



Introduction to Amazon Cloud & EC2 Overview

Yasser Quraishi, AWS Solutions Architect –
yquraish@amazon.com

Jeff Hosley, AWS Account Manager – jhosley@amazon.com

Agenda

- Introduction to AWS Cloud
- Global Reach
- EC2 Overview
- EC2 Details
- Pop quiz
- Hands-on Lab

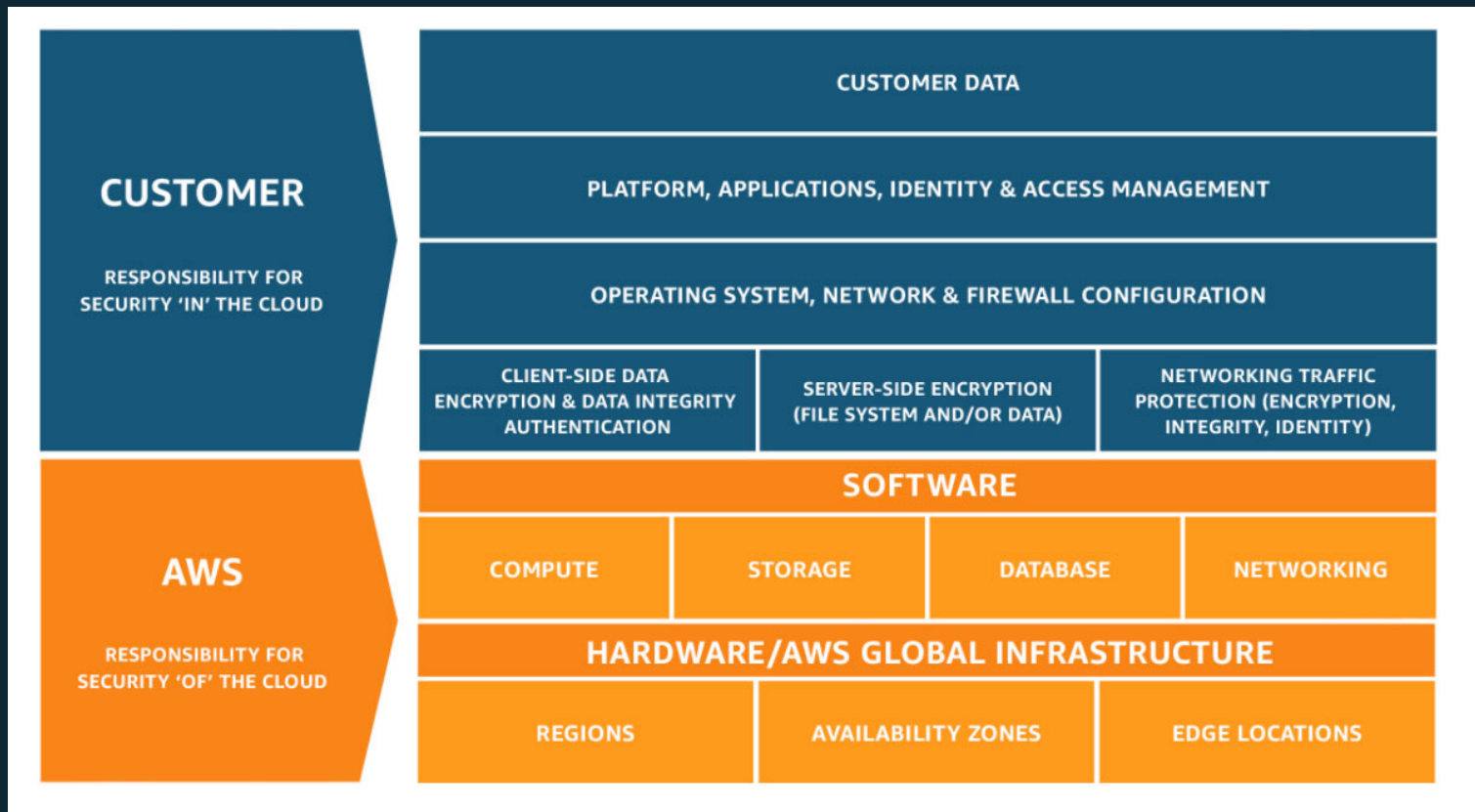
What is AWS?

AWS provides a highly reliable, scalable, low-cost infrastructure platform in the cloud that powers millions of businesses in 190 countries around the world.

Benefits

- Low Cost
- Elasticity & Agility
- Open & Flexible
- Secure
- Global Reach

Shared Responsibility Model





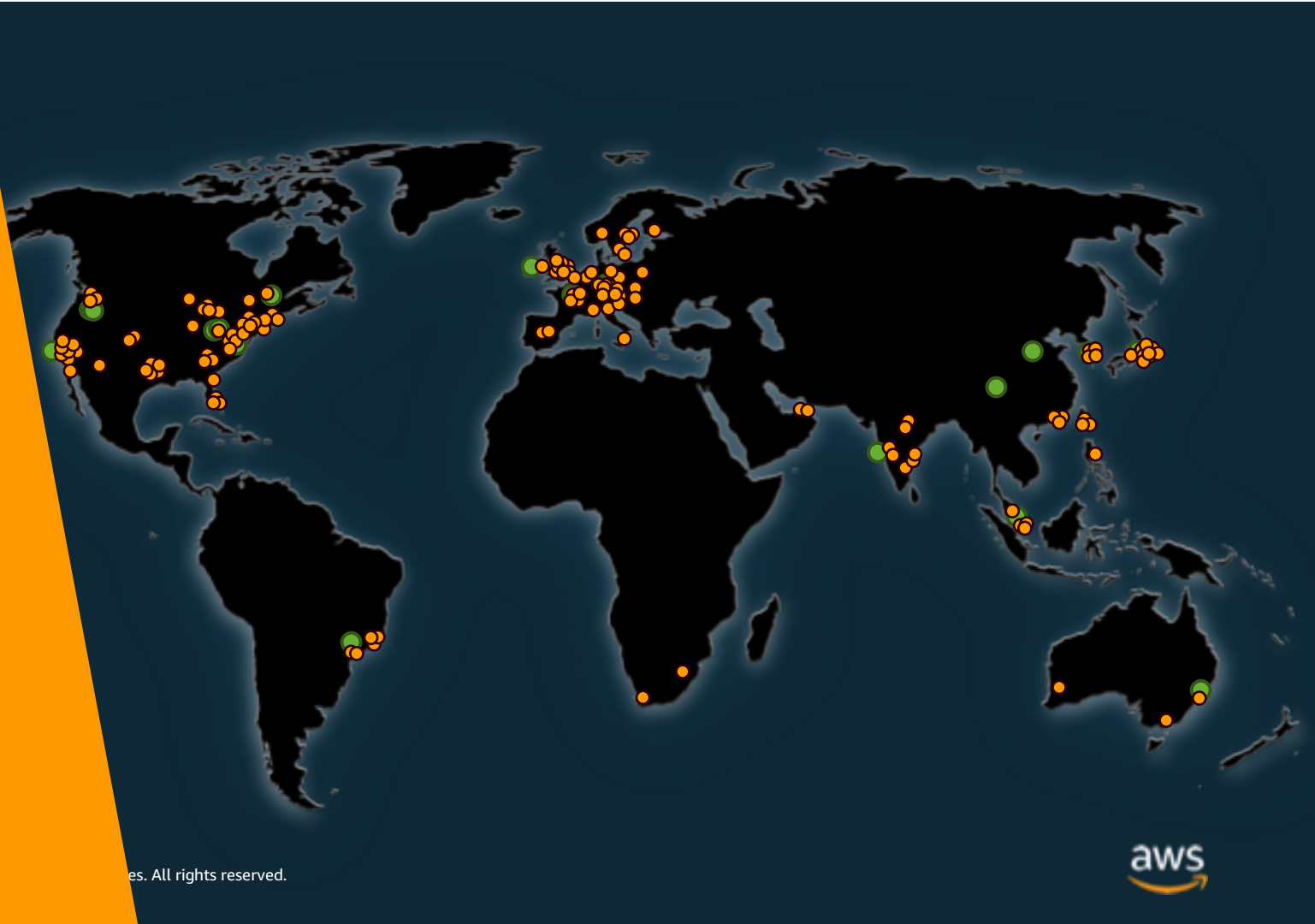
AWS Global Reach

22 Regions



180+

Amazon
CloudFront
Points of
Presence



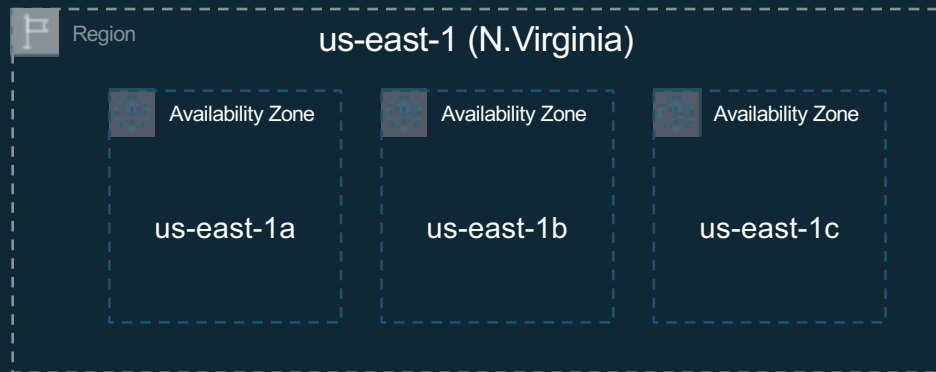
es. All rights reserved.



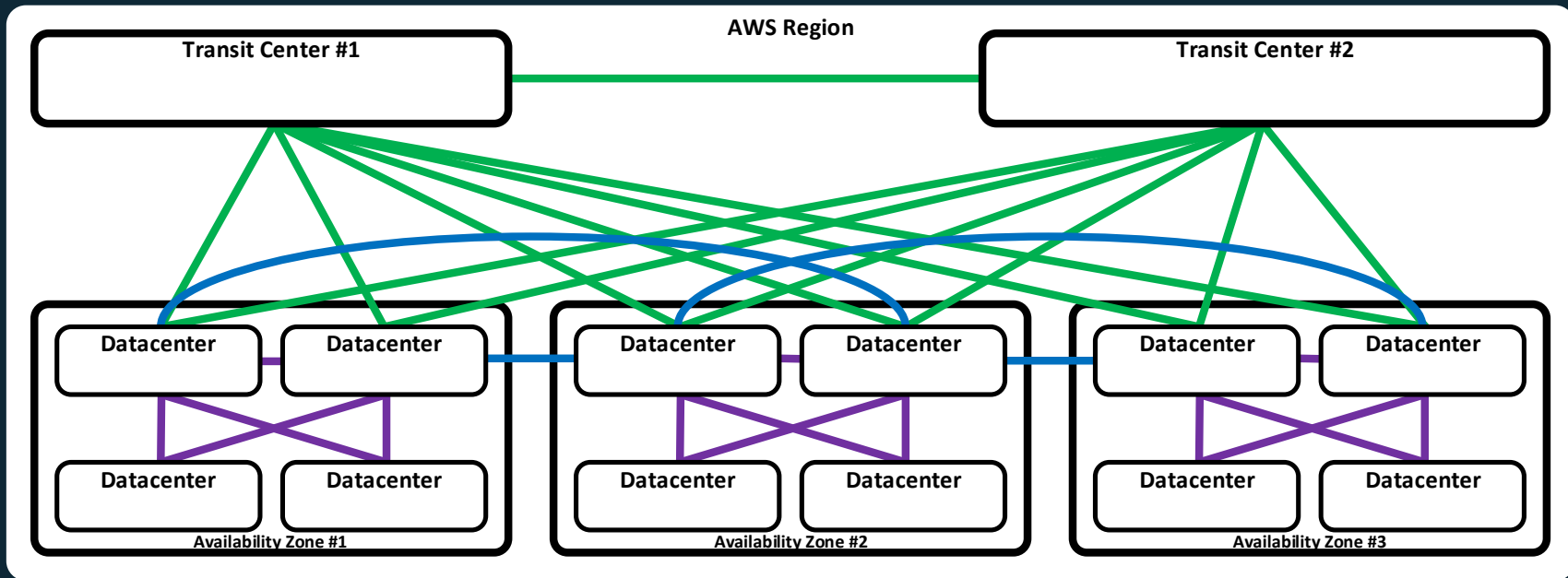


Availability Zones

- A Region is comprised of multiple Availability Zones
- Isolation from other Availability Zones (power, network, flood plains)
- Low latency (<10mS) direct connect between Availability Zones
- 1AZ can include multiple data centers
- Physical Separation < 100km



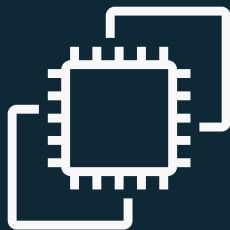
Availability Zones





EC2 Overview

Choices for Compute



Amazon EC2

Virtual server instances
in the cloud



Amazon ECS, EKS, and Fargate

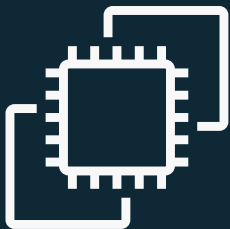
Container management
service for running
Docker on a managed
cluster of EC2



AWS Lambda

Serverless compute
for stateless code execution
in response to triggers

Amazon EC2



Amazon EC2

Linux | Windows

Arm and x86 architectures

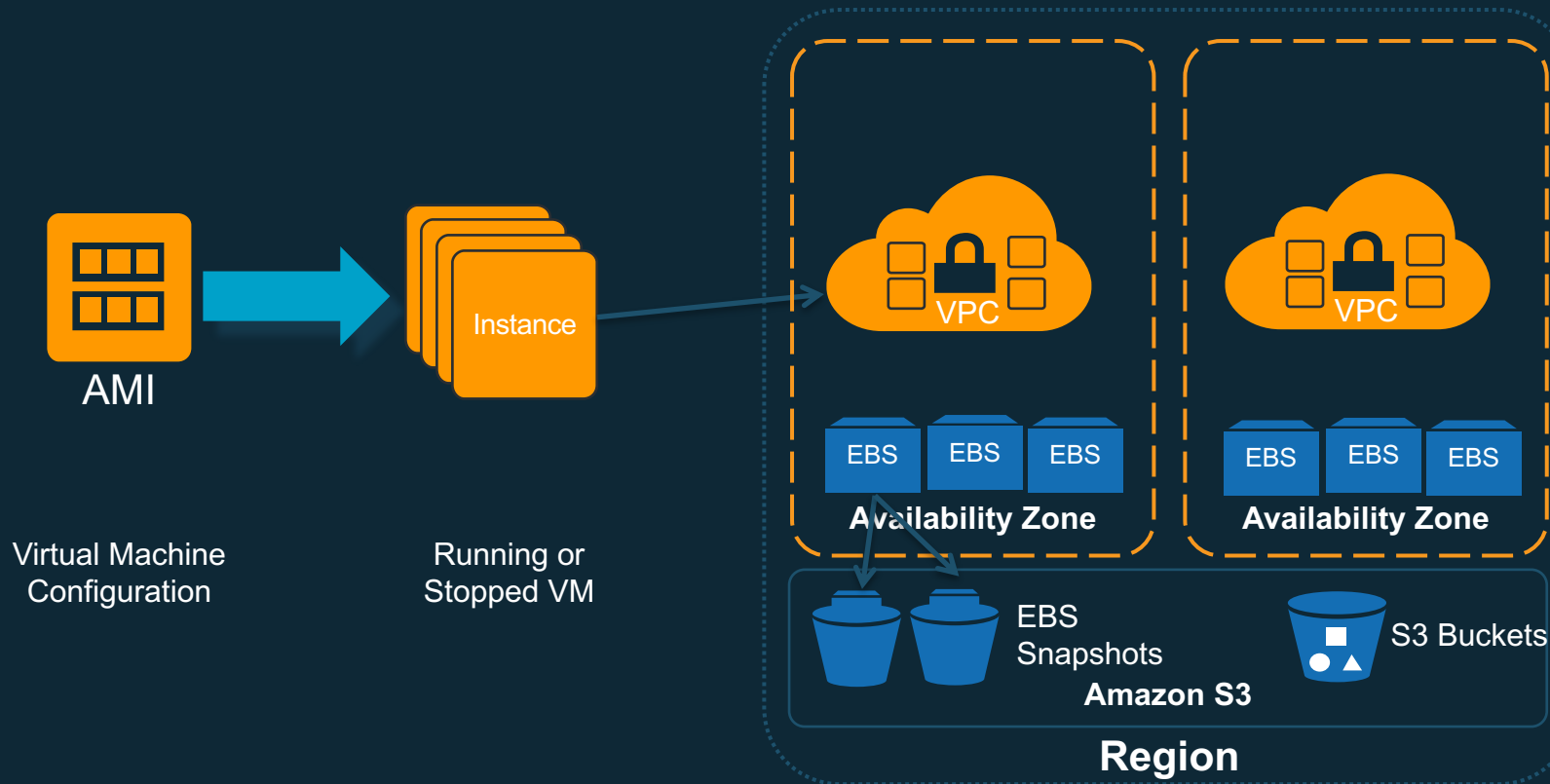
General purpose and workload optimized

Bare metal, disk, networking capabilities

Packaged | Custom | Community AMIs

Multiple purchase options: On-demand, RI, Spot

EC2 Terminology





Instance Types

Amazon
Lightsail

NEW

T3

M5

D2

H1

R5

R5
m

X1

X1e

I3

I3m

C5

G3

P3

F1

Z1d

Z1 dm

Virtual
Private
Servers

Burstable

General
Purpose

Dense
Storage

Big Data
Optimized

Memory Optimized

In-
memory

Memory
Intensive

High I/O

Bare
Metal
High I/O

Compute
Intensive

Graphics
Intensive

General
Purpose
GPU

FPGA

Compute and
Memory Intensive



EC2 Elastic GPUs

- Graphics acceleration for EC2 instances



EC2 Fleet

- Simplified provisioning
- Massive scale
- Flexible capacity allocation

EC2 Naming Explained



Instance generation

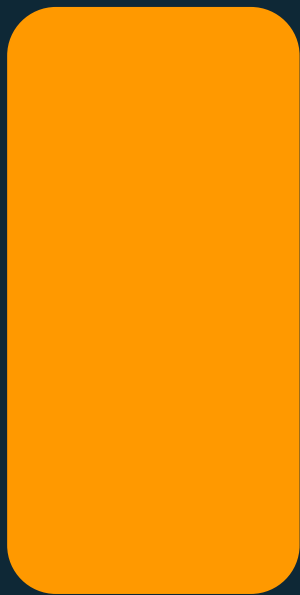
c5n.xlarge

Instance
family

Attribute

Instance size

Instance sizing



c4.8xlarge

≈



2 - c4.4xlarge

≈



4 - c4.2xlarge

≈



8 - c4.xlarge



EC2 Operating Systems Supported

Windows 2003R2/2008/2008R2/2012/2012R2/2016

Amazon Linux

Debian

Suse

CentOS

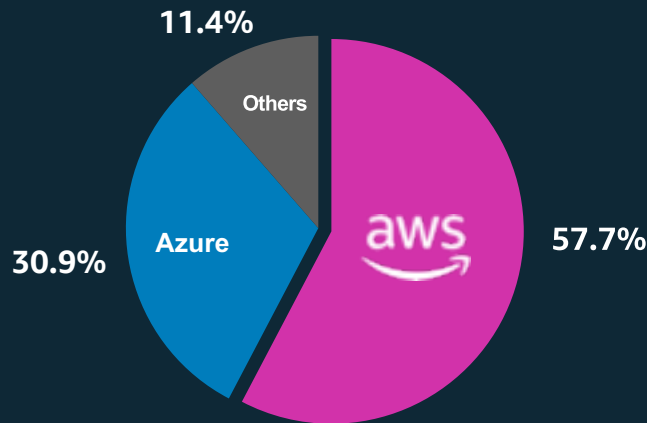
Red Hat Enterprise Linux

Ubuntu



for more OSes see: <https://aws.amazon.com/marketplace/b/2649367011>

Windows Licenses by Cloud Provider



Note: includes Windows instances deployed in the public cloud IaaS market during 2017 Source: IDC estimates 2018

IDC, Windows Server Operating Environment Market Update, Doc # US44217118, Aug 2018

https://d1.awsstatic.com/analyst-reports/IDC_Slide_WindowsonAWS_JM181015.pdf



Choose your processor and architecture



Intel® Xeon® Scalable
(Skylake) processor



NVIDIA V100 Tensor
Core GPUs



AMD EPYC processor



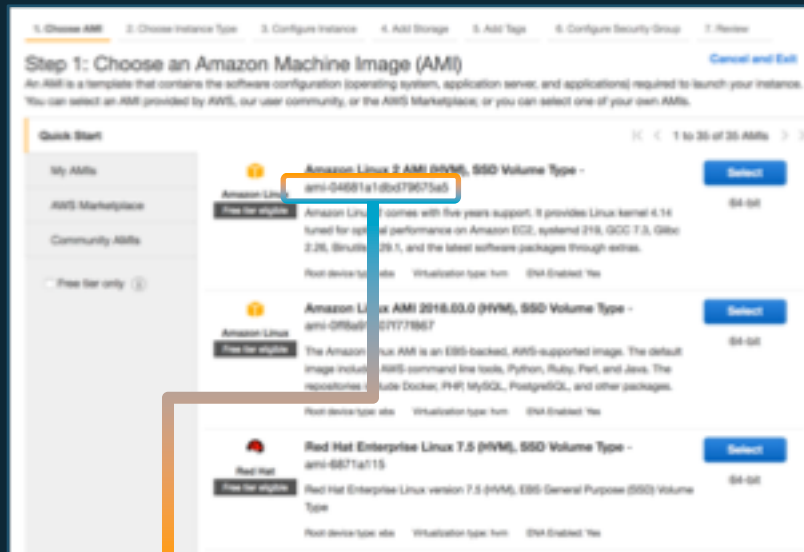
Amazon ARM based
Cloud Processor

Right compute for the right application and workload

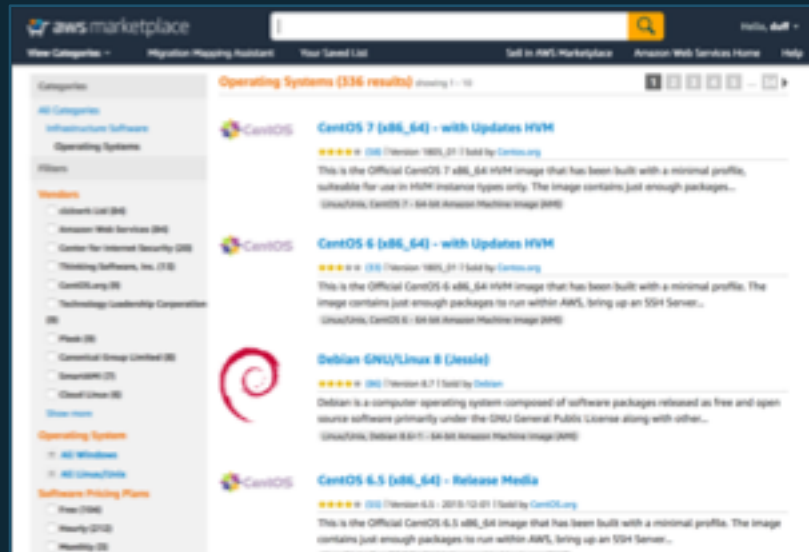
Choosing an AMI



AWS Console



AWS Marketplace



Use the AMI ID to launch through the API or AWS Command Line Interface (AWS CLI)

```
aws ec2 run-instances --image-id ami-04681a1dbd79675a5 --instance-type c4.xlarge --count 10 --key-name MyKey
```





Purchasing options at a glance

On-Demand Instances

Pay for compute capacity by the hour with no long-term commitments

For Spiky workloads or to define needs



Reserved Instances

Make a low, one-time payment and receive a significant discount on the hourly charge

For committed utilization



Spot Instances

Price fluctuates based on supply and demand

For time-insensitive or transient workloads



EC2 Options



Categories

General purpose
Burstable
Compute intensive
Memory intensive
Storage (High I/O)
Dense storage
GPU compute
Graphics intensive



Capabilities



Choice of processor
(AWS, Intel, AMD)

Fast processors
(up to 4.0 GHz)

High memory footprint
(up to 12 TiB)

Instance storage
(HDD and NVMe)

Accelerated computing
(GPUs and FPGA)



Networking
(up to 100 Gbps)

Bare Metal

Size
(Nano to 32xlarge)



Options

Elastic Block Store
Elastic Graphics
Elastic Inference

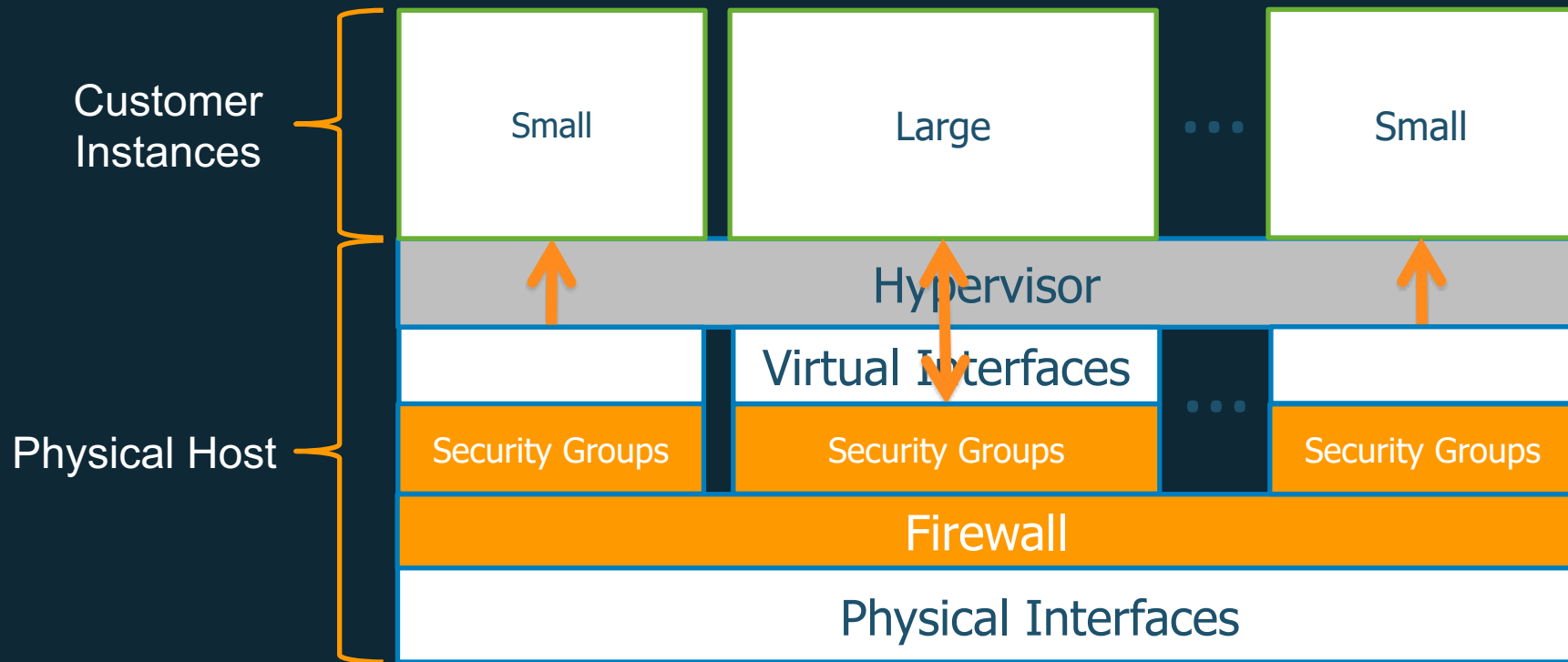


175
instance types
for virtually
every workload
and business need



EC2 Design

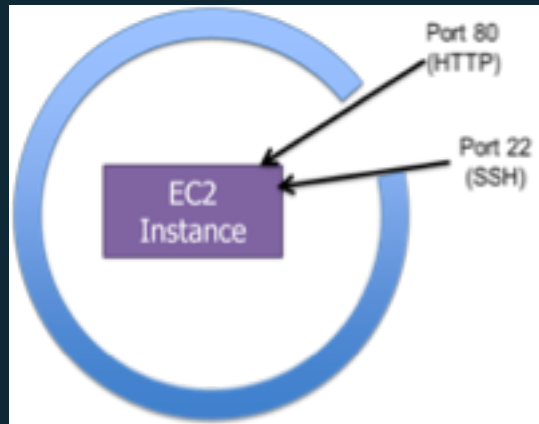
EC2 Host Virtualization



EC2 Security Groups

Security Group Rules

- Name
- Description
- Protocol
- Port range
- IP address, IP range, Security Group name



Tiered EC2 Security Groups

Hierarchical Security Group Rules

- Dynamically created rules
- Based on Security Group membership
- Create tiered network architectures



“Web” Security Group:

TCP 80 0.0.0.0/0

TCP 22 “Mgmt”

“App” Security Group:

TCP 8080 “Web”

TCP 22 “Mgmt”

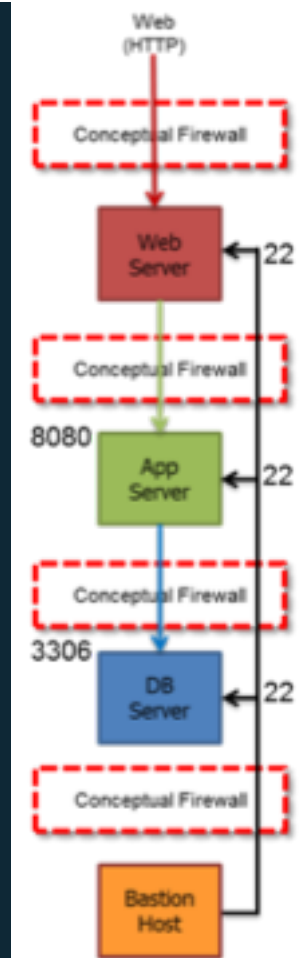
“DB” Security Group:

TCP 3306 “App”

TCP 22 “Mgmt”

“Mgmt” Security Group:

TCP 22 163.128.25.32/32




EC2-Specific Credentials

EC2 key pairs


- Linux – SSH key pair for first-time host login
- Windows – Retrieve Administrator password

Standard SSH RSA key pair

- Public/Private Keys
- Private keys are not stored by AWS



“Public Half” inserted by Amazon into each EC2 instance that you launch



“Private Half”
downloaded to your
desktop

Instance Metadata

<http://169.254.169.254/latest/meta-data/> contains a wealth of info

- ami-id
- ami-launch-index
- ami-manifest-path
- block-device-mapping/
- hostname
- instance-action
- ★ **instance-id**
- instance-type
- kernel-id
- local-hostname
- local-ipv4
- mac
- network/
- ★ **placement/availability-zone**
- profile
- public-hostname
- public-ipv4
- public-keys/

Pop Quiz

1. How many Regions does AWS have?
2. What are Availability Zones?
3. When should you use Reserved instances?
4. What is a Security Group?
5. What is a AMI?
6. What is a key pair?
7. How can an authorized user determine the IP address or hostname of an EC2 instance?

Any Questions?



Hands-on Lab #1:

<https://bit.ly/2Z13Ccn>