

Optimising with EC2 Placement Groups:

## 3 Types of Placement Groups



Cluster



Spread



Partition

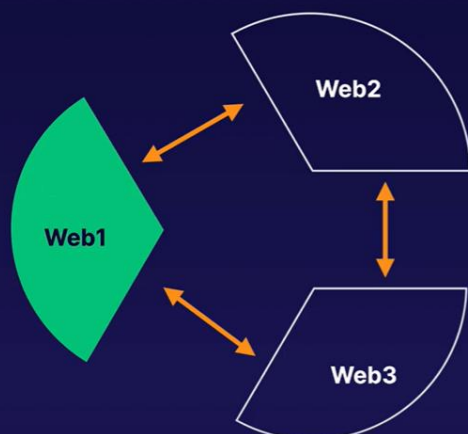
## Cluster Placement Groups

**Grouping of instances within a single Availability Zone.** Recommended for applications that need low network latency, high network throughput, or both.



• **Fact:**  
Only certain instance types can be launched into a cluster placement group.

## Spread Placement Groups



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.

# Partition Placement Groups

Each partition placement group has its own set of racks. Each rack has its own network and power source. No two partitions within a placement group share the same racks, allowing you to isolate the impact of hardware failure within your application.

EC2 DIVIDES EACH GROUP INTO LOGICAL SEGMENTS CALLED PARTITIONS.



## 3 Types of Placement Groups



### Cluster Placement Groups

Low network latency, high network throughput



### Spread Placement Groups

Individual critical EC2 instances



### Partition Placement Groups

Multiple EC2 instances; HDFS, HBase, and Cassandra

### Exam Tips

 A CLOUD GURU



A **cluster placement group** can't span multiple Availability Zones, whereas a spread placement group and partition placement group can.



Only **certain types of instances** can be launched in a placement group (compute optimized, GPU, memory optimized, storage optimized).



**AWS recommends homogenous instances** within cluster placement groups.



**You can't merge placement groups.**



You can **move an existing instance into a placement group**. Before you move the instance, the instance must be in the stopped state. You can move or remove an instance using the AWS CLI or an AWS SDK, but you can't do it via the console yet.

Solving Licensing Issues with Dedicated Hosts:

## DEDICATED HOSTS: EXAM TIPS

# Any question that talks about special licensing requirements.

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, and SUSE Linux Enterprise Server.

Timing Workloads with Spot Instances:

## WHEN TO USE SPOT INSTANCES

# Stateless, fault-tolerant, or flexible applications

Applications such as big data, containerized workloads, CI/CD, high-performance computing (HPC), and other test and development workloads.

To use **Spot Instances**, you must first decide on your maximum Spot price. The instance will be provisioned so long as the Spot price is **BELOW** your maximum Spot price.



The **hourly Spot price** varies depending on capacity and region.



If the Spot price goes above your maximum, you have **2 minutes** to choose whether to stop or terminate your instance.





You may also use a **Spot block** to stop your Spot Instances from being terminated even if the Spot price goes over your max Spot price. You can set Spot blocks for between **1 to 6 hours** currently.

## Spot Instances are useful for the following tasks:



Big data and analytics



Containerized workloads



CI/CD and testing



Image and media rendering



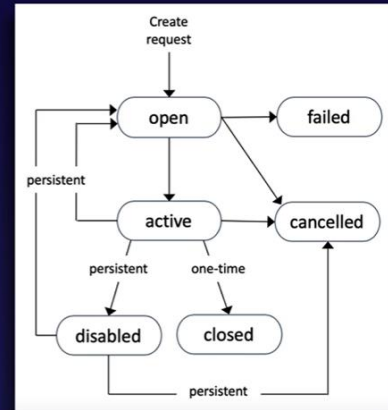
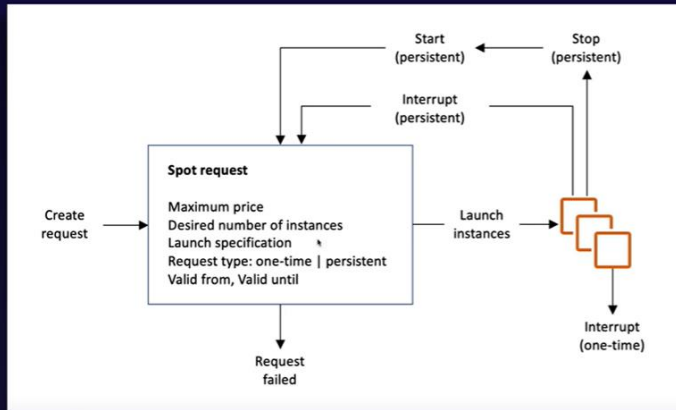
High-performance computing

## Spot Instances are **not** good for:

- ✓ Persistent workloads
- ✓ Critical jobs
- ✓ Databases



## Terminating Spot Instances

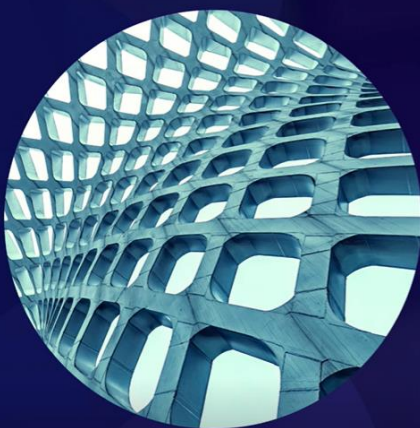


## Spot Fleets

**A Spot Fleet is a collection of Spot Instances and (optionally) On-Demand Instances.**

The **Spot Fleet** attempts to launch the number of Spot instances and On-Demand instances to meet the target capacity you specified in the Spot Fleet request. The request for Spot Instances is fulfilled if there is available capacity and the **maximum price you specified in the request exceeds the current Spot price**. The Spot Fleet also attempts to maintain its target capacity fleet if your Spot Instances are interrupted.

**Spot Fleets will try and match the target capacity with your price restraints.**



- 1 Set up different launch pools. Define things like **EC2** instance type, operating system, and Availability Zone.
- 2 You can have **multiple** pools, and the fleet will choose the best way to implement depending on the strategy you define.
- 3 Spot fleets will **stop launching instances** once you reach your price threshold or capacity desire.



Spot Instances save up to **90%** of the cost of On-Demand instances.



Useful for any type of computing where you don't need **persistent storage**.



You can block Spot Instances from terminating by using a **Spot block**.



A Spot Fleet is a collection of Spot Instances and (optionally) On-Demand instances.

