Universidad de las Ciencias Informáticas

Facultad 8

Título: Análisis, Diseño e Implementación de un Mercado de Datos para el Departamento de Población de la Oficina Nacional de Estadísticas.

Trabajo de Diploma para optar por el título de

Ingeniero en Ciencias Informáticas.

Autores:

Anelis Vargas Rio.

Mayte Machado Estévanez.

Tutores:

Ing. Mabel Medina Rodríguez.

Ing. Yanisbel González Hernández.

Ciudad de La Habana, junio 2010

Un ingeniero no es una copia,

es original y se atreve a cambiar una realidad,

no importa el tiempo o el espacio, todo es posible

mientras crea que es así.

Le dedico el presente trabajo a toda mi familia por el gran apoyo que me han brindado, en especial a mis padres y mis abuelos que son lo más grande que tengo.

Anelis

Quiero dedicar este trabajo de diploma:

A la persona más especial que existe en mi vida, mi madre.

A mi hermana para que siga adelante en sus estudios.

Mayte

Le agradezco a toda mi familia por haber confiado en mí y estar siempre a mi lado, en especial agradecerle a mis padres, mis hermanos, mis abuelos y a mi tía Gladis Vargas.

Les agradezco a mis tías Gladis Torres y Miriam ya que siempre estuvieron a mi lado cuando necesité una madre que me cuidara y me diera un consejo.

A mi madrasta Anita, a Yerlandis y a mi padrastro Yoel.

A todas mis amistades, todas, todas, en espacial a Maylén, Claritza y Yoangel.

A mi compañera de tesis Mayte, a mis tutoras Mabel y Yanisbel.

Al tribunal y oponente por ayudarnos a perfeccionar nuestro trabajo.

En general, a todo aquel de una forma u otra me apoyó y estuvo a mi lado cuando lo necesité.

A todos muchas gracias.

Anelis

Le quiero agradecer a mi madre, por brindarme

su incansable apoyo, preocupación, animación, amor y dedicación.

A toda mi familia que siempre ha confiado en mí como persona,

como estudiante y como profesional.

A todas las personas que me quieren y me brindaron su amor y apoyo incondicional.

A mis tutoras Mabel y Yanisbel por estar siempre dispuestas a ayudarme con el trabajo.

A mis compañeros y compañeras de la Universidad por colaborar en lo que pudieron.

A todos los que de una forma u otra, aunque fuera con su pensamiento, hayan favorecido la realización del trabajo.

A todos muchas gracias.

Mayte

Resumen

El presente trabajo de diploma está centrado en el área de los Mercados de Datos y en las técnicas de Procesamiento Analítico en Línea (OLAP por sus siglas en inglés) para el análisis de información estadístico en Cuba. Abarca un estudio detallado de la disciplina Demografía, además de investigar las metodologías, tendencias y herramientas para el desarrollo de este tipo de soluciones.

Como resultados se obtienen las estructuras dimensionales para el modelo estadístico de “Población” que contienen las dimensiones, jerarquías, niveles, atributos, tablas de hechos y medidas que garantizan los análisis estadísticos. Igualmente se detalla el negocio permitiendo identificar las reglas del mismo.

Se realiza el proceso de carga de los datos de las fuentes del Departamento de Población de la Oficina Nacional de Estadísticas (ONE) al Mercado de Datos presentado. En el trabajo se incluyen las estrategias de seguridad, respaldo y recuperación de datos y también se efectúan pruebas que validan la solución en cuestión.

Índice.

Índice.

INTRODUCCIÓN.

La estadística es comúnmente considerada como una colección de hechos numéricos expresados en términos de una relación dócil y que ha sido recopilada a partir de otros datos numéricos. (Hugo Morales Alejo, 2009)

Los babilonios, egipcios, chinos, mayas, incas y griegos, atesoraban y analizaban datos de su gobierno utilizando algún tipo de control de esta información, que podríamos definir como predecesor del estudio de la estadística. No es hasta el Siglo XVII que surge la "disciplina estadística". Por esta época se inicia el desarrollo de dos escuelas: la demográfica social y la enciclopédica matemática. La primera termina en la fundación de la demografía como disciplina y la segunda resultó la estadística en su concepción actual.

Durante las décadas de los treinta a los setenta del siglo XX, se introdujo la estadística en los centros de investigación y en la producción industrial, por lo que surgió una comunidad de profesionales de esta disciplina. Con la llegada e incremento de las computadoras, las técnicas para el manejo y aprovechamiento de la información se hacen indispensables.

El desarrollo de los paquetes computacionales estadísticos dentro de la industria del software en la década de los setenta, ochenta y noventa, hizo que los técnicos y profesionales tuviesen la posibilidad de aplicar la estadística sin tener que efectuar cálculos muy complicados. Hoy en día las técnicas y los métodos más complejos solo requieren de minutos de procesamiento computacional, además de permitir grandes facilidades de graficación. Esto ha posibilitado que muchos análisis estadísticos se puedan realizar de manera interactiva. También podemos observar como cada país en correspondencia con sus características y desarrollo adopta la vía más favorable, lo que determina que la forma de organización de la industria del software para regir su desarrollo no está definida.

En Cuba, la Oficina Nacional de Estadísticas (ONE), es la encargada de garantizar las estadísticas mediante el Sistema Estadístico Nacional. El Departamento de Población es una de las áreas de la ONE, que estudia los análisis demográficos del País. La demografía es una ciencia que tiene como objetivo, el análisis del volumen estructural y desarrollo de las poblaciones humanas desde el punto de vista cuantitativo. El estudio del estado, los cambios y la evolución de la población son fundamentales para planificar programas para el desarrollo de los pueblos. Algunas de las áreas que necesitan de estudios demográficos para realizar su labor de forma exitosa son:

1. Planificación de nuevos programas: salud, educación, seguridad, etc.
2. Evaluación del impacto de los programas existentes.
3. Distribución equitativa de los recursos.
4. Identificación de problemas y necesidades futuras.
5. Identificación del potencial de las localidades para el mercado de bienes y servicios.
6. Determinación de las características de los potenciales clientes.
7. Desarrollo de estrategias de mercadeo para nuevos productos.
8. Empleo de técnicas y modelos demográficos para explicar otros comportamientos. (econlink, 2009)

La Oficina Nacional de Estadística como se planteó anteriormente, es el órgano rector de la estadística en Cuba y tiene como objetivo fundamental captar, analizar y difundir los datos recogidos a lo largo y ancho de todo el País en áreas como: Educación, Salud, Transporte, Inmigración, Turismo, Medio Ambiente, Inversiones, Ocupación, Población, entre otras. Muchas de las necesidades informativas o análisis deseados sobre el área específica de Población no pueden ser realizadas, pues no cuentan con el soporte para llevarlas a cabo. Estos datos son almacenados en formatos de difícil acceso para su consulta, por lo que se hace muy complejo el proceso de acceder y divulgar dicha información.

La forma de almacenar, recuperar y presentar la información en el Departamento de Población en la ONE impide realizar los principales análisis y cruces de variables, indicadores, tasas, porcentajes y demás aspectos de interés; dificultando así la disponibilidad de información para órganos del estado y afectando el proceso de toma de decisiones.

Partiendo de lo anteriormente expuesto, se plantea el siguiente **problema de investigación:**

- ¿Cómo mejorar las tecnologías de almacenamiento y organización de datos para incrementar la eficiencia de los análisis de la información de Población almacenada en la ONE?

Para brindar solución a la interrogante anterior se traza el siguiente **objetivo general:**

- Desarrollar un Mercado de Datos para el control estadístico del área de Población en la ONE.

Para dar cumplimiento al objetivo general, los **objetivos específicos** propuestos son los siguientes:

- Realizar el estudio del arte del tema demografía e investigar las tecnologías y herramientas a utilizar.
- Analizar el modelo del Mercado de Datos para el área de Población de la Oficina Nacional de Estadísticas.
- Diseñar el Mercado de Datos del área de Población para el almacén de datos de la Oficina Nacional de Estadísticas.
- Cargar los clasificadores para el Mercado de Datos del área de Población de la Oficina Nacional de Estadísticas.
- Validar la solución desarrollada mediante la aplicación de pruebas de listas de chequeo de análisis, listas de chequeo de diseño, validación de los requerimientos por los clientes y casos de prueba de implantación al Mercado de Datos del área de Población de la Oficina Nacional de Estadísticas.

Por tanto, el **objeto de estudio** son las tecnologías y herramientas para Mercados de Datos.

A partir del objeto de estudio se determina como **campo de acción** las tecnologías y herramientas para Mercados de Datos de temas demográficos.

Como **posible resultado** del presente trabajo de diploma se obtendrá el Mercado de Datos para el área de Población de la ONE, contemplando todas las necesidades de información detectadas por parte de los involucrados.

Para cumplir con los objetivos específicos se proponen las siguientes **tareas científicas**:

1. Investigar los temas relacionados con la disciplina de demografía, tanto en el ámbito mundial como en Cuba, las tecnologías de almacenamiento y las herramientas necesarias para el desarrollo de la solución.
2. Definir la metodología a utilizar en el desarrollo de la solución.

3. Realizar entrevistas sistemáticas al personal especializado del Departamento de Población y del Departamento de Informática de la ONE.
4. Identificar estructuras de usuarios y permisos.
5. Definir temas de análisis.
6. Identificar necesidades de información, requisitos funcionales y no funcionales.
7. Modelar requerimientos.
8. Definir requisitos de entrada y de salida.
9. Elegir la granularidad del proceso del negocio.
10. Definir las dimensiones y los hechos del Mercado de Datos.
11. Estructurar el modelo dimensional y transformarlo al diseño físico.
12. Implementar la base de datos.
13. Montar los clasificadores para el Mercado de Datos del área de Población para el almacén de datos de la Oficina Nacional de Estadísticas.
14. Realizar pruebas mediante la aplicación de listas de chequeo de análisis, listas de chequeo de diseño, validación de los requerimientos por los clientes y casos de prueba de implantación, comprobando así la eficiencia del Mercado de Datos.

Para lograr el desarrollo del Mercado de Datos se determinan los métodos científicos, los cuales se dividen en métodos empíricos y teóricos.

Métodos Teóricos utilizados:

El **analítico – sintético** permite la división del fenómeno en sus múltiples relaciones y componentes, lo que facilita su estudio. Establece la unión entre las partes previamente analizadas, posibilita el descubrimiento de sus características generales y las relaciones entre ellas. Para este método se hace necesario partir de un análisis de toda la información que fue recopilada, mediante otros métodos de

investigación, como la entrevista; para luego centrarla y sintetizarla para elaborar el procedimiento adecuado.

El **histórico – lógico** posibilita documentar los temas relacionados con el desarrollo del Mercado de Datos, la disciplina de demografía y la metodología a utilizar.

El **modelado** define la arquitectura del sistema y posibilita determinar las dimensiones y los hechos del Mercado de Datos, así como los diferentes tipos de relaciones que se observan entre estos elementos. De esta manera, se conforma y estructura el modelo dimensional, transformándolo luego al diseño físico.

Método Empírico utilizado:

Entrevista: Entrevistas sistemáticas al personal especializado del Departamento de Población y del Departamento de informática de la ONE.

Estructura capitular

Capítulo 1: Fundamentación Teórica.

En este capítulo se realiza el estudio pertinente al tema demografía, su evolución y sistemas de control de datos estadísticos a nivel nacional y mundial. De igual forma se estudian las tecnologías de almacenamiento de datos y las herramientas para dar solución al problema planteado.

Capítulo 2: Análisis y Diseño.

En este capítulo se confecciona el análisis de la solución del Mercado de Datos del área de Población de la ONE. También se define el tema de análisis, los roles, permisos y las necesidades informativas de los usuarios. Además, se generan los diferentes requisitos, los casos de uso del sistema y se elabora el diseño lógico del Mercado de Datos.

Capítulo 3: Implementación y Prueba.

En este capítulo se realiza la implementación de la base de datos, se confecciona el modelo de datos y el modelo de despliegue. Igualmente se montan los clasificadores para el Mercado de Datos y se llevan a cabo las pruebas de volumen, carga y lista de chequeo.

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA.

Introducción.

En este capítulo se realiza un estudio sobre la disciplina de demografía, además se estudian las tecnologías de almacenamiento, para definir la adecuada en función del presente trabajo. Igualmente se analizan los modelos de base de datos, se aborda sobre los modelos de almacenamiento y se define la metodología a usar para el desarrollo del Mercado de Datos del Departamento de Población de la ONE. Además de investiga sobre las herramientas de integración de datos y las herramientas de modelado.

1.1 Demografía.

El interés por conocer el número de personas enmarcadas en un determinado espacio geográfico ha existido desde tiempos muy remotos. Dicho interés está vinculado a razones negativas, en la mayoría de los casos: cobros de impuestos, disponibilidad de hombres para la guerra, etc. La población pasa a ser objeto de estudio con carácter científico en la segunda mitad del siglo XVII al investigarse los bautismos y entierros durante un periodo de tiempo. De esta forma, surge la demografía como ciencia. Son varios los autores que han dado una definición de lo que es la demografía.

“La demografía es la ciencia que tiene como objeto de estudio las poblaciones humanas”. (Leggui., 1973)

“La demografía es una ciencia que tiene por objeto de estudio el volumen, la estructura y el desarrollo de las poblaciones humanas desde el punto de vista principalmente cuantitativo”. (NACIONES_UNIDAS, 1959)

“La demografía se resuelve en la descripción estadística de la población humana en lo que respecta: a su estado (cifra de población, distribución por sexo, por edad y estado civil, estadísticas de familia, etc.), en una fecha dada y a los hechos Demográficos (nacimientos, defunciones, celebración o disolución de uniones) que se producen en esas poblaciones.” (Pressart., 1970)

“La demografía es el estudio del tamaño, distribución geográfica y composición de la población, sus variaciones y causas pueden identificarse como fecundidad, mortalidad, movimientos territoriales (migraciones) y movilidad social (estados).” (Ducan, 1962)

Capítulo 1. Fundamentación Teórica.

"La ciencia demográfica es un sistema de conocimientos científicos. Su objeto de investigación es la población y su contenido, las leyes de su desarrollo, los cambios de las condiciones de trabajo y de la vida, o sea, la reproducción de la población en el amplio sentido del concepto (movilidad social, renovación natural de las generaciones, migraciones de sus distintas formas, cambios cuantitativos, la capacitación profesional y especializada, así como la salud)." (Revista de Ciencias Sociales, 1974)

La demografía se vale de la observación, la calorificación y el análisis de datos o información proveniente generalmente de la comunicación entre dos personas: el entrevistador y el entrevistado. Estos datos están sujetos a errores atribuibles al proceso de la comunicación, por lo que surge la necesidad de aplicar determinadas evaluaciones y ajustes a los mismos.

La disciplina demográfica mantiene una estrecha relación con ciencias como: Economía, Sociología, Geografía, Biología, Medicina, Matemática, etc. Esta relación se hace aún más evidente en el campo de los "Estudios sobre la Población", las variables demográficas (fecundidad, mortalidad, migraciones) resultan por lo general funciones de variables tales como: producto nacional, investigación, empleo (económicas), clima, relieve del terreno (geografías), tradiciones o costumbres (sociologías), etc.

Como se planteó anteriormente la demografía depende de la observación y posteriormente del registro de los sucesos que se verifican en las poblaciones humanas, de acuerdo a tiempo y espacio. El registro de los sucesos tiene categoría de estático: este incluye los censos de población y las encuestas por muestreo.

Esta disciplina aporta los datos necesarios para la elaboración de tasas y otros indicadores sanitarios, haciendo una relación entre la población afectada por un fenómeno de la salud y la población expuesta, manifestando los recursos sanitarios referidos a la población atendida. Realiza los estudios epidemiológicos que necesitan datos de la población y de su distribución según las características de la persona y el lugar; planifica y programa en salud pública la forma de precisar el volumen y la estructura de la población, para de esta manera prevenir su evolución en un tiempo determinado. Debido al gran volumen de información que maneja la ciencia demográfica, es de vital importancia su control y organización, haciéndose necesaria la existencia de sistemas que permitan y faciliten el control estadístico de este tipo de información.

1.2 Soluciones para el control estadístico de datos demográficos.

En el mundo actual existe un alto grado de competencia entre las grandes compañías, donde es menester para los directivos proponer innovadoras ideas dentro del ámbito gerencial. Esto provoca que las empresas perfeccionen las tecnologías para su uso y comercialización, enfocándolas en los clientes pues la actual economía mundial así lo requiere. La principal arma que se ha desarrollado para contrarrestar esta competencia son los sistemas de análisis de información que permiten estudiar los datos históricos y actuales. Dentro de estos sistemas las consultas se tornan más complejas, cuando se almacena todo lo histórico que posee la organización junto a la información diaria. Basadas en esta práctica, las compañías a nivel mundial han empezado a migrar hacia su actualización, para lograr un posicionamiento ventajoso en este sentido.

El notable desarrollo de la informática, particularmente en la aplicación de la computación a la investigación científica, plantea a los demógrafos la necesidad de aplicar estos nuevos métodos para ampliar la posibilidad de investigar temas, cuyo estudio no ha sido abordado por requerir amplios y extensos cálculos matemáticos. (Miro, 2006)

La aplicación Movimiento Natural de la Población, versión 1 (*MNPv1*), es una herramienta destinada a la gestión administrativa e informática de los datos del Movimiento Natural de la Población en Andalucía. Dicha gestión se inscribe en el marco de lo dispuesto en el convenio de colaboración del Instituto Nacional de Estadísticas (INE) y el Instituto de Estadística de Andalucía (IEA). (Mesa, 2002).

Otro sistema para el control estadístico de datos demográficos es el *openGis-EIEL* que ha sido diseñado para la gestión de Encuestas de Infraestructuras y Equitaciones Locales (EIEL). (Iguanahosting.com, 2009)

En nuestro País ha venido creciendo una formación tecnológica sobre el tema de los Almacén de Datos (en inglés *Data WareHouse* (DW)). No obstante, faltan varios aspectos por mejorar, pero se han visto pasos de avance dentro de esta rama.

El Departamento de Población de la ONE cuenta con el sistema de almacenamiento y análisis de datos (SIDEMO), el cual está desarrollado sobre *VisualPro* y corre sobre el Sistema de Procesamiento de Censo y Encuesta (CSpro). Este último es un paquete de software de dominio público para el tratamiento de datos de encuestas y censos. CSpro está financiado por la Oficina de Población de la Agencia de los

Estados Unidos para el desarrollo Industrial en muchos países del mundo. Otro sistema que se opera en el área de Población es el *DV survery* que es para el procesamiento de encuestas.

Anteriormente se han expuesto soluciones de almacenamiento de datos existentes en Cuba y en el mundo, las cuales presentan características y ventajas dignas de destacar. Sin embargo, estos almacenes a pesar de contar con propiedades avanzadas en el tema, no garantizan una solución a las necesidades informáticas detectadas en el presente trabajo, debido a que parte de ellos no están orientados a los mismos temas de interés, la mayoría son propietarios y traen costos significativos y lo que se desea es migrar completamente al país las soluciones libres. Además, el crecimiento de la industria del software de un país depende de su desarrollo económico y cultural. Las aplicaciones que existen actualmente en el País, no garantizan todos los análisis y reportes de temas demográficos que se desean realizar, es por esto que se determina efectuar un estudio de tecnologías de almacenamiento para seleccionar la que permita dar solución al problema planteado.

1.3 Tecnologías de almacenamiento de datos.

Las tecnologías de almacenamiento de datos poseen la capacidad de almacenar y recuperar información relacionada entre sí. Existen varios tipos de estas tecnologías, entre ellos tenemos las bases de datos, los Almacenes y Mercados de datos.

1.3.1 Bases de Datos.

Una base de datos se **define** como un depósito que permite guardar grandes volúmenes de información de forma organizada y posibilita su utilización fácilmente. Es un sistema de datos almacenados en discos, que permite el acceso directo a ellos. (Masadelante, 2009)

Los primeros ficheros de texto empezaron a guardarse secuencialmente y relacionados. Para la relación entre los ficheros se contaba con una aplicación, que tenía que actualizarse cuando se añadiera o cambiara sin depender de ningún dato. Existen bases de datos jerárquicas y de redes. *Bechman* a finales de los 60 y principios de los 70, aportó ideas para las bases de datos relacionales. En el año 76 *Chen* creó el primer sistema de bases de datos relacionales.

Entre las principales **características** de las bases de datos se destacan las siguientes: (MorenoOrtíz, 2000)

- Redundancia mínima.
- Integridad de los datos.
- Respaldo y recuperación.
- Seguridad de acceso y auditoría.
- Consultas complejas optimizadas.
- Independencia lógica y física de los datos.
- Acceso concurrente por parte de múltiples usuarios.
- Acceso a través de lenguajes de programación estándar.

La utilización de Bases de datos presenta **ventajas** y **desventajas** como: (Dr. Mario Piattini)

Ventajas	**Desventajas**
- Control sobre la redundancia de datos. - Consistencia de datos. - Compartición de datos. - Mantenimiento de estándares. - Mejora en la integridad de datos. - Mejora en la accesibilidad de datos. - Mejora en la productividad. - Mejora en el mantenimiento. - Mejora de los servicios de copia y seguridad.	- Complejidad. - Coste de equipamiento adicional. - Vulnerable a los fallos.

Tabla 1: Ventajas y desventajas de las Bases de Datos.

1.3.2 Almacenes de Datos.

Un almacén de datos es una colección de datos que tiene varias características. Una de ellas es que está orientado hacia la información y otra que se diseña para consultar eficientemente información relevante de la organización.

Los Almacenes de Datos han estado al frente de las solicitudes de las tecnologías de la información, desde la década de los noventa, como una forma para que las organizaciones utilicen la información digital en la planificación empresarial y en la toma de decisiones. Un entendimiento de la arquitectura del sistema de almacenamiento de datos es y seguirá siendo importante en las funciones y responsabilidades de la gestión de información.

Los Almacenes de Datos basados en los sistemas de información son el hogar de los datos, que se originan de una aplicación, ya sea de un sistema externo o una fuente. Optimizan la consulta de las bases de datos y las herramientas de informe.

Garantizan que el personal encargado del proceso de toma de decisiones de una empresa, extraiga la información de manera rápida y sencilla. Los Almacenes de Datos son analíticos y están orientados hacia temas específicos y organizados para las transacciones globales.

El equipo de Almacenes de Datos debe definir qué datos entran al almacén y en qué parte en particular se pueden encontrar. Algunos datos se internan en una organización, en casos como estos la información se puede obtener de otra fuente. La creación de programas de extracción para reunir datos en una zona de concentración fuera del almacén, queda en manos de un equipo de analistas y programadores; así se asegura que los datos no presenten errores para copiarlos luego en el almacén de datos. La fuente de extracción de datos, selección y proceso de transformación es la única para el almacenamiento de datos. Para el éxito de un proyecto de almacén de datos el análisis de datos de origen y el movimiento eficiente y preciso de los datos de origen son fundamentales en su entorno.

Los Almacenes de Datos están integrados, esto significa que se construyen mediante la integración de fuentes de datos múltiples y heterogéneos. En estos Almacenes de Datos se aplican técnicas de limpieza e integración. Se puede decir también que son no volátiles, por lo que los datos no cambian una vez que se

encuentran en el almacén, ya que la actualización de la base de datos operacional, no ocurre en el entorno del almacén de datos. Además, son variables en el tiempo porque los datos están asociados a un instante en el tiempo. (Sierra, 2009)

La tecnología de los Almacenes de Datos integra las técnicas de bases de datos y del análisis de los datos.

Los Almacenes de Datos permiten un análisis inmediato de los resultados esperados, además de poseer la capacidad de analizar y explorar las diferentes áreas de trabajo y la relación con el cliente. También facilitan la gestión y análisis de los recursos y relacionan departamentos empresariales. Los Almacenes de Datos son capaces de reaccionar rápidamente a los cambios del mercado.

Se construyen porque muchas compañías necesitan que la información sea comprensible para ampliar su negocio, estos muestran datos no filtrados, dispersos y nuevas formas de presentación.

Hay varias formas de desarrollar un almacén de datos. Sin embargo, hay diferentes aristas que necesitan ser consideradas: el alcance, la redundancia de datos y el tipo de usuario final.

1.3.3 Mercados de Datos.

Los Almacenes de Datos pueden ser divididos en unidades lógicas más pequeñas llamadas Mercados de Datos, estos son una versión especial de los Almacenes de Datos. Estas áreas contienen una visión de datos operacionales que ayudan a la toma de decisiones sobre las estrategias del negocio. La creación de un Mercado de Datos especifica la necesidad de los datos seleccionados, enfatizando el fácil acceso a una información importante.

Un Mercado de Datos es la vista de un almacén de datos que se define para satisfacer las necesidades de un área o sección de una empresa con menos información detallada y más agregaciones.

Los DW resuelven problemas de análisis de grandes cantidades de información en organizaciones donde una pequeña diferencia en el valor de una variable, puede afectar el resultado, perturbando el proceso de toma de decisiones y el estado financiero de la empresa. Mientras que los Mercados de datos han surgido para definir los requisitos más fácil y rápidamente; y resuelven aplicaciones a nivel departamental. (Lauro Soto, 2009)

Un Mercado de Datos centraliza la información de gestión, evitando respuestas distintas a una misma pregunta. También posibilita tener una visión global de la información en base a los conceptos de negocios que tratan los usuarios. Además, reduce el coste, posibilitando dedicar recursos a otras tareas; mejora la calidad de la gestión a partir de información relevante con un significado homogéneo y establece una base única del modelo de información de las empresas y organizaciones. (Sierra, 2009)

Existe un conjunto de interrogantes que resultan imprescindibles evaluar antes de decidir realizar la construcción de un Almacén de Datos o de un Mercado de Datos. (Lauro Soto, 2009)

- ¿La aproximación se realizará de arriba hacia abajo *(top-down)* o de abajo hacia arriba *(bottom-up)*?
- ¿Empresarial o departamental?
- ¿Cuál es primero, el Almacén de Datos o el Mercado de Datos?
- ¿Construir un piloto o directamente el Almacén de Datos completo?
- ¿Mercados de Datos dependientes o independientes?

A continuación se establece una comparación entre Almacenes de Datos y Mercados de Datos:

Almacén de Datos	**Mercado de Datos**
- Corporativo o red empresarial. - Unión de datos. - Datos recibidos del área de procesamiento. - Consulta sobre la presentación de recursos. - Estructura para vista corporativa de datos.	- Departamental. - Un simple proceso de negocio. - Unión en forma de estrella (hechos y dimensiones). - Tecnología para acceso a los datos y el análisis. - Estructura para adaptarse a la vista de datos departamentales.

Tabla 2. Almacén de datos vs Mercado de datos.

Por lo anteriormente planteado y teniendo en cuenta las características y ventajas de un Mercado de datos, se concluye que los mismos dan solución a los problemas de almacenamiento, recuperación y presentación de la información almacenada en el departamento de Población de la Oficina Nacional de Estadísticas, garantizando los principales análisis y cruces de variables .

Definido el Mercado de Datos cómo la tecnología de almacenamiento adecuada para la solución, se debe seleccionar una metodología que defina cómo llevar a cabo este tipo de desarrollo de software.

1.4 Metodologías para el desarrollo.

En la tecnología de Almacenes de Datos se han destacado un conjunto de metodologías que caracterizan y orientan todo el proceso de desarrollo. Existen criterios que han marcado fuertemente su tendencia, manejando a la comunidad mundial en este tema. Dichas metodologías son conocidas como la Metodología *Kimball* en honor a su creador *Ralph Kimball* y la de *Enmona* nombrada también por su creador *William H. Enmona*.

Basado en estas propuestas se han desarrollado metodologías que no se rigen estrictamente por una en específico, sino que presentan una selección de lo esencial de cada una y definen su propia metodología. A continuación se explican brevemente algunas de ellas:

- Metodología *SQLBI*, acreditada por *Microsoft* y orientada a *Microsoft SQL Server*, *SQL Server Analysis Services* y *Microsoft Suite foro Business Intelligence*.

- Metodología para el diseño Conceptual de Almacenes de Datos, presentada en la tesis de Doctorado de Leopoldo Zenaido Zepeda Sánchez. Aporta la incorporación de los Casos de Uso para guiar el proceso de desarrollo y define una serie de transformaciones para llevar desde un diagrama relacional a uno dimensional y obtener la estructura que conformará el repositorio de los datos. (Grupo_Profesionales, 2009)

Para definir la metodología propuesta de desarrollo a utilizar en la Línea de Almacenes de Datos e Inteligencia de Negocio (BI) del Centro de Tecnologías de datos (DATEC), se tomó como cimiento el estudio de la tesis de Doctorado de Leopoldo Zenaido Zepeda Sánchez anteriormente explicado y la Metodología para el diseño de *Kimball* se tuvo en cuenta por lo siguiente:

- Crea los conceptos de hechos y dimensiones, que garantiza que sea eficaz el proceso de toma de decisiones y proporciona mayor agilidad en el proceso de desarrollo.
- Propone incluir el almacén de datos a través de la construcción de los Mercados de datos departamentales, lo que forma una buena estrategia y coincide con la división lógica de las empresas, entidades, organismos, etc. (Grupo_Profesionales, 2009)

1.4.1 Justificación de la metodología a usar.

Basándose en su papel como órgano rector de la estadística en Cuba, la Oficina Nacional de Estadísticas merece la utilización de una metodología robusta que garantice integrar la información disponible actualmente de manera satisfactoria.

Por lo anteriormente planteado, asumiendo el estudio de las metodologías explicadas y teniendo en cuenta las características de la Universidad de Ciencias Informáticas (UCI) se adopta la metodología *Kimball* junto al estudio realizado en la tesis doctoral de Zepeda Sánchez, para enfrentar el desarrollo del Mercado de Datos del departamento de Población en la ONE. La unión de estas metodologías es una propuesta de DATEC que se ha utilizado en varios proyectos con resultados satisfactorios. Además de conseguir realizar el análisis de los datos que se encuentran integrados, con un mejor entendimiento por parte de los usuarios, en cuanto a la forma de almacenar la información. Es importante mencionar que es una propuesta de metodología resistente y adaptable ante los cambios.

Ya establecida la metodología de desarrollo para soluciones de Mercados de datos, es necesario definir el tipo de Modelo de Base de Datos que se adapte a la tecnología de almacenamiento seleccionada.

1.5 Modelos de Bases de Datos.

Un modelo de base de datos es una colección de conceptos definidos matemáticamente que ayuda a las propiedades estáticas y dinámicas de una aplicación. Un modelo se distingue de otro por el tratamiento que le brinda a las aplicaciones. (MorenoOrtíz, 2000)

Existen varios tipos de modelos de bases de datos, a continuación se explicarán: el Modelo Entidad – Relación y el Modelo Dimensional.

1.5.1 Modelo Entidad-Relación.

Un modelo entidad – relación (*"Entity Relationship", E-R* o Diagrama de Entidad Relación, *"DER"*) es un lenguaje para el modelado de datos de un sistema de información, que expresa las entidades más relevantes para el sistema, así como sus inter-relaciones y propiedades. Divide los datos en entidades moderadas donde cada una contiene una tabla física en la base de datos operacional.

Los modelos E-R no son meritorios para el diseño de Mercados de datos, ya que estos no garantizan la recuperación perfecta de la gran acumulación de información que se almacena. (Sierra, 2009)

1.5.2 Modelo Dimensional.

Como desigualdad con los sistemas de bases de datos más comunes que presentan un diseño de sus estructuras mediante el modelo Entidad-Relación, los Mercados de datos se diseñan mediante un modelo dimensional, el cual a diferencia del modelo E-R, presenta la información de manera más organizada, garantizando la velocidad y eficacia en la recuperación de datos.

Comúnmente la propuesta para desarrollar este tipo de modelo es la confeccionada por *Ralph Kimball* llamada "esquema estrella", esta consiste en una tabla central llamada "tabla de hechos" y un conjunto de tablas llamadas "dimensiones" que se relacionan a esta tabla central. Se le denomina Modelo estrella al esquema que representa las relaciones entre las tablas de dimensiones y la tabla central de hechos. Debe su nombre a su similitud con una estrella natural. (Sierra, 2009)

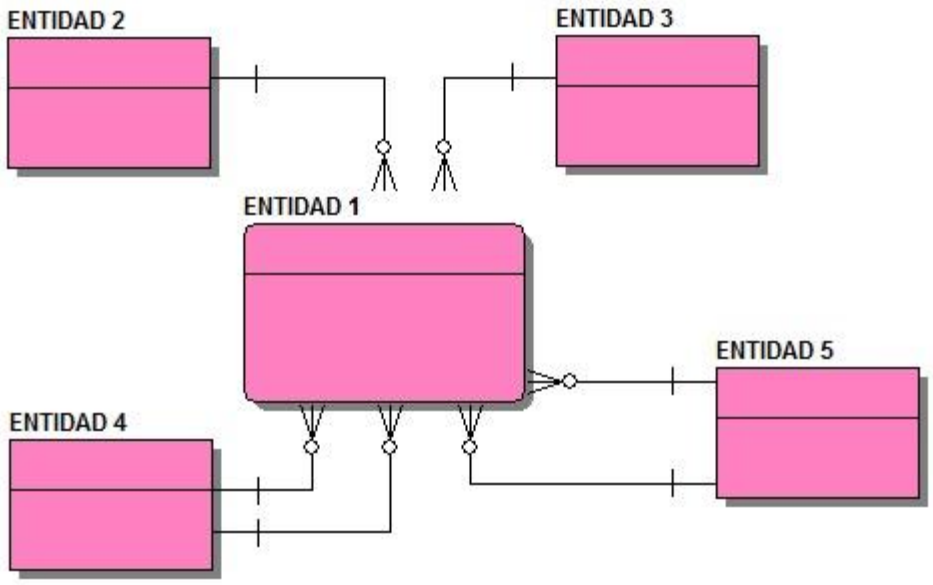

Figura 1. Representación de un Esquema Estrella.

Copo de Nieve (*Snowflake*, en inglés) es otra estructura que surge como producto de modificaciones realizadas al modelo estrella; el objetivo principal de la estructura Copo de Nieve es el ahorro de espacio de almacenamiento. Cuando los atributos de baja calidad se llevan a tablas diferentes se dice que la dimensión se encuentra en "*Snowflake*". Este tipo de estructura hace que las presentaciones sean más complejas, afectando el rendimiento de la recuperación de consultas.

Las subdimensiones se pueden utilizar de forma similar al Copo de Nieve, pero solo se recomienda utilizarlas cuando existe un grupo de atributos dentro de las dimensiones que es necesario separar. A esta estructura se le puede añadir la constelación de hechos teniendo como principal característica, múltiples tablas con dimensiones comunes, de esta manera, se logran utilizar varias medidas, separadas en diferentes tablas de hechos y definidas en las mismas dimensiones. (Sierra, 2009)

El modelo dimensional divide el universo de datos en dos grandes grupos: las medidas y las descripciones de estas medidas. Las medidas son generalmente numéricas, almacenadas en tablas de hechos y en las tablas de dimensiones se almacenan las descripciones textuales de los entornos. Las tablas de hechos son las que prevalecen en el modelo dimensional, englobando valores del negocio. Los hechos más

usuales son valores numéricos. Cada tabla es una interrelación muchos – muchos y posee dos o más llaves foráneas que acoplan con sus respectivas tablas de dimensiones. A continuación se establece una comparación entre el Modelo de Entidad - Relación y el Modelo Dimensional.

Modelo Entidad-Relación	Modelo Dimensional
- Presenta entidades con sus inter-relaciones. - Los datos están divididos en entidades moderadas, contenidas en una tabla física en la base de datos operacional. - No garantiza la recuperación perfecta del volumen de datos almacenados.	- Presenta tablas de hechos y dimensiones relacionadas a esta tabla de hechos. - Divide los datos en medidas y descripciones de dichas medidas. - Los valores del negocio se engloban en las tablas de hechos. - Cada tabla es una inter-relación muchos –muchos.

Tabla 3. Comparación entre Modelo Entidad – Relación y Modelo Dimensional

Partiendo del estudio realizado sobre los modelos de bases de datos, se define para la solución, el Modelo Dimensional, el cual se ajusta a las características de la tecnología de almacenamiento seleccionada.

Para diseñar el modelo de Base de Datos es necesario el uso de una herramienta de modelación a través de la cual se realice el diseño lógico del Mercado de Datos.

1.6 Herramientas de Modelado.

Una herramienta de modelado facilita llevar el diseño y construcción de la Base de Datos del nivel lógico a un nivel físico. Un modelo de datos describe los datos y sus relaciones, su significado y solidez.

1.6.1 *ERwin*.

ERwin es una de las herramientas *CASE* que se utiliza para el diseño de bases de datos, que brinda productividad en su diseño, generación, y mantenimiento de aplicaciones, desde un modelo lógico de los requerimientos de información, hasta el modelo físico perfeccionado con las características específicas de la base de datos diseñada. Además, *ERwin* permite visualizar la estructura, los elementos importantes y optimizar el diseño de la base de datos. Genera automáticamente las tablas y miles de líneas de *stored procedure* y *triggers* para las principales clases de bases de datos.

ERwin hace fácil el diseño de una base de datos. Los diseñadores de bases de datos sólo pulsan un botón para crear un gráfico del modelo E-R de todos los requerimientos de datos y capturar las reglas de negocio en un modelo lógico, mostrando todas las entidades, atributos, relaciones y llaves importantes.

La migración automática garantiza la integridad referencial de la base de datos. *ERwin* establece una conexión entre una base de datos diseñada y una base de datos no diseñada, permitiendo la transferencia entre ambas y la aplicación de la ingeniería inversa. Usando esta conexión, *ERwin* genera automáticamente tablas, vistas, índices, reglas de integridad referencial (llaves primarias, llaves foráneas), valores por defecto y restricciones de campos y dominios.

ERwin soporta principalmente bases de datos relacionales *SQL* y bases de datos que incluyen *Oracle, Microsoft SQL Server* y *Sybase*. El mismo modelo puede ser usado para generar múltiples bases de datos, o convertir una aplicación de una plataforma de base de datos a otra, por lo que es una herramienta que ayuda a alcanzar los resultados esperados en la construcción de un Almacén de Datos. (Michael Kornspan, 2010)

1.6.2 *Visual Paradigm*.

Visual Paradigm con el Lenguaje Unificado de Modelado (UML) es una herramienta que soporta el ciclo de vida completo del desarrollo de un software, es decir, que sobrelleva las distintas fases de un proyecto que son: análisis y diseño orientado a objeto, construcción, pruebas y despliegue.

El Modelado UML contribuye a la rápida construcción de aplicaciones con calidad, mejores y a un menor coste. Permite dibujar todos los tipos de diagramas de clases, código inverso, generar código desde

diagramas y generar documentación. La herramienta *UML CASE* también proporciona abundantes tutoriales, demostraciones interactivas y proyectos *UML*. (Free Download Manager.ORG., 2007)

Visual Paradigm es un tipo de las herramientas *CASE*, estas se clasifican en tres categorías o terminologías: *CASE* de Alto Nivel, *CASE* de Bajo Nivel y *CASE* Cruzado de Ciclo de vida.

- *CASE* de Alto Nivel son aquellas herramientas que automatizan o apoyan las fases iniciales del ciclo de vida del desarrollo de sistemas como la planificación, el análisis y el diseño de sistemas.
- *CASE* de Bajo Nivel son aquellas herramientas que automatizan o apoyan las fases finales o inferiores del ciclo de vida como el diseño detallado, la implantación y el soporte de sistemas.
- *CASE* Cruzado de Ciclo de vida son aquellas herramientas que apoyan actividades que tienen lugar a lo largo de todo el ciclo de vida, se incluyen actividades como la gestión de proyectos y la estimación.

Visto esto podemos argumentar que *Visual Paradigm* es una Herramienta *CASE* que se caracteriza por soportar las últimas versiones de *UML* y la Notación y Modelado de Procesos de Negocios, además de ser un generador de mapeo de objetos-relacionales para los lenguajes de programación *Java*, *.NET* y *PHP*.

Tiene conexión con *Rational Rose* en sus archivos de proyecto, además pueden ser importados a *Visual Paradigm UML* a través de esta importante característica. Para maximizar la interoperabilidad de los productos de *Visual Paradigm* con otras aplicaciones, se introdujo la importación y exportación de modelos de proyectos desde o hasta un formato *XML*. Los usuarios y proveedores de tecnologías pueden integrar esta herramienta en cada uno de sus modelos para utilizarlos en sus soluciones con un mínimo esfuerzo.

Visual Paradigm está integrado con varias herramientas *Java*, estas son: (Free Download Manager.ORG., 2007)

- *Eclipse/IBM WebSphere.*
- *JBuilder.*
- *NetBeans IDE.*
- *Oracle JDeveloper.*
- *BEA Weblogic.*

Algunas de las **características** que presenta *Visual Paradigm* son: (Free Download Manager.ORG., 2007)

- Ingeniería inversa.
- Generación de código.
- Ingeniería de ida y vuelta.
- Diagramas de flujos de datos.
- Soporte de *UML* versión 2.1.
- Generación de bases de datos.
- Editor de detalles de casos de uso.
- Diagramas de procesos del negocio.
- Ingeniería inversa de bases de datos.
- Distribución automática de diagramas.
- Ingeniería inversa *Java*, C++, esquemas *XML*, *XML, .NET, .exe, .dll.*
- Modelado colaborativo con *Concurrent Versions System (CVS)* y *Subversion*.

Visual Paradigm presenta varios **beneficios**, entre ellos se encuentran:

- Navegación intuitiva entre el código y el modelo.
- Poderoso generador de documentación y reportes *UML PDF/HTML/MS Word*.
- Demanda en tiempo real, modelo incremental de viaje redondo y sincronización de código fuente.
- Superior entorno de modelado visual.
- Soporte completo de notaciones *UML*.

Diagramas de diseño automático sofisticado. (Free Download Manager.ORG., 2007)

Visual Paradigm es la herramienta definida por la Dirección Técnica para el modelado conceptual de la información, pues la licencia de esta fue adquirida por la Universidad de las Ciencias Informáticas (UCI), por lo tanto es la que se utiliza para la solución.

Teniendo en cuenta la herramienta seleccionada, que soporta el modelado dimensional del Mercado de Datos, se requiere un sistema que permita crear una estructura de almacenamiento física para el diseño dimensional creado con *Visual Paradigm*.

1.7 Gestores de Base de Datos.

Un Sistema Gestor de base datos (SGBD) es un conjunto de programas que dan paso a la creación y mantenimiento de una base de datos, asegurando la integridad, confidencialidad y seguridad de la misma.

Los SGBDs permiten guardar información en bibliotecas para posteriores consultas. De manera más específica, cada empresa en particular guarda la información de sus clientes, proveedores y demás personal que intervienen en su desarrollo. La información de los SGBDs debe ser real y estructurada para poder recuperarla fácilmente, debe estar disponible en cualquier momento, ser consistente, no debe tener repeticiones ni redundancia y debe ser coherente. También un Sistema Gestor de base datos debe permitir el acceso de múltiples usuarios de la base de datos a la vez, a una misma información y realizar consultas complejas e imprevistas. Además debe existir un sistema de seguridad para los datos. Algunos de los Gestores de Base de Datos existentes son:

FoxPro permite hacer uso de los conocimientos de *Visual FoxPro* y habilidades para resolver tareas de manera eficiente y eficaz. Es un sistema orientado a objeto para la generación de base de datos y el desarrollo de aplicaciones. Ayuda a escribir código más rápido, con menos errores y para una gama amplia de usuarios. *FoxPro* crea formularios, base de datos, informes, vistas, proyectos, entre otros. Además realiza los elementos de un proyecto por separados. (Microsoft Corporation, 2010)

Microsoft SQLServer presenta aprisionamientos, operaciones de índice paralelas, espejos de bases de datos, gestión de recursos, comprensión de copias de seguridad y adición de memoria en cliente, además de contener fotografía instantánea de base de datos, información codificada de forma trasparente, revisión de cuentas seguras, *data-marts* escalables y reportes, comprensión de información, optimización de consultas, captura de datos modificados, vistas particionadas e índices alineados. También posee funciones analíticas avanzadas, algoritmos avanzados de minería de datos y servicios de integración, análisis y reportes. *Microsoft SQLServer* tiene un buen desempeño de la recopilación de información y tiene además estructuras de políticas de gestión. (Microsoft, 2008)

Oracle ofrece un rápido, fiable y seguro intercambio de información, análisis y extracción de datos a un bajo costo y redes escalables. Obtiene un rendimiento extremo y la escalabilidad de los Mercados de datos, gestiona la carga e integración de datos, mejora el rendimiento del Mercado de Datos, así como la disponibilidad y capacidad de administrar grandes tablas de partición. (Oracle, 2010)

PostgreSQL es un poderoso objeto de código abierto y un sistema de base de datos relacional. Cuenta con una arquitectura probada que se ha ganado una sólida reputación de confiabilidad, integridad de datos y corrección. Funciona en los principales sistemas operativos incluyendo a *Linux* y *Windows*.

Tiene soporte completo para claves foráneas, uniones, vistas, disparadores y procedimientos almacenados en varios idiomas. Además, es compatible con el almacenamiento de objetos binarios, incluyendo imágenes, sonidos y videos y tiene interfaces de programación nativo de *C/C++, Java, .Net, Perl, Python, Ruby, Tcl*, entre otros.

Cuenta con sofisticadas funciones como la Versión Multi-Control de Concurrencia (MVCC), *tablespaces*, replicación asíncrona, transacciones anidadas, un planificador de consultas sofisticadas y escribe por delante de registros para la tolerancia de fallos.

PostgreSQL ofrece muchas ventajas para una empresa o negocio sobre un sistema de base de datos. Modela el negocio más rentable con un despliegue a gran escala y presenta flexibilidad para la investigación e implementación de pruebas sin necesidad de incluir los costes de licencia adicionales.

El código fuente está disponible para todos sin costo alguno. Está disponible además para casi todas las marcas *Unix* y la compatibilidad para Windows. Usa la estrategia MVCC extremadamente sensible en entornos de alto volumen para *PostgreSQL*. (Group, 1996-2010.)

PostgreSQL requiere de clientes de Base de Datos que faciliten su utilización, entre los que se encuentra p*gAdmin III PostgreSQL Tools v 1.10.0.* El cual es el cliente de *PostgreSQL* más popular, con la base de datos de fuente abierta más avanzada del mundo. Está diseñado para responder a las necesidades de todos los usuarios. Además, su interfaz gráfica soporta todas las características de *PostgreSQL* y facilita la administración, p*gAdmin III PostgreSQL Tools* se desarrolla por la comunidad de expertos de *PostgreSQL* en todo el mundo y está disponible en más de una docena de idiomas. Es un software libre publicado bajo la licencia de *PostgreSQL*.

Por el estudio realizado anteriormente se define *PostgreSQL* como el gestor de base de datos a utilizar, pues garantiza la obtención de los resultados esperados en el desarrollo del Mercado de Datos de la solución. La información almacenada en el Gestor de Base de Datos requiere de un modo de almacenamiento que determine su estructura.

1.8 Modos de Almacenamiento de Datos.

Procesamiento Analítico Relacional en Línea (*Relational Online Analitical Process, ¨ROLAP¨*), Procesamiento Analítico Multidimensional en Línea *(Multidimensional Online Analitical Process, ¨MOLAP¨)* y Procesamiento Analítico Híbrido en Línea (*Hybrid Online Analitical Process,* ¨HOLAP¨) son tres modelos para el proceso analítico en línea (OLAP). En ellos el proceso de análisis se desarrolla de igual forma, variando en uno y otro caso, la metodología de almacenamiento. (Sierra, 2009)

1.8.1 ROLAP.

Los datos son almacenados en filas y columnas de forma relacional en el Procesamiento Analítico Relacional en Línea.

Los datos de los usuarios se presentan en forma de dimensiones de negocio. La semántica de las etiquetas de los metadatos es creada para encubrir las estructuras de almacenamiento y mostrar los datos dimensionales. Estas soportan el mapeo de las dimensiones a las tablas relacionales. Los metadatos son almacenados en tablas relacionales. Este modelo es utilizado fundamentalmente cuando la información no se consulta frecuentemente

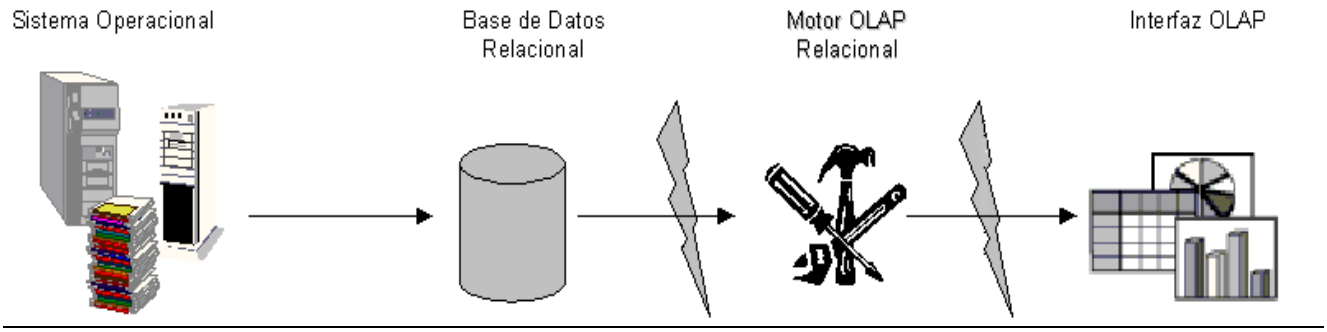

Figura 2. Modelo de almacenamiento ROLAP. (Gómez, y otros)

1.8.2 MOLAP.

El Procesamiento Analítico Multidimensional en Línea almacena los datos de forma dimensional a diferencia del ROLAP. Las estructuras de datos están fijas para la lógica, permitiendo que al tratar la información pueda estar basada en métodos determinados estableciendo las coordenadas del almacenamiento de datos.

Las estructuras de almacenamiento son una copia de la fuente de datos y físicamente permanecen en la misma estación de trabajo donde está instalada la herramienta *Data Warehousing*. Esto provoca que el acceso a la información sea de manera rápida y efectiva utilizándose el depósito donde el tiempo en velocidad de respuesta es crítico. (Sierra, 2009)

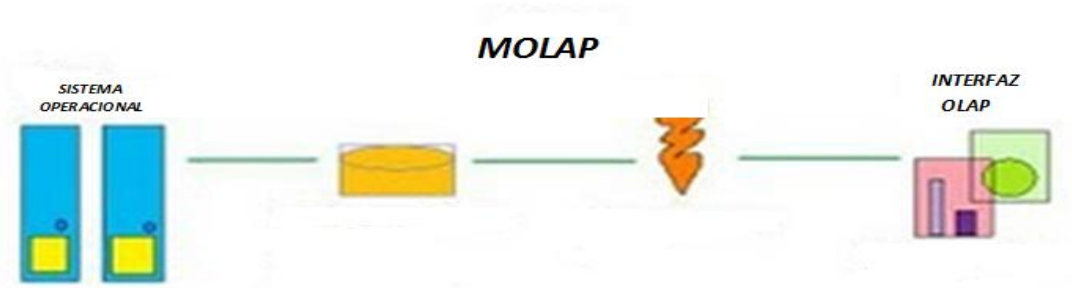

Figura 3. Modelo de almacenamiento MOLAP. (Gómez, y otros)

1.8.3 HOLAP.

El Procesamiento Analítico Híbrido en Línea es un híbrido entre los métodos de ROLAP y MOLAP que permite almacenar una parte de los datos como un sistema ROLAP y el resto como uno MOLAP. El control que ejerce el operador de la aplicación sobre este aprisionamiento varía de un producto a otro.

El aprisionamiento vertical almacena las agregaciones como un MOLAP para mejorar la velocidad de las consultas, detallando los datos en ROLAP para optimizar el tiempo en que se procesa en cubo. El aprisionamiento horizontal en el modo HOLAP almacena una sección de datos, generalmente los más recientes en modo MOLAP para mejorar la velocidad de consulta y los datos más antiguos en ROLAP. Se pueden almacenar algunos cubos en MOLAP y otros en ROLAP. (Sierra, 2009)

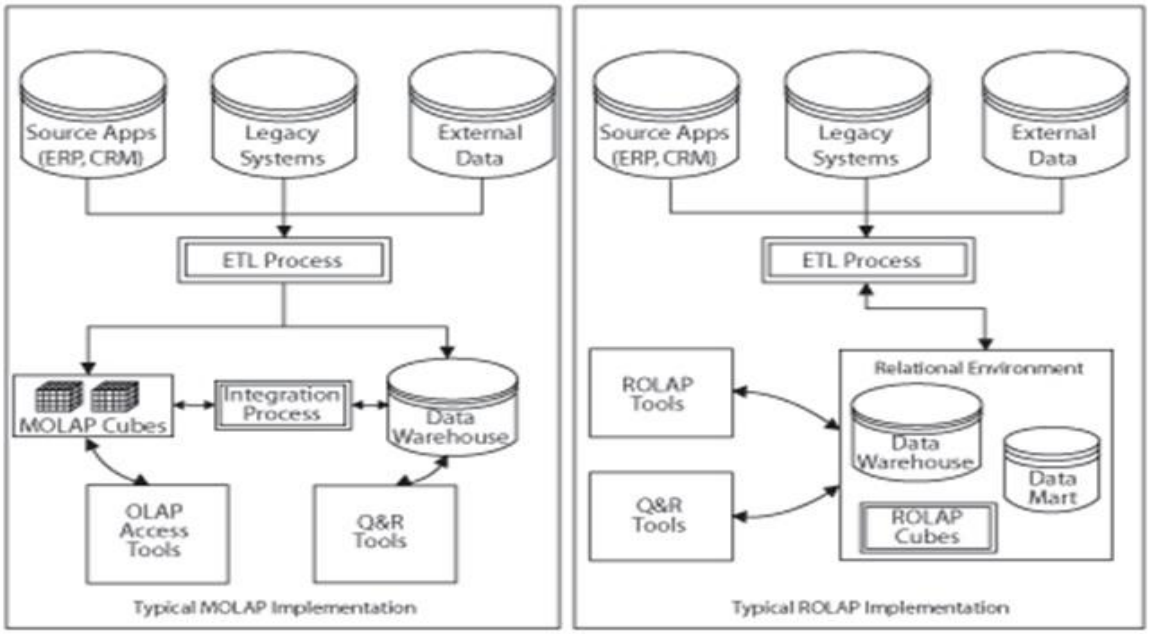

Figura 4. Modelo de almacenamiento HOLAP (Gómez, y otros)**.**

A continuación se establece una comparación entre MOLAP y ROLAP.

ROLAP	**MOLAP**
- Posee tablas relacionales que resumen los datos disponibles. - Tiene un volumen alto de datos. - Todos los datos de acceso están en la bodega de almacenamiento. - Utiliza *SQL* complejo para los datos del depósito. - Los cubos de datos son creados sobre la marcha por el servidor de	- Diversos resúmenes de datos en bases de datos propietarias. - Volumen de datos moderados. - Resúmenes de acceso a datos detallados en bases de datos Multidimensionales. - Crea cubos de datos prefabricados por el motor OLAP. - Usa tecnología propietaria para almacenar las vistas multidimensionales en arreglos.

análisis. - Muestra vistas multidimensionales en la capa de presentación. - Tiene un ambiente conocido. - Disponibilidad de herramientas. - Presenta limitaciones en funciones de análisis complejos. - Las agregaciones no son factibles.	- Utiliza una matriz de alta velocidad para la recuperación de datos. - Maneja poca tecnología de matriz de datos para gestionar los resúmenes. - Tiene una fuente de librería de funciones para el cálculo complejo. - Facilita el análisis independientemente de la cantidad de dimensiones.

Tabla 4. ROLAP versus MOLAP

Se utiliza ROLAP porque es el modo de almacenamiento establecido para *PostgreSQL*, siendo este el Gestor de Base de datos definido para la solución del Mercado de Datos del departamento de Población de la ONE.

Para garantizar la seguridad del presente trabajo es necesario utilizar una herramienta de control de versiones.

1.9 Herramienta para el control de versiones.

Subversion existe para ser universalmente reconocido y adoptado como herramienta de código abierto. Es un sistema de control centralizado que se caracteriza por su fiabilidad como refugio seguro para datos importantes. Su uso y modelo no son complejos y su capacidad es lo suficientemente amplia para apoyar las necesidades de una gran variedad de usuarios y proyectos, de individuos a escala de empresa y de grandes operaciones. (CollabNet, Inc., 2001 - 2009)

El cliente para Subversion que se utiliza en el trabajo es **TortoiseSVN**, el cual es un software de control de fuentes para Windows y fácil de utilizar. Es una herramienta de desarrollo que se aplica en diversos trabajos, además de tener un libre acceso. (CollabNet, Inc., 2001 - 2009)

Conclusiones.

En este Capítulo se estudiaron temas de demografía como: sus antecedentes, definiciones, características y otros aspectos importantes, además se realizó un análisis de la evolución y definición de las bases de datos. También se investigaron temas sobre los almacenes y Mercados de datos, como son: ventajas y características, dejando de manera clara que un Mercado de Datos es la solución para este trabajo. A partir del estudio realizado se definió que para darle solución al problema en cuestión, se adoptó el estudio de la metodología de *Kimball* junto a la investigación realizada en la tesis de Doctorado de Zepeda Sánchez como la metodología propuesta. Se abundó sobre el Modelo Entidad – Relación y el Modelo Dimensional, ambos modelos de bases de datos, donde se explicó el por qué se escogió el modelo dimensional. Se investigó sobre *ERwin*, y *Visual Paradigm*, dos herramientas de modelado, donde se estableció que la idónea para este trabajo es *Visual Paradigm* en su versión 6.4. Se plasmó también información sobre algunos de los gestores de bases de datos, determinando el uso de *PostgreSQL* v 8.4 y su cliente de base de datos p*gAdmin III PostgreSQL Tools v 1.10.0*. Para concluir se estudiaron algunos de los modos de almacenamiento (ROLAP, MOLAP, HOLAP), justificando la selección de ROLAP.

CAPÍTULO 2. ANÁLISIS Y DISEÑO.

Introducción.

En este capítulo se describe el negocio del Departamento de Población, se definen los temas de análisis, roles y permiso, las reglas del negocio y se identifican las necesidades del usuario detectadas en el área. También se muestran los requisitos funcionales, no funcionales, de información y multidimensionales. Además, se describen los casos de usos del sistema tanto los de información como los funcionales. Se confecciona la matriz BUS y el modelo de datos, además de identificarse y describirse las tablas de hechos y las dimensiones.

2.1 Descripción del negocio.

La Oficina Nacional de Estadísticas es la institución rectora de los temas estadísticos en Cuba y uno de sus objetivos es funcionar como un repositorio central, donde se lleven a cabo un conjunto de procesos que ejecutan y supervisan la gestión estadística del País.

Actualmente el mecanismo de información que se encuentra en ejecución está compuesto por diferentes fuentes. La diferencia existente entre estas está principalmente en los mecanismos de captación, en los períodos de captura (mensual, trimestral, semestral, anual, etc.), en las características específicas de la recopilación de información, en la estructura de las plantillas y otras.

La información estadística está recogida en modelos de manera organizada por cada centro informante. Dicha información está constituida bajo los diferentes niveles de estructura nacional que posee la ONE y donde cada uno de estos modelos incluye un conjunto de indicadores. Los indicadores relacionados se ajustan a cada centro informante en dependencia del trabajo que realice cada uno, ya sea social, de organismos estatales, económicos, etc.

El Departamento de Población es una de las áreas de la ONE, en ella se realiza el estudio de la población analizando su volumen y desarrollo. La información que se maneja en este departamento proviene de diferentes sectores como: el Ministerio de Salud Pública (MINSAP), los Bufetes Colectivos, Registros Civiles y el Departamento de Inmigración y Extranjería (DIE). Los datos que se procesan están en formatos DBF para Nacimientos, Defunciones, Defunciones Infantiles, Matrimonios, Divorcios, Migraciones Internas y Población.

Posterior a una descripción breve del negocio de la ONE se resalta que la información estadística captada en el área de Población se realiza a todos los niveles, almacenando desde los niveles más bajos de detalles hasta los más densos y complejos, con el objetivo de permitir la disponibilidad de la misma para su consulta con mayor rapidez y validez. Partiendo de la descripción del negocio para lograr el buen desarrollo de la solución, es necesario establecer los posibles temas de análisis.

2.2 Temas de análisis.

Los temas de análisis son áreas de una entidad o empresa a un alto nivel de información, que poseen objetivos o metas propias y son de gran importancia para el desarrollo de un Mercado de Datos. El cumplimiento de metas trazadas como son: la fiabilidad, la utilidad y el éxito de las estructuras, está dirigido hacia la realización completa y satisfactoria del Mercado de Datos. La propuesta de solución se enfoca en función de los cortes de información que comúnmente se realizan en la ONE. Esta oficina está interesada en analizar la población teniendo en cuenta varias perspectivas de análisis. El tema de análisis identificado en este trabajo es demografía. Las consultas y tratamiento de la información que giran en torno a este tema, requieren de niveles de acceso restringidos para garantizar su seguridad y rendimiento.

2.3 Roles y permisos.

Uno de los objetivos más importantes para el Departamento de Población de la ONE es contar con una aplicación distribuida a lo largo de todo el País, que permita que cualquier persona interesada en los cambios y desarrollo de la población pueda acceder a estos datos. El sistema cuenta con un rol de analista que consulta la información y un rol de administrador que es el encargado de extraer, transformar y cargar dicha información. Gran parte del funcionamiento de este sistema depende de las reglas del negocio establecidas.

2.4 Reglas del Negocio.

Las reglas del negocio son un conjunto de definiciones y restricciones establecidas por los clientes para darle tratamiento a los datos ya sea en su manipulación o interpretación. En el análisis se identificaron las siguientes reglas del negocio:

1. El código que identifica a los municipios está compuesto por cuatro dígitos, los dos primeros identifican la provincia y los dos restantes representan al municipio.

2. El número de carné de identidad posee once dígitos, los seis primeros representan la fecha de nacimiento de la persona, tomándose los dos primeros como el año, los dos seguidos como el mes y los dos que continúan como el día. Los otros cinco números restantes se analizan de la siguiente manera: el primero es el siglo en que nació la persona, si el número se encuentra del cero al cinco significa que la persona nació en el siglo XX, si está entre el seis y el ocho representa que la persona nació en el siglo XXI y si es el cero, la persona nació en el siglo XIX. Los tres dígitos que siguen identifican el sexo, este se determina si el número es impar el sexo es femenino y si es par el sexo es masculino y el último dígito es un módulo en que se guarda el número de carné de identidad.

3. Las tasas se calculan dividiendo el indicador entre la cantidad total de habitantes y multiplicándolo por 1000.

4. Las relaciones se establecen con la división de un indicador sobre otro multiplicado por 100.

5. El saldo migratorio es la diferencia entre la cantidad de inmigraciones y la cantidad de emigraciones.

6. La esperanza de vida se calcula a través de un instrumento que se denomina “tabla de vida”, la cual permite medir las probabilidades de muerte o de vida de una población en función de la edad, que junto con el sexo constituyen los dos atributos demográficos fundamentales de la misma.

7. La densidad de la población se calcula por la razón entre la cantidad de población y el área total de dicha población por cada 100 habitantes.

8. El movimiento natural no es más que la diferencia entre la cantidad de nacimientos y la cantidad de defunciones.

Después de definidas las reglas del negocio es importante determinar cuáles son los requisitos para el desarrollo de la solución.

2.5 Requerimientos.

Un requisito es una necesidad documentada sobre un contenido, forma o funcionalidad de un proceso o servicio. Existen cuatro categorías en que se clasifican los requisitos en este tipo de soluciones. Estas son: de información, funcionales, no funcionales y multidimensionales, estos últimos están asociados directamente a los requisitos informativos.

2.5.1 Requisitos de Información.

Los Requisitos de información constituyen las entradas fundamentales para futuros reportes. Para este trabajo los requisitos de información se agrupan por los distintos subsistemas que tiene el Departamento de Población de la ONE mencionadas anteriormente.

Nacimientos

RI 1. Obtener cantidad de nacimientos según la provincia de residencia de la madre.

RI 2. Obtener cantidad de nacimientos según el municipio de residencia de la madre.

RI 3. Obtener cantidad de nacimientos según la zona de residencia de la madre.

RI 4. Obtener cantidad de nacimientos según el mes de nacimiento.

RI 5. Obtener cantidad de nacimientos según la edad de la madre.

RI 6. Obtener cantidad de nacimientos según el nivel de escolaridad de la madre.

RI 7. Obtener cantidad de nacimientos según el tipo de embarazo.

RI 8. Obtener cantidad de nacimientos según el sexo del nacido.

RI 9. Obtener cantidad de nacimientos según el peso del nacido.

RI 10. Obtener cantidad de nacimientos según la provincia de ocurrencia del nacimiento.

Defunciones

RI 11. Obtener cantidad de defunciones según la provincia de residencia.

RI 12. Obtener cantidad de defunciones según el mes de ocurrencia.

RI 13. Obtener cantidad de defunciones según la edad.

RI 14. Obtener cantidad de defunciones según la causa de muerte.

RI 15. Obtener cantidad de defunciones según la ocupación.

RI 16. Obtener cantidad de defunciones según el municipio de residencia.

RI 17. Obtener cantidad de defunciones según la zona de residencia.

RI 18. Obtener cantidad de defunciones según el sexo.

RI 19. Obtener cantidad de defunciones de menores de 1 año.

Defunciones infantiles

RI 20. Obtener cantidad de defunciones perinatales, precoces y fetales tardías según la provincia de residencia de la madre.

RI 21. Obtener cantidad de defunciones perinatales, precoces y fetales tardías según la edad de la madre.

RI 22. Obtener cantidad de defunciones perinatales, precoces y fetales tardías según hijos vivos de la madre.

RI 23. Obtener cantidad de defunciones perinatales, precoces y fetales tardías según el municipio de residencia de la madre.

RI 24. Obtener cantidad de defunciones perinatales, precoces y fetales tardías según el sexo del nacido.

Matrimonios

RI 25. Obtener cantidad de matrimonios según la provincia de ocurrencia.

RI 26. Obtener cantidad de matrimonios según la provincia de residencia de los contrayentes.

RI 27. Obtener cantidad de matrimonios según la provincia de residencia de la mujer.

RI 28. Obtener cantidad de matrimonios según la provincia de residencia del hombre.

RI 29. Obtener cantidad de matrimonios según la edad combinada de los contrayentes.

RI 30. Obtener cantidad de matrimonios según la edad de la mujer.

RI 31. Obtener cantidad de matrimonios según estado conyugal anterior de los contrayentes.

RI 32. Obtener cantidad de matrimonios según estado conyugal anterior de la mujer.

RI 33. Obtener cantidad de matrimonios según estado conyugal anterior del hombre.

RI 34. Obtener cantidad de matrimonios según el nivel de escolaridad de los contrayentes.

RI 35. Obtener cantidad de matrimonios según el nivel de escolaridad de la mujer.

RI 36. Obtener cantidad de matrimonios según el nivel de escolaridad del hombre.

RI 37. Obtener cantidad de matrimonios según la zona de residencia de los contrayentes.

RI 38. Obtener cantidad de matrimonios según la ocupación combinada de los contrayentes.

RI 39. Obtener cantidad de matrimonios con extranjeros según país de residencia.

RI 40. Obtener cantidad de matrimonios con extranjeros según país de residencia del hombre.

RI 41. Obtener cantidad de matrimonios con extranjeros según país de residencia de la mujer.

RI 42. Obtener cantidad de matrimonios con extranjeros según la edad.

RI 43. Obtener cantidad de matrimonios con extranjeros según la edad de la mujer.

Divorcios

RI 44. Obtener cantidad de divorcios según provincia de residencia.

RI 45. Obtener cantidad de divorcios según provincia de residencia de la mujer.

RI 46. Obtener cantidad de divorcios según provincia de ocurrencia.

RI 47. Obtener cantidad de divorcios según mes de firmeza.

RI 48. Obtener cantidad de divorcios según la edad de la mujer al casarse.

RI 49. Obtener cantidad de divorcios según la edad del hombre al casarse.

RI 50. Obtener cantidad de divorcios según la edad de la mujer al divorciarse.

RI 51. Obtener cantidad de divorcios según la edad del hombre al divorciarse.

RI 52. Obtener cantidad de divorcios según el nivel de escolaridad de la mujer.

RI 53. Obtener cantidad de divorcios según el nivel de escolaridad del hombre.

RI 54. Obtener cantidad de divorcios según la ocupación de la mujer.

RI 55. Obtener cantidad de divorcios según la ocupación del hombre.

RI 56. Obtener cantidad de divorcios según la duración del matrimonio.

RI 57. Obtener cantidad de divorcios según el número de hijos.

RI 58. Obtener cantidad de divorcios según la edad de los hijos.

RI 59. Obtener cantidad de divorcios según el número de hijos procreadores.

Migraciones Internas

RI 60. Obtener cantidad de migraciones internas según el sexo.

RI 61. Obtener cantidad de migraciones internas según la provincia destino.

RI 62. Obtener cantidad de migraciones internas según el municipio destino.

RI 63. Obtener cantidad de migraciones internas según la zona destino.

RI 64. Obtener cantidad de migraciones internas según la provincia de procedencia.

RI 65. Obtener cantidad de migraciones internas según el municipio de procedencia.

RI 66. Obtener cantidad de migraciones internas según zona de procedencia.

RI 67. Obtener cantidad de migraciones internas según la edad.

RI 68. Obtener cantidad de migraciones internas según el Plan Turquino.

RI 69. Obtener cantidad de migraciones internas según el mes de ocurrencia.

RI 70. Obtener cantidad de migraciones internas según la situación de actividad.

Población

RI 71. Obtener población residente según el sexo.

RI 72. Obtener población residente según la edad

RI 73. Obtener población residente según la relación de masculinidad.

RI 74. Obtener población residente según la zona.

RI 75. Obtener población residente según la provincia.

RI 76. Obtener población residente según el municipio.

RI 77. Obtener población residente según la capital provincial.

RI 78. Obtener tasa de crecimiento natural según la cantidad total de habitantes.

RI 79. Obtener relación de masculinidad según la cantidad total de hombres.

RI 80. Obtener densidad de la población por Km² según la cantidad total de habitantes.

RI 81. Obtener relación de dependencia según el año.

RI 82. Obtener relación de dependencia según la provincia.

RI 83. Obtener evolución de la estructura según la edad.

RI 84. Obtener movimiento natural según la provincia.

RI 85. Obtener tasa de movimiento natural según la provincia.

RI 86. Obtener defunciones según la edad.

RI 87. Obtener defunciones según el sexo.

RI 88. Obtener defunciones de menores de un año según el sexo.

RI 89. Obtener esperanza de vida según el sexo.

RI 90. Obtener esperanza de vida según la edad.

RI 91. Obtener matrimonio según el estado conyugal anterior de los contrayentes.

RI 92. Obtener divorcio según la duración del matrimonio.

RI 93. Obtener migración interna según el sexo.

RI 94. Obtener migración interna según la provincia de procedencia.

RI 95. Obtener migración interna según la provincia destino

RI 96. Obtener saldo migratorio según la provincia.

RI 97. Obtener tasa de migración interna según la provincia.

RI 98. Obtener tasa de migración externa según la provincia.

RI 99. Obtener saldo migratorio según la provincia.

RI 100. Obtener tasa de saldo migratorio según la provincia.

2.5.2 Requisitos multidimensionales.

Los Requisitos multidimensionales se derivan de los requisitos de información, además contienen las variables de entrada y salida de los análisis, agrupadas según el tipo de información que se analiza.

Nacimientos

Variables de entrada.

- DPA.
- Temporal.
- Grupo de edades.
- Nivel de escolaridad.
- Tipo de embarazo.
- Sexo.
- Número de abortos anteriores de la madre.
- Cantidad de hijos de la madre.
- Lugar de ocurrencia.
- Peso del nacido.
- Total de embarazo.

Variable de salida.

- Cantidad de nacimientos.

Defunciones Infantiles

Variables de entrada.

- DPA.
- Grupo de edades.
- Causa de muerte.
- Temporal.
- Sexo.
- Lugar de ocurrencia.

Variables de salida.

- Cantidad de defunciones.

Defunciones

Variables de entrada.

- DPA.
- Grupo de edades.
- Causa de muerte.
- Ocupación
- Temporal.
- Sexo.
- Lugar de ocurrencia.

Variable salida.

- Cantidad de defunciones.

Matrimonios

Variables de entrada.

- DPA.
- Grupo de edades.
- Estado conyugal.
- Nivel de escolaridad.
- Ocupación.
- Temporal.
- Orden del matrimonio.

Variable de salida.

- Cantidad de matrimonios.

Divorcios

Variables de entrada.

- DPA.
- Temporal.
- Grupo de edades.
- Nivel de escolaridad.
- Ocupación.
- Duración del matrimonio.
- Cantidad de hijos del matrimonio.
- Sexo.

Variable de salida.

- Cantidad de Divorcios.

Migraciones Internas

Variable entrada.

- Sexo.
- DPA
- Grupo de edades.
- Ocupación.
- Temporal.
- Nivel de escolaridad.

Variable de salida.

- Cantidad de emigrantes.
- Cantidad de inmigrantes.

Población

Variables de entrada.

- DPA.
- Estado conyugal.

- Nivel de escolaridad.
- Temporal.
- Sexo.
- Ocupación.
- Grupo de edades.

Variables de salida.

- Cantidad de población.
- Tasa de crecimiento natural según la cantidad total de habitantes.
- Relación de masculinidad según la cantidad total de hombres.
- Densidad de la población por Km^2 según la cantidad total de habitantes.
- Relación de dependencia según edad
- Movimiento natural.
- Tasa de movimiento natural.
- Esperanza de vida.
- Saldo migratorio.
- Tasa de migración interna.
- Tasa de migración externa.
- Tasa de saldo migratorio.
- Tasa de Mortalidad.
- Tasa de Nacimientos.
- Tasa de Crecimiento Natural.

2.5.3 Requisitos funcionales.

Los requisitos funcionales están orientados a las necesidades de información de los usuarios, son las capacidades o condiciones que el sistema debe cumplir.

RF 1- El sistema debe permitir almacenar información histórica.

RF 2- El sistema debe permitir la obtención de la información sin necesidad de permisos.

RF 3- El sistema debe permitir la graficación de la información.

RF 4- El sistema debe garantizar la integración de los datos de las distintas fuentes.

RF 5- El sistema debe permitir cargar datos de DBF de nacimiento.

RF 6- El sistema debe permitir cargar datos de DBF de defunciones.

RF 7- El sistema debe permitir cargar datos de DBF de defunciones infantiles.

RF 8- El sistema debe permitir cargar datos de DBF de matrimonios.

RF 9- El sistema debe permitir cargar datos de DBF de divorcios.

RF 10- El sistema debe permitir cargar datos de DBF de migraciones internas.

RF 11- El sistema debe permitir cargar datos del Excel del censo de población.

RF 12- El sistema debe permitir transformar datos de DBF de nacimiento.

RF 13- El sistema debe permitir transformar datos de DBF de defunciones.

RF 14- El sistema debe permitir transformar datos de DBF de defunciones infantiles.

RF 15- El sistema debe permitir transformar datos de DBF de matrimonios.

RF 16- El sistema debe permitir transformar datos de DBF de divorcios.

RF 17- El sistema debe permitir transformar datos de DBF de migraciones internas.

RF 18- El sistema debe permitir transformar datos del Excel del censo de población.

RF 19- El sistema debe permitir extraer datos de DBF de defunciones.

RF 20- El sistema debe permitir extraer datos de DBF de defunciones infantiles.

RF 21- El sistema debe permitir extraer datos de DBF de matrimonios.

RF 22- El sistema debe permitir extraer datos de DBF de divorcios.

RF 23- El sistema debe permitir extraer datos de DBF de migraciones internas.

RF 24- El sistema debe permitir extraer datos de DBF de nacimiento.

RF 25- El sistema debe permitir extraer datos del Excel del censo de población.

2.5.4 Requisitos no funcionales.

Los Requisitos no funcionales son las propiedades o cualidades que el producto debe cumplir y hacen al producto atractivo, usable, rápido y confiable.

Requisitos de Usabilidad

RNF 1- El sistema debe de ser fácil de usar por los usuarios.

RNF 2- La información debe estar organizada por secciones según su tipo.

RNF 3- No deben existir más de tres niveles de navegación.

Fiabilidad

RNF 4- El repositorio de almacenamiento del proceso ETL debe de estar disponible 12 horas del día lo que representa el 50% horas uso.

RNF 5- El sistema de integración será accedido para su mantenimiento 1 vez por mes. En este plazo de mantenimiento se validará las estructuras de auditoría y se establecerán estrategias que permitan clasificar posibles errores y darles solución en caso de ser posible.

RNF 6- El sistema debe estar disponible 100% entre las 8:00 am y las 5:00 pm de lunes a viernes.

RNF 7- El tiempo medio de reparación en el proceso de integración depende en gran medida de la categoría del error o falla. El tiempo estimado es de 24 horas.

Eficiencia

RNF 8- El tiempo de respuesta debe ser en tiempo real (máximo 5 segundos).

RNF 9- El sistema debe permitir la concurrencia de varios usuarios sin que se afecte el tiempo de respuesta de las consultas.

RNF 10- El sistema debe permitir a varios usuarios acceder a la misma información al mismo tiempo.

Restricciones de diseño

RNF 11- El lenguaje para la programación del proceso de integración será SQL para realizar consultas a la base de datos y *JavaScript* para implementar algunas reglas de transformaciones.

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

RNF 12-Se dispondrá de una guía de ayuda sobre la navegación en la aplicación presentada.

Componentes Comprados

RNF 13- *Visual Paradigm* es la herramienta definida por la Dirección Técnica para el modelado conceptual de la información ya que la licencia de esta fue adquirida por la UCI.

Interfaz

RNF 14- Los reportes deben contar con una interfaz simple que facilite la interacción del usuario con la aplicación

RNF 15- Las interfaces de salida no serán cargadas con información innecesaria.

RNF 16-Los gráficos (un componente esencial en este tipo de solución) serán con los colores establecidos por la entidad ajustándose a los estándares establecidos de un buen diseño.

Interfaces Hardware

RNF 17- En el proceso de integración es necesaria la utilización de una memoria mínima de 1 GB para el proceso de transformación.

RNF 18- Se debe contar de un área de almacenamiento intermedio de 20 GB mínimos.

RNF 19- Para la visualización y la inteligencia de negocio se necesita una memoria de 1 GB.

RNF 20- Las estaciones de trabajo (PC clientes) deben contar con impresoras (para garantizar la impresión de las tablas de salida).

Interfaces Software

RNF 21- Se debe disponer de la instalación de la herramienta *Mondrian*.

RNF 22- Debe existir un navegador asociado al sistema operativo que se escoja para lograr que las interfaces web de las tablas de salida puedan visualizarse.

RNF 23- Es necesaria la instalación de *JDK**JRE* 1.5 para el uso de la herramienta *Mondrian*.

RNF 24- El lenguaje para la programación dentro del repositorio será *PostgreSQL* para realizar consultas a la base de datos e implementar las funciones necesarias.

RNF 25- El perfilado de datos se realizará con el *Talend Open Profiler*, herramienta libre especializada en estas funcionalidades.

Interfaces de Comunicación

RNF 26- La comunicación entre la base de datos de integración y el almacén de datos es a través del protocolo *TCP/IP*.

RNF 27- El sistema necesita estar conectado directamente a un dispositivo de red.

Requisitos de Licencia

RNF 28- La licencia de la herramienta a utilizar es adquirida por la UCI

El levantamiento de requisitos permite organizar la información según su tipo e identificar las operaciones que se deseen realizar sobre ella, posibilitando así la definición de los casos de uso.

Junto a las reglas del negocio se encuentran las necesidades del usuario, las cuales tributan información fundamental para el desarrollo del Mercado de Datos de la solución.

2.6 Necesidades de información.

Las necesidades de información definidas por el cliente son los análisis que estos deseen realizar sobre los datos, de ellas se derivan los requisitos de información.

2.7 Casos de uso del sistema.

Los casos de uso del sistema son las acciones que el sistema debe permitir, se clasifican en casos de uso de información y casos de uso funcionales. A través de estos los usuarios pueden consultar la información deseada.

2.7.1 Casos de uso de información.

Los casos de uso de información representan las consultas o análisis que los analistas pueden hacer con la información, teniendo en cuenta las diferentes aristas de análisis y resultados esperados. En un caso de uso se agrupan varios requisitos de información reuniendo sus variables de entrada y de salida.

En los casos de uso donde se consulta la información referente a nacimientos, defunciones, defunciones infantiles, matrimonios, divorcios, migraciones internas y población los analistas inician el proceso al solicitar la información de un tema determinado, el sistema muestra las opciones de reportes incluidos en dicho tema, el analista selecciona el que desea consultar y por último el sistema visualiza la información. Los casos de uso informativos identificados para la solución son los siguientes:

CUI 1. Consultar información de nacimientos.

CUI 2. Consultar información de defunciones.

CUI 3. Consultar información de defunciones infantiles.

CUI 4. Consultar información de matrimonios.

CUI 5. Consultar información de divorcios.

CUI 6. Consultar información de migraciones internas.

CUI 7. Consultar información de población.

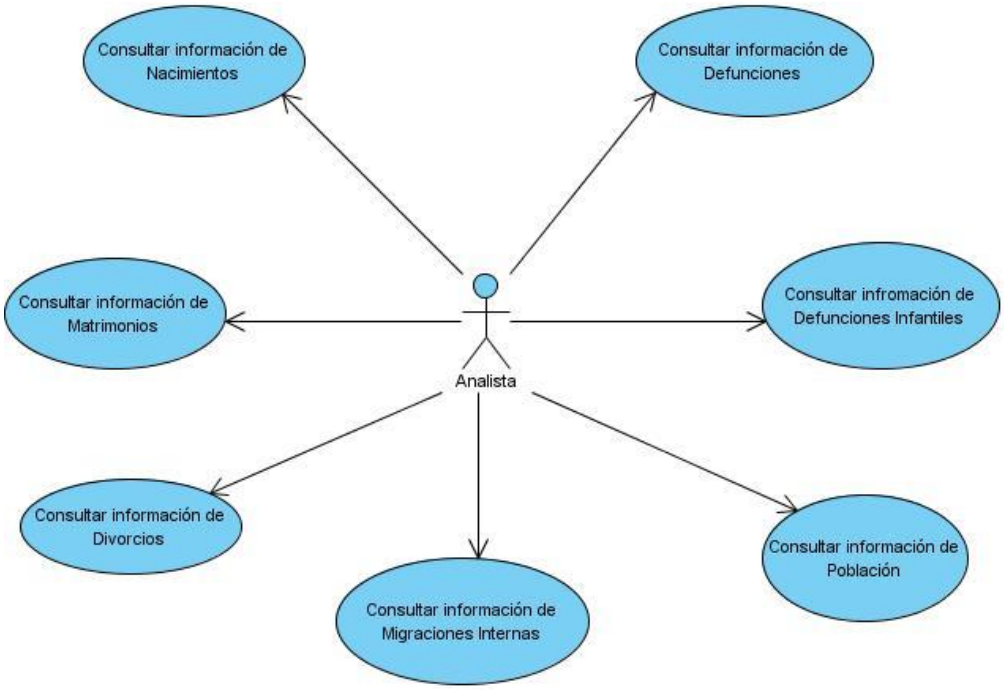

Figura 5. Diagrama de Casos de Uso de Información.

2.7.2 Casos de uso funcionales.

Los casos de uso funcionales son las acciones que el sistema debe realizar para satisfacer los requisitos funcionales. Los casos de uso de Extraer y de Transformar y cargar la información de nacimientos, defunciones, defunciones infantiles, matrimonios, divorcios, migraciones internas y población son iniciados por el administrador, el cual define el área de trabajo temporal. Luego realiza la extracción de la información, le hace los cambios y transformaciones pertinentes a los datos y posteriormente carga dichos datos en el almacén. Los casos de uso funcionales identificados son:

CUF 1. Extraer DBF de defunciones.

CUF 2. Extraer DBF de nacimientos.

CUF 3. Extraer DBF de defunciones infantiles.

CUF 4. Extraer DBF de matrimonios.

CUF 5. Extraer DBF de divorcios.

CUF 6. Extraer DBF de migraciones internas

CUF 7. Extraer Excel del censo de población.

CUF 8. Transformar y cargar DBF de nacimientos.

CUF 9. Transformar y cargar DBF de defunciones.

CUF 10. Transformar y cargar DBF de defunciones infantiles.

CUF 11. Transformar y cargar DBF de matrimonios.

CUF 12. Transformar y cargar DBF de divorcios.

CUF 13. Transformar y cargar DBF de migraciones internas.

CUF 14. Transformar y cargar Excel del censo de población.

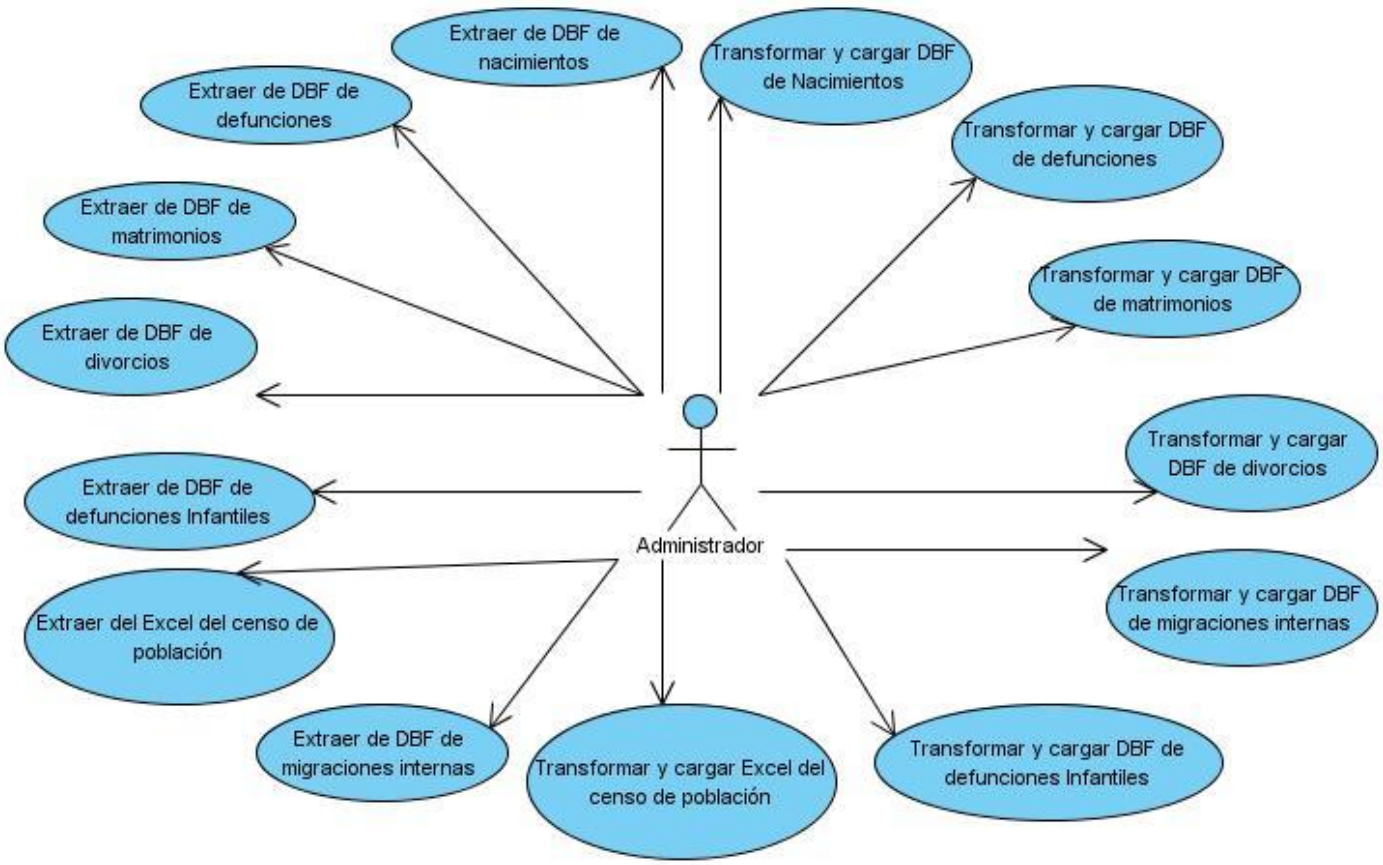

Figura 6. Diagrama de Casos de Uso Funcionales.

La información que se maneja en los casos de uso está relacionada con las tablas de hechos, las cuales engloban los valores del negocio descritos en las tablas de dimensiones.

2.8 Matriz BUS.

La matriz BUS es la relación existente de cada uno de los hechos con sus respectivas dimensiones en una matriz.

2.8.1 Tablas de Hechos.

- TH 1 – Población.
- TH 2 – Nacimientos.

- TH 3 – Defunciones.
- TH 4 – Defunciones infantiles.
- TH 5 – Matrimonios.
- TH 6 – Divorcios.
- TH 7 – Migraciones Internas.

2.8.2 Tablas de Dimensiones.

1. dim_abortos_anteriores_madre.
2. dim_cantidad_hijos.
3. dim_cantidad_hijos_matrimonio
4. dim_causa_de_muerte
5. dim_dpa
6. dim_duracion_matrimonio
7. dim_edad
8. dim_edad_infantil
9. dim_estado_conyugal
10. dim_lugar_ocurrencia
11. dim_nivel_escolaridad
12. dim_ocupacion
13. dim_orden_matrimonio
14. dim_peso_nacimiento
15. dim_semanas_gestación
16. dim_sexo
17. dim_temporal
18. dim_tipo_embarazo
19. dim_total_embarazos

	TH 1	***TH 2***	***TH 3***	***TH 4***	***TH 5***	***TH 6***	***TH 7***

1		X					
2		X					
3						X	
4			X	X			
5	X	X	X	X	X	X	X
6						X	
7	X	X	X		X	X	X
8				X			
9	X				X		
10		X	X	X			
11	X	X			X	X	X
12	X		X		X	X	X
13					X		
14		X					
15		X					
16	X	X	X	X			X
17	X	X	X	X	X	X	X
18		X					
19		X					

Tabla 5. Matriz BUS.

Las jerarquías, niveles y atributos de las tablas de hechos y dimensiones relacionadas anteriormente en la Matriz BUS, son descritas en el Modelo de Datos.

2.9 Modelo de Datos.

Al seleccionar las dimensiones candidatas para la solución, estas pasan a formar parte de las posibles dimensiones contenidas en el diseño. Entre las principales características de las dimensiones se encuentran la definición de las jerarquías, niveles y atributos.

2.9.1 Dimensiones identificadas.

1. dim_abortos_anteriores_madre.

   La dimensión describe los valores bajo los cuales puede clasificarse la información atendiendo al número de abortos anteriores de una madre.

   - numeroAbortos

2. dim_cantidad_hijos.

   La dimensión describe los valores bajo los cuales puede clasificarse la información atendiendo a la cantidad de hijos de una persona.

   - numeroHijos

3. dim_cantidad_hijos_matrimonio

   La dimensión describe los valores bajo los cuales puede clasificarse la información atendiendo a la cantidad de hijos de un matrimonio.

   - hijosMatrimonio

4. dim_causa_de_muerte

La dimensión describe los valores bajo los cuales puede clasificarse la información atendiendo a la causa de muerte de una persona.

- causasMuertes

5. dim_dpa

La dimensión describe los valores bajo los cuales puede clasificarse la información atendiendo al lugar de residencia de una persona o el lugar de ocurrencia de un hecho determinado según la división política administrativa.

- país -> provincia -> municipio ->municipioSuperficie -> zona

6. dim_duracion_matrimonio

La dimensión describe los valores bajo los cuales puede clasificarse la información atendiendo al tiempo de duración de un matrimonio.

- duracionMatrimonio

7. dim_edad

La dimensión describe los valores bajo los cuales puede clasificarse la información atendiendo a los grupos de edades de las personas.

- grupoEtario
- grupoQuinquenal
- grupoLaboral
- edad

8. dim_edad_infantil

La dimensión describe los valores bajo los cuales puede clasificarse la información atendiendo a la edad infantil de menores de 1 año (meses y días).

- grupoEdades
- meses

9. dim_estado_conyugal

La dimensión describe los valores bajo los cuales puede clasificarse la información atendiendo al estado conyugal de la persona.

- estadoConyugal
- estado_conyugal_codigo

10. dim_lugar_ocurrencia

La dimensión describe los valores bajo los cuales puede clasificarse la información atendiendo al sitio de ocurrencia de un hecho determinado (hospital, casa, etc.).

- lugarOcurrencia
- lugar_ocurrencia_codigo

11. dim_nivel_escolaridad

La dimensión describe los valores bajo los cuales puede clasificarse la información atendiendo al nivel de escolaridad alcanzado por una persona.

- nivelEscolaridad
- nivel_escolaridad_codigo

12. dim_ocupacion

La dimensión describe los valores bajo los cuales puede clasificarse la información atendiendo a la ocupación de una persona.

- gruposOcupacion
- estadoOcupacion

13. dim_orden_matrimonio

La dimensión describe los valores bajo los cuales puede clasificarse la información atendiendo al orden de matrimonio de una persona (primer, segundo, tercer, cuarto, etc.).

- ordenMatrimonio

14. dim_peso_nacimiento

La dimensión describe los valores bajo los cuales puede clasificarse la información atendiendo al peso del nacido.

- rangoPeso
- pesoInicio
- pesoFin

15. dim_semanas_gestación

La dimensión describe los valores bajo los cuales puede clasificarse la información atendiendo al número de semanas de gestación de la madre.

- semanas_gestacion_madre

16. dim_sexo

La dimensión describe los valores bajo los cuales puede clasificarse la información atendiendo al sexo de una persona.

- sexo
- sexo_codigo

17. dim_temporal

La dimensión describe los valores bajo los cuales puede clasificarse la información atendiendo al tiempo donde ocurre determinado hecho.

- anno -> mes

18. dim_tipo_embarazo

La dimensión describe los valores bajo los cuales puede clasificarse la información atendiendo al tipo de embarazo de la madre.

- tipoEmbarazo
- tipo_embarazo_codigo

19. dim_total_embarazos

La dimensión describe los valores bajo los cuales puede clasificarse la información atendiendo al total de embarazos que ha tenido la madre.

- numeroEmbarazos

2.9.2 Tablas de Hechos Identificadas.

En el subsistema Población se realizan los principales análisis de información y los reportes más generales de la disciplina demografía (tasas, porcientos, promedio, razón, etc.), estos análisis y reportes parten de la vinculación de los siguientes subsistemas: Divorcios, Nacimientos, Defunciones, Defunciones infantiles, Matrimonios y Migración Interna.

Tabla de Hechos Nacimientos *(Ver Anexo 1)*.

En esta tabla se encuentra el repositorio central de toda la información referente a los nacimientos.

Tabla de Hechos Defunciones *(Ver Anexo 2)*.

En esta tabla se encuentra el repositorio central de toda la información referente a las defunciones.

Tabla de Hechos Matrimonios *(Ver Anexo 3)*.

En esta tabla se encuentra el repositorio central de toda la información referente a los matrimonios.

Tabla de hechos Divorcios *(Ver Anexo 4)***.**

En esta tabla se encuentra el repositorio central de toda la información referente a los divorcios.

Tabla de Hechos Migración Interna *(Ver Anexo 5)***.**

En esta tabla se encuentra el repositorio central de toda la información referente a las migraciones internas.

Tabla de Hechos Defunciones Infantiles *(Ver Anexo 6)***.**

En esta tabla se encuentra el repositorio central de toda la información referente a las defunciones infantiles.

Tabla de Hechos Población *(Ver Anexo 7)***.**

En esta tabla se encuentra el repositorio central de toda la información referente a la población en general.

2.9.3 Medidas.

- Cantidad de nacimientos: Controla la cantidad de nacimientos en el País.
- Cantidad de defunciones: Controla la cantidad de defunciones en el País.
- Cantidad de defunciones infantiles: Controla la cantidad de defunciones infantiles en el País.
- Cantidad de matrimonios: Controla la cantidad de matrimonios en el País.
- Cantidad de divorcios: Controla la cantidad de divorcios en el País.
- Cantidad de inmigrantes: Controla la cantidad de inmigrantes en el País.
- Cantidad de emigrantes: Controla la cantidad de emigrantes en el País.
- Cantidad de población: Controla la cantidad de población en el País.

- Tasa de crecimiento natural según la cantidad total de habitantes: Controla cuando la tasa de una población aumenta (o disminuye) debido al incremento natural o a la emigración neta.
- Relación de masculinidad según la cantidad total de hombres: Controla el número de hombres por cada 100 habitantes.
- Densidad de la población por Km² según la cantidad total de habitantes: Controla la cantidad de habitantes por Km².
- Relación de dependencia según edad: Controla la razón de las personas que por su edad se definen como dependientes (menores de 15 años y mayores de 64) más las que se definen como económicamente productivas (15-64) dentro de una población.
- Movimiento natural: Controla el crecimiento de la población.
- Tasa de movimiento natural: Controla el crecimiento de la población por cada 1000 habitantes.
- Esperanza de vida: Controla la esperanza de vida de una población.
- Saldo migratorio: Controla la diferencia entre los emigrantes e inmigrantes de una población.
- Tasa de migración interna: Controla la cantidad de inmigrantes por cada 1000 habitantes.
- Tasa de migración externa: Controla la cantidad de emigrantes por cada 1000 habitantes.
- Tasa de saldo migratorio: Controla el saldo migratorio por cada 1000 habitantes.
- Tasa de Mortalidad: Controla la cantidad de defunciones por cada 1000 habitantes.
- Tasa de Nacimientos: Controla la cantidad de nacimientos por cada 1000 habitantes.
- Tasa de Crecimiento Natural: Controla el crecimiento natural por cada 1000 habitantes.

2.10 Esquema de seguridad.

A través de los roles se definen los niveles de acceso al sistema, respaldando así el esquema de seguridad del mismo. Para garantizar un alto grado de disponibilidad de la información y la seguridad en

un sistema, es necesario contar con dispositivos de seguridad y servidores de gestión y administración de bases de datos. El tiempo medio de reparación en el proceso de integración es de 24 horas. Además, el sistema será accedido para su mantenimiento una vez al mes.

2.11 Política de respaldo y recuperación.

La organización actualmente tiene definido que las salvas de toda la información que contiene la base de datos se realicen mensualmente, en un servidor local y en otro servidor fuera del establecimiento del servidor local con características iguales o similares. Las tablas de hechos involucradas en el proceso son: HECH_Nacimientos, HECH_Defunciones, HECH_Defunciones_Infantiles, HECH_Matrimonios, HECH_Divorcios, HECH_Migraciones Internas, HECH_Población.

Conclusiones.

En el capítulo 2 se definió demografía como tema de análisis, además de los roles del sistema que son el analista y el administrador, cada uno con sus respectivos permisos. Se detectaron 8 reglas del negocio. También se identificaron 25 requisitos funcionales, 28 no funcionales, 100 de información con sus requisitos multidimensionales. Además, se describieron los casos de usos del sistema, 7 casos de uso de información y 14 casos de uso funcionales. Se confeccionó la matriz BUS, el modelo de datos, se identificaron y describieron los 7 hechos y las 19 dimensiones.

CAPÍTULO 3. IMPLEMENTACIÓN Y PRUEBAS

Introducción.

En este capítulo se describe el modelo de datos físico, el cual muestra los esquemas, las tablas, restricciones, secuencias e índices de la estructura de datos. Además, se exponen los usuarios, roles y privilegios de las políticas de acceso a los objetos de la Base de Datos de demografía. Se establece la relación de los nomencladores del Mercado de Datos demográfico. También se crea una guía de implantación que contiene los requerimientos de software y los pasos para la instalación de la Base de Datos. Igualmente se realizan las siguientes pruebas: listas de chequeo de análisis, listas de chequeo de diseño, validación de los requerimientos por los clientes y casos de prueba de implantación.

3.1 Modelo de Datos Físico.

Los modelos de datos abarcan un grupo de operaciones fundamentales para la elaboración de consultas y reajuste de los datos. El almacenamiento de los datos se detalla a través del modelo físico.

En toda solución es importante conocer cómo está estructurada la base de datos para saber dónde encontrar la información recogida en la misma y la seguridad que brinda.

3.1.1 Estructuras de Datos.

Las estructuras de datos contienen elementos como son: Esquemas y Tablas, Restricciones y Secuencias e Índices. Los esquemas y las tablas están encargados de lograr una mejor organización en la base de datos, las restricciones y secuencias ofrecen una buena seguridad de los datos y los índices agilizan las operaciones que se realicen en la base de datos.

Esquemas y tablas

En el presente trabajo se presentan siete esquemas, en los cuales están organizadas las 26 tablas identificadas.

- Esquema ***defunción***.
- Esquema ***divorcio***.
- Esquema ***matrimonio***.

- Esquema ***migración***.
- Esquema ***nacimiento***.
- Esquema ***población***.
- Esquema ***dimensión***.

Tabla	***Esquema al que pertenece***	***Usuarios***	***Cantidad de Columnas***	***Descripción***
dim_abortos_anteriores_madre	dimensión	dba_demografia	2	Modela la dimensión dim_abortos_anteriores_madre.
dim_cantidad_hijos	dimensión	dba_demografia	2	Modela la dimensión dim_cantidad_hijos.
dim_cantidad_hijos_matrimonio	dimensión	dba_demografia	2	Modela la dimensión dim_cantidad_hijos_matrimonio.
dim_causa_de_muerte	dimensión	dba_demografia	2	Modela la dimensión dim_causa_de_muerte.
dim_dpa	dimensión	dba_demografia	5	Modela la dimensión dim_dpa.
dim_duracion_matrimonio	dimensión	dba_demografia	2	Modela la dimensión dim_duracion_matrimonio.
dim_edad	dimensión	dba_demografia	7	Modela la dimensión dim_edad.
dim_edad_infantil	dimensión	dba_demografia	3	Modela la dimensión dim_edad_infantil.
dim_estado_conyugal	dimensión	dba_demografia	2	Modela la dimensión dim_estado_conyugal.
dim_lugar_ocurrencia	dimensión	dba_demografia	2	Modela la dimensión dim_lugar_ocurrencia.

dim_nivel_escolaridad	dimensión	dba_demografia	2	Modela la dimensión dim_nivel_escolaridad.
dim_ocupacion	dimensión	dba_demografia	2	Modela la dimensión dim_ocupacion.
dim_orden_matrimonio	dimensión	dba_demografia	2	Modela la dimensión dim_orden_matrimonio.
dim_peso_nacimiento	dimensión	dba_demografia	4	Modela la dimensión dim_peso_nacimiento.
dim_semanas_gestacion	dimensión	dba_demografia	2	Modela la dimensión dim_semanas_gestacion.
dim_sexo	dimensión	dba_demografia	2	Modela la dimensión dim_sexo.
dim_temporal	dimensión	dba_demografia	5	Modela la dimensión dim_temporal.
dim_tipo_embarazo	dimensión	dba_demografia	2	Modela la dimensión dim_tipo_embarazo.
dim_total_embarazos	dimensión	dba_demografia	2	Modela la dimensión dim_total_embarazos.
dim_tipo_formalizacion	dimensión	dba_demografia	2	Modela al hecho hech_matrimonio.
hech_defuncion	defunción	dba_demografia	8	Modela al hecho hech_defuncion.
hech_defuncion_infantiles	defunción	dba_demografia	8	Modela al hecho hech_defuncion_infantiles.
hech_divorcio	divorcio	dba_demografia	15	Modela al hecho hech_divorcio.
hech_matrimonio	matrimonio	dba_demografia	18	Modela al hecho hech_matrimonio.
hech_migracion_interna	migración	dba_demografia	9	Modela al hecho hech_migracion_interna.

hech_nacimiento	nacimiento	dba_demografia	15	Modela al hecho hech_nacimiento.
hech_poblacion	población	dba_demografia	19	Modela al hecho hech_poblacion.

Tabla 6. Descripción de esquemas y tablas

Restricciones y Secuencias

Las restricciones son condiciones que obligan el cumplimiento de ciertas normas en una base de datos. Proporcionan un método para efectuar reglas, además limitan los datos que se pueden almacenar en las tablas y juegan el papel de rol a la hora de organizar mejor la información. Estas pueden ser establecidas por el administrador de la base de datos. Algunos ejemplos de restricciones son: la validación de un campo que sea únicamente de once dígitos y de tipo entero, las llaves primarias y foráneas de las tablas, entre otras.

Las secuencias son atributos que incrementan secuencialmente y paralelos al desarrollo del proceso. Ejemplo: las llaves primarias.

En el trabajo existen 19 llaves primarias y 73 foráneas, a continuación se muestran algunas de estas.

Llaves Primarias	dim_edad_id
	dim_nivel_escolaridad_id
	dim_cantidad_hijos_id
	dim_dpa_id
	dim_peso_nacimiento_id
	dim_edad_infantil_id
	dim_sexo_id
	dim_temporal_id

	dim_semanas_gestacion_id

Tabla 7. Llaves primarias

Llaves Foráneas	Refdim_causa_de_muerte204
	Refdim_dpa210
	Refdim_edad_205
	Refdim_sexo208
	Refdim_cantidad_hijos_matrimonio264
	Refdim_duracion_matrimonio213
	Refdim_temporal220
	Refdim_abortos_anteriores_madre234
	Refdim_ocupacion241

Tabla 8. Llaves Foráneas.

Índices

Los índices en una base de datos mejoran la velocidad de las operaciones que se realicen en ella, pues son estructuras de datos que permiten el acceso rápido a los registros de una tabla.

En el presente trabajo se identificaron 97 índices, algunos de estos son:

Índice	***Tabla***	***Esquema***	***Tipo de índice***	***Campos***
PK22	dim_abortos_anteriores_madre	dimensión	btree	dim_abortos_anteriores_madre_id
PK219	dim_cantidad_hijos	dimensión	btree	dim_cantidad_hijos_id

PK44	dim_dpa	dimensión	btree	dim_dpa_id
PK56	dim_duracion_matrimonio	dimensión	btree	dim_duracion_matrimonio_id
PK16	dim_edad	dimensión	btree	dim_edad_id
PK5	dim_sexo	dimensión	btree	dim_sexo_id
PK8	dim_temporal	dimensión	btree	dim_temporal_id
Ref10226	hech_nacimiento	nacimiento	btree	dim_semanas_gestacion_id
Ref26210	hech_defuncion	defunción	btree	dim_dpa_id
pk60_1	hech_defuncion_infantiles	defunción	btree	dim_causas_de_muertes_id, dim_edad_infantil_id, dim_temporal_id, dim_lugar_ocurrencia_id, dim_sexo_id, dim_dpa_id
Ref16262	hech_divorcio	divorcio	btree	dim_edad_conyugue1_id
Ref8220	hech_matrimonio	matrimonio	btree	dim_temporal_id
Ref16222	hech_migracion_interna	migración	btree	dim_edad_id
Ref5240	hech_poblacion	población	btree	dim_sexo_id

Tabla 9. Índices

Para el manejo de la base de datos es necesario determinar el nivel de acceso a los datos a través de los roles y permisos.

3.1.2 Roles y permisos.

Los roles se crean para garantizar la seguridad de la base de datos. De esta manera, se puede asegurar que un rol solo puede acceder a los datos si únicamente se encuentra autenticado en el sistema de base de datos.

Los roles y permisos identificados en el trabajo son los siguientes:

Roles	***Permisos***	***Descripción***
dba_demografia	Todos los permisos	Administra toda la base

		de datos.
analisis_demografia	Lectura (SELECT)	Analiza y visualiza los datos.
etl_demografia	Lectura y escritura (SELECT, INSERT, UPDATE, DELETE)	Analiza, visualiza y modifica los datos.

Tabla 10. Usuarios y permisos.

El llenado de la base de datos lo realiza el rol establecido, cargando los nomencladores en las tablas definidas.

3.1.3 Carga de nomencladores.

Los nomencladores son los atributos de las dimensiones que deben cargarse para que el Mercado de Datos esté listo y cargar los datos en las tablas de hecho.

En el trabajo se detectaron 18 nomencladores, algunos de los identificados son:

1. ***Nivel de escolaridad*** contiene el nivel de escolaridad de una persona determinada. Ejemplos: Primario, Técnico medio, Universitario, entre otros.
2. ***Grupo ocupacional*** contiene la ocupación de una persona determinada Ejemplos: Dirigente, Empleados de oficina, Obreros de máquina, entre otros.
3. ***Estado ocupacional*** contiene el cargo de la persona dentro de su ocupación. Ejemplos: Gerente o Director, Subdirector, entre otros.
4. ***Estado conyugal*** contiene el estado conyugal de una persona determinada. Ejemplos: Soltero(a), Casado(a), Divorciado(a), entre otros.
5. ***División política administrativa*** contiene la ubicación de una persona determinada o un lugar donde ocurre un hecho.
6. ***Sexo*** contiene el sexo de una persona determinada. Ejemplos: Femenino y Masculino.
7. ***Causa de muerte*** contiene la causa de muerte de una persona determinada. Ejemplos: Obesidad, Influenza y Neumonía, Tumores malignos, entre otros.

8. ***Lugar de ocurrencia*** contiene el sitio donde ocurrió determinado hecho. Ejemplos: Domicilio, Hospital, entre otros.
9. ***Tipo de embarazo*** contiene el tipo de embarazo de una mujer. Ejemplos: Normal, Riesgo, entre otros.
10. ***Edad*** contiene la edad de una persona. El rango oscila entre 1 y 120 años.

Para cargar exitosamente los nomencladores, montar la estructura, llenar las tablas de dicha estructura y definir la seguridad en la base de datos es necesario definir una guía de implantación, la cual refleja las normas a seguir.

3.2 Guía de Implantación.

La guía de implantación es un documento necesario en el trabajo, pues en ella se analizan todos los detalles de tipo técnico y organizativo.

3.2.1 Secuencia de Pasos.

1. Tiene que estar instalado *PostgreSQL 8.4* como gestor de base de datos.
2. Tiene que estar instalado el cliente del gestor de base de datos, en este caso *pgAdmin III PostgreSQL Tools v 1.10.0.*
3. Crear una nueva base de datos.
4. Crear los roles (dba_demografia, analisis_demografia, etl_demografia).
5. Cargar el script de lenguaje de datos (DDL).
6. Cargar el script lenguaje manipulación de datos (DML).
7. Cargar el script lenguaje control de datos en la base de datos creada (DCL).
8. Modificar los archivos de configuración de PostgreSQL para poder acceder a la base de datos desde otro puesto de trabajo.

Para verificar que el trabajo desarrollado cumple con las expectativas esperadas es necesario realizar validaciones y pruebas.

3.3 Validación y pruebas.

La validación y las pruebas que se le realizan a un producto, son fundamentales para garantizar que se cumpla con la exigencia del cliente y la eficiencia de la solución.

3.3.1 Listas de Chequeo de Análisis.

La lista de chequeo de análisis garantiza que los artefactos elaborados para el análisis cumplan con el formato determinado, que la información recogida esté organizada según se haya establecido y que sea la requerida para satisfacer el objetivo y alcance del artefacto.

- Lista de Chequeo Especificación de Requisitos.
- Lista de Chequeo Especificación de las áreas de la organización.
- Lista de Chequeo Herramienta para la recolección y análisis de la información.

3.3.2 Lista de Chequeo de Diseño.

Esta lista de chequeo garantiza que los artefactos elaborados para el diseño cumplan con la codificación establecida y que la información sea la requerida para satisfacer el objetivo y alcance del artefacto.

- Lista de Chequeo Modelo de Datos.

3.3.3 Validación de requisitos por el cliente.

Se realizó el encuentro con los clientes del presente trabajo, Diego Enrique González Galván, especialista de demografía del departamento Población en la ONE y Elena Leonila Fernández García, representante de la ONE en la UCI, luego de presentarle los principales requisitos de información, todos los posibles cruces de variables y las facilidades de cálculo de tasas, porcientos, relaciones y cantidades, los clientes estuvieron totalmente de acuerdo con la solución implementada.

3.3.4 Caso de prueba de implantación.

Escenarios del caso de prueba	***Pre condición***	***Resultado esperado***	***Pos condición***
Crear base de datos	*PostgreSQL 8.4*	Base de datos creada	Base de datos creada

	instalado como gestor de base de datos.	sobre el gestor de base de datos *PostgreSQL 8.4.*	
Roles y permisos	Base de datos creada en *pgAdmin III PostgreSQL Tools v 1.10.0* instalada como cliente de *PostgreSQL*.	Roles creados con sus respectivos permisos.	Base de datos creada con sus roles y permisos.
Cargar los Nomencladores	Base de datos creada con sus roles y permisos.	Campos de la base de datos llenos.	Base de datos lista para ser utilizada.

Tabla 11. Casos de prueba.

Conclusiones.

En este capítulo se describió el modelo físico del Mercado de Datos del Departamento de Población de la ONE, se detallaron los 7 esquemas existentes relacionados a las 27 tablas y los índices de la estructura de datos. Además, se especificaron dba_demografia, analisis_demografia y etl_demografia como los roles de la base de datos de demografía, cada uno de ellos con sus respectivos permisos. Se estableció la relación de 9 nomencladores especificando el tipo de información que manejan. También se creó la guía de implantación para la solución. Igualmente se realizaron las siguientes pruebas: El caso de prueba de implantación con resultados satisfactorios, Lista de Chequeo Especificación de Requisitos, de la cual se obtuvieron 27 conformidades y 1 no conformidad, Lista de Chequeo Especificación de las áreas de la organización, con 10 conformidades y 0 no conformidades y Lista de Chequeo Herramienta para la recolección y análisis de la información que presentó 25 conformidades y 1 no conformidad. En la validación de requisitos, el especialista del Departamento de Población de la ONE y la representante de dicha oficina en la UCI, ambos clientes del trabajo, estuvieron totalmente de acuerdo con la solución implementada.

Conclusiones del trabajo

Culminado el presente trabajo se concluye que:

- El estudio del arte garantizó la correcta selección de la tecnología de almacenamiento y de las herramientas utilizadas en la solución.
- El análisis de la solución permitió enfocar el desarrollo hacia las necesidades objetivas del cliente.
- El diseño lógico aseguró una correcta estructura dimensional.
- La solución obtenida garantizó un almacenamiento estructurado y funcional de la información, permitiendo la futura integración de los datos.
- Las pruebas realizadas validaron la solución en función de los requisitos y los artefactos generados, obteniéndose altos niveles de calidad y satisfacción del cliente.

Recomendaciones

- Mantener actualizado el Mercado de Datos ante los posibles cambios que ocurran para/con los datos almacenados.
- Integrar la información de los siete subsistemas a sus hechos correspondientes.
- Desarrollar la capa de visualización del Mercado de Datos.

Glosario de términos

- ***ONE***: Oficina Nacional de Estadísticas.
- ***INE***: Instituto Nacional de Estadísticas.
- ***IEA***: Instituto de Estadística de Andalucía.
- ***EIEL***: Encuestas de Infraestructuras y Equitaciones Locales.
- ***DW***: Almacenes de Datos.
- ***CSpro***: Sistema de Procesamiento de Censo y Encuesta.
- ***DATEC***: Centro de Tecnologías de datos.
- ***BI***: Inteligencia de Negocio.
- ***UCI***: Universidad de Ciencias Informáticas.
- ***SGBD***: Sistema Gestor de base datos.
- ***MVCC***: Versión Multi-Control de Concurrencia.
- ***ROLAP***: Procesamiento Analítico Relacional en Línea.
- ***MOLAP***: Procesamiento Analítico Multidimensional en Línea.
- ***HOLAP***: Procesamiento Analítico Híbrido en Línea.
- ***Matriz BUS***: Es la relación existente de cada uno de los hechos con sus respectivas dimensiones en una matriz.
- ***MINSAP***: Ministerio de Salud Pública
- ***DIE***: Departamento de Inmigración y Extranjería.
- ***RI***: Requisito de información.
- ***RF***: Requisito funcional.
- ***RNF***: Requisito no funcional.
- ***CUI***: Caso de uso informativo.
- ***CUF***: Caso de uso funcional.
- ***Índices***: Los índices en una base de datos mejoran la velocidad de las operaciones que se realicen en ella, pues son estructuras de datos que permiten el acceso rápido a los registros de una tabla.
- ***Secuencias***: Son atributos que incrementan secuencialmente y paralelo al desarrollo del proceso.