What Is EC2?

Amazon EC2 changes the economics of computing by allowing you to pay only for capacity that you actually use. Amazon EC2 provides developers the tools to build failure resilient applications and isolate themselves from common failure scenarios.

EC2 Options

- **On Demand** - allows you to pay a fixed rate by the hour (or by the second) with no commitment.

- **Reserved** - provides you with a capacity reservation, and offer a significant discount on the hourly charge for an instance.1 Year or 3 Year Terms

- **Spot** - enables you to bid whatever price you want for instance capacity, providing for even greater savings if your applications have flexible start and end times.

- **Dedicated Hosts** - Physical EC2 server dedicated for your use. Dedicated Hosts can help you reduce costs by allowing you to use your existing server-bound software licenses.

On Demand

A CLOUD GURU

- Perfect for users that want the low cost and flexibility of Amazon EC2 without any up-front payment or long-term commitment
- Applications with short term, spiky, or unpredictable workloads that cannot be interrupted
- Applications being developed or tested on Amazon EC2 for the first time

Reserved Instances

- Applications with steady state or predictable usage
- Applications that require reserved capacity
- Users can make up-front payments to reduce their total computing costs even further
  - Standard RIs (Up to 75% off on-demand)
  - Convertible RIs (Up to 54% off on-demand) feature the capability to change the attributes of the RI as long as the exchange results in the creation of Reserved Instances of equal or greater value.
  - Scheduled RIs are available to launch within the time window you reserve. This option allows you to match your capacity reservation to a predictable recurring schedule that only requires a fraction of a day, a week, or a month.

RESERVED

Spot Instances

- Applications that have flexible start and end times
- Applications that are only feasible at very low compute prices
- Users with an urgent need for large amounts of additional computing capacity

27. Let's Get Our Hands Dirty! Launch An EC2 Instance - Part 1

Dedicated Hosts

- Useful for regulatory requirements that may not support multi-tenant virtualization.
- Great for licensing which does not support multi-tenancy or cloud deployments.
- Can be purchased On-Demand (hourly.)
- Can be purchased as a Reservation for up to 70% off the On-Demand price.

27. Let's Get Our Hands Dirty! Launch An EC2 Instance - Part 1

EC2 Instance Types

Family	Speciality	Use case
F1	Field Programmable Gate Array	Genomics research, financial analytics, real-time video processing, big data etc
I3	High Speed Storage	NoSQL DBs, Data Warehousing etc
G3	Graphics Intensive	Video Encoding/ 3D Application Streaming
H1	High Disk Throughput	MapReduce-based workloads, distributed file systems such as HDFS and MapR-FS
T2	Lowest Cost, General Purpose	Web Servers/Small DBs
D2	Dense Storage	Fileservers/Data Warehousing/Hadoop
R4	Memory Optimized	Memory Intensive Apps/DBs
M5	General Purpose	Application Servers
C5	Compute Optimized	CPU Intensive Apps/DBs
P3	Graphics/General Purpose GPU	Machine Learning, Bit Coin Mining etc
X1	Memory Optimized	SAP HANA/Apache Spark etc

EC2 Instance Types

- How I remember them now;
  - **F** for FPGA
  - **I** for IOPS
  - **G** - Graphics
  - **H** - High Disk Throughput
  - **T** cheap general purpose (think T2 Micro)
  - **D** for Density
  - **R** for RAM
  - **M** - main choice for general purpose apps
  - **C** for Compute
  - **P** - Graphics (think Pics)
  - **X** - Extreme Memory

What is EBS?

Amazon EBS allows you to create storage volumes and attach them to Amazon EC2 instances. Once attached, you can create a file system on top of these volumes, run a database, or use them in any other way you would use a block device. Amazon EBS volumes are placed in a specific Availability Zone, where they are automatically replicated to protect you from the failure of a single component.

EBS Volume Types

- General Purpose SSD (GP2)
  - General purpose, balances both price and performance.
  - Ratio of 3 IOPS per GB with up to 10,000 IOPS and the ability to burst up to 3000 IOPS for extended periods of time for volumes at 3334 GiB and above.
- Provisioned IOPS SSD (IO1)
  - Designed for I/O intensive applications such as large relational or NoSQL databases.
  - Use if you need more than 10,000 IOPS.
  - Can provision up to 20,000 IOPS per volume.

A CLOUD GURU

EBS Volume Types

- Throughput Optimized HDD (ST1)
  - Big data
  - Data warehouses
  - Log processing
  - Cannot be a boot volume
- Cold HDD (SC1)
  - Lowest Cost Storage for infrequently accessed workloads
  - File Server
  - Cannot be a boot volume.

EBS Volume Types

- Magnetic (Standard)
  - Lowest cost per gigabyte of all EBS volume types that is bootable. Magnetic volumes are ideal for workloads where data is accessed infrequently, and applications where the lowest storage cost is important.

EC2 Exam Tips

- **On Demand** - allows you to pay a fixed rate by the hour (or by the second) with no commitment.

- **Reserved** - provides you with a capacity reservation, and offer a significant discount on the hourly charge for an instance.1 Year or 3 Year Terms

- **Spot** - enables you to bid whatever price you want for instance capacity, providing for even greater savings if your applications have flexible start and end times.

- **Dedicated Hosts** - Physical EC2 server dedicated for your use. Dedicated Hosts can help you reduce costs by allowing you to use your existing server-bound software licenses.

EC2 Exam Tips

A CLOUD GURU

If a Spot instance is terminated by Amazon EC2, you will not be charged for a partial hour of usage. However, if you terminate the instance yourself, you will be charged for the complete hour in which the instance ran.

Let's Get Our Hands Dirty! Launch An EC2 Instance - Part 2
Section 5, Lecture 28

Go to Dashboard

A CLOUD GURU

Lab Summary

- Termination Protection is turned off by default, you must turn it on.
- On an EBS-backed instance, the default action is for the root EBS volume to be deleted when the instance is terminated.
- EBS Root Volumes of your DEFAULT AMI's cannot be encrypted. You can also use a third party tool (such as bit locker etc) to encrypt the root volume, or this can be done when creating AMI's (lab to follow) in the AWS console or using the API.
- Additional volumes can be encrypted.

Security Group Lab

- All Inbound Traffic is Blocked By Default
- All Outbound Traffic is Allowed
- Changes to Security Groups take effect immediately
- You can have any number of EC2 instances within a security group.
- You can have multiple security groups attached to EC2 Instances
- Security Groups are **STATEFUL.**
  - If you create an inbound rule allowing traffic in, that traffic is automatically allowed back out again.
- You cannot block specific IP addresses using Security Groups, instead use Network Access Control Lists.
- You can specify allow rules, but not deny rules.

Volumes & Snapshots

- Volumes exist on EBS:
  - Virtual Hard Disk
- Snapshots exist on S3.
- Snapshots are point in time copies of Volumes.
- Snapshots are incremental — this means that only the blocks that have changed since your last snapshot are moved to S3.
- If this is your first snapshot, it may take some time to create.

Snapshots of Root Device Volumes

- To create a snapshot for Amazon EBS volumes that serve as root devices, you should stop the instance before taking the snapshot.
- However you can take a snap while the instance is running.
- **You can create AMI's from EBS-backed Instances and Snapshots.**
- You can change EBS volume sizes on the fly, including changing the size and storage type.
- Volumes will ALWAYS be in the same availability zone as the EC2 instance.
- To move an EC2 volume from one AZ/Region to another, take a snap or an image of it, then copy it to the new AZ/Region

Volumes vs Snapshots - Security

- Snapshots of encrypted volumes are encrypted automatically.
- Volumes restored from encrypted snapshots are encrypted automatically.
- You can share snapshots, but only if they are unencrypted.
  - These snapshots can be shared with other AWS accounts or made public.

EBS vs Instance Store

All AMIs are categorized as either backed by Amazon EBS or backed by instance store.

For EBS Volumes: The root device for an instance launched from the AMI is an Amazon EBS volume created from an Amazon EBS snapshot.

For Instance Store Volumes: The root device for an instance launched from the AMI is an instance store volume created from a template stored in Amazon S3.

# EBS vs Instance Store - Exam Tips

- Instance Store Volumes are sometimes called Ephemeral Storage.
- Instance store volumes cannot be stopped. If the underlying host fails, you will lose your data.
- EBS backed instances can be stopped. You will not lose the data on this instance if it is stopped.
- You can reboot both, you will not lose your data.
- By default, both ROOT volumes will be deleted on termination, however with EBS volumes, you can tell AWS to keep the root device volume.

Types Of Load Balancers

- Application Load Balancer
- Network Load Balancer
- Classic Load Balancer
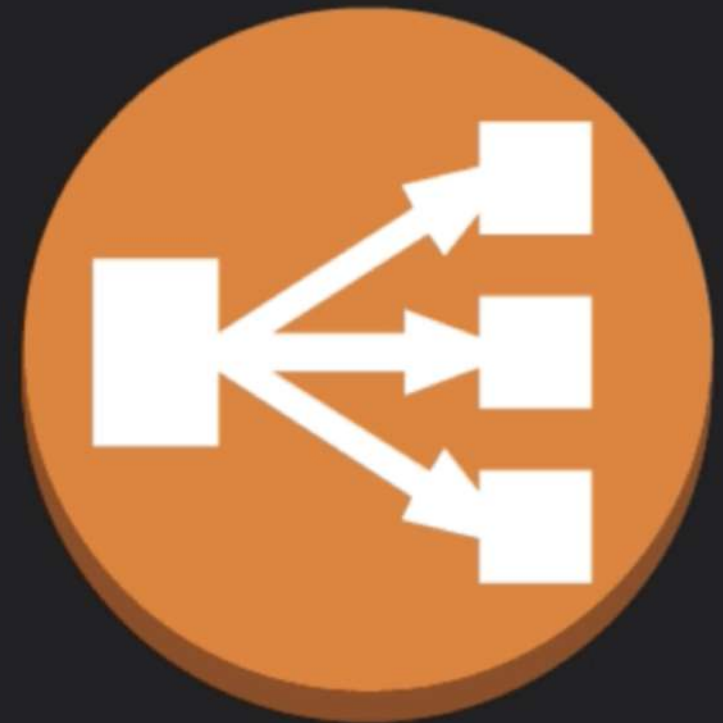
Application Load Balancers

Application Load Balancers are best suited for load balancing of HTTP and HTTPS traffic. They operate at Layer 7 and are application-aware. They are intelligent, and you can create advanced request routing, sending specified requests to specific web servers.

Network Load Balancer

Network Load Balancers are best suited for load balancing of TCP traffic where extreme performance is required. Operating at the connection level (Layer 4), Network Load Balancer are capable of handling millions of requests per second,while maintaining ultra-low latencies.

Use for extreme performance!

Classic Load Balancers

Classic Load Balancers are the legacy Elastic Load Balancers. You can load balance HTTP/HTTPS applications and use Layer 7-specific features, such as X-Forwarded and sticky sessions. You can also use strict Layer 4 load balancing for applications that rely purely on the TCP protocol.
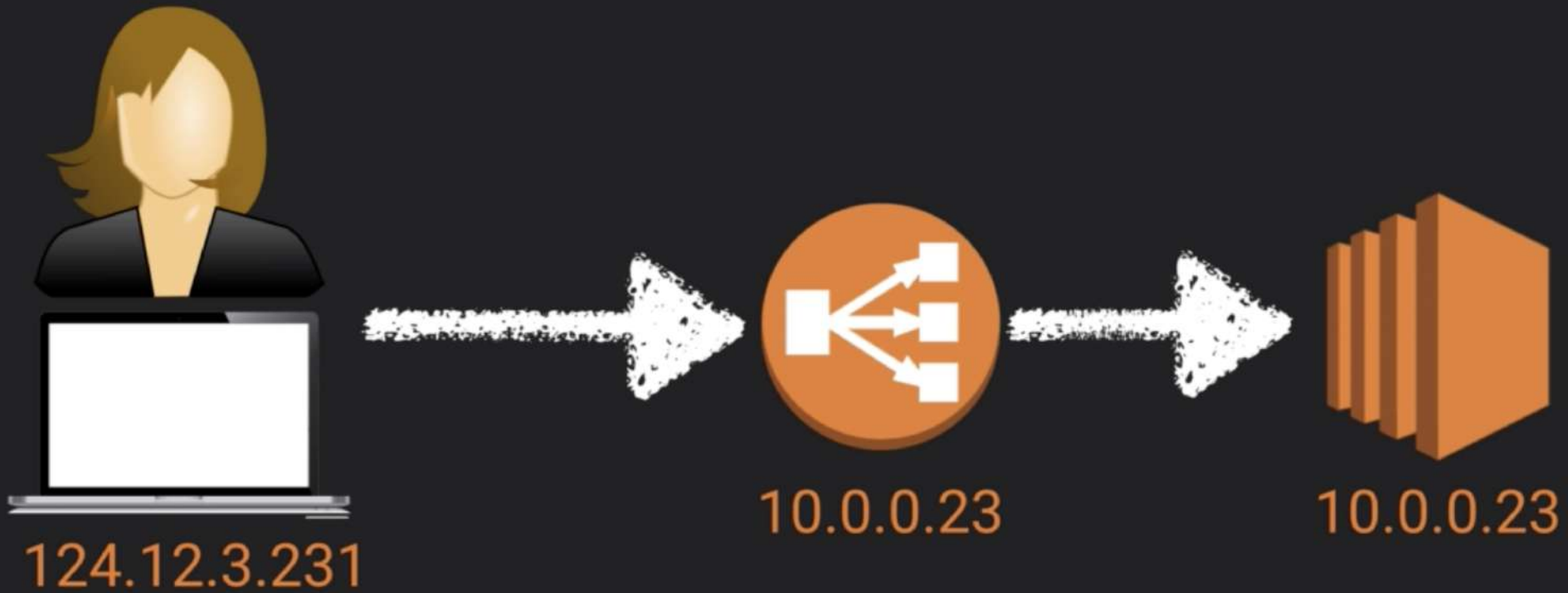
Load Balancer Errors

Classic Load Balancers - if your application stops responding, the ELB (Classic Load Balancer) responds with a 504 error.

This means that the application is having issues. This could be either at the Web Server layer or at the Database Layer.

Identify where the application is failing, and scale it up or out where possible.

ERROR
504

Gateway timeout.

ELB Exam Tips

- 3 Types of Load Balancers;
  - Application Load Balancers
  - Network Load Balancers
  - Classic Load Balancers
- 504 Error means the gateway has timed out. This means that the application not responding within the idle timeout period.
  - Trouble shoot the application. Is it the Web Server or Database Server?
- If you need the IPv4 address of your end user, look for the X-Forwarded-For header.

ELB Exam Tips

- 3 Types of Load Balancers;
  - Application Load Balancers
  - Network Load Balancers
  - Classic Load Balancers
- 504 Error means the gateway has timed out. This means that the application not responding within the idle timeout period.
  - Trouble shoot the application. Is it the Web Server or Database Server?
- If you need the IPv4 address of your end user, look for the X-Forwarded-For header.

Elastic Load Balancers

- Instances monitored by ELB are reported as ;
  InService , or OutofService
- Health Checks check the instance health by talking to it
- Have their own DNS name. You are never given an IP address.
- Read the ELB FAQ for Classic Load Balancers
- Want to deep dive on application load balancers? Check out our
  deep dive course!

Exam Tips

A CLOUD GURU

- Standard Monitoring = 5 Minutes
- Detailed Monitoring = 1 Minute

What can I do with Cloudwatch?

A CLOUD GURU

- Dashboards - Creates awesome dashboards to see what is happening with your AWS environment.
- Alarms - Allows you to set Alarms that notify you when particular thresholds are hit.
- Events - CloudWatch Events helps you to respond to state changes in your AWS resources.
- Logs - CloudWatch Logs helps you to aggregate, monitor, and store logs.

What is a Placement Group?

Two Types of Placement Groups;

- Clustered Placement Group

- Spread Placement Group

Clustered Placement Group

A cluster placement group is a grouping of instances within a single Availability Zone. Placement groups are recommended for applications that need low network latency, high network throughput, or both.

Only certain instances can be launched in to a Clustered Placement Group.

Spread Placement Group

A spread placement group is a group of instances that are each placed on distinct underlying hardware.
Spread placement groups are recommended for applications that have a small number of critical instances that should be kept separate from each other.

EC2 Placement Groups

- A clustered placement group can't span multiple Availability Zones.
- A spread placement group can.
- The name you specify for a placement group must be unique within your AWS account.
- Only certain types of instances can be launched in a placement group (Compute Optimized, GPU, Memory Optimized, Storage Optimized)
- AWS recommend homogenous instances within placement groups.
- You can't merge placement groups.
- You can't move an existing instance into a placement group. You can create an AMI from your existing instance, and then launch a new instance from the AMI into a placement group.

What is EFS

Amazon Elastic File System (Amazon EFS) is a file storage service for Amazon Elastic Compute Cloud (Amazon EC2) instances. Amazon EFS is easy to use and provides a simple interface that allows you to create and configure file systems quickly and easily. With Amazon EFS, storage capacity is elastic, growing and shrinking automatically as you add and remove files, so your applications have the storage they need, when they need it.

EFS Features

- Supports the Network File System version 4 (NFSv4) protocol
- You only pay for the storage you use (no pre-provisioning required)
- Can scale up to the petabytes
- Can support thousands of concurrent NFS connections
- Data is stored across multiple AZ's within a region
- Read After Write Consistency