



Título:

Mercado de datos Indicadores relacionados con el Comercio Agropecuario para la Oficina Nacional de Estadísticas.

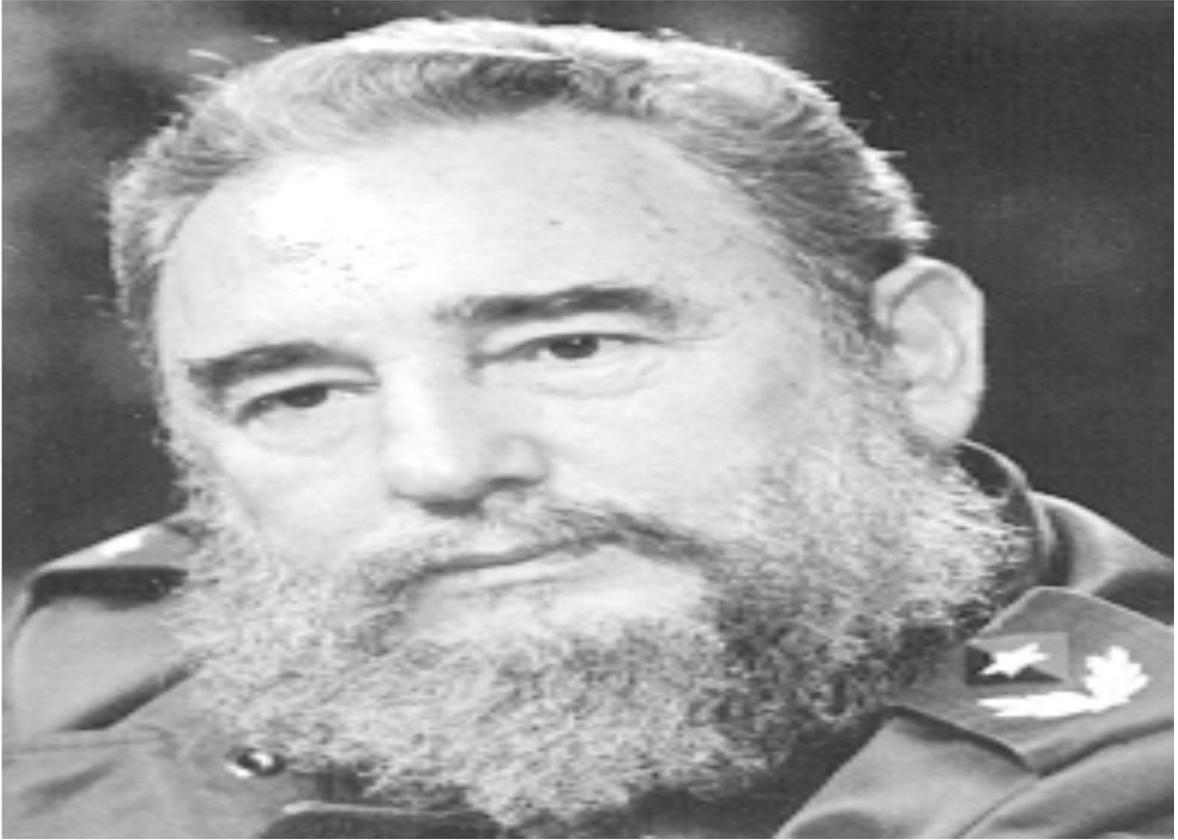
Trabajo de Diploma para optar por el título de Ingeniero en Ciencias Informáticas

Autor: Marianela Rodríguez Monges.

Tutor: Ing. Hamlet Martínez Espino.

Ciudad de La Habana, junio, 2010

Pensamiento



"... Siempre que existan tantas cabezas ardientes, llenas de fe y esperanza, tratando de hacer posible lo imposible, pensemos que todo no estará perdido..."

Fidel Castro Ruz

Agradecimientos

A mi madre, por ser esa persona tan especial en mi vida, por apoyarme en todo y por darme siempre ese voto de confianza y creer en mí.

A mi padre, por ser el inspirador número uno de lo que yo llamo la combinación perfecta.

A mi abuela Ñaño, porque desde el primer día que entré a la UCI me dijo que llegaría muy lejos, y aquí estoy.

A la gorda de Celsa, por quererme demasiado y mimarme con una niña.

A mi queridísima tía María, por pasarse 5 años atada a mis problemas sin protestar.

A mis primos, por ayudarme cuando pensé que todo estaba perdido.

A mi familia, por estar siempre conmigo en las buenas y las malas, por darme la fuerza necesaria para poseer las ganas de levantarme de cada caída y seguir.

A mis amigos, por aguantarme todos los días y hacer de este trabajo un logro de todos.

A mi tutor, porque aún sin conocer nada, aprendimos juntos que cuando se ponen ganas se puede salir adelante.

A mi oponente Mabel, porque de una manera u otra me regaló parte de su tiempo.

A mi tribunal, por darme consejos que hoy me hacen ver el camino más claro.

A todos los profesores que de una forma u otra contribuyeron a mi educación e instrucción.

A la Revolución, por todas las oportunidades.

Dedicatoria

A mi linda mamá, por ser esa estrella que alumbra mis noches cuando todo está triste y oscuro. Te quiero mi reina.

A mi papito el Gordo, por darme la oportunidad de demostrarle que todo se logra con voluntad humana.

A mi abuela Nãño, porque aunque ya no está conmigo, la siento tan cerca, como cuando me abrazaba de noche y me dormía. Para ti son todas mis batallas ganadas.

A mi tía María, porque me soporta y me quiere por encima de todo como su hija.

A mi abuela Celsa, por reír y llorar conmigo en las buenas y las malas.

A mis primos, por estar siempre a mi lado cuando los necesite.

A mi familia, por darme las fuerzas que buscaba cuando perdía las mías.

A mis amigos, esas personas que son el ingrediente perfecto cuando estás perdido y solo.

A mis profesores, por dedicar parte de su tiempo para hacer de mí una mejor persona.

Declaración de Autoría

Declaramos que somos los únicos autores del trabajo “Mercado de datos Indicadores relacionados con el Comercio Agropecuario para la Oficina Nacional de Estadísticas” y autorizamos a la Universidad de las Ciencias Informáticas a hacer uso del mismo en su beneficio.

Para que así conste firmamos la presente a los ____ días del mes de _____ del año 2010.

Autores:

Marianela Rodríguez Monges

Tutor:

Ing. Hamlet Martínez Espino

Datos de contacto

Síntesis del Tutor: Hamlet Martínez Espino.

Categoría Docente: Trabajador Docente.

Formación académica: Graduado en el 2009 de Ingeniero en Ciencias Informáticas de la Universidad de Ciencias Informáticas (UCI).

Centro laboral: Actualmente trabaja en el Departamento de Sistemas Digitales.

E-mail: hespino@uci.cu

Resumen

El presente trabajo surge a partir de la necesidad de la Oficina Nacional de Estadísticas (ONE) de centralizar la información mediante el mercado de datos para los Indicadores relacionados con el Mercado Agropecuario de la Actividad de Comercio que facilite la disponibilidad de dicha información. Esta oficina es el órgano rector de la estadística en Cuba, tiene como objetivo fundamental captar, analizar y difundir los datos recogidos a lo largo y ancho de todo el país. Con fin de optimizar la velocidad de respuesta del sistema a las consultas que se realicen sobre la información se tiene como objetivo crear un almacén de datos y dentro de este, un Mercado de datos para los indicadores del Comercio, fundamentalmente haciendo énfasis en el modelo 0335. Este modelo almacena las ventas de productos en el mercado agropecuario de una selección de ellos. Por tanto, todo el análisis, diseño y la posterior implementación de dicho mercado de datos constituyen la esencia de este trabajo. Definiéndose como metodología a usar, una adaptación de la popular metodología de Kimball, y las herramientas PostgreSQL 8.4.1 como sistema de gestor de bases de datos, Visual Paradigm 6.4 como herramienta de modelado, y como herramienta de administración PgAdmin 3.0. Lográndose al final una aplicación que permitirá a la ONE contar con una información estructurada y disponible para la toma de decisiones.

Palabras claves: Mercado de datos, Almacén de datos, Base de Datos (BD), Indicadores, Comercio, Agropecuario.

ÍNDICE

Capítulo 1: Fundamentación teórica.....	15
1.1. Modelo propuesto.	15
1.1.1. Indicadores.....	15
1.2. Aplicaciones existentes.	22
1.2.1. Softwares existentes a nivel Mundial.....	22
1.2.2. Sistemas existentes en Cuba.	23
1.3. Almacén de Datos.	23
1.3.1. Características del almacén de datos.....	25
1.3.2. Beneficios e inconvenientes del uso de los almacenes de datos.	25
1.3.3. Mercados de Datos.....	26
1.3.4. Metadatos.	27
1.3.5 Base de Datos.....	27
1.4 Modelos de Almacenamiento de Datos.	28
1.5 Modos de Almacenamiento OLAP.....	30
1.5.1 Tipos de OLAP.....	30
1.5.2 Comparación entre MOLAP y ROLAP.....	33
1.6. Fundamentación de la metodología a utilizar.....	34
1.6.1 Metodología utilizada en el centro.	36
1.7 Gestor de Bases de Datos	37
1.7.1 ¿Por qué PostgreSQL 8.4?	40
1.8. Herramienta de Administración.....	41
1.9 Herramientas de modelado.....	41
1.9.1 ¿Por que Visual Paradigm?	44
Capitulo 2: Análisis y Diseño del Mercado de Datos.	47
2.1. Análisis.	47
2.1.1. Definición del negocio.	48
2.1.2. Temas de Análisis.	49
2.1.3 Roles y permisos.....	49

2.1.4. Reglas del Negocio.	50
2.1.5 Necesidades de los usuarios	50
2.2 Requisitos de información y multidimensionales.....	50
2.2.1 Casos de uso del sistema.....	56
2.3 Diseño.	57
2.3.1 Matriz BUS.	57
2.3.2 Modelos de Datos.....	57
2.3.4 Tablas de hechos y medidas asociadas.	59
2.3.5 Esquema de seguridad.	60
2.3.6 Política de respaldo y recuperación.	60
Capítulo 3: Implementación y Validación de la solución.....	63
3.1. Modelo de Datos Físico.....	63
3.1.1. Estructuras de Datos.	64
3.1.2. Usuarios y privilegios	68
3.1.3. Carga de nomencladores.....	69
3.2. Guía de implantación	69
3.2.1. Requerimientos	69
3.2.2. Secuencia de pasos para la instalación de la BD	70
3.3. Validación y pruebas.....	71
3.3.1. Lista de chequeo de análisis	71
3.3.2. Lista de chequeo de diseño.....	71
3.3.3. Validación de requisitos por el cliente.....	72
3.3.4. Pruebas de implantación	72
Conclusiones Generales.....	75
Recomendaciones.....	76
Referencias.....	77
Bibliografías	78
Glosario de términos.....	80

Introducción.

Actualmente en el Mundo de hoy, constituyen parte indispensable en la vida del ser humano las tecnologías de la información. Teniendo en cuenta la enorme competencia que existe en el mercado internacional entre las compañías dedicadas al desarrollo de dichas tecnologías, ha surgido la necesidad de adaptarse constantemente a los cambios que ocurren en este campo. Las soluciones informáticas que se obtienen mediante estas tecnologías traen consigo mejoras para el ambiente en el cual se aplican, logrando así el ahorro de tiempo, recursos y aportando un sin número de ganancias.

Cuba, aún encontrándose en vías de desarrollo, no está ajeno a esta situación y se ha dado la ardua tarea de informatizarse. Para lograr este propósito, en el año 1997, se definieron los “Lineamientos estratégicos para la informatización de la sociedad cubana” por el Comité Ejecutivo del Consejo de Ministros de Cuba y dos años después, comienza un amplio programa de acciones. En víspera de cumplir dichos objetivos, se empieza un proceso continuo de transformación y desarrollo de la gestión de la información y el conocimiento para lograr la excelencia de las tareas, abarcando una serie de cuestiones que van desde los sistemas de información hasta la capacitación del capital humano, la seguridad informática y las redes.

El Ministerio de la Informática y las Comunicaciones fue el principal encargado de introducir al país dentro del mundo de la tecnología informática. Se llevó a cabo la digitalización de la red telefónica por parte de la Empresa de Telecomunicaciones de Cuba SA (ETECSA), permitiendo un incremento en la cantidad de usuarios de ese servicio.

Como parte de estos cambios las organizaciones del país se dieron a la tarea de actualizar, tanto la tecnología existente en sus centros, como la manera de recoger y gestionar la información. No es hasta 1976 que se crea el Comité Estatal de Estadística, el cual contaba, para el procesamiento de la información, con una red de procesamiento de datos a nivel nacional. Luego el 21 de abril de 1994 se crea la ONE, la cual procura responder a las nuevas exigencias de los usuarios, utilizando los avances tecnológicos que posibilitan una difusión más rápida de la información. La misma tiene como misión fundamental garantizar la producción de estadísticas de calidad a través del Sistema Estadístico Nacional ejerciendo una adecuada dirección, ejecución y control de la captación de las cifras económicas y sociales, así como su adecuada difusión de acuerdo con las necesidades de la economía y las demás necesidades del país en cuanto a información estadística. La visión de la ONE es

construir un sistema estadístico profesional capaz de responder con calidad y oportunidad a las necesidades de información estadística del país para cumplir las metas del desarrollo económico y social del mismo y su adecuado reflejo internacional, conscientes de ser útiles y necesarios a la sociedad.

Desde que se inicia la ONE, almacenaba toda la información relacionada con sus temas en cintas magnéticas. Más tarde cuando ocurre el derrumbe del campo socialista, se ve en la obligación de actualizar la tecnología, por lo que toda la información que tenía almacenada anteriormente se perdió. Luego de tratar varias veces de obtener una mejor gestión de las estadísticas se adopta un sistema basado en ficheros en el año 2000, dicho fichero va a incluir el Sistema de Gestión de Estadísticas económicas MICROSET y el Sistema de Gestión de Estadísticas de población SIDEMO/DV SURVEY. Estos sistemas de gestión no permiten realizar consultas a la información de manera eficaz ya que solo brindan la posibilidad de hacer reportes estáticos.

La ONE es el órgano rector de la estadística en Cuba, tiene como objetivo fundamental captar, analizar y difundir los datos recogidos a lo largo y ancho de todo el país. Para ello cuenta con una serie de modelos estadísticos en los cuales recoge información de todos los sectores de la economía y la sociedad. Cuenta con el modelo 0335 que será el modelo a analizar en este trabajo de diploma, el cual va a almacenar las ventas de productos en el mercado agropecuario de una selección de ellos. Las tecnologías y las herramientas utilizadas para almacenar, recuperar y presentar la información proveniente del MODELO 0335 impiden realizar los principales reportes y cruces de variables, indicadores, tasas, porcentajes y demás aspectos de interés; dificultando así la disponibilidad de información para Órganos del Estado y afectando el proceso de toma de decisiones.

Por tanto, después de analizar lo expuesto anteriormente surge el siguiente **problema a resolver**: ¿Cómo mejorar la tecnología de almacenamiento de la ONE para el Comercio Agropecuario que facilite un mejor análisis para la toma de decisiones?

Se definen como **objeto de estudio** los almacenes de datos y como **campo de acción** los mercados de datos para los Indicadores relacionados con el Comercio Agropecuario.

El **objetivo general** de esta investigación es desarrollar el análisis, diseño e implementación del mercado de datos Indicadores relacionados con Comercio Agropecuario para la ONE.

Como **objetivos específicos** se tienen:

- ✓ Elaborar el marco teórico de la Investigación acerca de las principales tendencias de implementación de los almacenes de datos, los mercados de datos y estudio del arte del tema Indicadores relacionados con el Mercado Agropecuario de la Actividad de Comercio
- ✓ Realizar el análisis y diseño del Modelo 0335 para el mercado de datos Indicadores relacionados con el Mercado Agropecuario de la Actividad de Comercio.
- ✓ Implementar el mercado de datos Indicadores relacionados con el Mercado Agropecuario de la Actividad de Comercio para el almacén de datos de la ONE.
- ✓ Validar la solución desarrollada mediante la realización de pruebas.

Para dar cumplimiento a los objetivos de esta investigación se plantean las siguientes **tareas de investigación:**

1. Estudiar los temas relacionados con el desarrollo de Mercados de Datos y la metodología a utilizar en el desarrollo.
2. Planificación y realización de entrevistas.
3. Identificación de la estructura de usuarios y permisos.
4. Identificación de las necesidades de información, requisitos funcionales y no funcionales.
5. Definición de requisitos de entrada y de salida.
6. Elección de la granularidad del proceso del negocio.
7. Definición de las dimensiones del Mercado de Datos y de los hechos asociados a las dimensiones definidas.
8. Estructuración del modelo dimensional y transformación del modelo dimensional al diseño físico.
9. Montaje de los clasificadores para mercado de datos Indicadores relacionados con el Mercado Agropecuario de la Actividad de Comercio para el almacén de datos de la ONE.
10. Realización de pruebas al Mercado de Datos.

A partir del conocimiento de todo esto se puede plantear la **idea a defender**: Si se desarrolla un mercado de datos para los Indicadores relacionados con el Mercado Agropecuario de la Actividad de Comercio se podrá contar con una información estructurada y disponible para la toma de decisiones.

Métodos Científicos.

Siendo parte de la investigación necesaria para la realización de cualquier trabajo se recomienda el uso de métodos científicos que servirán como una guía a seguir para realizar dicha investigación. Estos métodos no serán más que procedimientos que van a estudiar la sociedad, la realidad, la naturaleza y el pensamiento, para poder descubrir su esencia y sus relaciones.

Teóricos:

Histórico-Lógico: será utilizado en la investigación para:

- ✓ El estudio de los temas relacionados con el desarrollo del mercado de datos de los Indicadores relacionados con el Mercado Agropecuario de la Actividad de Comercio así como su evolución en la ONE.
- ✓ El estudio de la metodología a utilizar en el desarrollo.

Analítico - Sintético: Este método se utilizará durante la investigación para:

- ✓ Análisis de documentos, materiales, y temas relacionados a las mejores prácticas en el desarrollo de mercados de datos.
- ✓ Definir los temas de análisis.
- ✓ Definir los requisitos de entrada y de salida.

Modelación: se utilizará para:

- ✓ Dar la medida en que se logra la relación entre el modelo y el objeto, para ello se define la arquitectura del sistema, se definen las dimensiones del Mercado de Datos, los hechos asociados a las dimensiones definidas, se estructura el modelo dimensional y se transforma del modelo dimensional al diseño físico.

Empíricos:

Entrevista: este método se utiliza para:

- ✓ Obtener la información acerca de las necesidades de información, requisitos funcionales y no funcionales.

Este trabajo estará compuesto por **3 capítulos**:

Capítulo 1: Fundamentación Teórica.

En este capítulo se realiza un estudio del estado del arte sobre los Sistemas de Almacenes de Datos y los Mercados de Datos, sus principales metas, características y los elementos fundamentales que los componen. Se realiza un estudio de las fuentes de datos, las metodologías existentes y las principales herramientas para el desarrollo de los Almacenes de Datos a nivel nacional y mundial.

Capítulo 2: Análisis y Diseño de la solución.

Se abordará aspectos concernientes a la descripción e implementación de la solución. Se definen los temas de análisis, el ciclo completo del proceso del negocio, se identifica la estructura de usuarios y permisos así como las necesidades de información, requisitos funcionales y no funcionales, se modelan y validan los requerimientos, el modelo dimensional propuesto, tablas de hechos y además se definen los requisitos de entrada y de salida, así como la política de respaldo a utilizar en el sistema.

Capítulo 3: Implementación y Validación de la solución.

Contiene todos los elementos referentes a la implementación del sistema, las dimensiones y hechos que formarán parte de la posible solución, la descripción de los roles y permisos creados y los pasos para la implantación del sistema. Se implementa la BD, se montan los clasificadores y se detallarán las pruebas de rendimiento y las pruebas de integración. Se valida la solución a través del empleo de las listas de chequeo y la carta de aceptación del cliente.



Capítulo 1: Fundamentación teórica

Introducción

En este capítulo se realiza un estudio del estado del arte acerca del mercado de datos Indicadores relacionados con el Mercado Agropecuario de la Actividad de Comercio y la metodología a utilizar en el desarrollo del mismo. Se abordan algunos elementos teóricos acerca de la tecnología de almacenes de datos y los mercados de datos. Además, se hace referencia a la metodología que se utilizará, y las diferentes herramientas.

1.1. Modelo propuesto.

Existen múltiples modelos que forman el Sistema de Información Estadístico Nacional (SIEN) que son de gran importancia para el desarrollo de la rama en la cual se enmarcan. El presente trabajo tratará uno de los modelos que forman parte de los Clasificados según el Nomenclador de Actividades Económicas (NAE) vigente en el país, específicamente dentro del área de comercio, se presenta a continuación.

Modelo 0335: Ventas en el Mercado Agropecuario

Este modelo tiene como objetivo conocer las ventas, en físico y en valor, de productos agropecuarios que se realizan en el Mercado Agropecuario, con el fin de poder conocer el consumo de productos agropecuarios de la población a través de este sistema de ventas.

1.1.1. Indicadores

Los **indicadores** son conceptos que se van a “medir”. Una vez se han definido los indicadores de un Sistema de Calidad particular, se deben definir valores objetivos para los mismos en un determinado periodo. Esto permitirá, en primer lugar, establecer unos umbrales de Calidad que se deben alcanzar y, en segundo lugar, comprobar hasta qué punto se alcanzan esos niveles.

Para cada uno de los indicadores se deberá establecer la **métrica** o sistema de medida que se va a utilizar. Además de las métricas, se debe determinar **cada cuánto tiempo** se va a analizar el resultado de los indicadores. Por último, se deben establecer mecanismos que permitan determinar qué ha impedido - o favorecido - que se alcanzaran los valores previstos. (2010)

INDICADORES SELECCIONADOS DEL SECTOR AGROPECUARIO

SECCIÓN 1: AGRICULTURA NO CAÑERA

Filas

Aparecen relacionados los cultivos de los cuales se requiere información.

Transformaciones de pastos y forrajes (fila 53)

Es la superficie de pastos naturales que, mediante las labores de cultivo que se le ha realizado, ha mejorado sus condiciones productivas.

Producción de heno (fila 54)

Se anota la cantidad de heno producida durante el periodo que se informa.

Chapea total (fila 55)

Se anota la superficie de pastos a la que se le ha efectuado esta labor cultural durante el período que se informa.

Superficie sembrada (columna 1)

Se reportará el total del área sembrada. Este indicador es acumulado desde el inicio del año calendario (1 de enero) hasta la fecha en que concluye el periodo que se informa, teniendo en cuenta que la misma superficie física puede haber sido sembrada más de una vez en el periodo y se reportará tantas veces como haya sido sembrada dentro del periodo informado.

Superficie existente sembrada total (columna 2)

Se reportará la superficie total existente sembrada en el momento del cierre del periodo que se informa. Este indicador es una fotografía.

Superficie cosechada y en producción (columna 3)

Se informa como área cosechada, la superficie de los cultivos temporales a la cual se le ha recolectado el fruto agrícola o parte aprovechable de la planta. Cuando en los cultivos sea necesario efectuar más de un pase para la recogida de la cosecha, se considera la superficie recolectada, en el momento que haya recibido el último pase.

Es acumulado desde el inicio del año.

Para los cultivos permanentes se define como superficie existente en producción y es cuando el cultivo ya ha alcanzado la edad de fructificación y ha dado frutos aprovechables como mínimo una vez. Cuando en un año determinado no se obtiene producción por falta de agua, a consecuencia de plagas, o cualquier otra causa, se

considera dentro de esta categoría el área afectada. Es una fotografía del momento del cierre de la información.

La recolección, tanto de cultivos temporales como de cultivos permanentes, puede ser manual o mecanizada.

Producción total (columna 4)

Se informan los productos cosechados en su forma natural, antes de toda elaboración ulterior, independientemente del fin a que se destinen. No obstante, los cultivos que a continuación se relacionan deben considerarse como casos particulares y se expresarán en los términos especificados:

Ventas totales (columna 5)

Se anotarán las ventas efectuadas en el período a cualquier tipo de comprador, que fundamentalmente son las empresas de acopio y frutas selectas, empresas industriales, empresas exportadoras, empresas de beneficio y de gastronomía, la población - incluyendo las que se realizan en el mercado agropecuario -, así como las ventas a los trabajadores de la propia empresa, Unidades Básicas de Producción Cooperativa o Cooperativas de Producción Agropecuaria. Es decir, este indicador es la suma de todas las ventas ejecutadas en el período, independientemente de la etapa en que haya ocurrido la cosecha.

Incluye solo las ventas que corresponden a producciones propias, es decir, que siempre se reportarán por el productor.

En tal sentido, las empresas que suelen recibir de otros productores ciertas cantidades de productos para ser beneficiadas o recibir algún tratamiento industrial previo a su venta, no se adjudicarán las ventas que provengan de estas cantidades aunque sean sus comercializadores, pues corresponden a los productores que efectuaron las entregas para su beneficio o procesamiento industrial.

Ventas contratadas (columna 6)

Se anota la parte de las “ventas totales” que se efectúa en correspondencia con el pedido estatal a fin de cumplir o sobre cumplir un contrato, previamente firmado o surgido posteriormente, con cualquier entidad acopiadora autorizada. No se consideran ventas contratadas aquellas que se realizan a las entidades acopiadoras a precios diferenciados o mayores de los oficialmente vigentes y que se destinan al mercado agropecuario; ni las que se efectúan a otras entidades estatales si no responden al pedido estatal.

Las ventas contratadas se reportarán en los mismos términos que las ventas totales.

Ventas en el mercado agropecuario (columna 7)

Como un “de ello” de las ventas totales, en esta columna se consignará la parte de la producción reportada en el período que el propio productor vendió en el mercado

agropecuario o entregó a una entidad acopiadora o representante privado para su comercialización en ese lugar.

Ventas a trabajadores (columna 8)

Constituyen las ventas efectuadas por las propias entidades productoras estatales a sus trabajadores y las ventas efectuadas por las Unidades Básicas de Producción Cooperativa y las Cooperativas de Producción Agropecuaria a sus socios.

Las ventas a trabajadores se reportarán en los mismos términos que las ventas totales.

Autoconsumo (columna 9)

Se anota la parte de la producción que durante el período que se informa se destina al consumo de la propia entidad productora.

El autoconsumo se reportará en los mismos términos que las ventas totales.

SECCIÓN 2 GANADERÍA

Entregas a sacrificio (fila 01)

Se anota la cantidad de cabezas de ganado vendidas con destino al sacrificio para entregarlas a la industria o las que se pudieran sacrificar en la propia unidad productora, tanto para la entrega directa de carne a la población o al turismo o para el consumo interno, durante el período que se informa. Por lo tanto, la unidad que vende animales a otra para que esta última la entregue a la industria, no informa esta operación como entregas a sacrificio.

Entregas a sacrificio (producción de carne) (fila 02)

Se informa el peso en pie expresado en toneladas de los animales que se informan en la fila 01.

Ventas totales de carne (fila 03)

Se anota la parte de la producción de carne que se vende a cualquier tipo de comprador. Incluye las ventas contratadas de acuerdo con el pedido estatal, las ventas al Mercado Agropecuario, las ventas a trabajadores o cooperativistas de la propia entidad y otras ventas no especificadas. Por lo tanto, es la suma de todas las ventas efectuadas durante el período que se informa.

De ellas: contratadas (fila 04)

Se anota la parte de las ventas totales de carne que se efectúa en correspondencia con el pedido estatal a fin de cumplir o sobre cumplir un contrato, previamente firmado o surgido posteriormente, con cualquier entidad acopiadora autorizada.

No se consideran ventas contratadas aquellas que se realizan a las entidades acopiadoras a precios diferenciados o mayores que los oficialmente vigentes y que se destinan al Mercado Agropecuario; ni las que se efectúan a otras entidades estatales si no responden al pedido estatal.

De ellas: turismo (fila 05)

Se anota la parte de las ventas contratadas que durante el periodo que se informa se entrega con destino al turismo.

Ventas al mercado agropecuario (fila 06)

Es una desagregación de las ventas totales de carne (excepto vacuno y búfalo). Corresponde a la parte de la producción de carne que el propio productor vendió en el mercado agropecuario o entregó a una entidad acopiadora o representante privado para su comercialización en ese mercado.

A trabajadores (fila 07)

Se anota la carne que durante el periodo que se informa se vende a los trabajadores o cooperativistas de la propia entidad. Forma parte de las ventas totales.

Autoconsumo (fila 08)

Se anota el peso en pie de los animales que durante el período que se informa se destinan al consumo de las propias entidades productoras.

Recuperación de cueros (fila 09)

Se anota la cantidad de cueros obtenidos de los animales que se sacrifican en la entidad o los recuperados de animales muertos durante el período que se informa.

Ventas de cueros a la industria (fila 10)

Se anota la cantidad de cueros vendidos a la industria durante el período que se informa.

Hembras en ordeño (fila 11)

Se anota en las columnas respectivas el promedio de vacas, cabras y búfalas que se encuentran en ordeño. Este promedio se obtiene sumando las hembras que se ordeñan diariamente desde el primer día del año y dividiendo este resultado entre los días del período que se informa.

Producción de leche (fila 12)

Se anota la cantidad de leche obtenida de las hembras que se ordeñan durante el período que se informa.

Ventas totales de leche (fila 13)

Es la parte de la producción de leche que se vende a cualquier tipo de comprador. Incluye las ventas contratadas de acuerdo con el pedido estatal, las ventas a trabajadores o cooperativistas de la propia entidad y otras ventas no especificadas. Es decir, es la suma de todas las ventas efectuadas.

De ellas: contratadas (fila 14)

Se anota la parte de las ventas totales de leche que se efectúa en correspondencia con el pedido estatal a fin de cumplir o sobre cumplir un contrato previamente firmado o surgido posteriormente, con cualquier entidad acopiadora autorizada.

No se consideran ventas contratadas aquellas que se realizan a las entidades a precios diferenciados o mayores que los oficialmente vigentes y que se destinan al Mercado Agropecuario, ni las que se efectúan a otras entidades estatales si no responden al pedido estatal.

De ellas: a la industria (fila 15)

Se anotan aquellas ventas contratadas que se efectúan a las entidades industriales para su procesamiento o beneficio durante el período que se informa.

De ellas: a la población (fila 16)

Se anotan, de las contratadas, aquellas ventas de leche que se efectúan directamente a la población durante el período que se informa.

De ellas: a organismos (fila 17)

Se anotan, de las contratadas, aquellas ventas de leche que se realizan a cualquier organismo o dependencia estatal no considerada anteriormente, tales como Órganos Locales del Poder Popular, Ministerio de Educación, Ministerio de Salud Pública y otros.

A trabajadores (fila 18)

Se anotan las ventas de leche que se realizan a trabajadores o cooperativistas de la propia entidad productora durante el período que se informa. Forman parte de las ventas totales.

Autoconsumo (fila 19)

Se anota la parte de la producción de leche que durante el período que se informa se destina al consumo de la propia entidad productora.

Consumo animal (fila 20)

Es la parte de la producción de leche que se destina como alimento de los animales, excluye la directamente mamada por los terneros.

Insumo productivo (fila 21)

Se anota la parte de la producción de leche que durante el período que se informa se destina como materia prima para la elaboración de otros productos en la propia entidad productora.

Mermas y pérdidas (fila 22)

Comprende las mermas y pérdidas ocurridas durante el período después de registrada la producción y que ocurren antes de la entrega a cualquiera de sus destinos.

Producción de huevos (fila 23)

Se anota la cantidad de huevos obtenidos, independientemente de su destino, calidad, tamaño o estado (sano, cascado o roto).

Ventas totales de huevos (fila 24)

Se anota la parte de la producción de huevos que se vende a cualquier comprador. Incluye las ventas contratadas de acuerdo con el pedido estatal, las ventas al Mercado Agropecuario, las ventas a trabajadores o cooperativistas de la propia entidad y otras ventas no especificadas. Es decir, es la suma de todas las ventas efectuadas.

De ellas: contratadas (fila 25)

Se anota la parte de las ventas totales de huevos que se efectúa en correspondencia con el pedido estatal a fin de cumplir o sobre cumplir un contrato, previamente firmado o surgido posteriormente, con cualquier entidad acopiadora autorizada.

No se consideran ventas contratadas aquellas que se realizan a las entidades a precios diferenciados o mayores que los oficialmente vigentes y que se destinan al Mercado Agropecuario; ni las que se efectúan a otras entidades estatales si no responden al pedido estatal.

De ellas: turismo (fila 26)

Se anota la parte de las ventas contratadas que durante el periodo que se informa se entrega con destino al turismo.

Mercado agropecuario (fila 27)

Es una desagregación de las ventas totales de huevos y corresponde a la parte de la producción que el propio productor vendió en el mercado agropecuario o entregó a una entidad acopiadora o representante privado para su comercialización en ese mercado.

A trabajadores (fila 28)

Se anotan las ventas de huevos que se realizan a trabajadores o cooperativistas de la propia entidad productora durante el período que se informa. Forman parte de las ventas totales.

Autoconsumo (fila 29)

Se anota la parte de la producción de huevos que durante el período que se informa se destina al consumo de la propia entidad productora.

Existencia final del rebaño (fila 30)

Se anota la cantidad de animales existentes que al final del período informado forman parte del patrimonio de la empresa o entidad.

De ello: reproductoras (fila 31)

Se anota la cantidad de hembras aptas para la reproducción existente al final del período informado.

Nacimientos (fila 32)

Es el comienzo de la vida del animal por la expulsión completa o extracción a la madre de un producto de la concepción, independientemente de la duración de la gestación, según después de tal separación respira o muestra evidencia de vida, como el latido del corazón o un movimiento definitivo de los músculos voluntarios.

Muertes totales (fila 33)

Son aquellos animales en los que desaparece la vida natural o accidentalmente.

De ellas: crías (fila 34)

Es la parte de las muertes que corresponde a los animales que están en fase de cría.

Columnas (1 a la 8)

Aparecen relacionados los tipos de rebaño de los cuales se requiere información.

1.2. Aplicaciones existentes.

A nivel mundial son muchas las aplicaciones que se utilizan para el almacenamiento de datos, como de la misma manera existen aplicaciones dentro de Cuba para distintas Empresas que hacen uso de las tecnologías de almacenamiento, a continuación se realiza una breve caracterización de las más usadas.

1.2.1. Softwares existentes a nivel Mundial.

Existen softwares que son utilizados para llevar a cabo el control estadístico en las empresas. Estos lo realizan de una manera eficiente cumpliendo con los requerimientos de cada uno de los centros en los cuales son usados. En América Latina existen empresas como Telefónica de Argentina, Visa, Arcor, todas de Argentina; en México existen algunas como Walmart, Procter & Gamble, Whirpol, TV Azteca, Baxter, GNP, Warner Lambert y Sabre que también han venido incorporando el uso de los almacenes de datos para la toma de decisiones a nivel gerencial.

Se puede mencionar además a American Stores (Estados Unidos), Canadian Tyre (Canadá), Owens Corning Glass (Estados Unidos), Karsten Ping Golf Clubs que han obtenido grandes avances en este sentido. En Europa, pionero en este campo, existen empresas tales como Carrefour (España) WH Smith Books (Gran Bretaña), BonPreu (España), SAR Group (España), Great Universal (Gran Bretaña), Corte Inglés (Francia), Cortefiel (Francia), Eroski (Francia), Supermercados Casino (Francia), Otto Versand (Alemania), Helene Curtis (España).

El tema estadístico no ha estado exento a estas necesidades, en países como México, específicamente en el INEGI (Instituto Nacional de Estadísticas e Informática), se tiene la información almacenada en almacenes de datos para la toma de decisiones a nivel gubernamental. Mencionar también el Instituto Nacional de Estadística de Venezuela,

que posee un conjunto de Mercados de Datos para realizar el análisis de las operaciones estadísticas más importantes que utilizan (Pacheco, 2008)

1.2.2. Sistemas existentes en Cuba.

Cuba no ha quedado ajena a este desarrollo sobre Almacenes de Datos. Aunque todavía faltan muchos aspectos por mejorar ya se han visto algunos ejemplos que han dado pasos firmes dentro de la rama, destacando sin dudas el más sobresaliente, el DWS comercial de la Corporación CIMEX, la cual se dedica fundamentalmente a la Exportación e Importación de mercancías, formando parte de ella un conjunto de empresas que se encuentran enfocadas en diversos negocios, aquí se puede citar la red de Comercio Minorista y la Dirección de Logística, esta última dedicada al Comercio Mayorista. El mismo centra su atención en la actividad del comercio, principalmente en la gestión de inventario, permitiendo una gestión de compra-venta eficiente, con una finalidad de disminuir los costos, sin afectar al cliente, permitiendo prestaciones eficientes y con la calidad requerida, aumentando las utilidades de las mismas. En el XIII Concurso Nacional de Computación y en la Feria de Informática del 2002 se presentó un Almacén de Datos para CUBACEL desarrollado sobre plataforma Oracle con grandes resultados obtenidos a partir de su implantación. Existen otras entidades como UNION CUPET y COPEXTEL que en la actualidad se encuentran en el proceso de diseño e implementación de sus respectivos almacenes. (Varios, 2009)

Estos son los primeros pasos hacia un mayor desarrollo de los sistemas de almacenes de datos en el país, puesto que esta solución es a la vez eficiente y posible de realizar con los recursos que se tienen. Además, está ganando mucha fuerza en el mercado internacional y Cuba, si aspira a tener un puesto entre las naciones productoras de software en el mundo, debe estudiar todas las áreas de este mercado.

1.3. Almacén de Datos.

A partir de su aparición en la década de los 90's hasta la fecha la tecnología de Warehousing ha venido madurándose y posicionándose como la variante más acertada para la realización de análisis de información histórica.

Surgió con el objetivo de hacer consultable la información que se tiene de una empresa tanto de meses como de años anteriores. Es un almacenamiento simple, completo y consistente de datos obtenidos desde una variedad de fuentes, disponibles para el usuario final de forma tal que puedan entenderlos y utilizarlos. Facilita la integración de sistemas de aplicación no integrados. El Almacén de Datos está orientado a la toma de decisiones. Un buen diseño de la BD favorece el análisis y la

recuperación de datos para obtener una ventaja estratégica y para facilitar la toma de decisiones. (Reeves, 2008)

El Almacén de Datos está preparado para ser explotado mediante herramientas específicas que permiten la extracción de información significativa y patrones de comportamiento que permanecen ocultos en un enorme repositorio de datos. Un Almacén de Datos contiene la información de toda la empresa. Cualquier departamento puede acceder a la información de cualquier otro departamento mediante un único medio, así como obligar a que los mismos términos tengan el mismo significado para todos. En la actualidad, dado el desarrollo de las tecnologías de la información y las comunicaciones, las organizaciones a nivel mundial se han dado a la tarea de digitalizar su información, eliminando así su dependencia del trabajo manual y las grandes cantidades de documentos a procesar en caso de necesidad. Debido a esto la manera de gestionar los datos se vio obligada a cambiar, creándose nuevas dificultades a la hora de obtener la información correcta o en la cantidad de tiempo que requería el obtenerla. Una de las principales tareas en una organización es la toma de decisiones, que puede verse afectada si no se cuenta con una organización eficaz de los datos. A partir de este problema surge el concepto de Almacenes de datos, cuyo objetivo fundamental es el de ayudar a los usuarios a entender mejor el pasado, así como planear el futuro ya sea a corto, mediano o largo plazo. Esta tecnología no es más que la integración de las técnicas de BD con las de análisis de datos. (Varios, 2006-2010)

Momentos históricos importantes en los primeros años del desarrollo de los Almacenes de Datos:

- ✓ 1960s: El Gral. Mills y El Colegio Darmouth, durante un proyecto de investigación, crean los términos de dimensiones y hechos.
- ✓ 1883: Teradata crea un sistema de administración de BD diseñado específicamente para el soporte de decisiones.
- ✓ 1988: Barry Devlin y Paul Murphy publican el artículo “Una arquitectura para un sistema de negocios e información”, donde introducen el término “Business Data Warehouse”.
- ✓ 1991: Bill Inmon publica el libro “Building the Data Warehouse”.
- ✓ 1995: Es fundado el “Data Warehousing Institute”, que promueve el uso de los almacenes de datos.

- ✓ 1996: Ralph Kimball publica el libro "The Data Warehouse Toolkit".
- ✓ 1997: Oracle 8 es liberado con la opción para crear los esquemas de estrella.

1.3.1. Características del almacén de datos

El Almacén de Datos tiene como objetivo agrupar los datos de toda la empresa con el fin de facilitar su análisis, de forma que sean útiles para acceder y analizar información sobre la propia empresa. Agregar que según William H. Inmon un almacén de datos es "...una colección de datos orientados al tema, integrados, no volátiles e historiadados, organizados para el apoyo de un proceso de ayuda a la decisión". (Reeves, 2008)

- ✓ **Integrado**

Integra datos recolectados de diferentes sistemas operacionales o de fuentes externas en una estructura consistente, utilizándose a veces una estructuración con distintos niveles de detalle.

- ✓ **Temático**

La información se clasifica en base a los aspectos que son del interés de la empresa. Así los usuarios tendrán un mayor entendimiento y un mejor acceso a dicha información.

- ✓ **Historiado**

Los datos que se cargan en el almacén de datos no son más que los valores que toman las variables en el tiempo. Por esto es que se pueden realizar análisis de tendencias y comparaciones.

- ✓ **No volátiles**

La información existente en un almacén de datos no puede ser modificada, solo incrementada. Por tanto, el período que generalmente cubre un Almacén de Datos en cuanto a la información almacenada es de 2 a 10 años. (Velasco, 2007)

1.3.2. Beneficios e inconvenientes del uso de los almacenes de datos.

Beneficios.

- ✓ Acelera las consultas mediante la reducción del volumen de datos a recorrer.
- ✓ Estructura los datos para un mejor acceso por una herramienta.
- ✓ Divide los datos para imponer estrategias de control de acceso.

- ✓ Segmenta los datos en diferentes plataformas de hardware.
- ✓ Permite el acceso a los datos mediante un gran número de herramientas, lo que las independiza.

Inconvenientes.

- ✓ Baja estimación de los recursos necesarios para la carga de datos.
- ✓ Problemas ocultos en los sistemas fuente.
- ✓ Homogenización de los datos.
- ✓ Alto mantenimiento.
- ✓ Proyectos de larga duración.
- ✓ Complejidad de integración.

1.3.3. Mercados de Datos.

Al hacer un análisis de la bibliografía existente sobre la tecnología de Data Warehousing se puede constatar que existen autores que utilizan los conceptos de “Almacén de Datos” y “Mercados de Datos” indistintamente refiriéndose al mismo tema aunque realmente no definen un concepto idéntico. Los Mercados de Datos, también conocidos como DataMart (DM), son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones, mientras que un DW entrega información a nivel corporativo. Un concepto más amplio sería: el Data Mart es un sistema orientado a la consulta, es consultado mediante herramientas OLAP (On line Analytical Processing - Procesamiento Analítico en Línea) que ofrecen una visión multidimensional de la información. Sobre estas BD se pueden construir EIS (Executive Information Systems, Sistemas de Información para Directivos) y DSS (Decision Support Systems, Sistemas de Ayuda a la toma de Decisiones). En síntesis, se puede decir que los Data Mart son pequeños Data Warehouse centrados en un tema o un área de negocio específico dentro de una organización. (Peñaloza, 2008)

La visión de Inmon se basa en un enfoque descendente, propone construir primero el almacén de datos, y a partir de este los mercados de datos. 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 mercado de datos corporativo, convirtiéndose en un complejo empresarial de BD relacionales. A diferencia de la anterior, la propuesta de Kimball se basa en dividir el mundo de Inteligencia de Negocio entre los Hechos y las Dimensiones, ésta 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, está el hecho de poseer una arquitectura ascendente, plantea que se debe crear por cada departamento un conjunto de mercado de datos independientes orientados a los temas que estén relacionados con él. Se estima que un mercado de datos es una versión especial de almacén de datos, pues son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones. (Peñaloza, 2008)

1.3.4. Metadatos.

Dentro de los almacenes de datos se tiene otro concepto que no se debe dejar de mencionar que es el de metadatos, que no son más que datos relativos a datos. Estos permiten mantener información acerca de la procedencia de los datos, su fiabilidad, forma de cálculo, etc. (Rayner Huamantumba., 2007)

Sus objetivos principales son:

- ✓ Ayudar a los responsables técnicos del Almacén de Datos con respecto a la gestión de la información histórica, la administración del almacén y la elaboración de programas de extracción de los datos entre otros aspectos.
- ✓ Ayudar a los usuarios finales a entender la información contenida en el Almacén de Datos en su propio lenguaje de negocio, indicando que información hay y que significa, además de ayudar a construir consultas, informes y análisis.

1.3.5 Base de Datos.

Una BD se define como un fichero en el cual se almacena información en campos o delimitadores, teniendo acceso a ella posteriormente tanto de forma separada como de forma conjunta. Se usa normalmente para recoger grandes cantidades de información. El contenido de una BD engloba a la información concerniente de una organización, de tal manera que los datos estén disponibles para los usuarios. (Pérez, 2009)

Las BD tiene varias características, algunas de ellas son:

- ✓ Independencia de los datos: los datos no dependen del programa y por tanto cualquier aplicación puede hacer uso de los datos.
- ✓ Reducción de la redundancia: se llama redundancia a la existencia de duplicación de los datos, al reducir ésta al máximo se conseguirá un mayor aprovechamiento del espacio y además se evita que existan inconsistencias

entre los datos. Las inconsistencias se dan cuando se encuentran con datos contradictorios.

- ✓ Seguridad: un sistema de BD debe permitir un control sobre la seguridad de los datos.
- ✓ Se visualiza normalmente como una tabla de una hoja de cálculo, en la que los registros son las filas y las columnas son los campos.
- ✓ Permite la programación a usuarios avanzados y realizar un listado de la BD. (Pérez, 2009)

Las BD son uno de los recursos más significativos para el desarrollo de las empresas, demostrando así que son de gran influencia en el progreso de esta área, puesto que brindan un almacenamiento y acceso confiable en el uso de la información que se produce. Es una de las herramientas más importantes en la gestión de información.

1.4 Modelos de Almacenamiento de Datos.

En todo el trascurso de desarrollo de la esfera de las BD, se han utilizado varias formas de modelado para el diseño de las mismas. Dentro de los modelos más importantes se tienen:

Modelo relacional: El diagrama o modelo entidad-relación (DER), es un lenguaje para el modelado de datos de un sistema de información. Estos modelos expresan entidades más relevantes para el sistema, sus inter-relaciones y propiedades. Trabajan dividiendo los datos en muchas entidades discretas donde cada una se convierte en una tabla física en la BD operacional. Es el modelo conceptual más utilizado para el diseño conceptual de BD, está formado por un conjunto de conceptos que permiten describir la realidad mediante un conjunto de representaciones gráficas y lingüísticas. (S.L, 2007-2010)

Modelo dimensional: A diferencia de los clásicos sistemas de BD que presentan sus estructuras diseñadas mediante el modelo Entidad-Relación, los Almacenes de Datos se diseñan mediante un Modelo Dimensional. Poseen la misma información que el DER, pero la organiza de manera 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 cubra el universo de variantes que los usuarios necesiten, para consultar la información almacenada. En un modelo de datos dimensional los datos se organizan alrededor de los temas de la organización. La estructura de los datos que se maneja en este modelo son matrices multidimensionales o hipercubos. Un hipercubo consiste en un conjunto de celdas, las cuales se identifican en particular con la combinación de los miembros de las diferentes dimensiones y contiene el valor de la medida analizada por

dicha combinación de dimensiones. Este modelo está compuesto por dimensiones y hechos, las dimensiones no son más que la representación de cada uno de los ejes dimensionales. Suministran el contexto de cómo se obtienen las medidas de los hechos, se puede decir, que se utilizan para seleccionar y agrupar por niveles. Por su parte, los hechos no son más que los objetos que se van a analizar, poseen atributos llamados eje o síntesis y son de tipo cuantitativo, sus valores se obtienen generalmente por la aplicación de una función estadística que resume el conjunto de valores en un único valor. (S.L, 2007-2010)

El modelo dimensional implica tres variantes de modelamiento de acuerdo a la complejidad del sistema:

- ✓ **Esquema estrella:** es la técnica más común. En este esquema existe un único elemento central (tabla de hechos) conectado radialmente con las tablas de dimensiones, juntos forman una estructura que permite implementar mecanismos básicos para poder utilizarla con una herramienta de consultas OLAP.
- ✓ **Esquema copo de nieve (Snowflake):** es un esquema derivado del de estrella, donde las tablas de dimensiones se ramifican en más puntas. Para extraer datos de las tablas a veces hay que vincular muchas tablas en las sentencias SQL que puede llegar a ser muy complejo y difícil para mantener.
- ✓ **Esquema constelación:** está compuesto por una serie de esquemas de estrella, es decir, una tabla de hechos central con otras auxiliares y sus respectivas tablas de dimensiones. Tiene mucha flexibilidad y este factor es su grande virtud. Sin embargo, el problema es que cuando el número de las tablas vinculadas aumenta, la arquitectura puede llegar a ser muy compleja y difícil para mantener. (S.L, 2007-2010)

En este trabajo se utilizará el **modelo dimensional**, el cual se basa en un esquema estrella, soportando de medianas a grandes BD. Consume un alto grado de recursos por las necesidades de almacenamiento y se caracteriza por una velocidad muy buena cuando trabaja con conjuntos pequeños o grandes. Además, se hará uso del **esquema estrella**, siendo la técnica más común y permitiendo más comodidad a la hora de realizar el diagrama, por poseer un solo elemento central conectado radialmente con las dimensiones y por tener tiene mejor rendimiento y velocidad que el de constelación y copo de nieve.

1.5 Modos de Almacenamiento OLAP.

Las primeras soluciones OLAP (On Line Analytical Processing), estuvieron basadas en BD multidimensionales (MDDBS). Un cubo estructural (dos veces un hipercubo o un arreglo multidimensional) almacenaba los datos para que se puedan manipular intuitivamente y claramente ver las asociaciones a través de dimensiones múltiples, pero este enfoque tiene varias limitaciones:

- ✓ Las nuevas estructuras de almacenamiento de datos requieren BD propietarias. No hay realmente estándares disponibles para acceder a los datos multidimensionales.
- ✓ La segunda limitación de un MDDB concierne al desarrollo de una estructura de datos. Las compañías generalmente almacenan los datos de la empresa en BD relacionales, lo que significa que alguien tiene que extraer, transformar y cargar estos datos en el hipercubo

La tecnología de Procesamiento Analítico en línea (OLAP) permite un uso más eficaz de los almacenes de datos en línea, siendo el componente más poderoso y convirtiéndose en el motor de consultas especializado del almacén de datos. Consultas analíticas complejas e iterativas utilizadas generalmente para sistemas de ayuda para la toma de decisiones. (Mavilio.)

Las principales características de OLAP son:

- ✓ **Rápido:** proporciona la información al usuario a una velocidad constante (mayormente en cinco segundos o menos).
- ✓ **Análisis:** realiza análisis estadísticos y numéricos básicos de los datos.
- ✓ **Compartida:** permite compartir los datos potencialmente confidenciales a través de una gran cantidad de usuarios, implementando para esto los requerimientos de seguridad necesarios.
- ✓ **Multidimensional:** permite ver la información en determinadas vistas o dimensiones.
- ✓ **Información:** accede a todos los datos necesarios, donde quiera que ésta resida y mientras no esté limitada por el volumen. (no.3, Sept. 2007)

1.5.1 Tipos de OLAP.

- ✓ **ROLAP (Procesamiento Analítico en línea Relacional)**

En el Procesamiento Analítico Relacional en Línea (*Relational Online Analytical Process*, en inglés) los datos son almacenados en filas y columnas de forma relacional. Este modelo presenta los datos a los usuarios en forma de dimensiones de negocio. Con el fin de ocultar las estructuras de almacenamiento y presentar los datos dimensionalmente es creada la semántica de las etiquetas de los metadatos. Ellas soportan el mapeo de las dimensiones a las tablas relacionales. Estos metadatos también son almacenados en tablas 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 BD 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 cuáles los usuarios realizan los análisis OLAP (no.3, Sept. 2007) (ver figura 1)

Arquitectura del Sistema ROLAP



Figura 1 Arquitectura del Sistema ROLAP

Después de que el modelo de datos para el Almacén de Datos se ha definido, los datos se cargan desde el sistema operacional. Se ejecutan rutinas de BD para agregar el dato, si así es requerido por el modelo de datos. Se crean entonces los índices para optimizar los tiempos de acceso a las consultas.

Los usuarios finales ejecutan sus análisis multidimensionales, a través del motor ROLAP, que transforma dinámicamente sus consultas a consultas SQL. Se ejecutan estas consultas SQL en las BD relacionales, y sus resultados se relacionan mediante tablas cruzadas y conjuntos multidimensionales para devolver los resultados a los

usuarios. La arquitectura ROLAP es capaz de usar datos precalculados si estos están disponibles, o de generar dinámicamente los resultados desde los datos elementales si es preciso. Esta arquitectura accede directamente a los datos del almacén, y soporta técnicas de optimización de accesos para acelerar las consultas. Estas optimizaciones son: particionado de los datos a nivel de aplicación, soporte a la desnormalización y joins múltiples, entre otras. (no.3, Sept. 2007)

✓ **MOLAP (Procesamiento Analítico en línea Multidimensional)**

Un sistema MOLAP usa una BD propietaria multidimensional, en la que la información se almacena y visualiza multidimensionalmente. Este sistema utiliza una arquitectura de dos niveles: la BD multidimensionales y el motor analítico. La BD multidimensional es la encargada del manejo, acceso y obtención del dato; y el motor analítico es el responsable de la ejecución de los requerimientos OLAP. El nivel de presentación se integra con el de aplicación y proporciona una interfaz a través de la cual los usuarios finales visualizan los análisis OLAP.

El nivel de aplicación es el responsable de la ejecución de los requerimientos OLAP. El nivel de presentación se integra con el de aplicación y proporciona un interfaz a través del cual los usuarios finales visualizan los análisis OLAP. Una arquitectura cliente/servidor permite a varios usuarios acceder a la misma BD multidimensional. (no.3, Sept. 2007) (ver figura 2)

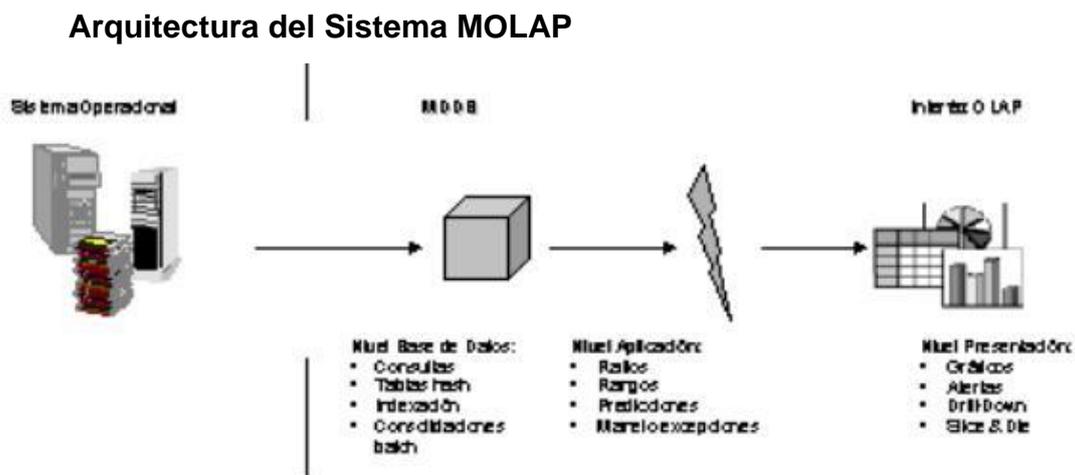


Figura 2 Arquitectura del Sistema MOLAP

La información procedente de los sistemas operacionales, se carga en el sistema MOLAP, mediante una serie de rutinas por lotes. Una vez cargado el dato elemental

en la BD multidimensional (MDDDB), se realizan una serie de cálculos por lotes, para calcular los datos agregados, a través de las dimensiones de negocio, rellorando la estructura MDDDB. Tras rellenar esta estructura, se generan unos índices y algoritmos de tablas hash para mejorar los tiempos de accesos a las consultas. Una vez que el proceso de compilación se ha acabado, la MDDDB está lista para su uso. Los usuarios solicitan informes a través del interface, y la lógica de aplicación de la MDDDB obtiene el dato. (no.3, Sept. 2007)

1.5.2 Comparación entre MOLAP y ROLAP.

Se han originado muchos debates acerca de dos tipos de almacenamiento MOLAP y ROLAP. Casi siempre, las implementaciones de MOLAP presentan el mejor rendimiento de la tecnología relacional, pero tiene problema de escalabilidad, por ejemplo: la adición de dimensiones a un esquema ya existente. Sin embargo, las implementaciones de ROLAP son más escalables y a menudo son más sociables debido a que se aprovechan de las inversiones de la tecnología de las BD relacionales. (no.3, Sept. 2007)

A continuación se muestran algunas diferencias entre los modelos MOLAP y ROLAP:

	MOLAP	ROLAP
	Multidimensional	Relacional
Datos	Detalle y pre calculados	Detalle y agregado
Estructura	Matrices comprimidas	Tablas relacionales
Administración	Especialista en BDMD	Administrador de BD
Acceso	Lenguaje especializado	SQL

Tabla 1 Comparación entre los modelos MOLAP y ROLAP.

Se decidió usar en la realización de este trabajo el modo de almacenamiento **ROLAP**, por generar dinámicamente los resultados desde la información elemental, soportando técnicas de optimización para acelerar las consultas como tablas particionadas,

brindando soporte a la desnormalización, múltiples reuniones, precalculado de datos, e índices.

1.6. Fundamentación de la metodología a utilizar.

Las **metodologías** no son más que el conjunto de procedimientos, técnicas, herramientas que ayuda a los desarrolladores a realizar un nuevo producto. Posterior a la creación del concepto de Almacén de datos, surgen dos tendencias de desarrollo, defendidas por los dos creadores de este concepto: William H. Inmon y Ralph Kimball respectivamente. (Rayner Huamantumba., 2007).

Según Bill Inmon, el almacén de datos es una parte del todo que conforma un sistema de inteligencia. Una entidad tiene un Almacén de Datos y los mercados de datos tienen como fuente de información ese almacén. Esta tendencia es conocida como "Top-Down". Para Ralph Kimball el almacén de datos se compone por el conjunto de mercados de datos que existan en la entidad donde esté implementado y la información siempre se almacena en un modelo dimensional. Esta tendencia es conocida como "Bottom-Up". No se puede decir que una de estas dos tendencias es mejor que la otra, puesto que los beneficios de cada una dependen del ambiente de la empresa en que se vayan a implantar. Tanto Inmon como Kimball coinciden en algo: los almacenes y mercados de datos independientes no satisfacen las necesidades de precisión y oportunidad de la información, como tampoco facilitan el acceso por parte del usuario ni el proceso de toma de decisiones. (Rayner Huamantumba., 2007)

A partir de estos dos conceptos, creados por las dos figuras más relevantes en el mundo de la tecnología de almacenes de datos, surgen las distintas metodologías a seguir para implementar con éxito un Almacén de Datos. A continuación se explican brevemente algunas de ellas.

- ✓ **Hefesto** permite la construcción de un almacén de datos de forma sencilla, ordenada e intuitiva. Lo que se busca con esta metodología es entregar una primera versión que satisfaga algunas de las necesidades iniciales, para demostrar las ventajas del uso del AD y motivar así al usuario.
- ✓ **SQLBI** es la metodología creada por Microsoft y orientada específicamente para sus productos (Microsoft SQL Server, etc.). Tiene como principales figuras a Marco Russo y Albert Ferrari.
- ✓ La **Metodología para el Diseño Conceptual de Almacenes de datos** fue descrita en la tesis de Doctorado de Leopoldo Zenaido Zepeda Sánchez y aporta la utilización de los Casos de Uso durante el desarrollo del proyecto. Así

mismo define un grupo de transformaciones que serán capaces de llevar un modelo relacional a uno dimensional, obteniendo las estructuras que formarán parte del Repositorio de Datos.

- ✓ La metodología **CRISP-DM** está descrita en términos de un modelo de proceso jerárquico que consta de un conjunto de tareas incluidas en cuatro niveles de abstracción que van de lo general a lo específico. El *nivel de fase* organiza el proceso de minería de datos por fases. El *nivel de tareas genéricas* es el encargado de cubrir, mediante un conjunto de tareas, todas las posibles situaciones de la minería de datos. El *nivel de tarea especializada* describe la realización de las tareas genéricas y el *nivel de instancia de proceso* es un registro de las acciones, decisiones y resultados de la minería de datos real. (Rayner Huamantumba., 2007)

En este caso para definir la metodología de desarrollo se tomó como base la **Metodología de Kimball** la cual se enfoca principalmente en el diseño de la BD que almacenará la información para la toma de decisiones. Propone ir construyendo el Almacén de Datos a través de la construcción de los Mercados de Datos departamentales, lo que constituye una estrategia buena y coincide con la división lógica de las empresas, entidades, organismos, etc. El diseño se basa en la creación de tablas de hechos (FACTS) que son tablas que contienen la información numérica de los indicadores a analizar, es decir la parte cuantitativa de la información. Debido a las características que posee se elige esta metodología. Las características son:

- ✓ Utiliza el concepto de Dimensiones y Hechos, lo cual facilita el proceso de desarrollo y lo hace más eficaz.
- ✓ Se basa en el paradigma de construcción de Almacenes de Datos “Bottom-Up”, que define la creación de los Mercados de datos departamentales y su posterior integración en un Almacén de datos de la organización. Esta estrategia coincide con la división lógica que se tiene en una organización.
- ✓ Es una metodología madura, bien definida y reconocida por el resto de la comunidad que se dedica al tema.
- ✓ Existe abundante documentación acerca de ella, además de grupos que se dedican a brindar ayuda del tema en la red. (Bernabeu., 2007)

1.6.1 Metodología utilizada en el centro.

La metodología que se propone utilizar en el presente trabajo, es desarrollada por la Línea de Soluciones de Almacenes de Datos e Inteligencia de Negocios, del Centro de Tecnologías y Análisis de Datos de la Universidad de las Ciencias Informáticas la misma se caracteriza por cubrir todas las fases del ciclo de vida por las que pasa la construcción de un almacén de datos, desde el levantamiento de información inicial hasta la implementación de la herramienta de Inteligencia de Negocio. Es una metodología compuesta, que reúne elementos de varias metodologías de desarrollo de proyectos de integración de datos. (Villa, 2009)

En la primera fase que desarrolla la metodología, se resalta el levantamiento de información a nivel de negocio para identificar los posibles indicadores y aspectos a medir en los análisis, que luego de algunas transformaciones se convierten en los requerimientos de información de entrada y de salida para la solución de integración. De forma paralela a esta actividad se lleva a cabo un estudio de las fuentes de datos que soportan los datos a cargar. Cumplidas estas dos tareas, se aprueba que la información levantada sobre las necesidades de los clientes esté realmente almacenada en las fuentes correspondientes, para a continuación, teniendo los requerimientos informativos correctamente definidos, proceder a diseñar la solución de BD. Una vez diseñada la estructura de BD, se realiza la carga de los datos desde las fuentes y posteriormente se implementan los requerimientos de inteligencia de negocio identificados en el levantamiento de información inicial. Las actividades y artefactos de la solución son realizados por 4 grupos que conforman la línea, especializados en componentes específicos de la solución. (Villa, 2009)

Los grupos de desarrollo estarán divididos en 4, ellos son:

- ✓ Grupo de Análisis.
- ✓ Grupos de Almacenes.
- ✓ Grupo de Integración de Datos
- ✓ Grupo de Inteligencia de Negocio.

Cada uno de estos grupos poseen sus objetivos específicos, artefactos y herramientas, así como, las personas que trabajan en ellos, protagonizan roles y responsabilidades diferentes. En el proceso de desarrollo, se tiene en cuenta el Flujo de Artefactos por fases del proyecto, la Distribución de Artefactos por fases y etapas de desarrollo del proyecto, el Cronograma Genérico gerencial y detallado de los

proyectos de soluciones de almacenes de datos e inteligencia de negocio. Vale aclarar, que todos los pasos y las fases que se aplican en esta metodología marcan pautas importantes en este proceso, de esta manera, se destacan las siguientes fases:

- ✓ Estudio preliminar.
- ✓ Fase requerimientos.
- ✓ Fase de Arquitectura y Diseño.
- ✓ Fase de Implementación.
- ✓ Fase de prueba.
- ✓ Fase de despliegue.
- ✓ Fase de soporte. (Con esta fase se le da cierre al proyecto desarrollado)(DATEC).

La ONE, basándose en su papel como órgano rector para analizar, recopilar y difundir toda la información dentro de todos los sectores del país, merece la utilización de una metodología robusta y madura que garantice el éxito de la integración de la información que actualmente disponen. Luego de analizar todo el conjunto de metodologías existentes para enfrentar el desarrollo del mercado de datos, se escoge esta metodología por política del proyecto, pues el centro ajustó la mundialmente conocida como metodología de Kimball a las necesidades del mismo, para un mejor desarrollo en tiempo real de sus productos.

1.7 Gestor de Bases de Datos

Un Sistema de Gestión de BD (SGBD) es el software que permite la utilización y actualización de los datos almacenados en una (o varias) base(s) de datos por uno o varios usuarios desde diferentes puntos de vista y a la vez. (2007)

A continuación se mencionan algunos de los diferentes gestores de BD, tales como:

Microsoft SQL Server (MySQL), propietario de Microsoft, pertenece a la familia de los sistemas de administración de BD, operando en una arquitectura cliente/servidor de gran rendimiento. Su desarrollo fue orientado para manejar grandes volúmenes de información, y un elevado número de transacciones. SQL Server es una aplicación completa que realiza toda la gestión relacionada con los datos. SQL Server permite la creación de procedimientos almacenados, los cuales consisten en instrucciones SQL

que se almacenan dentro de una BD, se trata de procedimientos que se guardan semicompileados en el servidor y que pueden ser invocados desde el cliente. Se ejecutan más rápido que instrucciones SQL independientes. SQL Server puede manejar perfectamente BD de terabytes con millones de registros y funciona sin problemas con miles de conexiones simultáneas a los datos. (Matos, 2001)

Oracle: Oracle Corporation es la primera compañía mundial proveedora de soluciones de software al mundo de la empresa. Con unos ingresos de 10.900 millones de dólares, la compañía ofrece su BD, herramientas y aplicaciones de gestión, junto con los correspondientes servicios de consultoría, formación y soporte, en más de 145 países de todo el mundo. En la actualidad es una de las corporaciones más completas, a nivel mundial, en el desarrollo de aplicaciones de este tipo, tanto para la optimización y mantenimiento de las BD en función de las mejoras de las consultas complejas y reportes, como en el desarrollo de cuadros de mandos integrales y sistemas estratégicos de negocios empresariales enfocados a la toma de decisiones a nivel empresarial. Una de las principales ventajas que presenta dicha compañía, es su arquitectura multiplataforma en el desarrollo de aplicaciones. Permite la realización de auditorías de los datos, el modelado relacional y dimensional, la gestión de datos y metadatos, opción avanzada de carga de datos, soporte de dimensiones lentamente cambiantes, traza lineal de principio a fin. Incluye el diseño de estructuras OLAP y relacionales, permitiendo la integración entre ambas funciones, así como la posibilidad de mantener un entorno integrado y el almacenamiento centralizado. El servidor Oracle brinda a los usuarios la ventaja de poder utilizar la herramienta Oracle Business Intelligence Suite, actualmente en su versión mejorada Oracle Business Intelligence Enterprise Edition Plus, para la realización de estudios de mercados e Inteligencia de Negocio. (2005-2010)

PostgreSQL 8.4, desde su creación ha alcanzado gran popularidad en todas las esferas en las que se ha aplicado. Esto se debe principalmente al hecho de que posee una instalación ilimitada, es decir, la licencia del software no tiene un costo asociado. PostgreSQL mejora la performance alrededor de un 20% en total en pruebas de sistemas OLTP (procesamiento de transacciones en línea) de alta gama. Los usuarios pueden ganar aún más en eficiencia en sistemas de análisis de datos (*data warehousing*). Los cambios incluyen ordenamientos más rápidos en disco y en memoria, mejor escalabilidad multi-procesador, mejores planes de ejecución en consultas sobre datos particionados, cargas masivas más rápidas y outer joins sumamente acelerados. El mismo trae consigo las ventajas adicionales siguientes:

- ✓ Modelos de negocios más rentables con instalaciones a gran escala.
- ✓ La organización no será auditada en ningún momento para verificar cumplimiento de licencia.
- ✓ Posibilidad de realizar investigaciones y desarrollo sin la necesidad de incurrir en costo de licenciamiento. (Varios, 2007)

También ofrece muchas ventajas para su compañía o negocio, respecto a otros sistemas de BD, como son:

Instalación ilimitada

Es frecuente que las BD comerciales sean instaladas en más servidores de lo que permite la licencia. Algunos proveedores comerciales consideran a esto la principal fuente de incumplimiento de licencia. Con PostgreSQL, nadie puede demandarlo por violar acuerdos de licencia, puesto que no hay costo asociado a la licencia del software.

Esto tiene varias ventajas adicionales:

- ✓ Modelos de negocios más rentables con instalaciones a gran escala.
- ✓ No existe la posibilidad de ser auditado para verificar cumplimiento de licencia en ningún momento.
- ✓ Flexibilidad para hacer investigación y desarrollo sin necesidad de incurrir en costos adicionales de licenciamiento. (Varios, 2007)

Extensible

El código fuente está disponible para todos sin costo. Si su equipo necesita extender o personalizar PostgreSQL de alguna manera, pueden hacerlo con un mínimo esfuerzo, sin costos adicionales.

Multiplataforma

PostgreSQL está disponible en casi cualquier Unix (34 plataformas en la última versión estable), y una versión nativa de Windows está actualmente en estado beta de pruebas.

Diseñado para ambientes de alto volumen

PostgreSQL usa una estrategia de almacenamiento de filas llamada MVCC para conseguir una mucha mejor respuesta. Los principales proveedores de sistemas de BD comerciales usan también esta tecnología, por las mismas razones.

En fin a continuación se muestra una lista breve de características técnicas que PostgreSQL ofrece:

- ✓ Cumple completamente con ACID
- ✓ Cumple con ANSI SQL

- ✓ Integridad referencial.
- ✓ Replicación (soluciones comerciales y no comerciales) que permiten la duplicación de BD maestras en múltiples sitios de réplica.
- ✓ Interfaces nativas para ODBC, JDBC, C, C++, PHP, Perl, TCL, ECPG, Python y Ruby.
- ✓ Reglas.
- ✓ Vistas.
- ✓ Unicode.
- ✓ Secuencias.
- ✓ Herencia.
- ✓ Procedimientos almacenados.
- ✓ Lenguajes procedurales.
- ✓ Bloqueo a nivel mejor-que-fila.
- ✓ Índices parciales y funcionales.
- ✓ Soporte para consultas con UNION, UNION ALL y EXCEPT.
- ✓ Extensiones para SHA1, MD5, XML y otras funcionalidades.
- ✓ Herramientas para generar SQL portable para compartir con otros sistemas compatibles con SQL.
- ✓ Sistemas de tipos de datos extensibles para proveer tipos de datos definidos por el usuario, y rápido desarrollo de nuevos tipos
- ✓ Funciones de compatibilidad para ayudar en la transición desde otros sistemas menos compatibles con SQL. (Varios, 2007)

1.7.1 ¿Por qué PostgreSQL 8.4?

Una vez analizados los sistemas gestores de BD se decidió utilizar en la elaboración del sistema planteado al *PostgreSQL* por todas las ventajas que este representa respecto a otros sistemas. A continuación se plantean algunas razones de las que se tuvieron en cuenta para la selección. *PostgreSQL* es un gestor de BD de tipo objeto relacional mientras que *Oracle* es sólo relacional, por ende *PostgreSQL* maneja herencia que es una ventaja. En *Oracle* no se pueden crear tipos de datos personalizados, mientras que en *PostgreSQL* sí. El *Oracle* tiene un coste muy elevado, en tanto el hecho de que *PostgreSQL* es un producto *Open Source*, sin costos de licencia lo hacen extremadamente atractivo. Girando la comparación ahora hacia *MySQL* se puede decir que este no considera las claves ajenas e ignora la integridad referencial, aspecto que se deja en manos del programador de la aplicación sin embargo *PostgreSQL* soporta transacciones, considera las claves ajenas, con comprobaciones de integridad referencial. *MySQL* es considerado más rápido, pero

también es cierto que por lograr la rapidez se sacrificaron aspectos muy importantes que lo hacen menos útil, también es importante destacar que *PostgreSQL* ha soportado el triple de peticiones simultáneas de lo que ha soportado *MySQL*. Mencionando otros puntos a favor del *PostgreSQL* lo primero, sin duda, es la economía tangible de decenas de miles de dólares en licencias de *software*. Otro punto importante es la simplificación del proceso de administración de licencias de *software*, que no es necesario cuando se usa *software* libre. Por otra parte, también ofrece beneficios como el aumento de la estabilidad y *performance* del sistema.

1.8. Herramienta de Administración.

PgAdmin 3.0: Es una herramienta de código abierto para la administración de BD PostgreSQL y derivados (EnterpriseDB Postgres Plus Advanced). Tiene además:

- ✓ Interfaz administrativa gráfica.
- ✓ Herramienta de consulta SQL (con un EXPLAIN gráfico).
- ✓ Editor de código procedural.
- ✓ Agente de planificación SQL/shell/batch.
- ✓ Administración de Slony-I.

PgAdmin se diseña para responder a las necesidades de la mayoría de los usuarios, desde escribir simples consultas SQL hasta desarrollar BD complejas. La interfaz gráfica soporta todas las características de PostgreSQL y hace simple la administración. Está disponible en más de una docena de lenguajes y para varios sistemas operativos, incluyendo Microsoft Windows, Linux, FreeBSD, Mac OSX y Solaris. Soporta versiones de servidores 7.3 y superiores. Versiones anteriores a 7.3 deben usar el PgAdmin II.

En la instalación del PgAdmin3, se siguen dos procesos, en Linux (Ubuntu y derivados de Debian), se debe instalar el paquete pgadmin3: `sudo apt-get install pgadmin3` y en Windows, pgAdmin viene incorporado con el instalador de PostgreSQL. Los instaladores, si bien son multilinguaje, generalmente hay que establecer qué lenguaje usará la interfaz. (ArPug, 2009)

1.9 Herramientas de modelado.

Las herramientas se ven simplemente como herramientas que cualquiera puede escoger y utilizar (como un martillo) para desarrollar un sistema de información, su selección e implementación casi siempre llevará a una reducida productividad y calidad. En resumen, las herramientas son un complemento de la caja de

herramientas del ingeniero del software, las cuales van a proporcionar al ingeniero la posibilidad de automatizar actividades manuales y de mejorar su visión general de la ingeniería, para asegurar la calidad de un producto desde su diseño antes de construirlo. (2009)

A continuación se mencionan algunas herramientas:

ER/STUDIO 2.5

Es una herramienta de modelado de datos fácil de usar y multinivel, para el diseño y construcción de BD a nivel físico y lógico. Direcciona las necesidades diarias de los administradores de BD, desarrolladores y arquitectos de datos, que construyen y mantienen aplicaciones de BD grandes y complejos. ER/Studio está equipado para crear y manejar diseños de BD funcionales y confiables. Ofrece fuertes capacidades de diseño lógico, sincronización bidireccional de los diseños físicos y lógicos, construcción automática de BD, documentación y fácil creación de reportes. (2009)

- ✓ Superior en la productividad utilizando un sistema de archivos de modelo: Permite a los usuarios crear múltiples modelos físicos e independientes entre sí, de un modelo lógico común dentro de la misma. Ofrece al usuario una solución completa y fácil de entender. Permite la actualización de los modelos lógicos, modelos físicos y BD en directo, sin tener que gestionar y sincronizar muchos archivos sueltos.
- ✓ Superior en trazabilidad con el linaje de los metadatos y análisis de impacto: Ofrece funciones para el mantenimiento de la trazabilidad entre el modelo lógico original y múltiples modelos físicos. Es capaz de mantener varios modelos dentro del mismo archivo, el análisis de impacto es mucho más fácil y más fiable.
- ✓ Superior en documentación y comunicación: Incluye un portal Web que permite a los desarrolladores y analistas de negocio, navegar, buscar e informar sobre el modelo. El portal cuenta con un repositorio de informes por separado que está optimizado para la búsqueda y presentación de informes.
- ✓ Superior en confiabilidad y promoción: Siempre ha estado centrada en ayudar a las empresas a diseñar, crear, optimizar y ejecutar sus aplicaciones de software y BD. Puede decidir qué productos apoyar y llevar a cabo lo que constituye una clara ventaja competitiva.
- ✓ Opciones de licencia flexibles que reducen drásticamente el coste total.

Visual Paradigm 3.4

Visual Paradigm para UML es una herramienta CASE profesional que utiliza UML como lenguaje de modelaje, que soporta el ciclo de vida completo del desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue. Permite dibujar todos los tipos de diagramas de clases, código inverso, generar código desde diagramas y generar documentación. Se integra con las herramientas Java Eclipse/IBM WebSphere, Jbuilder, NetBeans, Oracle, JDeveloper y BEA Weblogic. Está disponible en varias ediciones, cada una destinada a unas necesidades: Enterprise, Professional, Community, Standard, Modeler y Personal. El software de modelado UML ayuda a una más rápida construcción de aplicaciones de calidad, mejores y a un menor coste. Soporta un conjunto de lenguas, tanto en la generación de código y de ingeniería inversa en Java, C + +, CORBA IDL, PHP, XML Schema, Ada y Python. Además, apoya la generación de código C #, VB. NET, Objeto Definition Language (ODL), Flash ActionScript, Delphi, Perl, Objective-C, y Ruby.

Los desarrolladores pueden diseñar documentación del sistema con una plantilla de diseño. Los analistas del sistema pueden estimar las consecuencias de los cambios con los diagramas de análisis de impacto, tales como la matriz y el diagrama de análisis. Es una herramienta que posee una licencia gratuita y comercial. Tiene como características que es un producto de calidad, soporta aplicaciones Web, presenta generación de código para Java y exporta como HTML. Es fácil de instalar y actualizar y tiene compatibilidad entre ediciones. (Arg, 2010)

Además brinda una versión libre para uso no comercial.

También brinda características como:

- ✓ Producto de calidad.
- ✓ Soporta aplicaciones web.
- ✓ Las imágenes y reportes generados, no son de muy buena calidad.
- ✓ Generación de código para Java y exportación como HTML.
- ✓ Fácil de instalar y actualizar.
- ✓ Compatibilidad entre ediciones.
- ✓ Soporte de UML versión 2.1.

- ✓ Diagramas de Procesos de Negocio - Proceso, Decisión, Actor de negocio, Documento.
- ✓ Modelado colaborativo con CVS y Subversión.
- ✓ Interoperabilidad con modelos UML2 (metamodelos UML 2.x para plataforma Eclipse) a través de XML.
- ✓ Ingeniería inversa - Código a modelo, código a diagrama.
- ✓ Ingeniería inversa Java, C++, Esquemas XML, XML, NET exe/dll, CORBA IDL.
- ✓ Generación de código - Modelo a código, diagrama a código.
- ✓ Editor de Detalles de Casos de Uso, entorno todo-en-uno para la especificación de los detalles de los casos de uso, incluyendo la especificación del modelo general y de las descripciones de los casos de usos.
- ✓ Generación de código y despliegue de EJB's - Generación de beans para el desarrollo y despliegue de aplicaciones.
- ✓ Diagramas de flujo de datos.
- ✓ Soporte ORM - Generación de objetos Java desde la BD.
- ✓ Generación de BD. Transformación de diagramas de Entidad-Relación en tablas de BD.
- ✓ Ingeniería inversa de BD. Desde Sistemas Gestores de BD (DBMS) existentes a diagramas de Entidad-Relación.
- ✓ Generador de informes para generación de documentación.
- ✓ Distribución automática de diagramas - Reorganización de las figuras y conectores de los diagramas UML.
- ✓ Importación y exportación de ficheros XML. (Arg, 2010)

1.9.1 ¿Por que Visual Paradigm?

Visual Paradigm soporta la generación de código e ingeniería inversa en Java, C++, CORBA IDL, PHP, XML Schema, Ada y Python. Además ayuda la generación de código en C#, VB .NET, Object Definition Language (ODL), Flash ActionScript, Delphi, Perl, Objective-C, y Ruby, soporta la ingeniería inversa en clases de *Java*, *dll.NET* y

exe, *JDBC*, y el mapeo de archivos en *Hibernate*. Otro aspecto muy importante es la excelente documentación con que cuenta, en su sitio oficial se explican detalladamente todas las funcionalidades. Además, la herramienta es colaborativa pues soporta múltiples usuarios trabajando sobre el mismo proyecto, permite el control de versiones y genera la documentación en varios formatos. En definitiva *Visual Paradigm* es una herramienta *CASE* muy robusta, destacando además su usabilidad y portabilidad.

Conclusiones.

A partir del análisis del estado del arte y de la importancia que reviste, hoy más que nunca, la utilización de los almacenes de datos para ayudar en la toma de decisiones a partir de la información obtenida de un sector económico, se pudo establecer que ninguna de las soluciones estudiadas cubre aún las necesidades que tiene la Oficina Nacional de Estadística de Cuba. Por ende, se decide implementar una solución que facilite la disponibilidad de la información y dé respuesta al problema existente, para ello se obtuvieron los resultados siguientes:

- ✓ Se decidió que la metodología adecuada para resolver la problemática en materia, es una adaptación de la metodología de Kimball, desarrollada por el centro DATEC, puesto que la misma cumple con los requerimientos necesarios para la implementación del Mercado de Datos.
- ✓ La solución se diseñará según el esquema estrella.
- ✓ Se usará el modelo dimensional.
- ✓ Se utilizará el modo de almacenamiento ROLAP.
- ✓ Se definió como sistema gestor de BD PostgreSQL 8.4.1, principalmente por ser un sistema robusto y poseer licencia libre. Como herramienta de modelado Visual Paradigm 6.4. por ser es una herramienta CASE muy robusta, destacando además su usabilidad y portabilidad y como herramienta de administración PgAdmin 3.0.



Capítulo 2: Análisis y Diseño del Mercado de Datos.

Introducción.

En este capítulo se abordarán los aspectos pertenecientes al análisis y diseño de la solución, concretamente, a las características de las fuentes a integrar, definición de las áreas de análisis, análisis de datos, la arquitectura, el modelo dimensional propuesto, metadatos, tablas de hechos y tablas dimensionales.

También se hace un gran esfuerzo en los sistemas de almacenes de datos para que el trabajo realizado con los datos, no haga ningún tipo de daño o afectación a la integridad de los mismos, puesto que estos sistemas son actualizados con datos que pueden venir o proceder de fuentes muy diversas y variadas. Esta gran diversidad en las fuentes aumenta a un 100 % la probabilidad de que existan datos sucios o datos erróneos, es por eso que teniendo en cuenta que los almacenes de datos son utilizados para la toma de decisiones, se vuelve muy importante destacar que la calidad y claridad de sus datos es imprescindible para evitar las conclusiones equivocadas.

2.1. Análisis.

El análisis es la distinción y la separación de las partes de un todo hasta llegar a conocer sus principios o elementos. Hay varios tipos de análisis, uno de ellos es el análisis documental, que no es más que una forma de investigación técnica, un conjunto de operaciones intelectuales, que buscan describir y representar los documentos de forma unificada sistemática para facilitar su recuperación. Comprende el procesamiento analítico- sintético que, a su vez, incluye la descripción bibliográfica y general de la fuente, la clasificación, indización, anotación, extracción, traducción y la confección de reseñas. (Iglesias, 2007)

Por su parte el análisis de información, es una forma de investigación, cuyo objetivo es la captación, evaluación, selección y síntesis de los mensajes subyacentes en el contenido de los documentos, a partir del análisis de sus significados, a la luz de un problema determinado. Así, contribuye a la toma de decisiones, al cambio en el curso

de las acciones y de las estrategias. Es el instrumento por excelencia de la gestión de la información. (Iglesias, 2007)

2.1.1. Definición del negocio.

La ONE juega un papel de entidad rectora y coordinadora de los temas estadísticos en Cuba, por tal motivo va a funcionar como un repositorio central donde confluyen y vierten un conjunto de procesos que van a ejecutar y supervisar la gestión estadística del país. La misma integra su trabajo en un sistema eficiente conocido como Sistema Estadístico Nacional (SEN). Consta de una serie de funciones y atribuciones que se exponen a continuación:

- ✓ Organizar y aprobar la producción de las estadísticas centralizadas y territoriales.
- ✓ Dirigir metodológicamente la actividad estadística de los órganos, organismos, instituciones y entidades estatales, además de normar, velar y dictaminar acerca del funcionamiento de las estadísticas complementarias.
- ✓ Definir las atribuciones y responsabilidades de otros productores de estadísticas de interés nacional y su lugar en el sistema estadístico del país.
- ✓ Aprobar las normas metodológicas y de clasificación que se utilizan en las estadísticas centralizadas, territoriales y complementarias
- ✓ Identificar las unidades de observación estadística y captar, a través de la red territorial, la información correspondiente a las estadísticas centralizadas.
- ✓ Dirigir los procesos y ejecutar, según corresponda, los censos económicos y de población y encuestas económicas o sociales de carácter nacional. Aprobar la realización de este tipo de investigaciones estadísticas en el país.
- ✓ Llevar los registros estatales de empresas, unidades presupuestadas y otras entidades.
- ✓ Supervisar el trabajo estadístico de organismos y entidades, organizar la auditoría y comprobación estadísticas velando por la autenticidad de la información.
- ✓ Centralizar y emitir la estadística oficial del país.

El Modelo 0335 es el encargado de las estadísticas referentes al Comercio Agropecuario y el proceso de recogida y gestión de sus datos es el siguiente:

- ✓ El Modelo 0335 o Indicadores relacionados con el Comercio Agropecuario se recoge mensual por la Oficina Municipal de Estadísticas de las provincias. Aquí se conforma el modelo con los datos enviados por los organismos y se envía a la Oficina Provincial de Estadísticas.

- ✓ En la Oficina Provincial de Estadísticas se hace un resumen de los datos por municipio y se almacena en un fichero de extensión dbf y luego se colocan en un servidor, que tiene una carpeta definida para cada provincia.
- ✓ Este fichero llega a la ONE y se comienza con la limpieza de los datos por parte de los especialistas y se crea un archivo general con la información.
- ✓ Finalmente, la información se encuentra disponible mediante las publicaciones con que cuenta la ONE. En el caso del Modelo 0335 se tiene una publicación: mensual (resumen del año de los datos recogidos) y posee dos variantes: estatal y no estatal.

2.1.2. Temas de Análisis.

Las áreas de análisis son un punto importante en el desarrollo de los mercados de datos, pues estas van a marcar una gran diferencia entre las estructuras del mismo, buscando confiabilidad, fiabilidad, nuevas perspectivas y un buen uso. Dentro de la solución propuesta se acomodan dichas áreas de acuerdo con la situación de la organización, a la manera en que organizan la información, y respecto a sus necesidades:

- ✓ Inconsistencia de los datos almacenados en las estructuras propias de la ONE.
- ✓ Ineficiente manejo de datos de la institución para la obtención de información valiosa.
- ✓ Carencia de reportes flexibles con la información actualizada.

El área de análisis que se identifica es:

- ✓ Controles estadísticos de indicadores relacionados con el Mercado Agropecuario de la Actividad de Comercio para la ONE.

2.1.3 Roles y permisos.

Debido al intenso análisis del negocio realizado para definir quiénes son los que tienen acceso a la información contenida dentro de la BD que se implementará se definen dos roles “administrador y consultor”, y el nivel de acceso a los pedidos de información son públicos.

- ✓ Rol Administrador: es el encargado de introducir los datos de los indicadores en la BD, por lo que será capaz de incluir y ver la información.

- ✓ Rol Consultor: es la persona que accede a los datos para ver su contenido. No podrá modificar en forma alguna el contenido.

2.1.4. Reglas del Negocio.

A continuación se definen las reglas del negocio, que son las transformaciones que se le deben realizar a ciertos datos para obtener otros datos o los segmentos de un dato que pueden generar otros datos sin necesidad de ser introducidos por una persona. Especifican en un nivel adecuado de detalle lo que una organización debe hacer. Las reglas del negocio deben ser: declarativas, atómicas, independientes, distintas, expresadas en lenguaje natural y orientado al negocio.

- ✓ Precio Promedio = (Valor / Físico) * 100
- ✓ Valor corporativo = (U.B.P.C físico + C.PA físico) y (U.B.P.C valor + C.P.A valor)
- ✓ Valor no estatal = (U.B.P.C físico + C.P.A físico + Privado físico) y (U.B.P.C valor + C.P.A valor + Privado valor)

2.1.5 Necesidades de los usuarios.

El usuario necesita analizar la información del modelo 0335 "Indicadores del Comercio de la Actividad Agropecuaria" por provincias y municipios, por organismos de todos los datos que se solicitan en el modelo: Total(físico y valor), del total Organopónicos y huertos(físico y valor), Estatal(físico y valor), del estatal ventas de Acopo y Administrativas, U.B.P.C (físico y valor), C.P.A (físico y valor), E.J.T (físico y valor) y Privado (físico y valor).

2.2 Requisitos de información y multidimensionales.

Constituyen la entrada fundamental para el diseño de las estructuras del almacén. Parámetros de entrada y salida de las solicitudes de información de los clientes.

RI-1: Obtener el físico del total por centro informante y totalizado por los organismos, división político administrativa, fuente de productos e indicadores según el tiempo para la que se tiene información.

VE: Organismos, división político administrativa, temporal, centro informante, fuente de productos e indicadores.

VS: Físico del Total.

RI-2: Obtener el valor del total por centro informante y totalizado por los organismos, división político administrativa, fuente de productos e indicadores según el tiempo para la que se tiene información.

VE: Organismos, división político administrativa, temporal, centro informante, fuente de productos e indicadores.

VS: Valor del Total.

RI-3: Obtener el físico del total: Organopónicos y huertos por centro informante y totalizado por los organismos, división político administrativa, fuente de productos e indicadores según el tiempo para la que se tiene información.

VE: Organismos, división político administrativa, temporal, centro informante, fuente de productos e indicadores.

VS: Físico del total: Organopónicos y huertos.

RI-4: Obtener el valor del total: Organopónicos y huertos por centro informante y totalizado por los organismos, división político administrativa, fuente de productos e indicadores según el tiempo para la que se tiene información.

VE: Organismos, división político administrativa, temporal, centro informante, fuente de productos e indicadores.

VS: Valor del total: Organopónicos y huertos.

RI-5: Obtener el físico de Estatal por centro informante y totalizado por los organismos, división político administrativa, fuente de productos e indicadores según el tiempo para la que se tiene información.

VE: Organismos, división político administrativa, temporal, centro informante, fuente de productos e indicadores.

VS: Físico de Estatal.

RI-6: Obtener el valor de Estatal por centro informante y totalizado por los organismos, división político administrativa, fuente de productos e indicadores según el tiempo para la que se tiene información.

VE: Organismos, división político administrativa, temporal, centro informante, fuente de productos e indicadores.

VS: Valor de Estatal.

RI-7: Obtener el físico del estatal: Ventas de Acopio y Administrativas por centro informante y totalizado por los organismos, división político administrativa, fuente de productos e indicadores según el tiempo para la que se tiene información.

VE: Organismos, división político administrativa, temporal, centro informante, fuente de productos e indicadores.

VS: Físico del estatal: Ventas de Acopio y Administrativas.

RI-8: Obtener el valor del estatal: Ventas de Acopio y Administrativas por centro informante y totalizado por los organismos, división político administrativa, fuente de productos e indicadores según el tiempo para la que se tiene información.

VE: Organismos, división político administrativa, temporal, centro informante, fuente de productos e indicadores.

VS: Valor del estatal: Ventas de Acopio y Administrativas.

RI-9: Obtener el físico de U.B.P.C por centro informante y totalizado por los organismos, división político administrativa, fuente de productos e indicadores según el tiempo para la que se tiene información.

VE: Organismos, división político administrativa, temporal, centro informante, fuente de productos e indicadores.

VS: Físico de U.B.P.C

RI-10: Obtener el valor de U.B.P.C por centro informante y totalizado por los organismos, división político administrativa, fuente de productos e indicadores según el tiempo para la que se tiene información.

VE: Organismos, división político administrativa, temporal, centro informante, fuente de productos e indicadores.

VS: Valor de U.B.P.C

RI-11: Obtener el físico de C.P.A por centro informante y totalizado por los organismos, división político administrativa, fuente de productos e indicadores según el tiempo para la que se tiene información.

VE: Organismos, división político administrativa, temporal, centro informante, fuente de productos e indicadores.

VS: Físico de C.P.A

RI-12: Obtener el valor de C.P.A por centro informante y totalizado por los organismos, división político administrativa, fuente de productos e indicadores según el tiempo para la que se tiene información.

VE: Organismos, división político administrativa, temporal, centro informante, fuente de productos e indicadores.

VS: Valor de C.P.A

RI-13: Obtener el físico de E.J.T por centro informante y totalizado por los organismos, división político administrativa, fuente de productos e indicadores según el tiempo para la que se tiene información.

VE: Organismos, división político administrativa, temporal, centro informante, fuente de productos e indicadores.

VS: Físico de E.J.T

RI-14: Obtener el valor de E.J.T por centro informante y totalizado por los organismos, división político administrativa, fuente de productos e indicadores según el tiempo para la que se tiene información.

VE: Organismos, división político administrativa, temporal, centro informante, fuente de productos e indicadores.

VS: Valor de E.J.T

RI-15: Obtener el físico de Privado por centro informante y totalizado por los organismos, división político administrativa, fuente de productos e indicadores según el tiempo para la que se tiene información.

VE: Organismos, división político administrativa, temporal, centro informante, fuente de productos e indicadores.

VS: Físico de Privado.

RI-16: Obtener el valor de Privado por centro informante y totalizado por los organismos, división político administrativa, fuente de productos e indicadores según el tiempo para la que se tiene información.

VE: Organismos, división político administrativa, temporal, centro informante, fuente de productos e indicadores.

VS: Valor de Privado.

Requisitos Funcionales.

Son capacidades o condiciones que el sistema debe cumplir. Los requisitos funcionales permiten expresar específicamente las responsabilidades del sistema que se propone y permiten determinar de una manera clara, lo que el sistema debe hacer.

- ✓ Extraer información del DBF de los indicadores relacionados con el Mercado Agropecuario en la Actividad de Comercio.
- ✓ Transformar y cargar información del DBF de los indicadores relacionados con

el Mercado Agropecuario en la Actividad de Comercio.

Requisitos no funcionales.

Los requerimientos no funcionales son propiedades o cualidades que el producto debe tener. Debe pensarse como las características que hacen al producto atractivo, usable, rápido o confiable, están vinculados a requerimientos funcionales, es decir, una vez se conozca lo que el sistema debe hacer se puede determinar cómo ha de comportarse, qué cualidades debe tener o cuán rápido o grande debe ser.

Fiabilidad

Describe cuán fiables pueden ser los datos que se integran en el sistema en cuanto a:

- ✓ Disponibilidad.
- ✓ Tiempo medio de fallos.
- ✓ Tiempo medio de reparación.
- ✓ Máximo de errores.
- ✓ Exactitud.

Requisitos de fiabilidad

Disponibilidad

- ✓ El área de almacenamiento del proceso de ETL debe estar disponible 12 horas del día.
- ✓ El mantenimiento se realizará 1 vez por semana, realizándose chequeos para establecer los errores que existan y, de ser posible, arreglarlos. Esto se debe realizar preferentemente a finales de la semana.
- ✓ El sistema debe tener disponibilidad al 100% entre las 8:00 am y las 5:00 pm en los días laborables (lunes a viernes).

Tiempo medio de reparación

- ✓ El tiempo medio de reparación estimado es aproximadamente de 1 día luego del fallo.

Máximo de errores

- ✓ Por ser el proceso de integración un punto crítico para la calidad y exactitud de los datos almacenados se establece un máximo de errores de 0 errores/puntos de función.

Exactitud

- ✓ La salida de los datos debe ser 100% exacta, ya que de ahí dependen las decisiones que se tomen basándose en estos datos.

Eficiencia

La eficiencia del sistema es la que determina la calidad del mismo.

Requisitos de eficiencia

Tiempo de respuesta

- ✓ El sistema debe tener un tiempo de respuesta aproximado de 5 segundos, no debe excederse de 1 minuto.

Capacidad

- ✓ Durante la integración solo se conectará un usuario que deberá monitorear el proceso de integración.
- ✓ El sistema deberá permitir varios usuarios (mín. de 5) conectados sin que esto afecte la respuesta de las consultas.

Restricciones de diseño

- ✓ El lenguaje para realizar las consultas a la BD será SQL.

Componentes Comprados

- ✓ La herramienta CASE Visual Paradigm 6.4 no es un software libre pero se cuenta con la licencia de este programa.

Interfaz

Los requerimientos de interfaz son aquellos que estarán en contacto directo con el usuario, por lo tanto, incluyen el aspecto del sistema y el contenido que verá el usuario.

Interfaces de usuario

- ✓ El sistema debe tener una interfaz amigable y sencilla de utilizar, teniendo en cuenta que los usuarios finales no son personas adiestradas en el campo de la informática.
- ✓ Las interfaces de salida no deben tener otra información que no sea concerniente a lo que deben mostrar.

Interfaces Hardware

- ✓ Para el proceso de transformación es necesaria una memoria RAM de 512 MB como mínimo.
- ✓ Para el proceso de visualización e inteligencia de negocio se necesita una memoria RAM de 1 GB como mínimo, para garantizar el correcto funcionamiento del sistema cuando es accedido por varios usuarios simultáneamente.
- ✓ Se necesita un mínimo de 60 GB de capacidad del disco duro para el almacenamiento de la información.
- ✓ Se debe garantizar al menos una impresora para imprimir los reportes de salida.

Interfaces Software

- ✓ El lenguaje para realizar las consultas y la programación dentro del repositorio será el SQL.
- ✓ Se debe disponer de un navegador común, preferentemente asociado a un sistema operativo, para garantizar la visualización de las interfaces web de los reportes.
- ✓ La BD se montará con el gestor PostgreSQL 8.4.

Interfaces de Comunicación

- ✓ El sistema debe estar conectado a un dispositivo de red.
- ✓ La comunicación entre el almacén de datos y la BD de la integración será a través del protocolo TCP/IP.

Requisitos de Licencia

- ✓ El Visual Paradigm 6.4 es la única herramienta que no es libre, pero se cuenta con su licencia.

2.2.1 Casos de uso del sistema.

- ✓ Extraer información del DBF de los indicadores relacionados con el Mercado Agropecuario en la Actividad de Comercio.
- ✓ Transformar y cargar información del DBF de los indicadores relacionados con el Mercado Agropecuario en la Actividad de Comercio.
- ✓ Obtener información de los indicadores relacionados con el Mercado Agropecuario en la Actividad de Comercio.

2.3 Diseño.

En la solución se ha definido un mercado de datos donde convergen todas las dimensiones propuestas: Fuentes de productos, Temporal, Centro Informante, Indicadores, Organismos y DPA. Los hechos mensurables del mercado de datos, son los aspectos que se evalúan en el modelo estadístico en cuestión, en este caso se cuenta con un solo hecho: Indicadores relacionados con el Comercio Agropecuario.

2.3.1 Matriz BUS.

La Matriz BUS representa las relaciones existentes entre los hechos y las dimensiones del sistema diseñado. Para los modelos propuestos existen dimensiones comunes y específicas de cada modelo que se relacionan directamente con la tabla de hecho existente (ver tabla 2).

Hechos/dimensiones	Dim_temporal	Dim_centro_informante	Dim_indicadores	Dim_organismo	Dim_fuentes_producto	Dim_dpa
hec_indicadores_comercio_agropecuario	X	X	X	X	X	X

Tabla 2 Matriz Buz

2.3.2 Modelos de Datos.

El modelo de datos está integrado por todo lo que compone el modelo dimensional, su enfoque viene precedido por el esclarecimiento de las dimensiones, medidas, granularidad y hechos identificados, de esta manera, se procede a la estructuración del modelo dimensional que tendrá el sistema. Subrayando que por las necesidades actuales del negocio existen varios modelos que relacionan las dimensiones definidas y las medidas que se han detallado hasta el momento.

2.3.3 Dimensiones.

Detrás de haber declarado el grano del proceso a modelar se emprende la definición de las dimensiones candidatas que posteriormente, después de un hondo análisis, se convertirán en las dimensiones que contendrá la solución. Las dimensiones poseen entre sus características principales la definición de jerarquías entre sus atributos las que poseen como objetivo plasmar explícitamente la forma en que se puede consolidar. Las tablas de dimensiones son las compañeras integrales de las tablas de

hechos, ellas contienen la descripción textual del negocio. En el modelo dimensional, las tablas de dimensiones poseen varios atributos que en su conjunto definen una fila en la tabla de dimensión. Los atributos de las dimensiones sirven como fuente primaria de las restricciones de las consultas, agrupaciones y las etiquetas de los reportes. Ellos desempeñan un rol de vital importancia dentro del Almacén de Datos debido a que son las llaves que hacen el almacén de datos usable y entendible. Estos atributos son las llaves de entrada a los hechos o medidas almacenadas. La calidad de todo Almacén de Datos se mide por la definición de los atributos de las dimensiones. Su poder es directamente proporcional a la calidad y profundidad de estos atributos. (Kimball, 2002)

A continuación se describen las dimensiones y jerarquías que están relacionadas con el repositorio principal donde se va a almacenar la información.

1. Dimensión Temporal.

Esta es la dimensión más común e importante en los diseños de Mercados de Datos debido a que define una línea de tiempo por la que se regirá la investigación.

Jerarquía:

1. Anual-> Semestral->Trimestral->Mensual.

2. Dimensión Indicadores.

Esta dimensión describe el universo de valores bajo los cuales puede clasificarse la información atendiendo a los indicadores que se recogen en el modelo.

Jerarquía:

1. Temática.
2. Indicador.
 - Descripción del indicador.

3. Dimensión Organismo.

Esta dimensión describe el universo de valores bajo los cuales puede clasificarse la información atendiendo a los organismos.

4. Dimensión División Político Administrativa.

Esta dimensión describe el universo de valores bajo los cuales puede clasificarse la información atendiendo a la situación geográfica.

Jerarquía:

1. Provincia.
 - Provincia Código.
2. Municipio.

- ✓ Municipio Código.

5. Dimensión Centro Informante

Esta dimensión describe el universo de valores bajo los cuales puede clasificarse la información atendiendo a los centros informantes.

6. Dimensión Fuente de productos.

Esta dimensión describe el universo de valores bajo los cuales puede clasificarse la información atendiendo a las fuentes de productos.

2.3.4 Tablas de hechos y medidas asociadas.

Las tablas de hechos son las que almacenan las medidas numéricas. Son tablas primarias en el modelo dimensional, ya que almacenan los valores del negocio, cada tabla representa una relación de muchos a muchos, y contiene dos o más llaves extranjeras que se enlazan con sus respectivas tablas dimensiones.

Dada este tipo de información al almacenar, cada tabla de hecho generalmente define un mercado de datos determinado. Cada fila de esta tabla corresponde a un hecho determinado y a su vez cada conjunto de hechos dentro de la tabla de hechos referencian a la misma granularidad. La principal condición que debe cumplir las tablas de hechos es que el hecho debe almacenarse de tal forma que su valor sea numérico y a su vez sea aditivo para así poder realizar cálculos sobre él, ya sea por ciento, sumas, igualdades, etc. (Ponniiah., 2001)

A continuación se muestran la tabla de hecho que servirá de repositorio central a diferentes dimensiones.

Tabla de hecho Indicadores relacionados con el Comercio Agropecuario: donde se recoge todos los indicadores del Comercio Agropecuario.

Medidas

Las medidas numéricas definidas se almacenarán en la tabla de hecho, los extraídos de cada temática aparecen a continuación.

1. Total, físico: del total se obtendrá el físico de total
2. Total, valor: del total se obtendrá el valor de total.
3. De Total: Organopónicos y huertos, físico.

4. De Total: Organopónicos y huertos, valor.
5. Estatal, físico: se obtendrá el físico de estatal
6. Estatal, físico: se obtendrá el valor de estatal.
7. De Estatal: Ventas de Acopio y Administrativas, físico.
8. De Estatal: Ventas de Acopio y Administrativas, valor.
9. U.B.P.C, físico: se obtendrá el físico de U.B.P.C.
10. U.B.P.C, valor: se obtendrá el valor de U.B.P.C
11. C.P.A, físico: se obtendrá el físico de C.P.A.
12. C.P.A, valor: se obtendrá el valor de C.P.A.
13. E.J.T, físico: se obtendrá el físico de E.J.T.
14. E.J.T, físico: se obtendrá el físico de E.J.T.
15. Privado, físico: se obtendrá el físico de Privado.
16. Privado, valor: se obtendrá el valor de Privado.

2.3.5 Esquema de seguridad.

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

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

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

2.3.6 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 salvas: Mensualmente se realizan las salvas 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 de los indicadores relacionados con el Comercio
- ✓ Backups existentes: Actualmente no existen Backups en esta área.
 - Periodicidad de reemplazo de los Backups: Se realizan los reemplazos de Backups cada 1 año.
 - Periódicas de las pruebas a los Backups: El estado de los Backups se chequea mensualmente, mediante pruebas de rendimiento y flexibilidad.

Conclusiones.

En este capítulo se detallaron los aspectos principales dentro del proceso de análisis y diseño de la solución. Se definieron además las políticas de respaldo y recuperación y el esquema de seguridad. Concluyéndose los siguientes resultados:

- ✓ El trabajo con los códigos utilizados en la ONE garantiza la concordancia con las políticas de esta organización.
- ✓ Ciclo del Proceso del Negocio.
- ✓ Se definieron los roles Administrador y Consultor con sus respectivos permisos.
- ✓ Se definió el diseño lógico.
- ✓ Áreas de análisis: Comercio Agropecuario de la ONE.
- ✓ Se definieron los requisitos de información, los funcionales y los no funcionales.
- ✓ Se definieron 6 dimensiones y 1 hecho, los cuales cumplen con lo establecido por el cliente y satisfacen las necesidades del mismo.
- ✓ Se definieron 16 medidas, 3 medidas calculables y políticas de seguridad a seguir en el sistema.
- ✓ La arquitectura propuesta satisface el problema que se había planteado por parte de la ONE.
- ✓ La utilización de las políticas de respaldo garantizan la integridad y disponibilidad de los datos del mercado.



Capítulo 3: Implementación y Validación de la solución.

Introducción.

En este capítulo se contiene todos los elementos referentes a la implementación del sistema, las tablas que formarán parte de la posible solución, la descripción de los roles y permisos creados y los pasos para la implantación del sistema. También se incluirán los resultados de las pruebas ejecutadas al sistema.

3.1. Modelo de Datos Físico.

Los modelos de datos se pueden clasificar dependiendo de los tipos de conceptos que ofrecen para describir la estructura de la BD. Los modelos de datos de alto nivel, o modelos conceptuales, disponen de conceptos muy cercanos al modo en que la mayoría de los usuarios percibe los datos, mientras que los modelos de datos de bajo nivel, o modelos físicos, proporcionan conceptos que describen los detalles de cómo se almacenan los datos en el ordenador. Los conceptos de los modelos físicos están dirigidos al personal informático, no a los usuarios finales. Entre estos dos extremos se encuentran los modelos lógicos, cuyos conceptos pueden ser entendidos por los usuarios finales, aunque no están demasiado alejados de la forma en que los datos se organizan físicamente. Los modelos físicos poseen tablas de mantenimiento que habitualmente no se incluyen en el modelo lógico. Difieren mayormente en la especificación absoluta y puntualizada de las características físicas de la BD, comenzando por los tipos de datos hasta las tablas de segmentación, parámetros de almacenamiento de tablas y bandas de discos. Después que se ha cargado el modelo físico en la herramienta se comienzan las tareas de personalización de las estructuras físicas en función de la estandarización de formatos, nombres de objetos, la corrección de relaciones, y así sucesivamente con cada una de las tareas que permiten dejar a punto las estructuras dimensionales para ser desplegada en el SGBD (Sistema Gestor de Base Datos). En este sentido, en el proceso actual, se corrige la utilización de determinados nombres de atributos de dimensiones, tipos de datos, cantidad de caracteres que soporta, y la documentación de cada estructura y atributo, con el objetivo de que se concrete el modelo de datos.

3.1.1. Estructuras de Datos.

Esquemas:

Los esquemas no son más que formas de organización de la BD que pueden contener tablas, tipos de datos, funciones y operadores. A diferencia de la BD, los esquemas no están apartados entre sí, permitiéndole a un usuario acceder a cualquiera de ellos, si tiene los permisos adecuados. El uso de esquemas trae consigo un grupo de mejoras, las cuales son mencionadas a continuación:

- ✓ Organizan los objetos de la BD en grupos lógicos, facilitando su manejo.
- ✓ Permiten el uso de la BD por parte de muchos usuarios sin que estos interfieran entre sí.
- ✓ Permiten utilizar el mismo nombre para un objeto en esquemas diferentes sin ocasionar un conflicto en la BD.
- ✓ Organizan los objetos de la BD en grupos lógicos, facilitando su manejo.
- ✓ Permiten utilizar el mismo nombre para un objeto en esquemas diferentes sin ocasionar un conflicto en la BD.

Para el desarrollo del sistema propuesto en esta investigación se definieron 2 esquemas:

- ✓ Esq_dimensiones: contiene todas las tablas de dimensiones propuestas.
- ✓ Esq_hechos: contiene todas las tablas de hechos propuestas.

Tablas:

La solución cuenta con 7 tablas en total, 6 dimensiones y 1 hecho, distribuidas por los dos esquemas propuestos con anterioridad quedando de la siguiente manera:

Tabla	Esquema	Propietario	Descripción
dim_dpa	dimensión	Postgres	Contiene los datos de la División Político-Administrativa del país.
dim_indicadores	dimensión	Postgres	Contiene los datos de los indicadores concernientes a

			los productos del comercio agropecuario.
dim_organismo	dimensión	Postgres	Contiene los datos de los organismos: MINCIN, MINAG y MINAZ
dim_fuentes_productos	dimensión	Postgres	Contiene los datos de las fuentes de productos: U.B.P.C; C.P.A; E.J.T.
dim_temporal	dimensión	Postgres	Contiene los rangos de tiempo en que se recogen los datos.
dim_centro_informante	dimensión	Postgres	Contiene los datos de los centros informantes.
hec_indicadores_relacionados_comercio_agropecuario.	hechos	Postgres	Contiene las medidas concernientes a los productos del comercio agropecuario.

Tabla 3 Tablas creadas para la solución

Restricciones:

Las restricciones son condiciones que se le aplican a una BD para que cumpla con ciertos parámetros. Las mismas pueden ser creadas automáticamente al definir una tabla (en el caso de las llaves primarias) o ser introducidas por el programador de la BD cuando se busca algo específico, por ejemplo, que un campo tenga una longitud determinada o tenga la estructura de un ciclo. Existen 4 tipos de restricciones que son las más comunes:

- ✓ Clave foránea: Son las que referencian una clave de otra tabla, se usan para relacionar tablas diferentes.
- ✓ De unicidad: Implica que no debe haber 2 valores iguales en la misma columna.

- ✓ Clave primaria: Son valores que deben cumplir con un conjunto de restricciones: no tener valores nulos, ser únicos para cada tupla y ser necesarios (mínima).
- ✓ Valor no nulo: No debe existir ninguna casilla de la columna que esté vacía.

A continuación se especifican las restricciones tomadas en cuenta para la solución planteada.

Tabla DPA:

- ✓ Clave primaria: id_dpa.
- ✓ Secuencia: seq_dpa_id (define la secuencia de números, comenzando por el 1 y terminando en el 99999, que conformarán el id de la tabla DPA y se generará automáticamente).

Tabla Temporal:

- ✓ Clave primaria: temporal_id.
- ✓ Secuencia: seq_temporal_id (define la secuencia de números, comenzando por el 1 y terminando en el 99999, que conformarán el id de la tabla Temporal y se generará automáticamente).

Tabla Fuentes_prod:

- ✓ Clave primaria: id_fuentes_prod.
- ✓ Secuencia: seq_fuentes_prod_id (define la secuencia de números, comenzando por el 1 y terminando en el 99999, que conformarán el id de la tabla Fuentes_prod y se generará automáticamente).

Tabla Indicadores

- ✓ Clave primaria: id_indicador
- ✓ Secuencia: seq_indicador_id (define la secuencia de números, comenzando por el 1 y terminando en el 99999, que conformarán el id de la tabla Indicadores y se generará automáticamente).

Tabla Organismos:

- ✓ Clave primaria: id_organismo.

- ✓ Secuencia: seq_organismo_id (define la secuencia de números, comenzando por el 1 y terminando en el 99999, que conformarán el id de la tabla Organismos y se generará automáticamente).

Tabla Centro informantes:

- ✓ Clave primaria: id_centro_informante.
- ✓ Secuencia: seq_centro_informante_id (define la secuencia de números, comenzando por el 1 y terminando en el 99999, que conformarán el id de la tabla Centro_informantes y se generará automáticamente).

Tabla indicadores relacionados comercio agropecuario:

- ✓ Clave primaria: Conformada por el conjunto de llaves foráneas que contiene: id_dpa, temporal_id, id_centro_informante, id_fuentes_prod, id_organismo, id_indicador, id_indicador_relacionados_comercio_agropecuario.
- ✓ Claves foráneas:
 - id_dpa.
 - temporal_id.
 - id_centro_informante.
 - id_fuentes_prod.
 - id_organismo.
 - id_indicador.
 - id_indicador_relacionados_comercio_agropecuario.

Índices

Los índices dentro de la BD cumplen una función muy similar a los índices que se encuentran en los libros, estos van a acelerar el proceso de búsqueda. La utilización de los índices trae consigo ventajas tales como, evitar realizar un recorrido completo de la tabla en la que se busca la información. Así se mejora el rendimiento de la BD y la utilización de los recursos de la computadora, sobre todo en casos en los que las tablas tengan un gran tamaño. Con esto se logra además obtener tiempos de respuestas más rápidos en las consultas. La estrategia de indexado seleccionada para la solución es por claves primarias.

Índice	Tabla	Esquema	Tipo	Campos
PK6	dim_dpa	dimensión	btree	id_dpa
PK1	dim_indicador	dimensión	btree	id_indicador
PK5	dim_fuentes_prod	dimensión	btree	id_fuentes_prod
PK4	dim_temporal	dimensión	btree	temporal_id
PK2	dim_centro_informante	dimensión	btree	id_centro_informante
PK3	dim_organismo	dimensión	btree	id_organismo
PK7	hec_indicadores_relacionados_comercio_agropecuario	hechos	btree	id_dpa, temporal_id, id_indicador, id_fuentes_prod, id_centro_informante, id_organismo, id_indicador_relacionados_comercio_agropecuario

Tabla 4 Ejemplos de los índices seleccionados

3.1.2. Usuarios y privilegios

Usuarios y roles

Los usuarios definidos son:

- ✓ Rol Administrador: es el encargado de introducir los datos de los indicadores en la BD, por lo que será capaz de incluir y ver la información.
- ✓ Rol Consultor: es la persona que accede a los datos para ver su contenido. No podrá modificar en forma alguna el contenido.

Privilegios

Los privilegios establecidos se aplican a los objetos de tipo tabla y esquema. Ver en el *Anexo 2* un ejemplo del código del script que guarda los elementos de seguridad definidos.

Usuario	Permisos	Descripción
Postgres	Todos los permisos de la	Hace uso completo de la

	BD.	BD.
Administrador	Tiene permiso para extraer, transformar y cargar los datos de la BD.	Visualiza, extrae, inserta y actualiza los datos de la BD
Consultor	Lectura de los datos de la BD	Visualiza los datos de la BD

Tabla 5 Usuarios y privilegios establecidos

3.1.3. Carga de nomencladores

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

Para completar la solución se corresponde realizar el proceso de carga de los nomencladores que se recogen en el Modelo 0335, ellos son:

1. Nomenclador DPA. (ver anexo 7)
2. Nomenclador Centro Informante. (ver anexo 8)
3. Nomenclador Fuentes de productos. (ver anexo 9)
4. Nomenclador Indicadores. (ver anexo 10)
5. Nomenclador Organismos. (ver anexo 11)
6. Nomenclador Temporal. (ver anexo 12)

También, en el *Anexo 3* se pueden ver un ejemplo de los scripts utilizados para crear las tablas.

3.2. Guía de implantación

Para poder darle mantenimiento a la solución del presente trabajo es necesario crear una guía de implantación donde se analicen los detalles de tipo técnico y organizativo.

3.2.1. Requerimientos

- ✓ Software

- Se debe disponer de un navegador común, preferentemente asociado a un sistema operativo, para garantizar la visualización de las interfaces web de los reportes.
- La BD se montará con el gestor PostgreSQL 8.4.
- El lenguaje para realizar las consultas y la programación dentro del repositorio será el SQL.
- Se debe contar con la herramienta Mondrian instalada.
- Debe instalarse JDK\JRE 1.5 para poder utilizar Mondrian.

✓ Hardware

- Para el proceso de transformación es necesaria una memoria RAM de 512 MB como mínimo.
- Para el proceso de visualización e inteligencia de negocio se necesita una memoria RAM de 1 GB como mínimo, para garantizar el correcto funcionamiento del sistema cuando es accedido por varios usuarios simultáneamente.
- Se necesita un mínimo de 60 GB de capacidad del disco duro para el almacenamiento de la información.
- Se debe garantizar al menos una impresora para imprimir los reportes de salida.
- El sistema debe estar conectado directamente a la red.

3.2.2. Secuencia de pasos para la instalación de la BD

A continuación se describen los pasos a seguir para la instalación de la BD y su correcto funcionamiento.

1. Posterior a la instalación del PostgreSQL 8.4, se debe proceder a la creación de una nueva BD siguiendo los pasos:
 - a. Seleccionar BD/Editar/Nuevo objeto/Nueva BD.
 - b. Introducir los datos que se piden y crear la BD.
2. Seleccionar la BD creada e ir a la Herramienta de Consultas del PostgreSQL. Cargar el script denominado *dim ddl CAG.sql* y ejecutarlo. Esto creará el esquema "dimensión" con las tablas correspondientes.

3. Seleccionar la BD creada nuevamente e ir a la Herramienta de Consultas. Cargar el script denominado *hecho ddl CAG.sql* y ejecutarlo. Esto creará el esquema “hechos” con las tablas correspondientes.
4. Seleccionar la BD creada nuevamente e ir a la Herramienta de Consultas. Cargar el script denominado *dim dml CAG.sql* y ejecutarlo. Esto cargará los nomencladores de las tablas del esquema “dimensión”.
5. Seleccionar la BD creada nuevamente e ir a la Herramienta de Consultas. Cargar el script denominado *dcl DCL.sql* y ejecutarlo. Esto cargará los usuarios creados y establecerá los privilegios de éstos sobre los objetos de la BD.

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

3.3. Validación y pruebas

La validación y las pruebas que se llevan a cabo a un producto, se realizan para garantizar el cumplimiento de la exigencia del cliente y la calidad de la solución.

3.3.1. Lista de chequeo de análisis

Las listas de chequeo se ejecutan con el objetivo de garantizar que los artefactos que se fabriquen en el análisis puedan cumplir con un formato determinado y que toda la información recogida esté organizada según lo establecido. La lista es informativa y pretende cubrir todos los aspectos aplicables en los temas de análisis, particularmente para el área de Comercio, específicamente el departamento de Agropecuario de la ONE.

- Lista de Chequeo Especificación de Requisitos.
- Lista de Chequeo Especificación de las áreas de la organización.
- Lista de Chequeo Herramienta para la recolección y análisis de la información.

3.3.2. Lista de chequeo de diseño.

Esta lista de chequeo su objetivo es garantizar que los artefactos elaborados en el diseño y el modelo de datos cumplan con el formato que se tiene establecido. La lista es informativa y pretende cubrir todos los aspectos aplicables en los temas de diseño e implementación, particularmente para el área de Comercio, específicamente el departamento de Agropecuario de la ONE.

- Lista de Chequeo Modelo de Datos.

Los resultados después de la aplicación de las listas de chequeo a sus respectivos artefactos son los siguientes:

Cantidad de indicadores evaluados	Total de indicadores	Cantidad de indicadores evaluados de correctos	Cantidad de indicadores evaluados de incorrectos	Cantidad de indicadores que No Proceden	Evaluación del artefacto
Lista de chequeo Evaluación de las áreas de la organización	10	9	-	1	Bien
Lista de chequeo de la Herramienta para la recolección y análisis de la información	28	28	-	-	Bien
Lista de chequeo Especificación de Requisitos	30	30	-	-	Bien

Tabla 6 Resultados de las Listas de Chequeo de análisis aplicadas.

3.3.3. Validación de requisitos por el cliente.

Se ejecutó un encuentro con el cliente Elena Leonila Fernández García. Representante de la ONE en la UCI donde se mostraron los requisitos definidos así como el diseño del Modelo de Datos. Se discutieron los resultados obtenidos hasta el momento y el cliente expresó su aprobación con respecto a los requisitos y al Modelo de Datos.

3.3.4. Pruebas de implantación

Clases Válidas	Resultado Esperado	Resultado de la Prueba	Observaciones
Creación de nueva BD.	Nueva BD vacía y lista para recibir los datos.	Nueva BD vacía y lista para recibir los datos.	

Carga del script <i>dim ddl CAG.sql</i>	Creación del esquema "dimensión" con las tablas correspondientes.	Creación del esquema "dimensión" con las tablas correspondientes.	Las tablas dentro del esquema "dimensión" son: <ul style="list-style-type: none"> - Dim_dpa - Dim_temporal - Dim_fuentes_prod - Dim_indicadores - Dim_organismos - Dim_centro_informante
Carga del script <i>hecho ddl CAG.sql</i>	Creación del esquema "hechos" con las tablas correspondientes.	Creación del esquema "hechos" con las tablas correspondientes.	Las tablas dentro del esquema "hechos" son: <ul style="list-style-type: none"> - hecho_indicadores_relacionados_Comercio_Agropecuario
Carga del script <i>dim dml CAG.sql</i>	Carga de los nomencladores de las tablas del esquema "dimensión"	Carga de los nomencladores de las tablas del esquema "dimensión"	Los nomencladores se cargan sin presentar problema alguno.
Carga del script <i>dcl DCL.sql</i>	Establecimiento de los privilegios de los usuarios sobre los objetos de la BD.	Establecimiento de los privilegios de los usuarios sobre los objetos de la BD.	Deben otorgarse los permisos: ADMINISTRADOR (con permisos para visualizar, seleccionar, insertar, actualizar datos de las tablas) y CONSULTOR (con permisos para visualizar y seleccionar datos de las tablas).

Tabla 7 Pasos de implantación a los que se les realizaron las pruebas

Conclusiones

En este capítulo se describió el Modelo de Datos Físico, los pasos para la implantación del sistema y el proceso de validación y pruebas realizado. Concluyéndose que:

- ✓ Se figuró el modelo físico del mercado de datos.
- ✓ Se detallaron los 2 esquemas existentes relacionados a las tablas: hechos y dimensiones, y los índices de la estructura de datos, brindando así una correcta organización de los datos de la BD.
- ✓ Los usuarios y sus privilegios cumplen con lo establecido en los requisitos iniciales del sistema.
- ✓ Se detallaron los nomencladores especificando el tipo de información que manejan.
- ✓ Se crea la guía de implantación para la BD en PostgreSQL 8.4 y la herramienta de administración PgAdmin 3.0.
- ✓ Creación de la BD y roles con sus respectivos permisos.
- ✓ Los nomencladores cargados responden a lo especificado por el cliente.
- ✓ Equivalente se realizan las siguientes pruebas: lista de Chequeo Especificación de Requisitos, lista de Chequeo Especificación de las áreas de la organización, lista de Chequeo Herramienta para la recolección y análisis de la información y lista de Chequeo Modelo de Datos, verificándose el correcto funcionamiento de los pasos de implantación del sistema y obteniéndose resultados satisfactorios, cumpliendo así con las exigencias del cliente.

Conclusiones Generales

Con la culminación de este trabajo de diploma y tras realizar un amplio análisis de todo el tema relacionado con el Comercio Agropecuario se obtuvieron los resultados que se detallan a continuación:

- ✓ Se realizó un exhaustivo estudio del comportamiento de los mercados de datos a nivel mundial y nacional, y de la importancia que tiene la utilización de los almacenes de datos para ayudar en la toma de decisiones a partir de la información obtenida de un sector económico, obteniéndose un conjunto de información imprescindible para trazar las pautas hacia donde se debía encaminar el proyecto.
- ✓ La metodología utilizada en el trabajo, es una adaptación de la metodología de Kimball, desarrollada por DATEC, que cumple con los requerimientos necesarios para la implementación del Mercado de Datos.
- ✓ Se seleccionaron las herramientas necesarias para realizar el proyecto con la calidad requerida y que estuvieran a tono con las disposiciones existentes, siempre que fuera posible, sobre el uso de software libre y multiplataforma.
- ✓ Las estructuras dimensionales modeladas (tablas de hechos y dimensiones) abarcan la forma de organización del Modelo 0335 o de Indicadores relacionados con el Comercio Agropecuario y soportan el proceso de toma de decisiones.
- ✓ Los nomencladores cargados en la BD poseen la estructura especificada por el cliente.
- ✓ Se obtuvo una aplicación que permitirá a la Oficina Nacional de Estadísticas contar con una información estructurada y disponible para la toma de decisiones y a la cual se le practicaron pruebas estadísticas que permitieron un mayor análisis de los resultados, siendo los mismos satisfactorios.

Recomendaciones.

- ✓ Completar la construcción del Mercado de Datos para los Indicadores del Comercio Agropecuario e incorporarlo al Almacén de Datos de la ONE.
- ✓ Tomar el contenido de este trabajo como referencia para futuros sistemas de mercados y almacenes de datos que se desarrollen en el país.
- ✓ Realizar el proceso de extracción, transformación y carga de los datos para mostrar las estadísticas del Comercio Agropecuario.
- ✓ Extender el uso de la aplicación a otras instituciones que recopilen y organicen los datos estadísticos de los Indicadores relacionados con el Comercio Agropecuario.

Referencias

- Bernabeu., R. D. Hefesto. 2007.** *Metodología propia para la construcción de un Data Warehouse.* 2007.
- 2010.** Definición de indicadores-Calidad en ONG-. [En línea] 2010. [Citado el: 4 de Febrero de 2010.]
- 2009.** *Herramientas CASE.* 2009.
- Matos, R.M. 2001.** *Introducción al trabajo con Base de Datos. Asignatura de Sistemas de Gestión de Base de Datos.* 2001.
- Mavilio., Alfredo Sifontes.** *Diseño del Datamart del subsistema de Conducidos.*
- Rayner Huamantumba. 2007.** DataMart paso a paso. [En línea] 31 de Agosto de 2007. [Citado el: 14 de Febrero de 2010.] WWW.RUEDATECNOLOGICA.COM.
- Reeves, R. Kimball and L. 2008.** *The data warehouse lifecycle toolkit.* s.l. : Jon Wiley & Sons, 2008.
- 2007.** Sistemas Gestores de bases de datos. [En línea] 31 de Julio de 2007. [Citado el: 14 de Febrero de 2010.]
- Kimball, Ralph y Otros. 2002.** *The Data Warehouse Lifecycle Toolkit.* 2002.
- Ponniah. 2008.** *Data Warehousing Fundamentals.* 2008.
- Pacheco, Henry Jesus Mendoza. 2008.** *Tecniciencia.* 2008.
- S.L, Digitales. 2007-2010.** Portal en español. 2007-2010.
- 2010.** Technologies. 2010.
- Varios. 2006-2010.** ETL-Tools.Info. 2006-2010.
- . **2009.** Historia de la Informática en Cuba. 2009.
- Velasco, Roberto Hernando. 2007.** Almacén de Datos. 2007.
- Iglesias, Lic. María Elinor Dulzaides. 2007.** *Análisis.* 2007.
- Pérez, P. V. (2009).

Bibliografías

- Arias, Aliuska Sánchez Ibarria y Bettys Leydis. 2007.** *Sistema de análisis estadísticos de datos.* Habana : s.n., 2007.
- 2007.** *CARACTERÍSTICAS Y EVOLUCIÓN.* Habana : s.n., 2007.
- 2007.** *Curso Universitario-Sistemas de información en los sistemas de Salud.* Buenos Aires, Argentina : s.n., 2007.
- 1998.** Guía para la definición de indicadores. [En línea] Marzo de 1998.
- Mavilio., Alfredo Sifontes.** *Diseño del Datamart del subsistema de Conducidos.*
- **2002.** *Indicadores para la autoevaluación de instituciones de formación profesional inicial.* Bellaterra : s.n., 2002.
- 2008.** *Ofinas de control interno del sector Comercio, Industria y Turismo.* Bogota, D.C. : s.n., 2008.
- Ponniah. 2001.** Tablas de hechos y dimensiones. 2001.
- ArPug. 2009.** 2009.
- Copyright. 2005-2010.** Corporation Pentaho. 2005-2010.
- no.3, Ing. Investig. vol.26. Sept. 2007.** Ingeniería e Investigación. Bogotá : s.n., Sept. 2007.
- 2005-2010.** Oracle Corporation. 2005-2010.
- Pacheco, Henry Jesus Mendoza. 2008.** *Tecniciencia.* 2008.
- S.L, Digitales. 2007-2010.** Portal en español. 2007-2010.
- 2010.** Technologies. 2010.
- Varios. 2006-2010.** ETL-Tools.Info. 2006-2010.
- **2009.** Historia de la Informática en Cuba. 2009.
- Velasco, Roberto Hernando. 2007.** Almacén de Datos. 2007.
- Bernabeu., R. D. Hefesto. 2007.** *Metodología propia para la construcción de un Data Warehouse.* 2007.
- 2010.** Definición de indicadores-Calidad en ONG-. [En línea] 2010. [Citado el: 4 de Febrero de 2010.]
- 2009.** *Herramientas CASE.* 2009.
- Matos, R.M. 2001.** *Introducción al trabajo con Base de Datos. Asignatura de Sistemas de Gestión de Base de Datos.* 2001.
- Mavilio., Alfredo Sifontes.** *Diseño del Datamart del subsistema de Conducidos.*
- Rayner Huamantumba. 2007.** DataMart paso a paso. [En línea] 31 de Agosto de 2007. [Citado el: 14 de Febrero de 2010.] WWW.RUEDATECNOLOGICA.COM.

Reeves, R. Kimball and L. 2008. *The data warehouse lifecycle toolkit.* s.l. : Jon Wiley & Sons, 2008.

2007. Sistemas Gestores de bases de datos. [En línea] 31 de Julio de 2007. [Citado el: 14 de Febrero de 2010.]

Kimball, Ralph y Otros. 2002. *The Data Warehouse Lifecycle Toolkit.* 2002.

Ponniah. 2008. *Data Warehousing Fundamentals.* 2008.

Iglesias, Lic. María Elinor Dulzaides. 2007. *Análisis.* 2007.

—. 2007. *Gestor de Bases de Datos.* 2007.

Pérez, P. V. (2009).

Glosario de términos.

- ✓ **ANSI SQL 2003:** Estándar adoptado por la mayoría de los SGBD. Esta versión incluía algunas características de XML, cambios en las funciones y la estandarización del objeto *sequence* y las columnas auto numéricas.
- ✓ **CASE:** Ayuda a la Ingeniería de Software por computadora.
- ✓ **Centro informante:** Empresas, unidades presupuestadas, entidades subordinadas a las organizaciones políticas, sociales y de masas, cooperativas y unidades básicas de producción cooperativa que envían sus estadísticas a la ONE.¹
- ✓ **CU:** Caso de Uso.
- ✓ **CUBAENERGÍA:** Centro de la Información y Desarrollo de la Energía. Entidad presupuestada de investigación-desarrollo y servicios científico-técnicos en materia de energía y medio ambiente.
- ✓ **CUI:** Código Único de Indicador.
- ✓ **DATEC:** Centro de Tecnologías de Gestión de Datos.
- ✓ **Estadística:** Estudio de los datos cuantitativos de la población, de los recursos naturales e industriales, del tráfico o de cualquier otra manifestación de las sociedades humanas.²
- ✓ **GPL:** Licencia Pública General de GNU (compañía dedicada a la creación y promoción del software libre). Esta licencia contiene todo aquel software que se considera de libre distribución y de código abierto, permitiéndoles a sus usuarios modificar el programa sin incurrir en incumplimientos o faltas con respecto a su licencia.³
- ✓ **Herramientas CASE:** Herramienta de software que automatiza una parte del ciclo de desarrollo de este.
- ✓ **Indicador:** Herramientas para clarificar y definir, de forma más precisa, objetivos e impactos, son medidas verificables de cambio o resultado

¹ Tomado de: Sitio web oficial de la ONE. <http://www.one.cu> (febrero 2010)

² Tomado de: Microsoft Encarta 2009 (febrero 2010)

³ Tomado de: <http://www.gnu.org/licenses/licenses.html>

diseñadas para contar con un estándar contra el cual evaluar, estimar o demostrar el progreso con respecto a metas establecidas.⁴

- ✓ **Plugin:** Aplicación informática que, interactuando con otra aplicación, le aporta una función específica a ésta última.
- ✓ **TCP/IP:** Protocolo de control de transmisión/Protocolo de Internet. Protocolo utilizado para asignar una dirección única para cada computadora en la red, dividir los mensajes en paquetes, enrutar los datos en la red y detectar errores en dichos datos.⁵
- ✓ **Tupla:** Llamadas también registros o filas. Contienen la información relativa a una única entidad.⁶
- ✓ **UCI:** Universidad de Ciencias Informáticas.
- ✓ **DCL:** Lenguaje de Control de Datos.
- ✓ **DDL:** Lenguaje de Definición de Datos.
- ✓ **DML:** Lenguaje de Manipulación de Datos.
- ✓ **Script DCL:** Script que permite al administrador de la BD cargar los usuarios y privilegios. Incluye los tipos de consultas SQL **GRANT** y **REVOKE**.
- ✓ **Script DDL:** Script que permite definir los objetos de la BD. Incluye los tipos de consultas SQL **CREATE** y **ALTER**.
- ✓ **Script DML:** Script que permite manejar los datos de la BD. Incluye los tipos de consultas SQL **SELECT**, **INSERT**, **UPDATE** y **DELETE**.⁷

⁴ Tomado de: Organización de las Naciones Unidas (ONU). *Integrated and coordinated implementation and follow-up of major*. United Nations conferences and summits.

⁵ Tomado de: <http://es.kioskea.net/contents/internet/tcpip.php3>

⁶ Tomado de: Curso de MySQL.

http://www.google.com/cu/url?sa=t&source=web&ct=res&cd=3&ved=0CBgQFjAC&url=http%3A%2F%2Fmysql.conclase.net%2Fcurso%2Findex.php%3Fcap%3D003&rct=j&q=define+tupla&ei=qn3gS6HbNsOC8gbD48irBw&usq=AFQjCNHGzPo5_VSerNR-dOVRkZtjllhKAA

⁷ Tomado de: Introducción al SQL.

<http://cursos.atca.um.es/DBA9i1/sqlplus.html#tiposSentencias>.