

# ALMACÉN DE DATOS ESTADÍSTICOS DE LA ONE: DESARROLLO DE LA CAPA DE VISUALIZACIÓN DEL MERCADO DE DATOS DEMOGRAFÍA.

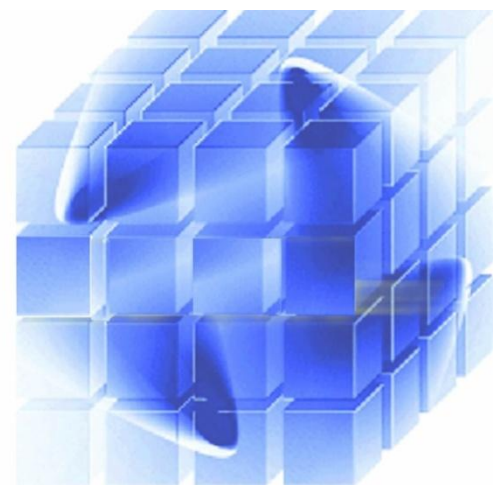
Trabajo de diploma para optar por el título de Ingeniero en Ciencias Informáticas

2009-2010

Universidad de las Ciencias Informáticas

**Autores:** Yoel Díaz Morales  
Ivan Martín Bravo

**Tutoras:** Yamila Mateu Romero  
Anisley Delfino Rodríguez



## DECLARACIÓN DE AUTORÍA

Declaramos ser los autores de la presente tesis y reconocemos a la Universidad de las Ciencias Informáticas, y al Centro de Tecnologías de Gestión de Datos (DATEC) los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste firmo la presente a los \_\_\_\_ días del mes de \_\_\_\_\_ del año \_\_\_\_\_.

**Ivan Martín Bravo**

\_\_\_\_\_

Firma del Autor

**Yoel Díaz Morales**

\_\_\_\_\_

Firma del Autor

**Anisley Delfino Rodríguez**

\_\_\_\_\_

Firma del Tutor

**Yamila Mateu Romero**

\_\_\_\_\_

Firma del Tutor

*Siempre que tengas un sueño estarás viviendo un por ciento más. Bueno, sería si lo consigues haciéndolo tu patria no más. Genial, si en camino a tu sueño entiendes, aplicas y profesas, que Patria es Humanidad.*

## AGRADECIMIENTOS

*De Ambos:*

*A nuestras tutoras Anisley y Yamila por dedicarnos tiempo y esfuerzo, por su asesoramiento en nuestras dudas, por sus críticas que ayudaron al perfeccionamiento del trabajo.*

*A Asnioby por su apoyo y colaboración en todo momento.*

**De Ivan:**

*A mis padres Isela y Efrén por el sacrificio, por su apoyo incondicional, por convertir cada obstáculo en un punto de apoyo.*

*A mi hermana Yanelvys por confiar en mí siempre, por ser la persona que más admiro en la vida, por ser la mejor hermana del mundo.*

*A mi amiga Nodalis por estar siempre, aún estando lejos, por ser la persona que es en mi vida.*

*A mis abuelos Irma y Rolín por tenerme siempre presente en sus vidas.*

*A mi familia por el apoyo que me han brindado durante el transcurso de mi vida.*

*Al 10, Jaime, Yunier, Santiago, Veitia, Jorge Félix, Pepe, Manfred, Pedry, Zussel y Darilkis por haber vivido juntos momentos alegres y difíciles para llegar hasta aquí.*

*De Yoel:*

*A Rafaela de la Caridad Morales Rodríguez, mi madre, por darme la vida.*

*A Joel Antonio Díaz Puente, mi padre, por enseñarme a vivir.*

*A Mi abuela Elena, mi tía Yusimí, mi abuela Teresa. A mi familia toda, a mi abuelo Antonio, a Richard, Vivian, Yamile, Oniria.*

*A mis Maestros todos, A Carmen Lidia, Dalia Gener, Isora Naranjo.*

*A Anisley Delfino Rodríguez.*

*A Pedro Javier, Kelly, Adonis, Dayniel, Manfred, Robier, Gonzalo, por ser los mejores compañeros que tuve, por valorarme con mis defectos y virtudes.*

*A Ivan mi compañero de tesis.*

*A José Antonio León Rosales y Lien Costales Leiva por hacer realidad el verso:*

*Tiene el leopardo un abrigo en su monte seco y pardo,*

*Yo tengo más que el leopardo, porque tengo un buen amigo.*

## DEDICATORIA

*A todos los niños, especialmente a dos.*

*Antonio Díaz Copa, mi hermano y Rachel Camila Miranda Díaz, mi prima.*

*A la memoria de mi abuelo José.*

*A todos los que aman mi deporte de verdad.*

*A mis primos todos.*

*A mi madre y a mi padre.*

10

*A la memoria de mi abuela Dulce.*

*A mis padres y mi hermana.*

*Ivan*

## RESUMEN

Algo peor que no tener información disponible es tener mucha y no saber qué hacer con ella. Con demasiada frecuencia, la transformación y el análisis de toda la información y los datos que las propias empresas generan se convierte en un verdadero problema y, por lo tanto, la toma de decisiones se vuelve un tema complicado.

El desarrollo de soluciones de Inteligencia de Negocios alcanza en nuestros días una elevada importancia debido a que las empresas en el mundo destinan recursos para tratar de obtener de sus datos la mayor información posible.

En Cuba la Oficina Nacional de Estadísticas (ONE) necesita de un mejor aprovechamiento de la información de los datos que almacenan. Estos datos son extraídos de las diferentes fuentes (hospitales, escuelas, etc.) que contribuyen de una manera u otra a que se tenga una visión del comportamiento de la demografía en Cuba.

El presente trabajo de diploma se enmarca en el área de Inteligencia de Negocios y sus técnicas y herramientas para la presentación de los datos contenidos en el Mercado de Datos Demografía que pertenece al Almacén de Datos Estadísticos de la ONE el cual forma parte del proyecto “Sistema de Información de Gobierno”. Comprende una revisión de tendencias y mejores prácticas para el desarrollo de este tipo de soluciones. La investigación toma como referencia la metodología formulada por el Centro de Tecnologías de Gestión de Datos (DATEC).

Como resultados se expone el diseño de un Cuadro de Mando Digital y la implementación de los módulos de reportes OLAP<sup>1</sup> que permiten el análisis de datos; la solución desarrollada se realizó usando herramientas libres.

**Palabras Claves:** datos, información, análisis, Inteligencia de Negocios, Almacén de Datos, Cuadro de Mando Digital, Procesamiento Analítico en Línea, Oficina Nacional de Estadísticas.

---

<sup>1</sup> Del inglés Online Analytical Processing, conocido por sus siglas OLAP



## ÍNDICE DE CONTENIDO

INTRODUCCIÓN .....	1
CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA .....	5
1.1 INTELIGENCIA DE NEGOCIOS. ....	5
1.1.1 Ejemplo del mundo real de aplicaciones de Inteligencia de Negocios .....	7
1.2 INTELIGENCIA INSTITUCIONAL .....	7
1.3 METODOLOGÍAS DE DESARROLLO. ....	8
1.4 HERRAMIENTAS DE INTELIGENCIA DE NEGOCIO.....	12
1.4.1 Procesamiento Analítico en Línea. ....	12
1.4.2 Cuadro de Mando Integral.....	13
1.4.3 Cuadro de Mando Digital.....	15
1.4.3.1 Indicadores Claves de Rendimiento y las Métricas .....	17
1.4.4 Aplicaciones de Reportes.....	17
1.5 HERRAMIENTAS DE SOFTWARE PARA INTELIGENCIA DE NEGOCIOS .....	18
1.5.1 Oracle Business Intelligence Suite Enterprise Edition Plus .....	18
1.5.2 IBM Cognos 8 Business Intelligence.....	19
1.5.3 MicroStrategy 9.....	20
1.5.4 Pentaho BI Suite 3.5 Community Edition.....	21
1.6 CONCLUSIONES DEL CAPÍTULO .....	22
CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA .....	24
2.1 ESTRUCTURA GENERAL DE LA SOLUCIÓN .....	24
2.2 ARQUITECTURA DE LA CAPA DE VISUALIZACIÓN DEL MERCADO DE DATOS DEMOGRAFÍA .....	25
2.3 DEFINICIÓN DE LA AUDIENCIA.....	28
2.4 COLORES E IMÁGENES VISUALES.....	28
2.5 TIPOS DE GRÁFICOS .....	28
2.6 NÚMERO DE VENTANAS DEL CMD .....	29
2.7 SIMETRÍA Y PROPORCIONES DEL CMD .....	29
2.8 INDICADORES CLAVES DE RENDIMIENTO .....	30
2.9 REGLAS DE NAVEGACIÓN.....	32

ÍNDICE DE CONTENIDO

<b>2.9.1 Agrupaciones de Información y Jerarquías</b> .....	33
<b>2.9.2 Elementos de las Reglas de Navegación</b> .....	34
<b>2.9.3 Metadatos del Dominio informativo</b> .....	36
<b>2.10 DISEÑO DE CUBOS MULTIDIMENSIONALES</b> .....	38
<b>2.11 DISTRIBUCIÓN POR ÁREAS DE ANÁLISIS</b> .....	39
<b>2.11.1 Mapa de Navegación</b> .....	40
<b>2.11.2 Descripción de las Áreas de Análisis</b> .....	40
<b>2.11.3 Descripción de los Libros de Trabajo</b> .....	41
<b>2.11.4 Reportes Candidatos: Población</b> .....	42
<b>2.11.5 Reportes Candidatos: Nacimientos</b> .....	42
<b>2.11.6 Reportes Candidatos: Defunciones</b> .....	43
<b>2.11.7 Reportes Candidatos: Divorcios</b> .....	44
<b>2.11.8 Reportes Candidatos: Matrimonios</b> .....	44
<b>2.11.9 Reportes Candidatos: Migraciones</b> .....	44
<b>2.12 GESTIÓN DE ALERTAS EN EL CUADRO DE MANDO DIGITAL</b> .....	45
<b>2.13 PLANTILLAS APROBADAS</b> .....	45
<b>2.14 CONCLUSIONES DEL CAPÍTULO</b> .....	50
<b>CAPÍTULO 3: VALIDACIÓN DE LA SOLUCIÓN</b> .....	<b>52</b>
<b>3.1 PRUEBAS</b> .....	52
<b>3.2 PRUEBAS APLICADAS</b> .....	52
<b>3.2.1 Prueba de Consistencia de los Datos</b> .....	52
<b>3.2.2 Prueba de las estructuras de pre-visualización o estructuras multidimensionales</b> .....	53
<b>3.3 VALIDACIÓN DE LOS REQUISITOS POR EL CLIENTE</b> .....	53
<b>3.4 CONCLUSIONES DEL CAPÍTULO</b> .....	54
<b>CONCLUSIONES GENERALES</b> .....	<b>55</b>
<b>RECOMENDACIONES</b> .....	<b>56</b>
<b>REFERENCIAS BIBLIOGRÁFICAS</b> .....	<b>57</b>
<b>BIBLIOGRAFÍA</b> .....	<b>59</b>
<b>GLOSARIO DE TÉRMINOS</b> .....	<b>62</b>

## INTRODUCCIÓN

La creciente evolución y desarrollo de las Ciencias de la Información le imponen al mundo una nueva forma de concepción para enfrentarse a los problemas que día a día se presentan. Esta nueva forma de conceptualizar las soluciones a estos inconvenientes se van fusionando al aumento de la explotación de las Tecnologías de la Información y las Comunicaciones (TIC) en la sociedad, mostrándose como un requisito indispensable para lograr un desarrollo sostenible e incremental.

El análisis de datos desempeña un papel cada vez más importante en nuestros días. Es por esto que los principales sectores de la sociedad en el mundo aplican la tecnología con el fin de obtener de ellos la mayor información posible.

El control de los datos estadísticos dentro de la infraestructura de un país constituye el eslabón principal para la toma de decisiones en los principales sectores socioeconómicos. La entidad rectora de este tema en el país es la ONE, la cual mediante su Sistema Estadístico Nacional (SEN), organiza, dirige, controla y regula esta actividad. La ONE tiene una estructura institucional distribuida territorialmente en las provincias y municipios del país. Existen 16 oficinas provinciales, una en cada provincia, una en el municipio especial Isla de la Juventud y otra adicional en Ciudad de La Habana; y 169 oficinas municipales, subordinadas a las oficinas provinciales, las cuales son las encargadas de interactuar directamente con los Centros Informantes conocido por sus siglas (CI), siendo estos el último eslabón de la cadena de la actividad estadística. Todas estas oficinas tienen atención administrativa y metodológica por la ONE. La información estadística en Cuba está agrupada en el SEN, que se divide en las siguientes áreas:

- ✓ Sistema de Información Estadística Nacional, conocido por sus siglas (SIEN): Incluye los formularios estadísticos recopilados por la ONE a través de sus dependencias y los ministerios con fines nacionales.
- ✓ Sistema de Información Estadística Territorial, conocido por sus siglas (SIET): Incluye los formularios estadísticos recopilados por la Oficina de Estadística Territorial y entidades territoriales con fines territoriales aprobados por la ONE.

## INTRODUCCIÓN

- ✓ Sistema de Información Estadística Complementaria, conocido por sus siglas (SIEC): Incluye los formularios estadísticos recopilados por todos los ministerios y entidades para sus propios fines aprobados por la ONE. [1]

Producto de la gran cantidad de información que la ONE recibe diariamente, se hace necesario el desarrollo de un Sistema de Integración y Almacenamiento de Datos. Es entonces que a un grupo de especialistas de la Universidad de las Ciencias Informáticas (UCI), en especial al Centro de Tecnologías de Gestión de Datos (DATEC) se le da la tarea de la construcción de un almacén de datos que integre la información que le suministran a la entidad todas sus fuentes. Actualmente dicho almacén se encuentra poblado, apareciendo dificultades para la presentación de la información esencial como consecuencia de los grandes volúmenes de datos almacenados, dificultades para definir y monitorear los Indicadores Claves de Rendimiento<sup>2</sup> que garanticen dar seguimiento y evolución a la estrategia de la entidad y garantizar un mayor análisis y velocidad de respuesta en el acceso a la información. Surge la necesidad de explotar el mismo con el objetivo de garantizar el aprovechamiento de la mayor información posible a extraer de los datos.

Teniendo en cuenta las dificultades planteadas anteriormente, se hace necesario explotar los datos almacenados y transformarlos en información útil a las personas encargadas de tomar decisiones en la definición de estrategias y políticas, para lograr realizar modelos eficaces que garanticen el avance de la sociedad cubana.

Por tanto, surge la necesidad de darle solución a la problemática planteada, definiendo como **problema a resolver** ¿Cómo presentar los datos integrados en el mercado de datos Demografía mediante la Inteligencia de Negocios<sup>3</sup> para el apoyo a la toma de decisiones?

Tomando como **objeto de estudio** la Inteligencia de Negocios, centrando su **campo de acción** en las técnicas y herramientas de Inteligencia de Negocios para el análisis y presentación de datos en Cuadros de Mando Digitales (CMD).

---

<sup>2</sup> Del inglés Key Performance Indicators, conocido por sus siglas KPIs

<sup>3</sup> Del inglés Business Intelligence, conocido por sus siglas BI

## INTRODUCCIÓN

Para dar respuesta al problema planteado se define como **objetivo general**: Diseño de un CMD e implementación de un módulo de reportes OLAP que permitan el monitoreo de las métricas y KPIs.

Para cumplir con el objetivo anteriormente expuesto y darle solución a la situación problemática que se plantea, se proponen las siguientes **tareas de investigación**:

- ✓ Elaboración del marco teórico de la investigación.
- ✓ Definición de los KPI.
- ✓ Definición de la arquitectura de la información.
- ✓ Definición del sistema de alertas.
- ✓ Modelación de la estructura multidimensional.
- ✓ Implementación del sistema de reportes que apoyan los indicadores claves de rendimiento.
- ✓ Validación de la solución propuesta.

Para dar solución al problema, se plantea la siguiente **idea a defender**: La mejor visualización de los datos integrados en el Almacén de Datos Estadísticos de la ONE debe contribuir al apoyo de la toma de decisiones ajustadas a la realidad del país.

Para dar cumplimiento a las tareas propuestas anteriormente se van a utilizar los métodos científicos de la investigación: **Teóricos** y **Empíricos**.

Entre los métodos teóricos utilizados para la investigación se emplean:

El método **Análisis Histórico – Lógico**, el mismo permite estudiar de forma analítica la trayectoria histórica real de las soluciones de Inteligencia de Negocios, su evolución y desarrollo.

El método **Analítico – Sintético**, el cual permite buscar la esencia de los fenómenos que describen la Inteligencia de Negocios, los rasgos que la caracterizan y distinguen, permitiendo la extracción de los elementos más importantes que se relacionan con el objeto de estudio.

De los métodos empíricos se utiliza:

El método **Entrevista**, el mismo permitió obtener información valiosa sobre las necesidades que existen en la ONE en cuanto al análisis y presentación de los datos almacenados en el almacén de estadísticas.

## INTRODUCCIÓN

El contenido de esta investigación está desglosado en 3 capítulos, conclusiones generales, recomendaciones, bibliografía utilizada, un glosario de términos y por último los anexos que complementan el cuerpo del trabajo y que son necesarios para su entendimiento.

En el capítulo 1 "**Fundamentación Teórica**" se incluyen los resultados del estudio del estado del arte de la investigación, se describen los conceptos principales que se van a tratar en el desarrollo del trabajo, selección de la metodología que se aplica, así como el análisis de las herramientas que se van a utilizar para el tratamiento y análisis de información.

En el capítulo 2 titulado "**Descripción de la Solución Propuesta**" se especifica la arquitectura de la Capa de Visualización del Almacén de Datos Estadísticos de la ONE, se modela el esquema multidimensional de la solución identificando los KPIs, se definen las plantillas que serán usadas por los reportes y se presenta el módulo de reportes construido que soportará el CMD.

En el capítulo 3 "**Validación de la Solución**" se describe la forma en que se validó que la solución cumpliera con los requerimientos y necesidades de los usuarios a los cuales está destinada la misma y que esta contribuyera con el apoyo a la toma de decisiones en la ONE.

## **CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA**

### ***Introducción***

En la época actual la información es la clave para obtener una ventaja competitiva en el mundo de los negocios. Para mantenerse competitiva una empresa, los gerentes y tomadores de decisiones requieren de un acceso rápido y fácil a información útil y valiosa. Lo que se busca es encontrar información que no solamente conteste preguntas de lo que pasó o lo que está sucediendo en el negocio sino que también a través de esta información se construyan modelos en los que se ayude a predecir futuros eventos. Una forma de solucionar este problema es por medio del uso de Inteligencia de Negocios.

Durante el desarrollo de este capítulo se describen aspectos teóricos relacionados con la evolución de las tecnologías utilizadas para el tratamiento y análisis de información existentes en el mundo, se presentan las herramientas apropiadas para la explotación de un almacén de datos, así como algunos conceptos importantes los cuales contribuyen a un mejor entendimiento de la investigación.

### ***1.1 Inteligencia de Negocios.***

El término Inteligencia de Negocios no escapa a la variedad de interpretaciones. Sin embargo, queda esencialmente claro que: “no es una metodología, software, sistema o herramienta específica, es más bien una colección de tecnologías que van desde arquitecturas para almacenar datos, metodologías, técnicas para analizar información y software, entre otros, con un fin común para el apoyo a la toma de decisiones”. [2]

De acuerdo con el Data Warehousing Institute, la definición de Inteligencia de Negocios es la siguiente:

“...Son los procesos, tecnologías, y herramientas que se necesitan para convertir los datos en información, la información en conocimiento, y el conocimiento en planes que impulsan acciones rentables para el negocio. La Inteligencia de Negocios abarca el almacenamiento de datos, herramientas analíticas, y contenido y gestión del conocimiento...” [3]

La Inteligencia de Negocios apoya a los tomadores de decisiones con la información correcta, en el momento y lugar correcto, lo que les permite tomar mejores decisiones de negocios. La información adecuada en el lugar y momento que se necesita incrementan la efectividad de cualquier empresa.

La Inteligencia de Negocios es conformada por todas las actividades relacionadas con la organización y entrega de información así como el análisis del negocio. Entre estas actividades se encuentran:

## CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA

- ✓ Sistemas de Almacenamiento de Datos (Almacén de Datos y Mercados de Datos).
- ✓ Aplicaciones Analíticas.
- ✓ Minería de Datos.
- ✓ Herramientas de Procesamiento Analítico de Datos (OLAP).
- ✓ Herramientas de Consulta y Reporte de Datos.
- ✓ Herramientas de Producción de Reportes Personalizados.
- ✓ Herramientas de Extracción, Transformación y Carga de Datos (ETL<sup>4</sup>).
- ✓ Herramientas de Administración de Sistemas.
- ✓ Portales de Información Empresarial.
- ✓ Sistemas de Bases de Datos.
- ✓ Sistemas de Administración del Conocimiento.

Estas soluciones proporcionan amplias ventajas al permitir que los datos se conviertan en un centro de beneficios, facilitando el análisis de información útil a las empresas, apoyando el proceso de toma de decisiones.

Las principales ventajas que ofrece una solución de Inteligencia de Negocio son:

- ✓ Control de costes, al tener un solo sistema que permite manejar fácilmente los distintos programas que se encuentran en los diferentes departamentos de su compañía.
- ✓ Mejora de la colaboración y la calidad de las decisiones, facilitando el acceso a la información en todos los niveles de la organización.
- ✓ Orienta las soluciones tecnológicas hacia el usuario, porque reduce los tiempos de aprendizaje mediante el uso de herramientas de uso cotidiano.
- ✓ Proporciona una profunda visión del negocio a través de un sistema integrado: Cuadros de Mando Integrales, Cuadros de Mando Digitales, Reportes, OLAP y Minería de Datos.

---

<sup>4</sup> Del inglés Extraction, Transformation and Loading, conocido por sus siglas ETL



- ✓ Asiste a los ejecutivos para planear y pronosticar el trabajo, presentando una descripción común de los procesos del negocio de una compañía.

### **1.1.1 Ejemplo del mundo real de aplicaciones de Inteligencia de Negocios**

Twentieth Century Fox utiliza Inteligencia de Negocios para predecir qué actores, argumentos y películas serán populares en cada vecindario. Evitando ciertos argumentos en cines específicos, la compañía tiene ahorros de aproximadamente 100 millones de dólares alrededor del mundo cada año. Esa misma tecnología la utilizan para seleccionar los "trailers" (cortos previos a la presentación de una película) alternativos para cada película en cada cine y así maximizar las ventas. Una película puede tener varios trailers diferentes, cada cual puede percibirse de diferente forma por cada tipo de audiencias. [4]

Así como este caso, existen una gran cantidad de casos de aplicación exitosa de sistemas de Inteligencia de Negocios, que han brindado una verdadera inteligencia al negocio, proporcionándole no solo una notoria eficiencia y ventaja competitiva, sino que a la larga puede ser la diferencia entre la supervivencia o desaparición de la empresa.

### **1.2 Inteligencia Institucional**

La Inteligencia Institucional es una especialización de la Inteligencia de Negocios, cuya principal diferencia radica en los indicadores que permiten medir el desempeño de las organizaciones y el Retorno de Inversión<sup>5</sup>. La información sobre el desempeño de una organización es la única fuente para evaluar su funcionamiento, y la herramienta primaria para planificar su futuro. Esta actividad no quedó restringida al mundo empresarial; para la administración pública fue creado el concepto Inteligencia Institucional, que gana en precisión y pertinencia al advertir en su desarrollo las complejidades y rasgos característicos de la tarea gubernamental.

Para las instituciones gubernamentales, el ROI no es únicamente un balance de resultados financieros. Su medición tiene indicadores tales como: más y mejor educación, más y mejores empleos, o reducción de la pobreza, por citar algunos ejemplos. De ahí que los datos que sirven para medir el desempeño de los

---

<sup>5</sup> Del inglés Return on Investment, conocido por sus siglas ROI

## CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA

gobiernos no se encuentran en las oficinas públicas, sino en la calle, en las ciudades y pueblos, o dicho de otro modo, en la satisfacción ciudadana y en una mejor calidad de vida para todos. [5]

Una segunda gran diferencia entre la implementación de una estrategia de Inteligencia de Negocios y una de Inteligencia Institucional, es la heterogeneidad de las fuentes de información. Para una empresa, cruzar diferentes datos sobre su operación puede resultar complejo, pero cuando se trata de organismos gubernamentales las fuentes son mucho más diversas. El gobierno es la suma de una gran variedad de dependencias, organismos, empresas paraestatales, organizaciones sociales y ciudadanas, entre otras, que trabajan en distintos ámbitos territoriales.

### **1.3 Metodologías de Desarrollo.**

El vocablo metodología, es la ciencia que estudia los métodos del conocimiento. Se refiere a los métodos o procedimientos de investigación que se siguen para alcanzar una gama de objetivos en una ciencia. Además, puede categorizarse como el conjunto de métodos que se rigen en una investigación científica o en una exposición doctrinal.

Existen 2 criterios bien identificados y que han marcado claramente su tendencia sirviéndole de guía a la comunidad mundial en cuanto al tema de metodologías de desarrollo para soluciones de Inteligencia de Negocios. Estas tendencias son las conocidas Metodología Kimball en honor de su creador Ralph Kimball<sup>6</sup> y Metodología de Inmon dada a su creador Bill H. Inmon<sup>7</sup>. Bill Inmon y Ralph Kimball son dos de las personalidades referentes y más influyentes en el área de Almacenes de Datos, y responsables de los dos enfoques a los que se hace referencia.

---

<sup>6</sup> Ralph Kimball, autor ampliamente publicado sobre el tema de almacén de datos y de inteligencia de negocios. Es considerado el "Gurú del Data Warehousing".

<sup>7</sup> Bill Inmon (1945), experto, orador y autor de los almacenes de datos, de renombre mundial. Es reconocido como "el padre del Data Warehousing".

## CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA

La principal diferencia que existe entre ambas tendencias está basada en la forma de enfrentar el problema. La visión de Inmon se basa principalmente en un enfoque descendente (top-down). Además, 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 del Almacén de Datos corporativo convirtiéndose en un complejo empresarial de bases de datos relacionales.

La metodología de Ralph Kimball se enfoca principalmente en el diseño de la base de datos que almacenará la información para la toma de decisiones. El diseño se basa en la creación de tablas de hechos, es decir, tablas que contengan la información numérica de los indicadores a analizar, o sea, la información cuantitativa de la información para la toma de decisiones. Las tablas anteriores se relacionan con tablas de dimensiones, las cuales contienen la información cualitativa, de los indicadores, es decir, toda aquella información que clasifique la información requerida, a este modelo de datos se le conoce como diseño estrella. [6]

Entre sus características principales tenemos su arquitectura ascendente (bottom-up) debido a que plantea que se debe crear por cada departamento un conjunto de mercados de datos independientes 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.

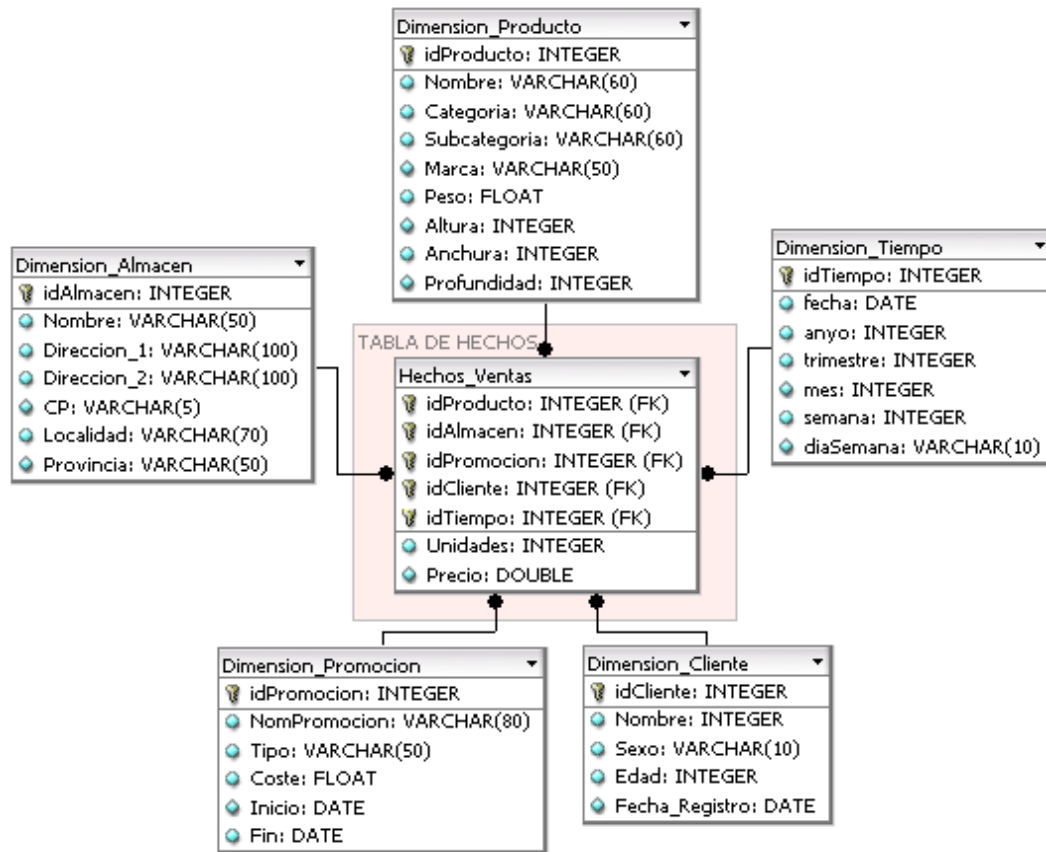


Figura 1 Ejemplo de modelo de datos en estrella

Basados en estas propuestas se han desarrollado un conjunto de metodologías que no siguen obligatoriamente una específica sino que realizan una selección de lo mejor de cada una y definen su propia metodología. A continuación se describen brevemente algunas de ellas.

Metodología **SQLBI**, avalada por Microsoft y orientada totalmente a sus herramientas: Microsoft SQL Server, SQL Server Analysis Services y su oferta más completa en este campo que es Microsoft Suite for Business Intelligence. [6]

La metodología **DM2** se basa en las necesidades de información a nivel gerencial, donde la información debe ser encarada como patrimonio de la empresa, accesible a quien la necesite. Por la propia naturaleza del ambiente, el modelo cumple con su objetivo (atender las necesidades de información del nivel gerencial y ejecutivo de una empresa), esta metodología se asemeja a la forma descendente que propone Inmon, y acorta en función razonable el tiempo entre el inicio del análisis y la implantación. Esta rapidez

## CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA

no solo es buena para el cliente sino que también es exigida y necesaria por el propio ambiente que lo rodea.

**CRISP-DM**, propuesta en 1996 como herramienta industrial y de aplicación neutral. Está descrita en términos de un modelo de proceso jerárquico, consistente en un conjunto de tareas descritas en cuatro niveles de abstracción (de lo general a lo específico): fase, tarea genérica, tarea especializada e instancia de procesos. Principalmente ha sido propulsada por el GIS (Grupo Interesado Especialmente en CRISP-DM, por sus siglas en inglés).

**Hefesto**, entre sus principales directrices plantea que la construcción e implementación de un Almacén de Datos puede adaptarse muy bien a cualquier ciclo de vida de desarrollo de software, con la salvedad de que para algunas fases en particular, las acciones que se han de realizar serán muy diferentes. Lo que se debe tener en cuenta, es no entrar en la utilización de metodologías que requieran fases extensas de reunión de requerimientos y análisis, fases de desarrollo monolítico que conlleve demasiado tiempo y fases de despliegue muy largas. Lo que se busca, es entregar una primera implementación que satisfaga una parte de las necesidades, para demostrar las ventajas del almacén de datos y motivar a los usuarios.

DATEC propone una metodología para este tipo de soluciones, adaptada a su proceso de desarrollo, el cual está basado en Líneas de Productos de Software, y a los lineamientos de calidad exigidos por la UCI, denominándose esta: **Metodología para el Desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio en DATEC**.

La misma cubre todas las fases por las que pasa la construcción de un almacén de datos, desde el levantamiento de información inicial hasta la Capa de Visualización. Es una metodología mixta que reúne elementos de varias metodologías de desarrollo de proyectos de integración de datos, toma como base la Metodología de Kimball. En una primera fase contempla el levantamiento de información a nivel de negocio para identificar los posibles indicadores y aspectos a medir en los análisis, que luego de algunas transformaciones se convierten en los requerimientos de información de entrada y de salida para la solución de integración. [7]

De forma paralela a esta actividad se lleva a cabo un estudio de las fuentes de datos que soportan los datos a cargar. Finalizadas estas dos tareas, se corrobora que la información levantada sobre las necesidades de los clientes esté realmente almacenada en las fuentes correspondientes, para posteriormente, teniendo los requerimientos informativos correctamente definidos, proceder a diseñar la

solución de almacén de datos. Una vez diseñada la estructura del almacén, se realiza la carga de los datos desde las fuentes y posteriormente se implementan los requerimientos de inteligencia de negocio identificados en el levantamiento de información inicial. Las actividades y artefactos de la solución son realizados por 4 grupos que conforman la línea, especializados en componentes específicos de la solución. [7]

Como se evidencia existen diversas metodologías que pretenden dar un acercamiento a una propuesta ideal para el desarrollo de almacenes. Cada autor la orienta a la optimización del rendimiento y a su visión de los principales procesos que se deben tener en cuenta para construir un almacén de datos flexible y dinámico. En dependencia de la problemática presentadas es la política de selección por una y otra metodología.

#### **1.4 Herramientas de Inteligencia de Negocio**

Es un tipo de software de aplicaciones diseñado para colaborar con la inteligencia de negocios en los procesos de las organizaciones; específicamente se trata de herramientas que asisten el análisis y la presentación de los datos. A continuación se presentarán algunas de estas herramientas y sus principales características.

##### **1.4.1 Procesamiento Analítico en Línea.**

El término OLAP, fue propuesto por Edgar Frank Codd<sup>8</sup>, el mismo define a “una tecnología que se basa en el análisis multidimensional de los datos y que le permite al usuario tener una visión más rápida e interactiva de los mismos”.

Lo que hace resaltar OLAP es su capacidad para realizar cálculos de múltiples dimensiones, lo que permite gran variedad de informes y análisis de grandes volúmenes de datos. Este análisis, también conocido como análisis del hipercubo, organiza la información según los parámetros que se consulten, de manera tal que a partir de estructuras multidimensionales que contienen los datos resumidos de Sistemas Transaccionales<sup>9</sup>, o de grandes bases de datos, se obtendrá la información requerida. [8]

---

<sup>8</sup> Edgar Frank Codd: Científico informático inglés (conocido por sus aportes a la teoría de bases de datos relacionales.

<sup>9</sup> Del inglés Online Transactional Processing, conocido por sus siglas OLTP

## CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA

Los sistemas OLAP deben permitir que no transcurra demasiado tiempo entre la necesidad de la información y el resultado, debe tener un lenguaje funcional del negocio, debe posibilitar ser manejado con sencillez, posibilitar también almacenar y compartir los informes y cálculos creados por los usuarios, para conducir a la generación de respuestas medibles para la toma de decisiones del negocio.

De esta manera, las principales características de OLAP son:

- ✓ **Rápido:** La primera regla de Codd se refiere a que el sistema debe ser capaz de responder de una forma rápida y ágil a la información que le sea solicitada por el usuario, el cual no deberá esperar más de cinco segundos a la hora de resolver peticiones sencillas y no más de veinte segundos en las peticiones complejas.
- ✓ **Análisis:** Significa que el sistema debe poder reflejar cualquier lógica del negocio para poder responder a las preguntas específicas y necesidades empresariales.
- ✓ **Compartido:** El sistema deberá proporcionar herramientas que garanticen la confidencialidad de los datos, seguridad de acceso por perfiles de los usuarios.
- ✓ **Multidimensional:** Este es el punto más importante que se podría definir como OLAP en sí mismo. La herramienta deberá proporcionar soporte a cada una de las múltiples jerarquías que puedan existir dentro de la organización de información. Son todos los datos e información derivada de este proceso de análisis, la cual nos permitirá la toma de decisiones.
- ✓ **Información:** Son todos los datos e información derivada de este proceso de análisis, la cual nos permitirá la toma de decisiones.

OLAP es un componente clave en las soluciones de Inteligencia de Negocios. Permite agilizar las consultas de grandes cantidades de datos, para ello utiliza estructuras multidimensionales o Cubos OLAP que contienen datos resumidos de grandes bases de datos. La puesta en práctica de esta técnica y un análisis acertado de la información que se está manejando permiten tomar decisiones más acertadas a la realidad objetiva del negocio.

### 1.4.2 Cuadro de Mando Integral

## CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA

El Cuadro de Mando Integral<sup>10</sup> es un modelo de gestión que ayuda a las organizaciones a transformar la estrategia en objetivos operativos, que a su vez constituyen la guía para la obtención de resultados del negocio y comportamientos estratégicamente alineados del personal de la empresa o institución.

Proporciona a los administradores una mirada global de las prestaciones del negocio. Es una herramienta de administración de empresas que muestra continuamente cuándo una compañía y sus empleados alcanzan los resultados definidos por el plan estratégico. También es una herramienta que ayuda a las compañías a expresar los objetivos e iniciativas necesarias para cumplir con la estrategia. Las ventajas que aporta a la organización el establecimiento de un Cuadro de Mando Integral son las siguientes:

- ✓ Clarifica la estrategia y consigue el consenso sobre ella.
- ✓ Comunica la estrategia a toda la organización.
- ✓ Facilita la convergencia entre los objetivos personales y departamentales con los objetivos de la organización.
- ✓ Vincula los objetivos estratégicos con los objetivos a largo plazo y los presupuestos anuales.
- ✓ Ofrece información sobre factores clave de éxito.
- ✓ Realiza revisiones estratégicas periódicas y sistemáticas.
- ✓ Se retroalimenta para aprender sobre la estrategia y mejorarla. [9]

El Cuadro de Mando Integral permite tanto guiar el desempeño actual como apuntar el desempeño futuro, usa perspectivas en cuatro categorías: desempeño financiero, conocimiento del cliente, procesos internos de negocios y aprendizaje y crecimiento para alinear iniciativas individuales, organizacionales y departamentales e identifica procesos enteramente nuevos para cumplir con objetivos del cliente y accionistas. “Cambia la manera en que se mide y gerencia un negocio”. [10]

Constituye una guía para transformar los propósitos y la estrategia de una organización en objetivos e indicadores concretos. Permite monitorear el estado del negocio y no solo es capaz de medir el momento actual sino los periodos anteriores lo que proporciona establecer una comparación de la evolución de la organización.

---

<sup>10</sup> Del inglés Balanced Scorecard, conocido por sus siglas BSC



### **1.4.3 Cuadro de Mando Digital.**

Un CMD es “una herramienta tecnológica capaz de entregar los datos que la mayoría de los usuarios de nivel estratégico y táctico de las organizaciones requieren, pero sólo los que ellos necesitan y cuando los necesitan”. [11]

Es “una pantalla de visualización de la información, necesaria para lograr uno o más objetivos, consolidada y organizada en una sola pantalla de modo que la información se pueda controlar de un vistazo”. [12]

Las organizaciones demandan elementos que permitan a sus ejecutivos gestionar en forma óptima la información valiosa de su negocio, en este contexto, el uso del CMD se ha convertido en una creciente necesidad en las empresas. Un CMD ofrece la posibilidad de analizar información crítica del negocio en un solo vistazo, valiéndose de adecuadas técnicas de diseño.

La síntesis de un proyecto de Inteligencia de Negocios suele ser la construcción de uno o más Cuadros de Mando. Aunque es solo la cima de todo el trabajo realizado es sin duda la parte más importante pues es lo que el cliente ve y percibe de todo el trabajo anterior.

Algunas de las ventajas de visualizar su operación en un CMD son:

- ✓ Posibilita una visión general e inmediata de la organización a través de informes y gráficos dinámicos.
- ✓ Traduce un modelo de negocios a indicadores claves de desempeño que facilita el consenso en toda la empresa, no sólo de la dirección, y establece los cimientos para obtener lo acordado.
- ✓ Visualiza información parametrizada y permite una avanzada capacidad de análisis en la monitorización y seguimiento de la actividad de las organizaciones.
- ✓ Exterioriza cómo las acciones diarias afectan a la organización no sólo a corto plazo, sino también en el futuro.
- ✓ Permite a la empresa enfocarse en las prioridades, eliminando la sobrecarga de información, presentando solo lo que es relevante. [13]

## CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA

Teniendo en cuenta las numerosas aplicaciones de los CMD dentro de las empresas se ha podido realizar una clasificación primaria que los enmarca en las siguientes categorías:

- ✓ **Cuadro de Mando Digital de Monitoreo de Procesos y Actividades:** Conocido también como Cuadro de Mando Digital Operacional puede implementarse para controlar procesos de negocios específicos o actividades generales. Para el monitoreo de operaciones, se debe mantener el conocimiento de los eventos que están cambiando constantemente y pueden requerir atención y respuesta inmediata, por lo que la visualización de estos Cuadros de Mando debe ser muy simple para no equivocarse su significado en una situación de urgencia.
- ✓ **Cuadro de Mando Digital Analítico:** Soporta análisis de datos por lo que la información exige mayor contexto, como las comparaciones enriquecedoras y evaluadores de función más sutiles. No necesita actualizaciones constantes de los datos y soporta interacciones como drill-down para permitir la exploración necesaria para dar sentido a la información.

Enmarcados en esta categoría se encuentran:

- ✓ **Cuadro de Mando Digital Divisional:** Muestra los parámetros de rendimiento y números específicos de áreas determinadas de la empresa. Son requeridos por los jefes de división y gerentes de operaciones.
- ✓ **Cuadro de Mando Digital de Aplicación:** Son en su mayoría incrustados en aplicaciones personalizadas para proporcionar métricas específicas definidas en la aplicación. Pueden ser tan diversos como las aplicaciones personalizadas y pueden ser definidos por el proveedor de la aplicación o el equipo responsable del desarrollo de la aplicación personalizada.

**Cuadro de Mando Digital para Propósitos Estratégicos:** Provee la visión general rápida que las personas que toman decisiones necesitan para monitorear la salud y las oportunidades del negocio. Se centran en medidas de alto nivel de rendimiento, incluyendo pronósticos futuros. No están diseñados para la interacción que podría ser necesaria para soportar un análisis más detallado, porque esto es rara vez la responsabilidad directa del gerente estratégico. [13]

El Cuadro de Mando Digital resuelve una necesidad crítica de información a los principales directivos de una organización, permite a través de indicadores monitorear factores clave que influirán en los resultados del futuro, derivados de la visión y estrategias de la organización.

### **1.4.3.1 Indicadores Claves de Rendimiento y las Métricas**

Las Métricas y los Indicadores Claves de Rendimiento son la base para construir un CMD de gran despliegue visual, ya que son las herramientas más eficaces para alertar a los usuarios en cuanto a dónde se encuentran parados en relación con los objetivos. Por ello es importante contar con una clara definición de estos elementos que constituyen la base del diseño del Cuadro de Mando. [14]

**Métrica:** es una medida numérica directa, que representa un conjunto de datos de negocios en la relación a una o más dimensiones.

Un KPI es simplemente un indicador que está vinculado a un objetivo. En la mayoría de los casos, un KPI es el estado del un indicador, es decir, si está por encima o por debajo de una meta predeterminada. Los KPIs generalmente se muestran como una tasa o porcentaje y están diseñados para permitir que un usuario de negocios pueda saber instantáneamente si están dentro o fuera de su plan sin que tenga que buscar información adicional. [14]

### **1.4.4 Aplicaciones de Reportes**

Tradicionalmente, los reportes han sido el medio principal para disponer de información. Tanto reportes en papel como en el escritorio, el usuario depende de ellos para comunicar a las personas lo que está ocurriendo en la empresa.

Las herramientas de consulta y reportes orientadas al usuario final son software que aísla la compleja capa técnica propia del lenguaje de sistemas, tal como lenguaje SQL, uniones de tablas y nombres crípticos, al organizar los datos de la terminología de negocios. [15]

Estas herramientas deben poseer algunas utilidades adicionales que faciliten la generación y publicación de reportes, entre ellas se encuentran:

- ✓ **Intuitivo:** Como cualquier herramienta de Inteligencia de Negocios, la característica común es su facilidad de uso e intuición. Con apoyo en interfaces gráficas y visuales, un usuario con una formación estándar podrá hacer uso de una herramienta de este tipo.

## CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA

- ✓ **Seguridad:** Deben brindar seguridad para el acceso a los reportes, tanto a nivel usuario como por grupos e, incluso, en el grado de profundidad de cada usuario a la información. Esto con la idea de que la información privada no sea accesible por cualquier persona.
- ✓ **Navegación:** La interrelación de reportes es también frecuente para generar sistemas basados en reportes y reportes auxiliares. La navegación entre ellos mediante ligas e hipervínculos es una funcionalidad más.
- ✓ **Reportes por Excepción:** Para no consultar una cantidad muy grande de información debe existir la posibilidad de brindar reportes por excepción, es decir, los que no cumplan con las condiciones de generalidad.

### **1.5 Herramientas de software para Inteligencia de Negocios**

En la actualidad se han desarrollado diversas herramientas con el fin de dar un acercamiento a la automatización del diseño, construcción, implementación y mantenimiento de los almacenes de datos. Existen compañías que han marcado hitos en este sentido y que trabajan en la producción de este tipo de soluciones. A continuación se mencionan algunas de las principales y se describen las principales características de sus principales productos.

#### **1.5.1 Oracle Business Intelligence Suite Enterprise Edition Plus**

El gigante Oracle, es una de las principales compañías de este tipo, su principal ventaja es su arquitectura multiplataforma en el desarrollo de sus aplicaciones. Con su producto Oracle Business Intelligence Suite, actualmente en su versión mejorada Oracle Business Intelligence Enterprise Edition Plus puso al mercado una poderosa herramienta que aborda todo el espectro de requisitos de Inteligencia de Negocios. La Suite cuenta con diferentes componentes algunos de los cuales se describen a continuación: [16]

- ✓ Oracle Business Intelligence Server: Es un servidor de Inteligencia de Negocios, diseñado para ser altamente dimensionable, posibilita hacer el valor de estas aplicaciones disponibles para la audiencia más grande posible. Permite mantener un entorno integrado y el almacenamiento

## CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA

centralizado en el servidor, posibilitando la realización de estudios de mercados e Inteligencia de Negocio, hojas de cálculos entre otras funcionalidades.

- ✓ Oracle Business Intelligence Dashboards: En el ambiente de Oracle BI Intelligence Dashboards, el usuario final trabaja con informaciones vivas, permite realizar mapas y gráficos en una pura arquitectura Web. El usuario tiene capacidad completa para navegar, modificar, e interactuar con estos resultados.
- ✓ Oracle Business Intelligence Answers: Permite a los usuarios interactuar con una vista lógica de la información, completamente ajenos de complejidad de estructura de datos también se pueden crear gráficas mediante el mismo.
- ✓ Oracle Delivers: Es una herramienta de alertas en tiempo real activadas bajo unos parámetros predeterminados. Estas alertas se pueden recibir a través de diferentes canales y formatos dependiendo de las preferencias del usuario. Están orientadas a generar acciones o medidas correctoras de forma automática.

Oracle Business Intelligence Enterprise Edition Plus ofrece una serie de elementos independientes que tienen a Oracle BI Server como pieza central y están diseñados para proporcionar una extensa y completa visibilidad de los datos de negocio a amplias audiencias de usuarios, pudiendo acceder de manera autónoma vía web a aquellas informaciones relevantes que les interesan.

### **1.5.2 IBM Cognos 8 Business Intelligence**

IBM<sup>11</sup> Cognos 8 Business Intelligence es una solución de Inteligencia de Negocios que proporciona una gama completa de capacidades de este tema como son: creación de informes, análisis, indicadores, cuadros de mando, planificación, elaboración de presupuestos y previsión, además de integración de datos, en una única arquitectura probada. Los principales componentes que conforman la solución se describen a continuación: [17]

---

<sup>11</sup> Del inglés International Business Machines (conocida coloquialmente como el Gigante Azul) , conocido por sus siglas IBM

## CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA

- ✓ IBM Cognos 8 Business Intelligence Reporting: Esta herramienta permite el acceso a una completa lista de tipos de informes de autoservicio, se adapta a cualquier fuente de datos, y opera desde una única capa de metadatos para ofrecer diversos beneficios como reportes en múltiples idiomas.
- ✓ IBM Cognos 8 Business Intelligence Dashboard: Permite realizar Cuadros de Mando del negocio que ofrecen información compleja de forma rápida. Muestra la información de los diversos sistemas y datos corporativos utilizando indicadores, mapas y gráficos.
- ✓ IBM Cognos 8 Business Intelligence Analysis: Posibilita la exploración guiada y el análisis de información relacionada con todas las dimensiones del negocio, con independencia de donde se encuentren almacenados los datos. Analiza y genera informes a partir de fuentes OLAP y fuentes de datos relacionales basadas en dimensiones.

### **1.5.3 MicroStrategy 9**

MicroStrategy provee software de Inteligencia de Negocios reconocidos mundialmente. El software de MicroStrategy permite realizar análisis ilimitado de datos y Minería de Datos, crear Informes de Negocio y Cuadros de Mando, predecir oportunidades de negocio, mejorar la gestión de operaciones y permitir la toma de decisiones a nivel corporativo.

Su suite de software más reciente se llama MicroStrategy 9, la misma es una plataforma de Inteligencia de Negocios totalmente integrada que hace que el análisis de datos sea más rápido, más sencillo y más cercano a los usuarios de negocio. Entre las principales características de esta plataforma se encuentran:  
[18]

- ✓ Incluye funcionalidades de reportes y analíticas, permitiendo a los usuarios ver los datos en una gran variedad de formatos, desde informes operacionales hasta visualizaciones de alta calidad gráfica, navegar hasta el máximo nivel de detalle para conocer las causas de posibles incidencias, y realizar consultas personalizadas.
- ✓ Gestionar el rendimiento mediante métricas aritméticas y estadísticas y exportar los datos a Excel o PDF.

- ✓ MicroStrategy 9 presenta mejoras de rendimiento y a la capacidad de proporcionar la potencia para realizar análisis interactivos a través de Cuadros de Mando y nuevas aplicaciones

#### **1.5.4 Pentaho BI Suite 3.5 Community Edition**

Pentaho BI Suite es una herramienta de Código Abierto perteneciente a la familia de aplicaciones de Inteligencia de Negocios, la misma ofrece una gama completa de funcionalidades de Inteligencia de Negocios incluyendo las capacidades de consultas y presentación de informes, análisis interactivo, Cuadros de Mando, ETL., Minería de Datos y una plataforma que la han convertido en la suite de código abierto más popular del mundo. La suite incluye: [19]

- ✓ Pentaho Reporting: Permite a las empresas acceder fácilmente a la información, darle formato y distribuirla a sus empleados, clientes y socios. Posibilita flexibilidad en la implementación de sus reportes.
- ✓ Para obtener la funcionalidad de procesamiento OLAP la suite utiliza otras dos aplicaciones: el servidor OLAP Mondrian, que combinado con Jpivot, permiten realizar consultas a Almacenes de Datos, que los resultados sean presentados mediante un navegador de modo que el usuario pueda realizar las actividades típicas de navegación. Mondrian utiliza MDX <sup>12</sup> como lenguaje de consulta, que fue un lenguaje propuesto por Microsoft. Funciona sobre las bases de datos estándar del mercado: Oracle, DB2, SQL-Server, MySQL, PostgreSQL... lo cual habilita y facilita el desarrollo del negocio basado en la plataforma Pentaho.
- ✓ Pentaho Schema Workbench: Es una herramienta para el desarrollo y prueba de cubos OLAP. 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. Con esta aplicación, se puede configurar una conexión con el modelo físico, para luego elaborar el esquema lógico de manera simple y efectiva. Para ello la herramienta ofrece un editor de esquemas con la fuente de datos subyacente para su validación.

---

<sup>12</sup>

*Multi-Dimensional eXpressions*, más conocida por sus siglas MDX.

## CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA

- ✓ Pentaho BI Server: Permite que puedan ejecutarse los informes y aplicaciones, se puede usar como base para construir un sistema propio de Inteligencia de Negocios con Pentaho.
- ✓ Pentaho Report Designer: Es una herramienta para generar todo tipo de documentos empresariales e informes de formato complejo, posee una interfaz unificada que permite un acceso rápido a los objetos de reportes de uso común, maximizando el área de diseño y una buena flexibilidad para la visualización de la información. Permite acceder a datos de múltiples fuentes y mostrarlos en cualquier parte del reporte.
- ✓ Pentaho Designer Studio: Es una herramienta que incluye una colección de editores, visores y módulos de administración integrados, que proporcionan un ambiente gráfico para construir y probar documentos.
- ✓ Plug-in de Pentaho para CMD: Es un API escrita en Java que sabe cómo manejar las solicitudes web para obtener un tablero de instrumentos y agregar funcionalidades personalizadas.

De manera general no existen grandes diferencias entre las propuestas Código Abierto y las propietarias. Todas éstas tienen en común que a partir de una recuperación de una base de datos ofrezcan la posibilidad de la realización de cálculos adicionales, manipulación de los datos y la presentación de la información ante los usuarios en formatos ya sean predefinidos o especificados dinámicamente por los mismos.

De los proveedores de software para Inteligencia de Negocios analizados solo Pentaho saca al mercado una versión gratis de sus herramientas, que sin embargo no cuenta con todas las funcionalidades que posee la versión comercial, este pequeño escenario ilustra lo que pasa actualmente en el mercado de este tipo de software, solo un restringido grupo de empresas desarrollan esta tecnología para ofertarla gratis o con un costo aceptable, el resto solo oferta sus productos a quienes puedan pagarlos, lo que provoca que solo las empresas poderosas puedan retroalimentarse de sus datos para trazar estrategias que los mantengan liderando el mercado y que las pequeñas y medianas empresas sean cada vez un eslabón más bajo en la cadena económica y sean adquiridas por las grandes transnacionales, monopolizando los precios del mercado e incluso incrementando el desempleo.

### **1.6 Conclusiones del Capítulo**



## CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA

Una vez culminado el desarrollo de este capítulo se concluye que:

- ✓ Se realizó un análisis de los principales conceptos que constituyen el estudio del estado del arte de la presente investigación, permitiendo una mejor comprensión de la misma.
- ✓ Se decidió adoptar la Metodología para el Desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio propuesta por DATEC la cual toma como base la metodología de Ralph Kimball, además de ser considerada la más adecuada para la situación que se nos plantea.
- ✓ Se utilizarán herramientas pertenecientes a Pentaho BI Suite 3.5 Community Edition como herramienta de desarrollo de la solución debido a la independencia tecnológica que existe hoy en el país, las posibilidades de análisis y presentación que ofrece la misma.

## **CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA**

### ***Introducción***

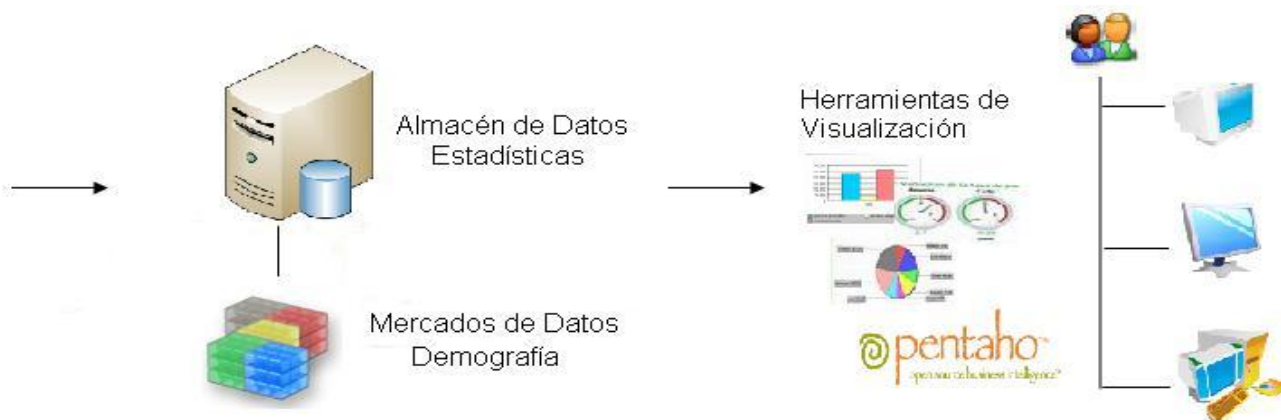
Durante el desarrollo de este capítulo se describe la capa de visualización del mercado de datos de Demografía, mostrándose las potencialidades de las herramientas seleccionadas en el capítulo anterior. Se presenta el diseño de un CMD, con su módulo de reportes OLAP, los principales KPIs y métricas que componen el mismo, además de especificar las áreas de análisis identificadas para agrupar los reportes construidos.

### ***2.1 Estructura General de la Solución***

La solución de Inteligencia de Negocio que se propone parte de los sistemas de origen de la ONE (bases de datos, ficheros de texto...) sobre los que anteriormente fue necesario aplicar una transformación estructural para optimizar su proceso analítico realizando una fase de ETL de los datos. La información resultante, ya unificada, depurada y consolidada, se almacena en el Mercado de Datos de Demografía sirviendo como base para la construcción del Almacén de Datos de Estadísticas. Los datos albergados en el almacén se explotan utilizando herramientas comerciales de análisis de datos para posteriormente ser mostrados a los usuarios encargados de la toma de decisiones en la empresa o cualquier usuario que requiera de dicha información.

La imagen que se muestra a continuación muestra la arquitectura general de la solución de Inteligencia de Negocios que se propone.

CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA



**Figura 2 Estructura General de la Solución de Inteligencia de Negocios**

### **2.2 Arquitectura de la Capa de Visualización del Mercado de Datos Demografía**

La arquitectura del CMD que se propone comprende tres capas partiendo de lo micro a lo macro; estas son Capa Superior, Capa Media y Capa Inferior.

**Capa Superior:** Monitorear y mostrar excepciones. En esta capa el CMD permitirá a los especialistas de la ONE tener un conjunto de métricas y KPIs en la pantalla para su fácil visualización. Los mismos se presentan por iconos visuales, gráficos, o tablas que son actualizados dependiendo de las necesidades del usuario (minutos, horas, días).

Para el desarrollo de esta capa se proponen las siguientes herramientas:

- ✓ Pentaho BI Server
- ✓ Pentaho Designer Studio
- ✓ Plug-in de Pentaho para CMD

**Capa Media:** Análisis y Exploración. En esta capa el usuario tendrá la posibilidad de hacer un análisis más avanzado arrastrando dimensiones y atributos mostrados en el módulo de reportes que se presenta. Estas funcionalidades avanzadas son utilizando cubos multidimensionales (OLAP).

En el desarrollo de esta capa se utilizaron las siguientes herramientas:

- ✓ Pentaho Schema Workbench
- ✓ Mondrian OLAP Server

## CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

### ✓ Apache Tomcat 5.5

Las mismas se utilizan para la presentación de los datos a los usuarios finales, permitiendo que a través de estos los mismos puedan traducir las estrategias de la organización en objetivos operacionales que permitan controlar el comportamiento de éstas mediante indicadores.

Para la realización del módulo de reportes OLAP que soporta el CMD se utilizó una arquitectura multiniveles o multi-tiers, en este tipo de arquitectura se suelen añadir niveles o software adicionales que se encargan de realizar funciones críticas dentro del sistema que se desarrolla. La arquitectura consta de tres capas o niveles fundamentales:

- ✓ Capa de Presentación
- ✓ Capa de Lógica de la Aplicación
- ✓ Capa de Acceso a Datos

En la capa de Presentación los usuarios podrán interactuar con la aplicación a través del protocolo http mediante el servidor web Apache Tomcat 5.5; en la capa de Lógica de Negocio se encuentra el servidor OLAP de Pentaho llamado Mondrian el cual consta de 4 capas lógicas: [20]

- ✓ Capa de Presentación: Es la encargada de gestionar lo que ve el usuario final en su monitor y cómo éste puede interactuar para formular nuevas consultas. Existen muchas formas de presentar los conjuntos de datos multidimensionales, como tablas pivotantes, gráficos, así como herramientas avanzadas de visualización como gráficos dinámicos y mapas interactivos.
- ✓ Capa Dimensional: Es la responsable de analizar gramaticalmente, validar y ejecutar las sentencias MDX que son formuladas por los usuarios.
- ✓ Capa de Estrella: Es la responsable de mantener un caché agregado. Una agregación es un conjunto de valores de medidas en memoria, calificadas por un conjunto de valores de columnas de dimensiones.
- ✓ Capa de Almacenamiento: Es la capa responsable de brindar celdas agregadas de datos y miembros de las tablas dimensionales. En esta capa Mondrian se comunica con el almacén de datos a través de un driver manager similar a como lo hace JDBC <sup>13</sup> para obtener del mismo los datos que requiera para satisfacer las consultas MDX que fueron formuladas por el usuario.

---

<sup>13</sup>

Java Database Connectivity, más conocida por sus siglas JDBC.

## CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

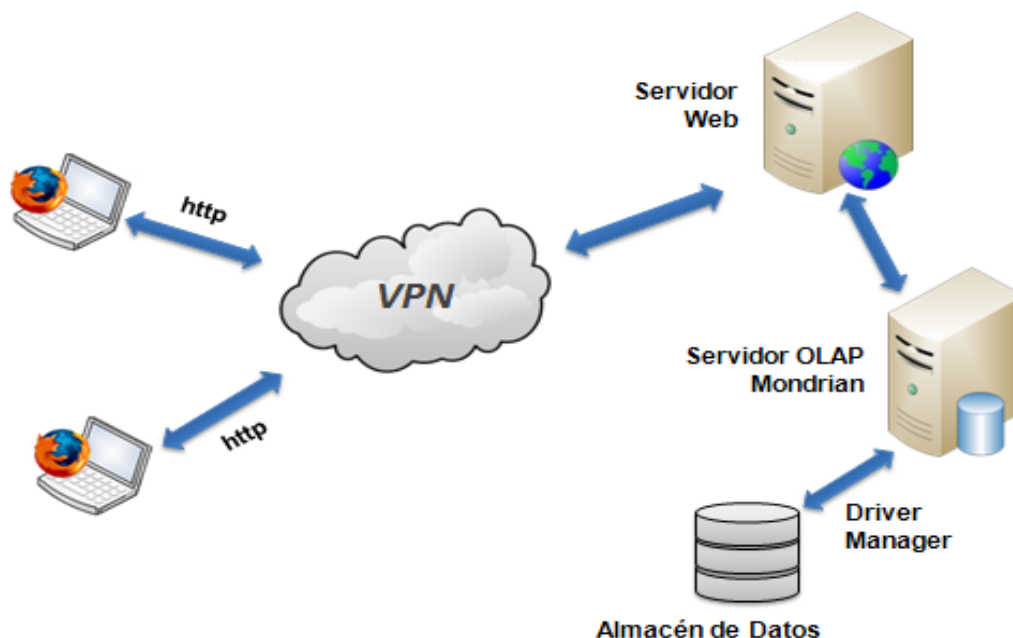
La interfaz de Mondrian se encuentra ubicada en la capa de Presentación de Mondrian y necesita del servidor Web Apache Tomcat para su correcto funcionamiento.

Estas capas pueden existir en una misma máquina, o pueden encontrarse distribuidas entre diferentes máquinas. Las capas 2 y 3, las que comprenden el servidor Mondrian, deben estar en la misma máquina. La capa de persistencia (capa 4) puede encontrarse separada, siendo accedida remotamente vía una conexión JDBC.

El uso de este tipo de arquitectura presenta algunas ventajas como por ejemplo:

- ✓ Las partes críticas de la aplicación se encuentran en el nivel medio, más cercanos a nivel de datos lo que posibilita un acceso más eficiente a los mismos.
- ✓ Solo los datos necesarios son transferidos al cliente, proporcionando menor carga de la red.
- ✓ Facilidad en el cambio de la base de datos.

A continuación se muestra una imagen que representa la arquitectura de la solución que se propone para la Capa de Visualización del Almacén de Datos de la ONE.



**Figura 3 Arquitectura del Módulo de Reportes OLAP**

## CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

**Capa Inferior:** Reportes y Datos Operacionales. En esta capa se presentarán informes planos con información detallada que servirán como soporte para la entrega de informes a ejecutivos.

Para el desarrollo de esta capa se propone la siguiente herramienta:

- ✓ Pentaho Report Designer

### **2.3 Definición de la Audiencia**

A través del CMD se especifican las informaciones relevantes que se le brindan a cada grupo de usuarios. Esta solución está orientada al análisis de comportamientos y tendencias de las principales áreas demográficas, de estas, los usuarios esperan obtener estadísticas que le ayudarán a la toma de decisiones basadas en la realidad del país.

La audiencia que tendrá el sistema dentro de la institución se clasifica de la siguiente forma:

- ✓ Estadísticos Nacionales: Personal de dominio avanzado en técnicas informáticas y alto conocimiento en análisis estadísticos.
- ✓ Estadísticos: No tienen amplio dominio de las técnicas informáticas, pero son especialistas en análisis estadístico.
- ✓ Ingenieros Informáticos: Especialistas con avanzado dominio en las técnicas y herramientas informáticas, pero que no tienen un profundo conocimiento para el análisis de datos estadísticos.
- ✓ Todo el Personal de la ONE: Todos los trabajadores de la institución que en determinado momento requieran consultar algún tipo de información que esté disponible en el CMD.

### **2.4 Colores e Imágenes Visuales**

Para el diseño del CMD fueron definidos diversos colores de acuerdo con el Manual de Entidad de la ONE con el fin de lograr una agradable vista del mismo y representar los distintos aspectos que contendrá este. Los mismos se mencionan a continuación: gris claro, azul claro, blanco y rojo.

### **2.5 Tipos de Gráficos**

## CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

El CMD contendrá los siguientes tipos de gráficos:

- ✓ Gráficos de Columnas: Mediante el mismo se muestran los cambios que han sufrido los datos en el transcurso de un período de tiempo determinado o ilustra las comparaciones entre elementos.
- ✓ Gráficos de Barras: Los gráficos de barras ilustran comparaciones entre elementos individuales.
- ✓ Gráficos de Pastel: Los gráficos circulares muestran el tamaño proporcional de los elementos que conforman un dato, en función de la suma de los elementos. Siempre mostrará una única serie de datos y es útil cuando se desea destacar un elemento significativo.
- ✓ Gráficos de Líneas: Los gráficos de líneas muestran una serie como un conjunto de puntos conectados mediante una sola línea. Los mismos se usan para representar grandes cantidades de datos que tienen lugar durante un período continuado de tiempo.

### **2.6 Número de Ventanas del CMD**

En las 6 ventanas con las que contará el CMD, se visualizarán cada uno de los KPIs definidos por la ONE, en forma de gráficos que fueron descritos anteriormente, los cuales brindan información de la evolución de los ratios en tiempo aproximadamente real.

De esta forma, un alto responsable o cualquier persona que lo necesite en la organización pueden ver al momento cómo evoluciona el negocio y decidir en qué áreas hay que incidir para corregir un comportamiento que podría desviar potencialmente los resultados esperados.

### **2.7 Simetría y Proporciones del CMD**

La simetría y las proporciones de las ventanas son elementos importantes para mantener proporciones adecuadas dentro de las aplicaciones de análisis de datos, teniendo en cuenta esto se utiliza la siguiente simetría.

## CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA



**Figura 4 Simetría**

### **2.8 Indicadores Claves de Rendimiento**

Para la realización de la Capa de Visualización del Mercado de Datos Demografía fue necesario definir un conjunto de Indicadores Claves de Rendimiento a través de los cuales se pueden realizar un conjunto de cálculos asociados a un grupo de medidas presentes en los diferentes cubos que se modelaron. Los indicadores se describen a continuación:

**Población Total:** Cantidad de habitantes que viven en una determinada región o país.

**Tasa de Crecimiento de la Población:** Es la tasa que está aumentando o disminuyendo una población en un año determinado, debido a un déficit de nacimiento en comparación con las muertes expresadas con un porcentaje de la población base.

**Tasa Global de Fecundidad:** Es el número medio de hijos que nacerían vivos durante la vida de una mujer.

**Tasa de Mortalidad Infantil:** Es el número de niños menores de un año que fallecen por cada 1000 nacidos vivos durante un año determinado.

**Grado de Envejecimiento:** Ocurre el envejecimiento cuando descenden las tasas de fecundidad en tanto permanece constante o mejora la esperanza de vida en edades más avanzadas.

**Porcentaje de Población en Edad Laboral:** Porcentaje de personas que se encuentran en edad laboral comprendida en mayores de 15 años y menores que 60.



CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

**Porcentaje de la Población en Edad Escolar:** Porcentaje de personas que se encuentran en edad escolar.

**Grado de Urbanización:** Aumento de la población que reside en zonas urbanas.

**Saldo Migratorio Interno:** La diferencia entre los inmigrantes y emigrantes entre las provincias del país.

Nombre del Indicador	Fórmula del Indicador
<b>Población Total</b>	Número Total de Personas (No tiene Formula)
<b>Tasa de Crecimiento de la Población</b>	$2/t(N_t - N_0) / (N_t + N_0) * 1000$ donde: t = tiempo transcurrido, generalmente un año. N <sub>t</sub> = Población al final del período N <sub>0</sub> = Población inicial
<b>Tasa Global de Fecundidad</b>	Número de nacimientos / Número de mujeres de 20 a 24 años de edad x K donde K = 1.000
<b>Tasa de Mortalidad Infantil</b>	Número de muertes de menores de 1 año durante un año determinado / Nacidos vivos totales de ese año x K donde K = 1.000
<b>Grado de Envejecimiento</b>	$(N_{60 y +} / N) * 1000$ donde: N <sub>60 y +</sub> = Población de 60 años y más N = Población total

CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

<b>Porcentaje de Población en Edad Laboral</b>	(Nm <sub>17-60</sub> +Nf <sub>17-55</sub> )/ N *100 donde: Nm <sub>17-60</sub> = Población masculina de 17 a 60 años. Nf <sub>17-55</sub> = Población femenina de 17 a 55 años. N = Población total.
<b>Porcentaje de Población en Edad Escolar</b>	(N <sub>6-23</sub> /N)*100 donde: N <sub>6-23</sub> = Población de 6 a 23 años N = Población total
<b>Grado de Urbanización</b>	Número de personas que viven en zonas urbanas / N x K donde K = 100 donde: N = Población total
<b>Saldo Migratorio Interno</b>	Número de Inmigrantes – Número de Emigrantes

### 2.9 Reglas de Navegación

A continuación se detallan los elementos que componen las estructuras de navegación de la información que será presentada en la Capa de Visualización del Mercado de Datos Demografía.

Las dimensiones poseen entre sus características principales la definición de jerarquías entre sus atributos las que poseen como objetivo plasmar explícitamente la forma en que se puede consolidar, ya sea mediante el uso de sumas, porcentos, máximos, mínimos, etc., la realización del proceso de análisis en línea de la información.

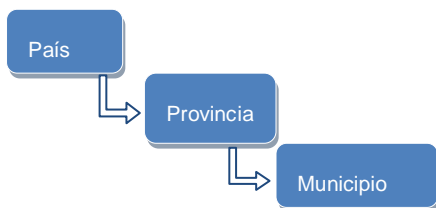
- ✓ DIM\_DPA: Representa formas de análisis de los datos por codificación político-administrativa.
- ✓ DIM\_EDAD: Representa formas de análisis de los datos por grupos de edades.
- ✓ DIM\_TEMPORAL: Representa formas de análisis de los datos por períodos de fechas.
- ✓ DIM\_SEXO: Representa formas de análisis de los datos por sexo.

## CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

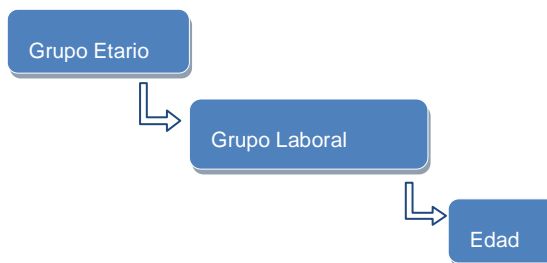
- ✓ DIM\_EDAD\_INFANTIL: Representa formas de análisis de los datos por edades infantiles comprende el período a partir del nacimiento hasta el primer año de vida.
- ✓ DIM\_NUMERO\_HIJOS: Representa formas de análisis de los datos por el número de hijos de la madre.
- ✓ DIM\_OCUPACION: Representa formas de análisis de los datos atendiendo a la ocupación de las personas.
- ✓ DIM\_TIPO\_EMBARAZO: Representa formas de análisis de los datos de acuerdo con el tipo de embarazo de la madre.
- ✓ DIM\_NIVEL\_ESCOLARIDAD: Representa formas de análisis de los datos atendiendo al nivel de escolaridad de la población cubana.
- ✓ DIM\_ESTADO\_CONYUGAL: Representa formas de análisis de los datos de acuerdo con el estado conyugal de las personas.

### 2.9.1 Agrupaciones de Información y Jerarquías

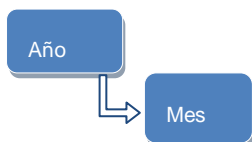
DIM\_DPA:



DIM\_EDAD:



DIM\_TEMPORAL



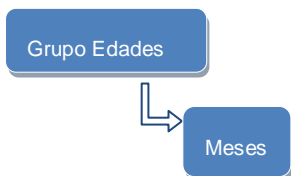
DIM\_SEXO



DIM\_EDAD\_INFANTIL

DIM\_NUMERO\_HIJOS

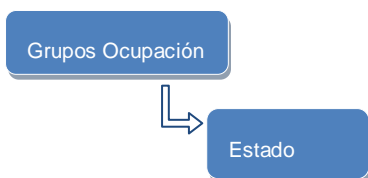
## CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA



DIM\_OCUPACION



DIM\_TIPO\_EMBARAZO



DIM\_NIVEL\_ESCOLARIDAD



DIM\_ESTADO\_CONYUGAL



### 2.9.2 Elementos de las Reglas de Navegación

DIM\_DPA:

- ✓ País: Nivel superior de la dimensión DPA, algunas solicitudes de información se representan a nivel de país para facilitar un correcto análisis de los datos.
- ✓ Provincia: Nivel inmediato inferior de País. Permite visualizar la información más específica a nivel de provincia.
- ✓ Municipio: Nivel inmediato inferior de Provincia. Permite visualizar la información más específica a nivel de municipio.

DIM\_EDAD:

- ✓ Grupo Etario: Nombre de la jerarquía que incluye dos niveles: Grupo Etario y Edad.
- ✓ Grupo Laboral: Nombre de la jerarquía que incluye dos niveles: Grupo Laboral y Edad.

DIM\_TEMPORAL:

## CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

- ✓ Año: Nivel superior de la dimensión TEMPORAL. Algunas solicitudes de información requieren ser representada por año.
- ✓ Mes: Nivel inmediato inferior de Año. Algunas solicitudes de la información requieren ser representadas por mes.

### DIM\_SEXO:

- ✓ Sexo: Algunas solicitudes de información requieren ser representadas por sexo.

### DIM\_EDAD\_INFANTIL:

- ✓ Grupo de Edades: Nivel superior de la dimensión EDAD\_INFANTIL. Algunas solicitudes de información requieren ser representada por diferentes grupos de edades infantiles.
- ✓ Meses: Nivel inmediato inferior de Grupo de Edades. Algunas solicitudes de la información requieren ser representadas por los meses de los nacidos.

### DIM\_NUMERO\_HIJOS:

- ✓ Número de Hijos: Algunas solicitudes de información requieren ser representadas por el número de hijos de la madre.

### DIM\_OCUPACION:

- ✓ Grupos de Ocupación: Nivel superior de la dimensión OCUPACION. Algunas solicitudes de información requieren ser representada por diferentes grupos de ocupación.
- ✓ Estado de Ocupación: Nivel inmediato inferior de grupos de Ocupación. Algunos reportes de información se muestran por el Estado de la Ocupación.

### DIM\_TIPO\_EMBARAZO:

- ✓ Tipo de Embarazo: Algunas solicitudes de información requieren ser representadas por el tipo de embarazo de la madre.

### DIM\_NIVEL\_ESCOLARIDAD:

- ✓ Nivel Escolaridad: Algunas solicitudes de información requieren ser representadas por el nivel de escolaridad que poseen las personas.

CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

DIM\_ESTADO\_CONYUGAL:

- ✓ Estado Conyugal: Algunas solicitudes de información requieren ser representadas por el estado conyugal de las personas.

**2.9.3 Metadatos del Dominio informativo**

Los datos provenientes de las fuentes integradas en el almacén de datos pueden clasificarse atendiendo a diferentes conceptos o variables. Una variable puede ser: edad, sexo, estado conyugal, nivel escolaridad, entre otros. El conjunto total de variables y sus cruces determinan el dominio informativo.

Las variables permiten analizar los diferentes escenarios que pueden tener los datos del almacén. Entre ellos tendríamos las posibles combinaciones de cruces entre las dimensiones en situaciones determinadas, con lo que se analizarían determinados cambios sociales.

Cada una de las conclusiones que aportan estos indicadores, demuestran y guían las decisiones a tomar en cuestiones demográficas. Entre las variables de más influencia en los análisis a realizar están las siguientes:

- Variable temporal: Se refiere a fechas y horarios.
- Variable espacial: Se refiere a la ubicación geográfica: provincia y municipio.

El dominio informativo identificado estuvo compuesto por las siguientes variables.

Variable Original	Metadato Asociado	Procedencia
año	<i>“Año”</i>	DIM_Temporal
mes	<i>“Mes”</i>	DIM_Temporal
provincia	<i>“Provincia”</i>	DIM_DPA
municipio	<i>“Municipio”</i>	DIM_DPA

CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

sexo	<i>“Sexo”</i>	DIM_Sexo
grupo_ocupacional	<i>“Grupo Ocupacional”</i>	DIM_Ocupación
ocupación	<i>“Ocupación”</i>	DIM_Ocupación
n_escolaridad	<i>“Nivel de Escolaridad”</i>	DIM_Nescolaridad
estado_conyugal	<i>“Estado Conyugal”</i>	DIM_Econyugal
grupo_laboral	<i>“Grupo Laboral”</i>	DIM_Edad
grupo_quinquenal	<i>“Grupo Quinquenal”</i>	DIM_Edad
grupo_etario	<i>“Grupo Etario”</i>	DIM_Edad
edad_laboral	<i>“Edad Laboral”</i>	DIM_Edad
edad	<i>“Edad”</i>	DIM_Edad
lugar_ocurrencia	<i>“Lugar de Ocurrencia”</i>	DIM_Locurrencia
semanas_gestación	<i>“Semanas Gestación”</i>	DIM_Sgestación
cant_abortos	<i>“Número de Abortos”</i>	DIM_Abortos_Anteriores
número_embarazos	<i>“Número Embarazos”</i>	DIM_Número_Embarazos
tipo_embarazo	<i>“Tipo de Embarazo”</i>	DIM_Tipo_Embarazo

CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

cantidad_hijos	<i>“Cantidad de Hijos”</i>	DIM_Cantidad_Hijos
gropos_ocupación	<i>“Grupos Ocupación”</i>	DIM_Ocupación_Madre
estado	<i>“Estado”</i>	DIM_Ocupación_Madre
causa_muerte	<i>“Causas de la Muerte”</i>	Dim_Causa_Muerte

**2.10 Diseño de Cubos Multidimensionales**

Para la implementación del módulo de reportes OLAP, es necesario la creación de los cubos multidimensionales los cuales se realizan utilizando la herramienta Pentaho Schema Workbench, la misma permite generar un fichero de configuración XML. En este fichero de esquema se pueden definir las dimensiones, los niveles de jerarquía de dimensiones, los hechos y conexión con el almacén que sirve los datos para el cubo OLAP.

Se modelaron 7 cubos multidimensionales, en los mismos se especificaron las dimensiones, así como un conjunto de características de estos. La siguiente imagen muestra el diseño utilizando la herramienta Pentaho Schema Workbench.



CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

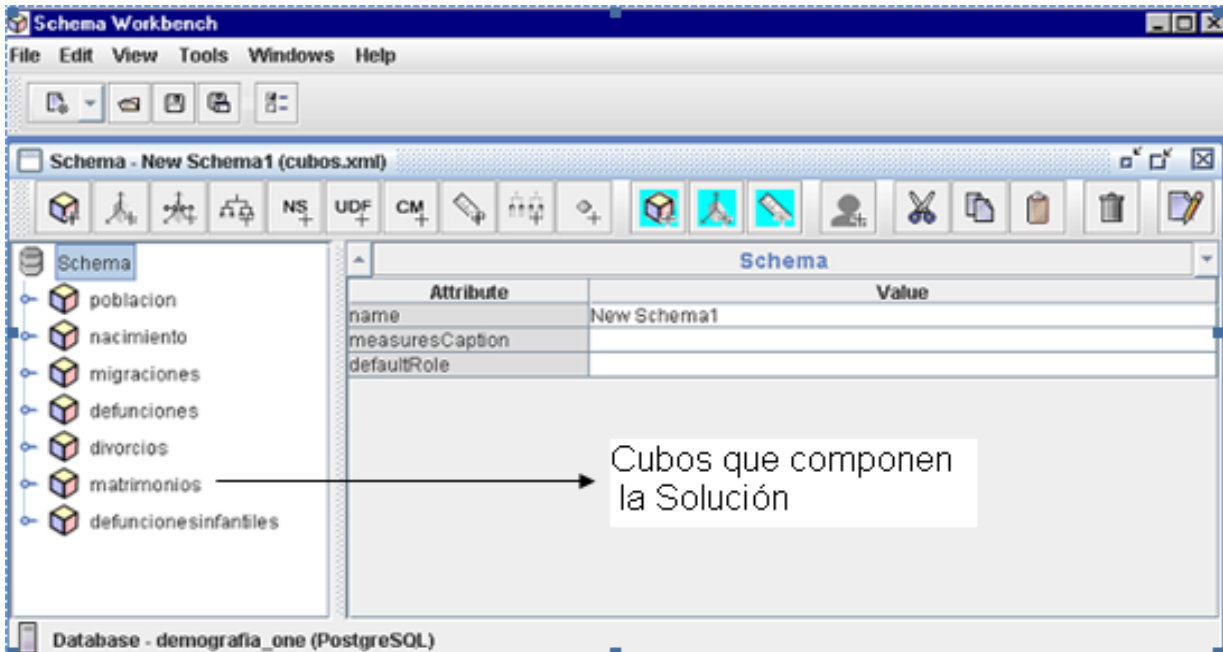


Figura 5 Ejemplo de diseño de los cubos utilizando Pentaho Schema Workbench

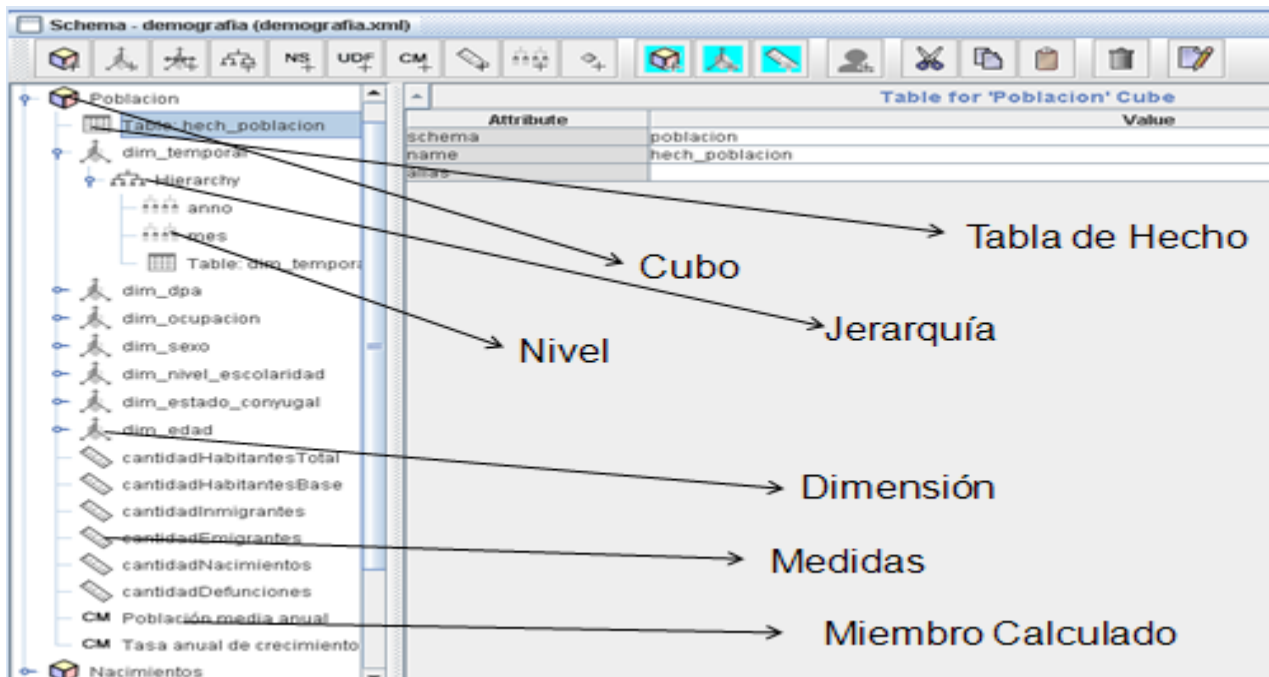


Figura 6 Definición de los elementos que componen los cubos

2.11 Distribución por Áreas de Análisis

## CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

A continuación se detallan los elementos que componen las estructuras de navegación de la información que será presentada en la capa de Visualización del Mercado de Datos Demografía; la misma contiene 6 Áreas de Análisis, igual cantidad de Libros de Trabajos y 60 reportes multidimensionales asociados a los libros de trabajos.

### 2.11.1 Mapa de Navegación

La siguiente imagen muestra el mapa de navegación del CMD en forma de árbol de secciones, niveles y contenidos relacionados con las áreas de análisis, libros de trabajo y reportes que se identificaron en las diferentes áreas de análisis que se describieron anteriormente.

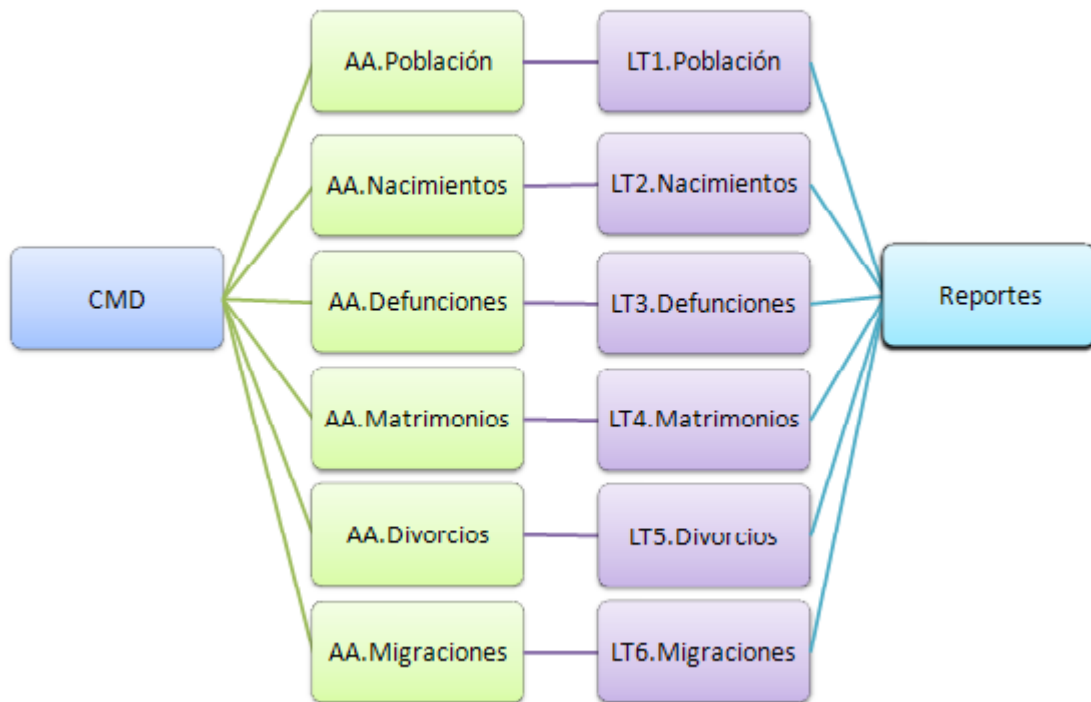


Figura 7 Mapa de Navegación del CMD

### 2.11.2 Descripción de las Áreas de Análisis

## CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

- ✓ A.A Población: Agrupa la información referente al comportamiento de la población cubana, concentra reportes estadísticos que apoyan la concepción de medidas y estrategias encaminadas al bienestar de la población.
- ✓ A.A Nacimientos: Agrupa la información referente al comportamiento de los nacimientos que ocurren en la población cubana, contiene reportes estadísticos que muestran el comportamiento de diferentes indicadores los cuales permiten adoptar políticas que contribuyan al mejoramiento de este sector.
- ✓ A.A Defunciones: Esta área agrupa los reportes relacionados con las defunciones que ocurren en el país, así como datos relacionados con las mismas.
- ✓ A.A Matrimonios: Contiene la información referente al comportamiento de los matrimonios en la población cubana, concentra reportes estadísticos que permiten analizar el comportamiento de diferentes indicadores relacionados con esta área.
- ✓ A.A Divorcios: Contiene la información referente de los divorcios en la población cubana, concentra reportes estadísticos que permiten analizar el comportamiento de diferentes indicadores relacionados con esta área.
- ✓ A.A Migraciones: Agrupa la información referente al comportamiento de las migraciones internas, contiene una serie de reportes estadísticos los cuales permiten obtener datos relacionados con esta área y que pueden ser de mucha importancia para la toma de decisiones en este sector.

### **2.11.3 Descripción de los Libros de Trabajo**

- ✓ LT.1 Población: Libro de trabajo contenido dentro del área de análisis de Población, contiene 10 reportes que permiten el análisis de datos correspondiente a este sector de la sociedad.
- ✓ LT.2 Nacimientos: Libro de trabajo contenido dentro del área de análisis de Nacimientos, soportado por 10 reportes relacionados con esta área.
- ✓ LT.3 Defunciones: Libro de trabajo contenido dentro del área de análisis de Defunciones, incluye un conjunto de 10 reportes los cuales ayudan a la toma de decisiones ajustadas a los principales problemas que se presenten en esta área.
- ✓ LT.4 Matrimonios: Libro de trabajo contenido dentro del área de análisis de Matrimonios, contiene 10 reportes que permiten el monitoreo de los principales KPIs incluidos en dicha área.

## CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

- ✓ LT.5 Divorcios: Libro de trabajo contenido dentro del área de análisis de Divorcios, incluye 5 reportes relacionados con esta área de análisis.
- ✓ LT.6 Migraciones Internas: Libro de trabajo contenido dentro del área de análisis de Migraciones, muestra mediante 5 reportes información relevante relacionada con las migraciones internas.

### **2.11.4 Reportes Candidatos: Población**

El Libro de Trabajo Población consta de 12 reportes o tablas de salidas agrupadas en un libro de trabajos.

A continuación se presentan los mismos:

- ✓ Población Total
- ✓ Tasa Anual de Crecimiento de la Población
- ✓ Población Media Anual Según Provincias
- ✓ Porcentaje de la Población en Edad Laboral
- ✓ Porcentaje de la Población en Edad Escolar
- ✓ Tasa de Nupcialidad de la Población
- ✓ Población Residente Según Provincia por Zonas
- ✓ Grado de Envejecimiento de la Población
- ✓ Grado de Urbanización de la Población
- ✓ Población Residente en las Capitales de Provincia
- ✓ Extensión Superficial y Densidad de la Población
- ✓ Tasa de Divorcialidad de la Población

### **2.11.5 Reportes Candidatos: Nacimientos**

## CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

El Libro de Trabajo Nacimientos está compuesto por 10 reportes a través de los cuales se pueden analizar el comportamiento de los principales indicadores que se relacionen con esta área. Los mismos se encuentran agrupados en un libro de trabajo y se mencionan a continuación:

- ✓ Total de Nacidos Vivos Según Provincia de Residencia de la Madre
- ✓ Total de Nacidos Vivos Según Mes de Ocurrencia
- ✓ Tasa Global de Fecundidad. Años 2007-2009
- ✓ Nacidos Vivos Según Peso al Nacer y Sexo, por Grupo de Edades de la Madre
- ✓ Nacidos Vivos Según Ocupación de la Madre, por Grupos de Edad
- ✓ Nacidos Vivos Según Estado Conyugal de la Madre, por Grupos de Edad
- ✓ Nacidos Vivos Según Edad Combinada de la Madre y el Padre
- ✓ Nacidos Vivos por Sexo y Total por Provincias
- ✓ Nacidos Vivos por Peso al Nacer. Años 2007-2009
- ✓ Nacidos Vivos Según Provincia de Residencia de la Madre por Sexo y Zona

### **2.11.6 Reportes Candidatos: Defunciones**

El Libro de Trabajo Defunciones consta de 10 reportes. A continuación se presentan los mismos:

- ✓ Defunciones Generales Según Provincia de Residencia por Sexo
- ✓ Tasa de Mortalidad Infantil Según Provincia de Residencia
- ✓ Defunciones Infantiles Según Provincia de Residencia
- ✓ Tasa de Mortalidad General Según Provincia de Residencia
- ✓ Defunciones Generales Según Mes de Ocurrencia. Año 2009
- ✓ Defunciones Infantiles Según Mes de Ocurrencia. Año 2009
- ✓ Defunciones Generales Según Provincia de Residencia por Mes de Ocurrencia

## CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

- ✓ Defunciones Generales Según Provincia, por Sitio de Ocurrencia
- ✓ Defunciones Infantiles Según Provincia de Residencia por Edad
- ✓ Defunciones Generales Según Grupos de Edad, por Sexo

### **2.11.7 Reportes Candidatos: Divorcios**

El Libro de Trabajo Divorcios consta de 5 reportes o tablas de salidas agrupadas en 1 libro de trabajo. A continuación se presentan los mismos:

- ✓ Divorcios Concedidos. Años 2007-2009
- ✓ Divorcios Concedidos Según Mes de Ocurrencia. Año 2009
- ✓ Divorcios Concedidos Según Provincia y Mes de Ocurrencia. Año 2009
- ✓ Divorcios Concedidos Según Nivel Educativo de la Mujer. Años 2007-2009
- ✓ Divorcios Concedidos Según Edad de la Mujer al Divorciarse. Años 2007-2009

### **2.11.8 Reportes Candidatos: Matrimonios**

El Libro de Trabajo Matrimonios consta de 10 reportes o tablas de salidas agrupadas en 1 libro de trabajo. A través de estos se pueden analizar información referente a este tema; los mismos se mencionan a continuación:

- ✓ Matrimonios Celebrados Según Mes de Ocurrencia. Año 2009
- ✓ Matrimonios Celebrados en Cuba. Años 2007-2009
- ✓ Matrimonios Celebrados Según Provincia de Residencia de la Mujer. Año 2009
- ✓ Matrimonios Celebrados Según Edad de la Mujer, por orden del Matrimonio. Años 2007-2009
- ✓ Matrimonios Celebrados Según Provincias, por Mes de Ocurrencia. Año 2009

### **2.11.9 Reportes Candidatos: Migraciones**

## CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

El Libro de Trabajo Migraciones consta de 5 reportes o tablas de salidas agrupadas en 1 libro de trabajo. A continuación se presentan los mismos:

- ✓ Saldo migratorio Interno por Provincias. Año 2009
- ✓ Saldo migratorio Interno por Sexo. Años 2007-2009
- ✓ Cantidad de Emigrantes Según Provincia por Sexo. Años 2007-2009
- ✓ Cantidad de Inmigrantes Según Provincia por Sexo. Años 2007-2009
- ✓ Movimiento Migratorio Interno por Provincia Según Nivel de Escolaridad. Año 2009

### **2.12 Gestión de Alertas en el Cuadro de Mando Digital**

El Cuadro de Mando Digital administra los KPIs predefinidos y provee alertas en tiempo real cuando esos KPIs necesitan atención, permite a los usuarios llevar un seguimiento del comportamiento de la organización respecto a sus objetivos estratégicos.

Los indicadores claves de rendimiento que contenidos en el CMD pueden tener interfaces fáciles de entender (indicadores de tendencias, semáforos, velocímetros) asociados a ellos. Si se produce una violación de los KPIs se pueden generar acciones como el cambio de color, la notificación por correo electrónico a las personas encargadas del monitoreo constante de esos indicadores, gracias a lo cual se logra un seguimiento de las métricas las veinticuatro horas del día.

Los colores que se usarán en la semaforización del CMD son el verde, el rojo y el amarillo donde el verde indica un avance favorable, el amarillo indica que vamos bien o a la mitad de nuestros objetivos y el rojo indica que algo no anda bien, que no se ha avanzado y que tal vez debería replantearse alguna estrategia.

El Cuadro de Mando proporcionará información personalizada a cada nivel de la compañía, asegurando unas operaciones más efectivas y una organización más eficiente.

### **2.13 Plantillas Aprobadas**

Después de varios encuentros con personal calificado de la ONE, y de la presentación de varias propuestas de Plantillas de Diseño, quedaron aprobadas las siguientes:

CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA



Figura 8 Portada del Módulo de Reportes de la Capa de Visualización de Demografía

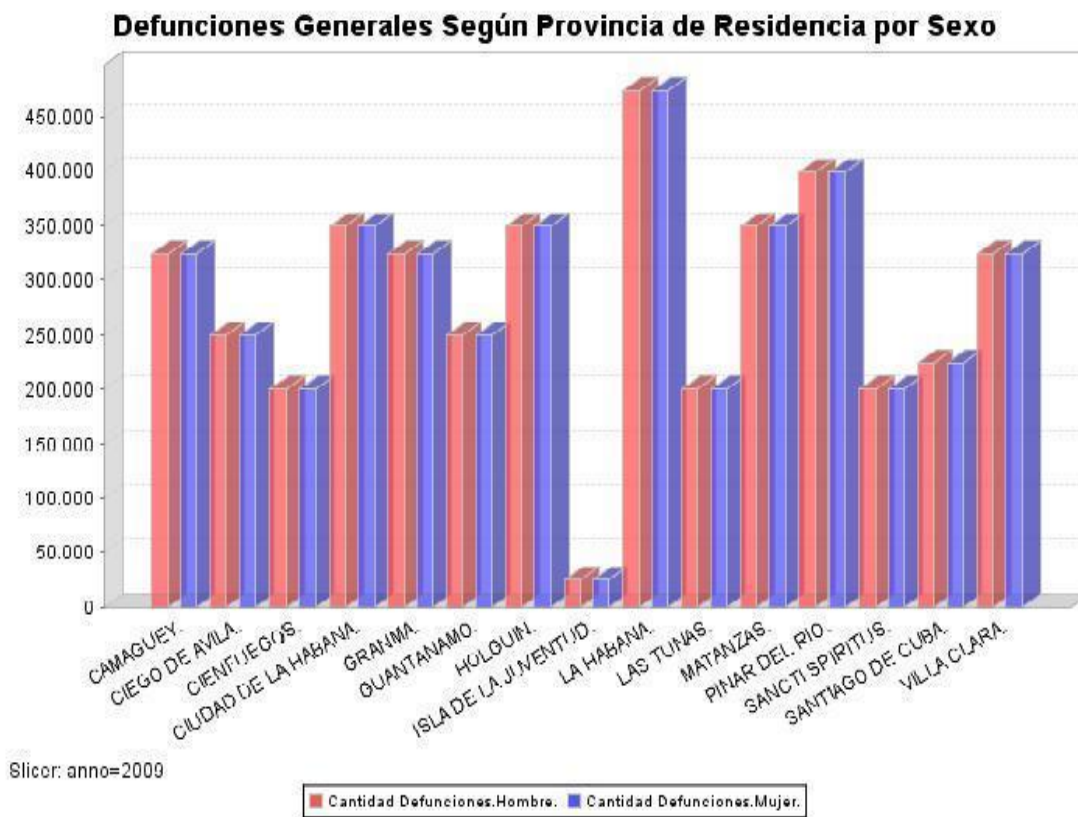


CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA



Área de Análisis: Defunciones

Fecha: Jueves, Mayo 20, 2010



Elaborado por el Centro Tecnologías de Gestión de Datos (DATEC)  
Universidad de las Ciencias Informáticas

Figura 9 Reporte Correspondiente al Área de Análisis de Defunciones

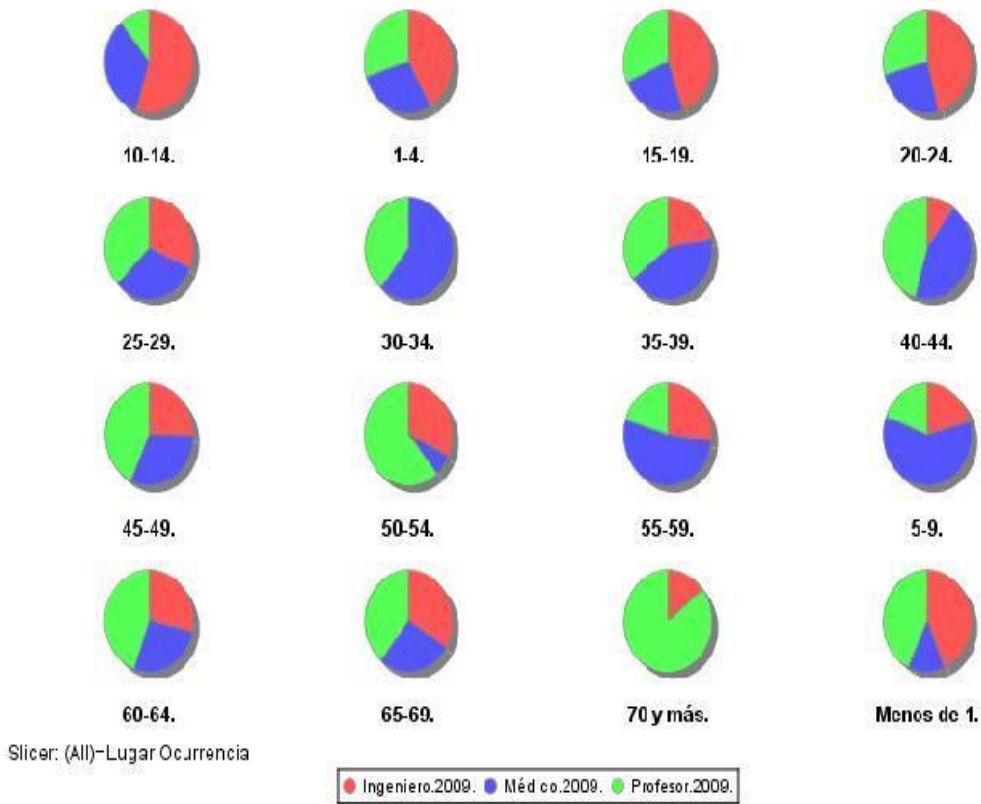
CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA



Área de Análisis: Nacimientos

Fecha: Jueves, Mayo 20, 2010

**Nacidos Vivos Según Ocupación de la Madre, por Grupos de Edad**



Elaborado por el Centro Tecnologías de Gestión de Datos (DATEC)  
Universidad de las Ciencias Informáticas

Figura 10 Reporte Correspondiente al Área de Análisis de Defunciones

CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA



Área de Análisis: Defunciones

Fecha: Jueves, Mayo 20, 2010

Tasa de Mortalidad Infantil. Año 2009



	Medidas
DPA	• Tasa de Mortalidad Infantil
+CAMAGUEY	653,784
+CIEGO DE AVILA	473,863
+CIENFUEGOS	754,656
+CIUDAD DE LA HABANA	611,993
+GRANMA	610,275
+GUANTANAMO	805,226
+HOLGUIN	531,321
+ISLA DE LA JUVENTUD	473,863
+LA HABANA	536,585
+LAS TUNAS	504,379
+MATANZAS	622,837
+PINAR DEL RIO	687,679
+SANCTI SPIRITUS	645,306
+SANTIAGO DE CUBA	257,677
+VILLA CLARA	763,993

Gráfico: [año =2009]

Elaborado por el Centro Tecnologías de Gestión de Datos (DATEC)  
Universidad de las Ciencias Informáticas

**Figura 11 Reporte Correspondiente al Área de Análisis de Defunciones**



Área de Análisis: Población

Fecha: Jueves, Mayo 20, 2010

Tasa Anual de Crecimiento de la Población. Año 2009



	Medidas
Cuba	Tasa Anual
+CAMAGUEY	38
+CIEGO DE AVILA	-265
+CIENFUEGOS	-512
+CIUDAD DE LA HABANA	456
+GRANMA	173
+GUANTANAMO	-57
+HOLGUIN	125
+ISLA DE LA JUVENTUD	-355
+LA HABANA	-33
+LAS TUNAS	145
+MATANZAS	173
+PINAR DEL RIO	289
+SANCTI SPIRITUS	-394
+SANTIAGO DE CUBA	-146
+VILLA CLARA	-296

Gráfico: [anno=2009]

Elaborado por el Centro Tecnologías de Gestión de Datos (DATEC)  
Universidad de las Ciencias Informáticas

**Figura 12 Reporte Correspondiente al Área de Análisis de Población**

### 2.14 Conclusiones del Capítulo

Una vez culminado el desarrollo de este capítulo se concluye que:

## CAPÍTULO 2: DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

- ✓ Se desarrolló la capa de visualización del mercado de datos de Demografía, mostrándose las potencialidades de las herramientas seleccionadas.
- ✓ Se identificaron y describieron los KPI del negocio.
- ✓ Se modeló el esquema multidimensional conformado por los diferentes cubos OLAP.
- ✓ Se determinaron los colores a usar en el diseño del CMD, así como la gestión de alertas en el mismo.
- ✓ Se identificaron las Áreas de Análisis, con los Libros de Trabajos y reportes candidatos contenidos.
- ✓ Se implementaron los reportes correspondientes a las áreas de análisis que se identificaron previamente.

## **CAPÍTULO 3: VALIDACIÓN DE LA SOLUCIÓN**

### ***Introducción***

En el desarrollo de este capítulo se describe el proceso seguido para la validación de la Capa de Visualización del Mercado de Datos Demografía. En este proceso se realizaron varias iteraciones, con diferente personal involucrado. Una vez concluida la solución se realizaron pruebas por parte de los usuarios funcionales del sistema, verificando la integridad de los datos de los reportes y el alcance del dominio informativo.

### ***3.1 Pruebas***

Las pruebas son un conjunto de actividades en las cuales un sistema o componente es ejecutado bajo unas condiciones o requerimientos específicos, donde los resultados son observados y registrados para dar una evaluación de algún aspecto del sistema o componente que es evaluado y determinar la calidad del mismo. En ellas se describen como comprobar cada versión operacional del sistema durante la integración del mismo, describiendo también como hacer las pruebas al sistema verificando que todos los requerimientos hayan sido implementados, determinado los defectos del mismo.

Es importante aclarar que las pruebas no pueden asegurar la ausencia de defectos; solo pueden demostrar que existen defectos en el software y que cada prototipo que se quiera entregar al final de una iteración debe ser probado y evaluado a diferentes niveles.

### ***3.2 Pruebas Aplicadas***

A la Capa de Visualización del Almacén de Datos de la ONE se le aplicaron un conjunto de pruebas, generales y específicas. Como principal objetivo de estas pruebas estaba: comprobar la consistencia e integridad de los datos presentados en los reportes, y el alcance del dominio informativo.

A continuación se detallan las pruebas realizadas, con sus resultados.

#### ***3.2.1 Prueba de Consistencia de los Datos***

### CAPÍTULO 3: VALIDACIÓN DE LA SOLUCIÓN

Esta prueba consiste en insertar un número finito de datos a través de proceso de ETL, buscando que se ejecute el ciclo completo hasta la capa de visualización, donde se verifica que los datos insertados sean los mismos que los visualizados en la aplicación del usuario final.

Para ejecutar esta prueba, se deben tener datos confiables y sin existencia de duplicidad, u otros errores de cálculos en las estructuras, o sea, ya validados con las ETL. Se mide con los criterios siguientes:

- ✓ Eficiente: El conjunto de datos de entrada debe el mismo que el de salida.
- ✓ Parcial: El conjunto de datos de salida es el 80 % del conjunto de datos de entrada.
- ✓ Deficiente: El conjunto de datos de salida es menor del 80 % del conjunto de datos de entrada.

El resultado obtenido fue: Eficiente

#### **3.2.2 Prueba de las estructuras de pre-visualización o estructuras multidimensionales**

Con esta prueba se verifica que las estructuras construidas respondan al dominio informativo definido en el alcance de la solución. Para ello, se debe contar con estructuras sólidas que respondan a las necesidades de información que se definieron en la fase de requerimientos. La forma de evaluar los resultados es la siguiente:

- ✓ Eficiente: Las estructuras cumplen con el alcance de información previsto
- ✓ Parcial: Las estructuras cumplen en un 80% con el alcance de información previsto
- ✓ Deficiente: Las estructuras cumplen menos del 80% con el alcance de información previsto.

El resultado obtenido fue: Eficiente

#### **3.3 Validación de los Requisitos por el Cliente**

Se realizó un encuentro con la compañera Elena Leonila Fernández García: cliente del presente trabajo, y el compañero Asnioby Hernández López, los cuales estuvieron de acuerdo con la propuesta de la solución presentada (consultar anexos 1 y 2).

### **3.4 Conclusiones del Capítulo**

Una vez finalizado el capítulo se concluye que:

- ✓ Se obtiene una aplicación funcional que cumple con los requisitos de los clientes.
- ✓ Se realizó la validación de la solución, incluyendo los tipos de pruebas realizadas y resultados obtenidos, buscando eficiencia y rendimiento.
- ✓ Los errores encontrados en las pruebas se fueron resolviendo a medida que se realizaban estas.
- ✓ Se referencia la certificación de calidad emitida por los especialistas de DATEC y el Acta de Aceptación de la Capa de Visualización firmada por el cliente.



## CONCLUSIONES GENERALES

Al finalizar el trabajo de diploma “Almacén de Datos de Estadísticas: Desarrollo de la Capa de Visualización del Mercado de Datos Demografía” se puede concluir que se ha cumplido con los objetivos generales planteados.

A partir de un estudio de las técnicas y herramientas existentes para Inteligencia de Negocios:

- ✓ Fueron seleccionadas las técnicas de diseño multidimensionales (OLAP), y las herramientas de consultas de datos, para el apoyo a la toma de decisiones.
- ✓ Se diseñó la estructura del CMD, así como el mapa de navegación del mismo.
- ✓ Se modelaron las estructuras dimensionales necesarias que abarcan la información relevante del Modelo Estadístico y soportan el proceso de toma de decisiones.
- ✓ Se construyeron los reportes multidimensionales, estructurados por áreas de análisis.
- ✓ Las pruebas realizadas con personal calificado en el tema de análisis de información permitieron validar la solución propuesta obteniendo resultados satisfactorios en cada una de ellas.

## RECOMENDACIONES

Con la culminación del presente trabajo, se da cumplimiento a los objetivos propuestos. Sin embargo, se realizan una serie de recomendaciones para enfocar el trabajo futuro en el mejoramiento de la propuesta realizada.

- ✓ Montar el módulo de reportes multidimensionales en el servidor de Pentaho.
- ✓ Implementar la capa superior del CMD.

## REFERENCIAS BIBLIOGRÁFICAS

1. Oficina Nacional de Estadísticas. [Online] [Cited: Diciembre 10, 2009.] <http://www.one.cu>.
2. **Caramazana, Alberto**. *Tecnologías y Metodologías para la Construcción de Sistemas de Gestión del Conocimiento*. Madrid : s.n., 2002.
3. **Armstrong, Smith**. *Oracle Discoverer 10g Handbook*. San Francisco, California : s.n., 2006.
4. Inteligencia de Negocios. [Online] [Cited: Enero 23, 2010.] <http://www.monografias.com>.
5. **Rivera, Sofía**. El Datawarehouse y el Business Intelligence en el Gobierno. [Online] 2006. [Cited: Enero 20, 2010.] <http://www.monografias.com>.
6. **Ferrari, Marco Russo and Alberto**. Introduction to the SQLBI Methodology. [Online] [Cited: Enero 25, 2010.] <http://www.sqlbi.com>.
7. **DATEC, Especialistas de**. *Metodología para el Desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocios en DATEC*. Ciudad de la Habana : s.n., 2010.
8. **Lanzillotta, Analía**. MasterMagazine. [Online] [Cited: Enero 11, 2010.] <http://www.mastermagazine.info>.
9. **Riviera, José**. El Cuadro de Mando Integral. [Online] [Cited: Enero 30, 2010.] <http://www.lasprovincias.es>.
10. **Press, Boston**. *The Balanced ScoreCard: Translating Strategy into Action*. Harvard Business School. 1996.
11. **Canney, Edward Restrepv**. [Online] [Cited: Febrero 1, 2010.] <http://www.igerencia.com>.
12. **Few, Stephen**. *Information Dashboard Design*. 2006.
13. **Rodríguez, Anisley Delfino**. *Cuadros de Mando Digitales, Soporte a la Toma de Decisiones*. Ciudad de La Habana : s.n., 2010.
14. Sixtina Consulting Group. *El Diseño del Dashboard: cómo incluir los KPIs(Indicadores Claves de Desempeño) y sus Métricas*. [Online] [Cited: Febrero 5, 2010.] <http://www.gestiopolis.com/administracion-estrategia/dashboard-kpi-metricas.htm#>.
15. BITAN. *Herramientas Reporteadoras*. [Online] [Cited: Febrero 5, 2010.] <http://www.g3advisoft.com> .

REFERENCIAS BIBLIOGRÁFICAS

16. Oracle. [Online] [Cited: Febrero 5, 2010.] <http://www.oracle.com>.
17. IBM Cognos. [Online] [Cited: Febrero 7, 2010.] <http://www-01.ibm.com>.
18. Microstrategy. [Online] [Cited: Febrero 2010, 7.] <http://www.microstrategy.com>.
19. Pentaho. [Online] [Cited: Febrero 2010, 2010.] <http://www.pentaho.com>.
20. **Torres, Lilliam Vega, Díaz, Luis Rojas and Villar, Cecilia Placeres.** *La inteligencia de Negocio. Su Implementación Mediante la Plataforma Pentaho.* Ciudad de la Habana : s.n., 2008.

## BIBLIOGRAFÍA

1. Oficina Nacional de Estadísticas. [Online] [Cited: Diciembre 10, 2009.] <http://www.one.cu>.
2. **Caramazana, Alberto.** *Tecnologías y Metodologías para la Construcción de Sistemas de Gestión del Conocimiento*. Madrid : s.n., 2002.
3. **Armstrong, Smith.** *Oracle Discoverer 10g Handbook*. San Francisco, California : s.n., 2006.
4. Inteligencia de Negocios. [Online] [Cited: Enero 23, 2010.] <http://www.monografias.com>.
5. **Rivera, Sofía.** El Datawarehouse y el Business Intelligence en el Gobierno. [Online] 2006. [Cited: Enero 20, 2010.] <http://www.monografias.com>.
6. **Ferrari, Marco Russo and Alberto.** Introduction to the SQLBI Methodology. [Online] [Cited: Enero 25, 2010.] <http://www.sqlbi.com>.
7. Pentaho Analysis Services: Mondrian Project. [Online] enero 25, 2008. [Cited: marzo 5, 2010.] <http://mondrian.pentaho.org/>.
8. Diez ventajas del software libre y propietario. [Online] [Cited: enero 20, 2010.] <http://www.abadiadigital.com/noticia2010.html>.
9. NEXTEL Engineering. *La Inteligencia de Negocios*. [Online] [Cited: enero 31, 2010.] <http://www.nexteleng.es>.
10. **DATEC, Especialistas de.** *Metodología para el Desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocios en DATEC*. Ciudad de la Habana : s.n., 2010.
11. **Lanzillotta, Analia.** MasterMagazine. [Online] [Cited: Enero 11, 2010.] <http://www.mastermagazine.info>.
12. **Riviera, José.** El Cuadro de Mando Integral. [Online] [Cited: Enero 30, 2010.] <http://www.lasprovincias.es>.
13. **Press, Boston.** *The Balanced ScoreCard: Translating Strategy into Action*. Harvard Business School. 1996.
14. **Canney, Edward Restrepv.** [Online] [Cited: Febrero 1, 2010.] <http://www.igerencia.com>.

15. **Few, Stephen.** *Information Dashboard Design.* 2006.
16. **Rodríguez, Anisley Delfino.** *Cuadros de Mando Digitales, Soporte a la Toma de Decisiones.* Ciudad de La Habana : s.n., 2010.
17. Sixtina Consulting Group. *El Diseño del Dashboard: cómo incluir los KPIs(Indicadores Claves de Desempeño) y sus Métricas.* [Online] [Cited: Febrero 5, 2010.] <http://www.gestiopolis.com/administracion-estrategia/dashboard-kpi-metricas.htm#>.
18. BITAN. *Herramientas Reporteadoras.* [Online] [Cited: Febrero 5, 2010.] <http://www.g3advisoft.com> .
19. Oracle. [Online] [Cited: Febrero 5, 2010.] <http://www.oracle.com>.
20. IBM Cognos. [Online] [Cited: Febrero 7, 2010.] <http://www-01.ibm.com>.
21. Microstrategy. [Online] [Cited: Febrero 2010, 7.] <http://www.microstrategy.com>.
22. Pentaho. [Online] [Cited: Febrero 2010, 2010.] <http://www.pentaho.com>.
23. **Torres, Lilliam Vega, Díaz, Luis Rojas and Villar, Cecilia Placeres.** *La inteligencia de Negocio. Su Implementación Mediante la Plataforma Pentaho.* Ciudad de la Habana : s.n., 2008.
24. **Viñegla, Dr. Alonso López.** Cuadro de Mando Integral. [Online] [Cited: febrero 15, 2010.] <http://cuadrodemandando.unizar.es>.
25. Cuadro de Mando Integral. [Online] [Cited: enero 24, 2010.] [http://www.fundibeq.org/metodologias/herramientas/cuadro\\_de\\_mando\\_integral.pdf](http://www.fundibeq.org/metodologias/herramientas/cuadro_de_mando_integral.pdf).
26. **Arias, Emilio.** Exito en la implantación de un sistema Business Intelligence. [Online] [Cited: marzo 20, 2010.] <http://www.monografias.com/trabajos29>.
27. Como crear un cubo multidimensional OLAP usando Pentaho. [Online] [Cited: abril 1, 2010.] <http://www.monografias.com/trabajos55>.
28. El Datawarehouse y el Business Intelligence en Gobierno. [Online] [Cited: diciembre 10, 2009.] <http://www.monografias.com/trabajos17>.
29. **Yáñez, Julián López.** *Inteligencia institucional. Hacia una ecología social de las organizaciones.* Barcelona : s.n., 2007.

BIBLIOGRAFÍA

30. La plataforma Pentaho Open Source Business Intelligence. [Online] [Cited: marzo 20, 2010.] <http://pentaho.almacen-datos.com/>.
31. Como instalar Pentaho. [Online] [Cited: marzo 17, 2010.] <http://www.biblogs.com/2007/05/28/como-instalar-pentaho/>.
32. Instalación y configuración del servidor Mondrian Pentaho. [Online] [Cited: marzo 1, 2010.] <http://pentaho.almacen-datos.com/mondrian.html>.

## GLOSARIO DE TÉRMINOS

**Almacén de Datos:** Es una colección de datos orientada a un determinado ámbito, integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza.

Bill Inmon define un almacén de datos en términos de las características del repositorio de datos:

- ✓ **Orientado a temas:** Los datos en la base de datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí.
- ✓ **Variante en el tiempo:** Los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones.
- ✓ **No volátil:** La información no se modifica ni se elimina, una vez almacenado un dato, éste se convierte en información de sólo lectura, y se mantiene para futuras consultas.
- ✓ **Integrado:** La base de datos contiene los datos de todos los sistemas operacionales de la organización, y dichos datos deben ser consistentes.

Ralph Kimball define un Almacén de Datos como: "una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis". Kimball además determinó que un almacén de datos no era más que: "la unión de todos los Mercados de Datos de una entidad".

**API:** Una interfaz de programación de aplicaciones es el conjunto de funciones y procedimientos (o métodos, si se refiere a programación orientada a objetos) que ofrece cierta biblioteca para ser utilizado por otro software como una capa de abstracción.

**Área de Análisis:** Se le denomina Área de análisis a la agrupación de información según su propósito, aunque el criterio depende de las necesidades de la institución o empresa donde se aplica el sistema. Permite restringir el número de usuarios que acceden a los datos.

**Cubo:** Colección de dimensiones y medidas en un área temática particular.

**Data Warehousing:** Se entiende por Data Warehousing el proceso de extraer y filtrar datos de las operaciones comunes de la organización, procedentes de los distintos sistemas de información operacionales y/o sistemas externos, para transformarlos, integrarlos y almacenarlos en un depósito o almacén de datos con el fin de acceder a ellos para dar soporte en el proceso de toma de decisiones de



una organización. Es decir, la finalidad es convertir los datos operacionales en información relacionada y estructurada, homogénea y de mayor calidad, identificada convenientemente y que se mantenga en el tiempo, o sea, los datos más recientes no sustituyen a los precedentes, pero tampoco se acumulan de cualquier manera, sino que se suelen mantener con un mayor nivel de detalle los datos actuales, y de manera más agregada los datos anteriores. Se pretende crear un círculo virtuoso para la información. Llegando a una conclusión operativa se definiría Data Warehousing como todas las técnicas u operaciones que se realizan sobre un almacén de datos.

**Dimensiones:** Las dimensiones de un cubo son atributos relativos a las variables, son las perspectivas de análisis de las variables. Son catálogos de información complementaria necesaria para la presentación de los datos a los usuarios, como por ejemplo: descripciones, nombres, zonas, rangos de tiempo, etc. Es decir, la información general complementaria a cada uno de los registros de la tabla de hechos.

**Drill-Down (Taladrar):** Usado comúnmente en los sistemas de información gerencial o de análisis de información, es la habilidad para poder navegar de lo general a lo particular en la información presentada.

**JDBC:** Es el acrónimo de Java Database Connectivity, un API que permite la ejecución de operaciones sobre bases de datos desde el lenguaje de programación Java independientemente del sistema operativo donde se ejecute o de la base de datos a la cual se accede utilizando el dialecto SQL del modelo de base de datos que se utilice. El API JDBC se presenta como una colección de interfaces Java y métodos de gestión de manejadores de conexión hacia cada modelo específico de base de datos. Un manejador de conexiones hacia un modelo de base de datos en particular es un conjunto de clases que implementan las interfaces Java y que utilizan los métodos de registro para declarar los tipos de localizadores a base de datos (URL) que pueden manejar.

**JPivot:** Es una librería de componentes JSP que se utiliza para construir tablas OLAP generadas de forma dinámica. Este tipo de tablas es de gran utilidad ya que permite mostrar los resultados de las consultas filtrando por los campos de la tabla de manera que se puedan quitar y poner distintos criterios de búsqueda de los datos, consiguiendo un amplio abanico de posibilidades.

**Libro de Trabajo:** Estructura organizativa que agrupa los reportes generados dentro de las áreas de análisis. Puede ser creado teniendo en cuenta criterio que permitan organizar la información: Emisor de los reportes, receptor del reporte, contenido, etc.

**Mercado de Datos:** Es un almacén de datos departamental, o sea, orientado a un sector determinado de la organización. Es también conocido: datamart.

**Mondrian:** Es un servidor OLAP escrito en Java. Permite analizar grandes cantidades de datos almacenados.

**Multidimensionalidad:** La información multidimensional se puede encontrar en hojas de cálculo, bases de datos, etc. Una herramienta de Inteligencia de Negocios debe de ser capaz de reunir información dispersa en toda la empresa e incluso en diferentes fuentes para así proporcionar a los departamentos la accesibilidad, poder y flexibilidad que necesitan para analizar la información.

**Procesamiento Analítico en Línea:** (OLAP), consiste en la realización de consultas a estructuras multidimensionales que contienen datos resumidos. Su objetivo fundamental es lograr altas velocidades en consultas complejas.

**Software Libre:** Software que, una vez obtenido, puede ser usado, copiado, estudiado, modificado y redistribuido libremente. El software libre suele estar disponible gratuitamente en Internet, o a precio del costo de la distribución a través de otros medios.

## ANEXOS

### Anexo1: Acta de aceptación del Componente de Visualización del Mercado de Datos Demografía.

Acta de aceptación del componente Capa de Visualización del Mercado de Datos Demografía perteneciente al Producto Almacén de Datos Estadísticos de la ONE de la Universidad de las Ciencias Informáticas.

Actualmente la Oficina Nacional de Estadísticas (ONE) tiene una gran dificultad consistente en que posee gran cantidad de información que necesita ser consultada por los especialistas de dicha entidad. Esta información se encuentra principalmente en ficheros xls los cuales se han guardados teniendo en cuenta la fecha a la que corresponde.

Esta situación ha provocado que se trabaje fundamentalmente con la información actual pero no se tiene una solución de software que permita realizar análisis corporativos sobre los datos almacenados.

En este sentido se contactó con el DATEC y se realizó la propuesta de crear un Almacén de Datos para lograr un mayor control sobre los datos relacionados con la Demografía en Cuba. Una vez construido y poblado dicho almacén se hace necesario explotar los datos que se encuentran almacenados en el mismo desarrollando la Capa de Visualización del Mercado de Datos Demografía.

A través de esta solución los usuarios podrán:

- ✦ Realizar análisis de comportamientos y tendencias de las principales áreas demográficas, para obtener estadísticas que le ayudarán a la toma de decisiones en Cuba.
- ✦ Acceder a la información de forma fácil, rápida e intuitiva.
- ✦ Consultar reportes de interés para los especialistas de la entidad que permitan tener una mayor visión del comportamiento de los principales indicadores de rendimientos de la entidad.

En general se estima que el componente realizado pone a los especialistas en un nivel superior en cuanto al uso de la información, lo cual facilitará el análisis de datos en la entidad.



Elena Leonila Fernández García

Gerente: Programa de Gestión de Información Nacional

Anexo 2: Ejemplo de Reporte del Área de Análisis de Población.

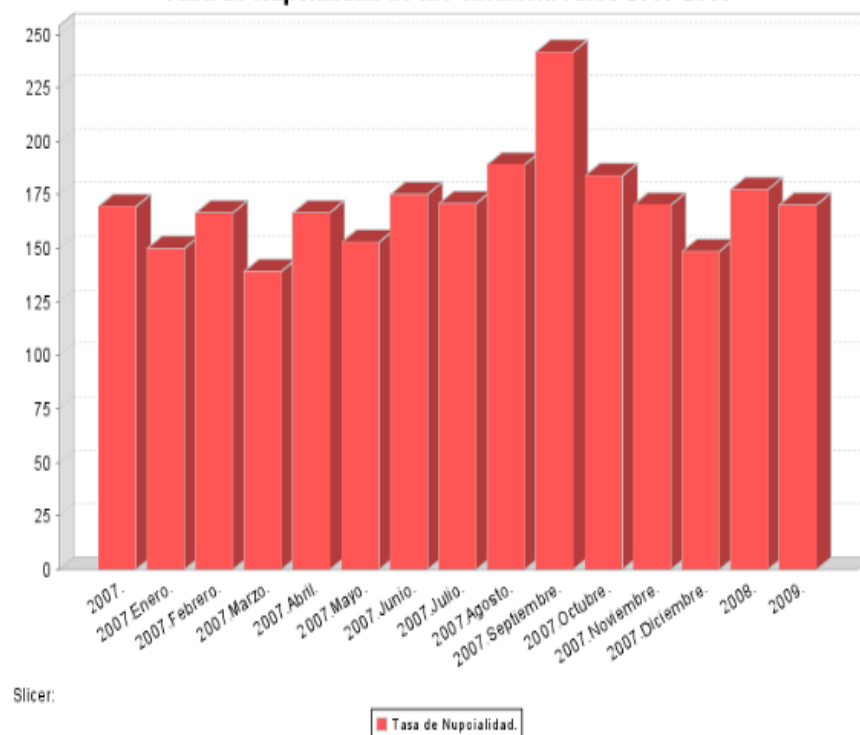


[Principal](#) | [Población](#) | [Nacimientos](#) | [Defunciones](#) | [Divorcios](#) | [Matrimonios](#) | [Migraciones](#)



	Medidas
Temporal	Tasa de Nupcialidad
2007	169,738
Enero	149,831
Febrero	166,417
Marzo	139,133
Abril	166,447
Mayo	153,097
Junio	175,056
Julio	170,832
Agosto	189,408
Septiembre	241,867
Octubre	184,138
Noviembre	170,638
Diciembre	148,601
2008	177,501
2009	170,084

Tasa de Nupcialidad de la Población. Años 2007-2009



Anexo 3: Ejemplo de Reporte del Área de Análisis de Nacimientos.



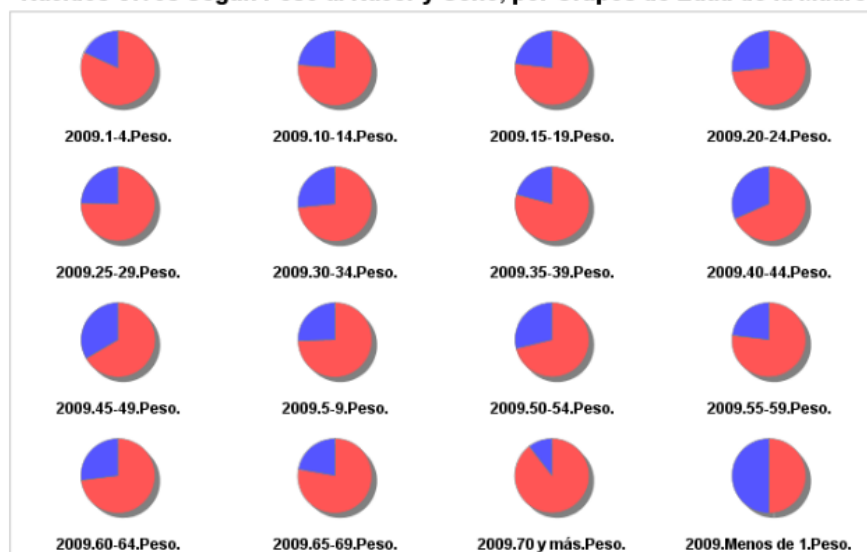
[Principal](#) | [Población](#) | [Nacimientos](#) | [Defunciones](#) | [Divorcios](#) | [Matrimonios](#) | [Migraciones](#)



Temporal	Grupo Quinquenal	Peso Nacimiento	Medidas	
			Cantidad Nacimientos Varones	Cantidad Nacimientos Hembras
2009	1-4	+Peso	126	28
	10-14	+Peso	1.172	359
	15-19	+Peso	881	266
	20-24	+Peso	1.994	725
	25-29	+Peso	1.101	357
	30-34	+Peso	1.689	612
	35-39	+Peso	903	237
	40-44	+Peso	705	325
	45-49	+Peso	1.293	646
	5-9	+Peso	1.117	380
	50-54	+Peso	791	321
	55-59	+Peso	1.206	357
	60-64	+Peso	1.350	495
	65-69	+Peso	1.201	343
	70 y más	+Peso	179	21
	Menos de 1	+Peso		

Gráfico:

**Nacidos Vivos Según Peso al Nacer y Sexo, por Grupos de Edad de la Madre**

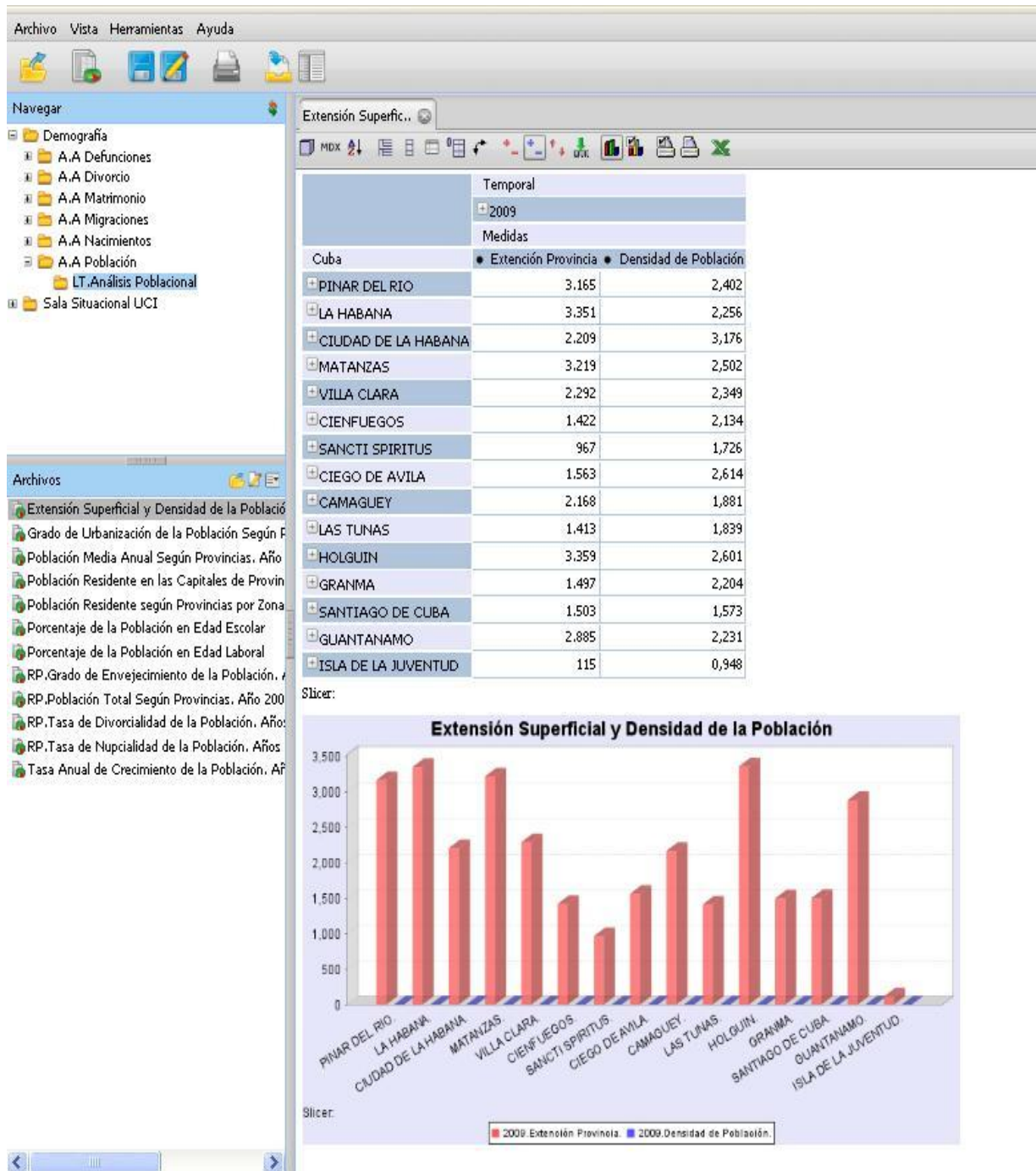


Slicer:



Elaborado por el Centro Tecnologías de Gestión de Datos (DATEC)  
Universidad de las Ciencias Informáticas

Anexo 4: Ejemplo de Reporte Montado en el servidor de Pentaho.





Anexo 5: Ejemplo de Reporte montado en el servidor de Pentaho (Población Media Anual).

