UNIVERSIDAD MAYOR DE SAN SIMON

FACULTAD DE CIENCIAS Y TECNOLOGÍA

CARRERA INGENIERÍA DE SISTEMAS- INGENIERÍA INFORMÁTICA

**TRABAJO DE DATA WAREHOUSE**

**DISEÑO DE DATA WAREHOUSE PARA LA CADENA DE HOTELES “EL VIAJERO”**

ELABORADO POR: Alexander Mamani Yucra

Alvaro Yapu Cossio

Gabriela Nathaly Torrico

PERIODO: 2-2019

# INTRODUCCIÓN

Un Data Warehouse (DWH) es un conjunto de datos orientados a temas, integrados, no volátiles e históricos, organizados de tal forma que sirven de apoyo a la toma de decisiones , dado que permiten analizar la información consolidada según diferentes puntos de vista. Dicho proceso de consolidación de información involucra actividades de extracción de diversas fuentes de datos, transformación de la información necesaria y finalmente su carga en el DWH. Usualmente se denomina a este proceso ETL, del inglés Extraction, Transformation and Loading.

Las transformaciones aplicadas a los datos provenientes de las distintas fuentes son básicamente de limpieza y de estructuración. Las transformaciones de limpieza son necesarias para asegurar la calidad de los datos finalmente almacenados en el DWH e incluye entre otros, la corrección de errores, eliminación de redundancia y resolución de inconsistencias, así como el asegurar las reglas de negocio definidas. Los cambios en la estructura se realizan para adecuar los esquemas a las funcionalidades de un DWH, e incluyen la adecuación al modelo de datos del DWH, cambios de formato, operaciones de agregación, entre otras.

Un data warehouse es de gran importancia pues puede ayudar a las organizaciones a extraer el máximo valor de los datos que se generan en el día a día, permitiendo además analizarlos y compararlos con sus valores históricos y no solo mediante datos actuales. Aun así, en algunos casos, sobre todo cuando hablamos de big data, el data warehouse puede tener ciertas limitaciones que sin embargo pueden ser resueltas gracias a su ductibilidad.

El uso de un DWH simplifica los procesos de toma de decisiones porque ofrece imágenes integradas de los datos. Facilita el proceso de comparación, proyección a futuro, relación con otros datos, muestra de indicadores, información consolidada, entre otras, el acceso a los datos es fácil y rápido, permitiendo a los usuarios hacer sus propias consultas, ayuda a mejorar el buen funcionamiento de los sistemas operacionales retroalimentando demandas para los sistemas transaccionales, además Un data warehouse se emplea para hacer el trabajo analítico, dejando las bases de datos transaccionales libres para centrarse en las transacciones, tiene la capacidad de analizar datos de múltiples fuentes y puede negociar las diferencias en cuanto a esquemas de almacenamiento utilizando procesos de ETL, al integrar datos de múltiples sistemas de origen, permite una visión central en toda la empresa, mantiene el historial de datos incluso si los sistemas transaccionales de origen no lo hacen, mejora los datos, proporcionando códigos y descripciones coherentes, incluso arreglando datos erróneos, presenta la información de la organización de forma coherente.

**JUSTIFICACION**

El presente proyecto explora nuevas herramientas del área informática, para que se optimicen la elaboración de informes para la toma de decisiones de las distintas áreas , es asi que empresas dedicadas a la hotelería tiene la gran necesidad de contar con este tipo de herramientas, es por ello que se plantea una solución basada en tecnologías de información mediante la implementación de un almacen de datos mejor llamado Data Warehouse, el mismo que nos permite obtener datos y generar información tratada y útil, la cual apoyara el análisis del negocio que se requieren.

Este proyecto ayudara en un mediano plazo a procesar, administrar y consultar volúmenes de informaicon, aumentar la eficiencia y competitividad, por otra parte, las empresas que actualmente no cuentan con este modelo encontrará una propuesta oportuna con visión en el futuro, al considerar adquiririr y desarrollar nuevas tecnologías que les permitan administrar sus recursos mediante la información que generan los sistemas de información.

Los nuevos retos que actualmente enfrenta una empresa hotelera en el país van encaminados a que las empresas se vuelven cada vez mas competitivas en el mundo laboral y el modelo copo de nieve de Data Warehouse seria una herramienta clave para analizar desde diversas perspectivas la información de las empresas.

Es por ello que la propuesta de este tipo de modelo es proveer el empleo de herramientas informáticas, conservación y explotación de la información histórica de los sistemas transaccionales, omitir registros innecesarios en las bases de datos y generar aquellos indispensables para la toma de decisiones.

# OBJETIVOS

## OBJETIVO GENERAL

* Construir Un Datawarehouse de la Base de Datos Relacional de la cadena de hoteles “EL VIAJERO” Usando El Modelo Estrella

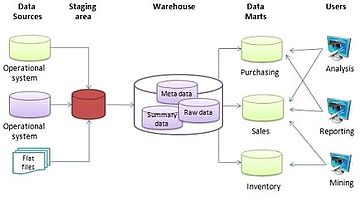
## OBJETIVOS ESPECÍFICOS

* Diseñar el Datawarehouse tomando la Base de Datos relacional usando el modelo de estrella
* Crear una data warehouse con 5 data marts
* Hacer un proceso de ETL para la extracción de datos para construir los data marts.
* Usar como mínimo 4 técnicas de transformación y 2 de scripting
* Crear una data warehouse con 5 data marts

# MARCO TEORICO

## DATAWAREHOUSE

Es la [integración](http://www.monografias.com/trabajos11/funpro/funpro.shtml) de datos consolidados, almacenados en un dispositivo de [memoria](http://www.monografias.com/trabajos13/memor/memor.shtml) no volátil, proveniente de múltiples y posiblemente diferentes [fuentes](http://www.monografias.com/trabajos10/formulac/formulac.shtml#FUNC) de datos. Con el propósito del análisis y a partir de este tomar decisiones en [función](http://www.monografias.com/trabajos7/mafu/mafu.shtml) de mejorar la gestión del negocio. Contiene un conjunto de cubos de datos que permiten a través de técnicas de OLAP consolidar, ver y resumir los datos acordes a diferentes dimensiones de estos.



## DATA MARTS

Es un subconjunto del Data Warehouse, usado normalmente para el análisis parcial de los datos. El objetivo de subdividir está dado por la complejidad computacional del análisis global de todas las dimensiones del Data Warehouse y por la necesidad de rapidez.

## DATA MINING

Es el descubrimiento de conocimiento oculto en las bases de datos. Relaciones entre estos y tendencias que permiten una toma de decisiones acertada. Incluye Asociación, Caracterización, Clasificación, Análisis de Series Cronológicas, etc.

## OLTP (Online Transaction Proccesing)

Se les llama así a las aplicaciones orientadas principalmente a la inserción, actualización y eliminación de datos, diseñada casi siempre usando el modelo Relacional. Estos sistemas están optimizados para realizar estas operaciones en un tiempo corto.

## OLAP (Online Analitical Proccesing)

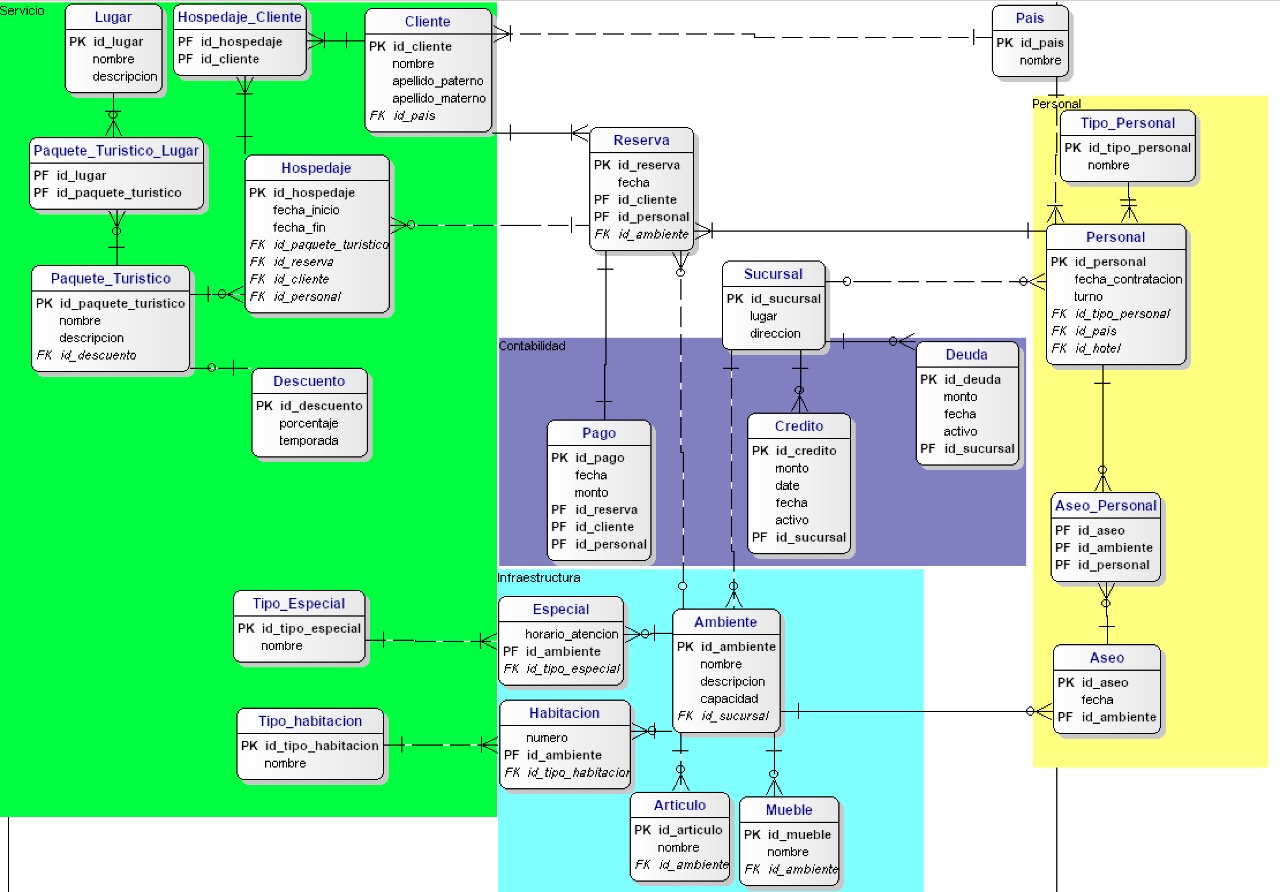
Son los sistemas que se usan para analizar los datos que las OLTP introducen en la Base de Datos. A diferencia de los primeros estos casi siempre usan el modelo multidimensional para organizar los datos en la Base de Datos ya que brindan mejores resultados a la hora del análisis de estos.

## MODELOS DE ALMACENAMIENTO

## Ventajas Modelo en estrella

* Este esquema es simple y veloz para ser usado en análisis multidimensionales. Permite acceder tanto a datos agregados como de detalle.
* El diseño de esquemas en estrella permite implementar la funcionalidad de una base de datos multidimensional utilizando una clásica base de datos relacional.
* Es su simple desde el punto de vista del usuario final. Las consultas no son complicadas, ya que las condiciones y los joins sólo involucran a la tabla de hechos y a las de dimensiones.
* Son mas simples de manejar que los modelos de copo de nieve.
* Es la opción con mejor rendimiento y velocidad pues permite indexar las dimensiones de forma individualizada sin que repercuta en el rendimiento de la base de datos en su conjunto.

**Modelo de Base de Datos realizado en Power Design**



**Requerimientos.-**

Estos son la configuración de la base de datos para las transformaciones

Base de datos de entrada:

* Host: localhost
* Port: 5432
* Database: data
* User: postgres
* Password: pepito1234

Base de datos de salida:

* Host: localhost
* Port: 5432
* Database: dataoutput
* User: postgres
* Password: pepito1234