Cloud Computing:
Locally Sub-Clouds instead of Globally One Cloud

Nawsher Khan, Universiti Malaysia Pahang, Malaysia

Noraziah Ahmad, University Ahmad Dahlan, Indonesia

Tutut Herawan, University Ahmad Dahlan, Indonesia

Zakira Inayat, University of Engineering & Technology, Peshawar, Pakistan

ABSTRACT

Efficiency (in term of time consumption) and effectiveness in resources utilization are the desired quality attributes in cloud services provision. The main purpose of which is to execute jobs optimally, i.e., with minimum average waiting, turnaround and response time by using effective scheduling technique. Replication provides improved availability and scalability; decreases bandwidth use and increases fault tolerance. To speed up access, file can be replicated so a user can access a nearby replica. This paper proposes architecture to convert Globally One Cloud to Locally Many Clouds. By combining replication and scheduling, this architecture improves efficiency and easy accessibility. In the case of failure of one sub cloud or one cloud service, clients can start using another cloud under "failover" techniques. As a result, no one cloud service will go down.

Keywords: *Cloud Computing, Effectiveness, Efficiency, Replication, Scheduling, Scheduling Techniques, Sub-Cloud*

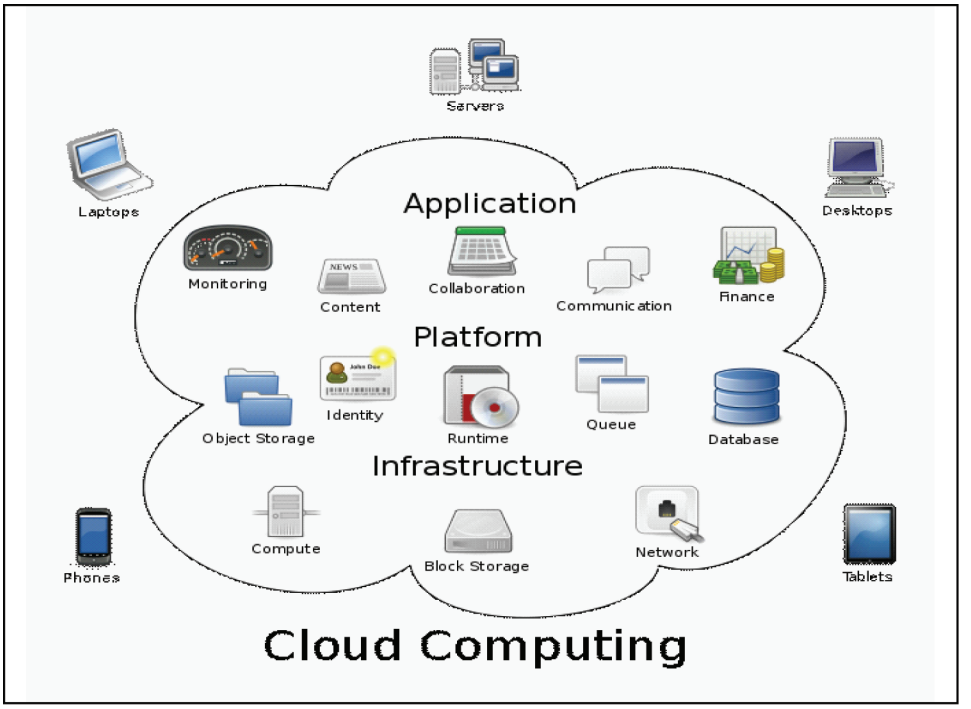
INTRODUCTION

Efficiency, scalability and easy accessibility are the key factors and should be the key features of cloud computing. From end-user computing, data storage and data transferring requirements are growing. Users demand for more capacity, more reliability and the capability to access information from anywhere in the world. Cloud services (computing, storage and transferring) meet this demand by providing transparent, easy and reliable solutions. Since late 2007 the concept of cloud computing was proposed (Weiss, 2007) and it has been utilized in many areas with some success (Brantner, Florescuy, Graf, Kossmann, & Kraska, 2008; Moretti, Bulosan, Thain, & Flynn, 2008). Cloud computing is deemed as the next generation of IT platforms that can deliver computing as a kind of utility (Buyya, Yeo, Venugopal, Broberg, & Brandic, 2009). Foster, Yong, Raicu, and Lu (2008) made a comprehensive comparison of grid computing and cloud computing.

By a cloud, we mean an infrastructure that provides resources and/or services over

DOI: 10.4018/ijcac.2012070103

Figure 1. Simple architecture of cloud computing



the Internet. A storage cloud provides storage services (block or file based services); a data cloud provides data management services (record-based, column-based or object-based services); and a compute cloud provides computational services. Often these are layered (compute services over data services over storage service) to create a stack of cloud services that serves as a computing platform for developing cloud-based applications. Examples include Google's Google File System (GFS), BigTable and MapReduce infrastructure (Dean & Ghemawat, 2004; Ghemawat, Gobioff, & Leung, 2003). Amazon's S3 storage cloud, SimpleDB data cloud, EC2 compute cloud (Amazon, 2009); and the open source Hadoop system (Borthakur, 2007; Dean & Ghemawat, 2008). Figure 1 shows the simple architecture of cloud computing.

For the majority of applications, databases are the preferred infrastructure for managing and archiving data sets, but as the size of the data set begins to grow larger than a few hundred terabytes, current databases become less competitive with more specialized solutions, such as the storage services (e.g., Borthakur, 2007; Dean & Ghemawat, 2008) that are parts of data clouds. For example, Google's GFS manages Petabytes of data (Hbase Development Team, 2009).

Cloud architectures are middleware services for different purposes i.e., resource allocation management, job scheduling, security, authorization and data management etc. When a user requests a file, a large amount of bandwidth could be spent to send the file from the server to the client and the delay or response time involved could be high (Bsoul, Al-Khasawneh, EddienAbdallah, & Kilani, 2011; Ben Charrada, Ounelli, & Chettaoui, 2010a). Besides that, maintaining local copies of data on each accessing site are cost prohibitive while storing all data in a centralized manner is impractical due to remote access latency