

@author: aar.velasquez@gmail.com

@date: 2020/10/19

Capítulo 24: Bases de datos NOSQL y sistemas de almacenamiento de Big Data.

La mayoría de los sistemas NOSQL son bases de datos distribuidas o sistemas de almacenamiento distribuidos, con un enfoque en semis- almacenamiento de datos estructurado, alto rendimiento, disponibilidad, replicación de datos y capacidad en lugar de un énfasis en la coherencia inmediata de los datos, lenguajes de consulta y almacenamiento de datos estructurados.

24.1 Introducción a los sistemas NOSQL

Aparición de sistemas NOSQL

■ Google desarrolló un sistema NOSQL patentado conocido como BigTable , que es utilizado en muchas de las aplicaciones de Google que requieren grandes cantidades de almacenamiento de datos edad, como Gmail, Google Maps e indexación de sitios web. Apache Hbase es un Sistema NOSQL de código abierto basado en conceptos similares. Innovación de Google condujo a la categoría de sistemas NOSQL conocidos como basados en columnas o amplia almacenes de columnas ; a veces también se les conoce como tiendas de familia de columnas .

Amazon desarrolló un sistema NOSQL llamado DynamoDB que está disponible a través de los servicios en la nube de Amazon. Esta innovación llevó a la categoría conocida como almacenes de datos de valor-clave o, a veces, almacenes de datos de tuplas clave o de objetos clave .

Facebook desarrollo un sistema NOSQL llamado **cassandra**

Otras empresas de software comenzaron a desarrollar sus propias soluciones y a crear están disponibles para los usuarios que necesitan estas capacidades, por ejemplo, MongoDB y CouchDB , que se clasifican como sistemas NOSQL basados en documentos o almacenes de documentos . Otra categoría de sistemas NOSQL son los sistemas NOSQL basados en gráficos,

Características de los sistemas NOSQL

1. Escalabilidad: como discutimos en la Sección 23.1.4, hay dos tipos de escalas capacidad en sistemas distribuidos: horizontal y vertical. En los sistemas NOSQL, Generalmente se usa la escalabilidad horizontal
2. Disponibilidad, replicación y consistencia eventual: muchas aplicaciones que utilizan sistemas NOSQL requieren una disponibilidad continua del sistema. Para acompañar Si hace esto, los datos se replican en dos o más nodos en un manual transparente. ner, de modo que si un nodo falla, los datos todavía están disponibles en otros nodos.
3. Modelos de replicación:

se utilizan dos modelos de replicación principales en el sistema NOSQL. tems: réplica maestro-esclavo y maestro-maestro.

Replicación maestro-esclavo.

requiere una copia para ser la copia maestra; todas las operaciones de escritura deben aplicarse a la copia maestra y luego se propaga a las copias esclavas, generalmente usando consistencia eventual (las copias esclavas eventualmente serán las mismas que las mas-ter copia). Para leer, el paradigma maestro-esclavo se puede configurar en varios formas.

La replicación maestro-maestro:

permite lecturas y escrituras en cualquiera de las réplicas, pero puede no garantizar que las lecturas en los nodos que almacenan copias diferentes vean el mismos valores.

4. Fragmentación de archivos:

en muchas aplicaciones NOSQL, archivos (o colecciones de datos objetos) pueden tener muchos millones de registros (o documentos u objetos), y miles de usuarios pueden acceder simultáneamente a estos registros. Entonces no es práctico para almacenar todo el archivo en un nodo.

5. Acceso a datos de alto rendimiento:

en muchas aplicaciones NOSQL, es necesario sario para encontrar registros u objetos individuales (elementos de datos) de entre los leones de registros de datos u objetos en un archivo.

Características de NOSQL relacionadas con modelos de datos y lenguajes de consulta.

1. No requerir un esquema : la flexibilidad de no requerir un esquema es logrado en muchos sistemas NOSQL al permitir semi-estructurado, auto- descripción de datos (consulte la Sección 13.1).
2. Lenguajes de consulta menos potentes: muchas aplicaciones que utilizan el sistema NOSQL Es posible que los tems no requieran un lenguaje de consulta potente como SQL
3. Control de versiones: algunos sistemas NOSQL proporcionan almacenamiento de varias versiones de los elementos de datos, con las marcas de tiempo de cuando se creó la versión de datos.

Categorías de sistemas NOSQL

1. Sistemas NOSQL basados en documentos: estos sistemas almacenan datos en forma de documentos que utilizan formatos conocidos, como JSON (Objeto JavaScript Notación). Los documentos son

accesibles a través de su ID de documento, pero también se pueden acceder rápidamente utilizando otros índices.

2. Almacenes de clave-valor NOSQL: estos sistemas tienen un modelo de datos simple basado en el acceso rápido por la clave al valor asociado con la clave; el valor puede ser un registro o un objeto o un documento o incluso tener datos más complejos estructura.
3. Sistemas NOSQL de columna ancha o basados en columnas: estos sistemas dividen una tabla por columna en familias de columnas (una forma de partición vertical; consulte Sección 23.2), donde cada familia de columnas se almacena en sus propios archivos. Ellos también permiten el versionado de valores de datos.
4. Sistemas NOSQL basados en gráficos: los datos se representan como gráficos y los nodos se pueden encontrar atravesando los bordes utilizando expresiones de ruta.
5. Sistemas NOSQL híbridos: estos sistemas tienen características de dos o más de las cuatro categorías anteriores.
6. Bases de datos de objetos: estos sistemas se discutieron en el Capítulo 12.
7. Bases de datos XML: discutimos XML en el Capítulo 13.