Large Data Center is the Wrong Direction for Internet Services

Position Paper 2

April 6, 2003

Abstract

Internet population increases very quickly. Internet service providers have to serve the increasing requests. Their solution is to build new large data centers. We argue that this is not the right direction for various reasons, such as resource and security. Finally we discuss an alternative solution, a distributed solution. We believe this is a better solution for its scalability and reliability.

1 Introduction

Internet services have become an important part of people's daily life. According to [10], the Internet population doubles every year in the past. They also report that a record half billion people worldwide now have home Internet access. The widespread of Internet has drawn a lot of attention from both the business and academic area. Many services have been provided in the Internet, such as shopping center, directory, search engine, and news center. These services are often hosted by some data centers.

A data center is a house or building to host computing devices. In [9], data centers are divided into three categories: corporate data center, co-location data center and managed data center. A corporate data center is owned by the corporation for the purpose of storing information and providing software applications specifically for that corporation. Co-location and managed data centers are two types of hosting facilities. In a co-location data center, companies rent space to host their computer equipments. The difference between co-location data center and managed data center is that in a managed data center, the data center provides other services such as management to monitor and maintain the computer equipment.

As the population of Internet users increases, the need for more service providers increases. According to [6, 4], their survey shows that both the number and the size of data centers will increase significantly: about 50% more than the previous year.

While people are pouring money and efforts into data centers, the purpose of this paper is just to remind a single fact: large data center is not the right direction for Internet Services. Before we build data centers, we have to be aware of this fact. Otherwise, it will not only be a waste of money, but also will affect people's Internet usage, which will lead to affect their daily life.

In the rest of this paper, in section 2, we will discuss some counter arguments. In section 3, we will discuss our arguments: why we have to worry about large data centers. In section 4, we will discuss an alternative approach. Finally we'll conclude based on the arguments.

2 Counter Arguments

The need for large data center can be summarized in two reasons: performance and manageability.

As the Internet population increases, the Internet service providers have to serve more requests than before. To satisfy the increasing amount of requests, the service providers' solution is to get more computers and resources, which results in building larger data centers.

Another reason for data center is easy manageability of centralized data center. Many corporations use data center to store their critical information. Data Centers enable the enterprises to manage their critical computing resources in controlled environments under centralized management. Centralized management also makes it possible for enterprises to operate continuously or according to their business needs.

Although these two arguments may seem reasonable at the first glance, they are not valid after a careful thinking.

2.1 Performance

The main reason for larger data center is to host more computer devices. The need of more computer devices is driven by the objective to satisfy more user requests. Although the intention is good, the solution is not in the right direction.

The reason is that this solution only solves this problem temporarily and it doesn't scale. By increasing the number of computer devices, it may satisfy current requests—and we can see even this is not true—in a long time, as the Internet population keeps increasing, it will soon be overloaded. With a centralized data center, we will never be able to catch up with the users. We need a better solution for this problem.

2.2 Manageability

Although a centralized point can be helpful to control the computing environment: such as restrict the users who can access the data center by using some physical boundary. This physical boundary may be helpful to protect some sensitive information in some sense. But in the Internet service scenario, the situation is different from traditional corporate data centers.

A traditional corporation needs to protect its private and sensitive information. While for Internet Services, the main purpose is service, such as searching, communication, and proxy. All these can be done better in a distributed way. While we remove this part and distribute it, the need for large data center will be greatly reduced.

3 The Arguments

The arguments against a large data center include the traditional arguments against centralized servers, such as single point failure and congestion. Because of the size of data centers, there are also some new arguments, such as availability, power, security and privacy issue.

3.1 Availability

Because the variety and population of Internet Service users, Internet services need to function continuously, which means that they need to maintain high availability. But to maintain the availability of a large data center is difficult. People now are still struggling in trying to provide 4 to 5 nines availability.

Moreover, the result for depending a data center too much will be disaster. Because the Internet is keeping growing, it's impossible to satisfy all users' request by using a single data center. And also it's impossible to predict how many users are going to use this service. This can be easily proved by the effect of 9/11: the news centers were unavailable for a long time period.

Another reason is for the cost of downtime: the larger a service is, the more it will cost if it is unavailable. According to the META Group, the average cost of an hour of downtime is estimated at $330,000. Strategic Research Corporation reports the financial impact of major outages is equivalent to US$6.5 million per hour for a brokerage operation, or US$2.6 million per hour for a credit-card sales authorization system. We can reduce the cost by distributing the service into multiple small parts.

3.2 Power

Another effect that often neglected by investors is the hidden cost, such as power consumption of a large data center.

The reason people seek for large data center is the cost. As the price of computers goes down, it is easy to build a large data center with a relatively low price and can have a relatively high performance.

However, the calculated cost has a big misunderstanding. They just calculate the immediate investment which include the resources. There is one hidden cost that they don't count, which is the power consumption.

According to [5], one data center(at full load) with computer floor 7,000 m^2 will consume electricity equivalent to 38,000 households and energy equivalent to 12,000 households. Power and energy account for large fraction of data center operating cost. The high power consumption means great investments in sophisticated racks, air conditioning and power generation infrastructure. The power consumption not only affects the corporation's bills, but also affects the environment and people's daily life, especially in an area where power is a scarce resource. For example, the power outages affect the business and daily life of California a lot[1].

3.3 Security

Security is always a big concern. With a centralized data center, the compromise or attack on it will have far more serious effects than a smaller one.

Although a normally functioning data center may provide a high availability, they are easily subject to attack. Depend too much on a centralized data center will affect not only the service provider's incoming, but also affect the user's daily life. Imagine how much researchers rely on google and other search engines, and imagine the effect when these search engines are attacked and unavailable for a long time period. The concern about security is never over-estimated, especially in modern highly connected world: one example is the wide spreading of viruses. Another example is the distributed denial of service(DDoS) attack on February 2000, which brings down some major web sites, such as Yahoo, CNN, Amazon, eBay. And now DDoS is still a threat to web sites[8].

3.4 Privacy

A centralized server will make it difficult to protect users' privacy. This is an issue that may frighten users away. Large data centers often analyze users' activities and profile their patterns. The results are a great property to the service providers but may be a potential violation to the users' privacy. This possible violation makes many users unwilling to use some Internet services.

4 An Alternative Approach

As we argue that large data center is not the right solution, we believe that a distributed solution is more feasible. The solution should rely on the existing infrastructure of Internet. Some researchers have already started on this direction, such as peer-to-peer auction[2], peer-to-peer file store[7], and peer-to-peer file sharing[3]. All these don't need a centralized server and can be deployed under the existing Internet. This eliminates the need of large data center and provide better performance, reliability and availability. We believe this is the right direction.

5 Conclusion

As Internet population increases, the need to satisfy their requests also increases. Many large data centers are built to satisfy these increasing requests. We argue that this is not the right direction for various reasons, such as resource and security. We believe that a distributed solution is better for the long run.

References

[1] Jonathan Angel. Emerging technology: Energy consumption and the new economy. January 2001.

[2] Marcus Fontoura, Mihail Ionesu, and Naftaly Minsky. Law-governed peer-to-peer auctions. May 2002.

[3] Gnutella. www.gnutella.com.

[4] Yankee Group. Executive summary of the u.s. collocation market: High-tech real estate heats up, 2000, www.yankeegroup.com, viewed 28 march 2001.

[5] Frank Hartkamp. Energy consumption of ict in the netherlands: perspective to the year 2010. In *Workshop IEA Impact of ICT on the energy system*, February 2002.

[6] Richard A Juarez, Chetan S. Karkhaniz Michael T. Alic, and Brett D. Johnson. Spacedexiii. hosting space: Not all space is created equalsmart, complex space takes center stage, robertson stephens, 29 january 2001.

[7] John Kubiatowicz, David Bindel, Yan Chen, Patrick Eaton, Dennis Geels, Ramakrishna Gummadi, Sean Rhea, Hakim Weatherspoon, Westly Weimer, Christopher Wells, and Ben Zhao. Oceanstore: An architecture for global-scale persistent storage. In *Proceedings of ACM ASPLOS*. ACM, November 2000.

[8] Robert Lemos. A year later, ddos attacks still a major web threat.

[9] Jennifer Mitchell-Jackson. Energy needs in an internet economy: a closer look at data centers.

[10] Nielsen/NetRatings. http://www.eratings.com/news/2002/20020306.htm.