

UNIVERSIDAD NACIONAL DEL ALTIPLANO - PUNO
FACULTAD DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA
ESCUELA PROFESIONAL DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA



**IMPLEMENTACIÓN DE DATA WAREHOUSE PARA SOPORTE
DE TOMA DE DECISIONES DE LA RED DE SALUD PUNO -
2013**

TESIS

PRESENTADA POR:

JAVIER ANGEL AYUNTA MAQUERA

PARA OPTAR EL TÍTULO PROFESIONAL DE:

INGENIERO ESTADÍSTICO E INFORMÁTICO

PUNO - PERÚ

2014



UNIVERSIDAD NACIONAL DEL ALTIPLANO - PUNO
FACULTAD DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA
ESCUELA PROFESIONAL DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA



IMPLEMENTACIÓN DE DATA WAREHOUSE PARA SOPORTE DE TOMA DE DECISIONES DE LA RED DE SALUD PUNO - 2013

TESIS


Presentada por:

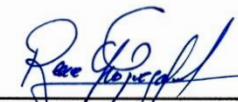
Bach. JAVIER ANGEL AYUNTA MAQUERA

A LA COORDINACIÓN DE INVESTIGACIÓN DE LA FACULTAD DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA DE LA UNIVERSIDAD NACIONAL DEL ALTIPLANO – PUNO, PARA OPTAR EL TÍTULO PROFESIONAL DE:


INGENIERO ESTADÍSTICO E INFORMÁTICO

APROBADA POR:

PRESIDENTE DEL JURADO : 
M.Sc. Alejandro Apaza Tarqui

PRIMER MIEMBRO : 
M.Sc. Remo Choquejahu Acero

SEGUNDO MIEMBRO : 
M.Sc. Freddy H. Willasante Saravia

DIRECTOR DE TESIS : 
M.Sc. Santos O. Morillos Valderrama

ASESOR DE TESIS : 
M.Sc. Confesor M. Vargas Valverde

2014

ÁREA: Informática
TEMA: Base de datos avanzados

DEDICATORIA

*A la memoria de mi Padre Gerónimo
Ayunta, a mi Madre María Maquera
y a mi querida hermana por su inagotable
apoyo espiritual y material que me dieron
durante toda mi formación profesional.*

AGRADECIMIENTOS

A todos mis familiares y amigos por sus orientaciones y generosidades. A mi escuela profesional de Ing. Estadística e Informática y a sus docentes por haberme formado en esta escuela.

ÍNDICE**DEDICATORIA AGRADECIMIENTOS INDICE****RESUMEN****ABSTRACT****INTRODUCCIÓN****CAPÍTULO I****PLAN DE INVESTIGACIÓN**

1.1	EI PROBLEMA	1
1.2	OBJETIVOS	2
1.3	HIPOTESIS	2
1.4	JUSTIFICACIÓN Y DELIMITACIÓN DE INVESTIGACIÓN.....	3

CAPÍTULO II**MARCO TEÓRICO**

2.1	ANTECEDENTES DE LA INVESTIGACIÓN	4
2.2	BASE TEORICA	8
2.3	DATA WAREHOUSE	30
2.4	SISTEMAS OPERACIONALES.....	42
2.5	DIMENSIONES	43
2.6	ETL (EXTRACT TRANSFORM LOAD).....	43
2.7	METADATA	44
2.8	MEDIDAS O MÉTRICAS	44
2.9	JERARQUÍAS DE DIMENSIONES Y NIVELES	45
2.10	MINERÍA DE DATOS	45

2.11	DATA WAREHOUSE Y DATA MINING.....	46
2.12	SISTEMA TRANSACCIONAL	46
2.13	SISTEMA DE GESTIÓN DE BASE DE DATOS	46
2.14	SQL (STRUCTURED QUERY LANGUAGE).....	47
2.15	OPERACIONALIZACION DE VARIABLES	47

CAPITULO III

MATERIALES Y MÉTODOS

3.1	POBLACIÓN.....	48
3.2	MUESTRA	48
3.3	MÉTODOS DE RECOPIACIÓN DE DATOS	48
3.4	MÉTODO DE TRATAMIENTO DE DATOS.....	50
3.5	MATERIAL EXPERIMENTAL	50
3.6	METODOLOGÍA DEL DESARROLLO	51

CAPITULO IV

RESULTADOS Y DISCUSIONES

4.1	ANÁLISIS Y COMPRENSIÓN DEL ENTORNO ANALÍTICO	70
4.2	PROTOCOLO DE ACCESO AL DATA WAREHOUSE	77
4.3	EVALUACIÓN DEL DATA WAREHOUSE.....	78
4.4	ANÁLISIS E INTERPRETACIÓN DE PRUEBA DE HIPÓTESIS.....	80
	CONCLUSIONES	82



SUGERENCIAS	83
BIBLIOGRAFÍA	84
ANEXOS	87

RESUMEN

Hoy en día las instituciones públicas y privadas utilizan distintos tipos de tecnología, especialmente el software para optimizar sus actividades, la información es muy importante ya que con esta se logra tomar decisiones que afectan de manera significativa dentro de la organización, actualmente hay muchos datos que aportan información en distintas instituciones, tal es el caso de la Red de Salud Puno que cuenta con un volumen de datos significativo, estos datos son de las atenciones que se brindan a la población que se encuentra en la jurisdicción de la Red de Salud Puno, por tal motivo se implementa el Data Warehouse que ayudara para la toma de decisiones dentro de la Red de Salud Puno. En este proyecto se plantea resolver los problemas de falta de información en tiempo real, y a si mismo diseñar un modelo que ayude en la toma de decisiones, en esta parte de esta investigación se trabajó con datos del Health Information System "HIS" que es actualmente el único sistema de recolección de datos de salud y así mismo se analizó los datos y sus tablas para así hacer el modelamiento de los datos. Para hacer esta implementación del Data Warehouse se aplico metodologías de desarrollo del Data Warehouse así como la herramienta de SQL Server 2012 con sus extensiones Data Tools y Analysis Services, también se optó a Microsoft Excel para hacer los reportes y gráficos de los datos que analizaremos según el requerimiento.

Palabra clave: Data Warehouse, HIS, red, decisiones, herramienta.

ABSTRACT

Today public and private institutions use different types of technology, especially software to optimize their activities, information is very important as this is achieved by making decisions that affect significant way within the organization, there are currently many data provide information in different institutions, as in the case of Red de Salud Puno has a significant volume of data, these data are of care that are provided to the population that is in the jurisdiction of Red de Salud Puno, for this reason the Data Warehouse to help for decision making within the Red de Salud Puno is implemented. In this project we plan to solve the problems of lack of information in real time, and himself to design a model to assist in decision making in this part of the research work was done with data from Health Information System "HIS" which is currently the unique collection of health data and the data itself and its tables are analyzed in order to make the modeling of the data. To make this implementation of the Data Warehouse we use methodology of Data Warehouse and SQL Server 2012 with its extensions Data Analysis Tools and Services, also opted for Microsoft Excel to make reports and charts to analyze data according to the requirement that this is acquired.

Keyword: Data Warehouse, HIS, Red, decisions, tool.

INTRODUCCIÓN

La información es realmente importante, desde muchos años atrás se ha ido recolectando mucha información dentro de las entidades, las entidades de Salud del sector privado y público necesitan de sistematizar información para agilizar sus labores y a si encaminarse en sus objetivos, en estos tiempos lo que se requiere es minimizar las labores obteniendo una buena producción, es por eso que un Data Warehouse es importante dentro de una entidad pública o privada ya que con este sistema se podrá agilizar y disminuir el tiempo de respuesta de las interrogantes que se le haga para una toma de decisiones.

Los datos almacenados por la entidad carecen de información si no se tienen los datos adecuados que sirva como base para tomar decisiones. Es por ello que es necesario que los datos históricos sean sometidos a un proceso de limpieza para garantizar su confiabilidad. Por eso se tiene que tener cuidado al momento de trabajar con los datos ya que un minúsculo error puede cambiar una decisión y esta de manera negativa a la entidad pública.

La Red de salud puno cuenta con información de atenciones, esta información se sistematiza dentro de un sistema (HIS) más aun, no se procesa de manera adecuada dicha información, teniendo este precedente importante para la toma de decisiones, se logra hacer una implementación de Data Warehouse en beneficio de la Red de Salud Puno. Es importante mencionar que los datos vienen de distintas fuentes, tablas, atributos etc. Simplemente a estos datos se le realiza las respectivas transformaciones para tener un buen resultado en nuestro Data Warehouse.

En el Capítulo I se da a conocer el problema que tiene la Red de Salud Puno, esto nos plantea los objetivos y justifica la investigación del Data Warehouse, siendo muy importante el cumplir con los objetivos dentro de esta investigación. El Capítulo II hace referencia al marco teórico que se presenta dentro de esta investigación, conceptos relacionados a Data Warehouse y al lugar donde se aplica la investigación, también hace referencia a aplicaciones de Data Warehouse en distintas áreas. En el Capítulo III se ve la metodología Dimensional Modeling (Modelado Dimensional) que se siguió para el Desarrollo del Data Warehouse así también como el alcance, los tratamientos de datos que se hizo, la población y el objeto en estudio que se plantea. En el Capítulo IV se discute los resultados obtenidos al implementar el Data Warehouse así como las evaluaciones que se realizó, seguido de las nuestras conclusiones y sugerencias que hace referencia que el sistema cumplió con nuestros objetivos planteados.

CAPÍTULO I

PLAN DE INVESTIGACIÓN

1.1 EI PROBLEMA

1.1.1 Formulación y Definición del Problema

Actualmente la Red de Salud – Puno cuenta con 11 Micro Redes a su cargo, estas micro Redes brindan servicios de atención básica en Salud a los pobladores de su jurisdicción, por lo cual uno de los problemas detectados es la falta de información en tiempo real , detallada y graficada con el que actualmente no cuenta el personal encargado de organizar campañas y solamente tienen acceso a esa información el personal de estadística quienes utilizan el “Health Información System (HIS)”, siendo de vital importancia la información contenida en las bases de datos de este sistema para la toma de decisiones.

La falta de organización en la información provoca cierta incertidumbre y una inapropiada toma de decisiones por parte del personal directivo, así como para los encargados de las estrategias, este problema se ha ido agravando conforme a los informes presentados por la oficina de Estadística e Informática y las diferentes estrategias con la que cuenta la Red de Salud Puno, al no coincidir la información este problema causó malestar en las evaluaciones semestrales que se realizan dentro de la institución pues genera disconformidad entre el personal

y la dirección.

Los reportes que se generan son en texto plano y no se puede modificar, esto hace que solamente el personal de estadística pueda trabajar con esa información, Teniendo en cuenta estas necesidades y expectativas se formula la siguiente interrogante.

¿De qué manera mejorará los reportes al implementar el Data Warehouse de soporte para toma de decisiones en la Red de Salud Puno -2013?

1.2 OBJETIVOS

1.2.1 Objetivo General

Implementar un Data Warehouse para la toma de decisiones eficientes de la Red de Salud Puno - 2013.

1.2.2 Objetivos Específicos

- ✓ Analizar y diseñar los datos utilizando herramientas y métodos del Data Warehouse.
- ✓ Implementar el Data Warehouse según las expectativas a nivel dirección para tener la información en tiempo real y confiable.
- ✓ Evaluar el Data Warehouse de acuerdo a la información que proporcione el mismo.

1.3 HIPOTESIS

Nuestra hipótesis es: el data warehouse ayuda de manera significativa para la toma de decisiones de la red de salud Puno - 2013.

1.4 JUSTIFICACIÓN Y DELIMITACIÓN DE INVESTIGACIÓN

1.4.1 Justificación de la Investigación

El presente trabajo obedece a la necesidad de tener información consistente en tiempo real de los datos que se almacenan en nuestro Data Warehouse. También se ha visto que la información que se tiene de los ¹HIS no ofrecen cierta conformidad por parte de las estrategias de la Red de Salud Puno, para esta investigación se tomó como referencia las evaluaciones semestrales que se realizan, así como su problemática, problemas dentro de la población entre muchos factores que son influyentes en una atención de calidad hacia la población. Estos problemas se nos harán mucho más sencillos de identificar teniendo de base un Data Warehouse y haciendo mucho más uso de este sistema se podrá realizar todo tipo de análisis y reportes.

1.4.2 Limitaciones de la Investigación

El primer limitante que se presentó en esta investigación es el acceso a los datos del HIS, porque el personal encargado del HIS no puede dar dicha información por ética de información, por otra parte otra de las limitantes fueron los datos mal sistematizados (mal ingresados) por parte del personal de Salud y de algunos errores del HIS, porque cada código no es ingresado por el Digitador solamente si tiene autorización del personal de Salud por eso se le considera una limitante ya que dentro de nuestro Data set hubo muchos registros sin Diagnostico y a si mismo se obtiene reportes solo de AIS Niño.

¹ HIS.- (Health Información System) Sistema de Información de Salud, es un sistema que se encarga de almacenar información de salud, y de todas las atenciones que se realiza a nivel nacional.

CAPITULO II

MARCO TEÓRICO

2.1 ANTECEDENTES DE LA INVESTIGACIÓN

Para reforzar los objetivos se han revisado trabajos previos que permiten conocer los avances que se han obtenido en este tipo de investigación. En este sentido, a continuación se presentan antecedentes que han servido como punto de partida para la misma.

1. PROYECTO: Construcción de una Data Warehouse de datos del medio ambiente para la toma de decisiones: aplicación a los datos hidrológicos Proyecto presentado por la Bach. María Magdalena Ydirim Universidad de las Américas, Puebla, Escuela de ingeniería, Departamento de Ingeniería de Sistemas Computacionales Cholula, Puebla, México.

La contribución del proyecto es SARP, un sistema de análisis de datos de ríos y presas del país, SARP constituye una Data Warehouse, lo que se denomina constelación, a partir de los datos obtenidos de las BANDAS (Sistemas Operacionales). Los Data Warehouse implementan sus esquemas multidimensionales sobre un sistema de administración de base de datos relacional (SGBDR) y están asociados a un motor de consulta analítica.

El motor de Consulta ofrece una interfaz gráfica para la expresión de consultas y se apoya en el lenguaje SQL para calcularlas, Cada Data Warehouse está asociada a un sistema integrado de datos que implementan módulos adaptados a la extracción de datos de las BANDAS y su integración con respecto a los esquemas multidimensionales.

- De este proyecto se extrajo el proceso de transformación y descripción de la metadata así como sus características.

2. PROYECTO: Incorporación de Soluciones OLAP en Entornos Empresariales.

Proyecto Realizado por el Ing. Patricia Andrea Zvenger, Universidad Nacional del Sur, Departamento de Ciencias e Ingeniería de la Computación.

Las bases de datos house-holding almacenaban información sobre la forma de vida de los clientes. Se mantenían clasificados por categorías para que los gerentes de negocios analizaran interrelaciones entre ellos y sus necesidades.

Luego surgió el data warehousing, que fue un paso más grande en el intento de integrar y cruzar la información de la organización destinada el soporte de decisiones. Este paso origino tareas y conceptos tales como reportes de ventas, indicadores clave, análisis de tendencias. Luego del surgimiento de la data warehousing comenzaron a surgir nuevas herramientas en el mercado y entre ellos una clase a la que nos ocuparemos en este proyecto, la herramienta OLAP (ON-LINE ANALITICAL PROCESSING).

CONCLUSIÓN. OLAP juega un rol importante dentro de las organizaciones, ya que ayuda a establecer, monitorear y continuar las medidas de performance

que permiten a los gerentes entender cómo se va desarrollando el negocio en el complejo entorno competitivo. Además permite establecer nuevas medidas de performance tradicionalmente, por tal motivo OLAP juega un papel muy importante y por eso se tomara en cuenta OLAP en la investigación.

3. PROYECTO: SAGITBA Sistema de Apoyo Gerencial para el Instituto Tecnológico de Buenos Aires

Proyecto realizado por el Ing. Javier Nader en el Instituto Tecnológico de Buenos Aires departamento de Ingeniería de Software, Buenos Aires, Argentina.

Con sistema del tipo “Inteligencia de Negocio” (Business Intelligence), se puede formular y responder a preguntas clave sobre el funcionamiento de la universidad accediendo directamente a los indicadores de éxito, señalar cuales son los factores que realmente inciden en el buen funcionamiento, detectar situaciones fuera de lo normal, encontrar factores que maximicen el beneficio y predecir el comportamiento futuro con un alto porcentaje de certeza.

Para esto es necesario implementar y brindar un sistema de Apoyo Gerencial. El sistema debe estar orientado a brindar información interrelacionada para quienes tiene responsabilidades en el ámbito estratégico y táctico de la organización.

Los usuarios de este sistema no son usuarios comunes, sino usuarios que toman decisiones y planifican día a día, a mediano plazo o a largo plazo, en donde la calidad de información tanto consolidada como detallada puede

predecir tendencias u comportamientos para la toma de decisiones proactivas.

4. PROYECTO: Aplicación de las Tecnologías Data Warehouse en el Contexto de la Empresa Turística de Alojamiento Hotelero.

Proyecto realizado por Antonia Gil Adilla, Rosario Berriel Martínez en Dpto. de Economía y Dirección de Empresas, Facultad de Ciencias Económicas y Empresariales Universidad de las Palmas de Gran Canaria.

Los cambios propiciados por el proceso de globalización de la economía, han configurado un nuevo entorno que está condicionando la actuación de las empresas del sector turístico, de tal modo, que la obliga a buscar nuevos métodos y técnicas que les ayuden a mejorar no solo la gestión sino también la toma de decisiones. Es necesario que centren sus esfuerzos en el desarrollo de una estrategia de integración de la información, que permita aprovechar el potencial que ofrece la tecnología de información (IT) para generar conocimiento y superar los inconvenientes planteados por los sistemas de gestión tradicionales. Una de las herramientas que podría ayudar a este objetivo, son la Data Warehouse (almacenes de datos) como tecnología idónea para incrementar el vacío de los datos que poseen y ayudar a afrontar las decisiones claves que afectan a su negocio.

Ante la situación descrita, y basándose en la revisión teórica realizada donde se pone en manifiesto las aplicaciones de las herramientas Data Warehouse y data Marts, consideramos que las IT citadas podría ser de gran ayuda para:

- 1) favorecer el proceso de toma de decisiones de áreas claves que hasta ahora no han contado con el soporte necesario, como podrían ser dirección,

ampliando la capacidad de análisis de estas.

2.2 BASE TEORICA

2.2.1 Base de Datos

Las bases de datos no son tan sólo una colección de archivos. Más bien, una base de datos es una fuente central de datos destinados a compartirse entre muchos usuarios para una diversidad de aplicaciones. El corazón de una base de datos lo constituye el sistema de administración de base de datos (dbms, datábase management system), el cual permite la creación, modificación y actualización de la base de datos, la recuperación de datos y la generación de informes y pantallas. La persona encargada de garantizar que la base de datos cumpla sus objetivos se conoce como administrador de base de datos.

2.2.2 Business Intelligence

Según, el término de BI (Business Intelligence) es usado por diferentes expertos y fabricantes de software para distinguir un amplio rango de tecnologías, plataformas de software, aplicaciones específicas y procesos. Se utiliza este término desde tres diferentes perspectivas:

- Tomar mejores decisiones rápidamente
- Convertir los datos en información
- Utilizar un método razonable para la gestión empresarial

El objetivo primario de la Inteligencia de Negocios es ayudar a las personas a

tomar decisiones que mejoren el rendimiento de la compañía e impulsen su ventaja competitiva en el mercado. Es decir, faculta a las organizaciones a tomar las mejores decisiones rápidamente. Para tomar mejores decisiones más rápidamente, los directivos y gerentes necesitan de información relevante y útil al alcance de la mano. Pero es común una larga brecha entre la información que los responsables en la toma de decisiones requieren, y las grandes cantidades de datos que las organizaciones recopilan cada día. Para saltar esta brecha, las organizaciones hacen significativas inversiones en desarrollar sistemas de BI para convertir los 12 datos originales en información de utilidad. Los sistemas de BI más efectivos tienen acceso a inmensas cantidades de datos para posteriormente entregar a los responsables en la toma de decisiones, información expresada de una forma que ellos pueden asimilar fácilmente. La inteligencia de Negocios puede ser definida como un método para la gestión empresarial, una forma de pensamiento organizacional y una filosofía de gestión. Tanto las personas como las organizaciones se interesan en la Inteligencia de Negocios, porque creen que el uso de un enfoque racional y basado en hechos a la hora de tomar decisiones resulta positivo en la medida que sea posible. El interés por adoptar el BI tiene las siguientes características:

- Buscar hechos (datos) que se puedan medir cuantitativamente acerca del negocio.
- Usar métodos organizados y tecnologías para analizar los hechos.
- Inventar o compartir modelos que expliquen las relaciones de causa y efecto entre las decisiones operativas y los efectos que éstas tienen en alcanzar los objetivos de negocio.

- Experimentar con métodos alternos y supervisar con retroalimentación sobre los resultados.
- Gestión de la empresa (decisiones e iniciativas) basadas en todas estas características.

Grafico 1 Esquema de Un Data Warehouse



Fuente: <http://www.dataprix.com/datawarehouse-vs-datamart>

2.2.2.1 Data Mart

Otro concepto de mucha importancia en lo que es BI es el de Data Mart. Un Data Mart es un subconjunto del Data Warehouse, con un alcance de contenido limitado. Éste se usa para un solo departamento de una organización y/o un problema particular de análisis dentro de la organización. Un Data Mart por sí solo, no es un Data Warehouse, ya que un Data Warehouse tiene más usuarios y más temas que un Data Mart, y provee una vista completa de las áreas funcionales.

De la organización. Un Data Mart, al igual que un Data Warehouse, consiste en una Base de Datos. Así mismo, define el Data Warehouse como un Repositorio

colectivo y centralizado que nutre o alimenta una serie de almacenes Que tienen una orientación específica o dominio específico, o tema específico, llamados Data Marts.

¿Por qué construir una data Mart?

Los data Mart no son siempre indispensables, más bien pueden añadir tiempo al proceso de actualización ya que se deben actualizar un número de sectores con nuevas data, sin embargo, los data Marts significan más rápidos queries (consultas) al haber menos data en cada uno de ellos.

2.2.2.2 Concepto de Data Marts

Existen varias definiciones que hacen referencia a la data Mart que a continuación se detallan.

Concepto 1.- El concepto de data Mart es una extensión natural de la Data Warehouse, y está enfocado a un departamento o área específica, como por ejemplo los departamentos de finanzas o marketing, permitiendo así un mejor control de la información que se está abarcando.

Toda empresa puede ser vista en base al proceso productivo que la sustenta. El resultado de los costos y beneficios de este proceso productivo forman una cadena de valor, donde cada eslabón (proceso de negocios) adiciona valor a la empresa. De esta forma es claro, que las empresas deben buscar optimizar cada uno de sus eslabones sin perder de vista la cadena total.

Al manejar eficientemente la información de cada área de la empresa, se pueden tomar mejores decisiones y así efectuar acciones apropiadas y finalmente

conseguir un mejor control sobre la producción empresarial.

En esta nueva tecnología cada eslabón de cadena de valor será representado por una base de datos multidimensional, la cual permite potencialmente administrar la etapa productiva que representa, la cadena de valor total será representada entonces por el conjunto de bases de datos multidimensionales asociadas a cada eslabón.

Concepto 2.- El Data Mart es un modelo multidimensional basado en tecnología OLAP que presenta a una área específica de la compañía, incluyendo las variables claves y los indicadores para el proceso de toma de decisiones. Una data Mart es simplemente un pequeño Data Warehouse. Usualmente la data Mart es un sector específico de data que es encontrado en la base de datos de la empresa.

Concepto 3.- Un depósito de datos, frecuentemente es parte de un conjunto mayor, diseñado para entender los requisitos específicos de un departamento o grupo dentro de la empresa.

Concepto 4.- Data Mart es una extensión natural del Data Warehouse, y está enfocado a un departamento o área específica, como por ejemplo los departamentos de finanzas o márketing, permitiendo así un mejor control de la información que está abarcando.

Este nuevo enfoque es el marketing basado en la relación con el cliente, también llamado marketing uno a uno. Este nuevo grado de relación se consigue con una gestión de la información más eficiente y rápida, junto con la capacidad de ajustar las campañas de marketing a cada individuo, detectando sus necesidades u

prediciendo su comportamiento.

Esta tecnología bien utilizada reduce la pérdida de cliente, aumenta la posibilidad de éxito, e incrementa los ingresos de la empresa. Al manejar eficientemente la información de cada área de la empresa, se pueden tomar mejores decisiones y así efectuar acciones apropiadas y finalmente conseguir un mejor control sobre la producción empresarial.

Concepto 5.- Un data Mart es una pequeña escala de un Data Warehouse que resuelve problemas de negocios de los usuarios. Contiene datos enfocados a las necesidades de un departamento específico como producción, ventas, marketing o finanzas, debido a la competencia y la necesidad de dar respuesta a los clientes han impulsado a las empresas a distribuir la toma de decisiones más cerca del punto contacto con los mismos.

Una vez contando con la base de información empresarial integrada, a partir de esta, se crean subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones, los datos existentes en este contexto puede ser sumariados, agrupados, explorados y reportados de múltiples formas para que diversos grupos de usuarios realicen la explotación de los mismos.

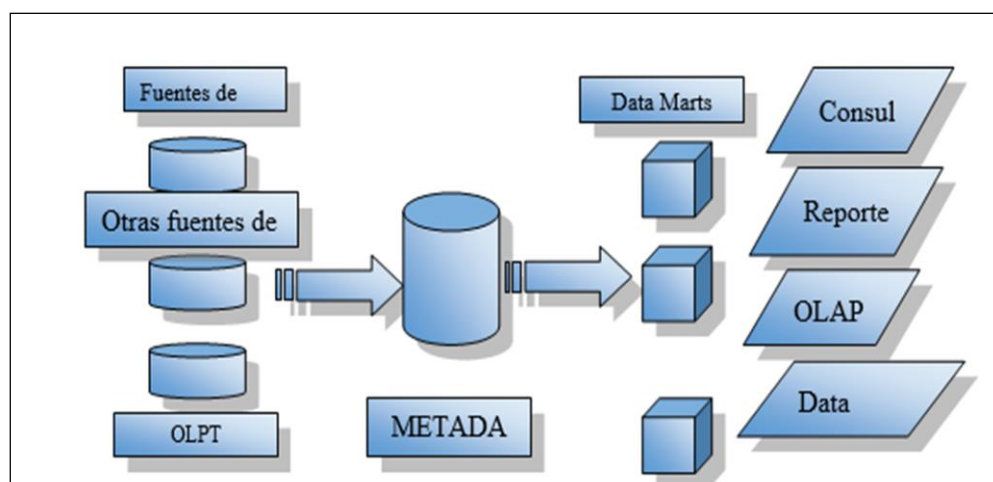
2.2.2.3 Estructura Descentralizada de una data Mart

El departamento de marketing, emprende el primer proyecto de Data Warehouse como una solución departamental, creando el primera data Mart de la empresa, visto el éxito del proyecto, otros departamentos, como el de riesgos, o el financiero se lanza a crear sus datas Marts. Marketing, comienza a usar otros

datos que también usan la data Marts de riesgo y financiero, y estos hacen lo propio.

Esto parece ser una decisión normal, puesto que las necesidades de información de todos los data Marts crecen conforme el tiempo avanza. Cuando esta situación evoluciona, la coordinación y gestión de información de todos los Data Marts es un Data Warehouse centralizado. Simplificada el crecimiento de una base de conocimientos a nivel de toda empresa.

Grafico 2 Esquema de un Data Warehouse



Fuente: www.intenego.blogspot.com.ar/2012/07/extraccion-transformacion-y-carga.html

2.2.2.3 Sistema de Soporte de Decisiones (SSD)

Técnicamente se puede definir como “sistema de cómputo a nivel directivo de una institución, que combinan los datos y modelos analíticos sofisticados para dar soporte a la toma de decisiones semi estructurados u no estructuradas”.

El SSD en sí, es un bloque de toma de decisiones sustentado en base de datos que quienes toman la decisiones puedan usar para apoyar el proceso de decidir.

En un sentido amplio, se define obtener de manera oportuna la información que

se requiere mediante el proceso de la toma de decisiones que se desarrolla en un ambiente de incertidumbre. Ayudan a la toma de decisiones de los administradores al combinar datos, modelos analíticos sofisticados y software amigable en un solo sistema poderoso que puede dar soporte a la toma de decisiones semi estructuradas o no estructuradas. El SSD está bajo control de usuario desde la concepción inicial a la implantación final.

Los sistemas de ayuda a la toma de decisiones (Decisión support System - SSD) posee como principal objetivo la complementación de las capacidades de decisión del ser humano, validándose de la potencia que adoptan las computadoras para el procesamiento de datos, los SSD deben mejorar el proceso de toma de decisiones, proporcionando información que actualmente no existe, como también mejorar en el acceso, la visualización el análisis de información. Por otro lado se proporcionan mejoras en el procedimiento deductivo partir de la información con la cual disponemos y ofreciendo mejor explicación a terceros sobre las decisiones tomadas.

La estrategia de la organización debe contemplar, como requisito indispensables para su éxito, la creación y mantenimiento de una base tecnológica adecuada que sustente al proceso decisorio, permitiéndose alcanzar las ventajas competitivas deseadas, así como crear las mayores sinergias favorables para la organización.

Una parte sustancial de sostén de la cultura organización es la información almacenada. A partir de la disfunción y del uso masivo de la tecnología de la información se ha creado y mantenido bases de datos, algunos de los cuales merecen el nombre de almacenes de datos o "Data Warehouse".

En los setenta, muchas empresas comenzaron a desarrollar sistemas de información muy diferentes de los sistemas de información para la administración (SIA) tradicionales. Esos nuevos sistemas eran más pequeños (en términos de mano de obra y costo), eran interactivos (pocos común en ese tiempo), y estaban diseñados para ayudar a los usuarios finales a utilizar datos y modelos para discutir y decidir (no resolver) problemas semi estructurados u no estructurados. A finales de los 80, estos primeros esfuerzos para ayudar a la toma de decisiones individuales se extendieron a los grupos y a las empresas.

La filosofía de un SSD es dar a los usuarios las herramientas necesarias para analizar bloques importantes de datos, usando modelos sofisticados fácilmente controlables de manera flexible, los SSD están diseñados para dar capacidades, no solo para responder a las necesidades de información.

Los SSD están mejor orientados que los sistemas SIA.- estos sistemas proporcionan a los administradores flujos rutinarios de datos y nos ayudan a la control general de la organización, en contraste, los SSD están muy enfocados a clases de decisiones específicas.

El SSD tiene como finalidad apoyar a la toma de decisiones mediante a la generación y evaluación sistemática de diferentes alternativas o escenarios de decisión, utilizando modelos y herramientas computacionales, Un SSD no soluciona problema, ya que solo apoya el proceso de la toma de decisiones, la responsabilidad de tomar una decisión, de optarla y de realizarla es de los administradores, no del SSD.

2.2.2.4 Características de los Sistemas de Soporte de Decisiones

Se toma las siguientes consideraciones:

- **Tipo de decisiones:** apoyar el proceso de toma de decisiones estructurada u no estructurada.
- **Frecuencia de uso:** tiene una utilización frecuente por parte de la administración media y alta para el desempeño de su función.
- **Variedad de usuarios:** puede emplearse por usuarios de diferentes áreas funcionales como ventas y producción, administración, finanzas y recursos humanos.
- **Flexibilidad:** permite acoplarse a una variedad determinada de estilos administrativos, autocrática, participativa, etc.
- **Desarrollo:** permite que el usuario desarrolle de manera directa los modelos de decisión sin participación operativa de profesionales en informática.
- **Interacción ambiental:** permite la posibilidad de interactuar con información externa como parte de los modelos de decisión.
- **Comunicación inter-organizacional:** facilita la comunicación de información relevante de los niveles altos a los niveles operativos y viceversa a través de gráficas.
- **Acceso a base de datos:** tiene la capacidad de acceder información de las bases de datos corporativos.
- **Simplicidad:** simple y fácil de aprender y utilizar por el usuario final.

2.2.3 Olap

En lo que es consultas para la presentación de los datos, es importante explicar el concepto de OLAP (Online Analytical Processing). OLAP proporciona un modelo de datos intuitivo y conceptual, para que los usuarios que no tengan experiencia como analistas puedan comprender y relacionar los datos mostrados. Este modelo es llamado análisis multidimensional, siendo habilitado para ver los datos a través de múltiples filtros, o dimensiones. Los sistemas OLAP organizan los datos directamente como estructuras multidimensionales, incluyendo herramientas fáciles de usar por usuarios para conseguir la información en múltiples y simultáneas vistas dimensionales. OLAP es también rápido para el usuario.

Rápidos tiempos de respuesta permiten que los gerentes y analistas puedan preguntar y resolver más situaciones en un corto período de tiempo. Una dimensión es una vista de los datos categóricamente consistente. Una característica de las dimensiones es la habilidad de hacer slice-and-dice. Slice (rebanada) y dice (cubo) hacen particiones de los datos en una base de datos multidimensional de acuerdo a los valores de ciertas dimensiones. Otra capacidad inherente en el diseño de OLAP es la rotación y anidamiento (Pivoting-and-Nesting) de las dimensiones. El pivote permite rotar los datos desde las columnas hasta las filas. También es importante mencionar el concepto de drill, que en los sistemas OLAP tiene un significado muy específico. Drill Down es la acción de seleccionar un miembro para ver el siguiente nivel inferior de detalle en la jerarquía. Drill up es seleccionar un miembro para ver el siguiente nivel superior, esto es, una acción de bottom-up. La mecánica o funcionamiento de las interfaces OLAP, especialmente pointing- and-clicking

(apuntar y seleccionar) para hacer drill-Down dentro de las capas de interés se hace posible por la velocidad con que las consultas son resueltas. Esta funcionalidad permite por completo a los gerentes y analistas un nuevo proceso para tratar con grandes cantidades de datos, un proceso conocido con el nombre de análisis ad hoc. En resumen, los sistemas OLAP organizan los datos por intersecciones multidimensionales. Esta organización, acompañada por una herramienta de interface para rotar y anidar dimensiones, permite a los usuarios visualizar rápidamente valores en detalle, patrones, variaciones y anomalías en los datos que estarían de otra manera ocultos por un análisis dimensional simple. A mayor número de dimensiones (dentro de los límites razonables), mayor es la profundidad del análisis.

2.2.4. Molap, Rolap, Holap

Existen variaciones de OLAP según la cantidad de datos y la eficiencia requerida. OLAP, no se recomienda para consultas complejas y que recorran muchas tablas. Una de estas variaciones es MOLAP (Multidimensional online analytical processing), según, los datos son colocados en estructuras especiales que se encuentran en un servidor central. MOLAP ofrece el mayor rendimiento de recuperación de información. Por otra parte, existe la solución ROLAP (Relational online analytical processing), permite tomar ventaja de uno de sus más grandes beneficios, el almacenamiento de inmensas cantidades de datos.

El rendimiento de recuperación de la información para ROLAP frecuentemente no es tan rápido como otras opciones de almacenamiento ROLAP es recomendado para consultas pesadas que no se usan muy a menudo. Finalmente existe HOLAP (Hybrid online analytical processing), que es un híbrido entre

MOLAP y ROLAP, y según, HOLAP no es realmente un modo diferente de almacenamiento de datos. Más bien es la habilidad para diseminar los datos a través de bases de datos relacionales y multidimensionales con la finalidad de obtener lo mejor de ambos sistemas.

- **Dimensión:** es un grupo de miembros consistentes categóricamente representados como una arista específica de un cubo OLAP, por ejemplo, el tiempo, clientes, productos.
- **Jerarquía:** es la organización de niveles dentro de una dimensión que refleje cómo los datos añadidos están agregados nivel a nivel, y el camino que permita hacer drill-Down de arriba abajo dentro de la dimensión. Por ejemplo: año, trimestre y mes.
- **Miembro:** es el nombre o etiqueta para cualquier miembro de cualquier nivel en una jerarquía. Los miembros inferiores son llamados algunas veces miembros hoja o miembros de nivel cero.
- **Generación jerárquica:** este término se utiliza para describir las relaciones entre miembros de una jerarquía. Lo más común es usar nombres de familia, como los siguientes:
 - ✓ **Hijo:** es un miembro directamente subordinado o por debajo de otro miembro en una jerarquía.
 - ✓ **Padre:** es un miembro que está directamente encima de otro miembro en una jerarquía.
 - ✓ **Hermano (Sibling):** es un miembro que está al mismo nivel de uno o más

miembros compartiendo el mismo padre.

✓ **Descendiente:** cualquier miembro en cualquier nivel en relación a otro miembro específico.

✓ **Ancestro:** cualquier miembro de cualquier nivel superior en relación a otro miembro.

2.2.5 Diseño Logico

La premisa básica en el diseño lógico es:

- Preparar a las bases multidimensionales para soportar el recupero de una gran cantidad de filas de datos en forma rápida.
- La mayoría de los analistas de negocios van a querer ver datos totalizado.

Estos datos en lo posible deben recalcularse y almacenarse de antemano para que esta recuperación sea rápida y eficiente.

- El diseño debe estar conducido por el acceso y por el uso, es decir, teniendo en cuenta que tipo de reportes o resúmenes son los más frecuentes, y cuales los más urgentes.
- Un diseño normalizado no es bueno, no solo por lo mencionado en la sección anterior, sino porque no resulta demasiado intuitivo para una persona de negocios, y podría volverse demasiado complejo.
- Todos los datos se incluyan ya deben existir en las fuentes de datos operacionales, o ser derivables a partir de ellos.

Las dos técnicas de diseño más populares son el esquema Star y el esquema Snowflake.

2.2.6 Esquema Star (Estrella)

Este esquema está formado por un elemento central consiste en una tabla llamada tabla de hechos, que está conectada a varias tablas de dimensiones.

2.2.7 Esquema Snowflake (Copo de nieve)

Es una variante al esquema estrella en la cual las tablas de dimensión están normalizadas, es decir, pueden incluir claves que apuntar a otras tablas dimensión.

2.2.8 Sistema Datawarehousing

Es un sistema informático capaz de ofrecer información para toma de decisiones, cuya pieza principal es un Data Warehouse.

2.2.9 Data Warehouse y Base de Datos Operacional

La mayoría de los sistemas operacionales están dirigidos a la carga de datos con cientos de miles de transacciones diarias y respectivas, que además requieren un tiempo de respuesta muy corto por parte de los usuarios.

Las bases de datos operacionales deben estar diseñadas con el objetivo de hacer esta tarea lo más eficiente posible. Otro objetivo del diseño es el ahorro en el espacio de almacenamiento.

Las aplicaciones destinadas al análisis de datos en cambio, están orientadas a la

Salida de información, ya sea en forma de reportes, cubos o respuestas a consultas mediante sistemas de consulta y reporte. Estas consultas y reportes no necesariamente se producirían en el mismo día, ni diariamente. No son tan previsibles y frecuentes como las transacciones en un sistema operacional, se podría consultar por un conjunto de datos en periodo mensuales, otros datos se pueden requerir en forma cuatrimestre. Etc.

2.2.10 Diferencia entre el Procesamiento OLAP y OLTP

La compra, venta, producción y distribución son ejemplos comunes de actividades de negocios de la operación diaria. Estas actividades constituyen el llamado procesamiento operación u OLTP (On Line Transactional Processing), y las aplicaciones que soportan este procesamiento se diseñan con orientación a la carga de datos. El planeamiento de recursos, planeamiento financiero, alianzas estratégicas, e iniciativas de mercado son ejemplos de actividades que generan y usan información basada en análisis y orientada a la toma de decisiones. Este tipo de actividades son soportadas por las aplicaciones de tipo OLAP, Podemos decir que las características que diferencia el procesamiento OLTP del procesamiento de tipo OLAP son las siguientes.

- Las aplicaciones OLTP son usadas por un gran número de usuarios concurrentes. Una aplicaciones OLAP, en cambio, no tiene que soportar este nivel de concurrencia, porque en general será usada solo a nivel gerencial o por los analistas de negocios.
- A diferencia del tiempo de respuesta a las consultas OLAP, el tiempo de respuesta de las operaciones en unas aplicaciones OLTP es crítico, debe de estar en el orden de los microsegundos, ya que un cliente puede estar

esperando la respuesta.

- El procesamiento OLTP tiende a concurrir en tiempo relativamente constantes, es decir, la frecuencia con la que se realizan la operación no varía demasiado. Por ejemplo, una aplicación de facturación recibe operaciones todo el tiempo pocos predecibles. Se pueden requerir algunos informes en forma mensual, o semanal, o cualquier momento.
- En una aplicación OLTP los datos se actualizan con la misma frecuencia con la que se leen. No es así en el caso del procesamiento OLAP, donde los datos se actualizan cada cierto periodo de tiempo cuando se hace el volcado de datos desde las bases de datos operacionales.
- Las consultas hechas a la aplicación OLTP actúa sobre una cantidad no muy extensa de datos, y la naturaleza de estas consultas no tiene complejidad. Por ejemplo, cuando se consulta sobre un determinado cliente para ver datos como su nombre y su domicilio. Esta consulta implica recuperar y comparar datos no derivados, es decir no se calculan en el momento. Este tipo de consultas en general se conoce y pertenece a un proceso dentro de la organización. En aplicaciones orientadas a la toma de decisiones en cambio, las consultas requieren el examen y recuperación de un cambio, las consultas requieren el examen y recuperación de una enorme cantidad de datos, incluso datos históricos.

Los usuarios de una aplicación OLAP genera consultas complejas y ad hoc, y a veces la respuesta a una consulta conduce a la formulación de otras más detallada, en general los usuarios comienzan buscando en los datos resumidos

y al identificar áreas de interés, comienzan a acceder al conjunto de datos detallado.

2.2.11 Operaciones Multidimensionales

Las operaciones multidimensionales se pueden agrupar en tres conjuntos básicos que se describen a continuación:

Operaciones de Slice & Dice (selección) no existe un concepto general en la definición de Slice & Dice, sin embargo, se acepta que hay tres operaciones incluidas en este conjunto.

- **Slice:** Seleccionar una tajada de un cubo tomando un determinado valor de dimensión de un cubo mayor, definiendo un sub cubo.
- **Dice:** tomar sesiones del cubo en función de un rango de valores de las dimensiones.
- **Rotación (pivoting):** permite visualizar diferentes planos de un cubo, en general el usuario indica la rotación arrastrando una dimensión a una nueva posición y la herramienta rota la perspectiva automáticamente.

2.2.12 Molap

Las base de datos multidimensionales usan estructuras de tipo arreglo para almacenar los datos, estas estructuras están indexadas con el fin de proveer un tiempo de acceso optimo a cualquier elemento. Se pueden distinguir dos enfoques en la forma de organizar estas estructuras, las MDDBs que se usan una arquitectura de hipercubo, y las que usan mutlicubos. La arquitectura de hipercubo almacena un único gran cubo en la cual cada medida esta referenciada

y totalizada en todas las dimensiones del mismo. En la arquitectura de multicubos los datos se guardan en más de un cubo.

2.2.13 Sistema

Un sistema es un conjunto de funciones, virtualmente referenciada sobre ejes, bien sean estos reales o abstractos. También suele definirse como un conjunto de elementos dinámicamente relacionados formando una actividad para alcanzar un objetivo operando sobre datos, energía o materia para proveer información.

2.2.14 Gestión

Generalmente una persona que “gestiona” es aquella que mueve todos los hilos necesarios para que ocurra determinada cosa o para que se logre determinado propósito. Existen varias acepciones para el término gestión: El diccionario Larousse la define así.

Del latín gestión. Acción y efecto de administrar. En el libro “Indicadores de Gestión”, publicado en 1999, lo define de la siguiente manera: “conjunto de decisiones y acciones que llevan al logro de objetivos previamente establecidos”.

De modo que la gestión, organizacionalmente hablando, se refiere al desarrollo de las funciones básicas de la administración: Planear, organizar, dirigir y controlar.

2.2.15 Control

El concepto de control es muy general y puede ser utilizado en el contexto organizacional para evaluar el desempeño general frente a un plan estratégico. La situación de estar bajo el dominio, dirección, regulación o comando, de algo o de alguien.

2.2.16 Software

Se conoce como software al equipamiento lógico o soporte lógico de una computadora digital; comprende el conjunto de los componentes lógicos necesarios que hacen posible la realización de tareas específicas, en contraposición a los componentes físicos del sistema, llamados hardware.

Tales componentes lógicos incluyen, entre muchos otros, aplicaciones informáticas como el procesador de textos, que permite al usuario realizar todas las tareas concernientes a la edición de textos o el software de sistema tal como el sistema (González, Marzo 2005) operativo, que, básicamente, permite al resto de los programas funcionar adecuadamente, facilitando la interacción con los componentes físicos y el resto de las aplicaciones, proporcionando también una interfaz para el usuario.

2.2.17 Ingeniería de Software

La ingeniería de sistemas se define como la aplicación efectiva de esfuerzos científicos y de ingeniería para transformar una necesidad operativa en una configuración definida de un sistema mediante el proceso iterativo de análisis de requisitos, la selección del concepto, y asignación, síntesis, soluciones de compromiso y optimización del diseño, prueba y evaluación. Entre sus características se incluye su estructura de arriba-abajo que ve el sistema como un todo; una orientación del ciclo de vida que considera todas las fases desde el diseño conceptual hasta la retirada del sistema; un enfoque interdisciplinario que incluya todas las disciplinas adecuadas de diseño de forma oportuna y concurrente; y la necesaria integración para asegurar que todos los objetivos de diseño se han cumplido de forma efectiva y eficaz. Está orientada al proceso e

incluye las provisiones esenciales de realimentación y control.

Se desea introducir aquí una definición clásica de ingeniería de sistemas de software como contrapunto a la anterior. Pressman la define enfatizando aspectos concretos de calidad: "establecimiento y uso de principios de ingeniería robustos, orientados a obtener software económico que sea fiable y funcione de manera eficiente sobre máquinas reales".

Pressman presupone que el sistema de software a realizar cumpla con la misión encomendada, satisfaga las expectativas del usuario, etc. Si comparamos esta definición con la de ingeniería de sistemas anterior, vemos que la de Pressman está enfocada al proceso de desarrollo y no tanto a la tecnología con la que se desarrolla; para él es más una ingeniería de proceso y no tanto de producto.

De su análisis se desprende que la ingeniería de sistemas de software es una especialización de la ingeniería de sistemas cuya creciente importancia está ligada a la de los sistemas de software en nuestra sociedad.

2.2.18 Tecnología de Software

Se define tecnología de software como un conjunto integrado de notaciones, herramientas y métodos, basados en unos sólidos fundamentos, que permiten el desarrollo de un producto software.

2.2.19 MySQL

MySQL es un sistema para la administración de bases de datos relacionales

(RDBMS) rápido y sólido. Las bases de datos permiten almacenar, buscar, ordenar y recuperar datos de forma eficiente. El servidor de MySQL, controla el acceso a los datos para garantizar el uso simultáneo de varios usuarios, para pronosticar acceso a dichos datos y para asegurarse de sólo obtienen acceso a ellos los usuarios con autorización. Por lo tanto, MySQL es un servidor multiusuario y de subprocesamiento múltiple. Utiliza SQL (del inglés Structure Query Lenguaje, Lenguaje de consulta estructurado), el lenguaje estándar para la consulta de bases de datos utilizado en todo el mundo. MySQL lleva disponible desde 1996 pero su nacimiento se remonta a 1979. Ha obtenido el galardón Choice Award del Linux Journal Readers en varias ocasiones.

MySQL se distribuye bajo un sistema de licencias dual. Puede utilizarlo bajo una licencia de código abierto (GNU GPL), que es gratuita mientras cumpla las condiciones de la misma. Si desea distribuir una aplicación que no sea GLP y que incluya MySQL, puede adquirir una licencia comercial.

2.2.20 Cubo Multidimensional

En una base de datos multidimensional, el modelo de datos está constituido por lo que se denomina un cubo multidimensional o simplemente Cubo. En un cubo la información se representa por medio de matrices multidimensionales o cuadros múltiples entradas, que no permite realizar distintas combinaciones de sus elementos para visualizar los resultados desde distintas perspectivas y variando los niveles de detalle. Esta estructura es independiente del sistema transaccional de las organizaciones, facilita y agiliza la consulta de información histórica ofreciendo la posibilidad de navegar y analizar los datos.

2.3 DATA WAREHOUSE

Hoy en día las empresas cuentan en su mayoría con la automatización de sus procesos, manejando gran cantidad de datos en forma centralizada y manteniendo sus sistemas en línea. En esta información descansa el know-how de la empresa, constituyendo en recurso corporativo primario y parte importante de su patrimonio.

El nivel competitivo alcanzado en la empresa les ha exigido desarrollar nuevas estrategias de gestión. En el pasado, las organizaciones fueron típicamente estructuradas en forma piramidal con información generada en su base fluyendo hacia lo alto y era en estrato de la pirámide más alto donde se tomaban decisiones a partir de la información proporcionada por la base, con un bajo aprovechamiento del potencial de esta información. Estas empresas han reestructurado y eliminado estratos de estas pirámides y ha autorizado a los usuarios de todos los niveles a tomar mayores decisiones y responsabilidades, sin embargo, sin información sólida para influenciar y apoyar las decisiones, la autorización no tiene sentido.

Esta necesidad de obtener información para una amplia variedad de individuos es la principal razón de negocios que conduce al concepto de Data Warehouse. El énfasis no está solo en llevar la información hacia lo alto sino que a través de la organización, para que todos los empleados que la necesiten la tengan a su disposición, el Data Warehouse convierte entonces los datos operacionales de una organización en una herramienta competitiva, por hacerlos disponibles a los empleados que lo necesiten para el análisis y toma de decisiones.

2.3.1 Características de un Data Warehouse

- **Orientado a temas o a sujetos:** Un primer aspecto de un Data Warehouse es que está orientado a los mayores sujetos de la empresa. El mundo operacional está diseñado alrededor de aplicaciones y funciones, como por ejemplo pagos, ventas y entregas de mercadería para una institución comercial. Un Data Warehouse está organizado alrededor de los motores sujetos, como clientes, vendedor, producto y actividades.

El mundo operación concierne al diseño de la base de datos y al diseño de procesos. Un Data Warehouse está enfocado en la modelización de los datos y el deseo de las base de datos exclusivamente. El diseño de procesos (en su forma clásica) no es parte del Data Warehouse.

- **Integrado:** El aspecto más importante del ambiente de un Data Warehouse es que sus datos están integrados. La integración se realiza cuando los datos son movidos del ambiente operacional, antes de entrar en el Warehouse.

Por ejemplo, un diseñador puede representar el sexo como “M” y “F”, otro puede representarlo como “0” y “1” o “X” y “Y” y otro usar las palabras completas como “masculino” y “femenino” no importa la fuente de la cual el campo sexo llegue al Data Warehouse, debe ser guardado en forma consistente, es decir que los datos deben estar integrados.

- **De tiempo variante:** los datos en el Data Warehouse son precisos para un cierto momento, no necesariamente ahora; El Data Warehouse copia datos de un largo horizonte de tiempo, las aplicaciones operacionales, por el contrario, contienen datos de intervalos de tiempo pequeños por cuestiones de

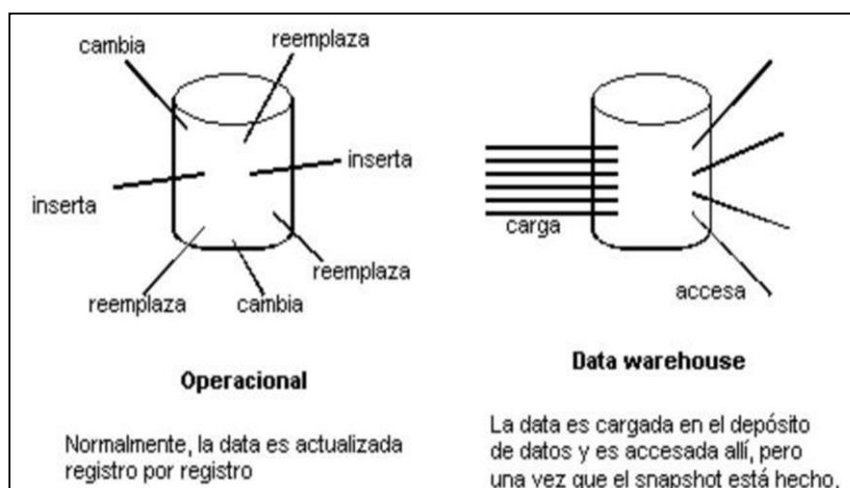
performance (tamaño chico de las tablas) Toda estructura clave en una Data Warehouse contiene implícita o explícitamente un elemento de tiempo. Esto no necesariamente pasa en el ambiente operacional.

En general, los datos de un Data Warehouse, una vez almacenados, no pueden ser modificados, no se permiten actualizaciones (adates). En el ambiente operacional, los datos, precisos al momento de acceso, pueden ser actualizados según sea necesario.

- **No volátil:** actualizaciones, inserciones y borradores son efectuados regularmente a los datos operacionales, en una base registro por registro (record by record). La manipulación de datos en una Data Warehouse es mucho más sencilla. Solo ocurren dos operaciones, la carga inicial, y el acceso a los datos. No hay necesidad de actualizaciones (en un sentido general)

Hay consecuencias muy importantes de esta diferencia de procesos con un sistema operacional, a nivel de diseño en un Data Warehouse no hay que controlar anomalías producidas por los “adates” ya que no hay “adates”, se puede tomar libertadas de diseño físico como optimizar el acceso a los datos y de normalización física, otra consecuencia es la simplicidad de la tecnología de Data Warehouse, en lo que respecta a back up, recuperación, integridad etc.

Grafico 3. Diferencia de sistema Operacional y Data Warehouse



Fuente: <https://pnfiequipo6.wordpress.com>

2.3.2 Funciones de Data Warehouse

En resumen se podría establecer que la Data Warehouse proporciona los siguientes beneficios a la gestión de la empresa

- Analiza la conducta de los clientes. Algunas de las preguntas son: ¿Cuál es el perfil de mis clientes?, ¿cómo es su comportamiento?, ¿Cuál es la rentabilidad que me deja?, ¿Cuál es el riesgo que corro con él? etc.
- Analiza la conducta de la empresa.
- Analiza la conducta del mercado, como un todo y por sectores.
- Localiza y planea como llegar a los clientes o sectores del mercado.
- Establece las nociones empresariales vitales que ayudan a controlar la participación en el mercado y elevar las utilidades.
- Establece la relación total con cada cliente para desarrollar las estrategias de propios adecuados con base en el uso real del producto y la experiencia del

cliente.

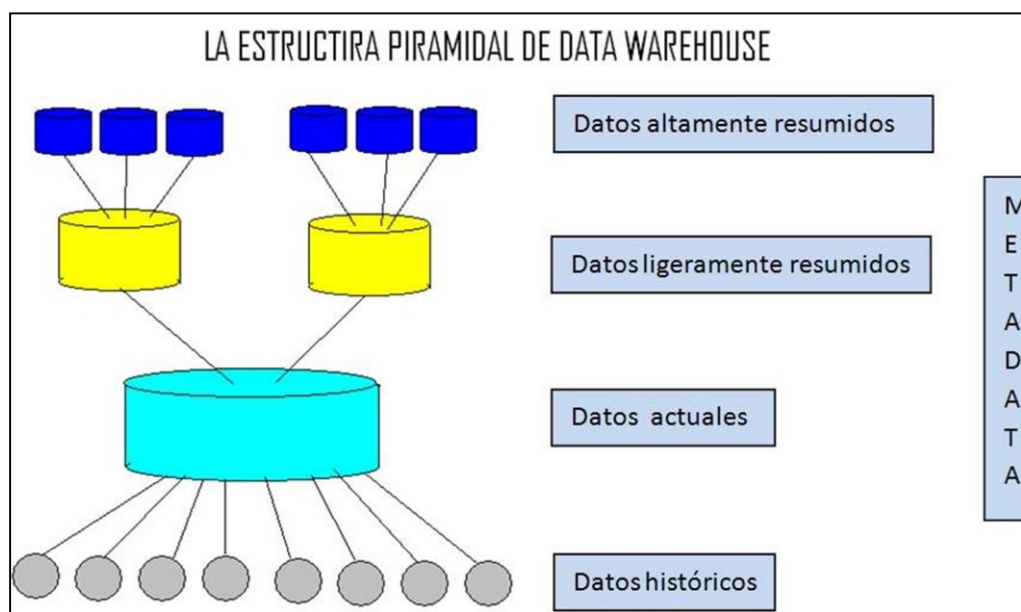
- Reduce los gastos promocionales e incrementa al mínimo tiempo la efectividad neta de las promociones en general.
- Realiza análisis estadísticas con perspectivas historias.

2.3.3 Estructura del Data Warehouse

La Data Warehouse tiene una estructura distinta. Hay niveles diferentes de esquematización y detalle que delimitan el Data Warehouse.

- **Detalle de datos actuales:** En gran parte, el interés más importante radica en el detalle de los datos actuales, debido a que refleja las ocurrencias más recientes, las cuales son de gran interés, es voluminoso ya que se almacena al más bajo nivel de granularidad casi siempre se almacena en discos el cual es de fácil acceso, aunque su administración sea costosa o compleja.
- **Detalle de datos antiguos:** La data antigua es aquella que se almacena sobre una forma de almacenamiento masivo. No es frecuentemente usada y se almacena a un nivel de detalle consistente con los datos detallados actuales, mientras no sea prioritario el almacenamiento de un medio de almacenaje alternativo, a causa del gran volumen de datos unidos al acceso no frecuente de los mismos, es poco usual el disco como medio de almacenamiento.
- **Datos ligeramente resumidos:** La data ligeramente resumida es aquella que proviene desde un bajo nivel de detalle encontrado al nivel de detalle.

Grafico 4 Estructura Piramidal de Data Warehouse



Fuente: <http://www.gestiopolis.com/canales/demarketing/articulos/68/mktinv1.htm>

- **Datos completamente resumidos:** el siguiente nivel de datos encontrados en el Data Warehouse es el de los datos completamente resumidos. Estos datos son compactos y fácilmente accesibles.

A veces se encuentran los ambientes de Data Warehouse y en otros, fuera del límite de la tecnología que ampara la Data Warehouse. (De todos modos, los datos completamente resumidos son parte del Data Warehouse sin considerar donde se alojar los datos físicamente).

- **Metadata:** el componente final de la Data Warehouse es el de la meta. De muchas maneras la Metadata se sitúa en una dimensión diferente al de otros datos de la Data Warehouse, debido a que su contenido no es timado directamente desde el ambiente operacional. La metadata juega un rol especial y muy importante en la Data Warehouse y es usado como:
 - Un directorio para ayudar al analista a ubica los contenidos del Data

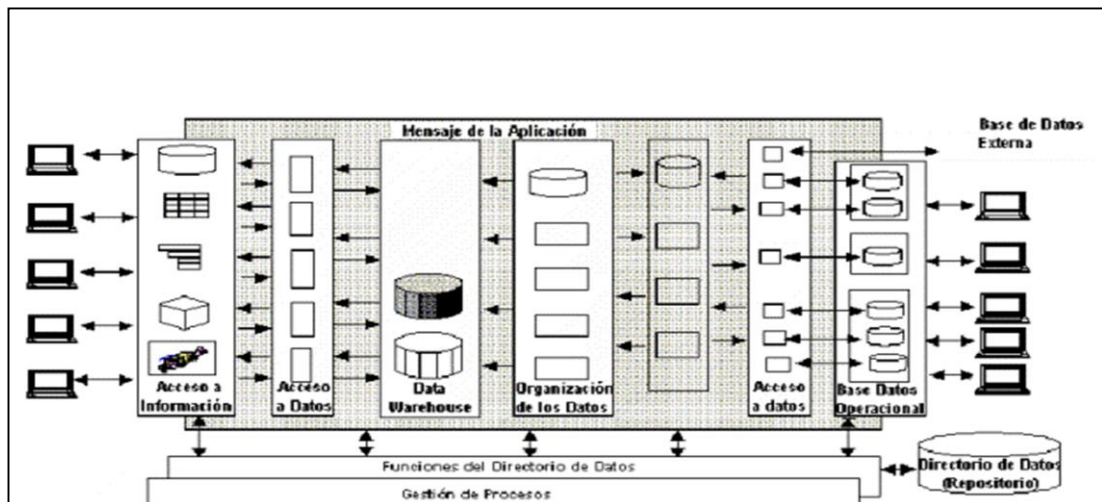
Warehouse.

- Una guía para mapón de datos cede como se transformar del ambiente operacional al del Data Warehouse.
- Un guía de los algoritmos usados para la esquematización entre el detalle de los datos actuales, con los datos ligeramente resumidos y estos, con los datos completamente resumidos, etc.
- La metadata juega un papel mucho más importante en un ambiente data warehousing que es un operacional clase.
- La metadata contiene (al menos): la estructura de los datos, los algoritmos usados para la esquematización y el mapping desde el ambiente operacional al Data Warehouse.

2.3.4 Arquitectura de un Data Warehouse

Una de las razones por las que el desarrollo de un Data Warehouse crece rápidamente, es que realmente es una tecnología muy entendible. De hecho, data warehousing puede representar mejor la estructura amplia de una empresa para administrar los datos informacionales dentro de la organización. A fin de comprender como se relacionan todos los componentes involucrados en una estrategia data warehousing es esencial tener una arquitectura Data Warehouse.

Grafico 5 Arquitectura de una Data Warehouse



Fuente: <http://www.monografias.com/trabajos90/datawarehouse-kimball-y-sql-2005/datawarehouse-kimball-y-sql-2005.shtml>

2.3.5 Arquitectura de un Data Warehouse

Una arquitectura de un Data Warehouse (Data Warehouse arquitectura - DWA) es una forma de representar la estructura total de datos, comunicación, procesamiento y presentación, que existe para los usuarios finales que disponen de una computadora dentro de la empresa.

- **Base de datos operacional/ nivel de base de datos externo:** los sistemas operacionales procesan datos para apoyar las necesidades operacionales críticas, para hacer eso, se han creado las bases de datos operacionales históricas que proveen una estructura de procesamiento eficiente, para un número relativamente pequeño de transacciones comerciales bien definidas.

Sin embargo, a causa del enfoque limitado de los sistemas operacionales, las bases de datos diseñadas para soportar estos sistemas, tienen dificultad al acceder a los datos para otra gestión o propósito informático. Esta dificultad en acceder a los datos operacionales es amplificada por el hecho que muchos de

estos sistemas tiene de 10 a 15 años de antigüedad. El tiempo de algunos de estos sistemas significa que la tecnología de acceso a los datos disponibles para obtener los datos operacionales, es así mismo antigua.

Ciertamente, la meta de la data Warehousing es liberar la información que es almacenada en bases de datos operacionales y combinarlas con la información desde otra fuente de datos, generalmente externa.

Cada vez más, las organizaciones grandes adquieren datos adicionales desde bases de datos externas. Esta información incluye tendencias demográficas, económicas, adquisitivas y competitivas (que pueden ser proporcionadas por instituciones oficiales u otros medios de información.

- **Nivel de acceso a la información:** el nivel de acceso a la información de la arquitectura del Data Warehouse, es el nivel del que el usuario final se encarga directamente, en particular, representa las herramientas que el usuario final normalmente usa día a día. Por ejemplo: Excel, Access, Sas, Spss.

Este nivel también incluye el hardware y software involucrados en mostrar información en la pantalla y emitir reportes de impresión, hojas de cálculo, gráficos y diagramas para el análisis de presentación. Hace dos décadas que el nivel de acceso a la información se ha expandido enormemente, especialmente a los usuarios finales quienes se han volcado a los PC monousuario y a los PC en redes.

Actualmente, existen herramientas más y más sofisticadas para manipular, analizar y presentar los datos, sin embargo, hay problemas significativos al tratar de convertir los datos tal como han sido recolectados y que se

encuentran contenidos en los sistemas operacionales en información fácil y transparente para las herramientas de los usuarios finales. Una de las claves para esto es encontrar un lenguaje de datos común que se puede usar a través de toda la empresa.

- **Nivel de acceso a los datos:** el nivel de acceso a los datos de la arquitectura Data Warehouse está involucrado con el nivel de acceso a la información conversando en el nivel operacional. En la red mundial de hoy, el lenguaje de datos común que ha surgido es SQL originalmente, SQL fue desarrollado por IBM como un lenguaje de consulta pero en los últimos veinte años ha llegado a ser el estándar para el intercambio de datos.

Uno de los adelantos claves de los últimos años ha sido el desarrollo de una serie de “filtros” de acceso a datos, tales como SQL para acceder a casi todos los sistemas de gestión de base de datos (Data Base) Management System – DBMSs) y sistemas de archivos de datos, relacionales o no. Estos filtros permiten a las herramientas de acceso a la información, acceder también a la data almacenada en sistemas de gestión de base de datos que tiene veinte años de antigüedad.

El nivel de acceso a los datos no solamente conecta DBMSs diferentes y sistemas de archivos sobre el mismo hardware, sino también a los fabricantes y protocolos de red. Una de las claves de una estrategia data warehousing es proveer a los usuarios finales con “acceso a datos universales”.

El acceso a los datos universales significa que, teóricamente por lo menos, los usuarios finales sin tener en cuenta la herramienta de acceso a la información o ubicación, deberían ser capaces de acceder a cualquier o todo

los datos en la empresa que es necesaria para ellos, para su trabajo.

El nivel de acceso a los datos entonces es responsable de la interfaces entre las herramientas de acceso a la información y las bases de datos operacionales. En algunos casos, esto es todo lo que un usuario final necesita. Sin embargo, en general, las organizaciones desarrollan un plan mucho más sofisticado para el soporte del data warehousing.

- **Nivel de directorio de datos (Metadata):** a fin de proveer el acceso a los datos universales, es absolutamente necesario mantener alguna forma de directorio de datos o repositorio de la información metadata. La metadata es la información alrededor de los datos dentro de la empresa.

A fin de tener un deposito totalmente funcional, es necesario tener una variedad de Metadata disponible, información sobre las visitas de datos de los usuarios finales e información sobre las bases de datos operacionales idealmente, los usuarios finales deberían de acceder a los datos desde el Data Warehouse(o desde las bases de datos operacionales), sin tener que conocer donde residen los datos o a la forma en que se han almacenados.

- **Nivel de gestión de procesos:** el nivel de gestión de procesos tiene que ver con la programación de diversas tareas que debe realizarse para contribuir y mantener la Data Warehouse y la información del directorio de datos. Este nivel puede depender del alto control de trabajo para muchos procesos (procedimientos) que deben ocurrir para mantener la Data Warehouse actualizado.
- **Nivel de mensaje de la aplicación:** el nivel de mensaje de la aplicación tiene

que ver con el transporte de la información alrededor de la red de la empresa. El mensaje de aplicación se refiere también como “subproducto”, pero puede involucrarse con solo protocolos de red. Puede usarse por ejemplo, para aislar aplicaciones operacionales o estrategias a partir del formato de datos exacto, recolectar transacciones o los mensajes y entregarlos a una ubicación segura en un tiempo seguro.

- **Nivel de Data Warehouse (físico):** en la Data Warehouse (núcleo) es donde ocurre la data actual, usada principalmente para usos estratégicos. de algunos casos, uno puede pensar del Data Warehouse simplemente como una vista lógica o virtual de datos. En muchos ejemplos, la Data Warehouse puede no involucrar almacenamiento de datos.

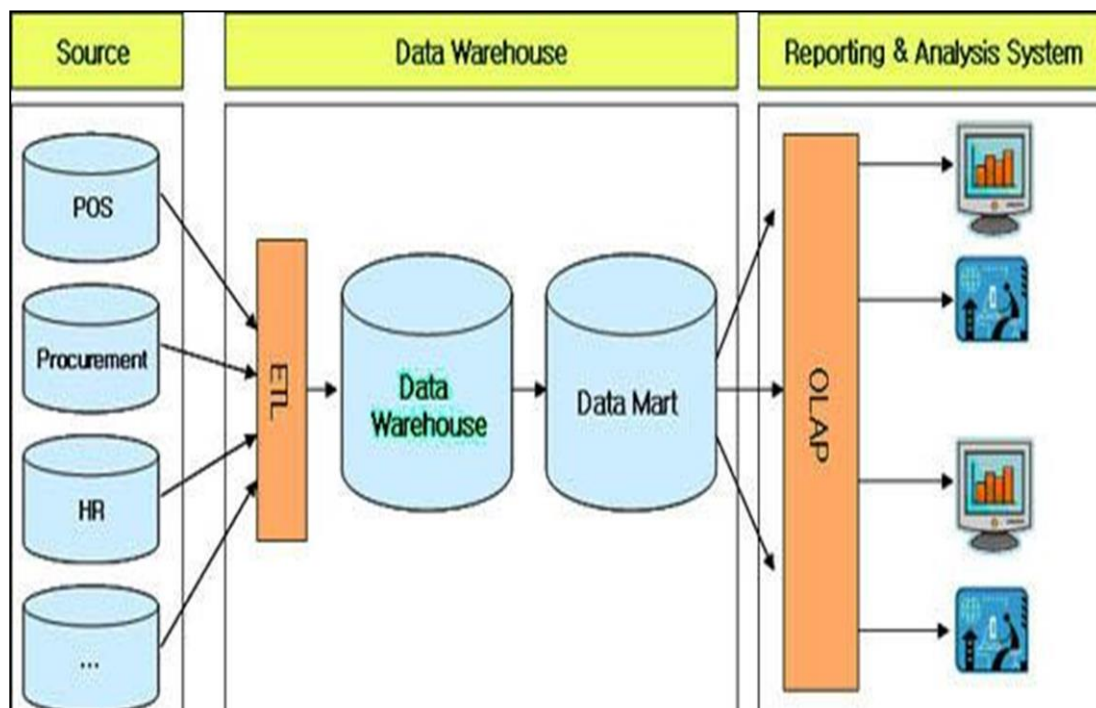
En una Data Warehouse físico, copias, en algunos casos, muchas copias de datos operacionales y/o externos, son almacenadas realmente en una forma que es fácil de acceder y es altamente flexible. Cada vez más, los Data Warehouse son almacenados sobre plataformas cliente/servidor, pero por lo general se almacenan sobre ³mainframes.

- **Nivel de organización de datos:** el componente final de la arquitectura Data Warehouse es la organización de los datos. Se llama también gestión de copia o replica, pero de hecho, incluye todos los procesos necesarios como seleccionar, editar, resumir, combinar y cargar datos en el depósito y acceder a la información desde bases de datos operacionales y/o externas la organización de datos involucra con

³ Mainframes.- es una computadora grande, potente y costosa usada principalmente por una gran compañía para el procesamiento de una gran cantidad de datos; por ejemplo, para el procesamiento de transacciones bancarias.

Frecuencia aun programación compleja, pero cada vez más están creándose las herramientas data.

Grafico 6 Organización de datos en Data Warehouse



Fuente:<http://www.monografias.com/trabajos75/inteligencia-negocios/inteligencia-negocios2.shtml>.

2.4 SISTEMAS OPERACIONALES

Los datos administrados por los sistemas de aplicación operacionales son la fuente principal de datos para el Data Warehouse las bases de datos operacionales se organizan como archivos indexados (UFAS, VSAN), Bases de datos de redes/jerárquicas (I-D-S/2, IMS, IDMS) o sistemas de base de datos relacionales (DB2, Oracle, Informe, etc).

2.5 DIMENSIONES

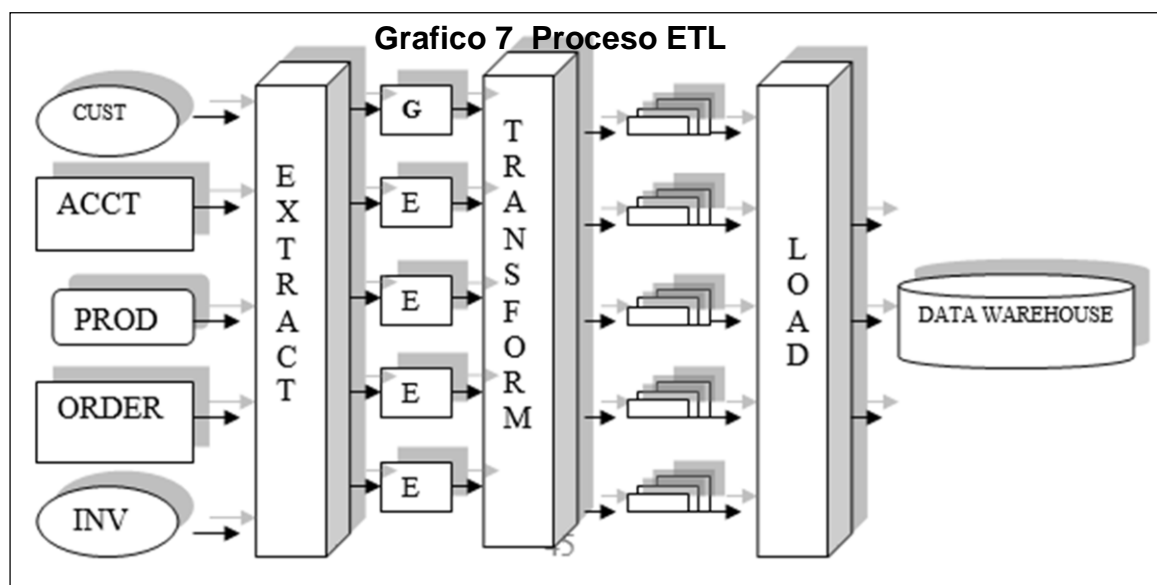
Son objetos del negocio con los cuales se puede analizar la tendencia y el comportamiento del mismo. Las definiciones de las dimensiones se basan en políticas de la compañía o el mercadeo, e indicar la manera en que la organización interpreta o clasifica su información para segmentarla el análisis de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

2.6 ETL (EXTRACT TRANSFORM LOAD)

La migración de los datos desde las fuentes operacionales a la Data Warehouse requiere la necesidad de procesos para extraer, transformar y cargar los datos, la actividad que se conoce como ETL.

La razón por la cual se requiere estos procesos se origina en una necesidad de reformatear, conciliar y limpiar los datos de origen. Sería conveniente que las transformaciones a los datos se apliquen solo una vez y que se reflejen en los archivos y bases de datos que actúan como fuente de datos.

La mayoría de los datos de origen son los datos operacionales actuales, para que de estos puedan ser datos históricos archivados, es importante considerar este tipo de datos al igual que los mencionados anteriormente ya que con estos datos históricos se podría realizar trabajos de minería de datos.



Fuente: <http://blog.gfi.es/metadata-etl-generico-para-bi-y-su-aplicacion-en-pentaho/>

2.7 METADATA

Otro paso necesario es crear la metadata. La metadata (es decir, datos acerca de datos) describe los contenidos del Data Warehouse. La metadata contiene de definiciones de los elementos de datos en el depósito, sistema de los elementos fuente como la data, se integra y transforma antes de ser almacenada en información similar.

2.8 MEDIDAS O MÉTRICAS

Son características cualitativas o cuantitativas de los datos que se desea analizar en las empresas. La medidas cuantitativas están dadas por valores cifras porcentuales, por ejemplo, la venta en dólares, cantidad de unidades en stock , cantidad de unidades de producto vendidas horas trabajadas el promedio de piezas producidas, el porcentaje de aceptación de un producto, el consumo de combustible de un vehículo, etc.

2.9 JERARQUÍAS DE DIMENSIONES Y NIVELES

Generalmente las dimensiones se estructuran en jerarquías, y cada jerarquía existen uno más niveles, los llamados niveles de agregación o simplemente niveles. Toda dimensión tiene por lo menos una jerarquía con un único nivel

2.10 MINERÍA DE DATOS

Las técnicas de minería de datos persiguen el descubrimiento automático del conocimiento contenido en la información almacenada de modo ordenado en grandes bases de datos. Estas técnicas tienen como objetivo descubrir patrones, perfiles y tendencias a través del análisis de los datos utilizando tecnologías de reconocimiento de patrones, redes neuronales, lógica difusa, algoritmos genéricos y otras técnicas avanzadas de análisis de datos.

Pero la minería de datos es solo una etapa del proceso de extracción de conocimiento a partir de datos (KDD) Knowledge Discover Data. Este proceso consta de varias fases como la preparación de datos (selección, limpieza y transformación), su exploración y auditoria. Minería de datos propiamente dicha (desarrollo de modelos y análisis de datos). Evaluación, difusión y utilización de modelos (outputs). Además, el proceso de extracción del conocimiento incorpora diferentes técnicas (árboles de decisión, regresión lineal, redes neuronales artificiales, técnicas bayesianas, máquinas de soporte vectorial, etc.) de campos diversos (aprendizaje automático e inteligencia artificial, estadística, base de datos, etc.) y aborda una tipología variada de problemas (clasificación, categorización, estimación/regresión, agrupamiento, etc.)

2.11 DATA WAREHOUSE Y DATA MINING

Una vez almacenada la información de un Data Warehouse (almacén de datos), se aplican sobre ella las técnicas de Data Mining (minería de datos) de modo óptimo. Los almacenes de datos constituyen una organización ideal de la información para aplicar sobre ella las técnicas de extracción del conocimiento o minería de datos.

2.12 SISTEMA TRANSACCIONAL

Un sistema de información es un conjunto de recursos humanos, materiales, financieros, tecnológicos, normativos y metodológicos, organizado para brindar a quienes paran y a quienes adoptan decisiones en una organización, la información que requieren para desarrollar sus respectivas funciones.

2.13 SISTEMA DE GESTIÓN DE BASE DE DATOS

Software que controla la organización, almacenamiento, recuperación, seguridad e integridad de los datos en una base de datos, acepta pedidos de datos desde un programa de aplicación y le ordena al sistema operativo transferir los datos apropiados.

Cuando se usa un sistema de gestión de base de datos, SGBD, (en inglés DBMS), los sistemas de información pueden ser cambiados más fácilmente a medida que cambien los requerimientos de la organización. Nuevas categorías de datos pueden agregarse a la base de datos sin dañar el sistema existente.

2.14 SQL (STRUCTURED QUERY LANGUAGE)

Lenguaje de interrogación normalizado para bases de datos relacionales. El SQL es un lenguaje de alto nivel, normalizado, que permite la consulta y actualización de los datos de BD relacionales. Se ha convertido en el estándar para acceder a BD relacionales. La primera versión se aprobó como norma ISO en 1987 y la segunda conocida como SQL2. Actualmente se trabaja con la norma SQL3 que soporta las bases de datos orientados a objetos.

2.15 OPERACIONALIZACION DE VARIABLES

Teniendo en cuenta nuestra hipótesis y nuestros objetivos tenemos el siguiente esquema de Operacionalización de variables.

CUADRO Nº 01
Operacionalización de variables

Variables	Dimensiones	Indicadores	Categorías	Instrumentos
variable independiente				
Implementación de Data Warehouse	Diseño del Sistema	Interfaz de usuario	Muy Bueno	Encuestas Estadísticas
			Bueno	
			Regular	
			Malo	
			Muy Malo	
		Facilidad de uso	Fácil	Encuestas Estadísticas
		Regular		
		Difícil		
	evaluación de sistema	Consistencia	Consistente	Encuestas Estadísticas
			Deficiente	
			Inconsistente	
		información a tiempo real	Muy Bueno	Encuestas Estadísticas
Bueno				
Regular				
Malo				
	Muy Malo			
variable dependiente				
Toma de decisiones red de salud puno	consultas en el sistema	tiempo de demora	Alta	observación
			Medio	
			Baja	
		Versatilidad	Alta	cuestionario y observación
			Medio	
			Baja	
	Eficacia	Alta	cuestionario y observación	
		Medio		
		Baja		

Fuente: Elaboración del investigador.

CAPÍTULO III

MATERIALES Y MÉTODOS

3.1 POBLACIÓN

Para este proyecto se toma como población a todos los usuarios involucrados en el manejo del Data Warehouse de la Red de Salud Puno – 2013, quienes realizan reportes, análisis de resultados y toma de decisiones.

3.2 MUESTRA

Se toma como muestra a 9 empleados de la Red de Salud Puno - 2013, quienes tienen a su cargo las diferentes Estrategias de Salud⁴ y entre ellos personal de la oficina de Estadística e Informática quienes son los usuarios frecuentes del Data Warehouse. Esta muestra nos llevara a realizar el análisis del Data Warehouse e interpretar sus resultados con métricas de Eficiencia, Confiabilidad, Mantenibilidad, Portabilidad y Funcionalidad.

3.3 MÉTODOS DE RECOPIACIÓN DE DATOS

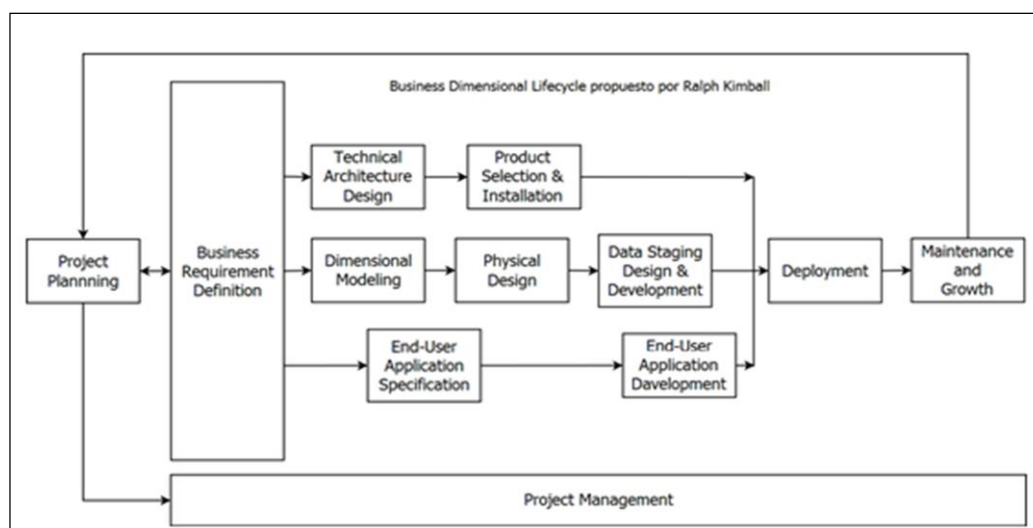
La metodología que se utilizo en este proyecto fue de Ralph Kimball, llamada “Modelo Dimensional (Dimensional Modeling)”, es un Ciclo de Vida Dimensional del Negocio (Business Dimensional Lifecycle). Este Modelo.

⁴ Estrategias de Salud: se considera a una unidad sobre problemas de salud que, requieren un abordaje integral, que tenga en cuenta todos los aspectos relacionados con la asistencia sanitaria así como la coordinación con otros departamentos o instituciones.

Dimensional se realizó en base a modelo de tablas y relaciones con el propósito de optimizar la toma de decisiones

El tamaño y el alcance del Data Warehouse se enfoca al Modelo Dimensional que está compuesto por una tabla con una llave combinada llamada tabla de hechos, y con un conjunto de tablas más pequeñas llamadas tablas de dimensiones⁵. Con este modelo se implemento el Data Warehouse, así como se realizo cambios en los Datos específicamente en su metadata. La expresión Ciclo de Vida se refiere a todos los pasos del proceso completo de desarrollo de Data Warehouse: planeación, diseño, codificación, prueba, implementación y administración, el ciclo de vida de Kimball es una metodología paso a paso para diseñar, desarrollar y desplegar Data Marts y Data Warehouse, y que es enriquecido además, por experiencias de otros coautores. Se muestra el proceso de construcción de inicio a fin de forma iterativa y también se muestra la utilización de las técnicas del modelado dimensional.

Grafico 8 ciclo de vida del Data Warehouse



Fuente:<http://marielitaortega.wordpress.com/2012/02/12/ciclo-de-vida-de-un-data-warehouse/>

⁵ <http://www.monografias.com/trabajos90/Data Warehouse-kimball-y-sql-2005/Data Warehouse-kimball-y-sql-2005.shtml#metodologia>

Esta metodología tiene como objetivo hacer que el Data Warehouse sea eficiente y se centra principalmente a una secuencia de procesos que al pasar el tiempo hacen que los datos sean versátiles y muy compactos, empieza con la planeación, gestión de proyectos y obtención de requerimientos. El Data Warehouse también soporta cambios inesperados en el comportamiento del usuario y es fácilmente extensible porque este modelo permite la extensión de nuevas tablas de dimensiones en base a nuevos requerimientos para insertar elementos o nueva data inesperada.

3.4 MÉTODO DE TRATAMIENTO DE DATOS

Para llevar a cabo la implementación del Data Warehouse se recolecto datos del HIS (Health Information System) que actualmente almacena todos los registros de información de las atenciones y atendidos de cada Puesto de Salud, a si mismo se normalizo los datos con el fin de tener mayor información dentro del Data Warehouse, estos Datos inicialmente estuvieron en formato BDF que posteriormente se introdujo en una Base de datos SQL Server 2012 puesto que el Data Warehouse se desarrollo en esta herramienta.

3.5 MATERIAL EXPERIMENTAL

Se tuvo como material experimental todos los datos del HIS (Health Information System) datos de los registros diarios de atenciones por Puesto de Salud recopilados durante el último año de la Red de Salud Puno - 2013, es necesario tener en cuenta que al tener estos datos se guarda estricta confidencialidad ya que cada registro representa el diagnostico a un paciente, este registro lleva su Documento de Identificación, Historia Clínica , Edad Sexo

y otras variables e indicadores exclusivamente de los pacientes, por tal motivo solo se usara los datos en carácter de Data Warehouse mas no de Minería de Datos, estos datos se procesaron con SQL Server 2012.

3.6 METODOLOGÍA DEL DESARROLLO

3.6.1 Análisis y Diseño de la Metodología del Data Warehouse.

La esencia del Data Warehouse en si es un almacén de datos que nos ayuda a tomar decisiones. La metodología de Ralph Kimball rescata y adecua al entorno del problema de la organización sin dejar de lado los importantes aportes de las metodologías de “Barry Devlin” y “William Inmon”, por esta razón se está aplicando esta metodología llamada “**dimension modeling**”.

Se determinó esta metodología porque resume el diagrama entidad relación del HIS y también esquematiza el Data Warehouse con nuestras Tablas de Hechos y Dimensiones, estas tablas se detallaran más adelante.

Condiciones a tener en cuenta para nuestro Data Warehouse.

- Se considera la base de datos HIS procesada, se entiende como procesada, que la base de datos este limpia y con campos que nos brinden información.
- Los servidores de las herramientas deben de estar activadas, así mismo deben tener contraseñas que sean de fácil acceso para no tener inconvenientes al momento de implementarlas.

Se estructura la metodología en etapas más que en pasos, pues los pasos pueden variar en cada proyecto, pero no así las etapas que se involucran en ella.

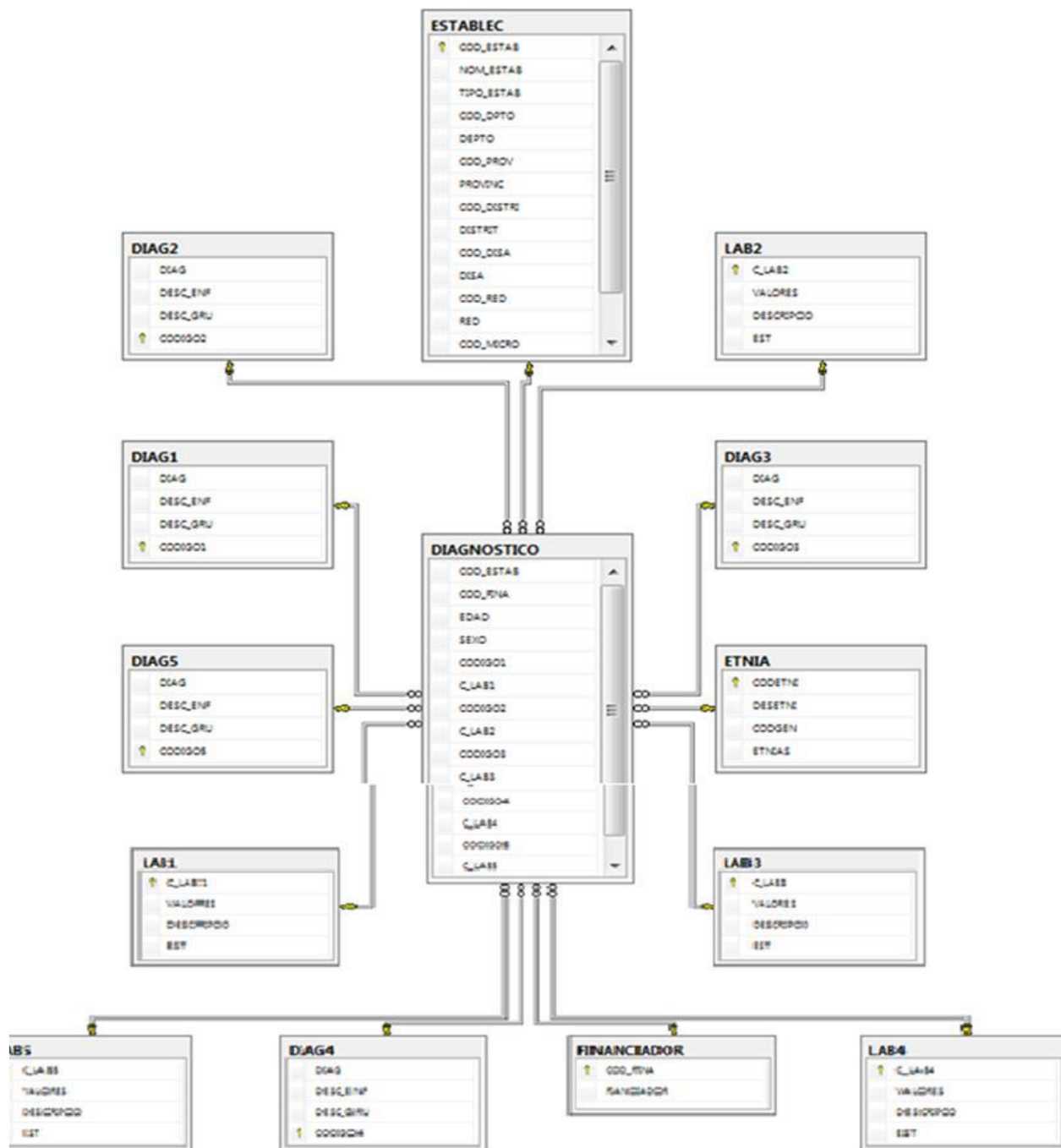
3.6.2 Implementación del Data Warehouse

Esta etapa establece nuestro marco de trabajo, estándares y procedimientos a seguir para la construcción del Data Warehouse, el objetivo de la arquitectura del Data Warehouse es integrar al Data Warehouse según la necesidad de la Red de Salud Puno, utilizamos la arquitectura de Ralph Kimball “arquitectura estructurada” ya que es sencilla y está acorde a nuestro tema de tesis.

En cuanto al modelo de datos del Data Warehouse se presenta el modelo dimensional que es más adecuado porque resume el modelo de entidad relación de la Red de Salud Puno haciendo que sea más eficiente en nuestros reportes.

Modelado Dimensional.- Utilizaremos el modelo estrella y no el modelo copo de nieve porque no tenemos volúmenes de datos que crecen exponencialmente pues solamente se trabaja con información de la Red de Salud Puno mas no con información Regional, en caso que la cantidad de datos que se prevé crezca demasiado, se puede contemplar el modelado copo de nieve o según la granularidad que se muestre en nuestro modelo.

Grafico 9 Esquema en Estrella



Fuente.- Elaboración del Investigador

Este diagrama es el Esquema en Estrella del Modelo Dimensional propuesto por Kimball, este grafico fue realizado con el Software SQL Server Management studio, se trabajó con esta arquitectura porque no tenemos una cantidad de Datos considerable y si es que este fuese el caso se tendrá que agregar nuevas

de dimensiones realizando con el enfoque copo de nieve.

Descripción del gráfico

El grafico cuenta con tablas hecho y tablas dimensiones, en este caso tenemos una tabla llamada hecho “Diagnostico” y nuestras tablas Dimensiones.

Tabla hechos.- Nuestra tabla de hechos es llamada “diagnostico” esta tabla tiene métricas, estas métricas puede ser sumarias y nos ayuda a ver el total de datos, que tendríamos dentro de nuestros cubos, las métricas de por si tiene características aritméticas que por defecto al sumar los datos y sus cantidades nos presenta los totales de registros o de códigos que se llenan en nuestro Data Warehouse.

Tabla de Dimensiones.- Llamamos Dimensiones a la tabla “Diag 1, Lab1, Etnia, Financiator”, estas Dimensiones nos permiten evaluar a nuestra tabla hechos y a su vez hacer posible la agrupación de ellos, teniendo jerarquías, grupos e identificando los tipos de enfermedades así como la ubicación de los establecimientos. También se puede decir que las dimensiones en general son indispensables para el análisis y lectura de nuestros reportes.

Grafico 10 Dimensión Establecimiento

Nombre de columna	Tipo de datos	Permitir val...
COD_ESTAB	int	<input type="checkbox"/>
NOM_ESTAB	varchar(100)	<input checked="" type="checkbox"/>
TIPO_ESTAB	int	<input checked="" type="checkbox"/>
COD_DPTO	int	<input checked="" type="checkbox"/>
DEPTO	varchar(50)	<input checked="" type="checkbox"/>
COD_PROV	int	<input checked="" type="checkbox"/>
PROVINC	varchar(100)	<input checked="" type="checkbox"/>
COD_DISTRI	int	<input checked="" type="checkbox"/>
DISTRIT	varchar(100)	<input checked="" type="checkbox"/>
COD_DISA	int	<input checked="" type="checkbox"/>
DISA	varchar(100)	<input checked="" type="checkbox"/>
COD_RED	int	<input checked="" type="checkbox"/>
RED	varchar(100)	<input checked="" type="checkbox"/>
COD_MICRO	int	<input checked="" type="checkbox"/>
MICRORED	varchar(100)	<input checked="" type="checkbox"/>

Fuente.- Elaboración del investigador en el Modelado Dimensional

Esta tabla es la tabla dimensión ESTABLEC esta tabla tiene como llave primaria a COD_ESTAB que luego pasara a ser llave foránea dentro de nuestra tabla de Hechos, también se tiene presente que esta tabla aporta información de lugar del centro de salud así como ubicación exacta.

Grafico 11 Dimensión Etnia

Nombre de columna	Tipo de datos	Permitir val...
CODETNI	int	<input type="checkbox"/>
DESETNI	varchar(20)	<input checked="" type="checkbox"/>
CODGEN	int	<input checked="" type="checkbox"/>
ETNIAS	varchar(20)	<input checked="" type="checkbox"/>

Fuente.- Elaboración del Investigador

Esta tabla Dimensión llamada ETNIA brinda información, como el paciente se siente identificado culturalmente si es que se considera Quechua u otra cultura peruana.


Grafico 12 Dimensión Financiador

	Nombre de columna	Tipo de datos	Permitir val...
	COD_FINA	int	<input type="checkbox"/>
	FIANCIADOR	varchar(100)	<input checked="" type="checkbox"/>

Fuente.- Elaboración del Investigador

Esta tabla lo que hace es dar información de Financiador ya sea del Essalud (Privado) o SIS (Publico) y no solamente si no que existe distinto tipo de financiamiento.


Grafico 13 Dimensión Diagnostico

	Nombre de columna	Tipo de datos	Permitir val...
	DIAG	varchar(50)	<input checked="" type="checkbox"/>
	DESC_ENF	varchar(150)	<input checked="" type="checkbox"/>
	DESC_GRU	varchar(150)	<input checked="" type="checkbox"/>
	CODIGO1	int	<input type="checkbox"/>

Fuente.- Elaboración del Investigador

Esta tabla nos ofrece información de el Diagnostico que se le proporcione al momento de su atención, así como el grupo de enfermedad a la que pertenece y la descripción de enfermedad.

Grafico 14 Dimensión Laboratorio (Dosis por Diagnostico)

	Nombre de columna	Tipo de datos	Permitir val...
	C_LAB1	int	<input type="checkbox"/>
	VALORES	varchar(4)	<input checked="" type="checkbox"/>
	DESCRIPCIO	varchar(100)	<input checked="" type="checkbox"/>
	EST	varchar(4)	<input checked="" type="checkbox"/>

Fuente.- Elaboración del investigador

Esta Tabla Dimensión Lab! Nos da información del número de dosis que se le receto por diagnostico así como la descripción del campo laboratorio según el registro HIS, también nos informa del estado del diagnóstico.

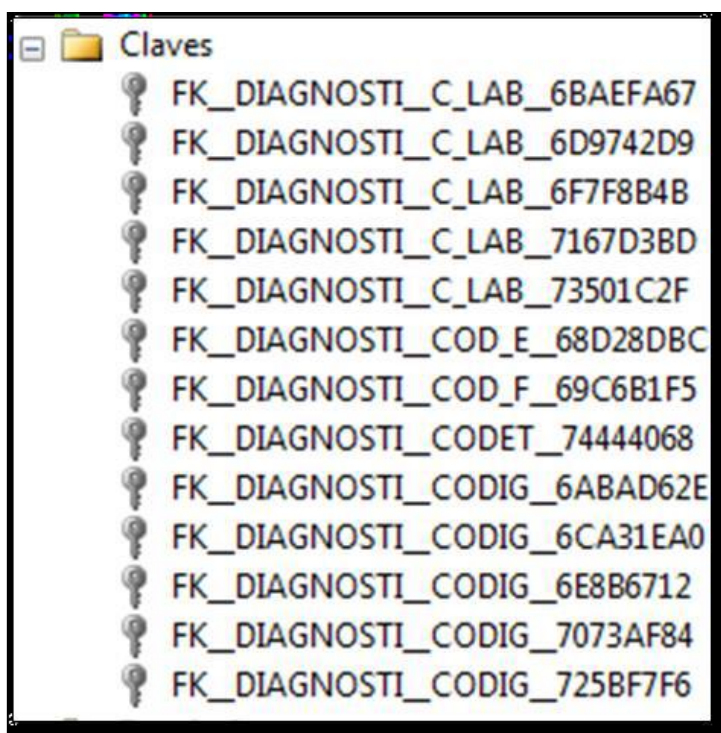
Grafico 15 Tabla de Hechos

Nombre de columna	Tipo de datos	Permitir val...
COD_ESTAB	int	<input type="checkbox"/>
COD_FINA	int	<input type="checkbox"/>
EDAD	int	<input type="checkbox"/>
SEXO	varchar(2)	<input type="checkbox"/>
CODIGO1	int	<input type="checkbox"/>
C_LAB1	int	<input type="checkbox"/>
CODIGO2	int	<input type="checkbox"/>
C_LAB2	int	<input type="checkbox"/>
CODIGO3	int	<input type="checkbox"/>
C_LAB3	int	<input type="checkbox"/>
CODIGO4	int	<input type="checkbox"/>
C_LAB4	int	<input type="checkbox"/>
CODIGO5	int	<input type="checkbox"/>
C_LAB5	int	<input type="checkbox"/>
CODETNI	int	<input type="checkbox"/>

Fuente.- Elaboración del Investigador.

Esta es nuestra tabla de Hechos y con ella podemos realizar nuestras métricas, cabe mencionar que esta tabla de hechos se estructura con sus llaves foráneas que son las siguientes:

Grafico 16 Claves Foráneas de la Tabla Hechos



Fuente.- Elaboración del Investigador.

Estas llaves foráneas nos ayudan a poder realizar nuestra búsqueda dentro de las dimensiones a las cuales se relacionan realizando esta tabla de hechos con SQL

```
CREATE TABLE DIAGNOSTICO(
```

```

COD_ESTAB INT FOREIGN KEY (COD_ESTAB) REFERENCES ESTABLEC NOT NULL, COD_FINA INT
FOREIGN KEY (COD_FINA) REFERENCES FINANCIADOR NOT NULL,
EDAD INT NOT NULL,
SEXO VARCHAR (2) NOT NULL,
CODIGO1 INT FOREIGN KEY (CODIGO1) REFERENCES DIAG1 NOT NULL, C_LAB1 INT FOREIGN KEY
(C_LAB1) REFERENCES LAB1 NOT NULL,
CODIGO2 INT FOREIGN KEY (CODIGO1) REFERENCES DIAG2 NOT NULL,
C_LAB2 INT FOREIGN KEY (C_LAB1) REFERENCES LAB2 NOT NULL, CODIGO3 INT FOREIGN KEY
(CODIGO1) REFERENCES DIAG3 NOT NULL,
C_LAB3 INT FOREIGN KEY (C_LAB1) REFERENCES LAB3 NOT NULL, CODIGO4 INT FOREIGN KEY
(CODIGO1) REFERENCES DIAG4 NOT NULL, C_LAB4 INT FOREIGN KEY (C_LAB1) REFERENCES LAB4
NOT NULL,
CODIGO5 INT FOREIGN KEY (CODIGO1) REFERENCES DIAG5 NOT NULL,
C_LAB5 INT FOREIGN KEY (C_LAB1) REFERENCES LAB5 NOT NULL, CODETNI INT FOREIGN KEY
(CODETNI) REFERENCES ETNIA NOT NULL,)
```

Donde se ve claramente las referencias de las tablas y las claves foráneas.

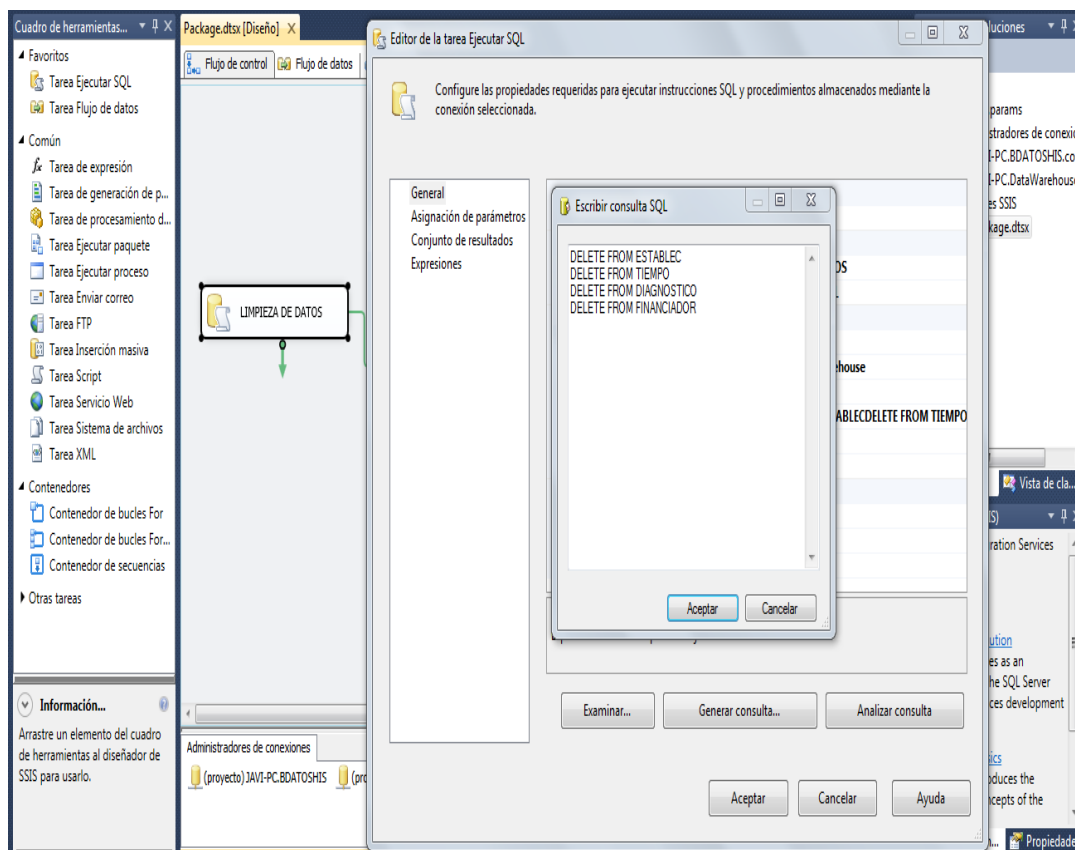
Enriquecimiento de los datos

En esta etapa utilizaremos La extensión de SQL Server llamada SQL Data Tools, una vez que se tiene en su lugar el modelo de datos del Data Warehouse es necesario poblar su estructura con sus datos respectivos, al decir poblar nos referimos a importar los datos a nuestro esquema dimensional, es necesario que dichos datos sufran procesos de transformación de limpieza, se extrae la data del HIS.

Esta es la etapa que más tarda y que más esfuerzo necesita, una consideración a tener en cuenta es la calidad de datos que se tiene que procesar ya que la calidad de datos hace que nuestro Data Warehouse sea más eficiente y que no se pueda perder ningún dato almacenado en nuestro Data Warehouse.

A continuación lo que se hace es limpiar nuestros datos, esto amerita que se tiene que eliminar cualquier dato que se tenga dentro de nuestras tablas de dimensiones, este procedimiento involucra trabajar dentro del data set y así poder extraer la data, primero eliminamos todos los datos de nuestras tablas dimensiones y nuestra tabla de hecho para así poblar nuestros Data Warehouse. Indicando las conexiones al origen en este caso es la misma conexión en SQL Cliente Services como clave de acceso por defecto de Windows.

Gráfico 17 Limpieza de Datos de Tabla de Dimensiones y Tabla de hechos



Fuente.- Elaboración del Investigador.

Se muestra el gráfico 12 donde se detalla y se ve mediante el código SQL la limpieza de nuestro Data Set con estos códigos.

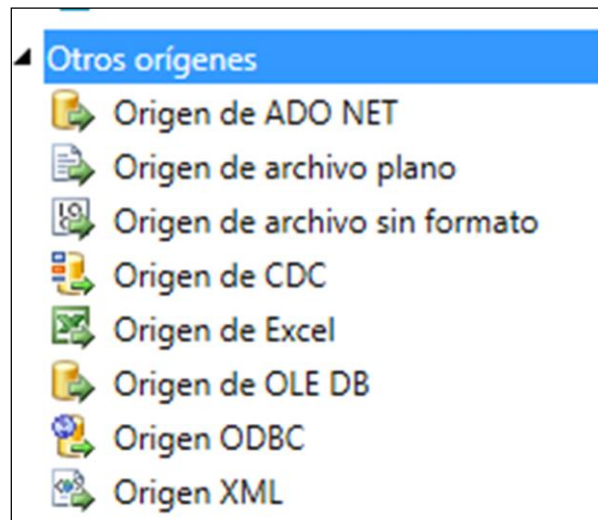
Ejemplo de limpieza de Datos: DELETE FROM ESTABLEC

DELETE FROM TIEMPO DELETE FROM DIAGNOSTICO DELETE FROM FINANCIADOR

Estos códigos nos indican que dentro de nuestras tablas Establec, Tiempo, Diagnostico, Financiadore, se eliminara todos los registros existentes, en eso consiste la limpieza de datos.

Luego de hacer la limpieza se tiene los orígenes de los datos que no es más que los datos del HIS

Grafico 18 Orígenes de Datos (ETL)

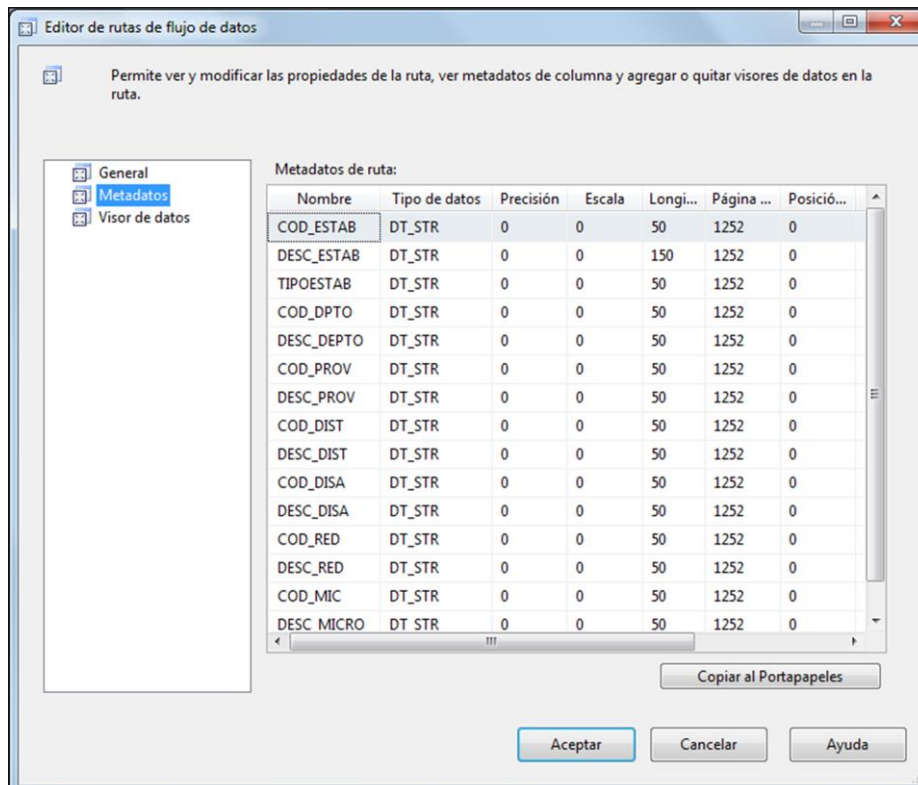


Fuente.- Elaboración del Investigador.

En este Grafico 13 se muestra que SQL Data Tools dispone de grupo de orígenes de datos disponibles que hace posible que sea eficaz la herramienta ya que podemos importar datos desde datos en Excel hasta datos en XML.

Nuestros orígenes serán nuestra base de datos HIS y sus respectivas tablas, que se tiene que exportar a una base de datos SQL, luego realizar mediante código SQL la transformación para luego tener un control en sus respectivas METADATAS que son importantes ya que sin ellas no había ningún tipo de transformación.

Grafico 19 Transformación de Metadatos para Dimensión Establecimiento



Permite ver y modificar las propiedades de la ruta, ver metadatos de columna y agregar o quitar visores de datos en la ruta.

Metadatos de ruta:

Nombre	Tipo de datos	Precisión	Escala	Longi...	Página ...	Posició...
COD_ESTAB	DT_STR	0	0	50	1252	0
DESC_ESTAB	DT_STR	0	0	150	1252	0
TIPOESTAB	DT_STR	0	0	50	1252	0
COD_DPTO	DT_STR	0	0	50	1252	0
DESC_DEPTO	DT_STR	0	0	50	1252	0
COD_PROV	DT_STR	0	0	50	1252	0
DESC_PROV	DT_STR	0	0	50	1252	0
COD_DIST	DT_STR	0	0	50	1252	0
DESC_DIST	DT_STR	0	0	50	1252	0
COD_DISA	DT_STR	0	0	50	1252	0
DESC_DISA	DT_STR	0	0	50	1252	0
COD_RED	DT_STR	0	0	50	1252	0
DESC_RED	DT_STR	0	0	50	1252	0
COD_MIC	DT_STR	0	0	50	1252	0
DESC MICRO	DT_STR	0	0	50	1252	0

Copiar al Portapapeles

Aceptar Cancelar Ayuda

Fuente.- Metadatos de Sql Server Tool.

En esta parte nos preguntamos si es que nuestra metadata está igual con nuestra metadata de nuestra tabla dimensión, se debe tener cuidado porque cuando se establece la transformación tiene que coincidir las características de nuestro metadata en nuestra tabla dimensiones.

Grafico 20 Cuadro de Asignaciones de la Tabla de Hechos

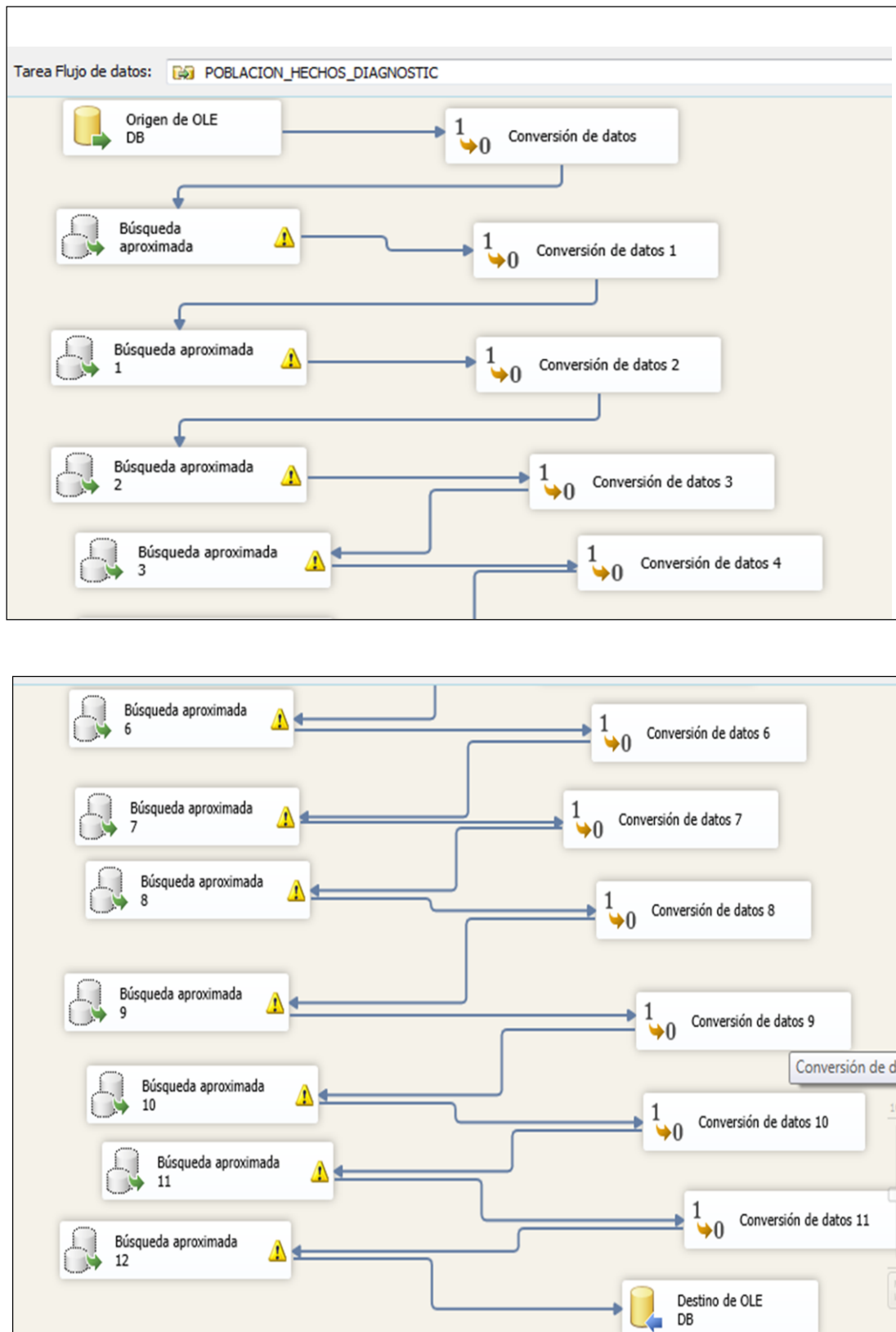
Columna de entrada	Columna de destino
COD_ESTAB	COD_ESTAB
Copia de DESC_ESTAB	NOM_ESTAB
Copia de TIPOESTAB	TIPO_ESTAB
Copia de COD_DPTO	COD_DPTO
Copia de DESC_DEPTO	DEPTO
Copia de COD_PROV	COD_PROV
Copia de DESC_PROV	PROVINC
Copia de COD_DIST	COD_DISTRI
Copia de DESC_DIST	DISTRIT

Fuente.- Sql Tools, Asignación y Poblamiento de Hechos

En este Grafico se muestra como se llena nuestra tabla hechos y las asignaciones correspondientes, cuando se ve que un atributo es copia de sí mismo esto quiere decir que este atributo ha sufrido una transformación y por ello se agiliza la implementación.

Se tiene que tener en consideración que en nuestra tabla de Hechos se encuentra nuestras métricas y nuestras FOREIGN KEY de todas nuestras tablas Dimensiones, esto nos ayudara a tener mejores reportes y que nuestro Data Warehouse sea óptima en desarrollo.

Grafico 21 Grafico de Asignaciones (ETL) y KEY a Tabla de Hechos



Fuente.- Sql Server tools, proceso Load a tabla Hechos

En este Grafico se muestra el origen de nuestro Data Set en este caso el HIS y las respectivas transformaciones que sufrió para llenar los datos en nuestra tabla Hechos, este esquema se realiza siempre y cuando tengan los Datos adecuados ya que este proceso es la base de todo nuestro Data Warehouse

Los procesos de conversión de datos nos ayuda a buscar nuestras Claves Foráneas que nos ayudaran en cruzar nuestras variables y así realizar nuestros reportes.

Luego de tener ya llenado nuestra tabla de hechos y dimensiones procedemos a la siguiente etapa que es la creación de cubos multidimensionales.

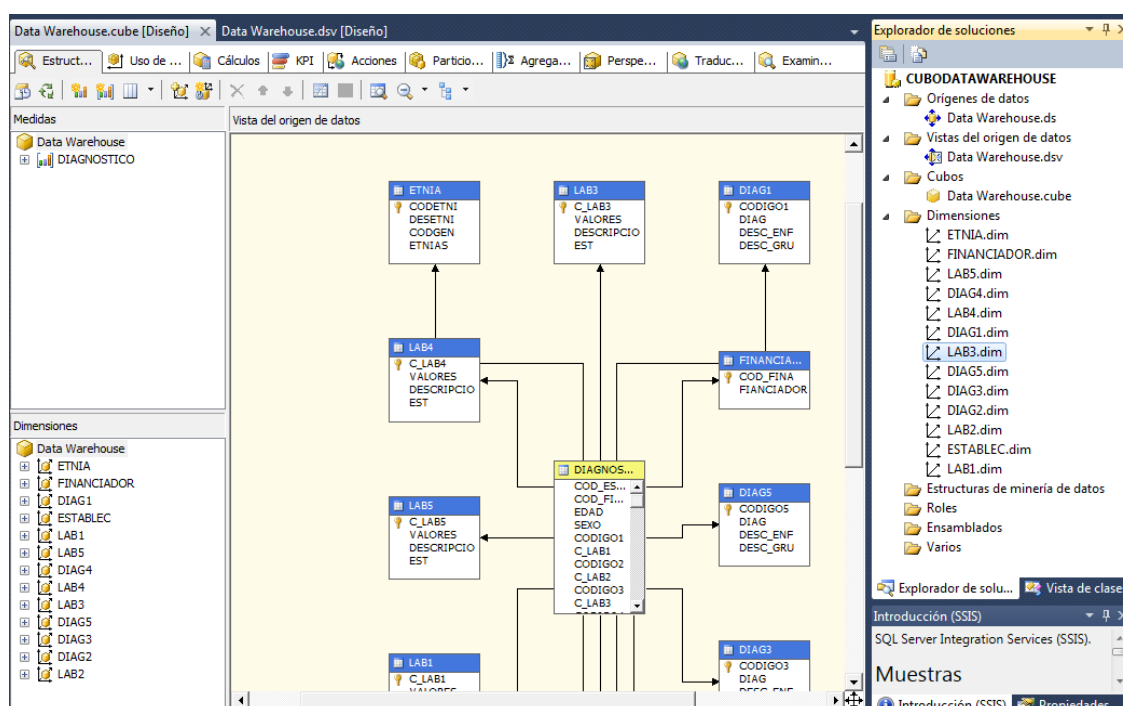
3.6.3 Diseñador de Cubos

Para desarrollar nuestro cubo utilizamos SQL ANALISYS SERVICES, la implantación del cubo no es más que la agrupación de la tabla de hechos y sus dimensiones en un solo campo haciendo posible que el reporte se pueda ver de distintas formas combinando muchas variables, para hacer posible el cubo se considera lo siguiente:

- 1.- Tabla de hechos con sus dimensiones.- estas dimensiones deben estar conectadas mediante un FOREIGN KEY (llave foránea) a la tabla de hechos para tener el más mínimo nivel de detalle.
- 2.- integración de los datos.- los datos tienen que estar debidamente integrados, teniendo cuidado en la metadata.

3.- servidores⁶: es común que Windows automáticamente desactive algunos de los servicios, se necesita que los servidores se encuentren habilitados y esto se hace entrando a administrador de tareas configuración de servicios y se activa los servicios de SQL Analsys Services.

Grafico 22. Entorno de Trabajo de SQL Analsys Services



Fuente.- Entorno de Trabajo de Sql Server 2012

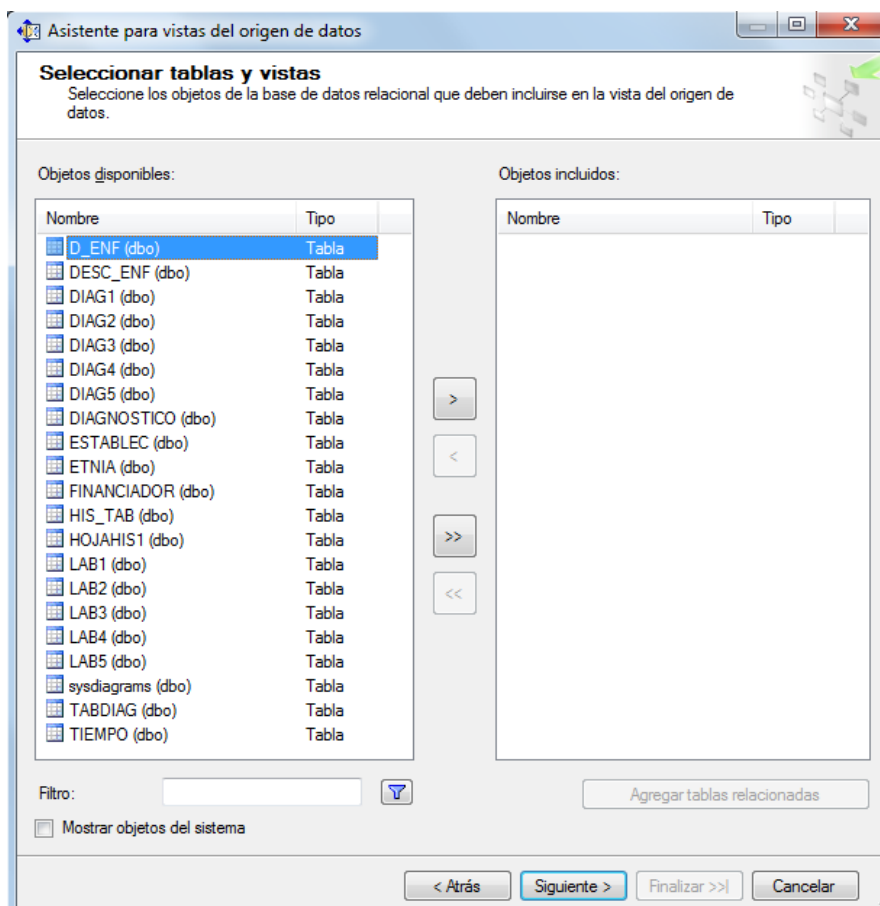
En el Grafico 22 se puede ver claramente las tablas de Dimensiones de color azul y la tabla de Hechos de color amarillo, y a si mismo a la derecha se ve el origen de los datos, las vistas de este origen y el cubo del Data Warehouse.

4. Orígenes de datos.- se refiere a nuestra tabla de hechos y sus dimensiones de donde se extraerá la data.

⁶ Servidor.- es un ordenador remoto que provee los datos solicitados por parte de los navegadores de otras computadoras.

5. Vista de orígenes de los datos.- se refiere a un marco visual de donde se extraerá la data, este Grafico se muestra a continuación.

Grafico 23 Vista de Orígenes de los Datos



Fuente.- Asistente para vistas de Orígenes de datos de Sql Server Data Tool

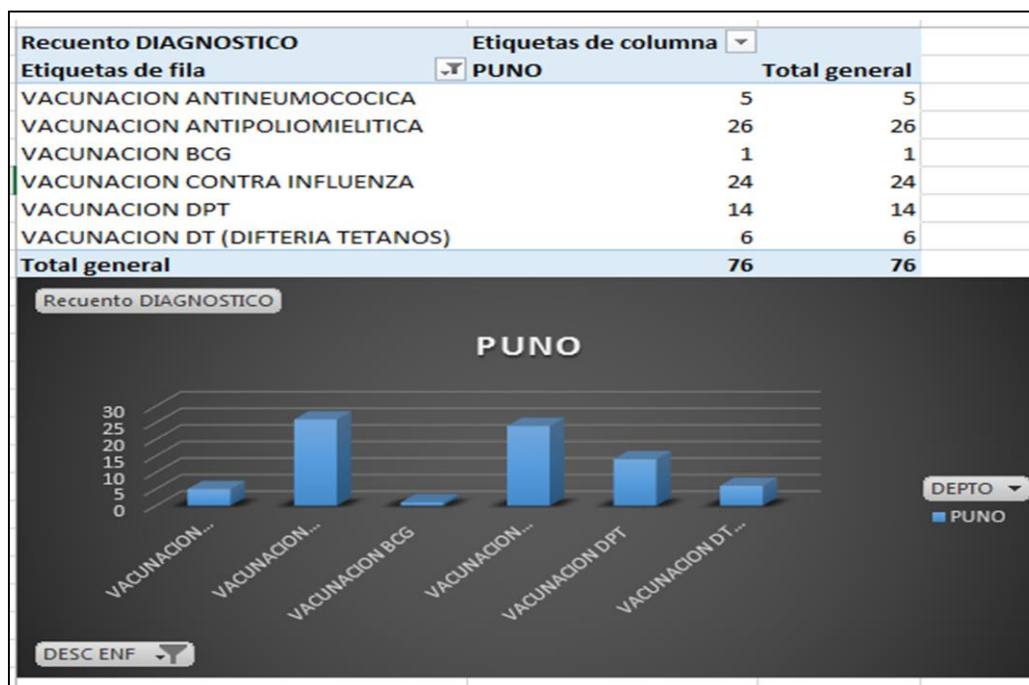
Se ve claramente todas las tablas que interactúan en el cubo, el cubo que se realizo tiene como nombre Data Warehouse Cube, teniendo ya y ejecutado este cubo se puede hacer los reportes que se requieren para nuestra toma de decisiones, los reportes originariamente se pueden realizar en SQL Reporting Services u otras softwares, para esta investigación se tomó en cuenta Microsoft Excel, porque los usuarios al utilizar los filtros y las tablas dinámicas harán posibles que los reportes sean mucho más explícitos y así se pueda tomar mejor las decisiones.

3.6.4 Reportes y Evaluación de información

Los usuarios accesan al Data Warehouse por medio de herramientas de productividad basadas en GUI (Graphical User Interface - Interfase gráfica de usuario). Pueden proveerse a los usuarios del data warehouse muchos de estos tipos de herramientas, tal es el caso de la Microsoft Excel con el que desarrollamos nuestros reportes y tomamos como interface de usuario.

Los reportes son fundamentales sobre todo al momento de dar información y ver la realidad de nuestra organización el uso de tablas dinámicas hacen que se puede realizar un cruce de variables y así poder tener datos en tiempo real y registros de actividad, dicho de esta manera se procede a realizar los reportes con la herramienta ya mencionado anteriormente.

Grafico 24 Tipo de vacunas y totales



Fuente.- Reporte de Microsoft Excel 2013 con tablas Dinámicas

El grafico 18 detalla los totales de las vacunas realizadas en la provincia de Puno

por día, para así tener conocimiento del número de dosis por vacuna. Se ve claramente que sobresale la Vacuna Antipoliomelitica con 26 dosis un indicador importante es la vacuna DT, tomando solamente este grafico se puede decir que se tiene que intensificar campañas de vacunas DT dentro de la población y así prever futuros problemas en salud.

La información que proporciona el Data Warehouse tiene sustento grafico y estadístico que hacen posible que la toma de decisiones sean más eficientes y confiables se puede tener un mayor control de metas alcanzadas por campaña de salud. Para tomar la decisión de realizar más campañas de salud es necesario tener en cuenta el tipo de variable al extraer y los indicadores a utilizar, podemos decir que nuestro Data Warehouse ayuda significativamente en la toma de decisiones.

Teniendo la información que proporciona el Data Warehouse se puede tomar decisiones comparando los resultados alcanzados con nuestras metas físicas⁷ y así decidir si es que se intensifican las campañas de salud o se realizan mas campañas en un determinado mes, esta informacion también nos permiten saber la cantidad de dosis por medicamento que se suministra y así comparar con la cantidades de dosis por medicamento que se adquiere para tener un mayor control. El interface de Nuestro Data Warehouse es Microsoft Excel ,para estos reportes puesto que es más sencillo de usar y no se requiere de conocimientos informáticos específicos, el Data Warehouse hace posible le manejo y reportes, haciendo también cruce de variables a las cuales se quiere sumarizar o realizar tablas estadísticas para tener una mejor visualización en información.

⁷ Metas físicas: es un objetivo propuesto en forma numérica a la que se alcanza en un determinado mes, semestre o año, que se propone una estrategia de salud.

CAPÍTULO IV

RESULTADOS Y DISCUSIONES

4.1 ANÁLISIS Y COMPRENSIÓN DEL ENTORNO ANALÍTICO

Para realizar la implementación del Data Warehouse se usó las bases de datos (HIS) estas bases de datos están hechas en el lenguaje de programación FOX PRO con extensión “.BDF”. De tal manera que al momento de extraer los datos estos pasaran con sus respectivas características, en el sistema HIS digitan los registros de atenciones, estos registros son llenados por el personal de salud y tienen 25 filas dentro de las cuales cada hoja HIS representa a una (UPS) Unidad Productora de Servicio, esto quiere decir que cada hoja puede ser de Medicina, enfermería, Obstetricia, Control y Crecimiento, Inmunizaciones, etc. Antes de digitar los datos dentro del sistema HIS se plantea realizar un control de calidad , estos controles se realizan al momento de recibir los registro hasta donde se sistematiza la información de tal manera que cuando se procesa y se realiza los reportes, estos puedan ser muy explícitos para la toma de decisiones.

Análisis de las tablas del HIS.

Como se conoce el sistema HIS está constituido por muchas tablas dentro de las cuales se resaltan las más importantes como las que se dan a continuación.

Base de Datos que contiene la base de datos HIS

NOMBRE DE ARCHIVO:BDATOSHIS	
Nombre de Tablas	Descripción
TABLAUPS	unidad productora de servicios
TABETNIA	identidad Cultural del Paciente
TABDIAG	Diagnósticos
T_ESTAB	Establecimiento
SERVSAL	Servicios de Salud
RED	Red
PROFESION	Profesión
SITUACIO	Situación de Salud
LOTES	Lores de HIS
HOJAHIS	Formato HIS
GRUPO	Grupo de Enfermedad
FINANCIA	Financiador de Salud
CODENF	Código de Enfermedad
CPT	Código de Clasificación de Enfermedad
COLEGIOS	Colegios Profesionales
CIE	Clasificación Internacional de Enfermedades

Fuente: base de datos HIS

Estas tablas brindan información importante para poder sacar reportes y por eso se detallan algunas. A continuación detallaremos las tablas más importantes que contiene el HIS

Tabla Unidad Productora de Servicio

Nombre del Archivo BdatosHIS			
Campos	Tamaño	Tipo de Dato	Descripción
cod_serv		INT	Identificador de UPS
desc_serv	60	VARCHAR	Descripción de la UPS
clave primaria : cod_serv clave foránea:			

Fuente: Tablas del HIS

Esta tabla contiene los datos de las Unidades Productoras de servicio, esto nos indica a qué tipo de agrupación pertenece la atención, es decir a que unidad pertenece ya sea (medicina, enfermería, odontología etc.)

Tabla Etnia Cultural

Nombre del Archivo BdatosHIS			
Campos	Tamaño	Tipo de Dato	Descripción
CODETNI		INT	Código de Etnia
DESCETNI	50	VARCHAR	Descripción de Etnia
CODGEN	10	VARCHAR	Código de Genero
ETNIAS	10	VARCHAR	Descripción General de Etnia
clave primaria : CODETNI			
clave foránea:			

Fuente: Tabla Etnia del HIS

En esta tabla se ve los campos para el llenado de la Tabla TABETNIA esta tabla da información de que manera el paciente se siente identificado, ya sea aymara, quechua, mestizo etc.

Tabla Establecimientos de Salud

Nombre del Archivo BdatosHIS			
Campos	Tamaño	Tipo de Dato	Descripción
COD_ESTAB		INT	Código de Establecimiento
DESC_ESTAB	100	VARCHAR	Descripción del Establecimiento
COD_2000		INT	Código de Ubicación
TIPOESTAB	10	VARCHAR	Tipo de Establecimiento
COD_DPTO	10	VARCHAR	Código de Departamento
COD_PROV	10	VARCHAR	Código de Provincia
COD_DIST	10	VARCHAR	Código de Distrito
COD_DISA	10	VARCHAR	Código de Dirección Racional de
COD_RED	10	VARCHAR	Código de Redes
COD_MIC	10	VARCHAR	Código de Microred
clave Primaria COD_ESTAB			
Clave Foránea			

Fuente: Tabla T_ESTAB de la Base de Datos HIS

En esta tabla se da a conocer información del lugar del establecimiento, el tipo de establecimiento si es Puesto de Salud o Centro de Salud, así como la ubicación del establecimiento, a que Dirección de Salud pertenece y la Micro Red a la cual pertenece.

Tabla Grupo de Enfermedad

Nombre del Archivo BdatosHIS			
Campos	Tamaño	Tipo de Dato	Descripción
COD_GRU	6	VARCHAR	Código de Grupo
COD_GRUX		INT	Código de Agrupación de Enfermedad
COD_CAP	80	VARCHAR	Código de Capítulos de Enfermedad
DESC_GRU	100	VARCHAR	Descripción de Grupo de enfermedad
clave primaria : COD_GRU			

Fuente: Base de Datos HIS

En esta tabla se da a conocer el Grupo de enfermedad, esta tabla agrupa las enfermedades según la codificación que el HIS le dio a este sistema.

Tabla Financiamiento

Nombre del Archivo BdatosHIS			
Campos	Tamaño	Tipo de Dato	Descripción
CF	11	INT	Código de Financiador
NF	50	VARCHAR	Descripción de Financiador de Salud
clave primaria : CF			
clave foránea:			

Fuente: Base de Datos HIS

En esta tabla lo que se ve es el tipo de financiamiento que se le da al paciente si es que es del sector público o privado. Por ejemplo Essalud, SIS, etc.

Tabla Código de Enfermedad según CIE- 10

Nombre del Archivo BdatosHIS			
Descripción			
Campos	Tamaño	Tipo de Dato	Descripción
CODIGO		INT	Código de Enfermedad Especifica
COD_CAT	10	VARCHAR	Código de Capitulo de enfermedad
COD_ENF	10	VARCHAR	Código de Enfermedad General
CLASE	2	VARCHAR	Código de Clase de Enfermedad
EST	2	VARCHAR	Código de Estado de Enfermedad
clave Primaria		CODIGO Clave Foránea	

Fuente: Base de Datos HIS

En esta tabla lo que se da a conocer es el código de enfermedad, es importante esta tabla ya que esta tabla aporta más información al sistema.

Tabla Diagnostico de Enfermedad

Nombre del Archivo BdatosHIS			
Campos	Tamaño	Tipo de Dato	Descripción
DIAG		INT	Código de Diagnostico
DESCCRIP	50	VARCHAR	Descripción del Código de Diagnostico
clave Primaria		DIAG Clave Foránea	

Fuente: Base de Datos HIS

Esta tabla son los Diagnósticos de las enfermedades y su descripción se relaciona con código de enfermedad.

Tabla Servicio de Salud

Nombre del Archivo BdatosHIS			
Campos	Tamaño	Tipo de Dato	Descripción
COD_SERVSA		INT	Código de Servicio
DESC_SERV	50	VARCHAR	Descripción de Servicio
COD_SERVSA Clave Foránea			

Fuente: Base de Datos HIS

Esta tabla son los servicios de salud que ofrece la Red de Salud Puno así como su descripción.

Tabla de la Red de Salud Puno

Nombre del Archivo BdatosHIS			
Campos	Tamaño	Tipo de Dato	Descripción
COD_DISA		INT	Código de Dirección de Salud
COD_RED		INT	Código de Red
DESC_RED	80	VARCHAR	Descripción de Red
clave Primaria COD_RED Clave Foránea COD_DISA			

Fuente: Base de Datos HIS

Esta tabla es la ubicación de la Dirección de Salud, esta tabla se relaciona mediante una llave foránea con T_ESTAB.

Análisis lógico.

Tomamos como ejemplo a la tabla establec

```

----- create table
ESTABLEC(
COD_ESTAB INT PRIMARY KEY NOT NULL, DESC_ESTAB VARCHAR
(100) NULL, COD_2000 INT NOT NULL
    
```

```
TIPO_ESTAB INT NULL, COD_DPTO VARCHAR(10) NULL, COD_PROV  
VARCHAR(10) NULL, COD_DIST VARCHAR(10) NULL, COD_DISA  
VARCHAR (10) NULL, COD_RED VARCHAR (100) NULL, COD_MICRO  
VARCHAR (10)INT NULL,)
```

El entorno del Data warehouse requiere de un Software diseñado para base de datos por ello se utilizó SQL Server 2012 para crear las bases de datos en las cuales se realiza la importación de ellos, SQL Server 2012 desde muchos años atrás trabajo en el entorno de Base de Datos y es confiable puesto que tiene experiencia y a su vez contiene muchos servidores para poder enlazarlos y esto hace más eficiente el trabajo.

Como SQL Server Tiene muchas Extensiones se utilizó para esta investigación las siguientes Extensiones:

- SQL Server Management Studio
- SQL Server Data Tools
- SQL Server Analysis Services

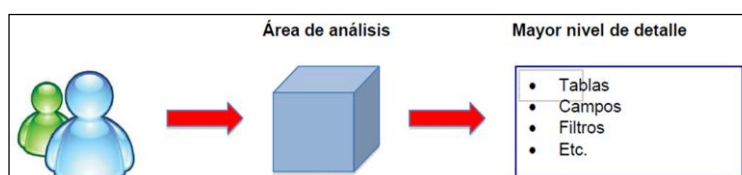
Estas Extensiones son necesarias para hacer nuestro Data Warehouse, nos permite realizar los procedimientos de ETL y Cubos, para realizar nuestros reportes utilizaremos Microsoft Excel, esto nos ayuda a visualizar mejor nuestros datos y así tomar una decisión con nuestra información.

4.2 PROTOCOLO DE ACCESO AL DATA WAREHOUSE

Para el acceso, visualización de información y elaboración de informes se define los perfiles de usuarios en función a los siguientes puntos

1. Seguridad de acceso a la aplicación, existen dos tipos de perfiles de usuarios
 - a. Usuarios de solo lectura.- estos usuarios solo podrán consultar informes/reportes, no pueden crear informes/reportes, no puede modificar informes/reportes.
 - b. Usuarios de lectura/ escritura.- estos usuarios podrán crear informes/reportes así como crear informes/ reportes y modificar estos informes/reportes.
2. Seguridad de acceso datos.

Se determina a nivel de detalle las áreas de análisis (medicamentos, estrategias, evaluaciones) que pueden visualizar cada perfil, así como tablas y campos y filtros sobre las mismas.



Rol 1: Lectura

Rol 2: Lectura / Escritura

Para la obtención de la información contenida en la Data se establece protocolos en el que deben intervenir todos los responsables de las distintas áreas de información contenidas en la base de datos. En tal sentido se organiza la estura

de funcionamiento con las siguientes asignaciones efectiva de responsabilidad que se definirán los dos tipos de usuario.

Rol 1: Lectura. Acceso a la información y obtención de informes. Las responsables de todas las estrategias creadas en la Red de salud puno.

Rol 2: Lectura y Escritura. Elaboración de informes y emisión, según la tipología de información esta tarea la lleva a cabo la Oficina de estadística e informática de la Red de salud puno que proporcionara el soporte necesario para la correcta gestión y administración de los datos. En cuanto a la seguridad de acceso, se cumple en los sistemas de Data Warehouse, que es preciso el implantar niveles de acceso a la información, realizando un plan completo de seguridad que contemple:

- Acceso a recursos de la red (local o intranet).
- Asignación de usuarios a grupos con perfiles de seguridad diferenciados.
- Asignación de niveles de autorización de aplicación a grupos de usuarios.
- Seguridad a nivel de Base de Datos, mediante los procedimientos provistos por las mismas.

4.3 EVALUACIÓN DEL DATA WAREHOUSE

Evaluamos al Data Warehouse utilizando la norma ISO 9116 “calidad de calidad de software” donde evaluamos principalmente:

- Funcionalidad.- Vemos que nuestro sistema satisface la necesidad de la Red de Salud puno dando soporte para toma de decisiones.

- **Confiabilidad.**- Se mantiene el sistema y los datos danto en el tiempo hasta que la institución lo requiera.
- **Usabilidad.**- El Data Warehouse es fácil de usar y fácil de aprender por tanto cumple con este ítem.
- **Eficiencia.**- El Data Warehouse es muy rápido en cuanto a los reportes y también en cuanto al cruce de variables para sacar los gráficos que ayudan en la toma de decisiones.
- **Mantenibilidad.**- El Data Warehouse es fácil de modificar siempre y cuando los datos lo requieran, como se dijo anteriormente utiliza el modelo en estrella pero puede ajustarse también al modelo copo de nieve y así poder disminuir el nivel de detalle.
- **Portabilidad.**- Es fácil de llevar desde un ambiente a otro no existe restricción en este sentido, cuadro donde se muestra de manera porcentual las características que cumple el Data Warehouse según a la norma ISO - 9126.

CUADRO N° 02
Ponderado según ISO – 9126

Características de ISO - 9126	porcentaje
Funcionalidad	80%
Confiabilidad	80%
Usabilidad	100%
Eficiencia	90%
Mantenibilidad	80%
Portabilidad	80%

Fuente.- Encuestas y Método de Observación

4.4 ANÁLISIS E INTERPRETACIÓN DE PRUEBA DE HIPÓTESIS

Se evaluarán los resultados mediante el procesamiento de las encuestas que se tomó a cada usuario y para nuestro desarrollo de hipótesis nos planteamos lo siguiente:

Ho: no hay ninguna diferencia al utilizar el data warehouse para la toma de decisiones de la red de salud puno – 2013.

Ha: el data warehouse ayuda significativamente en la toma de decisiones de la Red de salud puno 2013

Según nuestros datos de nuestra encuesta tenemos:

Según nuestra tabla y el ponderado de respuestas obtenidas por cada usuario final del Data Warehouse.

Título: Cuadro de promedio de respuestas de por cada encuesta realizada a los usuarios

preguntas	user 1	user 2	user 3	user 4	user 5	user 6	user 7	user 8	user 9
promedio	4,125	3,75	4,25	3,875	3,875	4,125	4	4	4

Fuente: Encuestas realizadas

Donde:

$$n = 9$$

$$x = 4$$

$$s = 0,153$$

Por lo cual aplicaremos el estadístico T-Student para validar nuestra hipótesis utilizando esta fórmula para validarla.

$$T_c = \frac{\bar{X} - u}{S/\sqrt{n}}$$

Utilizando un nivel de significancia de $\alpha = 0,05$

$$T_{n-1,0,05} \text{ entonces } T_{8,0,05} = 1,86$$

Realizando los cálculos

$$T_c = \frac{4 - 3}{\frac{0,153}{3}} = 2,17$$

Entonces según nuestra regla de decisión

$T_c > T$ tabular, entonces rechazamos la Hipótesis Nula y Aceptamos la alternativa.

Conclusión.- según nuestro resultado, se demuestra mediante este estadístico de prueba que nuestro Data Warehouse ayuda de forma eficiente en la toma de decisiones de la Red de Salud Puno – 2013. Hacemos la comprobación de este estadístico corroborando la eficacia en procesamiento de información y dando reportes de manera significativa para la toma de decisiones.

CONCLUSIONES

Se llegó a las siguientes conclusiones:

1. Se analizó y diseño el Data Warehouse según a las bases de datos HIS, el diseño del Data Warehouse podría tomar otro modelo, esto depende de la cantidad de data y la estructura de las bases de datos.
2. La implementación de Data Warehouse para toma de decisiones de la Red de salud Puno- 2013 hace reportes en tiempo real y minimiza el trabajo de estar juntando datos e transformándolos, Con el Data Warehouse optimiza el tiempo de respuesta de cualquier interrogante que se le haga así como brinda información sólida para la toma de decisiones.
3. El Data Warehouse Garantiza la confiabilidad de los datos ya que dentro del proceso se realizó la limpieza y transformación, gracias al Data Warehouse se puede estar seguro de las decisiones. El Data Warehouse es eficiente y de fácil uso no se necesita de personal capacitado para poder sacar informes, con tener conocimientos básicos de Microsoft Excel es suficiente, por tal motivo es amigable y de fácil uso.

SUGERENCIAS

1. Este sistema puede ser aplicable a otras redes de salud, siempre y cuando los campos sean iguales a los campos del HIS trabajados en el Data Warehouse. Probablemente sería necesario crear nuevas tablas y agregar nuevos procesos al ETL. Se deja abierta esta posibilidad
2. La dirección de la red de salud puno será capaz de planificar metas, tomar decisiones para prevenir eventos adversos, responder a situaciones imprevistas, cambios en la demanda de servicios mejorar la calidad de atención a los pacientes teniendo en estas comparaciones con cifras anteriores, siendo posible medir los cambios en los indicadores de calidad y eficiencia de gestión del hospital y analizando el impacto de sus decisiones de forma directa en los pacientes y en la red de salud puno.
3. Los datos dan información y estos obedecen a muchos cambios que tiene dentro de los procedimientos que se les hace es por eso que antes de realizar un Data Warehouse tiene que tener datos que sean normalizados para no tener mayores inconvenientes en cada proceso de ETL.

BIBLIOGRAFÍA

- ANDREA, Z. (2004). *Incorporación de Soluciones OLAP en entorno Empresariales*. España: NETSCUM.
- C. J. Date, S. L. (2010). *Sistema de Base de Datos*. España: Pearson.
- DIAZ, J. C. (2010). *Introducción a Business Intelligence*. Barcelona: UOC.
- GAUCHED, T. (2011). *Sql Server 2008 R2*. Barcelona: ENI.
- JAVIER, N. (2002). *Sistema de Apoyo Gerencial para el instituto tecnológico de Buenos aires*. Buenos Aires: amrtsolutions.
- MAGDALENA, Y. (2005). *Construcción de un data warehouse de datos del medio ambiente para la toma de decisiones*. Buenos Aires: UDLAPMX
- PÉREZ, C. (2004). *Minería de Datos Técnicas herramientas*. Madrid - España, Instituto de Estudios Fiscales: PARANINFGO CENGAGE LEARNING.
- TRUJILLO, J. (2011). *Diseño y Explotación de Almacén de Datos*. España: Club Universitario.
- VITT, L. (2002). *Business Inteligence*. INTERAMERICANA DE ESPAÑA, S.A.: MCGAW-HILL.
- WELLING, L. (2005). *Desarrollo Web con PHP y Mysql*. Madrid - España: ANAYA S.A.

REFERENCIA WEB SITE

CTIBUSIC S, HADZAGIC-CATIBUSIC F, ZUBCEVIC S. Medical data warehousing as a generator of system component for decision support in health care [en línea] text info [EEUU] MEDILINE ,octubre 2004.disponible en http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=15000000

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=15000000>

DURÁN VALVERDE FABIO. Estudio financiero-actuarial y de la gestión de EsSalud: análisis y recomendaciones técnicas. [en línea] text info [PERU-LIMA] Pubmed abril 2005. Disponible en

<http://www.oitandina.org.pe/documentos/informe_actuarial_esSalud_oit_25_may_05_final.pdf>

GIL PADILLA ANTONIA M, BERRIEL MARTINEZ ROSARIO.Aplicacion de las tecnologías Data Warehouse en el contexto de la empresa turística de alojamiento hotelero [en línea] text info text [España- Universidad de las Palmas de Gran Canaria] Dpto., de Economía y dirección de Empresas, Facultad de Ciencias Económicas y Empresariales–2004. Disponible en <http://www.turismo.uma.es/turitec/turitec99/pdf/bd2.pdf>

ALEJANDRO GUTIERREZ, REGINA MOTZ, BEATRIZ REVELLO, LYDIA SILVIA. Construcción de un sistema de apoyo a la toma de decisiones para el área gerencial del Hospital de Clínicas.[en línea] info text [Montevideo Uruguayu] Instituto de Computación, Facultad de Ingeniería, Universidad de la República - 2001. Disponible en web

<[http://www.fing.edu.uy/inco/grupos/csi/esp/
Proyectos/csic2000/pub/pub2_sistdwhc.pdf](http://www.fing.edu.uy/inco/grupos/csi/esp/Proyectos/csic2000/pub/pub2_sistdwhc.pdf)>

HUYNEN MMTE, VOLLEBREGT L, MARTENS P, BENAVIDES BM. The epidemiologic transition in Peru. [en linea] info text [Perú] Rev. Panam Salud Pública. 2005. Disponible en web <http://journal.paho.org/index.php?a_ID=252>

RONALD J LAGOE. Improving Outcomes With Community-wide Distribution of Health Care Data. Health Care Management Review. [en linea] info text [EEUU] disponible en web dede agosto 2007 <<http://www.hcmrjournal.com>>.

MA, YANQIANG ALLEN, PH.D . Determinants of hospital information system (HIS) integrity and hospital performance. [en linea] info text [Virginia Commonwealth University, 2003] disponible en web. <<http://had.vcu.edu>>

ANEXOS

Anexo 01

Diccionario de Datos

Nombre del Archivo BdatosHIS			
Descripción : VPLAZAS			
Campos	Tamaño	Tipo de Dato	Descripción
CODPSAL		INT	CODIGO DE PERSONAL
COD_2000		INT	CODIGO DE UBICACIÓN DE PERSONAL
clave Primaria		CODPSAL Clave Foránea	

Nombre del Archivo BdatosHIS			
Descripción : TABVAR			
Campos	Tamaño	Tipo de Dato	Descripción
TABVALUE		INT	CODIGO DE VALOR DE AMS
TABDESC	50	VARCHAR	DESCRIPCION DE LAS AMS
clave Primaria		TABVALUE Clave Foránea	

Nombre del Archivo BdatosHIS			
Descripción : TABTIPODOC			
Campos	Tamaño	Tipo de Dato	Descripción
COD_DOC		INT	CODIGO DE DOCUMENTO
DES_DOC	50	VARCHAR	DESCRIPCION DE DOCUMENTO
clave Primaria		COD_DOC Clave Foránea	

Nombre del Archivo BdatosHIS			
Descripción : TABUPS			
Campos	Tamaño	Tipo de Dato	Descripción
COD_SERVA		INT	CODIGO DE SERVICIO DE SALUD
DESC_Servs	50	VARCHAR	DESCRIPCION DE SERVICIO DE SALUD
Clave Primaria		COD_DOC	
Clave Foránea			

Nombre del Archivo BdatosHIS			
Descripción: TABETNIA			
Campos	Tamaño	Tipo de Dato	Descripción
CODETNI		INT	CODIGO DE ETNIA

DESETNI	50	VARCHAR	DESCRIPCION DE ETNIA
CODGEN		INT	CODIGO DE ETNIA GENERAL
ETNIAS	50	VARCHAR	DESCRIPCION DE ETNIA GENERAL
clave Primaria		CODETNI Clave Foránea	

Nombre del Archivo BdatosHIS			
Descripción: TABUBIGEO			
Campos	Tamaño	Tipo de Dato	Descripción
COD_UBI		INT	CODIGO DE UBIGEO
DESC_UBIGE	50	VARCHAR	DESCRIPCION DE UBIGEO
clave Primaria		COD_UBI Clave Foránea	

Nombre del Archivo BdatosHIS			
Descripción: TABDIAG			
Campos	Tamaño	Tipo de Dato	Descripción
DIAG		INT	DIAGNOSTICO
DESCIP	50	VARCHAR	DESCRIPCION DEL DIAGNOSTIVO
DIAG clave Primaria			
Clave Foránea			

Nombre del Archivo BdatosHIS			
Descripción: T_ESTAB			
Campos	Tamaño	Tipo de Dato	Descripción
COD_TSTAB		INT	CODIGO DE ESTABLECIMINIENTO
DESC_TESTAB	50	VARCHAR	TIPO DE ESTABLECIMIENTO
clave Primaria		COD_TSTAB	
Clave Foránea			

Nombre del Archivo BdatosHIS			
Descripción: SITUACION			
Campos	Tamaño	Tipo de Dato	Descripción
VALORES		INT	VALORES DE LAB
DESCRIPCION	50	VARCHAR	DESCRIPCION DE LAB
clave Primaria		VALORES Clave Foránea	

Nombre del Archivo BdatosHIS			
Descripción:SERVSAL			
Campos	Tamaño	Tipo de Dato	Descripción
COD_SERVA		INT	CODIGO DE SERVICIO DE SALUD
DESC_SERVS	50	VARCHAR	DESCRIPCION DE SERVICIO DE SALUD
clave Primaria		COD_SERVA Clave Foránea	

Nombre del Archivo BdatosHIS			
Descripción: SERV_CON			
Campos	Tamaño	Tipo de Dato	Descripción
COD_SERVSA		INT	COSIGO DE SERVICIO GENERAL
DESC_Servs	50	VARCHAR	DESCRIPCION DE SERVICIO DE SALUD GENERAL
clave Primaria		COD_SERVSA Clave Foránea	

Nombre del Archivo BdatosHIS			
Descripción: RGENERAL			
Campos	Tamaño	Tipo de Dato	Descripción
REPORTE		INT	CODIGO DEREPORTE
DESCRIP	50	VARCHAR	DESCRIPCION DE REPORTE
clave Primaria		REPORTE Clave Foránea	

Nombre del Archivo BdatosHIS			
Descripción: RED			
Campos	Tamaño	Tipo de Dato	Descripción
COD_DISA		INT	CODIGO DE DISA
COD_RED		INT	CODIGO DE RED
DESC_RED	50	VARCHAR	DESCRIPCION DE RED
clave Primaria		COD_RED Clave Foránea COD_DISA	

Anexo 02

Encuesta desarrollada para la eficiencia del Data Warehouse

Anexo 01

Apellido y Nombre _____

Cargo _____

EVALUACION DEL INTERFAZ

El manejo del Data Warehouse es

Muy fácil () Fácil () Regular ()
) Muy Difícil () La exactitud de

información es

Muy fácil () Fácil () Regular ()
) Muy Difícil () La disponibilidad de

información es

Muy bueno () Bueno () Regular ()
) Malo () Muy malo ()

El Data Warehouse le provee toda la información en tiempo real. Muy bueno

() Bueno () Regular ()
) Malo ()

Muy malo () La interfaz es:

Muy buena () Buena () Regular ()
) Mala () Muy mala ()

CONTENIDO Y LA FUNCIONALIDAD

Acorta el tiempo asociado a la toma de decisiones

Alta () Media () Baja () El Data

Warehouse es Eficaz en cuanto a los reportes.

Alta () Media () Baja () El Data Ware House es
Versátil con la información que tiene. Alta () Media ()
Baja ()