



Introduction to Amazon Cloud & EC2 Overview

Presenter Name, Email



Agenda

- Introduction to AWS Cloud
- Global Reach
- EC2 Overview
- EC2 Details

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

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, Smith, Dennis, Leong, Lydia, Bala, Raj, May 2018 G00336148
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: <https://www.gartner.com/doc/reprints?id=2G205FC&ct=150319&st=6>
Gartner does not endorse any vendor, product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner's research organization 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.

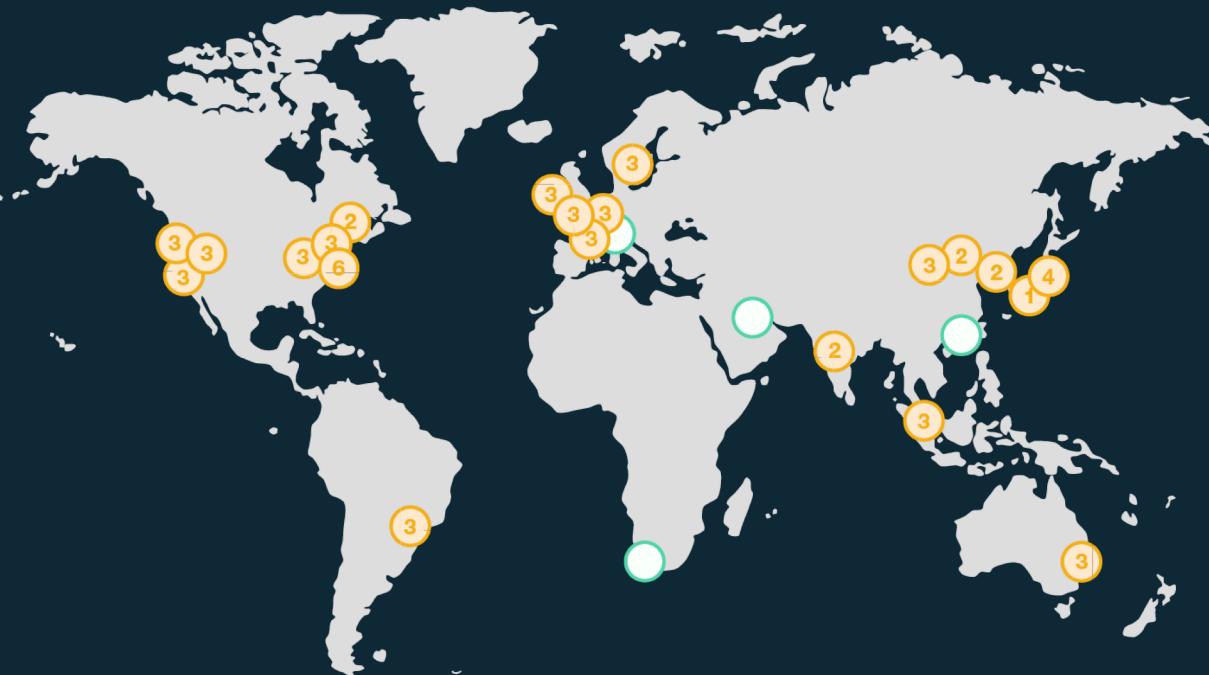
As of April 2018 © Gartner, Inc.

1

AWS Global Reach

AWS global reach

21 regions – 66 availability zones – 169 edge locations



Region & number of availability zones

AWS GovCloud (2)	EU
	Ireland (3)
US West	Frankfurt (3)
Oregon (4)	London (3)
Northern California (3)	Paris (3)
US East	Asia Pacific
N. Virginia (6), Ohio (3)	Singapore (2)
	Sydney (3), Tokyo (4),
Canada	Seoul (2), Mumbai (2)
Central (2)	China
South America	Beijing (2)
São Paulo (3)	Ningxia (2)

Announced regions

Bahrain, Hong Kong SAR, Sweden,
AWS GovCloud East

89

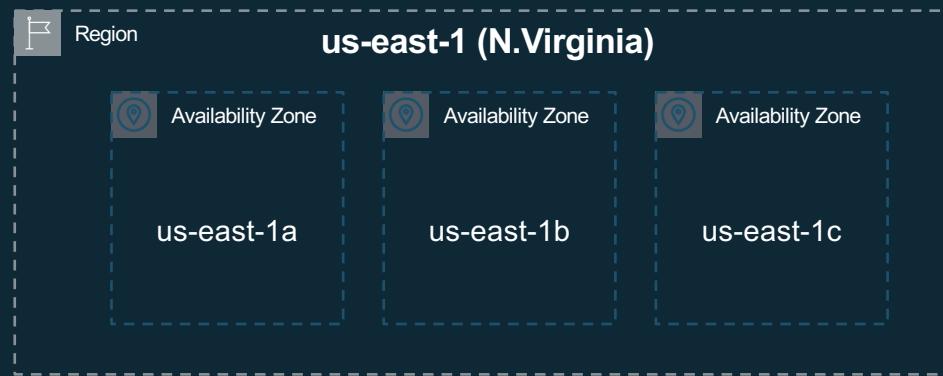
AWS Direct
Connect
locations





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

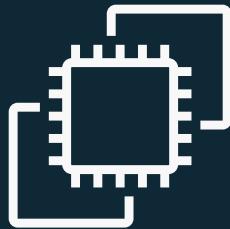




EC2 Overview



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

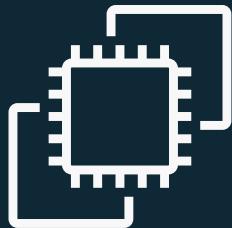


AWS Lambda

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

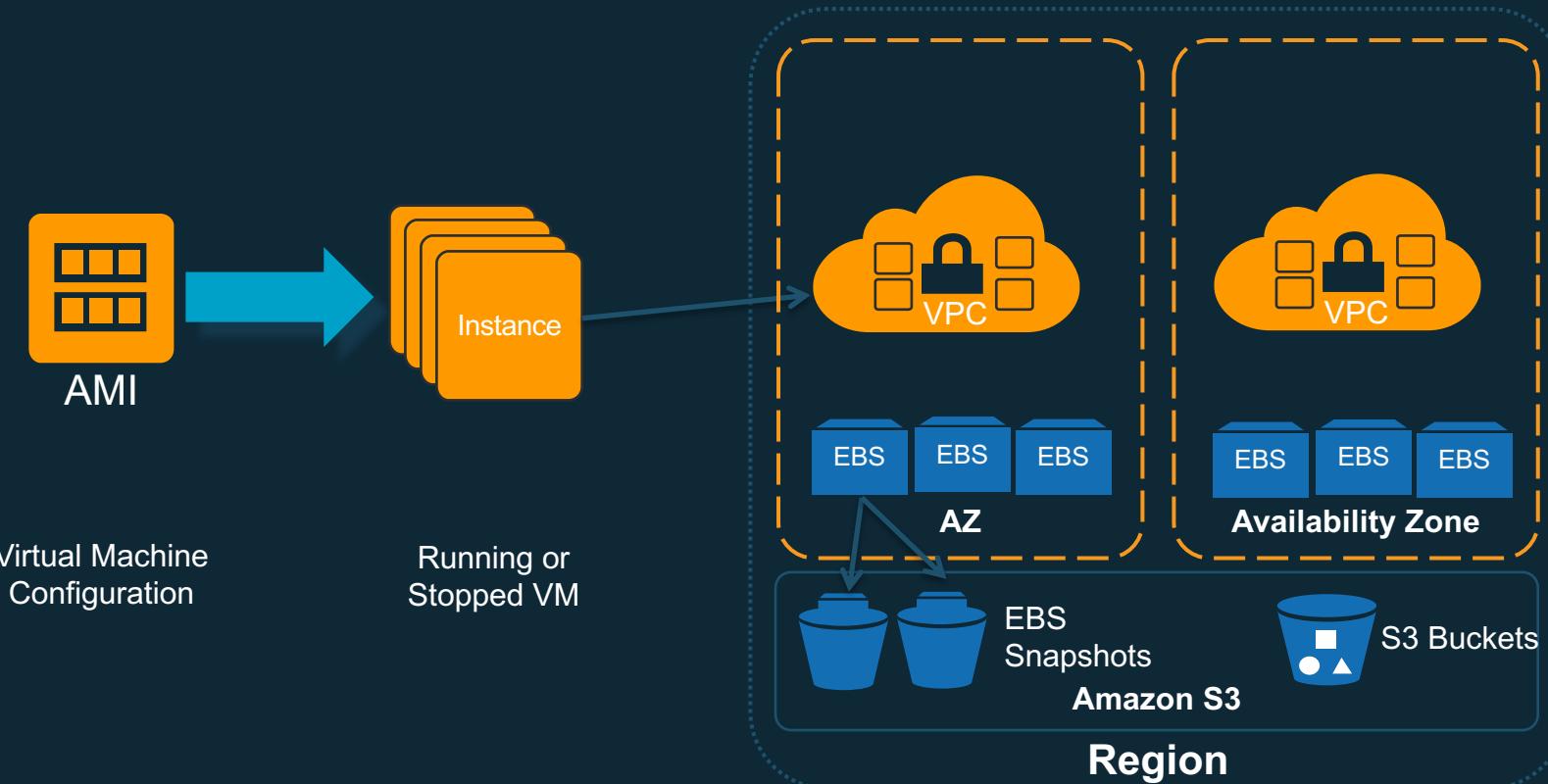
Bare metal, disk, networking capabilities

Packaged | Custom | Community AMIs

Multiple purchase options: On-demand, RI, Spot

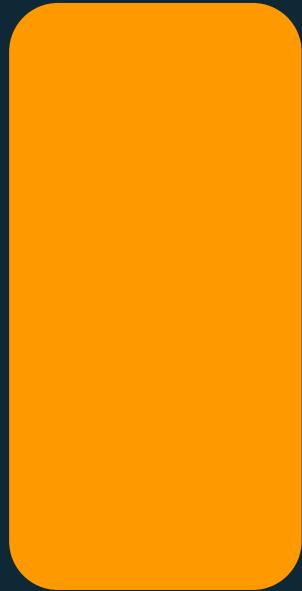


EC2 Terminology





Instance sizing



≈



≈



≈



c4.8xlarge

2 - c4.4xlarge

4 - c4.2xlarge

8 - c4.xlarge



EC2 Naming Explained

Instance generation

c5n.xlarge

The diagram illustrates the structure of an EC2 instance name, "c5n.xlarge". It features three yellow curly braces: one at the top level grouping "c5n" and ".xlarge"; one at the middle level grouping "c5n" and "Attribute"; and one at the bottom level grouping "family" and "Attribute".

Instance family Attribute Instance size



EC2 Operating Systems Supported

Windows 2003R2/2008/2008R2/2012/2012R2/2016/2019

Amazon Linux

Debian

Suse

CentOS

Red Hat Enterprise Linux

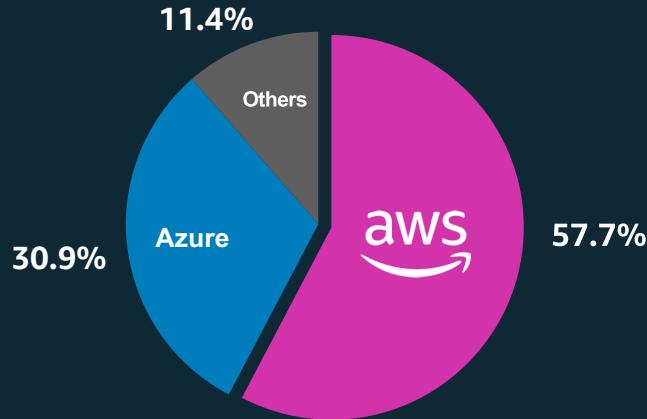
Ubuntu



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



Windows Licenses by Cloud Provider



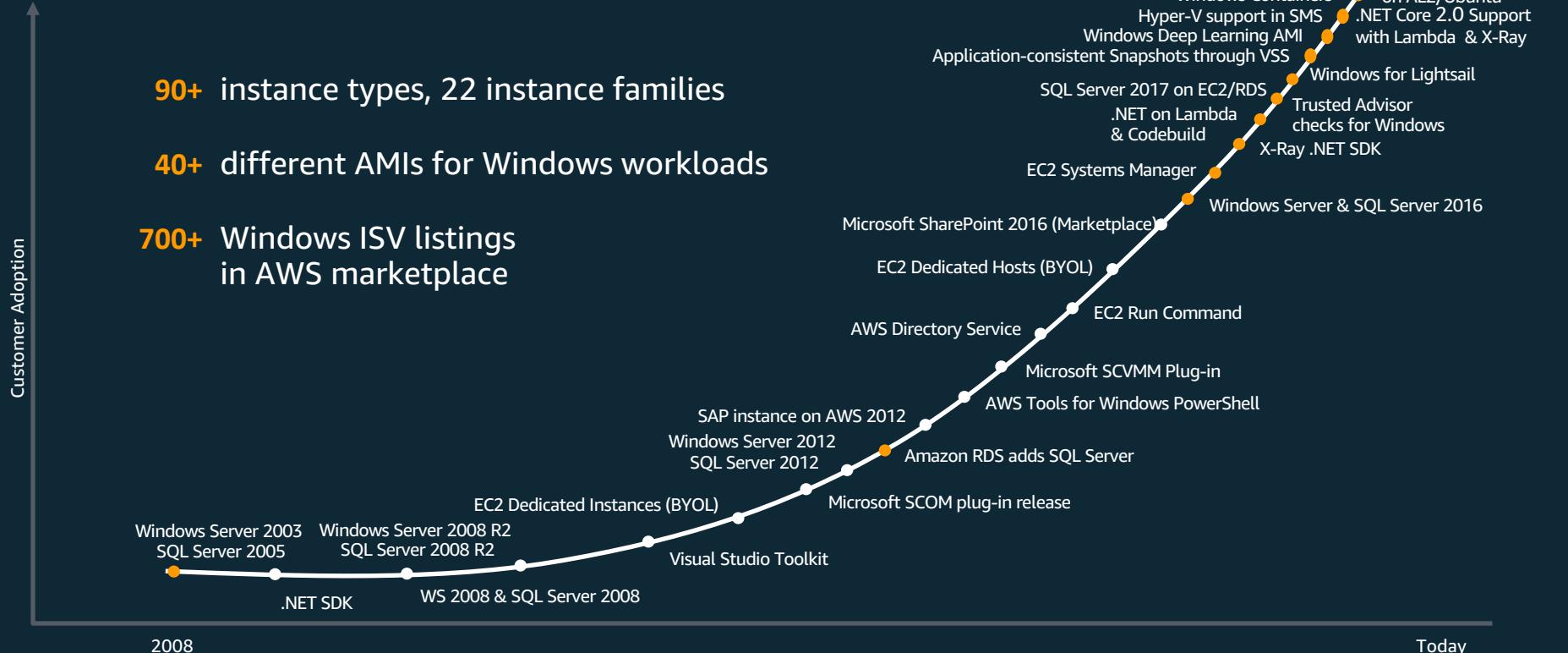
Note: Includes Windows instances deployed in the public cloud IaaS market during 2017. Source: DC estimates 2018

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

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

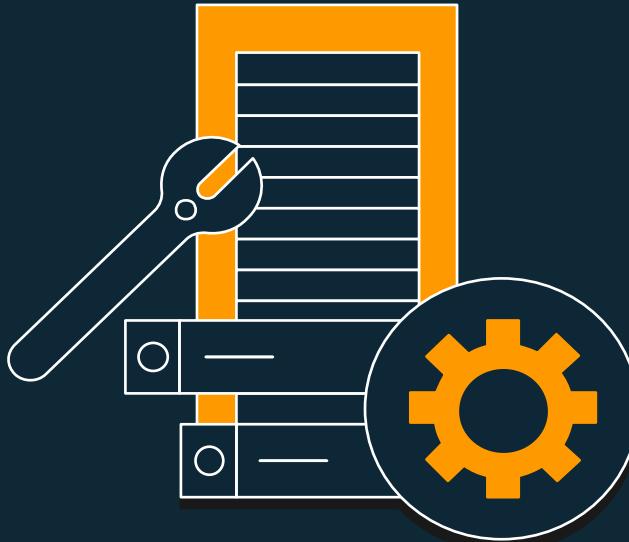
Innovation for Windows on AWS

Key Windows launches since 2008



Premier support for AWS customers running Microsoft workloads

- New support integration to improve customer experience
- AWS Support engineers can escalate directly to Microsoft Support on behalf of AWS customers





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

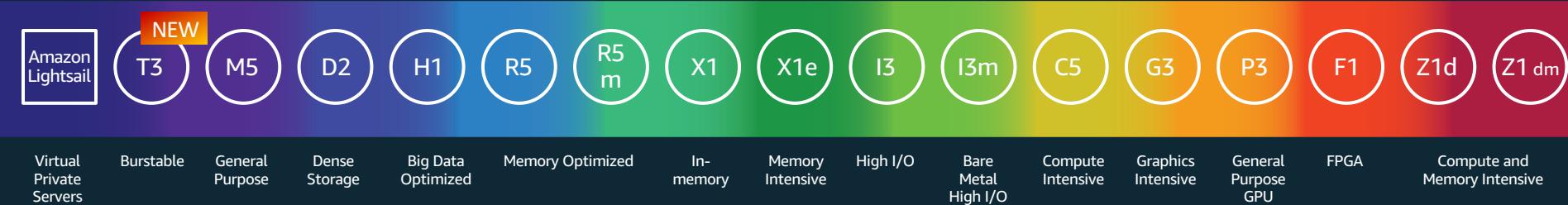
Launch multiple instances from a single AMI

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



Instance Types



EC2 Elastic GPUs

- Graphics acceleration for EC2 instances



EC2 Fleet

- Simplified provisioning
- Massive scale
- Flexible capacity allocation



Purchasing options at a glance

On-Demand Instances

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

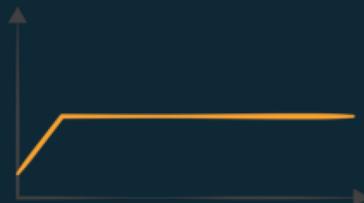
For Spiky workloads or to define needs



Reserved Instances

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

For committed utilization



Spot Instances

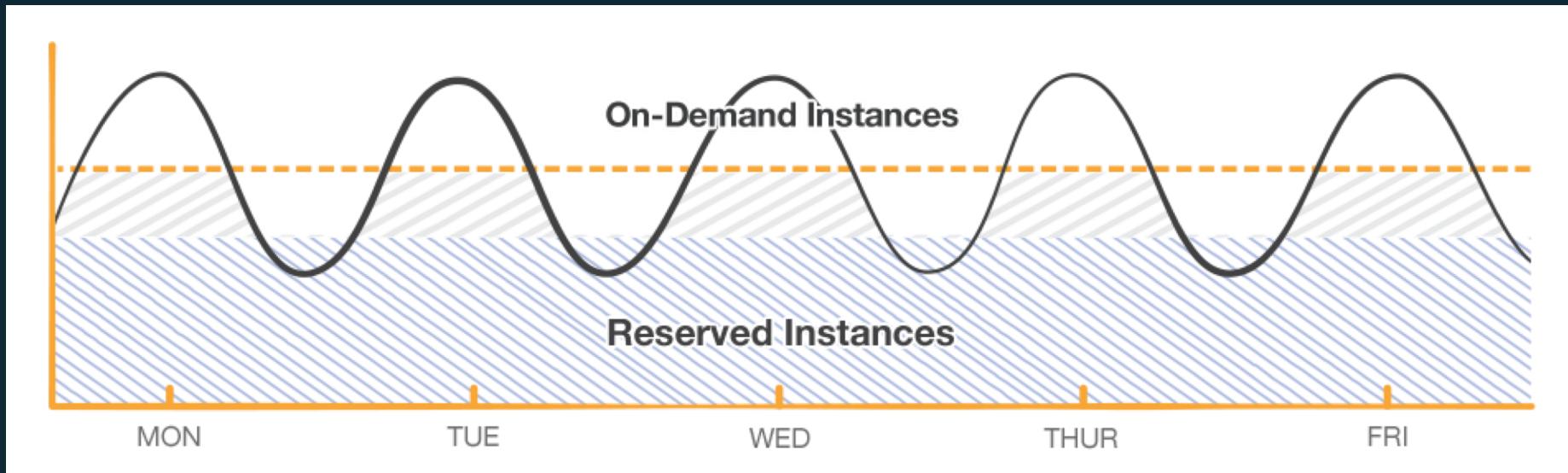
Bid for unused capacity, charged at a spot price which fluctuates based on supply and demand

For time-insensitive or transient workloads





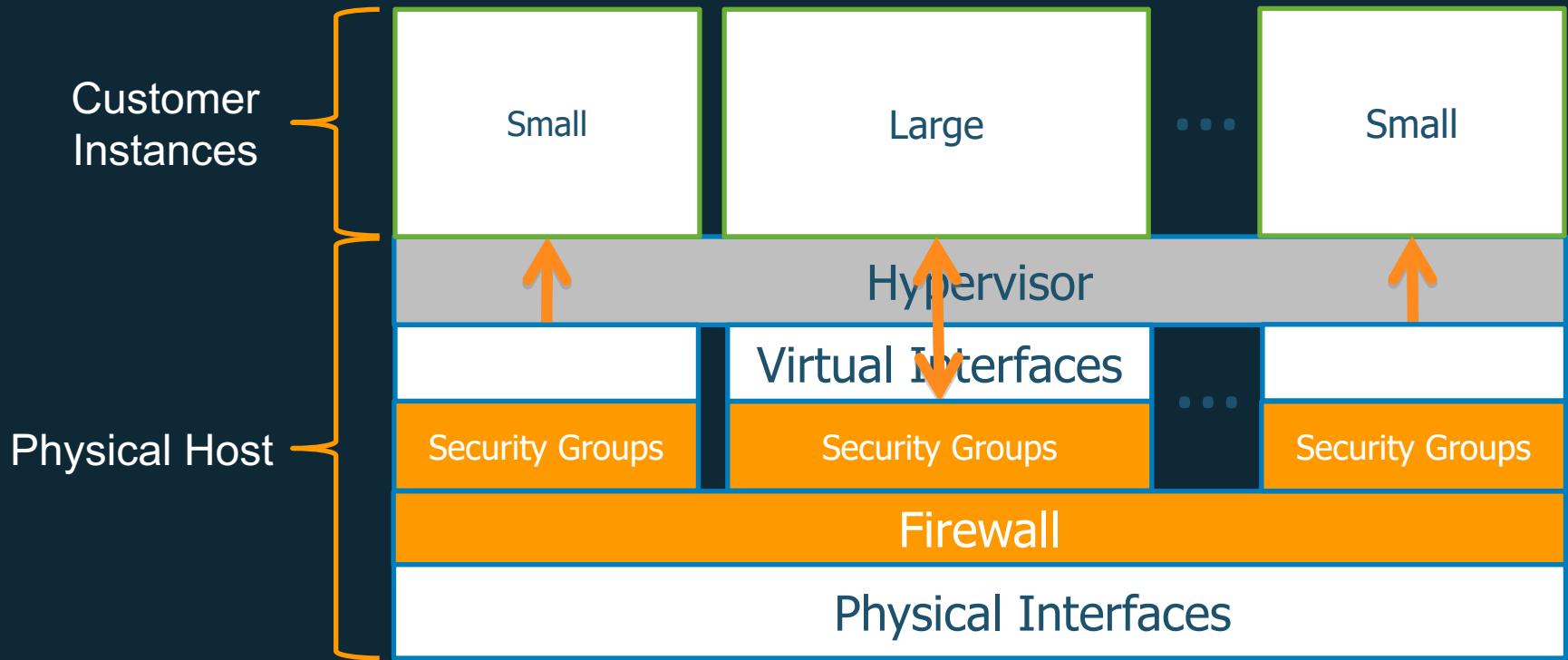
Layer your options





EC2 Design

EC2 Host Virtualization





Which hypervisor do we use?

Original: **Xen**

- Original hypervisor
- Consumed excessive resources
- Limited optimization

New (Nov/2017): **Custom KVM based hypervisor**

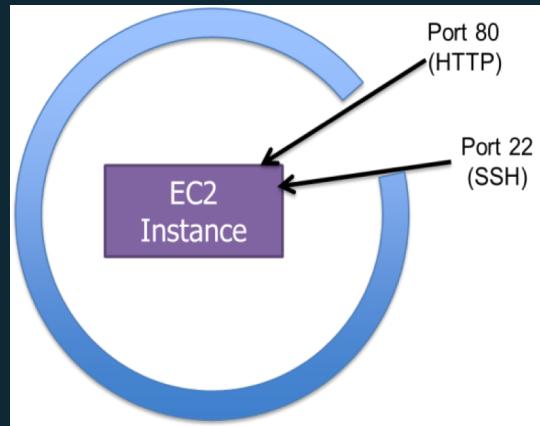
- Nitro instances
- Less server resources used, more resources for the customer
- AWS optimized

Bare metal: **No AWS provided hypervisor**

EC2 Security Groups

Security Group Rules

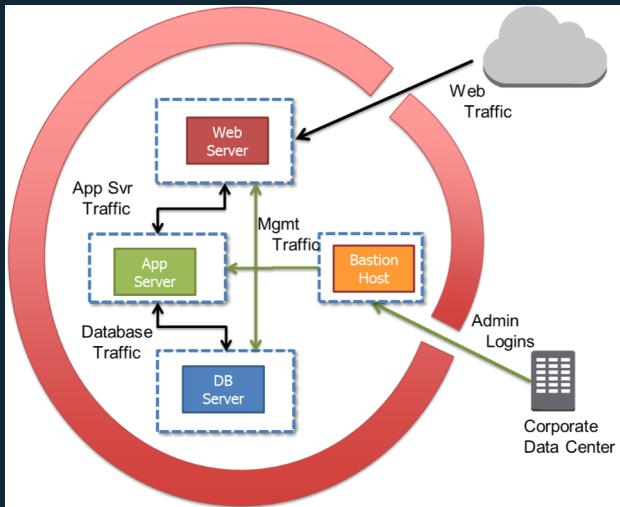
- Name
- Description
- Protocol
- Port range
- IP address, IP range, Security Group name



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:

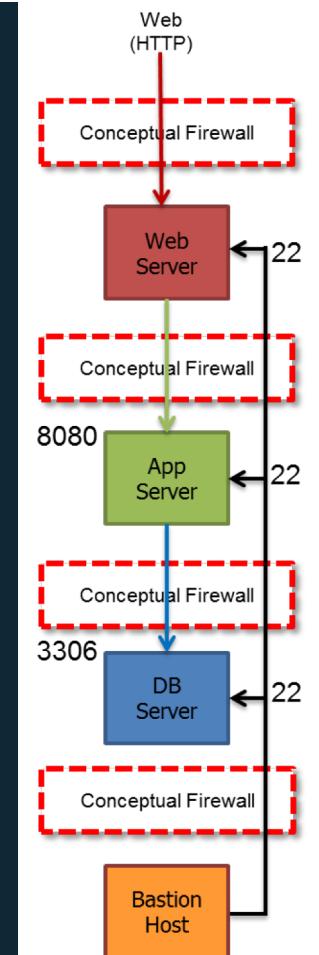
TCP 8080 "Web"
TCP 22 "Mgmt"

"DB" Security Group:

TCP 3306 "App"
TCP 22 "Mgmt"

"Mgmt" Security Group:

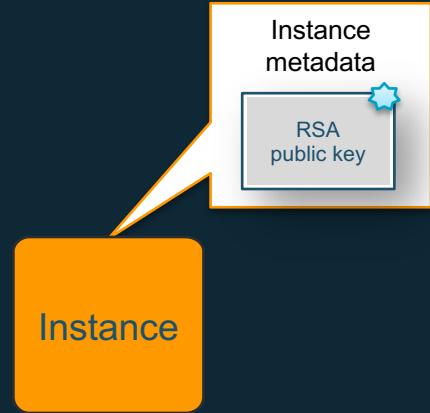
TCP 22 163.128.25.32/32



EC2 Instance access and Key Pairs

Linux launch (first boot)

- **Public key** made available through metadata
- Public key inserted into `~/.ssh/authorized_keys`
- User connects with SSH using their **private key**



EC2 Instance access and Key Pairs

Linux launch (first boot)

- **Public key** made available through metadata
- Public key inserted into `~/.ssh/authorized_keys`
- User connects with SSH using their **private key**

Windows launch (first boot sequence)

- **Public key** made available through metadata
- Sysprep
- Random Administrator password
- Password encrypted with public key
- User decrypts password with their **private key**

Instance

Instance metadata

RSA
public key

```
9/13/2011 9:55:18 PM: Waiting for meta-data accessibility...
9/13/2011 9:55:27 PM: Meta-data is now available.
<THUMBPRINT>44E515FED986668E107B2ADEB51B5FB1EF24E306E8</THUMBPRINT>
<RDPCERTIFICATE>
<cert>
<GhplGOqrJQmBJW4t1bqfNj946OckU19HrIz1InHTT25jVjAeuRF1Ifs9VS8VlxArIMs2tvTfbNSy+adM+6wRZ0dV8
</Password>
[REDACTED]
Initialization was successful.
9/13/2011 9:55:38 PM: Message: EC2Config Service is rebooting the instance. Please be patient.
```

System log
<Password>
aGlhplGOqrJQmBJW
...
K9gTD31Q==
</Password>

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/

Any Questions?

