



Universidad César Vallejo

Pregrado

# GESTIÓN DE DATOS E INFORMACIÓN II

30  
años

Licenciada por  
SUNEDU  
para que puedas  
salir adelante

## SESIÓN 08 INTEGRANDO DATOS I

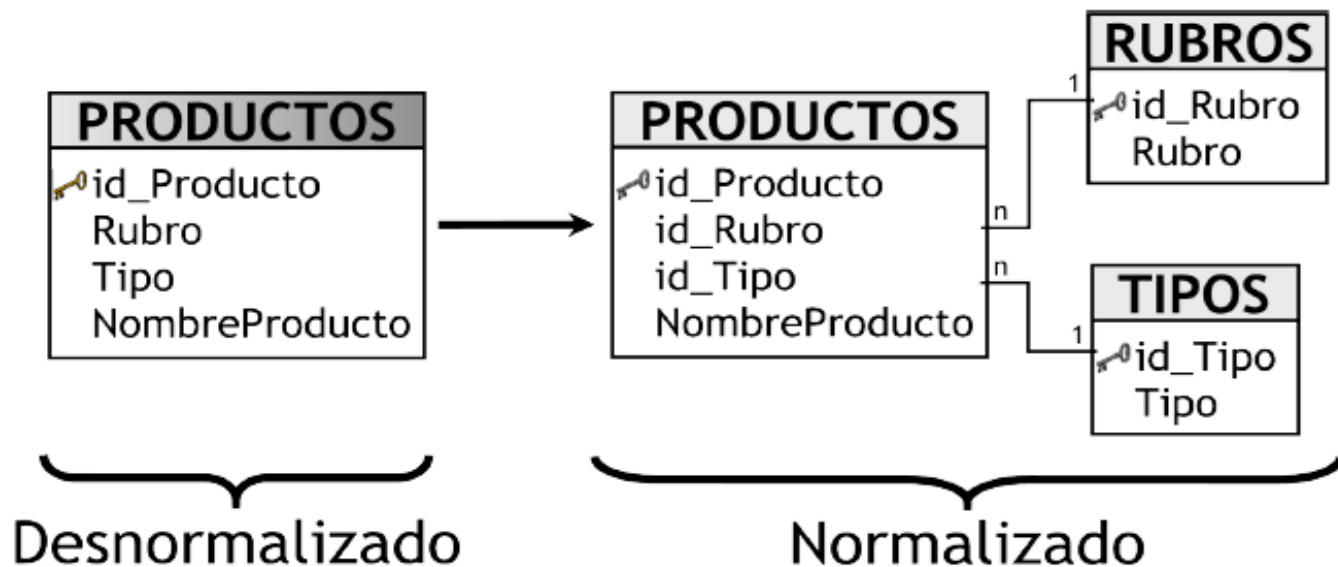


## DATAMART ESTRELLA

Consta de una tabla de hechos central y de varias tablas de dimensiones relacionadas a esta, a través de sus respectivas claves. ***Este modelo debe estar totalmente desnormalizado***, es decir que no puede presentarse en tercera forma normal (3ra FN).

Cuando se normaliza, se pretende eliminar la redundancia, la repetición de datos y que las claves sean independientes de las columnas, pero en este tipo de modelos se requiere no evitar precisamente esto.

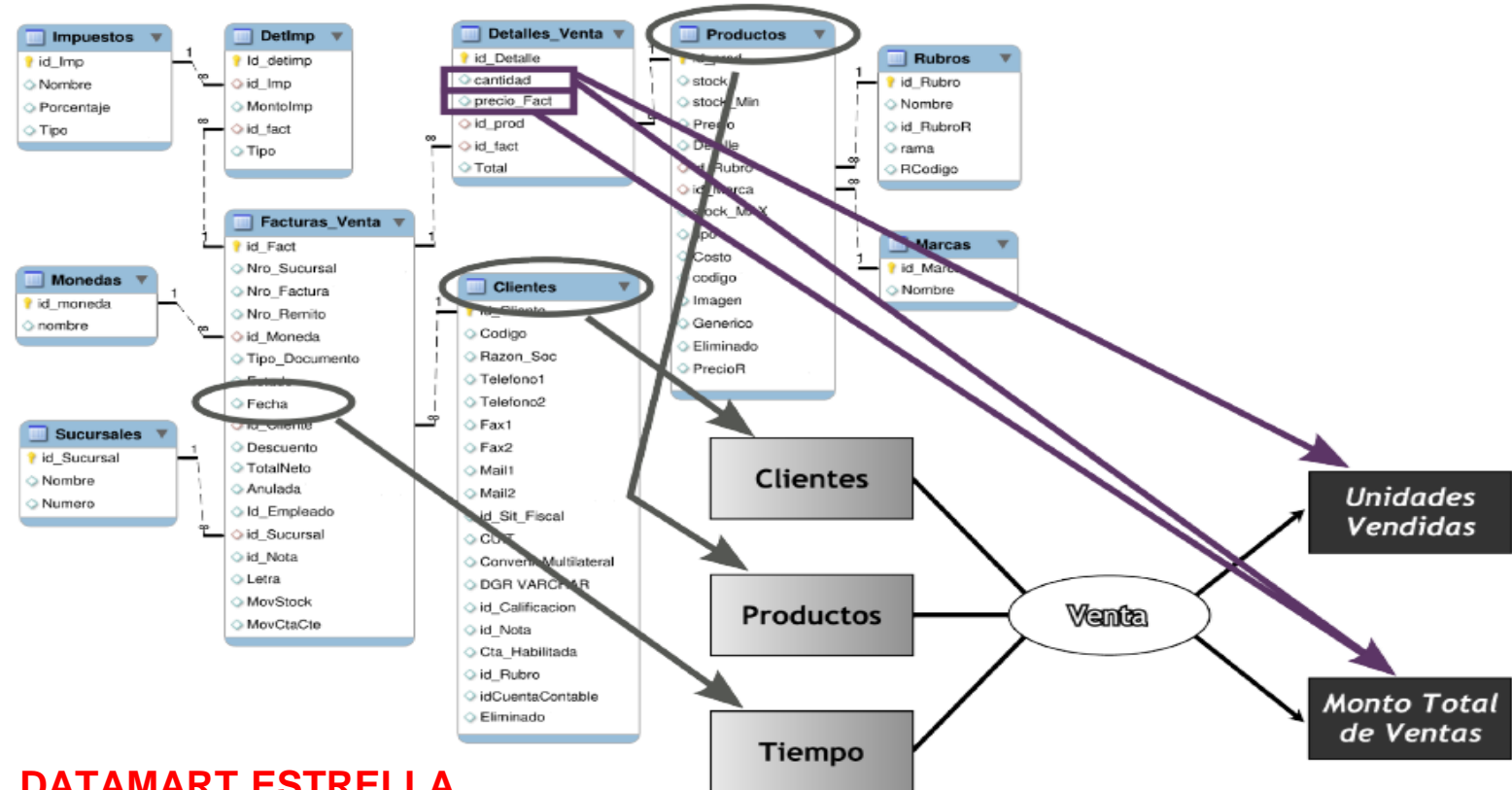
Las ventajas que trae aparejada la desnormalización, son las de obviar uniones (Join) entre las tablas cuando se realizan consultas, procurando así un mejor tiempo de respuesta y una mayor sencillez con respecto a su utilización. El punto en contra, es que se genera un cierto grado de redundancia, pero el ahorro de espacio no es significativo.



**DATAMART ESTRELLA**



# IMPLEMENTAR UN DATAMART ESTRELLA



**DATAMART ESTRELLA**



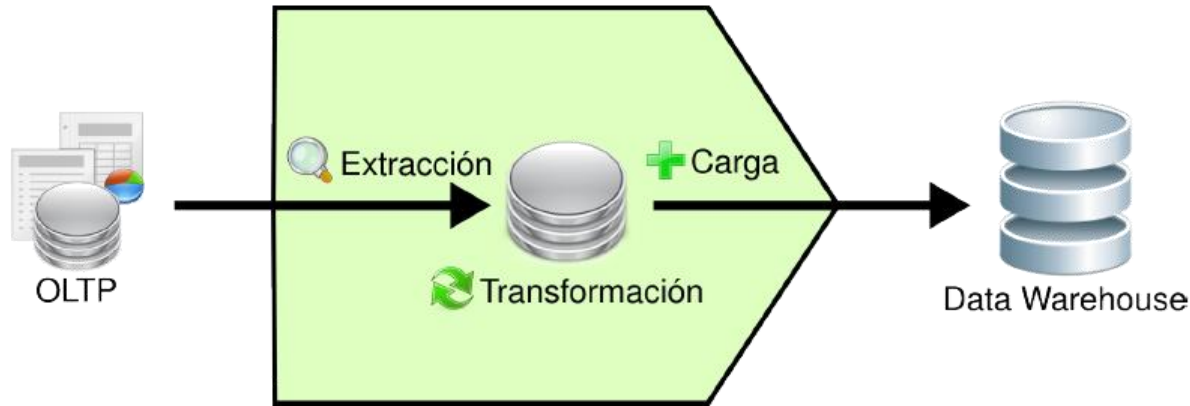
# PROCESO ETL

Pregrado

Ingeniería de  
Sistemas

Se extraen los datos relevantes desde los OLTP y se depositan en un almacenamiento intermedio.

- ✓ Se integran y transforman los datos, para evitar inconsistencias.
- ✓ Se cargan los datos desde el almacenamiento intermedio hasta el DW.





- ✓ Se exploran las diversas fuentes OLTP(Archivos de textos,Hipertextos, Hojas de cálculos, Bases de datos transaccionales ,Informes semanales, mensuales, anuales, etc.) que se tengan a disposición, y se extrae la información que se considere relevante al caso.
- ✓ Si los datos operacionales residen en un SGBD Relacional, el proceso de extracción se puede reducir a, por ej., consultas en SQL o rutinas programadas.
- ✓ Si se encuentran en un sistema no convencional o fuentes externas, ya sean textuales, hipertextuales, hojas de cálculos, etc., la obtención de los mismos puede ser un tanto más dificultoso, debido a que, por ej., se tendrán que realizar cambios de formato y/o volcado de información a partir de alguna herramienta específica.



# PROCESO DE TRANSFORMACIÓN

Pregrado

Ingeniería de  
Sistemas

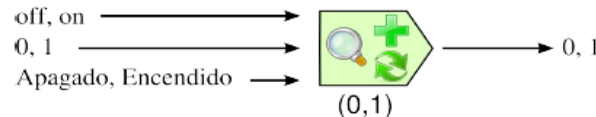
**Codificación.-** Una inconsistencia muy típica que se encuentra al intentar integrar varias fuentes de datos, es la de contar con más de una forma de codificar un atributo en común. Por ej., en el campo “estado”, algunos diseñadores completan su valor con “0” y “1”, otros con “Apagado” y “Encendido”, otros con “off” y “on”, etc. Lo que se debe realizar en estos casos, es seleccionar o recodificar estos atributos, para que cuando la información llegue al DW, esté integrada de manera uniforme.



OLTP



Data Warehouse





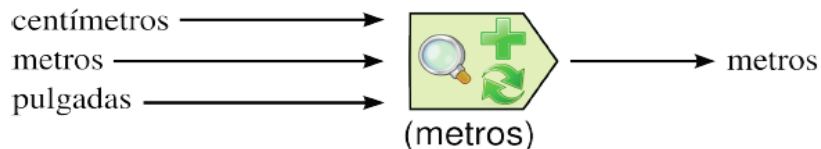
**Medida de atributos.-** Los tipos de unidades de medidas utilizados para representar los atributos de una entidad, varían considerablemente entre sí, a través de los diferentes OLTP. Por ej. al registrar la longitud de un producto determinado, de acuerdo a la aplicación que se emplee para tal fin, las unidades de medidas pueden ser explicitadas en centímetros, metros, pulgadas, etc.



OLTP



Data Warehouse







## PROCESO DE TRANSFORMACIÓN

**Convenciones de nombramiento.-** Usualmente, un mismo atributo es nombrado de diversas maneras en los diferentes OLTP. Por ej. , al referirse al nombre del proveedor, puede hacerse como “nombre”, “razón\_social”, “proveedor”, etc. Aquí, se debe utilizar la convención de nombramiento que para los usuarios sea más comprensible.



OLTP



Data Warehouse





## PROCESO DE CARGA

Pregrado

Ingeniería de  
Sistemas

**Carga inicial.-** Se refiere precisamente a la primera carga de datos que se le realizará al DW. Por lo gral., esta tarea consume un tiempo bastante considerable, ya que se deben insertar registros que han sido generados aproximadamente, y en casos ideales, durante más de 5 años.

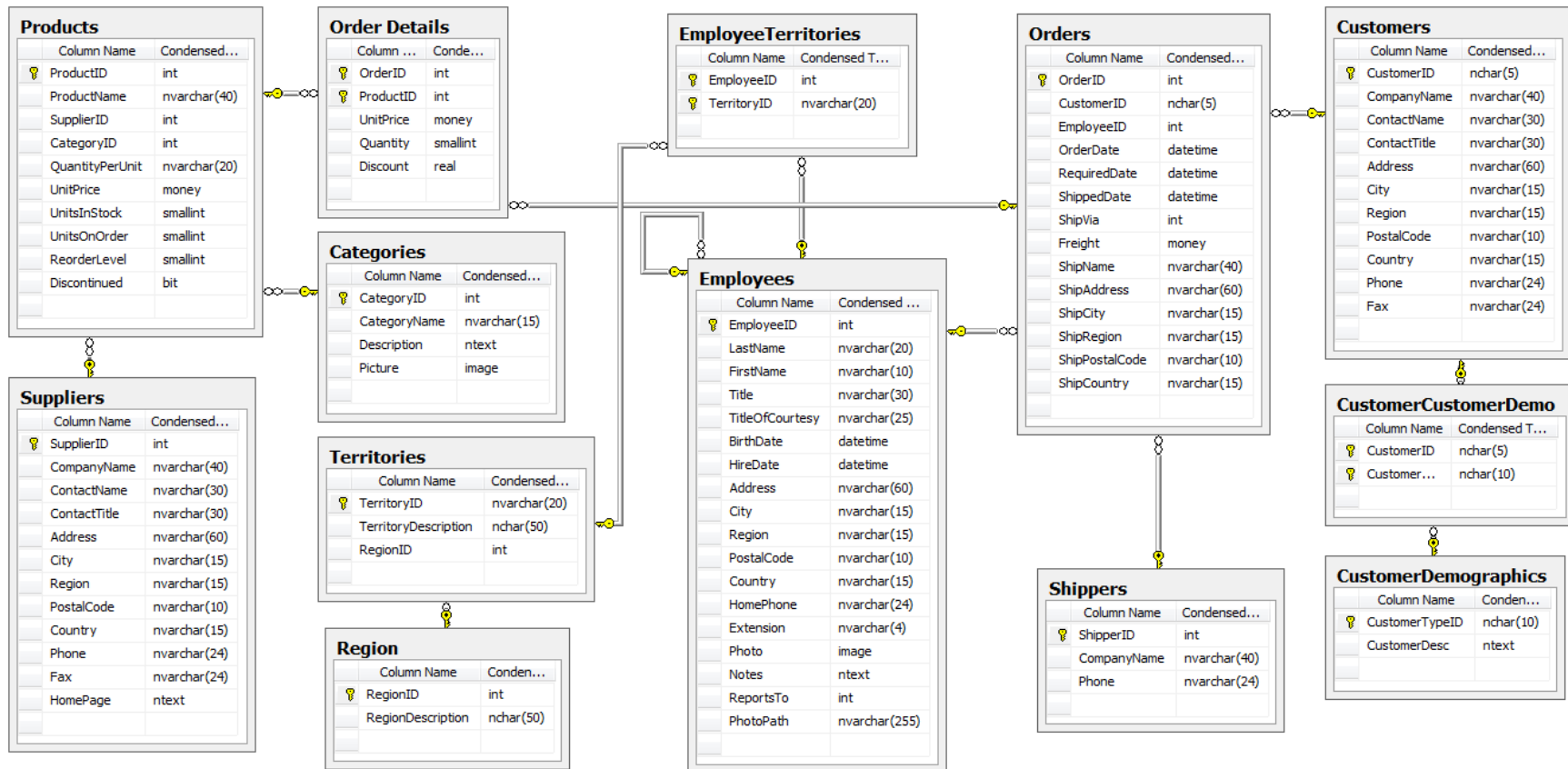
**Actualización o mantenimiento periódico.-** Los mantenimientos periódicos mueven pequeños volúmenes de datos, y su frecuencia está dada en función del gránulo del DW y los requerimientos de los usuarios. El objetivo de esta tarea es añadir al depósito aquellos datos nuevos que se fueron generando desde el último refresco.



# IMPLEMENTAR PROYECTO DATAMART

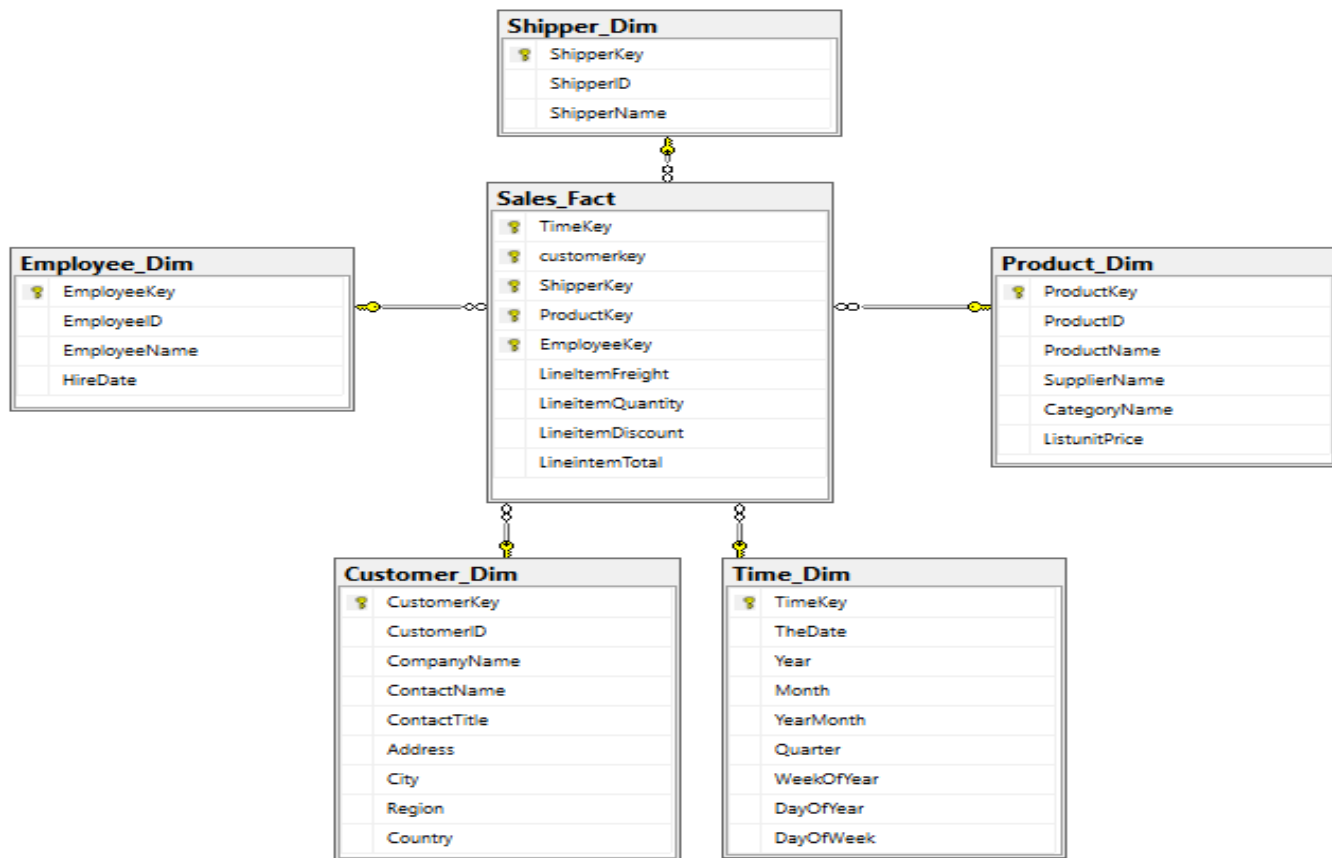
Pregrado

Ingeniería de  
Sistemas





# IMPLEMENTAR PROYECTO DATAMART NORTHWIND





## ¿QUÉ HEMOS APRENDIDO HOY?



Para que reflexionen y entiendan la importancia de los temas tratados y el mejoramiento de su propio proceso de aprendizaje.



Universidad **César Vallejo**

Licenciada por Sunedu  
para que puedas salir adelante