Definition of Attributes for Standard Benchmarking of Cloud Services

Prarthana Gupta, Brajesh Patel

Abstract*—Today's era is now embracing the cloud computing as a new emerging technology. As a matter of fact the cloud computing technologies are now playing a very vital role in almost every commercial as well as non-commercial field; including large or small-scale businesses, IT organizations, Government, Quasi-Government, Public and Private sectors. In a cloud environment the service provision is usually handled by large number of different service providers. As it is a growing market, these providers try to attract the customers by highlighting their strengths, and also undermining their areas of improvement. In this competitive era it's very difficult for the customer to evaluate or compare the Service Quality and Service Ratings for offered services. This tends to create even more complex situation in case of multi-cloud management environment when customers are required to choose different services from multi-cloud platform involving different service providers in order to fulfill their organizational needs. The Problem is to choose a service which provides optimum, cost-effective, highly available, reliable, portable, scalable, efficient and secured services in cloud environment. This problem becomes more severe in case of multi-cloud management platform, where complexity will be higher as involvement of different providers, result into a cumbersome process of choosing the best provider. If the customers are new user for this technology, it is very difficult task for choosing a provider. Even if customer is having some information regarding services or those who are familiar with the provider services, this problem of choosing best provider is also complex as it involves comparing all the services from each service providers individually, understanding all their security terms, rules and financial aspects separately. This process is thus time consuming, slow, complex and annoying for the customers especially for new users. There must be certain criteria's which forms the basis of Quality Rating Comparison Analysis of services between different service providers. If the standardization of the searched criteria is further worked-out, then it will help in creating a benchmarking for comparing services provided by different service providers. Also it will help in reducing the complexity and extra workload required by customers in searching best service provider, thus enabling faster and less time consuming approach.

Index Terms*— Cloud Computing, Cloud Standards, Cloud Service Providers, Cloud Benchmarking, Cloud Benchmarking Parameters, Cloud, Unified Service Catalogue, Service Demand Handler.

I. INTRODUCTION

In this Paper, we propose development of a comprehensive set of criteria required for standardized benchmarking of cloud service offerings. These criteria are obtained by analyzing cloud service providers services and customers.

Manuscript Received on May 07, 2015.

Prarthana Gupta, Student – M.Tech., Department of Computer Technology and Applications, Shriram Institute of Technology, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Jabalpur, India.

Brajesh Patel, Head, Department of Computer Science, Shriram Institute of Technology, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Jabalpur, India.

The criteria analyzed, so far includes the most important and common features which are supported by these service providers are : Availability of service-high availability, Reliability, Scalability, Elastic nature, Security and access control rules, portability, Bandwidth requirement, Throughput, Disaster Recovery mechanism, Metering billing and monitoring facility includes financial aspects, Quality of service, Operational efficiency. For evaluation of these criteria, information regarding them are collected and updated by service providers on real time basis. These obtained values are used for quality rating comparison analysis of - Cloud Service providers (QRCA-CSP). This will also helpful in reducing complexity of obtaining better services in case of multi-cloud environment where options for choosing services from providers are available. This will help the customer to initiate the computing by selecting services easier and faster thus reduces the burden of selection process.

II. CLOUD DEFINITION AND MEASUREMENT CRITERIA

For the purpose of this paper, we adopted the definition of cloud computing from the NIST Special Publication [(Mell & Grance, 2011)], which defines Cloud Computing as: Cloud computing is a model for enabling ubiquitous, convenient, 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. Three roles exist within cloud computing. Cloud-Providers sell computing or software services purchased by the End Customer/Consumer. A Cloud-Provider builds its cloud using products from various Hardware and Software Vendors. However, in all cases, Cloud Computing has five essential characteristics, namely

1. On-demand self-service, where a consumer can provision compute and storage capabilities without requiring human intervention from provider.
2. Broad network access, where a consumer can access compute and storage capabilities over the network.
3. Resource pooling, where a provider groups together resources such as CPU, memory, disk, and storage to serve multiple consumers.
4. Rapid elasticity, where resources used can be rapidly and in some cases automatically increased or decreased to handle demand.
5. Measure service, where the service used a consumer is metered.

A. Service Models for Cloud Computing:

There are three service models for cloud computing. They affect the definition of a System under Test (SUT) for any cloud benchmarks. The following are a brief description of the service models.

IaaS: The Service Provider gives the End-Consumer the capability to the provision processing, storage, network, and basic computing resources. They can also deploy and run arbitrary operating systems. The End-Consumer does not manage or control the underlying physical cloud infrastructure, but has control over the operating system, assigned storage, deployed applications, and limited control of select networking components (e.g., host firewalls).

PaaS: The Service Provider gives the End-Consumer the capability to deploy consumer created or acquired applications created using programming, languages, libraries, services, and tools supported by the Service Provider. The Service Provider retains control and manages the underlying cloud infrastructure, including network, servers, operating systems, and physical storage. End-Consumer has control over the deployed applications and configuration settings for the application-hosting environment

SaaS: The Service Provider gives End-Consumer the capability to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through either a thin client interface, such as a web browser (e.g., web-based email), or a program interface. The Service Provider retains control and manages the underlying cloud infrastructure, including individual applications, these application configurations, network, servers, operating systems, and physical storage. The End-Consumer might have limited control of user specific application configuration settings.

B. Deployment Models

- Private Cloud: The cloud is configured for exclusive use by one or more organizations
- Public Cloud: The cloud is configured for use by the general public
- Hybrid Cloud: The cloud is a composition of distinct infrastructures in order to retain the proprietary technology by the consumer.
- Community Cloud: The cloud is provisioned for exclusive use for a community of consumers

C. How to measure Cloud?

There are at least three categories interested in the development and use of a cloud benchmark. Between these three parties exists two relationship types.

Cloud Providers: build data centers using standard hardware, network, and management software. Amazon, IBM, HP, VMWare, Microsoft, Oracle and Terremark are examples of companies that provide public or enterprise IaaS, PaaS or SaaS clouds. They will be interested in publishing the benchmark results of their service offerings. They would also be interested in the benchmark results of hardware, software or network vendor products, and how these components support the performance of their service

Hardware & Software Vendors: provide the hardware (computers, blades, or servers) and software (virtualization, management, automation) products used to build the cloud. AMD, HP, Intel, IBM, VMWare, OpenStack, Oracle, Red Hat, and VMware are examples of companies that might publish benchmark results for their customers.

End Customers: cloud customers might use cloud benchmark results to help select a Cloud Provider. These are primarily businesses. While there exists several collaboration and social apps used by individuals, the recommendations in this document will not address the needs of this community. We restrict our attention to business users of the cloud. Based on the above, use of cloud benchmarks fall into two broad categories. A given workload can be tested either in the black-box or white-box context.

White Box Benchmark disclosures: Published by one or more Hardware/Software Vendors and used by Cloud Providers to determine the right set of hardware and software products to deploy in their cloud. Cloud-Providers can also publish results of the hardware and software components used in their cloud. In a White Box Benchmark Test, The SUT's exact engineering specifications is known and under the control of the tester. The benchmark results allow full comparisons, similar to existing benchmark results

Black Box Benchmark disclosures: published by Cloud Providers and used by End-Consumers to determine the appropriate cloud service provider for their application needs. In a Black Box Benchmark Test, The Cloud-Provider provides a general specification of the SUT, usually in terms of how the End-Consumer may be billed. For example, an end user of a cloud may rent a cloud platform under the description of "2 compute units." The exact hardware details corresponding to these compute units may not be known. Due to this vagueness or incompleteness, comparing benchmark results requires additional information not currently collected or mandated. For example, the benchmark disclosure may include additional conditions, such as time and place where measurements were collected or dynamic changes to the SUT during the collection period.

D. Cloud Computing Use Cases

The general class of applications either use cloud computing infrastructure or might transition all or subsets to cloud computing.

Typical Business Use Cases	**Typical Consumer Use Cases**
Data Analytics Data Warehousing / Mining Business OLTP High Performance Computing NoSQL Databases Memory Cloud	Social Networking & Collaboration Mail / Messaging Streaming Audio / Video Voice over IP Online Gaming

The following table shows which Processing Type is present in each Cloud Computing Use Case:

Process Type \ Cloud Use Case	Social Networking	Collaboration	Mail / Messaging (IM, Tweets)	Data Analytics	Data Warehousing / Mining	NoSQL Databases	Business OLTP	Memory Cloud	HPC	On-line Gaming	Streaming audio/video	VOIP
Data-Intensive / Planned Batch Jobs	×		×	×	×	×	×	×	×	×		
Processing Pipelines	×		×	×	×				×	×		×
Dynamic Websites	×									×		
Business Processing / OLTP / Mission Critical applications				×		×	×		×			
Latency Sensitive		×	×			×	×			×	×	×
Application Extensions / Backends for Mobile Communications		×	×	×						×	×	
Bandwidth and Storage Intensive		×	×	×	×	×	×	×	×	×	×	×
Mail Applications			×									
Others			×	×								

III. BENCHMARK CONSIDERATIONS

Defining SUT for a cloud benchmark is challenging due to conflicting goals of interested parties and different cloud service models. As discussed in earlier sections, an End-Consumer does not have knowledge of the physical infrastructure for IaaS, PaaS, or a SaaS service. However, Hardware and Software Vendors are one of the interested parties in a cloud benchmark. In order for a cloud benchmark to have any useful meaning for these vendors, the physical infrastructure should also be part of the results that a cloud benchmark reports. However, mandating the reporting of physical infrastructure results is problematic, because it will exclude benchmarking many existing public Cloud-Providers. The System under Test (SUT) comprises all components (cloud service, hardware, software, network connections within the SUT, and support services which are being tested by the cloud workload or required by the specific benchmark run rules. It does not include any client(s) or driver(s) necessary to drive the cloud workload or the network connections between the driver(s) and SUT. The actual set of SUT's constituent pieces differs based on the relationship between the SUT and the tester.

Black Box Cloud: The SUT consists of a description of the specific cloud offering used to run the workload with sufficient detail to meet Full Disclosure Report (FDR) Requirements as described in the appendix and the specific benchmark's reproducibility requirements on an instance of the same offering.

White Box Cloud: The SUT description can be more specific, similar to many existing SPEC benchmarks. These SUT descriptions consist of:

1. The host system(s) (including hardware and software) required to support the Workload and databases.
2. All network components (hardware and software) between host machines which are part of the SUT and all network interfaces to the SUT.
3. Components which provide load balancing within the SUT.

All software that is required to build, deploys, and run the specific benchmark workload.

IV. CLOUD METRICS

Metrics are used as the measurement criteria for the tests. The following metrics have been identified as being the key indicators of performance for most workloads in a Cloud environment. Note: Due to the breadth of cloud implementations, not all metrics will be applicable to each benchmark. For instance, if End-Consumer tests a public cloud, the density and power metrics are typically not measurable. The key metrics identified are:

Elasticity – relates to how quickly a service can adapt to the changing needs of the customer. Thus, a highly elastic system can scale to include newer instances, as well as quickly provision those instances

Provisioning Interval – is defined as the time needed to bring up or drop a resource. This is the time between initiating the request to bring up a new resource or to relinquish it, and when the resource is either ready to serve the first request or when it serves the first request.

Agility – relates to the ability to scale the workload and the ability of a system provisioned to be as close to the needs of the workload as possible

Scale up/Down – is defined as measurements of the system's ability to maintain a consistent unit completion time when solving increasingly larger problems only by adding a proportional amount of storage and computational resources—i.e., if we double the resources, can we solve a problem twice as large?

Elastic speedup – indicates whether adding SUT resources as the workload is running results in a corresponding decrease in the response time— i.e., if we double the resources, we can solve the same problem twice as fast?

Throughput – is the units of work processed by the system or cloud per unit time. For example, this is the number of tasks completed per minute when the Hadoop cluster is at 100 % utilization processing multiple Hadoop jobs

Response time – is the interval between when a request is made by a client or workload generator and when the response is received by the client. In addition to the client-server response time measurement, White Box benchmarks have the specific ability to rely on internally measure response times within a SUT as well as on its external workload generators. This is not applicable for Black Box benchmarks.

Variability – measures the repeatability of the test results. Many variables affect the repeatability and should be factored into defining the values for this metric. The variability metric for any parameter should be based on the standard deviation of the measurement. Variability could be measured against any of the following parameters: Variability with time or Variability with SUT location. The measured parameter should be collected in a number of iterations and the standard deviation or a metric related to standard deviation should be reported. Other relevant metrics include:

Durability – is defined as the probability of data loss. Depending upon the context, this entity could be a requirement on the system and not tested in the duration period of a normal benchmark run.

Reliability – is the ability of a system or component to perform its required functions under stated conditions for a specified period of time. The reliability of a system is usually measured by the probability of failures or by the mean time between failures (MTBF).

Availability – is the degree to which a system or component is operational and accessible when required for use. The time during which the system is not available is called downtime; the time during which the system is available is called uptime. A small uptime and a very small downtime combination may result in a high availability measure – which could be misleading. Therefore, the mean uptime is also often known as the Mean Time between Failures (MTBF), together with Mean Time to Repair (MTTR), and considered as better indicators for availability. Where

$$\text{Availability} = \frac{\text{MTBF}}{\text{MTBF+MTTR}}$$

Power Measurement – is defined as the total watts used by the SUT during the tests.

Price – We note that while price is an important metric in the context of cloud, this is somewhat temporal, and is clearly a non-engineering metric that may vary from customer to customer. The recommended way to incorporate this aspect is to include the Bill of Materials details as part of a benchmark disclosure. This list of model numbers, quantities, software and support information may be used by the benchmark consumer to construct the end price relevant to them. Inclusion of the exact price details in the benchmark disclosure is not recommended.

Density – measures how many instances of the workload can be run on the SUT before performance degrades below a specified QOS.

- For IaaS, it may refer to the number of virtual machines running on a physical hardware
- For PaaS, it may refer to the number of application servers running on a system.
- For SaaS, it may refer to the number of users that the system can service

V. BENCHMARK MEASUREMENT TOOLS

During the course of this research, we found various tools and frameworks that purportedly tested and measured cloud computing. The following are only some of the tools found in open source or restricted release forms. In spite of availability of tools, there is no one guideline or standards definition for benchmarking cloud services that has been universally accepted.

AEOLUS (Red Hat): Aeolus is an Open source project sponsored by Red Hat, designed as framework to create and manage an on-premise hybrid cloud Infrastructure-as-a-Service (IaaS). It provides self-service computing resources to users in a managed, governed, and secure way. You can deploy and manage applications on any type of server - physical, virtual, and public cloud

BITT (Intel): Intel® Benchmark Install and Test Tool (Intel® BITT) provides tools to install, configure, run, and analyse benchmark programs on small test clusters. It is implemented in python and uses gnu plot to generate performance plots. Intel® BITT currently runs on Linux and been used on Oracle VM, Amazon and various hardware platforms

CloudBench (IBM): IBM's CloudBench (CB) is a meta-benchmark framework designed for Infrastructure-as-a-Service (IaaS) clouds. It automates the execution, provisioning, data collection, management and other steps within an arbitrary number and types of individual benchmarks.

VI. SUMMARY AND NEXT STEPS

To summarize the current state, there are well documented processes and procedures established by various organizations, corporations and committees, and a comprehensive set of information available for public consumption on the benchmark test criteria, basis, execution and results. These data points are suitable and comprehensive, yet technical and inadequate for a consumer to make a decision. From our point of view, this is a one of the primary reasons why individual consumers tend to not embrace cloud to its full potential. However they function at best in isolation. Despite availability of tools, there is no one guideline or standards definition for benchmarking cloud services that has been universally accepted. It is our endeavour to highlight the fact that a simple decision process for a consumer starts by measurable and comparable facts, and for facts to be comparable, they should be measured on a common yardstick. Future work in this direction could be to:

1. Establish a common definition and methodology to measure cloud performance
2. Framework for the Cloud Benchmark - White box & Black box
3. Provide an API layer for extending benchmarking capabilities for all participants in Cloud development – service providers, hardware / software vendors, and consumers

REFERENCES

1. Berkeley RAD Lab Project at http://radlab.cs.berkeley.edu/wiki/Projects/Cloudstone
2. M. Armbrust and A. Fox et al. Above the clouds: A Berkeley view of cloud computing, Technical Report EECS-2009-28, UC Berkeley, 2009.
3. A. Li and X. Yang et al.: Shopping for a Cloud Made Easy, Proceeding of the 2nd USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 10), 2012.
4. T. Liu, Y. Li, and X. Li, A Collaborative Management as a Service Framework for Managing Internetware Systems, First Asia-Pacific Symposium on Internetware (Internetware 2009), 2009.
5. T. Wood and R. Shenoy et al. The Case for Enterprise-Ready Virtual Private Clouds,
6. jgaard-Hansen, K. ; Huan Cong Nguyen ; Schwefel, H. Session mobility solution for client-based application migration scenarios, Wireless On-Demand Network Systems and Services (WONS), 2011, p.76
7. David S. Linthicum, Cloud Computing and SOA Convergence in your Enterprise, Pearson, 2012
8. Mehrdad Mahdavi Boroujerdi, Soheil Nazem, Cloud Computing: Changing Cogitation about Computing, World Academy of Science, Engineering and Technology 58 2011
9. R. Buyya, C. S. Yeo, and S. Venugopal, "Market oriented Cloud Computing: Vision, hype, and reality for delivering it services as computing utilities", in Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications

10. Controlling Data in the Cloud: Outsourcing computation without Outsourcing Control, Richard Chow, Philippe Golle, Markus Jakobsson, Ryusuke Masuoka, Jesus Molina Elaine Shi, Jessica Staddon Parc, CCSW 2009, Chicago, Illinois, USA.
11. "Utility Based Cloud Power," (2011) Computers Journal, www.dirjournal.com
12. Chong, F. and G. Carraro (2011) "Architecture Strategies for Catching the Long Tail," Microsoft Corporation, April 2011
13. Miller, M. (2008) Cloud Computing: Web-Based Applications That Change the Way You Work and Collaborate Online, Indianapolis: Que Publishing, 2008
14. SPEC. (2011) Standard Performance Evaluation Corporation. [Online] http://www.spec.org/index.html
15. Stefan Bouckaert, Stephen C Philips, Jerker Wilander, Shafqat Ur Rehman, Walid Dabbous, Thierry Turletti – Benchmarking Computers and Computer Networks

AUTHORS PROFILE

Prarthana Gupta, She is currently pursuing Masters Degree in Computer Technology and Applications at Shriram Institute of Technology, India under Rajiv Gandhi Proudyogiki Vishwavidyalaya, India. Past experience includes five years of research and industry expertise in the field of Computer Technology, specializing in Cloud Enablement and Application in Heterogeneous Environments, Data Mining and Enterprise Data Modeling. She has over four years of academic experience in teaching graduate students in the subjects of Data Mining, Database Management & Algorithm Design at Mumbai University, India & University of Pune, India. She also specializes in experiential learning methodologies by incorporating global best practices and emerging technologies into everyday learning for her students. She has successfully incorporated the use of technology enabled collaborative learning for two batches at graduate level.

Prof. Brajesh Patel, Head, Computer Science Department, Shriram Institute of Technology, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Jabalpur, India. With great pleasure and deep sense of gratitude, I take this opportunity to express my sense of indebtedness to Prof. Brajesh Patel, Dept. Computer Science & Engineering., SRIT, Jabalpur, for his guidance, encouragement and whole hearted involvement in this paper, without which it would have been difficult for me to complete this work.