

Tema 2

Modelos NOSQL

¿Qué es una base de datos NOSQL?

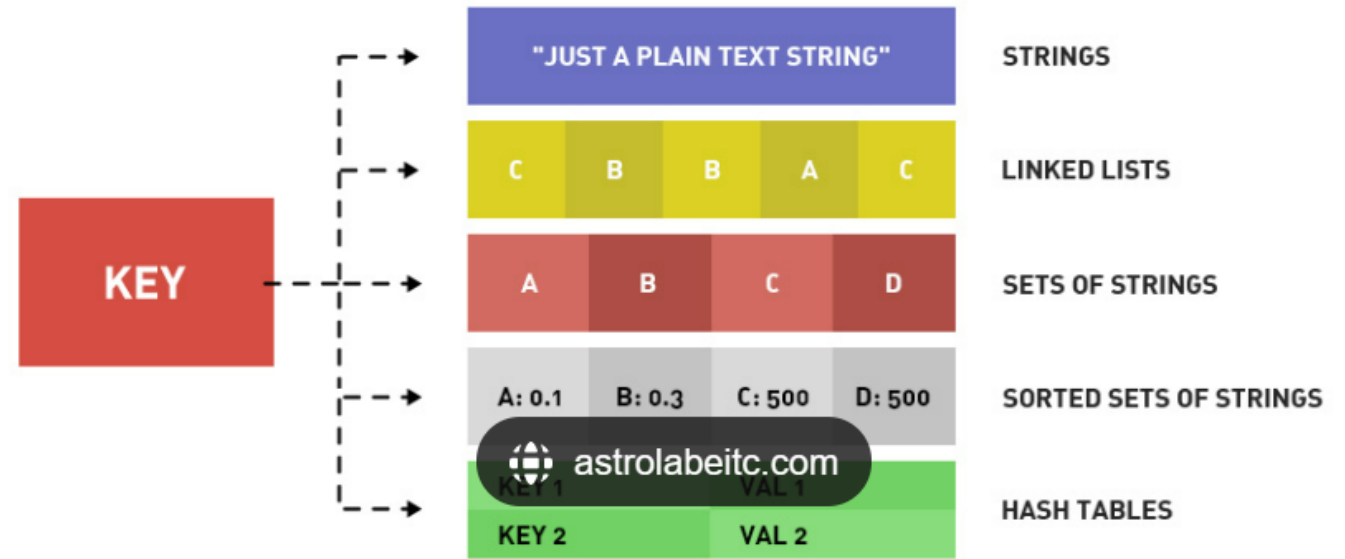
- NoSQL es una base de datos no relacional, lo que significa que permite diferentes estructuras que una base de datos SQL (no filas y columnas) y más flexibilidad para usar el formato que mejor se adapte a los datos.
- El término "NoSQL" no se acuñó hasta principios de la década de 2000. No significa que los sistemas no usen SQL, ya que las bases de datos NoSQL a veces admiten algunos comandos SQL. Más precisamente, "NoSQL" a veces se define como "no solo SQL".

Tipos de bases de datos NOSQL

- Key-Value store
- Document store
- Wide Column store
- Graph store
- In-memory-store

Tipos de bases de datos NOSQL

- Key-Value store
- Este modelo de datos sin esquema está organizado en un diccionario de pares clave-valor, donde cada elemento tiene una clave y un valor. Se usa comúnmente para almacenar en caché y almacenar información de la sesión del usuario, como carritos de compras. Sin embargo, no es ideal cuando necesita extraer varios registros a la vez.
- Redis ([Learn Redis with Free Online Courses](#) | [Redis University](#)) y Memcached son ejemplos de bases de datos de clave-valor de código abierto.



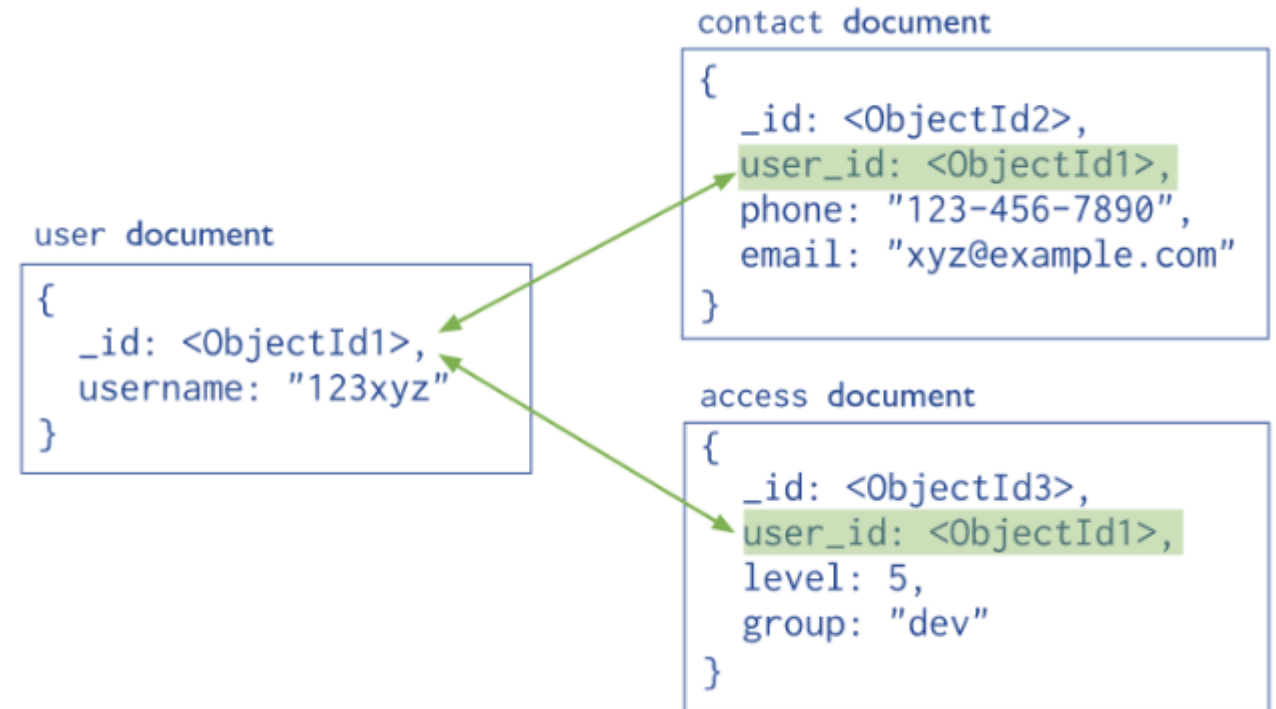
Tipos de bases de datos NOSQL

Document store

Almacenan datos como documentos y generalmente se almacenan en formatos JSON, XML o BSON.

Los desarrolladores también ganan más flexibilidad ya que los esquemas de datos no necesitan coincidir entre

Los casos de uso populares de bases de datos de documentos incluyen sistemas de gestión de contenido y perfiles de usuario. Un ejemplo de una base de datos orientada a documentos es MongoDB ([Free MongoDB Official Courses | MongoDB University](#)).



Tipos de bases de datos NOSQL

Wide-column store

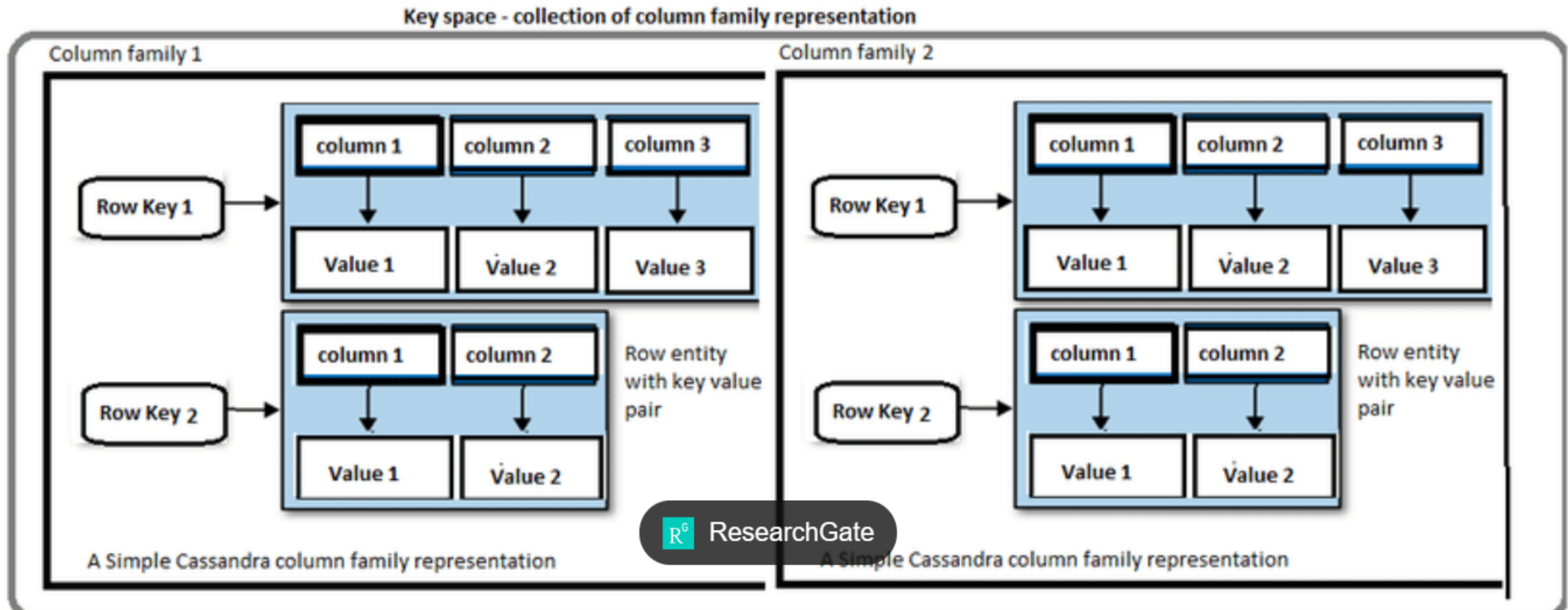
Estas bases de datos almacenan información en columnas, lo que permite a los usuarios acceder solo a las columnas específicas que necesitan sin asignar memoria adicional a datos irrelevantes.

Esta base de datos intenta resolver las deficiencias de los valores clave y los almacenes de documentos, pero dado que puede ser un sistema más complejo de administrar, no se recomienda su uso para equipos y proyectos más nuevos. Apache HBase y Apache Cassandra son ejemplos de bases de datos de columna ancha de código abierto.

ApacheHBase se basa en el sistema de archivos distribuidos de Hadoop que proporciona una forma de almacenar conjuntos de datos dispersos, que se usa comúnmente en muchas aplicaciones de big data.

ApacheCassandra ([Search | Datastax AbsorbLMS](#)), por otro lado, ha sido diseñado para administrar grandes cantidades de datos en múltiples servidores y agrupaciones que abarcan múltiples centros de datos. Se ha utilizado para una variedad de casos de uso, como sitios web de redes sociales y análisis de datos en tiempo real.

Modelo de datos de Cassandra



Tipos de bases de datos NOSQL

Graph store

Los elementos de datos se almacenan como nodos, aristas y propiedades. Cualquier objeto, lugar o persona puede ser un nodo. Una arista define la relación entre los nodos.

Las bases de datos de grafos se utilizan para almacenar y administrar una red de conexiones entre elementos dentro del grafo.

Neo4j es un servicio de base de datos basado en gráficos basado en Java con una edición comunitaria de código abierto

Comenzamos con el teorema CAP

- Tres características de los sistemas distribuidos:
- Consistencia
- Disponibilidad
- Tolerancia a las particiones



Consistencia

- Consistencia significa que todos los clientes ven los mismos datos al mismo tiempo, sin importar a qué nodo se conecten. Para que esto suceda, cada vez que se escriben datos en un nodo, deben reenviarse o replicarse instantáneamente a todos los demás nodos del sistema antes de que la escritura se considere "correcta".

Disponibilidad

- Disponibilidad significa que cualquier cliente que realice una solicitud de datos obtiene una respuesta, incluso si uno o más nodos están caídos. Otra forma de decir esto: todos los nodos de trabajo en el sistema distribuido devuelven una respuesta válida para cualquier solicitud, sin excepción.

Tolerancia a particiones

- Una partición es una interrupción de las comunicaciones dentro de un sistema distribuido: una conexión perdida o demorada temporalmente entre dos nodos. La tolerancia a la partición significa que el clúster debe continuar funcionando a pesar de cualquier cantidad de fallas en la comunicación entre los nodos del sistema.

CAP y las bases de datos NOSQL

- Las bases de datos NoSQL (no relacionales) son ideales para aplicaciones de red distribuidas. A diferencia de sus contrapartes SQL (relacionales) escalables verticalmente, las bases de datos NoSQL son escalables horizontalmente y están distribuidas por diseño; pueden escalar rápidamente a través de una red en crecimiento que consta de múltiples nodos interconectados.
- Bases de datos CP
- Bases de datos AP
- Bases de datos CA

CP

- Base de datos de CP: una base de datos de CP ofrece consistencia y tolerancia a la partición a expensas de la disponibilidad.
- Cuando se produce una partición entre dos nodos cualesquiera, el sistema tiene que cerrar el nodo no disponible (es decir, hacer que no esté disponible) hasta que se resuelva la partición.

AP

- Base de datos AP: una base de datos AP ofrece disponibilidad y tolerancia a la partición a expensas de la consistencia.
- Cuando se produce una partición, todos los nodos permanecen disponibles, pero los que están en el extremo equivocado de una partición pueden devolver una versión de datos más antigua que otros. (Cuando se resuelve la partición, las bases de datos de AP normalmente vuelven a sincronizar los nodos para reparar todas las inconsistencias en el sistema).

CA

- Base de datos de CA: una base de datos de CA ofrece consistencia y disponibilidad en todos los nodos.
- Sin embargo, no puede hacer esto si hay una partición entre dos nodos en el sistema y, por lo tanto, no puede ofrecer tolerancia a fallas.