**Lecture Notes for AWS Solution Architect**

**MODULE-3**

**Amazon Elastic Block Store**

 Amazon Elastic Block Store (EBS) is an easy to use, high-performance, block-storage service designed for use with Amazon Elastic Compute Cloud (EC2) for both throughput and transaction intensive workloads at any scale. A broad range of workloads, such as relational and non-relational databases, enterprise applications, containerized applications, big data analytics engines, file systems, and media workflows are widely deployed on Amazon EBS.

You can choose from six different volume types to balance optimal price and performance. You can achieve single-digit-millisecond latency for high-performance database workloads such as SAP HANA or gigabyte per second throughput for large, sequential workloads such as Hadoop. You can change volume types, tune performance, or increase volume size without disrupting your critical applications, so you have cost-effective storage when you need it.

Designed for mission-critical systems, EBS volumes are replicated within an Availability Zone (AZ) and can easily scale to petabytes of data. Also, you can use EBS Snapshots with automated lifecycle policies to back up your volumes in Amazon S3, while ensuring geographic protection of your data and business continuity.

**Benefits**

* **Performance for any workload**

EBS volumes are performant for your most demanding workloads, including mission-critical applications such as SAP, Oracle, and Microsoft products. SSD-backed options include a volume designed for high performance applications and a general-purpose volume that offers strong price/performance for most workloads. Customers who want to drive higher performance can attach their EBS volumes to Amazon EC2 R5b instances to get up to 60 Gbps bandwidth and 260K IOPS (input/output operations per second) of performance, the fastest block storage performance on EC2.

* **Highly available and durable**

Amazon EBS architecture offers reliability for mission-critical applications. EBS volumes are designed to protect against failures by replicating within the Availability Zone (AZ), offering 99.999% availability. EBS offers a high-durability volume (io2) for customers that need 99.999% durability, especially for their business-critical applications.

* **Cost-effective**

EBS offers six different volumes at various price points and performance benchmarks, enabling you to optimize costs and invest in a precise level of storage for your application needs. Options range from highly-cost-effective, dollar-per-gigabyte volumes to high-performance volumes with high IOPS and high throughput designed for mission-critical workloads.

* **Easy to Use**

Amazon EBS volumes are easy to create, use, encrypt, and protect. Elastic Volumes capability allows you to increase storage, tune performance up and down, and change volume types without any disruption to your workloads. EBS Snapshots allow you to easily take backups of your volumes for geographic protection of your data. Data Lifecycle Manager (DLM) is an easy-to-use tool for automating snapshot management without any additional overhead or cost.

* **Virtually unlimited scale**

Amazon EBS enables you to increase storage without any disruption to your critical workloads, build applications that require as little as a single GB of storage, or scale up to petabytes of data — all in just a few clicks. Snapshots can be used to quickly restore new volumes across a region's Availability Zones, enabling rapid scale.

* **Secure**

EBS is built to be secure for data compliance. Newly-created EBS volumes can be encrypted by default with a single setting in your account. EBS volumes support encryption of data at rest, data in transit, and all volume backups. EBS encryption is supported by all volume types, includes built-in key management infrastructure, and has zero impact on performance.

**Amazon EBS volume types**

* **Solid state drives (SSD)** — Optimized for transactional workloads involving frequent read/write operations with small I/O size, where the dominant performance attribute is IOPS.
* **Hard disk drives (HDD)** — Optimized for large streaming workloads where the dominant performance attribute is throughput.

**EC2 Placement Groups**

When you launch a new EC2 instance, the EC2 service attempts to place the instance in such a way that all of your instances are spread out across underlying hardware to minimize correlated failures. You can use placement groups to influence the placement of a group of interdependent instances to meet the needs of your workload. Depending on the type of workload, you can create a placement group using one of the following placement strategies:

**Cluster**– packs instances close together inside an Availability Zone. This strategy enables workloads to achieve the low-latency network performance necessary for tightly-coupled node-to-node communication that is typical of HPC applications.

**Partition**– spreads your instances across logical partitions such that groups of instances in one partition do not share the underlying hardware with groups of instances in different partitions. This strategy is typically used by large distributed and replicated workloads, such as Hadoop, Cassandra, and Kafka.

**Spread** – strictly places a small group of instances across distinct underlying hardware to reduce correlated failures.

**Instance States**

An Amazon EC2 instance transitions through different states from the moment you launch it through to its termination.

The following illustration represents the transitions between instance states.



| **Instance state** | **Description** | **Instance usage billing** |
| --- | --- | --- |
| pending | The instance is preparing to enter the running state. An instance enters the pending state when it launches for the first time, or when it is started after being in the stopped state. | Not billed |
| running | The instance is running and ready for use. | Billed |
| stopping | The instance is preparing to be stopped or stop-hibernated. | Not billed if preparing to stopBilled if preparing to hibernate |
| stopped | The instance is shut down and cannot be used. The instance can be started at any time. | Not billed |
| shutting-down | The instance is preparing to be terminated. | Not billed |
| terminated | The instance has been permanently deleted and cannot be started. | Not billed |

**Elastic Network Interface (ENI)**

An elastic network interface is a logical networking component in a VPC that represents a virtual network card. It can include the following attributes:

* A primary private IPv4 address from the IPv4 address range of your VPC
* One or more secondary private IPv4 addresses from the IPv4 address range of your VPC
* One Elastic IP address (IPv4) per private IPv4 address
* One public IPv4 address
* One or more IPv6 addresses
* One or more security groups
* A MAC address
* A source/destination check flag
* A description

You can create and configure network interfaces in your account and attach them to instances in your VPC. Your account might also have requester-managed network interfaces, which are created and managed by AWS services to enable you to use other resources and services. You cannot manage these network interfaces yourself.

This AWS resource is referred to as a network interface in the AWS Management Console and the Amazon EC2 API. Therefore, we use "network interface" in this documentation instead of "elastic network interface". The term "network interface" in this documentation always means "elastic network interface".

**Network interface basics**

You can create a network interface, attach it to an instance, detach it from an instance, and attach it to another instance. The attributes of a network interface follow it as it's attached or detached from an instance and reattached to another instance. When you move a network interface from one instance to another, network traffic is redirected to the new instance.

**Primary network interface**

Each instance has a default network interface, called the primary network interface. You cannot detach a primary network interface from an instance. You can create and attach additional network interfaces. The maximum number of network interfaces that you can use varies by instance type.

**Public IPv4 addresses for network interfaces**

In a VPC, all subnets have a modifiable attribute that determines whether network interfaces created in that subnet (and therefore instances launched into that subnet) are assigned a public IPv4 address. The public IPv4 address is assigned from Amazon's pool of public IPv4 addresses. When you launch an instance, the IP address is assigned to the primary network interface that's created.

When you create a network interface, it inherits the public IPv4 addressing attribute from the subnet. If you later modify the public IPv4 addressing attribute of the subnet, the network interface keeps the setting that was in effect when it was created. If you launch an instance and specify an existing network interface as the primary network interface, the public IPv4 address attribute is determined by this network interface.

**Elastic IP addresses for network interface**

If you have an Elastic IP address, you can associate it with one of the private IPv4 addresses for the network interface. You can associate one Elastic IP address with each private IPv4 address.

If you disassociate an Elastic IP address from a network interface, you can release it back to the address pool. This is the only way to associate an Elastic IP address with an instance in a different subnet or VPC, as network interfaces are specific to subnets.

**IPv6 addresses for network interfaces**

If you associate IPv6 CIDR blocks with your VPC and subnet, you can assign one or more IPv6 addresses from the subnet range to a network interface. Each IPv6 address can be assigned to one network interface.

All subnets have a modifiable attribute that determines whether network interfaces created in that subnet (and therefore instances launched into that subnet) are automatically assigned an IPv6 address from the range of the subnet. When you launch an instance, the IPv6 address is assigned to the primary network interface that's created.

**Termination behavior**

You can set the termination behavior for a network interface that's attached to an instance. You can specify whether the network interface should be automatically deleted when you terminate the instance to which it's attached.

**Source/destination checking**

You can enable or disable source/destination checks, which ensure that the instance is either the source or the destination of any traffic that it receives. Source/destination checks are enabled by default. You must disable source/destination checks if the instance if the instance runs services such as network address translation, routing, or firewalls.

**Monitoring IP traffic**

You can enable a VPC flow log on your network interface to capture information about the IP traffic going to and from a network interface. After you've created a flow log, you can view and retrieve its data in Amazon CloudWatch Logs.