

Storage and Archiving

EBS · INSTANCE STORE · EFS · S3 · S3 GLACIER · STORAGE GATEWAY · MIGRATION

STUDY & REVIEW

01 MUST KNOW

if you remember five things, remember these

- 01 Four storage categories:** AWS storage sorts into **block** (Amazon EBS), **object** (Amazon S3 and S3 Glacier), **file** (Amazon EFS and Amazon FSx), and **hybrid** (AWS Storage Gateway). Match the workload: one EC2 instance's disk → EBS; internet-accessible objects → S3; shared Linux file system → EFS; on-premises link → Storage Gateway.
- 02 EBS is AZ-scoped block storage:** Amazon EBS volumes are **network-attached**, **persistent** block storage for a single EC2 instance, and a volume must live in the **same Availability Zone** as the instance it attaches to. Back volumes up with **snapshots** — the first is full, every later one is **incremental** (changed blocks only) and stored in Amazon S3.
- 03 S3 is durable regional object storage:** Amazon S3 is **regionally** provisioned object storage kept in **buckets** with globally unique names. It is designed for **eleven 9s (99.99999999%) durability** and **99.99% availability**, holds an **unlimited** amount of data with up to **5 TB per object**, and keeps every object **private by default**.
- 04 Classes + lifecycle control cost:** S3 storage classes — **Standard**, **Intelligent-Tiering**, **Standard-IA**, **One Zone-IA**, and **Glacier** — trade retrieval cost against access frequency. **Object lifecycle** rules automate **transition** actions (move to a cheaper class) and **expiration** actions (delete), so data is stored cost-effectively across its life.
- 05 Glacier retrieval is by request:** Amazon S3 Glacier is extremely low-cost archive storage with retrieval **by request** (an asynchronous job), not instant access. Three options: **Expedited** (1–5 min), **Standard** (3–5 hr, the default), and **Bulk** (5–12 hr, the lowest cost).

02 KEY TERMS

TERM	DEFINITION
Amazon EBS	Elastic Block Store — network-attached, persistent block-storage volumes for a single EC2 instance in the same Availability Zone
EBS snapshot	Point-in-time, incremental backup of an EBS volume stored in Amazon S3; the first is full, later ones capture only changed blocks
EBS-optimized instance	An instance with a dedicated network connection to its EBS volumes that raises and stabilizes EBS I/O performance
Instance store	Temporary, non-persistent block storage physically attached to the host computer; fast, and reclaimed when the instance terminates
Amazon EFS	Elastic File System — a fully managed, elastic NFS v4 file system for Linux workloads, shareable by many EC2 instances at once
Mount target	An IP address for an NFSv4 endpoint in a VPC subnet where an EC2 instance mounts an EFS file system
Amazon S3	Simple Storage Service — regional object storage that keeps objects in buckets and serves them by key over the internet
Bucket	A container for S3 objects whose name is globally unique across all AWS accounts
S3 versioning	A bucket setting that keeps multiple versions of an object to guard against accidental overwrites and deletes; once enabled it can only be suspended
S3 Object Lock	Write once, read many (WORM) protection that blocks deletion or overwrite for a set period or indefinitely; enabled at bucket creation
Amazon S3 Glacier	Secure, extremely low-cost storage for long-term archive and backup; data is stored as archives inside vaults

TERM	DEFINITION
AWS Storage Gateway	Hybrid service that links on-premises applications to AWS storage through a file, volume, or tape gateway

03 PERSISTENT EBS VS. EPHEMERAL INSTANCE STORE both are block storage – one survives, one doesn't

AMAZON EBS <i>network-attached, persistent</i> - Attaches over the network to one instance in the same AZ Persists independently of the instance's life - Supports boot and data volumes - Backed up with incremental snapshots to Amazon S3 - Types: io1, gp2, st1, sc1	INSTANCE STORE <i>physically attached, temporary</i> Physically attached to the host computer Non-persistent — reclaimed when the instance terminates - Very fast (NVMe / SSD options) - For buffers, cache, scratch data - Instance-store-backed instances can't be stopped, only terminated
--	---

04 EBS VOLUME TYPES SSD = IOPS-intensive · HDD = throughput-intensive

TYPE	MEDIA	BEST FOR
Provisioned IOPS SSD (io1)	SSD	Highest performance; critical, I/O-intensive databases; extremely low latency
General Purpose SSD (gp2)	SSD	Default type; broad transactional workloads, dev/test, boot volumes
Throughput Optimized HDD (st1)	HDD	Frequently accessed, throughput-intensive work with large datasets and I/O sizes
Cold HDD (sc1)	HDD	Lowest cost per GB ; infrequently accessed, large cold datasets

05 AMAZON S3 STORAGE CLASSES trade retrieval cost against access frequency

STORAGE CLASS	DESIGNED FOR
S3 Standard	Frequently accessed, general-purpose data
S3 Intelligent-Tiering	Data with unknown or changing access patterns; moves objects automatically
S3 Standard-IA	Long-lived, infrequently accessed data
S3 One Zone-IA	Long-lived, infrequently accessed, non-critical data (stored in a single AZ)
S3 Glacier	Long-term archive and digital preservation

06 OBJECT LIFECYCLE → ARCHIVE one rule set moves objects to cheaper classes, then deletes them

01 Object created in **S3 Standard** → **02** Transition to **Standard-IA** (e.g. after 30 days) → **03** Archive to **S3 Glacier** (e.g. after 1 year) → **04** **Expiration** action deletes it at end of life

- ✗ “EBS snapshots live in S3, so I can grab them with the S3 tools.” — No. Snapshots sit in Amazon S3 infrastructure but are **managed by AWS and not directly accessible**; use the EBS tools (for example `create-volume` with `--snapshot-id`) to restore one to a new volume.
- ✗ “Any EBS volume can attach to any instance.” — A volume and its instance must be in the **same Availability Zone**. Snapshots are **Region-scoped** and must be **copied** to another Region before you can use them there.
- ✗ “Instance store keeps my data.” — Instance store is **non-persistent** and is reclaimed when the instance terminates. Instance-store-backed instances **cannot be stopped, only terminated** — the data is lost.
- ✗ “The first EBS snapshot and every later one are all full copies.” — Only the **first** snapshot is full. Every subsequent snapshot is **incremental**, copying just the blocks that changed since the previous one.
- ✗ “I can switch S3 versioning back off.” — Once **enabled**, versioning on a bucket can only be **suspended**, never disabled. A suspended bucket keeps existing versions but behaves like an unversioned one for new writes.
- ✗ “In Compliance mode an admin can still delete a locked object.” — In **Compliance** mode **no one — not even the AWS account root user** — can overwrite or delete the object during its retention period. Object Lock must be enabled **when the bucket is created**.
- ✗ “Glacier reads are instant like S3.” — Retrieval is **by request** and asynchronous. **Standard** retrieval (the default) takes **3–5 hours**; only **Expedited** is 1–5 minutes. There is also **no console support** for direct archive operations.
- ✗ “gp2 versus io1 doesn't matter, and EBS sits on the host like instance store.” — **gp2** (General Purpose SSD) is the **default**; **io1** (Provisioned IOPS SSD) is the highest-performance type. EBS is **network-attached**, while **instance store** is **physically attached** to the host.

08 SELF-QUIZ

answers are upside-down below

- Q1** To which compute resource does an Amazon EBS volume attach, and in what location must the two match?
- Q2** Which EBS volume type is the default for EC2 instances, and which is the highest-performance option?
- Q3** After the first EBS snapshot of a volume, what do subsequent snapshots capture, and where are all of them stored?
- Q4** What happens to instance store data when the instance terminates, and can an instance-store-backed instance be stopped?
- Q5** Which protocol and operating-system family does Amazon EFS serve, and how many instances can use one file system at once?
- Q6** State S3's designed durability, its designed availability, and the maximum size of a single object.
- Q7** Once versioning is enabled on an S3 bucket, what are the only two states the bucket can be in afterward?
- Q8** Which S3 storage class automatically moves objects to the most cost-effective tier, and after how many idle days does it demote an object?
- Q9** Name Amazon S3 Glacier's three archive retrieval options in order from fastest to slowest.
- Q10** Which three AWS data transfer and migration services move data to AWS, and which one uses shipped physical appliances?

ANSWER KEY (rotate the page)

Q1 A single Amazon EC2 instance, in the same Availability Zone. **Q2** General Purpose SSD (gp2) is the default; Provisioned IOPS SSD (io1) is the highest performance. **Q3** Only the blocks that changed (they are incremental); all snapshots are stored in Amazon S3. **Q4** The data is lost (instance store is reclaimed); the instance can only be terminated, not stopped. **Q5** The NFS (v4) protocol for Linux workloads; many EC2 instances can access it concurrently. **Q6** Eleven 9s (99.999999999%) durability, 99.99% availability, and up to 5 TB per object. **Q7** Versioning enabled or versioning suspended — it can never be disabled again. **Q8** S3 Intelligent-Tiering; after 30 consecutive days without access. **Q9** Expedited (1–5 min), Standard (3–5 hr), Bulk (5–12 hr). **Q10** AWS Transfer for SFTP, AWS DataSync, and AWS Snowball; Snowball uses physical appliances.