

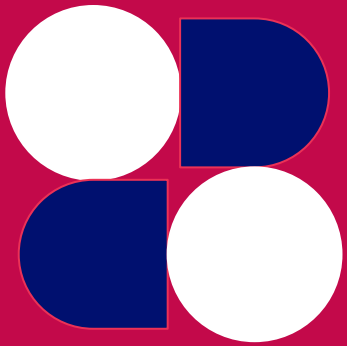


Programa PUCP DATA DRIVEN

Formación tecnológica

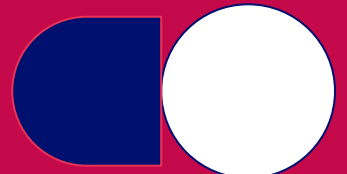
Dirección Académica de Planeamiento y Evaluación
Dirección de Gestión del Talento Humano





SESIÓN 11:

Introducción al modelado de datos





Al finalizar la sesión, lograrás:

- Comprender los conceptos básicos de modelado de datos
- Identificar y establecer relaciones entre tablas
- Aplicar buenas prácticas en el modelado de datos
- Desarrollar un modelo de datos básico



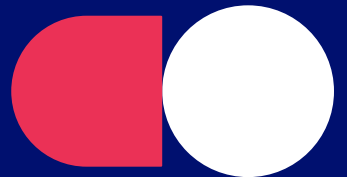
Contenido:



- 1 ¿Qué es un modelo de datos?
- 2 Conceptos básicos de modelado de datos
- 3 Diferenciación entre OLAP (Online Analytical Processing) y OLTP (Online Transactional Processing).



1. ¿Qué es un modelo de datos?





1. ¿Qué es un modelo de datos?

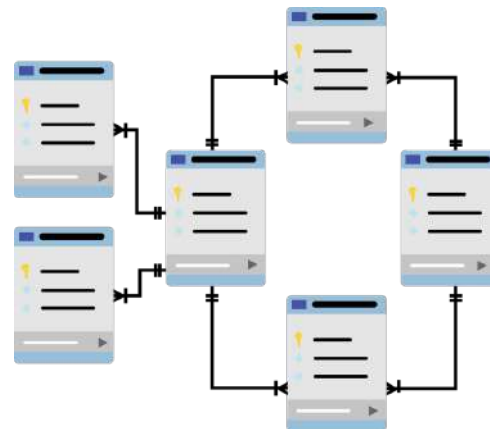


Definición:

- Proceso de estructuración y organización de datos para su análisis.
- Involucra la creación de tablas y la definición de relaciones entre ellas.

Importancia:

- Facilita la comprensión y el análisis de grandes volúmenes de datos.
- Mejora la precisión y eficiencia de los informes y dashboards.





1. Objetivos del Modelado de Datos



Organización:

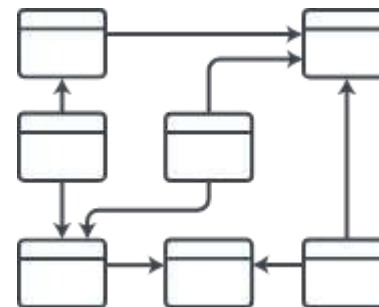
- Crear una estructura lógica y coherente para los datos.

Integridad:

- Asegurar la consistencia y exactitud de los datos.

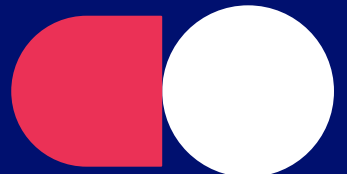
Eficiencia:

- Optimizar el rendimiento de consultas y cálculos en los informes.





2. Conceptos básicos de modelado de datos





2. Tablas



- **Tablas de Hechos:**
Contienen datos transaccionales como ventas, transacciones o registros de eventos.
Ejemplo: tabla de ventas.
- **Tablas de Dimensiones:**
Contienen datos descriptivos relacionados con las tablas de hechos, como productos, clientes o fechas.
Ejemplo: tabla de productos.

FactResellerSales	
🔑	SalesOrderNumber
🔑	SalesOrderLineNumber
	ClaveProducto
	OrderDateKey
	DueDateKey
	ShipDateKey
	ResellerKey
	EmployeeKey
	SalesTerritoryKey
	OrderQuantity
	TotalProductCost
	ImporteVentas

ProductoDim	
🔑	ClaveProducto
	ProductAlternateKey
	...

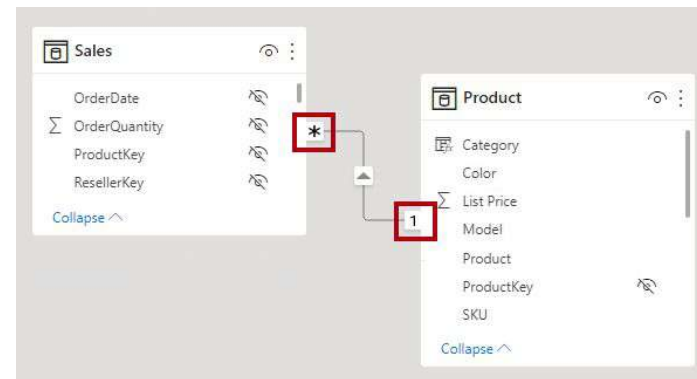


2. Relaciones - Definición



En el contexto del modelado de datos, una relación es un vínculo establecido entre dos tablas que define cómo los datos de una tabla se conectan con los datos de otra tabla.

Las relaciones son fundamentales para integrar y analizar datos provenientes de diferentes fuentes o tablas dentro de un mismo modelo de datos.



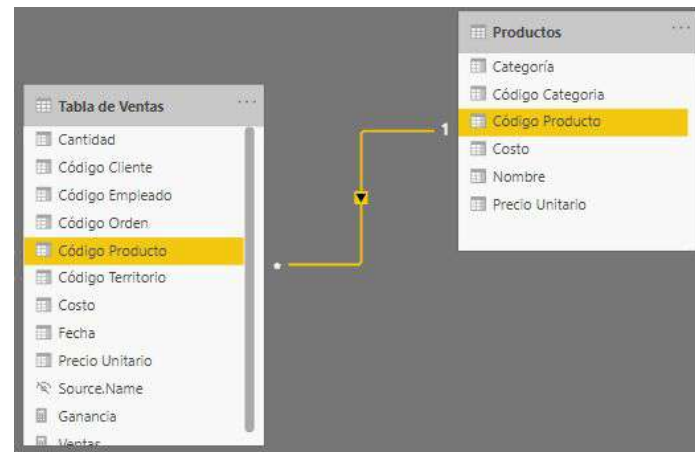


2. Relaciones - Importancia



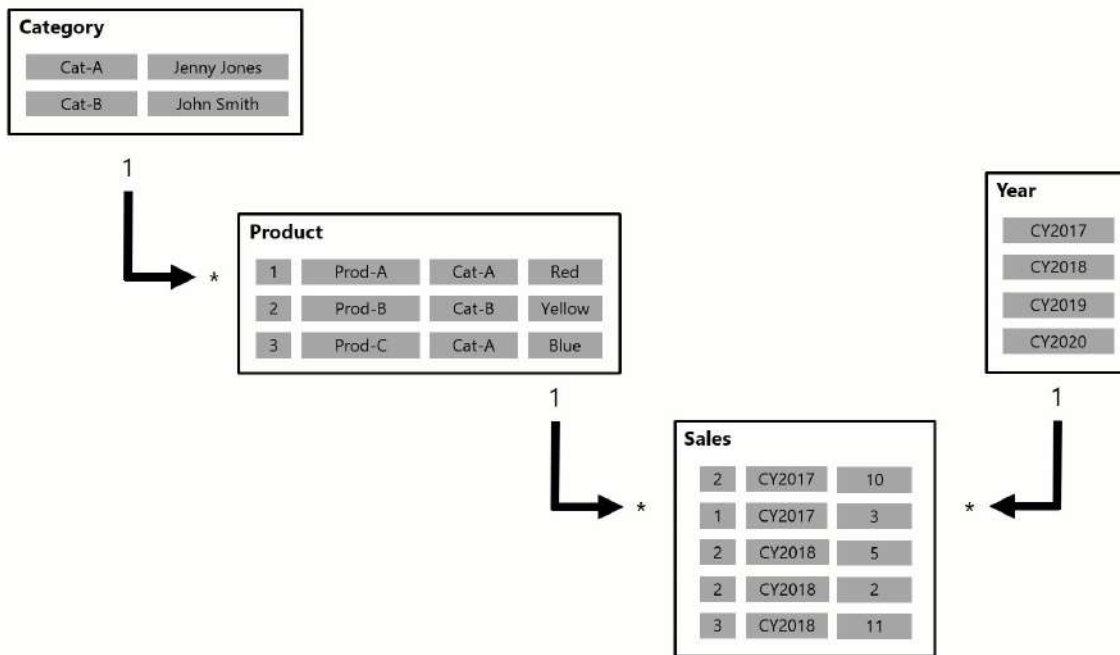
Las relaciones permiten a los analistas y a las herramientas de BI, como Power BI, combinar y comparar datos de manera efectiva. Sin relaciones, cada tabla se trataría de manera aislada, lo que dificultaría la creación de informes coherentes y significativos que abarquen múltiples aspectos del negocio.

Una relación de modelo propaga los filtros aplicados en la columna de una tabla de modelos a otra tabla de modelos. Los filtros se propagarán siempre que haya una ruta de relación que seguir, lo que puede implicar la propagación a varias tablas.





2. Relaciones - Importancia

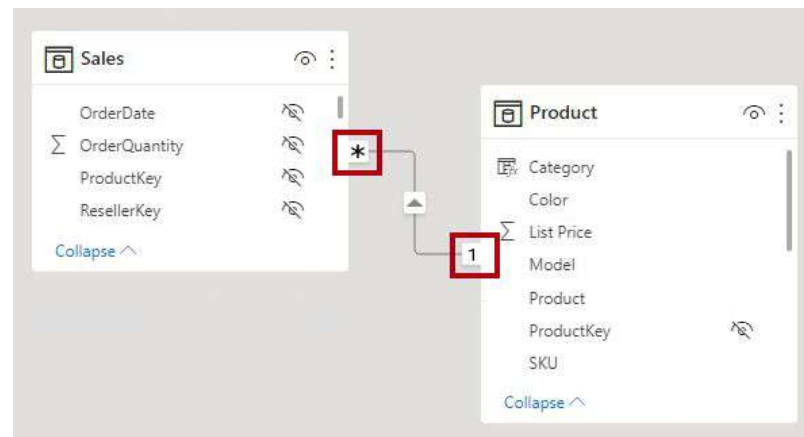




2. Llaves primarias y foráneas



- **Llave Primaria:** Un identificador único para cada fila en una tabla.
Por ejemplo, ID_Producto en una tabla de productos.
- **Llave Foránea:** Un campo en otra tabla que hace referencia a la llave primaria.
Por ejemplo, ID_Producto en una tabla de ventas que indica qué producto se vendió.

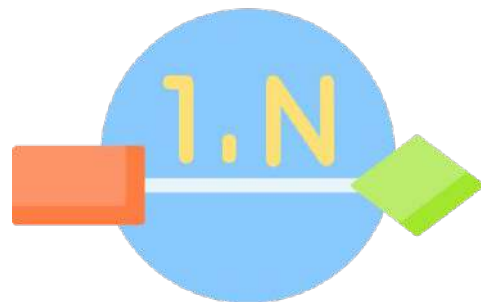




2. Cardinalidad



- **Uno a Uno (1:1):** Cada fila en una tabla tiene una única fila correspondiente en la otra tabla.
- **Uno a Muchos (1:*):** Una fila en una tabla puede tener múltiples filas correspondientes en la otra tabla. Es la más común.
- **Muchos a Muchos (*:*):** Más complejo, donde varias filas en una tabla pueden corresponder a varias filas en la otra tabla. A menudo se evita en modelos simples.





2. Llaves primarias y foráneas



Tabla Estudiantes

ID_Estudiante	Nombre
1	Juan Pérez
2	Ana Gómez
3	Luis García

¿Cuál sería la llave primaria y llave foránea?

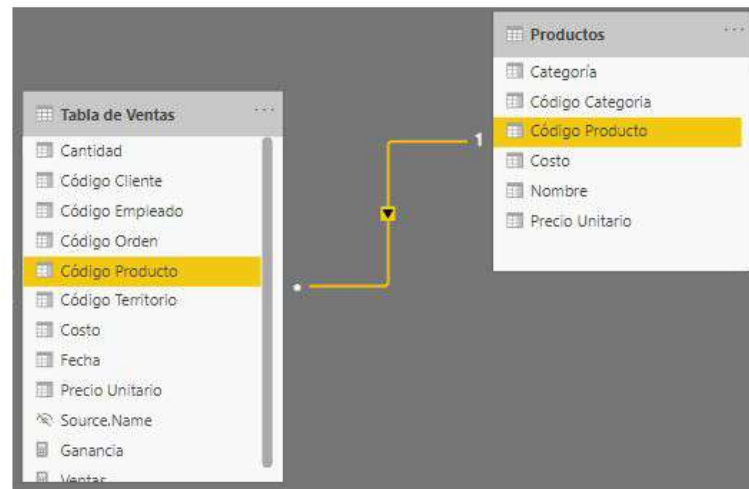
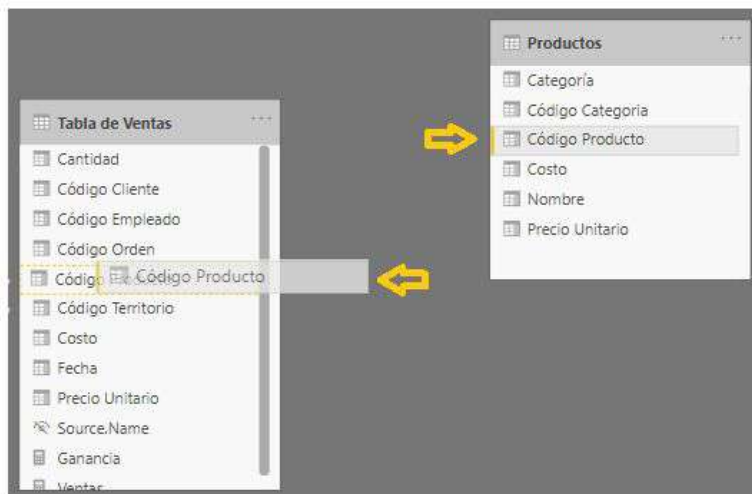
Tabla Clases

¿Qué relación podríamos establecer entre ambas tablas?

ID_Clase	Nombre_Clase	ID_Estudiante
101	Matemáticas	1
102	Historia	2
103	Ciencias	1
104	Educación Física	3

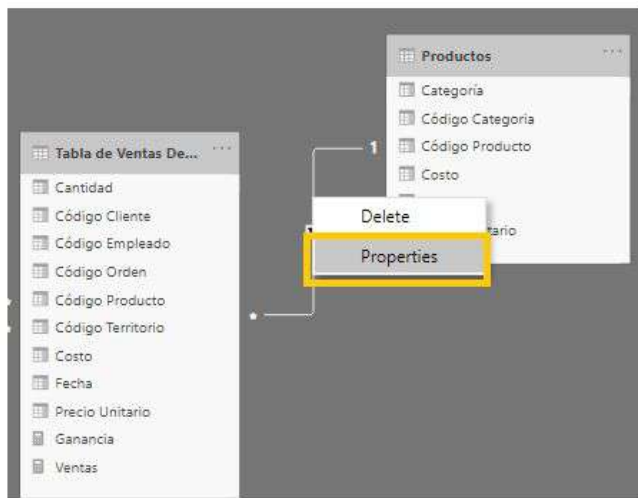


2. Ejemplo de relación - Power BI





2. Ejemplo de relación - Power BI



Edit relationship

Select tables and columns that are related.

Tabla de Ventas Detalle

Código Orden	Fecha	Código Territorio	Código Cliente	Código Producto	Código Empleado	Precio Unitario
1000	1/1/2017	10	3	100	3	
1002	1/3/2017	10	5	100	1	
1003	1/4/2017	10	13	100	4	

Productos

Código Producto	Nombre	Código Categoría	Categoría	Precio Unitario	Costo
100	Product HHYDP	20	Beverages	15	5
103	Product KSBRM	20	Beverages	12	2
101	Product RECZE	21	Condiments	12	4

Cardinality: Many to one (*:1)

Cross filter direction: Single

☒ Make this relationship active

☐ Assume referential integrity

☐ Apply security filter in both directions

OK Cancel



2. Ejemplo de relación - Power BI



Create relationship

Select tables and columns that are related.

Productos

Código Producto	Nombre	Código Categoría	Categoría	Precio Unitario	Costo
100	Product HHYDP	20	Beverages	15	5
103	Product KSBHM	20	Beverages	12	2
101	Product RECZE	21	Condiments	12	4

Tabla de Ventas

Source.Name	Código Orden	Fecha	Código Territorio	Código Cliente	Código Producto	Código Empleado
2017.xlsx	1000	1/1/2017	10	3	100	
2017.xlsx	1002	1/3/2017	10	5	100	
2017.xlsx	1003	1/4/2017	10	13	100	

Cardinality

One to many (1:*)

Cross filter direction

Single

☒ Make this relationship active

☐ Assume referential integrity

☐ Apply security filter in both directions

OK

Cancel



2. Ejemplo de relación - Power BI



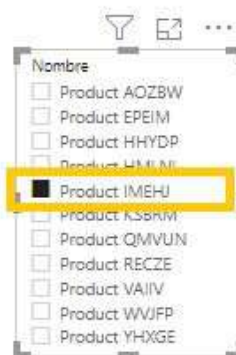
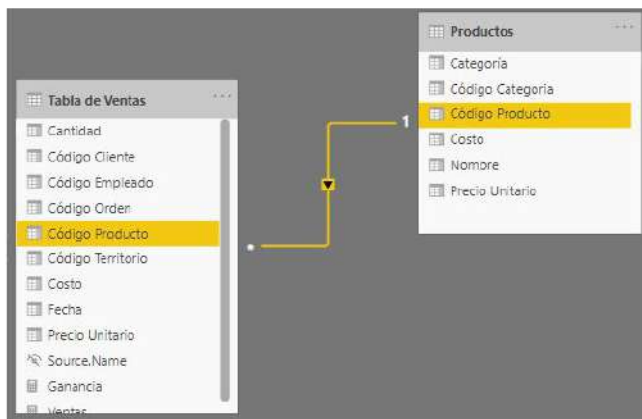
- **Dirección de filtro cruzado:** Las relaciones del modelo se definen con una dirección de filtro cruzado las cuales dependen del tipo de cardinalidad:

Tipo de Cardinalidad	Opciones de filtro cruzado
Uno a Varios	• Único • Ambos
Uno a Uno	• Ambos
Varios a Varios	• Único (Definiendo que tabla filtra al otro) • Ambos



2. Ejemplo de relación - Power BI

- **Único:** Esta dirección es la más común, la cual significa que el filtro viaja en una única dirección hacia la otra tabla.



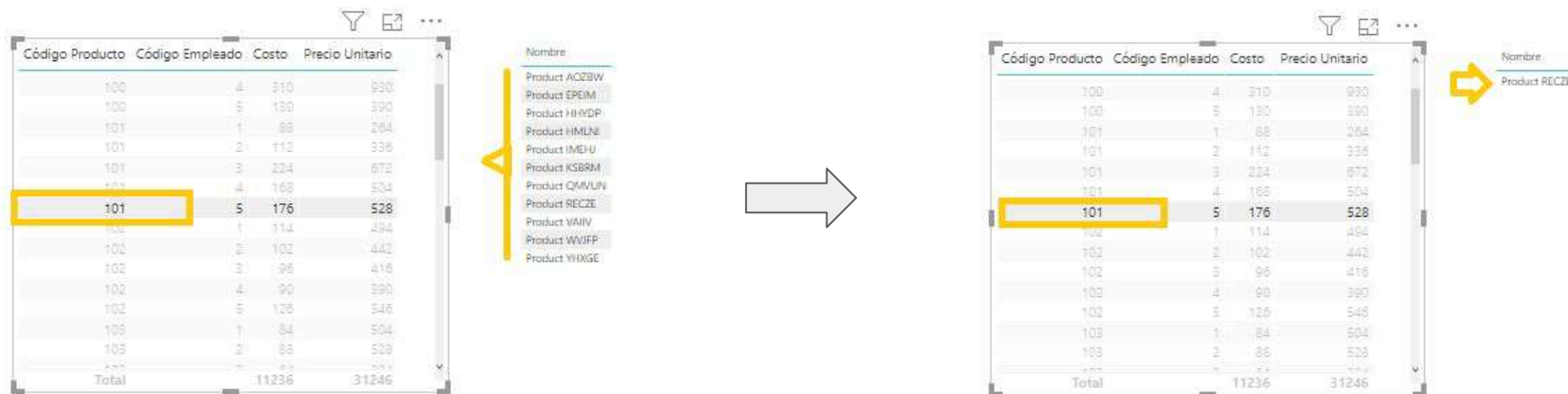
Código Producto	Código Empleado	Costo	Precio Unitario
102	1	114	494
102	2	102	442
102	3	96	416
102	4	90	390
102	5	126	546
Total		528	2288



2. Ejemplo de relación - Power BI

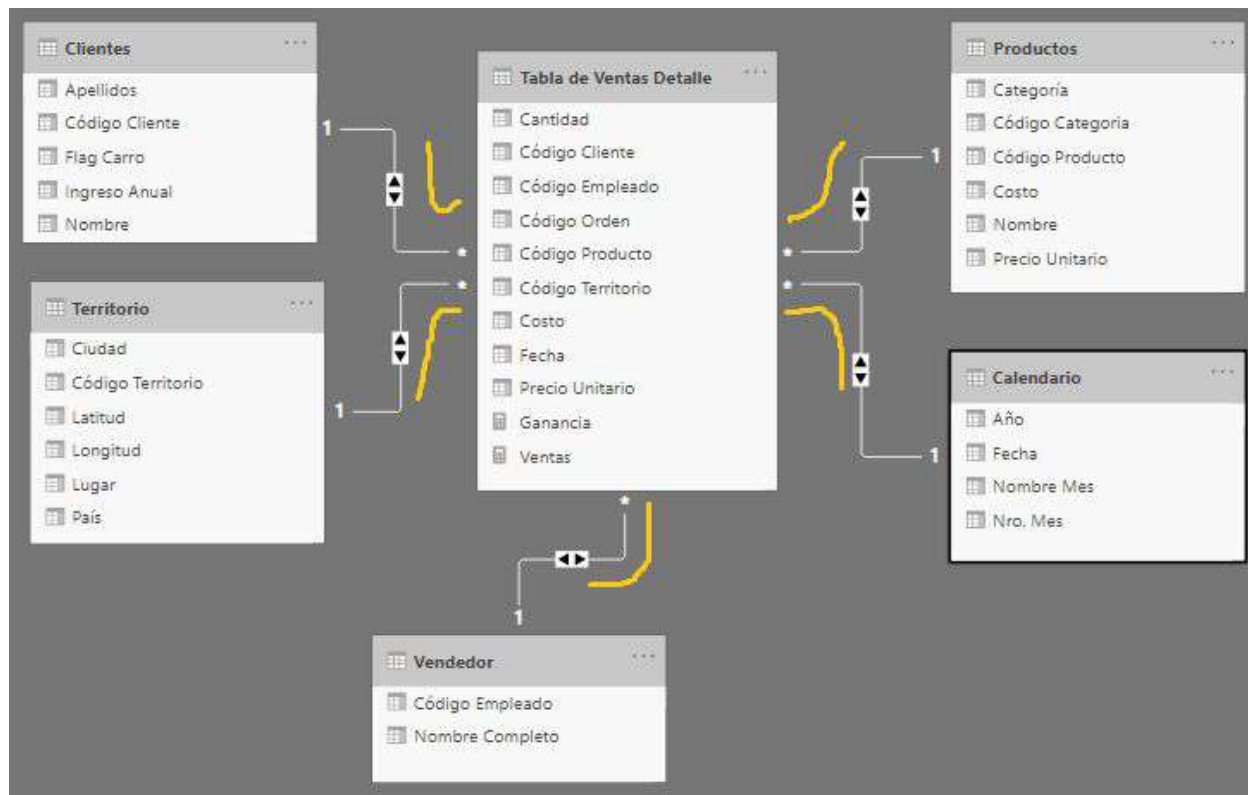


- **Ambos:** Esta dirección implica que la relación entre las tablas viaje en ambos sentidos, es también conocida como “Bidireccional”.





2. Ejemplo de relación - Power BI





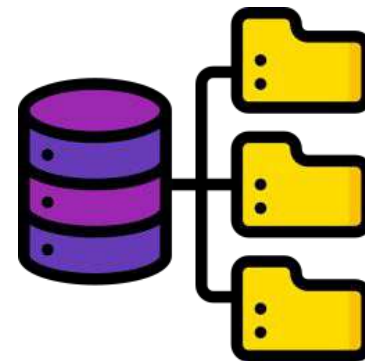
2. Ejercicio práctico - Tablas y Relaciones



Contexto: Una universidad necesita diseñar una base de datos para gestionar la información sobre **estudiantes, cursos, profesores y matrículas**. Los estudiantes deben diseñar las tablas necesarias, definir las llaves primarias y foráneas, y establecer las relaciones entre las tablas.

Instrucciones

1. Leer el enunciado del problema.
2. Diseñar las tablas necesarias.
3. Identifique tablas de hechos y dimensiones.
4. Definir las llaves primarias para cada tabla.
5. Definir las llaves foráneas para establecer relaciones entre las tablas.
6. Dibujar un diagrama simple que muestre las tablas y sus relaciones.





2. Ejercicio práctico - Tablas y Relaciones



Requisitos del Sistema

- Estudiantes:
Cada estudiante tiene un ID único, nombre, y correo electrónico.
- Cursos:
Cada curso tiene un ID único, nombre del curso, y número de créditos.
- Profesores:
Cada profesor tiene un ID único, nombre, y departamento.
- Matrículas:
Los estudiantes se matriculan en varios cursos.
Cada matrícula debe registrar la fecha de inscripción y la calificación obtenida.
Un curso puede tener varios estudiantes matriculados, y un estudiante puede estar matriculado en varios cursos.



2. Ejercicio práctico - Tablas y Relaciones



Ejemplo de Solución

Tabla Estudiantes

Atributo	Tipo
ID_Estudiante	INT
Nombre	VARCHAR
Correo	VARCHAR

Tabla Cursos

Atributo	Tipo
ID_Curso	INT
Nombre_Curso	VARCHAR
Créditos	INT

Tabla Profesores

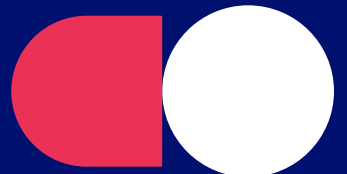
Atributo	Tipo
ID_Profesor	INT
Nombre	VARCHAR
Departamento	VARCHAR

Tabla Matrícula

Atributo	Tipo
ID_Matricula	INT
ID_Estudiante	INT
ID_Curso	INT
Fecha_Inscripcion	DATE
Calificacion	FLOAT



**3. Diferenciación entre OLAP
(Online Analytical Processing) y
OLTP (Online
Transactional Processing).**





3. ¿Qué es OLAP?



OLAP (Online Analytical Processing) es un tipo de sistema de base de datos diseñado para:

- **Análisis de Datos:** Facilita la consulta y análisis de grandes volúmenes de datos históricos.
- **Toma de Decisiones:** Ayuda a los gerentes y analistas a tomar decisiones informadas basadas en datos.





3. OLAP - Características clave



- Consultas Complejas: Realiza cálculos como sumas, promedios y otras agregaciones.
- Datos Históricos: Utiliza datos que han sido recopilados a lo largo del tiempo.
- Usuarios: Principalmente utilizado por analistas de negocio y gerentes que necesitan crear informes detallados y gráficos.
- Ejemplos: Sistemas de inteligencia de negocios (BI) como data warehouses.





3. ¿Qué es OLTP?



OLTP (Online Transactional Processing) es un tipo de sistema de base de datos diseñado para:

- **Gestión de Transacciones:** Maneja las operaciones diarias de una empresa, como ventas, reservas, y registros de clientes.
- **Alta Velocidad y Precisión:** Optimizado para procesar rápidamente un gran número de transacciones simples.

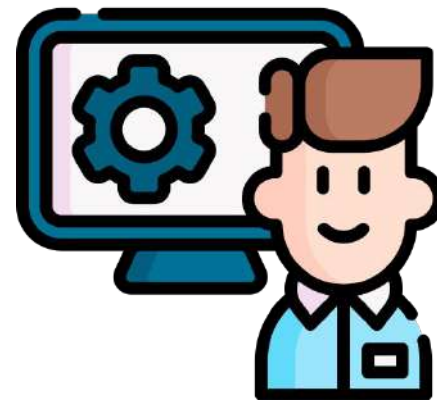




3. OLTP - Características clave



- Consultas Simples: Maneja operaciones rápidas como insertar, actualizar y borrar datos.
- Datos Actuales: Trabaja con datos en tiempo real, como las transacciones del día.
- Usuarios: Utilizado por el personal operativo, como cajeros y empleados que necesitan registrar y acceder a datos rápidamente.
- Ejemplos: Sistemas de punto de venta (POS), sistemas de reservas en línea.





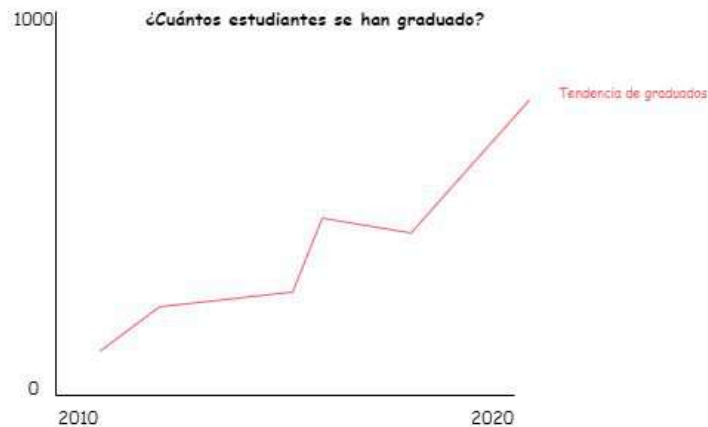
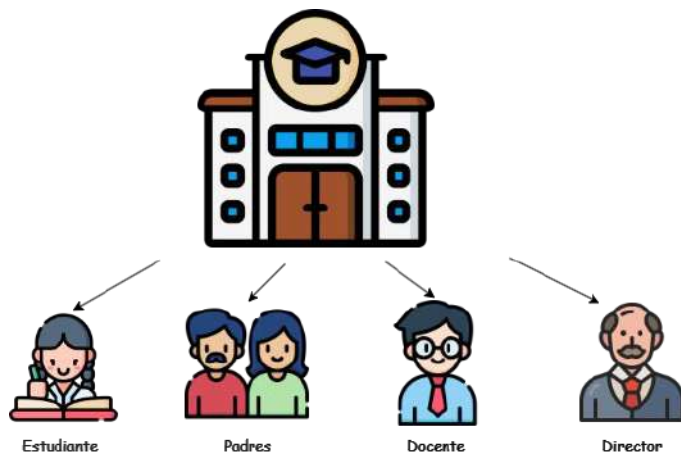
3. OLAP vs. OLTP



OLAP	OLTP
Maneja todos los datos históricos	Maneja datos operacionales recientes
El objetivo es tomar decisiones a partir de grandes fuentes de datos	El objetivo es realizar las operaciones diarias
Usa consultas complejas	Usa consultas simples
Menores velocidades de procesamiento	Mayores velocidades de procesamiento
Solo requiere operaciones de lectura	Requiere operaciones de lectura/escritura



3. Ejemplo OLAP vs. OLTP



Si un estudiante se graduó hace más de 5 años,
¿Nos importa conservar su información?



Almacenes de Datos



3. Ejemplo OLAP vs. OLTP

OLTP (Online Transaccional Processing)



¿Cómo le va a nuestro hijo en la escuela?



Bien! Obtuvo una **A+** en mi clase este semestre

OLAP (Online Analytical Processing)



¿Le irá bien a nuestro hijo en esta escuela?



Bueno, según la tasa de aprobación del **95%** que hemos mantenido en los últimos 5 años, le debería ir bien en esta escuela.

Debido a que el tipo de datos almacenados en una base de datos es diferente a un almacén de datos, tiene sentido que el tipo de información que una base de datos puede proporcionar difiera de la información que puede proporcionar un almacén de datos.

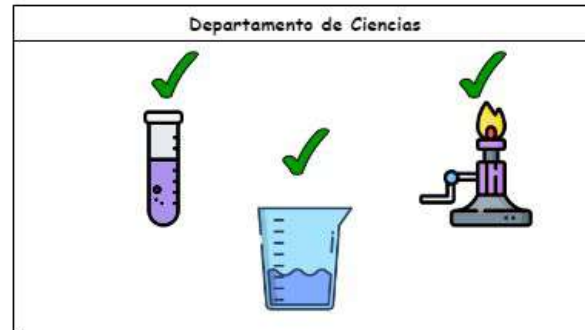


3. Ejemplo OLAP vs. OLTP

Database (OLTP)



Data Warehouse (OLAP)



Digamos que a la escuela le recortaron los fondos y el director de repente tiene que reducir el personal de la escuela. Saben que el departamento de matemáticas ha recibido muchas quejas últimamente, por lo que el director simplemente pregunta a la base de datos de la escuela: **"¿De qué maestro debería deshacerme según la cantidad de quejas?"**

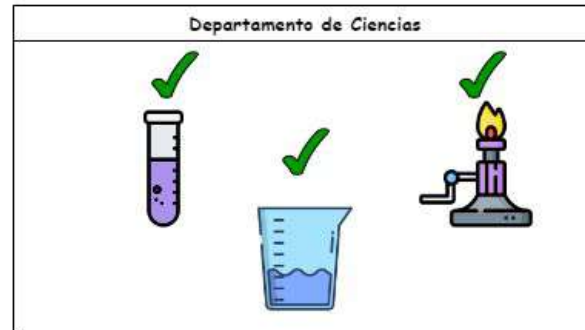


3. Ejemplo OLAP vs. OLTP

Database (OLTP)



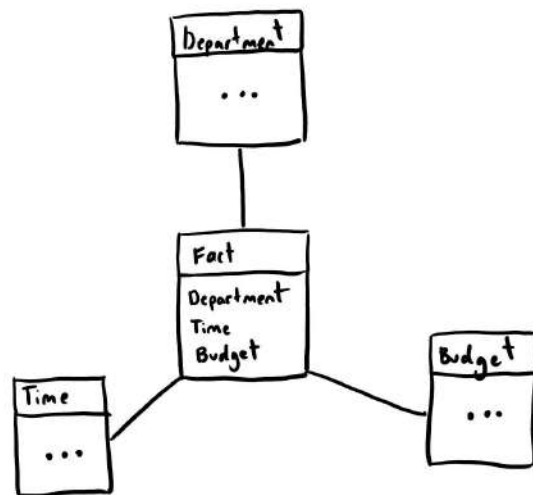
Data Warehouse (OLAP)



○ en un escenario más feliz, la escuela recibe una subvención de financiación, por lo que ahora tienen que elegir en qué departamento invertir. Digamos que el almacén de datos está diseñado para poder responder exactamente a eso.



3. Ejemplo OLAP vs. OLTP

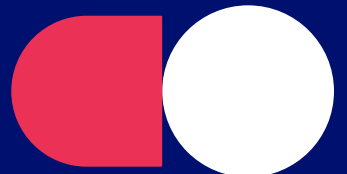


Star Schema

Al agregar los datos de sus clases (departamentos de educación) y departamentos de finanzas en el almacén de datos, el director puede ejecutar consultas OLAP para descubrir que el departamento de ciencias no ha invertido mucho y necesita más inversión.



4. Actividad grupal





Repaso



slideswith.com/PREPASOSII

Leer



Analiza



Medir



Describir





Respuesta



- El sistema de gestión de inscripciones de la universidad registra cada vez que un estudiante se inscribe en un curso. ¿Qué tipo de sistema es este?

R// OLTP

- El departamento de análisis académico de la universidad quiere analizar las calificaciones de los estudiantes de los últimos diez años para identificar tendencias y mejorar la calidad educativa. ¿Qué tipo de sistema es más adecuado para este análisis?

R// OLAP

- El departamento de análisis académico de la universidad quiere analizar las calificaciones de los estudiantes de los últimos diez años para identificar tendencias y mejorar la calidad educativa. ¿Qué tipo de sistema es más adecuado para este análisis?

R// Gestión de transacciones diarias



Respuesta



- El departamento de TI de la universidad utiliza un sistema que organiza los datos de las inscripciones, calificaciones y asistencia para realizar análisis complejos. ¿Qué tipo de sistema están utilizando?

R// OLAP

- En la base de datos de la universidad, cada estudiante tiene un identificador único llamado ID_Estudiente. ¿Qué es ID_Estudiente en este contexto?

R// Una llave primaria

- En la base de datos de la universidad, la tabla Inscripciones contiene el campo ID_Curso que se refiere al ID_Curso en la tabla Cursos. ¿Qué es ID_Curso en la tabla Inscripciones?

R// Una llave foránea



Respuesta



- En una relación uno a muchos (1:M) entre Profesores y Cursos, ¿cuántos cursos puede enseñar un profesor?

R// Varios

- . En la universidad, cada estudiante puede estar inscrito en varios cursos, y cada curso puede tener varios estudiantes inscritos. ¿Qué tipo de relación es esta?

R// Muchos a muchos (M:M)

- En la base de datos de la universidad, los estudiantes pueden tener varias actividades extracurriculares y cada actividad puede tener muchos estudiantes participantes. ¿Cuál sería una posible tabla intermedia en este sistema?

R// Participaciones



Gracias

