PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

FACULTAD DE CIENCIAS E INGENIERÍA

ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN DATAMART PARA EL SOPORTE DE TOMA DE DECISIONES Y EVALUACIÓN DE LAS ESTRATEGIAS SANITARIAS EN LAS DIRECCIONES DE SALUD

Tesis para optar por el Título de Ingeniero Informático, que presenta el bachiller:

Carmen Pamela Rosales Sedano

ASESOR: Carla Basurto Figueroa

Lima, Enero del 2009

Resumen

El presente proyecto de tesis tiene como objetivo la implementación de un datamart que permita apoyar la toma de decisiones necesarias para cumplir con los objetivos específicos de cada estrategia sanitaria nacional dentro de las direcciones de salud.

Se tomó como estrategia sanitaria piloto, la estrategia de *Alimentación y Nutrición Saludable*. Para ella, se realizó la captura de requerimientos, análisis, diseño y construcción del datamart.

Los resultados obtenidos son los reportes analíticos útiles para tomar decisiones de forma rápida y confiable.

Dedicatoria:

Este proyecto de tesis está dedicado a mis padres por brindarme su apoyo incondicional y por creer en mí cada día.

Agradecimientos:

A Dios, por acompañarme siempre

A mis padres Carmen y Víctor, por su infinito amor y su confianza en mí

A mis hermanos, por su comprensión

A mis profesores, por sus consejos

Y a mis mejores amigos, por brindarme su amistad sincera.

Tabla de Contenido

Índice de Figuras

Índice de Cuadros

Introducción

En nuestro país, uno de los servicios importantes que, sin embargo, no cuenta con el apoyo suficiente que debería recibir, es el área de salud, en este hay tres problemas principales que el estado debe afrontar: problemas sanitarios, del sistema de salud y de los factores determinantes (medio ambiente, nutrición, educación, entre otros).

El ministerio de salud (**MINSA**) ha emprendido un nuevo plan concertado de salud para lograr disminuir y eliminar los problemas antes mencionados, parte del plan es mejorar el sistema de salud allí donde el adecuado manejo de información ayudaría al mejor control y seguimiento de acciones que se están realizando. En general, el MINSA tiene como obligación establecer las políticas necesarias para llevar a cabo todo tipo de acciones y además ha establecido ***10 estrategias sanitarias nacionales*** con la finalidad de dar énfasis a temas fundamentales para la salud del país, a través del abordaje, control, reducción erradicación o prevención de los daños-riesgos priorizados.

Uno de los objetivos principales del plan concertado es el de lograr la total descentralización, es decir que cada gobierno regional vele por el cumplimiento del plan y las estrategias sanitarias dentro de su jurisdicción. Esta misión esta a cargo de la Dirección de Salud (DISA), la cual controla y dirige a centros de salud designados.

El proyecto se centra en el punto del manejo eficiente de información de cada estrategia sanitaria dentro de cada una de las DISAs del país.

1. Capítulo 1: Generalidades

Este capítulo tratará sobre la definición del problema que se desea solucionar, marco conceptual necesario para entender el contexto, gestión del proyecto, estado del arte y descripción de la solución.

1.1. Definición del Problema

El sistema de salud en el Perú ha sido manejado de forma centralizada por muchos años, siendo el centro la capital, Lima, sin embargo esto ha demostrado que los resultados no son satisfactorios sobretodo ha causado ineficiente distribución de recursos médicos y personal. Como parte de la modernización del estado peruano, se ha incluido la descentralización del sector de salud en respuesta al problema, para llevarlo a cabo se ha tenido que realizar análisis de ventajas y desventajas de dicho proceso, así como la estimación de costos y la anticipación de problemas que se pudieran presentar (problemas socio-económicos, ambientales, políticos.).

Aunque la descentralización parece ser una buena solución, no todas las actividades pueden se descentralizadas, entre estas están las que establecen normas de control de calidad de servicios, de enfermedades transmisibles, de salud ocupacional y otras políticas administrativas, en la actualidad el encargado de estas actividades es el ministerio de salud (MINSA).

Aquellas actividades de salud que sí son factibles de ser descentralizadas, serán ejecutadas por cada uno de los gobiernos regionales, específicamente por las

Direcciones Regionales de Salud (DISA) que son encargadas de dicho sector.

El tema de descentralización se encuentra en el Plan Nacional Concertado de Salud, en uno de sus lineamientos de política denominado "*La Descentralización de la función salud a nivel del Gobierno Regional y Local*" señala que objetivo estratégico es que los gobiernos regionales y locales ejerzan plenamente sus funciones en materia de salud.

A continuación se muestran las metas al 2011 del lineamiento:

- Culminar el proceso de transferencia de las funciones de salud a los gobiernos regionales
- Incrementar las capacidades de gobierno suficientes y necesarias para la gestión en salud en el nivel regional y local.

El estado ha iniciado el proceso de descentralización de salud siguiendo el lineamiento mencionado, sin embargo aún falta la total transferencia de funciones a los gobiernos regionales y locales. Se muestra a continuación aquellos puntos importantes que todavía no han sido totalmente implementados [ART01]:

- Formular, aprobar, ejecutar, evaluar, dirigir, controlar y administrar las políticas de salud de la región en concordancia con las políticas naciones y los planes sectoriales.
- Coordinar las acciones de salud integral en el ámbito regional
- Promover y ejecutar en forma prioritaria las actividades de promoción y prevención de la salud.
- Organizar, implementar y mantener los servicios de salud para la prevención, protección, recuperación y rehabilitación en materia de salud, en coordinación con los gobiernos locales.
- Conducir y ejecutar coordinadamente con los órganos competentes la prevención y control de riesgos y daños de emergencia y desastres.

El objetivo de la descentralización es hacer un estado al servicio de los ciudadanos y que los servicios que presta sean cada vez mejores a través del ejercicio del gobierno en los niveles nacional, regional, local.

Como parte del desarrollo del objetivo, el MINSA ha definido las estrategias sanitarias, las cuales se definirán en la sección siguiente, las que deberán ser controladas y monitoreadas de forma descentralizada por cada DISA.

Actualmente cada DISA cuenta con un área central de estadística la cual

proporciona solo reportes estáticos a todas las estrategias sanitarias que alberga. Sin embargo el proceso de solicitud puede demorar cierto tiempo ya que el área de estadística, además, tiene que atender otras solicitudes de las demás áreas que se encuentran en la DISA. Es por ello que la toma de decisiones se ve retrasada por no contar con los reportes en el tiempo estimado.

Para hacer posible que el objetivo y las metas del lineamiento se cumplan se requiere un sistema de información que permita el monitoreo y la evaluación de las estrategias dentro de cada dirección regional siguiendo siempre las políticas fijadas en el Plan Concertado, de esta manera se podrá controlar los avances que se realicen y el cumplimiento de las políticas. La inteligencia de negocios será de utilidad para la evaluación de los resultados de cada estrategia sanitaria de las DISA's.

Para el alcance del proyecto de fin de carrera, solamente se realizará la solución para la estrategia sanitaria de *Alimentación y Nutrición Saludable*.

1.2. Marco conceptual del problema

En esta sección procederemos a explicar los conceptos mínimos requeridos para llevar a cabo el presente trabajo de fin de carrera.

1.2.1. Conceptos: Inteligencia de Negocios

Inteligencia de Negocios

Es un enfoque para la gestión empresarial que permite a una organización definir que información es útil y relevante para la toma de decisiones corporativas. Inteligencia de Negocios es un esquema polifacético que fortalece a las organizaciones para tomar mejores decisiones mas rápidamente, convertir los datos en información y usar una estrategia inteligente para la gestión empresarial [BIB01].

Data Warehouse

Es un almacén o repositorio para los datos. Muchos expertos definen el data warehouse como un almacén de datos centralizados que introduce datos en un almacén de datos específico llamado datamart. Otros aceptan una amplia definición de data warehouse, como un conjunto integrado de datamarts.
Es utilizado para el proceso de toma de decisiones gerenciales. [BIB01]

Datamart

Es un conjunto de datos que son estructurados de forma que facilite su posterior análisis. Un datamart contiene la información referente a un área en particular, con datos relevantes que provienen de las diferentes aplicaciones operacionales. Los datamarts pueden ser de diversas bases de datos OLAP dependiendo del tipo de análisis que se quiera desarrollar

Explotación de datos

La explotación de la información se realiza a través de un amplio conjunto de herramientas de consulta y análisis de la información. Estas herramientas de explotación son sistemas que ayudan al usuario a la exploración de los datos y generación de vistas de información. Se dividen en reporteadores, sistemas de análisis multidimensional, sistemas de apoyo a la toma de decisiones y sistemas de información ejecutiva.

ETL (extract - transformation - load)

ETL significa en español *extracción, transformación y carga* de los datos.

La extracción es el primer paso en el proceso de obtención de datos en el entorno del data warehouse. Extracción significa leer, entender los datos fuentes y copiar los datos necesarios para el DWH en el *Staging Area* (área de ETL) para su manipulación posterior.

Una vez que la data es extraída del *Staging Area*, hay numerosas y potenciales transformaciones como la limpieza de datos (corrección de datos escritos de forma errónea, resolviendo conflictos, colocación en formatos nuevos, etc.), combinando datos de múltiples fuentes, eliminando datos duplicados y asignando *warehouse keys* (llaves primarias en el data warehouse).

Finalmente con todas las transformaciones, los datos son cargados en el área de presentación del data warehouse [BIB01].

1.2.2. Conceptos: Análisis Dimensional

Fact Table

Es la tabla central del esquema en estrella que representa datos numéricos en el contexto de las entidades del negocio. La tabla de hechos esta constituida por medidas y por foreign keys La fact table expresa la relación muchos a muchos entre las dimensiones dentro del modelo dimensional. [BIB01]

Dimensiones

Son objetos del negocio con los cuales se puede analizar la tendencia y el comportamiento del mismo. Las definiciones de las dimensiones se basan en políticas de la compañía, e indican la manera en que la organización interpreta o clasifica su información para segmentar el análisis facilitando la observación de los datos. [BIB01]

Medidas o métricas

Son características cualitativas o cuantitativas de los objetos que se desean analizar en las empresas. Las medidas cuantitativas están dadas por valores o cifras porcentuales. Por ejemplo las ventas en dólares, cantidad de unidades en stock, cantidad de unidades de productos vendidos, etc. [BIB01]

Esquema Estrella

Este esquema está formado por un elemento central que consiste en una tabla llamada la Tabla de Hechos (Fact Table), que está conectada a varias tablas de dimensiones.

Las tablas de hechos contienen los valores precalculados que surgen de totalizar valorizar operacionales atómicos según las distintas dimensiones, tales como clientes, productos o periodos de tiempo. Se presenta un ejemplo, el cual representa un evento crítico y cuantificable en el negocio como ventas o costos. Su clave está compuesta por las claves primarias de las tablas de dimensión relacionadas. [BIB01]

Esquema copo de nieve

La figura que se muestra a continuación, presenta una variante del esquema estrella en el cual las tablas de dimensión están normalizadas, es decir, pueden incluir claves que apuntan a otras tablas de dimensión. Las ventajas de esta normalización son la reducción del tamaño y redundancia en las tablas de dimensión y un aumento de la flexibilidad en la definición de dimensiones.

Sin embargo, el incremento en la cantidad de tablas hace que se necesiten más operaciones de unión para responder a las consultas, lo que empeora el rendimiento, además del mantenimiento que requieren las tablas adicionales. [BIB01]

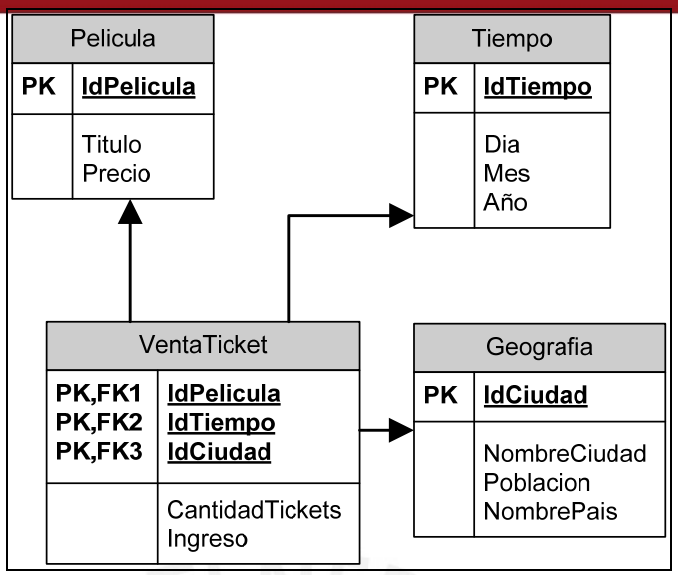


Figura 1.1 Modelo Estrella

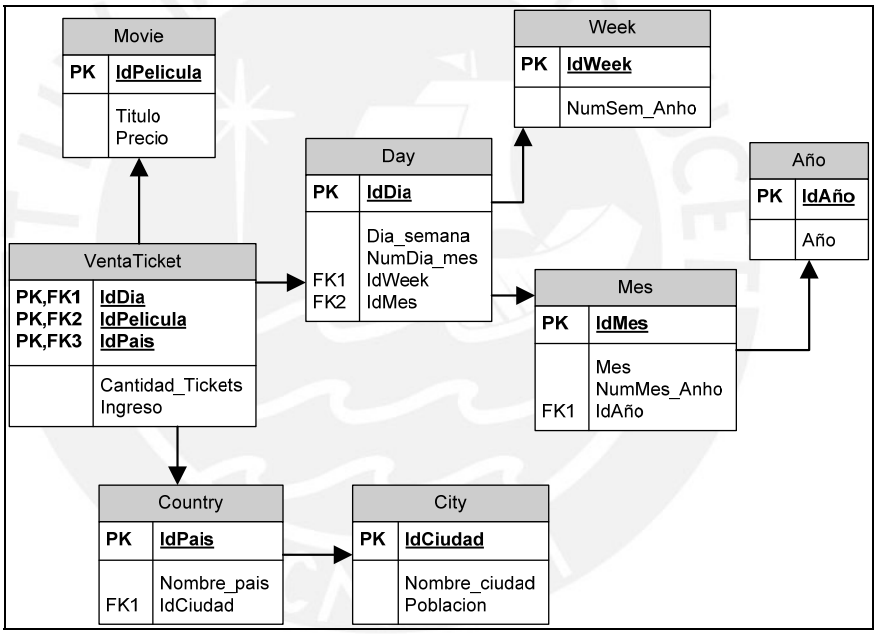


Figura 1.2 Esquema Copo de Nieve

1.2.3. Conceptos: Software Libre

Open Source

Es el software que, una vez obtenido, puede ser usado, copiado, estudiado, modificado y redistribuido libremente. El software libre suele estar disponible gratuitamente, pero no se debe asociar software libre a software gratuito, o a precio del costo de la distribución a través de otros medios; sin embargo no es obligatorio que sea así y, aunque conserve su carácter de libre, puede ser vendido comercialmente

Open Source Business Intelligence (OSBI)

Es una solución de Inteligencia de negocios basada en tecnologías open source

Sistemas OLTP vs sistemas OLAP

Sistemas OLTP *(On-Line Transaction Processing)*

Mientras su tecnología subyacente ha cambiado dramáticamente con el tiempo, las bases de datos operacionales mantienen su misma funcionalidad básica: capturar, actualizar, almacenar y recuperar archivos de datos, dichas base de datos están estructuradas con el propósito de dar apoyo a las operaciones diarias procesando transacciones; ellas no han sido diseñadas para desarrollar análisis de negocio. [BIB02]

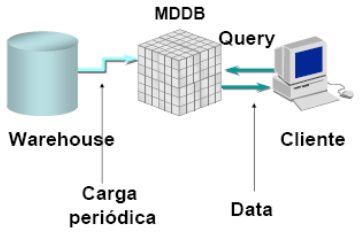
Sistemas OLAP *(On-Line Analytical Processing)*

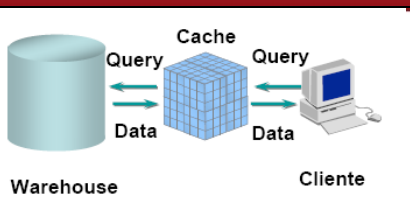
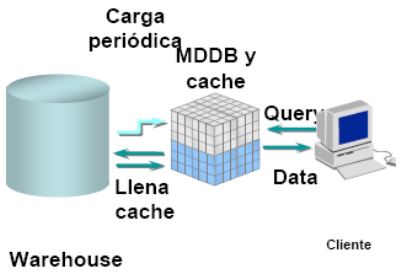
OLAP es una tecnología de análisis de datos que presenta una visión multidimensional lógica de los datos en el data warehouse. La visión es independiente de cómo se almacenan los datos, comprende siempre la consulta interactiva y el análisis de los datos con rapidez, de modo que el proceso de análisis no se vea interrumpido.

Los sistemas OLAP organizan los datos directamente como estructuras multidimensionales, incluye herramientas fáciles de usar por usuarios para conseguir la información en múltiples y simultáneas vistas dimensionales.

OLAP genera rápidos tiempos de respuesta los cuales permiten a los gerentes y analistas preguntar y resolver más situaciones en un corto período de tiempo

El motor de cálculo de OLAP organiza los datos en una forma que permite a los analistas escribir sencillas y directas fórmulas que se ejecutan a través de múltiples dimensiones. Cuenta con tres formas principales de procesamiento: [CLA01]

Forma de Procesamiento	Diagrama
MOLAP (Multidimensional OLAP) *La data es pre-agregada y almacenada en estructuras propietarias conocidas como "Cubos OLAP".* *Tiene un tiempo de respuesta muy bueno para consultas interactivas.*	

ROLAP (Relational OLAP) *Obtiene los datos del data warehouse y los almacena en cubos temporales* *Su tiempo de respuesta es menor*	
HOLAP (Hybrid OLAP) *HOLAP consiste en diseminar los datos a través de bases de datos relacionales y multidimensionales con la finalidad de obtener lo mejor de ambos sistemas*	

Cuadro 1.1 Procesamiento OLAP

1.2.4. Conceptos: Sector Salud

Dirección de Salud

MINSA define a las Direcciones de Salud (DISAs) en [URL07] como los órganos desconcentrados que ejercen la autoridad de salud por delegación de la Alta Dirección y tienen a su cargo, las siguientes funciones generales en sus respectivas jurisdicciones:

a) Implementar la visión, misión, política, objetivos y normas sectoriales, en su jurisdicción.
b) Brindar, en forma eficaz y oportuna, la asistencia, apoyo técnico y administrativo a la gestión de las Direcciones de Red de Salud y de los Hospitales bajo su dependencia y jurisdicción.
c) Mantener informadas a las entidades públicas y organizaciones en general, que desarrollen actividades afines para el Sector Salud sobre los dispositivos legales para la Salud, evaluando su cumplimiento.

Estrategias sanitarias

La priorización de problemas específicos de salud requiere que se aborden por estrategias sistematizadas con enfoque de Salud Pública. La mayoría de dichas prioridades (riesgos y daños) son las Estrategias Sanitarias Nacionales.

Las estrategias sanitarias atraviesan todas las etapas de vida de la persona, no tienen estructura orgánica, requieren de mecanismos propios para el seguimiento de su evolución epidemiológica y de los procesos clave para la producción de

servicios que se relacionan con estos problemas.

El desarrollo de las Estrategias Sanitarias considera:

- Definición precisa de las prioridades sanitarias nacionales y regionales, a partir de un análisis cuidadoso de la salud nacional y regional.
- Identificación del problema de salud, considerando todas las evidencias disponibles de sus factores críticos que influyen en su origen y perpetuación
- Plan de acción intersectorial, como producto del proceso de concertación, que incluye la definición de responsabilidades a todo nivel. [ART03]

Las direcciones regionales deben aplicar dichas estrategias en la población con sus respectivas líneas de acción de cada una. [URL06]

El ministerio de salud ha establecido 10 estrategias sanitarias, se muestra un cuadro con las respectivas estrategias y sus órganos responsables

Estrategia Sanitaria Nacional	**Órganos responsables**
Inmunizaciones	Dirección General de Salud de las Personas
Prevención y Control de Enfermedades Metaxénicas y otras Transmitidas por Vectores	Dirección General de Salud de las Personas
Prevención y Control de Infecciones de Transmisión Sexual y VIH-SIDA	Dirección General de Salud de las Personas
Prevención y Control de Tuberculosis	Dirección General de Salud de las Personas
Salud Sexual y Salud Reproductiva	Dirección General de Salud de las Personas
Prevención y Control de Daños No Transmisibles	Dirección General de Salud de las Personas
Accidentes de Tránsito	Oficina General de Defensa Nacional
Salud de los Pueblos Indígenas	Centro Nacional de Salud Intercultural del Instituto Nacional de Salud
Alimentación y Nutrición Saludable	Centro Nacional de Alimentación y Nutrición del Instituto Nacional de Salud
Salud Mental y Cultura de Paz	Dirección General de Promoción de la Salud

Cuadro 1.2 Estrategias Sanitarias Nacionales

Estrategia de Alimentación y Nutrición Saludable

La Estrategia Sanitaria "Alimentación y Nutrición Saludable" es una de las 10 estrategias del Ministerio de Salud que integra intervenciones (consejería nutricional, sesiones demostrativas, campañas, charlas, talleres) y acciones

priorizadas dirigidas a la reducción de la morbi-mortalidad materna e infantil y a la reducción de las deficiencias nutricionales. Debe coordinar, supervisar y monitorear las diversas actividades relacionadas a la alimentación y nutrición que ejecutan los establecimientos de salud. Su objetivo general mejorar el estado nutricional de la población peruana a través de acciones integradas de salud y nutrición, priorizadas los grupos vulnerables y en pobreza extrema y exclusión.
La estrategia de Alimentación, se propone metas anuales las cuales deben intentar cumplir.

El Modelo de Atención Integral de Salud (MAIS)

Es la forma de aplicar las acciones de salud tomando como eje central las necesidades de salud de las personas en el contexto de la familia y de la comunidad, antes que a los daños o enfermedades específicas. El modelo de atención integral establece la visión multidimensional y biosicosocial de las personas e implica la provisión continua y coherente de acciones dirigidas al individuo, a su familia y a su comunidad desarrollada en corresponsabilidad por el sector salud, la sociedad y otros sectores, para la promoción, prevención, recuperación y rehabilitación de la salud, con la finalidad de mejorar el estado de salud para el desarrollo sostenible.

Las estrategias sanitarias nacionales son parte del Modelo de Atención Integral, conforman uno de los ejes brindando los contenidos de la mayoría de los cuidados esenciales, resolviendo el “qué y cómo lograr la meta”.

Los Paquetes de Atención Integral de Salud:

El Paquete de Atención Integral de Salud, es un conjunto de cuidados esenciales que requiere la persona para satisfacer sus necesidades de salud, brindados por el personal de salud, la propia persona, familia, los agentes comunitarios y otros actores sociales de la comunidad. [URL17]

Programa PANTBC

Es el Programa de Alimentación y Nutrición para el Paciente Ambulatorio con Tuberculosis y Familia. Este programa forma parte del conjunto de intervenciones del Estado Peruano para apoyar la alimentación de grupos vulnerables de la población.

El Programa PANTBC permite ejecutar actividades de complementación alimentaria-nutricional y evaluación nutricional, para contribuir a la recuperación integral del paciente ambulatorio con tuberculosis, y a la protección de su familia. Este programa se integra y complementa con otras actividades que desarrollan los establecimientos de salud en atención a la enfermedad.

Tiene como objetivo principal: contribuir a la recuperación del paciente tuberculoso ambulatorio y proteger a su familia mediante el desarrollo de actividades educativas, evaluación nutricional y el aporte de una canasta de alimentos que brinde el 28% de los requerimientos calóricos y el 38% de los requerimientos protéicos de la familia compuesta por un paciente y dos contactos.

Apoya a los pacientes y familia (dos personas) con diagnóstico de tuberculosis pulmonar o extrapulmonar, que reciben tratamiento ambulatorio en el Programa de Control de Tuberculosis (PCT), de los establecimientos de salud del MINSA.

HIS

El sistema transaccional utilizado para el proyecto de fin de carrera es HIS (en el idioma original, Health Information System).

HIS es una herramienta indispensable que garantiza, el adecuado registro de las actividades de salud, contribuyendo a mejorar la calidad del registro de datos, homogenizando criterios, incorporando nuevas formas de registro y consolidándolo como única fuente de información, con el propósito de instrumentalizar el soporte para la toma de decisiones.

HIS, permite adecuarse a la situación actual de la organización del Sistema de Servicios de Salud, Estrategias Sanitarias Nacionales, Etapas de Vida y Componentes Especiales.

Tiene como finalidad servir como fuente de información básica de la atención ambulatoria diaria brindada a las personas que acuden a los establecimientos de salud, de las Direcciones de Salud y Direcciones Regionales de Salud del País. También sirve de fuente básica de información para la vigilancia epidemiológica en cuanto a morbilidad y de las actividades preventivo-promocionales, realizadas tanto a nivel familiar como en grupos organizados de la comunidad. [ART07]

1.3. Plan de Proyecto

En esta sección se explicará los pasos a seguir para la planeación del proyecto. Se

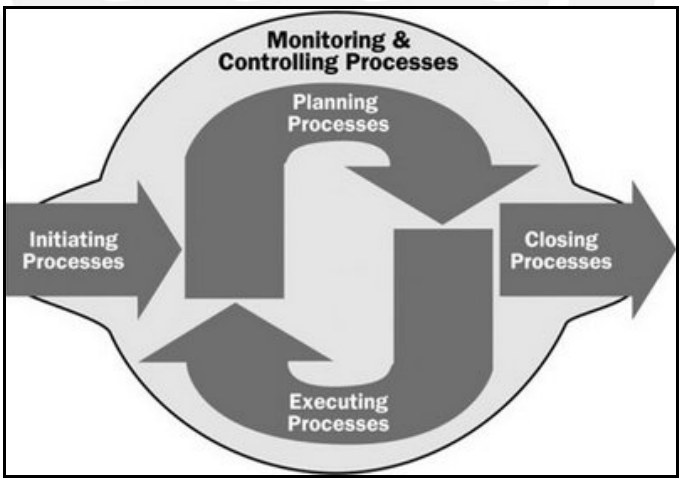
basa en las mejores prácticas establecidas en el PMBOK del PMI

Se definirán algunos conceptos respecto a la dirección de proyectos

Dirección de Proyectos (Project Management)

Es la aplicación de conocimientos, habilidades, herramientas y técnicas con el fin de cumplir los requerimientos de un proyecto. Se logra mediante los procesos de dirección de proyectos: inicio, planeamiento, ejecución, monitoreo y control, y cierre

- **Iniciación.** Definen y autorizan el proyecto o una fase del proyecto.
- **Planificación.** Definen y refinan los objetivos y los planes de acción requeridos para lograr los objetivos y el alcance del proyecto.
- **Ejecución.** Integra a las personas y otros recursos para llevar a cabo el plan de proyecto.
- **Monitoreo y Control.** Mide y monitorea el progreso para identificar desviaciones del plan de proyecto, y así tomar acciones correctivas cuando sea necesario.
- **Procesos de Cierre.** Formaliza la aceptación del producto, servicio o resultado, y conduce al proyecto o una fase del proyecto a su final.



Fuente: A Guide to the Project Management Body of Knowledge (PMBOK® Guide) - Third Edition

Figura 1.3 Procesos de dirección de proyectos

Áreas de Conocimiento.

Para la dirección de proyectos se tiene 44 procesos agrupados en 9 áreas de conocimientos.

- ***Integración del proyecto (Project Integration Management)***

Consiste en los procesos para asegurar la coordinación, unificación y definición de elementos del proyecto.

- ***Alcance del proyecto (Project Scope Management)***

Incluye los procesos para asegurar que el proyecto contiene todo el trabajo necesario para completar el proyecto satisfactoriamente. En el proyecto a desarrollar se estableció como alcance el análisis, diseño, construcción y pruebas del *datamart* para la estrategia sanitaria: "Alimentación y Nutrición Saludable", es decir hasta obtener el producto que vendría a ser los reportes analíticos según los requerimientos que solicita la estrategia.

Con respecto al tiempo, el proyecto tendrá una duración estimada de 12 meses.

- ***Costes del proyecto (Project Cost Management)***

La dirección de costos del proyecto incluye los procesos de planeamiento, estimación y control de costos, de manera que se asegure que el proyecto sea completado dentro del presupuesto establecido.

El proyecto a desarrollar será utilizando en gran parte herramientas libres por lo que no habría problemas por las licencias de estas herramientas y aquellas que requieran de licencia, estas serán proporcionadas por la universidad.

- ***Calidad del proyecto (Project Quality Management)***

La dirección de calidad del proyecto incluye los procesos necesarios para asegurar que el proyecto satisfaga las necesidades para la que se ha llevado a cabo. Incluye todas las actividades de la dirección que determinan las políticas de la calidad, objetivos y responsabilidades, así como su desarrollo a través de la planificación, el control y el aseguramiento de la calidad. Por ejemplo, las reuniones con el asesor del tema sirven para realizar correcciones y/o modificaciones, esto es para garantizar que se está cumpliendo con lo planeado en el proyecto.

- ***Comunicaciones del proyecto. (Project Communications Management)***

La dirección de comunicaciones del proyecto incluye los procesos requeridos para asegurar la precisa y apropiada generación, recolección, distribución, almacenamiento y obtención de la información del proyecto. Provee vínculos críticos entre las personas y la información que son necesarios para una comunicación exitosa entre miembros del proyecto, usuarios, asesores, etc. En el

proyecto a realizar no se aplica esta área del conocimiento.

- ***Riesgos del proyecto. (Project Risk Management)***

La dirección de riesgos incluye los procesos relacionados con la identificación, análisis y respuesta a los riesgos del proyecto. Su objetivo es incrementar los efectos positivos de los distintos eventos y minimizar las consecuencias de sus efectos negativos.

- ***Recursos Humanos del proyecto (Project Human Resource Management)***

La dirección de recursos humanos del proyecto incluye los procesos que organizan y administran al equipo del proyecto. Asigna roles y responsabilidades a los miembros que conforman el equipo para lograr completar el proyecto. Debido a que el proyecto será realizado por una única persona, esta área de conocimiento no se aplica.

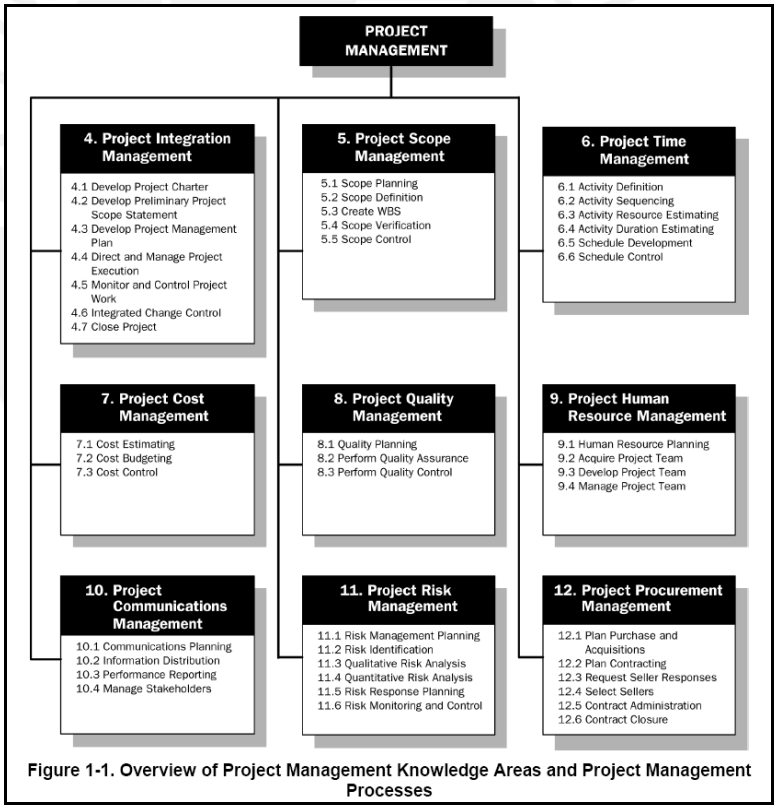


Figure 1-1. Overview of Project Management Knowledge Areas and Project Management Processes

Fuente: A Guide to the Project Management Body of Knowledge (PMBOK® Guide) - Third Edition

Figura 1.4 Áreas de Conocimiento de la Dirección de Proyectos

- ***Aprovisionamiento del proyecto (Project Procurement Management)***

La dirección de aprovisionamiento del proyecto incluye los procesos para comprar o adquirir los productos, servicios o resultados necesarios para la realización del

trabajo. En el presente proyecto no se adquirirá materiales extras, por lo tanto esta área del conocimiento no aplica.

- ***Plazos del proyecto (Project Time Management)***

La dirección de plazos del proyecto incluye los procesos necesarios para asegurar la conclusión del proyecto en los tiempos establecidos. Con ayuda de un diagrama de Gantt se puede gestionar el tiempo para cada fase que ha sido definida en el WBS (The Work Breakdown Structure).

WBS es una herramienta que define un proyecto en grupos de elementos de trabajo (subdivisión del esfuerzo) para así organizar y definir el trabajo total del alcance del proyecto a realizar.

A continuación se presenta el diagrama WBS para el proyecto de tesis:

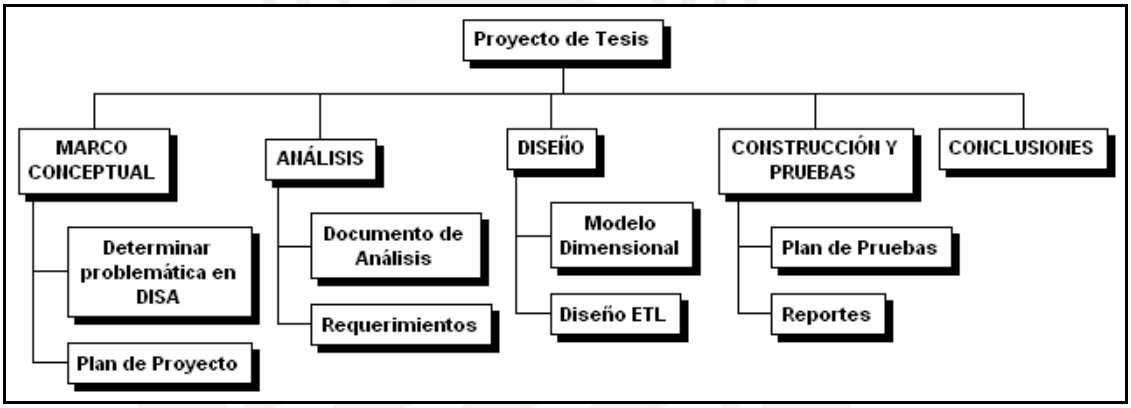


Figura 1.5 Proyecto de Tesis - WBS

Para poder realizar un correcto seguimiento a los plazos establecidos se ha elaborado el Diagrama de Gantt, el cual se encuentra en el *Anexo 1*.

1.4. Estado del Arte

Se presenta aquí las principales soluciones actuales en herramientas de extracción y explotación de datos así como otros trabajos con datamarts.

1.4.1. Herramientas de extracción de datos

En esta sección se expondrán los requisitos para la elección de la herramienta ETL y luego realizar la comparación de características y ventajas entre las herramientas que existen en el mercado.

Requerimientos a cumplir:

1. Forma parte de una plataforma integrada de BI Forma parte de una plataforma integrada de BI.
2. Es multiplataforma.
3. Limpieza de datos.
4. Trabajo con más de una fuente de datos.
5. Filtrado de datos
6. Interfaz gráfica para el desarrollo del proceso ETL.
7. Monitoreo y diseño del proceso ETL.
8. Programación del proceso ETL para que se ejecute automáticamente.
9. Trabajo con las principales bases de datos del mercado como: Microsoft SQL Server, Oracle, PostgreSQL, MySQL entre otras.
10. Soporte de la herramienta.

- **Oracle Warehouse Builder**

Descripción

Oracle Warehouse Builder (OWS) es una herramienta gráfica de Oracle destinada a la extracción, transformación y carga de los datos (ETL); al modelamiento relacional y dimensional; y a la administración de todo el ciclo de vida de los datos y metadatos.

OWS está diseñado para base de datos Oracle.

Requerimientos de hardware y software

OWS está escrito en Java y actualmente se encuentra en la versión 11g.

Además, OWS es multiplataforma, los sistemas operativos que soporta son: Microsoft Windows, Linux/Unix y Solaris.

A continuación se muestra los requerimientos de hardware para MS-Windows y Linux/UNIX

Sistema Operativo	*Requerimiento*
MS-Windows	Las arquitecturas de 32bit y 64bit son soportadas. Se requiere que la computadora cuente con un mínimo de 850MB de espacio de disco y 768MB de memoria disponible.
Linux/UNIX	Se requiere que la computadora cuente con un mínimo de 1100MB de espacio de disco y 768MB de memoria disponible.

Cuadro 1.3 Requerimientos de HW y SW - OWS

Características y ventajas

Permite realizar las siguientes actividades:

- Data Profiling: Permite descubrir y cuantificar los defectos de los datos antes y durante el proceso de creación del data warehouse o aplicación de inteligencia de negocios.
- Modelado de datos relacionales y estructuras dimensionales
- Soporte para el manejo de las cambios: Permite almacenar y gestionar tanto los datos actuales como históricos a través del tiempo en el data warehouse.
- Integración con Oracle OLAP: Permite cargar directamente la fuente de datos desde la opción OLAP, eliminando el uso de un área temporal para data relacional.
- Permite planificar la ejecución automática de los procesos ETL.
- Integración con con los principales ERPs del mercado tales como SAP; E-Business Suite y PeopleSoft.
- Administración de Seguridad: Permite controlar la seguridad de acuerdo a la metadata del perfil de usuario.
- Auditoría de datos.
- Limpieza de datos para maximizar la calidad de la información
- Diseño y gestión de metadatos corporativos.

➢ **IBM Websphere DataStage**

Descripción

IBM ® WebSphere DataStage ® es una herramienta ETL que integra los datos de múltiples y grandes volúmenes de fuentes de datos y metadatos.

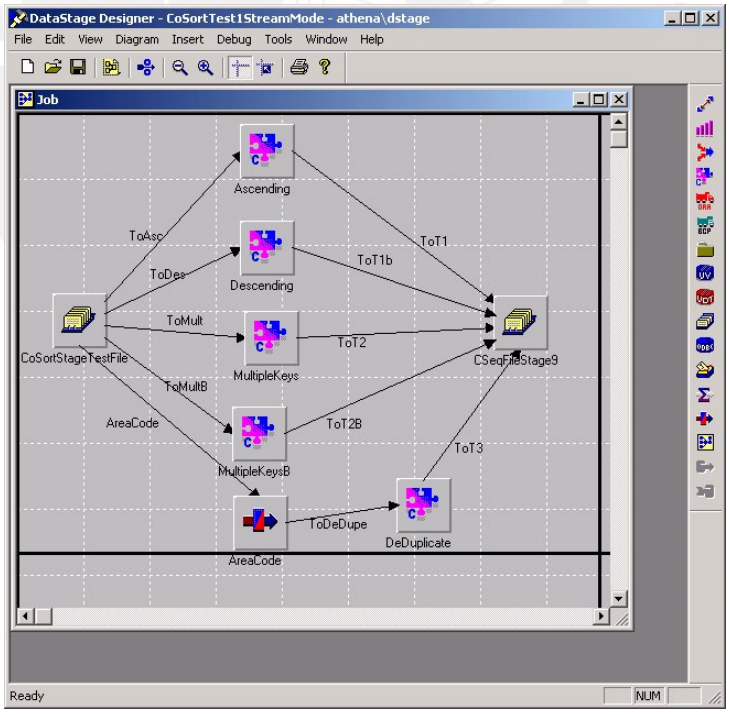
DataStage gestiona los datos que llegan en tiempo real, así como los datos recibidos a diario, semanal o mensual

Requerimientos de software

IBM WebSphere DataStage es multiplataforma, soporta los sistemas operativos Microsoft Windows, Linux (Red Hat).

Características y ventajas

- Brinda soporte para la extracción, integración y transformación de grandes cantidades de datos, con estructuras de datos que van desde simple a muy complejas
- Ofrece una plataforma escalable que permite a las compañías resolver grandes problemas de negocios a través del alto rendimiento en procesamiento de datos masivos.
- Usa notaciones gráficas para construir la solución de integración de datos.
- Soporta un número ilimitado de fuentes de datos heterogéneas, prácticamente todas las bases de datos como Oracle, IBM DB2, IBM Informix, Sybase, Teradata y Microsoft SQL Server, además incluye otras fuentes como: archivos de texto, complejas estructuras de datos en XML, sistemas ERP como SAP y PeopleSoft. [URL10]



Fuente: http://www.mainstream.co.il/ComplexSortJob.jpg

Figura 1.6: IBM WebSphere DataStage

➢ **SQL Server Integration Services (SSIS)**

Descripción

SSIS es un componente de Microsoft SQL Server 2005 y 2008, está disponible únicamente en sus versiones “Estándar” y “Enterprise”.

SSIS ha reemplazado a la característica del SQL Server 2000, Data Transformation Service (DTS). Provee una plataforma para diseñar y generar soluciones de integración de datos y aplicaciones workflows.

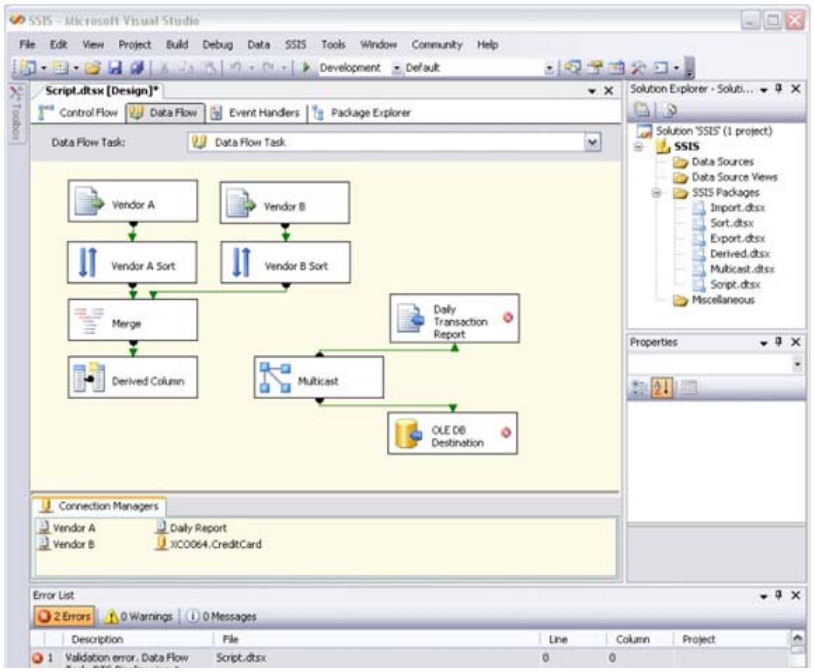
Su principal característica es ser una herramienta flexible para el proceso de ETL.

Requerimientos de software

El sistema operativo que lo soporta es Microsoft Windows

Características y ventajas

- SSIS puede extraer y transformar datos de una amplia variedad de fuentes como archivos XML, archivos planos y bases de datos relacionales para luego cargar los datos en uno o más destinos.
- SSIS posee un amplio rango de componentes como lookups, ordenamiento, agregación, combinaciones entre otros. Que son usados para el diseño del ETL.
- Provee de modo debug gráfico.[URL20]



Fuente: http://www.programminghelp.com/

Figura 1.7: SQL Server Integration Services

➢ **Kettle, Pentaho (Actualmente Data Integration)**

Descripción

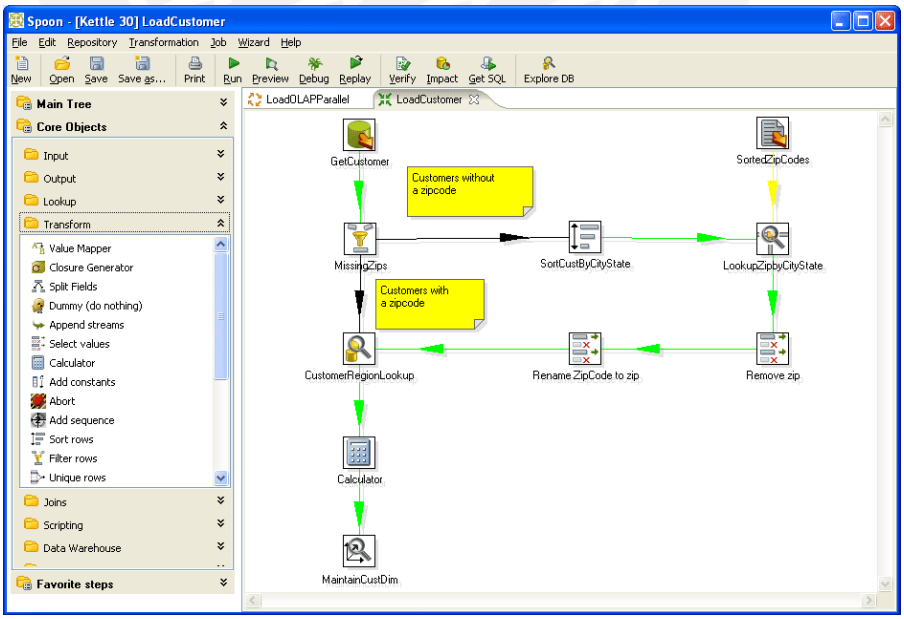
Kettle Pentaho es una herramienta que permite implementar el proceso de ETL, además forma parte de la plataforma Open Source Pentaho Business Intelligence.

Requerimientos de software

Es multiplataforma, soporta sistemas operativos MS-Windows, Linux/UNIX.

Características y Beneficios:

- Soporta las siguientes base de datos, desde MySQL, Oracle, AS/400, MS Access, MS SQL Server, IBM DB2, PostgreSQL, Intersystems, Caché, Informix, Sybase, Gupta SQL Base, dBase III, IV o 5, Firebird SQL, MaxDB (SAP DB), Hypersonic, Generic, CA Ingress, SAP R/3 System.
- La creación de transformations o jobs (secuencia de pasos para realizar la carga de una dimensión o tabla de hechos) se pueda realizar a través del método drag-and-drop que provee el ambiente gráfico de Kettle.
- Permite la migración de datos entre diferentes bases de datos.
- Exporta e importa archivos Excel, txt, entre otros.
- Cuenta con licencia Open Source.



Fuente: http://www.pentaho.com/products/

Figura 1.8: Transformation en Kettle-Pentaho (Data Integration)

- Cuenta con cuatro herramientas indispensables para el desarrollo de los procesos ETL:
    - Spoon: Diseña transformaciones ETTL usando el entorno gráfico.
    - PAN: ejecuta las transformaciones diseñadas por Spoon.
    - CHEF: crea los jobs o trabajos de ETL.
    - Kitchen: ejecuta los jobs o trabajos ETL.

Análisis y Comparación

Luego de revisar las principales características de las herramientas de extracción candidatas: Oracle Warehouse Builder, IBM Websphere DataStage, SSIS y Kettle-Pentaho, se realizó un cuadro comparativo (Cuadro 1.4) con los requerimientos que debe cumplir la herramienta seleccionada.

Para ello se le ha asignado una calificación a cada requerimiento que va de 0 (no cumple) a 5 (cumple totalmente).

Se muestra que Pentaho Kettle obtuvo mayor puntaje además tiene una ventaja sobre las otras que es contar con licencia open source, a pesar que actualmente no existe un vasto soporte, la documentación que se encuentra disponible de forma oficial cubre la mayor parte de las principales necesidades del proyecto de tesis.

Recientemente, el personal de Pentaho está brindando soporte técnico, actualizaciones y/o mejoras aunque no está cubierto por la licencia open source.

La herramienta Kettle, al igual que las otras herramientas, es intuitiva y de fácil aprendizaje ya que presenta un entorno gráfico para el diseño del ETL.

Criterio	Pentaho Kettle	SSIS	IBM Websphere DataStage	Oracle Warehouse Builder
Forma parte de una plataforma integrada de BI	5	5	5	5
Es multiplataforma	5	3	5	5
Limpieza de datos	5	5	5	5
Trabajo con más de una fuente de datos.	4	5	5	5
Filtrado de datos	4	4	4	4
Interfaz gráfica para el desarrollo del proceso ETL	5	4	5	3
Permitir el monitoreo del proceso ETL, así como el diseño del mismo.	4	3	3	3
Programación del proceso ETL para que se ejecute automáticamente	5	4	4	5
Trabajo con las principales bases de datos del mercado como: Microsoft SQL Server 2005 y Oracle Database9g.	5	5	4	5
Soporte de la herramienta.	4	5	5	5
Puntaje Total	**46**	**43**	**45**	**45**

Cuadro 1.4 Cuadro Comparativo – Herramientas Explotación

1.4.2. Herramientas de explotación de datos

Se tienen también diversas herramientas de explotación. Desde el momento en que la mayoría de los ejecutivos están aplicando sus propias pruebas y preparan sus propios informes, es imperativo que estas herramientas sean, al mismo tiempo, de

fácil acceso y simples de usar para la exploración de los datos.

A continuación se muestran los requisitos necesarios para la elección de la herramienta y luego una comparación entre algunas de las principales herramientas.

Requerimientos a cumplir:

- Análisis de la información en todas las dimensiones
- Generación de reportes gráficos en base a la información
- Exportación a archivos Excel para visualizar la información
- Drill and Down: Detalle de los datos accediendo directamente a las fuentes multidimensionales.
- Filtros y búsquedas personalizables.
- Slice and Dice: Operaciones de corte y rotación de datos.
- Soporte de la herramienta.
- Ser multiplataforma.

➢ **Business Objects OLAP Intelligence**

Es una herramienta que permite acceder y analizar los datos almacenados en el servidor OLAP.

Características:

- Provee una interfaz basada en web que permite a los usuarios seleccionar dimensiones y miembros de un panel de consultas, dando la posibilidad interactuar directamente con la información y generar consultas espontáneas que permitan descubrir anomalías.
- Provee una interfaz intuitiva y rica en funciones para catalogar, filtrar y realizar rápidos cálculos sobre la información. A su vez permite esconder dimensiones que no sean relevantes para el análisis de la información.
- Acceso a servidores OLAP de Microsoft, Hyperion y SAP, así como a servidores de datos multidimensionales como Microsoft SQL Server Analysis Services, Hyperion Essbase, IBM DB2 OLAP, y SAP BW.
- Entorno gráfico y asistente para desarrollo de hojas de trabajo para OLAP, donde se puede colocar la información en el formato que se crea conveniente para luego ser publicada para otros usuarios del sistema.
- Opciones de drill-down de información sumarizada a reportes detallados. Esto permite ampliar el nivel de análisis de los usuarios, permitiéndoles visualizar un mayor detalle de la información con que se cuenta para descubrir alguna anomalía en la misma.

- Interfaz con Excel para mostrar la información desde este sistema de hoja de cálculo. [URL21]

➢ **Microsoft AS (Analysis Services)**

Es una aplicación que permite ofrecer una visión unificada e integrada de todos los datos del negocio como base para la elaboración de todos los informes y análisis OLAP.

Características:

- Modelo Dimensional Unificado: El UDM (Unified Dimensional Model) es el repositorio central de metadatos del Analysis Services. En este se definen las entidades empresariales, lógica empresarial, cálculos y medidas que sirven de origen para todos los informes, hojas de cálculo, exploradores OLAP y aplicaciones de análisis. Los usuarios pueden emitir consultas en el UDM mediante diversas herramientas de cliente, como Microsoft Excel.
- Almacenamiento en cache proactivo: Cuenta con almacenamiento en caché proactivo, que permite combinar las ventajas de las actualizaciones en tiempo real con el rendimiento que brinda la arquitectura MOLAP. Esto permite mantener automáticamente la caché mientras los datos se van modificando en los orígenes de datos subyacentes. Además se logra obtener un rendimiento sobresaliente en las consultas.
- Procesamiento analítico en línea (OLAP): permite obtener acceso a datos organizados y agregados de orígenes de datos empresariales, como por ejemplo almacenamientos de datos, en una estructura multidimensional denominada cubo. El Analysis Services proporciona herramientas y características para OLAP que puede utilizar para diseñar, implementar y mantener cubos y otros objetos compatibles. [URL18]

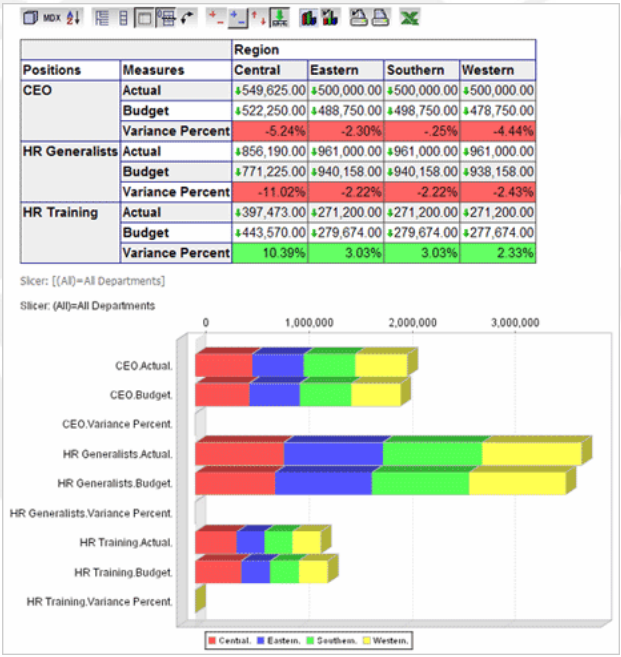
➢ **Pentaho Open BI Suite - Pentaho Analysis**

Parte de la suite Pentaho BI.

Pentaho Análisis suministra a los usuarios un sistema avanzado de análisis de información. Con uso de las tablas dinámicas (pivot tables, crosstabs), generadas por Mondrian y JPivot, el usuario puede navegar por los datos, ajustando la visión de los datos, los filtros de visualización, añadiendo o quitando los campos de agregación. [URL09]

Características:

- Genera informes de forma ágil y permite a través de Pentaho Reporting la distribución de los resultados del análisis en múltiples formatos, incluyendo la opción de imprimir o exportar a formato PDF, XLS, HTML y texto.
- Para el análisis toma en cuenta los 6 elementos básicos del sistema OLAP: Dimensiones, valores, jerarquías, niveles, atributos e indicadores.
- Al utilizar la arquitectura MOLAP esta permite tener datos agregados o pre-calculados, ya que estos residen en el mismo formato multidimensional.
- Funcionalidad de procesamiento, donde dos aplicaciones ayudan a obtener los procesos OLAP, estos son el servidor OLAP Mondrian, y el Jpivot, gracias a estos se pueden realizar querys a Datamart, visualización de resultados mediante un browser, permitir el drill-down.



		Region			
Positions	Measures	Central	Eastern	Southern	Western
CEO	Actual	549,625.00	500,000.00	500,000.00	500,000.00
	Budget	522,250.00	488,750.00	498,750.00	478,750.00
	Variance Percent	-5.24%	-2.30%	-.25%	-4.44%
HR Generalists	Actual	856,190.00	961,000.00	961,000.00	961,000.00
	Budget	771,225.00	940,158.00	940,158.00	938,158.00
	Variance Percent	-11.02%	-2.22%	-2.22%	-2.43%
HR Training	Actual	397,473.00	271,200.00	271,200.00	271,200.00
	Budget	443,570.00	279,674.00	279,674.00	277,674.00
	Variance Percent	10.39%	3.03%	3.03%	2.33%

Fuente: http://www.pentaho.com/images/snap_analysis_olap.png

Figura 1.9: Ejemplo de Reporte de Análisis – Pentaho Análisis.

- Suministra a los usuarios un sistema avanzado de análisis de información. Con uso de las tablas dinámicas (pivot tables, crosstabs), generadas por Mondrian (Servidor OLAP escrito en Java) y JPivot, el usuario puede navegar por los datos, ajustando la visión de los datos, los filtros de visualización, añadiendo o quitando los campos de agregación.
- Permite trabajar sobre diferentes bases de datos, como Oracle, DB2, SQL-Server, MySQL, PostgreSQL, etc., esto a través de Mondrian.

- Puede ser utilizada en UNIX, LINUX y Microsoft Windows.

Análisis y Comparación

Se consideraron tres herramientas candidatas: Pentaho Analysis, Business Objects OLAP y Microsoft Análisis Services.
Al igual que para la selección de herramientas ETL, se ha asignado una calificación por cada requerimiento que debe cumplir la herramienta de 0 (no cumple) a 5 (cumple totalmente).

La herramienta que obtuvo más puntaje, según el cuadro comparativo (Cuadro 1.5), fue Pentaho Análisis y tiene la ventaja de formar parte de la plataforma integral Pentaho BI Suite con licencia Open Source. Como las demás herramientas comparadas, cuenta con una interfaz gráfica para fácil manejo por parte de los usuarios

Otra ventaja es que si se reduce la inversión en licencias, se puede destinar ese dinero a mejorar el análisis, el diseño, la formación de usuarios, la toma de decisiones. La DISA podrá administrar el dinero ahorrado para enfocarse en solucionar los verdaderos problemas, es decir los de salud.
Como una desventaja seria el tema del soporte técnico, pues el que brinda Pentaho tiene un costo y es este el que la DISA tendrá que asumir en caso se presente algún problema con el sistema.

Criterio	Pentaho Analysis	Business Objects	Analysis Services
Análisis de la información en todas las dimensiones	5	5	5
Generación de reportes gráficos	5	5	5
Exportación a archivos Excel para visualizar la información	5	5	5
Drill and Down	5	5	4
Filtros y búsquedas personalizables.	4	3	4
Slice and Dice: Operaciones de corte y rotación de datos	5	5	5
Soporte de la herramienta	4	5	5
Ser Multiplataforma	5	4	3
Puntaje Total	**38**	**37**	**36**

Cuadro 1.5 Cuadro Comparativo – Herramientas Explotación

En el capítulo 2, se realizará el análisis tecnológico, económico y de tiempo de la solución y se comprobará la selección de la herramienta.

1.4.3. Tesis y productos similares

A continuación se muestran algunos trabajos de tesis con datamarts y productos que existen en el mercado relacionadas a mejoras en el sector salud.

- **Tema: Una metodología para sectorizar pacientes en el consumo de medicamentos aplicando datamart y datamining en un hospital**
  Autores: Iván Tapia Rivas, Maria Ruiz Rivera y Edgar Ruiz Lizama

Propone un método para el análisis de datos en la forma con que se consumen los medicamentos en un hospital peruano a fin de poder identificar algunas realidades o características no observables que producirían desabastecimiento o insatisfacción del paciente, el cual servirá como una herramienta para la toma de decisión sobre el abastecimiento de medicamentos en el hospital. Aquí se complementan técnicas

de datamart, de extracción y carga de datos, así como algoritmos de minería de datos como K-means para sectorizar los consumos de medicamentos mencionados. [TES02]

- **Tema: Análisis, diseño e implantación e impacto de un datamart para el área de ventas de empresas de sector farmacéutico**
  Autor: Moro Ríos, Christian.

El objetivo de esta tesis fue realizar un datamart en el área de ventas de gran facilidad de uso, aprendizaje, y que tenía costos reducidos en aspectos de hardware y software.

Dicho datamart se desarrolló con el propósito de mejorar la toma de decisiones de la alta gerencia ya que era indispensable contar con información altamente disponible y de mejor calidad, que es normalmente brindado por sistemas de data warehouse.

Como es sabido se había desarrollado gran competitividad entre farmacéuticas transaccionales, así mismo se contaba con bajo presupuestos para proyectos locales, debido a la baja rentabilidad de la industria farmacéutica en nuestros mercados comparativamente con otras zonas del mundo, tal como Europa o Asia. Toda la compañía farmacéutica, tenía el gran compromiso de tomar las mejores decisiones en sus distintas áreas, eran operaciones de uno o varios países en las que no se podía tener conocimiento o cercanía de los clientes por los escasos recursos humanos y las grandes distancias.
Se determinó que era indispensable información altamente disponible y de mejor calidad, normalmente brindada por los sistemas de data warehouse.

Impacto estratégico luego de la implantación del datamart:
Se obtuvo una organización muy ágil donde el análisis de información es completamente asequible por los empleados de todo nivel, y este fue usado como una ventaja competitiva en el muy ágil mercado farmacéutico.
Se mejoró notablemente los procesos de toma de decisiones, la comprensión del negocio, las áreas de responsabilidad de los empleados, el control de gestión e identificación de las oportunidades de mercado que permitían mejorar la rentabilidad y ventas de la compañía. [TES01]

➢ **Empresa: Med-Vantage®**

Med-Vantage ha desarrollado varias aplicaciones complementarias y métodos de negocios. Además proporciona aplicación propietaria para realizar los procesos de ETL (extraer, transformar y cargar datos) la cual mejora el rendimiento de la producción, validación de datos y la transformación de un gran conjunto de datos administrativos en concretos resultados

Métricas de acciones y resultados

- Costo de atención y reportes analíticos de calidad con múltiple niveles, además de permitir manejar la información (drill and downs)
- Registro de pacientes y recordatorios para realizar seguimientos

Presentación de resultados integrados y personalizados

- Aplicaciones de soporte a decisiones de salud para médicos y miembros
- Navegación web personalizada
- Recopilación de datos y la integración de contenido de salud
- Informes de análisis estadísticos [URL13]

➢ **Empresa: Planwatch**

Análisis Completo del Plan de Beneficios de Salud

Planwatch ofrece un análisis financiero completo de los programas de beneficio de salud (planes de salud) actualmente en curso (ofrecidos actualmente) incluyendo los 25 reportes de nivel superior (top 25) acerca de datos demográficos, diagnóstico / indicadores de estilo de vida, uso del beneficio o plan, proveedor / utilización de la red, y más.

Modelamiento de Plan y Motor de Readjudicación

Calcule el impacto de las modificaciones propuestas al diseño de un plan con cambios simulados con respecto al historial de reclamos y/o quejas. Readjudicación completa usando cantidades modificadas deducibles y co-pagos, porcentajes de co-seguros, máximo nivel de beneficio, y más.

Informes interactivos

Enfoque orientado al informe de análisis avanzado. Planwatch combina la familiaridad de los reportes de negocios estándar con la potencia del software inteligente para negocios modernos. El resultado: una herramienta práctica tanto para informes simples como lo es para las tareas de análisis puro.

Reclamos al Datamart, almacenado y gestionado plenamente

La información se extrae de los sistemas de procesamiento de reclamos TPA,

transformado en una estructura optimizada para el análisis, y se aloja en nuestro centro de datos gestionados. Se ingresa de manera segura desde cualquier equipo con conexión a Internet y navegador web. La gestión de servicios incluye la supervisión, monitoreo, backup automático y soporte. [URL12]

1.5. Descripción y sustentación de la solución

La solución al problema de la descentralización del sector salud con respecto al manejo de información de las 10 estrategias sanitarias dentro de las direcciones de salud es el desarrollo de un sistema de información que permita supervisar, controlar, evaluar los avances y resultados de estas a través del fácil acceso a los datos provenientes de todos los centros de salud asignados a cada DISA. De esta forma se llevara a cabo una estandarización, es decir que cada DISA del país presente la misma forma de manejo y control de las 10 estrategias.

El sistema de información del proyecto de fin de carrera consiste en una solución de inteligencia de negocios que es el análisis, diseño y construcción de un datamart para la evaluación de resultados de las 10 estrategias sanitarias a través de la generación de reportes según requerimientos específicos para cada una de las estrategias y con ello, como se mencionó, permitir predecir que nuevas acciones tomar para mejorar. La información inicial será obtenida del sistema OLTP llamado HIS (Health Information System) en la cual se recogen las consultas médicas y sus diagnósticos realizados en los centros de salud de cada dirección regional, dicha información será dividida y clasificada por estrategia sanitaria y posteriormente será evaluada según sus líneas de acción (requerimientos y acciones).
Para el proyecto de fin de carrera, únicamente se realizará la implementación de la estrategia piloto (“Alimentación y Nutrición Saludable”,) y esta hace uso de un sistema transaccional adicional que maneja la información del programa PANTBC mencionado en el punto 1.2.4.

En general, cada estrategia sanitaria será analizada de forma independiente siguiendo sus requerimientos específicos. La implementación de las demás estrategias se realizará tomando como base la estrategia piloto.

La herramienta que se utilizará para la implementación del datamart, luego del análisis comparativo realizado en el punto 1.4, es Pentaho BI Suite versión 1.6 ya

que consta de licencia libre y todos los módulos necesarios para la extracción y explotación de datos.

2. Capítulo 2: Análisis

En este capítulo se tratará de la definición de la metodología para implementación de un proyecto de Inteligencia de negocios, además de la determinación de requerimientos y análisis dimensional.

2.1. Definición de la metodología

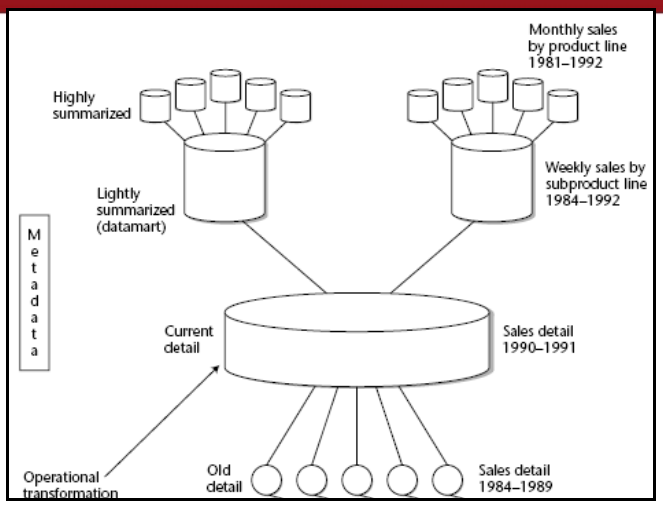
A continuación se explicará las dos metodologías más utilizadas para la construcción del datamart:

2.1.1. El modelo de Inmon

Trata del manejo de una arquitectura de múltiples niveles en el cual el DWH colecciona datos de diversas fuentes, luego estos son integrados y finalmente distribuidos a subconjuntos denominados **datamarts** (enfoque “Top – Down”).

Estructura del Data warehouse

En la figura 2.1 se muestran diferentes niveles de detalle. Hay un antiguo y actual nivel de detalle, un nivel de datos ligeramente sumarizados (nivel datamart) y un nivel de datos altamente sumarizados.



Fuente: The Data Warehouse Toolkit – The Complete Guide to Dimensional Modeling

Figura 2.1 Estructura del DW

Orientación por temas

El DWH está orientado a la división por temas o áreas que se han definido en el modelo de datos del proyecto. Por ejemplo: cliente, producto, cuenta, transacción o actividad.

Diseño del Datawarehouse

Granularidad

Es el aspecto más importante en el diseño del DWH. El mayor detalle es el más bajo nivel de granularidad y el menor detalle es el más alto nivel.

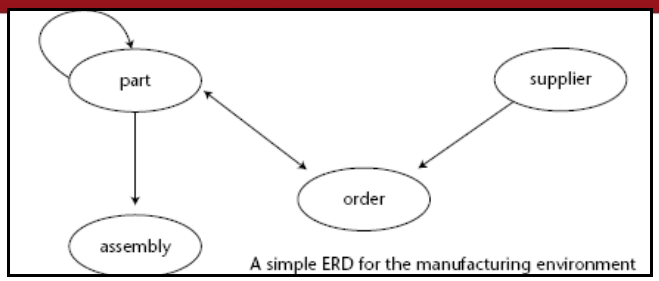
Es el tema más crítico en el diseño del entorno del DWH porque afecta el volumen de datos que residirá en el DWH y el tipo de consultas que serán respondidas.

Modelo de Datos del Data Warehouse

Existen tres niveles de modelamiento de datos:

- *Alto nivel, llamado ERD (Entity Relationship Diagram)*

Consiste en entidades y relaciones. El nombre de la entidad está encerrada por un óvalo, las relaciones entre entidades está representado por flechas. Un ERD corporativo está compuesto por varios ERD's los cuales reflejan diferentes puntos de vistas de las personas de la corporación.

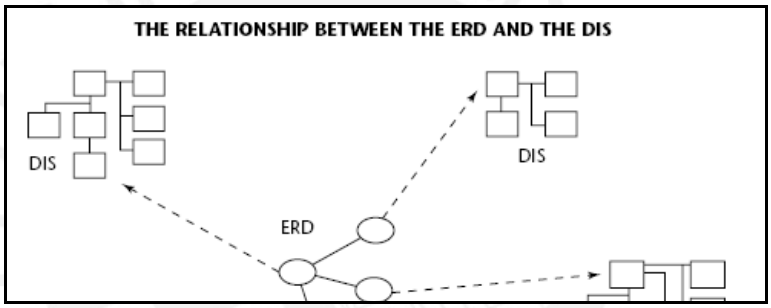


Fuente: The Data Warehouse Toolkit – The Complete Guide to Dimensional Modeling

Figura 2.2. ERD

- *Nivel Medio, llamado DIS (Data Item Set)*

Luego de definir el alto nivel del modelo de datos, se establece el siguiente nivel: DIS. Por cada área principal o entidad identificada en el nivel anterior, se crea un nivel medio para expandir los atributos.



Fuente: The Data Warehouse Toolkit – The Complete Guide to Dimensional Modeling

Figura 2.3. Relación entre ERD y DIS

- *Nivel Bajo, llamado Modelo Físico (Physical Model)*

Creado a partir del nivel medio con la finalidad de extenderlo incluyendo llaves y características físicas del modelo. A este punto, el modelo de datos físicos luce como una serie de tablas llamadas tablas relacionadas.

Proceso de ETL

El DWH es cargado con los datos obtenidos en el entorno operaciones, mas antes dichos datos deben pasar por un complejo proceso de conversión, reformato e integración para así recién ser parte del entorno del DWH.

Acceso Indirecto al DWH

Un enfoque distinto para el diseño de base de datos en el contexto de data warehousing es el enfoque multidimensional. Este enfoque es aplicado exclusivamente en datamarts no datawarehouses..

Se deben obtener los datos del DWH para formar el datamart. En el DWH, los datos

son muy granulares, en cambio en el datamart son más compactos y sumarizados.

2.1.2. Metodología de Ralph Kimball

El modelo de Kimball deja de lado la necesidad de un data warehouse debido a que la mayoría de usuarios desea obtener datos detallados, Kimball argumenta que es mejor almacenar los datos en datamarts independientes y lógicamente conectados usando dimensiones. Para la optimización de consultas y mejorar la facilidad de uso de datamarts, Kimball propone el modelo de datos como esquema estrella.

Componentes del DWH

Existen cuatro componentes importantes para la creación del entorno del DWH. Los componentes se muestran en la figura 2.4, y estos son:

- Sistemas fuentes operacionales (Operational Source Systems)
- Área de limpieza de datos (Data Staging Area)
- Área de presentación de datos (Data Presentation Area)
- Herramientas de acceso a datos (Data Access Tools)

Operational Source Systems (OLTP)

Son los sistemas de registro que capturan todas las transacciones del negocio. Sus prioridades son el procesamiento y la disponibilidad. Consultas contra estos sistemas son limitados.

Data Staging Area

Área de almacenamiento y conjunto de procesos ETL (extract- transformation-load)
Extracción (1er Paso): Obtener datos para el DWH, significa leer y entender los datos de las fuentes y luego copiar los necesarios para el DWH.

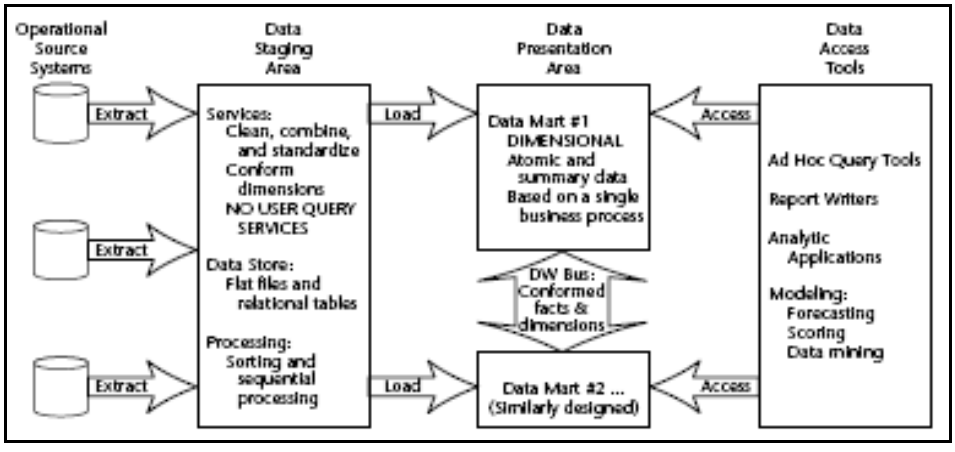


Figura 2.4. Elementos básicos del DWH según Kimball

Transformación (2do Paso): Existen potenciales transformaciones (corregir errores de escritura, conflictos de dominio, elementos perdidos, datos duplicados, asignar DWH keys), combinando datos de múltiples fuentes.

Carga (3er Paso): La carga de datos en el DWH, usualmente toma la forma de presentar tablas dimensionales que aseguren la calidad para cargar cada datamart.

Data Presentation Area

Área donde los datos están organizados, almacenados y disponibles para consultas de los usuarios, reportes y otras aplicaciones analíticas

Es todo lo que los usuarios ven y tocan a través de las herramientas de acceso. Presenta una serie de datamarts integrados en el que cada uno presenta datos de un solo proceso de negocio. Se utiliza el modelamiento dimensional como técnica para entregar los datos a los usuarios, aquí no se busca remover datos redundantes sino más bien rapidez de acceso en las consultas.

Data Access Tools

Puede ser tan simple como un “ad hoc” query (consulta creada para obtener información de acuerdo a la necesidad que surja.) hasta un complejo data mining. Por definición, todas las herramientas de acceso consultan los datos en el área de presentación. El uso de estas herramientas permite a los usuarios tomar decisiones analíticas. Ejm : OLAP, MOLAP, datamart, ad hoc querys, etc.

Pasos para el proceso de diseño dimensional

Paso 1: Selección del proceso de negocio

El primer paso es decidir qué proceso o procesos de negocio se va a modelar para combinar y entender los requerimientos con los datos disponibles.

Un proceso es una actividad normal de negocio realizada en la organización que típicamente está soportada por un sistema OLTP.

Paso 2: Declarar la granularidad

Una vez que se ha identificado el proceso de negocio se debe definir la granularidad. ¿Qué nivel de detalle debería estar disponible en el modelo dimensional?.

Este paso es sumamente importante ya que de él dependerá el modelo de datos y

la implementación del datamart.

Paso 3: Elección de las dimensiones

Se deben elegir las dimensiones que se aplicarán para cada tabla de hechos, las dimensiones deben responder a la pregunta: ¿Cómo los usuarios describen los datos que resultan del proceso de negocios?.

Paso 4: Identificar los hechos

Identificar los hechos numéricos que formarán parte de la tabla de hechos. Los hechos son determinados por esta pregunta: "¿Qué estamos midiendo?". Los usuarios están muy interesados en el análisis de estas medidas de la ejecución de procesos de negocio.

Todos los hechos candidatos deben ser fieles al grano definido en el paso 2. Los hechos que pertenecen a un grano diferente deberán estar en una tabla de hechos separada

2.1.3. Elección de la metodología

Para poder elegir la metodología a seguir durante todo el proyecto es necesario revisar las diferencias y semejanzas que existen.

A continuación se presenta un cuadro comparativo entre la metodología de Kimball e Inmon.

	Kimball	**Inmon**
Objetivo	Todas las empresas necesitan almacenar, analizar e interpretar los datos que van generando y acumulando, para luego tomar decisiones críticas que les permitan maximizar la prosperidad. Para ello se necesita un sistema que les ayude a entender los datos y logren cumplir sus objetivos, de esta forma nace la idea de "implementar un data warehouse".	
Diseño del Data Warehouse	Utiliza el enfoque "Bottom – Up"	Utiliza el enfoque "Top – Down"
Enfoque	Tiene un enfoque por procesos que son manejados por las diferentes áreas del proceso. Trata de responder necesidades específicas según el tema.	Tiene un enfoque global de toda la empresa. No esta basado en requerimientos específicos.
Tiempo de implementación	Ya que primero se implementan los datamarts, el tiempo de	Debido a que se implementa por completo el DWH se demanda

del DWH	implementación es rápido. Sin embargo se tiene que tener cuidado ya que si se trabaja de forma independiente cada datamart el entorno del DWH se desintegraría rápidamente.	mucho más tiempo.
Costos	Implementar cada datamart permite que la solución no presente un alto costo.	Se replican grandes cantidades de datos por tanto los costos aumentan.
Modelo de Datos	Kimball propone usar el modelamiento dimensional: Esquema estrella. Identificación de dimensiones y hechos.	Inmon propone tres niveles en el modelo de datos del DWH: *-Alto nivel, ERD (Entity Relationship Diagram)* *-Nivel Medio, DIS (Data Item Set)* *-Nivel Bajo, llamado Modelo Físico (Physical Model)* Sin embargo, menciona que para implementar los datamarts debe hacerse con modelamiento dimensional.

La metodología de Inmon tiene mayor alcance ya que es un enfoque global, sin embargo para el presente proyecto de tesis solo se requiere la construcción de un datamart que albergue datos acerca de las estrategias sanitarias es así que la metodología de Kimball es la que mejor se adapta a estas necesidades, además el tiempo y costos con los que se cuentan son limitados.

Luego de elegir la metodología para la construcción, se decidió seguir la siguiente estrategia para la ejecución del proyecto la cual consiste en la realización de los siguientes procesos: análisis dimensional, diseño de extracción y diseño de explotación.

Estos procesos se presentan como resumen tomando como referencia las mejores prácticas recomendadas para la implantación de soluciones de inteligencia de negocios, al no existir un estándar para este tipo de implantaciones. Estos son:

Diagnóstico

A través de este proceso se busca identificar la información necesaria para la toma de decisiones, es decir los requerimientos tanto funcionales como no funcionales.

Corresponde al entregable de la etapa de *assess* de la figura 2.5.

En el proyecto, este diagnóstico se realizó mediante entrevistas y revisión de documentos de la estrategia piloto para determinar sus requerimientos.

Análisis Dimensional

Se detallan los elementos que contendrán el datamart de las estrategias sanitarias.

Corresponde al entregable de la etapa de *plan* de la figura 2.5

Esta etapa consistió en el análisis de los diagramas y diccionarios de base de datos de los sistemas transaccionales como HIS y del manejo del programa de PANTBC para así identificar las dimensiones y las tablas de hechos tomando como base los requerimientos solicitados por la estrategia piloto.

En esta fase se elaborará el *Anexo de Análisis (Anexo 2)*.

Diseño de extracción

Se realiza la carga de datos de los sistemas transaccionales mencionados en la etapa anterior, archivos internos u otro tipo de fuentes externas.

En esta fase se elaborará el *Anexo de Extracción (Anexo 3).*

Diseño de explotación

Se realiza el diseño que tendrán los reportes analíticos como: filtros, gráficos, dimensiones y tablas de hechos involucrados.

Según los requerimientos funcionales se elaborarán los reportes.

En esta fase se elaborará el *Anexo de Explotación (Anexo 4).*

Implementación

Mediante de este proceso se logra la emisión de consultas y reportes que presenten la información de manera integrada a través de distintas áreas de la organización, lo cual sirva de ayuda para la toma de decisiones.

El usuario podrá acceder a los datos manipulando los cubos OLAP generados para obtener diferentes perspectivas de análisis, los cuales se reflejan en los reportes analíticos.

En esta fase se elaborará el *Anexo de Construcción (Anexo 5)* que incluye las pruebas de los reportes analíticos.

2.2. Identificación de Requerimientos

Esta sección corresponde a la etapa de diagnóstico de la figura 2.5.

Los requerimientos deben ser identificados según estrategia sanitaria nacional, como se mencionó en el capítulo 1, existen 10 estrategias y cada una merece un análisis por separado.

La captura de requerimientos se hizo a través de entrevistas a los encargados principales de la estrategia a implementar tomando como dirección de salud piloto a la DISA I del Callao, además de documentación acerca de los problemas principales de salud.

Los requerimientos están asociados a las necesidades de todos los actores de la estrategia y se las ha asignado un nivel de prioridad con el fin de priorizar todos aquellos que sean los más importantes y solucionen los problemas encontrados.

2.2.1. Requerimientos Funcionales

Requerimientos Generales

Aunque las estrategias sanitarias demanden requerimientos distintos, presentan ciertos requerimientos generales:

- Toda estrategia está regida por líneas de acción establecidas por el MINSA y deben ser cumplidas a cabalidad. Además debe guiarse según los lineamientos del modelo integral de salud es decir que los requerimientos que se indiquen para cada una de las estrategias deben estar orientados a lograr una adecuada gestión de prestación de salud.
- Para todas las estrategias se realizarán reportes que midan y evalúen la ganancia en salud y bienestar de las personas, de las familias y comunidades. Comparación entre valores de situación inicial con situación deseada y real.

Requerimientos Estrategia Sanitaria Alimentación y Nutrición Balanceada

Se listan los requerimientos que el sistema debe satisfacer para cumplir con el objetivo de la estrategia:

Requerimiento Funcional	Prioridad	Dificultad
El sistema permitirá generar un reporte en el que se	2	1

muestre la evolución de número de casos de actividades o enfermedades a través del tiempo. Es mostrará en el reporte *Evolución de actividades CIE 10 en el tiempo*		
El sistema permitirá generar indicadores de desnutrición según establecimientos de salud.	1	2
El sistema permitirá reflejar la cantidad total de personas de la población que padecen enfermedades relacionados a la alimentación como anemia, hipotiroidismo, bocio, enanismo, desnutrición, retraso, sobrepeso, obesidad, entre otras establecidas por el CIE 10, para cada establecimiento de salud. (prevalencia) Se mostrará a través del reporte de *Prevalencia de Enfermedades*	1	3
El sistema permitirá comparar el estado de cada establecimiento en el aspecto de nutrición para así saber cuales son aquellos que se encuentra en riesgo o necesitan más apoyo.	1	2
El sistema permitirá controlar la entrega de alimentos (en unidades p.e. Kg) a los establecimientos de salud para los pacientes que padecen de TBC (tuberculosis). Se mostrará a través del reporte de *Evolución de entrega de alimento en el tiempo.*	1	2
El sistema permitirá controlar la cantidad de alimentos, destinado a pacientes con TBC, según estados (entregado, deteriorado, robado, perdido) para cada establecimientos de salud Se mostrará a través del reporte de *Estado de alimentos entregados*	1	2
El sistema permitirá el análisis de número de casos de pacientes con TBC ingresantes al programa de PANTBC según su condición nutricional (bajo peso, normal, sobrepeso) Se mostrará a través del reporte de *Condición de pacientes según IMC*	1	3
El sistema permitirá la comparación entre número de	1	3

pacientes con TBC, pacientes que ingresan al programa de PANTBC y pacientes estimados a ingresar al programa de PANTBC. Además muestra indicadores como Atendidos por PANTBC / Estimados. Se mostrará a través del reporte de *Consolidado de atenciones PANTBC*		
El sistema permitirá el análisis del número de casos de egresos de pacientes del programa de PANTBC según condición (recuperado, abandono, fallecido, transferido) Se mostrará a través del reporte de *Condición de egresos de pacientes*	1	3
El sistema permitirá controlar el número de intervenciones realizadas para los pacientes con TBC como consejerías, entrega de canastas de alimentos, charlas educativas y visitas domiciliarias. Se mostrará a través del reporte *Actividades realizadas*	1	1
El sistema permitirá controlar el total de pacientes asistentes con TBC a las actividades realizadas por el programa de PANTBC Se mostrará a través del reporte *Asistencia de pacientes a actividades*	2	1

Leyenda Prioridad:

Calificación	**Significado**
1	Importante
2	Regular
3	Prescindible

Leyenda Dificultad:

Calificación	**Significado**
1	Fácil
2	Regular
3	Difícil

La dificultad esta medida por el tiempo que se requiere tanto para el análisis como para la elaboración de la extracción de los datos necesarios de los sistemas OLTP (HIS y/o PANTBC) para generar el reporte analítico.

2.2.2. Requerimientos no funcionales

Cada usuario será identificado según estrategia sanitaria que tiene a su cargo, además tendrá un determinado rol y privilegios con el fin de asegurar la información.

La seguridad del sistema es un factor crítico pues la información que se maneja es sensible es decir representa información acerca de la realidad de salud en la que se encuentran las personas y no debe ser manipulada por usuarios no autorizados.
Como el software que se utilizará es de código abierto, no se incurrirá en gastos por licencias de software; sin embargo se requiere un servidor de aplicaciones, un servidor de base de datos y un dominio.

La página de administración de Pentaho, herramienta de BI, consiste en un portal web que se utilizará para análisis en el datamart, es desde allí que el usuario con suficiente privilegio generará los reportes y realizará la evaluación de las estrategias.

El uso de la herramienta es intuitivo, debe cumplir:

- La herramienta de explotación seleccionada debe permitir realizar las técnicas de consulta multidimensional como Slicing, Dicing, Drilling, etc. Además, debe permitir la creación de reportes personalizados por el usuario final.

- Exportación de reportes a archivos de formatos estándar como archivo de texto (.txt) o Excel (.xls).

- Mostrar la información del datamart en base a los siguientes 3 tipos: Reportes tradicionales (basados en columnas), Tablas pivote (Matriz de doble entrada) y Gráficos.

Entre otros requerimientos no funciones de la herramienta se tiene:

- *Disponibilidad***:** El sistema al ser de consultas y análisis no impacta a las operaciones diarias que se realizan ya sea en la DISA como en los establecimientos de salud, es por ello que su disponibilidad puede ser menor a 24 horas los 7 días de la semana. Además el sistema no se

encontrar disponible los días en los que se realice la actualización de datos del datamart.

- *Utilidad:* El sistema será apropiado para el uso de los distintos perfiles, de manera que sirvan como soporte de las labores de todas personas que interactúen con él.

- *Escalabilidad:* Las herramientas de negocios permiten cualquier tipo de consultas y reportes que más adelante se podrían solicitar.

2.3. Análisis de la Solución

En esta fase luego de haber identificado los requerimientos para la estrategia piloto se propone una solución adecuada cuya implementación conduzca a la solución del problema.

2.3.1. Consideraciones sobre el sistema

Para el análisis de la viabilidad del sistema se ha considerado tres factores importantes:

Factor tecnológico:

En la DISA donde se realizó la captura de información cuentan con computadoras Pentium III y IV. Para Pentaho es necesario en caso el sistema operativo sea Windows que sea XP o versión superior, para Linux o Mac de preferencia deberían ser de las últimas versiones, además el equipo debe soportar JRE (Java Runtime Environment) y poseer por recomendación 1GB de memoria RAM.
Aquellos usuarios que realizarán el análisis y toma de decisiones requerirán de equipo con los requerimientos sugeridos para Pentaho, pues esto traerá mejoras en rapidez de obtención de resultados.
Cada estrategia sanitaria cuenta con un área propia o compartida con otra estrategia sin embargo cada una trabaja de forma independiente y con su respectivo equipo.

Respecto a la base de datos que se utilizara en la solución, se eligió Postgres principalmente por ser de código abierto y soporta todo lo necesario para la

solución como:

- Consultas complejas incluyendo subconsultas
- Integridad referencial
- Triggers
- Vistas
- PL/pgSQL: Lenguaje de procedimientos almacenados
- Capacidad de multiplataforma, puede ser ejecutado en sistemas operativos Windows y Linux.

Otra ventaja es que cuenta con herramientas graficas para la administración de la base de datos.

Como el alcance del proyecto de tesis no abarca el tema de implantación, no se hará un análisis exhaustivo en este factor.

Factor Económico:

Otro factor importante que se debe considerar es el factor económico en el desarrollo del sistema, por lo que las herramientas a emplearse son de código abierto pues el presupuesto designado al área de salud es limitado y debe ser destinado a otros fines (compra de medicinas, pago de personal, etc.).

El sistema de información o producto puede incurrir en ciertos gastos de mantenimiento como soporte de Pentaho en caso se presente problemas durante su uso. Estos también deben ser considerados y asumidos por la administración de la estrategia sanitaria o la DISA.

Factor Tiempo:

El tiempo designado para la total realización del proyecto de tesis también influye en la implementación, es por ello que el énfasis del sistema irá más enfocado a la funcionalidad que al diseño gráfico que, por ejemplo, podría tener el portal web de la solución Pentaho.

2.3.2. Actores del sistema

Se tienen los principales actores que harán uso de la herramienta:

Encargado Principal de la Estrategia Sanitaria

Son los usuarios que realizarán el análisis y toma de decisiones de las mejores acciones a realizar dentro de su estrategia a cargo.

Asistente

Participante del análisis y encargado de la generación de reportes.

2.3.3. Análisis técnico y económico

Para la elección de las herramientas a utilizarse para la implementación del sistema se ha considerado lo siguientes criterios:

- Curva de aprendizaje: Se requiere herramientas cuyo aprendizaje no sea tan complejo, para evitar retrasos en los tiempos establecidos.
- Lenguaje multiplataforma
- Herramientas libres, de fácil acceso
- Beneficios y características que poseen que los ponen en mejor ventajas frente a otros (Para ello en el capítulo 1 se realizaron análisis comparativos)

2.3.4. Definición del sistema

Se explicará la definición del sistema, esto servirá como base para su diseño, pues permitirá modelar y definir una arquitectura que soporte todos los requerimientos funcionales del sistema y sobretodo a los no funcionales.

En el presente capitulo únicamente se definirá el sistema en relación al análisis, en los capítulos posteriores se explicara la extracción (ETL) y la explotación de los datos (definición de reportes analíticos) del sistema.

Modelo Dimensional

Se muestran y explican todas las dimensiones y tablas de hechos utilizadas para la estrategia sanitaria piloto.

Las dimensiones que conforman los esquemas estrellas para esta estrategia son:

N°	Dimensiones
1.	Actividad_Enfermedad
2.	Establecimiento_Salud
3.	IMC
4.	Tiempo
5.	Paciente
6.	Condicion_Egreso
7.	Actividad_Capacitación
8.	Alimento

Actividad_Enfermedad

Esta dimensión contiene los datos de las actividades y enfermedades.

Están organizados y clasificados según el CIE 10 (Clasificación Estadística Internacional de Enfermedades y otros Problemas de Salud).

Aquellas relacionadas con la estrategia de nutrición son: talleres, planes familias, enfermedades consecuentes de la desnutrición, sobre peso, obesidad entre otras.

Establecimiento_Salud
Esta dimensión contiene los datos de ubicación y características del establecimiento de salud (hospital o postas médicas).

IMC
Esta dimensión contiene datos como talla, edad y peso para así establecer una condición del paciente con respecto a su estado de nutrición.
En realidad esta dimensión es una mini-dimensión es decir permite el manejo de rangos.

Tiempo
Esta dimensión contiene los niveles en tiempo en el cual se harán los filtros de las consultas.

Mini_Paciente
Esta dimensión contiene rangos y atributos en los cuales puede encontrarse un paciente que se atiende en cualquiera de los establecimientos de salud.

Condicion_Egreso
Esta dimensión representa las condiciones de egreso de los pacientes con TBC (Fallecido, Recuperado, Abandono, Traslado).

Actividad_Capacitación
Esta dimensión representa todas las actividades que realiza la estrategia de nutrición dirigidas a los pacientes con tuberculosis.

Alimento
Esta dimensión contiene los datos de los alimentos que son entregados en canastas a los pacientes con tuberculosis. Estos son: cereal, menestra, leche evaporada, leche en polvo, pescado grated y aceite.

Tabla de Hechos (Facts)

Para la estrategia de alimentación y nutrición saludable, luego de haber analizado las dimensiones, se determinaron 6 tablas de hechos. La tabla de hechos IMC_PANTBC deriva de la tabla Atenciones_PANTBC por tanto se colocó en el mismo esquema.

En total son 6 esquemas estrellas las cuales serán necesarias para cumplir los requerimientos funcionales identificados.

N°	Tema	Facts
1	FT_Diagnostico	Fact 1
2	Atenciones_PANTBC	Fact 2
3	Alimentación_PANTBC	Fact 3
4	Actividades_PANTBC	Fact 4
5	Egresos_PANTBC	Fact 5
6	*IMC_PANTBC*	Fact 6

A continuación se presentan las descripciones de cada una de las tablas de hechos:

FT_Diagnóstico

Permite mostrar las actividades-enfermedades por establecimiento de salud, por paciente, por tiempo (fechas), relacionado a nutrición y alimentación.

Atenciones_PANTBC

Permite mostrar las atenciones brindadas por el PANTBC a los pacientes con tuberculosis que siguen el programa PCT (Plan de control de tuberculosis, elaborada por la estrategia de tuberculosis)

Se obtendrán las medidas por establecimiento de salud y por tiempo

IMC_PANTBC

Esta tabla de hechos deriva de Atenciones_PANTBC pues también indica el número de atenciones sin embargo utiliza una dimensión más (IMC), de esta forma se reconoce el estado nutricional de los pacientes al ingresar al plan PANTBC.

Egresos_PANTBC

Permite determinar la cantidad de pacientes que han egresado del tratamiento de un determinado establecimiento de salud y en un tiempo indicado.

Alimentación_PANTBC

Permite determinar los movimientos y estados de los alimentos entregados. Se obtendrán las medidas por tiempo, por establecimiento de salud, por alimento.

Actividades_PANTBC

Permite determinar las actividades que realiza PANTBC con la finalidad de capacitar a los pacientes por tiempo, por establecimiento salud.

Facts vs. Dimensiones

En el cuadro a continuación se muestran las dimensiones que utiliza cada tabla de hechos indicada en el punto anterior

	Tabla de Hechos (Facts)					
Dimensiones	FT_Diagnostico	IMC_PANTBC	Atenciones_PANTBC	Egresos_PANTBC	Alimentación_PANTBC	Actividades_PANTBC
Actividad_Enfermedad	✓					
Establecimiento_Salud	✓	✓	✓		✓	✓
IMC		✓				
Tiempo	✓	✓	✓	✓	✓	✓
Mini_Paciente	✓					
Condicion_Egreso				✓		
Actividad_Capacitación						✓
Alimento					✓	

Cuadro 2.1 Facts vs Dimensiones

El análisis dimensional completo se encuentra en el anexo 2 (Análisis Dimensional)

3. Capítulo 3: Diseño

En este capítulo se tratará de la arquitectura planteada para realizar el proceso de extracción, también se explica el proceso de explotación. Se complementa el capítulo con ejemplos de los procesos mencionados.

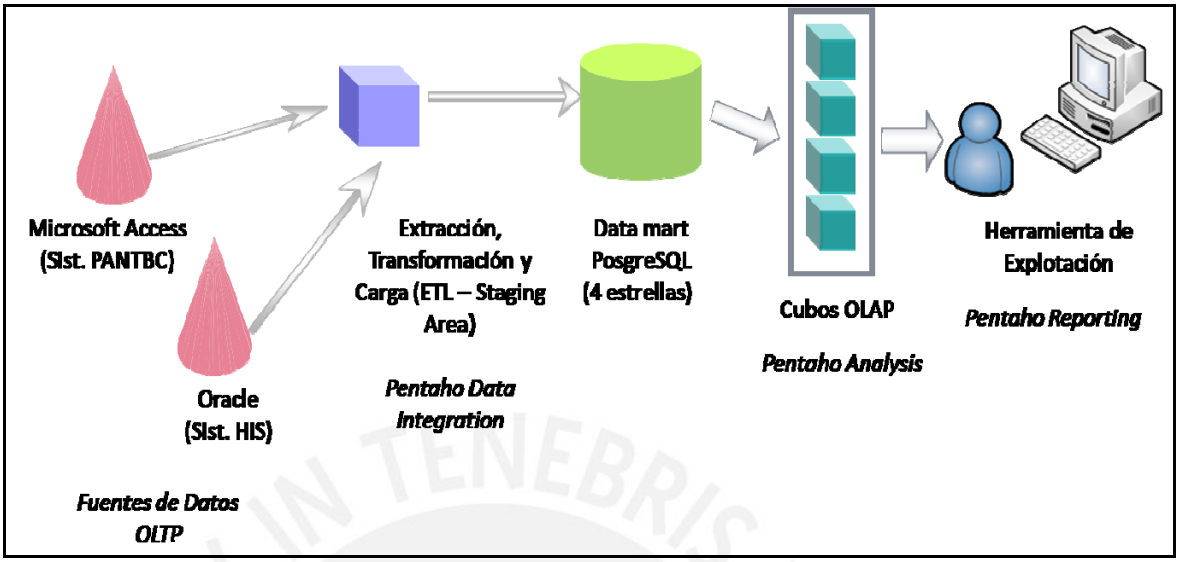
3.1. Arquitectura de la solución

En la figura 3.1 se muestra la arquitectura de la solución. En ella, la fuente de datos para el datamart serán las bases de datos transaccionales de los sistemas OLTP HIS - Oracle (instalado en cada establecimiento de salud) y PANTBC – Microsoft Access (utilizado en la estrategia de *Alimentación y Nutrición Saludable* de la DISA).

Para el proceso ETL se contará con un repositorio (Staging Area) en la cual se realizarán todas las operaciones y transformaciones que necesitan los datos previamente a ser almacenados al datamart, para ello se utilizará la herramienta Pentaho Data Integration (Kettle).

El datamart estará formado por 5 esquemas estrellas.
Como herramienta OLAP se empleará Pentaho Analysis (Mondrian) y WorkBench, donde se construirán los cubos.
Los usuarios tendrán acceso a los cubos mediante una herramienta de explotación,

Pentaho Reporting y/o Microsoft Office Excel.



Fuente: Elaboración propia

Figura 3.1 Arquitectura de extracción

3.2. Proceso de Extracción

Para la extracción se tomó en cuenta que la información será obtenida a partir de la base de datos del sistema OLTP PANTBC y de Oracle para el caso del sistema OLTP HIS. La base multidimensional que se utilizará será una base de datos en PostgreSQL.

Para realizar la carga de los datos se seguirá el siguiente proceso

- Carga directa de la fuente hacia una tabla temporal.
- Aplicar procesos de transformación y limpieza de datos sobre los registros en la tabla temporal.
- Cargar los registros de la tabla temporal hacia el destino en el datamart.
- Borrar la tabla temporal.

A continuación se muestra un ejemplo de carga de la dimensión ACTIVIDAD_ENFERMEDAD.

Actividad_Enfermedad

Descripción
Representa la carga de las tablas CAPITULO, GRUPO, CATEGORIA y CIE

provenientes del sistema OLTP HIS

Descripción de Tablas Fuentes

Tipo de Fuente	Nombre de Tabla	Descripción
OLTP HIS	CAPITULO	Contiene el número de capítulo y su descripción de acuerdo al libro CIE 10. El CIE 10 contiene 21 capítulos
OLTP HIS	GRUPO	Contiene los grupos de diagnóstico según la CIE, el orden es correlativo dentro de cada capítulo. Los grupos pertenecen a un capítulo respectivo
OLTP HIS	CATEGORIA	Contiene las categorías de diagnóstico según el CIE 10, están predefinidas y su código está formado de 3 caracteres. Las categorías pertenecen a un grupo respectivo
OLTP HIS	CIE	Contiene la lista de de enfermedades, diagnósticos o actividades establecidas por la Clasificación Internacional de Enfermedades (CIE). Pertenecen a una categoría respectiva.

Cuadro 3.1 Dimensión Actividad_Enfermedad – Descripción de Tablas Fuentes

Estandarización de Datos y Limpieza de Datos

Se indica el formato que tendrán los campos de la tabla final

Nombre	Llave	Tipo	Formato	Limpieza	Valor por Defecto
IdActEnfermedad	PK	Integer	Número correlativo	No debe ser nulo	Ninguno
Codigo		Varchar(7)	<A-Z><xy><z> A-Z: Valor entre A y Z XY: Numeración según categoría Z: Numeración correlativa	No debe ser nulo	Ninguno
Descripción		Varchar(255)	Texto		Ninguno
Categoría		Varchar(255)	Texto		Ninguno
Grupo		Varchar(255)	Texto		Ninguno
Capítulo		Varchar(255)	Texto		Ninguno

Cuadro 3.2 Dimensión Actividad_Enfermedad – Limpieza de Datos

Fuentes de Datos

Tabla:	CAPITULO			
Nombre	**Llave**	**Tipo**	**Formato**	**Consideración Importante**
COD_CAP	PK	Varchar(2)	Texto (00)	No debe ser nulo Tiene ceros a la izquierda

Tabla:	CAPITULO			
Nombre	**Llave**	**Tipo**	**Formato**	**Consideración Importante**
DESC_CAP		Varchar(255)	Texto	

Tabla:	GRUPO			
Nombre	**Llave**	**Tipo**	**Formato**	**Consideración Importante**
COD_GRU	PK	Varchar(2)	Texto (00)	No debe ser nulo Tiene ceros a la izquierda
COD_CAP	FK	Varchar(2)	Texto (00)	Tiene ceros a la izquierda
DESC_GRU		Varchar(255)	Texto	

Tabla:	CATEGORIA			
Nombre	**Llave**	**Tipo**	**Formato**	**Consideración Importante**
COD_CAT	PK	Varchar(3)	<A-Z><xy> A-Z: Valor entre A y Z XY: Numeración según categoría	No debe ser nulo
DESC_CAT		Varchar(255)	Texto	
COD_CAP	FK	Varchar(2)	Texto (00)	Tiene ceros a la izquierda
COD_GRU	FK	Varchar(2)	Texto (00)	Tiene ceros a la izquierda

Tabla:	CIE			
Nombre	**Llave**	**Tipo**	**Formato**	**Consideración Importante**
COD_CAT	FK	Varchar(3)	<A-Z><xy> A-Z: Valor entre A y Z XY: Numeración según categoría	
COD_ENF	PK	Varchar(13)	Texto	No debe ser nulo
DESCRIPCIO		Varchar(255)	Texto	

Cuadro 3.3 Dimensión Actividad_Enfermedad – Tablas Fuentes

Tabla Destino

Se indica descripción general del mapeo.

Tabla:	tmpActividad_Enfermedad	
Campo	**Tipo**	**Mapeo**
IdActDiagnostico	Integer	Autosecuencial
Codigo	Varchar(7)	Mayúscula(CIE.COD_ENF)
Descripción	Varchar(255)	Mayuscula(CIE.descripcio)
Categoría	Varchar(255)	Mayuscula(CATEGORIA.desc_cat)
Grupo	Varchar(255)	Mayuscula(GRUPO.desc_grupo)
Capítulo	Varchar(255)	Mayuscula(CAPITULO.desc_cap)

Cuadro 3.4 Dimensión Actividad_Enfermedad – Tabla Destino

Proceso

Se indica el proceso a seguir para la carga final de datos

1. Borrar Tablas Temporal

Eliminar la tabla temporal tmpActividad_Enfermedad

2. Cargar registros de la tablas CAPITULO, GRUPO, CATEGORIA, CIE

Se extrae el atributo DESCRIPCIO, COD_ENF de la tabla CIE y se cargan a la tabla temporal, también se extrae el atributo COD_CAT y se obtiene la descripción en el atributo DESC_CAT de la tabla CATEGORIA y se carga a la tabla temporal. Además se extrae el atributo COD_CAP y COD_GRU de la tabla CATEGORIA para obtener sus descripciones en las tablas CAPITULO y GRUPO respectivamente, estas son cargadas a la tabla temporal según el mapeo.

3. Carga de la Dimensión

Tomar los valores de la tabla temporal y llevarla a la dimensión Actividad_Enfermedad. En caso que sean nuevas actividades o diagnósticos estos deben ser insertados

A continuación se muestra un ejemplo de carga de la fact table ALIMENTACION_PANTBC

Alimentacion_PANTBC

Descripción

Representa la carga de las tablas MOVALIPANTC y LINMOVALI

Descripción de Tablas Fuentes

Tipo de Fuente	Nombre de Tabla	Descripción
OLTP - PANTBC	MOVALIPANTBC	Datos sobre el movimiento y manejo de alimentos
OLTP - PANTBC	LINMOVALI	Líneas de la tabla MOVALIPANTBC que incluye datos (cantidad) de un determinado alimento

Cuadro 3.5 Fact Table Alimentacion_PANTBC – Tablas Fuentes

Estandarización de Datos y Limpieza de Datos

Nombre	Llave	Tipo	Formato	Limpieza	Valor por Defecto
IdEstSalud	PK	Integer	Número	No debe ser nulo	Ninguno
IdTiempo		Integer	Número		Ninguno
IdAlimento		Integer	Número		Ninguno
CantDeteriorada		Decimal	Número		Ninguno
CantPerdida		Decimal	Número		Ninguno
CantRobada		Decimal	Número		Ninguno
CantEntregada		Decimal	Número		Ninguno
CantTotal		Decimal	Número		Ninguno
PorDeteriorada		Decimal	Número		Ninguno
PorPerdida		Decimal	Número		Ninguno
PorRobada		Decimal	Número		Ninguno
PorEntregada		Decimal	Número		Ninguno

Cuadro 3.6 Fact Table Alimentacion_PANTBC – Limpieza de Datos

Fuentes de Datos

Tabla:	MOVALIMENTO			
Nombre	**Llave**	**Tipo**	**Formato**	**Consid. Importante**
IDMOVALI	PK	Integer	Número correlativo	No puede ser nulo
COD_ESTAB		Varchar(9)	Texto	No puede ser nulo
ANO		Integer (0000)	Número	
MES		Integer (00)	Número	

Tabla:	LINMOVALI			
Nombre	**Llave**	**Tipo**	**Formato**	**Consideración Importante**
IDMOVALI	PK/FK	Integer	Número correlativo	No puede ser nulo
IDLINMOVALI	PK	Integer	Número correlativo	No puede ser nulo
ALIMENTO		Varchar(50)	Texto	No puede ser nulo
DETERIORADO		Decimal	Número	
PERDIDO		Decimal	Número	
ROBADO		Decimal	Número	
ENTREGADO		Decimal	Número	

Cuadro 3.7 Fact Table Alimentacion_PANTBC – Tablas Fuentes

Tabla Destino

Se indica descripción general del mapeo.

Tabla:	**tmpAlimentacion_PANTBC**	
Campo	**Tipo**	**Mapeo**
IdEstSalud	Integer	Se obtiene el atributo COD_ESTAB de la tabla EGRESOPANTBC y se hace un lookup a la dimension ESTABLECIMIENTO y de esta se obtiene el IdEstSalud
IdTiempo	Integer	Se obtienen atributos ANO, MES y con DIA=0 de la tabla EGRESOPANTBC y se hace un lookup con la dimensión TIEMPO y de esta se obtiene el IdTiempo.
IdAlimento	Integer	Se obtiene el atributo ALIMENTO de la tabla LINMOVALI, se hace un lookup con la dimensión ALIMENTO y de esta se obtiene el IdAlimento
CantDeteriorada	Integer	Según cálculo
CantPerdida	Integer	Según cálculo
CantRobada	Integer	Según cálculo
CantEntregada	Integer	Según cálculo
CantTotal	Decimal	Según cálculo
PorDeteriorada	Decimal	Según cálculo
PorPerdida	Decimal	Según cálculo
PorRobada	Decimal	Según cálculo
PorEntregada	Decimal	Según cálculo

Cuadro 3.8 Fact Table Alimentacion_PANTBC – Tabla Destino

Proceso

1. Cargar registros de tabla MOVALIMENTO, LINMOVALI

En el punto anterior se explicó como se obtendrán las llaves primarias de la fact table, estas serán cargadas a la tabla temporal tmpAlimentacion_PANTBC

Para hallar las medidas (CANTX) se realizará lo siguiente:

```
SELECT      SUM (LIN.DETERIORADO), SUM (LIN.PERDIDO),
            SUM(LIN.ROBADO),   SUM(ENTREGADO)
FROM MOVALIMENTO M
LEFT JOIN LINMOVALI LIN ON LIN. IDMOVALI = M. IDMOVALI
WHERE -- FILTROS
GROUP BY [A.MES], [A.ANO], [..]
```

Para hallar los porcentajes se realizará lo siguiente:

```
SELECT      (SUM(LIN.xxxx) /(SUM (LIN.DETERIORADO)+ SUM (LIN.PERDIDO)+
            SUM(LIN.ROBADO) +   SUM(ENTREGADO))*100
FROM MOVALIMENTO M
LEFT JOIN LINMOVALI LIN ON LIN. IDMOVALI = M. IDMOVALI
WHERE -- FILTROS
GROUP BY [A.MES], [A.ANO], [..]
```

Donde LIN.xxxx puede representar: DETERIORADO, PERDIDO, ROBADO, ENTREGADO, y luego esta cantidad es dividida por el total de alimento.

2. Carga de la Fact Table

Todos los datos cargados en la tabla temporal son llevados a la fact table ALIMENTACION_PANTBC.

Para un mayor detalle de la extracción, revisar el Anexo 3 (Diseño de Extracción).

3.3. Proceso de Explotación

En la explotación se muestran aquellos reportes que servirán de ayuda a las direcciones de salud para la toma de decisiones.

El ejemplo a continuación, trata sobre la evolución de actividades o enfermedades a través de tiempo, de esta forma se puede medir si el número de casos que se presentan aumentan o disminuyen, además se puede realizar el contraste entre los diferentes establecimientos de salud, red, microrred (dependiendo del filtro),

Diseño:

	Bocio difuso no tóxico – Casos											
Estab. De Salud	**Ene.**	**Feb.**	**Mar.**	**Abr.**	**May.**	**Jun.**	**Jul.**	**Ago.**	**Set.**	**Oct.**	**Nov.**	**Dic.**
E.S. Apoyo San Jose	100	122	120	112	106	103	90	98	103	115	99	87
E.S. Carmen de la Legua	23	30	21	35	25	26	40	30	32	36	37	32
E.S. Villa Sr. Milagros	2	4	7	0	0	1	2	8	4	3	2	0
E.S.La Mar	75	77	80	81	88	91	92	71	69	83	74	73

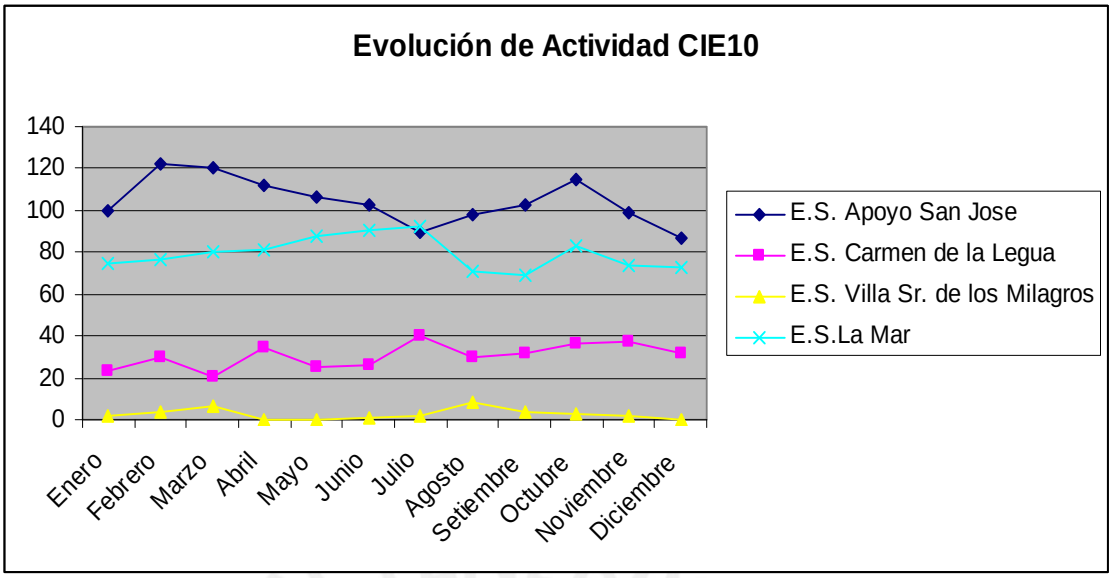


Figura 3.2 Ejemplo de Reporte – Diseño de Explotación

Tipo: Reporte tipo líneas.

Filas:

No.	Dimensión	Nivel / Categoría
1	Establecimiento_Salud	Establecimiento de Salud
2	Establecimiento_Salud	Microrred
3	Establecimiento_Salud	Red
4	Establecimiento_Salud	Distrito
5	Establecimiento_Salud	Provincia

Columnas:

No.	Dimensión	Nivel / Categoría
1	Actividad_Enfermedad	CIE10
2	Actividad_Enfermedad	Categoría
3	Actividad_Enfermedad	Grupo
4	Tiempo	Tiempo

Medida:

No.	Medida	Formato
1	NumDiagnosticos	Integer

Filtro:

Indica el filtro que va a tener el reporte

No.	Operación
1	Tiempo : Meses , Trimestres o Años
2	DISA= <Nombre de la DISA a ingresar>
3	Indicar Red o Redes a evaluar
4	Indicar Microrred o microrredes a evaluar
5	Indicar Establecimiento o establecimientos a evaluar
6	Indicar solo una Actividad_Enfermedad a evaluar

Para un mayor detalle del diseño y muestra de los reportes, consultar el anexo 4 (Diseño de Explotación).

4. Capítulo 4: Construcción

Este capítulo tratará sobre los pasos principales realizados para la configuración de las herramientas utilizadas así como de los resultados obtenidos durante la construcción del datamart.

4.1. Configuración del software

Las configuraciones que se muestran a continuación fueron realizadas en el sistema operativo Linux, distribución Ubuntu 7.10 Gutsy Gibbon. Sin embargo, únicamente la forma de instalar o ejecutar las aplicaciones a utilizar varía de acuerdo al sistema operativo, las configuraciones tanto de la base de datos y aplicaciones de Pentaho son independientes de la plataforma.

4.1.1. Configuración de la base de datos

Descarga e Instalación

Se utilizará PostgreSQL como motor de base de datos, tanto para las transaccionales, repositorio de ETL y dimensional.

Primero, se debe descargar e instalar PostgresSQL, para ello se realiza los siguientes pasos:

Instalar los siguientes paquetes:

- postgresql-client-8.2 (parte cliente)
- postgresql-8.2 (parte servidor)

- pgadmin3 (cliente gráfico para una fácil interacción con el servidor, para el proyecto se utilizo pgAdminIII)

En la aplicación gráfica pgadmin3 se configurarán los siguientes datos:

Nombre	:	Se ingresa el nombre del servidor
Servidor	:	Se ingresa la dirección ip o url del servidor, para este caso se utiliza de forma local (localhost)
Puerto, BD de Mantenimiento	:	Son valores por defecto, puerto 5432
Nombre y contraseña de usuario	:	Se ingresan los datos del usuario creados anteriormente.

Finalmente se tiene todo correctamente configurado y se continúa con el proceso de creación de base de datos pero de modo gráfico o ejecutando scripts de creación.

4.1.2. Configuración de Pentaho

Primero se procederá a descargar todos los aplicativos (incluido la plataforma) necesarios desde la sección de descarga de la página oficial de Pentaho, según el sistema operativo en que se va a trabajar.
Se muestran todas las configuraciones para cada uno de los productos de Pentaho.

Plataforma Pentaho

Se instalará la plataforma Pentaho, la versión demo ya que la solución se encuentra pre-configurada. Se configurarán algunos archivos principales para poder personalizar nuestra solución.
El paquete a instalar es: pentaho_demo_mysql5-1.6.0.GA.863.tar.gz para Linux, este se descomprime en una ubicación fácil de recordar. Se debe agregar el driver necesario para la conexión a la base de datos postgres.

Desplegar solución

Para ver la solución, se abre un navegador de preferencia y se ingresa la dirección de la siguiente forma http://<servidor>:8080/pentaho/
El servidor que se está utilizando es local y el puerto de conexión es 8080 por tanto la dirección queda así: http://localhost:8080/pentaho/

Configuración de archivos

- Archivo: myfirst-ds.xml
- Archivo web.xml
- Archivo: jboss-web.xml

La configuración de archivos puede verse en el anexo 5 (Construcción).

Kettle (Data Integration)

Kettle o Pentaho Data Integration es el aplicativo para realizar el proceso de ETL (Extracción, Transformación y Carga).

Una vez que haya sido descargado Kettle se descomprime el archivo, en la carpeta descomprimida se encuentran una serie de ejecutables (.exe para Windows y .sh para Linux), para realizar el proceso de ETL únicamente se utilizará Spoon.
Desde una consola de Linux se ejecuta el comando: `$sudo ./spoon.sh`

Se muestra la siguiente ventana:

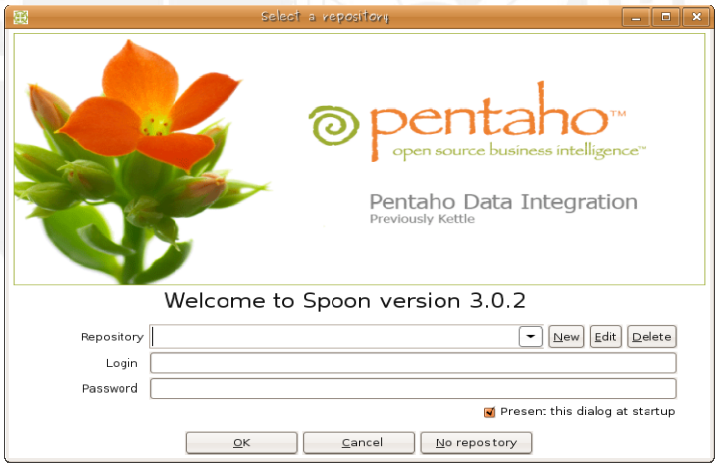


Figura 4.3. Kettle – Pantalla Inicio

En ella se definirá el repositorio en el cual se guardan todos los objetos de Kettle, para esto se da clic al botón “New” y aparecerá una ventana de diálogo que solicita información de la conexión.

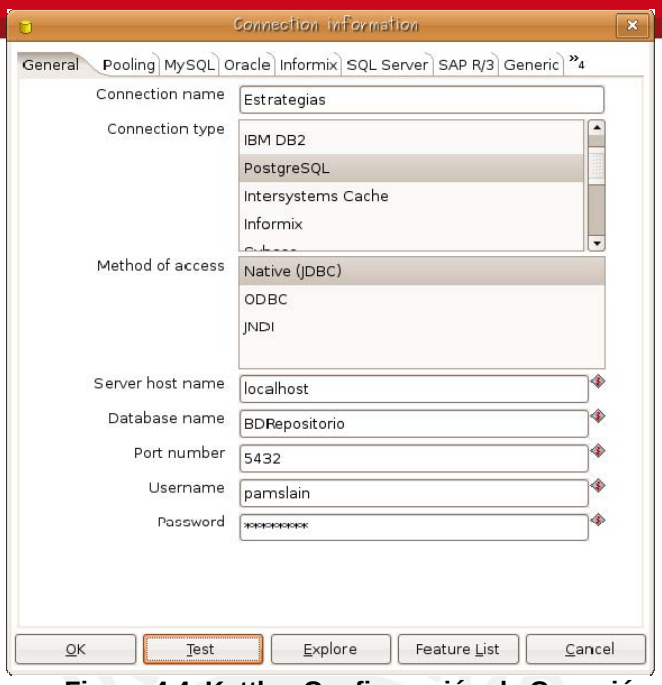


Figura 4.4. Kettle - Configuración de Conexión

Nombre : Se ingresa el nombre del servidor

Servidor : Se ingresa la dirección ip o url del servidor, para este caso se utiliza de forma local (localhost)

Puerto, BD de Mantenimiento : Son valores por defecto, puerto 5432

Nombre y contraseña de usuario : Se ingresan los datos del usuario creados anteriormente.

Cube Designer (Analysis)

Con Cube Designer se crearán los cubos OLAP representados por ficheros de configuración XML llamado también fichero del esquema Cubo Mondrian (Mondrian Cube Scheme), en él estarán definidos las dimensiones, hechos y conexión a la base de datos que sirve para el cubo OLAP.

Antes de configurar el Cube Designer, se debe establecer la contraseña de publicación en el portal Pentaho, para ello se debe modificar el archivo publisher_config.xml ubicado en *pentaho-demo/jboss/server/default/deploy*.
En medio de las etiquetas <publisher-password></publisher-password> se ingresa la nueva contraseña.

Ahora, una vez que ha sido descargado correctamente la aplicación Cube Designer, se debe ejecutar de la siguiente manera.

Abrir una consola y colocarse en la ubicación del Cube Designer, luego ingresar

```
$sudo ./start.sh
```

Aparecerá la siguiente ventana, la cual es el primer paso para la creación del cubo. Para esta sección, únicamente se explicará la configuración de la fuente

Se selecciona el botón “Add” (Añadir) de la ventana y aparecerá una ventana de diálogo solicitando datos de la conexión JNDI.

Figura 4.5. Cube Designer – Descripción del cubo

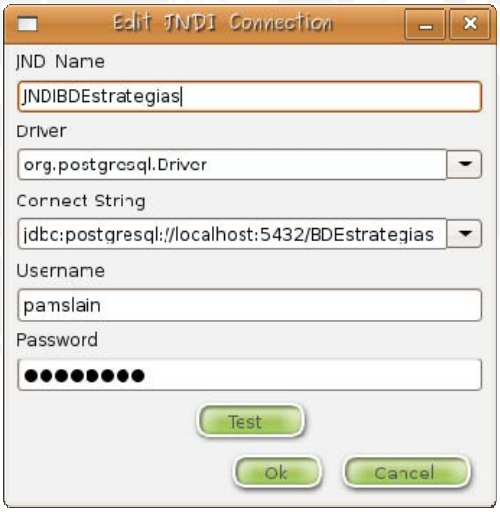


Figura 4.6. Cube Designer – Conexión JNDI

JNDI Name	:	Nombre JNDI, descripción
Driver	:	Driver utilizado para la base de datos fuente
Connect String	:	Cadena de conexión a base de datos
Username	:	Nombre de usuario de base de datos
Password	:	Contraseña de base de datos

Se prueba la conexión presionando el botón “Test”, si se obtiene el mensaje de éxito se podrá continuar con los siguientes pasos.

4.2. Construcción de procesos de carga

Ya teniendo el aplicativo Kettle – Spoon debidamente configurado, se procederá a realizar los procesos de carga de las dimensiones y tabla de hechos.

Para poder realizar la carga se crea un “transformation” (transformación) el cual está formado por “steps” (pasos)
Un “step” puede brindar un amplio rango de funcionalidad desde leer de archivos de texto hasta implementar cambios en las dimensiones.

Añadir *steps*

Para añadir “steps” al “transformation” simplemente se arrastran de la barra lateral y se colocan en el área de trabajo, luego se deben configurar y organizar.
Las relaciones entre “steps” se denominan “hops”.

Ejecutar *transformation*

Cuando se ha finalizado la modificación al “transformation” se debe ejecutar con el botón “Run o Start” del menú principal.
Aparecerá un registro con los pasos realizados y sus estados, además de presentarse errores se podrán verificar también en el registro.

4.2.1. Carga de dimensión: ACTIVIDAD_ENFERMEDAD

Tablas utilizadas: CAPITULO, GRUPO, CATEGORIA, CIE del sistema OLTP HIS*Tabla destino:* ACTIVIDAD_ENFERMEDAD de la base de datos dimensional.
Descripción:
En la figura 4.7 se muestra el proceso de carga. Se obtienen de las tablas CAPITULO, GRUPO, CATEGORIA y CIE sus campos de descripción y el código CIE10 (código internacional de enfermedad), luego se genera el código identificador para la tabla dimensional ACTIVIDAD_ENFERMEDAD. Antes de insertar estos valores, se debe realizar un filtrado para evitar que los campos sean nulos.

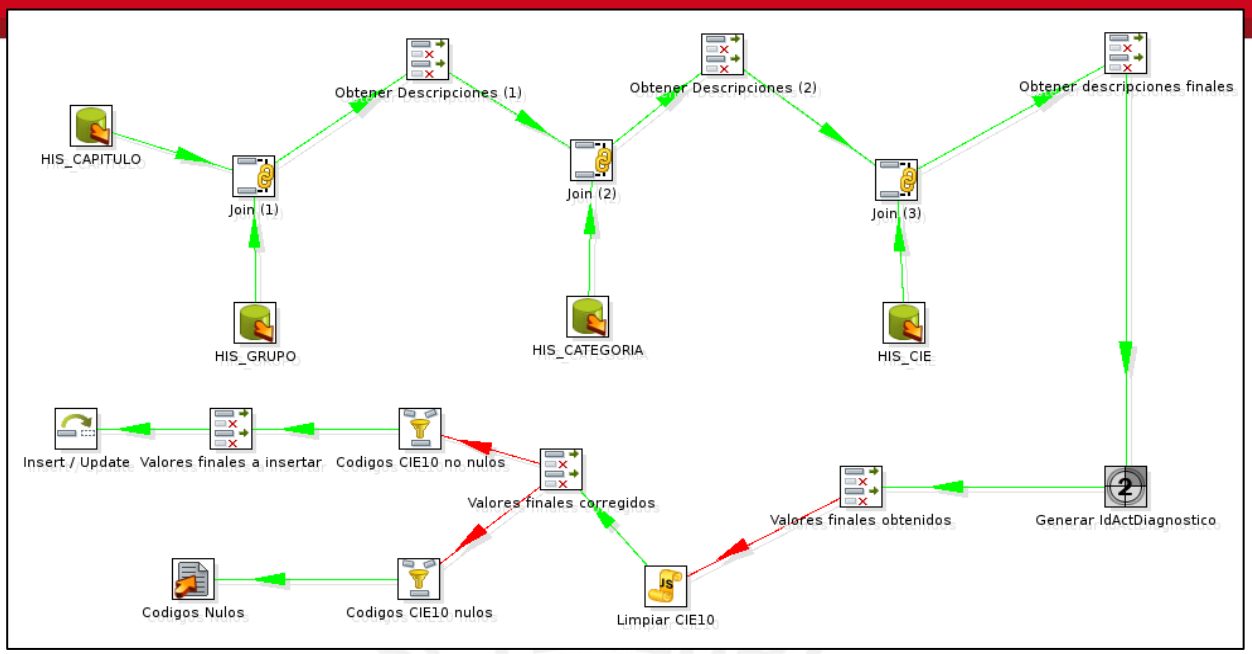


Figura 4.7. Kettle – Carga dimensión ACTIVIDAD_ENFERMEDAD

La carga de las dimensiones y fact tables se encuentra en el anexo 5 (Construcción).

4.3. Construcción de reportes

4.3.1. Reporte: Prevalencia de Enfermedades

Descripción del Reporte

Este reporte muestra la prevalencia de enfermedades, es decir el número de casos de diagnósticos detectados así como el porcentaje que representa dentro de un determinado establecimiento de salud.

Elección de dimensiones y tabla de hechos

Para este reporte se colocaron en el área de trabajo:

Dimensiones

- MINI_PACIENTE
- ACTIVIDAD_ENFERMEDAD
- TIEMPO
- ESTABLECIMIENTO_SALUD

Tabla de hechos

- FT_DIAGNÓSTICO

Se seleccionan los atributos que se desean mostrar en el reporte, marcando los checkboxs de las dimensiones y tabla de hecho.

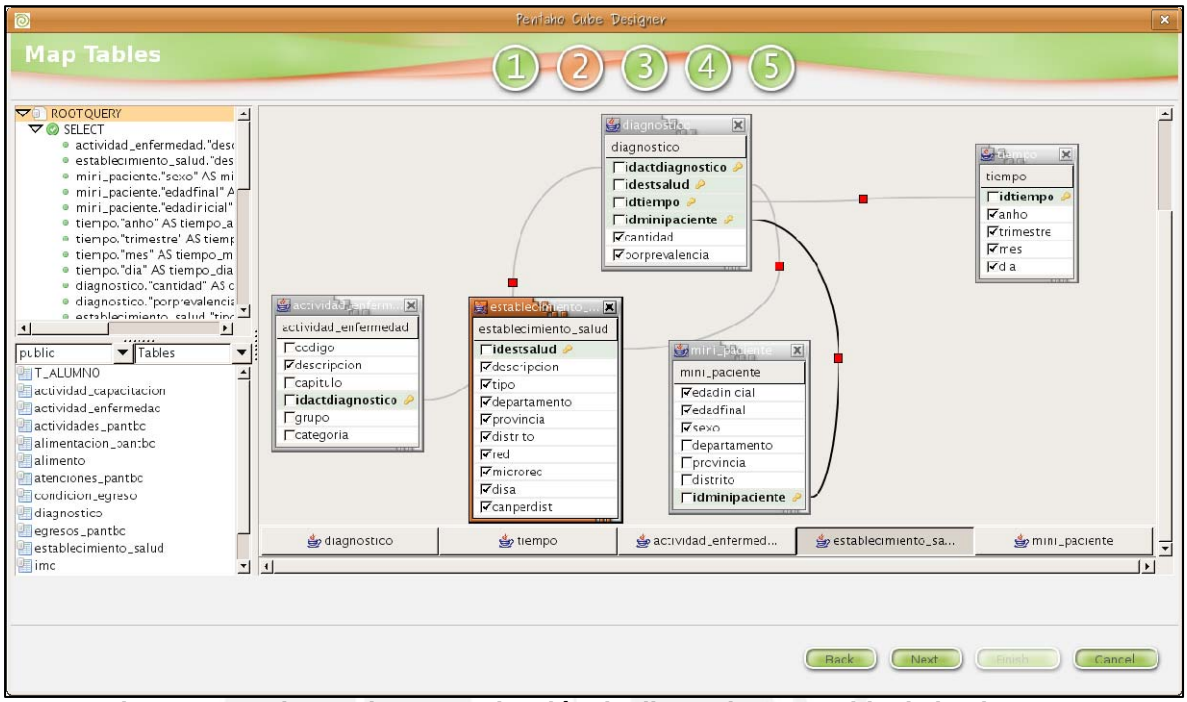


Figura 4.8. Cube Designer – Selección de dimensiones y tabla de hecho

Creación de las medidas

Las dos medidas que se evaluarán en el reporte son:

- Cantidad: Indica la cantidad de casos atendidos por enfermedad.
- Prevalencia: Indica en porcentaje los casos atendidos por enfermedad.

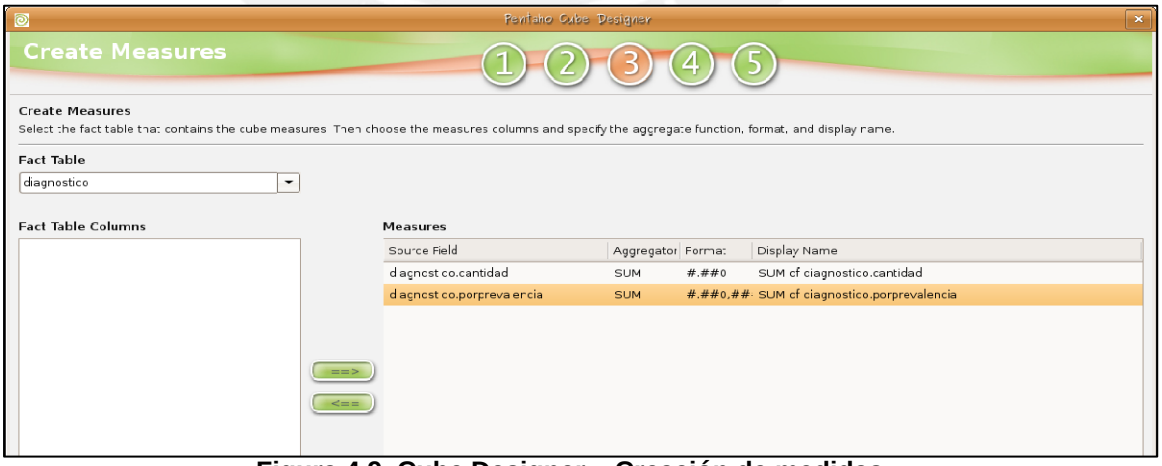


Figura 4.9. Cube Designer – Creación de medidas

Creación de dimensiones

Se debe crear y configurar las dimensiones (jerarquías) en el cubo multidimensional OLAP.

Publicación de Cubo

Se deben colocar los datos de publicación en el portal Pentaho.

Publish Password: Contraseña establecida en archivo de publicación del portal Pentaho.

Server Userid, Server Password: Datos del usuario por defecto del portal Pentaho.

Además se genera el archivo XML del cubo creado, este debe ser guardado en caso se desee modificar luego.

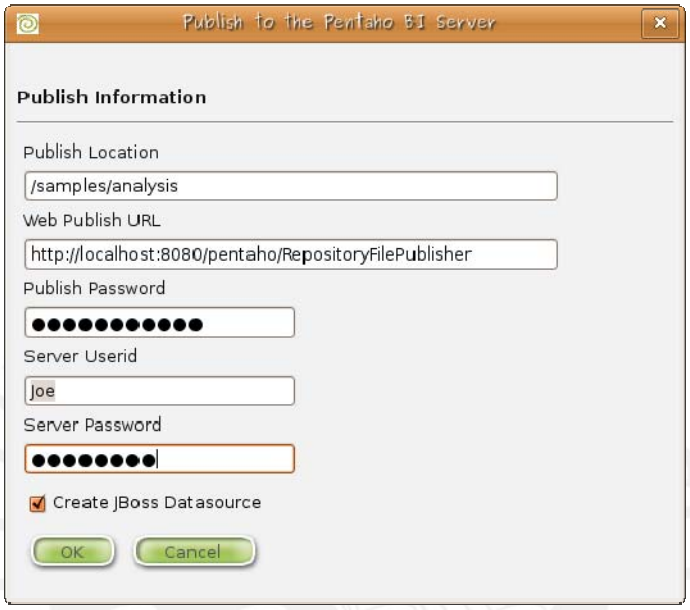


Figura 4.10. Cube Designer – Publicación de Cubo

4.4. Ejecución de pruebas de proceso de carga

Todas las pruebas que se realizaron fueron de forma local utilizando en su mayor parte datos reales y actualizados que se encuentran de libre disposición en el portal del MINSA con dichos datos se llenaron las tablas de los sistemas transaccionales. Los datos sensibles como diagnósticos o información de pacientes fueron generados por fines prácticos.

Se muestra la ejecución del “transformation” mencionado anteriormente, en ella se observa las actividades que ha realizado cada “step” como: Copiado, Lectura, Escritura, Datos de Ingreso, Datos de Salida, Actualizaciones, Rechazos y Errores.

El “transformation” ha leído de las tablas fuentes del sistema OLTP HIS, las ha enlazado adecuadamente y finalmente las ha insertado o actualizado en la tabla destino.

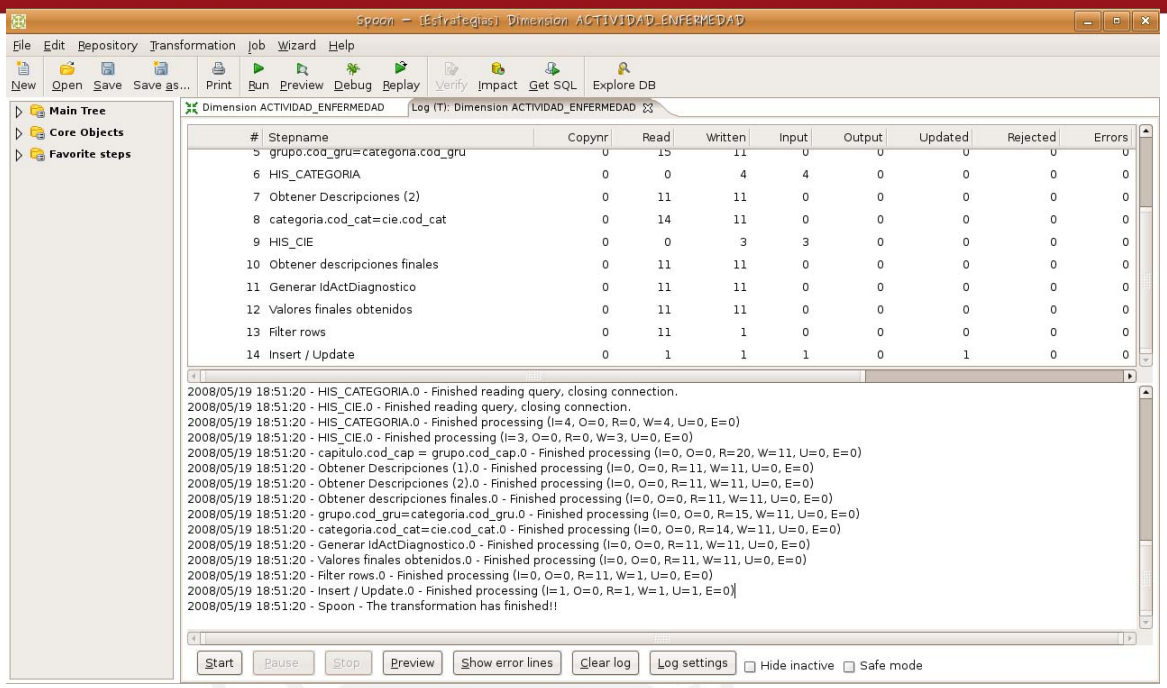


Figura 4.11. Kettle – Ejecución de carga

El resultado de la ejecución es el ingreso o actualización de los datos en la tabla ACTIVIDAD_ENFERMEDAD de la base de datos dimensional.

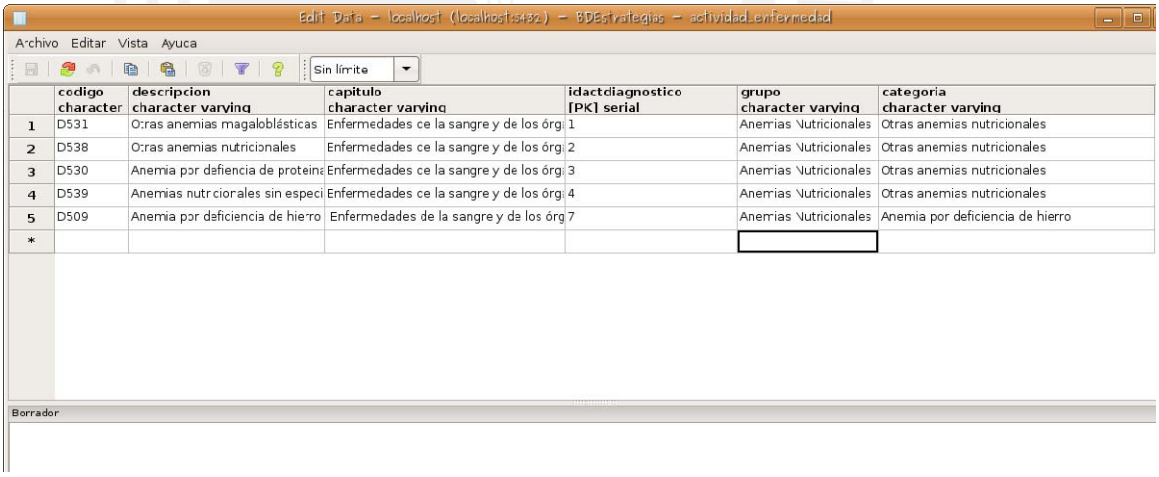


Figura 4.12. Kettle – Resultado de ejecución

4.5. Ejecución de reportes

Luego de haber indicado las dimensiones, tabla de hechos, medidas y publicado el reporte, se puede ver los resultados ingresando al portal de la plataforma Pentaho en la sección de “Soluciones” -> “Analisis” -> “Prev_Enf” (nombre establecido en el cubo).

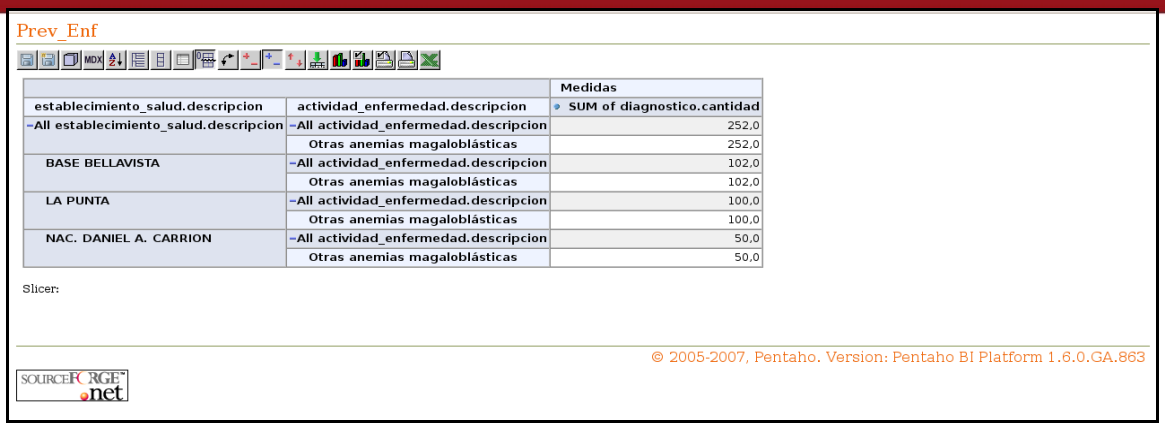


Figura 4.13. Reporte de Prevalencia de Enfermedades (1)

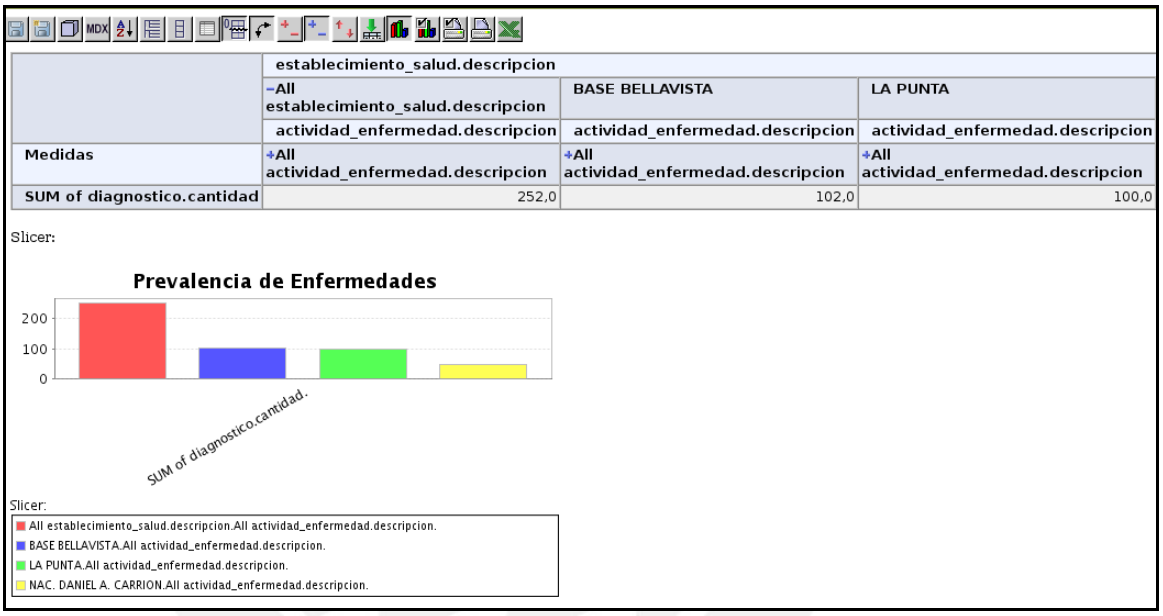


Figura 4.14. Reporte de Prevalencia de Enfermedades (2)

En la figura 4.14 se aprecia la prevalencia de la enfermedad (número de diagnósticos): “Otras Anemias megaloblásticas” en 3 establecimientos de salud.

En el anexo 5 (construcción) se encuentran todos los reportes generados por las herramientas de Pentaho así como las pruebas realizadas.

5. Capitulo 5: Observaciones, conclusiones y recomendaciones

En este último capítulo se tratará sobre las observaciones del proyecto, conclusiones y recomendaciones para mejorar el proyecto realizado.

5.1. Observaciones

Para la captura de requerimientos y necesidades se solicitó la planificación de entrevistas en la DISA del CALLAO (LIMA I) con los encargados de la estrategia de Alimentación y Nutrición Balanceada, así se obtuvieron datos reales para iniciar el análisis y diseño del datamart que permita facilitar la generación de reportes que ellos requieren.

En la construcción del datamart se utilizó Linux como sistema operativo (Ubuntu 7.10, distribución libre) así como herramientas libres (productos de Pentaho). El aprendizaje en el manejo de tales herramientas demandó tiempo, pues a pesar de existir documentación oficial, aún son relativamente nuevas y por tanto no había información necesaria respecto a problemas específicos que podrían presentarse en su uso.

Una observación importante es que, sobretodo, se aprendió a plantear un proyecto de inteligencia de negocios, pues se aplicó la teoría aprendida en clases a través

del uso de las herramientas.

5.2. Conclusiones

El objetivo principal de este proyecto de fin de carrera es el desarrollo del datamart para la estrategia de Alimentación y Nutrición Balanceada, el cual sería empleado en las diferentes Direcciones de Salud (DISA) de los gobiernos regionales del país. Al finalizar este proyecto, se obtiene la implementación del datamart que cumple con los requerimientos establecidos.

La generación de reportes por parte del datamart para cada estrategia sanitaria permite el ahorro de tiempo, pues actualmente cada estrategia debe solicitar a un área central (área de estadística) lo cual implica sobrecarga en dicha área.
Además los reportes que dicha área entrega son “estáticos” y en un formato definido, en cambio con el datamart los involucrados podrán colocar los filtros deseados y realizar cambios como modificación en el nivel de detalle, selección de determinadas dimensiones, límite de valores, entre otros.

El uso de herramientas libres, a pesar de su curva de aprendizaje durante la implementación, en modo usuario es de fácil interacción. Otra ventaja es que no hay una inversión fuerte de dinero pues cuenta con licencias libres, de esta forma el dinero se invierte en las necesidades más urgentes que son las soluciones a los problemas de salud. Como se menciono en el análisis, el gasto que la DISA podría incurrir seria en soporte técnico de Pentaho durante la fase de uso del sistema.

5.3. Recomendaciones y trabajos futuros

Se recomienda la realización de un mayor número de pruebas para lograr una mayor confiabilidad de la información que se obtiene, así como un estudio más a fondo de todas las funciones que brinda Pentaho para lograr un mejor análisis de los reportes generados.

Los cubos OLAP creados, en formato XML, durante el proyecto con la herramienta Cube Designer de Pentaho podrían optimizarse, como por ejemplo permitir mayor flexibilidad u obtener cálculos más complejos, con la herramienta libre llamada

Schema WorkBench pues esta brinda mayores opciones frente al Cube Designer.

Como trabajos futuros, queda finalizar el análisis, diseño y construcción del datamart con las 9 estrategias sanitarias nacionales restantes basándose en la estrategia piloto. Para ello deberá realizarse necesariamente, al igual que en el proyecto, entrevistas con los involucrados, sin embargo ya teniendo una base se logrará estandarizar el manejo de la toma de decisiones de las estrategias.

También, sería aconsejable mejorar la apariencia física (interfaz gráfica) del portal web de la solución Pentaho, traducción de las opciones ya sea a través de configuración de archivos (pues originalmente se encuentra en inglés) y establecer una debida clasificación o división por estrategia dentro del portal.

Bibliografía

1. [BIB01] Ralph Kimball, “The Data Warehouse Toolkit”, 2002.
2. [BIB02] Elizabeth Vitt, Michael Luckevich y Stacia Misner, “Business Intelligence Técnicas de análisis para la toma de decisiones estratégicas”, España 2002
3. [BIB03] William H. Inmon, “Building the Data Warehouse”, 2005
4. [ART01] Plan Nacional Concertado de Salud 2007
5. [ART02] The Olap Survey 6
6. [ART03] Memoria 2001-2006 MINSA
7. [CLA01] Diapositivas del Curso de Inteligencia de Negocios 2007-2
8. [TES02] Ivan Tapia, María Ruiz Rivera, Edgar Ruiz, “Una metodología para sectorizar pacientes en el consume de medicamentos aplicando datamart y datamining en un hospital”
9. [TES01] Morote Ríos Christian, Tesis: Análisis, diseño, implementación e impacto estratégico de un datamart para el área de ventas de empresas del sector farmacéutico
10. [URL01]http://sisbib.unmsm.edu.pe/BVRevistas/situa/2001_n17/ventajas.htm
11. [URL02]www.TodoBI.com
12. [URL03]www.Stratebi.com
13. [URL04]www.Pentaho.com *Pagina Oficial de Pentaho*
14. [URL05]www.minsa.gob.pe *Ministerio de Salud del Peru*
15. [URL06] http://www.minsa.gob.pe/portal/03esn/default2.asp *Estrategias Sanitarias*
16. [URL07] http://www.minsa.gob.pe/portalminsa/directorioinstitucional/default.asp *Direcciones Regionales*
17. [URL08]http://www.oracle.com/technology/products/warehouse/index.html
18. [URL09]http://pentaho.almacen-datos.com/
19. [URL10]http://www306.ibm.com/software/data/integration/datastage/features.html?S_CMP=wspace
20. [URL11]http://www.oracle.com/technology/products/discoverer/htdocs/oraclebi_discoverer_1012_fov.htm#porta
21. [URL12] http://plan-watch.com/ *Plan Watch*
22. [URL13] http://www.medvantage.com/aboutus.htm *MedVentage*
23. [ART04] Getting Started with the BI Platform
24. [ART05] Pentaho Cube Designer Guide
25. [ART06] Spoon User Guide

26. [ART07] Manual General de HIS

27. [URL14] http://www.adrformacion.com/cursos/javaser/leccion3/tutorial2.html

28. [URL15] http://www.huihoo.org/jboss/online_manual/3.0/ch07s22.html

29. [URL16] JNDI: http://programacion.com/tutorial/jndi/6/

30. [URL17] Ministerio de Salud - Guía Nacional de Operativización del Modelo de Atención Integral de Salud http://www.cimfweb.org/bn_admin/include/images/pdf/goapsperu.pdf

31. [URL18] Analysis Services http://technet.microsoft.com/es-es/library/ms166350.aspx

32. [URL19] Plataforma Pentaho Open Source – Business Intelligence http://egkafati.bligoo.com/content/view/219538/La_plataforma_Pentaho_Open_Source_Business_Intelligence.html

33. [URL20] MSDN SQL Server Integration Services http://msdn.microsoft.com/en-us/library/ms141026.aspx

34. [URL21] Business Objects OLAP Intelligence http://www.businessobjects.com/global/pdf/products/queryanalysis/olap_intelligence.pdf