



U N I V E R S I D A D  
**COMPLUTENSE**  
M A D R I D

# Introducción a NoSQL

Gestión de la Información en la Web  
Enrique Martín - [emartinm@ucm.es](mailto:emartinm@ucm.es)  
Grados de la Fac. Informática

# ¿Qué son las BD NoSQL?

- Más que una definición se encuentran **características comunes**:
- Usan un **modelo de datos alternativo** al modelo **relacional** de las BD tradicionales. Por ello suelen **carecer** de algunas propiedades como:
  - **No garantizan seguridad en transacciones**: Si se está realizando una actualización y llega una consulta no se sabe si tendrá en cuenta los nuevos datos o no.
  - **No garantizan consistencia**: no todas soportan comprobaciones de integridad en los datos.

# ¿Qué son las BD NoSQL?

- **Carecen de esquema fijo**, lo que aporta flexibilidad.
- Soportan un **volumen de datos muy grande** (GB, TB, PB) y se ejecutan bien en entornos distribuidos (clústeres).
- Suelen ser de **código abierto**.
- Surgen **a partir del año 2000**.

# ¿Por qué surgen?

- Durante muchos años las bases de datos relacionales han sido las bases de datos más utilizadas.
- Poseen muy buenas cualidades debido a su eficiencia y al manejo de transacciones, sin embargo se pueden identificar 2 “limitaciones”
  - **Diferencia entre la representación** de datos en memoria y la de la base de datos (*impedance mismatch*)
  - **No suelen soportar cantidades de datos muy grandes**, ya que no suelen ejecutarse en sistemas distribuidos.

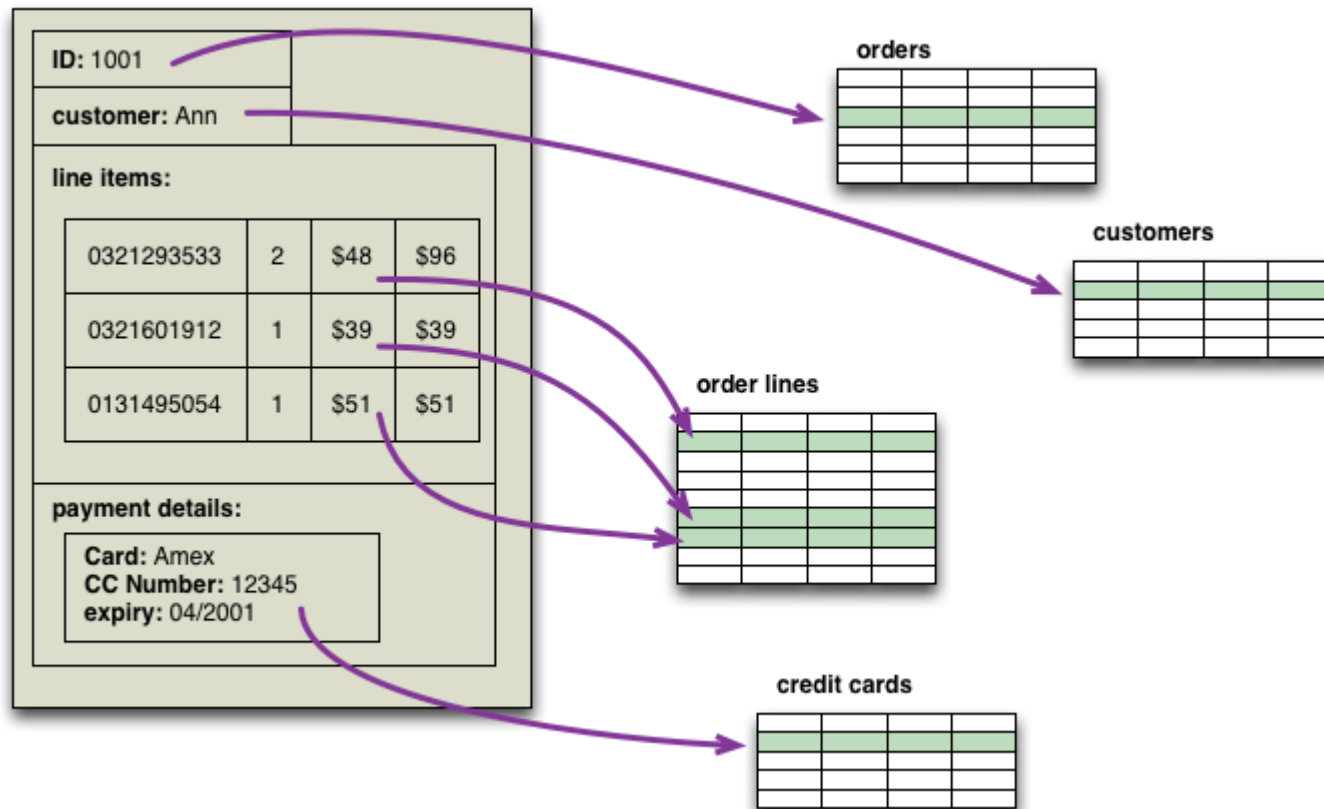
# Diferencias de representación

- En memoria, un pedido podría ser un **objeto** conteniendo la siguiente información:
  - Dirección de envío
  - Dirección de facturación
  - Datos de pago
  - Detalles de los productos
- En nuestra aplicación queremos que toda esa información sea almacenada y accedida de manera conjunta, ya que está íntimamente relacionada.

# Diferencias de representación

- Al utilizar bases de datos relacionales estamos forzados a utilizar su modelo de representación basado en tuplas: filas en distintas tablas:
  - Direcciones: tabla de usuarios, direcciones y código postal
  - Datos de pago: tabla de tarjetas de crédito y cuentas Paypal
  - Detalles de los productos: tabla de pedidos
- Para almacenar un pedido tendríamos que traducirlo en **distintas filas para las distintas tablas**. Lo mismo para consultarlo.

# Diferencias de representación



# Diferencias de representación

- Esta diferencia entre la representación en memoria (la que usa la aplicación) y la representación en la base de datos hace que sea necesario dedicar **esfuerzo** a esta **fase de traducción**.
- Las bases de datos NoSQL, al **carecer de esquema fijo**, te permitirían almacenar los datos de una manera mucho más cercana a cómo están representados en **memoria**, ahorrando tiempo durante el desarrollo.



# Volumen de datos

- Actualmente la velocidad a la que se generan los datos ha crecido de manera descomunal:
  - Datos procedentes de entidades públicas: climatología, ambiental, economía, etc.
  - Datos procedentes de usuarios: búsquedas, flujo de clics, preferencias. Etc.
  - Datos procedentes del propio negocio: pedidos, productos, tiendas.

# Volumen de datos

- Las bases de datos relacionales soportan gran cantidad de datos, pero no fueron diseñadas para ejecutarse en entornos **distribuidos** lo que al final limita su crecimiento.
- Se pueden extender a este tipo de entornos, pero algunas características como la seguridad de las **transacciones** o la comprobación de restricciones de **integridad** requieren un tratamiento delicado.

# Volumen de datos

- Las bases de datos **NoSQL** surgen con la idea de **manejar grandes volúmenes de datos** desde el inicio.
- Para ellos proporcionan **soporte para distribución** de manera sencilla, a costa de sacrificar algunas características de las bases de datos relacionales:
  - No comprueban la integridad.
  - Se pierden garantías de las transacciones.
  - Durante algún tiempo el estado de la base de datos puede no ser consistente.

# Familias de BBDD NoSQL

- Existen diferentes modelos de representación de datos dentro de las bases de datos NoSQL. Los más importantes son:
  - Bases de datos orientadas a **documento**: **MongoDB**, CouchDB.
  - Bases de datos **clave-valor**: Riak, Redis, Dynamo, Voldemort.
  - Bases de datos basadas en **columnas**: HBase, Cassandra, Hypertable, SimpleDB
  - Bases de datos de **grafos**: Neo4J, Infinite Graph.