

database2

by Yinjie Zhang

Submission date: 06-Jan-2023 11:59AM (UTC+0000)

Submission ID: 193796423

File name: 40673_Yinjie_Zhang_database2_474125_2006813569.docx (547.17K)

Word count: 2510

Character count: 14099



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

Database Systems

by

Yinjie Zhang

Lecturer: Dr Stephen Hole

December 2022

directory

1. introduction	3
2. Comparison of two Nosql databases.....	3
2.1.mongodb	3
2.1.1.Introduction	3
2.1.2. scheme	4
2.2.redis	6
2.2.1.Introduction	6
2.2.2. scheme	7
2.3.compare	7
2. sql database	8
3. Comparison of sql and nosql	9
3.1. Relational database	9
3.2.Nosql	10
3.3. Comparison.....	11
4. Combine sql and nosql.....	11
Reference	13

1.introduction

AE (Abertawe Entertainment), based in Swansea, sells a large range of text, ebook and audio books. Given the company's readiness to expand its business and the volume of transactions, traditional relational databases have become a bottleneck in system performance. A new data volume needs to be customized to handle the high concurrency, so as to prepare for future sec-kill activities.

2. Comparison of two Nosql databases

2. 1. mongodb

2.1.1.Introduction

MongoDB is an open source, high performance, schema-free document database, which is designed to simplify development and facilitate expansion, and is one of the NoSQL database products. Is the non-relational database that most closely resembles a relational database (MySQL). It supports a very loose data structure in a Json-like format called BSON, so it can store complex data types while being quite flexible. A record in MongoDB is a document, which is a data structure composed of fields and value pairs. A MongoDB document is similar to a JSON object in that a document is considered an object. The data type of the field is character, and its values can include other documents, plain arrays, and document arrays in addition to the basic types.

MongoDB has the following application scenarios:

- 1) For social scenes, MongoDB storage is used to store user information and the information of users' published circle of friends, and nearby people and places are realized through geographical location index.
- 2) In game scenarios, MongoDB is used to store game user information. Users' equipment and credits are directly stored in the form of embedded documents, which is convenient for query, efficient storage and access.
- 3) In logistics scenario, the order information is stored in MongoDB, and the order status is constantly updated in the process of transportation. It is stored in the form of MongoDB embedded array, and all the changes of the order can be read out in a query.
- 4) In the Internet of Things scenario, MongoDB is used to store information of all connected intelligent devices and log information reported by devices, and these information is analyzed in multiple dimensions.
- 5) Live video, use MongoDB to store user information, like interactive information, etc.

In conclusion, the common features of data manipulation in these application

scenarios are:

- 1) Large amount of data.
- 2) Frequent writes (both reads and writes are frequent).
- 3) Low value data, transactional requirements are not high.

In the selection of architecture, in addition to the above three characteristics, if you still hesitate to choose it? Some questions to consider are: ① Applications do not require transaction and complex join support ② New applications, requirements change, data models are uncertain, and rapid iterative development is desired. ③ Applications require read/write QPS of 2000-3000 + (higher is acceptable). ④ Applications require TB or even PB data stores. ⑤ Applications develop rapidly and need to be able to scale rapidly horizontally. ⑥ The application requires that the stored data not be lost. ⑦ Application needs a lot of location query, text query.

MongoDB has the following features:

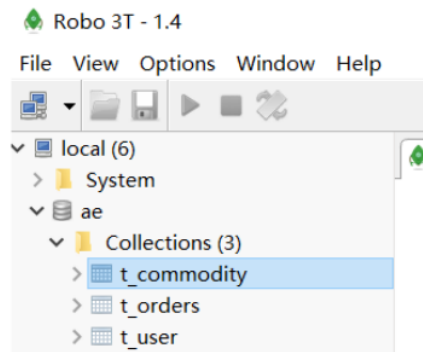
- 1) High performance: MongoDB provides high-performance data persistence. In particular, support for embedded data models reduces I/O activity on database systems. Indexes support faster queries and can contain keys from embedded documents and arrays. (Text index solves the search requirement, TTL index solves the automatic expiration requirement of historical data, and geographical location index can be used to build various O2O applications.) Multiple engines, including mmapv1, wiredtiger, mongorocks (rocksdb), and in-memory, support various scenarios. Gridfs addresses file storage needs.
- 2) High availability: The replica tool of MongoDB is called replica set, which provides automatic failover and data redundancy.
- 3) High scalability: MongoDB provides horizontal scalability as part of its core functionality. Sharding distributes data across a cluster of machines. Since 3.4, MongoDB supports the creation of data regions based on slice keys. In a balanced cluster MongoDB directs reads and writes covered by a zone to only those pieces within that zone.
- 4) Rich query support: MongoDB supports a rich query language that supports read and write operations (CRUDs) such as data aggregation, text search, and geospatial queries.
- 5) Other features: patternless (dynamic mode), flexible document model.

2.1.2. scheme

The main problems AE company is facing now are large sales volume, massive data and high concurrency. Traditional relational databases are difficult to meet, and

MongoDB is a good choice here.

Without considering some complicating factors, I have designed three collections as follows:



Although mongodb is a nosql database, it is the non-relational database that most resembles a relational database. Its collection can be thought of as a table in a relational database.

First, the user table. This form records the user's registration information, as well as the balance of the account.

```
{
  "name" : "tom",
  "address" : "ny",
  "telnum" : "911",
  "datajoin" : "2022-1-1",
  "money" : "1000"
}
```

The second and most crucial table is the order table. High concurrency is mainly a large number of create orders and modify orders. MongoDB is directly stored in memory, which eliminates the io operation with disk and greatly speeds up the storage speed. Besides, MongoDB provides a series of complex query methods, making it able to perform some complex query operations like relational databases, which is very suitable for the current target.

```
{
  "uid" : "63b67618b383e8070e7a7e8a",
  "create_time" : "2023-1-5-15:34",
  "tile" : "math",
  "author" : "yinjie zhang",
  "genre" : "academic",
  "publisher" : "aaa",
  "release_date" : "2030-1-1",
  "price" : "100",
  "stock_amount" : "0.01",
  "blob" : "null",
  "invoice_id" : "123"
}
```

```
}
```

The adaptation records the order id, the user id of the purchase, when the order was created, and some other detailed order messages. The company can look up the assets owned by a user (by indexing the user id) through an order.

Finally, there is the inventory table, which is also a table affected by high concurrency, and users need to deduct inventory for purchases.

```
{
```

```
    "commodity_discription"          :          "Math,author:yinjie  
zhang,genre:acdamic,publisher:aaa,release_date:2030-1-1,price:100",
```

```
    "commodity_flag" : "true",
```

```
    "commodity_num" : "1000",
```

```
    "commodity_picture" : "D:picture"
```

```
}
```

The table mainly records the description of the product information, whether it is on the shelf, the stock and the address of the picture. It is mainly used for front - end page display and post - state management.

2. 2. redis

2.2.1.Introduction

Redis is a key - value NoSql database. Key-value means that data stored in Redis is stored in the form of key.value pairs, and value can be in various forms, including strings. Values. Even json.

Redis has the following characteristics:

- 1) key-value: value supports multiple data structures and provides rich functions
- 2) Single-threaded, each command is atomic
- 3) Low latency, fast (memory based. IO multiplexing. Good coding).
- 4) Support for data persistence
- 5) Supports primary and secondary clusters
- 6) Supports multi-language clients

Redis is a key-value database. key is usually String, but there are various types of value:

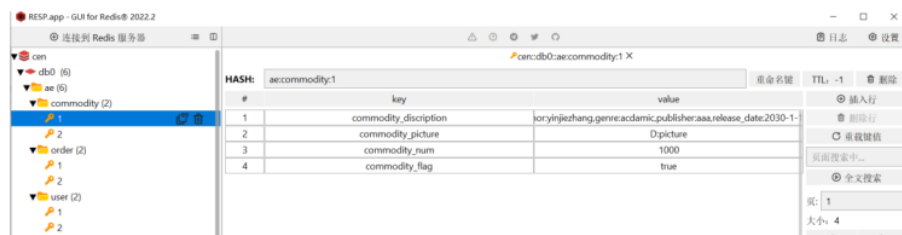
String	hello world
Hash	{name: "Jack", age: 21}
List	[A -> B -> C -> C]
Set	{A, B, C}
SortedSet	{A: 1, B: 2, C: 3}
GEO	{A: (120.3, 30.5)}
BitMap	0110110101110101011
HyperLog	0110110101110101011

2.2.2. scheme

Here I have redis installed in Linux. First log in to redis from the console.

```
root@hadoop100 ~]# redis-cli
127.0.0.1:6379> auth zyj990127
OK
127.0.0.1:6379> ping
PONG
127.0.0.1:6379>
```

Redis doesn't have tables like MySQL or collections like mongodb. We can distinguish this by prefixing the key. As shown below, there are three keys: ae:user:id, ae:order:id and ae:commodity:id.



The rest of the design is not much different from the mongodb design above.

2. 3. compare

1) Memory management: All Redis data is stored in memory and written to disk regularly. When the memory is insufficient, the specified LRU algorithm can be selected to delete the data. MongoDB data is preferentially stored in the memory. When the memory is insufficient, only hot data is stored in the memory and other data is stored on disks.

2) Data structure: Redis supports a variety of data structures, including hash, set, and list. MongoDB data structure is relatively simple, but it supports rich data expression, indexes, most similar to relational databases, and supports very rich query languages.

3) Data volume and performance: When physical memory is sufficient, performance, redis>mongodb>mysql. Data volume, mysql>mongodb> redis.

4) reliability: Since version 1.8, mongodb uses binlog (MySQL also uses the binlog mode) to support persistence, improving reliability. Redis relies on snapshots for persistence; AOF enhances reliability; While improving reliability, the access performance is affected. MongoDB is better than Redis in reliability.

In conclusion, redis is better for cache. For example, user tables, user assets, inventory, etc. are stored in Mysql. When the project is started, redis cache is carried out on key hot commodities and so on for hot start. The order is stored in redis when it is created and then synchronized to mysql at free time. In practice, a web project is usually composed of multiple caches, local caches, nginx, redis, and so on. redis is commonly used for cache.

MongoDB is the most appropriate for this application. The amount of data in this application is large, but it does not exceed the upper limit required for MongoDB. Moreover, this application requires some complex queries, and MongoDB is very suitable for it. If you need to join the sec-kill service in the future, you can use redis for cache.

2.sql database

Traditional database solution, here I choose MySQL database. Without complicated considerations, I set up four tables.

t_user				
Uid	int	0	Primary key	
Name	varchar	255		
Address	varchar	255		
TelNum	varchar	13		Telephone number
DateJoin	datetime	0		
Money	varchar	255		

t_order_master				
Order_id	int	0	Primary key	Order id
Order_sn	bigint	0		Order number
Uid	int	0		User number

Pay_method	tinyint	0		Payment method: 1 Alipay, 2 wechat and 3 bank cards
Create_time	datetime	0		Order creation time
Invoice_id	varchar	255		Invoice title

t_order_detail

Order_sn	bigint	255	Primary key	Order number
title	varchar	255		
author	varchar	255		
genre	varchar	255		
publisher	varchar	255		
Release_date	datetime	0		
pirce	varchar	255		
Stock_amount	varchar	255		
blob	varchar	255		

t_commodity

commodity_id	int	0	Primary key	Inventory id
commodity_description	varchar	255		Product description
commodity_flag	tinyint	0		If 1 hits the shelves, 0 drops
commodity_num	int	0		inventory
commodity_picture	varchar	255		Product picture address

3.Comparison of sql and nosql

3.1. Relational database

Relational database and nosql database have their own usage, and the difference is relatively large. Different businesses use different databases, either relational or nosql. Only by fully understanding the characteristics of both and performing their respective duties can the advantages of database management be better played. Some of the advantages and disadvantages of relational databases and the advantages and disadvantages of nosql data are described below.

Advantages:

- 1) Easy maintenance: relational database has a table structure, similar to matrix, unified format, easy to manage. SQL statements are more convenient to operate.
- 2) sql statement easy to get started: For computer students, the first database contact is relational database and sql language. Because sql language is easy to understand and learn, you can carry out some complex additions and deletions.
- 3) Capable of complex CRUD: multiple tables can be linked for queries, and there is a transaction mechanism with high data security.

Disadvantages:

- 1) When large amounts of data are processed efficiently. Because relational databases perform disk io operations, read and write speeds are slow.
- 2) Once the table structure of the relational database is determined, it is difficult to modify later, and the flexibility is relatively low. The table needs to be carefully designed at the beginning. It is best not to modify the table structure in the project.
- 3) Similar to the first rule, the relational database is not applicable to high-concurrency scenarios due to disk I/O limitations, resulting in poor performance.

3. 2. Nosql

Advantages:

- 1) More flexible data storage: Compared to the fixed table structure of traditional relational databases, nosql databases have more flexible storage methods. NoSql, on the other hand, has no strict constraints on the database format and is often loose and free in form. It can be key-value, document, or even graph format.
- 2) High scalability: Because of the fixed table structure, it is difficult to fragment data in relational database, and thus cannot do horizontal expansion of data. However, NoSQL databases can scale horizontally and can store a certain amount of different types of data.
- 3) Support the ultimate consistency of data: NoSQL database is based on the CAP model, but also support the BASE distributed theory, the purpose is to achieve the ultimate consistency of data.

Disadvantages:

- 1) Cannot use SQL syntax: SQL syntax is familiar to many beginners. And different non-relational database query syntax is very different, all kinds of miscellaneous, so that the learning cost is larger.
- 2) Lack of feature support: Most nosql data has no transaction mechanism, and the new version of mongodb transaction is not a true transaction mechanism. Data security and final consistency requirements are lower.
- 3) Most NoSQL databases are still in their infancy: whether it is nosql database books, online courses, or technical documents, the number of users and so on are not as good as relational databases. A large number of undergraduates are not familiar with nosql databases.

3.3. Comparison

Looking at the advantages and disadvantages of appeals, relational and non-relational application scenarios can be obtained.

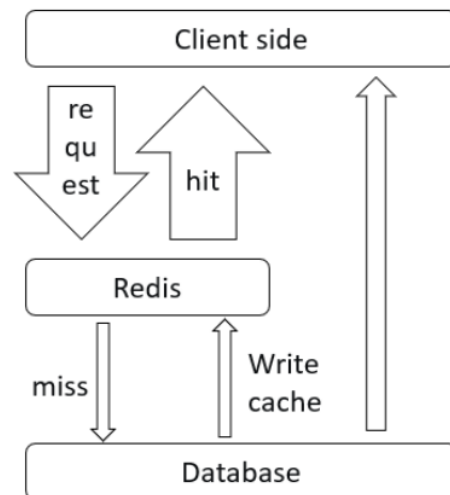
Relational database is suitable for fixed data structure, huge amount of data, CRUD(add, delete, update and check) more frequently, and high requirements for data security.

One of the advantages of Nosql databases is that they are fast. It is suitable for high concurrency and frequent deletion, of course, it also comes with some disadvantages, such as no transaction mechanism and so on.

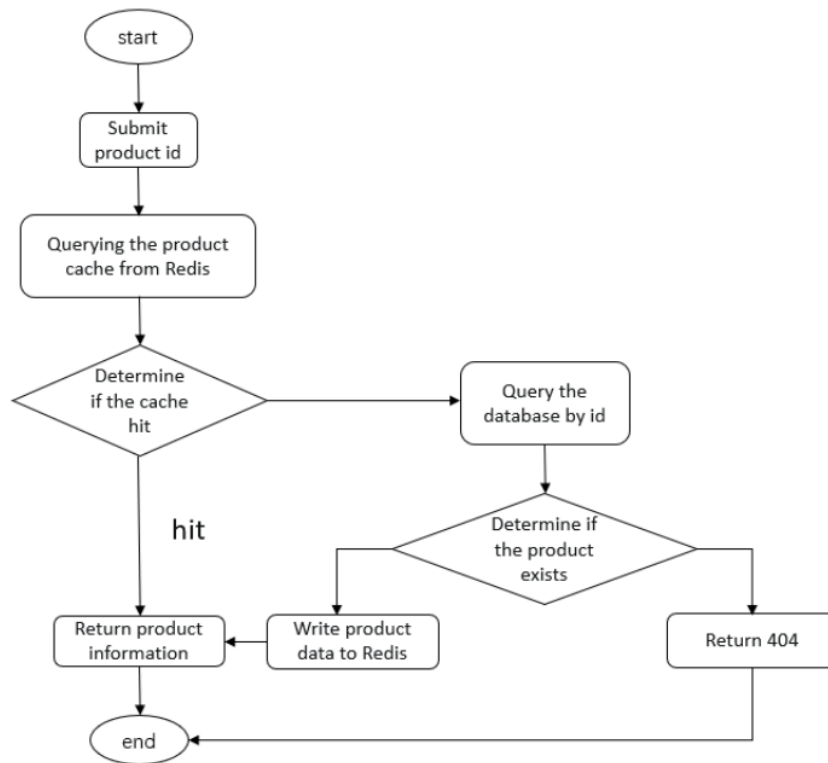
4.Combine sql and nosql

In practice, you don't use a single database. Usually relational databases are used together with nosql data. Let's take MySQL and redis as an example.

When we query the merchant information, we are directly operating the query from the database. Querying the database directly is slow due to the on-disk io, so we need to add a cache. The approximate flow of adding cache via redis is as follows:



We can use this idea to query for product information. This is done as shown in the following figure:



Only by giving full play to the advantages of relational data and the advantages of nosql database, can the performance of the whole project be the best.

Reference

- [1] Makris Antonios; Tserpes Konstantinos; Spiliopoulos Giannis; Zisis Dimitrios; Anagnostopoulos Dimosthenis. MongoDB Vs PostgreSQL: A comparative study on performance aspects[J]. Geoinformatica, 2021, Vol. 25(2): 243-268
- [2] Muñoz-Sánchez, Pablo D.; Fernández Candel, Carlos Javier; García-Molina, Jesús Joaquín; Sevilla Ruiz, Diego. Managing Physical Schemas in MongoDB Stores[J]. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2020, Vol. 12584: 162-172
- [3] William Schultz; Tess Avitabile; Alyson Cabral. Tunable Consistency in MongoDB[J]. Proceedings of the VLDB Endowment, 2019, Vol. 12(12)
- [4] Dirk Eddelbuettel. Redis for Market Monitoring[J]. 2022.
- [5] A High Performance Memory Key-Value Database Based on Redis. A High Performance Memory Key-Value Database Based on Redis[J]. Journal of Computers, 2019, Vol. 14(3): 170-183
- [6] Bagga, Simmi; Sharma, Anil. A Comparative Study of NoSQL Databases[J]. Lecture Notes in Electrical Engineering, 2021, Vol. 701: 51-61
- [7] Sambuu, Uyanga; Namsraidorj, Munkhtsetseg; Batbayar, S.; Bat-Ulzii, Sh. A Comparative Study of NoSQL and Relational Database[J]. Smart Innovation, Systems and Technologies, 2021, Vol. 212: 116-122

FINAL GRADE

GENERAL COMMENTS

Instructor

53/100

Generic Assessment Criteria	Marks available	Marks awarded
1. Engagement with Literature Skills	20	9
You have a Reference section, however you are not using it in the main body of the text through the use of citation. Use references and citations to support your text and demonstrate ownership of the claims being made. This is a must for a Masters' level assessment. Some good work on the document schema in the report. Relate all the NoSQL database to the Abertawe Entertains case study. This will make the report less generic and more coherent.		
2. Knowledge and Understanding Skills	20	11
Knowledge has a reasonably well-defined focus, which is reasonably extensive, coherent and		

detailed, with a critical awareness of current problems. A good document schema has been supplied in the coding files. You could have included an Entity Relationship Diagram for the relational schema.		
3. Cognitive and Intellectual Skills	15	8
Proposes adequate but limited new schemas for the relational and document based database systems.		
4. Practical Application Skills	35	20
You have implemented and deployed in both database models. The document implementation is applicable to mostly meet the requirements of the case study. A reasonable relational database implementation but without Functions, Views, Procedures, Function, Triggers, Foreign Keys, Constraints and Sequences.		
5. Transferable Skills for Life and Professional Practice	10	5
Mostly coherent, organised work, in a suitable structure and is for the most part clearly expressed. Provide more depth in your discussion and reference the whole of the document.		

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

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13
