Analysis of Performance Optimization Principles and Models in Web

Wang Xin

***Abstract*—The Internet high speed development, causes Web the optimized question to be getting more and more prominent, therefore the Web performance optimizes into inevitably. the first principle of Web Performance Optimization is to understand, to know that income will have to pay, and return is diminishing; Simultaneously the probability will decrease Web the performance, and will start from the highest level to optimize obtained biggest. Web Technical models to improve the performance are: sharing costs, high-speed caching, profiles, parallel processing, simplified treatment; Especially critical to the performance of Web database, Such as to analyze and research from the cache, statements, tables, connection pooling, query, index, and several other aspects. Based on this study, given the crucial Web performance optimization recommendations, which improve the performance of Web usage, accelerate the efficient use of Internet has an important significance.**

***Index Terms*—Web performance; optimization; principle; model; Database**

I. Introduction

Now, Web has grown from a novelty into an essential tool to disseminate information. However, with the continuous development of Internet, Web performance problems more seriously, this is more a wealth of information and modern services making. Therefore, we must optimize the performance of the Web. Performance optimization in the Web, there are some general principles applicable to a variety of specific solutions and unified pattern. Following these principles and patterns will be analyzed and summarized.

II. The Principle Of Optimal Performance Web

A. Has obtained, must have loses.

Only after doing research to see if I can improve the performance of a system. But you also risk the risk that: Put a lot of time wasted on research, the last only a can not do anything to improve the performance of the conclusions, especially in the tight financial constraints and a greater risk of the case. You must carefully analyze the difficulty of the problem and possible size of the harvest. If the gains can be large, then it should be determined to find a solution to improve performance, or not worth the time[1].

Manuscript received March 17, 2011

Wang Xin, School of Computer and Communication Engineering, Weifang University, Shandong 261061, China(email: www268@126.com).

B. The understanding is the base element

Enters a house when darkness, will have the possibility to hit hurts itself. If turned on the light to be many easily. The lamp has given you the information, enables you to optimize the path. Optimize Web performance is the same reason. Consideration of issues in the mind the more clearly, the problem easier to solve. By now, instructed your light was along with the workstation, the software, and the router reference manual. Therefore must first do the matter is reads the reference manual and understands it.

When the reference manual has reduces your pain the key, they appear are rare. By changing the setting and measuring performance is feasible, but to figure out why the difference between different settings, and their relationship to other settings and the relationship between subsystems is even more important. This knowledge can be found in the reference manual. Somewhere in the increase in performance may cause performance degradation elsewhere. If you do not know what to pay, then you will not know whether it is worthwhile to modify the system[1].

C. Has obtained must pay

Although the key of optimal performance is to do more with less, but any increase in performance, as little effort to understand and solve this problem, will have to be paid. In some cases, you have to buy new hardware, re-planning system architecture, may also reduce system portability, maintainability, security, reliability, or increase development time. You can let the system bus running day and night to improve server performance, but such systems are also more likely to collapse. To improve performance, you can remove the firewall and any encryption device, but you will also be exposed, more vulnerable to external attack. In fact the optimization of known usage patterns will inevitably hurt some other different usage patterns.

D. The repayment is decreasing progressively

The size of the investment can be seen that when the hardware has been optimized to complete the task. When it comes to optimizing the system, it is usually easy to find the most simple problem and fix it. With the optimization of the system to increase the performance becomes more difficult and more dependent on the particular configuration and usage patterns [2]. When does not match the return on investment and optimize over. Some clues to explain the optimization process has ended, at least so in the following cases:

1) Users no longer aware of any performance improvement.
2) Changed the good programming style for the performance, making the code can not be transplanted, and difficult to maintain.

3) When you consider programs written in assembly language.
4) The total cost per page view than a person employed in accordance with customer's request to call them and fax costs.
5) When you're bored time.

Optimization goal is to change, because the configuration, usage patterns and available components is constantly changing, so impossible to achieve and maintain optimal performance. In the long run, using standard protocols and API to optimize better than the use of proprietary or solution of their own invention. As optimizing the performance of the efforts can be rewarded, while the system between the different generations have portability, and will have a more long-term investment returns.

E. Portability will reduce the performance

Has the conflict existence between the performance and the probability. The best performance only after the optimization of a special environment to get, but portability was defined in many different environments to determine the function. Because they can not optimize all of the circumstances, it is sometimes necessary to choose the performance for portability.

Fully optimized software is limited to a specific platform, because it must take full advantage of the platform's available features, such as: dedicated CPU registers and system calls used. On the other hand, portable software does not take full advantage of the characteristics of a specific platform, because if so there is not portable. At another level, in a particular mode can be used with the optimization method, when the environment changes, performance is likely to reduce rather than boost performance[3].

If the portability of software is not particularly valuable, then the loss of portability of the software on the optimization will not have much impact. But that is not the case. Portability of the source code level that when it is run on other platforms without rewriting the code, simply re-compile it. This saves development costs and provide a larger market. Portability of the object code level, such as Java and Smalltalk, for developers and users is the ideal goal, developers can focus on writing code instead of busy porting the code, and users can choose their favorite platform . For users, Limits does not have any advantage in a platform. Follows the public network standard to be possible to obtain another kind of probability, like this itself cannot transplant the software may at least with other computer correspondence. The emergence of Web HTTP protocol is the direct cause of portability. HTTP may not all computers have brought to the optimum performance, but since it is in so many machines have been realized, that it provides the value of information sharing is very important. On any Web server can be any browser, because they both use the same language. I want to say here is that you can use to improve performance in exchange for portability, but it is only an ideal of the transaction, eventually you will pay a very high price.

F. The safety protection and the performance have the conflict

System security is another limitation, and all restrictions will reduce the freedom to optimize performance: SSL connection establishment takes time, the firewall will slow down the transmission speed of data packets, enter the password for the user's speed will slow down. Security is necessary, but its performance is often very significant.

G. Abstracts has the conflict with the performance

In the "higher" level of abstraction for programming, more details need to be considered, it will only get the performance in general, and will not be "the best performance." Whether the use of high-level programming language, or automatically generated SQL, so that the details of the system optimization procedures will lost simplicity to understand and control. Sometimes this good, but sometimes not.

H. Memory structure and mode of certain

Web can be seen as simply the slowest and most expensive kind of memory. Although the Web is not really in your computer, and in most cases it is just read-only, But it actually matches with memory other parts of structures[4].

Memory structures are different at every level of cost performance, price is often associated directly with the access speed. Recently used data are cached normally by the next faster level. The purpose of caching is to use the fastest memory the best performance, Let not the least use of high-speed cache.

If memory access is completely random, then the cache does not help performance, high-speed cache memory will continue to be covered in the random part of the data, so you are less likely to access the same data twice within a short time. However, there is a pattern of memory access, a recent memory address being accessed and the adjacent memory address also may be accessed immediately, this model is called the address of relevance. This is why some algorithm of the recent visit of the memory address and the address adjacent to the contents of the memory cache up can effectively improve the performance actually[5]. For example: It is effective to use this mode in the Unix file system buffer and cache Web browser.

III. The Optimized Goal Is The Overall Performance

When there is a Web system is not the bottleneck, they were considered to be in optimal condition. However, this did not mention the total throughput of the system—a very low system throughput can be in a technically optimized state. Optimization goal is to not waste any capacity, in other words, to achieve the ultimate state of each component at the same time. Some parts of the system is more durable than some other parts of little significance.

There is also a fact is: optimize the least effective system is the most likely to be optimized. When a component used up all the time, it is a clear need to focus on solving this problem. In handling the problems one by one, and again a round of this cycle, and so on until all components are equal when the issue so far. At this point, the end of optimization.

Because the Web system is usually dynamic, ensuring at all times to identify the weakest link in a timely manner is almost impossible. Optimal performance is not to spend all the time to track and Trace volatile bottleneck. to determine the

appropriate performance was more meaningful, even if some components have not been fully utilized. However, when most of the components are not fully utilized, there can be a problem.

In the Web, the smaller the packet arrival time will be faster. Regardless of how to optimize your system, too much content will the performance of the system crashes. Therefore, to maintain the content concise. Do not to receive arbitrarily large data from the user, but should be a reasonable place to cut it [1].

IV. Starts From The Highest Level To Optimize Obtained Biggest

It is difficult to make performance of the system optimization through modify the parameters repeated, only a careful analysis of system architecture and removed every part of the processing steps that can be canceled in order to optimize the performance of the system. To gain the most, must first analyze system's architecture in the highest level. This can also reduce the risk of the work, if only in small details of the level of optimization, then it may just be futile, because later you may find that the entire processing steps can be canceled

V. Web Improve Performance Model

The performance improvement may according to the model grouping, this way be more advantageous than each one concrete proposal to the work. Below is about the performance improvement technical model analysis.

A. Cost-sharing

For the purpose of economy, performance improvements usually involve how the cost sharing between multiple transaction processing:

1) HTTP allows a single TCP connection for downloading multiple files. This feature is known as the "persistent connection". Therefore, a TCP connection establishment and cancellation of more than one file involved, not only has a relationship with a file.
2) Design image map is another. Not to send multiple small images, but to send a large image. That if the original image is clickable, So by having this big picture into a clickable image of the way, you can also get the same functionality.
3) ⑶ Java.jar document with this similar, the java.class document package, like this may download them through a TCP connection, but is not establishes an independent TCP connection to each kind of document. This kind of situation negative influence is the kind which the .jar document will possibly contain some you always not to use.

A. Caching Technology

The idea of caching technology is simple: those frequently accessed data at hand. Only when some data than other data access is more frequent in practice, the high speed cache technology only then has an effect.

1) By the following approach to the storage space for more good performance: Run the most popular input program input offline data to the CGI program, and will cache all the results together. Then, users can quickly access to static HTML without having to generate dynamic HTML.
2) Increase the memory can reduce the server to find content on the disk needs, thereby reducing the access time.
3) Web proxy server caches some of the most popular Web pages, so that organizations can reduce the load on Internet access, while also reducing the time to access these pages.

B. Parallel processing

Web services have many problems due to multiple entities at the same time solving the same problem:

1) Netscape and some other browser can open multiple connections to the server, to issue multiple requests in parallel, and want the server to determine the order of one of the most effective services to these requests, rather than a random order so that customers request.
2) The Java procedure benefits from the multithreading, when some threads are blocked, the multithreading allows other threads to continue to carry out. For example: A Java application procedure's user registers when needs to fill in some things on the screen, then another different thread may use this opportunity to download other kind of documents. Only if the order is indeed very important, otherwise do not carry on the serial processing easily.
3) Symmetric multiprocessing hardware can map multiple threads to multiple CPU, and can execute code in parallel.

C. Profile

Profile can be used to discover the reality of some use patterns, use it to find bottlenecks in your code, you can also optimize the usage patterns. We can follow the following "Amdahl's recommendations" to quickly solve commonly encountered problems.

1) Discovers from configuration files' code most often the code which visits, in order to maximum limit optimize these codes.
2) Can be configured for the user, and web site settings will be closer in their place based on the information.
3) Write down the user download time, and assuming about what they have access to throughput, to adjust the content to make them more suitable for the user's access type.

D. Using known information

Do not underestimate the value of information, even the most trivial information:

1) You knew that the next visit is possibly on a HTML page's image, as the matter stands, theoretically speaking, the Web server may to the HTML page analysis, prefetch the image.
2) Once has used some connection, will have the possibility also to use to connect. Therefore HTTP has the lasting connection.
3) If the Web server can identify a particular user's usage patterns, you can optimize these models, you can prepare in advance of its contents, or the content needs to be dynamically generated.

E. Simplified processing

Sometimes simplifies the matter processes and reduces its scope to be possible to bring many harvests:

1) Built-in modem is not line connection between the system bus and the modem, so fast and cheap. You can not just because there is no connections between them will buy the wrong connections.
2) HTML content to maintain a small and simple, get rid of frames and tables, as little as possible to retain the image, which can greatly reduce the download time. Yahoo content is like that.
3) Use only static content do not have to CGI, so to reduce the cost of flexibility can greatly reduce service response time.

VI. Web Database Selection Strategy

People's great interest on the Web is due to relatively cheap and easy access to the global database on the Internet. Much of the information is on the host or in a relational database management system.

At present has three kind of standard database classes, each kind has the different request:

1) Inquiry of read-only database (such as AltaVista) is the individual query.
2) Carries on the very complex inquiry in the mass data, usually stems from the marketing need. This is called the data mining. A very famous data mining example: The grocery store has collected all goods sales situation, discovered that the beer and the handkerchief sell together frequently. Before nobody notes this, but both have sold out and needs to purchase, this becomes meaningful. As discovery result, the grocery store puts the beer and the handkerchief on the same place routinely. The data mining is read-only, moreover inquires is very usually complex, must spend the very long time, therefore we did not suggest that carries on the public Web visit.
3) Transaction processing, such as online credit card verification and sales, or bank account access. Transaction soon became an important and valuable Web applications.

The scalability and ability in the three database access are different. Read only access classes can easily be expanded by copying the database. Data collection is usually no need to extend the database, because few users will make such inquiries. Transaction database is the most difficult to expand, because at any time it can only copy the data to a host, this will cause a very significant bottleneck.

Planning and optimizing the database is a big area, which is much larger than to optimize the Web services all the work done together.

Choose from traditional performance SQL database is sometimes low, but the programming is easy, a low-capacity site can consider using them.

For those who only need a simple way to check the small data set, the best option is to all of the data downloaded to the client-side in HTML form, allowing users to use the browser's search function to obtain the corresponding row. On small data sets for complex queries, consider writing a Java applet and downloaded with the data, the program represents a user's query interface, you can simplify the query.

To the client side access speed, if the data set is too big, one procedure is in the server end with conventional CGI, the server API model or Java servlet carries on the inquiry. The Unix grep order is quite effective, very easy to use in CGI. Sometimes, inquires a simple ASCII data file to be able to gain the higher investment rate of return compared to any form database, because programs is very easy. Perl has the Hash table which very easy to use, to the peaple who compiled CGI with C, the Unix ndbm document also has similar function to the Hash table. C programmers can read a map to the memory and the memory structure directly as a binary file. If the cost of start CGI can be to reduce as a server API module or written as a background program running, then the performance of this approach will be very good.

Finally, if you need to use SQL for complex queries, but with relatively small data set, you can consider using MiniSQL, it is also known as mSQL. the mSQL performance is very good, moreover supports a large subset of ANSI SQL. using MySQL in small database is another good choice, moreover free.

VII. Database Performance Optimization Strategy

A. Database connection caching

We know that the user will access the database through the Web server, Web application and database server need to first establish a connection,, then can deposit and withdraw the data, after processing had ended, this kind of connection is closed. Each time the user visit needs to be redundant such step. As a result of database connection process dissipation system resources, moreover the time expenses are also quite big, particularly uses when public gateway connection CGI (Common Gateway Interface) carries on the connection, the efficiency is especially low, usually in the situation, the connection process to the system response time's influence is very big[6].

The database connection cahe is refers to between the Web server and the database server establishes the regular connection, when user need visit database, uses the connection which directly these already existed, after the operation had ended, the connection still maintained, but does not close, as the matter stands, the user visit database's step is simplified, then lift system's efficiency.

B. Anticipates statement and binding variable

The pre-analytical statement of database can be stored in a specific location together with variables. These variables are called " binding variable." Anticipates statements have much higher performance compared to the statement which analysis and optimization in the implementation, but the statement at the beginning of the establishment have a certain overhead[7].

When the inquiry to be carried out almost exactly the same and only the value of the query not the same, that is, not change the structure of the table, the use of pre-treatment of the statement is the best. However, the storage cost of statement pre-treatment is relatively high, so only suitable for

one-time use, not suitable for circulation.

C. To form non-normalized

Sometimes, the purpose of improving performance can be achieved for simply putting the data with common feature of into the same table, it will avoid the high cost of joint operations, but also makes simple work of writing queries. Because only need to query a table, eliminating the query trouble for joint multiple table. On the other hand, the form of non-regularization increases the possibility of inconsistent data, so as not to mistakenly think that the different data in the table are the same. However, the database administrator will work more difficult because the form of non-normalized[8].

D. Good query mechanism

A good mechanism can reduce the workload of the database query. Now the database contains a lot of optimization procedures, which are divided into two categories: rule-based optimizer and cost-based optimizer. Rule-based optimizer to optimize under a specific set of rules, the cost-based optimizer pay attention to the cost of the actual time for a particular query. We can add some comments to optimize the program in SQL statement on the following principles[1]:

1) The most common query results will be cached up.
2) To query and update firstly for some more restrictive operations, and the remaining part of the data to be processed will be less, thus speeding up the speed.
3) It is best to maintain a large time interval between database operations, that a small number of large query execution, rather than a large number of smaller queries.
4) Pre-compiled on the table.
5) To limit what you really want to lock on the data bits. If all of the query are locked to the same table, then these queries can only serial execution, performance will decrease.
6) Note that a bad SQL query will lead to the collapse of the entire database. So do not let the public unlimited access to the database, even on the intranet should be limited.

E. The rational allocation of connection pool

For a large-capacity site, the usage of connection pool is necessary. Cache and connection pool are important techniques in data access, It has a huge increase to the performance of accessing the database in some cases, and have been generally supportive of the database field. Conceives this kind of situation: You need to drink a glass of water, of course, the sooner the better. Usually, a water's production including from the water source extraction, purifies through the pipeline transmission and the equipment, before reaching your drinking water container. The above process is must, but is not each water production must be redundant one time the above process[9]. You can use a larger container to contain lot of water, the cost only is transferring water from the large container to the cup; You may also when the massive waters used, only need turn on the water valve, but does not need to lay down the pipeline to water source temporarily and the to purchase treated water equipment. Therefore, the Government will be laying pipelines and construction of water treatment station, to complete the more difficult work, achieve the purpose of sharing resources, and you can for your own needs, with containers to have a particular purpose water. Caching and connection pool have many similarities with the above specific vessel and the transmission pipeline, they all reached the same goal: to meet the desires of the user under the premise of shared resources as much as possible in order to improve overall system performance.

Because the establishment of database connection is very time-consuming, it can not to establish a database connection to the Web site for each visit. If you are using a single application server (such as Weblogic), you will need to configure the connection pool initial capacity as the maximum. This is because when the need to increase capacity of the connection pool, create a new connection takes a long time. If the start connection pool is set to maximum, then the growth of the connection pool without having to wait the time required. However, the shortcomings of this database is the need to use more resources, and also may use other applications resources.

F. Do not create a cursor in a loop

We know that the cursor is a memory region which store the query results, create cursors costly, it should be careful not to create a cursor inside the loop.

G. Multi- tier system

The setup of browser / Web server / database is three-tier structure, in a two-tier structure, the database is also server, it can get better performance for a small amount of users, but scalability is not good. To three-tier system, if does not make the plan to them, they can not take full advantage of the benefits of three-layer protocol[10].

When the user are many, the three-tier system can reuse the business objects in Web server or application server, Its read and write operations without having to immediately access the database again, this can greatly improve the performance. We can connect multiple databases into one database by a middle layer, which can be distributed application database. And the middle tier processing monitor can dramatically improve performance through access to the database, so do not open and close the connection for each query.

H. Row-level locking

Compared with the table-level locking, row-level locking can greatly improve performance.

I. Timely index

Only to establish the related index for which hope to query. Otherwise, continue to create and update the index will be a waste of time and disk space.

J. Integrated Web server / database

Some of the database itself is the HTTP server, thus canceled layer between the client and the database. Moreover, like this may also dynamic construct HTML, such as CGI, but also to maintain the status of the transaction. They can be set in pairs all requests to use the same database connection, compared to open a connection for each request, this approach can greatly improve performance[11]. Side effect is that they are dedicated, not easy to expand. For these mixed server written a application does not run on other servers.

Database allows network access to other databases, but you no longer deal with the performance advantages of only a process. Here are some Web server / database:

The IBM Merchant server uses DB2.

Web Datalade server uses the Informix database.

NS LiveWire server uses the Informix database, now also uses the Oracle database.

Oracle WebServer server uses the Oracle database.

Sybase Web SQL server uses the Sybase database.

VIII. KEY RECOMMENDATIONS

Read the fantasy manual。

For simplicity.

Minimize I / O.

In the possible situation, to transmit merely has the change content.

Whenever possible, use high-speed cache.

Best to share skills, rather than hide them as secret.

Using a connection pool.

If the data quantity is not big, then please consider that chooses the non-RDBMS system.

Create the index.

For complex queries, the first need to make the query for the most restrictive part.

IX. SUMMARY

Of course, to make the Web the fastest way is to do nothing. That is, if you can remove a part of the system, then this part should be removed. One way is to observe whether there is some redundancy, if any, are removed. But there is a better way, that is not possible to consider the use of certain equipment. Your users will be running their own Web server, you may not need to use for a specific business Web. This Web performance issues was also gone.

REFERENCES

[1] Patrick Killelea. “Web Performance Tuning”. Tsinghua University Press. 2003

[2] Cao Dongqi. “Network Design Fundamental”. Beijing Hope Electronic Press, 2000

[3] Wu Gongqi. “Key technology e-commerce”. Economic Science Press.2002

[4] Kohavi R Mining. “E-commerce data:the good,the bad,and the ugly”. 2001

[5] Shahabi C.Banaei-Kashani F. “A framework for efficient and anonymous Web usage mining based on client-side tracking”. 2002

[6] Spiliopoulou M.Mobasher B.Berendt B. “ A framework for the evaluation of session reconstruction heuristics in Web-usage analysis” 2003(2)

[7] Srivastava J.Cooley B.Deshpande M. “ Web usage mining:discovery and applications of usage patterns from Web data”. 2000(2)

[8] J · S · park ，M · S · Chen and R S · Yu · “Using a Hash · Based Method with Transaction Trimming for Mining Association Rules” · IEEE Tram On Knowledge and Data Engineering · 1997 · 9 ，9(5)：813～825

[9] IT · Yah ， M · Jacobsen ， H · Garcia-Molina,and U · Dayaln · “From user access patterns to dynamic hypertext linking” · 1 999

[10] Michele Facca F.Luca Lanzi P. “ Recent developments in Web usage mining research”. 2003

[11] Gaochuanshan, Qiansongrong. “Data communications and computer networks”. Beijing. Higher Education Press. 2001

Wang Xin, Born in 1969, Male, Master of Engineering, Associate Professor of Computer and Communication Engineering College, Weifang University, Main research directions is E-commerce.