Volume 2 No. 3, March 2012 ISSN 2223-4985

International Journal of Information and Communication Technology Research



http://www.esjournals.org

# Architecting the Cloud: Prospects and Problems

**Philip Achimugu, Oluwatolani Oluwagbemi, Victor Popoola**
Department of Computer and Information Science, Lead City University, Ibadan.

## ABSTRACT

Highly-available and scalable web hosting can be a complex and expensive venture. Traditional scalable web architectures have not only needed to implement complex solutions to ensure high levels of reliability, but have also required an accurate forecast of traffic to provide a good quality of service (QoS) to clients. Dense peak traffic periods and wild swings in traffic patterns result in low utilization rates of expensive hardware, yielding high operating costs to maintain idle hardware, as well as an inefficient use of capital for underutilized hardware. As a result, this paper discusses the merits of cloud computing as well as the challenges that could be encountered when hosting computing resources in the cloud.

**Keywords:** *Scalability, Web, Architectures, QoS, Traffic*

## 1. INTRODUCTION

Cloud computing is a mechanism for enabling efficient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. The world-wide web (www) has proven to be the fundamentals of cloud computing. Many businesses rely on it to provide their clients with immediate access to information. However, to retain a large number of clients, it is important to guarantee a reasonable access performance regardless of the request load that is addressed to the system. Web application hosting systems therefore need the ability to scale their capacity according to business needs.

As the web is increasingly becoming more "dynamic", the content is produced by programs that execute at the time a request is made and is often customized based on several factors like a user's preferences and the previous content the user has viewed [1]. Dynamic content allows creation of rich interactive applications like social networks, bulletin boards, civic emergency management, and e-commerce applications, which represent the future landscape of Web applications.

Generally, web applications are typically deployed on a three-tiered server-side architecture consisting of one or more instances of: *a web server*, *an application server*, and *a database server* [2]. The web server manages the HTTP (Hyper-Text Transfer Protocols) interactions, the application server runs the application code, and the database server houses the application's database. All these servers are referred to as home *server(s)*. (Figure 1 shows the resulting architecture.)

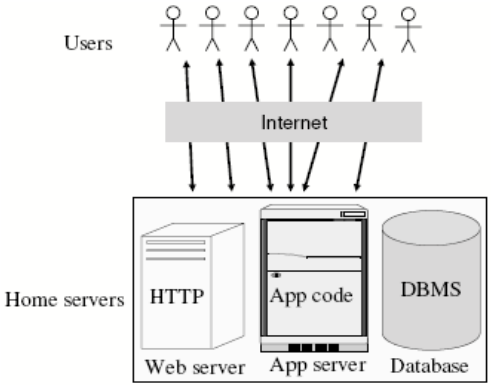


**Figure 1: Traditional Web-Based Architecture**
**Source: [2]**

The key to scalability is to ensure that the home server(s) remains lightly loaded even at high request rates. Because an application's web servers and application servers do not carry any persistent state, they can be replicated so that each replica remains lightly loaded even at high request rates. Alternatively, an application can use a Content Distribution Network that executes application code to scale its web and application server.

The main challenge is to design a *Database Scalability Service (DBSS)*, which can effectively offload at least some of the database work from the home infrastructure's database server(s) into an interoperable rescue database system.

More so, many research efforts have been made to provide scalable infrastructures for static content. However, scaling web applications that dynamically generate content remains a challenge. This research therefore, seeks to explore how cloud computing can help organizations provide good

quality of service for various web-based applications and also identify some of the challenges that could militate against efficient web-based applications.

## 2. SCALABILITY CHALLENGES IN WEB-BASED APPLICATIONS

Dynamic Web applications are characterized by capabilities for personalization and distributed updating of data [4]. These features, coupled with the often sensitive nature of the associated data, create new systems and security challenges in building an effective scalability service for dynamically-generated Web content. Web applications are typically deployed on a three-tiered server-side architecture consisting of one or more instances each of: a web server, an application server, and a database server. Most advanced Web applications rely on the database server(s) for the bulk of the data management tasks, and indeed the database servers often become the bottleneck in terms of maximum supportable load. Consequently, an effective scalability service would have to offload at least some of the database work from the home organization [5]. However, that task is encumbered by two major difficulties inherent in dynamic Web applications:

a. Most advanced Web applications require strong consistency for their most important data. It is well-known that maintaining strong consistency among replicas in a distributed setting presents significant scalability challenges.
b. Administrators are typically reluctant to cede ownership of data and permit data updating to take place outside the home organization. This reluctance arises with good reason, due to the security concerns, data corruption risks, and cross organizational management difficulties entailed. Difficulty firstly precludes caching techniques based entirely on timed data expiration, i.e., *time-to-live* (TTL) protocols [2], which are the norm in current approaches to scaling the delivery of static Web content. The growing number of dynamic Web applications with sensitive and mission-critical data requires a high degree of data fidelity and rely on systems that adhere to the transactional model of consistency. Secondly, in many cases, capabilities for data updating simply must remain within the boundaries of the home organization, and decentralized data updating is simply unacceptable.

## 3. BUSINESS BENEFITS OF CLOUD COMPUTING

There are some clear business benefits to building applications in the cloud. A few of these are listed below:

a. *Almost zero upfront infrastructure investment*: If one have to build a large-scale system; it may cost a fortune to invest in real estate, physical security, hardware (racks, servers, routers, backup power supplies), hardware management (power management, cooling), and operations personnel. Because of the high upfront costs, the project would typically require several rounds of management approvals before the project could even get started. Now, with utility-style cloud computing, there is no fixed cost or startup cost.
b. *Just-in-time Infrastructure*: In the past, if your application became popular and your systems or your infrastructure did not scale you became a victim of your own success. Conversely, if you invested heavily and did not get popular, you became a victim of your failure. By deploying applications in-the-cloud with just-in-time self-provisioning, you do not have to worry about pre-procuring capacity for large-scale systems. This increases agility, lowers risk and lowers operational cost because you scale only as you *grow* and only pay for what you use.
c. *More efficient resource utilization:* System administrators usually worry about procuring hardware (when they run out of capacity) and higher infrastructure utilization (when they have excess and idle capacity). With the cloud, they can manage resources more effectively and efficiently by having the applications request and relinquish resources on-demand.
d. *Usage-based costing*: With utility-style pricing, you are billed only for the infrastructure that has been used. You are not paying for allocated but unused infrastructure. This adds a new dimension to cost savings. You can see immediate cost savings (sometimes as early as your next month's bill) when you deploy an optimization patch to update your cloud application. For example, if a caching layer can reduce your data requests by 70%, the savings begin to accrue immediately and you see the reward right in the next bill. Moreover, if you are building platforms on the top of the cloud, you can pass on the same flexible, variable usage-based cost structure to your own customers.
e. *Reduced time to market*: Parallelization is the one of the great ways to speed up processing. If one compute-intensive or data-intensive job that can be run in parallel takes 500 hours to process on one machine, with cloud architectures, it would be possible to spawn and launch 500 instances and process the same job in 1 hour. Having available an elastic infrastructure provides the application with the ability to exploit parallelization in a cost-effective manner reducing time to market.

### 3.1 Technical Benefits of Cloud Computing

Some of the technical benefits of cloud computing includes:

a. ***Automation:*** *"Scriptable infrastructure":* One can create repeatable build and deployment systems by leveraging programmable (API-driven) infrastructure.
b. ***Auto-scaling:*** You can scale your applications up and down to match your unexpected demand without any human intervention. Auto-scaling encourages automation and drives more efficiency.
c. ***Proactive Scaling:*** Scale your application up and down to meet your anticipated demand with proper planning understanding of your traffic patterns so that you keep your costs low while scaling.
d. ***More Efficient Development lifecycle****:* Production systems may be easily cloned for use as development and test environments. Staging environments may be easily promoted to production.
e. ***Improved Testability:*** Never run out of hardware for testing. Inject and automate testing at every stage during the development process. You can spawn up an "instant test lab" with pre-configured environments only for the duration of testing phase.
f. ***Disaster Recovery and Business Continuity:*** The cloud provides a lower cost option for maintaining a fleet of DR servers and data storage. With the cloud, you can take advantage of geo-distribution and replicate the environment in other location within minutes.
g. ***"Overflow" the traffic to the cloud:*** With a few clicks and effective load balancing tactics, you can create a complete overflow-proof application by routing excess traffic to the cloud.

## 3.2 Challenges of Cloud Computing

The major factors militating against cloud computing environment are complex but solvable. These factors must be taken into cognizance before designing and deploying any application in the cloud. They include:

a. ***Heterogeneity:*** Various entities in the system must be able to interoperate with one another, despite differences in hardware architectures, operating systems, communication protocols, programming languages, software interfaces, security models, and data formats.
b. ***Transparency:*** The entire system should appear as a single unit and the complexity and interactions between the components should be typically hidden from the end user.
c. ***Fault tolerance and failure management:*** Failure of one or more components should not bring down the entire system, and should be isolated.
d. ***Scalability:*** The system should work efficiently with increasing number of users and addition of a resource should enhance the performance of the system.
e. ***Concurrency:*** Shared access to resources should be made possible.
f. ***Openness and Extensibility:*** Interfaces should be cleanly separated and publicly available to enable easy extensions to existing components and add new components.
g. ***Migration and load balancing:*** Allow the movement of tasks within a system without affecting the operation of users or applications, and distribute load among available resources for improving performance.
h. ***Security:*** Access to resources should be secured to ensure only known users are able to perform allowed operations.

## 4. CONCLUSION

As we have noted thus far, cloud computing is fast becoming an important part of peoples' lives as a result of innovations in the recent past in the area of Web-based applications, and will continue to make a serious impact in the future. Emerging technologies such as Grids will drive the next wave of innovation enabling the creation of applications that deliver IT as the 5th utility after water, electricity, gas, and the telephone. In summary, cloud computing is a very broad area with vast potential to improve efficiency of business processes and quality of life.

## REFERENCES

[1] Wallach, D., Burrows, M., Chandra, T., Fikes, A., and Gruber, R., (2006). Bigtable: A distributed storage system for structured data. In Proc. OSDI.

[2] Olston, C., Manjhi, A., Garrod, C., Ailamaki, A., Maggs, C., and Mowry, T., (2005). A scalability service for dynamic web applications. In Proc. Conf. on Innovative Data Systems Research.

[3] Chang, F., Dean, J., Ghemawat, S., Hsieh, W., Buyya, R., (1999). *High Performance Cluster Computing*, Prentice Hall, USA.

[4] Couloris, G., Dollimore, J., and Kinberg, T., (2001). *Distributed Systems – Concepts and Design*, 4th Edition, Addison-Wesley, Pearson Education, UK, 2001.

[5] Tanenbaum, A., and Van Steen, M., (2002). *Distributed Systems: Principles and Paradigms*, Prentice Hall, Pearson Education, USA.