



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

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

What sets AWS apart?

-  Experience Building and managing cloud since 2006
-  Service Breadth & Depth 145+ services to support any cloud workload
-  Pace of Innovation History of rapid, customer-driven releases
-  Global Footprint 21 regions, 64 availability zones, 150+ edge locations
-  Pricing Philosophy 73 proactive price reductions to date
-  Ecosystem Thousands of consulting/system integrator & technology partners

Experience with Operational Reliability

Our goal is to make our operational performance indistinguishable from perfect. We are driven to remove any all causes of failure.

- We have spent over a decade building the world's most reliable, secure, scalable, and cost-effective infrastructure.
- Service SLAs between 99.9% and 100% availability. Amazon S3 is designed for 99.999999999% durability.
- Availability Zones exist on isolated fault lines, flood plains, and electrical grids to substantially reduce the chance of simultaneous failure.
- The AWS Service Health Dashboard provides 24/7 visibility in the real-time operational status of all services around the globe.

Pricing Philosophy

High volume / low margin businesses are in our core DNA

Trade CapEX for
variable expense

Our economies of
scale provide us
with lower costs

Pricing model
choice to support
variable and
stable workloads

Save more money
as you grow bigger

Pay for what
you use

73 price
reductions
since 2006

On-demand
Reserved Instances
Spot

Tiered pricing
Volume discounts
Custom pricing

Customer obsessed



90%
of roadmap originates with customer requests
and are designed to meet specific needs



"Performance, reliability, and responsiveness are fundamental to our customer experience, and T3 instances help us to deliver on that customer promise while also controlling our costs."

—Heroku

AWS positioned as a leader in the Gartner Magic Quadrant for Cloud Infrastructure as a Service

AWS is positioned highest in execution and furthest in vision within the Leaders Quadrant



Gartner, Magic Quadrant for Cloud Infrastructure as a Service, Worldwide, 2018, Dennis Lewis, Lydia Rataj, May 2018. G00838148
This graphic was published by Gartner, Inc. as part of a larger research document and should be evaluated in the context of the entire document. The Gartner document is available upon request from AWS. ©2018 Gartner, Inc. or its affiliates. All rights reserved. Gartner is a registered trademark and does not advise technology users to select only those vendors with the highest ratings or other designations. Gartner research publications consist of the opinions of Gartner's research organizations and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

1

AWS Global Reach

21
Regions



150+

Amazon
CloudFront
Points of
Presence



89

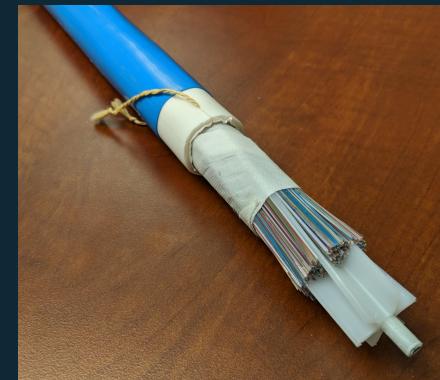
AWS Direct
Connect
locations





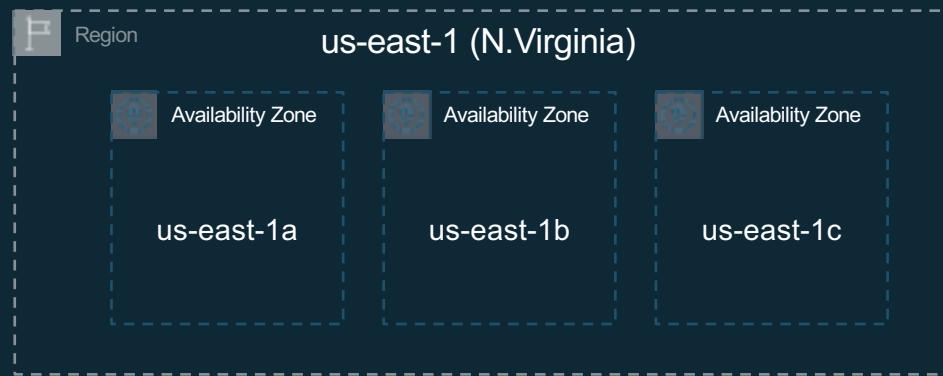
Intra & inter-AZ connectivity

- Dark fiber “spans”
 - Optimized for low-latency & physical diversity
- Amazon controlled infrastructure
- Geospatial coordinates
- Dense wavelength division multiplexing (DWDM)

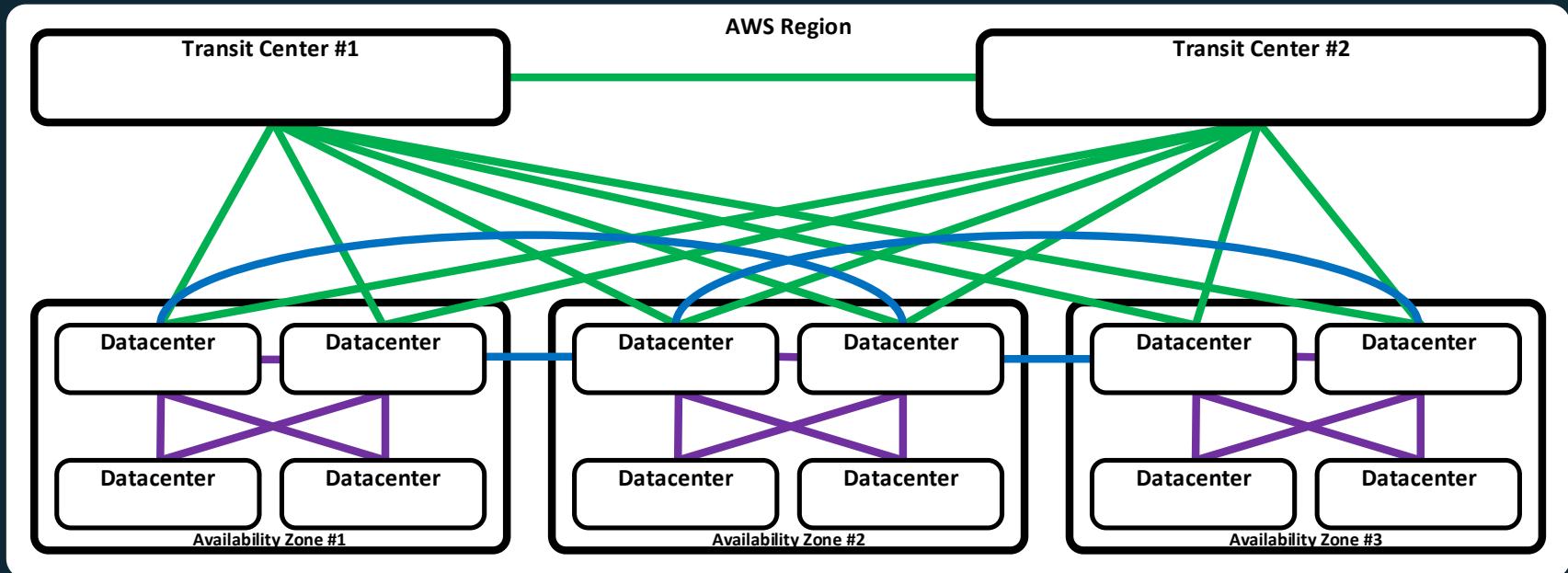


Availability Zones

- 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

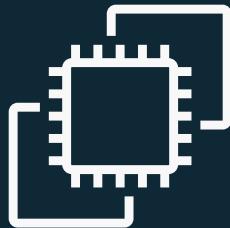




EC2 Overview



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



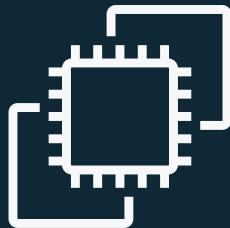
AWS Lambda

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

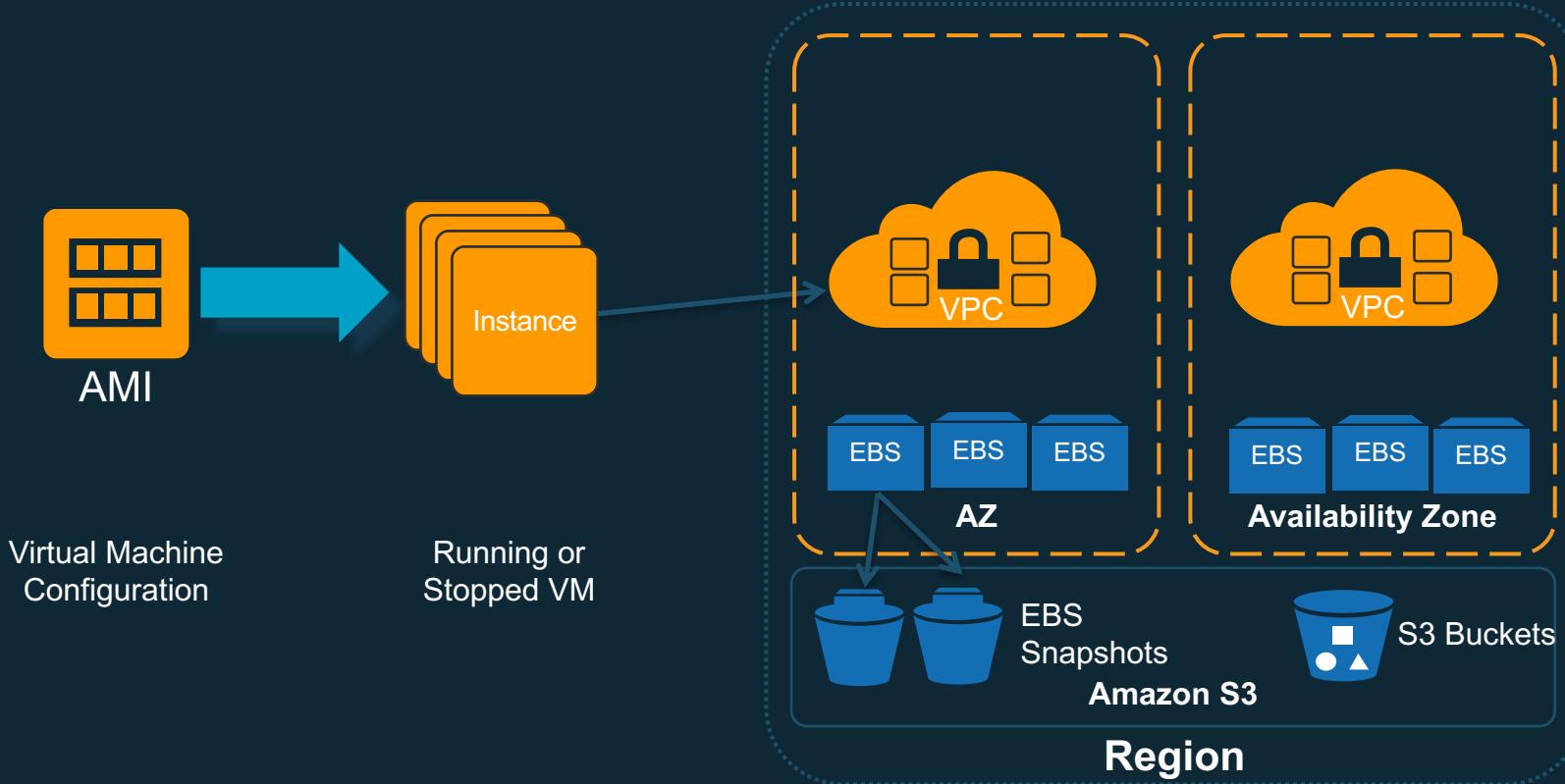
Bare metal, disk, networking capabilities

Packaged | Custom | Community AMIs

Multiple purchase options: On-demand, RI, Spot



EC2 Terminology





What's a virtual CPU? (vCPU)

- A vCPU is typically a hyper-threaded physical core*
- Divide vCPU count by 2 to get core count
- On Linux, "A" threads enumerated before "B" threads
- On Windows, threads are interleaved

- Cores by Amazon EC2 & RDS DB Instance type:
<https://aws.amazon.com/ec2/virtualcores/>



Memory and Storage

What's a GiB?

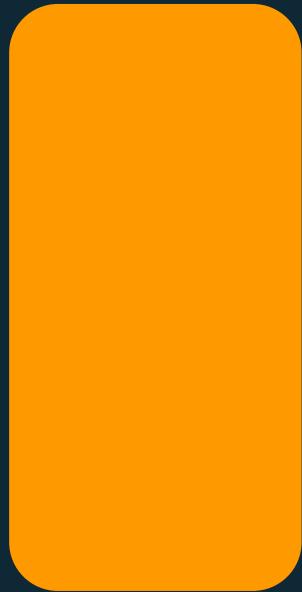
- Memory is presented as GibiBytes (GiB) and not Gigabytes (GB)
- $256 \text{ GiB} = 275 \text{ GB}$

What about storage?

- Storage is independent of compute
- You allocate drives known as EBS volumes
- Max 16 TiB per volume



Instance sizing



≈



≈



≈



c4.8xlarge

2 - c4.4xlarge

4 - c4.2xlarge

8 - c4.xlarge



Instance Types



EC2 Elastic GPUs

- Graphics acceleration for EC2 instances



EC2 Fleet

- Simplified provisioning
- Massive scale
- Flexible capacity allocation



EC2 Naming Explained

Instance generation

c5n.xlarge

The diagram illustrates the structure of an EC2 instance name. The prefix 'c' is labeled 'Instance family'. The digit '5' is labeled 'Attribute'. The suffix 'xlarge' is labeled 'Instance size'.



EC2 Operating Systems Supported

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

Amazon Linux

Debian

Suse

CentOS

Red Hat Enterprise Linux

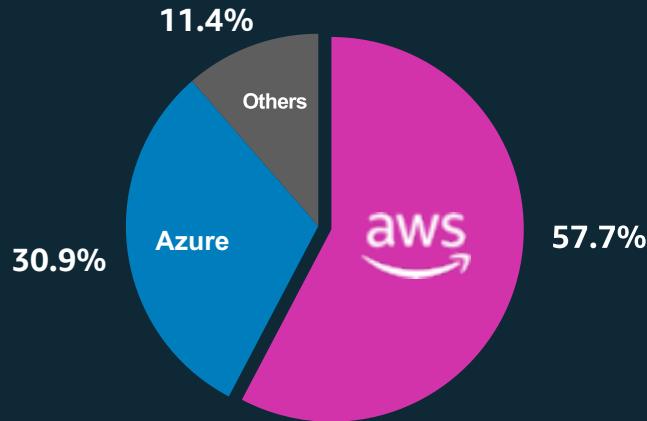
Ubuntu



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: DC 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





What is an Amazon Machine Image (AMI)?

Provides the information required to launch an instance

Launch multiple instances from a single AMI

An AMI includes the following

- A template for the root volume (for example, operating system, applications)
- Launch permissions that control which AWS accounts can use the AMI
- Block device mapping that specifies volumes to attach to the instance



Choosing an AMI

AWS Console

Step 1: Choose an Amazon Machine Image (AMI)

An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. You can select an AMI provided by AWS, our user community, or the AWS Marketplace, or you can select one of your own AMIs.

Quick Start

My AMIs

AWS Marketplace

Community AMIs

Free for me only

Amazon Linux 2 AMI HVM, SSD Volume Type - ami-04681a1dbd79675a5

Amazon Linux comes with five years support. It provides Linux kernel 4.14, tuned for optimal performance on Amazon EC2, systemd 218, GCC 7.3, Glibc 29.1, and the latest software packages through extras.

Root device type: /dev/sda1 Virtualization type: hvm (Not enabled: Yes)

Select

Amazon Linux AMI 2018.03.0 (HVM), SSD Volume Type - ami-0f8eef57

The Amazon Linux AMI is an EBS-backed, AWS-supported image. The default image includes AWS-command line tools, Python, Ruby, Perl, and Java. The repositories include Docker, PHP, MySQL, PostgreSQL, and other packages.

Root device type: /dev/sda1 Virtualization type: hvm (Not enabled: Yes)

Select

Red Hat Enterprise Linux 7.5 (HVM), SSD Volume Type - ami-6871a115

Red Hat Enterprise Linux version 7.5 (HVM), EBS-General Purpose (SSD) Volume Type

Root device type: /dev/sda1 Virtualization type: hvm (Not enabled: Yes)

Select

AWS Marketplace

aws marketplace

View Categories | Migration Mapping Assistant | Your Saved List

Sell in AWS Marketplace | Amazon Web Services Home | Help

Operating Systems (336 results) Showing 1 – 10

Categories

All Categories

Infrastructure Software

Operating Systems

Filters

Version

- CentOS 7 (x86_64) - with Updates HVM
- CentOS 6 (x86_64) - with Updates HVM
- Debian GNU/Linux 8 (Jessie)
- CentOS 6.5 (x86_64) - Release Media

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 c4.8xlarge --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

Bid for unused capacity, charged at a spot price which 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 FPGAs)
- Networking**
(up to 100 Gbps)

- Bare Metal
- Size
(Nano to 32xlarge)

Options

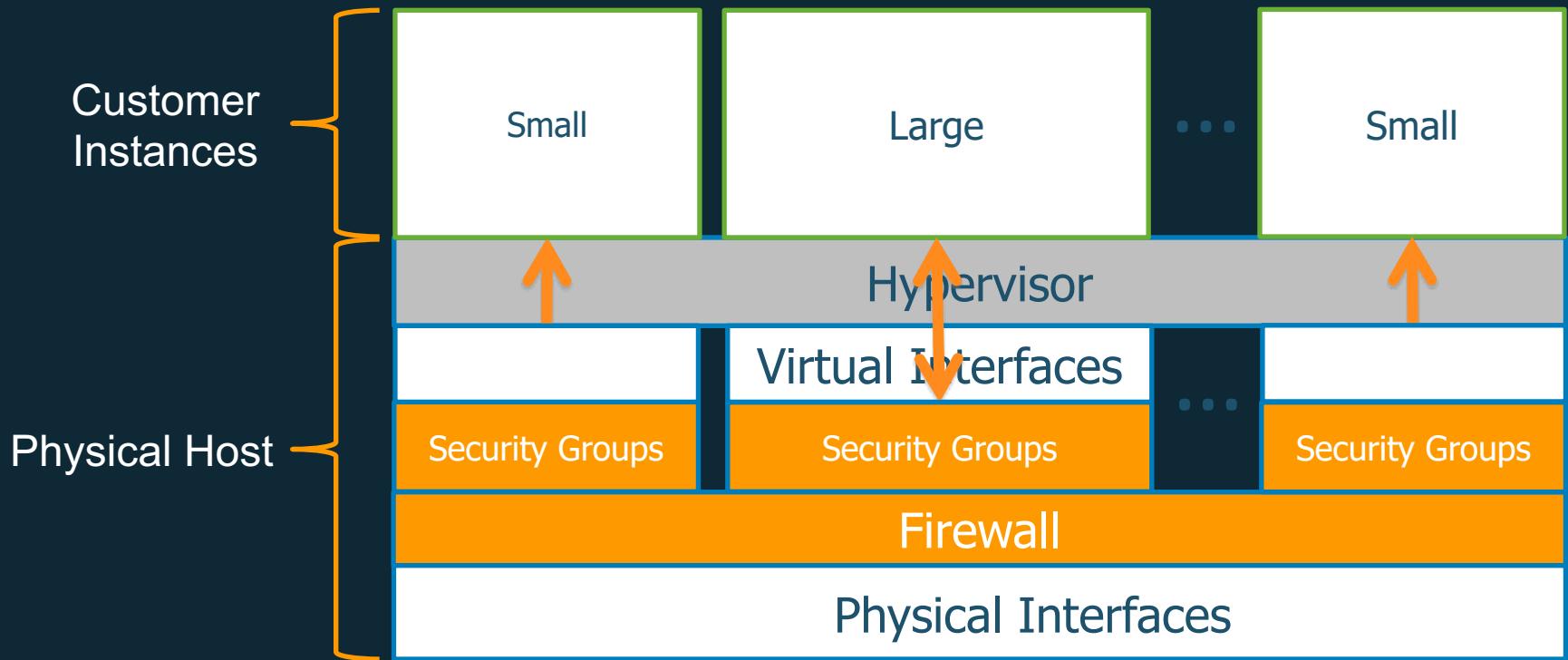
- Elastic Block Store
- Elastic Graphics
- Elastic Inference

175 instance types
for virtually
every workload
and business need



EC2 Design

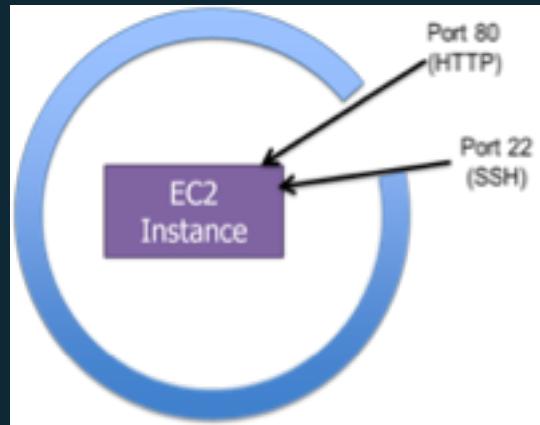
EC2 Host Virtualization



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:

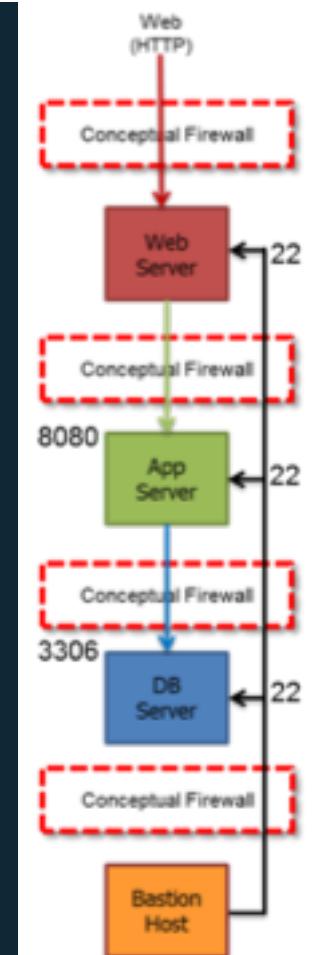
TCP 8080 “Web”
TCP 22 “Mgmt”

“DB” Security Group:

TCP 3306 “App”
TCP 22 “Mgmt”

“Mgmt” Security Group:

TCP 22 163.128.25.32/32



EC2 IP Addressing

Default VPC	Virtual Private Cloud
Dynamic Private IP	Dynamic or Static Private IP Address
Dynamic Public IP	None by default (can be created with publicIP=true)
Optional Static Public IP (EIP)	Optional Static Public IP (EIP)
AWS-provided DNS names <ul style="list-style-type: none">• Private DNS name• Public DNS name	AWS-provided public DNS lookup AWS-provided private DNS names Customer-controlled DNS options

EC2-Specific Credentials

EC2 key pairs

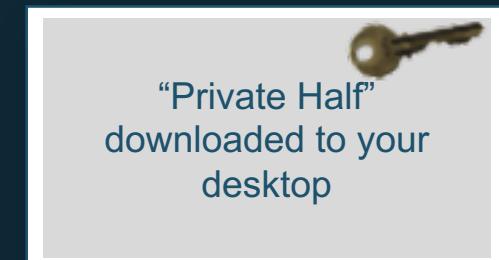
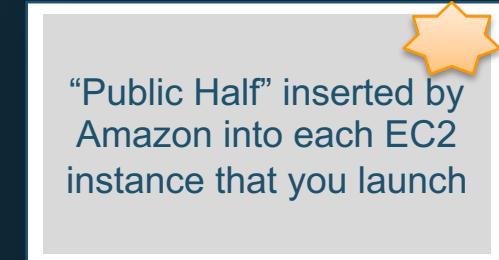
- Linux – SSH key pair for first-time host login
- Windows – Retrieve Administrator password

Standard SSH RSA key pair

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

AWS approach for providing initial access to a generic OS

- Secure
- Personalized
- Non-generic (NIST, PCI DSS)



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. What is a AMI?
4. What is a key pair?
5. How can an authorized user determine the IP address or hostname of an EC2 instance?

Any Questions?

