



Facultad 8

Análisis, Diseño e Implementación del mercado de datos para los indicadores relacionados con el control de la energía para la Oficina Nacional de Estadísticas.

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

Autor:

Franklin Lopez Castillo

Tutor:

Ing. José Leandro Cabrera Pérez

La Habana, Cuba

Junio 2010

“Año 52 de la Revolución”

Declaración de autoría

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

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

Autor:

Franklin Lopez Castillo

Tutor:

Ing. José Leandro Cabrera Pérez

Datos de contacto

Tutor: Ing. José Leandro Cabrera Pérez

Formación académica:

- ✓ 2007 Graduado de Ingeniero en Ciencias Informáticas en la Universidad de las Ciencias Informáticas (UCI).
- ✓ Categorizado como Instructor.
- ✓ Líder del Proyecto Sistema Integral de Béisbol de la facultad 8.
- ✓ Profesor de la asignatura Sistemas de Bases de Datos con más de 3 años de experiencia en el tema.

Centro laboral: Departamento de la Especialidad, ubicada en la Facultad 8 de la UCI.

Correo electrónico: jcabrera@uci.cu

Agradecimientos

Se me hace bastante difícil agradecer a tantas personas que de un modo u otro me ayudaron en mi carrera profesional. Haré lo posible por no olvidar a nadie que haya contribuido a cumplir uno de mis sueños.

A mis padres Margarita y Francisco, que me han dado todo el apoyo y amor del mundo para que yo saliera adelante en mis estudios y me indicaron el camino a seguir en todo momento.

A nuestro Comandante en Jefe Fidel Castro Ruz y a nuestra Revolución por dejarme realizar mis sueños de ser un profesional y haber pasado 5 grandes e inolvidables años en esta Universidad de excelencia.

A nuestro decano Pedro Luis Basulto por permitirme culminar mis estudios universitarios, mil gracias por ayudarnos cuando lo necesitamos.

A mis hermanos Francis e Inalvis, que me dieron fuerza para continuar y siempre han estado al tanto de mi vida.

A mi tutor y amigo Ing. José Leandro Cabrera quien me ayudó en la elaboración de la tesis siempre que lo necesité y en todo lo que estuvo a su alcance.

A mis tíos: Ulises, Maricel y Arelis, por brindarme su apoyo y ayuda incondicional en toda mi vida universitaria.

A Yunier, que quiero como mi hermano por estar siempre al tanto de mí, por su preocupación y apoyo siempre que lo necesité.

A todos mis amigos, que compartieron conmigo en la universidad y estuvieron presentes para mí en todo momento, por su ayuda, disposición y preocupación.

A mi grupo 8502 y todos los estudiantes que compartieron conmigo por ser el mejor grupo de todos los tiempos... disculpen por no mencionar nombres, pues sería injusto mencionarlos y se me olvidé alguno pero sepan que los quiero mucho y los voy a extrañar cantidad, solo les deseo que les vaya bien en su vida social así como profesional.

Y para terminar a todas aquellas personas que han colaborado con el desarrollo de mi tesis y me han dado de una manera u otra su ayuda en estos años de mi carrera.

Dedicatoria

A mis padres porque todo lo que soy se lo debo a ustedes, por ayudarme en todo momento y confiar siempre en mí. Aquí está el fruto de su sacrificio y amor inmenso.

A mis hermanos, por estar a mi lado siempre y ser los ejemplos de mi vida pues sin su apoyo no hubiese sido posible que estuviera presentando esta tesis.

Resumen

En el presente trabajo se muestra el resultado del análisis, diseño e implementación del mercado de datos de indicadores relacionados con el control de la energía para la Oficina Nacional de Estadísticas (ONE). En esta oficina existe una gran cantidad de información digitalizada relacionada con la energía de muchos años. Dicha información es almacenada en formatos de difícil acceso para su consulta, por tal motivo se hace necesario desarrollar el mercado de datos para los indicadores relacionados con el control de la energía y de esta forma se facilita la disponibilidad de la misma. El mercado de datos se realizará utilizando una metodología adaptada de la metodología de Kimball y lo propuesto por Leopoldo Zenaido Zepeda Sánchez en su Tesis de Doctorado, orientando el trabajo a los casos de uso y fortaleciendo la etapa del levantamiento de requisitos. Utilizando la herramienta CASE Visual Paradigm y el gestor de base de datos PostgreSQL.

Palabras Claves: Almacén de Datos, Mercado de Datos, Indicadores.

Índice de contenidos

Introducción	11
CAPÍTULO # 1 FUNDAMENTACIÓN TEÓRICA	15
1.1 Introducción	15
1.2 Almacén de Datos.....	15
1.3 Bases de datos multidimensionales.....	17
1.4 Proceso OLAP	18
1.5 Mercado de Datos.....	19
1.6 Indicadores	21
1.7 Herramientas de Ingeniería de Software Asistida por Computación (CASE)	21
1.7.1 Embarcadero ER/Studio.....	21
1.7.2 EasyCase Profesional.....	22
1.7.3 Visual Paradigm	22
1.7.4 Case Studio 2	23
1.8 Gestores de Bases de datos.....	24
1.8.1 PostgreSQL	24
1.8.2 Microsoft SQL Server	26
1.8.3 MySQL	27
1.8.4 Oracle	28
1.8.5 Comparación entre los SGBD	30
1.9 Metodologías de desarrollo de Mercado de Datos	31
1.9.1 Metodología para el Diseño Conceptual de Almacenes de Datos	31
1.9.2 Metodología de Ralph Kimball.....	31
1.9.3 Metodología de W. H. Inmon.....	32
1.9.4 Selección de la metodología a utilizar	33
1.10 Conclusiones parciales del capítulo.....	34
CAPÍTULO # 2: ANÁLISIS Y DISEÑO DE LA SOLUCIÓN	35
2.1 Introducción	35

Índice de contenidos

2.2 Análisis.....	35
2.2.1 Definición del Negocio	35
2.2.2 Tema de análisis	35
2.2.3 Roles y permisos.....	35
2.2.4 Reglas del negocio.....	36
2.2.5 Necesidades de los usuarios	37
2.2.6 Casos de uso del sistema	50
2.2.7 Diseño	52
2.2.8 Esquema de Seguridad.....	66
2.2.9 Política de respaldo y recuperación.....	67
2.3 Conclusiones parciales	68
CAPÍTULO # 3: IMPLEMENTACIÓN Y PRUEBAS DE LA SOLUCIÓN.....	69
3.1 Introducción del capítulo	69
3.2 Modelos de Datos Físico	69
3.2.1 Estructuras de Datos.....	71
3.2.2 Usuarios y Privilegios.....	73
3.2.3 Carga de nomencladores	74
3.3 Guía de Implantación.....	75
3.3.1 Requerimientos.....	75
3.3.2 Secuencia de Pasos	76
3.4 Validación y pruebas.....	76
3.4.1 Listas de Chequeo Análisis	76
3.4.2 Validación de requisitos por el cliente.....	77
3.4.3 Pruebas de Implantación.....	77
3.5 Conclusiones parciales	78
Conclusiones generales.....	79
Recomendaciones	80
Referencias bibliográficas	81

Índice de contenidos

Bibliografía	84
Glosario de términos	85
Anexos.....	86

Índice de tablas

Tabla 1 Comparación SGBD	31
Tabla 2 Roles y permisos	36
Tabla 3 Matriz BUS	53
Tabla 4 Tabla descriptiva de la dimensión CAE	54
Tabla 5 Tabla descriptiva de la dimensión NAE	55
Tabla 6 Tabla descriptiva de la dimensión DPA	56
Tabla 7 Tabla descriptiva de la dimensión Centro informante	56
Tabla 8 Tabla descriptiva de la dimensión Esfera	57
Tabla 9 Tabla descriptiva de la dimensión Forma de financiamiento	58
Tabla 10 Tabla descriptiva de la dimensión Forma organizativa	58
Tabla 11 Tabla descriptiva de la dimensión Indicador	59
Tabla 12 Tabla descriptiva de la dimensión Organismo	60
Tabla 13 Tabla descriptiva de la dimensión Subordinación.....	60
Tabla 14 Tabla descriptiva de la dimensión Temporal.....	61
Tabla 15 Tabla descriptiva de la dimensión Sindicato.....	62
Tabla 16 Hecho balance de consumo de los portadores energéticos	62
Tabla 17 Tabla de especificación de las medidas	66
Tabla 18 Esquemas y Tablas.....	72
Tabla 19 Llaves primarias	72
Tabla 20 Llaves foráneas	72
Tabla 21 Índices de las tablas del mercado de datos	73
Tabla 22 Usuarios, roles y permisos.....	74
Tabla 23 Nomencladores	75
Tabla 24 resultados de la aplicación de las listas de chequeo.....	77
Tabla 25 Pruebas de implantación	78

Introducción

Actualmente la energía es imprescindible en nuestras vidas. Todo el sistema de vida moderno está basado en la disposición de abundante energía a bajo costo. Su consumo ha ido creciendo continuamente paralelamente a los cambios de los hábitos de vida y las formas de organización social. En el mundo de hoy el derroche es inmenso y se lucha por su uso racional. En Cuba, el programa de la Revolución Energética ideado por el líder cubano Fidel Castro surgió con el propósito de disminuir los gastos en el consumo eléctrico y prepararse mejor para enfrentar las consecuencias de la actual crisis global capitalista. El órgano encargado del control de la energía es la Oficina Nacional de Estadísticas (ONE); órgano rector de la estadística en Cuba. Su objetivo fundamental es captar, analizar y difundir los datos recogidos a lo largo y ancho de todo el país. Para ello tiene el modelo 5073 en el cual se almacenan los datos relacionados con el balance de consumo de portadores energéticos. Estos datos son almacenados en formatos de difícil acceso para su consulta, por lo que se hace muy complejo el proceso de acceder y divulgar dicha información.

La forma de almacenar, recuperar y presentar la información proveniente del modelo 5073, el cual impide realizar de manera eficiente los principales reportes, cruces de variables, indicadores, tasas y porcentajes, dificultando así la disponibilidad de información para Órganos del Estado y afectando la toma de decisiones.

La Universidad de las Ciencias Informáticas, universidad creada en el calor de la batalla de ideas en la capital cubana, su misión fundamental es el desarrollo de software y servicios informáticos a partir de la vinculación estudio-trabajo, cuenta con el centro Tecnología de Datos (DATEC). Dicho centro se encarga, según el área de conocimientos que cubre, de dar solución a problemas como los existentes en la ONE en busca de mejoras cuantitativas.

Esta investigación surge con el propósito de dar solución a los problemas existentes por lo que se plantea el siguiente **problema a resolver**: ¿Cómo centralizar la información en una estructura que facilite la disponibilidad de los datos en la Oficina Nacional de Estadísticas?

Tomando como **idea a defender** que si se desarrolla un mercado de datos para los indicadores relacionados con el control de la energía se podría contar con una información centralizada y disponible

para los Órganos del Estado a la hora de la toma de decisiones. Definiendo como **objeto de estudio** los almacenes de datos y **el campo de acción** los mercados de datos estadísticos.

Para dar solución a la situación descrita y al problema planteado se trazó como **objetivo general** desarrollar el análisis, diseño e implementación del mercado de datos Indicadores relacionados con el control de la energía para la Oficina Nacional de Estadísticas, **enmarcado en los objetivos específicos:**

- 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 acerca de ellos.
- Realizar el análisis del Modelo 5073 de la Oficina Nacional de Estadísticas.
- Diseñar el mercado de datos Indicadores relacionados con el control de la energía para el almacén de datos de la Oficina Nacional de Estadísticas.
- Implementar y cargar los clasificadores para mercado de datos Indicadores relacionados con el control de la energía para el almacén de datos de la Oficina Nacional de Estadísticas.
- Validar la solución desarrollada mediante la realización de pruebas.

Para dar cumplimiento a los objetivos se trazaron las siguientes **tareas:**

- Estudio de los temas relacionados con el desarrollo de Mercados de Datos para definir los conceptos fundamentales del tema.
- Estudio de las metodologías para definir cual se va a utilizar en el desarrollo del mercado de datos.
- Planificación y realización de entrevistas.
- Identificación de la estructura de usuarios y permisos.
- Definición de los temas de análisis.
- Identificación de las necesidades de información, requisitos funcionales y no funcionales.
- Modelación de los requerimientos.
- Validación de requerimientos.
- Definición de requisitos de entrada y de salida.
- Elección de la granularidad del proceso del negocio.
- Definición de las dimensiones del Mercado de Datos (MD).
- Definición de los hechos asociados a las dimensiones definidas.
- Estructuración del modelo dimensional.

- Transformación del modelo dimensional al diseño físico.
- Implementación de la base de datos.
- Montaje de los clasificadores para mercado de datos Indicadores relacionados con el control de la energía para el almacén de datos de la Oficina Nacional de Estadísticas.
- Realización de pruebas al mercado de datos.

Para dar solución a las tareas se hizo necesaria la utilización de **métodos científicos** para garantizar que la investigación se realizara con la calidad requerida, por lo que se utilizaron los siguientes métodos:

Dentro de los **métodos Teóricos**:

Histórico-Lógico: Este método se utiliza para el estudio de los temas relacionados con el desarrollo de los mercados de datos, su evolución y para la metodología a utilizar en el desarrollo.

Analítico – Sintético: Se utilizó para el análisis de documentos, materiales, y temas relacionados a las mejores prácticas en el desarrollo de mercados de datos. Permitió definir los conceptos fundamentales del tema y se usó para especificar la metodología a utilizar y las dimensiones del mercado de datos.

Modelación: Se utilizó para la modelación del mercado de datos.

Dentro de los **métodos Empíricos** se utilizó la **Entrevista** para obtener la información acerca de las necesidades de información, requisitos funcionales y no funcionales.

El presente trabajo está dividido en tres capítulos:

Capítulo 1: Fundamentación Teórica.

En este capítulo se realiza un estudio del estado del arte acerca de los mercados de datos, las metodologías, herramientas y gestores de base de datos más convenientes a utilizar en el desarrollo del mercado de datos para los indicadores relacionados con el control de la energía.

Capítulo 2: Análisis y diseño de la solución.

Se definen los temas de análisis, 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 y además se definen los requisitos de entrada y de salida.

Capítulo 3: Implementación y Pruebas de la solución.

En este capítulo será elegida la granularidad del proceso del negocio, se definen las dimensiones y los hechos asociados a las dimensiones definidas y se estructura el modelo dimensional. Se logrará transformar el modelo dimensional al diseño físico, se implementa la base de datos, se ejecuta el montaje

de los clasificadores y se realizan las pruebas al mercado de datos.

CAPÍTULO # 1

FUNDAMENTACIÓN TEÓRICA

1.1 Introducción

En este capítulo se realiza un estudio del estado del arte acerca de los mercados de datos. Recoge un estudio de algunas metodologías existentes para el desarrollo del mercado de datos así como un estudio de las herramientas que pudieran emplearse para la construcción de la propuesta.

1.2 Almacén de Datos

Según Ralph Kimball:

Es una fuente de datos de la empresa que puede ser consultada.

- No debe ser organizada con ayuda del modelo entidad/relación
- Es frecuentemente modificada, a partir de datos correctos. (Kimball, 1998)

Según W. H. Inmon:

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

Se puede definir a un almacén de datos como una colección de datos interesantes para la empresa, organizados, integrados y orientados al negocio. Utilizándolo se puede brindar al usuario una visión histórica a la hora de la toma de decisiones.

Entre sus características se encuentran: (Inmon, 1992)

- Es una colección de datos orientado a temas debido a que:

Fundamentación Teórica

- Está organizado en torno a las materias de negocio: cliente, producto y ventas.
- El modelado y el análisis de los datos se enfoca de cara a la toma de decisiones, no para las operaciones del día a día o en el procesamiento de transacciones.
- Proporciona una vista simple y concisa de las materias de negocio excluyendo aquellos datos que no se necesitan para el proceso de toma de decisiones.
- Es una colección de datos integrada porque:
 - Se construye mediante la integración de fuentes de datos múltiples y heterogéneos.
- Bases de datos relacionales, ficheros planos, registros de transacciones on-line
 - Se aplican técnicas de limpieza e integración
 - Asegurar la consistencia en el nombrado, estructuras codificadas, medidas de los atributos, y demás aspectos entre las múltiples bases de datos.
- Cuando los datos se mueven al almacén, éstos se tienen que transformar y varía en el tiempo ya que:
 - El horizonte de tiempo para el almacén es significativamente más largo que el de los sistemas operacionales.
 - Base de datos operacional: valor actual.
 - Almacén: proporciona información desde una perspectiva histórica. (5- 10 años)
 - Cada clave del almacén contiene una referencia a la fecha explícita o implícitamente.
- Es una colección de datos no volátil debido a que:
 - Es un almacén de datos físicamente separado del entorno operacional.
 - Las actualizaciones de la base de datos operacional no ocurren en el entorno del almacén.

- No se requieren mecanismos de control de la concurrencia y recuperación.
- Se requieren dos operaciones nada más:
 - Carga de los datos y acceso a datos.

Posee varias ventajas como las que a continuación se muestran: (Orallo, José Hernández, 2003)

- Aporta facilidad e inmediatez en el manejo de la información.
- Logra la cohesión de los departamentos de una empresa.
- Se basan en la estructura de información, es decir, almacenan la información homogénea y fiable en una estructura basada en la consulta y el tratamiento jerarquizado de la misma, y en un entorno diferenciado de los sistemas operacionales.
- Decisiones soportadas por datos fiables.
- Rentabilidad de las inversiones.
- Aumentar la competitividad en los nuevos entornos hostiles.
- Aumento de la productividad de los técnicos de dirección.

1.3 Bases de datos multidimensionales

Los almacenes de datos son soportados por un modelo multidimensional, a diferencia de las bases de datos relacionales que se sustentan en el modelo Entidad-Relación. El modelo multidimensional guarda la información de manera que sea de fácil entendimiento por parte del usuario y que garantice a su vez, la rápida y eficiente ejecución de las consultas. Esta estructura de almacenamiento tiene forma de cubo multidimensional, donde las aristas representan las dimensiones, y las intersecciones entre estas aristas son los hechos.

La **dimensión** constituye una vista de datos categóricamente consistente. Todos los miembros de una dimensión pertenecen a un mismo grupo. Es una entidad independiente dentro del modelo multidimensional de una organización, que sirve como llave de búsqueda (actuando como índice), o como

mecanismo de selección de datos. (Cristhian Herrera, 2007)

Los **hechos** son medidas numéricas, las cuales describen los resultados (rendimiento) del negocio. Deben ser de un valor cuantificable ó medible. (Cristhian Herrera, 2007)

Además tener presente que la **granularidad** indica el grado de detalle asociado a un hecho particular. El primer gran factor decisivo de la granularidad es el tiempo, ya que mientras menor sea el intervalo de tiempo, mayor será el grado de detalle obtenido. La granularidad depende directamente del número de dimensiones que se asocian a la tabla de hechos. Se deben considerar otros factores como la carga del procesador, espacio de almacenamiento y satisfacer a cabalidad los requerimientos del cliente. (Cristhian Herrera, 2007)

Una **medida** es el valor que toma determinada variable que se está analizando. Son resultados cuantificables, o indicadores clave de desempeño usados para determinar el éxito de una operación de negocios. Orientan las respuestas a preguntas relacionadas con cuestiones numéricas como la cantidad, valor o costo. (Juan David Peña Rivera, Jesús Armando Suárez Daza, 2005)

Una **jerarquía** constituye un grupo de atributos que siguen un orden preestablecido. Las jerarquías definen cómo los datos son sumariados desde los niveles más bajos hacia los más altos. Una dimensión típica soporta una o más jerarquías naturales. Una jerarquía puede pero no exige contener todos los valores existentes en la dimensión. Es un conjunto de atributos descriptivos que permite que a medida que se tenga una relación de muchos a uno se ascienda en la jerarquía.

1.4 Proceso OLAP¹

Es un proceso analítico de datos online que permite al usuario seleccionar y extraer la información desde diferentes puntos de vista. Se traduce en un conjunto de técnicas y estándares definidos por organismos de estandarización.

Pueden ser definidos algunos tipos de OLAP dependiendo de las técnicas que se utilicen a la hora de obtener los datos y la forma en la que están estructurados, etcétera. Dentro de estos tipos están las herramientas que permitirán realizar análisis de los datos, ellas son: (Borja, 2002)

¹ Proceso Analítico en Línea

- **Herramientas MOLAP** (Multidimensional On-line Analytical Process). Son herramientas que acceden a datos que no están almacenados en registros de tablas, sino que almacenan los datos en arrays de varias dimensiones, llamados cubos. Estos cubos utilizan índices para optimizar el acceso a los datos.
- **Herramientas ROLAP** (Relational On-line Analytical Process). Son herramientas OLAP que crean vistas multidimensionales extrayendo los datos de bases de datos SQL ordinarias, es decir, relacionales. Estas herramientas simulan los datos multidimensionales usando sofisticadas técnicas de indexación, cachés, metadata, etc.
- **Herramientas HOLAP** (Hybrid On-line Analytical Process). Nos permiten un análisis híbrido de la información, es decir que une lo mejor de los dos tipos anteriores. El análisis HOLAP nos ayudará a reducir costes de hardware ya que necesitaremos menos disco que en las bases de datos relacionales. Además, la respuesta de las consultas sobre las bases de datos multidimensionales son más rápidas que sobre las relacionales. Como aspecto negativo, los datos multidimensionales deben ser cargados antes de ser consultados y refrescados cuando se actualizan los datos de la organización.

1.5 Mercado de Datos

Se puede definir un mercado de datos como:

Una versión especial de almacén de datos. Son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones. (Outeiro, 2007)

Un subconjunto de los datos de un almacén de datos es la forma de información resumida que soporta los requerimientos de un departamento o función de negocio particular. (Cristhian Herrera, 2007)

Por tanto se puede concluir que un mercado de datos es un subconjunto de un almacén de datos especializado en una parte específica de información de dicho almacén. Es un repositorio de información específica del mercado de datos.

Entre las características destacan: (Cristhian Herrera, 2007)

- Usuarios limitados.
- Área específica.
- Tiene un propósito específico.
- Tiene una función de apoyo.
- Mejora el tiempo de respuesta al usuario final debido a la reducción en el volumen de información a ser accedido.
- Provee datos apropiadamente estructurados para satisfacer los requerimientos de las herramientas de acceso de usuario final.
- Los mercados de datos normalmente usan menos datos de tal manera que la tarea de carga de datos es más fácil, y de esta manera la implementación de un mercado de datos es más simple comparado con un almacén de datos corporativo.
- El costo de implementación de un mercado de datos normalmente es menor que el requerido para establecer un almacén de datos.

Entre sus ventajas se encuentran: (Ing. Bernabeu Ricardo Dario, 2009)

- Fácil acceso a los datos que se necesitan frecuentemente.
- Crea vista colectiva para grupo de usuarios.
- Facilidad de creación.
- Costo inferior al de la aplicación de un almacén de datos completo.
- Son simples de implementar.
- Conllevan poco tiempo de construcción y puesta en marcha.
- Permiten manejar información confidencial.
- Reflejan rápidamente sus beneficios y cualidades.
- Reducen la demanda del depósito de datos.

A continuación se presentan algunas diferencias entre los mercados de datos y los almacenes de datos: (Cristhian Herrera, 2007)

- Los mercados de datos se centran solamente en los requerimientos de los usuarios asociados con un departamento o función de negocio.

- Los mercados de datos normalmente no contienen datos operacionales detallados a diferencia de los almacenes de datos.
- Debido a que los mercados de datos contienen menos información comparado con los almacenes de datos, son más fácilmente entendibles y navegables.
- El mercado de datos es departamental y más sencillo de usar.

1.6 Indicadores

No existe una definición oficial de indicadores por parte de algún organismo nacional o internacional, sólo algunas referencias que los describen como herramientas para clarificar y definir, de forma más precisa, objetivos e impactos. Son medidas verificables de cambio o resultado, diseñada para contar con un estándar contra el cual evaluar, estimar o demostrar el progreso con respecto a metas establecidas. Facilitan el reparto de insumos, produciendo productos y alcanzando objetivos. (Pérez, 2002)

Son variables que sirven para medir los cambios. (Bobadilla, 2009)

Se puede decir que los indicadores son puntos de referencia que sirven para brindar información y permitir el seguimiento del desarrollo de un proceso en estimación.

1.7 Herramientas de Ingeniería de Software Asistida por Computación (CASE)

Las Herramientas CASE son un conjunto de programas y ayudas que dan asistencia a los analistas, ingenieros de software y desarrolladores, durante todos los pasos del ciclo de vida de desarrollo de un software. (EAFIT, 2005)

Las herramientas CASE, están tomando cada vez más relevancia en la planeación y ejecución de proyectos que involucran sistemas de información, pues suelen inducir a los usuarios a la correcta utilización de metodologías que le ayudan a llegar con facilidad a los productos de software construidos. Todas las herramientas CASE prestan un soporte a un lenguaje de modelado para acompañar la metodología. Entre ellas se encuentran las que a continuación se explican.

1.7.1 Embarcadero ER/Studio

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

1.7.2 EasyCase Profesional

EasyCASE Profesional es un producto para la generación de esquemas de base de datos e ingeniería reversa, trabaja para proveer una solución comprensible para el diseño, consistencia y documentación del sistema en conjunto. (Camaleon, 2010)

Esta herramienta permite automatizar las fases de análisis y diseño dentro del desarrollo de una aplicación, posibilitando las aplicaciones desde el procesamiento de transacciones, hasta la aplicación de bases de datos de cliente/servidor, así como sistemas a tiempo real. (Camaleon, 2010)

EasyCASE Profesional, es una herramienta multi-usuario, es ideal para aquellos que necesitan compartir datos y trabajar en un proyecto con otros departamentos. El equipo completo puede acceder a proyectos localizados en el servidor de la red concurrentemente. Para asegurar la seguridad de los datos existe el diagrama y diccionario de los datos que bloquean por niveles al registro, al archivo y al proyecto, así como niveles de control de acceso. (Camaleon, 2010)

1.7.3 Visual Paradigm

Visual Paradigm para UML² es una herramienta CASE profesional que utiliza UML como lenguaje de modelaje. Soporta el ciclo de vida completo del desarrollo de software, entre ellos 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. Soporta un conjunto de lenguajes, tanto en la generación de código e ingeniería inversa en Java, C + +, CORBA IDL, PHP, XML Schema, Ada y Python. Además, apoya la generación de código C #, VB. NET, Object Definition Language (ODL), Flash ActionScript, Delphi, Perl, Objective-C, y Ruby. (Medina, 2009)

1.7.4 Case Studio 2

Es un programa muy usado en la modelación de base de datos que permite crear y mantener diagramas de entidad relación (DER) y diagramas de flujo de datos (DFD). En él se crean scripts SQL de forma automática que funcionan sobre diferentes sistemas de base de datos y se pueden utilizar plantillas para sistemas con características similares. Presenta dos variantes, una gratuita que es bastante completa con solo limitaciones de pago y otra versión completa que ofrece grandes y perfeccionadas funcionalidades. Trata aspectos como contenidos y dominios para dar mejor seguridad al sistema. Contiene catálogos para guardar el modelo de diseño. Brinda soporte para ingeniería inversa que permite identificar y estructurar base de datos ya existentes para poder trabajar con ellas. Facilita el trabajo de documentación al generar informes detallados en HTML³ y RTF⁴, así como la documentación de diagramas de flujo de datos exportables en formato XML⁵. Exporta para varios gestores de base de datos como son, Oracle, MySQL, PostgreSQL, Access, MS SQL, Max DB, Firebird. Tiene una interfaz gráfica muy intuitiva y fácil de usar. Es soportado en diferentes sistemas operativos, Windows 95/98/Me/NT/2000 /XP. Al crear el diagrama entidad relación (ERD) considera cada base de datos de opciones tales como integridad referencial, restricciones, dominios y disparadores. (Rodríguez, 2009)

² Unified Modeling Language (Lenguaje Unificado de Modelado)

³ Hyper Text Markup Language (Lenguaje de Marcado de Hipertexto)

⁴ Rich Text Format (formato de texto enriquecido)

⁵ Extensible Markup Language (Lenguaje de Etiquetado Extensible)

Después de haber realizado un estudio de varias herramientas CASE se determinó que la herramienta CASE Visual Paradigm se utilizará para el desarrollo del mercado de datos debido a que es una herramienta rápida y sencilla. En la UCI es una herramienta bastante conocida y aunque no es libre la universidad paga licencia para su uso. A continuación se muestra un estudio de varios sistemas gestores de bases de datos que se utilizan en el desarrollo de los mercados de datos

1.8 Gestores de Bases de datos

Un Sistema Gestor de Base de Datos (SGBD) es un conjunto de programas que permiten crear y mantener una base de datos, asegurando su integridad, confidencialidad y seguridad. Es una aplicación que permite a los usuarios definir, crear y mantener la base de datos (BD) y proporciona un acceso controlado a la misma. Todo SGBD debe proporcionar los siguientes servicios: (Medina, 2009)

- Creación y definición de la base de datos.
- Manipulación de los datos.
- Acceso controlado a los datos mediante mecanismos de seguridad.
- Mantener integridad y consistencia de los datos.
- Acceso compartido a la base de datos.
- Mecanismos de copias de respaldo y recuperación.

1.8.1 PostgreSQL

Es un gestor de base de datos relacional de código abierto orientado a objetos, publicado bajo la licencia BSD⁶ y creado en software libre por una comunidad de desarrolladores y organizaciones comerciales. El PostgreSQL permite que mientras varios usuarios accedan a una tabla, otros pueden trabajar en ella sin necesidad de bloqueos. Cuenta con una amplia variedad de tipos y un entorno administrativo amigable y

⁶ Berkeley Software Distribution (Distribución de Software Berkeley)

fácil de usar. Es un sistema objeto-relacional ya que incluye la herencia, tipos de datos, funciones, restricciones, disparadores, reglas e integridad transaccional. Es el sistema libre más avanzado soportando la gran mayoría de las transacciones SQL. (Rodríguez, 2009) Su versión más reciente posee incorporadas varias mejoras entre las que están: (Casanova, 2009)

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

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

Entre sus ventajas se pueden encontrar las siguientes: (Rodríguez, 2009)

- Soporta múltiples tipos de datos como los tipos de fecha, monetarios, elementos gráficos, datos de redes y también creación de tipos de datos propios.
- Permite la creación de funciones personalizadas y el manejo y configuración de disparadores.
- Posee una gran escalabilidad.
- Es capaz de ajustarse al número de CPU y la cantidad de memoria que ofrece el sistema de forma óptima, siendo capaz de soportar mayor cantidad de peticiones simultáneas sin provocar errores.
- Hace uso de rollback's⁷.
- Subconsultas y transacciones.
- Comprueba la integridad referencial y almacena procedimientos en su propia base de datos.

Entre sus desventajas están: (Rodríguez, 2009)

⁷ Operación que devuelve a la base de datos a algún estado previo.

- La velocidad de respuestas que ofrece este gestor es un poco deficiente.
- Tienen un límite de 8k por filas que se pueden aumentar hasta 32 lo que trae consigo que disminuya su rendimiento.
- Es más lento que los otros gestores de base de datos.
- Consume gran cantidad de recursos cuando se realizan muchas transacciones de datos.

El PostgreSQL es un gestor de base de datos que cuenta con casi todas las características de los gestores comerciales, haciendo de él una buena alternativa GPL (General Public License). (Rodríguez, 2009)

1.8.2 Microsoft SQL Server

Es un sistema de gestión de base de datos relacional que utiliza lenguaje SQL, a través del cual se puede acceder a grandes cantidades de datos de manera simultánea. Constituye la alternativa de Microsoft y es un sistema muy utilizado en aplicaciones de escritorio y otros sistemas de gestión (Rodríguez, 2009)

Posee servicios de transformación de datos, pues introduce un rediseño completo para proporcionar una plataforma de extracción, transformación y carga de los datos integral. Brinda servicios de presentación de informes que posibilitan a los negocios integrar de forma sencilla datos, ya sean fuentes heterogéneas o almacenes de datos en informes ricos, interactivos y gestionables. (Bolaños, 2006)

Tiene varias ventajas de trabajo como por ejemplo: (Rodríguez, 2009)

- Es un sistema muy costoso.
- Es el primero en soportar la interfaz de acceso OLE DB⁸ y ADO⁹.
- Cuando el SQL Server no tiene tareas del usuario que procesar comienza la llamada tarea de limpieza, lo que hace mejorar el funcionamiento del CPU.

⁸ Object Linking and Embedding for Databases ("Enlace e incrustación de objetos para bases de datos")

⁹ ActiveX Data Objects, mecanismos que usan los programas de computadoras para comunicarse con las bases de datos.

- Necesita menor limpieza en las memorias intermedias durante el procesamiento de transacciones. Registra las transacciones de tal forma que las actualizaciones se pueden reducir al último estado en caso del servidor fallar.
- Soporta transacciones.
- Soporta procedimientos almacenados, triggers¹⁰, etcétera.
- Cuenta con una gran estabilidad y seguridad.
- Escalabilidad.
- Tiene un entorno gráfico administrativo muy potente.

Entre sus desventajas se encuentran: (Rodríguez, 2009)

- Presenta bloqueos a nivel de página.
- Dispositivos con crecimiento manual.
- Hay que pagar licencia para su uso.
- Soportado solo por el sistema operativo Windows.

1.8.3 MySQL

Es un sistema de gestión de base de datos relacional cuyo principal objetivo de creación fue la velocidad. Cuenta con una licencia para uso en software privativo. El MySQL es propietario y su diseño multihilo y multiusuario le permite una gran carga de trabajo, procesando gran cantidad de información. (Rodríguez, 2009)

Hace funcionar almacenes de datos que contengan terabytes de información. En entornos de almacenes de datos este gestor ofrece diversas fortalezas para escalar hasta el rango de los terabytes, tanto para un solo servidor como para varios. (Oracle, 2008)

Posee muchas ventajas como las que se muestran a continuación: (Rodríguez, 2009)

¹⁰ Un **trigger** (o disparador) en una base de datos, es un procedimiento que se ejecuta cuando se cumple una condición establecida al realizar una operación de inserción (INSERT), actualización (UPDATE) o borrado (DELETE).

- Consume pocos recursos por lo que puede ser ejecutado en máquinas de bajo rendimiento.
- Para lograr mejor velocidad se eliminaron las subselecciones y las transacciones.
- Ofrece mejor rendimiento al ser usado en aplicaciones web, en Drupal y con su conexión programada en el lenguaje PHP.
- Los sistemas se pueden conectar usando socket TCP/IP (Protocolo de control de transmisión/Protocolo de Internet) en cualquier plataforma.
- Posee múltiples motores de almacenamiento que permiten escoger la más adecuada para cada tabla.
- Logra mayor cantidad de transacciones por segundo debido a un sistema de agrupación de transacciones.
- El uso de índices y optimización de consultas agilizan el trabajo en bases de datos muy grandes ya que esto trae consigo que se gestione gran cantidad de información.
- Soporta una gran cantidad de tipos de datos diferentes, puede ser utilizada en el desarrollo de bases de datos que gestionen gran volumen de información.

Al igual que otros sistemas tiene desventajas, ellas son:

- Puede provocar problemas de integridad y la información se puede dañar en entornos donde se gestione mucha información simultáneamente.
- Altas concurrencias en las modificaciones de datos.
- Es privativo, lo que trae consigo que cualquier distribución de un producto debe ser bajo la licencia de GNU GPL o la licencia comercial de MySQL.
- Carece de transacciones rollback`s y subconsultas.

1.8.4 Oracle

Es uno de los manejadores de base de datos más potentes que existe actualmente. Brinda las funcionalidades básicas de un gestor sin olvidar las opciones de subconsultas. Es muy vendido a nivel mundial y su alto precio es debido a que Oracle es un manejador de base de datos muy poderoso y solo

está al alcance de medianas o grandes empresas como las multinacionales. Consume gran cantidad de recursos en el servidor que trae consigo que se necesite de un buen sistema de hardware, RAM (Memoria de acceso aleatorio), microprocesador, disco duro. Es capaz de almacenar gran cantidad de información, ejemplo de ellos es la existencia de servidores que gestionan hasta muchos teras de capacidad. En el desarrollo de aplicaciones web no está tan extendido debido que es un sistema muy caro a diferencias de otros sistemas de gestión de base de datos como MySQL y SQL Server. Oracle se basa en tecnología cliente-servidor, cuenta con herramientas necesarias para su utilización en el servidor así como con herramientas de programación básicas de Oracle. Para su desarrollo se utiliza el lenguaje PL/SQL y SQL para crear formularios. (Rodríguez, 2009)

Oracle permite mejorar el rendimiento, la disponibilidad y la facilidad de gestión de los almacenes de datos mediante la partición de las tablas grandes. Posibilita la búsqueda de información y patrones ocultos en el almacén de datos. Ejecuta los análisis más rápidos con vistas materializadas y tablas organizadas. (Oracle. World's Largest Data Warehouse, 2006)

Algunas de sus ventajas son: (Rodríguez, 2009)

- Permite crear los formularios de forma similar a otras herramientas de programación conocidas como Visual Basic, Visual C.
- Puede ejecutarse en varias plataformas, desde una PC hasta un servidor que cuente con diferentes sistemas operativos.
- Usa particiones para mejorar la eficiencia de replicación y algunas versiones admiten administración de base de datos distribuidas.
- Es un sistema muy complejo que permite desarrollar diseños con triggers y procedimientos almacenados con una integridad referencial bastante potente.
- Capaz de almacenar una amplia variedad de información y recuperarla de forma rápida y fácil. Permite a las organizaciones conocer más de su negocio.
- Soporte de transacciones.

- Estabilidad.
- Escalabilidad.

Como otros sistemas de gestión de base de datos cuenta con ciertas desventajas, y esto trae consigo que no sea factible su utilización en algún software, estas desventajas son: (Rodríguez, 2009)

- Es muy riesgoso el proceso de controlador de versiones ya que es bastante engorroso y puede perderse o dañarse la información.
- Las últimas versiones presentan algunos fallos en el sistema de almacenamiento por causa de las extensiones orientadas a objetos.
- Una de sus mayores desventajas radica en su precio, incluso las licencias son excesivamente caras.
- Otro problema es la necesidad de ajustar el sistema a las necesidades del cliente, no basta con instalar el Oracle en un servidor y conectar las aplicaciones al cliente, ya que si es mal configurado puede resultar extremadamente lento.

1.8.5 Comparación entre los SGBD

A continuación se realiza una comparación entre los SGBD estudiados:

Criterios	MySQL	SQLServer	PostgreSQL	Oracle
Sistemas Operativos	Linux/Windows	Windows	Linux/Windows	Windows, Unix , Macintosh y Mainframes
Consumo de recursos	Consume pocos recursos	Consume pocos recursos	Consume pocos recursos	Consume gran cantidad de recursos
Velocidad	Rápido	Rápido	Lento	Rápido

Costo	Muy Costoso	Muy Costoso	Gratis	Muy Costoso
Rendimiento	Alto	Alto	Alto	Alto
Comercialización	Privativo	Privativo	Libre	Privativo
Volumen de Datos	Alto	Alto	Alto	Alto
Integridad	Baja	Alta	Alta	Alta

Tabla 1 Comparación SGBD

Después de realizar un análisis de los diferentes sistemas de gestión de base de datos en la tabla anterior se decidió escoger a PostgreSQL en su versión 8.4 porque es más ventajoso, ya que es multiplataforma y no se necesita pagar una licencia para utilizarlo. Además posee una alta integridad de datos y se adapta al sistema que se necesita para el desarrollo de la aplicación.

1.9 Metodologías de desarrollo de Mercado de Datos

Existen diversas metodologías que pretenden dar un acercamiento a una propuesta ideal para el desarrollo de almacenes. Cada autor la orienta a la optimización del rendimiento y a su visión de los principales procesos que se deben tener en cuenta para construir un almacén de datos flexible y dinámico. A continuación se describen brevemente algunas de ellas.

1.9.1 Metodología para el Diseño Conceptual de Almacenes de Datos

Esta metodología es presentada en la Tesis de Doctorado de Leopoldo Zenaido Zepeda Sánchez. Aporta como aspecto novedoso la incorporación de los casos de uso para guiar el proceso de desarrollo. Define una serie de transformaciones para llevar desde un diagrama relacional a uno dimensional y así obtener las estructuras que conformarán el Repositorio de Datos. (Sánchez, 2008)

1.9.2 Metodología de Ralph Kimball

Se basa en dividir el mundo de inteligencia de negocio entre los hechos y las dimensiones. Es eficaz y conduce a una solución completa en un corto período de tiempo. Tiene abundante documentación y se puede encontrar una respuesta a casi todas las preguntas que se puedan tener. Entre sus características principales se encuentra que su arquitectura es ascendente (bottom-up). Esta metodología plantea que se debe crear por cada departamento un conjunto de MD independientes orientados a los temas que estén relacionados con él. Además según la metodología de Ralph Kimball, el almacén de datos es la unión de todos los mercados de datos de una entidad. (Curto, 2008)

En el ciclo de vida de esta metodología se ilustra el flujo general de implementación de un almacén de datos y se identifica la secuencia de tareas ordenadas y actividades principales que deben suceder concurrentemente. También se plantean que muchas necesidades deben ser acomodadas para lograr una única necesidad de la organización y no todos los detalles de las tareas del ciclo de vida deben ser ejecutadas en todos los proyectos. (Díaz, 2007)

1.9.3 Metodología de W. H. Inmon

Se basa en un enfoque descendente (top-down), propone construir primero el almacén de datos y a partir de este último los Mercados de Datos (MD). Propone 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 DW Corporativo, convirtiéndose en un complejo empresarial de bases de datos relacionales. Inmon afirma que la creación de una base de datos relacional con una leve normalización necesita ser las bases para los MD, por esta razón no se crean los MD directamente desde los Sistemas de Procesamiento de Transacciones en Línea (OLTP) a través de un área de ensayo, sino que se realizan a partir de la arquitectura relacional de los datos corporativos. (DATEC, 2009)

Esta metodología consta de dos fases:

Getting Data In: incluye los procesos ETL¹¹ de las fuentes de origen (que incluyen Data Quality, Data Staging u otros procesos), creación del almacén de datos corporativo y, en caso de ser necesario, ODS y Staging Area.

Getting Information Out: procesos ETL que distribuyen la información a los diferentes mercados de datos que responden necesidades diferentes (lo que se llama Data Delivery). Las herramientas de explotación de datos atacan los mercados de datos y en casos excepcionales al mismo almacén de datos corporativo. (Curto, 2009)

1.9.4 Selección de la metodología a utilizar

Para definir la metodología de desarrollo a utilizar se tomó una adaptada de la metodología de Kimball y lo propuesto por Leopoldo Zenaido Zepeda Sánchez en su Tesis de Doctorado por los siguientes aspectos:

- Crea los conceptos de hechos y dimensiones, lo que indudablemente es muy eficaz en el proceso de la toma de decisiones y proporciona mayor agilidad en el proceso de desarrollo.
- Propone ir construyendo el almacén de datos a través de la construcción de los Mercados de Datos departamentales, lo que constituye una estrategia buena y coincide con la división lógica de las empresas, entidades, organismos, etcétera.
- Existe abundante documentación sobre la misma, la respuesta a todas las dudas y preguntas que puedan surgir se pueden encontrar en la web y a través de los servicios que brindan el grupo creador de la metodología.
- Es una metodología madura y reconocida por el resto de la comunidad dedicada al tema. Tiene bien definidas las etapas, actividades, artefactos y roles.

¹¹ Extracción, Transformación y Carga

- Como complemento y fortaleciendo la etapa del levantamiento de requisitos, se tomó lo planteado por Leopoldo en su Tesis de Doctorado, orientando así el trabajo a los Casos de Uso. De esta forma se logra estar más alineado con las tendencias y normas de la UCI.

Esta metodología presenta un ciclo de vida bien definido donde existen una serie de flujos de trabajo que tienen objetivos específicos, estos últimos son: (DATEC, 2009)

- Estudio Preliminar o Planeación.
- Requerimientos.
- Arquitectura y Diseño.
- Implementación.
- Prueba.
- Despliegue.
- Soporte y Mantenimiento.
- Gestión y Administración del Proyecto.

1.10 Conclusiones parciales del capítulo

En este capítulo se realizó un estudio de las tecnologías y tendencias actuales en cuanto al desarrollo de mercados de datos. Se analizaron las principales herramientas, gestores de base de datos y metodologías que han alcanzado un gran auge actualmente, teniendo en cuenta el conocimiento previo sobre dichas tecnologías. Debido a que la UCI, al igual que Cuba, presenta una gran tendencia al software libre, las tecnologías que fueron seleccionadas se identifican por ser en su mayoría software libre y multiplataformas.

CAPÍTULO # 2: ANÁLISIS Y DISEÑO DE LA SOLUCIÓN

2.1 Introducción

En este capítulo se tratan los aspectos relacionados con la definición de los temas de análisis. Se identifica la estructura de usuarios y permisos así como las necesidades de información. Son definidos los requisitos funcionales, no funcionales y los multidimensionales (requisitos de entrada y de salida). Además se modelan y validan los requerimientos.

2.2 Análisis

2.2.1 Definición del Negocio

La Oficina Nacional de Estadísticas es una entidad creada para proponer, organizar y ejecutar, según corresponda, la aplicación de la política estatal en materia de estadística del país. Para organizar mejor su trabajo cuenta con la Dirección de Industria y Medio Ambiente, la cual presenta el departamento de Energía donde se encuentra el personal especializado solamente en este tema. Las personas que laboran en este departamento se encargan de analizar todos los datos recogidos a todo lo largo y ancho del país relacionados con la energía.

2.2.2 Tema de análisis

Después de haber analizado los aspectos relacionados con la presente investigación se decidió tomar como tema de análisis: Energía, ya que tributa a la investigación desarrollada y está acorde al objetivo general planteado.

2.2.3 Roles y permisos

Análisis y Diseño de la solución

En este aspecto se definen los roles y permisos en dependencia de la actividad que vaya a desarrollar cada rol en el mercado de datos, ya sea lectura o escritura y depende además del tipo de información que vaya a manejar.

Roles	Permisos	
	Lectura	Escritura
Administrador	X	X
Analista	X	

Tabla 2 Roles y permisos

Leyenda:

Roles:

- **Administrador:** persona autorizada y capacitada para introducir y consultar la información almacenada en el mercado de datos.
- **Analista:** persona autorizada y capacitada solo para consultar la información almacenada en el mercado de datos.

2.2.4 Reglas del negocio

En el libro Database Design for Mere Mortals de Michael J. Hernández se definen las reglas de negocio con la siguiente frase:

"Las reglas de negocio imponen restricciones o limitaciones sobre ciertos aspectos de una base de datos basadas en la manera en que la organización percibe o utiliza sus datos."

El presente trabajo presenta las siguientes reglas del negocio:

- El primer y segundo número de la dimensión CAE significan el sector, del primero al cuarto significa la rama y los seis en su conjunto la sub rama.
- En la dimensión DPA el código está compuesto por cuatro dígitos: los dos primeros identifican la provincia y los dos restantes representan al municipio.

- La tasa de crecimiento se calcula dividiendo el consumo acumulado real y el consumo acumulado del año anterior (Columna 16 / Columna 17).

2.2.5 Necesidades de los usuarios

Las necesidades de los usuarios constituyen la información que exige conocer el cliente para generar posteriormente los reportes relacionados con el balance del consumo de los portadores energéticos. A continuación se plasman éstas:

Obtener la demanda del mes siguiente a la entrega, el plan operativo del mes, el inventario inicial y final físico, las compras a CUPET y las compras al cargar otras entradas y salidas. El consumo directo e indirecto, el consumo, el consumo acumulado real y consumo acumulado del año anterior, el entregado para consumo, la carga para el próximo mes, el saldo final y el recibido del que efectúa carga.

2.2.5.1 Requisitos de información

- RI-1 Obtener la demanda del mes siguiente a la entrega dado:
 - o un organismo.
 - o una provincia
 - o la forma de financiamiento.
 - o el CAE¹².
 - o el tipo de consumo.
 - o el centro informante.
 - o la esfera.
 - o la subordinación.
 - o un mes.
 - o un indicador.

¹² Clasificador de Actividades Económicas

- el NAE¹³.
- un sindicato.
- RI-2 Obtener el plan operativo del mes dado:
 - un organismo.
 - una provincia.
 - la forma de financiamiento.
 - el CAE.
 - el tipo de consumo.
 - el centro informante.
 - la esfera.
 - la subordinación.
 - un mes.
 - un indicador.
 - el NAE.
 - un sindicato.
- RI-3 Obtener el inventario inicial físico dado:
 - un organismo.
 - una provincia.
 - la forma de financiamiento.
 - el CAE.
 - el tipo de consumo.
 - el centro informante.
 - la esfera.

¹³ Nomenclador de Actividades Económicas.

Análisis y Diseño de la solución

- la subordinación.
 - un mes.
 - un indicador.
 - el NAE.
 - un sindicato.
- RI-4 Obtener el inventario final físico dado:
- un organismo.
 - una provincia.
 - la forma de financiamiento.
 - el CAE.
 - el tipo de consumo.
 - el centro informante.
 - la esfera.
 - la subordinación.
 - un mes.
 - un indicador.
 - el NAE.
 - un sindicato.
- RI-5 Obtener el consumo directo dado:
- un organismo.
 - una provincia.
 - la forma de financiamiento.
 - el CAE.
 - el tipo de consumo.

- el centro informante.
 - la esfera.
 - la subordinación.
 - un mes.
 - un indicador.
 - el NAE.
 - un sindicato.
- RI-6 Obtener el consumo indirecto dado:
- un organismo.
 - una provincia.
 - la forma de financiamiento.
 - el CAE.
 - el tipo de consumo.
 - el centro informante.
 - la esfera.
 - la subordinación.
 - un mes.
 - un indicador.
 - el NAE.
 - un sindicato.
- RI-7 Obtener las compras a CUPET¹⁴ dado:
- un organismo.
 - una provincia.

¹⁴ Cubapetróleo.

Análisis y Diseño de la solución

- la forma de financiamiento.
 - el CAE.
 - el tipo de consumo.
 - el centro informante.
 - la esfera.
 - la subordinación.
 - un mes.
 - un indicador.
 - el NAE.
 - un sindicato.
- RI-8 Obtener las otras entradas dado:
- un organismo.
 - una provincia.
 - la forma de financiamiento.
 - el CAE.
 - el tipo de consumo.
 - el centro informante.
 - la esfera.
 - la subordinación.
 - un mes.
 - un indicador.
 - el NAE.
 - un sindicato.

Análisis y Diseño de la solución

- RI-9 Obtener las otras salidas dado:
 - o un organismo.
 - o una provincia.
 - o la forma de financiamiento.
 - o el CAE.
 - o el tipo de consumo.
 - o el centro informante.
 - o la esfera.
 - o la subordinación.
 - o un mes.
 - o un indicador.
 - o el NAE
 - o un sindicato.

- RI-10 Obtener el consumo acumulado real dado:
 - o un organismo.
 - o una provincia.
 - o la forma de financiamiento.
 - o el CAE.
 - o el tipo de consumo.
 - o el centro informante.
 - o la esfera.
 - o la subordinación.
 - o un mes.
 - o un indicador.

Análisis y Diseño de la solución

- el NAE.
- un sindicato.
- RI-11 Obtener el consumo acumulado año anterior dado:
 - un organismo.
 - una provincia.
 - la forma de financiamiento.
 - el CAE.
 - el tipo de consumo.
 - el centro informante.
 - la esfera.
 - la subordinación.
 - un mes.
 - un indicador.
 - el NAE.
 - un sindicato.
- RI-12 Obtener el saldo final dado:
 - un organismo.
 - una provincia.
 - la forma de financiamiento.
 - el CAE.
 - el tipo de consumo.
 - el centro informante.
 - la esfera.
 - la subordinación.

- un mes.
- un indicador.
- el NAE.
- un sindicato.
- RI-13 Obtener la carga para el próximo mes dado:
 - un organismo.
 - una provincia.
 - la forma de financiamiento.
 - el CAE.
 - el tipo de consumo.
 - el centro informante.
 - la esfera.
 - la subordinación.
 - un mes.
 - un indicador.
 - el NAE.
 - un sindicato.
- RI-14 Obtener las compras al cargar dado:
 - un organismo.
 - una provincia.
 - la forma de financiamiento.
 - el CAE.
 - el tipo de consumo.
 - el centro informante.

Análisis y Diseño de la solución

- la esfera.
 - la subordinación.
 - un mes.
 - un indicador.
 - el NAE.
 - un sindicato.
- RI-15 Obtener el recibido del que efectúa carga dado:
- un organismo.
 - una provincia.
 - la forma de financiamiento.
 - el CAE.
 - el tipo de consumo.
 - el centro informante.
 - la esfera.
 - la subordinación.
 - un mes.
 - un indicador.
 - el NAE.
 - un sindicato.
- RI-16 Obtener el consumo dado:
- un organismo.
 - una provincia.
 - la forma de financiamiento.
 - el CAE.

- el tipo de consumo.
 - el centro informante.
 - la esfera.
 - la subordinación.
 - un mes.
 - un indicador.
 - el NAE.
 - un sindicato.
- RI-17 Obtener el entregado para consumo dado:
- un organismo.
 - una provincia.
 - la forma de financiamiento.
 - el CAE.
 - el tipo de consumo.
 - el centro informante.
 - la esfera.
 - la subordinación.
 - un mes.
 - un indicador.
 - el NAE.
 - un sindicato.

2.2.5.2 Requisitos funcionales

1. Extraer datos del Excel Balance de consumo de los portadores energéticos.
2. Transformar y cargar datos del Excel Balance de consumo de los portadores energéticos.

3. Extraer datos del Excel Forma de financiamiento.
4. Transformar y cargar datos del Excel Forma de financiamiento.
5. Extraer datos del Excel Organismo.
6. Transformar y cargar datos del Excel Organismo.
7. Extraer datos del Excel CAE.
8. Transformar y cargar datos del Excel CAE.
9. Extraer datos del Excel NAE.
10. Transformar y cargar datos del Excel NAE.
11. Extraer datos del Excel Entidad que consume.
12. Transformar y cargar datos del Excel Entidad que consume.
13. Extraer datos del Excel Centro informante.
14. Transformar y cargar datos del Excel Centro informante.
15. Extraer datos del Excel Sindicato.
16. Transformar y cargar datos del Excel Sindicato.
17. Extraer datos del Excel Indicador.
18. Transformar y cargar datos del Excel Indicador.
19. Extraer datos del Excel DPA¹⁵.
20. Transformar y cargar datos del Excel DPA.
21. Extraer datos del Excel Subordinación.
22. Transformar y cargar datos del Excel Subordinación.
23. Extraer datos del Excel Esfera.
24. Transformar y cargar datos del Excel Esfera.

¹⁵ División Política Administrativa.

2.2.5.3 Requisitos no funcionales

Estos requisitos no son más que las propiedades o cualidades que el producto debe tener.

2.2.5.3.1 Usabilidad

Por las características de los sistemas de almacenamiento y análisis de datos el/los usuario(s) final(es) debe(n) tener conocimiento sobre el trabajo con las herramientas de integración de datos.

2.2.5.3.1.1 Requisitos de Usabilidad

El sistema debe contar con un diseño sencillo, además de una estructura y distribución que permita trabajar con rapidez y eficiencia.

2.2.5.3.2 Fiabilidad

La integridad de los datos es fundamental para lograr el éxito de la solución de almacenamiento y análisis de datos.

- Disponibilidad.
- Máximo de errores.

2.2.5.3.2.1 Requisitos de Fiabilidad

Disponibilidad:

- El área de almacenamiento del proceso de ETL (Extracción, Transformación y Carga) debe estar disponible las 12 horas del día, lo que representa 50% de las horas de uso.
- El sistema de integración será accedido para su mantenimiento 1 vez por mes. En este plazo de mantenimiento se validarán las estructuras de auditoría y se establecerán estrategias que permitirán clasificar posibles errores y darles solución en caso de ser posible.
- El sistema debe estar disponible 100% entre las 8:00 am y las 5:00 pm de lunes a sábado.

Máximo de errores:

- La cantidad de errores en el proceso de integración define la calidad de los datos que se están almacenando; es por ello que es crítico, definiéndose 0 errores/puntos de función.

2.2.5.3.3 Eficiencia

La eficiencia del sistema constituye un elemento clave para la calidad requerida de las funciones que deberá realizar el sistema solicitadas por los clientes.

2.2.5.3.3.1 Requisitos de Eficiencia

Utilización de recursos:

- El tiempo de respuesta debe ser de 0.5 segundos en tiempo real.
- El sistema debe permitir 50 usuarios conectados simultáneamente sin que se afecte el tiempo de respuesta.

Capacidad:

En el proceso de integración solo tendrá conectado un usuario que tendrá la tarea de vigilar el proceso de integración de datos.

2.2.5.3.4 Restricciones de diseño

El producto se desarrollará con PostgreSQL 8.4.

2.2.5.3.5 Requisitos para la documentación de usuarios en línea y ayuda del sistema

Se dispondrá de una guía de ayuda sobre componentes del sistema, además de una gran documentación sobre herramientas ETL presentadas en el sitio de la comunidad de Pentaho.

2.2.5.3.6 Componentes Comprados

Todas las herramientas utilizadas están licenciadas bajo GPL, y son Open Source y totalmente gratis, así como ampliamente reutilizables y extensibles.

2.2.5.3.7 Interfaz

Los requerimientos de interfaz de usuario se centran en la presentación de la información de cara al cliente.

2.2.5.3.7.1 Interfaces de usuario

- Los reportes estadísticos deben contar con una interfaz simple que facilite la interacción usuario-aplicación.
- Las interfaces de salida no serán cargadas con información innecesaria.
- Los gráficos serán con los colores establecidos por la ONE ajustándose a los estándares establecidos de un buen diseño.

2.2.5.3.7.2 Interfaces Hardware

- En el proceso de integración es necesario la utilización de una memoria mínima de 1 GB para el proceso de transformación.
- Se debe contar de un área de almacenamiento intermedio de 80 GB como mínimo.
- Para la visualización y la inteligencia de negocio se necesita una memoria de 1 GB.
- Las estaciones de trabajo (pc clientes) deben contar con impresoras (para garantizar la impresión de las tablas de salida).

2.2.5.3.7.3 Interfaces de Comunicación

- La comunicación entre la base de datos de integración y el almacén de datos es a través del protocolo TCP/IP.
- El sistema necesita estar conectado directamente a un dispositivo de red.

2.2.5.3.8 Requisitos de Licencia

Las herramientas a utilizar se encuentran bajo licencia GPL.

2.2.6 Casos de uso del sistema

2.2.6.1 Caso de uso de información

CU_Consultar balance de consumo de los portadores energéticos.

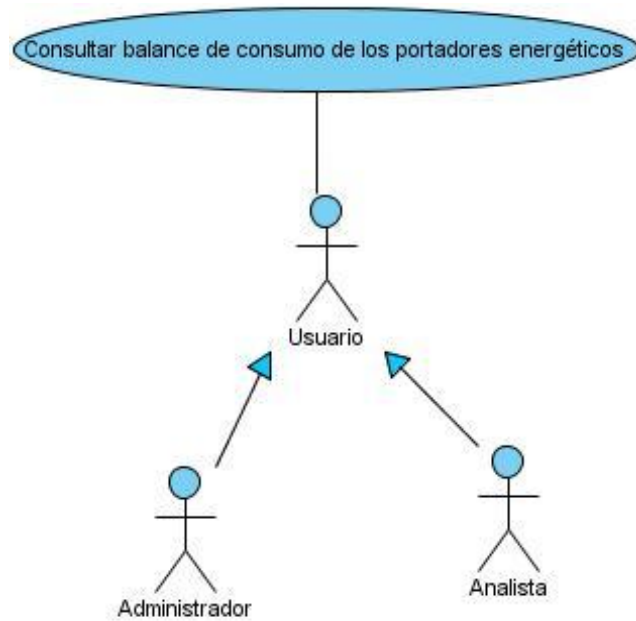


Figura 1: Diagrama de caso de uso informativo

2.2.6.2 Casos de uso funcionales

- Extraer del Excel Balance de consumo de los portadores energéticos.
- Transformar y cargar del Excel Balance de consumo de los portadores energéticos.
- Extraer del Excel Forma de financiamiento.
- Transformar y cargar del Excel Forma de financiamiento.
- Extraer del Excel Organismo.
- Transformar y cargar del Excel Organismo.
- Extraer del Excel CAE.
- Transformar y cargar del Excel CAE.
- Extraer del Excel NAE.
- Transformar y cargar del Excel NAE.
- Extraer del Excel Tipo de consumo.

- Transformar y cargar del Excel Tipo de consumo.
- Extraer del Excel Centro informante.
- Transformar y cargar del Excel Centro informante.
- Extraer del Excel Sindicato.
- Transformar y cargar del Excel Sindicato.
- Extraer del Excel Indicador.
- Transformar y cargar del Excel Indicador.
- Extraer del Excel DPA.
- Transformar y cargar del Excel DPA.
- Extraer del Excel Subordinación.
- Transformar y cargar del Excel Subordinación.
- Extraer del Excel Esfera.
- Transformar y cargar del Excel Esfera.

2.2.7 Diseño

2.2.7.1 Matriz BUS

El propósito de la Matriz Dimensional es obtener un modelo lógico inicial, donde quedan identificados todos los hechos y dimensiones relacionadas con éstos. La matriz recoge la relación que existe entre los hechos y las dimensiones, es decir con qué dimensiones está vinculado el hecho.

Como dimensiones están:

1. DPA
2. Indicador
3. CAE
4. NAE

Análisis y Diseño de la solución

5. Centro informante
6. Temporal
7. Variante
8. Organismo
9. Forma de financiamiento
10. Esfera
11. Subordinación
12. Sindicato

Como hecho a analizar se plantea el siguiente:

1. Balance consumo portadores energéticos (hech_balance_consumo_portadores_energ)

Dimensión/ Hecho	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5	Dim 6	Dim 7	Dim 8	Dim 9	Dim 10	Dim 11	Dim 12
1	X	X	X	X	X	X	X	X	X	X	X	X

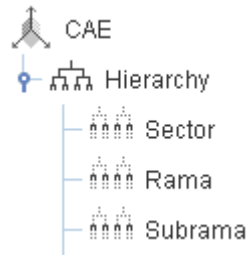
Tabla 3 Matriz BUS

2.2.7.2 Modelo de Datos

A continuación se presenta una breve descripción de las tablas de las dimensiones, la tabla de hechos y las medidas relacionadas con la tabla de hechos. Todas ellas conforman el modelo de datos.

2.2.7.2.1 Especificación de las dimensiones

2.2.7.2.1.1 Dimensión DIM_CAE

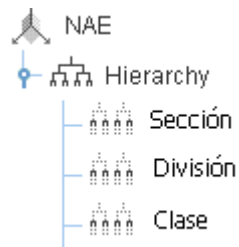


Esta dimensión posee los siguientes atributos:

Nombre del atributo	Descripción
cae_id	Es la llave primaria de la dimensión.
sector_codigo	Almacena el código oficial de los sectores establecidos dentro del CAE
sector_descripcion	Almacena la descripción de los sectores establecidos dentro del CAE
rama_codigo	Almacena el código oficial de las ramas establecidos dentro del CAE
rama_descripcion	Almacena la descripción de las ramas establecidas dentro del CAE
subrama_codigo	Almacena el código oficial del CAE.
subrama_descripcion	Almacena la descripción de las subramas establecidas dentro del CAE.

Tabla 4 Tabla descriptiva de la dimensión CAE

2.2.7.2.1.2 Dimensión DIM_NAE

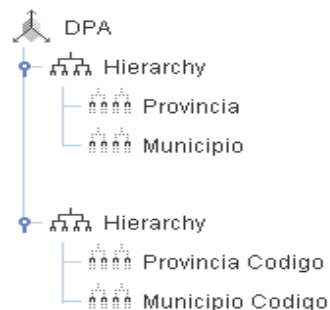


La dimensión posee los siguientes atributos:

Nombre del atributo	Descripción
nae_id	Es la llave primaria de la dimensión.
seccion_codigo	Almacena el código oficial de las secciones establecidas dentro del NAE
seccion_descripcion	Almacena la descripción de las secciones establecidas dentro del NAE
division_codigo	Almacena el código oficial de las divisiones establecidas dentro del NAE
division_descripcion	Almacena la descripción de las divisiones establecidas dentro del NAE
clase_codigo	Almacena el código oficial de las clases establecidas dentro del NAE.
clase_descripcion	Almacena la descripción de las clases establecidas dentro del NAE.

Tabla 5 Tabla descriptiva de la dimensión NAE

2.2.7.2.1.3 Dimensión DIM_DPA

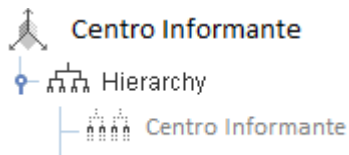


Los atributos de esta dimensión son:

Nombre del atributo	Descripción
dpa_id	Es la llave primaria de la dimensión.
prov_codigo	Almacena el código oficial de las provincias establecidos dentro de la DPA
prov_nombre	Almacena el nombre de las provincias establecidas dentro de la DPA
mun_codigo	Almacena el código oficial de los municipios establecidos dentro de la DPA
mun_nombre	Almacena el nombre de los municipios establecidas dentro de la DPA

Tabla 6 Tabla descriptiva de la dimensión DPA

2.2.7.2.1.4 Dimensión DIM_CENTRO_INFORMANTE



A continuación se muestra una tabla con los atributos pertenecientes a esta dimensión:

Nombre del atributo	Descripción
ci_id	Es la llave primaria de la dimensión.
ci_codigo	Almacena el código oficial de los centros informantes establecidos a nivel nacional
ci_descripcion	Almacena el nombre de los centros informantes establecidos a nivel nacional

Tabla 7 Tabla descriptiva de la dimensión Centro informante

2.2.7.2.1.5 Dimensión DIM_ESFERA



La tabla muestra los atributos de esta dimensión:

Nombre del atributo	Descripción
esfera_id	Es la llave primaria de la dimensión.
esfera_codigo	Almacena el código oficial de las esferas establecidas a nivel nacional
esfera_desc	Almacena el nombre de las esferas establecidas a nivel nacional Ejemplo: Esfera Productiva

Tabla 8 Tabla descriptiva de la dimensión Esfera

2.2.7.2.1.6 Dimensión DIM_FINANCIAMIENTO



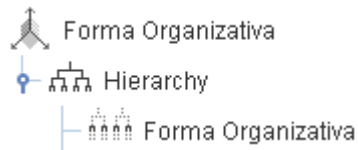
La tabla muestra los atributos de esta dimensión:

Nombre del atributo	Descripción
ff_id	Es la llave primaria de la dimensión.
ff_codigo_gral	Almacena el código oficial de la forma de financiamiento 2 establecida a nivel nacional
esfera_descripcion_gral	Almacena la descripción oficial de la forma de financiamiento 2 establecida a nivel nacional
ff_codigo	Almacena el código oficial de la forma de financiamiento 1 establecida a nivel nacional

	nacional
esfera_descripcion	Almacena la descripción oficial de la forma de financiamiento 1 establecida a nivel nacional

Tabla 9 Tabla descriptiva de la dimensión Forma de financiamiento

2.2.7.2.1.7 Dimensión DIM_FORGANIZATIVA



La tabla muestra los atributos de esta dimensión:

Nombre del atributo	Descripción
forganizativa_id	Es la llave primaria de la dimensión.
forganizativa_codigo	Almacena el código oficial de las formas organizativas establecidos a nivel nacional
forganizativa_descripcion	Almacena el nombre de las formas organizativas establecidas a nivel nacional

Tabla 10 Tabla descriptiva de la dimensión Forma organizativa

2.2.7.2.1.8 Dimensión DIM_INDICADOR



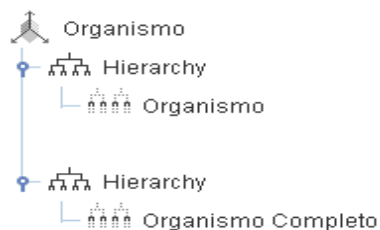
La tabla muestra los atributos de esta dimensión:

Análisis y Diseño de la solución

Nombre del atributo	Descripción
indicador_id	Es la llave primaria de la dimensión.
Indicador_clasificado	Almacena si el indicador es clasificado o no.
tematica_codigo	Almacena el código oficial de las temáticas establecidos a nivel nacional
tematica_descripcion	Almacena el nombre de las temáticas establecidas a nivel nacional
indicador_cui	Almacena el código oficial de los indicadores establecidos a nivel nacional en el Clasificador Único de Indicadores
indicador_descripcion	Almacena el nombre oficial de los indicadores establecidos a nivel nacional en el Clasificador Único de Indicadores
indicador_um	Almacena la unidad de medida oficial de los indicadores establecidos a nivel nacional en el Clasificador Único de Indicadores
indicador_agregable	Almacena la si un indicador determinado puede es aditivo.
indicador_tipo	Almacena el tipo de indicadores establecidos a nivel nacional en el Clasificador Único de Indicadores

Tabla 11 Tabla descriptiva de la dimensión Indicador

2.2.7.2.1.9 Dimensión DIM_ORGANISMO



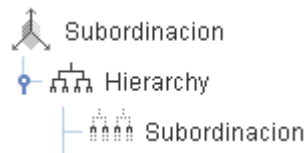
La tabla muestra los atributos de esta dimensión:

Nombre del atributo	Descripción
---------------------	-------------

organismo_id	Es la llave primaria de la dimensión.
organismo_codigo	Almacena el código oficial de los organismos establecidos a nivel nacional
organismo_descripcion	Almacena el nombre de los organismos establecidos a nivel nacional
organismo_desc_corta	Almacena las siglas de los organismos establecidos a nivel nacional

Tabla 12 Tabla descriptiva de la dimensión Organismo

2.2.7.2.1.10 Dimensión DIM_ SUBORDINACION

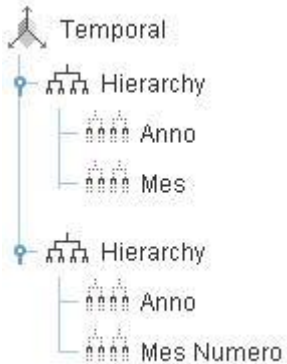


La tabla muestra los atributos de esta dimensión:

Nombre del atributo	Descripción
subord_id	Es la llave primaria de la dimensión.
subord_codigo	Almacena el código oficial de las subordinaciones establecidas a nivel nacional
subord_descripcion	Almacena el nombre de las subordinaciones establecidas a nivel nacional

Tabla 13 Tabla descriptiva de la dimensión Subordinación

2.2.7.2.1.11 Dimensión DIM_TEMPORAL

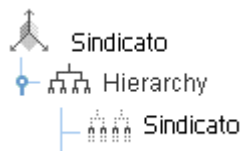


La tabla muestra los atributos de esta dimensión:

Nombre del atributo	Descripción
temporal_id	Es la llave primaria de la dimensión.
anno	Almacena el año de la información
mes_numero	Almacena el número del mes de la información
mes_descripcion	Almacena el nombre del mes de la información
mes_codigo	Almacena el código del mes de la información

Tabla 14 Tabla descriptiva de la dimensión

2.2.7.2.1.12 Dimensión DIM_SINDICATO



La tabla muestra los atributos de esta dimensión:

Nombre del atributo	Descripción
sindicato_id	Es la llave primaria de la dimensión.

Análisis y Diseño de la solución

sindicato_codigo	Almacena el código del sindicato
sindicato_descripcion	Almacena el nombre del sindicato. Ej.: Sindicato Nacional Trabajadores Azucareros

Tabla 15 Tabla descriptiva de la dimensión Sindicato

2.2.7.3 Tabla de hechos

Atributos	ID	Tipo de dato
DIM_CAE	CAE_ID	integer
DIM_SUBORDINACION	SUBORD_ID	integer
DIM_FFINANCIAMIENTO	FF_ID	integer
DIM_ENTIDAD_CONSUME	ENTIDAD_CONSUME_ID	integer
DIM_ORGANISMO	ORGANISMO_EST_ID ORGANISMO_REG_ID	integer
DIM_DPA	DPA_EST_ID DPA_REG_ID	integer
DIM_CENTRO_INFORMANTE	CI_ID	integer
DIM_NAE	NAE_ID	integer
DIM_TEMPORAL	TEMPORAL_ID	integer
DIM_INDICADOR	INDICADOR_ID	integer
DIM_SINDICATO	SINDIC_ID	integer
DIM_ESFERA	ESFERA_ID	integer

Tabla 16 Hecho balance de consumo de los portadores energéticos

2.2.7.4 Medidas

Nombre de la medida	Descripción
---------------------	-------------

Análisis y Diseño de la solución

Demanda del mes siguiente a la entrega	Cantidad solicitada por la entidad a su organismo superior para el mes siguiente con relación a la presentación del modelo.
Plan operativo del mes	Plan operativo de consumo del mes que se analiza, aprobado para la entidad por el MEP ¹⁶ que tiene en cuenta la capacidad de consumo, entendiéndose como tal el inventario disponible al cierre del mes anterior y la autorización de consumo del mes.
Inventario inicial físico	Se reportará el volumen real existente, medido físicamente a principios del mes en los tanques de almacenamiento, debiendo coincidir las magnitudes con las informadas como inventario final del mes anterior. Solo lo reportarán las entidades que tienen depósitos para almacenar físicamente combustibles y lubricantes, tanto para su consumo como la venta en servicentros.
Compras a CUPET	Cantidad física recibida por la entidad proveniente de cualquier distribuidora de la Unión Cuba Petróleo o CUBALUB, tanto para el consumo (Variante 1) como para la venta de los servicentros (Variante 2). Se utilizará como documento primario la factura oficial establecida por el sistema de control interno del MFP (SNC-105), emitida por el proveedor y que acompaña la entrega física del producto.
Otras entradas	Cantidad física recibida en los depósitos propios de almacenamiento de la entidad, de cualquier producto proveniente de otra entidad no perteneciente a la Unión Cuba Petróleo, utilizándose como fuente un documento oficial establecido en el sistema de control interno del MFP que acompañe la entrega del mismo. Se reporta también la compra de los lubricantes adquiridos por fuentes diferentes a CUBALUB. Solo lo reportarán las entidades que tienen depósitos para almacenar físicamente combustibles y lubricantes.
Consumo directo	Consumo correspondiente a todos los equipos de la entidad a los que se les

¹⁶ Ministerio de Economía y Planificación

Análisis y Diseño de la solución

	sirve el producto de sus depósitos propios, excluyendo lo adquirido por tarjetas magnéticas, en cualquier tipo de moneda, en los servicentros de las Cadenas establecidas al efecto. Solo lo reportarán las entidades que tienen depósitos para almacenar físicamente combustibles y lubricantes.
Consumo indirecto	Consumo correspondiente a todos los equipos de la entidad a los que se les sirve el producto desde depósitos ajenos a la entidad y por medios diferentes a la tarjeta magnética.
Otras salidas	Entrega física de combustibles y lubricantes extraídos de los depósitos propios por un consumidor a otro perteneciente o no a su organismo (Variante 1) o vendido por los servicentros (Variante 2). Solo lo reportarán las entidades que tienen depósitos para almacenar físicamente combustibles y lubricantes.
Inventario final físico	Volumen real existente, medido físicamente a fines de cada mes, en los tanques de almacenamiento destinados al consumo y que corresponde al saldo resultante de las operaciones realizadas por consumir (variante 1) o vender (variante 2). Solo lo reportarán las entidades que tienen depósitos para almacenar físicamente combustibles y lubricantes. Nota: es calculable, pero no para el negocio del trabajo. Es un valor que agrega el que llena el modelo.
Compras al cargar	Se reporta por las entidades que asumen la carga de las tarjetas en los centros de carga establecidos a tales efectos de las tarjetas propias o de otros consumidores. Constituye el equivalente físico de combustible cargado en tarjetas en el mes que se informa y puede corresponder tanto a la asignación de ese período como a la cantidad comprada a cuenta de anticipar, parcial o totalmente, lo asignado para el mes siguiente. Se refiere a lo adquirido en cualquier tipo de moneda.
Recibido del que efectúa	Equivalente físico de combustible total recibido por una entidad en las tarjetas magnéticas que, siendo de su propiedad, son cargadas por otra entidad de su

Análisis y Diseño de la solución

carga	propio organismo. Por tal motivo, no incluye aquellas cantidades de combustibles que son recibidas excepcionalmente, en tarjetas propiedad de otros centros al que se le preste algún servicio. Se refiere a lo adquirido en cualquier tipo de moneda.
Consumo	Se anota el gasto físico realizado utilizando las tarjetas magnéticas, propias de la entidad que reporta, en cualquier tipo de moneda. Por lo tanto, no incluye el combustible utilizado por el centro informante para realizar una actividad por la que recibió combustible, en tarjeta magnética propiedad de otra entidad. Corresponde a la suma de los litros de combustible que aparecen registrados en los vales del mes que se reporta, como resultado del combustible despachado a todos los vehículos de la entidad, en los servicentros establecidos al efecto. Dicha cantidad será transferida al modelo en miles de litros.
Entregado para consumo	Constituye el equivalente físico de combustible total entregado a otra entidad consumidora, a través de las tarjetas previamente cargadas por el informante como responsable de dicha carga. Se refiere a lo adquirido en cualquier tipo de moneda.
Saldo final	Cantidad de combustible físico, equivalente al saldo final en tarjetas magnéticas, no utilizado una vez concluido el mes. Es la suma del saldo final que en valor aparece en el último vale correspondiente a cada tarjeta al concluir el mes, llevado a litros según el precio del combustible que se reporta y que finalmente se registrará en el modelo en miles de litros. Esta cantidad constituye a su vez la existencia final en el registro de la entidad para el próximo mes.
Carga para el próximo mes	Cantidad de combustible físico equivalente que se reporta como saldo final en tarjeta, pero que fue cargado parcial o totalmente a cuenta de lo asignado para el próximo mes.

Análisis y Diseño de la solución

Consumo acumulado real	Se anotan de forma acumulada las cantidades destinadas en el año actual a satisfacer los consumos de los productos utilizados por la entidad independientemente de la unidad monetaria mediante la cual se adquiere el producto (moneda nacional o divisas).
Consumo acumulado año anterior	Se anota el consumo real acumulado registrado hasta igual período del año anterior, considerando esta columna bajo las mismas especificaciones que aparecen en la columna 16. Esta columna coincide con la cifra reportada en la columna de consumo real del modelo de igual período del año anterior o con la reportada en los ajustes correspondientes al mismo, siempre y cuando la entidad no haya experimentado cambios institucionales, en cuyo caso se reporta esta columna ajustándola a la estructura actual.

Tabla 17 Tabla de especificación de las medidas

2.2.8 Esquema de Seguridad

El esquema de seguridad estará protegido por los niveles de acceso al medio, particularmente 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 base de datos.
- **Presentación:** Consolas de administración.

La solución debe tener una Interfaz de Administración Gráfica (GUI) en la consola de administración, la cual proporciona una visualización de la topología de red en el editor de políticas, que no es más que una representación de los objetos de red definidos en el editor de políticas y la relación entre dichos objetos y la red. Para la auditoría del tráfico entrante y saliente, así como para la generación de informes de eventos e intentos de ataque, se debe utilizar un registro del tipo LOG FILE como un evento que deba ser registrado. El formato de las entradas del registro pedidas por una regla es determinado por el tipo del

registro especificado en la ella; todo esto sin alterar el rendimiento de acceso a los equipos protegidos por la solución.

La configuración y pruebas de los equipos deberán efectuarse en las instalaciones de la ONE, conjuntamente con un personal técnico de Informática, de acuerdo a las especificaciones de configuración y pruebas que serán elaboradas por la DIE.

El mantenimiento preventivo deberá efectuarse mediante una visita semestral durante el período de vigencia de la garantía técnica. El mantenimiento correctivo deberá efectuarse anualmente, a requerimiento de la ONE con tiempos de respuesta por soporte técnico de 4 horas y de solución de 48 horas los 365 días del año, de fácil ubicación vía e-mail o telefónico.

Se deben considerar actualizaciones de software: parches, sistema operativo, etcétera, los mismos que deben ser recomendados por los fabricantes de los productos entregados.

2.2.9 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:** Las salvas se efectúan mensualmente de la información total que posea la base de datos, así lo tiene definido actualmente la organización, certificando en todo momento que exista una copia estricta de la información que está presente en el servidor.
- **Tablas involucradas:** Las tablas que se involucran en la realización de las salvas son la tabla de hecho Balance de consumo de portadores energéticos y las tablas de las dimensiones que se realizan una sola vez.
- **Backups existentes:** Actualmente no existen Backups en esta área.

1. Periodicidad de reemplazo de los Backups: Se realizan los reemplazos de Backups cada 1 año.

2. Periódicas de las pruebas a los Backups: El estado de los Backups se chequea mensualmente, mediante pruebas de rendimiento y flexibilidad.

2.3 Conclusiones parciales

En el presente capítulo se realizó un estudio más detallado relacionado con el tema de energía para desarrollar la solución, por lo que se define como tema de análisis Energía. Se identificaron las necesidades de los usuarios, los requerimientos funcionales y no funcionales, las reglas del negocio necesarias para el desarrollo del mercado de datos y los roles del sistema: el analista y el administrador, cada uno con sus permisos definidos. Además se confeccionó toda la parte del diseño con la descripción de las tablas de hechos y dimensiones. Se realizó la matriz BUS y el modelo de datos del mercado.

CAPÍTULO # 3: IMPLEMENTACIÓN Y PRUEBAS DE LA SOLUCIÓN

3.1 Introducción del capítulo

En este capítulo se describe el modelo de datos físico, el cual en su estructura de datos muestra los esquemas, las tablas, restricciones y secuencias, las llaves primarias y foráneas así como los índices de la combinación de datos. Se muestran los usuarios, roles y permisos de los usuarios al mercado de datos. Además se crea la guía de implantación en la cual a groso modo se describe el proceso para la instalación del mercado de datos y se dan los requerimientos de hardware y software necesarios para el sistema. Se validan los requisitos por el cliente.

3.2 Modelos de Datos Físico

Un modelo de datos es un lenguaje utilizado para la representación de una base de datos. Por lo general, permite detallar las estructuras de datos, las restricciones de integridad y las operaciones de manejo de datos (agregado, borrado, modificado y recuperación de los datos). El modelo de datos físico es una estructura de datos de bajo nivel implementadas dentro del propio manejador.

Como resultado se obtienen los scripts: DDL¹⁷, DML¹⁸ y DCL¹⁹, que contienen la estructura, carga de datos y políticas de seguridad del mercado de datos. A continuación se presentan ejemplos de dichos scripts así como la estructura que presentan los datos:

- **DDL_Dimensiones el cual contiene la estructura física de las dimensiones del mercado de datos. Por ejemplo:**

- o `CREATE TABLE "Dimensiones".dim_sindicato (`

¹⁷ Lenguaje de Definición de Datos.

¹⁸ Lenguaje de Manipulación de Datos.

¹⁹ Lenguaje de Control de Datos.

Implementación y pruebas de la solución

- "sindicato_ID" integer DEFAULT
nextval(("Dimensiones"."dim_sindicato_sindicato_id_seq"::text)::regclass) NOT NULL,
 - "SINDICATO_CODIGO" character(2) NOT NULL,
 - "SINDICATO_DESCRIPCION" character varying(45) NOT NULL
 -) WITHOUT OIDS;
- **DDL_Hecho para crear la estructura física del hecho en el mercado de datos. Por ejemplo:**
- ALTER TABLE ONLY hech_balance_consumo_portadores_energ
 - ADD CONSTRAINT "PK5" PRIMARY KEY (organismo_est_id, organismo_reg_id, cae_id, nae_id, ci_id, entidad_consume_id, "sindicato_ID", indicador_id, dpa_est_id, dpa_reg_id, subordinacion_id, esfera_id, ff_id, temporal_id);
- DML_completo se utiliza para cargar los nomencladores del mercado de datos. Un ejemplo del contenido del script se pone a continuación:
- INSERT INTO dim_cae (cae_id, "SECTOR_DESCRIPCION", "SECTOR_CODIGO", "RAMA_CODIGO", "RAMA_DESCRIPCION", "SUBRAMA_CODIGO", "SUBRAMA_DESCRIPCION")

VALUES (528, 'Industria', '01', '0101', 'Energía eléctrica', '010100', 'Energía eléctrica');
- DCL_Dimensiones que contienen los aspectos de seguridad de la base de datos para las dimensiones. Un ejemplo del contenido se pone a continuación:
- ALTER TABLE "Dimensiones".dim_cae OWNER TO "Administrador";
 - GRANT ALL ON TABLE "Dimensiones".dim_cae TO "Administrador";
 - GRANT SELECT ON TABLE "Dimensiones".dim_cae TO "Consultante";
- GRANT SELECT, UPDATE, INSERT, DELETE ON TABLE "Dimensiones".dim_cae TO "Integrador";
- **DCL_Hecho que contienen los aspectos de seguridad de la base de datos para el hecho.**
- A continuación se muestra un fragmento del código:

Implementación y pruebas de la solución

- ALTER TABLE "Hechos".hech_balance_consumo_portadores_energ OWNER TO "Administrador";
- GRANT ALL ON TABLE "Hechos".hech_balance_consumo_portadores_energ TO "Administrador";
- GRANT SELECT ON TABLE "Hechos".hech_balance_consumo_portadores_energ TO "Consultante";

```
GRANT SELECT, UPDATE, INSERT, DELETE ON TABLE "Hechos".hech_balance_consumo_portadores_energ TO "Integrador";
```

3.2.1 Estructuras de Datos

Una estructura de datos es la forma de organizar los datos con el objetivo de facilitar su manejo, por tanto, para el desarrollo del mercado de datos están estructurados por tablas y esquemas, donde los esquemas están organizados por hechos y dimensiones y las tablas están organizadas por políticas de indexado e índices bien definidos.

A continuación se muestra una tabla con los esquemas y tablas del mercado de datos, las restricciones, secuencias y los índices que se utilizan para el desarrollo del mismo:

Esquema	Tablas
Dimensiones	dim_ffinanciamiento
Dimensiones	dim_esfera
Dimensiones	dim_subordinacion
Dimensiones	dim_temporal
Dimensiones	dim_dpa
Dimensiones	dim_indicador
Dimensiones	dim_centro_informante
Dimensiones	dim_empresa_consume
Dimensiones	dim_nae

Implementación y pruebas de la solución

Dimensiones	dim_cae
Dimensiones	dim_organismo
Dimensiones	dim_sindicato
Hechos	hech_balance_consumo_portadores_energ

Tabla 18 Esquemas y Tablas

Restricciones y Secuencias

Una restricción es un estado que exige el respeto de ciertas condiciones en la base de datos. Proporcionan una técnica para efectuar políticas en el mercado de datos. Las restricciones limitan los datos que pueden ser almacenados en las tablas. Habitualmente se especifican utilizando expresiones que dan como resultado un valor booleano, mostrando si los datos satisfacen la restricción o no. No son parte formal del modelo, pero se incluyen porque juegan el rol de organizar mejor los datos. El presente trabajo contiene las siguientes llaves primarias y llaves foráneas:

Llaves Primarias	cae_id
	ci_id
	dpa_id
	entidad_consume_id
	esfera_id
	ff_id
	indicador_id
	nae_id
	organismo_id
	sindicato_ID
	subordinacion_id
	temporal_id

Tabla 19 Llaves primarias

Llaves foráneas	Refdim_cae10
	Refdim_centro_informante12
	Refdim_dpa17
	Refdim_dpa18
	Refdim_entidad_consume13
	Refdim_esfera21
	Refdim_ffinanciamiento22
	Refdim_indicador15
	Refdim_nae11
	Refdim_organismo7
	Refdim_organismo9
	Refdim_sindicato14
Refdim_subordinacion20	
Refdim_temporal19	

Tabla 20 Llaves foráneas

Índices

Implementación y pruebas de la solución

El índice es una estructura de datos que incrementa la rapidez de las instrucciones, permitiendo un rápido acceso a los registros de una tabla. Al aumentar radicalmente la rapidez de acceso, se suelen utilizar sobre aquellos campos en los cuales se hagan búsquedas habituales. Funciona similar al índice de un libro, almacenando parejas de elementos: el elemento que se desea indexar y su posición en la base de datos. Se utiliza para la creación del mercado de datos la política de indexado por llave primaria.

Índice	Esquema	Tabla	Tipo de índices
PK4	Dimensiones	dim_organismo	btree
PK5	Hechos	hech_balance_consumo_ portadores_energ	btree
PK6	Dimensiones	dim_cae	btree
PK7	Dimensiones	dim_nae	btree
PK8	Dimensiones	dim_centro_informante	btree
PK9	Dimensiones	dim_entidad_consume	btree
PK10	Dimensiones	dim_sindicato	btree
PK11	Dimensiones	dim_indicador	btree
PK13	Dimensiones	dim_dpa	btree
PK14	Dimensiones	dim_temporal	btree
PK15	Dimensiones	dim_subordinacion	btree
PK16	Dimensiones	dim_esfera	btree
PK17	Dimensiones	dim_ffinanciamiento	btree

Tabla 21 : Índices de las tablas del mercado de datos

3.2.2 Usuarios y Privilegios

Implementación y pruebas de la solución

Los usuarios son creados para mantener el control de las personas que acceden al sistema por lo que a cada uno de ellos se le asignan roles y privilegios a la hora de manejar la información que se encuentra en el sistema, la tabla muestra a continuación los usuarios que participan en el sistema y los roles que desempeña el mercado de datos:

Usuarios	Roles	Permisos	Descripción
Consultante	Usuario	Lectura	Analiza y visualiza los datos.
Administrador	Administrador	Control Total	Administra la base de datos.

Tabla 22 Usuarios, roles y permisos

3.2.3 Carga de nomencladores

Los nomencladores constituyen las dimensiones encontradas después de haber analizado de manera profunda el negocio. La carga se realiza para crear los valores que se utilizarán luego en la realización del mercado de datos y es hecha desde el gestor de bases de datos, debido a que posee implementada una funcionalidad que permite realizar el trabajo de manera fácil, rápida y segura. Los datos a cargar se encuentran recogidos en los archivos de tipo excel (.xls).

A continuación se muestra una tabla con los nomencladores (especificados anteriormente en la especificación de las dimensiones) y el fichero que contiene los valores a cargar:

Nomenclador	Fichero
dim_ffinanciamiento	FFIN.xls
dim_esfera	Esfera.xls
dim_subordinacion	Subordinacion.xls
dim_temporal	Temporal.xls
dim_dpa	DPA.xls
dim_centro_informante	CENTRO_INFORMANTE.xls
dim_empresa_consume	Entidad_consume.xls

Implementación y pruebas de la solución

dim_nae	NAE.xls
dim_cae	CAE.xls
dim_organismo	Organismos.xls
dim_sindicato	Sindicato.xls
dim_indicador	Indicadores.xls

Tabla 23 Nomencladores

3.3 Guía de Implantación

Consiste en una guía para la gestión del riesgo de la seguridad de la información y sirve, por tanto, de apoyo a la implantación del mercado de datos de manera correcta.

3.3.1 Requerimientos

A continuación se plasman los requisitos de software y hardware necesarios para un correcto funcionamiento del mercado de datos:

Software

- Sistema Operativo: Windows XP Service Pack 2
- SGBD: PostgreSQL 8.4
- Herramienta de administración: PgAdmin III.

Hardware

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

3.3.2 Secuencia de Pasos

1. Realizar el diseño del almacén de datos en la herramienta de modelado Visual Paradigm y generar el script del mercado de datos.
2. Instalar el SGBD PostgreSQL 8.4 y el PgAdmin III como herramienta de administración.
3. Se crea una nueva base de datos utilizando el PgAdmin III y se carga el script DDL_Dimensiones generado en el Visual Paradigm, el cual contiene la estructura física de las dimensiones del mercado de datos.
4. Se carga el script DDL_Hecho para crear la estructura física del hecho en la base de datos.
5. Se carga el script DML_completo y se ejecuta para cargar los nomencladores del mercado de datos y ya se cuenta con una base de datos creada estructuralmente con todos sus nomencladores cargados.
6. Luego se procede a cargar los script DCL_Dimensiones y DCL_Hecho que contienen los aspectos de seguridad de la base de datos y ya se cuenta con una BD disponible para ser utilizada en los procesos de extracción, transformación y carga.

Después de cargar todos los scripts antes mencionados se realizaron pruebas al mercado de datos para garantizar que cumplía con los requerimientos del usuario y validar de esta manera la solución propuesta.

3.4 Validación y pruebas

Las pruebas se realizan para garantizar que el producto que se va a entregar esté con la calidad requerida. Luego de verificar si es válido, el producto es validado para su posterior entrega al usuario. El presente trabajo cuenta con listas de chequeo para el control de la calidad.

3.4.1 Listas de Chequeo Análisis

Se entiende por lista de chequeo (cheks-list) a un listado de preguntas, en forma de cuestionario, que sirve para verificar el grado de cumplimiento de determinadas reglas establecidas a priori, con un fin establecido. (Bichachi, 2006)

Las listas puestas a continuación tienen el propósito de garantizar que los artefactos del análisis posean la calidad requerida. Ellas son:

Implementación y pruebas de la solución

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

Como resultado de la aplicación de las listas de chequeo se obtuvo lo siguiente (Ver gráfica de resultados en Anexos):

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

Tabla 24 resultados de la aplicación de las listas de chequeo

3.4.2 Validación de requisitos por el cliente

Se realizó el encuentro con la compañera Elena Leonila Fernández García: cliente del presente trabajo, y el compañero Asnioby Hernández López, los cuales estuvieron de acuerdo con la propuesta de la solución presentada.

3.4.3 Pruebas de Implantación

Casos de Prueba	Pre condición	Pos condición	Resultado esperado
-----------------	---------------	---------------	--------------------

Implementación y pruebas de la solución

Crear base de datos	PostgreSQL 8.4 instalado como gestor de base de datos.	Base de datos creada.	Base de datos creada sobre el gestor de base de datos PostgreSQL 8.4.
Roles y permisos	Base de datos creada.	Base de datos creada con sus roles y permisos.	Roles creados con los respectivos permisos.
Cargar los Nomencladores	Base de datos creada con sus roles y permisos.	Base de datos lista para ser utilizada.	Campos de la base de datos llenos.

Tabla 25 Pruebas de implantación

3.5 Conclusiones parciales

Al concluir el presente capítulo se concluyó que:

- Con la estructura de esquemas y tablas definidas se logró una mejor organización de los datos. Los usuarios creados con sus privilegios cumplen con lo requerido por los usuarios en los requisitos iniciales del sistema.
- La guía de implantación de la BD definida logra la creación correcta de la misma.
- Los nomencladores cargados responden a las especificaciones realizadas por el cliente.
- Las pruebas realizadas confirman el buen funcionamiento de los pasos para la implantación del sistema.

Conclusiones generales

En este trabajo se realizó un estudio de las tecnologías y tendencias actuales en cuanto al desarrollo de mercado de datos. Se analizaron las principales herramientas CASE, SGBD y metodologías que han adquirido un gran auge hoy en día. También se especificaron los artefactos generados a lo largo de la investigación. De todo este estudio realizado se concluyó lo siguiente:

- El uso de los mercados de datos permitió una mejor organización de la información referente al balance de consumo de portadores energéticos, además se centraliza dicha información y se logra mayor rapidez en la realización de las búsquedas de información por parte del personal. Quedó demostrado que es una solución viable para la Oficina Nacional de Estadísticas.
- El análisis del modelo 5073 permitió que se confeccionaran temas de interés para lograr un mejor entendimiento del negocio.
- El diseño e implementación del mercado de datos para los indicadores relacionados con el control de la energía permitió plasmar de manera organizada los hechos, dimensiones y medidas asociadas del mercado.
- Los nomencladores cargados en la BD poseen la estructura especificada por el cliente.
- Las pruebas realizadas al mercado de datos permitieron validar el producto y los artefactos generados, además garantizaron la calidad de dichos artefactos como parte de la solución.
- Las metas propuestas fueron alcanzadas y se le dio cumplimiento al objetivo fundamental de esta investigación.

Recomendaciones

Después de haber concluido el presente trabajo se recomienda:

- Continuar el estudio de las mejoras prácticas en el desarrollo de los mercados de datos.
- Generalizar la aplicación de los mercados de datos a todas las empresas de nuestro país que lo necesiten.

Referencias bibliográficas

Bernabeu, R. D. 2007. Hefesto: Metodología propia para la construcción de un DataWarehouse. [En línea] 2007. [Citado el: 10 de febrero de 2010.] <http://www.dataprix.com/es/hefesto-metodologia-propia-para-la-construccion-un-data-warehouse>.

Camaleon. 2010. CENTROCAMALEON.COM. [En línea] 2010. [Citado el: 12 de febrero de 2010.] <http://www.centrocamaleon.com/v1/?p=233>.

Cristhian Herrera. 2007. Autentia S.L. *Autentia S.L.* [En línea] 30 de 10 de 2007. [Citado el: 4 de marzo de 2010.] <http://www.adictosaltrabajo.com/tutoriales/tutoriales.php?pagina=datawarehouse4>.

Curto, J. 2008. CIF vs MD: Dos enfoques clásicos en el diseño de la arquitectura de una Data Warehouse. [En línea] 2008. <http://bi-businessintelligence.blogspot.com/2009/01/cif-vs-md-dos-enfoques-clasicos-en-el.html>.

DATEC, Colectivo de Autores de. 2009. METODOLOGÍA PARA EL DESARROLLO DE SOLUCIONES DE ALMACENES DE DATOS E INTELIGENCIA DE NEGOCIO EN CENTALAD. Ciudad de la Habana : s.n., 2009.

EAFIT, Universidad. 2005. redalyc.uaemex.mx. [En línea] 2005. [Citado el: 12 de febrero de 2010.] <http://redalyc.uaemex.mx/redalyc/pdf/215/21513706.pdf>.

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

—. 1992. *“Building the Data Warehouse”* . 1992.

Inmon, W. H. 1995. "What is a Data Warehouse?" 1995.

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

Referencias bibliográficas

Medina, Raimundo Adrian Acosta. 2009. Sistema para la gestión de la información referente a los problemas tanto de software como de hardware que existen en las áreas productivas de la UCI. Ciudad de La Habana : s.n., 2009.

Outeiro, Ronda de. 2007. Sinnexus. [En línea] 2007. [Citado el: 20 de febrero de 2010.] http://www.sinnexus.com/business_intelligence/datawarehouse.aspx.

Pérez, Angélica Rocío Mondragón. 2002. Instituto Nacional de Estadística y Geografía(INEGI). [En línea] 2002. [Citado el: 21 de febrero de 2010.] <http://www.inegi.org.mx/inegi/contenidos/espanol/prensa/contenidos/articulos/economicas/indicadores.pdf>.

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

Sánchez, L. Z Zepeda. 2008. Metodología para el Diseño Conceptual de Almacenes de Datos. Valencia : s.n., 2008.

Sons, Jon Wiley &. 2008. Kimball, R., Reeves, L., Ross, M., Thornthwaite, W. The data warehouse lifecycle toolkit. 2008.

Borja, Ronda. 2002. iWorld - La empresa multidimensional: OLAP . [En línea] 2002. [Citado el: 15 de marzo de 2010.] <http://www.idg.es/iWorld/articulo.asp?id=143456>.

Ing. Bernabeu Ricardo Dario. 2009. Dataprix. [En línea] 17 de Enero de 2009. [Citado el: 20 de febrero de 2010.] <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/i-data-warehousing-investigacion-y-sistematizacion-concepto-13>.

Referencias bibliográficas

Juan David Peña Rivera, Jesús Armando Suárez Daza. 2005. Pontificia Universidad Javeriana. *Tesis "Utilización de información histórica para decisiones empresariales"*. [En línea] Junio de 2005. [Citado el: 10 de marzo de 2010.] <http://www.javeriana.edu.co/biblos/tesis/ingenieria/Tesis204.pdf>.

Orallo, José Hernández. 2003. Análisis y Extracción de Conocimiento en Sistemas de Información: Datawarehouse y Datamining. [En línea] 2003. [Citado el: 18 de febrero de 2010.] <http://users.dsic.upv.es/~jorallo/cursoDWDM>. **Díaz, Mg. Samuel Oporto. 2007.** Metodologías para el Data Warehousing. [En línea] 2007. [Citado el: 18 de marzo de 2010.] www.wiphala.net/courses/dwh/ICSI243/.../class_12_lifecycle_dwh.ppt.

Outeiro, Ronda de. 2007. Sinnexus. [En línea] 2007. [Citado el: 10 de marzo de 2010.] http://www.sinnexus.com/business_intelligence/datamart.aspx.

Bobadilla, Dra. Faria Bernui. 2009. Evaluación de la Calidad - Indicadores. [En línea] 2009. [Citado el: 1 de abril de 2010.] http://sisbib.unmsm.edu.pe/BVrevistas/anales/v58_n1/ecalidad1.htm.

Curto, Josep. 2009. BI - Business Intelligence - CIF vs MD : Dos enfoques clásicos en el diseño de la arquitectura de un Data Warehouse. [En línea] 12 de enero de 2009. [Citado el: 28 de marzo de 2010.] <http://bi-businessintelligence.blogspot.com/2009/01/cif-vs-md-dos-enfoques-clasicos-en-el.html>.

Bolaños, S. (2006). *Universidad Autónoma del Estado de Morelos*. Recuperado el 3 de abril de 2010, de <http://www.uaem.mx/posgrado/mcruz/cursos/miic/mysql16.pdf>

Casanova, J. (1 de Julio de 2009). *PostgreSQL*. Recuperado el 25 de febrero de 2010, de <http://archives.postgresql.org/pgsql-es-fomento/2009-07/msg00000.php>

Oracle. (2008). *Ooracle*. Recuperado el 2010 de abril de 15, de <http://www.oracle.com/us/products/mysql/index.html>

Oracle. World's Largest Data Warehouse. (2006). Recuperado el 20 de febrero de 2010, de <http://www.sun.com/service/refarch/datawarehouse/>

Bibliografía

DATEC, Colectivo de Autores de. 2009. Metodología para el desarrollo de soluciones de almacenes de datos e inteligencia de negocio en DATEC. Ciudad de la Habana : s.n., 2009.

EAFIT, Universidad. 2005. redalyc.uaemex.mx. [En línea] 2005. [Citado el: 20 de febrero de 2010.] <http://redalyc.uaemex.mx/redalyc/pdf/215/21513706.pdf>.

Inmon, W. H. 1992. "Using the Data Warehouse" . 1992.

Inmon, W. H. 1992. "Building the Data Warehouse" . 1992.

Inmon, W. H. 1995. "What is a Data Warehouse?" 1995.

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

Sánchez, L. Z Zepeda. 2008. Metodología para el Diseño Conceptual de Almacenes de Datos. Valencia : s.n., 2008.

Sons, Jon Wiley &. 2008. Kimball, R., Reeves, L., Ross, M., Thornthwaite, W.The data warehouse lifecycle toolkit. 2008.

Glosario de términos

- **DATEC:** Centro de Tecnologías de Gestión de Datos.
- **DCL:** Lenguaje de Control de Datos.
- **DDL:** Lenguaje de Definición de Datos.
- **DML:** Lenguaje de Manipulación de Datos.
- **MOLAP:** Procesamiento Analítico Multidimensional en Línea.
- **ROLAP:** Procesamiento Analítico Relacional en Línea.
- **HOLAP:** Procesamiento Analítico Híbrido en Línea.

Anexos

Anexo 1 Lista de chequeo Evaluación de áreas de la organización

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

crítico	1. ¿Se definió el alcance de las áreas de la organización?	0			
crítico	2. ¿Se definió el objetivo principal de las áreas de la organización?	0			
	3. ¿Se definieron las características de la organización?	0			
	4. Se utilizó un esquema para representar la información obtenida en la evaluación del área?	0			
Semántica del documento					
Peso	Indicadores a Evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
Crítico	1. ¿No se han identificado errores ortográficos?	0			

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

Anexo 2 Lista de chequeo Herramienta para la recolección y análisis de la información

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

crítico	3. ¿Está el documento acorde con la plantilla estándar del proyecto o del expediente de proyecto utilizado?	0			
Crítico	4. ¿Contiene las secciones obligatorias definidas en el expediente utilizado por el proyecto?	0			
Elementos definidos por la metodología					

Peso	Indicadores a Evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
Crítico	¿Ha sido identificado el tema de análisis?	0			
	¿Se ha descrito brevemente el tema de análisis identificado?	0			
Crítico	¿Las necesidades del cliente han sido identificadas?	0			
	¿Los niveles de acceso para cada pedido de información se han identificado correctamente?	0			

	¿Han sido 0 identificados los roles que tendrán acceso a ver la información de cada pedido?			
Critico	¿Se ha 0 determinado la periodicidad con que se recoge la información de cada pedido?			
Critico	¿Han sido 0 clasificados los pedidos de información de acuerdo a la información que manejan?			

Critico	¿Los pedidos de información se han agrupado en casos de uso informativos, según un criterio que los describa?	0			
Critico	¿Han sido determinados los criterios para agrupar los pedidos de información en casos de uso del sistema?	0			
Critico	¿Se han agrupado los pedidos de información en casos de uso del sistema?	0			
Critico	¿Todas las variables de entrada han sido identificadas?	0			

Critico	¿Se han identificado jerarquías de las variables de entrada que las poseen?	0			
Critico	¿Se han determinado los niveles de cada jerarquía?	0			
Critico	¿Se ha reflejado a que caso de uso del sistema pertenece cada variable de entrada?	0			
	¿Se ha reflejado el tema de análisis al que pertenece cada variable de entrada?	0			

Critico	¿Todas las variables de salida han sido identificadas?	0			
Critico	¿Se ha definido la unidad de medida en que se va a medir cada variable de salida?	0			
Critico	¿Se ha determinado el tipo de variable para cada una de las variables de salida identificadas?	0			
	¿Se ha identificado alguna variable calculable?	0			

Critico	¿Se ha determinado la fórmula para calcular todas aquellas variables calculables que fueron identificadas?	0			
Critico	¿Ha sido identificado el tipo de aditividad que presenta cada variable de salida?	0			
Critico	¿Se identificaron los tipos de operaciones aditivas que se pueden realizar?	0			

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

Anexo 3 Lista de chequeo Especificación de Requisitos

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

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

	3. ¿Se ha asegurado de capturar lo esencial de cada requisito registrado?	0			
	4. ¿Están todos los requisitos redactados de forma simple y clara para aquellos que vayan a consultarlo en un futuro?	0			
	5. ¿Debería especificarse algún requisito con más detalle?	0			
	6. ¿Debería especificarse algún requisito con menos detalles?	0			

	7. ¿Ha definido el flujo de información de manera adecuada para el problema de dominio? (entrevistas, encuestas, equipos de discusión, grupos de trabajo).	0			
crítico	8. ¿Ha identificado todas las funciones que el usuario debe hacer?	0			
	9. ¿Ha definido los límites del sistema?	0			

crítico	10. ¿Todos los requisitos identificados se centran en lo que el sistema debe hacer y no como el sistema debe hacerlo?	0			
crítico	11. ¿Se han enumerado los requisitos incluso los que se derivan de otros requisitos?	0			
	12. ¿No aparece un mismo requisito en más de un lugar del documento de especificación?	0			

	13. ¿No existe contradicción entre lo especificado por un requisito y lo especificado por otro?	0			
crítico	14. ¿Cada requisito es verificable? (Un requisito se dice que es verificable si existe algún proceso no excesivamente costoso por el cual una persona o una máquina pueda chequear que el software satisface dicho requerimiento).	0			

Crítico	15. ¿Han sido definidos todos los datos de entrada y salida?	0			
	16. ¿Cada funcionalidad del sistema ha sido representada gráficamente?	0			
crítico	17. ¿Ha identificado los requerimientos de software y de hardware?	0			
crítico	18. ¿Han sido identificadas las restricciones de diseño e implementación?	0			

crítico	19. ¿Han sido identificadas las restricciones de interfaz externa?	0			
crítico	20. ¿Los requerimientos de soporte y usabilidad se han identificados?	0			Los requisitos de usabilidad se identificaron, pero de soporte no se detectó ningún requisito.
crítico	21. ¿Se han identificado los requerimientos de seguridad (confidencialidad, integridad, disponibilidad)?	0			

	22. ¿Se puede trazar cada requisito al origen en el entorno del problema, (caso de uso del sistema)?	0			
crítico	23. ¿Todos los cambios en los requisitos han sido controlados?	0			
	24. ¿Existe correspondencia entre el modelo de caso de uso sistema y las especificaciones de requerimientos?	0			
Semántica del documento					
Peso	Indicadores a Evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios

Crítico	1. ¿Ha identificado errores ortográficos?	0			
Crítico	2. ¿Se entiende claramente lo que se ha especificado en el documento?	0			
	3. ¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?	0			
	4. ¿El total de páginas que aparecen en las reglas de confidencialidad coincide con el total de páginas que tiene el documento?	0			

Anexo 4 Resultados de la aplicación de las listas de chequeo a los artefactos.

