

Universidad de las Ciencias Informáticas

Facultad 6



Título: Sistema de Información de Gobierno. Mercado de datos Plan Turquino.

Trabajo de Diploma para Optar por el título de Ingeniero en Ciencias Informáticas.

Autores:

Maray García Méndez.

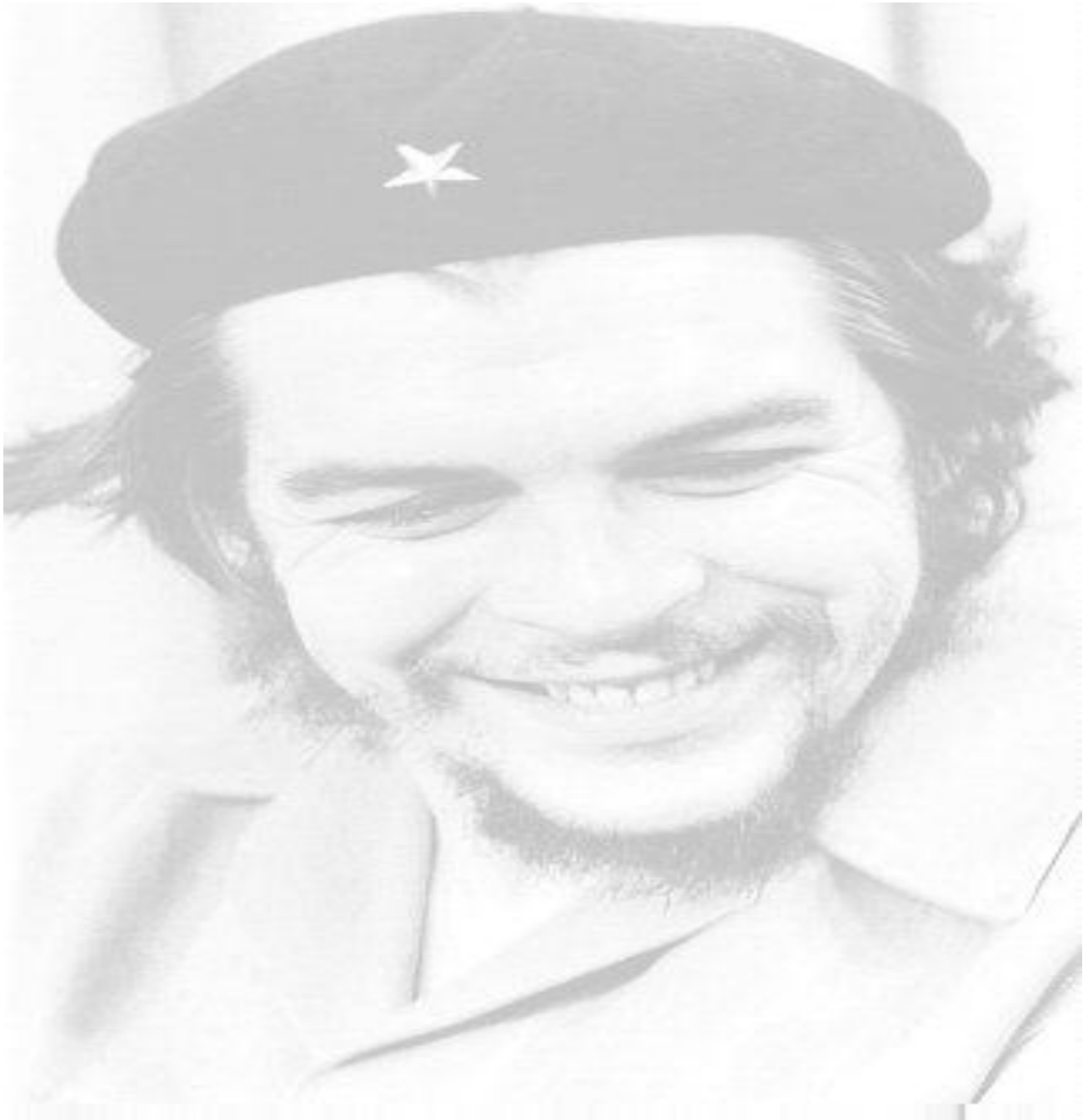
Edelis Castilla Rodríguez.

Tutores:

Ing. Irina Collazo Ávila.

CIUDAD DE LA HABANA

CUBA, Junio 2011.



"Si el presente es de lucha, el futuro es nuestro."

Che

DECLARACIÓN DE AUTORÍA

Declaramos ser autores de la presente tesis y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

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

Firma del Autor
Maray García Méndez

Firma del Autor
Edelis Castilla Rodríguez

Firma del Tutor
Ing. Irina Collazo Ávila

DATOS DE CONTACTO

Ing. Irina Collazo Ávila

Ingeniera en Ciencias Informáticas, graduación del 2007. Imparte la asignatura de Matemática en la facultad 6 y pertenece a la línea de almacenes del centro DATEC.

e-mail: icollazo@uci.cu

DEDICATORIA

De Maray:

Esta tesis está dedicada a la memoria de mi papá, que lamentablemente no pudo ver realizado mis sueños, pero sé que estaría muy orgulloso de su niña y que desde donde está siempre me apoyó y siguió mis pasos.

A mi mamá por confiar siempre en mí, por dándome cariño, amor y todo el apoyo del mundo, te quiero. A mis abuelos maternos y paternos que gracias a dios los tengo a todos a mi lado, siempre al tanto de su nieta más chiquita. A mi abuelo Marcelo que aparte de ser abuelo es un padre para mí. Y en especial a mi abuelita querida Olga, que es lo que más quiero en el mundo, por cuidar mucho de mí, darme fuerza y estar siempre a mi lado. A mi hermana Marbely por estar siempre al tanto de mí. A mi sobrina que es lo más lindo que tengo. A Maidoly que ha sido otra hermana para mí, hemos compartido tantas cosas juntas, en los momentos buenos y malos siempre ha estado a mi lado. A mis suegros Ana y Vicente que han sido otra familia y siempre me aconsejaron y animaron a seguir hasta el final. A mi novio que ha estado a mi lado en estos 5 años de carrera dándome ánimo y sobre todo mucho amor, con él he aprendido muchas cosas, te amo. En general a toda mi familia que siempre me apoyaron y confiaron en mí.

De Edelis:

Dedico esta tesis a mis padres, a mi hermano y a toda mi familia porque son la razón de mi vida, porque han sido la luz que ha iluminado mi camino en toda mi trayectoria de estudio, por la confianza que pusieron en mí cuando lo creía todo perdido, por su amor y comprensión incondicional y porque a ellos le debo lo que soy hoy.

AGRADECIMIENTOS

De Maray:

Agradezco a todos los que de una forma u otra tuvieron que ver con el resultado de esta tesis.

A mi familia por el apoyo que siempre me dieron.

A mis compañeros del grupo que nos estuvimos ayudando durante estos meses.

A la tutora Irina, por toda la ayuda que nos brindó. Por permitirnos ocupar su tiempo, y siempre estar dispuesta a colaborar con nosotros, muchas gracias. También agradecerle a los profesores del tribunal y al oponente por brindarnos su apoyo.

Por último pero no menos importante a mis amigas y amigos que conocí a lo largo de esta carrera. En especial a los que estuvieron a mi lado desde primer año compartiendo momentos buenos y malos, Betty, Ayleneis, Alejandro, Raiza y Edelis, esa chiquitica pero de corazón grande que siempre me ha comprendido y ha aguantado esos momentos de mal carácter por lo que uno pasa. Nunca olvidaré este tiempo de mi vida en la universidad porque gracias a eso conocí lo bello que es la amistad y que no importa de qué lugar seas. Siempre estarán en mi corazón. Gracias a todos por brindarme su amistad.

AGRADECIMIENTOS

De Edelis:

Agradezco a mis padres y familiares por el apoyo que me han dado siempre, por darme ánimos en días de tristeza y por estar a mi lado cuando los necesito. Un beso enorme para ellos, los quiero mucho

A mi madrina Sari por guiar mi camino y cada paso que doy.

Agradezco a mi compañera de tesis la enorme paciencia que ha tenido conmigo, porque de no haber sido por ella quizás hoy no me sintiera así de feliz con nuestros resultados. Por entenderme y ayudarme cuando me sentía muy insegura de mi misma y por ser como es, tan especial y sencilla.

Le agradezco muchísimo a esas personitas que día a día conviven conmigo y que son como hermanos para mí: Bety que me ha calmado en momentos difíciles, Ayle que con su enorme sonrisa me ha alegrado siempre, Ale porque nunca me dejó sola y me ha cuidado mucho y a Mara por ser mi conciencia aquí en la escuela. Los extrañaré mucho.

También agradezco a Oscar que aunque no se encuentra aquí ahora me supo ayudar siempre, a Julio, a Ricardo por haber estado en las buenas y malas conmigo brindándome apoyo y cariño, a mis compañeros de grupo tanto los del actual como con los que compartí en otros años, a todos mis profesores que a lo largo de mi carrera he conocido, a los del proyecto, a mi tutora Irina, a Yonelbys y a todos los integrantes del tribunal.

De forma general agradezco a todos los que me apoyaron, a los que han compartido conmigo, a los que hicieron posible que hoy fuera ingeniera y a la vida por permitirme vivir y haberlos conocido.

Resumen

Resumen

Este Trabajo de Diploma surge en el marco de trabajo del proyecto DATEC, perteneciente a la facultad 6 de la Universidad de las Ciencias Informáticas, en colaboración con la Oficina Nacional de Estadísticas. El proyecto tiene como objetivo desarrollar un almacén de datos llamado Sistema de Información de Gobierno, el cual gestiona información asociada a varios mercados de datos. A partir de la creación de estos mercados de datos, surge la necesidad de integrar la información de los datos estadísticos del país referentes a los indicadores de las zonas montañosas, incluyendo la Ciénaga de Zapata, proceso que resultaba algo engorroso de resolver. Para la creación de este mercado de datos, se hizo necesario realizar un proceso de extracción, transformación y carga de los datos, además de visualizar los mismos. Con el resultado de la aplicación se ha mejorado la calidad de las consultas realizadas a los datos de una forma más rápida y fácil. La implementación de la aplicación hará posible una mejora total en cuanto a condiciones de trabajo de los usuarios, evitándoles el desgaste y agotamiento que produce el procesamiento manual de la información.

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

Índice

Índice

| | |
|--|----|
| Introducción | 1 |
| Capítulo I: Fundamentación Teórica | 5 |
| Introducción | 5 |
| 1.1 Marco Teórico | 5 |
| 1.1.1 Almacenes de Datos..... | 5 |
| 1.1.2 Base de datos multidimensionales | 6 |
| 1.1.3 Proceso OLAP | 7 |
| 1.1.4 Mercado de datos..... | 7 |
| 1.1.5 Metadatos | 8 |
| 1.1.6 Granularidad | 8 |
| 1.1.8 Inteligencia de Negocio..... | 10 |
| 1.2 Marco conceptual teórico | 10 |
| 1.3 Herramientas y Metodología | 11 |
| 1.3.1 Metodología de Desarrollo..... | 11 |
| 1.3.2 Herramienta de Desarrollo. | 13 |
| 1.3.3 Herramienta de Modelado | 14 |
| 1.3.4 Herramienta de perfilado de datos..... | 14 |
| 1.3.5 Herramienta de ETL..... | 15 |
| 1.3.6 Herramientas de BI | 15 |
| 1.4 Conclusiones parciales..... | 16 |
| Capítulo II Análisis y diseño del mercado de datos | 17 |
| Introducción | 17 |
| 2.1 Estudio preliminar del negocio..... | 17 |
| 2.2 Definición del negocio | 17 |
| 2.3 Roles y permisos | 18 |
| 2.4 Reglas de Negocio | 19 |
| 2.5 Necesidades de información | 21 |

Índice

| | |
|---|----|
| 2.5.1 Requisitos de información..... | 21 |
| 2.5.2 Requisitos multidimensionales | 22 |
| 2.5.3 Requisitos funcionales | 23 |
| 2.5.4 Requisitos no funcionales | 24 |
| 2.6 Descripción de Casos de Uso | 26 |
| 2.6.1 CU Extraer los datos. | 27 |
| 2.6.2 CU Realizar Transformación y Carga de los datos. | 28 |
| 2.6.3 CU Autenticar Usuario..... | 28 |
| 2.7 Desarrollo de la matriz BUS..... | 30 |
| 2.8 Identificación de dimensiones, hechos y medidas..... | 31 |
| 2.8.1 Dimensiones identificadas | 31 |
| 2.8.2 Hechos identificados | 34 |
| 2.8.3 Medidas | 35 |
| 2.9 Modelo de datos | 36 |
| 2.10 Conclusiones parciales | 37 |
| Capítulo III Implementación del Mercado de Datos. | 38 |
| Introducción | 38 |
| 3.1 Modelo de datos físico | 38 |
| 3.1.1 Esquemas y tablas | 38 |
| 3.2 Implementación del subsistema de integración | 40 |
| 3.2.1 Arquitectura | 40 |
| 3.2.2 Perfilado de datos..... | 40 |
| 3.2.3 Extracción de los datos | 42 |
| 3.2.4 Limpieza y Transformación de los datos | 42 |
| 3.2.5 Carga de los datos..... | 42 |
| 3.2.6 Implementación del Job | 43 |
| 3.3 Implementación del subsistema de visualización | 44 |
| 3.3.1 Implementación de los cubos OLAP | 44 |

Índice

| | |
|---|----|
| 3.3.2 Navegación de la capa de visualización | 45 |
| 3.3 Conclusiones parciales | 48 |
| Capítulo IV: Validación y pruebas del Mercado de Datos | 49 |
| Introducción | 49 |
| 4.1 Lista de chequeo..... | 49 |
| 4.1.1 Elementos que forman parte de la estructura de la lista de chequeo..... | 49 |
| 4.2 Pruebas | 53 |
| 4.2.1 Modelo V..... | 53 |
| 4.2.2 Casos de Prueba | 55 |
| 4.3 Conclusiones parciales..... | 56 |
| Conclusiones | 57 |
| Recomendaciones | 58 |
| Referencias bibliográficas | 59 |
| Bibliografía..... | 60 |
| Anexos..... | 62 |
| Glosario de términos..... | 63 |

Índice de Figuras

Índice de Figura

| | |
|---|----|
| Fig. 1 Granularidad | 9 |
| Fig. 2 Diagrama CU del Sistema | 30 |
| Fig. 3 Diseño del modelo de datos | 37 |
| Fig. 4 Estructura física de la BD | 39 |
| Fig. 5 Arquitectura de la integración de datos | 40 |
| Fig. 6 Ejemplo de transformación..... | 43 |
| Fig. 7 Job | 44 |
| Fig. 8 Diseño de los cubos..... | 45 |
| Fig. 9 Estructura de la navegación..... | 47 |
| Fig. 10 Ejemplo del reporte Agricultura, Ganadería, Silvicultura..... | 47 |
| Fig. 11 Rol Administrador Fig. 12 Rol Analista | 48 |
| Fig. 13 Modelo de Ciclo de Vida en V | 54 |
| Fig. 14 Ejemplo del Caso de Prueba basado en CU (Industria) | 56 |

INTRODUCCIÓN

Introducción

El desarrollo de la Industria del Software, en los últimos años, se ha convertido en un elemento esencial para el incremento de la productividad en diferentes renglones en diversos países, siendo los software estadísticos una fuerte base para apoyar estas esferas. Actualmente se lleva a cabo en la mayoría de las sociedades, la informatización masiva de las mismas, en la que algunos buscan soluciones rentables a sus problemas, apoyándose principalmente en la informática, la cual se ha convertido en el principal soporte para tener el control de grandes volúmenes de información.

El uso de aplicaciones estadísticas ha permitido identificar problemas antes que estos ocurran, evaluar nuevos procedimientos y su impacto en la producción así como organizar, resumir, presentar y analizar datos relativos a un conjunto de objetos, personas y procesos mediante un análisis más detallado de los mismos, permitiendo elevar el nivel de eficiencia y eficacia en la toma de decisiones estratégicas para el desarrollo de la sociedad actual.

Nuestro país no está ajeno a esto, es por ello que el 21 de abril de 1994 fue creada la Oficina Nacional de Estadísticas (ONE), órgano rector de las estadísticas en Cuba. Su principal objetivo se fundamenta en captar, analizar y difundir los datos recogidos a lo largo y ancho de todo el país. Para ello cuenta con una serie de modelos estadísticos en los cuales se recoge información de todos los sectores de la economía y la sociedad. El uso de herramientas informáticas en este sector facilita los cálculos matemáticos y la elaboración de gráficos, lo cual permite a la máxima dirección del país la confección de las estrategias nacionales a mediano y largo plazo.

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. (1)

INTRODUCCIÓN

Por lo engorroso que se hacía realizar este trabajo, se le da la tarea a la Universidad de las Ciencias Informáticas (UCI), específicamente al Centro de Tecnologías de Gestión de Datos (DATEC) perteneciente a la facultad 6, de crear mercados de datos para facilitar la gestión de información. Actualmente en la ONE existe un almacén de datos llamado Sistema de Información de Gobierno (SIGOB) que está compuesto por varios mercados de datos. Uno de estos está asociado al área Plan Turquino encargada de recoger los datos estadísticos en las zonas montañosas del país, incluyendo la Ciénaga de Zapata.

En el curso 2009-2010 se crea una primera versión de lo que sería el mercado de datos para el Plan Turquino, pero dicho estudio no responde a todos los intereses por parte del cliente, debido a que se quedó sin completar unas de las dimensiones recogidas en el levantamiento de requisitos, por lo que se hace necesario realizar un nuevo análisis del proceso de negocio para su completa validación, unido a que la aplicación existente no cuenta con una capa de inteligencia del negocio ni con un proceso de integración de datos.

Por todo lo anteriormente planteado surge como **Problema de la Investigación:** ¿Cómo contribuir a la toma de decisiones en el área Plan Turquino del Sistema de Información de Gobierno?

Definiéndose como **Objeto de Estudio** los Almacenes de Datos (AD) e Inteligencia de Negocio (BI), y como **Campo de Acción** los Mercado de Datos (MD) y la capa de visualización para el área Plan Turquino del Sistema de Información de Gobierno.

Determinando como **Objetivo General:** Desarrollar el mercado de datos Plan Turquino del Sistema de Información de Gobierno que contribuya a la toma de decisiones.

Como **Objetivos Específicos** se tienen:

1. Refinar el análisis y diseño del mercado de datos del área Plan Turquino.
2. Implementar el mercado de datos del área Plan Turquino.
3. Implementar la capa de visualización de datos del área Plan Turquino.
4. Validar el mercado de datos del área Plan Turquino.

Para cumplir con los **objetivos específicos**, se trazaron las siguientes **Tareas de la Investigación:**

INTRODUCCIÓN

1. Caracterización de las metodologías, herramientas y tecnologías a utilizar en el desarrollo de almacenes de datos.
2. Refinamiento de los requisitos del mercado de datos.
3. Redefinición de los casos de usos del mercado de datos.
4. Redefinición de los hechos, las medidas y las dimensiones del mercado de datos.
5. Restructuración del modelo de datos.
6. Redefinición de la arquitectura del mercado de datos.
7. Restructuración del subsistema de integración.
8. Restructuración del subsistema de visualización.
9. Restructuración de los casos de pruebas.
10. Implementación del subsistema de integración.
11. Implementación del subsistema de visualización.
12. Aplicación de las listas de chequeo.
13. Aplicación de los casos de pruebas.

Este documento estará estructurado de la siguiente forma:

Capítulo 1: Fundamentación teórica de los almacenes de datos.

Este capítulo aborda varios elementos teóricos referentes a los Almacenes de Datos y Mercados de Datos. Además se hace referencia a las principales características de las herramientas que serán utilizadas en el transcurso de la investigación así como las principales razones que se tuvieron en cuenta para la selección de la metodología a utilizar.

Capítulo 2: Análisis y diseño del mercado de datos Plan Turquino.

En este capítulo se abordará todo lo referente al proceso de análisis y diseño del mercado de datos, donde se hará un estudio preliminar del negocio para facilitar una mejor comprensión del desarrollo de la solución. También se definirán los requisitos de información, funcionales, no funcionales y multidimensionales, así como las reglas del negocio. Se realizarán las descripciones de los casos de usos,

INTRODUCCIÓN

se desarrollará la matriz BUS, además de identificar los hechos, medidas y dimensiones. Se definirán los reportes candidatos y el diseño del proceso de integración.

Capítulo 3: Implementación del mercado de datos Plan Turquino.

En este capítulo se abordará todo lo referente al proceso de integración de datos que contiene los tres tipos de actividades de carácter general como son: extracción, transformación y carga en el mercado de datos. Además de implementar el subsistema de visualización, modelo de datos, cubos OLAP y reportes candidatos.

Capítulo 4: Validación y prueba del Mercado de Datos.

En este capítulo se realizarán las validaciones al sistema y las pruebas pertinentes que comprueben la calidad del producto, como son los casos de prueba y la lista de chequeo.

Capítulo I: Fundamentación Teórica

Introducción

Este capítulo aborda varios elementos teóricos referentes a los Almacenes de Datos y Mercados de Datos. Además se hace referencia a las principales características de las herramientas que serán utilizadas en el transcurso de la investigación así como las principales razones que se tuvieron en cuenta para la selección de la metodología a utilizar.

1.1 Marco Teórico

1.1.1 Almacenes de Datos

Un almacén de datos (AD) es una base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes distintas, para luego procesarla permitiendo su análisis desde infinidad de perspectivas y con grandes velocidades de respuesta. (2)

A principios de los años noventa, William H. Inmon, quien se considera uno de los padres de los AD, definió: “Un Almacén de Datos, es una colección de datos orientados a temas, integrados, variables en el tiempo y no volátiles, que se utilizan como apoyo al proceso de toma de decisiones”. (2)

Según definió el propio William H. Inmon, un AD se caracteriza por ser:

- **Temático:** la información se clasifica en función de los aspectos que son de interés para la organización. Se organizan por temas para facilitar el acceso y el entendimiento por parte de los usuarios finales.
- **Integrado:** los datos almacenados en el AD deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas. La información suele estructurarse también en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.
- **Histórico:** el tiempo es parte implícita de la información contenida en un AD. En los sistemas operacionales, los datos siempre reflejan el estado de la actividad del negocio en el momento presente. Por el contrario, la información almacenada sirve, entre otras cosas, para realizar análisis

Capítulo I Fundamentación Teórica

de tendencias. Por lo tanto, el AD se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.

- **No volátil:** el almacén de información de un AD existe para ser leído, pero no modificado. La información es por tanto permanente, significando la actualización del AD la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía.

1.1.2 Base de datos multidimensionales

Las bases de datos multidimensionales, proveen una estructura que permite tener acceso a los datos, para explorar y analizar sus relaciones, y resultados consiguientes. Estas se pueden visualizar como un cubo multidimensional, en donde las variables asociadas existen a lo largo de varios ejes o dimensiones, y la intersección de las mismas representa la medida, indicador o el hecho que se está evaluando. (3)

Las **Tablas de hechos** contienen los hechos, medidas o indicadores que serán utilizados por los analistas de negocio para apoyar el proceso de toma de decisiones. El registro del hecho posee una clave primaria que está compuesta por las claves primarias de las tablas de dimensiones relacionadas a este.

Las **Tablas de dimensiones** definen como están los datos organizados lógicamente y proveen el medio para analizar el contexto del negocio. Representan los ejes del cubo, y los aspectos de interés, mediante los cuales el usuario podrá filtrar y manipular la información almacenada en la tabla de hechos.

Existen tres formas de modelar un AD:

- **Esquema estrella:** consta de una tabla de hechos central y de varias tablas de dimensiones relacionadas a esta, a través de sus respectivas claves. (3)
- **Esquema copo de nieve:** es un esquema derivado del de estrella, donde las tablas de dimensiones se ramifican en más puntas. (3)
- **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. (3)

1.1.3 Proceso OLAP

Para crear el MD de un área funcional de la empresa es preciso encontrar la estructura óptima para el análisis de su información, estructura que puede estar montada sobre una base de datos de Procesamiento Analítico en Línea (OLAP).

- **OLAP:** Permite un uso más eficaz de los almacenes de datos para el análisis en línea, lo que proporciona respuestas rápidas a consultas analíticas complejas e iterativas. Los modelos de datos multidimensionales de OLAP y las técnicas de agregados de datos organizan y resumen grandes cantidades de datos para que puedan ser evaluados con rapidez mediante el análisis en línea y las herramientas gráficas. Los sistemas OLAP proporcionan la velocidad y la flexibilidad necesaria para dar apoyo al analista en tiempo real, es por esto que se define ésta estructura para la construcción del MD.

Existen tres tipos de OLAP:

- **R-OLAP (Procesamiento Analítico en línea Relacional)**
Arquitectura de Base de Datos Multidimensional en la que los datos se encuentran almacenados en una Base de Datos Relacional, normalmente con en forma de estrella (copo de nieve, araña). (4)
- **M-OLAP (Procesamiento Analítico en línea Multidimensional)** Arquitectura de Base de Datos Multidimensional en la que los datos se encuentran almacenados en una Base de Datos Multidimensional, que mejora los tiempos de acceso a costa de mayores necesidades de almacenamiento y retardos en las modificaciones. (4)
- **H-OLAP (Procesamiento Analítico en línea Híbrido)**
Arquitectura que combina las tecnologías ROLAP y MOLAP. En HOLAP, el soporte de almacenamiento de datos y el motor de generación de vistas contienen elementos de ambas tecnologías. Pretende combinar las ventajas de cada una sin sus inconvenientes. (4)

1.1.4 Mercado de datos

Un mercado de datos corresponde a un pequeño almacén de datos, específico a un área de negocio o departamento de la empresa. (5)

Capítulo I Fundamentación Teórica

Traen consigo el concepto de almacén de datos. Tienen mayor utilidad para grupos más pequeños que deseen acceder a datos detallados. El motivo por el cual se crean MD es el crecimiento que tiene el almacén, y así facilitar su construcción y utilización.

Características que presentan los MD:

- Se centran en los requisitos de los usuarios, asociados a un departamento o área de negocio concreto.
- Como diferencia con los almacenes de datos, los mercados no contienen datos operacionales detallados.
- Son más sencillos a la hora de utilizarlos y comprender sus datos, debido a que la cantidad de información que contienen es mucho menor que en los almacenes de datos.

1.1.5 Metadatos

Uno de los componentes más importantes de la arquitectura de un almacén de datos son los metadatos. Se define comúnmente como "información sobre información" o "datos sobre los datos", en el sentido de que se trata de datos que describen cuál es la estructura de los datos que se van a almacenar y cómo se relacionan. (6)

Estos pueden describir colecciones de objetos y también los procesos en los que están involucrados, detallando cada uno de los eventos, sus componentes y cada una de las restricciones que se les aplican. Los metadatos definen las relaciones entre los objetos, como las tuplas en una base de datos o clases en orientación a objetos, generando estructuras.

1.1.6 Granularidad

La granularidad representa el nivel de detalle o resumen de las unidades de datos en el almacén de datos. Mientras más detalle exista, más bajo será el nivel de granularidad (Fig. 1.a) y a menor detalle, más granularidad (Fig. 1.b). (1)

Capítulo I Fundamentación Teórica



FIG. 1 GRANULARIDAD

Mientras mayor sea el nivel de detalle de los datos, se tendrán mayores posibilidades analíticas, ya que los mismos podrán ser resumidos. Es decir, los datos que posean granularidad fina (nivel de detalle) podrán ser resumidos hasta obtener una granularidad media o gruesa. No sucede lo mismo en sentido contrario, ya que los datos almacenados con granularidad media podrán resumirse, pero no tendrán la facultad de ser analizados a nivel de detalle.

1.1.7 Extracción, Transformación y Carga (ETL).

Los procesos de extracción, transformación y carga, generalmente conocidos por sus siglas en inglés (*Extraction, Transformation and Loading*) (ETL) son importantes ya que constituyen la forma en que los datos se guardan en un almacén de datos (o en cualquier base de datos). Los cuales son preparados de la forma más adecuada para facilitar el análisis.

Según **Ralph Kimball** los procesos fundamentales para la integración de los datos en un almacén son: (7)

- **Extracción:** se extraen los datos de los sistemas transaccionales hacia el área de preparación del AD, donde son transformados y limpiados.
- **Transformación:** una vez que la información es extraída hacia el área de tráfico de datos, pasa por procesos de transformación y limpieza, seleccionándose únicamente los campos necesarios para el AD. Estas transformaciones son necesarias para asegurar la calidad de los datos que

Capítulo I Fundamentación Teórica

serán almacenados en el AD e incluyen, entre otras, la corrección de errores, fallas de integridad referencial, eliminación de redundancia y la resolución de inconsistencias.

- **Carga:** consiste en mover los datos desde las fuentes operacionales o el almacenamiento intermedio hasta el AD y cargar los datos en las correspondientes estructuras de datos.

1.1.8 Inteligencia de Negocio

La Inteligencia de Negocios (BI según sus siglas en inglés) es el conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento mediante el análisis de datos existentes en una organización o empresa (8).

Este conjunto de herramientas y metodologías tienen en común las siguientes características:

- **Accesibilidad a la información.** Los datos son la fuente principal de este concepto. Lo primero que deben garantizar estos tipos de herramientas y técnicas es el acceso de los usuarios a los datos con independencia de la procedencia de estos.
- **Apoyo en la toma de decisiones.** Se busca ir más allá en la presentación de la información, de manera que los usuarios tengan acceso a herramientas de análisis que les permitan seleccionar y manipular sólo aquellos datos que les interesen.
- **Orientación al usuario final.** Se busca independencia entre los conocimientos técnicos de los usuarios y su capacidad para utilizar estas herramientas.

1.2 Marco conceptual teórico

En la versión anterior al mercado de datos Plan Turquino. Las estructuras dimensionales (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 poseen la estructura especificada por el cliente. Cuenta con una descripción del proceso de carga de la información a un modelo físico.

Capítulo I Fundamentación Teórica

Esta primera versión no se concluyó producto de que unas de las dimensiones recogidas en el levantamiento de requisitos no finalizó, además de que no cuenta con una capa de inteligencia del negocio ni con un proceso de integración de datos.

1.3 Herramientas y Metodología

El Grupo de Desarrollo de Soluciones para la línea de Almacenes de Datos del proyecto DATEC, después de haber hecho un estudio pertinente y profundo acerca de las diferentes metodologías y herramientas existentes en el mundo para la implementación de los almacenes de datos, definió las que a continuación se exponen, para su uso en esta investigación.

1.3.1 Metodología de Desarrollo

➤ **Metodología para el Desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio en DATEC.**

Esta metodología es una adaptación que hizo el Grupo de Desarrollo de Soluciones para la línea de Almacenes de Datos del proyecto DATEC (9). Basándose en las características que presenta la metodología de Kimball, que son las siguientes.

- Utiliza los conceptos Hechos y Dimensiones, lo que indudablemente es muy eficaz en el proceso de la toma de decisiones y proporciona mayor agilidad en el proceso de desarrollo.
- Propone ir construyendo el AD a través de la construcción de los MD departamentales.
- Existe abundante documentación sobre la misma.
- Es una metodología madura y reconocida por el resto de la comunidad dedicada al tema. Tiene bien definidas las etapas, actividades, artefactos y roles.

A partir de lo planteado en la metodología seleccionada como base y teniendo en cuenta las características propias de la UCI y DATEC, se tuvo en cuenta su ciclo de vida en los siguientes flujos de trabajo:

- **Estudio Preliminar o Planeación:** Se llevan a cabo las tareas de planeación del proyecto, se definen los objetivos, el alcance preliminar, los costos estimados y los recursos necesarios.
- **Requerimientos:** Se definen los requerimientos a partir de una comparación de las necesidades y las reglas del negocio con los elementos disponibles en las fuentes.

Capítulo I Fundamentación Teórica

- **Arquitectura y Diseño:** Se definen las estructuras de almacenamiento, se diseñan las reglas de extracción, transformación y carga, así como 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 con las particiones y agregaciones correspondientes según la solución en desarrollo. Se crea el Área Temporal de Almacenamiento, se ejecutan las reglas de ETL. Se configuran e implementan las herramientas de BI para obtener los reportes, gráficos, mapas y otros que cubran los requerimientos firmados con el cliente final.
- **Prueba:** Se realizan varias pruebas, comenzando por las Pruebas de Unidad hasta la Aceptación con el cliente final.
- **Despliegue:** Se realizan dos etapas, la primera es un Despliegue Piloto, donde se configuran los servidores necesarios y se instalan las herramientas según la Arquitectura definida, se cargan una muestra de los datos, con el fin de demostrar que la solución funciona. Una vez aceptada por el cliente, se realiza la carga histórica de los datos.
- **Soporte y Mantenimiento:** Después de la solución implantada se brindan servicios que, pueden ser desde un soporte en línea, vía telefónica, web, correo u otras, hasta el acompañamiento junto al cliente.
- **Gestión y Administración del Proyecto:** Transcurre a lo largo de todo el ciclo de vida, es llevado a cabo por el Grupo de Dirección del Proyecto, es aquí donde se controla, gestiona y chequea todo el desarrollo, los gastos, las utilidades, los recursos, las adquisiciones, y demás actividades relacionadas con la gestión de proyecto.

Los grupos de trabajo que presenta la línea, está estructurada de la siguiente forma:

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

1.3.2 Herramienta de Desarrollo.

➤ **Sistemas gestores de bases de datos**

Hoy en día las organizaciones enfrentan numerosos desafíos de datos, tales como: la necesidad de tomar decisiones más rápidas y orientadas a datos, la necesidad de aumentar la productividad y flexibilidad del personal de desarrollo, la presión ejercida para reducir los presupuestos generales de informática, a la vez que escalan la infraestructura para satisfacer las exigencias cada vez mayores; por lo que se han creado diferentes herramientas que posibilitan la construcción de un almacén de datos, conocidas como Sistemas Gestores de Bases de Datos (SGBD).

➤ **PostgreSQL 8.4.0**

Contiene una gran cantidad de mejoras, por lo que la administración, consulta y programación en PostgreSQL será más fácil y productiva que en versiones anteriores. Es una versión libre, que continúa con el rápido desarrollo de la base de datos de código abierto a escala mundial. Aumenta el rendimiento de las aplicaciones, entre las principales mejoras se encuentran:

- Restauración de base de datos en procesos paralelos, que acelera la recuperación de un respaldo hasta 8 veces.
- Privilegios por columna, que permiten un control más granular de datos confidenciales.
- Configuración de ordenamiento configurable por base de datos, lo cual hace a PostgreSQL más útil en entornos con múltiples idiomas.

Ventajas

- Mejor soporte que los proveedores comerciales.
- Ahorros considerables en costos de operación.
- Estabilidad y confiabilidad.
- Extensible.
- Multiplataforma.
- Diseñado para ambientes de alto volumen.

La herramienta que utiliza PostgreSQL para la administración de bases de datos es PgAdmin III.

1.3.3 Herramienta de Modelado

Las herramientas nombradas CASE (Ingeniería de Software Asistida por Computadora), son útiles para la modelación de proyectos, flujos de información, entidades de datos, así como requerimientos del sistema, facilitando el proceso de planeación. En estas se puede diseñar el modelo lógico de la solución para posteriormente generar el físico, lo que proporcionará ganancias en tiempo de desarrollo ya que permite la generación del script de la base datos de forma automática. La herramienta de modelado usada se describe a continuación:

➤ **Visual Paradigm 6.4**

Ventajas:

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

Esta herramienta posee las siguientes características:

- Disponibilidad en múltiples plataformas.
- Mayor rapidez en la construcción de aplicaciones de calidad.
- Permite dibujar todos los tipos de diagramas de clases.
- Generación de código inverso, código para Java y exportación como HTML.
- Soporta aplicaciones web.
- Fácil de instalar y actualizar.
- Exportación de imágenes jpg, png y svg.

1.3.4 Herramienta de perfilado de datos

El perfilado de datos es un paso vital ya que a la hora de plantear un análisis de los datos de origen, los mismos se encuentran en muchas ocasiones que realmente no se saben qué se va a preguntar, ni donde pueden residir algunos problemas. El mismo analiza los datos de los sistemas para entender su contenido, estructura, calidad y dependencias.

➤ **DataCleaner 1.5.3**

Capítulo I Fundamentación Teórica

DataCleaner es una aplicación de código abierto para el perfilado, la validación y comparación de datos. Estas actividades ayudan a administrar y supervisar la calidad de los datos con el fin de garantizar que la información es útil y aplicable a su situación de negocio.

Es la alternativa gratuita al software de gestión de datos maestros metodológicos, almacenamiento de datos de proyectos de la investigación estadística y la preparación para la **extracción, transformación y carga** (ETL) de datos.

1.3.5 Herramienta de ETL

➤ Pentaho Data Integration 3.0.1

Pentaho Data Integration o Kettle es una herramienta multiplataforma, fácil de instalar y configurar. Usa tecnologías estándar como Java, XML, Java Script. Proporciona la extracción de gran alcance, Transformación y Carga de datos (ETL) utilizando un enfoque innovador, orientado a los metadatos. Presenta una probada arquitectura escalable y basada en estándares. Tiene un integrado único que permite a los desarrolladores y usuarios finales trabajar sin problemas. Reduce drásticamente el tiempo y la dificultad de construir y desplegar aplicaciones de BI.

Pentaho Data Integration incluye:

- Amplia fuente de datos de apoyo incluyendo aplicaciones empaquetadas, más de 30 plataformas de código abierto y propietario de base de datos, archivos planos y documentos de Excel.
- Unificado de ETL, modelado y visualización de entorno de desarrollo para el diseño de aplicaciones de BI.

1.3.6 Herramientas de BI

➤ Pentaho Analysis Services 3.0.4

Pentaho Analysis Services también conocido como Mondrian es un servidor OLAP que permite a los usuarios de negocio analizar grandes cantidades de datos en tiempo real. Es una de las aplicaciones más importantes de la plataforma Pentaho BI, la misma gestiona comunicación entre una aplicación OLAP (escrita en Java) y la base de datos con los datos fuente.

➤ Pentaho Schema Workbench 3.2.0

Capítulo I Fundamentación Teórica

Pentaho Schema Workbench es la herramienta gráfica que permite crear, modificar y publicar un esquema de Mondrian, y además permite publicarlos al servidor BI para que puedan ser utilizados en los análisis por los usuarios de la plataforma. Es un programa java multiplataforma.

Es una herramienta orientada al desarrollador conocedor de la estructura de un esquema de Mondrian. Tiene dos áreas: la zona en la que se muestra la estructura jerárquica del esquema OLAP y la zona de edición de las propiedades de cada elemento.

➤ **Pentaho BI Server 3.6.0**

La aplicación más conocida de la Plataforma de BI es la Pentaho BI Server que funciona como una red basada en sistema de gestión de informe, el servidor de integración de aplicaciones y un motor de flujo de trabajo ligero (secuencias de acción). Está diseñado para integrarse fácilmente en cualquier proceso. Provee la arquitectura y la infraestructura necesarias para crear soluciones completas para problemas de negocios.

➤ **Apache Tomcat 5.5**

Apache Tomcat es desarrollado en un entorno abierto y participativo y publicado bajo la licencia Apache versión 2. Los usuarios disponen de libre acceso a su código fuente y a su forma binaria en los términos establecidos en la Apache Software Licence.

1.4 Conclusiones parciales

En este capítulo se realizó un estudio detallado de los elementos principales de la fundamentación teórica. Después de un análisis pertinente se definió como metodología de desarrollo, la Metodología para el Desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio en DATEC.

Dentro de las diferentes herramientas que existen, el centro definió como gestor de bases de datos PostgreSQL 8.4.0 y como herramienta de modelado Visual Paradigm 6.4. Para el desarrollo del proceso ETL se seleccionó Pentaho Data Integration 3.0.1 y como perfilado de datos DataCleaner 1.5.3. Para el desarrollo de la inteligencia de negocio se definieron Pentaho Analysis Services 3.0.4, Pentaho Schema Workbench 3.2.0, Pentaho BI Server 3.6.0 y Apache Tomcat 5.5.

Capítulo II Análisis y diseño del Mercado de Datos

Capítulo II Análisis y diseño del mercado de datos

Introducción

En este capítulo se abordará todo lo referente al proceso de análisis y diseño del mercado de datos, donde se hará un estudio preliminar del negocio para facilitar una mejor comprensión del desarrollo de la solución. También se definirán los requisitos de información, funcionales, no funcionales y multidimensionales, así como las reglas del negocio. Se realizarán las descripciones de los casos de usos, desarrollo de la matriz BUS, además de identificar los hechos, medidas y dimensiones. Se definirán los reportes candidatos y el diseño del proceso de integración de datos.

2.1 Estudio preliminar del negocio

En el curso 2009-2010 se le asigna la tarea a la UCI de crear un mercado de datos al área específica del Plan Turquino para SIGOB. Debido a ello se crea una primera versión del mercado de datos Plan Turquino. Se contaba con 23 indicadores registrados en el modelo 0024-03 recogido en las publicaciones de la ONE que es la encargada de la estadística general en el país. Esta tarea pasó a ser tema de un trabajo de diploma a la cual se le dio cumplimiento en todo el proceso de análisis y diseño. Por motivos de tiempo, parte de la implementación del indicador servicios no quedó totalmente satisfecha por parte del cliente, pues se tuvo que apresurar la entrega del producto y algunos de los indicadores anteriormente mencionados no se incluyeron. Actualmente se orientó redefinir el análisis y diseño de todo el negocio e incluir en una nueva versión del producto, los procesos de ETL y BI.

2.2 Definición del negocio

La ONE es la institución rectora de la estadística en nuestro país, en ella se recogen informaciones de las diferentes esferas de la sociedad. Estos datos no están integrados de la mejor forma, por lo que se hace necesario crear mercados de datos para lograr una mejor organización del almacén de datos, donde cada mercado constituye un área temática de dicho almacén. Existen diferentes modelos que guardan esta información según su tipo, con una frecuencia (anual, semestral y trimestral). En el modelo 0024-03 se recoge la información de los datos estadísticos referente al Plan Turquino de la siguiente forma:

Capítulo II Análisis y diseño del Mercado de Datos

- Los datos de los Indicadores se recogen cada tres meses por la Oficina Municipal de Estadísticas de cada provincia perteneciente. Aquí se acumulan los datos enviados por los organismos y luego 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 guardan en ficheros de extensión dbf y se envían a la ONE.
- Cuando llega este a su destino, los especialistas realizan una limpieza de los datos y se crea un archivo general con la información.
- Finalmente la información se encuentra disponible en las diferentes publicaciones con que cuenta la ONE. En el caso del Modelo 0024-03 se encuentra en la publicación trimestral y posee dos variantes: estatal y no estatal.

2.3 Roles y permisos

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

- ✓ **Analista:** Encargado de realizar los reportes y consultas sobre la información.
- ✓ **Administrador del Sistema:** Encargado de gestionar los permisos, usuarios y roles, además de los reportes OLAP.
- ✓ **Administrador ETL:** Encargado de realizar el proceso de extracción, transformación y carga de los datos en el mercado de datos.

| Roles | Permisos | |
|---------------------------|----------|-----------|
| | Lectura | Escritura |
| Analista | X | |
| Administrador del Sistema | X | X |
| Administrador ETL | X | X |

Capítulo II Análisis y diseño del Mercado de Datos

2.4 Reglas de Negocio

Las Reglas del Negocio se basan en la extracción de información originada de las políticas, reglas y regulaciones del negocio y de la descripción del flujo.

- ✓ **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%.

$$\text{Crecimiento} = (\text{real_año_actual} / \text{real_año_anterior}) * 100 (1)$$

- ✓ **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%.

$$\text{Cumplimiento} = (\text{real_año_actual} / \text{plan_año_actual}) * 100 (1)$$

Otras Reglas

R1. La siembra de alevines en la acuicultura tiene que ser mayor o igual que la siembra de alevines del MINAG.

R2. En la acuicultura, espejos de agua tiene que ser mayor o igual que espejos de agua del MINAG.

R3. La producción de pescado en la acuicultura tiene que ser mayor o igual que la producción de pescado del MINAG.

R4. En el turismo, las habitaciones de los días existentes tienen que ser mayor o igual que las habitaciones de los días ocupadas (extranjeros).

R5. La suma de las viviendas del sector estatal y cooperativo, y las viviendas de población tienen que ser igual a las viviendas terminadas en total.

R6. La suma de las viviendas del sector estatal y cooperativo del total del MINAG, MICONS, CAP y otros, tienen que ser igual a las viviendas del sector estatal y cooperativo.

R7. La suma de conservación y rehabilitación del sector estatal y cooperativo de las viviendas, y de la población, tiene que ser igual a la conservación y rehabilitación total.

R8. La suma del comercio interior, los servicios a la población tradicional y paralelo, tiene que ser igual al servicio a la población total.

R9. La suma del comercio interior, el comercio minorista de alimentos, industriales y otras ventas, tiene que ser igual al total de las ventas del comercio minorista.

Capítulo II Análisis y diseño del Mercado de Datos

R10. La suma de las ventas en el comercio interior de la gastronomía comestibles y otras ventas, tiene que ser igual que las ventas totales de la gastronomía.

R11. La suma del mantenimiento y conservación de viales en el MINAG, MITRANS, CAP y otros, tiene que ser mayor o igual que el mantenimiento y conservación de viales.

R12. El mantenimiento y conservación de viales, el funcionamiento de la divisa, tiene que ser mayor o igual que el funcionamiento de divisas del combustible. R13. En la construcción de viales, el funcionamiento en divisas, tiene que ser mayor o igual que el funcionamiento en divisas del combustible.

R14. Los servicios prestados en comunicaciones, el funcionamiento de divisas, tiene que ser mayor o igual que el funcionamiento de divisas en las inversiones.

R15. La suma de las inversiones del medio ambiente de la cuenca hidrográfica, suelos, atmósfera, silvicultura, eliminación de fuentes contaminantes y otros, tiene que ser igual a las inversiones del medio ambiente del MINAG.

R16. La suma de las inversiones del medio ambiente de la cuenca hidrográfica, suelos, atmósfera, silvicultura, eliminación de fuentes contaminantes y otros, tiene que ser igual a las inversiones del medio ambiente del financiamiento en divisas del MINAG.

R17. La suma de las inversiones del medio ambiente de la cuenca hidrográfica, suelos, atmósfera, silvicultura, eliminación de fuentes contaminantes y otros, tiene que ser igual a las inversiones del medio ambiente de otros OACE.

R18. La suma de las inversiones del medio ambiente de la cuenca hidrográfica, suelos, atmósfera, silvicultura, eliminación de fuentes contaminantes y otros, tiene que ser igual a las inversiones del medio ambiente del financiamiento en divisas de otros OACE.

R19. La suma de los expedientes aprobados en el perfeccionamiento empresarial, en el sector agroforestal, industrial, comercio, servicios y otros sectores, tiene que ser igual a los expedientes aprobados del perfeccionamiento empresarial.

Además de los atributos calculables, en la ONE 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). (1)

Capítulo II Análisis y diseño del Mercado de Datos

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. (1)

2.5 Necesidades de información

Después de un análisis detallado del negocio, se definen cómo necesidades de información, a los indicadores del Plan Turquino, conteniendo las medidas que se aglomeran en el Modelo 0024-03: la existencia al final del período, el real del año anterior y el real y plan del año actual. Por la gran cantidad de indicadores es necesario la agrupación de los mismos en 3 grupos específicos: agricultura-ganadería-silvicultura, industria y servicios. Debido a las necesidades del cliente estos parámetros se medirán de acuerdo al tiempo, a la variante del sector (estatal o no estatal) y a la provincia. De esta forma se describen a continuación los requisitos de información, multidimensionales, funcionales y no funcionales del sistema.

2.5.1 Requisitos de información

Posteriormente al análisis realizado de los problemas existentes en la ONE y a las entrevistas llevadas a cabo con el cliente, se definieron los requisitos de información de la investigación. 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, ganadería y silvicultura por provincia.
- RI 2. Obtener existencia al final del período de los productos de la agricultura, ganadería y silvicultura por el tiempo.
- RI 3. Obtener existencia al final del período de los productos de la agricultura, ganadería y silvicultura por variante de sector.
- RI 4. Obtener real del año anterior de los productos de la agricultura, ganadería y silvicultura por provincia.
- RI 5. Obtener real del año anterior de los productos de la agricultura, ganadería y silvicultura por el tiempo.
- RI 6. Obtener real del año anterior de los productos de la agricultura, ganadería y silvicultura por variante de sector.

Capítulo II Análisis y diseño del Mercado de Datos

- RI 7. Obtener plan del año actual de los productos de la agricultura, ganadería y silvicultura por provincia.
- RI 8. Obtener plan del año actual de los productos de la agricultura, ganadería y silvicultura por el tiempo.
- RI 9. Obtener plan del año actual de los productos de la agricultura, ganadería y silvicultura por variante de sector.
- RI 10. Obtener real del año actual de los productos de la agricultura, ganadería y silvicultura por provincia.
- RI 11. Obtener real del año actual de los productos de la agricultura, ganadería y silvicultura por el tiempo.
- RI 12. Obtener real del año actual de los productos de la agricultura, ganadería y silvicultura 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.

2.5.2 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. (1)

Capítulo II Análisis y diseño del Mercado de Datos

VE:

- ✓ dpa
- ✓ variante_sector
- ✓ temporal_trimestre
- ✓ productos_agricultura_ganadería_silvicultura
- ✓ productos_industria
- ✓ servicios

VS:

- ✓ existencia_final_periodo
- ✓ real_año_anterior
- ✓ real_año_actual
- ✓ plan_año_actual

2.5.3 Requisitos funcionales

Los requisitos funcionales no es más, que la funcionalidad que debe cumplir el sistema, de acuerdo con las necesidades y especificaciones del cliente.

RF 1 Insertar usuario

RF 2 Eliminar usuario

RF 3 Insertar rol

RF 4 Eliminar rol

RF 5 Insertar reporte OLAP

RF 6 Modificar reporte OLAP

RF 7 Eliminar reporte OLAP

RF 8 Autenticar usuario

RF 9 Extraer los datos

RF 10 Realizar Transformación y Carga de los datos

RF 11 Mostrar editor MDX

RF 12 Configurar tabla OLAP

RF 13 Mostrar propiedades

RF 14 Suprimir filas / columnas vacías

Capítulo II Análisis y diseño del Mercado de Datos

RF 15 Mostrar gráfico

RF 16 Configurar impresión

RF 17 Exportar a PDF y Excel

2.5.4 Requisitos no funcionales

Usabilidad

Requisito 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.
- Dominar elementos básicos sobre base de datos.
- Los usuarios deben recibir capacitación de las herramientas utilizadas.

Fiabilidad

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

Requisito 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**

Capítulo II Análisis y diseño del Mercado de Datos

- El tiempo medio de reparación estimado es aproximadamente de 1 día luego del fallo.
- **Máximo de errores**
 - Por ser el proceso de integración un punto crítico para la calidad y exactitud de los datos almacenados se establece un máximo de errores de 0 errores/puntos de función.
- **Exactitud**
 - La salida de los datos debe ser 100% exacta, ya que de ahí dependen las decisiones que se tomen basándose en estos datos.

Eficiencia

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

Requisito de eficiencia

- **Tiempo de respuesta**
 - El sistema debe tener un tiempo de respuesta aproximado de 1 minuto, no debe excederse de 2 minutos.
- **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

Requisito de diseño

- El Sistema Gestor de Base de Datos que se utilizará será PostgreSQL versión 8.4.

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

- Se elaborará un manual de usuario que guiará la ejecución del usuario teniendo en cuenta cada funcionalidad.

Componentes comprados

- La herramienta CASE Visual Paradigm no es un software libre pero se cuenta con una licencia adquirida previamente por la UCI.

Interfaz

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

Capítulo II Análisis y diseño del Mercado de Datos

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 de 512 MB como mínimo.
- Para el proceso de visualización e inteligencia de negocio se necesita una memoria 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 para el almacenamiento de la información.
- Se debe garantizar al menos una impresora, para imprimir los reportes de salida.

Interfaces de 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 al sistema operativo, para garantizar la visualización de las interfaces web de los reportes.
- La base de datos se acoplará 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

- De las herramientas utilizadas, Visual Paradigm 6.4 es la única herramienta que no es libre, pero se cuenta con su licencia previamente adquirida por la universidad.

2.6 Descripción de Casos de Uso

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

Capítulo II Análisis y diseño del Mercado de Datos

2.6.1 CU Extraer los datos.

| | |
|---|--|
| Caso de Uso: | Extraer los datos |
| 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. Finaliza cuando se almacenan los datos en un área temporal. |
| Precondiciones: | Disponibilidad de las fuentes. |
| Referencias | RF 9 |
| Prioridad | Crítico |
| Flujo Normal de Eventos | |
| Acción del Actor | Respuesta del Sistema |
| 1. Interactúa con la herramienta PDI. | 2. Muestra el área de trabajo. |
| 3. Selecciona el componente de entrada de datos XBase. | 4. Muestra una ventana para introducir la dirección del fichero fuente. |
| 5. Especifica la dirección del fichero fuente. | |
| 6. Selecciona la opción pre visualizar. | 7. Muestra los datos cargados del fichero fuente. |
| 8. Presiona el botón aceptar y finaliza el caso de uso. | |
| Prototipo de Interfaz | |
| Flujos Alternos | |
| Acción del Actor | Respuesta del Sistema |
| | 1.1 No responde a la solicitud de conexión. |
| 2. Notifica el error al administrador. | |
| 3. Si hay control de cambios, verifica si hay modificaciones. En caso de no haber, va al paso 6 flujo normal. En caso de que haya, va al paso 5 del flujo normal. | |
| Prototipo de Interfaz | |
| Poscondiciones | Datos disponibles para transformar. |

Capítulo II Análisis y diseño del Mercado de Datos

2.6.2 CU Realizar Transformación y Carga de los datos.

| | |
|------------------------|--|
| Caso de Uso: | Realizar Transformación y Carga de los datos. |
| 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. Cargar los datos en el mercado y finaliza el CU. |
| Precondiciones: | Extracción completada en área temporal. La estructura del almacén debe estar disponible. |
| Referencias | RF 10 |
| Prioridad | Crítico |

Flujo Normal de Eventos

| Acción del Actor | Respuesta del sistema |
|--|--|
| 1. Selecciona estructuras a transformar. | |
| 2. Carga datos seleccionados en memoria. | |
| 3. Aplica transformaciones pertinentes y genera datos. | |
| 4. Selecciona "Ejecutar transformación". | 5. Ejecuta la transformación. |
| | 6. Almacena los datos en la base de datos. |

Prototipo de Interfaz

| | |
|-----------------------|---------------------------------|
| Poscondiciones | Datos transformados y cargados. |
|-----------------------|---------------------------------|

2.6.3 CU Autenticar Usuario

| | |
|---------------------|--|
| Caso de Uso: | Autenticar Usuario. |
| Tipo: | Funcional |
| Actores: | Analista |
| Resumen: | El caso de uso se inicia cuando el Analista se autentica. Finaliza cuando el analista se autentica y tiene permiso para entrar al sistema. |

Capítulo II Análisis y diseño del Mercado de Datos

| | |
|-------------------------------------|--|
| Precondiciones: | Sistema on-line |
| Referencia | RF 8 |
| Prioridad: | Crítico |
| Flujo Normal de Eventos | |
| 1. El analista accede al sistema. | 2. El sistema muestra la vista para introducir los datos. |
| 3. El Analista introduce los datos. | 4. El sistema valida los datos. |
| | 5. Si los datos son válidos, el sistema da los permisos para entrar a la aplicación. |
| Flujo Alterno | |
| Acción del Actor | Respuesta del Sistema |
| | 4.1 El sistema muestra un mensaje de error y regresa al punto 2 del flujo normal de eventos. |
| Poscondiciones | Usuario autenticado listo para operar en el sistema. |

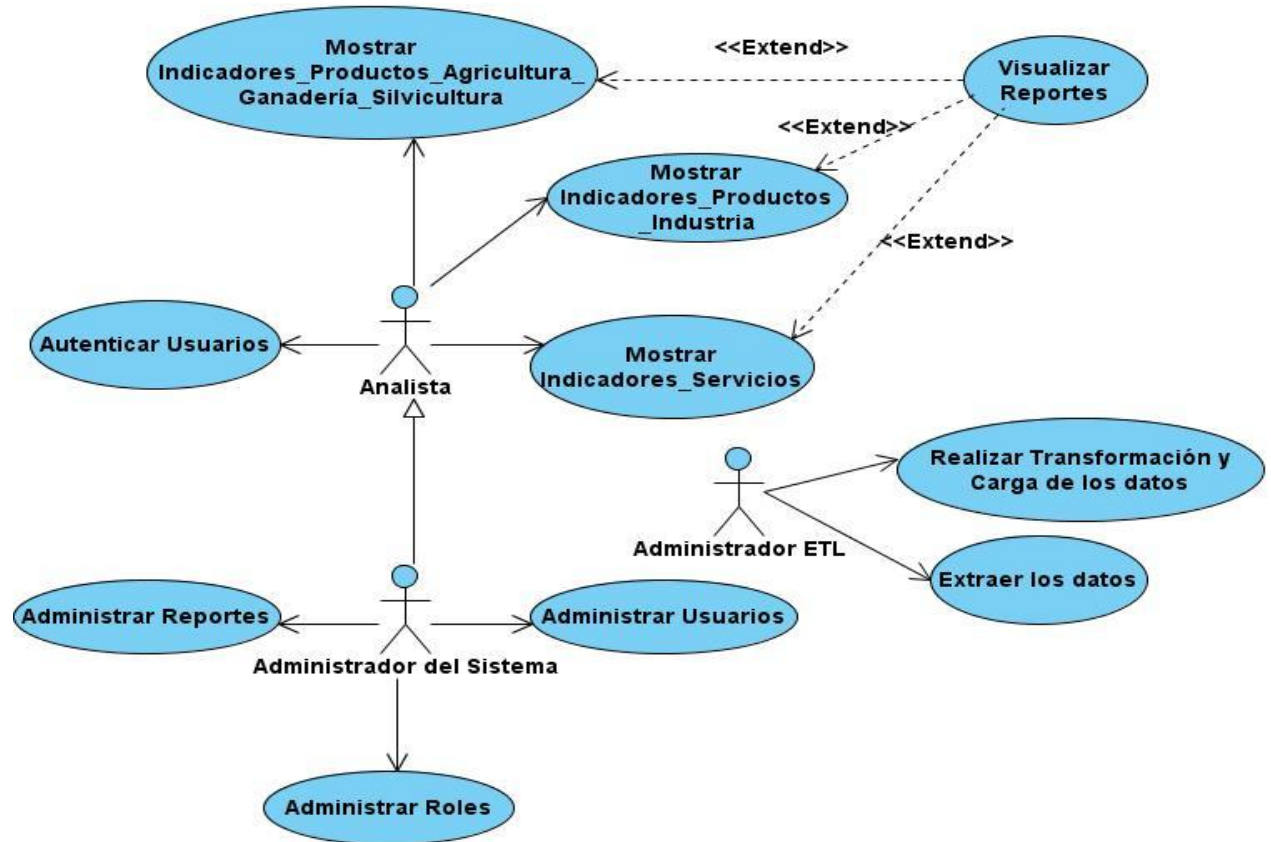


FIG. 2 DIAGRAMA CU DEL SISTEMA

2.7 Desarrollo de la 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_trimestre, variante_sector y dpa están relacionadas con los 3 hechos existentes. En el caso de las dimensiones indic_prod_agric_ganad_silvicult, indic_prod_industria e indic_servicios solamente se relacionan cada uno de los hechos de igual nombre respectivamente.

Dimensiones: (1)

D1: dim_dpa

D2: dim_temporal_trimestre

Capítulo II Análisis y diseño del Mercado de Datos

D3: dim_variante_sector

D4: dim_indic_prod_agric_ganad_silvicult

D5: dim_indic_prod_industria

D6: dim_indic_servicios

| Hechos/dimensiones | D1 | D2 | D3 | D4 | D5 | D6 |
|---------------------------------------|----|----|----|----|----|----|
| hech_indic_prod_agric_ganad_silvicult | X | X | X | X | | |
| hech_indic_prod_industria | X | X | X | | X | |
| hech_indic_servicios | X | X | X | | | X |

2.8 Identificación de dimensiones, hechos y medidas

2.8.1 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_dpa

Esta dimensión representa la situación geográfica del país con respecto a la división política administrativa.

Provincia -> Municipios

| dim_dpa | | |
|---------------------------|-------------|-------------------------|
| +dim_dpa_id | int4 | Nullable = false |
| provincia_codigo | char(2) | Nullable = false |
| municipio_codigo | varchar(4) | Nullable = false |
| provincia_nombre | varchar(25) | Nullable = false |
| municipio_nombre | varchar(30) | Nullable = false |
| provincia_descripcion | varchar(45) | Nullable = false |
| municipio_descripcion | varchar(45) | Nullable = false |
| municipio_extension | float4 | Nullable = false |
| municipio_ext_cayosady | float4 | Nullable = false |
| municipio_ext_tierrafirme | float4 | Nullable = false |
| dpa_anno_nombre | varchar(4) | Nullable = false |

Capítulo II Análisis y diseño del Mercado de Datos

✓ dim_temporal_trimestre

Esta dimensión representa a los valores anuales y trimestrales que posee el modelo. Presenta un nivel denominado “año” y dentro la categoría “trimestre”.

Anno -> Trimestre

| dim_temporal_trimestre | | |
|-----------------------------------|-------------|-------------------------|
| +dim_temporal_trimestre_id | int4 | Nullable = false |
| trimestre_codigo | varchar(6) | Nullable = false |
| año_codigo | varchar(4) | Nullable = false |
| semestre_codigo | varchar(6) | Nullable = false |
| trimestre_numero | int4 | Nullable = false |
| semestre_numero | int4 | Nullable = false |
| año_numero | int4 | Nullable = false |
| trimestre_nombre | varchar(20) | Nullable = false |
| semestre_nombre | varchar(20) | Nullable = false |
| año_nombre | char(4) | Nullable = false |

✓ dim_variante_sector

Esta dimensión representa la clasificación de la información teniendo en cuenta la variante del sector, estatal o no estatal, del modelo.

| dim_variante_sector | | |
|--------------------------------|-------------|-------------------------|
| +dim_variante_sector_id | int4 | Nullable = false |
| variante_sector_codigo | int4 | Nullable = false |
| variante_sector_nombre | varchar(35) | Nullable = false |
| variante_sector_descripcion | varchar(45) | Nullable = false |

✓ dim_indic_prod_agric_ganad_silvicult

Esta dimensión representa la información de los indicadores que se recogen para los productos de la agricultura, ganadería, silvicultura, acuicultura, apicultura y avicultura.

Capítulo II Análisis y diseño del Mercado de Datos

| dim_indic_prod_agric_ganad_silvicult | | |
|---|--------------|-------------------------|
| +dim_indic_prod_agric_ganad_silvicult_id | int4 | Nullable = false |
| indicador_codigo | varchar(11) | Nullable = false |
| indicador_nombre | varchar(45) | Nullable = false |
| tematica_nombre | varchar(35) | Nullable = false |
| indicador_descripcion | varchar(200) | Nullable = false |
| indicador_um | varchar(15) | Nullable = false |
| subtematica_nombre | varchar(45) | Nullable = false |
| indicador_agregable | varchar(2) | Nullable = false |
| seguridad_codigo | varchar(1) | Nullable = false |
| seguridad_nombre | varchar(10) | Nullable = false |
| tematica_descripcion | varchar(50) | Nullable = false |
| subtematica_descripcion | varchar(50) | Nullable = false |

✓ dim_indic_prod_industria

Esta dimensión representa la información de los indicadores que se recogen para los productos de la industria.

| dim_indic_prod_industria | | |
|-------------------------------------|--------------|-------------------------|
| +dim_indic_prod_industria_id | int4 | Nullable = false |
| indicador_codigo | varchar(11) | Nullable = false |
| indicador_nombre | varchar(45) | Nullable = false |
| tematica_nombre | varchar(35) | Nullable = false |
| indicador_descripcion | varchar(200) | Nullable = false |
| indicador_um | varchar(15) | Nullable = false |
| subtematica_nombre | varchar(45) | Nullable = false |
| indicador_agregable | varchar(2) | Nullable = false |
| seguridad_codigo | varchar(1) | Nullable = false |
| seguridad_nombre | varchar(10) | Nullable = false |
| tematica_descripcion | varchar(50) | Nullable = false |
| subtematica_descripcion | varchar(50) | Nullable = false |

✓ dim_indic_servicios

Esta dimensión representa la información de los indicadores que se recogen para los servicios, por ejemplo, el programa de empleo, cultura, deporte, salud, turismo, entre otros.

Capítulo II Análisis y diseño del Mercado de Datos

| dim_indic_servicios | | |
|-------------------------|--------------|-------------------------|
| +dim_serv_id | int4 | Nullable = false |
| indicador_codigo | varchar(11) | Nullable = false |
| indicador_nombre | varchar(45) | Nullable = false |
| tematica_nombre | varchar(35) | Nullable = false |
| indicador_descripcion | varchar(200) | Nullable = false |
| indicador_um | varchar(15) | Nullable = false |
| subtematica_nombre | varchar(45) | Nullable = false |
| indicador_agregable | varchar(2) | Nullable = false |
| seguridad_codigo | varchar(1) | Nullable = false |
| seguridad_nombre | varchar(10) | Nullable = false |
| tematica_descripcion | varchar(50) | Nullable = false |
| subtematica_descripcion | varchar(50) | Nullable = false |

2.8.2 Hechos identificados

El modelo tiene tres hechos, debido a que se agrupan los indicadores en tres categorías, y estos a su vez se relacionan con las dimensiones DPA, temporal_trimestre y variante_sector.

✓ hech_indic_prod_agric_ganad_silvicult

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, ganadería y silvicultura.

| hech_indic_prod_agric_ganad_silvicult | | |
|---|-------------|-------------------------|
| +prod_agric_ganad_sivicult_id | int4 | Nullable = false |
| existencia_final_periodo | float4 | Nullable = true |
| real_anno_anterior | float4 | Nullable = true |
| real_anno_actual | float4 | Nullable = true |
| plan_anno_actual | float4 | Nullable = true |
| <i>#dim_variante_sector</i> <i>dim_variante_sector_id</i> | <i>int4</i> | <i>Nullable = false</i> |
| <i>#dim_temporal_trimestre</i> <i>dim_temporal_trimestre_id</i> | <i>int4</i> | <i>Nullable = false</i> |
| <i>#dim_dp</i> <i>dim_dpa_id</i> | <i>int4</i> | <i>Nullable = false</i> |
| <i>#dim_indic_prod_agric_ganad_silvicult</i> <i>dim_indic_prod_agric_ganad_silvicult_id</i> | <i>int4</i> | <i>Nullable = false</i> |

✓ hech_indic_prod_industria

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

Capítulo II Análisis y diseño del Mercado de Datos

| hech_indic_prod_industria | | |
|--|-------------|-------------------------|
| +dim_indic_prod_industria_id | int4 | Nullable = false |
| plan_anno_actual | float4 | Nullable = true |
| real_anno_actual | float4 | Nullable = true |
| real_anno_anterior | float4 | Nullable = true |
| #dim_variante_sectordim_variante_sector_id | int4 | Nullable = false |
| #dim_temporal_trimestredim_temporal_trimestre_id | int4 | Nullable = false |
| #dim_dpاديم_dpa_id | int4 | Nullable = false |
| #dim_indic_prod_industriadim_indic_prod_industria_id | int4 | Nullable = false |

✓ hech_indic_servicios

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.

| hech_indic_servicios | | |
|--|-------------|-------------------------|
| +dim_serv_id | int4 | Nullable = false |
| real_anno_anterior | float4 | Nullable = true |
| real_anno_actual | float4 | Nullable = true |
| plan_anno_actual | float4 | Nullable = true |
| #dim_variante_sectordim_variante_sector_id | int4 | Nullable = false |
| #dim_temporal_trimestredim_temporal_trimestre_id | int4 | Nullable = false |
| #dim_dpاديم_dpa_id | int4 | Nullable = false |
| #dim_indic_serviciosdim_serv_id | int4 | Nullable = false |

2.8.3 Medidas

Hecho indic_produc_ganad_agric_silvicult (1)

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

Hecho indic_produc_industria (1)

Capítulo II Análisis y diseño del Mercado de Datos

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

Hecho `indic_servicios` (1)

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

2.9 Modelo de datos

El modelo de datos contiene 6 dimensiones y 3 hechos, los cuales se relacionan con las dimensiones de igual nombre y, además, con las dimensiones `dpa`, `temporal_trimestre` y `variante_sector` como se muestra en la siguiente figura.

Capítulo II Análisis y diseño del Mercado de Datos

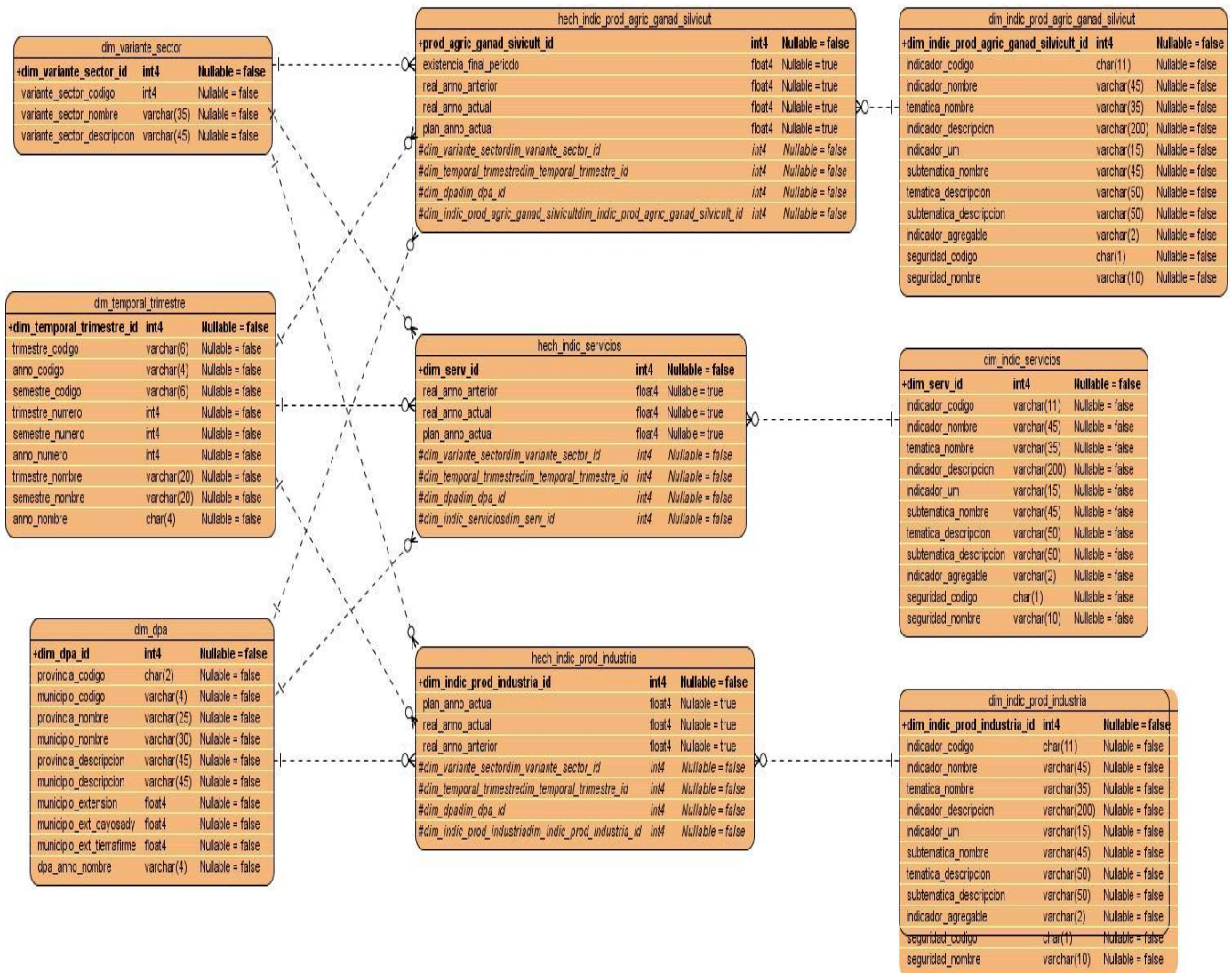


FIG. 3 DISEÑO DEL MODELO DE DATOS

2.10 Conclusiones parciales

En este capítulo se llevó a cabo todo el proceso de redefinición del análisis y diseño del negocio. En las entrevistas realizadas con el cliente se definieron los requisitos de información, funcionales y no funcionales. Se establecieron las principales reglas del negocio requeridas. También se confeccionó la matriz bus con el objetivo de poder observar en esta las relaciones que existen entre los hechos y las dimensiones. Se identificaron las dimensiones, hechos y medidas para el diseño del modelo de datos.

Capítulo III Implementación del Mercado de Datos

Capítulo III Implementación del Mercado de Datos.

Introducción

En este capítulo se abordará todo lo referente al proceso de integración de datos que contiene tres tipos de actividades de carácter general como son: extracción, transformación y carga en el mercado de datos. Además de implementar el subsistema de visualización, modelo de datos, cubos OLAP y reportes candidatos.

3.1 Modelo de datos físico

El modelo de datos es un lenguaje utilizado para la descripción de una base de datos, es una representación física que se hace con el propósito de estudiar detalladamente el comportamiento de la estructura de los datos y de cómo se relacionan entre ellos. El mismo presenta tres sublenguajes:

- Lenguaje de Definición de Datos o DDL (Data Definition Language), cuya función es describir, de una forma abstracta, las estructuras de datos y las restricciones de integridad. (10)
- Lenguaje de Manipulación de Datos o DML (Data Manipulation Language), que se orienta a describir las operaciones de manipulación de los datos. (10)
- Lenguaje de Control de Datos o DCL (Data Control Language), lenguaje que proporcionan los SGBD para controlar los usuarios de la base de datos. (10)

3.1.1 Esquemas y tablas

El esquema permite analizar y organizar la información de los datos contenidos en una BD. Dentro de ellos pueden estar implícitos tablas, funciones, operadores y tipos de datos. En dependencia de los permisos asignados el usuario podrá tener acceso a ellos. El mercado de datos Plan Turquino está compuesto por 2 esquemas:

- esquema dimensiones: contiene todas las tablas de las dimensiones del mercado de datos que se encuentran definidas en la BD de la ONE.
- esquema mart_turquino: contiene todas las tablas de hechos y las dimensiones propias del mercado de datos Plan Turquino.

Capítulo III Implementación del Mercado de Datos

La estructura física de la BD contiene 9 tablas en total: 6 tablas de dimensiones y 3 tablas de hechos quedando de la siguiente forma:

| Esquemas | Tablas |
|---------------|---------------------------------------|
| dimensiones | dim_dpa |
| | dim_temporal_trimestre |
| mart_turquino | dim_variante_sector |
| | dim_indic_prod_agric_ganad_silvicult |
| | dim_indic_prod_industria |
| | dim_indic_servicios |
| | hech_indic_prod_agric_ganad_silvicult |
| | hech_indic_prod_industria |
| | hech_indic_servicios |

La siguiente figura muestra como quedaría conformada la base de datos.

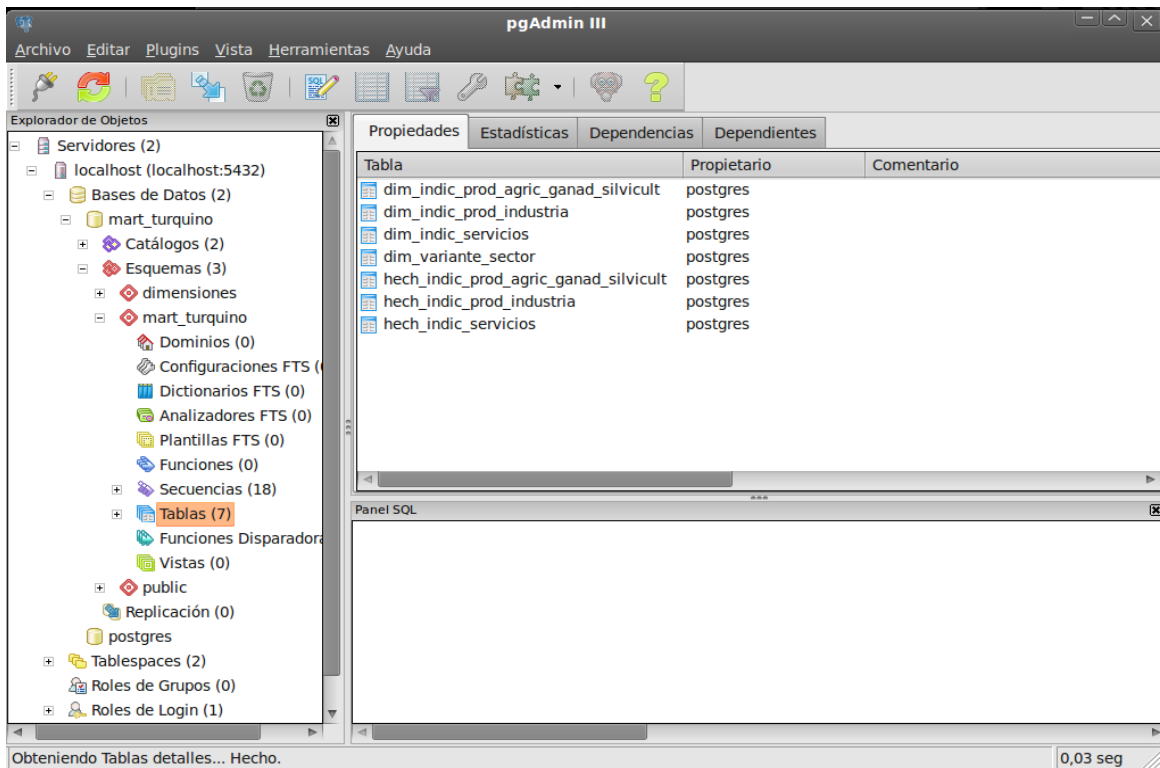


FIG. 4 ESTRUCTURA FÍSICA DE LA BD

3.2 Implementación del subsistema de integración

3.2.1 Arquitectura

La arquitectura es un conjunto de componentes relacionados y constituye el principal elemento sobre el cual se sustenta el desarrollo del software.

En el proceso de integración de los datos, la arquitectura sería la siguiente:

- Fuente de Datos: Es el archivo de extensión DBF del Modelo 0024-03.
- Área temporal: Punto intermedio entre la fuente y el mercado, es donde se realiza la integración y transformación de los datos.
- Mercado de Datos: Destino donde son cargados los datos para su futuro análisis.

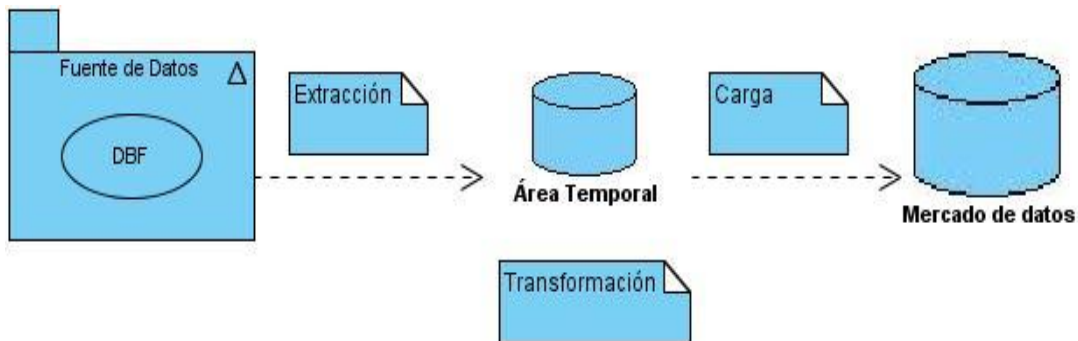


FIG. 5 ARQUITECTURA DE LA INTEGRACIÓN DE DATOS

3.2.2 Perfilado de datos

El perfilado de datos es una de las primeras tareas que se suelen abordar en procesos de Calidad de Datos, y consiste en realizar un primer análisis sobre los datos de origen, normalmente sobre tablas, con el objetivo de empezar a conocer su estructura, formato y nivel de calidad. (11)

En dependencia de los resultados arrojados se establecen las reglas para posteriormente realizar la integración de los datos. La herramienta utilizada fue Data Cleaner en su versión 1.5.4.

- ✓ Análisis de todas las tuplas de la tabla.

Capítulo III Implementación del Mercado de Datos

Standard measures

| | v01 | dpa | fil | c01 | c02 | c03 | c04 |
|---------------|------|------|----------|--------|-----------|----------|----------|
| Row count | 4714 | 4714 | 4714 | 4714 | 4714 | 4714 | 4714 |
| Null values | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Empty values | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Highest value | 2 | 1410 | 00001810 | 8435.0 | 1106097.2 | 309520.0 | 308336.0 |
| Lowest value | 1 | 0102 | 00000010 | 0.0 | 0.0 | 0.0 | 0.0 |

- ✓ Análisis de los valores numéricos.

Number analysis

| | c03 | c04 | c01 | c02 |
|--------------------|-------------|-------------|-----------|-------------|
| Highest value | 309520 | 308336 | 8435 | 1106097,2 |
| Lowest value | 0 | 0 | 0 | 0 |
| Sum | 5722861,1 | 6058806,2 | 296560 | 7213732,3 |
| Mean | 1214,01 | 1285,28 | 62,91 | 1530,28 |
| Geometric mean | 0 | 0 | 0 | 0 |
| Standard deviation | 8135,8 | 8361,02 | 379,91 | 17924,69 |
| Variance | 66191247,33 | 69906641,62 | 144332,06 | 321294638,1 |

- ✓ Análisis de los valores de tipo string.

String analysis

| | dpa | v01 | fil |
|------------------|-------|------|-------|
| Char count | 18856 | 4714 | 37712 |
| Max chars | 4 | 1 | 8 |
| Min chars | 4 | 1 | 8 |
| Avg chars | 4 | 1 | 8 |
| Max white spaces | 0 | 0 | 0 |
| Min white spaces | 0 | 0 | 0 |
| Avg white spaces | 0 | 0 | 0 |
| Uppercase chars | 0% | 0% | 0% |
| Lowercase chars | 0% | 0% | 0% |
| Non-letter chars | 100% | 100% | 100% |
| Word count | 4714 | 4714 | 4714 |
| Max words | 1 | 1 | 1 |
| Min words | 1 | 1 | 1 |

Capítulo III Implementación del Mercado de Datos

3.2.3 Extracción de los datos

En el proceso de la extracción de los datos es donde se trasladan los datos de los archivos DBF para el área temporal, los cuales contienen todo lo referente al mercado de datos Plan Turquino del Modelo 0024 que serán almacenados en las 3 tablas de hechos.

3.2.4 Limpieza y Transformación de los datos

La limpieza está dentro del proceso de transformación de los datos de una BD, constituye un paso de gran importancia y se realiza con el objetivo de corregir o eliminar datos erróneos. Permite identificar datos incompletos, incorrectos e inexactos y luego substituir, modificar o eliminar estos datos sucios. De los errores más comunes se encuentran los valores nulos. El resultado de este transcurso es la carga al mercado de datos.

3.2.5 Carga de los datos

La carga de los datos constituye el paso final de este proceso, estos son cargados al mercado para ser utilizados posteriormente con éxito.

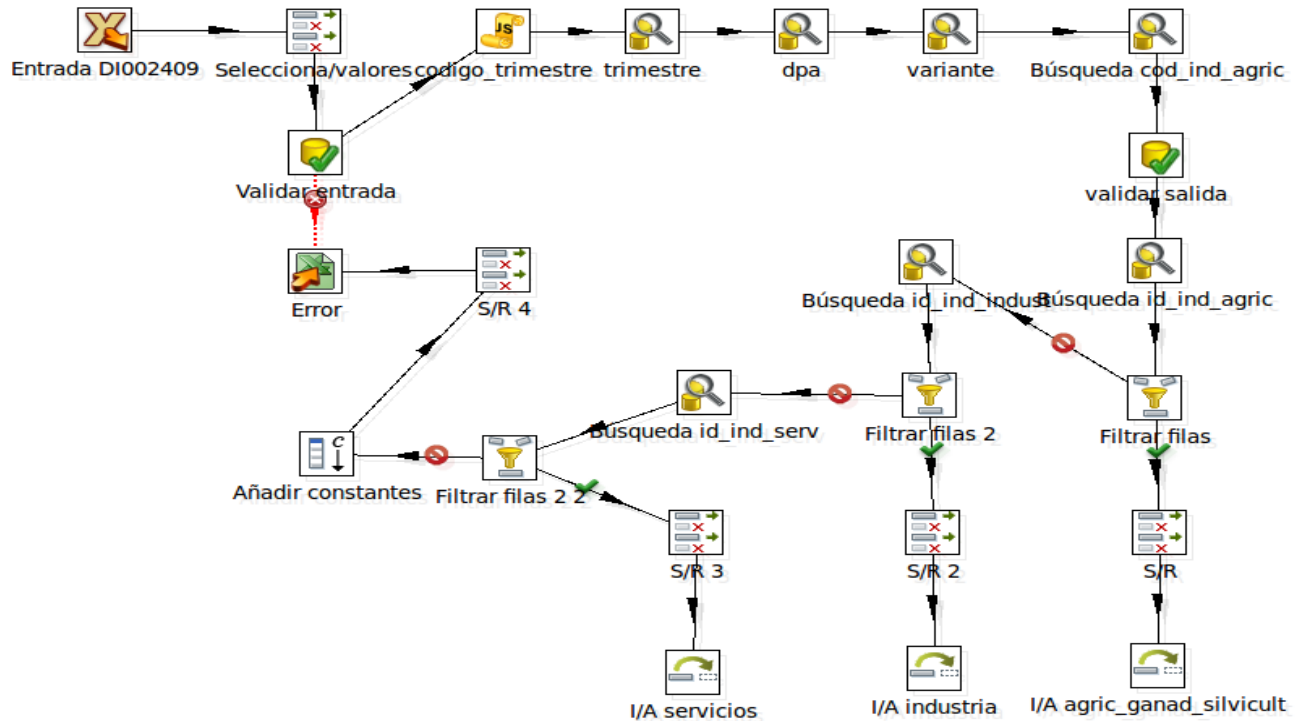


FIG. 6 EJEMPLO DE TRANSFORMACIÓN

La figura 6 muestra un ejemplo de una de las transformaciones, la cual corresponde a la carga de los tres hechos que contiene el modelo de datos. En la misma se abarcan todos los procesos de la integración de datos: la extracción, la limpieza y carga de los mismos. La entrada de los datos corresponde a la extracción. Luego se seleccionan del DBF los datos que se van a utilizar. Después se validan que existan los datos en la fuente, es decir que no haya ningún campo nulo, en caso de encontrar alguno es enviado a un archivo Excel para luego ser tratado. Posteriormente se obtiene el id de cada dimensión, se valida que estos existan y de no ser así en algún caso es enviado al archivo Excel. Finalmente se procede a la carga de las tablas de los hechos.

3.2.6 Implementación del Job

El job (trabajo) es un proceso que agrupa transformaciones a un alto nivel, con una secuencia lógica donde primero son cargadas las dimensiones y luego los hechos.



FIG. 7 JOB

3.3 Implementación del subsistema de visualización

3.3.1 Implementación de los cubos OLAP

Es precisa la creación de los cubos multidimensionales para la implementación del subsistema de visualización, porque es en ellos donde se definen las dimensiones, los niveles de jerarquía de las dimensiones y las medidas. Este proceso se realiza a través de la herramienta Pentaho Schema Workbench, el cual genera un fichero XML, que contendrá todos los cubos OLAP.

En el diseño del esquema Turquino se modelaron 3 cubos OLAP: `hech_indic_prod_industria`, `hech_indic_servicios` y `hech_indic_prod_agric_ganad_silvicult`, donde cada uno está formado por 6 dimensiones, 3 de estas son comunes para todos, además de que contienen 6 medidas: `real_anno_actual`, `real_anno_anterior`, `plan_anno_actual`, `existencia_final_periodo` que solo le corresponde al cubo `hech_indic_prod_agric_ganad_silvicult`, `crecimiento` y `cumplimiento`, estas dos últimas son medidas calculables. La siguiente imagen muestra el diseño de los cubos.

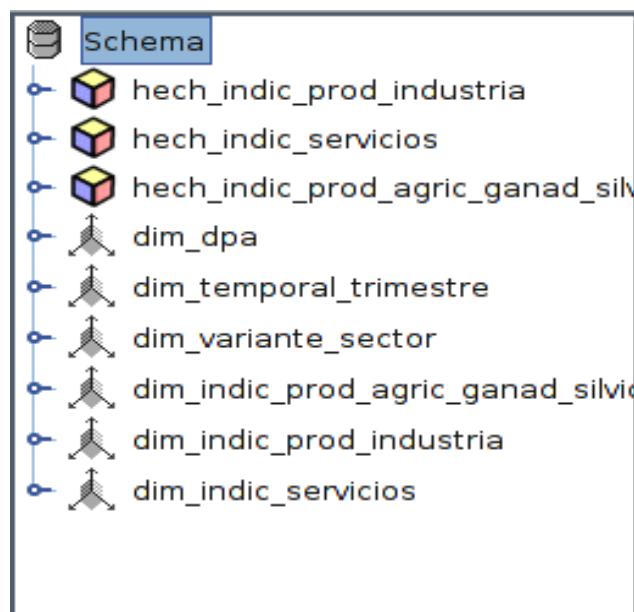


FIG. 8 DISEÑO DE LOS CUBOS

3.3.2 Navegación de la capa de visualización

El mapa de navegación constituye la estructura gráfica en que está organizada la información. El mercado de datos Plan Turquino está compuesto por 1 Área de Análisis General (A.A.G), 1 Área de Análisis (A.A) y 3 Libros de Trabajo (L.T). A continuación se detalla la representación en la capa de visualización.

Descripción del Área de Análisis General (A.A.G)

A.A.G SIGOB: Concentra la información de todos los mercados de datos de la ONE, formando el almacén SIGOB.

Descripción del Área de Análisis (A.A)

A.A Plan Turquino: Concentra la información referente a los indicadores de Agricultura, Ganadería y Silvicultura, Industria y Servicios. Además de contener los reportes estadísticos correspondiente de cada indicador.

Capítulo III Implementación del Mercado de Datos

Descripción de los Libros de Trabajo

L.T Agricultura, Ganadería, Silvicultura: Concentra la información referente a los indicadores de Agricultura, Ganadería y Silvicultura, contiene 1 reporte que permite realizar un análisis general de los datos.

L.T Industria: Concentra la información referente a los indicadores de Industria y contiene 1 reporte que permite realizar un análisis general de los datos.

L.T Servicios: Concentra la información referente a los indicadores de Servicios y contiene 1 reporte que permite realizar un análisis general de los datos.

Descripción de los Reportes del L.T Agricultura, Ganadería, Silvicultura:

T.S 1: Agricultura, Ganadería, Silvicultura: Muestra los indicadores referentes a Agricultura, Ganadería y Silvicultura en el Sector Estatal y No Estatal, por provincias y municipios correspondientes del país en el Cuarto Trimestre del año 2009.

Descripción de los Reportes del L.T Industria

T.S 1: Industria: Muestra los indicadores referentes a la Industria en el Sector Estatal y No Estatal, por provincias y municipios correspondientes del país en el Cuarto Trimestre del año 2009.

Descripción de los Reportes del L.T Servicios

T.S 1: Servicios: Muestra los indicadores referentes a los Servicios en el Sector Estatal y No Estatal, por provincias y municipios correspondientes del país en el Cuarto Trimestre del año 2009.

El mercado de datos Plan Turquino se compone por 3 reportes en total, los cuales fueron distribuidos dentro de los 3 libros de trabajo. Para la elaboración de estos se utilizaron como herramientas: el Mondrian para la visualización de los reportes, utilizando el Tomcat como servidor web y el Pentaho BI Server, donde finalmente se estructura el mapa de navegación junto con todos los reportes candidatos, como se muestra en las siguientes figuras:

Capítulo III Implementación del Mercado de Datos



FIG. 9 ESTRUCTURA DE LA NAVEGACIÓN

| | | | | Tiempo | | |
|----------------------------------|---|------|---------------------|--------------------------|-------------------|-----------------|
| | | | | Cuarto Trimestre | | |
| | | | | Medidas | | |
| División Política Administrativa | Indicador Agricultura, Ganadería y Silvicultura | UM | Variante del sector | Existencia final periodo | Real año anterior | Plan año actual |
| Cuba | Produc de pescado | t | Estatad | 0,0 | 1798,4 | 1258,7 |
| | | | No Estadad | 0,0 | 15,6 | 9,2 |
| | Produc de pesc MINAG | t | Estatad | 0,0 | 431,0 | 387,5 |
| | | | No Estadad | 0,0 | 13,9 | 9,2 |
| | Distrib de pescado | t | Estatad | 0,0 | 554,9 | 558,6 |
| | | | No Estadad | 0,0 | 5,2 | 5,2 |
| | Financ en divisas | MCUC | Estatad | 0,0 | 12,9 | 2,1 |
| | Siemb de alevines | MU | Estatad | 0,0 | 333,5 | 380,0 |
| | Siemb de alev MINAG | MU | Estatad | 0,0 | 56,0 | 100,0 |
| | Espejos de agua | ha | Estatad | 0,0 | 6753,0 | 5391,2 |
| No Estadad | | | 0,0 | 21,8 | 19,8 | |

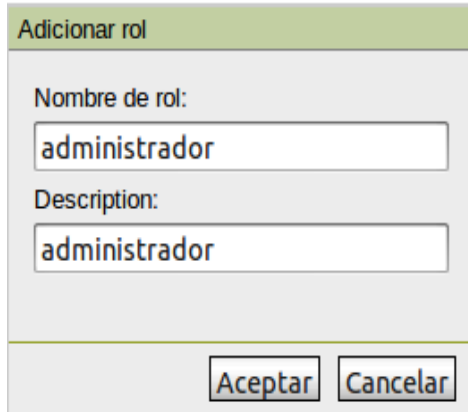
FIG. 10 EJEMPLO DEL REPORTE AGRICULTURA, GANADERÍA, SILVICULTURA

3.3.3 Configuración de la seguridad de los usuarios

Durante la implementación del subsistema de visualización del mercado de datos Plan Turquino se crearon 2 usuarios y roles los cuales tienen diferentes permisos de acceso a la información, proporcionando una mayor seguridad al sistema.

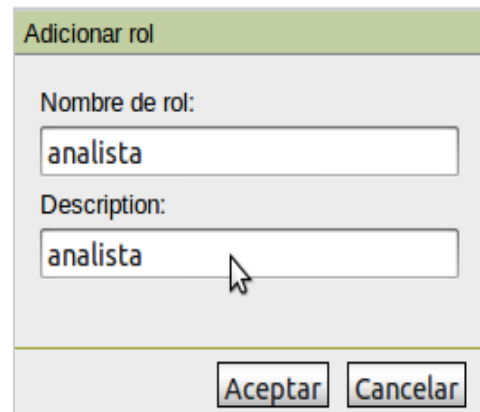
Capítulo III Implementación del Mercado de Datos

- El rol de administrador tiene todos los permisos de la aplicación y posee el usuario administrador.
- El rol de analista tiene permiso de solo lectura y posee el usuario analista.



The screenshot shows a web form titled "Adicionar rol". It has two input fields: "Nombre de rol:" with the text "administrador" and "Description:" with the text "administrador". At the bottom, there are two buttons: "Aceptar" and "Cancelar".

FIG. 11 ROL ADMINISTRADOR



The screenshot shows a web form titled "Adicionar rol". It has two input fields: "Nombre de rol:" with the text "analista" and "Description:" with the text "analista". A mouse cursor is pointing at the "Description:" field. At the bottom, there are two buttons: "Aceptar" and "Cancelar".

FIG. 12 ROL ANALISTA

3.3 Conclusiones parciales

En este capítulo se llevó a cabo todo el proceso de carga del modelo de datos, quedó estructurada la BD por 2 esquemas y 9 tablas. Se implementó el subsistema de integración donde se obtuvieron 6 transformaciones logrando de esta forma, la carga de los hechos de manera satisfactoria. Además se modeló el esquema multidimensional a través del diseño de los 3 cubos OLAP. Finalmente se implementó el subsistema de visualización en el que fueron realizados 3 reportes.

Capítulo IV Validación y prueba del Mercado de Datos

Capítulo IV: Validación y pruebas del Mercado de Datos

Introducción

En este capítulo se realizarán las validaciones al sistema y las pruebas pertinentes que comprueben la calidad del producto, como son los casos de prueba y la lista de chequeo.

4.1 Lista de chequeo

Las listas de chequeo se aplican con el objetivo de evaluar la calidad de los artefactos que se generan en la solución. Mediante estas se pueden cubrir todos los aspectos aplicables en los temas de análisis del Plan Turquino.

La lista de chequeo contiene diferentes indicadores a evaluar, los cuales se encuentran distribuidos en tres secciones fundamentales:

- ✓ **Estructura del documento:** abarca todos los aspectos definidos por el expediente de proyecto o el formato establecido por el proyecto.
- ✓ **Indicadores definidos:** abarca todos los indicadores a evaluar durante la etapa de desarrollo del mercado.
- ✓ **Semántica del documento:** contempla todos los indicadores a evaluar respecto a la ortografía, redacción y demás.

4.1.1 Elementos que forman parte de la estructura de la lista de chequeo

- ✓ **Peso:** define si el indicador a evaluar es crítico o no.
- ✓ **Indicadores a evaluar:** son los indicadores a evaluar en las secciones Estructura del documento, Elementos definidos por el modelo de desarrollo y Semántica del documento.
- ✓ **Evaluación (Eval):** es la forma de evaluar el indicador. El mismo se evalúa de 1 en caso de que exista alguna no conformidad sobre el indicador y 0 en caso de que el indicador revisado no presente inconformidades.
- ✓ **N.P. (No Procede):** se usa para especificar que el indicador no es necesario evaluarlo en ese caso.

Capítulo IV Validación y prueba del Mercado de Datos

- ✓ **Cantidad de elementos afectados:** especifica la cantidad de errores encontrados sobre el mismo indicador.
- ✓ **Comentario:** especifica los señalamientos o sugerencias que quiera incluir la persona que aplica la lista de chequeo. Pueden o no existir señalamientos o sugerencias.

Una vez aplicada la lista de chequeo se detectan los indicadores evaluados de mal y con el objetivo de darles solución se especifican en una tabla de no conformidades (NC), la cual presenta la siguiente estructura:

- ✓ **No.:** es un número consecutivo e indica la cantidad de no conformidades identificadas.
- ✓ **Elemento de evaluación:** se refiere a un número que identifica al elemento de evaluación para el cual se corresponden los indicadores identificados.
- ✓ **No conformidad:** especifica la no conformidad a la que se refiere.
- ✓ **Fase correspondiente:** especifica la fase del procedimiento a la que corresponde la no conformidad encontrada.
- ✓ **Significación:** especifica si la no conformidad es o no significativa, dependiendo si el indicador es o no crítico.
- ✓ **Recomendación:** especifica si la no conformidad es una recomendación, es decir que no es de obligatorio cumplimiento que se solucione por parte de los diseñadores.
- ✓ **Estado NC:** especifica el estado de solución en que se encuentra la no conformidad, puede ser: pendiente o solucionada.
- ✓ **Respuesta del equipo de desarrollo:** si es necesario se especifica la respuesta que le da el equipo de desarrollo a la no conformidad.

Los resultados después de la aplicación de la lista de chequeo al desarrollo del mercado de datos Plan Turquino, son los siguientes:

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

Capítulo IV Validación y prueba del Mercado de Datos

| crítico | 1. ¿Están los documentos en correspondencia con las planillas estándares que han sido definidas por el proyecto? | 0 | | 0 | |
|---|--|------|------|---------------------------------|-------------|
| Indicadores definidos en el desarrollo | | | | | |
| Peso | Indicadores a evaluar | Eval | (NP) | Cantidad de elementos afectados | Comentarios |
| crítico | 1. ¿Se efectuó un estudio previo de la entidad cliente? | 0 | | 0 | |
| crítico | 2. ¿Se realizó el diseño del modelo correspondiente al mercado de datos Plan Turquino unido con el cliente y los especialistas del centro? | 0 | | 0 | |
| crítico | 3. ¿Se ejecutó el proceso de Extracción, Transformación y Carga correspondiente al mercado de datos? | 0 | | 0 | |
| crítico | 4. ¿Las transformaciones se pueden ejecutar desde cualquier PC? | 0 | | 0 | |
| crítico | 5. ¿Se realizó el proceso de Inteligencia de Negocio correspondiente al mercado de datos Plan Turquino? | 0 | | 0 | |
| crítico | 6. ¿Los reportes que se muestran | 0 | | 0 | |

Capítulo IV Validación y prueba del Mercado de Datos

| | | | | | |
|----------------|--|---|--|---|--|
| | en la capa de visualización se corresponden con las necesidades del negocio identificadas? | | | | |
| crítico | 7. ¿Se realizaron los diseños de los casos de prueba? | 0 | | 0 | |
| crítico | 8. ¿En dependencia del rol asignado al usuario, el sistema es capaz de garantizar la confidencialidad y seguridad de acceso a los datos? | 0 | | 0 | |
| crítico | 9. ¿Los cambios en los datos se reflejan automáticamente en los reportes de forma instantánea? | 0 | | 0 | |

Semántica del documento

| Peso | Indicadores a evaluar | Eval | (NP) | Cantidad de elementos afectados | Comentarios |
|----------------|---|-------------|-------------|--|---|
| crítico | 1. ¿Se han identificado errores ortográficos en los entregables? | 1 | | 5 | Problemas con las palabras en cuanto al género. |
| crítico | 2. ¿Se deduce claramente lo que se ha especificado en el documento? | 1 | | 2 | |
| | 3. ¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente | 0 | | 0 | |

Capítulo IV Validación y prueba del Mercado de Datos

| | | | | | |
|--|------------------|--|--|--|--|
| | en dicha página? | | | | |
|--|------------------|--|--|--|--|

Tabla 1: Lista de chequeo

Como resultado de la Lista de Chequeo se detectaron 7 no conformidades que fueron solucionadas.

4.2 Pruebas

Las pruebas se realizan con el objetivo de comprobar, verificar y revelar la calidad de un producto. Su uso consiste en identificar posibles fallos de implementación, calidad o usabilidad de un programa. Para determinar el nivel de calidad se deben efectuar estas pruebas que permitan comprobar el grado de cumplimiento del sistema diseñado.

4.2.1 Modelo V

El modelo en V es una variación del modelo en cascada que muestra cómo se relacionan las actividades de prueba con el análisis y el diseño. Como se muestra en la Figura 13, la codificación forma el vértice de la V, con el análisis y el diseño a la izquierda y las pruebas y el mantenimiento a la derecha. La unión mediante líneas discontinuas entre las fases de la parte izquierda y las pruebas de la derecha representa una doble información. Por un lado sirve para indicar en qué fase de desarrollo se deben definir las pruebas correspondientes. Por otro sirve para saber a qué fase de desarrollo hay que volver si se encuentran fallos en las pruebas correspondientes. Por lo tanto el modelo en V hace más explícita parte de las iteraciones y repeticiones de trabajo que están ocultas en el modelo en cascada. Mientras el foco del modelo en cascada se sitúa en los documentos y productos desarrollados, el modelo en V se centra en las actividades y la corrección. (12)

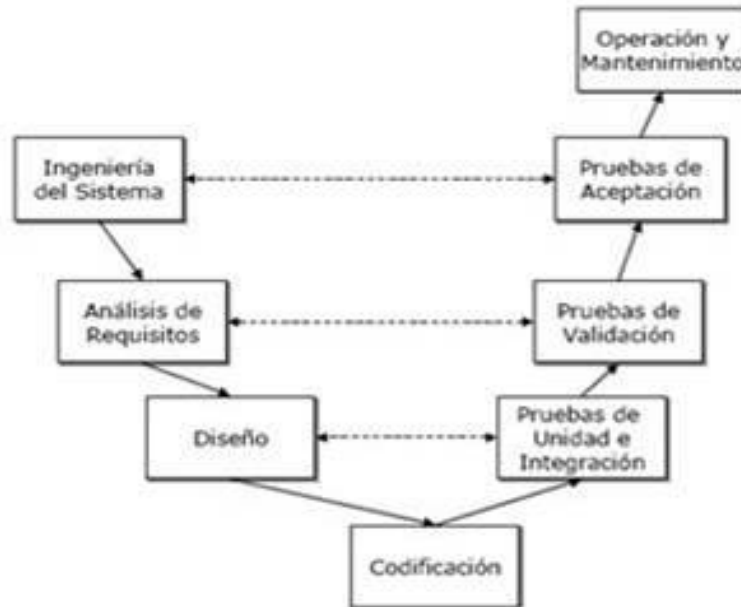


FIG. 13 MODELO DE CICLO DE VIDA EN V

El centro de DATEC realiza ciertas pruebas para comprobar el estado del producto utilizando el Modelo V, como se muestra a continuación:

- ✓ **Prueba unitaria:** son utilizadas para asegurar que cada uno de los módulos funcione correctamente por separado. Estas pruebas son diseñadas y ejecutadas por el desarrollador una vez terminado el desarrollo de cada componente.
- ✓ **Prueba de integración:** se ejecutan para probar todos los elementos del sistema como un todo. Estas pruebas son diseñadas y ejecutadas por el desarrollador cuando la solución está completa junto a los especialistas del centro.
- ✓ **Prueba del sistema:** son las pruebas que se realizan para determinar el correcto funcionamiento de un sistema y su cumplimiento contra las especificaciones del producto. Estas pruebas son ejecutadas por el desarrollador o equipo de pruebas en un ambiente controlado.
- ✓ **Prueba de aceptación:** son las pruebas que realiza el especialista para determinar si el resultado final del producto está en correspondencia con las necesidades del cliente. En caso de que los resultados sean satisfactorios quedarán oficializados en una carta donde se acepte que el producto está listo para ser utilizado.

Capítulo IV Validación y prueba del Mercado de Datos

Resultados de las pruebas

- ✓ En las pruebas unitarias que se le realizó al proceso de integración de datos, los errores más comunes fueron: que no todos los datos de la fuente eran cargados, debido a que existían en la misma id duplicados. Por lo que se hizo necesario modificar estos id para que los datos fueran cargados satisfactoriamente.
- ✓ En la prueba de integración realizada por el especialista del centro, se integró el mercado de datos Plan Turquino al almacén SIGOB satisfactoriamente.
- ✓ En las pruebas del sistema se verificaron que los casos de pruebas coincidieran con la aplicación. Los mismos fueron revisados por los especialistas del centro, donde detectaron ocho no conformidades que posteriormente fueron resueltas de una forma satisfactoria. Actualmente CALISOFT se encuentra en el proceso de liberación del mercado de datos.
- ✓ En la prueba de aceptación nos entrevistamos con el cliente para comprobar que los requisitos de información estuvieran en correspondencia con el resultado de la aplicación. Además de verificar que los datos cargados fueran los correctos. Se detectó que existían en las medidas, algunos valores redondeados. Estos fueron corregidos logrando así la aceptación del cliente.

4.2.2 Casos de Prueba

Para determinar la calidad del mercado de datos, se realizaron tres casos de prueba, cuyos resultados se recogen en los documentos *05_Caso de Prueba basado en CU (Servicios).doc*, *05_Caso de Prueba basado en CU (Agric, Ganad y Silv).doc* y *05_Caso de Prueba basado en CU (Industria).doc*, que se encuentran en el expediente proyecto. Los mismos tienen como propósito validar los casos de uso de información, aunque pueden existir más casos de prueba en dependencia de la complejidad de los casos de uso que hayan sido identificados. Esta prueba permite demostrar que los requisitos de información identificados durante la fase de análisis se cumplen satisfactoriamente.

A continuación se muestra un fragmento del caso de prueba correspondiente al caso de uso Mostrar Indicadores Producto Industria.

Capítulo IV Validación y prueba del Mercado de Datos

| Escenario | Descripción | Variable 1 | Variable 2 | Respuesta del sistema | Flujo central |
|-----------------|---|--|--|--|---|
| EC Industria | 1.1: Permite visualizar el reporte con las variables presentes en el mismo. | <ul style="list-style-type: none"> ✓ División Política Administrativa ✓ Tiempo ✓ Variante del sector ✓ Indicador Industria | <ul style="list-style-type: none"> ✓ Real año anterior ✓ Real año actual ✓ Plan año actual ✓ PorCiento del crecimiento ✓ PorCiento del cumplimiento | Se muestra la tabla con la descripción correspondiente a cada indicador. | <ul style="list-style-type: none"> Se abre la aplicación Se autentifica Se entra al sistema Se despliega hacia la derecha el componente ubicado en el lateral izquierdo que contiene el navegador Se selecciona el área de análisis de A.A Plan Turquino Se selecciona el libro de trabajo L.T Industria En la parte inferior izquierda se selecciona el reporte deseado En el área de trabajo se visualiza la tabla correspondiente al reporte |

FIG. 14 EJEMPLO DEL CASO DE PRUEBA BASADO EN CU (INDUSTRIA)

4.3 Conclusiones parciales

En este capítulo se aplicaron las pruebas pertinentes para verificar la calidad del sistema, el correcto funcionamiento de la implementación del mismo y las listas de chequeo en vísperas del buen funcionamiento y estado del mercado de datos.

Conclusiones

Conclusiones

Se puede concluir que la investigación realizada cumplió satisfactoriamente con todos los objetivos trazados al inicio de la misma, debido a que:

- Se logró realizar una correcta selección de las herramientas para la implementación del mercado de datos permitiendo que los resultados fueran eficientes y favorables.
- Se logró que los hechos y dimensiones incluidos en el modelo 0024-03 referente a los indicadores del Plan Turquino soportaran todo el proceso de toma de decisiones.
- Se implementaron las vistas de análisis del mercado de datos, las cuales estuvieron en correspondencia con las necesidades del cliente.
- Las pruebas que se realizaron al sistema permitieron validar todos los artefactos correspondientes a ellas, demostrando así la calidad de la solución final.

Recomendaciones

Recomendaciones

- En versiones posteriores se exhorta a crear en la capa de visualización un reporte que contenga una ficha técnica, la cual le permita al cliente obtener una descripción más detallada de los indicadores.
- Se propone para la mejora del Mercado de Datos la creación de nuevas funcionalidades en la parte de Inteligencia de Negocio como es la minería de datos.

Referencias bibliográficas

Referencias bibliográficas

1. **Fonseca, Karina Fernández Fernández y Carlos Navarro.** *Análisis, diseño e implementación del mercado de datos para los indicadores sobre el Plan Turquino para la Oficina Nacional de Estadísticas.* 2010.
2. **Inmon, William H.** "What is a Data Warehouse?". [ed.] Prism. 1995.
3. **Scribd.** [En línea] 2011. <http://es.scribd.com/doc/55234594/36/Base-de-datos-multidimensional>
4. **Base de Datos Oracle.** [En línea] Manuel de la Herrán Gascón, 1999-2004.
<http://www.redcientifica.com/oracle/c0001p0005.html>
5. **Martí, José Luis.** *Inteligencia de Negocios.*
6. **HiperTexto.** [En línea] María Jesús Lamarca Lapuente.
<http://www.hipertexto.info/documentos/metadatos.htm>
7. **Kimball, Ralph.** "The Data Warehouse Lifecycle". 2002.
8. **Huamantumba, Rayner.** *Datamart Paso a Paso.* 2007
9. **autores, Colectivo de.** *Metodología para el desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio en CENTALAD.*2009.
10. **Definición.de.** [En línea] 2008. <http://definicion.de/modelo-de-datos>.
11. **Dataprix.** [En línea] <http://www.dataprix.com/category/integracion-datos/perfilado-datos>.
12. **Spanishpmo.** [En línea] <http://spanishpmo.com/index.php/ciclos-de-vida-modelo-en-v>.

Bibliografía

Bibliografía

1. **Apache Tomcat**. [En línea] 1999-2011. <http://tomcat.apache.org>.
2. **autores, Colectivo de**. *Metodología para el desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio en CENTALAD*.2009.
3. **Base de Datos Oracle**. [En línea] Manuel de la Herrán Gascón, 1999-2004.
<http://www.redcientifica.com/oracle/c0001p0005.html>
4. **DataCleaner**. [En línea] 2010. <http://datacleaner.eobjects.org>.
5. **Dataprix**. [En línea] <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/arquitectura-del-data-warehouse/34-datawarehouse-manager#x1-480003.4.4.5>.
6. **Dataprix**. [En línea] <http://www.dataprix.com/category/integracion-datos/perfilado-datos>.
7. **Definición.de**. [En línea] 2008. <http://definicion.de/modelo-de-datos>
8. **El Rincon del BI**. [En línea] Roberto Espinosa Milla. <http://churriwifi.wordpress.com/2010/07/04/17-3-preparando-el-analisis-dimensional-definicion-de-cubos-utilizando-schema-workbench/>.
9. **Fonseca, Karina Fernández Fernández y Carlos Navarro**. *Análisis, diseño e implementación del mercado de datos para los indicadores sobre el Plan Turquino para la Oficina Nacional de Estadísticas*. 2010.
10. **Hidalgo-Gato, MSc. Ing. Gisel González**. *Almacén de Datos para la Gestión Contable de la EMPAI*. Matanzas, Cuba : s.n., 2009-2010.
11. **HiperTexto**. [En línea] María Jesús Lamarca Lapuente.
<http://www.hipertexto.info/documentos/metadatos.htm>
12. **Huamantumba, Rayner**. *Datamart Paso a Paso*. 2007
13. **Inmon, William H**. "What is a Data Warehouse?". [ed.] Prism. 1995.
14. **Kimball, Ralph**. "The Data Warehouse Lifecycle". 2002.
15. **Lular**. [En línea] 2010. <http://www.lular.info/a/Internet/2010/11/Que-es-MOLAP.html>.
16. **Martí, José Luis**. *Inteligencia de Negocios*.
17. **ONE**. [En línea] 2006-2011. <http://www.one.cu>.
18. **Pentaho**. [En línea] 2005-2011. http://www.pentaho.com/products/data_integration.
19. **Pentaho**. [En línea] 2005-2011. http://community.pentaho.com/faq/bi_platform.php.
20. **Sánchez, Leopoldo Zenaido Zepeda**. *Metodología para el Diseño*. Valencia : s.n., Marzo de 2008.
21. **Scribd**. [En línea] 2011. <http://es.scribd.com/doc/55234594/36/Base-de-datos-multidimensional>

Bibliografía

22. **Sinnexus**. [En línea] 2007-2010. http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx.
23. **Sinnexus**. [En línea] 2007-2010.
http://www.sinnexus.com/business_intelligence/datawarehouse.aspx
24. **Zapata, Luis Giraldo y Yulian**. *Herramientas de Desarrollo de Ingeniería de SW para Linux*. 24 de Septiembre de 2005.

Anexos

Anexos

Glosario de Términos

Glosario de términos

- **A.A:** Área de Análisis.
- **A.A.G:** Área de Análisis General.
- **BI:** Bussines Intelligence.
- **CU:** Caso de Uso.
- **CASE:** Ingeniería de Software Asistida por Computadora.
- **CAP:** Consejo de la Administración Provincial.
- **DATEC:** Centro de Tecnologías de Gestión de Datos.
- **CALISOFT:** Calidad del Software.
- **DCL:** Lenguaje de Control de Datos.
- **DDL:** Lenguaje de Definición de Datos.
- **DML:** Lenguaje de Manipulación de Datos.
- **DPA:** División Política Administrativa.
- **Estadísticas:** Ciencia que se ocupa de la recogida y ordenación de datos numéricos, para obtener a partir de ellos conclusiones basadas en el cálculo de probabilidades.¹
- **ETL:** Extraction, Transformation, Load.
- **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.²
- **L.T:** Libro de Trabajo.
- **MINAG:** Ministerio de la Agricultura.
- **MICONS:** Ministerio de la Construcción.
- **MITRANS:** Ministro del Transporte.
- **Plan Turquino:** Se crea en Cuba con el fin de lograr un desarrollo integral y sostenible de las zonas montañosas.

¹ Tomado de: Kalipedia. <http://www.kalipedia.com/glosario/estadistica.html?x=1515>

² Tomado de: Organización de las Naciones Unidas (ONU). *Integrated and coordinated implementation and follow-up of major*. United Nations conferences and summits.

Glosario de Términos

- **OACE:** Organismo de la Administración Central del Estado.
- **SIGOB:** Sistema de Información de Gobierno.
- **SGBD:** Sistema de Gestor de Base de Datos.
- **Tupla:** Llamadas también registros o filas. Contienen la información relativa a una única entidad.³
- **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.⁴
- **T.S:** Tabla de Salida.
- **UCI:** Universidad de las Ciencias Informáticas.
- **UML:** Lenguaje de Modelado Unificado.

³ Tomado de: MySQL con Clase. <http://mysql.conclase.net/curso/index.php?cap=003>

⁴ Tomado de: Kioskeda.net. <http://es.kioskea.net/contents/internet/tcpip.php3>