AWS Solutions Architect Associate

Session 501

Compute: EC2

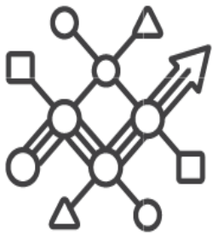
July/2024

Elastic Compute Cloud – EC2

Web hosting

Databases

Authentication



Anything a server can do

AMI

Instance

T3 instance

EC2 Benefits

Amazon EC2 can solve some problems that are more difficult with a on-premises server.

When using disposable resources

Data-driven decisions

Quick iterations

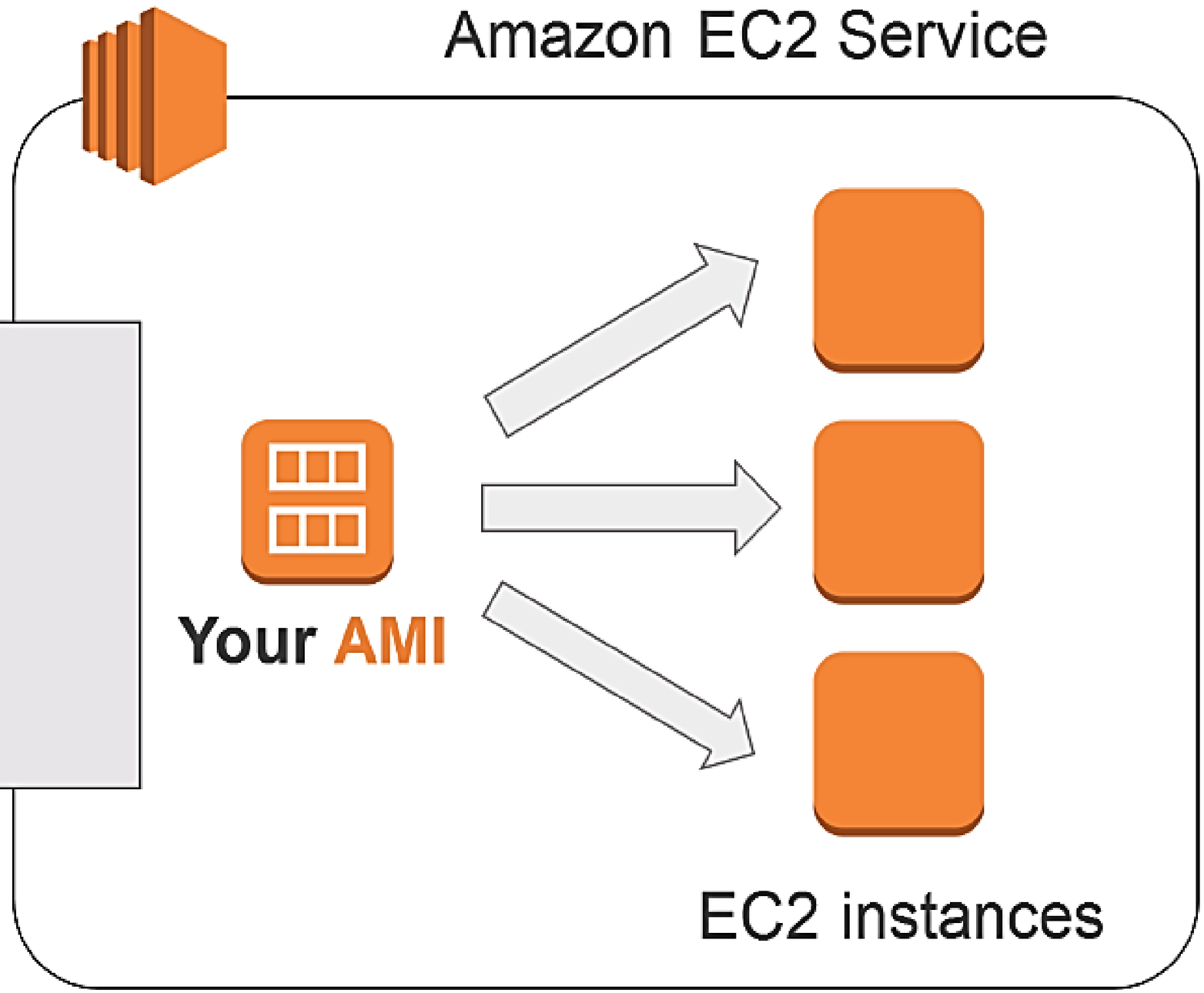
Free to make mistakes

Amazon Machine Image - AMI

AMIs include:

A template for the root volume
(Copy of the boot drive)
Launch permissions

A block device mapping



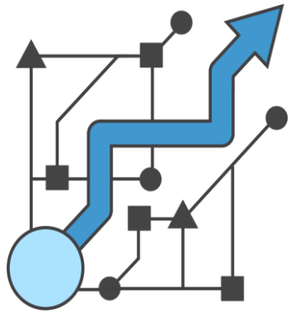
AMI = Amazon Machine Image

Sources of AMI

Three ways to get your **AMI**

Pre-Built

Marketplace



Create your own

AMI Benefits

Repeatability

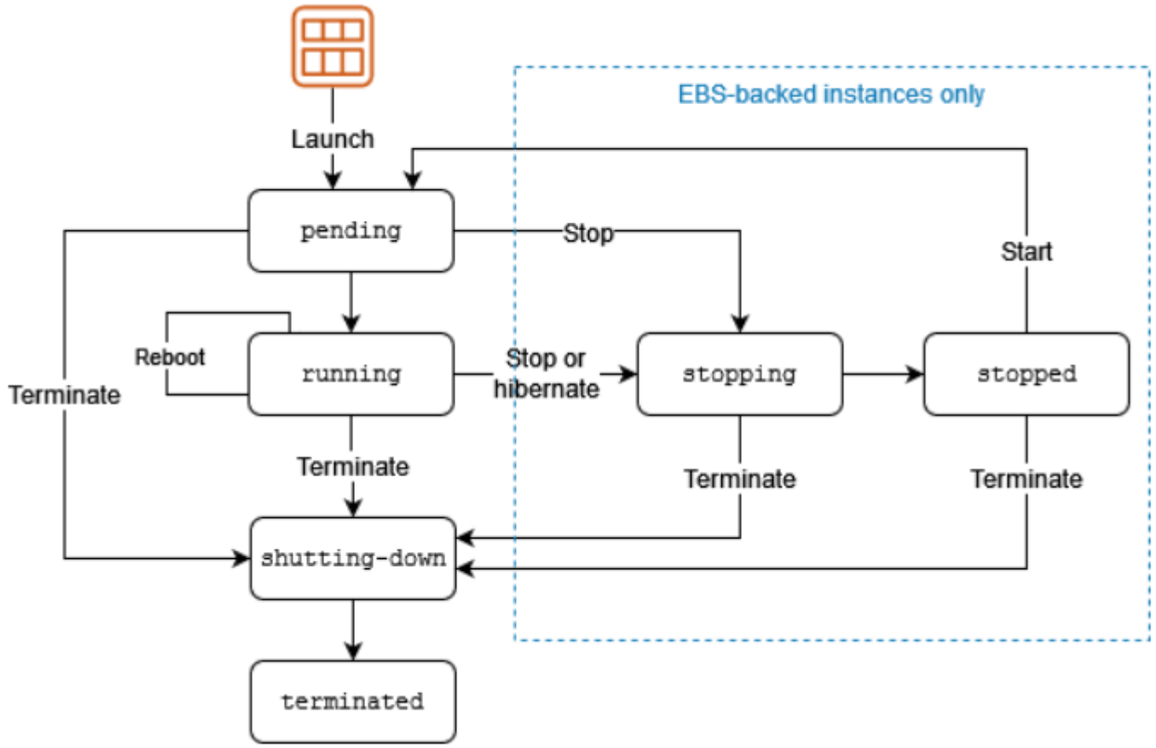
Reusability

Recoverability

Marketplace Activities

Backups

Instance Lifecycle

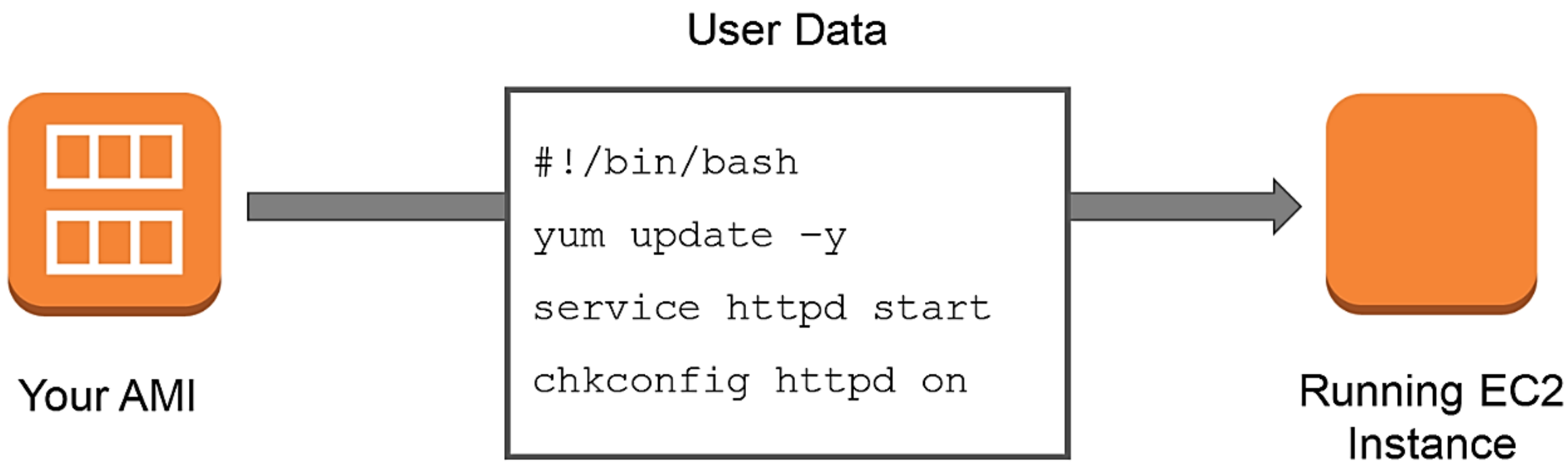


Additional charges can apply: EBS, EIP.
Hibernation is very similar to Stop when the difference of Ram contens is saved on disk.
IPv4 and IPv6 is retain on stopped status.

Instance state	Description	Instance usage billing
`pending`	The instance is preparing to enter the `running` state. An instance enters the `pending` state when it is launched or when it is started after being in the `stopped` state.	Not billed
`running`	The instance is running and ready for use.	Billed
`stopping`	The instance is preparing to be stopped.	Not billed
`stopped`	The instance is shut down and cannot be used. The instance can be started at any time.	Not billed
`shutting-down`	The instance is preparing to be terminated.	Not billed
`terminated`	The instance has been permanently deleted and cannot be started.	Not billed **Note** Reserved Instances that applied to terminated instances are billed until the end of their term according to their payment option. For more information, see Reserved Instances for Amazon EC2 overview

Taken from https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ec2-instance-lifecycle.html (18/07/2024)

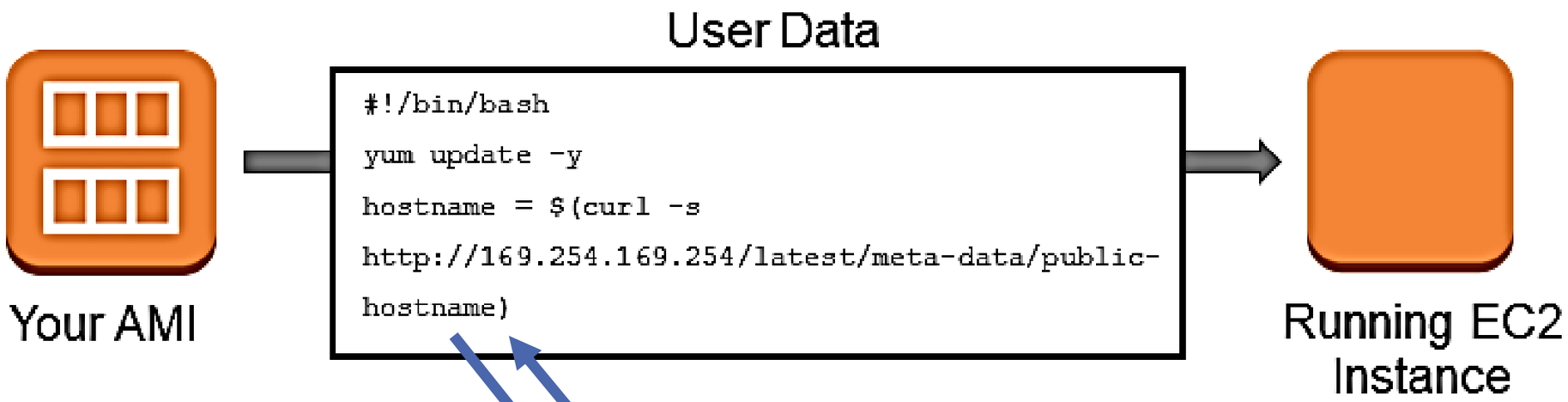
User Data



Also called bootstrap script. Work on Windows and Linux.

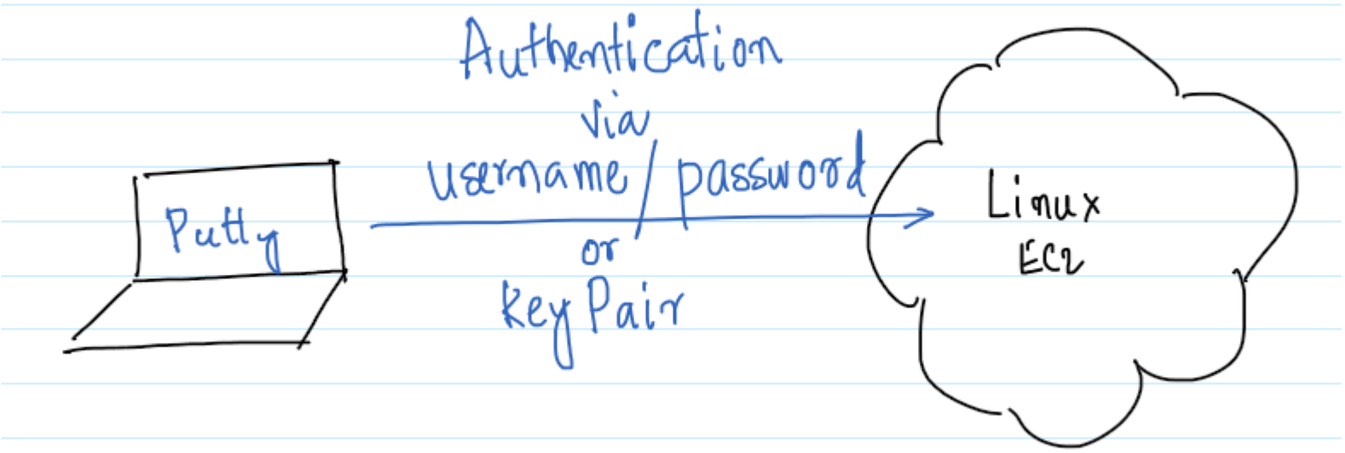
MetaData URL

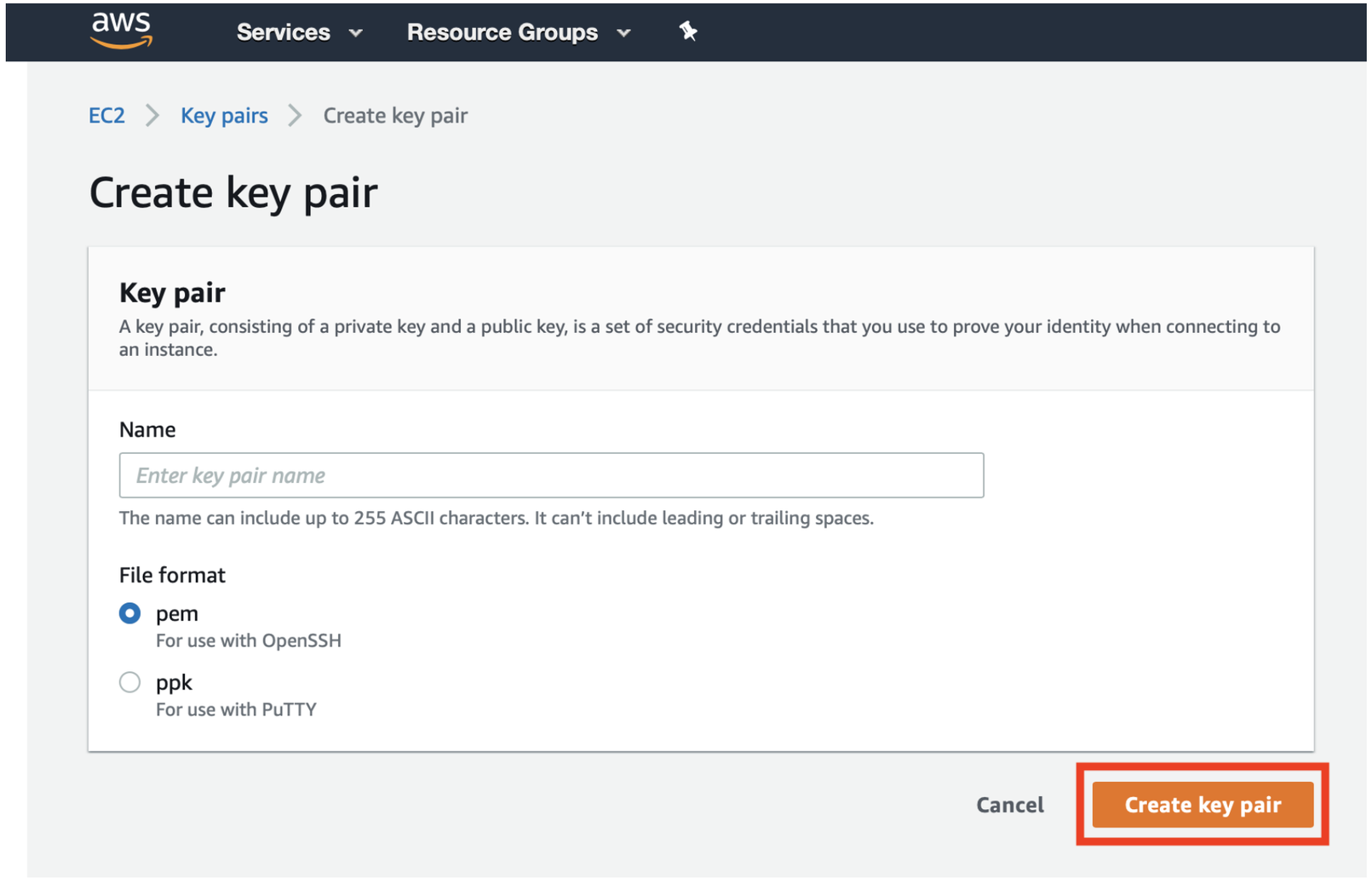
Metadata URL: To get on running instances. 169.254.169.254



Metadata	Value
instance-id	i-1234567890abcdef0
mac	00-1B-63-84-45-E6
public-hostname	ec2-203-0-113-25.compute-1.amazonaws.com
public-ipv4	67.202.51.223
local-hostname	ip-10-251-50-12.ec2.internal
local-ipv4	10.251.50.12

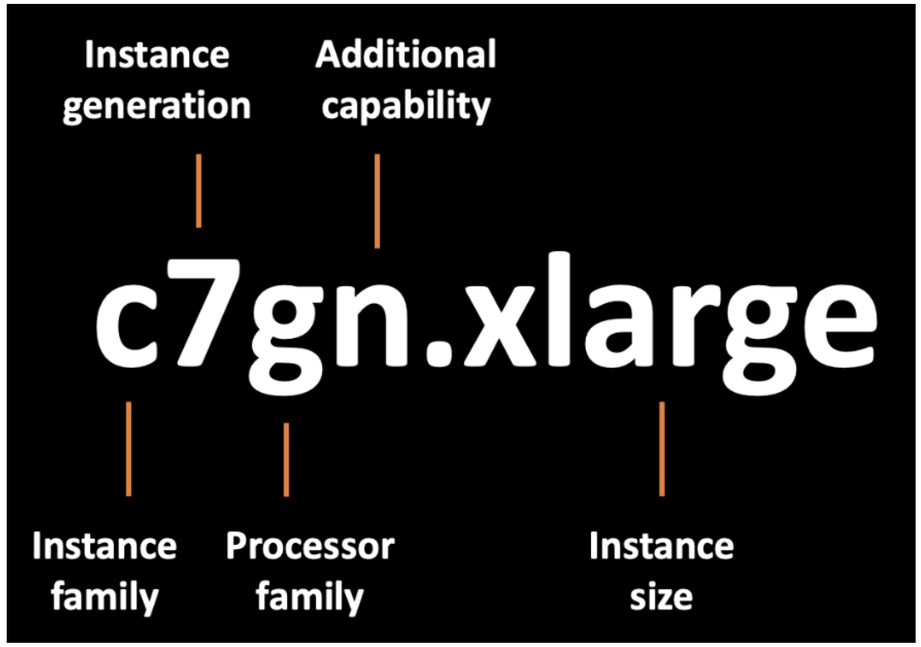
KeyPair





Mandatory to the lab. Read https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/connect-to-linux-instance.html (18/07/2024)

Instance Type Naming

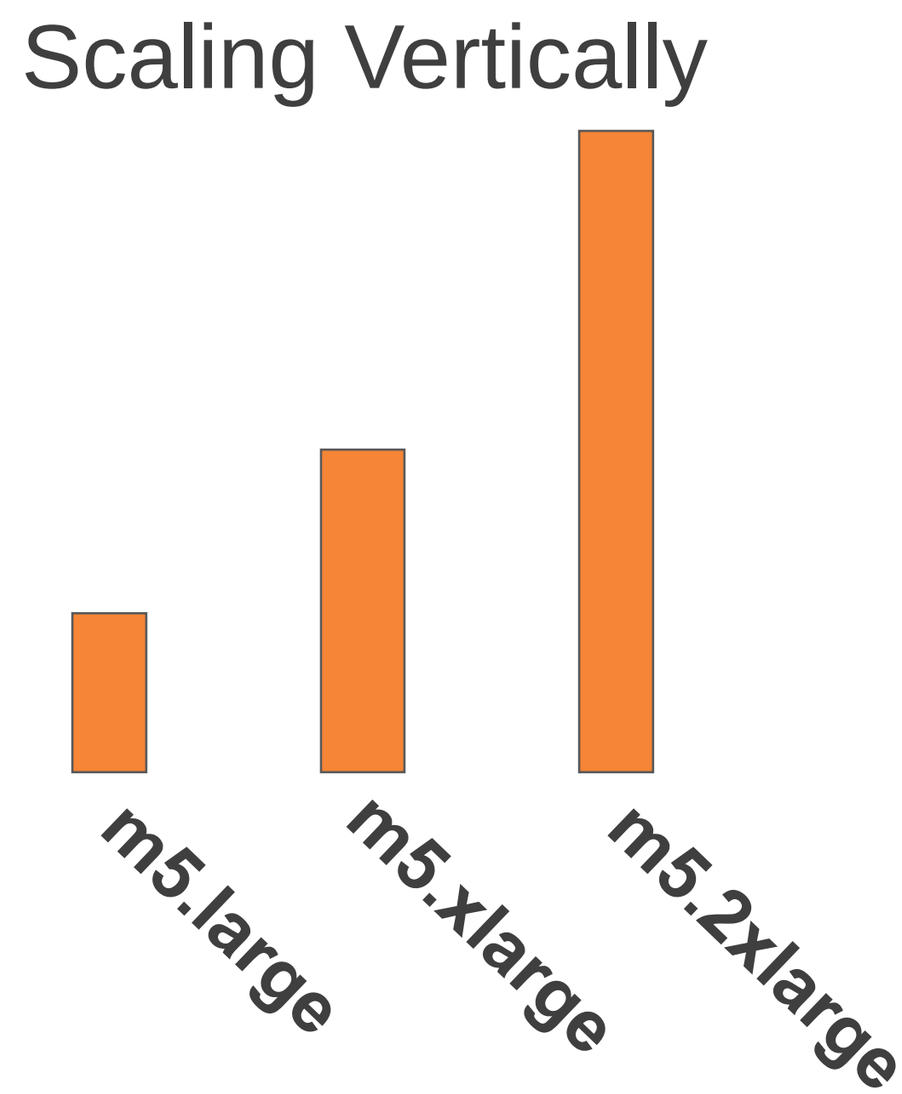


Taken from https://docs.aws.amazon.com/ec2/latest/instancetypes/instance-type-names.html (18/07/2024)

Instance families	Processor families	Additional capabilities
• **C** – Compute optimized • **D** – Dense storage • **F** – FPGA • **G** – Graphics intensive • **Hpc** – High performance computing • **I** – Storage optimized • **Im** – Storage optimized (1 to 4 ratio of vCPU to memory) • **Is** – Storage optimized (1 to 6 ratio of vCPU to memory) • **Inf** – AWS Inferentia • **M** – General purpose • **Mac** – macOS • **P** – GPU accelerated	• **a** – AMD processors • **g** – AWS Graviton processors • **i** – Intel processors	• **b** – Block storage optimization • **d** – Instance store volumes • **e** – Extra storage or memory • **flex** – Flex instance • **n** – Network and EBS optimized • **q** – Qualcomm inference accelerators • **z** – High performance

Model Example

Model	vCPU
m5.large	2
m5.xlarge	4
m5.2xlarge	8
m5.4xlarge	16
m5.12xlarge	48
m5.24xlarge	96

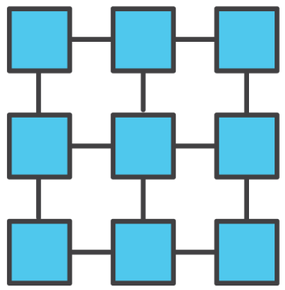


EC2 Types

Choosing the correct type is very important for:

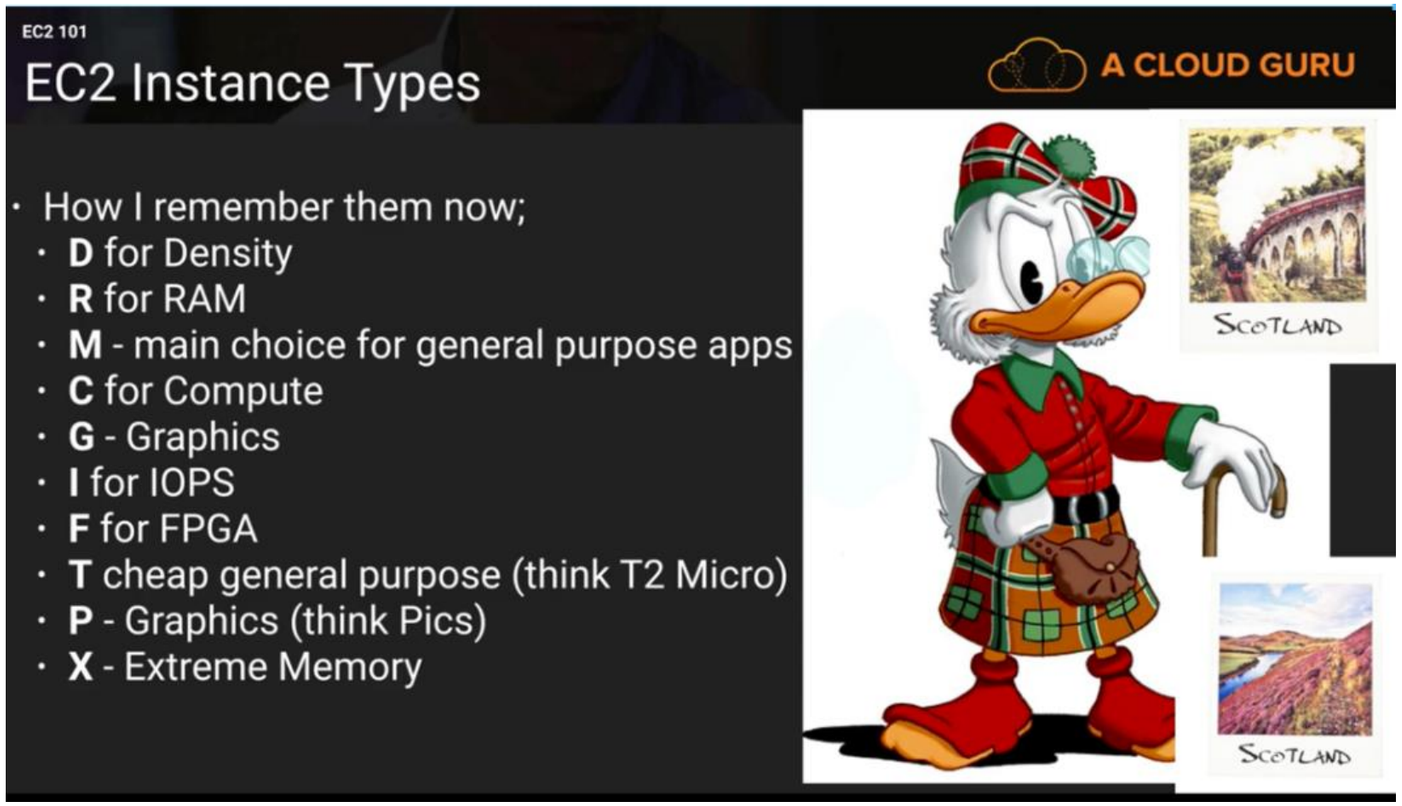
T
M
C
X
R
P
F
H
I
D
G

Efficient utilization of your instances



Reducing unneeded cost

https://www.ec2instances.info/



EC2 Instance Type - Example

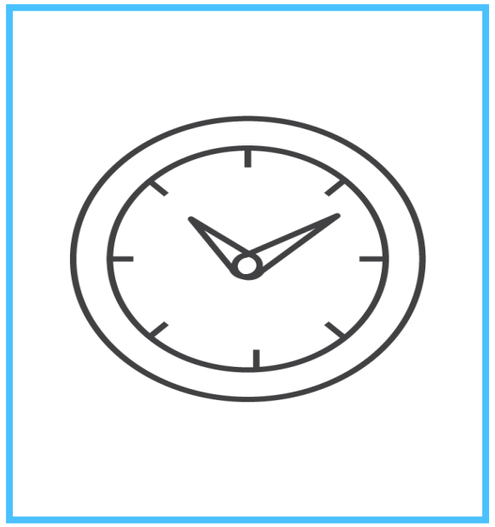
aws EC2 instance types

	General Purpose		Compute Optimized	Memory Optimized		Accelerated Computing	Storage Optimized		
Type	t2	m5	c5	r4	x1e	p3	h1	i3	d2
Description	Burstable, good for changing workloads	Balanced, good for consistent workloads	High ratio of compute to memory	Good for in-memory databases	Good for full in-memory applications	Good for graphics processing and other GPU uses	HDD backed, balance of compute and memory	SDD backed, balance of compute and memory	Highest disk ratio
Mnemonic	**t** is for **tiny** or **turbo**	**m** is for **main** or happy **medium**	**c** is for **compute**	**r** is for **RAM**	**x** is for **xtreme**	**p** is for **pictures**	**h** is for **HDD**	**i** is for **IOPS**	**d** is for **dense**

ParkMyCloud

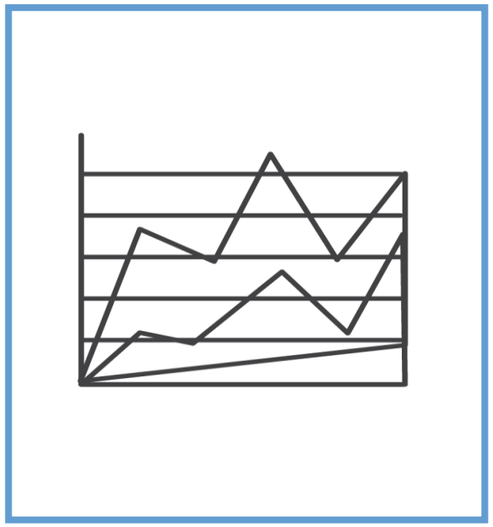
Taken from https://www.business2community.com/product-management/ec2-instance-types-comparison-and-how-to-remember-them-02097611 (18/07/2024)

On-Demand Instances



Savings Plan

Spot Instances



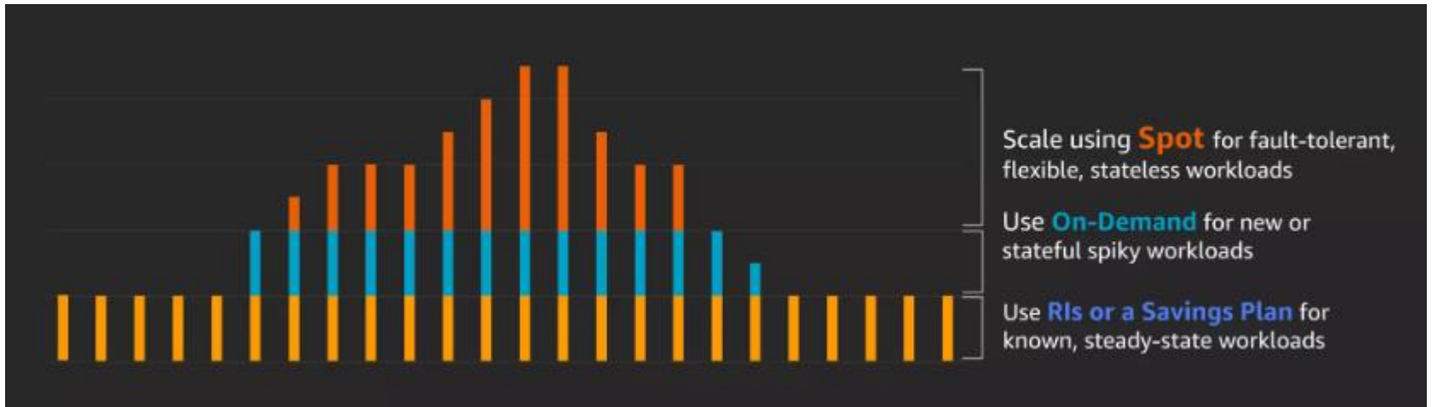
Modification using Savings Plan for Reserved Instances on Nov/2019.

Taken from https://techcrunch.com/2019/11/07/aws-announces-new-savings-plans-to-reduce-complexity-of-reserved-instances/ and https://www.gorillastack.com/news/aws-savings-plans-reserved-instances/ on 20/05/2020

More info at SAP at AWS at https://www.slideshare.net/AmazonWebServices/track-3-session-5-amazon-ec2#29 (18/07/2024)

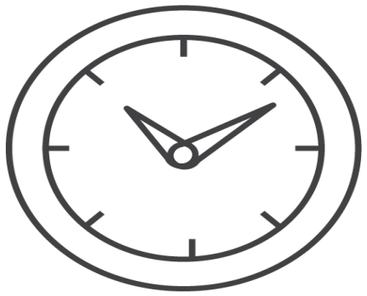
AWS Savings Plans and Reserved Instances Comparison

Unit	Reserved Instance	EC2 Instance Savings Plan	Compute Savings Plan
Average 1y Discount	38%	29%	29%
Average 3y Discount	58%	58%	51%
Instance Family	Fixed	Fixed	Flexible
Instance size	Fixed (except linux)	Flexible	Flexible
Geography	1 Region	1 Regions	Flexible
OS	Fixed	Flexible	Flexible
Service	EC2 / RDS	EC2	EC2 / Fargate



OnDemand Instances

On-Demand Instances



- Pay for compute capacity per second (Amazon Linux and Ubuntu) or by the hour (all other OS)
- No long-term commitments
- No upfront payments
- Increase or decrease your compute capacity depending on the demands of your application

Solves the need for immediate compute capacity

Savings Plan

Savings Plan

Comparing Reserved Instances & Savings Plans

Standard RI
AZ, size (Linux)
Discount up to 72%

Convertible RI
AZ, size, family, OS, tenancy
Discount up to 66%

EC2 Savings Plan
AZ, size, OS, tenancy
Discount up to 72%

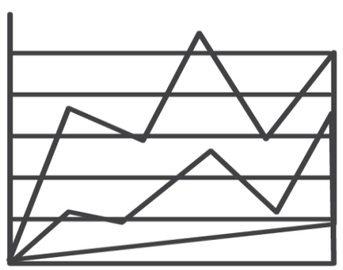
Compute Savings Plan
AZ, size, family, OS, tenancy, region, service
Discount up to 66%

UNIT	STANDARD RESERVED INSTANCE	EC2 INSTANCE SAVINGS PLAN	CONVERTIBLE RESERVED INSTANCE	COMPUTE SAVINGS PLAN
OS	Fixed	Automatically Flexes	Automatically Flexes	Automatically Flex
GEOGRPHY	Region-specific	Region-specific	Region-specific	Available in all regions
AVERAGE DISCOUNT/1 YEAR	38%	38%	29%	29%
AVERAGE DISCOUNT/3 YEARS	58%	58%	51%	51%
INSTANCE FAMILY	Fixed	Fixed	Fixed	Flexible
INSTANCE SIZE	Fixed	Fixed	Fixed	Flexible
SERVICE USE	EC2 only	EC2 only	EC2 only	EC2, Fargate, and Lambda

Taken from https://jason-6.medium.com/aws-ri-vs-savings-plan-10bb683896d9 , https://www.cloudforecast.io/blog/savings-plans-vs-reserved-instances/ and https://aws.amazon.com/ec2/pricing/reserved-instances/pricing/ (18/07/2024)

Spot Instances

Spot Instances

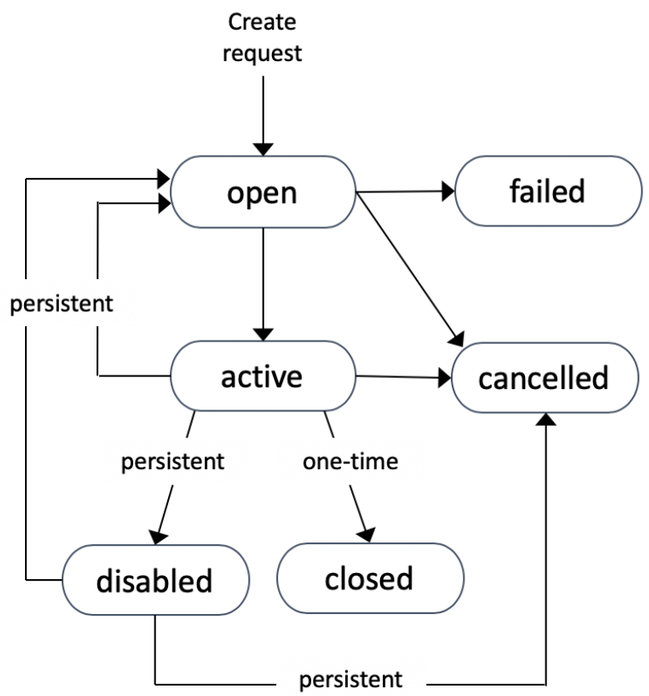


- Bid for unused Amazon EC2 capacity
- Prices controlled by AWS based on supply and demand
- Termination notice provided 2 minutes prior to termination

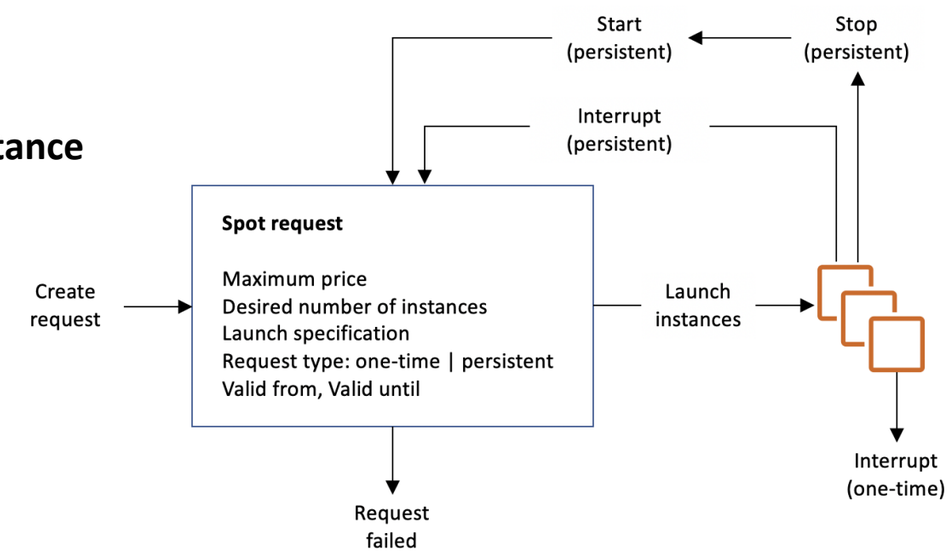
Can provide the steepest discounts as long as your workloads withstand starting and stopping

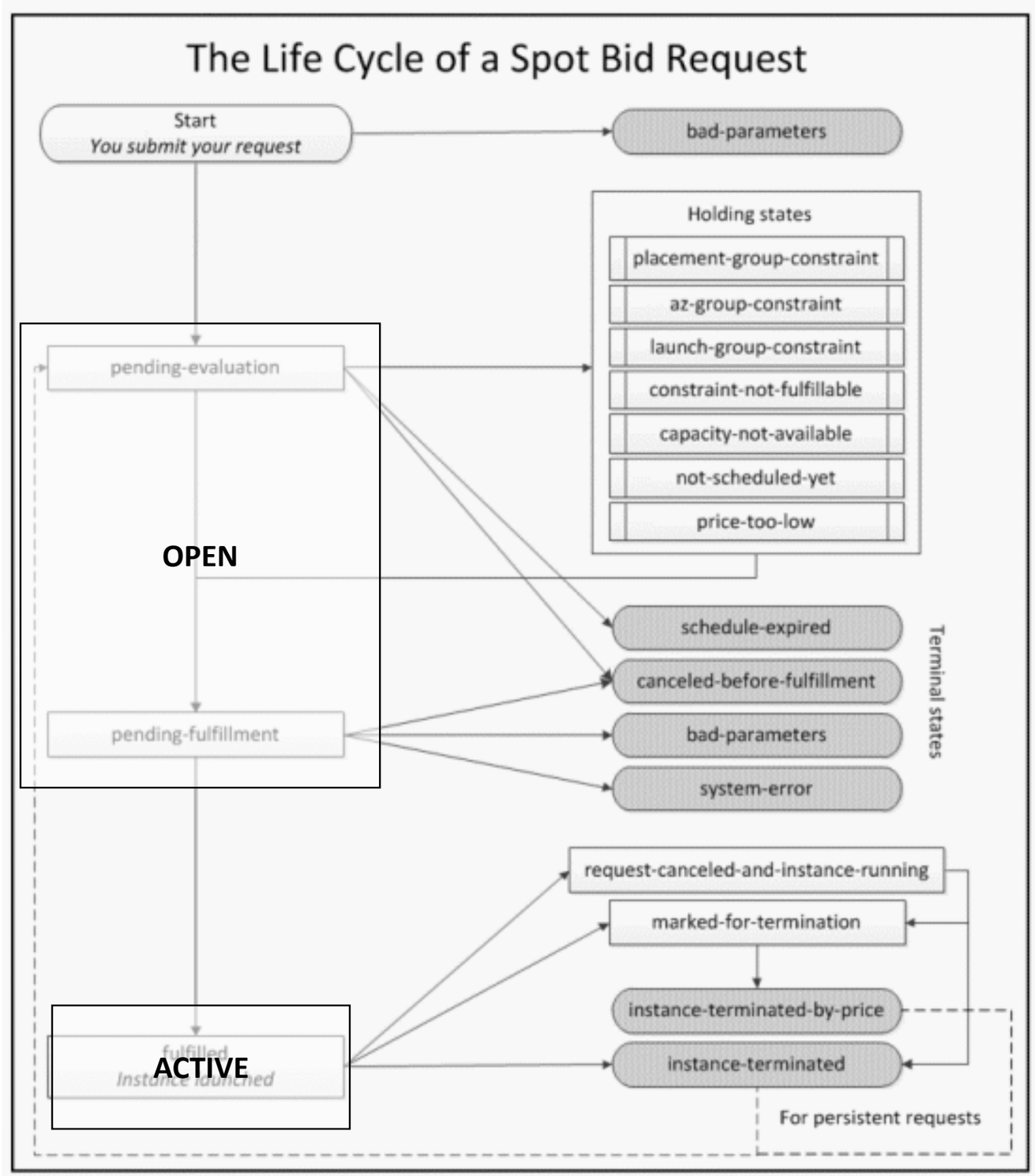
Spot Lifecycle

Spot Bid Request Lifecycle



Spot Instance Lifecycle



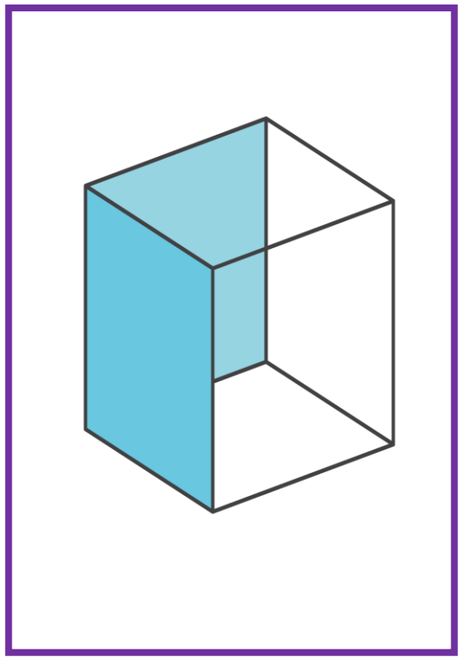


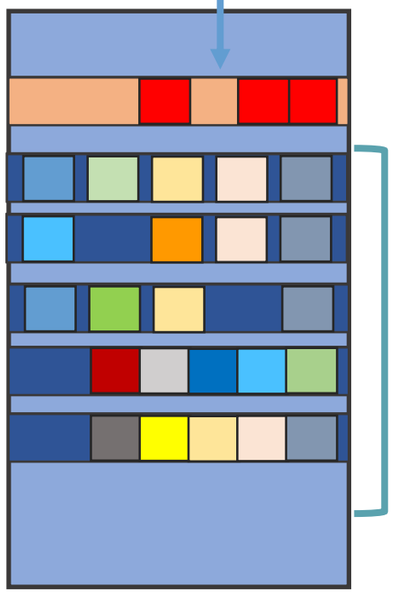
Taken from https://aws.amazon.com/blogs/aws/amazon-ec2-spot-instance-bid-status/ and https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/spot-requests.html (18/07/2024)

Dedicated Instances

Dedicated instances are physically isolated from other AWS accounts

Dedicated Instances





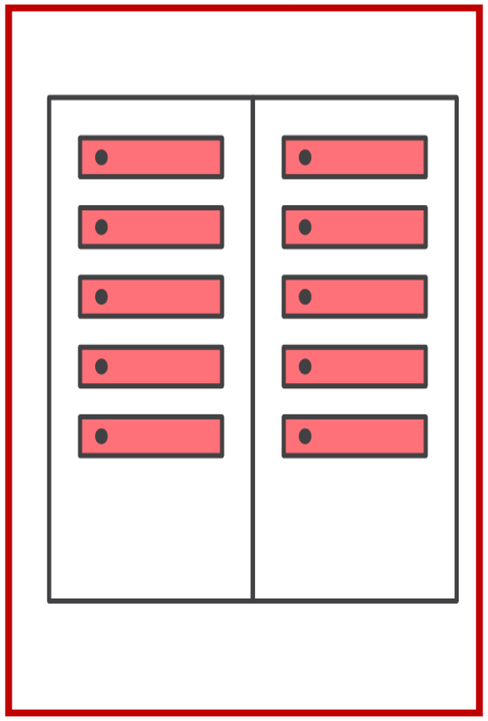
Helps meet requirements for regulatory compliance or software license use

You share your dedicated instances with other of your instances (shared) on the same server.

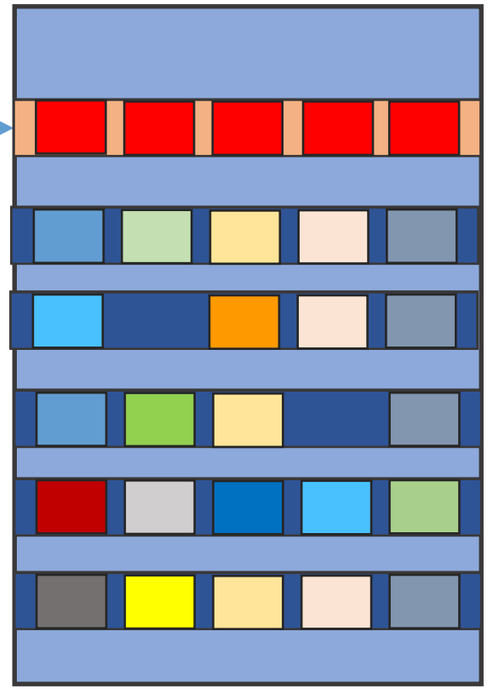
Dedicated Hosts

A *dedicated host* is a full physical server with EC2 instance capacity fully dedicated to your use.

Dedicated Hosts



Host ID: h-039725dyhe980010



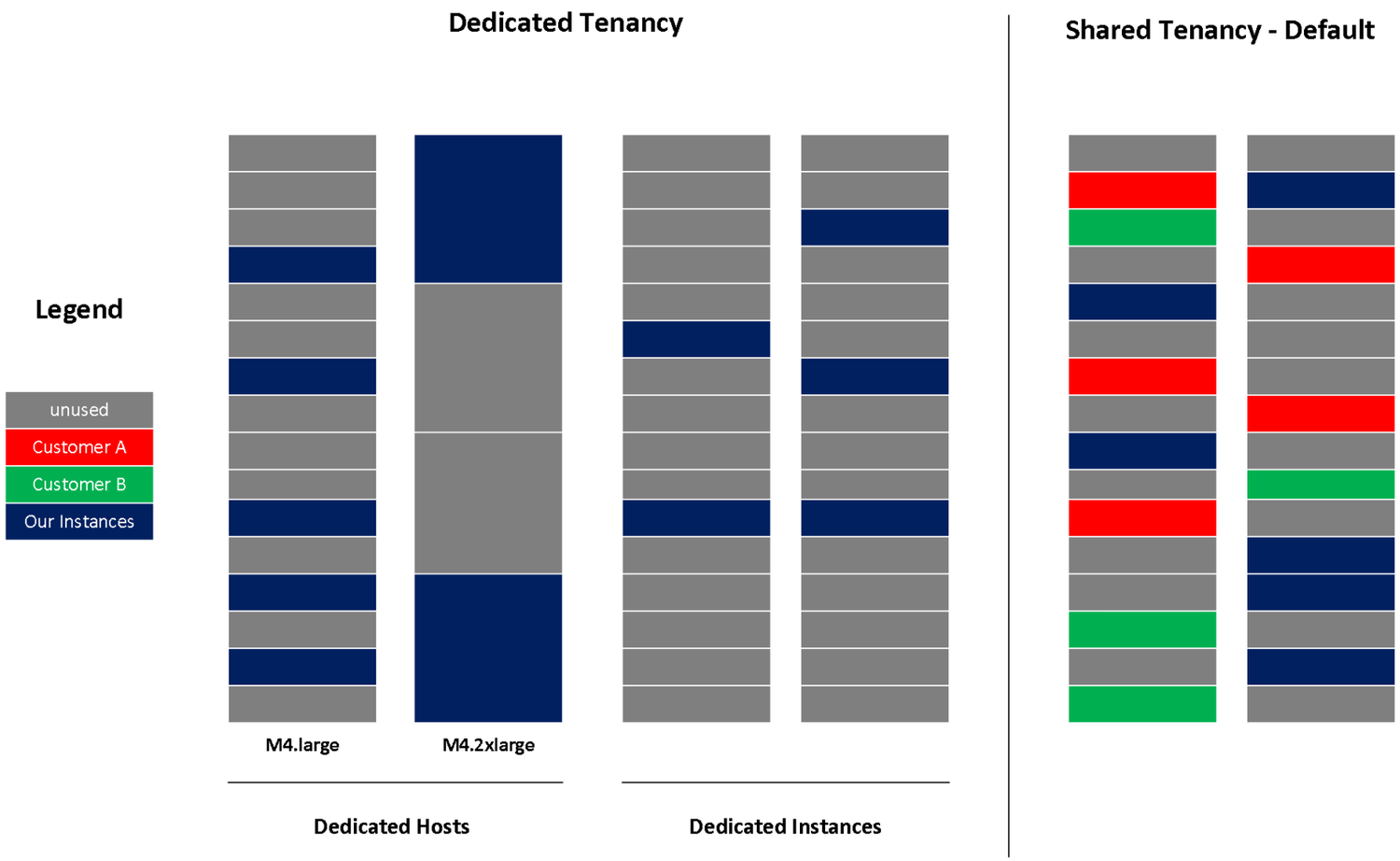
Helps meet *strict* requirements for regulatory compliance or software license use

AWS give you a machine with a number of sockets and cores to used, you choose the family and you deploy the number of that instances than you need.
Scenarios: BYOL

Taken from https://aws.amazon.com/es/ec2/dedicated-hosts/ and https://aws.amazon.com/ec2/dedicated-hosts/pricing/?nc1=h_ls on 11/05/2020

Dedicated Host vs Instances

	Only your AWS account on the hardware?	Description
Default	No	Your instance runs on shared hardware.
Dedicated Instance	Yes	Runs on a non-specific piece of hardware.
Dedicated Host	Yes	Runs on a specific piece of hardware of your choosing, over which you receive greater control.

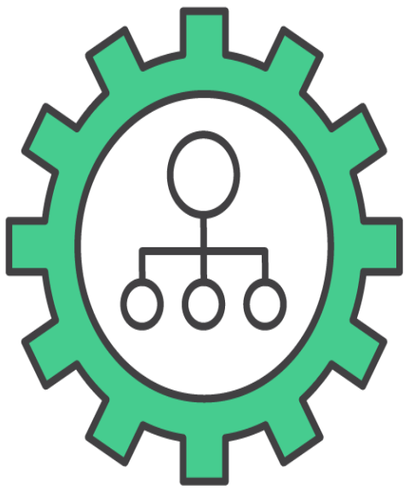


Taken from https://www.trek10.com/blog/dedicated-hosts-and-dedicated-instances (13/10/2020) and https://theithollow.com/2017/10/16/understanding-aws-tenancy/ on 11/05/2020

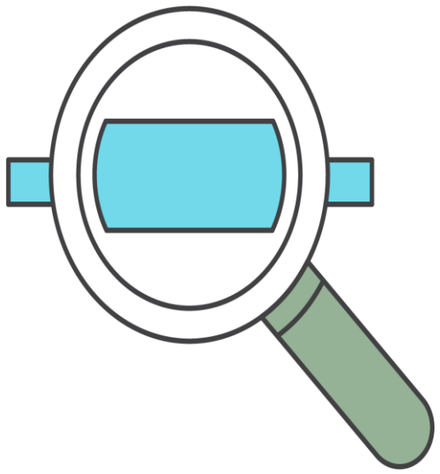
Tags

Assign metadata **tags** to your AWS resources to help you:

Manage



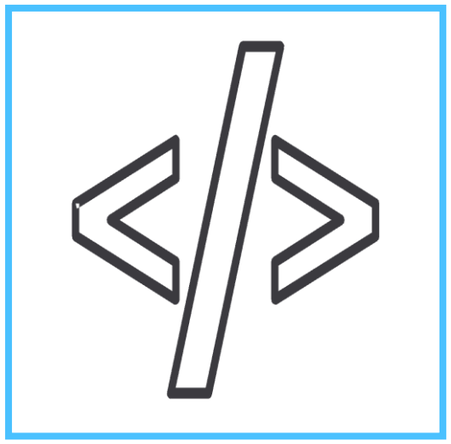
Search



Filter

Tag Resource Groups

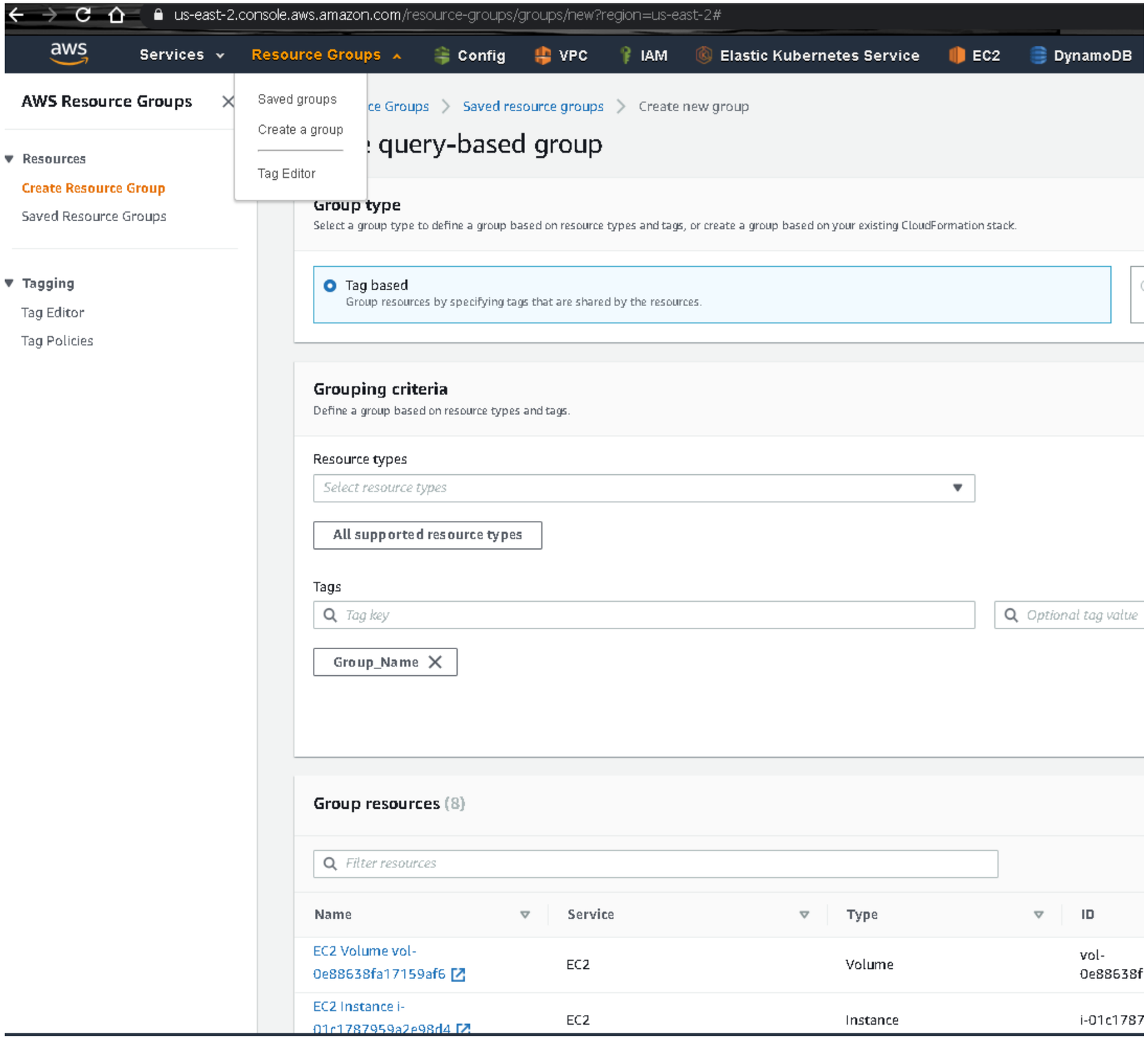
Tag Management / Resource Groups



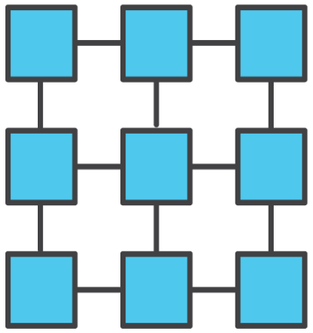
- Standardized, case-sensitive format for tags
- Implement automated tools to help manage resource tags
- Favor using too many tags rather than too few
- Remember, it’s easy to modify tags
- Examples: App Version, ENV, DNS Name, App Stack Identifier

Helps you to understand what your resources are doing and their cost impact.

Resource Groups



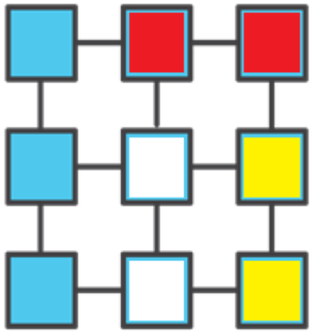
Placement Group Types



Cluster Placement Groups

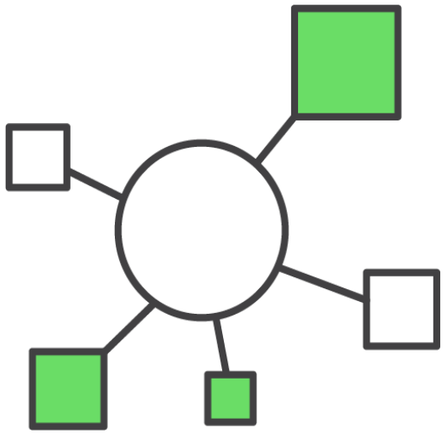
Does your compute layer require the lowest latency and highest packet-per-second network performance possible?
Running on the same AZ.

Tip: Add new instance? Up to capacity of server no for initial requirement.
Can be merged? No. Its only for initial requirement-



Partition Placement Groups

Have you run on spread deployment on distributed workloads?
Running in logical servers groups called Partitions which resided on several racks depends on partitions.



Spread Placement Groups

Do you have applications that have a small number of critical instances that should be kept separate from each other? Can be different AZs.