INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 7, Issue 3 - V7I3-1467)
Available online at: https://www.ijariit.com

A study on Cloud Benchmarking

Suman M. J.
sumanj.sit19@rvce.edu.in
R V College of Engineering, Bengaluru, Karnataka

Sharadadevi K.
sharadadeviks@rvce.edu.in
R V College of Engineering, Bangaluru, Karnataka

ABSTRACT

Sharing of files is a common functionality in cloud storage systems, which enable many users to concurrently use files in shared folders or workspaces. Current cloud storage systems, on the other hand, can deliver unexpected results to end users in the face of concurrency. Our research's aim is to identify the findings provided by existing cloud storage systems in concurrent environments and make these results publicly available, which could help both cloud storage end-users and service providers. Because of the large number of cloud computing providers and the various services each provides, cloud users need benchmark data that addresses the unique characteristics of the cloud computing world, such as dynamic scaling. Cloud Cmp, Cloud Stone, Hi Bench, YCSB, and Cloud Suite are five common tools that present workloads and/or methods for quantitatively comparing cloud computing offerings.

Keywords– *Cloud computing, Workload, Benchmarking, Performance evaluation*

1. INTRODUCTION

The benefits of the dynamically flexible, pay-as-you-go model have fueled the rise in popularity of cloud computing in recent years. This allows businesses to concentrate on delivering services to their customers by using computing resources as a utility. Organizations eliminate major capital expenditures by eliminating the need for on-premises equipment and instead concentrate resources on quicker implementation. The pay-as-you-go model allows a company to expand naturally in response to consumer demand. Since cloud computing services scale elastically, using it eliminates the risk of over provisioning, wasting resources during off-peak hours, and under provisioning, losing out on potential customers. Start-up success stories like Instagram, which grew its user base to over 150 million in less than four years using only public cloud technologies, demonstrate the potential for rapid growth that cloud computing offers. With so many cloud storage providers and so many different services to choose from, a customer looking for the best option for their business needs benchmark data that clearly addresses the cloud computing environment's unique characteristics. A benchmark should accurately reflect the workload that the user intends to run.

2. BACKGROUND

2.1 Cloud Computing

Cloud computing is a large-scale, distributed computing paradigm that relies on economies of scale to operate. Cloud computing service providers provide external customers with abstracted, virtualized, dynamically scalable, and controlled services on demand through the Internet. Compute, storage, and networking are examples of these capabilities. Cloud computing companies take advantage of economies of scale by assembling large datacenters with tens of thousands of servers serving a diverse customer base. Large-scale operations bear maintenance costs more efficiently due to the advantages of increased equipment usage, bulk discounts on imported equipment, and lower cooling and powering costs. As Internet users produce larger sets of data to be processed, the demand for large-scale computing capacity continues to rise.

2.2 Virtualization

Virtualization is an essential part of cloud computing because it allows you to pool and dynamically allocate hardware resources. A hypervisor is based on a server in a datacenter that acts as a host machine and can run several instances of virtual machines or guest machines at the same time. These virtual machines are operating system instances that are loaded into respective host machines by a separate controlling computer. A cloud storage provider unifies the datacenter's resources into an encapsulated pool that can be distributed and released according to user demand, with the controlling device handling the computing resources of several servers.

2.3 Services

The cloud computing services that providers provide are classified by the NIST into three service models: infrastructure-as-a-service (IaaS), platform-as-a-service (PaaS), and software-as-a-service (SaaS).

- An IaaS offers access to unified services such as computation, storage, and networking in the form of instances. Providers provide ad hoc computing services for a fee dependent on use. These resources are distributed as on-demand instances that are treated similarly to physical hardware. When scaling is necessary, the user is responsible for requesting and initializing new instances.
- A PaaS offers many of the same services as an IaaS, but in an integrated ecosystem that reduces the development burden while also limiting features. PaaS providers have a limited set of computing and storage resources that can be accessed through APIs in a more restricted setting. Many application-specific resources, such as web hosting, data processing, and business analytics, are pre-built and accessible to users.
- SaaS applications, such as e-mail and Google Docs, are special-purpose software services that end users access remotely. They are often installed with PaaS and IaaS software, but the end-user is unaware of the implementation specifics.

2.4 Map Reduce

End-users are taking advantage of cloud computing's large horizontal scaling to process vast sets of data, a service that was historically only open to users with a dedicated datacenter. A parallel processing platform used by several cloud-based batch-processing projects is Apache Hadoop [4], an opensource version of Google's MapReduce [2] and GFS [3].

2.5 Task of Benchmarking

The precise meaning of "best" is determined by the benchmarking purpose and is the first question to be addressed when creating a new benchmark. Florescu and Kossmann propose looking at the properties and constraints of the structures to be benchmarked as a systematic approach to addressing this question [5]. Lower priority properties create restrictions, while the top priority property must be optimized. As a result, a benchmark can be thought of as a way to clearly identify these goals and constraints. The benchmark's aim is to report on how well various systems perform in terms of the optimized priority under the constraints.

3. CLOUD BENCHMARKING TOOLS

3.1 Cloud Cmp

Cloud Cmp is a proposed method for estimating the performance and cost of a legacy application running in the cloud without having to port and deploy the application. Cloud Cmp uses a three-phase approach to achieve this goal: service benchmarking, application workload compilation, and performance prediction. Six cloud providers (including Google App Engine [6], Amazon AWS [7], Microsoft Azure [6], Go Grid [4], and Rackspace [5]) are chosen for the service benchmarking process based on their ability to provide cloud computing services needed for web application creation on the cloud. Access to an elastic compute cluster, persistent storage, intra-cloud networking, and wide-area distribution networking are among the cloud computing services available. The efficiency and cost of each cloud service was measured by completing a set of benchmarking tasks that test each of the cloud computing service's characteristics.

- SPECjvm2008 [6] Java tasks were used to measure the performance of various elastic compute clusters. Because of Java's portability, it was chosen for the tasks. The cost effectiveness of each cluster was calculated by the cost per mission, while the success of each cluster was measured by the completion time of each task.
- Elastic computing cluster scaling – The delay between the time an instance was requested and the time it was ready was used to calculate scaling. Since not all services support scaling through instance request, this metric's applicability is restricted.
- Persistent storage services – The latency to insert or fetch a random to and from a data table was calculated to test the efficiency of a persistent storage service. Table sizes of 1000 and 100,000 entries were used in the experiment. The results revealed that table size and activity had a major impact on success.

Table 1. Comparison of cloud benchmarking tools

	Cloud Cmp	Cloud Stone	Hi Bench	YCSB	Cloud Suite
Cost	Cost per task	Cost per user	Not covered	Not covered	Not covered
Service	IaaS, PaaS	IaaS	PaaS	PaaS	PaaS
Scaling	Latency to allow new instance.	Load balancer – Apache default	None specific	Scaleup, Elastic speed up	None specific
Storage	Latency to fetch a random entry from a pre-defined table.	User's choice of relational table.	Aggregated bandwidth delivered by HDFS.	Adjust possible operations, data size, and distribution to target specific workloads.	Uses YCSB to asses serving systems.
Networking	Intra-cloud-TCP throughput between instances	None specific.	None specific.	None specific.	None specific.
Performance	Latency of various SPECjvm2008 tasks	Response time of request made by load generator	Speed – job running time	Read/Update Latency	Execution cycle profile.
Test Environment	Multiple instance types.	Amazon EC2 instance types	Hadoop cluster	Data serving system	Server

3.2 Cloud Stone

Cloud Stone is a toolkit for analysing a standard social networking website's workload. Cloud Stone's mission is to provide developers with resources to examine various implementation decisions that have an effect on the efficiency and cost of running a

social networking website. Currently, these tools can only be used on a cloud service that supports Amazon EC2 instances. Olio, automation software for running Olio tests, and a technique for computing a suggested metric are the three components of Cloud Stone. Olio includes two full implementations of a social event calendar app and makes use of Faban [7], a time-varying workload generator. Both PHP and Ruby-on-Rails implementations of the application have an identical user interface, enabling a direct comparison of each development stack. Faban simulates multiple users at the same time by running concurrent agents on multiple computers under the supervision of a single central coordinator. During a run, the central coordinator may also adjust the number of active users. Faban also records the latency of each request as well as details on usage. Selecting a configuration for the Olio deployment, selecting a workload profile to be created by Faban, and deploying the instances are all part of running an experiment with CloudStone. The performance of an Olio configuration can vary based on the various tuning mechanisms provided by each implementation, such as database caching, load balancing, and so on. It is proposed that the experiment's findings be expressed in terms of dollars per consumer per month.

3.3 Hi Bench

Hi Bench is a benchmark suite that focuses on Hadoop system components. Hadoop's parallel computing component (MapReduce) and database component are completely exercised by using a variety of practical workloads (HDFS). Microbenchmarks, web search tasks, machine learning tasks, and HDFS benchmarks are among the benchmarking tasks chosen. Sort [7], Word Count [2], and Tera Sort [2] are examples of micro-benchmarks. Sort is intended to represent a type of MapReduce problem that transforms a data set. It simply sorts a large array of data. Word Count, on the other hand, is meant to describe a class that collects a small amount of data from a large data set. Another sorting operation, Tera Sort, but with a larger data set. To build their input datasets, all of the micro benchmarks use Hi Bench software.

3.4 Yahoo! Cloud Serving Benchmark

Yahoo created the Yahoo! Cloud Serving Benchmark (YCSB) to test their PNUTS [2] serving scheme. This benchmark focuses on scalable serving systems that enable data to be read and written. The role of benchmarking serving systems is divided into two levels by YCSB. Tier 1 refers to overall efficiency as determined by a request's latency while the database is busy. The latency of a request is tracked as the throughput is increased to measure the balance of throughput and latency. This is accomplished by examining the effect of adding more machines to the system on its efficiency. The system's ability to scale up well is described by its latency remaining constant over multiple tests with increasing workload and server count.

3.5 Cloud Suite

Cloud Suite is a set of benchmarking tasks that were developed to identify inefficiencies in the microarchitecture of modern server CPUs used in cloud computing. Benchmarking tasks have been described as some of the more common tasks performed by cloud computing. Data serving, MapReduce, media streaming, SAT solving, web hosting, and web search were among the activities.

4. CONCLUSION

Cloud computing allows businesses to scale more effectively to the size of their customer base, giving them a competitive edge if the right services are chosen. We have introduced available benchmarking tools for cloud computing services in this paper. Cloud Cmp proposes a method for evaluating a provider's individual cloud computing services. Cloud Stone is a Web 2.0 application testing platform that includes a social networking application with simulated user interaction. Hi Bench is a tool that gathers practical workloads for MapReduce processing. With generated workloads, YCSB tests the efficiency and scalability of serving systems. Finally, Cloud Suite proposes workloads to model the behaviour of typical tasks in a cloud computing environment.

5. ACKNOWLEDGEMENT

I take this opportunity to thank my guide, Prof. Sharadadevi K., Assistant Professor, Department of Information Science & Engineering, RV COLLEGE OF ENGINEERING, Bengaluru for her valuable suggestions, support and regular source of encouragement and assistance throughout this project. and senior authorities whose constant encouragement made it possible for me to take up the challenge of doing this project.

6. REFERENCES

[1] Li, Ang, et al. "CloudCmp: comparing public cloud providers." Proceedings of the 10th ACM SIGCOMM conference on Internet measurement. ACM, 2010.
[2] Sobel, Will, et al. "Cloudstone: Multi-platform, multilanguage benchmark and measurement tools for web 2.0." Proc. of CCA. 2008.
3] Huang, Shengsheng, et al. "The HiBench benchmark suite: Characterization of the MapReduce-based data analysis." Data Engineering Workshops (ICDEW), 2010 IEEE 26th International Conference on. IEEE, 2010.
[4] Cooper, Brian F., et al. "Benchmarking cloud serving systems with YCSB. "Proceedings of the 1st ACM symposium on Cloud computing. ACM, 2010.
[5] Ferdman, Michael, et al. "Clearing the clouds: a study of emerging scale-out workloads on modern hardware." ACM SIGARCH Computer Architecture News. Vol. 40. No. 1. ACM, 2012.
[6] Hennessy, John L., and David A. Patterson. Computer architecture: a quantitative approach. Elsevier, 2012.
[7] Mell, Peter, and Timothy Grance. "The NIST definition of cloud computing (draft)." NIST special publication 800.145 (2011).

