SA Group

Assignment 1

Tom Moulard (16920041) *Long Wei*

Mathis Raguin (16920040)

Louis Tamagny (16920043)

Will Detlor (16920031)

Yuhui Zhang (14301027)

20 march 2017

Contents

1 Introduction

1.1 Subject Presentation

Architecture design of Web Architecture

2 Architecture Model

3 Hight Performance Architecture

A small website, such as a personal website,which can be accomplished by the simplest HTML, and then with some pictures to make it more beautiful. For such websites, all the pages are stored in a directory, so the website's requirements on the system architecture, performance are very simple. With the continuous enrichment of the Internet business, website-related technology through these years of development, has been subdivided into a very fine aspects. Especially for large sites, the use of technology is involved in a very wide range, From hardware to software, programming language, database, webserver, firewall and other fields all have a very high demand.

Large sites, such as portals. When they face a large number of user access, high concurrent requests, and the basic solution focused on several aspects: to use high-performance servers, high-performance database, efficient programming language, as well as high-performance Web containers. But in addition to these aspects, there's no the fundamental solution to large sites facing high load and high concurrent problems.Through reading papers, we summarized the points as follows:

3.1 HTML static

As we all know, with the highest efficiency, but the smallest consumption is pure static html pages, so we can try our best to adopt the way by use static html to implement our website. The simplest method is actually the most effective way. But for the sites that have a large number of content and are updated frequently, we can not manual to modify these content, so the information release system CMS was born, such as the portal site of the news channel, is through the information release system to manage and implement. Information release system can achieve the simplest function that Input the information and then generate static pages automatically, and it also can be used to manage channels, manage authority,Automatic crawling and other functions. for a large website, with a set of efficient, manageable CMS is essential.

In addition to the portal and information release websites, for the highly interactive community type of site, as much as possible static is a necessary method to improve the performance. Make the article and and post static, modify them and the re-static is also a strategy that has been used heavily.

At the same time, html static is also a means of some of the cache strategy. For the applications that use the database query frequently but the content update rarely, can be considered to implement by html static method. Such as forum's public settings information, these information can be managed through backstage and re-stored to database, and these information is called heavily but the update frequency is low, so we can consider this part of the contents static when background update to avoid a large number of databases access request.

3.2 Separate picture and server

As we all know, for the Web server, whether Apache, IIS or other containers, the picture is the most resource-consuming, so we need to separate the picture and the page, which is a strategy that a large site will adopt. These kind of website has Independent image server, and even a lot of picture servers. This kind of architecture can reduce the pressure that provides page access requests on the server system, and can ensure that the system will not crash due to image problems. On the application server and the picture server, different configuration optimizations can be made, such as apache when configuring ContentType can support less, as little as possible LoadModule, to ensure higher system consumption and execution efficiency.

3.3 Database cluster and library table hash

Large sites all have complex applications, these applications must use the database, so when they in the face of a large number of visits, the database bottleneck will soon be able to show up, then a database will soon be unable to meet the application, so we need to use the database Cluster or library table hash. In the database cluster, many databases have their own solutions, Oracle, Sybase and so on all have a good program. Even Master/Slave MySQL provided is a similar program. The database cluster mentioned above is limited by the type of DB used in terms of architecture, cost, and expansion, so we need to consider improving the system architecture from the view of the application. Library table hashes are the most common and effective solution.

We install the function module to separate the database in the application to separate the database, different modules corresponding to a different database or table, and then according with a certain strategy or function to make a smaller database hash. Such as user table, In accordance with the user ID table hash, so that can improve the performance of low-cost system and get very good scalability. Sohu's forum is adopt this way to hash database and table in accordance with the plate and ID for the post and the user, and ultimately a simple configuration in the configuration file will allow the system at any time to add a low-cost database to complement the system performance.

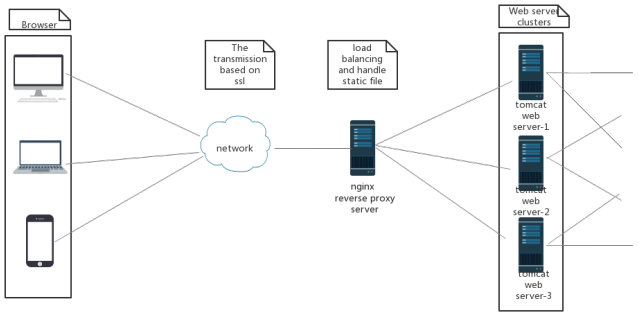

Figure 1: Database

3.4 Cache

Cache is very important for web site architecture and web development. Apache provides its own cache module, you can also use the external Squid module cache, these two methods can effectively improve the ability of access to the response for Apache

Cache for the development of web program, Memory Cache is a common cache interface can be used in web development which is provided by Linux, such as when use Java to develop we can call MemoryCache to implement data cache and communication sharing, some large communities have used such structure. In addition, the basic language has its own cache modules and methods, PHP has Pear Cache module, Java has more.

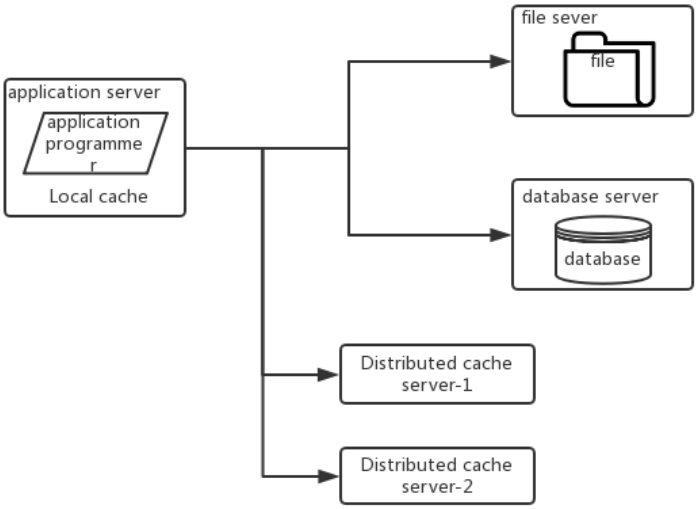

Figure 2: Cache

3.5 Mirror

Mirroring is often used in large-scale Web sites to improve performance and data security, mirroring technology can solve access speed difference that brought by the different network access providers and regions, such as the difference between ChinaNet and EduNet promote a lot of sites to build a mirror site in educational network.

3.6 Load balancing

Load balancing will be the ultimate solution for large sites to address high load access and a large number of concurrent requests. Load balancing technology have developed for many years, there are many professional service providers and products can choose, of which there are two architectures can give you to refer.

3.6.1 Hardware four-tier exchange

The fourth layer exchange uses the header information of the third layer and the fourth layer packet, according to the application interval identification service flow and then assigns the service flow of the entire interval segment to the appropriate application server. The fourth layer exchange function is like a virtual IP, pointing to the physical server. It communicates services that are subject to a variety of protocols, including HTTP, FTP, NFS, Telnet, or other protocols. These services on the basis of physical servers, and need for complex load balancing algorithm. In the IP world, the service type is determined by the terminal TCP or UDP port address. The application interval in the Layer 4 exchange is determined by the source and terminal IP addresses, TCP and UDP ports. In the field of hardware four-layer exchange products, there are some well-known products can be chosen, such as Alteon, F5, etc., these products are expensive, but value for money, they can provide very good performance and very flexible management capabilities. Yahoo China had nearly 2,000 servers using three or four Alteon to get.

3.6.2 Software four-tier exchange

We all know the principle of the hardware four-tier switch, based on the OSI model to achieve the four-tier exchange of software also came into being, this solution to achieve the same principle, but with slightly worse performance. But to meet a certain amount of pressure is ease, some people say that the software is actually more flexible, and the ability depend on you configure absolutely. For software four-tier exchange, We can use LVS which is used commonly in Linux to solve. LVS is Linux Virtual Server, it provides a heartbeat based heartbeat real-time disaster response solution to improve the system's robustness, while for flexible virtual VIP Configuration and management functions, can meet a variety of application requirements at the same time, which is essential for distributed systems. Typical strategy for using load balancing is to build a squid cluster on the basis of software or hardware switching. This idea is adopted on many large sites, including search engines. Software four-tier exchange is cost-effective, have high performance, expansion, the structure to increase or decrease the nodes are very easy at any time inside and outside.

4 High Availability Architecture

5 Scalable Architecture

6 Security Architecture

7 API ?

8 Front End Architecture ?

9 Conclution

List of Figures