



UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS

FACULTAD # 8

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

**ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DEL MERCADO DE DATOS SOBRE LAS
INVERSIONES PARA LA OFICINA NACIONAL DE ESTADÍSTICAS**

AUTORES:

Yanelis Rosales Leyva

Leonardo Rodríguez Labrada

TUTOR:

Ing. Karenia Donatien Goliath

Ciudad Habana, Junio 2010

“Año 52 de la Revolución”

*“Nunca consideres el estudio como una obligación,
sino, como una oportunidad para penetrar
en el bello y maravilloso mundo del saber.”*

Albert Einstein



Dedicatoria

A mi mamá por ser mi inspiración, por depositar toda su confianza en mí y por tener la paciencia y fuerza de esperar en todos estos años que hemos pasados separadas.

A mi papá por ser mi guía y ejemplo, por ser exigente cuando tiene que serlo y por depositar toda su confianza en mí.

A mi hermano, que aunque no sepa mucho de la vida, esa es la razón de vivir de nosotros.

A los tres: los amo y hoy ya su negrita es ingeniera.

Todo este esfuerzo es para ustedes.

Yanelis.

A mi abuelo Miguel Ángel, porque me dio las fuerzas necesarias para terminar mi carrera.

Leonardo.

Agradecimientos

A mi mamá y a mi papá por estar siempre conmigo cuando hace falta, sin ustedes no hubiera sido posible el éxito de mi carrera.

A toda mi familia, en especial a todos mis tíos, primas y primos, mis abuelos, gracias por estar siempre pendientes de mí y mis abuelos que no están, donde quieran que estén sé que estarán orgullosos de mí.

A mi novio Yordanis que a pesar de la distancia, nos hemos demostrado que con amor se puede todo. Gracias por apoyarme y ser el sostén de mi vida al estar lejos de mis padres.

A mis amigos, que siempre compartieron momentos tristes y alegres conmigo, que lloramos y reímos juntos, en especial Mery, Ayerim, Héctor, Palmero, Rosio, Ivet, Isa, Farita, Marlon, Ernesto, Michel, Elizabeth, Nory, Anny, y Sury. Independiente de las vueltas que de la vida, seguiremos siendo amigos para siempre.

A mi compañero de tesis, por saber ser perseverante y fuerte para lograr todos nuestros objetivos. Gracias por tu ánimo para hoy en día haber logrado nuestra meta.

A la familia de mi novio, especialmente a mi suegra Danis, por su constante preocupación, por todos sus consejos y poder contar con su apoyo. Gracias por estar presente.

A mi tutora Karenia por su dedicación y su paciencia.

A todos los que han hecho posible que hoy en día yo sea lo que soy, en especial a las profesoras Yosety, Arlenis, Lombillo, al profe Dosagues, y a nuestro decano. Gracias a todos.

A mi tribunal en especial a mi oponente Yaneisy, por corregirnos cada detalle y estar dispuestos cada vez que hiciera falta. Gracias por su apoyo.

A la Revolución y a nuestros comandantes Fidel y Raúl Castro por darnos la oportunidad de estudiar en esta universidad y formarnos como profesionales.

Yanelis.

AGRADECIMIENTOS

A mis tres padres: a mi mamá, a mi papá y a mi padrastro, a quienes tengo que agradecerles muchísimo.

A toda mi familia, por servirme de apoyo y por estar siempre pendientes a mis tutores.

A mis amigos que han compartido conmigo estos 5 años.

A mi compañera de tesis por ser una buena compañera de tesis.

A mi tutora Karenia por corregirnos cada detalle en todo el recorrido de nuestra tesis.

A las profesoras Yosety y Arlenis, sin su ayuda ni siquiera hubiera comenzado a realizar mi trabajo de diploma.

A mi tribunal, en especial a mi oponente Yaneisy por ser una persona tan comprensiva.

A la Revolución y a nuestros comandantes Fidel y Raúl por haber creado esta universidad de la que fui parte.

Leonardo.

DECLARACIÓN DE AUTORÍA

Declaración de Autoría

Declaramos que somos los únicos autores del trabajo “Análisis, diseño e implementación del mercado de datos sobre las inversiones para la Oficina Nacional de Estadísticas” y autorizamos a 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 2010.

Autores:

Yanelis Rosales Leyva

Leonardo Rodríguez Labrada

Tutor:

Ing. Karenia Donatien Goliath

Resumen

El presente trabajo se desarrolla a partir de la necesidad de reconstruir o modificar el sistema para recoger la información que existe en la Oficina Nacional de Estadísticas. Teniendo en cuenta dicha necesidad, se toma como decisión la creación de un Almacén de Datos y dentro del mismo un Mercado de Datos sobre las inversiones, para perfeccionar el sistema que se encuentra en estos momentos y así acelerar la velocidad de respuesta del mismo en cuanto a las consultas sobre la información existente. En esta investigación se recoge el análisis, diseño e implementación del antes mencionado Mercado de Datos.

Índice de Contenidos

Introducción	1
Capítulo 1: Fundamentación Teórica	5
1.1 Introducción	5
1.2 Conceptos fundamentales asociados al negocio	5
1.2.1 Almacén de Datos	5
1.2.1.1 Definiciones generales	5
1.2.1.2 Características	6
1.2.1.3 Beneficios	6
1.2.2 Modelo Dimensional	7
1.2.2.1 Conceptos asociados al modelo dimensional	7
1.2.2.2 Esquemas que se encuentran dentro del modelo dimensional	8
1.2.3 Proceso OLAP	9
1.2.4 Mercado de datos	10
1.2.4.1 Definiciones generales	10
1.2.4.2 Características	11
1.2.4.3 Beneficios	11
1.2.5 Diferencias entre Mercado de Datos y Almacén de Datos	12
1.2.6 Modelos estadísticos del área temática de Inversiones. (Modelo Sien, 2009)	12
1. 3 Tendencias y Tecnologías Actuales	13
1.3.1. Gestor de base de datos	13
1.3.1.1 Microsoft SQL Server	13
1.3.1.2 Oracle	14
1.3.1.3 PostgreSQL	15
1.3.2 Herramientas CASE	16

1.3.2.1 EasyCASE	17
1.3.2.2 ERwin.....	18
1.3.2.3 Embarcadero ER/Studio	19
1.3.2.4 Visual Paradigm	20
1.3.3 Metodologías.....	21
1.3.3.1 Metodología para el Diseño Conceptual de Almacenes de Datos.....	21
1.3.3.2 Metodología de Inmon.....	21
1.3.3.3 Metodología de Kimball.....	22
1.3.3.4 Metodología para el Desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio (BI) en la Línea de Almacenes de Datos e Inteligencia de Negocio (DW&BI) de DATEC.....	22
1.4 Conclusiones.....	26
Capítulo 2: Análisis y diseño del mercado de datos	27
2.1 Introducción	27
2.2 Análisis	27
2.2.1 Definición del Negocio	27
2.2.2 Temas de análisis	27
2.2.3 Roles y permisos.....	28
2.2.4 Reglas del negocio.....	28
2.2.5 Necesidades de los usuarios.	29
2.2.6 Requisitos de información.....	30
2.2.7 Requisitos Multidimensionales.....	31
2.2.8 Requisitos funcionales.....	33
2.2.9 Requisitos no funcionales.	34
2.2.10 Casos de uso del sistema.....	35
2.2.10.1 Casos de uso de información.	35

2.2.10.2 Casos de uso funcionales	35
2.3 Diseño	36
2.3.1 Matriz BUS	36
2.3.2 Modelo de Datos	37
2.3.2.1 Dimensiones	37
2.3.2.2 Tablas de hechos.....	41
2.3.2.3 Medidas	46
2.3.4 Esquema de Seguridad.....	47
2.3.5 Política de respaldo y recuperación.....	48
2.4 Conclusiones.....	48
Capítulo 3: Implementación y Pruebas.....	49
3.1 Introducción del capítulo.....	49
3.2 Modelos de Datos Físico.....	49
3.2.1 Estructuras de Datos	49
3.2.2 Esquemas y tablas.....	49
3.2.3 Restricciones y Secuencias.....	50
3.2.4 Índices.....	52
3.3 Usuarios y Permisos	53
3.4 Carga de nomencladores.....	54
3.5 Guía de Implantación.....	54
3.5.1 Secuencia de Pasos	54
3.6 Validación y pruebas.....	55
3.6.1 Listas de Chequeo de Análisis.....	55
3.6.2 Validación de requisitos por el cliente.....	55
3.6.3 Lista de Chequeo de Diseño	56
3.6.4 Pruebas de Implantación	56

INDICE DE CONTENIDOS

3.7 Conclusiones.....	57
Conclusiones generales.....	58
Recomendaciones	59
Referencias Bibliográficas.....	60
Anexos.....	64

Índice de Figuras

Figura 1 Diferencias entre Mercado de Datos y Almacén de Datos.	12
Figura 2 : Relación de los grupos con los flujos de trabajo.	23
Figura 3 Tabla del Hecho Ámbito Ambiental.	42
Figura 4 Tabla del hecho Cuencas Hidrográficas.	43
Figura 5 Tabla del hecho Principales bahías.	44
Figura 6 Tabla del hecho Ejecución física.	45
Figura 7 Tabla del hecho Ejecución física específica.	46

Índice de Tablas

Tabla 1: Matriz BUS.....	36
Tabla 2 Esquemas y Tablas.....	50
Tabla 3: Llaves primarias relacionadas con la BD.....	51
Tabla 4: Llaves foráneas de la tabla de hecho ejecución física	51
Tabla 5 Índices de las tablas de la BD.....	52
Tabla 6 Usuarios y roles de la BD.....	53
Tabla 7 Resultados de las listas de chequeo aplicadas.....	56

Introducción

Durante las décadas de los treinta a los setenta del siglo XX, se introdujo la estadística en los centros de investigación y en la producción industrial. En Cuba el órgano rector de la estadística es la Oficina Nacional de Estadística (ONE). Ésta tiene como objetivo fundamental captar, analizar y difundir los datos recogidos a lo largo y ancho de todo el país en áreas como: Educación, Salud, Transporte, Inmigración, Turismo, Medio Ambiente, Ocupación, Demografía e Inversiones.

En el año 2002 se crea la Universidad de las Ciencias Informáticas (UCI), la cual tiene como misión:

1. Formar profesionales, comprometidos con su Patria, altamente calificados en la rama de la informática.
2. Producir software y servicios informáticos, a partir de la vinculación estudio-trabajo como modelo de formación.

En la UCI se encuentra el Centro de Tecnologías y Gestión de Datos (DATEC), el cual desarrolla diferentes herramientas entre las que se encuentran las de almacenamiento de datos. Muchos de los software que se construyen son utilizados por los diferentes centros del país, como la ONE, ayudándoles a mejorar el almacenamiento de información.

El Departamento de Industria y Medio Ambiente, forma parte de la ONE, dicho departamento tiene especialistas que se encargan de atender las inversiones, los cuales brindan información de los recursos que constituyen reservas técnicas y financieras de los programas medioambientales. Permitiéndole a los distintos órganos del estado mantener el control sobre los beneficios que deban capitalizar los fondos de reserva y además la incorporación de recursos frescos al portafolio de inversiones.

La ONE cuenta con una serie de modelos estadísticos, en los cuales se recoge la información de todos los sectores de la economía y la sociedad. Para el trabajo con las inversiones existen los modelos 5702, 5708 y 1006; que se dedican a brindar información del plan de la ejecución física de las inversiones, al comportamiento de las inversiones seleccionadas y a las encuestas de los gastos de inversión en actividades de protección para el medio ambiente respectivamente.

Muchas de las necesidades informativas o análisis deseados sobre las inversiones se dificultan pues no cuentan con el soporte para llevarlos a cabo. Debido a que el almacenamiento de estos datos está en formatos de difícil acceso para su consulta, hace compleja su utilización, imposibilitando divulgar la información que contienen.

Partiendo de lo anteriormente expuesto, se plantea el siguiente **problema de la investigación**:

¿Cómo lograr la integración de los datos para desarrollar un sistema sobre los análisis de la información referida a las inversiones que se almacenan en la ONE?

Para brindar solución al problema de la investigación, se plantea el siguiente **objetivo general**: desarrollar un mercado de datos (MD) para ayudar al análisis de las estadísticas en la temática de inversiones por parte de los especialistas de la ONE.

Para darle cumplimiento al objetivo general se trazaron los siguientes **objetivos específicos**:

1. Elaborar el marco teórico de la investigación acerca de las principales tendencias de implementación de los almacenes de datos, los mercados de datos y estudio del arte sobre inversiones.
2. Realizar el análisis de los Modelos relacionados con las inversiones de la ONE.
3. Diseñar el MD de las inversiones para el almacén de datos(DW) de la ONE.
4. Implementar y cargar los clasificadores para el MD de las inversiones para el almacén de datos de la ONE.
5. Validar la solución desarrollada mediante la realización de pruebas, dentro de las que se encuentran Lista de chequeo especificación de requisitos, Lista de Chequeo Modelo de Datos y Caso de prueba para la guía de implantación.

Por tanto, el **objeto de estudio** son los Almacenes de Datos. La investigación se enmarca en el siguiente **campo de acción**: construcción de un MD.

Idea a defender: si se lograra efectuar el análisis, diseño e implementación del MD para los indicadores sobre las inversiones y toda la documentación correspondiente, se podría obtener una integración más completa de la información y se tendría una mejor disponibilidad de ésta, lo que favorecería al proceso de toma de decisiones mediante la informatización de los datos.

Para poder lograr el cumplimiento de los objetivos se proponen las siguientes **tareas científicas**:

1. Estudio y definición de conceptos sobre los temas relacionados con el desarrollo del MD.
2. Estudio y definición de conceptos sobre los temas relacionados con la disciplina de inversiones tanto en el ámbito mundial como en Cuba.
3. Estudio y definición de conceptos sobre la metodología a utilizar en el desarrollo.
4. Realización de entrevistas sistemáticas al personal especializado del área temática de inversiones y de la dirección de informática de la ONE.
5. Identificación de la estructura de usuarios y permisos.
6. Definición de temas de análisis.
7. Identificación de necesidades de información, requisitos funcionales y no funcionales.
8. Modelación de requerimientos.
9. Validación de requerimientos.
10. Definición de requisitos de entrada y de salida.
11. Elección de la granularidad del proceso del negocio.
12. Definición de las dimensiones del MD.
13. Definición de los hechos asociados a las dimensiones definidas.
14. Estructuración del modelo dimensional.
15. Transformación del modelo dimensional al diseño físico.
16. Implementación de la Base de Datos(BD).
17. Montaje de los clasificadores para MD del área de inversiones para el DW de la ONE.
18. Realización de las pruebas al MD.

Métodos Científicos:

Métodos teóricos

- Histórico-Lógico: se realizó un estudio del estado de arte de los almacenes de datos y mercados de datos con el objetivo de conocer más acerca de su evolución y desarrollo en el transcurso de la historia, además de conocer cómo trabaja la ONE el tema de inversiones, para tener una mejor idea del ambiente en el que se desarrollará la aplicación.
- Analítico-Sintético: se hizo un estudio de la bibliografía referente a diferentes conceptos asociados al tema de almacén de datos y mercados de datos. Además de hacer un

análisis de las tecnologías y herramientas utilizadas en la investigación basándose en la revisión de documentos e informes.

- Modelación: se utilizó a la hora del diseño de la BD del MD como posible solución a las dificultades existentes en este entorno, siendo éste de gran importancia para la modelación de los requerimientos y así mostrar con mayor claridad la forma en que se presentará la solución.

Método empírico

- Entrevista: se realizaron entrevistas a especialistas de la ONE para obtener información del proceso de negocio que actualmente se lleva a cabo en dicha institución.

El presente trabajo está compuesto por tres capítulos:

Capítulo 1 - Fundamentación teórica: incluye el estado del arte del tema tratado, describe el objeto de estudio, el campo de acción y la situación problemática presente. Además, se selecciona la plataforma, lenguaje, herramientas y metodología con las que se llevará a cabo la propuesta.

Capítulo 2 - Análisis y diseño del MD: se identifican las especificaciones de requisitos que se debe tener en cuenta para el desarrollo del sistema, además una breve descripción de cómo se construirá la propuesta de solución.

Capítulo 3 – Implementación y validación del MD: se definen los elementos relacionados a la implementación del sistema, las dimensiones y sus hechos asociados. También se implementa la BD , se montan los clasificadores, se realizan las pruebas y la validación de la solución.

Capítulo 1: Fundamentación Teórica

1.1 Introducción

En este capítulo se exponen los conceptos fundamentales asociados al negocio, los cuales permitirán una mejor comprensión del entorno del objeto de estudio, por parte de las personas implicadas en el desarrollo de la aplicación. Además, se hace un análisis de las tendencias actuales en las metodologías y tecnologías, para el desarrollo de sistemas informáticos, que pudieran ser adecuadas para el desarrollo del MD.

1.2 Conceptos fundamentales asociados al negocio

1.2.1 Almacén de Datos

1.2.1.1 Definiciones generales

A continuación se darán algunos conceptos dados por diferentes especialistas sobre el tema de Almacén de Datos (del inglés *Data Warehouse*, DW).

John Edwards en su artículo de CIO Magazine, define DW como: "Un Data Warehouse toma información de múltiples sistemas de BD y la almacena de una manera que está diseñada para dar a los usuarios acceso más rápido, más fácil y más flexible a los aspectos claves". (Shinin, 2006)

W.H. Inmon que es considerado como el padre de las bases de datos, en 1992 plantea que: "Un DW es una colección de datos orientados a temas, integrados, no-volátiles y variante en el tiempo, organizados para soportar necesidades empresariales". Existen muchas definiciones para DW, pero la más conocida es ésta. (Forno, 2009)

En 1993, Susan Osterfeldt publica una definición de DW: "Yo considero al DW como algo que provee dos beneficios empresariales reales: Integración y Acceso de datos. DW elimina una gran cantidad de datos inútiles y no deseados, como también el procesamiento desde el ambiente operacional clásico". (Romero, 2006)

Ralph Kimball define un DW como: "una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis". (Lara, 2009)

Un DW es "un conjunto integrado de bases de datos que se diseña y utiliza para apoyar a la toma de decisiones y en él cada unidad de datos es relevante en algún instante de tiempo. Un

Data Warehouse es un enorme almacén de datos que contiene información no sólo de bases de datos relacionales, sino de otras fuentes relacionadas con la actividad de la empresa y cuya finalidad no sólo se centra en el almacenamiento de esos datos, sino en su análisis para la obtención de información estructurada y en definitiva útil para la toma de decisiones” (DELGADO, 1999)

Se puede decir que un DW es una colección de datos que apoya a la toma de decisiones, está alimentado de numerosas fuentes que están transformadas en grupos de información orientada a algunos temas específicos. Además se puede decir que es un repositorio o un lugar de depósito de información integrada que sirve para el acceso o facilitar búsquedas.

1.2.1.2 Características

Un Almacén de Datos tiene las siguientes características: (Velasco R. H., 2009)

- Organizado en torno a temas.
- La información se clasifica en base a los conceptos que son de interés para la empresa.
- Integrado.
- Es el aspecto más importante. La integración de datos consiste en convenciones de nombres, codificaciones consistentes y medida uniforme de variables.
- Dependiente del tiempo.

Esta dependencia aparece de tres formas:

- La información representa los datos sobre un horizonte largo de tiempo.
- Cada estructura clave contiene (implícita o explícitamente) un elemento de tiempo (día, semana, mes).
- La información, una vez registrada correctamente, no puede ser actualizada.
- No volátil.

El DW sólo permite cargar nuevos datos y acceder a los ya almacenados, pero no permite ni borrar ni modificar los datos.

1.2.1.3 Beneficios

Un almacén de datos puede dar sin lugar a dudas importantes beneficios, su utilización permitirá que la información de gestión sea: accesible, correcta, uniforme y actualizada. Estas características con otra serie de aspectos dan lugar a la obtención de un conjunto de beneficios, que podríamos resumir de la siguiente manera: (López D. A., 2005)

- Simplifica los procesos de toma de decisiones porque ofrece imágenes integradas de los datos.
- Facilita el proceso de comparación, proyección a futuro, relación con otros datos, muestra de indicadores e información consolidada.
- El acceso a los datos es fácil y rápido, permitiendo a los usuarios hacer sus propias consultas.
- Ayuda a mejorar el buen funcionamiento de los sistemas operacionales retroalimentando demandas para los sistemas transaccionales.

1.2.2 Modelo Dimensional

El modelo dimensional es el más usado en los sistemas de almacenes de datos. Está basado en dimensiones, las cuales representan categorías de información, atributos que representan un único nivel dentro de una dimensión, pueden existir jerarquías de atributos las cuales expresan relaciones entre diferentes atributos y por último tablas de hechos, las cuales contienen datos de interés que presentan un nivel de granularidad. (Vázquez, 2009)

1.2.2.1 Conceptos asociados al modelo dimensional

Hechos

Son medidas numéricas, las cuales describen los resultados (rendimiento) del negocio. Un hecho debe ser un valor cuantificable o medible. Por ejemplo el número de unidades vendidas, ingreso por ventas. (Herrera, 2007)

Dimensiones

Las dimensiones representan cada uno de los ejes en un espacio multidimensional. Suministran el contexto en el que se obtienen las medidas de un hecho. Algunos ejemplos son: módulo, punto, condición, técnico y tiempo. Las mismas se utilizan para seleccionar y agrupar los datos en un nivel de detalle deseado. Los componentes de una dimensión se denominan niveles y se organizan en jerarquías. La dimensión tiempo puede tener niveles día, mes y año. (Fránquiz, 2009)

Granularidad

Es el nivel más bajo de información que será almacenado en la tabla de hechos e indica el grado de detalle asociado a un hecho particular. La granularidad depende directamente del

número de dimensiones que se asocian a la tabla de hechos. El primer paso para diseñar una tabla de hechos es determinar la granularidad.

El gran factor y el más decisivo de la granularidad es el tiempo, ya que mientras menor sea el intervalo de tiempo, mayor será el grado de detalle obtenido. Se deben considerar otros factores como la carga del procesador, espacio de almacenamiento y satisfacer a cabalidad los requerimientos del cliente.

Jerarquía

Es un conjunto de atributos descriptivos que se rigen por un orden preestablecido. Una jerarquía implica una organización de niveles dentro de una dimensión, con cada nivel representando el total agregado de los datos del nivel inferior. Además, definen como los datos son sumados desde los niveles más bajos hasta los más altos. Una jerarquía puede contener todos los valores que se encuentran en una dimensión, pero no es necesario. (Herrera, 2007)

1.2.2.2 Esquemas que se encuentran dentro del modelo dimensional

Esquema Estrella

Se conoce además como modelo multidimensional, pues su estructura base es similar: una tabla central y un conjunto de tablas que la atienden radialmente. Su nombre se deriva del hecho que su diagrama forma una estrella, con puntos radiales desde el centro. Este modelo también es asimétrico pues hay una tabla dominante en el centro con varias conexiones a otras tablas. (Herrera, 2007)

Esquema copo de nieve (snowflake)

Es una estructura más compleja que el esquema de estrella. Se da cuando las dimensiones se implementan con más de una tabla de datos. Aunque puede reducir espacio por la mínima redundancia de datos, tiene la contrapartida de peores rendimientos al tener que crear más tablas de dimensiones y más relaciones entre las tablas lo que tiene un impacto directo sobre el rendimiento. (Vázquez, 2009)

Además se puede decir que es una variante del esquema de estrella, las tablas de dimensión están normalizadas, es decir, pueden incluir claves que apuntan a otras tablas de dimensión. (Rivas, 2006)

Esquema constelación

Este esquema se puede decir que es más complejo que los demás, pues contiene múltiples tablas de hechos que comparten tablas de dimensiones que se visualizan como una constelación de hechos. (Vázquez, 2009)

Para la selección del esquema que se encuentra dentro del modelo dimensional se tuvo en cuenta que:

- El sistema de constelación es más complejo que el esquema copo de nieve y estrella, pues tiene muchas tablas de hechos que comparten tablas de dimensiones.
- El esquema copo de nieve es más complejo y ocupa menos espacio que el esquema estrella.
- El esquema estrella es el más simple y tiene mejor rendimiento y velocidad que los anteriores.

El espacio en disco no suele ser el problema y sí el rendimiento, por tanto se escoge el esquema estrella.

1.2.3 Proceso OLAP

La tecnología de Procesamiento Analítico en línea se caracteriza por ser un análisis multidimensional de datos corporativos, que soportan los análisis del usuario y unas posibilidades de navegación, seleccionando la información a obtener.

Existen diferentes tipos de OLAP: (Torres, 2007)

- **ROLAP (Procesamiento Analítico en línea Relacional):** es capaz de usar datos pre calculados si éstos están disponibles, o de generar dinámicamente los resultados desde los datos elementales si es preciso. Esta arquitectura accede directamente a los datos del almacén de datos y soporta técnicas de optimización de accesos para acelerar las consultas.
- **MOLAP (Procesamiento Analítico en línea Multidimensional):** usa unas BD multidimensionales para proporcionar el análisis, su principal premisa es que está mejor implantado almacenando los datos multidimensionalmente.
- **HOLAP (Procesamiento Analítico en línea Híbrido):** permiten un análisis híbrido de la información, uniendo lo mejor de las arquitecturas ROLAP y MOLAP. Utiliza tecnología

multidimensional para un mejor desempeño. Cuando se necesita llegar a la información detallada, usa técnicas de datos relacionales para llegar a ésta.

1.2.4 Mercado de datos

1.2.4.1 Definiciones generales

Es una versión especial de un almacén de datos. Como los DW, los MD contienen una visión de datos operacionales que ayudan a decidir sobre estrategias de negocio basadas en el análisis de tendencias y experiencias pasadas. La creación de un MD es específica para una necesidad de datos seleccionados, enfatizando el fácil acceso a una información relevante. Se destaca por una definición de requerimientos más fácil y rápida. También se simplifica el desarrollo de todo el mecanismo de su BD y con ello baja substancialmente todo el coste del proyecto, así como su duración. Normalmente, un MD resuelve aplicaciones a nivel departamental, aunque en ocasiones se desarrolla una aplicación que integre todas ellas. (Soto, 2009)

Son un conjunto de modelos de negocio relacionados entre sí. Los MD son diseñados para satisfacer las necesidades específicas de grupos comunes de usuarios (divisiones geográficas y divisiones organizacionales). Son generalmente, subconjuntos del almacén de datos, pero pueden también integrar un número de fuentes heterogéneas, e inclusive ser más grandes, en volumen de datos, que el propio almacén central.

El concepto MD es una extensión natural del DW y está enfocado a un departamento o área específica, como por ejemplo los departamentos de Finanzas o Marketing. Permitiendo así un mejor control de la información que se está abarcando. (Torres, 2007)

Es una BD departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Dispone de una estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. Puede ser alimentado desde los datos de un DW, o integrar por sí mismo un compendio de distintas fuentes de información. (Díaz, 2009)

Se puede resumir, basado en los conceptos antes mencionados, como un MD, un conjunto de modelos de negocio relacionados que son creados para satisfacer las necesidades específicas de un grupo de usuarios.

1.2.4.2 Características

Entre las características de los mercados de datos se destacan: (Herrera, 2007)

- Posee usuarios limitados.
- Tiene un área específica.
- Tiene un propósito específico.
- Posee una función de apoyo.
- Da a los usuarios acceso a los datos que les hace falta para analizarlos con más frecuencia.
- Provee los datos de una manera que coincida con la vista colectiva de los datos por un grupo de usuarios en un departamento o función de negocio.
- Mejora el tiempo de respuesta al usuario final por causa de la reducción en el volumen de información a ser accedido.
- Provee datos estructurados apropiadamente para satisfacer los requerimientos de las herramientas de acceso de usuario final.
- Usan menos datos de manera tal que la tarea de carga de datos es más fácil y de esta forma la implementación de un MD es más simple comparado con un almacén de datos corporativo.
- El costo de implementación de un MD normalmente es menor que el requerido para establecer un almacén de datos.

1.2.4.3 Beneficios

Existen muchos beneficios de los MD, entre los más importantes están: (Bernabeu, 2009)

- Fácil acceso a los datos que se necesitan de forma frecuente.
- Crea vista colectiva para grupos de usuarios.
- Mejora el tiempo de respuesta al usuario final.
- Facilidad de creación.
- Costo inferior de la aplicación de un completo de DW.
- Son simples de implementar.
- Conllevan poco tiempo de construcción y puesta en marcha.
- Permiten manejar información confidencial.

- Reflejan rápidamente sus beneficios y cualidades.
- Reducen la demanda del depósito de datos.

1.2.5 Diferencias entre Mercado de Datos y Almacén de Datos

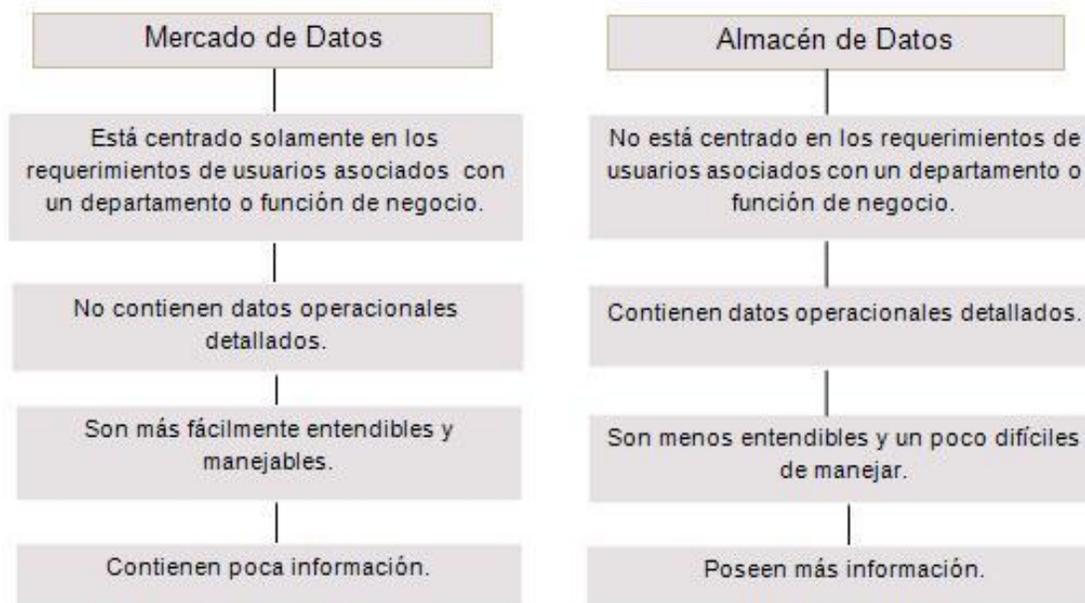


Figura 1 Diferencias entre Mercado de Datos y Almacén de Datos.

1.2.6 Modelos estadísticos del área temática de Inversiones. (Modelo Sien, 2009)

- Control del plan de la ejecución física de las inversiones (5702). El objetivo de este modelo es conocer detalladamente el comportamiento de la ejecución física de las inversiones con independencia del pago de los trabajos, constituyendo la única fuente para determinar la ejecución de inversiones desde el nivel de centro informante hasta la nación, además de brindar un conjunto de indicadores que, de forma definitiva, permiten evaluar el cumplimiento del plan de ejecución de inversiones de acuerdo con sus diferentes niveles de desagregación y ser uno de los elementos de cálculo de las inversiones en proceso.
- Comportamiento de las inversiones seleccionadas (5708). El objetivo de este modelo es controlar las inversiones y programas seleccionados a partir del plan emitido por el Ministerio de Economía y Planificación. Éste permite conocer, entre otros aspectos, la

marcha del cumplimiento del plan anual, incluidas las causas que lo afectan, así como de los fondos básicos puestos en explotación.

- Encuesta sobre los gastos de inversión en actividades de protección para el medio ambiente (1006). El objetivo de este modelo es conocer de forma detallada el comportamiento de la ejecución física de los gastos de inversiones para la protección del medio ambiente, en la medida que se produce el avance físico de la inversión, independientemente que se haya realizado o registrado el pago por estos trabajos, constituyendo la fuente para evaluar, desde el nivel de centro informante hasta la nación, la ejecución física de las inversiones, de acuerdo con los principales problemas ambientales.

En esta etapa no se estudia el modelo 5708, solamente va a ser objeto de estudio los modelos 1006 y 5702.

1. 3 Tendencias y Tecnologías Actuales

1.3.1. Gestor de base de datos

Es una colección de programas cuyo objetivo es servir de interfaz entre la BD, el usuario y las aplicaciones. Se compone de un lenguaje de definición de datos, de un lenguaje de manipulación de datos y de un lenguaje de consulta. Un Sistema Gestor de Base de Datos (SGBD) permite definir los datos a distintos niveles de abstracción y manipular dichos datos, garantizando la seguridad e integridad de los mismos. (MENDEZ, 2009)

Para el desarrollo del MD del área de inversiones es necesario realizar un estudio de los diferentes gestores existentes, teniendo en cuenta sus características principales, valorando sus ventajas y desventajas, para determinar cuál de ellos es más factible.

1.3.1.1 Microsoft SQL Server

Microsoft con su software SQL Server nos ofrece una plataforma de gestión de datos muy óptima, al cual podemos acceder desde cualquier lugar y en cualquier momento. Con SQL Server se puede almacenar datos estructurados, semi-estructurados, no estructurados y documentos, tales como las imágenes; de forma directamente en la BD.

Además posee los más altos niveles de seguridad, fiabilidad y escalabilidad, para obtener los mejores resultados en aplicativos empresariales; pues con éste se puede reducir el costo y el tiempo para la gestión de datos y el desarrollo de aplicativos. (Rodríguez, 2008)

Ventajas:

- Escalabilidad: se adapta a las necesidades de la empresa, soportando desde unos pocos usuarios a varios. Empresas centralizadas u oficinas distribuidas, replicando cientos de sitios.
- Potencia: Microsoft SQL Server es la mejor BD para Windows NT Server.
- Posee los mejores registros del Protocolo de Transmisión de Control (TCP) tanto en transacciones totales como en coste por transacción.
- Gestión: con un completo interfaz gráfico que reduce la complejidad innecesaria de las tareas de administración y gestión de la BD.
- Orientada al desarrollo: Visual Basic, Visual C++, Visual J++, Visual Interdev, Microfocus Cobol y muchas otras herramientas son compatibles con Microsoft SQL Server. (Janny, 2006)

Walter Donnetta, Gerente de Sistemas, Grupo Sancor Seguros, plantea: “Con SQL Server 2008 tenemos una plataforma ágil, segura y fácilmente adaptable a los cambios que se producen constantemente en el mercado”. (MENDEZ, 2009)

1.3.1.2 Oracle

Oracle es un SGBD que presenta las siguientes características principales: (Velasco, 2005)

- Entorno cliente/servidor.
- Gestión de grandes BD.
- Usuarios concurrentes.
- Alto rendimiento en transacciones.
- Sistemas de alta disponibilidad.
- Disponibilidad controlada de los datos de las aplicaciones.
- Adaptación a estándares de la industria, como SQL-92.
- Gestión de la seguridad.
- Autogestión de la integridad de los datos.
- Opción distribuida.
- Portabilidad.

- Compatibilidad.
- Conectabilidad.
- Replicación de entornos.

Oracle gestiona la seguridad de la BD usando: (Velasco, 2005)

- Usuarios y esquemas de la BD.
- Privilegios.
- Roles.
- Ajustes de rendimiento y cuotas.
- Límites sobre los recursos.
- Auditoría.

Cada usuario tiene un dominio de seguridad, que determina cosas como: (Velasco, 2005)

- Acciones (privilegios y roles) disponibles para el usuario.
- Cuotas sobre *tablespaces*.
- Límites en los recursos del sistema.

1.3.1.3 PostgreSQL

PostgreSQL es un sistema de gestión de BD objeto-relacional, distribuido bajo licencia de software libre permisiva y con su código fuente disponible libremente. Es el sistema de gestión de BD de código abierto más potente del mercado y en sus últimas versiones no tiene nada que envidiarle a otras BD comerciales.

PostgreSQL utiliza un modelo cliente/servidor y usa multiprocesos en vez de multihilos para garantizar la estabilidad del sistema. (Martinez, 2010)

PostgreSQL es el SGBD de software libre que más se acerca a SGBDs más experimentados en *Business Intelligence*, como Oracle y Microsoft SQL Server. Sin embargo, para optimizarlo para trabajar con DW hay que ajustar algunos parámetros del servidor a mano. Además, existe una especie de distro de PostgreSQL llamada *Bizgre* que está más enfocada al trabajo con DW. (Aleman, 2009)

Características de PostgreSQL

PostgreSQL posee muchas características entre las que se destacan: (Martínez D. P., 2010)

- Soporta diferentes tipos de datos como son: fecha, monetarios, elementos gráficos, datos sobre redes y cadenas de bits, además del soporte para los tipos base y permite la creación de tipos propios.
- Incorpora una estructura de datos de arreglos.
- Soporta el uso de índices, reglas y vistas.
- Incluye herencia entre tablas y entre gestores de objetos- relacionales.
- Permite la declaración de funciones propias, así como la definición de disparadores.
- Permite la gestión de diferentes usuarios, como también los permisos asignados a cada usuario.

Mejoras de PostgreSQL 8.4 (Casanova, 2009)

- Restauración de BD 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 BD, lo cual hace a PostgreSQL más útil en entornos con múltiples idiomas.
- Nuevas herramientas de monitoreo de consultas que le otorgan a los administradores mayor información sobre la actividad del sistema.

Selección del Sistema Gestor de Bases de Datos

SQL Server y PostgreSQL son potentes SGBD, ambos pueden ser ejecutados sobre la plataforma más usada en todo el mundo: Windows; aunque SQL Server tiene la ventaja en este sentido, pues permite su ejecución sobre un mayor número de sistemas operativos. Se dice escoger PostgreSQL porque brinda mayor ventaja su utilización, ya que no es propiedad de ninguna compañía y eso favorece en este sentido su elección, debido a las condiciones de nuestro país, pues es un producto de una comunidad de desarrolladores.

1.3.2 Herramientas CASE

CASE es una sigla, que corresponde a las iniciales de: **C**omputer **A**ided **S**oftware **E**ngineering; y en su traducción al Español significa Ingeniería de Software Asistida por Computación.

El concepto de CASE es muy amplio y una buena definición genérica, que pueda abarcar esa amplitud de conceptos, sería: la aplicación de métodos y técnicas a través de las cuales se

hacen útiles a las personas comprender las capacidades de las computadoras, por medio de programas, de procedimientos y su respectiva documentación.

Las herramientas CASE representan una forma que permite modelar los procesos de negocios de las empresas y desarrollar los sistemas de información gerenciales. (Perissé, 2001)

1.3.2.1 EasyCASE

EasyCASE es la base de los productos para procesos, modelamiento de datos y eventos e ingeniería de BD y genera esquemas de BD e ingeniería reversa. Esta herramienta permite automatizar las fases de análisis y diseño dentro del desarrollo de una aplicación, para poder crear las aplicaciones eficazmente, desde el procesamiento de transacciones a la aplicación de BD de cliente/servidor, así como sistemas de tiempo real.

EasyCASE Profesional, una herramienta multi-usuario, es ideal para aquellos que necesitan compartir datos y trabajar en un proyecto con otros departamentos. Para asegurar la seguridad de los datos, existe el diagrama y diccionario de los datos que bloquean por niveles al registro, al archivo, al proyecto y niveles de control de acceso. (Martínez, 2010)

Bases de datos que soporta:

- Oracle
- Paradox
- Progress
- SQLBase
- SQL Server
- Sybase
- Watcom SQL
- Access
- ANSI SQL
- Clipper
- dBASE III, IV, V
- DB2
- FoxPro
- Informix (Martínez, 2010)

1.3.2.2 ERwin

ERwin es una herramienta de BD que le ayuda a diseñar, generar y mantener aplicaciones de BD de calidad y alto rendimiento. Desde un modelo lógico de sus requerimientos de información y reglas del negocio que definen su BD, hasta un modelo físico, optimizado por las características específicas de su BD de destino, ERwin le permite visualizar la estructura adecuada, los elementos clave y un diseño optimizado de su BD. ERwin genera tablas automáticamente y miles de líneas de *stored procedures* y código *trigger* para las principales bases de datos. Su tecnología "complete-compare" permite el desarrollo interactivo, de manera que su modelo está siempre sincronizado con su BD. A través de la integración con los ambientes de desarrollo líderes en la industria, ERwin también acelera la creación de aplicaciones data-centric. (Nazario, 2002)

Software para Aplicaciones Compatibles (Martínez, 2010)

- NetDynamics
- PowerBuilder
- PROGRESS
- Visual Basic

Bases de Datos Compatibles (Martínez, 2010)

- CA-Clipper
- CA-OpenIngres
- DB2 for MVS
- DB2 for OS/390
- DB2 UDB
- dBASE
- FoxPro
- HiRD
- Informix
- InterBase
- Microsoft Access
- Microsoft SQL Server
- Oracle

- Paradox
- Rdb
- red Brick Warehouse
- SAS
- SQL Anywhere
- SQLBase
- Sybase
- Teradata

Sistemas Operativos Compatibles (Martínez, 2010)

- Windows NT
- Windows 95
- Windows 98

Requerimientos Técnicos (Martínez, 2010)

Mínimo 10 MB de espacio de disco duro, 16 MB RAM (32 MB RAM recomendado para modelos largos).

1.3.2.3 Embarcadero ER/Studio

Es una herramienta de modelado de datos fácil de usar y multinivel, para el diseño y construcción de bases de datos a nivel físico y lógico. Direcciona las necesidades diarias de los administradores de bases de datos, desarrolladores y arquitectos de datos que construyen y mantienen aplicaciones de bases de datos grandes y complejas. Soporta BD entre las que están Oracle, Sybase System, Microsoft SQL Server 6.5, Microsoft Visual FoxPro. (López, 2009)

ER/Studio posee numerosas funcionalidades que aceleran el ciclo de diseño, entre las que están: (López, 2009)

- Capacidad fuerte en el diseño lógico.
- Sincronización bidireccional de los diseños lógico y físico.
- Construcción automática de BD.
- Reingeniería inversa de BD.
- Documentación basada en HTML.
- Un repositorio para el modelado.

ER/Studio provee los siguientes beneficios: (Chavez, 2008)

- Una interface de maquinaria de búsqueda que resulta inmediatamente familiar a los usuarios finales de negocio.
- Reporteo de autoservicio que aumenta el conocimiento y productividad del trabajador con documentación bajo demanda.
- Distribución vía web que permite estandarización global y la utilización de activos de datos a cualquier organización.
- Instalación e implementación simple que permite la comunicación con los metadatos en minutos.

1.3.2.4 Visual Paradigm

Es una herramienta de modelado UML profesional que soporta el ciclo de vida completo del desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue.

Además, es una herramienta colaborativa, es decir, soporta múltiples usuarios trabajando sobre el mismo proyecto; genera la documentación del proyecto automáticamente en varios formatos como Web o .Pdf y permite control de versiones.

Cabe destacar igualmente su robustez, usabilidad y portabilidad.

En definitiva, Visual Paradigm es una herramienta muy a tener en cuenta a la hora de ponerse manos a la obra con un proyecto importante. (Pavón, 2007)

Esta herramienta tiene **características** entre las que se encuentran: (Zapata, 2005)

- Producto de calidad.
- Soporta aplicaciones web.
- Varios idiomas.
- Generación de código para Java y exportación como HTML.
- Fácil de instalar y actualizar.
- Compatibilidad entre ediciones.

Selección de la herramienta CASE a utilizar

Teniendo en cuenta las herramientas estudiadas, se ha llegado a la conclusión de que la más óptima para realizar el diseño del MD es Visual Paradigm, permitiendo un fácil acceso a cualquier BD. Realiza de una forma íntegra los planes de construcción del software. Tiene una comparación comprensiva entre los modelos de los datos y la BD soportando también su

diseño para los DW. Además soporta la separación del modelo lógico y del físico, soporta sincronización de modelo a modelo y modelo a script. Visual Paradigm genera de forma automática códigos desde diagramas además brinda la posibilidad de generar documentación y es compatible con el gestor de BD PostgreSQL.

1.3.3 Metodologías

Las metodologías son parte del proceso de investigación que posibilita la sistematización de los métodos y de las técnicas necesarias para llevarlas a cabo.

Hoy en día existen numerosas propuestas metodológicas que inciden en distintas dimensiones del proceso de desarrollo de un MD. A continuación se profundizará un poco acerca de estas metodologías, para una mejor comprensión del estudio de la investigación.

1.3.3.1 Metodología para el Diseño Conceptual de Almacenes de Datos

Esta metodología fue resultado de la tesis doctoral de Leopoldo Zenaido Zepeda Sánchez, desarrollada en España en marzo del 2008 se enmarca en el área de diseño de almacenes de datos con la metodología para el diseño conceptual. Este método es empleado con una metodología compuesta y consiste en tres fases.

La primera está dedicada a examinar el esquema Entidad –Relación de la BD operacional, generando los esquemas multidimensionales candidatos para el DW.

En la segunda fase, los requisitos de usuarios son recogidos por medio de entrevistas. El propósito de las entrevistas es obtener información acerca de las necesidades de análisis de los usuarios. Como base para esta fase, adaptamos un método de elicitación de requisitos basado en metas.

La tercera fase, contrasta la información obtenida en la segunda fase, como los esquemas multidimensionales candidatos formados en la primera fase, generando así, la mejor solución que mejor reflejan los requisitos de usuarios. (Giménez, 2008)

1.3.3.2 Metodología de Inmon

La metodología desarrollada por Bill Inmon, es una metodología descendente (*top-down*) a la hora de diseñar un almacén de datos, ya que de esta forma se consideraran mejor todos los datos corporativos. Propone construir primero el DW y a partir de este los MD y además

plantea la creación de un repositorio de datos corporativos como fuente de información consolidada, persistente, histórica y de calidad.

1.3.3.3 Metodología de Kimball

Kimball aseguró en 1997 su modelo multidimensional era "la única manera viable de diseñar BD destinadas a su uso directo por parte de un usuario final". (González, 2006)

La metodología desarrollada por Ralph Kimball, de arquitectura ascendente (*bottom-up*), se enfoca principalmente en el diseño de la BD que almacenará la información para la toma de decisiones. Este diseño se basa en la creación de tablas de hechos, que son los que contienen la información numérica de los indicadores a realizar, o sea la información cuantitativa para la toma de decisiones, que dichas tablas se relacionan las tablas de dimensiones, las cuales contienen la información cualitativa de los indicadores, es decir, toda aquella información que clasifique la información requerida.

1.3.3.4 Metodología para el Desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio (BI) en la Línea de Almacenes de Datos e Inteligencia de Negocio (DW&BI) de DATEC.

A continuación se describe el ciclo de vida de la metodología, explicando cada uno de los flujos de trabajo, desde la concepción del proyecto, hasta la entrega al cliente final. Se describen los grupos creados, sus actividades fundamentales, los roles que participan y se mencionan los artefactos elaborados.

Grupos/ Flujos	Estudio Preliminar	Requerimientos	Arquitectura y Diseño	Implementación	Prueba	Despliegue	Soporte y Mantenimiento
Análisis	Responsable	Responsable	Participa	No Participa	Responsable	No Participa	No Participa
Almacén	Participa	No Participa	Responsable	Responsable	Participa	Responsable	Participa
ETL	Participa	No Participa	Responsable	Responsable	Participa	Responsable	Participa
BI	Participa	No Participa	Responsable	Responsable	Participa	Responsable	Participa
Dirección	Responsable	Responsable	Responsable	Responsable	Responsable	Responsable	Participa

Legenda:
 Responsable
 Participa
No Participa

Figura 2 : Relación de los grupos con los flujos de trabajo.

Ciclo de vida de la metodología

Flujos de Trabajo:

- Estudio Preliminar o Planeación: se hace un estudio detallista en la entidad cliente. Esto incluye un diagnóstico de información, de datos y de infraestructura tecnológica, todo esto se basa en determinar qué es lo que se desea construir y qué condiciones existen para el desarrollo y montaje de la misma. Además, se llevan a cabo tareas de planeación del proyecto, se definen los objetivos, el alcance preliminar, los costos estimados y los recursos necesarios. Los resultados más importantes son: Ficha Técnica, Proyecto Técnico, Cronograma General de Ejecución, Roles y Responsabilidades, Presupuesto, Importe de la Solución y Contrato.
- Requerimientos: es llevado a cabo por el grupo de Análisis, se enfoca hacia dos direcciones, la primera identifica las necesidades de información y reglas de negocio y la otra hace un levantamiento minucioso de cada una de las distintas fuentes de datos a integrar. Además, se definen los requerimientos a partir de una comparación de las necesidades y las reglas del negocio con los elementos disponibles en las fuentes. El resultado más importante es la lista de Requerimientos.

- **Arquitectura y Diseño:** aquí participan los tres grupos fundamentales Extracción, Transformación y Carga (ETL), Almacén de Datos (DW) e Inteligencia de Negocios (BI). Se especifican las estructuras de almacenamiento, se diseñan las reglas de extracción, transformación y carga, así como la arquitectura de información que regirá al desarrollo de la solución. Los resultados más importantes son: Arquitectura del Sistema, Modelo de Datos del Repositorio Corporativo y las Reglas de Extracción, Transformación y Carga de Datos.
- **Implementación:** participan los tres grupos de desarrollo (ETL, DW y BI). Se lleva a cabo el diseño físico del Repositorio de Datos, se crean las estructuras de almacenamiento con las particiones y agregaciones adecuadas para la solución del problema en desarrollo. Se crea el Área Temporal de Almacenamiento, donde se hacen ajustes para integrar la información necesaria a través de las reglas de ETL. Se configuran e implementan las herramientas de BI para obtener los reportes, gráficos, mapas que cubran los requerimientos firmados con el Cliente Final. Los resultados más importantes son: Repositorio de Datos, Área Temporal de Almacenamiento, Reglas de ETL, configuración y personalización de las Herramientas de BI.
- **Prueba:** aquí se realizan pruebas como son las Pruebas de Unidad, que son las que se llevan a cabo por los propios desarrolladores de cada uno de los grupos, luego las Pruebas de Integración y Sistema, hasta las Pruebas de Aceptación con el cliente final. Los principales resultados son: Solicitudes de Cambios, No Conformidades y Acta de Aceptación.
- **Despliegue:** este flujo consiste consta de dos etapas, la primera es un Despliegue Piloto, donde se configuran los servidores necesarios y se instalan las herramientas según la arquitectura definida, se cargan muestras de los datos en un ambiente controlado, con el fin de demostrarle al cliente final que la solución funciona, luego se realiza la carga histórica de los datos y es aquí donde se lleva a cabo la Capacitación y Transferencia Tecnológica. El resultado fundamental es la solución desplegada en el entorno real y el correcto funcionamiento.
- **Soporte y Mantenimiento:** en este flujo de trabajo ya la solución del problema está implantada y en explotación, según el contrato firmado y las condiciones de soporte

establecidas, son los servicios que brindan, que pueden ser desde un soporte en línea, vía telefónica, web o correo.

- Gestión y Administración del Proyecto: este flujo transcurre a lo largo de todo el ciclo de vida, es llevado a cabo por el Grupo de Dirección del Proyecto, es aquí donde se controla, gestiona y chequea todo el desarrollo, los gastos, las utilidades, los recursos y las adquisiciones relacionadas con la gestión del proyecto.

Grupos de Trabajo

Estas metodologías están estructuradas por cuatro grupos, cada uno de estos realiza actividades específicas y bien delimitadas según sus responsabilidades dentro de un proyecto para darle cumplimiento a todas las tareas asignadas, además tienen la función de generar sus propios artefactos.

- Grupo de Análisis.
- Grupo de Almacenes de Datos (DW).
- Grupo de Extracción, Transformación y Carga (ETL).
- Grupo de Inteligencia de Negocio (BI).
- Grupo de Dirección.

Selección de la metodología a utilizar

Teniendo en cuenta las metodologías antes mencionadas, se considera que la Metodología para el Desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio (BI) en la Línea de Almacenes de Datos e Inteligencia de Negocio (DW&BI) de DATEC, es la más adecuada para la realización del MD para el área temática de inversiones en la ONE, ya que ésta usa como base la Metodología de Kimball, que está basada en crear los conceptos de los hechos y dimensiones, se puede llegar a la solución en un corto período de tiempo. Además propone ir construyendo el DW a través de la construcción de los MD departamentales, lo que constituye una estrategia buena y coincide con la división lógica de las empresas, entidades y organismos, teniendo además abundante documentación. Como complemento a la misma y fortaleciendo la etapa del levantamiento de requisitos, se tomó lo planteado por Leopoldo Zenaido Zepeda Sánchez en su Tesis de Doctorado, orientando así el trabajo a los Casos de Uso y se logra estar más alineado con las tendencias y normas de la Universidad.

1.4 Conclusiones

En este capítulo se hizo un estudio del estado del arte de las tecnologías que se utilizarán, además se eligieron las herramientas, metodologías utilizadas, así como el gestor de BD para un mejor desarrollo del MD.

Teniendo en cuenta el análisis realizado previamente, sobre las tendencias que existen en la actualidad acerca de las tecnologías y herramientas más usadas en el campo de la informática, se toma como decisión desarrollar un MD, aprovechando las potencialidades que brinda el gestor de BD PostgreSQL, se propone además, para garantizar la realización de un MD adecuado, el uso de la herramienta Visual Paradigm, por las facilidades que ésta brinda y todo este proceso será controlado y orientado por la Metodología para el Desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio (BI) en la Línea de Almacenes de Datos e Inteligencia de Negocio (DW&BI) de DATEC.

Capítulo 2: Análisis y diseño del mercado de datos

2.1 Introducción

El siguiente capítulo está enmarcado en exponer las características del sistema propuesto para que cumpla con las expectativas del cliente y usuarios finales. Inicialmente se realiza una descripción detallada de la definición del negocio, los temas de análisis, se tienen en cuenta los distintos roles tratados así como los permisos. Además, se hace una definición de las reglas del negocio para que los clientes tengan un mejor conocimiento de todo el proceso de desarrollo del MD. Se exponen los requisitos de información, los funcionales, no funcionales, multidimensionales y los casos de uso del sistema. También se realiza la matriz BUS, modelo de datos, esquema de seguridad así como las políticas de respaldo y recuperación.

2.2 Análisis

2.2.1 Definición del Negocio

La ONE es una entidad creada para proponer, organizar y ejecutar la aplicación de la política estatal en materia de estadística del país. Está compuesta por distintos departamentos o áreas las cuales cumplen con diversas funciones. Una de las áreas es el Departamento de Industria y Medio Ambiente, donde existen dos tipos de reportes, algunos son publicados en el sitio de la ONE y otros se les facilita a las personas que necesitan esta información. Este departamento consta además de especialistas que se encargan de emitir publicaciones que son utilizadas para la toma de decisiones. De estos especialistas existen algunos que se dedican al área de industria y otro, que es un especialista adscrito a la inversión, generalmente el director, se dedica al medio ambiente por la importancia que éste tiene para nuestras vidas.

2.2.2 Temas de análisis

Los especialistas de la ONE que se encargan de las inversiones realizan un grupo de tareas entre las que se destacan:

- Brindar la información de las encuestas sobre los gastos de inversión en actividades de protección para el medio ambiente. Entre sus objetivos fundamentales están: brindar información y conocer detalladamente el comportamiento de dichas actividades, en la medida que se produce el avance físico de la inversión, independientemente que se haya

realizado o registrado el pago por estos trabajos, constituyendo la fuente para evaluar, desde el nivel de centro informante hasta la nación.

- Examinar los gastos de protección ambientales, en cuencas hidrográficas de interés nacional, donde se anota el total de los valores ejecutados vinculados en las cuencas hidrográficas de interés nacional.
- Brindar información sobre los gastos de protección ambiental en las principales bahías, donde se anota el total de los valores ejecutados.
- Brindar información sobre el control del plan de la ejecución física de las inversiones, que tiene como objetivo principal conocer de forma detallada el comportamiento de la ejecución física de las inversiones, con independencia del pago de los trabajos. Esto constituye la única fuente para determinar la ejecución de inversiones desde el nivel del centro informante hasta la nación de forma definitiva, permiten evaluar el cumplimiento del plan de ejecución de inversiones de acuerdo con sus diferentes niveles de desagregación y ser uno de los elementos de cálculo de las inversiones en proceso.

2.2.3 Roles y permisos

Consultante: es la persona que utilizará el sistema para buscar información relacionada con los indicadores de inversiones, realizará los reportes y consultas a los datos relacionados con los temas de análisis correspondientes a los modelos de inversiones. Tiene permisos solo de Select.

Especialista de ETL: está estrechamente relacionado con las actividades relacionadas a los procesos de ETL. Tiene permisos de Select, Insert, Update, Delete, Refresh y Trigger.

2.2.4 Reglas del negocio

El código de la dimensión Clasificador de Actividad Económica (CAE), consta de la siguiente regla de negocio:

- El primer y segundo número significan el sector, del primero al cuarto significa la rama y los 6 en su conjunto la sub rama.

La dimensión indicador consta de la siguiente regla del negocio:

- En la dimensión indicador de protección del medio ambiente, el valor que toma el indicador protección y recuperación de las especies y los hábitats es la suma de la

Protección de los recursos forestales y protección y recuperación de playas y zonas costeras, incluyendo bahías.

- El valor que toma la protección de los recursos forestales es la suma de los bosques de conservación y protección y de los bosques productores.
- En el modelo 5702 en el indicador llamados otros, se refiere aquellas partidas de inversión que no correspondan a: construcción civil, equipos y al montaje de los mismos. Dentro de estos trabajos se encuentran los gastos de preparación, capacitación, adiestramiento, documentación técnica y de proyecto, promoción y comercialización, administración y gastos requeridos para las pruebas y puesta en explotación. Ampara igualmente, trabajos de prospección geológica, perforación de pozos de petróleo y gas, acumulación en la agricultura, silvicultura y ganadería, dotación de libros, obras de artes plásticas y aplicadas y otros objetos valiosos, animales para exhibición, herramientas necesarias para la habilitación inicial de las inversiones y el incremento del capital de trabajo durante el período establecido en el dictamen de aprobación del estudio de factibilidad. Su valor es la suma de la inversión físicamente tangible que registran las entidades para llevar a cabo las actividades operativas y la inversión físicamente intangible que comprende los bienes no materiales que registran las entidades para llevar a cabo las actividades operativas. El valor de la inversión físicamente tangible es la suma de la inversión de la acumulación agrícola y de la inversión de la acumulación silvícola. (Modelo Sien, 2009)

2.2.5 Necesidades de los usuarios.

Después de realizar un análisis minucioso del negocio se definen las siguientes necesidades existentes en la ONE en la temática de inversiones:

- Obtener el control de inversiones en plan de la ejecución física de inversiones, totales y en CUC por empresa, por organismos, por forma de financiamiento, por sindicatos, por DPA, por CAE, por subordinación, por empresa en perfeccionamiento, por indicadores y un trimestre dado.
- Obtener gastos de inversiones en las principales bahías por empresa, por organismos, por forma de financiamiento, por sindicatos, por DPA, por CAE, por subordinación, por empresa en perfeccionamiento, por indicadores y un trimestre dado.

- Obtener gastos de inversiones en las principales cuencas por empresa, por organismos, por forma de financiamiento, por sindicatos, por DPA, por CAE, por subordinación, por empresa en perfeccionamiento, por indicadores y un trimestre dado.
- Obtener gastos de inversiones en áreas ambientales por empresa, por organismos, por forma de financiamiento, por sindicatos, por DPA, por CAE, por subordinación, por empresa en perfeccionamiento, por indicadores y un trimestre dado.

2.2.6 Requisitos de información.

Luego de realizar un análisis de los problemas que existen en la ONE con respecto a las inversiones se definieron los siguientes requisitos de información:

- Obtener la ejecución real de la prevención de la contaminación atmosférica por modificación de procesos, en el tratamiento de los gases de escape y el aire de ventilación, en el empleo de fuentes renovables de energía y ahorro de energía, en la sustitución y eliminación de sustancias agotadoras de la capa de ozono, en equipos e instalaciones para el control, medición y análisis de la contaminación atmosférica, en otras actividades, en modificaciones preventivas en el origen, en la construcción de dispositivos antiruidos y antivibraciones, en equipos e instalaciones para la medición y el control del ruido y las vibraciones, en la prevención de la contaminación por modificación de procesos, en la prevención de la contaminación en aguas superficiales, en la limpieza de las masas de agua, en el tratamiento de las aguas residuales, en el tratamiento de las aguas de enfriamiento, en las redes de alcantarillado, el suministro y calidad de agua potable, equipos e instalaciones para el control, medición y análisis de la calidad y contaminación de las aguas, en otras actividades, en la prevención de la generación de residuos por modificación de procesos, en la recogida y transporte de los residuos, en el tratamiento y disposición final de residuos peligrosos, en el tratamiento y disposición final de residuos no peligrosos, en equipos e instalaciones para la medición, caracterización, monitoreo y control de residuos sólidos, en la prevención de la infiltración de contaminantes en suelos y aguas subterráneas, en la limpieza de suelos, en la protección de los suelos contra la erosión y otros tipos de degradación física, en la prevención de la salinización del suelo y su rehabilitación, en actividades y medidas encaminadas a la medición y control de la calidad y contaminación de los suelos.

- Obtener la ejecución real en la protección y recuperación de las especies y los hábitats, en la protección de los recursos forestales, en los bosques de conservación y protección, en los bosques productores, en la protección y recuperación de playas y zonas costeras (incluye bahías), en los equipos e instalaciones para la medición, monitoreo y análisis, en otras actividades, en protección contra las radiaciones, en la investigación y desarrollo, en otras actividades de protección del medio ambiente.
- Obtener el valor total de la ejecución real, obtener la ejecución de la gestión de las aguas, en recursos forestales y en la protección de los suelos en las cuencas de Cuyaguaje, Ariguanabo, Almendares-Vento, Zaza, Hanabanilla, Cauto, Toa, Guaso-Guantánamo, Mayarí.
- Obtener el valor total de la ejecución real y la ejecución de la gestión de las aguas, en recursos forestales y en los residuos en las bahías de La Habana, Cienfuegos, Santiago de Cuba, Nipe y Matanzas.
- Obtener el valor del plan total, el total real acumulado, el plan total en CUC y el real acumulado en CUC de la fabricación de las viviendas, en la construcción de edificios no residenciales, en otras construcciones, en el transporte, en maquinaria y equipos agrícolas, en maquinaria y equipos industriales, en la informática, en equipos importados, en proyectos, en construcción y montaje, en equipos, en equipos importados, en proyectos, en la acumulación agrícola, en la acumulación silvícola, en la acumulación ganadera .

2.2.7 Requisitos Multidimensionales.

Los requisitos multidimensionales que se detallan en el MD sobre las inversiones son los que se muestran a continuación:

- Obtener el valor de las inversiones en las actividades protección del aire y el clima, reducción del ruido y las vibraciones, gestión de las aguas, en residuos, protección y rehabilitación de los suelos, protección de la biodiversidad y los paisajes, protección contra las radiaciones, investigación y desarrollo y otras actividades de protección del medio ambiente, teniendo en cuenta las siguientes variable de entradas:

VE: empresa

VE: esfera

VE: organismo

VE: CAE

VE: indicador de protección del medio ambiente

VE: temporal

VE: DPA

VE: forma de financiamiento

VE: sindicato

VE: subordinación

VE: empresa en perfeccionamiento

- Obtener el valor total de la ejecución real, la ejecución de la gestión de las aguas y la ejecución en recursos forestales y en la protección de los suelos en las principales cuencas teniendo en cuenta las siguientes variables de entradas:

VE: empresa

VE: esfera

VE: organismo

VE: CAE

VE: indicador de las principales cuencas

VE: temporal

VE: DPA

VE: forma de financiamiento

VE: sindicato

VE: subordinación

VE: empresa en perfeccionamiento

- Obtener el valor total de la ejecución real, de la ejecución de la gestión de las aguas, en la ejecución de recursos forestales y en las principales bahías, teniendo en cuenta las siguientes variables de entradas:

VE: empresa

VE: esfera

VE: organismo

VE: CAE

VE: indicador de las principales bahías

VE: temporal

VE: DPA

VE: forma de financiamiento

VE: sindicato

VE: subordinación

VE: empresa en perfeccionamiento

- Obtener el valor de inversión del plan total, el total real acumulado, el plan total en CUC y el real acumulado en CUC en construcción y montaje, equipos, otros y fondos básicos puestos en explotación, teniendo en cuenta las siguientes variables de entradas:

VE: empresa

VE: esfera

VE: organismo

VE: CAE

VE: indicadores del control de la ejecución física

VE: temporal

VE: DPA

VE: forma de financiamiento

VE: sindicato

VE: subordinación

VE: empresa en perfeccionamiento

2.2.8 Requisitos funcionales

En el MD sobre las inversiones los requisitos funcionales son los que se detallan a continuación:

- Extraer DBF de empresa en perfeccionamiento.
- Transformar y cargar DBF de empresa en perfeccionamiento.
- Extraer DBF de indicador.
- Transformar y cargar DBF de indicador.
- Extraer DBF de CAE.
- Transformar y cargar DBF de CAE.
- Extraer DBF de organismo.
- Transformar y cargar DBF de organismo.
- Extraer DBF de DPA.
- Transformar y cargar DBF de DPA.

- Extraer DBF de empresa.
- Transformar y cargar DBF de empresa.
- Extraer DBF de esfera.
- Transformar y cargar DBF de esfera.
- Extraer DBF de sindicato.
- Transformar y cargar DBF de sindicato.
- Extraer DBF de subordinación.
- Transformar y cargar DBF de subordinación.

2.2.9 Requisitos no funcionales.

Los requisitos no funcionales son propiedades o cualidades que el producto debe tener, a continuación se muestran los que se definieron en la investigación:

RNF1: Software

- Sistema Operativo Windows 98 o superior o cualquiera de las distribuciones de Linux.

RNF2: Hardware

- Procesador Pentium III o superior.
- 1 GB como mínimo de RAM para garantizar el correcto funcionamiento del sistema al ser accedido por varios usuarios.
- 80 GB de capacidad en disco duro para el almacenamiento de información.
- Se deberá garantizar una impresora para imprimir los reportes.

RNF3: Restricciones en el diseño y la implementación

- El producto se desarrollará usando PostgreSQL 8.4.
- La herramienta manejadora de datos será PgAdmin3 1.10.
- Para el modelado de datos se utilizará Visual Paradigm.

RNF4: Apariencia o interfaz externa

- El sistema debe poseer una interfaz con un diseño sencillo, eficaz y flexible de forma tal que le facilite al usuario el uso con un sistema como éste.

RNF5: Usabilidad

- El sistema será usado por una persona que gestione la información relacionada con los indicadores estadísticos, además debe tener un diseño del modelo físico sencillo, donde se hace necesario que los usuarios tengan una noción del manejo de las computadoras y los sistemas operativos para que se pueda trabajar con rapidez y eficiencia.

RNF6: Portabilidad

- El producto será usado bajo los sistemas operativos Windows y Linux.

RNF7: Rendimiento

- Para cargar los datos y ejecutar las consultas el tiempo debe ser corto.
- La BD debe gestionar toda la información y dar respuesta a las solicitudes lo más rápido y eficientemente posible.
- El tiempo de reinicio después de un fallo tiene que ser corto y con probabilidad baja de dañar los datos.

RNF8: Confiabilidad

- Disponible en todo momento, debido a la importancia de la información que se gestiona.
- Los reportes y estadísticas que se obtendrán del producto deben ser 100% precisos y reales.
- El producto debe ser capaz de restaurarse o reponerse de las fallas de forma rápida.
- El sistema debe tener la capacidad de mantener la integridad de los datos.

2.2.10 Casos de uso del sistema.

2.2.10.1 Casos de uso de información.

- Conocer la ejecución real de las inversiones para las actividades de protección del medio ambiente.
- Conocer la ejecución real de las inversiones para las actividades de protección ambiental en cuencas hidrográficas de interés nacional.
- Conocer la ejecución real para las actividades de protección ambiental en las principales bahías.
- Conocer le ejecución física de las inversiones para las actividades físicas de las inversiones total y en CUC por diferentes tipos de inversiones.

2.2.10.2 Casos de uso funcionales.

- Extraer DBF de empresa en perfeccionamiento.
- Transformar y cargar DBF de empresa en perfeccionamiento.
- Extraer DBF de indicador.
- Transformar y cargar DBF de indicador.
- Extraer DBF de CAE.

- Transformar y cargar DBF de CAE.
- Extraer DBF de organismo.
- Transformar y cargar DBF de organismo.
- Extraer DBF de DPA.
- Transformar y cargar DBF de DPA.
- Extraer DBF de empresa.
- Transformar y cargar DBF de empresa.
- Extraer DBF de esfera.
- Transformar y cargar DBF de esfera.
- Extraer DBF de sindicato.
- Transformar y cargar DBF de sindicato.
- Extraer DBF de subordinación.
- Transformar y cargar DBF de subordinación.

2.3 Diseño

2.3.1 Matriz BUS

A continuación se muestra la matriz BUS que no es más que la representación de las relaciones existentes entre los hechos y las dimensiones del MD diseñado:

Tabla 1: Matriz BUS.

Hechos	Dimensiones															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
hech_Ámbito_ambiental	X	X	X	X	X	X	X	X	X	X			X			
Hech_Ejecución física	X	X	X	X	X	X	X	X	X	X				X		
Hech_Ejecución física específica	X	X	X	X	X	X	X	X	X	X					X	X
hech_En cuencas hidráulicas	X	X	X	X	X	X	X	X	X	X		X				
Hech_En las principales bahías	X	X	X	X	X	X	X	X	X	X	X					

2.3.2 Modelo de Datos

A continuación se expone el modelo de datos compuesto por los hechos, dimensiones y medidas seleccionados según los modelos estudiados en la investigación.

2.3.2.1 Dimensiones

Luego de un análisis de los modelos existentes en el área de inversiones se definieron las dimensiones que se mencionarán a continuación:

- DIM_CAE: esta dimensión describe los valores asociados a los clasificadores de las actividades económicas. Presenta los niveles sector, rama y subrama.
sector->rama ->subrama.

Dim_CAE		
+cae_id	integer(10)	Nullable = f...
sector_descripcional	varchar(255)	Nullable = true
subrama_descripcional	varchar(255)	Nullable = true
rama_código	integer(10)	Nullable = true
rama_descripcional	varchar(255)	Nullable = true
subrama_código	integer(10)	Nullable = true
sector_código	integer(10)	Nullable = true

- DIM_SINDICATO: esta variable de entrada describe los valores asociados a toda la información referente a todos los sindicatos de todo el país.

Dim_Sindicato		
+sindicato_id	integer(10)	Nullable = f...
sindicato_cod	integer(10)	Nullable = true
sindicato_desc	varchar(255)	Nullable = true

- DIM_EMPRESA EN PERFECCIONAMIENTO: esta variable de entrada describe los valores asociados a toda la información referente a la empresa en perfeccionamiento.

Dim_Emp_Perfeccionamiento		
+emp_perfeccionamiento_id	integer(10)	Nullable ...
emp_perfeccionamiento_cod	integer(10)	Nullable = ...
descripción	varchar(255)	Nullable = ...

- DIM_DPA: esta variable de entrada describe los valores asociados a la división política administrativa. Presenta un nivel denominado provincia y otro municipio.
provincia-> municipio.

Dim_DPA		
+dpa_id	integer(10)	Nullable = f...
prov_código	integer(10)	Nullable = true
prov_nombre	varchar(255)	Nullable = true
munic_código	integer(10)	Nullable = true
mun_nombre	varchar(255)	Nullable = true

- DIM_EMPRESA: esta variable de entrada describe los valores asociados a toda la información referente a la empresa.

Dim_Empresa		
+empresa_id	integer(10)	Nullable = f...
empresa_código	integer(10)	Nullable = true
empresa_descripcional	varchar(255)	Nullable = true

- DIM_ESFERA: esta dimensión describe los valores asociados a toda la información referente a la esfera.

Dim_Esfera		
+esfera_id	integer(10)	Nullable = f...
esfera_desc	varchar(255)	Nullable = true
esfera_código	integer(10)	Nullable = true

- DIM_FFINANCIAMIENTO : esta variable de entrada describe los valores asociados a toda la información referente a la forma de financiamiento. Presenta los niveles forma de financiamiento1 y forma de financiamiento2.

forma de financiamiento1-> forma de financiamiento2.

Dim_FFinanciamiento		
+ff_id	integer(10)	Nullable = f...
ff_código_gral	integer(10)	Nullable = true
ff_descripción_gral	varchar(255)	Nullable = true
ff_código	integer(10)	Nullable = true
ff_descripción	varchar(255)	Nullable = true

- DIM_ORGANISMO : esta variable de entrada describe los valores asociados a toda la información referente al organismo.

Dim_Organismo		
+organismo_id	integer(10)	Nullable = false
organismo_cod	integer(10)	Nullable = true
organismo_código	integer(10)	Nullable = true
organismo_descr_corta	varchar(255)	Nullable = true

- DIM_SUBORDINACION: esta variable de entrada describe los valores asociados a toda la información referente a la subordinación.

Dim_Subordinación		
+subordinación_id	integer(10)	Nullable = false
subordinación_código	integer(10)	Nullable = true
subordinación_descripción	varchar(255)	Nullable = true

- DIM_TEMPORAL: esta variable de entrada describe los valores asociados a toda la información referente al tiempo. Presenta los niveles trimestre y mes.
trimestre-> mes.

Dim_Temporal		
+temporal_id	integer(10)	Nullable = f...
anno	integer(10)	Nullable = true
trimestre_número	integer(10)	Nullable = true
trimestre_descripción	varchar(255)	Nullable = true

- DIM_INDICADOR DE LAS PRINCIPALES BAHIAS: esta variable de entrada describe los valores asociados a toda la información referente a las principales bahías. Presenta los niveles temática e indicador.
temática-> indicador.

Dim_Indicador de las principales bahías		
+indicador_id	integer(10)	Nullable = f...
indicador_descripción	varchar(255)	Nullable = true
indicador_tipo	integer(10)	Nullable = true
temática_descripción	varchar(255)	Nullable = true
indicador_CUI	integer(10)	Nullable = true
indicador_UM	integer(10)	Nullable = true
indicador_agragable	binary(1)	Nullable = true

- DIM_INDICADOR DE LAS PRINCIPALES CUENCAS: esta variable de entrada describe los valores asociados a toda la información referente a las principales cuencas. Presenta los niveles temática e indicador.
temática-> indicador.

Dim_Principales Cuencas		
+indicador_id	integer(10)	Nullable = f...
indicador_descripción	varchar(255)	Nullable = true
indicador_tipo	integer(10)	Nullable = true
temática_descripción	varchar(255)	Nullable = true
indicador_CUI	integer(10)	Nullable = true
indicador_UM	integer(10)	Nullable = true
indicador_agragable	binary(1)	Nullable = true

- DIM_INDICADOR DE PROTECCION DE MEDIO AMBIENTE: esta variable de entrada describe los valores asociados a toda la información referente a las actividades relacionadas con la protección del medio ambiente. Presenta los niveles temática e indicador.

temática-> indicador.

Dim_Indicador de protección al medio ambiente		
+indicador_id	integer(10)	Nullable = f...
indicador_descripción	varchar(255)	Nullable = true
indicador_tipo	integer(10)	Nullable = true
temática_descripción	varchar(255)	Nullable = true
indicador_CUI	integer(10)	Nullable = true
indicador_UM	integer(10)	Nullable = true
indicador_agragable	binary(1)	Nullable = true

- DIM_INDICADOR DEL CONTROL DE LA EJECUCION FISICA: esta variable de entrada describe los valores asociados a toda la información referente al control de la ejecución física, específicamente los valores de cada rama. Presenta los niveles temática e indicador.

temática-> indicador.

Dim_Indicador del control de la ejecución física		
+indicador_id	integer(10)	Nullable = f...
indicador_descripción	varchar(255)	Nullable = true
indicador_tipo	integer(10)	Nullable = true
temática_descripción	varchar(255)	Nullable = true
indicador_CUI	integer(10)	Nullable = true
indicador_UM	integer(10)	Nullable = true
indicador_agragable	binary(1000)	Nullable = true

- DIM_INDICADOR DEL CONTROL DE LA EJECUCION FISICA ESPECIFICA: esta variable de entrada describe los valores asociados a toda la información referente al control de la ejecución física pero basada en los valores de cada subrama. Presenta los valores temática e indicador.

temática-> indicador.

Dim_Indicador del control de la ejecución física esp		
+indicador del control de la ejecución física esp	integer(10)	Nullable = false
indicador_descripción	varchar(255)	Nullable = true
indicador_tipo	integer(10)	Nullable = true
indicador_CUI	integer(10)	Nullable = true
indicador_UM	integer(10)	Nullable = true
indicador_agragable	binary(1)	Nullable = true

- DIM_TEMPORAL_ANNUAL: esta variable describe los valores asociados a toda la

información referente al tiempo, específicamente al último trimestre del año. Presenta los niveles trimestre y mes.

trimestre-> mes.

Dim_Temporal_Annual		
+temporal_annual_id	integer(10)	Nullable = ...
anno	integer(10)	Nullable = t...

2.3.2.2 Tablas de hechos.

En la ONE se realizan los análisis más importantes de información y los reportes más universales de la disciplina inversiones, estos análisis y reportes parten de los siguientes subsistemas: Organismo, Esfera, CAE, Empresa, Indicadores, Temporal, DPA, Forma de financiamiento, Sindicato, Empresa en perfeccionamiento, Subordinación.

Tabla de Hecho Ámbito Ambiental.

En esta tabla se encuentra toda la información referente al ámbito ambiental.

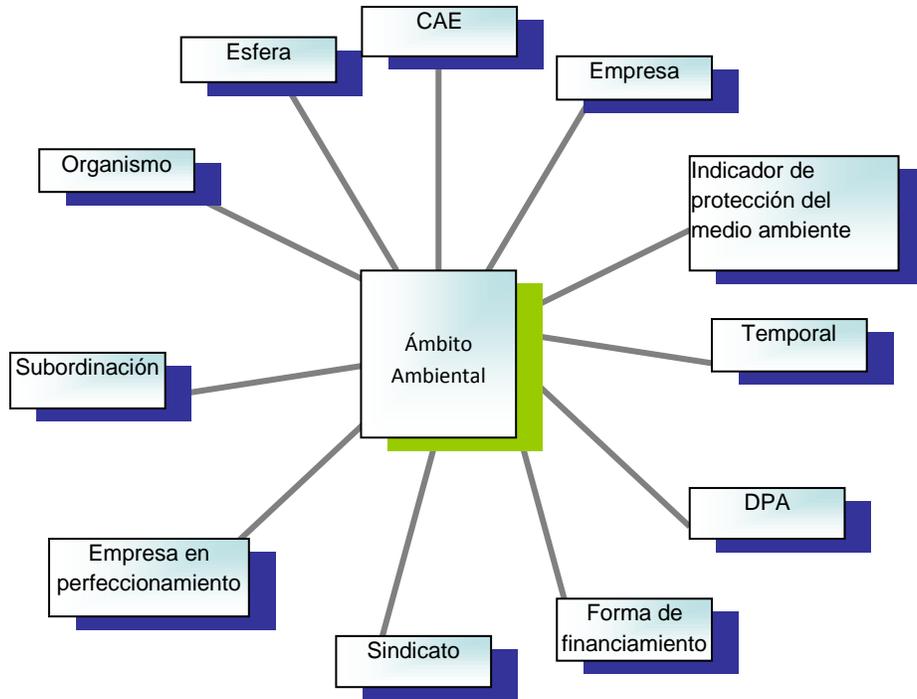


Figura 3 Tabla del Hecho Ámbito Ambiental.

En esta tabla se encuentra toda la información referente a las cuencas hidráulicas.

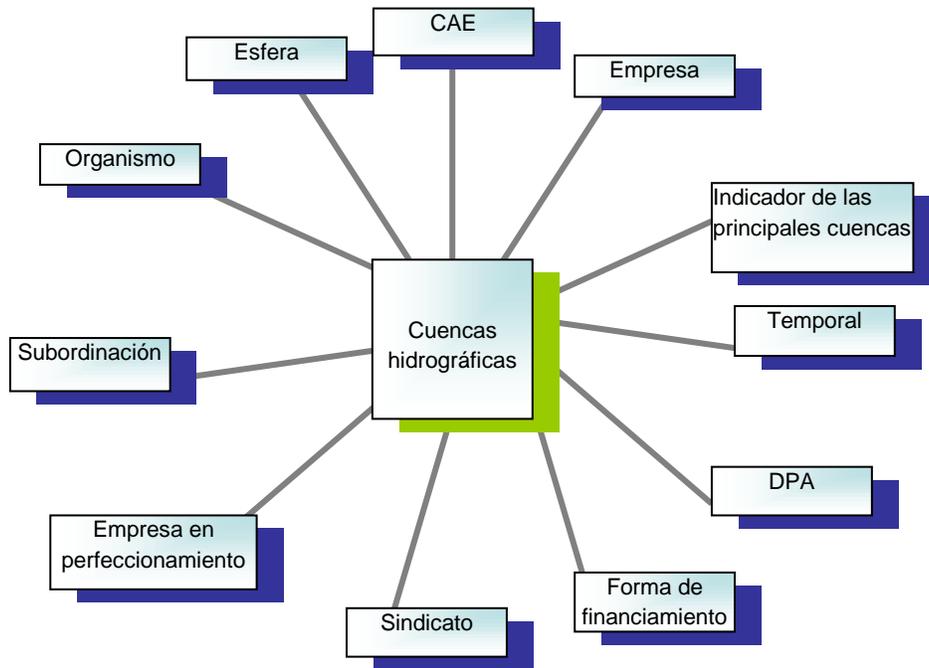


Figura 4 Tabla del hecho Cuencas Hidrográficas.

Tablas de Hecho principales bahías.

En esta tabla se encuentra toda la información referente a todas las principales bahías.

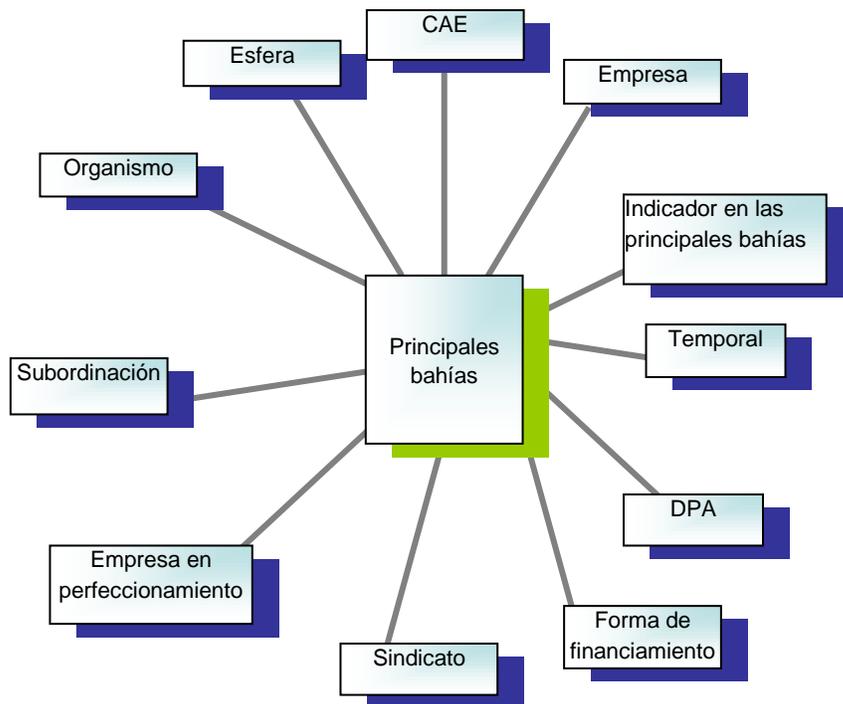


Figura 5 Tabla del hecho Principales bahías.

Tablas de Hecho ejecución física.

En esta tabla se encuentra toda la información referente a toda la ejecución física.

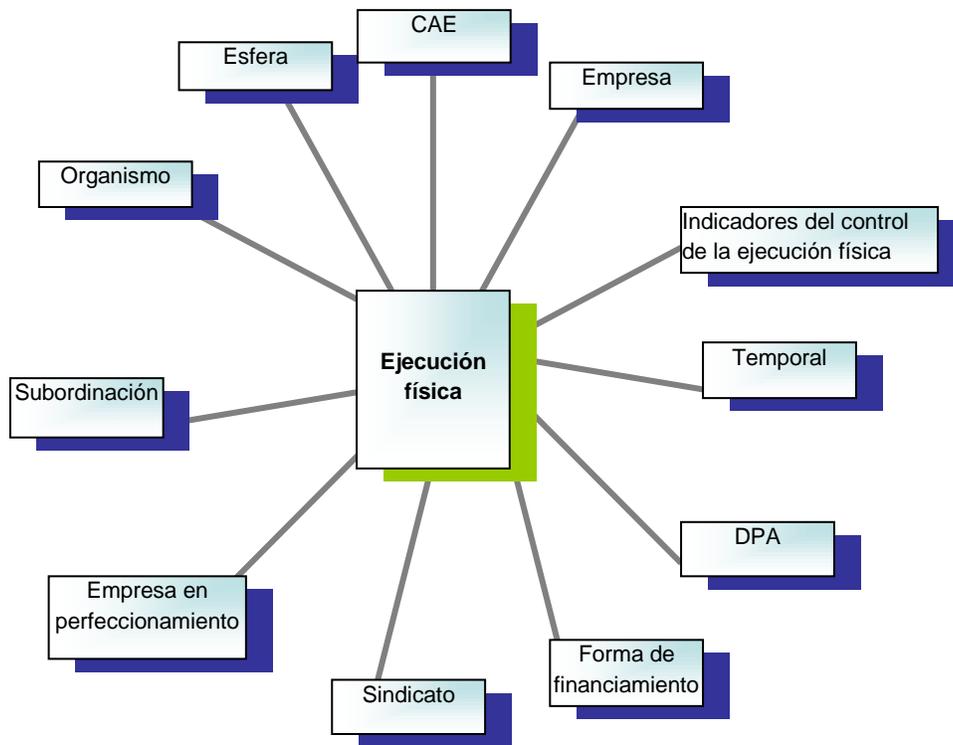


Figura 6 Tabla del hecho Ejecución física.

Tablas de Hecho Ejecucion_fisica_especifica.

En esta tabla se encuentra toda la información referente a toda la ejecución física específica.

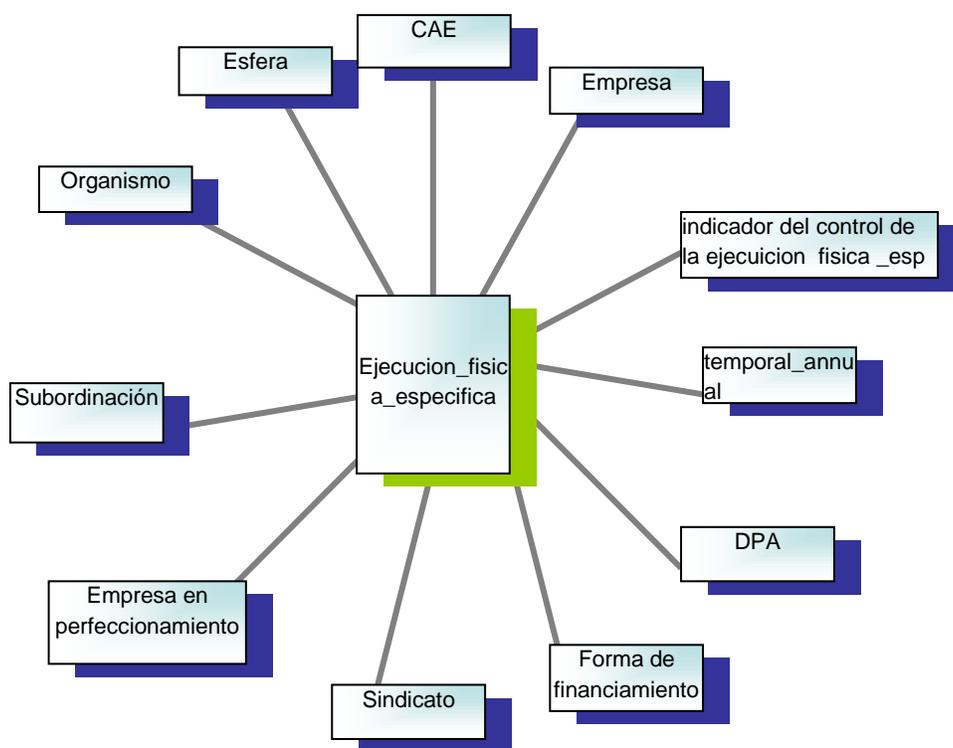


Figura 7 Tabla del hecho Ejecución física específica.

2.3.2.3 Medidas

Las medidas que están relacionadas con el hecho_ principales bahías:

- Ejecución en recursos forestales.
- Ejecución en residuos.
- Ejecución en gestión de las aguas.
- Total de ejecución real.

Las medidas que están relacionadas con el hech_En cuencas hidráulicas:

- Ejecución en recursos forestales.
- Ejecución en protección de los suelos.

- Ejecución en gestión de las aguas.
- Total de ejecución real.

Las medidas que están relacionadas con el hech_Ejecución física:

- Plan_Total.
- Real_Acumulado.
- Plan_Total.
- Real_Acomolado_CUC

Las medidas que están relacionadas con el hech_Ámbito_ambiental:

- existencia_final_período

2.3.4 Esquema de Seguridad

El esquema de seguridad estará respaldado por los niveles de acceso al sistema, específicamente por los roles definidos.

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

- Funcionamiento: Dispositivos de seguridad (firewall, inspectores de contenido, sensores).
- Servidores: Servidor de gestión de administración y de BD.
- Presentación: Consolas de administración.

La solución debe tener una Interfaz de Administración Gráfica (GUI) en la consola de administración, la cual proporciona una visualización de la topología de red en el editor de políticas. La visualización de la topología es una representación de los objetos de red definidos en el editor de políticas y la relación entre estos objetos y la red. Para la auditoria del tráfico entrante y saliente, así como para la generación de informes de eventos e intentos de ataque, se debe utilizar un registro del tipo LOG FILE como un evento que deba ser registrado. El formato de las entradas del registro pedidas por una regla, es determinado por el tipo del registro especificado en la regla, todo esto sin alterar el rendimiento de acceso a los equipos protegidos por la solución.

La configuración y pruebas de todos los equipos deberán efectuarse en las instalaciones de la ONE en la Ciudad de la Habana, conjuntamente con personal técnico de informática, de acuerdo con las especificaciones de configuración.

El mantenimiento preventivo deberá efectuarse mediante una visita semestral durante el período de vigencia de la garantía técnica.

El mantenimiento correctivo deberá efectuarse anualmente, a requerimiento de la ONE con tiempos de respuesta por soporte técnico de 4 horas y de solución de 48 horas los 365 días del año, de fácil ubicación vía e-mail o telefónica.

Se debe considerar actualizaciones de software: parches y sistemas operativos, los mismos que deben ser recomendados por los fabricantes de los productos entregados.

2.3.5 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:

1) Periodicidad de las salvas: Las salvas se efectúan mensualmente de la información total que posea la BD, así lo tiene definido actualmente la organización, certificando en todo momento que exista una copia estricta de la información que está presente en el servidor.

2) Tablas involucradas: Las tablas que se involucran en la realización son Principales Bahías, En Cuencas Hidráulicas, Ejecución Física, Ejecución Física Específica y Ámbito Ambiental.

3) Salvas (Backups) existentes: Actualmente no existen backups en esta área.

3.1) Periodicidad de reemplazo de los backups: se realizan los reemplazos de backups cada un año.

3.2) Periódicas de las pruebas a los backups: El estado de los backups se chequea mensualmente, mediante pruebas de rendimiento y flexibilidad.

2.4 Conclusiones

En este capítulo se realizó un estudio detallado del negocio vigente en sobre las inversiones y se comenzó a desarrollar la propuesta de solución del sistema. Con este propósito se identificaron los requisitos funcionales, no funcionales, multidimensionales y de información, además de los temas de análisis, reglas del negocio que se deberán tener en cuenta, los roles y permisos, las necesidades de usuarios, los casos de uso del sistema así como la matriz BUS, el modelo de datos, las políticas de respaldo y recuperación y los esquemas de seguridad.

Capítulo 3: Implementación y Pruebas

3.1 Introducción del capítulo

En este capítulo se describe el modelo de datos físico, donde se muestran los esquemas, las tablas, restricciones, secuencias e índices de la estructura de datos. Se describen los usuarios, roles y privilegios de la política de acceso a los objetos de la BD de inversiones. También se crea una guía de implantación que posee los pasos para la instalación de la BD y los requerimientos de software. Además, se realizan las siguientes pruebas: lista de chequeo de análisis, lista de chequeo de diseño, validación de los requerimientos por los clientes y casos de prueba de implantación.

3.2 Modelos de Datos Físico

Un modelo de datos es un tipo de lenguaje que se orienta principalmente a describir una BD. Estos se pueden clasificar dependiendo de los tipos de conceptos que ofrecen para describir la estructura de la BD.

Además, contienen un grupo de operaciones esenciales para la elaboración de las consultas y reajustes de los datos. En el modelo físico se detalla el almacenamiento de los datos.

3.2.1 Estructuras de Datos

Las estructuras de datos son una colección de datos cuya organización se caracteriza por las funciones de acceso que se usan para almacenar y acceder a elementos individuales de datos; se pueden definir también como una forma de organizar un conjunto de datos elementales con el objetivo de facilitar su manipulación.

Los datos están estructurados por esquemas para una mejor organización y también por tablas ya que éstas presentan las restricciones y están organizadas por los índices definidos y las políticas de indexado.

3.2.2 Esquemas y tablas

En una BD, los esquemas definen sus tablas, sus campos en cada tabla y las relaciones entre cada campo y cada tabla. Estos esquemas son almacenados en un Diccionario de Datos.

Este trabajo cuenta con 3 esquemas que constituyen formas de organización de la BD que pueden contener tablas, tipos de datos, funciones y operadores. Además, en el desarrollo de la solución se encuentran 21 tablas.

A continuación se mostrará una representación de la forma en que están agrupadas las tablas por esquemas.

Tabla 2 Esquemas y Tablas.

Esquema	Tablas
Dimensiones	dim_cae
Dimensiones	dim_dpa
Dimensiones	dim_empresa
Dimensiones	dim_esfera
Dimensiones	dim_organismo
Dimensiones	dim_emp_perfeccionamiento
Dimensiones	dim_indicador del control de la ejecución física _esp
Dimensiones	dim_ffinanciamiento
Dimensiones	dim_indicador de las principales bahías
Dimensiones	dim_indicador de las principales cuencas
Dimensiones	dim_indicador de protección del medio ambiente
Dimensiones	dim_indicadores del control de la ejecución física
Dimensiones	dim_sindicato
Dimensiones	dim_subordinación
Dimensiones	dim_temporal
Dimensiones	dim_temporal_annual
Hechos_Ejecución_física	Hechos_Ejecución_física
Hechos_Ejecución_física	Hech_Ejecución física específica
Hechos_Medio_Ambiente	Hech_En las principales bahías
Hechos_Medio_Ambiente	hech_En cuencas hidráulicas
Hechos_Medio_Ambiente	hech_Por ámbito ambiental

3.2.3 Restricciones y Secuencias

Las restricciones son las condiciones que exigen la obediencia de ciertas normas en una BD. Éstas pueden ser establecidas por el administrador de la BD. Un ejemplo de restricción podría ser: validar que un campo tenga cierta cantidad de caracteres y sea de un tipo específico.

Las secuencias son atributos que se van incrementando secuencialmente durante la entrada de datos a la BD, un ejemplo de estos tipos de atributos son las llaves primarias.

En este trabajo existen 65 llaves foráneas y 16 llaves primarias, a continuación se mostrarán algunas de ellas:

Tabla 3: Llaves primarias relacionadas con la BD.

Llaves primarias	Indicador_control _ejecución_fisica esp_id
	cae_id
	dpa_id
	emp_perfeccionamiento_id
	empresa_d
	esfera_id
	principales cuencas_id
	ff_id
	ejecución fisica_id
	principales bahias_id
	organismo_id
	sindicato_id
	subordinación_id
	temporal_id
	temporal_anual_id
medio ambiente_id	

Tabla 4: Llaves foráneas de la tabla de hecho ejecución física

Llaves Foráneas	dim_indicador_control_ejecucion__fisica__esp103
	dim_cae105
	dim_dpa107
	dim_dpa114
	dim_emp_perfeccionamiento104
	dim_empresa108
	dim_esfera109
	dim_ffinanciamiento112
	dim_organismo106

	dim_organismo115
	dim_sindicato113
	dim_subordinacion111
	dim_temporal_annual116

3.2.4 Índices

Los índices optimizan la velocidad de las operaciones que se realizan en una BD, ya que se puede acceder de forma rápida a los registros de una tabla. Contiene claves generadas a partir de una o varias columnas de la tabla o la vista. Estas claves están almacenadas en una estructura que permite que se busque de forma rápida y eficiente la fila o filas asociadas a los valores de cada clave.

Para esta solución se utilizó como estrategia de indexado por claves primarias, quedando un total de 21 índices, a continuación se muestran algunas de ellos:

Tabla 5 Índices de las tablas de la BD

Índice	Esquema	Tabla	Tipo de índices
PK_dim_ejecuc_fis	Dimensiones	Dim_indicador_del_control_de_la_ejecucion_fisica	btree
PK_hech_ejec_fis_esp	Hechos_Ejecucion_fisica	Hech_Ejecucion_fisica_especifica	btree
PK_hech_ejec_fis	Hechos_Ejecucion_fisica	Hech_Ejecucion_fisica	btree
PK_hech_bahias	Hechos_Medio_Ambiente	Hech_En las principales_bahias	btree
PK_dim_c	Dimensiones	dim_cae	btree

ae			
PK_dim_d pa	Dimensiones	dim_dpa	btree
PK_dim_e mp_perf	Dimensiones	dim_emp_perfeccionamiento	btree
PK_dim_e mp	Dimensiones	dim_empresa	btree
PK_dim_e sfera	Dimensiones	dim_esfera	btree

3.3 Usuarios y Permisos

Usuarios y Roles

Los usuarios y roles se crean con el propósito de garantizar la seguridad de la BD. De esta forma, se asegura que un usuario puede tener acceso a los datos sólo si encuentra autenticado en la BD.

A continuación se muestran los usuarios, roles y permisos de la BD:

Tabla 6 Usuarios y roles de la BD.

Usuarios	Roles	Permisos	Descripción
Especialista de ETL	Extrae, transforma y carga los datos.	<i>Select, Insert, Update, Delete, Refresh y Trigger</i>	Extrae, transforma y carga los datos.
Consultante	Usuario	<i>Select.</i>	Analiza y visualiza los datos.
DBA	Administrador	<i>Owner, Select, Update, Insert, Delete, Refresh y Trigger.</i>	Administra la BD.

3.4 Carga de nomencladores

1. El clasificador de la actividad económica contiene el nombre de sector, nombre de la rama y el nombre de la actividad económica al que pertenecen, también cuenta con un código que los identifica.
2. El nomenclador de la DPA contiene el nombre del municipio, el nombre de la provincia al que pertenece y un código que los identifica.
3. El nomenclador de la empresa en perfeccionamiento contiene un código por los cuales se agrupan las empresas y una descripción de lo que significa cada código.
4. El nomenclador empresa contiene en la descripción de la empresa, el nombre de la empresa y el código que identifica cada empresa.
5. El nomenclador esfera contiene en la descripción de la esfera el nombre de la esfera a la que pertenece cada centro informante.
6. El nomenclador organismo contiene en la descripción del organismo el nombre del mismo y en el alias las siglas que los identifica.
7. El nomenclador sindicato contiene en la descripción del sindicato el nombre de todos los sindicatos del país.
8. El nomenclador subordinación contiene el nombre de una descripción a la cual se encuentra subordinado cada centro informante de país.

3.5 Guía de Implantación

Las guías de implantación contienen los pasos necesarios para implantar cualquier sistema informático, que no son más que las necesidades que presenta el software para su posterior instalación. Para poder darle mantenimiento a la solución del presente trabajo es necesario crear una guía de implantación donde se analicen los detalles de tipo técnico y organizativo.

3.5.1 Secuencia de Pasos

1. Debe estar instalado el gestor de BD PostgreSQL 8.4.
2. Debe estar instalada una herramienta de administración de BD, la cual ya fue definida, el PgAdmin.
3. Se debe crear una BD nueva utilizando la herramienta de administración de BD deseada.
4. Luego de tener la BD nueva lista, se carga el script `bd_dim_ddl` para crear la estructura

física de las dimensiones del mercado de datos.

5. Luego de tener la BD nueva lista, se carga el script `bd_Hech_fisico_ddl` para crear la estructura física de los hechos del mercado de datos.
6. Luego de tener la BD nueva lista, se carga el script `bd_Hech_ambiente_ddl` para crear la estructura física de los hechos del mercado de datos.
7. Teniendo la estructura de la BD, están creadas las condiciones para correr el script `bd_dim_dml`, quedando así cargados los nomencladores en el almacén.
8. Al estar la BD creada estructuralmente con los nomencladores cargados se debe correr el script `bd_dim_dcl`.
9. Al estar la BD creada estructuralmente con los nomencladores cargados se debe correr el script `bd_Hech_fisico_dcl`.
10. Al estar la BD creada estructuralmente con los nomencladores cargados se debe correr el script `bd_Hech_ambiente_dcl`.

Luego de aplicados estos pasos la BD se encuentra en disposición de ser utilizada para el posterior proceso de ETL.

3.6 Validación y pruebas

La validación y las pruebas que se llevan a cabo a un producto, se realizan para garantizar el cumplimiento de la exigencia del cliente y la calidad de la solución. Existen varios métodos para validar el buen funcionamiento de los MD, uno de estos es el de implantación.

3.6.1 Listas de Chequeo de Análisis

Las listas de chequeo se realizan con el objetivo de garantizar que los artefactos que se elaboren en el análisis puedan cumplir con un formato determinado y que toda la información recogida esté organizada según lo establecido.

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

3.6.2 Validación de requisitos por el cliente

Se realizó un encuentro con los clientes del presente trabajo, Elena Leonila Fernández García y Asnioby Hernández López, los cuales estuvieron conformes con la propuesta de la solución presentada. Durante todo el proceso efectuado en la investigación los clientes estuvieron

supervisando cada avance, estando satisfechos con el análisis, diseño e implementación de la solución que se generó al finalizar dichas fases.

3.6.3 Lista de Chequeo de Diseño

El objetivo de la lista de chequeo es garantizar que los artefactos elaborados en el diseño y el modelo de datos cumplan con el formato que se tiene establecido. Además, es específica para cada sistema y pretende cubrir todos los aspectos aplicables en los temas de diseño e implementación. A continuación se menciona este tipo de lista.

- Lista de Chequeo Modelo de Datos.

3.6.4 Pruebas de Implantación

<u>Casos de Prueba</u>	<u>Precondición</u>	<u>Pos condición</u>	<u>Resultado esperado</u>
Crear BD.	PostgreSQL 8.4 instalado como gestor de BD.	BD creada.	BD creada sobre el gestor de BD PostgreSQL 8.4.
Roles y permisos	BD creada e instalada la herramienta de administración.	BD creada con sus roles y permisos.	Roles creados con los respectivos permisos.
Cargar los Nomencladores	BD creada con sus roles y permisos.	BD lista para ser utilizada.	Campos de la BD llenos.

Tabla 7 Resultados de las listas de chequeo aplicadas.

Cantidad de indicadores evaluados	Total de indicadores	Cantidad de indicadores evaluados de correctos	Cantidad de indicadores evaluados de incorrectos	Cantidad de indicadores que No Proceden	Evaluación del artefacto
Lista de chequeo Evaluación de las áreas de la organización.	10	10	0	-	Bien

Lista de chequeo de la Herramienta para la recolección y análisis de la información.	29	27	-	2	Bien
Lista de chequeo Especificación de Requisitos.	30	30	0	-	Bien
Lista de Chequeo Modelo de Datos.	14	14	0	-	Bien

3.7 Conclusiones

En este capítulo se describió el modelo físico del MD, los pasos para la implantación del sistema y el proceso de validación y pruebas realizado. Basado a lo antes mencionado se puede concluir que:

- La estructura de esquemas y tablas especificada brinda una correcta organización de los datos de la BD.
- Los usuarios y sus privilegios cumplen con lo establecido en los requisitos iniciales del sistema.
- Los pasos especificados para la implantación de la BD logran la correcta creación de los objetos, carga de nomencladores y creación de usuarios y sus privilegios.
- Los nomencladores cargados dan respuesta a lo que especifica el cliente.
- Las pruebas realizadas verifican el correcto funcionamiento de los pasos de implantación del sistema.

Conclusiones generales

Se puede plantear que la investigación cumplió todos los objetivos planteados cuando se comenzó la misma. Además se puede decir que:

- El estudio de las principales tendencias de implementación de los MD permitió escoger las herramientas factibles para el diseño e implementación del sistema.
- El análisis de los Modelos del DW de las inversiones proporcionó que se obtuvieran los indicadores que luego se utilizarían para el diseño del MD.
- La implementación y carga de los clasificadores para el MD de las inversiones garantizó que el área de inversiones cuente con una integración más completa de la información y una mejor disponibilidad de la misma.
- La validación de la propuesta mediante las pruebas hizo posible que la propuesta contara con la calidad requerida.

Recomendaciones

Los objetivos generales de este trabajo fueron alcanzados, pero durante su desarrollo han surgido ideas que serían recomendables para su futuro perfeccionamiento:

- Perfeccionar la construcción del MD para las inversiones e incorporarlo al DW de la ONE.
- Que se logre una integración de todos los MDs para que se pueda trabajar con más información y de forma más organizada.
- Que se tome el contenido de esta investigación como referencia para próximos sistemas de mercados y almacenes de datos que se desarrollen.

Referencias Bibliográficas

(2008). Obtenido de

<http://www.gestiopolis.com/recursos/experto/catsexp/pagans/ger/14/ratioseguimiento.htm>.

(2010). Obtenido de nocturnabsas.com.ar

Alvarez, S. (31 de julio de 2007). *DesarrolloWeb.com*. Obtenido de <http://DesarrolloWeb.com>.

Ayala., M. C. (5 de Febrero de 2010). Obtenido de http://www.guia-ubuntu.org/index.php?title=PgAdmin_III

Bernabeu, R. D. (2009). Obtenido de <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/i-data-warehousing-investigacion-y-sistematizacion-concepto-13>.

BRAKETT. (s.f.). Obtenido de

http://www.igac.gov.co:8080/igac_web/UserFiles/File/ciaf/TutorialSIG_2005_26_02/paginas/bd_conceptosgenerales.htm#Data%20Warehouse

Caprino, J. L. (s.f.). Obtenido de

<http://www.tecnomarkets.com/servicios/tecnonews/research62.htm>

Casanova, J. (1 de Julio de 2009). Obtenido de <http://archives.postgresql.org/pgsql-es-fomento/2009-07/msg00000.php>

centrocamaleon.com. (s.f.). Obtenido de <http://www.centrocamaleon.com/v1/?p=233>

Chavez, M. P. (19 de Agosto de 2008). Obtenido de <http://www.embarcadero.com/press-releases/embarcadero-lanza-er-studio-enterprise-portal>

Curto, J. (28 de Noviembre de 2006). Obtenido de

<http://informationmanagement.wordpress.com/2006/11/28/dw-definiciones-de-inmon-y-kimball/>

definicion.de. (2008). Obtenido de <http://definicion.de/metodologia/>.

DELGADO. (1999). Obtenido de

http://www.igac.gov.co:8080/igac_web/UserFiles/File/ciaf/TutorialSIG_2005_26_02/paginas/bd_conceptosgenerales.htm#Data%20Warehouse

Díaz, S. O. (2009). Obtenido de www.wiphala.net/courses/dwh/ICSI243/2009-II/class/class_11_bi.ppt

Forno, J. P. (2009). *Taller Marketing Intelligence*. Obtenido de

http://www.formulisa.cl/cms/index.php?option=com_docman&task=doc_download&gid=7&Itemid=159

REFERENCIAS BIBLIOGRÁFICAS

- Fránquiz, I. M. (Marzo de 2009). Obtenido de <http://www.ahciet.net/actualidad/revista/r.aspx?ids=10796&ids2=21861>
- gestiopolis.com*. (2008). Obtenido de <http://www.gestiopolis.com/recursos/experto/catsexp/pagans/ger/14/ratioseguimiento.htm>.
- Giménez, L. Z. (2008). *Documat*. Obtenido de <http://documat.unirioja.es/servlet/tesis?codigo=18136>
- González, J. F. (19 de noviembre de 2006). *Sistemas Decisionales, algo más que Bu'siness Intelligence*. Obtenido de <http://sistemasdecisionales.blogspot.com/2006/11/kimball-1-inmon-1-usabilidad-vs.html>
- Herramientas CASE*. (s.f.). Obtenido de <http://www.pol.una.py/archivos/IngeInfo/ingeSoftI/MaterialPrimeraC.pdf>.
- Herrera, C. (2007).
- Herrera, C. (30 de 10 de 2007). *Adictos al trabajo*. Obtenido de <http://www.adictosaltrabajo.com/tutoriales/tutoriales.php?pagina=datawarehouse4>
- INMON. (1993). Obtenido de http://www.igac.gov.co:8080/igac_web/UserFiles/File/ciaf/TutorialSIG_2005_26_02/paginas/bd_conceptosgenerales.htm#Data%20Warehouse
- INMON. (1993). Obtenido de http://www.igac.gov.co:8080/igac_web/UserFiles/File/ciaf/TutorialSIG_2005_26_02/paginas/bd_conceptosgenerales.htm#Data%20Warehouse
- Inmon, B. (2003). *Wiley-Mastering Data Warehouse Design-Relational and Dimensional Techniques*.
- Janny. (02 de junio de 2006). *Sistemas Manejadores de Bases de Datos*. Obtenido de <http://janeidi.blogspot.com/>
- Lara, A. (2009). *Bases de Datos: Conceptos y Software*. Obtenido de <http://www.wwf-mar.org/events/2008-2009/seminars-workshops/files/DATA%20WORKSHOP.PRESENTATIONS/ARNOLD/Bases%20de%20Datos%20-%20Conceptos%20y%20software.pptx>
- López, D. A. (2005). *Beneficios del Data Warehouse*. Obtenido de <http://impreso.elnuevodiario.com.ni/2005/07/11/informatica/39878>
- López, E. (29 de Julio de 2009). Obtenido de <http://www.gratisprogramas.org/descargar/embarcadero-erstudio-v802-build-5991-rapidshare/>

REFERENCIAS BIBLIOGRÁFICAS

- Martínez, D. P. (2010). *danielpecos.com*. Obtenido de http://danielpecos.com/docs/mysql_postgres/x15.html
- Martínez, R. (2010). *postgresql-es.org*. Obtenido de http://www.postgresql-es.org/sobre_postgresql
- Martínez, S. C. (21 de Enero de 2010). *centrocamaleon.com*. Obtenido de <http://www.centrocamaleon.com/v1/?p=233>
- MENDEZ, F. (2009). *Sistema Gestor de Bases de Datos (SGBD)*. Obtenido de <http://ferchomendez06.blogspot.com/2009/04/sistema-gestor-de-bases-de-datos-sgbd.html>
- Mundo-Contact*. (agosto de 2008). Obtenido de www.embarcadero.com/products/er_studio_portal.
- Nazario, M. A. (2002). *AllFusion ErwinData Modeler*.
- netpecos.org*. (s.f.). Obtenido de http://www.netpecos.org/docs/mysql_postgres/index.html
- Pavón, E. L. (2007). Obtenido de <http://sliion2000.blogspot.com/2007/04/visual-paradigm-una-herramienta-de-lo.html>
- Perissé, M. C. (2001). *Una Metodología Simplificada*. Argentina.
- postgresql.org*. (2009). Obtenido de http://www.postgresql.org/about/press/presskit84.html.es#original_release.
- Ramírez, A. (22 de marzo de 2004). *Subversion*. Obtenido de <http://www.scribd.com/doc/2606327/subversion-spanish>
- Rivas, I. G. (2006). *Una metodología para sectorizar pacientes en el consumo de medicamentos aplicando Datamart y Datamining en un Hospital Nacional*.
- Rodríguez, H. A. (2006). *Sistema Manejador de Base de Datos*. Obtenido de <http://www.mailxmail.com/curso-procesamiento-datos-oracle/sistema-manejador-base-datos>
- Rodriguez, J. (16 de Noviembre de 2008). *Legox*. Obtenido de [url:http://www.legox.com/software/sql-server-2008/](http://www.legox.com/software/sql-server-2008/)
- Romero, A. L. (2006). *Bodegas de Datos: Consideraciones de Implementación, Seguridad y Control*. Obtenido de <http://www.acis.org.co/fileadmin/Conferencias/ConfAlfonsoLuqueNov30-06.pdf>
- Shinin, M. V. (2006). *Análisis diseño e implementación de un datamart para el área de sismología del departamento de geofísica de la escuela politécnica nacional*. Obtenido de <http://bibdigital.epn.edu.ec/bitstream/15000/177/1/CD-0566.pdf>

REFERENCIAS BIBLIOGRÁFICAS

Sitio de Descargas. (2007). Obtenido de [http://www.freedownloadmanager.org/es/downloads/Paradigma_Visual_para_UML_\(M%C3%8D\)_14720_p/](http://www.freedownloadmanager.org/es/downloads/Paradigma_Visual_para_UML_(M%C3%8D)_14720_p/)

Soto, P. L. (2009). *Mitecnologico*. Obtenido de <http://www.mitecnologico.com/Main/MercadosDeDatosDataMart>

todoexpertos.com. (s.f.). Obtenido de <http://www.todoexpertos.com/categorias/tecnologia-e-internet/bases-de-datos/oracle/respuestas/14706/vetajas-y-desventajas>

Torres, L. (12 de Octubre de 2007). Obtenido de <http://www.gravitar.biz/index.php/bi/bi-terminologia-1/>

Vázquez, L. C. (27 de 08 de 2009). *Tecnología e Internet*. Obtenido de <http://gestiopolis.com>

Velasco, R. H. (2009). Obtenido de <http://www.rhernando.net>

Velasco, R. H. (2005). *Características principales de Oracle*. Obtenido de <http://www.rhernando.net/modules/tutorials/doc/bd/oracle.html>

Zapata, L. G. (24 de Septiembre de 2005). Obtenido de http://hugolopez.phi.com.co/docs/download/file=Giraldo-Zapata-Herramientas%20de%20ISW.pdf,_id=17

(2009). *Modelo Sien*.

Aleman, J. M. (29 de mayo de 2009). *Javadabadu!!! Desarrollo de aplicaciones y sistemas de software con tecnología Java*. Obtenido de <http://javadabaduuu.blogspot.com/2009/05/parentesis-datawarehouse-con-mysql-y.html>

Anexos

Anexo 1 Listas de Chequeo del Análisis.

Lista de Chequeo Evaluación de las áreas de la organización.

Estructura del documento					
Peso	Indicadores a Evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
Crítico	1. ¿Está el documento acorde con la plantilla estándar del proyecto o del expediente de proyecto?	0			
Crítico	2. ¿Contiene las secciones obligatorias definidas en el expediente? (Ver Expediente de Proyecto)	0			
Elementos definidos por la metodología					
Peso	Indicadores a Evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
Crítico	1. ¿Se definió el	0			

	alcance de las áreas de la organización?				
Crítico	2. ¿Se definió el objetivo principal de las áreas de la organización?	0			
	3. ¿Se definieron las características de la organización?	<u>0</u>			
	4. Se utilizó un esquema para representar la información obtenida en la evaluación del área?	<u>0</u>			

Semántica del documento

Peso	Indicadores a Evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
Crítico	1. ¿No se han identificado errores ortográficos?	0			

Crítico	2. ¿Se entiende claramente lo que se ha especificado en el documento?	0			
	3. ¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?	0	NP		
	4. ¿El total de páginas que aparecen en las reglas de confidencialidad coincide con el total de páginas que tiene el documento?	0			

Lista de Chequeo Especificación de Requisitos

Estructura del documento					
Peso	Indicadores a Evaluar	Eval	(NP)	Cantidad de elementos	Comentarios

afectados					
Crítico	1. ¿Está el documento acorde con a la plantilla estándar del proyecto o del expediente de proyecto utilizado?	0			
Crítico	2. ¿Contiene las secciones obligatorias definidas en el expediente? (Ver Expediente de Proyecto del Departamento)	0			
Elementos definidos por la metodología					
Peso	Indicadores a Evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
	1. ¿Ha identificado a la (s) persona(s) que lo ayudará a especificar los requisitos?	0			
	2. ¿Ha solicitado la	0			

	participación de diferentes personas para poder definir los requisitos en diferentes puntos de vista?				
	3. ¿Se ha asegurado de capturar lo esencial de cada requisito registrado?	0			
	4. ¿Están todos los requisitos redactados de forma simple y clara para aquellos que vayan a consultarlo en un futuro?	0			
	5. ¿Debería especificarse algún requisito con más detalle?	0			
	6. ¿Debería especificarse algún requisito con menos detalles?	0			
	7. ¿Ha definido el flujo de información de	0			

	manera adecuada para el problema de dominio? (entrevistas, encuestas, equipos de discusión, grupos de trabajo).				
Crítico	8. ¿Ha identificado todas las funciones que el usuario debe hacer?	0			
	9. ¿Ha definido los límites del sistema?	0			
Crítico	10. ¿Todos los requisitos identificados se centran en lo que el sistema debe hacer y no como el sistema debe hacerlo?	0			
Crítico	11. ¿Se han enumerado los requisitos incluso los que se derivan de otros requisitos?	0			
	12. ¿No aparece un mismo requisito en más de un lugar del documento de especificación?	0			

	13. ¿No existe contradicción entre lo especificado por un requisito y lo especificado por otro?	0			
Crítico	14. ¿Cada requisito es verificable? (Un requisito se dice que es verificable si existe algún proceso no excesivamente costoso por el cual una persona o una máquina pueda chequear que el software satisface dicho requerimiento).	0			
Crítico	15. ¿Han sido definidos todos los datos de entrada y salida?	0			
	16. ¿Cada funcionalidad del sistema ha sido representada gráficamente?	0			
Crítico	17. ¿Ha identificado los requerimientos de software y de	0			

	hardware?				
Crítico	18. ¿Han sido identificadas las restricciones de diseño e implementación?	0			
Crítico	19. ¿Han sido identificadas las restricciones de interfaz externa?	0			
Crítico	20. ¿Los requerimientos de soporte y usabilidad se han identificados?	0			
Crítico	21. ¿Se han identificado los requerimientos de seguridad (confidencialidad, integridad, disponibilidad)?	0			
	22. ¿Se puede trazar cada requisito al origen en el entorno del problema, (caso de uso del sistema)?	0			
Crítico	23. ¿Todos los cambios en los requisitos han	0			

	24. ¿Existe correspondencia entre el modelo de caso de uso sistema y las especificaciones de requerimientos?	0			
Semántica del documento					
Peso	Indicadores a Evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
Crítico	1. ¿Ha identificado errores ortográficos?	0			
Crítico	2. ¿Se entiende claramente lo que se ha especificado en el documento?	0			
	3. ¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?	0			
	4. ¿El total de páginas que aparecen en las reglas de confidencialidad coincide con el total de páginas que tiene el	0			

	documento?				
--	------------	--	--	--	--

Lista de Chequeo de la Herramienta para la recolección y análisis de la información.

Estructura del documento					
Peso	Indicadores a Evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
Crítico	¿Está el documento acorde con la plantilla estándar del proyecto o del expediente de proyecto utilizado?	0			
Crítico	¿Contiene las secciones obligatorias definidas en el expediente utilizado por el proyecto?	0			
Elementos definidos por la metodología					
Peso	Indicadores a Evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
Crítico	¿Ha sido identificado el tema de análisis?	0			
	¿Se ha descrito brevemente el tema de análisis	0			

	identificado?				
Crítico	¿Las necesidades del cliente han sido identificadas?	0			
	¿Los niveles de acceso para cada pedido de información se han identificado correctamente?	0			
	¿Han sido identificados los roles que tendrán acceso a ver la información de cada pedido?	0			
Crítico	¿Se ha determinado la periodicidad con que se recoge la información de cada pedido?	0			
Crítico	¿Han sido clasificados los pedidos de información de acuerdo a la información que manejan?	0			
Crítico	¿Los pedidos de información se han agrupado en casos de uso informativos, según un	0			

	critério que los describa?				
Crítico	¿Han sido determinados los criterios para agrupar los pedidos de información en casos de uso del sistema?	0			
Crítico	¿Se han agrupado los pedidos de información en casos de uso del sistema?	0			
Crítico	¿Todas las variables de entrada han sido identificadas?	0			
Crítico	¿Se han identificado jerarquías de las variables de entrada que las poseen?	0			
Crítico	¿Se han determinado los niveles de cada jerarquía?	0			
Crítico	¿Se ha reflejado a que caso de uso del sistema pertenece cada variable de entrada?	0			
	¿Se ha reflejado el tema de análisis al que pertenece cada variable de entrada?	0			

Crítico	¿Todas las variables de salida han sido identificadas?	0			
Crítico	¿Se ha definido la unidad de medida en que se va a medir cada variable de salida?	0			
Crítico	¿Se ha determinado el tipo de variable para cada una de las variables de salida identificadas?	0			
	¿Se ha identificado alguna variable calculable?				No se aplica pues mi negocio no presenta.
Crítico	¿Se ha determinado la fórmula para calcular todas aquellas variables calculables que fueron identificadas?				No se aplica pues mi negocio no presenta.
Crítico	¿Ha sido identificado el tipo de aditividad que presenta cada variable de salida?	0			
Crítico	¿Se identificaron los tipos de operaciones aditivas	0			

	que se pueden realizar?				
Crítico	¿Se ha especificado a que caso de uso del sistema pertenece cada variable de salida?	0			
	¿Se ha especificado a que tema de análisis pertenece cada variable de salida?	0			
Semántica del documento					
Peso	Indicadores a Evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
Crítico	¿Se han identificado errores ortográficos en el Excel?	0			
Crítico	¿Se entiende claramente todo lo especificado en el Excel?	0			

Anexo 2 Lista de Chequeo del Diseño: Modelo de Datos.

Lista de Chequeo Modelo de Datos.

Estructura del documento					
Peso	Indicadores a Evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios

Crítico	¿Está el documento acorde con la plantilla estándar del proyecto o del expediente de proyecto utilizado?	0			
Crítico	¿Contiene las secciones obligatorias definidas en el expediente utilizado por el proyecto?	0			
Peso	Indicadores a Evaluar	Eval (NP)	Cantidad elementos afectados	de	Comentarios
Crítico	¿Las necesidades del cliente han sido identificadas?	0			
	¿Han sido identificados los hechos y dimensiones?	0			
Crítico	¿Se ha confeccionado la Matiz Bus con todas las relaciones existentes entre hechos y dimensiones?	0			
Crítico	¿Han sido especificadas las dimensiones?	0			
Crítico	¿Se ha definido el gránulo?	0			
Crítico	¿Han sido especificados los	0			

	hechos?				
Crítico	¿Se ha ubicado físicamente los archivos?	0			
Crítico	¿Se ha diseñado la estructura física de los datos?	0			
Crítico	¿Se ha identificado el plan de indexado?	0			
Crítico	¿Se ha analizado los resultados de estimación por tablas?	0			
Semántica del documento					
Peso	Indicadores a Evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
Crítico	¿Se han identificado errores ortográficos en el Excel?	0			
Crítico	¿Se entiende claramente todo lo especificado en el Excel?	0			

Anexo 3 Resultados de las pruebas realizadas.

