



# Módulo 8

## Limpieza de datos con SQL



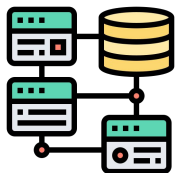
## AGENDA DE LA CLASE

- ✓ Bases de datos en el análisis computacional de datos
- ✓ SQL Vs Hojas de Cálculos
- ✓ Comandos SQL para limpieza de datos



# 01 - Bases de datos en el análisis computacional de datos

Las **bases de datos** permiten a los analistas **manipular, almacenar y procesar datos**. Esto les ayuda a buscar datos de manera mucho más eficiente para obtener la mejor información.



Una **base de datos relacional** es una base de datos que contiene una serie de **tablas** que se pueden **conectar para mostrar relaciones**.



Las tablas de una base de datos relacional están **conectadas** por los campos que tienen en común mediante **claves**.

# 01 - Bases de datos en el análisis computacional de datos



Una **clave primaria** es un identificador que hace referencia a una columna en la que cada valor es único.

El valor asignado a la clave primaria en una fila determinada debe ser **único en toda la tabla**.



Una **clave externa** es un campo en una tabla que es una clave primaria en otra tabla.

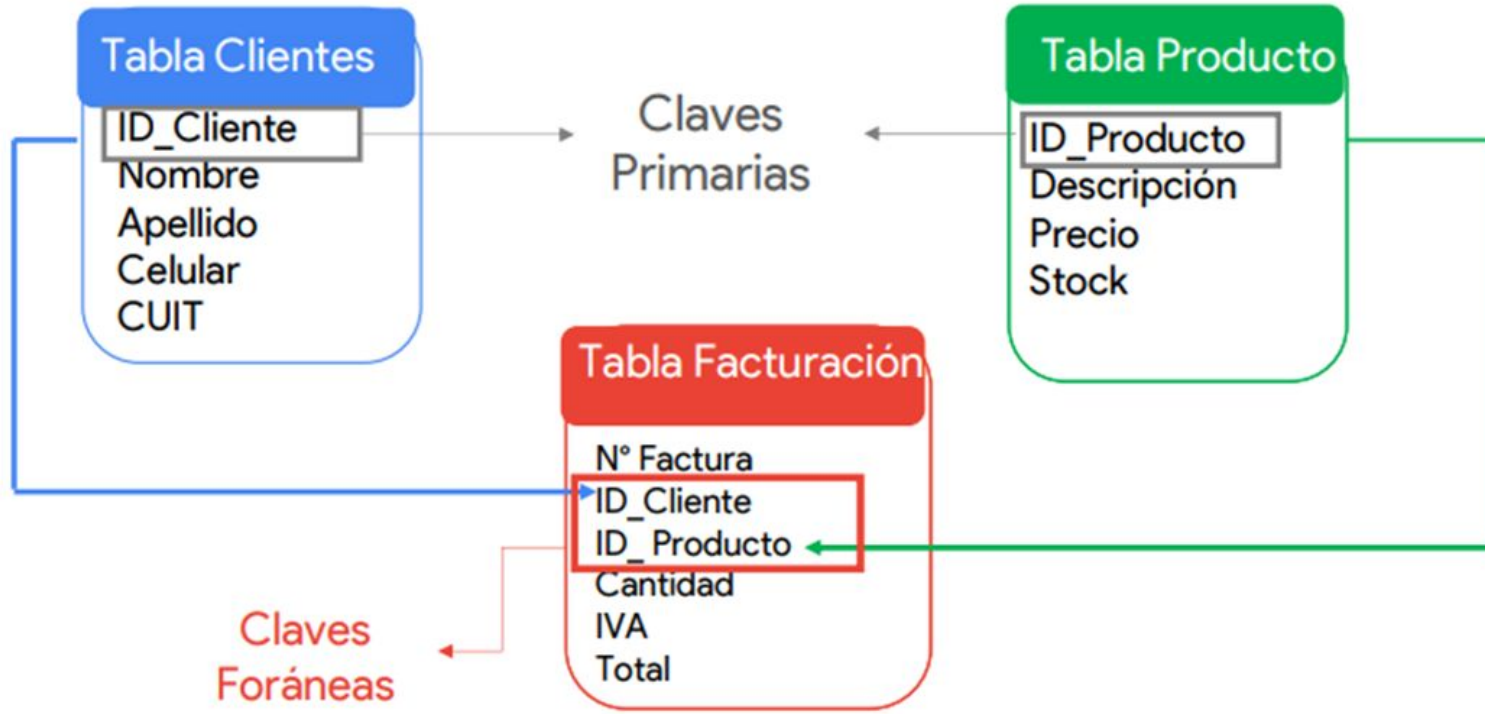
Una tabla debe tener solo una clave primaria, pero puede tener **varias claves externas**.



Una **clave compuesta** es un tipo de clave primaria que se construye a partir de varias columnas de una tabla.

*Por ejemplo, cuando se juntan las columnas `customer_id` y `location_id` para dar una clave primaria.*

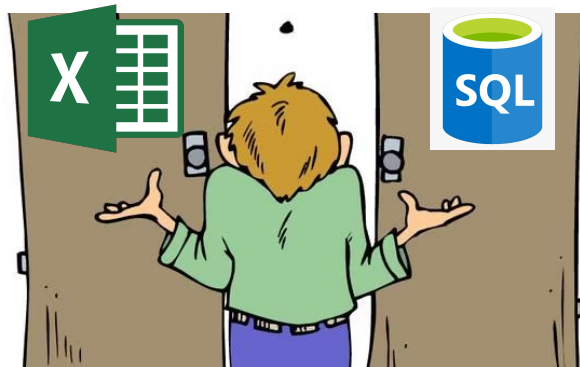
## 01 - Bases de datos en el análisis computacional de datos



## 02 - SQL Vs Hojas de Cálculos

Clase 2

### HOJAS DE CÁLCULO



### SQL

- Son programas
- Información local
- Base limitada capacidad (Max. 1 millón de registros o filas)
- Ingreso manual
- Datos estructurados en filas y columnas.
- Un usuario a la vez
- Controlado por el usuario
- Visualización con gráficos
- MEJOR PARA EL TRABAJO SOLOS

- Es un lenguaje
- Interacción con BD externas
- Bases de datos extensas (no reduce la velocidad de consulta)
- Ingreso de datos por Querys
- Tablas relacionadas
- Usuarios simultáneos
- Controlado por el sistema de gestión de bases de datos
- MEJOR PARA EL TRABAJO COLABORATIVO

## 03 - Comandos SQL para limpieza de datos

○ **Insertar** registros faltantes:

```
INSERT INTO (columna 1, columna 2,...)
VALUES (VALOR 1, VALOR 2, ...)
```

○ **Actualizaciones** de valores:

```
UPDATE Base
SET Campo a actualizar = 'nuevo valor'
WHERE ...
```

○ **Cantidad de caracteres**  
de un campo:

```
SELECT LENGHT (Campo a medir) AS NOMBRE_COLUMNA
FROM ...WHERE
```

○ **Eliminación duplicados:**

```
SELECT DISTINCT ...
FROM ...WHERE
```

○ **Extracción de caracteres**  
específicos:

```
SUBSTR (NombreCampo, posición inicial,
cantidad de caracteres) AS...
```

○ **Juntar campos:**

```
SELECT CONCAT (NombreCampo1,
NombreCampo2) AS...
```

## 03 - Comandos SQL para limpieza de datos

○ Eliminar espacios en blanco:

```
TRIM (NombreCampo)
```

○ Máximos y mínimos de un campo:

```
SELECT  
MAX (NombreCampo) AS...,  
MIN (NombreCampo) AS..
```

○ Cambiar formato de campo:

```
SELECT  
CAST (NombreCampo AS  
Tipo de campo)  
FROM...
```

○ Rellenar campos:

```
SELECT  
COALESCE (Nombre  
campo1, Nombre campo2)  
AS Tipo de campo  
FROM...
```

○ Borrar registros:

```
DELETE Base  
WHERE condición para el  
borrado
```





¿PREGUNTAS?





**¡Muchas gracias!**