datos y la realización de complejos análisis de datos multidimensionales de manera más eficiente a los usuarios finales. Las bases de datos OLAP, están diseñadas para agilizar la recuperación de datos. (Luján-Mora, 2007)

Algunas de las características de OLAP son:

⁴ OLAP: Procesamiento Analítico en Línea. OnLine Analytical Processing, por sus siglas en inglés.

- ✓ Posibilita que el análisis multidimensional que se haga sea dinámico, de esta forma, permite que los usuarios finales puedan realizar actividades de navegación y analíticas.
- ✓ Brinda consultas rápidas, independientemente de la complejidad y del tamaño de la base de datos. (Luján-Mora, 2007)

Se utilizan varios modos de Procesamiento Analítico en Línea para almacenar los datos, entre los que se pueden mencionar:

ROLAP⁵: para proporcionar los análisis OLAP esta arquitectura accede a los datos almacenados en un datawarehouse. Los datos son acumulados en filas y columnas de forma relacional, y se les presentan a los usuarios en forma de dimensiones del negocio. ROLAP guarda la información en bases de datos relacionales, aprovechando así la tecnología relacional, permitiendo usar la integridad y seguridad de los sistemas gestores de bases de datos relacionales y gratuitos, y es capaz de manejar grandes volúmenes de datos. (Luján-Mora, 2007)

MOLAP⁶: usa bases de datos multidimensionales para almacenar los datos, presentando un mejor rendimiento que la tecnología relacional en el procesamiento de las consultas, ofrece una mayor flexibilidad y rapidez de acceso para realizar el análisis de los datos. (Luján-Mora, 2007)

HOLAP⁷: es una combinación de ambas arquitecturas, que recoge las mejores características de cada una de ellas. Este modelo posee dos tipos de particionamiento: el vertical y el horizontal. (Luján-Mora, 2007)

1.6 Sistemas gestores de bases de datos

Los SGBD⁸ son los que se encargan de manipular de forma sencilla, ordenada y clara, un conjunto de datos específicos relacionados con un tema y que pasan a formar parte de la información relevante para un buen manejo de los datos.

Los sistemas gestores de bases de datos se pueden clasificar en tres grupos teniendo en cuenta la disponibilidad en el mercado:

⁵ ROLAP: Procesamiento Analítico Relacional en Línea. Relational OnLine Analytical Processing, por sus siglas en inglés.

⁶ MOLAP Procesamiento Analítico Multidimensional en Línea. Multidimensional OnLine Analytical Processing, por sus siglas en inglés.

⁷ HOLAP: Procesamiento Analítico Híbrido en Línea. Hybrid Online Analytical Processing, por sus siglas en inglés.

⁸ SGBD: Sistema Gestor de Bases de Datos.

Capítulo 1: Fundamentación teórica

Libres, entre los que se encuentran: MySQL, PostgreSQL, Firebird, Salirte, DB2 Express-C y Apache Derby.

No libres, entre los que se encuentran: Advantage Database, dBase, FileMaker, Fox Pro, IBM DB2, IBM Informix, Internase de CodeGear, filial de Borland, MAGIC, Microsoft Access, Microsoft SQL Server, NexusDB, Open Access Oracle, Paradox, PervasiveSQL, Progress (DBMS), Sybase ASE, Sybase ASA Sybase IQ, WindowBase, IBM IMS y CA-IDMS.

No libres y gratuitos, entre los que se encuentran: Microsoft SQL Server Compact Edition Basica, Sybase ASE Express Edition para Linux (edición gratuita para Linux) y Oracle Express Edition 10.

A continuación se presentan algunos de estos sistemas con las más notables de sus características:

MySQL: es un sistema gestor de bases de datos relacionales, sencillo de usar y rápido, la empresa que lo desarrolla pertenece a Sun Microsystems que ahora forma parte de Oracle. Es un software de código abierto que consume pocos recursos, brinda mayor rendimiento y mejores utilidades de administración, se integra perfectamente con PHP, no tiene límites para el tamaño de los registros y permite un mejor control de los usuarios. Hace funcionar datawarehouse que contengan terabytes de información. En entornos de datawarehousing, este gestor ofrece diversas fortalezas para escalar hasta el rango de los terabytes, tanto para un único servidor como para varios. Gran parte de las utilidades de MySQL no están documentadas, lo que representa una desventaja para su utilización, además no permite realizar restricciones de integridad, procedimientos almacenados y vistas. (Toledo, 2008)

Microsoft SQL Server: es un sistema gestor de bases de datos relacionales que se basa en el lenguaje SQL, ofrece grandes cantidades de datos de manera simultánea a cierto número de usuarios, presenta gran estabilidad, seguridad y escalabilidad; soporta procedimientos almacenados e incluye además un potente entorno gráfico de administración, permitiendo administrar información de otros servidores de datos. Permite la creación de procedimientos almacenados, es decir, son instrucciones que se almacenan dentro de la base de datos. Posee servicios de transformación de datos, pues introduce un rediseño completo para proporcionar una plataforma de extracción, transformación y carga de los datos integral. Brinda servicios de reporting que posibilitan a los negocios integrar de forma sencilla datos desde fuentes heterogéneas y datawarehouses en informes ricos, interactivos y gestionables. Pero es válido destacar que el costo de las licencias comparadas con otros competidores es su desventaja, ya que varían

significativamente debido a la cantidad de servidores a utilizar y al número de usuarios conectados. (Bolaños, 2006)

Oracle: es un sistema gestor de bases de datos que se basa en la tecnología cliente/servidor, es un producto que se comercializa a nivel mundial, aunque tiene un precio muy elevado. Permite mejorar el rendimiento, la disponibilidad y la facilidad de gestión de los almacenes de datos, mediante la partición de las tablas grandes. Posibilita la búsqueda de información y patrones ocultos en el datawarehouse. Ejecuta los análisis más rápidos con vistas materializadas y tablas organizadas. Su principal desventaja se encuentra en la gran cantidad de memoria RAM⁹ que utiliza para la instalación y utilización del software. (Oracle, 2008)

1.6.1 Selección del gestor de bases de datos a utilizar

PostgreSQL: es un motor muy potente de bases de datos, su código fuente está disponible libremente, presenta funcionalidades y prestaciones similares a otros gestores, es más completo que MySQL, pues permite realizar restricciones de integridad, procedimientos almacenados y vistas materializadas.

Para el desarrollo del almacén se utiliza como sistema gestor PostgreSQL 8.4, por ser de código abierto, el más potente del mercado; el mismo proporciona un análisis más sencillo de los datos a través de funcionalidades avanzadas y mejora el rendimiento de las aplicaciones. Funciona muy bien con grandes cantidades de datos y permite una alta concurrencia de usuarios accediendo a la vez. (PostgreSQL, 2009)

Entre las mejoras más populares están:

- ✓ Restauración de base de datos en procesos paralelos, que acelera la recuperación de un respaldo hasta 8 veces.
- ✓ Privilegios por columna, que permiten un control más granular de datos confidenciales.
- ✓ Configuración de ordenamiento configurable por base de datos, lo cual hace a PostgreSQL más útil en entornos con múltiples idiomas.
- ✓ Nuevas herramientas para monitoreo de consultas que le otorgan a los administradores mayor información sobre la actividad del sistema. (PostgreSQL, 2009)

⁹ RAM: Memoria de Acceso Aleatorio. Random Access Memory, por sus siglas en inglés.

Estas mejoras permiten que los usuarios construyan almacenes de datos aún más grandes, pues admiten que los recorridos de tablas grandes se acarreen unos a otros, reduciendo enormemente la entrada/salida total requerida. (PostgreSQL, 2009)

1.7 Herramientas CASE

Las herramientas CASE¹⁰ son un conjunto de programas que brindan a los desarrolladores asistencia para el desarrollo del ciclo de vida de un proyecto. Con la introducción de estas herramientas se han podido confeccionar y modificar los diagramas que mejoran la calidad y productividad del software. Tienen como objetivo fundamental automatizar los aspectos clave en el proceso de desarrollo.

Dentro de las características más importantes de las herramientas CASE están:

- ✓ Alto rendimiento.
- ✓ Calidad de software desarrollado.
- ✓ Archivos de datos de conocimiento.
- ✓ Facilidades de uso. (Herramientas CASE y RAD¹¹, 2006)

Entre las herramientas CASE más utilizadas se pueden mencionar:

EasyCASE: herramienta que permite automatizar dentro del desarrollo de una aplicación las fases de análisis y diseño, trabaja para proveer soluciones comprensibles para esta última. Permite la comunicación de las ideas gráficamente para que sean fáciles de comprender y posibilita la captura de los detalles de un sistema. No permite la especificación de los elementos borrados y modificados, lo cual constituye una desventaja. (Domínguez y Ligeró, 2006)

Oracle Designer: es un juego de herramientas que posibilita la rápida construcción de aplicaciones flexibles y gráficas. Guarda todas las definiciones que necesita el usuario para la rápida construcción de aplicaciones. Habilita la captura del diseño de los sistemas existentes. El mayor inconveniente es su alto precio. (Oracle, 2003)

Erwin: herramienta para el diseño de las bases de datos que brinda gran productividad en su diseño, permite ver elementos importantes y la estructura de la base de datos, generando de forma automática las

¹⁰ CASE: Herramientas de Ingeniería de Software Asistida por Computación. Computer Aided Software Engineering, por sus siglas en inglés.

¹¹ RAD: Desarrollo Rápido de Aplicaciones. Rapid Application Development, por sus siglas en inglés.

tablas, haciendo posible la transferencia de datos entre un par de ellas. No visualiza los nombres de las interrelaciones del modelo, lo cual evidencia un inconveniente para su uso. (Danysoft, 2007)

ER/Studio: es una herramienta para el modelado de las bases de datos fácil de usar, ofrece capacidades de diseño lógico y la construcción automática de las bases de datos, posee abundante documentación y fácil creación de los reportes. La realización de los diagramas es rápida y clara, permitiendo la documentación del mismo. ER/Studio permite realizar cambios en los datos del modelo de diseño de forma directa a la base de datos, pero no es una herramienta libre, característica que dificulta su utilización en la solución. (Embarcadero, 2005)

1.7.1 Selección de las herramientas CASE a utilizar

Se utiliza para modelar el negocio el Visual Paradigm en su versión 6.4, pues soporta el ciclo de vida completo del desarrollo del software y para generar un producto de mayor calidad el diseño se centra en los casos de uso y se enfoca en el negocio. Hay que señalar su robustez, usabilidad y portabilidad. Está diseñada para dar soporte a arquitectos de sistemas, diseñadores, desarrolladores, analistas de procesos de negocio y modeladores de datos en los procesos de desarrollo de software. La Universidad de las Ciencias informáticas enseña la asignatura de Ingeniería de Software mediante el empleo de esta herramienta luego de haber pagado la licencia de la misma. (Giraldo y Zapata ,2005)

1.8 Metodologías de desarrollo

Muchos han sido los estudios realizados sobre las metodologías para contribuir en la elaboración de los sistemas, para lo cual se realiza un estudio minucioso acerca de las metodologías a aplicar en el desarrollo de un proyecto productivo, conociendo que estas brindan un conjunto de documentos que ayudan a los desarrolladores para realizar un nuevo software y a perfeccionar los que están ya implementados.

Son diversas las metodologías que se pueden utilizar para el desarrollo de soluciones de almacenes de datos, y muchas las diferencias entre una y otra para realizar la selección de la correcta, e incluso a veces se definen nuevas metodologías a partir de una selección de las características esenciales de algunas de ellas.

A continuación se describen brevemente cada una de estas metodologías:

Capítulo 1: Fundamentación teórica

Metodología SQLBI¹²: está orientada totalmente a sus herramientas Microsoft SQL Server, SQL Server Analysis Services, avalada por la Microsoft y en este campo su oferta más completa es Microsoft Suite for BI. (Dataprix, 2008)

Metodología DM2¹³: se basa en las necesidades que existan de información a nivel gerencial, donde la misma debe ser accesible únicamente a quien la necesite, acorta de manera razonable el tiempo que existe entre el análisis y la implementación, esta rapidez es buena para el cliente y también es necesaria y exigida por el propio ambiente que lo rodea. (Ideasoft, 2008)

Metodología CRISP-DM¹⁴: herramienta industrial y aplicación neutral propuesta en 1996, está descrita en un modelo de proceso jerárquico, sus tareas se describen en 4 niveles de abstracción; va de lo más general a lo específico. (Dataprix, 2006)

Metodología Hefestos: plantea que la construcción e implementación de un almacén de datos se puede adaptar al ciclo de vida de desarrollo de software, aunque algunas fases tienen particularidades en las acciones a realizar, buscando entregar una primera parte de la implementación que satisfaga las necesidades, y manteniendo motivados a los usuarios. (Dataprix, 2008)

Metodología para el Diseño Conceptual de Almacenes de Datos: tiene como aspecto novedoso la incorporación de los casos de uso para guiar el proceso de desarrollo, definiendo además varias transformaciones para llevar del modelo relacional al dimensional, obteniendo las estructuras que posteriormente conforman el repositorio de datos. (Zenaido, 2008)

Metodología Kimball: se centra esencialmente en el diseño de la base de datos y este a su vez se enfoca en la creación de las tablas de hechos. Esta metodología tiene gran aceptación en el mundo para el desarrollo de una solución de Inteligencia de Negocios, permite que su solución se realice en un período corto de tiempo. (Inteligencia de negocios, 2007)

¹² SQLBI: Firma consultora con una experiencia fuerte y única en Inteligencia de Negocios, en torno a las tecnologías de Microsoft SQL Server.

¹³ DM: Minería de Datos. Data Mining, por sus siglas en inglés.

¹⁴ CRISP-DM: Interprofesionales Norma Proceso de Minería de Datos. CRoss-Industry Standard Process for Data Mining, por sus siglas en inglés.

1.8.1 Selección de la metodología a utilizar

Se tomó como base la Metodología Kimball para definir la que se va a utilizar en la Línea Almacenes de Datos y BI de DATEC¹⁵, pues la misma reúne los siguientes elementos:

- ✓ Facilita el proceso de toma de decisiones, lo cual hace que muestre lo indudablemente eficaz que es, proporcionando una mayor agilidad en el proceso de desarrollo y la creación de los conceptos de hechos y dimensiones.
- ✓ A través de la construcción de los mercados de datos departamentales propone ir construyendo el almacén de datos, lo cual coincide con la división lógica de los organismos, entidades y empresas, lo que a su vez constituye una buena estrategia.
- ✓ El uso de esta metodología facilita el trabajo de los desarrolladores, porque cuenta con una abundante documentación, pudiendo aclarar todas las dudas y preguntas a través de la web, pues el grupo creador de la metodología brinda muy buenos servicios.
- ✓ Esta metodología es muy reconocida por el resto de la comunidad que se dedica a este tema. Tiene claramente definidos los roles, actividades, artefactos y etapas de desarrollo.

Como complemento a la misma y fortaleciendo la etapa del levantamiento de requisitos, se tomó lo planteado por Leopoldo Zenaído Zepeda Sánchez en su tesis de doctorado, orientando así el trabajo a los casos de uso, lográndose estar más alineado con las tendencias y normas de la universidad.

En el ciclo de vida tienen lugar los siguientes flujos de trabajo:

- ✓ Estudio preliminar o planeación.
- ✓ Requerimientos.
- ✓ Arquitectura y diseño.
- ✓ Implementación.
- ✓ Prueba.
- ✓ Despliegue.

¹⁵ DATEC: Centro de Tecnología de Datos.

- ✓ Soporte y mantenimiento.
- ✓ Gestión y administración del proyecto.

La línea de almacenes tiene una estructura conformada por cuatro grupos de trabajo, donde cada uno realiza actividades específicas y bien delimitadas según sus responsabilidades dentro de un proyecto. Independiente a estos grupos, para el desarrollo de un proyecto, también se conforma otro que tiene como tarea fundamental la dirección del mismo. Cada grupo se especializa en un conjunto de actividades generales que contienen actividades específicas que tributan al desarrollo del proyecto, generan sus propios artefactos y se dividen por roles para darle cumplimiento a todas las tareas.

Esta metodología está estructurada por grupos:

- ✓ Grupo de análisis.
- ✓ Grupo de DWH¹⁶.
- ✓ Grupo de ETL¹⁷.
- ✓ Grupo de BI.
- ✓ Grupo de dirección.

1.9 Conclusiones del capítulo

Al realizar un estudio detallado acerca de los principales conceptos relacionados con el tema a desarrollar y teniendo en cuenta el creciente desarrollo de la tecnología, se ha decidido desarrollar un mercado de datos utilizando como herramienta de modelado el Visual Paradigm en su versión 6.4, debido a las características que presenta; por las nombradas mejoras que posee se usa como gestor de bases de datos PostgreSQL en su versión 8.4 y la metodología para el desarrollo de Soluciones de Almacenes de Datos y BI de DATEC.

¹⁶ DWH: Almacenes de Datos. Datawarehouse, por sus siglas en inglés.

¹⁷ ETL: Extracción, Transformación y Carga. Extraction, Transformation and Load, por sus siglas en inglés.

Capítulo 2

Análisis y diseño de la solución

2.1 Introducción

En este capítulo se definen temas de interés para la correcta comprensión del negocio, dándose a conocer las principales necesidades de información, así como los requerimientos funcionales y no funcionales, además de los requisitos de información y los multidimensionales, realizándose la modelación y validación de los mismos. De igual manera se definen los hechos y dimensiones del mercado de datos.

2.2 Análisis

Se realiza un estudio de los principales aspectos a tener en cuenta para lograr una mejor comprensión del negocio.

2.2.1 Definición del negocio

La ONE es la entidad creada para proponer, ejecutar y organizar, según corresponda, la aplicación de la política estatal en materia de estadísticas del país. La misma está delimitada por varias áreas que se encargan de recopilar, analizar y difundir la información recogida a lo largo y ancho de todo el país. Para ello cuenta con el área de Estadísticas sociales, que comprende el departamento que lleva el mismo nombre, en el cual se recogen todos los datos sobre los ocupados, permitiendo conocer la ocupación total y por sexo del país (sector Estatal Civil y no Estatal) por territorio puro, así como su desglose por tipos de contratos de trabajo, mayores y menores de la edad laboral que trabajan, categoría ocupacional y sexo.

2.2.2 Temas de análisis

El tema de análisis a desarrollar en la solución es el comportamiento estadístico de los indicadores específicos de la ocupación.

2.2.3 Roles y permisos

En la ONE no necesariamente el personal que trabaja en el área de ocupación tiene acceso a la información. Por lo que se definen los siguientes roles y permisos:

- ✓ Administrador: es el encargado de trabajar en la extracción, transformación y carga de los datos. Tiene permiso de lectura y escritura.
- ✓ Analista: es el encargado de analizar la información y hace uso de la misma para los cálculos estadísticos. Tiene permiso de lectura.

El nivel de acceso es público, ya que no es estrictamente necesario que los usuarios estén dentro de estas clasificaciones para acceder a la información.

2.2.4 Reglas del negocio

A través de la información almacenada en el modelo es de fácil consulta realizar una serie de cálculos como los que se mencionan a continuación:

- ✓ La cantidad de trabajadores del sexo masculino, se obtiene mediante la diferencia entre el total de trabajadores y la cantidad de mujeres.
- ✓ La cantidad de trabajadores contratados por tiempo determinado del sexo masculino, se obtiene mediante la diferencia entre el total de trabajadores contratados por tiempo determinado y la cantidad de mujeres contratadas por tiempo determinado.
- ✓ La cantidad de trabajadores mayores de la edad laboral del sexo masculino, se obtiene mediante la diferencia entre el total de trabajadores mayores de la edad laboral y la cantidad de mujeres mayores de la edad laboral.
- ✓ La cantidad de trabajadores menores de la edad laboral del sexo masculino, se obtiene mediante la diferencia entre el total de trabajadores menores de la edad laboral y la cantidad de mujeres menores de la edad laboral.
- ✓ La cantidad de trabajadores operarios del sexo masculino, se obtiene mediante la diferencia entre el total de trabajadores operarios y la cantidad de mujeres operarias.
- ✓ La cantidad de trabajadores técnicos del sexo masculino, se obtiene mediante la diferencia entre el total de trabajadores técnicos y la cantidad de mujeres técnicas.
- ✓ La cantidad de trabajadores administrativos del sexo masculino, se obtiene mediante la diferencia entre el total de trabajadores administrativos y la cantidad de mujeres administrativas.

Capítulo 2: Análisis y diseño de la solución

- ✓ La cantidad de trabajadores de servicios del sexo masculino, se obtiene mediante la diferencia entre el total de trabajadores de servicios y la cantidad de mujeres de servicios.
- ✓ La cantidad de trabajadores dirigentes del sexo masculino, se obtiene mediante la diferencia entre el total de trabajadores dirigentes y la cantidad de mujeres dirigentes.
- ✓ El por ciento de los trabajadores del sexo femenino que están ocupados, se obtiene mediante la división de la cantidad de mujeres y el total de trabajadores por 100.
- ✓ El por ciento de los trabajadores del sexo masculino que están ocupados, se obtiene mediante la diferencia entre el total de trabajadores y la cantidad de mujeres / el total de trabajadores por 100.
- ✓ El por ciento de los trabajadores operarios, se obtiene mediante la división entre el total de trabajadores operarios y el total de trabajadores por 100.
- ✓ El por ciento de los trabajadores operarios del sexo femenino, se obtiene mediante la división entre cantidad de mujeres operarias y el total de trabajadores por 100.
- ✓ El por ciento de los trabajadores operarios del sexo masculino, se obtiene mediante la diferencia entre el total de trabajadores operarios y el total de trabajadores operarios de sexo femenino / el total de trabajadores por 100.
- ✓ El por ciento de los trabajadores técnicos, se obtiene mediante la división entre el total de trabajadores técnicos y el total de trabajadores por 100.
- ✓ El por ciento de los trabajadores técnicos del sexo femenino, se obtiene mediante la división entre la cantidad de técnicos de sexo femenino y el total de trabajadores por 100.
- ✓ El por ciento de los trabajadores técnicos del sexo masculino, se obtiene mediante la diferencia entre el total de trabajadores técnicos y la cantidad de mujeres técnicas / el total de trabajadores por 100.
- ✓ El por ciento de los trabajadores administrativos, se obtiene mediante la división entre el total de trabajadores administrativos y el total de trabajadores por 100.
- ✓ El por ciento de los trabajadores administrativos del sexo femenino, se obtiene mediante la división entre la cantidad de mujeres que ocupan cargos administrativos y el total de trabajadores por 100.

- ✓ El por ciento de los trabajadores administrativos del sexo masculino, se obtiene mediante la diferencia entre el total de trabajadores administrativos y la cantidad de mujeres que ocupan cargos administrativos / el total de trabajadores por 100.
- ✓ El por ciento de los trabajadores de servicios, se obtiene mediante la división entre el total de trabajadores de servicios y el total de trabajadores por 100.
- ✓ El por ciento de los trabajadores de servicios del sexo femenino, se obtiene mediante la división entre el total de mujeres que ocupan cargos de servicios y el total de trabajadores por 100.
- ✓ El por ciento de los trabajadores de servicios del sexo masculino, se obtiene mediante la diferencia entre el total de trabajadores de servicios y el total de mujeres que ocupan cargos de servicios / el total de trabajadores por 100.
- ✓ El por ciento de los trabajadores dirigentes, se obtiene mediante la división entre el total de trabajadores dirigentes y el total de trabajadores por 100.
- ✓ El por ciento de los trabajadores dirigentes del sexo femenino, se obtiene mediante la división entre el total de mujeres que ocupan cargos de dirigentes y el total de trabajadores por 100.
- ✓ El por ciento de los trabajadores dirigentes del sexo masculino, se obtiene mediante la diferencia entre el total de trabajadores dirigentes y el total de mujeres que ocupan cargos de dirigentes / el total de trabajadores por 100.

2.2.5 Necesidades de los usuarios

Son las especificidades que necesita conocer el usuario para generar posteriormente los reportes. Entre estas necesidades se pueden mencionar las siguientes: la cantidad total de trabajadores, los que se contrataron por un tiempo determinado, así como los trabajadores mayores y menores de la edad laboral, los trabajadores por categoría ocupacional y sexo.

2.2.6 Requisitos

Los requisitos o requerimientos son los que de forma general especifican lo que debe hacer la solución.

2.2.6.1 Requisitos funcionales

Los requisitos funcionales son las condiciones o capacidades que la solución debe cumplir. La solución en cuestión debe permitir realizar dichas funciones:

R1- Extraer información del archivo DBF¹⁸ del número de trabajadores por categoría ocupacional y sexo.

R2- Transformar y cargar la información del archivo DBF del número de trabajadores por categoría ocupacional y sexo.

2.2.6.2 Requisitos no funcionales

Los requisitos no funcionales son las cualidades o propiedades que la solución debe tener.

✓ Disponibilidad:

RNF1- El sistema debe estar disponible el tiempo que se especifique y en óptimas condiciones para su uso.

✓ Confiabilidad:

RNF1- El sistema debe ser capaz de mantener la integridad de los datos.

RNF2- La herramienta de implementación debe tener soporte para la recuperación de los datos ante algún fallo.

✓ Fiabilidad:

RNF1- El acceso a la información debe estar disponible el tiempo especificado y se tienen en cuenta los permisos establecidos.

RNF2- Al sistema se le dará mantenimiento al menos una vez al mes.

✓ Eficiencia:

RNF1- El tiempo de respuesta de las consultas debe ser rápido.

¹⁸ DBF: extensión del fichero.

RNF2- El sistema debe soportar varios usuarios conectados a la vez, sin afectar el tiempo de respuesta del sistema.

- ✓ Portabilidad:

RNF1- El sistema podrá funcionar en cualquier sistema operativo, pues el SGBD es multiplataforma.

- ✓ Rendimiento de las consultas y de la carga:

RNF1- El tiempo de respuesta de las consultas debe ser rápido, al igual que el de la carga de los datos, pero manteniendo la integridad y consistencia de los mismos.

2.2.6.3 Requisitos de información

Los requerimientos de información les brindan a los usuarios determinadas informaciones de interés, como es el caso de obtener información sobre el número de trabajadores ocupados por categoría ocupacional y sexo, que se desglosa en los siguientes requisitos:

R1- Obtener la cantidad total de trabajadores.

R2- Obtener la cantidad de trabajadores del sexo femenino.

R3- Obtener la cantidad de trabajadores del sexo masculino.

R4- Obtener la cantidad total de trabajadores contratados por tiempo determinado.

R5- Obtener la cantidad de trabajadores contratados por tiempo determinado del sexo femenino.

R6- Obtener la cantidad de trabajadores contratados por tiempo determinado del sexo masculino.

R7- Obtener la cantidad total de trabajadores mayores de la edad laboral.

R8- Obtener la cantidad de trabajadores mayores de la edad laboral del sexo femenino.

R9- Obtener la cantidad de trabajadores mayores de la edad laboral del sexo masculino.

R10- Obtener la cantidad total de trabajadores menores de la edad laboral.

R11- Obtener la cantidad de trabajadores menores de la edad laboral del sexo femenino.

R12- Obtener la cantidad de trabajadores menores de la edad laboral del sexo masculino.

Capítulo 2: Análisis y diseño de la solución

- R13- Obtener la cantidad total de trabajadores operarios.
- R14- Obtener la cantidad de trabajadores operarios del sexo femenino.
- R15- Obtener la cantidad de trabajadores operarios del sexo masculino.
- R16- Obtener la cantidad total de trabajadores técnicos.
- R17- Obtener la cantidad de trabajadores técnicos del sexo femenino.
- R18- Obtener la cantidad de trabajadores técnicos del sexo masculino.
- R19- Obtener la cantidad total de trabajadores administrativos.
- R20- Obtener la cantidad de trabajadores administrativos del sexo femenino.
- R21- Obtener la cantidad de trabajadores administrativos del sexo masculino.
- R22- Obtener la cantidad total de trabajadores de servicios.
- R23- Obtener la cantidad de trabajadores de servicios del sexo femenino.
- R24- Obtener la cantidad de trabajadores de servicios del sexo masculino.
- R25- Obtener la cantidad total de trabajadores dirigentes.
- R26- Obtener la cantidad de trabajadores dirigentes del sexo femenino.
- R27- Obtener la cantidad de trabajadores dirigentes del sexo masculino.
- R28- Obtener el por ciento de los trabajadores del sexo femenino que están ocupados.
- R29- Obtener el por ciento de los trabajadores del sexo masculino que están ocupados.
- R30- Obtener el por ciento de los trabajadores operarios.
- R31- Obtener el por ciento de los trabajadores operarios del sexo femenino.
- R32- Obtener el por ciento de los trabajadores operarios del sexo masculino.
- R33- Obtener el por ciento de los trabajadores técnicos.
- R34- Obtener el por ciento de los trabajadores técnicos del sexo femenino.

R35- Obtener el por ciento de los trabajadores técnicos del sexo masculino.

R36- Obtener el por ciento de los trabajadores administrativos.

R37- Obtener el por ciento de los trabajadores administrativos del sexo femenino.

R38- Obtener el por ciento de los trabajadores administrativos del sexo masculino.

R39- Obtener el por ciento de los trabajadores de servicios.

R40- Obtener el por ciento de los trabajadores de servicios del sexo femenino.

R41- Obtener el por ciento de los trabajadores de servicios del sexo masculino.

R42- Obtener el por ciento de los trabajadores dirigentes.

R43- Obtener el por ciento de los trabajadores dirigentes del sexo femenino.

R44- Obtener el por ciento de los trabajadores dirigentes del sexo masculino.

2.2.6.4 Requisitos multidimensionales

Se caracterizan por poseer variables de entrada y salida para la obtención de la información.

A continuación se introducen los requerimientos con sus respectivas variables de entrada y salida.

R1- Obtener la cantidad total de trabajadores.

VE¹⁹: contratados por tiempo determinado y por edad y categoría ocupacional.

VS²⁰: cantidad total de trabajadores.

R2- Obtener la cantidad de trabajadores del sexo femenino.

VE: contratados por tiempo determinado y por edad y categoría ocupacional.

VS: cantidad de trabajadores del sexo femenino.

R3- Obtener la cantidad de trabajadores del sexo masculino.

VE: contratados por tiempo determinado y por edad y categoría ocupacional.

¹⁹ VE: Variables de Entrada.

²⁰ VS: Variables de Salida.

Capítulo 2: Análisis y diseño de la solución

VS: cantidad de trabajadores del sexo masculino.

R4- Obtener la cantidad total de trabajadores contratados por tiempo determinado.

VE: contratados por tiempo determinado y por edad.

VS: cantidad total de trabajadores contratados por tiempo determinado.

R5- Obtener la cantidad de trabajadores contratados por tiempo determinado del sexo femenino.

VE: contratados por tiempo determinado y por edad.

VS: cantidad de trabajadores contratados por tiempo determinado del sexo femenino.

R6- Obtener la cantidad de trabajadores contratados por tiempo determinado del sexo masculino.

VE: contratados por tiempo determinado y por edad.

VS: cantidad de trabajadores contratados por tiempo determinado del sexo masculino.

R7- Obtener la cantidad total de trabajadores mayores de la edad laboral.

VE: contratados por tiempo determinado y por edad.

VS: cantidad total de trabajadores mayores de la edad laboral que trabajan.

R8- Obtener la cantidad de trabajadores mayores de la edad laboral del sexo femenino.

VE: contratados por tiempo determinado y por edad.

VS: cantidad de trabajadores mayores de la edad laboral del sexo femenino.

R9- Obtener la cantidad de trabajadores mayores de la edad laboral del sexo masculino.

VE: contratados por tiempo determinado y por edad.

VS: cantidad de trabajadores mayores de la edad laboral que trabajan del sexo masculino.

R10- Obtener la cantidad total de trabajadores menores de la edad laboral.

VE: contratados por tiempo determinado y por edad.

VS: cantidad total de trabajadores menores de la edad laboral.

Capítulo 2: Análisis y diseño de la solución

R11- Obtener la cantidad de trabajadores menores de la edad laboral del sexo femenino.

VE: contratados por tiempo determinado y por edad.

VS: cantidad de trabajadores menores de la edad laboral del sexo femenino.

R12- Obtener la cantidad de trabajadores menores de la edad laboral del sexo masculino.

VE: contratados por tiempo determinado y por edad.

VS: cantidad de trabajadores menores de la edad laboral del sexo masculino.

R13- Obtener la cantidad total de trabajadores operarios.

VE: categoría ocupacional.

VS: cantidad total de trabajadores operarios.

R14- Obtener la cantidad de trabajadores operarios del sexo femenino.

VE: categoría ocupacional.

VS: cantidad de trabajadores operarios del sexo femenino.

R15- Obtener la cantidad de trabajadores operarios del sexo masculino.

VE: categoría ocupacional.

VS: cantidad de trabajadores operarios del sexo masculino.

R16- Obtener la cantidad total de trabajadores técnicos.

VE: categoría ocupacional.

VS: cantidad total de trabajadores técnicos.

R17- Obtener la cantidad de trabajadores técnicos del sexo femenino.

VE: categoría ocupacional.

VS: cantidad de trabajadores técnicos del sexo femenino.

R18- Obtener la cantidad de trabajadores técnicos del sexo masculino.

Capítulo 2: Análisis y diseño de la solución

VE: categoría ocupacional.

VS: cantidad de trabajadores técnicos del sexo masculino.

R19- Obtener la cantidad total de trabajadores administrativos.

VE: categoría ocupacional.

VS: cantidad total de trabajadores administrativos.

R20- Obtener la cantidad de trabajadores administrativos del sexo femenino.

VE: categoría ocupacional.

VS: cantidad de trabajadores administrativos del sexo femenino.

R21- Obtener la cantidad de trabajadores administrativos del sexo masculino.

VE: categoría ocupacional.

VS: cantidad de trabajadores administrativos del sexo masculino.

R22- Obtener la cantidad total de trabajadores de servicios.

VE: categoría ocupacional.

VS: cantidad total de trabajadores de servicios.

R23- Obtener la cantidad de trabajadores de servicios del sexo femenino.

VE: categoría ocupacional.

VS: cantidad de trabajadores de servicios del sexo femenino.

R24- Obtener la cantidad de trabajadores de servicios del sexo masculino.

VE: categoría ocupacional.

VS: cantidad de trabajadores de servicios del sexo masculino.

R25- Obtener la cantidad total de trabajadores dirigentes.

VE: categoría ocupacional.

Capítulo 2: Análisis y diseño de la solución

VS: cantidad total de trabajadores dirigentes.

R26- Obtener la cantidad de trabajadores dirigentes del sexo femenino.

VE: categoría ocupacional.

VS: cantidad de trabajadores dirigentes del sexo femenino.

R27- Obtener la cantidad de trabajadores dirigentes del sexo masculino.

VE: categoría ocupacional.

VS: cantidad de trabajadores dirigentes del sexo masculino.

R28- Obtener el por ciento de los trabajadores del sexo femenino que están ocupados.

VE: contratados por tiempo determinado y por edad y categoría ocupacional.

VS: por ciento de los trabajadores del sexo femenino que están ocupados.

R29- Obtener el por ciento de los trabajadores del sexo masculino que están ocupados.

VE: contratados por tiempo determinado y por edad y categoría ocupacional.

VS: por ciento de los trabajadores del sexo masculino que están ocupados.

R30- Obtener el por ciento de los trabajadores operarios.

VE: categoría ocupacional.

VS: por ciento de los trabajadores operarios.

R31- Obtener el por ciento de los trabajadores operarios del sexo femenino.

VE: categoría ocupacional.

VS: por ciento de los trabajadores operarios del sexo femenino.

R32- Obtener el por ciento de los trabajadores operarios del sexo masculino.

VE: categoría ocupacional.

VS: por ciento de los trabajadores operarios del sexo masculino.

Capítulo 2: Análisis y diseño de la solución

R33- Obtener el por ciento de los trabajadores técnicos.

VE: categoría ocupacional.

VS: por ciento de los trabajadores técnicos.

R34- Obtener el por ciento de los trabajadores técnicos del sexo femenino.

VE: categoría ocupacional.

VS: por ciento de los trabajadores técnicos del sexo femenino.

R35- Obtener el por ciento de los trabajadores técnicos del sexo masculino.

VE: categoría ocupacional.

VS: por ciento de los trabajadores técnicos del sexo masculino.

R36- Obtener el por ciento de los trabajadores administrativos.

VE: categoría ocupacional.

VS: por ciento de los trabajadores administrativos.

R37- Obtener el por ciento de los trabajadores administrativos del sexo femenino.

VE: categoría ocupacional.

VS: por ciento de los trabajadores administrativos del sexo femenino.

R38- Obtener el por ciento de los trabajadores administrativos del sexo masculino.

VE: categoría ocupacional.

VS: por ciento de los trabajadores administrativos del sexo masculino.

R39- Obtener el por ciento de los trabajadores de servicios.

VE: categoría ocupacional.

VS: por ciento de los trabajadores de servicios.

R40- Obtener el por ciento de los trabajadores de servicios del sexo femenino.

VE: categoría ocupacional.

VS: por ciento de los trabajadores de servicios del sexo femenino.

R41- Obtener el por ciento de los trabajadores de servicios del sexo masculino.

VE: categoría ocupacional.

VS: por ciento de los trabajadores de servicios del sexo masculino.

R42- Obtener el por ciento de los trabajadores dirigentes.

VE: categoría ocupacional.

VS: por ciento de los trabajadores dirigentes.

R43- Obtener el por ciento de los trabajadores dirigentes del sexo femenino.

VE: categoría ocupacional.

VS: por ciento de los trabajadores dirigentes del sexo femenino.

R44- Obtener el por ciento de los trabajadores dirigentes del sexo masculino.

VE: categoría ocupacional.

VS: por ciento de los trabajadores dirigentes del sexo masculino.

2.2.7 Casos de uso del sistema

Un caso de uso es una secuencia de transacciones que son desarrolladas por un sistema en respuesta a un evento que inicia un actor sobre el propio sistema.

2.2.7.1 Casos de uso funcionales

Extraer información del archivo DBF del número de trabajadores por categoría ocupacional y sexo.

Transformar y cargar información del archivo DBF del número de trabajadores por categoría ocupacional y sexo.

2.2.7.2 Caso de uso informativo

Obtener información sobre el número de trabajadores ocupados por categoría ocupacional y sexo.

2.2.8 Diagrama de casos de uso

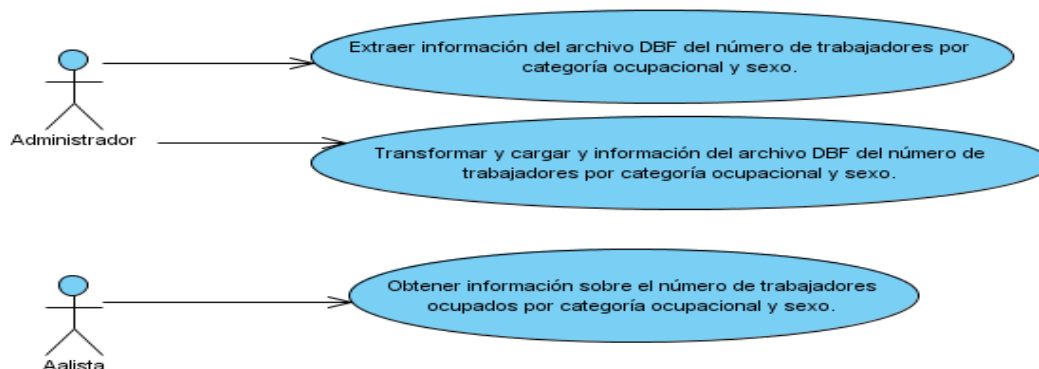


Figura 1: Diagrama de casos de uso del sistema

2.2.9 Descripción textual de los casos de uso

Tabla 1: Descripción del caso de uso Extraer información del archivo DBF del número de trabajadores por categoría ocupacional y sexo

Caso de uso:	Extraer información del archivo DBF del número de trabajadores por categoría ocupacional y sexo.
Tipo:	Funcional
Actores:	Administrador
Resumen:	El administrador inicia el caso de uso para extraer información del archivo DBF del número de trabajadores por categoría ocupacional y sexo.
Precondiciones:	Disponibilidad de las fuentes
Referencias	R1
Prioridad	Crítico
Complejidad	Alta

Flujo normal de eventos	
Sección “Extraer información”	
Acción del actor	Respuesta de la fuente de ocupación
1. Realizar la conexión al archivo DBF del número de trabajadores por categoría ocupacional y sexo.	2. Responde a la solicitud.
3. Seleccionar estructuras o archivos a extraer.	4. Muestra estructuras o archivos a extraer.
5. Realizar la extracción.	6. Actualiza los datos.
7. Termina la conexión.	
Prototipo de interfaz	
Flujos alternos	
Acción del actor	Respuesta del sistema
	1. No responde a la solicitud.
2. En caso de error se le notifica al administrador del DBF del número de trabajadores por categoría ocupacional y sexo.	3. Notifica al administrador.
4. En el control de cambio se verifica si hay modificaciones, en caso de haberlas regresa al paso 3 del flujo normal, en caso contrario regresa al 7.	5. Verifica modificaciones.

Capítulo 2: Análisis y diseño de la solución

Prototipo de Interfaz

Poscondiciones	Datos del archivo DBF disponibles para ser transformados.
-----------------------	-----------------------------------------------------------

Tabla 2: Descripción del caso de uso Transformar y cargar información del archivo DBF del número de trabajadores por categoría ocupacional y sexo

Caso de uso:	Transformar y cargar información del archivo DBF del número de trabajadores por categoría ocupacional y sexo.
Tipo:	Funcional
Actores:	Administrador
Resumen:	El administrador inicia el caso de uso para transformar y cargar información del archivo DBF del número de trabajadores por categoría ocupacional y sexo.
Precondiciones:	Extracción completada en área temporal. Estructuras del almacén disponibles.
Referencias	R2
Prioridad	Crítico
Complejidad	Alta

Flujo normal de eventos

Sección "Transformar y cargar información"

Acción del actor	Respuesta de la fuente de ocupación
1. Seleccionar la estructura a transformar.	2. Responde a la solicitud.
3. Cargar los datos seleccionados.	4. Muestra los datos seleccionados.

Capítulo 2: Análisis y diseño de la solución

5. Aplicar transformaciones y generar los datos.	6. Actualiza los datos.
7. Cargar los datos en el almacén.	8. Carga los datos.
9. Actualizar las tablas de control de cambio. (Opcional)	
10. Actualizar las vistas materializadas. (Opcional)	
<i>Prototipo de interfaz</i>	
Flujos alternos	
Acción del actor	Respuesta del sistema
	1. No responde a la solicitud.
2. En caso de error se le notifica al administrador del DBF del número de trabajadores por categoría ocupacional y sexo.	3. Notifica al administrador.
4. En el control de cambio se verifica si hay modificaciones, en caso de haberlas regresa al paso 3 del flujo normal, en caso contrario regresa al 7.	5. Verifica modificaciones.
<i>Prototipo de interfaz</i>	
Poscondiciones	Datos del archivo DBF del número de trabajadores por categoría ocupacional y sexo, transformados y cargados.

Capítulo 2: Análisis y diseño de la solución

Tabla 3: Descripción del caso de uso Obtener información sobre el número de trabajadores ocupados por categoría ocupacional y sexo

Caso de uso:	Obtener información sobre el número de trabajadores ocupados por categoría ocupacional y sexo.	
Tipo:	Informativo	
Actores:	Analista	
Resumen:	El analista inicia el caso de uso para obtener información sobre el número de trabajadores ocupados por categoría ocupacional y sexo.	
Precondiciones:	Compleitud del almacén. Carga de los datos.	
Referencias	R1 al R27	
Prioridad	Crítico	
Flujo normal de eventos		
Sección “Obtener información”		
Acción del actor	Respuesta del negocio	
1. Entrar al sistema.	2. Muestra opciones de reporte del número de trabajadores por categoría ocupacional y sexo.	
3. Seleccionar reportes del número de trabajadores por categoría ocupacional y sexo.	4. Muestra tipo de reporte del número de trabajadores por categoría ocupacional y sexo.	
Opciones de reportes de ocupaciones		

Capítulo 2: Análisis y diseño de la solución

Perspectivas de análisis	Posibles resultados	
	VARIABLES DE SALIDA	PERIODICIDAD
Variables de entrada relacionadas con el caso de uso: Esfera, forma organizativa, subordinación, temporal, DPA ²¹ , sector ocupación, categoría ocupacional, contratados por tiempo determinado y por edad, empresa, NAE ²² , CAE ²³ , organismo y forma de financiamiento.	Cantidad total de trabajadores.	Anual
	Cantidad de trabajadores del sexo femenino.	Anual
	Cantidad de trabajadores del sexo masculino.	Anual
<i>Prototipo de interfaz</i>		
Poscondiciones	Mostrar la cantidad total de trabajadores ocupados.	

2.3 Diseño

Se tiene en cuenta el análisis realizado para comenzar a diseñar la solución.

2.3.1 Matriz dimensional

Esta matriz está afiliada a la relación que existe entre el hecho y las dimensiones asociadas al mismo.

Lista de hecho:

- ✓ Número de trabajadores ocupados.

Lista de dimensiones:

²¹ DPA: Distribución Política Administrativa.

²² NAE: Nomenclador de Actividades Económicas.

²³ CAE: Clasificador de Actividades Económicas.

Capítulo 2: Análisis y diseño de la solución

- ✓ Esfera.
- ✓ Forma organizativa.
- ✓ Subordinación.
- ✓ Temporal.
- ✓ DPA.
- ✓ Sector ocupación.
- ✓ Categoría ocupacional.
- ✓ Contratados por tiempo determinado y por edad.
- ✓ Empresa.
- ✓ NAE.
- ✓ CAE.
- ✓ Organismo.
- ✓ Forma de financiamiento.

Tabla 4: Matriz dimensional

Hecho/ Dimen sión	DIM _1	DIM _2	DIM _3	DIM _4	DIM _5	DIM _6	DIM _7	DIM _8	DIM _9	DIM _10	DIM _11	DIM _12	DIM _13
H1	x	x	x	x	x	x	x	x	x	x	x	x	x

2.3.2 Modelo de datos

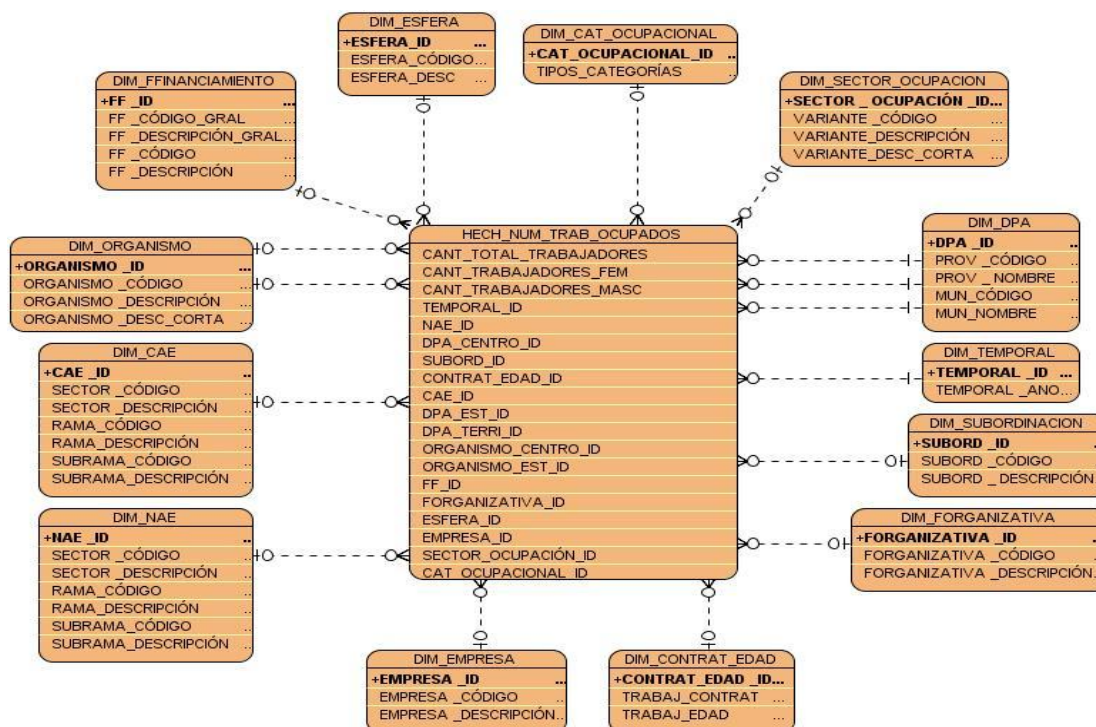


Figura 2: Modelo dimensional

Dimensiones identificadas

DIM_SECTOR_OCUPACIÓN

Tabla 5: Descripción de la dimensión Sector ocupación

Nombre del atributo	Descripción del atributo
SECTOR_OCUPACIÓN_ID	Llave primaria de la dimensión.
VARIANTE_CÓDIGO	Almacena el código de los municipios establecidos nacionalmente.
VARIANTE_DESCRIPCIÓN	Almacena el nombre de las variantes ocupacionales

Capítulo 2: Análisis y diseño de la solución

	establecidas nacionalmente.
VARIANTE_DESC_CORTA	Almacena las siglas de las variantes ocupacionales establecidas nacionalmente.

DIM_CAT_OCUPACIONAL

Tabla 6: Descripción de la dimensión Categoría ocupacional

Nombre del atributo	Descripción del atributo
CAT_OCUPACIONAL_ID	Llave primaria de la dimensión.
TIPOS_CATEGORÍAS	Almacena la clasificación de las categorías existentes (operarios, técnicos, administrativos, de servicios, dirigentes).

DIM_CONTRAT_EDAD

Tabla 7: Descripción de la dimensión Contratados por tiempo determinado y por edad

Nombre del atributo	Descripción del atributo
CONTRAT_EDAD_ID	Llave primaria de la dimensión.
TRABAJ_CONTRAT	Almacena los trabajadores contratados por un tiempo determinado.
TRABAJ_EDAD	Almacena los trabajadores que son mayores y menores de la edad laboral.

Para ver las demás tablas dirigirse a:

https://repositorio.datec.prod.uci.cu/svn/cbd_almacenes/Proyectos/8.%20Data%20Marts%20ONE/2.%20Data%20Mart%20Ocp/

Tabla de hecho identificada

HECH_NUM TRAB_OCUPADOS
CANT_TOTAL TRABAJADORES
CANT TRABAJADORES_FEM
CANT TRABAJADORES_MASC
TEMPORAL_ID
NAE_ID
DPA_CENTRO_ID
SUBORD_ID
CONTRAT_EDAD_ID
CAE_ID
DPA_EST_ID
DPA_TERRI_ID
ORGANISMO_CENTRO_ID
ORGANISMO_EST_ID
FF_ID
FORGANIZATIVA_ID
ESFERA_ID
EMPRESA_ID
SECTOR_OCUPACIÓN_ID
CAT_OCUPACIONAL_ID

Figura 3: Hecho identificado

Tabla 8: Descripción de los atributos de la tabla de hecho identificada

Nombre de los atributos	Descripción de los atributos
TEMPORAL_ID	Almacena la llave primaria de la dimensión.
NAE_ID	Almacena la llave primaria de la dimensión.
CAE_ID	Almacena la llave primaria de la dimensión.
DPA_CENTRO_ID	Almacena el lugar donde radica el centro.
DPA_EST_ID	Almacena el lugar de dónde es el centro.
DPA_TERRI_ID	Almacena la distribución ocupacional del centro.

Capítulo 2: Análisis y diseño de la solución

ORGANISMO_CENTRO_ID	Almacena el organismo al cual pertenece el centro.
ORGANISMO_EST_ID	Almacena el organismo al cual pertenece el centro, después de la reestructuración.
FF_ID	Almacena la llave primaria de la dimensión.
FORGANIZATIVA_ID	Almacena la llave primaria de la dimensión.
SUBORD_ID	Almacena la llave primaria de la dimensión.
ESFERA_ID	Almacena la llave primaria de la dimensión.
EMPRESA_ID	Almacena la llave primaria de la dimensión.
SECTOR_OCUPACIÓN_ID	Almacena la llave primaria de la dimensión.
CAT_OCUPACIONAL_ID	Almacena la llave primaria de la dimensión.
CONTRAT_EDAD_ID	Almacena la llave primaria de la dimensión.

Medidas identificadas

Cantidad total de trabajadores (CANT_TOTAL_TRABAJADORES).

Cantidad de trabajadores del sexo femenino (CANT_TRABAJADORES_FEM).

Cantidad de trabajadores del sexo masculino (CANT_TRABAJADORES_MASC).

Tabla 9: Descripción de las medidas de la tabla de hecho identificada

Nombre de las medidas	Descripción de las medidas
CANT_TOTAL_TRABAJADORES	Almacena la cantidad total de trabajadores.

CANT_TRABAJADORES_FEM	Almacena la cantidad de trabajadores del sexo femenino.
CANT_TRABAJADORES_MASC	Almacena la cantidad de trabajadores del sexo masculino.

2.3.3 Esquema de seguridad

El esquema de seguridad está respaldado por los niveles de acceso, específicamente por los roles definidos.

La arquitectura del sistema de seguridad y alta disponibilidad debe ser de 3 capas:

- ✓ Funcionamiento: dispositivos de seguridad (firewall, inspectores de contenido, sensores).
- ✓ Servidores: servidor de gestión de administración y de base de datos.
- ✓ Presentación: consolas de administración.

La configuración y prueba de todos los equipos debe efectuarse en las instalaciones de la ONE en la Ciudad de La Habana, conjuntamente con el personal técnico de informática, de acuerdo con las especificaciones de configuración y prueba. El mantenimiento preventivo debe efectuarse mediante una visita quincenal durante el período de vigencia de la garantía técnica.

2.3.4 Política de respaldo y recuperación

La política de respaldo y recuperación que utiliza la solución es sencilla pero a la vez sólida, por ello se miden 3 puntos esenciales:

- ✓ Periodicidad de las salvadas: mensualmente se realizan las salvadas de toda la información contenida en la base de datos. La organización lo tiene definido de esta manera, certificando en todo momento la existencia de una copia escrita de la información que está presente en el servidor.
- ✓ Tablas involucradas: las tablas que se involucran son la tabla de hecho del número de trabajadores por categoría ocupacional y sexo, además de las tablas de las dimensiones asociadas al mismo.
- ✓ Backups existentes: actualmente no existen backups en esta área.

2.4 Conclusiones parciales

Con la culminación de este capítulo se identificaron los principales aspectos que contribuyen a la correcta implementación de la base de datos, pues se definió el tema de análisis, las principales necesidades de información de los usuarios, así como los requerimientos funcionales y no funcionales, además de los requisitos de información y los multidimensionales, realizándose la modelación y validación de los mismos. De igual manera se definieron las dimensiones y el hecho del mercado de datos.

Capítulo 3 Implementación y prueba

3.1 Introducción

En este capítulo se hace una transformación del modelo dimensional al modelo físico, así como la selección de la granularidad del proceso de negocio, la identificación de los clasificadores y finalmente la implementación de la base de datos, concluyendo la misma con un proceso de prueba.

3.2 Implementación

Se parte del diseño obtenido para realizar la implementación de la solución.

3.2.1 Modelo de datos físico

Un modelo de datos físico es un modelo abstracto que describe cómo deben ser representados y usados los datos, estos están dirigidos al personal informático, no a los usuarios finales. Para ver el modelo físico de la solución debe dirigirse a la siguiente dirección:

https://repositorio.datec.prod.uci.cu/svn/cbd_almacenes/Proyectos/8.%20Data%20Marts%20ONE/2.%20Data%20Mart%20Ocp/

3.2.2 Estructuras de datos

Las estructuras se pueden definir como una forma de organizar un conjunto de datos elementales con el objetivo de facilitar su manipulación. Estas estructuras están compuestas por diferentes niveles. Por debajo de estas hay un nivel en el cual se sitúan discos, archivos, particiones y espacios de tablas; por lo que el uso adecuado de los mismos, influye en la factibilidad de la solución. En el proceso de desarrollo, el particionamiento de las tablas es fundamental, en función de lograr una mejor organización de la información y velocidad de recuperación.

3.2.2.1 Esquemas y tablas

El esquema es generalmente almacenado en un diccionario de datos. Aunque usualmente el esquema es definido en un lenguaje de base de datos, el término se usa a menudo para referirse a una representación gráfica de la estructura de los datos.

Capítulo 3: Implementación y prueba

Las tablas por su parte se refieren al tipo de modelado de datos, donde se guarda la información recogida por un programa.

Se definieron 2 esquemas para el almacenamiento de las tablas para garantizar la estructura de los datos:

- ✓ El esquema Dimension está conformado por las tablas de categoría ocupacional, contratados por tiempo determinado y por edad, empresa, forma organizativa, forma de financiamiento, sector ocupación, subordinación, entre otras.
- ✓ El esquema Hecho está conformado por la tabla del número de trabajadores ocupados.

Tabla 10: Descripción de esquemas y tablas

Tablas	Esquemas	Propietario	Descripción
DIM_CAE	Dimension	Postgres	Esta tabla contiene la descripción y código de los clasificadores de actividades económicas.
DIM_CATEGORÍA_OCUPACIONAL	Dimension	Postgres	Esta tabla contiene las diferentes categorías ocupacionales.
DIM_CONTRAT_EDAD	Dimension	Postgres	Esta tabla contiene los trabajadores contratados por un tiempo determinado, así como los que son mayores y menores de la edad laboral.
DIM_DPA	Dimension	Postgres	Esta tabla contiene la división política administrativa de todo el país.
DIM_EMPRESA	Dimension	Postgres	Esta tabla contiene el

Capítulo 3: Implementación y prueba

			código y la descripción de la empresa que pertenece al centro informante.
DIM _ ESFERA	Dimension	Postgres	Esta tabla contiene el nombre de las esferas establecidas nacionalmente.
DIM _ FFINANCIAMIENTO	Dimension	Postgres	Esta tabla contiene las diferentes formas de financiamiento, así como su código y descripción.
DIM _ FORGANIZATIVA	Dimension	Postgres	Esta tabla contiene las diferentes formas organizativas, así como su código y descripción.
DIM _ NAE	Dimension	Postgres	Esta tabla contiene la descripción de la subrama, así como el código del sector de los nomencladores de actividades económicas.
DIM_ ORGANISMO	Dimension	Postgres	Esta tabla contiene la descripción del organismo, así como su código.
DIM _ SECTOR _ OCUPACIÓN	Dimension	Postgres	Esta tabla contiene las variantes, así como su descripción y código.
DIM_ SUBORDINACIÓN	Dimension	Postgres	Esta tabla contiene

			códigos, así como una descripción del tipo de subordinación.
DIM_TEMPORAL	Dimension	Postgres	Esta tabla contiene los rangos de tiempo en que se recogen los datos.
HECH_NUM_TRAB_OCUPADOS	Hecho	Postgres	Esta tabla contiene todas las dimensiones del modelo, así como las diferentes medidas.

3.2.2.2 Restricciones y secuencias

Las restricciones son reglas que deben mantener los datos almacenados en la base de datos para que cumplan con ciertos parámetros. Las mismas pueden ser creadas automáticamente al definir una tabla (en el caso de las llaves primarias) o ser introducidas por el programador de la base de datos cuando se busca algo específico. Existen 4 tipos de restricciones que son las más comunes:

- ✓ Clave foránea: es la que referencia una clave de otra tabla, se usa para relacionar tablas diferentes.
- ✓ Clave primaria: es el valor que debe cumplir con un conjunto de restricciones: no tener valores nulos, ser únicos para cada tupla y ser necesarios (mínima).
- ✓ De unicidad: implica que no debe haber dos valores iguales en la misma columna.
- ✓ Valor no nulo: no debe existir ninguna casilla de la columna que esté vacía. (Casares, 1999-2010)

En la solución existen 16 llaves foráneas, a continuación se muestran algunos ejemplos de las mismas:

Capítulo 3: Implementación y prueba

Tabla 11: Descripción de llaves foráneas

Llaves Foráneas	Tabla	Esquema	FK Tabla
RefDIM_CAE12	HECH_NUM_TRAB_OCUPADOS	Hecho	Dimension.DIM CAE
RefDIM_CAT_OCUPACIONAL 17	HECH_NUM_TRAB_OCUPADOS	Hecho	Dimension. DIM CAT_OCUPACIONAL
RefDIM_CONTRAT_EDAD18	HECH_NUM_TRAB_OCUPADOS	Hecho	Dimension.DIM CONTRAT_EDAD
RefDIM_DPA30	HECH_NUM_TRAB_OCUPADOS	Hecho	Dimension.DIM_ DPA
RefDIM_DPA31	HECH_NUM_TRAB_OCUPADOS	Hecho	Dimension.DIM_ DPA
RefDIM_DPA32	HECH_NUM_TRAB_OCUPADOS	Hecho	Dimension.DIM_ DPA
RefDIM_EMPRESA9	HECH_NUM_TRAB_OCUPADOS	Hecho	Dimension. DIM EMPRESA

Las secuencias son atributos que van a ir incrementándose secuencialmente durante la entrada de información a la base de datos, en este caso las llaves primarias serían estos tipos de valores.

En la solución existen 13 llaves primarias, a continuación se muestran algunos ejemplos de las mismas.

Tabla 12: Descripción de llaves primarias

Llave Primaria	Nombre	Tipo de datos	Única
✓	CAE _ ID	serial	✓
	SUBRAMA_DESCRIPCIÓN	text	

--	--	--	--

Llave Primaria	Nombre	Tipo de datos	Única
✓	CAT_OCUPACIONAL_ID	serial	✓
	TIPOS _ CATEGORÍAS	text	

3.2.2.3 Índices

El índice de una base de datos es una estructura de datos que mejora la velocidad de las operaciones, permitiendo un rápido acceso a los registros de una tabla, guardando parejas de elementos: el elemento que se desea indexar y su posición en la base de datos. Para buscar un elemento que esté indexado, exclusivamente hay que buscar en el índice dicho elemento, para una vez encontrado, devolver el registro que se encuentre en la posición marcada por el índice. (MSDN Microsoft Corporation, julio de 2009-2010)

En la solución se identificaron 14 índices, a continuación se muestran ejemplos de los mismos.

Tabla 13: Descripción de los índices

Índice	Tablas	Esquemas	Campos
PK20	DIM_CAE	Dimension	CAE_ID
PK23	DIM_CAT_OCUPACIONAL	Dimension	CAT_OCUPACIONAL_ID
PK25	DIM_CONTRAT_EDAD	Dimension	CONTRAT_EDAD_ID
PK15	DIM_DPA	Dimension	DPA_ID
PK18	DIM_EMPRESA	Dimension	EMPRESA _ ID

PK11	DIM_ESFERA	Dimension	ESFERA_ID
PK22	DIM_FFINANCIAMIENTO	Dimension	FF_ID

3.2.2.4 Descripción del artefacto DDL

EL DDL²⁴ es un lenguaje proporcionado por el sistema de gestión de base de datos que permite a los usuarios de la misma llevar a cabo las tareas de definición de las estructuras que almacenan los datos, así como de los procedimientos o funciones que permitan consultarlos.

El lenguaje de programación SQL, el más difundido entre los gestores de bases de datos, admite las siguientes sentencias de definición: CREATE, DROP y ALTER, cada una de las cuales se puede aplicar a las tablas, vistas, procedimientos almacenados y triggers de la base de datos.

Otras que se incluyen dentro del DDL, pero que su existencia depende de la implementación del estándar SQL que lleve a cabo el gestor de base de datos son GRANT y REVOKE, las cuales sirven para otorgar permisos o quitarlos, ya sean a usuarios específicos o a un rol creado dentro de la base de datos. (Quintero, 2008)

El artefacto DDL está ubicado en el repositorio en la carpeta de nombre Script, el fichero DDL_Ocupaciones.sql.

3.2.3 Usuarios y privilegios

En cuanto a los usuarios y privilegios hay que tener en cuenta la existencia de permisos comunes entre roles, por lo que es necesario agrupar a estos en grupos de trabajo, debido a que se dificulta este tipo de asignación de diferentes usuarios sobre un mismo objeto; es por eso que Postgres brinda la posibilidad de agrupar a estos según las necesidades de permisos y accesos que necesita cada rol.

3.2.3.1 Usuarios

La definición de los usuarios y roles que interactúan con la base de datos permite lograr una mejor protección y seguridad de la misma, los usuarios son los siguientes:

²⁴ DDL: Lenguaje de Definición de Datos. Data Definition Language, por sus siglas en inglés.

- ✓ Administrador: es el especialista encargado de brindar mantenimiento a la base de datos, así como la administración y configuración de la misma, garantizando un funcionamiento óptimo.
- ✓ Analista: es el encargado de consultar la información.
- ✓ Integrador: es el especialista encargado de todo el proceso de extracción, transformación y carga de los datos de la base de datos.
- ✓ Postgres: es el súper-usuario de la base de datos.

3.2.3.2 Privilegios

Los privilegios asignados a cada usuario dependen exclusivamente del rol que desempeñan:

- ✓ Administrador: se le asigna el derecho de lectura y escritura de la base de datos.
- ✓ Analista: este usuario únicamente tiene derecho a la lectura de los datos contenidos en el mercado de datos.
- ✓ Integrador: por el rol que desempeña este usuario tiene derecho de lectura y escritura de los datos almacenados.
- ✓ Postgres: hace completo uso de la base de datos, tiene permisos de lectura y escritura.

3.2.3.3 Descripción del artefacto DCL

El artefacto DCL²⁵ está directamente relacionado con la seguridad del sistema y constituye un apoyo para las labores de administración de la base de datos, mediante un subconjunto de instrucciones SQL. Quien genera dichas acciones es el DCL, que se encarga de reflejar todo el proceso de acceso a la base de datos.

En PostgreSQL, la ejecución de un comando DCL forma parte de una transacción, por lo que puede ser deshecha mediante el comando rollback. (Quintero, 2008)

El artefacto DCL está ubicado en el repositorio en la carpeta de nombre Script, el fichero DCL_Ocupaciones.sql.

²⁵ DCL: Lenguaje de Control de Datos. Data Control Language, por sus siglas en inglés.

3.2.4 Carga de nomencladores

La fase de carga es el momento en que los datos provenientes de la fase anterior son incluidos en el sistema de destino. Dependiendo de los requerimientos de la organización, este proceso puede abarcar una amplia variedad de acciones diferentes; en algunas bases de datos, se sobrescribe la información antigua con nuevos datos; almacenes de datos, por ejemplo, mantienen un historial de los registros de manera que se pueda hacer una auditoría de los mismos y disponer de un rastro del comportamiento de un determinado valor a lo largo del tiempo.

La fase de carga interactúa directamente con la base de datos de destino. Al realizar esta operación se aplican todas las restricciones que se hayan definido en ésta, por ejemplo, valores únicos, integridad referencial, campos obligatorios, rangos de valores. Estas restricciones están bien definidas, en tanto garantizan la calidad de los datos en el proceso ETL.

3.2.4.1 Nomenclador

Los nomencladores son los que facilitan la flexibilidad del sistema, además de tener organizada la información y así aseguran la consistencia de los datos. Este es un instrumento desarrollado que se enmarca dentro de la propuesta de transmisión electrónica de datos estandarizados.

- ✓ El nomenclador DPA contiene la distribución del país en provincias y municipios.
- ✓ El nomenclador categoría ocupacional contiene las diferentes clasificaciones ocupacionales.
- ✓ El nomenclador sector ocupación contiene las variantes ocupacionales establecidas en el país.
- ✓ El nomenclador organismo contiene la descripción del centro informante, así como el código del mismo.
- ✓ El nomenclador empresa contiene los códigos y las descripciones de las diferentes empresas del país.
- ✓ El nomenclador subordinación contiene una descripción a la cual se encuentra subordinado cada centro informante del país.
- ✓ El nomenclador esfera contiene el nombre y el código de las esferas establecidas nacionalmente.
- ✓ El nomenclador temporal contiene la periodicidad con la que se van a realizar los reportes.

3.2.4.2 Descripción del artefacto DML

DML²⁶ es un lenguaje proporcionado por el sistema de gestión de base de datos que permite a los usuarios llevar a cabo las tareas de consulta o manipulación de la información, organizados por el modelo de datos adecuado. El más utilizado hoy día es el SQL, usado para recuperar y manipular datos. También contiene herramientas de acceso a los SGBD, las cuales sirven para escribir e interpretar las instrucciones SQL o DDL, que se envían al motor de base de datos para ser ejecutadas.

Ejemplos de DML: Select, Insert, Delete, Update.

Los DML se clasifican en dos grandes grupos:

- ✓ Los lenguajes de consulta procedimentales: el usuario da instrucciones al sistema para que realice una serie de procedimientos u operaciones en la base de datos para calcular un resultado final.
- ✓ Los lenguajes de consulta no procedimentales: el usuario describe la información deseada sin un procedimiento específico para obtener esa información. (Quintero, 2008)

El artefacto DML está ubicado en el repositorio en la carpeta de nombre Script, el fichero DML_Ocupaciones.sql.

3.3 Guía de implantación

Las guías de implantación se centran en desarrollar el sistema a nivel organizativo y en simplificar la gestión de la información. En concreto, el proceso de planificación estratégica y de gestión recoge la estructura de planes, así como las operaciones de la formulación, el desarrollo y la revisión de la estrategia, para la accesibilidad de la información y la simplificación de la gestión en el proceso de obtención de entrada de datos

3.3.1 Requerimientos

Un requerimiento es una necesidad documentada sobre el contenido, forma o funcionalidad de un producto o servicio. Se utilizan como datos de entrada en la etapa de diseño del producto.

Los requisitos que debe tener el servidor en cuanto a:

²⁶ DML: Lenguaje de Manipulación de Datos. Data Manipulation Language, por sus siglas en inglés.

- ✓ Software: sistema operativo Windows XP Service Pack2, Postgre8.4, PgAdmin III.
- ✓ Hardware: 1 GB de memoria RAM, 80 GB de disco duro.

3.3.2 Secuencia de pasos

- ✓ Debe de estar instalado PostgreSQL 8.4 como gestor de base de datos.
- ✓ Debe de estar instalada una herramienta de administración de base de datos, en este caso PgAdmin III.
- ✓ Se debe crear la nueva base de datos utilizando la herramienta de administración.
- ✓ Se crean los roles y permisos de los usuarios.
- ✓ Se carga el script DDL_Ocupaciones.sql.
- ✓ Se carga el script DML_Ocupaciones.sql.
- ✓ Se carga el script DCL_Ocupaciones.sql.
- ✓ Se verifica que la base de datos funcione correctamente.

3.4 Validación y pruebas

A través del proceso de validación se puede garantizar la realización de un producto que cumpla con las especificaciones y atributos de calidad, concluyendo con un proceso de prueba que permita verificar y revelar la calidad del software.

3.4.1 Listas de chequeo del análisis

Las listas de chequeo se basan en un listado de preguntas en forma de cuestionario que sirven para verificar el grado de cumplimiento de determinadas reglas establecidas con un fin determinado, garantizando que la información recogida cumpla con el formato establecido y que la misma esté bien organizada.

- ✓ Lista de chequeo de evaluación de las áreas de la organización.
- ✓ Lista de chequeo de la herramienta para la recolección y análisis de la información.
- ✓ Lista de chequeo de especificación de requisitos.

Después de aplicadas las listas de chequeo del análisis se detectó una no conformidad de tipo error ortográfico de peso crítico, la cual fue corregida.

3.4.2 Validación de requisitos por el cliente

El proceso de validación fue satisfactorio, ya que el cliente estuvo de acuerdo con los resultados obtenidos con respecto al desarrollo del producto. Este fue realizado el día 5 de mayo del 2010, por la especialista Elena Leonila Fernández García y el Máster Asnioby Hernández López.

3.4.3 Lista de chequeo del diseño

A raíz del cuestionario realizado por los diseñadores se arribaron a determinados resultados que dan cumplimiento a los aspectos reflejados en la lista de chequeo que aparece en el Anexo 1:

- ✓ El nombre de las tablas de las dimensiones está estructurado de la siguiente manera: DIM_NOMBRE.
- ✓ El nombre de la tabla de hecho está estructurado de la siguiente manera: HECH_NOMBRE.
- ✓ Todas las tablas de las dimensiones cuentan con un identificador o llave primaria que es única.
- ✓ El tipo de dato de los identificadores debe ser de tipo serial para que se generen secuencialmente.
- ✓ El nombre de los usuarios creados en el modelo de datos debe coincidir con el nombre de los esquemas.
- ✓ Al cargar los nomencladores todos sus campos deben estar llenos, pues no deben existir tablas vacías.

3.4.4 Pruebas de implantación

Las pruebas las realizan los técnicos de sistemas y de operación, que forman parte del grupo de usuarios técnicos que han recibido la formación necesaria para llevarlas a cabo.

La finalidad de las pruebas de implantación es doble:

- ✓ Comprobar el funcionamiento correcto en el entorno de operación.
- ✓ Permitir que el usuario determine, desde el punto de vista de operación, la aceptación del sistema instalado en su entorno real, según el cumplimiento de los requisitos especificados.

Tabla 14: Descripción de los casos de prueba

Casos de prueba	Precondición	Poscondición	Resultado esperado
Crear base de datos	PostgreSQL 8.4 instalado como gestor de base de datos.	Base de datos creada.	Base de datos creada sobre el gestor de base de datos PostgreSQL 8.4.
Crear roles y permisos	Base de datos creada e instalada la herramienta de administración PgAdmin III.	Base de datos creada con sus roles y permisos.	Roles creados con los respectivos permisos.
Cargar nomencladores	Base de datos creada con sus roles y permisos.	Base de datos lista para ser utilizada.	Campos de la base de datos llenos.

3.5 Conclusiones parciales

Al culminar el capítulo han quedado definidos aspectos de gran importancia para el funcionamiento de la solución, así como la definición de la estructura de los datos para proveerle seguridad a la información almacenada en la base de datos, se cargaron todos los nomencladores de las dimensiones asociadas al hecho, se validaron los requerimientos por parte de los clientes y se realizaron pruebas de implantación para verificar la validez de la solución desarrollada.

Conclusiones generales

A partir de un estudio exhaustivo para la selección de la tecnología para el análisis, diseño e implementación del mercado de datos para los indicadores específicos de la ocupación para la Oficina Nacional de Estadística quedó seleccionada como herramienta de modelado el Visual Paradigm en su versión 6.4, el sistema gestor de bases de datos fue PostgreSQL en su versión 8.4 y como metodología la definida por la Línea de Almacenes de Datos y BI del Centro de Tecnología de Datos. El análisis del modelo 5200-03 permitió conocer la forma de tratar el indicador, donde se confeccionaron temas de interés para lograr una mejor comprensión del negocio. Se diseñó el mercado de datos, donde de forma clara quedaron plasmadas las dimensiones, las medidas y el hecho asociado al mercado. Se implementaron y cargaron los clasificadores para el mercado de datos y se validó la solución desarrollada mediante la realización de pruebas.

Con la construcción de la solución propuesta se concluye que este mercado de datos para los indicadores específicos de la ocupación contribuye a mejorar la realización de los principales reportes y cruces de variables, indicadores, tasas, porcentajes y demás aspectos de interés, facilitando la disponibilidad de información para los Órganos del Estado y el proceso de toma de decisiones.

Recomendaciones

Se recomienda:

- ✓ Culminar la construcción del mercado de datos de los indicadores específicos de la ocupación.
- ✓ Integrar la solución desarrollada al almacén de datos de la Oficina Nacional de Estadística.

Bibliografía

1. Qué es un Data Mart?[Consultado el: 15 de febrero del 2010]. Mercado De Datos. Data Mart. [Online] Disponible en: <http://www.mitecnológico.com/Main/MercadosDeDatosDataMart>.
2. Metodologías de desarrollo[Consultado el: 15 de febrero del 2010]. BIBLIOTECA VIRTUAL de Derecho, Economía y Ciencias Sociales. [Online] Disponible en: <http://www.eumed.net/libros/2009c/584/metodologías%20de%20desarrollo%20en%20el%20contexto%20mundial.htm>.
3. Metodologías [Consultado el: 15 de febrero del 2010]. Metodologías de Desarrollo de Software. [Online] Disponible en: <http://alarcos.inf-cr.uclm.es/doc/ISOFTWAREI/Tema04.pdf>.
4. Equipo de desarrollo de DATEC. METODOLOGÍA PARA EL DESARROLLO DE SOLUCIONES DE ALMACENES DE DATOS E INTELIGENCIA DE NEGOCIOS.
5. El Datawarehouse: nueva perspectiva de consulta para las empresas[Consultado el: 15 de febrero del 2010]. El Datawarehouse: nueva perspectiva de consulta para las empresas. [Online] Disponible en: <http://www.datawarehouse.com/trabajos16/datawarehouse/datawarehouse.shtml#INTRO>.
6. Ochoa, Darián González. Trabajo de Diploma: Diseño e Implementación de un Almacén de Datos Operacionales para la Corporación CIMEX. 2009.
7. —. The Greenplum Data Warehouse Powered by Sun[Consultado el: 21 de febrero del 2010]. Oracle. [Online] Disponible en: <http://www.sun.com/solutions/landing/infrastructure/dwa/benefits.jsp>.
8. INE [Consultado el 16 de marzo del 2010].INE Instituto Nacional de Estadística. [Online] Disponible en: <http://www.ine.es/ine/eline.htm>.
9. MEDITECH. Almacén de datos[Consultado el: 26 de febrero del 2010]. MEDITECH. [Online] Disponible en: http://www.meditech.com/MeditechEnEspañol/Pages/ES_DataRepository.htm.
10. Microsoft. Almacén de datos de ejemplo AdventureWorksDW[Consultado el: 26 de febrero del 2010]. [Online] consultado en: <http://msdn.microsoft.com/es-es/library/ms124623.aspx>.
11. Corporación CIMEX. Sistema DataWarehouse Comercial de la Corporación CIMEX[Consultado el: 22 de febrero del 2010].
12. García, Alejandro González. DESARROLLO METODOLÓGICO PARA LA IMPLEMENTACIÓN DEL PROCESO DE DATAWAREHOUSING EN CUBAENERGÍA[Consultado el: 23 de febrero del 2010].

.docStoc. [Online] Disponible en: <http://www.docstoc.com/docs/23518404/El-uso-y-aplicaci%C3%B3n-de-la-energ%C3%ADa-nuclear-con-fines-pac%C3%ADficos->.

13. Marquéz, María Mercedes Andrés[Consultado el: 1 de abril del 2010]. [Online] Disponible en: <http://www3.uji.es/~mmarques/f47/apun/node4.html>.

14. Microsoft[Consultado el: 2 de abril del 2010]. [Online] Disponible en: <http://www.microsoft.com/spain/sql/productinfo/features/top30features.msp>.

15. Warp[Consultado el: 2 de abril del 2010]. [Online] Disponible en: <http://warp.es/mysql/productos/razones/>.

Referencias bibliográficas

1. Inmon, W. H. (2006) Definiciones de almacén de datos[Consultado el: 15 de febrero del 2010]. Information Management. [Online] Disponible en: <http://informationmanagement.wordpress.com/2006/11/28/dw-definiciones-de-inmon-y-kimball/>.
2. Kimball, Ralph. (2006) Definiciones de almacén de datos. Information Management[Consultado el: 15 de febrero del 2010]. [Online] Disponible en: <http://informationmanagement.wordpress.com/2006/11/28/dw-definiciones-de-inmon-y-kimball/>.
3. Sobre PostgreSQL. (2009)[Consultado el: 25 de febrero del 2010]. PostgreSQL-es.org. [Online] Disponible en: <http://archives.postgresql.org/pgsql-es-fomento/2009-07/msg00000.php>.
4. Orallo, José Hernández. (2003) Análisis y Extracción de Conocimiento en Sistemas de Información:Datawarehouse y Datamining[Consultado el: 26 de febrero del 2010]. [Online] 2003. Disponible en: <http://users.dsic.upv.es/~jorallo/cursoDWDWM/>.
5. CASE Y RAD. (2006)[Consultado el: 26 de febrero del 2010]. [Online] Disponible en: <http://www.fortunecity.es/imaginapoder/apunte/434/caseyrad2.htm#Definici%F3n>.
6. Verástegui, Hazbleydi C. (2007) Modelo Dimensional de Datos[Consultado el: 31 de marzo del 2010]. [Online] Disponible en: http://api.ning.com/files/ei6*2AatbAfAFDwyLdYnMfJ7ZXvd3xwXakwY*glIpDQk4Ykcb3K9af3G*OvF06uLSTXQWlltUimXRHhSRU8enhjUX2QDs-kR/MODELADODIMENSIONALDEDATOS_V2.pdf.
7. Castillo, Carlos. (2008) Sistemas de información 2[Consultado el: 31 de marzo del 2010]. [Online] Disponible en: http://www.tejedoresdelweb.com/wiki/images/a/a5/Basesdatos_teo5_modelo_relacional.pdf.
8. Rubia, Jorge Moral. (2007) Introducción a los almacenes de datos[Consultado el: 26 de marzo del 2010]. [Online] 2007.
9. Luján-Mora, Sergio. (2007)[Consultado el: 31 de marzo del 2010]. [Online] Disponible en: http://www.google.com.cu/url?sa=t&source=web&ct=res&cd=9&ved=0CB8QFjAI&url=http%3A%2F%2Fgplsi.dlsi.ua.es%2Falmacenes%2Fver.php%3Fpdf%3D76&rct=j&q=OLAP+%2B+almacenes+da+datos&ei=Wq_PS6vqAYGclgfqpaGiCw&usg=AFQjCNEXe8nRP__kw3oD7NpjlvvnDbWrw.
10. Toledo, Enríquez. (2008)[Consultado el: 1 de abril del 2010]. [Online] Disponible en: <http://www.uaem.mx/posgrado/mcruz/cursos/miic/MySQL.pdf>.
11. Bolaños, Susana. (2006)[Consultado el: 3 de abril del 2010]. [Online] Disponible en: <http://www.uaem.mx/posgrado/mcruz/cursos/miic/mysql16.pdf>

12. Oracle. World's Largest Data Warehouse. (2003)[Consultado el: 20 de febrero del 2010]. Oracle. [Online] Disponible en: <http://www.sun.com/service/refarch/datawarehouse/>.
13. INE (2010) [Consultado el 30 de marzo del 2010].INE Instituto Nacional de Estadística. [Online] Disponible en: <http://www.ine.es/>.
14. INE (2008) [Consultado el 30 de marzo del 2010].INE Instituto Nacional de Estadísticas. [Online] Disponible en: http://www.ine.cl/canales/chile_estadistico/home.php.
15. Oracle. (2008)[Consultado el: 3 de abril del 2010]. [Online] Disponible en: <http://www.oracle.com/newsletters/sap/volumes/volume12-sp.pdf>.
16. Danysoft. (2007)[Consultado el: 28 de marzo del 2010]. [Online] Disponible en: <http://www.danysoft.com/bol/ca-erwin.htm>.
17. Embarcadero. (2005)[Consultado el: 28 de marzo del 2010]. [Online] Disponible en: <http://www.embarcadero.com/>.
18. Giraldo, Luis y Zapata, Yuliana. (2005) Herramientas de Desarrollo de Ingeniería de SW[Consultado el: 28 de marzo del 2010]. [Online] Disponible en: <http://www.embarcadero.com/>.
19. Dataprix. (2008)[Consultado el: 28 de marzo del 2010]. [Online] Disponible en: http://www.google.com/cu/url?sa=t&source=web&ct=res&cd=1&ved=0CAYQFjAA&url=http%3A%2F%2Fwww.dataprix.com%2Fprintpdf%2F288&rct=j&q=Metodolog%C3%ADa+SQLBI&ei=Nb_PS5L3NYGdIgfCroSiCw&usq=AFQjCNGD0prG8PrOfwBRU4r5Pwq4CtYiMg.
20. Ideasoft. (2006)[Consultado el: 27 de marzo del 2010]. [Online] Disponible en: www.ideasoft.com.uy/consult/dm2.htm.
21. Dataprix. (2006)[Consultado el: 28 de marzo del 2010]. [Online] Disponible en: <http://www.dataprix.com/en/la-metodolog%C3%AD-crisp-dm>.
22. Dataprix. (2008)[Consultado el: 25 de marzo del 2010]. [Online] Disponible en: <http://www.dataprix.com/es/data-warehousing-hefesto>.
23. Zepeda, Leopoldo Zenaido. (2008) Metodología para el Diseño Conceptual de Almacenes de Datos[Consultado el: 2 de marzo del 2010]. [Online] Disponible en: <http://www.dsic.upv.es/docs/bib-dig/tesis/etd-04032008-120043/tesisprimeraversionLeopoldo.pdf>.
24. Inteligencia de Negocios Datos. (2007)[Consultado el: 2 de marzo del 2010]. [Online] Disponible en: <http://www.gopac.com.mx/v3/gopacbi/quees.asp>.

25. Sobre PostgreSQL. (2009)[Consultado el: 16 de febrero del 2010]. PostgreSQL-es.org. [Online] Disponible en: http://www.postgresql-es.org/sobre_postgresql.
26. Coloma, Miguel Ángel Domínguez y Salas, Emilio Ligeró. (2006)[Consultado el: 29 de marzo del 2010]. [Online] Disponible en: <http://kybele.escet.urjc.es/documentos/HC/Exposiciones/EasyCASE.pdf>.
27. Cabrera, María Evelia Casales. 2009. Data Warehouse.
28. Casares, C. (1999-2010). programatium.com. Retrieved 4 5, 2010, from programatium.com: <http://www.programatium.com/manuales/sql/moddat007.htm>
29. MSDN © 2010 Microsoft Corporation. (julio de 2009-2010). msdn. Retrieved 4 5, 2010, from msdn: <http://msdn.microsoft.com/es-es/library/ms175049.aspx>
30. Juan Bernardo Quintero, D. M. (Julio 2008). DIRECTRICES PARA LA CONSTRUCCIÓN DE ARTEFACTOS DE PERSISTENCIA EN EL PROCESO DE DESARROLLO DE SOFTWARE. Revista EIA, ISSN 1794-1237 Número 9, p. 77-90. , 88.

Anexos

Anexo 1

Preguntas del cuestionario:

¿El nombre de las tablas de las dimensiones posee la estructura DIM_NOMBRE?

¿El nombre de la tabla de hecho posee la estructura HECH_NOMBRE?

¿Las tablas de las dimensiones cuentan con una llave primaria o identificador?

Comprobar que los tipos de datos de los identificadores sean de tipo serial.


¿En la tabla de hecho aparecen las medidas y las llaves foráneas?

¿Todos los campos están llenos al cargar los nomencladores?

Verificar que los nombres de los usuarios creados en el modelo de datos coincidan con los esquemas creados.

Anexo 2

Modelo 5200-03

		Sistema de Información de Estadística Nacional (SIEN)		NÚMERO DE TRABAJADORES POR CATEGORÍA OCUPACIONAL Y SEXO				INFORME AL CIERRE DE: 30 de septiembre del año: ____				MODELO No. 5200-03 Página ____ de ____ ANUAL							
Centro informante:						Código del centro informante:				Variante:				UNIDAD DE MEDIDA: Uno					
MUNICIPIOS	CÓDIGO D.P.A.	NÚMERO DE TRABAJADORES SEGÚN REGISTRO																	
		TOTAL		CONTRATADOS POR TIEMPO DETERMINADO		MAYORES DE LA EDAD LABORAL QUE TRABAJAN		MENORES DE LA EDAD LABORAL QUE TRABAJAN		CATEGORÍA OCUPACIONAL									
		Operarios	Técnicos	Administrativos	De servicios	Dirigentes													
A	B	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18
Total																			
Suma de control																			

Glosario de términos

Alter: modificar, cambiar, alterar, transformar.

Ámbito: espacio comprendido dentro de límites determinados.

Agregación: unión o adición de algo a un todo.

BMMsoft: compañía desarrolladora de software para empresas.

Código abierto: término con el que se conoce al software distribuido y desarrollado libremente.

Create: crear.

Datamart: mercado de datos.

Datawarehouse: almacén de datos.

Desagregación: separación o disgregación.

Drop: eliminar.

Escrutinio: cómputo de los votos en las elecciones o en otro acto análogo.

Grant: conceder, otorgar.

Iceberg: gran masa de hielo.

Ingente: muy grande, enorme.

Microsoft: empresa estadounidense, dedicada al sector de la informática.

Reporting: presentación de informes.

Revoke: anular, renunciar.

Rollback: retroceso.

Sybase: compañía de software que se dedica a la producción de servicios relacionados con la gestión de información.

SQL Server: gestor de bases de datos.

Triggers: desencadenantes.

OPINIÓN DEL USUARIO DEL TRABAJO DE DIPLOMA

El Trabajo de Diploma, titulado "Análisis, diseño e implementación del mercado de datos para los indicadores específicos de la ocupación para la Oficina Nacional de Estadística", fue realizado en el Centro de Tecnología de Datos. Esta entidad considera que, en correspondencia con los objetivos trazados, el trabajo realizado le satisface

- Totalmente
 Parcialmente en un ____ %

Los resultados de este Trabajo de Diploma le reportan a esta entidad los beneficios siguientes (cuantificar):

Los directivos pueden dedicar más tiempo al análisis de los datos en vez de a su búsqueda debido a que se aceleran las consultas al reducir la cantidad de datos a recorrer. La acumulación de datos les permitirá detectar tendencias y hacer previsiones.

Y para que así conste, se firma la presente a los 3 días del mes de junio del año 2010.

Elena L. Fernández García

Representante de la entidad


Firma

Coordinadora del proyecto

