



**UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MÉXICO**

---

**Centro Universitario UAEM Nezahualcóyotl**

# **Licenciatura en Ingeniería en Sistemas Inteligentes**

**Unidad de Aprendizaje:  
DATA WAREHOUSE I**

**Tema:**

**Arquitectura para la implantación del proceso de Data  
Warehouse y Diseño del Data Warehouse**

**Dra. Carmen Liliana Rodríguez Páez**

**[clrodriguezp@uaemex.mx](mailto:clrodriguezp@uaemex.mx)**



## Rector

Doctor en Educación **Alfredo Barrera Baca**

Maestra en Salud Pública

**María Estela Delgado Maya**

Secretaria de Docencia

Doctor en Ciencias e Ingeniería

**Carlos Eduardo Barrera Díaz**

Secretario de Investigación y Estudios Avanzados

Doctor en Ciencias Sociales

**Luis Raúl Ortiz Ramírez**

Secretario de Rectoría

Doctor en Arte

**José Édgar Miranda Ortiz**

Secretario de Difusión Cultural

Maestra en Comunicación

**Jannet Valero Vilchis**

Secretaria de Extensión y Vinculación

Maestro en Economía

**Javier González Martínez**

Secretario de Administración

Doctor en Ciencias de la Computación

**José Raymundo Marcial Romero**

Secretario de Planeación y Desarrollo Institucional

Maestra en Lingüística Aplicada

**María del Pilar Ampudia García**

Secretaria de Cooperación Internacional

Doctora en Ciencias Sociales y Políticas

**Gabriela Fuentes Reyes**

Abogada General

Licenciado en Comunicación

**Gastón Pedraza Muñoz**

Director General de Comunicación Universitaria

Maestro en Relaciones Interinstitucionales

**Jorge Bernáldez García**

Secretario Técnico de la Rectoría

Maestra en Administración Pública

**Guadalupe Ofelia**

Directora General de Centros Universitarios

**Santamaría González**

y Unidades profesionales

Maestro en Administración

**Ignacio Gutiérrez Padilla**

Contralor

Maestro en Derecho **Juan Carlos Medina Huicochea**

Encargado del Despacho de la Dirección del Centro Universitario UAEM Nezahualcóyotl

Maestro en Ciencias

**José Antonio Castillo Jiménez** Subdirector Académico

Licenciado en Economía

**Ramón Vital Hernández** Subdirector Administrativo

Doctora en Ciencias Sociales

**María Luisa Quintero Soto** Coordinadora de Investigación y Estudios Avanzados

Licenciado en Administración de Empresas

**Víctor Manuel Durán López** Coordinador de Planeación y Desarrollo Institucional

Doctor en Relaciones Internacionales

**Rafael Alberto Durán Gómez** Coordinador de la Licenciatura en Comercio Internacional

Maestra en Ciencias

**Gabriela Kramer Bustos** Coordinadora de la Licenciatura en Educación para la Salud

Doctor en Ingeniería de los Sistemas

**Ricardo Rico Molina** Coordinador de la Licenciatura en Ingeniería en Sistemas Inteligentes

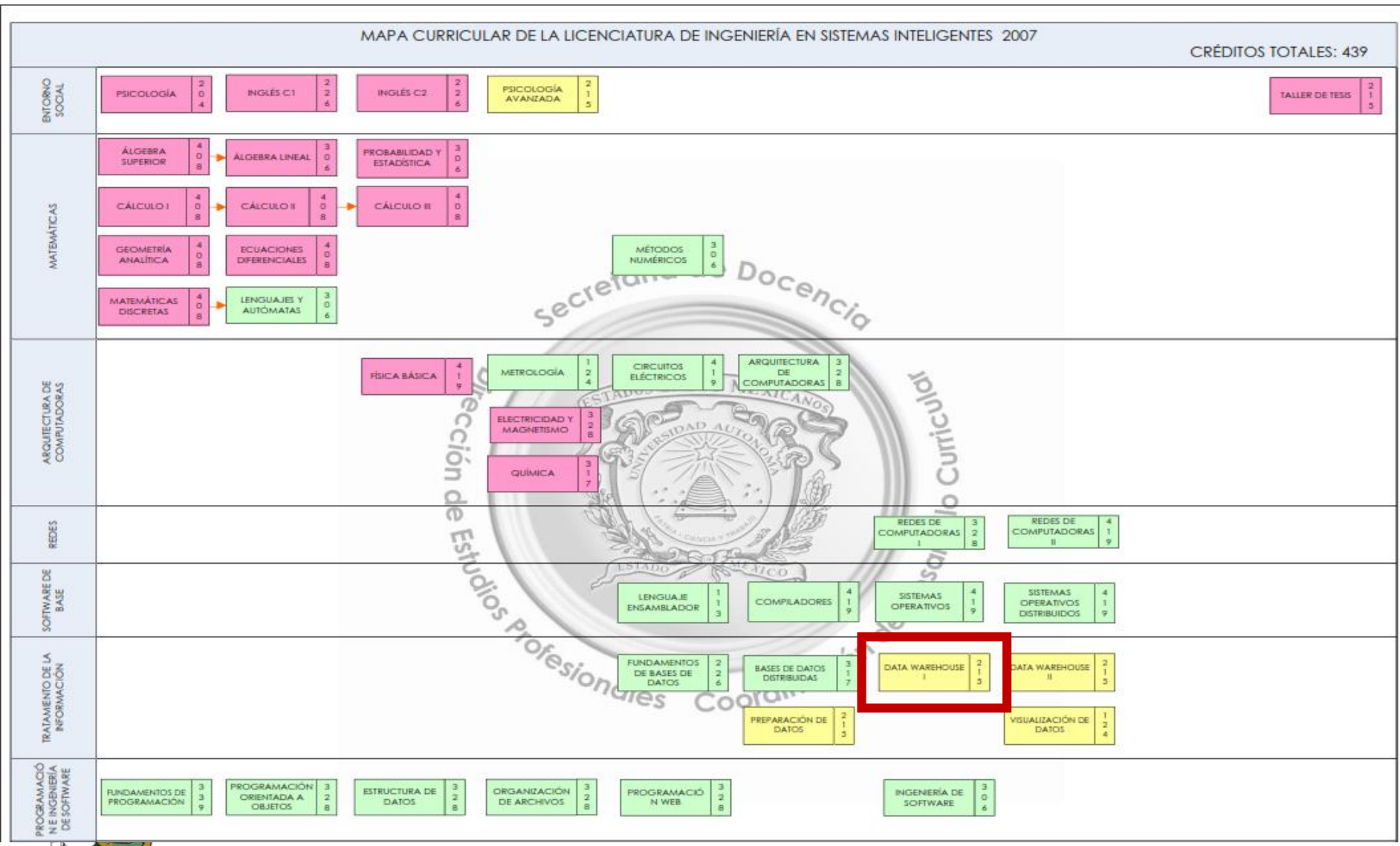
Doctor en Urbanismo

**Noé Gaspar Sánchez** Coordinador de la Licenciatura en Ingeniería en Transporte

Maestro en Ciencias de la Computación

**Erick Nicolás Cabrera Álvarez** Coordinador de la Licenciatura en Seguridad Ciudadana

# Ubicación de la asignatura Data Warehouse I dentro del programa de la Lic. en Ingeniería en Sistemas Inteligentes



La Unidad de Aprendizaje (UA) de *Data Warehouse I* tiene como área curricular el tratamiento de la información y forma parte del núcleo sustantivo.

El material de visión proyectable de la Unidad de Aprendizaje Data Warehouse I es un documento que contiene los temas: ***Arquitectura para la implantación del proceso de Data Warehouse, Elementos de un Data Warehouse y Inteligencia de Negocios.*** Correspondientes a la unidad I. del programa por competencia.

Permitirá a los alumnos la comprensión de los temas, y será una guía de apoyo para su uso en la elaboración del proyecto final de la materia.

Al término de la unidad, el alumno comprenderá los fundamentos teóricos básicos del Data Warehouse (DW) para el análisis, mapeo, y transformación de datos útil en la toma de decisiones.

## Unidad 1. Introducción a los Conceptos del Data Warehouse

**Objetivo:** Conocer los conceptos básicos de Data Warehouse

**Contenidos:**

- 1.0. Introducción
- 1.1 Antecedentes sobre el concepto de Data Warehouse
- 1.2 ¿Qué es un Data Warehouse?
- 1.3 ¿Por qué es necesario un Data Warehouse?
- 1.4 Tipos de Data Warehouse
- 1.5 Arquitectura para la implantación del proceso de Data Warehouse
- 1.6. Diseño de un data warehouse

## Unidad 2. Análisis del entorno Data Warehouse

**Objetivo:** Conocer las características de un Data Warehouse

**Contenidos:**

- 2.1. Conceptos y funcionalidades básicas
- 2.3. Tendencias tecnológicas y de mercado
- 2.4. Relación del data warehouse con la Inteligencia del Negocio

## Unidad 3. Aspectos técnicos en el proceso de creación y explotación del Data Warehouse

**Objetivo:** Identificar indicadores y perspectivas de realización para el desarrollo de un Data Warehouse

**Contenidos:**

- 3.1. Análisis de Requerimientos de la empresa
- 3.2. Análisis de los OLTP
- 3.3. Factores relevantes en el proceso de adquisición

## Unidad 4. Diseño del modelo conceptual del Data Warehouse

**Objetivo:** Diseñar el modelo conceptual del Data Warehouse

**Contenidos:**

- 4.1 Diseño del pliego de prescripciones técnicas particulares
- 4.2 Modelo Conceptual DW

## Unidad 5. Modelo lógico del Data Warehouse

**Objetivo:** Crear el modelo conceptual ampliado y el modelo lógico

**Contenidos:**

- 5.1. Modelo Conceptual ampliado
- 5.2. Modelo lógico

## Data Warehouse I

### Unidad 1. Introducción a los Conceptos del Data Warehouse

**Objetivo: Conocer los conceptos básicos de Data Warehouse**

**Contenidos:**

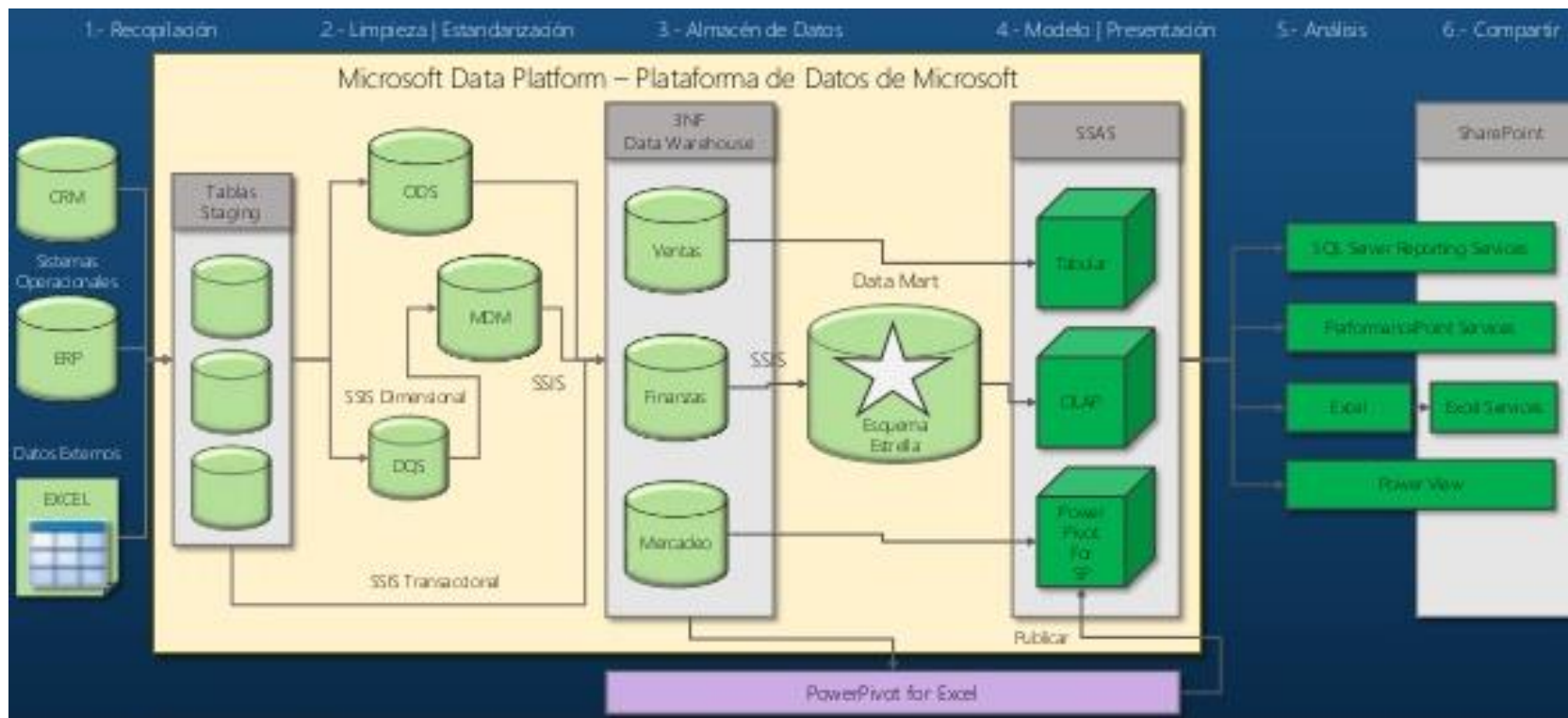
- 1.0. Introducción
- 1.1 Antecedentes sobre el concepto de Data Warehouse
- 1.2 ¿Qué es un Data Warehouse?
- 1.3 ¿Por qué es necesario un Data Warehouse?
- 1.4 Tipos de Data Warehouse
- 1.5 Arquitectura para la implantación del proceso de Data Warehouse
- 1.6 Diseño de un data warehouse

**1.5 Arquitectura para la implantación del proceso de Data Warehouse**

**1.6 Diseño del Data Warehouse**

# Arquitectura Data Warehouse

- Base de datos operacional / Nivel de base de datos externo
- Nivel de acceso a la información
- Nivel de acceso a los datos
- Nivel de directorio de datos (Metadata)
- Nivel de gestión de proceso
- Nivel de mensaje de la aplicación
- Nivel de data warehouse
- Nivel de organización de datos

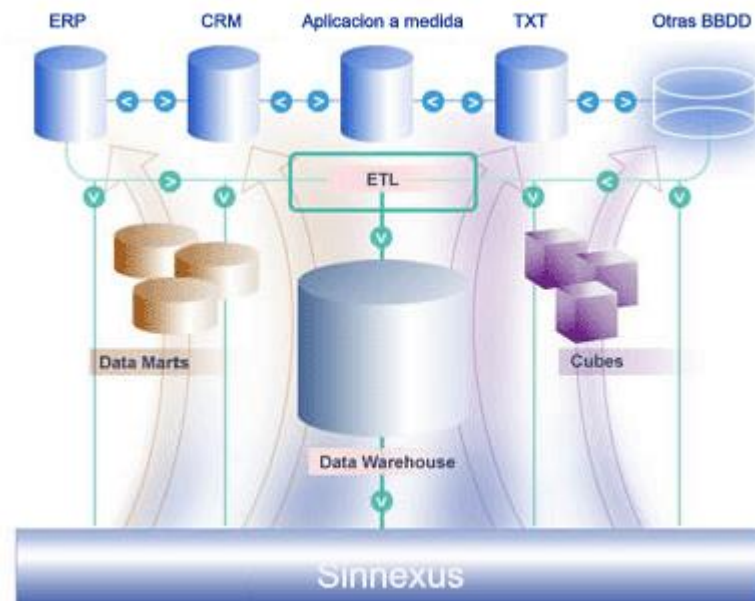


**Figura 1. Arquitectura del Data Warehouse**

## 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.

- Ciertamente, la meta del data warehousing es liberar la información que es almacenada en bases de datos operacionales y combinarla con la información desde otra fuente de datos, generalmente externa.



# Nivel de acceso a la información

- El nivel de acceso a la información de la arquitectura 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, Lotus 1-2-3, Focus, Access, SAS, etc.

- Una de las claves para esto es encontrar un lenguaje de datos común que puede usarse 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 para relacionarse con el nivel operacional.
- El lenguaje de datos común es el SQL.

- 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 todos los datos en la empresa que es necesaria para ellos, para hacer su trabajo



- El nivel de acceso a los datos es responsable de la interfase 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.
- 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.

75	84	1/31/1980	Thursday	d:\hrp2\1082	16	Amber Glenn-Thomas	02/06/14	Zeutschel	d:\7027	Rinchen Dalim	6/3/2014	LIB-OPC01	Zake Graves	5/11/2014	1
75	85	2/1/1980	Friday	d:\hrp2\1083	20	Amber Glenn-Thomas	02/06/14	Zeutschel	d:\7027	Rinchen Dalim	6/3/2014	LIB-OPC01	Zake Graves	5/11/2014	1
75	86	2/4/1980	Monday	d:\hrp2\1084	12	Amber Glenn-Thomas	02/06/14	Zeutschel	d:\7027	Rinchen Dalim	6/3/2014	LIB-OPC01	Zake Graves	5/11/2014	1
75	87	2/5/1980	Tuesday	d:\hrp2\1085	16	Amber Glenn-Thomas	02/06/14	Zeutschel	d:\7027	Rinchen Dalim	6/4/2014	LIB-OPC01	Zake Graves	5/11/2014	1
75	88	2/6/1980	Wednesday	d:\hrp2\1086	16	Amber Glenn-Thomas	02/06/14	Zeutschel	d:\7027	Rinchen Dalim	6/4/2014	LIB-OPC01	Zake Graves	5/11/2014	1
75	89	2/7/1980	Thursday	d:\hrp2\1087	16	Amber Glenn-Thomas	02/06/14	Zeutschel	d:\7027	Rinchen Dalim	6/4/2014	LIB-OPC01	Zake Graves	5/11/2014	1
75	name	Winter 1980		d:\hrp2\1088	24	Amber Glenn-Thomas	02/06/14	Zeutschel	d:\7027	Rinchen Dalim	6/4/2014	LIB-OPC01	Zake Graves	5/11/2014	1
75	90	2/8/1980	Friday	d:\hrp2\1089	28	Amber Glenn-Thomas	02/06/14	Zeutschel	d:\7027	Rinchen Dalim	6/4/2014	LIB-OPC01	Zake Graves	5/11/2014	1
75	91	2/11/1980	Monday	d:\hrp2\1090	12	Amber Glenn-Thomas	02/06/14	Zeutschel	d:\7027	Rinchen Dalim	6/4/2014	LIB-OPC01	Zake Graves	5/11/2014	1
75	92	2/12/1980	Tuesday	d:\hrp2\1091	20	Amber Glenn-Thomas	02/06/14	Zeutschel	d:\7027	Rinchen Dalim	6/4/2014	LIB-OPC01	Zake Graves	5/11/2014	1
75	93	2/13/1980	Wednesday	d:\hrp2\1092	12	Amber Glenn-Thomas	02/06/14	Zeutschel	d:\7027	Rinchen Dalim	6/4/2014	LIB-OPC01	Zake Graves	5/11/2014	1
75	94	2/14/1980	Thursday	d:\hrp2\1093	20	Zake Graves	05/05/14	Zeutschel	d:\7028	Rinchen Dalim	6/4/2014	LIB-OPC01	Zake Graves	5/22/2014	1
75	95	2/15/1980	Friday	d:\hrp2\1094	20	Zake Graves	05/05/14	Zeutschel	d:\7028	Rinchen Dalim	6/4/2014	LIB-OPC01	Zake Graves	5/22/2014	1
75	96	2/16/1980	Monday	d:\hrp2\1095	16	Zake Graves	05/05/14	Zeutschel	d:\7028	Rinchen Dalim	6/4/2014	LIB-OPC01	Zake Graves	5/22/2014	1
75	97	2/19/1980	Thursday	d:\hrp2\1096	12	Zake Graves	05/05/14	Zeutschel	d:\7028	Rinchen Dalim	6/4/2014	LIB-OPC01	Zake Graves	5/22/2014	1
75	98	2/20/1980	Friday	d:\hrp2\1097	16	Zake Graves	05/05/14	Zeutschel	d:\7028	Rinchen Dalim	6/4/2014	LIB-OPC01	Zake Graves	5/22/2014	1
75	99	2/21/1980	Monday	d:\hrp2\1098	16	Zake Graves	05/05/14	Zeutschel	d:\7028	Rinchen Dalim	6/4/2014	LIB-OPC01	Zake Graves	5/22/2014	1
75	100	2/22/1980	Tuesday	d:\hrp2\1099	16	Zake Graves	05/05/14	Zeutschel	d:\7028	Rinchen Dalim	6/4/2014	LIB-OPC01	Zake Graves	5/22/2014	1
75	101	2/23/1980	Wednesday	d:\hrp2\1100	12	Zake Graves	05/05/14	Zeutschel	d:\7028	Rinchen Dalim	6/4/2014	LIB-OPC01	Zake Graves	5/22/2014	1
75	102	2/24/1980	Thursday	d:\hrp2\1101	20	Zake Graves	05/05/14	Zeutschel	d:\7028	Rinchen Dalim	6/4/2014	LIB-OPC01	Zake Graves	5/22/2014	1
75	103	2/27/1980	Monday	d:\hrp2\1102	12	Zake Graves	05/05/14	Zeutschel	d:\7028	Rinchen Dalim	6/4/2014	LIB-OPC01	Zake Graves	5/22/2014	1
75	104	2/28/1980	Tuesday	d:\hrp2\1103	12	Zake Graves	05/05/14	Zeutschel	d:\7028	Rinchen Dalim	6/4/2014	LIB-OPC01	Zake Graves	5/22/2014	1
75	105	2/29/1980	Wednesday	d:\hrp2\1104	16	Zake Graves	05/05/14	Zeutschel	d:\7028	Rinchen Dalim	6/4/2014	LIB-OPC01	Zake Graves	5/22/2014	1
75	106	2/3/1980	Monday	d:\hrp2\1105	12	Zake Graves	05/05/14	Zeutschel	d:\7028	Rinchen Dalim	6/4/2014	LIB-OPC01	Zake Graves	5/22/2014	1

METADATA



- A fin de tener un depósito totalmente funcional, es necesario tener una variedad de metadata disponibles, información sobre las vistas 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 dónde residen los datos o 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 deben realizarse para construir y mantener el data warehouse y la información del directorio de datos.

# Nivel de Gestión de Procesos

- El nivel de gestión de procesos tiene que ver con la programación de diversas tareas que deben realizarse para construir y mantener el data warehouse y la información del directorio de datos.

- Este nivel puede depender del alto nivel de control de trabajo para muchos procesos (procedimientos) que deben ocurrir para mantener el data warehouse actualizado.

# Nivel de Mensaje de la Aplicación

- El nivel de mensaje de la aplicación tiene que ver con el transporte de información alrededor de la red de la empresa.
- El mensaje de aplicación se refiere también como "subproducto", pero puede involucrar sólo protocolos de red.

- Puede usarse por ejemplo, para aislar aplicaciones operacionales o estratégicas a partir del formato de datos exacto, recolectar transacciones o los mensajes y entregarlos a una ubicación segura en un tiempo seguro

# Nivel Data Warehouse (Físico)

- En el data warehouse (núcleo) es donde ocurre la data actual, usada principalmente para usos estratégicos.
- Se puede pensar del data warehouse simplemente como una vista lógica o virtual de datos.
- En muchos ejemplos, el data warehouse puede no involucrar almacenamiento de datos.

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

# 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 réplica.
- 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 frecuencia una programación compleja.
- Están creándose las herramientas data warehousing para ayudar en este proceso.

# 1.6 Diseño del Data Warehouse

Para ello vamos a suponer que estamos usando la arquitectura de Ralph Kimball que es la mas vista en la practica de una empresa.

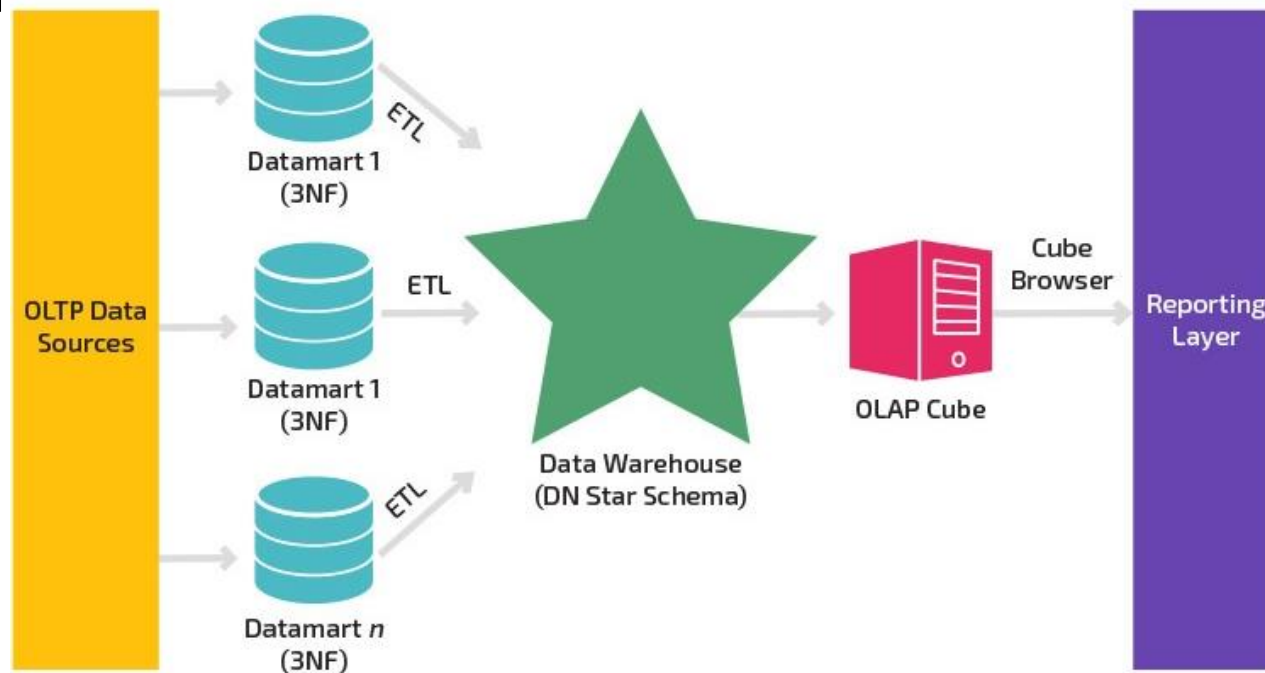


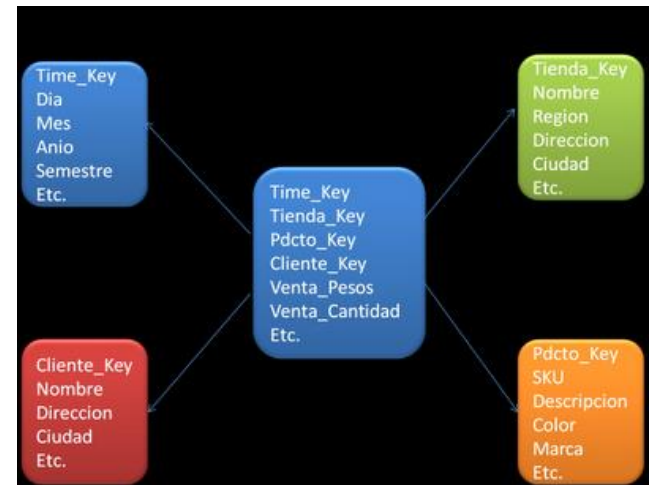
Figura 2. Arquitectura Ralp Kimball

# Modelo dimensional

Disciplina de modela de datos alternativa al modelo E-R-

Maneja un formato simétrico que permite:

- Facilidad de entendimiento
- Eficiencia en consultas
- Resistencia al cambio, es ultima quiere decir que cambios futuros a la estructura de datos, requiera menos trabajo y manipulación de la base de datos.



Se basa en dos tipos de tablas:

- Hechos
- Dimensiones

Estas tablas forman un esquema en estrella, el cual representa un diagrama por cada proceso de negocio. Un proceso de negocio vendría ser conjunto de actividades definidas dentro de la empresa en un tema definido.

# Hechos

Las tablas de hechos son el centro del modelo relacional. Estas contienen mediciones cuantitativas del negocio que quiere decir valores numéricos o aditivos los que se les puede aplicar sumatorias y promedios que son el resultado de los análisis que hagamos del proceso de negocio.

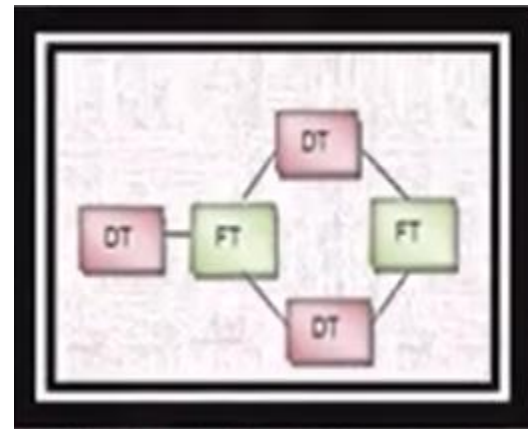
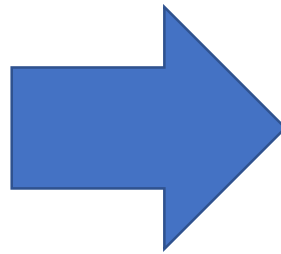
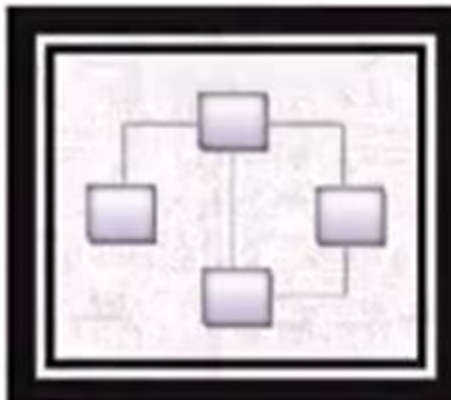
Estas tablas no tienen llaves primarias sino que se encuentran formadas por las llaves foráneas completando así relaciones N:M. relaciones de muchos a muchos dependiendo de las dimensiones.

# Dimensiones

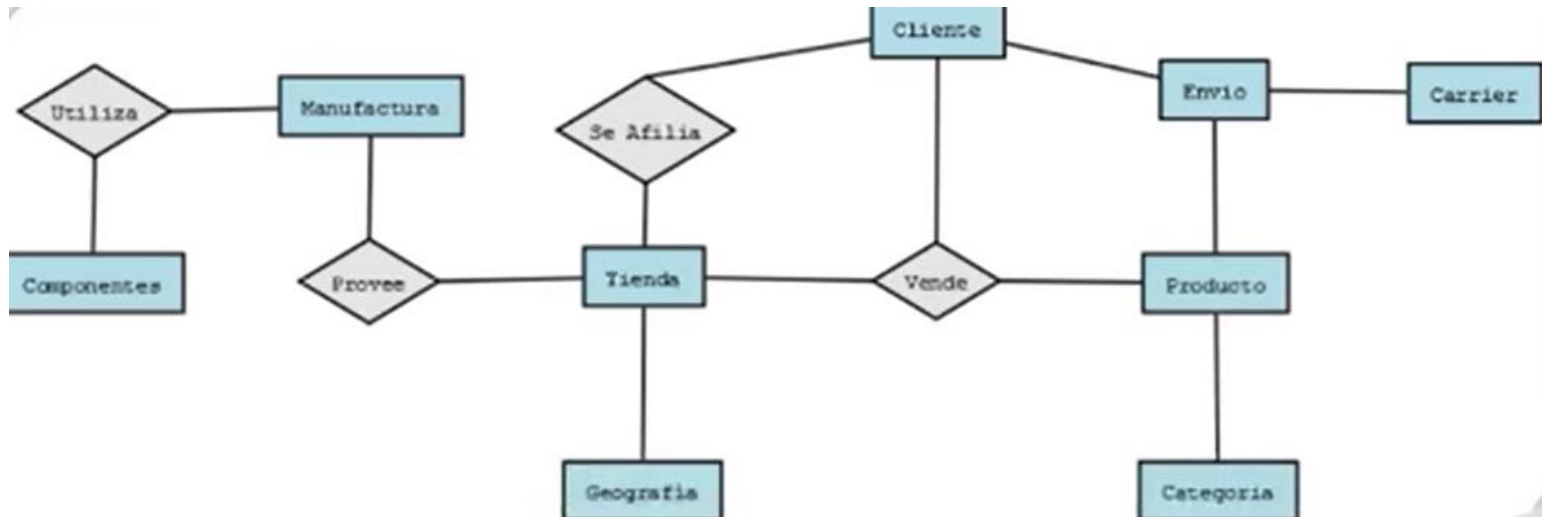
- Se puede decir que es el segundo componente, las tablas de dimensiones, de donde obtienen los datos.
- Estas son las que tienen la información extraída de los sistemas transaccionales.

- Sus llaves primarias definen la integridad referencial del DW. Esto quiere decir cada una de estas dimensiones tiene una llave primaria única que se utiliza para relacionarse entre los diferentes hechos.
- Además contiene campos que funcionan como filtros y agrupadores para aplicar funciones de agregación.

# Paso a paso de como transformar el modelo ER al modelo dimensional



Se parte del esquema entidad relación que puede ser por ejemplo una tienda.



**1. Separar el diagrama ER en sub-diagramas por cada proceso de negocio.**

## 2. Obtener los valores cuantificables de cada sub diagrama.

Consiste en datos que podemos extraer de cada una de estas relaciones.

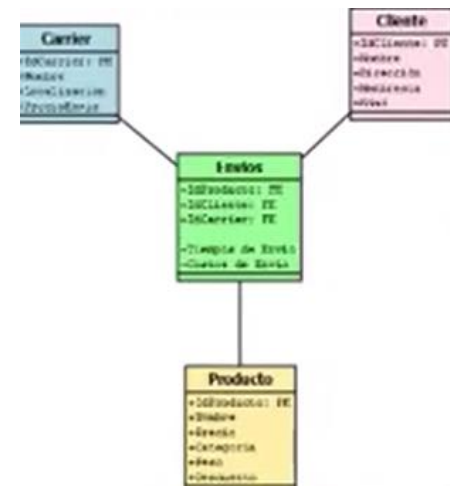


Estos valores cuantificables que sacamos de cada su diagrama vendrían hacer los hechos para el diagrama en estrella.

Luego que tenemos estos hechos se pasa al paso 3.

### 3. La transformación de las tablas involucradas en cada proceso de negocio en tablas “planas” con llaves primarias únicas y atómicas.

Teniendo los hechos de los envíos podemos reducir las tablas de dimensiones producto, Carrier y cliente. En su forma mas simple.



Con los datos necesarios para poder medir los hechos de los envíos y entonces así completar el esquema estrella.

Cuando una tabla de dimensión se conecta hechos de distintos procesos de negocios se les llama dimensiones conformadas.

Esto quiere decir que estas dimensiones se comparten entre diferentes esquemas y por lo tanto deben representar el mismo objeto para ello volvemos al mismo concepto de arquitectura de bus: estándar de dimensiones sobre Data Marts para unificar los datos.

Esto quiere decir que estas dimensiones se comparten entre diferentes esquemas y por lo tanto deben representar el mismo objeto para ello volvemos al mismo concepto de arquitectura de bus: estándar de dimensiones sobre Data Marts para unificar los datos.

Por ejemplo tenemos una dimensión de productos que existe entre distintos data Marts (entre distintos de departamentos) estos deben ponerse de acuerdo para estandarizar el formato de este y que sea compatible a través de todos los data marts.

Con el objetivo de que ahora el DW realice Querys a cualquiera de estos productos su información sea compatible.

Una vez que completamos el diseño del DW se procede a cargar datos en el para ello se tiene el proceso de ETL.

Este proceso tiene tres partes:

- La extracción de datos de bases de datos transaccionales
- La transformación
- El transporte a sus data marts finales

## En la parte de la extracción de datos de bases de datos transaccionales:

El primer paso que se tiene es identificar cuales datos han cambiado desde la ultima vez que fue hecha una extracción. En el caso de ser la primera vez se cargan todos los datos.

Se hace una lectura de los datos se obtienen los agregados eso quiere decir se realizan operaciones de sumas y promedios y algunas otras funciones del mantenimiento del metadato.

## En La transformación de los datos:

Que incluye una limpieza e integración, esto quiere decir que si tenemos el mismo dato pero en diferente formatos en varias bases de datos lo que se hace es transformarlo a un único significado.

Por ejemplo, si tenemos nombres de países o numeraciones en diferentes bases transformamos todos estos nombres a un solo idioma.

## La etapa de transporte

Los datos son insertados en las bases columnares de la data mart se realizan pruebas de carga, gestión de errores, y una vez que esto esta completado tenemos listos los datos para empezar a realizar análisis.

## Acceso a datos

Es la ultima capa, es la razón de ser de todo el DW, que es la capa de acceso a datos, esta se puede implementar como una interfaz de reportes, ya sea web, o por medio de una aplicación, la cual llame a querys en el DW que se conecten entre los diferentes Data marts para obtener información que decíamos. Esta información puede darse en tablas, gráficos, o en otra representación grafica que se desee.

# Herramientas

Para finalizar se puede nombrar algunas de las herramientas para crear y diseñar Data Warehouse:

Base de datos columnares que facilite consultas para análisis:

- **Infobright**
- **Luciddb**
- **DB2**

Paquetes de software mas completos:

- **MicroStrategy**
- **Teradata**
- **Microsoft BI**
- **Oracle warehouse builder**

Qué poseen módulos para crear reportes y optimizar consultas y otros tipos de cosas que van mas allá de crear bases de datos.

Ian H. Witten, Eibe Frank., (2011) Data Mining Morgan Kaufman, pp629.  
Clasificación: Qa 76.9.D343 W58 2011.

Pang-Ning Tan, Michael Steinbach., (2006) Introduction To Data Mining  
Pearson pp760. Clasificación: Qa 76.9d343 T35 2006.

Jawei Han, Michael Kamber Jian Pei., (2011) Data Mining Concepts And  
Techniques Morgan Kaufman pp703 Clasificación: Qa 76.9.D343 H36 2011.

Thomas C, Hammergren., (2001) Data Warehousing For Dummies ,

Krish Krishnan., (2001) Data Warehousingin The Age Of Big Data

Richard O. Duda., (2001) Pattern Classification pp 654

Paulraj, Ponniah., (2010) Data Warehousing Fundamentals, Wiley



***GRACIAS POR SU ATENCIÓN***