

Universidad de las Ciencias Informáticas

Facultad 2



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

**Mercado de datos para el análisis y la toma de decisiones referente al
consumo de medicamentos en el Departamento de Farmacoepidemiología del
Ministerio de Salud Pública**

Autor (es): Elizabet Córdova Viera

Yaquelín Córdova Viera

Tutor (es): Ing. Velmour Muñoz Casals

Ing. Arieskien Mendoza Guerra

Cotutor: Ing. Guillermo Suárez Lorenzo

La Habana, Junio de 2014

“Año 56 de la Revolución.”



SOCRATES

“El conocimiento es la virtud y sólo si se sabe se puede divisar el bien”.

Sócrates

Declaración de Autoría

Declaración de Autoría

Declaramos que somos las únicas autoras de este trabajo y autorizamos a la Universidad de las Ciencias Informáticas para que haga uso del mismo en su beneficio.

Para que así conste firmamos la presente a los 20 días del mes de junio del año 2014.

Yaquelín Córdova Viera

Autor

Elizabeth Córdova Viera

Autor

Ing. Velmour Muñoz Casals

Tutor

Ing. Arieskien Mendoza Guerra

Tutor

Ing. Guillermo Suárez Lorenzo

Cotutor

Datos de Contacto

Datos de Contacto

Ing. Velmour Muñoz Casals (viazul@transnet.cu): graduada de Ingeniería en Ciencias Informáticas en la Universidad de las Ciencias Informáticas (UCI) en el año 2007. Ha sido tutora de 10 tesis de grado, oponente de 5 y ha pertenecido a varios tribunales de tesis en los que ha desempeñado en algunos con el rol de presidente y en otros de secretaria. Fue profesora por 6 años de la UCI obteniendo excelente durante los seis años en su evaluación profesoral. En estos momentos ocupa el cargo de Especialista en Ciencias Informáticas en la Agencia de Viajes Viazul de La Habana.

Ing. Arieskien Mendoza Guerra (amendoza@uci.cu): graduado de Ingeniería en Ciencias Informáticas en la Universidad de las Ciencias Informáticas (UCI) en el año 2007. Profesor de la UCI. Pertenece al Centro de Informática Médica (CESIM), específicamente al departamento de Sistemas Especializados en Salud (SES).

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), dentro de él al departamento de Sistemas Especializados en Salud (SES).

Dedicatoria

Dedicamos este trabajo a nuestros padres que siempre nos han apoyado, y ofrecido su cariño incondicional. Gracias mami y papi por su ayuda y perseverancia, por no rendirse nunca y enseñarnos a luchar por lo que queremos, cuando de objetivos e ideales se trata. Gracias por haber reído con nosotras cuando alcanzamos buenos resultados y por sentir dolor ante cada situación inesperada. A la memoria de nuestra tía Clara que todo la vida nos quiso como sus hijas y que siempre estuvo pendiente de nuestros estudios. Hoy somos 5/5 de ingenieras y eso ha sido gracias a su cuidado y dedicación.

A nuestro hermano Alejandro va dedicado especialmente este trabajo, por siempre protegernos y por su preocupación ante nuestros resultados académicos. También a Samuelito por ayudarnos tanto.

A nuestros amigos, a nuestra familia en general y a todos los que durante estos 5 años de carrera estuvieron a nuestro lado para ofrecernos su cariño y su ayuda.

Gracias por su apoyo, Elizabet y Yaquelin.

Agradecimientos

Agradecimientos

A todos los profesores que durante 5 años de carrera contribuyeron a nuestra formación. A nuestros compañeros de aula y a las muchachitas del apartamento, especialmente a Ariannis por ser nuestra mejor amiga y por no fallarnos nunca. A nuestra tutora por siempre defendernos y por dedicarnos tiempo cada vez que lo necesitamos. A Guillermo, Angélica y Polanco por su apoyo incondicional. También agradecemos a todas las personas que nos ayudaron en la realización de este trabajo.

Elizabet y Yaquelín Córdova Viera

Resumen

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: económico, social, político, científico, destacándose por su amplio impacto el sector de la salud.

Con el objetivo de llevar el control de los medicamentos consumidos en Cuba, se registra en el Módulo Consumo de Medicamentos (MCM) del Sistema para el Control Farmacológico (Synta), los medicamentos que circulan cada año en el país y su consumo. Este módulo brinda un conjunto de reportes específicos, pero estos no satisfacen las necesidades de análisis de información de los especialistas del Ministerio de Salud Pública (MINSAP), por lo que no pueden realizar una buena toma de decisiones.

El presente trabajo de diploma propone el diseño de un mercado de datos (MD) haciendo uso de la plataforma de Inteligencia de Negocios proporcionada por Pentaho; así como la metodología Hefesto para su construcción de forma sencilla, ordenada e intuitiva.

Luego de creado el MD este se integró al MCM del sistema Synta, permitiéndole a los especialistas del MINSAP efectuar consultas especializadas respecto al consumo de medicamentos, pudiendo obtener una representación gráfica de los reportes que precisan para una mejor comprensión y análisis a la hora de tomar decisiones.

Palabras claves: *consumo de medicamentos, inteligencia de negocios, mercado de datos.*

Índice de Contenidos

Índice de Contenidos

Introducción	1
Capítulo 1: Fundamentación Teórica	6
1.1 Inteligencia de negocios	6
1.1.1 Definición de inteligencia de negocios.....	6
1.1.2 Proceso de inteligencia de negocios	6
1.1.3 Beneficios de la inteligencia de negocios	7
1.2 Almacén de datos.....	8
1.2.1 Definición de almacén de datos.....	8
1.2.2 Características de los almacenes de datos	8
1.3 Mercado de datos.....	9
1.3.1 Definición de mercado de datos	9
1.3.2 Arquitectura de los mercados de datos	9
1.3.3 Ventajas de los mercados de datos.....	10
1.3.4 Diferencias entre almacenes y mercados de datos	10
1.4 Sistemas existentes	11
1.4.1 Nivel internacional	11
1.4.2 Nivel nacional.....	12
1.5 Procesamiento Transaccional en Línea.....	14
1.6 Integración de Datos	14
1.6.1 Proceso de Extracción, Transformación y Carga (ETL).....	15
1.7 Modelado Dimensional	15
1.7.1 Definición de cubo multidimensional.....	15
1.7.2 Ventajas del uso de cubos multidimensionales	16
1.7.3 Procesamiento Analítico en Línea	17
1.7.4 Tipos de modelado dimensional.....	19
1.8 Metodologías para el desarrollo de un mercado de datos	22
1.8.1 Metodología Kimball.....	22
1.8.2 Metodología Inmon.....	23
1.8.3 Metodología Hefesto	23
1.9 Herramientas propuestas para el desarrollo del mercado de datos	25

Índice de Contenidos

1.9.1	Herramientas de modelado	26
1.9.2	Sistemas Gestores de Bases de Datos	27
1.9.3	Herramienta de administración de base de datos.....	29
1.9.4	Herramientas de Inteligencia de Negocios	29
1.9.5	Servidores de Aplicaciones	33
Capítulo 2: Análisis y diseño del mercado de datos		35
2.1	Propuesta del sistema	35
2.2	Fases de la metodología Hefesto	35
2.2.1	Análisis de los requerimientos	36
2.2.2	Análisis de los OLTP	39
2.2.3	Modelo Lógico del mercado de datos	42
Capítulo 3: Integración de datos		48
3.1	Integración de datos.....	48
3.1.1	Carga Inicial	48
3.1.2	Carga Incremental.....	53
Capítulo 4: Proceso Analítico en Línea		55
4.1	Proceso de análisis de la información del mercado de datos.....	55
4.1.1	Configuración de la conexión en Schema Workbench.....	55
4.1.2	Determinar jerarquías de las tablas de dimensiones	56
4.1.3	Medidas	56
4.1.4	Consultas multidimensionales	58
4.1.5	Resultados del cliente JPivot.....	59
4.1.6	Integración del mercado de datos con el Módulo Consumo de Medicamentos.....	62
Capítulo 5: Pruebas		64
5.1	Pruebas al mercado de datos.....	64
5.1.1	Pruebas de rendimiento	64
Conclusiones		67
Recomendaciones		68
Referencias Bibliográficas.....		69
Bibliografía.....		71
Anexos.....		75

Índice de Figuras

Figura 1: Fases del proceso de inteligencia de negocios	7
Figura 2: Arquitectura dependiente	9
Figura 3: Arquitectura independiente	10
Figura 4: Cubo multidimensional	16
Figura 5: Esquema en Estrella	20
Figura 6: Esquema Copo de Nieve	20
Figura 7: Esquema Constelación de Hechos	21
Figura 8: Pasos de la metodología Hefesto.....	24
Figura 9: Modelo conceptual.....	39
Figura 10: Modelo conceptual ampliado.....	42
Figura 11: Dimensión “d_medicamento”	43
Figura 12: Diseño de las tablas de hechos.....	44
Figura 13: Uniones de las dimensiones (modelo lógico)	46
Figura 14: Transformación de la dimensión “d_medicamento”	48
Figura 15: Consulta SQL para obtener los medicamentos que serán analizados.....	49
Figura 16: Consulta SQL para obtener los campos asociados a los medicamentos.....	49
Figura 17: Transformación de la tabla “hecho_pacientes_por_medicamento_controlado”	50
Figura 18: Consulta SQL para obtener los pacientes que serán analizados.....	51
Figura 19: SQL para obtener los campos asociados a los pacientes	51
Figura 20: Tratamiento de fechas	52
Figura 21: Ejecución del trabajo (Carga inicial)	53
Figura 22: Ejecución del trabajo (Carga incremental).....	54
Figura 23: Configuración de la conexión en Schema Workbench	55
Figura 24: Jerarquía de la dimensión Tiempo	56
Figura 25: Medida cantidad de medicamentos consumidos	57
Figura 26: Medida calculable porcentaje de pacientes.....	57
Figura 27: Consulta MDX para el cubo “Registro_Medicamentos”	58
Figura 28: Consulta MDX para el cubo “Pacientes_por_Medicamento_Controlado”	59

Índice de Figuras

Figura 29: Información filtrada por el cubo “Registro_Medicamentos”	59
Figura 30: Información filtrada por el cubo “Pacientes_por_Medicamento_Controlado”	60
Figura 31: Información en gráfica de barras.....	61
Figura 32: Información en gráfica de pastel	61
Figura 33: Reporte utilizando el visor JPivot desde el MCM.....	63
Figura 34: Configuración del Plan de Pruebas de rendimiento a la base de datos	65
Figura 35: Resultados del Plan de Prueba para 25 usuarios.....	65
Figura 36: Resultado del Plan de Prueba para 50 usuarios	65
Figura 37: Resultado del Plan de Prueba para 100 usuarios	66

Índice de Tablas

Índice de Tablas

Tabla 1: Características técnicas de los sistemas analizados	13
Tabla 2: Conformación de indicadores.....	40
Tabla 3: Campos que contendrá la perspectiva “Medicamento”	41
Tabla 4: Descripción de la tabla “hecho_pacientes_por_medicamento_controlado”	45

Introducción

El volumen y variedad de la información que se encuentra en bases de datos digitales ha crecido exponencialmente en las últimas décadas, de tal forma que las empresas e instituciones en el mundo se han visto en ocasiones abarrotadas de datos históricos que no aprovechan al máximo. Esta información, bien tratada y analizada, puede reportar grandes beneficios a las organizaciones, lo que ha hecho necesario la creación de tecnologías que permitan su organización y procesamiento, posibilitando extraer conocimiento útil de la información almacenada.

La información adecuada en el lugar y momento oportuno incrementa la efectividad de cualquier negocio, pudiendo analizar sus bienes o datos acumulados y así obtener el conocimiento para apoyar la toma de decisiones. Esto se puede lograr con el uso de la inteligencia de negocios (en lo adelante IN), término que en 1989 Howard Dresner¹ definió como "... un conjunto de conceptos y métodos para mejorar el proceso de decisión utilizando un sistema de soporte basado en hechos..." (1). Mediante el uso de la IN se logra una unión entre el mundo de los datos y el de los negocios, con una solución basada en almacenes de datos (2), que según Ralph Kimball es una copia de los datos transaccionales específicamente estructurada para la consulta y el análisis (3).

Existen muchas empresas u organizaciones que implementan almacenes de datos para lograr el éxito en sus negocios de mercado (4). Ejemplo en Uruguay, para evaluar el desempeño de las instituciones de salud; en Perú, para mejorar la calidad de la atención a los pacientes. También en Cuba para el control de los recursos humanos de la salud y para análisis de la información referente a las reacciones adversas a medicamentos. Esto presupone que el uso de esta tecnología es de gran utilidad para apoyar el proceso de toma de decisiones en diferentes tipos de organizaciones.

En el sector de la salud pública cubana, específicamente en el Departamento de Farmacoepidemiología (órgano rector de los centros provinciales y municipales encargados de implementar la estrategia de Farmacovigilancia²) del Ministerio de Salud Pública (MINSAP), implementar una solución de este tipo, es

¹ Howard Dresner: autoridad bien conocida y autor en las áreas de inteligencia de negocios. Director de investigaciones de Dresner Advisory Services.

²Farmacovigilancia: conjunto de procedimientos que tienen por objeto la identificación cuantitativa del riesgo y la valoración cualitativa clínica de los efectos del uso agudo o crónico de los fármacos en el conjunto de la población o en subgrupos específicos de ella.

Introducción

conveniente debido al gran número de información que se gestiona. Además de la necesidad de contar con información consistente, confiable y oportuna para la toma de decisiones.

En el Departamento de Farmacoepidemiología del MINSAP se lleva a cabo el control de los medicamentos consumidos anualmente, y se realizan estudios de consumo para tener una noción del comportamiento de los medicamentos en la nación. Estos estudios ayudan a la planificación, teniendo en cuenta la correspondencia entre las Reacciones Adversas a Medicamentos (RAM³) ocurridas y el consumo de medicamentos (5). También se realizan cuatro procesos fundamentales: Reacciones Adversas a Medicamentos, Tarjeta Control, Recetas Médicas y Consumo de Medicamentos.

Con el objetivo de gestionar la información que se maneja en dicho departamento, es desarrollado en el Centro de Informática Médica (CESIM), perteneciente a la Universidad de las Ciencias Informáticas (UCI), el Sistema para el Control Farmacológico (Synta), específicamente en el departamento. Sistemas Especializados en Salud (SES) de este centro. Synta cuenta con cinco módulos: Reacciones Adversas a Medicamentos (MRAM), Tarjeta Control, Control de Recetas Médicas, Consumo de Medicamentos (MCM) y Administración. El sistema cuenta también con un subsistema independiente que ayuda a manejar la información común entre los componentes: el sistema Nomencladores (5).

El MCM permite registrar el consumo de medicamentos año por año ocurrido en Cuba, pudiendo obtener comparaciones de cómo se ha comportado la población con respecto a este factor. Además permite identificar qué medicamentos son los más consumidos, y cómo se debe comportar el consumo de los mismos en los años venideros.

Los especialistas del MINSAP registran en formato digital (Excel) el consumo de medicamentos en un período de tiempo determinado. Estos documentos Excel son procesados por el MCM. Una vez que la información se encuentra almacenada en la base de datos, le permite a los especialistas generar un conjunto de reportes, como se listan a continuación:

- Medicamentos consumidos en un período de tiempo.
- Consumo de un medicamento determinado.
- Visualizar gráficamente el consumo de medicamentos de un año respecto a otro.

³ RAM: reacciones perjudiciales e inesperadas, a medicamentos administrados a las dosis habituales con fines terapéuticos.

Introducción

A pesar de que el MCM gestiona un conjunto de información, no le permite a los especialistas en esta área realizar una buena toma de decisiones, pues no se puede analizar el consumo de medicamentos por:

- Provincia, sexo y la raza de los pacientes en un período de tiempo (mes, trimestre, anual). Al poder realizar un análisis detallado y preciso del consumo de medicamentos, por la unión de varias características se pueden tomar medidas de forma rápida sobre los pacientes que presentan características similares.
- Grupos etarios⁴ (de cero a quince años, de dieciséis a treintainueve, de cuarenta a cincuenta y de sesenta años en adelante). Una vez identificados los medicamentos más consumidos, se le debe dar seguimiento, según las características específicas de cada grupo de edad. De esta forma se puede detectar el abuso en el consumo de un determinado medicamento, para poder sacarlo de circulación o emitir una circular para esos grupos etarios.
- Fabricante y producción (nacional o internacional). Al poder analizar los datos históricos del consumo de medicamentos, por estas perspectivas, se puede planear la compra de medicamentos y sus cantidades, contando con una base científica.
- Otro conjunto de información necesaria para que los especialistas puedan realizar un análisis detallado de cómo se comporta el consumo de medicamentos en el país.

Debido al gran cúmulo de información recopilada en formato Excel, a los especialistas les resulta complejo el análisis de la misma para la creación de reportes que no ofrece el MCM, pues se debe consultar más de un documento Excel al mismo tiempo, lo cual trae consigo el agotamiento físico que conlleva al error humano. También trae como consecuencia el atraso en la disponibilidad de la información obtenida, careciendo los reportes de la calidad requerida, ocasionando entonces el retraso en el proceso de toma de decisiones.

A partir de lo anteriormente planteado se identifica como **problema a resolver**:

¿Cómo contribuir al análisis para la toma de decisiones referente al consumo de medicamentos en el Departamento de Farmacoepidemiología del Ministerio de Salud Pública?

El problema se encuentra enmarcado en el **objeto de estudio**: proceso de análisis para toma de decisiones en el sector de la salud, delimitándose como **campo de acción**: el análisis para la toma de

⁴ Etario: perteneciente o relativo a la edad de una persona.

Introducción

decisiones referente al consumo de medicamentos en el Departamento de Farmacoepidemiología del Ministerio de Salud Pública.

Se propone como **objetivo general**: desarrollar un mercado de datos, que contribuya al análisis para la toma de decisiones referente al consumo de medicamentos en el Departamento de Farmacoepidemiología del Ministerio de Salud Pública.

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

1. Analizar sistemas existentes que permiten la toma de decisiones en el sector de la salud, tanto a nivel internacional como nacional, obteniéndose similitudes y diferencias con la investigación.
2. Valorar las herramientas, tecnologías y metodologías, seleccionándose las necesarias para el desarrollo del mercado de datos.
3. Desarrollar el mercado de datos teniendo en cuenta los requerimientos y tecnologías identificadas, para solucionar el problema planteado.
4. Integrar el mercado de datos al Módulo Consumo de Medicamentos del Sistema para el Control Farmacológico, obteniéndose un único sistema que gestione y analice el consumo de medicamentos del país.

Para realizar la presente investigación se tuvieron en cuenta los siguientes métodos científicos:

Teóricos:

- **Histórico-Lógico**: se utilizó al realizar una valoración sobre los almacenes de datos existentes y las metodologías para su desarrollo, enmarcados en la toma de decisiones en el sector de la salud, así como al analizar el funcionamiento de los mismos.
- **Analítico-Sintético**: se utilizó durante todo el proceso investigativo, haciendo un estudio de la bibliografía relacionada con el tema y a partir del análisis realizado se seleccionó una síntesis de lo estudiado.
- **Inductivo-Deductivo**: se utilizó al analizar el mercado de datos desarrollado para el Módulo Reacciones Adversas a Medicamentos logrando obtener las diferencias y similitudes con el mercado de datos desarrollado para el Módulo Consumo de Medicamentos del Sistema para el Control Farmacológico.
- **Modelación**: se utilizó en la confección de modelos y diagramas que se generan en la metodología utilizada para el desarrollo de la propuesta de solución.

Introducción

Empíricos:

- **Entrevista:** dentro de la entrevista se utilizó la estructurada (ver Anexo 1), logrando identificar las características y requerimientos necesarios para la construcción del mercado de datos.

Una vez cumplido el objetivo general se espera el siguiente resultado:

Un mercado de datos que comprenda la información referente al consumo de medicamentos del país, con el propósito de contribuir al análisis estadístico de la información, para la toma de decisiones en el Departamento de Farmacoepidemiología del Ministerio de Salud Pública.

El presente trabajo está estructurado en cuatro capítulos. A continuación se expone una breve descripción de cada uno de ellos:

- **Capítulo 1: Fundamentación Teórica.** En este capítulo se definen los conceptos fundamentales relacionados con la investigación. Además se realiza un análisis sobre los almacenes y mercados de datos existentes a nivel nacional e internacional. Por último se definen la arquitectura, metodología y herramientas a utilizar.
- **Capítulo 2: Análisis y diseño.** En este capítulo se detallan las necesidades y perspectivas del cliente en cuanto al negocio en general. Además se llevan a cabo las tres primeras etapas que propone la metodología Hefesto para el diseño del mercado de datos, obteniéndose el modelo lógico.
- **Capítulo 3: Integración de datos.** En este capítulo se expone la última etapa de la metodología Hefesto, todo lo relacionado con la construcción de los procesos de extracción, transformación y carga del mercado de datos, empleándose la herramienta Pentaho Data Integration.
- **Capítulo 4: Proceso Analítico en Línea.** En este capítulo se procede a diseñar el cubo multidimensional haciendo uso de la herramienta Schema Workbench, dentro de esta se definen las dimensiones con sus jerarquías correspondientes. También se elaboran las medidas para dar respuesta a las necesidades del cliente y se visualizan los resultados mediante tablas de datos y gráficas, utilizando el visor web JPivot desde el MCM.
- **Capítulo 5: Pruebas.** En este capítulo se realizan varias pruebas de rendimiento al mercado de datos, con el objetivo evaluar su desempeño. Para esto se utiliza la herramienta Jkarta-jmeter y las cantidades de 25, 50 y 100 usuarios accediendo simultáneamente.

Capítulo 1: Fundamentación Teórica

Se definen los conceptos fundamentales relacionados con la investigación. Además se realiza un análisis sobre los almacenes y mercados de datos existentes a nivel nacional e internacional. Por último se definen la arquitectura, metodología y herramientas a utilizar.

1.1 Inteligencia de negocios

En la era de la información, las organizaciones tienen a su disposición grandes cantidades de datos, recolectadas en sistemas transaccionales. Dichos sistemas, son esenciales para las operaciones del negocio.

Disponer de datos no es lo mismo que disponer de información. Los datos se convierten en información cuando se pueden utilizar para responder preguntas del negocio, de tal manera que se pueda comprender mejor el funcionamiento del mismo. La inteligencia de negocios (conocida en inglés como Business Intelligence) permite responder a tales cuestiones, por lo que los decisores pueden responder rápidamente ante los cambios en el entorno del negocio (6).

1.1.1 Definición de inteligencia de negocios

Para extraer una cierta inteligencia o conocimiento de las organizaciones, nace en los años 90` la Inteligencia de Negocios (IN en lo adelante) como apoyo al proceso de toma de decisiones.

Se puede definir la IN, como un concepto que integra por un lado el almacenamiento y por otro el procesamiento de grandes cantidades de datos, con el principal objetivo de transformarlos en conocimiento y en decisiones en tiempo real, a través de un sencillo análisis y exploración. Inteligencia de negocios es el proceso de convertir datos en conocimiento y el conocimiento en acción, para la toma de decisiones (7).

1.1.2 Proceso de inteligencia de negocios

El proceso de IN está dividido en cinco fases, las cuales son explicadas teniendo como referencia el siguiente gráfico, que sintetiza todo el proceso:

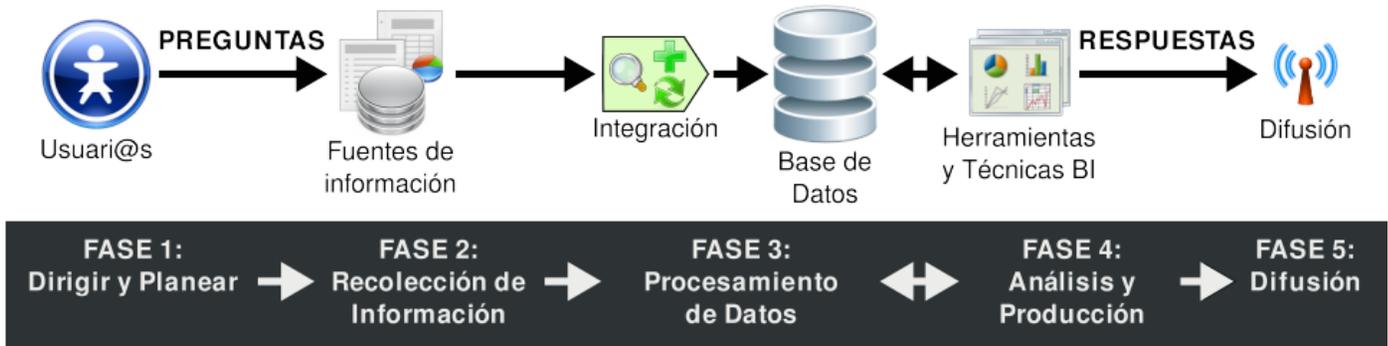


Figura 1: Fases del proceso de inteligencia de negocios

- **Fase 1: Dirigir y Planear.** En esta fase es donde se recolectan los requerimientos de información específicos de los diferentes usuarios para así entender sus necesidades, luego en conjunto con ellos se generan las preguntas que les ayudarán a alcanzar sus objetivos.
- **Fase 2: Recolección de Información.** En esta fase es donde se realiza el proceso de extraer de las diferentes fuentes de información de la empresa, los datos que serán necesarios para encontrar las respuestas a las preguntas planteadas en el paso anterior.
- **Fase 3: Procesamiento de Datos.** En esta fase es donde se integran y cargan los datos en un formato utilizable para el análisis. Esto puede realizarse mediante la creación de una nueva base de datos, agregando datos a una base de datos ya existente o bien consolidando la información.
- **Fase 4: Análisis y Producción.** En esta fase se procederá a trabajar sobre los datos extraídos e integrados, utilizando herramientas propias de la IN. Como resultado final de esta fase se obtendrán las respuestas a las preguntas de la fase 1 mediante la creación de reportes, indicadores de rendimiento, gráficos estadísticos, etc.
- **Fase 5: Difusión.** En esta fase y última, se le entregarán a los usuarios que lo requieran las herramientas necesarias, que les permitirán revisar los datos de manera rápida y sencilla (8).

1.1.3 Beneficios de la inteligencia de negocios

- Integración de datos entre los diferentes sistemas de información existentes en una organización.
- Proporciona herramientas de análisis para establecer comparaciones y tomar decisiones.

- Permite a los usuarios no depender de reportes o informes programados, ya que los mismos serán generados de manera dinámica.
- Los usuarios podrán consultar y analizar los datos de manera sencilla e intuitiva.

1.2 Almacén de datos

Debido a que para llevar a cabo la IN, es necesario gestionar datos guardados en diversos formatos y fuentes, para luego depurarlos e integrarlos, además de almacenarlos en un sólo destino o base de datos que permita su posterior análisis y exploración, es de vital importancia contar con un almacén de datos, en lo adelante AD.

1.2.1 Definición de almacén de datos

Un AD puede verse como una bóveda donde están almacenados todos los datos necesarios para realizar las funciones de gestión de las empresas, de manera que puedan utilizarse fácilmente según se necesiten, permitiendo satisfacer las necesidades de información de los analistas y usuarios que toman decisiones (8).

Según John Edwards⁵: *“un AD toma información de múltiples sistemas y bases de datos, y la almacena para dar a los usuarios acceso rápido, fácil y flexible a los aspectos claves de la organización”* (7).

1.2.2 Características de los almacenes de datos

- **Orientado a temas:** los AD se construyen orientados a la información más relevante de una organización. Los temas a analizar por cada organización son particulares de cada una de ellas.
- **Integrado:** la integración implica que todos los datos de diversas fuentes deben ser consolidados en una instancia antes de ser agregados al AD. A este proceso se le conoce como Extracción, Transformación y Carga (8).
- **Variante en el tiempo:** los datos almacenados en la base de datos, son relativos a un período de tiempo y estos deben ser integrados periódicamente (9), con el fin de poder identificar los distintos valores que han ido tomando los datos a lo largo de su historia.
- **No volátil:** para realizar análisis de información y tomar decisiones se necesita una base de datos estable, por lo que una vez introducidos los datos en un AD pasan a ser de sólo lectura, es decir, los

⁵John Edwards: autor de numerosos artículos y noticias relacionados con la inteligencia de negocios y múltiples áreas de la tecnología. Sus trabajos aparecen publicados en sitios web patrocinados por Oracle, Cisco, IBM, Dell, entre otros.

datos no pueden ser ni modificados ni eliminados. Las únicas dos acciones que se permiten realizar son la carga de datos y el acceso a la información (7).

1.3 Mercado de datos

Los AD juegan un papel fundamental como herramientas de apoyo al proceso de toma de decisiones. Sin embargo, para que puedan tomar ventajas de los recursos de información de una manera satisfactoria, es conveniente que las empresas cuenten con herramientas que le permitan almacenar y analizar la información de un determinado departamento dentro de la organización. Entre estas herramientas se destacan los mercados de datos.

1.3.1 Definición de mercado de datos

Un mercado de datos (MD en lo adelante) es una base de datos departamental, especializada en el almacenamiento de los datos de un área de 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 un departamento. Un MD puede ser alimentado desde los datos de un AD o integrar por sí mismo un compendio de distintas fuentes de datos (10).

1.3.2 Arquitectura de los mercados de datos

De acuerdo a las operaciones que se requieran desarrollar, los MD pueden adoptar las siguientes arquitecturas:

- **Dependientes** (o Top-Down, en inglés): primero se define el AD y luego se desarrollan, construyen y cargan los MD a partir del mismo. En la siguiente figura se detalla esta arquitectura:

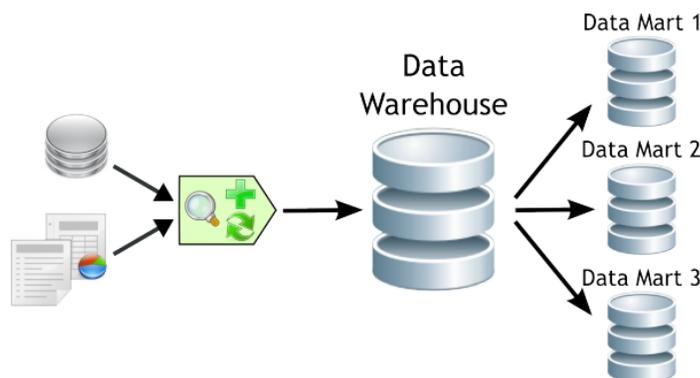


Figura 2: Arquitectura dependiente

- **Independientes** (o Button Up, en inglés): en esta arquitectura, se definen previamente los MD y luego se integran en un AD centralizado (8). La siguiente figura presenta esta implementación:

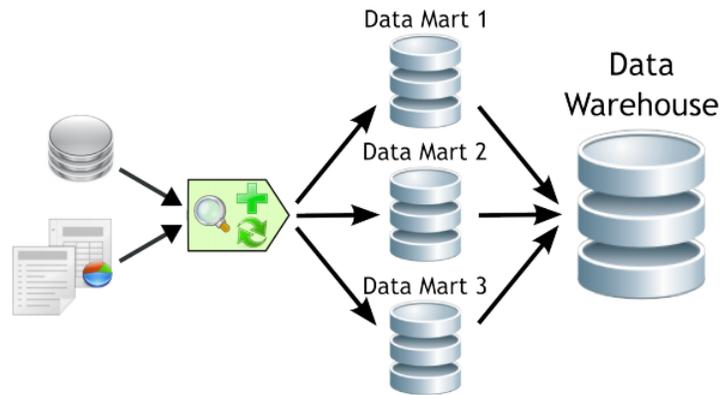


Figura 3: Arquitectura independiente

1.3.3 Ventajas de los mercados de datos

- Son simples de implementar.
- Conllevan poco tiempo de construcción y puesta en marcha.
- Integran diferentes fuentes de datos en información orientada a la toma de decisiones.
- Ofrecen rapidez en las consultas.
- Permiten ejecutar consultas en Lenguaje de Consulta Estructurado (Structure Query Language, SQL⁶, por sus siglas en inglés) y/o en Lenguaje de Expresiones Multidimensionales (Multidimensional Expressions, MDX⁷, por sus siglas en inglés) sencillas.

1.3.4 Diferencias entre almacenes y mercados de datos

La principal diferencia entre un AD y un MD es su alcance. El ámbito de un AD es la organización en su conjunto. En cambio, un MD está pensado para cubrir las necesidades de un grupo de trabajo dentro de la organización. Este último es el almacén natural para los datos departamentales (10). Es menos costoso de diseñar e implementar que un AD. Por tanto, no se puede pensar en un MD en términos de un AD más pequeño, porque no es su tamaño lo que lo define, sino su objetivo en la organización.

⁶ SQL: lenguaje formal declarativo para manipular la información de una base de datos.

⁷ MDX: lenguaje de consulta para bases de datos multidimensionales sobre cubos OLAP.

La información que se desea gestionar está orientada al Departamento de Farmacoepidemiología del MINSAP, los medicamentos y su consumo, por lo que se hace necesario la implementación de un MD independiente y no un AD.

1.4 Sistemas existentes

A nivel mundial son muchas las empresas que han optado por los AD y MD para el almacenamiento y tratamiento de su información. Ejemplos, el Sistema Nacional de Información Médico-Asistencial de Uruguay, para evaluar el desempeño de las instituciones de salud; también en Cuba para analizar la información histórica de las reacciones adversas a medicamentos. Convirtiéndose estos sistemas en una vía para ayudar a los ejecutivos en la toma decisiones apropiadas.

1.4.1 Nivel internacional

Sistema Nacional de Información Médico-Asistencial de Uruguay (SINADI): software creado en 1991, que recoge y procesa información médico-asistencial de todas las instituciones prestadoras de servicios de salud Uruguay. Este sistema permite generar automáticamente informes mensuales correspondientes al desempeño de cada una de las instituciones, haciendo uso de un AD. Desde el punto de vista operativo, las instituciones hospitalarias envían mensualmente en formato electrónico (planillas Excel) los datos requeridos por el Ministerio de Salud Pública (MSP), estos son cargados en un AD (desarrollado con herramientas de la empresa IBM Cognos, Power Play, etc.) y posteriormente se generan un conjunto de reportes en Word que son analizados por la Dirección General del MSP. Entre los reportes se encuentran: total de consultas de urgencias, porcentaje de cesáreas, estadía promedio en el hospital, etc.

La información procesada de cada institución es enviada a la misma, comparada con los indicadores del resto de las instituciones del país. Al no obtenerse indicadores vinculados a la calidad de los servicios y ante la necesidad de ampliar y mejorar la calidad de los datos recibidos por el SINADI, se le definieron cambios a este, que están en vías de implementación. Los nuevos reportes están referidos fundamentalmente al análisis de indicadores como: índice de reingresos, tiempo de demora de una cirugía coordinada, entre otros. Este sistema no es de código abierto y el costo de la licencia establecido por IBM es alto (11).

Almacén de datos de Soporte de Decisiones para un Hospital del Sistema de Salud Público de Perú (ADDHP): decide, basándose en datos históricos, la calidad del servicio otorgado en el hospital. Permite a

los directivos del hospital Santa Rosa tomar decisiones para prevenir eventos adversos, responder a situaciones imprevistas y a cambios en la demanda de los servicios. Permite además mejorar la calidad de la atención a los pacientes teniendo en cuenta comparaciones con cifras anteriores, siendo posible medir los cambios en los indicadores de calidad y eficiencia de gestión del hospital.

Para la construcción de este almacén de datos se utilizó la plataforma de inteligencia de negocios Pentaho y como gestor de base de datos PostgreSQL. Fue desarrollado en el año 2008 en la Universidad Católica de Lima, Perú (12).

1.4.2 Nivel nacional

Almacén de datos para el control de los Recursos Humanos de la Salud (ADRHS): permite a los especialistas del Observatorio Nacional de Recursos Humanos de la Salud⁸ (RRHH) contar con un AD que interrelacione los temas de colaboración médica y postgrado, para obtener una visión de la información histórica de los colaboradores y su formación docente. Este AD permite conocer la cantidad, porcentaje y total de recursos humanos del sector de salud, clasificándolos por diversos criterios (ejemplos: cantidad de RRHH que se encuentran colaborando en un país determinado y cuentan con la categoría de máster o doctores en ciencia, total de RRHH que se encuentran en espera de misión y son especialistas en primer grado, porcentaje que representa los RRHH en misión con respecto a los que se encuentran en espera de misión y cuentan con alguna categoría científica).

Posibilita además que los especialistas puedan ver los resultados gráficamente, para una mejor comprensión y estudio a la hora de tomar decisiones sobre la asignación de recursos humanos donde realmente sea necesario, según su formación postgraduada y experiencia en la actividad de colaboración. Para el proceso de integración de datos se utilizó Pentaho Data Integration, para la estructuración de la información el esquema en estrella y como metodología Hefesto (13). El almacén de datos fue desarrollado por la Facultad 7 de la Universidad de las Ciencias Informáticas en el año 2009.

Mercado de datos para la toma de decisiones referente a las Reacciones Adversas a Medicamentos en el Ministerio de Salud Pública desde el sistema Synta (MDRAM): este MD permite a los farmacoepidemiólogos del MINSAP contar con una herramienta para realizar el análisis de la información

⁸Observatorio Nacional de Recursos Humanos de la Salud: departamento del MINSAP encargado de supervisar y controlar los recursos humanos del sector de la salud en Cuba.

Capítulo 1: Fundamentación Teórica

referente a las RAM ocurridas en Cuba. Además ofrece reportes que facilitan una mejor comprensión de los datos (mediante gráficos y tablas) a la hora de tomar decisiones.

Se utilizó como metodología Hefesto, como herramienta de integración de datos Pentaho Data Integration para la extracción, transformación y carga de la información, y el esquema en estrella para el modelado dimensional (14). El mercado de datos fue desarrollado por la Facultad 7 de la Universidad de las Ciencias Informáticas en el año 2013.

Una vez realizado el análisis de un grupo de sistemas, tanto nacionales como internacionales, examinando sus potencialidades y características técnicas (ver Tabla 1), se puede afirmar que ninguno de los sistemas analizados satisface en su totalidad las necesidades de la investigación. Estos posibilitan el análisis de información y la toma de decisiones en el sector de la salud, pero no están orientados a analizar la información referente al consumo de medicamentos. Además algunos tienen licencia propietaria, lo que conlleva al pago de altos precios por las mismas. Por lo antes planteado es necesario implementar un sistema que satisfaga las necesidades de la investigación, permitiendo a los especialistas del MINSAP realizar el análisis estadístico de la información y tomar decisiones apropiadas sobre el consumo de medicamentos del país.

No obstante el análisis de estos sistemas posibilitó un mayor acercamiento a los AD y MD obteniendo nuevos conocimientos sobre su diseño e implementación. Permitió además conocer técnicas y herramientas a utilizar en el desarrollo del mercado de datos y la construcción del diseño de la propuesta de solución.

Sistemas		Herramientas de IN		Modelado dimensional		Metodología	
		Propietaria	Libre	Estrella	Constelación	Hefesto	Propia
Internacionales	SINADI	X		X			X
	ADDHP		X		X		X
Nacionales	ADRHS		X	X		X	
	MDRAM		X	X		X	
Sistema a desarrollar	MD: consumo medicamentos		X		X	X	

Tabla 1: Características técnicas de los sistemas analizados

1.5 Procesamiento Transaccional en Línea

El Procesamiento Transaccional en Línea (OLTP, por sus siglas en inglés) representa toda aquella información transaccional que genera la empresa en su accionar diario. Entre los OLTP más habituales que pueden existir en una organización se encuentran:

- Archivos de textos
- Hojas de cálculos
- Informes semanales, mensuales, anuales
- Bases de datos transaccionales (8)

Estas últimas están orientadas al procesamiento de transacciones, involucrando operaciones de inserción, modificación y borrado de datos. Además el acceso a los datos está optimizado para tareas frecuentes de lectura y escritura y tomar decisiones diarias en una organización (15). Sin embargo, tienen la desventaja de no ser eficientes en el manejo de grandes volúmenes de información, ni en consultas de mucha complejidad.

En cambio, los AD y MD están diseñados para llevar a cabo procesos de consultas y análisis (basados en el historial de una empresa) para luego tomar decisiones estratégicas. En la mayoría de los casos, la información histórica proviene de múltiples fuentes y generalmente se encuentra en formatos distintos, por lo que es necesario llevar a cabo el proceso de integración de datos, lo que permite unificar y sincronizar estos.

1.6 Integración de Datos

La Integración de Datos agrupa una serie de técnicas o subprocesos que se encargan de llevar a cabo todas las tareas relacionadas con la extracción, manipulación, control, depuración y carga de datos. Es decir, todas las tareas que se realizarán desde que se toman los datos de las diferentes fuentes hasta que se cargan en el MD. Estos subprocesos son la Extracción, Transformación y Carga (ETL, por sus siglas en inglés) (7).

1.6.1 Proceso de Extracción, Transformación y Carga (ETL)

El proceso ETL permite a las organizaciones explorar las diversas fuentes de datos que se tengan a disposición y extraer la información que se considere relevante, luego transformarla para resolver posibles problemas de fiabilidad y finalmente, se procede a su carga en el MD (7).

Las funciones específicas de la ETL son:

- **Extracción:** es el proceso de lectura de datos de los sistemas operacionales o cualquier otra fuente de información, como *por ejemplo: Excel*.
- **Transformación:** es el proceso de convertir, filtrar y validar los datos extraídos de las fuentes de información, dejándolos listos para que puedan ser cargados en el MD.
- **Carga:** consiste en almacenar los datos transformados y validados en la base de datos final, *por ejemplo el MD (8)*.

Un buen proceso ETL permite una integración consistente de los datos. Al culminar este proceso, los datos serán: precisos, completos, creíbles, interpretables y accesibles. Este proceso no es una simple migración, pues asegura la calidad de los datos, los prepara, limpia, transforma y organiza, brindando a quienes utilicen estos datos un conocimiento que facilitará la toma de decisiones.

1.7 Modelado Dimensional

Modelado dimensional, es el nombre que recibe una técnica utilizada especialmente en los MD. La idea fundamental es que los datos del negocio pueden ser representados como cubos de datos.

1.7.1 Definición de cubo multidimensional

Un cubo multidimensional, convierte los datos planos que se encuentran en filas y columnas en una matriz de n dimensiones. Este modelo permite el acceso flexible a los datos, para explorar y analizar sus relaciones y consiguientes resultados. Además la información va a almacenarse a través de tablas de dimensiones y tablas de hechos, presentándola de una manera estándar, sencilla y sobre todo intuitiva para los usuarios (7). A continuación se muestra un ejemplo de cubo multidimensional:

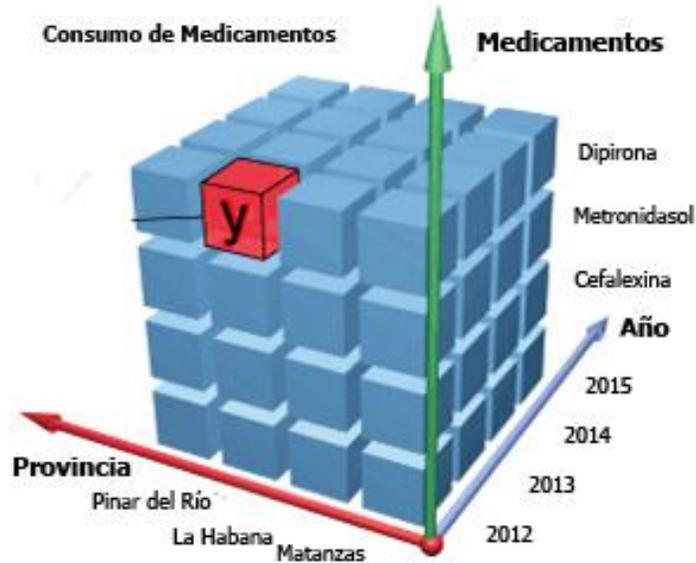


Figura 4: Cubo multidimensional

Los objetos más importantes que se pueden incluir en un cubo dimensional son los siguientes:

- **Hecho:** evento específico que constituye la unidad fundamental de análisis de datos, para la toma de decisiones. *Ejemplo: Consumo de Medicamentos.* Cada hecho representa una tabla en el MD.
- **Dimensiones:** es una entidad del negocio respecto a la cual se pueden calcular medidas. *Ejemplos: medicamentos, provincias, tiempo.* Las tablas que almacenan las dimensiones se denominan tablas de dimensiones, estas definen como están los datos organizados lógicamente.
- **Medidas:** valores cuantitativos que almacenan las métricas del negocio. Están representados por columnas numéricas en las tablas de hechos. *Ejemplo: cantidad.*

En los cubos cada celda contiene un valor y las aristas definen dimensiones naturales de análisis.

1.7.2 Ventajas del uso de cubos multidimensionales

- Se presentan los datos desde diferentes perspectivas de análisis.
- Posibilitan la adición de dimensiones y hechos, sin que esto implique volver a cargar los datos ya almacenados.
- Pueden responder con rapidez a consultas de los usuarios.

1.7.3 Procesamiento Analítico en Línea

Los cubos son elementos claves en el Procesamiento Analítico en Línea (On Line Analytical Processing, OLAP por sus siglas en inglés), una tecnología que provee el rápido acceso a los datos en un MD.

Los sistemas OLAP son una tecnología de software⁹ para análisis en línea, administración y ejecución de consultas, que permiten inferir información del comportamiento del negocio. Este análisis suele implicar, generalmente la lectura de grandes cantidades de datos para llegar a extraer información útil, proporcionando respuestas rápidas a consultas analíticas utilizadas generalmente en sistemas de apoyo a la toma de decisiones (16). Estos sistemas son típicos de los MD, y pueden organizar y presentar información en el formato (tablas de datos y/o gráficas) que mejor se ajuste a las necesidades de los usuarios, de forma que estos puedan entender mejor la información que se desea mostrar.

En síntesis, los sistemas OLAP se caracterizan por brindar un análisis multidimensional de los datos, soportando las consultas de los usuarios y ofreciendo una navegación que permite seleccionar la información a obtener.

Categorías de los sistemas OLAP

- OLAP Relacional (Relational On Line Analytic Processing, ROLAP, por sus siglas en inglés)
- OLAP Multidimensional (Multidimensional On Line Analytic Processing, MOLAP, por sus siglas en inglés)
- OLAP Híbrido (Hybrid On Line Analytic Processing, HOLAP, por sus siglas en inglés)

OLAP Relacional

Los sistemas ROLAP acceden directamente a los datos almacenados en un MD, proporcionando los análisis para el OLAP. Los datos son almacenados en filas y columnas de forma relacional, y se les presentan a los usuarios en forma de dimensiones del negocio (8). ROLAP guarda la información en bases de datos relacionales, aprovechando así la tecnología relacional, permitiendo usar la integridad y seguridad de los sistemas gestores de bases de datos.

⁹ Software: es el conjunto de programas de cómputo, procedimientos, reglas, documentación y datos asociados que forman parte de las operaciones de un sistema de cómputo.

Ventajas de los sistemas ROLAP

- Manejan grandes cantidades de datos.
- Los cubos multidimensionales se generan dinámicamente al instante de realizar las consultas, haciendo de esta manera el manejo de cubos transparente a los usuarios.
- No tienen limitaciones en cuanto al uso de dimensiones.

Desventajas de los sistemas ROLAP

- El rendimiento puede ser lento, pues los datos de los cubos se deben calcular cada vez que se ejecuta una consulta sobre ellos.

OLAP Multidimensional

El objetivo de los sistemas MOLAP es almacenar físicamente los datos en estructuras multidimensionales de manera que la representación externa y la interna de la base de datos coincidan (7).

Ventajas de los sistemas MOLAP

- Debido a que poseen datos precalculados, ofrecen respuestas rápidas a las consultas de los usuarios.

Desventajas de los sistemas MOLAP

- Son poco flexibles, pues al modificar un cubo es necesario recalcularlo totalmente, para que se reflejen las modificaciones llevadas a cabo.
- Se precisa más espacio físico para almacenar datos precalculados.
- Limitado en cuanto a la cantidad de datos que se puede manejar, siendo poco adecuado para muchas dimensiones.

OLAP Híbrido

Los sistemas HOLAP, constituye un sistema híbrido entre MOLAP y ROLAP, que combina estas dos implementaciones con la finalidad de obtener lo mejor de ambos sistemas (16). Los datos precalculados se almacenan en estructuras multidimensionales y los de menor nivel de detalle en estructuras relacionales.

Ventajas de los sistemas HOLAP

- Reduce los costos de hardware ya que se necesita menos espacio en disco que en las bases de datos relacionales.
- Las respuestas de las consultas sobre las bases de datos multidimensionales son más rápidas que sobre las relacionales.

Desventajas de los sistemas HOLAP

- Hay que realizar un buen análisis para identificar los diferentes tipos de datos que irán en cada estructura dimensional.
- No es recomendable para personas con poca experiencia en el tema de los AD y MD.

Para el procesamiento analítico en línea se utiliza el sistema ROLAP, puesto que el MD está soportado por PostgreSQL, el cual es un gestor de base de datos relacional. Además los datos se guardan en filas y columnas de forma relacional, y se les muestran a los usuarios en forma de dimensiones del negocio. Para lograr esto se utilizan diferentes esquemas o variantes de modelado dimensional: en estrella, copo de nieve o constelación, los cuales permiten, junto con otras herramientas, gestionar y administrar los datos de manera multidimensional.

1.7.4 Tipos de modelado dimensional

Los MD implican tres tipos posibles de modelado, los cuales permiten realizar consultas de soporte de decisiones:

- Esquema en Estrella
- Esquema Copo de Nieve
- Esquema Constelación de Hechos

Esquema en Estrella

Este esquema consta de una tabla de hechos central y de varias tablas de dimensiones relacionadas a esta, a través de sus respectivas claves. 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. Las tablas de dimensiones solo se relacionan con la tabla de hechos (17). En la siguiente figura se puede apreciar un esquema en estrella estándar:

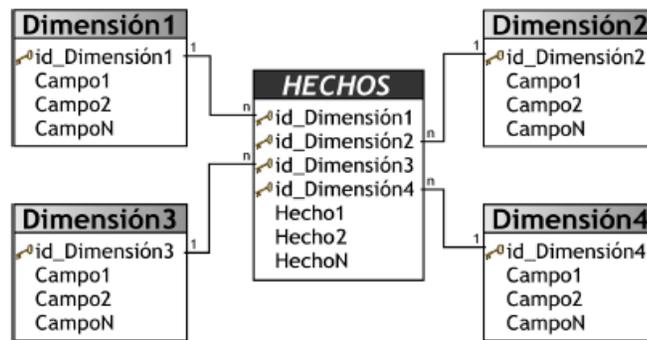


Figura 5: Esquema en Estrella

Características del esquema en estrella

- Su diseño es fácilmente modificable.
- Simplifica el análisis.
- Facilita la interacción con herramientas de consulta y análisis.

El esquema en estrella es el más simple de interpretar y optimiza los tiempos de respuestas ante las consultas de los usuarios, sin embargo es el menos robusto para la carga y es el más lento de construir.

Esquema Copo de Nieve

El esquema copo de nieve representa una extensión del modelo en estrella cuando las tablas se organizan en jerarquías de dimensiones. En este esquema (ver Figura 6) existe una tabla de hechos central que está relacionada con una o más tablas de dimensiones, quienes a su vez pueden estar relacionadas o no con una o más tablas de dimensiones (16).

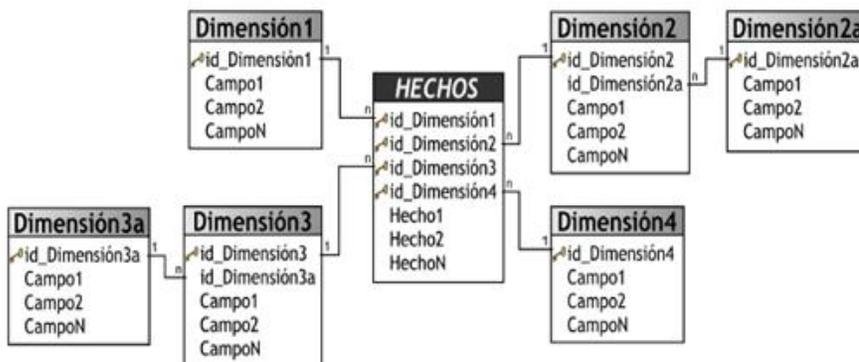


Figura 6: Esquema Copo de Nieve

Características del esquema copo de nieve

- Posee mayor complejidad en su estructura, que el esquema en estrella.
- Hace una mejor utilización del espacio.
- Es muy útil en tablas de dimensiones con muchos registros.

Este esquema puede llegar a ser inmanejable de existir muchas tablas de dimensiones, cada una de ellas con varias jerarquías. Además su desempeño puede verse reducido si existen muchas uniones y relaciones entre tablas.

Esquema Constelación de Hechos

Este modelo está compuesto por una serie de esquemas en estrella y tal como se puede apreciar en la Figura 7, está formado por una tabla de hechos principal y por una o más tablas de hechos auxiliares, las cuales pueden ser sumalizaciones¹⁰ de la principal. Dichas tablas yacen en el centro del modelo y están relacionadas con sus respectivas tablas de dimensiones.

No es necesario que las diferentes tablas de hechos compartan las mismas tablas de dimensiones, ya que, las tablas de hechos auxiliares pueden vincularse con solo algunas tablas de dimensiones asignadas a la tabla de hechos principal, y también pueden hacerlo con nuevas tablas de dimensiones (16).

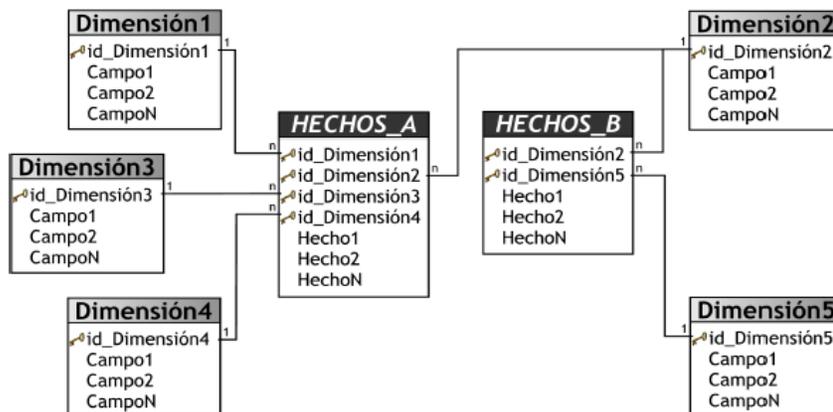


Figura 7: Esquema Constelación de Hechos

¹⁰Sumarización: 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.

Características del esquema constelación de hechos

- Permite tener más de una tabla de hechos, por lo cual se pueden analizar más aspectos claves del negocio con un mínimo esfuerzo adicional de diseño.
- Una misma tabla de dimensión puede relacionarse con varias tablas de hechos, contribuyendo a la reutilización de las tablas de dimensiones.

Se propone utilizar el esquema constelación de hechos pues al estar compuesto de varias estrellas es fácil de interpretar y modificar. Además es el modelo que brinda mayor flexibilidad a la hora de diseñar el modelo dimensional, ofreciendo un buen tiempo de respuesta para las consultas de los usuarios.

1.8 Metodologías para el desarrollo de un mercado de datos

Una metodología de desarrollo de software es un conjunto de procedimientos, técnicas y herramientas que ayudan a los desarrolladores a realizar un nuevo producto de software. Una de las peculiaridades de los AD y MD, es que modelos y metodologías tradicionales no son en ocasiones apropiadas para su diseño, por lo que han surgido metodologías propias para este tipo de sistemas.

Existen varias metodologías que han sido reconocidas en el área de los MD, entre estas se encuentran:

- Kimball
- Inmon
- Hefesto

1.8.1 Metodología Kimball

La metodología Kimball, llamada así en homenaje a su autor Ralph Kimball¹¹, está basada en una arquitectura independiente, donde los datos se almacenan en varios MD satisfaciendo las necesidades departamentales, para luego construir el AD de la organización (10). El ciclo de vida de Kimball ilustra el flujo general de implementación de un MD, identificando una secuencia de tareas ordenadas y actividades principales que deben suceder concurrentemente; dejando claro que se debe hacer, pero no como lograrlo, lo que provoca demora en los resultados.

¹¹ Ralph Kimball: autor ampliamente publicado sobre el tema de almacén de datos. Considerado el principal promotor del enfoque dimensional para el diseño de almacenes de datos.

1.8.2 Metodología Inmon

Esta metodología la definió Bill Inmon¹² en el año 1992, en su libro “Building the Data Warehouse”. El enfoque de Inmon está basado en una arquitectura dependiente, donde los datos son cargados desde los sistemas operacionales hacia el AD empresarial y entonces subdivididos en los MD (15). Esta metodología puede tener una implementación tardada, por lo que no es muy recomendable para proyectos sencillos pues va de lo general, el AD, a lo más específico, el MD.

1.8.3 Metodología Hefesto

Hefesto es una metodología, cuya propuesta está fundamentada en una amplia investigación, comparación de metodologías existentes y experiencias adquiridas en procesos de confección de MD. 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 (7). Puede ser embebida en cualquier ciclo de vida de desarrollo de software que no requiera de fases extensas de requerimientos y análisis, esta metodología se encuentra dividida en 4 fases, las cuales pueden resumirse en el siguiente gráfico.

¹² Bill H. Inmon: reconocido como el padre de los almacenes de datos.



Figura 8: Pasos de la metodología Hefesto

- **Fase 1: Análisis de los requerimientos.** 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 MD.
- **Fase 2: Análisis de los OLTP.** Se analizan los OLTP, conformando los indicadores y estableciendo las correspondencias con los datos fuentes, de esta manera se obtienen los diferentes niveles de granularidad¹³ y se construye el modelo conceptual ampliado.

¹³Granularidad: describe el nivel de detalle con que se desea analizar la información presente en una base de datos.

- **Fase 3: Modelo lógico del mercado de datos.** Se construye el modelo lógico del MD, 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.
- **Fase 4: Integración de datos.** Por último, utilizando el proceso ETL, se definen políticas y estrategias para la carga inicial del MD y su respectiva actualización (8).

Características de la metodología Hefesto

- 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.
- Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.
- Es independiente de las herramientas que se utilicen para su implementación.
- Cuando se culmina con una fase, los resultados obtenidos se convierten en el punto de partida para llevar a cabo el paso siguiente.

Después del análisis de las metodologías antes expuestas, se propone utilizar la metodología Hefesto para guiar el desarrollo del MD. Esta metodología permite construirlo de forma sencilla, ordenada e intuitiva. Además es una metodología bien fundamentada y explícita, que facilita el proceso de implementación del MD, pues permite guiarse por una serie de pasos lógicos relacionados sólidamente durante todas las etapas de desarrollo.

1.9 Herramientas propuestas para el desarrollo del mercado de datos

En este epígrafe se realiza una investigación acerca de las herramientas necesarias para llevar a cabo el proceso de construcción de un mercado de datos, las cuales gracias a este análisis arrojan un conjunto de características que se mencionarán posteriormente, con las cuales se realizará este proceso desde la adquisición de los datos hasta la visualización de los resultados.

Entre las diferentes herramientas se encuentran:

- Herramientas de modelado: Visual Paradigm, Enterprise Architect.
- Sistemas Gestores de Base de Datos: MySQL, PostgreSQL.
- Herramienta de Administración de Base de Datos: PgAdmin III.
- Herramientas de Inteligencia de Negocios: Oracle Business Intelligence Suite, JasperSoft Business Intelligence Suite, Pentaho Open Source Business Intelligence. Esta última se compone de Pentaho

Data Integration como herramienta para el proceso ETL, Schema Workbench y Pentaho Analyses Services para el trabajo con los cubos.

- Servidores de aplicaciones: JBoss, Apache Tomcat.

1.9.1 Herramientas de modelado

Las herramientas de Ingeniería de Software Asistidas por Computadoras (CASE, por sus siglas en ingles), son aplicaciones computacionales que soportan y ayudan al proceso de análisis y desarrollo de software. Estas herramientas facilitan el progreso de los procesos y permiten lograr mayor calidad en el producto final (18).

Existen diversas herramientas que potencian el modelado, entre las más conocidas se puede mencionar:

- Visual Paradigm
- Enterprise Architect

Visual Paradigm

De las herramientas CASE orientadas al Lenguaje Unificado de Modelado (UML, por sus siglas en ingles), Visual Paradigm soporta el ciclo de vida completo del desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue. Permite dibujar todos los tipos de diagramas de clases, código inverso, generar código desde diagramas y generar documentación. Es empleado para aplicaciones grandes y complejas que presentan un enfoque orientado a objetos. Presenta licencia gratuita y comercial (código cerrado y propietario). Consume gran cantidad de recursos del ordenador.

Enterprise Architect 7.1

Enterprise Architect es una herramienta CASE profesional y efectiva, la cual ofrece una trazabilidad¹⁴ completa desde el análisis de requerimientos y los artefactos¹⁵ de diseño, hasta la implementación y el despliegue. Provee una sobrecarga de instalación baja, un rendimiento brillante y una interfaz intuitiva (18). Permite generar documentación flexible y de alta calidad, así como código para el software que se

¹⁴ Trazabilidad: proceso prefijado que se lleva a cabo para determinar los diversos pasos que recorre un producto, desde su nacimiento hasta su ubicación final. Registro de huellas que deja un producto mientras transita por la cadena antes de llegar al final.

¹⁵Artefactos: son productos tangibles del proyecto. Las cosas que el proyecto produce o usa para componer el producto final (modelos, documentos, código, ejecutables).

está construyendo e ingeniería inversa. Es una herramienta de construcción y modelado de software orientada a objetos, siendo flexible para las plataformas de Windows y Linux, ofreciendo bajos costos de licencias.

Se elige para realizar el diseño de los diferentes artefactos del MD la herramienta Enterprise Architect, pues provee beneficios significativos para ayudar a construir modelos de sistemas de software rigurosos y donde es posible mantener la trazabilidad de manera consistente, soportando este proceso en un ambiente fácil de usar, rápido y flexible. Se cuenta además con una amplia documentación sobre ella pues es la herramienta definida en las políticas del departamento Sistemas Especializados en Salud.

1.9.2 Sistemas Gestores de Bases de Datos

Los Sistemas Gestores de Base de Datos (SGBD) son un tipo de software específico que permiten a los usuarios procesar, describir, administrar y recuperar los datos almacenados en una base de datos. Además proporcionan un conjunto de programas, procedimientos y lenguajes, que permiten a los distintos usuarios realizar sus tareas con los datos, garantizando la seguridad e integridad de los mismos (19).

Entre los SGBD más conocidos se encuentran:

- MySQL
- PostgreSQL

MySQL

Es un SGBD relacional, bajo la Licencia Pública General (GPL, por sus en inglés) de la GNU. MySQL fue creada por la empresa sueca MySQL AB. Está enfocado a aplicaciones web de lectura mayormente, utilizando consultas sencillas.

Características de MySQL

- Soporta gran cantidad de tipos de datos para las columnas.
- Gran portabilidad entre sistemas.
- Soporta hasta 32 índices por tabla.
- Gestión de usuarios y contraseñas, manteniendo un muy buen nivel de seguridad en los datos.

Ventajas de utilizar MySQL

- Lo mejor de MySQL es su velocidad a la hora de realizar las operaciones sobre las bases de datos.

Desventajas de utilizar MySQL

- No es viable su uso con grandes bases de datos, a las que se accede continuamente, ya que no implementa una buena escalabilidad.

PostgreSQL 9.1

PostgreSQL es un motor de base de datos relacional, no privativo, orientado a objetos, publicado bajo la licencia Distribución de Software Berkeley (BSD, por sus siglas en inglés).

Características de PostgreSQL

- Ha sido diseñado, programado y construido de tal manera que trabaje con los lenguajes más populares: C/C++, Java, PHP, .Net y Python.
- Disponible en los principales sistemas operativos: Linux, Unix, Mac OS, Windows.
- Soporta un número ilimitado de bases de datos.
- Documentación muy organizada, pública y libre, con comentario de los propios usuarios.
- Altamente adaptable a las necesidades del cliente.

Ventajas de utilizar PostgreSQL

- Posee muy buena escalabilidad, siendo capaz de soportar gran cantidad de peticiones simultáneamente de manera correcta.
- Es capaz de ajustarse a la cantidad de memoria que posea el sistema, de forma óptima.
- Es capaz de unir grandes cantidades de tablas eficientemente.

Desventajas de utilizar PostgreSQL

- La velocidad de respuesta que ofrece este gestor puede parecer más lenta con respecto a MySQL, pero funciona rápido con bases de datos grandes.

Cada uno de estos gestores es idóneo para ciertos campos y se trata de escoger el más conveniente para este caso. Se selecciona PostgreSQL para la gestión de los datos en el MD, pues se requiere una base de datos robusta para la carga de grandes cantidades de datos, a los que se acceda continuamente. Además

porque es de código abierto¹⁶ (open source, en inglés), está disponible en múltiples plataformas y cuenta con varias herramientas gráficas de alta calidad para administrar las bases de datos, entre ellas PgAdmin III. Brinda además estabilidad, integridad y seguridad, siendo estas características esenciales para el manejo de información desde la base de datos fuente al MD. Está definida en las políticas del departamento Sistemas Especializados en Salud.

1.9.3 Herramienta de administración de base de datos

El software de administración de base de datos es la herramienta que controla la creación, el mantenimiento y uso de las bases de datos de una organización.

PgAdmin III 1.14

Es una herramienta de código abierto muy avanzada, que se utiliza para administrar y desarrollar bases de datos en PostgreSQL. Se diseñó para responder a las necesidades de la mayoría de los usuarios, desde escribir simples consultas SQL, hasta desarrollar bases de datos complejas. Está disponible en más de una docena de idiomas y para varios sistemas operativos, incluyendo Microsoft Windows, Linux, Mac OS y Solaris (20).

Se selecciona PgAdmin III como herramienta de administración de base de datos a utilizar por ser multiplataforma, de código abierto y desarrollada por la comunidad de expertos de PostgreSQL.

1.9.4 Herramientas de Inteligencia de Negocios

Las herramientas de inteligencia de negocios se encargan de gestionar la extracción de los datos y su automatización, para mostrar una visión general de todos los procesos de la entidad, facilitar el análisis y la presentación de los datos, apoyando la toma de decisiones (21).

Entre las herramientas de inteligencia de negocios más populares están:

- Oracle Business Intelligence Suite
- Pentaho Open Source Business Intelligence
- JasperSoft Business Intelligence Suite

¹⁶ Código abierto: es el software que está licenciado de tal manera que los usuarios puedan estudiar, modificar y mejorar su diseño mediante la disponibilidad de su código fuente.

Oracle Business Intelligence Suite

El Oracle Business Intelligence Suite es una plataforma de inteligencia de negocios completa e innovadora que permite crear soluciones de IN desde fuentes de datos heterogéneas para la distribución de información, con sistemas Oracle¹⁷ y que no son de Oracle. Cubre un amplio espectro de necesidades de inteligencia de negocios, incluidos los análisis predictivos en tiempo real, entre otras. Es una herramienta de software propietario (22).

Pentaho Open Source Business Intelligence

La corporación Pentaho es el patrocinador primario y propietario del proyecto Pentaho Open Source Business Intelligence, el cual es una iniciativa en curso de la comunidad Open Source, que provee a las empresas de mejores soluciones para cubrir sus necesidades de IN (24).

Las soluciones de Pentaho están escritas en Java. Esta plataforma (o suite, en inglés) incluye un motor de soluciones que integra reportes, análisis para el OLAP, integración de datos, entre otros. Esta herramienta posee como principal característica que es multiplataforma y de código abierto, además de poseer abundante documentación disponible. Esta suite se ha convertido en la líder mundial y en la más ampliamente utilizada herramienta de IN de código abierto.

Pentaho Open Source Business Intelligence es una plataforma que está compuesta por una serie de herramientas entre las que se encuentran:

- Pentaho Data Integration
- Schema Workbench
- Pentaho Analyses Services

Pentaho Data Integration 4.2

Pentaho Data Integration (PDI), es el componente de la suite de Pentaho responsable de los procesos de ETL. Este se encarga de abrir, limpiar e integrar la información disponible en diferentes fuentes de datos y ponerla en manos del usuario. Es conocido también como Kettle, es una herramienta que permite definir transformaciones de forma gráfica, interconectando bloques que tienen diversas funciones (24). Provee un

¹⁷ Oracle: una de las mayores compañías de software del mundo, sus sistemas van desde bases de datos hasta herramientas de gestión.

ambiente de diseño intuitivo y soporta conexión a bases de datos en Oracle, MySQL, PostgreSQL, entre otras.

Características de Pentaho Data Integration

- Entorno gráfico de desarrollo con indicadores de las transformaciones.
- Basado en metadatos¹⁸.
- Basado en dos tipos de objetos: transformaciones (colección de pasos en un proceso de ETL) y trabajos (colección de transformaciones).
- Los proyectos realizados con Kettle son muy fáciles de mantener.

Schema Workbench 3.2

Es un entorno visual para la confección y prueba de esquemas de cubos OLAP. Esta herramienta se utiliza para la creación de los archivos XML¹⁹ que se usan para la construcción de los cubos (25). Tiene dos áreas: la zona en que se muestra la jerarquía del esquema OLAP y la zona de edición de las propiedades de cada elemento. Su desarrollo fue implementado en Java.

Características de Schema Workbench

- Diseñador intuitivo de esquemas multidimensionales.
- Permite probar consultas dimensionales en lenguaje MDX contra el esquema de base de dato en pantalla.

Pentaho Analyses Services 3.3

Es conocido como Mondrian. Esta herramienta permite ejecutar consultas MDX sobre un mercado de datos, y que los resultados sean presentados mediante un navegador; de modo que el usuario pueda interactuar con la información. Mondrian es una de las aplicaciones más importantes de la plataforma Pentaho. Es un servidor para OLAP de código abierto, que permite el análisis de grandes volúmenes de información almacenada en un MD, además es un motor ROLAP desarrollado en Java (26).

¹⁸ Metadatos: son datos que describen o dan información de otros datos, es decir brindan información de localización, estructura y significado de los mismos.

¹⁹ XML: son las siglas en inglés de Extensible Markup Language, es un metalenguaje extensible de etiquetas desarrollado por World Wide Web Consortium.

Características de Pentaho Analyses Services

- Exploración dimensional de los datos.
- Permite ser instalado en servidores de aplicación como Apache Tomcat y JBoss.
- Es el motor de la mayoría de las soluciones de IN.
- Cuenta con una cache que le proporciona velocidad.

Para acceder a las funcionalidades que brinda Mondrian se requiere de un cliente, es el caso de JPivot.

JPivot

Visor web OLAP de código abierto, que se utiliza para construir tablas generadas de forma dinámica y permite a los usuarios realizar consultas OLAP por medio del lenguaje MDX (7). Permite realizar operaciones tales como: apreciar los datos en un mayor detalle, bajando por la jerarquía de una dimensión (Drill Down, en inglés) o apreciar los datos en menor nivel de detalle, subiendo por la jerarquía de una dimensión (Drill Up, en inglés).

JasperSoft Business Intelligence Suite

JasperSoft es la empresa que ha construido la plataforma de inteligencia de negocios JasperSoft BI Suite, la cual es de código abierto. Jasper ha integrado en su solución proyectos preexistentes y consolidados pero no los ha absorbido. Esta estrategia le hace depender de Talend en cuanto a solución ETL y de Mondrian como motor OLAP.

JasperSoft Business Intelligence Suite cuenta con varias herramientas, entre las que se encuentran:

- Jasper ETL (Talend Studio)
- Mondrian
- Jasper Analysis

Jasper ETL

Es en realidad Talend Studio, es la otra gran solución ETL junto a Kettle, pero desarrollada por la empresa Talend. Esta no ha sido absorbida por Jasper y Talend sigue siendo una empresa que ofrece sus productos de forma independiente. Talend Studio está orientado al tipo de usuario programador, con un nivel de conocimientos técnicos superior al requerido por Kettle, por lo que la curva de aprendizaje es más

larga. Tiene una exigencia muy alta de recursos del ordenador, pues está basado en Eclipse. Es más complicado depurar el flujo que con Kettle.

Mondrian

El motor OLAP que utiliza la plataforma es Mondrian, ya referido en el apartado de Pentaho.

Jasper Analysis

Es el visor web para las soluciones JasperSoft BI Suite, el cual está basado en JPivot con pequeñas modificaciones.

A partir del estudio de las principales herramientas de inteligencia de negocios se decide utilizar Pentaho Open Source Business Intelligence como plataforma de IN. Se escoge por estar disponible en los principales sistemas operativos, por ser fácil de instalar y configurar, por ser de código abierto y por poseer amplia documentación disponible. Además ofrece soluciones propias para desarrollar, mantener y explotar soluciones de inteligencia de negocios. Existen experiencias previas en el uso de esta herramienta en la Universidad de las Ciencias Informáticas.

1.9.5 Servidores de Aplicaciones

Un servidor web es un programa que corre sobre el servidor que escucha las peticiones HTTP²⁰ que le llegan y las satisface. Dependiendo del tipo de petición, el servidor web buscará una página web o bien ejecutará un programa en el servidor. De cualquier modo, siempre devolverá algún tipo de resultado HTML al cliente o navegador que realizó la petición.

Entre los servidores de aplicaciones estudiados se encuentra:

- JBoss
- Apache Tomcat

JBoss

Es el servidor de aplicaciones de Edición de Empresas Java (J2EE, por sus siglas en inglés) más ampliamente desarrollado del mercado. Combina una arquitectura orientada a servicios revolucionaria,

²⁰ HTTP: son las siglas en inglés de Hypertext Transfer Protocol, en español protocolo de transferencia de hipertexto. Método más común de intercambio de información en la world wide web, mediante el cual se transfieren las páginas web a un ordenador.

con una licencia de código abierto (27). JBoss puede ser descargado, utilizado, incrustado y distribuido sin restricciones por la licencia.

Características principales de JBoss

- Alto desempeño.
- Arquitectura modular.
- Confiable a nivel de empresa.

Apache Tomcat

Servidor web escrito en Java, funciona en cualquier sistema operativo que disponga de una máquina virtual Java (28). Esta herramienta puede ser utilizada por plataformas como: Windows, Linux, con sus ficheros de configuración específicos, y consejos paso a paso para implementar y correr aplicaciones web.

Características de Apache Tomcat

- Es de código abierto.
- No requiere mucha memoria para iniciar.
- Amplia documentación en la web.

Después del estudio realizado sobre los distintos servidores de aplicaciones, se selecciona para el despliegue del mercado de datos el Apache Tomcat 7.0, pues es de código abierto e implementado con tecnología Java. Además facilita la integración con Apache 2.2 que es el servidor web donde se encuentra el sistema Synta, debido a que ambos servidores están desarrollados por la fundación Apache.

La realización de una investigación detallada sobre los MD existentes en el área de la salud, tanto a nivel internacional como nacional, permitió identificar y analizar las tecnologías existentes para su desarrollo. Sobre la base de dicho análisis se decidió utilizar como herramienta de modelado Enterprise Architect, como gestor de base de datos PostgreSQL, como administrador de base de datos pgAdmin III, como plataforma de inteligencia de negocios Pentaho Open Source Business Intelligence y como metodología Hefesto. Además se seleccionó el Schema Workbench y el Mondrian pertenecientes a la plataforma de Pentaho para el trabajo con los cubos. También se eligió Pentaho Data Integration como herramienta para el proceso ETL, JPivot para visualizar la información gestionada del MD y Apache Tomcat como servidor web.

Capítulo 2: Análisis y diseño del mercado de datos

En este capítulo se detallan las necesidades y perspectivas del cliente en cuanto al negocio en general. Además se llevan a cabo las tres primeras etapas que propone la metodología Hefesto para el diseño del mercado de datos, obteniéndose el modelo lógico.

2.1 Propuesta del sistema

Se propone el desarrollo de un mercado de datos que permita el análisis estadístico de la información y contribuya a la toma de decisiones referente al consumo de medicamentos, en el Departamento de Farmacoepidemiología del Ministerio de Salud Pública. El MD obtendrá la información de la base de datos del sistema Synta, específicamente de tres módulos. Del Módulo Consumo de Medicamentos se extraen los medicamentos que han circulado cada año en el país y su consumo. La información de los pacientes inscritos al consumo de un medicamento controlado, es tomada del Módulo Tarjeta Control; y debido a que gran cantidad de información está nombrada, es necesario obtenerla del sistema Nomencladores. Ejemplo de esta información son los medicamentos y las provincias. Para la carga del MD es necesario hacer la triangulación entre estos tres módulos. Posteriormente el MD se integrará al Módulo Consumo de Medicamentos, permitiéndole a los especialistas contar con un sólo sistema que gestione y analice el consumo de medicamentos del país. Entre los reportes que se esperan obtener están los siguientes:

- Cantidad de pacientes que consumen medicamentos controlados en un período de tiempo.
- Porcentaje de pacientes que consumen medicamentos controlados, por sexo, raza y grupo etario, en un período de tiempo determinado.
- Total de consumo de un medicamento en una provincia y año determinado.
- Cantidad de medicamentos consumidos por fabricante y producción en un año determinado.

2.2 Fases de la metodología Hefesto

La metodología Hefesto está compuesta de cuatro fases las cuales facilitan el proceso de construcción de un MD, partiendo de la recolección de las necesidades de información de los usuarios, posteriormente la confección del esquema lógico y concluyendo con los respectivos procesos de extracción, transformación y carga. A continuación se ejemplifican las tres primeras etapas que componen esta metodología. El último de los procesos es explicado en el capítulo tres.

2.2.1 Análisis de los requerimientos

Se identifican los requerimientos de los usuarios a través de preguntas que permitan entender el negocio que se evalúa. Luego, se analizan estas preguntas a fin de identificar cuáles serán los indicadores y perspectivas que serán tomados en cuenta en la construcción del MD. Finalmente se confecciona un modelo conceptual en donde se visualiza el resultado obtenido en este paso (7).

Identificar preguntas

Se realizó una entrevista al cliente donde se identificaron las necesidades de información de este, los resultados esperados y los reportes que necesitan para las actividades dentro del departamento de Farmacoepidemiología del MINSAP, de manera que se apoye el proceso de toma de decisiones.

De la entrevista realizada se obtienen las preguntas claves del negocio. Los especialistas desean conocer:

- Cantidad de pacientes que consumen medicamentos controlados, por sexo y raza en un período de tiempo.
- Cantidad de pacientes que consumen medicamentos controlados, por grupo etario en un período de tiempo.
- Cantidad de pacientes que consumen medicamentos controlados, por provincia.
- Cantidad de pacientes que consumen medicamentos controlados por fabricante en un período de tiempo.
- Porcentaje de pacientes que consumen medicamentos controlados, por sexo, raza y grupo etario en un período de tiempo.
- Porcentaje de pacientes que consumen medicamentos controlados por provincia y fabricante en un período de tiempo.
- Cantidad de pacientes que consumen medicamentos controlados.
- Cantidad de medicamentos consumidos en una provincia y año determinado.
- Cantidad de medicamentos consumidos por fabricante y producción (nacional o internacional).
- Total de consumo de un medicamento en una provincia y año determinado.
- Total de consumo de un medicamento por fabricante y producción.

Identificar indicadores y perspectivas

Una vez que se establecieron las preguntas del negocio, se procedió a su descomposición en identificadores y perspectivas de análisis.

Los indicadores son valores numéricos y representan lo que se desea analizar concretamente. En cambio, las perspectivas se refieren a los objetos mediante los cuales se quiere examinar los indicadores, con el fin de responder a las preguntas planteadas (8).

A continuación se detallan los indicadores y perspectivas identificados:

Cantidad de pacientes que consumen medicamentos controlados, por sexo, raza y tiempo.

Indicador

Perspectivas

Cantidad de pacientes que consumen medicamentos controlados, por grupo etario y tiempo.

Indicador

Perspectivas

Cantidad de pacientes que consumen medicamentos controlados, por provincia.

Indicador

Perspectivas

Cantidad de pacientes que consumen medicamentos controlados por fabricante y tiempo.

Indicador

Perspectivas

Porcentaje de pacientes que consumen medicamentos controlados, por sexo, raza y grupo etario en

Indicador

Perspectivas

un período de tiempo.

Perspectivas

Porcentaje de pacientes que consumen medicamentos controlados por provincia, fabricante y tiempo.

Indicador

Perspectivas

Capítulo 2: Análisis y diseño

Cantidad de pacientes que consumen medicamentos controlados.
Indicador *Perspectivas*

Cantidad de medicamentos consumidos en una provincia y año determinado.
Indicador *Perspectivas*

Cantidad de medicamentos consumidos por fabricante y producción.
Indicador *Perspectivas*

Total de consumo de un medicamento en una provincia y año determinado.
Indicador *Perspectivas*

Total de consumo de un medicamento por fabricante y producción.
Indicador *Perspectivas*

Modelo conceptual

En esta etapa se construye el modelo conceptual de los datos del MD, a partir de los indicadores y perspectivas obtenidas en el paso anterior. Este modelo permite observar con claridad el alcance del sistema y obtener un alto nivel de definición de los datos.

La representación del modelo conceptual puede observarse en la Figura 9, siendo analizados los indicadores (derecha) y perspectivas (izquierda) desde los medicamentos controlados y los pacientes que consumen estos medicamentos.

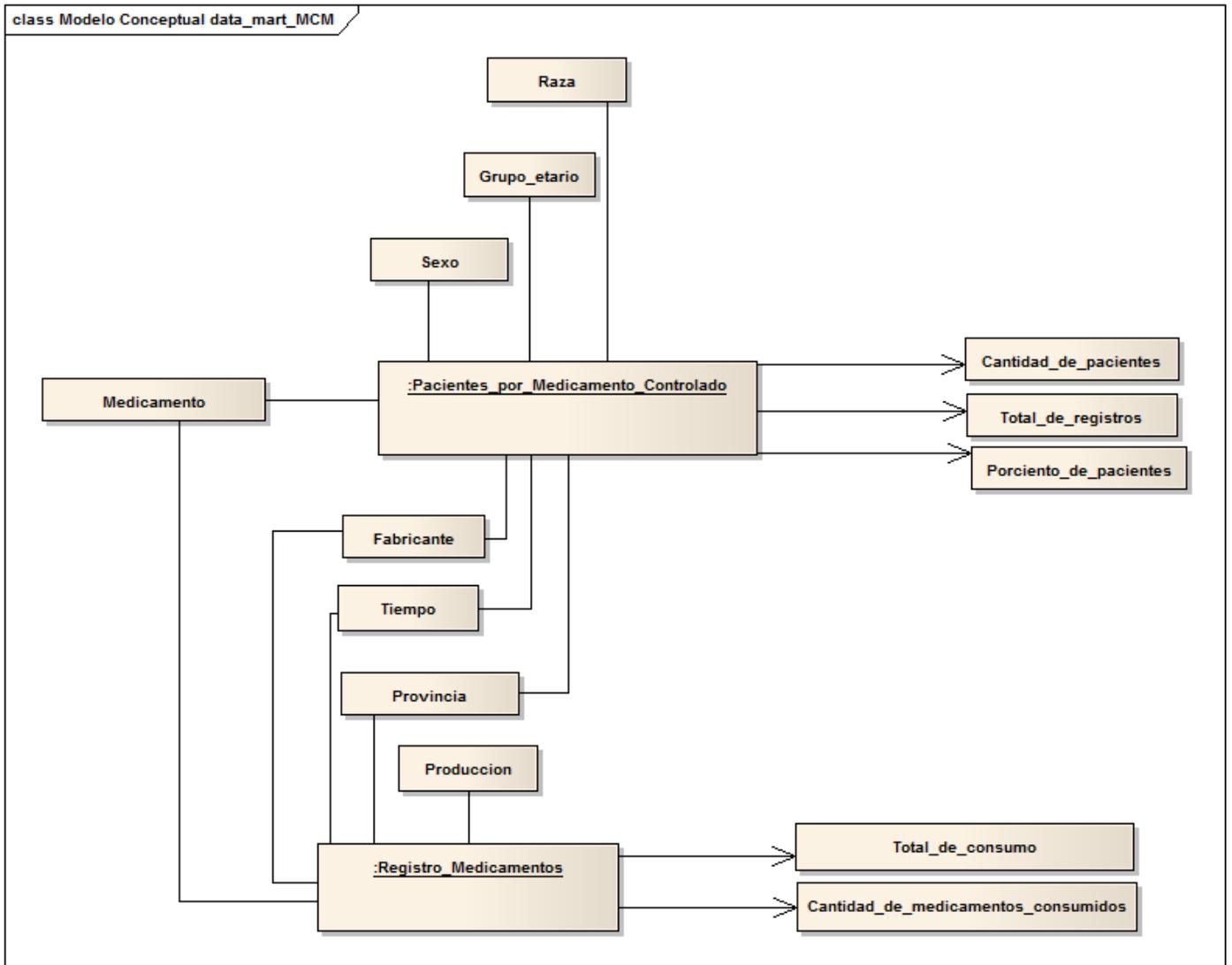


Figura 9: Modelo conceptual

2.2.2 Análisis de los OLTP

Se analizan las fuentes OLTP 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. Luego se definen qué campos se incluirán en cada perspectiva y se amplía el modelo conceptual con la información obtenida en este paso (7).

Conformar indicadores

En este paso se explica cómo se calculan los indicadores y sus respectivas funciones de sumalización.

No.	Indicador	Función Sumarización	Descripción
1	Cantidad de pacientes	DISTINCT COUNT	Cuenta el total de pacientes diferentes.
	Total de registros	count	Cuenta el total de registros
2	Porcentaje de pacientes	(Cantidad de pacientes * 100)/Total de registros)	Calcula el porcentaje de pacientes que consumen un medicamento controlado.
3	Total de consumo	SUM	Suma los consumo de un medicamento.
4	Cantidad de medicamentos	DISTINCT COUNT	Cuenta el total de medicamentos diferentes.

Tabla 2: Conformación de indicadores

Establecer correspondencias

Se examinan las fuentes OLTP disponibles que contienen la información requerida, así como sus características, identificando las correspondencias entre el modelo conceptual y las fuentes de datos. Logrando que todos los elementos del modelo conceptual estén correspondidos en los OLTP.

Las correspondencias que se establecen entre la base de datos del sistema Synta y el mercado de datos fueron las siguientes:

- El campo `id_medicamento` de la tabla `tb_c_r_medicamento_cuadro_basico` se relaciona con la perspectiva “Medicamento” en dependencia del tipo de medicamento (controlado o no controlado).
- El campo `valor` de la tabla `tb_n_nomenclador_valor` se relaciona con la perspectiva “Sexo”, cuando el campo `nombre` de la tabla `tb_n_campo` sea igual a `sexo`.
- El campo `tb_tc_grupo_etario_id` de la tabla `tb_tc_certificado` se relaciona con la perspectiva “Grupo_Etario”.
- El campo `valor` de la tabla `tb_n_nomenclador_valor` se relaciona con la perspectiva “Raza”, cuando el campo `nombre` de la tabla `tb_n_campo` sea igual a `raza`.

Capítulo 2: Análisis y diseño

- El campo valor de la tabla tb_n_nomenclador_valor se relaciona con la perspectiva “Fabricante”, cuando el campo nombre de la tabla tb_n_campo sea igual a fabricante.
- El campo fecha de la tabla tb_c_cuadro_basico y el campo id_mes de la tabla tb_c_r_consumo_mes_provincia se relacionan con la perspectiva “Tiempo”.
- El campo id_provincia de la tabla tb_c_r_consumo_mes_provincia se relaciona con la perspectiva “Provincia”.
- El campo producción de la tabla tb_c_r_medimento_cuadro_basico se relaciona con la perspectiva “Producción”.

Los campos tb_tc_grupo_etario_id, id_medimento, id_provincia, id_mes están nombrados.

Nivel de granularidad

Después de haber establecido las relaciones con los OLTP, se seleccionan los campos que contendrá cada perspectiva, ya que a través de estos se examinan y filtran los indicadores. Se presentan a los usuarios los datos disponibles para cada perspectiva y se escogen los más relevantes para consultar y analizar. Esta acción determina la granularidad de la información en el MD. A continuación la tabla 2 muestra los campos que contendrá la perspectiva “Medimento”, en el Anexo 2 se muestran los campos que contendrán las demás perspectivas identificadas:

Nombre: Perspectiva “Medimento”		
Atributos	Tipo	Descripción
id_medimento	Integer	Identificador del medicamento.
nombre_medimento	Character Varing	Nombre del medicamento.
cantidad_presentacion	Integer	Número de medida del medicamento.
nombre_unidad_medida	Character Varing	Unidad de medida en que se presenta el medicamento.
nombre_forma_farmaceutica	Character Varing	Forma farmacéutica en que se presenta el medicamento.
nombre_via_administracion	Character Varing	Vía de administración del medicamento.
controlado	Character Varing	Tipo de medicamento (controlado o no).

Tabla 3: Campos que contendrá la perspectiva “Medimento”

Modelo Conceptual ampliado

En este paso se amplía el modelo conceptual, colocando bajo cada perspectiva los campos seleccionados y bajo cada indicador su respectiva función de sumarización.

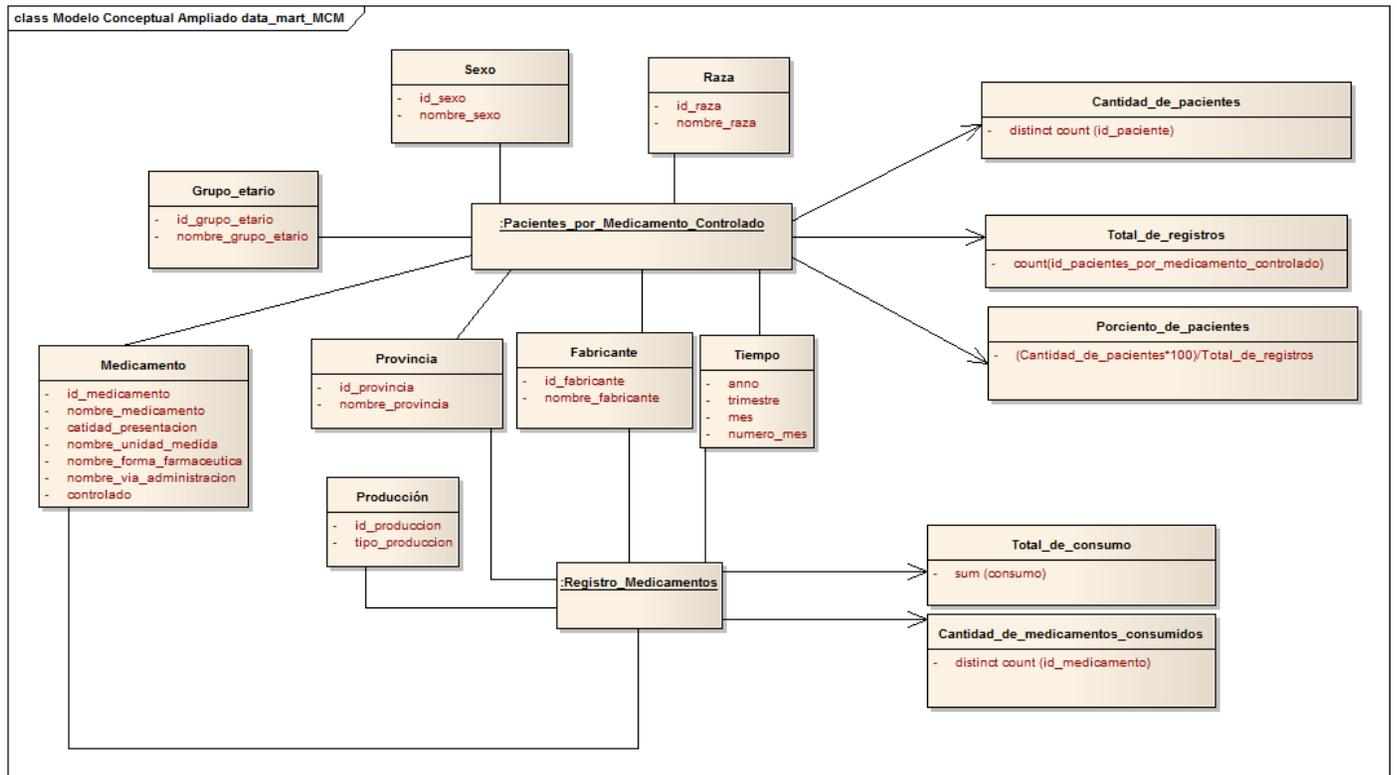


Figura 10: Modelo conceptual ampliado

2.2.3 Modelo Lógico del mercado de datos

A continuación se confecciona el modelo lógico de la estructura del MD, teniendo como base el modelo conceptual que ya ha sido creado. 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 las tablas de hechos. Finalmente, se realizan las uniones pertinentes entre las tablas (8).

Tipo de Modelo Lógico del MD

El tipo de esquema que se utiliza para contener la estructura del MD es el modelo constelación de hechos, explicado anteriormente, debido a sus ventajas con respecto a otros esquemas.

Tablas de dimensiones

En este paso se diseñan las tablas de dimensiones que forman parte del MD. Cada perspectiva definida en el modelo conceptual constituye una dimensión. Para ello se toma cada perspectiva con sus campos relacionados y se elige un 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 siguiente figura se muestra la transformación desde la perspectiva “Medicamento” a la dimensión “d_medicamento”. Para ver las transformaciones de las restantes dimensiones consultar el Anexo 3.

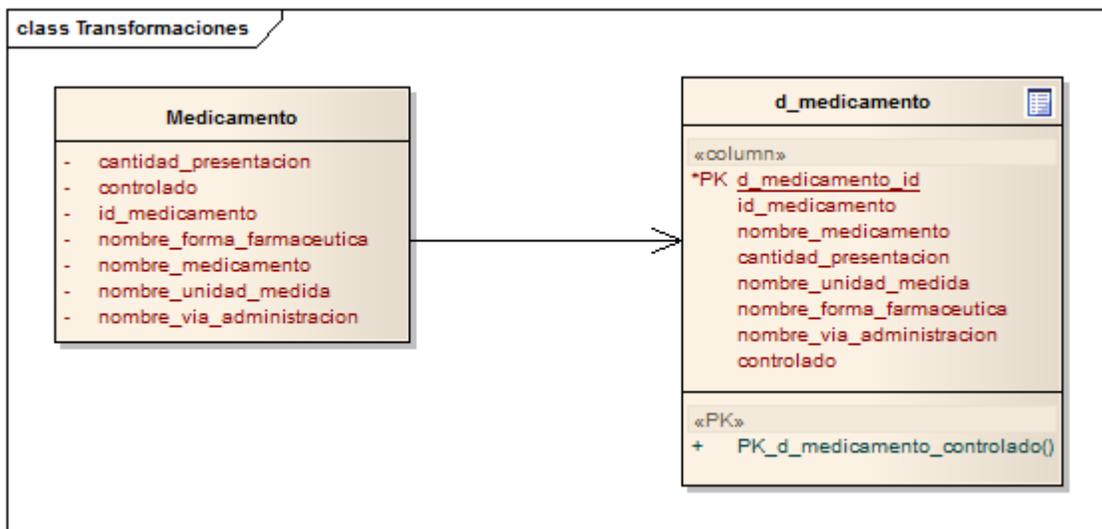


Figura 11: Dimensión “d_medicamento”

Tablas de hechos

Las tablas de hechos se confeccionan teniendo en cuenta el análisis de las preguntas realizadas en pasos anteriores y sus respectivos indicadores y perspectivas. Cada tabla de hechos posee un nombre que la identifica, sus indicadores correspondientes, y su clave está conformada por una clave definida y la

Capítulo 2: Análisis y diseño

combinación de las claves que hereda de las tablas de dimensiones y puentes²¹ relacionados. Se le asignan los nombres de “hecho_pacientes_por_medicamento_controlado” y “hecho_registro_medicamentos” a las tablas de hechos del MD.

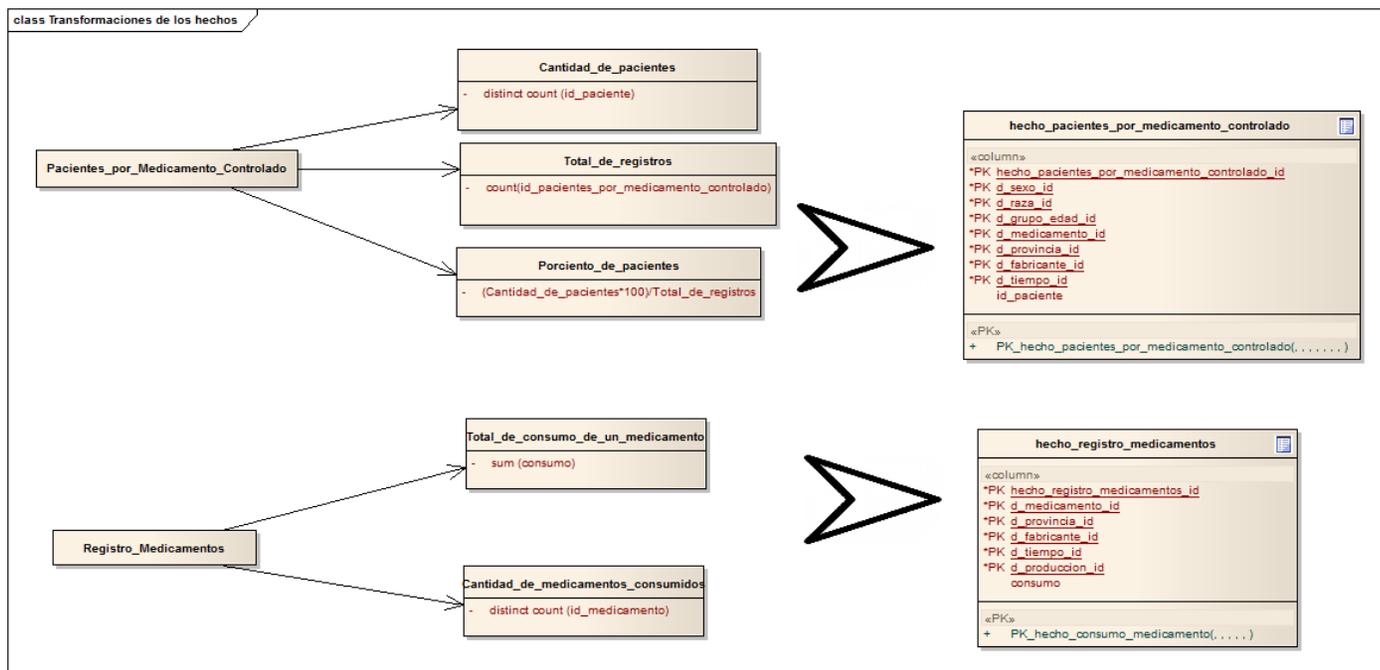


Figura 12: Diseño de las tablas de hechos

La tabla 3 muestra la descripción del “hecho_pacientes_por_medicamento_controlado” y en el Anexo 4 se observa la descripción de las demás tablas de hechos y puentes:

Nombre: “hecho_pacientes_por_medicamento_controlado”		
Atributos	Tipo	Descripción
hecho_pacientes_por_medicamento_controlado_id	Integer	Identificador del hecho.
d_sexo_id	Integer	Identificador que proviene de la tabla “d_sexo”.
d_raza_id	Integer	Identificador que proviene de la tabla “d_raza”.
d_grupo_etario_id	Integer	Identificador que proviene de la tabla

²¹ Puente: dimensión creada a partir de la relación de muchos a muchos entre un hecho y una dimensión.

		"d_grupo_etario".
d_medamento_id	Integer	Identificador que proviene de la tabla "d_medamento".
d_provincia_id	Integer	Identificador que proviene de la tabla "d_provincia".
d_fabricante_id	Integer	Identificador que proviene de la tabla "d_fabricante".
d_tiempo_id	Integer	Identificador que proviene de la tabla "d_tiempo".
id_paciente	Integer	Identificador del paciente.

Tabla 4: Descripción de la tabla "hecho_pacientes_por_medamento_controlado"

Uniones

Se realizan las uniones correspondientes entre las tablas de dimensiones y las tablas de hechos como se muestra a continuación.

Capítulo 2: Análisis y diseño

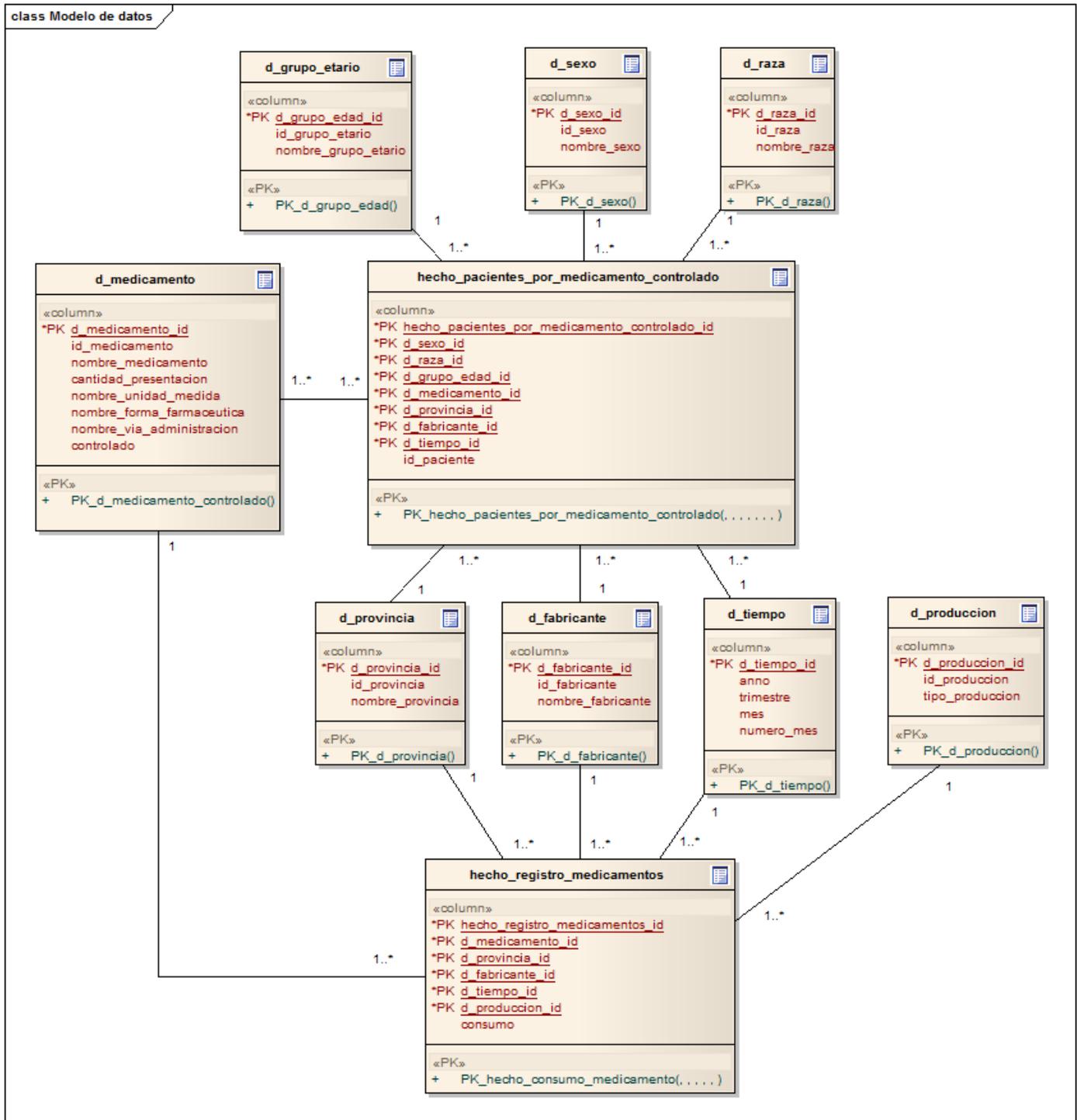


Figura 13: Uniones de las dimensiones (modelo lógico)

Capítulo 2: Análisis y diseño

La obtención de las preguntas claves del negocio e identificación de los indicadores y perspectivas de análisis, dieron paso a la elaboración del modelo conceptual y conceptual ampliado del MD, haciendo uso de la metodología Hefesto. Además de la identificación de las dimensiones, hechos y medidas, realizándose las uniones pertinentes entre las tablas.

Capítulo 3: Integración de datos

En este capítulo se expone la última etapa de la metodología Hefesto, todo lo relacionado con la construcción de los procesos de extracción, transformación y carga del mercado de datos, empleándose la herramienta Pentaho Data Integration.

3.1 Integración de datos

Una vez diseñado el MD se procede a poblarlo con datos, utilizando el proceso de extracción, transformación y carga. Luego se definen reglas y estrategias para su carga inicial y actualización.

3.1.1 Carga Inicial

En este paso se realiza la carga inicial al MD, poblando el modelo de datos que se construyó anteriormente, para lo cual se debe llevar adelante una serie de tareas básicas: primeramente se cargarán los datos de las tablas de dimensiones, luego los de las tablas de hechos y por último la información del puente, teniendo en cuenta la correspondencia entre cada elemento. Para esto es necesario realizar las transformaciones oportunas, las cuales constituyen un elemento imprescindible dentro de este paso.

Transformaciones de las dimensiones y los hechos

Se detallan las transformaciones llevadas a cabo en el proceso ETL, haciendo uso de la herramienta Pentaho Data Integration. Entre los componentes más utilizados se encuentran: entrada tabla, ordenar filas, juntar filas, búsqueda de valores en base de datos e insertar/actualizar.

Seguidamente se muestra la transformación de la dimensión “d_medicamento” (ver Figura 14). Las transformaciones del resto de las dimensiones se pueden consultar en el Anexo 5:

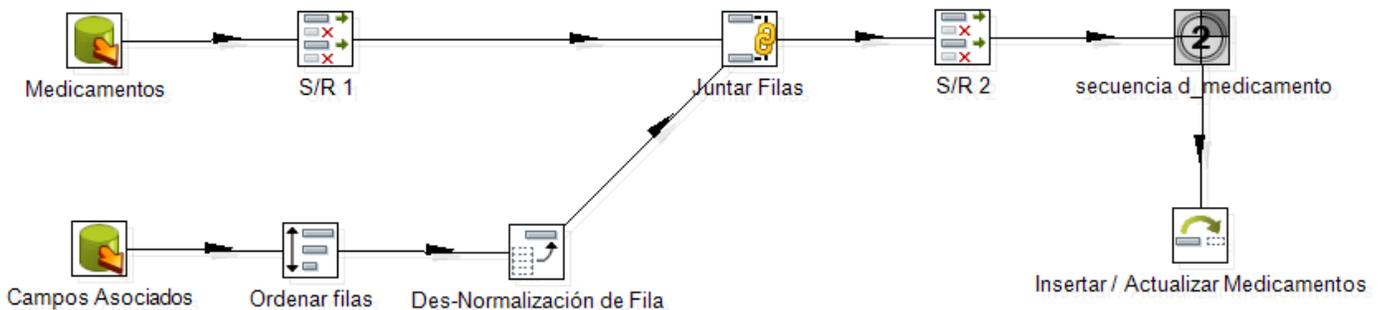


Figura 14: Transformación de la dimensión “d_medicamento”

Capítulo 3: Integración de datos

- Utilizando el componente Entrada Tabla (Medicamentos), se explora la fuente de datos, entiéndase, la base de datos del sistema Synta, y se extrae la información que se considera relevante para satisfacer los requisitos del cliente. Para esto se crea una consulta SQL como se muestra a continuación, obteniéndose los medicamentos que serán objetos de análisis:

```
SELECT DISTINCT
  tb_n_nomenclador.nombre,
  tb_n_nomenclador.root_id,
  tb_n_nomenclador.id
FROM
  tb_n_nomenclador
WHERE
  (tb_n_nomenclador.id <> tb_n_nomenclador.root_id AND
  tb_n_nomenclador.root_id = (SELECT tb_n_nomenclador.root_id FROM tb_n_nomenclador WHERE tb_n_nomenclador.nombre = 'FNM') AND
  tb_n_nomenclador.id IN (SELECT tb_c_r_medicamento_cuadro_basico.id_medicamento FROM tb_c_r_medicamento_cuadro_basico))
```

Figura 15: Consulta SQL para obtener los medicamentos que serán analizados

- A la par de esto, se obtienen los campos asociados a los medicamentos, valiéndose del componente Entrada Tabla (Campos Asociados). La siguiente consulta SQL permite obtener esta información:

```
SELECT
  tb_n_nomenclador.id,
  tb_n_nomenclador.nombre,
  tb_n_campo.nombre,
  tb_n_nomenclador_valor.valor
FROM
  tb_n_nomenclador
  INNER JOIN tr_n_campo_nomenclador ON (tb_n_nomenclador.id = tr_n_campo_nomenclador.tb_n_nomenclador_id)
  INNER JOIN tb_n_campo ON (tr_n_campo_nomenclador.tb_n_campo_id = tb_n_campo.id)
  INNER JOIN tb_n_nomenclador_valor ON (tr_n_campo_nomenclador.id = tb_n_nomenclador_valor.tb_n_nomenclador_id)
WHERE
  (tb_n_nomenclador.id <> tb_n_nomenclador.root_id) and tb_n_nomenclador.root_id = (SELECT
  tb_n_nomenclador.root_id
FROM
  tb_n_nomenclador
WHERE
  (tb_n_nomenclador.nombre = 'FNM'))
```

Figura 16: Consulta SQL para obtener los campos asociados a los medicamentos

- Luego con el componente Ordenar Filas se ordenaron los campos obtenidos por: identificador, nombre del medicamento, nombre del campo asociado y valor.
- Posteriormente, manipulando el componente Des-Normalización de Fila se filtran los campos asociados siguientes: cantidad de presentación, unidad de medida, forma farmacéutica, vía de administración, fabricante y controlado.
- Utilizando a continuación el componente Juntar Filas se unieron las informaciones provenientes de ambos flujos de información, a través del identificador del medicamento.
- Recurriendo además al componente Seleccionar/Renombrar Valores (S/R 1 y S/R 2) se renombraron los medicamentos y sus campos relacionados respectivamente.

Capítulo 3: Integración de datos

- Finalmente con el componente Añadir Secuencia (secuencia d_med medicamento) se genera la llave primaria de la dimensión medicamento; y se cargan los datos en esta dimensión haciendo uso del componente Insertar/Actualizar (Insertar/Actualizar Medicamentos).

Una vez realizadas las transformaciones de las dimensiones se procede a transformar los hechos. Se puede observar la transformación del “hecho_pacientes_por_med medicamento_controlado” en la siguiente imagen. Consultar el Anexo 6 para ver las restantes transformaciones de los hechos y puente:

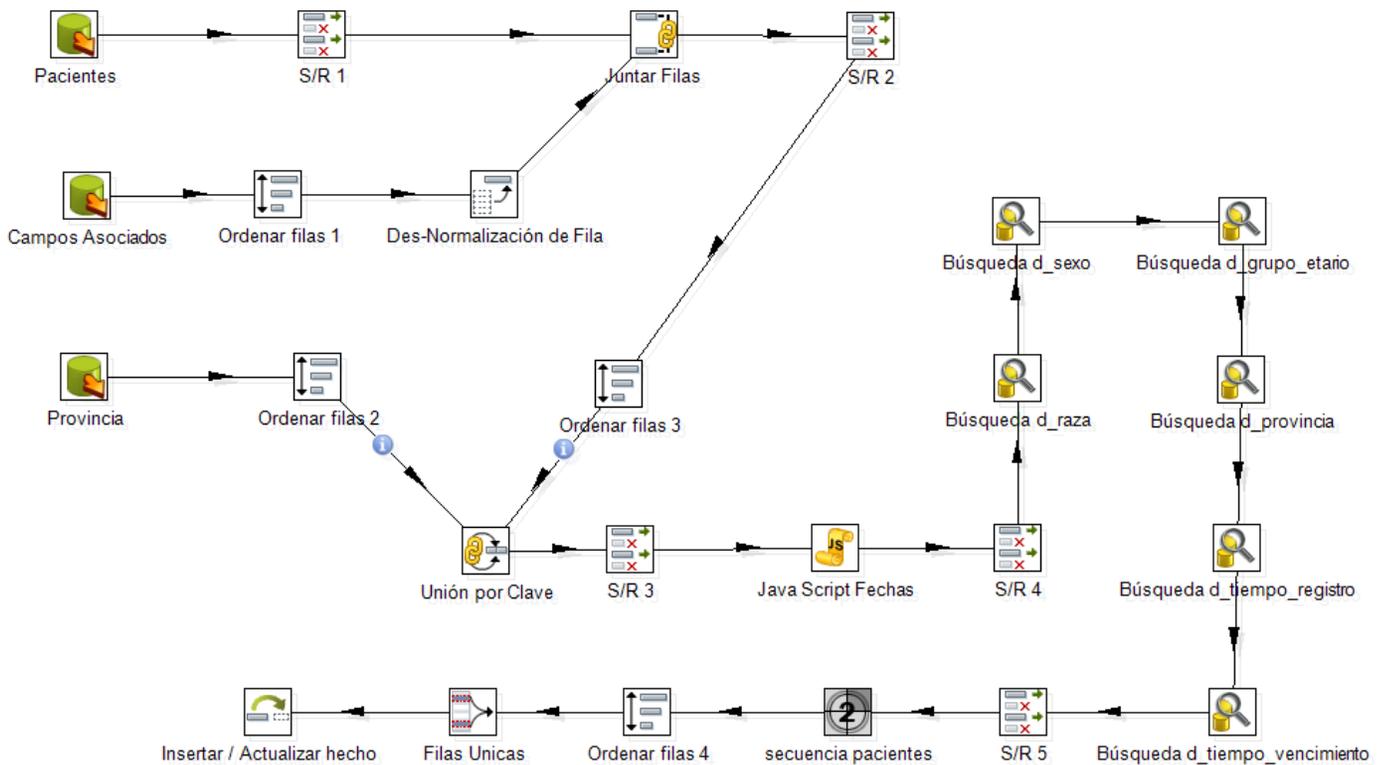


Figura 17: Transformación de la tabla “hecho_pacientes_por_med medicamento_controlado”

- Se utiliza el componente Entrada Tabla (Pacientes) para obtener los identificadores de los pacientes que serán analizados. Para esto es necesario realizar la siguiente consulta SQL a la fuente de información:

```
SELECT DISTINCT
  `tb_n_nomenclador`.`id`
FROM
  `tb_tc_certificado`
  INNER JOIN `tb_n_nomenclador` ON (`tb_tc_certificado`.`tb_tc_paciente_id` = `tb_n_nomenclador`.`id`)
```

Figura 18: Consulta SQL para obtener los pacientes que serán analizados

- A la vez se manipula el componente Entrada Tabla (Campos Asociados), con el objetivo de obtener los campos asociados a los pacientes. A continuación se muestra la consulta SQL que obtiene esta información:

```
SELECT
  `tb_n_nomenclador_valor`.`valor`,
  `tb_n_campo`.`nombre`,
  `tb_n_nomenclador`.`id`,
  `tb_n_nomenclador1`.`id`
FROM
  `tr_n_campo_nomenclador`
  INNER JOIN `tb_n_nomenclador` ON (`tr_n_campo_nomenclador`.`tb_n_nomenclador_id` = `tb_n_nomenclador`.`id`)
  INNER JOIN `tb_n_nomenclador_valor` ON (`tr_n_campo_nomenclador`.`id` = `tb_n_nomenclador_valor`.`tr_n_campo_nomenclador_id`)
  INNER JOIN `tb_tc_certificado` ON (`tb_n_nomenclador`.`id` = `tb_tc_certificado`.`tb_tc_paciente_id`)
  INNER JOIN `tb_n_campo` ON (`tr_n_campo_nomenclador`.`tb_n_campo_id` = `tb_n_campo`.`id`)
  INNER JOIN `tb_n_nomenclador` `tb_n_nomenclador1` ON (`tb_n_nomenclador_valor`.`valor` = `tb_n_nomenclador1`.`nombre`)
WHERE
  (`tb_n_campo`.`nombre` = 'Sexo' OR
  `tb_n_campo`.`nombre` = 'Raza') AND
  `tb_n_nomenclador`.`deleted_at` IS NULL
ORDER BY
  `tb_n_nomenclador`.`id`
```

Figura 19: SQL para obtener los campos asociados a los pacientes

- Seguidamente se ordenaron los campos utilizando el componente Ordenar Filas (Ordenar Filas 1), por: identificador del paciente, identificador del sexo y la raza, nombre del campo asociado y valor.
- Después, manejando el componente Des-Normalización de Fila se filtraron los campos asociados raza y sexo para cada paciente.
- Posteriormente se utilizó el componente Juntar Filas, el cual es el responsable de realizar el producto cartesiano de los elementos que tiene como entrada, uniéndolos por el identificador del paciente.
- También se usa el componente Seleccionar/Renombrar Valores (S/R 1 y S/R 2), posibilitando escoger y renombrar los campos que continuarán hacia el siguiente paso.
- Además con el componente Entrada Tabla (Provincia) se logra obtener la provincia, el grupo etario, la fecha de registro y la fecha de vencimiento para cada paciente.
- Luego con el componente Ordenar Filas (Ordenar Filas 2 y Ordenar Filas 3) se ordenaron ambos flujos de información por el identificador del paciente.

Capítulo 3: Integración de datos

- Valiéndose del componente Unión por Clave se unieron uno y otro flujo de datos por el identificador del paciente, empleando el tipo de unión INNER. Obteniéndose así cada paciente con sus campos asociados más el identificador de su grupo etario, de su provincia, y la fecha de registro y vencimiento. Estos campos son renombrados utilizando el componente Seleccionar/Renombrar Valores (S/R 3).
- Para el tratamiento de fechas se recurre al componente Valor Java Script Modificado (Java Script Fechas), el cual permite desglosar cada elemento de las fechas. Utilizándose además el componente Seleccionar/Renombrar Valores (S/R 4) se convierten las fechas tratadas a valores enteros.

```
Script1 ✖  
//Script here  
  
var fr_temp=fecha_registro;  
var fr=fr_temp.split("/");  
  
var fr_anno=parseInt(fr[0]);  
var fr_mes=parseInt(fr[1]);  
  
var fv_temp=fecha_vencimiento;  
var fv=fv_temp.split("/");  
  
var fv_anno=parseInt(fv[0]);  
var fv_mes=parseInt(fv[1]);
```

Figura 20: Tratamiento de fechas

- Inmediatamente se utiliza el componente Búsqueda de Valores en Base de Datos (Búsqueda d_raza, Búsqueda dsexo, Búsqueda d_grupo_etario, Búsqueda d_provincia, Búsqueda d_tiempo_registro, d_tiempo_vencimiento), el cual se encargará de que los datos a integrar sean correctos. Se renombran los campos utilizando el componente Seleccionar/Renombrar Valores (S/R 5).
- Consecutivamente se genera la llave primaria del hecho pacientes por medicamento controlado, con el componente Añadir Secuencia (secuencia pacientes).
- Después se ordenan las filas ascendentemente basándose en los valores de los campos, para esto se utiliza el componente Ordenar Filas (Ordenar filas 4). Cuando las filas quedan ordenadas se usa el componente Filas Únicas, el cual elimina filas duplicadas y deja solo ocurrencias únicas de los pacientes.
- Por último se cargan los datos en el hecho haciendo uso del componente Insertar/Actualizar (Insertar/Actualizar hecho).

El último de los subprocesos ETL es la carga de la información al mercado de datos, para esto se diseña un trabajo²² de automatización denominado carga inicial (ver Figura 21). Esta carga inicial se refiere precisamente a la primera carga de datos que se le realiza al mercado de datos. Entre los componentes más usados se encuentran: inicio de un trabajo y ejecución de una transformación. El trabajo cuenta también, con el componente Dummy para el manejo de errores en cualquier etapa, que informa mediante el componente Display MsgBoxInfo (Mensaje de error y Mensaje de carga exitosa), al usuario, sobre las acciones realizadas. Fue necesario además crear un trabajo para añadir al depósito aquellos datos nuevos que hayan sido generados después de la carga inicial (o de la última actualización), este trabajo se denomina carga incremental (ver Figura 22).

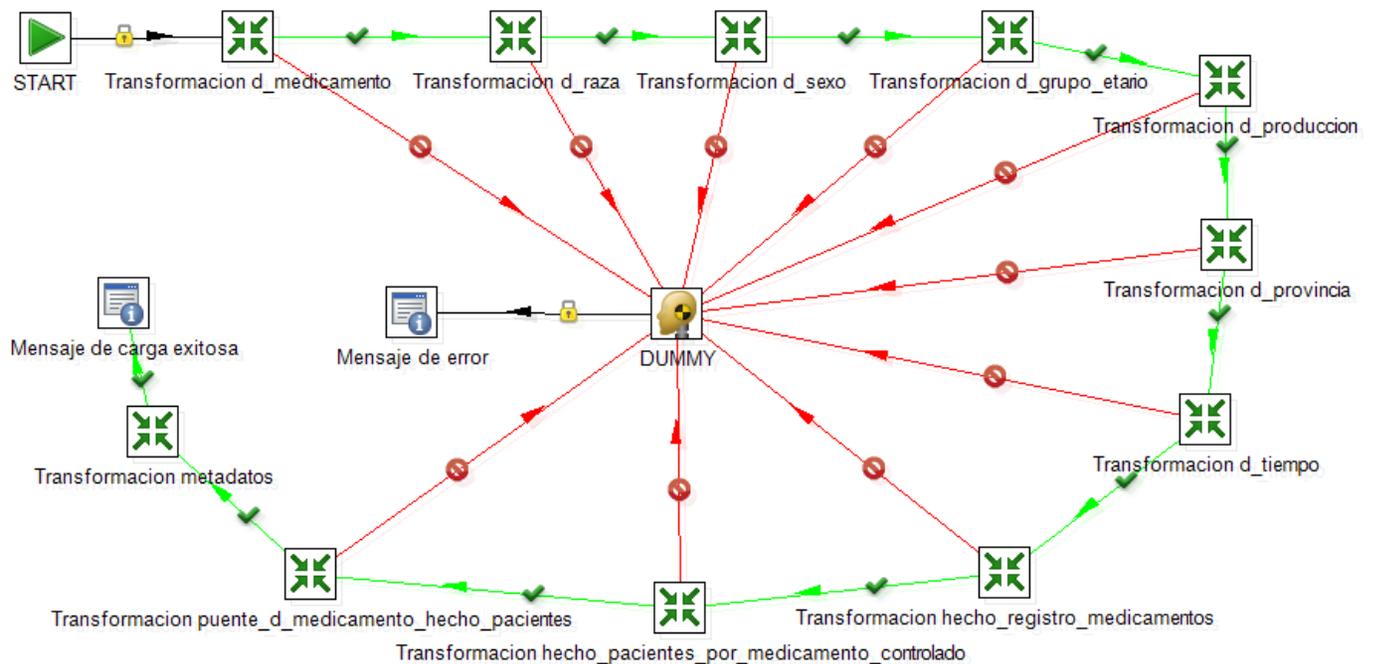


Figura 21: Ejecución del trabajo (Carga inicial)

3.1.2 Carga Incremental

Tiene la función de realizar los mantenimientos periódicos que mueven pequeños volúmenes de datos, y su frecuencia está dada por el gránulo y los requerimientos de los usuarios. La misma se llevó a cabo

²² Trabajo: coordinación y ejecución de transformaciones.

Capítulo 3: Integración de datos

mediante la programación del trabajo siguiente. Consultar el Anexo 7 para ver la transformación de los metadatos.

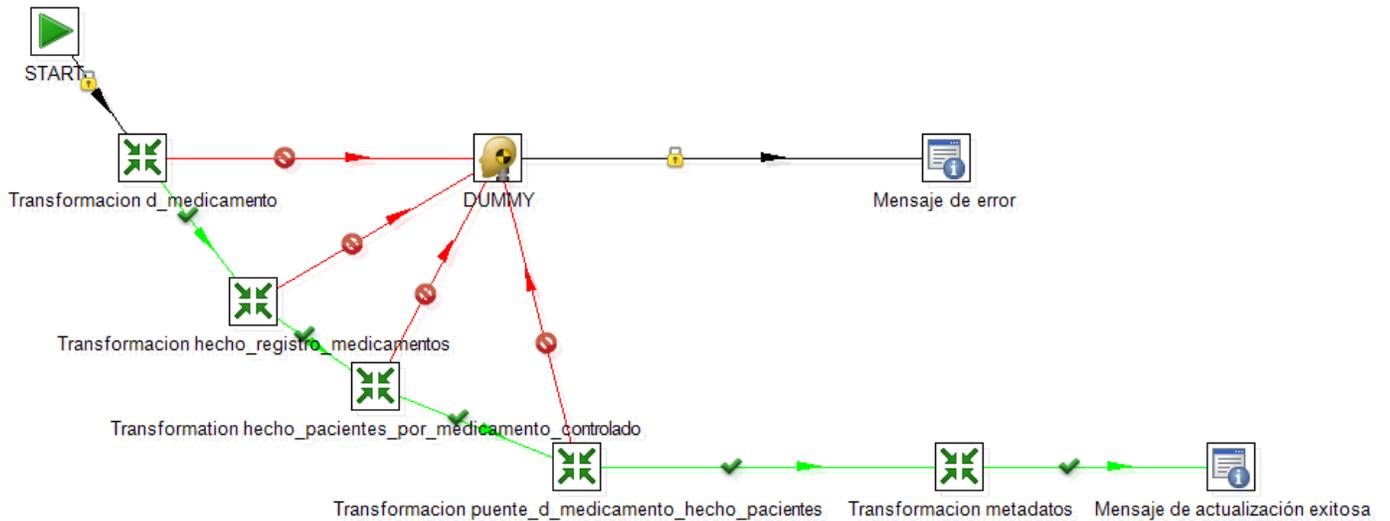


Figura 22: Ejecución del trabajo (Carga incremental)

En este capítulo se expusieron los pasos realizados para desarrollar el proceso de extracción, transformación y carga de los datos en el mercado, obteniéndose el diseño de la carga inicial e incremental. Se definieron además cada uno de las transformaciones de las dimensiones y hechos, logrando la integración de datos.

Capítulo 4: Proceso Analítico en Línea

En este capítulo se procede a diseñar el cubo multidimensional haciendo uso de la herramienta Schema Workbench, dentro de esta se definen las dimensiones con sus jerarquías correspondientes. También se elaboran las medidas para dar respuesta a las necesidades del cliente y se visualizan los resultados mediante tablas de datos y gráficas, utilizando el visor web JPivot desde el MCM.

4.1 Proceso de análisis de la información del mercado de datos

4.1.1 Configuración de la conexión en Schema Workbench

Una de las primeras cosas que se debe configurar al utilizar Schema Workbench, es la ubicación de los archivos .jar que se emplearán en las conexiones con las distintas bases de datos. Para el caso de PostgreSQL es "postgresql-9.1-901.jdbc4" (archivo ejecutable Jar), el cual se debe ubicar en la carpeta drivers.

Una vez que se ingresa en la herramienta, se establece la conexión al mercado de datos. Se introducen los campos, especificando el nombre de la base de datos, el usuario y la contraseña, como se muestra en la Figura 23.

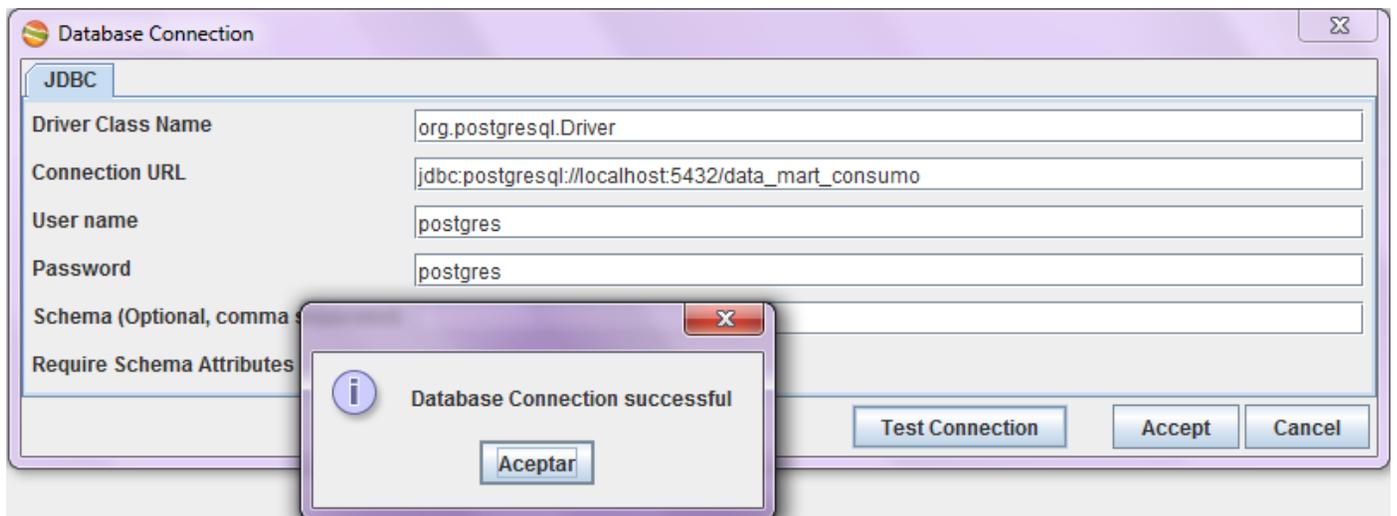


Figura 23: Configuración de la conexión en Schema Workbench

4.1.2 Determinar jerarquías de las tablas de dimensiones

Posteriormente se diseñan los cubos de datos, los cuáles se componen de las dimensiones, sobre las cuáles es necesario establecer las jerarquías. Determinar la jerarquía de cada una de las tablas de dimensiones resulta de gran importancia para organizar los niveles dentro de una dimensión, permitiendo un orden lógico de lo que se analiza, ya sea de un nivel general a uno específico o viceversa (29).

En la Figura 24 se muestra la jerarquía de la dimensión Tiempo. Para visualizar las jerarquías de las otras dimensiones, ver Anexo 8.

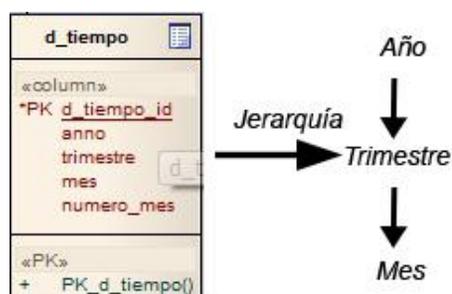


Figura 24: Jerarquía de la dimensión Tiempo

4.1.3 Medidas

Una vez establecidas las jerarquías de cada una de las dimensiones se elaboran las medidas en busca de las respuestas de lo que desean conocer los especialistas del MINSAP. Para ello se utilizan las consultas MDX que se ejecutan sobre el mercado de datos para obtener los indicadores o medidas que contribuirán a la toma de decisiones, como: total de registros, cantidad y porcentaje de pacientes que consumen medicamentos controlados, así como la cantidad de medicamentos consumidos en un período de tiempo y el total de consumo de un medicamento determinado.

A continuación se muestran dos tipos de medidas (ver Figura 25 y 26), para ver los restantes indicadores consultar el Anexo 9. En la primera se muestra la cantidad de medicamentos consumidos, la cual se obtiene aplicando la función que cuenta los registros distintos (distinct-count) sobre la columna `d_medimento_id` de la tabla "hecho_registro_medicamentos". 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 entre el total de registros).

Capítulo 4: Proceso Analítico en Línea

The screenshot shows the 'Schema - EsquemaConsumo_de_Medicamentos (esquema_consumo6.xml)' window. On the left, a tree view shows the 'CuboRegistro_Medicamentos' cube with dimensions: 'd_provincia', 'd_produccion', 'd_medicamento', 'd_tiempo', 'd_fabricante', and measures: 'Total_de_consumo' and 'Cantidad_medicamentos_consumidos'. The right pane shows the 'Measure' properties for 'Cantidad_medicamentos_consumidos'.

Attribute	Value
name	Cantidad_medicamentos_consumidos
description	
aggregator	distinct-count
column	d_medicamento_id
formatString	
visible	<input checked="" type="checkbox"/>
datatype	Integer
formatter	
caption	Cantidad de medicamentos consumidos

Figura 25: Medida cantidad de medicamentos consumidos

The screenshot shows the 'Schema - EsquemaConsumo_de_Medicamentos (esquema_consumo6.xml)*' window. On the left, a tree view shows the 'CuboPacientes_por_Medicamento_Controlado' cube with dimensions: 'd_provincia', 'd_raza', 'd_sexo', 'd_registro', 'd_vencimiento', 'd_grupo_etario', 'd_medicamento_controlado', 'd_fabricante', and measures: 'Cantidad_de_pacientes', 'Total_de_registros', and 'Porciento_de_pacientes'. The right pane shows the 'Calculated Member for 'CuboPacientes_por_Mec' properties for 'Porciento_de_pacientes'.

Attribute	Value
name	Porciento_de_pacientes
description	
caption	PorCiento(%)
dimension	Measures
visible	<input checked="" type="checkbox"/>
formula formulaElement...	$(([Measures].[Cantidad_de_pacientes]*100)/[Measures].[Total_de_registros])$
formatString	#.00

Figura 26: Medida calculable porciento de pacientes

4.1.4 Consultas multidimensionales

Las expresiones multidimensionales se utilizan en la inteligencia de negocios para generar reportes para la toma de decisiones basados en datos históricos. Las MDX permiten consultar objetos multidimensionales (como los cubos “Registro_Medicamentos” y “Pacientes_por_Medicamento_Controlado”) y devolver un conjunto de celdas, que es el resultado de tomar un subconjunto de las celdas del cubo original. Dichas consultas calcularán los datos numéricos y estos serán añadidos en los resultados a visualizar, utilizando el motor OLAP.

Una vez diseñado las dimensiones y cubos, se realizan consultas MDX a través de la herramienta Mondrian. En la siguiente consulta se seleccionan las medidas: total de consumo de un medicamento, cantidad de medicamentos consumidos (en las columnas) y las dimensiones: provincia, producción, medicamento y tiempo (en las filas), obteniéndose la información a través del cubo de datos “Registro_Medicamentos”.

```
select { [ Measures ] . [ Total_de_consumo ] ,  
        [ Measures ] . [ Cantidad_medicamentos_consumidos ] } on columns,  
      { ( [ d_provincia ] , [ d_produccion ] , [ d_medicamento ] , [ d_fabricante ] ,  
        [ d_tiempo ] ) } on rows  
from CuboRegistro_Medicamentos
```

Figura 27: Consulta MDX para el cubo “Registro_Medicamentos”

En la consulta que se muestra a continuación se seleccionan las medidas: cantidad de pacientes y porcentaje de pacientes (en las columnas) y las dimensiones d_provincia, d_raza, dsexo, d_grupo_etario, d_medicamento_controlado (en las filas), obteniéndose la información a través del cubo “Pacientes_por_Medicamento_Controlado”.

```
select { [ Measures ] . [ Cantidad_de_pacientes ] ,  
        [ Measures ] . [ Total_de_registros ] ,  
        [ Measures ] . [ Porciento_de_pacientes ] } on columns,  
        { ( [ d_provincia ] , [ d_raza ] , [ dsexo ] , [ d_grupo_etario ] ,  
          [ d_med medicamento_controlado ] , [ d_fabricante ] ) } on rows  
from CuboPacientes_por_Medicamento_Controlado
```

Figura 28: Consulta MDX para el cubo “Pacientes_por_Medicamento_Controlado”

4.1.5 Resultados del cliente JPivot

Una vez diseñados los cubos “Registro_Medicamentos” y “Pacientes_por_Medicamento_Controlado”, son almacenados en un archivo XML que se cargará en el Mondrian y será utilizado para realizar el Proceso Analítico en Línea. Posteriormente la información será visualizada por el cliente JPivot, a través de la web.

JPivot permite construir tablas generadas de forma dinámica, consiguiendo un amplio abanico de posibilidades de análisis. Este tipo de tablas es de gran utilidad, ya que permiten mostrar los resultados de las consultas filtrando por diversos campos, de manera que se puedan adicionar y eliminar distintos criterios de búsqueda. Además los resultados pueden ser mostrados mediante gráficas y exportados a formato pdf o Excel.

En la Figura 29 se muestra la cantidad de medicamentos consumidos en la provincia de Artemisa, filtrando estos por tipo de producción.

Provincias		Producción	Medidas
- Todos	+ Todos		Cantidad de medicamentos consumidos
Artemisa	- Todos		10
		Internacional	7
		Nacional	3

Figura 29: Información filtrada por el cubo “Registro_Medicamentos”

Capítulo 4: Proceso Analítico en Línea

En la Figura 30 se muestra la cantidad de pacientes, total de registros y porcentaje de pacientes que consumen medicamentos controlados, filtrando dichas cantidades por raza y sexo.

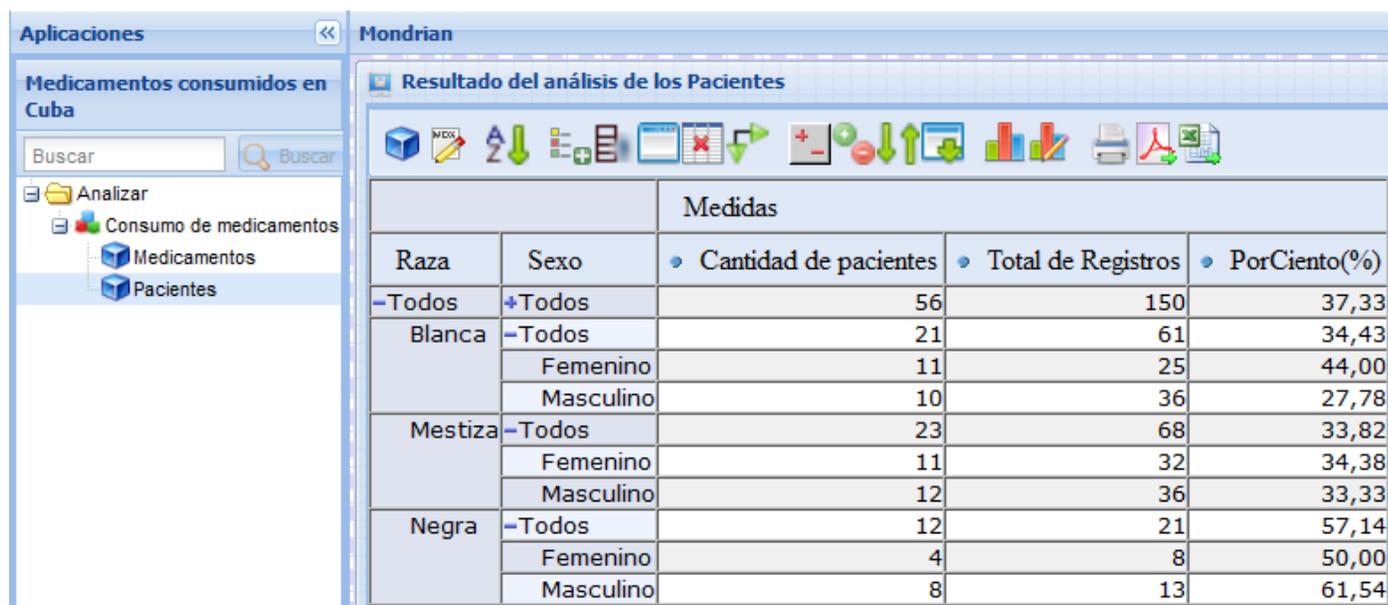


Figura 30: Información filtrada por el cubo “Pacientes_por_Medicamento_Controlado”

Una vez visualizadas las tablas de datos, JPivot permite mostrar estas a través de gráficas para que el usuario tenga la posibilidad de observarlas desde otras perspectivas, para un mejor entendimiento de ellas, apoyando el proceso de toma de decisiones. La información del reporte mostrado en la Figura 30 se puede observar en forma de gráficas de barras y de pastel en las Figuras 31 y 32 respectivamente.

Capítulo 4: Proceso Analítico en Línea

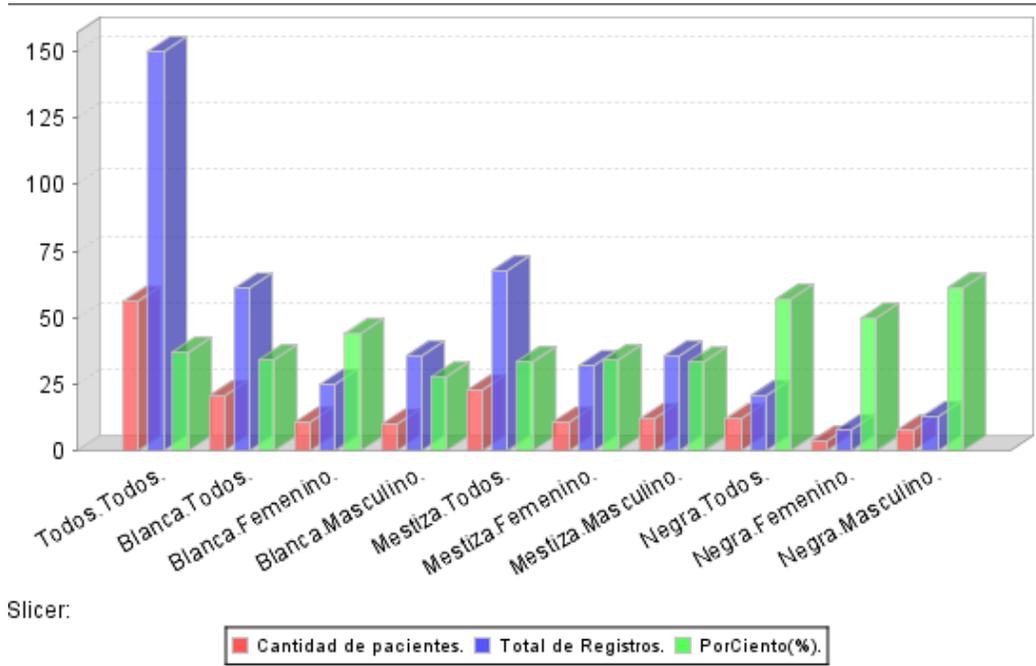


Figura 31: Información en gráfica de barras

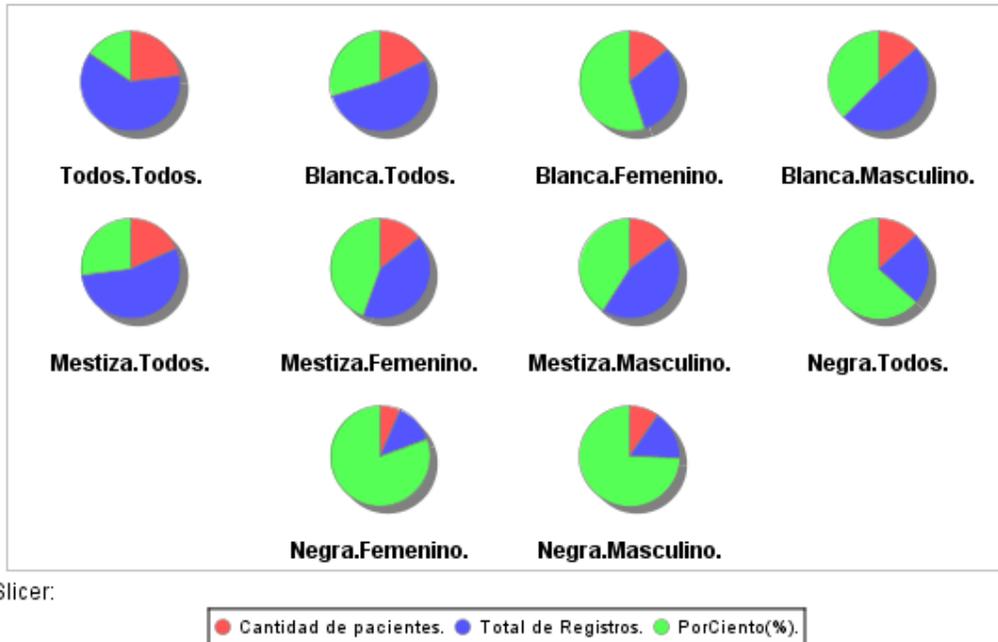


Figura 32: Información en gráfica de pastel

4.1.6 Integración del mercado de datos con el Módulo Consumo de Medicamentos

Para realizar la integración del mercado de datos con el MCM del sistema Synta, se realizaron los siguientes pasos:

1. Se analizaron las características que presenta el MCM, para poder visualizar la información que brinda el cliente JPivot desde dicho módulo.
2. Para configurar el cliente JPivot, se modificaron los archivos `datasources.xml` y `mondrian.properties`. En estos archivos se especificó la dirección de la base de datos (mercado de datos), el usuario, la contraseña y el driver (para PostgreSQL) de conexión, además del catálogo que hace referencia a los cubos "Registro_Medicamentos" y "Pacientes_por_Medicamento_Controlado".
3. En el MCM se creó una plantilla siguiendo las políticas del sistema Synta, que permitió la visualización de la información a través del cliente JPivot. Debido a que el servidor web de Synta es Apache 2.2 y JPivot utiliza el Apache Tomcat 7.0, fue necesario utilizar una etiqueta HTML (`iframe`), para mostrar el visor dentro del módulo. A esta etiqueta se le especificó la dirección URL del cliente web, el alto y el ancho en que debía ser mostrado.

```
'<iframe src="localhost:8080/mondrian?query=default" style="width:100%;height:100%;border:none;">
</iframe>'
```

4. Luego se habilitó un vínculo en la pestaña Reportes del MCM, para acceder a las funcionalidades que brinda el mercado de datos (desde el visor web JPivot), a través de una plantilla creada en el área de reportes del módulo, que contiene el vínculo de acceso al reporte que se hace referencia.
5. Se realizaron algunas modificaciones en el diseño de la herramienta JPivot, para proporcionarle una apariencia similar a las interfaces del MCM del Synta. Como parte de este paso, se modificaron también los nombres de las funcionalidades, para una mayor comprensión por parte de los usuarios.

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

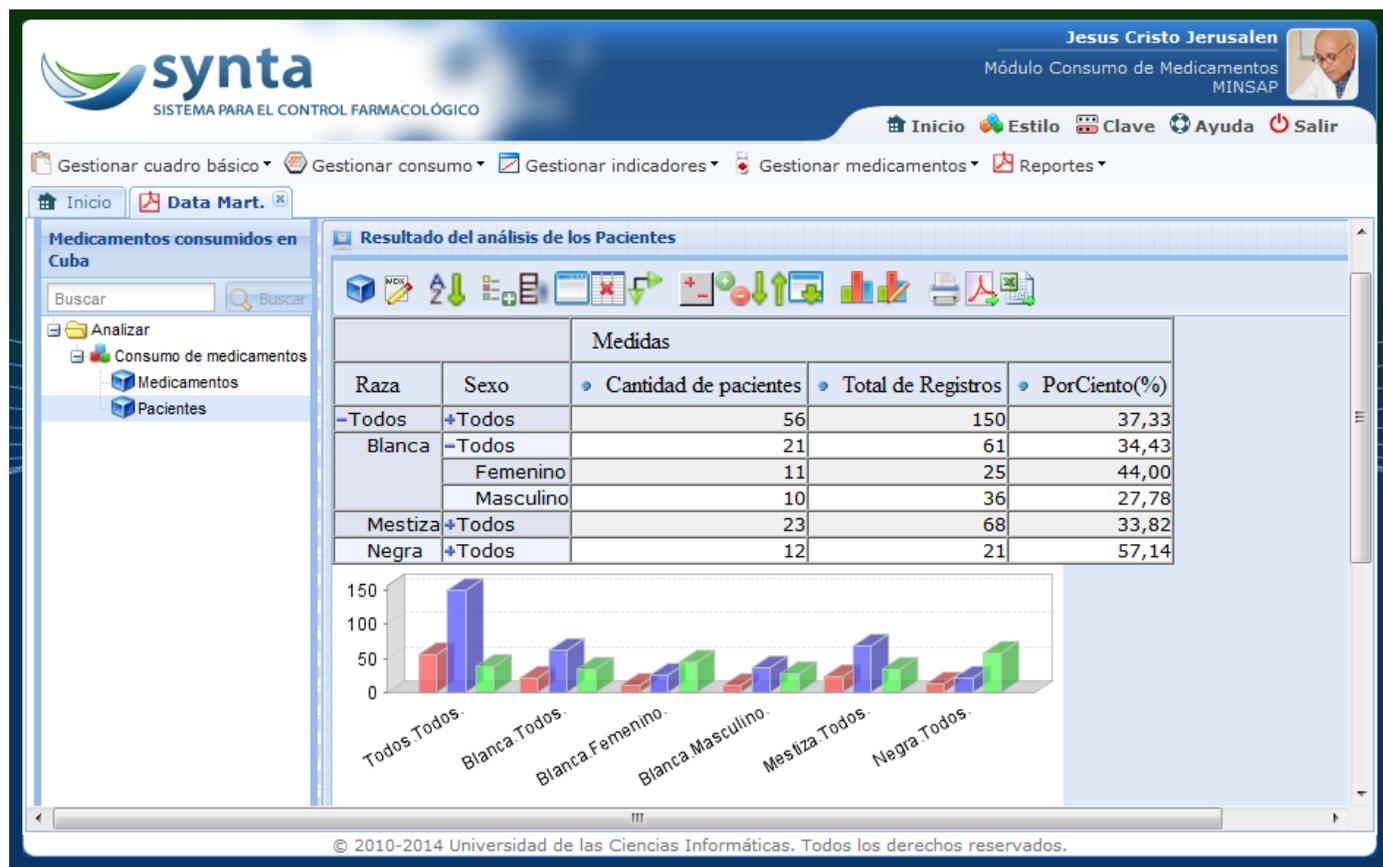


Figura 33: Reporte utilizando el visor JPivot desde el MCM

En el desarrollo de este capítulo se utilizó la herramienta Schema Workbench para la diseño de los cubos de datos, que consiste en la definición de las dimensiones, la jerarquización de sus atributos e implementación de los indicadores. Con el proceso analítico en línea se apoyará la toma de decisiones referente al consumo de medicamentos, pues los usuarios podrán ver los informes en forma de tablas y gráficos desde diferentes perspectivas, satisfaciendo así las necesidades de información de los especialistas del MINSAP.

Capítulo 5: Pruebas

En este capítulo se realizan varias pruebas de rendimiento al mercado de datos, con el objetivo evaluar su desempeño. Para esto se utiliza la herramienta Jakarta-jmeter y las cantidades de 25, 50 y 100 usuarios accediendo simultáneamente.

5.1 Pruebas al mercado de datos

Las pruebas de software son los procesos que permiten verificar y revelar la calidad de un producto. Son utilizadas para identificar posibles fallos de implementación, calidad y usabilidad en la solución.

Para determinar el nivel de calidad de un software se deben efectuar pruebas que permitan comprobar el grado de cumplimiento de las especificaciones del sistema. Para la validación del mercado de datos se aplicaron varias pruebas de rendimiento a la base de datos.

5.1.1 Pruebas de rendimiento

Estas pruebas se realizan para determinar lo rápido que realiza una tarea un sistema en condiciones particulares de trabajo. También puede servir para validar y verificar otros atributos de la calidad del sistema, tales como la escalabilidad, fiabilidad y uso de los recursos.

Existen aplicaciones que se encargan de simular situaciones de carga de usuarios sobre cualquier sistema. Dichas aplicaciones se encargan de realizar peticiones a un sistema y medir todos los factores que intervienen en la respuesta del mismo, para proporcionar medidas de tiempo, carga y rendimiento. Una de estas aplicaciones más famosas, de carácter gratuito y código abierto, es Jkarta-jmeter. Aplicación del proyecto Apache Jakarta, escrita en Java que permite realizar pruebas de rendimiento y funcionalidad sobre aplicaciones de tipo cliente/servidor escritas en cualquier lenguaje. Puede ser utilizada para realizar pruebas sobre conexiones de bases de datos JDBC, HTTP, FTP, entre otras.

Para la configuración y realización de las pruebas se utilizó la herramienta Jkarta-jmeter en su versión 2.4, en una computadora con un microprocesador Intel (R) Core 2 Dúo, memoria RAM de 2 GB y 120 GB de disco duro.

Plan de prueba

Se configuró un plan de pruebas (ver Figura 34) para comprobar el rendimiento de la base de datos, configurándose las peticiones JDBC para un acceso de 25, 50 y 100 usuarios conectados concurrentemente al depósito de datos.

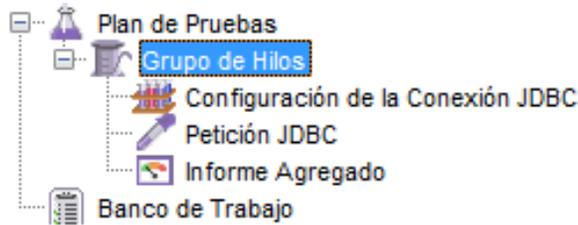


Figura 34: Configuración del Plan de Pruebas de rendimiento a la base de datos

Resultados de la prueba

Una vez realizadas las pruebas de rendimiento a la base de datos, se obtuvieron los tiempos de respuesta 1.1/s, 3.4/s y 4.0/s para 25, 50 y 100 accesos de usuarios respectivamente. En las figuras 35, 36 y 37 se muestran los resultados arrojados.

Label	# Muestras	Media	Mediana	Linea de 90%	Min	Máx	% Error	Rendimiento	Kb/sec
Petición JDBC	10	7	3	44	2	44	0,00%	1,1/seg	,1
TOTAL	10	7	3	44	2	44	0,00%	1,1/seg	,1

Figura 35: Resultados del Plan de Prueba para 25 usuarios

Label	# Muestras	Media	Mediana	Linea de 90%	Min	Máx	% Error	Rendimiento	Kb/sec
Petición JDBC	20	4	2	6	1	47	0,00%	3,4/seg	,4
TOTAL	20	4	2	6	1	47	0,00%	3,4/seg	,4

Figura 36: Resultado del Plan de Prueba para 50 usuarios

Capítulo 5: Pruebas

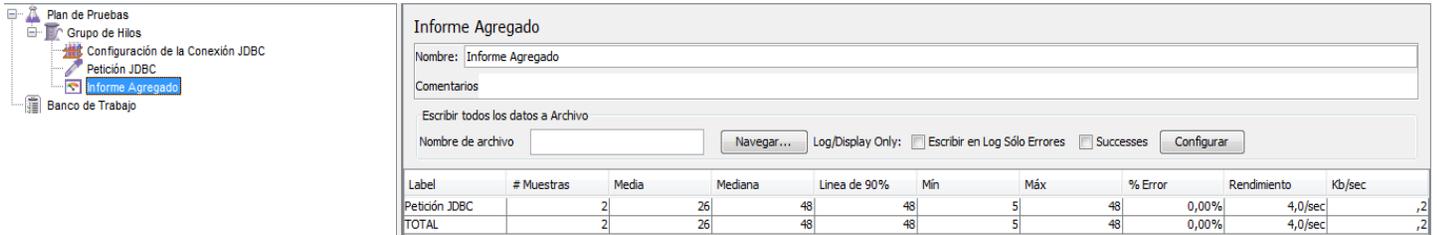


Figura 37: Resultado del Plan de Prueba para 100 usuarios

Luego de concluidas las pruebas se puede observar que el tiempo de respuesta es aceptable debido a la cantidad de usuarios conectados concurrentemente, así como, a las características que posee la computadora donde se realizó la prueba. Las pruebas realizadas responden a las necesidades requeridas.

Conclusiones

Después de haber desarrollado el mercado de datos para el Módulo Consumo de Medicamentos del Sistema para el Control Farmacológico, se llegan a las siguientes conclusiones:

- La realización de un análisis sobre los almacenes y mercados de datos como sistemas que apoyan el proceso de toma de decisiones en el área de la salud, tanto a nivel internacional como nacional, permitió identificar las tendencias actuales que presentan estos en cuanto a diseño y construcción. Como resultado no se encontró un sistema que cumpliera con las características de la investigación.
- La elaboración de un conjunto de preguntas a los especialistas, permitió identificar los principales requerimientos de información. Esto posibilitó la obtención de los indicadores y perspectivas, que constituyó el punto de partida para el desarrollo del sistema.
- La implementación del sistema utilizando la metodología Hefesto, permitió una construcción sencilla, ordenada e intuitiva, siendo esta metodología la más adecuada para personas que entran por primera vez en el mundo del diseño de los mercados de datos.
- Se realizó la integración del mercado de datos al Módulo Consumo de Medicamentos, permitiendo a los especialistas tomar decisiones a partir de la información obtenida en un único sistema, basadas en reportes, tablas de datos y gráficas, además de exportar los informes a formato Excel y PDF.
- La visualización de la información del mercado de datos con una apariencia similar al Módulo Consumo de Medicamentos, demostró que con JPivot se pudo lograr un diseño homogéneo con respecto a las interfaces del Sistema para el Control Farmacológico.
- La realización de pruebas de rendimiento demostró el correcto funcionamiento del mercado de datos al acceder simultáneamente 25, 50 y 100 usuarios.

Recomendaciones

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

- Dar continuidad a la investigación, con el fin de obtener las decisiones precisas referentes al consumo de medicamentos del país.

Referencias Bibliográficas

1. Transforminf Knowledge Into Action. *Artículos BI en la Práctica*. [En línea] 2 de junio de 2013. [Citado el: 1 de Noviembre de 2013.] <http://www.siskle.com/spanish/articulo01.html>.
2. *Business Intelligence y Data Warehousing en Windows*. s.l. : DanySoft, 2005. pág. 325.
3. **Curto, Josep**. DW: definiciones de Inmon y Kimball. *DW: definiciones de Inmon y Kimball*. [En línea] 28 de Noviembre de 2006. [Citado el: 4 de Noviembre de 2013.] [http://informationmanagement.wordpress.com/2006/11/28/dw-definiciones-de-inmon-y-kimball/..](http://informationmanagement.wordpress.com/2006/11/28/dw-definiciones-de-inmon-y-kimball/)
4. **Brito Sarasas, Ing. Raycos**. *Minería de Datos aplicada a la Gestión Docente del Instituto Superior Politécnico Jose Antonio Echeverría*. CUJAE. 2008. pág. 100, Tesis de Maestría.
5. **Mojena Alpizar, Ing. José, Flores Moreno, Ing. Dannier y Arencia Morales, Ing. Annia**. *Sistema para el Control de Medicamentos*. UCI. 2010. pág. 10.
6. **Mazariegos Rabanales, Jorge Armín**. *Desarrollo de un Data Mart de Información Académica de Estudiantes de la Escuela de Ciencia y Sistemas de la Facultad de Ingeniería de la USAC*. Guatemala : s.n., Agosto, 2007. pág. 85, Tesis de diploma.
7. **Bernabeu Ricardo, Ing. Dario**. *Data Warehousing: Investigación y Sistematización de Conceptos – Hefesto: Metodología propia para la Construcción de un Data Warehouse*. Córdoba, Argentina : s.n., 2010. pág. 148. Vol. 2.
8. —. *Data Warehousing: Investigación y Sistematización de Conceptos - Hefesto: Metodología propia para la Construcción de un Data Warehouse*. Córdoba, Argentina : s.n., 21 de Abril de 2009. pág. 135. Vol. 1.1.
9. **Lazaro, Rigoberto**. Data Warehouse. *Data Warehouse*. [En línea] 18 de septiembre de 2012. [Citado el: 02 de enero de 2014.] http://rigobertlc.blogspot.com/2012_09_01_archive.html..
10. Grupo de Autores Sinnexus. *Sinnexus*. [En línea] 5 de febrero de 2012. [Citado el: 4 de enero de 2014.] http://www.sinnexus.com/business_intelligence/datamart.aspx/.
11. **Vázquez, Rodolfo**. 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: 4 de enero de 2013.] http://www.femisaluddigital.org.uy/admin/files/femi/from_old_website/LibroSIS_SIGs_Data_Warehouses.doc.
12. **Villanueva Ojeda, Álvaro**. *Almacén de datos de Soporte de Decisiones para un Hospital del Sistema de Salud Público*. Universidad Católica . Lima, Perú : s.n., 2008. pág. 82, Tesis de diploma.
13. **Ramón Cueto, Ariagna y Díaz García, Joannis**. *Implementación de un datawarehouse para el control de los Recursos Humanos de Salud*. UCI. 2009. pág. 80, Tesis de diploma.
14. **Bello Sosa, Aliandy y Lóriga Salas, Dennier**. *Data Mart para la toma de decisiones referente a las Reacciones Adversas a Medicamentos en el Ministerio de Salud Pública desde el producto Synta*. UCI. La Habana : s.n., 2013. pág. 80, Trabajo de diploma.

Referencias bibliográficas

15. **Ilbay Yupa, Elsi Elizabet.** *Propuesta metodológica para aplicar Business Intelligence Caso Práctico "COHERVI.SA"*. Escuela Ingeniería en Sistemas. Riobamba, Ecuador : s.n., Octubre, 2009. pág. 311, Tesis de diploma.
16. **Inmon, Bill.** About Bill Inmon, The Father of Data Warehousing. Biography. [En línea] [Citado el: 8 de diciembre de 2012.] <http://www.inmoncif.com/about/>.
17. **Tamayo, Marysol y Moreno, Francisco Javier.** Análisis del modelo de almacenamiento MOLAP frente al modelo de almacenamiento ROLAP. [En línea] 12 de Diciembre de 2008. [Citado el: 6 de Enero de 2014.] http://www.scielo.org.co/scielo.php?pid=S0120-56092006000300016&script=sci_arttext&lng=pt..
18. Sparx. Enterprise Architect. *Enterprise Architect*. [En línea] Sparx, 2008. [Citado el: 14 de enero de 2014.] <http://www.sparxsystems.es/New/products/ea.html..>
19. **Sánchez Asenjo, Jorge.** Sistema Gestores de Base de Datos. *Sistema Gestores de Base de Datos*. [En línea] 2009. [Citado el: 15 de enero de 2014.] <http://ubuntuone.com/p/sqt/>.
20. Pgdming Official Site. [En línea] [Citado el: 17 de enero de 2014.] <http://www.arpug.com.ar/trac/wiki/PgAdmin..>
21. Herramientas ETL. [En línea] [Citado el: 18 de enero de 2014.] <http://www.carlosproal.com/dw/dw05.html..>
22. SOAction. [En línea] [Citado el: 18 de enero de 2014.] <http://soaction.sisorg.com.mx/solucion2.html>.
23. Microsoft Office. [En línea] [Citado el: 19 de enero de 2014.] <http://office.microsoft.com/es-hn/support/las-diez-ventajas-principalesde-microsoft-business-intelligence-HA010165509.aspx>.
24. La plataforma Pentaho Open Source Business Intelligence. . [En línea] [Citado el: 20 de enero de 2014.] <http://pentaho.almacen-datos.com/>.
25. **Caralt, Jordi Conesa.** *Introducción a la Business Intelligence. Primera edición en lengua castellana*. Barcelona : s.n., 2010.
26. Pentaho Architecture Layers of a Mondrian System. [En línea] [Citado el: 21 de enero de 2014.] <http://mondrian.pentaho.org/documentation/architecture.php>.
27. **Autores, C. d.** *JBoss Enterprise Middleware*. 2007.
28. Apache Tomcat. [En línea] [Citado el: 22 de enero de 2014.] <http://www.agapea.com/libros/Tomcat-6-0-La-guia-definitiva-isbn-8441524319-i.htm..>
29. **Juárez, Zoraya Catalina Flores.** OLAP (OnLine Analytical Processing). <http://basesdatoscms.files.wordpress.com/2012/09/olap-resumen1.pdf>. [En línea] septiembre de 2012. [Citado el: 20 de 4 de 2014.]

Bibliografía

1. Apache Tomcat. [Online] [Cited: enero 20, 2014.] <http://tomcat.apache.org/>.
2. **Autores, C. d.** *JBoss Enterprise Middleware*. 2007
3. Apache Tomcat. [Online] [Cited: enero 22, 2014.] <http://www.agapea.com/libros/Tomcat-6-0-La-guia-definitiva-isbn-8441524319-i.htm..>
4. *Business Intelligence y Data Warehousing en Windows*. s.l. : DanySoft, 2005. p. 325.
5. **Brito Sarasas, Ing. Raycos.** *Minería de Datos aplicada a la Gestión Docente del Instituto Superior Politécnico Jose Antonio Echeverría*. CUJAE. 2008. p. 100, Tesis de Maestría.
6. **Bernabeu Ricardo, Ing. Dario.** *Data Warehousing: Investigación y Sistematización de Conceptos – Hefesto: Metodología propia para la Construcción de un Data Warehouse*. Córdoba, Argentina : s.n., 2010. p. 148. Vol. 2.
7. **Bello Sosa, Aliandy and Lóriga Salas, Dennier.** *Data Mart para la toma de decisiones referente a las Reacciones Adversas a Medicamentos en el Ministerio de Salud Pública desde el sistema Synta*. UCI. La Habana : s.n., 2013. p. 80, Trabajo de diploma.
8. **Bartomeus Rojas, Pol.** Portal d'accés obert al coneixement de la UPC. Portal d'accés obert al coneixement de la UPC. [Online] septiembre 2009. [Cited: enero 7, 2014.] <http://upcommons.upc.edu/pfc/bitstream/2099.1/7461/4/59744.pdf.txt..>
9. **Curto, Josep.** DW: definiciones de Inmon y Kimball. *DW: definiciones de Inmon y Kimball*. [Online] Noviembre 28, 2006. [Cited: Noviembre 4, 2013.] [http://informationmanagement.wordpress.com/2006/11/28/dw-definiciones-de-inmon-y-kimball/..](http://informationmanagement.wordpress.com/2006/11/28/dw-definiciones-de-inmon-y-kimball/)
10. **Caralt, Jordi Conesa.** *Introducción a la Business Intelligence. Primera edición en lengua castellana*. Barcelona : s.n., 2010.
11. —. *Data Warehousing: Investigación y Sistematización de Conceptos - Hefesto: Metodología propia para la Construcción de un Data Warehouse*. Córdoba, Argentina : s.n., 21 de Abril de 2009. p. 135. Vol. 1.1.
12. **García Cabrera, Anabel and Garay Polanco, Leodanny Wuilber.** *Alamcén de datos para el ciclo logístico de los medios*. La Habana : s.n., 2013. Tesis de diploma
13. Grupo de Autores Sinnexus. *Sinnexus*. [Online] febrero 5, 2012. [Cited: enero 4, 2014.] http://www.sinnexus.com/business_intelligence/datamart.aspx/.
14. Herramientas ETL. [Online] [Cited: enero 18, 2014.] <http://www.carlosproal.com/dw/dw05.html..>
15. **Hallo, Maria.** *Análisis, diseño, e implementación de un Data Mart para el área de Sismología del Departamento de Geofísica de la Escuela Politécnica Nacional*. Quito : s.n., marzo 2006. p. 90, Tesis de diploma.
16. **H. Inmon, William.** *Building the data warehouse*. 3. John Wiley and Sons : s.n., 2002. p. 234.

Bibliografía

17. **Hernández, Martha Denia.** Procedimiento para el desarrollo de un sistema de inteligencia de negocios en la gestión de ensayos clínicos en el Centro de Inmunología Molecular. [Online] 22, Febrero 4, 2011. [Cited: Enero 10, 2014.] <http://www.acimed.sld.cu/index.php/acimed/rt/printerFriendly/208/168>. ISSN:2307-2113.
18. **Ibay Yupa, Elsi Elizabet.** *Propuesta metodológica para aplicar Business Intelligence Caso Práctico "COHERVI.SA"*. Escuela Ingeniería en Sistemas. Riobamba, Ecuador : s.n., Octubre, 2009. p. 311, Tesis de diploma.
19. **Inmon, Bill.** About Bill Inmon, The Father of Data Warehousing. Biography. [Online] [Cited: diciembre 8, 2012.] <http://www.inmoncif.com/about/>.
20. **Juárez, Zoraya Catalina Flores.** OLAP (OnLine Analytical Processing). <http://basesdatoscms.files.wordpress.com/2012/09/olap-resumen1.pdf>. [Online] septiembre 2012. [Cited: 4 20, 2014.]
21. [jpivot.sourceforge.net](http://jpivot.sourceforge.net/tags/jpivot-tags-en.html). jpivot.sourceforge.net. [Online] [Cited: abril 23, 2014.]
22. Kettle. [Online] [Cited: enero 20, 2014.] <http://www.gravitar.biz/index.php/bi/introduccionpentaho-parte-1/>.
23. **Kimball, Ralph.** *The Data Warehouse ETL Toolkit. Practical Techniques for extracting, cleaning, conforming and delivering data.* s.l. : WILEY PUBLISHING, INC, 2004.
24. **Kimball, Ralph.** *The Data Warehouse Lifecycle Toolkit.* Nueva York: Wiley : s.n., 1998. p. 120.
25. **Lazaro, Rigoberto.** Data Warehouse. *Data Warehouse.* [Online] septiembre 18, 2012. [Cited: enero 02, 2014.] http://rigobertic.blogspot.com/2012_09_01_archive.html.
26. La plataforma Pentaho Open Source Business Intelligence. . [Online] [Cited: enero 20, 2014.] <http://pentaho.almacen-datos.com/>.
27. Microsoft Office. [Online] [Cited: enero 19, 2014.] <http://office.microsoft.com/es-hn/support/las-diez-ventajas-principalesde-microsoft-business-intelligence-HA010165509.aspx>.
28. msdn.microsoft. [Online] microsoft. [Cited: abril 21, 2014.] <http://msdn.microsoft.com/es-es/library/ms145514.aspx>.
29. **Martínez Arol, Alfredo.** OLAP y el diseño de cubos. . [Online] <http://gestiopolis.com/canales8/ger/olap-online-analytic-processing.html>.
30. **Mata Gomez, Greybi.** *Almacén de datos para el Sistema Informativo Automatizado de los Cuadros.* UCI. 2012. p. 80, Tesis de diploma.
31. **Mestre Morales, Yanetsy.** *Propuesta de un Data Mart para el análisis de la deuda externa en el Banco Nacional de Cuba.* La Habana : s.n., Junio 2010. p. 80, Trabajo de diploma.
32. **Mojena Alpizar, Ing. José, et al., et al.** *Sistema para el Control de Medicamentos.* UCI. 2010. p. 10.

Bibliografía

33. **Mazariegos Rabanales, Jorge Armín.** *Desarrollo de un Data Mart de Información Académica de Estudiantes de la Escuela de Ciencia y Sistemas de la Facultad de Ingeniería de la USAC.* Guatemala : s.n., Agosto, 2007. p. 85, Tesis de diploma.
34. **Ordaz, Blvd. Diaz.** Ordaz, Blvd. Diaz. . *Características de Pentaho.* [Online] 2013. [Cited: enero 20, 2014.] [http://www.gravitar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/..](http://www.gravitar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/)
35. Pentaho Mondrian Documentation. [Online] [Cited: enero 21, 2014.] <http://mondrian.pentaho.com/documentation/workbench.php>.
36. Pentaho Architecture Layers of a Mondrian System. [Online] [Cited: enero 21, 2014.] <http://mondrian.pentaho.org/documentation/architecture.php>.
37. Pentaho BI . [Online] [Cited: enero 21, 2014.] <http://es.scribd.com/doc/109789629/Pentaho-Bi..>
38. Pentaho BI Suite Enterprise Edition. [Online] [Cited: enero 20, 2014.] http://www.pentaho.com/products/data_services/?hp=y
39. **Peregrin Tamayo, Neyaris and Casaña Vinagera, Virgen.** *Proceso de migración de datos hacia un Data Warehouse para el módulo Análisis Químico del proyecto LIMS de Calidad.* UCI. 2009. Tesis de diploma.
40. PostgreSQL. [Online] [Cited: enero 16, 2014.] <http://www.todoprogramas.com/programa/postgresql..>
41. Pgadmin Official Site. [Online] [Cited: enero 17, 2014.] <http://www.arpug.com.ar/trac/wiki/PgAdmin..>
42. **Ramón Cueto, Ariagna and Díaz García, Joannis.** *Implementación de un datawarehouse para el control de los Recursos Humanos de Salud.* UCI. 2009. p. 80, Tesis de diploma.
43. **Sinnexus.** Bases de Datos OLTP y OLAP. *Sinergia e Inteligencia de Negocios.* [Online] 2007. http://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx
44. **Sanz, Miguel Rodríguez.** 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. [Online] Julio 22, 2010. [Cited: Enero 10, 2014.] <http://es.scribd.com/doc/52203545/16/Metodologia-propuesta-por-Bill-Inmon..>
45. Sparx. Enterprise Architect. *Enterprise Architect.* [Online] Sparx, 2008. [Cited: enero 14, 2014.] <http://www.sparxsystems.es/New/products/ea.html..>
46. **Sánchez Asenjo, Jorge.** Sistema Gestores de Base de Datos. *Sistema Gestores de Base de Datos.* [Online] 2009. [Cited: enero 15, 2014.] [http://ubuntuone.com/p/sqt/.](http://ubuntuone.com/p/sqt/)
47. SOAction. [Online] [Cited: enero 18, 2014.] <http://soaction.sisorg.com.mx/solucion2.html>.
48. **Sánchez, José A Alférez.** INSTALACIÓN, CONFIGURACIÓN Y ADMINISTRACIÓN DEL SERVIDOR DE APLICACIONES JBOSS. [Online] enero 2011. [Cited: enero 22, 2014.] <http://www.alferez.es/documentos/Jboss.pdf..>
49. Transforminf Knowledge Into Action. *Artículos BI en la Práctica.* [Online] junio 2, 2013. [Cited: Noviembre 1, 2013.] <http://www.siskle.com/spanish/articulo01.html>.

Bibliografía

50. **Torres Galvez, Ariel.** *Desarrollo de un Data Mart para la obtención de las razones financieras de los subsistemas de cobros y pagos, caja y banco del proyecto ERP-Cuba.* La Habana : s.n., Junio, 2011. p. 80, Tesis de diploma.
51. **TANDRÓN, IVÁN MAYKEL.** *Técnicas y herramientas de extracción, transformación y carga de datos aplicadas a la seguridad ciudadana.* Las Villas : s.n., 2008.
52. **Tamayo, Marysol y Moreno, Francisco Javier.** Análisis del modelo de almacenamiento MOLAP frente al modelo de almacenamiento ROLAP. [Online] Diciembre 12, 2008. [Cited: Enero 6, 2014.] http://www.scielo.org.co/scielo.php?pid=S0120-56092006000300016&script=sci_arttext&lng=pt..
53. **Vázquez, Rodolfo.** Sistemas de Información para la Gestión de Instituciones de Salud. *Teoría y aplicaciones en el Uruguay.* [Online] octubre 21, 2003. [Cited: enero 4, 2013.] http://www.femisaluddigital.org.uy/admin/files/femi/from_old_website/LibroSIS_SIGs_Data_Warehouses.doc
54. **Villanueva Ojeda, Álvaro.** *Almacén de datos de Soporte de Decisiones para un Hospital del Sistema de Salud Público.* Universidad Católica . Lima, Perú : s.n., 2008. p. 82, Tesis de diploma.
55. **Vega Torres, Ing. Lilian, Rojas Díaz, Ing. Luis y Placeres Villar, Lic. Cecilia.** 2008. La inteligencia de negocios. [Online] 2008 www.redciencia.info.ve/memorias/ProyProsp/trabajos/l3.doc
56. **Velasco, Roberto Hernando.** [Online] 2010 <http://www.rhernando.net/modules/tutorial/doc/bd/dw.html>.
57. **Velasco, Roberto Hernando.** [Online] 2010 <http://www.rhernando.net/modules/tutorials/doc/bd/oracle.html>

Anexos

Anexo 1. Entrevista realizada a la directora del Dpto. de Farmacoepidemiología del MINSAP: Dr. Cristina Lara Bastanzuri

- ¿Qué indicadores desea medir del consumo de medicamentos?
- ¿Qué características del consumo de medicamentos, usted desea analizar?
- ¿Una vez que usted analiza la información del consumo de medicamentos, cómo le gustaría visualizarla? ¿Gráficas de pastel, de barra, etc.?
- ¿Le gustaría que la información analizada fuera posible guardarla en formato .pdf o Excel?

Anexo 2. Perspectivas

Campos que contendrá la perspectiva “Sexo”

Nombre: Perspectiva “Sexo”		
Atributos	Tipo	Descripción
id_sexo	Integer	Identificador del sexo.
nombre_sexo	Character Varing	Nombre del sexo.

Campos que contendrá la perspectiva “Grupo_Etario”

Nombre: Perspectiva “Grupo_Etario”		
Atributos	Tipo	Descripción
id_grupo_etario	Integer	Identificador del grupo etario.
nombre_grupo_etario	Character Varing	Nombre de las clasificaciones de las edades de las personas.

Campos que contendrá la perspectiva “Tiempo”

Nombre: Perspectiva “Tiempo”		
Atributos	Tipo	Descripción
anno	Character Varing	Número del año.
trimestre	Character Varing	Nombre del trimestre.
mes	Character Varing	Nombre del mes.
código	Integer	Número del mes.

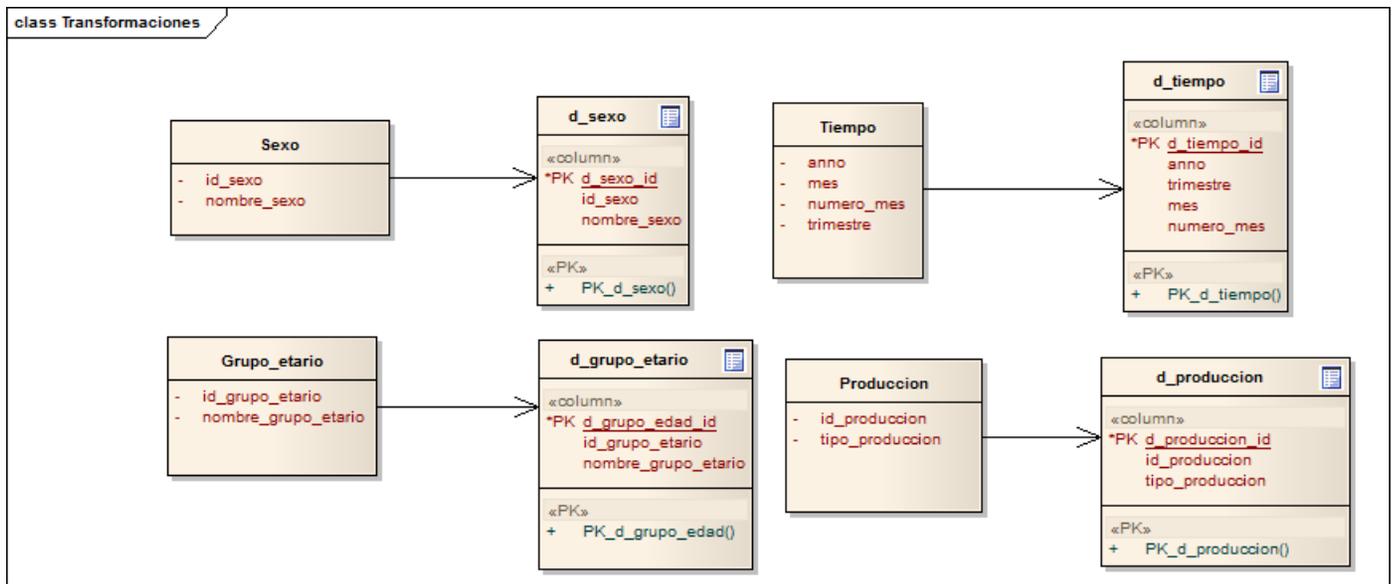
Campos que contendrá la perspectiva “Producción”

Nombre: Perspectiva “Producción”		
----------------------------------	--	--

Anexos

Atributos	Tipo	Descripción
id_produccion	Integer	Identificador de la producción.
tipo_produccion	Character Varing	Tipo de producción.

Anexo 3. Tablas de dimensiones



Anexo 4. Descripción de las tablas de hechos y puentes

Tabla “hecho_registro_medicamentos”

Nombre: “hecho_registro_medicamentos”		
Atributos	Tipo	Descripción
hecho_registro_medicamentos_id	Integer	Identificador del hecho.
d_medicamento_id	Integer	Identificador que proviene de la tabla “d_medicamento”.
d_provincia_id	Integer	Identificador que proviene de la tabla “d_provincia”.
d_fabricante_id	Integer	Identificador que proviene de la tabla “d_fabricante”.
d_tiempo_id	Integer	Identificador que proviene de la tabla “d_tiempo”.
d_produccion_id	Integer	Identificador que proviene

Anexos

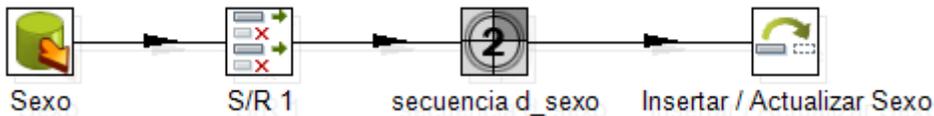
		de la tabla "d_produccion".
consumo	Integer	Cantidad consumida de un medicamento determinado.

Tabla "puente_d_medicamento_hecho_pacientes"

Nombre: "puente_d_medicamento_hecho_pacientes"		
Atributos	Tipo	Descripción
hecho_pacientes_por_medicamento_controlado_id	Integer	Identificador del hecho.
d_medicamento_id	Integer	Identificador que proviene de la tabla "d_medicamento".

Anexo 5. Transformaciones de las dimensiones

Dimensión "d_sexo"



Dimensión "d_grupo_etario"



Dimensión "d_tiempo"



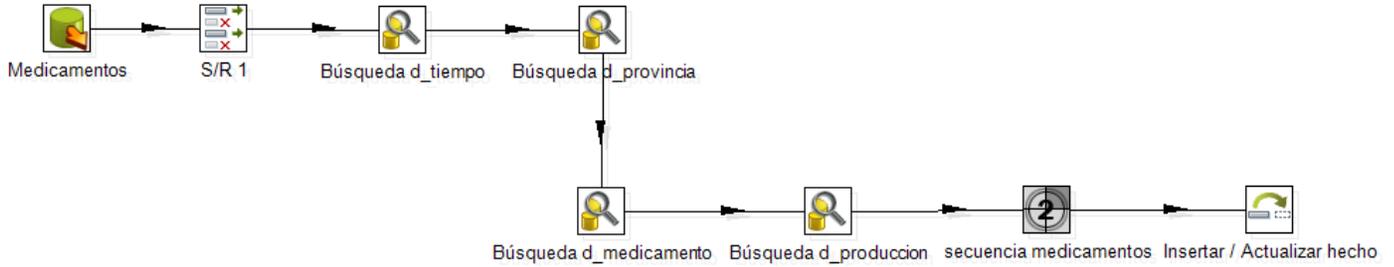
Dimensión "d_produccion"



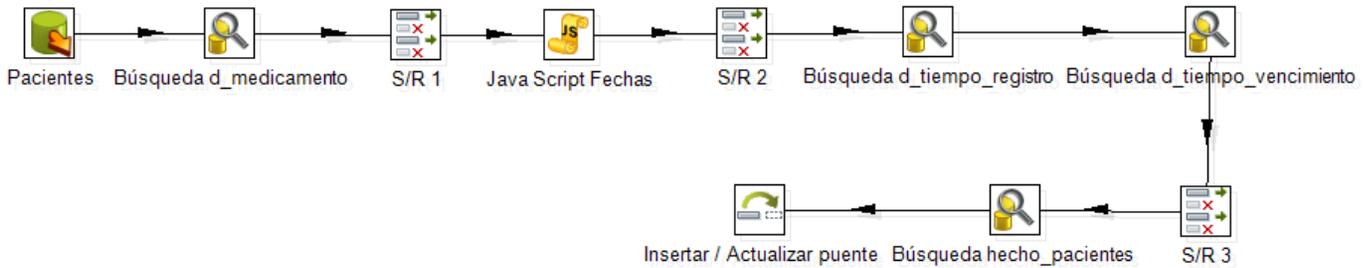
Anexo 6. Transformaciones de los hechos y los puentes

Anexos

Tabla “hecho_registro_medicamentos”



Puente entre la dimensión “d_medicamento” y la tabla “hecho_pacientes_por_medicamento controlado”



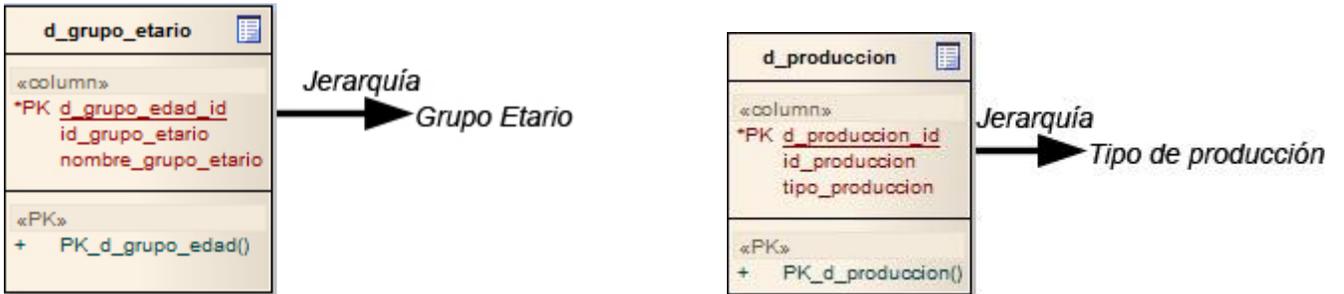
Anexo 7. Metadatos

Tabla “metadatos”



Anexo 8. Jerarquías de los atributos





Anexo 9. Medidas del mercado de datos

Medida “Cantidad_de_pacientes” del cubo “Pacientes_por_Medicamento_Controlado”

The screenshot shows the configuration of a cube in a BI tool. The main window displays the cube 'CuboPacientes_por_Medicamento_Controlado' with the following dimensions:

- d_provincia
- d_raza
- d_sexo
- d_registro
- d_vencimiento
- d_grupo_edad
- d_medicamento_controlado
- d_fabricante
- Cantidad_de_pacientes** (selected measure)
- Total_de_registros

The right-hand pane shows the properties for the selected measure:

Attribute	Value
name	Cantidad_de_pacientes
description	
aggregator	distinct-count
column	id_paciente
formatString	
visible	<input checked="" type="checkbox"/>
datatype	Integer
formatter	
caption	Cantidad de pacientes