BDA 4-5

1 Hello everyone, I am Haiying Che, from Institute of Data Science and knowledge Engineering

School of Computer Science, in Beijing Institute of Technology , from this session ,

we start to learn Data storing system, and in this session we discuss about Four main types NoSQL Database .

2

NoSQL Databases are mainly categorized into four types: Key-value pair, Column-oriented, Graph-based and Document-oriented.

Every category has its unique attributes and limitations. None of the above-specified database is better to solve all the problems.

Users should select the database based on their product needs.

3

The key-value database is based on key-value pairs (key, value) to store and query data.

Data is stored in key/value pairs. It is designed in such a way to handle lots of data and heavy load.

The basic idea is that the data value is queried by the key, the key can be a string type, and the value can be any type of data, such as integer, character, array, list, collection, JSON, BLOB(Binary Large Objects)etc.

The underlying structure can use the hash table to index and manage the keys to support fast query.

But the key-value database does not support queries based on data values

It is one of the most basic NoSQL database example.

This kind of NoSQL database is used as a collection, dictionaries, associative arrays, etc.

Key value stores help the developer to store schema-less data. They work best for shopping cart contents.

Advantages of key-value database is Good scalability, flexibility, and high performance when writing in large numbers.

And disadvantage of key-value database is it Can't store structured information and conditional query is inefficient

4

Column-oriented databases work on columns and are based on BigTable paper by Google.

The column storage database adopts the column storage structure, which is also called DSM (decomposition storage model),

which means that the data is retrieved, stored and managed based on the range (column).

Column storage still uses keys, but the keys point to columns.

The column storage database supports high compression ratios.

The query on data fields under large-scale data is very efficient, but it is not suitable for real-time deletion or updating of entire records and does not support data tables. join operation

Every column is treated separately. Values of single column databases are stored contiguously.

Column based NoSQL database deliver high performance on aggregation queries like SUM, COUNT, AVG, MIN etc. as the data is readily available in a column.

Column-based NoSQL databases are widely used to manage data warehouses, business intelligence, CRM, Library card catalogs.

Advantages of column storage database is Fast search speed, strong scalability, easy distributed extension, low complexity.

And disadvantage of column storage database is Fewer features, most do not support strong transaction consistency .

5

Document-Oriented NoSQL DB stores and retrieves data as a key value pair but the value part is stored as a document.

The document is stored in JSON or XML formats. The value is understood by the DB and can be queried.

The document database is organized around a series of semantically self-contained documents to organize data management.

Documents have no schema, which means that documents do not require a specific structure. which make it flexible.

A document database is actually a collection of documents. A document is actually a data record.

This record can "self-describe" the type and content of the data contained.

XML documents, HTML documents and JSON documents fall into this category.

The data record contained in each document is a collection of a series of data items.

Each data item has a name and a corresponding value. The value can be a simple data type such as a string, a number, and a date, or it can be a complex type, such as ordered lists and associated objects.

Document databases are particularly suitable for managing document-oriented data or similar semi-structured data, such as websites with many read and write operations in the background,

applications that use JSON data structures, or use nested structured non-standardized data

Relational Vs. Document

In this diagram on your left you can see we have rows and columns, and in the right, we have a document database which has a similar structure to JSON.

Now for the relational database, you must know what columns you have and so on.

However, for a document database, you have data store like JSON object.

The document type is mostly used for CMS content management systems, blogging platforms, real-time analytics & e-commerce applications.

It should not use for complex transactions which require multiple operations or queries against varying aggregate structures.

Advantages of Document DB is Good performance (high concurrency), high flexibility, low complexity, and flexible data structure. Provides embedded document function to store frequently queried data in the same document. Indexes can be built based on either keys or content.

And disadvantage is Lack of uniform query syntax.

6

Graph database

Describing social relations and other data as points (vertex) and edges (Edge) and their properties, each graph can be regarded as a structured data.

The dots here represent entities, such as persons, companies, accounts or any other data items, similar to data records in a relational database or files in a document database;

"edges" represent the relationship between points; "attributes" are It is the point-related feature that the user pays attention to.

Relational databases imply the association relationship between different data in the data range .

while in the graphic database, This relationship is directly defined and stored in the database,

Traversing relationship is fast as they are already captured into the DB, and there is no need to calculate them.

so the graph database is good at handling highly related data, suitable for social networking, pattern recognition, dependency analysis, recommendation systems, and path finding, and other problems that can be expressed as relationship graphs.

Advantages of graph database is High flexibility, supporting for complex graphics algorithms, and it can be used to build complex relationship map

And disadvantage of graph database is High complexity, can only support a certain data size .

7

Document database includes mongoDB，CouchDB etc.

Graph database includes Neo4j, infiniteGraph etc.

Column database includes Redis, DynamoDB etc.

Key value database includes Hbase Cassandra, etc.

8 In this session we learned Four main types NoSQL Database.

thank you for your attention, if you have any question, feel free to contact me.