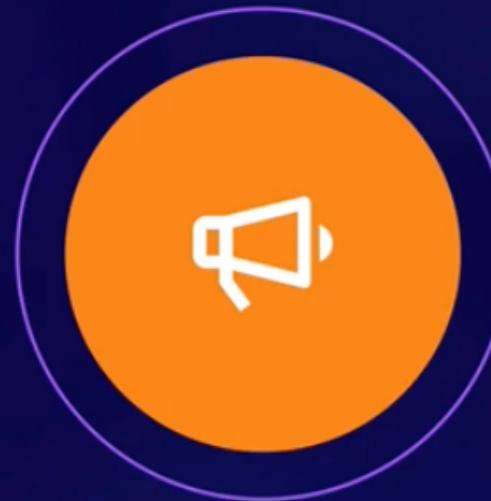


EC2 Pricing Options



On-Demand

Pay by the hour or the second, depending on the type of instance you run.



Reserved

Reserved capacity for 1 or 3 years. Up to 72% discount on the hourly charge.



Spot

Purchase unused capacity at a discount of up to 90%. Prices fluctuate with supply and demand.



Dedicated

A physical EC2 server dedicated for your use. The most expensive option.

Reserved Instances



Predictable Usage

Applications with steady state or predictable usage.



Standard RIs

Up to 72% off the on-demand price.



Specific Capacity Requirements

Applications that require reserved capacity.



Convertible RIs

Up to 54% off the on-demand price. Has the option to change to a different RI type of equal or greater value.



Pay up Front

You can make upfront payments to reduce the total computing costs even further.



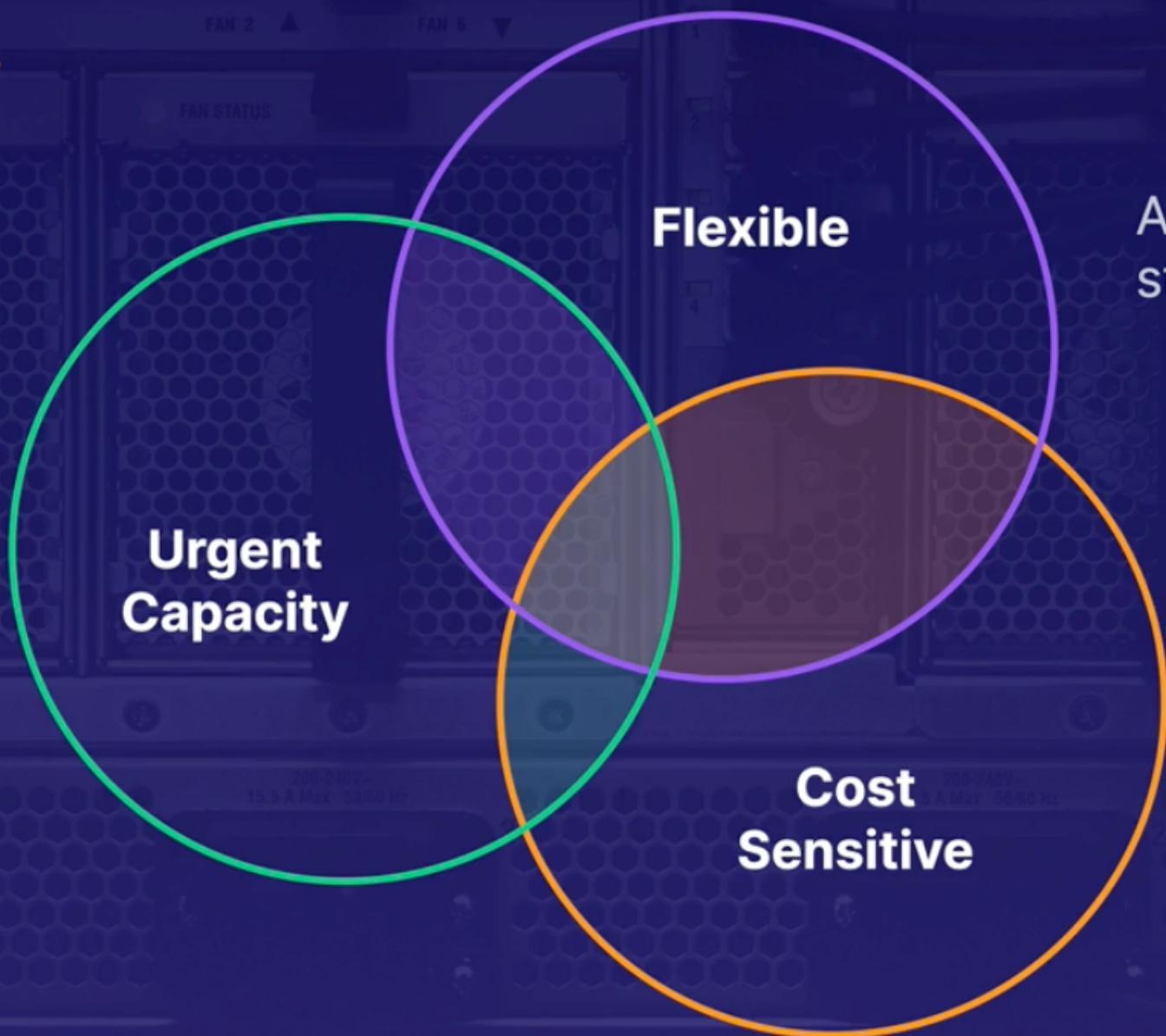
Scheduled RIs

Launch within the time window you define. Match your capacity reservation to a predictable recurring schedule that only requires a fraction of a day, week, or month.

When to Use Spot Instances

Image rendering.
Genomic sequencing.
Algorithmic trading engines.

Users with an urgent need for large amounts of additional computing capacity.



Applications that have flexible start and end times.

Applications that are only feasible at very low compute prices.

Dedicated Hosts



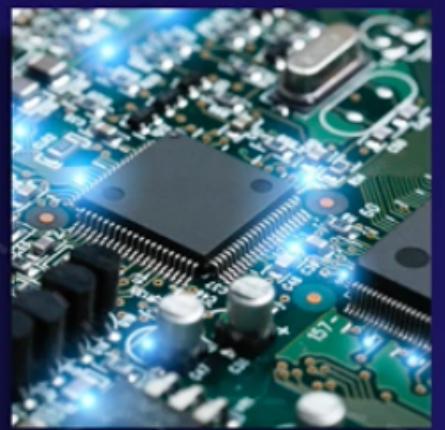
Compliance

Regulatory requirements that may not support multi-tenant virtualization.



On-Demand

Can be purchased on-demand (hourly).



Licensing

Great for licensing that does not support multi-tenancy or cloud deployments.



Reserved

Can be purchased as a reservation for up to 70% off the on-demand price.

Security Groups

Security Groups

Security groups are **virtual firewalls for your EC2 instance**. By default, everything is blocked.

TO LET EVERYTHING IN: 0.0.0.0/0



In order to be able to **communicate to your EC2 instances via SSH/RDP/HTTP**, you will need to **open up the correct ports**.

Bootstrap Scripts

A script that runs when the instance first runs

```
#!/bin/bash  
yum install httpd -y  
#installs apache  
yum service httpd start  
#starts apache
```

Adding these tasks at boot time
adds to the amount of time it takes to boot the instance.
However, it allows you to **automate the installation** of applications.

Retrieving Metadata

```
curl http://169.254.169.254/latest/meta-data/local-ipv4
[ec2-user@ip-172-31-31-0 ~]$ curl http://169.254.169.254/latest/meta-
data/
ami-id
ami-launch-index
ami-manifest-path
block-device-mapping/
events/
 hibernation/
hostname
identity-credentials/
instance-action
instance-id
instance-life-cycle
instance-type
local-hostname
local-ipv4
mac
managed-ssh-keys/
metrics/
network/
placement/
profile
public-hostname
public-ipv4
public-keys/
reservation-id
security-groups
```

Retrieving Metadata

Using the `curl` command, we can query metadata about our EC2 instance.

For different scenarios on the exam, choose the correct networking device.

1 ENI

For basic networking. Perhaps you need a separate management network from your production network or a separate logging network, and you need to do this at a low cost. In this scenario, use multiple ENIs for each network.

3 EFA

For when you need to accelerate High Performance Computing (HPC) and machine learning applications or if you need to do an OS-bypass. If you see a scenario question mentioning HPC or ML and asking what network adapter you want, choose EFA.

2 Enhanced Networking

For when you need speeds between 10 Gbps and 100 Gbps. Anywhere you need reliable, high throughput.



3 Types of Placement Groups



Cluster



Spread



Partition

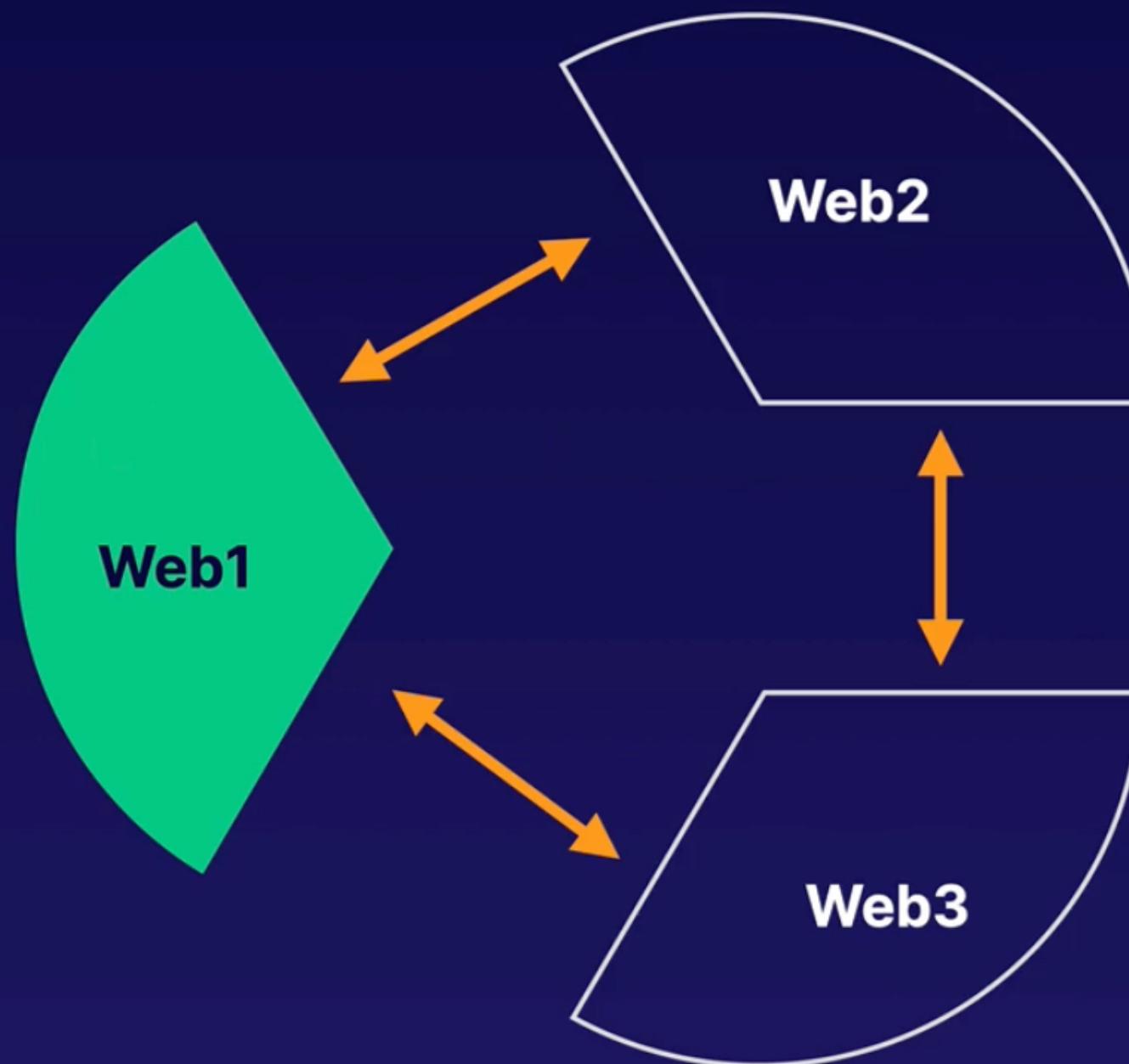
Cluster Placement Groups

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.



Partition 1



Partition 2



Partition 3

Exam Tips

-  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.

Spot Prices

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.

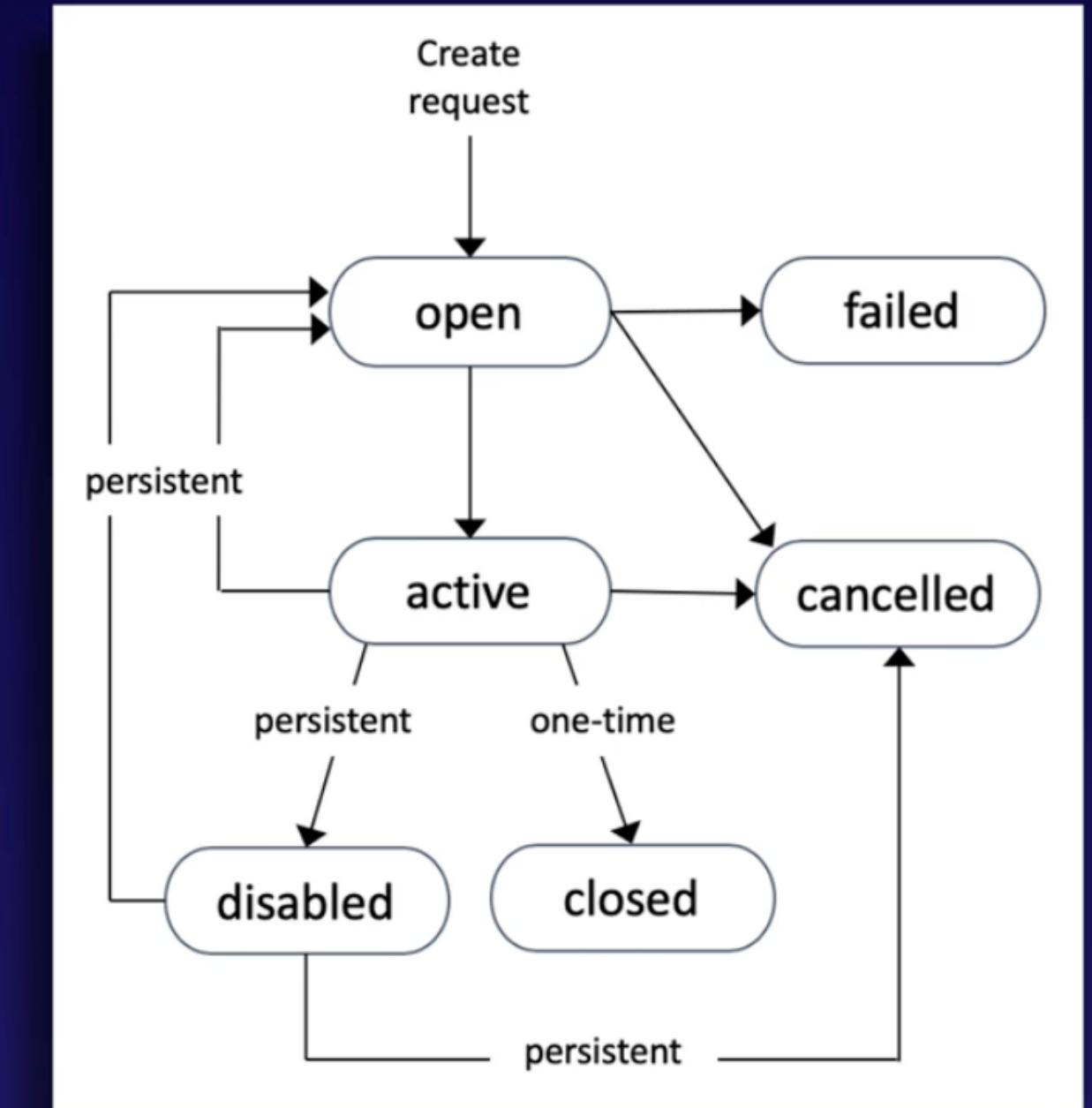
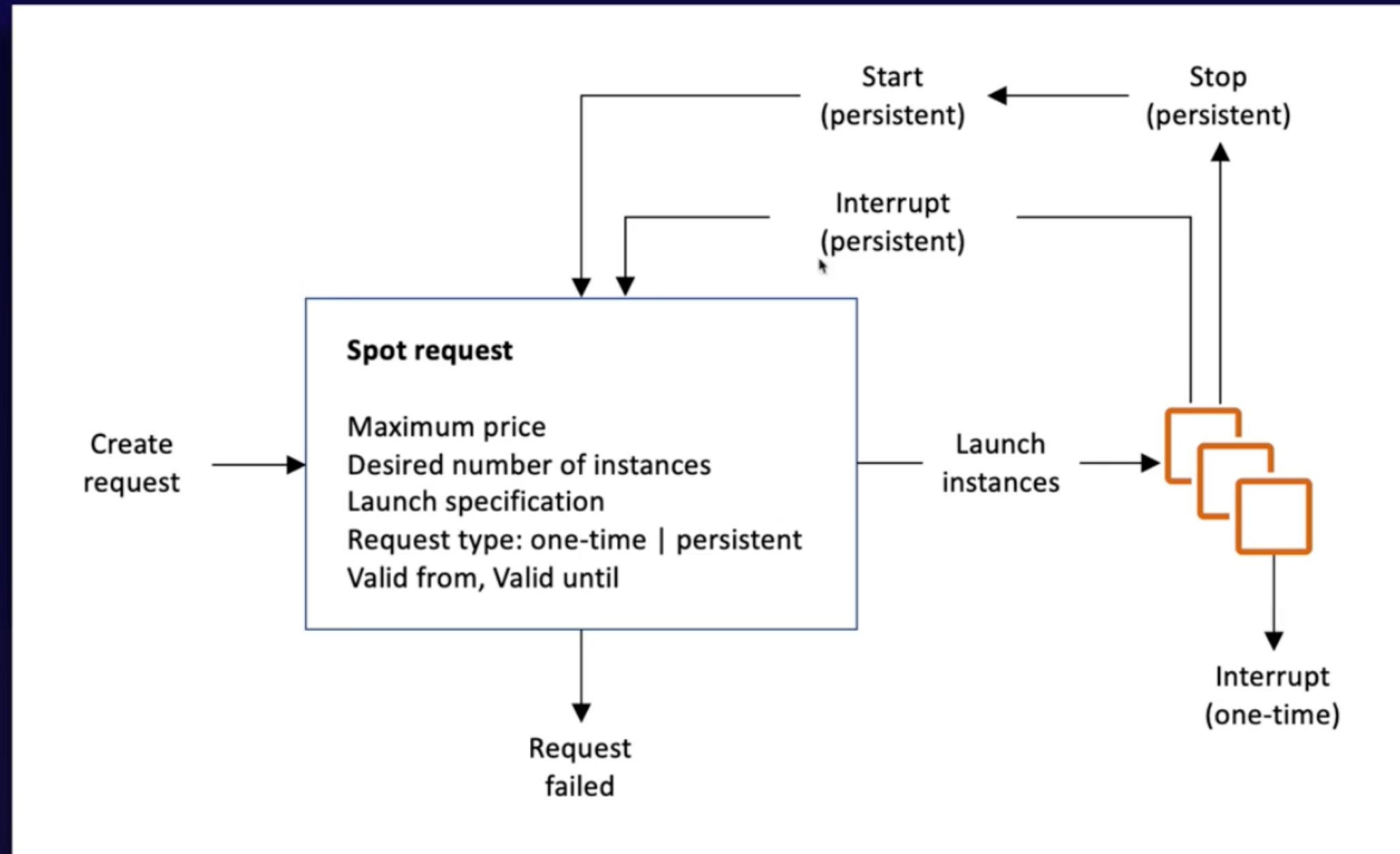


The **hourly Spot price** varies depending on capacity and region.



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

Terminating Spot Instances



What Are Spot Fleets?

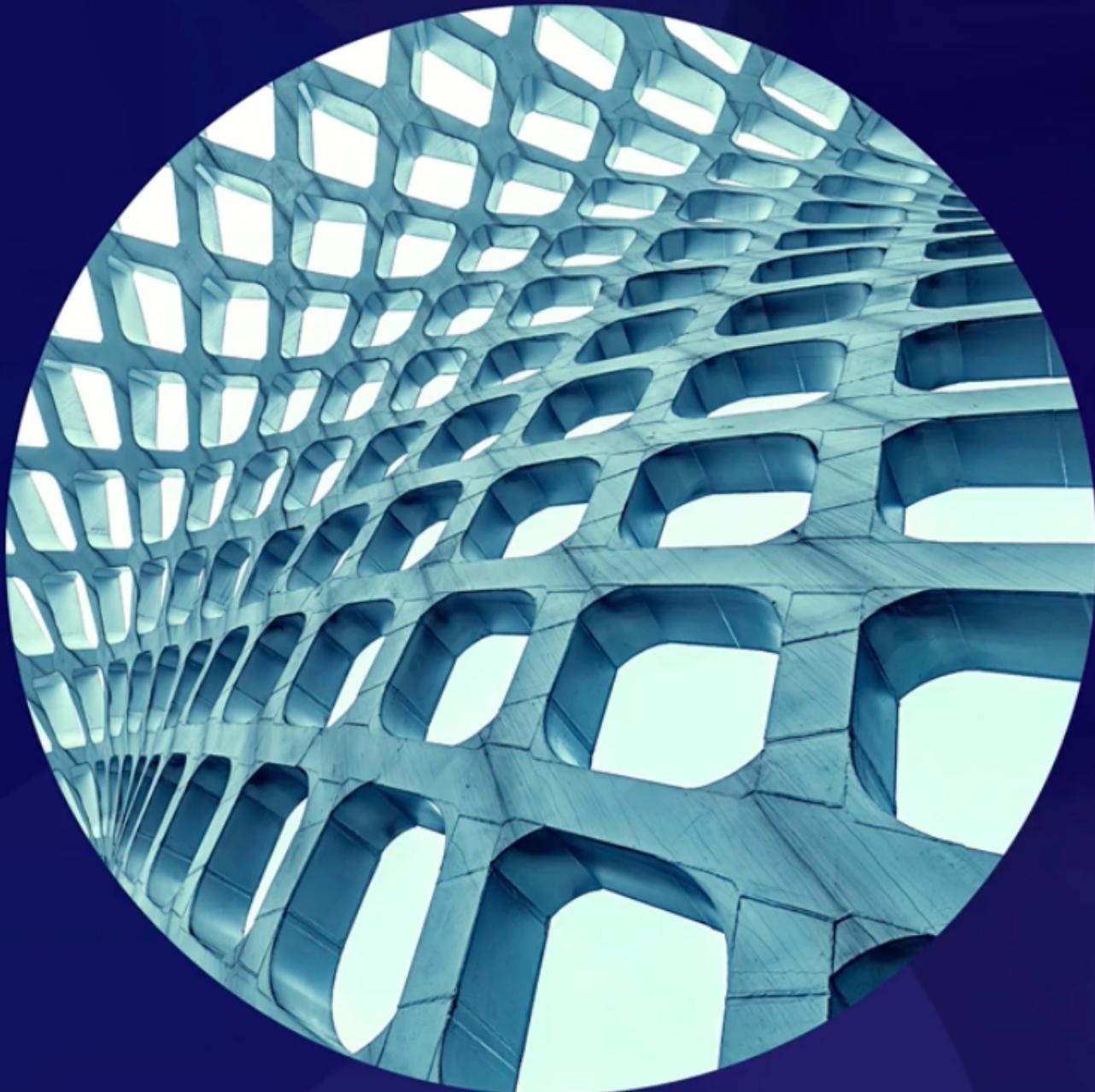
Spot Fleets

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.

Launch Pools

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.

Strategies

You can have the following strategies with Spot Fleets.

✓ **capacityOptimized**

The Spot Instances come from the pool with optimal capacity for the number of instances launching.

✓ **lowestPrice**

The Spot Instances come from the pool with the lowest price. This is the default strategy.

✓ **diversified**

The Spot Instances are distributed across all pools.

✓ **InstancePoolsToUseCount**

The Spot Instances are distributed across the number of Spot Instance pools you specify. This parameter is valid only when used in combination with `lowestPrice`.