



**Universidad de las Ciencias Informáticas**

**Facultad 7**

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

**Desarrollo de un Data Mart para la toma de decisiones referente a  
las Reacciones Adversas a Medicamentos, en el Ministerio de Salud  
Pública, desde el producto Synta**

**Autores:** Aliandy Sosa Bello

Dennier Salas Lóriga

**Tutores:** MSc. Annia Arencibia Morales

Ing. Darien Castellano Pérez

**Cotutor:** Ing. Guillermo Suárez Lorenzo

**La Habana, junio de 2013**

**“Año 55 de la Revolución”**



*"Los estudiantes son en su mayoría revolucionarios. Revolucionarios por naturaleza, porque pertenecen a ese estrato de jóvenes que se abren a la vida y que adquieren todos los días conocimientos nuevos"*

*Ernesto Che Guevara*

### Datos de Contacto

MSc. Annia Arencibia Morales (aarencibia@uci.cu): graduada de Ingeniería en Ciencias Informáticas en la Universidad de las Ciencias Informáticas en el año 2007. Pertenece al Centro de Informática Médica (CESIM), actualmente posee la categoría docente de profesor Asistente y es Máster en Informática Aplicada. Profesora de Metodología de la Investigación Científica e imparte Postgrados de Ingeniería de Software. Pertenece al Departamento de Sistemas de Apoyo a la Salud.

Ing. Darien Castellano Pérez (dcperez@uci.cu): graduado de Ingeniería en Ciencias Informáticas en la Universidad de las Ciencias Informáticas en el año 2009. Es profesor de Programación III en la Facultad 7 y trabaja como adjunto al Centro de Informática Médica (CESIM), específicamente en el Departamento de Sistemas de Apoyo a la Salud.

Ing. Guillermo Suárez Lorenzo (glorenzo@uci.cu): graduado de Ingeniería en Ciencias Informáticas en la Universidad de las Ciencias Informáticas en el año 2011. Pertenece al Centro de Informática Médica (CESIM), específicamente en el Departamento de Sistemas de Apoyo a la Salud.

### Agradecimientos

*A mi madre por ser la persona más importante en mi vida; por su cariño, amor, dedicación y apoyo incondicional durante todas las etapas de mi vida. Gracias a ti hoy estoy aquí.*

*A mis hermanos, que ingenuamente han sido mis principales impulsores, han hecho que me esfuerce cada día más para poder ser un ejemplo a seguir.*

*A mis abuelos por sus consejos y por todos los valores que han inculcado en mí; doy gracias por poder contar con su compañía.*

*A mi novia por todo su apoyo, cariño, amor, comprensión y compañía en todo momento. Gracias por existir.*

*A mis suegros por todo su apoyo y por tratarme como un hijo.*

*A mi familia por todo su apoyo y preocupación durante toda mi carrera.*

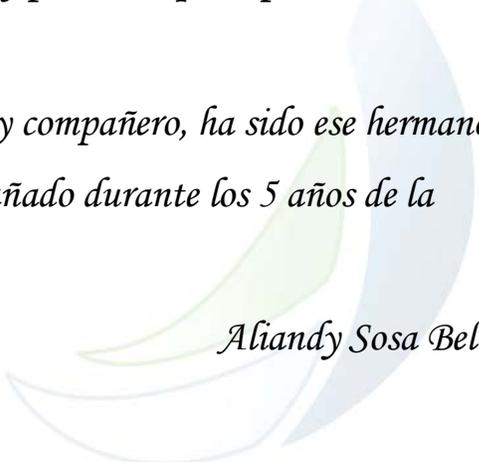
*A mi compañero de tesis por su comprensión y apoyo en cada paso que dimos durante la confección de esta tesis.*

*A mis tutores por todo el tiempo dedicado durante el desarrollo de esta tesis.*

*A Guillermo por todas esas noches interminables de trabajo y por tener que soportarme, con incondicional paciencia, durante el desarrollo de esta tesis.*

*A Vladimir por todo el tiempo compartido. Más que amigo y compañero, ha sido ese hermano que me ha apoyado, aconsejado, guiado y muchas veces regañado durante los 5 años de la carrera.*

*Aliandy Sosa Bello*



*A mi madre por ser mi amiga, por hacer de mí una mejor persona a través de sus consejos, por su enseñanza y amor, por ser una madre excepcional.*

*A mi padre por brindarme los recursos necesarios y estar a mi lado apoyándome y aconsejándome siempre.*

*A mi padrasto por ser un padre para mí, por apoyarme y confiar en mí como persona.*

*A mis hermanos Dariel, Danel y Tahimí por estar siempre presentes, por seguirme y apoyarme.*

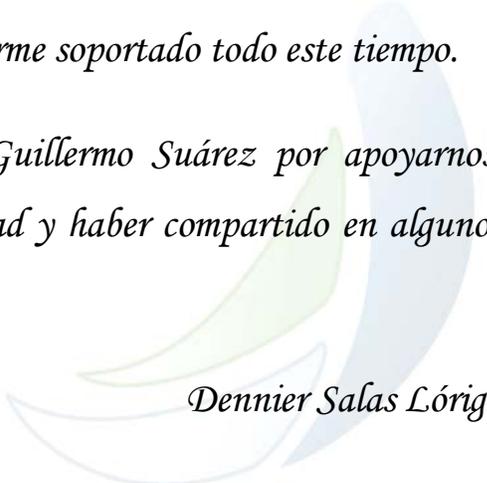
*A mi novia Lisandra por su cariño y amor, por apoyarme, por estar siempre ahí en todo momento.*

*A Luis, Meybel, Alejandro, Rogeidis, Victor, Chuchy, Dainier, Reinaldo que más que amigos han sido hermanos, me siento muy orgulloso de haberlos conocidos y haber compartido juntos momentos buenos y malos.*

*A mi compañero de tesis por haber sido tan paciente y haberme soportado todo este tiempo.*

*A mis tutores Annia Arencibia, Darien Castellano y Guillermo Suárez por apoyarnos, guiarnos y confiar en nosotros, además de tener su amistad y haber compartido en algunos momentos.*

*Dennier Salas Lóriga*



**Dedicatoria**

*Esta tesis está dedicada a toda mi familia;  
a toda esa gente que siempre confió en mí, especialmente  
a mi mamá y hermanos.*

*Aliandy Sosa Bello*

*La presente tesis se la dedico a mi querida mamá,  
a mi papá, a mis hermanos, a mi padrasto, a mi querida novia,  
a toda mi familia, a mis amigos y a todos los que de una forma u otra  
han contribuido a mi formación como profesional y como persona.*

*Dennier Salas Lóriga*



### Resumen

Existe un gran volumen de información informatizada en todos los sectores, tanto económicos como sociales, donde se destaca por su vital importancia el sector de la Salud. En Cuba se registra un cúmulo elevado de información referente a las Reacciones Adversas a Medicamentos (RAM) ocurridas en el país, haciendo uso del Módulo de Reacciones Adversas a Medicamentos (MRAM) del producto Sistema para el Control Farmacológico (Synta). Este sistema le brinda al farmacoepidemiólogo un conjunto de reportes específicos, pero no le permite efectuar consultas especializadas respecto a las RAM, no pudiendo realizar de esta forma una buena toma de decisiones.

Para dar solución al problema planteado se definió crear un Data Mart (DM), sistema pensado para contribuir en el proceso de toma de decisiones en un área determinada, haciendo uso de la metodología Hefesto y la herramienta STPivot, para mostrar la información; así como Pentaho Data Integration para la extracción, transformación y carga de la información de la base de datos del MRAM hacia el DM.

Una vez creado el DM, es incorporado al MRAM del producto Synta, permitiendo a los farmacoepidemiólogos contar con una sola herramienta, así como realizar análisis de la información haciendo uso de reportes. Estos reportes facilitan una mejor comprensión y análisis de los datos (mediante gráficas y tablas) a la hora de tomar decisiones.

**Palabras clave:** *Data Mart, Reacciones Adversas a Medicamentos, toma de decisiones*



**Índice**

Introducción .....5

Capítulo 1: Fundamentación teórica de la investigación.....11

1.1 Introducción a Data Warehouse .....11

1.2 Data Mart .....12

1.3 ¿Por qué usar un Data Mart y no un Data Warehouse? .....13

1.4 Extracción, Transformación y Carga.....14

1.5 Sistemas de Procesamiento Transaccional en Línea y Procesamiento Analítico en Línea .....14

    1.5.1 Procesamiento Analítico en Línea .....14

1.6 Modelado dimensional .....15

    1.6.1 Esquema en Estrella .....16

    1.6.2 Esquema Copo de Nieve.....17

    1.6.3 Esquema de Constelación de Hechos.....18

1.7 Metodologías para el diseño e implementación de un Data Mart.....19

    1.7.1 Kimball .....19

    1.7.2 Inmon .....19

    1.7.3 Hefesto 2.0.....20

1.8 Sistemas existentes .....22

    1.8.1 Nivel internacional .....23

    1.8.2 Nivel nacional.....23

1.9 Herramientas para la construcción del Data Mart .....24

    1.9.1 Sistemas Gestores de Base de Datos .....24

    1.9.2 Herramientas de Integración de Datos .....25

    1.9.3 Herramientas para el Proceso Analítico en Línea.....27

    1.9.4 Herramienta de modelado .....29

1.10 Servidores de Aplicaciones .....30

Capítulo 2: Propuesta de diseño e implementación del Data Mart .....32

2.1 Propuesta del sistema .....32

2.2 Fases de la metodología Hefesto .....32

2.2.1	Fase 1: Análisis de los requerimientos .....	32
2.2.2	Fase 2: Análisis de los OLTP .....	36
2.2.3	Fase 3: Modelo lógico del Data Mart .....	42
2.2.4	Fase 4: Integración de datos .....	46
Capítulo 3: Proceso Analítico en Línea del Data Mart .....		48
3.1	Proceso de análisis de la información del Data Mart .....	48
3.1.1	Determinar jerarquías.....	48
3.1.2	Medidas .....	49
3.1.3	Consultas multidimensionales .....	50
3.1.4	Resultados del cliente STPivot.....	50
3.1.5	Integración del Data Mart con el Módulo de Reacciones Adversas a Medicamentos .....	54
Conclusiones .....		56
Recomendaciones .....		57
Referencias Bibliográficas.....		58
Bibliografía.....		64
Anexos.....		70



## Índice de Figuras

Figura 1. Data Mart dependiente.....	13
Figura 2. Data Mart independiente.....	13
Figura 3. Modelo dimensional.....	16
Figura 4. Esquema Estrella.....	17
Figura 5. Esquema Copo de Nieve.....	18
Figura 6. Esquema Constelación de Hechos.....	18
Figura 7. Metodología Hefesto.....	21
Figura 8. Modelo conceptual.....	36
Figura 9. Modelo conceptual ampliado.....	42
Figura 10. Dimensión RAM.....	43
Figura 11. Diseño de la tabla de hechos.....	44
Figura 12. Uniones de las dimensiones.....	45
Figura 13. Transformación RAM.....	46
Figura 14. Ejecución del trabajo.....	47
Figura 15. Jerarquía de los atributos.....	48
Figura 16. Medida total de registros.....	49
Figura 17. Medida calculable porcentaje de pacientes.....	49
Figura 18. Reporte de la dimensión medicamentos.....	51
Figura 19. Información filtrada.....	52
Figura 20. Información en gráfica de barras.....	53
Figura 21. Información en gráficas de pastel.....	53
Figura 22. Reporte del STPivot desde el MRAM.....	55



### Introducción

El desarrollo de la informática y las comunicaciones ha producido grandes avances tecnológicos, provocando a nivel mundial un aumento considerable de la información; nombrándose la época actual como “Era de la Información”. La confluencia de una multitud de tendencias durante la última década como: redes sociales, aplicaciones informáticas, caída del coste de la banda ancha, interconexión de objetos a través de la red y computación en la nube, ha tendido a que se produzcan grandes cantidades de datos que necesiten ser captados y procesados para un mejor análisis. Todo ese volumen de datos es almacenado, principalmente para utilizarlo en dependencia de las necesidades sociales y para extraer la información útil en la toma de decisiones. (Rizo, et al., 2008 p. 2)

El volumen y variedad de la información que se encuentra informatizada en bases de datos digitales, ha crecido considerablemente en la última década. El área de la extracción de conocimiento de base de datos, ha adquirido recientemente una importancia científica y económica inusual. Estos cambios en el almacenamiento de la información traen consigo que las formas de extraer el conocimiento también varíen, siendo más eficientes, con un mayor entorno de descubrimiento y con consultas más inductivas<sup>1</sup>. (Orallo, y otros, 2010 pág. 4)

Es imprescindible convertir la información existente en experiencia, conocimiento y sabiduría, para posibilitar una buena toma de decisiones, especialmente en grandes organizaciones y proyectos científicos que tributan a la salud. La búsqueda de información relevante siempre es útil en: la administración empresarial, el control de la producción, el análisis de los mercados, el diseño en ingeniería y la exploración científica, porque pueden ofrecer las respuestas más apropiadas a las necesidades de información. (Rodríguez, et al., 2002)

El análisis de la información se ve afectado por la existencia de grandes volúmenes de datos; este fenómeno no solo ocurre a nivel internacional. En Cuba existen innumerables sistemas que necesitan de una correcta manipulación de los datos. Con la creación del Ministerio de la Informática y las Comunicaciones (MIC), se han consolidado los esfuerzos en función de alcanzar un nivel adecuado de

---

<sup>1</sup> Inductivas: se aplica al razonamiento que a partir de una serie de hechos particulares establece una ley general.

informatización de la sociedad, como uno de los objetivos estratégicos fundamentales para el desarrollo del país. La informatización de la sociedad cubana, se considera como el proceso de utilización ordenada y masiva de las Tecnologías de la Información y las Comunicaciones (TIC), para satisfacer las necesidades de información y conocimiento en todas las esferas de la sociedad como: la educación, la cultura y la salud. (Ferrera, et al., 2010)

En Cuba, el Sistema Nacional de Salud (SNS) debe contar con información confiable, consistente y oportuna para la toma de decisiones y el mejoramiento de los procesos médicos asistenciales, para así garantizar el incremento en la calidad y seguridad de la atención médica a la población. Durante los últimos años, el Ministerio de Salud Pública (MINSAP) y la Universidad de las Ciencias Informáticas (UCI), han venido desarrollando sistemas encaminados a lograr la informatización de los procesos en el sector de la salud, poniendo como centro al paciente y utilizando las TIC como factor esencial en el desarrollo de estos sistemas. (Ramos, et al., 2009 p. 1)

Para el desarrollo de sistemas, que tienen como objetivo informatizar los procesos del sector de la salud, la UCI cuenta con el Centro de Informática Médica (CESIM), perteneciente a la Facultad 7. Este contiene varios departamentos, dentro de ellos el Departamento de Sistemas de Apoyo a la Salud (SAS). En dicho departamento se desarrolla el Sistema para el Control Farmacológico (Synta), creado con el objetivo de informatizar el Departamento de Farmacoepidemiología (del MINSAP): órgano rector de los centros provinciales y municipales encargados de implementar la estrategia de Farmacovigilancia<sup>2</sup>. El objetivo de dicho departamento, es recopilar información que conduzca a la protección de la población y mantener la seguridad y eficacia en el consumo de los medicamentos. Synta cuenta con seis módulos: Módulo Control de Recetas Médicas, Módulo Consumo de Medicamentos, Nomencladores, Módulo Tarjeta de Control, Administración y el Módulo de Reacciones Adversas a Medicamentos (MRAM).

El MRAM del producto Synta, permite registrar la información de las Reacciones Adversas a Medicamentos (RAM)<sup>3</sup> ocurridas en Cuba, llevando el control de los medicamentos que más inciden negativamente en la población. La información proporcionada es recolectada, investigada y valorada por

---

<sup>2</sup> Farmacovigilancia: conjunto de actividades que tienen por objeto la identificación y la cuantificación del riesgo de los efectos indeseados producidos por los medicamentos.

<sup>3</sup> RAM: reacciones perjudiciales e inesperadas, a medicamentos administrados a las dosis habituales con fines terapéuticos. (OMS, 2008)

los farmacoepidemiólogos (en todos los niveles de dirección del MINSAP). El módulo, dentro de sus funcionalidades permite: la creación de reportes, gráficas de comportamiento de un medicamento dado y comparación entre varios de ellos, entre otras funcionalidades.

Actualmente la información es registrada por el médico o profesional sanitario que detectó la RAM, haciendo uso de una planilla (Notificación de sospecha de Reacción Adversa a Medicamentos por el profesional de salud), distribuida por el MINSAP. El farmacoepidemiólogo de cada Unidad de Salud (US) es el encargado de transcribir la información, de dichas planillas, a formato digital (*excel*). Posteriormente este documento *excel* es procesado por el MRAM. Una vez que la información se encuentra dentro del sistema, los especialistas pueden crear un conjunto de reportes, siendo estos:

- ✓ RAM por medicamento
- ✓ Medicamento por RAM
- ✓ Reporte anual de las RAM (grupos etarios<sup>4</sup>, medicamentos nocivos, grupos de medicamento ATC<sup>5</sup>, entre otros)

A pesar que el módulo le proporciona a los especialistas esta información. El mismo no brinda el soporte necesario para generar un grupo de reportes de vital importancia, para realizar un análisis detallado y preciso del comportamiento de las RAM en un momento determinado, ya sea por valores específicos o por la unión de varias características de estas al mismo tiempo. Algunos de los reportes que los farmacoepidemiólogos necesitan, son:

- ✓ Analizar cómo afecta un medicamento determinado a los tres grupos etarios definidos en Cuba (niños: de cero a quince años, adultos: de dieciséis a cincuenta y nueve y geriátricos: de sesenta en adelante). Una vez identificados los medicamentos, se le debe dar seguimiento, según las características específicas de cada grupo de edad (los niños y los geriátricos tienen prioridad por ser los dos grupos más vulnerables). De esta forma se puede conocer qué RAM ha ocurrido por el consumo de estos medicamentos, para poder sacar de circulación dicho medicamento o emitir una circular para esos grupos de edades.

---

<sup>4</sup> Etario: perteneciente o relativo a la edad de una persona.

<sup>5</sup> Sistema de Clasificación Anatómica, Terapéutica, Química (ATC): índice de sustancias farmacológicas y medicamentos, organizados según grupos terapéuticos.

- ✓ Se precisa un análisis de los medicamentos que están provocando más RAM, analizando específicamente de estos el lote, permitiendo sacarlos de circulación en un tiempo determinado, para evitar que afecte demasiado a la población (este análisis actualmente se hace, pero no se pueden tomar medidas a tiempo).
- ✓ Es necesario llevar el control de las RAM ocasionadas por los medicamentos consumidos por pacientes con un Antecedente Patológico (APP) determinado (hipertensión, asma, diabetes, entre otros); identificando si algunas de estas RAM no son frecuentes, para realizar un análisis de inmediato.
- ✓ Las embarazadas, puérperas<sup>6</sup> o lactantes, deben tener un seguimiento especial respecto a las RAM ocurridas durante estas etapas (algunas pueden ocurrir por el consumo inadecuado de un medicamento, la inadecuada vacunación en una etapa de gestación o simplemente porque un medicamento está ocasionando RAM no previstas). Estos pacientes presentan condiciones únicas, sobre las que se deben tomar medidas de forma rápida, que permitan salvaguardar tanto la vida de la madre, como la del niño.

El personal encargado de realizar estos reportes no cuenta con mucha experiencia en el análisis de dicha información (puede llegar a tener un gran volumen); trayendo consigo el agotamiento físico (el procesamiento se hace de forma manual, consultando más de un *excel* al mismo tiempo) y el error humano. Por lo tanto se puede obtener información errónea o falsas alarmas en algún caso determinado. Además se produce retraso en la toma de decisiones (estratégicas y tácticas que tiene como objetivo principal preservar la vida humana), debido a que no se obtienen los reportes en el tiempo requerido.

Por lo anteriormente planteado se identifica como **problema a resolver**: ¿Cómo facilitar la toma de decisiones referente a las Reacciones Adversas a Medicamentos, en el Ministerio de Salud Pública, desde el producto Synta? Se define como **objeto de estudio**: proceso de toma de decisiones referente a las Reacciones Adversas a Medicamentos, delimitándose como **campo de acción**: la toma de decisiones referente a las Reacciones Adversas a Medicamentos por parte de los farmacoepidemiólogos del Ministerio de Salud Pública.

---

<sup>6</sup> Puérpera: mujer que acaba de parir.

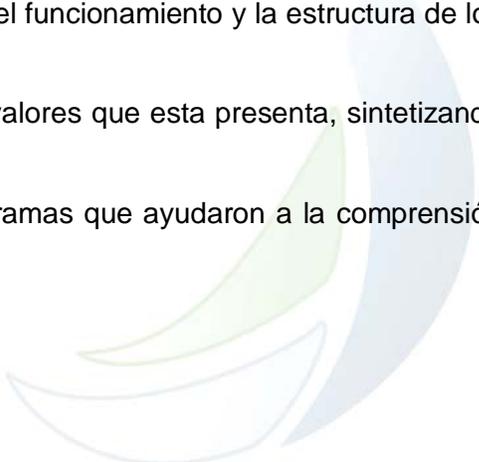
Se propone como **objetivo general**: desarrollar un Data Mart para la toma de decisiones referente a las Reacciones Adversas a Medicamentos por parte de los farmacoepidemiólogos, del Ministerio de Salud Pública, desde el Módulo de Reacciones Adversas a Medicamentos del producto Synta.

Para dar cumplimiento al objetivo general planteado, se definen las siguientes **tareas de la investigación**:

- ✓ Analizar Data Marts y Data Warehouse a nivel internacional y nacional que permitan la toma de decisiones en el sector de la salud, estableciendo similitudes y diferencias con la investigación en curso.
- ✓ Realizar entrevistas al cliente, definiendo los requisitos funcionales para la implementación del Data Mart.
- ✓ Valorar las herramientas, tecnologías y metodologías, seleccionando las necesarias para el desarrollo del Data Mart.
- ✓ Desarrollar el Data Mart teniendo en cuenta las tecnologías identificadas, para solucionar la problemática planteada.
- ✓ Integrar el Data Mart al Módulo de Reacciones Adversas a Medicamentos, siguiendo las pautas de desarrollo establecidas para el producto Synta.

Los métodos científicos utilizados en el trabajo de investigación son:

### Teóricos:

- ✓ **Histórico-Lógico**: se utilizó para realizar la valoración de Data Marts (DM) existentes, enmarcados en la toma de decisiones en el sector de la salud y analizar el funcionamiento y la estructura de los mismos.
  - ✓ **Analítico-Sintético**: se utilizó para analizar las RAM y los valores que esta presenta, sintetizando los necesarios para la creación del DM.
  - ✓ **Modelación**: se utilizó en la confección de modelos y diagramas que ayudaron a la comprensión de los procesos a desarrollados.
- 

### Empírico:

**Entrevistas:** dentro de la entrevista se utilizó la estructurada (ver Anexo 1), para determinar los requerimientos necesarios para la construcción del DM.

Una vez que se cumpla el objetivo general de la investigación, se espera obtener un DM que contenga la información de las RAM ocurridas en Cuba, que le facilite a los farmacoepidemiólogos, del MINSAP, la toma de decisiones.

El presente trabajo está estructurado en tres capítulos, distribuidos de la siguiente manera:

- ✓ **Capítulo 1: Fundamentación teórica de la investigación.** En este capítulo se realiza un análisis de DM y Data Warehouses (DWH) a nivel internacional y nacional que permiten la toma de decisiones en el sector de la salud. Además se describe la arquitectura, herramientas, metodologías y lenguajes de programación propuestos para solucionar el problema planteado.
- ✓ **Capítulo 2: Propuesta de diseño e implementación del Data Mart.** En este capítulo, se describen los pasos a seguir para el diseño del Data Mart, haciendo uso de la metodología Hefesto.
- ✓ **Capítulo 3: Proceso Analítico en Línea del Data Mart.** En este capítulo, se explica la construcción del cubo de datos, definiéndose las jerarquías y atributos de las dimensiones y la implementación de los valores numéricos (medidas). Además se visualizan los resultados obtenidos desde el MRAM, mediante la herramienta STpivot, que permite consultar y analizar la información almacenada mediante una tabla de datos y gráficas.



## Capítulo 1: Fundamentación teórica de la investigación

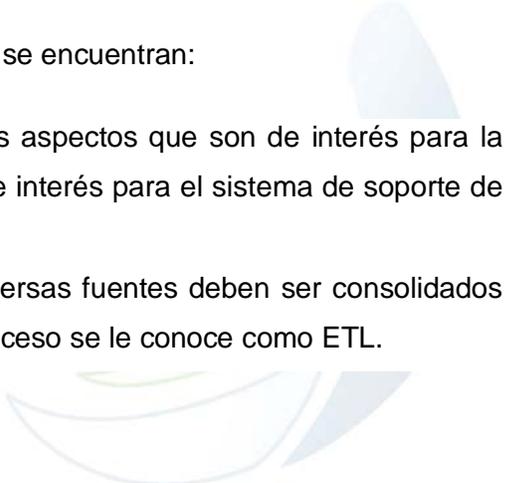
Se realizó un análisis de los DM, exponiendo su definición y características. Además se describe la arquitectura, herramientas, metodologías y lenguajes de programación utilizados para dar solución al problema planteado.

### 1.1 Introducción a Data Warehouse

Con el aumento gradual de la información en las instituciones, se hace necesaria la utilización de herramientas para la toma de decisiones en cualquier área funcional y la aplicación de técnicas estadísticas de análisis, de ahí la importancia de adentrarse en el conocimiento de los almacenes de datos o Data Warehouse.

- ✓ Según W. H. Inmon (considerado por muchos el padre del DWH): es un conjunto de datos orientados por temas, integrados, variantes en el tiempo y no volátiles, que tienen por objetivo dar soporte a la toma de decisiones. (Curto, 2006)
- ✓ Según Ralph Kimball (considerado el principal promotor del enfoque dimensional para el diseño de DWH): es una copia de los datos transaccionales específicamente estructurada para la consulta y el análisis. (Curto, 2006)
- ✓ Según Susan Osterfeldt: “Yo considero al DWH como algo que provee dos beneficios empresariales reales: Integración y Acceso de datos. El DWH elimina una gran cantidad de datos no deseados, como también el procesamiento desde el ambiente operacional clásico”. (Febles, y otros, 2011)

Dentro de las principales características de un DWH (Lazaro, 2012) se encuentran:

- ✓ Orientado al tema: la información se clasifica en base a los aspectos que son de interés para la organización. Se excluye toda la información que no será de interés para el sistema de soporte de decisiones.
  - ✓ Integrado: la integración implica que todos los datos de diversas fuentes deben ser consolidados en una instancia antes de ser agregados al DWH. A este proceso se le conoce como ETL.
- 

- ✓ De tiempo variante: los datos almacenados en la base de datos, son relativos a un período de tiempo y estos deben ser integrados periódicamente. Este sello de tiempo permite realizar pronósticos y análisis a partir de una base estadística de información.
- ✓ No volátil: los datos, una vez cargados en el DWH, no sufrirán cambios. Solamente estos datos serán consultados mediante gráficas y reportes. La información es útil para el análisis y la toma de decisiones solamente cuando es estable.

Partiendo de lo antes planteado, se resume que un DWH esencialmente posibilita la extracción de datos de sistemas operacionales y otras fuentes externas. Permite la integración y homogenización de los datos de toda una empresa, abasteciendo de información a la misma; siendo esta transformada y resumida<sup>7</sup> de forma flexible. Tienen como principal objetivo servir de ayuda en el proceso de toma de decisiones estratégicas y tácticas.

## 1.2 Data Mart

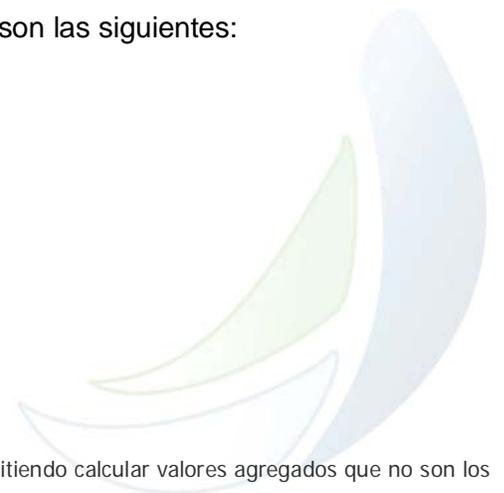
Un DM es una versión especial de un DWH; una base de datos departamental, especializada en el almacenamiento de los datos de un área del negocio específica. Se caracteriza por disponer de una estructura de datos pensada para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. Un DM puede ser alimentado desde los datos de un DWH, o integrar por sí mismo un compendio de distintas fuentes de información. En síntesis los DM son pequeños DWH centrados en un tema o un área de negocio específico dentro de una organización. (Dario, 2010 pág. 73)

Las características principales de un DM (Ramos, 2009 pp. 20-23) son las siguientes:

- ✓ Alcance departamental
- ✓ Optimizado para consultas específicas
- ✓ Altamente resumido
- ✓ Específicas funciones del negocio
- ✓ Orientada a un grupo de usuarios

---

<sup>7</sup> La sumarización o agregación, muestra los datos de una manera más resumida, permitiendo calcular valores agregados que no son los datos directos registrados, sino datos derivados de ellos.



Se definen dos tipos de Data Mart (HALLO, et al., 2006 p. 56), los dependientes y los independientes:

- ✓ Dependientes: son los que se construyen a partir de un DWH central, y reciben sus datos de un repositorio empresarial central (ver Figura 1).

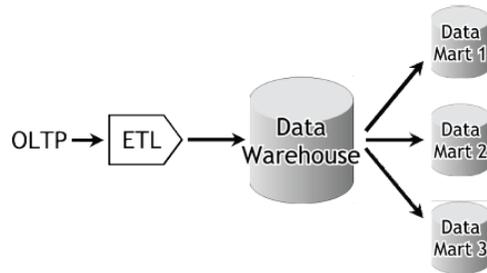


Figura 1. Data Mart dependiente (Bernabeu, 2009 pág. 60)

- ✓ Independientes: son aquellos DM que no dependen de un DWH central. Estos pueden recibir los datos directamente del ambiente operacional (ver Figura 2).

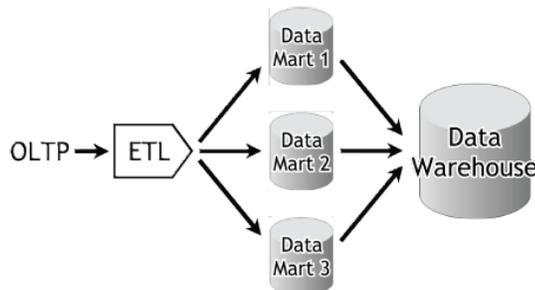


Figura 2. Data Mart independiente (Bernabeu, 2009 pág. 60)

### 1.3 ¿Por qué usar un Data Mart y no un Data Warehouse?

Un DM está pensado para cubrir las necesidades de un grupo de trabajo o de un determinado departamento dentro de la organización, en cambio, el ámbito de un DWH es la organización en su conjunto. El costo del uso de un DM es inferior al de aplicar un DWH y conllevan a un menor tiempo de construcción y puesta en marcha. La información que se desea gestionar está orientada al Departamento de Farmacoepidemiología, los medicamentos y sus respectivas reacciones adversas, por lo que se propone un DM independiente y no un DWH.

## 1.4 Extracción, Transformación y Carga

La Extracción, Transformación y Carga (ETL, por sus siglas en inglés), son tres funciones que se combinan en una sola herramienta y extraen datos de las diversas fuentes que se requieran, los transforman para resolver posibles problemas de inconsistencias entre los mismos y finalmente, después de haberlos depurado se procede a su carga en el Data Mart (Dario, 2010 pág. 21).

Las funciones específicas de ETL (Dario, 2010 pág. 27) son:

- ✓ Extracción: es el proceso de lectura de datos de los sistemas operacionales o cualquier otra fuente de información, como por ejemplo: *excel*, eXtensible Markup Language (XML) y DWH.
- ✓ Transformación: es el proceso de convertir, filtrar, validar y aplicar reglas de negocio a los datos extraídos de las fuentes de almacenamiento.
- ✓ Carga: es el proceso de escritura de los datos en un DWH, DM, XML o cualquier otra base de datos de destino.

## 1.5 Sistemas de Procesamiento Transaccional en Línea y Procesamiento Analítico en Línea

Los sistemas de bases de datos operacionales están dirigidos a realizar transacciones y a la resolución de consultas, por esta razón se denominan sistemas de Procesamiento Transaccional en Línea (OLTP, por sus siglas en inglés). Una transacción es un proceso atómico que debe ser validado o abortado, y que puede involucrar operaciones de consulta, inserción, modificación y borrado de datos. Estos sistemas cubren la mayor parte de la operativa del día a día de una empresa, como la gestión de pedidos, la realización de inventario, la gestión de la contabilidad, entre otros. Por otra parte, los DM se sitúan en los entornos decisionales de las empresas, que se encuentran totalmente separados de los sistemas operacionales y cuya finalidad es proporcionar datos para su análisis y posterior toma de decisiones; suelen denominarse sistemas de Procesamiento Analítico en Línea (OLAP, por sus siglas en inglés). Estos sistemas OLAP pueden organizar y presentar datos en el formato que mejor se ajuste a las necesidades de los usuarios. (Bartomeus, 2009)

### 1.5.1 Procesamiento Analítico en Línea

Las funcionalidades de los sistemas OLAP se caracterizan por brindar un análisis multidimensional de los datos, que soportan los análisis de los usuarios y ofrecen una navegación que permite seleccionar la

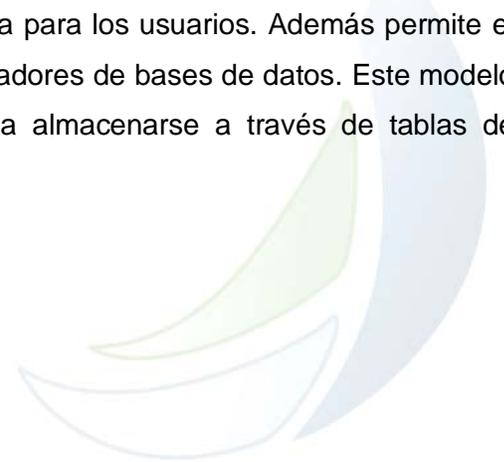
información a obtener. Existen varias arquitecturas para los sistemas OLAP como: OLAP multidimensional (MOLAP, por sus siglas en inglés) y OLAP relacionales (ROLAP, por sus siglas en inglés). (Molier, 2005 p. 173)

- ✓ **Sistemas ROLAP:** accede a los datos almacenados en el DM para proporcionar los análisis OLAP, su principal premisa es extraer la información de bases de datos relacionales (Molier, 2005 p. 174). ROLAP utiliza una arquitectura de tres niveles. La base de datos relacional maneja el almacenamiento de los datos, el motor OLAP proporciona la funcionalidad analítica y alguna herramienta especializada es empleada para el nivel de presentación. (Tamayo, et al., 2006)
- ✓ **Sistemas MOLAP:** usa bases de datos multidimensionales para proporcionar el análisis, su principal premisa es que el OLAP esta mejor implementado almacenando la información en una base de datos multidimensional (Molier, 2005 p. 175). MOLAP utiliza una arquitectura de dos niveles: la base de datos multidimensional para el manejo, acceso y obtención de los datos y el motor analítico, que es el responsable de la ejecución de las consultas OLAP y de proporcionar una interfaz a través de la cual los usuarios finales visualizan los análisis OLAP. (Tamayo, et al., 2006)

Para la construcción del DM se selecciona el sistema ROLAP, puesto que los datos se extraen de una base de datos relacional soportada por un sistema gestor de base de datos PostgreSQL.

## 1.6 Modelado dimensional

Modelado dimensional, es el nombre que recibe una técnica utilizada especialmente en DM. Presenta la información de una manera estándar, sencilla y sobre todo intuitiva para los usuarios. Además permite el acceso a la información mucho más rápido por parte de los manejadores de bases de datos. Este modelo brinda una búsqueda rápida de los datos y la información va a almacenarse a través de tablas de dimensiones y hechos. (Bernabeu, 2009 pág. 27)



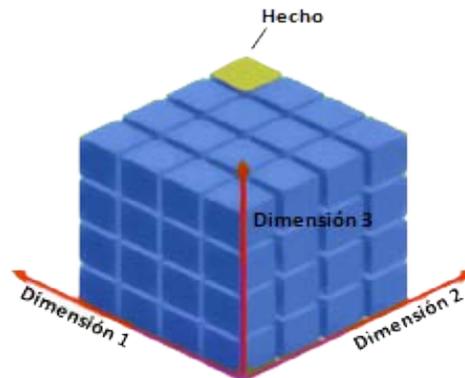


Figura 3. Modelo dimensional (Dario, 2010 pág. 36)

- ✓ **Dimensiones:** son objetos del negocio con los cuales se puede analizar la tendencia y el comportamiento del mismo. Las definiciones de las dimensiones se basan en políticas de la compañía, e indican la manera en que la organización interpreta o clasifica su información para segmentar el análisis, facilitando la observación de los datos. (Kimball, 2002)
- ✓ **Hechos:** es un concepto de interés primario para el proceso de toma de decisiones, corresponde a eventos que ocurren dinámicamente en el negocio de la empresa y contiene los hechos, indicadores o medidas del negocio que se desean analizar. (Kimball, 2002)
- ✓ **Medidas:** son características cuantitativas de los objetos que se desean analizar en las empresas. Las medidas cuantitativas están dadas por valores o cifras porcentuales, por ejemplo: las ventas en dólares, cantidad de unidades principales, cantidad de unidades de productos vendidos, entre otros. (Kimball, 2002)

Las bases de datos dimensionales implican tres variantes posibles de modelación, las cuales se mencionan a continuación:

- ✓ Esquema estrella
- ✓ Esquema copo de nieve
- ✓ Esquema constelación

## 1.6.1 Esquema en Estrella

Contiene en el medio una tabla de hechos que se encuentra conectada a una o varias tablas de dimensiones. Las tablas de dimensiones solo se relacionan con la tabla de hechos y no existen relaciones

entre dimensiones. Las tablas de dimensiones tendrán siempre una clave primaria simple, mientras que en la tabla de hechos, la clave principal estará compuesta por las claves principales de las tablas dimensionales o una propia (ver Figura 4) (Tamayo, et al., 2006). Entre sus ventajas más significativas está que es el esquema más simple de interpretar, posee los mejores tiempos de respuesta, su diseño es fácil de modificar y simplifica el análisis. Sin embargo, es el menos robusto para la carga y es el más lento de construir. (Dario, 2010 p. 38)

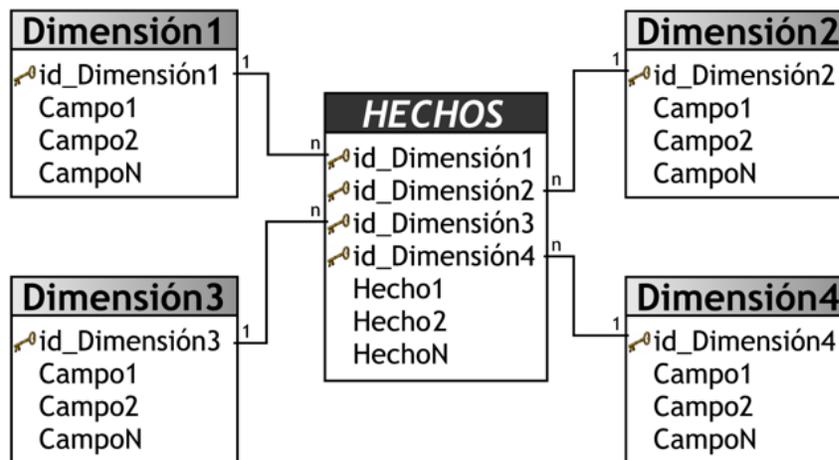


Figura 4. Esquema Estrella (Bernabeu, 2009)

## 1.6.2 Esquema Copo de Nieve

Es una extensión del esquema estrella, donde se presentan ramificaciones a partir de las tablas de dimensiones y no solo a partir de la tabla de hechos. En esta forma de esquema, las tablas de dimensión pueden tener relación con otras tablas de dimensión (Tamayo, et al., 2006). Tiene como desventaja que de existir muchas tablas de dimensiones, cada una de ellas con varias jerarquías, pueden crearse bastantes tablas llegando a ser inmanejables. Además su desempeño puede verse reducido si existen muchas uniones y relaciones entre tablas (ver Figura 5). (Dario, 2010 p. 39)

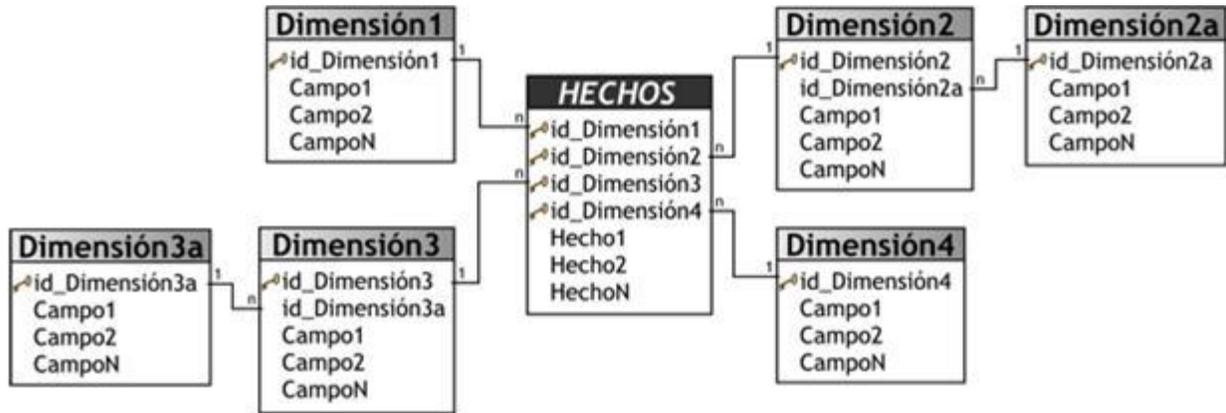


Figura 5. Esquema Copo de Nieve (Bernabeu, 2009)

### 1.6.3 Esquema de Constelación de Hechos

Este modelo está compuesto por una serie de esquemas en estrella y lo conforman una tabla de hechos principal y una o más tablas de hechos auxiliares, que están relacionadas con sus respectivas tablas de dimensiones, vinculándose las tablas de hechos auxiliares con algunas dimensiones asignadas a la tabla de hecho principal y también con nuevas tablas de dimensiones (Tamayo, et al., 2006). Este esquema se destaca, porque al tener más de una tabla de hechos facilita el análisis de más aspectos importantes del negocio y contribuye a la reutilización de las tablas de dimensiones, ya que pueden utilizarse para varias tablas de hechos (ver Figura 6). (Dario, 2010 p. 41)

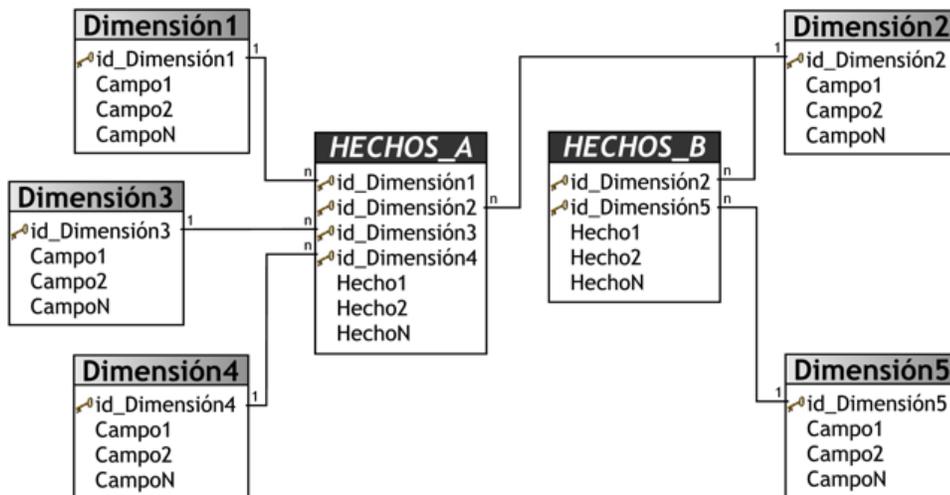


Figura 6. Esquema Constelación de Hechos (Bernabeu, 2009)

Se propone utilizar el esquema estrella, pues es ideal por su simplicidad y velocidad para ser usado en análisis multidimensionales. Además, el diseño de esquema en estrella, permite implementar la funcionalidad de una base de datos multidimensional, utilizando una base de datos relacional. Otra razón para utilizar este tipo de esquema, es su sencillez desde el punto de vista del usuario final. Las consultas no son complicadas. Por último, es el esquema con mejor rendimiento y velocidad, pues permite indexar las dimensiones de forma individualizada sin que repercuta en el rendimiento de la base de datos en su conjunto.

## 1.7 Metodologías para el diseño e implementación de un Data Mart

Con el desarrollo y avance en el nivel de conocimientos en los procesos de construcción de los DM, han surgido un conjunto de metodologías que agilizan y rigen los procesos de desarrollo de dichos sistemas. A continuación se explican algunas de las metodologías más utilizadas para la construcción de un DM:

- ✓ Kimball
- ✓ Inmon
- ✓ Hefesto

### 1.7.1 Kimball

La metodología Kimball, llamada así en homenaje a su autor Ralph Kimball, está basada en una arquitectura ascendente (Bottom-Up, en inglés); donde se definen previamente los DM y luego se integran en un DWH centralizado. La idea es construir DM independientes para evaluar las ventajas del sistema a medida que se avanza. El ciclo de vida de Kimball, ilustra el flujo general de implementación de un DM, identificando una secuencia de tareas ordenadas y actividades principales que deben suceder concurrentemente. Esta metodología acomoda las necesidades de las empresas, para lograr una única necesidad de la organización y no todos los detalles de las tareas del ciclo de vida deben ser ejecutados en todos los proyectos (Rivadera, 2010 págs. 57-59). La metodología de Kimball ofrece una amplia manera de abordar los elementos para las etapas de desarrollo, y deja claro qué se debe hacer, pero no cómo lograrlo, lo que provoca demoras en los resultados. (Ramírez, 2011)

### 1.7.2 Inmon

La metodología Inmon, llamada así en homenaje a su autor Bill Inmon, está basada en una arquitectura descendente (Top-Down, en inglés); parte de la premisa de que los DM se deben crear después de haber

terminado el DWH completamente (Ramírez, 2011). Trata el manejo de una arquitectura de múltiples niveles, en el cual el DWH colecciona datos de diversas fuentes, luego estos son integrados y finalmente distribuidos a los DM (Sedano, 2009 p. 33). De forma contraria a la de Kimball, esta metodología puede tener una implementación lenta, y es recomendada cuando se hace demasiado difícil representar el modelo a través de dimensiones y la complejidad de la solución se hace demasiado grande. No es recomendable para proyectos sencillos pues va de lo general (el DWH), a lo más específico (el DM). (Sanz, 2010 p. 24)

### 1.7.3 Hefesto 2.0

Hefesto es una metodología, cuya propuesta está fundamentada en una amplia investigación, comparación de metodologías existentes y experiencias propias adquiridas en procesos de confección de DM. La idea principal es comprender que se realizará en cada paso para no caer en la rutina de tener que seguir un método al pie de la letra sin saber exactamente qué se está haciendo; ni por qué (Bernabeu, 2010 pág. 85). La metodología Hefesto propone cuatro pasos fundamentales para su aplicación, cada uno con una serie de puntos que se deben realizar y que se resumen en el siguiente gráfico:





Figura 7. Metodología Hefesto (Bernabeu, 2010 pág. 87)

1. Se comienza recolectando las necesidades de información de los usuarios y se obtienen las preguntas claves del negocio. Luego, se deben identificar los indicadores resultantes de los interrogativos y sus respectivas perspectivas de análisis, mediante las cuales se construirá el modelo conceptual de datos del DM.
2. Se analizan los OLTP, conformando los indicadores y estableciendo las correspondencias con los datos fuentes, de esta manera se obtienen los diferentes niveles de granularidad<sup>8</sup> y se construye el modelo conceptual ampliado.

<sup>8</sup> Granularidad: describe el nivel de detalle con que se desea analizar la información presente en una base de datos.

3. Se construye el modelo lógico del depósito, donde se define cuál será el tipo de esquema que se implementará. Seguidamente, se confeccionan las tablas de dimensiones y las tablas de hechos, para luego efectuar sus respectivas uniones.
4. Por último, utilizando técnicas de limpieza y calidad de datos, procesos ETL, entre otros, se definen políticas y estrategias para la carga inicial del DM y su respectiva actualización. (Dario, 2010)

Esta metodología cuenta con las siguientes características (Bernabeu, 2009):

- ✓ Los objetivos y resultados esperados en cada fase se distinguen fácilmente y son sencillos de comprender.
- ✓ Se basa en los requerimientos de los usuarios, por lo cual su estructura es capaz de adaptarse con facilidad y rapidez ante los cambios en el negocio.
- ✓ Reduce la resistencia al cambio, involucrando a los usuarios finales en cada etapa para que tomen decisiones respecto al comportamiento y las funciones del DM.
- ✓ Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.
- ✓ Es independiente de las herramientas que se utilicen para su implementación.
- ✓ Es independiente de la estructura física que contenga el almacén de datos y de su respectiva distribución.
- ✓ Cuando se culmina con una fase, los resultados obtenidos se convierten en el punto de partida para llevar a cabo el paso siguiente.
- ✓ Se aplica tanto para DWH como para DM.

Luego del análisis realizado, se define Hefesto como metodología a seguir; esta permite la construcción del DM de forma sencilla, ordenada e intuitiva. Hefesto es una metodología ideal para las personas que entran por primera vez en el mundo de los DM, pues permite construirlos de manera metódica; guiándose por pasos lógicos, relacionados durante todas las etapas del proceso de confección. Dicha metodología agiliza el proceso de desarrollo del DM, ya que propone no entrar en fases extensas de reunión de análisis y requerimientos, ni fases de despliegue muy largas. (Bernabeu, 2009)

## 1.8 Sistemas existentes

El mundo empresarial ha seguido muy de cerca el desarrollo de las nuevas tecnologías, dada la necesidad constante de consumir información y hacer un mejor uso de los datos ante un mercado

competitivo. Es por ello que hoy en día, la mayoría de las empresas líderes en sus distintas categorías, cuentan con un sistema informático de apoyo a la toma de decisiones, tanto a nivel internacional, como nacional.

### 1.8.1 Nivel internacional

**Sistema de Información para la Gestión de Instituciones de Salud** en Uruguay: tiene el propósito de apoyar la toma de decisiones en todos los niveles de una organización, haciendo que se disponga en tiempo real de la información, logrando que el personal hospitalario no tenga que dedicarse a tareas administrativas repetitivas. (Vázquez, et al., 2003 p. 5)

**Data Warehouse de Soporte de Decisiones para un Hospital del Sistema de Salud Público** de Perú: sirve de apoyo en el proceso de toma de decisiones del directorio del hospital Santa Rosa, el cual, decide (basándose en base de datos históricos y cuadros generados en línea sobre la optimización del uso del personal) la calidad de servicio otorgada; identifica fallas en los procesos y realiza auditorías y notificaciones en tiempo real, utilizando Pentaho como suite de Inteligencia de Negocio<sup>9</sup>. (Ojeda, 2008 p. 3)(sic)

### 1.8.2 Nivel nacional

**Data Mart para la Unidad Central de Cooperación Médica (UCCM)**: ofrece reportes sobre las colaboraciones médicas en el exterior, brinda fácil acceso a los datos que necesitan frecuentemente los directivos de la UCCM y permite que exista una fuente única de información relevante para la toma de decisiones, mejorando el tiempo de respuesta. Los directivos obtendrán los valores numéricos que muestren lo que desean analizar concretamente y podrán obtener una representación gráfica para una mejor comprensión y estudio a la hora de tomar decisiones. (López, et al., 2013 p. 1)

**Data Warehouse para los Ensayos Clínicos del Centro de Inmunología Molecular**: brinda facilidades a los especialistas, para tratar los datos relacionados con los ensayos clínicos, así como posibilita obtener, mediante la interfaz web, el análisis estadístico sobre eficacia y seguridad del producto para el tratamiento del cáncer en diferentes localizaciones. La información de los ensayos clínicos se gestiona en el Centro de

---

<sup>9</sup> Inteligencia de Negocio: conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento mediante el análisis de datos existentes en una organización o empresa.

Inmunología Molecular con la intención de almacenar gran cantidad de información, para permitir la integración y contribuir a la toma de decisiones de la institución. (Ramírez, 2011)

**Data Mart para el Sistema Integral de Gestión de Medicamentos:** brinda reportes y acceso eficiente a la información que necesitan los directivos de la Industria Química Farmacéutica de Cuba (QUIMEFA), además de contener información consolidada que apoya a la correcta toma de decisiones. Para la confección del Data Mart se utilizó una metodología híbrida que engloba a Hefesto y Kimball. (Aguilar, 2011 p. VII)

El análisis detallado de las herramientas similares que realizan el proceso de toma de decisiones, a nivel internacional y nacional, brinda un acercamiento a los procesos de ETL y las herramientas OLAP, permitiendo un mayor conocimiento respecto a la construcción de los Data Marts. Estos sistemas ofrecen una visión detallada que contribuye al desarrollo del presente trabajo, aunque no estén enfocados precisamente al tema de las RAM.

### 1.9 Herramientas para la construcción del Data Mart

De acuerdo a las características que presenta el proceso de construcción de un DM, este debe estar soportado por un conjunto de herramientas, donde se establece una cooperación entre ellas, para transitar por las diferentes etapas del proceso de análisis de los datos: desde la adquisición hasta la visualización de los resultados. Entre los diferentes tipos de herramientas se encuentran:

- ✓ Sistemas Gestores de Base de Datos: PostgreSQL
- ✓ Herramientas de Integración de Datos: Clover ETL, Enhydra Octopus, Pentaho Data Integration
- ✓ Herramientas para el Proceso Analítico en Línea: Jedox Palo, Olap4J, Mondrian

#### 1.9.1 Sistemas Gestores de Base de Datos

Los Sistemas Gestores de Base de Datos (SGBD) son un tipo de software específico, dedicados a servir de interfaz entre: la base de datos, el usuario y las aplicaciones que lo utilizan. Las principales funciones que debe cumplir un SGBD, se relacionan con: la creación y mantenimiento de la base de datos, el control de accesos, la manipulación de datos de acuerdo con las necesidades del usuario, el cumplimiento de las normas de tratamiento de datos, evitar redundancias e inconsistencias y mantener la integridad. (Asenjo, 2009 p. 11)

## PostgreSQL 8.3

Es un servidor de base de datos relacional, distribuido bajo Licencia Distribución de Software Berkeley (BSD, por sus siglas en inglés) (PostgreSQL, 2013). A continuación se enumeran las principales características (Pecos, 2010) de este gestor de bases de datos:

- ✓ Soporta distintos tipos de datos y permite la creación de tipos propios
- ✓ Incorpora una estructura de arreglos (*array*) de datos
- ✓ Soporta el uso de índices, reglas y vistas
- ✓ Incluye herencia entre tablas
- ✓ Permite la gestión de diferentes usuarios y los permisos asignados a cada uno de ellos
- ✓ Soporta casi toda la sintaxis SQL (incluyendo subconsultas, transacciones, tipos y funciones definidas por el usuario)
- ✓ Cuenta con un amplio conjunto de enlaces con lenguajes de programación (C, C++, Java, Python y PHP)

Esta herramienta es utilizada por estar definida en el documento de arquitectura de SAS.

## 1.9.2 Herramientas de Integración de Datos

Las herramientas de ETL, son las encargadas de extraer datos de diferentes fuentes (base de datos, *excel*, XML y otras). Estas herramientas también permiten transformar y limpiar datos que contengan errores, así como cargar los datos a cualquier sistema de almacenamiento, como por ejemplo un DWH o un DM.

Las principales herramientas de integración de datos son:

- ✓ Clover ETL
- ✓ Enhydra Octopus
- ✓ Pentaho Data Integration

### Clover ETL

Es una herramienta de código abierto, basada en Java, con un marco de integración de datos que puede ser utilizado para transformar la información. Se puede utilizar de forma independiente como una

aplicación de servidor o puede ser empotrada en otras aplicaciones como una biblioteca de la transformación. (Luis, 2008)

Características principales de Clover ETL (Luis, 2008):

- ✓ Contiene una paleta de más de cuarenta componentes especializados de transformación
- ✓ Soporta los principales estándares de la industria de base de datos (Oracle, MS SQL, DB2, entre otras) y varias variantes de código abierto (MySQL, PostgreSQL)
- ✓ Lee y escribe datos en XML, *excel* y datos de longitud variable
- ✓ Lectura y escritura de datos a través de protocolos (FTP, SFTP, HTTP y HTTPS)
- ✓ Se puede ejecutar en plataformas como: Windows, Linux, Solaris, entre otras

## Enhydra Octopus

Es una herramienta de ETL basada en Java. Puede conectarse a cualquier base de datos usando JDBC y realizar transformaciones definidas en un archivo XML. Es compatible con los modelos de datos, por generar Identificadores de Objetos (OIDs, por sus siglas en inglés) para los nuevos objetos. Las claves naturales pueden utilizarse para insertar o actualizar los datos existentes y crear relaciones con los identificadores de objetos. (Softpedia, 2012)

Características principales de Enhydra Octopus (Softpedia, 2012):

- ✓ Su característica principal, es el requerimiento de que para acceder a las fuentes de datos que lo utilicen debe existir un manejador JDBC disponible.
- ✓ Es una herramienta poderosa capaz de: normalizar datos, crear llaves artificiales, tablas y llaves primarias.
- ✓ Todos los trabajos que se ejecutan en Octopus, son independientes de los gestores de bases de datos.

## Pentaho Data Integration 4.2

Pentaho Data Integration (PDI), ofrece capacidades de ETL mediante un enfoque innovador de metadatos. Presenta un ambiente intuitivo y gráfico, con un diseño de arrastrar y soltar y una arquitectura: probada, escalable y basada en estándares. (Pentaho, 2013)

Uno de los objetivos de la integración de datos, es hacer el proceso de ETL fácil de generar, mantener y desplegar. Pentaho Data Integration se compone de cuatro herramientas (Bernabeu, 2010):

- ✓ SPOON: permite diseñar de forma gráfica la transformación.
- ✓ PAN: ejecuta un conjunto de transformaciones diseñadas con SPOON, conocidas como trabajos, creando dependencias entre dichas transformaciones.
- ✓ CHEF: permite, mediante una interfaz gráfica, diseñar la carga de datos, incluyendo un control de estado de los trabajos.
- ✓ KITCHEN: permite ejecutar los trabajos por lotes diseñados con CHEF.

Características principales de PDI (Pentaho, 2013a):

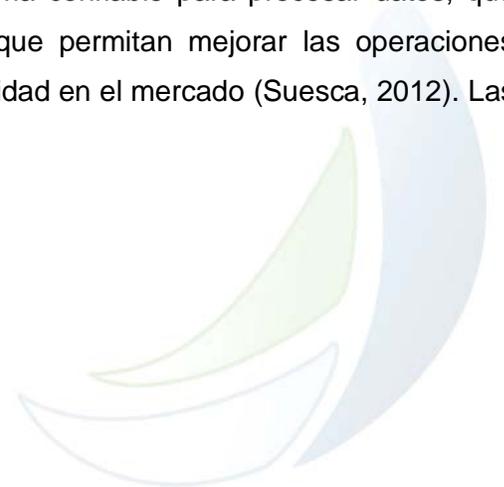
- ✓ Tiene una interfaz visual con indicadores de las transformaciones.
- ✓ Es una aplicación escrita en Java, con algunas características avanzadas escritas en JavaScript
- ✓ Está basado en la utilización de metadatos.
- ✓ Como soporte se encuentran los foros y la comunidad Pentaho.
- ✓ Con respecto a escalabilidad, soporta la arquitectura de procesamiento en paralelo para distribuir las tareas de ETL a través de múltiples servidores.
- ✓ Permite realizar la limpieza de datos.

Para el proceso de ETL, se escoge la herramienta Pentaho Data Integration, pues cuenta con características que facilitan la interacción con los datos. Además de que existe una experiencia previa, en el trabajo con esta herramienta, en la Universidad de las Ciencias Informáticas.

### 1.9.3 Herramientas para el Proceso Analítico en Línea

Las herramientas OLAP, proporcionan a las compañías un sistema confiable para procesar datos, que luego serán utilizados para llevar a cabo análisis e informes, que permitan mejorar las operaciones productivas, tomar decisiones inteligentes y optimizar la competitividad en el mercado (Suesca, 2012). Las principales herramientas de procesamiento de la información son:

- ✓ Jedox Palo
- ✓ Olap4j
- ✓ Mondrian



## **Jedox Palo**

Es un servidor de bases de datos multidimensionales, orientado a celdas, capaz de centralizar y administrar un número infinito de hojas de cálculo. El sistema opera en tiempo real, soporta la consolidación de jerarquías, así como numerosas funciones de inteligencia empresarial y es un servidor de código abierto. Esta herramienta es de gran utilidad para todos aquellos que manejen gran cantidad de datos en *excel*. (Morales, 2012)

## **Olap4J**

Es una interfaz de aplicación para el ambiente Java, que soporta la creación, almacenamiento y administración de datos para una aplicación OLAP (Morales, 2012). Permite consultar metadatos, y ejecutar las consultas de Expresiones Multidimensionales (MDX, por sus siglas en inglés), desde el entorno Java. Es la primera interfaz de aplicación Java que ha permitido a los programadores conectarse a servidores OLAP de diferentes proveedores. El primer piloto de aplicación es un consumidor de servicios web de Lenguaje de Marcado Extensibles de Análisis (XMLA, por sus siglas en inglés) genérico. Es capaz de conectarse a muchos motores de bases de datos en el mercado, a través de servicios web de estilo Protocolo de Acceso a Objetos Simple (SOAP, por sus siglas en inglés). (Humble, et al., 2011)

## **Mondrian 3.5**

Es una de las aplicaciones más importantes de la plataforma Pentaho. Es un servidor OLAP de código abierto que gestiona comunicación entre una aplicación OLAP (escrita en Java) y la base de datos, con los datos fuente (BI, 2011). Se considera además un motor de ROLAP desarrollado en Java, que permite analizar grandes conjuntos de datos, que se encuentran almacenados en el DM, pues se encarga de recibir consultas dimensionales en lenguaje de MDX y devolver los datos del cubo que correspondan a la consulta. El cubo se representa como un conjunto de metadatos, que definen cómo se han de mapear estas consultas dimensionales a sentencias SQL, para obtener de la base de datos, la información necesaria para satisfacer la consulta dimensional. Utiliza una memoria *cache*, para almacenar los resultados de las consultas a las que se acceden múltiples veces. (Mondrian, 2007)

## **STPivot**

Es un visor web OLAP, de código abierto, creado sobre la base del visor JPivot. STPivot es una librería de componentes Servidor de Páginas Java (JSP, por sus siglas en inglés), que se utiliza para construir tablas OLAP, generadas de forma dinámica, y permite a los usuarios realizar consultas OLAP por medio del

lenguaje MDX. El objetivo de STPivot es mejorar la experiencia del usuario, de Jpivot, mediante el empleo de la librería jQuery y el conjunto de tecnologías Ajax. (STPivot, 2011)

## **Schema Workbench 3.2**

Es un entorno visual para el desarrollo y prueba de cubos OLAP. Se utiliza para la creación de los archivos XML que se usan para la construcción de los cubos. Con esta aplicación, se puede configurar una conexión JDBC con el modelo físico, para elaborar el esquema lógico de manera simple y efectiva. Para ello el entorno ofrece un editor de esquemas, con la fuente de datos subyacente para su validación. Permite la ejecución de consultas MDX contra el esquema y la base de datos. (Vega Torres, et al., 2008)

Luego de analizar las herramientas OLAP, se selecciona el servidor Mondrian y la herramienta Schema Workbench, por ser las recomendadas por el Centro de Tecnologías de Gestión de Datos (DATEC) perteneciente a la UCI. Este centro cuenta con varios años de experiencia en el desarrollo de DM, contando con un gran cúmulo de bibliografía del tema. El cliente STPivot se utiliza debido a que posee una interfaz similar a Synta, además de contar con características que permiten la integración a dicho producto.

## **1.9.4 Herramienta de modelado**

Las herramientas de Ingeniería de Software Asistida por Computadora (CASE, por sus siglas en inglés) pueden ayudar en los aspectos del ciclo de vida de desarrollo del software. Estas herramientas permiten realizar el análisis y diseño orientado a objetos y abstraer el código fuente, a un nivel donde la arquitectura y el diseño se tornan más obvios y más fáciles de entender y modificar. (Sparx, 2008)

## **Enterprise Architect 7.1**

Es una herramienta de uso muy sencillo, que aborda el diseño y análisis utilizando el Lenguaje Unificado de Modelado (UML, por sus siglas en inglés) y cubre el desarrollo de software, desde la captura de requerimientos, a lo largo de las etapas de análisis, diseño, implementación, pruebas y mantenimiento. Es una herramienta multi-usuario, diseñada para ayudar a construir software robusto y fácil de mantener. Además, permite generar documentación e informes flexibles y de alta calidad. Ofrece beneficios significativos para ayudar a construir modelos rigurosos de software, donde es posible mantener la trazabilidad de manera consistente. Proporciona trazabilidad completa desde el análisis de requerimientos y los artefactos de diseño, hasta la implementación y el despliegue. (Sparx, 2008)

Esta herramienta es utilizada por estar definida en el documento de arquitectura de SAS.

## 1.10 Servidores de Aplicaciones

Los servidores de aplicaciones ofrecen una mayor capacidad de expansión, al incrementar el desempeño. Proporcionan servicios que soportan la ejecución y disponibilidad de las aplicaciones desplegadas, además de una estructura en tres capas que permite conformar el sistema de forma eficiente. Brinda la posibilidad de mejorar tres aspectos fundamentales en una aplicación: la alta disponibilidad, la escalabilidad y el mantenimiento del sistema.

Entre los servidores estudiados se encuentran:

- ✓ JBoss
- ✓ Apache Tomcat

### JBoss

Es un servidor de aplicaciones de Edición de Empresas Java (J2EE, por sus siglas en inglés) de código abierto. Al estar basado en Java, JBoss puede ser utilizado en cualquier sistema operativo que lo soporte, haciendo posible desarrollar e implementar aplicaciones Java innovadoras y escalables. (Sánchez, 2011)

Características principales (Sánchez, 2011) de JBoss:

- ✓ Código abierto
- ✓ Escalable
- ✓ Alto desempeño
- ✓ Producto de licencia de código abierto sin coste adicional
- ✓ Confiable a nivel de empresa
- ✓ Orientado a arquitectura de servicios
- ✓ Flexibilidad consistente

### Apache Tomcat 7.0

Es un servidor web que funciona como contenedor de servlets<sup>10</sup>, desarrollado en código abierto por Apache Software Foundation. Tomcat implementa las especificaciones de las tecnologías servlets Java y de JSP. (López, 2012)

---

<sup>10</sup> Servlets: pequeños programas que se ejecutan en el contexto de un navegador web.

Las características principales (Apache, 2011) de Apache Tomcat:

- ✓ Servidor de aplicaciones de código abierto
- ✓ Fácil integración con Apache
- ✓ Muy estable en sistemas UNIX
- ✓ Buena documentación en la Web
- ✓ No requiere mucha memoria para iniciar
- ✓ Es gratuito

Se selecciona como servidor de aplicaciones para la construcción del DM, Apache Tomcat 7.0, pues contiene características primordiales para su utilización, como por ejemplo: permite la inclusión de ficheros externos dentro de una aplicación, facilita la integración con Apache 2.2 (servidor de aplicación donde se encuentra el producto Synta) debido a que ambos están desarrollados por Apache Software Foundation.



### Capítulo 2: Propuesta de diseño e implementación del Data Mart

En este capítulo se describen los pasos a seguir para el diseño del DM haciendo uso de la metodología seleccionada. Se identifican las necesidades de información del cliente, permitiendo el análisis de los requerimientos, desde la identificación de indicadores y perspectivas hasta los procesos de ETL.

#### 2.1 Propuesta del sistema

Se propone el desarrollo de un DM, que permita a los farmacoepidemiólogos, la toma de decisiones respecto a las RAM ocurridas en Cuba. El DM obtendrá la información (registrada por tres módulos) de la base de datos del producto Synta. La información referente a las RAM, es tomada a partir de los registros del MRAM; debido a que un conjunto de esta información es nomenciada, es necesario obtenerla de Nomencladores. Ejemplo de esta información son los medicamentos, de los que además de su información propia (forma farmacéutica, vía de administración, presentación, etc.), se necesita conocer su clasificación VEN<sup>11</sup> (registrada por el Módulo Consumo de Medicamentos). Para el llenado del DM es necesario hacer la triangulación entre los tres módulos. El DM se integrará al MRAM, usando la herramienta STPivot (para visualizar la información), permitiendo a los farmacoepidemiólogos contar con un solo sistema que gestione las RAM. Entre los reportes que se esperan obtener están los siguientes:

- ✓ Total de RAM por sexo, causalidad
- ✓ Cantidad de pacientes afectados por una RAM, por raza y especialidad
- ✓ Total de RAM por medicamentos y grupo etario

#### 2.2 Fases de la metodología Hefesto

La metodología Hefesto cuenta con cuatro fases que se encargan de describir cada proceso de desarrollo de un DM, a continuación se abordarán los elementos que componen cada una de ellas.

##### 2.2.1 Fase 1: Análisis de los requerimientos

Se identifican los requerimientos de los usuarios a través de preguntas, que permitan dar cumplimiento a los objetivos trazados. Se analizan las preguntas con el propósito de identificar cuáles serán los

---

<sup>11</sup> VEN: clasificación de los medicamentos, teniendo en cuenta si son: vitales, no vitales, esenciales o especiales.

perspectivas e indicadores que serán utilizados para la construcción del DM y finalmente se confecciona un modelo conceptual donde se visualiza el resultado obtenido en este paso.

### Identificar preguntas

Para el desarrollo de este primer paso de la metodología, se realizó una entrevista a un grupo de farmacoepidemiólogos del MINSAP, con el propósito de obtener las necesidades de información de los usuarios, los resultados esperados y los reportes que se consideren importantes para el cliente.

De la entrevista realizada se obtienen las preguntas claves del negocio, donde se desea conocer:

- ✓ La cantidad de pacientes, afectados por una RAM, por sexo y raza, en un período de tiempo determinado
- ✓ La cantidad de pacientes, afectados por una RAM, por grupo etario y edad en un período de tiempo
- ✓ La cantidad de pacientes, afectados por una RAM, por antecedentes patológicos
- ✓ La cantidad de pacientes, afectados por una RAM, por localización
- ✓ La cantidad de pacientes, afectados por una RAM, por medicamentos
- ✓ El porcentaje de pacientes, afectados por una RAM, por medicamentos
- ✓ El porcentaje de pacientes, afectados por una RAM, por sexo, raza, grupo etario y edad en un período de tiempo determinado
- ✓ El total de registros RAM, por niveles de atención, especialidad y localización, en un período de tiempo
- ✓ El total de registro RAM, por frecuencia, desenlace, causalidad y severidad en un período de tiempo
- ✓ El total de registro RAM, por medicamentos
- ✓ El total de registro RAM, por secuencia temporal y datos registrados
- ✓ La cantidad de medicamentos, por registro RAM, en un período de tiempo
- ✓ La cantidad de medicamentos que han ocasionado una RAM a un paciente determinado
- ✓ La cantidad de medicamentos que han ocasionado RAM, según su clasificación VEN

### Identificar indicadores y perspectivas

#### Indicadores:

Los indicadores son valores numéricos y representan lo que se desea analizar concretamente, por ejemplo: saldos, promedios, cantidades, sumatorias, fórmulas, porcentos, entre otras. (Dario, 2010)

#### Perspectivas:

Las perspectivas son objetos mediante los cuales se quiere examinar los indicadores, con el objetivo de responder a las preguntas planteadas, por ejemplo: clientes, proveedores, sucursales, países, productos, rubros, entre otras. (Dario, 2010)

A partir de las preguntas identificadas en la entrevista realizada se obtuvieron los siguientes indicadores con sus correspondientes perspectivas:

Cantidad de pacientes, afectados por una RAM, por sexo y raza, en un período de tiempo determinado.

Indicador

Perspectiva

Cantidad de pacientes, afectados por una RAM, por grupo etario y edad, en un período de tiempo.

Indicador

Perspectiva

Cantidad de pacientes, afectados por una RAM, por antecedentes patológicos.

Indicador

Perspectiva

Cantidad de pacientes, afectados por una RAM, por localización.

Indicador

Perspectiva

Cantidad de pacientes, afectados por una RAM, por medicamentos.

Indicador

Perspectiva

Porcentaje de pacientes, afectados por una RAM, por medicamentos.

Indicador

Perspectiva

## Capítulo 2: Propuesta de diseño e implementación del Data Mart

Porcentaje de pacientes, afectados por una RAM, por sexo, raza, grupo etario y edad.

Indicador

Perspectiva

Total de registros RAM por niveles de atención, especialidad y localización, en un período de tiempo.

Indicador

Perspectiva

Total de registros RAM por frecuencia, desenlace, causalidad y severidad, en un período de tiempo.

Indicador

Perspectiva

Total de registros RAM, por medicamentos.

Indicador

Perspectiva

Total de registros RAM, por secuencia temporal y datos registrados.

Indicador

Perspectiva

Cantidad de medicamentos, por registro RAM, en un período de tiempo.

Indicador

Perspectiva

Cantidad de medicamentos que han ocasionado una RAM a un paciente determinado.

Indicador

Perspectiva

Cantidad de medicamentos que han ocasionado RAM, por clasificación VEN.

Indicador

Perspectiva

### Modelo conceptual

Luego de identificar los indicadores y perspectivas, se confecciona el modelo conceptual de los datos del DM, con el objetivo de facilitar la comprensión del alcance del sistema y obtener un alto nivel de definición de los datos. En la Figura 8 se representan a la izquierda las perspectivas y a la derecha los indicadores, siendo estos analizados desde los registros de las RAM.

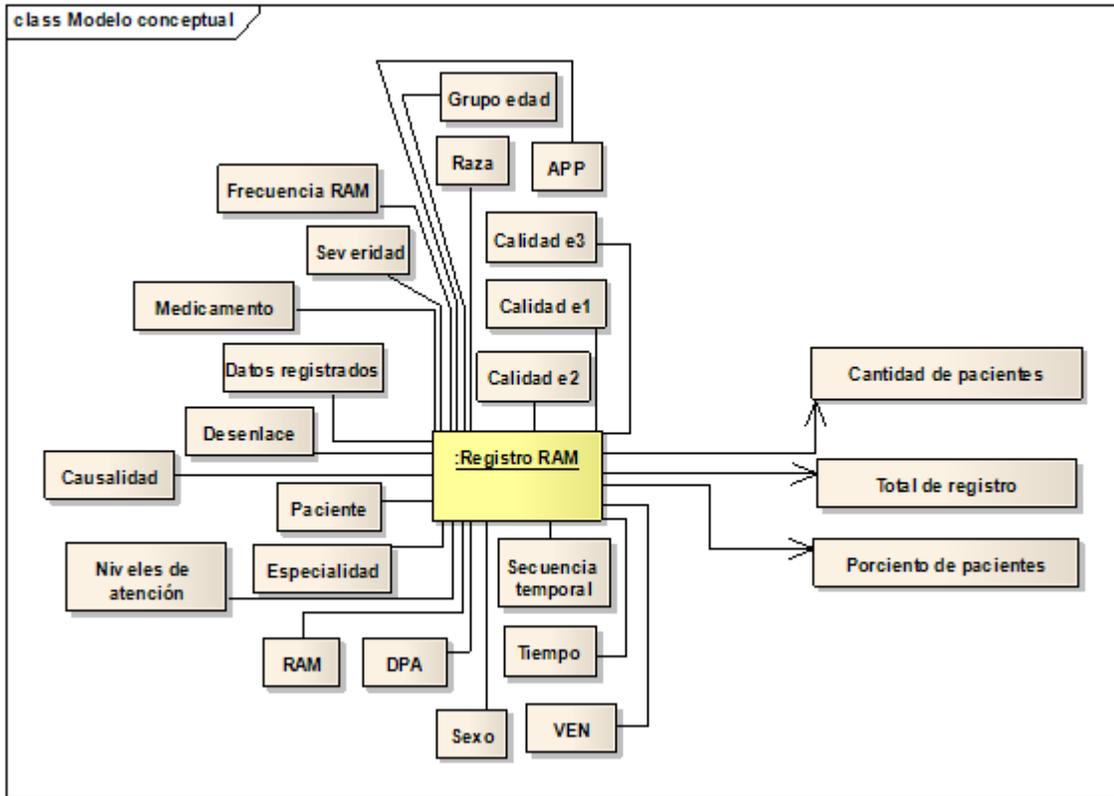


Figura 8. Modelo conceptual (elaboración propia)

### 2.2.2 Fase 2: Análisis de los OLTP

Las fuentes OLTP, se analizan para determinar cómo se calculan los indicadores y para establecer las respectivas correspondencias, entre el modelo conceptual creado en el paso anterior y las fuentes de datos. Se definen los campos que se incluyen en cada perspectiva y finalmente, se amplía el modelo conceptual con la información obtenida en este paso.

#### Conformar indicadores

En este paso se obtienen los elementos que componen a los indicadores con sus respectivas fórmulas.

Tabla 1. Confirmación de indicadores (elaboración propia)

	Indicador	Función de sumarización	Aclaración
1	Cantidad de pacientes	distinct-count	Cuenta el total de pacientes diferentes
2	Total de registros	count	Cuenta el total de registros
3	Porcentaje de pacientes	(Cantidad de pacientes * 100) / Total de registros	Calcula el porcentaje de pacientes registrados por RAM

### Establecer correspondencias

El objetivo de este paso, es examinar los OLTP disponibles, que contengan la información requerida, para poder identificar las correspondencias entre el modelo conceptual y las fuentes de datos.

Para establecer las correspondencias entre la base de datos de Synta y el DM, las relaciones identificadas fueron las siguientes:

- ✓ Los campos “mes\_entrada\_id” y “año” de la tabla “tb\_ra\_datos\_generales”, se relaciona con la perspectiva “Tiempo”.
- ✓ El campo “sexo\_id”, “raza\_id”, “id” de la tabla “tb\_ra\_paciente”, se relaciona con la perspectiva “Sexo” y “Raza”, “Paciente” respectivamente.
- ✓ El campo “antecedente\_id” de la tabla “tb\_ra\_antecedentes\_patologicos”, se relaciona con la perspectiva “APP”.
- ✓ El campo “edad\_paciente” de la tabla “tb\_ra\_registro\_ram”, se relaciona con la perspectiva “Grupo edad”.
- ✓ El campo “medicamento\_id” de la tabla “tb\_ra\_r\_medicamentos\_nomenclados”, se relaciona con la perspectiva “Medicamentos”.
- ✓ El campo “clasificacion\_ven” de la tabla “tb\_c\_r\_medicamento\_cuadro\_basico”, se relaciona con la perspectiva “VEN”.
- ✓ El campo “ram\_id” de la tabla “tb\_ra\_r\_ram\_asociada”, se relaciona con la perspectiva “RAM”.
- ✓ Los campos “frecuencia\_ram\_id”, “desenlace\_id”, “severidad\_id”, “causalidad\_id” y “secuencia\_temporal\_id” de la tabla “tb\_ra\_registro\_ram”, se relacionan con las perspectivas “Frecuencia RAM”, “Desenlace”, “Severidad”, “Causalidad” y “Secuencia temporal”, respectivamente.

- ✓ Los campos “baja”, “suspendio\_medificacion”, “re\_exposicion”, “recuperado”, de la tabla “tb\_ra\_registro\_ram” y los campos “urgencia”, “ingreso”, “reposo” y “peligro\_vida” de la tabla “tb\_ra\_r\_datos\_de\_ram”, se relacionan con la perspectiva “Datos registrados”.
- ✓ El campo “nivel\_de\_atencion\_id” de la tabla “tb\_ra\_otros\_datos\_historico”, se relaciona con la perspectiva “Niveles de atención”.
- ✓ Los campos “provincia\_id” y “municipio\_id” de la tabla “tb\_ra\_datos\_generales”, se relacionan con la perspectiva “Localización”.
- ✓ El campo “especialidad\_id” de la tabla “tb\_ra\_notificador”, se relaciona con la perspectiva “Especialidad”.
- ✓ Los campos “e1”, “e2” y “e3” de la tabla “tb\_ra\_registro\_ram”, se relacionan con las perspectivas “Calidad e1”, “Calidad e2” y “Calidad e3” respectivamente.

Para un mejor entendimiento y comprensión de la relación entre los atributos de la base de datos y las perspectivas, se realiza una tabla con la relación existente entra ellas (ver Anexo 2).

### Nivel de granularidad

Luego de establecer las relaciones con los OLTP, se seleccionan los campos que contendrá cada perspectiva, pues será por lo que se examinarán y filtrarán los indicadores. Para ello, se presentan a los usuarios, los datos de análisis disponibles para cada perspectiva y se decide cuáles son más relevantes para consultar los indicadores y cuáles no. La selección de los campos que integran cada perspectiva, determina la granularidad de la información encontrada en el DM. Los resultados obtenidos son los siguientes:

#### Perspectiva “Tiempo”:

- ✓ “Mes”: referido al nombre del mes
- ✓ “Trimestre”
- ✓ “Año”

#### Perspectiva “Sexo”:

- ✓ “id\_sexo”
- ✓ “nombre\_sexo”: referido al nombre del sexo



### Perspectiva “Raza”:

- ✓ “id\_raza”
- ✓ “nombre\_raza”: nombre de la raza de las personas

### Perspectiva “Grupo edad”:

- ✓ “id\_grupo\_etario”
- ✓ “nombre\_grupo\_etario”: nombre de las clasificaciones de las edades de las personas
- ✓ “edad”

### Perspectiva “APP”:

- ✓ “id\_app”
- ✓ “nombre\_app”: nombre de los antecedentes patológicos de los pacientes

### Perspectiva “Paciente”:

- ✓ “id\_paciente”
- ✓ “embarazada”: condición de embarazo del paciente (embaraza o no embarazada)
- ✓ “lactante”: condición de lactante del paciente (lactante o no lactante)
- ✓ “puérpera”: si el paciente se encuentra recién parida (puérpera o no puérpera)

### Perspectiva “Medicamentos”:

- ✓ “id\_medicamento”
- ✓ “nombre\_medicamento”: nombre de los medicamentos
- ✓ “cantidad\_presentacion”: número de medida del medicamento
- ✓ “nombre\_unidad\_medida”: unidad de medida en que se presenta el medicamento
- ✓ “nombre\_forma\_farmaceutica”: forma en que se presenta el medicamento
- ✓ “nombre\_via\_administracion”: vía de administración del medicamento
- ✓ “nombre\_fabricante”: nombre del fabricante del medicamento

### Perspectiva “VEN”:

- ✓ “id\_ven”
- ✓ “nombre\_ven”: nombre de la clasificación de los medicamentos (VEN)

### Perspectiva “RAM”:

- ✓ “id\_ram”
- ✓ “nombre\_ram”: nombre de la RAM
- ✓ “principal”: condición que se le da a la RAM de principal o no principal
- ✓ “nombre\_sistema\_organos”: nombre del sistema de órganos afectado por la RAM principal

### Perspectiva “Frecuencia RAM”:

- ✓ “id\_frecuencia\_ram”
- ✓ “nombre\_frecuencia\_ram”: frecuencia con que ocurre la RAM

### Perspectiva “Desenlace”:

- ✓ “id\_desenlace”
- ✓ “nombre\_desenlace”: desenlace que provoca la RAM

### Perspectiva “Severidad”:

- ✓ “id\_severidad”
- ✓ “nombre\_severidad”: severidad de la RAM

### Perspectiva “Causalidad”:

- ✓ “id\_causalidad”
- ✓ “nombre\_causalidad”: causa que ocasiona la RAM

### Perspectiva “Secuencia temporal”:

- ✓ “id\_secuencia\_temporal”
- ✓ “nombre\_secuencia\_temporal”: duración de la RAM en un paciente

### Perspectiva “Datos registrados”:

- ✓ “peligro\_vida”: especifica si la RAM puso en peligro la vida del paciente
- ✓ “suspension\_medicamento”: condición de suspensión del medicamento sospechoso cuando ocurre la RAM
- ✓ “re\_exposicion”: condición de haber ingerido nuevamente el medicamento sospechoso que ocasionó la RAM
- ✓ “recuperado”: situación de mejora del paciente (recuperado o no recuperado)

- ✓ “ingreso”: condición que especifica si el paciente requirió ingreso
- ✓ “atencion\_urgencia”: especifica si el paciente requirió atención de urgencia
- ✓ “reposo\_mas\_de\_tres\_dias”: especifica si el paciente requirió de reposo por más de tres días
- ✓ “baja\_laboral”: especifica si el paciente, por causa de la RAM, requirió de baja laboral

### **Perspectiva “Calidad e1”:**

- ✓ “e1”: calidad del registro de la RAM

### **Perspectiva “Calidad e2”:**

- ✓ “e2”: calidad del registro de la RAM

### **Perspectiva “Calidad e3”:**

- ✓ “e3”: calidad del registro de la RAM

### **Perspectiva “Niveles de atención”:**

- ✓ “id\_niveles\_atencion”
- ✓ “nombre\_niveles\_atencion”: nombre del nivel de atención en que fue atendido el paciente

### **Perspectiva “Localización”:**

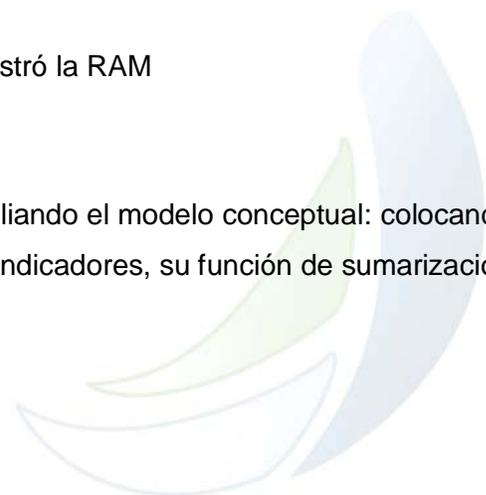
- ✓ “nombre\_provincia”: nombre de la provincia en que fue atendido el paciente
- ✓ “nombre\_municipio”: nombre del municipio en que fue atendido el paciente

### **Perspectiva “Especialidad”:**

- ✓ “id\_especialidad”
- ✓ “nombre\_especialidad”: especialidad de la persona que registró la RAM

### **Modelo conceptual ampliado**

Los resultados obtenidos en los pasos anteriores, se grafican, ampliando el modelo conceptual: colocando en cada perspectiva los campos de interés seleccionados y en los indicadores, su función de sumarización (ver Figura 1).



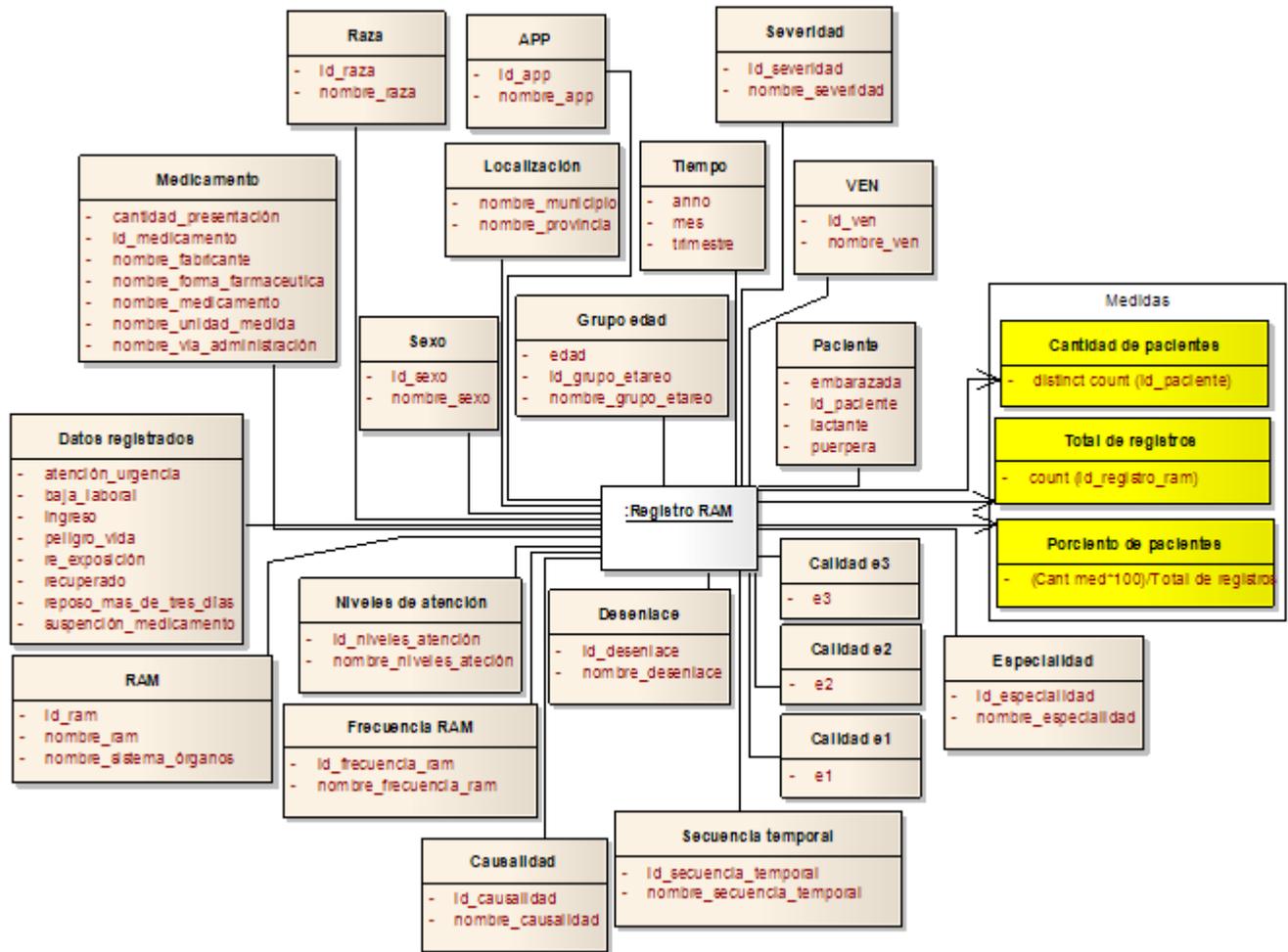


Figura 9. Modelo conceptual ampliado (elaboración propia)

### 2.2.3 Fase 3: Modelo lógico del Data Mart

Teniendo como base el modelo conceptual que ha sido creado, se confecciona el modelo lógico de la estructura del DM. Para hacer esta confección, se define el tipo de modelo que se utilizará y luego se llevan a cabo las acciones propias del caso, para diseñar las tablas de dimensiones y de hechos. Finalmente, se realizan las uniones pertinentes entre estas tablas.

### Tipo de modelo lógico del Data Mart

Para obtener la estructura del DM se utiliza el esquema en estrella, explicado en el capítulo anterior, debido a sus características, ventajas y diferencias con los otros esquemas.

### Tablas de dimensiones

En este paso se diseñan las tablas de dimensiones que forman parte del DM. Cada perspectiva definida en el modelo conceptual constituye una tabla de dimensión. Para ello se toma cada perspectiva con sus campos relacionados y se elige el nombre que identifica la tabla de dimensión. Se añade un campo que represente su clave principal y se redefinen los nombres de los campos que lo necesiten. En la Figura 10 se muestra la transformación desde la perspectiva “RAM” a la dimensión “dim\_ram”. Para visualizar las transformaciones de las restantes dimensiones consultar el Anexo 3.

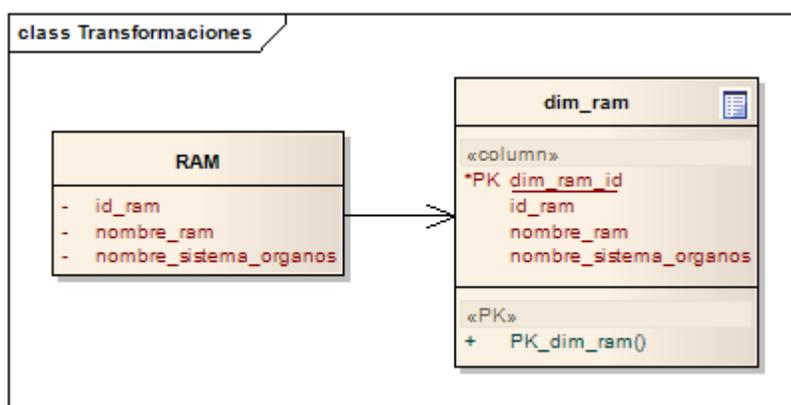


Figura 10. Dimensión RAM (elaboración propia)

### Tablas de hechos

En este paso, se define una tabla de hechos, que contiene los indicadores de estudio obtenidos anteriormente. Se asigna el nombre “hecho\_registro\_ram” a la tabla de hechos, de forma tal que representa la información analizada. Se define la clave primaria de la tabla (es la combinación de una clave primaria definida, con las demás claves que herede de sus relaciones con las dimensiones y los puentes<sup>12</sup>) (ver Figura 11).

<sup>12</sup> Puente: dimensión creada a partir de la relación de muchos a muchos entre un hecho y una dimensión.

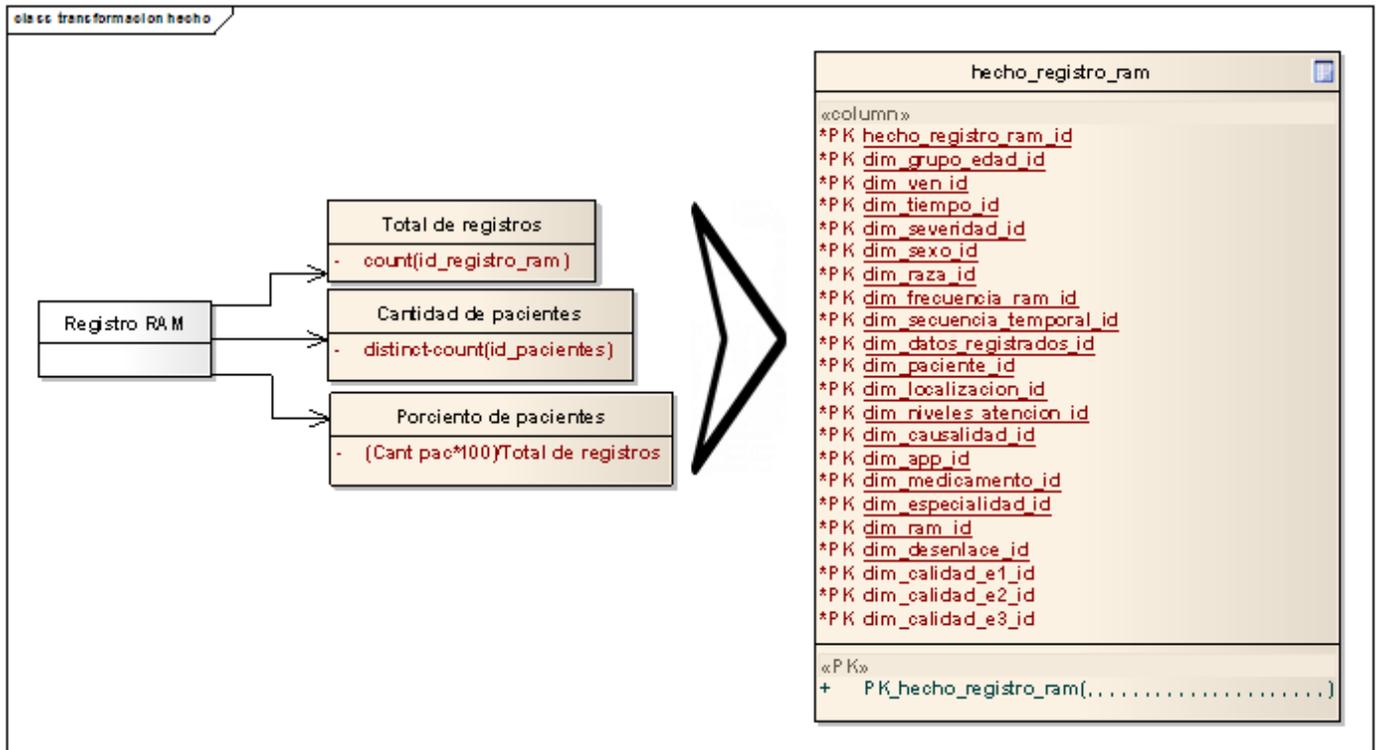


Figura 11. Diseño de la tabla de hechos (elaboración propia)

### Uniones

En este paso se realizan las uniones correspondientes entre las tablas de dimensiones y la tabla de hecho, mostrando de forma detallada las relaciones entre ellas (ver Figura 12).



## Capítulo 2: Propuesta de diseño e implementación del Data Mart

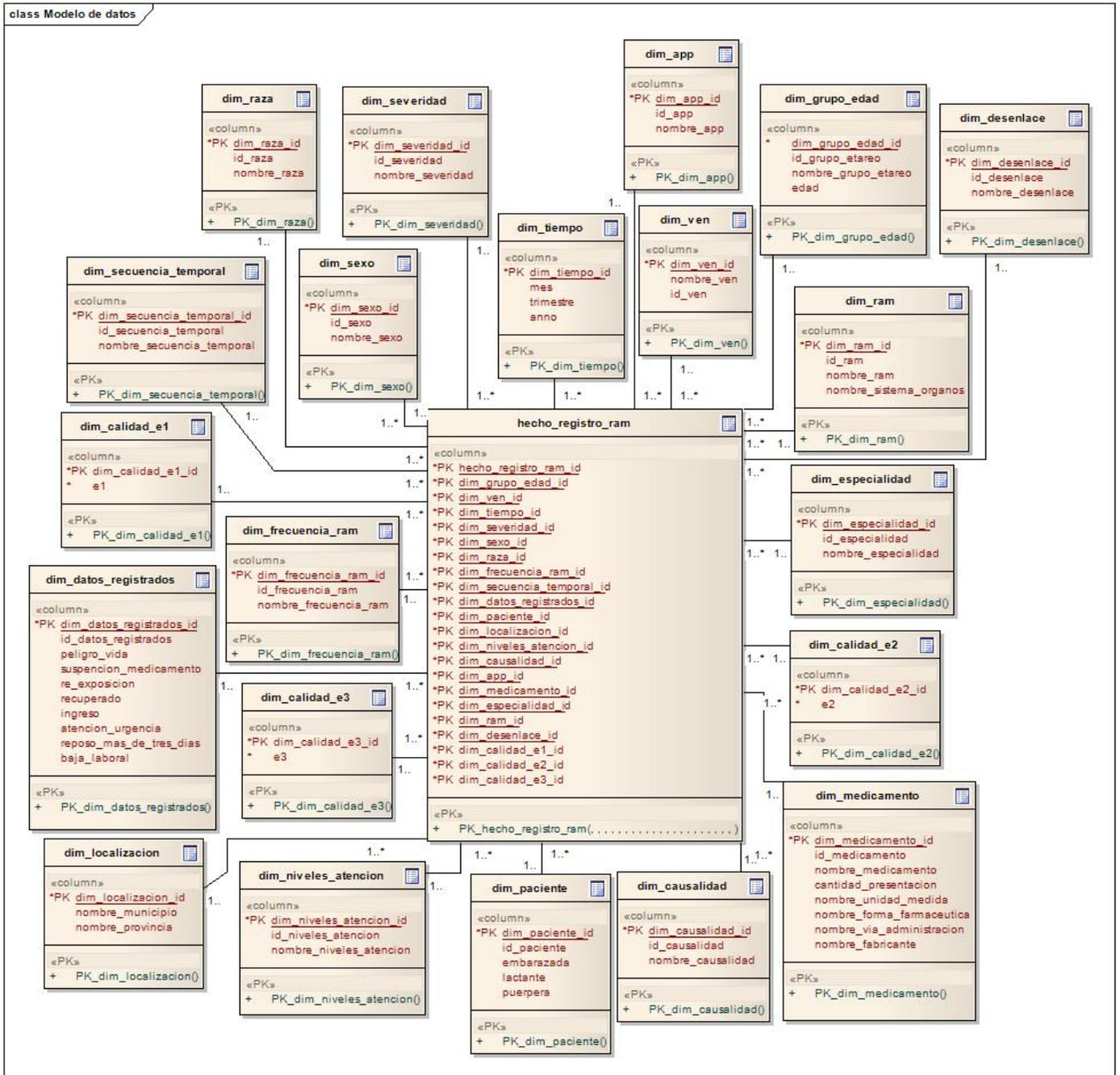


Figura 12. Uniones de las dimensiones (elaboración propia)

### 2.2.4 Fase 4: Integración de datos

Una vez construido el modelo lógico, se procede a poblarlo con datos, utilizando técnicas de limpieza y calidad de datos (es el proceso de transformación donde se realiza una revisión de la información por si presenta errores como: falta de caracteres, campos vacíos, entre otros). Luego se definen las reglas y políticas para su respectiva actualización, así como los procesos que la llevarán a cabo.

#### Carga inicial

Para llevar a cabo el soporte de la carga inicial, se emplea la herramienta antes descrita: Pentaho Data Integration, donde se realiza una serie de pasos para poblar de datos al DM, iniciando este proceso a partir de la confección de las transformaciones. A continuación se especifica la transformación RAM (ver Figura 13), las demás transformaciones se pueden consultar en el Anexo 4.

En el caso de la dimensión RAM, el proceso inicia insertando en la dimensión: el identificador nombre, si es la reacción principal o no y el sistema de órganos que afecta. Para obtener dichos atributos se procedió de la siguiente manera:

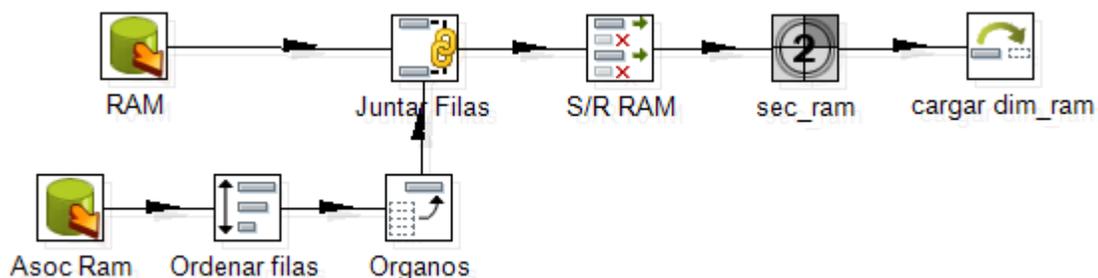


Figura 13. Transformación RAM (elaboración propia)

- ✓ Se utilizó un componente Entrada Tabla (“RAM”) que permitió obtener la información de las RAM desde la base de datos de Synta; de forma paralela se obtuvieron los campos asociados (Entrada Tabla “Asoc Ram”) y se ordenaron (por el identificador y nombre) los campos asociados obtenidos, a través del componente “Ordenar filas”.
- ✓ A través de un componente Des-Normalización de filas (“Organos”), se obtuvo el campo asociado sistema de órganos que afecta la RAM. Luego con el componente “Juntar Filas” se unificó (por “id\_ram”) la información que proveniente del componente “RAM”, con la que se obtuvo de “Asoc Ram”.

## Capítulo 2: Propuesta de diseño e implementación del Data Mart

- ✓ A través de un componente Selecciona/Renombrar Valores (“S/R RAM”), se renombraron las variables que contienen los valores, para que llegaran al almacén en la forma deseada.
- ✓ Luego con un componente Añadir secuencia (“sec\_ram”), se generó la clave principal de la dimensión y con un componente Insertar/Actualizar (“cargar dim\_ram”) se adicionaron los datos en la dimensión RAM.

Para la carga del hecho y los puentes, se realiza un trabajo que tiene como objetivo ejecutar la carga de todas las dimensiones del DM. Para iniciar este proceso se utiliza el componente “Start”, el cual inicializa la ETL por la dimensión “app” y culmina con la transformación “puente\_med\_ram”. El trabajo cuenta, para el manejo de errores en cualquier etapa, con el componente “DUMMY”, que informa mediante los componentes Display MsgBox Info (“Mensaje de error” y “Mensaje\_final”), al usuario, sobre las acciones realizadas (ver Figura 14).

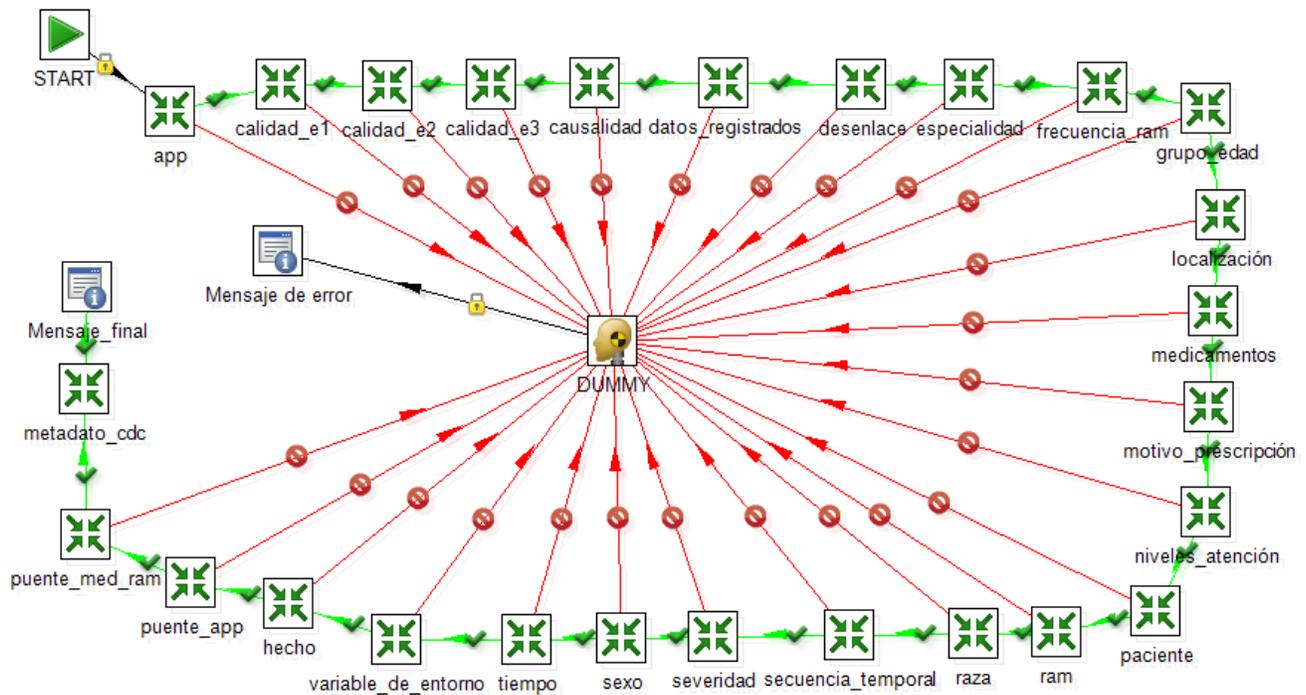


Figura 14. Ejecución del trabajo (elaboración propia)

## Capítulo 3: Proceso Analítico en Línea del Data Mart

En el presente capítulo se explica la construcción del cubo de datos, donde se definen las jerarquías y los atributos de las dimensiones y la implementación de los valores numéricos (medidas). Además se visualizan los resultados obtenidos desde el MRAM, mediante el visor web STPivot, que permite consultar la información almacenada y analizarla mediante tablas de datos y gráficas.

### 3.1 Proceso de análisis de la información del Data Mart

#### 3.1.1 Determinar jerarquías

El cubo dimensional se compone de las dimensiones, sobre las que es necesario establecer las jerarquías (representan relaciones lógicas entre dos o más atributos). La principal ventaja de manejar jerarquías, reside en poder analizar los datos, desde su nivel más general al más detallado o del menor nivel de detalle hacia el más general. (Juárez, 2012 p. 11)

Las jerarquías poseen las siguientes características (Juárez, 2012 p. 11):

- ✓ Pueden existir varias en una misma dimensión
- ✓ Están compuestas por dos o más niveles
- ✓ Se tiene una relación de “uno-muchos” o “padre-hijo” entre atributos consecutivos de un nivel superior y uno inferior

En la Figura 15 se muestran las jerarquías de los atributos de algunas de las dimensiones, que componen al DM. Para visualizar las jerarquías de las otras dimensiones, ver Anexo 5.

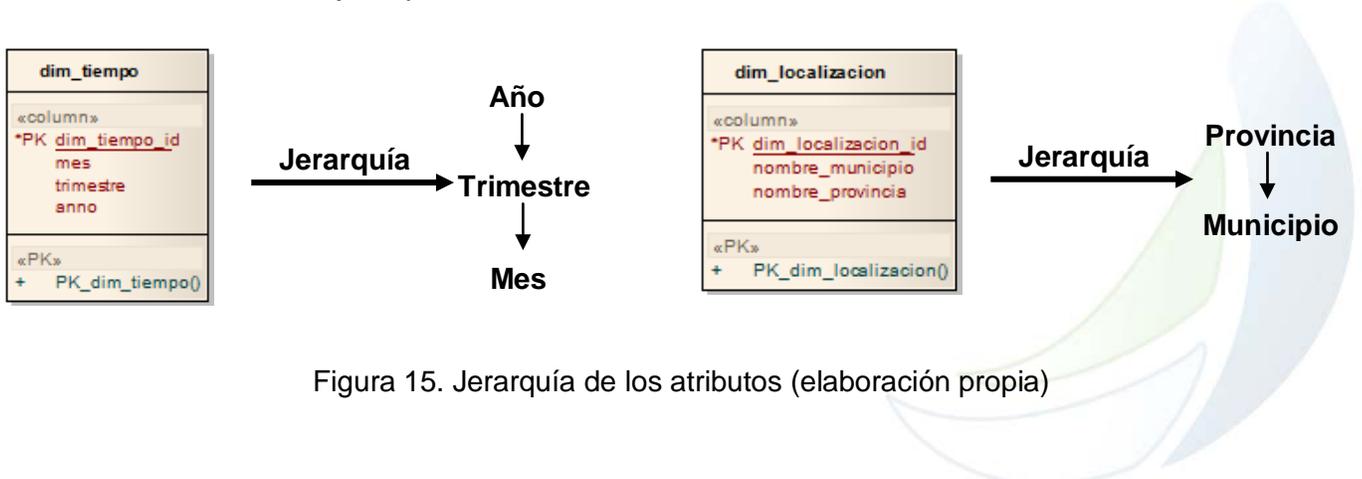


Figura 15. Jerarquía de los atributos (elaboración propia)

### 3.1.2 Medidas

Luego de establecidas las jerarquías, de los atributos de cada una de las dimensiones, se elaboran las medidas para dar respuesta a los intereses de los farmacoepidemiólogos. Para ello se utilizan las consultas MDX, que se ejecutan sobre el DM para obtener los indicadores que contribuirán a la toma de decisiones. Con la herramienta Workbench se diseña el cubo de datos, agregándole las medidas y formulando las consultas para obtener: el total de registro y la cantidad y el porcentaje de los pacientes afectados por una RAM.

A continuación se muestran dos tipos de medidas (ver Figura 2 y 17). En la primera se muestra el total de registros, el cual se obtiene aplicando la función de conteo (count) sobre el id de la tabla de hechos ("hecho\_registro\_ram\_id"). La segunda medida es un miembro calculable; es utilizada para obtener el porcentaje de pacientes. Este cálculo se realiza a través de una fórmula (cantidad de pacientes multiplicado por cien y dividido por el total de registros).

Measure for 'Ram' Cube	
Attribute	Value
name	total_registro
description	
aggregator	count
column	hecho_registro_ram_id
formatString	
visible	<input checked="" type="checkbox"/>
datatype	Numeric
formatter	
caption	Total Registros

Figura 16. Medida total de registros (elaboración propia)

Calculated Member for 'Ram' Cube	
Attribute	Value
name	porciento
description	
caption	PorCiento(%)
dimension	Measures
visible	<input checked="" type="checkbox"/>
formula   formulaElement...	([Measures].[cant_pacientes])/([Measures].[total_registro])
formatString	#.00%

Figura 17. Medida calculable porcentaje de pacientes (elaboración propia)

### 3.1.3 Consultas multidimensionales

Las MDX permiten consultar objetos multidimensionales (como el cubo de RAM) y devolver conjuntos de celdas multidimensionales, que contengan los datos del cubo. Las consultas calcularán los datos numéricos y estos serán añadidos en los resultados a visualizar, utilizando la herramienta Mondrian.

Ejemplos de consultas multidimensionales:

En la siguiente consulta se seleccionan los indicadores o medidas: total de registros, cantidad y porcentaje de pacientes (en las columnas) y los medicamentos (en las filas), obteniendo la información a través del cubo de datos “Ram”.

```
select {[Measures].[total_registro], [Measures].[cant_pacientes], [Measures].[porcentaje]} ON  
COLUMNS, {[Medicamentos.Medicamentos]} ON ROWS from [Ram]
```

En la consulta que se muestra a continuación, se seleccionan los indicadores o medidas: total de registros, cantidad y porcentaje de pacientes (en las columnas) y la jerarquía medicamentos relacionados con las RAM, las que pueden ser de tipo principal o no principal (en las filas), obteniendo la información a través del cubo de datos “Ram”.

```
select {[Measures].[total_registro], [Measures].[cant_pacientes], [Measures].[porcentaje]} ON  
COLUMNS, Hierarchize({([Medicamentos.Medicamentos].[Medicamentos], [ram.Tipos de RAM].[Tipo  
de RAM]))} ON ROWS from [Ram]
```

### 3.1.4 Resultados del cliente STPivot

Una vez diseñado el cubo “Ram”, el mismo es almacenado en un archivo XML que se cargará en el Mondrian y será utilizado para realizar el Proceso Analítico en Línea. Una vez realizados estos pasos, la información será visualizada por el cliente STPivot.

El cliente STPivot muestra un menú con varias opciones, permitiendo:

- ✓ Filtrar los campos
- ✓ Añadir o quitar campos
- ✓ Emplear filtros de visualización
- ✓ Observar la información desde diferentes perspectivas

- ✓ Mostrar los datos mediante tablas y gráficas
- ✓ Exportar los datos a formato *pdf* o *excel*

En la Figura 3 se muestran los medicamentos que se consideran sospechosos en reacciones adversas, determinando cuántas reacciones ha provocado cada medicamento y a cuántos pacientes ha afectado; así como el porcentaje que representan dichos pacientes.

Medicamentos	Total Registros	Cant de pacientes	PorCiento(%)
Medicamentos	30	7	23,33%
+ FENTANYL 0,05 MG/ML x3 ML	10	5	50,00%
+ FLUOTANO FCO X 250 ML (HALOTANO)	2	1	50,00%
+ ISOFLUORANO 100% SOL 250ML FCO	8	5	62,50%
+ KETAMINA 10 MG BBO. X 20 ML (CETAMIN)	6	3	50,00%
+ KETAMINA 50 MG X 10 ML BBO. (CETAMIN)	2	2	100,00%
+ OXIDO NITROSO KG	2	2	100,00%

Figura 18. Reporte de la dimensión medicamentos (elaboración propia)

En la Figura 19 se observa otro ejemplo de reporte OLAP: las RAM ocurridas por sexo, determinando la cantidad de RAM registradas y la cantidad de pacientes por sexo, así como el porcentaje que representan.

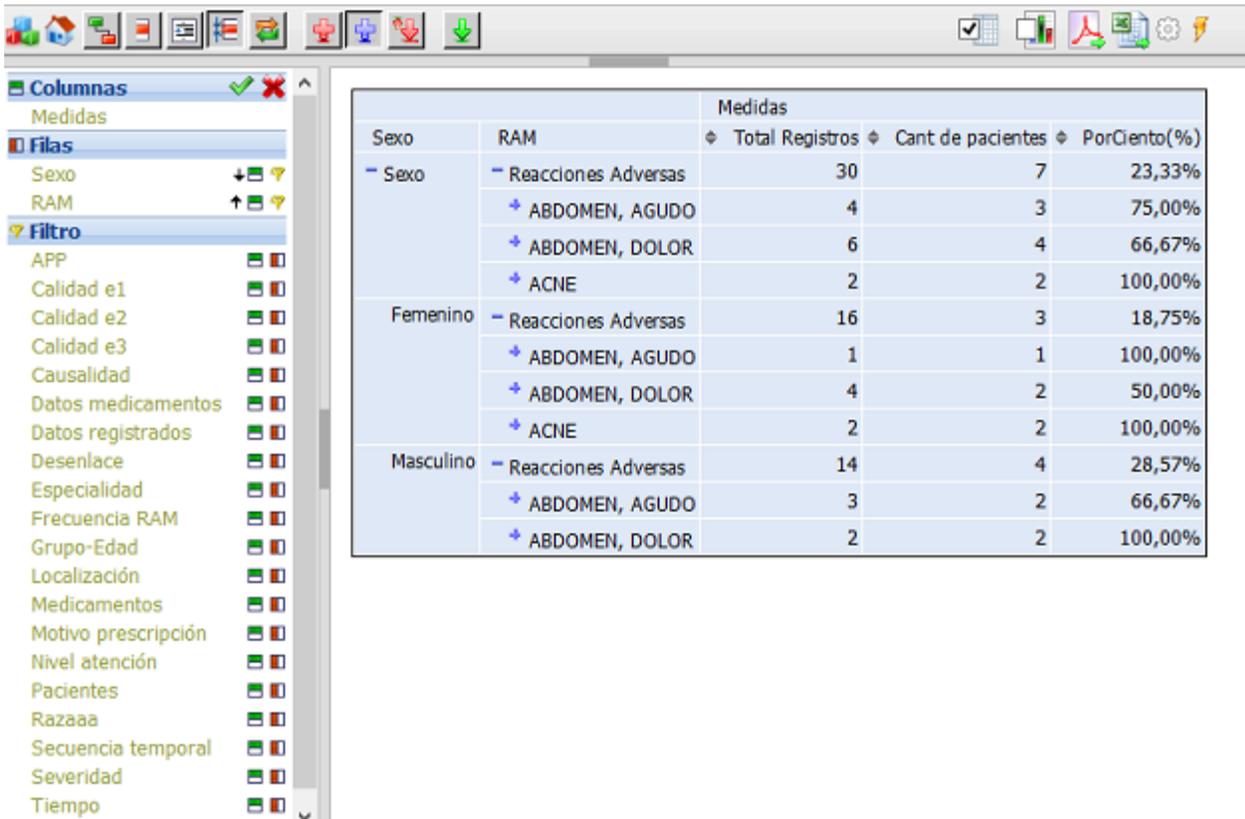
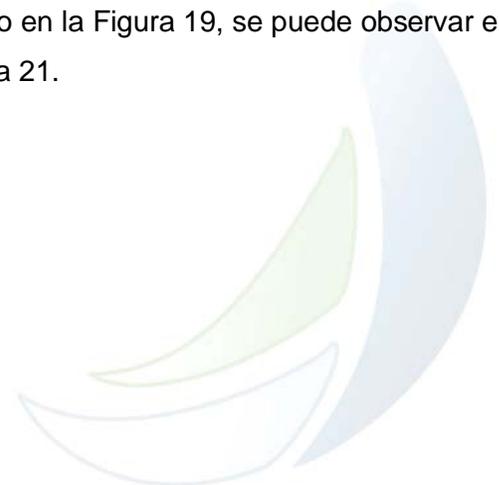


Figura 19. Información filtrada (elaboración propia)

El cliente STPivot también permite mostrar la información a través de gráficas, para que el usuario tenga la posibilidad de observarla desde otra perspectiva, para un mejor entendimiento y comprensión para el apoyo a la toma de decisiones. La información del reporte mostrado en la Figura 19, se puede observar en forma de gráficas: de barras en la Figura 20 y de pastel en la Figura 21.



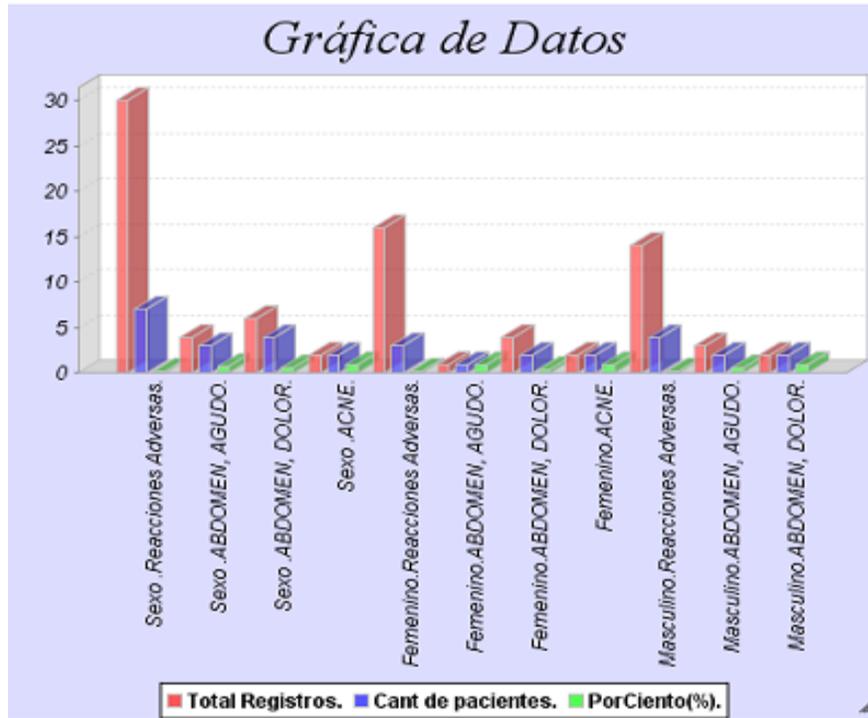


Figura 20. Información en gráfica de barras (elaboración propia)

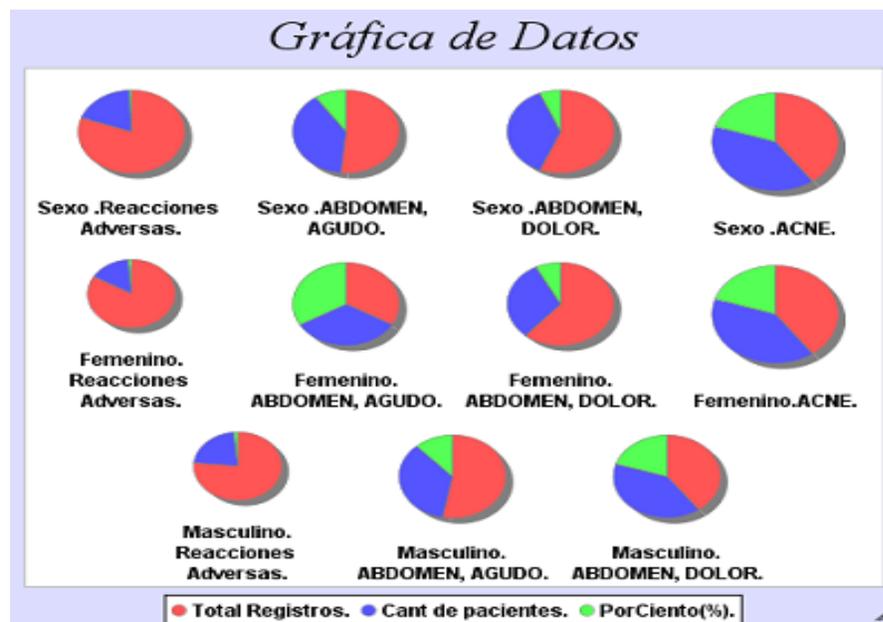


Figura 21. Información en gráficas de pastel (elaboración propia)

### 3.1.5 Integración del Data Mart con el Módulo de Reacciones Adversas a Medicamentos

Para realizar la integración del DM con el MRAM del producto Synta, se realizaron los siguientes pasos:

1. Se analizaron las características que presenta el MRAM, para poder visualizar la información que brinda el cliente STPivot desde dicho módulo.
2. Para configurar el cliente STPivot, se modificaron los archivos web.xml y datasource.xml. En web.xml se le agregaron los *servlets* correspondientes a STPivot (para que las peticiones fuesen atendidas por este visor, de forma correcta). Luego se configuró el modo de conexión, en el archivo datasource.xml, donde se especificó: la dirección de la base de datos del DM, el usuario, la contraseña y el *driver* (para PostgreSQL) de conexión, además del catálogo que hace referencia al cubo "Ram".
3. En el MRAM se creó una plantilla siguiendo las políticas del proyecto Synta, que permitió la visualización de la información a través del cliente STPivot. Debido a que el servidor web de Synta es Apache 2.2 y STPivot utiliza el Apache Tomcat 7.0, fue necesario utilizar una etiqueta HTML ("iframe"), para mostrar el visor dentro del MRAM. A esta etiqueta se le especificó la dirección URL del cliente web y el ancho y el alto en que debía ser mostrado.

```
'<iframe src="http://localhost:8080/jpivot/STPivot?query=default" style="width:100%;height:100%">
</iframe>'
```

4. Luego se habilitó un vínculo, para acceder a las funcionalidades que brinda el DM (desde el visor web STPivot), a través de una plantilla creada en el área de reportes del MRAM, que contiene el vínculo de acceso al reporte, al que se hace referencia.
5. Se realizaron algunas modificaciones en el diseño de la herramienta STPivot, para proporcionarle una apariencia similar a las interfaces del MRAM de Synta. Como parte de este paso, se modificaron las hojas de estilo en cascada (CSS, por sus siglas en inglés), para cambiar los elementos de los reportes. También se modificaron los nombres de las funcionalidades, para una mayor comprensión por parte del usuario.

A continuación, se muestra el resultado de la integración, una vez realizados los pasos anteriores.

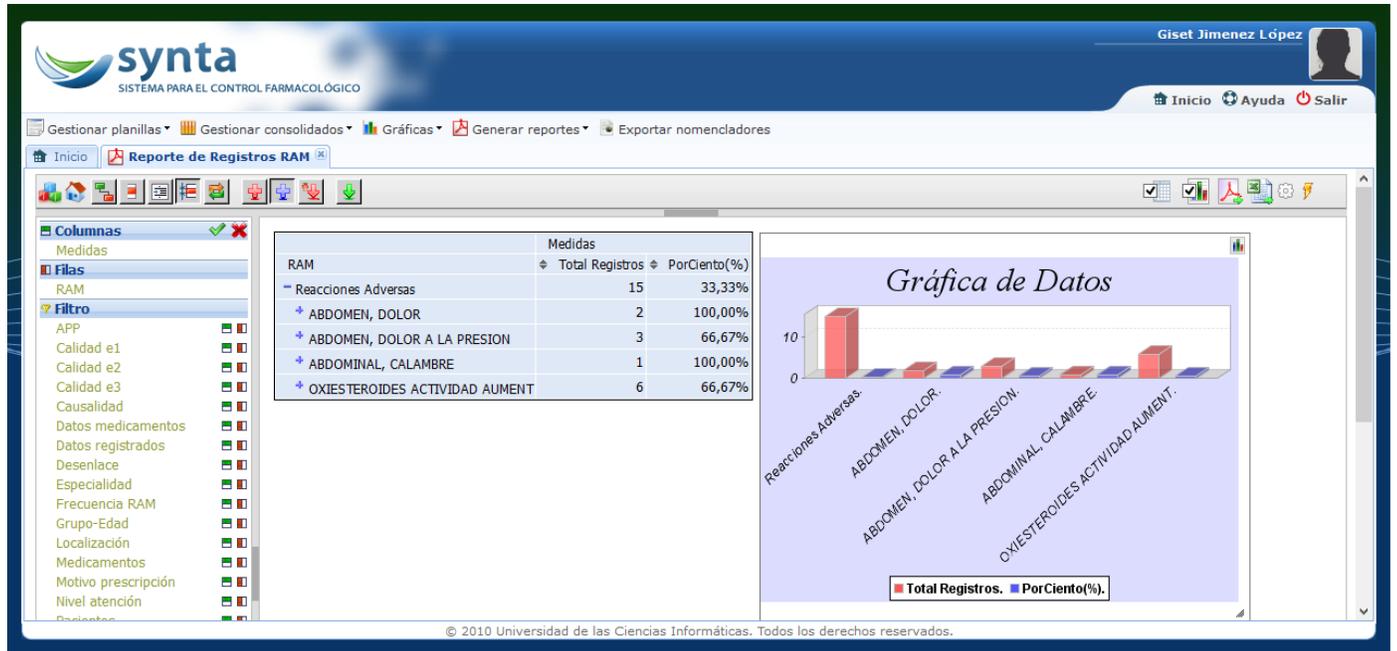


Figura 22. Reporte del STPivot desde el MRAM (elaboración propia)

### Conclusiones

Al término de la investigación en curso, se concluye lo siguiente:

- ✓ Se realizó un análisis del estado del arte de los principales sistemas de soporte a la toma de decisiones a nivel internacional y nacional, haciendo énfasis en Data Warehouses y Data Marts. Aunque no se encontró un sistema en el área de la salud, que cumpliera con las características necesarias de la investigación, el análisis permitió identificar las tendencias actuales en cuanto al diseño y construcción de sistemas de soporte a la toma de decisiones.
- ✓ Se realizó un conjunto de preguntas a los farmacoepidemiólogos, identificando los principales requerimientos de información. Este paso es fundamental en la construcción del Data Mart, pues permitió la obtención de los indicadores y perspectivas, que constituyen el punto de partida en este tipo de sistema.
- ✓ Al utilizar las herramientas, tecnologías y metodologías seleccionadas para el desarrollo del Data Mart, se demostró que son factibles para personas que tienen poca experiencia en este campo.
- ✓ Se elaboró el diseño y la implementación del sistema utilizando la metodología Hefesto, la cual permitió, a través de cada una de sus fases, el desarrollo del Data Mart de forma sencilla, ordenada e intuitiva; posibilitando una mejor comprensión a las personas con poca experiencia en el desarrollo de Data Marts.
- ✓ Al integrar el Data Mart al Módulo de Reacciones Adversas a Medicamentos, se demostró que es posible incorporar el cliente STPivot a aplicaciones web, a pesar de que funcionen en servidores web diferentes.



### Recomendaciones

Luego de la experiencia adquirida durante la realización de la investigación, se recomienda:

- ✓ Mejorar las interfaces del cliente STPivot, en cuanto al estilo de: la tipografía, imágenes, color y forma de las tablas; para lograr un diseño homogéneo con respecto al Módulo de Reacciones Adversas a Medicamentos del producto Synta.



### Referencias Bibliográficas

1. (Aguilar, 2011) Aguilar, Yuniesky Nueva. 2011. *Construcción de un Data Mart para el Sistema Integral de Gestión de Medicamentos*. La Habana : Trabajo de Diploma, 2011.
2. (Apache, 2011) The Apache Software Foundation. 2011. Apache Tomcat. *Apache Tomcat*. [En línea] noviembre de 2011. [Citado el: 11 de febrero de 2013.] <http://tomcat.apache.org/>.
3. (Asenjo, 2009) Asenjo, Jorge Sánchez. 2009. Sistema Gestores de Base de Datos. *Sistema Gestores de Base de Datos*. [En línea] 2009. [Citado el: 4 de febrero de 2013.] <http://ubuntuone.com/p/sqt/>.
4. (Bartomeus, 2009) Bartomeus, Pol Rojas. 2009. Portal d'accés obert al coneixement de la UPC. Portal d'accés obert al coneixement de la UPC. [En línea] septiembre de 2009. [Citado el: 25 de enero de 2013.] <http://upcommons.upc.edu/pfc/bitstream/2099.1/7461/4/59744.pdf.txt>.
5. (Bernabeu, 2009) Bernabeu, R. Dario. 2009. DataPRIX Datawarehouse Manager. *DataPRIX Datawarehouse Manager*. [En línea] DataPRIX, 22 de enero de 2009. [Citado el: 7 de mayo de 2013.] [www.dataprix.net/gl/datawarehouse-manager-34#x1-550003.4.7.1](http://www.dataprix.net/gl/datawarehouse-manager-34#x1-550003.4.7.1).
6. (Bernabeu, 2009) Dario, Bernabeu Ricardo. 2009. *HEFESTO: Metodología propia para la Construcción de un Data Warehouse*. 1ra. Córdoba : s.n., 2009. pág. 27. s.n.
7. (Bernabeu, 2010) Bernabeu, R. Dario. 2010. pentaho community. *Introducción a Spoon*. [En línea] 4 de enero de 2010. [Citado el: 14 de marzo de 2013.] <http://wiki.pentaho.com/pages/viewpage.action?pageId=14844841>.
8. (BI, 2011) BI. 2011. El servidor OLAP Mondrian. *El servidor OLAP Mondrian*. [En línea] 2011. [Citado el: 8 de marzo de 2013.] <http://pentaho.almacen-datos.com/mondrian.html>.
9. (Curto, 2006) Curto, Josep. 2006. DW: definiciones de Inmon y Kimball. DW: definiciones de Inmon y Kimball. [En línea] Information Management, 28 de 11 de 2006. [Citado el: 20 de 10 de 2012.] <http://informationmanagement.wordpress.com/2006/11/28/dw-definiciones-de-inmon-y-kimball/>.

10. (Dario, 2010) Dario, Bernabeu Ricardo. 2010. Hefesto. [En línea] 7 de 2010. <http://es.scribd.com/doc/49131495/HEFESTO-V2-0>.
11. (Febles, y otros, 2011) Febles, Juan Pedro, Senti, Vivian Estrada y Pérez, Efraín Sánchez. 2011. Biblioteca Virtual de las Ciencias en Cuba. Biblioteca Virtual de las Ciencias en Cuba. [En línea] 2011. [Citado el: 5 de 11 de 2012.] <http://www.bibliociencias.cu/gsd/collect/libros/index/assoc/HASH0106/b6fac6b9.dir/doc.pdf>.
12. (Ferrera, et al., 2010) Ferrera, Miriam Cecilia Gómez y Domínguez, Joel Fuentes. 2010. La Tecnología informática en la sociedad cubana. [ed.] Juana Ivis Imamura Díaz. Ciudad de la Habana : s.n., 2010. <http://www.pedagogiaprofesional.rimed.cu/especial%20/Joe.htm>.
13. (HALLO, et al., 2006) HALLO, MARIA, NARANJO, MICHAEL WLADIMIR VIZUETE y SHININ, CARLOS PATRICIO YELA. 2006. DSpace Escuela Politécnica Nacional. DSpace Escuela Politécnica Nacional. [En línea] marzo de 2006. [Citado el: 12 de enero de 2013.] <http://bibdigital.epn.edu.ec/handle/15000/177>.
14. (Hinz, et al., 2011) Hinz, Stefan, y otros. 2011. Capítulo 7. Optimización de MySQL. *Capítulo 7. Optimización de MySQL*. [En línea] 2011. [Citado el: 10 de febrero de 2013.] <http://dev.mysql.com/doc/refman/5.0/es/mysql-optimization.html>.
15. (Humble, et al., 2011) Humble, Charles y Hirsch, Jai. 2011. Olap4j 1.0: a Java API for OLAP Servers. *Olap4j 1.0: a Java API for OLAP Servers*. [En línea] 24 de junio de 2011. [Citado el: 6 de mayo de 2013.] <http://www.infoq.com/news/2011/06/olap4j>.
16. (Juárez, 2012) Juárez, Zoraya Catalina Flores. 2012. OLAP (OnLine Analytical Processing). *OLAP (OnLine Analytical Processing)*. [En línea] septiembre de 2012. [Citado el: 12 de mayo de 2013.] <http://basesdatoscms.files.wordpress.com/2012/09/olap-resumen1.pdf>.
17. (Kimball, 2002) Kimball, Ralph. 2002. The Data Warehouse Toolkit. s.l. : Wiley Computer Publishing, 2002. ISBN-10: 0471200247.
18. (Lazaro, 2012) Lazaro, Rigoberto. 2012. Data Warehouse. Data Warehouse. [En línea] 18 de septiembre de 2012. [Citado el: 12 de diciembre de 2012.] [http://rigobertlc.blogspot.com/2012\\_09\\_01\\_archive.html](http://rigobertlc.blogspot.com/2012_09_01_archive.html).
19. (López, 2012) López, Francisco. 2012. Hispasec. *Hispasec*. [En línea] 7 de noviembre de 2012. [Citado el: 10 de marzo de 2013.]

- <http://unaaldia.hispasec.com/2012/12/solucionadas-tres-vulnerabilidades-en.html>.
20. (López, et al., 2013) *Datamart para la Unidad Central de Cooperación Médica*. López, Leydis Hidalgo y Cartaya, Alier Caballero. 2013. 2, La Habana : s.n., 15 de febrero de 2013, Serie Científica de la Universidad de las Ciencias Informáticas, Vol. 6, pág. 10. ISSN: 2306-2495.
21. (Luis, 2008) Luis, Jose. 2008. Clover.ETL . *Clover.ETL* . [En línea] 21 de octubre de 2008. [Citado el: 21 de febrero de 2013.] <http://macloc.blogspot.com/2008/10/cloveretl.html>.
22. (Molier, 2005) Molier, Francisco Javier López. 2005. Grupo A y B Informatica :Bloque específico. [ed.] S.L. EDITORIAL MAD. Valencia : EDITORIAL MAD, S.L., 2005. págs. 173,174,175. Vol. II. ISBN Obra Completa:84-665-2080-5.
23. (Mondrian, 2007) Mondrian Project. 2007. Mondrian. Pentaho Analysis Services: Mondrian Project. *Mondrian. Pentaho Analysis Services: Mondrian Project*. [En línea] 4 de septiembre de 2007. [Citado el: 20 de febrero de 2013.] <http://mondrian.pentaho.com/>.
24. (Morales, 2012) Morales, Ana Cristina González. 2012. Herramientas y suites de IE. *Herramientas y suites de IE*. [En línea] 18 de junio de 2012. [Citado el: 20 de enero de 2013.] [http://catarina.udlap.mx/u\\_dl\\_a/tales/documentos/lis/gonzalez\\_m\\_ac/capitulo3.pdf](http://catarina.udlap.mx/u_dl_a/tales/documentos/lis/gonzalez_m_ac/capitulo3.pdf).
25. (Morales, 2012) Morales, Ana Cristina González. 2012. Herramientas y suites de IE. *Herramientas y suites de IE*. [En línea] 18 de junio de 2012. [Citado el: 20 de enero de 2013.] [http://catarina.udlap.mx/u\\_dl\\_a/tales/documentos/lis/gonzalez\\_m\\_ac/capitulo3.pdf](http://catarina.udlap.mx/u_dl_a/tales/documentos/lis/gonzalez_m_ac/capitulo3.pdf).
26. (Ojeda, 2008) Ojeda, Álvaro Villanueva. 2008. Análisis, Diseño e Implementación de un DataWarehouse de Soporte de Decisiones para un Hospital del Sistema de Salud Público. *Análisis, Diseño e Implementación de un DataWarehouse de Soporte de Decisiones para un Hospital del Sistema de Salud Público*. [En línea] 2008. [Citado el: 7 de mayo de 2013.] [http://tesis.pucp.edu.pe/repositorio/bitstream/handle/123456789/333/VILLANUEVA\\_%c3%81LVARO\\_AN%c3%81LISIS\\_DISE%c3%91O\\_E\\_IMPLEMENTACI%c3%93N\\_DE\\_UN\\_DATAWAREHOUSE\\_DE\\_SOPORTE\\_DE\\_DECISIONES\\_PARA\\_UN\\_HOSPITAL\\_DEL\\_SISTEMA\\_DE\\_SALUD\\_P%c3%9aBLICO.pdf?sequence=1](http://tesis.pucp.edu.pe/repositorio/bitstream/handle/123456789/333/VILLANUEVA_%c3%81LVARO_AN%c3%81LISIS_DISE%c3%91O_E_IMPLEMENTACI%c3%93N_DE_UN_DATAWAREHOUSE_DE_SOPORTE_DE_DECISIONES_PARA_UN_HOSPITAL_DEL_SISTEMA_DE_SALUD_P%c3%9aBLICO.pdf?sequence=1).

27. (Orallo, y otros, 2010) Orallo, José Hernández y Ramírez, César Ferri. 2010. Minería de Datos y Extracción de Conocimiento de Bases de Datos. Minería de Datos y Extracción de Conocimiento de Bases de Datos. [En línea] 2010. [Citado el: 20 de junio de 2012.] <http://users.dsic.upv.es/~jorallo/docent/doctorat/t2a.pdf>.
28. (Pecos, 2010) Pecos, Daniel. 2010. PostGreSQL vs. MySQL. *PostGreSQL vs. MySQL*. [En línea] 2010. [Citado el: 20 de marzo de 2012.] [http://www.danielpecos.com/docs/mysql\\_postgres/x57.html](http://www.danielpecos.com/docs/mysql_postgres/x57.html).
29. (Pentaho, 2013) Pentaho. 2013. pentaho. *Pentaho Data Integration*. [En línea] 2013. [Citado el: 5 de diciembre de 2012.] <http://www.pentaho.com/explore/pentaho-data-integration/>.
30. (Pentaho, 2013a) Pentaho .2013a. pentaho. *Pentaho and the Future of Analytics*. [En línea] 2013a. [Citado el: 14 de febrero de 2013.] <http://www.pentaho.com/explore/>.
31. (PostgreSQL, 2013) PostgreSQL Global Development Group. 2013. PostgreSQL 8.3.23 Documentation. *PostgreSQL 8.3.23 Documentation*. [En línea] 2013. [Citado el: 2 de mayo de 2013.] <http://www.postgresql.org/docs/8.3/static/intro-what-is.html>.
32. (Ramírez, 2011) Procedimiento para el desarrollo de un sistema de inteligencia de negocios en la gestión de ensayos clínicos en el Centro de Inmunología Molecular. Ramírez, Martha Denia Hernández. 2011. 4, La Habana : Editorial Ciencias Médicas, 2011, Información en Ciencias de la Salud, Vol. 22. <http://www.acimed.sld.cu/index.php/acimed/rt/prinFRIENDLY/208/168>. ISSN:2307-2113
33. (Ramos, et al., 2009) Ramos, Ariel Delgado, Hernández, Mirna Cabrera y Díaz, Alfredo Rodríguez. 2009. Informática en Salud 2009. Informática en Salud 2009. [En línea] Infomed, febrero de 2009. [Citado el: 12 de 10 de 2012.] <http://informatica2009.sld.cu/Members/mirnacabrera/estrategia-de-informatizacion-del-sistema-nacional-de-salud/>.
34. (Ramos, 2009) Ramos, Jonathan David Nima. 2009. Enciclopedia y Biblioteca Virtual. Enciclopedia y Biblioteca Virtual. [En línea] 2009. [Citado el: 10 de 1 de 2013.] <http://eumed.net/>.
35. (Rizo, et al., 2008) Rodríguez Sanz, Miguel. 2010. Análisis y diseño de un Data Mart para el seguimiento académico de alumnos en un entorno universitario. [En línea] 22 de 7 de 2010. [Citado el: 30 de 11 de 2010.] <http://es.scribd.com/doc/52203545/16/Metodologia-propuesta-por-Bill-Inmon>.

36. (Rivadera, 2010) Rivadera, Gustavo R. 2010. *La metodología de Kimball para el diseño de almacenes de datos (Data warehouses)*. Universidad Católica de Salta. Argentina : s.n., 2010. De Curso. <http://www.ucasal.net/templates/unid-academicas/ingenieria/apps/5-p56-rivadera-formateado.pdf>.
37. (Rodríguez, et al., 2002) Aplicación de la minería de datos en la bioinformática. Rodríguez, Juan Pedro Febles y Pérez, Abel González. 2010. 2, Ciudad de la Habana : Scielo, 2010, Vol. 10. ISSN 1024-9435.
38. (Sánchez, 2011) Sánchez, José A Alférez. 2011. INSTALACIÓN, CONFIGURACIÓN Y ADMINISTRACIÓN DEL SERVIDOR DE APLICACIONES JBOSS. *INSTALACIÓN, CONFIGURACIÓN Y ADMINISTRACIÓN DEL SERVIDOR DE APLICACIONES JBOSS*. [En línea] enero de 2011. [Citado el: 5 de mayo de 2013.] <http://www.alferez.es/documentos/Jboss.pdf>.
39. (Sanz, 2010) Sanz, Miguel Rodríguez. 2010. Análisis y diseño de un Data Mart para el seguimiento académico de alumnos en un entorno universitario. Análisis y diseño de un Data Mart para el seguimiento académico de alumnos en un entorno universitario. [En línea] 22 de julio de 2010. [Citado el: 30 de febrero de 2013.] <http://es.scribd.com/doc/52203545/16/Metodologia-propuesta-por-Bill-Inmon>.
40. (Sedano, 2009) Sedano, Carmen Pamela Rosales. 2009. Repositorio Digital de Tesis PUCP. Repositorio Digital de Tesis PUCP. [En línea] enero de 2009. [Citado el: 9 de diciembre de 2013.] Tesis. [http://tesis.pucp.edu.pe/repositorio/bitstream/handle/123456789/1379/ROSALES\\_SEDANO\\_CARMEN\\_DATAMART\\_TOMA\\_DECISIONES.pdf?sequence=1](http://tesis.pucp.edu.pe/repositorio/bitstream/handle/123456789/1379/ROSALES_SEDANO_CARMEN_DATAMART_TOMA_DECISIONES.pdf?sequence=1).
41. (Softpedia, 2012) Softpedia. 2012. Enhydra Octopus. *Enhydra Octopus*. [En línea] diciembre de 2012. [Citado el: 20 de marzo de 2013.] <http://www.softpedia.es/programa-Enhydra-Octopus-149669.html>.
42. (STPivot, 2011) STPivot. 2011. Strate bi open business intelligence. *STPivot*. [En línea] 2011. [Citado el: 18 de mayo de 2013.] <http://www.stratebi.com/stpivot>.
43. (Sparx, 2008) Sparx. 2008. Enterprise Architect. *Enterprise Architect*. [En línea] Sparx, 2008. [Citado el: 19 de marzo de 2013.] <http://www.sparxsystems.es/New/products/ea.html>.
44. (Suesca, 2012) Suesca, Fabian. 2012. Prezi. *Prezi*. [En línea] 18 de octubre de 2012. [Citado el: 8 de mayo de 2013.] <http://prezi.com/hlwhjshhamga/bases-de-datos->

multidimensionales/.

45. (Tamayo, et al., 2006) Tamayo, Marysol y Moreno, Francisco Javier. 2006. Análisis del modelo de almacenamiento MOLAP frente al modelo de almacenamiento ROLAP. [En línea] 12 de 2008. [http://www.scielo.org.co/scielo.php?pid=S0120-56092006000300016&script=sci\\_arttext&tlng=pt](http://www.scielo.org.co/scielo.php?pid=S0120-56092006000300016&script=sci_arttext&tlng=pt).
46. (Vázquez, et al., 2003) Vázquez, Rodolfo, y otros. 2003. Sistemas de Información para la Gestión de Instituciones de Salud. Teoría y aplicaciones en el Uruguay. *Sistemas de Información para la Gestión de Instituciones de Salud. Teoría y aplicaciones en el Uruguay*. [En línea] 21 de octubre de 2003. [Citado el: 10 de mayo de 2013.] [http://www.femisaluddigital.org.uy/admin/files/femi/from\\_old\\_website/LibroSIS\\_-\\_SIGs\\_Data\\_Warehouses.doc](http://www.femisaluddigital.org.uy/admin/files/femi/from_old_website/LibroSIS_-_SIGs_Data_Warehouses.doc).



## Bibliografía

1. **Aguilar, Yuniesky Nueva. 2011.** *Construcción de un Data Mart para el Sistema Integral de Gestión de Medicamentos.* La Habana : Trabajo de Diploma, 2011.
2. *Análisis del modelo de almacenamiento MOLAP frente al modelo de almacenamiento ROLAP .* **Tamayo, Marysol y Moreno, Francisco Javier. 2006.** 3, Bogotá : s.n., 12 de 2006, Ingeniería e Investigación, Vol. 26. ISSN 0120-5609.
3. *Análisis del modelo de almacenamiento MOLAP frente al modelo de almacenamiento ROLAP.* **Tamayo, Marysol y Moreno, Francisco Javier. 2006.** 3, Colombia : Redalyc.org, 3 de diciembre de 2006, Sistema de Información Científica, Vol. 26.  
<http://www.redalyc.org/articulo.oa?id=64326317>. ISSN (Versión impresa): 0120-5609.
4. **Apache. 2011.** The Apache Software Foundation. *Apache Tomcat.* [En línea] noviembre de 2011. [Citado el: 11 de febrero de 2013.] <http://tomcat.apache.org/>.
5. *Aplicación de la minería de datos en la bioinformática.* **Rodríguez, Juan Pedro Febles y Pérez, Abel González. 2002.** [ed.] Scielo. 2, Ciudad de La Habana : s.n., marzo-abril de 2002, ACIMED, Vol. 10. [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1024-94352002000200003](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352002000200003). ISSN 1024-9435.
6. **Asenjo, Jorge Sánchez. 2009.** Sistema Gestores de Base de Datos. *Apuntes Completos Sistemas de Gestores de Bases de Datos.* [En línea] 2009. [Citado el: 4 de febrero de 2013.] <http://ubuntuone.com/p/sqf/>.
7. **Bartomeus, Pol Rojas. 2009.** Portal d'accés obert al coneixement de la UPC. *Estudio comparativo de bases de datos analíticas.* [En línea] septiembre de 2009. [Citado el: 25 de enero de 2013.] <http://upcommons.upc.edu/pfc/bitstream/2099.1/7461/4/59744.pdf.txt>.
8. **Bernabeu, R. Dario. 2009.** DataPrix Datawarehouse Manager. *DataPrix Datawarehouse Manager.* [En línea] DataPrix, 22 de enero de 2009. [Citado el: 7 de mayo de 2013.] [www.dataprix.net/gl/datawarehouse-manager-34#x1-550003.4.7.1](http://www.dataprix.net/gl/datawarehouse-manager-34#x1-550003.4.7.1).
9. —. **2010.** pentaho community. *Introducción a Spoon.* [En línea] 4 de enero de 2010. [Citado el: 14 de marzo de 2013.] <http://wiki.pentaho.com/pages/viewpage.action?pageId=14844841>.
10. **Bernabeu, Ricardo Dario. 2010.** Data Prix. *Data Prix.* [En línea] julio de 2010. [Citado el: 25 de marzo de 2013.] <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/-metodologia-hefesto/52-descripcion>.

11. —. **2010**. *DATA WAREHOUSING: Investigación y Sistematización de Conceptos - HEFESTO: Metodología propia para la Construcción de un Data Warehouse*. Córdoba : s.n., 2010. pág. 160. s.n.
12. —. **2009**. *HEFESTO: Metodología propia para la Construcción de un Data Warehouse*. 1ra. Córdoba : s.n., 2009. pág. 27. s.n.
13. **BI, Colectivo de autores de. 2011**. El servidor OLAP Mondrian. *El servidor OLAP Mondrian*. [En línea] 2011. [Citado el: 8 de marzo de 2013.] <http://pentaho.almacen-datos.com/mondrian.html>.
14. **Cueto Ramón, Ariagna. García Díaz, Joannis. 2008**. *Implementación de un Data Warehouse para el control del Recurso Humano de la Salud*. Ciudad de la Habana : s.n., 2008.
15. **Curto, Josep. 2006**. DW: definiciones de Inmon y Kimball. *DW: definiciones de Inmon y Kimball*. [En línea] Information Management, 28 de 11 de 2006. [Citado el: 20 de 10 de 2012.] <http://informationmanagement.wordpress.com/2006/11/28/dw-definiciones-de-inmon-y-kimball/>.
16. **Dario, Bernabeu Ricardo. 2010**. Hefesto. [En línea] 7 de 2010. <http://es.scribd.com/doc/49131495/HEFESTO-V2-0>.
17. —. **2010**. *HEFESTO: Metodología propia para la Construcción de un Data Warehouse V2.0*. Córdoba : s.n., 2010. pág. 162. s.n..
18. —. **2009**. Metodología Hefesto. *Metodología Hefesto*. [En línea] 7 de mayo de 2009. [Citado el: 20 de marzo de 2013.] <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/-metodologia-hefesto/51-introduccion>.
19. *Datamart para la Unidad Central de Cooperación Médica*. **López, Leydis Hidalgo y Cartaya, Alier Caballero. 2013**. 2, La Habana : s.n., 15 de febrero de 2013, Serie Científica de la Universidad de las Ciencias Informáticas, Vol. 6, pág. 10. ISSN: 2306-2495.
20. **Díaz, Josep Curto. 2010**. OLAP. [En línea] 6 de diciembre de 2010. [Citado el: 10 de febrero de 2013.] [http://www.tecnologiahechapalabra.com/tecnologia/glosario\\_tecnico/articulo.asp?i=5249](http://www.tecnologiahechapalabra.com/tecnologia/glosario_tecnico/articulo.asp?i=5249).
21. **Díaz, Samuel Oporto. 2010**. Documents & Resources for Small Businesses & Professionals. *Documents & Resources for Small Businesses & Professionals*. [En línea] 30 de agosto de 2010. [Citado el: 15 de enero de 2013.] <http://www.docstoc.com/docs/52595481/Metodolog%C3%ADas-para-el-Data-WareHousing>.
22. —. **2010**. Metodologías para el Data WareHousing. *Metodologías para el Data WareHousing*. [En línea] 30 de 8 de 2010. [Citado el: 10 de 9 de 2012.] [www.docstoc.com/docs/.../Metodologas-para-el-Data-WareHousing](http://www.docstoc.com/docs/.../Metodologas-para-el-Data-WareHousing).

23. **Febles, Juan Pedro, Senti, Vivian Estrada y Pérez, Efraín Sánchez. 2011.** Biblioteca Virtual de las Ciencias en Cuba. *Importancia de la utilización de un Data Warehouse (DW) en las empresas.* [En línea] 2011. [Citado el: 5 de 11 de 2012.]  
<http://www.bibliociencias.cu/gsd/collect/libros/index/assoc/HASH0106/b6fac6b9.dir/doc.pdf>.
24. **HALLO, MARIA, NARANJO, MICHAEL WLADIMIR VIZUETE y SHININ, CARLOS PATRICIO YELA. 2006.** *Análisis, Diseño e Implementación de un Data Mart para el Área de Sismología del Departamento de Geofísica de la Escuela Politécnica Nacional.* 2006. pág. 169.
25. **Hinz, Stefan, y otros. 2011.** Capítulo 7. Optimización de MySQL. *Capítulo 7. Optimización de MySQL.* [En línea] 2011. [Citado el: 10 de febrero de 2013.]  
<http://dev.mysql.com/doc/refman/5.0/es/mysql-optimization.html>.
26. **Humble, Charles y Hirsch, Jai. 2011.** Olap4j 1.0: a Java API for OLAP Servers. *Olap4j 1.0: a Java API for OLAP Servers.* [En línea] 24 de junio de 2011. [Citado el: 6 de mayo de 2013.]  
<http://www.infoq.com/news/2011/06/olap4j>.
27. **Interactiva, Saludalia. 2008.** Reacciones adversas a medicamentos. [En línea] 10 de 2008.  
<http://www.saludalia.com/alergias/reacciones-adversas-medicamentos>.
28. **Juárez, Zoraya Catalina Flores. 2012.** OLAP (OnLine Analytical Processing). *OLAP (OnLine Analytical Processing).* [En línea] septiembre de 2012. [Citado el: 12 de mayo de 2013.]  
<http://basesdatoscms.files.wordpress.com/2012/09/olap-resumen1.pdf>.
29. **Kimball, Ralph. 2002.** *The Data Warehouse Toolkit.* s.l. : Wiley Computer Publishing, 2002. ISBN-10: 0471200247.
30. *La Tecnología informática en la sociedad cubana.* **Ferrera, Miriam Cecilia Gómez y Domínguez, Joel Fuentes. 2010.** [ed.] Juana Ivis Imamura Díaz. Ciudad de la Habana : s.n., 2010, Pedagogía Profesional, Vols. Edición especial,2.  
<http://www.pedagogiaprofesional.rimed.cu/especial%202/Joe.htm>. ISSN 1684-5765.
31. **Lazaro, Rigoberto. 2012.** Data Warehouse. *Data Warehouse.* [En línea] 18 de septiembre de 2012. [Citado el: 12 de diciembre de 2012.]  
[http://rigobertlc.blogspot.com/2012\\_09\\_01\\_archive.html](http://rigobertlc.blogspot.com/2012_09_01_archive.html).
32. **López, Francisco. 2012.** Hispasec. *Hispasec.* [En línea] 7 de noviembre de 2012. [Citado el: 10 de marzo de 2013.] <http://unaaldia.hispasec.com/2012/12/solucionadas-tres-vulnerabilidades-en.html>.
33. **Luis, Jose. 2008.** M@LLOC. *Consejos sobre programación y algunas cosas sobre el mundo del IT.* [En línea] 21 de octubre de 2008. [Citado el: 21 de febrero de 2013.]  
<http://maclloc.blogspot.com/2008/10/cloveretl.html>.

34. **Méndez, Geidy Acosta. 2013.** Data Mart para la dirección de salud en Cuba. *Data Mart para la dirección de salud en Cuba*. [En línea] 2013. [Citado el: 10 de mayo de 2013.] <http://www.informaticahabana.cu/content/data-mart-health-management-cuba>.
35. **Molier, Francisco Javier López. 2005.** *Grupo A y B Informatica :Bloque específico*. [ed.] S.L. EDITORIAL MAD. Valencia : EDITORIAL MAD, S.L., 2005. págs. 173 - 175. Vol. II. ISBN Obra Completa:84-665-2080-5.
36. **Mondrian. 2007.** Mondrian. Pentaho Analysis Services: Mondrian Project. *Mondrian. Pentaho Analysis Services: Mondrian Project*. [En línea] 4 de septiembre de 2007. [Citado el: 20 de febrero de 2013.] <http://mondrian.pentaho.com/>.
37. **Morales, Ana Cristina González. 2012.** Herramientas y suites de IE. *Herramientas y suites de IE*. [En línea] 18 de junio de 2012. [Citado el: 20 de enero de 2013.] [http://catarina.udlap.mx/u\\_dl\\_a/tales/documentos/lis/gonzalez\\_m\\_ac/capitulo3.pdf](http://catarina.udlap.mx/u_dl_a/tales/documentos/lis/gonzalez_m_ac/capitulo3.pdf).
38. —. **2012.** Herramientas y suites de IE. *Herramientas y suites de IE*. [En línea] 18 de junio de 2012. [Citado el: 8 de marzo de 2013.] [http://catarina.udlap.mx/u\\_dl\\_a/tales/documentos/lis/gonzalez\\_m\\_ac/capitulo3.pdf](http://catarina.udlap.mx/u_dl_a/tales/documentos/lis/gonzalez_m_ac/capitulo3.pdf).
39. **Moran, Doug. 2012.** Pentaho Data Integration (Kettle) Tutorial. *Pentaho Data Integration (Kettle) Tutorial*. [En línea] 28 de enero de 2012. [Citado el: 10 de diciembre de 2012.] <http://wiki.pentaho.com/display/eai/pentaho+data+integration+%28kettle%29+tutorial>.
40. **Ojeda, Álvaro Villanueva. 2008.** *Análisis, Diseño e Implementación de un DataWarehouse de Soporte de Decisiones para un Hospital del Sistema de Salud Público*. FACULTAD DE CIENCIAS E INGENIERÍA, PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ. Perú : s.n., 2008. pág. 159, Tesis para optar por el Título de Ingeniero Informático.
41. **OMS. 2008.** Organización Mundial de Salud. *Nota descriptiva N°293* . [En línea] octubre de 2008. [Citado el: 12 de mayo de 2012.] <http://www.who.int/mediacentre/factsheets/fs293/es/>.
42. **Orallo, José Hernández y Ramírez, César Ferri. 2010.** Minería de Datos y Extracción de Conocimiento de Bases de Datos. *Minería de Datos y Extracción de Conocimiento de Bases de Datos*. [En línea] 2010. [Citado el: 20 de junio de 2012.] <http://users.dsic.upv.es/~jorallo/docent/doctorat/t2a.pdf>.
43. **Ordaz, Blvd. Diaz. 2013.** Características. *Características*. [En línea] 2013. [Citado el: 6 de mayo de 2013.] <http://www.gravitar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/>.
44. **Pecos, Daniel. 2010.** PostGreSQL vs. MySQL. *PostGreSQL vs. MySQL*. [En línea] 2010. [Citado el: 20 de marzo de 2012.] [http://www.danielpecos.com/docs/mysql\\_postgres/x57.html](http://www.danielpecos.com/docs/mysql_postgres/x57.html).

45. **Pedro Febles, Dr. Juan, y otros. 2011.** Biblioteca Virtual de las Ciencias en Cuba. *Biblioteca Virtual de las Ciencias en Cuba*. [En línea] 2011. [Citado el: 1 de 11 de 2012.] <http://www.bibliociencias.cu/gsd/collect/libros/index/assoc/HASH0106/b6fac6b9.dir/doc.pdf>.
46. **Pentaho. 2013.** pentaho. *Pentaho Data Integration*. [En línea] 2013. [Citado el: 5 de diciembre de 2012.] <http://www.pentaho.com/explore/pentaho-data-integration/>.
47. —. **2013a.** pentaho. *Pentaho and the Future of Analytics*. [En línea] 2013a. [Citado el: 14 de febrero de 2013.] <http://www.pentaho.com/explore/>.
48. **PostgreSQL. 2013.** PostgreSQL 8.3.23 Documentation. *PostgreSQL 8.3.23 Documentation*. [En línea] 2013. [Citado el: 2 de mayo de 2013.] <http://www.postgresql.org/docs/8.3/static/intro-what-is.html>.
49. *Procedimiento para el desarrollo de un sistema de inteligencia de negocios en la gestión de ensayos clínicos en el Centro de Inmunología Molecular.* **Ramírez, Martha Denia Hernández. 2011.** 4, La Habana : Ciencias Médicas, 2011, Vol. 22. ISSN:2307-2113.
50. **Ramos, Ariel Delgado, Hernández, Mirna Cabrera y Díaz, Alfredo Rodríguez. 2009.** Informática en Salud 2009. *Informática en Salud 2009*. [En línea] febrero de 2009. [Citado el: 12 de 10 de 2012.] [informatica2009.sld.cu/Members/mirnacabrera/estrategia-de-informatizacion-del-sistema-nacional-de-salud/at\\_download/trabajo](http://informatica2009.sld.cu/Members/mirnacabrera/estrategia-de-informatizacion-del-sistema-nacional-de-salud/at_download/trabajo).
51. **Ramos, Jonathan David Nima. 2009.** *Soluciones de Inteligencia de Negocios a su alcance: Fundamentos y casos de aplicación*. Lima : s.n., 2009. pág. 54.
52. **Rivadera, Gustavo R. 2010.** *La metodología de Kimball para el diseño de almacenes de datos (Data warehouses)*. Universidad Católica de Salta. Argentina : s.n., 2010. De Curso. <http://www.ucasal.net/templates/unid-academicas/ingenieria/apps/5-p56-rivadera-formateado.pdf>.
53. **Rizo Rizo, MSc. Emma R, y otros. 2009.** Biblioteca Virtual de las Ciencias en Cuba. *Biblioteca Virtual de las Ciencias en Cuba*. [En línea] 2009. [Citado el: 16 de 10 de 2012.] <http://www.bibliociencias.cu/gsd/collect/libros/index/assoc/HASH0106/b6fac6b9.dir/doc.pdf>.
54. **Rizo, Emma R Rizo, y otros. 2008.** Biblioteca Virtual de las Ciencias en Cuba. *Biblioteca Virtual de las Ciencias en Cuba*. [En línea] 2008. [Citado el: 16 de octubre de 2012.] <http://www.bibliociencias.cu/gsd/collect/libros/index/assoc/HASH0106/b6fac6b9.dir/doc.pdf>.
55. **Rodríguez Sanz, Miguel. 2010.** Análisis y diseño de un Data Mart para el seguimiento académico de alumnos en un entorno universitario. [En línea] 22 de 7 de 2010. [Citado el: 30 de 11 de 2010.] <http://es.scribd.com/doc/52203545/16/Metodologia-propuesta-por-Bill-Inmon>.

56. —. **2010**. e-archivo. *e-archivo*. [En línea] 22 de 7 de 2010. [Citado el: 5 de 1 de 2013.] <http://e-archivo.uc3m.es/handle/10016/9856>.
57. **Saludalia. 2008**. Reacciones adversas a medicamentos. *Reacciones adversas a medicamentos*. [En línea] 25 de octubre de 2008. [Citado el: 2 de junio de 2012.] <http://www.saludalia.com/alergias/reacciones-adversas-medicamentos>.
58. **Sánchez, José A Alférez. 2011**. INSTALACIÓN, CONFIGURACIÓN Y ADMINISTRACIÓN DEL SERVIDOR DE APLICACIONES JBOSS. *INSTALACIÓN, CONFIGURACIÓN Y ADMINISTRACIÓN DEL SERVIDOR DE APLICACIONES JBOSS*. [En línea] enero de 2011. [Citado el: 5 de mayo de 2013.] <http://www.alferez.es/documentos/Jboss.pdf>.
59. **Sanz, Miguel Rodríguez. 2010**. ANÁLISIS Y DISEÑO DE UN DATA MART PARA EL SEGUIMIENTO ACADÉMICO DE ALUMNOS EN UN ENTORNO UNIVERSITARIO. Universidad Carlos III de Madrid Escuela Politécnica Superior Ingeniería en Informática. 2010. pág. 136, Proyecto Fin de Carrera.
60. **Sedano, Carmen Pamela Rosales. 2009**. ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN DATAMART PARA EL SOPORTE DE TOMA DE DECISIONES Y EVALUACIÓN DE LAS ESTRATEGIAS SANITARIAS EN LAS DIRECCIONES DE SALUD. PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ. Perú : s.n., 2009. pág. 81, Tesis.
61. **Softpedia. 2012**. Softpedia. *Enhydra Octopus 3.2-2*. [En línea] diciembre de 2012. [Citado el: 20 de marzo de 2013.] <http://www.softpedia.es/programa-Enhydra-Octopus-149669.html>.
62. **Sparx. 2008**. Enterprise Architect. *Enterprise Architect*. [En línea] Sparx, 2008. [Citado el: 19 de marzo de 2013.] <http://www.sparxsystems.es/New/products/ea.html>.
63. **STPivot. 2011**. Strate bi open business intelligence. *STPivot*. [En línea] 2011. [Citado el: 18 de mayo de 2013.] <http://www.stratebi.com/stpivot>.
64. **Suesca, Fabian. 2012**. Prezi. *Prezi*. [En línea] 18 de octubre de 2012. [Citado el: 8 de mayo de 2013.] <http://prezi.com/hlwhjshhamga/bases-de-datos-multidimensionales/>.
65. **Vázquez, Rodolfo, y otros. 2003**. *Sistemas de Información para la Gestión de Instituciones de Salud*. Uruguay : s.n., 2003. pág. 30.
66. **Vega Torres, Ing. Lilliam, Rojas Díaz, Ing. Luis y Placeres Villar, Lic. Cecilia. 2008**. La inteligencia de negocio. [En línea] 2008. [www.redciencia.info.ve/memorias/ProyProsp/trabajos/l3.doc](http://www.redciencia.info.ve/memorias/ProyProsp/trabajos/l3.doc).

## Anexos

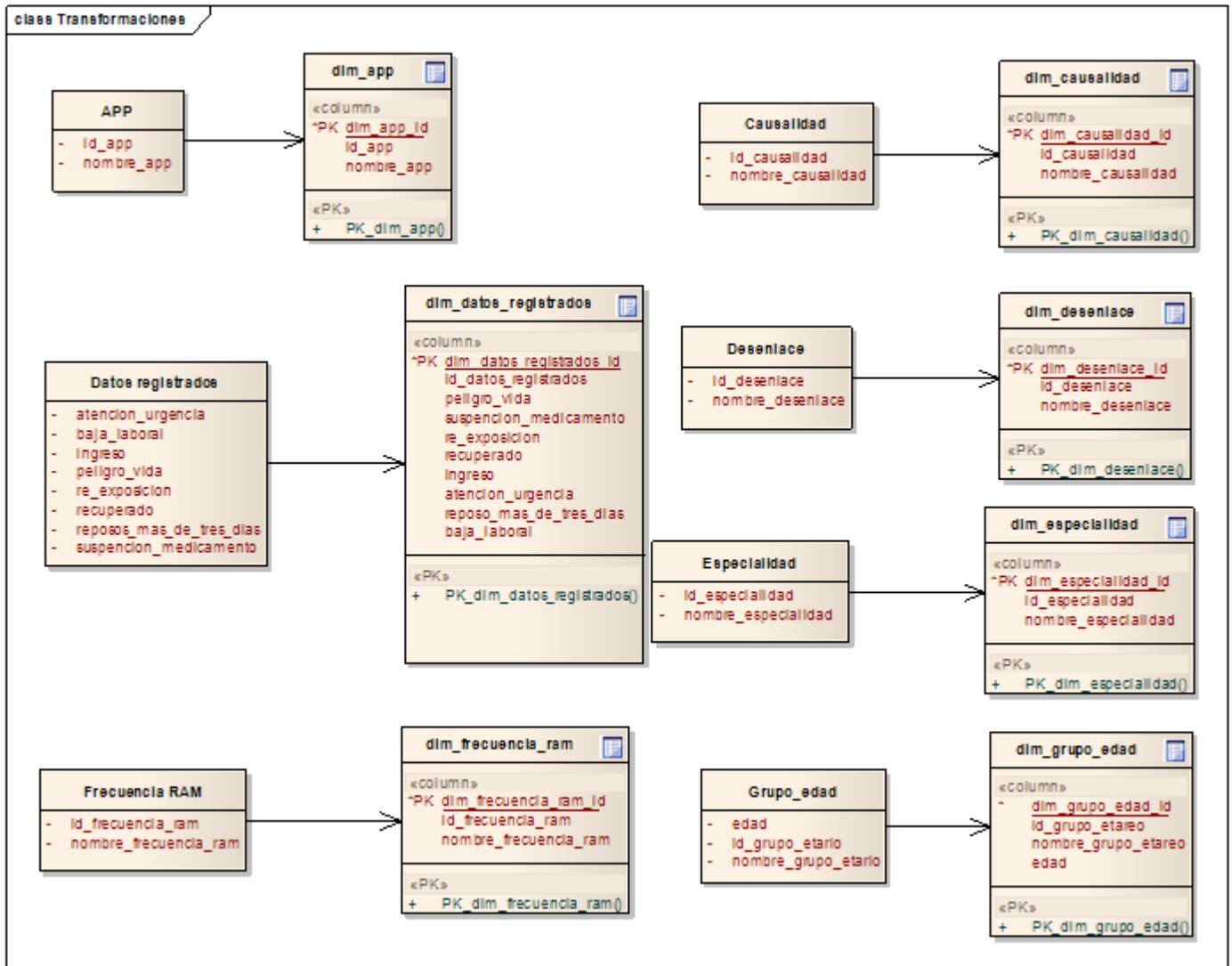
### Anexo 1. Entrevista realizada a los farmacoepidemiólogos

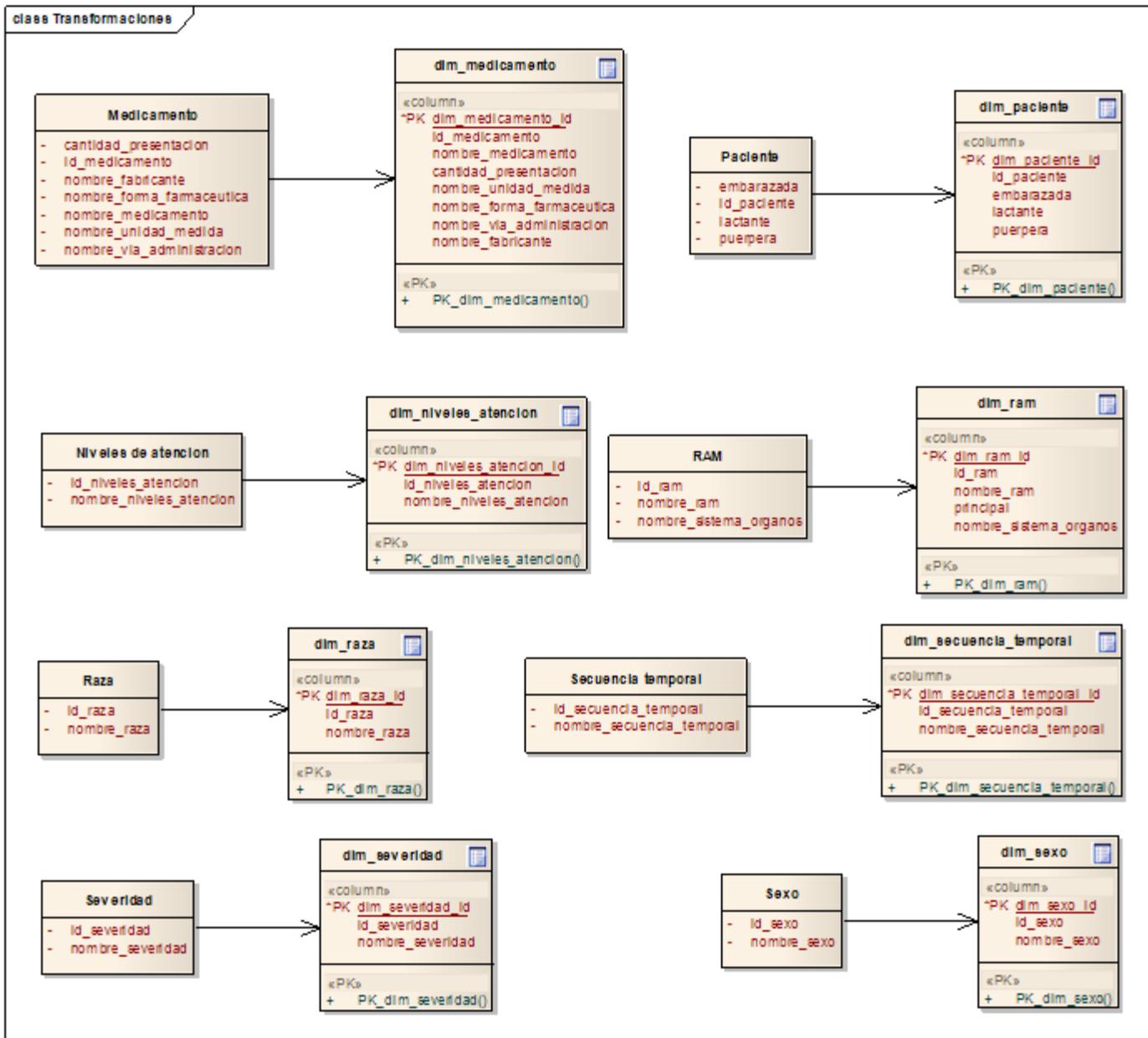
- ¿Qué indicadores desea medir de las RAM?
- ¿Qué características de las RAM, Ud. desea analizar?
- ¿Podría explicar qué información de la RAM desea triangular?
- ¿Una vez que Ud. analiza la información de las RAM, como le gustaría visualizarla? ¿Gráficas de pastel, de barra, lista de registros, etc.?

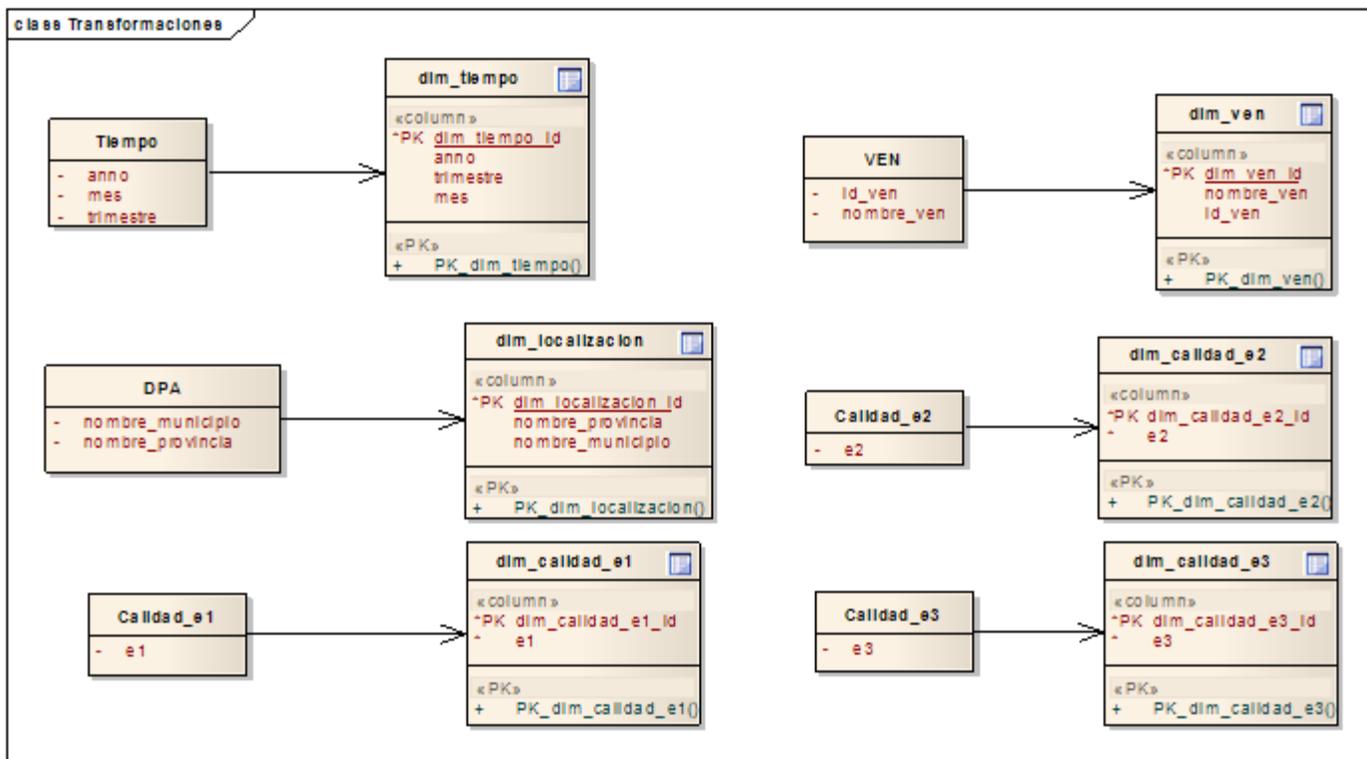




### Anexo 3. Tablas de dimensiones

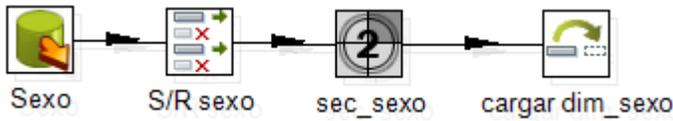






## Anexo 4. Transformaciones

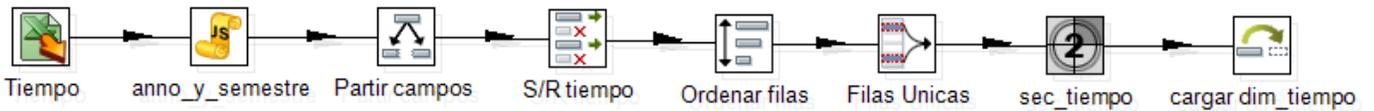
### Dimensión Sexo



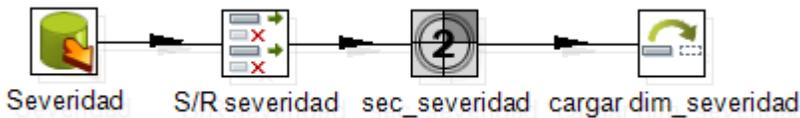
### Dimensión Raza



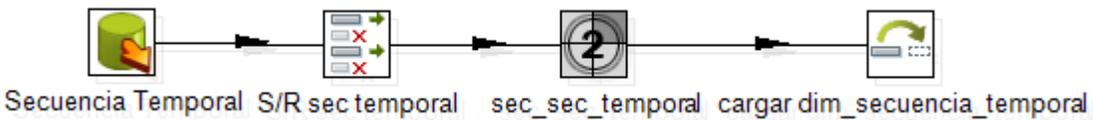
### Dimensión Tiempo



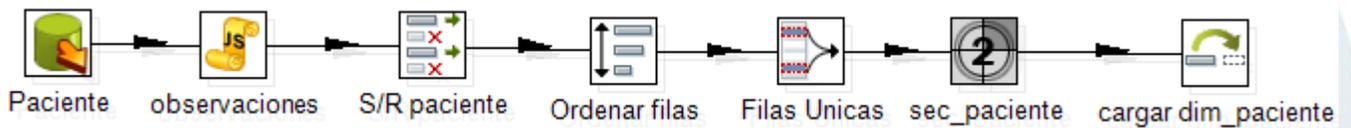
### Dimensión Severidad



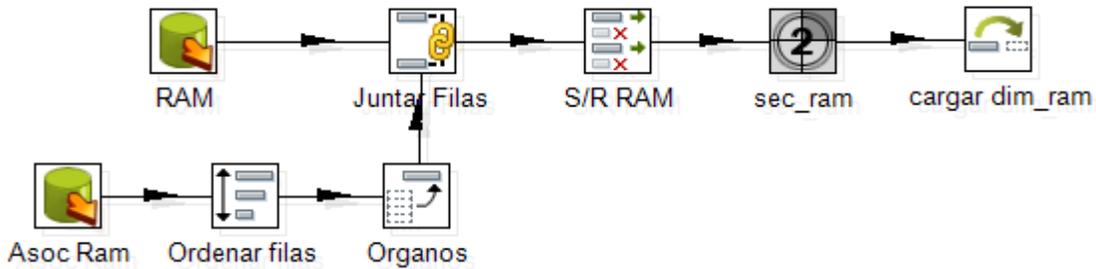
### Dimensión Secuencia Temporal



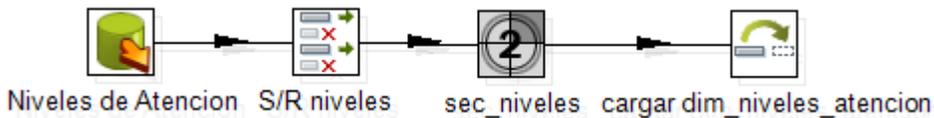
### Dimensión Paciente



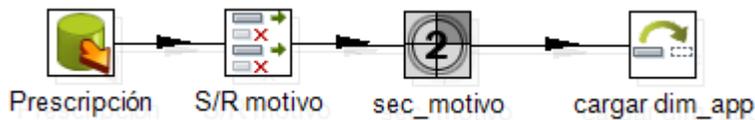
**Dimensión RAM**



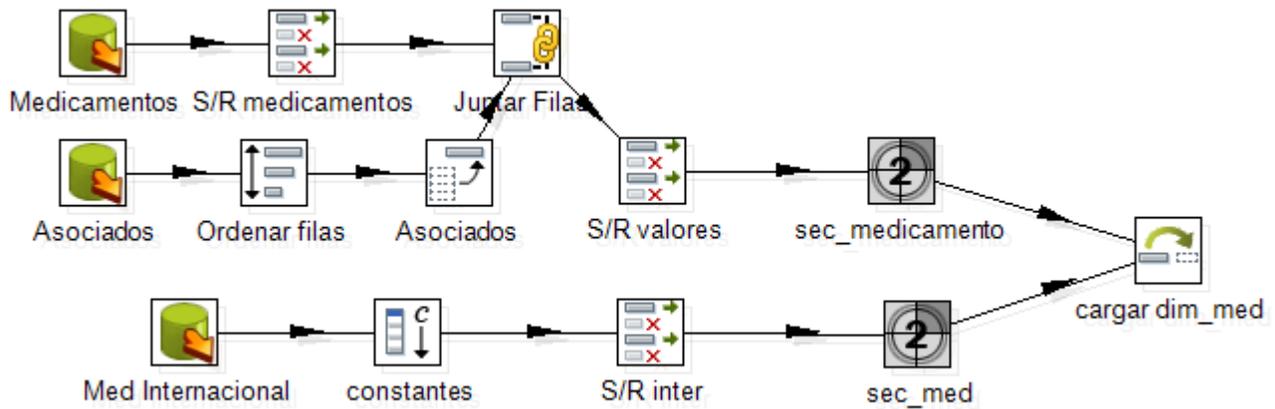
**Dimensión Niveles de atención**



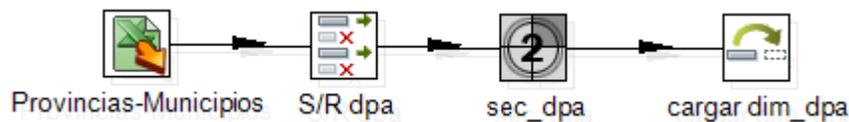
**Dimensión Motivo de prescripción**



**Dimensión Medicamento**



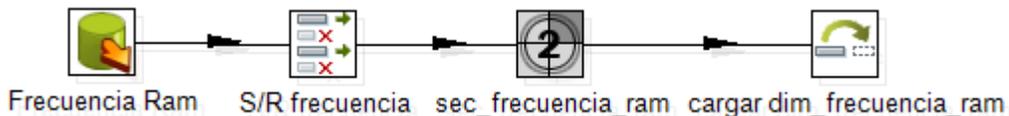
**Dimensión Localización**



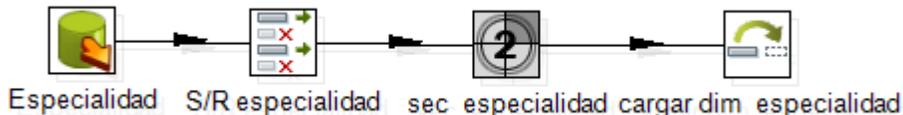
**Dimensión Grupo etario-edad**



**Dimensión Frecuencia RAM**



**Dimensión Especialidad**



**Dimensión APP**



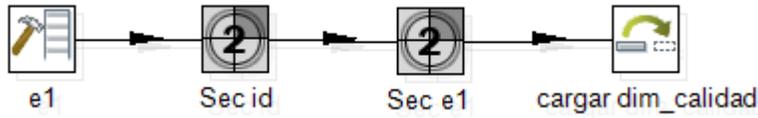
**Dimensión Causalidad**



**Dimensión Desenlace**



**Dimensión Calidad e1**



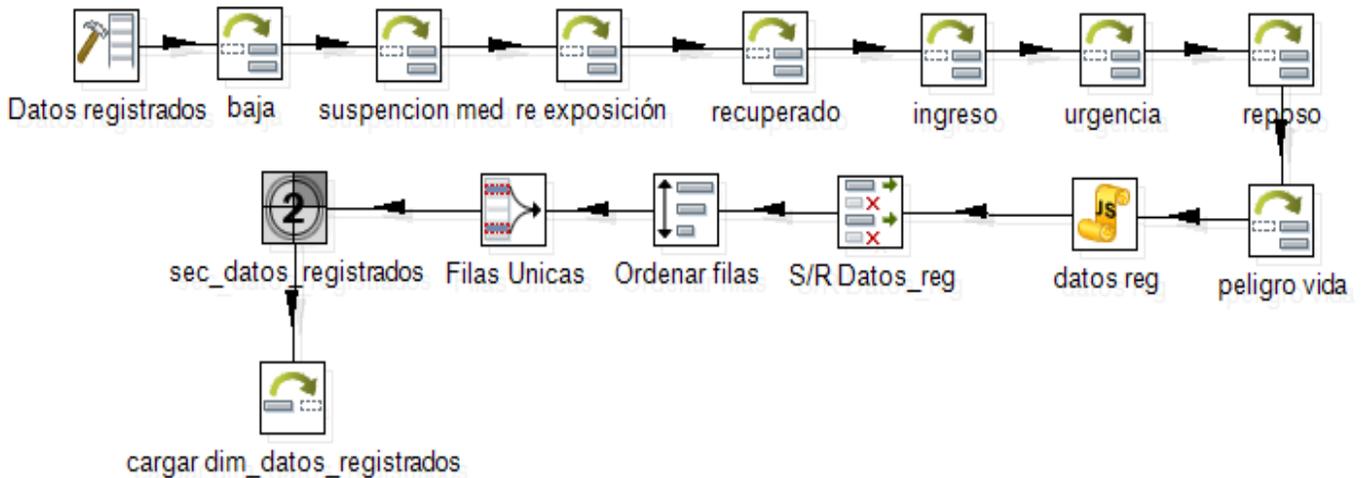
**Dimensión Calidad e2**



**Dimensión Calidad e3**



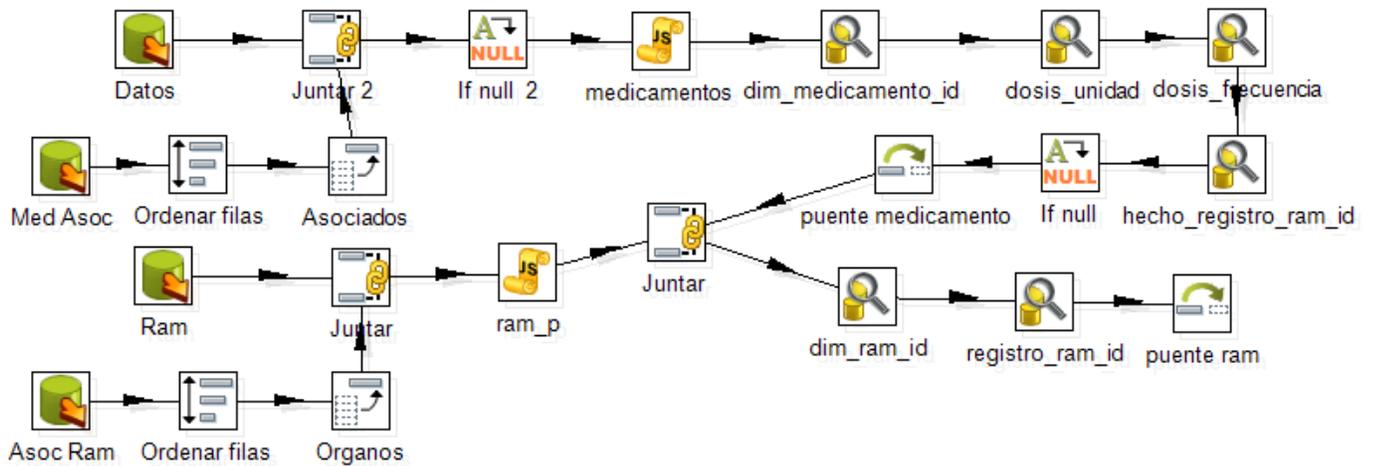
**Dimensión Datos registrados**



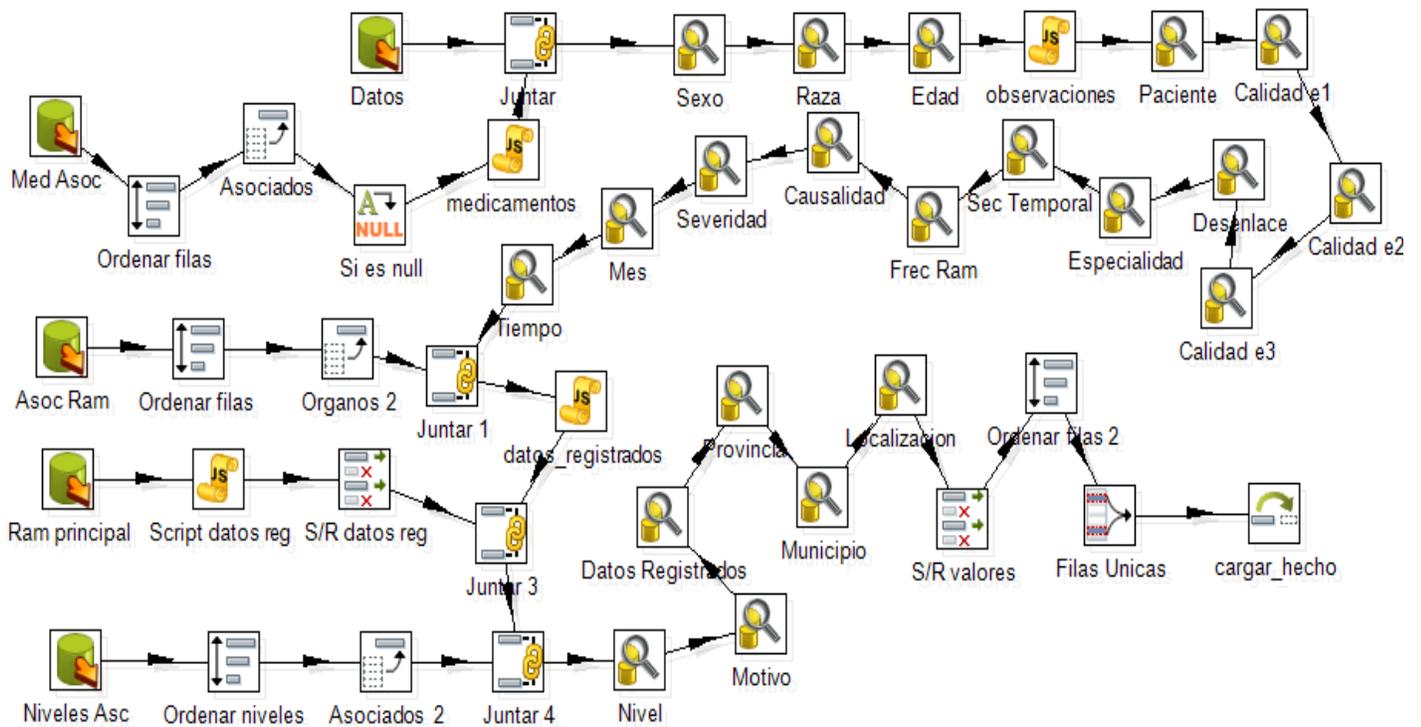
**Dimensión Puente APP**



### Dimensión Puente Medicamento-RAM



### Carga del hecho



## Anexo 5. Jerarquía de los atributos

