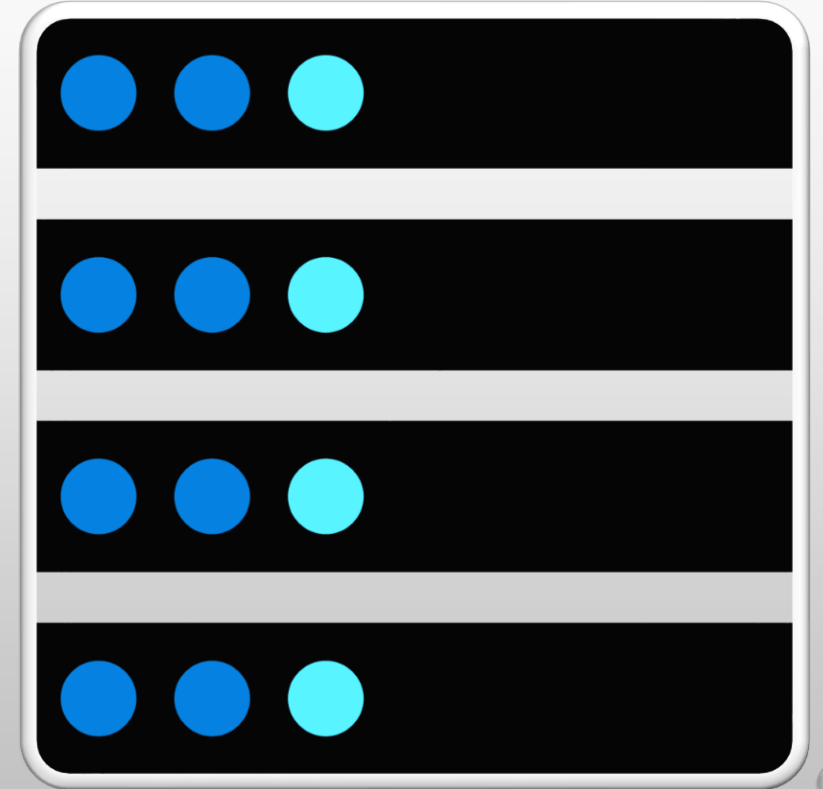
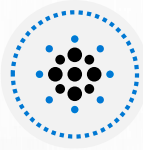


CONCEPTOS BÁSICOS DE DATOS



CONTENIDOS



Definición y almacenamiento de los datos.



Características de datos relacionales y no relacionales.



Describir y diferenciar las cargas de trabajo de datos.



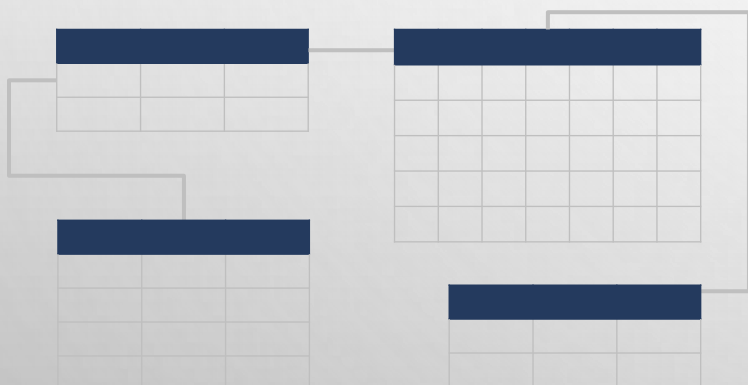
Describir y diferenciar datos por lotes y en Streaming.

¿QUÉ SON LOS DATOS?

- *ES UNA COLECCIÓN DE HECHOS, NÚMEROS, DESCRIPCIONES, OBJETOS, ALMACENADOS DE FORMA ESTRUCTURADA, SEMIESTRUCTURADA Y NO ESTRUCTURADA.*

Estructurados

Tabla



Se almacenan en una base de datos relacional como SQL Server o Azure SQL Database.

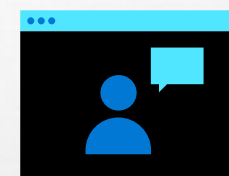
Semiestructurados

```
## Document 1 ## {
  "customerID": "103248",
  "name": { "first": "AAA",
    "last": "BBB" }, "address": {
    "street": "Main Street",
    "number": "101", "city":
    "Acity", "state": "NY" },
  "ccOnFile": "yes",
  "firstOrder": "02/28/2003" }
## Document 2 ## {
  "customerID": "103249",
  "name": { "title": "Mr",
    "forename": "AAA",
    "lastname": "BBB" },
  "address": { "street":
    "Another Street", "number":
    "202", "city": "Bcity",
    "county": "Gloucestershire",
    "country-region": "UK" },
  "ccOnFile": "yes" }
```



Es información que no residen en una base de datos relacional.

No estructurados

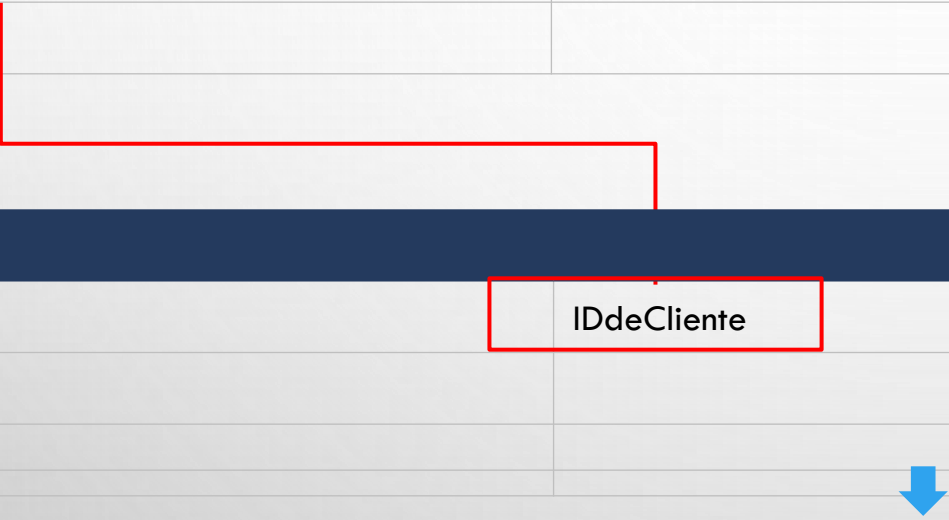


Los archivos de audio y vídeo y los archivos de datos binarios pueden no tener una estructura específica.

ALMACENAMIENTO DE DATOS TRANSACCIONALES FRENTE A ANALÍTICOS

Procesamiento transaccional en línea (OLTP)

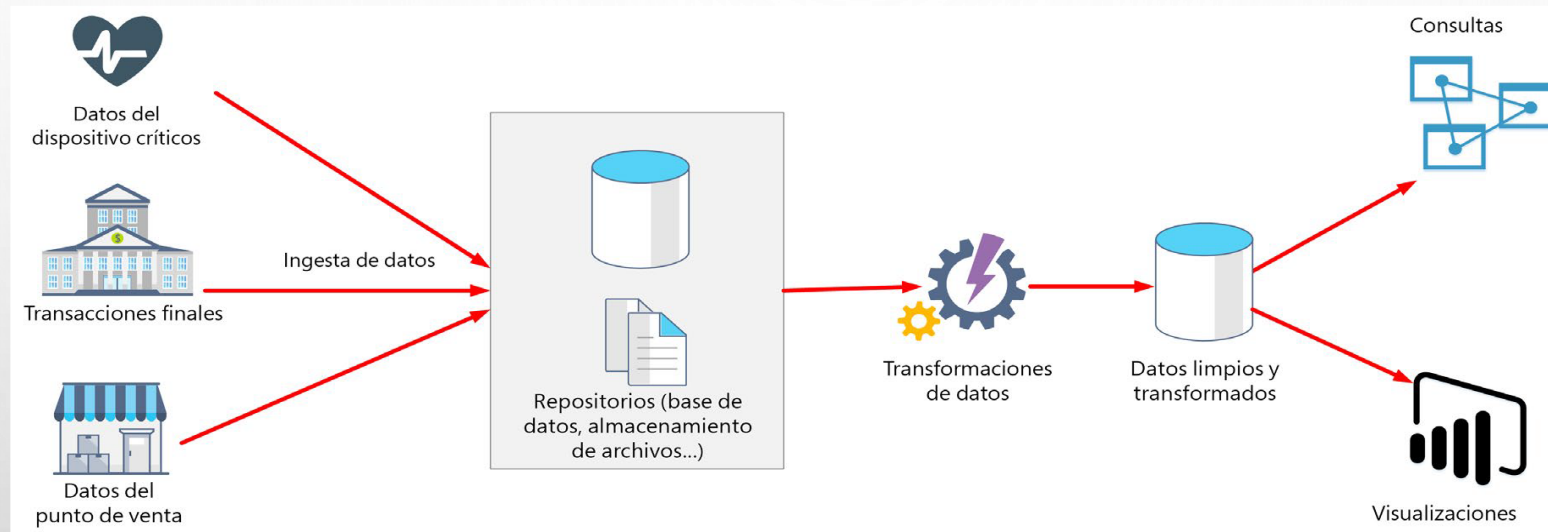
Cliente		
IDdeCliente	CustomerName	CustomerPhone
Pedidos		
OrderID	IDdeCliente	OrderDate



Registra *transacciones*. Una transacción podría ser financiera, como el movimiento de dinero entre cuentas en un sistema bancario, o podría ser parte de un sistema minorista que haga un seguimiento de los pagos de bienes y servicios por parte de la clientela. Piense en una transacción como una unidad de trabajo pequeña y discreta.

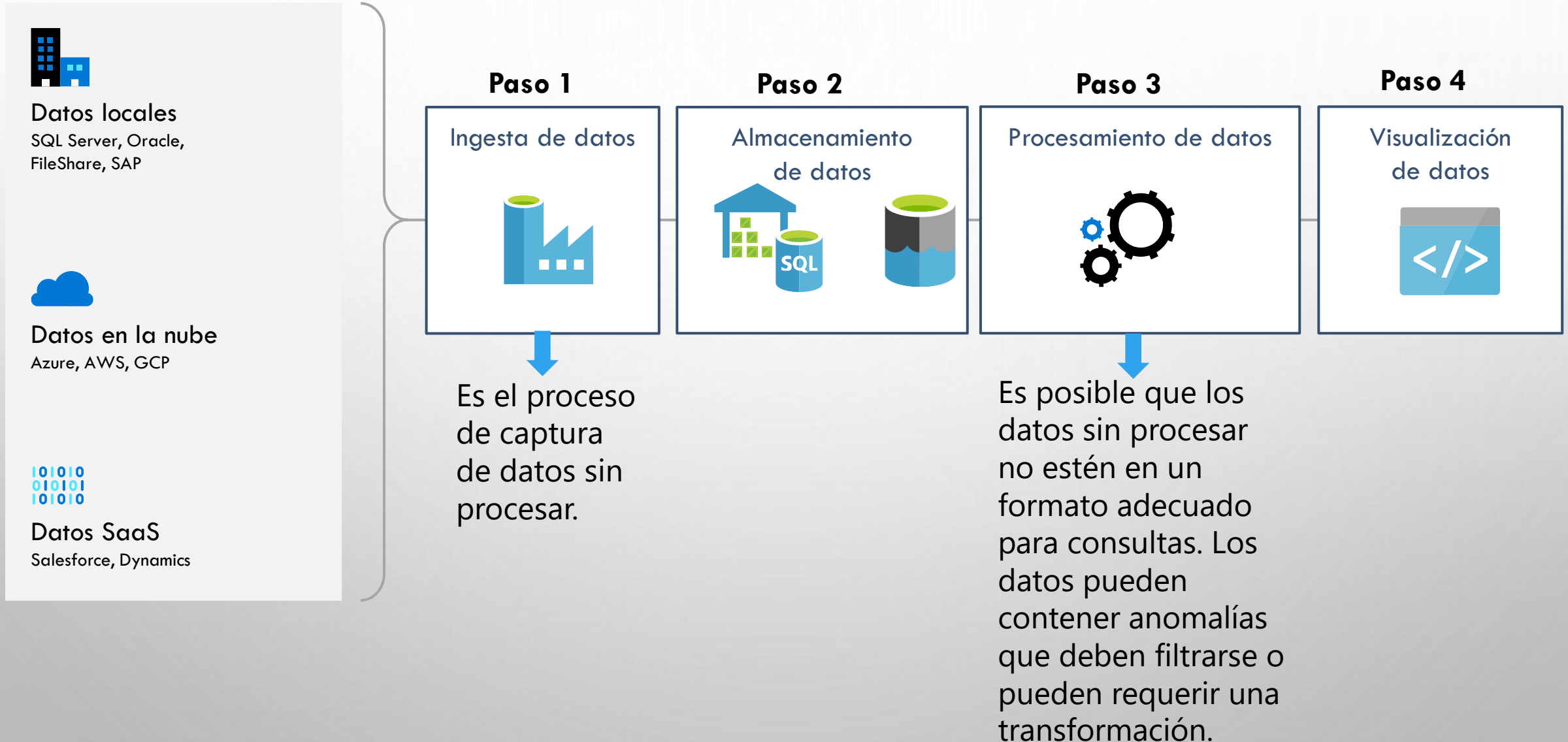
ALMACENAMIENTO DE DATOS TRANSACCIONALES FRENTE A ANALÍTICOS

Procesamiento analítico en línea (OLAP)



Está diseñado para ayudar a las personas usuarias empresariales que necesitan consultar datos y obtener una vista *general* de la información contenida en una base de datos.
Los sistemas analíticos se dedican a recopilar datos sin procesar y a usarlos para generar conclusiones.

PASO A PASO DE UN SISTEMA DE ANALÍTICO

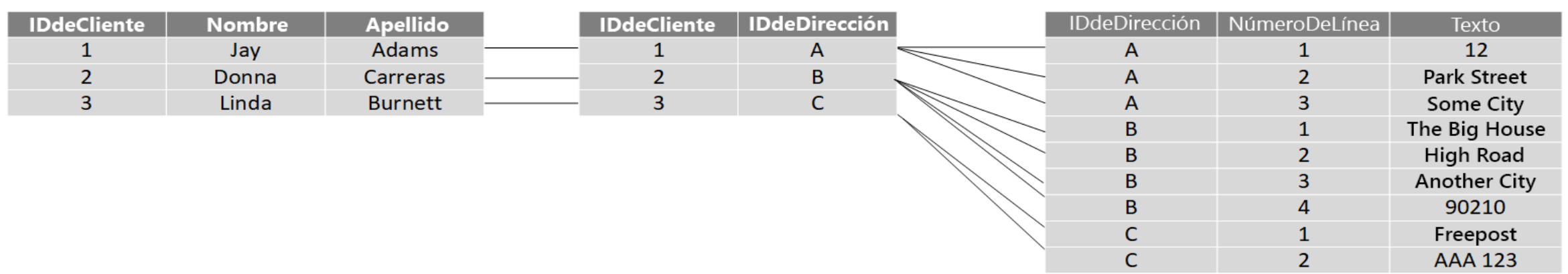
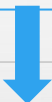


IDENTIFICACIÓN DE TIPOS DE DATOS Y ALMACENAMIENTO DE DATOS

Procesamiento Relacionales

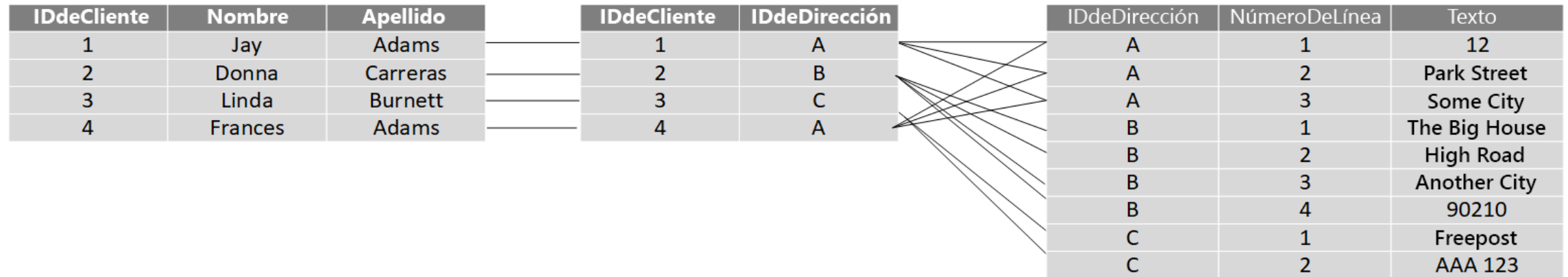
Las bases de datos relacionales proporcionan quizá el modelo más comprendido para conservar datos. La estructura simple de tablas y columnas las hace fáciles de usar inicialmente, pero la estructura rígida puede causar algunos problemas.

Esto se puede resolver mediante un proceso llamado **normalización**. Regularmente, el resultado final del proceso de normalización es que los datos se dividen en una gran cantidad de tablas estrechas y bien definidas (una tabla *estrecha* es una tabla con pocas columnas), con referencias de una tabla a otra, como se muestra en la imagen siguiente.



IDENTIFICACIÓN DE TIPOS DE DATOS Y ALMACENAMIENTO DE DATOS

Procesamiento No relacionales

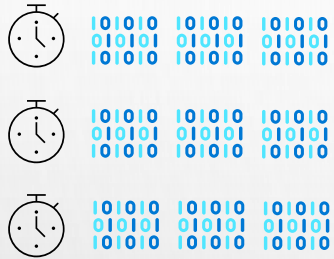


Le permiten almacenar datos en un formato que se asemeja más a la estructura original. Por ejemplo, en una base de datos de documentos, podría almacenar los detalles de cada clientela en un solo documento, como se muestra en la imagen arriba. Recuperar la información de la población de clientes, incluida la dirección, es cuestión de leer un solo documento. Sin embargo, existen algunas desventajas al utilizar una base de datos de documentos.

En una base de datos de documentos, en caso de que las personas clientas tengan la misma dirección se duplicaría en los documentos de ambos. Esta duplicación no solo aumenta el almacenamiento requerido, sino que también, puede hacer que el mantenimiento sea más complejo (si la dirección cambia, debe modificarla en dos documentos).

DIFERENCIA ENTRE: DATOS POR LOTES/DATOS DE STREAMING

Lote



Los elementos de datos recién llegados se recopilan en un grupo. El momento exacto en que se procesa cada grupo, se puede determinar de varias formas.

Un ejemplo: es la forma en que las compañías de tarjetas de crédito controlan la facturación. La clientela no recibe una factura por cada compra con tarjeta de crédito por separado, sino una factura mensual por todas las compras de ese mes.



Streaming



En el procesamiento de flujo, cada nuevo dato se procesa cuando llega. Por ejemplo, la ingesta de datos es inherentemente en cada proceso de streaming, se manejan datos en tiempo real.

Un ejemplo: una institución financiera rastrea los cambios en el mercado de valores en tiempo real, calcula el valor en riesgo y reequilibra automáticamente las carteras, según los movimientos del precio de las acciones.

COMPRENDER LAS DIFERENCIAS ENTRE DATOS POR LOTES Y EN STREAMING

- **Ámbito de los datos:** el procesamiento por lotes puede procesar todos los datos del conjunto de datos. El procesamiento de secuencias generalmente solo tiene acceso a los datos más recientes recibidos, o dentro de una ventana de tiempo gradual (los últimos 30 segundos, por ejemplo).
- **Tamaño de los datos:** el procesamiento por lotes es adecuado para manejar grandes conjuntos de datos de manera eficiente. El procesamiento de secuencias está destinado a registros individuales o *microlotes* que constan de pocos registros.
- **Rendimiento:** la latencia para el procesamiento por lotes suele ser de unas pocas horas. El procesamiento de secuencias suele ocurrir de inmediato, con una latencia del orden de segundos o milisegundos. La latencia es el tiempo que tardan en recibirse y procesarse los datos.
- **Análisis:** por lo general, usa el procesamiento por lotes para realizar análisis complejos. El procesamiento de secuencias se utiliza para funciones de respuesta simples, agregados o cálculos, como media acumulada.