

UNIVERSIDAD GALILEO

Postgrado en Análisis y Predicción de Datos

Curso: Diseño y Construcción de Data Warehouses Horario: Lun: 18:00 – 21:00

Tutor: Msc. José Rolando Lucero Morataya



**Diseño y gestión de Data Warehouses**

**Semana 4**

*Galileo*  
UNIVERSIDAD

Integrantes:

Andrea María Hernández Marroquín – 240110074

Otto Francisco Chamo Cheley – 19001395

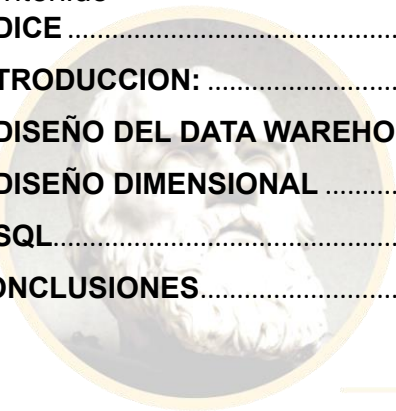
Edgar Geovany Ocaña Orozco – 24010004

Jose Eduardo Calderon Lopez - 14001764

Guatemala, 24 de marzo de 2024

## INDICE

Contenido	
<b>INDICE</b> .....	2
<b>INTRODUCCION:</b> .....	3
<b>DISEÑO DEL DATA WAREHOUSE</b> .....	4
<b>DISEÑO DIMENSIONAL</b> .....	5
<b>SQL</b> .....	6
<b>CONCLUSIONES</b> .....	8



*Galileo*  
UNIVERSIDAD

## INTRODUCCION:

El diseño dimensional en un Data Warehouse (DW) es un enfoque de modelado que se utiliza para organizar y estructurar los datos de manera que sean óptimos para el análisis y la generación de informes. A diferencia de los modelos relacionales tradicionales, que están diseñados para transacciones y operaciones de base de datos, el diseño dimensional se centra en facilitar el análisis de datos mediante la creación de estructuras fáciles de entender y consultar.

En un diseño dimensional, los datos se organizan en dos tipos principales de tablas: tablas de hechos y tablas de dimensiones. Las tablas de hechos contienen métricas o medidas cuantitativas que se están analizando, mientras que las tablas de dimensiones contienen descripciones o atributos que proporcionan contexto a las métricas. Este enfoque permite a los usuarios realizar consultas complejas y analizar datos desde diferentes perspectivas.

La presente tarea se presenta un detallado diagrama dimensional para una empresa que busca analizar información de un supermercado para poder tomar decisiones.

## **DISEÑO DEL DATA WAREHOUSE**

Para realizar el diseño dimensional para un Supermercado se propone lo siguiente la creación de las siguientes tablas:

1. Tabla customer\_dimension

Se decidió por construir una dimensión que se lleva registro de los datos del cliente, que contenga información sobre el segmento y código postal de donde hace la compra.

2. Tabla ship\_mode\_dimension

Se construyo una tabla que lleve registro de los tipos modos de entrega para cada uno de los pedidos, para lograr identificar cual es el más usado y, si eventualmente el negocio llega a ofrecer una nueva alternativa.

3. Tabla geography\_dimension

Esta tabla llevara el registro de todos los países y diferentes regiones a las que se ha enviado un cargamento.

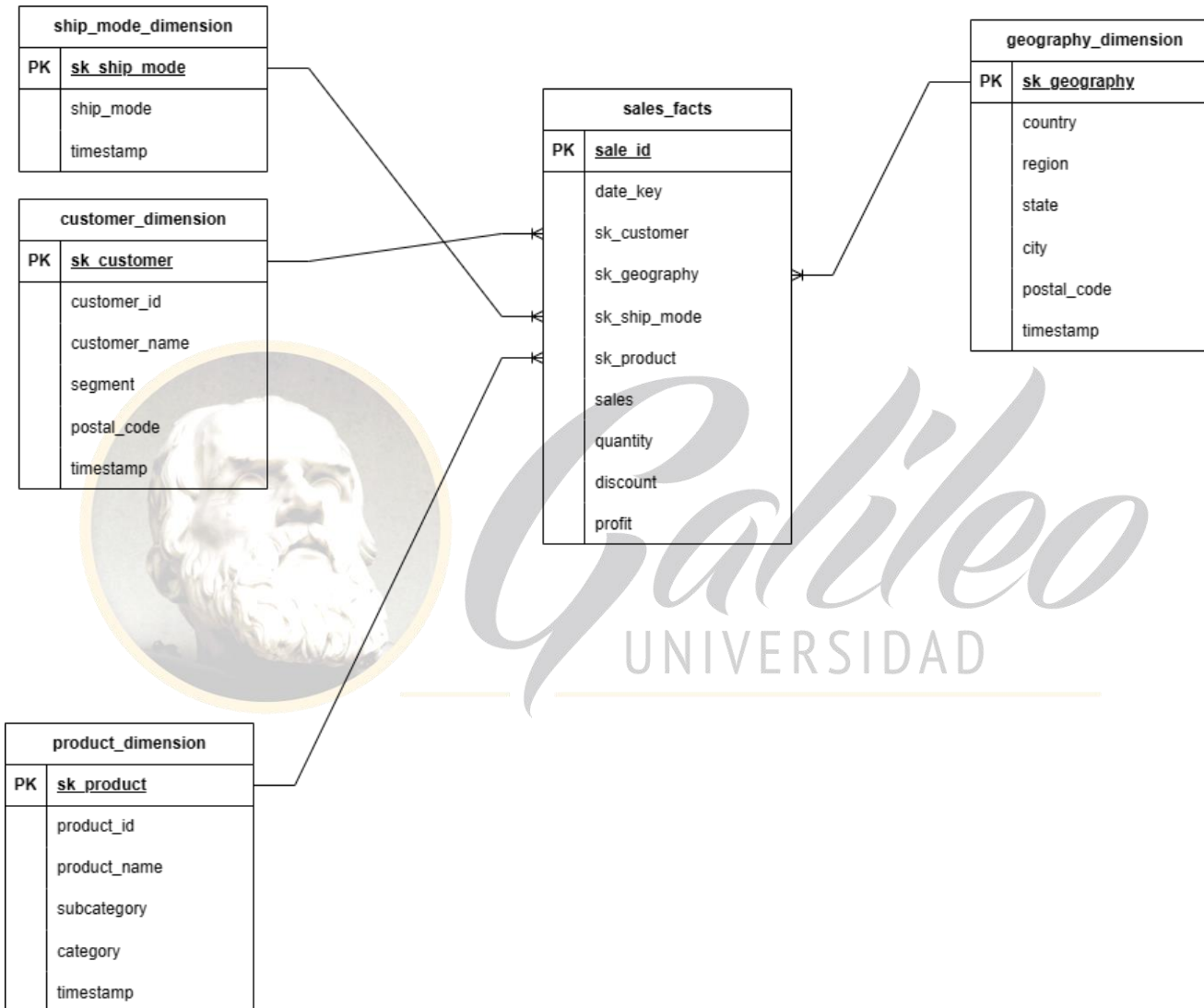
4. Tabla producto\_dimension

Una tabla que lleva registro de todos los productos que tiene la empresa y, que ira creciendo según la empresa ofrezca nuevos productos.

5. Tabla sales\_facts

La tabla de hechos que lleva registro de todas las variables que buscamos analizar, lleva registro de cada una de las ventas, cantidad, descuentos y ganancias obtenidas.

## DISEÑO DIMENSIONAL



## SQL

```
CREATE DATABASE IF NOT EXISTS retail_sales;

USE retail_sales;

CREATE TABLE customer_dimension (
    sk_customer    INT NOT NULL auto_increment PRIMARY KEY,
    customer_id    VARCHAR (255) NULL,
    customer_name  VARCHAR (255) NULL,
    segment        VARCHAR (255) NULL,
    postal_code    int NULL,
    timestamp      DATETIME
);

CREATE TABLE ship_mode_dimension
(
    sk_ship_mode    INT NOT NULL auto_increment PRIMARY KEY,
    ship_mode       VARCHAR (255) NULL,
    timestamp       DATETIME
);

CREATE TABLE geography_dimension
(
    sk_geography    INT NOT NULL auto_increment PRIMARY KEY,
    country         VARCHAR (255) NULL,
    region          VARCHAR (255) NULL,
    state           VARCHAR (255) NULL,
    city            VARCHAR (255) NULL,
    postal_code     int NULL,
    timestamp       DATETIME
);

CREATE TABLE product_dimension
(
    sk_product      INT NOT NULL auto_increment PRIMARY KEY,
    product_id      VARCHAR (255) NULL,
    product_name    VARCHAR (255) NULL,
    subcategory     VARCHAR (255) NULL,
    category        VARCHAR (255) NULL,
    timestamp       DATETIME
);
```

```
CREATE TABLE sales_facts
(
  sale_id          VARCHAR (255) NOT NULL,
  date_key         INT NOT NULL,
  sk_customer      INT NOT NULL,
  sk_geography     INT NOT NULL,
  sk_ship_mode     INT NOT NULL,
  sk_product       INT NOT NULL,
  sales            DECIMAL (18, 2) NULL,
  quantity         INT NULL,
  discount         DECIMAL (18, 2) NULL,
  profit          DECIMAL(18, 2) NULL,
  FOREIGN KEY (sk_customer) REFERENCES customer_dimension (sk_customer),
  FOREIGN KEY (sk_geography) REFERENCES geography_dimension(sk_geography),
  FOREIGN KEY (sk_ship_mode) REFERENCES ship_mode_dimension(sk_ship_mode),
  FOREIGN KEY (sk_product) REFERENCES product_dimension(sk_product)
);
```



*Galileo*  
UNIVERSIDAD

## CONCLUSIONES

Con el diseño del Data Warehouse anterior se desea convertir los datos en información valiosa y accionable para la toma de decisiones empresariales. Al proporcionar una estructura clara y organizada para los datos, se facilita el análisis, se mejora el rendimiento y se impulsa la colaboración en toda la organización.

