

Universidad de las Ciencias Informáticas

Facultad 8

“Análisis, Diseño e Implementación del mercado de datos indicadores relacionados con la ciencia e innovación tecnológica para la Oficina Nacional de Estadísticas”

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

Autor:

Yunior Ricardo Hechavarría

Tutores:

Ing. Rosalba Carralero Medina

Ing. Guillermo Solenzal Fernández

Ciudad de La Habana, junio de 2010

“Año 52 de la Revolución”

Frase

“La verdad se robustece con la investigación y la dilación; la falsedad, con el apresuramiento y la incertidumbre”.



Tácito (55-115) Historiador Romano

DECLARACIÓN DE AUTORÍA

Declaro que soy el único autor del trabajo análisis, diseño e implementación del mercado de datos para los indicadores relacionados con la ciencia e innovación tecnológica para la Oficina Nacional de Estadísticas y autorizo a la Facultad 8 de la Universidad de las Ciencias Informáticas a hacer uso del mismo en su beneficio.

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

Yunior Ricardo Hechavarría
Autor

Ing. Guillermo Solenzal Fernández
Tutor

Ing. Rosalba Carralero Medina
Tutor

Agradecimientos

Dedico esta investigación a quien me ha dado el placer de ser el hijo más dichoso del mundo y que sin su apoyo no estuviese aquí: a mi queridísima Madre.

A mi padre por forjarme con carácter y hacerme un hombre de bien.

A mi hermano por hacerme pensar tanto en él cuando no podía a veces ni seguir escribiendo y ser lo más lindo que dio a luz mi Madre.

A Isa y Nene: mis guías y formadores desde mi infancia, y que tanto corrieron a trabajos voluntarios conmigo.

Para mi padrino y mis hermanos de santo: Dixon, Rolando y mi hermano Yasmani, por confiar tanto en mí.

A mis tutores por su apoyo en todo momento.

A Fneida por ser una verdadera amiga y por alentarme siempre.

A los que siempre confiaron en mí y ayudaron a que esto saliera: Franklin, Pantaleón, Meylin, Viquillon, Dayana, Flder a todos los compañeros del aula por estos años tan productivos.

RESUMEN

En el presente trabajo se muestra el resultado del análisis y diseño e implementación del mercado de datos Indicadores relacionados con ciencia e innovación tecnológica para la Oficina Nacional de Estadísticas (ONE) donde existe una gran cantidad de información digitalizada relacionada con la ciencia e innovación de muchos años, dicha información es almacenada en formatos de difícil acceso para su consulta, por eso se hace necesario desarrollar una mercado de datos que facilite la disponibilidad de esta información. Se realiza la solución utilizando una metodología propia del proyecto conocida como Metodología para el desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio (DW&BI) apoyándose en la herramienta CASE Visual Paradigm y el Sistema Gestor de Base de Datos será PostgreSQL en su versión 8.4.

Palabras claves: Mercado de Datos, Almacenes de Datos.

ÍNDICE DE CONTENIDOS

INTRODUCCIÓN.....	1
CAPÍTULO 1 FUNDAMENTACIÓN TEÓRICA	5
1.1 Introducción del capítulo	5
1.2 Sistemas que realizan funciones similares.	5
1.3 Almacenes de Datos (AD)	6
1.3.1 Características del AD	6
1.3.2 Tipos de Modelos de Almacenamiento	7
1.4 OLAP	10
1.5 Mercado de Datos.....	13
1.5.1 Ventajas de usar un MD	13
1.5.2 Desventajas de usar un MD.....	13
1.6 Sistema Gestor de Bases de Datos	14
1.6.1 MySQL.....	15
1.6.1.2 Ventajas	15
1.6.1.3 Desventajas	16
1.6.2 Oracle	16
1.6.2.1 Ventajas	16
1.6.2.2 Desventajas.....	17
1.6.3 Microsoft SQL Server	17
1.6.3.1 Ventajas	17
1.6.3.2 Desventajas.....	18
1.6.4 PostgreSQL	18

1.6.4.1 Ventajas	18
1.6.4.2 Desventajas.....	18
1.6.5 Valoración del gestor escogido.....	19
1.7 Metodologías	20
1.7.1 Kimball	20
1.7.2 Immon.....	20
1.7.3 Metodología para el Diseño Conceptual de Almacenes de Datos.....	21
1.7.4 Valoración de la metodología escogida	21
1.8 Herramientas CASE.....	23
1.8.1 Embarcadero Erwin Studio	24
1.8.2 Case Studio 2	24
1.8.3 ER/Studio.....	25
1.8.4 Visual Paradigm	25
1.8.5 Valoración de la Herramienta CASE escogida	26
1.9 Conclusiones del Capítulo	26
CAPÍTULO 2 ANÁLISIS Y DISEÑO.....	27
2.1 Introducción del Capítulo	27
2.2 Definición del Negocio	27
2.3 Tema de análisis.....	27
2.4 Necesidades de los usuarios	28
2.5 Requisitos de información.....	28
2.6 Requisitos Multidimensionales (RM).....	28
2.7 Requisitos funcionales (RF).....	29
2.8 Requisitos no funcionales	30

2.9 Casos de uso del sistema	32
2.10 Matriz BUS.....	36
2.11 Modelo de Datos.....	37
2.11.1 Dimensiones.....	38
2.11.2 Tablas de hechos	39
2.11.3 Medidas.....	40
2.12 Esquema de Seguridad	40
2.13 Política de respaldo y recuperación.....	41
2.14 Conclusiones del Capítulo	41
CAPÍTULO 3 IMPLEMENTACIÓN Y PRUEBA.....	42
3.1 Introducción del capítulo	42
3.2 Modelos de Datos Físico	42
3.2.1 Estructura de Datos.....	42
3.2.1.1 Esquemas y Tablas.....	42
3.2.1.2 Restricciones y Secuencias.....	43
3.2.1.3 Índices	45
3.2.2 Usuarios y Privilegios	45
3.2.3 Carga de nomencladores	46
3.3 Guía de Implantación.....	47
3.3.1 Secuencia de Pasos.....	47
3.3.2 Requerimientos	48
3.4 Validación y pruebas.....	48
3.4.1 Listas de Chequeo del Análisis.....	48
3.4.2 Lista de Chequeo del Diseño.....	49

3.4.3 Validación de requisitos por el cliente	49
3.4.4 Pruebas de Implantación	49
3.5 Conclusiones del Capítulo	50
CONCLUSIONES GENERALES	51
RECOMENDACIONES.....	52
REFERENCIAS BIBLIOGRÁFICAS.....	53
ANEXOS.....	55

ÍNDICE DE TABLAS

Tabla 1 Comparación entre los SGBD.....	19
Tabla 2 Comparación entre la Metodología de Inmon y Kimball.....	21
Tabla 3 Descripción del Caso de Uso: Analizar Trabajadores Físicos.	33
Tabla 4 Descripción del Caso de Uso Funcional: Extraer datos según su nivel educacional.	35
Tabla 5 Descripción del Caso de Uso Funcional: Transformar y cargar según su nivel educacional.....	36
Tabla 6 Tabla de Hechos.....	39
Tabla 7 Esquemas y Tablas	43
Tabla 8 Restricciones y Secuencias	44
Tabla 9 índices	45
Tabla 10 Roles y Privilegios.....	46
Tabla 11 Pruebas de Implantación	50
Tabla 12 Lista de Chequeo Evaluación de áreas de la organización.	60
Tabla 13 Lista de Chequeo Especificación de Requisitos	65
Tabla 14 Lista de Chequeo de la Herramienta para la recolección y análisis de la información...	69
Tabla 15 Lista de Chequeo Modelo de Datos.....	71

ÍNDICE DE ILUSTRACIONES

Ilustración 1 Estructura de un cubo OLAP..... 8

Ilustración 2 Tipo de Almacenamiento MOLAP 11

Ilustración 3 Tipo de Almacenamiento ROLAP 12

Ilustración 4 Tipo de Almacenamiento HOLAP 12

Ilustración 5 Diagrama de Caso de Uso Informativo: Analizar Trabajadores Físicos. 32

Ilustración 6 Diagrama del Caso de Uso Funcional: Extraer, transformar y cargar datos según su nivel educacional. 34

Ilustración 7 Matriz Bus..... 37

Ilustración 8 Modelo de Datos Físico 38

Ilustración 9 Opinión del Usuario 55

Ilustración 10 Diagrama de Casos de Uso Funcional 56

Ilustración 11 Diagrama de Casos de Uso Informativo..... 57

INTRODUCCIÓN

La informática desde sus inicios ha impulsado el desarrollo intelectual en el mundo, propiciando así un salto inigualable tecnológicamente y extendiéndose hasta los lugares más recónditos; además ha desencadenado una nueva revolución guiada por los avances en la ciencia y las telecomunicaciones.

Cuba trabaja para alcanzar un desarrollo informático de punta, guiado por la telefonía inalámbrica, Internet, la televisión digital y la computación, y en apoyo a esto trabaja en la informatización y digitalización de no solo las empresas cubanas sino también de toda la sociedad. Bajo esta ideología surge La Universidad de las Ciencias Informáticas (UCI), convirtiéndose en un motor impulsor dentro de todos los ámbitos informáticos, pues desarrolla la industria del software cubano, contribuye a la economía del país y propicia la creación de Centros de Desarrollo como lo es el de Tecnología de Datos (DATEC) :quien crea productos y brinda servicios relacionados con las bases de datos, análisis de datos y atiende directamente a la Oficina Nacional de Estadísticas (ONE) quien se encarga de guardar gran cantidad de información .

Situación Problémica: La ONE es el órgano rector de la estadística en Cuba y tiene como objetivo fundamental captar, analizar y difundir los datos recogidos a lo largo y ancho de todo el país; para ello cuenta con diferentes publicaciones que recogen una serie de datos que reflejan los indicadores relacionados con la ciencia e innovación tecnológica para el análisis de los resultados socioeconómicos. Estos datos son almacenados en formatos de difícil acceso para su consulta, por lo que se hace muy complejo el proceso de acceder y divulgar dicha información. La forma de almacenar, recuperar y presentar la información proveniente de las publicaciones dificulta la integración, centralización y la disponibilidad de la información lo que afecta el proceso de toma de decisiones a los Órganos del Estado. A partir de lo anteriormente planteado surge el siguiente **Problema de la Investigación:** ¿Cómo centralizar la información en una estructura que permita la disponibilidad de dicha información en la Oficina Nacional de Estadísticas?

EL **Objeto de Estudio:** Los Almacenes de Datos y el **Campo de Acción:** Mercados de Datos para los indicadores sobre ciencia e innovación tecnológica.

El **Objetivo General** es: Desarrollar el MD¹ para los indicadores sobre ciencia e innovación tecnológica. Este objetivo se desglosa en los siguientes **Objetivos Específicos:**

¹ Mercado de Datos

1. Elaborar el marco teórico de la Investigación acerca de las principales tendencias de implementación de los almacenes de datos, los mercados de datos.
2. Realizar el análisis de los documentos Excel relacionados con los indicadores sobre ciencia e innovación tecnológica.
3. Diseñar el mercado de datos para los indicadores sobre ciencia e innovación tecnológica para el almacén de datos de la ONE.
4. Implementar y cargar los clasificadores para el mercado de los indicadores sobre ciencia e innovación tecnológica para el almacén de datos de la ONE.
5. Validar la solución desarrollada mediante la realización de pruebas.

La **Idea a Defender** de la presente investigación se enmarcaría en que si se realizara el análisis, diseño e implementación del MD para los indicadores sobre ciencia e innovación tecnológica y toda la documentación correspondiente, se lograría una integración más completa de la información, se tendría una mejor disponibilidad de esta, lo que beneficiaría al proceso de toma de decisiones mediante la informatización de los datos. Para desarrollar la investigación se tendrán en cuenta **Métodos Científicos**, pues servirán como una guía para realizar dicha investigación. Estos métodos son procedimientos que estudian la realidad, la sociedad, la naturaleza y el pensamiento para poder descubrir su esencia y sus relaciones.

Como **Métodos Teóricos** se utilizarán:

- **Análisis histórico-lógico:** se utilizará mediante el estudio del estado del arte de los Almacenes de Datos y los Mercados de Datos con el fin de analizar su evolución y desarrollo, así como de la Oficina Nacional de Estadísticas (ONE) para una mayor comprensión del entorno en el que se aplicará el producto final.
- **Analítico-Sintético:** se utilizará durante la investigación de los temas relacionados con los AD, MD y lo relacionado con las herramientas a utilizar para la confección del producto final. Estará presente además durante la identificación de las necesidades de información y el levantamiento de requisitos.
- **Modelación:** se utilizará durante la implementación del Mercado de Datos como posible solución a los problemas existentes.

Como **Métodos Empíricos** se utilizará:

- **Entrevista:** se utilizará mediante conversatorios con el cliente para definir lo que necesita y así definir los requisitos.

Para dar cumplimiento a los objetivos se plantean las siguientes **Tareas:**

1. Definir los conceptos fundamentales relacionados con el desarrollo de MD.
2. Analizar las metodologías de desarrollo de AD para definir cual se va a utilizar.
3. Planificar y realizar entrevistas.
4. Identificar la estructura de usuarios y permisos.
5. Definir los temas de análisis.
6. Identificar las necesidades de información, requisitos funcionales y no funcionales.
7. Modelar los requerimientos.
8. Validar los requerimientos.
9. Definir requisitos de entrada y de salida.
10. Elegir la granularidad del proceso del negocio.
11. Definir las dimensiones del Mercado de Datos (MD).
12. Definir los hechos asociados a las dimensiones definidas.
13. Estructurar el modelo dimensional.
14. Transformar el modelo dimensional al diseño físico.
15. Implementar la Base de Datos.
16. Montar los clasificadores para el mercado de datos indicadores relacionados con el control de la ciencia e innovación tecnológica para el almacén de datos de la Oficina Nacional de Estadística.
17. Realizar pruebas al Mercado de Datos.

Estructura capitular

Capítulo 1: Fundamentación Teórica.

En este capítulo se analizan elementos teóricos relacionados con las características y estructuras de los Gestores de Bases de Datos y la tecnología de AD. Ofrece una descripción de las herramientas que se utilizarán para dar solución al problema planteado.

Capítulo 2: Análisis y Diseño.

En este capítulo se definen los temas de análisis, el negocio, se identifica la estructura de usuarios y permisos así como las necesidades de información, requisitos multidimensionales, funcionales y no funcionales. Se modelan y validan los requerimientos y además se definen los requisitos de entrada y de salida. Se definen los casos de uso y se construye la matriz bus. Además se construye el modelo de datos, el esquema de seguridad y la política de respaldo y recuperación. Se elige la granularidad del proceso del negocio, se definen las dimensiones del MD y los hechos asociados a las dimensiones definidas, se estructura el modelo dimensional, se transforma el modelo dimensional al diseño físico.

Capítulo 3: Implementación y Prueba.

En este capítulo se implementa la Base de Datos, se realiza el montaje de los clasificadores para el Mercado de Datos y se realizan las pruebas al Mercado de Datos.



CAPÍTULO 1 FUNDAMENTACIÓN TEÓRICA

1.1 Introducción del capítulo

En este capítulo se realiza un estudio del estado del arte acerca del MD indicadores relacionados con la ciencia e innovación tecnológica. Además se hace un estudio de algunas metodologías existentes para el desarrollo de un MD, así como un bosquejo de la herramienta que pudiera emplearse para la construcción de la solución propuesta.

1.2 Sistemas que realizan funciones similares.

La mayoría de las empresas a nivel nacional y mundial apuestan por mejorar la forma de almacenamiento de la información que manejan; pues en ocasiones acceder ella o consultarla se hace engorroso. La utilización de los MD como solución eminente a este problema que no está solo en pequeñas empresas sino que interesa hasta a las más grandes compañías. Es un paso para fomentar, conocer y centralizar los datos que cada una posee. Para el desarrollo de la investigación se realiza un estudio exhaustivo de algunas de estas empresas y así saber como se desarrollan los MD relacionados con ciencia e innovación tecnológica a lo largo del país y del mundo.

Después de una búsqueda centrada principalmente en el tema de ciencia e innovación tecnológica no se ha encontrado ningún MD ni AD relacionados con el tema que aborda al trabajo, de ahí que no se mencione ningún ejemplo en este epígrafe.

1.3 Almacenes de Datos (AD)

La informática desde sus inicios acondicionó la forma futura en que se almacenarían y centralizarían los datos. Propició un avance ascendente hasta llegar a la creación de los AD, que ocupan hoy un lugar trascendental en cada renglón de la economía y la sociedad.

Según Ralph Kimball:

- ✓ Es una fuente de datos de la empresa que puede ser consultada.
- ✓ No debe ser organizada con ayuda del modelo entidad/relación

Es frecuentemente modificada, a partir de datos correctos. (Kimball, 1998)

Según W. H. Inmon:

“Un almacén de datos es una colección de datos orientada a temas o materias, integrada, variable en el tiempo y no volátil que será utilizada fundamentalmente en el proceso de toma de decisiones.” (Inmon, 1992)

En resumen un AD es una base de datos (BD), donde los datos están bien definidos, dispuestos y agrupados al negocio, y el atributo tiempo mantiene un equilibrio entre la información y el momento de su creación. No cambia constantemente los datos.

1.3.1 Características del AD

Los AD presentan características que los identifican tales como:

- **Integrados:** Se construye mediante la integración de fuentes de datos múltiples, y heterogéneas. Se aplican técnicas de limpieza e integración.
- **Orientados a temas:** Está organizado en torno a las materias de negocio: cliente, producto, ventas. El modelado y el análisis de los datos se enfocan de cara a la toma de decisiones, no para las operaciones del día a día o en el procesamiento de transacciones. Proporciona una vista simple y concisa de las materias del negocio excluyendo aquellos datos que no se necesitan para el proceso de toma de decisiones.

- **Variables en el tiempo:** Cada clave del almacén contiene una referencia a la fecha explícita o implícitamente y el horizonte de tiempo para el almacén es significativamente más largo que el de los sistemas operacionales.
- **No volátil:** Es un almacén de datos físicamente separado del entorno operacional. Las actualizaciones de la BD operacional no ocurren en el entorno del almacén. No se requieren mecanismos de control de la concurrencia y recuperación. Se requieren dos operaciones nada más: carga de los datos y acceso a datos. (Inmon, 1992)

1.3.2 Tipos de Modelos de Almacenamiento

Los modelos de almacenamiento ayudan a estructurar los datos y para un mejor entendimiento de sus características se describen a continuación tres de estos: el relacional, el dimensional y el multidimensional.

Modelo Relacional

La estructura fundamental del modelo relacional es la relación, es decir una tabla bidimensional constituida por filas (tuplas) y columnas (atributos). Las relaciones representan las entidades que se consideran interesantes en la BD. Cada instancia de la entidad encontrará sitio en una tupla de la relación, mientras que los atributos de la relación representan las propiedades de la entidad. (Quiroz, 2003)

Modelo Dimensional

A diferencia de los clásicos sistemas de BD que presentan sus estructuras diseñadas mediante el modelo Entidad-Relación, los AD se diseñan mediante un Modelo Dimensional. Poseen la misma información que el DER² pero la organiza de forma diferente para garantizar la velocidad y eficiencia en la recuperación de la misma. Una de sus características principales es que no necesita una predefinición de los reportes debido a que se diseñan de forma tal, que cubran el universo de variantes que los usuarios necesitan para consultar la información almacenada. En la ilustración 1 se muestra la estructura espacial que posee este tipo de diseño.

² Modelo Entidad –Relación

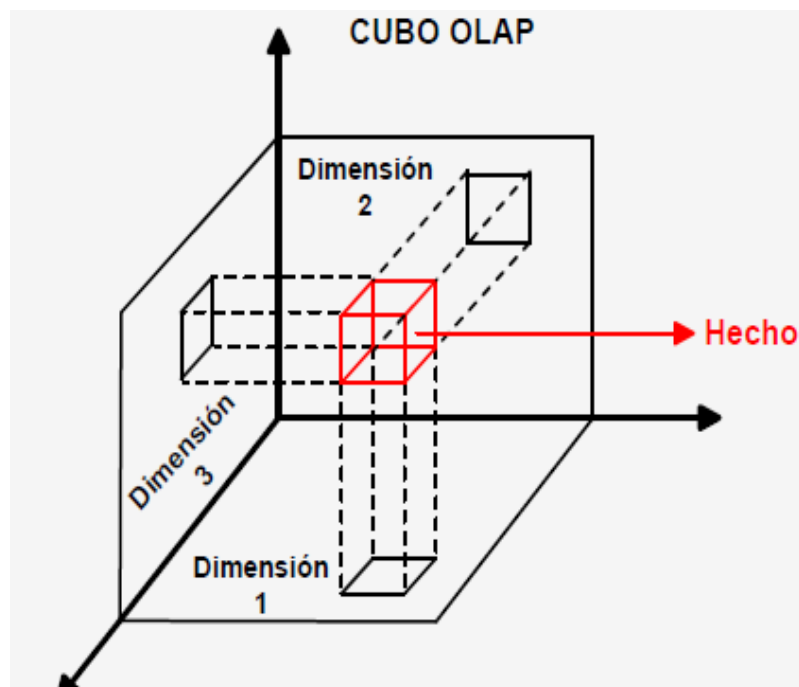


Ilustración 1 Estructura de un cubo OLAP.

El modelo dimensional divide el mundo de los datos en dos grandes conjuntos: las medidas y las descripciones del entorno de estas medidas. Las medidas, que generalmente son numéricas, se almacenan en las tablas de hechos y las descripciones de los entornos que son textuales se almacenan en las tablas de dimensiones. Las tablas de hechos son las tablas primarias en el modelo dimensional y contienen los valores del negocio. Los hechos más comunes son valores numéricos. (Sierra, 2009)

Los elementos de estas tablas se pueden definir de la siguiente manera:

- **Hecho:** Representa un evento o actividad específica, tiene dimensiones y medidas. Un ítem de negocio, una transacción o un evento que tiene significancia para el negocio. Corresponden a una colección de ítems de datos y datos de contexto. Son aquellos datos que residen en una tabla de hechos y que son utilizados para crear indicadores, a través de sumalizaciones preestablecidas. Un hecho debe estar relacionado al menos con una dimensión: “El tiempo”. (Díaz, 2009)
- **Dimensión:** Son las característica de un hecho que permite su análisis posterior, en el proceso de toma de decisiones. Determina el contexto del hecho (quién participó, cuándo y dónde pasó y su tipo). Es una entidad de negocios respecto a la cual se deben calcular las métricas (clientes, productos, tiempo). Tienden a ser discretas y jerárquicas <país, región, departamento, provincia,

distrito>. Es una colección de miembros o unidades o individuos del mismo tipo que permiten categorizar un hecho. (Díaz, 2009)

- **Medida:** Es un atributo numérico de un hecho que representa el performance o comportamiento del negocio relativo a la dimensión. Representan los valores que son analizados. Deben ser numéricas, porque estos valores son las bases de las cuales el usuario puede realizar cálculos. Cruzan todas las dimensiones en todos los niveles. Si la medida es no numérica debemos codificarla a un valor numérico y cuando tengamos que exponerla decodificarla para mostrarla con el valor original. (Díaz, 2009)
- **Jerarquía:** Una jerarquía representa una relación lógica entre los datos de una dimensión. Se presentan al interior de una dimensión. Pueden existir varios niveles (dos o más). Relación “1-n” o “padre-hijo” entre atributos consecutivos de un nivel superior y uno inferior. Es una clasificación u ordenación de abstracciones. (Díaz, 2009)
- **Granularidad:** La granularidad es el nivel de detalle en que se almacena la información. A mayor nivel de detalle, mayor posibilidad analítica, ya que los mismos podrán ser resumidos o sumariados. Los datos con granularidad fina (nivel de detalle) podrán ser resumidos hasta obtener una granularidad media o gruesa. No sucede lo mismo en sentido contrario. (Díaz, 2009)

Modelo Multidimensional

En un modelo de datos multidimensional los datos se organizan alrededor de los temas de la organización. La estructura de datos manejada en este modelo son matrices multidimensionales o hipercubos. Un hipercubo consiste en un conjunto de celdas, cada una se identifica por la combinación de los miembros de las diferentes dimensiones y contiene el valor de la medida analizada para dicha combinación de dimensiones. (Tamayo, 2006)

Este modelo puede representarse en esquemas tales como:

- **Esquema en estrella:** Está formado por una tabla de hechos y una tabla para cada dimensión.
- **Esquema copo de nieve:** Es una variante del esquema en estrella que presenta las tablas de dimensión normalizadas.
- **Constelación de hechos:** Son varios esquemas en estrella o copo de nieve que comparten dimensiones. (Perdomo, 2009)

1.4 OLAP

Los sistemas OLAP³ son BD orientadas al procesamiento analítico. Este análisis suele implicar generalmente, la lectura de grandes cantidades de datos para llegar a extraer algún tipo de información útil: tendencias de ventas, patrones de comportamiento de los consumidores, elaboración de informes complejos. En este sistema:

- El acceso a los datos suele ser de sólo lectura. La acción más común es la consulta, con muy pocas inserciones, actualizaciones o eliminaciones.
- Los datos se estructuran según las áreas del negocio, y los formatos de los datos están integrados de manera uniforme en toda la organización.
- El historial de datos es a largo plazo, normalmente de dos a cinco años.
- Las bases de datos OLAP se suelen alimentar de información procedente de los sistemas operacionales existentes, mediante un proceso de extracción, transformación y carga (ETL). (Sinexus, 2007)

1.4.1 Tipos de almacenamiento OLAP

Son sistemas que permiten recolectar y organizar la información analítica realmente necesaria y disponer inmediatamente de ella en diversos formatos (tablas, gráficos, reportes). Analizan los datos desde diferentes perspectivas (dimensiones) del negocio y soportan análisis complejos de grandes volúmenes de datos.

MOLAP

La arquitectura MOLAP⁴ usa unas BD multidimensionales para proporcionar el análisis, su principal premisa es que el OLAP está mejor implantado almacenando los datos multidimensionalmente. Un sistema MOLAP usa una BD propietaria multidimensional, donde la información se almacena multidimensionalmente para ser visualizada en varias dimensiones de análisis.

³ Procesamiento Analítico en Línea

⁴ Procesamiento Analítico Multidimensional en Línea.

El sistema MOLAP utiliza una arquitectura de dos niveles: las bases de datos multidimensionales y el motor analítico. La BD multidimensional es la encargada del manejo, acceso y obtención del dato. (Sinexus, 2007)

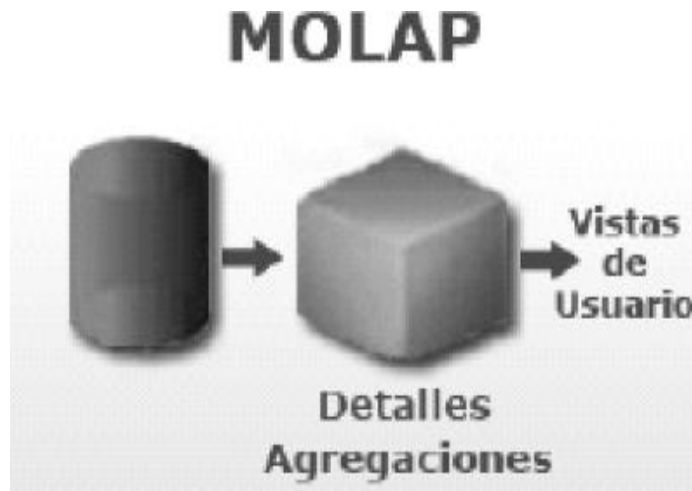


Ilustración 2 Tipo de Almacenamiento MOLAP

ROLAP

La arquitectura ROLAP⁵, accede a los datos almacenados en un AD para proporcionar los análisis OLAP. La premisa de los sistemas ROLAP es que las capacidades OLAP se soportan mejor contra las bases de datos relacionales. El sistema ROLAP utiliza una arquitectura de tres niveles. La BD relacional maneja los requerimientos de almacenamiento de datos, y el motor ROLAP proporciona la funcionalidad analítica.

El nivel de BD usa bases de datos relacionales para el manejo, acceso y obtención del dato. El nivel de aplicación es el motor que ejecuta las consultas multidimensionales de los usuarios.

El motor ROLAP se integra con niveles de presentación, a través de los cuales los usuarios realizan los análisis OLAP. Después de que el modelo de datos para el AD se ha definido, los datos se cargan desde el sistema operacional. Se ejecutan rutinas de bases de datos para agregar el dato, si así es requerido por el modelo de datos. Además se crean los índices para optimizar los tiempos de acceso a las consultas. (Sinexus, 2007)

⁵ Procesamiento Analítico Relacional en Línea.



Ilustración 3 Tipo de Almacenamiento ROLAP

HOLAP

Un desarrollo un poco más reciente ha sido la solución HOLAP⁶, la cual combina las arquitecturas ROLAP y MOLAP para brindar una solución con las mejores características de ambas: desempeño superior y gran escalabilidad. Un tipo de HOLAP mantiene los registros de detalle (los volúmenes más grandes) en la BD relacional, mientras que mantiene las agregaciones en un almacén MOLAP separado. (Sinexxus, 2007)



Ilustración 4 Tipo de Almacenamiento HOLAP

⁶ Procesamiento Analítico Híbrido en Línea.

1.5 Mercado de Datos

Es un subconjunto de datos de un almacén relativo a los requisitos de un departamento o área de negocio concreto. Este subconjunto de datos puede funcionar de forma autónoma, o bien enlazado al Almacén de Datos. (Cabrera, 2009)

En esencia un MD⁷ es una BD creada para un departamento específico, encargado de un área de negocio determinada.

1.5.1 Ventajas de usar un MD

Los MD presentan las siguientes ventajas:

- Mayor rapidez de consulta.
- Consultas SQL y/o MDX⁸ sencillas.
- Validación directa de la información.
- Facilidad para la historización de los datos.
- Permiten manejar información confidencial.
- Reflejan rápidamente sus beneficios y cualidades.
- Reducen la demanda del depósito de datos. (Sinexus, 2007)

1.5.2 Desventajas de usar un MD

- Se pierde capacidad de procesamiento debido al crecimiento de los datos.
- Los usuarios necesitan acceder a varios mercados de datos.
- Compleja administración del mercado de datos.
- Dificultad para construir debido al corto plazo de desarrollo.

⁷ Mercado de Datos.

⁸ Las expresiones multidimensionales (MDX es el acrónimo de Multi-Dimensional Expressions) es un lenguaje de consulta para bases de datos multidimensionales sobre cubos OLAP.

1.6 Sistema Gestor de Bases de Datos

Se define un Sistema Gestor de Bases de Datos o SGBD, también llamado DBMS (Data Base Management System) como una colección de datos relacionados entre sí, estructurados y organizados, y un conjunto de programas que acceden y gestionan esos datos. (McGraw-Hill, 2008)

En efecto un SGBD es una aplicación constituida por datos donde los usuarios definen, crean y mantienen la información coleccionada dentro de una DB y se gestiona por programas específicos .

Un Sistema de Gestión de Base de Datos se divide en módulos que tratan cada una de las responsabilidades del sistema general. Los componentes funcionales de un SGBD incluyen:

- **Procesador de Consultas:** Traduce sentencias en un lenguaje de consultas a instrucciones de bajo nivel que entiende el gestor de la BD.
- **Gestor de la Base de Datos:** Proporciona la interface entre los datos de bajo nivel almacenados en la BD, los programas de aplicación y las consultas que se hacen en el sistema.
- **Gestor de Archivos.** Gestiona la asignación de espacio en la memoria del disco y de las estructuras de datos usadas para representar la información almacenada en disco.
- **Pre compilador del Lenguaje de Manipulación de Datos DML:** Convierte las sentencias en DML⁹ incorporadas en un programa de aplicación en llamadas normales a procedimientos en el lenguaje principal.
- **Compilador del Lenguaje de Definición de Datos DDL:** Convierte sentencias en DDL¹⁰ en un conjunto de tablas metadatos o "datos sobre datos".
- **Gestor del Diccionario de Datos:** Almacena metadatos sobre la estructura de la BD. (Capetillo, 2010)

Los SGBD se desglosan en:

- ✓ SGBD Libres.
- ✓ SGBD Proprietarios.

Dentro de estos dos grupos se analizan brevemente:

SGBD Proprietarios:

⁹ Lenguaje de Manipulación de Datos

¹⁰ Lenguaje de Definición de Datos

- ✓ Oracle (Oracle Corp.)
- ✓ MySQL (SUN Microsystems)
- ✓ DB2 (IBM)
- ✓ InterBase (CodeGear)
- ✓ JavaDB (SUN Microsystems)
- ✓ SQL Server (Microsoft Corp.)

SGBD Libres:

- ✓ PostgreSQL
- ✓ FireBird
- ✓ SQLite (Valmaseda, 2009)

Dentro de los Gestores de Bases de Datos existentes en el mundo sobresalen Oracle, MySQL, Microsoft SQL Server y PostgreSQL. A continuación se proponen algunas características de los mismos.

1.6.1 MySQL

Es un sistema de gestión de BD relacional, cuyo principal objetivo de creación fue la velocidad. Cuenta con una licencia para uso en software privativos, de esta forma las empresas que deseen adquirirla para el desarrollo de software privados es necesario comprarla con este fin. El MySQL es propietario, el cual está sustentado por una empresa MySQL AB privada que posee la mayor parte del derecho del código. Su diseño multi-hilo y multiusuario le permite una gran carga de trabajo, procesando gran cantidad de información.

Es la BD libre más popular de Internet. Es usada actualmente por muchas empresas debido a su fiabilidad y facilidad de uso, por lo que persigue cubrir todas las necesidades del cliente, principalmente la velocidad. Funciona sobre varias plataformas, permitiendo su uso en disímiles lenguajes de programación al acceder a las BD, esta cuenta con una gran cantidad de librerías y herramientas que alcanzan una mayor popularidad entre los sistemas de gestión de BD. Es desarrollado cumpliendo con el estándar SQL, logrando así una gran fiabilidad por los desarrolladores. (Rodríguez, 2009)

1.6.1.2 Ventajas

- ❖ Consume pocos recursos por lo que puede ser ejecutado en máquinas de bajo rendimiento.

- ❖ Ofrece mejor rendimiento al ser usado en aplicaciones web, en Drupal y con su conexión programada en el lenguaje PHP.
- ❖ Los sistemas se pueden conectar usando socket TCP/IP¹¹ en cualquier plataforma. (Rodríguez, 2009)

1.6.1.3 Desventajas

- ❖ Puede provocar problemas de integridad y la información se puede dañar en entornos donde se gestione mucha información simultáneamente, altas concurrencias en las modificaciones de datos.
- ❖ Es privativo lo que trae consigo que cualquier distribución de un producto debe de ser bajo la licencia de GNU PGL¹² o la licencia comercial de MySQL.
- ❖ Su escalabilidad no es muy buena y por eso no es recomendable usarlo en BD muy grandes. (Rodríguez, 2009)

1.6.2 Oracle

Es uno de los manejadores de BD más potente que existe actualmente. Brinda las funcionalidades básicas de un gestor sin olvidar las opciones de subconsultas, no es lento al ser ejecutadas las mismas. Es muy vendido a nivel mundial, y su alto precio es debido a que Oracle es un manejador de BD muy poderoso y solo está al alcance de medianas o grandes empresas como las multinacionales. Consume gran cantidad de recursos en el servidor que trae consigo que se necesite de un buen sistema de hardware, RAM, microprocesador, disco duro. Es capaz de almacenar gran cantidad de información, existen servidores que gestionan hasta muchos teras de capacidad. En el desarrollo de aplicaciones web no está tan extendido debido a que es un sistema muy caro a diferencias de otros sistemas de gestión de BD como MySQL, SQL Server. Oracle se basa en tecnología cliente-servidor, cuenta con herramientas necesarias para su utilización en el servidor, así como con herramientas de programación básicas de Oracle. (Rodríguez, 2009)

1.6.2.1 Ventajas

- ❖ Permite crear los formularios de forma similar a otras herramientas de programación conocidas como Visual Basic, Visual C.

¹¹ Protocolo de control de transmisión/Protocolo de Internet.

¹² Licencia Pública General GNU.

- ❖ Puede ejecutarse en varias plataformas, desde una PC hasta un servidor que cuente con diferentes sistemas operativos.
- ❖ Usa particiones para mejorar la eficiencia de replicación y algunas versiones admiten administración de bases de datos distribuidas.
- ❖ Es un sistema muy complejo que permite desarrollar diseños y procedimientos almacenados con una integridad referencial bastante potente.
- ❖ Capaz de almacenar una amplia variedad de información y recuperarla de forma rápida y fácil. Permite a las organizaciones conocer más de su negocio.

1.6.2.2 Desventajas

- ❖ Es muy riesgoso el proceso de controlador de versiones, ya que es muy engorroso y puede perderse o dañarse la información.
- ❖ Las últimas versiones presentan algunos fallos en el sistema de almacenamiento por causa de las extensiones orientadas a objetos.
- ❖ Una de sus mayores desventajas radica en su precio, incluso las licencias son excesivamente caras. (Rodríguez, 2009)

1.6.3 Microsoft SQL Server

Es un SGBD relacional que utiliza lenguaje SQL, a través del cual se puede acceder a grandes cantidades de datos de manera simultánea. Constituye la alternativa de Microsoft, es un sistema muy utilizado en aplicaciones de escritorio y otros sistemas de gestión, preferentemente con productos de Microsoft.

1.6.3.1 Ventajas

- ❖ Es un sistema barato.
- ❖ Cuando el SQL Server no tiene tareas del usuario que procesar comienza la llamada tarea de limpieza lo que hace mejorar el funcionamiento del CPU, necesita menor limpieza en las memorias intermedias durante el procesamiento de transacciones.
- ❖ Registra las transacciones de forma tal que las actualizaciones se pueden reducir al último estado en caso del servidor fallar.
- ❖ Soporta transacciones.

1.6.3.2 Desventajas

- ❖ Presenta bloqueos a nivel de página.
- ❖ Dispositivos con crecimiento manual.
- ❖ Su licencia es propietaria.
- ❖ Soportado por el sistema operativo Windows. (Rodríguez, 2009)

1.6.4 PostgreSQL

Es un servidor de BD relacional de código abierto orientado a objetos, publicado bajo la licencia BSD¹³ y creado en software libre por una comunidad de desarrolladores y organizaciones comerciales. El PostgreSQL permite que mientras unos usuarios accedan a una tabla otros pueden trabajar en ella sin necesidad de bloqueos. Cuenta con un entorno administrativo amigable y fácil de usar. Es un sistema objeto-relacional, ya que incluye la herencia, tipos de datos, funciones, restricciones, disparadores, reglas e integridad transaccional.

Es el sistema libre más avanzado, soporta la gran mayoría de las transacciones SQL. Trabaja con varios lenguajes de programación como son: C, C++, Java, Python, PHP.

1.6.4.1 Ventajas

- ❖ Soporta múltiples tipos de datos: como los tipos de fecha, monetarios, elementos gráficos, datos de redes y también creación de tipos de datos propios.
- ❖ Permite la creación de funciones personalizadas y el manejo y configuración de disparadores.
- ❖ Posee una gran escalabilidad.
- ❖ Es capaz de ajustarse al número de CPU y la cantidad de memoria que ofrece el sistema de forma óptima, siendo capaz de soportar mayor cantidad de peticiones simultáneas sin provocar errores.
- ❖ Comprueba la integridad referencial y almacena procedimientos en su propia BD.

(Rodríguez, 2009)

1.6.4.2 Desventajas

- ❖ La velocidad de respuestas que ofrece este gestor es un poco deficiente.

¹³ Distribución de Software Berkeley.

Capítulo 1: Fundamentación Teórica

- ❖ Tienen un límite de 8k por filas que se pueden aumentar hasta 32, lo que trae consigo que disminuya su rendimiento.
- ❖ Consume gran cantidad de recursos cuando se realizan muchas transacciones de datos.

Criterios	MySQL	SQL Server	PostgreSQL	Oracle
Sistemas Operativos	Linux/Windows	Windows	Linux/Windows	Windows, Unix , Macintosh y Mainframes
Consumo de recursos	Consume pocos recursos	Consume pocos recursos	Consume pocos recursos	Consume gran cantidad de recursos
Velocidad	Rápido	Rápido	Lento	Rápido
Costo	Muy Costoso	Muy Costoso	Gratis	Muy Costoso
Rendimiento	Alto	Alto	Alto	Alto
Comercialización	Privativo	Privativo	Libre	Privativo
Volumen de Datos	Alto	Alto	Alto	Alto
Integridad	Baja	Alta	Alta	Alta

Tabla 1 Comparación entre los SGBD

1.6.5 Valoración del gestor escogido

Después de un exhaustivo análisis de los gestores de bases de datos, viendo sus ventajas y desventajas, se utiliza PostgreSQL en su versión 8.4, pues se centra en la restauración de la BD usando procesos paralelos, acelerando la recuperación de un respaldo hasta en 8 veces respecto a la versión anterior. Tiene privilegios por columna, para poder controlar el acceso a un nivel de detalle mayor. Presenta configuración de idioma y ordenamiento por BD. Contiene nuevas herramientas para monitorear las consultas, entregando mayor información a los administradores para saber lo que está sucediendo en la

BD y mejoras en procedimientos almacenados, por ejemplo: usar valores por omisión en la declaración de parámetros o listas de argumentos de largo variable, al más puro estilo C++/Java. Además de las nuevas características, esta nueva versión incluye fuertes mejoras en el rendimiento, sobre todo en operaciones complejas. (Catrin, 2009)

1.7 Metodologías

Las metodologías son un conjunto de pasos y procedimientos que deben seguirse para desarrollar un software. A continuación se reflejan algunas de ellas.

1.7.1 Kimball

Se basa en dividir el mundo de Inteligencia de Negocio entre los Hechos y las Dimensiones, esta es eficaz y conduce a una solución completa en un corto período de tiempo. Además, tiene abundante documentación y se puede encontrar una respuesta a casi todas las preguntas que se puedan tener. Entre sus características principales están que su arquitectura es ascendente, plantea que se debe crear por cada departamento un conjunto de MD independientes, orientados a los temas que estén relacionados con él. (Curto, 2009)

1.7.2 Inmon

La visión de Inmon se basa en un enfoque descendente, propone construir primero el Almacén de Datos y a partir de este los MD, plantea la creación de un repositorio de datos corporativo como fuente de información consolidada, persistente, histórica y de calidad. Al ser construido descendentemente los Mercados de Datos se nutren del AD Corporativo, convirtiéndose en un complejo empresarial de bases de datos relacionales. Inmon afirma que la creación de una BD relacional con una leve normalización necesita ser las bases para los MD, por lo que no se crean los MD directamente desde los Sistemas de Procesamiento de Transacciones en Línea (OLTP) a través de un área de ensayo. (Inmon, 2005)

Metodología y Arquitectura	Inmon	Kimball
Visión global	Top-Down	Bottom-Up

Estructura	El AD empresarial alimenta” las Base de Datos Departamentales.	Los AD modelan un sólo proceso de negocio.
Complejidad del método	Bastante complejo	Bastante simple
Metodología de Desarrollo	Derivada de la metodología en espiral.	Proceso de 4 pasos
Diseño	Bastante riguroso	Bastante ligero
Orientación de los Datos	Orientado a los datos o tema	Orientado a proceso
Herramientas	Tradicional	Modelo dimensional, surgido a partir del modelo relacional
Accesibilidad del usuario final	Baja	Alta

Tabla 2 Comparación entre la Metodología de Inmon y Kimball.

1.7.3 Metodología para el Diseño Conceptual de Almacenes de Datos

Creada en la Tesis de Doctorado de Leopoldo Zenaido Zepeda Sánchez, se destaca por la introducción de los Casos de Uso para guiar el proceso de desarrollo. Además procede con transformaciones para convertir un diagrama relacional a uno dimensional para conformar el repositorio de datos con las estructuras obtenidas. Aporta además ventajas debido a que el esquema multidimensional final se encuentra fortificado con la Base de Datos Operacional por lo que el proceso de ETL (Extracción, Transformación y Carga) es más factible. Unido a esto emplea una notación estandarizada para construir el esquema multidimensional. (Sánchez, 2008)

1.7.4 Valoración de la metodología escogida

Para definir 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) creada por DATEC¹⁴, se utiliza como base la Metodología Kimball por los siguientes elementos:

¹⁴ Centro de Tecnología de Datos.

Capítulo 1: Fundamentación Teórica

- Crea los conceptos de Hechos y Dimensiones, lo que indudablemente es muy eficaz en el proceso de la toma de decisiones y proporciona mayor agilidad en el proceso de desarrollo.
- Propone ir construyendo el Almacén de Datos a través de la construcción de los Mercados de Datos departamentales, lo que constituye una buena estrategia y coincide con la división lógica de las empresas, entidades y organismos.
- Es una metodología madura y reconocida por el resto de la comunidad dedicada al tema. Tiene bien definidas las etapas, actividades, artefactos y roles.

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 para lograr un mejor alineado con las tendencias y normas de la Universidad.

Dentro del ciclo de vida de esta metodología existen una serie de flujos de trabajo mencionados a continuación:

Estudio Preliminar o Planeación: Se realiza un estudio minucioso en la entidad cliente. Esto incluye un diagnóstico de información, de datos y de infraestructura tecnológica.

Requerimientos: Es llevado a cabo por el Grupo de Análisis, se realiza en dos direcciones, una, identificando las necesidades de información y reglas de negocio; y la otra haciendo un levantamiento detallado de cada una de las distintas fuentes de datos a integrar.

Arquitectura y Diseño: Aquí participan los tres grupos fundamentales: Extracción, Transformación y Carga, Almacenes de Datos e Inteligencia de Negocio. En la definición de la arquitectura participan los arquitectos de cada uno de los grupos mencionados y en el diseño participan igualmente los tres grupos pero se incrementa considerablemente la cantidad de personas, todo depende de la complejidad de la solución.

Implementación: Participan los tres grupos de desarrollo (ETL¹⁵, DWH y BI¹⁶). Se lleva a cabo el diseño físico del Repositorio de Datos, se crean las estructuras de almacenamiento con las particiones y agregaciones correspondientes según la solución en desarrollo.

¹⁵ Extracción, Transformación y Carga.

¹⁶ Inteligencia de Negocios.

Prueba: Se realizan varias pruebas, comenzando por las Pruebas de Unidad, llevadas 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.

Despliegue: Este flujo 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 carga una muestra de los datos en un ambiente controlado, con el fin de demostrarle al Cliente Final que la solución funciona.

Soporte y Mantenimiento: pueden ser desde un soporte en línea, vía telefónica, web, correo u otras, hasta el acompañamiento junto al cliente.

Gestión y Administración del Proyecto: es llevado a cabo por el Grupo de Dirección del Proyecto. Gestiona y chequea todo el desarrollo, los gastos, las utilidades, los recursos, las adquisiciones, y demás actividades relacionadas con la gestión de proyecto.

Como se ha descrito, en cada flujo intervienen grupos específicos, cada uno con actividades y responsabilidades concretas, a continuación se mencionan los mismos, quienes son los encargados de todo el trabajo:

- ✓ Grupo de Análisis.
- ✓ Grupo de Almacenes de Datos (DWH).
- ✓ Grupo de ETL.
- ✓ Grupo de BI.
- ✓ Grupo de Dirección.

1.8 Herramientas CASE

Actualmente la tendencia es hacia un mundo heterogéneo, en el cual convivan diversos productos que se complementen y en ese contexto contar con herramientas de desarrollo abiertas con conectividad a diversas plataformas. Las herramientas CASE¹⁷ participan en el modelado de proyectos, flujos de información, entidades de datos y requerimientos del sistema; facilitan el proceso de planeación y mucha de la información que se captura durante esta fase se usará también durante el desarrollo y el mantenimiento del ciclo de vida del software. Las herramientas CASE son un conjunto de programas y

¹⁷ Ingeniería de Software Asistida por Computación.

ayudas que dan asistencia a los analistas, ingenieros de software y desarrolladores, durante todos los pasos del ciclo de vida de desarrollo de un software. A continuación se describen algunas de las herramientas CASE:

1.8.1 Embarcadero Erwin Studio

Es una herramienta para la arquitectura y modelamiento de BD. Provee de mejores reportes y mejor soporte para la implementación de prácticas y estándares en modelos complejos que contienen metadatos críticos. Mejora la calidad del modelo de datos y reduce los riesgos asociados a la seguridad de la información. Es una herramienta multinivel destinada para la construcción del modelo físico y lógico. Está equipado para crear y manejar diseños de BD grandes y complejos. Cuenta con grandes capacidades de diseño lógico, construcción automática de la BD, permite realizar ingeniería inversa partiendo del modelo físico al modelo lógico. Brinda mayor soporte en la integración de almacenamiento de datos diseñados para visualizar, documentar y compartir el conocimiento. Soporta como sistemas de gestión de BD a Oracle, Sybase System, Microsoft SQL Server, IBM D B/2 Universal, InterBase 4, Microsoft Access y Microsoft Visual Fox Pro. (Rodríguez, 2009)

1.8.2 Case Studio 2

Es un programa muy usado en la modelación de BD, permite crear y mantener DER¹⁸ y DFD¹⁹, en él se crean scripts SQL de forma automática que funcionan sobre diferentes sistemas de BD, además se pueden utilizar plantillas para sistemas con características similares. Presenta dos variantes, una gratuita que es bastante completa, tan sólo presenta limitaciones de pago, y otra versión completa que ofrece grandes y completas funcionalidades. Trata aspectos como contenidos y dominios para dar mejor seguridad al sistema. Contiene catálogos para guardar el modelo de diseño. Brinda soporte para ingeniería inversa, que permite identificar y estructurar BD ya existentes para poder trabajar con ellas. Facilita el trabajo de documentación al generar informes detallados en HTML²⁰ y RTF²¹, además de la documentación de diagramas de flujo de datos exportables en formato XML²². Exporta para varios

¹⁸ Diagramas de Entidad Relación.

¹⁹ Diagramas de Flujo de Datos.

²⁰ Lenguaje de Marcado de Hipertexto.

²¹ Formato de Texto Enriquecido.

²² Lenguaje de Anotación Extensible.

gestores de BD como son, Oracle, MySQL, PostgreSQL, Access, MS SQL, Max DB, Firebird. Tiene una interfaz gráfica muy intuitiva y fácil de usar. Soportado en diferentes sistemas operativos, Windows 95/98/Me/NT/2000 /XP. Al crear el DER, considera cada BD de opciones tales como integridad referencial, restricciones, dominios y disparadores. (Rodríguez, 2009)

1.8.3 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 lógico y físico. Direcciona las necesidades diarias de los administradores de bases de datos, desarrolladores y arquitectos de datos que construyen y mantienen aplicaciones de BD grandes y complejas. El mismo ofrece funcionalidades tales como: capacidad fuerte en el diseño lógico, sincronización bidireccional de los diseños lógicos y físicos, construcción automática de BD, ingeniería inversa, documentación basada en HTML, un repositorio para el modelado.

ER/Studio es una excelente herramienta para el modelado de bases de datos, una de sus grandes potencialidades es la cantidad de SGDB con los cuales es compatible entre ellos se pueden citar:

Oracle 7.3®, 8.x & 9i

Sybase® System 11.9.2, 12.x & 12.5

Microsoft® SQL Server 6.5, 7 & 2000

Open Systems, OS/390® & AS/400 4.5

SQL Anywhere™ and Watcom™ SQL

1.8.4 Visual Paradigm

Es una herramienta CASE de modelado que utiliza UML como lenguaje de modelado profesional, soporta el ciclo de vida completo del desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue. Permite realizar ingeniería tanto directa como inversa. La herramienta es 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 el control de versiones. Cabe destacar igualmente su robustez, usabilidad y portabilidad. En definitiva, Visual Paradigm es una herramienta a tener en cuenta a la hora de ponerse manos a la obra con un proyecto importante. (Rodríguez, 2009)

1.8.5 Valoración de la Herramienta CASE escogida

Para un mejor modelamiento del diseño y así prever la eficiente construcción del MD se utilizará Visual Paradigm debido a que:

Es una herramienta de diseño que soporta todos los diagramas UML, SysML²³, esquemas y diagramas de entidad-relación. Genera código Java. Permite soportar a arquitectos de sistemas, diseñadores, desarrolladores, analistas de procesos de negocio y modeladores de datos en los procesos de desarrollo de software. No es libre pero la universidad paga por su licencia debido a que la utiliza y es muy popular por su uso.

1.9 Conclusiones del Capítulo

En este capítulo se refleja la investigación sobre el tema de ciencia e innovación tecnológica, detectándose que no se ha encontrado ningún MD ni AD relacionados con el tema que aborda el trabajo. Además se describe el uso de las diferentes tecnologías, metodologías y estándares para el posterior desarrollo del mercado de datos indicadores sobre ciencia e innovación tecnológica, quedando definido para la realización del producto final la utilización de PostgreSQL en su versión 8.4 como GBD. Además se toma la metodología definida por DATEC nombrada: 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) luego de un estudio realizado por la institución y tomando como base en la creación de esta, la Metodología Kimball y como complemento a la misma y fortaleciendo la etapa del levantamiento de requisitos lo planteado por Leopoldo Zenaido Zepeda Sánchez en su Tesis de Doctorado. Se dispone de Visual Paradigm como Herramienta CASE a utilizar y se propone de manera general, una solución que integra los elementos estudiados y se adapta a la arquitectura definida para el desarrollo del proyecto.

²³ Por sus siglas en inglés (System Modeling Language): es un lenguaje de especificación de sistemas. Este lenguaje es un subconjunto ampliado de UML 2.0



CAPÍTULO 2 ANÁLISIS Y DISEÑO

2.1 Introducción del Capítulo

En este capítulo se tiene en cuenta la definición del negocio, el tema de análisis. Se definen los roles y permisos; así como las reglas del negocio y las necesidades de los usuarios; también los requisitos de: información, funcionales, no funcionales y multidimensionales. Se conforman los casos de uso, se construye la matriz bus, el modelo datos, se definen: las dimensiones, los hechos y las medidas. Se realiza el esquema de seguridad y se describe la política de respaldo y recuperación.

Análisis

El análisis consiste en obtener una visión del sistema y se interesa por los requisitos funcionales. Es descrito en el lenguaje de los desarrolladores y analiza con profundidad los requisitos funcionales. Esboza de forma clara cómo llevar a cabo las funciones dentro del sistema, incluida la funcionalidad significativa para la arquitectura; sirve como una primera aproximación al diseño.

2.2 Definición del Negocio

La ONE es la entidad encargada de recopilar, organizar y ejecutar en correspondencia con la política estatal en materia de estadística del país. Está compuesta por diversas direcciones como lo es la de Estadísticas Sociales, encomendada a atender mediante especialistas dedicados al estudio de la información estadística toda la información relacionada con los indicadores y los temas de la ciencia e innovación tecnológica.

2.3 Tema de análisis

El tema de análisis es el centro de la investigación y constituye para la definición de todos los aspectos relacionados con la solución un patrón a seguir. El seleccionado en el trabajo es: Ciencia y Tecnología.

2.4 Necesidades de los usuarios

Las necesidades de los usuarios se enfocan en la publicación de los indicadores de ciencia e innovación tecnológica.

2.5 Requisitos de información

Los requisitos de información (RI) describen qué información debe almacenar el sistema para satisfacer las necesidades de los clientes y usuarios. Identifican los conceptos relevantes sobre los que se debe almacenar información y los datos específicos que son de interés. A continuación se mencionan los RI:

- Obtener la cantidad de trabajadores según su nivel educacional por año.
- Obtener la cantidad de trabajadores según categoría ocupacional por año.
- Obtener gastos por tipos de actividades por año.
- Obtener gastos por fuente financiamiento por año.
- Obtener inversiones ejecutadas por componentes por año.
- Obtener registro de patentes de invenciones por año.
- Obtener registro de patentes de modelos industriales por año.
- Obtener cantidad de grados científicos otorgados por año.
- Obtener coeficiente de invenciones por año.

2.6 Requisitos Multidimensionales (RM)

Los RM describen en síntesis las entradas y salidas dirigidas al diseño del mercado, los definidos en este trabajo se describen a continuación:

VE: variable entrada **VS:** variable salida.

- **RM-1** Obtener la cantidad de trabajadores según su nivel educacional por año.
VE: Nivel Educacional, año.
VS: Cantidad de Trabajadores.
- **RM- 2** Obtener la cantidad de trabajadores según categoría ocupacional por año.
VE: Categoría Ocupacional, año.

VS: Cantidad de Trabajadores.

- **RM- 3** Obtener gastos por tipos de actividades por año.

VE: Tipo de Actividad, año.

VS: Gasto.

- **RM-4** Obtener gastos por fuente de financiamiento por año.

VE: Fuente Financiamiento, año.

VS: Gasto.

- **RM-5** Obtener inversiones ejecutadas por año.

VE: Componente, año.

VS: Inversión.

- **RM-6** Obtener registro de patentes de invenciones por año.

VE: Estado Legal, Tipo Solicitud, año.

VS: Cantidad de Solicitudes de registro de invenciones.

- **RM-7** Obtener registro de patentes de modelos industriales por año.

VE: Estado Legal, Tipo Solicitud, año.

VS: Cantidad de Solicitudes de Registro de Modelos Industriales.

- **RM-8** Obtener cantidad de grados científicos otorgados por año.

VE: año

VS: Cantidad de grados científicos otorgados.

- **RM-9** Obtener coeficiente de invenciones por año.

VE: año.

VS: Cantidad de solicitudes nacionales por cada 100 000 habitantes.

2.7 Requisitos funcionales (RF)

Los RF describen los casos de uso que los diferentes actores utilizan de los servicios proporcionados por el sistema. Cada requisito funcional identifica el evento de activación, los pres y post-condiciones y los

pasos que componen el caso de uso, así como las posibles excepciones. Los RF son capacidades o condiciones que el sistema debe cumplir. A continuación se mencionan todos los RF definidos:

Extraer información del Excel:

- Trabajadores físicos en la actividad de ciencia y tecnología según nivel educacional.
- Trabajadores físicos en la actividad de ciencia y tecnología según categoría ocupacional.
- Gasto Total en actividades de ciencia y tecnología por tipo de actividades.
- Gastos corrientes en actividades de ciencia y tecnología por fuente de financiamiento.
- Inversiones ejecutadas en la actividad de ciencia y tecnología por componentes.
- Registro de patentes de invenciones presentadas en Cuba.
- Registro de patentes de modelos industriales presentados en Cuba.

Transformar y cargar del Excel:

- Trabajadores físicos en la actividad de ciencia y tecnología según nivel educacional.
- Trabajadores físicos en la actividad de ciencia y tecnología según categoría ocupacional.
- Gasto Total en actividades de ciencia y tecnología por tipo de actividades.
- Gastos corrientes en actividades de ciencia y tecnología por fuente de financiamiento.
- Inversiones ejecutadas en la actividad de ciencia y tecnología por componentes.
- Registro de patentes de invenciones presentadas en Cuba.
- Registro de patentes de modelos industriales presentados en Cuba.

2.8 Requisitos no funcionales

Los RNF²⁴ describen aquellas características no funcionales que los clientes y usuarios desean que tenga el sistema a desarrollar. Son propiedades o cualidades que el producto debe tener. Todos los analizados se exponen a continuación :

RNF1: Restricciones en el diseño y la implementación

²⁴ Requisitos No Funcionales

- El producto se desarrollará con PostgreSQL 8.4.y Visual Paradigm será el que modele todo el diseño.

RNF2: Apariencia o interfaz externa

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

RNF3: Usabilidad

- Diseño orientado a la simplicidad, con mínimas distracciones y una clara arquitectura de la información.
- Las personas que tramiten los reportes deberán tener conocimientos básicos de computación.
- Debe presentar interfaz amigable.

RNF4: Portabilidad

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

RNF5: Rendimiento

- Para cargar los datos y ejecutar las consultas el tiempo no debe exceder de los 5 segundos.
- 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.

RNF6: 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.

2.9 Casos de uso del sistema

Los CUS²⁵ son artefactos narrativos que describen, bajo la forma de acciones y reacciones, el comportamiento del sistema desde el punto de vista del usuario. Es un proceso que da un resultado de valor para un actor determinado y una secuencia de actividades a automatizar.

Casos de Uso de Información

El trabajo cuenta con 4 casos de uso de información y 14 funcionales. A continuación se describe 1 CUI²⁶ y 2 CUF²⁷.

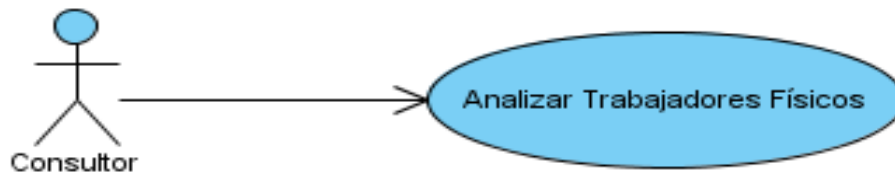


Ilustración 5 Diagrama de Caso de Uso Informativo: Analizar Trabajadores Físicos.

Caso de Uso:	Analizar Trabajadores Físicos
Tipo:	Información
Actores:	Administrador, Consultor
Resumen:	Se analiza el caso de uso Trabajadores Físicos y con él sus variables: de entrada, de salida y su periodicidad.
Precondiciones:	Compleitud del almacén. Carga de los Datos.

²⁵ Casos de Uso del Sistema.

²⁶ Caso de Uso Informativo.

²⁷ Caso de Uso Funcional.

Referencias	Requisitos de Información: Obtener la cantidad de trabajadores según su nivel educacional por año, obtener la cantidad de trabajadores según categoría ocupacional por año y Obtener cantidad de grados científicos otorgados por año.	
Prioridad	Crítico	
Flujo Normal de Eventos		
Sección “”		
Acción del Actor	Respuesta del Sistema	
1. Entra al sistema.	1.1 Muestra opciones de reportes.	
2. Selecciona el reporte Trabajadores Físicos	2.1 Muestra tipos de reportes Trabajadores Físicos.	
Opciones de reportes		
Perspectivas de análisis	Posibles resultados	
	Variables de Salida	Periodicidad
Variables de entrada disponibles relacionadas con el Caso de Uso Trabajadores Físicos. <ul style="list-style-type: none"> ✓ Nivel Educacional ✓ Categoría Ocupacional ✓ Temporal 	-cantidad trabajadores -cantidad de grados científicos otorgados	-anual
Prototipo de Interfaz		
Pos condiciones	Disponibilidad de opciones (cruces de variables) de reportes relacionados con el hecho Trabajadores Físicos.	

Tabla 3 Descripción del Caso de Uso: Analizar Trabajadores Físicos.

Casos de Uso Funcionales

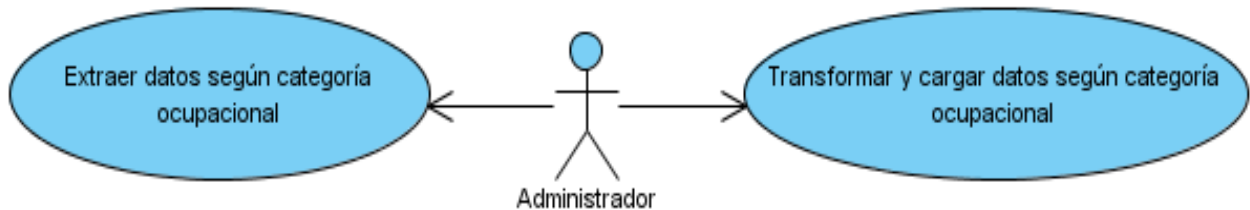


Ilustración 6 Diagrama del Caso de Uso Funcional: Extraer, transformar y cargar datos según su nivel educacional.

Caso de Uso:	Extraer datos del documento según nivel educacional.
Tipo:	Funcional
Actores:	Administrador
Resumen:	Se procede a la extracción de los datos.
Precondiciones:	Disponibilidad de las fuentes.
Referencias	Requisito Funcional: Extraer información del Excel Trabajadores físicos en la actividad de ciencia y tecnología según nivel educacional.
Flujo Normal de Eventos	
Sección “”	
Acción del Actor	Respuesta del Sistema
1. Realiza la conexión al Excel Trabajadores Físicos en la actividad de ciencia y tecnología según categoría ocupacional.	1.1 Responde a la solicitud de conexión.
2. Selecciona estructuras o archivos a extraer.	
3. Realiza la extracción.	
Prototipo de Interfaz	
Flujos Alternos	
Acción del Actor	Respuesta del Negocio
	1.2 No responde a solicitud de conexión.

1. Notifica el error a administrador del Excel Trabajadores Físicos en la actividad de ciencia y tecnología según categoría ocupacional.	
Flujos Alternos	
2. Si hay control de cambios, verifica si hay modificaciones. En caso de que no haya, va al paso 3 del flujo normal.	
Flujos Alternos	
3. En caso de que haya, va al paso 2 del flujo normal.	
Prototipo de Interfaz	
Pos condiciones	Datos disponibles para transformar.

Tabla 4 Descripción del Caso de Uso Funcional: Extraer datos según su nivel educacional.

Caso de Uso:	Transformar y cargar datos su nivel educacional.	
Tipo:	Funcional	
Actores:	Administrador	
Resumen:	Se procede a la transformación y carga del documento Excel.	
Precondiciones:	Extracción completada en área temporal. Estructuras del almacén disponibles.	
Referencias	Requisito Funcional: Transformar y cargar datos del Excel Trabajadores Físicos en la actividad de ciencia y tecnología según su nivel educacional.	
Prioridad	Crítico	
Flujo Normal de Eventos		
Sección ""		
Acción del Actor.	Respuesta del Sistema	
1. Selecciona estructuras del área temporal a transformar.		
2. Carga datos seleccionados en memoria.		

3. Aplica transformaciones pertinentes y genera datos de auditoría.	
4. Carga datos en el almacén.	
Prototipo de Interfaz	
Flujos Alternos	
Acción del Actor	Respuesta del Negocio
Prototipo de Interfaz	
Pos condiciones	Datos del Excel Trabajadores Físicos en la actividad de ciencia y tecnología según su nivel educacional.

Tabla 5 Descripción del Caso de Uso Funcional: Transformar y cargar según su nivel educacional.

Diseño

El diseño es un refinamiento del análisis que tiene en cuenta los requisitos no funcionales y cómo cumple el sistema sus objetivos. El diseño debe ser suficiente para que el sistema pueda ser implementado sin ambigüedades. En el diseño se modela el sistema incluyendo la arquitectura, para que soporte los requisitos, incluyendo los no funcionales y las restricciones que se le suponen.

2.10 Matriz BUS

El propósito de la Matriz Dimensional es obtener un modelo lógico inicial, donde queden identificadas todos los hechos y dimensiones relacionadas entre sí.

Dimensiones:

- 1. Nivel Educacional
- 2. Categoría Ocupacional
- 3. Tipo Actividades
- 4. Fuente Financiamiento
- 5. Componentes
- 6. Tipo Solicitud
- 7. Estado Legal Temporal
- 8. Temporal

Hechos	Dim-1	Dim-2	Dim-3	Dim-4	Dim-5	Dim-6	Dim-7	Dim-8
Trabajadores Físicos	x	x						x
Gasto			x	x				x
Inversiones Ejecutadas					x			x
Registro Patentes						x	x	x
Grado Científico								x
Coeficiente Invenciones								x

Ilustración 7 Matriz Bus

2.11 Modelo de Datos

Dispuesta la información y establecida la relación entre los hechos y las dimensiones en la matriz bus dentro del modelo de datos se describe cómo se almacenan los datos en el ordenador, el formato de los registros, la estructura de los ficheros y los métodos de acceso utilizados.

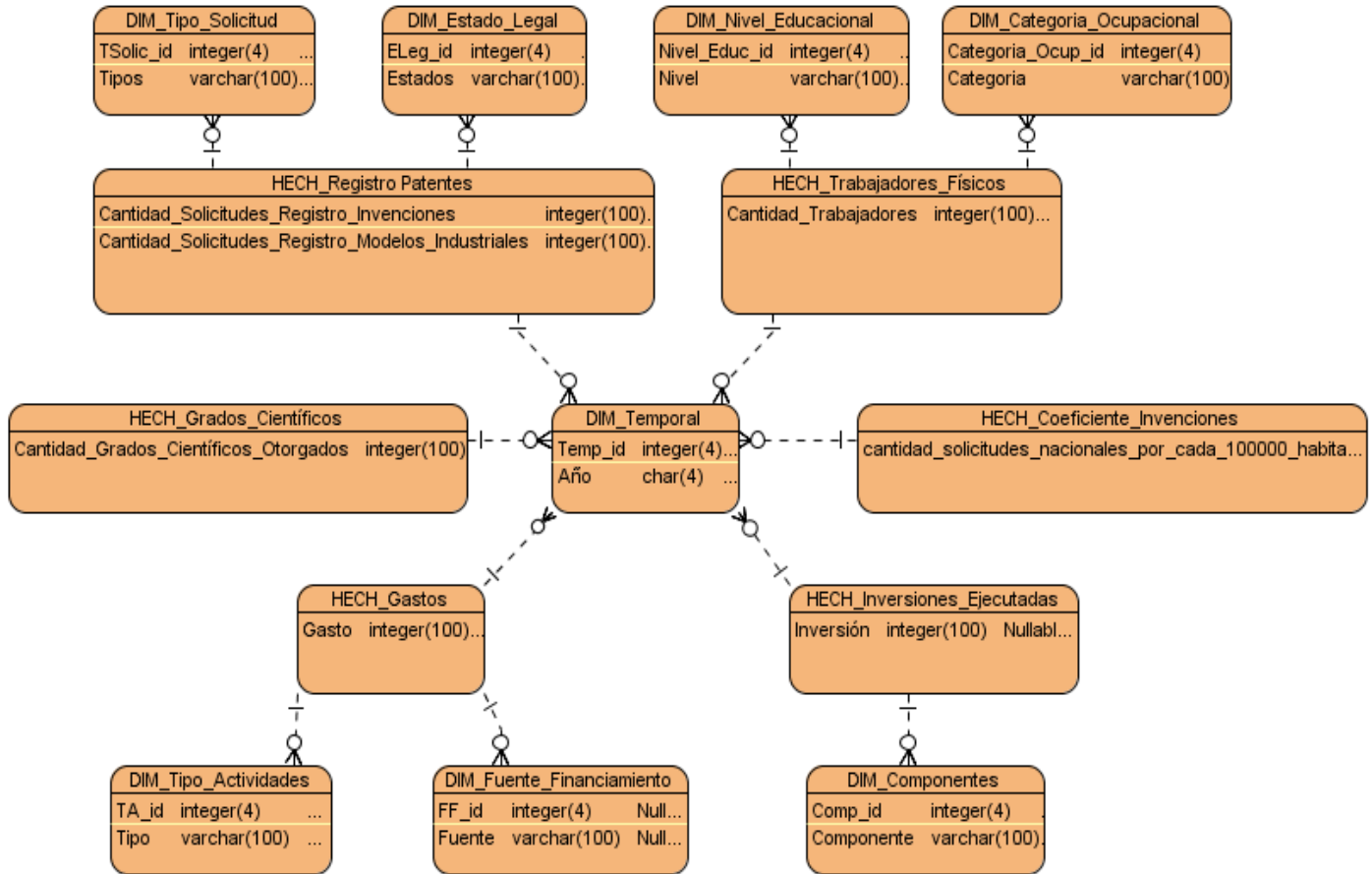


Ilustración 8 Modelo de Datos Físico

2.11.1 Dimensiones

Las dimensiones son las características de un hecho, que permiten su posterior análisis en el proceso de toma de decisiones; una entidad de negocio respecto a la cual se deben calcular las métricas. Las dimensiones definidas son las siguientes:

- Dimensión DIM_Temporal
- Dimensión DIM_Nivel_Educacional
- Dimensión DIM_Categoria_Ocupacional
- Dimensión DIM_Tipo_Actividades
- Dimensión DIM_Fuente_Financiamiento
- Dimensión DIM_Componentes

- Dimensión DIM_Estado Legal
- Dimensión DIM_Tipo _Solicitud

2.11.2 Tablas de hechos

Las tablas de hechos contienen las dimensiones y las medidas asociadas al mismo. La tabla de hechos tiene una clave primaria compuesta por la clave primaria de las tablas de dimensiones relacionadas a este. (Díaz, 2009)

Hechos	Dimensiones Asociadas	Medidas
Registro Patentes	-Estado Legal -Tipo Solicitud	-Cantidad solicitudes registro de invenciones. --Cantidad solicitudes registro de modelos industriales.
Gastos	-Tipo Actividades - Fuente Financiamiento	-Gastos
Inversiones Ejecutadas	- Inversiones	-Inversión
Trabajadores Físicos	-Nivel Educativo -Categoría Ocupacional	-Cantidad trabajadores
Grado Científico		-Cantidad grados científicos otorgados
Invenciones		-Cantidad de solicitudes nacionales por cada 100 000 habitantes

Tabla 6 Tabla de Hechos

2.11.3 Medidas

Las medidas son los valores de datos numéricos que se analizan, son las variables de salida en el diseño, de ahí que representen lo contable que se necesita conocer. Las mencionadas a continuación son las definidas:

- Cantidad de Trabajadores.
- Gasto.
- Cantidad de solicitudes de registro de invenciones.
- Cantidad de solicitudes de registro de modelos industriales.
- Inversión
- Cantidad de grados científicos otorgados.
- Coeficiente invenciones.

2.12 Esquema de Seguridad

El esquema de seguridad representa el respaldado por los niveles de acceso, específicamente por los roles definidos. La arquitectura del sistema de seguridad y la alta disponibilidad debe ser 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 configuración y prueba de todos los equipos deberá efectuarse en las instalaciones de la Oficina Nacional de Estadísticas en Ciudad de La Habana, conjuntamente con personal técnico especializado, de acuerdo a las especificaciones de configuración y prueba. 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 Oficina Nacional de Estadísticas, 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 en las actualizaciones de software, que deben ser los recomendados por los fabricantes de los productos entregados.

2.13 Política de respaldo y recuperación.

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

- **Periodicidad de las salvallas:** Mensualmente se realizan las salvallas de toda la información contenida en la BD. La organización lo tiene definido de esta manera, certificando en todo momento la existencia de una copia escrita de la información que está presente en el servidor.
- **Tablas involucradas:** Las tablas que se involucran en la realización son: tabla de hechos coeficiente de invenciones, tabla de hechos gastos, tabla de hechos grados científicos, tabla de hechos inversiones ejecutadas, tabla de hechos registro de patentes y tabla de hechos trabajadores físicos. Además de las tablas de las dimensiones que se realizan una vez.
- **Backups existentes:** Actualmente no existen backups en esta área.

2.14 Conclusiones del Capítulo

El desarrollo de este capítulo prepara las bases para dar paso a la implementación del MD, puesto que se define: el negocio y sus reglas, el tema de análisis, los roles y permisos; así como las necesidades de los usuarios. También los requisitos: de información, funcionales, no funcionales y multidimensionales. Se construyen los casos de uso y la matriz bus, el modelo de datos y se definen: las dimensiones, los hechos y las medidas.



CAPÍTULO 3 IMPLEMENTACIÓN Y PRUEBA

3.1 Introducción del capítulo

En este capítulo se tiene en cuenta la estructura de los datos, tanto los esquemas como las tablas, los índices, restricciones y secuencias. Además se definen los roles y los privilegios. Se cargan los nomencladores, se establece una guía de implantación y se realizan las validaciones y las pruebas.

3.2 Modelos de Datos Físico

El Modelo de Datos es una manera de organizar todo lo relacionado con la BD, es un sintetizador para establecer y mantener todas las condiciones acordes dentro del funcionamiento del sistema y se encuentra ubicado en la siguiente dirección:

https://repositorio.prod.uci.cu/svn/cbd_almacenes/Proyectos/8. Data Marts ONE/11. Data Mart Ciencia, Inn y Tecn

3.2.1 Estructura de Datos

En su totalidad, los datos están estructurados por esquemas que los organizan y además presentan tablas que se encargan de darle especificación a las restricciones, presentando los índices con sus políticas.

3.2.1.1 Esquemas y Tablas

Un esquema es la definición de una estructura. Determina la identidad de la relación y qué tipo de información podrá ser almacenada dentro de ella y las tablas en las bases de datos, se refiere al tipo de modelado de datos, donde se guardan los datos recogidos por un programa, en este caso dimensiones y hechos.

Esquemas	Tablas
Dimensiones	Dim_Categoria_Ocupacional
Dimensiones	Dim_Nivel_Educacional
Dimensiones	Dim_Componentes
Dimensiones	Dim_Estado_Legal
Dimensiones	Dim_Fuente_Financiamiento
Dimensiones	Dim_Temporal
Dimensiones	Dim_Tipo_Actividades
Dimensiones	Dim_Tipo_Solicitud
Hechos	Hech_Coeficiente_Invenciones
Hechos	Hech_Gastos
Hechos	Hech_Grados_Científicos
Hechos	Hech_Inversiones_Ejecutadas
Hechos	Hech_Registro_Patentes
Hechos	Hech_Trabajadores_Físicos

Tabla 7 Esquemas y Tablas

3.2.1.2 Restricciones y Secuencias

Una restricción es una especificación o condición que obliga el cumplimiento de ciertas circunstancias en la BD. Algunas no son determinadas por los usuarios, sino que son definidas por el administrador del sistema. Las restricciones no son parte formal del modelo relacional, pero son incluidas porque juegan el rol de organizar mejor los datos.

Capítulo 3: Implementación y Prueba

Llave primaria: Una clave primaria es una clave única elegida entre todas las candidatas que define unívocamente a todos los demás atributos de la tabla, para especificar los datos que serán relacionados con las demás tablas. La forma de hacer esto es por medio de claves foráneas.

Llave foránea: Una clave foránea es una referencia a una clave en otra tabla. Las claves foráneas no necesitan ser claves únicas en la tabla donde están y sí a donde están referenciadas.

El trabajo cuenta con: 13 llaves foráneas y 8 llaves primarias, la tabla siguiente es la muestra de algunas de estas:

LLAVES PRIMARIAS	Categoria_Ocup_id
	Comp_id
	ELeg_id
	TSolic_id
	FF_id
	Nivel_Educ_id
	Temp_id
	TA_id
LLAVES FORÁNEAS	RefDIM_Temporal691
	RefDIM_Fuente_Financiamiento411
	RefDIM_Tipo_Actividades401
	RefDIM_Componentes571
	RefDIM_Estado_Legal431
	RefDIM_Tipo_Solicitud441
	RefDIM_Categoria_Ocupacional391
	RefDIM_Nivel_Educacional381

Tabla 8 Restricciones y Secuencias

3.2.1.3 Índices

Son una estructura de datos que mejora la velocidad de las operaciones, permitiendo un rápido acceso a los registros de una tabla. Al aumentar drásticamente la velocidad de acceso, se suelen usar sobre aquellos campos sobre los cuales se hagan frecuentes búsquedas. En la siguiente tabla se describen varios de los índices utilizados en la BD.

Índice	Esquema	Tabla
PK26	Dimensión	Dim_Nivel_Educacional
PK28	Dimensión	Dim_Estado_Legal
PK48	Dimensión	Dim_Temporal
PK39	Hechos	Hech_Trabajadores_Físicos
PK42	Hechos	Hech_Gastos
PK43	Hechos	Hech_Registro_Patentes
PK44	Hechos	Hech_Inversiones_Ejecutadas

Tabla 9 índices

3.2.2 Usuarios y Privilegios

En este aspecto se definen los roles y privilegios de cada rol en dependencia de la actividad que vaya a desarrollar en el MD, ya sea: administrar, seleccionar, insertar, cargar y eliminar, depende además del tipo de información que vaya a manejar. Posterior al análisis del negocio se definen dos roles “Administrador y Consultor”.

Capítulo 3: Implementación y Prueba

El nivel de acceso a los pedidos de información es público. Los roles y privilegios se describen a continuación:

Roles:

- **Administrador:** persona autorizada para administrar la información así como introducir y cambiar datos. Se encarga de trabajar directamente con las bases de datos.
- **Consultor:** persona capacitada para consultar los datos, seleccionarlos y trabajar con ellos.

Privilegios:

- **Seleccionar:** Solo puede consultar, es decir, leer la información.
- **Administrar:** Puede consultar y modificar la información. Tiene control total sobre la BD.
- **Insertar:** Introduce datos al sistema.
- **Cargar:** Carga datos al sistema.
- **Eliminar:** Elimina dimensiones, hechos, tablas.

Roles	Privilegios				
	Administrar	Seleccionar	Insertar	Cargar	Eliminar
Administrador	x	x	x	x	x
Consultor		x			

Tabla 10 Roles y Privilegios

3.2.3 Carga de nomencladores

La carga de los nomencladores es el proceso en donde se carga en las BD las tablas mencionadas a continuación desde los documentos Excel definidos:

- 1) El nomenclador temporal almacena el año de la información.
- 2) El nomenclador Nivel Educativo almacena los niveles: superior, medio y otros niveles.
- 3) El nomenclador Categoría Ocupacional almacena las categorías establecidas: dirigentes, técnicos, administrativos, obreros y de servicio.
- 4) El nomenclador Tipo Actividades almacena los tipos establecidos: investigación y desarrollo, conjuntamente con otras actividades científicas y tecnológicas.
- 5) El nomenclador Fuente Financiamiento almacena las fuentes establecidas: presupuesto del estado, financiamiento empresarial y otros financiamientos.
- 6) El nomenclador Componentes almacena los componentes establecidos: construcción y montaje, equipos y otros.
- 7) El nomenclador Tipo Solicitud almacena los tipos establecidos de solicitud: solicitudes nacionales y solicitudes extranjeras.
- 8) El nomenclador Estado Legal almacena los estados establecidos: registros concedidos, solicitudes denegadas y otros estados.

3.3 Guía de Implantación

Para darle cumplimiento a la solución del trabajo es necesario establecer una guía de implantación para lograr un equilibrio tanto entre los detalles técnicos como en los organizativos.

3.3.1 Secuencia de Pasos

Para llevar a cabo la instalación de la BD se tienen en cuenta una serie de pasos, tales como:

- 1) Crear una nueva BD utilizando como SGBD PostgreSQL en su versión 8.4.
- 2) Cargar el script DDL²⁸ Dimensiones.sql que contiene los roles definidos y ejecutarlo.
- 3) Cargar el script denominado DDL Hechos.sql y ejecutarlo.

²⁸ Data Definition Language, DDL por sus siglas en inglés. Estructura de la Base de Datos.

- 4) Cargar el script denominado DML²⁹ Dimensiones.sql y ejecutarlo.
- 5) Cargar el script denominado DCL³⁰.sql y ejecutarlo.

3.3.2 Requerimientos

Software

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

Hardware

- Memoria: 1 GB
- Red: 100.0 Mbps
- Almacenamiento: 80 GB.

3.4 Validación y pruebas

La validación es la comprobación de que la solución está acorde a las necesidades y exigencias de los clientes, mientras que las pruebas se realizan para tener bien definida la calidad de la solución.

3.4.1 Listas de Chequeo del Análisis

Para concretar y propiciar un buen desarrollo en el trabajo, se crean listas de chequeo del análisis quienes propician que los artefactos construidos en el análisis cumplan con los parámetros establecidos y estén dentro del formato correspondiente en cada caso.

Las listas de chequeo que se utilizadas son las siguientes:

- Herramienta para recolección y el análisis de la información.
- Áreas de la Organización.
- Especificación de requisitos.

²⁹ Data Manipulation Language, DML por sus siglas en inglés. Las asistencias INSERT.

³⁰ Data Control Language, DCL por sus siglas en inglés. Lenguaje de control de datos. Lenguaje que controla los usuarios de la base de datos y las estrategias de seguridad.

Después de haber aplicado las listas de chequeo anteriormente mencionadas se concluye con que no existe ninguna no conformidad relacionada con indicadores con peso crítico y la cantidad de elementos afectados de un indicador que no sea crítico no es mayor que dos, demostrándose la calidad de la solución.

3.4.2 Lista de Chequeo del Diseño.

La lista de chequeo del diseño es la encargada de procurar que el artefacto construido en el diseño verifique que todas las condiciones estén en orden y tenga la estructura adecuada. Para cumplimentar lo antes planteado se cuenta con las lista de chequeo:

- Modelo de Datos.

Posterior a la aplicación de esta lista de chequeo fue verificado que no existe ninguna no conformidad, manifestándose así que la solución está a plenitud y con toda la calidad requerida.

3.4.3 Validación de requisitos por el cliente

Para aprobar la propuesta de solución se realizó una reunión en donde asistió la cliente: Elena Fernández García, quedando conforme con la propuesta mostrada y satisfecha con el análisis y diseño definido. Dando paso a la realización de las pruebas.

3.4.4 Pruebas de Implantación

<u>Casos de Prueba</u>	<u>Pre condición</u>	<u>Pos condición</u>	<u>Resultado esperado</u>
Crear BD	PostgreSQL 8.4 instalado como gestor de base de datos.	Base de datos creada.	Base de datos creada sobre el gestor de base de datos PostgreSQL 8.4.
Roles y permisos	BD creada e instalada la herramienta de administración, PgAdmin III.	Base de datos creada con sus roles y permisos.	Roles creados con los respectivos permisos.

Cargar Nomencladores	BD creada con sus roles y permisos.	BD lista para ser utilizada.	Campos de la BD llenos.
----------------------	-------------------------------------	------------------------------	-------------------------

Tabla 11 Pruebas de Implantación

3.5 Conclusiones del Capítulo

En este capítulo mediante la estructura de esquemas y tablas definida se brinda una correcta organización de los datos. Con la creación de los usuarios y sus privilegios se cumple con lo establecido en los requisitos iniciales del sistema. Al definir los pasos para la implantación de la BD se logra la correcta creación de la misma. Los nomencladores cargados responden a las especificaciones realizadas por el cliente. Las pruebas realizadas verifican el correcto funcionamiento de los pasos de implantación del sistema y luego de la aplicación de las listas de chequeo en donde no existe ninguna no conformidad se demuestra que la solución está a plenitud y con toda la calidad requerida.

CONCLUSIONES GENERALES

A partir de los objetivos planteados en esta investigación donde se realiza la creación de un MD para la ONE relacionado con los indicadores sobre ciencia e innovación tecnológica se arribó a los resultados esperados:

- La elaboración del marco teórico desarrollado por una vasta investigación propició un total entendimiento de las tendencias de la implementación de los AD y los MD.
- El análisis de los todos los documentos Excel relacionados con la ciencia e innovación tecnológica permitió que se identificaran eficientemente todos los requisitos funcionales, informativos y multidimensionales.
- El diseño del MD para los indicadores sobre ciencia e innovación tecnológica se desarrolló satisfactoriamente obteniéndose un modelamiento que soporta los requisitos funcionales definidos para el almacén de datos de la ONE.
- La implementación y carga de los clasificadores para el MD de los indicadores sobre ciencia e innovación tecnológica para el AD de la ONE produjo una integración más completa de la información y una mejor disponibilidad de la misma.
- La validación de la solución desarrollada mediante la realización de pruebas demuestra que el sistema cumple con las expectativas del cliente.

Por todo lo anteriormente expuesto se concluye que para el presente trabajo se han cumplido satisfactoriamente los objetivos trazados.

RECOMENDACIONES

Se recomienda concluir la implementación del MD indicadores relacionados con la ciencia e innovación tecnológica para la ONE e integrar la solución desarrollada al AD de la ONE.

REFERENCIAS BIBLIOGRÁFICAS

Cabrera, María Evelia Casales. 2009. [Online] enero 2009. [Cited: marzo 27, 2010.] <http://hp.fciencias.unam.mx/~alg/bd/dwh.pdf>

Capetillo, Cristian Martínez. 2010. idalicius. [En línea] 2 de febrero de 2010. [Citado el: 2 de marzo de 2010.] <http://www.idalicius.net/2010/02/sistema-gestion-de-bases-de-datos-sgbd.html>

Catrin, Franco. 2009. fayerwayer. [Online] julio 1, 2009. [Cited: marzo 24, 2010.] <http://www.fayerwayer.com/2009/07/postgresql-84-mantiene-distancia-con-sus-rivales/>.

Díaz, Mg. Samuel Oporto. 2009. www.wiphala.net. [En línea] 22 de agosto de 2009. [Citado el: 4 de marzo de 2010.] www.wiphala.net/courses/dwh/ICSI243/2009-II/class/class_11_bi.ppt

Editorial McGraw-Hill . 2008. mailxmail.com. [En línea] 28 de octubre de 2008. [Citado el: 15 de febrero de 2010] www.mailxmail.com/curso-sistemas-bases-datos/sgbd-sistemas-gestores-bases-datos-introduccion .

Frade, Diego Orlando Abril. 2007. scielo. [En línea] abril de 2007. [Citado el: 23 de febrero de 2010.] http://www.scielo.org.co/scielo.php?pid=S0120-56092007000100008&script=sci_arttext.

Inmon, W. H. 1992. *“Using the Data Warehouse”* . 1992.

Kimball, Ralph. 1998. *“The data warehouse toolkit: the complete guide to dimensional modeling”*. 1998.

Perdomo, Mariluz Hernández. 2009. *Guía para la optimización de servidores de bases de datos de PostgreSQL*. Habana : s.n., 2009.

Quiroz, Javier. 2003. ENEGI. [En línea] 2003. [Citado el: 21 de febrero de 2010.] <http://www.inegi.org.mx/inegi/contenidos/espanol/prensa/Contenidos/Articulos/tecnologia/relacional.pdf>.

Rodríguez, Alexei Domínguez. 2009 . *Diseño e implementación de la Base de Datos del Sistema de planificación del Entrenamiento deportivo de Judo Femenino*. Habana : s.n., 2009 .

Rodríguez., Raúl Mató. 2009. *Diseño de Bases de Datos para la Empresa de Gas Manufacturado*. Habana : s.n., 2009.

Sánchez, Leopoldo Zenaido Zepeda. 2008. *Metodología para el Diseño Conceptual de*. Valencia : s.n., 2008.

Sierra, Julio Ernesto Ortiz. 2009. *Diseño e Implementación de un Mercado de Datos para la Oficina Nacional de Estadísticas.* Habana : s.n., 2009.

sinexus. 2007. Sinexus. [En línea] 2007. [Citado el: 5 de febrero de 2010.] http://www.sinnexus.com/business_intelligence/datamart.aspx.

Sinexus. 2007. Sinexus. [En línea] 2007. [Citado el: 13 de febrero de 2010.] http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx.

Siskle. 2009. kle. [En línea] 19 de julio de 2009. [Citado el: 15 de febrero de 2010.] <http://www.siskle.com/spanish/articulo03.html>.

Tamayo, Marysol. 2006. scielo. [En línea] diciembre de 2006. [Citado el: 20 de enero de 2010.] http://www.scielo.org.co/scielo.php?pid=S0120-56092006000300016&script=sci_arttext

Valmaseda, Marcos Ortíz. 2009. *O2PMigration:Herramienta para la migración de bases de datos Oracle a PostgreSQL.* Habana : s.n., 2009.

ANEXOS

Anexo 1 Opinión del usuario del Trabajo de Diploma

	<p style="text-align: right; margin: 0;"><i>Vicerrectoría de Formación Dirección de Formación del Profesional Departamento de Práctica Profesional</i></p>
<p style="text-align: center; margin: 0;">OPINIÓN DEL USUARIO DEL TRABAJO DE DIPLOMA</p>	
<p>El Trabajo de Diploma, titulado "Análisis, diseño e implementación del mercado de datos para los indicadores sobre Ciencia e Innovación Tecnológica para la Oficina Nacional de Estadísticas", fue realizado en la Universidad de las Ciencias Informáticas para la Oficina Nacional de Estadísticas. Esta entidad considera que, en correspondencia con los objetivos trazados, el trabajo realizado le satisface</p>	
<p> <input checked="" type="checkbox"/> Totalmente <input type="checkbox"/> Parcialmente en un ____ % </p>	
<p>Los resultados de este Trabajo de Diploma le reportan a esta entidad los beneficios siguientes (cuantificar):</p>	
<p>Los directivos pueden dedicar más tiempo al análisis de los datos en vez de a su búsqueda debido a que se aceleran las consultas al reducir la cantidad de datos a recorrer. La acumulación de datos les permitirá detectar tendencias y hacer previsiones.</p>	
<p>Y para que así conste, se firma la presente a los <u>3</u> días del mes de <u>Junio</u> del año <u>2010</u>.</p>	
<p>_____ Elena L. Fernández García Representante de la entidad</p>	<p>_____ Coordinadora del proyecto Cargo</p>
<p>_____  Firma</p>	<p>_____  Cuiño</p>

Ilustración 9 Opinión del Usuario

Anexo 2 Diagrama de Casos de Uso Funcionales.



Ilustración 10 Diagrama de Casos de Uso Funcional

Anexo 3 Diagrama de Casos de Uso Informativo

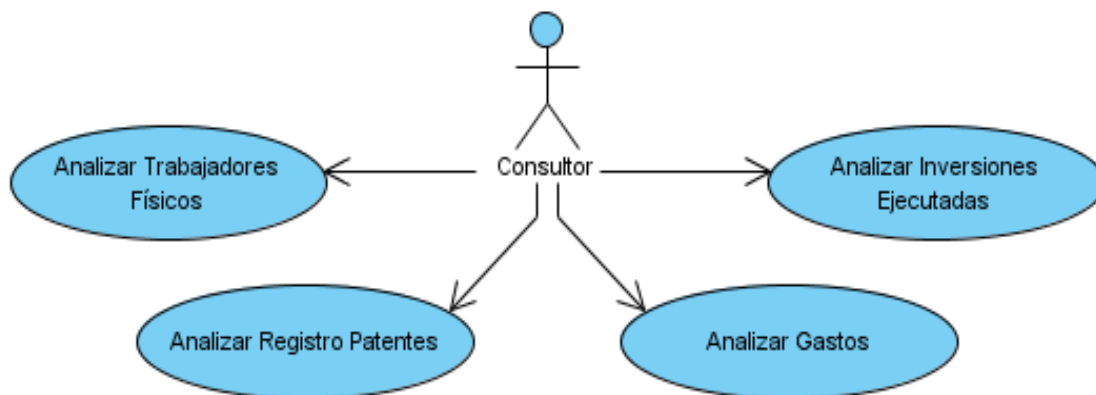


Ilustración 11 Diagrama de Casos de Uso Informativo

Anexo 3 Listas de Chequeo del Análisis.

Lista de Chequeo Evaluación de 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 alcance de las áreas de la organización?	0			
Crítico	2. ¿Se definió el objetivo principal de las áreas de la organización?	0			
	3. ¿Se definieron las características de la organización?	0			
	4. ¿Se utilizó un esquema para representar la información	0			

	obtenida en la evaluación del área?				
Semántica del documento					
Peso	Indicadores Evaluar	a	Eval	(NP)	Cantidad de Comentarios elementos afectados
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		0		

	total de páginas que tiene el documento?				
--	--	--	--	--	--

Tabla 12 Lista de Chequeo Evaluación de áreas de la organización.

Lista de Chequeo Especificación de Requisitos

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

				elementos afectados	
1.	¿Ha identificado a la (s) persona(s) que lo ayudará a especificar los requisitos?	0			
2.	¿Ha solicitado la participación de diferentes personas para poder definir los requisitos en diferentes puntos de vista?	0			
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	0			

	requisito con menos detalles?				
	7. ¿Ha definido el flujo de información de manera adecuada para el problema de dominio? (entrevistas, encuestas, equipos de discusión, grupos de trabajo).	0			
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	0			

	documento de especificación?				
	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	0			

	software y de 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 sido controlados?	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 documento?	0			

Tabla 13 Lista de Chequeo Especificación de Requisitos

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 identificado?	0			
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			
Critico	¿Se ha determinado la periodicidad con que se recoge la información de cada pedido?	0			
Critico	¿Han sido clasificados los pedidos de información de acuerdo a la información que manejan?	0			
Critico	¿Los pedidos de información se han agrupado en casos de uso informativos, según un criterio que los describa?	0			
Critico	¿Han sido determinados los criterios para agrupar los pedidos de información en casos de uso del sistema?	0			
Critico	¿Se han agrupado los pedidos de información en casos de uso del sistema?	0			
Critico	¿Todas las variables de entrada han sido identificadas?	0			
Critico	¿Se han identificado jerarquías de las variables de entrada que las poseen?	0			

Critico	¿Se han determinado los niveles de cada jerarquía?	0			
Critico	¿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			
Critico	¿Todas las variables de salida han sido identificadas?	0			
Critico	¿Se ha definido la unidad de medida en que se va a medir cada variable de salida?	0			
Critico	¿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 el negocio no presenta.
Critico	¿Se ha determinado la fórmula para calcular todas aquellas variables calculables que fueron identificadas?				No se aplica pues el negocio no presenta.
Critico	¿Ha sido identificado el tipo de aditividad que presenta cada variable de salida?	0			
Critico	¿Se identificaron los tipos de operaciones aditivas que se pueden realizar?	0			
Critico	¿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
Critico	¿Se han identificado errores ortográficos en el Excel?	0			
Critico	¿Se entiende claramente todo lo especificado en el Excel?	0			

Tabla 14 Lista de Chequeo de la Herramienta para la recolección y análisis de la información.

Anexo 4 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 de elementos afectados	Comentarios

Critico	¿Las necesidades del cliente han sido identificadas?	0			
	¿Han sido identificados los hechos y dimensiones?	0			
Critico	¿Se ha confeccionado la Matriz Bus con todas las relaciones existentes entre hechos y dimensiones?	0			
Critico	¿Han sido especificadas las dimensiones?	0			
Critico	¿Se ha definido el granulo?	0			
Critico	¿Han sido especificados los hechos?	0			
Critico	¿Se ha ubicado físicamente los archivos?	0			
Critico	¿Se ha diseñado la estructura física de los datos?	0			
Critico	¿Se ha identificado el plan de indexado?	0			
Critico	¿Se ha analizado los resultados de estimación por tablas?	0			

Semántica del documento

Peso	Indicadores a Evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
------	-----------------------	------	------	---------------------------------	-------------

Critico	¿Se han identificado errores ortográficos en el Excel?	0			
Critico	¿Se entiende claramente todo lo especificado en el Excel?	0			

Tabla 15 Lista de Chequeo Modelo de Datos.