

UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS

Facultad 8

Análisis, diseño e implementación del mercado de
datos para los indicadores sobre el Plan Turquino
para la Oficina Nacional de Estadísticas.

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

Autores: Karina Fernández Fernández
Carlos Navarro Fonseca

Tutor: Ing. José Rodríguez Alarcón

CENTALAD

Junio 2010

“Año 52 de la Revolución”

*“Sólo aquellos que se arriesgan a ir muy lejos, pueden llegar
a saber lo lejos que pueden ir”*

T. S. Elliot

Declaración de autoría

Declaramos que somos los únicos autores del trabajo “Análisis, diseño e implementación del mercado de datos para los indicadores sobre el Plan Turquino para la Oficina Nacional de Estadísticas” y autorizamos a la Facultad 8 de la Universidad de las Ciencias Informáticas a hacer uso del mismo en su beneficio.

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

Autores: _____

Karina Fernández Fernández

Carlos Navarro Fonseca

Tutor: _____

Ing. José Rodríguez Alarcón

Agradecimientos

De Karina:

Ante todo agradezco a mis padres Raisa Fernández y Héctor Fernández Menduiña, que son la principal razón por la cual ingresé en la UCI, y por darme su apoyo y ayuda durante mis años de estudio.

A mi novio y compañero de tesis que me apoyó, me dio las fuerzas para salir de los momentos malos y me regaló muchos momentos buenos.

Agradezco además a los demás miembros de mi familia y a todas las amistades, especialmente a Adriana, que me ayudó a despejar en los momentos de tensión.

Agradezco también a mis suegros, Marelis y Sergio Mario, que siempre me aconsejaron y animaron a seguir hasta el final.

Finalmente, pero no menos importante, agradezco a nuestra clienta Elena, que nos recibió todas las veces que fuimos a verla con dudas.

De Carlos:

Agradezco a mis padres Marelis Fonseca Vizcaíno y Sergio Mario Navarro Navarro por todo el apoyo, amor y cariño que me han dado durante mi formación.

Agradezco a mi familia en general por haber confiado en mí y por toda la ayuda que me ha brindado.

Agradezco también a mi novia y compañera de tesis por todo el apoyo que me ha dado durante estos 5 años de carrera.

Agradezco a Raisa y a Héctor por haberme acogido en la familia como tal y por toda la confianza que han depositado en mí durante todo este tiempo.

Agradezco a mis amistades que durante todo este tiempo han compartido conmigo momentos buenos y no tan buenos, a las personas que de una forma u otra me han brindado su ayuda durante estos 5 años.

Dedicatoria

Karina:

Dedico este título a mi novio, a mi familia, a mis amigos y sobre todo a mis padres que son lo más importante en mi vida.

Carlos:

Les dedico este título a mis padres, a mi familia, a mi novia y amigos, ellos son la razón por la que se ha hecho posible este sueño.

Resumen

El presente trabajo surge a partir de la necesidad de la Oficina Nacional de Estadísticas de actualizar el sistema para almacenar la información que se recoge en dicho centro. En aras de optimizar la velocidad de respuesta del sistema a las consultas que se realicen sobre la información, se tiene como tarea la creación de un Almacén de Datos, y dentro de este, un Mercado de Datos para los indicadores que se recogen en las zonas montañosas del país. El análisis, diseño y posterior implementación de dicho Mercado de Datos constituye la esencia de este trabajo.

Palabras claves: Almacén de Datos, Mercado de Datos, indicadores.

ÍNDICE

Introducción	1
Capítulo 1: Fundamentación teórica	5
1.1. Introducción	5
1.2. Marco teórico.....	5
1.2.1. Almacenes de datos	5
1.2.2. Características del AD	6
1.2.3. Beneficios del uso de los AD	6
1.2.4. Principales procesos de un AD.....	6
1.2.5. Inteligencia de Negocio	7
1.2.6. Bases de datos multidimensionales	7
1.2.7. Proceso OLTP	8
1.2.8. Proceso OLAP	8
1.2.9. Tipos de OLAP	9
1.2.10. OLTP contra OLAP	9
1.2.11. Mercado de Datos	10
1.2.12. Metadatos.....	10
1.2.13. Granularidad.....	11
1.2.14. Estructura final de un AD.....	11
1.2.15. Modelo 0024-03 para los Indicadores del Plan Turquino	12
1.3. Tecnologías y tendencias actuales.....	12
1.3.1. Sistemas existentes a nivel mundial.....	12
1.3.2. Fundamentación de la metodología a utilizar	13
1.3.3. Tecnologías y herramientas propuestas para el desarrollo del sistema	15
1.3.3.a Sistema Gestor de Bases de Datos	16
Oracle.....	16
MySQL	16

SGBD a utilizar en el presente trabajo: PostgreSQL.....	17
1.3.3.b Herramientas CASE	18
EasyCASE	18
ER/Studio 8.0	18
Rational Rose.....	19
Herramienta de modelado a utilizar: Visual Paradigm.....	19
1.4. Conclusiones parciales.....	20
Capítulo 2: Análisis y diseño de la solución	21
2.1. Introducción	21
2.2. Análisis	21
2.2.1. Definición del negocio	21
2.2.2. Tema de análisis	21
2.2.3. Roles y permisos.....	22
2.2.4. Reglas del negocio.....	23
2.2.5. Necesidades de los usuarios.....	26
2.2.5.a. Requisitos de información	27
2.2.5.b. Requisitos multidimensionales	28
2.2.5.c. Requisitos funcionales.....	29
2.2.5.d. Requisitos no funcionales.....	29
2.2.5.e. CU del sistema	31
2.3. Diseño	34
2.3.1. Matriz BUS	34
2.3.2. Modelo de datos.....	35
2.3.2.a. Dimensiones identificadas.....	35
2.3.2.b. Hechos identificados	37
2.3.2.c. Medidas.....	38
2.3.3. Diseño de las dimensiones y la tabla de hechos.....	38
2.3.4. Esquema de seguridad	39

2.3.5. Política de respaldo y recuperación.	40
2.4. Conclusiones parciales.....	40
Capítulo 3: Implementación y pruebas.....	42
3.1. Introducción.....	42
3.2. Modelo de Datos Físico.....	42
3.2.1. Estructuras de Datos.....	42
3.2.2. Usuarios y privilegios.....	47
3.2.3. Carga de nomencladores.....	47
3.3. Guía de implantación.....	50
3.3.1. Requerimientos.....	50
3.3.2. Secuencia de pasos para la instalación de la BD.....	50
3.4. Validación y pruebas.....	51
3.4.1. Lista de chequeo de análisis.....	51
3.4.2. Validación de requisitos por el cliente.....	52
3.4.4. Pruebas de implantación.....	52
3.5. Conclusiones parciales.....	53
Conclusiones.....	54
Recomendaciones.....	55
Referencias bibliográficas.....	56
Bibliografía.....	57
Anexos.....	59
Anexo 1 Tabla de indicadores y sus respectivos CUI.....	59
Anexo 2 Modelo estadístico del Plan Turquino.....	63
Anexo 3 Ejemplo de script de seguridad aplicado a la tabla dim_dpa.....	64
Anexo 4 Ejemplo de script DDL utilizado para crear la tabla dim_dpa.....	65
Anexo 5 Ejemplo de script DML utilizado para cargar los nomencladores de la tabla dim_dpa.....	66
Anexo 6 Lista de Chequeo de la Especificación de requisitos.....	67

Anexo 7 Lista de Chequeo para la Evaluación de las áreas de la organización	75
Anexo 8 Lista de Chequeo de la Herramienta para la recolección y análisis de la información	78
Anexo 9 Resultados de la aplicación de las Listas de Chequeo a los artefactos del análisis	83

ÍNDICE DE FIGURAS

Fig. 1 Granularidad & detalle	11
Fig. 2 Metodología de Kimball	14
Fig. 3 Relación de los Grupos de Trabajo con los diferentes Flujos del ciclo de vida.....	15
Fig. 4 Tablas de códigos de la ONE para las provincias y sus municipios.	26
Fig. 5 Estructura del CPCU v1.1.....	26
Fig. 6 Diagrama de CU del sistema	34
Fig. 7 Diseño de las dimensiones y tablas de hechos de la solución	39
Fig. 8 Privilegios del usuario ADMINISTRADOR_ETL.....	64
Fig. 9 Privilegios del usuario AnalistaPT.....	64

ÍNDICE DE TABLAS

Tabla 1 OLTP & OLAP.....	10
Tabla 2 Matriz BUS.....	35
Tabla 3 Esquema de seguridad propuesto	40
Tabla 4 Tablas creadas para la solución	43
Tabla 5 Ejemplos de los índices seleccionados.....	47
Tabla 6 Usuarios y privilegios establecidos	47
Tabla 7 Resultados de las Listas de Chequeo de análisis aplicadas.....	52
Tabla 8 Pasos de implantación a los que se les realizaron las pruebas	53

Introducción

En el mundo actual las tecnologías de la información forman parte indispensable en la vida del ser humano. Debido a la gran competencia existente en el mercado internacional entre las compañías dedicadas al desarrollo de dichas tecnologías, surge la necesidad de adaptarse constantemente a los cambios que ocurren en este campo. Las soluciones informáticas traen mejoras considerables para el entorno en el cual se aplican, ahorrando tiempo, recursos y generando ganancias.

Cuba, como país en vías de desarrollo, no está ajena a esta situación, por lo que se ha dado a la tarea de informatizar al país. Con este fin, en el año 1997, se definieron los “Lineamientos estratégicos para la informatización de la sociedad cubana” por el Comité Ejecutivo del Consejo de Ministros de Cuba. A partir de ahí se realiza una serie de acciones en pos de lograr estos objetivos. El Ministerio de la Informática y las Comunicaciones fue el principal encargado de introducir al país dentro del mundo de la tecnología informática.

Como parte de estos cambios las organizaciones del país se dieron a la tarea de actualizar, tanto la tecnología existente en sus centros, como la manera de recoger y gestionar la información.

En el año 1976 se crea el Comité Estatal de Estadísticas, que contaba, para el procesamiento de la información, con una red de procesamiento de datos a nivel nacional. Pero no es hasta el 21 de abril de 1994 que, como resultado de la reorganización de los organismos de la Administración Central del Estado, se crea la Oficina Nacional de Estadísticas (ONE). La misión de dicha organización es garantizar la producción de estadísticas de calidad mediante una correcta captación de las cifras económicas y sociales, así como su difusión de acuerdo a las necesidades del país.

En sus inicios la ONE almacenaba la información en cintas magnéticas. Luego del derrumbe del campo socialista, debido a la situación del país, esta organización se vio obligada a utilizar una tecnología diferente, por lo que toda la información anterior se perdió. Tras varios intentos por mejorar la gestión de las estadísticas se adopta un sistema basado en ficheros en el año 2000. Este incluye el Sistema de Gestión de Estadísticas económicas MICROSET (una aplicación de los años 70) y el Sistema de Gestión de Estadísticas de población SIDEMO/DV SURVEY. Cada centro informante envía los respectivos ficheros a la ONE para que ésta, mediante un grupo de trabajo conformado por especialistas e informáticos, procese dicha información. Los sistemas de gestión que se utilizan actualmente no permiten realizar consultas a la información de manera eficiente ya que solo brindan la posibilidad de hacer reportes estáticos.

Todo esto da paso a la siguiente **situación problemática**: La ONE tiene una gran cantidad de información digitalizada referente a los indicadores sobre el Plan Turquino, recogidos en años anteriores en el MODELO 0024-03. Esta información, para ser accedida, necesita de un especialista en informática que la introduzca en un fichero. Las tecnologías y las herramientas utilizadas para almacenar, recuperar y presentar la información proveniente del MODELO 0024-03 dificultan la integración, centralización y disponibilidad de la información lo cual limita el proceso de toma de decisiones para los órganos del estado.

Por lo anteriormente expuesto se plantea el siguiente **problema a resolver**: ¿Cómo desarrollar un sistema para la integración de datos de los indicadores de las zonas montañosas?

Se definen como **objeto de estudio** los Almacenes de Datos (AD) y como **campo de acción** los Mercados de Datos (MD) estadísticos.

El **objetivo general** de esta investigación es desarrollar el análisis, diseño e implementación del MD para los Indicadores sobre el Plan Turquino para la ONE.

Como **objetivos específicos** se tienen:

1. Elaborar el marco teórico de la Investigación acerca de las principales tendencias de implementación de los Almacenes de Datos, los Mercados de Datos y estudio del arte del tema referente a los Indicadores sobre el Plan Turquino.
2. Realizar el análisis del Modelo 0024-03 de la ONE.
3. Diseñar el MD para los Indicadores sobre el Plan Turquino para el AD de la ONE.
4. Implementar y cargar los clasificadores para el MD de los Indicadores sobre el Plan Turquino para el AD de la ONE.
5. Validar la solución desarrollada mediante la aplicación de Listas de Chequeo y Casos de Prueba como parte del proceso de pruebas.

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

1. Estudio de los temas relacionados con el desarrollo de MD.
2. Estudio de la metodología a utilizar en el desarrollo.
3. Planificación y realización de entrevistas.
4. Identificación de la estructura de usuarios y permisos.
5. Definición de temas de análisis.
6. Identificación de necesidades de información, requisitos funcionales y no funcionales.
7. Modelación de requerimientos.
8. Validación de requerimientos.
9. Definición de requisitos de entrada y de salida.
10. Elección de la granularidad del proceso del negocio.
11. Definición de las dimensiones del MD.
12. Definición de los hechos asociados a las dimensiones.

13. Estructuración del modelo dimensional.
14. Transformación del modelo dimensional al diseño físico.
15. Implementación de la Base de Datos.
16. Montaje de los clasificadores para el MD de los Indicadores sobre el Plan Turquino para el AD de la ONE.
17. Aplicación de las Listas de Chequeo y Casos de Prueba como parte del proceso de pruebas a realizar al MD.

Conociendo todo esto se puede plantear la **idea a defender**: Si se realizara el análisis, diseño e implementación del MD para los Indicadores sobre el Plan Turquino y toda la documentación correspondiente, se lograría una integración más completa de la información, se tendría una mejor disponibilidad de ésta, lo que beneficiaría al proceso de toma de decisiones mediante la informatización de los datos.

Métodos científicos

Como parte de la investigación necesaria para la realización de cualquier trabajo se recomienda la utilización de métodos científicos que servirán como una guía para realizar dicha investigación. Estos métodos no son más que procedimientos que estudian la realidad, la sociedad, la naturaleza y el pensamiento, para poder descubrir su esencia y sus relaciones.

Como **métodos teóricos** se utilizarán:

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

Como **método empírico** se utilizará:

- Entrevista: se aplicará mediante conversatorios con el cliente para comprender lo que este necesita y así definir los requisitos.

El siguiente trabajo estará compuesto por **3 capítulos**:

Capítulo 1: Fundamentación teórica

El capítulo 1 aborda algunos elementos teóricos acerca de la tecnología de AD y los MD. Además se hace referencia a la estructura y principales características de los gestores de bases de datos que se utilizan para el trabajo con estos sistemas y a las herramientas que se van a utilizar. Se exponen además las razones por las cuales se escogió la metodología que se utilizará a lo largo del trabajo.

Capítulo 2: Análisis y diseño de la solución

El capítulo 2 contiene la descripción de los pasos a seguir durante el análisis y el diseño de la solución. Se aborda la especificación de los requisitos que debe cumplir el sistema, así como la realización de un modelo del MD a partir de los indicadores que se seleccionaron. Se realiza el diseño de lo que constituirá el sistema.

Capítulo 3: Implementación y pruebas

El capítulo 3 contiene todos los elementos referentes a la implementación del sistema, las dimensiones y hechos que formarán parte de la posible solución, la descripción de los roles y permisos creados y los pasos para la implantación del sistema. Además se incluirán los resultados de las pruebas realizadas al sistema.

Capítulo 1: Fundamentación teórica

1.1. Introducción

En el siguiente capítulo se abordan algunos elementos teóricos acerca de la tecnología de AD y los MD. Además se hace referencia a la estructura y principales características de los gestores de bases de datos y a las herramientas que se van a utilizar. Se exponen además las razones por las cuales se escogió la metodología que se utilizará a lo largo del trabajo.

1.2. Marco teórico

1.2.1. Almacenes de datos

En la actualidad, dado el vertiginoso desarrollo de las tecnologías de la información y las comunicaciones, las organizaciones a nivel mundial se han dado a la tarea de digitalizar su información, eliminando así su dependencia del trabajo manual y las grandes cantidades de documentos a procesar. Debido a esto fue necesario cambiar la manera de gestionar los datos, creándose nuevas dificultades a la hora de obtener la información correcta o en la cantidad de tiempo que requería el proceso de obtención.

Una de las principales tareas en una organización es la toma de decisiones, que puede verse afectada si no se cuenta con una organización eficaz de los datos. A partir de este problema surge el concepto de Almacenes de Datos, cuyo objetivo fundamental es el de ayudar a los usuarios a entender mejor el pasado, así como planear el futuro ya sea a corto, mediano o largo plazo. Esta tecnología no es más que la integración de las técnicas de bases de datos con las de análisis de datos. Según Ralph Kimball, un AD es “una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis”. (1)

Momentos históricos importantes en los primeros años del desarrollo de los AD:

- 1960s: El Gral. Mills y el Colegio Darmouth, durante un proyecto de investigación, crean los términos de dimensiones y hechos.
- 1983: Teradata crea un sistema de administración de bases de datos diseñado específicamente para el soporte de decisiones.
- 1988: Barry Devlin y Paul Murphy publican el artículo “Una arquitectura para un sistema de negocios e información”, donde introducen el término “Business Data Warehouse”.
- 1991: Bill Inmon publica el libro “Building the Data Warehouse”.
- 1995: Es fundado el “Data Warehousing Institute”, que promueve el uso de los almacenes de datos.
- 1996: Ralph Kimball publica el libro “The Data Warehouse Toolkit”.
- 1997: Oracle 8 es liberado con la opción para crear los esquemas de estrella.

1.2.2. Características del AD

Según William H. Inmon un AD es "...una colección de datos orientados al tema, integrados, no volátiles e historiadados, organizados para el apoyo de un proceso de ayuda a la decisión" (2). Seguidamente se explica el por qué de cada característica.

- **Temático**

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

- **Integrado**

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

- **No volátiles**

La información existente en un AD no puede ser modificada, solo incrementada. Por tanto el período que generalmente cubre un AD en cuanto a la información almacenada es de 2 a 10 años.

- **Historiado**

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

1.2.3. Beneficios del uso de los AD

- Permiten disponer de la información necesaria en poco tiempo.
- Los directivos pueden dedicar más tiempo al análisis de los datos en vez de a su búsqueda.
- Cuando se aplica al marketing permite profundizar en el comportamiento del cliente.
- Debido a la acumulación de datos y experiencia se pueden detectar tendencias y hacer previsiones.
- Facilitan la toma de decisiones estratégicas.

1.2.4. Principales procesos de un AD

Los principales procesos que conforman un AD son: extracción, elaboración y carga. Son generalmente conocidos por sus siglas en inglés ETL (Extraction, Transformation and Loading).

Extracción: obtención de la información de las distintas fuentes tanto internas como externas.

Elaboración: Luego de realizarse el proceso de extracción los datos provenientes de las diferentes fuentes pueden ser incoherentes, tener errores o estar incompletos. Con esto se busca obtener datos lo más precisos, completos, consistentes, interpretables y accesibles posible. Después del proceso de limpieza se lleva a cabo la integración de los datos con el propósito de eliminar problemas de redundancia e identificar las fuentes de datos más fiables. Una vez realizado el proceso de extracción y limpieza se procede a transformar los datos para de esta forma estandarizar los códigos, corregir los datos, eliminar registros duplicados y usar conversiones y combinaciones para generar nuevos campos.

Carga: organización y actualización de los datos y los metadatos en la base de datos.

Si no se realiza un correcto proceso de ETL se pudieran obtener datos incorrectos, lo que afectaría el proceso de toma de decisiones; es por eso que este proceso constituye aproximadamente un 70% del trabajo de la construcción de un AD.

1.2.5. Inteligencia de Negocio

La Inteligencia de Negocio (BI por sus siglas en inglés) surge “para aprovechar toda la información operacional y convertirla en *información estratégica*, útil para tomar decisiones de negocio.” (3)

La función de un sistema de BI es la de “convertir” la información de una enorme cantidad de datos que solo son útiles para el trabajo diario, en datos que pueden proporcionar conocimiento y ayudar a tomar decisiones que, de otra forma, serían muy complicadas. Con una herramienta de BI se pueden generar reportes, informes, gráficos y todo lo necesario para que los interesados puedan analizar y comprender los datos fácilmente y, a partir de ahí y sin perder mucho tiempo, llegar a conclusiones o realizar decisiones en su área de trabajo.

1.2.6. Bases de datos multidimensionales

Los AD son soportados por un modelo multidimensional, a diferencia de los sistemas de bases de datos relacionales que se sustentan en el modelo Entidad-Relación. El modelo multidimensional almacena la información de forma tal que sea de fácil entendimiento por parte del usuario y que, a su vez, garantice 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.

Las **tablas de hechos** son los elementos primarios del modelo dimensional y contienen los valores del negocio. Una fila de la tabla de hechos corresponde a una medida, que no es más que la intersección entre todas las dimensiones. (1)

Las **tablas de dimensiones** son las “compañeras integrales” de las tablas de hechos, y contienen las descripciones textuales del negocio. Las dimensiones juegan un papel vital dentro de los AD puesto que son la fuente de casi todos los reportes y consultas que se le pueden realizar a éste. Debido a esto se les considera la llave para hacer un AD usable y comprensible. (1)

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

- **Esquema estrella:** es la técnica más común. En este esquema existe un único elemento central (tabla de hechos) conectado radialmente con las tablas de dimensiones.
- **Esquema copo de nieve (Snowflake):** es un esquema derivado del de estrella, donde las tablas de dimensiones se ramifican en más puntas.
- **Esquema constelación:** está compuesto por una serie de esquemas de estrella, es decir, una tabla de hechos central con otras auxiliares y sus respectivas tablas de dimensiones.

1.2.7. Proceso OLTP

OLTP (Procesamiento de Transacciones en Línea) permite tener tiempos de respuesta rápidos a la gran cantidad de transacciones que se realizan en una organización. Sin embargo no son eficientes cuando se trata de grandes cantidades de información ni de consultas muy complejas. Estos sistemas están diseñados para soportar actualizaciones debido a que los datos están normalizados, es decir, se han eliminado las redundancias y cualquier problema que pueda crear dificultades al actualizar cierto dato.

Sin embargo, para analizar los datos los usuarios deben examinarlos registro a registro, tomando así grandes cantidades de tiempo cuando se trata de mucha información. Por lo tanto no es un sistema óptimo para ser utilizado en la toma de decisiones. (4)

1.2.8. Proceso OLAP

La tecnología de Procesamiento Analítico en Línea (OLAP) permite un uso más eficaz de los AD en línea, lo que proporciona respuestas rápidas a consultas analíticas complejas e iterativas utilizadas generalmente para sistemas de ayuda para la toma de decisiones. OLAP presenta los datos de manera natural, de manera que los usuarios puedan entender mejor la información y esta se aproveche correctamente.

Las principales características de OLAP son:

- **Rápido:** proporciona la información al usuario a una velocidad constante (mayormente en cinco segundos o menos).
- **Análisis:** realiza análisis estadísticos y numéricos básicos de los datos.

- **Compartida:** permite compartir los datos potencialmente confidenciales a través de una gran cantidad de usuarios, implementando para esto los requerimientos de seguridad necesarios.
- **Multidimensional:** permite ver la información en determinadas vistas o dimensiones.
- **Información:** accede a todos los datos necesarios, donde quiera que estos residan, y mientras no esté limitada por el volumen. (4)

1.2.9. Tipos de OLAP

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

Organización física que se implementa sobre tecnología relacional, pero que tiene algunas facilidades para mejorar el rendimiento. Accede directamente a los datos del AD y soporta técnicas de optimización de accesos para acelerar las consultas, por ejemplo, partición de datos a nivel de aplicación, soporte a la desnormalización y *joins* múltiples. Esta organización tiene como desventaja que es más lenta que las demás estrategias de almacenaje.

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

Usa las bases de datos multidimensionales bajo la premisa que OLAP está mejor implantado almacenando los datos multidimensionalmente. Tiene el mejor tiempo de respuesta, excelente rendimiento y compresión de datos. Tiene la desventaja que es una solución particular, para soluciones con volúmenes de información y cantidad de dimensiones más bien modestos.

- **HOLAP (Procesamiento Analítico en línea Híbrido)**

Combina las arquitecturas ROLAP y MOLAP brindando así lo mejor de las dos: un desempeño superior y gran escalabilidad. Mantiene los registros de detalle en la base de datos relacional y las agregaciones en un almacén MOLAP separado. Es usado generalmente para cubos que requieren rápida respuesta con una gran cantidad de datos. (4)

1.2.10. OLTP contra OLAP

En la tabla siguiente se exponen las diferencias principales entre los dos tipos de procesos, OLTP y OLAP.

OLTP	OLAP
Acceso de miles de usuarios	Acceso de un número de usuarios moderado
Modificación de los datos	No se modifican los datos

Representan el estado de una organización	Guardan el historial de una organización
Estructuras de BD complejas	Estructuras de BD simples
Se eliminan las redundancias (más eficiencia en las transacciones, más dificultades para la navegación del usuario)	Datos redundantes (acceso fácil para el usuario y buenos tiempos de respuesta)

Tabla 1 OLTP contra OLAP

1.2.11. Mercado de Datos

Los AD son los contenedores de la información de todos los departamentos de una organización o empresa, pero por razones de conveniencia para simplificar el trabajo a la hora de construir un AD y agilizar el acceso a su información, surge el concepto de Mercado de Datos. Se dice que un MD es un almacén de datos históricos relativos a un departamento de una organización, por ejemplo al de finanzas o marketing. Por lo tanto un AD puede estar formado por varios MD de los diferentes departamentos de la empresa en que esté implementado. Sin embargo, no siempre es necesaria la existencia de un AD para que existan mercados de datos. En el caso de que la información de la organización no sea suficientemente grande, se puede obviar el AD y crear solo los MD pertinentes a los departamentos que los necesitan.

Beneficios de la utilización de los MD:

- Acelera las consultas al reducir la cantidad de datos a recorrer.
- Estructura los datos para su adecuado acceso por una herramienta.
- Divide los datos para imponer estrategias de control de acceso.
- Segmenta los datos en diferentes plataformas de hardware.
- Permite el acceso a los datos mediante un gran número de herramientas del mercado, logrando así independencia de estas.

1.2.12. Metadatos

Otro concepto dentro de los AD es el de metadatos, que no son más que datos relativos a datos. Estos permiten mantener información acerca de la procedencia de los datos, su fiabilidad, forma de cálculo, etc.

Sus objetivos principales son:

- Ayudar a los usuarios finales ayudándolos a entender la información contenida en el AD en su propio lenguaje de negocio, indicando que información hay y que significa, además de ayudar a construir consultas, informes y análisis.

- Ayudar a los responsables técnicos del AD con respecto a la gestión de la información histórica, la administración del almacén y la elaboración de programas de extracción de los datos entre otros aspectos.

1.2.13. Granularidad

El concepto de granularidad es el que hace posible la arquitectura que se utiliza en el entorno de los AD. Se refiere al nivel de detalle o resumen de las unidades de datos en el AD. Mientras más detalle exista, más bajo será el nivel de granularidad (Fig. 1.a) y a menor detalle, más granularidad (Fig. 1.b). (5)



Fig. 1 Granularidad y detalle

Una organización donde esté implantado un AD puede verse beneficiada en muchas maneras de este, ya que debido a la granularidad de los datos, estos pueden ser reutilizados y servir para los propósitos y necesidades de diferentes personas y áreas de la empresa. Además, el concepto de granularidad provee flexibilidad al contenido del AD y a la manera de ver los datos de este. Como un edificio que, con una sólida base, logra mantenerse de pie durante un terremoto, un AD con una correcta granularidad como cimiento, es capaz de enfrentarse a los cambios que se puedan presentar en el futuro y al uso diario.

1.2.14. Estructura final de un AD

Los datos provenientes de las diferentes fuentes de información, tanto internas de la organización, como externas a esta, son captados y pasan por el proceso de ETL. Este proceso se realiza gracias al gestor de cargas, que se encarga, posterior a la extracción, de la limpieza y filtrado de los datos. La información limpia pasa a ser cargada en el AD, incluyéndose además los metadatos que son los que contienen la información sobre los datos del almacén. Finalmente los interesados pueden realizar consultas, análisis, estudios, reportes, etc, con la información contenida dentro del AD.

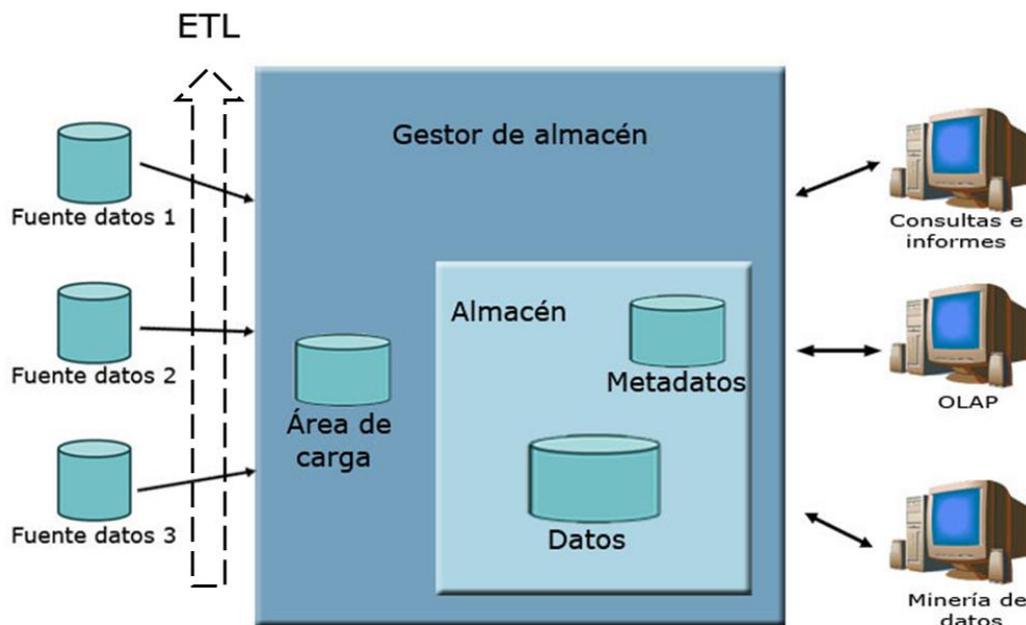


Fig. 2 Estructura general de un AD.

1.2.15. Modelo 0024-03 para los Indicadores del Plan Turquino

La información recopilada por la ONE es clasificada y recogida en los llamados “modelos”. El Modelo 0024-03 es el encargado de contener las estadísticas de los indicadores del Plan Turquino, tanto para los que caracterizan el desarrollo productivo como social. Dicho modelo contiene información de los 53 municipios de las 10 provincias que pertenecen a este plan (incluyéndose la Ciénaga de Zapata por razones prácticas). En el Anexo 2 se puede ver el facsímil del modelo.

1.3. Tecnologías y tendencias actuales

1.3.1. Sistemas existentes a nivel mundial

La información estadística, debido a su gran volumen y complejidad, así como a la importancia que tiene para los estudios de tendencias, toma de decisiones o análisis de datos, se convierte en una aplicación perfecta para la tecnología de AD. En estos momentos a nivel mundial existen varios sistemas de AD y MD estadísticos. A continuación se exponen ejemplos de ello:

- El 11 de febrero del presente año se puso en línea el Almacén Central de Datos de la ONE para el Sistema Estadístico Nacional de República Dominicana. Este AD servirá de apoyo a la recolección, procesamiento, análisis y difusión de la información estadística del país. Además ofrecerá servicios de información basándose en modelos analíticos y de minería de datos. Actualmente el almacén recoge la información del Censo Nacional de Población y Vivienda del 2002, del Directorio de Establecimientos Económicos 2009, de los nacimientos, defunciones, matrimonios y divorcios registrados entre los años 2001-2008, de las encuestas de Ingresos y gastos de los Hogares 2007, entre otras.

- El Almacén de Datos para el Análisis y Difusión de la Información Estadística del Turismo en España o DATATUR es un AD estadístico que se encarga de la recolección de todos los datos concernientes al área del turismo español. Sus principales objetivos son conservar y mantener al día los datos estadísticos de esta área, facilitar la actualización de manera ágil y sin errores, cubrir las demandas externas de información, entre otros aspectos.
- El Instituto Nacional de Estadísticas de Chile posee una AD implementado con varios MD que contienen la información acerca de los empleos, la industria, minería, las ventas, entre otros indicadores.
- El INEGI (Instituto Nacional de Estadísticas e Informática) de México tiene su información almacenada en un AD cuyo propósito es integrar dicha información en un repositorio para su consulta y análisis, para permitir a sus usuarios tomar decisiones en el contexto del Sistema Nacional de Estadísticas e Información Geográfica. (6)

1.3.2. Fundamentación de la metodología a utilizar

Como se ha visto ya a lo largo del desarrollo de la informática, las nuevas formas de crear software a menudo vienen acompañadas de sus respectivas metodologías, que no son más que las guías para el proceso de creación de dicho software.

Posterior a la creación del concepto de AD, surgieron dos tendencias de desarrollo, defendidas por los dos creadores de este concepto: William H. Inmon y Ralph Kimball.

Según Bill Inmon, el AD es una parte del todo que conforma un sistema de inteligencia. Una entidad tiene un AD y los MD tienen como fuente de información ese almacén. Esta tendencia es conocida como “Top-Down”.

Para Ralph Kimball el AD se compone por el conjunto de MD que existan en la entidad donde esté implementado y la información siempre se almacena en un modelo dimensional. Esta tendencia es conocida como “Bottom-Up”.

No se puede decir que una de estas dos tendencias es mejor que la otra, puesto que los beneficios de cada una dependen del ambiente de la empresa en que se vayan a implantar. En ocasiones la mejor estrategia es la combinación de estas dos tendencias, todo depende de lo que se quiera implementar y del negocio que se tenga. Tanto Inmon como Kimball coinciden en algo: los almacenes y mercados de datos independientes no satisfacen las necesidades de precisión y oportunidad de la información, como tampoco facilitan el acceso por parte del usuario ni el proceso de toma de decisiones.

Tras analizar las ventajas y desventajas de cada metodología existente para el trabajo con AD, el Grupo de Desarrollo de Soluciones para la línea de Almacenes de Datos del proyecto DATEC realizó una adaptación de la Metodología de Kimball para regir su proceso de desarrollo. Esto se debe a las siguientes características que posee esta metodología:

- Utiliza el concepto de Dimensiones y Hechos, lo cual facilita el proceso de desarrollo y lo hace más eficaz.
- Se basa en el paradigma de construcción de AD “Bottom-Up”, que define la creación de los MD departamentales y su posterior integración en un AD de la

organización. Esta estrategia coincide con la división lógica que se tiene en una organización.

- Es una metodología madura, bien definida y reconocida por el resto de la comunidad que se dedica al tema.
- Existe abundante documentación acerca de ella, además de grupos que se dedican a brindar ayuda del tema en la red.

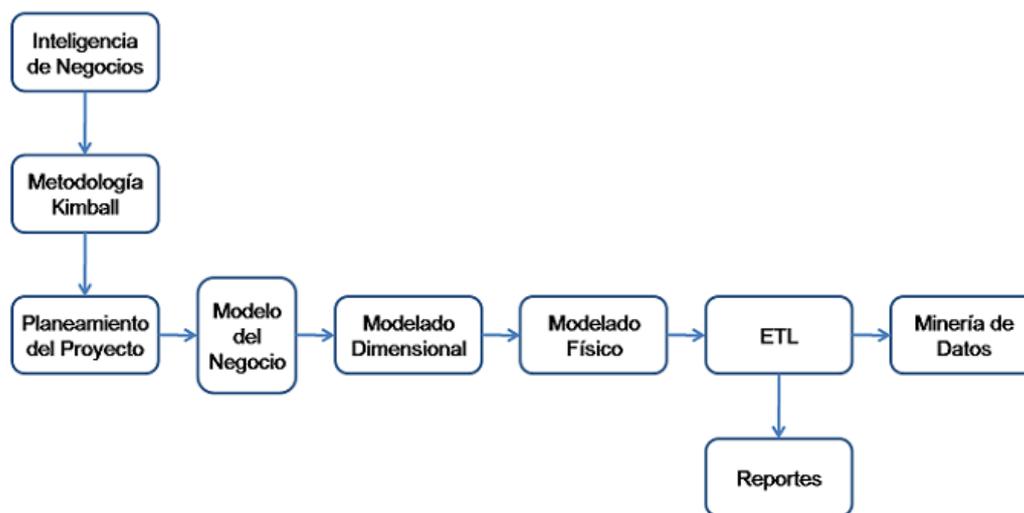


Fig. 2 Metodología de Kimball

Para complementar esta metodología se le agrega lo planteado por Leopoldo Zenaido Zepeda Sánchez en su Tesis de Doctorado “Metodología para el Diseño Conceptual de Almacenes de Datos”, al utilizar Casos de Uso para reforzar la etapa de levantamiento de requisitos, con lo que se logra además estar más a la par de las tendencias utilizadas en la Universidad. (7)

A partir de lo planteado en las metodologías-base descritas anteriormente y teniendo en cuenta las características propias de la UCI y DATEC se crea la metodología que rige el flujo de esta investigación. Durante su ciclo de vida se tienen los siguientes flujos de trabajo:

- **Estudio Preliminar o Planeación:** Se realiza el estudio de la entidad cliente, la planeación del proyecto, se definen los objetivos, el alcance preliminar, los costos estimados y otras actividades.
- **Requerimientos:** Se realiza en dos direcciones, una, mediante la identificación de las necesidades de información y reglas del negocio; y la otra con un levantamiento detallado de las fuentes de datos a integrar. Después se procede a la definición de los requerimientos.
- **Arquitectura y Diseño:** Se definen las estructuras de almacenamiento, se diseñan las reglas de extracción, transformación y carga y se define la arquitectura de información que regirá el desarrollo de la solución.
- **Implementación:** Se diseña físicamente el Repositorio de Datos, se crean las estructuras de almacenamiento, el Área Temporal de Almacenamiento, se ejecutan las reglas de ETL y se configuran e implementan las herramientas de BI para la

obtención de los elementos (reportes, gráficos, etc.) que se acordaron con el cliente final.

- **Prueba:** Se realizan las pruebas al sistema desde las Pruebas de Unidad hasta las de Aceptación con el cliente final.
- **Despliegue:** Se realiza un Despliegue Piloto en el cual se configuran los servidores y se instalan las herramientas y se carga una muestra de los datos para demostrar que el sistema funciona. Posterior a la aceptación del cliente se realiza la carga de los datos y la Capacitación y Transferencia Tecnológica.
- **Soporte y Mantenimiento:** Tras la implantación de la solución se brindan los servicios de soporte en línea, vía telefónica, web u otras según el contrato firmado y las condiciones de soporte establecidas.
- **Gestión y Administración del Proyecto:** A lo largo del ciclo de vida se realizan actividades de control, gestión y chequeo del desarrollo, los gastos, las utilidades, los recursos y demás actividades por parte del Grupo de Dirección del Proyecto.

Los Grupos de Trabajo que intervienen en el desarrollo del proyecto son (7):

- Grupo de Análisis.
- Grupo de Almacenes de Datos.
- Grupo de Extracción, Transformación y Carga.
- Grupo de Inteligencia de Negocio.
- Grupo de Dirección.

Grupos/ Flujos	Estudio Preliminar	Requerimientos	Arquitectura y Diseño	Implementación	Prueba	Despliegue	Soporte y Mantenimiento
Análisis							
Almacén							
ETL							
BI							
Dirección							

Leyenda:

	Responsable
	Participa
	No participa

Fig. 3 Relación de los Grupos de Trabajo con los diferentes Flujos del ciclo de vida.

1.3.3. Tecnologías y herramientas propuestas para el desarrollo del sistema

1.3.3.a Sistema Gestor de Bases de Datos

Oracle

Ventajas

- Oracle es el motor de base de datos más usado a nivel mundial.
- Posee un lenguaje de diseño de Bases de Datos (BD) muy completo (PL/SQL) que le permite implementar diseños complejos, con triggers y procedimientos almacenados, manteniendo una integridad bastante potente.
- Oracle está más orientado hacia Internet y posee amplia bibliografía en este medio.
- Para el trabajo con BD multidimensionales posee grandes facilidades a la hora de crear las dimensiones y hechos, brindando además estereotipos propios para estos elementos.
- Permite la creación de las jerarquías y niveles de manera sencilla.
- Posee herramientas dirigidas explícitamente al desarrollo de BD multidimensionales.
- Permite trabajar directamente con el código o mediante el uso de los *wizards* o ayudantes gráficos.
- Las últimas versiones de Oracle permiten dividir las BD de gran tamaño en elementos menores. Esta posibilidad se denomina Particionamiento y es considerada como una de las herramientas de más utilidad del Oracle.

Desventajas

- El mayor inconveniente de Oracle es su precio, ya que incluso las licencias de Personal Oracle son excesivamente caras.
- Consume muchos recursos de la computadora como RAM y espacio en el disco duro.

MySQL

Ventajas

- Es un SGBD muy usado debido a que se encuentra bajo la licencia GPL, es decir, es un software libre.
- Posee gran velocidad al realizar las operaciones y no necesita muchos recursos de la computadora en la que se ejecuta.
- No posee límites en los tamaños de los registros y es fácil de instalar y configurar.
- Tiene una baja probabilidad de corrupción en los datos.

- Posee herramientas para el trabajo con BD multidimensionales que permiten crear copias funcionales de la BD, chequear la integridad de esta y otras posibilidades.

Desventajas

- No es viable a la hora de utilizarlo con grandes cantidades de información a la que se acceda constantemente pues no posee una buena escalabilidad. Esto lo hace desventajoso a la hora del trabajo con BD multidimensionales.

SGBD a utilizar en el presente trabajo: PostgreSQL

Ventajas

El Sistema Gestor de Bases de Datos a utilizar en el presente trabajo será PostgreSQL 8.4.1, que desde su creación ha alcanzado gran popularidad en todas las esferas en las que se ha aplicado. Esto se debe principalmente a las siguientes características:

- Posee una instalación ilimitada y se encuentra bajo la licencia GPL, es decir, la licencia del software no tiene un costo asociado. Así nadie puede demandar al que use PostgreSQL por incumplimiento de licencia, trayendo las ventajas adicionales siguientes:
 - Modelos de negocios más rentables con instalaciones a gran escala.
 - La organización no será auditada en ningún momento para verificar cumplimiento de licencia.
 - Posibilidad de realizar investigaciones y desarrollo sin la necesidad de incurrir en costo de licenciamiento.
- Este gestor de bases de datos es altamente confiable y estable a la vez que ahorra considerablemente en costos de operación, manteniendo las características y rendimiento de otros productos similares.
- Es considerado actualmente el mejor SGBD libre del mundo.
- Permite el trabajo con la BD utilizando tanto código como los *wizards* o ayudantes gráficos.

La versión 8.4 de este gestor trae como ventajas (8):

- Una mayor velocidad en la restauración de bases de datos.
- Privilegios por columnas que permiten un control más granular de los datos confidenciales.
- Nuevas herramientas de monitoreo de consultas que permiten un mayor conocimiento del sistema por parte de los administradores.
- Facilidad en el análisis de los datos y consultas más complejas mediante funcionalidades avanzadas del estándar ANSI SQL 2003.

- Mejor rendimiento de la aplicación durante las consultas complejas.
- Permite dividir las tablas de hechos y dimensiones, tanto horizontal como verticalmente, en tablas más pequeñas. Esto se denomina Particionamiento de la BD y constituye una de las grandes ventajas dentro de los SGBD.
- Muchas de las herramientas usadas durante el proceso de ETL son compatibles con PostgreSQL.

1.3.3.b Herramientas CASE

EasyCASE

Ventajas

- Permite automatizar las fases de análisis y diseño dentro del desarrollo de una aplicación.
- Es un producto para la generación de esquemas de bases de datos e ingeniería inversa y trabaja para proveer una solución comprensible para el diseño.
- Es fácil de utilizar gracias a que posee una GUI (Interfaz Gráfica de Usuario) y produce prototipos rápidos y funcionales de la solución.
- Es compatible con muchos SGBD entre los que se encuentra el PostgreSQL.
- Posee niveles de seguridad para el trabajo con los registros, archivos y para el proyecto como tal.

Desventajas

- Esta herramienta no permite especificar los borrados y modificados antes de crear el código SQL.

ER/Studio 8.0

Ventajas

- ER/Studio es una herramienta visual de diseño y construcción de bases de datos tanto a nivel físico como lógico.
- Es fácil de usar además de poseer una extensa bibliografía en la red, además de la suya propia.

Posee las siguientes características:

- Una fuerte capacidad en el diseño lógico.
- Sincronización bidireccional de los diseños lógico y físico.
- Construcción automática de Bases de Datos.
- Reingeniería inversa de BD.

- Capacidad de importación y exportación de metadatos.
- Un repositorio para el modelado.
- Estereotipos propios para el trabajo con las tablas de hechos y dimensiones.
- Compatible con muchos SGBD entre los que se encuentra el PostgreSQL.

Desventajas

- Es un software propietario de Embarcadero Technologies Inc. por lo que se hace necesario poseer una licencia para poder utilizarlo.

Rational Rose

Ventajas

- Esta herramienta permite:
 - Crear, ver y modificar los componentes de un modelo.
 - Manipular los componentes de un modelo.
 - Habilita asistentes para crear clases y provee plantillas de código.
 - Admite la integración con otras herramientas de desarrollo.
- Posee sus propios estereotipos, aunque no diferencia las BD multidimensionales de las relacionales.
- Es altamente reconocido a nivel mundial.
- Soporta el desarrollo del ciclo de vida completo del software.

Desventajas

- Esta herramienta posee licencia privada de Rational Software Corporation.
- Necesita mucha memoria para poder ejecutarse de manera eficiente.

Herramienta de modelado a utilizar: Visual Paradigm

Ventajas

- Visual Paradigm es una herramienta profesional UML que soporta el ciclo de vida completo el desarrollo del software.
- Esta herramienta es fácil de usar, instalar y actualizar; además posee características como:

- Modelado colaborativo con CVS y Subversion.
- Ingeniería inversa: código a modelo, código a diagrama.
- Generación de código: modelo a código, diagrama a código.
- Editor de detalles de Casos de Uso.
- Generación de BD.
- Reorganización de las figuras y conectores de los diagramas UML.
- Editor de figuras.

Desventajas

- Una desventaja de esta herramienta es que posee licencia privada de Visual Paradigm International Limited, pero la UCI posee dicha licencia.

1.4. Conclusiones parciales

En este capítulo se llevó a cabo una profundización de los elementos fundamentales de la fundamentación teórica, además de exponer las herramientas y metodologías que harán posible la realización del trabajo. Luego del análisis pertinente se escogió la metodología definida por el proyecto DATEC por ser la más apropiada para el proceso que se llevará a cabo. Se definieron además como sistema gestor de bases de datos PostgreSQL 8.4.1, principalmente por ser un sistema robusto y poseer licencia libre, y como herramienta de modelado el Visual Paradigm 6.4.

Capítulo 2: Análisis y diseño de la solución

2.1. Introducción

El siguiente capítulo contiene la descripción de los pasos a seguir durante el análisis y el diseño de la solución. Aborda la especificación de los requisitos que debe cumplir el sistema, así como la realización de un modelo del MD a partir de los indicadores que se seleccionaron. Se realiza el diseño de lo que constituirá el sistema.

2.2. Análisis

2.2.1. Definición del negocio

La ONE, por ser la institución rectora de las estadísticas en Cuba, tiene como principal objetivo el de actuar como un almacén donde se gestione y guarde toda la información estadística del país. Con este fin la ONE cuenta con diferentes modelos que recogen la información según su tipo. Estos modelos se diferencian entre sí por la frecuencia con que se recogen (anual, trimestral, etc.), por los datos que se miden y por la estructura de la plantilla, entre otros aspectos. En el Modelo 0024-03 se recogen las estadísticas referentes al Plan Turquino y el proceso de recogida y gestión de sus datos es el siguiente:

- El Modelo 0024-03 o de Indicadores del Plan Turquino se recoge cada tres meses por la Oficina Municipal de Estadísticas de las provincias pertenecientes al Plan. Aquí se conforma el modelo con los datos enviados por los organismos y se envía a la Oficina Provincial de Estadísticas.
- En la Oficina Provincial de Estadísticas se hace un resumen de los datos por municipio y se almacena en un fichero de extensión dbf y luego se colocan en un servidor, que tiene una carpeta definida para cada provincia.
- Este fichero llega a la ONE y se comienza con la limpieza de los datos por parte de los especialistas y se crea un archivo general con la información.
- Finalmente la información se encuentra disponible mediante las publicaciones con que cuenta la ONE. En el caso del Modelo 0024-03 se tienen dos publicaciones: trimestral y anual (resumen del año de los datos recogidos) y posee dos variantes: estatal y no estatal.

2.2.2. Tema de análisis

Un tema de análisis no es más que la división o categorización de la información de una organización según las diferentes temáticas que contenga. Se elige un tema de análisis basándose en los objetivos que persigan las áreas de la organización. En el caso de esta investigación que se centra en el Modelo 0024-03 solamente, es decir, que no contiene otros tipos de información que no tengan relación con los indicadores del Plan Turquino, se tiene un tema de análisis denominado “Comportamiento de los Indicadores del Plan Turquino”.

Dentro de este tema se recogen los datos siguientes:

- Agricultura no cañera
- Ganado
- Forestal
- Turismo
- Industria alimenticia
- Industrias locales
- Industria básica
- Programa de empleo
- Viviendas
- Cultura
- Deporte
- Salud
- Comercio interior
- Energía
- Comunidades a solucionar abasto de agua
- Construcciones
- Transportación de pasajeros
- Viales
- Comunicaciones
- Investigaciones
- Inversiones del medio ambiente
- Perfeccionamiento empresarial

2.2.3. Roles y permisos

Para el acceso a la información contenida dentro de la base de datos que se implementará se definieron dos roles principales:

- Analista: es el encargado de realizar los análisis de la información, los reportes y consultas a los datos.
- Administrador de ETL: es la persona que administra el almacén y se encarga de todas las actividades referentes al proceso de ETL.

2.2.4. Reglas del negocio

A continuación se definen las reglas del negocio, o sea, las transformaciones que se le deben realizar a ciertos datos para obtener otros datos o los segmentos de un dato que pueden generar otros datos sin necesidad de ser introducidos por una persona.

- Crecimiento: Representa el porcentaje del crecimiento del año actual con respecto al anterior. El resultado se debe devolver con una cifra decimal; en caso de devolver un entero se deja un cero después de la coma. Ej. 75,6% ó 55,0%.

$$\text{Crecimiento} = (\text{real_año_actual} / \text{real_año_anterior}) * 100$$

- Cumplimiento: Representa el porcentaje del cumplimiento del trimestre del año actual. El resultado se debe devolver con una cifra decimal; en caso de devolver un entero se deja un cero después de la coma. Ej. 75,6% ó 55,0%.

$$\text{Cumplimiento} = (\text{real_año_actual} / \text{plan_año_actual}) * 100$$

- Variación contra plan: Representa la diferencia entre el real existente y el plan para el período que se está analizando. El resultado se debe devolver con una cifra decimal; en caso de devolver un entero se deja un cero después de la coma. Ej. 75,6 ó 97,0.

$$\text{Variación contra plan} = \text{real_año_actual} - \text{plan_año_actual}$$

- Variación contra año actual: Representa la diferencia entre el real existente y el real del año anterior para el período que se está analizando. El resultado se debe devolver con una cifra decimal; en caso de devolver un entero se deja un cero después de la coma. Ej. 75,6 ó 97,0.

$$\text{Variación contra año actual} = \text{real_año_actual} - \text{real_año_anterior}$$

Además de los atributos calculables anteriormente descritos, como parte del proceso de recogida y clasificación de datos que se lleva a cabo en la ONE, y para facilitar este proceso, se utilizan códigos para así evitar la pérdida de tiempo y los errores que podrían ser causados por el trabajo con los nombres de los parámetros. Esto se aplica tanto a los indicadores que se recogen en el Modelo 0024, como a los nombres de las provincias y municipios que componen la División Político-Administrativa (DPA).

En el caso de los datos de la DPA, se utiliza un código de 4 dígitos, los 2 primeros para identificar la provincia y los 2 últimos para los municipios. De esta manera queda un código que identifica a la vez la provincia y el municipio que pertenece a dicha provincia. (9) Estos códigos son los siguientes:

PINAR DEL RÍO		
COD_PROV	COD_MUN	DESCRIPCIÓN
01	0101	SANDINO
01	0102	MANTUA
01	0103	MINAS DE MATAHAMBRE
01	0104	VIÑALES
01	0105	LA PALMA
01	0106	BAHIA HONDA
01	0107	CANDELARIA
01	0108	SAN CRISTOBAL
01	0109	LOS PALACIOS
01	0110	CONSOLACION DEL SUR
01	0111	PINAR DEL RIO
01	0112	SAN LUIS
01	0113	SAN JUAN Y MARTINEZ
01	0114	GUANES

LA HABANA		
COD_PROV	COD_MUN	DESCRIPCIÓN
02	0201	MARIEL
02	0202	GUANAJAY
02	0203	CAIMITO
02	0204	BAUTA
02	0205	SAN ANTONIO DE LOS BAÑOS
02	0206	BEJUCAL
02	0207	SAN JOSE DE LAS LAJAS
02	0208	JARUCO
02	0209	SANTA CRUZ DEL NORTE
02	0210	MADRUGA
02	0211	NUEVA PAZ
02	0212	SAN NICOLAS
02	0213	GUINES
02	0214	MELENA DEL SUR
02	0215	BATABANO
02	0216	QUIVICAN
02	0217	GUIRA DE MELENA
02	0218	ALQUIZAR
02	0219	ARTEMISA

CIUDAD DE LA HABANA		
COD_PROV	COD_MUN	DESCRIPCIÓN
03	0301	PLAYA
03	0302	PLAZA DE LA REVOLUCION
03	0303	CENTRO HABANA
03	0304	LA HABANA VIEJA
03	0305	REGLA
03	0306	LA HABANA DEL ESTE
03	0307	GUANABACOA
03	0308	SAN MIGUEL DEL PADRON
03	0309	DIEZ DE OCTUBRE
03	0310	CERRO
03	0311	MARIANAO
03	0312	LA LISA
03	0313	BOYEROS
03	0314	ARROYO NARANJO
03	0315	COTORRO

MATANZAS		
COD_PROV	COD_MUN	DESCRIPCIÓN
04	0401	MATANZAS
04	0402	CARDENAS
04	0403	VARADERO
04	0404	MARTI
04	0405	COLON
04	0406	PERICO
04	0407	JOVELLANOS
04	0408	PEDRO BETANCOURT
04	0409	LIMONAR
04	0410	UNION DE REYES
04	0411	CIENAGA DE ZAPATA
04	0412	JAGUEY GRANDE
04	0413	CALIMETE
04	0414	LOS ARABOS

VILLA CLARA		
COD_PROV	COD_MUN	DESCRIPCIÓN
05	0501	CORRALILLO
05	0502	QUEMADO DE GUINES
05	0503	SAGUA LA GRANDE
05	0504	ENCRUCIJADA
05	0505	CAMAJUANI
05	0506	CAIBARIEN
05	0507	REMEDIOS
05	0508	PLACETAS
05	0509	SANTA CLARA
05	0510	CIFUENTES
05	0511	SANTO DOMINGO
05	0512	RANCHUELO
05	0513	MANICARAGUA

CIENFUEGOS		
COD_PROV	COD_MUN	DESCRIPCIÓN
06	0601	AGUADA DE PASAJEROS
06	0602	RODAS
06	0603	PALMIRA
06	0604	LAJAS
06	0605	CRUCES
06	0606	CUMANAYAGUA
06	0607	CIENFUEGOS
06	0608	ABREUS

SANCTI SPIRITUS		
COD_PROV	COD_MUN	DESCRIPCIÓN
07	0701	YAGUAJAY
07	0702	JATIBONICO
07	0703	TAGUASCO
07	0704	CABAIGUAN
07	0705	FOMENTO
07	0706	TRINIDAD
07	0707	SANCTI SPIRITUS
07	0708	LA SIERPE

CIEGO DE ÁVILA		
COD_PROV	COD_MUN	DESCRIPCIÓN
08	0801	CHAMBAS
08	0802	MORON
08	0803	BOLIVIA
08	0804	PRIMERO DE ENERO
08	0805	CIRO REDONDO
08	0806	FLORENCIA
08	0807	MAJAGUA
08	0808	CIEGO DE AVILA
08	0809	VENEZUELA
08	0810	BARAGUA

CAMAGUEY		
COD_PROV	COD_MUN	DESCRIPCIÓN
09	0901	CARLOS M. DE CESPEDES
09	0902	ESMERALDA
09	0903	SIERRA DE CUBITAS
09	0904	MINAS
09	0905	NUEVITAS
09	0906	GUAIMARO
09	0907	SIBANICU
09	0908	CAMAGUEY
09	0909	FLORIDA
09	0910	VERTIENTES
09	0911	JIMAGUAYU
09	0912	NAJASA
09	0913	SANTA CRUZ DEL SUR

LAS TUNAS		
COD_PROV	COD_MUN	DESCRIPCIÓN
10	1001	MANATI
10	1002	PUERTO PADRE
10	1003	JESUS MENENDEZ
10	1004	MAJIBACOA
10	1005	LAS TUNAS
10	1006	JOBABO
10	1007	COLOMBIA
10	1008	AMANCIO

HOLGUÍN		
COD_PROV	COD_MUN	DESCRIPCIÓN
11	1101	GIBARA
11	1102	RAFAEL FREIRE
11	1103	BANES
11	1104	ANTILLA
11	1105	BAGUANOS
11	1106	HOLGUIN
11	1107	CALIXTO GARCIA
11	1108	CACOCUM
11	1109	URBANO NORIS
11	1110	CUETO
11	1111	MAYARI
11	1112	FRANK PAIS
11	1113	SAGUA DE TANAMO
11	1114	MOA

GRANMA		
COD_PROV	COD_MUN	DESCRIPCIÓN
12	1201	RIO CAUTO
12	1202	CAUTO CRISTO
12	1203	JIGUANI
12	1204	BAYAMO
12	1205	YARA
12	1206	MANZANILLO
12	1207	CAMPECHUELA
12	1208	MEDIA LUNA
12	1209	NIQUERO
12	1210	PILON
12	1211	BARTOLOME MASO
12	1212	BUEY ARRIBA
12	1213	GUISA

SANTIAGO DE CUBA		
COD_PROV	COD_MUN	DESCRIPCIÓN
13	1301	CONTRAMAESTRE
13	1302	MELLA
13	1303	SAN LUIS
13	1304	SEGUNDO FRENTE
13	1305	SONGO - LA MAYA
13	1306	SANTIAGO DE CUBA
13	1307	PALMA SORIANO
13	1308	TERCER FRENTE
13	1309	GUAMA

GUANTÁNAMO		
COD_PROV	COD_MUN	DESCRIPCIÓN
14	1401	EL SALVADOR
14	1402	GUANTANAMO
14	1403	YATERAS
14	1404	BARACOA
14	1405	MAISI
14	1406	IMIAS
14	1407	SAN ANTONIO DEL SUR
14	1408	MANUEL TAMES
14	1409	CAIMANERA
14	1410	NICETO PEREZ

ISLA DE LA JUVENTUD		
COD_PROV	COD_MUN	DESCRIPCIÓN
99	9901	ISLA DE LA JUVENTUD

Fig. 4 Tablas de códigos de la ONE para las provincias y sus municipios.

Debido a la necesidad de introducir un código único para cada indicador de los contenidos dentro del Modelo 0024-03 se estudió el Clasificador Central de Productos de Cuba (CPCU) en su versión 1.1. Este clasificador está basado en el Clasificador Central de Productos (CPC), que sirve de marco para la comparación internacional de estadísticas en materia de bienes, servicios y activos.

Tomando como base el CPC y adaptándolo a las características y necesidades de Cuba, se crea el CPCU v1.1. A partir del último nivel jerárquico de la CPC se agregan cuatro dígitos para la codificación cubana, dos para las partidas y dos para las subpartidas, quedándose el código cubano con un total de nueve dígitos. (10)

Seguidamente se muestra un ejemplo del CPCU para que se comprenda mejor su estructura.

Sección	0	Productos de la agricultura, la silvicultura y la pesca
División	01	Productos de la agricultura, la horticultura y la jardinería comercial
Grupo	013	Frutas y Nueces
Clase	0132	Frutas cítricas, frescas o secas
Subclase	01320.0000	Frutas cítricas, frescas o secas
	01320.0100	Naranja Dulce
Partida	01320.0200	Toronja
	01320.0300	Limón
	01320.0900	Las demás frutas cítricas, frescas o secas
Subpartida	01320.0901	Limas
	01320.0902	Mandarinas

Fig. 5 Estructura del CPCU v1.1.

Los códigos de los indicadores, por tener una cantidad significativa, se pueden encontrar en el Anexo 1 de este documento.

2.2.5. Necesidades de los usuarios

Después de un análisis del negocio que toma lugar en la ONE se define como la necesidad de los usuarios la publicación de los Indicadores del Plan Turquino, conteniendo los parámetros que se recogen en el modelo: la existencia al final del período, el real del año anterior y el real y plan del año actual. Los diferentes indicadores se dividieron en 3 grupos según los datos que representan: indicadores de los productos de la agricultura y la ganadería, de los productos de la industria y de los servicios. Finalmente y según lo especificado por el cliente estos parámetros se medirán de acuerdo al tiempo, a la variante del sector (estatal o no estatal) y a la provincia. Partiendo de esto se describen los requisitos de información, multidimensionales, funcionales y no funcionales del sistema.

2.2.5.a. Requisitos de información

Luego de un análisis de los problemas que existen en la ONE con respecto a la centralización y disponibilidad de los datos concernientes al Modelo 0024-03, y a los conversatorios llevados a cabo con el cliente, se definieron los requisitos de información de la investigación. Todos los requisitos pertenecen al tema de análisis “Comportamiento de los Indicadores del Plan Turquino” y su información es pública. Hay un total de 30 requisitos de información que abarcan los datos que se recogen en el Modelo 0024-03, estos requisitos son:

RI 1. Obtener existencia al final del período de los productos de la agricultura y la ganadería por provincia.

RI 2. Obtener existencia al final del período de los productos de la agricultura y la ganadería por el tiempo.

RI 3. Obtener existencia al final del período de los productos de la agricultura y la ganadería por variante de sector.

RI 4. Obtener real del año anterior de los productos de la agricultura y la ganadería por provincia.

RI 5. Obtener real del año anterior de los productos de la agricultura y la ganadería por el tiempo.

RI 6. Obtener real del año anterior de los productos de la agricultura y la ganadería por variante de sector.

RI 7. Obtener plan del año actual de los productos de la agricultura y la ganadería por provincia.

RI 8. Obtener plan del año actual de los productos de la agricultura y la ganadería por el tiempo.

RI 9. Obtener plan del año actual de los productos de la agricultura y la ganadería por variante de sector.

RI 10. Obtener real del año actual de los productos de la agricultura y la ganadería por provincia.

RI 11. Obtener real del año actual de los productos de la agricultura y la ganadería por el tiempo.

RI 12. Obtener real del año actual de los productos de la agricultura y la ganadería por variante de sector.

RI 13. Obtener real del año anterior de los productos de la industria por provincia.

RI 14. Obtener real del año anterior de los productos de la industria por el tiempo.

RI 15. Obtener real del año anterior de los productos de la industria por variante de sector.

RI 16. Obtener plan del año actual de los productos de la industria por provincia.

- RI 17. Obtener plan del año actual de los productos de la industria por el tiempo.
- RI 18. Obtener plan del año actual de los productos de la industria por variante de sector.
- RI 19. Obtener real del año actual de los productos de la industria por provincia.
- RI 20. Obtener real del año actual de los productos de la industria por el tiempo.
- RI 21. Obtener real del año actual de los productos de la industria por variante de sector.
- RI 22. Obtener real del año anterior de los servicios por provincia.
- RI 23. Obtener real del año anterior de los servicios por el tiempo.
- RI 24. Obtener real del año anterior de los servicios por variante de sector.
- RI 25. Obtener plan del año actual de los servicios por provincia.
- RI 26. Obtener plan del año actual de los servicios por el tiempo.
- RI 27. Obtener plan del año actual de los servicios por variante de sector.
- RI 28. Obtener real del año actual de los servicios por provincia.
- RI 29. Obtener real del año actual de los servicios por el tiempo.
- RI 30. Obtener real del año actual de los servicios por variante de sector.

2.2.5.b. Requisitos multidimensionales

Los requisitos multidimensionales no son más que las variables de entrada y de salida de la solución. Estos se definen a partir de los requisitos de información anteriormente descritos.

VE:

- productos_agricultura_ganadería
- productos_industria
- servicios
- provincia
- tiempo
- variante_sector

VS:

- existencia_final_período
- real_año_anterior
- plan_año_actual

- real_año_actual

2.2.5.c. Requisitos funcionales

RF 1. Extraer, transformar y cargar los datos de las provincias.

RF 2. Extraer, transformar y cargar los datos de los indicadores.

2.2.5.d. Requisitos no funcionales

Usabilidad (11)

Requisitos de usabilidad

- El sistema debe contar con un diseño del modelo físico sencillo, con una estructura y distribución que permita trabajar con rapidez y eficiencia.

Fiabilidad (11)

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

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

Requisitos de fiabilidad

- **Disponibilidad**
 - El área de almacenamiento del proceso de ETL debe estar disponible 12 horas del día.
 - El mantenimiento se realizará 1 vez por semana, realizándose chequeos para establecer los errores que existan y, de ser posible, arreglarlos. Esto se debe realizar preferentemente a finales de la semana.
 - El sistema debe tener disponibilidad al 100% entre las 8:00 am y las 5:00 pm en los días laborables (lunes a viernes).
- **Tiempo medio de fallos**
 - El tiempo medio de fallos del proceso de integración es de 7 días.
- **Tiempo medio de reparación**
 - El tiempo medio de reparación estimado es aproximadamente de 1 día luego del fallo.
- **Máximo de errores**
 - Por ser el proceso de integración un punto crítico para la calidad y exactitud de los datos almacenados se establece un máximo de errores de 0 errores/puntos de función.
- **Exactitud**
 - La salida de los datos debe ser 100% exacta, ya que de ahí dependen las decisiones que se tomen basándose en estos datos.

Eficiencia (11)

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

Requisitos de eficiencia

- **Tiempo de respuesta**
 - El sistema debe tener un tiempo de respuesta aproximado de 5 segundos, no debe excederse de 1 minuto.
- **Capacidad**
 - Durante la integración solo se conectará un usuario que deberá monitorear el proceso de integración.
 - El sistema deberá permitir varios usuarios (mín. de 5) conectados sin que esto afecte la respuesta de las consultas.

Restricciones de diseño (11)

- El lenguaje para realizar las consultas a la base de datos será SQL.

Componentes Comprados (11)

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

Interfaz (11)

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

Interfaces de usuario

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

Interfaces Hardware

- Para el proceso de transformación es necesaria una memoria RAM de 512 MB como mínimo.
- Para el proceso de visualización e inteligencia de negocio se necesita una memoria RAM de 1 GB como mínimo, para garantizar el correcto funcionamiento del sistema cuando es accedido por varios usuarios simultáneamente.
- Se necesita un mínimo de 60 GB de capacidad del disco duro para el almacenamiento de la información.
- Se debe garantizar al menos una impresora, principalmente de tipo Laser jet 1100 que son las utilizadas en estos momentos, para imprimir los reportes de salida.

Interfaces Software

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

Interfaces de Comunicación

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

Requisitos de Licencia (11)

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

2.2.5.e. CU del sistema

De acuerdo a los requisitos funcionales que se definieron anteriormente, los CU del sistema son:

✓ **CU Extraer datos del xls DPA**

Caso de Uso:	Extraer datos del xls DPA	
Tipo:	Funcional	
Actores:	Administrador ETL	
Resumen:	El CU inicia cuando el actor accede a la fuente de información con los datos de DPA, selecciona los datos a extraer y realiza la extracción.	
Precondiciones:	Existencia de las fuentes	
Referencias		
Prioridad		
Complejidad		
Flujo Normal de Eventos		
Sección “Extraer datos del xls DPA”		
	Acción del Actor	Respuesta de la fuente.
	1. Accede al archivo xls que contiene la información de la DPA.	1.1 Abre el archivo.
	2. Selecciona información a extraer.	
	3. Realiza la extracción.	
	4. Cierra el archivo.	
Flujos Alternos		

Acción del Actor	Respuesta de la fuente
	1.1 No puede abrir el archivo.
2. Notifica el error al usuario.	
Poscondiciones	Datos disponibles para transformar.

✓ **CU Extraer datos del xls Productos_Indicadores**

Caso de Uso:	Extraer datos del xls Productos_Indicadores
Tipo:	Funcional
Actores:	Administrador ETL
Resumen:	El CU inicia cuando el actor accede a la fuente de información con los datos de los productos e indicadores, selecciona los datos a extraer y realiza la extracción.
Precondiciones:	Existencia de las fuentes
Referencias	
Prioridad	
Complejidad	

Flujo Normal de Eventos

Sección “Extraer datos del xls Productos_Indicadores”

Acción del Actor	Respuesta de la fuente.
1. Accede al archivo xls que contiene la información de los Productos e Indicadores.	1.1 Abre el archivo.
2. Selecciona información a extraer.	
3. Realiza la extracción.	
4. Cierra el archivo.	

Flujos Alternos

Acción del Actor	Respuesta de la fuente
	1.1 No puede abrir el archivo.
2. Notifica el error al usuario.	
Poscondiciones	Datos disponibles para transformar.

✓ **CU Transformar y cargar datos del xls DPA**

Caso de Uso:	Transformar y cargar datos del xls DPA
Tipo:	Funcional
Actores:	Administrador ETL

Resumen:	El CU inicia cuando el actor selecciona las estructuras a transformar, carga los datos y les realiza las transformaciones necesarias. Seguidamente vuelve a cargar los datos en el almacén.	
Precondiciones:	Extracción completada en área temporal. La estructura del almacén debe estar disponible.	
Referencias		
Prioridad		
Complejidad		
Flujo Normal de Eventos		
Sección “Transformar y cargar datos del xls DPA”		
	Acción del Actor	Respuesta del almacén.
	1. Selecciona estructuras del xls de DPA a transformar.	
	2. Carga datos seleccionados en memoria.	
	3. Aplica transformaciones pertinentes y genera datos de auditoría.	
	4. Carga datos en el almacén.	
Flujos Alternos		
	Acción del Actor	Respuesta del almacén
Poscondiciones	Datos del xls DPA transformados y cargados.	

✓ **CU Transformar y cargar datos del xls Productos_Indicadores**

Caso de Uso:	Transformar y cargar datos del xls Productos_Indicadores	
Tipo:	Funcional	
Actores:	Administrador ETL	
Resumen:	El CU inicia cuando el actor selecciona las estructuras a transformar, carga los datos y les realiza las transformaciones necesarias. Seguidamente vuelve a cargar los datos en el almacén.	
Precondiciones:	Extracción completada en área temporal. La estructura del almacén debe estar disponible.	
Referencias		
Prioridad		
Complejidad		
Flujo Normal de Eventos		
Sección “Transformar y cargar datos del xls Productos_Indicadores”		
	Acción del Actor	Respuesta del almacén.

1. Selecciona estructuras del xls de Productos_Indicadores a transformar.	
2. Carga datos seleccionados en memoria.	
3. Aplica transformaciones pertinentes y genera datos de auditoría.	
4. Carga datos en el almacén.	
Flujos Alternos	
Acción del Actor	Respuesta del almacén
Poscondiciones	Datos del xls Productos_Indicadores transformados y cargados.

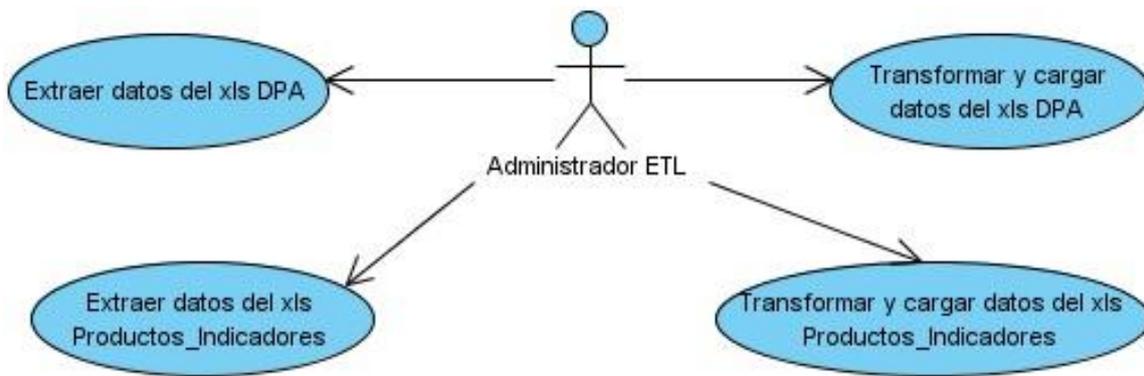


Fig. 6 Diagrama de CU del sistema

2.3. Diseño

2.3.1. Matriz BUS

La Matriz BUS representa las relaciones existentes entre los hechos y las dimensiones del sistema diseñado. En el caso de este modelo las dimensiones Temporal, Variante_sector y DPA están relacionadas con los 3 hechos existentes. En el caso de las dimensiones Indicadores_productos_agricultura_ganadería, Indicadores_productos_industria e Indicadores_servicios solamente se relacionan cada uno de los hechos de igual nombre respectivamente.

Dimensiones:

D1: Temporal

D2: Variante_sector

D3: DPA

D4: Indicadores_productos_agricultura_ganadería

D5: Indicadores_productos_industria

D6: Indicadores_servicios

Hechos/dimensiones	D1	D2	D3	D4	D5	D6
hec_indicadores_produc_ganad_agric	X	X	X	X		
hec_indicadores_produc_ind	X	X	X		X	
hec_indicadores_serv	X	X	X			X

Tabla 2 Matriz BUS

2.3.2. Modelo de datos

A continuación se explica el modelo de datos, compuesto por las dimensiones y hechos seleccionados y las medidas definidas según el Modelo 0024-03.

2.3.2.a. Dimensiones identificadas

Luego del análisis del Modelo 0024-03 y su contenido, así como de los reportes que se realizan de este, se definieron las dimensiones que intervendrán en el sistema a diseñar.

✓ dim_temporal

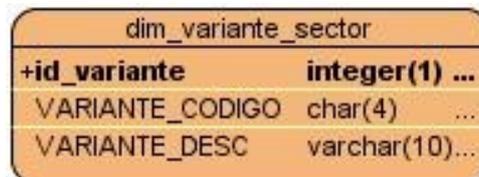
Esta dimensión describe el universo de valores bajo los cuales puede clasificarse la información atendiendo a los valores anuales y trimestrales que posee el modelo. Presenta un nivel denominado “año” y dentro la categoría “trimestre”.

Anno -> Trimestre



✓ dim_variante_sector

Esta dimensión describe el universo de valores bajo los cuales puede clasificarse la información atendiendo a la variante, ya sea estatal o no estatal, del modelo.



✓ dim_DPA

Esta dimensión describe el universo de valores bajo los cuales puede clasificarse la información atendiendo a la situación geográfica. Presenta un nivel “provincia” y dentro de este se encuentra “municipios”, representando la actual división del país en provincias y sus municipios.

Provincia -> Municipios

dim_DPA		
+id_dpa	integer(4)	Nullable = false
PROV_CODIGO	char(2)	Nullable = true
PROV_NOMBRE	varchar(25)	Nullable = true
MUN_CODIGO	char(4)	Nullable = true
MUN_NOMBRE	varchar(25)	Nullable = true

✓ **dim_indicadores_productos_agricultura_ganadería**

Esta dimensión describe el universo de valores bajo los cuales puede clasificarse la información atendiendo a los indicadores que se recogen para los productos de la agricultura, ganadería, acuicultura, apicultura, avicultura y forestal.

dim_indicadores_productos_agricultura_ganaderia		
+id_indicador_prod_agric_ganad	integer(4)	Nullable = false
INDICADOR_PROD_AGRIC_GANAD_DESCRIPCION	varchar(100)	Nullable = true
TEMATICA_PROD_AGRIC_GANAD_DESCRIPCION	varchar(100)	Nullable = true
SUBTEMATICA_PROD_AGRIC_GANAD_DESCRIPCION	varchar(100)	Nullable = true
INDICADOR_PROD_AGRIC_GANAD_CUI	char(11)	Nullable = true
INDICADOR_PROD_AGRIC_GANAD_UM	char(10)	Nullable = true

✓ **dim_indicadores_productos_industria**

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

dim_indicadores_productos_industria		
+id_indicador_prod_ind	integer(4)	...
INDICADOR_PI_DESCRIPCION	varchar(100)	...
TEMATICA_PI_DESCRIPCION	varchar(100)	...
SUBTEMATICA_PI_DESCRIPCION	varchar(100)	...
INDICADOR_PI_CUI	char(11)	...
INDICADOR_PI_UM	char(10)	...

✓ **dim_indicadores_servicios**

Esta dimensión describe el universo de valores bajo los cuales puede clasificarse la información atendiendo a los indicadores que se recogen para los servicios, por ejemplo, el programa de empleo, cultura, deporte, entre otros.

dim_indicadores_servicios		
+id_indicador_serv	integer(4)	...
INDICADOR_SERV_DESCRIPCION	varchar(100)	...
TEMATICA_SERV_DESCRIPCION	varchar(100)	...
SUBTEMATICA_SERV_DESCRIPCION	varchar(100)	...
INDICADOR_SERV_CUI	char(11)	...
INDICADOR_SERV_UM	char(10)	...

2.3.2.b. Hechos identificados

El modelo consta de 3 hechos debido a la necesidad de separar los indicadores en 3 categorías, y estos a su vez se relacionan con las dimensiones DPA, Temporal y Variante_sector.

✓ **hec_indicadores_produc_ganad_agric**

Se relaciona con la dimensión de igual nombre y se utiliza para el trabajo con los datos de todos los indicadores relacionados con los productos de la agricultura y la ganadería.

hec_indicadores_produc_ganad_agric	
#id_variante	integer(1) ...
#temporal_id	integer(4) ...
#id_dpa	integer(4) ...
#id_indicador_prod_agric_ganad	integer(4) ...
existencia_final_periodo	decimal(18, 0)...
real_anno_anterior	decimal(18, 0)...
real_anno_actual	decimal(18, 0)...
plan_anno_actual	decimal(18, 0)...

✓ **hec_indicadores_produc_ind**

Se relaciona con la dimensión de igual nombre y se utiliza para el trabajo con los datos de todos los indicadores relacionados con la industria básica, alimenticia y local.

hec_indicadores_produc_ind	
#id_indicador_prod_ind	integer(4) ...
#id_variante	integer(1) ...
#temporal_id	integer(4) ...
#id_dpa	integer(4) ...
real_anno_anterior	decimal(18, 0)...
plan_anno_actual	decimal(18, 0)...
real_anno_actual	decimal(18, 0)...

✓ **hec_indicadores_produc_serv**

Se relaciona con la dimensión de igual nombre y se utiliza para el trabajo con los datos de todos los indicadores relacionados con los servicios, es decir, la educación, cultura, deporte, turismo, entre otros.

hec_indicadores_serv		
#id_variante	integer(1)	...
#temporal_id	integer(4)	...
#id_dpa	integer(4)	...
#id_indicador_serv	integer(4)	...
real_anno_anterior	decimal(18, 0)	...
plan_anno_actual	decimal(18, 0)	...
real_anno_actual	decimal(18, 0)	...

2.3.2.c. Medidas

✓ Hecho indicadores_produc_ganad_agric

- existencia_final_período: representa la cantidad existente tanto de animales como de productos agrícolas al final del período al cual pertenecen los datos (un trimestre o año específico).
- real_anno_anterior: representa la cantidad real existente al final del mismo período, pero del año anterior al que se representa en el reporte.
- plan_anno_actual: representa la cantidad que se espera obtener (meta) en el período que se está evaluando del año actual.
- real_anno_actual: representa la cantidad real que se tiene en el período que se está evaluando del año actual.

✓ Hecho indicadores_produc_ind

- real_anno_anterior: representa la cantidad real existente al final del mismo período, pero del año anterior al que se representa en el reporte.
- plan_anno_actual: representa la cantidad que se espera obtener (meta) en el período que se está evaluando del año actual.
- real_anno_actual: representa la cantidad real que se tiene en el período que se está evaluando del año actual.

✓ Hecho indicadores_serv

- real_anno_anterior: representa la cantidad real existente al final del mismo período, pero del año anterior al que se representa en el reporte.
- plan_anno_actual: representa la cantidad que se espera obtener (meta) en el período que se está evaluando del año actual.
- real_anno_actual: representa la cantidad real que se tiene en el período que se está evaluando del año actual.

2.3.3. Diseño de las dimensiones y la tabla de hechos

El diseño final quedaría como se ilustra en la figura 7, con los 3 hechos relacionados con las dimensiones de igual nombre y, además, con las dimensiones Temporal, DPA y Variante_sector.

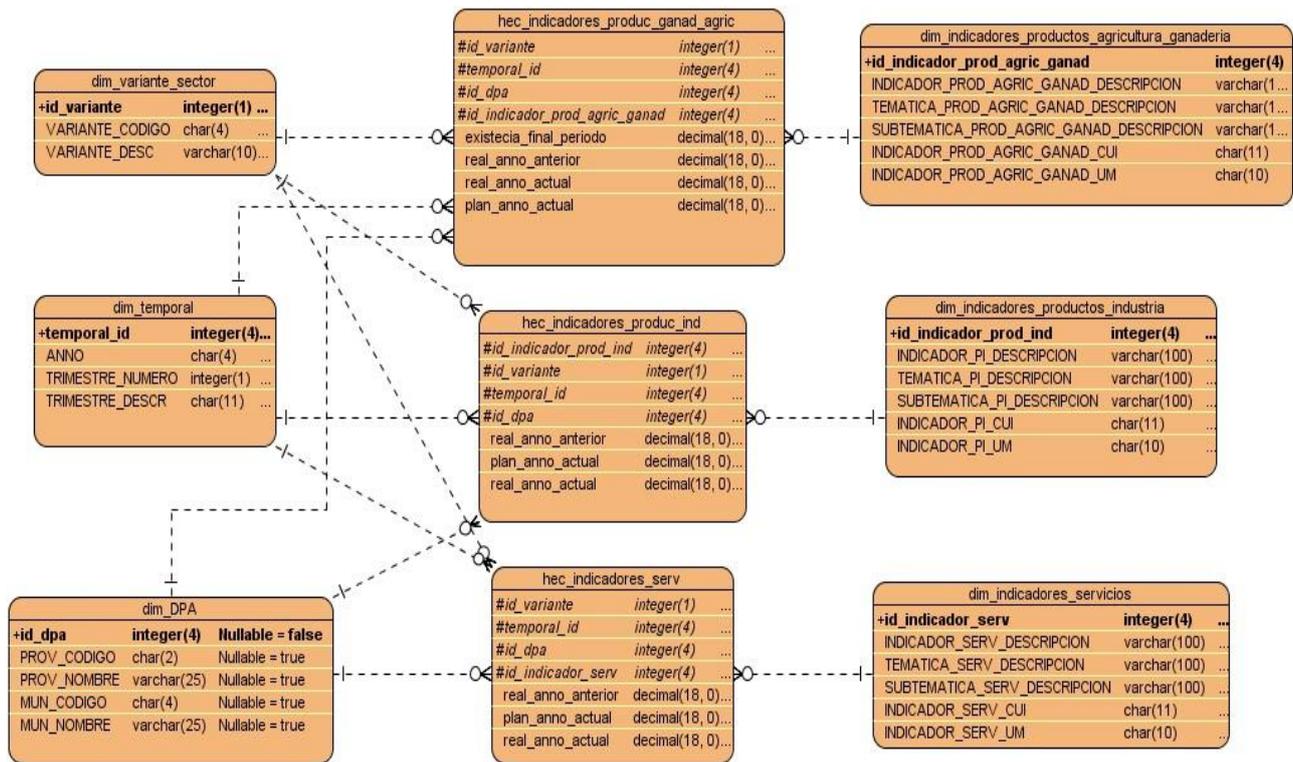


Fig. 7 Diseño de las dimensiones y tablas de hechos de la solución

2.3.4. Esquema de seguridad

El esquema de seguridad pretende administrar los permisos que tendría cada tipo de usuario a la hora de acceder a la información del sistema. Los datos recogidos en el Modelo 0024-03 son públicos, por lo que no se necesitan muchas restricciones a la hora de definir el acceso a dicha información.

ROL	BD	Esquemas	Tablas
Arq. De Datos	ONE_dwh (o), ONE_meta (o), ONE_middle (o)	Todos	Todos
ETL	ONE_middle (rwu)	Dimensiones (u), Indicadores_plan_Turquino (u)	Dim_* (rwu), Hech_* (rwu)
SCN	ONE_dwh (rwu)	Dimensiones (u), Mart_Indicadores_Plan_Turquino (u), Configuración (u)	Dim_* (r), Hech_* (r), Vm_* (rwu)
ANÁLISIS	ONE_dwh (r)	Dimensiones (u), Mart_Indicadores_Plan_Turquino (u)	Dim_* (r), Hech_* (r), Vm_* (r)

RÉPLICA	ONE_dwh ()	Réplica () Dimensiones (), Mart_Indicadores_Plan_Turquino ()	Dim_* (), Hech_* ()
----------------	------------	---	------------------------

Tabla 3 Esquema de seguridad propuesto

LEYENDA

- O -> Owner
- U -> Update
- R -> Read
- E -> Execute
- W -> Write

2.3.5. Política de respaldo y recuperación.

Ante la posibilidad de fallos o pérdida de los datos y para evitar además la existencia de información que no sea de utilidad para los especialistas de la ONE, se proponen las siguientes medidas:

- ✓ Realizar copias de respaldo (backups) de la información cada 3 meses, debido a que es este el período de tiempo establecido por la ONE para captar la información recogida en el modelo. Para esto se definirá una estructura de carpetas de la manera Modelo-> Año->Trimestre, para garantizar la rápida ubicación y fácil comprensión de la copia, en caso de necesitarse. (12)
- ✓ Para evitar que el almacén crezca indefinidamente se propone, de acuerdo con el cliente, que la información se almacene durante 10 años. Después de este período de tiempo, los datos deben almacenarse de forma parecida a una copia de respaldo, pero llegando solamente al nivel de Modelo-> Año. (12)
- ✓ En caso de ocurrir una eventualidad y afectarse la información contenida en el MD, se debe realizar un informe notificando lo ocurrido. Luego se procede a restaurar la información, utilizando la última copia existente, en un plazo máximo de 24 horas para no afectar el trabajo de los especialistas o de los usuarios que necesitan de esta información.

2.4. Conclusiones parciales

En este capítulo se describieron los aspectos principales dentro del proceso de análisis y diseño de la solución. Se definieron además las políticas de respaldo y recuperación y el esquema de seguridad. De todo este proceso se concluye lo siguiente:

- Las dimensiones y tablas de hechos definidos cumplen con lo establecido por el cliente y satisfacen las necesidades de este.

- El trabajo con los códigos utilizados en la ONE garantiza la concordancia con las políticas de esta organización.
- La arquitectura propuesta satisface el problema que se había planteado por parte de la ONE.
- La utilización de las políticas de respaldo garantizan la integridad y disponibilidad de los datos del mercado.

Capítulo 3: Implementación y pruebas

3.1. Introducción

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

3.2. Modelo de Datos Físico

3.2.1. Estructuras de Datos

- Esquemas

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

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

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

- esq_dimension: contiene todas las tablas de dimensiones propuestas.
- esq_hechos: contiene todas las tablas de hechos propuestas.

- Tablas

La solución cuenta con 9 tablas en total, 6 dimensiones y 3 hechos, distribuidas por los dos esquemas propuestos con anterioridad quedando de la siguiente manera:

Tabla	Esquema	Propietario	Descripción
dim_dpa	dimension	ADMINISTRADOR_ETL	Contiene los datos de la División Político-Administrativa del país.
dim_indicadores_productos_agricultura_ganad	dimension	ADMINISTRADOR_ETL	Contiene los datos de los indicadores

eria			concernientes a los productos de la agricultura y la ganadería.
dim_indicadores_productos_industria	dimensi on	ADMINISTRADOR_ ETL	Contiene los datos de los indicadores concernientes a los productos de la industria básica, local y alimenticia.
dim_indicadores_servicios	dimensi on	ADMINISTRADOR_ ETL	Contiene los datos de los indicadores concernientes a los servicios (cultura, educación, vivienda, etc.).
dim_temporal	dimensi on	ADMINISTRADOR_ ETL	Contiene los rangos de tiempo en que se recogen los datos.
dim_variante_sector	dimensi on	ADMINISTRADOR_ ETL	Contiene la variante de sector: estatal o no estatal.
hec_indicadores_producto_ganad_agric	hechos	ADMINISTRADOR_ ETL	Contiene las medidas concernientes a los productos de la agricultura y la ganadería.
hec_indicadores_producto_ind	hechos	ADMINISTRADOR_ ETL	Contiene las medidas concernientes a los productos de la industria.
hec_indicadores_serv	hechos	ADMINISTRADOR_ ETL	Contiene las medidas concernientes a los servicios.

Tabla 4 Tablas creadas para la solución

- Restricciones

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

- De unicidad: Implica que no deben haber 2 valores iguales en la misma columna.

- Valor no nulo: No debe existir ninguna casilla de la columna que esté vacía.
- Clave primaria: Son valores que deben cumplir con un conjunto de restricciones: no tener valores nulos, ser únicos para cada tupla y ser necesarios (minimal).
- Clave foránea: Son las que referencian una clave de otra tabla, se usan para relacionar tablas diferentes. (13)

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

- Tabla DPA:
 - Clave primaria: id_dpa.
 - De unicidad: munic_codigo.
 - Valor no nulo: todos los campos.
- Tabla Temporal:
 - Clave primaria: temporal_id.
 - Valor no nulo: todos los campos.
- Tabla Variante_sector:
 - Clave primaria: id_variante.
 - De unicidad: variante_codigo.
 - Valor no nulo: todos los campos.
- Tabla Indicadores_productos_agricultura_ganadería:
 - Clave primaria: id_indicador_prod_agric_ganad.
 - De unicidad: indicador_cui.
 - Valor no nulo: todos los campos.
- Tabla Indicadores_productos_industria:
 - Clave primaria: id_indicador_prod_ind.
 - De unicidad: indicador_cui.
 - Valor no nulo: todos los campos.
- Tabla Indicadores_servicios:
 - Clave primaria: id_indicador_serv.

- De unicidad: indicador_cui.
 - Valor no nulo: en este caso no se aplicó la restricción de valor no nulo a todos los campos porque esta tabla aún no está completa y el definir esta restricción puede ocasionar molestias a la hora de terminar de introducir los datos. Sin embargo en el futuro, cuando se termine la carga de los nomencladores, se deben definir todos los campos de la tabla como no nulos.
- Tabla hec_indicadores_prod_agric_gan:
- Clave primaria: Conformada por el conjunto de llaves foráneas que contiene: id_dpa, temporal_id, id_variante, id_indicador_prod_agric_ganad.
 - Claves foráneas:
 - id_dpa
 - temporal_id
 - id_variante
 - id_indicador_prod_agric_ganad
- Tabla hec_indicadores_prod_ind:
- Clave primaria: Conformada por el conjunto de llaves foráneas que contiene: id_dpa, temporal_id, id_variante, id_indicador_prod_ind.
 - Claves foráneas:
 - id_dpa
 - temporal_id
 - id_variante
 - id_indicador_prod_ind
- Tabla hec_indicadores_serv:
- Clave primaria: Conformada por el conjunto de llaves foráneas que contiene: id_dpa, temporal_id, id_variante, id_indicador_serv.
 - Claves foráneas:
 - id_dpa
 - temporal_id
 - id_variante

- id_indicador_serv

- Índices

Los índices dentro de las BD cumplen una función muy similar a los índices que se encuentran en los libros: acelerar el proceso de búsqueda. La utilización de los índices trae consigo ventajas como evitar realizar un recorrido completo de la tabla en la que se busca la información. Así se mejora el rendimiento de la BD y la utilización de los recursos de la computadora, sobre todo en casos en los que las tablas tengan un gran tamaño. Con esto se logra además obtener tiempos de respuesta más rápidos en las consultas. (14)

La estrategia de indexado seleccionada para esta solución es por claves primarias, quedando un total de 21 índices.

Índice	Tabla	Esquema	Tipo	Campos
PK_DPA	dim_dpa	dimension	btree	id_dpa
PK_PAG	dim_indicadores_productos_agricultura_ganaderia	dimension	btree	id_indicador_prod_agric_ganad
PK_PIND	dim_indicadores_productos_industria	dimension	btree	id_indicador_prod_ind
PK_SERV	dim_indicadores_servicios	dimension	btree	id_indicador_Serv
PK_TEMP	dim_temporal	dimension	btree	temporal_id
PK_VA	dim_variante_sector	dimension	btree	id_variante
PK_HEC_PA	hec_indicadores_produc_ganad_agric	hechos	btree	id_dpa, temporal_id, id_variante, id_indicador_prod_agric_ganad
Ref11	hec_indicadores_produc_ganad_agric	hechos	btree	id_dpa
Ref24	hec_indicadores_produc_ganad_agric	hechos	btree	id_indicador_prod_agric_ganad
Ref32	hec_indicadores_produc_ganad_agric	hechos	btree	temporal_id
Ref43	hec_indicadores_produc_ganad_agric	hechos	btree	id_variante
PK_HEC_PIN	hec_indicadores_produc_ind	hechos	btree	id_variante, id_dpa, temporal_id, id_indicador_prod_ind

Tabla 5 Ejemplos de los índices seleccionados

3.2.2. Usuarios y privilegios

- Usuarios y roles

Los usuarios definidos son:

- Administrador_ETL: es el encargado de realizar las actividades de ETL por lo que tiene todos los permisos para realizar las acciones de inserción, selección, actualización y eliminación sobre todas las tablas de la BD.
- AnalistaPT: es la persona encargada de realizar las consultas y reportes con la información de la BD por lo que solo tendrá los permisos para seleccionar los datos de todas las tablas.

- Privilegios

Los privilegios establecidos se aplican a los objetos de tipo tabla y esquema. Ver en el *Anexo 3* un ejemplo del código del script que guarda los elementos de seguridad definidos e imágenes de estos elementos.

Usuario	Permisos	Descripción
postgres	Todos los permisos	Administrador de la BD.
ADMINISTRADOR_ETL	Lectura y escritura	Visualiza, extrae, inserta y actualiza los datos.
AnalistaPT	Lectura	Visualiza los datos.

Tabla 6 Usuarios y privilegios establecidos

3.2.3. Carga de nomencladores

Para completar la solución se debe realizar el proceso de carga de los nomencladores que se recogen en el Modelo 0024-03. A continuación se describen los campos de cada una de las tablas de la solución. Además, en los *Anexos 4* y *5* se pueden ver ejemplos de los scripts utilizados para crear las tablas y cargar los nomencladores respectivamente.

✓ **Tabla DPA**

Llave primaria	Nombre	Tipo de datos	Ejemplo	Descripción
	id_dpa	integer	15	Identificador único de la tabla.
	PROV_CODIGO	char(2)	2	Código único de las provincias.
	PROV_NOMBRE	varchar(25)	La Habana	Nombres de las provincias.

	MUNIC_CODIGO	char(4)	201	Código único de los municipios.
	MUN_NOMBRE	varchar(25)	Mariel	Nombres de los municipios.

✓ **Tabla Temporal**

Llave primaria	Nombre	Tipo de datos	Ejemplo	Descripción
☑	temporal_id	integer	1	Identificador único de la tabla.
	ANNO	char(4)	2009	Año.
	TRIMESTRE_NUMERO	integer	1	Número de trimestre (1, 2, 3 o 4).
	TRIMESTRE_DESCRIPCION	char(15)	Trimestre 1	Descripción del trimestre.

✓ **Tabla Variante_sector**

Llave primaria	Nombre	Tipo de datos	Ejemplo	Descripción
☑	id_variante	integer	1	Identificador único de la tabla.
	VARIANTE_CODIGO	char(4)	1	Código correspondiente a cada variante.
	VARIANTE_DESC	varchar(10)	Estatad	Variante de sector (estadad o no estadad).

✓ **Tabla indicadores_productos_agricultura_ganaderia**

Llave primaria	Nombre	Tipo de datos	Ejemplo	Descripción
☑	id_indicador_prod_agric_ganad	integer	7	Identificador único de la tabla.
	INDICADOR_DESCRIPCION	varchar(100)	Hortalizas	Nombre del indicador.
	TEMATICA_DESCRIPCION	varchar(100)	Plan Turquino	Nombre de la temática.
	SUBTEMATICA_DESCRIPCION	varchar(100)	Tubérculos y raíces	Nombre de la subtemática.

	INDICADOR_CUI	char(11)	1239011000	Código único de cada indicador.
	INDICADOR_UM	char(10)	t	Unidad de medida del indicador.

✓ **Tabla indicadores_productos_industria**

Llave primaria	Nombre	Tipo de datos	Ejemplo	Descripción
✓	id_indicador_prod_ind	integer	1	Identificador único de la tabla.
	INDICADOR_DESCRIPCION	varchar(100)	Pan	Nombre del indicador.
	TEMATICA_DESCRIPCION	varchar(100)	Plan Turquino	Nombre de la temática.
	SUBTEMATICA_DESCRIPCION	varchar(100)	Productos de Industria Alimentaria	Nombre de la subtemática.
	INDICADOR_CUI	char(11)	2343000000	Código único de cada indicador.
	INDICADOR_UM	char(10)	t	Unidad de medida del indicador.

✓ **Tabla indicadores_servicios**

Llave primaria	Nombre	Tipo de datos	Ejemplo	Descripción
✓	id_indicador_Serv	integer	14	Identificador único de la tabla.
	INDICADOR_DESCRIPCION	varchar(100)	Viviendas Terminadas-Población	Nombre del indicador.
	TEMATICA_DESCRIPCION	varchar(100)	Plan Turquino	Nombre de la temática.
	SUBTEMATICA_DESCRIPCION	varchar(100)	Viviendas	Nombre de la subtemática.
	INDICADOR_CUI	char(11)	53111011150	Código único de cada indicador.
	INDICADOR_UM	char(15)	U	Unidad de medida del indicador.

Los nomencladores de la tabla indicadores_servicios no están completos debido a que la búsqueda de los CUI por parte del cliente presenta demoras motivado por la gran cantidad de estos. Además existen dificultades para asignar los CUI para los indicadores que no los presentan.

3.3. Guía de implantación

3.3.1. Requerimientos

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

- Hardware
 - Para el proceso de transformación es necesaria una memoria RAM de 512 MB como mínimo.
 - Para el proceso de visualización e inteligencia de negocio se necesita una memoria RAM de 1 GB como mínimo, para garantizar el correcto funcionamiento del sistema cuando es accedido por varios usuarios simultáneamente.
 - Se necesita un mínimo de 60 GB de capacidad del disco duro para el almacenamiento de la información.
 - Se debe garantizar al menos una impresora, principalmente de tipo Laser jet 1100 que son las utilizadas en estos momentos, para imprimir los reportes de salida.
 - El sistema debe estar conectado directamente a la red.

3.3.2. Secuencia de pasos para la instalación de la BD

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

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

3. Seleccionar la BD creada nuevamente e ir a la Herramienta de Consultas. Cargar el script denominado *DDL Hechos PT.sql* y ejecutarlo. Esto creará el esquema “hechos” con las tablas correspondientes.
4. Seleccionar la BD creada nuevamente e ir a la Herramienta de Consultas. Cargar el script denominado *DML Dimensiones PT.sql* y ejecutarlo. Esto cargará los nomencladores de las tablas del esquema “dimension”.
5. Seleccionar la BD creada nuevamente e ir a la Herramienta de Consultas. Cargar el script denominado *DCL PT.sql* y ejecutarlo. Esto establecerá los privilegios de los usuarios sobre los objetos de la BD.

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

3.4. Validación y pruebas

3.4.1. Lista de chequeo de análisis

Las listas de chequeo se utilizan con la finalidad de contar con una referencia de los puntos eficientes y deficientes que tenga el elemento al que se le aplica la lista. Las listas que se aplicaron son:

- ✓ Lista de chequeo Evaluación de las áreas de la organización: se aplica al artefacto de análisis Evaluación de las áreas de la organización.
- ✓ Lista de chequeo de la Herramienta para la recolección y análisis de la información: se aplica al artefacto de análisis Herramienta para la recolección y análisis de la información.
- ✓ Lista de chequeo Especificación de Requisitos: se aplica al artefacto de análisis Especificación de Requisitos.

Los resultados después de la aplicación (ver en el *Anexo 9* una representación gráfica de los resultados) de las listas de chequeo a sus respectivos artefactos son los siguientes:

Listas de Chequeo	de	Total de indicadores	Cantidad de indicadores evaluados de correctos	Cantidad de indicadores evaluados de incorrectos	Cantidad de indicadores que No Proceden	Evaluación del artefacto
Lista de chequeo Evaluación de las áreas de la organización		10	8	1	1	Bien
Lista de chequeo de la Herramienta para la	la	28	26	-	2	Bien

recolección y análisis de la información					
Lista de chequeo Especificación de Requisitos	30	29	1	-	Bien

Tabla 7 Resultados de las Listas de Chequeo de análisis aplicadas

Estos resultados demuestran que los artefactos a los que se les aplicaron las Listas de Chequeo poseen la calidad requerida por el cliente y cumplen con los parámetros que fueron establecidos por el proyecto DATEC.

3.4.2. Validación de requisitos por el cliente

Se realizó un encuentro con la Especialista de la ONE Elena Leonila Fernández García y Asnioby Hernández López, Máster en Ciencias, donde se expusieron los requisitos definidos así como el diseño del Modelo de Datos. Se discutieron los resultados obtenidos hasta el momento y los clientes expresaron su aprobación con respecto a los requisitos y al Modelo de Datos.

3.4.4. Pruebas de implantación

Se le aplicó un proceso de pruebas a los pasos para la implantación de la BD cuyos resultados se recogen en el documento *CP para la Guía de implantación.doc*. Luego de realizarse el proceso de implantación la BD fue creada con éxito, incluida la creación de las tablas de dimensiones y hechos, la carga de los nomencladores y la otorgación de los permisos para los usuarios definidos.

Con estos resultados se puede afirmar que los artefactos (scripts) generados para creación de las tablas, la carga de datos y los elementos de seguridad funcionan correctamente.

Clases Válidas	Resultado Esperado	Resultado de la Prueba	Observaciones
Creación de nueva BD.	Nueva BD vacía y lista para recibir los datos.	Nueva BD vacía y lista para recibir los datos.	
Carga del script <i>DDL Dimensiones PT.sql</i>	Creación del esquema "dimension" con las tablas correspondientes.	Creación del esquema "dimension" con las tablas correspondientes.	Las tablas dentro del esquema "dimension" son: <ul style="list-style-type: none"> - Dim_dpa - Dim_temporal - Dim_variante_sector - Dim_indicadores_prod_agric_ganad - Dim_indicadores_prod_ind - Dim_indicadores_serv
Carga del script <i>DDL</i>	Creación del esquema	Creación del esquema	Las tablas dentro del esquema "hechos" son:

<i>Hechos PT.sql</i>	“hechos” con las tablas correspondientes.	“hechos” con las tablas correspondientes.	<ul style="list-style-type: none"> - hec_indicadores_produc_ganad_agric - hec_indicadores_produc_ind - hec_indicadores_serv
Carga del script <i>DML Dimensiones PT.sql</i>	Carga de los nomencladores de las tablas del esquema “dimension”	Carga de los nomencladores de las tablas del esquema “dimension”	Los nomencladores de la tabla indicadores_servicios no están completos dado que la búsqueda de los CUI por parte del cliente presenta demoras debido a la gran cantidad de estos.
Carga del script <i>DCL PT.sql</i>	Establecimiento de los privilegios de los usuarios sobre los objetos de la BD.	Establecimiento de los privilegios de los usuarios sobre los objetos de la BD.	Deben otorgarse los permisos: ADMINISRADOR_ETL (con permisos para visualizar, seleccionar, insertar, actualizar y eliminar datos de las tablas) y AnalistaPT (con permisos para visualizar y seleccionar datos de las tablas).

Tabla 8 Pasos de implantación a los que se les realizaron las pruebas

3.5. Conclusiones parciales

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

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

Conclusiones

Se puede concluir que la investigación cumplió los objetivos planteados al inicio de la misma. Además se puede decir que:

- La solución del problema planteado mediante el uso de la tecnología de Mercados de Datos ha demostrado tener resultados satisfactorios.
- Las herramientas seleccionadas soportaron el proceso de diseño e implementación de la solución.
- Las estructuras dimensionales modeladas (tablas de hechos y dimensiones) abarcan la forma de organización del Modelo 0024-03 o de Indicadores del Plan Turquino y soportan el proceso de toma de decisiones.
- Los nomencladores cargados en la BD poseen la estructura especificada por el cliente.
- Las pruebas realizadas validan los artefactos sometidos a ellas y, por lo tanto, la calidad de dichos artefactos como parte de la solución.

Recomendaciones

- Completar la construcción del MD para los Indicadores del Plan Turquino e incorporarlo al AD de la ONE.
- Tomar el contenido de este trabajo como referencia para futuros sistemas de mercados y almacenes de datos que se desarrollen en el país.

Referencias bibliográficas

1. **Ralph Kimball, Margy Ross.** "The Data Warehouse toolkit" Segunda edición. [Online] 2002.
2. **Inmon, William H.** "*What is a Data Warehouse?*". [ed.] Prism. 1995.
3. **Bustos, Jorge.** Business Intelligence y Data Warehousing en Windows. *Danysoft.com*. [Online] 2005.
4. **Nader, Ing. Javier.** "*Sistema de apoyo gerencial universitario*" Tesis de Magister en Ingeniería de Software. 2003.
5. **Inmon, William H.** "*Building the Data Warehouse*" Cuarta edición. [ed.] John Wiley & Sons Inc. 2005.
6. **INEGI.** Consulta interactiva de datos en la página de INEGI en Internet. Antecedentes e importancia. [Online].
7. **autores, Colectivo de.** "METODOLOGÍA PARA EL DESARROLLO DE SOLUCIONES DE ALMACENES DE DATOS E INTELIGENCIA DE NEGOCIO EN CENTALAD". 2009.
8. **Bartolini, Gabriele.** [Online] Noviembre 2009. www.2ndQuadrant.it.
9. **ONE.** *Clasificadores, DPA.xls*.
10. **ONE.** Clasificadores y codificadores. [Online] www.one.cu.
11. **DATEC.** *Especificación de requisitos para el Almacén de Datos de la ONE*. 2009.
12. **Sierra, Julio E. Ortiz.** *Diseño e implementación de un Mercado de Datos para la Oficina Nacional de Estadísticas*. UCI. 2009.
13. Base de Datos relacional. *Rincón del vago*. [Online] http://pdf.rincondelvago.com/base-de-datos-relacional_1.html.
14. Ventajas y desventajas de los índices (Bases de Datos). [Online] Junio 2007. <http://www.tufuncion.com/>.

Bibliografía

1. **Zorrilla, Marta.** *"Data Warehouse y OLAP"*. Universidad de Cantabria : s.n., 2007/08.
2. **Sánchez, Leopoldo Zenaido Zepeda.** *"Metodología para el Diseño Conceptual de Almacenes de Datos"*. Univ. Politécnica de Valencia : s.n., 2008.
3. ONE pone en línea Almacén Central de Datos del Sistema Estadístico Nacional. [Online] 2010. www.one.gob.do.
4. ONE. [Online] 2010. www.one.cu.
5. **Galdiano, José Lázaro.** DATATUR, Almacén de Datos para el Análisis y Difusión de la Información Estadística del Turismo en España. [Online] www.iet.tourspain.es.
6. "BI en la práctica". SIS KLE. [Online] Agosto 20, 2009. www.siskle.com/spanish/index.html.
7. **Vaisman, Alejandro.** *"La investigación en OLAP y Data Warehousing: pasado, presente y futuro"*. Universidad de Buenos Aires/Universidad de Chile : s.n., 2006.
8. **Sierra, Julio E. Ortiz.** *Diseño e implementación de un Mercado de Datos para la Oficina Nacional de Estadísticas*. UCI. 2009.
9. **DATEC.** *Especificación de requisitos para el Almacén de Datos de la ONE*. 2009.
10. **ONE.** Clasificadores y codificadores. [Online] www.one.cu.
11. **Bartolini, Gabriele.** [Online] Noviembre 2009. www.2ndQuadrant.it.
12. **ONE.** *Clasificadores, DPA.xls*.
13. **autores, Colectivo de.** "METODOLOGÍA PARA EL DESARROLLO DE SOLUCIONES DE ALMACENES DE DATOS E INTELIGENCIA DE NEGOCIO EN CENTALAD". [Online] 2009.
14. **INEGI.** Consulta interactiva de datos en la página de INEGI en Internet. Antecedentes e importancia. [Online]
15. **Inmon, William H.** *"Building the Data Warehouse" Cuarta edición*. [ed.] John Wiley & Sons Inc. 2005.
16. **Nader, Ing. Javier.** *"Sistema de apoyo gerencial universitario" Tesis de Magister en Ingeniería de Software*. 2003.
17. **Bustos, Jorge.** Business Intelligence y Data Warehousing en Windows. *Danysoft.com*. [Online] 2005.
18. **Inmon, William H.** *"What is a Data Warehouse?"*. [ed.] Prism. 1995.

19. **Ralph Kimball, Margy Ross.** "The Data Warehouse toolkit" Segunda edición. [Online] 2002.

20. Base de Datos relacional. *Rincón del vago*. [Online] http://pdf.rincondelvago.com/base-de-datos-relacional_1.html.

21. Ventajas y desventajas de los índices (Bases de Datos). [Online] Junio 2007. <http://www.tufuncion.com/>.

22. **Vegas, Jesús.** Introducción al SQL. [Online] <http://cursos.atica.um.es/DBA9i1/sqlplus.html#tiposSentencias>.

Anexos

Anexo 1 Tabla de indicadores y sus respectivos CUI.

Indicador_descripción	Indicador_um	Indicador_cui
Café- Superficie existente en producción	ha	1611000001
Café- Acopio	t	1611000002
Cacao- Superficie existente en producción	ha	1614000001
Cacao- Acopio	t	1614000002
Malanga	t	1240020000
Malanga- Superficie existente sembrada	ha	1240020003
Ñame	t	1240040000
Ñame- Superficie existente sembrada	ha	1240040003
Hortalizas	t	1239011000
Cítricos	t	1320000000
Cítricos- Superficie existente en producción	ha	1320000001
Frutales	t	1310000000
Frutales- Superficie existente en producción	ha	1310000001
Plátano total	t	1310010000
Plátano total- Superficie existente en producción	ha	1310010001
Vacuno-Existencia de rebaño	Miles cabz	2111010001
Vacuno-Producción de leche	Miles L	2111010002
Porcino-Existencia de reproductoras	Miles cabz	2121000001
Porcino-Entregas a sacrificio	cabz	2121000002
Porcino- Entrega a sacrificio (producción de carne)	t	2121000003
Existencia de rebaño ovino	Miles cabz	2112000001
Existencia de rebaño caprino	Miles cabz	2112000002
Producción de leche de cabra	L	2112000003
Entregas a sacrificio (producción de carne Ovino-Caprino)	t	2112000004
Existencia de rebaño equino	cabz	2113010000
Existencia de rebaño asnal	cabz	2113020000
Existencia de rebaño mular	cabz	2113030000

Existencia de yeguas en reproducción con asnal	cabz	2113020005
Existencia de aves ponedoras	Miles U	2122000001
Producción de huevos	Miles U	2920000000
Entregas a sacrificio (producción de carne de aves)	t	2122000005
Entrega de aves semirrusticas	Miles cabz	2122000003
Colmenas existentes	U	2930000001
Producción de miel	t	2930000002
Producción de cera	kg.	2980010000
Producción de propóleos	kg.	2930000004
Madera aserrada de producción nacional	Miles m ³	31100000001
Entregas a exportaciones de resina de pino	t	3230000001
Total plantado	ha	1510000001
Total plantado	Miles U	1510000002
Manejo de la regeneración natural en las plantas	ha	1510000003
Siembra de alevines	MU	4120010011
Siembra de alevines por el MINAG	MU	4120010012
Espejos de agua	ha	4120010013
Espejos de agua del MINAG	ha	4120010014
Producción de pescado	t	4120010000
Producción de pescado por el MINAG	t	4120010001
Distribución de pescado	t	4120010002
Producción de pescado- Financiamiento en divisas	MCUC	4120010003
Turistas días extranjeros	Miles tur/d	63000101010
Habitaciones días existentes	U	63000102000
Habitaciones días ocupadas (extranjeros)	U	63000102011
Presupuesto Ingresos y Gastos		63000103000
Presupuesto Ingresos y Gastos-Fuentes	MCUC	63000103010
Presupuesto Ingresos y Gastos-Destinos	MCUC	63000103020
Pan	t	23430000000
Repostería seca	t	23430010000
Otros Productos de la Ind. Alim. Exc. Pan y Repostería seca	t	20000950010

Financiamiento en divisas para la Industria Alimentaria	MCUC	20000950080
Productos de las Industrias Locales		20000900000
Valor de la producción de las Industrias Locales	MP	20000900010
Financiamiento en divisas para las Industrias Locales	MCUC	20000900080
Recultivación de áreas afectadas por la minería	ha	1510090001
Keroseno	t	33340010000
Alcohol	hL	24110010000
Programa de empleos		405000000
Nuevos empleos	U	405020000
Viviendas-Terminadas total	U	53111011100
Viviendas Terminadas-Sector estatal y cooperativo	U	53111011110
Viviendas Terminadas- MINAG total	U	53111011111
Viviendas Terminadas-MICONS	U	53111011112
Viviendas Terminadas- CAP	U	53111011113
Viviendas Terminadas- Otros(exc. MINAG,MICONS, CAP)	U	53111011114
Viviendas Terminadas- Población	U	53111011150
Conservación y rehabilitación total	U	53111020000
Conservación y rehabilitación-Sector estatal y cooperativo	U	53111020010
Conservación y rehabilitación-Población	U	53111020050
Eliminación de pisos de tierra total	U	53111020001
Financiamiento en divisas total para construcción y reparación de viviendas	MCUC	53111000080
Financiamiento en divisas de la Educ. Superior	MCUC	92300000080
Educación General		92000000000
Financiamiento en divisas de la Educ. General	MCUC	92000000080
Actividades y/o festivas culturales	U	96220000010
Financiamiento en divisas para la Cultura	MCUC	96200000080
Actividades y/o festivas deportivos	U	96500000010
Financiamiento en divisas para el Deporte	MCUC	96500000080
Nacidos vivos	U	93200700000
Defunciones (niños menores de un año)	U	93200801000
Financiamiento en divisas para la Salud	MCUC	93200000080

Financiamiento para inversiones en divisas para la Salud	MCUC	93200000081
Servicios a la población (total)	MP	62000000010
Servicios a la población Tradicional	MP	62000000011
Paralelo	MP	62000000012
Ventas Totales del Comercio Minorista	MP	402013000
Comercio Minorista de Alimentos	MP	62120000000
Comercio Minorista de Productos Industriales	MP	62130000000
Otras ventas minoristas(esc. Alimentos e industriales)	MP	62130090010

Anexo 3 Ejemplo de script de seguridad aplicado a la tabla dim_dpa

ALTER TABLE dimension.dim_dpa OWNER TO "ADMINISTRADOR_ETL";

GRANT SELECT, UPDATE, INSERT, DELETE ON TABLE dimension.dim_dpa TO "ADMINISTRADOR_ETL";

GRANT SELECT ON TABLE dimension.dim_dpa TO "AnalistaPT";

	Object Name	OWN	SEL	INS	UPD	DEL
	dimension.dim_dpa	●	●	●	●	●
	dimension.dim_indicadore	●	●	●	●	●
	dimension.dim_indicadore	●	●	●	●	●
	dimension.dim_indicadore	●	●	●	●	●
	dimension.dim_temporal	●	●	●	●	●
	dimension.dim_variante_s	●	●	●	●	●
	hechos.hec_indicadores_	●	●	●	●	●
	hechos.hec_indicadores_	●	●	●	●	●
	hechos.hec_indicadores_	●	●	●	●	●
	dimension	●				
	hechos	●				

Fig. 8 Privilegios del usuario ADMINISTRADOR_ETL

	Object Name	OWN	SEL	INS	UPD	DEL
	dimension.dim_dpa		●			
	dimension.dim_indicadore		●			
	dimension.dim_indicadore		●			
	dimension.dim_indicadore		●			
	dimension.dim_temporal		●			
	dimension.dim_variante_s		●			
	hechos.hec_indicadores_		●			
	hechos.hec_indicadores_		●			
	hechos.hec_indicadores_		●			

Fig. 9 Privilegios del usuario AnalistaPT

Anexo 4 Ejemplo de script DDL utilizado para crear la tabla dim_dpa

```
CREATE SCHEMA dimension AUTHORIZATION "ADMINISTRADOR_ETL";
SET check_function_bodies = false;
--
-- Structure for table dim_dpa (OID = 37815) :
--
SET search_path = dimension, pg_catalog;
CREATE TABLE dimension.dim_dpa (
    id_dpa integer NOT NULL,
    "PROV_CODIGO" character(2) NOT NULL,
    "PROV_NOMBRE" character varying(25) NOT NULL,
    "MUNIC_CODIGO" character(4) NOT NULL,
    "MUN_NOMBRE" character varying(25) NOT NULL
) WITHOUT OIDS;
```

Anexo 5 Ejemplo de script DML utilizado para cargar los nomencladores de la tabla dim_dpa

```
INSERT INTO dim_dpa (id_dpa, "PROV_CODIGO", "PROV_NOMBRE",  
"MUNIC_CODIGO", "MUN_NOMBRE")
```

```
VALUES (1, '1', 'Pinar del Rio', '101', 'Sandino');
```

```
INSERT INTO dim_dpa (id_dpa, "PROV_CODIGO", "PROV_NOMBRE",  
"MUNIC_CODIGO", "MUN_NOMBRE")
```

```
VALUES (2, '1', 'Pinar del Rio', '102', 'Mantua');
```

```
INSERT INTO dim_dpa (id_dpa, "PROV_CODIGO", "PROV_NOMBRE",  
"MUNIC_CODIGO", "MUN_NOMBRE")
```

```
VALUES (3, '1', 'Pinar del Rio', '103', 'Minas de Matahambre');
```

```
INSERT INTO dim_dpa (id_dpa, "PROV_CODIGO", "PROV_NOMBRE",  
"MUNIC_CODIGO", "MUN_NOMBRE")
```

```
VALUES (5, '1', 'Pinar del Rio', '105', 'La Palma');
```

```
INSERT INTO dim_dpa (id_dpa, "PROV_CODIGO", "PROV_NOMBRE",  
"MUNIC_CODIGO", "MUN_NOMBRE")
```

```
VALUES (7, '1', 'Pinar del Rio', '107', 'Candelaria');
```

Anexo 6 Lista de Chequeo de la Especificación de requisitos

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

	consultarlo en un futuro?				
	5. ¿Debería especificarse algún requisito con más detalle?	0	-	-	Los requisitos se trataron de llevar a la forma más específica posible.
	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	-	-	Primero se realizaron las entrevistas con el cliente, luego se consultaron los posibles requisitos con las analistas del proyecto y posteriormente se plasmaron en el documento como requisitos ya aprobados.
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	0	-	-	-

	entre lo especificado por un requisito y lo especificado por otro?				
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?	1	-	1	-
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	0	-	-	Los requisitos de usabilidad se identificaron, pero de soporte no se detectó ningún requisito.

	identificado?				
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 especificacion	0	-	-	-

	es de requerimiento s?				
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 confidencialida d coincide con el total de páginas que tiene el documento?	0	-	-	-

Anexo 7 Lista de Chequeo para la Evaluación de las áreas de la organización

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

Semántica del documento

Peso	Indicadores a Evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
Crítico	5. ¿No se han identificado errores ortográficos?	0	-	-	-
Crítico	6. ¿Se entiende claramente lo que se ha especificado en el	0	-	-	-

	documento?				
	7. ¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?	-	NP	-	-
	8. ¿El total de páginas que aparecen en las reglas de confidencialidad coincide con el total de páginas que tiene el documento?	0	-	-	-

Anexo 8 Lista de Chequeo de la Herramienta para la recolección y análisis de la información

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

Critico	¿Las necesidades del cliente han sido identificadas?	0	-	-	-
	¿Los niveles de acceso para cada pedido de información se han identificado correctamente?	0	-	-	-
	¿Han sido identificados los roles que tendrán acceso a ver la información de cada pedido?	-	NP	-	La información es pública por lo que no se necesita especificar el rol o grupo que tiene acceso al elemento.
Critico	¿Se ha determinado la periodicidad con que se recoge la información de cada pedido?	0	-	-	-
Critico	¿Han sido clasificados los pedidos de información de acuerdo a la información que manejan?	0	-	-	-
Critico	¿Los pedidos de información se han agrupado en casos de uso informativos, según un criterio que	0	-	-	-

	los describa?				
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	0	-	-	-

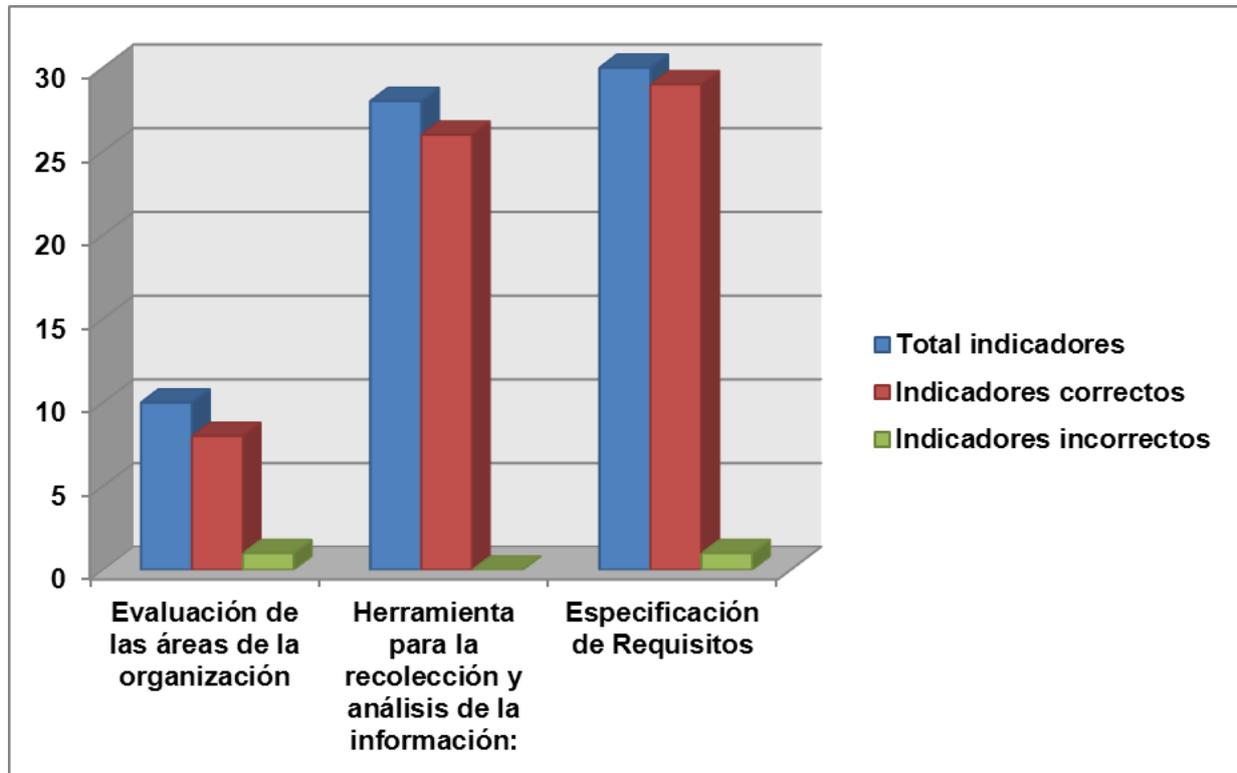
	variable de entrada?				
Critico	¿Todas las variables de salida han sido identificadas?	0	-	-	-
Critico	¿Se ha definido la unidad de medida en que se va a medir cada variable de salida?	0	-	-	-
Critico	¿Se ha determinado el tipo de variable para cada una de las variables de salida identificadas?	0	-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

	tipos de operaciones aditivas que se pueden realizar?				
Critico	¿Se ha especificado a que caso de uso del sistema pertenece cada variable de salida?	-	NP	-	En este caso los CU del sistema no tiene relación con las variables de salida, ya que se centran en el proceso de carga de los datos.
	¿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	-	-	Algunas palabras como, por ejemplo, los nombres de las variables, no tienen tildes o la letra "ñ" debido a que pueden ocasionar problemas durante la carga de la BD.
Critico	¿Se entiende claramente todo lo especificado en el Excel?	0	-	-	-

Anexo 9 Resultados de la aplicación de las Listas de Chequeo a los artefactos del análisis



Glosario de términos

- **ANSI SQL 2003:** Estándar adoptado por la mayoría de los SGBD. Esta versión incluía algunas características de XML, cambios en las funciones y la estandarización del objeto *sequence* y las columnas auto numéricas.
- **CASE:** Ayuda a la Ingeniería de Software por computadora.
- **Centro informante:** Empresas, unidades presupuestadas, entidades subordinadas a las organizaciones políticas, sociales y de masas, cooperativas y unidades básicas de producción cooperativa que envían sus estadísticas a la ONE.¹
- **CU:** Caso de Uso.
- **CUBAENERGÍA:** Centro de la Información y Desarrollo de la Energía. Entidad presupuestada de investigación-desarrollo y servicios científico-técnicos en materia de energía y medio ambiente.
- **CUI:** Código Único de Indicador.
- **DATEC:** Centro de Tecnologías de Gestión de Datos.
- **DCL:** Lenguaje de Control de Datos.
- **DDL:** Lenguaje de Definición de Datos.
- **DML:** Lenguaje de Manipulación de Datos.
- **Estadística:** Estudio de los datos cuantitativos de la población, de los recursos naturales e industriales, del tráfico o de cualquier otra manifestación de las sociedades humanas.²
- **GPL:** Licencia Pública General de GNU (compañía dedicada a la creación y promoción del software libre). Esta licencia contiene todo aquel software que se considera de libre distribución y de código abierto, permitiéndoles a sus usuarios modificar el programa sin incurrir en incumplimientos o faltas con respecto a su licencia.³
- **Herramientas CASE:** Herramienta de software que automatiza una parte del ciclo de desarrollo de este.
- **Indicador:** Herramientas para clarificar y definir, de forma más precisa, objetivos e impactos, son medidas verificables de cambio o resultado diseñadas para contar con un estándar contra el cual evaluar, estimar o demostrar el progreso con respecto a metas establecidas.⁴
- **Plan Turquino:** se crea en Cuba con el fin de lograr un desarrollo integral y sostenible de las zonas montañosas. Aspectos fundamentales del Plan Turquino

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

² Tomado de: Microsoft Encarta 2009 (febrero 2010)

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

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

han sido: el desarrollo social y cultural, la elevación de las condiciones de vida de los pobladores de las montañas, el fortalecimiento de los asentamientos humanos en estos territorios para ayudar a alcanzar las metas económicas, la conservación del medio ambiente y la defensa del país desde las montañas.⁵

- **Plugin:** Aplicación informática que, interactuando con otra aplicación, le aporta una función específica a ésta última.
- **Script DCL:** Script que permite al administrador de la BD cargar los usuarios y privilegios. Incluye los tipos de consultas SQL **GRANT** y **REVOKE**.
- **Script DDL:** Script que permite definir los objetos de la BD. Incluye los tipos de consultas SQL **CREATE** y **ALTER**.
- **Script DML:** Script que permite manejar los datos de la BD. Incluye los tipos de consultas SQL **SELECT**, **INSERT**, **UPDATE** y **DELETE**.⁶
- **TCP/IP:** Protocolo de control de transmisión/Protocolo de Internet. Protocolo utilizado para asignar una dirección única para cada computadora en la red, dividir los mensajes en paquetes, enrutar los datos en la red y detectar errores en dichos datos.⁷
- **Tupla:** Llamadas también registros o filas. Contienen la información relativa a una única entidad.⁸
- **UCI:** Universidad de Ciencias Informáticas.

⁵ Tomado de: Sitio web "50 años de Revolución cubana". <http://www.revolucioncubana.cip.cu> (febrero 2010)

⁶ Tomado de: Introducción al SQL. <http://cursos.atika.um.es/DBA9i1/sqlplus.html#tiposSentencias>.

⁷ Tomado de: <http://es.kioskea.net/contents/internet/tcpip.php3>

⁸ Tomado de: Curso de MySQL.
http://www.google.com/cu/url?sa=t&source=web&ct=res&cd=3&ved=0CBgQFjAC&url=http%3A%2F%2Fmysql.conclase.net%2Fcurso%2Findex.php%3Fcap%3D003&rct=j&q=define+tupla&ei=qn3gS6HbNsOC8gbD48iRbW&usq=AFQjCNHGzPo5_VSerNR-dOVRkZtjilhKAA