

Elastic Block Store (EBS) and Elastic File System (EFS)

>EBS Overview

EBS Volume Types: Solid State Disk

Essentially EBS are storage volumes you can attach to your EC2 Instances. Amazon Elastic Block Store (Amazon EBS) provides block level storage volumes for use with EC2 instances. EBS volumes behave like raw, unformatted block devices. You can mount these volumes as devices on your instances.

- EBS volumes can be encrypted, which, when you're working with sensitive data, can make a big difference.
- EBS volumes can be moved around, mounted on other instances, and, as you've seen, converted to AMIs.
- EBS is a storage device (Called a volume) That can be attached to (or removed from) your instance.
- Tied to one AZ and can only be attached to one instance in the same AZ
- Long-term data storage
- EBS volumes can only be mounted or attached to one EC2 instance at a given time
 - It's possible to convert any EC2 instance into an AMI by creating a snapshot from the EBS volume used with an instance and then creating an image from the snapshot. The resulting image will be available as an AMI either in the AMIs menu on the EC2 Dashboard or on the My AMIs tab in the Choose An Amazon Machine Image page of the instance launch process.
- EBS storage volumes can be encrypted and are more like physical hard drives in the flexibility of their usage.
- Database instances use EBS storage. Both the Standard and Memory Optimized instance class types are EBS-optimized, meaning they provide dedicated bandwidth for transfers to and from EBS storage.

EBS is used for mission critical where it is great for production workloads. Made to be highly available where it is automatically replicated within a single AZ to protect against hardware failures. And is scalable where it increases capacity and change the volume type with no downtime or performance impact.

- General Purpose SSD volumes (gp2 and gp3) balance price and performance for a wide variety of transactional workloads. These volumes are ideal for use cases such as boot volumes, medium-size single instance databases, and development and test environments.
- Provisioned IOPS SSD volumes (io1 and io2) are designed to meet the needs of I/O-intensive workloads that are sensitive to storage performance and consistency. They provide a consistent IOPS rate that you specify when you create the volume. This enables you to predictably scale to tens of thousands of IOPS per instance. Additionally, io2 volumes provide the highest levels of volume durability.
- Throughput Optimized HDD volumes (st1) provide low-cost magnetic storage that defines performance in terms of throughput rather than IOPS. These volumes are ideal for large, sequential workloads such as Amazon EMR, ETL, data warehouses, and log processing.
- Cold HDD volumes (sc1) provide low-cost magnetic storage that defines performance in terms of throughput rather than IOPS. These volumes are ideal for large, sequential, cold-data workloads.

If you require infrequent access to your data and are looking to save costs, these volumes provides inexpensive block storage.

- **AMI Overview**

Amazon machine image, which represents a customization of EC2 instances. An Amazon Machine Image (AMI) is a supported and maintained image provided by AWS that provides the information required to launch an instance. You must specify an AMI when you launch an instance. You can launch multiple instances from a single AMI when you require multiple instances with the same configuration. You can use different AMIs to launch instances when you require instances with different configurations.

- Built for a specific region (and can be copied across regions)
- Launch EC2 instances from a public AMI, your own AMI, or AWS marketplace
- AMI Process (From an EC2 Instance)

i. Start an EC2 and customize it > Stop the instance (For data integrity) > Build an AMI - this will also create EBS snapshots > launch instances from other AMIs

- **Instance Store**

An instance store provides temporary block-level storage for your instance. This storage is located on disks that are physically attached to the host computer. Instance stores are ideal for temporary storage of information that changes frequently, such as buffers, caches, scratch data, and other temporary content, or for data that is replicated across a fleet of instances, such as a load-balanced pool of web servers.

- Network drives with good but “limited” performance
- If you need high-performance hardware disk, use EC2 instance store (Ephemeral – lasting for a very short time.)
 - The virtual devices for instances store volumes are... Ephemeral.)
- Better I/O performance
 - Input / Output
 - In computing, input/output is the communication between an information processing system, such as a computer, and the outside world, possibly a human or another information processing system.
- EC2 instance store lose their storage if they’re stopped
- Good for buffer / cache / scratch data / temporary content

- **EBS Volume Types**

https://www.youtube.com/watch?v=XvC3C9rOs_k

<https://www.youtube.com/watch?v=lahRGMRiK1c> (IOPS explained)

- EBS Volumes comes in 6 types

i. **Gp2 / gp3** (SSD): General purpose SSD volume that balances price and performance for a wide variety of workloads

❖ IOPS = Input/output operations per second is an input/output performance measurement used to characterize computer storage devices like hard disk drives, solid state drives, and storage area networks

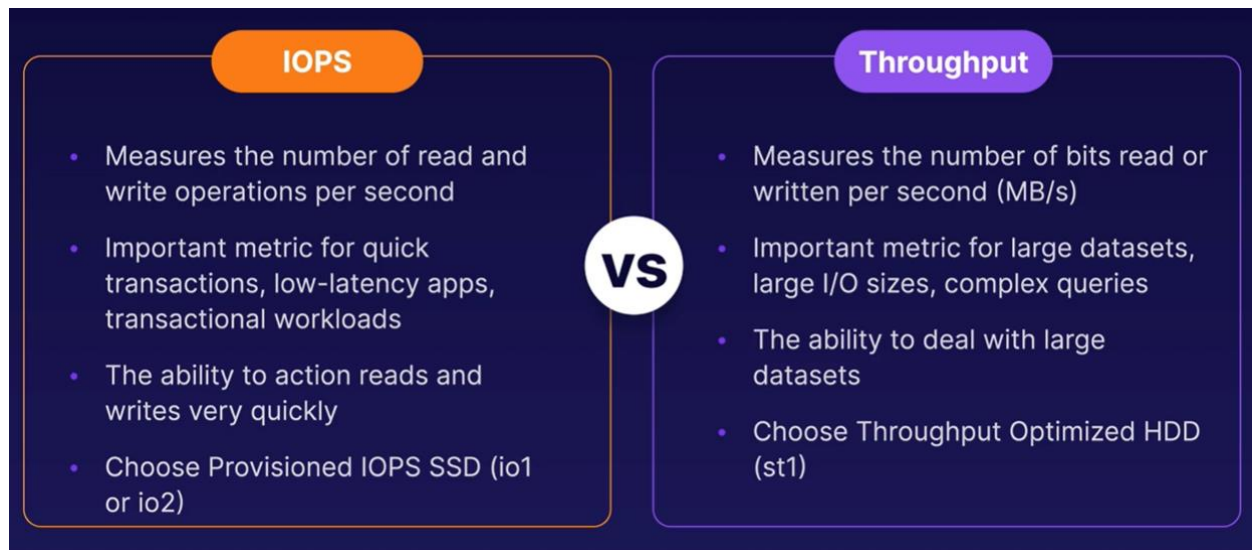
- Cost effective storage, low-latency
- 1 GB-16TB
- **GP2 (Only these 2 can be used as boot volumes)**
 - Small GP2 volumes can burst IOPS to 3,000
 - Size of the volume and IOPS are linked, max IOPS is 16,000
 - 3 IOPS per GB means at 5,334 GB we are at the max IOPS
- **GP3**

- Baseline of 3,000 IOPS and throughput of 125 MB/s
 - Can increase IOPS up to 16,000 and throughput up to 1000 MB/s independently
 - 4 times faster than max throughput of gp2 volumes
- ii. **io1 / io2 (SSD):** Highest-performance SSD volume for mission critical low-latency or high-throughput workloads.
- Provisioned IOPS (PIOPS) SSD
 - Critical business applications with sustained IOPS performance
 - Or applications that need more than 16k IOPS
 - Great for databases workloads
 - **io1/io2 (4 GB - 16 TB)**
 - Max PIOPS: 64,000 for Nitro EC2 instances & 32,000 for other
 - Can increase PIOPS independently from storage size
 - io2 have more durability and more IOPS per GB (At the same price as
 - **io2 Block express (4 GB - 64 TB)**
 - Sub-millisecond latency
 - Max PIOPS: 256,000 with an IOPS:GB of 1,000:1
 - Supports EBS multi-attach
- ii. **St1 (HDD):** Low cost HDD volume designed for frequently accessed, throughput-intensive workloads
 - Big data, data warehouses, log processing
 - Somewhat reminds me a little about AWS Redshift
 - Baseline throughput of 40 MB/s per TB
 - Ability to burst up to 250 MB/s per TB
 - Max throughput 500 MB - Max IOPS 500
- iii. **Sc1 (HDD):** Lowest cost HDD volume designed for less frequently accessed workloads.
 - For data that is infrequently accessed
 - Scenarios where lowest cost is imported
 - Max throughput 250 MB - max IOPS 250
- HDD cannot be a boot volume. Max throughput is 125 MB to 16 TB

Tip... When it doubt, always consult the AWS documentation

<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-volume-types.html>

- Provisioned IOPS (PIOPS) SSD
 - Critical business applications with sustained IOPS performance
 - Or applications that need more than 16k IOPS
 - Great for databases workloads



Things to remember when studying - If you need to have a higher speed EBS volume, you need to use an EC2 Nitro type of instance to make sure you're leveraging the increased level of IOPS.

- EBS Multi-Attach

Multi-Attach makes it easier for you to achieve higher application availability in clustered Linux applications that manage concurrent write operations.

- Attach the same EBS volume to multiple EC2 instances in the same AZ
 - **ONLY IF... it is part of the i01 and i02 family**
- Each instance has full read & write permissions to the volume

- EBS Encryption

Use Amazon EBS encryption as a straight-forward encryption solution for your EBS resources associated with your EC2 instances. With Amazon EBS encryption, you aren't required to build, maintain, and secure your own key management infrastructure. Amazon EBS encryption uses AWS KMS keys when creating encrypted volumes and snapshots.

- When you create an encrypted EBS volume, you get the following:
 - Data at rest is encrypted inside the volume
 - All the data in flight moving between the instance and the volume is encrypted
 - All snapshots are encrypted
 - All volumes created from the snapshot
 - Encryption and decryption are handled transparently (you have nothing to do)
 - Encryption has a minimal impact on latency
 - EBS encryption leverages keys from KMS (AES-256)
 - Copying an unencrypted snapshot allows encryption
 - Snapshots of encrypted volumes are encrypted
 - To encrypt and unencrypted EBS volume
1. Create an EBS snapshot of the volume
 2. Encrypt the EBS snapshot (using copy)
 3. Create new EBS volume from the snapshot (the volume will also be encrypted)
 4. Now you can attach the encrypted volume to the original instance

>Volumes and snapshots

EBS Snapshots

You can back up the data on your Amazon EBS volumes to Amazon S3 by taking point-in-time snapshots. Snapshots are incremental backups, which means that only the blocks on the device that have changed after your most recent snapshot are saved. This minimizes the time required to create the snapshot and saves on storage costs by not duplicating data.

- Making a snapshot aka a backup... of your EBS volume
- Not necessary to detach volume to do snapshot, but recommended
- Can copy snapshots across AZ or Region
- EBS Snapshot archive

i. Move a snapshot to an “archive tier” that is 75% cheaper

ii. Takes within 24 to 72 hours for restoring the archive

- Recycle Bin (Delete) for EBS snapshots

. Setup rules to retain deleted snapshots so you can recover them after an accidental deletion

EBS Volume

i. Network drive (Not physical)

- Uses the network to communicate the instance, which means there might be a bit of latency
- Can detach from an EC2 instance and attached to another one quickly

. Locked to an Az

- An EBS volume in us-east-1a cannot be attached to us-east-1b
- To move a volume across, you first need to snapshot it

. Have a provisioned capacity

- Get billed for all provisioned capacity
- You can increase the capacity of the drive over time
 - **EBS - Delete on termination attribute**
- By default it is checked for the root volume but not for a new volume
- By default, the root EBS volume is deleted (Attribute enabled)
- By default, any other attached EBS volume is not deleted (Attribute disabled)

Snapshots exist on S3, Snapshots are point in time, Snapshots are incremental

EBS volumes will always be in the same AZ as the EC2 instance to which it is attached. You can resize EBS volumes on the fly. You can also change volume types on the fly as well.

>Protecting EBS Volumes with encryption

EBS Encryption

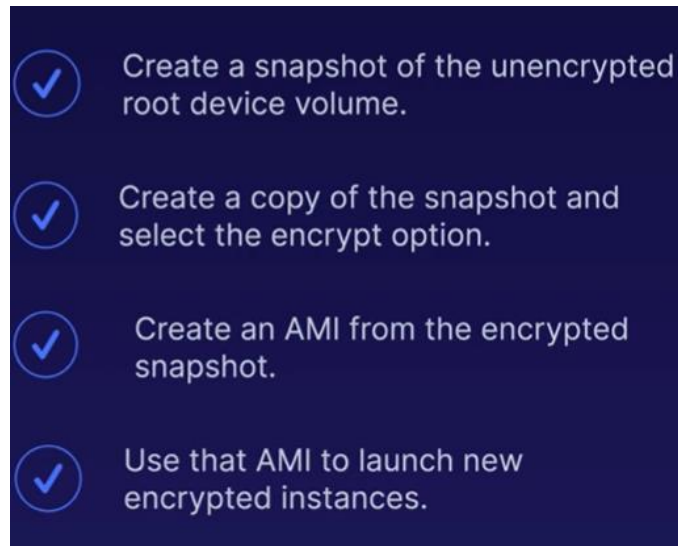
encrypts your volume with a data key using the industry-standard AES-256 algorithm. This uses AWS Key Management Service (AWS KMS) customer master keys (CMK) when creating encrypted volumes and snapshots.

EBS Encryption

- Handled transparently
 - Encryption and decryption are handled transparently
- Latency

- Encryption has a minimal impact on latency
- Copying
 - Copying an unencrypted snapshot allows encryption
- Snapshots
 - Snapshots of encrypted volumes are encrypted
- Root Device Volumes
 - You can now encrypt root device volumes upon creation

Down below are the 4 steps to encrypt an unencrypted volume



>EC2 Hibernation

Hibernation gives you the ability to launch EC2 instances, set them up as desired, hibernate them, and then quickly bring them back to life when you need them. Applications pick up exactly where they left off instead of rebuilding their memory footprint. Hibernation saves the contents from the instance memory (RAM) to your AWS EBS root volume.

- We can stop/terminate instances
 - Stop - the data on disk (EBS) is kept intact in the next start
 - Terminate - any EBS volumes (root) also set-up to be destroyed is lost
- Down below are the process on start
 - First start: the OS boots & the EC2 user data script is run (bootstrap scripts)
 - Following that the OS boots up
 - Then the application starts, which caches gets warmed up, which will take some time
- Introducing EC2 Hibernate
 - The RAM state is preserved
 - Instance boot is much faster (the OS is not stopped/restarted)
 - The RAM state is written to a file in the root EBS volume
 - The root EBS volume MUST be encrypted

<https://aws.amazon.com/about-aws/whats-new/2020/12/ec2-hibernation-feature-is-now-available-in-aws-govcloud-us->

[regions/#:~:text=Hibernation%20gives%20you%20the%20ability,of%20rebuilding%20their%20memory%20footprint.](#)

So in the course he explains how when an EC2 instance is running. Down below is the order

1. EC2 Instance running which data is in RAM
2. Starts the hibernation process
3. The instance will go in the stopping state and the RAM is going to be dumped into the EBS volume
4. Then the instance is shut down and the RAM disappears, because as soon as you stop an instance the RAM goes away.
5. But the EBS volume still contains the dump of the RAM
6. And then when your instance is started, then the RAM is going to be loaded from disk onto the EC2 instance memory.
7. And that means that it is just as is
8. Your EC2 instance never got stopped.

So, the use cases for hibernation is if you want to have long-running processes and never actually stop them, or if you want to save the RAM state, or if you just want to reboot, like boot up fast, and you have services that take time to initialize. So you actually don't want them to initialize, you want them to stay up and running even if you hibernate your instance.

Good to know

- Instance Ram Size - must be less than 150 GB
- No supported for bare metal instances
- AMI - Amazon Linux 2, Linux AMI, Ubuntu, RHEL, CentOS & Windows
- Works with the root volume which must be encrypted

>EFS

- Amazon Elastic File System (Amazon EFS) is a simple, serverless, set-and-forget elastic file system that makes it easy to set up, scale, and cost-optimize file storage in AWS.
 - Amazon EFS file systems can automatically scale from gigabytes to petabytes of data without needing to provision storage. Tens, hundreds, or even thousands of compute instances can access an Amazon EFS file system at the same time, and Amazon EFS provides consistent performance to each compute instance. Amazon EFS is designed to be highly durable and highly available. With Amazon EFS, there is no minimum fee or setup costs, and you pay only for what you use.
 - EFS has a **higher price point than EBS**
 - Scales automatically as you add files to it
 - Can leverage EFS-IA for cost savings
 - Mounting 100s of instances across AZ
 - EFS share website files (WordPress)
 - **ONLY for Linux instances (POSIX)**
- For EFS, you get **billed only for what you use on your EFS, whereas for EBS, you have to provision in advance.**
- EFS is really for a network file system to be mounted across multiple instances

EFS is great for content management systems as you can easily share content between EC2 instances. Also great for web servers, just having a single folder structure for your website.

- Managed NFS (Network file system) that can be mounted on many EC2
- EFS works with EC2 instances in multi-AZ
- Highly available, scalable, expensive (3x gp2), pay per use
- Uses NFSv4.1 protocol
- Compatible with Linux based AMI (Not windows)
- Encryption at rest using KMS
- Uses security group to control access to EFS
- Performance & Storage Class
 - EFS Scale
 - 1000s of concurrent NFS clients, 10 GB+ /s throughput
 - Grow Petabyte-scale network file system, automatically
 - Performance Mode (Set at EFS creation time)
 - General purpose (Default): latency-sensitive use cases (Web server, CMS, etc..)
 - Max I/O - higher latency, throughput highly parallel (big data, media processing)
 - Throughput mode
 - Bursting (1 TB = 50MB/s + burst of up to 100MB/s)
- EFS - Storage Classes
 - Storage Tiers (Lifecycle management feature - move file after N days)
 - Standard: for frequently accessed files
 - Infrequent access (EFS-IA) with a lifecycle policy
 - Availability and durability
 - Regional: Multi-AZ, great for prod
 - One Zone: One AZ, great for dev, backup enabled by default, compatible with IA (EFS One Zone IA)

For a scenario questions based around highly scalable shared storage using NFS, think EFS

>FSx Overview

➤ Windows

- FSx for Windows File Server provides a fully managed native Microsoft Windows File System
 - This makes it easier to move windows-based applications that require file storage to AWS



Amazon FSx for Lustre

- Fully managed file system that is optimized for compute-intensive workloads
- Basically anything to do with **AI and machine learning**
- Launch and run Lustre file system that can process massive datasets at up to hundreds of gigabytes per second
- Easy and cost effective to launch, run, and scale the world's most popular high-performance file system
- Can be linked to Amazon S3 Buckets

AMI: EBS vs Instance Store

- AMI (Amazon Machine Image)
 - Provides information required to launch your instance

5 things you can base your AMI on



- Instance Store
 - This is storage physically attached to the host PC
 - This is temp since data loss occurs when the EC2 is stopped.
 - Even when you put it in a hibernate or a terminate state. Data will be loss.
 - Recommended for temp storage needs. Nothing valuable should NOT be in here.

>AWS Backup

AWS backup allows you to consolidate your backups across multiple AWS services, such as EC2, EBS, EFS, Amazon FSx for Lustre, and many others. IT can include other services such as database technologies like RDS and Dynamo DB.

- Benefits of AWS Backup
 - Central Management
 - Use a single, central backup console, allowing you to centralize your backups across multiple AWS services and multiple AWS accounts.
 - Automation
 - You can create automated backup schedules and retention policies. You can also create lifecycle policies, allowing you to expire unnecessary backups after a period of time.
 - Improved Compliance
 - Backup policies can be enforced while backups can be encrypted both at rest and in transit, allowing alignment to regulatory compliance.

Exam Tips

1. Learning EBS: SSD Volumes
 - GP2
 - Suitable for boot disk and general applications
 - Up to 16,000 IOPS per volume
 - GP3

- High performance applications
 - Predictable 3000 IOPS baseline performance and 125 MiB/s regardless of volume size
- io1
 - Suitable for OLTP and latency sensitive apps
 - 50 IOPS/GiB
 - Up to 64,000 IOPS per volume
- io2
 - 500 IOPS/GiB
 - Latest generation provisioned IOPS volume
- 2. EBS: HDD Volumes
 - St1
 - Throughput optimized HDD
 - Suitable for big data, data warehouses and ETL
 - Max throughput is 500 MB/s per volume
 - Cannot be a boot volume
 - Sc1
 - Cold HDD
 - Max throughput of 250 MB/s per volume
 - Less frequently accessed data
 - Lowest cost
 - Cannot be boot volume
- 3. EBS Volumes and Snapshots
 - Volumes exist on EBS, whereas snapshots exist on S3
 - Snapshots are point-in-time photographs of volumes and are incremental in nature
 - The first snapshot will take some time to create
 - You can resize EBS volumes on the fly as well as changing the volume types
- 4. Encrypted Volumes
 - Data at rest is encrypted inside the volume
 - All data in flight moving between the instance and the volume is encrypted
 - All snaps are encrypted
 - All volumes created from the snapshot are encrypted
 - The 4 steps to encrypted volumes
- 5. EC2 Hibernation
 - Preserves the in-memory RAM on persistent storage (EBS)
 - Much faster to boot up (stored in RAM)
- 6. EFS
 - Supports the network file system version 4 protocol
 - Only pay for the storage you use
 - Can scale up to petabytes
 - Read-after-write consistency

- Stored across multiple AZs within a region
7. FSx
- EFS: When you need distributed, highly resilient storage for Linux instances and Linux-Based applications
 - Amazon FSx for Windows: When you need centralized storage for Windows-based applications, such as SharePoint, Microsoft SQL Server, Workspaces, or any other native Microsoft application
 - Amazon FSx for Lustre: When you need high speed, high-capacity distributed storage. For applications that do high performance computing (HPC). Remember that FSx for Lustre can store data directly on S3
8. AWS Backup
- Consolidation: Use AWS backup to backup AWS services, such as EC2, EBS, EFS, Amazon FSx for Lustre, Amazon FSx for Windows File Server, and AWS storage gateway
 - Organizations: Use this in conjunction with AWS backup to backup your different AWS services across multiple AWS accounts.
 - Benefits: Gives you centralized control, letting you automate your backups and define lifecycle policies for your data and audit them once completed