

Universidad de las Ciencias Informáticas

Facultad 6



Título: Sistema de Información de Gobierno. Mercado de datos Servicios comunales y personales.

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

Autores:

Beatriz Gago Alonso

Alejandro Pérez Pérez

Tutores:

Ing. Haymée Llerena Esperón

Ing. Arodys E. Dominguez Vaillant

La Habana, Cuba

Junio 2011

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

Beatriz Gago Alonso.

Firma del autor

Alejandro Pérez Pérez.

Firma del tutor

Ing. Haymeé Llerena Esperón.

Firma del tutor

Ing. Arodys E. Dominguez Baillant.

Datos de contacto

Profesor Instructor: Ing. Haymée Llerena Esperón
Universidad de las Ciencias Informáticas, La Habana, Cuba.
Email: hllerena@uci.cu

Profesor Instructor: Ing. Arodys E. Dominguez Vaillant
Universidad de las Ciencias Informáticas, La Habana, Cuba.
Email: adominguez@uci.cu

Dedicatoria

Beatriz

Le dedico esta tesis a mi madre porque es la estrella de mi vida, mi amiga, lo más grande que tengo, a mi padre por darme su cariño y comprensión.

A mis hermanos José Luis y Andrés, por ser mi ejemplo a seguir.

A mis abuelas Estrella y María por todas las cosas que me han enseñado y por ser tan especiales y a mi sobrino José Andrés.

Alejandro

A la memoria de mi abuelo, a mi madre, a mi tía Neni, a mi abuela, a Marilín, a Arnaldito y a mi hermano René.

Agradecimientos

Beatriz

A todas las personas que he conocido durante estos cinco años, que me han ofrecido su amistad y ayuda incondicional.

A los profesores y compañeros del proyecto Almacenes de datos, por siempre estar dispuestos a atenderme y aclararme cualquier duda, A mis amigos de los antiguos grupos 6107, 6302 a los que siempre recuerdo con mucho cariño.

A los tutores Arodys y Haymée por habernos dedicado parte de su tiempo, ayudándonos en todo lo que estuvo en sus manos. Al tribunal y la oponente Yetel, por sus críticas constructivas.

A mi compañero de tesis y amigo, Alejandro, muchas gracias por la amistad que me has ofrecido a lo largo de estos 5 años.

A mis compañeras de apartamento (Elisa, Raiza, Betty Lara), a las amigas que a lo largo de estos cinco años han estado siempre a mi lado, en los momentos buenos y malos, a Aylenis por que gracias a ella la vida en la UCI ha sido más fácil, tener su amistad es un privilegio para mí, a Maray por los gestos que has tenido conmigo, a Edelis que aunque eres pequeña de altura, tienes un corazón enorme, a Leidiana por estar siempre dispuesta a escucharme y darme un buen consejo. Tengo muy buenos recuerdos gracias a ustedes.

Le agradezco a mi suegra Lourdes por sus consejos cuando más los necesité y por ocuparse de mí como una madre. A Janet, Aday y René muchas gracias por el cariño que me han dado, a mi novio Javier, por aceptarme como soy y brindarme su apoyo, amor y comprensión.

Le agradezco a toda mi familia por confiar en mí, a mi hermano Andy por haberse ocupado de mí como si fuera un padre, gracias porque siempre has estado ahí para mí.

A mis padres porque me hacen sentir fuerte y segura en los momentos difíciles, por todo el amor que me dan y por todo el sacrificio que han hecho para que yo pudiera graduarme.

Alejandro

Quiero expresar mis agradecimientos a la persona más importante de mi vida: mi madre, quien me dio la vida y ha guiado mi camino con su amor y dedicación. Por todo su apoyo y confianza depositada en mí, por la educación que me dio, y todo su sacrificio por convertirme en un hombre de bien.

A toda mi familia por ser el motor impulsor en cada aspecto de mi vida. A mi abuela por su cariño y preocupación.

A mi tía Marlenes por ser como una madre para mí, y estar presente cada vez que la he necesitado.

A mis hermanos Marilín y René, gracias a ellos he aprendido el verdadero valor de la amistad.

A mi compañera de tesis Betty, por todos los momentos que pasamos juntos trabajando y estudiando, y por permitirme ser parte de su vida durante toda la carrera.

A Edelis, a Maray, a Aylenis a Magnolia, les agradezco la oportunidad de haber convivido con ustedes, y ser como una familia para mí.

A todas las personas que de una forma u otra me han ayudado a lo largo de estos cinco cursos. A los muchachos del grupo de almacenes de datos, que han sido de gran apoyo. A Yulio, Laritza, Leidiana, Sucl, Patricia, Elisa, Wendy.

A los tutores por su tiempo dedicado a la revisión del documento de tesis y sus consejos.

A todos muchas gracias.

Resumen

La Oficina Nacional de Estadísticas (ONE) es el órgano rector de la estadística en Cuba, su principal objetivo es captar, analizar y difundir los datos recogidos a lo largo y ancho de todo el país. Para la recolección de los datos, existen modelos estadísticos donde se recoge toda la información de los diferentes sectores, tanto económicos como sociales; entre los que se encuentra Servicios comunales y personales. El presente trabajo surge a partir de la necesidad que tiene este centro de analizar el comportamiento de los indicadores de esta área, para así apoyar la toma de decisiones respecto a la actividad de limpieza e higiene comunal.

En aras de optimizar la velocidad de respuesta del sistema a las consultas que se realicen sobre la información, se tiene como tarea la creación de un almacén de datos, y dentro de este, un mercado de datos para los indicadores Servicios comunales y personales. El análisis, diseño e implementación de dicho mercado de datos constituyen la esencia de este trabajo.

Palabras claves: estadística, decisiones, almacén de datos, mercado de datos, servicios comunales.

Índice

Introducción	1
Estructura de la Tesis	4
Capítulo 1: Fundamentos teóricos de los almacenes de datos	5
1.1 Introducción	5
1.2 Almacenes de datos	5
1.2.1 Definición de almacén de datos	5
1.2.2 Características de los almacenes de datos.....	6
1.2.3 Ventajas y desventajas de los almacenes de datos.....	6
1.2.4 Procesos de un almacén de datos.....	7
1.2.5 Inteligencia de negocios.....	8
1.3 Tendencias actuales de los almacenes de datos	8
1.3.1 Tendencias actuales de los almacenes de datos en el mundo	8
1.3.2 Tendencias actuales de los almacenes de datos en Cuba	9
1.4 Modelo multidimensional	9
1.5 Proceso OLAP	10
1.5.1 Características de OLAP	11
1.5.2 Tipos de OLAP	11
1.6 Mercado de datos	12
1.6.1 Beneficios del uso de los mercados de datos.....	12
1.7 Metodologías para el desarrollo del almacén de datos	13
1.7.1 Metodología utilizada en el centro DATEC	13
1.8 Tecnologías y herramientas para el desarrollo	14
1.8.1 Sistemas Gestores de Bases de Datos (SGBD)	14
1.8.2 Herramienta de Modelado	15

1.8.3 Herramientas para la Integración de datos (ETL).....	16
1.8.4 Herramientas para la Inteligencia de negocios (BI).....	18
1.9 Conclusiones	20
Capítulo 2: Análisis y diseño del mercado de datos Servicios comunales y personales	22
2.1 Introducción	22
2.2 Estudio preliminar del negocio	22
2.3 Necesidades de información.....	22
2.4 Especificación de requerimientos	22
2.4.1 Requisitos de información.....	22
2.4.2 Requisitos funcionales	23
2.4.3 Requisitos no funcionales	24
2.4.4 Requisitos Multidimensionales	27
2.4.5 Reglas del negocio	28
2.5 Modelo de casos de uso del sistema.....	29
2.5.1 Actores del sistema	29
2.5.2 Diagrama de casos de uso del sistema.....	29
2.5.3 Especificación de casos de uso del sistema.....	29
2.6 Especificación del modelo dimensional.....	32
2.6.1 Tablas de dimensiones	32
2.6.2 Tablas de hechos	34
2.6.3 Medidas	34
2.6.4 Matriz BUS	35
2.7 Modelo físico	35
2.8 Estrategia de respaldo y recuperación.....	36
2.9 Conclusiones	37
Capítulo 3: Implementación del mercado de datos Servicios comunales y personales.....	38

3.1	Introducción	38
3.2	Roles y privilegios	38
3.3	Implementación del modelo de datos	39
3.4	Perfilado de los datos	39
3.5	Arquitectura de integración	40
3.6	Implementación de los procesos de ETL	41
3.6.1	ETL de dim_dpa	41
3.6.2	ETL de dim_temporal_trimestre	41
3.6.3	ETL de dim_indicadores_comunales	41
3.6.4	ETL de la tabla de hechos	42
3.6.5	Implementación de los trabajos	43
3.7	Proceso de Inteligencia de negocios	44
3.7.1	Implementación de los cubos OLAP	44
3.7.2	Arquitectura de información	45
3.7.3	Reportes candidatos	46
3.8	Mapa de Navegación	49
3.9	Conclusiones	50
Capítulo 4: Validación del mercado de datos Servicios comunales y personales		51
4.1	Introducción	51
4.2	Calidad del software	51
4.3	Factores que determinan la calidad del software	51
4.4	Pruebas	51
4.4.1	Pruebas aplicadas al mercado de datos	52
4.5	Conclusiones	60
Conclusiones Generales		61

Recomendaciones	62
Referencias	63
Bibliografía	64
Anexos.....	67
Glosario de Términos	68

Introducción

En la sociedad actual el gran avance tecnológico se debe fundamentalmente al desarrollo de las tecnologías de la información y las comunicaciones (TIC), es indiscutible que la informática ha tomado el control en cuanto a adelantos técnicos se refiere. Gracias a estos avances, el trabajo con grandes cantidades de información se puede hacer prácticamente sin esfuerzo y en un tiempo mínimo. Esta ciencia sirve de apoyo a otras ramas como es la Estadística, llevar el control de la misma es una actividad de gran importancia, ya que puede contribuir en la toma de decisiones, así como estudiar el comportamiento de diversas esferas tanto sociales como económicas.

Cuba no está exenta a esta revolución tecnológica, siendo el desarrollo de la Industria del Software un medio importante para incrementar el desarrollo económico del país. El uso del Software en la Estadística es primordial a la hora de resolver los grandes problemas en los que se ven envueltas muchas instituciones que requieren el manejo de grandes flujos de informaciones, ayudan en el procesamiento de los datos, ya que en la mayoría de los casos facilita los cálculos matemáticos y la elaboración de gráficos. Una de las instituciones que está trabajando para mitigar este tipo de problemas es la Oficina Nacional de Estadística, órgano rector de la estadística en Cuba, su principal objetivo se fundamenta en captar, analizar y difundir los datos recogidos a lo largo y ancho de todo el país. La Oficina Nacional de Estadísticas integra el trabajo estadístico del país en un sistema eficiente, el Sistema Estadístico Nacional (SEN). En el SEN participan coordinadamente los órganos, organismos, instituciones y entidades que elaboran estadísticas en el país. Está integrado por 3 subsistemas: Sistema de Información de Estadística Nacional (SIEN), Sistema de Información de Estadística Territorial (SIET), Sistema de Información de Estadística Complementaria (SIEC).

En la Universidad de las Ciencias Informáticas (UCI) se creó el Centro de Tecnología de Gestión de Datos (DATEC), este se rige por un conjunto de políticas y procedimientos para la obtención de productos de software empaquetados, encargado de desarrollar productos y brindar servicios relacionados con el tratamiento y almacenamiento de datos. Es el responsable de organizar de manera eficiente los grandes volúmenes de información existentes en las diferentes áreas de la ONE. Dicho centro desarrolla un almacén de datos que es parte del proyecto “Sistema de Información de Gobierno” (SIGOB), el cual está formado por varios mercados de datos que se especializan en distintas áreas.

Para la recolección de los datos en la ONE, existen modelos estadísticos donde se recoge información de todos los sectores de la economía y la sociedad cubana, entre los que se encuentra el de los Indicadores de Servicios comunales y personales, el cual presenta toda la información estadística de esta área en Cuba, mostrando el comportamiento de los principales indicadores de la actividad de limpieza e higiene comunal, áreas verdes, producción de flores, desechos sólidos, equipos de recolección, barrido de calles y otros.

La información del área de Servicios comunales y personales en la Oficina Nacional de Estadísticas se recoge en un sistema muy antiguo. Los datos enviados por las oficinas regionales no están en un formato adecuado y no poseen ningún tipo de integridad. La no existencia de una forma adecuada de almacenar, recuperar y presentar la información dificulta analizar los principales reportes y cruces de variables, indicadores, porcentajes y demás aspectos de interés; obstaculizando la disponibilidad de información para Órganos del Estado y afectando el proceso de toma de decisiones. Existen deficiencias en la gestión de los datos tales como limitaciones para recuperar indicadores desde distintas perspectivas de análisis, las múltiples versiones de los datos crean confusión a los analistas, el proceso de recuperación y elaboración de informes es costoso en esfuerzo y tiempo.

Partiendo de lo anteriormente expuesto, se plantea el siguiente **Problema científico**: ¿cómo contribuir a la toma de decisiones en el área Servicios comunales y personales del Sistema de Información de Gobierno?

Para dar solución a la problemática planteada anteriormente se ha definido como **Objetivo General**: Desarrollar un mercado de datos para el área de Servicios comunales y personales del Sistema de Información de Gobierno que contribuya a la toma de decisiones.

Por tanto el **Objeto de Estudio** son los almacenes de datos e inteligencia de negocios y el **campo de acción**: mercado de datos y capa de visualización para el área de Servicios comunales y personales del Sistema de Información de Gobierno.

Para lograr el objetivo general anteriormente expuesto se trazan los siguientes **Objetivos Específicos**:

- ✓ Realizar el análisis y diseño del mercado de datos del área Servicios comunales y personales.
- ✓ Implementar el mercado de datos del área Servicios comunales y personales.
- ✓ Implementar la capa de visualización de datos.

- ✓ Validar el mercado de datos del área Servicios comunales y personales.

Para darle solución a los objetivos específicos se realiza la siguiente planificación de las **Tareas de Investigación**.

- ✓ Caracterización de las metodologías, herramientas y tecnologías a utilizar en el desarrollo de almacenes de datos.
- ✓ Levantamiento de requisitos.
- ✓ Descripción de los casos de uso del mercado de datos.
- ✓ Definición de los hechos, las medidas y las dimensiones del mercado de datos.
- ✓ Diseño del modelo de datos.
- ✓ Definición de la arquitectura del mercado de datos.
- ✓ Diseño del subsistema de integración.
- ✓ Diseño del subsistema de visualización.
- ✓ Diseño de los casos de pruebas.
- ✓ Implementación del almacén.
- ✓ Implementación del subsistema de integración.
- ✓ Implementación del subsistema de visualización.
- ✓ Aplicación de las listas de chequeo.
- ✓ Aplicación de los casos de pruebas.

Estructura de la Tesis

El documento está estructurado por 4 capítulos.

Capítulo 1: Fundamentos teóricos de los almacenes de datos.

Estará referido a los fundamentos teóricos de los almacenes de datos y los mercados de datos, sus características, así como sus ventajas y desventajas. Se realizará también una investigación acerca de las metodologías existentes y las principales herramientas para el desarrollo de los almacenes de datos.

Capítulo 2: Análisis y diseño del mercado de datos Servicios comunales y personales.

En este capítulo se efectúa un estudio preliminar del negocio para identificar las necesidades del cliente. Además, se realiza la descripción de los casos de uso, se identifican las dimensiones, hechos y medidas con que cuenta el negocio y se desarrolla el modelo de datos.

Capítulo 3: Implementación del mercado de datos Servicios comunales y personales.

Se implementa el proceso de extracción, transformación y carga de los datos. Se desarrolla el subsistema de visualización de los datos donde se diseñan los cubos multidimensionales, se desarrolla la estructura de navegación y se implementan los reportes candidatos.

Capítulo 4: Validación del mercado de datos Servicios comunales y personales.

En el capítulo se exponen las pruebas realizadas al mercado de datos, así como los resultados obtenidos en cada una de ellas luego de su aplicación. Dichas pruebas son realizadas para identificar posibles fallos en el producto.

Capítulo 1: Fundamentos teóricos de los almacenes de datos

1.1 Introducción

Este capítulo estará referido a los fundamentos teóricos de los almacenes de datos y los mercados de datos, sus características, así como sus ventajas y desventajas. Se realizará también una investigación acerca de las metodologías existentes y las principales herramientas para el desarrollo de los almacenes de datos.

1.2 Almacenes de datos

En la actualidad toda empresa necesita herramientas confiables que le permitan actuar de manera correcta en las operaciones que realiza. Las decisiones tomadas deben ser rápidas y basadas en buenos cimientos para lo que se necesita de hechos y cifras que deben ser manejados y analizados en el menor tiempo posible debido a que la competencia en los negocios crece rápidamente. El poseer conocimientos correctos significa tener respuestas correctas y trazarse estrategias en beneficio de la empresa.

Las tareas de recolectar, procesar, limpiar y transformar la información no es una tarea sencilla si se tiene en cuenta que una empresa cuenta con distintas áreas, las cuales forman un todo y juegan un papel fundamental para el proceso de toma de decisiones.

El componente de Inteligencia de negocios que resuelve este problema en el manejo y análisis de información es el almacén de datos.

¿Por qué utilizarlos?

Para transformar los datos en información, ante el alcance limitado de los sistemas tradicionales y a la vez brindar mayor credibilidad en la transformación de los datos operacionales, para unificar criterios y así aumentar la productividad.

1.2.1 Definición de almacén de datos

Los almacenes de datos son considerados una potente herramienta para la consulta y el análisis de información. Le brindan un gran apoyo a las empresas en cuanto a comodidad del trabajo, facilitándoles el proceso de toma de decisiones.

Los almacenes de datos han sido definidos de diversas maneras:

- ✓ Concepto definido por Bill Inmon: el almacén de datos es una colección de datos orientados a un tema, integrados, no volátiles e historizados, organizados para dar soporte al proceso de ayuda a la toma de decisiones. (1)
- ✓ Definición de Ralph Kimball, el cual es otro conocido autor en el tema de los almacenes de datos: "una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis". (2)
- ✓ Un Almacén de Datos es una gran colección de datos que recoge información de múltiples sistemas fuentes u operacionales dispersos, y cuya actividad se centra en el análisis de la información. Una vez reunidos los datos de los sistemas fuentes se guardan durante mucho tiempo, lo que permite el acceso a datos históricos; así los almacenes de datos proporcionan al usuario una interfaz consolidada única para los datos, lo que hace más fácil escribir las consultas para la toma de decisiones.(3)

En esta investigación se considera el concepto enunciado por Bill Inmon ya que este, en la opinión de los autores, engloba todos los aspectos de los almacenes de datos.

1.2.2 Características de los almacenes de datos

Orientado a temas: Los datos son almacenados por materias o temas. Estos se organizan desde la perspectiva del usuario final.

Integrado: Integra datos recogidos de diferentes sistemas fuentes de la organización. La integración de datos se muestra de muchas maneras: en convenciones de nombres consistentes, en la medida uniforme de variables, entre otros.

Variante en el tiempo: Los datos son relativos a un periodo de tiempo y deben ser incrementados periódicamente.

No volátil: Los datos almacenados no son actualizados, sólo son incrementados.

1.2.3 Ventajas y desventajas de los almacenes de datos

Un almacén de datos permite administrar la información de manera correcta, accesible, uniforme y

actualizada, lo que favorece que los costos para la toma de decisiones sean menores, aumentando la flexibilidad en el entorno de trabajo.

Ventajas

- Permiten ejecutar o procesar gran cantidad de información.
- Se puede acceder a la información de forma rápida y flexible.
- Capacidad de analizar y explorar las diferentes áreas de trabajo.
- Permiten realizar los planes de la empresa de forma más efectiva, lo que ayuda a mejorar las relaciones con el cliente.
- Facilidades en la gestión y análisis de recursos.
- Debido a la acumulación de datos y experiencia se pueden detectar tendencias y hacer previsiones.

Desventajas

- El almacén de datos realiza la reestructuración de los sistemas operacionales, lo que implica altos costos.
- Es necesario de aplicaciones, sistemas, y almacenamientos específicos.

1.2.4 Procesos de un almacén de datos

Existe un conjunto de procesos básicos detrás de una arquitectura de almacén de datos. A continuación se mencionan brevemente.

Extracción: El proceso de extracción consiste en estudiar y entender los datos fuente, tomando aquellos que son de utilidad para el almacén.

Transformación: Una vez que los datos son extraídos estos se transforman a una forma presentable para los usuarios. Este proceso incluye corrección de errores, resolución de problemas de dominio, borrado de campos que no son de interés, agregación de información.

Carga: Al terminar el proceso de transformación, se cargan los datos en el almacén.

1.2.5 Inteligencia de negocios

Inteligencia de negocios 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. (4)

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

- **Accesibilidad a la información:** Los datos son la fuente principal de este concepto. Lo primero que debe garantizar este tipo de herramientas y técnicas será el acceso de los usuarios a los datos con independencia de la procedencia de estos. (5)
- **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. (5)
- **Orientación al usuario final:** Se busca independencia entre los conocimientos técnicos de los usuarios y su capacidad para utilizar estas herramientas. (5)

1.3 Tendencias actuales de los almacenes de datos

1.3.1 Tendencias actuales de los almacenes de datos en el mundo

Muchas son las empresas en el mundo que han optado por los almacenes de datos para el almacenamiento y tratamiento de su información, convirtiéndose así en la herramienta idónea para ayudar a los ejecutivos a tomar las decisiones apropiadas, que les permitan seguir compitiendo en el mercado.

- La ONE de la República Dominicana para el Sistema Estadístico Nacional pone en práctica el 11 de febrero del 2010 el Almacén Central de Datos que recoge la información del Censo Nacional de Población y Vivienda del 2002, del Directorio de Establecimientos Económicos 2009, de los nacimientos, defunciones, matrimonios y divorcios registrados entre los años 2001-2008, de las encuestas de ingresos y gastos de los hogares del 2007.
- El Instituto de Estudios Turísticos implementó un almacén de datos para el análisis y difusión de la información estadística del turismo en España. El mismo pone al alcance de los analistas una herramienta que permite analizar la información y trabajar con ella de un modo más sencillo que con las tradicionales aplicaciones estadísticas, se publica a través de Internet para que quien la

necesite pueda disponer de ella.

- El departamento financiero de la empresa Multicines S.A en Ecuador implementó un mercado de datos cuyo objetivo es facilitar la toma de decisiones mediante la generación de reportes que permitirán contestar preguntas relacionadas con las ventas realizadas en las distintas dependencias del cine.

1.3.2 Tendencias actuales de los almacenes de datos en Cuba

En Cuba todavía no existe un gran auge en cuanto a almacenes de datos se refiere, sin embargo hay algunas empresas que han dado grandes pasos de avances con el propósito de mejorar la toma de decisiones.

- En la Feria Informática 2002 se presentó un almacén de datos desarrollado por y para CubaCel, basado en ORACLE.
- La corporación CIMEX tiene implementado un almacén de datos comercial que centra su atención en la gestión de inventario, permitiendo una compra-venta eficiente, con la finalidad de disminuir los costos, sin afectar al cliente, permitiendo prestaciones eficientes y con la calidad requerida.
- Actualmente la Oficina Nacional de Estadísticas de Cuba se encuentra inmersa en el desarrollo del almacén de datos Sistema de Información de Gobierno, el cual recogerá la información de los modelos estadísticos de las diferentes áreas atendidas por la institución.

1.4 Modelo multidimensional

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

Las tablas de hechos son el elemento primario del modelo dimensional y contienen los valores del negocio. Una fila de la tabla de hechos corresponde a una medida, que no es más que la intersección entre todas las dimensiones. (6)

Las tablas de dimensiones son las “compañeras integrales” de las tablas de hechos, y contienen las descripciones textuales del negocio. Las dimensiones juegan un papel vital dentro de los almacenes de datos puesto que son la fuente de casi todos los reportes y consultas que se le pueden realizar a este. Debido a esto se les considera la llave para hacer un almacén de datos usable y comprensible. (6)



Ilustración 1: Cubo dimensional de tres dimensiones

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

Esquema en estrella: El esquema de estrella representa una tabla de hechos central relacionada con varias tablas de dimensiones. Es el modelo más conocido y el resto de los modelos resultan modificaciones de él.

Esquema copo de nieve: Se da cuando alguna de las dimensiones se relaciona con una o más tablas de datos. Su objetivo es normalizar las tablas para reducir el espacio de almacenamiento al eliminar la redundancia de datos.

Esquema constelación: Está compuesto por una serie de esquemas de estrella.

1.5 Proceso OLAP

La tecnología de Procesamiento Analítico en Línea OLAP (Online Analytical Processing) permite un uso más eficaz de los almacenes de datos para el análisis de datos en línea, lo que proporciona respuestas rápidas a consultas analíticas complejas e iterativas utilizadas generalmente para sistemas de ayuda en la toma de decisiones. El OLAP presenta los datos a los usuarios a través de un modelo de datos intuitivo y

natural. Con este estilo de navegación, los usuarios finales pueden ver y entender más efectivamente la información de sus bases de datos, permitiendo así a las organizaciones reconocer mejor el valor de sus datos.

El OLAP acelera la entrega de información a los usuarios finales que ven estas estructuras de datos como cubos denominados multidimensionales debido a que la información es vista en varias dimensiones. Esta entrega es optimizada ya que se prepararan algunos valores calculados en los datos por adelantado, en vez de realizar el cálculo al momento de la solicitud. La combinación de navegación fácil y rápida permite a los usuarios ver y analizar información eficientemente. El resultado final: se pasa más tiempo analizando los datos y menos tiempo analizando las bases de datos.

1.5.1 Características de OLAP

- **Rápido:** proporciona la información al usuario a una velocidad constante. La mayoría de las peticiones se deben de responder al usuario en cinco segundos o menos.
- **Análisis:** realiza análisis estadísticos y numéricos básicos de los datos, predefinidos por el desarrollador de la aplicación o definido por el usuario.
- **Compartida:** implementa los requerimientos de seguridad necesarios para compartir datos potencialmente confidenciales a través de una gran población de usuarios.
- **Multidimensional:** llena la característica esencial del OLAP, que es ver la información en determinadas vistas o dimensiones.
- **Información:** acceden a todos los datos y a la información necesaria y relevante para la aplicación, donde sea que esta resida y no esté limitada por el volumen.

1.5.2 Tipos de OLAP

ROLAP

La arquitectura ROLAP (Relational On-line Analytical Processing), accede a los datos almacenados en un almacén de datos para proporcionar los análisis OLAP. La premisa de los sistemas ROLAP es que las capacidades OLAP se soportan mejor contra las bases de datos relacionales. La arquitectura ROLAP es

capaz de usar datos precalculados si estos están disponibles, o de generar dinámicamente los resultados desde los datos elementales si es preciso. (7)

MOLAP

La arquitectura MOLAP (Multidimensional On-line Analytical Processing) usa bases de datos multidimensionales para proporcionar el análisis, su principal premisa es que el OLAP está mejor implantado almacenando los datos multidimensionalmente. Un sistema MOLAP usa una base de datos propietaria multidimensional, donde la información se almacena multidimensionalmente, para ser visualizada en varias dimensiones de análisis. La base de datos multidimensional es la encargada del manejo, acceso y obtención de los datos. (8)

HOLAP

La arquitectura HOLAP (Hybrid On-line Analytical Processing) combina las arquitecturas ROLAP y MOLAP brindando así lo mejor de las dos: un desempeño superior y gran escalabilidad. Mantiene los registros de detalle en la base de datos relacional y las agregaciones en un almacén MOLAP separado. Es usado generalmente para cubos que requieren rápida respuesta con una gran cantidad de datos. (9)

1.6 Mercado de datos

Un mercado de datos es una versión de un almacén de datos. Como los almacenes de datos, los mercados de datos contienen una visión de datos operacionales que ayudan a decidir sobre estrategias de negocio basadas en el análisis de tendencias y experiencias pasadas. La diferencia principal es que la creación de un mercado de datos es específica para una necesidad de datos seleccionados, enfatizando el fácil acceso a una información relevante.

1.6.1 Beneficios del uso de los mercados de datos

- Dado que un mercado de datos soporta menos usuarios que un almacén de datos se puede optimizar para recuperar más rápidamente los datos que necesitan los usuarios.
- Menores cantidades de datos indican que se procesan antes, tanto las cargas de datos como las consultas.

- Los costos que implica la construcción de un mercado de datos son mucho menores a los de la implementación de un almacén de datos.

1.7 Metodologías para el desarrollo del almacén de datos

La metodología es una de las etapas específicas de un trabajo o proyecto que nace a partir de una posición teórica y conlleva a una selección de técnicas concretas o métodos, de cómo se va a realizar la investigación.

Debido a que el uso de los almacenes de datos ha alcanzado un nivel significativo en los últimos años, se hace necesario disponer de una metodología robusta, capaz de guiar todo el ciclo de vida de los mismos, permitiendo así una mejor organización del trabajo y obtener resultados satisfactorios.

Existen varias metodologías para el desarrollo de un almacén de datos, entre ellas se encuentran la metodología Inmon y la metodología de Kimball, que deben su nombre a dos grandes personalidades del mundo de los almacenes de datos: Bill Inmon y Ralph Kimball.

- **Metodología de Bill Inmon:** plantea que el almacén de datos es una parte del todo que conforma un sistema de inteligencia. Una entidad tiene un almacén de datos y los mercados de datos tienen como fuente de información ese almacén. Esta tendencia es conocida como “Top-Down”.
- **Metodología de Ralph Kimball:** el almacén de datos se compone por el conjunto de mercados de datos que existan en la entidad donde esté implementado y la información siempre se almacena en un modelo dimensional. Esta tendencia es conocida como “Bottom-Up”.

1.7.1 Metodología utilizada en el centro DATEC

La metodología que se utiliza en esta investigación, es la definida por la línea de soluciones de almacenes de datos e inteligencia de negocios, del Centro de tecnologías y gestión de datos (DATEC), de la Universidad de las Ciencias Informáticas (UCI). La cual es una adaptación de la metodología Kimball a partir de las necesidades productivas de la UCI.

La metodología Kimball crea los conceptos de hechos y dimensiones, lo que indudablemente es muy eficaz en el proceso de la toma de decisiones y proporciona mayor agilidad en el proceso de desarrollo. Propone ir construyendo el almacén de datos a través de la construcción de los mercados de datos

departamentales, lo que constituye una estrategia buena y coincide con la división lógica de las empresas, entidades y organismos. Existe abundante documentación sobre la misma, la respuesta a todas las dudas y preguntas que puedan surgir se pueden encontrar en la web, a través de los servicios que brinda el grupo creador de la metodología. Es una metodología madura y reconocida por el resto de la comunidad dedicada al tema. Tiene bien definidas las etapas, actividades, artefactos y roles. Como complemento a la misma y fortaleciendo la etapa del levantamiento de requisitos; se tomó lo planteado por Leopoldo Zenaido Zepeda Sánchez en su Tesis de Doctorado, orientando así el trabajo a los casos de uso y se logra estar más alineado con las tendencias y normas de la UCI. (10)

1.8 Tecnologías y herramientas para el desarrollo

1.8.1 Sistemas Gestores de Bases de Datos (SGBD)

PostgreSQL 8.4.0: PostgreSQL es un sistema de gestión de bases de datos objeto-relacional, distribuido bajo licencia BSD y con su código fuente disponible libremente. Es el sistema de gestión de bases de datos de código abierto más potente del mercado.

Su desarrollo comenzó hace más de 15 años, y durante este tiempo, estabilidad, potencia, robustez, facilidad de administración e implementación de estándares han sido las características que más se han tenido en cuenta durante su desarrollo. PostgreSQL funciona muy bien con grandes cantidades de datos y una alta concurrencia de usuarios accediendo a la vez al sistema.

Características

- Integridad referencial.
- Unicode.
- Juegos de caracteres internacionales.
- Múltiples métodos de autenticación.
- Acceso cifrado vía SSL.
- Completa documentación.
- Disponible para Linux y UNIX en todas sus variantes (AIX, BSD, HP-UX, SGI IRIX, Mac OSX,

Solaris, Tru64) y Windows 32/64bit.

- Altamente adaptable a las necesidades del cliente.

PgAdmin 1.10

Pgadmin 1.10 es una aplicación gráfica para gestionar el gestor de bases de datos PostgreSQL, siendo la más completa y popular con licencia de código abierto. Está escrita en C++ usando la librería gráfica multiplataforma wxWidgets, lo que permite que se pueda usar en Linux, FreeBSD, Solaris, Mac OS X y Windows. Es capaz de gestionar versiones a partir de la de PostgreSQL 7.3 ejecutándose en cualquier plataforma.

Está diseñado para responder a las necesidades de todos los usuarios, desde escribir consultas SQL simples hasta desarrollar bases de datos complejas. El interfaz gráfico soporta todas las características de PostgreSQL y facilita en gran medida la administración. La aplicación también incluye un editor SQL con resaltado de sintaxis, un editor de código de la parte del servidor, un agente para lanzar scripts programados. La conexión al servidor puede hacerse mediante conexión TCP/IP y puede cifrarse mediante SSL para mayor seguridad.

1.8.2 Herramienta de Modelado

Visual Paradigm 6.4: Visual Paradigm para UML ayuda a una rápida construcción de aplicaciones de calidad, mejores y a un menor coste. Permite dibujar todos los tipos de diagramas de clases, código inverso, generar código desde diagramas y generar documentación. La herramienta UML CASE también proporciona abundantes tutoriales de UML, demostraciones interactivas de UML y proyectos UML.

Características:

- Producto de calidad.
- Soporta aplicaciones web.
- Varios idiomas.
- Generación de código para Java y exportación como HTML.
- Fácil de instalar y actualizar.
- Compatibilidad entre ediciones.

- Soporte de UML versión 2.1.
- Diagramas de procesos de negocio: proceso, decisión, actor de negocio, documento.
- Ingeniería inversa código a modelo, código a diagrama.
- Ingeniería inversa Java, C++, esquemas XML, XML, NET exe/dll, CORBA IDL.
- Generación de código: modelo a código, diagrama a código.
- Editor de detalles de casos de uso, entorno todo en uno para la especificación de los detalles de los casos de uso, incluyendo la especificación del modelo general y de las descripciones de los casos de usos.
- Diagramas de flujo de datos.
- Generación de bases de datos. Transformación de diagramas de entidad-relación en tablas de base de datos.
- Ingeniería inversa de bases de datos. Desde Sistemas Gestores de Bases de Datos existentes a diagramas de entidad-relación.
- Generador de informes para generación de documentación.
- Distribución automática de diagramas, reorganización de las figuras y conectores de los diagramas UML.
- Importación y exportación de ficheros XMI.

1.8.3 Herramientas para la Integración de datos (ETL)

Pentaho Data Integration 3.0.1 (Kettle)

Pentaho Data Integration abre, limpia e integra la información y la pone en manos del usuario. Provee una consistencia, una sola versión de todos los recursos de información, que es uno de los más grandes desafíos para las organizaciones hoy en día. Pentaho Data Integration permite una poderosa ETL (Extracción, Transformación y Carga).

El uso de Kettle permite evitar grandes cargas de trabajo manual frecuentemente difícil de mantener y de desplegar.

Propiedades básicas

Aparte de ser de código abierto y sin costes de licencia, las características básicas de esta herramienta son:

- Entorno gráfico de desarrollo.
- Uso de tecnologías estándar: Java, XML, JavaScript.
- Fácil de instalar y configurar.
- Multiplataforma: Windows, Macintosh, Linux.
- Permite llegar a un resultado final de forma más efectiva.
- Reduce riesgos y costos de implementación.
- Permite a los desarrolladores enfocarse en funcionalidad.
- Gran facilidad de mantenimiento.
- Gran flexibilidad a la hora de realizar las transformaciones.

Data Cleaner 1.5.3

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 (DMM), metodologías, almacenamiento de datos, proyectos de investigación estadística, la preparación para el proceso ETL y mucho más.

Data Cleaner puede acceder y analizar prácticamente cualquier almacén de datos, incluyendo:

- Bases de datos como Oracle, Microsoft SQL Server, PostgreSQL, MySQL, OpenOffice (ODB)
- Los archivos separados por comas y separados por tabuladores (.csv / .tsv)
- Hojas de cálculo Excel (.Xls)
- Archivos XML

1.8.4 Herramientas para la Inteligencia de negocios (BI)

Pentaho Analysis Services 3.0.4 (Mondrian)

Pentaho Analysis Services, también conocido como Mondrian, es el motor OLAP de la suite de pentaho que permite analizar grandes cantidades de datos en tiempo real. Los usuarios exploran los datos del negocio internándose en la información y la tabulación cruzada con respuestas a la velocidad del pensamiento en consultas analíticas.

El Mondrian puede ser integrado independientemente en cualquier otra plataforma, y de hecho es el componente, junto con Data Integration que más se utiliza independientemente. Es un motor Hybrid OLAP que combina la flexibilidad de los motores ROLAP con una caché que le proporciona velocidad.

Ventajas

- Es un motor ampliamente utilizado y consolidado en entornos JAVA.
- Es el motor de facto de la mayoría de soluciones de BI de código abierto.

Desventajas

- No permite write-back como sí lo permiten los motores MOLAP como Palo o Analysis Services.

Pentaho BI Server 3.6.0

Pentaho BI Server es una aplicación 100% Java2EE que permite gestionar todos los recursos de BI. Cuenta con una interfaz de usuario donde se encuentran disponibles todos los informes, vistas OLAP y cuadros de mando. Así mismo como el acceso a una consola de administración que permitirá gestionar y supervisar tanto la aplicación como los usuarios, qué informes consulta cada usuario, cuándo se han consultado, el rendimiento de la aplicación.

Ventajas

- Aplicación Java2EE 100% extensible, adaptable y configurable.
- La gestión de la configuración, tanto de la instalación inicial como del mantenimiento está muy bien resuelta.

- Se integra con la mayoría de los entornos y se puede comunicar con otras aplicaciones vía webservices.
- Integra todos los recursos informacionales en una única plataforma de explotación.
- Proporciona mucha libertad al usuario y a los desarrolladores para crear contenidos nuevos.

Desventajas

- Los cuadros de mandos son complejos de realizar y ofrecen poca flexibilidad.
- La traducción al español no es al 100%.

Pentaho Workbench 3.2.0

Es un entorno visual para el desarrollo y prueba de cubos OLAP. Si bien la definición del XML para esquemas Mondrian no es extremadamente compleja, en la práctica resulta engorroso recordar cada uno de los elementos junto a sus atributos y sub-elementos. Con esta aplicación, se puede configurar una conexión JDBC como el modelo físico, para luego elaborar el esquema lógico de manera simple y efectiva. Para ello el entorno ofrece un editor de esquemas con la fuente de datos subyacente para su validación. Permite la ejecución de consultas MDX contra el esquema y la base de datos y la navegación por la base de datos subyacente.

Apache Tomcat 5.5

Es un servidor HTTP y un contenedor de servlets. En sus inicios existió la percepción de que el uso de Tomcat de forma autónoma era sólo recomendable para entornos de desarrollo, con requisitos mínimos de velocidad y gestión de transacciones. Hoy en día ya no existe esa percepción y Tomcat es usado como servidor web autónomo en entornos con alto nivel de tráfico y alta disponibilidad.

Apache Tomcat permite escribir y desplegar aplicaciones web complejas de forma sencilla, proporcionando el soporte para características de desarrollo que de otra forma tendrían que ser desplegadas manualmente. Resuelve de forma más eficiente las opciones de autenticación de usuarios, los registros de log, las herramientas de fortificación del servidor, así como la infraestructura y los filtrados a nivel de aplicación.

Dado que Tomcat fue escrito en Java, funciona en cualquier sistema operativo que disponga de la máquina virtual Java.

Pentaho Report Designer 3.6.1

El Pentaho Report Designer es una herramienta independiente que forma parte de la unidad de reportes de Pentaho (Pentaho Reporting), que simplifica el proceso de generación de reportes, permitiendo la creación de informes sofisticados y ricos visualmente, basados en el proyecto de reportes de Pentaho JFreeReport.

El diseñador de reportes ofrece un entorno gráfico familiar, con herramientas intuitivas y fáciles de utilizar, y una estructura de reporte bastante acertada y flexible para darle libertad al diseñador de generar reportes que se adapten totalmente a su necesidad.

Características

- Diseñador gráfico basado en “arrastrar y soltar” (drag & drop) que provee completo control de acceso a los datos, agrupaciones, cálculos, gráficas y formato para reportes de alta resolución.
- Asistente paso a paso integrado, que guía a los diseñadores de reportes durante el proceso de diseño.
- Plantillas de reportes aceleran el proceso de generación, proporcionando un aspecto consistente y atractivo.
- Opciones de salida flexibles incluyendo los populares formatos Adobe PDF, HTML, Microsoft Excel.

1.9 Conclusiones

En este capítulo se exponen los conceptos básicos y se describen las herramientas que se utilizan para la construcción de un mercado de datos. Entre estos conceptos se incluye la “Metodología para el desarrollo de soluciones de almacenes de datos e inteligencia de negocios en DATEC”, que se usa en el desarrollo del software de esta investigación. Esta metodología es una adaptación a la propuesta por Kimball. Además, el modelo de datos utilizado es el esquema en estrella y el modo de almacenamiento que se aplica es el ROLAP.

Finalmente, las herramientas seleccionadas son: PostgreSQL 8.4.0, PgAdmin 1.10, Visual Paradigm 6.4, Pentaho Data Integration 3.0.1, Data Cleaner 1.5.3, Pentaho Analysis Services 3.0.4, Pentaho BI Server 3.6.0, Pentaho Workbench. 3.2.0, Apache Tomcat 5.5, Pentaho Report Designer 3.6.1.

Capítulo 2: Análisis y diseño del mercado de datos Servicios comunales y personales

2.1 Introducción

En este capítulo se efectúa un estudio preliminar del negocio para identificar las necesidades del cliente. Además, se realiza la descripción de los casos de uso, se identifican las dimensiones, hechos y medidas con que cuenta el negocio y se desarrolla el modelo de datos.

2.2 Estudio preliminar del negocio

La ONE es la institución rectora de las estadísticas en Cuba, tiene como principal objetivo gestionar y difundir toda la información estadística del país. La información estadística está captada en modelos de manera organizada por cada centro informante. Dicha información está constituida bajo los diferentes niveles de estructura nacional que posee la ONE y donde cada uno de estos modelos incluye un conjunto de indicadores. Los indicadores relacionados se ajustan a cada centro informante en dependencia del trabajo que realice cada uno, ya sea social, estatal, económico o de otra índole.

En el área de Servicios comunales y personales se realiza un estudio de todas las labores de limpieza e higiene que se realizan en el país, así como los servicios que ofrece la dirección de servicios comunales a la población. Los datos referentes a este sector se captan en el modelo 0901 “Indicadores seleccionados de Servicios comunales”.

2.3 Necesidades de información

Luego de un estudio del negocio, en la ONE se define como necesidad de información el análisis y difusión de los indicadores seleccionados relacionados con el área de Servicios comunales y personales, los cuales son captados en el modelo 0901.

2.4 Especificación de requerimientos

2.4.1 Requisitos de información

Los requisitos de información son aquellos que se corresponden con las necesidades de los usuarios y

reflejan la información deseada. Se definen a partir de los elementos disponibles en las fuentes. Para el levantamiento de los requisitos se realizaron entrevistas con el cliente, a través de la representante de la ONE en la UCI. A continuación se muestran todos los requisitos de información identificados para el área de Servicios comunales:

Ri1: Obtener total de áreas de calles barridas, por división política administrativa (DPA), en un tiempo dado.

Ri2: Obtener total de áreas verdes por concepto de áreas verdes existentes, áreas verdes atendidas por comunales, mantenimiento de áreas verdes, por DPA, en un tiempo dado.

Ri3: Obtener volumen total de desechos sólidos recolectados, por DPA, en un tiempo dado.

Ri4: Obtener total de flores por concepto de flores producidas y flores comercializadas, por DPA, en un tiempo dado.

Ri5: Obtener promedio de equipos trabajando (tractores, colectores, tracción animal), por DPA, en un tiempo dado.

Ri6: Obtener total de servicios necrológicos por concepto de servicios fúnebres, inhumaciones, exhumaciones, promedio de carros fúnebres trabajando, por DPA, en un tiempo dado.

Ri7: Obtener la ficha técnica de todos los indicadores de servicios comunales y personales.

Ri8: Obtener el total de limpieza e higiene por concepto de volumen de desechos sólidos recolectados, área de calles aptas para barrer, área de calles barridas, por DPA, dado un tiempo.

Ri9: Obtener el total de todos los indicadores de la actividad comunales, por DPA, dado un tiempo.

2.4.2 Requisitos funcionales

Los requisitos funcionales describen las funcionalidades que el sistema debe hacer:

RF1: Adicionar usuario.

RF2: Eliminar usuario.

RF3: Adicionar rol.

RF4: Eliminar rol.

RF5: Autenticar usuario.

RF6: Realizar transformación y carga de los datos de comunales.

RF7: Extraer datos de comunales.

RF8: Insertar reporte.

RF9: Modificar reporte.

RF10: Eliminar reporte.

RF11: Configurar elementos del cubo.

RF12: Ordenar resultados.

RF13: Mostrar padres.

RF14: Mostrar propiedades.

RF15: Suprimir filas.

RF16: Invertir eje.

RF17: Detallar miembros.

RF18: Entrar en detalles del reporte.

RF19: Mostrar datos de origen.

RF20: Mostrar gráfico.

RF21: Configurar gráfico.

RF22: Editar consulta MDX.

RF23: Exportar reporte a .pdf.

RF24: Exportar reporte a .xls.

RF25: Imprimir reporte.

2.4.3 Requisitos no funcionales

Los requisitos no funcionales son las cualidades o propiedades que el producto debe tener.

Requisitos Usabilidad

RNF 1. Cumplir con las pautas de diseño de las interfaces.

El sistema debe tener una interfaz gráfica uniforme a través del mismo incluyendo pantallas, menús y opciones. Las pautas de diseño se realizarán siguiendo los lineamientos de la arquitectura de información.

RNF 2. Mostrar los títulos de los componentes que aparezcan en la interfaz del sistema en el idioma español.

Los títulos de los componentes de la interfaz para interactuar con los usuarios deben ser en idioma español, además deben tener una apariencia uniforme en todo el sistema.

RNF 3. Diseño y navegabilidad de los reportes del almacén de datos de manera sencilla.

Un usuario con conocimientos básicos del sistema podrá diseñar un reporte de manera sencilla sin necesidad de ser un experto en las herramientas requeridas para ello. Además, podrá realizar agrupaciones de variables en los reportes de manera dinámica en la misma pantalla de trabajo, agilizando la navegabilidad de los usuarios en un reporte.

Requisitos Fiabilidad

RNF 4. Asegurar recuperación ante fallos.

El sistema debe ser capaz de recuperarse ante un fallo, teniendo en cuenta la complejidad y naturaleza de este.

RNF 5. Garantizar la persistencia de la información.

Para garantizar la persistencia de la información se realizará un respaldo total de los datos del almacén.

Restricciones de diseño

RNF 6. Lograr que los elementos definidos en el almacén tengan una estructura homogénea.

Las estructuras del almacén de datos se nombrarán de una manera estándar teniendo en cuenta el tipo de estructura que se maneje. Se definen convenciones de nombrado con el objetivo de manejar un vocabulario común en el almacén de datos que permita un entendimiento claro y conciso de las estructuras por parte de los desarrolladores que interactúen con el almacén de datos.

RNF 7. Utilizar el lenguaje de programación definido durante la investigación.

Para la programación en el almacén de datos se utilizará SQL como lenguaje dentro del Sistema Gestor de Base de Datos y para la implementación de los procesos de integración de datos el lenguaje JavaScript. Además, se utilizará el lenguaje MDX para realizar las consultas.

Requisitos para la documentación de usuarios y ayuda del sistema

RNF 8. Confeccionar manual de usuario

El sistema debe estar acompañado de un documento que guiará la ejecución del usuario teniendo en cuenta cada funcionalidad.

Requisitos de seguridad

RNF 9. Sesión de usuarios

Los permisos correspondientes al usuario autenticado se activarán una vez que este se autentique y en caso de cambiar, tendrá acceso sólo a la información que le compete de acuerdo con sus privilegios.

Requisitos de software

RNF 10. Utilizar el Sistema Gestor de Base de datos definido durante la investigación.

El Gestor de Bases de Datos para implementar el almacén es el PostgreSQL. Como Interfaz de Administración del Gestor de Base de Datos se usará el PgAdmin.

RNF 11. Utilizar la herramienta de integración de datos definida durante la investigación.

Para el proceso de integración de datos se utilizará la herramienta Pentaho Data Integration.

RNF 12. Utilizar las herramientas para la implementación de la capa de inteligencia de negocios definidas durante la investigación.

De la suite de Pentaho, se usarán los siguientes componentes.

- ✓ Schema Workbench: Es la herramienta gráfica que se utiliza para construir el esquema multidimensional que soportará la creación de los reportes multidimensionales.

- ✓ Pentaho BI Server: Es el servidor de visualización de la suite, que se encarga de visualizar los reportes y controlar el acceso a la información.

Las estaciones de trabajo deben contar además con:

- Firefox 2.0 o superior.

Se requiere la instalación de la Java Virtual Machine 6.0 para el uso de las herramientas anteriores.

Requisitos de hardware

RNF 13: Garantizar características mínimas de hardware para las PC clientes.

- ✓ 1 GB de RAM o superior.
- ✓ Procesador Intel Pentium 4.
- ✓ 80 GB de disco duro.

RNF 14: Garantizar características mínimas de hardware para los servidores.

- ✓ 2 GB de RAM o superior.
- ✓ Procesador Intel Dual Core.
- ✓ 160 GB de disco duro.

2.4.4 Requisitos Multidimensionales

Los requisitos multidimensionales son las variables de entrada y las variables de salida de la solución, definidos a partir de los requisitos de información.

Variables de Entrada:

- ✓ Indicadores (Indicadores de Servicios comunales y personales).
- ✓ Temporal_trimestre (Tiempo).
- ✓ DPA (División política administrativa).

Variables de salida.

- ✓ Real_anno_actual (datos correspondientes al año en curso).
- ✓ Real_anno_anterior (datos acumulados hasta igual período del año anterior).

2.4.5 Reglas del negocio

Las reglas del negocio fueron identificadas en el levantamiento de información y en el análisis de las fuentes, son la entrada fundamental para los procesos de diseño del almacén, ETL y BI.

Datos calculables:

- **R1:** La variable 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. La fórmula para calcularla es $(\text{real_anno_actual} / \text{real_anno_anterior}) * 100$.

Otras reglas del Negocio:

- **R2:** El área de calles aptas para barrer debe ser menor o igual que el área total de calles existentes.
- **R3:** El volumen total de desechos sólidos urbanos debe ser menor o igual que el volumen total de desechos sólidos recolectados.
- **R4:** El volumen total de desechos sólidos reciclados debe ser menor o igual que el volumen total de desechos sólidos recolectados.
- **R5:** El total de vertederos con tratamiento sanitario debe ser menor o igual que el total de vertederos.
- **R6:** El total de vertederos con relleno sanitario manual debe ser menor o igual que el total de vertederos.
- **R7:** Las inhumaciones en tierra deben ser menores o iguales que las inhumaciones.
- **R8:** Una vez cargados los datos en el almacén, no pueden existir campos nulos.

- **R9:** Para la variable 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 al municipio con su respectiva provincia.

2.5 Modelo de casos de uso del sistema

2.5.1 Actores del sistema

Un actor es alguien o algo que interactúa con el sistema; es quien utiliza el sistema.

Administrador de ETL: Realiza los procesos de ETL (Extracción, Transformación y Carga).

Analista: Analiza y consulta la información de los indicadores Servicios comunales y personales.

Administrador: Administra los roles, usuarios y los reportes.

2.5.2 Diagrama de casos de uso del sistema

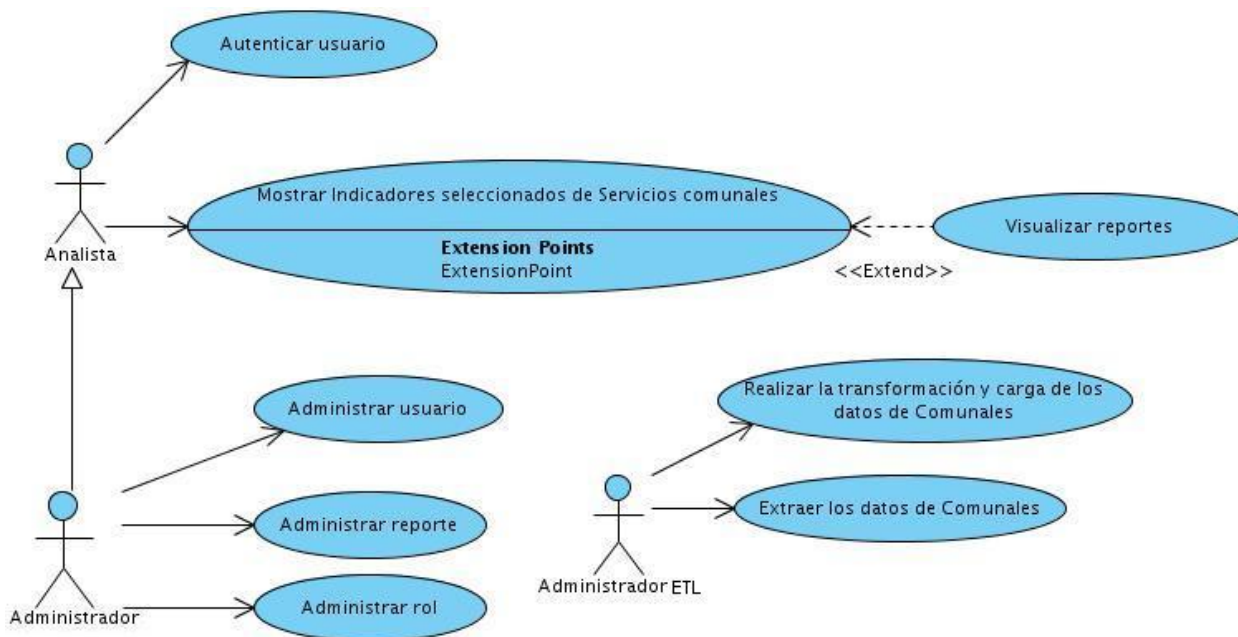


Ilustración 2: Diagrama de casos de uso del sistema

2.5.3 Especificación de casos de uso del sistema

- ✓ Mostrar indicadores seleccionados de Servicios comunales

Nombre del Caso de uso	Mostrar indicadores seleccionados de Servicios comunales.	
Tipo:	Información	
Actor(es):	Analista, Administrador	
Resumen:	El caso de uso se inicia cuando el usuario desea consultar la información relacionada con los indicadores de Servicios comunales. Luego de seleccionar el reporte deseado, el sistema muestra la información contenida en él. El caso de uso finaliza cuando el usuario termina el análisis de la información de los indicadores de Servicios comunales.	
Precondiciones:	El almacén debe tener todos los datos cargados. El usuario se ha autenticado en el sistema. Disponibilidad de los reportes.	
Referencias:	Ri1, Ri2, Ri3, Ri4, Ri5, Ri6, Ri7, Ri8, Ri9, CU Visualizar reporte.	
Prioridad:	Crítico	
Flujo normal de eventos		
Acción del actor	Respuesta del sistema	
1. El usuario selecciona el área de análisis general A.A.G SIGOB.	2. El sistema muestra las áreas de análisis que están contenidas en el A.A.G SIGOB.	
3. El usuario selecciona el área de análisis A.A Servicios comunales.	4. El sistema muestra el libro de trabajo contenido dentro del área de análisis A.A Servicios comunales.	
5. El usuario selecciona el libro de trabajo L.T Indicadores de Servicios comunales.	6. El sistema muestra los reportes contenidos en el libro de trabajo L.T Indicadores de Servicios comunales.	
7. El usuario selecciona el reporte deseado.	8. El sistema muestra la información del reporte seleccionado y brinda la posibilidad al usuario de visualizar el reporte desde diferentes perspectivas de análisis. Ir al caso de uso Visualizar reporte. Finaliza el caso de uso.	
Opciones de reportes		
Entradas	Posibles resultados	
	Salidas	Periodicidad

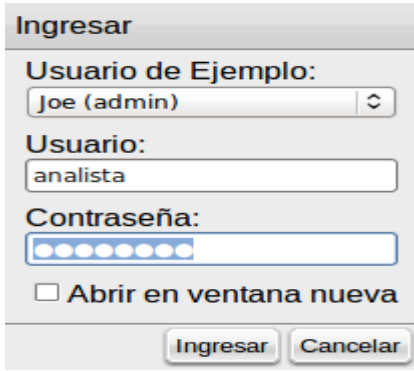
<p>Variables de entradas relacionadas con el CU Analizar Indicadores seleccionados de Servicios comunales:</p> <ul style="list-style-type: none"> ✓ Indicadores ✓ DPA ✓ Temporales 	<p>Variables de salida disponibles en el caso de uso Analizar Indicadores seleccionados de Servicios comunales:</p> <ul style="list-style-type: none"> ✓ Año actual ✓ Año anterior 	<p>Rango de tiempo en que se solicitan las variables de salida:</p> <ul style="list-style-type: none"> ✓ Trimestral
---	--	--

Prototipo de interfaz

		Q4		
		Medidas		
Indicadores	DPA	● Año Actual	● Año Anterior	● Crecimiento
Área de calles barridas	Pinar del Río	38905,5	38993,1	99,8
	La Habana	58226,7	59109,7	98,5
	Ciudad de La Habana	542439,2	575150,4	94,3
	Matanzas	113204,2	107913,3	104,9
	Villa Clara	69995,3	63768,8	109,8

✓ Autenticar usuario

Caso de Uso:	Autenticar usuario
Tipo:	Funcional
Actores:	Analista, Administrador
Resumen:	El CU comienza cuando el actor desea autenticarse, introduce los datos, al ser validados tiene los permisos para entrar al sistema.
Precondiciones:	Usuario registrado en la base de datos.
Referencias	RF5
Prioridad	Crítico
Complejidad	

Flujo Normal de Eventos	
Acción del Actor	Respuesta del Sistema.
1. El usuario entra a la aplicación.	2. Muestra la vista para autenticarse.
3. Introduce usuario y contraseña.	4. Valida que el usuario y la contraseña sean correctos.
	5. El sistema da los permisos para entrar a la aplicación.
<p>Prototipo de Interfaz</p> 	
Flujos Alternos	
	4.1 Si los datos no son válidos, muestra un mensaje de error y regresa al paso 2.

2.6 Especificación del modelo dimensional

Para realizar el modelo dimensional se debe tener en cuenta el análisis de la información por medio de los reportes que se deseen visualizar. Por lo tanto, al modelar un mercado o almacén de datos, se debe tener como objetivo la información deseada en los reportes.

2.6.1 Tablas de dimensiones

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

- ✓ **Dim_indicadores_comunales:** esta dimensión contiene la información de los indicadores de Servicios comunales y personales.

dim_indicadores_comunales		
+dim_indicadores_comunales_id	int4	Nullable = false
indicador_codigo	varchar(11)	Nullable = false
indicador_ficha_tecnica	varchar(500)	Nullable = false
indicador_nombre	varchar(60)	Nullable = false
indicador_descripcion	varchar(200)	Nullable = false
indicador_um	char(25)	Nullable = false
tematica_codigo	varchar(4)	Nullable = false
tematica_descripcion	varchar(100)	Nullable = false
indicador_agregable	char(1)	Nullable = false
tematica_nombre	varchar(60)	Nullable = false
division_codigo	varchar(5)	Nullable = false
division_nombre	varchar(45)	Nullable = false
division_descripcion	varchar(150)	Nullable = false
grupo_codigo	varchar(6)	Nullable = false
grupo_nombre	varchar(45)	Nullable = false
grupo_descripcion	varchar(150)	Nullable = false
seguridad_codigo	varchar(1)	Nullable = false
seguridad_nombre	varchar(25)	Nullable = false

Ilustración 3: Dimensión indicadores comunales

- ✓ **Dim_dpa:** contiene información sobre la división política administrativa del país, dividida en provincias y municipios. La jerarquía permite navegar desde las provincias hasta cada uno de sus municipios.

dim_dpa		
+dim_dpa_id	int4	Nullable = fa...
municipio_extension	float4	Nullable = true
municipio_ext_cayosady	float4	Nullable = true
municipio_ext_tierrafirme	float4	Nullable = true
municipio	char(10)	Nullable = true
provincia_codigo	char(2)	Nullable = false
provincia_nombre	varchar(25)	Nullable = false
provincia_descripcion	varchar(45)	Nullable = false
municipio_codigo	char(4)	Nullable = false
municipio_nombre	varchar(30)	Nullable = false
municipio_descripcion	varchar(45)	Nullable = false

Ilustración 4: Dimensión división política administrativa

- ✓ **Dim_temporal_trimestre:** contiene información del tiempo en que se realizan los reportes por parte de los centros informantes. La jerarquía permite navegar desde los años hasta los trimestres.

dim_temporal_trimestre		
+dim_temporal_trimestre_id	int4	Nullable = false
semestre_nombre	char(10)	Nullable = false
semestre_codigo	varchar(6)	Nullable = false
semestre_numero	int4	Nullable = false
anno_codigo	varchar(4)	Nullable = false
anno_nombre	char(4)	Nullable = false
anno_numero	int4	Nullable = false
trimestre_codigo	varchar(6)	Nullable = false
trimestre_nombre	varchar(11)	Nullable = false
trimestre_numero	int4	Nullable = false

Ilustración 5: Dimensión temporal trimestre

2.6.2 Tablas de hechos

Las tablas de hechos contienen las medidas y las dimensiones asociadas al mismo. La tabla de hechos tiene una clave primaria compuesta por la clave primaria de las tablas de dimensiones relacionadas a esta.

Luego de analizar el modelo, se llegó a la conclusión de que con una tabla de hechos sería suficiente. En ella se agrupan todos los indicadores de Comunales, ya que estos se relacionan con las mismas dimensiones.

- ✓ **hech_series_comunales:** representa el valor (en cifras) de los reportes realizados por los centros informantes sobre los indicadores de Servicios comunales.

hech_series_comunales		
+#dim_dpa_id	int4	Nullable = false
+#dim_temporal_trimestre_id	int4	Nullable = false
+#dim_indicadores_comunales_id	int4	Nullable = false
real_anno_actual	float4	Nullable = false
real_anno_anterior	float4	Nullable = false

Ilustración 6: Hecho series comunales

2.6.3 Medidas

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

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

2.6.4 Matriz BUS

La Matriz BUS representa las relaciones existentes entre los hechos y las dimensiones.

Dimensiones/hechos	hech_series_comunales
Dim_indicadores_comunales	X
Dim_dpa	X
Dim_temporal_trimestre	X

2.7 Modelo físico

Luego de haber determinado las dimensiones y la tabla de hechos, describir cada una de ellas y sus jerarquías, se procedió a crear el modelo físico, desplegando físicamente en el gestor de base de datos cada uno de los objetos creados. En este caso consiste en un diagrama en estrella conformado por una tabla de hechos y tres dimensiones.

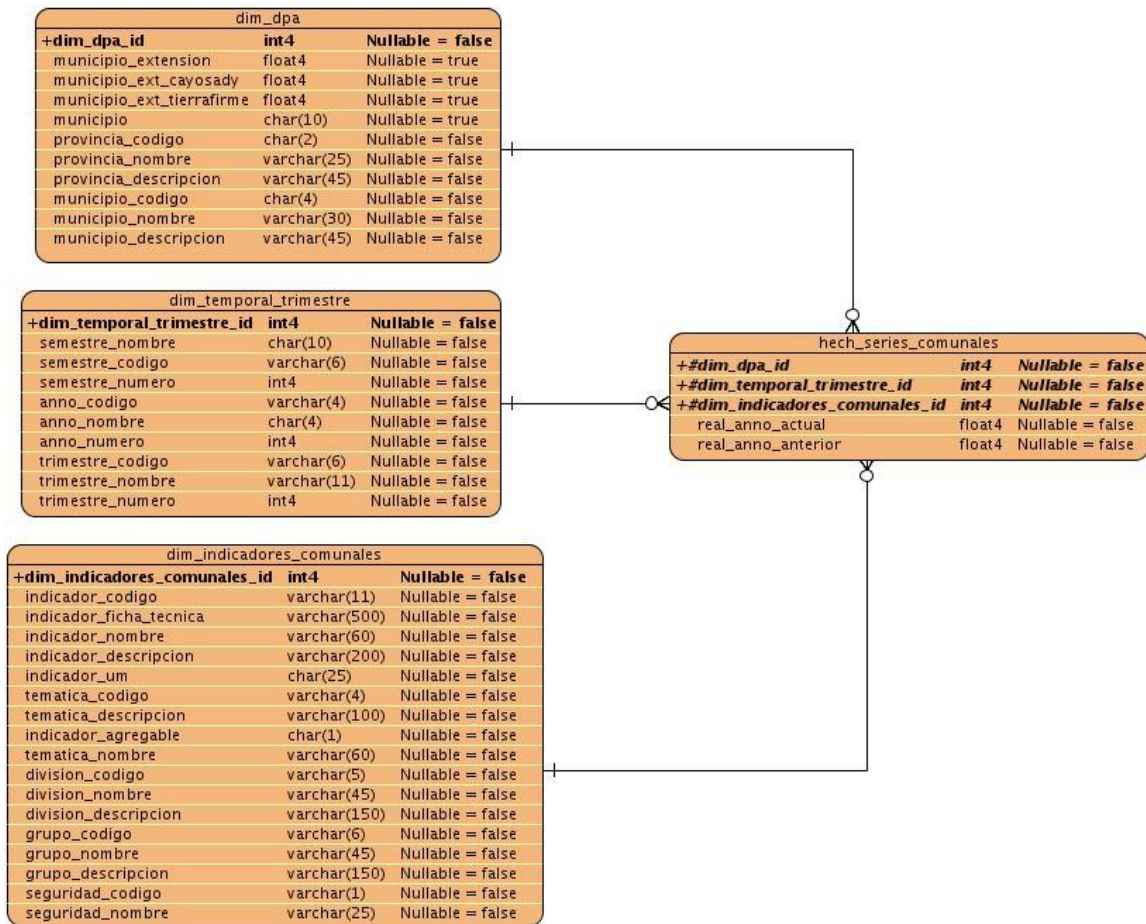


Ilustración 7: Modelo Físico

2.8 Estrategia de respaldo y recuperación

En cualquier sistema informático juega un papel fundamental que el mismo esté preparado para recuperarse ante la pérdida de información. Los mercados de datos no están exentos de dicha problemática. Para contrarrestar la pérdida de información y poder poner en marcha nuevamente el mercado de datos se definieron las siguientes políticas:

- ✓ Hacer una copia total de la base de datos del mercado de datos cada vez que se realicen las cargas de los ficheros fuentes. En este caso sería trimestral.
- ✓ Se le realizará salvase a todas las tablas del mercado, tanto a las dimensiones como al hecho.

- ✓ Se guardarán las salvas en discos duros por un período de un año.
- ✓ Deben ejecutarse pruebas de restauración a las salvas guardadas para verificar que contienen la información correcta.

2.9 Conclusiones

En el desarrollo de este capítulo se definió el análisis y diseño del mercado de datos, paso fundamental para poder realizar su correcta implementación. Se efectuó un estudio preliminar del negocio y las necesidades de información del cliente. Se identificaron los requisitos de información, funcionales y no funcionales. Se construye la matriz bus y se definen las dimensiones, los hechos y las medidas que cumplen con lo establecido por el cliente. De esta forma, se da cumplimiento al primer objetivo específico: realizar el análisis y diseño del mercado de datos del área Servicios comunales y personales.

Capítulo 3: Implementación del mercado de datos Servicios comunales y personales

3.1 Introducción

En este capítulo se documenta todo lo referente a la implementación del mercado de datos Servicios comunales y personales. Se implementa el proceso de extracción, transformación y carga de los datos. Se desarrolla el subsistema de visualización de los datos (BI), donde se diseñan los cubos OLAP, se implementan los reportes candidatos y se desarrolla la estructura de navegación.

3.2 Roles y privilegios

No existe ningún sistema que pueda considerarse libre de riesgos, pueden calificarse de fiables pero nunca de seguros. Juega un papel fundamental definir correctamente los roles existentes, así como los privilegios de los mismos, a continuación se describen cuáles son los roles que interactúan en el mercado de datos y los privilegios de cada uno de ellos.

- ✓ Administrador de ETL: es el encargado de realizar los procesos de extracción, transformación y carga de datos hacia el mercado, tiene permisos de lectura y escritura.
- ✓ Administrador de la base de datos: tiene la responsabilidad de mantener funcional la base de datos que conforma el mercado de datos. Tiene control total sobre el mercado, puede insertar, actualizar, crear objetos, eliminar, dar permisos, y demás tareas de administración.
- ✓ Analista de información: es el encargado de trabajar con los reportes, tiene permisos de sólo lectura sobre los mismos.
- ✓ Administrador del Portal: administra el Pentaho BI Server, crea usuarios, reportes, define privilegios y otras tareas de administración sobre la herramienta de BI.

3.3 Implementación del modelo de datos

En la implementación del modelo de datos se definieron 2 esquemas para agrupar las dimensiones y los hechos.

Esquema dimensiones: Contiene las tablas de las dimensiones que son comunes con la BD general de la ONE.

Esquema mart_comunales: Contiene el hecho y la dimensión de la base de datos del mercado de datos Servicios comunales y personales así como la tabla correlacionadora del modelo propuesto para el mercado de datos.

Esquemas	Tablas
dimensiones	dim_dpa
dimensiones	dim_temporal_trimestre
mart_comunales	dim_indicadores_comunales
mart_comunales	hech_indicadores_comunales
mart_comunales	correlacionador

3.4 Perfilado de los datos

El perfilado de datos es el proceso de examinar los datos que existen en las fuentes de origen de una organización y recopilar estadísticas e información sobre los mismos. El propósito de dichas estadísticas es: (11)

- ✓ Determinar qué datos pueden ser usados para otros propósitos.
- ✓ Conseguir métricas de calidad de datos que incluyen si los datos cumplen los estándares de la organización.
- ✓ Reducir el riesgo de integrar información a nuevas aplicaciones dado que se conoce su estado.
- ✓ Permitir hacer un seguimiento de la calidad de datos.

- ✓ Entender problemas derivados de los datos en proyectos que hagan uso intensivo de los mismos.

La herramienta utilizada en esta investigación para el perfilado de los datos fue el Data Cleaner en su versión 1.5.4. En dependencia de los resultados obtenidos se establecen las reglas para la integración de los datos.

En la siguiente tabla se muestra el resultado obtenido a partir de la realización del perfilado de datos a las fuentes de información. Las tablas SE090109 y DI090109 contienen la información de los indicadores de Servicios comunales, correspondiente al tercer y cuarto trimestre del año 2009 respectivamente.

Los atributos c01 y c02, corresponden a las medidas real_anno_actual y real_anno_anterior de la tabla de hechos, respectivamente. En las columnas Min y Máx se guarda la longitud mínima y máxima de la variable dpa, mientras que para las variables numéricas, esto representa los valores mínimos y máximos que pueden tomar.

Tabla	Atributo	Tipo de dato	Total de valores	Distintos	Nulos	Únicos	Duplicados	Mín	Máx
DI090109	dpa	char	4824	156	0	0	156	4	4
	c01	float	4824	1387	0	998	389	0	113204,2
	c02	float	4824	1394	0	1001	393	0	107913,3
SE090109	dpa	char	3491	157	0	1	156	4	4
	c01	float	3491	1699	0	1206	493	0	758693,6
	c02	float	3491	1751	0	1302	449	0	661381,8

3.5 Arquitectura de integración

El proceso de integración de los datos se divide en los procesos de extracción, transformación y carga. Para este proceso se extraen los datos que se encuentran en la base de datos de la ONE y en los ficheros fuentes. Se realizan las transformaciones requeridas, en caso de encontrarse errores en los datos el flujo se desvía hacia un fichero. Una vez comprobado que la información es correcta se cargan en el mercado de datos.



Ilustración 8: Arquitectura de integración

3.6 Implementación de los procesos de ETL

3.6.1 ETL de dim_dpa

Este proceso tiene como objetivo extraer los datos de la tabla dim_dpa de la base de datos de la ONE y cargar dicha información en la tabla de igual nombre del modelo propuesto para la solución del mercado de datos.



Ilustración 9: Proceso de carga de la dimensión DPA

3.6.2 ETL de dim_temporal_trimestre

Este proceso tiene como objetivo extraer los datos de la tabla dim_temporal_trimestre de la base de datos de la ONE y cargar dicha información en la tabla de igual nombre del modelo propuesto para la solución del mercado de datos.



Ilustración 10: Proceso de carga de la dimensión temporal trimestre

3.6.3 ETL de dim_indicadores_comunales

Este proceso tiene como objetivo extraer los datos de los indicadores de Servicios comunales y personales, existentes en un fichero de formato xls. Para ello se valida que los campos provenientes del

fichero Excel contengan valores erróneos que no coincidan con los definidos en la tabla destino, dim_indicadores_comunales del modelo de datos propuesto.



Ilustración 11: Proceso de carga de la dimensión dim_indicadores_comunales

3.6.4 ETL de la tabla de hechos

Este proceso tiene como objetivo extraer, transformar (si fuese necesario) y cargar los datos fuentes a la tabla de hechos hech_series_comunales.

Los datos para cargar la tabla de hecho provienen de los ficheros dbf pertenecientes a los trimestres de Septiembre y Diciembre (se nombra el trimestre por el mes en el cual finaliza).

Para realizar esta transformación se utilizó la tabla correlacionador, esta es independiente al modelo de datos, pero se utiliza para extraer la información necesaria de los indicadores desde los ficheros dbf y se validó que los campos llaves de las dimensiones no fueran nulos.

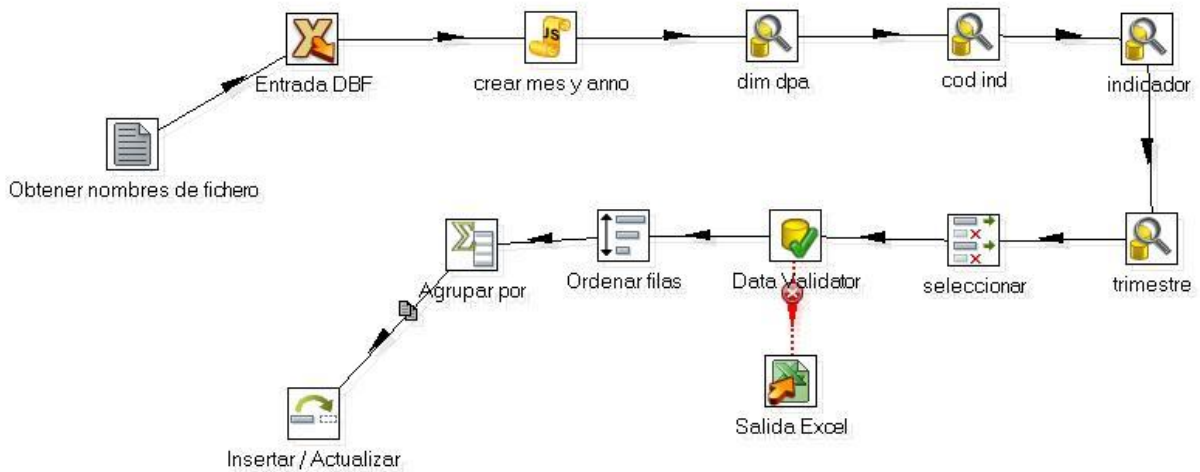


Ilustración 12: Transformación del hecho series comunales

3.6.5 Implementación de los trabajos

El trabajo se encarga de ejecutar todas las transformaciones en un orden lógico definido, primero las dimensiones y después los hechos.

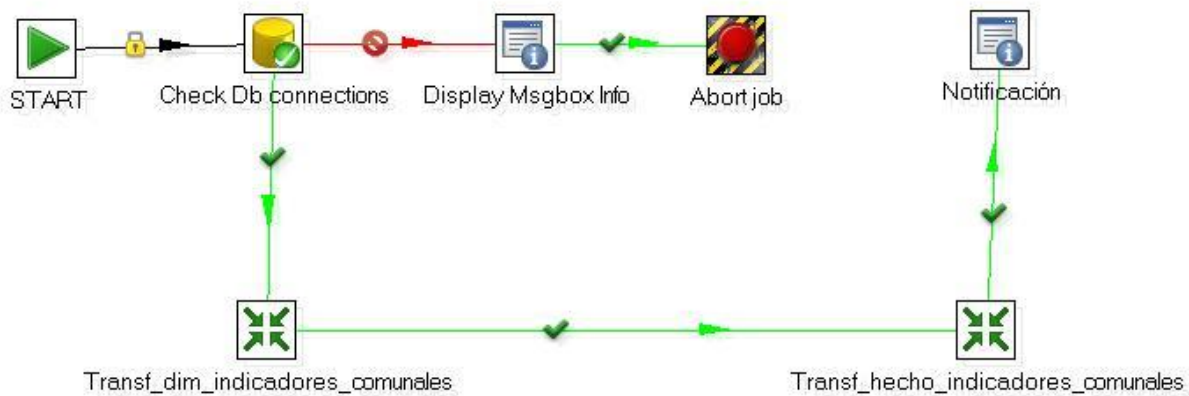


Ilustración 13: Trabajo del mercado de datos Servicios comunales y personales

3.7 Proceso de Inteligencia de negocios

El proceso de inteligencia de negocios permite una vez capturada toda la información del área Servicios comunales y personales obtener el conocimiento del negocio facilitando establecer estrategias, conocer fortalezas y debilidades, además de facilitar la toma de decisiones, por lo que es de vital importancia que este proceso se realice correctamente.

3.7.1 Implementación de los cubos OLAP

Los cubos OLAP permiten tener una visión dinámica del análisis de datos, el usuario puede consultar de una forma rápida y precisa la información contenida en su base de datos.

Con el cubo multidimensional, se pueden representar tabulaciones multidimensionales escogiendo varias métricas a la vez. De esta forma, pueden hallarse totales, medias, porcentajes y ranking de un cruce de dimensiones, con la posibilidad de recalcularlo tantas veces como modificaciones o actualizaciones sufra la base de datos.

En la implementación del cubo multidimensional para el mercado de datos Servicios comunales y personales se utilizó la herramienta Pentaho Schema Workbench que genera un fichero de configuración XML. Se modeló un cubo OLAP compuesto por tres dimensiones y tres medidas, una de ellas es calculable.

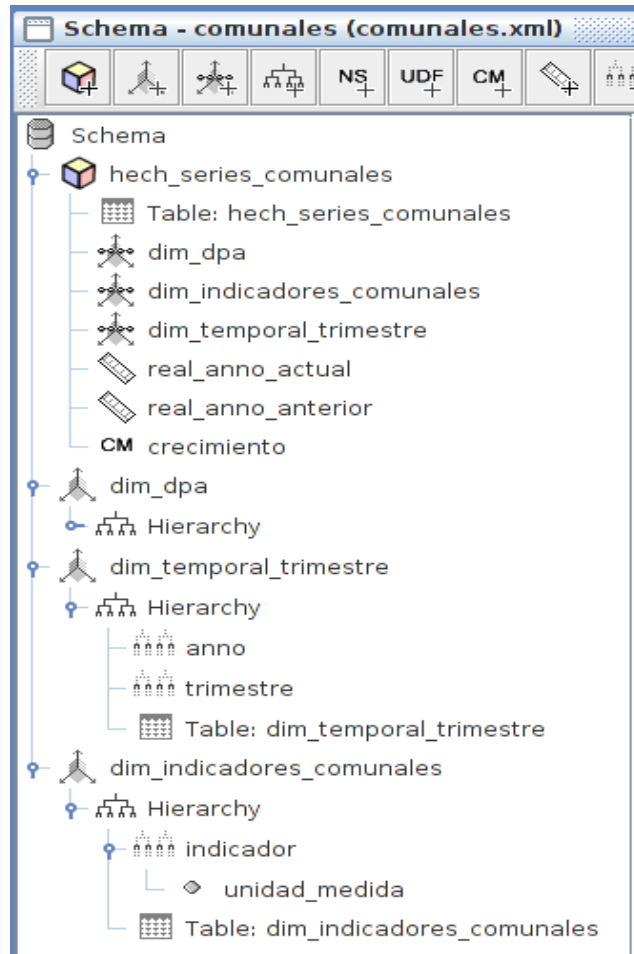


Ilustración14: Cubo multidimensional

3.7.2 Arquitectura de información

La estructura de navegación que se presenta en la capa de visualización está compuesta por un área de análisis que agrupa la información referente al área Servicios comunales y personales, y un libro de trabajo que contiene 9 reportes.

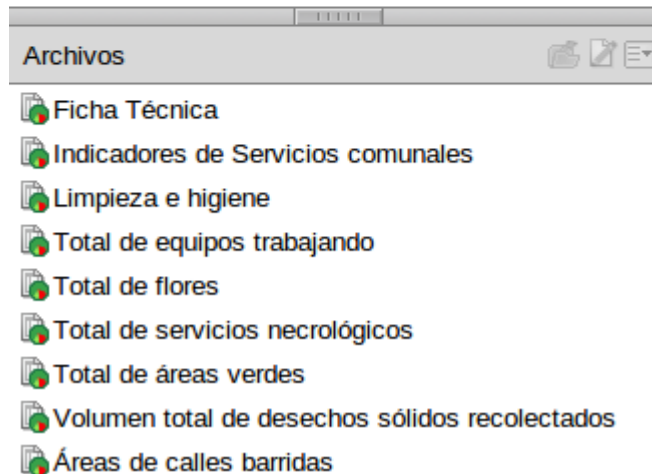
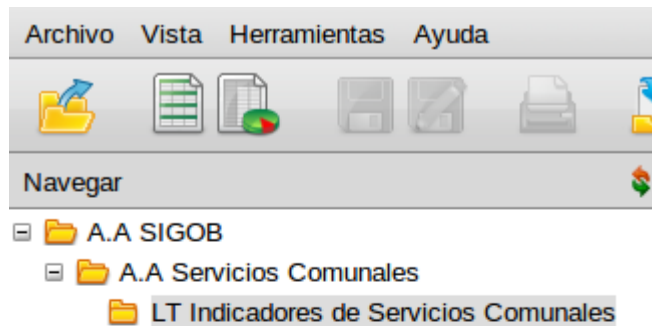


Ilustración 15: Capa de visualización

3.7.3 Reportes candidatos

Los reportes candidatos son la manera en la que el cliente necesita que se muestre la información. Fueron seleccionados después de un análisis detallado del modelo 0901, Indicadores seleccionados de Servicios comunales, y especificados los requisitos de información.

Su implementación está hecha sobre Mondrian, con el uso de su interfaz gráfica es posible visualizar estos reportes y analizar la información con tablas de datos o gráficas de diversos tipos. El reporte Ficha

técnica fue desarrollado en la herramienta Pentaho Report Designer 3.6.1, por tener la particularidad de que no es necesario mostrarlo desde diversas perspectivas de análisis, sólo muestra la descripción de cada indicador con su respectiva unidad de medida. A continuación se mencionan los reportes candidatos que fueron identificados:

Reportes del Libro de trabajo Indicadores servicios comunales:

- Ficha técnica.
- Indicadores de Servicios comunales.
- Limpieza e higiene.
- Total de equipos trabajando.
- Total de flores.
- Total de servicios necrológicos.
- Total de áreas verdes.
- Volumen total de desechos sólidos.
- Área de calles barridas.

En el siguiente reporte se muestra la información del indicador: Área de calles barridas, por provincias en el cuarto trimestre del año 2009.

		Temporales		
		Q4		
		Measures		
Indicadores	DPA	● Año Actual	● Año Anterior	● Crecimiento
Área de calles barridas	▣ Cuba	1458503.4	1444472.5	101.0
	▣ Pinar del Río	38905.5	38993.1	99.8
	▣ La Habana	58226.7	59109.7	98.5
	▣ Ciudad de La Habana	542439.2	575150.4	94.3
	▣ Matanzas	113204.2	107913.3	104.9
	▣ Villa Clara	69995.3	63768.8	109.8
	▣ Cienfuegos	62086.8	49178.3	126.2
	▣ Sancti Spíritus	79482.6	81531.2	97.5
	▣ Ciego de Ávila	41337.1	40679.9	101.6
	▣ Camagüey	52376.3	55934.7	93.6
	▣ Las Tunas	46302.9	45806.8	101.1
	▣ Holguín	53973.7	54222.5	99.5
	▣ Granma	129893.8	129728.7	100.1
	▣ Santiago de Cuba	121469.7	96168.2	126.3
	▣ Guantánamo	39841.9	38732.7	102.9
▣ Isla de La Juventud	8967.7	7554.2	118.7	

Ilustración 16: Reporte área de calles barridas

En el reporte Ficha técnica se muestra la descripción de cada indicador del área de Servicios comunales y personales con su unidad de medida.

Ficha Técnica de los indicadores de Servicios Comunes

Indicador	Unidad Medida	Ficha Técnica
Área total de calles existentes	Miles m2	Constituye el área total de calles que existen, con condiciones o sin condiciones para ser barridas.
Área de calles aptas para barrer	Miles m2	Constituye el área total de calles que, por sus condiciones (asfaltadas y con contenes), se haya determinado por los organismos competentes que están aptas para barrer, independientemente de que se les preste el servicio o no.
Área de calles barridas	Miles m2	Comprende el área total de calles a las que se le prestó el servicio de barrido, ya sea manual o mecánico. Debe acumularse todas las veces que se barren en el periodo informado.
Población urbana con servicio de recogida de basura	Miles hab.	Se refiere al total de la población urbana que recibe el servicio de recogida de basura.
Volumen total de desechos sólidos recolectados	Miles m3.	Constituye el total de desechos domésticos y otros desechos similares recolectados. Incluye desechos de hogares, comercios, pequeños negocios, organismos, e instituciones (escuelas, hospitales, etc.). Incluye los desechos de los servicios municipales como desechos de parques y

Ilustración 17: Reporte Ficha técnica

Para visualizar los restantes reportes remitirse al Anexo 1.

3.8 Mapa de Navegación

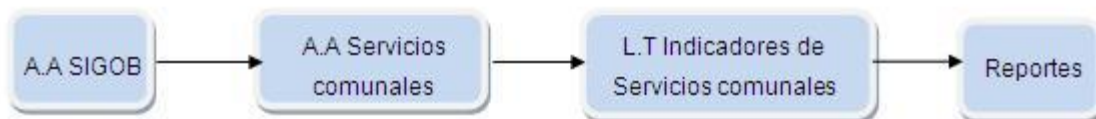


Ilustración 18: Mapa de navegación

- ✓ **Área de Análisis General:** Es el área más abarcadora donde se agrupan los temas de análisis, en

el caso específico de la solución propuesta constituye el almacén de datos SIGOB.

- ✓ **Área de Análisis:** Es la forma de agrupación de la información según su propósito, en el caso específico de la solución propuesta constituye las secciones en las que está dividido el almacén de datos.
- ✓ **Libro de Trabajo:** Constituye las categorías en los que se agrupan los reportes dentro del área de análisis.
- ✓ **Reportes:** Es la unidad básica utilizada para mostrar la información almacenada.

3.9 Conclusiones

En este capítulo se describieron los procesos de ETL que fueron implementados en el Pentaho Data Integration. Además, se detalla el modelo de datos físico desarrollado en la herramienta PgAdmin 3.0. Luego, se presenta el diseño del cubo OLAP que fue descrito utilizando el Pentaho workbench 3.2.0. Finalmente, se muestra la capa de visualización en el Pentaho BI Server 3.6.0.

Una vez terminada la implementación de la aplicación se concluye que:

- ✓ La estructura de esquemas y tablas definida brinda una correcta organización de los datos de la base de datos.
- ✓ Los usuarios y sus privilegios cumplen con lo establecido para velar por la seguridad del sistema.
- ✓ El cubo OLAP permite tener una visión dinámica de los datos del área Servicios comunales y personales de la ONE.

De esta forma, se da cumplimiento al segundo y tercer objetivo específico: realizar la implementación del mercado de datos del área Servicios comunales y personales e implementar la capa de visualización de datos.

Capítulo 4: Validación del mercado de datos Servicios comunales y personales

4.1 Introducción

En el capítulo se exponen las pruebas realizadas al mercado de datos, así como los resultados obtenidos en cada una de ellas luego de su aplicación. Dichas pruebas son realizadas para identificar posibles fallos en el producto.

4.2 Calidad del software

La calidad del software es una preocupación a la que se dedican muchos esfuerzos. Todo proyecto tiene como objetivo producir software de la mejor calidad posible, que cumpla con las expectativas de los usuarios. A continuación se presenta el concepto de calidad del software enunciado por Pressman: Concordancia con los requisitos funcionales y de rendimiento explícitamente establecidos con los estándares de desarrollo explícitamente documentados y con las características implícitas que se espera de todo software desarrollado profesionalmente. (12)

4.3 Factores que determinan la calidad del software

Existen varios factores a tener en cuenta para garantizar que el producto cumpla con la calidad requerida: (13)

Corrección: El grado en que una aplicación satisface sus especificaciones y consigue los objetivos del cliente.

Eficiencia: La cantidad de recursos hardware y software que necesita una aplicación para realizar las operaciones con los tiempos de respuesta adecuados.

Flexibilidad: El esfuerzo requerido para modificar una aplicación en funcionamiento

Facilidad de prueba: El esfuerzo requerido para probar una aplicación de forma que cumpla con lo especificado en los requisitos.

4.4 Pruebas

Las pruebas son el instrumento adecuado para determinar el estado de la calidad de un producto de software. En este proceso se ejecutan pruebas dirigidas a componentes del software o al sistema de software completo, con el objetivo de medir el grado en que este cumple con los requerimientos. Existen varios tipos de pruebas entre los que se encuentran:

- ✓ **Pruebas unitarias:** Verifican el funcionamiento aislado de componentes de software que pueden ser probados de forma separada. (14)
- ✓ **Pruebas 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.
- ✓ **Pruebas de 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.
- ✓ **Pruebas de aceptación:** son las pruebas que se realizan conjuntamente con el cliente, para validar que el sistema cumple con sus necesidades.

4.4.1. Pruebas aplicadas al mercado de datos

Pruebas unitarias

Las pruebas unitarias se realizaron internamente, a medida que fueron implementados el subsistema de integración y la capa de visualización. Durante el desarrollo de los procesos de ETL, se verificó que la carga de los datos se realizara correctamente, en este proceso el error más común fue la existencia de valores nulos, la solución fue enviar estos datos a un fichero Excel, para posteriormente ser tratados por algún especialista de la ONE.

La capa de visualización se validó a través de las consultas realizadas, teniendo en cuenta que las mismas cumplieran con lo definido en los requisitos de información, mostrando correctamente lo establecido por el cliente. Detectándose como no conformidad que no se mostraron los totales de las provincias, en ninguna de las vistas de análisis. Este error fue solucionado al modificar las consultas correctamente.

Casos de Prueba

Se diseñaron los casos de prueba basados en caso de uso, para validar el caso de uso de información. Esta prueba permite demostrar que los requisitos de información identificados durante la fase de análisis se cumplen satisfactoriamente. A continuación, se muestran las tablas presentes en el caso de prueba correspondiente al caso de uso: Mostrar indicadores seleccionados de Servicios comunales.

Escenario (SC): Reportes candidatos:

Escenario	Descripción	Perfiles de análisis	Indicadores a medir	Respuesta del sistema	Flujo central
EC.1.1: Ficha técnica	Permite visualizar cada indicador con su unidad de medida y descripción. Para que la descripción del reporte pueda visualizarse correctamente en la aplicación, se necesita la versión 7.0 de Pentaho BI Server	NA	NA	Se muestra la tabla con la descripción correspondiente a cada indicador	<p>Se abre la aplicación</p> <p>Se autentica</p> <p>Se entra al sistema</p> <p>Se despliega hacia la derecha el componente ubicado en el lateral izquierdo que contiene el navegador</p> <p>Se selecciona el área de análisis de AA. Servicios comunales</p> <p>Se selecciona el libro de trabajo L.T Indicadores de servicios comunales</p> <p>En la parte inferior izquierda se selecciona el reporte deseado</p>
EC.1.2: Indicadores de servicios comunales	Permite visualizar los reportes con las variables presentes en los mismos	Temporales DPA	Año actual Año anterior Crecimiento	Se muestra la tabla con las variables correspondientes a cada escenario	En el área de trabajo se visualiza la tabla correspondiente al reporte
EC.1.3: Limpieza e higiene					

EC.1.4: Total de equipos trabajando					
EC.1.5: Total de flores					
EC.1.6: Total de servicios necrológicos					
EC.1.7: Total de áreas verdes					
EC.1.8: Volumen total de desechos sólidos recolectados					
EC 1.9: Área de calles barridas					

SC Descripción de las variables

No	Nombre del campo	Clasificación	Valor nulo	Descripción
1	DPA	Lista desplegable	No	Muestra las provincias con sus respectivos municipios

2	Temporales	Lista desplegable	No	Muestra los años y sus trimestres
3	Año actual	Valor fijo	Sí	Representa la cantidad real que se tiene en el período que se está evaluando del año actual
4	Año anterior	Valor fijo	Sí	Representa la cantidad real existente al final del mismo período, pero del año anterior al que se representa en el reporte
5	Crecimiento	Valor fijo calculable	Sí	La variable crecimiento representa el porcentaje de crecimiento del año actual con respecto al anterior

Después de aplicada esta prueba por parte de los especialistas del centro se encontró una no conformidad.

Tipo de error	Etapas	Descripción	Respuesta
Redacción	Iteración 1	El nombre del A.A análisis Servicios Comunales, va con minúscula comunales, al igual que el nombre del libro de trabajo L.T Indicadores de Servicios Comunales, va con minúscula servicios comunales	Los nombres fueron cambiados en la aplicación, dejando con mayúscula sólo la primera letra del nombre del área de análisis y del libro de trabajo

Luego de haberse desarrollado las pruebas por el centro, se realiza la entrega de la solución a calidad UCI donde se detectaron 5 no conformidades, dos de las cuales se describen a continuación:

Tipo de error	Etapas	Descripción	Respuesta
Correspondencia con otro artefacto	Iteración 1	En la tabla 4. Descripción de las variables. En el apartado (Valor nulo) aparece que no existen valores nulos para el nombre de los campos (Año actual, Año anterior, Crecimiento) cuando sí hay valores nulos	Para los campos: <ul style="list-style-type: none"> ✓ Año actual ✓ Año anterior ✓ Crecimiento Quedó definido que sí

			existen valores nulos
Correspondencia con otro artefacto	Iteración 1	En la Tabla 3. SC Mostrar Indicadores de Servicios Comunes en el apartado (Indicadores a medir) No coincide el nombre de las Medidas con el de la aplicación	El nombre de las medidas fue cambiado en el cubo multidimensional

Listas de chequeo

La lista de chequeo es un documento que contiene un conjunto de parámetros a medir sobre un aspecto determinado con el objetivo de verificar el grado de cumplimiento de determinadas reglas. Es un instrumento de medición y evaluación que consiste básicamente en un formulario de preguntas referentes al software y a los artefactos que se generan.

La lista de chequeo definida para la evaluación del mercado de datos Servicios comunales y personales, analiza 14 indicadores, de los cuales siete son críticos. Estos indicadores 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 por la etapa:** abarca todos los indicadores a evaluar durante la etapa de diseño.
- **Semántica del documento:** contempla todos los indicadores a evaluar respecto a la ortografía, redacción y otros aspectos de forma y estilo.

A continuación se muestra la lista de chequeo para evaluar el diseño realizado en la presente investigación:

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

Crítico	¿Los entregables contienen las secciones obligatorias de la plantilla estándar definidas para un expediente de proyecto? (portada, control de versiones, reglas de confidencialidad, tabla de contenidos y contenido). Ver expediente de proyecto.				
Indicadores definidos en el desarrollo					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
	¿Se utilizó un lenguaje cuyas sentencias son expresables mediante una sintaxis bien definida?				
Crítico	¿Los reportes son configurables a través de la interfaz del sistema?				
	¿La interfaz está orientada a facilitar el uso de las funciones del sistema por parte de los usuarios?				
Crítico	¿No existen restricciones para construir cubos OLAP con dimensiones y niveles de agregación ilimitados?				
Crítico	¿Los usuarios son capaces de manipular los resultados de manera que se ajusten a sus necesidades, conformando nuevos reportes?				
	¿El sistema responde de una forma rápida a la información que le sea solicitada por el usuario?				

	¿El sistema refleja cualquier lógica del negocio para poder responder a preguntas específicas?				
Crítico	¿El sistema garantiza la confidencialidad y seguridad de acceso a los datos por rol de los usuarios?				
	¿Los datos e información derivados del proceso de análisis realizado mediante la aplicación, apoyan la toma de decisiones en la Institución?				
Crítico	¿Los cambios en los datos se reflejan automáticamente en los reportes de forma instantánea?				
Semántica del documento					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
Crítico	¿Se han identificado errores ortográficos en los entregables?				
Crítico	¿Se entiende claramente lo que se ha especificado en el documento?				
	¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?				

Luego de haberse aplicado las listas de chequeo, por los especialistas del centro a los artefactos se obtuvieron los siguientes resultados:

En el artefacto Especificación de requisitos se detectaron dos no conformidades, en la semántica del documento, en un indicador crítico, La siguiente figura muestra gráficamente el comportamiento de los indicadores evaluados:

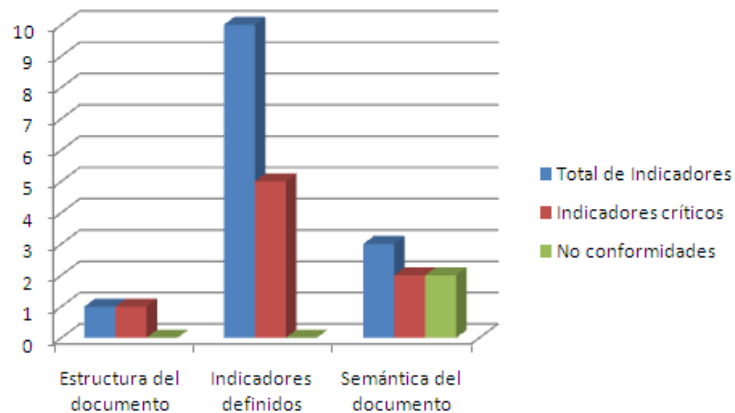


Ilustración 19: Resultados del artefacto Especificación de requisitos

En el artefacto Modelo de casos de uso se detectó una no conformidad, en la semántica del documento, la cual no es significativa y se solucionó inmediatamente. La siguiente figura muestra gráficamente el comportamiento de estos indicadores luego de aplicada la lista de chequeo:

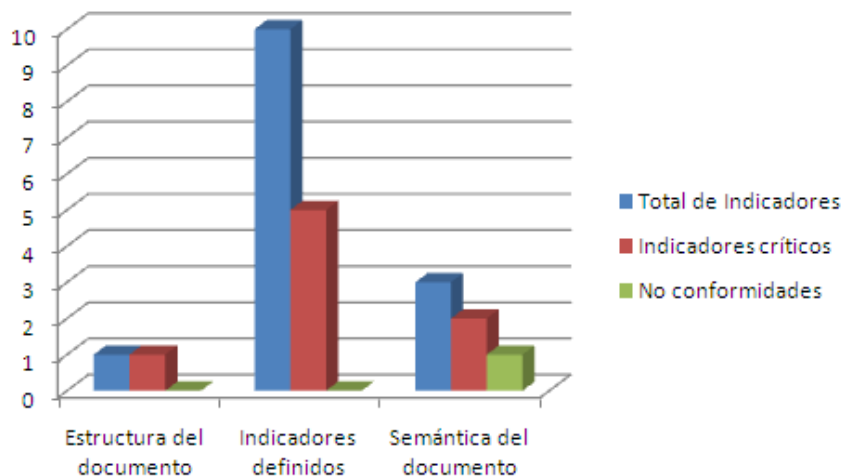


Ilustración 20: Resultados del artefacto Modelo de casos de uso

Pruebas de aceptación

Las pruebas de aceptación al sistema son realizadas por el cliente, con ellas se verifica que los requerimientos definidos se cumplan y que la aplicación esté funcionando correctamente. Estas pruebas se realizaron con la especialista de la ONE: Elena Leonila Fernández García, la cual representa a dicha entidad en la UCI.

Los resultados arrojados luego de ejecutar dichas pruebas fueron satisfactorios, ya que se confirmó que el mercado cumple todas las necesidades de información identificadas con anterioridad. Dichos resultados quedaron oficializados con la carta emitida por el cliente, donde acepta que el producto está listo para ser utilizado.

4.5 Conclusiones

En el capítulo, se detallaron las pruebas realizadas para la evaluación del mercado de datos, las cuales arrojaron resultados satisfactorios. Para la validación de la interfaz se utilizaron los casos de prueba diseñados, así como la aplicación de una lista de chequeo garantizado que los artefactos generados hayan sido correctamente elaborados. También, se realizaron pruebas de aceptación conjuntamente con el cliente para confirmar que el producto satisface sus necesidades, demostrando de esta forma, el correcto funcionamiento del mercado de datos.

De esta manera, se da cumplimiento al cuarto objetivo específico: realizar la validación del mercado de datos del área Servicios comunales y personales.

Conclusiones Generales

El desarrollo de almacenes de datos para la gestión y el procesamiento de grandes volúmenes de datos sigue siendo una línea de investigación abierta, con una gran cantidad de aplicaciones en el mundo empresarial. Un ejemplo de tales aplicaciones, se pudo encontrar en el área de Servicios comunales y personales de la Oficina Nacional de Estadísticas de Cuba.

Al inicio de esta investigación, los datos manejados por el área antes mencionada se almacenaban y procesaban de una manera muy ineficiente con el uso de un sistema muy antiguo. En tal sentido, el objetivo general de esta investigación estuvo enfocado a la solución de este problema que estaba enfrentando la ONE.

Una vez finalizada la investigación se concluye que:

- ✓ La solución del problema planteado mediante el uso de la tecnología de mercados de datos mostró resultados satisfactorios ya que permitió almacenar, recuperar y presentar la información, permitiendo analizar los principales reportes y cruces de variables.
- ✓ Las herramientas seleccionadas soportaron el proceso de diseño e implementación de la solución.
- ✓ Las estructuras dimensionales modeladas (tablas de hechos y dimensiones) abarcan la forma de organización del Modelo 0901 “Indicadores seleccionados de Servicios comunales” y soportan el proceso de toma de decisiones.
- ✓ Las pruebas realizadas validan los artefactos sometidos a ellas y, por lo tanto, la calidad de dichos artefactos como parte de la solución.

De esta manera, los datos estadísticos del área Servicios comunales de la ONE podrán ser procesados de manera eficiente con el sistema propuesto en esta investigación.

Recomendaciones

- ✓ Que se carguen los datos históricos, antes del 2009 y del 2010, para ver cómo se comportan los indicadores de servicios comunales.
- ✓ Se recomienda que se ponga en explotación el mercado de datos para probar las funcionalidades del mismo y detectar nuevas funcionalidades.

Referencias

- 1 **Inmon, William H.** *Building the Data Warehouse*. s.l. : Wiley, 2005.
- 2 **EcuRed.** EcuRed. [En línea] 2010. [Citado el: 8 de 6 de 2011.] http://www.ecured.cu/index.php/Almacen_de_Datos.
- 3 **Velasco, Roberto Hernando.** 2009. Almacenes de Datos. Almacenes de Datos. [En línea] 2009. [Citado el: 16 de 11 de 2010.] <http://www.rhernando.net/modules/tutorials/doc/bd/dw.html>.
- 4 **Alcocer, Alberto.** Societic. [En línea] 29 de 04 de 2008. [Citado el: 4 de 10 de 2010.] <http://www.societic.com/2008/04/hello-world/>.
- 5 **Onuva.** [En línea] [Citado el: 18 de 1 de 2011.] <http://www.onuva.com/es/BI>.
- 6 **Zorrilla, Marta.** 2008. Data Warehouse y OLAP. Universidad de Cantabria : s.n., 2008.
- 7 **Cornejo Lopez, Nydia Lisset.** [En línea] 08 de 2003. [Citado el: 5 de 06 de 2011.] http://biblioteca.usac.edu.gt/tesis/08/08_6793.pdf.
- 8 **Sinnexus.** Sinnexus. [En línea] [Citado el: 1 de 12 de 2010.] http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx.
- 9 **Hernandez, Katy L. Martinez.** Scribd. [En línea] 2001. [Citado el: 10 de 1 de 2011.] <http://www.scribd.com/doc/50422074/Taller-2-Topicos-by-Katy-Martinez>.
- 10 **DATEC.** 2010. Metodología para el desarrollo de soluciones de almacenes de datos e inteligencia de negocio en centalad. Habana : s.n., 2010.
- 11 **Curto Díaz, Josep.** Informationmanagement. [En línea] 30 de 07 de 2010. [Citado el: 9 de 06 de 2011.] <http://informationmanagement.wordpress.com/category/data-profiling/>.
- 12 **Pressman, R. S.** scribd.com. Recuperado el 25 de 4 de 2011, de <http://www.scribd.com/doc/7978336/Ingenieria-de-Software-Un-Enfoque-Practico-Pressman-5th-Ed>
- 13 **Veliz Monteagudo, Mijail.** *Data Warehouse, más allá del almacén de datos*. La Habana : s.n., 2003.
- 14 **Blanco Bueno, Carlos.** [En línea] [Citado el: 7 de 06 de 2011.] <http://ocw.unican.es/enseanzas-tecnicas/ingenieria-del-software-ii/materiales/tema1-pruebasSistemasSoftware.pdf>.

Bibliografía

1. **Andrés Andrade, Christian y Valverde, Verónica Alexandra.** *Desarrollo de un Datamart para el departamento financiero de la empresa Multicines SA.* Quito : s.n., 2007.
2. **Alcocer, Alberto.** Societic. [En línea] 29 de 04 de 2008. [Citado el: 4 de 10 de 2010.] <http://www.societic.com/2008/04/hello-world/>.
3. **Blanco Bueno, Carlos.** [En línea] [Citado el: 7 de 06 de 2011.] <http://ocw.unican.es/enseanzas-tecnicas/ingenieria-del-software-ii/materiales/tema1-pruebasSistemasSoftware.pdf>.
4. **Casas, Alfonso.** PCWorld.es. [En línea] 30 de 6 de 2010. [Citado el: 2 de 12 de 2010.] <http://www.idg.es/pcworldtech/Disponibile-Apache-Tomcat-7.0/doc96785-actualidad.htm>.
5. **Cornejo Lopez, Nydia Lisset.** [En línea] 08 de 2003. [Citado el: 5 de 06 de 2011.] http://biblioteca.usac.edu.gt/tesis/08/08_6793.pdf.
6. **Curto Díaz, Josep.** Informationmanagement. [En línea] 30 de 07 de 2010. [Citado el: 9 de 06 de 2011.] <http://informationmanagement.wordpress.com/category/data-profiling/>.
7. DataCleaner. [En línea] [Citado el: 6 de 12 de 2010.] <http://datacleaner.eobjects.org>.
8. **DATEC.** *Metodología para el desarrollo de soluciones de almacenes de datos e inteligencia de negocio en centalad.* La Habana : s.n., 2010.
9. **EcuRed.** EcuRed. [En línea] 2010. [Citado el: 8 de 6 de 2011.] http://www.ecured.cu/index.php/Almacen_de_Datos.
10. Entorno Virtual de Aprendizaje. [En línea] 2010. [Citado el: 3 de 12 de 2010.] http://eva.uci.cu/file.php/624/2._Clases/Semana_1/2da_frecuencia/MApoyo/C2_DER.pdf
11. freedownloadmanager.org. [En línea] [Citado el: 1 de 12 de 2010.] [http://www.freedownloadmanager.org/es/downloads/Paradigma_Visual_para_UML_\[cuenta_de_Plataforma_de_Java_14715_p/](http://www.freedownloadmanager.org/es/downloads/Paradigma_Visual_para_UML_[cuenta_de_Plataforma_de_Java_14715_p/).
12. Gravatar. [En línea] [Citado el: 2 de 12 de 2010.] <http://www.gravatar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/>.
13. **Hernando Velasco, Roberto.** [En línea] 2009. [Citado el: 16 de 11 de 2010.] <http://www.rhernando.net/modules/tutorials/doc/bd/dw.html>.
14. **Ibarra, María de los Angeles.** *Procesamiento analítico en línea.* 2005.

15. **Inmon, William H.** *Building the Data Warehouse*. s.l. : Wiley, 2005.
16. **Kimball, R.** *El Juego de Herramientas del Almacén de Datos*. s.l. : John Wiley & Sons, 1996.
17. Kioskea. [En línea] [Citado el: 20 de 11 de 2010.]
<http://es.kioskea.net/contents/entreprise/datawarehouse-datamart.php3>.
18. **Martinez Hernandez, Katy L.** Scribd. [En línea] 2001. [Citado el: 10 de 1 de 2011.]
<http://www.scribd.com/doc/50422074/Taller-2-Topicos-by-Katy-Martinez>.
19. **Martinez, Rafael.** PostgreSQL-es. [En línea] 2010. [Citado el: 1 de 11 de 2010.]
http://www.postgresql-es.org/sobre_postgresql.
20. **Nader, Javier.** *Sistema de Apoyo Gerencial Universitario*. 2002.
21. **Najera Cano, Fernando P.** TortoiseSVN. [En línea] 2010. [Citado el: 7 de 12 de 2010.]
http://tortoisesvn.net/docs/release/TortoiseSVN_es/tsvn-introduction.html#tsvn-intro-about.
22. Office. [En línea] [Citado el: 16 de 2 de 2011.] <http://office.microsoft.com/es-es/excel-help/informacion-general-sobre-el-procesamiento-analitico-en-linea-olap-HP010177437.aspx>.
23. ONE. [En línea] [Citado el: 10 de 11 de 2010.] <http://www.one.cu/atribucionesyfunciones.htm>.
24. Onuva. [En línea] [Citado el: 18 de 1 de 2011.] <http://www.onuva.com/es/BI>.
25. Onuva. [En línea] [Citado el: 20 de 05 de 2011.]
http://www.onuva.com/files/pentaho/Manual_PRD.pdf.
26. **Padrón Torres, Liudmila.** Mailxmail. [En línea] [Citado el: 24 de 11 de 2010.]
<http://www.mailxmail.com/curso-almacenes-datos-importancia-estandar>.
27. **Pérez Lamancha, Beatriz.** *Tesis de Maestría en Informática: Proceso de Testing Funcional Independiente*. Montevideo : s.n., 2006.
28. **Pressman, R S.** Scribd. [En línea] [Citado el: 25 de 4 de 2011.]
<http://www.scribd.com/doc/7978336/Ingenieria-de-Software-Un-Enfoque-Practico-Pressman-5th-Ed>.
29. **Rubio Navarro, Jorge y Salinas, José Manuel.** *Almacén de Datos para el Análisis y Difusión de la Información Estadística del Turismo en España*. 2007.
30. Scribd. [En línea] [Citado el: 20 de 03 de 2011.] <http://es.scribd.com/doc/52203545/16/Metodologia-propuesta-por-Bill-Inmon>.
31. Sinnexus. [En línea] [Citado el: 1 de 12 de 2010.]
http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx.

32. Stratebi. [En línea] 2010. [Citado el: 2 de 12 de 2010.] http://www.stratebi.es/todobi/jun10/Comparativa_OSBI.pdf.
33. Ubicuos. [En línea] [Citado el: 1 de 11 de 2010.] <http://www.ubicuos.com/2010/05/08/como-instalar-pgadmin-3>.
34. **Vega Lebrún, Carlos.** Eumed. [En línea] [Citado el: 3 de 06 de 2011.] http://www.eumed.net/libros/2008a/351/Certificacion_de_la_Calidad.htm.
35. **Veliz Monteagudo, Mijail.** *Data Warehouse, más allá del almacén de datos.* La Habana: s.n., 2003.
36. **Vizete Naranjo, Michael Wladimir y Yela Shinin, Carlos Patricio.** *Análisis, diseño e implementación de un datamart para el Área de Sismología del Departamento de Geofísica de la Escuela Politécnica Nacional.* Quito: s.n., 2006.
37. **Zorrilla, Marta.** *Data Warehouse y OLAP.* Universidad de Cantabria: s.n., 2008.

Anexos

Anexo 1

Reporte: Indicadores de Servicios comunales.

Muestra los indicadores de Servicios comunales, por DPA, en un tiempo.

		Temporales		
		Q4		
		Medidas		
DPA	Indicadores	● Año Actual	● Año Anterior	● Crecimiento
<input checked="" type="checkbox"/> Cuba	Áreas verdes existentes	27690,7	26818,0	103,3
	Áreas verdes atendidas por comunales	15381,5	14821,2	103,8
	Áreas verdes con mantenimiento aplicado	375636,9	354681,7	105,9
	Fomento de áreas verdes	211,6	750,1	28,2
	Reconstrucción de áreas verdes	1240,6	1294,9	95,8
	Plantas sembradas en viveros	808733,0	789300,2	102,5
	Plantas logradas en viveros	739254,8	725579,3	101,9
	Áreas sembradas de flores	113,1	124,3	91,0
	Flores producidas	1641,1	1641,2	100,0
	Flores comercializadas	2359,1	2342,2	100,7
	Total de equipos trabajando	173,2	184,7	93,8
	Promedio de colectores trabajando	16,0	13,5	118,5
	Promedio de tractores	27,4	29,0	94,5
	Promedio de camiones abiertos trabajando	49,3	55,5	88,8
	Promedio de carros de tracción animal trabajando	549,8	600,1	91,6
	Promedio de pipas trabajando	6,6	6,3	104,8
	Promedio de bulldozers	5,1	3,9	130,8
	Área total de calles existentes	3337,9	3305,1	101,0

Glosario de Términos

A.A.G: Área de Análisis General.

A.A: Área de Análisis

BI: Business Intelligence (Inteligencia de Negocios).

CASE: Ingeniería de Software Asistida por Computadora

DATEC: Centro de Tecnología y Gestión de Datos

ETL: Extract, transform and load (Extracción, Transformación y Carga)

HOLAP: Hybrid Online Analytical Processing (Procesamiento Analítico en Línea Híbrido).

JDBC: Java Database Connectivity.

L.T: Libro de Trabajo

MOLAP: Multidimensional Online Analytical Processing (Procesamiento Analítico en Línea Multidimensional).

OLAP: Online Analytical Processing (Procesamiento Analítico en Línea)

ONE: Oficina Nacional de Estadísticas.

ROLAP: Relational OnLine Analytical Processing (Procesamiento Analítico en Línea Relacional).

SEN: Sistema Estadístico Nacional.

SIEC: Sistema de Información Estadística Complementaria.

SIEN: Sistema de Información de Estadísticas Nacional.

SIET: Sistema de Información de Estadísticas Territorial.

SIGOB: Sistema de Información de Gobierno.

SQL: Lenguaje de consulta estructurado, es un lenguaje de base de datos normalizado.¹

SSL: Secure Sockets Layer (Protocolo de capa de conexión segura).

¹ Tomado de: World Wide Web Consortium. <http://www.maestrosdelweb.com/editorial/tutsql1/>

TCP/IP: Protocolo de control de transmisión/Protocolo de Internet.²

TIC: Tecnologías de la Informática y las Comunicaciones.

UCI: Universidad de las Ciencia Informáticas.

UML: Lenguaje de Modelado Unificado.

Unicode: proporciona un número único para cada carácter, sin importar la plataforma, sin importar el programa, sin importar el idioma.³

XML: Lenguaje de Etiquetado Extensible.

² Tomado de: Kioskea.net. <http://es.kioskea.net/contents/internet/tcpip.php3>

³ Tomado de: The Unicode Consortium. <http://unicode.org/standard/translations/spanish.html>