



FACULTAD 6

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



---

Mercado de Datos Morbilidad y Mortalidad para las instituciones de salud.

---



Autor(es): Juan Pablo Caballero Romero

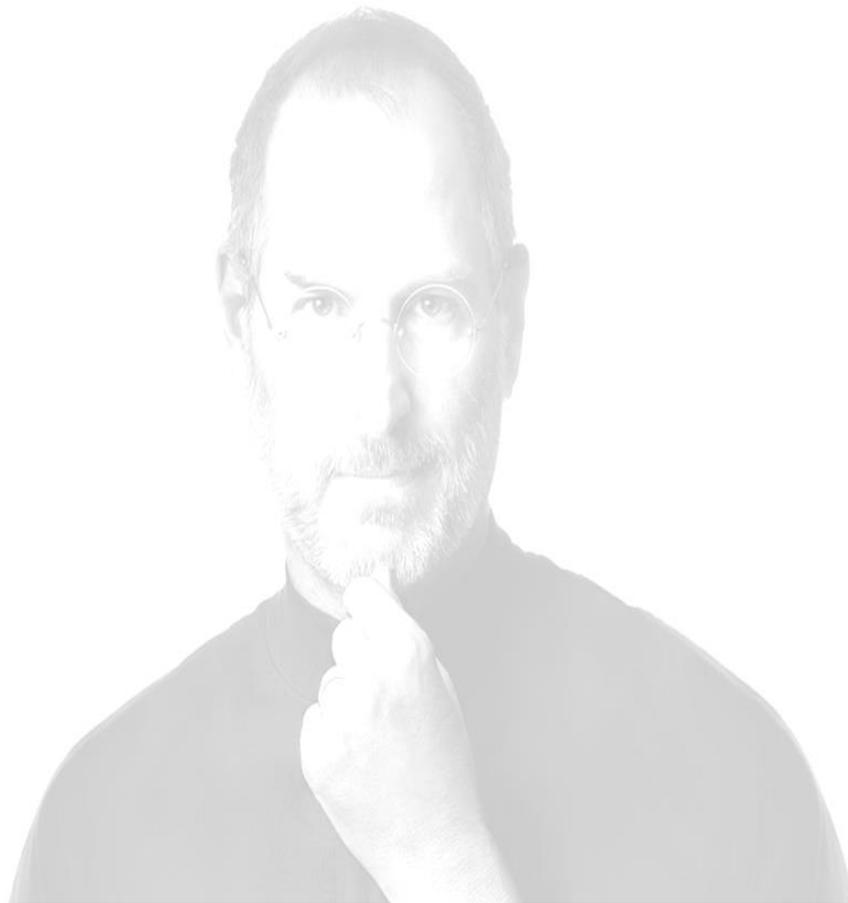
Eduardo Machado Soler

Tutor(es): Ing. Amarelys Alvarez Pérez

Ing. Asdrubal A. Nicot García

La Habana, Julio de 2016  
"Año 58 de la Revolución"

Pensamiento



*Aquellos que están locos como para pensar que pueden cambiar el mundo, son aquellos que lo hacen.*

*Steve Jobs*

## **Declaración de autoría**

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

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

**Eduardo Machado Soler**

**Juan Pablo Caballero Romero**

---

Firma del autor

---

Firma del autor

**Amarelys Alvarez Pérez**

**Asdrubal A. Nicot García**

---

Firma de la tutora

---

Firma del tutor

## **Datos de contacto**

**Tutora: Ing. Amarelys Alvarez Pérez**

Especialidad de graduación: Ingeniería en Ciencias Informáticas

Categoría docente: -

Categoría científica: -

Años de experiencia en el tema: 4

Años de graduado: 4

Correo Electrónico: [aaperez@uci.cu](mailto:aaperez@uci.cu)

**Tutor: Ing. Asdrubal A. Nicot García**

Especialidad de graduación: Ingeniería en Ciencias Informáticas

Categoría docente: Instructor

Categoría científica: -

Años de experiencia en el tema: 2

Años de graduado: 9

Correo Electrónico: [aangarcia@uci.cu](mailto:aangarcia@uci.cu)

## **Agradecimientos**

### **Juan Pablo**

Quisiera agradecerles a mis padres por haberme ayudado durante toda mi vida y convertirme en un profesional. A mi hermano por estar ahí siempre en los momentos difíciles. A mi familia en general por ayudarme de una forma u otra. Agradecerle también a mi compañero de tesis que a pesar que no nos conocíamos hemos hecho una buena amistad y logramos sacar adelante este trabajo. A Claudia por haberme hecho pasar dos años maravillosos a su lado y a pesar que no estuviéramos juntos me siguió apoyando.

### **Eduardo**

Me gustaría agradecerle a mi papá por haberme hecho el hombre que soy hoy y poder dedicarle este trabajo. A mi mamá por estar en todos los momentos de mi vida apoyándome y animándome a seguir por muy malos que fueran estos, por haber sido madre y padre a la vez sin pedir nada a cambio. A mis abuelos que los quiero con la vida por estar siempre atentos a mi carrera y brindarme todo el apoyo que necesitaba para continuar. A toda mi familia por estar siempre ahí en los momentos que más lo necesitaba. A Yudi por haberme soportado durante cuatro años y haberse convertido en una persona muy especial para mí. A Hectorín porque más que un amigo ha sido un hermano. A Juan Pablo por haber compartido conmigo unos de los años más importantes de su vida.

En general queremos agradecerles a nuestros tutores, en especial a Amarelys por haber sido con nosotros más que tutora una hermana mayor. A Henry, el Barra, la Flaki, Elian, el cuso, Migue, los Carlos, el lokiño, Ray, Daldis, Rolo, Eric, Bidot, Luis Alberto, el Lachy, Aris y muchos más que no mencionamos porque la lista es muy larga e igual de una forma u otra nos ayudaron a convertirnos en profesionales.

**Dedicatoria**

*A nuestros padres, familiares y amigos por brindarnos todo el apoyo necesario para cumplir esta meta.*

## **Resumen**

En la actualidad la información histórica que almacena el Sistema de Gestión Hospitalaria XAVIA HIS para el cálculo de los indicadores Morbilidad y Mortalidad, utiliza fórmulas complejas, por lo que su disponibilidad se ve afectada por el tiempo de ejecución de las consultas realizadas al sistema, lo que conlleva a la pérdida de datos de relevante importancia para los diferentes estudios de la salud. Con el propósito de facilitar a los especialistas hacer análisis históricos y de tendencias sobre los datos, la investigación tiene como objetivo general desarrollar un mercado de datos para contribuir a la toma de decisiones sobre los indicadores relacionados con la Morbilidad y Mortalidad en las instituciones de salud. Para su posterior implementación se documentan las herramientas, tecnologías y metodología empleadas en el desarrollo de la investigación. Se realiza el análisis, diseño e implementación de los subsistemas de almacenamiento, integración y visualización, haciendo uso de Visual Paradigm Community Edition 8.0, PostgreSQL 9.2 y las herramientas de la Suite de Pentaho. Se llevaron a cabo pruebas de integración, unitarias y de aceptación al mercado de datos para comprobar el correcto funcionamiento del mismo. Finalmente se obtuvo como resultado un mercado de datos poblado con toda la información asociada a los términos Morbilidad y Mortalidad.

**Palabras claves:** almacén de datos, mercado de datos, morbilidad, mortalidad

## **Summary**

At present the historical information stored Hospital Management System XAVIA HIS for calculating the Morbidity and Mortality indicators, uses complex formulas, so its availability is affected by the runtime of queries to the system, which leads to data loss of significant importance for different health studies. In order to make it easier for specialists to historical and trends data analysis, research general objective is to develop a market data to assist decision making on indicators related morbidity and mortality in health institutions . For further implementation tools, technologies and methodology used in the development of research are documented. Analysis, design and implementation of storage subsystems, integration and visualization is performed, using Visual Paradigm Community Edition 8.0, PostgreSQL 9.2 and tools Pentaho Suite. Integration, acceptance and unit testing to verify the correct operation took place. Finally obtained as a result a data mart populated with all the information associated with the terms Morbidity and Mortality.

**Keywords:** data warehouse, data mart, mortality, morbidity.

## Índice

Introducción: .....	1
Capítulo 1: Fundamentación teórica metodológica para el MD Morbilidad y Mortalidad .....	5
1.1 Definición de AD.....	5
1.2 Características de los AD .....	5
1.3 Ventajas y desventajas de los AD .....	6
1.4 Mercado de datos.....	6
1.4.1 Etapas de desarrollo de un MD .....	7
1.5 Modelado multidimensional de los AD .....	8
1.6 MD precedentes para la salud .....	10
1.7 Procesamiento analítico en línea.....	11
1.8 Metodologías para el desarrollo de los AD .....	12
1.9 Herramientas utilizadas para el desarrollo del MD.....	14
1.9.1 Herramienta de modelado.....	14
1.9.2 Sistema gestor de BD .....	14
1.9.3 Herramientas para los procesos de extracción, transformación y carga.....	14
1.9.4 Herramientas para el proceso de Inteligencia de negocios (BI).....	15
Conclusiones del capítulo.....	16
Capítulo 2: Análisis y diseño del MD Morbilidad y Mortalidad.....	17
2.1 Necesidades del negocio.....	17
2.2 Especificación de los requisitos .....	17
2.2.1 Requisitos de información .....	17
2.2.2 Requisitos funcionales .....	19
2.2.3 Requisitos no funcionales .....	19
2.3 Diagrama de casos de uso de información y casos de uso funcionales.....	21
2.3.1 Casos de uso de información.....	21

2.3.2 Casos de uso funcionales .....	21
2.3.3 Actores del MD Morbilidad y Mortalidad .....	21
2.3.4 Diagrama de casos de uso.....	22
2.3.5 Especificación del CU Presentar información de mortalidad poblacional.....	24
2.4 Reglas del negocio.....	25
2.5 Definición de la arquitectura base del MD .....	26
2.6 Diseño del MD.....	27
2.6.1 Subsistema de almacenamiento .....	27
2.6.2 Matriz dimensional o matriz bus del MD Morbilidad y Mortalidad .....	29
2.6.3 Modelo de datos .....	30
2.6.4 Política de respaldo y recuperación .....	32
2.7 Subsistema de integración .....	32
2.7.1 Diseño de las transformaciones .....	32
2.7.2 Estrategias generales de integración de datos.....	33
2.8 Diseño del subsistema de visualización.....	34
2.8.1 Arquitectura de la información.....	34
2.8.2 Diseño de los cubos OLAP .....	35
2.9 Esquema de seguridad.....	36
Conclusiones del capítulo.....	37
Capítulo 3: Implementación del MD Morbilidad y Mortalidad .....	38
3.1 Implementación del subsistema de almacenamiento.....	38
3.1.1 Estándares de codificación .....	39
3.2 Implementación del subsistema de integración .....	40
3.2.1 Implementación de las transformaciones .....	40
3.2.2 Implementación de los trabajos.....	42
3.3 Implementación del subsistema de visualización.....	42
3.3.1 Representación de la arquitectura de la información.....	43

3.4 Pruebas aplicadas al MD Morbilidad y Mortalidad .....	45
3.5 Herramientas de pruebas .....	47
3.6 Resultados de las pruebas .....	48
Conclusiones del capítulo.....	49
Conclusiones .....	50
Recomendaciones .....	51
Referencias Bibliográficas.....	52
Bibliografía.....	55
Glosario de Términos.....	58
Anexos.....	59

## Índice de figuras

Fig 1. Esquema estrella .....	9
Fig 2. Esquema copo de nieve .....	10
Fig 3. Esquema constelación de hechos .....	10
Fig 4. Diagrama de CU del MD Morbilidad y Mortalidad .....	23
Fig 5. Arquitectura del MD Morbilidad y Mortalidad .....	26
Fig 6. Modelo de datos del MD Morbilidad y Mortalidad .....	31
Fig 7. Tipos de datos de la fuente .....	32
Fig 8. Diseño de las transformaciones .....	33
Fig 9. Mapa de navegación .....	35
Fig 10. Diseño de los cubos del MD Morbilidad y Mortalidad .....	36
Fig 11. Transformación de la dimensión servicio .....	41
Fig 12. Transformación del hecho índice de masa corporal .....	41
Fig 13. Trabajo para la carga de las dimensiones .....	42
Fig 14. Trabajo para la carga de los hechos .....	42
Fig 15. Arquitectura de la información del MD Morbilidad y Mortalidad .....	43
Fig 16. VA Obesidad .....	44
Fig 17. Dashboard sobre el índice de masa corporal .....	44
Fig 18. Modelo V .....	46
Figura 19 Carta de aceptación .....	60

## Índice de Tablas

Tabla 1. Actores del sistema .....	22
Tabla 2. CU Presentar información de mortalidad poblacional .....	24
Tabla 3. Hechos y dimensiones del MD Morbilidad y Mortalidad.....	27
Tabla 4. Hechos y medidas.....	28
Tabla 5. Matriz Bus del MD Morbilidad y Mortalidad .....	29
Tabla 6. Esquema de seguridad subsistema de almacenamiento .....	36
Tabla 7. Esquema de seguridad subsistema de visualización.....	37
Tabla 8. Esquemas y tablas.....	39
Tabla 9. Estándares de codificación del MD Morbilidad y Mortalidad .....	39
Tabla 10. Caso de prueba cantidad de personas obesas.....	47
Tabla 11. NC por iteración .....	48
Tabla 12 NC pruebas de aceptación.....	49

## **Introducción:**

Con la informatización de la sociedad, ha aumentado a nivel mundial la cantidad de información generada diariamente en el sector empresarial. La necesidad de análisis, procesamiento y generación de informes sobre los datos almacenados, han traído como consecuencia la búsqueda de nuevos métodos para mejorar los procesos de negocio en las organizaciones, instituciones y entidades.

En la actualidad, las instituciones de salud plantean nuevas interrogantes para evitar, tratar o aliviar daños en la sociedad. La disponibilidad de información estadística confiable, oportuna, accesible, sencilla y de cada vez más amplio conocimiento, permite llevar a cabo actividades encaminadas a mejorar la salud de la población. Basado en una adecuada administración de los datos para proveer información clara a manos de decisores, hoy en día se hace uso de los sistemas que dan soporte al proceso de toma de decisiones. Dentro de esta clasificación se encuentran los Almacenes de Datos, una tecnología capaz de convertir datos crudos en información valiosa para su posterior uso o divulgación.

En Cuba, existen entidades encargadas de la producción de *software* con vista a la informatización del país, dentro de los que se destaca por el desarrollo de aplicaciones y servicios informáticos, a partir de la vinculación estudio-trabajo como modelo de formación, la Universidad de las Ciencias Informáticas (UCI). La UCI cuenta con varios centros de investigación y desarrollo, entre ellos, el Centro de Tecnologías de Gestión de Datos (DATEC) y el Centro de Informática Médica (CESIM). El primero se encarga de desarrollar productos y brindar servicios relacionados con el tratamiento y almacenamiento de datos y el otro se dedica al desarrollo de soluciones informáticas para el sector de la salud.

Uno de los productos que desarrolla el centro CESIM, es el Sistema de Información Hospitalaria XAVIA HIS. Orientado a satisfacer las necesidades de generación de información, para almacenar, procesar, reinterpretar datos médicos y administrativos está implantado en las instituciones de salud de la República Bolivariana de Venezuela.

Dentro de los procesos que se manejan en XAVIA HIS, se encuentran los relacionados con los términos médicos Morbilidad y Mortalidad. El primero informa la proporción de personas que sufren una enfermedad en un espacio y tiempo acotados, mientras que el segundo designa un número proporcional de muertes en una población y tiempo determinado.

El sistema de gestión hospitalaria presenta una Base de Datos (BD) relacional con más de 3000 tablas agrupadas en 40 esquemas, de los cuales se utilizan siete esquemas y 30 tablas para el análisis de los indicadores Morbilidad y Mortalidad. Puede ser accedido desde diferentes centros hospitalarios, o cada centro de salud contar con el *software* para simultáneamente poblar el servidor central, motivo por el cual se incrementa considerablemente el volumen de información. Este elemento ocasionó que actualmente el

módulo Estadísticas de XAVIA HIS presente problemas en el rendimiento para generar reportes. El cálculo de los indicadores utiliza fórmulas complejas, necesitando la unión de varias tablas para obtener los valores numéricos que permiten realizar las operaciones matemáticas.

Como consecuencia de lo antes expuesto, la disponibilidad de la información es afectada por la demora de ejecución de las consultas realizadas a la BD del sistema. Además, los datos se encuentran en un ambiente operacional y esto impide que se tenga la información correcta en el momento oportuno.

Aunque XAVIA HIS contempla información histórica, hoy en día los especialistas no tienen la posibilidad de obtener reportes que muestren los datos históricos almacenados sobre Morbilidad y la Mortalidad poblacional. De este modo se impide la realización de análisis estadísticos complejos entre diferentes indicadores, análisis históricos y de tendencias, lo que conlleva a la pérdida de datos útiles y valiosos para la toma de decisiones.

Por la situación anteriormente descrita se tiene el siguiente **problema de investigación**: ¿Cómo contribuir a la toma de decisiones sobre los indicadores relacionados con la Morbilidad y Mortalidad en las instituciones de salud?

Para establecer los límites del trabajo de diploma se define como **objeto de estudio**: los Almacenes de Datos, y como **campo de acción**: Mercado de Datos Morbilidad y Mortalidad para las instituciones de salud. Para darle solución al problema de la investigación se propone como **objetivo general** del presente trabajo: desarrollar un mercado de datos para contribuir a la toma de decisiones sobre los indicadores relacionados con la Morbilidad y Mortalidad en las instituciones de salud.

Con el propósito de cumplir el objetivo general, se tienen los siguientes **objetivos específicos**:

- ✓ Elaborar el marco teórico metodológico de la investigación, relacionado con el desarrollo de Almacenes de Datos (AD) y Mercado de Datos (MD).
- ✓ Realizar el análisis y diseño del MD Morbilidad y Mortalidad para las instituciones de salud.
- ✓ Implementar el MD Morbilidad y Mortalidad para las instituciones de salud.
- ✓ Probar el MD Morbilidad y Mortalidad para las instituciones de salud.

Después de los objetivos específicos mencionados anteriormente se tienen las siguientes **preguntas de la investigación**:

- ✓ ¿Cómo elaborar el marco teórico metodológico de la investigación relacionado con el desarrollo de AD y MD?
- ✓ ¿Cómo realizar el análisis y diseño del MD Morbilidad y Mortalidad para las instituciones de salud?
- ✓ ¿Cómo implementar el MD Morbilidad y Mortalidad para las instituciones de salud?
- ✓ ¿Cómo probar el MD Morbilidad y Mortalidad para las instituciones de salud?

A partir de las preguntas científicas se trazan como **tareas de la investigación**:

- ✓ Elaboración del estado del arte sobre el desarrollo de AD y MD en el sector de la salud para establecer los fundamentos teóricos metodológicos de la investigación.
- ✓ Caracterización de la metodología, herramientas y tecnologías a utilizar para guiar el proceso de desarrollo del MD.
- ✓ Levantamiento de requisitos del MD Morbilidad y Mortalidad para determinar las necesidades de información del cliente.
- ✓ Descripción de los casos de uso del MD para especificar cada una de las funcionalidades del sistema.
- ✓ Definición de los hechos, las medidas y las dimensiones del MD para determinar los elementos que conforman el modelo lógico de los datos.
- ✓ Diseño del modelo de datos para identificar los elementos que forman parte del modelo físico de los datos.
- ✓ Definición de la arquitectura del MD para identificar los subsistemas que componen la solución.
- ✓ Implementación del subsistema de almacenamiento para definir la estructura de la BD del MD Morbilidad y Mortalidad.
- ✓ Diseño del subsistema de integración para guiar la carga de las tablas, hechos y dimensiones.
- ✓ Diseño del subsistema de visualización para definir la capa de presentación de los datos.
- ✓ Diseño de los casos de pruebas para identificar los elementos que deben estar disponibles en el MD resultante.
- ✓ Implementación del subsistema de integración para poblar el MD.
- ✓ Implementación del subsistema de visualización para mostrar los reportes a los usuarios finales.
- ✓ Aplicación de pruebas al MD para garantizar el correcto funcionamiento del sistema.

### **Métodos Teóricos:**

**Histórico lógico:** se utilizó para profundizar en las distintas etapas del desarrollo de los AD, teniendo en cuenta las metodologías y herramientas utilizadas, además del análisis de soluciones similares. Con dicha información se pudo conformar el marco teórico de la investigación.

**Análisis-Síntesis:** se empleó con el objetivo de analizar la necesidad del trabajo a partir de la problemática existente. Permitted estudiar las características de las metodologías, herramientas y definiciones vinculadas para el desarrollo de AD. Se obtuvo el conocimiento y los elementos necesarios para el desarrollo de la solución según sus particularidades.

**Modelación:** con la utilización de este método se define la tabla de hechos del MD y sus dimensiones asociadas. Identificada la relación entre las tablas se estructura el modelo dimensional.

**Métodos empíricos:**

**Entrevista:** la entrevista se utiliza como una técnica para recopilar información. Con el objetivo de entender de qué trata el Sistema XAVIA HIS se realizó una entrevista a los especialistas del centro CESIM. Mediante este método se pudo alcanzar una mayor comprensión acerca de los términos médicos Morbilidad y Mortalidad y de la BD del sistema.

**Capítulo 1: Fundamentación teórica metodológica para el MD Morbilidad y Mortalidad.**

Se expone el marco conceptual de la solución; con la descripción de AD desarrollados para el área de la salud y la introducción de conceptos e ideas que se manejarán a lo largo de la investigación. Además, se caracterizan las herramientas y la metodología que serán utilizados en el desarrollo del MD Morbilidad y Mortalidad para las instituciones de salud.

**Capítulo 2: Análisis y diseño del MD Morbilidad y Mortalidad.**

Enmarca la descripción de la solución propuesta. En un primer momento se describe el diseño del diagrama de casos de uso del sistema, el modelo de datos y los subsistemas de integración y visualización. Se expondrá la manera en que se propone la construcción del MD, que precisa obtener un sistema bien estructurado, robusto y capaz de responder a las necesidades de los usuarios finales, en aras de alcanzar las metas y objetivos proyectados con el desarrollo de la investigación.

**Capítulo 3: Implementación y pruebas al MD Morbilidad y Mortalidad.**

En este capítulo se describen los aspectos relacionados con la construcción de la solución propuesta. Se implementan los subsistemas de almacenamiento, integración y visualización, incluyendo la creación de las estructuras para la navegación y el análisis de los datos. Luego se realizan pruebas de aceptación, integración y unitarias al MD Morbilidad y Mortalidad para constatar que cumple con las expectativas del cliente.

## Capítulo 1: Fundamentación teórica metodológica para el MD Morbilidad y Mortalidad.

Este capítulo define lo que es un AD desde diferentes puntos de vista según autores que han trabajado con el tema. Se plasman sus características principales, ventajas y desventajas, así como la definición de MD y sus etapas de desarrollo. También se describe el tipo de modelado, las metodologías y herramientas que se utilizan para el desarrollo de los AD. Se mencionan, además, MD precedentes para contribuir a la toma de decisiones en el área de salud.

### 1.1 Definición de AD

Una de las definiciones sobre AD, es la de William Harvey Inmon, quien define: un AD es una colección de datos orientada al negocio, integrada, variante en el tiempo y no volátil para el soporte del proceso de toma de decisiones de la gerencia. (Bernabeu, 2010)

Según Ralph Kimball, que es otro conocido autor en el tema de los AD, define un AD como: “una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis” (Kimball, y otros, 2002). Fue Kimball quien determinó que un AD no era más que: “la unión de todos los MD de una entidad”, defiende por tanto una metodología ascendente (*bottom-up*) a la hora de diseñar un AD. (Kimball, 2009)

A partir de lo planteado por los dos autores antes mencionados, se puede llegar a la conclusión que un AD reúne la información de múltiples áreas de una empresa para conformar diversos MD los cuales componen el AD. Para luego realizar un análisis de la misma con el objetivo de apoyar el proceso de toma de decisiones.

### 1.2 Características de los AD

Los AD se han convertido en una de las principales estrategias e inversiones de negocios para el apoyo a la toma de decisiones, sus principales características son:

- ✓ *Organizado en torno a temas.* La información se clasifica en base a los aspectos que son de interés para la empresa.
- ✓ *Integrado.* La integración de datos es el aspecto más importante, consiste en convenciones de nombres, codificaciones consistentes, medida uniforme de variables, etc.
- ✓ *Dependiente del tiempo.* Esta dependencia aparece de tres formas:
  - La información representa los datos sobre un horizonte largo de tiempo.
  - Cada estructura clave contiene (implícita o explícitamente) un elemento de tiempo (día, semana, mes, etc.).

- La información, una vez registrada correctamente, no puede ser actualizada.
- ✓ *No volátil*. El AD sólo permite cargar nuevos datos y acceder a los ya almacenados, pero no permite ni borrar ni modificar los datos. (Hernado, 2008)

### 1.3 Ventajas y desventajas de los AD

Los AD son muy utilizados hoy en día por las grandes empresas, dando la oportunidad de unir la información de distintas bases de datos y de esta forma combinarlo para proporcionar respuestas útiles que abarquen a toda la organización.

#### Ventajas

- ✓ Proporciona una herramienta para la toma de decisiones en cualquier área funcional, basándose en información integrada y global del negocio.
- ✓ Facilita la aplicación de técnicas estadísticas de análisis y modelización para encontrar relaciones ocultas entre los datos del almacén; obteniendo un valor añadido para el negocio de dicha información.
- ✓ Proporciona la capacidad de aprender de los datos del pasado y de predecir situaciones futuras en diversos escenarios.
- ✓ Simplifica dentro de la empresa la implantación de sistemas de gestión integral de la relación con el cliente.
- ✓ Supone una optimización tecnológica y económica en entornos de Centro de Información, estadística o de generación de informes con retornos de la inversión espectaculares. (Sinnexus, 2016)

#### Desventajas

- ✓ Infravaloración de los recursos necesarios para la captura, carga y almacenamiento de los datos.
- ✓ Infravaloración del esfuerzo necesario para su diseño (integración, consolidación) y creación.
- ✓ Incremento continuo de los requerimientos de los usuarios.
- ✓ Privacidad de los datos. (Celma, 2002)

### 1.4 Mercado de datos

Los MD, son un subconjunto de datos de un AD donde se almacenan la mayoría de las actividades de análisis que se llevarán a cabo, en el entorno de la inteligencia de negocio. (Group, 2015)

Inmon plantea que “Un MD es una solución que, compartiendo tecnología con el almacén de datos (pero con contenidos específicos, volumen de datos más limitado y un alcance histórico menor), permita dar soporte a una empresa pequeña, un departamento o área de negocio de una empresa grande”. (Inmon, 2005)

Se puede concluir que un MD es un AD histórico relativo a un departamento de una organización. La mayor diferencia entre los almacenes y los MD, es el ámbito de la información que contienen, debido a que en los MD es más pequeño y los datos se obtienen de un menor número de fuentes, por tanto, provoca que el tiempo de desarrollo sea menor. Son una alternativa de solución, al igual que los AD, pues el diseño y la construcción son similares.

#### **1.4.1 Etapas de desarrollo de un MD**

✓ Análisis y diseño

Toda fase inicial de un proyecto de gran alcance como son los almacenes de datos, debe contar con bases sólidas que le permitan avanzar hacia las otras fases de desarrollo con un mayor nivel de seguridad y organización. Para desarrollar un almacén de datos es necesario reconocer las necesidades analíticas que tiene una organización y establecer los objetivos a cumplir. Este es el primer paso a seguir, el análisis y la comprensión del nuevo entorno, el entorno analítico.

✓ Proceso de extracción, transformación y carga (ETL)

El proceso de integración de los datos tiene como objetivo fundamental unificar los datos pertenecientes a bases de datos fuente, archivos u otros sistemas de almacenamiento. Este proceso se lleva a cabo con el fin de organizar el flujo de los datos entre diferentes sistemas en una organización y aporta los métodos y herramientas necesarias para mover datos desde múltiples fuentes a un AD, reformatearlos, limpiarlos y cargarlos en otra BD, MD.

El ETL se divide a su vez en tres subprocesos.

- **Extracción:** en esta etapa se obtiene la información desde los sistemas de origen. Las fuentes normalmente se encuentran en diferentes formatos como son BD relacionales o ficheros planos, además pueden incluir otras estructuras diferentes. Cuando los datos son extraídos estos se convierten a un formato listo para comenzar la transformación.
- **Limpieza y transformación:** los datos provenientes de distintas fuentes pueden estar incoherentes, tener errores o estar incompletos. Para esto se realiza un proceso de limpieza que elimina todos estos errores e inconsistencias. Luego de que los datos se encuentran limpios y homogenizados se procede a transformar los datos para de esta forma estandarizar los códigos, corregir los datos, eliminar registros duplicados y usar conversiones y combinaciones para generar nuevos campos.
- **Carga:** en esta fase los datos son organizados y actualizados en la BD. En algunas bases de datos la información antigua se sobrescribe con nuevos datos. Los almacenes de datos mantienen un historial de los registros de manera que se pueda hacer una auditoría de los

mismos y disponer de un rastro de toda la historia de un valor a lo largo del tiempo. Si no se realiza un correcto proceso de ETL se pudieran obtener datos incorrectos lo que afectaría el proceso de toma de decisiones, es por eso que este proceso constituye aproximadamente un 70% del trabajo de la construcción de un AD. (Reyes, y otros, 2007)

✓ Inteligencia de negocio

La inteligencia de negocios es un conjunto de herramientas que permiten a través de consultas, reportes y el análisis de los datos, brindar a los usuarios la información de forma sintetizada con el objetivo de facilitar la toma de decisiones. Se utiliza además para comprender, mejorar el rendimiento y reducir los costes e identificar nuevas oportunidades de negocio. Las aplicaciones de la inteligencia de negocios incluyen las actividades de apoyo a las decisiones, consulta y presentación de informes, procesamiento analítico en línea (OLAP), el análisis estadístico, el pronóstico y minería de datos. Una definición más formal sería:

Son los procesos, tecnologías, y herramientas que se necesitan para convertir los datos en información, la información en conocimiento, y el conocimiento en planes que impulsan acciones rentables para el negocio. La inteligencia de negocios abarca el almacenamiento de datos, herramientas analíticas, contenido y gestión del conocimiento. (Ortiz, 2007)

### 1.5 Modelado multidimensional de los AD

A diferencia de los clásicos sistemas de BD que presentan sus estructuras diseñadas mediante el modelo Entidad-Relación (E-R) los AD se diseñan mediante un modelo multidimensional compuestos por dimensiones y hechos. Poseen la misma información que el de E-R, pero la organiza de forma diferente para garantizar la velocidad y eficiencia en la recuperación de la misma.

**Dimensiones:** las dimensiones definen como están los datos organizados lógicamente y proveen el medio para analizar el contexto del negocio. Representan los ejes del cubo y los aspectos de interés mediante los cuales el usuario podrá filtrar y manipular la información almacenada en los hechos. (Reategui, 2016).

**Hechos:** los hechos son datos instantáneos en el tiempo, que son filtrados, agrupados y explorados a través de condiciones definidas en las dimensiones. Los datos presentes en los hechos constituyen el volumen del AD, y pueden estar compuestos por millones de registros dependiendo de su granularidad y de los intervalos de tiempo de los mismos. Los más importantes son los de tipo numérico. El registro del hecho posee una clave primaria que está compuesta por las claves primarias de las dimensiones relacionadas a este. (Reategui, 2016)

**Medidas:** son los atributos numéricos asociados a los hechos (lo que realmente se mide). Por ejemplo: volumen de las ventas, coste asociado a un producto, número de transacciones efectuadas, porcentaje de beneficios. (Schupmann, 2002) y se pueden clasificar como:

- Aditivas, pueden ser combinadas a lo largo de cualquier dimensión. Por ejemplo “temperatura”, que puede estar dada por las dimensiones estación, región y fecha.
- Semi-aditivas, pueden no ser combinadas a lo largo de una o más dimensiones. Por ejemplo “nómina” que puede estar dada por las dimensiones empleados y tiempo, pero no producto.
- No aditivas, no pueden combinarse a lo largo de ninguna dimensión. Por ejemplo “cantidad de producto”, que únicamente puede estar dada por la dimensión producto. (Universidad de Catarina, 2016)

Para el modelado de los datos en un AD existen tres variantes de modelación, las cuales se detallan a continuación:

- **Esquema estrella:** es la técnica más común. Posee un solo hecho que se relaciona con todas las dimensiones. (Figura 1)

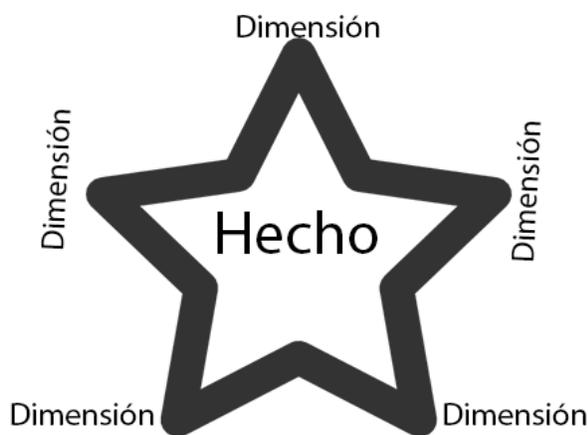


Fig 1. Esquema estrella

- **Esquema copo de nieve:** es una ampliación del modelo estrella, las dimensiones son normalizadas en dependencia de la jerarquía existente, pero la estructura de una única tabla de hechos se mantiene. (Figura 2)

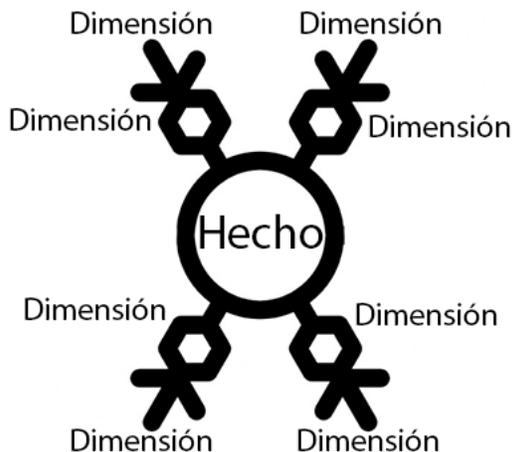


Fig 2. Esquema copo de nieve

- **Esquema constelación de hechos:** está compuesto por diversos esquemas de estrella, con la particularidad de que varias tablas de hechos comparten algunas tablas de dimensiones. (Figura 3)

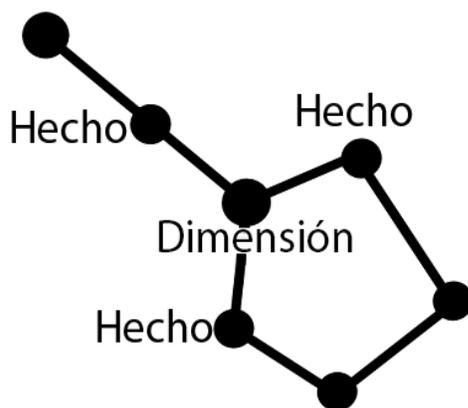


Fig 3. Esquema constelación de hechos

### 1.6 MD precedentes para la salud

Las distintas instituciones han recurrido a las ventajas de los MD para lograr brindar un mejor servicio a la población. Algunos ejemplos se evidencian a continuación:

- **Data Mart Biologics (BIA-MART):** ha sido creada específicamente para satisfacer la necesidad de la rápida expansión de la atención médica digital. BIA-MART es una herramienta de gestión de datos que de forma segura y eficiente recopila datos de los sistemas de atención de salud críticos normalmente aislados, ya sea en el hospital, en consultorios médicos o parte de una infraestructura regional. Los datos que se han encerrado en aplicaciones individuales, tales como admisión, descarga, transferencia, orden de entrada, farmacia, laboratorio, imágenes digitales, registros médicos, y gestión financiera se puede acceder a través

de una interfaz web habilitada fácilmente configurable, liberado para el análisis y presentación de informes. Una vez que los procesos de carga se hacen, sus datos viven en su propio MD, listo para su uso con cualquier informe o paquete gráfico (Bialogics, 2015).

- AD para los Ensayos Clínicos (EC) que se gestionan en el Centro de Inmunología Molecular: es un AD que se desarrolla para integrar y centralizar los EC que se gestionan en diferentes centros inmunológicos de Cuba. Permite la realización de análisis estadísticos complejos dentro de un mismo EC o entre diferentes EC para la toma de decisiones. (UCI, 2010)

Después del análisis de estos sistemas y para los procesos que fueron desarrollados, ninguno de ellos abarca los indicadores de morbilidad y mortalidad del modo que se necesita para trabajar en los centros hospitalarios que utilizan XAVIA HIS.

### 1.7 Procesamiento analítico en línea

Las bases de datos del procesamiento analítico en línea (OLAP) facilitan las consultas de la inteligencia empresarial. OLAP es una tecnología de base de datos que se ha perfeccionado para efectuar consultas e informes, en vez de procesar transacciones. El origen de datos de OLAP son las bases de datos de procesamiento transaccional en línea (OLTP, *Online Transactional Processing*) que se suelen almacenar en almacenes de datos. Los datos OLAP se derivan de estos datos históricos y se han agregado a estructuras que permiten realizar análisis complejos. Además, se organizan jerárquicamente y se almacenan en cubos en vez de en tablas. Se trata de una tecnología compleja que usa estructuras multidimensionales para proporcionar un acceso rápido a los datos con el fin de efectuar un análisis. Esta organización permite que un informe de tabla dinámica o gráfico dinámico muestre resúmenes de alto nivel, como totales de ventas en todo un país o región, así mismo los detalles de los sitios en los que las ventas son especialmente altas o bajas. (Microsoft, 2016).

Existen tres tipos de modelos OLAP:

**MOLAP:** la arquitectura MOLAP usa unas bases de datos multidimensionales para proporcionar el análisis, su principal premisa es que el OLAP está mejor implantado almacenando los datos multidimensionalmente. Un sistema MOLAP usa una base de datos propietaria multidimensional, en la que la información se almacena multidimensionalmente, para ser visualizada en varias dimensiones de análisis.

**ROLAP:** la arquitectura ROLAP, accede a los datos almacenados en un AD para proporcionar los análisis OLAP. La premisa de los sistemas ROLAP es que las capacidades OLAP se soportan mejor contra las bases de datos relacionales. El sistema ROLAP utiliza una arquitectura de tres niveles. La base de datos relacional maneja los requerimientos de almacenamiento de datos, y el motor ROLAP proporciona la funcionalidad analítica. El nivel de base de datos usa bases de datos relacionales para el manejo, acceso y

obtención del dato. El nivel de aplicación es el motor que ejecuta las consultas multidimensionales de los usuarios.

**HOLAP:** un desarrollo un poco más reciente ha sido la solución OLAP híbrida (HOLAP), la cual combina las arquitecturas ROLAP y MOLAP para brindar una solución con las mejores características de ambas: desempeño superior y gran escalabilidad. Un tipo de HOLAP mantiene los registros de detalle (los volúmenes más grandes) en la base de datos relacional, mientras que mantiene las agregaciones en un almacén MOLAP separado. (Sinnexus, 2016)

De los tres tipos de sistemas OLAP mencionados anteriormente se decidió utilizar el sistema ROLAP por las características anteriormente descritas. También se tuvo en cuenta que el MD se construye sobre un Sistema Gestor de Base de datos (SGBD) relacional.

### **1.8 Metodologías para el desarrollo de los AD**

Para el diseño y la construcción de cualquier AD, se deben seguir una serie de pasos, que pueden ser definidos como procesos que abarca una metodología confiable. En las tecnologías de AD se puede evidenciar dos grandes tendencias a nivel mundial, ambas sustentadas en las manos de sus principales creadores, la primera conocida como metodología de Inmon y la segunda, conocida como metodología Kimball, ambas con un sin número de seguidores.

Inmon es considerado el padre de los AD, pues fue quien primero introdujo el término, este, trata la construcción de los almacenes con un enfoque descendente (top-Down) donde los pequeños almacenes departamentales (MD) se nutrirán del AD, donde se encuentren los datos de forma consistente e histórica. Por su parte resulta tentadora la propuesta que ofrece Kimball quién es considerado un sabio en el tema de tecnologías de almacenamiento de datos, cuando de interesante manera separa la Inteligencia del Negocio entre el hecho y sus dimensiones, esto de manera general es muy eficaz y conduce a respuestas en muy pequeñas cantidades de tiempo. (PentahoCorporation., 2005)

Por otro lado se encuentra la metodología de HEFESTO cuya propuesta se fundamenta en una amplia investigación, comparación de metodologías existentes y la experiencia en la elaboración de AD. La ventaja principal de esta metodología es que especifica puntualmente los pasos a seguir en cada fase a diferencia de otras metodologías que mencionan los procesos, más no explican cómo realizarlos. (Dario, 2010)

DATEC propone la Metodología de Desarrollo para Proyectos de Almacenes de Datos que abarca todas las fases del desarrollo de un AD, esta toma como fundamento la metodología de Kimball, teniendo en cuenta que esta propone los conceptos de hechos y dimensiones, lo cual apoya la toma de decisiones y también en el proceso de desarrollo. (Hernández, 2013)

Plantea que la construcción del AD sea a medida que se construyan las BD departamentales, lo que da la impresión de organización en la empresa, entidad u organización. Consta de una abundante documentación, posibilitando la aclaración de las dudas que vayan surgiendo relacionadas a ese tema. Además cuenta con un amplio reconocimiento por los especialistas que se desempeñan en esta disciplina, pues constituye una metodología madura y que tiene bien definidas las etapas, actividades, artefactos y roles. Además esta metodología se ajusta a las condiciones y características de la producción en DATEC y la UCI. (Gonzalez Pompa, y otros, 2013)

Durante el ciclo de vida de la misma se destacan las siguientes fases de desarrollo:

- Estudio Preliminar o Planeación: se realiza el estudio de la entidad cliente, la planeación del proyecto, se definen los objetivos, el alcance preliminar, los costos estimados y otras actividades.
- Requerimientos: se realiza en dos direcciones, una, mediante la identificación de las necesidades de información y reglas del negocio; y la otra con un levantamiento detallado de las fuentes de datos a integrar. Después se procede a la definición de los requerimientos.
- Arquitectura: se definen las estructuras de almacenamiento, se diseñan las reglas de extracción, transformación y carga, definiéndose la arquitectura de información que regirá el desarrollo de la solución.
- Diseño e Implementación: se diseña físicamente el repositorio de datos, se crean las estructuras de almacenamiento, el área temporal de almacenamiento, se ejecutan las reglas de ETL y se configuran e implementan las herramientas de inteligencia de negocios para la obtención de los elementos que se acordaron con el cliente final.
- Prueba: se realizan las pruebas al sistema desde las Pruebas de Unidad hasta las de Aceptación con el cliente final.
- Despliegue: se realiza un despliegue piloto en el cual se configuran los servidores y se instalan las herramientas y se carga una muestra de los datos para demostrar que el sistema funciona. Posterior a la aceptación del cliente se realiza la carga de los datos así como la Capacitación y Transferencia Tecnológica.
- Soporte y Mantenimiento: tras la implantación de la solución se brindan los servicios de soporte en línea, vía telefónica, web u otras según el contrato firmado y las condiciones de soporte establecidas.

Gestión y Administración del Proyecto: constituye un flujo de trabajo que se ejecuta a lo largo de todo el ciclo de vida del proyecto. Está compuesto por un grupo de procesos que se encargan de mantener la adecuada gestión del proyecto a partir de la aplicación de conocimientos, habilidades, herramientas y técnicas. (Hernández, 2013)

Para el desarrollo del MD Morbilidad y Mortalidad se utilizará esta metodología y no transitará por todas las fases de su ciclo de vida, solo llegará hasta la fase de pruebas. Atendiendo a que las dos fases posteriores a esta son ejecutadas por el personal especializado de DATEC.

## **1.9 Herramientas utilizadas para el desarrollo del MD**

Para el desarrollo de un MD es necesaria la correcta selección de las herramientas a utilizar; basado en las empleadas por el equipo de desarrollo de DATEC para la creación de los AD, se seleccionaron las que a continuación se detallan, teniendo como premisa que son herramientas bajo licencia de software libre.

### **1.9.1 Herramienta de modelado**

**Visual Paradigm for UML Community Edition 8.0:** es una plataforma de modelado, diseñada para apoyar a los arquitectos de sistemas, desarrolladores y diseñadores UML (*Unified Modeling Language*) para acelerar el proceso de análisis y diseño de aplicaciones empresariales complejas. (Soft112, 2015)

### **1.9.2 Sistema gestor de BD**

**PostgreSQL 9.2:** es un potente sistema de BD, de código abierto objeto-relacional. Cuenta con más de 15 años de desarrollo activo y una arquitectura probada que se ha ganado una sólida reputación por su fiabilidad, integridad de datos y corrección. Se ejecuta en todos los sistemas operativos, incluyendo Linux, UNIX (AIX, BSD, HP-UX, SGI IRIX, Mac OS X, Solaris, Tru64), y Windows. Tiene soporte completo para claves foráneas, combinaciones, vistas, *triggers* y procedimientos almacenados (en varios idiomas). (Group, 2015)

**PgAdmin 1.16:** está diseñado para responder a las necesidades de todos los usuarios, desde escribir consultas SQL sencillas hasta el desarrollo de bases de datos complejas. La interfaz gráfica soporta todas las características de PostgreSQL y hace fácil administración. PgAdmin es desarrollado por una 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 PostgreSQL. (2015)

### **1.9.3 Herramientas para los procesos de extracción, transformación y carga**

**Pentaho Data Integration (PDI) v5.4:** es una herramienta flexible que le permite recoger datos de fuentes dispares, tales como bases de datos, archivos y aplicaciones, y convertir los datos en un formato unificado que sea accesible y relevante para los usuarios finales. PDI ofrece el motor de extracción, transformación y carga (ETL) que facilita el proceso de capturar los datos correctos, la limpieza de la información, y el almacenamiento de los datos utilizando un formato coherente y uniforme.

PDI ofrece apoyo a los que cambian lentamente dimensiones y llave sustituta para el almacenamiento de datos, permite la migración de datos entre bases de datos y aplicaciones, es lo suficientemente flexible como para cargar conjuntos de datos gigantes, y se puede sacar el máximo provecho de la nube, agrupado, y entornos de procesamiento masivamente paralelo. (2015)

**DataCleaner 3.1.2:** es un motor de procesamiento que fue construido para tareas altamente interactivas, de rendimiento y flexibilidad. Es una herramienta de análisis de calidad de la información que le permite realizar perfiles de datos y validación. El monitoreo es un aspecto central del DataCleaner para establecer el punto de partida, los objetivos, y para asegurar un proceso de seguimiento de las cuestiones de calidad de datos. Además encuentra las pautas, valores perdidos, juegos de caracteres y otras características de los valores de la información. (Data Cleaner Team, 2016)

#### **1.9.4 Herramientas para el proceso de Inteligencia de negocios (BI)**

**Pentaho Schema Workbench 3.10:** esta herramienta de la suite Pentaho tiene como objetivo facilitar la tarea de diseño de cubos OLAP. Su sencilla interfaz permite modelar un XML con el diseño del cubo a través de opciones lógicas e intuitivas que no requieren de un manejo avanzado de este formato de archivo. Dentro de sus características se destacan:

- Diseñador intuitivo de esquemas OLAP
- Permite crear, editar, actualizar y publicar esquemas OLAP para ser desplegados por aplicaciones de visualización Pentaho.
- Acelera de manera considerable la construcción e implementación de este tipo de soluciones. (PentahoCorporation., 2005)

**Pentaho BI Server 5.4:** la Plataforma Pentaho BI Server provee el soporte y la infraestructura necesarios para crear soluciones de BI a problemas de negocios. El marco proporciona los servicios básicos, incluidos autenticación, registro, auditoría, servicios web y motor de reglas. La plataforma también incluye un motor de solución que integra vistas de análisis, tableros de comandos y componentes de minería de datos. El diseño modular y arquitectura basada en *plug-in* permite a todos o parte de la plataforma estar embebida en aplicaciones de terceros por los usuarios finales, así como fabricantes de equipos originales. La Plataforma Pentaho BI Server funciona como un sistema basado en administración web de informes, el servidor de integración de aplicaciones y un motor de flujo de trabajo ligero (secuencias de acción.) Está diseñado para integrarse fácilmente en cualquier proceso de negocio.

Ventajas:

- Aplicación Java2EE 100% extensible, adaptable y configurable.
- Administra y programa vistas de análisis.

- Administra seguridad de usuarios.
- Se integra con la mayoría de entornos y se puede comunicar con otras aplicaciones vía *web services*. (PentahoCorporation., 2005)

**Pentaho Report Designer 5.4:** es una herramienta capaz de transformar los datos en información significativa. Se puede crear HTML, Excel, PDF, texto o informes impresos. Utilizando el motor de informes le da una flexibilidad sin igual para crear informes que se adaptan a los datos, ya que casi todas las propiedades se pueden calcular durante la generación del informe. Los informes pueden incluir datos de prácticamente cualquier fuente de datos debido a la gran selección de fuentes de datos, incluyendo bases de datos SQL, fuentes de datos OLAP e incluso la herramienta ETL Pentaho Data-Integración. (PentahoCorporation., 2005)

### **Conclusiones del capítulo**

En este capítulo se analizaron los principales conceptos relacionados con los AD y MD. Se logró caracterizar la metodología, tecnologías y herramientas utilizadas para el modelado e implementación de este tipo de solución. A partir del estudio de algunos sistemas implementados relacionados con el tema, se llega a la conclusión que ninguno presenta las características necesarias para ser usado por XAVIA HIS. Para el desarrollo del MD Morbilidad y Mortalidad la definición de Ralph Kimball es el paradigma que se toma en cuenta por las características de la presente investigación. Se decidió utilizar la Metodología de Desarrollo para Proyectos de Almacenes de Datos ya que abarca todas las etapas por las que transita un AD, además de que la misma tiene como base la metodología propuesta por Kimball. Se seleccionó el método de modelado ROLAP por las ventajas que ofrece a la hora de realizar el modelado y la posterior implementación de los procesos ETL. Para el modelado se decidió utilizar la herramienta Visual Paradigm for UML Community Edition 8.0. PostgreSQL en su versión 9.2 es el gestor de BD que se selecciona, con la herramienta de administración de BD PgAdmin 1.16. Para los procesos de ETL se seleccionó la herramienta Pentaho Data Integration (PDI) v5.4 y para el perfilado de los datos DataCleaner versión 3.1.2. La implementación del subsistema de visualización contará con la herramienta Pentaho Schema Workbench 3.10 para la modelación de los cubos OLAP, Pentaho BI Server 5.4 para la presentación de la información y el Report Designer 5.4 para la creación de reportes operacionales. Todas las herramientas a utilizar describen un entorno de trabajo de código abierto.

## Capítulo 2: Análisis y diseño del MD Morbilidad y Mortalidad.

Este capítulo abordará aspectos relacionados con los requerimientos del sistema, en él se describe el diseño del diagrama de casos de uso del sistema, el modelo de datos y los subsistemas de integración y visualización. De manera general, aborda el resultado del análisis y el diseño del MD Morbilidad y Mortalidad para las instituciones de salud. Se expondrá la manera en que se propone la construcción del MD, que precisa obtener un sistema bien estructurado, robusto y capaz de responder a las necesidades de los usuarios finales, en aras de alcanzar las metas y objetivos proyectados con el desarrollo de la investigación.

### 2.1 Necesidades del negocio

Para la presentación de la información relacionada con la Morbilidad y Mortalidad poblacional en el Sistema de gestión hospitalaria XAVIA HIS, se utilizan fórmulas complejas, necesitando la unión de varias tablas para obtener los valores numéricos que permiten realizar las operaciones matemáticas. Como consecuencia de lo antes expuesto, la disponibilidad de la información es afectada por la demora de ejecución de las consultas realizadas a la BD del sistema. Además, los datos se encuentran en un ambiente operacional y esto impide que se tenga la información correcta en el momento oportuno. Aunque el sistema presenta información histórica almacenada necesaria para la toma de decisiones, actualmente no se tiene la posibilidad de obtener reportes basados en el análisis de datos históricos sobre la morbilidad y la mortalidad poblacional. La propuesta de la solución está enfocada a resolver este problema a partir de analizar con exactitud las necesidades de los usuarios. Se realizó una entrevista a varios especialistas del área y se examinó la fuente de datos con el fin de analizar toda la información correspondiente a la misma, donde se definieron los objetivos para el desarrollo de la solución (ver Anexo 1).

### 2.2 Especificación de los requisitos

Los requisitos para un sistema son descripciones de lo que el sistema debe hacer: el servicio que ofrece y las restricciones en su operación. Tales requisitos reflejan las necesidades de los clientes por un sistema que atienda cierto propósito, como sería controlar un dispositivo, colocar un pedido o buscar información. (Sommerville, 2011)

#### 2.2.1 Requisitos de información

Los requisitos del sistema son versiones extendidas de los requisitos del usuario que los ingenieros de *software* usan como punto de partida para el diseño del sistema. Añaden detalles y explican cómo el sistema debe brindar los requisitos del usuario. Se pueden usar como parte del contrato para la implementación del sistema y, por lo tanto, deben ser una especificación completa y detallada de todo el sistema (Sommerville, 2011). A continuación, se muestran los requisitos de información identificados durante el proceso de análisis

según los temas de análisis definidos. Para la determinación de los requisitos del MD Morbilidad y Mortalidad se realizó una consulta a distintas fuentes de datos tales como: Indicadores básicos para el análisis del estado de salud de la población del Ministerio de Salud Pública de la República de Cuba, así como el libro Indicadores de gestión para establecimientos de atención médica. Para una mayor comprensión de estos requerimientos referirse al artefacto Requisitos de Software.

- RI1:** Obtener el total de defunciones por tiempo, dpa, sexo, edad, raza, causa y servicio.
- RI2:** Obtener la media anual de defunciones por tiempo y un dpa dado.
- RI3:** Obtener la tasa bruta de mortalidad por causa, raza, tiempo, dpa dado el tiempo y un dpa.
- RI4:** Obtener la tasa neta de mortalidad por causa, raza, tiempo, dpa dado el tiempo y un dpa.
- RI5:** Obtener el porcentaje de muertes antes de 24h del total de defunciones por tiempo, dpa, sexo, edad, raza, causa y servicio.
- RI6:** Obtener el porcentaje de muertes antes de 48h del total de defunciones por tiempo, dpa, sexo, edad, raza, causa y servicio.
- RI7:** Obtener la tasa de mortalidad específica por causa, raza, tiempo, dpa dado el tiempo y un dpa.
- RI8:** Obtener la mortalidad infantil por tiempo, dpa, sexo, raza, causa y servicio dado el tiempo y un dpa.
- RI9:** Obtener la tasa de mortalidad neonatal por tiempo, dpa, sexo, raza, causa y servicio dado el tiempo y un dpa.
- RI10:** Obtener la tasa de mortalidad neonatal precoz por tiempo, dpa, sexo, edad por tiempo, dpa, sexo, raza, causa y servicio dado el tiempo y un dpa.
- RI11:** Obtener la tasa de mortalidad pos neonatal por tiempo, dpa, sexo, raza, causa y servicio dado el tiempo y un dpa.
- RI12:** Obtener la tasa de mortalidad perinatal por tiempo, dpa, sexo, raza, causa y servicio dado el tiempo y un dpa.
- RI13:** Obtener la tasa de mortalidad de la niñez por tiempo, dpa, sexo, raza, causa y servicio dado el tiempo y un dpa.
- RI14:** Obtener la tasa de letalidad por causa, raza, tiempo, dpa dado el tiempo y un dpa.
- RI15:** Obtener la tasa de morbilidad general por tiempo, dpa, sexo, raza y causa dado el tiempo y un dpa.
- RI16:** Obtener la morbilidad proporcional por tiempo, dpa, sexo, raza y causa dado el tiempo y un dpa.
- RI17:** Obtener la tasa de incidencia por tiempo, dpa, sexo, raza y causa dado el tiempo y un dpa.
- RI18:** Obtener la tasa de prevalencia por tiempo, dpa, sexo, raza y causa dado el tiempo y un dpa.
- RI19:** Obtener el índice de bajo peso al nacer por tiempo, dpa, sexo y raza dado el dpa.
- RI20:** Obtener cantidad de personas obesas por tiempo, dpa, sexo y raza dado el dpa.
- RI21:** Obtener cantidad de personas con sobrepeso por tiempo, dpa, sexo y raza dado el dpa.
- RI22:** Obtener el total de enfermos por tiempo, dpa, sexo, raza y causa dado el dpa.

### **2.2.2 Requisitos funcionales**

Los requisitos funcionales para un sistema refieren lo que el sistema debe hacer. Tales requisitos dependen del tipo de software que se esté desarrollando, de los usuarios esperados del software y del enfoque general que adopta la organización cuando se escriben los requerimientos. Al expresarse como requerimientos del usuario, los requisitos funcionales se describen por lo general de forma abstracta que entiendan los usuarios del sistema. (Sommerville, 2011) Se mencionan a continuación los requisitos funcionales (RF) identificados:

**RF1:** Autenticar usuario.

**RF2:** Adicionar rol.

**RF3:** Eliminar rol.

**RF4:** Modificar rol.

**RF5:** Adicionar usuario.

**RF6:** Eliminar usuario.

**RF7:** Modificar usuario.

**RF8:** Adicionar reporte.

**RF9:** Eliminar reporte.

**RF10:** Modificar reporte.

**RF11:** Extraer datos de la fuente.

**RF12:** Realizar transformación de los datos.

**RF13:** Cargar los datos al MD Morbilidad y Mortalidad.

**RF14:** Realizar operaciones sobre el reporte.

### **2.2.3 Requisitos no funcionales**

Los requerimientos no funcionales, como indica su nombre, son requerimientos que no se relacionan directamente con los servicios específicos que el sistema entrega a sus usuarios. Pueden relacionarse con propiedades emergentes del sistema, como fiabilidad, tiempo de respuesta y uso de almacenamiento. De forma alternativa, pueden definir restricciones sobre la implementación del sistema, como las capacidades de los dispositivos o las representaciones de datos usados en las interfaces con otros sistemas. Los requerimientos no funcionales, como el rendimiento, la seguridad o la disponibilidad, especifican o restringen por lo general características del sistema como un todo. Los requerimientos no funcionales a menudo son más significativos que los requerimientos funcionales individuales. (Sommerville, 2011) Para una mayor comprensión de los mismos y conocimiento de todos los requisitos no funcionales (RNF) remitirse al artefacto: Especificacion\_de\_requisitos\_de\_software.doc del expediente de proyecto.

## Usabilidad

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

El sistema debe tener una interfaz gráfica uniforme que incluya pantallas, menús y opciones. Las pautas de diseño se realizarán siguiendo la arquitectura de información definida.

## Confiabilidad

**RNF 6.** Asegurar la disponibilidad del sistema.

El sistema debe estar disponible durante el horario de trabajo. En caso de fallo, la recuperación del servicio deberá ser de un período de tiempo de una hora o menos.

## Eficiencia

**RNF 9.** Garantizar la conexión de más de 150 usuarios concurrentemente.

El sistema no debe colapsar ante la presencia de al menos 150 conexiones concurrentes. El servicio que brinda el MD no debería verse afectado ante una gran carga de peticiones.

## Soporte

**RNF 11.** Lograr la homogeneidad de la estructura de los elementos definidos en el mercado.

Las estructuras del MD deben tener un nombre estándar teniendo en cuenta el tipo de estructura que sea. En el documento Estándares de Codificación se definen convenciones de nombrado con el objetivo de manejar un vocabulario común en todo el MD, permitiendo un entendimiento claro y conciso por parte de los desarrolladores.

**RNF 12.** Disponibilidad de tutoriales de uso de las herramientas utilizadas en la confección de las tablas de salida.

Garantizar la disponibilidad de tutoriales de uso de las herramientas utilizadas en la confección de los reportes para lograr un dominio más específico de cada herramienta.

## Restricciones de diseño

**RNF 14.** Utilizar los lenguajes de programación definidos durante la investigación.

Como lenguaje dentro del sistema gestor de BD para la programación en el MD se utilizará PL/pgSQL. En la implementación de los procesos de integración de datos se utilizará el lenguaje JavaScript. También se hará uso del lenguaje MDX para realizar las consultas.

## Interfaz

**RNF 15.** Acceso al sistema.

El usuario deberá acceder a la aplicación mediante el protocolo HTTP, usando preferiblemente el navegador web Firefox 40.0.

## 2.3 Diagrama de casos de uso de información y casos de uso funcionales.

### 2.3.1 Casos de uso de información

Los casos de uso informativos describen los requisitos de información agrupados según el tema de análisis. A continuación se describe cada caso de uso de información (CUI) definido para el MD Morbilidad y Mortalidad.

**CUI1:** Presentar información de morbilidad poblacional: muestra los reportes de los indicadores relacionados con la morbilidad poblacional.

**CUI2:** Presentar información de mortalidad poblacional: muestra los reportes de los indicadores relacionados con la mortalidad poblacional.

### 2.3.2 Casos de uso funcionales

Los casos de usos funcionales (CUF) se basan en la gestión de los roles, reportes, usuarios y la autenticación de los usuarios, además de las consultas en la BD (Sommerville, 2011). A continuación se describen los casos de uso del MD Morbilidad y Mortalidad.

**CUF1.** Autenticar usuario: se realiza la autenticación de los usuarios en el sistema.

**CUF2.** Gestionar rol: permite adicionar, modificar, eliminar y mostrar roles.

**CUF3.** Gestionar usuario: permite adicionar, modificar, eliminar y mostrar usuarios.

**CUF4.** Gestionar reportes: permite adicionar, modificar, eliminar y mostrar reportes.

**CUF5.** Extraer los datos de la fuente: se realiza la extracción de los datos de las fuentes.

**CUF6.** Realizar la transformación y carga de los datos: se realiza la transformación y carga de los datos necesarios para la construcción del MD Morbilidad y Mortalidad.

### 2.3.3 Actores del MD Morbilidad y Mortalidad

Un actor es una agrupación uniforme de personas, sistemas o máquinas que interactúan con el sistema. Los actores se representan con dibujos simplificados de personas, llamados en inglés “*stick man*” (hombres de palo). (Ceria, 2016)

Tabla 1. Actores del sistema

Actor	Responsabilidad
<b>Administrador</b>	Inserta y elimina los usuarios en el sistema. Inserta roles que van a ser asignados a los usuarios y elimina los innecesarios en el sistema. Inserta y modifica vistas de análisis que estarán disponibles en el sistema y elimina aquellos innecesarios. Consulta información de las vistas de análisis, los usuarios que han accedido al sistema y la información consultada por estos.
<b>Administrador ETL</b>	Realiza la extracción de los datos de los ficheros fuentes, así como la transformación y la carga de los datos hacia el MD Morbilidad y Mortalidad.
<b>Especialista</b>	Consulta la información de las vistas de análisis disponibles en el sistema y realiza operaciones sobre los reportes generados.

### 2.3.4 Diagrama de casos de uso

Los diagramas de caso de uso brindan un panorama bastante sencillo de una interacción. El modelado de casos de uso se utiliza ampliamente para apoyar la adquisición de requerimientos. Un caso de uso puede tomarse como un simple escenario que describa lo que espera el usuario de un sistema. Cada caso de uso representa una tarea discreta que implica interacción externa con un sistema. (Sommerville, 2011)

Para un mayor entendimiento referirse al artefacto Especificación de Casos de Uso (CU).

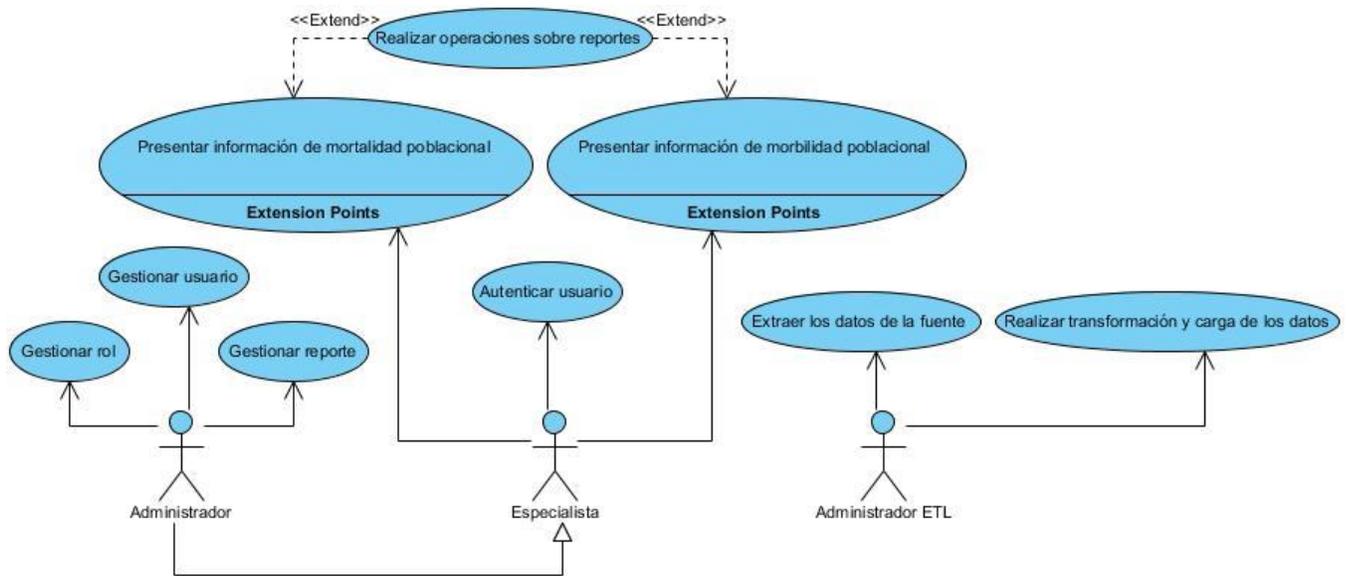


Fig 4. Diagrama de CU del MD Morbilidad y Mortalidad

### Patrones de CU utilizados

Los patrones de CU son comportamientos que deben existir en el sistema, ayudan a describir qué es lo que el sistema debe hacer, es decir, describen el uso del sistema y cómo este interactúa con los usuarios. Estos patrones son utilizados generalmente como plantillas que describen como debería ser estructurados y organizados los casos de uso. Son patrones que capturan mejores prácticas para modelar casos de uso. (Larman, 2003) En el MD Morbilidad y Mortalidad se evidencian los patrones de CU que a continuación se mencionan:

**CRUD Total:** el patrón CRUD total o completo consiste en un caso de uso para administrar la información, permite modelar las diferentes operaciones para administrar una entidad de información, tales como crear, leer, modificar y eliminar. (Larman, 2003)

**Múltiples actores rol común:** a un CU ingresan más de dos actores y estos tienen un rol común. (Larman, 2003)

**Concordancia de adición:** en el caso de este patrón alternativo, la subsecuencia común de casos de uso, extiende los casos de uso compartiendo la subsecuencia de acciones. Los otros casos de uso modelan el flujo que será expandido con la subsecuencia. Este patrón es preferible usarlo cuando otros casos de uso se encuentran propiamente completos, o sea, que no requieren de una subsecuencia común de acciones para modelar los usos completos del sistema. (Larman, 2003)

**2.3.5 Especificación del CU Presentar información de mortalidad poblacional**

Tabla 2. CU Presentar información de mortalidad poblacional

<b>Objetivo</b>	El objetivo principal es presentar toda la información relacionada con la mortalidad poblacional.	
<b>Actores</b>	Administrador, Especialista "(Inicia)".	
<b>Resumen</b>	El caso de uso inicia cuando el especialista o el administrador desean hacer un análisis de la información relacionada con la mortalidad poblacional. Seleccionan el reporte que desean ver, el sistema muestra la información contenida en él. El caso de uso finaliza cuando el actor termina el análisis de la información solicitada.	
<b>Complejidad</b>	Media.	
<b>Prioridad</b>	Alta.	
<b>Precondiciones</b>	Que el especialista o el administrador estén autenticados en el sistema. Que los datos solicitados estén cargados en el MD. Que los reportes relacionados con la información solicitada estén creados.	
<b>Referencias</b>		
<b>Flujo de eventos</b>		
<b>Curso Normal de los Eventos</b>		
<b>Acciones del Actor</b>	<b>Respuesta del Sistema</b>	
1. El especialista o administrador inicia el caso de uso al entrar al sistema  3. Realiza la selección del reporte de mortalidad que va a presentar la información.	2. El sistema muestra los reportes asociados a la mortalidad poblacional.  4. Muestra el reporte seleccionado por el especialista o el administrador.  5. El caso de uso termina al salir de la sesión o cerrar el sistema.	
<b>Relaciones</b>	<b>CU incluidos</b> N/A	
	<b>CU extendidos</b>	Realizar operaciones sobre reporte.
<b>Requisitos no funcionales</b>	N/A	
<b>Asuntos pendientes</b>	N/A	

## 2.4 Reglas del negocio

Las reglas de negocios (o las directivas empresariales) definen y controlan la estructura, el funcionamiento y la estrategia de una organización. Pueden estar formalmente definidas en manuales de procedimiento, contratos o acuerdos, o bien pueden existir como conocimiento o experiencia que tienen los empleados. Además son dinámicas, están sujetas a cambios en el tiempo y pueden encontrarse en todo tipo de aplicaciones (Microsoft, 2016). Las mismas son una entrada fundamental para los procesos de diseño del almacén, ETL y BI. Por esta razón la identificación de las mismas es de suma importancia para lograr el objetivo a cumplir y constituye un eslabón esencial dentro del sistema. Para más detalles dirigirse al artefacto Reglas del Negocio y Transformación. A continuación se muestran algunas reglas del negocio definidas para el MD Morbilidad y Mortalidad.

**Reglas de almacenamiento:** definen características específicas del almacenamiento de alguna variable. (Hernández, 2013)

- ✓ No pueden existir campos con valores negativos.
- ✓ Los id de las dimensiones no pueden tomar valores nulos.
- ✓ Las personas con sexo y/o raza desconocidos se les inserta el valor "desconocido" en el campo correspondiente.

**Reglas de variables:** definen las variables calculables que son objeto de análisis. (Hernández, 2013)

- ✓ Para el cálculo de la tasa de letalidad se tiene en cuenta el número de defunciones entre el total de enfermos multiplicado por K.
- ✓ La tasa de mortalidad infantil depende del total de defunciones de niños menores de un año en un periodo, dividido entre los nacidos vivos en el año multiplicado por K.
- ✓ Para determinar la tasa bruta de mortalidad se tiene el total de defunciones por año entre el total de la población multiplicado por K.

El valor de la variable K para el cálculo de las tasas relacionadas con morbilidad y mortalidad poblacional puede tomar valores potencias de 10 hasta 10000000.

**Reglas de visualización:** implican alguna condición para la visualización de una variable. (Hernández, 2013)

- ✓ Para la presentación de las variables de tipo *float* se deben visualizar solo con dos lugares decimales después de la coma.

## 2.5 Definición de la arquitectura base del MD

La arquitectura del MD Morbilidad y Mortalidad quedó estructurada de manera tal que se compone por la fuente de datos Salud (representa la BD del sistema XAVIA HIS) y tres subsistemas bases, los cuales se pueden apreciar en la figura 5 y se describen a continuación:

- Subsistema de integración: encargado de la extracción, limpieza e integración de la información para su posterior carga al MD.
- Subsistema de almacenamiento: encargado de almacenar todos los datos del MD en los diferentes hechos y dimensiones definidas.
- Subsistema de visualización: encargado de consultar los datos almacenados en el almacén, con el objetivo de mostrarlos a los usuarios finales en los distintos gráficos y reportes, contribuyendo a la toma de decisiones.

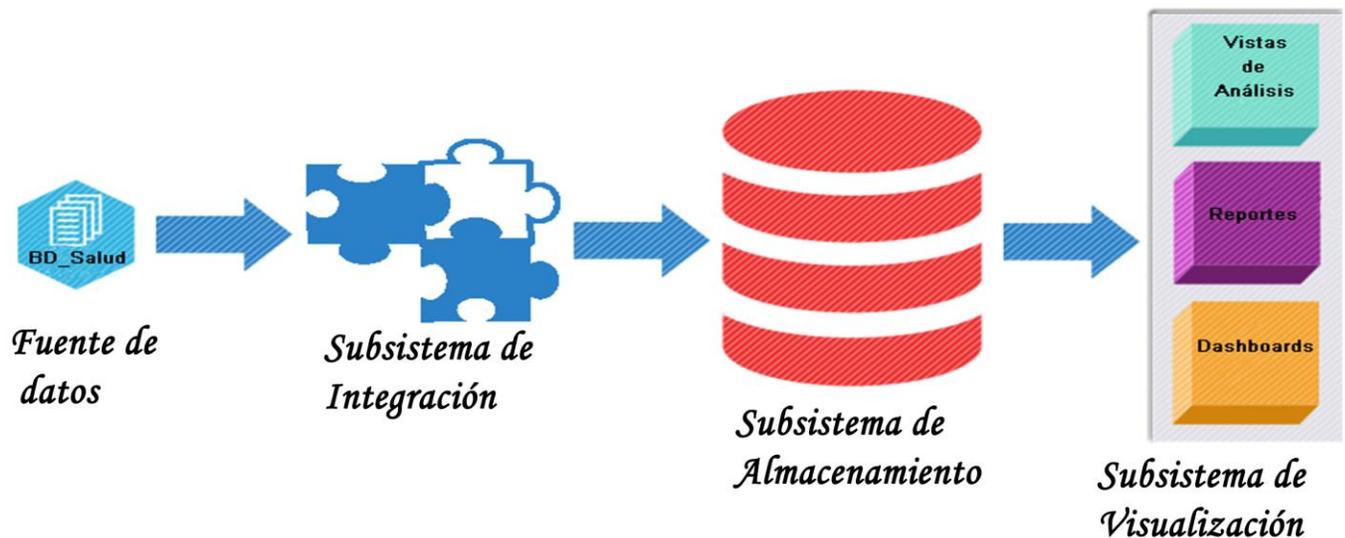


Fig 5. Arquitectura del MD Morbilidad y Mortalidad

Primeramente se encuentra el subsistema de integración que obtiene la información de la BD del Sistema XAVIA HIS, en el cual se realiza el proceso de integración que consiste en la carga, limpieza y transformación de los datos para su almacenamiento posterior. Luego el subsistema de almacenamiento recibe la información integrada y la almacena en la BD xavia\_his\_dwh, la cual es soportada por el SGBD PostgreSQL 9.2 y es administrada por los usuarios que tienen acceso a la información mediante pgAdminIII 1.16.1. Finalmente, en el Subsistema de visualización se muestra la información a través de vistas de

análisis y reportes operacionales mediante la herramienta Pentaho BI Server 5.4, la cual es accedida por los usuarios autorizados a la visualización de las vistas de análisis.

## 2.6 Diseño del MD

Para el diseño del MD Morbilidad y Mortalidad se realiza el modelo dimensional de los datos, el cual contiene los hechos y dimensiones, así como las relaciones entre las mismas.

### 2.6.1 Subsistema de almacenamiento

En el diseño del subsistema de almacenamiento se definieron ocho dimensiones y cuatro hechos. Los mismos almacenarán toda la información relacionada con la morbilidad y mortalidad en las instituciones de salud. A continuación, se presenta una tabla con los hechos y dimensiones y su respectiva descripción.

Tabla 3. Hechos y dimensiones del MD Morbilidad y Mortalidad.

Dimensión/Hecho	Descripción
dim_servicio	Se refiere al servicio o atención que brinda la institución de salud al paciente.
dim_dpa	Es donde se encontrará la División Política Administrativa de Venezuela, desde país hasta la entidad hospitalaria, incluyendo el municipio, estado y localidad.
dim_causa	Se almacenará la información referente a las enfermedades definidas por la OMS. Estará constituida por una jerarquía que abarca cinco niveles: capítulo, grupo, categoría, subcategoría y enfermedad.
dim_raza	Almacenará todas las razas existentes en la fuente de datos Salud. Incluirá la definición de desconocido, en caso de que sea imposible determinar a simple vista la raza de la persona (lesiones o quemaduras graves en todo el cuerpo).
dim_edad	Presentará los grupos etarios donde se encuentran comprendidas las edades de las personas que son atendidas en las diferentes instituciones de salud.
dim_tiempo	Como su nombre lo indica se va a almacenar la información referente al tiempo (año, trimestre, mes,

	semana y día). Presentará cinco jerarquías con sus niveles correspondientes.
dim_sexo	Almacenará todas las razas existentes en la fuente de datos Salud. También se incluye la definición de desconocido, en caso de que sea imposible determinar a simple vista el sexo de la persona (pérdida o quemaduras graves en los órganos genitales).
dim_persona	Contendrá la cédula de la persona para lograr que no existan muertes repetidas en el MD morbilidad y mortalidad.
hech_mortalidad	Presentará los indicadores referentes a la mortalidad de una población en un tiempo y espacio determinados.
hech_morbilidad	Agrupará indicadores referentes a la morbilidad de una población en un tiempo y espacio acotados. Exceptuando los indicadores de obesidad y bajo peso.
hech_bajo_peso	Almacenará el indicador relacionado con los nacidos vivos con menos de 2500g.
hech_masa_corporal	Presentará la cantidad de personas obesas (con IMC > 27kg) y personas con sobrepeso (con IMC entre 25kg y 26,9kg).

Para el MD Morbilidad y Mortalidad se identificaron 21 medidas de las cuales seis son calculables y las restantes son físicas y todas son no aditivas excepto la cantidad de personas obesas y el total de la población. A continuación se muestra una tabla con las medidas agrupadas por los hechos del MD Morbilidad y Mortalidad.

Tabla 4. Hechos y medidas

Hechos	Medidas
<b>hech_mortalidad</b>	cant_defunciones_menores_1
	nacidos_vivos_anno
	muertes_nacimiento_27dias
	muertes_nacimiento_6dias
	muertes_nacimiento_11meses

	defunciones_1_4_años
	poblacion_1_4_años
	defunciones_1_11meses
	total_defunciones
	cant_defunciones_antes_48h
	cant_defunciones_antes_24h
<b>hech_morbilidad</b>	total_poblacion
	total_egresos
	total_enfermos
	casos_nuevos_enfermedad
	cant_casos_existentes
	egresos_48h_mas
<b>hech_bajo_peso</b>	cant_nacidos_vivos
	total_nacidos_vivos
<b>hech_indice_masa_corporal</b>	cantidad_personas_obesas
	cant_sobre_peso

### 2.6.2 Matriz dimensional o matriz bus del MD Morbilidad y Mortalidad

La matriz bus es una herramienta clave de diseño que representa los procesos de negocio centrales de la organización y la dimensionalidad asociada. Es el modelo arquitectónico que transmite la perspectiva estratégica de arriba hacia abajo para asegurar que los datos en el entorno de AD / BI se pueden integrar en toda la empresa, mientras que la entrega de abajo hacia arriba se produce ágil, centrándose en un solo proceso de negocio a la vez. (Kimball Group, 2016)

Tabla 5. Matriz Bus del MD Morbilidad y Mortalidad

Dimensiones	Hechos			
	hech_Indice_masa_corporal	hech_bajo_peso	hech_morbilidad	hech_mortalidad
<b>dim_dpa</b>	X	X	X	X
<b>dim_causa</b>			X	X
<b>dim_edad</b>	X		X	X
<b>dim_persona</b>				X

<b>dim_raza</b>	X	X	X	X
<b>dim_servicio</b>			X	X
<b>dimsexo</b>	X	X	X	X
<b>dim_tiempo</b>	X	X	X	X

En el MD Morbilidad y Mortalidad la matriz bus permitió reconocer que los cuatro hechos identificados comparten dimensiones, sin embargo, no existen hechos solapados, es decir, hechos que compartan exactamente las mismas dimensiones. Se identificó además que de las ocho dimensiones cuatro son compartidas por los cuatro hechos.

### **2.6.3 Modelo de datos**

Una vez definidas en el negocio las dimensiones y los hechos se conforma el modelo de datos del MD Morbilidad y Mortalidad cuyo objetivo es representar la estructura lógica de la información, evidenciando además las relaciones existentes entre los hechos y las tablas dimensiones. A continuación, se muestra en la Figura 8 el modelo de datos dimensional propuesto para el desarrollo del MD Morbilidad y Mortalidad.

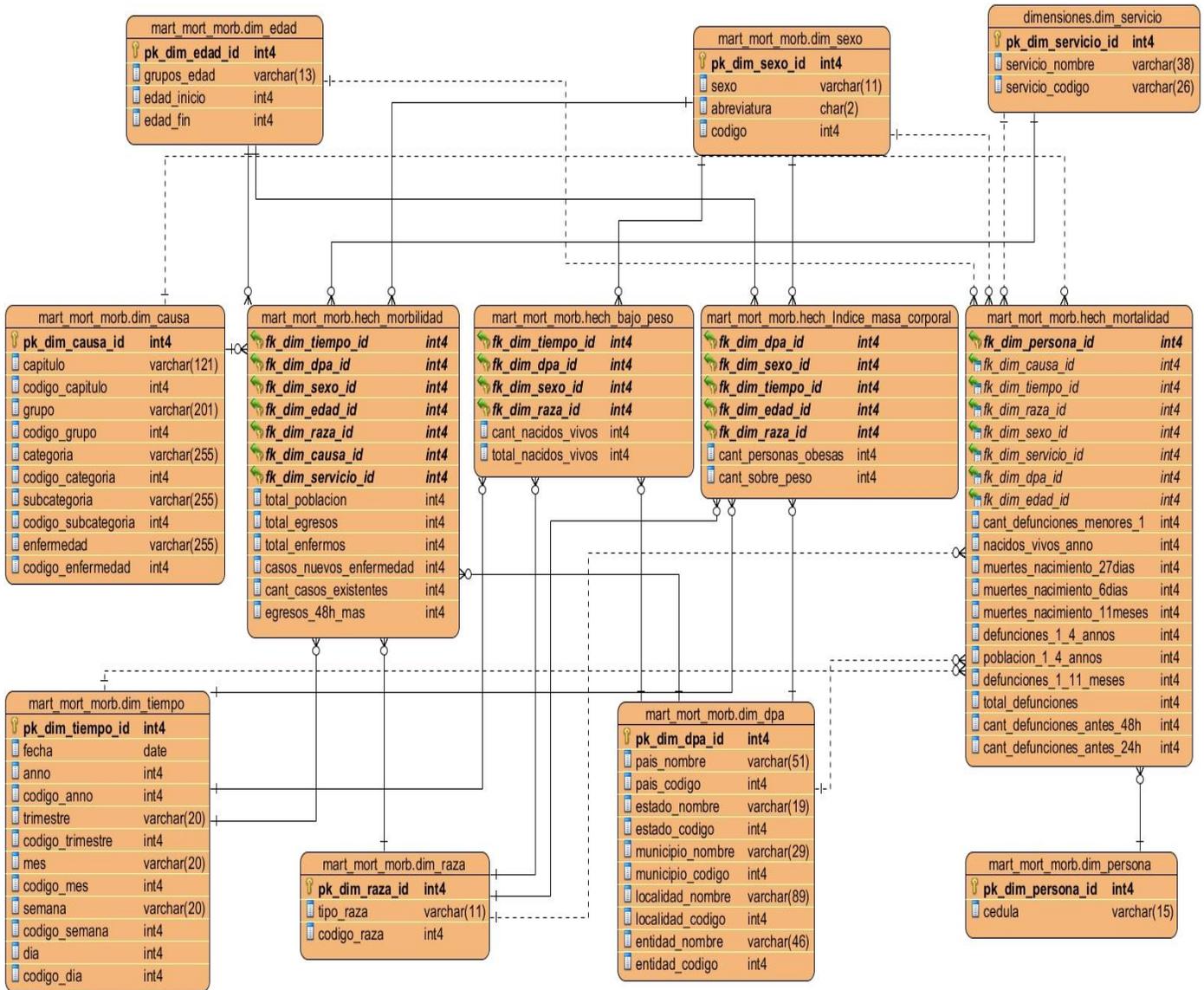


Fig 6. Modelo de datos del MD Morbilidad y Mortalidad

### Topología de esquema

La topología seleccionada para el diseño del modelo de datos físico y lógico del MD Morbilidad y Mortalidad fue el esquema Constelación de Hechos, que es utilizado cuando existen hechos que comparten algunas dimensiones. Los cuatro hechos identificados (hech\_indice\_masa\_corporal, hech\_bajo\_peso, hech\_morbilidad y hech\_mortalidad) presentan dimensiones comunes.

### 2.6.4 Política de respaldo y recuperación

En el MD Morbilidad y Mortalidad la política de respaldo y recuperación se lleva a cabo teniendo en cuenta tres aspectos principales: tablas involucradas, periodicidad de las salvas y backups existentes.

- **Tablas involucradas:** en la realización de las salvas se involucran todas las tablas del MD.
- **Periodicidad de las salvas:** las salvas de la información contenida en el MD se harán en correspondencia con la periodicidad con la que se accede a la información. La salva de la información asociada a los cuatro hechos se hará con una frecuencia mensual.
- **Backups existentes:** cada año se realizarán cuatro salvas, una salva trimestral. En esta salva se encuentra la información asociada a los indicadores referentes a los términos morbilidad y mortalidad al cierre de cada trimestre del año.

### 2.7 Subsistema de integración

El diseño del subsistema de integración es fundamental para un MD, este incluye el perfilado de datos con el fin de supervisar la calidad, estado y estructura de los datos provenientes de la fuente. Después de realizar el perfilado de datos a la fuente de datos del MD Morbilidad y Mortalidad se pudo observar que los datos encontrados son: textos y números enteros, además no existen valores negativos y se encontraron valores vacíos. Permite además verificar los distintos tipos de datos que contiene la fuente de información del MD Morbilidad y Mortalidad.

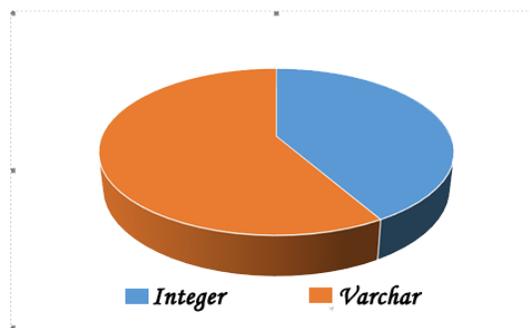


Fig 7. Tipos de datos de la fuente

#### 2.7.1 Diseño de las transformaciones

Las transformaciones son el elemento fundamental del proceso de ETL, en su diseño se detalla cada uno de los pasos a seguir para efectuar la carga de las dimensiones y los hechos en el MD. A continuación, en la figura 8, se muestra de forma general el diseño de las transformaciones del MD Morbilidad y Mortalidad.

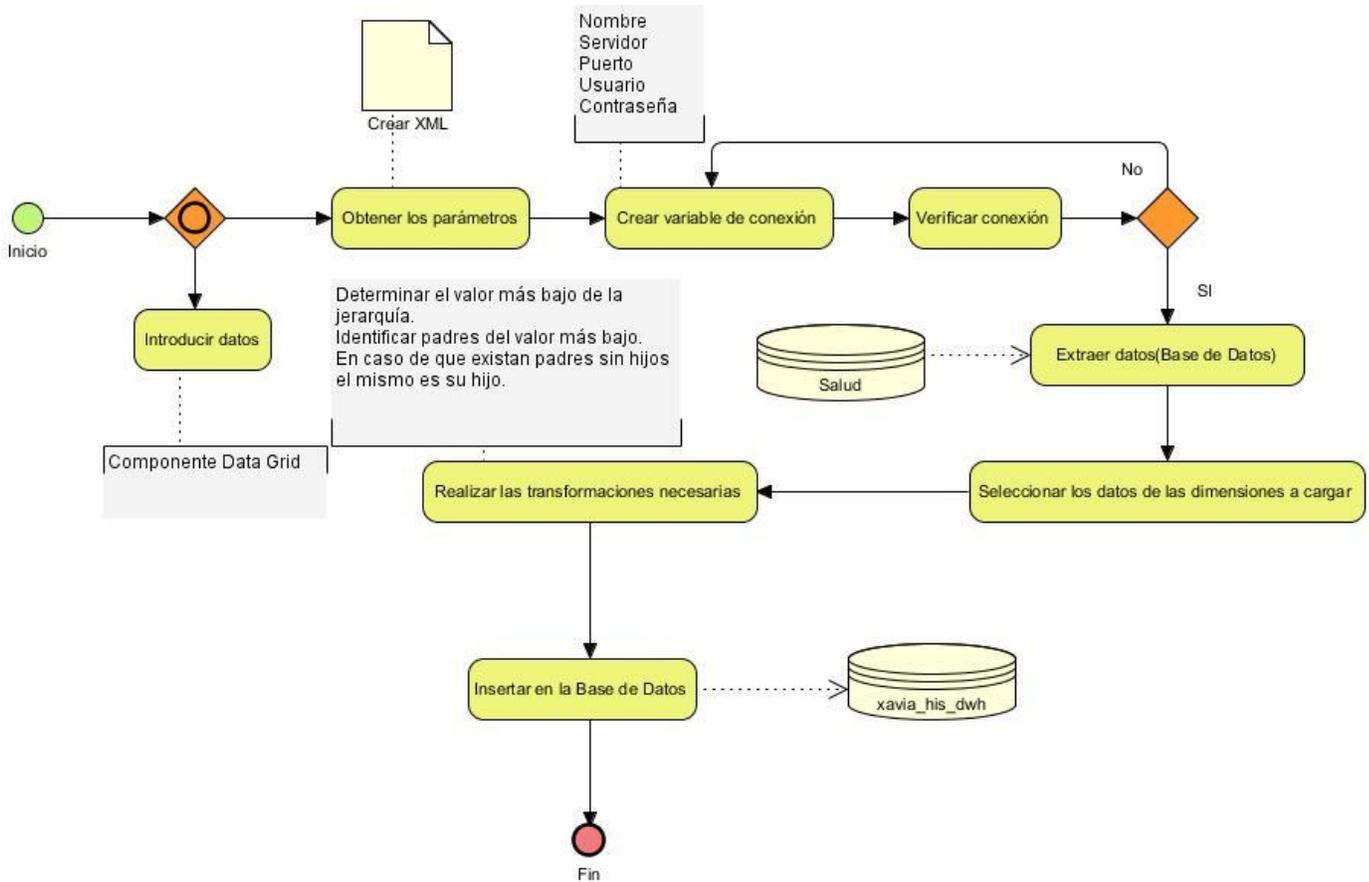


Fig 8. Diseño de las transformaciones

### 2.7.2 Estrategias generales de integración de datos

Para el desarrollo del MD se utiliza como estrategia de integración de datos ETL. La primera parte del proceso ETL consiste en localizar y **extraer** los datos, ya que en la mayoría de los proyectos de almacenamiento se fusionan datos provenientes de diferentes sistemas de origen. Los formatos de las fuentes normalmente se encuentran en bases de datos relacionales o ficheros planos, pero pueden incluir también bases de datos no relacionales y otras estructuras o formatos diferentes. En este caso los datos provienen de una única fuente que sería la BD Salud. La extracción convierte los datos a un formato homogéneo y consolidado para iniciar el proceso de **transformación**. Un requerimiento importante que se debe exigir a la tarea de extracción es que ésta cause el menor impacto en el sistema de origen. Si los datos a extraer son muchos, el sistema de origen se podría ralentizar e incluso colapsar, provocando que éste no pueda utilizarse con normalidad para su uso cotidiano. La fase de **carga** es el momento en el cual los datos de la fase anterior (transformación) son cargados en el sistema de destino BD xavia\_his\_dwh. Este proceso puede abarcar una amplia variedad de acciones diferentes. En algunas bases de datos se sobrescribe la información antigua con nuevos datos. En la mayoría de los casos, se generan AD o MD en dónde se

mantiene un historial de los registros de manera que se pueda hacer una auditoría de los mismos y disponer de un rastro de toda la historia de un valor a lo largo del tiempo. (BUYTO, 2016)

## 2.8 Diseño del subsistema de visualización

El diseño del Subsistema de visualización se realiza con el objetivo de organizar las vistas de análisis, reportes operacionales y *dashboards* o tableros de control por áreas de análisis, facilitando al usuario una búsqueda rápida de la información.

**Vistas de análisis (VA):** informe que expresa mediante tablas los valores de los indicadores analizados

**Reportes operacionales:** los reportes operacionales tienen por objetivo proporcionar soporte para la toma de decisiones en un entorno con un movimiento potencialmente rápido, dando una imagen detallada del presente y el futuro inmediato, de forma que las acciones individuales pueden administrarse en tiempo real o casi en tiempo real. (eHow, 2016)

**Dashboards:** es una herramienta muy poderosa y completa que combina a la perfección con el extremo delantero de las fuentes de datos y componentes personalizados. (PentahoCorporation., 2005) Se utiliza mayormente para resaltar los indicadores más importantes de un negocio a través de gráficas y tablas.

### 2.8.1 Arquitectura de la información

Con la creación del MD se podrá almacenar la información asociada a los términos morbilidad y mortalidad, para ser analizada por las personas capacitadas y sirviendo de apoyo al proceso de toma de decisiones en las instituciones de salud. En la investigación se identificó un área de análisis que se divide en dos libros de trabajo que incluyen los reportes, vistas de análisis y *dashboards* definidos por el cliente.

- **Área de análisis general (A.A.G):** contiene todas las áreas de análisis que abarca el Sistema de Gestión Hospitalaria XAVIA HIS.
- **Área de análisis (AA):** son las secciones en las que se divide el AD, las que contienen los diferentes libros de trabajo.
- **Libro de trabajo (LT):** son las diferentes categorías en las que se pueden agrupar cada una de las vistas de análisis, reportes y *dashboards* que se implementen.

En la siguiente imagen se muestra el mapa de navegación del MD Morbilidad y Mortalidad. (Figura 10)

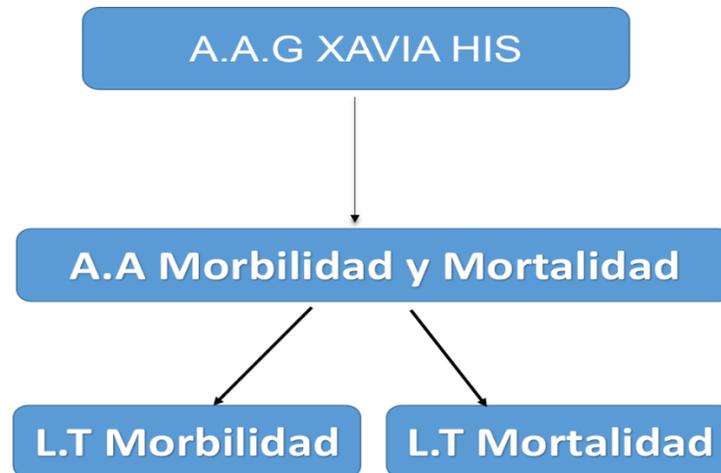


Fig 9. Mapa de navegación

### **AA. Morbilidad y Mortalidad**

**LT Morbilidad:** contiene todas las vistas de análisis, reportes y *dashboards* asociados con la morbilidad poblacional, teniendo en cuenta las aristas de análisis necesarias para la presentación de la información.

**LT Mortalidad:** contiene todas las vistas de análisis, reportes y *dashboards* asociados con la mortalidad poblacional, teniendo en cuenta las aristas de análisis necesarias para la presentación de la información.

Los dos libros de trabajos identificados, contendrán su información asociada teniendo en cuenta el tiempo, la División Política Administrativa, el servicio prestado, la raza, la enfermedad o causa que da un diagnóstico, el sexo y la edad de las personas.

#### **2.8.2 Diseño de los cubos OLAP**

En el MD Morbilidad y Mortalidad se definieron cuatro cubos multidimensionales y un cubo virtual con el objetivo de calcular medidas que utilizan otras medidas existentes en hechos diferentes. Los mismos fueron realizados con el Pentaho Schema Workbench 3.10. En la solución propuesta se identificaron los siguientes cubos: bajo peso, índice de masa corporal, mortalidad, morbilidad, además del cubo virtual y en total existen 18 medidas calculables. La figura 11 muestra el diseño de los cubos identificados.



Fig 10. Diseño de los cubos del MD Morbilidad y Mortalidad

## 2.9 Esquema de seguridad

En el MD Morbilidad y Mortalidad el esquema de seguridad está respaldado por los niveles de acceso, específicamente por los roles definidos. De esta forma cada usuario accede solo a las funcionalidades que le corresponden.

### Seguridad en el subsistema de almacenamiento

Para la seguridad en el subsistema de almacenamiento se hizo un diseño de forma tal que el rol Administrador posee acceso total a la BD y el Administrador de ETL se encargaría de los procesos de extracción, transformación y carga. En la Tabla 5 se muestra el esquema de seguridad definido para el subsistema de almacenamiento.

Tabla 6. Esquema de seguridad subsistema de almacenamiento

Roles	Permisos
<b>Administrador</b>	<ul style="list-style-type: none"> <li>• Gestión de usuarios.</li> <li>• Gestión de reportes.</li> <li>• Gestión de roles.</li> </ul>
<b>Administrador ETL</b>	<ul style="list-style-type: none"> <li>• Administración de la BD</li> <li>• Extracción y limpieza de datos</li> <li>• Transformación y carga de la información.</li> </ul>

### Seguridad en el subsistema de visualización

Para la seguridad de la aplicación se definió el rol Administrador, que posee total acceso al área de análisis general instituciones de salud, además de ser el encargado de la creación de nuevos reportes, usuarios, así como asignarles los roles y permisos a los mismos y el rol especialista consulta las vistas de análisis correspondientes a cada uno de los libros de trabajo del área de análisis Morbilidad y Mortalidad. En la Tabla 6 se muestra el esquema de seguridad definido para el Subsistema de Visualización.

Tabla 7. Esquema de seguridad subsistema de visualización

Roles	Permisos
<b>Administrador</b>	<ul style="list-style-type: none"> <li>• Administración del Área Análisis General instituciones de salud.</li> <li>• Gestión de roles</li> <li>• Gestión de usuarios</li> <li>• Gestión de reportes</li> </ul>
<b>Especialista</b>	<ul style="list-style-type: none"> <li>• Visualización de las vistas de análisis, reportes candidatos y <i>dashboards</i> del Área Análisis Morbilidad y Mortalidad.</li> </ul>

### Conclusiones del capítulo

Con la culminación de este capítulo se identificaron los principales aspectos que contribuyen a la correcta implementación de la solución. Se definieron las necesidades de información de los usuarios, así como los requisitos del sistema, realizándose la modelación de los mismos a través del diagrama de casos de uso. Se diseñaron además los subsistemas de almacenamiento, integración y visualización quedando estructurado el MD Morbilidad y Mortalidad.

## Capítulo 3: Implementación del MD Morbilidad y Mortalidad

El presente capítulo está encaminado a la implementación de los diferentes aspectos relacionados con los procesos de integración de datos, con el propósito de brindar una mayor comprensión de las estrategias y procedimientos utilizados. Además, se abordan elementos relacionados con la implementación de la capa de visualización, incluyendo la creación de las estructuras necesarias para la navegación y el análisis de los datos.

### 3.1 Implementación del subsistema de almacenamiento

Una vez realizado el diseño del modelo de datos dimensional se procedió con la transformación al modelo físico. El mismo detalla de forma sencilla como estarán almacenados los datos y las respectivas relaciones entre los hechos, dimensiones y medidas que lo componen.

#### Estructura de los datos

Una estructura de datos es, a grandes rasgos, una colección de datos (normalmente de tipo simple) que se caracterizan por su organización y las operaciones que se definen en ellos. Por tanto, una estructura de datos vendrá caracterizada tanto por unas ciertas relaciones entre los datos que la constituyen, como por las operaciones posibles en ella. (Universidad de Valencia, 2016)

**Esquemas:** El esquema de una BD describe la estructura de la misma, en un lenguaje formal soportado por un SGBD. En una BD relacional, el esquema define sus tablas, sus campos en cada tabla y las relaciones entre cada campo y cada tabla. (Instituto Tecnológico de Piedras Negras, 2016)

En el proceso de implementación se determinó que se iban a utilizar dos esquemas:

- **Esquema dimensiones:** este esquema contiene las dimensiones comunes del AD Salud.
- **Esquema mart\_mort\_morb:** en este esquema se encuentran los hechos y dimensiones específicas del MD Morbilidad y Mortalidad.

**Tablas:** las tablas de una BD conforman la estructura principal de la misma, se podría decir que se trata de los cimientos de esta estructura. Cada tabla contiene la información que utilizarán el resto de objetos de la BD: consultas, formularios, informes. ( Instituto Nacional de Tecnologías Educativas y de Formación del Profesorado, 2016) La solución propuesta contiene un total de dos tablas contenidas en el esquema dimensiones y once tablas contenidas en el esquema mart\_mort\_morb, para un total de trece tablas que se representan en la siguiente tabla que describe dicha estructura.

Tabla 8. Esquemas y tablas

No.	Esquema	Tablas
1	dimensiones	dim_dpa
		dim_servicio
2	mart_mort_morb	dim_causa
		dim_edad
		dim_persona
		dim_raza
		dim_sexo
		dim_tiempo
		hech_bajo_peso
		hech_indice_masa_corporal
		hech_mortalidad
		hech_morbilidad
tasa		

### 3.1.1 Estándares de codificación

Los estándares de codificación se utilizan para lograr un entendimiento entre las partes implicadas en un proyecto. Para el desarrollo del MD Morbilidad y Mortalidad fueron utilizados varios estándares de codificación utilizados por los especialistas en DATEC. A continuación se exponen algunos de ellos:

Tabla 9. Estándares de codificación del MD Morbilidad y Mortalidad

Tipo de objeto	Función	Nomenclatura	Descripción
<b>Esquema</b>	Dimensiones compartidas	dimensiones	Esquema donde se organizan las dimensiones compartidas por varios MD.
<b>Tablas</b>	Dimensiones	dim_[nombre]	Dimensiones utilizadas. Ejemplo: dim_raza, dim_sexo

	Hechos	hech_[nombre]	Hechos que definen las principales medidas requeridas para calcular indicadores y otras medidas derivadas. Ejemplo: hech_mortalidad, hech_morbilidad.
--	--------	---------------	--

### 3.2 Implementación del subsistema de integración

La información histórica del sistema XAVIA HIS se encuentra almacenada en una BD relacional soportada por el SGBD PostgreSQL en su versión 9.2. Para lograr la correcta integración de los datos se debe realizar primeramente un análisis de la fuente de datos con el objetivo de detectar errores y entradas duplicadas. Posteriormente se lleva a cabo el proceso de integración el cual consta de tres etapas fundamentales extracción, transformación y carga. La transformación de los datos nos permite filtrar información, combinar y ordenar los datos para su posterior inserción en la BD xavia\_his\_dwh del MD Morbilidad y Mortalidad.

#### 3.2.1 Implementación de las transformaciones

La etapa de transformación aplica un conjunto de reglas para transformar los datos desde el origen al destino. Esto incluye la conversión de los datos medidos usando las mismas unidades para que más adelante se puedan unir. La etapa de transformación también requiere unir datos de varias fuentes, la generación de agregados, la generación de llaves subrogadas, clasificando, derivando nuevos valores calculados, y aplicando las reglas de validación avanzada. (Javlin Data Solution, 2015)

En la presente investigación se realizó una transformación para cada una de las cargas de las tablas pertenecientes a los esquemas que componen el MD Morbilidad y Mortalidad. Para la carga de las dimensiones se realizaron las transformaciones pertinentes, a partir de los indicadores relacionados con la Morbilidad y la Mortalidad en la fuente de datos. Estas dimensiones contienen los datos que son necesarios para poblar la BD destino xavia\_his\_dwh para posteriormente realizar las transformaciones de los hechos. Para realizar la carga de la dimensión servicio (dim\_servicio) (ver figura 12) el primer paso es la extracción de los datos de la fuente Salud, luego se crea un componente Data Grid para agregar datos a la BD destino que no existen en la fuente, en este caso sería el valor “ninguno” con el objetivo de que no existan valores vacíos en dicha dimensión, este componente se une con el Extraer a través de un Append. Posteriormente con el componente Añadir Secuencia se agrega un id que va a ser el de la dimensión. El componente S/R se utiliza para verificar los datos que van a ser utilizados en el Filtrar fila para verificar si el código es nulo,

en caso de que así sea se añade una secuencia para agregar un código e insertarlo, en caso contrario se inserta directamente.

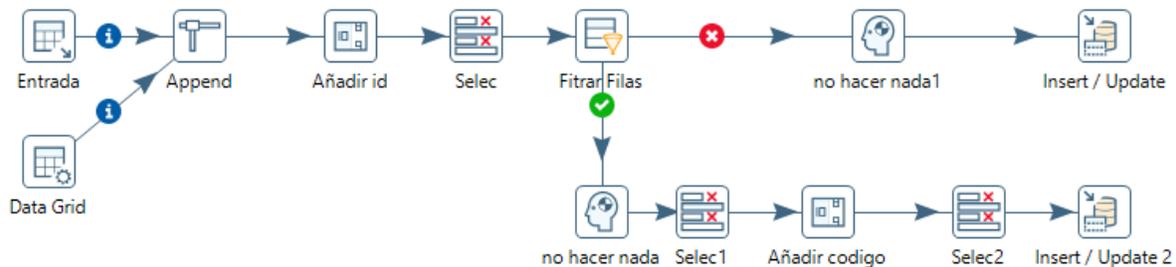


Fig 11. Transformación de la dimensión servicio.

Para la carga del hecho índice de masa corporal (`hech_indice_masa_corporal`) (ver figura 13) se realizaron una serie de transformaciones que se describen a continuación. Primeramente se extraen los datos de la fuente. Luego con el componente selecciona/renombra (S/R) se cambia el tipo de dato a las fechas para poder realizar cálculos con las mismas con el componente Calculadora. Una vez realizado esto se filtran las filas, se ordenan y posteriormente se obtiene la cantidad de personas obesas y con sobrepeso. Se unen los dos flujos y se hace la búsqueda de los id para la inserción en la BD `xavia_his_dwh`. En caso de que la dimensión raza contenga valores nulos se cambian por el valor “desconocida” y luego se hace un tratamiento de errores para verificar que no exista ningún valor nulo y proceder con la inserción de los datos.

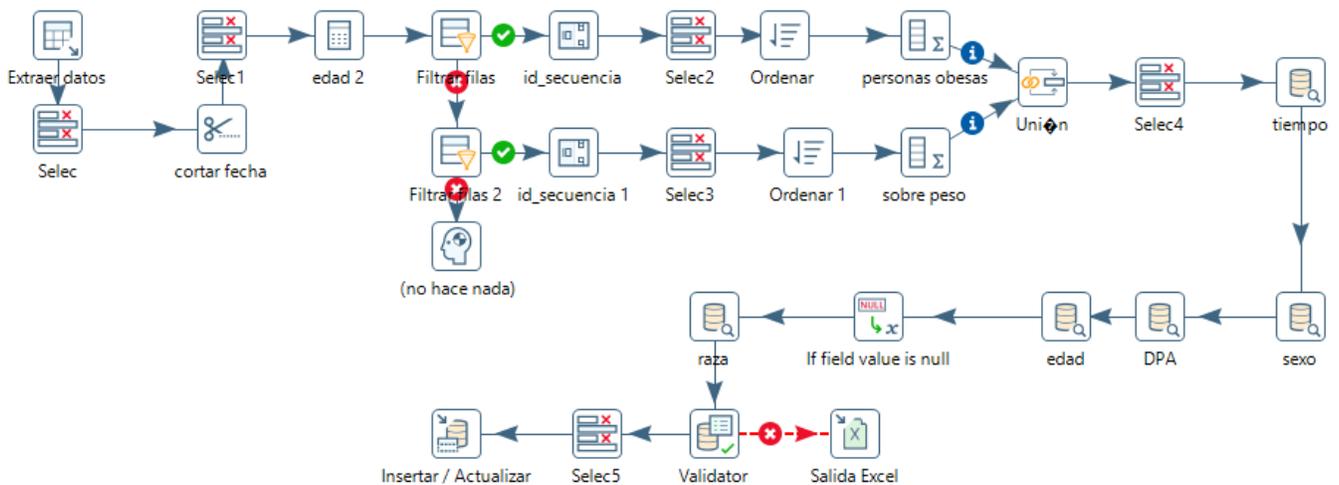


Fig 12. Transformación del hecho índice de masa corporal

### 3.2.2 Implementación de los trabajos

En la presente investigación se implementaron dos trabajos, uno para la carga de las ocho dimensiones (ver figura 13) y otro para la carga de los cuatro hechos (ver figura 14).

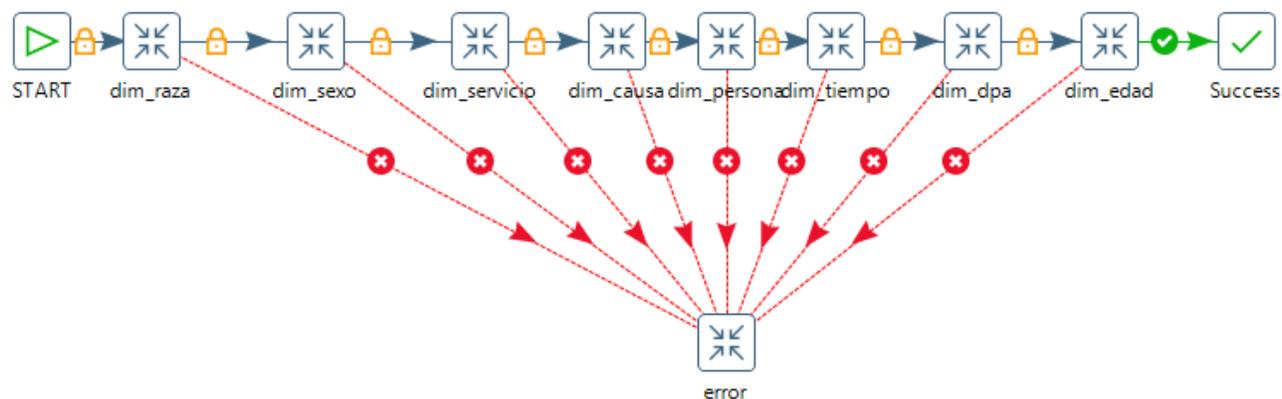


Fig 13. Trabajo para la carga de las dimensiones

En este trabajo se ejecutan todas las transformaciones implementadas para cada dimensión del MD Morbilidad y Mortalidad.

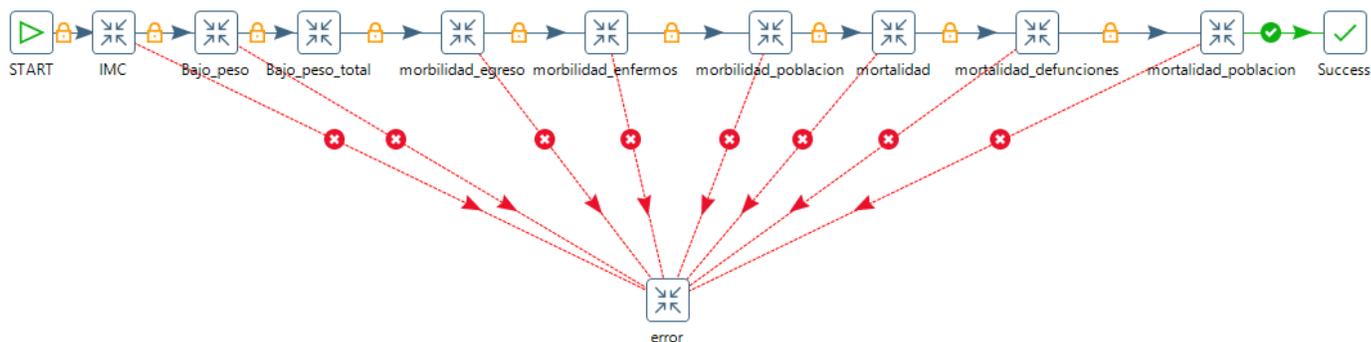


Fig 14. Trabajo para la carga de los hechos

En este trabajo se ejecutan todas las transformaciones implementadas para cada hecho del MD Morbilidad y Mortalidad.

### 3.3 Implementación del subsistema de visualización

Una vez cargados los datos al MD Morbilidad y Mortalidad se procede al modelado de los cubos OLAP y la creación de los reportes candidatos. Una vez implementados se puede analizar la información almacenada en el MD para que los especialistas puedan tomar las decisiones pertinentes.

### 3.3.1 Representación de la arquitectura de la información

El MD Morbilidad y Mortalidad está compuesto por un área de análisis general que sería XAVIA HIS, luego estaría el área de análisis Morbilidad y Mortalidad. Dentro de la misma se encuentran dos libros de trabajo, uno asociado a la mortalidad y el otro a la morbilidad, los cuales contienen las vistas de análisis, reportes y *dashboards* correspondientes a cada uno. A continuación, se muestra una figura con la Arquitectura de Información del MD Morbilidad y Mortalidad:

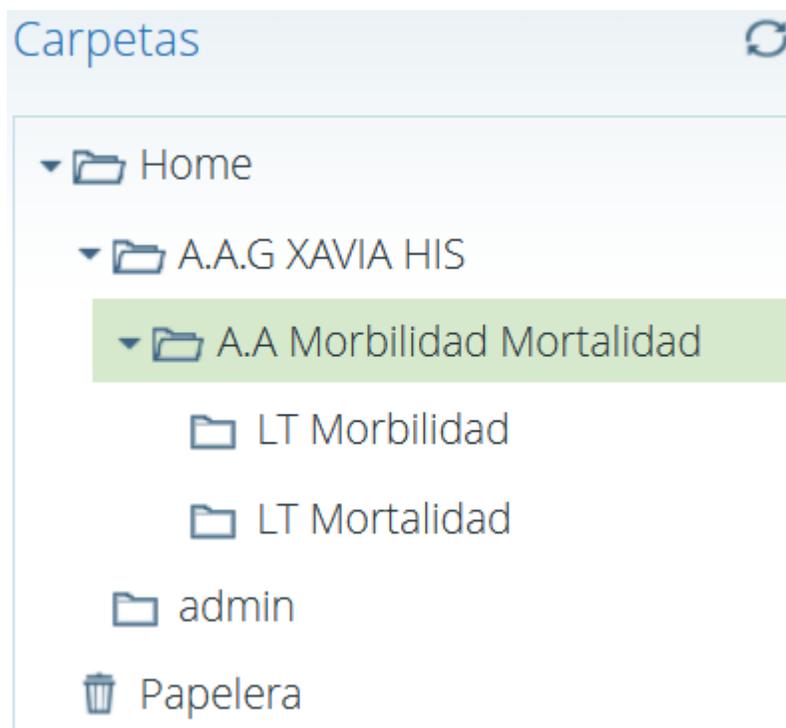


Fig 15. Arquitectura de la información del MD Morbilidad y Mortalidad

**A.A Mortalidad Morbilidad:** contiene la información asociada a los términos morbilidad y mortalidad.

#### Descripción de los libros de trabajo:

LT\_Morbilidad: presenta dos vistas de análisis, cuatro reportes operacionales y tres *dashboards* referentes a los indicadores asociados al término morbilidad.

LT Mortalidad: presenta dos vistas de análisis, diez reportes operacionales y un *dashboard* referentes a los indicadores asociados al término mortalidad. A continuación, se muestra la vista de análisis VA Obesidad que se encuentra en el LT Morbilidad y está asociado al hecho índice de masa corporal (ver figura 16):

			Medidas	
Grupo edad	Raza	Sexo	● Cantidad de personas obesas	● Cantidad de personas con sobrepeso
+ < 1 año	Blanca	Femenino	92	19
+ 25 a 29	Blanca	Femenino	34	10
+ 30 a 34	Blanca	Femenino	418	151
+ 40 a 44	Blanca	Femenino	2	1
		Masculino	14	14
+ 45 a 49	Blanca	Femenino	90	61
		Masculino	8	7
+ 50 a 54	Blanca	Femenino	3	4
		Masculino	34	10

Fig 16. VA Obesidad

En esta vista de análisis se evidencian la cantidad de personas obesas y con sobrepeso por grupos de edad, raza y sexo con el objetivo de analizar las tendencias de este padecimiento.

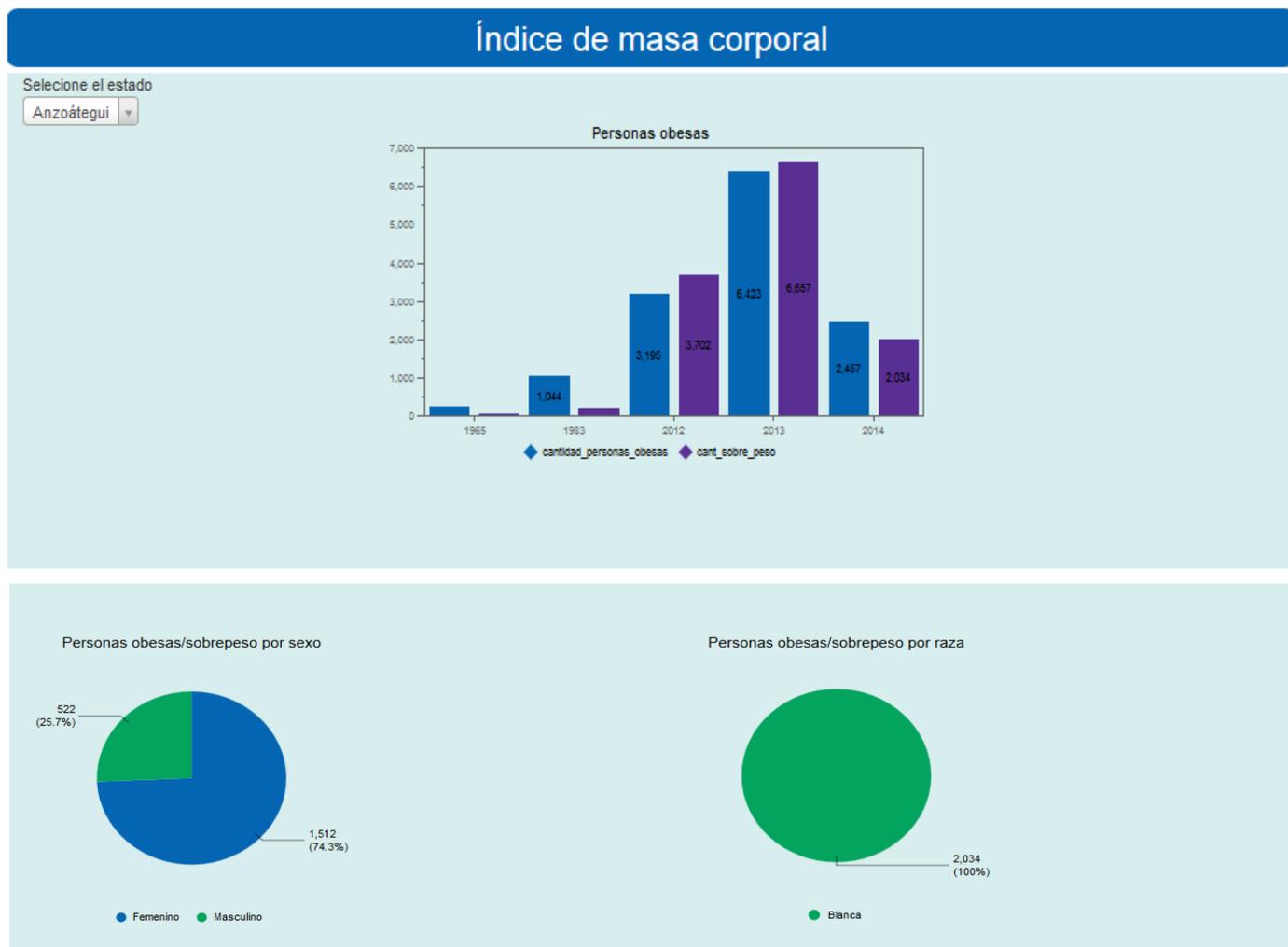


Fig 17. Dashboard sobre el índice de masa corporal

Los *dashboards* se realizan teniendo en cuenta los indicadores más importantes del negocio. En este *dashboard* (figura 17) se evidencian gráficas acerca de la cantidad de personas obesas y con sobrepeso por dpa, tiempo, raza y sexo con el objetivo de analizar las tendencias de este padecimiento.

### 3.4 Pruebas aplicadas al MD Morbilidad y Mortalidad

Las pruebas del software son un elemento crítico para la garantía de calidad del software y representa una revisión final de las especificaciones, del diseño y de la codificación. La creciente percepción del software como un elemento del sistema y la importancia de los costes asociados a un fallo del propio sistema, motivan la creación de pruebas minuciosas y bien planificadas. (Pressman, 2002)

Durante el desarrollo de la solución propuesta en la presente investigación se utilizó el Modelo V, para verificar que el producto posea la calidad requerida y brinde ayuda en el desarrollo de las pruebas. El modelo V demuestra cómo se relacionan las actividades de prueba con las de análisis y diseño. La punta de la V es la codificación, con el análisis y el diseño a la izquierda y la prueba y el mantenimiento a la derecha. El modelo V sugiere que la prueba unitaria y de integración sea hecha también para verificar el diseño, verificando que todos los aspectos del diseño del programa se han implementado correctamente en el código, la prueba del sistema debe verificar el diseño del sistema, asegurando que todos los aspectos del diseño están correctamente implementados. La prueba de aceptación que es dirigida por el cliente, valida los requerimientos. La vinculación entre la parte izquierda y derecha del modelo V implica que si se encuentran problemas durante la verificación y la validación entonces el lado izquierdo de la V puede ser ejecutado nuevamente para solucionar el problema y mejorar los requerimientos, el diseño y el código antes de retomar las pruebas del lado derecho (ver figura 17). (Rojas Ricardo, y otros, 2012)



Fig 18. Modelo V

### Pruebas unitarias

Las pruebas unitarias son el proceso de probar los componentes individuales del sistema. Es un proceso de pruebas de defectos, por lo que su objetivo es encontrar defectos en estos componentes. (Pressman, 2002)

### Pruebas de integración

El proceso de integración del sistema implica construir este a partir de sus componentes y probar el sistema resultante para encontrar problemas que pueden surgir debido a la integración de los componentes. Los componentes que se integran pueden ser componentes comerciales, componentes reutilizables que han sido adaptados a un sistema en particular o componentes nuevos desarrollados. Las pruebas de integración comprueban que estos componentes realmente funcionan juntos, son llamados correctamente y transfieren los datos correctos en el tiempo preciso a través de sus interfaces. (Pressman, 2002)

### Pruebas de aceptación

Es donde los usuarios reales o potenciales de un sistema prueban el sistema en su propio entorno. Para productos de software, el "usuario" puede ser un grupo interno de marketing, que decide si el software se comercializa, se lanza y se vende. Las pruebas de aceptación se efectúan cuando el cliente prueba de

manera formal un sistema para decidir si debe aceptarse del proveedor del sistema, o si se requiere más desarrollo. (Sommerville, 2011)

### 3.5 Herramientas de pruebas

#### Casos de prueba

El objetivo del proceso de diseño de casos de prueba, es crear un conjunto de casos de prueba que sean efectivos descubriendo defectos en los programas y muestren que el sistema satisface sus requerimientos. Para diseñar un caso de prueba, se selecciona una característica del sistema o componente que se está probando. A continuación, se selecciona un conjunto de entradas que ejecutan dicha característica y se documenta las salidas esperadas o rangos de salida. (Pressman, 2002)

En la presente investigación se desarrollaron 21 casos de prueba, uno por cada pedido de información identificado en la etapa de análisis y diseño. A continuación, se muestra en la tabla 10, el caso de prueba correspondiente al CU “Mostrar información relacionada con la morbilidad poblacional” donde se verifica el reporte VA Cantidad de personas obesas

Tabla 10. Caso de prueba cantidad de personas obesas

Escenario	Descripción	Variables de entrada	Variables de salida	Respuesta del sistema	Flujo central
Cantidad de personas obesas y con sobrepeso	Muestra la cantidad de personas con índice de masa corporal (IMC) mayor que 27 (personas obesas) y la cantidad de personas con IMC mayor que 25 y menor que 26,9 (personas con sobrepeso).	grupo edad, raza, sexo, dpa, tiempo	Cantidad de personas obesas y con sobrepeso	Se muestra la información correspondiente al escenario	El usuario se autentica, selecciona examinar archivo/Inicio/ A.A.G XAVIA HIS/ A.A Morbilidad y Mortalidad/L.T Morbilidad/ Doble clic sobre el reporte a visualizar.

### 3.6 Resultados de las pruebas

#### Pruebas unitarias y de integración

Al término de la implementación del MD Morbilidad y Mortalidad se le realizaron pruebas unitarias y de integración con el objetivo de encontrar problemas en el sistema. Las mismas arrojaron en su totalidad siete no conformidades (NC), que fueron resueltas satisfactoriamente en el tiempo establecido. A continuación se muestra el comportamiento de las NC detectadas en cada iteración:

Tabla 11. NC por iteración

No. Iteración	Tipo de NC	Impacto	Descripción
<b>Iteración 1</b>	Funcionalidad	Alto	NC 1: Existen errores lógicos en la implementación de las transformaciones, ocasionando que la información que se cargue no sea la correcta.
	Error Técnico	Medio	NC 2: La conexión no se encuentra de forma genérica en todas las transformaciones, trayendo como consecuencia que no exista una estandarización de las conexiones definidas en las transformaciones.
	Funcionalidad	Alto	NC 3: Las transformaciones no están implementadas con la mayor optimización, lo que da lugar a que el tiempo de carga de los datos sea mucho mayor.
	Ortografía	Medio	NC 4: Los reportes operacionales presentan errores ortográficos, no brindándole al especialista la información con la calidad requerida.
	Funcionalidad	Alto	NC 5: La información de los reportes operacionales no estaba ubicada en tiempo y espacio, o sea, no se incluían las dimensiones tiempo y dpa, entorpeciendo el correcto análisis de la información.
	Funcionalidad	Alto	NC 6: El reporte media anual de defunciones muestra valores incorrectos, no mostrando la información adecuada a los especialistas.
	Funcionalidad	Alto	NC7: Existen filas vacías en los reportes operacionales, perdiendo sentido el análisis de los datos.
<b>Iteración 2</b>	Error Técnico	Medio	NC 2 sin solucionar.
	Ortografía	Medio	NC 4 sin solucionar.

<b>Iteración 3</b>	Se solucionaron todas las no conformidades.
--------------------	---

### Pruebas de aceptación

Se realizaron pruebas de aceptación al MD con el cliente arrojando una no conformidad la cual fue resuelta en el tiempo establecido.

Tabla 12 NC pruebas de aceptación

No. Iteración	Tipo de NC	Impacto	Descripción
<b>Iteración 1</b>	Funcionalidad	Alto	NC 1: Los reportes operacionales no muestran la información a nivel de municipio, localidad y entidad, provocando pérdida de información valiosa en estos niveles del dpa.
<b>Iteración 2</b>	Se solucionaron todas las no conformidades.		

Después de las dos iteraciones que se realizaron en la aceptación del sistema se emite una carta de aceptación la cual devela que el sistema satisface las necesidades del mismo (ver Anexo 2).

### Conclusiones del capítulo

En el capítulo fueron tratados temas que dieron paso a la implementación del MD. Se realizó la transformación al modelo físico de la BD, facilitando una implementación de la BD consecuente con el diseño multidimensional. La estructura de los datos quedó organizada mediante 13 tablas agrupadas en dos esquemas. Se realizó la carga de los datos de la fuente, quedando poblada la BD. En el proceso de BI se diseñaron cuatro cubos OLAP y un cubo virtual, un A.A y dos libros de trabajo, así como, 22 reportes candidatos, quedando así implementado el subsistema de visualización. Una vez concluido el MD Morbilidad y Mortalidad y aplicadas las pruebas de calidad al mismo, se logró constatar que mediante la identificación de las no conformidades y su posterior solución, el MD quedó totalmente funcional. Además de lograr que la aplicación desarrollada cumpliera con los objetivos definidos por el cliente.

## Conclusiones

Al término de este trabajo, se puede concluir que se elaboró el MD Morbilidad y Mortalidad arrojando resultados satisfactorios e influyendo en el proceso de toma de decisiones en el sector de la salud. Los siguientes resultados demuestran el cumplimiento de los objetivos propuestos en la investigación:

- ✓ El estudio teórico sobre los AD y MD se sustentó en algunas de las herramientas, las metodologías y las tecnologías a utilizar para el desarrollo del MD, lo que posibilita en esta investigación asumir para el desarrollo del MD Morbilidad y Mortalidad, la Metodología de Desarrollo para Proyectos de Almacenes de Datos basada en la propuesta de Kimball, pues esta presenta todas las etapas por las que transita un AD.
- ✓ Se realizó el análisis y diseño del MD Morbilidad y Mortalidad para las instituciones de salud. Se identificaron 14 requisitos funcionales, 22 de información, 24 requisitos no funcionales y 24 reglas del negocio. Posteriormente se diseñaron los subsistemas de almacenamiento, integración y visualización, para emplearlos en la posterior implementación del MD Morbilidad y Mortalidad.
- ✓ Se implementó el MD Morbilidad y Mortalidad para las instituciones de salud haciendo uso de las herramientas descritas. Como resultado se obtuvieron 16 reportes operacionales, cuatro vistas de análisis y cuatro *dashboards*.
- ✓ Se realizaron pruebas de aceptación, unitarias y de integración al MD encontrando siete no conformidades, las cuales se solucionaron en el tiempo establecido asegurando la calidad del producto y su aceptación por parte del cliente, pues permite realizar análisis estadísticos complejos entre diferentes indicadores, análisis históricos y de tendencias.

## Recomendaciones

- ✓ Incluir en el MD desarrollado los indicadores relacionados con la morbilidad y mortalidad materna a partir de los datos que se almacenan en el Sistema de Gestión Hospitalaria XAVIA HIS. De este modo se abarcan ampliamente todos los indicadores relacionados con los términos médicos de Mortalidad y Morbilidad.
- ✓ Desplegar el MD en las instituciones de salud para contribuir al proceso de toma de decisiones en los temas referidos a mortalidad y morbilidad poblacional.
- ✓ Ofrecer el servicio de soporte técnico al MD para garantizar su correcto funcionamiento y la disponibilidad e integridad de sus datos.
- ✓ Para futuras versiones del MD Morbilidad y Mortalidad para las instituciones de salud, diseñar e implementar el sistema de metadatos para la auditoría de los datos.
- ✓ Añadir a la estrategia de ETL los elementos necesarios para la carga incremental, de este modo se capturan los datos que cambiaron en la fuente desde la última extracción.

## Referencias Bibliográficas

- Instituto Nacional de Tecnologías Educativas y de Formación del Profesorado. 2016.** <http://www.ite.educacion.es/>. <http://www.ite.educacion.es/>. [En línea] 15 de Mayo de 2016. [http://www.ite.educacion.es/formacion/materiales/93/cd/m2\\_1/index.html](http://www.ite.educacion.es/formacion/materiales/93/cd/m2_1/index.html).
- Bernabeu, Ricardo. 2010.** *Investigación y Sistematización de Conceptos*. 2010.
- Bialogics. 2015.** Bialogics. [En línea] 27 de Noviembre de 2015. <http://www.bialogics.com/new-page/>.
- BUYTO. 2016.** <http://www.buyto.es>. <http://www.buyto.es>. [En línea] 6 de Mayo de 2016. <http://www.buyto.es/general-business-intelligence/datamining-transformacion-de-datos-en-business-intelligence>.
- Celma, Matilde. 2002.** *Almacenes de datos*. 2002.
- Ceria, Santiago. 2016.** <http://www-2.dc.uba.ar/>. <http://www-2.dc.uba.ar/>. [En línea] 6 de Mayo de 2016. [http://www-2.dc.uba.ar/materias/isoft1/2001\\_2/apuntes/CasosDeUso.pdf](http://www-2.dc.uba.ar/materias/isoft1/2001_2/apuntes/CasosDeUso.pdf).
- Dario, Ing. Bernabeu Ricardo. 2010.** *HEFESTO: Metodología para la Construcción de un Data Warehouse*. Córdoba : s.n., 2010.
- Data Cleaner Team. 2016.** <http://datacleaner.org/>. <http://datacleaner.org/>. [En línea] 6 de Mayo de 2016. <https://datacleaner.org/resources/docs/3.1.2/pdf/datacleaner-reference.pdf>.
- eHow. 2016.** <http://www.ehowenespanol.com>. <http://www.ehowenespanol.com>. [En línea] 2016. [Citado el: 15 de Mayo de 2016.] [http://www.ehowenespanol.com/diferencia-reportes-analiticos-operativos-info\\_204841/](http://www.ehowenespanol.com/diferencia-reportes-analiticos-operativos-info_204841/).
- García Sérven, José R. 1992.** *Indicadores de gestión para establecimientos de atención médica*. Caracas : Disinlimed, 1992.
- Gonzalez Pompa, Yisel de los Angeles y Rosales Gonzalez, Maria Teresa. 2013.** E-gov. *E-gov*. [En línea] 2013. [http://www.egov.ufsc.br/portal/sites/default/files/mercado\\_de\\_datos\\_cooperacion\\_internacional.pdf](http://www.egov.ufsc.br/portal/sites/default/files/mercado_de_datos_cooperacion_internacional.pdf).
- Group, Global Development. 2015.** PostgreSQL. *PostgreSQL*. [En línea] 26 de 11 de 2015. <http://www.postgresql.org>.
- Hernado, Roberto. 2008.** sitio web de Procastrina. [En línea] 2008. <http://www.rhernando.net/>.
- Hernández, Yanisbel González. 2013.** *METODOLOGÍA DE DESARROLLO PARA PROYECTOS DE ALMACENES DE DATOS*. 2013. pág. 85.
- Inmon, W.H. 2005.** *Building the Data Warehouse*. s.l. : Wiley Publishing, 2005.

- Instituto Tecnológico de Piedras Negras. 2016.** <http://www.itpn.mx/>. <http://www.itpn.mx/>. [En línea] 9 de Mayo de 2016. <http://www.itpn.mx/recursositcs/4semestre/tallerdebasededatos/Unidad%20II.pdf>.
- Javlin Data Solution. 2015.** <http://www.dataintegration.info>. [En línea] 2015. [Citado el: 6 de Mayo de 2016.] <http://www.dataintegration.info/etl>.
- Kimball Group. 2016.** <http://www.kimballgroup.com>. <http://www.kimballgroup.com>. [En línea] 6 de Mayo de 2016. <http://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/kimball-data-warehouse-bus-architecture/>.
- Kimball, Ralph. 2009.** *The Data Warehouse Toolkit*. 2009.
- Kimball, Ralph y Ross, Margv. 2002.** *The Data Warehouse Toolkit*. EUA : Wiley Publishing Inc, 2002.
- Kimball, Ralph, y otros. 2008.** *The Data Warehouse Lifecycle Toolkit*. Indiana : Wiley Publishing, Inc, 2008.
- Larman, Craig. 2003.** *UML y patrones*. Madrid : Pearson, 2003.
- Microsoft. 2016.** <https://msdn.microsoft.com/>. <https://msdn.microsoft.com/>. [En línea] 6 de Mayo de 2016. <https://msdn.microsoft.com/es-es/library/aa577691.aspx>.
- . 2016.** <https://support.office.com>. <https://support.office.com>. [En línea] 2016. [Citado el: 16 de Noviembre de 2016.] [https://support.office.com/es-es/article/Informaci%C3%B3n-general-sobre-el-procesamiento-anal%C3%ADtico-en-l%C3%ADnea-OLAP-15d2cdde-f70b-4277-b009-ed732b75fdd6#bmwhat\\_is\\_on\\_line\\_analytical\\_processing](https://support.office.com/es-es/article/Informaci%C3%B3n-general-sobre-el-procesamiento-anal%C3%ADtico-en-l%C3%ADnea-OLAP-15d2cdde-f70b-4277-b009-ed732b75fdd6#bmwhat_is_on_line_analytical_processing).
- Ortiz, Marta Cecilia. 2007.** *La inteligencia de negocios aplicada a las organizaciones en Latinoamérica*. 2007.
- 2015.** Pentaho. [En línea] 26 de 11 de 2015. <http://www.pentaho.com/>.
- PentahoCorporation. 2005.** Pentaho BI Suite Enterprise Edition. [En línea] 2005. [www.pentaho.com](http://www.pentaho.com).
- 2015.** PgAdmin. [En línea] 26 de 11 de 2015. <http://www.pgadmin.org/index.php>.
- Pressman, Roger S. 2002.** *Ingeniería del software: un enfoque práctico*. Madrid : McGraw Hill, 2002.
- Reategui, Franco Ushiñahua. 2016.** <http://es.scribd.com>. <http://es.scribd.com>. [En línea] 6 de Mayo de 2016. <https://es.scribd.com/doc/55234594/Data-Warehousing-Hefesto>.
- Reyes, Mario Roberto y Rosales, Pablo Augusto. 2007.** *DESARROLLO DE UN DATAMART DE INFORMACIÓN ACADÉMICA DE ESTUDIANTES DE LA ESCUELA DE CIENCIAS*. Guatemala : s.n., 2007.
- Rojas Ricardo, Edgar y Pedraza González, Yaneisy. 2012.** [publicaciones.uci.cu](http://publicaciones.uci.cu). [publicaciones.uci.cu](http://publicaciones.uci.cu). [En línea] 15 de Diciembre de 2012. [Citado el: 23 de Mayo de 2016.] <http://publicaciones.uci.cu/index.php/SC/article/viewFile/1037/602>. 2343.
- Schupmann, Viv. 2002.** *Oracle9i Data Warehousing Guide, Release 2 (9.2)*. 2002.

**Sinnexus. 2016.** <http://www.sinnexus.com/>. *http://www.sinnexus.com/*. [En línea] 2016. [Citado el: 2016 de Diciembre de 4.] [http://www.sinnexus.com/business\\_intelligence/olap\\_avanzado.aspx](http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx).

**Soft112. 2015.** Soft112. [En línea] 26 de 11 de 2015. <http://visual-paradigm-for-uml-standard.soft112.com/>.

**Sommerville, Ian. 2011.** *Ingeniería de Software*. México : Pearson Education, Inc, 2011.

**UCI. 2010.** *Diseño de un Datawarehouse para los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular*. 2010.

**Universidad de Catarina. 2016.** <http://catarina.udlap.mx/>. *http://catarina.udlap.mx/*. [En línea] 6 de Mayo de 2016. [http://catarina.udlap.mx/u\\_dl\\_a/tales/documentos/lis/perez\\_p\\_a/capitulo2.pdf](http://catarina.udlap.mx/u_dl_a/tales/documentos/lis/perez_p_a/capitulo2.pdf).

**Universidad de Valencia. 2016.** <http://robotica.uv.es>. *http://robotica.uv.es*. [En línea] 6 de Mayo de 2016. <http://robotica.uv.es/pub/Libro/PDFs/CAPI5.pdf>.

## Bibliografía

**Instituto Nacional de Tecnologías Educativas y de Formación del Profesorado. 2016.** <http://www.ite.educacion.es/>. <http://www.ite.educacion.es/>. [En línea] 15 de Mayo de 2016. [http://www.ite.educacion.es/formacion/materiales/93/cd/m2\\_1/index.html](http://www.ite.educacion.es/formacion/materiales/93/cd/m2_1/index.html).

**Bernabeu, Ricardo. 2010.** *Investigación y Sistematización de Conceptos*. 2010.

**Bialogics. 2015.** Bialogics. [En línea] 27 de Noviembre de 2015. <http://www.bialogics.com/new-page/>.

**BUYTO. 2016.** <http://www.buyto.es>. <http://www.buyto.es>. [En línea] 6 de Mayo de 2016. <http://www.buyto.es/general-business-intelligence/datamining-transformacion-de-datos-en-business-intelligence>.

**Celma, Matilde. 2002.** *Almacenes de datos*. 2002.

**Ceria, Santiago. 2016.** <http://www-2.dc.uba.ar/>. <http://www-2.dc.uba.ar/>. [En línea] 6 de Mayo de 2016. [http://www-2.dc.uba.ar/materias/isoft1/2001\\_2/apuntes/CasosDeUso.pdf](http://www-2.dc.uba.ar/materias/isoft1/2001_2/apuntes/CasosDeUso.pdf).

**Dario, Ing. Bernabeu Ricardo. 2010.** *HEFESTO: Metodología para la Construcción de un Data Warehouse*. Córdoba : s.n., 2010.

**Data Cleaner Team. 2016.** <http://datacleaner.org/>. <http://datacleaner.org/>. [En línea] 6 de Mayo de 2016. <https://datacleaner.org/resources/docs/3.1.2/pdf/datacleaner-reference.pdf>.

**eHow. 2016.** <http://www.ehowenespanol.com>. <http://www.ehowenespanol.com>. [En línea] 2016. [Citado el: 15 de Mayo de 2016.] [http://www.ehowenespanol.com/diferencia-reportes-analiticos-operativos-info\\_204841/](http://www.ehowenespanol.com/diferencia-reportes-analiticos-operativos-info_204841/).

**García Sérven, José R. 1992.** *Indicadores de gestión para establecimientos de atención médica*. Caracas : Disinlimed, 1992.

**Gonzalez Pompa, Yisel de los Angeles y Rosales Gonzalez, Maria Teresa. 2013.** E-gov. *E-gov*. [En línea] 2013. [http://www.egov.ufsc.br/portal/sites/default/files/mercado\\_de\\_datos\\_cooperacion\\_internacional.pdf](http://www.egov.ufsc.br/portal/sites/default/files/mercado_de_datos_cooperacion_internacional.pdf).

**Group, Global Development. 2015.** PostgreSQL. *PostgreSQL*. [En línea] 26 de 11 de 2015. <http://www.postgresql.org>.

**Hernado, Roberto. 2008.** sitio web de Procastrina. [En línea] 2008. <http://www.rhernando.net/>.

**Hernández, Yanisbel González. 2013.** *METODOLOGÍA DE DESARROLLO PARA PROYECTOS DE ALMACENES DE DATOS*. 2013. pág. 85.

**Inmon, W.H. 2005.** *Building the Data Warehouse*. s.l. : Wiley Publishing, 2005.

- Instituto Tecnológico de Piedras Negras. 2016.** <http://www.itpn.mx/>. <http://www.itpn.mx/>. [En línea] 9 de Mayo de 2016. <http://www.itpn.mx/recursositcs/4semestre/tallerdebasededatos/Unidad%20II.pdf>.
- Javlin Data Solution. 2015.** <http://www.dataintegration.info>. [En línea] 2015. [Citado el: 6 de Mayo de 2016.] <http://www.dataintegration.info/etl>.
- Kimball Group. 2016.** <http://www.kimballgroup.com>. <http://www.kimballgroup.com>. [En línea] 6 de Mayo de 2016. <http://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/kimball-data-warehouse-bus-architecture/>.
- Kimball, Ralph. 2009.** *The Data Warehouse Toolkit*. 2009.
- Kimball, Ralph y Ross, Margv. 2002.** *The Data Warehouse Toolkit*. EUA : Wiley Publishing Inc, 2002.
- Kimball, Ralph, y otros. 2008.** *The Data Warehouse Lifecycle Toolkit*. Indiana : Wiley Publishing, Inc, 2008.
- Larman, Craig. 2003.** *UML y patrones*. Madrid : Pearson, 2003.
- Microsoft. 2016.** <https://msdn.microsoft.com/>. <https://msdn.microsoft.com/>. [En línea] 6 de Mayo de 2016. <https://msdn.microsoft.com/es-es/library/aa577691.aspx>.
- . 2016.** <https://support.office.com>. <https://support.office.com>. [En línea] 2016. [Citado el: 16 de Noviembre de 2016.] [https://support.office.com/es-es/article/Informaci%C3%B3n-general-sobre-el-procesamiento-anal%C3%ADtico-en-l%C3%ADnea-OLAP-15d2cdde-f70b-4277-b009-ed732b75fdd6#bmwhat\\_is\\_on\\_line\\_analytical\\_processing](https://support.office.com/es-es/article/Informaci%C3%B3n-general-sobre-el-procesamiento-anal%C3%ADtico-en-l%C3%ADnea-OLAP-15d2cdde-f70b-4277-b009-ed732b75fdd6#bmwhat_is_on_line_analytical_processing).
- Ortiz, Marta Cecilia. 2007.** *La inteligencia de negocios aplicada a las organizaciones en Latinoamérica*. 2007.
- 2015.** Pentaho. [En línea] 26 de 11 de 2015. <http://www.pentaho.com/>.
- PentahoCorporation. 2005.** Pentaho BI Suite Enterprise Edition. [En línea] 2005. [www.pentaho.com](http://www.pentaho.com).
- 2015.** PgAdmin. [En línea] 26 de 11 de 2015. <http://www.pgadmin.org/index.php>.
- Pressman, Roger S. 2002.** *Ingeniería del software: un enfoque práctico*. Madrid : McGraw Hill, 2002.
- Reategui, Franco Ushiñahua. 2016.** <http://es.scribd.com>. <http://es.scribd.com>. [En línea] 6 de Mayo de 2016. <https://es.scribd.com/doc/55234594/Data-Warehousing-Hefesto>.
- Reyes, Mario Roberto y Rosales, Pablo Augusto. 2007.** *DESARROLLO DE UN DATAMART DE INFORMACIÓN ACADÉMICA DE ESTUDIANTES DE LA ESCUELA DE CIENCIAS*. Guatemala : s.n., 2007.
- Rojas Ricardo, Edgar y Pedraza González, Yaneisy. 2012.** [publicaciones.uci.cu](http://publicaciones.uci.cu). [publicaciones.uci.cu](http://publicaciones.uci.cu). [En línea] 15 de Diciembre de 2012. [Citado el: 23 de Mayo de 2016.] <http://publicaciones.uci.cu/index.php/SC/article/viewFile/1037/602>. 2343.
- Schupmann, Viv. 2002.** *Oracle9i Data Warehousing Guide, Release 2 (9.2)*. 2002.

**Sinnexus. 2016.** <http://www.sinnexus.com/>. *http://www.sinnexus.com/*. [En línea] 2016. [Citado el: 2016 de Diciembre de 4.] [http://www.sinnexus.com/business\\_intelligence/olap\\_avanzado.aspx](http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx).

**Soft112. 2015.** Soft112. [En línea] 26 de 11 de 2015. <http://visual-paradigm-for-uml-standard.soft112.com/>.

**Sommerville, Ian. 2011.** *Ingeniería de Software*. México : Pearson Education, Inc, 2011.

**UCI. 2010.** *Diseño de un Datawarehouse para los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular*. 2010.

**Universidad de Catarina. 2016.** <http://catarina.udlap.mx/>. *http://catarina.udlap.mx/*. [En línea] 6 de Mayo de 2016. [http://catarina.udlap.mx/u\\_dl\\_a/tales/documentos/lis/perez\\_p\\_a/capitulo2.pdf](http://catarina.udlap.mx/u_dl_a/tales/documentos/lis/perez_p_a/capitulo2.pdf).

**Universidad de Valencia. 2016.** <http://robotica.uv.es>. *http://robotica.uv.es*. [En línea] 6 de Mayo de 2016. <http://robotica.uv.es/pub/Libro/PDFs/CAPI5.pdf>.

## Glosario de Términos

**HTTP:** siglas de *Hyper Text Transfer Protocol* (protocolo de transferencia de hipertexto). Se utiliza para la comunicación *web*, permitiendo que los usuarios finales accedan a la aplicación e interactúen con la misma.

**UML:** lenguaje Unificado de Modelado (UML, por sus siglas en inglés, *Unified Modeling Language*). Es un lenguaje gráfico para visualizar, especificar, construir y documentar un sistema.

**Cubos:** son subconjuntos de datos de un almacén de datos, organizado y sumariado dentro de una estructura multidimensional.

**SQL:** siglas de *Structured Query Language* (lenguaje estructurado de consulta). Es un lenguaje de alto nivel, no procedural, normalizado, que permite la consulta y actualización de los datos de bases de datos relacionales.

**Dashboard:** esta herramienta **permite visualizar el problema y favorecer la toma de decisiones** orientada a mejorar los posibles errores que podamos estar cometiendo.

**GB:** gigabyte, es una unidad de almacenamiento de información.

**RAM:** memoria de acceso aleatorio (*Random Access Memory*, RAM)

**Java:** es un lenguaje de programación de propósito general, concurrente y orientado a objetos.

**Java EE:** Enterprise Edition o Java EE, es una plataforma de programación, parte de la plataforma java para desarrollar y ejecutar software de aplicaciones en el lenguaje de programación Java.

## Anexos

### Anexo 1

Entrevistas realizadas a los trabajadores del centro de CESIM

1- ¿Cuáles son los objetivos y funciones de los centros de salud? ¿Qué situación existe en la actualidad para el análisis de la información?

Para el área a analizar:

2- ¿Qué área de análisis es de prioridad para los centros de salud? ¿Cuáles son los procesos que se manejan en esta área de acuerdo al sistema de gestión hospitalaria XAVIA HIS?

3- ¿Qué tipos de reportes se obtienen del sistema?

4- ¿Con que frecuencia se obtiene la información?

5- ¿Cantidad de información que se maneja en el sistema?

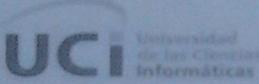
6- ¿Que indicadores y nomencladores se tienen en cuenta para el análisis de la información?

7- ¿Cuáles son los reportes más solicitados por los especialistas?

8- ¿Qué tipos de análisis les gustaría realizar sobre los indicadores mencionados anteriormente?

9- ¿Cómo les sería más fácil la presentación de la información en pantalla?

## Anexo 2


**Mercado de Datos Morbilidad y Mortalidad para las Instituciones de Salud**

**ACTA DE ACEPTACIÓN**

Entre el Centro de Informática Médica, representado en este acto por el ciudadano cubano Alain Ramos Medina, mayor de edad, en su condición de Jefe de Proyecto del Sistema de Información Hospitalaria XAVIA HIS, facultado para este acto, que en lo sucesivo se denominará la "PARTE CLIENTE", por una parte; y por la otra el equipo desarrollador del Mercado de Datos Morbilidad y Mortalidad para las Instituciones de Salud, representado en este acto por los ciudadanos Juan Pablo Caballero Romero y Eduardo Machado Soler, ciudadanos cubanos, mayores de edad, quienes actúan en su condición de Tesistas, suficientemente facultados para este acto, que en lo sucesivo se denominará la "PARTE PROVEEDORA"; acuerdan expresamente que:

La Parte Cliente, luego de concluir el proyecto Mercado de Datos Morbilidad y Mortalidad para las Instituciones de Salud, determina que el mismo se efectuó satisfactoriamente.

Y para que así conste se suscribe la presente acta en La Habana a los 6 días del mes de junio de 2016.

Entrega:	Recibe:
Juan Pablo Caballero Romero 	Alain Ramos Medina
Eduardo Machado Soler 	
Cargo: Tesistas	Cargo: Jefe de Proyecto

Figura 19 Carta de aceptación