An Introduction to AWS – EC2 (Elastic Compute Cloud)

Rahul Saini[1], Rachna Behl[2]

[1]*Research Scholar, MRIIRS Faridabad, Haryana, India*
[2]*Asstant Professor, MRIIRS Faridabad, Haryana, India*
[1]rahulhbc@yahoo.in , [2]rachnabhel.fet@mriu.edu.in

Abstract—**Most recent two decades IT is following conventional methodologies for dealing with its framework since the beginning of Cloud administrations, for example, Amazon Web Services, Google Cloud Platform, IBM Cloud and a lot all the more giving on the web foundation administrations. Associations have acknowledged an adjustment in their working model, if an association searches for an extension to expand its registering power, they just get it online by starting a virtual machine on the cloud. Virtual machines can be immediately propelled and shutdown through programming interfaces, offering adaptability to the client rather than customary methodologies.**

There were times when one was restricted by the boundaries of a machine e.g.; a data scientist has a large scale of data and would like to perform some analysis, however, encounters an error as below while uploading this data.

> *data <- read.csv (file='no_of_bug.csv')*

Error: cannot allocate vector of size 500.0 Mb

The error is the result of the unavailability of the RAM since the operating system does not have enough RAM available as there are many solutions available of such problems following the traditional approach one should upgrade the RAM of the machine with the restriction to the RAM up-gradation limit to the machine's compatibility. However, there is another way by introducing the concept of virtualization in the could with more RAM and CPU.

This paper explores the concept of Virtual machines on AWS EC2 as they are called instances with many advantages few of them includes being highly scalable (one can choose instances with more CPU and RAM etc.) most reliable and scalable easy to use for users as pay for what being used also allows a selection of different platform (OS) as an outcome of this paper will create an understating of the working fundamental of AWS EC2.

IndexTerms—**Amazon EC2, Cloud Computing, Elastic Compute Cloud**

I. INTRODUCTION

ITS impression has been altogether changed since the distributed computing idea being presented as a pay-more only as costs arise administrations. To have an application utilizing Amazon Web Service's (AWS) Elastic Compute Cloud carries numerous progressions to the improvement, sending and upkeep forms as these progressions might be extremely gigantic for certain associations anyway EC2 guarantees expanded adaptability, simplicity of organization, versatility and the huge help for the outsider administrations in contrast with the conventional methodologies.

A couple of years back a help administrations association expected to have and keep up its Incident Management System with an immense venture including programming buys. As the Cloud processing stage has caught the market so broadly now association can undoubtedly move their product benefits all the more safely to the web.

Right now, will dive into the working key of AWS's Elastic Compute Cloud (EC2) as the working basics are fundamentally unique concerning the conventional methodologies, for example, inhouse framework the board that is generally acknowledged anyway these distinctions may prompt inventive sort of changes and requires various instruments to pick up perceivability basic the cloud base foundation.

National Institute of Standard and Technology defined Cloud Computing:

"Cloud Computing concept enables on-demand network access to a shared pool of configurable computing resources such as (e.g. servers, storage, application, network, and services) that can be rapidly conducted and released with minimal management effort or interaction with the service provider. This cloud model promotes availability and is composed of five essential characteristics, three service models and four deployment models." [3]

II. WHAT IS AMAZON EC2

Foundation as-an administration customers are offered a wide assorted variety of Cloud assets from numerous, circulated Cloud suppliers (e.g., Amazon EC2 [5], and RackSpace [6], IBM Smart Cloud[7]) provided at various hourly cost rates. Besides, comparable Cloud assets might be valued diversely by each Cloud supplier. Besides, Cloud buyers may demand heterogeneous arrangements of Cloud assets that may not be accessible in a solitary Cloud supplier. Along these lines, self-rulingly completing asset assignment from numerous and self-intrigued Cloud suppliers while testing hourly cost rates related to Cloud assets is important to give proficient (i.e., with low portion costs) asset allotment administrations to purchasers in a unique way. This burdens the requirement for the operator worldview. Operators are self-governing issue solvers that can demonstrate and work together deftly and self-interestedly among one another.

A famous methodology in cloud-based administrations is to permit clients to make and offer virtual pictures with dif-

ferent clients. For instance, a client who has made an inheritance Linux Debian Sarge picture may choose to make this picture open so that different clients can undoubtedly re-use it.

Notwithstanding user shared pictures, the cloud specialist organization may likewise give redone open pictures dependent on the regular needs of their clients (e.g., a Ubuntu web server picture that has been pre-arranged with MySQL, PHP and an Apache). This permits the clients to just launch and start new servers, without the problem of putting in new programming themselves.

Amazon Elastic Compute Cloud (AWS EC2) provides scalable computing capacity in the Amazon Web Services cloud. Using AWS EC2 eliminates the need to invest in hardware so that one can develop and deploy applications faster. Amazon EC2 can be used to launch as many or required virtual machines as per need with security configuration and networking and storage management.

For example, AWS powers the biggest website in the world which is Netflix example.
- EC2 is one of the most popular of AWS offering
- It mainly consists in the capacity of:
  - Renting virtual machine (EC2)
  - Storing data on virtual devices (EBS)
  - Distribute load across machines (ELB) Elastic Load Balancer
  - Scaling the services using an auto-scaling group (ASG)
- Renting a virtual server on cloud is the key to understand how the cloud works

In this thesis, we are going to learn more about Amazon Web Services EC2.

EC2 is the main service to provide machines on-demand and nearly all AWS accounts use EC2 instances to a various extent knowing EC2 is the key thing to understand how the cloud works.
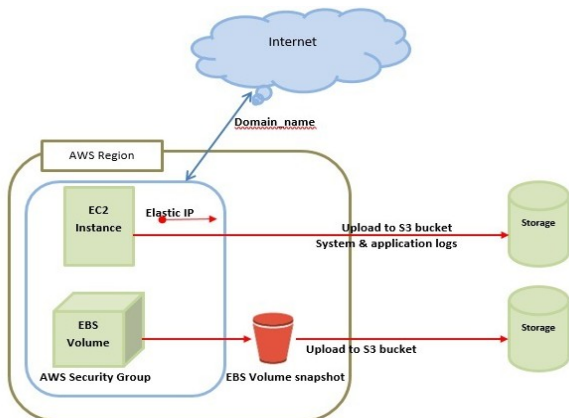

Fig 1: Simple Storage Services (S3) Diagram

Amazon Elastic Compute Cloud (Amazon EC2) gives an adaptable processing limit in the Amazon Web Services (AWS) cloud. Utilizing Amazon EC2 takes out your need to put resources into equipment in advance so that you can create and convey applications quicker. You can utilize

Amazon EC2 to dispatch the same number of or as not many virtual servers as you need, design security and organizing, and oversee capacity. Amazon EC2 empowers you to scale up or down to deal with changes in necessities or spikes in prominence, lessening your need to gauge traffic.

EC2 permit clients to utilize virtual machines of various setups according to their prerequisite. It permits different setup alternatives, mapping of individual servers, different valuing choices, and so on.

We will talk about these in detail in the AWS Products area. Following is the diagrammatic portrayal of the design.

It permits clients to store and recover different sorts of information utilizing API calls. It doesn't contain any figuring component.

Amazon EC2 (Elastic Compute Cloud) is a web administration interface that gives a resizable figure limit in the AWS cloud. It is intended for engineers to have unlimited authority over web-scaling and processing assets.

EC2 examples can be resized and the number of cases scaled up or down according to our prerequisite. These occasions can be propelled in at least one geological areas or districts, and Availability Zones (AZs). Every area involves a few AZs in particular areas, associated with low inertness arranges in a similar locale.

III.   IN-DEPTH WITH AMAZON ELATIC COMPUTE CLOUD

A、   Classification of Elastic Compute Cloud

Amazon EC2 gives a wide determination of case types advanced to fit distinctive use cases. Occurrence types involve differing blends of CPU, memory, stockpiling, and systems administration limit and give you the adaptability to pick the fitting blend of assets for your applications. Each occurrence type incorporates at least one example sizes, permitting you to scale your assets to the prerequisites of your objective remaining task at hand.

1. Different type of EC2 instances

Amazon has thought of a wide scope of Instances that are intended to satisfy the wide assortment of necessities of an

	Type	Description	Mnemonic
General Purpose	A1	Good for scale-out workloads, supported by ARM	a is for ARM processor - or as light as A1 steak sauce
	T-family: T3, T3a, T2	Burstable, good for changing worloads	t is for tiny or turbo
	M family: M6g, M5, M5a, M5n, M4	Balanced, good for consistent workloads	m is for main or happy medium
Compute Optimized	C-family: C5, C5n, C4	High ratio of compute to memory	c is for compute
Memory Optimized	R-family: R5, R5a, R5n, R4	Good for in-memory databases	r is for RAM
	X1-family: X1e, X1	Good for full-in memory application	x is for xtreme
	High memory	Good for large in-memory databases	High memory is for high memory
	Z1d	Both high compute and high memory	z is for zippy
Accelerated Optimized	P-family: P3, P2	Good for graphics processings and other GPU uses	p is for pictuers
	InF1	Support machine learning inference applications	inf is for inference
	G-family: G4, G3	Accelerate machine learning inference and graphics-intensive workloads	g is for graphics
	F1	Customizable hardware acceleration	f is FPGA or feel as in hardware
Storage Optimized	I-family: I3, I3en	SDD-backed, balance of compute and memory	I is for IOPS
	D2	Highest disk ratio	d is for dense
	H1	HDD-backed, balance of compute and memory	H is for HDD

Fig 2: Different types of EC2 instances

association. We have recorded distinctive AWS EC2 occurrence Types underneath with the highlights. How about we find a workable pace each EC2 example type in a superior way.

This diagram shows a brisk rundown of the EC2 instances, below there is a short description of each category of the EC2 instances.

2. Issues with EC2 performance

I. EC2 instances can be optimized with effective utilization and effective services used however in how the resources are delivered and utilized, EC2 instances can be optimized.

II. This is the original EC2 architecture, you see everything is on the server, Network, Storage, Management, security, and monitoring which means it is taking the resources from the server.
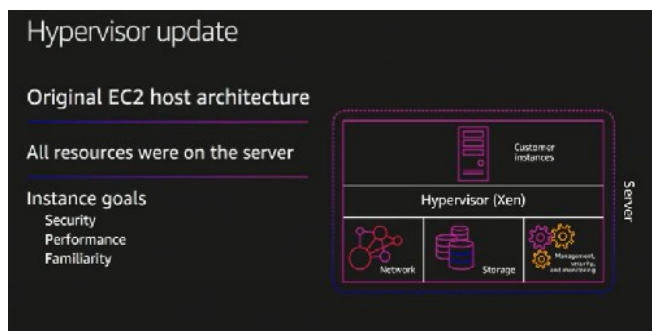

Fig 3: EC2 Architecture

That means that we can utilize the resources in such a way where we can make some of the instance free.

"Launching new instance and running tests in parallel is easy however there is no substitute for measuring the performance of your full application."

III. Choosing a right AMI plays an important role in the performance of the EC2 instances since the AMI includes such information about the launching environment of the instance that may disrupt the performance of your EC2 instance such as template of the root volume which includes (operating system, application), launching permission that control which AWS account to be used and block device mapping that specifies volume to attached to the instance.

Some of the tips to choose the right AMI and OS, choose the latest OS level your tool or application supports
The kernel should be at 3.10 or higher

- As much as 40% performance management
- Should not be using a 2.6 or older kernel

Minimum recommended OS*
- The most recent version of Amazon Linux 2 or Amazon Linux AMI
- Ubuntu version 16.04 or latest LTS release provided by AWS
- Red Hat Enterprise Linux version 7.4
- CentOS 7 version 1708_11
- SUSE Linux Enterprise Server 12 SP2

- FreeBSD 11.1 or later (does not support F1 instances)

IV. Processor state control also important to improve the performance of the EC2 instances there are two types of state control instances

C-state - entering the deeper idle state, allows active cores to achieve higher clock frequency, but the deeper idle state requires more time it exit may not be appropriate for latency-sensitive workloads, Windows: no option to control c-state

P-state - controls the CPU's ability to change frequency, including enabling or disabling Turbo boost

V. Amazon Elastic Compute Cloud (AWS EC2) provides scalable computing capacity in the Amazon Web Services cloud. Using AWS EC2 eliminates the need to invest in hardware so that one can develop and deploy applications faster. Amazon EC2 can be used to launch as many or required virtual machines as per need with security configuration and networking and storage management.

3. Challenges with EC2 instances

We have studied EC2 instances performance depends upon multiple factors in the cloud architecture. Our objective is to identify the factors which may impact the performance of the instances.

- Storage Latency

The problem that we faced more frequently with the EC2 instances such as Storage latency, In AWS EC2 there are two types of storage volume in Elastic Block Storage (EBS) as Standard volumes and Provisioned volume the basic difference between these volumes is provisioned volume perform much faster than the standard volume for e.g: an operation demands a throughput at the rate of 10,000 IPOS (input/output per second), considering the several bands only provisioned volume can get you that.

Increasing the IPOS may lead to other issues such as EBS do not support the given rate of the IPOS, the tasks many begin to queueing up or application may get fail.

To handle such issues we should keep track of IPOS, there is a feature in AWS called VolumeQueueLength which tracks the number of IOPS operations in the queue.

- EM2 Memory Leaks

There is a high probability that instances may run out of memory. The chances are high as EC2 instances may not include the swap volumes resulted unable to free the memory with applications at the time it is being allocated to the new application. Such cases can be considered a good example during the heavy traffic on the application, the instances may freeze the memory while the application is going through with the peak load, therefore, the purpose of EC2 instances are defeated in the first place.

Such scenarios are difficult to handle however adding up more EC2 instances may get beneficial to optimize the performance of your application, however, on the other hand, killing these processes may help however increases the risk or a failover of your application as a vital process may get killed. Adding more EC2 instances could be expensive however this is one of the better ways of doing it.

• Idle EC2 Instances

At launch all the EC2 instances are idle or the scenario could be anything for e.g: an instance having an average CPU utilization less than 2% in last week or the network I/O is less than 10 MB last week so defining a threshold of an instance to consider an idle instance depends upon the configuration of the instance or the need of your organization.

There is a tool that has been launched with the new features that provide the recommendations to optimize the EC2 instances. The main purpose of the tool to monitor the utilization of your EC2 instances upon defining the thresholds of the instances any breaches to the set values such as CPU utilization is lower than 1%, the toll suggests to terminate the instance since this is no longer active.

IV. Conclusion

This paper elaborates an overview of Amazon EC2 instances which cover the types of EC2 instances as described in Fig:2 and major performance issues with EC2 instaces are effective utilization, creating and using an EC2 instances AMI plays vital role in performance of an instances.

Challenges with EC2 instances is explained, a practitioner to consider which contributes effectivtly to improve the performance of EC2 instacnes.

References

[1] Alexis Lê-Quô, Mike Fiedler, and Carlo Cabanilla, "AWS EC2 Performance Problems," Datadog Inc., March 2013.

[2] Marco Balduzzi, Jonas Zaddach, Davide Bazartti, Engin Kirda, "A Security Analysis of Amazon's Elastic Compute Cloud Services", EURECOM: 2012.

[3] M. SEAGRAVE, "How cloud computing shaping the IT job market," Guardian, https://careers.guardian.co.uk/careers-blog/it-job-market-cloud-computing, accesses, April 2012.

[4] B. Simpson, et al, "Title of paper goes here if known," unpublished.

[5] Amazon elastic compute cloud, May 2011. http://aws.amazon.com/security

[6] Cloud computing, cloud hosting and online storage by rackspace hosting, May 2011. http://www.rackspace.com/cloud/

[7] Ibm smartcloud, May 2011.

[8] https: //www.ibm.com/cloud-computing/us/en/#!iaas

[9] https://aws.amazon.com/about-aws/

[10] https://www.visual-paradigm.com/guide/cloud-services-architecture/what-is-aws-architecture/

[11] Cloud computing, cloud hosting and online storage by rackspace hosting, May 2011. http://www.rackspace.com/cloud/

[12] M. Young, *The Technical Writer's Handbook*, Mill Valley, CA: University Science, 1989.