# Cloud Computing: Workload Forecasting

Sabnam Nayak[1],
[1,] Department of Computer Science and Engineering,
Ganga Institute of Technology and Management,
Jhajjar- 124104, INDIA

Shalini Yadav[2]
[2] Assistant Professor
Department of Computer Science and Engineering,
Ganga Institute of Technology and Management,
Jhajjar- 124104, India

**Cloud computing has revolutionized the entire web. To use cloud environment efficiently one need to understand workload and capacity requirements of cloud. Fluctuating workloads can result in mismanagement of cloud resources. Workload prediction is becoming very crucial, as almost every organization is moving towards cloud. In this paper workload prediction, its importance and steps involved for predicting workload are discussed.**

***Index Terms— E-Business, IAAS, PAAS, SAAS***

## I. INTRODUCTION

According to Gartner reports, cloud-based security service market has increased to $3.1B in 2015 from $2.1 B in 2013 globally [1]. Cloud computing popularity is resulted from its ability to self-host different services. The cloud computing has brought a revolution in the way information technology services are invented, deployed, developed, updated, scaled, maintained and paid. Computing today reflects a major change in the world of technology — on one hand, computers are becoming exponentially more and more powerful and the per-unit cost is falling rapidly [2,3]. On the other hand, as computing is becoming more pervasive, there is an increasing difficulty in management of whole infrastructure [4]. The cloud computing brings together all the functionalities of existing IT services with reduction in upfront costs of computing [5].

The cloud computing has become global. The cloud computing has made its impact on each and every aspect of business within every organization. Every organization is using cloud as it provides the benefits of cost saving and efficiency. Third party cloud service providers manage the services of cloud. It has resulted in simplified and enhanced IT management and maintenance because of central resource management. Some of the recent trends of cloud computing technology are as follow [6] [7].

- Hybrid clouds. Hybrid clouds infrastructure combines private cloud features like security with scalable, powerful and cost-effective features of public cloud. In this type of cloud system IT managers have vast choices for solutions. As hybrid models are becoming main stream, more companies will adapt this cloud deployment model.

- BYOD. Large numbers of customer are using mobile devices "bring your own device". End users use mobile devices in order to place large amount of data into cloud services. Cloud services can be used for syncing, storing and streaming of data. Different researches must be taken in order to find a suitable way of integrating personal cloud services in a

- BYOD environment with tools of Mobile Device Management.

- PaaS solutions provide more efficient development and testing environment. This results in cost reduction and increase accuracy of application development environment. As per International Data Corporation, one of the prominent analyst organization [7], the PaaS market is expected to increase to $14 billion globally by 2017 from its current cost of $3.7 billion.

- Big data. Many organizations believe that merging big data and cloud computing is very much beneficial. The emergence of big data analytics as a service will offer organizations an achievable and scalable tool for competing with global market.

- Graphics intensive applications. Initially requirement of graphic intensive applications includes considerable hardware infrastructure as investment. This reality is changed by cloud computing. There are a large number of cloud-based graphics technologies from famous graphics companies that helps end users in running high-end graphic intensive applications with help of different web browsers.

- Identity protection. Security is most important aspect of cloud computing. As almost every organization is moving information and data on cloud servers, this aspect requires major attention. Researches are going on for finding solutions of identity management based on cloud-based systems.

- Web-powered apps. Scalability and efficiency are among main benefits of cloud computing hence there is a need of cloud-based applications that are multiple platforms compatible and can run using multiple browsers.

Cloud computing effects almost every department of organizations in one way or another. It has influenced almost every technology like big data, security and protection, graphic intensive layout etc. Due to this rapid growth of cloud computing workload forecasting plays pivotal role. User behavior largely impacts the utilization of IT resources. The workload can be calculated in the form of request, loads, traffic, amount of data stored etc.

Workload refers to IT resources utilization depending on the type of application. The users are interested to pay for amount of resources used by them. In the next session characteristics of workload prediction involved in cloud computing are discussed.

## II. STUDY ON WORKLOAD PREDICTION

Workload with respect to cloud computing is an extent to which cloud consumers put their systems, resources and information on the cloud. As for example, a desktop workload supports a large number of users logging to desktop interactive sessions. Different applications on cloud have different characteristics and metrics. These applications and their requirements are discussed in table 1 below.

| SNO | Type of application on cloud | Example | Service parameters of cloud required |
|---|---|---|---|
| 1. | Web sites | Social networking, informational etc. | High bandwidth, large storage. |
| 2. | Scientific applications | Numerical analysis, Bioinformatics etc. | High computing capability. |
| 3. | Business software | ERP, data warehouse, SAS, Kenexa (hiring) | Customer support, availability. |
| 4. | Performance testing | Large workload is simulated to test the performance. | High computing capability |
| 5. | Development and testing | Rational Software Architect, Microsoft Visual Studio etc. | Flexibility, understandability and rich set of infrastructure services. |
| 6. | Mobile services | Mobile applications | High availability required. |
| 7. | E-commerce | Shopping | Variable computing load, example during holiday times |
| 8. | Highly Graphics intensive applications. | Virtualization and animation | High Network bandwidth, |

TABLE 1: APPLICATIONS AND SERVICE PARAMETERS REQUIRED

Depending on type of application role of cloud and workload can be decided. Wang [8] highlights the role of cloud in case of ecommerce applications. The aim of this paper is to provide basic idea of building e-business model on the cloud.

### A. *Types of workload*

Fehling, C [9] has discussed different types of workload in terms of elasticity as well as types. This paper acts as a base in understanding different types of workload involved. As users are more interested in paying for services used by them hence there is need of rapid elasticity that is growing and reducing resources allocated as per requirements. In the static scaling case, physical servers are provisioned beforehand hence the time taken for ordering, setting and starting them is not responsive enough for handling faulty prediction. Hence required resource is not available on time resulting in under provisioned of applications. In order to handle this inflexibility all resources need to be provisioned from starting and are difficult to decommission when workload decreases. This leads to over provisioned application load after the peak. It has direct effect on properties and cost of hosted applications as under provisioned applications results in poor performance and over provisioned results in high cost. As compared to static scaling, elastic scaling can provision resources more flexibility. Hence in case of elastic scaling there is no problem of resource over provisioning and under provisioning but this will be architecture depended [9].

Different types of workload according to Fehling, C [9] are as follow:

1. Static workload. This type of workload will change minimally with time. In this type of workload there is no explicit necessity for adding or removing resources. Examples include small applications used by organizations internally. In applications based on static workload of cloud, elastic scaling is not of much benefit due to constant need of required resources. The cost benefits available with cloud computing are limited in case of static workload. In some cases the homogenization of IT resources and elasticity in failure cases provides necessary benefits of a cloud.
2. Periodic workload: This type of workload has recurring peaks. Example includes monthly bills, weekly report etc. From a cloud service consumer's viewpoint the cost-saving ability of periodic workload allows decommissioning of resources as it provides pay per use pricing.
3. Once-in-a-lifetime workload: This type of workload has a peak once. In this workload IT resources have an equal utilization throughout but will rise only once in a lifetime.

### B. *Workload based on service model*

Workload need also depends on type of service model [9]. Some key points are discussed below.

- Infrastructure as a Service (IaaS): Applications experiences different amount of workloads that result in different utilizations of resources. Especially, in case of periodic workload, which is the special case of once-in-a- lifetime workloads IT resources need to be provisioned flexibly. One common example is an environment hosted by servers. With varying workloads the required number of servers can be adjusted. An IaaS provider offers physical and virtual hardware that can be used using self-service interfaces of cloud service providers. On these IT resources customer is able to install their operating system and other required software.
- Platform as a Service (PaaS): IaaS cloud environment can adjust number of servers according to requirements as discussed previously but this means customer need to manage their own operating

systems and execution environment. If many customers require similar kind of environment there will be redundant installations on cloud resulting in inefficiency. Further the management of these environments needs to be handled by each customer of IaaS providers leading to more inefficiency. A cloud providers having Platform as service managed these difficulties of operating system and execution environment. In PaaS environment Cloud providers expects that each customer execution to be handled in different way. Different PaaS offerings like Google App Engine helps in deployment of different web applications. Others like Cordys provides elastic platform for deployment of business processes. Message oriented middleware is offered by Amazon simple queue services in the cloud.

- Software as a Service: Small and medium enterprises may not have the manpower to develop consumer software applications. If these applications are

installed by customer then the complexity to manage them still need to be handled by customer. SaaS model offers a complete software applications environment to consumers who may use it as required using self-service interface. The access is billed according to pay per usage basis. Example includes Saleforce.com that offer Web-based customer relationship management software as a SaaS cloud service offering.

Different properties of cloud computing like access using network, self-service on-demand, pay-as-per-utilization, resource sharing and elasticity decides how to allocate IT resources to consumers. It is essential to understand properties and their delivery issues. When considering application workloads, different dimensions have to be considered. Different service parameters are also discussed in table 1. Additional dimensions to consider are as follow [10]:

- Legacy applications. Legacy applications have specific requirements. These requirements are not compliant with the Cloud specifications.
- Front and back-office standard applications. These applications required further assessment. They can fulfill the requirements and specifications of cloud providers easily s compared to legacy applications.
- Online workloads or Batch workloads. Online workloads or Batch workloads required continuous assessment as batch workload required a lot of storage and memory for calculations as compared to online workloads.
- Workload frequency. During certain time period workload increases like in seasonal businesses experience high workload during specific time period as compared to complete year.
- Workload cost. The cost involved in hardware, operating system, software needs to be analyzed when putting together different business cases together.

There are two types of workloads:

- Workloads that can be executed at any time in batch mode and
- Workloads that need to be executed in real time.

Planning and analyzing depends on type of workload in used as every type of application and workload has different requirements. In next session steps involved in workload prediction as discussed by Almedia [11] [12] is presented. Workload prediction using neural network is presented in [13]. Also review of workload prediction for e-business website is presented in [14]. In [15][16][17][18][19][20] different features of cloud, cloud market characteristics along with market survey, and case study with respect to Amazon cloud is presented.

## III. Steps for Workload prediction Activity

Workload prediction and capacity planning is very much essential for management of quality of service of applications. Instead of depending on intuitions and different rule of thumb it is better to understand, analyze and plan according to type of applications. There are number of steps which are to be followed for workload prediction. Almedia [11] [12] discusses these steps required for capacity planning and workload prediction with respect to e-business websites. Following are general steps, which need to be followed for workload prediction.

### A. Environment Identification

The first step for workload prediction to be followed is acquiring in-depth understanding of the service architecture. This step results in an orderly description of the Web environment, its services and components. This initial phase consists of understanding kind of hardware and software resources required. It also deals with protocols that are used by environment. For example in case of e-business, website on cloud environment includes identification of peak usage time, management structures, and service-level agreements. Different parameters considered are mean time, service rate, load time etc.

### B. Characterization of workload

The process of explaining complete system workload in terms of components involved comes under workload characterization. In this step each component is studied in detail. Each workload component is decomposed into further basic components if required. In case of e-business websites basic components of workload are characterized by transaction arrival rates and service demand parameters for each resource. In Web-based environments, users' interaction is done through sessions. Navigational pattern is different for each user. Examples of e-business functions as requested by an online shopper includes browsing the catalog, searching for products and services based on entered keywords, selecting products for obtaining more detailed knowledge about product, addition in the shopping cart, user registration, and checkout.

### C. *Workload Model Parameter*

In this step values of different workload parameters are obtained. Performance measurement is done with help of different parameters that are selected to monitor workload characteristics of environment under study from different reference point. For example, transactions logs and server accesses are main source of information.

### D. *Forecasting Workload*

The fourth step deals with forecasting the expected workload for Web services on cloud computing. Demand forecasting is crucial for maintenance of quality of service. It plays critical role for the operation of Web services as high demand for the new product can slows down Web servers. The responsiveness and speed, starts degrading as more customers tried to access the web service. This scenario highlights significance of effective forecasting and planning for Web environments. A good forecast is combination of different scenarios and assumptions.

### E. *Performance Model Development*

Different quantitative techniques and analytical models are required for developing performance models of Web services. Performance models are used for predicting performance when different aspect of the workload is changed. There are two types of models: simulation model and analytical models. Analytical models are the specification of the interaction between the various components of a Web system. In simulation model, program is used to depict behavior of the actual system.

### F. Validating Performance Model

Once the model construction, parameterization, and solution is done it must be validated. Calculating difference between actual and expected errors validates performance model. Errors must be in acceptable range as decided by modeler. Various metrics like throughput, response time, and resource utilization etc. must be calculated with certain acceptable margin of errors.

### G. *Predictions of Service Performances*

Prediction is key to capacity planning. Model must be able to forecast effects of change in different parameter of web applications.

### H. *Analyzing Future Scenarios*

In this step, different combinations of architectures are analyzed for finding optimized one. In this expected workload and the quality of service perceived by customers are analyzed. End results of this step should help management in decision-making about different business goal and plan for future.

Workload prediction activity is needed in order to avoid drawbacks of inadequacy of capacity planning activity and to improve cost benefit ratio. Above-mentioned steps present fundamentals to basic capacity planning activity.

## IV. CONCLUSION AND FUTURE SCOPE

Almost every business is using cloud. So it has become very important to provide workload forecasting. Inaccuracy in workload prediction can lead to over provisioned or under provisioned of resources. If cloud workload is forecasted accurately cost benefit ratio can be optimized for almost every organization. Also workload prediction of applications differs from one to another. Workload requirements for banking and finance differ from that of e-business workload requirements.

In future work can be done for merging different cloud computing with any of recent technology. Also steps discussed by Almedia [12] can be act as base for forecasting capacity of any technology on cloud. Different tools like Matlab, excel etc. can be used for forecasting. Planning and analysis of workload depends on the application

## REFERENCES

[1] "Gartner Says Cloud Computing Will Be As Influential As E-business". Gartner. Retrieved 2010-08-22.
[2] S. Hackett, Managed Services: An Industry Built on Trust, IDC, 2008.
[3] J.T. Hamill, R.F. Deckro, J.M.K. Jr., Evaluating information assurance strategies, Decision Support Systems 39 (3) (2005) 463–484.
[4] P. Roehrig, New Market Pressures Will Drive Next-Generation IT Services Outsourcing, Forrester Research, Inc., 2009.
[5] J. Staten, Hollow Out The MOOSE: Reducing Cost With Strategic Rightsourcing, Forrester Research, Inc., 2009.
[6] Mike Solomonov "Cloud Computing Trends and Latest Stats"April 3, 2014
[7] John Grady "Major Current Trends in Cloud Computing" Mar 03, 2014 5:46 PM PDT
[8] Xiaofeng Wang "Research on E Commerce development model in cloud computing environment": 2012 International Conference on System Science and Engineering.
[9] Fehling, C.; Leymann, F.; Retter, R; Schupeck, W; Arbitter,P. "cloud computing fundamentals" Cloud Computing Patterns, 21 DOI 10.1007/978-3-7091-1568-8_2, Springer-Verlag Wien 2014
[10] Maximizing the Value of Cloud for Small-Medium Enterprises: Applicable Workloads for Cloud by The Open Group Platinum.
[11] Virgilio A.F. Almeida, Daniel A.Menasce "capacity planning: An Essential Tool for Managing Web Services".
[12] Virgilio A.F. Almeida "capacity planning for Web Services Techniques and Methodology".
[13] Sahi S.K. and Dhaka V.S. "Study on predicting for workload of cloud services using Artificial Neural Network": Computing for Sustainable Global Development (INDIACom), 2015 2nd International, pp. 331 – 335, Publisher IEEE
[14] Supreet Kaur Sahi and Dr. V.S.Dhaka. Article: A Review on Workload Prediction of Cloud Services. International Journal of Computer Applications 109(9): 1-4, January 2015
[15] Amazon.com: Elastic Compute Cloud (EC2). http://aws.amazon.com/ec2/. Accessed June 2013
[16] "Launch of IBM Smarter Computing". Retrieved 1 March 2011.
[17] NIST Definition of Cloud Computing, Peter Mell and Tim Grance, July 2009: www.nist.gov/itl/cloud/upload/cloud-def-v15.pdf.
[18] Rizwan Mian, Patrick Martin ;"Executing data-intensive workloads in a Cloud ", 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing ;78-0-7695-4691-9/12 © 2012 IEEE DOI 10.1109/CCGrid.2012.18.
[19] Rajkumar Buyya, Chee Shin Yeo, Srikumar Venugopal, James Broberg, Ivona Brandic "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility", Future Generation Computer Systems 25 (2009) 599–616, journal homepage: www.elsevier.com/locate/fgcs
[20] Kim Weins "Cloud Computing Trends: 2014 State of the Cloud Survey"(April 02, 2014) retrieved from http://www.rightscale.com/blog/cloud-industry-insights/cloud-computing-trends-2014-state-cloud-survey