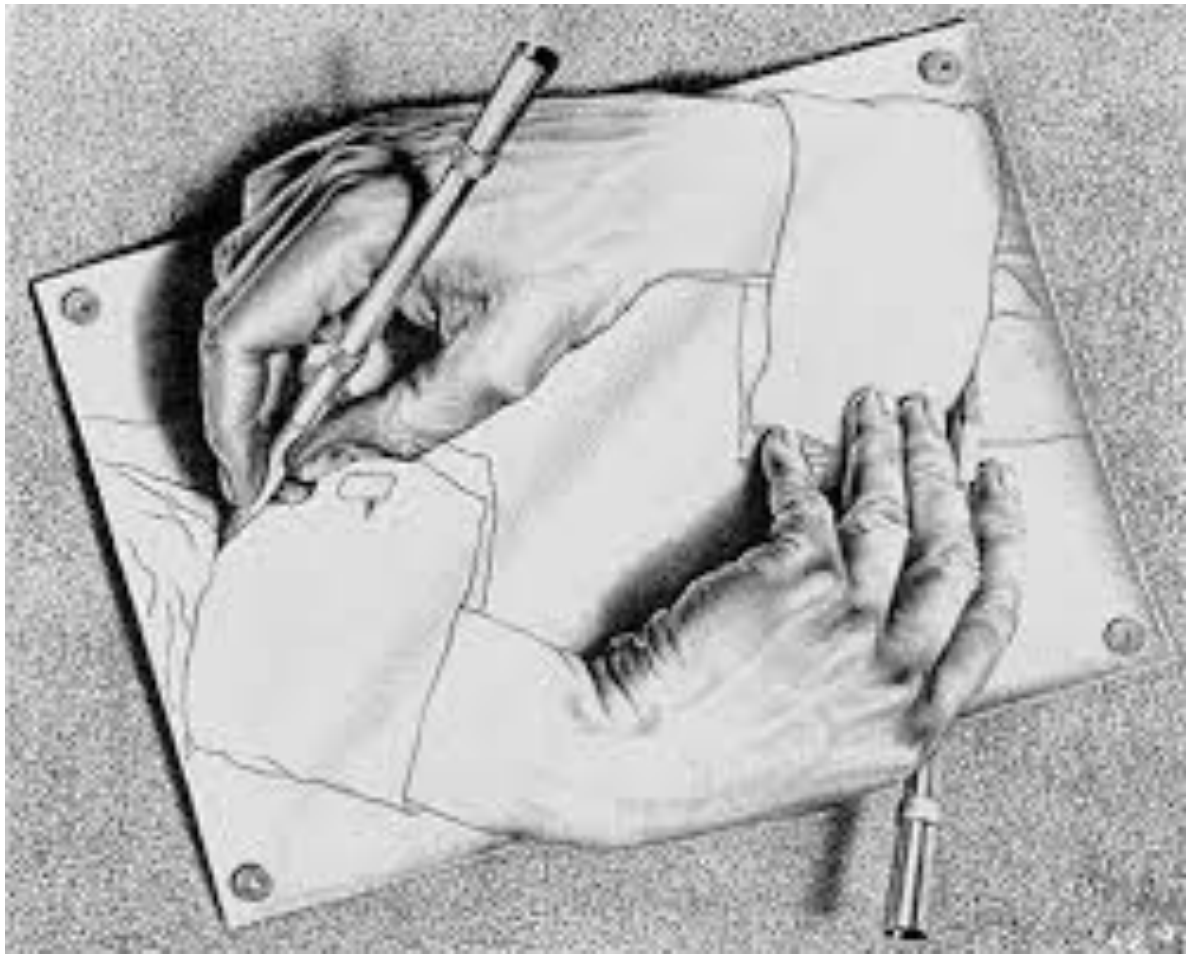
1 Introduction to EC2

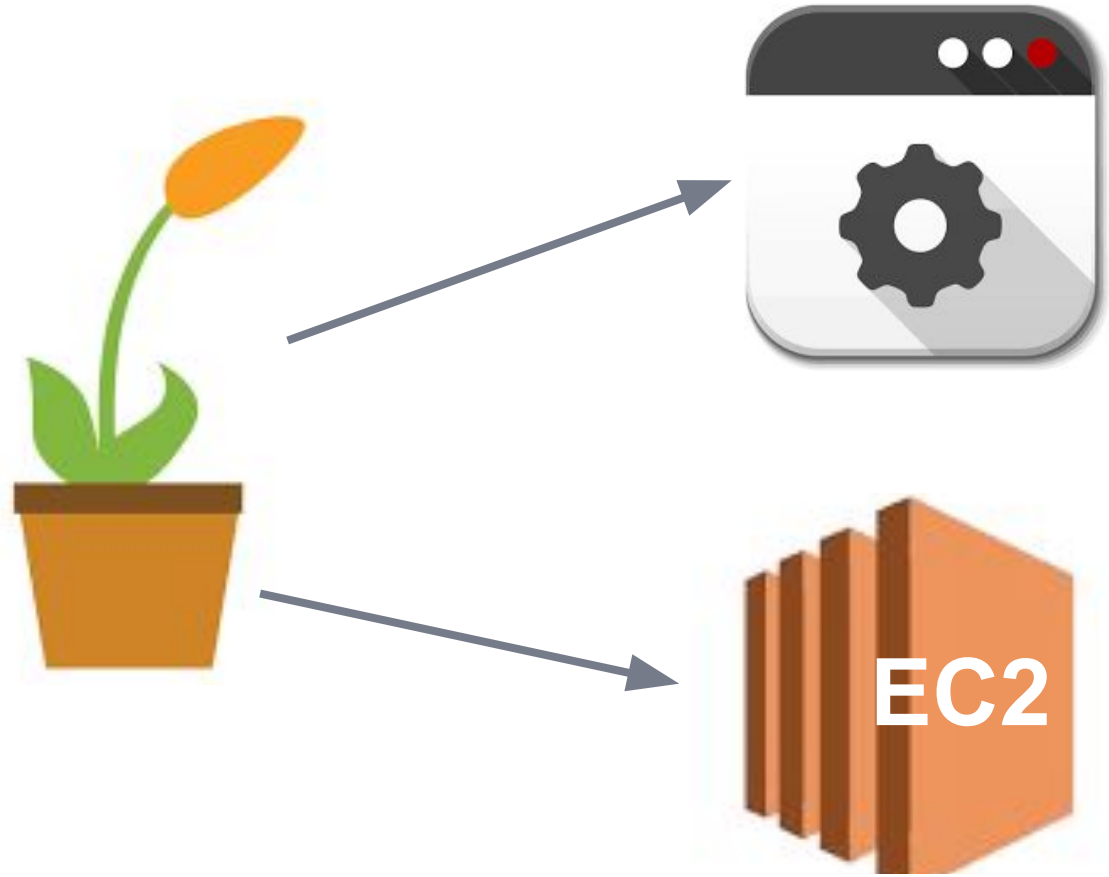
Today's Takeaways

- Introduction to EC2
- EC2 Instance Types
- Creating an EC2 instance



Introduction to EC2

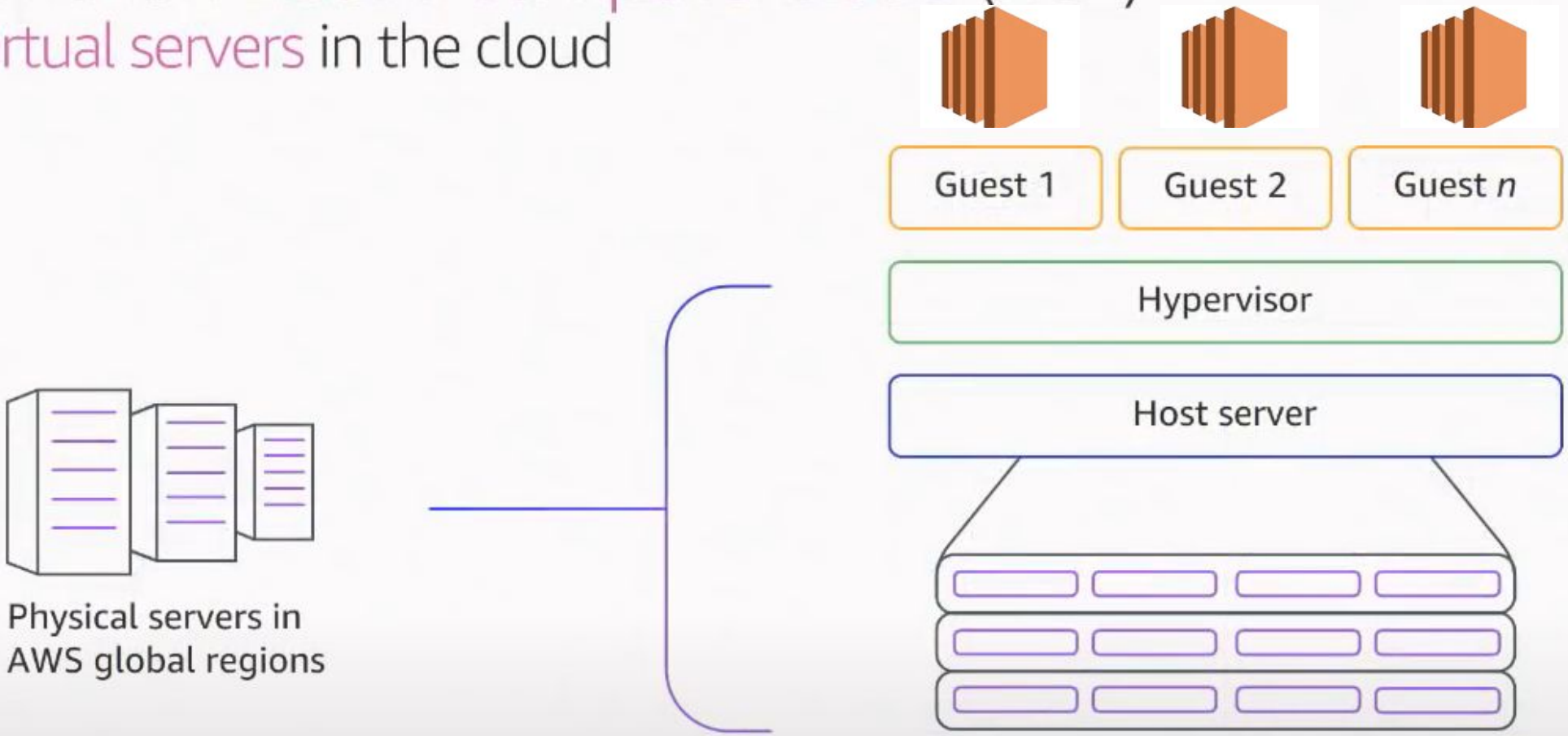
What is EC2?



- EC2 stands for Elastic Compute Cloud in AWS.
- EC2 is a service that allows you to run application programs in the computing environment.
- EC2 is a web service that provides secure, resizable compute capacity in the cloud. It is designed to make web-scale cloud computing easier for developers.

Amazon Elastic Compute Cloud (EC2)

Virtual servers in the cloud

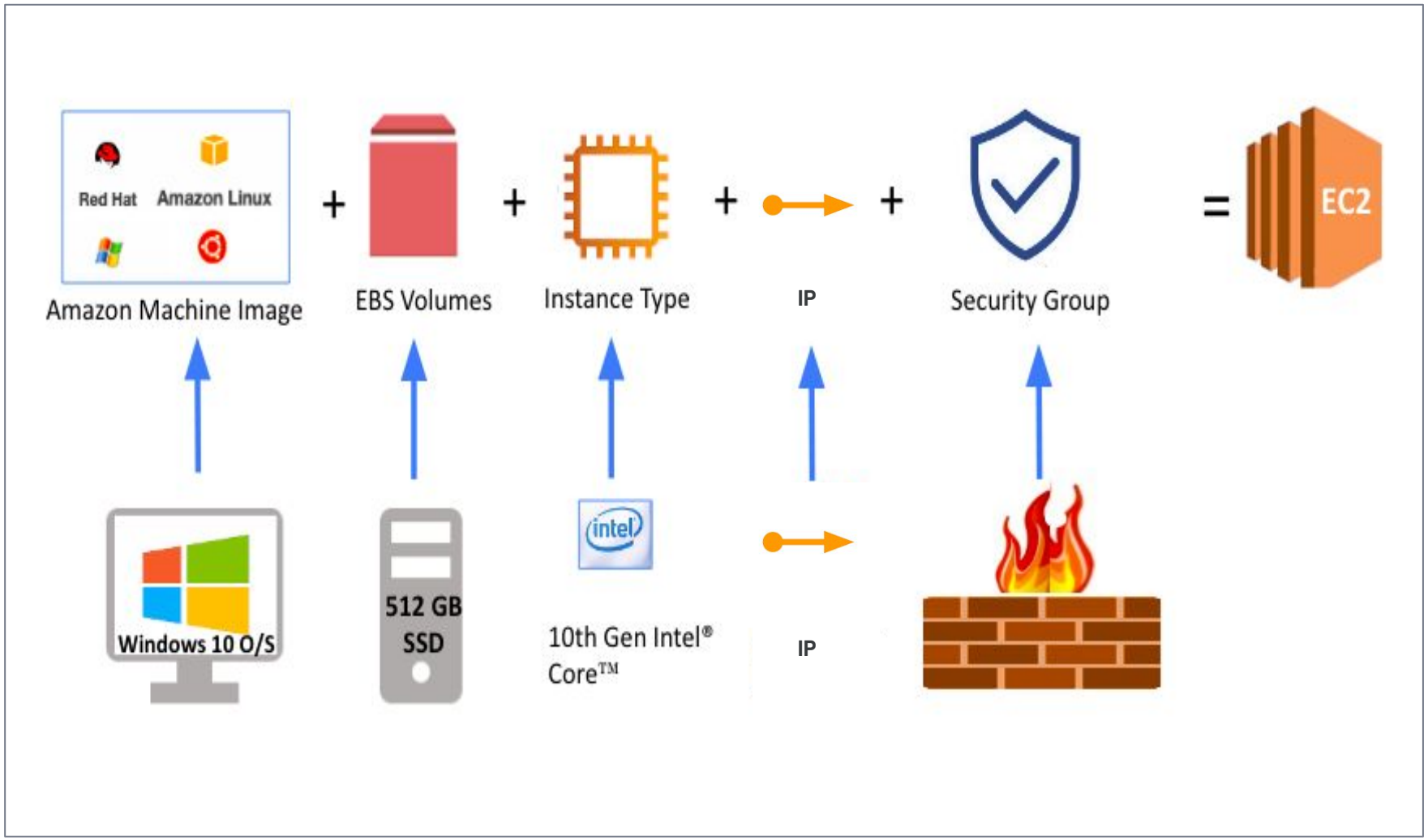


Introduction to EC2

What is EC2?

- In fact, EC2 is a kind of computer such as your desktop in your home.Components of the EC2 are similar to conventional computer devices.

- Each EC2 component refers to one of the conventional computer parts such as Operating System, Hard Disk and Intel/AMD processors, etc.



Introduction to EC2

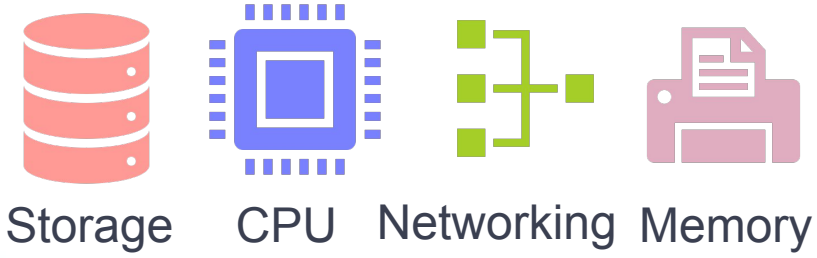
EC2 Features

- Pay as you go,
- Setup and ready to use within 1 minute,
- CPU, Memory and Storage Capacity needs can be arranged within minutes,
- Create, Stop or Terminate instances via EC2 console easily.

2 Types of Instances

EC2 Instances

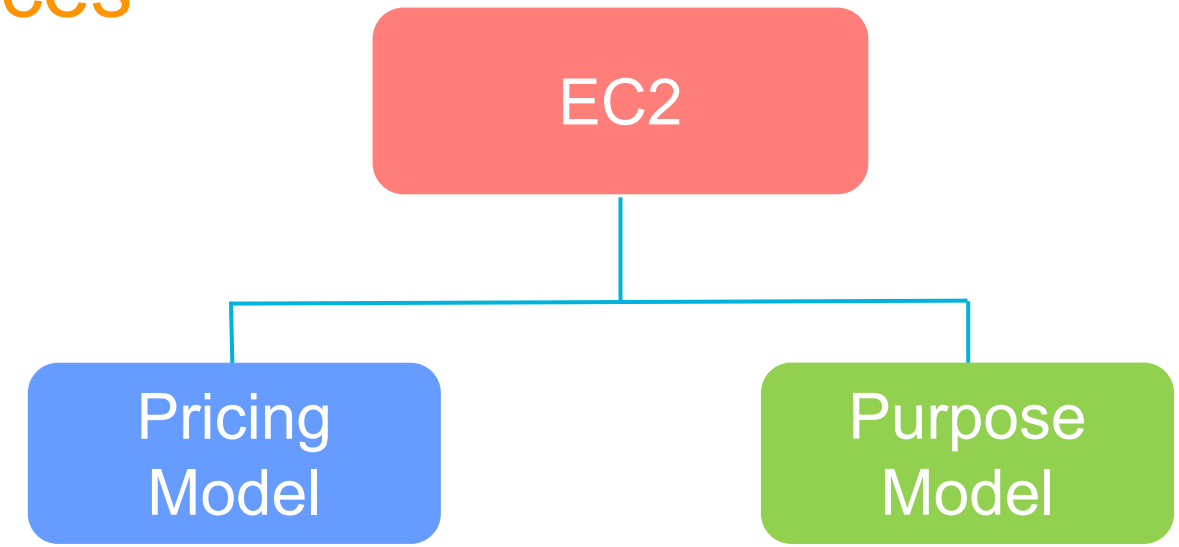
Types of Instances



- Amazon EC2 provides a wide selection of instance types optimized to fit different use cases.
- Instance types comprise varying combinations of CPU, memory, storage, and networking capacity

EC2 Instances

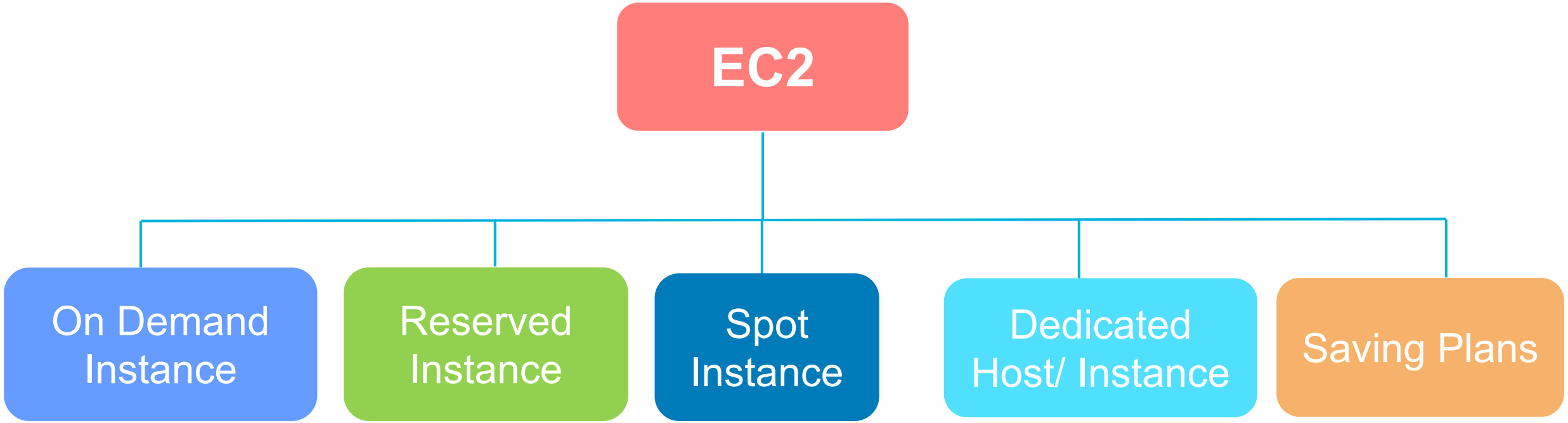
Types of Instances



- Instance types are grouped into a variety of families based on target application profiles and pricing options. It is possible to categorize EC2 types under two main perspective :
- These are Pricing Model and Purpose Model.

EC2 Instances

Pricing Model of Instances



When we look at the pricing perspective, AWS offers 5 different types of instance pricing.

* Capacity Reservation

EC2 Instances

On Demand Instances

- You pay for compute capacity by the “hour “or the “second”
- No commitments
- No upfront payments
- You can increase or decrease your compute capacity
- Pre-estimated

EC2 Instances

On Demand Instances

On-Demand instances are recommended for:

- Users that prefer 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

EC2 Instances

On Demand Pricing

- t2.micro in us-east-1 (N.Virginia)
- cost : $ 0.0116/hour

- 25 seconds usage--->>> $ 0.0116 / 60= $ 0.00019 (min 60 seconds)
- 60 seconds usage--->>> $ 0.0116 / 60= $ 0.00019 (min 60 seconds)
- 30 minutes usage--->>> $ 0.0116 / 2= $ 0.0058)
- 1 month usage---->>> $ 0.0116 * 24 *30 = $8.47)

* Add Pricing of EBS Volume + IPV4 (after February 1, 2024)

EC2 Instances

Reserved Instances (RI)

- Reserved Instances provide you with a significant discount (up to 75%) compared to On-Demand instance pricing.
- It is a tariff that takes advantage of the discounted price by giving AWS a 1 or 3-year commitment.

EC2 Instances

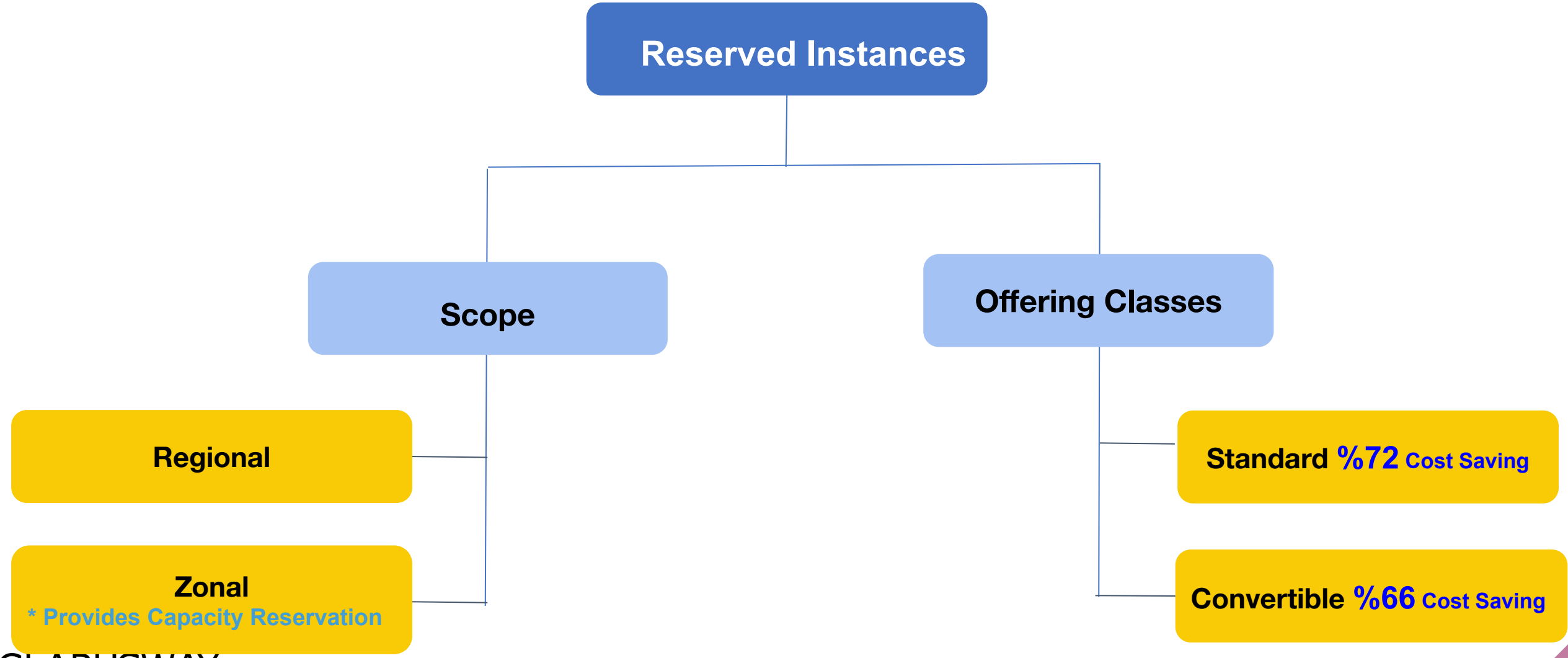
Reserved Instances (RI)

Reserved Instances are recommended for:

- Applications with steady state usage
- Applications that may require reserved capacity
- Customers that can commit to using EC2 over a 1 or 3 year term to reduce their total computing costs

EC2 Instances

Reserved Instances (RI)



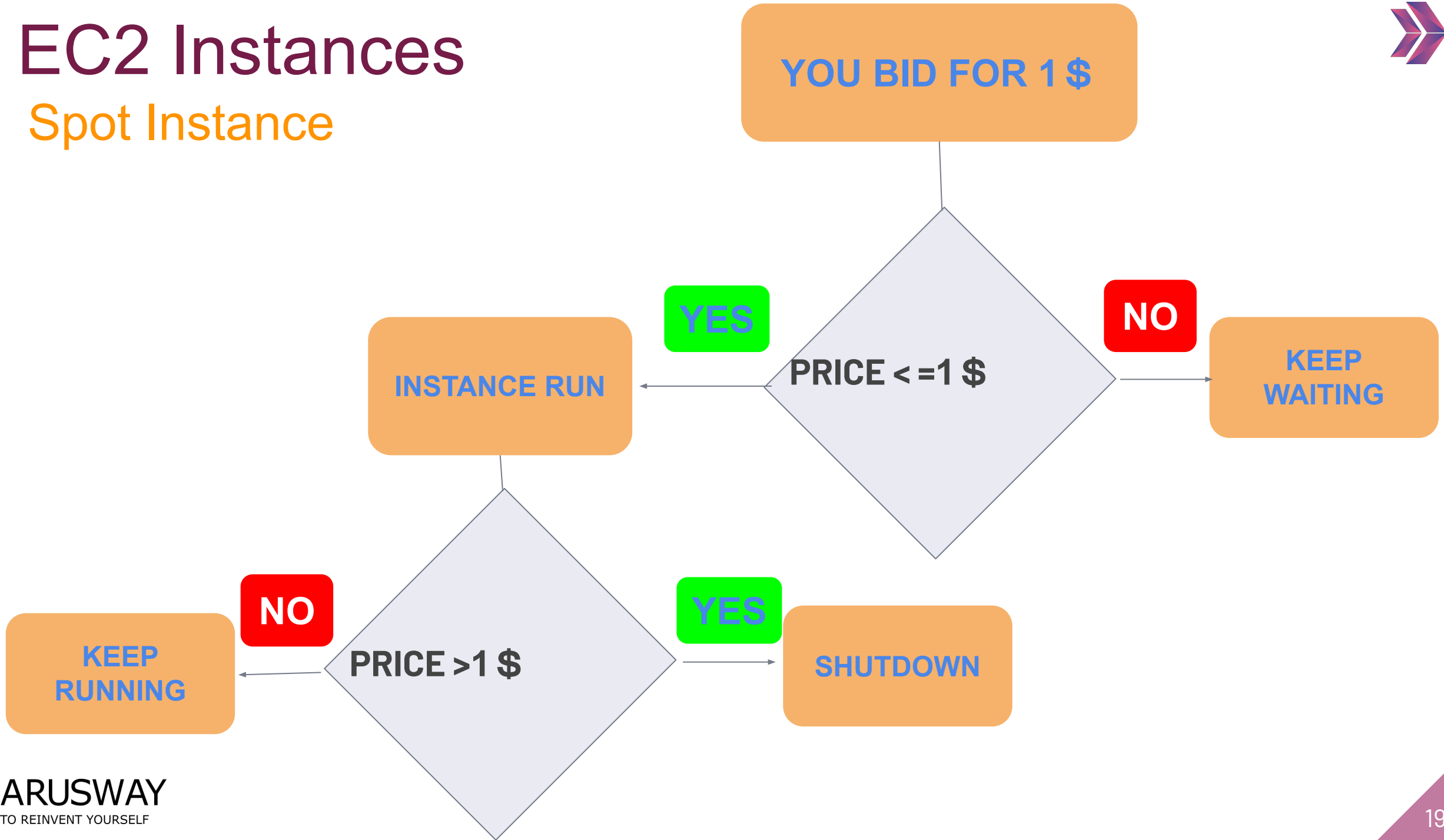
EC2 Instances

Spot Instance

- In Spot Instance, you can enter a purchase order by setting a target price.
- The machine runs when the current price falls below the target price.
- The machine automatically shuts down if the price exceeds that target price.
- You can save up to 90% cost advantage.

EC2 Instances

Spot Instance



EC2 Instances

Spot Instance vs. On Demand Price

45 minutes

Spot Instance

Pay for 45 minutes

???????

EC2 Instances

Spot Instance

Spot instances are recommended for:

- Applications that have flexible start and end times
- Non-continuity jobs such as testing

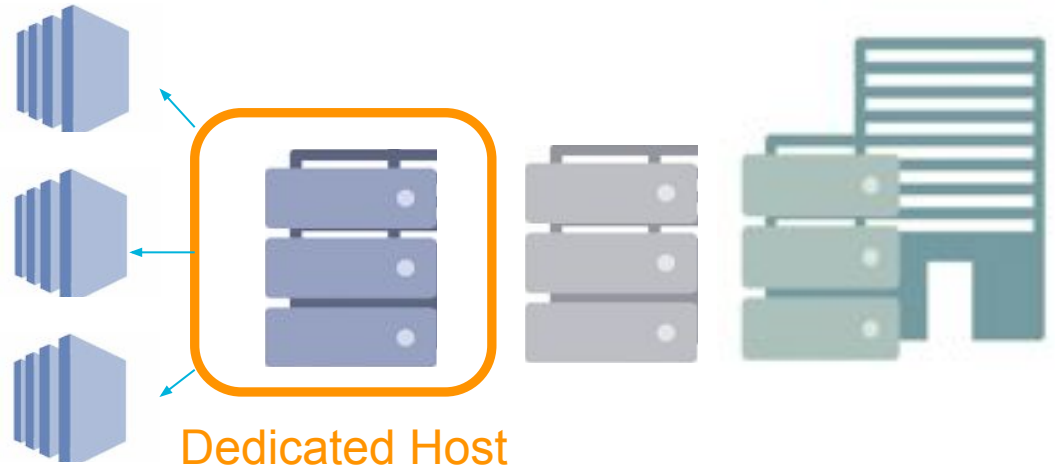
EC2 Instances

Dedicated Host/Instance

A Dedicated Host is a physical server the whole capacity of with EC2 instance is dedicated to your use.

Not only your instances are reserved but also they physically separated from the other servers.

A Dedicated Host consists of Dedicated Instance capacities according to your needs. You may choose to buy a Dedicated Host or only one Dedicated Instance also.



EC2 Instances

Saving Plans

ON DEMAND

5000 HOURS OF USAGE

1500$

SAVING PLAN

5000 HOURS OF USAGE

1000$

EC2 Instances

Saving Plans

EC2 Instance Savings Plans

- %72 Cost Saving
- You can change:
  - Size (within the instance family!!)
  - Operating system
  - Tenancy

Compute Savings Plans

- %66 Cost Saving
- You can change:
  - Instance family,
  - Size
  - Operating system,
  - Tenancy,
  - Other compute resources: Fargate and Lambda

EC2 Instances

Capacity Reservation

	Capacity Reservations	Zonal Reserved Instances	Regional Reserved Instances	Savings Plans
Term	No commitment required. Can be created and canceled as needed.	Requires a fixed one-year or three-year commitment		
Capacity benefit	Capacity reserved in a specific Availability Zone.		No capacity reserved.	
Billing discount	No billing discount. †	Provides a billing discount.		
Instance Limits	Your On-Demand Instance limits per Region apply.	Default is 20 per Availability Zone. You can request a limit increase.	Default is 20 per Region. You can request a limit increase.	No limit.

While Capacity Reservation enables reserving physical capacity for a specific AWS resource, Reserved Instance and Saving plan allows reserving a specific AWS resource for a certain period at a predetermined price, thereby reducing costs. Both strategies cater to different usage scenarios and requirements.

EC2 Instances Recap

Capacity Reservation

Dedicated Host/Instance

Spot

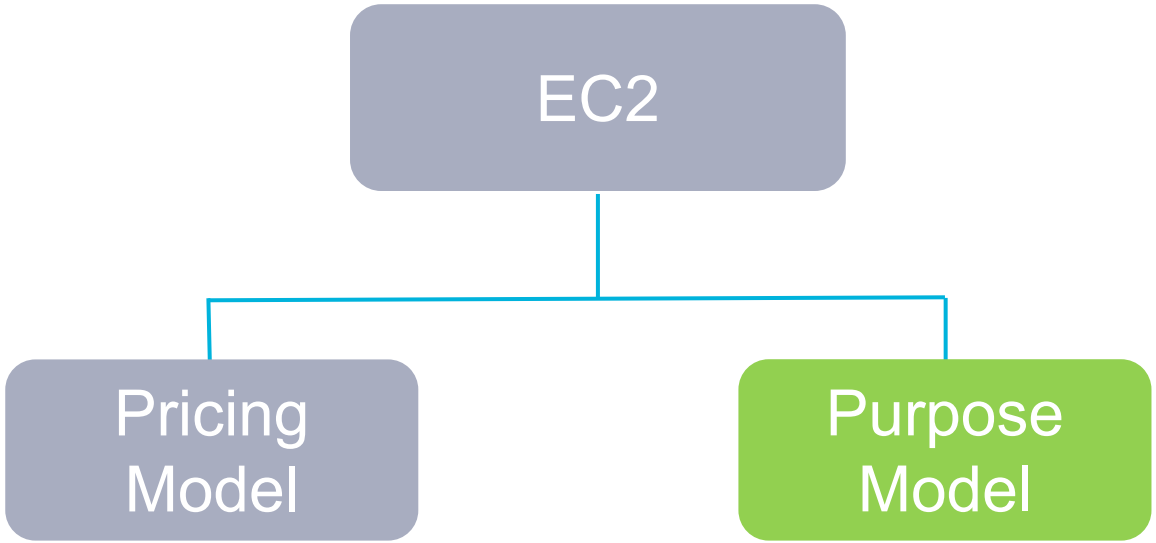
On Demand

Reserved

Saving Plan

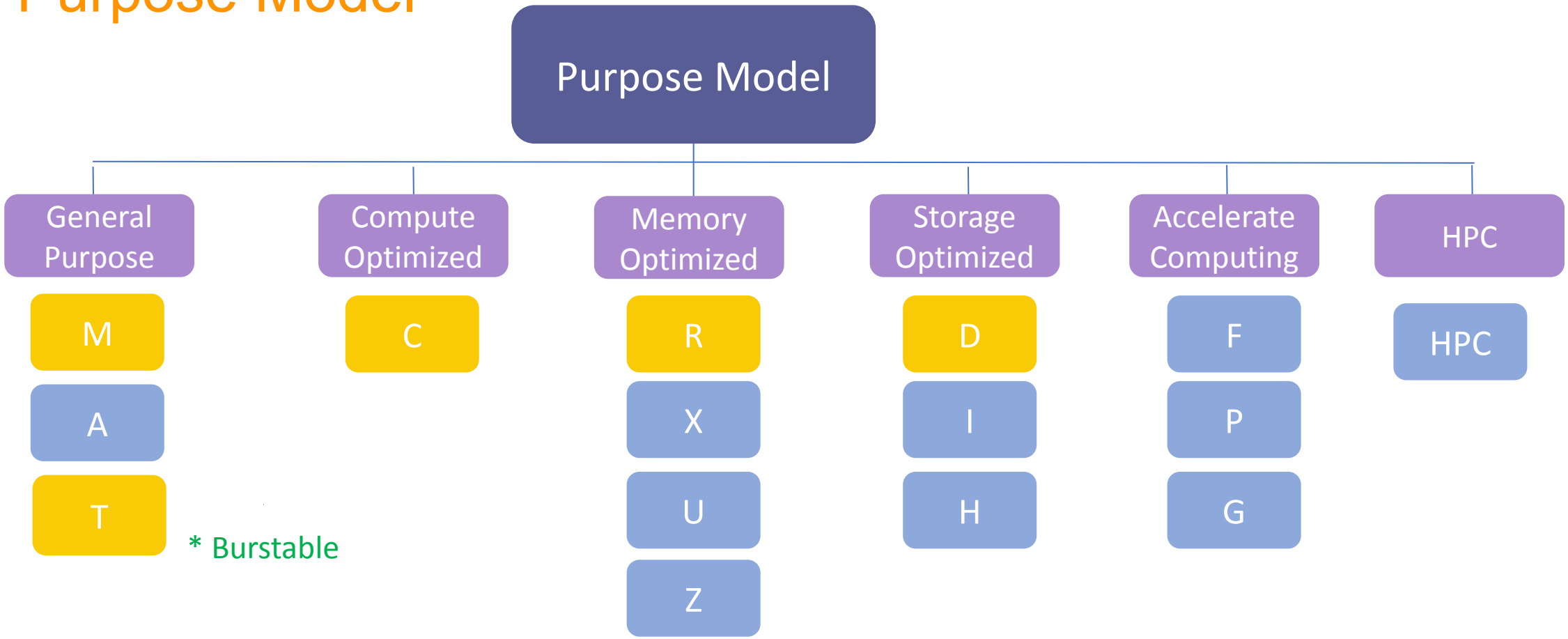
EC2 Instances

Types of Instances Recap



EC2 Instances

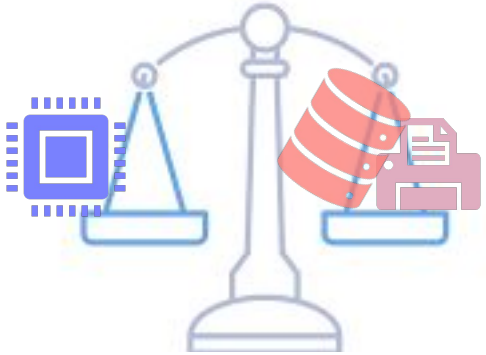
Purpose Model



AWS offers 15 different types of virtual machines in 6 categories

EC2 Instances

General Purpose

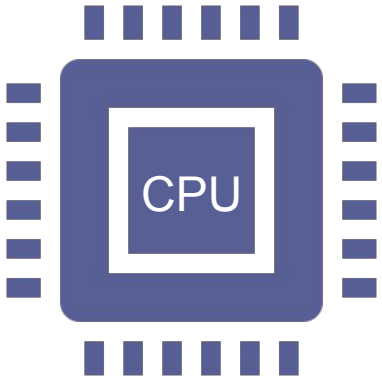


General Purpose

- General purpose instances provide a balance of compute, memory and networking resources, and can be used for a variety of diverse workloads.
- There are T, M and A options that we can use for standard and application needs.
- This is the most commonly used instance type and ideal for web servers.

EC2 Instances

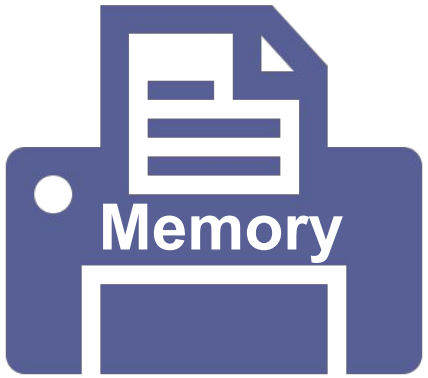
Compute Optimized



- Compute Optimized instances are ideal for compute bound applications that benefit from high performance processors.
- Instances belonging to this family are well suited for batch processing workloads, media transcoding, high performance web servers, dedicated gaming server, etc.

EC2 Instances

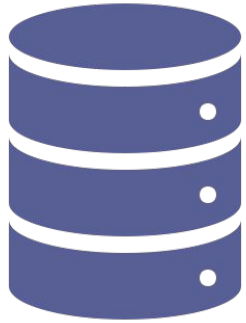
Memory Optimized



- Memory optimized instances are used in situations requiring a high-performance database, real-time large data analytics, and high memory usage.
- There are R, X, Z and U type instances in this category.

EC2 Instances

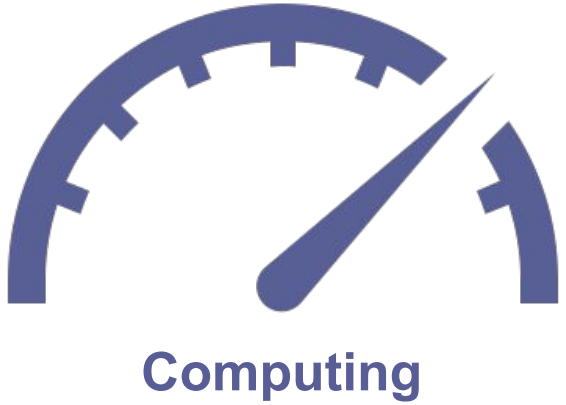
Storage Optimized



- Storage optimized instances are designed for workloads that require high, sequential read and write access to very large data sets on local storage.
- It is the best used for the fast disk structures we need in NoSQL databases or data warehouse solutions.
- There are D, H and I type of instances in this category.

EC2 Instances

Accelerated Computing



- Preferred when you need machine learning, deep learning calculation, and analysis.
- There are F, P and G type of instances in this category.

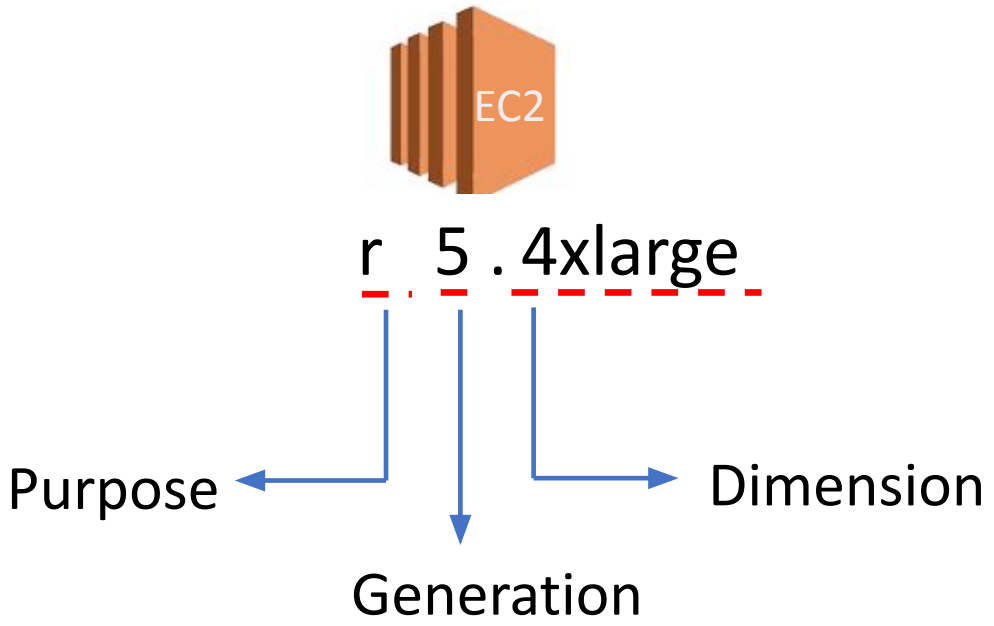
EC2 Instances

HPC Optimized

- HPC instances are ideal for applications that benefit from high-performance processors such as large, complex simulations weather forecasting, molecular dynamics and deep learning workloads.

EC2 Instances

Instance Coding



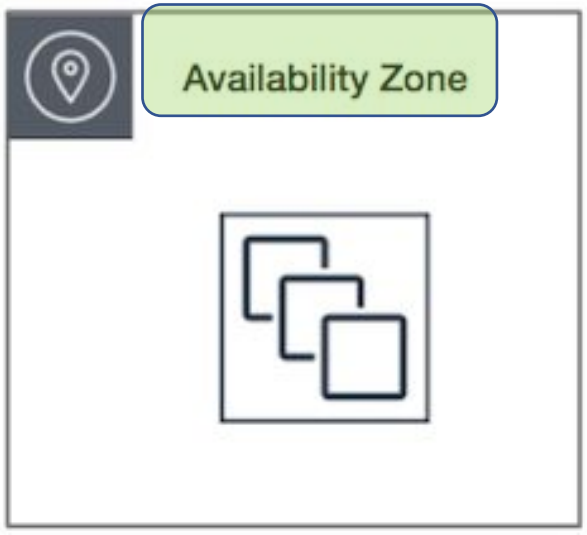
- R refers to its purpose. It means this EC2 is Memory Optimized instance.
- 5 refers to instance generation. For example, the last generation of the r-family is 5.
- 4xlarge refers to dimension of instance. AWS has built servers of various sizes to suit every need in instance families. For example, the r5-family has 8 different sizes starting from **large to 24xlarge.**
- Not all models have instances in every generation and size.

EC2 Instances

Placement Group

1 AZ

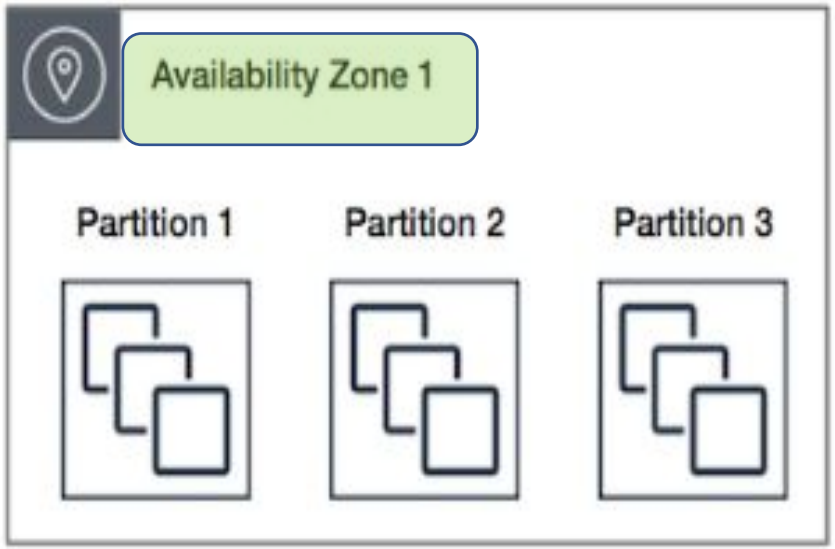
Cluster placement



Exam Tip: low latency& high performance

Multiple AZ in Single Region

Partition placement



-Hadoop, Cassandra, and Kafka-
-Prevent correlated failures

Multiple AZ in Single Region

Spread placement

High Availability

- **Prevent simultaneous failure**

Introduction to EC2

Let's get our hands dirty!

- Introduction of EC2 console
- Creating an EC2 instance
- Creating an EC2 instance with user data
- Working with Instance Actions

THANKS!

Any questions?

You can find me at:

- @osvaldo
- osvaldo@clarusway.com