



# INSTITUTO POLITECNICO NACIONAL



Unidad Profesional Interdisciplinaria de  
Ingeniería Ciencias Sociales y Administrativas

Lic. Administración Industrial  
**TECNOLOGÍAS DE INFORMACIÓN**

## UNIDAD TEMÁTICA III

Diseño de bases de datos

### **Equipo 3**

Del Río García Ricardo Isaac

Dorantes Prado Dulce Xcaret

Estela Castrejón Ayleen Guadalupe

Estrella Cárdenas Fernanda

Fernández Moreno Karina

## Índice:

Unidad Temática III “Diseño de bases de datos” .....	3
3.1. Bases de datos .....	3
3.1.1. Esquemas de Base de Datos .....	3
3.1.2. Atributos de los Datos .....	4
3.2. Tipos de Bases de Datos.....	5
3.2.1. Jerárquicas.....	7
3.2.2. Transaccionales.....	8
3.2.3. Relacionales.....	8
3.2.4. Multidimensionales .....	9
3.2.5. De red y Orientadas a Objetos .....	10
3.3. Estructuras de las bases de datos .....	10
3.3.1. Tipos de estructuras de bases de datos.....	14
3.4. Procesamiento de bases de datos en línea.....	20
3.4.1. On-Line Analytical Processing.....	27
3.4.2. On-Line Transaction Processing.....	29
3.5. Bases de datos en CLOUD .....	34
3.5.1. Automatizadas .....	36
3.5.2. Gestionadas .....	36
3.5.3. Autónomas .....	36
Material de apoyo:.....	37
Bibliografía:.....	38

## Unidad Temática III “Diseño de bases de datos”

### 3.1. Bases de datos

Una base de datos es un conjunto de archivos similares que están relacionados entre sí.



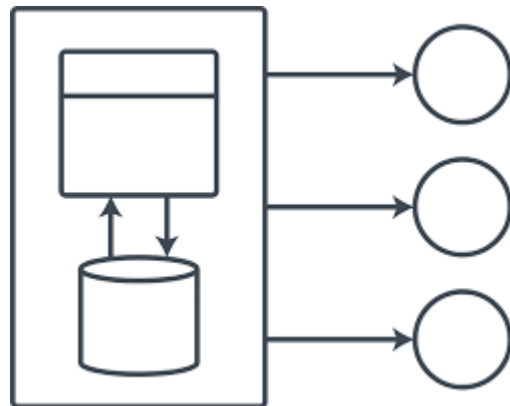
Una base de datos es un conjunto de información organizada y estructurada que se almacena de forma permanente en un sistema informático. En una base de datos, la información se guarda en tablas que están relacionadas entre sí a través de ciertos campos comunes. Las bases de datos son utilizadas para almacenar y gestionar grandes cantidades de datos de manera eficiente, permitiendo realizar consultas, análisis y generación de informes de forma rápida y sencilla.

Una base de datos es una recopilación organizada de información o datos estructurados, que normalmente se almacena de forma electrónica en un sistema informático. Normalmente, una base de datos está controlada por un sistema de gestión de bases de datos (DBMS). En conjunto, los datos y el DBMS, junto con las aplicaciones asociadas a ellos, reciben el nombre de sistema de bases de datos, abreviado normalmente a simplemente base de datos.

Los datos de los tipos más comunes de bases de datos en funcionamiento actualmente se suelen utilizar como estructuras de filas y columnas en una serie de tablas para aumentar la eficacia del procesamiento y la consulta de datos. Así, se puede acceder, gestionar, modificar, actualizar, controlar y organizar fácilmente los datos. La mayoría de las bases de datos utilizan un lenguaje de consulta estructurada (SQL) para escribir y consultar datos (¿Qué es una base de datos?, s/f).

#### 3.1.1. Esquemas de Base de Datos

Un esquema de base de datos representa la configuración lógica de todo o parte de una base de datos relacional. Puede existir de dos formas: como representación visual y como un conjunto de fórmulas conocidas como restricciones de integridad que controlan una base de datos. Estas fórmulas se expresan en un lenguaje de definición de datos, tal como SQL. Como parte de un diccionario de datos, un esquema de base de datos indica cómo las entidades que conforman la base de datos se



relacionan entre sí, incluidas las tablas, las vistas, los procedimientos almacenados y mucho más.

Típicamente, un diseñador de bases de datos crea un esquema de base de datos para ayudar a los programadores cuyo software interactuará con la base. Al proceso de crear un esquema de base de datos se le llama modelado de datos. Al seguir el enfoque de tres esquemas para el diseño de bases de datos, este paso seguiría la creación de un esquema conceptual. Los esquemas conceptuales se enfocan en las necesidades informativas de una organización, más que en la estructura de una base de datos.

### **Hay dos tipos principales de esquemas de bases de datos:**

1. Un esquema lógico de base de datos expresa las restricciones lógicas que se aplican a los datos almacenados. Puede definir las restricciones de integridad, las vistas y las tablas.
2. Un esquema físico de base de datos dispone cómo se almacenan los datos físicamente en un sistema de almacenamiento en términos de archivos e índices.

En el nivel más básico, un esquema de base de datos indica qué tablas o relaciones componen la base de datos, así como los campos incluidos en cada tabla. Por lo tanto, los términos diagrama de esquema y diagrama de relaciones de entidades con frecuencia son intercambiables (¿Qué es un esquema de base de datos?, s/f).

Un esquema de base de datos define cómo se organizan los datos dentro de una base de datos relacional. Esto incluye restricciones lógicas como, por ejemplo, nombres de tablas, campos, tipos de datos y las relaciones entre estas entidades. Los esquemas suelen utilizar representaciones visuales para comunicar la arquitectura de la base de datos, convirtiéndose en la base de la disciplina de gestión de datos de una organización. Este proceso de diseño de esquemas de bases de datos también se conoce como modelado de datos.

Estos modelos de datos cumplen una variedad de funciones, como usuarios de bases de datos, administradores de bases de datos y programadores. Por ejemplo, puede ayudar a los administradores de bases de datos a gestionar los procesos de normalización para evitar la duplicación de datos. Alternativamente, puede permitir que los analistas naveguen por estas estructuras de datos para realizar informes u otros análisis comerciales valiosos. Estos diagramas actúan como documentación valiosa dentro del sistema de gestión de bases de datos (DBMS), asegurando la alineación entre varios stakeholders (¿Qué es un esquema de base de datos?, s/f).

#### **3.1.2. Atributos de los Datos**

Un atributo de los datos es una característica o cualidad que describe una entidad o elemento en un conjunto de datos.

Un atributo es el nombre de una columna de una relación y representan a los atributos del modelo entidad-relación (s/f-b).

Un atributo es un dato que describe una entidad. Por ejemplo, en una base de datos de clientes, los atributos pueden ser nombre, dirección y número de teléfono. En una base de datos de productos, los atributos pueden ser el nombre, el precio y la fecha de fabricación. Cada atributo tiene un tipo de datos específico, como una cadena, un número entero o una fecha.



Los atributos son la columna de entrada para cualquier base de datos. Se utiliza para crear registros y para almacenar información específica. La información almacenada en un atributo puede variar dependiendo de la aplicación. Por ejemplo, una base de datos de inventario para una tienda de ropa tendrá atributos como el tamaño, la marca, el color, el precio y la cantidad.

Los atributos tienen un papel importante en el diseño de una base de datos. Deben ser relevantes para la información que se está almacenando y estar relacionados entre sí. Esto significa que cada atributo debe estar relacionado con la información general de la base de datos. Por ejemplo, en una base de datos de inventario, los atributos deben estar relacionados con los productos que se están almacenando.

Existen varios tipos de atributos. Los atributos simples contienen un solo valor o dato. Los atributos compuestos almacenan varios valores de datos, como los atributos de la dirección postal. Los atributos multi-valuados almacenan múltiples valores para los mismos atributos.

Los atributos son una parte integral de cualquier base de datos. Estos son los bloques de construcción para almacenar la información correcta. Asegúrate de elegir los atributos adecuados para que tu base de datos funcione correctamente. Esto te ayudará a mantener tu base de datos organizada y eficiente (García, 2023).

### 3.2. Tipos de Bases de Datos

Una base de datos (BBDD) es un conjunto de datos organizado de tal forma que permite obtener rápidamente varios tipos de información. Por supuesto, el procesamiento de los datos dependerá de los objetivos que se tengan.

Todas las bases de datos ordenan y clasifican los datos para después poder investigar y analizar los mismos. Hay softwares especiales que trabajan con BBDD

conocidos como sistemas gestores de bases de datos (SGBD). Lógicamente, hay distintos softwares para cada tipo de base de datos.

Gracias a ellos se puede almacenar y acceder rápidamente a los datos estructurados alojados en una BBDD. Así, los SGBD cuentan con características propias en función del tipo de BBDD.

### Existen diversos tipos de bases de datos los cuales son:

- Base de datos relacional
- Base de datos distribuida
- Base de datos orientada a objetos
- Base de datos gráfica
- Base de datos NoSQL

### Base de datos relacional

Las **bases de datos relacionales** son las más comunes en la actualidad. Consisten en una serie de contenedores de información denominados tablas, en los que se almacenan registros. Los registros serían las filas de las tablas mientras que sus campos de éstos serían las columnas.



La característica más importante de las bases de datos relacionales es que los datos de las tablas pueden estar relacionados entre sí por medio de índices que nos permiten pivotar de una tabla a otra u otras.

### Base de datos distribuida

Las **bases de datos distribuidas** son aquellas en las que la información no reside en un único servidor de bases de datos, sino que existen diversos servidores, a menudo en redes distintas, que mantienen una porción de los datos.

Debido a su característica distribuida, las bases de datos están alojadas en diversos nodos y se produce un trasiego de la información continuo entre ellos. Sin embargo, los clientes pueden acceder a los datos y recuperar la información sin necesidad de preocuparse de que ésta se encuentre distribuida entre varias máquinas.

### Base de datos orientada a objetos

Son un modelo de bases de datos en las que la información se almacena mediante objetos. **Los objetos son conjuntos heterogéneos de datos**, tan complejos como sea necesario para modelizar aquella información que se necesite.

Para los que vienen del área de la programación podemos entenderlas como bases de datos que consiguen almacenar registros igual que si se tratase de objetos de la Programación Orientada a Objetos. Al almacenarse la información organizada en objetos resulta muy adecuado para el trabajo directo desde **lenguajes de programación** capaces de usar este paradigma.

### **Base de datos gráfica**

Son un tipo de bases de datos donde **se guardan datos y sus relaciones**. Sin embargo, las relaciones podemos considerarlas en este modelo de base de datos como ciudadanos de primer nivel, ya que se almacenan tal cual en la base de datos y no mediante índices como en las relacionales.

Las bases de datos gráficas permiten realizar recorridos muy rápidos entre los distintos nodos, ya que las relaciones apuntan directamente a los nodos a los que se puede llegar desde otro nodo y **no requieren cálculos y aplicación de uniones**. Este tipo de bases de datos es frecuente en motores sociales o en motores de recomendación.

### **Base de datos NoSQL**

Las bases de datos NoSQL (Not only SQL) son aquellas en las que **no se usan tablas sino colecciones de elementos**. Los elementos almacenados en las colecciones pueden ser heterogéneos, de modo que en una colección podemos almacenar registros con juegos de datos distintos entre sí. Además, este tipo de base de datos no suele usar SQL para realizar consultas, sino programación funcional para hacer filtrados y otros tipos de operaciones.

Las NoSQL son otro modelo de bases de datos muy popular, sin llegar al grado de las relacionales, pero muy frecuente en la actualidad en muchos tipos de aplicaciones donde se requiere variabilidad entre la información almacenada en las colecciones, así como velocidad de recuperación de la información (de Zúñiga, 2023).

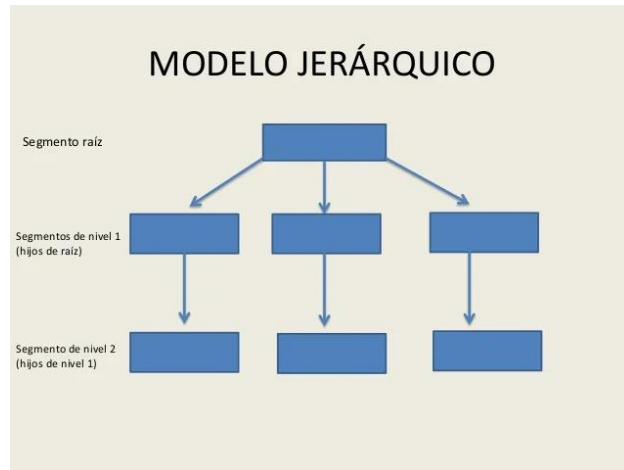
#### **3.2.1. Jerárquicas**

Las BBDD jerárquicas **almacenan la información en una estructura jerárquica** o con un orden de importancia. Así, los datos se organizan en una figura parecida a un árbol invertido con segmentos conocidos como nodos y ramas, que contienen información de interés. Estos pueden ser de tres categorías:

- **Padre:** nodo de que se desprenden descendientes. Todos los padres se sitúan al mismo nivel y tienen la misma importancia.
- **Hijo:** nodo que depende de un padre, es decir, se deriva del anterior.
- **Raíz:** origen de los datos, al no tener padre. Se ubica en el nivel superior del árbol y todos los nodos salen de ella.

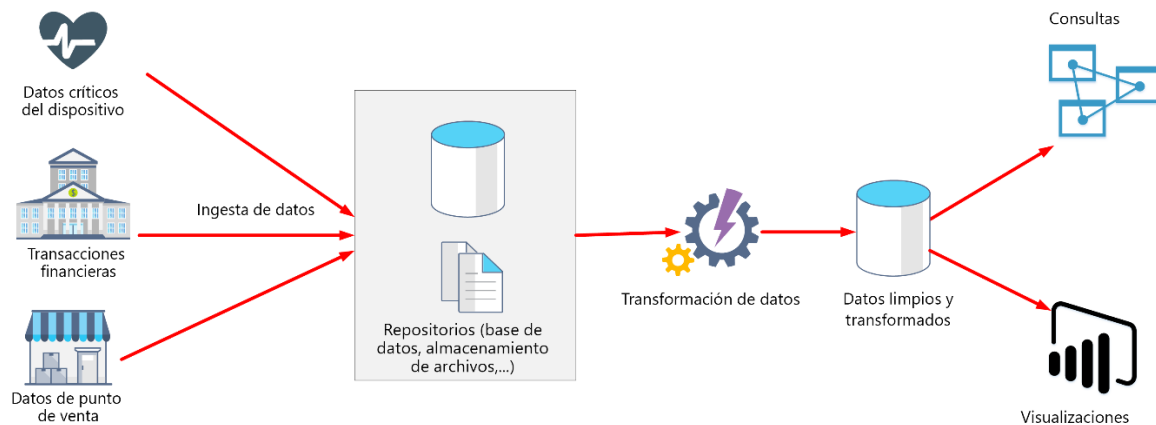
Entre las **principales características** de uno de los tipos de bases de datos más usadas se encuentran las siguientes:

- Se puede compartir el acceso y la información con varios usuarios.
- Los datos son independientes.
- Es complicado modificarla, puesto que es una estructura rígida.
- Se necesita gran conocimiento de las unidades de información.
- Los nodos lejanos de la raíz son de difícil acceso, por lo que se requiere tiempo.



### 3.2.2. Transaccionales

Las BBDD transaccionales son las encargadas de enviar y recibir datos a gran velocidad. Es raro que los usuarios “normales” las utilicen, puesto que están dirigidas a ciertos sectores como los sistemas bancarios, en los que por ejemplo se registran operaciones inmediatas entre cuentas con los correspondientes datos de dichas operaciones.



### 3.2.3. Relacionales

Las BBDD relacionales son, en la actualidad, **uno de los tipos de bases de datos más utilizados**. El lenguaje predominante en ellas es el Structured Query Language, más conocido como SQL.

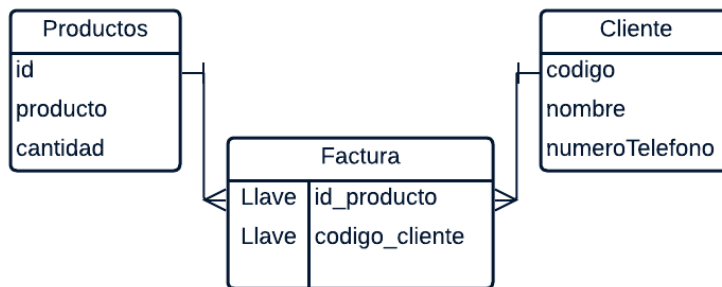
Los datos **se almacenan en registros organizados en tablas**, por lo que pueden asociar y cruzar los elementos con facilidad. Es una base de datos recomendable



si los datos tienen un margen de error nulo y no necesitan modificaciones continuas.

**Sus principales características son:**

- Pueden ser utilizadas por cualquier usuario.
- Su gestión es fácil.
- Se puede acceder rápidamente a los datos.
- Garantiza la total consistencia de los datos, sin posibilidad de error.



Uno de los servicios de base de datos relacional, inteligente y escalable creado para la nube que ofrecemos en Inteliquia es **Azure SQL Database**.

Sus **características basadas en inteligencia artificial**

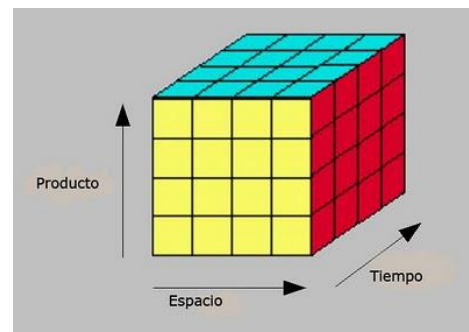
**automatizadas** optimizan el rendimiento y la durabilidad.

Además, sus opciones de proceso sin servidor y almacenamiento en el nivel Hiperescala modifican de forma automática la escala de los recursos a petición, lo que permite centrarse en la **creación de nuevas aplicaciones sin tener que preocuparse por el tamaño del almacenamiento** ni la administración de los recursos.

### 3.2.4. Multidimensionales

Otro de los tipos de bases de datos más usados. Las BBDD multidimensionales se utilizan para funciones concretas. Lo que las separa de las bases de datos relacionales solo se aprecia a nivel conceptual, puesto que en las multidimensionales **los campos o atributos de una tabla pueden ser de dos tipos:**

- Pueden representar dimensiones dentro de una tabla de datos.
- Pueden representar las métricas que se pretenden obtener.



Algunas de sus **principales características** son:

- No emplean ninguna jerarquía.
- Facilitan la búsqueda y la modificación posterior.

- Usan un espacio menor de almacenamiento.
- Tienen acceso a grandes cantidades de información.

### **3.2.5. De red y Orientadas a Objetos**

Las BBDD de red son una variación de la anterior. Su principal diferencia radica en la composición del nodo, ya que en este modelo pueden tener varios padres. Entre sus principales contras es que es difícil modificarlas y adaptarlas al tener una estructura compleja.

Las BBDD orientadas a objetos son de las más modernas, especialmente por su gran capacidad y potencia. Una de sus principales características es que en ellas no se guarda información detallada sobre el objeto, ya que se almacena por completo al mismo.

Cada objeto tiene características propias que le permiten diferenciarse de otros similares. Sus ventajas son claras:

- Admiten más cantidad de contenido.
- Permiten que el usuario tenga más información de primera mano.

(Tipos de base de datos, s/f).

## **3.3. Estructuras de las bases de datos**

Tablas, registros, campos son las tres estructuras principales de una base de datos.

Recordando que una base de datos es una estructura creada para guardar toda la información de una parte de la organización.

### **Tablas**

Dentro de cada base de datos se crean primeramente las tablas las tablas son otras estructuras siempre dentro de base de datos nunca fuera de ellas que guardan un subconjunto del todo.

### **Registros y columnas**

Dentro de cada tabla todos los registros tienen el mismo formato es decir tienen los mismos campos. Hasta ahora, la estructura de la base de datos, tablas y campos las podremos definir, en una base de datos real, por medio de comandos DDL porque estamos hablando de la estructura que tendrá la base de datos, independientemente de que, si guardará muchos o pocos datos, todos tendrán la misma estructura (DBA, 2017).

Dentro de una base de datos, los datos relacionados se agrupan en tablas, cada una de ellas consiste en filas (también llamadas "tuplas") y columnas, como una hoja de cálculo.

Cada fila de una tabla se llama "registro". Los registros incluyen datos sobre algo o alguien, como un cliente específico. En cambio, las columnas (también conocidas como "campos" o "atributos") contienen un único tipo de información que aparece en cada registro, como las direcciones de todos los clientes enumerados en la tabla.

Nombre	Apellido	Edad	Código postal
Roger	Williams	43	34760
Jerrica	Jorgensen	32	97453
Samantha	Hopkins	56	64829

Con el fin de que los datos sean consistentes de un registro al siguiente, asigna el tipo de datos apropiado a cada columna. Los tipos de datos comunes incluyen:

CHAR - una longitud específica de texto.

VARCHAR - texto de longitudes variables.

TEXT - grandes cantidades de texto.

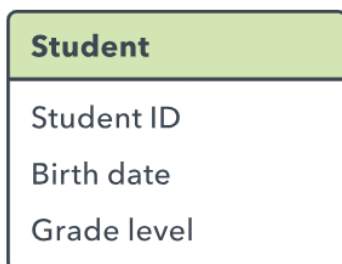
INT - número entero positivo o negativo.

FLOAT, DOUBLE - también puede almacenar números de punto flotante.

BLOB - datos binarios.

Algunos sistemas de gestión de bases de datos también ofrecen el tipo de datos denominado "Autonumeración", que genera automáticamente un número único en cada fila.

A los efectos de crear una visión general de la base de datos, conocida como un diagrama entidad-relación, no incluiremos las tablas reales, sino que cada tabla se convertirá en un recuadro del diagrama. El título de cada recuadro debería indicar que describen los datos en la tabla, mientras que los atributos están enumerados a continuación, del siguiente modo:



Por último, se debe decidir qué atributo o atributos funcionarán como clave primaria para cada tabla, si procede. Una clave primaria (PK) es un identificador único para una entidad determinada, esto significa que puedes seleccionar un cliente concreto incluso si solo conoces ese valor.

Los atributos seleccionados como claves primarias deben ser únicos, inalterables y estar siempre presentes (nunca NULL o vacíos). Por este motivo, los números de pedido y los nombres de usuario son excelentes claves primarias, mientras que los números de teléfono o direcciones postales no lo son. También puedes usar múltiples campos conjuntamente como la clave primaria (esto se denomina "clave compuesta").

Cuando llegue el momento de crear la base de datos real, ubicarás la estructura de datos lógicos y la estructura de datos físicos en el lenguaje de definición de datos admitido por nuestro sistema de gestión de base de datos. En este punto, también deberías calcular el tamaño aproximado de la base de datos para asegurarte de tener el nivel de rendimiento y el espacio de almacenamiento necesarios (Tutorial de estructura y diseño de bases de datos, s/f).

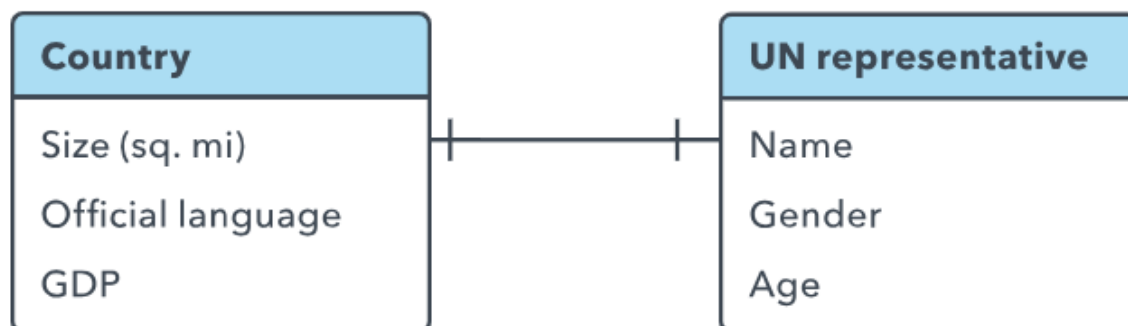
### Creación de relaciones entre entidades

Cuando tus tablas de base de datos se conviertan en tablas, estarás listo para analizar las relaciones entre esas tablas. La cardinalidad se refiere a la cantidad de elementos que interactúan entre dos tablas relacionadas. Identificar la cardinalidad te ayuda a asegurarte de que has dividido los datos en tablas de la forma más eficiente.

Cada entidad puede, potencialmente, tener una relación con todas las demás, pero por lo general esas relaciones pueden ser de uno de tres tipos:

#### Relaciones uno a uno

Si hay una única instancia de la Entidad A para cada instancia de la Entidad B, se dice que tienen una relación de uno a uno (a menudo se escribe 1:1). Puedes indicar este tipo de relación en un diagrama ER mediante una línea con un guión en cada extremo:



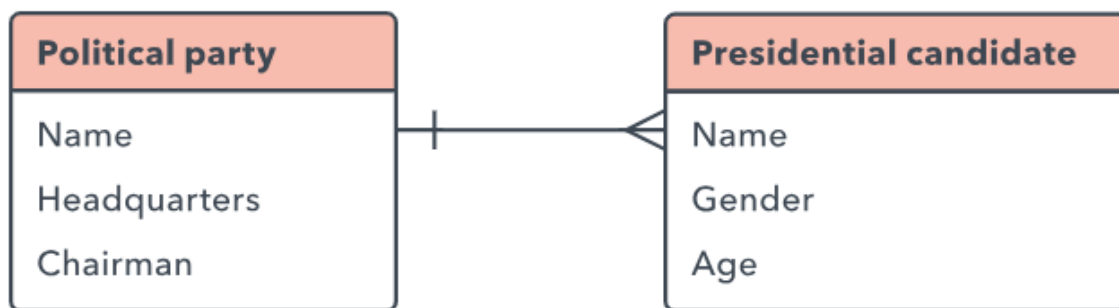
A menos que tengas un buen motivo para no hacerlo, una relación 1:1 generalmente indica que la mejor opción sería combinar los datos de las dos tablas en una sola tabla.

Sin embargo, quizás desees crear tablas con una relación de uno a uno en una serie particular de circunstancias. Si tienes un campo con datos opcionales, como "descripción", que está en blanco para muchos registros, puedes mover todas las descripciones a su propia tabla, eliminando espacio vacío y mejorando el rendimiento de la base de datos.

Para garantizar que los datos coincidan correctamente, luego tendrías que incluir al menos una columna idéntica en cada tabla, lo más probable es que sea la clave primaria.

### Relaciones uno a muchos

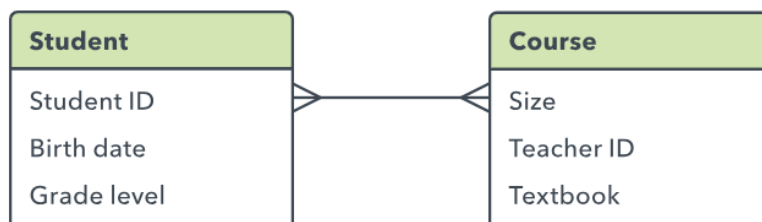
Estas relaciones suceden cuando un registro de una tabla está asociado a múltiples entradas en otra tabla. Por ejemplo, un solo cliente puede haber solicitado múltiples pedidos o una persona haberse llevado muchos libros de la biblioteca a la vez. Las relaciones uno a muchos (1:M) se indican con lo que se denomina "notación patas de gallo" como en el siguiente ejemplo:



Para implementar una relación uno a muchos (1:M) mientras preparas una base de datos, simplemente agrega la clave primaria de "un" lado de la relación como un atributo en la otra tabla. Cuando una clave primaria se detalla en otra tabla de esta manera, se denomina "clave extranjera". La tabla en el lado "1" de la relación es considerada una tabla principal respecto de la tabla secundaria que se encuentra del otro lado.

### Relaciones muchos a muchos

Cuando múltiples entidades de una tabla se pueden asociar a múltiples entidades de otra tabla, se dice que tienen una relación de muchos a muchos (M:N). Esto puede suceder en el caso de estudiantes y clases, ya que un estudiante puede inscribirse en muchas clases, y una clase puede tener numerosos estudiantes.

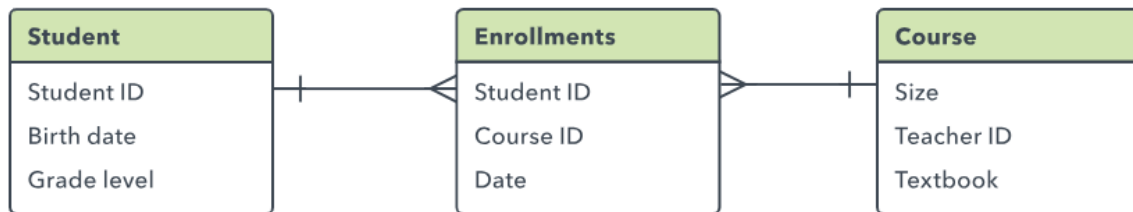


En un diagrama ER, estas relaciones se representan con estas líneas:

Lamentablemente, no es posible implementar directamente este tipo de relación en una base de datos. En cambio, debes dividirlo en dos relaciones uno a muchos.

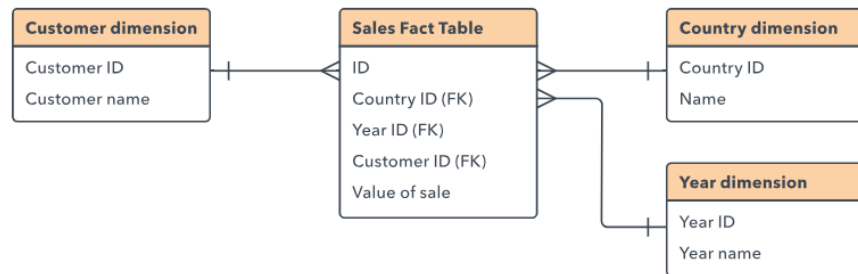
Para ello, debes crear una nueva entidad entre esas dos tablas. Si la relación M:N existe entre ventas y productos, quizás llames a esa nueva entidad "productos vendidos", ya que mostraría los contenidos de cada venta. Tanto las tablas de ventas como de productos tendrían una relación 1:M con "productos vendidos". Esta clase de entidad intermedia se llama "tabla de enlaces", "entidad asociativa" o "tabla de unión" en diversos modelos.

Cada registro de la tabla de enlaces se correspondería con dos de las entidades de las tablas contiguas (también puede incluir información adicional). Por ejemplo, una tabla de enlaces entre estudiantes y clases podría verse así:



## Datos multidimensionales

Algunos usuarios quizás deseen acceder a múltiples dimensiones de un único tipo de dato, especialmente en las bases de datos OLAP. Por ejemplo, quizás deseen conocer las ventas por cliente, estado y mes. En esta situación, lo mejor es crear una tabla de datos central que actúe de referencia para las otras tablas de cliente, estado y mes, de este modo:



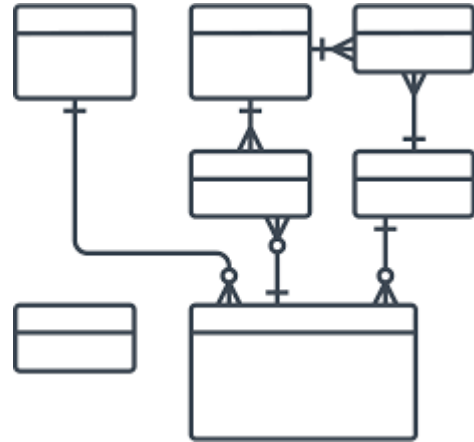
(*Tutorial de estructura y diseño de bases de datos, s/f*).

### 3.3.1. Tipos de estructuras de bases de datos

Hay muchos tipos de modelos de bases de datos. Algunos de los más comunes incluyen:

- Modelo de base de datos jerárquico
- Modelo relacional

- Modelo de red
- Modelo de base de datos orientado a objetos
- Modelo entidad-relación
- Modelo de documentos
- Modelo entidad-atributo-valor
- Esquema de estrella
- Modelo relacional de objetos, que combina los dos que forman su nombre

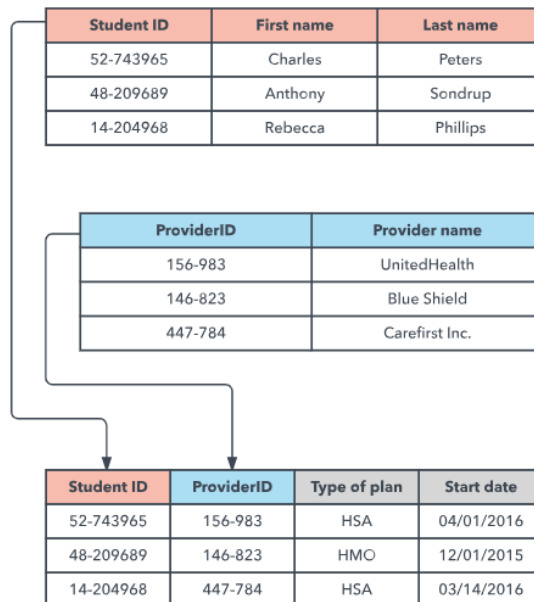


## Modelo relacional

Siendo el modelo más común, el modelo relacional ordena los datos en tablas, también conocidas como relaciones, cada una de las cuales se compone de columnas y filas. Cada columna enumera un atributo de la entidad en cuestión, por ejemplo, precio, código postal o fecha de nacimiento. En conjunto, a los atributos en una relación se los llama dominio. Se elige un atributo particular o combinación de atributos como clave primaria, a la cual se puede hacer referencia en otras tablas, en donde se la denomina clave externa.

Cada fila, también denominada tupla, incluye datos sobre una instancia específica de la entidad en cuestión, por ejemplo, un empleado específico.

El modelo también representa los tipos de relaciones entre esas tablas, incluidas las relaciones uno a uno, uno a muchos y muchos a muchos. Este es un ejemplo:

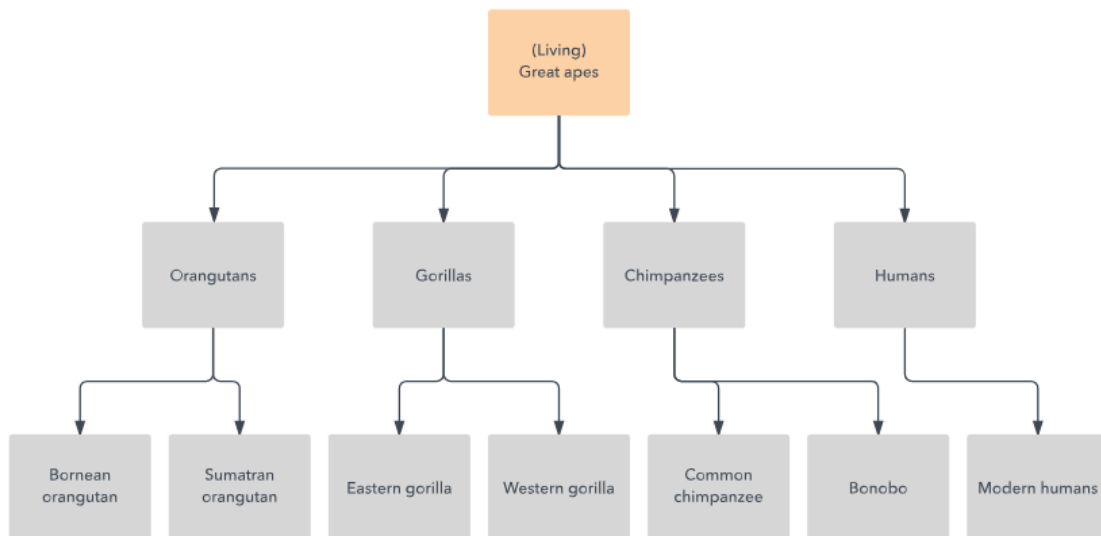


Dentro de la base de datos, las tablas se pueden normalizar, es decir, hacer que cumplan las reglas de normalización que hacen a la base de datos flexible, adaptable y escalable. Al estar normalizada, cada porción de los datos es atómica, es decir, está dividida en partes útiles lo más pequeñas posibles.

Las bases relacionales generalmente se escriben en Lenguaje de Consulta Estructurado (SQL, por sus siglas en inglés). El modelo fue introducido por E. F. Codd en 1970.

## Modelo jerárquico

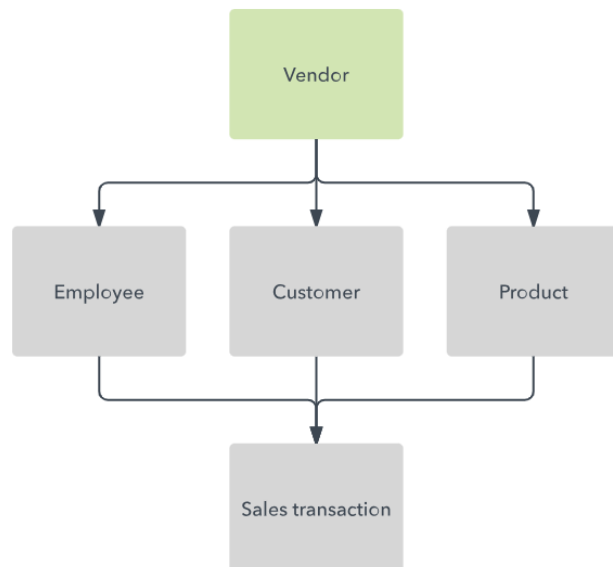
El modelo jerárquico organiza los datos en una estructura de árbol, en la que cada registro tiene un único elemento o raíz. Los registros del mismo nivel se clasifican en un orden específico. Ese orden se usa a manera de orden físico para almacenar la base de datos. El modelo es bueno para describir muchas relaciones del mundo real.



Este modelo fue usado principalmente por los Sistemas de Gestión de Información de IBM en las décadas de 1960 y 1970, pero raramente se los ve en la actualidad debido a ciertas ineficiencias operativas.

## Modelo de red

El modelo de red se basa en el modelo jerárquico, permitiendo relaciones de muchos a muchos entre registros vinculados, lo que implica registros principales múltiples. Basado en la teoría matemática de conjuntos, el modelo se construye con conjuntos de registros relacionados. Cada conjunto consiste en un registro propietario o principal y uno o más registros miembros o secundarios. Un registro puede ser miembro o secundario en múltiples conjuntos, permitiendo que este modelo represente relaciones complejas.





Fue muy popular en la década de 1970 después de que fue definido formalmente por la Conference on Data Systems Languages (CODASYL).

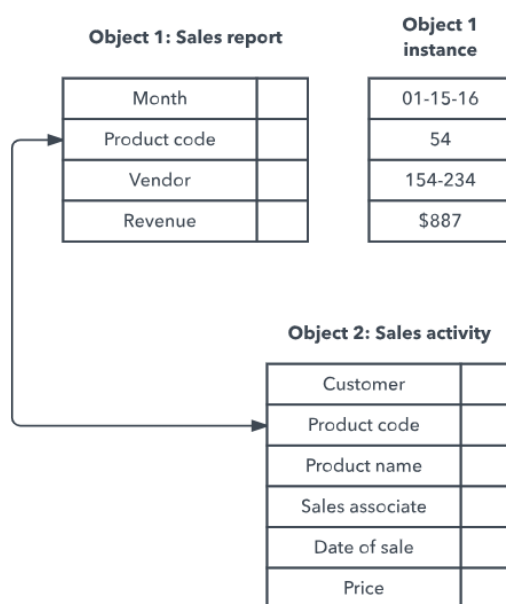
### Modelo de base de datos orientado a objetos

Este modelo define una base de datos como una colección de objetos, o elementos de software reutilizables, con funciones y métodos relacionados. Hay varios tipos de bases de datos orientadas a objetos:

Una base de datos multimedia incorpora elementos multimedia, tales como imágenes, que no se podrían almacenar en una base de datos relacional.

Una base de datos de hipertexto permite que cualquier objeto se conecte a cualquier otro objeto. Es útil para organizar lotes de datos disímiles, pero no es ideal para análisis numérico.

El modelo de base de datos orientado a objetos es el mejor modelo conocido de base de datos posrelacional ya que incorpora tablas, pero no se limita a ellas. A dichos modelos también se los conoce como modelos de base de datos híbridos.



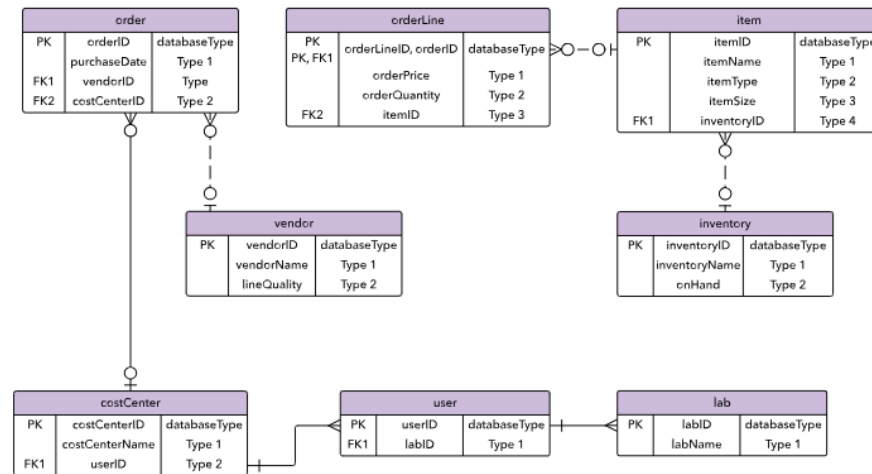
### Modelo relacional de objetos

Este modelo de base de datos híbrido combina la sencillez del modelo relacional con parte de la funcionalidad avanzada del modelo de base de datos orientado a objetos. En esencia, permite a los diseñadores incorporar objetos en una estructura familiar de tablas.

Entre los idiomas e interfaces de llamada se incluyen SQL3, lenguajes de proveedores, ODBC, JDBC e interfaces de llamada patentadas que son extensiones de lenguajes e interfaces usadas por el modelo relacional.

## Modelo entidad-relación

Este modelo capta las relaciones entre entidades del mundo real de forma muy similar al modelo de red, pero no está directamente ligado a una estructura física de la base de datos. En cambio, se lo usa para diseñar una base de datos conceptualmente.



Aquí, a las personas, lugares y cosas, acerca de las cuales se almacenan puntos de datos, se las denomina entidades, cada una de las cuales tiene ciertos atributos que en conjunto forman su dominio. La cardinalidad, o relaciones entre entidades, también se representa en diagramas.

Una forma común del diagrama entidad-relación ER es el esquema de estrella, en el cual una tabla central de información se conecta a múltiples tablas dimensionales.

## Otros modelos de bases de datos

Otros diversos modelos de bases de datos se han usado y siguen usándose en la actualidad.

### Modelo de archivo invertido

Una base de datos construida con estructura de archivo invertido se diseña para facilitar búsquedas rápidas de texto completo. En este modelo, el contenido de datos se indexa como una serie de claves en una tabla de búsqueda, donde los valores indican la ubicación de los archivos asociados. Esta estructura puede proporcionar un informe casi instantáneo de, por ejemplo, datos masivos y análisis estadístico.

Este modelo ha sido usado por el sistema de gestión de bases de datos ADABAS de Software AG desde 1970, y se sigue utilizando en la actualidad.

### Modelo plano

El modelo plano es el modelo de datos más antiguo y sencillo. Simplemente presenta todos los datos en una sola tabla, que consiste en columnas y filas. Para acceder a los datos o manipularlos, la computadora tiene que leer todo el archivo

plano y traerlo a la memoria, lo que vuelve a este modelo ineficiente para todo, excepto para los conjuntos de datos más pequeños.

### **Modelo multidimensional**

Esta es una variación del modelo relacional diseñado para facilitar un mejor procesamiento analítico. Aunque el modelo relacional está optimizado para procesamiento de transacciones en línea (OLTP), este modelo está diseñado para procesamiento analítico en línea (OLAP).

Cada celda de una base de datos dimensional contiene datos acerca de las dimensiones registradas por la base de datos. Visualmente, es como una colección de cubos, en lugar de tablas bidimensionales.

### **Modelo semiestructurado**

En este modelo, los datos estructurales comúnmente contenidos en el esquema de la base de datos están incrustados con los datos mismos. Aquí la diferencia entre los datos y el esquema es imprecisa en el mejor de los casos. Este modelo es útil para describir sistemas, tales como ciertas fuentes de datos basadas en web, las cuales tratamos como bases de datos, pero que no podemos limitar con un esquema. También es útil para describir interacciones entre bases de datos que no se apegan al mismo esquema.

### **Modelo de contexto**

Este modelo puede incorporar elementos de otros modelos de bases de datos, según sea necesario. Improvisa elementos de modelos orientados a objetos, semiestructurados y de red.

### **Modelo asociativo**

Este modelo divide todos los puntos de datos en base a si describen una entidad o una asociación. En este modelo, una entidad es todo lo que existe de manera independiente, mientras que una asociación es algo que solo existe en relación con algo más.

El modelo asociativo estructura los datos en dos grupos:

- Un grupo de elementos, cada uno con un identificador único, un nombre y un tipo.
- Un grupo de enlaces, cada uno con un identificador único y los identificadores únicos de una fuente, verbo u objetivo. La información almacenada tiene que ver con la fuente y cada uno de los tres identificadores pueden hacer referencia ya sea a un enlace o a un elemento.

Otros modelos menos comunes de bases de datos incluyen:

- Modelo semántico, el cual incluye información acerca de cómo los datos almacenados se relacionan con el mundo real.
- Base de datos XML, la cual permite que los datos sean especificados e incluso almacenados en formato XML.
- Gráfico etiquetado.
- Triplestore.

## **Modelos de bases de datos NoSQL**

Además del modelo de base de datos de objetos, otros modelos distintos al tipo SQL han surgido en contraste con el modelo relacional:

El modelo de base de datos gráfico, el cual es aún más flexible que el modelo de red, permitiendo que cualquier nodo se conecte a cualquier otro.

El modelo multivalor, el cual se desprende del modelo relacional, permitiendo que los atributos contengan una lista de datos en lugar de un solo punto de datos.

El modelo de documentos, el cual se diseña para almacenar y administrar documentos o datos semiestructurados, en lugar de datos atómicos.

## **Bases de datos en la web**

La mayoría de los sitios web dependen de cierto tipo de base de datos para organizar y presentar datos a los usuarios. Cada vez que alguien usa las funciones de búsqueda en estos sitios, sus términos de búsqueda son convertidos en consultas para que un servidor de bases de datos los procese. Generalmente, programas intermedios conectan el servidor web con la base de datos.

La amplia presencia de las bases de datos les permite ser usadas casi en cualquier campo, desde compras en línea hasta el enfoque detallado en un segmento de votantes como parte de una campaña política. Diversas industrias han desarrollado sus propias normas para el diseño de bases de datos, desde el transporte aéreo hasta la manufactura de vehículos (*Qué es un modelo de base de datos, s/f*).

### **3.4. Procesamiento de bases de datos en línea**

El procesamiento analítico en línea (OLAP) es una tecnología de software que puede utilizar para analizar datos empresariales desde diferentes puntos de vista. Las organizaciones recopilan y almacenan datos de múltiples fuentes de datos, como sitios web, aplicaciones, medidores inteligentes y sistemas internos. OLAP combina y agrupa estos datos en categorías para proporcionar información procesable para la planificación estratégica. Por ejemplo, un minorista almacena datos sobre todos los productos que vende, como el color, el tamaño, el costo y la ubicación. El minorista también recopila datos de compra de los clientes, como el nombre de los artículos pedidos y el valor total de las ventas, en un sistema diferente. OLAP combina los conjuntos de datos para responder preguntas como

qué productos de color son más populares o cómo la colocación de productos afecta a las ventas.



El procesamiento analítico en línea (OLAP) ayuda a las organizaciones a procesar y beneficiarse de una cantidad cada vez mayor de

información digital. Algunos de los beneficios de OLAP son los siguientes.

### **Toma de decisiones más rápida**

Las empresas utilizan OLAP para tomar decisiones rápidas y precisas a fin de mantenerse competitivas en una economía acelerada. Hacer consultas analíticas en varias bases de datos relacionales lleva mucho tiempo porque el sistema de computación busca en varias tablas de datos. Por otro lado, los sistemas OLAP calculan previamente e integran los datos para que los analistas empresariales puedan generar informes más rápido y cuando sea necesario.

### **Asistencia al usuario sin conocimientos técnico**

Los sistemas OLAP facilitan el análisis de datos complejos para los usuarios empresariales sin conocimientos técnicos. Los usuarios empresariales pueden crear cálculos analíticos complejos y generar informes en lugar de aprender a operar las bases de datos.

### **Vista de datos integrada**

OLAP proporciona una plataforma unificada para las unidades empresariales de marketing, finanzas, producción y otras. Los administradores y tomadores de decisiones pueden ver el panorama general y resolver los problemas de manera efectiva. Pueden llevar a cabo análisis hipotéticos, que muestran el impacto de las decisiones tomadas por un departamento en otras áreas de la empresa.

Los sistemas de procesamiento analítico en línea (OLAP) almacenan datos multidimensionales al representar la información en más de dos dimensiones o categorías. Los datos bidimensionales incluyen columnas y filas, pero los multidimensionales tienen varias características. Por ejemplo, los datos multidimensionales para la venta de productos pueden tener las siguientes dimensiones:

- Tipo de producto
- Ubicación
- Duración

Los ingenieros de datos crean un sistema OLAP multidimensional que consta de los siguientes elementos.

### **Almacenamiento de datos**

Un almacenamiento de datos recopila información desde distintos orígenes, incluidas las aplicaciones, los archivos y las bases de datos. Procesa la información mediante varias herramientas para que los datos estén listos con fines analíticos. Por ejemplo, el almacenamiento de datos puede recopilar información de una base de datos relacional que almacena datos en tablas de filas y columnas.

### **Herramientas ETL**

Las herramientas de extracción, transformación y carga (ETL) son procesos de bases de datos que recuperan, modifican y preparan automáticamente los datos en un formato adecuado con fines analíticos. Los almacenamientos de datos utilizan los procesos ETL para convertir y estandarizar la información procedente de diferentes orígenes antes de ponerla a disposición de las herramientas OLAP.

### **Servidor OLAP**

Un servidor OLAP es la máquina subyacente que alimenta el sistema OLAP. Utiliza herramientas ETL para transformar la información en las bases de datos relacionales y prepararlas para las operaciones de OLAP.

### **Base de datos OLAP**

Una base de datos OLAP es una base de datos independiente que se conecta al almacenamiento de datos. Los ingenieros de datos a veces utilizan una base de datos OLAP para evitar que el almacenamiento de datos se vea sobrecargado por el análisis OLAP. También utilizan una base de datos OLAP para facilitar la creación de modelos de datos OLAP.

### **Cubos OLAP**

Un cubo de datos es un modelo que representa una matriz multidimensional de información. Si bien es más fácil visualizarlo como un modelo de datos tridimensional, la mayoría de los cubos de datos tienen más de tres dimensiones. Un cubo OLAP, o hipercubo, es el término para los cubos de datos de un sistema OLAP. Los cubos OLAP son rígidos, ya que no se pueden cambiar las dimensiones ni los datos subyacentes una vez modelados. Por ejemplo, si se agrega la dimensión de almacenamiento a un cubo con las dimensiones de producto, ubicación y tiempo, tendrá que remodelarlo completamente.

### **Herramientas analíticas OLAP**

Los analistas empresariales utilizan las herramientas OLAP para interactuar con el cubo OLAP. Llevan a cabo operaciones como *slicing* (dividir), *dicing* (cortar)

y *pivoting* (rotar) para obtener una visión más profunda de la información específica dentro del cubo OLAP.

### **¿Cómo funciona OLAP?**

Un sistema de procesamiento analítico en línea (OLAP) funciona recopilando, organizando, agregando y analizando datos mediante los siguientes pasos:

1. El servidor OLAP recopila datos de varios orígenes, incluidas las bases de datos relacionales y los almacenamientos de datos.
2. A continuación, las herramientas de extracción, transformación y carga (ETL) limpian, agregan, calculan previamente y almacenan datos en un cubo OLAP según el número de dimensiones especificadas.
3. Los analistas empresariales utilizan herramientas OLAP para consultar y generar informes a partir de los datos multidimensionales del cubo OLAP.

OLAP utiliza expresiones multidimensionales (MDX) para consultar el cubo OLAP. MDX es una consulta, como SQL, que proporciona un conjunto de instrucciones para manipular bases de datos.

Los sistemas de procesamiento analítico en línea (OLAP) funcionan de tres formas principales.

### **MOLAP**

El procesamiento analítico multidimensional en línea (MOLAP) implica la creación de un cubo que representa datos multidimensionales de un almacenamiento de datos. El sistema MOLAP almacena en el hipercubo los datos calculados previamente. Los ingenieros de datos utilizan MOLAP porque este tipo de tecnología OLAP proporciona un análisis rápido.

### **ROLAP**

En lugar de utilizar un cubo de datos, el procesamiento analítico relacional en línea (ROLAP) permite a los ingenieros hacer análisis de datos multidimensionales en una base de datos relacional. En otras palabras, los ingenieros de datos utilizan consultas SQL para buscar y recuperar información específica según las dimensiones requeridas. ROLAP es adecuado para analizar datos extensos y detallados. Sin embargo, ROLAP tiene un rendimiento de consultas lento en comparación con MOLAP.

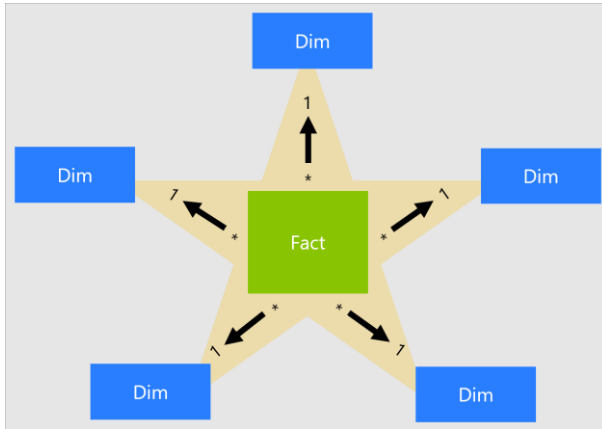
### **HOLAP**

El procesamiento analítico híbrido en línea (HOLAP) combina MOLAP y ROLAP para ofrecer lo mejor de ambas arquitecturas. HOLAP permite a los ingenieros recuperar rápidamente los resultados analíticos de un cubo de datos y extraer información detallada de las bases de datos relacionales.

## ¿Qué es el modelado de datos en OLAP?

El modelado de datos es la representación de estos en almacenamientos o bases de datos de procesamiento analítico en línea (OLAP). El modelado de datos es esencial en el procesamiento analítico relacional en línea (ROLAP) porque los analiza directamente desde la base de datos relacional. Almacena datos multidimensionales como un esquema en estrella o de copo de nieve.

### Esquema en estrella



El esquema en estrella consta de una tabla de hechos y de varias tablas de dimensiones. La tabla de hechos es una tabla de datos que contiene valores numéricos relacionados con un proceso empresarial y la tabla de dimensiones contiene valores que describen cada atributo de la tabla de hechos. La tabla de hechos se refiere a tablas de dimensiones con claves externas: identificadores únicos que se

correlacionan con la información correspondiente de la tabla de dimensiones.

En un esquema de estrella, una tabla de hechos se conecta a varias tablas de dimensiones para que el modelo de datos parezca una estrella. El siguiente es un ejemplo de una tabla de datos para la venta de productos:

- ID de producto
- ID de ubicación
- ID del vendedor
- Cantidad de ventas

El ID del producto indica al sistema de base de datos que recupere información de la tabla de dimensiones del producto, que puede tener el siguiente aspecto:

- ID de producto
- Nombre del producto
- Tipo de producto
- Costo del producto

Del mismo modo, el ID de ubicación apunta a una tabla de dimensiones de ubicación, que puede consistir en lo siguiente:

- ID de ubicación



- País
- Ciudad

La tabla del vendedor puede tener el siguiente aspecto:

- ID del vendedor
- Nombre
- Apellido
- Correo electrónico

### Esquema de copo de nieve

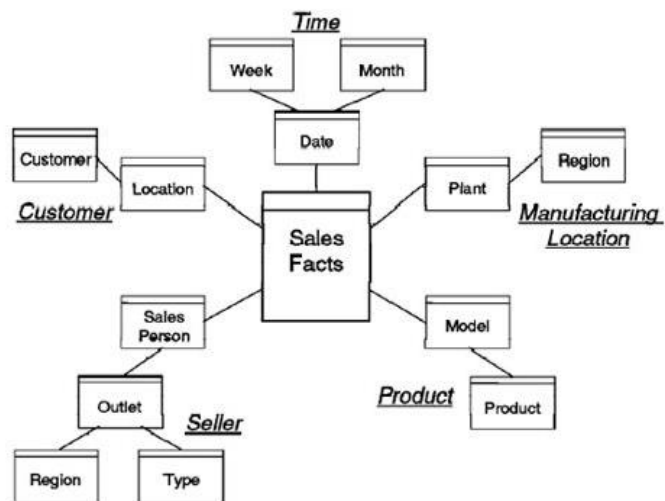
El esquema de copo de nieve es una extensión del esquema en estrella. Algunas tablas de dimensiones pueden conducir a una o más tablas secundarias. Esto da como resultado que, cuando se juntan las tablas de dimensiones, se forme algo parecido a un copo de nieve.

Por ejemplo, la tabla de dimensiones del producto puede contener los siguientes campos:

- ID de producto
- Nombre del producto
- ID del tipo de producto
- Costo del producto

El ID del tipo de producto se conecta a otra tabla de dimensiones, como se muestra en el siguiente ejemplo:

- ID del tipo de producto
- Nombre del tipo
- Versión
- Variante



## **¿Qué son las operaciones OLAP?**

Los analistas de negocios llevan a cabo varias operaciones analíticas básicas con un cubo multidimensional de procesamiento analítico en línea (MOLAP).

### **Consolidación**

En la consolidación, el sistema de procesamiento analítico en línea (OLAP) resume los datos de atributos específicos. En otras palabras, muestra datos menos detallados. Por ejemplo, puede ver las ventas de productos según Nueva York, California, Londres y Tokio. Una operación de consolidación proporcionaría una vista de los datos de ventas en función de países, como EE. UU., Reino Unido y Japón.

### **Profundización**

La profundización es lo opuesto a la operación de resumen. Los analistas empresariales bajan en la jerarquía de conceptos y extraen los detalles que necesitan. Por ejemplo, pueden pasar de ver los datos de ventas por años a visualizarlos por meses.

### **División**

Los ingenieros de datos usan la operación de división para crear una vista bidimensional a partir del cubo OLAP. Por ejemplo, un cubo MOLAP ordena los datos según los productos, las ciudades y los meses. Al dividir el cubo, los ingenieros de datos pueden crear una tabla similar a una hoja de cálculo compuesta de productos y ciudades para un mes específico.

### **Corte**

Los ingenieros de datos utilizan la operación de corte para crear un subcubo más pequeño a partir de un cubo OLAP. Determinan las dimensiones necesarias y construyen un cubo más pequeño a partir del hipercubo original.

### **Rotación**

La operación de rotación implica girar el cubo OLAP a lo largo de una de sus dimensiones para obtener una perspectiva diferente del modelo de datos multidimensional. Por ejemplo, un cubo OLAP tridimensional tiene las siguientes dimensiones en los ejes respectivos:

- Eje X: producto
- Eje Y: ubicación
- Eje Z: tiempo

Luego de rotarlo, el cubo OLAP tiene la siguiente configuración:

- Eje X: ubicación

- Eje Y: tiempo
- Eje Z: producto

## ¿Cómo se compara OLAP con otros métodos de análisis de datos?

### Minería de datos

La minería de datos es una tecnología de análisis que procesa grandes volúmenes de datos históricos para encontrar patrones e información. Los analistas empresariales utilizan herramientas de minería de datos para descubrir relaciones en ellos y hacer predicciones precisas de las tendencias futuras.

### OLAP y la minería de datos

El procesamiento analítico en línea (OLAP) es una tecnología de análisis de bases de datos que implica consultar, extraer y estudiar datos resumidos. Por otro lado, la minería de datos implica analizar en profundidad la información no procesada. Por ejemplo, los especialistas en marketing podrían usar herramientas de minería de datos para analizar el comportamiento de los usuarios a partir de los registros de cada visita que hagan al sitio. A continuación, podrían utilizar el software OLAP para inspeccionar esos comportamientos desde varios ángulos, como la duración, el dispositivo, el país, el idioma y el tipo de navegador.

### OLTP

El procesamiento de transacciones en línea (OLTP) es una tecnología de datos que almacena información de forma rápida y confiable en una base de datos. Los ingenieros de datos utilizan herramientas OLTP para almacenar datos transaccionales, como registros financieros, suscripciones a servicios y comentarios de los clientes, en una base de datos relacional. Los sistemas OLTP implican la creación, actualización y eliminación de registros en tablas relacionales.

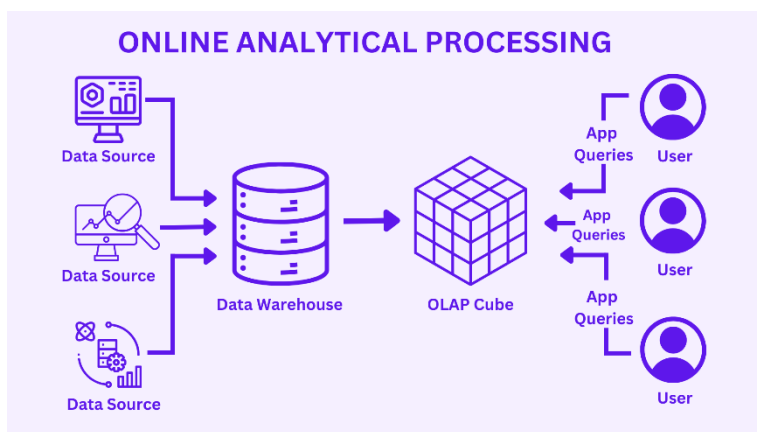
### OLAP y OLTP

OLTP es ideal para gestionar y almacenar varios flujos de transacciones en bases de datos. Sin embargo, no puede hacer consultas complejas desde la base de datos. Por lo tanto, los analistas empresariales utilizan un sistema OLAP para analizar datos multidimensionales. Por ejemplo, los científicos de datos conectan una base de datos OLTP a un cubo OLAP basado en la nube para hacer consultas de computación intensivas en los datos históricos (s/f-a).

#### 3.4.1. On-Line Analytical Processing

Online Analytical Processing o Procesamiento analítico en línea (OLAP) El procesamiento analítico en línea (OLAP) es una tecnología que organiza grandes bases de datos empresariales y proporciona análisis complejo. Se puede utilizar para realizar consultas analíticas complejas sin afectar negativamente los sistemas transaccionales.

Las bases de datos que utiliza una empresa para almacenar todas sus transacciones y registros se llaman bases de datos de procesamiento de transacciones en línea (OLTP). Normalmente, estas bases de datos tienen registros que se introducen uno cada vez. A menudo contienen una gran cantidad de información de valor para la organización. Sin embargo, las bases de datos que se usan para OLTP no se diseñaron para el análisis. Por lo tanto, obtener respuestas de estas bases de datos es costoso en términos de tiempo y esfuerzo. Los sistemas OLAP se han diseñado para ayudar a extraer de los datos esta información de inteligencia empresarial con un alto rendimiento. Esto se debe a que las bases de datos OLAP se optimizan para cargas de trabajo grandes en lecturas y pequeñas en escrituras (Procesamiento analítico en línea (OLAP), s/f).



El procesamiento analítico en línea (online analytical processing, OLAP) es un método informático que permite a los usuarios extraer y consultar datos de manera fácil y selectiva para analizarlos desde diferentes puntos de vista. Las consultas de inteligencia empresarial OLAP a

menudo ayudan en el análisis de tendencias, informes financieros, previsión de ventas, presupuestos y otros propósitos de planificación (De, 2021).

### Cómo funcionan los sistemas OLAP

Para facilitar este tipo de análisis, los datos se recopilan de múltiples fuentes de datos y se almacenan en almacenes de datos, luego se limpian y organizan en cubos de datos. Cada cubo OLAP contiene datos categorizados por dimensiones (como clientes, región geográfica de ventas y período de tiempo) derivados de tablas dimensionales en los almacenes de datos. Luego, las dimensiones se completan con miembros (como nombres de clientes, países y meses) que están organizados jerárquicamente. Los cubos OLAP a menudo se resumen previamente en todas las dimensiones para mejorar drásticamente el tiempo de consulta en las bases de datos relacionales.

Luego, los analistas pueden realizar cinco tipos de operaciones analíticas OLAP contra estas bases de datos multidimensionales:

- Enrollar (roll-up). También conocida como consolidación o desglose, esta operación resume los datos a lo largo de la dimensión.
- Profundizar (drill-down). Esto permite a los analistas navegar más profundamente entre las dimensiones de los datos, por ejemplo,

profundizando desde "período de tiempo" a "años" y "meses" para trazar el crecimiento de las ventas de un producto.

- Cortar (slice). Esto permite que un analista tome un nivel de información para mostrar, como, por ejemplo, "ventas en 2017".
- Cubo o dado (dice). Esto permite que un analista seleccione datos de múltiples dimensiones para analizar, como, por ejemplo, "ventas de pelotas de playa azules en Iowa en 2017".
- Pivotar (pivot). Los analistas pueden obtener una nueva vista de los datos girando los ejes de datos del cubo.

El software OLAP luego localiza la intersección de dimensiones, como todos los productos vendidos en la región oriental por encima de un precio determinado durante un período de tiempo determinado, y los muestra. El resultado es la "medida"; cada cubo OLAP tiene al menos una o quizás cientos de medidas, que se derivan de la información almacenada en tablas de hechos en el almacén de datos.

### **Tipos de sistemas OLAP**

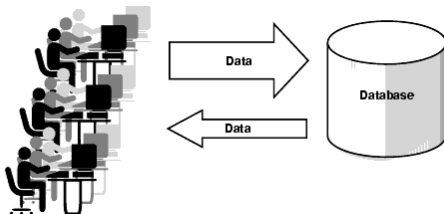
Los sistemas OLAP (procesamiento analítico en línea) generalmente se dividen en uno de tres tipos:

**OLAP multidimensional (MOLAP)** es OLAP que indexa directamente en una base de datos multidimensional.

**OLAP relacional (ROLAP)** es OLAP que realiza un análisis multidimensional dinámico de datos almacenados en una base de datos relacional.

**OLAP híbrido (HOLAP)** es una combinación de ROLAP y MOLAP. HOLAP fue desarrollado para combinar la mayor capacidad de datos de ROLAP con la capacidad de procesamiento superior de MOLAP (De, 2021).

### **3.4.2. On-Line Transaction Processing**

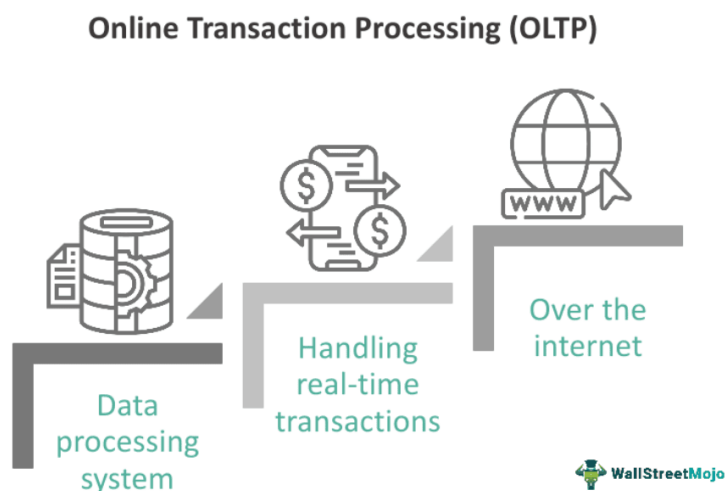


OLTP o procesamiento de transacciones en línea es un tipo de procesamiento de datos que consiste en ejecutar una serie de transacciones que ocurren simultáneamente: banca en

línea, compras, ingreso de pedidos o envío de mensajes de texto, por ejemplo. Estas transacciones tradicionalmente se denominan transacciones económicas o financieras, registradas y aseguradas para que una empresa pueda acceder a la información en cualquier momento con fines contables o de presentación de informes.

En el pasado, OLTP se limitaba a interacciones del mundo real en las que se intercambiaba algo: dinero, productos, información, solicitud de servicios, etc. Pero la definición de transacción en este contexto se ha ampliado a lo largo de los años,

especialmente desde la llegada de Internet, para abarcar cualquier tipo de interacción o compromiso digital con una empresa que pueda activarse desde cualquier parte del mundo y a través de cualquier sensor conectado a la web.



También incluye cualquier tipo de interacción o acción, como descargar archivos PDF en una página web, ver un video específico o activadores de mantenimiento automático o comentarios en canales sociales que pueden ser críticos para que una empresa los registre para brindar un mejor servicio a sus clientes.

La definición principal de transacciones (económicas o financieras) sigue siendo la base de la mayoría de los sistemas OLTP, por lo que el procesamiento de transacciones en línea generalmente implica insertar, actualizar y/o eliminar pequeñas cantidades de datos en un almacén de datos para recopilar, administrar y proteger esas transacciones. Normalmente, una aplicación web, móvil o empresarial rastrea todas esas interacciones o transacciones con clientes, proveedores o socios y las actualiza en la base de datos OLTP. Estos datos de transacciones almacenados en la base de datos son críticos para las empresas y se utilizan para generar informes o se analizan para tomar decisiones basadas en datos.

Las empresas suelen tener dos tipos de capacidades de procesamiento de datos: OLTP y OLAP.

### **OLTP frente a OLAP**

Aunque suenan similares y ambos son sistemas de procesamiento de datos en línea, existe una gran diferencia entre los dos.

OLTP permite la ejecución en tiempo real de un gran número de transacciones por parte de un gran número de personas, mientras que el procesamiento analítico en línea (OLAP) normalmente implica consultar estas transacciones (también denominadas registros) en una base de datos con fines analíticos. OLAP ayuda a las empresas a extraer información de los datos de sus transacciones para poder utilizarla para tomar decisiones más informadas.

La siguiente tabla muestra una comparación entre los sistemas OLTP y OLAP.

<b>sistemas OLTP</b>	<b>sistemas OLAP</b>
Permitir la ejecución en tiempo real de un gran número de transacciones de bases de datos por parte de un gran número de personas.	Generalmente implican consultar muchos registros (incluso todos los registros) en una base de datos con fines analíticos.
Requiere tiempos de respuesta ultrarrápidos	Requerir tiempos de respuesta que sean órdenes de magnitud más lentos que los requeridos por OLTP
Modifique pequeñas cantidades de datos con frecuencia y, por lo general, implica un equilibrio de lecturas y escrituras.	No modifique datos en absoluto; las cargas de trabajo suelen ser de lectura intensiva
Utilice datos indexados para mejorar los tiempos de respuesta	Almacene datos en formato de columnas para permitir un fácil acceso a una gran cantidad de registros
Requerir copias de seguridad de bases de datos frecuentes o simultáneas	Requiere copias de seguridad de la base de datos mucho menos frecuentes
Requiere relativamente poco espacio de almacenamiento	Normalmente tienen importantes requisitos de espacio de almacenamiento, porque almacenan grandes cantidades de datos históricos.
Por lo general, ejecuta consultas simples que involucran solo uno o varios registros.	Ejecute consultas complejas que involucren una gran cantidad de registros

Entonces, OLTP es un sistema de modificación de datos en línea, mientras que OLAP es un sistema de almacenamiento de datos multidimensionales históricos en línea que se utiliza para recuperar grandes cantidades de datos con fines analíticos. OLAP generalmente proporciona análisis de datos capturados por uno o más sistemas OLTP.

### **Requisitos para un sistema OLTP**

La arquitectura más común de un sistema OLTP que utiliza datos transaccionales es una arquitectura de tres niveles que normalmente consta de un nivel de presentación, un nivel de lógica empresarial y un nivel de almacén de datos. El nivel de presentación es la interfaz, donde la transacción se origina a través de una interacción humana o es generada por el sistema. El nivel lógico consta de reglas que verifican la transacción y garantizan que todos los datos necesarios para completarla estén disponibles. El nivel del almacén de datos almacena la transacción y todos los datos relacionados con ella.

Las principales características de un sistema de procesamiento de transacciones en línea son las siguientes:

- **Cumplimiento de ACID:** los sistemas OLTP deben garantizar que toda la transacción se registre correctamente. Una transacción suele ser la ejecución de un programa que puede requerir la ejecución de múltiples pasos u operaciones. Puede estar completo cuando todas las partes involucradas reconocen la transacción, o cuando se entrega el producto/servicio, o cuando se realiza una cierta cantidad de actualizaciones en las tablas específicas de la base de datos. Una transacción se registra correctamente sólo si se ejecutan y registran todos los pasos involucrados. Si hay algún error en alguno de los pasos, se debe cancelar toda la transacción y eliminar todos los pasos del sistema. Por lo tanto, los sistemas OLTP deben cumplir con propiedades atómicas, consistentes, aisladas y duraderas (ACID) para garantizar la precisión de los datos en el sistema.
  - Atómico: Los controles de atomicidad garantizan que todos los pasos de una transacción se completen exitosamente como grupo. Es decir, si alguno de los pasos entre las transacciones falla, todos los demás pasos también deben fallar o revertirse. La finalización exitosa de una transacción se llama compromiso. El fracaso de una transacción se llama aborto.
  - Consistente: la transacción preserva la coherencia interna de la base de datos. Si ejecuta la transacción por sí sola en una base de datos que inicialmente es consistente, cuando la transacción termina de ejecutarse, la base de datos vuelve a ser consistente.
  - Aislada: la transacción se ejecuta como si se ejecutara sola, sin otras transacciones. Es decir, el efecto de ejecutar un conjunto de transacciones es el mismo que ejecutarlas una a la vez. Este comportamiento se denomina serialización y normalmente se implementa bloqueando filas específicas de la tabla.
  - Durable: Los resultados de la transacción no se perderán en caso de falla.
- **Concurrencia:** los sistemas OLTP pueden tener poblaciones de usuarios enormemente grandes, y muchos usuarios intentan acceder a los mismos datos al mismo tiempo. El sistema debe garantizar que todos estos usuarios que intentan leer o escribir en el sistema puedan hacerlo simultáneamente. Los controles de concurrencia garantizan que dos usuarios que accedan a los mismos datos en el sistema de base de datos al mismo tiempo no podrán cambiar esos datos, o que un usuario tenga que esperar hasta que el otro haya terminado de procesar antes de cambiar ese dato.



- **Escala:** los sistemas OLTP deben poder ampliarse y reducirse instantáneamente para administrar el volumen de transacciones en tiempo real y ejecutar transacciones simultáneamente, independientemente de la cantidad de usuarios que intenten acceder al sistema.
- **Disponibilidad:** un sistema OLTP debe estar siempre disponible y listo para aceptar transacciones. La pérdida de una transacción puede provocar una pérdida de ingresos o tener implicaciones legales. Debido a que las transacciones se pueden ejecutar desde cualquier parte del mundo y en cualquier momento, el sistema debe estar disponible las 24 horas del día, los 7 días de la semana.
- **Alto rendimiento y tiempo de respuesta corto:** los sistemas OLTP requieren tiempos de respuesta de nanosegundos o incluso más cortos para mantener productivos a los usuarios empresariales y satisfacer las crecientes expectativas de los clientes.
- **Confiabilidad:** los sistemas OLTP generalmente leen y manipulan pequeñas cantidades de datos altamente selectivas. Es fundamental que en un momento dado los datos de la base de datos sean fiables y dignos de confianza para los usuarios y las aplicaciones que acceden a esos datos.
- **Seguridad:** debido a que estos sistemas almacenan datos de transacciones de clientes altamente confidenciales, la seguridad de los datos es fundamental. Cualquier incumplimiento puede resultar muy costoso para la empresa.
- **Recuperabilidad:** Los sistemas OLTP deben tener la capacidad de recuperarse en caso de cualquier falla de hardware o software.

### **Bases de datos para cargas de trabajo OLTP**

Las bases de datos relacionales se crearon específicamente para aplicaciones de transacciones. Incorporan todos los elementos esenciales necesarios para almacenar y procesar grandes volúmenes de transacciones, al mismo tiempo que se actualizan continuamente con nuevas características y funcionalidades para extraer más valor de estos ricos datos de transacciones. Las bases de datos relacionales están diseñadas desde cero para proporcionar la mayor disponibilidad posible y el rendimiento más rápido. Proporcionan simultaneidad y cumplimiento ACID para que los datos sean precisos, estén siempre disponibles y sean fácilmente accesibles. Almacenan datos en tablas después de extraer las relaciones entre los datos para que cualquier aplicación pueda utilizarlos, garantizando una única fuente de verdad (What is OLTP?, s/f).

### 3.5. Bases de datos en CLOUD

Una base de datos en la nube es una base de datos que se implementa, entrega y accede a la nube. Las bases de datos en la nube organizan y almacenan datos estructurados, no estructurados y semiestructurados, al igual que las bases de datos locales tradicionales. Sin embargo, también proporcionan muchos de los mismos beneficios de la computación en la nube, incluida la velocidad, la escalabilidad, la agilidad y los costos reducidos.



Una base de datos en la nube es una base de datos creada para ejecutarse en un entorno de nube pública o híbrida a fin de ayudar a organizar, almacenar y administrar datos dentro de una organización. Las bases de datos en la nube se pueden ofrecer como una base de datos como servicio administrado (DBaaS) o implementada en una máquina virtual (VM) basada en la nube, y autoadministrada por un equipo de TI propio.

#### Tipos de bases de datos en la nube

Al igual que las bases de datos locales tradicionales, las bases de datos en la nube se pueden clasificar en bases de datos relacionales y no relacionales.

- Las **bases de datos relacionales en la nube** consisten en una o más tablas de columnas y filas, que te permiten organizar los datos en relaciones predefinidas para comprender la relación lógica de los datos. Por lo general, estas bases de datos usan un esquema de datos fijo, y puedes usar el lenguaje de consulta estructurado (SQL) para consultar y manipular datos. Son muy coherentes, confiables y son más adecuados para manejar grandes cantidades de datos estructurados.

Entre los ejemplos de bases de datos relacionales, se incluyen SQL Server, Oracle, MySQL, PostgreSQL, Spanner y Cloud SQL.

- Las **bases de datos no relacionales en la nube** almacenan y administran datos no estructurados, como mensajes de texto y correos electrónicos para mensajes móviles, documentos, encuestas, archivos de rich media y datos de sensores. No siguen un esquema definido claramente, como las bases de datos relacionales, y te permiten guardar y organizar la información sin importar su formato.

Algunos ejemplos de bases de datos no relacionales incluyen MongoDB, Redis, Cassandra, Hbase y Cloud Bigtable.

#### ¿Por qué usar una base de datos en la nube?

En la actualidad, la cantidad de datos generados y recopilados crece de manera exponencial. No solo es más variado, sino también muy dispar. Ahora los datos pueden residir en bases de datos locales y en aplicaciones y servicios distribuidos en la nube, lo que dificulta la integración mediante enfoques tradicionales. Además, el procesamiento de datos en tiempo real se vuelve esencial para el éxito comercial, ya que las demoras y retrasos en la entrega de datos a aplicaciones esenciales pueden tener consecuencias catastróficas.

A medida que se acelera la adopción de la nube y la forma en que usamos los datos, las bases de datos heredadas enfrentan desafíos significativos.

Las bases de datos en la nube ofrecen flexibilidad, confiabilidad, seguridad, asequibilidad y mucho más. Proporcionar una base sólida para crear aplicaciones empresariales modernas. En particular, pueden adaptarse con rapidez a las cargas de trabajo y las demandas cambiantes sin aumentar la carga de trabajo de los equipos ya cargados.

### **Ventajas de las bases de datos en la nube**

- Reducción de la sobrecarga operativa
- Mayor agilidad y escalabilidad
- Disminuir el costo total de propiedad (TCO)
- Opciones flexibles de base de datos
- Datos seguros y protegidos
- Mejor confiabilidad

### **Base de datos en la nube**

Una base de datos en la nube es una base de datos desarrollada, implementada y a la que se accede en un entorno de nube, como una nube privada, pública o híbrida.

Existen estos dos modelos principales de implementación de bases de datos en la nube:

#### **Base de datos tradicional**

Es muy similar a una base de datos de gestión interna, salvo debido al aprovisionamiento de su infraestructura. En este caso, una organización compra espacio de máquina virtual de un proveedor de servicios en la nube y la base de datos se implementa en la nube. Los desarrolladores de la organización emplean un modelo de DevOps o el personal de TI tradicional controla la base de datos. La organización se encarga de la supervisión y la gestión de la base de datos.

#### **Base de datos como servicio (DBaaS)**

Cuando una organización contrata a un proveedor de servicios en la nube a través de un servicio de suscripción según una cuota. El proveedor de servicios ofrece diferentes tareas operativas, de mantenimiento, administrativas y de gestión de

bases de datos en tiempo real para el usuario final. La base de datos se ejecuta en la infraestructura del proveedor del servicio. Este modelo de uso suele incluir la automatización en las zonas de suministro, copia de seguridad, escalabilidad, alta disponibilidad, seguridad, aplicación de parches y supervisión del estado. El modelo DBaaS es el que aporta el mayor valor a las organizaciones, ya que les permite emplear una gestión de base de datos subcontratada y optimizada mediante la automatización de software, en lugar de contratar y gestionar expertos internos.

### **Opciones de gestión de las bases de datos en la nube**

Las empresas tienen diferentes opciones a la hora de administrar sus bases de datos en la nube. Los estilos de gestión de bases de datos se pueden generalizar en las siguientes cuatro categorías:

#### **Bases de datos en la nube autogestionadas**

En este modelo, una organización ejecuta su base de datos en la nube, pero gestiona la propia base de datos utilizando recursos internos y sin que el proveedor de servicios en la nube integre ninguna automatización. Este modelo ofrece algunos de los beneficios estándar de ubicar una base de datos en la nube (incluidas una flexibilidad y agilidad mejoradas), pero la organización aún conserva la responsabilidad y el control sobre la gestión de las bases de datos (¿Qué es una base de datos en la nube?, s/f).

##### **3.5.1. Automatizadas**

#### **Bases de datos en la nube automatizadas**

En este modelo, las empresas utilizan interfaces de programación de aplicaciones (API) de servicios en la nube para bases de datos a fin de colaborar con operaciones de ciclo de vida, pero mantienen el acceso a los servidores de la base de datos y controlan la configuración de la base de datos y los sistemas operativos. Los servicios de base de datos automatizados presentan SLA limitados y, por lo general, no incluyen actividades planificadas, como la aplicación de parches y el mantenimiento (¿Qué es una base de datos en la nube?, s/f).

##### **3.5.2. Gestionadas**

#### **Bases de datos en la nube gestionadas**

Este modelo es similar al de las bases de datos en la nube automatizadas, con la excepción de que el proveedor de servicios en la nube no permite que los consumidores accedan a los servidores que alojan la base de datos. Los ajustes se limitan a las configuraciones en la nube suministradas por proveedores, ya que los usuarios finales no pueden instalar su propio software (¿Qué es una base de datos en la nube?, s/f).

##### **3.5.3. Autónomas**

#### **Bases de datos autónomas en la nube**

Este es un nuevo modelo operativo "manos libres" en el que la automatización y el aprendizaje automático eliminan el trabajo humano relacionado con la gestión de la base de datos y el ajuste del rendimiento. Algunos de los servicios son SLA en aplicaciones clave para la actividad comercial, como las operaciones sin tiempo de inactividad para tareas (previstas y no previstas) de bases de datos y ciclo de vida de servicios (¿Qué es una base de datos en la nube?, s/f).

**Material de apoyo:**

- [https://youtu.be/knVwokXITGI?si=DM\\_-U1vt\\_Y0oiTcT](https://youtu.be/knVwokXITGI?si=DM_-U1vt_Y0oiTcT)
- <https://youtu.be/6pM0Rmj57Vk?si=zxC4kMT3wsQdHqKU>
- [https://youtu.be/6S8A-1jBD5Y?si=F7\\_z6GwGkL1AMcLh](https://youtu.be/6S8A-1jBD5Y?si=F7_z6GwGkL1AMcLh)
- <https://youtu.be/o5q0IJvbqWU?si=-qBHso5gE6alGMmN>
- [https://youtu.be/WU1tUV\\_krtA?si=N4yuiyERfTA4P9gE](https://youtu.be/WU1tUV_krtA?si=N4yuiyERfTA4P9gE)
- [https://youtu.be/Zh8\\_TC9ukss?si=Q6\\_MtWLFT7vipe3q](https://youtu.be/Zh8_TC9ukss?si=Q6_MtWLFT7vipe3q)

## Bibliografía:

- Se hizo uso de Inteligencia Artificial (IA)
- DBA. (2017, enero 17). *Estructuras principales de una base de datos*. DBA dicit; DBA. <https://dbadixit.com/estructuras-principales-una-base-datos/>
- De, C. (2021, julio 19). *OLAP o procesamiento analítico en línea*. ComputerWeekly.es; TechTarget. <https://www.computerweekly.com/es/definicion/OLAP-o-procesamiento-analitico-en-linea>
- de Zúñiga, F. G. (2023, junio 15). *Tipos de bases de datos que existen*. Blog de arsys.es; Arsys. <https://www.arsys.es/blog/tipos-de-bases-de-datos-que-existen>
- García, F. (2023, enero 14). ▷ *Que es un Atributo en BASE DE DATOS*. Codigosql.top. [https://codigosql.top/bases-de-datos/atributo/?expand\\_article=1](https://codigosql.top/bases-de-datos/atributo/?expand_article=1)
- *Procesamiento analítico en línea (OLAP)*. (s/f). Microsoft.com. Recuperado el 4 de abril de 2024, de <https://learn.microsoft.com/es-es/azure/architecture/data-guide/relational-data/online-analytical-processing>
- *Qué es un esquema de base de datos*. (s/f). Lucidchart. Recuperado el 4 de abril de 2024, de <https://www.lucidchart.com/pages/es/que-es-un-esquema-de-base-de-datos>
- *¿Qué es un esquema de base de datos?* (s/f). Ibm.com. Recuperado el 4 de abril de 2024, de <https://www.ibm.com/mx-es/topics/database-schema>
- *Qué es un modelo de base de datos*. (s/f). Lucidchart. Recuperado el 4 de abril de 2024, de <https://www.lucidchart.com/pages/es/que-es-un-modelo-de-base-de-datos>
- *¿Qué es una base de datos?* (s/f). Oracle.com. Recuperado el 4 de abril de 2024, de <https://www.oracle.com/mx/database/what-is-database/>
- *¿Qué es una base de datos en la nube?* (s/f). Oracle.com. Recuperado el 4 de abril de 2024, de <https://www.oracle.com/mx/database/what-is-a-cloud-database/>
- *Tipos de base de datos*. (s/f). Intelequia. Recuperado el 4 de abril de 2024, de <https://intelequia.com/es/blog/post/tipos-de-base-de-datos>
- *Tutorial de estructura y diseño de bases de datos*. (s/f). Lucidchart. Recuperado el 4 de abril de 2024, de <https://www.lucidchart.com/pages/es/tutorial-de-estructura-y-diseno-de-bases-de-datos>
- *What is OLTP?* (s/f). Oracle.com. Recuperado el 4 de abril de 2024, de <https://www.oracle.com/database/what-is-oltp/>
- (S/f-a). Amazon.com. Recuperado el 4 de abril de 2024, de <https://aws.amazon.com/es/what-is/olap/>

- (S/f-b). Www.uv.mx. Recuperado el 4 de abril de 2024, de [https://www.uv.mx/personal/ermeneses/files/2020/09/Clase6-ModeloDeDatos\\_parteIII\\_ModeloRelacional.pdf](https://www.uv.mx/personal/ermeneses/files/2020/09/Clase6-ModeloDeDatos_parteIII_ModeloRelacional.pdf)