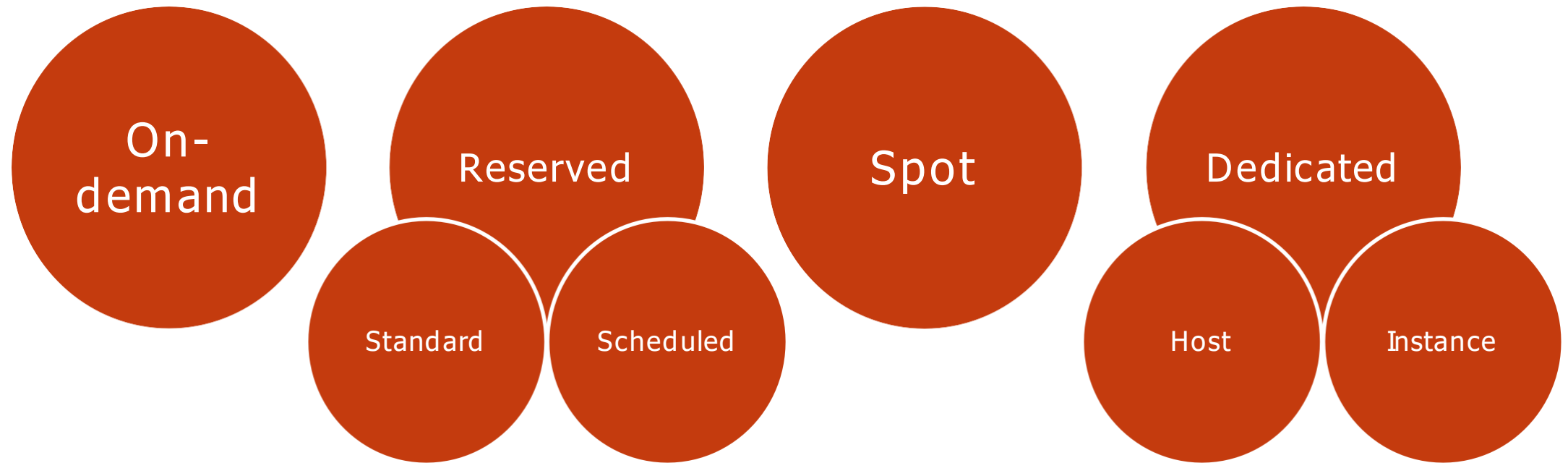


Understanding and Using Elastic Cloud Compute (EC2)



EC2 Instance Types





On-demand:

- Low cost and flexibility with no up front cost
- Ideal for auto scaling groups and unpredictable workloads
- Dev/test

Reserved Instances:

- Steady state and predictable usage
- Applications that need reserved capacity
- Upfront payments reduce hourly rate
- Scheduled Ris match your capacity reservation to a predictable recurring schedule



Spot:

- Flexible start and end times
- Grid computing and HPC
- Very low hourly compute cost

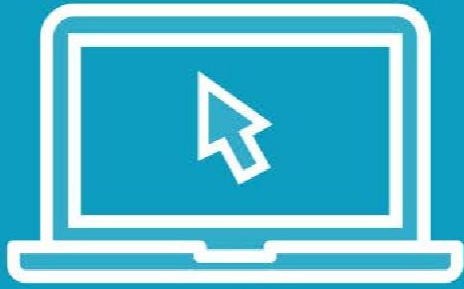
Dedicated:

- Predictable performance
- Complete isolation
- Most expensive

EC2 Instance Family

Family	Generation	Smallest	Largest
General Purpose (GP)	t2	t2.micro	t2.2xlarge
	m4	m4.large	m4.16xlarge
	m3	m3.medium	m3.2xlarge
Compute Optimized	c4	c4.large	c4.8xlarge
	c3	c3.large	c3.8xlarge
Memory Optimized	r3	r3.large	r3.8xlarge
	r4	r4.large	r4.16xlarge
	x1	x1.16xlarge	x1.32xlarge
Storage Optimized	i2	i2.xlarge	i2.8xlarge
	d2	d2.2xlarge	d2.8xlarge
Accelerated Computing	g2	g2.2xlarge	g2.8xlarge
	p2	p2.xlarge	P2.16xlarge

Demo



Create EC2 instances

Standard Reserved Instances Attributes

All Upfront
(up to 68% over a 3 year term)

Partial Upfront
(up to 60% over a 3 year term)

No Upfront
(up to 30% over a 1 year term)

Requires a commitment
(1 or 3 years)

Scheduled Reserved Instances Attributes

Accrue charges hourly, billed
in monthly increments over
the term

Have a 1 year term
commitment

Modifying Your RIs

Switch Availability Zones within the same region

Change the instance size within the same instance type

Instance type modifications are supported only for Linux. Due to licensing differences, Linux RIs cannot be modified to RedHat or SUSE

You cannot change the instance size of Windows Reserved Instances

Instance size	Normalization factor
micro	0.5
small	1
medium	2
large	4
xlarge	8
2xlarge	16
4xlarge	32
8xlarge	64
10xlarge	80

Example 1

Quantity of 2 xlarge instances = $2 \times 8 = 16$

Can be changed to:

- 4 large instances: $16/4=4$
- 8 medium instances: $16/2=8$
- 16 small instances: $16/1=16$

Example 2

Quantity of 8 small instances = $8 \times 1 = 8$

Can be changed to:

- 2 large instances: $8/4=2$
- 1xlarge instance: $8/1=8$

Modification Requests



The diagram consists of three identical rectangular boxes arranged horizontally. Each box is divided into two horizontal sections. The top section of each box is a solid light gray rectangle. The bottom section is white and contains text. The first box on the left has a green border and the text 'Console'. The middle box has a maroon border and the text 'interface'. The third box on the right has a blue border and the text 'Amazon EC2 API'.

Console

interface

Amazon EC2 API

High Performance Computing (HPC)



HPC used by oil & gas, pharmaceuticals, research, automotive, and other industries

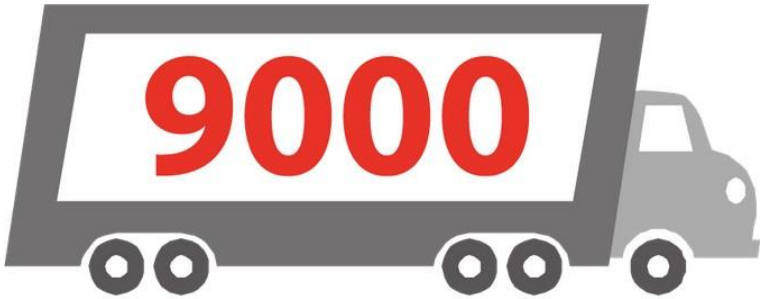
Batch processing of compute intensive workloads

Requires high performance CPU, network, and storage

Jumbo Frames are typically required

- HPC workloads typically need access to a shared filesystem, and will use a lot of disk I/O

Jumbo Frames



Help significantly because they can carry up to 9000 bytes of data

Supported on AWS through enhanced networking

- Enhanced networking is enabled through single root I/O virtualization (SR-IOV) on supported instances
- Enhanced networking is only supported on Hardware Virtualization (HVM) instances. Not supported on Paravirtualized (PV) instances

Jumbo Frames

Enabling Enhanced Networking on Linux Instances in a VPC:

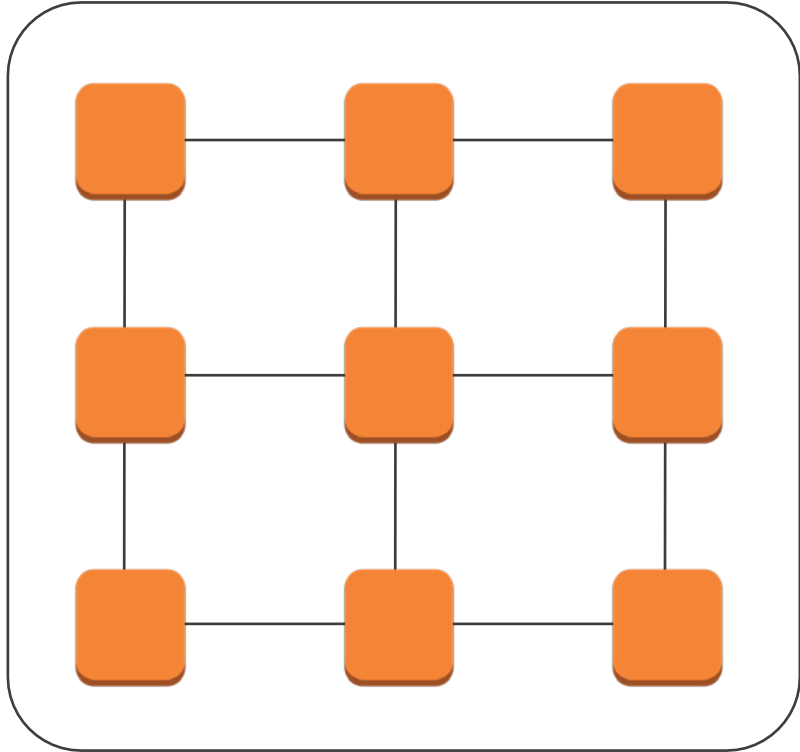
<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/enhanced-networking.html>

Enabling Enhanced Networking on Windows Instances in a VPC:

<http://docs.aws.amazon.com/AWSEC2/latest/WindowsGuide/enhanced-networking.html>



Placement Groups



A logical grouping of instances in a single availability zone (AZ)

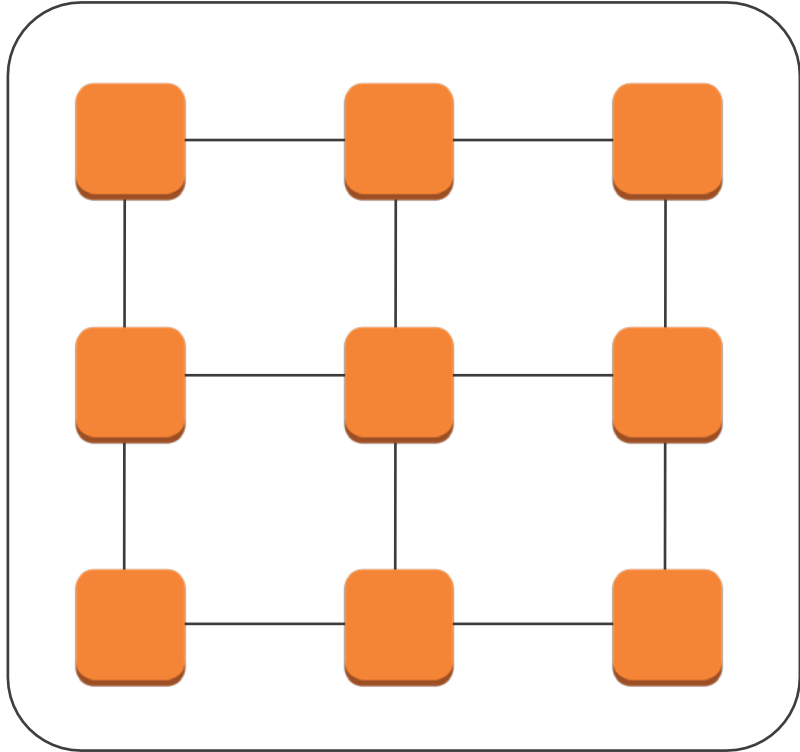
Can't span multiple availability zones

Name must be unique across AWS account

Recommended for application that benefit from low latency, high bandwidth or both

Only supported instances that support enhanced networking can be launched into a placement group (C3, C4, D2, I2, M4, E3)

Placement Groups



Existing instances cannot be moved into a placement group

Placement groups cannot be merged

Can span peered VPCs but you will not get full-bisection bandwidth between instances

Reserved instances are supported on an instance level but you cannot explicitly reserved capacity for a placement group

Placement Groups

Placement Groups and supported instances:

<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/placement-groups.html>



Summary



EC2 instance types

EC2 reserved instances

High performance computing

Placement groups