IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.9634

WORLD OF SERVERS IN AMAZON WEB SERVICES (AWS) CLOUD

R Bharath[1], Dr. R Savitha[2]

[1]Student, Master of Computer Applications, RV College of Engineering, Bengaluru, India.

[2]Associate Professor, RV College of Engineering, Bengaluru, India.

Abstract: Cloud computing is one of the trending technologies in the 21st century. Cloud computing has adapted all the upcoming changes for their sustainability. A decade before in 2006 Amazon web services stepped in the market with Elastic Compute Cloud (EC2) service having a vision to rent the infrastructure for software applications. The application data has to be securely stored using elastic Block storage. Each EC2 instance resembles an independent computing machine. AWS offers these services based on a pay-per-use model. There exists a huge performance difference among on-premises instances and the EC2. The outcome of synopsis is to improve the performance of the EC2 and volume through various techniques such as Monitor Memory Consumption at the OS Level, Selecting the Right Storage and Instance Types, replace a degraded EBS Volume and Manually Kill Leaking Processes.

Keywords: Elastic compute cloud, performance EC2 detection and avoidance

I. INTRODUCTION

It is difficult to stay on top of all the new developments in the digital industry. New developments and inventions appear almost daily, and it can be tricky to stay on top of everything. Cloud computing isn't necessarily a new trend; however, some companies have started using it for their services. Several aspects of daily life have been changed and transformed by this innovative digital solution. Its impact on the data industry and end users cannot be overstated. From early start-ups until established businesses, cloud computing has helped minimize the costs and increase offerings/profits. This is because they no longer require extra hardware and software to procure in order to balance the peak workloads/requests hitting on the applications. Cloud computing refers to the use of hardware and software that are delivered via a network to manage applications and its data. It has gained broad popularity due to the profound problems. Cloud Computing refers to the abstraction of a rather complex infrastructure that enables software, hardware, computation, and remote services to be performed

The servers in AWS Cloud offer the client with the platform for designing, developing and deploying the application. The service providers (AWS) provide flexibility for utilizing the resources based on the client's requirements. The AWS Cloud services serve to provide tailored requirements. The services are instantiated based on demand and provide full control over the instantiated environment. When opting for such a cloud architecture solution, the provisioned infrastructure, platform and software totally depends on the requirement. The cost is directly proportional to the services used. There are various service models which can be utilized based on the client's needs and requirements.

PaaS (Platform as a Service) – It is one of the service models in cloud computing that offers a platform/environment as a service. This environment can be utilized by the developers for developing the application. The platform as a service provides all the dependencies, libraries and software for developing the application.

SaaS (Software as a Service) – SaaS is one of the service models that offers software as a service. The software such as word, excel etc are rented to the client for the usage. The SaaS helps to pay based on the usage. These applications are accessible via network / Internet.

IaaS (Infrastructure as a Service) – IaaS is a prominent service model that offers computing, storage, and networking resources for application development, deployment and execution. These services are on a pay-as-you-go basis model. Infrastructure as a service provides virtual computing systems so-called servers. These machines are used to host and develop applications. These machines are ubiquitous, which means they are accessible anywhere over a network connection (Internet).

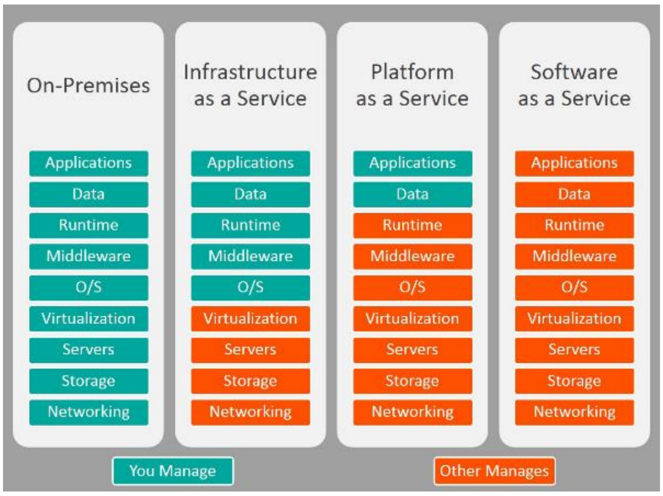

Fig 1: Cloud Service Model and shared responsibility over computing resources

AWS provides a vast variety of services for various requirements. To implement the IaaS, AWS provides a service called EC2 (Elastic compute cloud). Elastic compute offers a virtual server, it could be operated via the AWS management console. It enables clients to host their applications within their environment. Virtual machines (VMs) can be deployed in virtually an unlimited number of instances. AWS leverages the client to host their Elastic Compute Cloud in any public AWS region.

Amazon Web Service provider provides various types of Elastic compute cloud instance types, these instance types are of different families. The Compute capacity differs from families to families. It consists of varieties of configuration of memory, networking resources and CPU to address the client needs.

Often the Elastic compute cloud is created using the templates and these templates are called Amazon Machine Images. These templates are configured with the Operating system environment and few other software's. The Client can select their Amazon Machine Images from the amazon marketplace. Amazon Machine Images provide the initial software configuration of an instance. The client can have Linux AMI, Windows AMI, and Ubuntu AMI etc. as a platform for his server.

There are various types of Instance types which Refers to the hardware capabilities of an instance such as memory optimized, accelerated computing, General purpose, Compute optimized, and storage optimized.

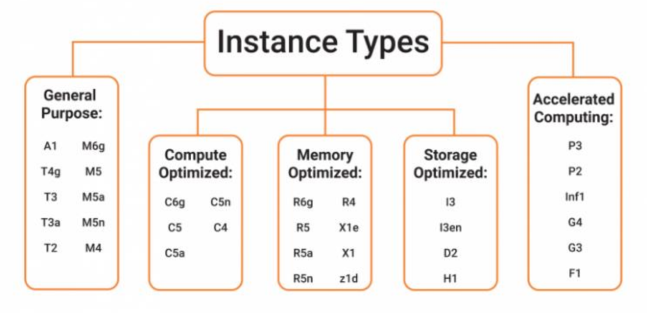

Fig 2: Classification of Instance types in AWS

The above-mentioned Instance types utilized as per the application requirements. The performance of each Instance type differs.

1. The General-Purpose Instance type provides the instance that consists of computing, memory and networking resources at a balanced quantity/equal quantity and could server large set of applications. A1, T4g, M6g, M5, M5a, M5n, T3, T3a, T2, and M4 instance types belong to it.

2. Compute optimized types are used at compute intense workload such as batch processing, data analytics, and machine learning, The compute-optimized instances provide high performance with minimal cost.

3. Storage optimized instance type provides Instances that are memory optimized and deliver excellent performance for applications processing large data sets in memory

4. The accelerated computing instances employ the co-processors or the hardware accelerators to execute some functions that are graphics processing, pattern matching, decimal number calculations etc.

DOI: 10.17148/IARJSET.2022.9634

II. LITERATURE SURVEY

Several related works have already been done in the field of Cloud Computing with respect to AWS and implemented till now. Some of them are:

Due to their extensive use by individual users, researchers, and businesses for their daily work, public cloud monitoring and analysis is gaining traction. presenting an algorithm for effectively resizing a collection of operating Amazon EC2 instances to optimize cost and utilization. The Cost and Usage Optimization process takes data about the Elastic compute cloud instances such as number of servers, instance types and utilization of CPU and suggests a new set of the Elastic compute cloud instances that relatively helps the client to reduce their cost and increase their performance. Smart cloud Monitoring tool was developed for gathering and analysing monitoring data from Amazon Web Services [1]

The Cloud providers such as Amazon web services offer Elastic block storage volume discount to the client that has a huge Elastic compute cloud instance reservation over that particular time slot. There are some tasks that are delay intense that means they often don't tolerate any latency. The above situation provides the cloud brokerage a platform to schedule the volumes for the delay tolerant jobs. The concept behind the scheduling is to create numerous task bundles, each with a discount. Let's consider the model, in which each job has the same processing and delay-tolerant time, and then we suggest a dynamic programming solution. The model is then extended to a heterogeneous model, in which the job processing time and work deadline might have any value. [2]

In recent years, the amount of available information has increased due to an increase of information delivery and data storage solutions via broadband networks. The goal of this research paper is to collect the resources that are not sufficient from cloud systems while performing tasks at a local system. However, the scalable resources can be done in their own local cluster and dynamically acquiring of resources can be done via cloud computing. performing compute intense operations would decrease the execution time while executing the tasks over compute optimized instances. Moreover, they have measured basic performance on Amazon Elastic Compute Cloud as basic data of a real cloud computing system. [3]

III. TECHNICAL SIGNIFICANCE

Issues in Current EC2:

There are many issues that were identified in this field and still the process continued however there are few unresolved issues in EC2 (Elastic compute Cloud).

- Uncertain EBS Disk I/O
- Mismatch of ECU and extensive utilization of CPU
- Ran out of Elastic compute cloud Instance Memory
- Load balancer traffic latency
- Interruption due to AWS maintenance

AWS Cloud IOPS abbreviated as input/output operations per second. Standard EBS volumes in AWS cloud can deliver 100 IOPS. If you have purchased 4,000 IOPS per volume, provisioned IOPS volumes can supply that throughput. You can predict the volume to produce between 90% and 100% of its provisioned IOPS 99.9% of the time over a given year, only when these conditions are met.

Using appropriate instance types that suit the application size.

it is often observed through the average queue length, the application delivers enough requests to the volumes. read and write operations that are applied on 64kb or less should expect ¼ of the supplied IOPS. only if the block size is 64KB.

These stringent requirements need to come from a networked storage service, yet they are still somewhat limiting. Sometimes the input/output operations land up in waiting for the CPU to execute the transactions when the IOPS reaches the limit, and there is a drastic increase in Volume Queue Length. The operating system can see this delay through several I/O related indicators. (e.g., percentage of CPU spent waiting for input/output operations). At that moment, your application will most likely run at the same speed as the EBS volumes.

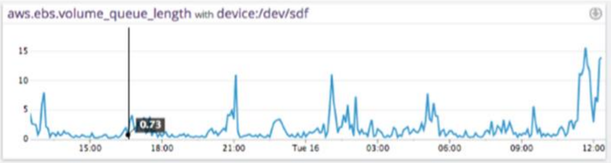

Fig 3: Volume Queue Length for an EBS Volume.

Why It has Occurred:
Two Main reasons are:

1. Due to the extreme slow of the Elastic Block Volume
2. Due to the sharing of the storage devices over the network.

The above reason can be easily overcome by replacing the solid-state disc with provisioned IOPS volumes that provide better performance. however, we can avoid expecting 100 IOPS from normal EBS volume that can be enhanced through RAID pool EBS volumes.
Latency and IOPS are, for the most part, tightly connected on a dedicated storage system. Latency will gradually grow as you raise the number of IOPS until the storage bus or drives themselves are saturated. Pushing more IOPS past the threshold of saturation just generates a backlog in front of the storage system (usually at the operating system layer).

Problem Avoidance and Resolutions
1. Choosing the correct Instance types and the storage based on the application:- The Client application that is hosted in the Elastic compute cloud puts on data to the storage. Therefore it is required to attach the volume with the EC2.however this will not avoid network constraints that would arise from various other sources. it would often reduce the input and output mismatch that are mostly occurred by misconfigurations.
2. Utilizing the instance storage rather than EBS:- Instance storage produces higher input/output performance rate compared to the other Elastic Block storage volumes. Using the instance storage would drastically enhance the performance of the IO operations. Instance storage is another version of storage solution that stores within the server. Earlier The instance block is not flexible/versatile because it restricts instance storage capacity but now the instance storage block can be expanded based on the end user requirement.
3. Prime your EBS volumes: - The initial access to almost any block on EBS will be at a 50 percent throughput in IOPS, thus if you want maximum performance instantly, make sure all blocks have been visited once.
4. Buying of Provisioned IOPS: - AWS offers the client to purchase provisioned IOPS that guarantees to transmit the application input and output operations over the network with the sufficient IOPS. The AWS limit to reserved instances because of the storage in the AWS Provisioned IOPS packages.
5. Replace a degraded EBS Volume: - Arrange the Elastic Block storage volume according to the RAID. However the architect can replace or remove the non-functional volumes. The decommissioned volume snapshot is taken for data backup and later the snapshot can be used to instantiate a new Elastic Block Storage volume. It would consume more bandwidth and increase the storage latency.

IV. TOOLS AND TECHNOLOGIES USED

SOFTWARE
GitBash 2.36.1v is a tool that provides a Linux terminal for the Windows machine. This terminal is very useful in order to connect to the Elastic Compute Cloud. The Key pair which is a private key is in the format of .ppk has to be converted into .ppm. GitBash is such a tool which accepts .ppk format file, converts to .ppm format and allows to connect to remote servers or remote Elastic Compute Cloud Instances. GitBash also list the window file system in the Linux terminal and also allows for easy navigation through Linux commands.
Putty :- Putty is one of the most utilized tools for implementing SSH over a windows machine. It is extensively used for connecting to remote servers. Not only does it support SSH, it provides extended support to various other protocols. Putty is also used for generating keys such as Hash keys. Putty enables the communication from the Linux server to non - Linux servers.

V. CONCLUSION

This paper presents various instance types and their performance issues. Each has its own advantages and drawbacks. Depending on what level of security, the client needs, The Application demands one has to make the right choice. Elastic Compute Cloud and volumes must face problems with the performance, Uncertain EBS Disk I/O. But if one can go through this problem, it can provide a very good performance and efficient utilization. People can avoid the problem by Selecting the Right Storage and Instance Types, Purchase Provisioned IOPS and Replace a degraded EBS Volume. Hence It will improve the standard of the utilization. The volume of various types can be used to optimize the volume Queue length with the available IOPS.

VI. ACKNOWLEDGEMENT

The satisfaction that accompanies the success of any work would be incomplete unless we mention the name of the person who made it possible.
I express my sincere thanks and wholehearted credit to our guide **Dr. R Savitha**, Assistant Professor, Department of MCA, R.V. College of Engineering ®, Bengaluru for her constant encouragement, support and guidance during the project work.

REFERENCES

[1]. Ning Wang and Jie Wu “M Optimal Cloud Instance Acquisition via IaaS Cloud Brokerage with Volume Discount" in IEEE/ACM 26th International Symposium on Quality of Service (IWQoS). Date Added to IEEE Xplore: 24 January 2019 DOI: 10.1109/IWQoS.2018.8624186.

[2]. S. Namasudra “Cloud computing: A new era” in Journal of Fundamental and Applied Sciences. eISSN: 1112-9867 Vol. 10 No. 2 (2018) Published 2018-05-29.

[3]. Okwoli Mercy Enefu “G The Adoption of Cloud Computing Technology for Library Services in the National Open University of Nigeria Library” eISSN: 1596-5422 Vol. 15 No. 1-2 (2015) Published 2016-10-03.

[4]. Edje E. Abel “A survey on the utilization of cloud computing services for academic learning” in Journal of the Nigerian Association of Mathematical Physics eISSN: 1116-4336 Published 2020-06-02 Vol. 49 No. 1 (2019)..

[5]. S.A. Akinboro, U.J. Asanga and M.O. Abass “Privacy enforcement on subscribers data in cloud computing” eISSN: 2467-8821 Published 2021-10-18 Vol. 40 No. 2 (2021) DOI: 10.4314/njt.v40i2.17.

[6]. M Xingwang Huang Lingqing Chena and “Improved firefly algorithm with courtship learning for unrelated parallel machine scheduling problem with sequence- dependent setup times” in Journal of Cloud Computing 11, Article number: 9 (2022).

[7]. Ezugwu AE, Akutsah F (2018) An improved firefly algorithm for the unrelated parallel machines scheduling problem with sequence-dependent setup times. IEEE Access 6:54,459–54,478.

[8]. Chang PC, Chen SH (2011) Integrating dominance properties with genetic algorithms for parallel machine scheduling problems with setup times. Appl Soft Comput 11(1):1263–1274.

[9]. Arnaout JP (2020) A worm optimization algorithm to minimize the makespan on unrelated parallel machines with sequence-dependent setup times. Ann Oper Res 285(1):273–293.

[10]. Jovanovic R, Voß S (2021) Fixed set search application for minimizing the makespan on unrelated parallel machines with sequence-dependent setup times. Appl Soft Comput 110:107,521.L. H. Iwaya, A. Ahmad and M. A. Babar, "Security and Privacy for mHealth and uHealth Systems: A Systematic Mapping Study," in IEEE Access, vol. 8, pp. 150081- 150112, 2020, doi: 10.1109/ACCESS.2020.3015962.

[11]. Qiang S (2020) A hybrid multi-objective teaching-learning-based optimization algorithm for unrelated parallel machine scheduling problems. Control Theory Appl 37(10):2242–2256