

Universidad de las Ciencias Informáticas

Facultad 6



*“Sistema de Información de Gobierno. Mercado de
datos Comunicaciones”*

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

Autores

Laidy Jiménez Pérez

Leandro Carmenaty Pupo

Tutores

Ing. Roberto Tellez Ibarra

Ing. Esley León Valdés

Ciudad de La Habana, junio de 2011

Nuestra recompensa se encuentra en el esfuerzo y no en el resultado.

Un esfuerzo total es una victoria completa.

Mahatma Gandhi

DECLARACIÓN DE AUTORÍA

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

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

Laidy Jiménez Pérez

Leandro Carmenaty Pupo

Firma del Autor

Firma del Autor

Ing. Esley León Valdés

Ing. Roberto Tellez Ibarra

Firma del Tutor

Firma del Tutor

Datos del Contacto

Tutor: Ing. Esley León Valdés.

Profesión: Instructor Recién Graduado, DATEC, Universidad de las Ciencias Informáticas, Habana, Cuba.

Correo electrónico: elvaldez@uci.cu.

Tutor: Ing. Roberto Tellez Ibarra.

Profesión: Instructor Recién Graduado, DATEC, Universidad de las Ciencias Informáticas, Habana, Cuba.

Correo electrónico: rtibarra@uci.cu.

Dedicatoria

“A mis padres, y a mi familia de la UCI”

Leandro Carmenaty Pupo

“A mis padres, en especial a mi mamá y a todos los que me ayudaron a construir mi sueño”

Laidy Jiménez Pérez

Resumen

La necesidad de desarrollar un mercado de datos en el área de Comunicaciones para la Oficina Nacional de Estadísticas e Información, surge por las dificultades que existen en el momento de almacenar y/o gestionar los datos que se generan. Además, contribuye a que el trabajo de los especialistas sea más rápido y eficaz en el proceso de la toma de decisiones. Inicialmente se fundamentan los resultados de la investigación, haciendo referencia a los principales conceptos, herramientas y metodologías utilizadas en el desarrollo de la aplicación. Se realiza la descripción del negocio y el levantamiento de los requerimientos, lo que ayudan al análisis de los indicadores del mercado de datos y al diseño final. Se genera el modelo dimensional dándole paso al proceso más importante, la implementación del mercado de datos y la carga de los clasificadores. Por último, se valida la solución desarrollada aplicando el modelo V, las listas de chequeo y los casos de prueba. Con la culminación del trabajo se obtiene una versión del mercado de datos que posibilitará la mejor organización y análisis de las estadísticas que se manejan en el área de Comunicaciones de la Oficina Nacional de Estadísticas e Información.

Índice

<i>Introducción</i>	1
<i>Capítulo 1: Fundamentación teórica de los almacenes de datos</i>	4
1.1. Introducción	4
1.2. Almacenes de datos	4
1.2.1. Características de los almacenes de datos	5
1.2.2. Arquitectura de un almacén de datos	5
1.2.3. Mercado de datos	6
1.2.4. Componentes de un almacén de datos	7
1.2.5. Topología de un almacén de datos	7
1.2.6. Modelo multidimensional	8
1.2.7. Procesos OLAP y OLTP	8
1.3. Inteligencia de negocio	11
1.4. Integración de datos	10
1.5. Metodología de desarrollo	13
1.6. Herramientas	15
1.6.1. Gestor de base de datos	15
1.6.2. Herramientas Case de Modelado	16
1.6.3. Herramienta software para ETL	17
1.6.4. Herramienta de software de inteligencia de negocios	17
1.7. Conclusiones del capítulo	18
<i>Capítulo II: Análisis y diseño del mercado de datos Comunicaciones</i>	19
2.1. Introducción	19
2.2. Análisis de la solución	19

2.2.1.	Definición del negocio	19
2.2.2.	Temas de análisis	20
2.2.3.	Reglas del negocio	20
2.2.4.	Requerimientos de información	21
2.2.5.	Requerimientos funcionales	23
2.2.6.	Requerimientos no funcionales	24
2.2.7.	Diagrama de casos de uso del sistema	25
2.3.	Diseño de la solución	26
2.3.1.	Matriz multidimensional	26
2.3.2.	Modelo de datos	27
2.4.	Política de respaldo y recuperación	30
2.5.	Esquema de seguridad	31
2.6.	Conclusiones del capítulo	32

<i>Capítulo III: Implementación del mercado de datos Comunicaciones</i>	33
3.1. Introducción	33
3.2. Implementación de la base de datos	33
3.2.1. Estructura de los datos	33
3.2.2. Esquemas y Tablas	34
3.2.3. Usuarios y roles	35
3.3. Implementación del subsistema de integración	36
3.3.1. Arquitectura del subsistema de integración	36
3.3.2. Perfilado de Datos	36
3.3.3. Extracción, Transformación y Limpieza de los Datos	37
3.3.4. Carga	37
3.3.5. Trabajo	38
3.4. Implementación de subsistema de visualización	38
3.4.1. Esquema OLAP	38
3.4.2. Navegación de la capa de visualización	40
3.4.3. Configurar la seguridad de los usuarios	44
3.5. Conclusiones del capítulo	45
<i>Capítulo 4: Pruebas del mercado de datos Comunicaciones</i>	46
4.1. Introducción	46
4.2. Pruebas	46
4.3. Lista de Chequeo	48
4.4. Casos de prueba	49
4.5. Prueba de sistema	50
4.6. Prueba de aceptación	51

4.7. Conclusiones del capítulo	51
<i>Recomendaciones</i>	53
<i>Referencias bibliográficas</i>	54
<i>Anexos</i>	58
Anexo 1: Diseño del Modelo Lógico	60
Anexo 2: Análisis de todas las tuplas del perfilado de datos	60
Anexo 3: Análisis de los valores de tipo string del perfilado de datos	61
Anexo 4: Análisis de los valores de tipo numérico del perfilado de datos	61
Anexo 5: Lista de Chequeo	62
<i>Glosario de términos</i>	65

Índice de tabla

Tabla 1 Descripción de los actores.....	26
Tabla 2 Descripción de los casos de usos.....	26
Tabla 3 Matriz BUS	27
Tabla 4 Permiso del usuario en la base de datos	31
Tabla 5 Permiso de los roles en la aplicación.....	31
Tabla 6 Acceso a los elementos de la aplicación.....	31
Tabla 7: Estructura de nombrado	34
Tabla 8 Esquema de la base de datos	35

Índice de Figuras

Figura 1: Diagrama de caso de uso	¡Error! Marcador no definido.
Figura 2 Modelo de datos	28
Figura 3 Arquitectura de Integración	36
Figura 4 Transformación de la carga de los hechos que se muestran con información mensual.	37
Figura 5 Trabajo de la carga de los hecho	38
Figura 6 Diseño del esquema	39
Figura 7: Diseño de las dimensiones	¡Error! Marcador no definido.
Figura 8: Cubo año radiodifusión	40
Figura 9 : Cubo mes prensa.....	40
Figura 10: Cubo semestre prensa.....	40
Figura 11 : Ejemplo del reporte Distribución de Prensa.....	44
Figura 12 Adicionar rol analista	¡Error! Marcador no definido.
Figura 13 Adicionar rol administrador	¡Error! Marcador no definido.
Figura 14: Modelo v	47
Figura 15 Comportamiento de los indicadores	49
Figura 16: No conformidades de los casos de prueba	50

Introducción

La acelerada evolución de las tecnologías ha venido a desempeñar un papel muy importante en el mundo empresarial, donde las organizaciones buscan soluciones tecnológicas que les brinden mayor eficiencia, flexibilidad y rapidez, con el fin de obtener una ventaja competitiva que les permita permanecer en la élite del mercado mundial.

La estadística tiene un papel muy importante en el mercado mundial, ya que es la disciplina que se encarga de la recopilación, organización, presentación, análisis e interpretación de los datos con el fin de realizar una mejor toma de decisiones. Como resultado de la generación de grandes volúmenes de información se ha visto impedido el almacenamiento y análisis estadístico de los datos.

Cuba se ha trazado la tarea de informatizar los distintos sectores de la sociedad y diversas ramas de la economía nacional, con el fin de alcanzar un mayor perfeccionamiento empresarial. En este sentido, la Universidad de las Ciencias Informáticas (UCI) posee un papel fundamental, ya que desde su fundación ha sido uno de los pilares fundamentales de la industria del software cubano.

Durante los últimos años diversas instituciones cubanas, en conjunto con la UCI, han venido desarrollando sistemas encaminados a la informatización de distintos sectores. Un ejemplo de estas instituciones es la Oficina Nacional de Estadísticas e Información (ONEI), entidad encargada de recopilar, analizar y difundir los datos provenientes de todos los sectores estatales del país.

La ONEI como parte de su funcionamiento procesa un gran volumen de datos, los cuales son recopilados por diferentes modelos estadísticos que provienen de diversas ramas de la economía. Dentro de ellas, se encuentra el departamento de Comunicaciones, que manipula información de varios indicadores como: la prensa, la telefonía y la radiodifusión, entre otros.

Los formatos en que se encuentra almacenada la información son de difícil acceso, lo que provoca que el proceso de recuperación y elaboración de informes sea más costoso en tiempo y esfuerzo, aumentando así las probabilidades de que se produzcan errores estadísticos. Al mismo tiempo, los sistemas informáticos existentes no son capaces de dar seguimiento al proceso de toma de decisiones, ni de realizar reportes actualizados y flexibles. Por otra parte la información digital se tiene en documentos en formato Microsoft Office Excel (xls), Microsoft Office Word (doc), y otros, lo que requiere de un alto nivel de conocimiento tanto informático como estadístico para realizar el análisis de la misma. Además, se generan ficheros anualmente lo que aumenta el volumen de información por lo que el proceso de toma de decisiones se hace más engorroso y lento.

Debido a esta situación surge como ***problema de la investigación***: ¿cómo contribuir a la toma de

decisiones en el área de las Comunicaciones del Sistema de Información del Gobierno?

Teniendo como **objeto de estudio** los almacenes de datos, enmarcado en **el campo de acción** el mercado de datos para el área Comunicaciones del Sistema de Información del Gobierno.

Para dar solución al problema de la investigación se plantea como **objetivo general**: Desarrollar el mercado de datos Comunicaciones del Sistema de Información de Gobierno que contribuya a la toma de decisiones.

A partir del análisis realizado, se definen como **objetivos específicos**:

- ✓ Realizar el análisis y diseño del mercado de datos del área Comunicaciones.
- ✓ Implementar el mercado de datos del área Comunicaciones.
- ✓ Validar el mercado de datos del área Comunicaciones.

Para dar cumplimiento a los objetivos anteriormente mencionados se plantean como **tareas de la investigación**:

- ✓ Caracterización de las metodologías, herramientas y tecnologías a utilizar en el desarrollo de almacenes de datos.
- ✓ Levantamiento de requerimientos.
- ✓ 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 de la base de datos.
- ✓ Implementación del subsistema de integración.
- ✓ Implementación del subsistema de visualización.
- ✓ Aplicación de las listas de chequeo.
- ✓ Aplicación de los casos de pruebas.

El presente documento está estructurado en cuatro capítulos, en los que se describen los métodos y procedimientos a seguir para dar cumplimiento a los objetivos trazados.

En el Capítulo 1 **“Fundamentación teórica”**: se realiza un estudio del negocio para el desarrollo de un mercado de datos para el área Comunicaciones, así como de las principales definiciones, metodologías y herramientas que se utilizan para su desarrollo.

En el Capítulo 2 **“Análisis y diseño del mercado de datos Comunicaciones”** se hace referencia al flujo de trabajo de análisis y diseño, en el cual se definen los requerimientos del sistema, se diseña el diagrama de casos de uso del sistema; así como el subsistema de integración, almacenamiento y visualización.

En el Capítulo 3 **“Implementación del mercado de datos Comunicaciones”** se implementa el subsistema de implementación, almacenamiento y visualización, con el objetivo de darle solución a los requerimientos del sistema.

En el Capítulo 4 **“Pruebas del mercado de datos Comunicaciones”** se valida el mercado de datos a través de las listas de chequeo y los casos de prueba.

Capítulo 1: Fundamentación teórica

1.1. Introducción

En el capítulo se realiza un estudio del negocio para el desarrollo del mercado de datos Comunicaciones, así como de las principales definiciones, metodologías y herramientas que se utilizan para su desarrollo.

1.2. Almacenes de datos

El concepto de almacén de datos o datawarehouse (DWH), se comienza a utilizar a mediados de la década de 1980 cuando investigadores de la International Business Machines (IBM), se dieron a la tarea de proveer un modelo de arquitectura del flujo de datos de los sistemas operacionales a los ambientes de soporte de decisiones, ya que en su ausencia se requería una enorme redundancia de los datos, para respaldar los sistemas de toma de decisiones.

Existen dos puntos de vista, considerados entre los más importantes a la hora de implementar un almacén de datos, planteados por dos de las figuras precursoras de la disciplina inteligencia de negocios: Bill Inmon y Ralph Kimball.

Ellos plantean lo siguiente:

✓ **Bill Inmon:** el almacén de datos se considera sólo una parte de un sistema de inteligencia de negocios. Una empresa tiene un DWH, y los mercados de datos o Data Marts (MD) tienen como fuente de información ese almacén. Esta aproximación también es conocida como "Top-Down".(1)

✓ **Ralph Kimball:** el DWH es la unión de todos los mercados de datos o Data Marts generados en una empresa. La información siempre se almacena en un modelo multidimensional. Otra forma de denominar esta aproximación es como "Bottom-up".(2)

Estos fueron los principios que sirvieron para implementar los almacenes de datos en los orígenes, y que aún continúan siendo los más utilizados por su conceptualización. Han surgido nuevos paradigmas, que con la ayuda de las nuevas tecnologías han evolucionado estos puntos de vista, pero todos parten del estudio de los dos anteriormente expuestos.

En resumen, un almacén de datos es una colección de datos orientados a temas, integrado, no volátil, de tiempo variante, que se usa para el apoyo del proceso de toma de decisiones gerenciales.

1.2.1. Características de los almacenes de datos

De acuerdo con la definición de los almacenes de datos estos se caracterizan por ser: (3)

- ✓ **Orientados a temas:** el principal objetivo de este punto, es lograr un mejor entendimiento por parte de los usuarios finales. Desde el entorno operacional se va filtrando la información, hasta quedarse sólo la que ayuda en el proceso de generación de conocimiento.
- ✓ **Integrado:** deben eliminarse las inconsistencias existentes en las fuentes de datos, para lograr una integración sólida a la hora de cargarlos al almacén.
- ✓ **No volátil:** se conserva la información y sólo se actualiza.
- ✓ **Variante en el tiempo:** el almacén se carga con valores variables en el tiempo que permiten comparaciones, y entre otras cosas, análisis de tendencias.

Otra característica del DWH, es que contiene metadatos, es decir, datos sobre los datos. Los metadatos permiten saber la procedencia de la información, su periodicidad de actualización, su fiabilidad y forma de cálculo.

1.2.2. Arquitectura de un almacén de datos

Los almacenes de datos presentan la siguiente arquitectura: (4)

- ✓ **Sistemas Operacionales:** es la fuente de datos donde se nutre el almacén. Pueden estar organizados en archivos indexados, bases de datos jerárquicas o en sistemas de base de datos relacionales.
- ✓ **Extracción, Transformación y Carga:** el primer paso es la extracción de los datos de los sistemas operacionales; luego se realizan las transformaciones necesarias para que sólo quede la información que el cliente desee y en el formato correspondiente; por último esta información “limpia” es cargada al almacén.
- ✓ **Creación de los metadatos:** es el proceso de crear la información sobre los datos que van a ser almacenados. Describe cómo se almacenan, se integran y se transforman estos datos.
- ✓ **Acceso de usuarios finales:** los usuarios finales acceden a la información del DWH a través de una interfaz gráfica. Estas pueden incluir software de consultas, generadores de reportes,

procesamientos analíticos en línea, entre otras funciones, dependiendo de lo que desee el cliente.

- ✓ **Plataforma del almacén:** casi siempre se trata de un sistema de base de datos relacional, dependiendo del volumen de información. De ser muy grande, se puede utilizar un bloque de servidores UNIX con multiprocesadores simétricos (SMP)¹ o un servidor con procesadores paralelos masivos (MPP)².

1.2.3. Mercado de datos

Los mercados de datos son bases de datos departamentales, especializadas en el almacenamiento de información de un área de negocio específica. Se caracterizan por disponer de una estructura óptima de los datos, para analizar la información al detalle, desde todas las perspectivas que afecten a los procesos de dicha área. Los datos que contienen pueden ser explorados, agrupados y propagados de diversas formas, y así los usuarios pueden realizar su explotación de acuerdo con sus necesidades. Además de que estos están orientados a un tema específico, lo que implica que se obtengan de un menor número de fuentes, y así el tiempo de desarrollo sea menor. Un mercado de datos puede ser alimentado desde los datos de un almacén, o integrar por sí mismo un compendio de distintas fuentes de información, son fáciles de entender y navegar. (5)

Características de los mercados de datos

Se caracterizan por: (2)

- ✓ Se centran en los requerimientos de los usuarios asociados a un departamento o área de negocios concreta.
- ✓ No contienen datos operacionales detallados.
- ✓ Son sencillos a la hora de utilizarlos y comprender sus datos.
- ✓ El diseño del mercado de datos, se realiza siguiendo una estructura consistente para las necesidades de los usuarios.
- ✓ Contiene sólo el mínimo de información histórica.

¹ SMP es un tipo de arquitectura de ordenadores en que dos o más procesadores comparten una única memoria central.

² MPP es una arquitectura de ordenadores donde la memoria está distribuida entre los procesadores.

- ✓ Facilita el acceso de las herramientas de consulta y divide los datos para controlar mejores accesos, debido a que frecuentemente hay grupos de usuarios que sólo acceden a un subconjunto concreto de datos.

1.2.4. Componentes de un almacén de datos

El almacén de datos está formado por diversos componentes, los cuales se describen a continuación:
(6)

- ✓ **Tablas de Hechos:** representan la ocurrencia de un determinado proceso dentro de la organización. Es en ellas donde se almacenan las medidas numéricas del negocio, donde cada medida es tomada de la intersección de las dimensiones que la definen, idealmente están compuestas por valores numéricos, continuamente evaluados y aditivos.
- ✓ **Tablas de Dimensiones:** contienen los detalles de los valores que se encuentran asociados a las tablas de hechos. Están compuestas por atributos dimensionales que son fundamentalmente textos descriptivos, estos desempeñan un papel determinante, puesto que son la fuente de gran parte de las necesidades que deben cubrirse. Además, sirven de restricciones en la mayoría de las consultas que realizan los usuarios.
- ✓ **Medida:** es un atributo numérico de un hecho, que constituye el comportamiento del negocio relativo a la dimensión. Representan los valores que son analizados y las bases de los cálculos realizados por los usuarios.
- ✓ **Jerarquía:** una jerarquía representa una relación lógica entre los datos de una dimensión, y su representación se encuentra en su interior. Pueden existir varios niveles con una relación "1-n" entre atributos consecutivos de un nivel superior y uno inferior. Es una clasificación u ordenación de abstracciones.

1.2.5. Topología de un almacén de datos

Existen varias topologías o esquemas para el modelado de los datos en un almacén, estas son: (7)

- ✓ **Esquema de Estrella:** su estructura es similar a la de una estrella. La tabla de hecho está en el centro de la estrella, y están relacionadas con ella de forma radial todas las tablas de dimensiones, las cuales no se relacionan entre sí. En este esquema no existen caminos

alternativos en las dimensiones, es decir, las tablas de dimensiones tienen sólo la conexión a la tabla de hecho.

- ✓ **Esquema de Copo de Nieve:** es parecido al de estrella, pero existen jerarquías en las dimensiones. Aparecen cuando las dimensiones se normalizan y las tablas de hechos ya no serían las únicas que se relacionan con las dimensiones ya que existen relaciones en las dimensiones. Existe el copo de nieve completo que es cuando todas las tablas están normalizadas y el copo de nieve parcial que es cuando sólo aparecen algunas normalizadas.
- ✓ **Esquema de Constelación:** para cada esquema estrella o copo de nieve en el almacén de datos es posible construir un esquema de constelación de hechos. Este esquema es más complejo que las otras arquitecturas, debido a que contiene múltiples tablas de hechos. Con esta solución las tablas de dimensiones, pueden estar compartidas en más de una tabla de hechos. La debilidad que presenta este tipo de arquitectura, es que cuando el número de tablas vinculadas aumenta, la arquitectura puede llegar a ser muy compleja y difícil de mantener.

1.2.6. Modelo multidimensional

El Modelo de Entidad Relación (MER) es un lenguaje para el modelado de datos de sistemas de información. Está basado en una percepción del mundo real, que consta de un conjunto de objetos básicos llamados entidades, con sus atributos y de las interrelaciones que existen entre estos objetos, donde las entidades se convierten en una tabla física en la base de datos operacional, y las interrelaciones son asociaciones o conexiones que existen entre dos o más entidades.(7)

El Modelo Multidimensional (MM) contiene la misma información que un MER pero la manera de organizarla es de forma diferente, con el objetivo de garantizar mayor comprensión por parte del usuario y menor tiempo de respuesta al ejecutar las consultas. Mediante su utilización se disminuye la cantidad de tablas y relaciones entre ellas, lo que agiliza el acceso a los datos. Dada estas características, para la implementación del mercado de dato se utiliza el modelo multidimensional.

1.2.7. Procesos OLAP y OLTP

Los sistemas orientados al Proceso de Transacciones En Línea (OLTP) se caracterizan por un gran número de transacciones que pueden involucrar operaciones de: inserción, eliminación y actualización. En estos sistemas los datos son actuales, además de que no son necesariamente uniformes, por lo que existe falta de compatibilidad. Son principalmente orientados a servicios o a servicios web, su

arquitectura es de tipo cliente-servidor y las principales aplicaciones son en cajeros automáticos, comercio electrónico, supermercados, entre otros.

Estos sistemas son utilizados en las bases de datos operacionales y no en los almacenes de datos, por lo que se necesita un sistema que les permita a las compañías convertir la información en conocimiento, con el fin de obtener apoyo en el proceso de toma de decisiones. Es aquí donde surge la interrogante ¿cómo obtener la información precisa, fiable y rápida? como respuesta a esta interrogante se introduce el término de Procesamiento Analítico en Línea (OLAP).

El origen de la tecnología OLAP se acuñó por el investigador Ted Codd en 1993, el cual se conoce como “el padre de las bases de datos relacionales”. OLAP se basa en el modelo multidimensional de datos, es decir, clasifican los datos en diferentes dimensiones y pueden ser vistas unas con otras en distintas combinaciones, para obtener diversos análisis de los datos lo que le permite a los usuarios seleccionar y extraer la información desde diferentes puntos de vista. Esta tecnología tiene como principal característica que es muy rápida a la hora de ejecutar sentencias de tipo SELECT lo que permite mostrar la información de forma ágil, además de que procesa las transacciones de un negocio en tiempo real.

Existen varios modos de procesamiento analítico en línea (OLAP), dentro de los cuales se destacan:

- ✓ **MOLAP (OLAP multidimensional):** estos sistemas usan una base de datos multidimensional, en la cual es almacenada la información que puede ser visualizada en varias dimensiones de análisis. Su arquitectura está dividida en dos niveles: las bases de datos multidimensionales y el motor analítico. Su característica principal es que los datos son pre-calculados y luego son almacenados con el objetivo de optimizar el tiempo de respuesta. Estos sistemas posibilitan hacer cálculos más complicados y muchos utilizan técnicas de comprensión de los datos, para disminuir el espacio de almacenamiento.(8)
- ✓ **ROLAP (OLAP Relacional):** son sistemas en los cuales los datos se encuentran almacenados en una base de datos relacional. Utiliza una arquitectura de tres niveles: el nivel de base de datos que usa bases de datos relacionales para el manejo, acceso y obtención del dato, el nivel de aplicación que es el motor que ejecuta las consultas multidimensionales de los usuarios, el cual se integra con el nivel de presentación a través del cual los usuarios realizan los análisis OLAP. Estos sistemas son capaz de usar datos pre-calculados si estos están disponibles y de generar dinámicamente los resultados. Además que accede directamente a los datos del almacén de datos, y soporta técnicas de optimización de accesos para acelerar las consultas.

Estas optimizaciones son, entre otras, particionado de los datos al nivel de aplicación, soporte a la des-normalización y uniones múltiples.(9)

- ✓ **HOLAP (OLAP híbrida):** este sistema combina las arquitecturas ROLAP y MOLAP para brindar una solución con las mejores características de ambas: desempeño superior y gran escalabilidad. Además de que mantiene los volúmenes más grandes en la base de datos relacional y las agregaciones en un almacén MOLAP separado.

Para la construcción del mercado de datos Comunicaciones se utilizará el sistema ROLAP ya que maneja grandes volúmenes de datos. Los datos se almacenan en una base de datos relacional estándar que puede ser accedida por cualquier herramienta de generación de informes de Lenguaje de Consulta Estructurado (SQL). Estas herramientas no tienen que ser necesariamente de tipo OLAP.

1.3. Integración de datos

El objetivo fundamental de la integración de datos, es permitir el rápido desarrollo de aplicaciones que requieran diversas fuentes de datos. Este objetivo encierra un gran desafío, ya que abarca desde identificar la mejor fuente de datos hasta la creación de una adecuada interfaz para presentarlos. Consta de cuatro fases principales:

- ✓ **Entendimiento:** en esta fase la primera tarea es buscar relación entre los datos y su principal objetivo es entenderlos y comprenderlos.
- ✓ **Normalización:** aprovecha la fase anterior para determinar cuál es la mejor forma de integrar esos datos, limpiarlos o repararlos.
- ✓ **Especificación:** en esta fase se documenta todo lo relacionado con el proceso de integración, es la parte de configuración de un motor de ejecución para realizar la integración deseada.
- ✓ **Ejecución:** es donde realmente ocurre la integración, que puede realizarse a través de la materialización, federación o indexación. (10)

Las principales tecnologías para realizar una integración de datos son:

- ✓ **Herramienta de Integración Funcional (EAI):** para soluciones simples, ideal para cuando el número de aplicaciones que deben intercambiar información es reducido y estático. Son un

conjunto de adaptadores y enrutadores, que mueven transacciones en forma de mensajes entre aplicaciones interconectadas.

- ✓ **Herramientas de Extracción Transformación y Carga (ETL):** las bases de datos pueden ser heterogéneas y son integradas y materializadas en un almacén de datos. El almacén se puede ver a su vez, como almacenes más pequeños interrelacionados, cada uno relacionado a un área específica, por lo que pueden estar separados físicamente y se debe garantizar su sincronización. Generalmente la actualización se realiza por las noches, para aprovechar el poco tráfico en la red y la frecuencia puede ser establecida por el administrador.
- ✓ **Captura, Transformación y Flujo (CTF):** se realiza un intercambio directo entre bases de datos, donde las transformaciones son muy elementales comparadas con ETL.
- ✓ **Sistema para el Manejo de Datos Federados (EII):** se puede tener acceso a diferentes fuentes de datos sin tener que cargarlos primero. Se enfoca más a las consultas, y los datos necesitan estar actualizados.

Se selecciona la tecnología de ETL por las funcionalidades que brinda en el momento de la integración y, principalmente, en la limpieza de los datos, muy necesarias en el proceso que se desarrolla.

1.4. Inteligencia de negocio

Las grandes empresas tienen grandes volúmenes de datos disponibles, y con las presiones competitivas en los mercados actuales necesitan emplear sus propios datos, para comprender su negocio y mantenerse en un alto nivel competitivo. Hoy en día las organizaciones, toman decisiones estratégicas, que generalmente están basadas en enormes volúmenes de información, registrados en bases de datos operacionales o de otros tipos de fuentes de datos. La recopilación y análisis de la información, dado su carácter heterogéneo y su volumen, se convierten usualmente en un problema para las entidades. Todo proceso de toma de decisiones implica aceptar un riesgo, lo importante es cómo minimizarlo, a partir de esto surge el término inteligencia de negocio o Business Intelligence (BI). La inteligencia de negocios es la habilidad para transformar los datos en información, y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones.

Los datos son la materia bruta, la unidad mínima, los cuales por si solos son irrelevantes. En el momento que se le atribuye algún significado pasan a convertirse en información, la cual no es más que un conjunto de datos procesados que tienen un significado. Una vez obtenida la información es interpretada por los especialistas. (5)

También se podría definir desde el punto de vista tecnológico, donde la inteligencia de negocios es el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales, los cuales no se encuentren integrados en información estructurada, para su explotación directa o para su análisis y conversión en conocimiento, dando así soporte a la toma de decisiones sobre el negocio. (5)

Existen muchas maneras de analizar la información, y por este motivo existen un conjunto de soluciones que resuelven las diferentes necesidades analíticas.

Características de la Inteligencia de Negocio

- ✓ **Apoyo en la toma de decisiones:** se busca ir más allá en la presentación de la información, de manera que los usuarios tengan acceso a herramientas de análisis que les permitan seleccionar y manipular sólo aquellos datos que les interesen.
- ✓ **Orientación al usuario final:** se busca independencia entre los conocimientos técnicos de los usuarios y su capacidad para utilizar estas herramientas.

Ventajas de la inteligencia de negocios

- ✓ Las tecnologías de BI permiten reunir, normalizar y centralizar toda la información de la empresa, mediante un almacén de datos, permitiendo así su explotación. De esta forma, los departamentos comerciales, operativos y financieros basan las decisiones estratégicas en la misma información.
- ✓ Permite descubrir información no evidente para las aplicaciones, es decir, en las aplicaciones de gestión pueden contener pautas de comportamiento, tendencias, evoluciones del mercado, cambios en el consumo o en la producción, que resulta prácticamente imposible reconocer y con las herramientas de BI se pueden descubrir.
- ✓ Posibilita que una empresa aprenda de su historia y de sus mejores prácticas, y que pueda evitar tropezarse de nuevo con los mismos errores del pasado.
- ✓ Permite crear, manejar y monitorear las métricas y los objetivos estratégicos propuestos para poder detectar a tiempo las desviaciones, adoptando las acciones oportunas para corregirlas.
- ✓ Ayuda a las compañías a reducir los ciclos de desarrollo de productos, agilizar operaciones y mejorar relaciones con clientes y proveedores.

La inteligencia de negocios forma parte hoy en día en una solución completa que responde a varias preguntas: ¿qué está ocurriendo?, ¿por qué ocurre?, ¿qué ocurriría?, ¿qué debería hacer?, ¿qué camino se debe seguir?, la cual se ha convertido en un factor estratégico para las compañías, generando una potencial ventaja competitiva en el mercado mundial, proporcionando información privilegiada para responder a los problemas del negocio, entrada a nuevos mercados, promociones u ofertas de productos, control financiero, optimización de costos, planificación de la producción, análisis de perfiles de clientes y la rentabilidad de un producto concreto.

1.5. Metodología de desarrollo

Las metodologías de desarrollo de software son un conjunto de procedimientos, técnicas, herramientas y un soporte documental que ayuda a los desarrolladores a realizar un nuevo producto de software. Estas surgieron a raíz de la necesidad de controlar y documentar proyectos cada vez más complejos y con mayores requerimientos de seguridad y fiabilidad. Su función está guiada a mejorar aplicaciones que conduzcan a una mejor calidad y mantener un proceso de desarrollo controlado.

Existen diversas metodologías para la construcción de almacenes de datos, donde cada una está orientada a la optimización del rendimiento, y a la visión de los principales procesos que se deben tener en cuenta para construir un almacén de datos flexible y dinámico.

Existen dos personas que han marcado un gran impacto en la implementación de los almacenes y han sido guía para la elaboración de otras metodologías, estos son Ralph Kimball y Bill Inmon.

Bill Inmon se basa en un enfoque descendente (top-down), el cual propone construir primero el almacén de datos y a partir de este los mercados de datos. Plantea la creación de un repositorio de datos corporativo como fuente de información consolidada, persistente, histórica y de calidad. Al ser construido descendentemente, los mercados de datos se nutren de los almacenes de datos corporativos, convirtiéndose en un complejo empresarial de base de datos relacionales. Inmon afirma que la creación de una base de datos relacional con una leve normalización, necesita ser la base para los mercados de datos. Por lo que los mismos se crean a partir de la arquitectura relacional de los datos corporativos.

Por otro lado Ralph Kimball se basa en un enfoque ascendente (bottom-up), plantea que se debe crear por cada departamento un conjunto de mercados de datos independientes, y orientados a los temas que estén relacionados con él. Y el almacén de datos es la unión de todos los mercados de datos de una entidad. Entre sus principales características se encuentran que divide el mundo de inteligencia de negocios entre los hechos y las dimensiones, lo cual hace que esta metodología sea eficaz y que conduzca a una solución completa en un corto período de tiempo. Además que 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.

Existen diversas metodologías que pretenden dar una propuesta ideal para el desarrollo de almacenes de datos. El Centro de Tecnologías y Gestión de Datos (DATEC) de la Universidad de las Ciencias Informáticas, tomó como base el ciclo de vida de Kimball por tener bien definidas las etapas, actividades, artefactos y roles; además de que el flujo de datos parte de la fuente, de la cual se realiza la construcción de los distintos mercados de datos y luego la integración de los mismos en el almacén. Por otra parte los conceptos de hechos y dimensiones juegan un papel muy eficaz en el proceso de toma de decisiones, lo que proporciona mayor agilidad en el desarrollo.

A partir de la metodología seleccionada como base se conformó un modelo para el desarrollo de los almacenes. Este modelo cuenta con cinco grupos de trabajos, los mismos son: Análisis, Almacenes, ETL, BI y por último el de Dirección, donde cada uno realiza actividades específicas y bien delimitadas según sus responsabilidades dentro de un proyecto.

El proceso de desarrollo de esta modelo está compuesto por ocho flujos de trabajos, los cuales son:

- ✓ **Estudio Preliminar o Planeación:** se realiza un estudio de lo qué se desea construir y qué condiciones existen para su desarrollo y montaje.
- ✓ **Requerimiento:** se definen las reglas del negocio y necesidades de la información, además de un levantamiento detallado de cada una de las distintas fuentes de datos a integrar.
- ✓ **Arquitectura y Diseño:** se definen las estructuras de almacenamiento, se diseñan las reglas de extracción, transformación y carga, así como la arquitectura de información que regirá el desarrollo de la solución.
- ✓ **Implementación:** se lleva a cabo el diseño físico del repositorio de datos, se ejecutan las reglas de extracción, transformación y carga, haciendo los ajustes para integrar la información necesaria y se configuran e implementan las herramientas de BI para obtener los reportes, gráficos, mapas y otros que cubran los requerimientos.
- ✓ **Prueba:** se realizan varias pruebas comenzando por pruebas de unidad, pruebas de integración y sistema, hasta pruebas de aceptación con el cliente.
- ✓ **Despliegue:** se despliega la solución en el entorno real y en correcto funcionamiento.

- ✓ **Soporte y Mantenimiento:** son los servicios que se brindan ya una vez implantada la solución.
- ✓ **Gestión y Administración del Proyecto:** es donde se controla, gestiona y chequea todo el desarrollo, los gastos, las utilidades, los recursos, las adquisiciones, y demás actividades relacionadas con la gestión de proyecto.

Este modelo permite organizar el proceso de desarrollo de los almacenes de datos e inteligencia de negocios, la cual define grupos de trabajo, artefactos, roles, responsabilidades, entre otros aspectos que contribuyen a una mayor rapidez y calidad en el desarrollo de las soluciones.

1.6. Herramientas

Se realiza una descripción de las herramientas propuestas por el centro DATEC para la construcción del mercado de datos Comunicaciones.

1.6.1. Gestor de base de datos

Postgre 8.4: es un sistema de gestión de bases de datos objeto-relacional (ORDBMS) basado en el proyecto POSTGRES, de la Universidad de Berkeley, que ha venido siendo desarrollado de varias formas desde la década de 1980. Es una derivación libre (OpenSource)³, y utiliza el lenguaje SQL. Fue el pionero en muchos de los conceptos existentes en el sistema objeto-relacional actual. Incluye características de la orientación a objetos, como puede ser la herencia, tipos de datos, funciones, restricciones, disparadores, reglas e integridad transaccional. Es un potente motor de bases de datos, que tiene prestaciones y funcionalidades equivalentes a otros gestores comerciales. (11)

Características

- ✓ Tienen la función de mantener la integridad y consistencia en la BD.
- ✓ Soporta operadores, funciones, métodos de acceso y tipos de datos definidos por el usuario.
- ✓ Incorpora funciones de diversa índole: manejo de fechas, geométricas, orientadas a operaciones con redes.

³ OpenSource (Código abierto) es un software libre que su código está disponible.

✓ **Ventajas**

- ✓ El código fuente de PostgreSQL está disponible bajo una licencia de código abierto liberal. Esta licencia le da la libertad para usar, modificar y distribuir PostgreSQL en cualquier forma, abierto o de código cerrado.
- ✓ PostgreSQL ha sido diseñado y creado para tener un mantenimiento y ajuste mucho menor que otros productos, conservando todas las características, estabilidad y rendimiento.
- ✓ Es multiplataforma, está disponible en las plataformas UNIX, Windows y Linux.
- ✓ Ejecuta los procedimientos almacenados en más de una docena de idiomas de programación, como Java, Perl, Python, Ruby, Tcl, C / C + +.
- ✓ Se integra perfectamente con sistemas OLAP para mostrar los reportes.

PgAdminIII 1.10: es una herramienta de código abierto muy avanzada y se utiliza para administrar y desarrollar bases de datos en PostgreSQL. Se diseña para responder a las necesidades de la mayoría de los usuarios, desde escribir simples consultas SQL, hasta desarrollar bases de datos complejas. La interfaz gráfica soporta todas las características de PostgreSQL y facilita la administración. Está disponible en más de una docena de lenguajes y para varios sistemas operativos, incluyendo Microsoft Windows, Linux, FreeBSD, Mac OSX y Solaris. (12)

1.6.2. Herramientas Case de Modelado

Visual Paradigm for UML 6.4: es una herramienta que soporta el ciclo de vida completo del desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue. Utiliza el Lenguaje de Modelado Unificado (UML), lo que ayuda a una rápida construcción de aplicaciones de calidad, y a un menor costo. Permite dibujar todos los tipos de diagramas de clases, código inverso, generar código desde diagramas y generar documentación. Posee abundante documentación y tutoriales. Ofrece un completo conjunto de herramientas de los equipos de desarrollo de software, necesario para la captura de requerimientos, la planificación de programas y de controles, la clase de modelado, modelado de datos. (13)

Características

- ✓ Es multiplataforma, está disponible en las plataformas Windows y Linux.

- ✓ No sólo realiza diagramas de UML, sino también posibilita modelar hardware, dominio-específico, el software.
- ✓ Soporta ingeniería inversa y generación de código.
- ✓ Permite la generación de bases de datos (transformación de modelos Entidad-Relación en tablas de bases de datos).

1.6.3. Herramienta software para ETL

Pentaho Data Integration 4.1.0 (Kettle): herramienta utilizada para el proceso de Extracción, Transformación y Carga de datos, de código abierto, que reúne un conjunto de componentes que permiten modelar y ejecutar transformaciones sobre flujos de datos. Ofrece una presentación gráfica, amigable y sólida, es escalable. (14)

DataCleaner 1.5.3: es una aplicación de código abierto para el perfilado, validación y comparación de datos. Estas actividades ayudan a administrar y monitorear la calidad de la información, en el orden de garantizar la utilidad de la información y de la aplicación para el negocio. Es una alternativa libre para la metodología de administración de datos, para proyectos de almacenes de datos, búsquedas estadísticas, para actividades de preparación de ETL. (15)

1.6.4. Herramienta de software de inteligencia de negocios

Pentaho Schema Workbench 3.2: es una herramienta para el desarrollo y prueba de esquemas multidimensionales. La definición del XML no es extremadamente compleja, pero en la práctica resulta engorroso recordar cada uno de los elementos junto a sus atributos y sub-elementos, tal y como se encuentran en el almacén. Esta aplicación permite la ejecución de las consultas MDX⁴ (). (16)

Pentaho BI Server 3.6: ofrece soporte y una infraestructura para solucionar los problemas de negocios con soluciones de inteligencia empresarial. Incluye servicios básicos como autenticación, registro, auditoría, servicios web y motor de reglas. También un motor de soluciones unido a un generador de reportes, tablero de comandos, análisis y componentes de minería de datos. Funciona como un sistema de administración web de informes, un servidor de integración de aplicaciones y un

⁴ MDX (lenguaje de expresiones multidimensionales) para consultar datos multidimensionales o para crear expresiones MDX para su uso en un cubo.

motor de flujo de trabajo ligero. Su diseño permite la integración de forma muy fácil a cualquier proceso del negocio.

Mondrian 3.0.4: permite realizar consultas al almacén de datos, y que los resultados sean presentados mediante un navegador; de modo que el usuario pueda interactuar con la información. Comprende cuatro capas, vistas desde el nivel del usuario final: la capa de presentación, la capa dimensional, la capa de estrella y la de almacenamiento. La capa de presentación brinda variadas formas de visualizar los datos, como gráficas de pastel, barras o líneas, y la exportación de las mismas. La dimensional analiza gramaticalmente, valida y ejecuta sentencias MDX. La capa de estrella es la responsable de mantener la caché y la de almacenamiento se encuentra representada por un sistema gestor de base de datos. Se pueden encontrar en el mismo servidor o en varios, aunque la capa dimensional y la de estrella si deben estar en el mismo servidor. (17)

Apache Tomcat 5.5: es un servidor web de aplicaciones. Gestiona solicitudes y respuestas http. Es de código abierto, implementado con tecnología Java bajo la licencia de Apache 2, que le permite funcionar en cualquier sistema operativo donde se encuentre la maquina virtual de java. (18)

1.7. Conclusiones del capítulo

Una vez culminado el desarrollo de este capítulo se decide realizar un mercado de datos para dar solución a la problemática y para la realización del producto se selecciona:

- ✓ Como metodología, la de Desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocios de DATEC.
- ✓ Como herramientas Case, Visual Paradigm para el lenguaje de modelado. PostgreSQL como gestor de base de datos. Para la implementación del subsistema de integración el Pentaho Data Integration (Kettle) y el DataCleaner para la realización del perfilado de los datos. Para el diseño de los cubos multidimensionales el Pentaho Schema Workbench, el Apache Tomcat como servidor web y el Mondrian y el Pentaho BI Server para la implementación de la capa de visualización.

Capítulo II: Análisis y diseño del mercado de datos Comunicaciones

2.1. Introducción

En este capítulo se realizará un estudio preliminar del negocio, donde quedarán definidas las reglas del negocio, los requerimientos de información, requerimientos funcionales y no funcionales, las políticas de respaldo y el esquema de seguridad, así como también se diseñará el modelo de casos de uso del sistema, el modelo de datos y la matriz multidimensional.

2.2. Análisis de la solución

El análisis es la fase determinante para entender los requerimientos de la organización, en el cual se definen las estructuras de almacenamiento, se diseñan las reglas de extracción, transformación y carga, así como la arquitectura de información que regirá el desarrollo de la solución.

2.2.1. Definición del negocio

La Oficina Nacional de Estadística e Información (ONEI), funciona como repositorio central donde interactúan varios procesos, que ejecutan y controlan la gestión estadística de Cuba. Esta entidad opera en diferentes áreas de la economía y la sociedad, utilizando diversos modelos para la gestión de la información.

El modelo 0009 de la ONEI almacena los datos provenientes de la esfera de las comunicaciones, conformada por tres indicadores: la prensa, la radiodifusión y la telefonía. De la prensa se recoge información sobre los periódicos, tabloides, además de la correspondencia, dígame cartas y telegramas. De la telefonía se obtienen datos sobre los servicios públicos de telecomunicaciones mediante la operación, instalación, explotación, comercialización y mantenimiento de las redes públicas, que satisfacen las necesidades de los usuarios y la población. La radiodifusión está dividida en indicadores de radio y de televisión, de la que se maneja la potencia, la disponibilidad tanto en el ámbito internacional como nacional. Toda esta información es recopilada con una periodicidad mensual, semestral y anual. Los datos que son recopilados mensuales y semestrales, son enviados por los centros informantes el día ocho de cada mes al municipio, mientras, los anuales son enviados el día 20 de enero, los cuales se transfieren posteriormente a la provincia, esta, a su vez se los transmite a la ONEI Nacional, donde son analizados por los especialistas a partir de la segunda quincena de cada mes.

2.2.2. Temas de análisis

Las áreas de análisis son la forma en que se dividen o clasifican la información referente a una determinada organización, según los objetivos que se quieran lograr. De acuerdo con los datos que se manejan en la ONEI, en la presente investigación se define como área de análisis las Comunicaciones dividida en tres libros de trabajo prensa, telefonía y radiodifusión.

2.2.3. Reglas del negocio

Las reglas del negocio (RN) describen políticas que deben cumplirse o condiciones que deben satisfacerse por lo que regulan algún aspecto del negocio (19). Es decir, estas reglas son políticas, normas, operaciones, definiciones y restricciones presentes en la institución y que son de vital importancia para alcanzar sus objetivos.

La ONEI para la realización de los cálculos estadísticos en la esfera de las Comunicaciones definió 16 reglas del negocio, con el objetivo de regular la fuente de los datos y obtener un mejor manejo de la información. A continuación se describen estas reglas:

RN1. Distribución de prensa total debe ser mayor o igual que Distribución de prensa de periódicos, más Distribución de prensa de revista, más Distribución de prensa de tabloides.

RN2. Líneas telefónicas fijas en servicio nacionalmente debe ser mayor que Líneas telefónicas fijas en servicio nacionalmente de líneas telefónicas residenciales.

RN3. Total de ingresos en moneda nacional por servicios de telefonía pública debe ser mayor o igual que Total de ingresos en moneda nacional por servicios de telefonía pública de tarjeta en moneda nacional (MN) más Total de ingresos en moneda nacional por servicios de telefonía pública de teléfonos monederos, más Total de ingresos en moneda nacional por servicios de telefonía pública de los centros agentes, más Total de ingresos en moneda nacional por servicios de telefonía pública de los centros comunitarios.

RN4. Abonados móviles del sistema celular debe ser mayor o igual que Abonados móviles del sistema celular de la población.

RN5. Total de ingresos en moneda nacional por servicios de telefonía fija debe ser mayor o igual que Total de ingresos en moneda nacional por servicios de telefonía fija de la población.

RN6. Total de ingresos en moneda nacional por servicios de telefonía fija celular debe ser mayor o igual que Total de ingresos en moneda nacional por servicios de telefonía fija celular de la población.

RN7. Total de ingresos en moneda nacional por servicio postal debe ser mayor o igual que Total de ingresos en moneda nacional por servicio postal de la población.

RN8. Total de ingresos en divisas por servicios de telefonía fija debe ser mayor o igual que Total de ingresos en divisas por servicios de telefonía fija de la población.

RN9. Total de ingresos en divisas por servicios de telefonía móvil debe ser mayor o igual que Total de ingresos en divisas por servicios de telefonía móvil de la población.

RN10. Total de ingresos en divisas por servicios de telefonía pública debe ser mayor o igual que Total de ingresos en divisas por servicios de telefonía pública por tarjetas CUC.

RN11. Total de ingresos en divisas por servicio postal debe ser mayor o igual que Total de ingresos en divisas por servicio postal.

RN12. Facturación total de ventas efectuadas por comercio electrónico debe ser mayor o igual que Facturación total de ventas efectuadas por comercio electrónico en divisa.

RN13. Gasto total en adquisiciones realizadas por comercio electrónico debe ser mayor o igual que Gasto total en adquisiciones realizadas por comercio electrónico en divisa.

RN14. No se admiten datos nulos.

RN15. Variación va a ser igual que Real Año Anterior menos Real Año Actual.

RN16. Crecimiento va a ser igual que Real Año Anterior dividido entre Real Año Actual y el resultado multiplicado por 100.

2.2.4. Requerimientos de información

Los requerimientos de información representan las necesidades de información del cliente. Estos son visualizados en las vistas de análisis. De acuerdo con la investigación del negocio se definieron 96 requerimientos de información los cuales se encuentran distribuidos en los tres indicadores de análisis: la prensa, la radiodifusión y la telefonía, los mismos se encuentran referenciados en el artefacto Especificación de Requerimientos. A continuación se describe algunos de los requerimientos de información por cada indicador de análisis.

Prensa

RI 1: Obtener la distribución total de la prensa mensualmente a partir de su centro informante.

RI 2: Obtener la distribución total de periódicos mensualmente a partir de su centro informante.

RI 3: Obtener la distribución total de revistas nacionales mensualmente a partir de su centro informante.

RI 4: Obtener la distribución total de tabloides mensualmente a partir de su centro informante.

RI 60: Obtener el total de ingresos en divisa por servicios postal semestralmente a partir de su centro informante.

RI 61: Obtener el total de ingresos en divisa por servicios postal a la población semestralmente a partir de su centro informante.

Radiodifusión

RI8: Obtener la potencia radiada de radiodifusión mensualmente a partir de su centro informante.

RI9: Obtener la potencia radiada de radiodifusión de ondas medias mensualmente a partir de su centro informante.

RI 10: Obtener la potencia radiada de radiodifusión de frecuencia modulada mensualmente a partir de su centro informante.

RI 11: Obtener la potencia radiada de radiodifusión internacional mensualmente a partir de su centro informante.

RI 12: Obtener la potencia radiada de televisión mensualmente a partir de su centro informante.

RI 92: Obtener el porcentaje de la población cubierta por la difusión de radio anualmente de entrada a partir de su centro informante.

RI 93: Obtener el porcentaje de la población cubierta por la difusión de televisión anualmente de entrada a partir de su centro informante.

RI 94: Obtener el porcentaje de territorio cubierto por la difusión de radio anualmente de entrada a partir de su centro informante.

RI 95: Obtener el porcentaje de territorio cubierto por la difusión de televisión anualmente de entrada a partir de su centro informante.

Telefonía

RI 26: Obtener el total de abonados del sistema celular mensualmente a partir de su centro informante.

RI 27: Obtener el total de abonados fijos del sistema celular mensualmente a partir de su centro informante.

RI 28: Obtener el total de abonados móviles del sistema celular mensualmente a partir de su centro informante.

RI 29: Obtener el total de abonados móviles del sistema celular de población mensualmente a partir de su centro informante.

RI 44: Obtener el total de de ingresos en moneda nacional por servicios de telefonía fija semestralmente a partir de su centro informante.

RI 45: Obtener el total de de ingresos en moneda nacional por servicios de telefonía fija a la población semestralmente a partir de su centro informante.

RI 85: Obtener el total de tráfico internacional anualmente a partir de su centro informante.

RI 86: Obtener el total de tráfico internacional de salida anualmente a partir de su centro informante.

RI 87: Obtener el total de tráfico internacional de entrada anualmente a partir de su centro informante.

RI 88: Obtener el total de tráfico internacional revertido anualmente de entrada a partir de su centro informante.

2.2.5. Requerimientos funcionales

Los requerimientos funcionales son capacidades o condiciones que el sistema debe cumplir (19), y que son muy importantes para satisfacer las expectativas del cliente. Se definieron como requerimientos funcionales:

RF 1: Autenticar usuario.

RF 2: Extraer datos.

RF 3: Realizar transformación y carga de los datos.

RF 4: Insertar usuario.

RF 5: Eliminar usuario.

RF 6: Insertar roles.

RF 7: Eliminar roles.

RF 8: Insertar reporte.

RF 9: Eliminar reporte.

RF 10: Modificar reporte

RF 11: Configurar elementos del cubo.

- RF 12:** Mostrar padres.
- RF 13:** Mostrar propiedades.
- RF 14:** Suprimir filas y columnas vacías.
- RF 15:** Invertir eje.
- RF 16:** Detallar miembros.
- RF 17:** Mostrar gráfico.
- RF 18:** Configurar gráfico.
- RF 19:** Editar consulta MDX.
- RF 20:** Exportar reporte a .pdf.
- RF 21:** Exportar reporte a .xls.

2.2.6. Requerimientos no funcionales

Los requerimientos no funcionales son propiedades o cualidades que el producto debe tener (19). Para el desarrollo del mercado de datos se identificaron 12 requerimientos no funcionales: de usabilidad, software, hardware, fiabilidad, seguridad entre otros, los cuales se hace referencia en el artefacto Especificación de Requerimientos. A continuación se muestra el listado de los requerimientos no funcionales.

- RNF 1.** Cumplir con las pautas de diseño de las interfaces.
- RNF 2.** Mostrar los títulos de los componentes que aparezcan en la interfaz del sistema en el idioma español.
- RNF 3.** Realizar diseño y navegabilidad de los reportes del almacén de datos de manera sencilla.
- RNF 4.** Garantizar la persistencia de la información.
- RNF 5.** Lograr que los elementos definidos en el almacén tengan una estructura homogénea.
- RNF 6.** Utilizar el lenguaje de programación definido durante la investigación.
- RNF 7.** Confeccionar manual de usuario.
- RNF 8.** Realizar sección de usuario.
- RNF 9.** Utilizar el Sistema Gestor de Base de datos definido durante la investigación.
- RNF 10.** Utilizar la herramienta de integración de datos definida durante la investigación.
- RNF 11.** Utilizar las herramientas para la implementación de la capa de inteligencia de negocios definidas durante la investigación.
- RNF 12.** Garantizar características mínimas de hardware.

2.2.7. Diagrama de casos de uso del sistema

Los diagramas de casos de uso son la representación de la relación entre los casos de uso, que no son más que la captura de los requerimientos y los actores que intervienen en el sistema. Después de un análisis del levantamiento de requerimientos, se define el siguiente diagrama de casos de uso del sistema:

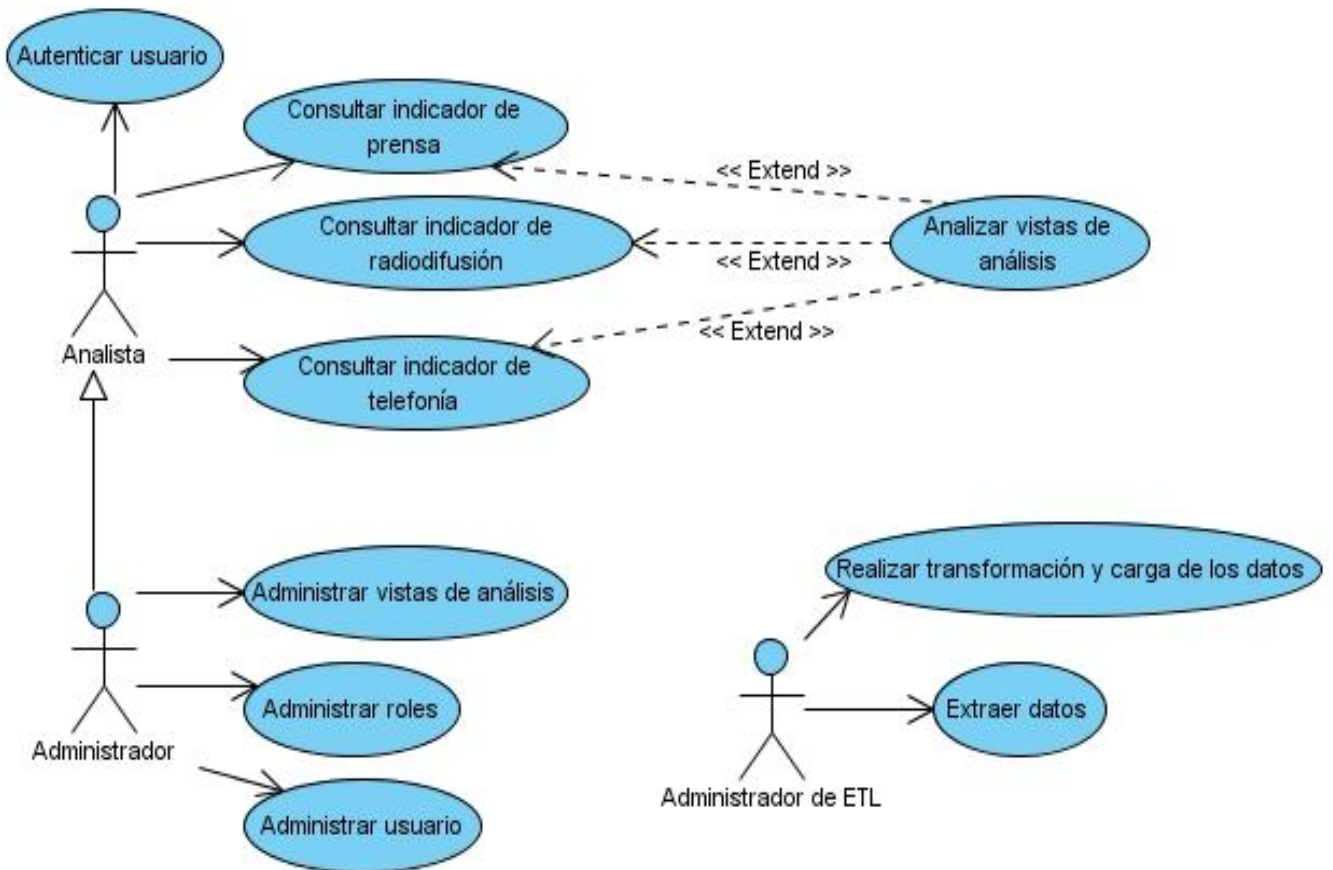


Figura 1: Diagrama de casos de uso del sistema

Actores	Descripción
Analista	Es el responsable de analizar la información de los diferentes indicadores de Comunicaciones.
Administrador	Es el responsable de administrar las vistas de análisis, los usuarios y los roles, además de analizar los diferentes indicadores de Comunicaciones.
Administrador ETL	Es el responsable de la extracción, transformación y

	carga de los datos.
--	---------------------

Tabla 1 Descripción de los actores

Caso de Uso	Descripción
Autenticar Usuario	Realiza la autenticación de los usuarios en el sistema.
Consultar indicador de prensa	Visualiza todas las vistas de análisis del indicador de prensa.
Consultar indicador de telefonía	Visualiza todas las vistas de análisis del indicador de telefonía.
Consultar indicador de radiodifusión	Visualiza todas las vistas de análisis del indicador de radiodifusión.
Analizar vistas de análisis	Brinda las diferentes funcionalidades en las que se pueden analizar las vistas de análisis.
Realizar transformación y carga de los datos	Realiza la transformación y carga de los datos.
Extraer datos	Realiza la extracción de los datos.
Administrar roles	Elimina e inserta los roles que interactúan con el sistema.
Administrar vistas de análisis	Elimina, inserta y modifica las vistas de análisis que se visualizan.
Administrar usuario	Elimina e inserta los usuarios.

Tabla 2 Descripción de los casos de usos

2.3. Diseño de la solución

El diseño es el refinamiento del análisis. El cual identifica los aspectos lógicos del diseño del almacén de datos, como la definición de las tablas y claves, el diseño del esquema de seguridad.

2.3.1. Matriz multidimensional

La matriz multidimensional es la representación de la relación que existe entre los hechos y las dimensiones. Se define como la habilidad para describir y seguir la vida tanto de una dimensión como de un hecho, la cual permite determinar el impacto que provocaría un cambio durante el desarrollo del sistema.

Dimensiones

Hechos

D1: dim_indicador_general

H1: hech_temp_sem_pren

D2: dim_provincia

H2: hech_temp_sem_telf

D3: dim_entidad

H3: hech_temp_mes_pren

D4: dim_temporal_mes

H4: hech_temp_mes_telf

D5: dim_temporal_semestre

H5: hech_temp_mes_rad

D6: dim_temporal_anno

H6: hech_temp_anno_telf

H7: hech_temp_anno_rad

Dimensiones / Hechos	H1	H2	H3	H4	H5	H6	H7
D1	x	x	x	x	x	x	x
D2	x	x	x	x	x	x	x
D3	x	x	x	x	x	x	x
D4			x	x	x		
D5	x	x					
D6						x	x

Tabla 3 Matriz multidimensional

2.3.2. Modelo de datos

El modelo de datos está integrado por las dimensiones, medidas y hechos identificados. Para el diseño del modelo de datos se utilizó el esquema constelación de hechos ya que todas las tablas de hechos tienen dimensiones en común. De acuerdo con las necesidades actuales del negocio se diseñó el siguiente modelo de datos:

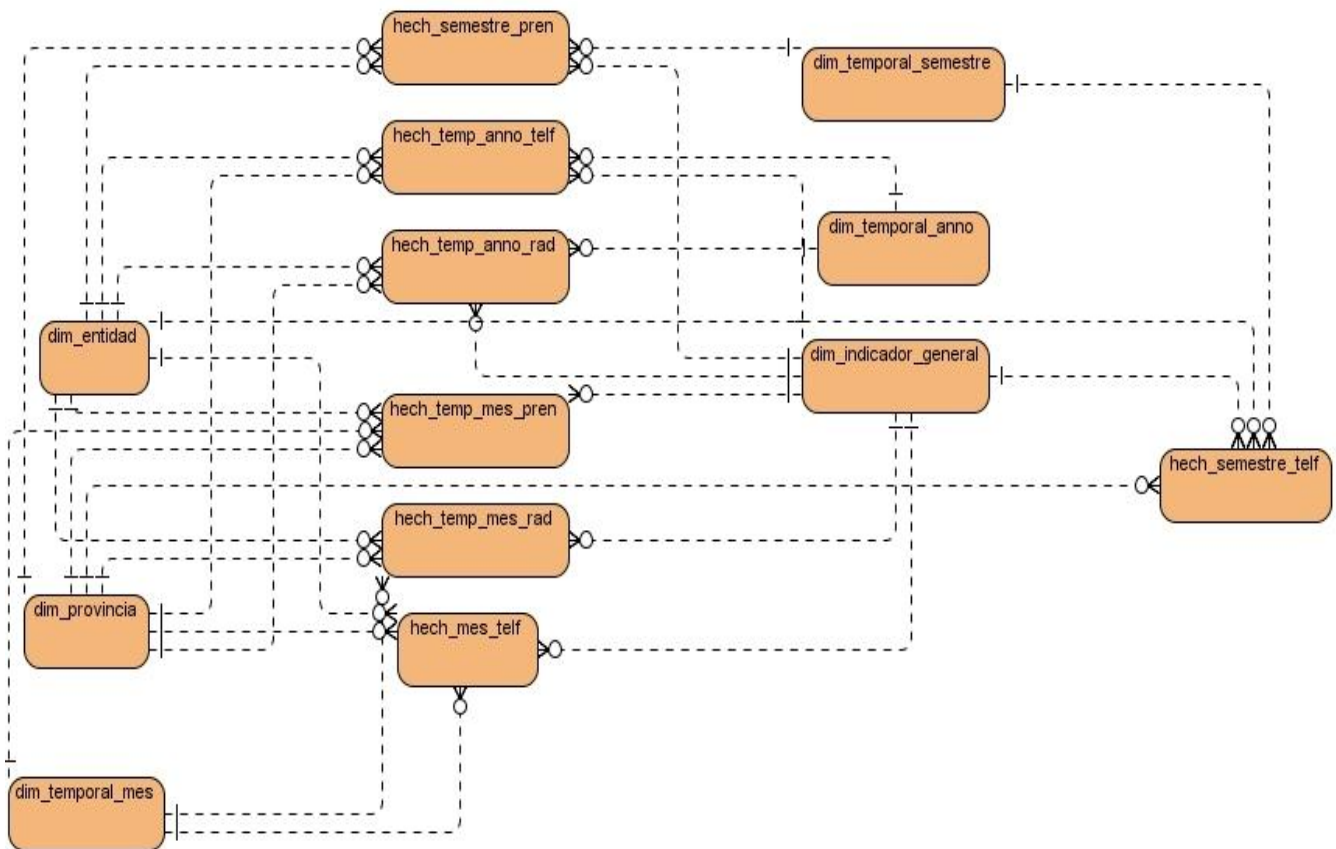


Figura 2 Modelo de datos

Dimensiones

Después de haber realizado el análisis de los modelos propuestos, fue posible la definición de las dimensiones anteriormente mencionadas. A continuación se explica en qué consiste cada una de ellas con sus respectivas jerarquías.

- ✓ . Dimensión dim_indicador_general: es la dimensión que contiene todos los indicadores de Comunicaciones, tanto de prensa y telefonía como de radiodifusión.

Jerarquía: Indicador

- ✓ Dimensión dim_provincia: es la dimensión que contiene las provincias del país.

Jerarquía: Provincia

- ✓ Dimensión dim_entidad: es la dimensión que contiene todas las entidades.

Jerarquía: Entidad

- ✓ Dimensión dim_temporal_mes: es una dimensión temporal donde se registra los meses, el semestre y el año al que pertenece el mes almacenado

Jerarquía: Año
Año → Semestre
Año → Semestre → Mes

- ✓ Dimensión dim_temporal_semestre: es una dimensión temporal donde se registra los semestres y el año al que pertenece el semestre almacenado.

Jerarquía: Año
Año → Semestre

- ✓ Dimensión dim_temporal_anno: es una dimensión temporal donde se registran los años.

Jerarquía: Año

Hechos

Al realizar el análisis del modelo se definen siete tablas de hechos, debido a que la información que es recogida en cada una de estas es distinta y difieren en el tiempo.

- ✓ Hecho temporal semestral de prensa (hech_temp_sem_pren): en el hecho se recoge la información semestral de los indicadores de prensa.
- ✓ Hecho temporal semestral de telefonía (hech_temp_sem_telf): en el hecho se recoge la información semestral de los indicadores de telefonía.
- ✓ Hecho temporal mes de prensa (hech_temp_mes_pren): en el hecho se recoge la información mensual de los indicadores de prensa.
- ✓ Hecho temporal mes de telefonía (hech_temp_mes_telf): en el hecho se recoge la información mensual de los indicadores de telefonía.
- ✓ Hecho temporal mes de radiodifusión (hech_temp_mes_rad): en el hecho se recoge la información mensual de los indicadores de radiodifusión.
- ✓ Hecho temporal anual de telefonía (hech_temp_anno_telf): en el hecho se recoge la información mensual de los indicadores de telefonía.
- ✓ Hecho temporal mes de radiodifusión (hech_temp_anno_rad): en el hecho se recoge la información anual de los indicadores de radiodifusión.

Medidas

Las medidas representan los valores que son analizados, y tienen como características que son numéricas y permiten el cruce de variables en todos los niveles. En el modelo de datos se definen según las necesidades de los clientes cuatro medidas, las cuales se encuentran en todas las tablas de hechos.

- ✓ Real año actual (real_anno_actual): la medida mide el comportamiento de los indicadores en el año actual.
- ✓ Real año anterior (real_anno_anterior): la medida mide el comportamiento de los indicadores en los años anteriores.
- ✓ Variación es igual a Real Año Anterior menos Real Año Actual.
- ✓ Crecimiento es igual a Real Año Anterior entre Real Año Actual y el resultado multiplicado por 100.

2.4. Política de respaldo y recuperación

En el mercado de datos de Comunicaciones la política de respaldo y recuperación que se empleará está condicionada por tres puntos fundamentales:

- ✓ Periodicidad de salvadas del sistema: estas se realizarán cada vez que es recogida la información, es decir, mensual, semestral y anual. Además de que se verifica que exista una copia de toda la información almacenada.
- ✓ Tablas involucradas: Las tablas que se involucran en la realización son: hech_temp_sem_pren, hech_temp_sem_telef, hech_temp_mes_pren, hech_temp_mes_telf, hech_temp_mes_rad, hech_temp_anno_telf, hech_temp_anno_rad.
- ✓ Salvadas existentes: las salvadas se realizan con el mismo período que es recopilada la información. La información que se recoge mensualmente se le hacen salvadas parciales. En los meses de junio y diciembre se realizan salvadas totales donde la información es almacenada en los servidores de la ONEI, además de que se almacena en otro servidor ubicado fuera del edificio de la entidad.

2.5. Esquema de seguridad

La seguridad para el mercado de datos Comunicaciones, está determinada mayormente por los niveles de acceso al sistema. Dicha seguridad se rige fundamentalmente por los roles y permisos que los usuarios poseen en su interacción con la base de datos y la aplicación.

Seguridad en la base de datos

Para la interacción de los usuarios con la base de datos se definió el actor:

Actor	Permiso
Administrador ETL	Tiene total acceso a los datos de la base de datos del sistema.
Administrador	Tiene total acceso los datos, los esquemas, en fin a la estructura de la base de datos del sistema.

Tabla 4 Permiso del usuario en la base de datos

Seguridad en la aplicación

Actualmente las aplicaciones desplegadas en el servidor de inteligencia de negocios de Pentaho BI-server muestran un sucesivo incremento, así como los usuarios que tienen acceso a estas. Como consecuencia de esto se definen los siguientes roles:

Roles	Permisos
Administrador	Tiene acceso total al Área de Análisis General (A.A.G) Sistema de Información de Gobierno (SIGOB).
Analista	Tiene acceso al Área de Análisis (A.A) Comunicaciones.

Tabla 5 Permiso de los roles en la aplicación

Elemento de aplicación	Roles con acceso
A.A.G	➤ Administrador
A.A .Comunicaciones	➤ Administrador ➤ Analista

Tabla 6 Acceso a los elementos de la aplicación

Además de los niveles de acceso y roles antes definidos para interactuar con la aplicación, la plataforma Pentaho BI tiene incluida su propia seguridad, la misma se divide en tres áreas fundamentales, estas son:

- ✓ **Seguridad de acceso a datos de objetos:** Incluye usuarios, contraseñas, autorizaciones permitidas, recursos web y protección a datos.(20)
- ✓ **Autenticación:** Tiene que ver con el procesamiento de información interactiva de inicio de sesión (por ejemplo nombre de usuario y contraseña) comparándola con la información recuperada del almacén de datos de seguridad.(20)
- ✓ **Autorización de recursos web (URL):** Brinda protección a las URL para responder a cada usuario si pueden o no acceder a una determinada página. Esto es decidido por el administrador de recursos web, el cual le brinda a cada usuario autenticado un permiso de seguridad, delimitando las páginas a las que tiene acceso y a las que no.(20)

2.6. Conclusiones del capítulo En este capítulo se definieron los principales procesos que se llevan a cabo en el desarrollo del análisis y diseño del mercado de datos Comunicaciones, en los cuales:

- ✓ Se determinaron 16 reglas del negocio, 96 requerimientos de información, 21 requerimientos funcionales y 12 no funcionales con los que debe cumplir el sistema.
- ✓ Se diseñó el modelo de casos de uso del sistema donde se identificaron los actores, roles y permisos con los cuales constarán cada uno de ellos, además de los casos de uso.
- ✓ Se especificaron seis dimensiones, siete hechos y un correlacionador_comunicaciones por los cuales está compuesto el modelo dimensional.
- ✓ Se establecieron las políticas de seguridad que deben cumplir los empleados de la institución.

Capítulo III: Implementación del mercado de datos Comunicaciones

3.1. Introducción En este capítulo se implementarán los subsistemas de integración donde se realizarán las transformaciones necesarias para la carga de los datos, los cuales se llevan a cabo en el proceso de ETL. Además se le dará solución a los requerimientos de información a través de la implementación del subsistema de visualización.

3.2. Implementación de la base de datos

3.2.1. Estructura de los datos El modelo de datos es la representación de un diseño de datos, el cual permite describir las estructuras y la forma en que estos se relacionan. Los modelos físicos describen cómo se almacenan los datos en el ordenador: el formato de los registros, la estructura de los ficheros (desordenados, ordenados) y los métodos de acceso utilizados (índices).

La estructura física del mercado de datos Comunicaciones esta nombrada de acuerdo a un estándar de codificación definido por el departamento de almacenes para la implementación de la base datos, donde todas las llaves primarias que contienen las tablas de dimensiones son números enteros de tipo serial y las llaves de las tablas de hechos son la unión de las llaves foráneas de las dimensiones que se relacionan con el mismo.

En la siguiente tabla se muestra los estándares de nombrados utilizados para el mercado de datos.

Estructura	Descripción	Ejemplo
Tablas de hechos	Todas las tablas de hechos tendrán una cadena que demuestra que son hechos y el concepto que describen.	hech_<concepto>
Tablas de dimensiones	Todas las tablas de dimensiones tendrán una cadena que demuestra que son dimensiones y el concepto que describen.	dim_<concepto>
Llaves primarias	Todas las llaves primarias tendrán una cadena, que demuestre la tabla a la que pertenece seguido por _id donde demuestra que son llaves	<nombre_tabla>_id

	primarias.	
Atributos compuestos	En los atributos donde el nombre es compuesto se debe especificar el primer componente del atributo separado del segundo por un caracter de -.	<Primer nombre>_<Segundo nombre>

Tabla 7: Estructura de nombrado

Índices

Los sistemas que contienen grandes volúmenes de datos pueden producir demoras ante una consulta. La mayoría de estas consultas incluirán probablemente operaciones de unión entre tablas muy grandes, lo cual puede resultar extremadamente costoso. Para ganar en eficiencia a la hora de realizar estas operaciones se han investigado y creado técnicas especializadas que hoy ofrecen varios gestores, como los índices.

Las políticas de indexado son restricciones que el programador de base de datos define a la hora de indexar o crear un índice. Un índice es una estructura de disco asociada con una tabla o una vista que acelera la recuperación de filas de la tabla o de la vista. Contiene claves generadas a partir de una o varias columnas de la tabla o la vista. Dichas claves están almacenadas en una estructura que permite que SQL Server busque de forma rápida y eficiente la fila o filas asociadas a los valores de cada clave.

La solución posee como indexado el que trae por defecto el gestor PostgreSQL, para la búsqueda de datos utiliza las llaves primarias, foráneas y campos únicos. Los índices son de tipo “b-tree” (Árbol B) lo que implica que cualquier búsqueda que se realice utilizando las llaves, se optimiza mediante este método, además de que en las búsquedas se pueden utilizar operaciones <, <=, =, >=, > siempre que un atributo indexado estén involucrado.

3.2.2. Esquemas y Tablas Los esquemas representan agrupaciones de elementos de una la base de datos (BD), estos contienen funciones, operadores y tipos de datos. Para el desarrollo del mercado de datos se definieron tres esquemas:-

- ✓ Esquema dimensiones: contiene las seis tablas de dimensiones.
- ✓ Esquema mart_comunicaciones: contiene siete tablas de hechos.
- ✓ Esquema stg_etl: contiene la tabla comunicaciones_correlacionador.

Esquemas	Tablas
dimensiones	dim_entidad
	dim_indicador_general
	dim_provincia
	dim_temporal_anno
	dim_temporal_semestre
	dim_temporal_mes
mart_comunicaciones	hech_temp_anno_rad
	hech_temp_anno_telf
	hech_temp_sem_pren
	hech_temp_sem_telf
	hech_temp_mes_pren
	hech_temp_mes_telf
	hech_temp_mes_rad
stg_etl	correlacionador_comunicaciones

Tabla 8 Esquema de la base de datos

3.2.3. Usuarios y roles Postgresql ofrece la posibilidad de agrupar a los usuarios en roles, según las necesidades de permisos y accesos de cada rol, para realizar su función como trabajador del sistema. Los usuarios y roles definidos en la base de datos contribuyen a garantizar la seguridad de la misma. En la implementación de la base de datos se definieron tres usuarios donde se le asignaron los siguientes privilegios:

- ✓ **Administrador del sistema:** tiene acceso a la base de datos en su totalidad, dígase administración y configuración tanto de la base de datos como de los usuarios restantes. Este desempeña el rol de administrador y tiene los privilegios de leer, insertar, eliminar, modificar y actualizar los datos, los esquemas y las tablas almacenados en la base de datos.
- ✓ **Administrador de ETL:** tiene acceso solamente a los datos de la base de datos. Desempeña el rol de administrador de ETL y tiene los privilegios de leer, insertar, eliminar y actualizar solamente los datos almacenados en la base de datos.
- ✓ **Analista:** tiene acceso a consultar la información de la base de datos. Desempeña el rol de analista del sistema y tiene privilegio de lectura a los datos almacenados en la base de datos.

3.3. Implementación del subsistema de integración

3.3.1. Arquitectura del subsistema de integración

La arquitectura del proceso de desarrollo de un software, es la representación de la estructura del sistema al nivel más alto y consiste en un grupo de patrones que sirven de guía para la elaboración de un software. No se recomienda comenzar la construcción sin tener definida la arquitectura.

En el caso de la integración de datos de ETL, la arquitectura queda de la siguiente forma:

- ✓ **Fuente de Datos:** son los archivos de extensión DBF que contienen la información.
- ✓ **Área temporal:** es el área donde se realiza la integración y transformación de los datos: punto intermedio entre la fuente y el mercado.
- ✓ **Mercado de Datos:** destino donde son cargados los datos para su futuro análisis.

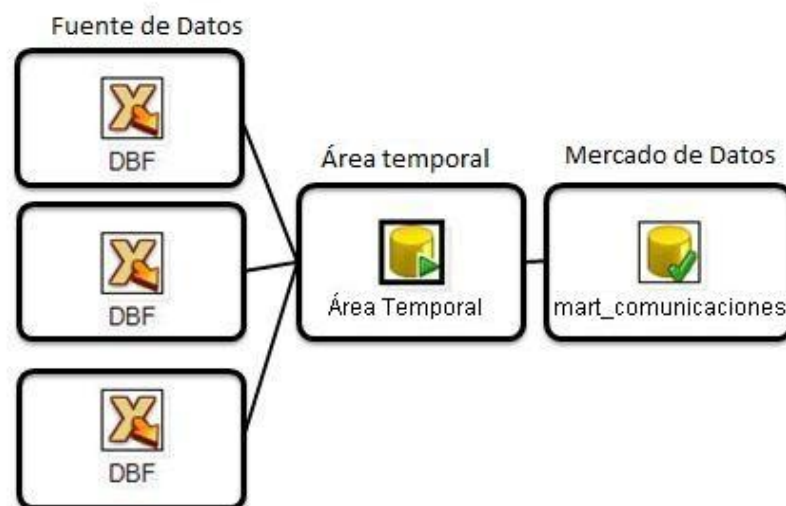


Figura 3 Arquitectura de Integración

3.3.2. Perfilado de Datos

Es un proceso que se recomienda realizar antes del proceso de integración. Consiste en un análisis de la fuente para conocer el estado en que se encuentran los datos, calidad de los mismos y su estructura. Del resultado se establecen reglas para luego realizar la integración de los datos. La herramienta que se utilizó fue Data Cleaner. Después de realizar el perfilado de los datos se obtuvieron los siguientes resultados:

- ✓ No se encontraron valores nulos.

- ✓ Los tipos de datos identificados fueron varchar e integer.
- ✓ Existen valores duplicados y únicos.

3.3.3. Extracción, Transformación y Limpieza de los Datos

La extracción de los datos es el proceso donde se llevan los datos de los archivos dbf, para el área temporal para adaptarlos al modelo establecido. Los archivos contienen todos los datos referentes al mercado de datos Comunicaciones que serán almacenados en las siete tablas de hechos.

La transformación y limpieza es de gran importancia, porque es la etapa donde se garantiza el resultado final de cómo se van a mostrar los datos. Se aplican las reglas de transformación definidas a partir de las reglas de transformación identificadas; se detectan otras posibles deficiencias de la fuente y se corrigen. Entre los errores más comunes se encuentran: valores nulos, combinación de datos, entre otros. El resultado es la carga al mercado de datos.

3.3.4. Carga

Es el paso final de este proceso, los datos son cargados al mercado para ser utilizados de forma satisfactoria.

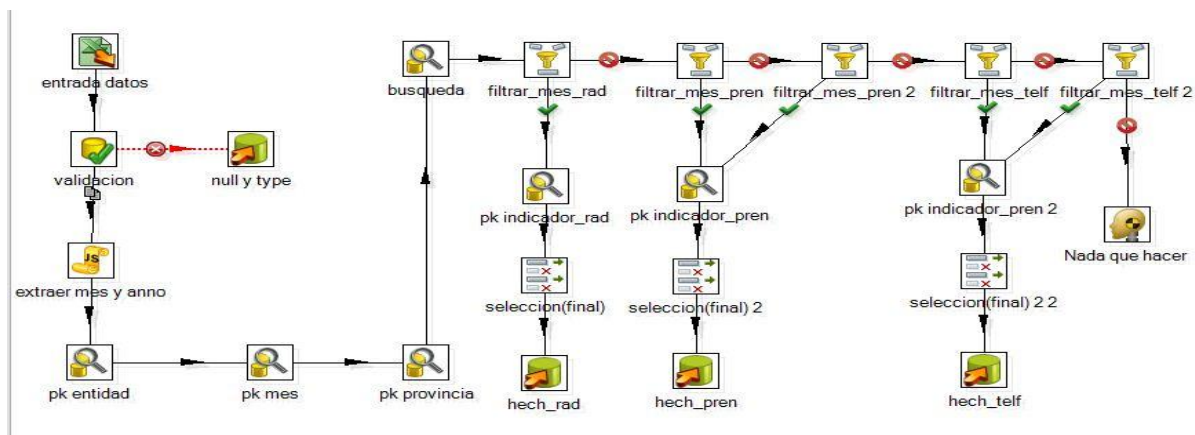


Figura 4 Transformación de la carga de los hechos que se muestran con información mensual.

La figura 4 muestra un ejemplo de una de las transformaciones implementadas que corresponde a la carga de los datos mensuales, la misma abarca todos los procesos de la integración de datos: la extracción, la limpieza y carga de los mismos.

Una transformación es el elemento básico para el proceso de ETL en la herramienta seleccionada. Está compuesta por pasos, que son la unidad más pequeña de la transformación, enlazados entre sí por saltos, que es por donde pasa el flujo de la información. Es un conjunto de metadatos en XML

(Extensible Markup Language) que son interpretados por el motor del Kettle dándole las instrucciones para llevar a cabo las acciones configuradas.

La entrada de datos corresponde a la extracción. Luego se valida que los datos no contengan valores nulos y que corresponda a los tipos de datos identificados en el perfilado. En caso de encontrar algún valor nulo, es almacenado en un archivo excel que está localizado en una carpeta que se encuentra donde mismo están las transformaciones, llamada Error. Del nombre del archivo se obtiene el código del año y el mes, que sirven para buscar las llaves primarias en las dimensiones correspondientes, así como en las dimensiones provincia e indicador general. Finalmente se carga esta información para el mercado de datos.

3.3.5. Trabajo

Un trabajo es un conjunto sencillo o complejo de tareas con el objetivo de realizar una acción determinada. Pueden realizarse acciones de envío de email, así como validaciones de conexión a base de datos y repositorios. También puede ejecutarse transformaciones donde hayamos diseñado una secuencia de instrucciones. Se encuentra en un nivel superior a las transformaciones.

La forma en que se realiza la carga de los datos hacia el mercado se muestra en la figura 5. Luego de asegurarse que la conexión al mercado está en perfecto estado, se comienza a realizar la carga del mismo mediante la ejecución de las transformaciones previamente diseñadas.

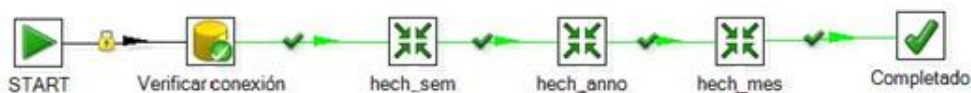


Figura 5 Trabajo de la carga de los hecho

3.4. Implementación de subsistema de visualización

3.4.1. Esquema OLAP

Para la implementación del subsistema de visualización, es necesaria la creación de los cubos multidimensionales, en los cuales se definen las dimensiones, los niveles de jerarquía de las dimensiones, las medidas físicas, las medidas calculables, las propiedades. Esto se realiza a través de la herramienta Pentaho Schema Workbench, donde la misma genera un fichero XML, el cual contendrá la estructura de todos los cubos OLAP.

Capítulo III: Implementación del mercado de datos Comunicaciones

En la realización del diseño de los cubos, se modeló el esquema `mart_comunicaciones` integrado por siete cubos OLAP: `anno_rad`, `anno_telf`, `mes_pen`, `mes_rad`, `mes_telf`, `sem_pren`, `sem_telf`, donde cada uno está formado por cuatro dimensiones. Las dimensiones entidad, indicador y provincia son comunes para todos. La dimensión indicador contiene propiedades, unidad de medida (um), ya que además de visualizar la descripción del indicador es necesario mostrar la unidad de medida de cada indicador. Todos los cubos están compuestos por cuatro medidas: `real_anno_actual`, `real_anno_anterior`, `variación` y `crecimiento`, las dos últimas son medidas calculables, y se calculan de acuerdo a la RN15 y RN16 definidas en la etapa de análisis. Las siguientes imágenes muestran una vista de la aplicación.

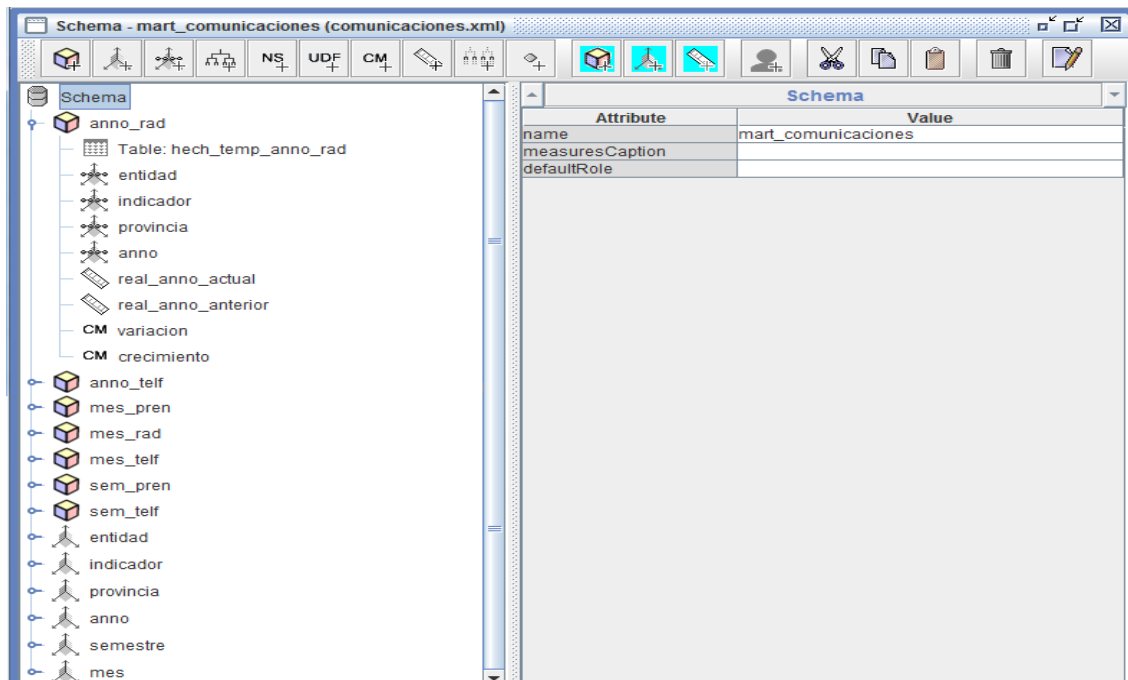


Figura 6 Diseño del esquema

Los cubos `anno_rad` y `anno_telf` comparten las mismas dimensiones, al igual que `mes_pen`, `mes_telf` y `mes_rad`, esto mismo sucede con los cubos `sem_pren` y `sem_telf`. Las siguientes imágenes muestran la estructura de algunos cubos:

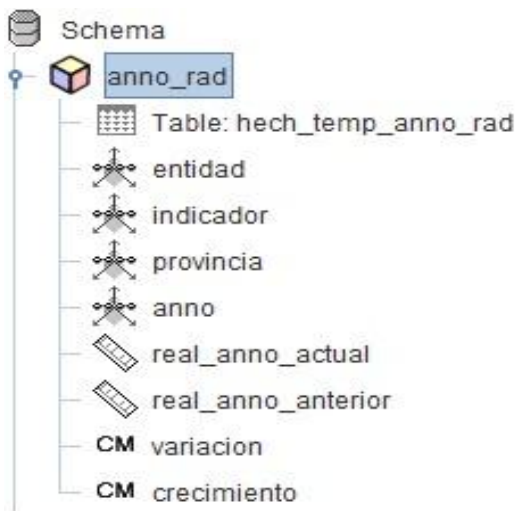


Figura 7: Cubo año radiodifusión

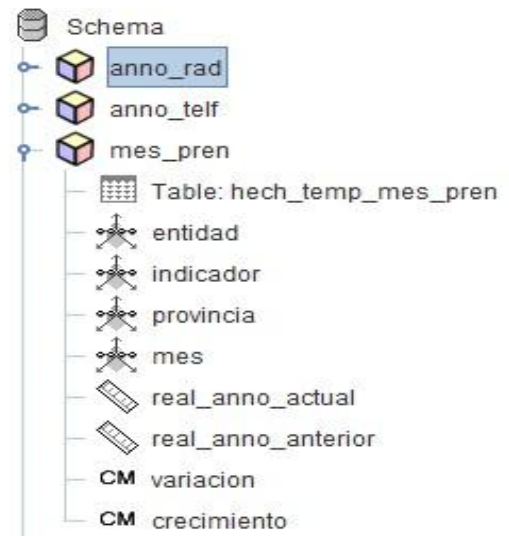


Figura 8 : Cubo mes prensa



Figura 9: Cubo semestre prensa

3.4.2. Navegación de la capa de visualización

El mapa de navegación es la representación gráfica en que está organizada la información. El mercado de datos Comunicaciones está compuesto por un Área de Análisis General (A.A.G), un Área de

Capítulo III: Implementación del mercado de datos Comunicaciones

Análisis (A.A), tres Libros de Trabajo (L.T) y 16 Tablas de Salida (TS). A continuación se detalla la estructura en la cual será presentado en la capa de visualización.

Descripción del A.A.G.

- ✓ **A.A.G. SIGOB:** agrupa toda la información de los mercados de datos de la ONEI, formando el almacén SIGOB.

Descripción del A.A.

- ✓ **A.A. Comunicaciones:** agrupa toda la información referente a los indicadores de prensa, radiodifusión y telefonía. Además contiene reportes estadísticos que muestran el comportamiento de los diferentes indicadores.

Descripción de los L.T.

- ✓ **L.T. Prensa:** agrupa la información referente a los indicadores de prensa, contiene cuatro reportes que permiten realizar un análisis general de datos.
- ✓ **L.T. Radiodifusión:** agrupa la información referente a los indicadores de radiodifusión, contiene tres reportes que permiten realizar un análisis general de datos.
- ✓ **L.T. Telefonía:** agrupa la información referente a los indicadores de telefonía, contiene nueve reportes que permiten realizar un análisis general de datos.

Descripción de las vistas de análisis del L.T. Prensa.

- ✓ **T.S 1: Distribución de prensa:** muestra la distribución de prensa de todas las provincias del país en los meses del año 2010.
- ✓ **T.S 2: Total de ingreso en CUC y MN por servicio postal:** muestra el total de ingreso en divisa y moneda nacional por servicio postal de todas las provincias del país en los semestres del año 2010.

- ✓ **T.S 3: Unidades de correos:** muestra el total de unidades de correos que existen en todas las provincias del país en los meses del año 2010.
- ✓ **T.S 4: Ventas:** muestra el total de venta de sobres y sellos que ocurren en todas las provincias del país en los meses del año 2010.

Descripción de las vistas de análisis del L.T. Radiodifusión.

- ✓ **T.S 1: Confiabilidad de servicio:** muestra la confiabilidad de servicio en todas las provincias del país en los meses del año 2010.
- ✓ **T.S 2: Porcentaje:** muestra el porcentaje de radio y televisión de todas las provincias del país del año 2010.
- ✓ **T.S 3: Potencia radiada:** muestra la potencia de radio y televisión de todas las provincias del país en los meses del año 2010.

Descripción de las vistas de análisis del L.T. Telefonía.

- ✓ **T.S 1: Completamiento de llamadas:** muestra la cantidad de llamadas ya sean locales o a larga distancia en todas las provincias del país en el año 2010.
- ✓ **T.S 2: Facturación y gasto:** muestra la facturación y gastos de comercio electrónico en todas las provincias del país en los semestres del año 2010.
- ✓ **T.S 3: Línea telefónica:** muestra la cantidad de líneas telefónicas fijas de todas las provincias del país semestralmente en el año 2010.
- ✓ **T.S 4: Otros:** muestra el promedio de interrupciones pendientes, la cantidad de teléfonos reparados, entre otros indicadores de telefonía de todas las provincias del país en los meses del año 2010.

- ✓ **T.S 5: Total de abonados:** muestra el total de abonados fijos y móviles del sistema celular de todas las provincias del país en los meses del año 2010.
- ✓ **T.S 6: Total de ingreso en CUC:** muestra el total de ingreso en divisas por servicios de telefonía en todas las provincias del país semestralmente en el año 2010.
- ✓ **T.S 7: Total de ingreso en MN:** muestra el total de ingreso en Moneda Nacional por servicios de telefonía en todas las provincias del país semestralmente en el año 2010.
- ✓ **T.S 8: Total de estaciones públicas:** muestra el total de estaciones públicas en todas las provincias del país en el año 2010.
- ✓ **T.S 9: Tráfico internacional:** muestra el tráfico internacional de salida, de entrada y revertido en todas las provincias del país en el año 2010.

El mercado de datos Comunicaciones está compuesto por 16 vistas de análisis en total, las cuales fueron distribuidas dentro de los tres libros de trabajo. Para la realización de estas vistas se utilizó el Mondrian, el Tomcat y el Pentaho BI Server.

Todas las vistas de análisis contienen la misma estructura, en las filas se encuentra la dimensión provincia que agrupa a la dimensión indicador, mientras en las columnas están la dimensión entidad, la dimensión de tiempo a la que pertenece la vista y las cuatro medidas. En la siguiente imagen se representa un ejemplo de cómo es la estructura de las vistas de análisis.

Entidad		Real Año Actual	Real Año Anterior	Variación	Crecimiento
EMPRESA DE CORREOS DE CUBA, ECC					
Mes					
Junio					
Medidas					
Provincia	Indicador	Real Año Actual	Real Año Anterior	Variación	Crecimiento
Provincia	Distribución de prensa total	119.856,3	23.645,2	96.211,1	506,9%
	Distribución de prensa total de periódicos	113.290,5	22.628,7	90.661,8	500,6%
	Distribución de prensa total de revistas nacionales	6.565,8	1.006,9	5.558,9	652,1%
	Distribución de prensa total de tabloides	0	9,6	-9,6	0,0%

Filtro:



Figura 10 : Ejemplo del reporte Distribución de Prensa

3.4.3. Configurar la seguridad de los usuarios

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

- ✓ El rol administrador tiene todos los permisos de la aplicación y posee el usuario administrador del sistema. Este rol tiene acceso a la carpeta de área de A.A.G.SIGOB y a todas las carpetas que contiene.

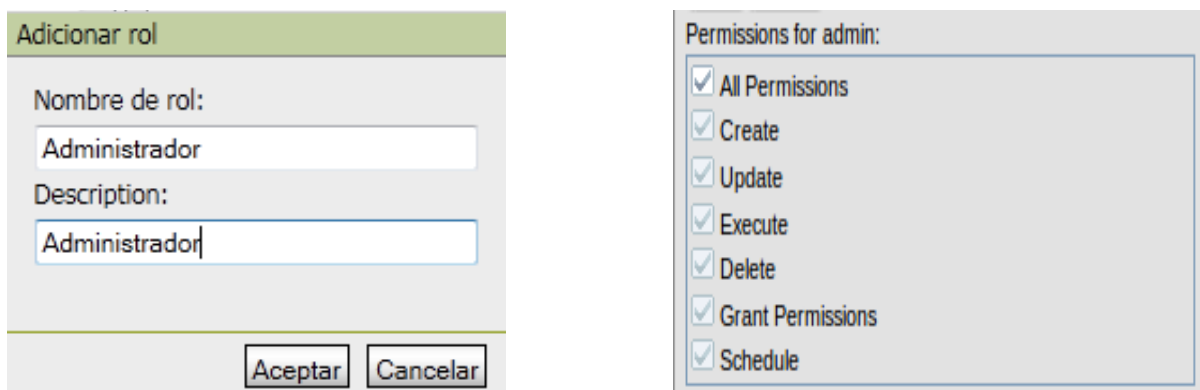


Figura 11: Crear rol y permiso al administrador

El rol analista tiene permiso de sólo lectura y posee el usuario analista del sistema. Este rol tiene acceso a la carpeta de área de A.A.Comunicaciones y a todas las carpetas que contiene.

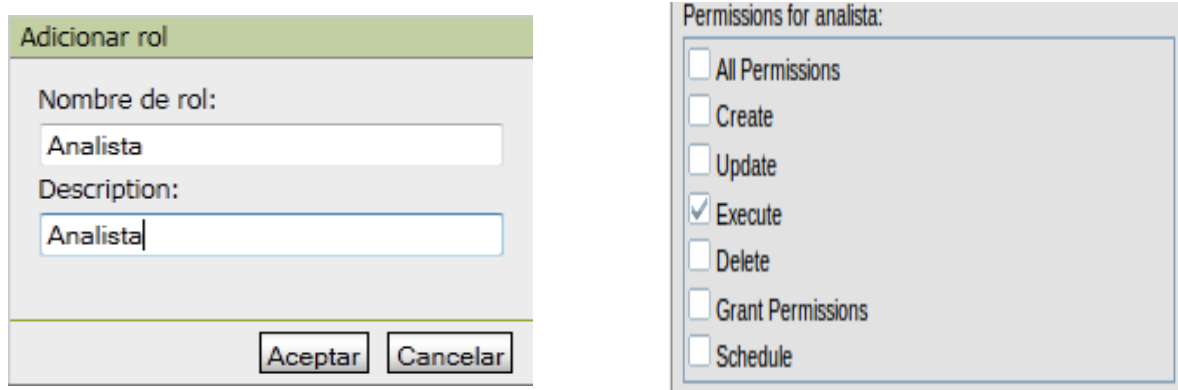


Figura 12: Adicionar rol y permiso al analista

3.5. Conclusiones del capítulo

En el desarrollo de este capítulo se realizó la implementación del mercado de datos Comunicaciones donde se efectuó:

- ✓ La carga del modelo de datos, donde la BD quedó estructurada por tres esquemas y 14 tablas.
- ✓ Se implementó el subsistema de integración, en el cual se obtuvieron tres transformaciones y un trabajo, logrando así la carga de los hechos satisfactoriamente.
- ✓ Se modeló el esquema multidimensional a través del diseño de los siete cubos OLAP, a partir de estos se implementó el subsistema de visualización en el que fueron realizados 16 vistas de análisis.

Capítulo 4: Pruebas del mercado de datos Comunicaciones

4.1. Introducción

Una vez realizada la implementación de la aplicación, comienza la validación y prueba del mercado de datos Comunicaciones, donde se verifica que se cumplan todos los requerimientos definidos en el negocio.

4.2. Pruebas

El creciente desarrollo de software implica una serie de actividades por lo que existe una mayor probabilidad que aparezcan errores. Para garantizar la calidad del producto surgen las pruebas.

Las pruebas, son la actividad en la cual un sistema o componente es ejecutado bajo unas condiciones o requerimiento específicos, los resultados son observados y registrados. Esta actividad no garantiza la ausencia de defecto; sólo puede demostrar que existen defectos y nos permite evaluar la calidad de software.

Para evaluar la calidad del mercado de datos Comunicaciones se utilizó el Modelo V el cual fue definido por el Centro de Tecnologías de Gestión de Datos (DATEC) con el fin de lograr que el producto cumpla con las especificaciones del negocio y tenga una mayor calidad.

El modelo V muestra cómo se relacionan las actividades de prueba con el análisis y el diseño. La fase de análisis y diseño se encuentra a la izquierda del modelo el cual es relacionado con las pruebas que se encuentran a la derecha a través de los planes de pruebas. La codificación forma el vértice de la V, el mismo es la etapa de implementación del mercado de datos. Este modelo permite que cada vez que se termina una etapa de desarrollo se le apliquen pruebas, garantizando una mayor calidad del producto para la próxima etapa. En la siguiente figura queda estructurado el modelo V.

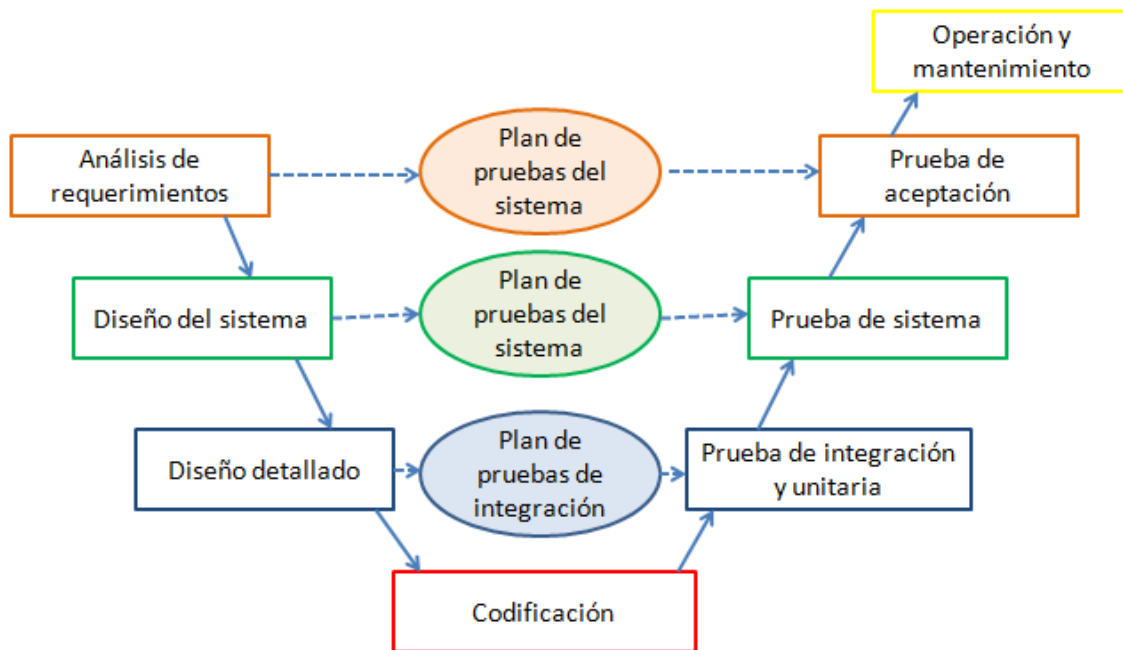


Figura 13: Modelo V

- ✓ **La prueba unitaria:** es el proceso de probar los componentes individuales de la solución. El propósito es descubrir las diferencias entre los artefactos y el comportamiento real de la aplicación.
- ✓ **La prueba de integración:** son las pruebas que se realizan para determinar la integración de los componentes dentro de un sistema y evaluar su correcta interface, funcionalidad y desempeño. Estas pruebas son diseñadas y ejecutadas por el desarrollador cuando la solución está completa.
- ✓ **Prueba del sistema:** se refiere al comportamiento del sistema integrado. La prueba de sistema se aplica generalmente para probar 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:** se realizan para probar que el sistema cumpla con los requerimientos especificados por el cliente.

Para realizar las pruebas al mercado de datos Comunicaciones se elaboró una lista de chequeo y se diseñaron tres casos de prueba.

4.3. Lista de Chequeo

La lista de chequeo es un listado de preguntas, en forma de cuestionario que se utiliza para verificar el grado de cumplimiento de determinadas reglas establecidas. Tiene como objetivo evaluar la calidad del mercado de datos Comunicaciones, donde se mide una serie de indicadores que se encuentran implicados en el proceso de la creación de la capa de integración y visualización del sistema, además de medir, la calidad de los artefactos generados en la realización del producto.

Los indicadores que contiene, se encuentran distribuidos en tres secciones fundamentales:

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

Los elementos que forman parte de la estructura de la lista de chequeo son:

- ✓ **Peso:** define si el indicador a evaluar es crítico o no, el mismo se define con una C si es crítico.
- ✓ **Indicadores a evaluar:** son los indicadores a evaluar en las secciones Estructura del documento, Semántica del documento e Indicadores definidos por la etapa
- ✓ **Evaluación (Eval):** es la forma de evaluar el indicador en cuestión, el mismo se evalúa de 1 en caso de que exista alguna dificultad sobre el indicador y 0 en caso de que el indicador revisado no presente problemas.
- ✓ **N.P. (No Procede):** se usa para especificar que el indicador no es necesario evaluarlo en ese caso.
- ✓ **Cantidad de elementos afectados (CEA):** especifica la cantidad de errores encontrados sobre el mismo indicador.
- ✓ **Comentario (Comt):** especifica los señalamientos o sugerencias que quiera incluir la persona que aplica la lista de chequeo, pueden o no existir señalamientos o sugerencias.

Una vez definida la estructura de la lista de chequeo esta es aplicada al mercado de datos, donde todos los indicadores son evaluados (Anexo 7). Esto tuvo como resultado que de los 25 indicadores de los cuales 11 son críticos se encontraron tres no conformidades una por cada sección, las cuales fueron resueltas. En la siguiente gráfica se muestra el comportamiento de la lista de chequeo.

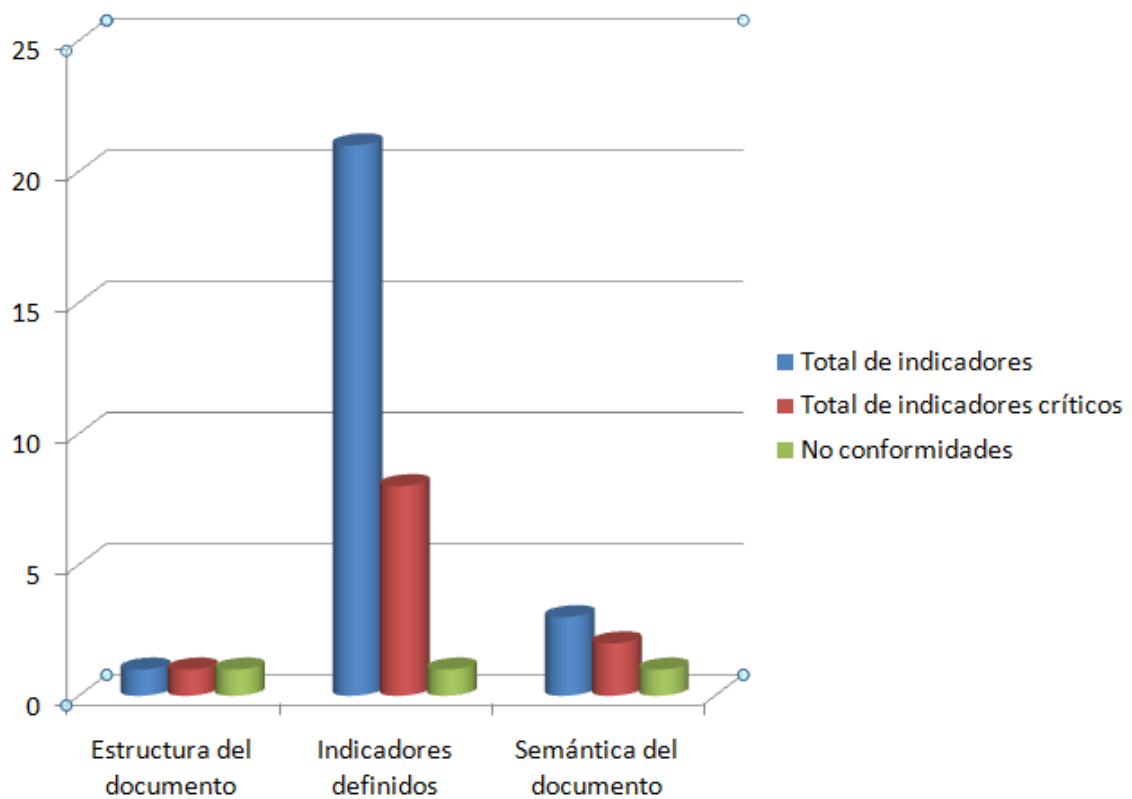


Figura 14 Comportamiento de los indicadores

4.4. Casos de prueba

Los casos de prueba permiten verificar la calidad del software, son utilizados para identificar posibles fallos de implementación y comprobar el grado de cumplimiento de las especificaciones del sistema. Está formado por una sección:

- ✓ **Reportes candidatos:** verifica que los reportes se visualicen con las variables del mismo.

En el mercado de datos Comunicaciones se aplicaron tres casos de prueba, uno a cada caso de uso de información identificado en la etapa de análisis. Al aplicar los casos de pruebas se identificaron ocho no conformidades, las cuales ya fueron resueltas, en la siguiente gráfica se muestra la cantidad de no conformidades por cada caso de prueba.

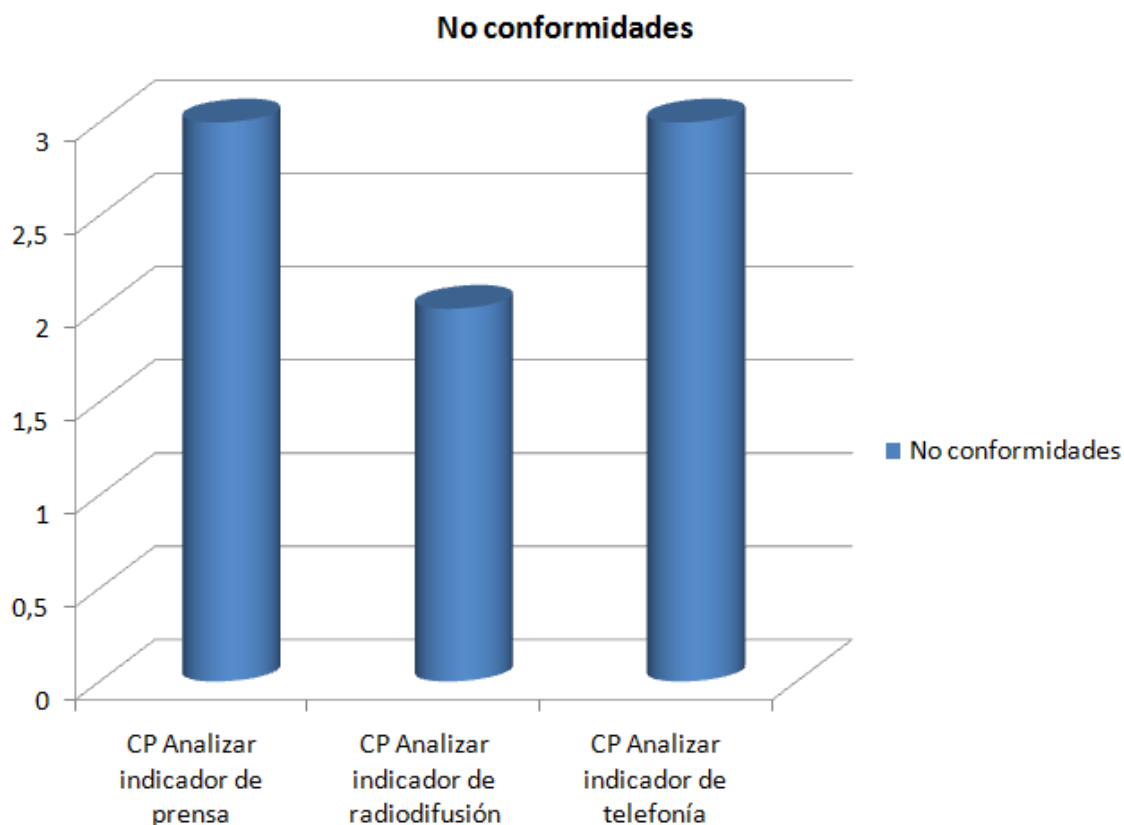


Figura 15: No conformidades de los casos de prueba

4.5. Prueba de sistema

Al mercado de datos se le hicieron pruebas del sistema las que fueron realizadas por los especialistas del centro DATEC donde se encontró una no conformidad la cual afectaba el funcionamiento del producto, esta fue resuelta.

También se le realizaron pruebas de control de acceso al sistema donde se valida que los usuarios establecidos tengan permiso de acceder y los permisos para realizar las operaciones previamente establecidas. Las mismas arrojaron un satisfactorio resultado donde se pudo comprobar la correcta autenticación y autorización al sistema.

Finalmente el producto fue liberado por parte de los especialistas de calidad de software de la UCI (Calisoft) con cero no conformidades, lo cual demuestra la calidad del mercado de datos.

4.6. Prueba de aceptación

Las pruebas de aceptación son la demostración de que la solución cumple con las necesidades del cliente y con los requerimientos definidos en la fase de análisis. Durante estas pruebas se realizaron varias entrevistas con el cliente, donde finalmente el producto fue aprobado.

4.7. Conclusiones del capítulo

En este capítulo se realizó la validación y prueba del mercado de datos Comunicaciones, donde se realizaron varias pruebas en las cuales se obtuvieron diferentes resultados:

- ✓ En la aplicación de la lista de chequeo se encontraron tres no conformidades.
- ✓ En la aplicación de los casos de prueba se encontraron ocho no conformidades.
- ✓ En las pruebas de sistema se encontró una no conformidad.
- ✓ En la prueba de aceptación se aprobó el mercado de datos Comunicaciones.
- ✓ Todas las no conformidades fueron resueltas satisfactoriamente.

Conclusiones

Conclusiones

Una vez concluida la implementación del Mercado de Datos Comunicaciones se llegaron a las siguientes conclusiones generales:

- ✓ Se realizó el análisis y diseño del mercado de datos Comunicaciones acorde con lo previsto en el negocio.
- ✓ Se implementaron las estructuras dimensionales, dando soporte al proceso de integración de los datos de manera satisfactoria.
- ✓ Los reportes fueron desarrollados en su totalidad lo que permite su visualización, contribuyendo a la toma de decisiones.
- ✓ Se realizaron con éxito las pruebas pertinentes para la validación del producto.

Recomendaciones

Se recomienda:

- ✓ Integrar el mercado de datos Comunicaciones con el Sistema Informático de Gestión Estadística (SIGE).
- ✓ Poner en explotación el mercado de datos para probar las funcionalidades del mismo.
- ✓ Desplegar el producto final hasta las provincias de forma tal que se pueda realizar un mejor análisis de los datos y por consiguiente una mejor toma de decisiones.

Referencias bibliográficas

1. **Inmon, William H.** *Building the data warehouse*. 2002.
2. **Kimball, Ralph.** *The Data Warehouse Toolkit*. 2009.
3. **Bill Inmon vs. Ralph Kimball.** *1keydata*. [Online] [Cited: Octubre 27, 2010.] <http://www.1keydata.com/datawarehousing/inmon-kimball.htm>.
4. **Cabrera, María Evelia Casales.** *Data Wharehouse*. 2009.
5. **Sinnexus.** [Online] [Cited: Octubre 27, 2010.] <http://www.sinnexus.com>.
6. **Ibarra, María De Los Ángeles.** *Procesamiento Analítico en Línea*. 2005.
7. **Torres, Lilian Vegas and Luis, Rojas Diaz y Cecilia, Placeres Villar.** *La inteligencia de negocio. Su implementación mediante la plataforma Pentaho*. La Habana : s.n., 2009.
8. **Gascón, Manuel de la Herrán.** Eside. [Online] [Cited: Noviembre 3, 2010.] <http://www.eside.deusto.es/profesores/mherran/>.
9. **TODO BI, "Informe Business Intelligence".** *Recopilación de los mejores artículos de inteligencia de negocio del 2006. Abril*. 2006.
10. **Integración de Datos.** [Online] [Cited: Noviembre 7, 2010.] <http://www.datati.es/cuadros-demandos-%C2%BFque-integracion-de-datos-es-la-mas-adecuada/> .
11. **Postgres.** *Postgres Official Site*. [Online] [Cited: Noviembre 10, 2010.] <http://postgres.org> .
12. **Pgadmin.** *Pgadmin Official Site*. [Online] [Cited: Noviembre 10, 2010.] <http://pgadmin.org>.
13. **Visual Paradigm.** [Online] [Cited: Octubre 13, 2010.] <http://www.visual-paradigm.com>.
14. **Transformación grafica ETL con Spoon – un componente de Pentaho Data Integration (Kettle).** [Online] [Cited: Octubre 27, 2010.] Transformación grafica ET <http://www.gravitar.biz/index.php/bi/introduccion-pentaho-parte-1/#comment-46>.
15. **Data Cleaner.** *Data Cleaner Official Site*. [Online] [Cited: Octubre 25, 2010.] <http://datacleaner.eobjects.org/>.
16. **Mondrian Schema Workbench.** [Online] [Cited: Noviembre 3, 2010.] <http://wiki.pentaho.com/display/PMOPEN/Mondrian+Schema+Workbench>.
17. **Pentaho Architecture Layers of a Mondrian System.** [Online] [Cited: Octubre 27, 2010.] <http://mondrian.pentaho.org/documentation/architecture.php>.
18. **Apache Tomcat.** [Online] [Cited: Noviembre 10, 2010.] <http://tomcat.apache.org>.
19. **Modelamiento del Negocio.** [Online] [Cited: Octubre 26, 2010.] http://eva.uci.cu/file.php/102/Curso_2010-

Referencias bibliográficas

2011/Clases/Semana_04/Conferencia_6/Materiales_complementarios/Fase_de_Inicio._Disciplina_de_Modelamiento_del_Negocio.pdf.

20. **Plataforma Pentaho BI**. Seguridad. [Online] [Cited: Octubre 27, 2010.]

<http://wiki.pentaho.com/display/ServerDoc1x/03.+Platform+Security+Implementation>.

Bibliografía

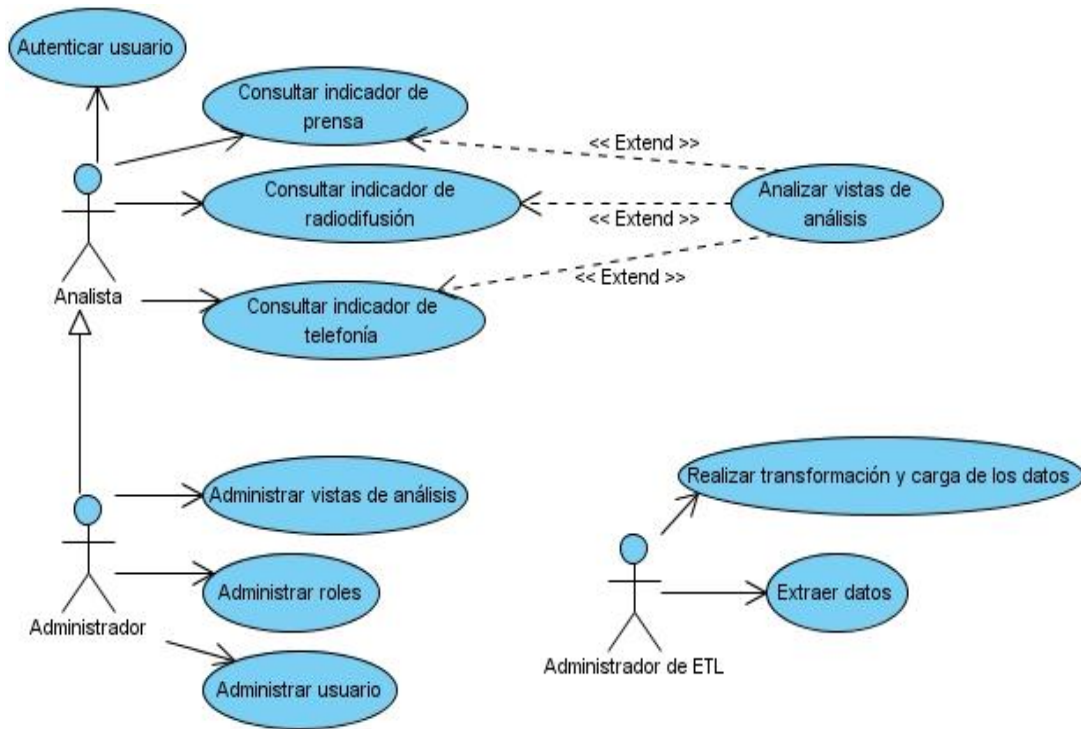
1. Apache Tomcat. [En línea] [Citado el: 10 de Noviembre de 2010.] <http://tomcat.apache.org>.
2. Bill Inmon vs. Ralph Kimball. 1keydata. [En línea] <http://www.1keydata.com/datawarehousing/inmon-kimball.htm>.
3. Cabrera, María Evelia Casales. Data Wharehouse. 2009.
4. Data Cleaner. Data Cleaner Official Site. [En línea] [Citado el: 25 de Octubre de 2010.] <http://datacleaner.eobjects.org/>
5. Gascón, Manuel de la Herrán. Eside. [En línea] [Citado el: 3 de Noviembre de 2010.] <http://www.eside.deusto.es/profesores/mherran/>.
6. Ibarra, María De Los Ángeles. Procesamiento Analítico en Línea. 2005.
7. Inmon, William H. Building the data warehouse. 2002.
8. Integración de Datos. [En línea] [Citado el: 7 de Noviembre de 2010.] <http://www.datati.es/cuadros-de-mandos-%C2%BFque-integracion-de-datos-es-la-mas-adecuada/>.
9. Kimball, Ralph. The Data Warehouse Toolkit. 2009.
10. Modelamiento del Negocio. [En línea] [Citado el: 26 de Octubre de 2010.] http://eva.uci.cu/file.php/102/Curso_2010-2011/Clases/Semana_04/Conferencia_6/Materiales_complementarios/Fase_de_Inicio._Disciplina_de_Modelamiento_del_Negocio.pdf.
11. Mondrian Schema Workbench. [En línea] [Citado el: 3 de Noviembre de 2010.] <http://wiki.pentaho.com/display/PMOPEN/Mondrian+Schema+Workbench>.
12. Pentaho Architecture Layers of a Mondrian System. [En línea] [Citado el: 27 de Octubre de 2010.] <http://mondrian.pentaho.org/documentation/architecture.php>.
13. Pgadmin. Pgadmin Official Site. [En línea] [Citado el: 10 de Noviembre de 2010.] <http://pgadmin.org>.

Bibliografía

14. Plataforma Pentaho BI. Seguridad. [En línea] [Citado el: 27 de Octubre de 2010.] <http://wiki.pentaho.com/display/ServerDoc1x/03.+Platform+Security+Implementation>.
15. Postgres. Postgres Official Site. [En línea] [Citado el: 10 de Noviembre de 2010.] <http://postgres.org>.
16. Sinnexus. [En línea] [Citado el: 27 de Octubre de 2010.] <http://www.sinnexus.com>.
17. TODO BI, "Informe Business Intelligence". Recopilación de los mejores artículos de inteligencia de negocio del 2006. Abril. 2006.
18. Torres, Lilian Vegas y Luis, Rojas Diaz y Cecilia, Placeres Villar. La inteligencia de negocio. Su implementación mediante la plataforma Pentaho. La Habana : s.n., 2009.
19. Transformación grafica ETL con Spoon – un componente de Pentaho Data Integration (Kettle). [En línea] [Citado el: 27 de Octubre de 2010.] Transformación grafica ETL <http://www.gravitar.biz/index.php/bi/introduccion-pentaho-parte-1/#comment-46>.
20. Visual Paradigm. [En línea] [Citado el: 13 de Octubre de 2010.] <http://www.visual-paradigm.com>.

Anexos

Anexo 1: Diagrama de casos de uso del sistema



Anexo 2: Descripción del CU Analizar indicador de prensa

Caso de Uso:	Analizar indicadores de prensa.
Tipo:	Información.
Actores:	Analista.
Resumen:	El CU inicia cuando el actor entra al sistema, selecciona el reporte de prensa que desea analizar. El caso de uso termina cuando se muestra el reporte correspondiente.
Precondiciones:	El usuario debe estar autenticado y se debe a ver realizado la carga de los datos.
Referencias:	RI1, RI2, RI3, RI4, RI5, RI6, RI7, RI33, RI34, RI35, RI36, RI37, RI38, RI47, RI48, RI49, RI50, CU9.
Prioridad	Crítico.

Flujo Normal de Eventos

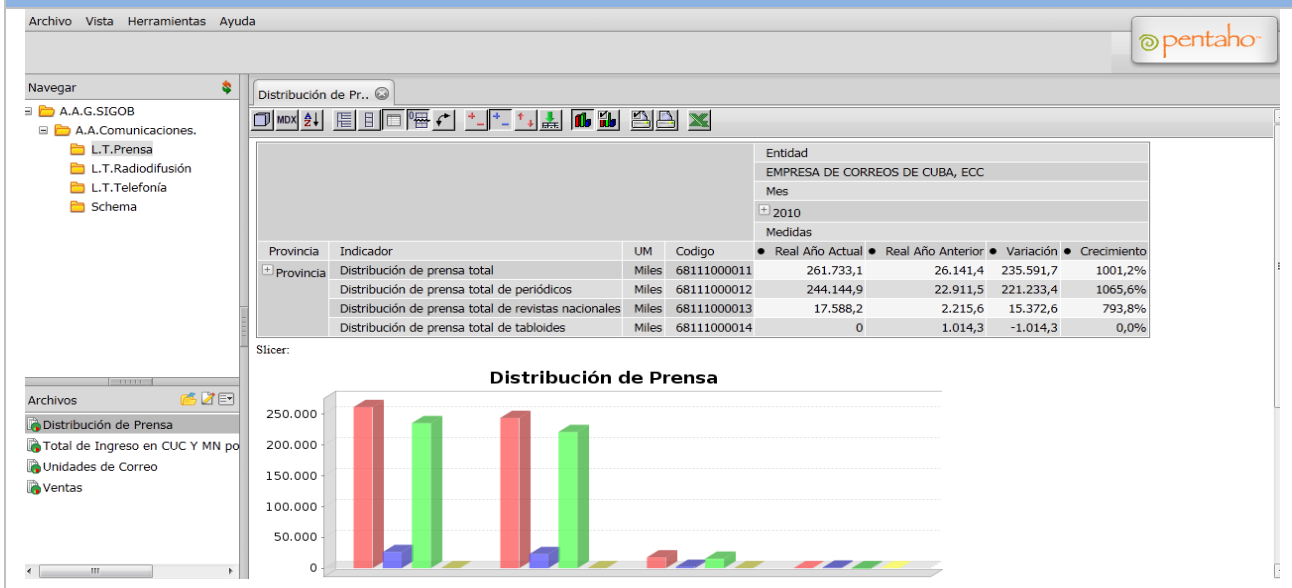
Acción del Actor	Respuesta del Sistema
1. El actor se autentica y entra al sistema.	2. El sistema muestra interfaz principal con las A.A.G

	existentes.
3. El actor selecciona el A.A.G.SIGOB que desea analizar.	4. El sistema muestra las A.A que contiene el A.A.G seleccionado.
5. El actor selecciona el A.A.Comunicaciones que desea analizar.	6. El sistema muestra los L.T que contiene esta A.A.
7. El actor selecciona el L.T.Prensa que desea analizar.	8. El sistema muestra opciones del reporte.
9. Selecciona reporte de prensa.	10. Muestra la información contenida en el reporte de prensa seleccionado y brinda la posibilidad al analista de hacerle cambios al reporte para analizarlo de otra vista de análisis. Ir CU9. Finaliza el caso de uso.

Opciones de Reportes de Prensa

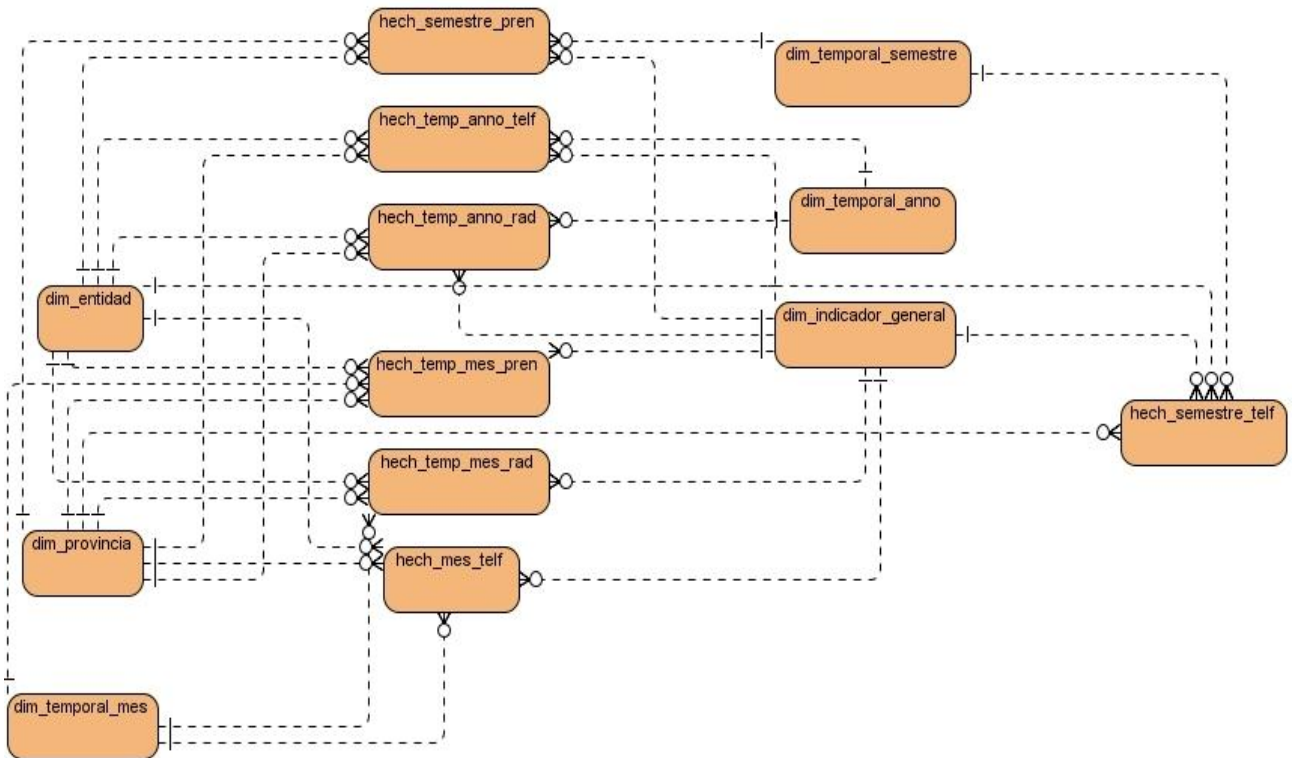
Entradas	Posibles resultados	
	Salidas	Periodicidad
Variables de entrada disponibles relacionadas con el caso de uso Analizar Indicador de Prensa: <ul style="list-style-type: none"> ✓ Indicadores ✓ Provincias ✓ Entidades ✓ Meses ✓ Semestres 	Variables de salida disponibles en el caso de uso Analizar Indicador de Prensa: <ul style="list-style-type: none"> ✓ <i>Real año actual</i> ✓ <i>Real año anterior</i> ✓ <i>Variación</i> ✓ <i>Crecimiento</i> 	Rango de tiempo en que se solicitan las variables de salida: <ul style="list-style-type: none"> ✓ Mensual. ✓ Semestral.

Prototipo de Interfaz



Poscondiciones Los reportes de prensa han sido analizados por el analista.

Anexo 3: Diseño del Modelo Lógico



Anexo 4: Análisis de todas las tuplas del perfilado de datos

Standard measures																									
	c02	org	ff2	for	mod	ram	sin	uni	nae	ep	v01	fil	esf	sub	fpr	ff1	pro	v02	c01	cae	emp	sec	dpa	eat	
Row count	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078
Null values	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Empty values	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Highest value	8626775.0	999	4	99.000909	9999	99	999189999	03	3	00007090	2	4	2	3	99	003540160.0	999999999001	99	9901	3					
Lowest value	0.0	100	1	2.000909	0101	01	001010211	00	1	00004795	1	0	1	1	01	00	0.0	01010000171	01	0108	1				

Anexo 5: Análisis de los valores de tipo string del perfilado de datos

String analysis																							
	org	ff2	for	mod	ram	sin	uni	nae	ep	v01	fil	esf	sub	fpr	ff1	pro	v02	cae	emp	sec	dpa	eat	
Char count	12234	4078	8156	24468	16312	8156	12234	24468	8156	4078	32624	4078	4078	4078	4078	8156	8156	24468	20390	8156	16312	4078	
Max chars	3	1	2	6	4	2	3	6	2	1	8	1	1	1	1	2	2	6	5	2	4	1	
Min chars	3	1	2	6	4	2	3	6	2	1	8	1	1	1	1	2	2	6	5	2	4	1	
Avg chars	3	1	2	6	4	2	3	6	2	1	8	1	1	1	1	2	2	6	5	2	4	1	
Max white spaces	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Min white spaces	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Avg white spaces	0	0	0,92	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Uppercase chars	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	
Lowercase chars	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	
Non-letter chars	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	
Word count	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	4078	
Max words	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
Min words	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	

Anexo 6: Análisis de los valores de tipo numérico del perfilado de datos

Number analysis		
	c02	c01
Highest value	8626775	3540160
Lowest value	0	0
Sum	26250347	13287153,1
Mean	6437,06	3258,25
Geometric mean	0	0
Standard deviation	196192,24	73107,07
Variance	38491395742,77	5344643909,28

Anexo 7: Lista de Chequeo

Estructura del documento					
Peso	Indicadores a evaluar	Eval	NP	CEA	Comt
C	1. ¿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)	1		4	
Indicadores definidos por la etapa					
Peso	Indicadores a evaluar	Eval	NP	CEA	Comt
	1. ¿La arquitectura satisface las necesidades del proyecto?	0		0	
	2. ¿La arquitectura soporta el incremento del proyecto?	0		0	
	3. ¿Se utilizó el menor número de transformaciones posibles al cargar los datos hacia el área de trabajo?	0		0	
C	4. ¿Se creó el Modelo Físico a partir del Modelo Lógico?	0		0	
C	5. ¿Cumple la implementación del proceso de ETL con la arquitectura definida?	0		0	
	6. ¿Se tuvo en cuenta los formatos fuentes y tipos de datos de las perspectivas de análisis?	0		0	
	7. ¿Se realiza una limpieza de los datos antes de realizar la carga de los mismos?	0		0	
	8. ¿La arquitectura satisface las necesidades del proyecto?	0		0	

	9. ¿Se utilizó un lenguaje cuyas sentencias son expresables mediante una sintaxis bien definida?	0			
	10. ¿Se realizó una interfaz amigable para hacer consultas?	0		0	
C	11. ¿Los reportes son configurables a través de la interfaz del sistema?	0		0	
	12. ¿El rendimiento de los reportes no se afecta cuando el número de dimensiones del modelo se incrementa?	0		0	
	13. ¿Presenta la capacidad de crear todo tipo de dimensiones con funcionalidades aplicables de una dimensión a otra?	0		0	
	14. ¿La interfaz está orientada a facilitar el uso de las funciones del sistema por parte de los usuarios?	0		0	
C	15. ¿No existen restricciones para construir cubos OLAP con dimensiones y niveles de agregación ilimitados?	0		0	
C	16. ¿Los usuarios son capaces de realizar nuevos reportes?	1		x	
	17. ¿El sistema responde de una forma rápida y veraz a la información que le sea solicitada por el usuario?	0		0	
C	18. ¿El sistema refleja cualquier lógica del negocio para poder responder a preguntas específicas?	0		0	
C	19. ¿El sistema garantiza la confidencialidad y seguridad de acceso a los datos por rol de los usuarios?	0		0	
	20. ¿Los datos e información derivados del proceso de análisis realizado mediante la aplicación, apoyan la toma de decisiones en la Institución?	0		0	

C	21. ¿Los cambios en los datos se reflejan automáticamente en los reportes de forma instantánea?	0		0	
Semántica del documento					
Peso	Indicadores a evaluar	Eval	NP	CEA	Comt
C	1. ¿Se han identificado errores ortográficos en los entregables?	1		1	
C	2. ¿Se entiende claramente lo que se ha especificado en el documento?	0		0	
	3. ¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?	0		0	

Glosario de términos

- **BD:** Base de Datos, conjunto de datos almacenados de forma organizada que pertenecen a un mismo contexto.
- **BI (Business Intelligence):** Inteligencia de Negocio, habilidad para transformar los datos en información, y la información en conocimiento.
- **DWH (Datawarehouse):** Almacén de Datos, colección de datos orientados a temas, integrado, no volátil, de tiempo variante, que se usa para el apoyo del proceso de toma de decisiones gerenciales. (Ver Capítulo 1).
- **ETL:** Extracción, Transformación y Carga, proceso que se realiza en el desarrollo de un almacén de datos.
- **HOLAP:** Procesamiento Analítico En Línea Híbrido, es una unión de ROLAP y MOLAP.
- **MD:** Mercado de Datos, bases de datos departamentales, especializadas en el almacenamiento de información de un área de negocio específica.
- **MDX:** Expresiones Multidimensionales, lenguaje de consulta a bases de datos multidimensionales a cubos OLAP.
- **MER:** Modelo Entidad-Relación, es utilizado para modelar sistemas de información.
- **MM:** Modelo Multidimensional, basa su funcionalidad en súper consultas que van encaminada a la toma de decisiones.
- **MOLAP:** Procesamiento Analítico Multidimensional En Línea, es una alternativa a la tecnología OLAP pero utiliza una matriz de almacenamiento multidimensional optimizada.
- **OLAP:** Procesamiento Analítico En Línea, solución utilizada para agilizar consultas a grandes cantidades de datos.
- **OLTP:** Procesamiento de Transacciones En Línea, es un sistema que facilita y administra aplicaciones transaccionales.
- **ROLAP:** Procesamiento Analítico En Línea Relacional, es una alternativa a la tecnología OLAP pero utiliza un sistema relacional clásico.
- **UML:** Lenguaje Unificado de Modelado, lenguaje gráfico para visualizar, especificar, construir y documentar un sistema
- **XML:** Lenguaje de Marcas Extensibles, propone como un estándar para el intercambio de información estructurada entre diferentes plataformas.