



विश्वजीवनमृतं ज्ञानम्

**ATAL BIHARI VAJPAYEE- INDIAN INSTITUTE OF
INFORMATION TECHNOLOGY AND MANAGEMENT
GWALIOR**

CASE STUDY

On

**AMAZON ELASTIC COMPUTE
CLOUD
(EC2)**

Submitted To:

Dr. Neetesh Kumar

Submitted By:-

Amal Shaji (2017BCS-010)

Sandarbh Yadav (2017BCS-027)

Vaibhav Garg (2017BCS-038)

1. Explain the component of Amazon Elastic Compute Cloud (EC2) architecture.

Ans:

Amazon Elastic Compute Cloud (Amazon EC2) provides scalable computing capacity in the Amazon Web Services (AWS) cloud

The **key components** of Amazon EC2 are:

- Instances
 - Instances are basically servers that are hosted in AWS cloud using EC2 service.
- Images
 - AMI provides template with OS and application pre-configured to reduce the instance deployment.
 - Bundle Tasks – used for bundling windows instance to create a windows instance store-backed AMI
- Elastic Block store-
 - A block storage system used to store persistent data
 - It has three types of volume, i.e. General Purpose (SSD), Provisioned IOPS (SSD), and Magnetic
- Network & security
 - A firewall that enables you to specify the protocols, ports, and source IP ranges that can reach your instances using security groups.
- Load balancing
 - Load Balancing distributes the incoming application or network traffic across multiple targets, such as Amazon EC2 instances, containers, and IP addresses, in multiple Availability Zones.
- Auto scaling
 - AWS Auto Scaling lets you build scaling plans that automate how groups of different resources respond to changes in demand
- Monitoring
 - Amazon CloudWatch monitors your AWS resources and the applications you run on AWS in real time. You can use CloudWatch to collect and track metrics
 - AWS publishes events for each region for any news, update, maintenance or disaster

- System manager
 - AWS Systems Manager is a collection of capabilities for configuring and managing your Amazon EC2 instances, on-premises servers and virtual machines, and other AWS resources at scale
- Region & availability zone
 - Amazon EC2 is hosted in multiple locations worldwide. These locations are composed of regions and Availability Zones. Each region is a separate geographic area. Each region has multiple, isolated locations known as Availability Zones.

2. What are the current pricing and configuration available for GPU computing at EC2?

Ans:

Currently, in amazon EC2 there are different instances of GPU computing on demand which range from low-tier to enterprise-level.

The different instances available at EC2 now are:

p3.2xlarge, p3.8xlarge, and p3.16xlarge have 2.3 GHz (base) and 2.7 GHz (turbo) Intel Xeon E5-2686 v4 processors.

p3dn.24xlarge has 2.5 GHz (base) and 3.1 GHz (sustained all-core turbo) Intel Xeon P-8175M processors

Instance	GPUs	vC PU	Mem (GiB)	GPU Mem (GiB)	GPU P2P	Storage (GB)	Dedicated EBS Bandwidth	Networking Performance
p3.2xlarge	1	8	61	16	-	EBS-Only	1.5 Gbps	Up to 10 Gigabit
p3.8xlarge	4	32	244	64	NVLink	EBS-Only	7 Gbps	10 Gigabit
p3.16xlarge	8	64	488	128	NVLink	EBS-Only	14 Gbps	25 Gigabit
p3dn.24xlarge	8	96	768	256	NVLink	2 x 900 NVMe SSD	19 Gbps	100 Gigabit

Instance	GPU s	vCPU	Mem (GiB)	GPU Memory (GiB)	Network Performance
p2.xlarge	1	4	61	12	High
p2.8xlarge	8	32	488	96	10 Gigabit
p2.16xlarge	16	64	732	192	25 Gigabit

Instance	AWS Inferenia Chips	vCP U	Mem (GiB)	Inferentia Chip-to-Chip Interconnect	Storag e (GB)	Dedicated EBS Bandwidth	Network Performance
inf1.xlarge	1	4	8	N/A	EBS-o nly	Up to 3.5 Gbps	Up to 25 Gigabit
inf1.2xlarge	1	8	16	N/A	EBS-o nly	Up to 3.5 Gbps	Up to 25 Gigabit
inf1.6xlarge	4	24	48	Yes	EBS-o nly	3.5 Gbps	25 Gigabit
inf1.24xlarge	16	96	192	Yes	EBS-o nly	14 Gbps	100 Gigabit

	Instance	GPU s	vCPU	Mem (GB)	GPU Memory (GB)	Instance Storage (GB)	Network Performance (Gbps)
Single GPU VMs	g4dn.xlarge	1	4	16	16	125	Up to 25
	g4dn.2xlarge	1	8	32	16	225	Up to 25
	g4dn.4xlarge	1	16	64	16	225	Up to 25
	g4dn.8xlarge	1	32	128	16	1x900	50
	g4dn.16xlarge	1	64	256	16	1x900	50
Multi GPU VMs	g4dn.12xlarge	4	48	192	64	1x900	50
	g4dn.metal*	8	96	384	128	2x900	100

Instance	GPU s	vCPU	Mem (GiB)	GPU Memory (GiB)	Network Performance
g3s.xlarge	1	4	30.5	8	Up to 10 Gigabit

g3.4xlarge	1	16	122	8	Up to 10 Gigabit
g3.8xlarge	2	32	244	16	10 Gigabit
g3.16xlarge	4	64	488	32	25 Gigabit

3. Write down the stepwise procedure to access the Amazon EC2 services? By running a program on AWS: Matrix Multiplication on 1, 2 and 4 VMs.

Ans:

Stepwise procedure to access Amazon EC2

Step 1: Create a key pair

- a. Open AWS management console and click Create Key pair
- b. Name the keypair and click on Create

Step 2: Enter the CodeDeploy Console

- a. In the console click Get Started Now
- b. Select Sample Deployment and click Next Step

Step 3: Launch a VM

- a. Configure the instance setting with options from operating system, instance type, key pair name, tag key and value respectively.
- b. Click launch instances

Step 4: Name the application and review.

- a. In the *Application Name* box, enter *HelloWorld* as the name for your sample application and click Next Step.
- b. Review information about your application revision, such as its location and description.
- c. Click Next Step

Step 5: Create a Deployment Group

- a. The *Key* and *Value* columns should be auto populated with the values from Step 1. And click Next Step

Step 6: Create a service role

Step 7: Deploy the application

- a. For deploying more than one instances(VM's) we need to modify deployment configuration in custom deployment configuration
- b. Fill the details and click Next step
- c. Review the details of deployment and click deploy now

Step 8: For connecting more than one ec2 instances

- a. Click on EC2 Service on AWS console
- b. In the left panel, click on *Security Groups*
- c. Click on the button *Create Security Group*
- d. Put-in the *name* and *description* of your choosing
- e. Click on the tab *Inbound* and click on *Add Rule*(here, you are adding which port should accept connections.)
- f. Set the following details: -
 - Type: Custom TCP Rule
 - Protocol: TCP
 - Port Range: 1000 [Or any other port on which your program runs]
 - Source: External IP from where *Program1* can be accessed. It can be "Everywhere", "My IP" or a "Custom IP"
- g. Click on the tab *Outbound* and click on *Add Rule*
- h. Repeat Step f, if you want outbound communication.
- i. Repeat these steps on Instance2 and you will be good to go.
- j. Clean up your instances to avoid further charges
- k. Select the boxes of each Amazon EC2 instance to terminate. Select Actions, Instance State, and click Terminate.

4. Discuss the concept of auto-scaling in AWS.

Ans:

Auto scaling in AWS allows user to scale their Amazon EC2 instances up or down automatically as per the instructions set by the user. Parameters like minimum and maximum number of instances are set by the user. Using this, the number of Amazon EC2 instances you're using increases automatically as the demand rises to maintain the performance, and decreases automatically as the demand decreases to minimize the cost.

Auto Scaling is particularly effective for those applications that fluctuate on hourly, daily, or weekly usage. Auto Scaling is enabled by Amazon CloudWatch and is available at no extra cost. AWS CloudWatch can be used to measure CPU utilization, network traffic, etc. AWS Auto Scaling monitors your applications and automatically adjusts capacity to maintain steady, predictable performance at the lowest possible cost.

We can activate the auto-scaling feature using AWS Management Console, Command Line Interface (CLI), or SDK. Auto Scaling helps you ensure that user has the correct number of Amazon EC2 instances available to handle the load for the application. Amazon EC2 Auto Scaling helps maintaining the application availability through fleet management for EC2 instances, which detects and replaces unhealthy instances.

5. Explain about Amazon EC2 Two New Allocation Strategies: On-Demand Prioritized List, and Lowest Price.

Ans:

Amazon EC2 simplifies the provisioning of EC2 capacity across different instance types, Availability Zones (AZs) and across On-Demand, Reserved Instances (RIs) and Spot purchase models, optimizing scale, performance and cost. Allocation strategies lets you determine how EC2 should select from instances and AZs you have specified to fulfil the desired capacity. Until today, you had the option of selecting the single cheapest Instance Pool for Spot and

On-Demand, or diversifying Spot Instances across multiple instance types and AZs in an EC2.

Now, you can use Prioritized List to specifically determine the order in which EC2 attempts to fulfil you're On-Demand capacity. EC2 will attempt to launch all capacity using the instance with the highest priority first, and if all your capacity cannot be fulfilled using your highest priority instance then EC2 will attempt launching capacity using the second priority instance type. You can define a priority for all instances you specify in your EC2.

Now, you can also balance desired cost and availability in the context of applications by directing EC2 to evenly deploy Spot capacity across the N lowest priced Instance Pools. For example, if you are running batch processing, you may prefer to set N to just 2 to maximize savings while attempting to ensure your queue always has compute capacity. However, if you are running a web service you may set N to 10 to minimize the impact of any Spot Instance Pool becoming temporarily unavailable.

6. What is AWS IoT Device Defender? explain

Ans:

AWS IoT Device Defender is a fully managed service that helps you secure your fleet of IoT devices and features are:

- AWS IoT Device Defender Continuously audits your IoT configurations to make sure that they aren't deviating from security best practices
- A configuration is a set of technical controls you set to help keep information secure when devices are communicating with each other and the cloud
- AWS IoT Device Defender makes it easy to maintain and enforce IoT configurations, such as ensuring device identity, authenticating and authorizing devices and encrypting devices data.
- If something doesn't look right, AWS IoT device defender sends out an alert so you can take action to remediate the issue. For example,. Traffic spikes in outbound traffic might indicate that a device is participating in a DDoS attack.

7. What are new solutions/techniques for AWS Deep Learning services?

Ans:

AWS Deep learning can be used to build, train and deploy deep learning applications in the cloud at any scale. The AMI's come pre-installed with open-source deep learning frameworks including TensorFlow, Apache, MXNet, PyTorch. We can train custom models, experiments with new algorithms, and learn new deep learning skills and techniques. There is no additional charge to use the AMIs. the AMIs offer GPU acceleration through pre-configured CUDA and cuDNN drivers, as well as CPU acceleration through intel MKL-DNN drivers.

Steps to achieve so:-

1. Sign into the AWS Management console and search EC2
2. choose the launch instance button.
3. Select the AWS deep learning AMI
4. on the details page, Choose Continue
5. select an instance type (Gpu compute, Memory optimized, Accelerated services etc..) and launch it.
6. create new private key file
7. find the instances public DNS and copy it
8. start a notebook and transfer the code to amazon EC2 instance

8. What do you mean by Container concept in Cloud Computing? How Amazon Elastic Container Service has added support for application deployments?

Ans:

Containers offer a logical packaging mechanism in which applications can be abstracted from the environment in which they actually run. This decoupling allows container-based applications to be deployed easily and consistently, regardless of whether the target environment is a private data centre, the public cloud, or even a developer's personal laptop. Containerization provides a clean separation of concerns, as developers focus on their application logic and dependencies, while IT operations teams can focus on deployment and management without bothering with application details such as specific software versions and configurations specific to the app. Containers are often compared with virtual machines (VMs). The concept relating to the container in cloud computing is beneficial to the enterprises. Then there is the virtual machine, which is low cost and offers internal as well as cloud portability. Some big organizations comprise of both the virtual machines and cloud containers. The enterprises have to evaluate the pros and cons of each of these and deploy the most apt one.

The code of the application can be bundled with configurations and dependencies in a systematic manner. By this, the user is assured of reliability, consistency, and quickness regardless of the distributed platform. The container usage in online services benefits storage with cloud computing information security, availability and elasticity.

ECS supports Fargate(container running engine) to provide serverless compute for containers, this is an added facility for developers to manage the containers created in virtual machines instances. So, we can rest ourselves in focussing on infrastructure and system deployment which is now easier. We can also scale the application hassle-free which also reduces maintenance costs.

9. How do you understand Amazon Polly Launches Brand Voice? How does it work? Discuss its methodology.

Ans:

Amazon Polly is a service that turns text into lifelike speech. Amazon Polly includes dozens of lifelike voices and support for multiple languages, so you can select the ideal voice and distribute your speech-enabled applications in many geographies. Amazon Polly is a secure service that delivers all of these benefits at high scale and at low latency. Polly's Text-to-Speech (TTS) service uses advanced deep learning technologies to synthesize speech that sounds like a human voice. When combined with Amazon Lex, developers can create full-blown Voice User Interfaces for their applications. Amazon Polly target users are the brands that create voice-interface for the customers which in telephone interaction or while building order management systems which are based on speech-output.

10. How AWS is Introducing the Desktop Client for AWS Client VPN? Discuss.

Ans:

Historically, AWS has had VPN functionality. However, this has been focused on site-to-site IPsec tunnels that connect static locations, such as corporate offices or data centres to the cloud. If we wished to establish a TLS based VPN where the client endpoint could be anywhere, we've had to setup an EC2 instance based solution. That's where Client VPN Endpoints comes in.

AWS Client VPN is a managed client-based VPN service that enables you to securely access your AWS resources and resources in your on-premises network. With Client VPN, you can access your resources from any location using an OpenVPN-based VPN client. This is a fully-managed elastic VPN service that scales the number of connections up and down according to demand. These clients provide basic connectivity but lack advanced authentication and management capabilities.

AWS Client VPN for Desktop is built by AWS and works seamlessly with AWS Client VPN endpoints. Customers now have an end-to-end solution designed, built, and operated by AWS.

The client applications are available at no charge, and can be used to establish connections to any AWS region where you have an AWS Client VPN endpoint, making it even easier for customers to connect their Windows and MacOS clients to AWS, with the launch of the desktop client by AWS.