Massive Concurrent Data Processing Technology in the Web System
Fen ZHOU and Yu-rong GUAN
Department of Computer Science, HuangGang Normal University,
HuangGang, Hubei 438000 China

Keywords: Web system, Concurrent data, Performance optimization.

Abstract. Starting with the high concurrent, massive data requirements in the web system, the analysis of the system structure of the cache technology, module processing technology, thread technology and the appropriate application of indexing technology to solve the large number of concurrent data real-time processing problems. To meet the high load, high concurrency, large amount of data to improve the overall system performance requirements for the effective solution to large quantities of concurrent data processing provides a convenience.

Introduction
At present, the major sites of electricity spike and snapped; for us, are no strangers. However, from a technical point of view, this is a huge test for the web system. When a web system receives tens of thousands or even more requests within a second, system optimization and stability are critical. Therefore, in the design of Web systems, we must consider in the case of high concurrent data volume, can reduce the time of data query and improve query efficiency. With the rapid increase in information data collection station, together with the information data collection site must be synchronized to upload information, a large number of data surges need to be quickly processed. So the design of a fast and accurate data processing technology has become an imminent problem.

Key Technologies

Data Caching Technology
Data caching technology is used to improve the performance. It will access the high frequency of data or high cost of data stored in memory. The core module of the system can fully reflect the overall performance of the system. Caching technology solves the following problems:

Performance: the corresponding data is stored to avoid duplication of data creation, processing and transmission, can effectively improve performance.

Stability: multiple requests for the same data, logical functions, and user interface are common.

Availability: sometimes, the provision of data information services may be unexpectedly stopped, this time the cache technology can be in a certain period of time can still provide normal support for end users, improve the system availability.

Treading Technology
Multi-threaded concurrent improves the concurrent processing capabilities, such as the server asynchronous processing or the user registration. It needs to add points to the user, etc. These tasks may be more time-consuming, so through multi-threaded asynchronous implementation of the general queue with the background thread, such as regular clean up the database Connection pool connection expired session expired cache and so on.

Indexing Technology
Indexing the database design in a Web system requires consideration of the following principles:

Suitable for the establishment of the case of clustering, including columns are often sorted by the order to return a range of data, a small number of different values.
Foreign key column, primary key column. The situation for creating no clustered indexes includes large numbers of different values, frequently updated columns, and frequent modification of index columns.

In the case of small amounts of data, with the aggregate index than the use of the primary key for order by when the speed.

**Web System Design**

In the web system design process, there may be similar to 5w per second high concurrent functional requirements, in the process, the entire Web system encountered a lot of problems and challenges. If the Web system does not do targeted optimization, it will easily fall into the abnormal state. We are now discussing together, optimize the ideas and methods.

**Design a Reasonable Request Interface**

In order to achieve this as fast as possible, the back-end storage of the interface using memory-level operation will be better. Still directly for the storage of MySQL and the like is not appropriate, if there is such a complex business needs, are recommended to use asynchronous write.

![Web Interface Design](image)

**Overload and Overload Protection**

If the system "avalanche", rush to restart the service, cannot solve the problem. The most common phenomenon is that after the start up, immediately hang up. At this time, it is best to refuse the traffic at the entrance level and then reboot. If it is redis / me cache this service is also linked, restart the need to pay attention to "warm", and it may take a long time. If the system is detected full load status, deny the request is also a protective measure. Setting up filtering at the front end is the easiest way, but this is done by the user "what is meant" by the user. More appropriately, the overload protection is set at the CGI entry level and quickly returns the customer's direct request.

**Attack and Defense**

Web system design, when received a "massive" data request, in fact, the water inside is great. Many users, in order to "grab" to the goods, will use the "brush ticket tool" and other types of auxiliary tools to help them send as many requests to the server as possible. There are also some high-level users, making powerful automatic request script. The reason for this approach is also very simple, that is, in the spike and snapped up the request, the number of their own request more, the higher the probability of success. These are the "means of cheating", but there are "offensive" there is "defense."
For the same account, one-time issue of multiple requests for attack. This kind of request in some do not do data security processing system, may also cause another kind of destruction, causes some judgment condition to be bypassed. For example, a simple collection of logic, first determine whether the user involved in the record, if not then to receive success, and finally written to participate in the record. This is a very simple logic, but in the case of high concurrency, there are deep loopholes. Multiple concurrent requests through the load balancing server, assigned to multiple Web servers within the network, they first send a query request to the storage, and then, in a request to successfully write the record of the time difference, the other request to query the results are "not involved in the record". Here, there is a risk that logical judgment is bypassed.

**High Data Security under the Concurrent**

We know that in the multi-threaded write the same file, there will be a "thread security" problem (multiple threads running the same code at the same time, if the results of each run and single-threaded results are the same, the results and expectations The same, is thread safe). If the MySQL database, you can use it comes with the lock mechanism is a good solution to the problem, but in large-scale concurrent scenes, is not recommended to use MySQL. Spike and snapped the scene, there is another problem, that is, "super", if in this regard control inadvertently, will produce too much to send the situation. We have heard of some of the electricity business to engage in snapping activities, buyers successfully photographed, the business does not recognize the order is valid, refused to ship. The problem here, perhaps not necessarily business treacherous, but the system technical level there is the risk of overweight caused.

![Figure 2. Pessimistic lock thinking.](image)

**Summary**

The Internet is the rapid development of the use of Internet services, the more users, the web system in the high-concurrent scenes has become more and more. While the specific technical solutions to our problem may vary widely, the challenges are similar, so the idea of solving the problem is similar. Only continue to technology innovation and improvement to keep up with the trend of social development. The computer software designs more in line with people's needs, more and more powerful, should be our goal. And in the development of software design also advocated the introduction of new social trends such as "people-oriented" to improve the computer software in the community of service value.

**Acknowledgement**

This work was supported by the guidance project of science and technology research plan of Hubei Provincial Department of Education, project number is B2017193.
References
