S3 101:

S3 is build of eleven 9 availability

S3 has:

* Tiered Storage
* Lifecycle management – if a file is x old, do y with it
* Versioning
* Encryption
* MFA Delete
* Secure Data using ACLs and Bucket policies

1. S3 Standard – eleven 9 availability where the loss of 2 facilities concurrently can be sustained
2. S3 IA- Infrequently accessed. Lower charged storage that is accessed infrequently. Retrieval fee is charged
3. S3 One Zone IA – lower cost option for infrequently accessed data where multiple AZs are not required
4. S3 Intelligent Tiering – designed to optimize cost by automatically moving your data to appropriate S3 Tier
5. S3 Glacier – Low cost storage for archiving. Retrieval time is minutes to hours
6. S3 Glacier Deep Archive – Cheapest storage. Retrieval hours on the order of 12 hours



Know – First byte latency

For S3, you’re charged for

* Storage
* number of requests
* storage management pricing
* data transfer pricing
* Transfer Acceleration
  + The user uploads file to the Amazon edge location and that data is routed to the data center over the Amazon backbone network path
* Cross Region Replication Pricing

**Exam Tips:**

* S3 is Object Based (vs block based – os and db)
* Files can be 0 to 5Tb
* Unlimited storage
* Buckets
* S3 is universal namespace (converts to a DNS name)
* Can’t install OS on S3
* Successful uploads will generate HTTP 200 code
* Can turn on MFA Delete

Key fundamentals of S3:

* Key
* Value
* Version ID
* Metadata
* Sub resources
* Access Control Lists (bucket and object level)
* Torrent

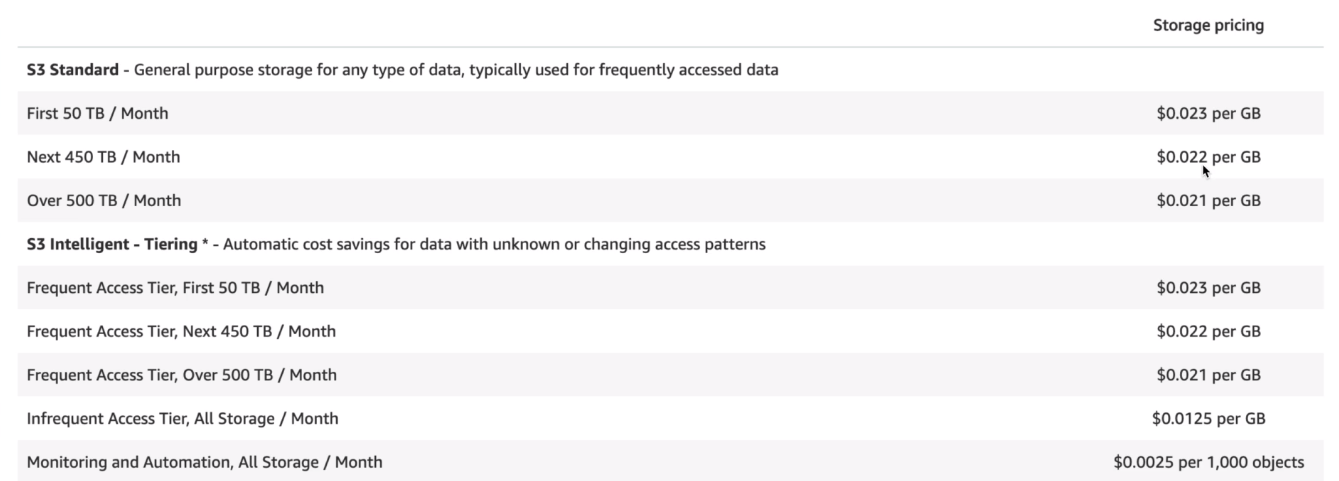
Read after Write consistency for PUTS of new Objects

Eventual Consistency for overwrite PUTS and DELETE – can take time to propagate

Different Storage classes (Table Above)

Go read the S3 FAQs before the Exam!

S3 Pricing Tiers:



Size of s3 bucket is sum total of all the versions of a file it contains.

**S3 Versioning Version Control Tips:**

* Stores all versions of object
* Great backup tool
* Once enabled, versioning cannot be disabled, only suspended.
* Integrates with lifecycle rules
* Versioning MFA Delete capability exists

Different ways to share s3 buckets across accounts:

* Using bucket policies and IAM (applies across entire bucket). Programmatic access only
* Using Bucket ACLs and IAM (Individual objects). Prog accesso nly
* Cross-account IAM Roles. Prog and Console access

**s3 replication** doesn't replicate old objects

s3 replication doesn't replicate deletes

cross region rep (CRRO) - versioning must be enabled on source and destination buckets

**S3 Transfer Acceleration:**

**Cloudfront is a CDN:**

* Edge location is a location where content is cached
* Origin is the location of the data
* Distribution is name given to CDN which consists of various Edge locations
* Web Distribution – used for websites
* RTMP – used for media streaming

Edge locations are not just read only – can write to them

Objects are cached for the life of the TTL (Time to live)

You can clear cached objects from an edge location, but are charged for them.

**Snowball**:

Petabyte-scale transport solution. Getting data into and out of AWS.

Can have 50TB or 80TB.

Snowball Edge is 100TB data transfer service with on-board compute capabilite.s

Storage Gateway (store stuff in s3):

Types:

* File Gateway (NFS and SMB)
* Volume Gateway (iSCSI)
  + Stored Volumes (OS or hdd images) – Entire Dataset is stored on site and asynchronously backed up to S3
  + Cached Volumes – entire dataset is stored on S3 and most frequently accessed data is cached on site
* Tape Library

**Athena vs. Macie**

* Athena: Interactive query service which enables you to analyze and query data in S3 using standard SQL
  + Serverless, pay per query and per TB scanned
  + Don’t need to set up complex Extract/Transform/Load (ETL) processes
  + Works directly with data stored in s3
  + USE CASES: query log files in s3,
    - generate business reports on data stored in s3
    - analyze aws cost and usage reports
    - run queries on click-stream data
* Macie: Security service which uses NLP to discover classify and protect sensitive data in S3
  + Use AI to recognize if s3 objects contain PII
  + Dashboards, reporting, and alters
  + Works directly with data stored in s3
  + Can analyze Cloudtrail logs
  + Great for PCI-DSS and preventing ID theft

IAM and S3 Summary:

* Mostly covered above. Again read s3 FAQ

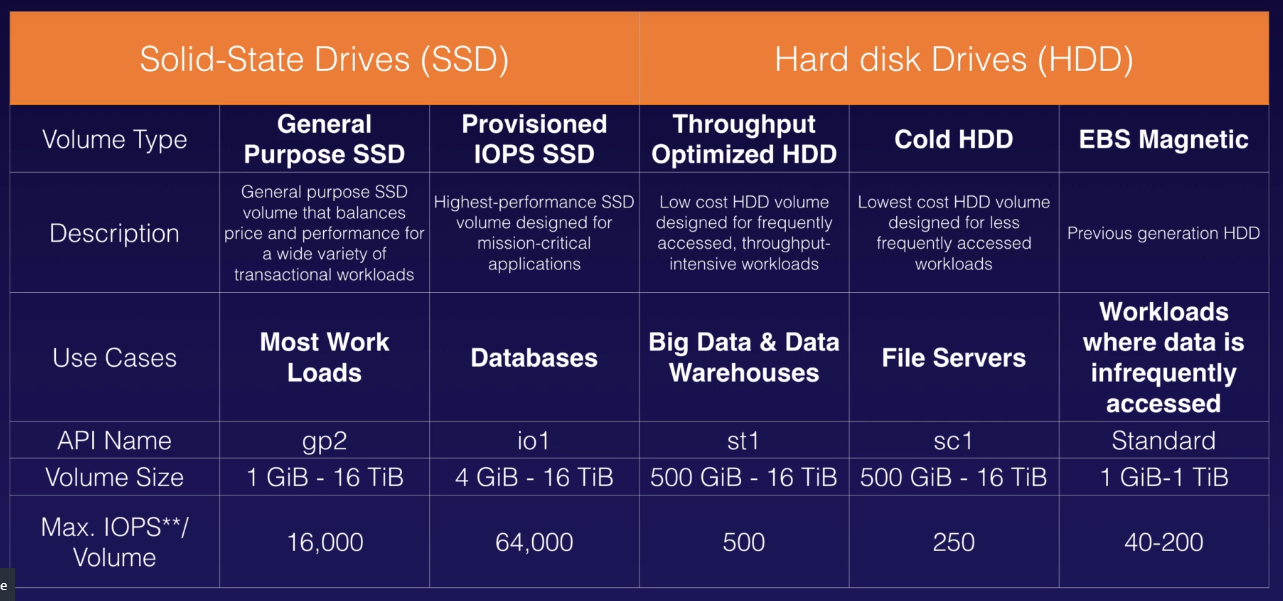
**EC2**

*Security Groups* are stateful. NACLs (VPC section) are stateless. In security groups, when an inbound rule is created, and outbound rule is created automatically (stateful).

* You need proper security group setting on the side of the request originator (responses come free)
* You cannot **block** IP addresses using security groups (use NACLs instead)

*Elastic Block Storage (EBS)* –

* Each EBS volume is replicated automatically within its AZ

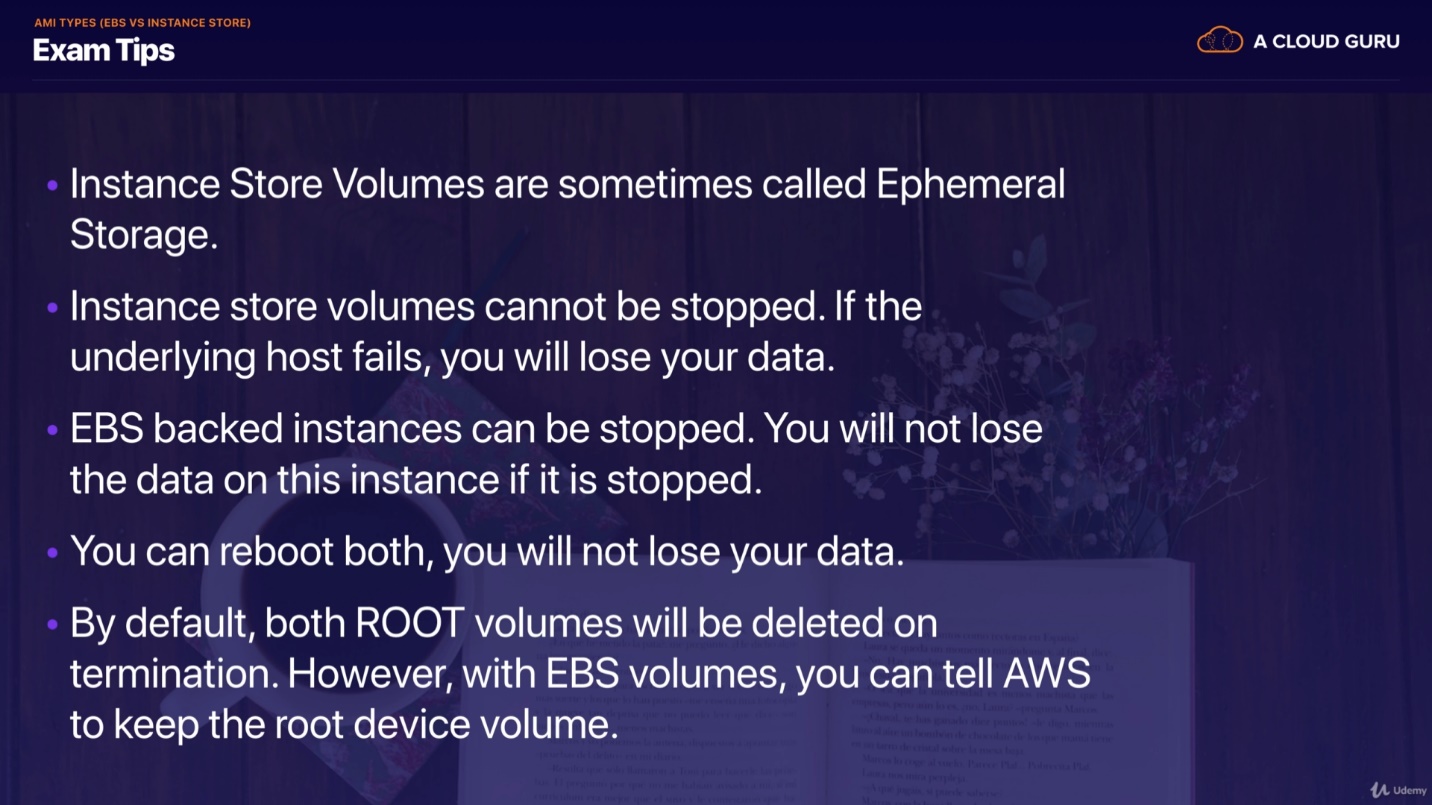


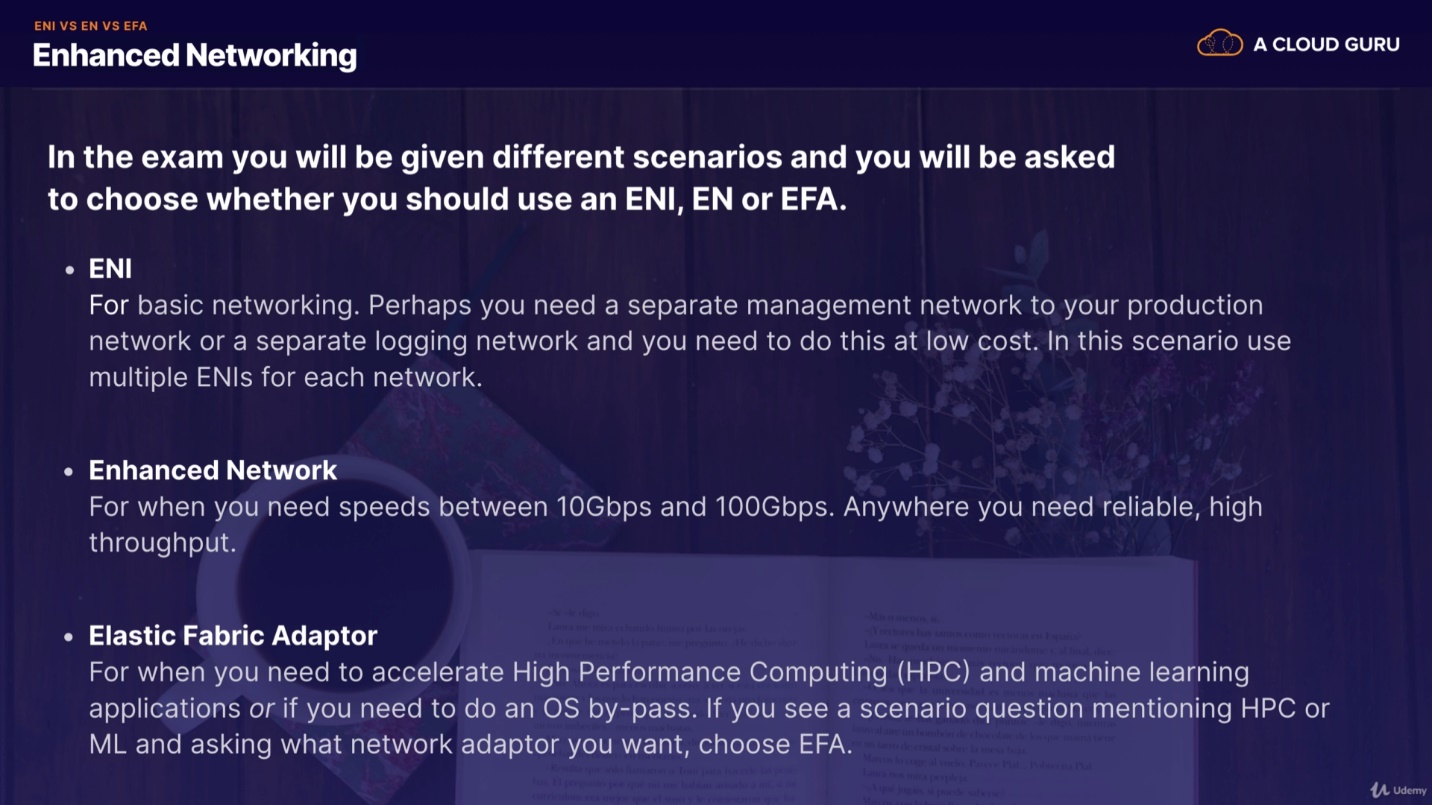
\*\* know the API name row for above table \*\*

* The EBS volume attached to an instance needs to be in the same AZ
* To move an EBS volume from one AZ to another, create a snapshot and create an AMI. Use the AMI to launch instance in different AZ
  + Can copy AMI by selecting AMI > Actions > Copy AMI > Fill out Destination, Name, Description

AMI types (EBS vs. instance store)

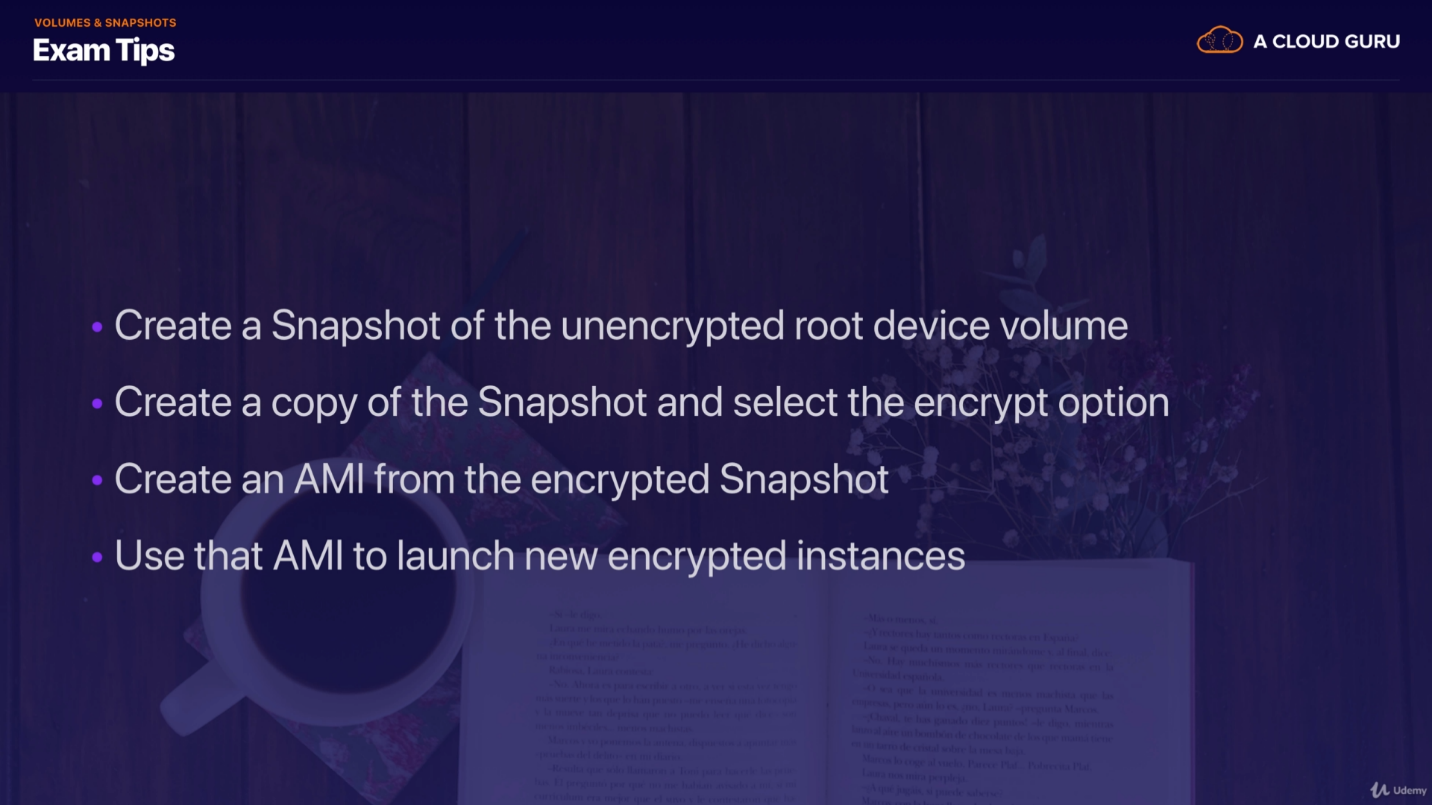
* AMIs are categorized as backed by Amazon EBS or instance store
* The root device of an ec2 is EBS if created from EBS snapshot, and it is and instance store volume if created from a template stored in S3







Encrypting an unencrypted root volume



**Spot instances:**

Spot fleet is a collection of spot instances

You can block spot instances from terminating using spot block

**CloudWatch vs. CloudTrail:**

Cloud watch is resource performance monitoring whereas Cloud Trail is monitoring (auditing) AWS Console and API calls (records IP, time, etc.)

* Cloudwatch can monitor at a minimum of once per minute

Stopped at: Bootstrap/bash scripting