

data

by Shizhao Zhang

Submission date: 18-Jan-2023 04:27PM (UTC+0000)

Submission ID: 195498132

File name: 39084_Shizhao_Zhang_data_474125_2013692637.doc (144K)

Word count: 2536

Character count: 13912

Abstract

The main content of this article is to compare different types of NoSQL databases, understand the use proportion of different types of NoSQL databases, and then compare NoSQL with traditional relational databases to determine when to use the most suitable database. Finally, a database supporting JSON is established, including four tables. The database can basically meet the company's requirements of Abertawa, focusing on users and their assets as well as the transaction of books.

Key word: different types of NoSQL databases, JSON

Content

Abstract	1
I. Different types of NoSQL databases	3
1.1 Document-Oriented DB	3
1.2 Key-Value DB	3
1.4 Wide Column Store DB	4
II. NoSQL and Relational Database	4
2.1 Relational Database	5
2.2 NoSQL	5
2.3 Use of different databases	6
III. For Abertawe Entertainment	6
IV. Summary	8
Reference:	9

I. Different types of NoSQL databases

1.1 Document-Oriented DB

The biggest feature of Document-Oriented database is that data is stored in the form of documents. Each document has its own data unit, and it is a collection for storing a series of similar data. Each data will also have a corresponding value, which may be a complex type such as association object and sequence table, or a very simple type such as date, time, number and string. Document is the smallest unit in data storage. Its attributes in the same table are not necessarily the same, and data storage methods include JSON, XML, and so on.

Its main databases include CouchDB and MongoDB.

Many companies are using these databases. The most famous one is NBC News, which uses RavenDB.

The main scenarios used by Document-Oriented database are:

It is used for enterprises to store logs. In an enterprise, every application will generate logs every day. Document-Oriented database is very suitable because the logs that each application tries to use are also different attributes, and Document-Oriented database does not require that the attributes of documents be consistent, so it can be used to store different logs. The weak schema structure of Document-Oriented database can be used to add new metrics or store different metrics without changing the schema,

The inapplicable situation of Document-Oriented database is that when facing different documents and being required to add transactions, Document-Oriented database does not allow adding transactions.

1.2 Key-Value DB

The key value database is similar to the hash table in c language. It is a data structure that can be accessed directly according to the key value. In other words, it accesses records by mapping the key value to a location in the table to speed up the search. Deleting, querying or adding data will be easier because of the key value. Its main purpose is to facilitate the access of the primary key and make a series of operations more convenient.

Among them, the products that mainly use Key-Value database are Amazon's Dynamo, Riak and Memcached.

And now several well-known companies are also using Key-Value database, among which Github is using Riak, Instagram and StackOverflow are using Redis, and Twitter and YouTube are using Memcached. The above companies have a market value of more than \$1 billion and are using Key-Value database, which shows its excellent performance.

In fact, there are many scenarios. In most cases, if you choose to use the Key-Value database as long as the information is related to the sequence (key), it will be easy for you to list your operations in the shopping cart and conversation, etc.

There are also scenarios that do not apply: for example, key-value does not support rollback,

which makes it easy to lose data in the event of a failure, and it is also a relatively bad choice to use the Key-Value database instead of querying information through the key value. The database itself does not have the function of value query, and the Key-Value database cannot perform data association through multiple keys. In these cases, the Key-Value database is not suitable.

1.3 Graph-Oriented DB

Like the name of image database, its storage method is to convert data into images for storage. The vertex represents the entity, and the edge represents the connection between each entity. For example, z, w and x are three entities. If there is a connection between them, each entity point will have two edges to connect them. The main products of the graph data include OrientDB and Infinite Graph, etc. The more famous companies are Adobe which uses Neo4J and T-Mobile which also uses Neo4J.

The image database is applicable to the image database if the main emphasis of some data is relational. The image database is not suitable for data model. Due to the limitations of the image database itself, it is difficult to spread itself to the whole image, and the scope of application is small.

1.4 Wide Column Store DB

The column storage database, as its name implies, stores data in the column family. In this way, the data of the relevant column family can be queried together. For example, for a person, his age, name and gender are usually in a column family. When you query the person class, you can see the information in a column family, but what books he bought will not be in the column family.

Its famous products are HBase and Cassandra. Many famous companies are applying these products, such as Instagram and Ebay, and NASA are using Cassandra.

It is suitable for blogs. Because we have many column families to store information, including one category, one label, and others. Log is also very applicable. You can write information to your own column family in each application.

The inapplicable scenario is model design. We can't know the query method before the schema design. That is, we can't determine what information will be placed in what column family. Once the query information changes, the column family will also change. Therefore, model design is not suitable for column storage databases.

II. NoSQL and Relational Database

2.1 Relational Database

Relational database is a database management system that uses relational schema to store data.

A table is the most basic data organization unit in a relational database. Each table contains several data fields and columns. The columns represent the data properties of the data, and the rows represent the data instances..

In a relational database, tables can be linked by a relationship, that is, the values of columns in one table are the same as those in other tables. Therefore, a relationship between tables and tables that can be queried with SQL statements is established.

At present, the main functions of relational databases include Oracle, PostgreSQL, SQL Server, MySQL, etc. These systems have efficient, stable and safe data storage and management functions, and support SQL for data query and management.

2.2 NoSQL

NoSQL means "not only SQL", and its full name is "Not Only SQL". This is an unrelated and unstructured database management system.

NoSQL database has no fixed format and format compared with traditional relational databases such as MySQL and Oracle. Instead, they use unstructured and unrelated data models, such as key-value pairs, documents, charts, etc.

NoSQL database is used to handle large-scale, high-concurrency, high-availability data storage, analysis, social networks, games, Internet of Things and other fields.

Common NoSQL databases include Couchbase, MongoDB, Redis, Cassandra, etc

The design idea of NoSQL database is to distribute data to a large number of computers, thus making the data model and system structure simpler, thus improving the performance and scalability of the system.

There are mainly the following types of data models:

Key-Value Store: stores data in key-value pairs, and uses keys to quickly query data. The commonly used key values are Redis and Riak.

File storage: store data in a file containing several fields and values. Common file databases are MongoDB and Couchbase

Column storage: stores data in a column, which contains several columns and values. The commonly used column databases are Cassandra and Hbase.

Graph storage: store data in a graph. Each node and edge has an attribute. Common chart databases include Neo4j and OrientDB

NoSQL has the characteristics of high performance, scalability and high availability. They are generally distributed, able to store and read data on multiple computers, and improve availability through load balancing and replication.

However, NoSQL also has some shortcomings, that is, compared with relational databases, its data consistency and transaction processing ability are poor.

2.3 Use of different databases

The approximate rate of all stored data in NoSQL is suitable for the following situations:

The first is to process unstructured data such as text, photos, audio, and web pages,

The second is to process data with very large capacity, such as the reading and analysis of massive data.

Third, data processing in a decentralized environment.

Fourth, data processing in the case of high performance requirements and low latency requirements.

Finally, in general, there are some scenarios where data needs to be processed flexibly and high divergence is required.

In contrast, the application of large probability of relational databases is as follows:

First, it is to process data that can be structured, such as store orders, sorting out customers, etc.

Secondly, there are high requirements for data classification consistency.

Finally, it is necessary to deal with the relationship that is difficult to be sorted out quickly and have higher requirements for dealing with transactions.

The most well-known database among database types is relational database, which is very applicable in many scenarios. Generally speaking, when there is obvious relationship between data, relational database is the most appropriate.

1. When there is a large amount of interrelated data to be stored, such as customer, customer asset information, etc. 2. When it is necessary to optimize and retrieve the data, it is applicable to the relational database 3. When it is necessary to support transaction processing, it can ensure the consistency and integrity of the data 3. When it is necessary to perform a certain degree of management operations and data queries on the SQL language 4. When it is necessary to ensure the relationship between tables.

III. For Abertawe Entertainment

Alberta Entertainment is headquartered in Swansea, but they are now considering how to expand the company and set up branches across Wales. Therefore, it faces the problem of system development. Its system needs to be optimized for performance, such as adding indexes to improve the function of the whole system to meet the needs of branches in the future. In the future, its data will not only exist in a single chip or program, but will be scattered in various

branches. We will establish a distributed database, so that all branches can operate data more conveniently and transparently in all places, so that users can query data at any time and place, and the database distribution is very transparent.

In order to improve the query efficiency, I set the following two points

-- Create an index on the title column of assets table

```
CREATE INDEX assets_title_index ON assets (title);
```

-- Create an index on the genre column of assets table

```
CREATE INDEX assets_genre_index ON assets (genre);
```

In order to speed up our search for personal assets, we have created an index of assets -- title. In order to facilitate our understanding of the subject matter of books, we have created an index of gene.

Insert ten examples into the icon:

```
INSERT INTO users (name, address, telephone_number, date_joined)
```

```
VALUES ('John Smith', '123 Main St, Swansea, UK', '555-555-5555', '2022-01-01'),  
      ('Jane Doe', '456 Park Ave, Swansea, UK', '555-555-5556', '2022-01-02'),  
      ('Bob Johnson', '789 Elm St, Swansea, UK', '555-555-5557', '2022-01-03'),  
      ('Emily Davis', '321 Oak St, Swansea, UK', '555-555-5558', '2022-01-04'),  
      ('Michael Brown', '159 Pine St, Swansea, UK', '555-555-5559', '2022-01-05'),  
      ('Jessica Wilson', '753 Cedar St, Swansea, UK', '555-555-5560', '2022-01-06'),  
      ('Matthew Moore', '951 Birch St, Swansea, UK', '555-555-5561', '2022-01-07'),  
      ('Daniel Taylor', '135 Maple St, Swansea, UK', '555-555-5562', '2022-01-08'),  
      ('David Anderson', '246 Elm St, Swansea, UK', '555-555-5563', '2022-01-09'),  
      ('Maria Garcia', '369 Oak St, Swansea, UK', '555-555-5564', '2022-01-10');
```

```
INSERT INTO assets (title, author, genre, publisher, release_date, price, stock_amount, blob)
```

```
VALUES ('War and Peace', 'Leo Tolstoy', 'Historical Fiction', 'CymruBooks', '2022-05-01', 19.99,  
100, null),  
      ('Pride and Prejudice', 'Jane Austen', 'Romantic Fiction', 'CymruBooks', '2022-03-01',  
14.99, 50, null),  
      ('The Odyssey', 'Homer', 'Epic Poem', 'CymruBooks', '2022-02-01', 12.99, 75, null),  
      ('In Cold Blood', 'Truman Capote', 'Non-Fiction', 'CymruBooks', '2022-01-01', 9.99, 25,  
null),  
      ('The Great Gatsby', 'F. Scott Fitzgerald', 'Tragedy', 'CymruBooks', '2022-04-01', 17.99,  
100, null),  
      ('The Catcher in the Rye', 'J.D. Salinger', 'Coming-of-Age', 'CymruBooks', '2022-06-01',  
12.99, 50, null),  
      ('The Lord of the Rings', 'J.R.R. Tolkien', 'Fantasy', 'CymruBooks', '2022-07-01', 19.99, 75,  
null),  
      ('The Hitchhiker Guide to the Galaxy', 'Douglas Adams', 'Science Fiction', 'CymruBooks',  
'2022-08-01', 14.99, 100, null),  
      ('The Diary of a Young Girl', 'Anne Frank', 'Autobiography', 'CymruBooks', '2022-09-01',
```

11.99, 50, null),
('Moby-Dick', 'Herman Melville', 'Adventure Fiction', 'CymruBooks', '2022-10-01', 17.99,
75, null));
Shown below:

asset_id	title	author	genre	publisher	release_date	price	stock_amount	blob
1	War and Peace	Leo Tolstoy	Historical Fiction	CymruBooks	2022-05-01 00:00:00	19.99	100	(Null)
2	Pride and Prejudice	Jane Austen	Romantic Fiction	CymruBooks	2022-03-01 00:00:00	14.99	50	(Null)
3	The Odyssey	Homer	Epic Poem	CymruBooks	2022-02-01 00:00:00	12.99	75	(Null)
4	In Cold Blood	Truman Capote	Non-Fiction	CymruBooks	2022-01-01 00:00:00	9.99	25	(Null)
5	The Great Gatsby	F. Scott Fitzgerald	Tragedy	CymruBooks	2022-04-01 00:00:00	17.99	100	(Null)
6	The Catcher in the Rye	J.D. Salinger	Coming-of-Age	CymruBooks	2022-06-01 00:00:00	12.99	50	(Null)
7	The Lord of the Rings	J.R.R. Tolkien	Fantasy	CymruBooks	2022-07-01 00:00:00	19.99	75	(Null)
8	The Hitchhiker Guide to	Douglas Adams	Science Fiction	CymruBooks	2022-08-01 00:00:00	14.99	100	(Null)
9	The Diary of a Young Gir	Anne Frank	Autobiography	CymruBooks	2022-09-01 00:00:00	11.99	50	(Null)
10	Moby-Dick	Herman Melville	Adventure Fiction	CymruBooks	2022-10-01 00:00:00	17.99	75	(Null)

user_id	name	address	telephone_number	date_joined	company_id
1	John Smith	123 Main St, Swansea, U	555-555-5555	2022-01-01 00:00:00	(Null)
2	Jane Doe	456 Park Ave, Swansea, I	555-555-5556	2022-01-02 00:00:00	(Null)
3	Bob Johnson	789 Elm St, Swansea, UK	555-555-5557	2022-01-03 00:00:00	(Null)
4	Emily Davis	321 Oak St, Swansea, UK	555-555-5558	2022-01-04 00:00:00	(Null)
5	Michael Brown	159 Pine St, Swansea, U	555-555-5559	2022-01-05 00:00:00	(Null)
6	Jessica Wilson	753 Cedar St, Swansea, I	555-555-5560	2022-01-06 00:00:00	(Null)
7	Matthew Moore	951 Birch St, Swansea, U	555-555-5561	2022-01-07 00:00:00	(Null)
8	Daniel Taylor	135 Maple St, Swansea,	555-555-5562	2022-01-08 00:00:00	(Null)
9	David Anderson	246 Elm St, Swansea, UK	555-555-5563	2022-01-09 00:00:00	(Null)
10	Maria Garcia	369 Oak St, Swansea, UK	555-555-5564	2022-01-10 00:00:00	(Null)

According to the situation of the company, we can use the corresponding select statement to help us speed up the search of data. For example, we can use it if we need to find science fiction

SELECT * FROM assets WHERE genre = 'Non-Fiction';

localhost
test1121
public
运行
停止
解释

```

1 SELECT * FROM assets WHERE genre = 'Non-Fiction';
2 |

```

信息
摘要
结果 1

asset_id	title	author	genre	publisher	release_date	price	stock_amount	blob
4	In Cold Blood	Truman Capote	Non-Fiction	CymruBooks	2022-01-01 00:00:00	9.99	25	(Null)

IV. Summary

This time I met Abertawa enterprise, which is very suitable for distributed database. Distributed database is a data storage system based on multiple nodes. The amount of data in this event is large, and a single computer cannot process it at the same time. The distributed database can

store data on multiple computers, thus increasing the storage capacity of data.

In addition, there is a large amount of parallel access. The distributed database can allocate access requests on multiple computers, thus increasing the system's throughput. "... When the headquarters has high requirements for strong consistency of data in various sub-departments, the distributed database can ensure that the data on multiple computers is always consistent. At the same time, the security requirements of enterprise data are very high. The distributed database can store data on different computers, thus enhancing the security of data. At the same time, the availability of data is also a necessary condition for the company. It can store data on multiple computers, so that local users can access data locally.

Reference:

- [1] Dou Liang, Xu Fei, Gu Lintao. Teaching reform of database system practice course under the background of big data [J]. Computer Education, 2018 (04): 117-120. DOI: 10.16512/j.cnki.jsjy.2018.04.028.
- [2] Jin Cheqing, Qian Weining, Zhou Minqi, Zhou Aoying. Benchmark of data management system evaluation: from traditional database to emerging big data [J]. Journal of Computer Science, 2015,38 (01): 18-34
- [3] Shen Derong, Yu Ge, Wang Xite, Nie Tiezheng, Kou Yue. Research review of NoSQL system supporting big data management [J]. Journal of Software, 2013,24 (08): 1786-1803
- [4] Zhou Silin, Xie Conghua, Zhou Bei. Research on the teaching mode of Database Application Technology based on independent learning [J]. Software Guide, 2013,12 (01): 198-199
- [5] Jiang Yongsheng. Comparative study of online autonomous learning and traditional teaching methods based on college computer basic courses [J]. Computer Knowledge and Technology, 2012,8 (08): 1870-1872
- [6] Xia Ying. Course construction and practice of advanced database system technology for graduate students [J]. Computer Education, 2015 (23): 4-6. DOI: 10.16512/j.cnki.jsjy.2015.23.002
- [7] Liu Bingbing, Meng Xiaofeng, Shi Yingjie. CloudBM: Cloud Data Management System Test

Benchmark [J]. Computer Science and Exploration, 2012,6 (06): 504-512

[8] Wang Jinglei. On the application technology and maintenance of computer database [J]. Science and Technology Perspectives, 2021 (21): 126-127. DOI: 10.19694/j.cnki.issn2095-24572021.21.46

FINAL GRADE

GENERAL COMMENTS

Instructor

65/100

Generic Assessment Criteria	Marks available	Marks awarded
1. Engagement with Literature Skills	20	
Poor engagement with essential reading. No evidence of wider reading. Reliance on inappropriate sources, and/or indiscriminate use of sources. Heavily reliant on information gained through class contact. Inconsistent and weak use of referencing. More credible sources needed, but overall a good amount of references. Would have liked more depth into the different types of NoSQL databases.	10	
2. Knowledge and Understanding Skills	20	

The investigation would have benefitted from examples and real world scenarios and a better definitive conclusion.	10	
3. Cognitive and Intellectual Skills	15	
An interesting approach to the design of the NoSQL implementation, I wonder why you chose to nest each entity rather than a separate collection for each. An issue that could come from this is a document is limited to around 100mb so its possible the document would become too big, the structure of each entity would be fine to be distributed into individual collections Both schemas are applicable	10	
4. Practical Application Skills	35	
Good clean code		

	25	
5. Transferable Skills for Life and Professional Practice	10	
No cover page or introduction or page numbers Use of first person is to be avoided in academic writing	6	

Assessment Mark (Assessment marks are subject to ratification at the Exam Board. These comments and marks are to give feedback on module work and are for guidance only until they are confirmed.)

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7
