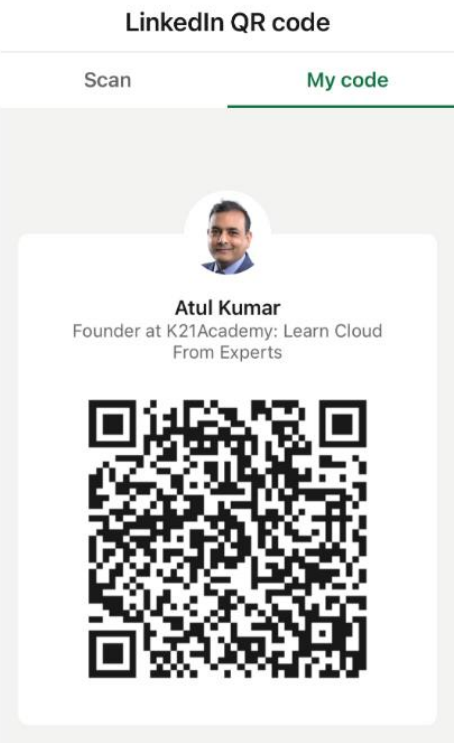
K21Academy
Learn Cloud from Experts

amazon web services

Amazon EC2

Atul Kumar
Author & Cloud Expert

Atul Kumar



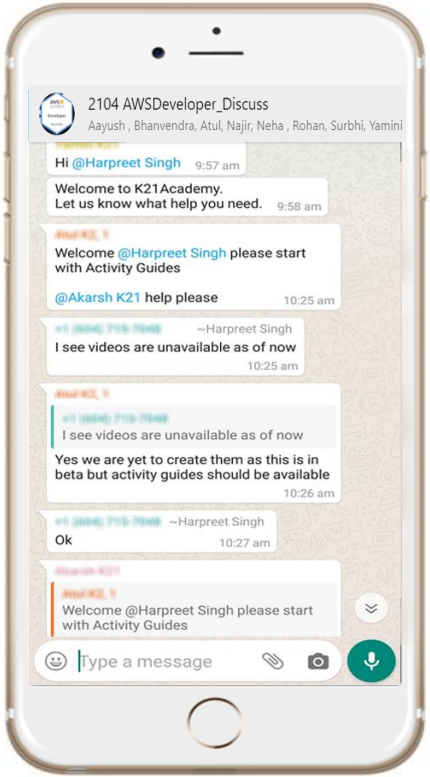
- Author & Cloud Architect
- 21+ Years working in IT & Certified Cloud Architect
- Helped **6000+ individuals** to learn Cloud & Cloud Native

WhatsApp & Ticketing System

K21Academy
Learn Cloud from Experts

support@k21academy.com

zendesk



Module Agenda

Module: Agenda

- Amazon Elastic Compute Cloud (EC2) and its benefits
- Amazon Machine Image (AMI) Security Groups In AWS
- Authentication Through Key-Pair
- Hosting a Website inside EC2
- Creating A Custom AMI
- Hardware Tenancy - Shared v/s Dedicated
- Networking Layer In EC2: VPC
- Elastic Network Interface and Its Attributes
- Different Categories Of IP Address

Module: Agenda

- Public IP v/s Elastic IP
- AWS Storage Services and How to select them
- Instance Store
- Elastic Block Store (EBS), its features and Volume types
- Solid State Drive: General Purpose SSD and Provisioned IOPS
- Hard Disk Drive: Throughput Optimized HDD and Cold HOD
- EBS Snapshots
- Elastic File System (EFS) and its Features
- EBS v/s EFS
- Cost Optimization

K21Academy
Learn Cloud from Experts

Amazon Elastic Compute Cloud (EC2)

Amazon EC2

- EC2 is a web service that provides servers in the cloud which are customized as per need
- It is highly ***scalable*** and works on ***pay-as-you-go*** model

Benefits of EC2

Tags

K21Academy
Learn Cloud from Experts

- A tag is a label that you assign to an AWS resource.
- Used to manage AWS assets.
- Tags are just arbitrary name/value pairs that you can assign to virtually all AWS assets to serve as metadata.
- Each tag consists of a key and an optional value, both of which you define.
- Tagging strategies can be used for cost allocation, security, automation, and many other uses. For example, you can use a tag in an IAM policy to implement access control.
- Enforcing standardized tagging can be done via AWS Config rules or custom scripts. For example, EC2 instances not properly tagged are stopped or terminated daily.
- Most resources can have up to 50 tags.

Metadata and User Data

K21Academy
Learn Cloud from Experts

- User data is data that is supplied by the user at instance launch in the form of a script.
- Instance metadata is data about your instance that you can use to configure or manage the running instance.
- User data is limited to 16KB.
- User data and metadata are not encrypted.
- **Instance metadata is available at http://169.254.169.254/latest/meta-data/** (the trailing "/" is required).

Metadata and User Data

K21Academy
Learn Cloud from Experts

- Instance user data is available at: http://169.254.169.254/latest/user-data.
- The I**P address 169.254.169.254 is a link-local address** and is valid only from the instance.
- On Linux you can use the curl command to view metadata and user data, e.g. “curl http://169.254.169.254/latest/meta-data/”.
- The **Instance Metadata Query tool** allows you to query the instance metadata without having to type out the full URI or category names.

K21Academy
Learn Cloud from Experts

Amazon Machine Image (AMI)

AMI

K21Academy
Learn Cloud from Experts

- **Amazon Machine Image (AMI)** provides the information required to launch an instance.
- AMIS are region specific, if you need to use an AMI in another region you can copy an AMI into the destination region via **Copy AMI**.
- You can create an AMI from an existing EC2 instance that's either running or stopped.
- **Community AMI** are free AMIs maintained by the community.
- **AWS Marketplace** free or paid subscription AMIs maintained by vendors.

AMI

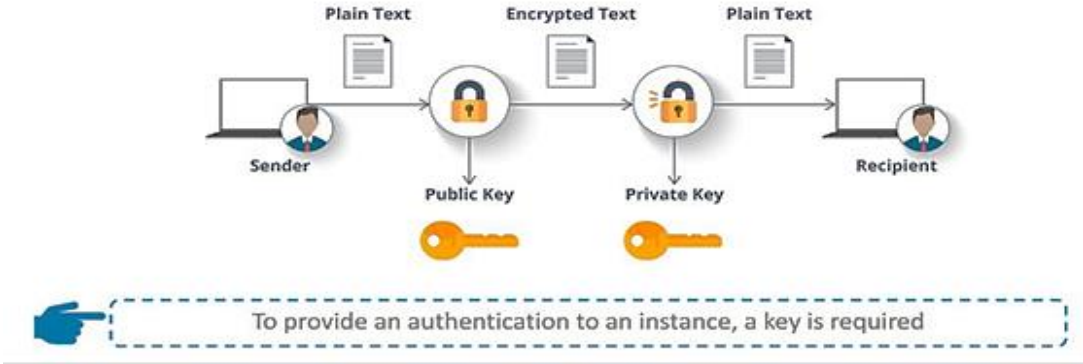
- AMIs have an **AMI ID**. The same AMI e.g. (Amazon Linux 2) will vary in both AMI ID and options e.g., Architecture options in different regions.
- An AMI holds the following information:
  - A template for the root volume for the instance (EBS Snapshot or Instance Store template) e.g., an operating system, an application server, and applications.
  - Launch permissions that control which AWS accounts can use the AMI to launch instances.
  - A block device mapping that specifies the volumes to attach to the instance when it's launched.
- AMIs are regional. You can only launch an AMI from the region in which it is stored. However, you can copy AMI's to other regions using the console, command line, or the API.

K21Academy
Learn Cloud from Experts

Authentication Through Key-Pair

Key Pair

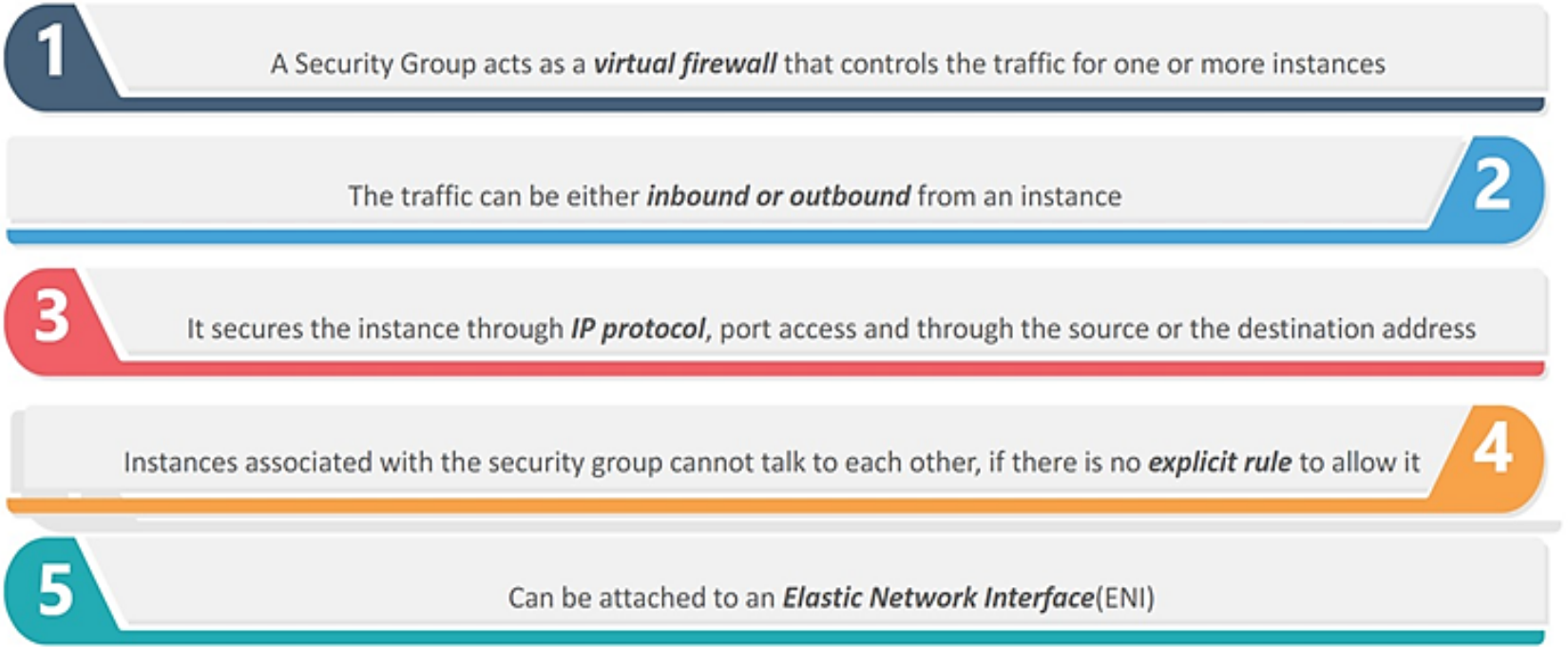
- Public key + Private Key = Key Pair
- Amazon EC2 uses public and private key cryptography to encrypt and decrypt information while connecting to EC2.
- Public key is used to encrypt a data, while the recipient uses a private key to decrypt the same.
- AWS issues .pem file, a client needs to convert it to a format which is recognized by client software.



K21Academy
Learn Cloud from Experts

AWS
Security Groups

Security Groups



K21Academy
Learn Cloud from Experts

AWS
Network ACL

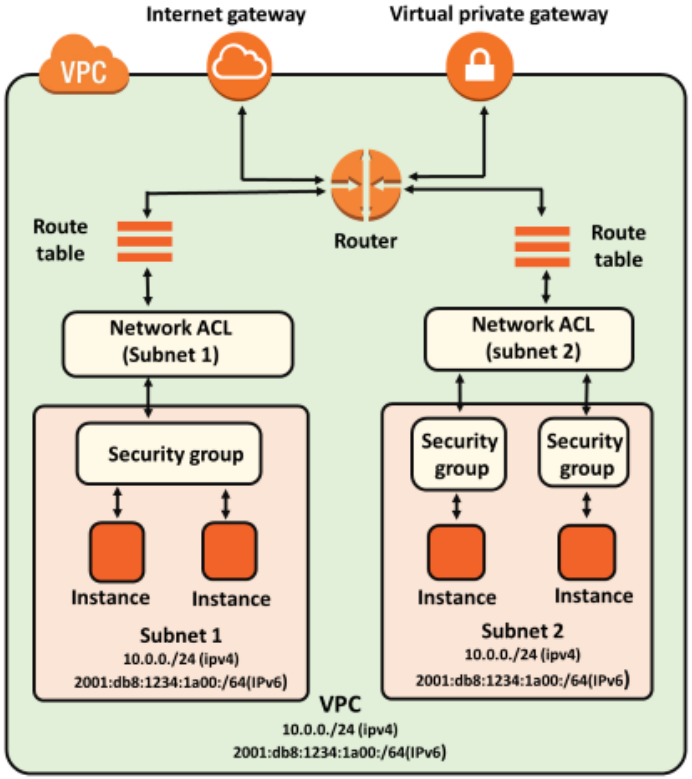
Network ACL

- NACL also adds an additional layer of security associated with subnets that control both inbound and outbound traffic at the subnet level.
- Maximum number of rules that exist per NACL: 20

K21Academy
Learn Cloud from Experts

Difference Between Security Group And Network ACL

Security Group Vs Network ACL



Security Group Vs Network ACL

K21Academy
Learn Cloud from Experts

Security Group	Network ACL
Operates at the instance level (first layer of defense)	Operates at the subnet level (second layer of defense)
Supports allow rules only	Supports allow rules and deny rules
Is stateful: Return traffic is automatically allowed, regardless of any rules	Is stateless: Return traffic must be explicitly allowed by rules
We evaluate all rules before deciding whether to allow traffic	We process rules in number order when deciding whether to allow traffic
Applies to an instance only if someone specifies the security group when launching the instance, or associates the security group with the instance later on	Automatically applies to all instances in the subnets it's associated with (backup layer of defense, so you don't have to rely on someone specifying the security group)

K21Academy
Learn Cloud from Experts

EC2
Hardware Tenancy

Hardware Tenancy – Shared Vs Dedicated

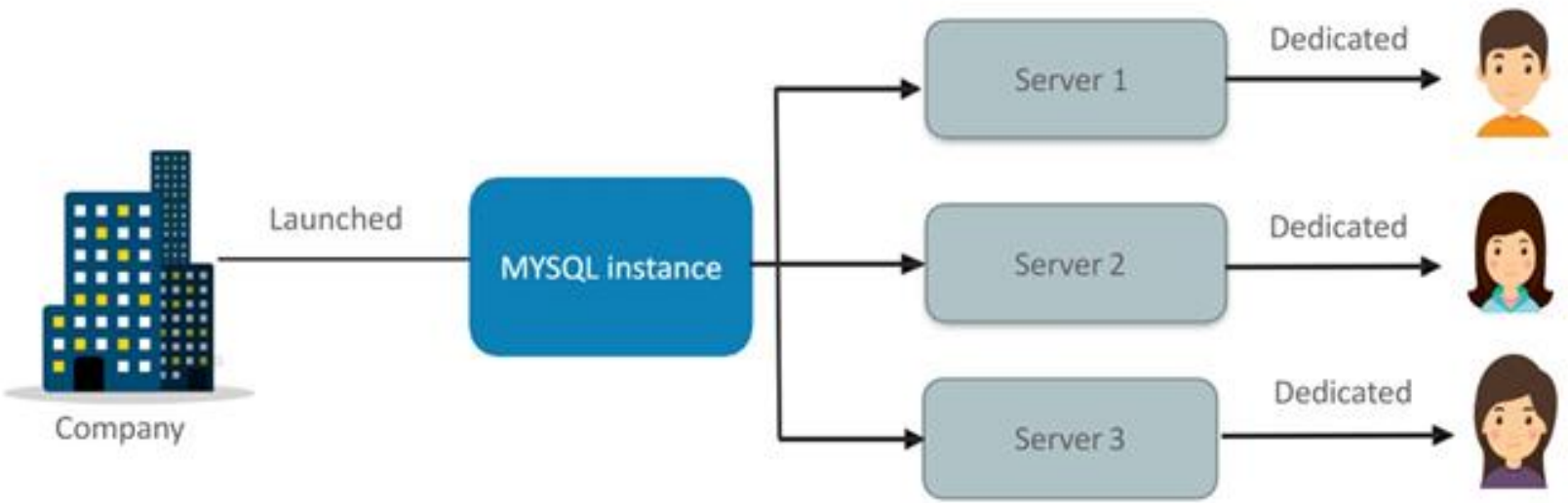
K21Academy
Learn Cloud from Experts

- Tenancy determines the owner of a resource.
- AWS provides two types of the tenancy to comply with your Organization Regulatory Security.

Shared	Dedicated
A single physical machine runs multiple instances which are launched by several AWS customers	When an instance is launched, it will ensure that it will run only on single-tenant hardware
All the customers are served from the same common hardware infrastructure	Each customer gets his own machine to run their instance

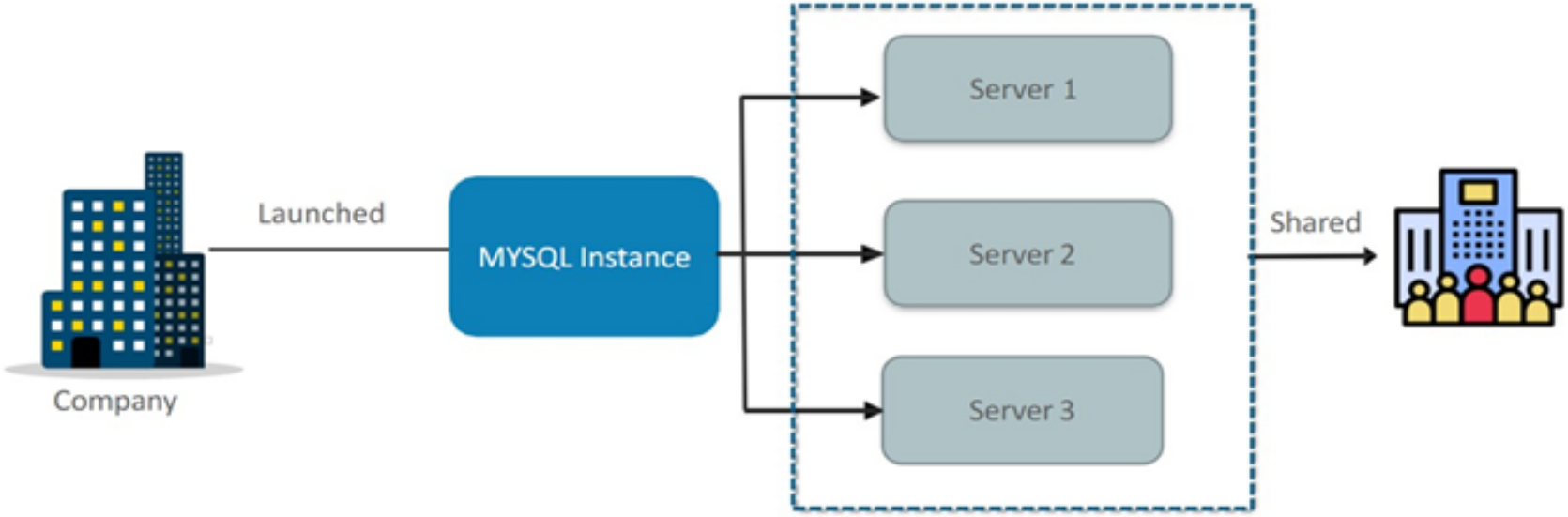
Hardware Tenancy - Dedicated

If you are opting for ***Dedicated Instance***, Each user will get separate hardware.



Hardware Tenancy - Shared

For ***Shared Instance***, Servers will be shared across users.



K21Academy
Learn Cloud from Experts

EC2 Purchasing Options

EC2 Pricing Model

K21Academy
Learn Cloud from Experts

On-Demand Least Commitment

- low cost and flexible
- only pay per hour
- short-term, spiky, unpredictable workloads
- cannot be interrupted
- For first time apps

Spot upto 90% Biggest Savings

- request spare computing capacity
- flexible start and end times
- Can handle interruptions (server randomly stopping and starting)
- For non-critical background jobs

Reserved upto 75% off Best Long-term

- steady state or predictable usage
- commit to EC2 over a 1 or 3 year term
- Can resell unused reserved instances

Dedicated Most Expensive

- Dedicated servers
- Can be on-demand or reserved (upto 70% off)
- When you need a guarantee of isolate hardware (enterprise requirements)

EC2 On-Demand Instance

K21Academy
Learn Cloud from Experts

When you launch an EC2 instance it is by default using **On-Demand** Pricing

On-demand has **no up-front payment** and **no long-term commitment**

You are charged by the **hour** or by the **minute** (varies based on EC2 Instance Types)

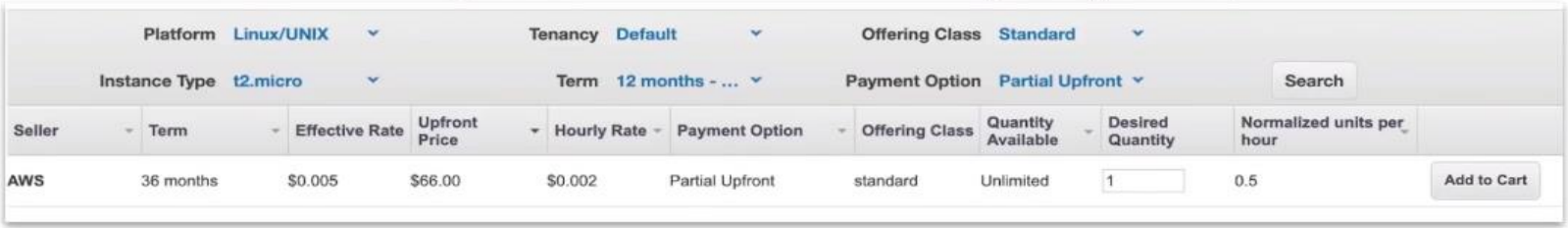
On-Demand is for applications where the workload is for **short-term**, **spikey** or **unpredictable**.
When you have a **new app** for development or you want to run experiment.

EC2 Reserved Instances(RI)

K21Academy
Learn Cloud from Experts

Designed for applications that have a **steady-state**, **predictable usage**, or require **reserved capacity**.

Reduced Pricing is based on **Term** x **Class Offering** x **Payment Option**



Standard Up to **75%** reduced pricing compared to on-demand.
Cannot change RI Attributes.

Convertible Up to **54%** reduced pricing compared to on-demand.
Allows you to change RI Attributes if greater or equal in value.

Scheduled You reserve instances for specific time periods eg. once a week for a few hours. Savings vary

Terms
You commit to a **1 Year** or **3 Year** contract.
The longer the term the greater savings.

Payment Options
All Upfront, **Partial Upfront**, and **No Upfront**
The greater upfront the great the savings

RIs can be **shared between multiple accounts** within an org

Unused RIs can be sold in the **Reserved Instance Marketplace**

EC2 Reserved Instances(RI)

K21Academy
Learn Cloud from Experts

- When you purchase a Reserved Instance, you determine the scope of the Reserved Instance. The scope is either regional or zonal.
- **Regional**: When you purchase a Reserved Instance for a Region, it's referred to as a *regional* Reserved Instance.
- **Zonal**: When you purchase a Reserved Instance for a specific Availability Zone, it's referred to as a *zonal* Reserved Instance.
- The scope does not affect the price.

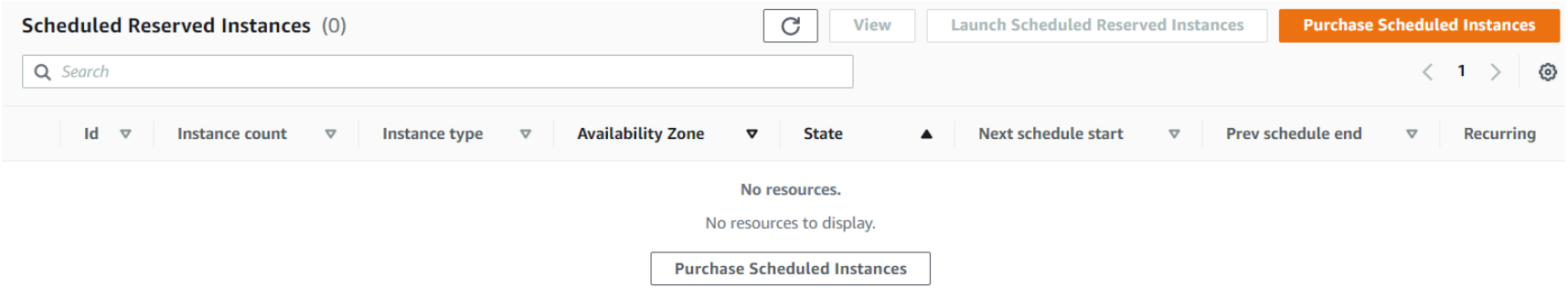
EC2 Reserved Instances(RI)

K21Academy
Learn Cloud from Experts

	Regional Reserved Instances	Zonal Reserved Instances
Ability to reserve capacity	A regional Reserved Instance does *not* reserve capacity.	A zonal Reserved Instance reserves capacity in the specified Availability Zone.
Availability Zone flexibility	The Reserved Instance discount applies to instance usage in any Availability Zone in the specified Region.	No Availability Zone flexibility—the Reserved Instance discount applies to instance usage in the specified Availability Zone only.
Instance size flexibility	The Reserved Instance discount applies to instance usage within the instance family, regardless of size.	No instance size flexibility—the Reserved Instance discount applies to instance usage for the specified instance type and size only.
Queuing a purchase	You can queue purchases for regional Reserved Instances.	You can't queue purchases for zonal Reserved Instances.

Scheduled Reserved Instances

- Scheduled Reserved Instances, you can reserve capacity that is scheduled to recur daily, weekly, or monthly, with a specified start time and duration, for a one-year term.
- After you complete your purchase, the instances are available to launch during the time windows that you specified.



Scheduled Reserved Instances

K21Academy
Learn Cloud from Experts

- **Instance attributes**

  A Reserved Instance has four instance attributes that determine its price.
- **Instance type**: For example, m4.large. This is composed of the instance family (for example, m4) and the instance size (for example, large).
- **Region**: The Region in which the Reserved Instance is purchased.
- **Tenancy**: Whether your instance runs on shared (default) or single-tenant (dedicated) hardware. For more information, see Dedicated Instances.
- **Platform**: The operating system; for example, Windows or Linux/Unix..

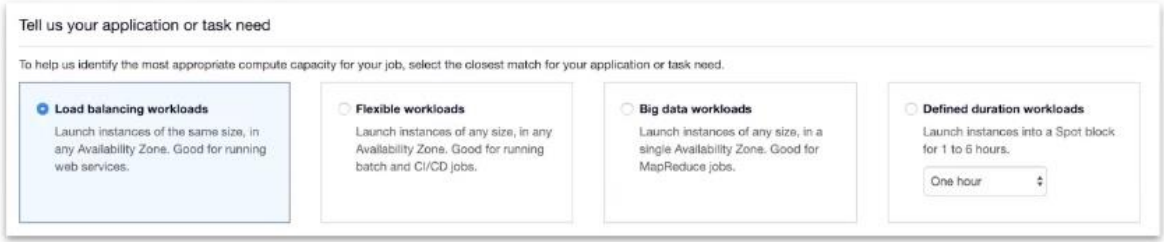
EC2 Spot Instances

K21Academy
Learn Cloud from Experts

AWS has **unused compute capacity** that they want to maximize the utility of their idle servers.
It's like when a hotel offers discounts for to fill vacant suites or planes offer discount to fill vacant seats.

Spot Instances provide a discount of **90%** compared to On-Demand Pricing
Spot Instances can be terminated if the computing capacity is needed by on-demand customers.

Designed for applications that have flexible start and end times or applications that are only feasible at **very low** compute costs.



AWS Batch is an easy and convenient way to use Spot Pricing

Termination Conditions
Instances can be terminated by AWS **at anytime**
If your instance is **terminated by AWS**, **you don't get charged** for a partial hour of usage.
If **you terminate** an instance **you will still be charged** for any hour that it ran.

Dedicated Host vs Dedicated Instance

K21Academy
Learn Cloud from Experts

- **EC2 Dedicated Host :** An Amazon EC2 Dedicated Host is a physical server with EC2 instance capacity fully dedicated to your use. Dedicated Hosts allow you to use your existing per-socket, per-core, or per-VM software licenses, including Windows Server, Microsoft SQL Server, SUSE, and Linux Enterprise Server.
- Support for multiple instance sizes on the same Dedicated Host is available for the following instance families: A1, C5, M5, R5, C5n, R5n, and M5n. Other instance families support only a single instance size on the same Dedicated Host.

Dedicated Host vs Dedicated Instance

K21Academy
Learn Cloud from Experts

- **Dedicate Instances:** Dedicated Instances are Amazon EC2 instances that run in a virtual private cloud (VPC) on hardware that's dedicated to a single customer. Dedicated Instances that belong to different AWS accounts are physically isolated at a hardware level, even if those accounts are linked to a single payer account. However, Dedicated Instances may share hardware with other instances from the same AWS account that are not Dedicated Instances.

Dedicated Host vs Dedicated Instance

K21Academy
Learn Cloud from Experts

	Dedicated Host	Dedicated Instance
Billing	Per-host billing	Per-instance billing
Visibility of sockets, cores, and host ID	Provides visibility of the number of sockets and physical cores	No visibility
Host and instance affinity	Allows you to consistently deploy your instances to the same physical server over time	Not supported
Targeted instance placement	Provides additional visibility and control over how instances are placed on a physical server	Not supported
Automatic instance recovery	Supported. For more information, see Host recovery.	Supported
Bring Your Own License (BYOL)	Supported	Not supported

On Demand Capacity Reservation

K21Academy
Learn Cloud from Experts

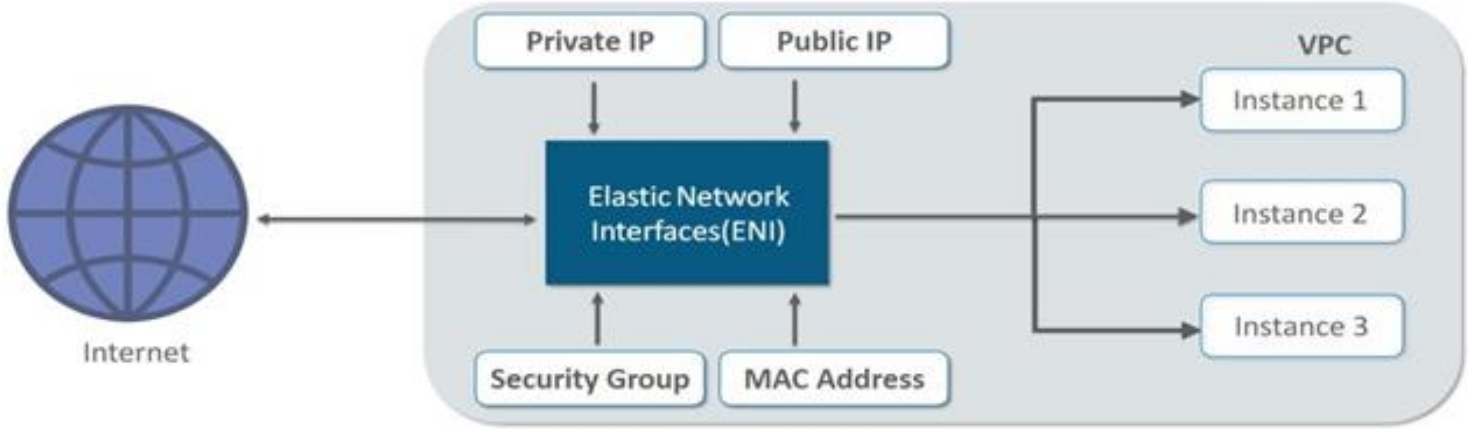
- On-Demand Capacity Reservations enable you to reserve compute capacity for your Amazon EC2 instances in a specific Availability Zone for any duration. This gives you the ability to create and manage Capacity Reservations independently from the billing discounts offered by Savings Plans or regional Reserved Instances.
- When you create a Capacity Reservation, you specify:
  - The Availability Zone in which to reserve the capacity
  - The number of instances for which to reserve capacity
  - The instance attributes, including the instance type, tenancy, and platform/OS

K21Academy
Learn Cloud from Experts

Amazon EC2 Networking Layer

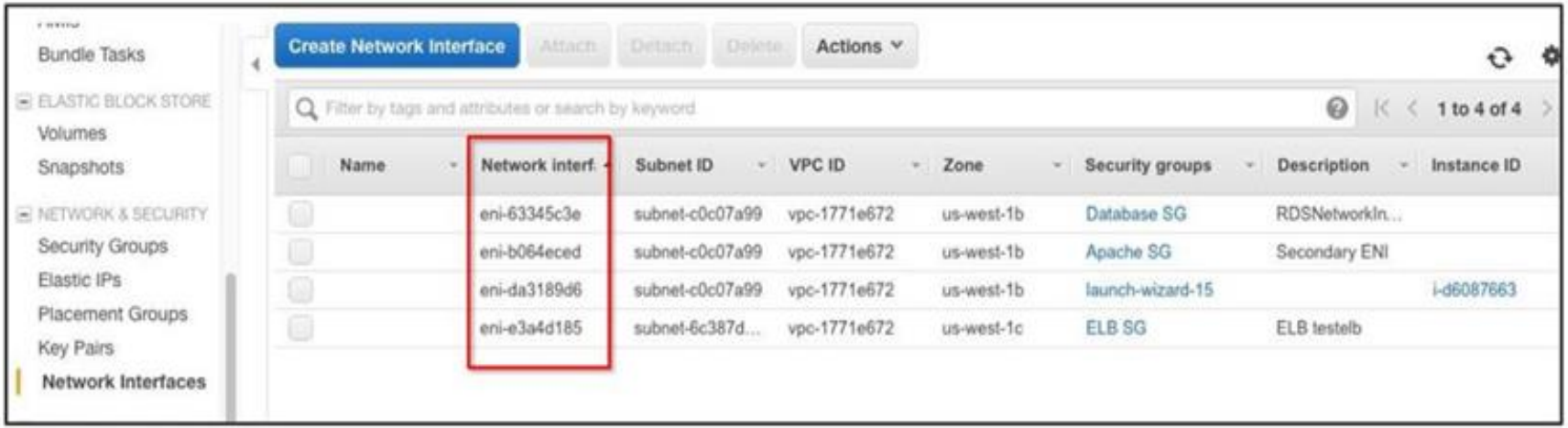
Elastic Network Interface

- An Elastic Network Interface (ENI) is a virtual network interface which acts as a point of interface between VM and network by attaching a public IP, private IP, security groups and many more to your instance



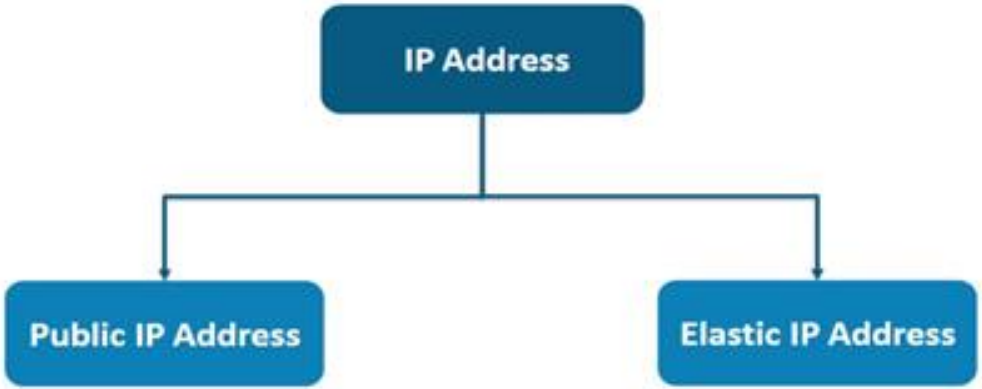
Why it is Elastic?

- Virtual network interface can be attached to multiple instance in a VPC
- An ENI detached from an EC2 instance can be attached to another instance
- Can be created using the Amazon EC2 console or the command line



Different Category Of IP Address

- In AWS, specific IP Address is given to each Instance which helps to communicate it to the server and to establish the connection between the machines

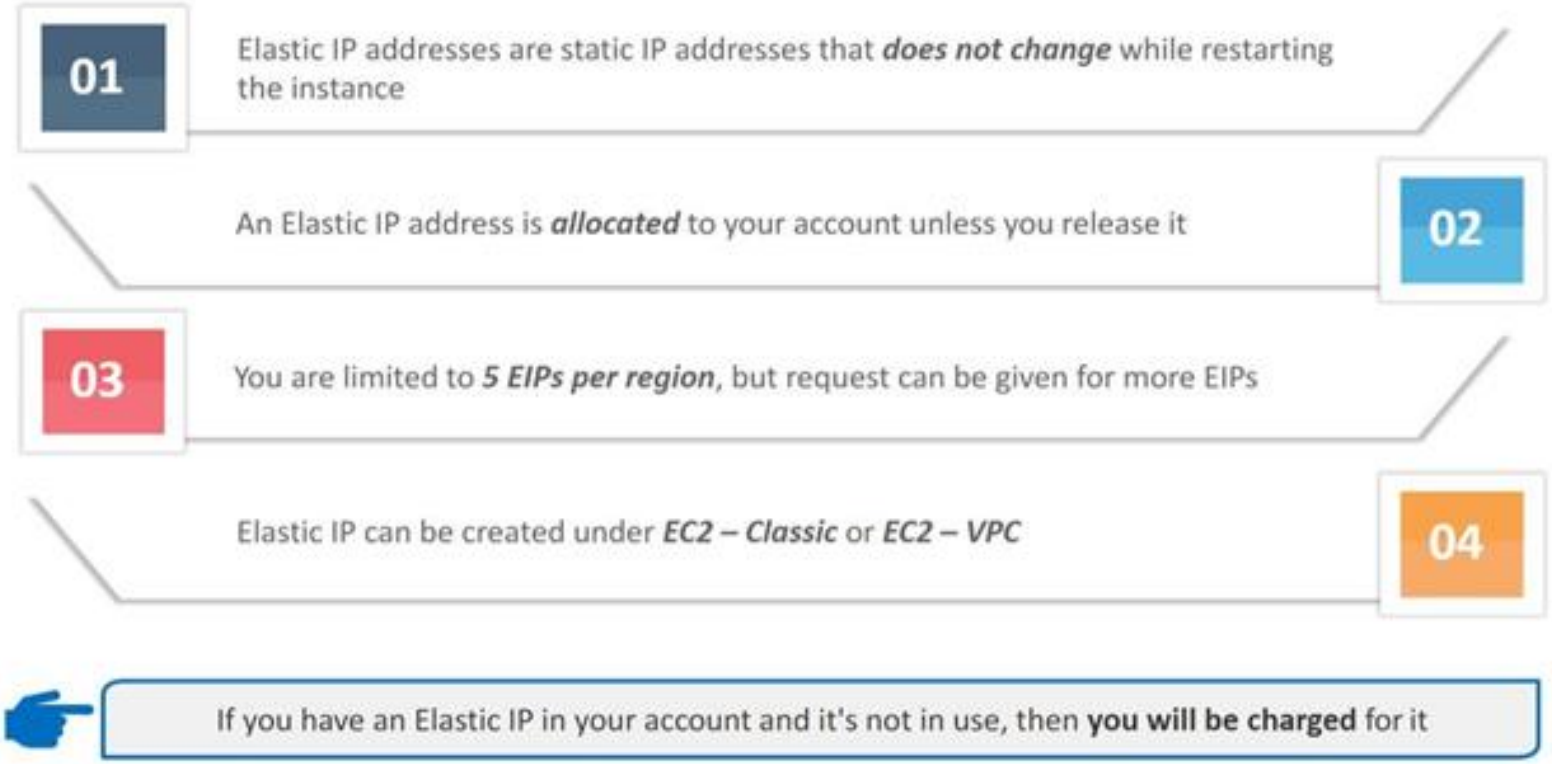


Public IP Vs Elastic IP

K21Academy
Learn Cloud from Experts

Public IP	Elastic IP
It is assigned to your launched instance.	It is assigned to your AWS account.
when an instance is terminated the public IP attached to it gets released and further when you relaunch the same instance new IP address is assigned.	Elastic IP do not change and they remain same even if you terminate the instance and later again restart the same instance.

Elastic IP Address



Elastic Network Adapter (ENA)

K21Academy
Learn Cloud from Experts

- **Enhanced networking provides higher bandwidth, higher packet-per-second (PPS) performance, and consistently lower inter-instance latencies**.
- Enhanced networking is enabled using an Elastic Network Adapter (ENA).
- If your packets-per-second rate appears to have reached its ceiling, you should consider moving to enhanced networking because you have likely reached the upper thresholds of the VIF driver.
- AWS currently supports enhanced networking capabilities using SR-IOV.

Elastic Network Adapter (ENA)

K21Academy
Learn Cloud from Experts

- SR-IOV provides direct access to network adapters, provides higher performance (packets-per-second) and lower latency.
- Must launch an HVM AMI with the appropriate drivers.
- Only available for certain instance types.
- Only supported in VPC.

Elastic Fabric Adapter (EFA)

K21Academy
Learn Cloud from Experts

- An Elastic Fabric Adapter is an AWS Elastic Network Adapter (ENA) with added capabilities.
- An EFA can still handle IP traffic, but also supports an important access model commonly called OS bypass.
- This model allows the application (most commonly through some user-space middleware) access the network interface without having to get the operating system involved with each message.
- Elastic Fabric Adapter (EFA) is a network interface for Amazon EC2 instances that enables customers to run applications requiring high levels of inter-node communications at scale on AWS.

Elastic Fabric Adapter (EFA)

K21Academy
Learn Cloud from Experts

- Its custom-built operating system (OS) bypass hardware interface enhances the performance of inter-instance communications, which is critical to scaling these applications.
- With EFA, High Performance Computing (HPC) applications using the Message Passing Interface (MPI) and Machine Learning (ML) applications using NVIDIA Collective Communications Library (NCCL) can scale to thousands of CPUs or GPUs.
- As a result, you get the application performance of on-premises HPC clusters with the on-demand elasticity and flexibility of the AWS cloud.
- EFA is available as an optional EC2 networking feature that you can enable on any supported EC2 instance at no additional cost.

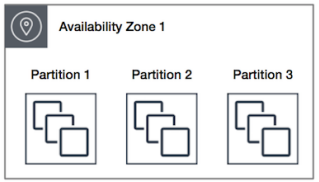
K21Academy
Learn Cloud from Experts

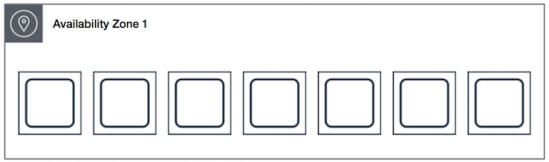
Amazon EC2 Placement Groups

Placement Groups

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

There is no charge for creating a placement group.





Limitations & Rules

K21Academy
Learn Cloud from Experts

General rules and limitations

Before you use placement groups, be aware of the following rules:

- The name that you specify for a placement group must be unique within your AWS account for the Region.
- You can't merge placement groups.
- An instance can be launched in one placement group at a time; it cannot span multiple placement groups.
- On-Demand Capacity Reservation provide a capacity reservation for EC2 instances in a specific Availability Zone. The capacity reservation can be used by instances in a placement group. However, it is not possible to explicitly reserve capacity for a placement group.
- You cannot launch Dedicated Hosts in placement groups.

Rules & limitations: Cluster

K21Academy
Learn Cloud from Experts

The following rules apply to **cluster placement groups**:

- Instances in a cluster placement group you must use the following supported instance types:
  - Current generation instances, except for burstable performance instances (for example, T2) and Mac1 instances.
  - The following previous generation instances: A1, C3, cc2.8xlarge, cr1.8xlarge, G2, hs1.8xlarge, I2, and R3.
- A cluster placement group can't span multiple Availability Zones.
- The maximum network throughput speed of traffic between two instances in a cluster placement group is limited by the slower of the two instances. For applications with high-throughput requirements, choose an instance type with network connectivity that meets your requirements.

Rules & limitations: Cluster

- For instances that are enabled for enhanced networking, the following rules apply:
  - Instances within a cluster placement group can use up to 10 Gbps for single-flow traffic. Instances that are not within a cluster placement group can use up to 5 Gbps for single-flow traffic.
  - Traffic to and from Amazon S3 buckets within the same Region over the public IP address space or through a VPC endpoint can use all available instance aggregate bandwidth.
- You can launch multiple instance types into a cluster placement group. However, this reduces the likelihood that the required capacity will be available for your launch to succeed. We recommend using the same instance type for all instances in a cluster placement group.
- Network traffic to the internet and over an AWS Direct Connect connection to on-premises resources is limited to 5 Gbps.

Rules & limitations: Partition

K21Academy
Learn Cloud from Experts

The following rules apply to **partition placement** groups:

- A partition placement group supports a maximum of seven partitions per Availability Zone. The number of instances that you can launch in a partition placement group is limited only by your account limits.
- When instances are launched into a partition placement group, Amazon EC2 tries to evenly distribute the instances across all partitions. Amazon EC2 doesn't guarantee an even distribution of instances across all partitions.
- A partition placement group with Dedicated Instances can have a maximum of two partitions.

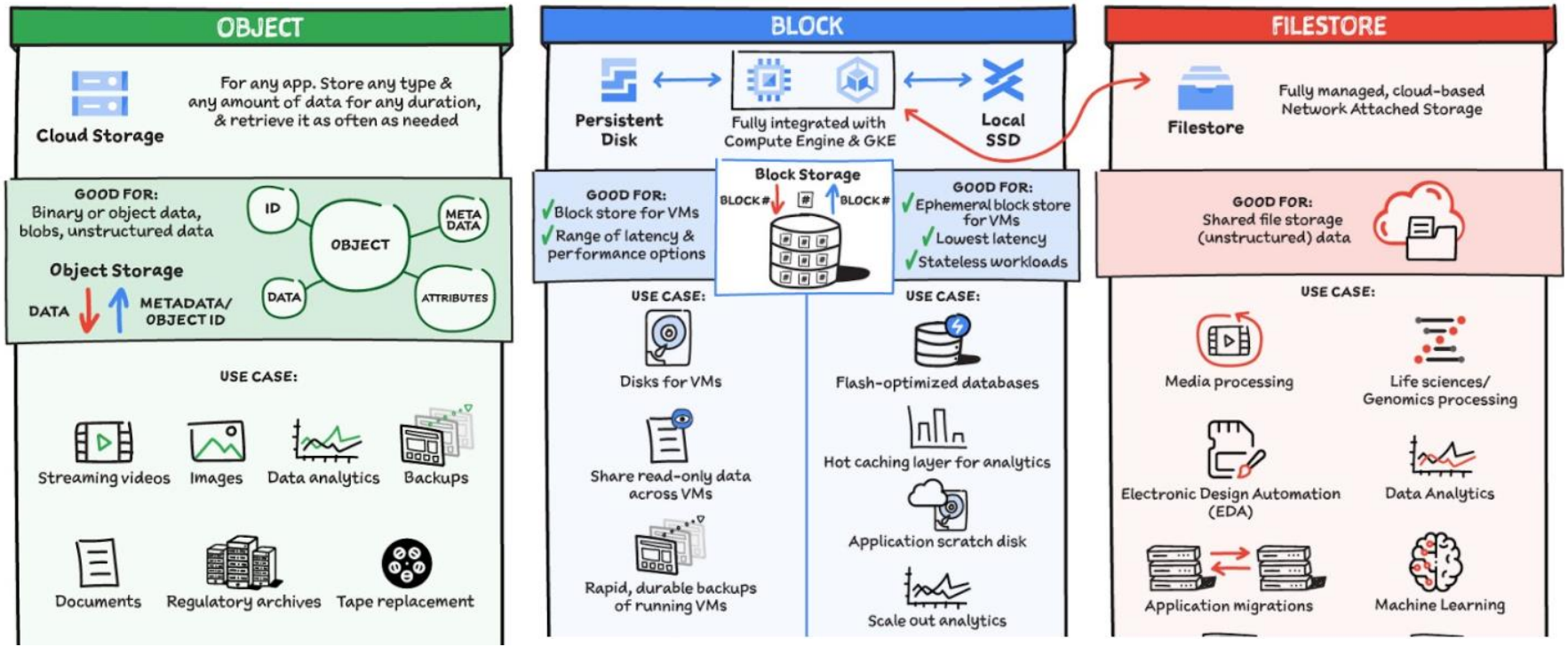
Rules & limitations: Spread

- A **spread placement group** supports a maximum of seven running instances per Availability Zone. For example, in a Region with three Availability Zones, you can run a total of 21 instances in the group (seven per zone)
- Spread placement groups are not supported for Dedicated Instances.

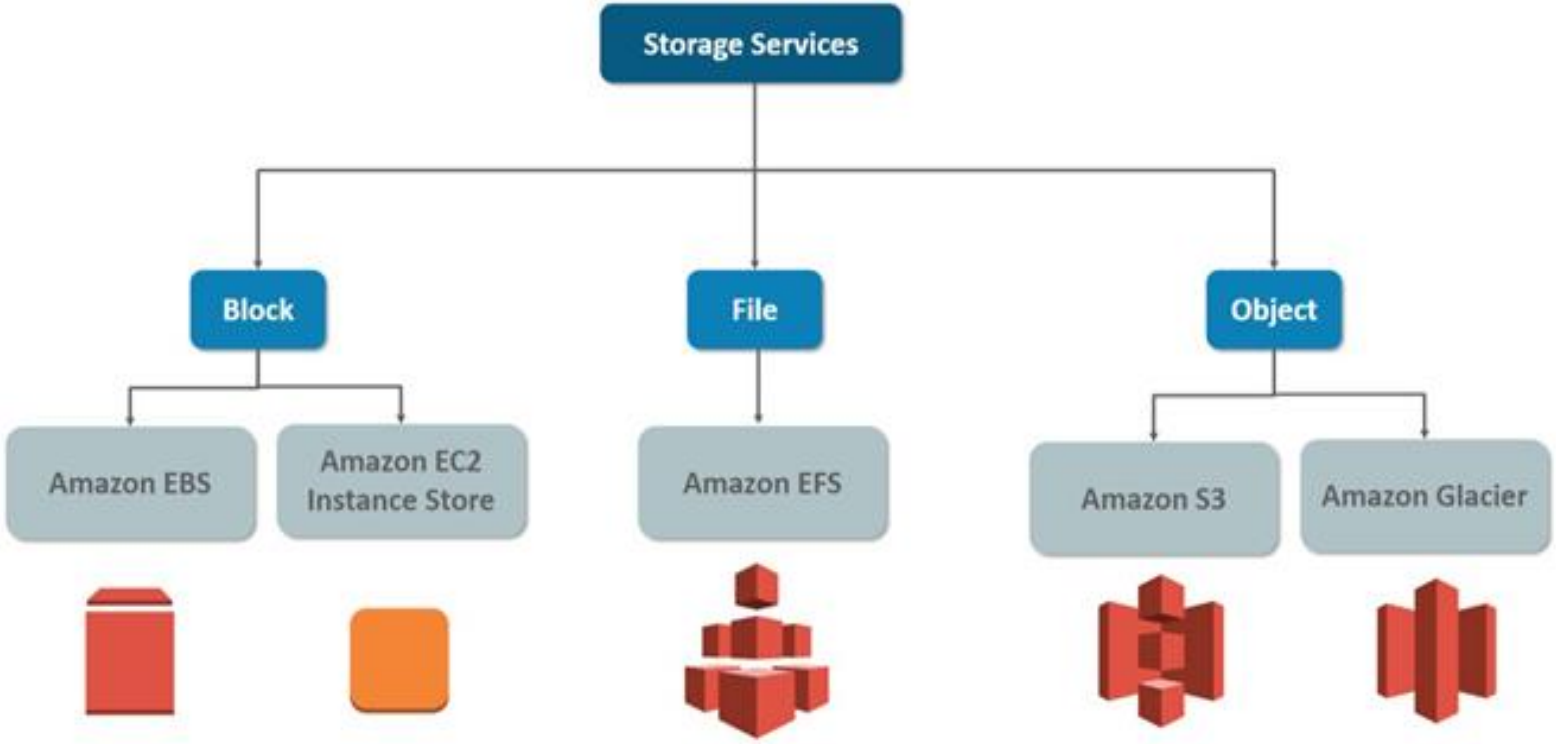
K21Academy
Learn Cloud from Experts

Amazon Storage Services

Storage Types



AWS Storage Services



Storage Services

K21Academy
Learn Cloud from Experts

- **What is object storage?**

  Object storage, also known as object-based storage, is a flat structure in which files are broken into pieces and spread out among hardware. In object storage, the data is broken into discrete units called objects and is kept in a single repository, instead of being kept as files in folders or as blocks on servers. Used Case : BigData, WebApps, Backup Archives

Storage Services

K21Academy
Learn Cloud from Experts

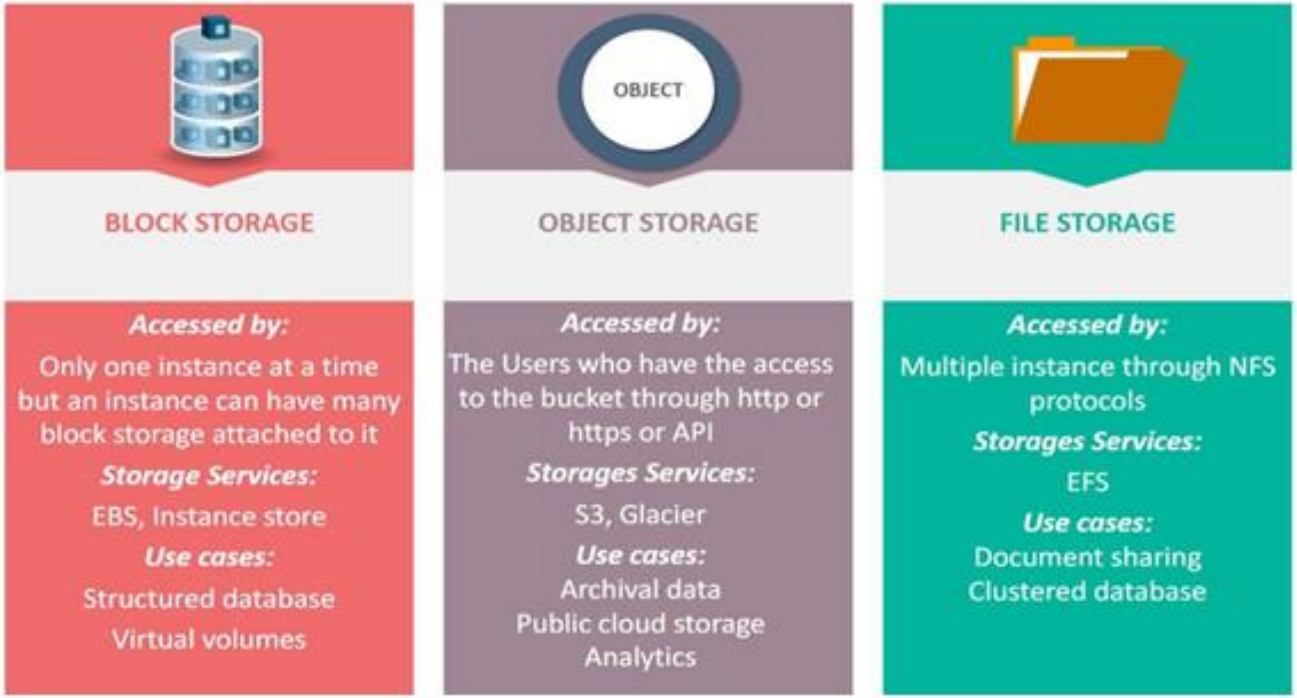
- **What is block storage?**

  Block storage chops data into blocks—get it?—and stores them as separate pieces. Each block of data is given a unique identifier, which allows a storage system to place the smaller pieces of data wherever is most convenient. **Use Cases : Databases, Email servers, RAID**

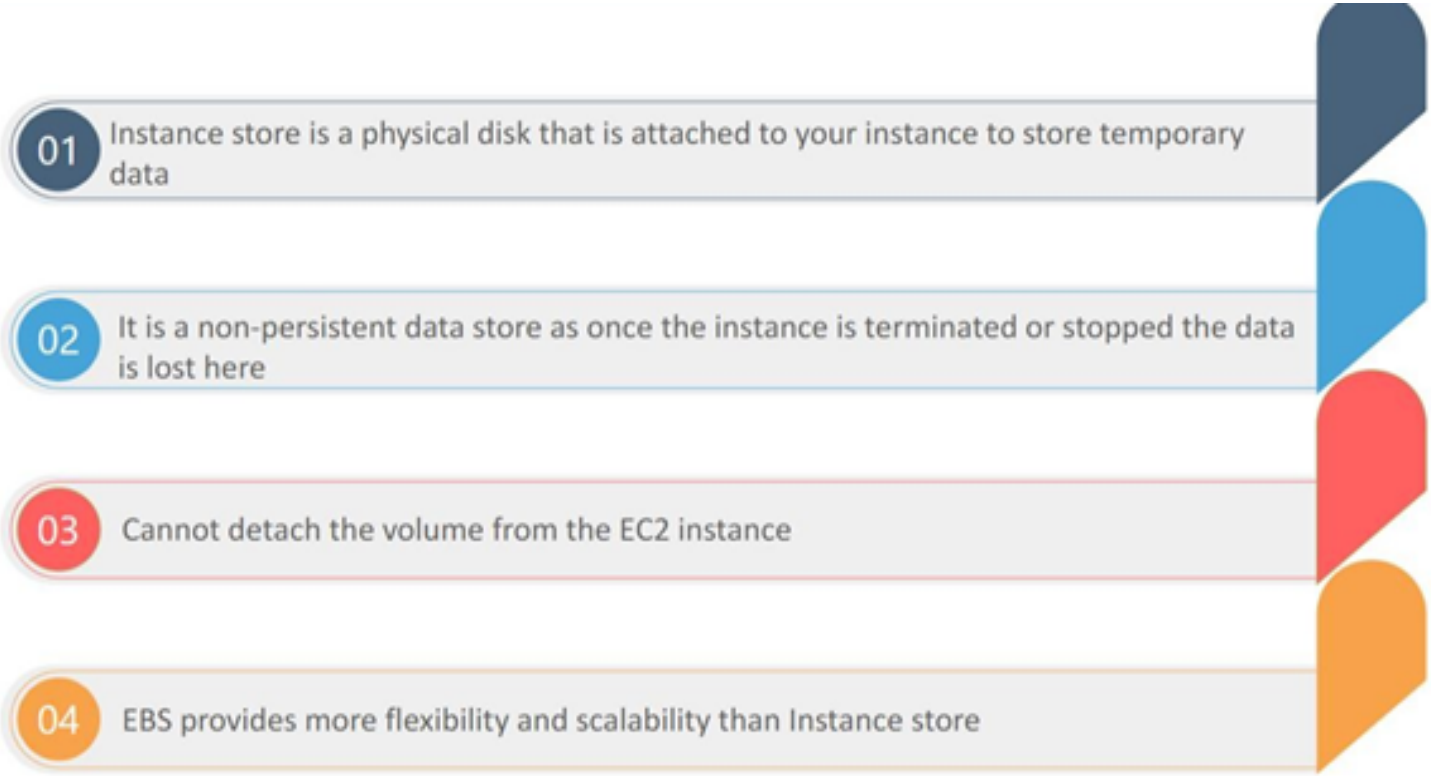
- **What is file storage?**

  Data is stored as a single piece of information inside a folder, just like you'd organize pieces of paper inside a manila folder. Used Case **File sharing, Local archiving, Data protection**

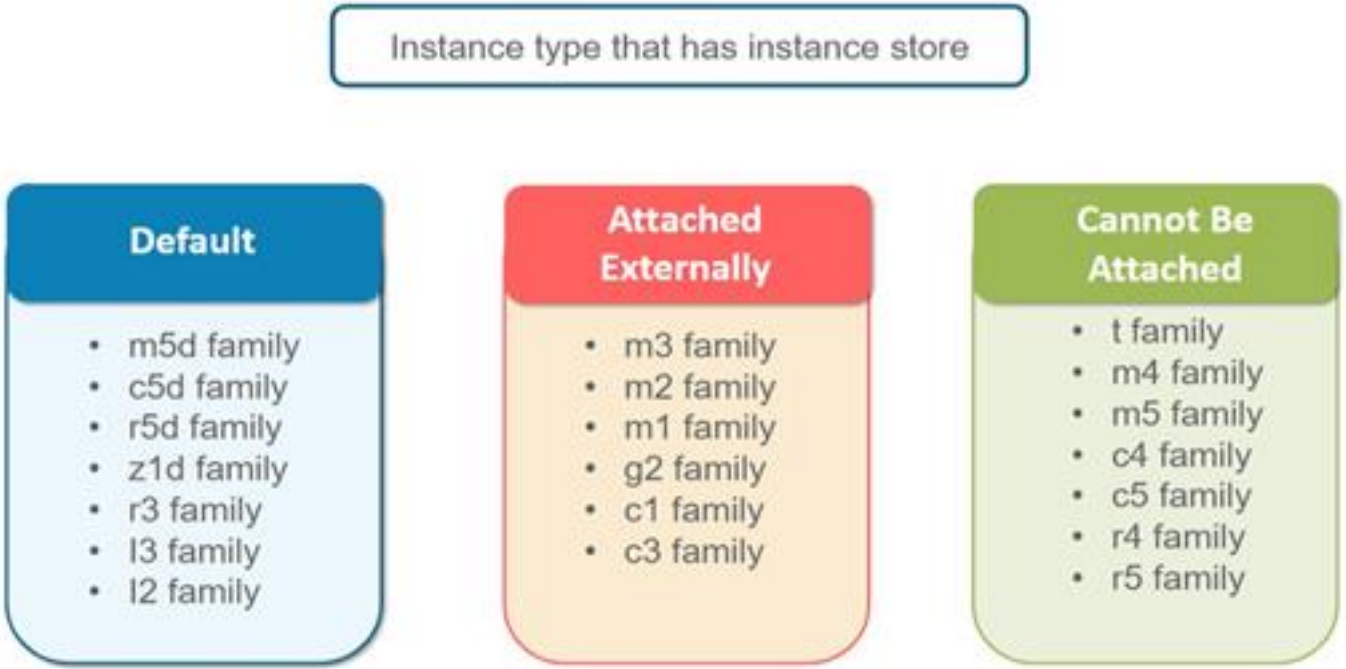
How To Select Storage Service?



Instance Storage



Instance Storage

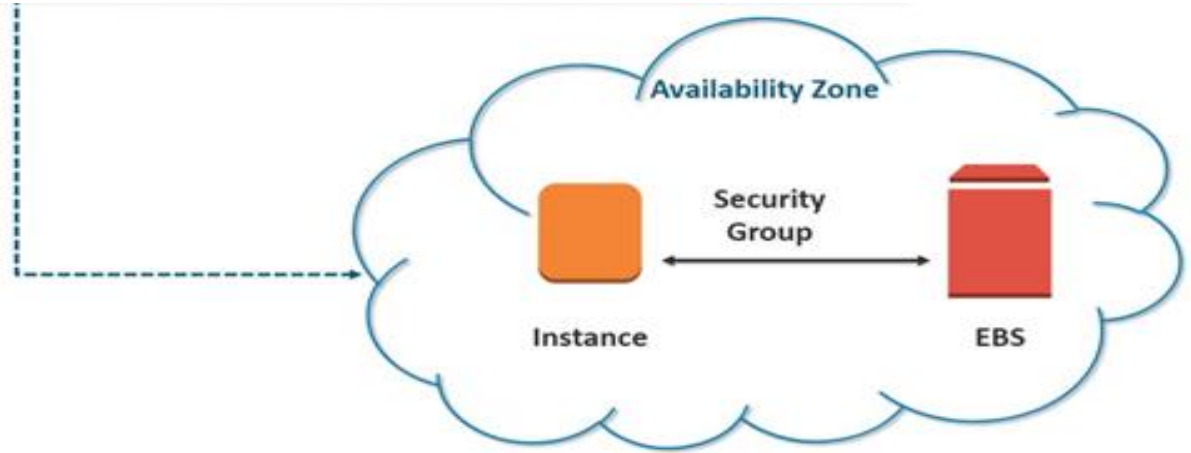


K21Academy
Learn Cloud from Experts

Elastic Block Storage (EBS)

What Is EBS?

- EBS is the logical volumes to use it with the EC2 instances
- This type of storage is used, when the data needs to be accessed quickly and required for the long time
- Lifetime of the EBS is not dependent on the EC2 instance
- Volume and instance must be in the same Availability Zone

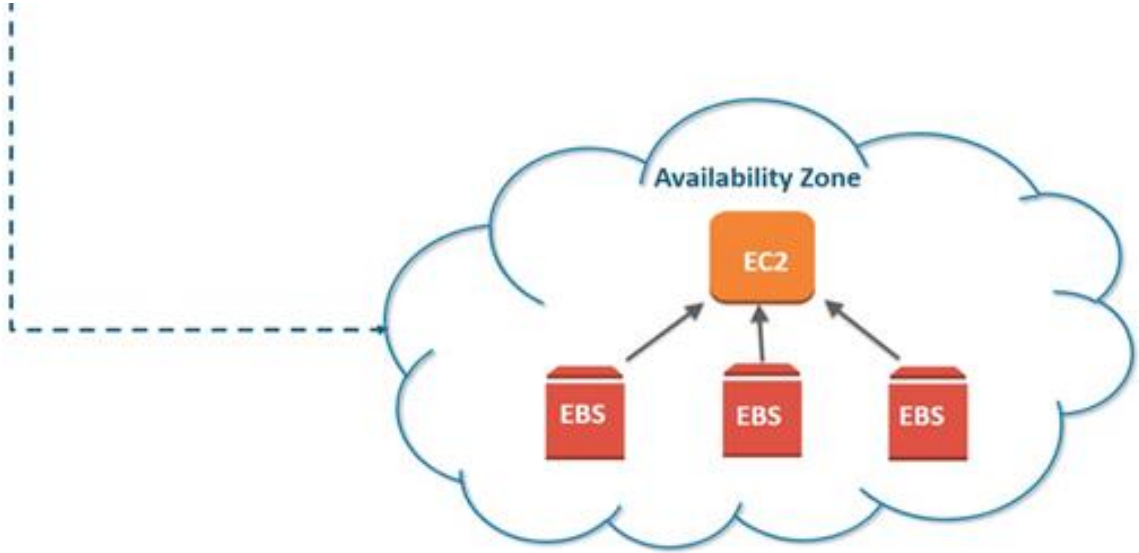


EBS

- A volume can be attached with only one instance at a time
- It can be detached and attached between the instances in the same Availability Zone

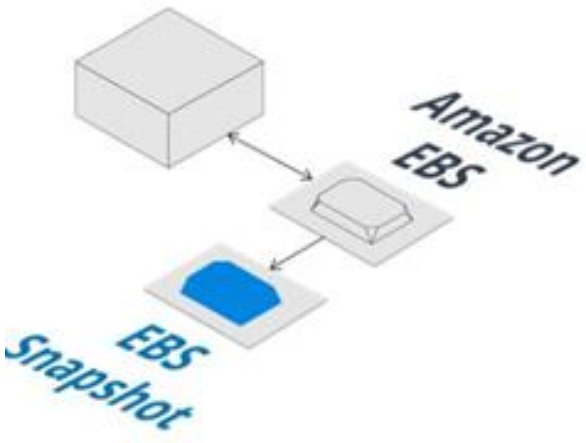
EBS

- Any number of EBS volume can be attached with EC2 instance



EBS Features

EBS is a persistent storage for EC2



Feature	Details
High performance file system	Mount EBS as drives and format as required
Flexible size	Volumes from 1GB to 1TB in size
Secure	Private to your instances
Available	Replicated within an Availability Zone
Backups	Volumes can be snapshotted for point in time restore
Monitoring	Detailed metrics captured via Cloud Watch

EBS Volume Types

K21Academy
Learn Cloud from Experts

The volumes types fall into these categories:

- 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.

- Previous generation — Hard disk drives that can be used for workloads with small datasets where data is accessed infrequently, and performance is not of primary importance. We recommend that you consider a current generation volume type instead.

SSD vs HDD

K21Academy
Learn Cloud from Experts

Solid state drives (SSD)

- The SSD-backed volumes provided by Amazon EBS fall into these categories:
- General Purpose SSD — Provides a balance of price and performance. We recommend these volumes for most workloads.
- Provisioned IOPS SSD — Provides high performance for mission-critical, low-latency, or high-throughput workloads.

Hard disk drive (HDD)

- The HDD-backed volumes provided by Amazon EBS fall into these categories:
- Throughput Optimized HDD — A low-cost HDD designed for frequently accessed, throughput-intensive workloads.
- Cold HDD — The lowest-cost HDD design for less frequently accessed workloads.

GP vs IO SSD

K21Academy
Learn Cloud from Experts

General Purpose SSD			Provisioned IOPS SSD	
Volume type	gp3	gp2	io2	io1
Durability	99.8% - 99.9% durability (0.1% - 0.2% annual failure rate)	99.8% - 99.9% durability (0.1% - 0.2% annual failure rate)	99.999% durability (0.001% annual failure rate)	99.8% - 99.9% durability (0.1% - 0.2% annual failure rate)
Use cases	•Low-latency interactive apps •Development and test environments		Workloads that require sub-millisecond latency, and sustained IOPS performance or more than 64,000 IOPS or 1,000 MiB/s of throughput	•Workloads that require sustained IOPS performance or more than 16,000 IOPS •I/O-intensive database workloads
Volume size	1 GiB - 16 TiB		4 GiB - 64 TiB	4 GiB - 16 TiB
Max IOPS per volume (16 KiB I/O)	16,000		256,000	64,000 †
Max throughput per volume	1,000 MiB/s	250 MiB/s *	4,000 MiB/s	1,000 MiB/s †
Amazon EBS Multi-attach	Not supported		Not supported	Supported
Boot volume	Supported			

Throughput Optimized vs Cold HDD

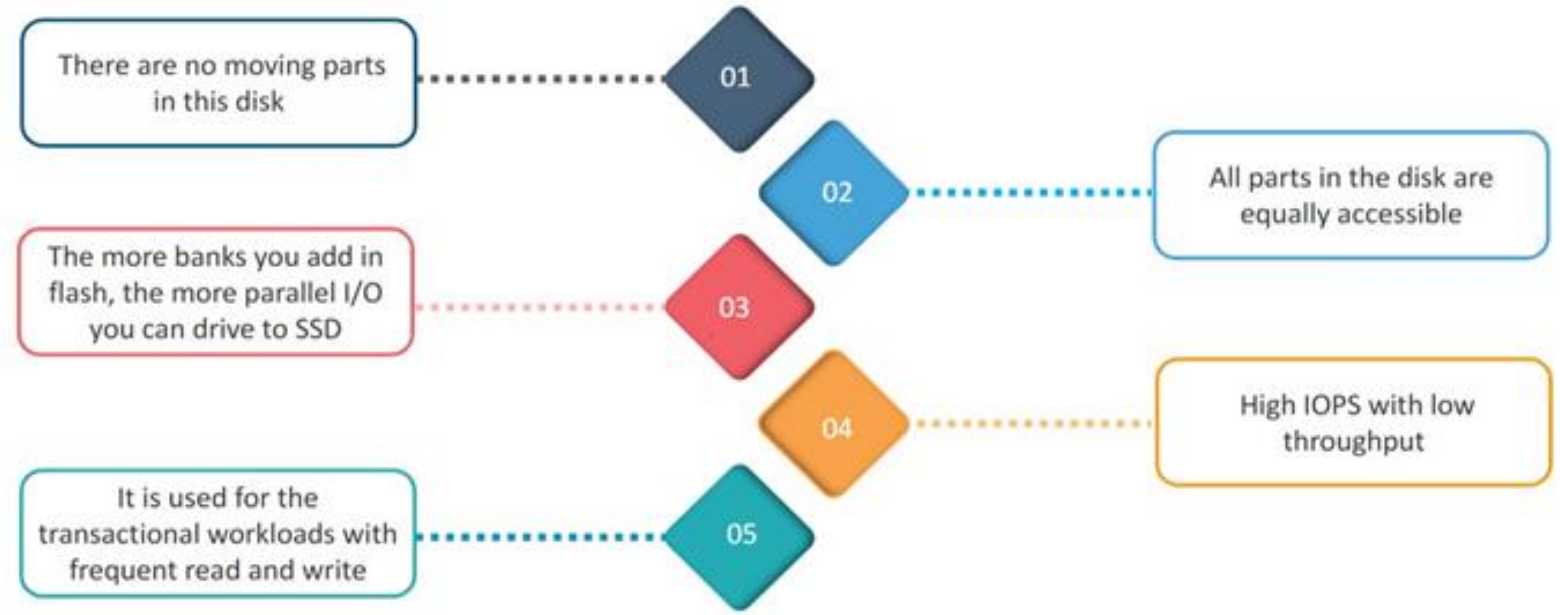
K21Academy
Learn Cloud from Experts

	Throughput Optimized HDD	Cold HDD
Volume type	st1	sc1
Durability	99.8% - 99.9% durability (0.1% - 0.2% annual failure rate)	99.8% - 99.9% durability (0.1% - 0.2% annual failure rate)
Use cases	•Big data •Data warehouses •Log processing	•Throughput-oriented storage for data that is infrequently accessed •Scenarios where the lowest storage cost is important
Volume size	125 GiB - 16 TiB	125 GiB - 16 TiB
Max IOPS per volume (1 MiB I/O)	500	250
Max throughput per volume	500 MiB/s	250 MiB/s
Amazon EBS Multi-attach	Not supported	Not supported
Boot volume	Not supported	Not supported

K21Academy
Learn Cloud from Experts

Solid State Drive And It's Types

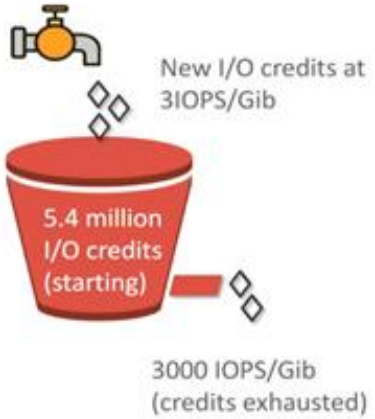
Solid State Drive (SSD)



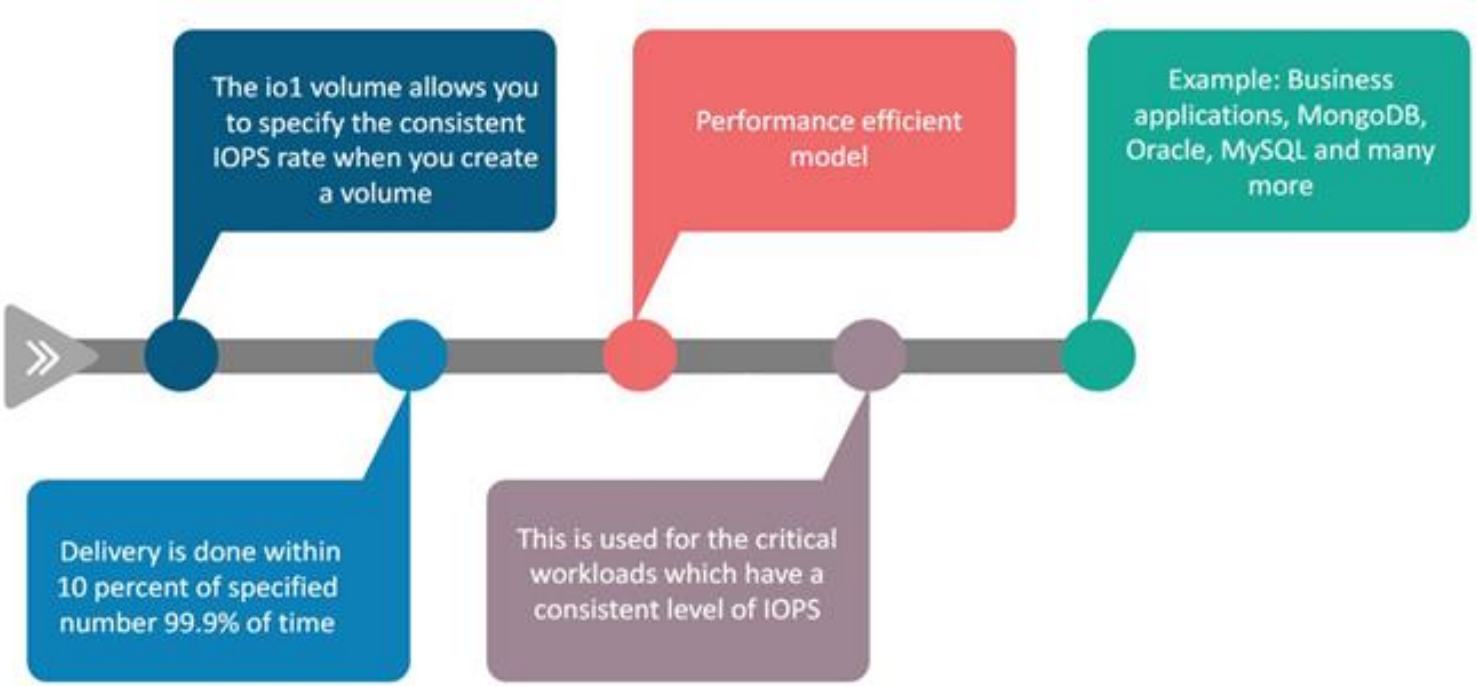
General Purpose SSD (gp2)

K21Academy
Learn Cloud from Experts

- gp2 consists of the token bucket, which is constantly accumulating 3IOPS/GiB/sec
- The bucket can consist a maximum of 5.4 million I/O credits
- It starts the execution only after the bucket is full
- Once the bucket is full, it executes 3000 IOPS /Gib which is called a BURST
- Cost efficient model data type



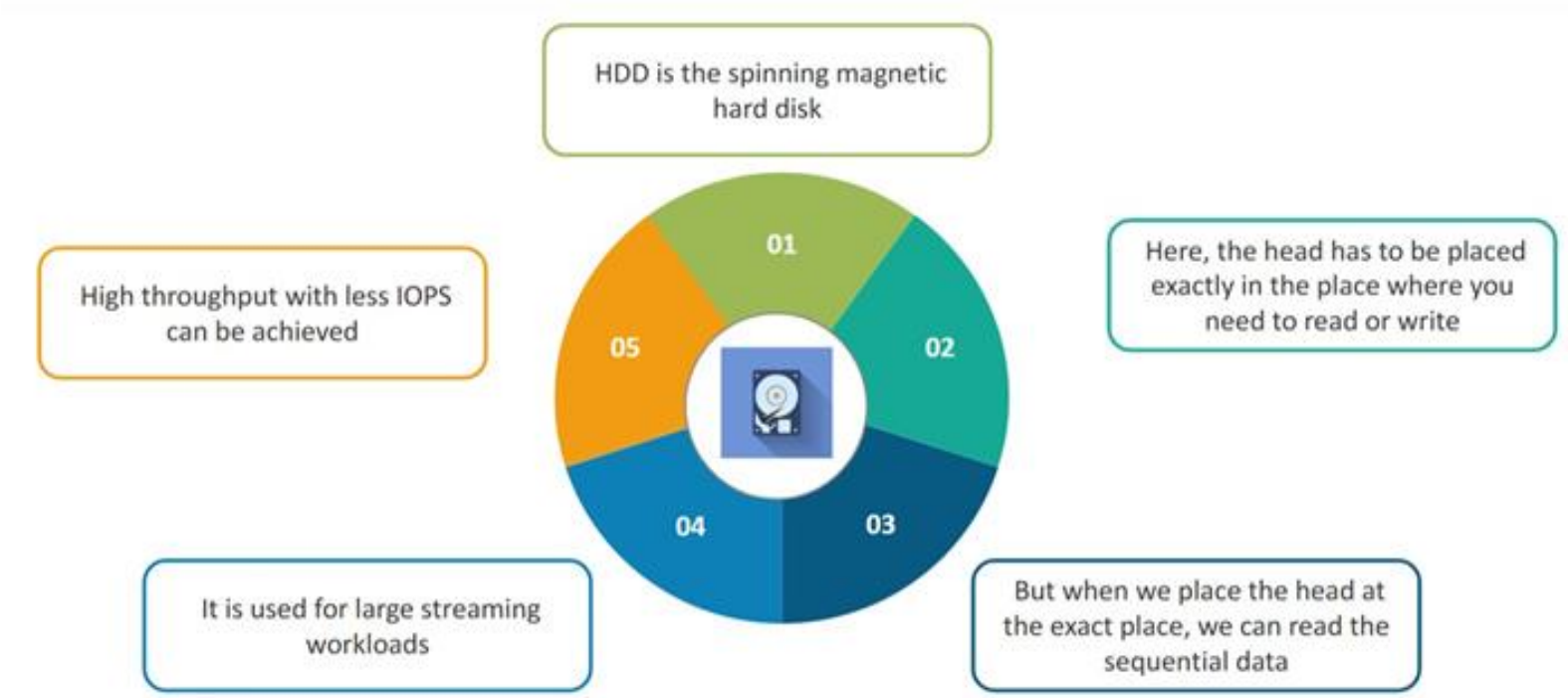
Provisioned IOPS (io1)



K21Academy
Learn Cloud from Experts

Hard Disk Drive And It's Types

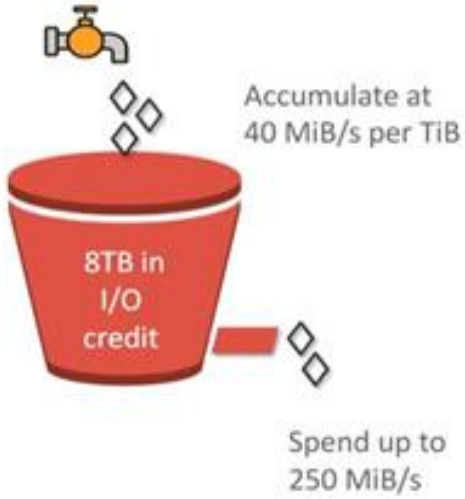
Hard Disk Drive (HDD)



Throughput Optimized HDD (st1)

K21Academy
Learn Cloud from Experts

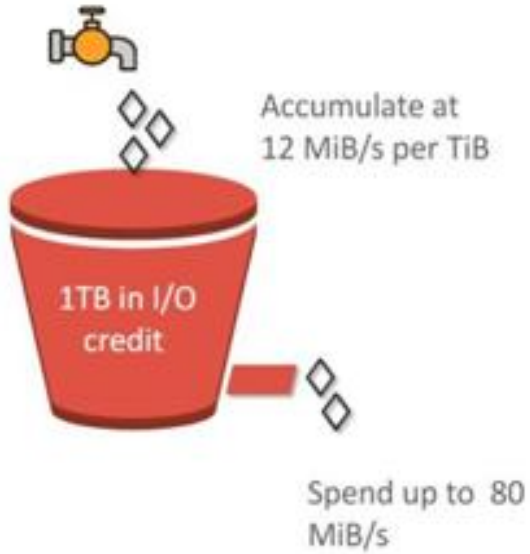
- st1 consist of the bucket, which scales with the volume
- Bigger the volume, bigger the bucket
- The throughput of the volume will be increased with the size



Cold HDD (sc1)

K21Academy
Learn Cloud from Experts

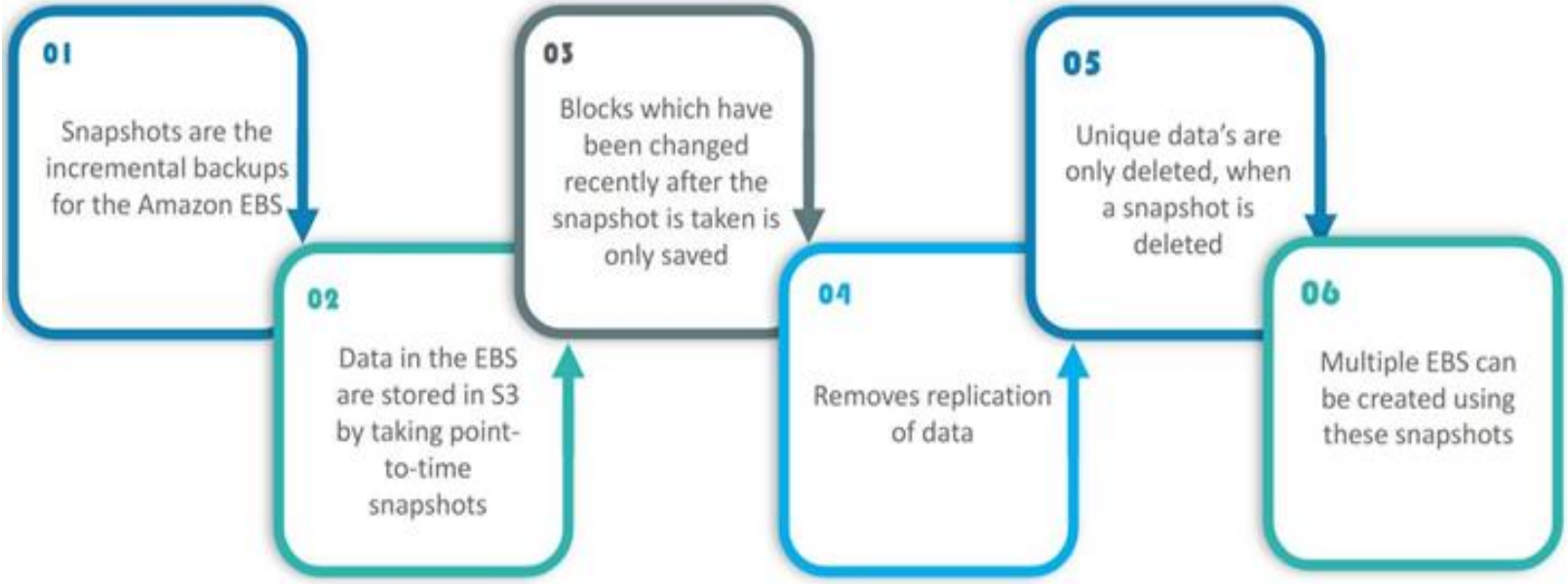
- sc1 is a low cost magnetic disk
- It is used for the ***infrequent access*** of the data
- Cost effective model compared to st1



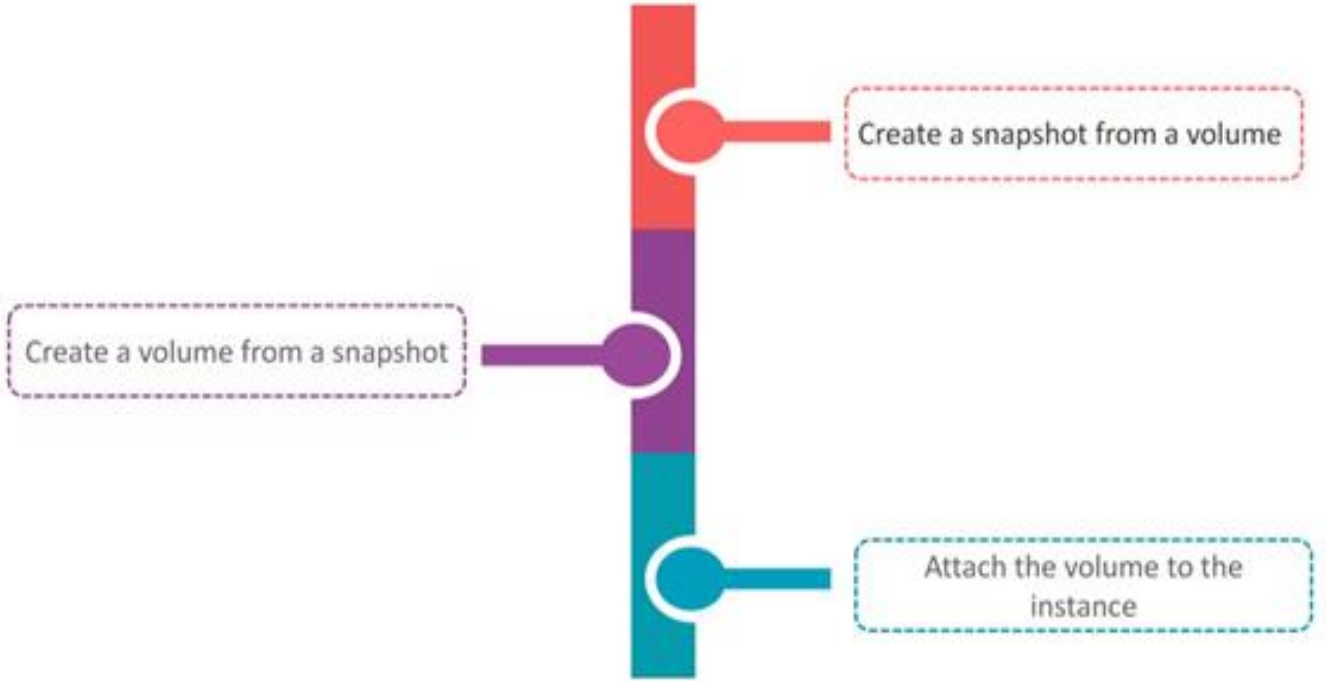
K21Academy
Learn Cloud from Experts

EBS Snapshots

Snapshots



Create A Snapshot

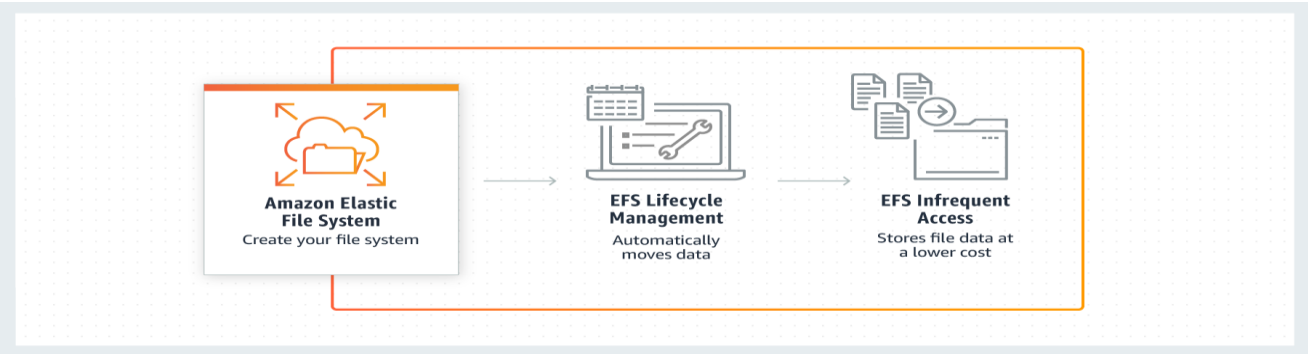


K21Academy
Learn Cloud from Experts

Amazon Elastic File System (EFS)

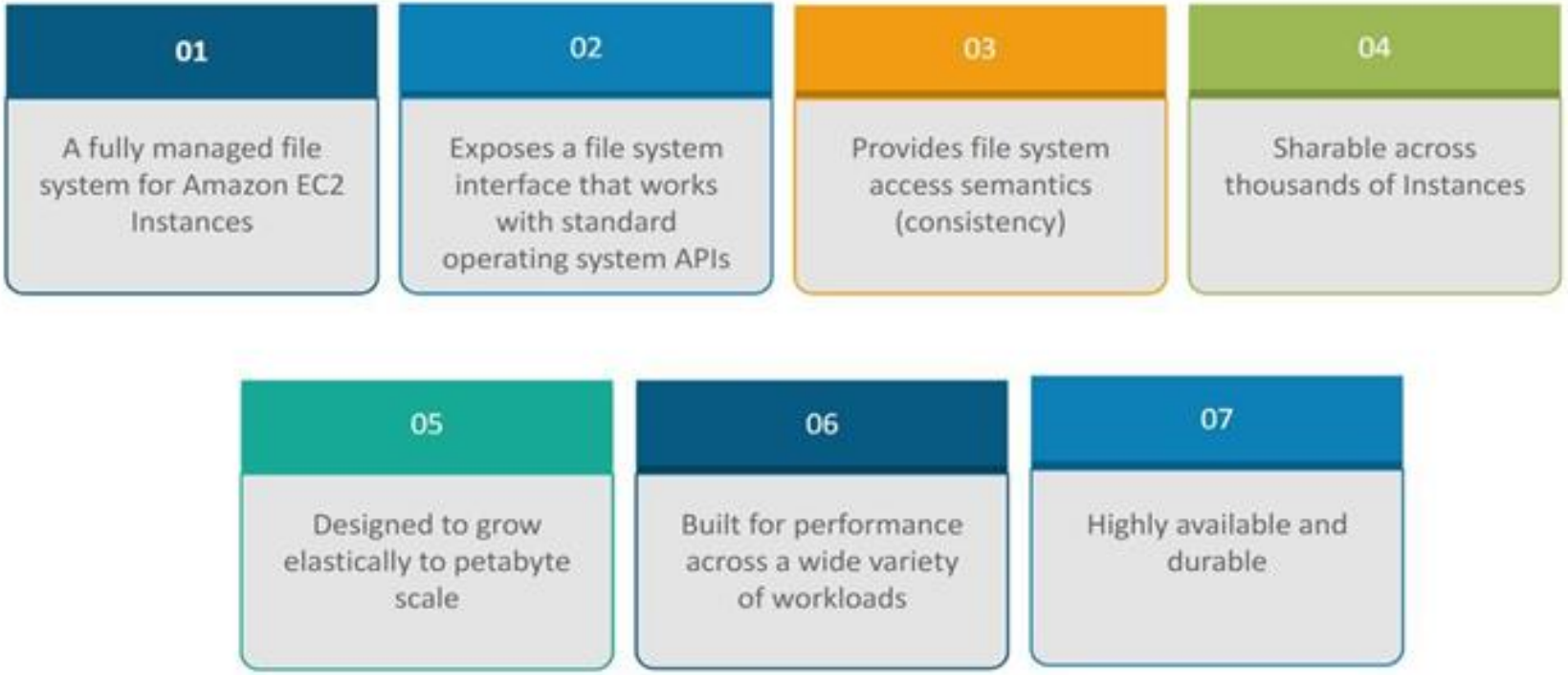
Amazon Elastic File System (EFS)

- An Amazon EFS Elastic file system is accessed by EC2 instances running inside one of your VPCs. Instances connect to a file system by using a network interface called a mount target. Each mount target has an IP address, which we assign automatically or you can specify.
- Amazon EFS file systems can automatically scale from gigabytes to petabytes of data without needing to provision storage, With Amazon EFS, there is no minimum fee or setup costs, and you pay only for what you use.



Disadvantage Of On-Premise File Storage

Amazon EFS?



Amazon EFS Features

K21Academy
Learn Cloud from Experts

Simple

Elastic

Scalable

Highly Durable and Available

- Fully managed
  - No hardware, network, file layer
  - Create a scalable file system in seconds
- Seamless integration with existing tools and apps
  - NFS v4.1—widespread, open
  - Standard file system access semantics
  - Works with standard OS file system APIs
- Simple pricing = simple forecasting

Amazon EFS Features

K21Academy
Learn Cloud from Experts

Simple

Elastic

Scalable

Highly Durable and Available

- File systems grow and shrink automatically as you add and remove files
- No need to provision storage capacity or performance
- You pay only for the storage space you use, with no minimum fee

Amazon EFS Features

K21Academy
Learn Cloud from Experts

- Simple
- Elastic
- **Scalable**
- Highly Durable and Available

- File systems can grow to petabyte scale
- Throughput and IOPS scale automatically as file systems grow
- Consistent low latencies regardless of file system size
- Support for thousands of concurrent NFS connections

Amazon EFS Features

K21Academy
Learn Cloud from Experts

- Simple
- Elastic
- Scalable
- Highly Durable and Available

- Designed to sustain AZ offline conditions
- Superior to traditional Network Attached Storage(NAS) availability models
- Appropriate for production/tier 0 applications

K21Academy
Learn Cloud from Experts

Difference Between EFS & EBS

EFS Vs EBS

K21Academy
Learn Cloud from Experts

Features	EBS	EFS
Storage Size	Maximum storage size of 16 TB	No limitation
File Size	No Limitation	Single file can have maximum 52TiB
Performance	Without stopping instance volume can be scaled manually	It supports up to 7000 file system operations per second
Data Store	Data is stored in same Availability Zone and can be replicated within the same AZs	Data is stored in region and replicated within the region
Date Access	Can Be accesses from only one EC2	Can be accessed from 1-10 EC2 instance from multi AZs parallel
Availability Zone Failure	Without point-in-time backup it will fail	Can survive
Permissions	Supports ex3 and ext4 and other various file systems	EFS can be used as NFS for on-premise servers too using AWS Direct Connect
Encryption	KMS Managed Keys	SKMS Managed Keys and AES 256

K21Academy
Learn Cloud from Experts

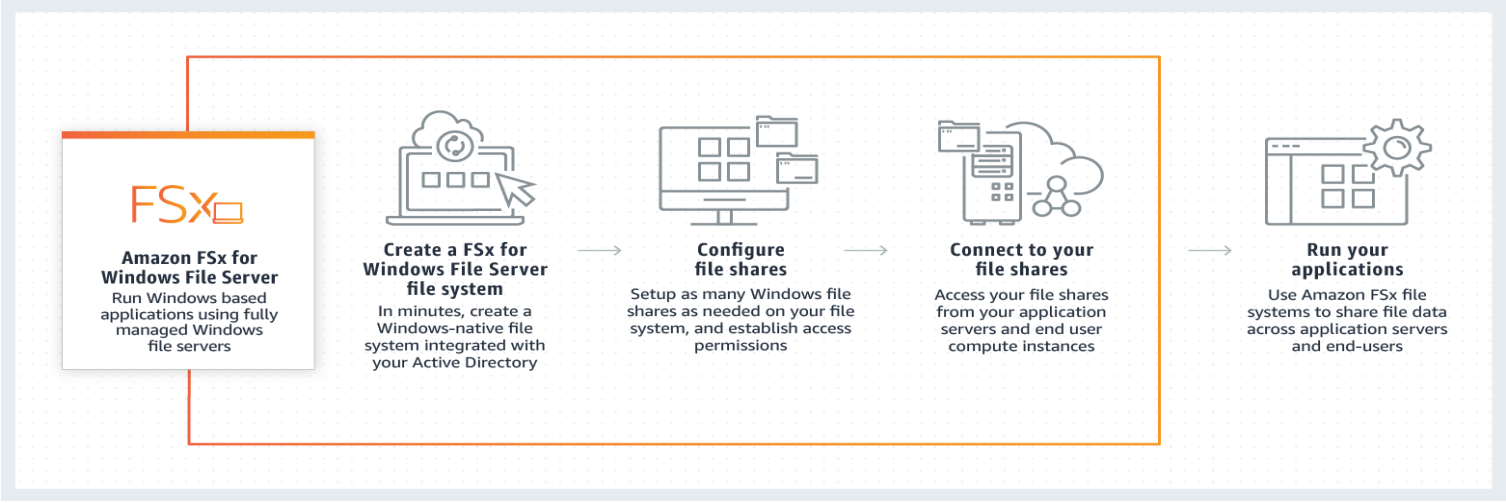
Amazon FSx File System

Amazon FSx

Amazon FSx provides you with two file systems to choose from

1) Amazon FSx for Windows File Server for enterprise workloads

2) Amazon FSx for Lustre for high-performance workloads.



Amazon FSx for Windows File Server for Enterprise Workloads

K21Academy
Learn Cloud from Experts

- Amazon FSx for Windows File Server provides simple, fully managed, highly reliable file storage that's accessible over the industry-standard Server Message Block (SMB) protocol.
- Built on Windows Server, providing full SMB support and a wide range of administrative features like user quotas, data deduplication, and end-user file restore.
- Provides file storage that is accessible from Windows, Linux, and MacOS compute instances and devices running on AWS or on-premises.

Amazon FSx for Windows File Server for Enterprise Workloads

K21Academy
Learn Cloud from Experts

- Integrates with Microsoft Active Directory (AD) to support Windows-based environments and enterprises.
- Offers single-AZ and multi-AZ deployment options, SSD and HDD storage options, and provides fully managed backups.
- All file system data is automatically encrypted at rest and in transit.
- You can dynamically scale your file system to fit your storage and throughput needs.

Amazon FSx for Lustre for High-Performance Workloads.

K21Academy
Learn Cloud from Experts

- Amazon FSx for Lustre makes it easy and cost-effective to launch and run the world's most popular high-performance file system. Use it for workloads where speed matters, such as machine learning, high-performance computing (HPC), video processing, and financial modelling.
- Allows your workloads to process data with consistent sub-millisecond latencies, up to hundreds of gigabytes per second of throughput, and up to millions of IOPS.
- POSIX-compliant, so you can use your current Linux-based applications without having to make any changes, providing a native file system interface that works as any file system does with your Linux operating system.

Amazon FSx for Lustre for High-Performance Workloads.

- Supports multiple deployment options for short-term and long-term data processing.
- Seamlessly integrated with Amazon S3 (connect your S3 data sets to your FSx for Lustre file system, run your analyses, write results back to S3, and delete your file system), Amazon Sage Maker, Amazon Elastic Kubernetes Service (EKS), and AWS Parallel Cluster Accessible from on-premises over Direct Connect and VPN connections.

K21Academy
Learn Cloud from Experts

Amazon FSx Lustre vs Amazon EFS

Amazon FSx Lustre vs Amazon EFS

K21Academy
Learn Cloud from Experts

Use cases of Amazon FSx Lustre

- Use Amazon FSx for Lustre for workloads where speed matters, such as Machine learning, high performance computing (HPC), video processing, financial modeling, genome sequencing, and electronic design automation (EDA).

Use Cases of Amazon EFS

- Amazon EFS is designed to provide performance for a broad spectrum of workloads and applications, including Big Data and analytics, media processing workflows, content management, web serving, and home directories.

K21Academy
Learn Cloud from Experts

Amazon EC2 Cost Optimization

Cost Optimization

- Instance are provided here On demand
- Pay only for EC2 instance you use
- There will not be any upfront charges
- Prices will be decided by AWS and it will be displayed on the AWS Website
- It is charged in hours or seconds for the services you are using
- It frees you from the planning, purchasing and maintaining hardware

Cost Optimization

- Capacity reservation for EC2 instance is done priory
- The reserved Instance is for customers with predictable workloads
- Payment option available in reserved instance: all upfront, partial up front or no upfront
- It is 75% cheaper than On-Demand Instance
- Price of the reserved instance varies with the Availability Zone

Cost Optimization

- In spot instance, the spot price that is in effect for the time period your instances are running is paid
- The spot instances offer spare Compute capacity that optimizes your cost and scales your application throughput up to 10x in the same budget
- This is suitable for the workloads which are not critical and are tolerant of interruption

Lab Exercises

K21Academy
Learn Cloud from Experts

Creating and Managing Volumes and Snapshots

- Creating & managing volumes
- Creating & managing snapshots

K21Academy
Learn Cloud from Experts

amazon web services

Table of Contents

Lab Exercises

K21Academy
Learn Cloud from Experts

Attach & Mount EBS Volume to Linux EC2 Instance

- Launch Linux EC2 Instance
- Create Empty EBS Volume
- Attach EBS Volume to Instance
- Format & Mount EBS

K21Academy
Learn Cloud from Experts

amazon web services

Contents

Lab Exercises

K21Academy
Learn Cloud from Experts

Attach & Mount EBS Volume to EC2 Windows Instance

- Launch Windows EC2 Instance
- Create Empty EBS Volume
- Attach EBS Volume to Instance
- Format & Mount EBS

Contents

Lab Exercises

K21Academy
Learn Cloud from Experts

Creating Windows File Systems Using Amazon FSx

- Creating FSx

K21Academy
Learn Cloud from Experts

amazon web services

Contents

Lab Exercises

Mount Elastic File System (EFS) on EC2 Instances

- Launching EC2
- Creating EFS

Contents

Lab Exercises

K21Academy
Learn Cloud from Experts

SSH into EC2 Linux Instance via Username & Password

K21Academy
Learn Cloud from Experts

amazon web services

- Launching Linux EC2 Instance & connect via putty
- Create use & password for authentication
- Enable password authentication in SSHD
- Login via username and password

Contents

Quiz

K21Academy
Learn Cloud from Experts

To help you manage your Amazon EC2 instances, you can assign your own metadata in the form of ____.
A. Tags
B. Wildcards
C. Certificates
D. Notes

Correct Answer: A

Explanation: Tagging is a key part of managing an environment. Even in a lab, it is easy to lose track of the purpose of resources, and tricky determine why it was created and if it is still needed. This can rapidly translate into lost time and lost money.

Quiz

K21Academy
Learn Cloud from Experts

When creating a new security group, all inbound traffic is allowed by default.

A. True
B. False

Correct Answer: B

Explanation : There are slight differences between a normal 'new' Security Group and a 'default' security group in the default VPC. For a 'new' security group nothing is allowed in by default.

Find Us


https://www.facebook.com/K21Academy


http://twitter.com/k21Academy


https://www.linkedin.com/company/k21academy

https://www.youtube.com/k21academy


https://www.instagram.com/k21academy/