

DW y ETL

OBJETIVO DW





ÍNDICE

1. ¿QUÉ ES UN DATAWAREHOUSE?
2. ANALIZANDO NUESTRO ORIGEN DE DATOS
3. CONSTRUYENDO NUESTRO DW
4. DISEÑO DE LA ODS Y DEL DW
5. AL TAJO... DIGOOO, AL TALEND

¿QUÉ ES UN DATAWAREHOUSE?



UN DW ES UN **ALMACÉN DE DATOS** CUYO OBJETIVO ES CONTENER DATOS QUE SON NECESARIOS PARA UNA ORGANIZACIÓN

OBJETIVOS PRINCIPALES DEL DATAWAREHOUSE

- INTEGRACIÓN DE LOS DATOS
- SEGMENTACIÓN DE LOS DATOS SEGÚN SU PROPÓSITO DE DIVULGACIÓN

¿QUÉ ES UN DATAWAREHOUSE?

ODS - El almacén operacional de datos es un contenedor de datos activos, diseñado para integrar datos de múltiples fuentes. Con una ventana de refresco pequeña o incluso en tiempo real, este almacén guarda una copia de los datos de los sistema origen, utilizándose como capa intermedia para un datawarehouse.

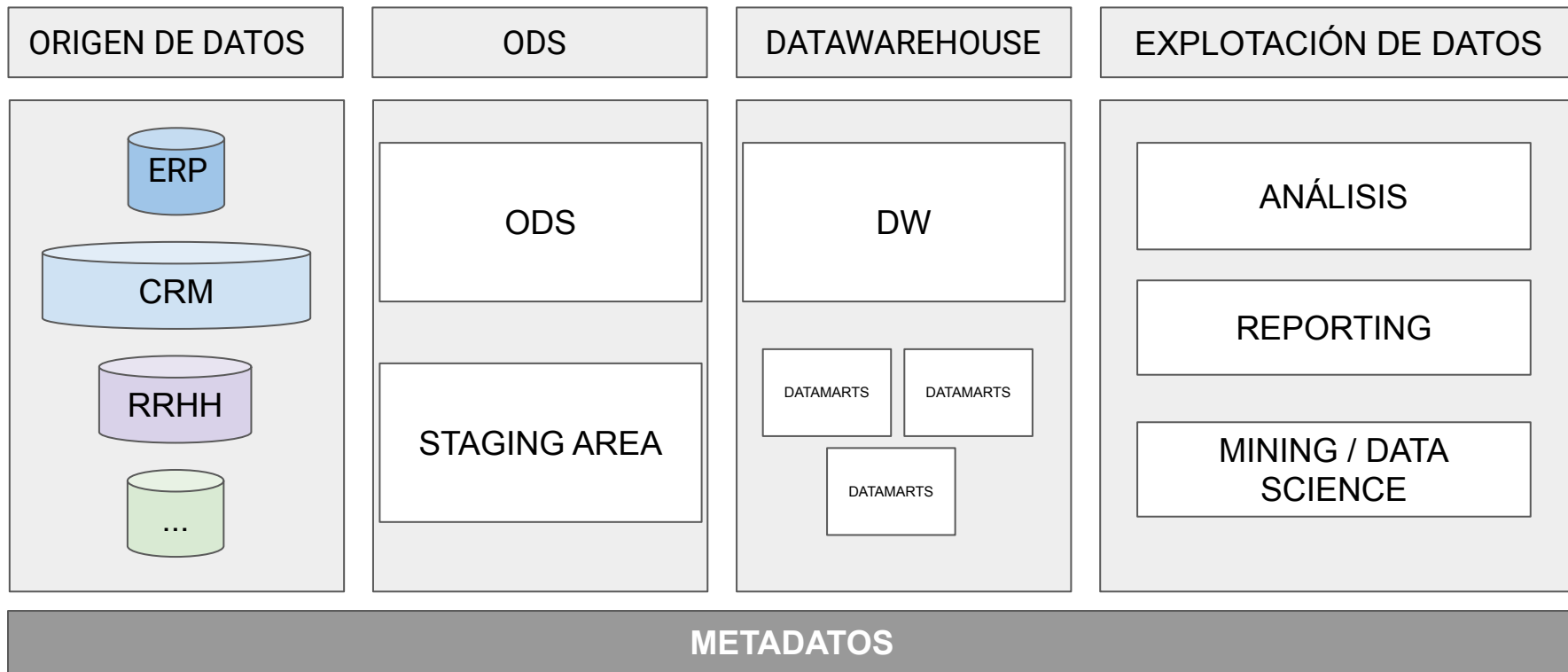
STAGING AREA - Es otro sistema intermedio cuyo objetivo es facilitar la extracción y la transformación de los datos. A diferencia del ODS, este sistema suele ser volátil, normalmente borrado tras la carga de datos final en el DW, o tras una fase de validación.

¿QUÉ ES UN DATAWAREHOUSE?

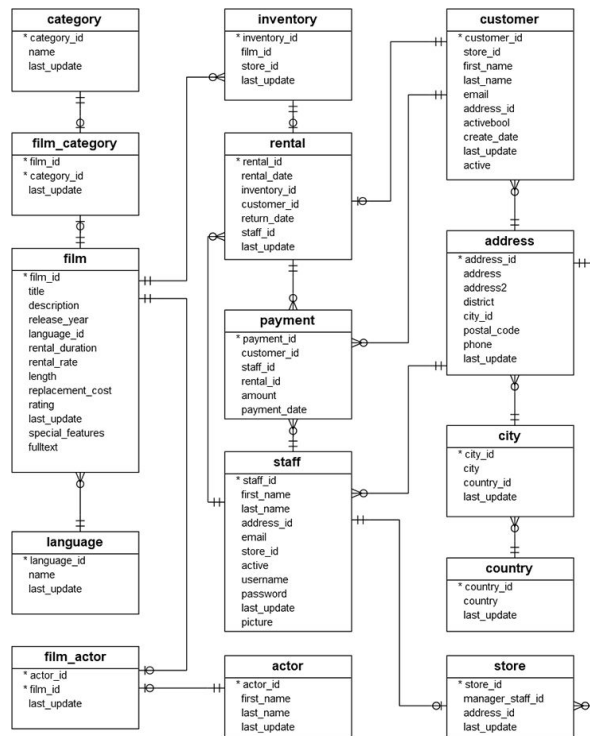
DATAMARTS - Llamamos Data Marts a los subconjuntos de datos de un DW.
Ejemplos: Data Mart Contable, Data Mart Recibos...

METADATOS - Se define comúnmente como “datos acerca de los datos”. El metadato documenta y responde a preguntas como, qué tablas existen en un DW, que columnas se utilizan y qué tipo de datos contienen, qué procesos explotan estas tablas y datos, o donde se está visualizando dicha información.

¿QUÉ ES UN DATAWAREHOUSE?



ANALIZANDO NUESTRO ORIGEN DE DATOS



NUESTRO SAMPLE DE DATOS
CONTIENE LA BASE DE DATOS DE UNA
EMPRESA FICTICIA QUE SE DEDICA A
ALQUILER DE PELIS

ANALIZANDO NUESTRO ORIGEN DE DATOS

LA BASE DE DATOS CONTIENE 15 TABLAS.

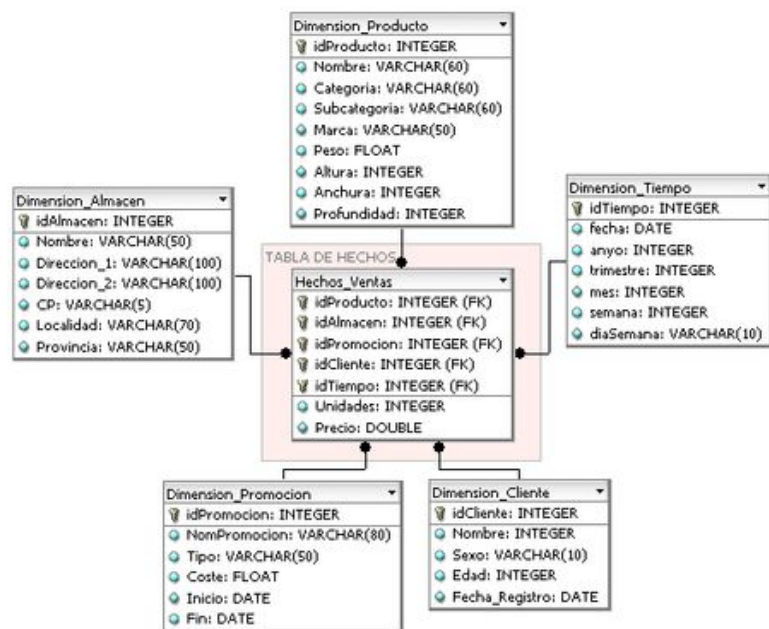
- actor – Datos de los actores (nombre, apellido...)
- film – Datos de las pelis (título, año, duración, rating...)
- film_actor – Tabla de relación de pelis y actores
- category – Datos de las categorías de las pelis
- film_category – Tabla de relación entre pelis y categorías
- store – Datos de los almacenes (manager, dirección...)
- inventory – Datos de inventario (peli, almacén donde se encuentra...)
- rental – Datos de los alquileres (fecha alquiler y devolución, peli, personal...)
- payment – Datos del pago del cliente (cantidad, fecha de pago...)
- staff – Datos del personal (datos personales, almacén donde trabaja...)
- customer – Datos de los clientes
- address – Datos de direcciones de clientes y staff de almacenes
- city – Datos de poblaciones (población, país)
- country – Datos de países

CONSTRUYENDO NUESTRO DW

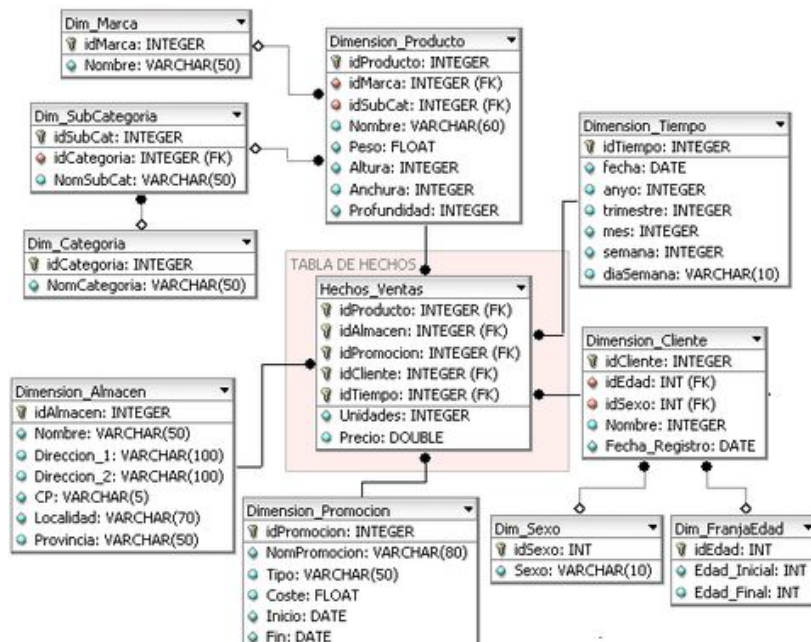
DIMENSIONES - Representan factores por lo que se analiza un determinado área del negocio. Son tablas normalmente pequeñas que sirven como maestros de información.
Ejemplo: tiempo, producto, tipo de cliente...

HECHOS - Las tablas de hechos (fact tables) son el eje, o la tabla principal de un modelo dimensional. Contienen campos clave que se unen a las tablas de Dimensión, y métricas con lo que queremos medir o analizar (cantidad, importe, precio, margen, nº de operaciones... Por lo general, se trata de tablas con gran volumen de datos.
Ejemplos: análisis de ventas, operaciones, llamadas de un call center...

CONSTRUYENDO NUESTRO DW



MODELO DE DATOS
EN ESTRELLA



MODELO DE DATOS
EN COPO DE NIEVE

CONSTRUYENDO NUESTRO DW

ANÁLISIS EVOLUCIÓN GLOBAL

1. Evolución de los **alquileres en el tiempo**, pudiendo ver de manera anual, trimestral o mensual.
2. El modelo debe permitir analizar la **evolución del nº de alquileres e importe**, pudiendo desglosar el dato por cliente, almacen, staff y categoría de pelis.

CONSTRUYENDO NUESTRO DW

ANÁLISIS POR CLIENTE

1. El modelo debe permitir **analizar la evolución de los clientes** de Netflix:
 - a. Cuales han generado más ingresos
 - b. Cuales son los clientes morosos (con pagos pendientes)
 - c. Cuales son los que tienen pelis por devolver
2. Debo poder ver **qué clientes puedo considerar VIP** según su gasto mensual en la empresa.
3. Permitir visualizar las **películas y/o categorías más vistas** por los clientes.
4. Permitir **análisis global y también desglosado por país y ciudad**.
5. **Top clientes por mes**, por almacén, ciudad o país.

CONSTRUYENDO NUESTRO DW

ANÁLISIS POR STAFF/ALMACÉN

1. **Qué empleados generan más ingresos.** Ranking Top empleados del mes. Visión por almacén, ciudad, país.
2. **Almacenes que más ingresos** generan.
3. **Análisis de alquileres por staff**, almacén y categoría de pelis. ¿Qué categorías alquila más un empleado concreto? ¿Qué empleado alquila más de cierta categoría de pelis?
4. Empleados que no generan ingresos y quien es su manager de almacén.

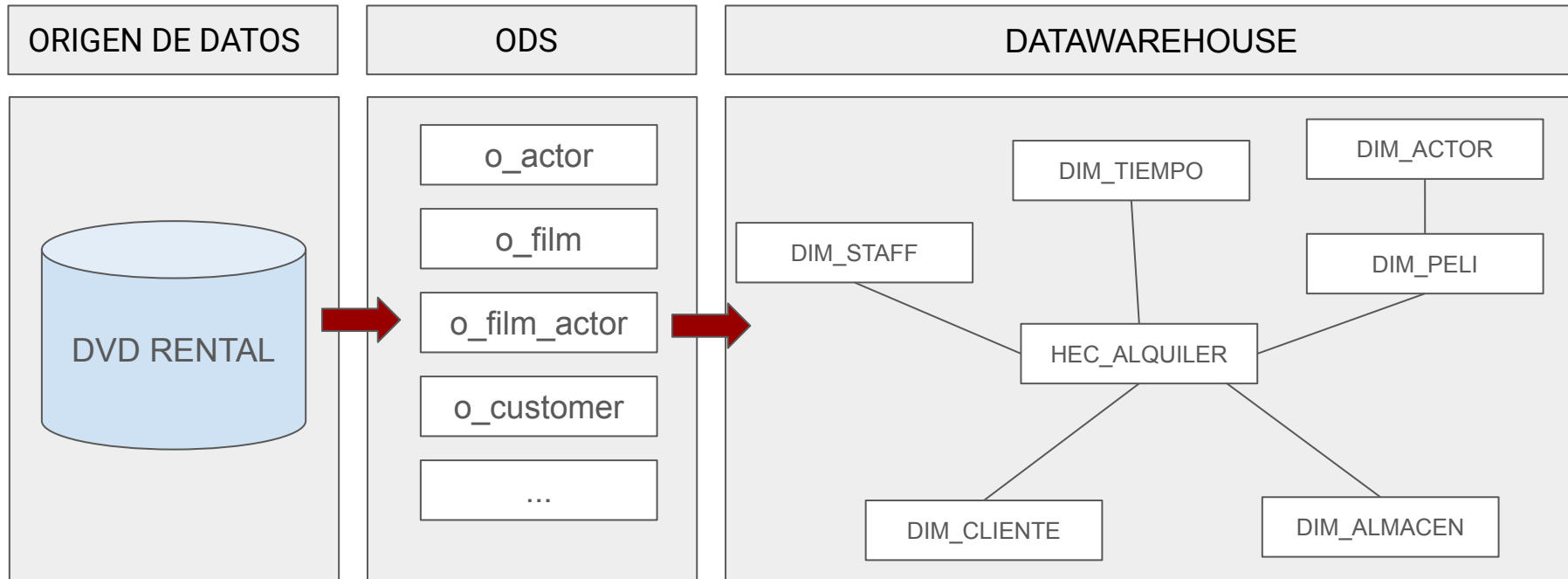


CONSTRUYENDO NUESTRO DW

ANÁLISIS DE FECHA

- 1.Cuál es la distribución del nº de clientes alquilando por mes, día, hora.
2. Qué periodos del año, mes tienen más actividad de alquiler

CONSTRUYENDO NUESTRO DW



DISEÑO DE LA ODS Y DEL DW

ANÁLISIS

1. **Definir alcance del proyecto** de integración de datos para disponer de una visión clara de los requisitos que deben cumplir los procesos ETL.
2. **Evaluar y analizar los datos origen.** Ubicación de los datos en origen, análisis de volumetrías

DISEÑO

1. **Diseño de los Modelos Destino.** Diseño del Modelo Lógico y Físico de datos del sistema destino.
2. **Diseño Mapa Fuente-Destino** de los datos. Diseño detallado de los mapas lógicos y técnicos de carga de los campos origen y destino.
3. **Diseño del Plan de Pruebas.**

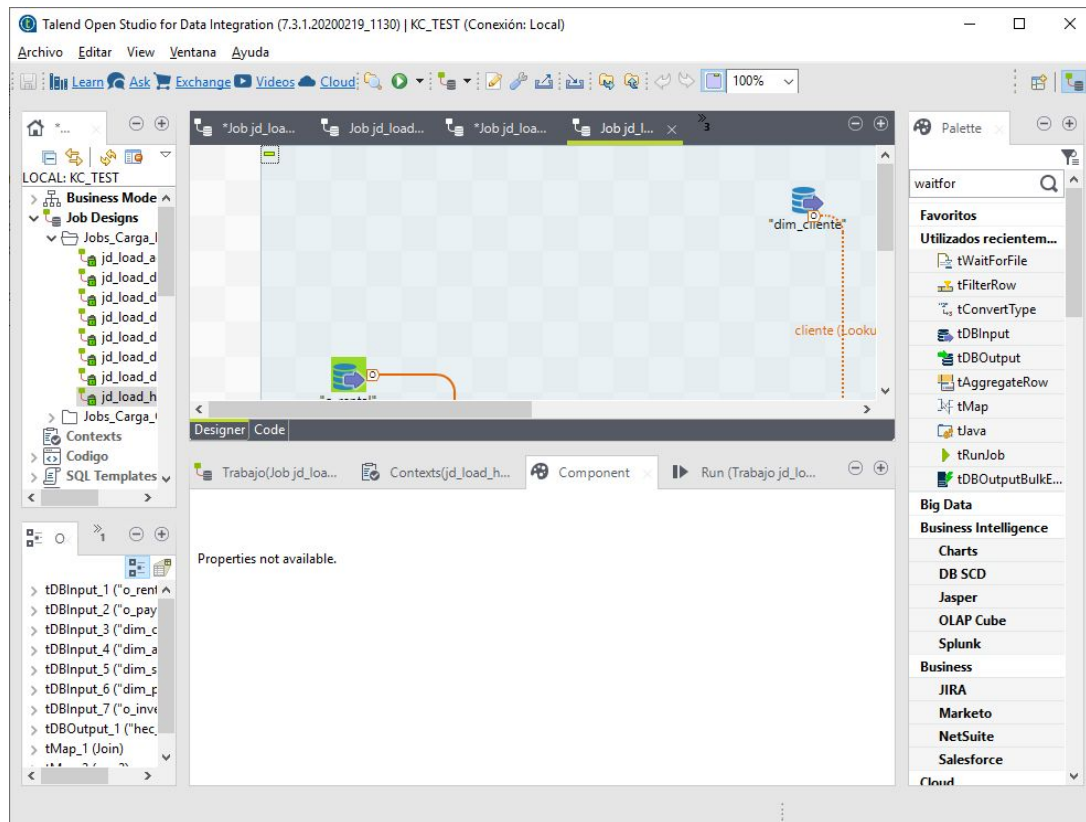
CONSTRUCCIÓN

1. **Construcción de los procesos ETL.**
2. **Ejecución y carga** de los datos.
3. **Ejecución del Plan de Pruebas** para certificar que los procesos está bien construidos y los datos corresponden a los requisitos establecidos.

IMPLANTACIÓN

1. **Implantación de los procesos ETL.** Traslado de los procesos al entorno de producción (real).
2. **Carga de datos históricos** (si procede)
3. **Planificación y ejecución** de los procesos.

AL TAJO... DIGOOO, AL TALEND





KEEPCODING

Tech School

Madrid | Barcelona | Bogotá

Jon Pineda Galvez

jpinedagalvez@gmail.com

<https://www.linkedin.com/in/jonathanpq>