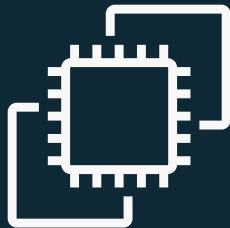




Compute



Choices for Compute



Amazon EC2

Virtual server instances
in the cloud



Amazon ECS, EKS, and Fargate

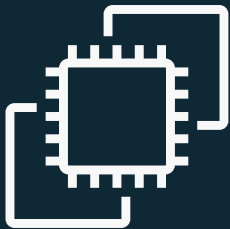
Container management
service for running
Docker on a managed
cluster of EC2



AWS Lambda

Serverless compute
for stateless code execution
in response to triggers

Amazon EC2



Amazon EC2

Linux | Windows

Arm and x86 architectures

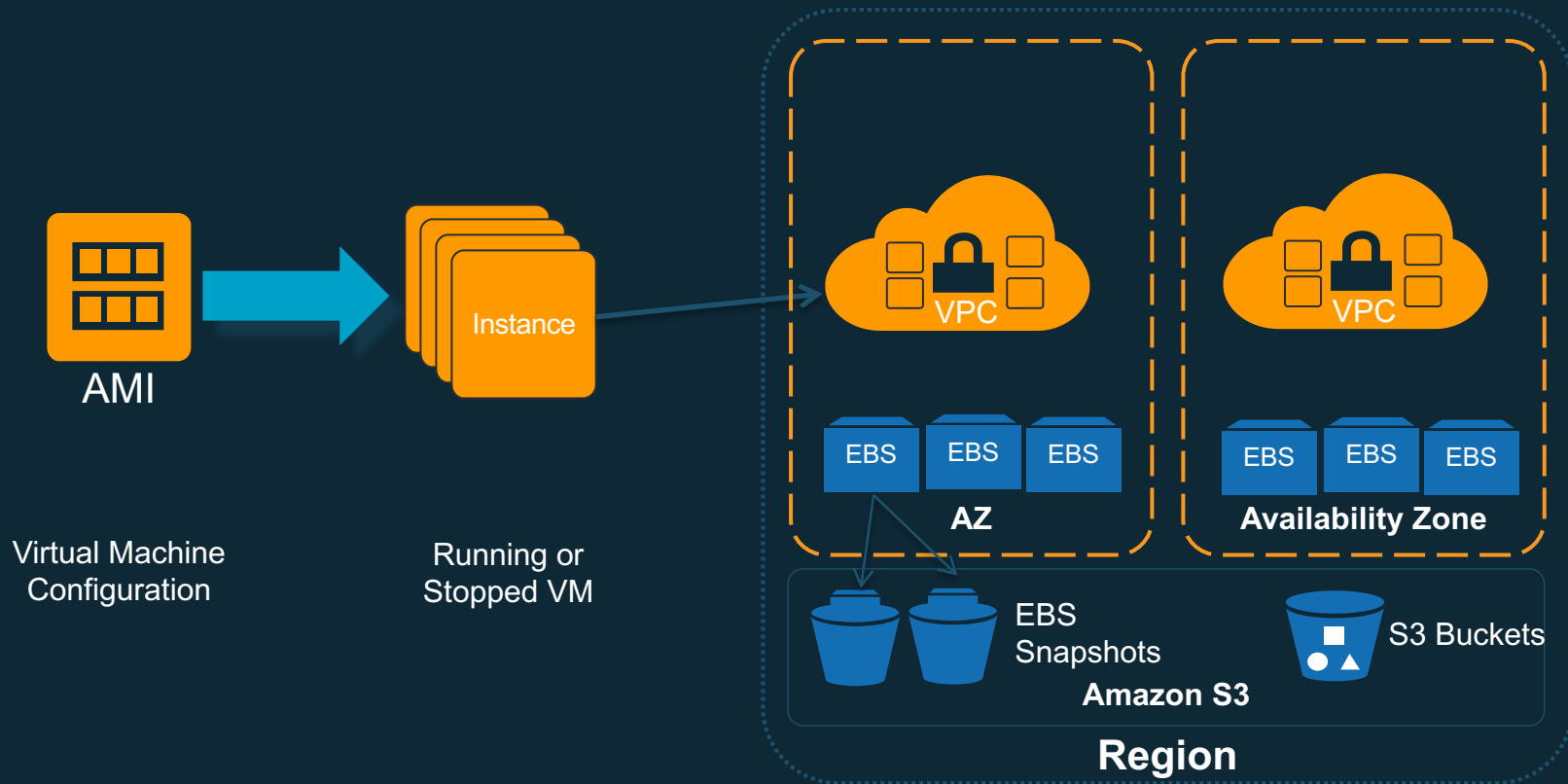
General purpose and workload optimized

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/>

EC2 Naming Explained



Instance generation

c5n.xlarge

Instance
family

Attribute

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>

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

1. Choose AMI 2. Choose Instance Type 3. Configure Instance 4. Add Storage 5. Add Tags 6. Configure Security Group 7. Review

Cancel and Exit

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 tier only ⓘ

AMI	Free tier eligible	64-bit	Select
Amazon Linux 2 AMI (HVM), SSD Volume Type - ami-04681a1dbd79675a5	<input checked="" type="checkbox"/>	64-bit	Select
Amazon Linux 2 comes with five years support. It provides Linux kernel 4.14 tuned for optimal performance on Amazon EC2, systemd 219, GCC 7.3, Glibc 2.26, Binutils 2.29.1, and the latest software packages through extras.			
Root device type: ebs Virtualization type: hvm ENA Enabled: Yes			
Amazon Linux 2018.03.0 (HVM), SSD Volume Type - ami-0ff8a910777f667	<input checked="" type="checkbox"/>	64-bit	Select
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: ebs Virtualization type: hvm ENA Enabled: Yes			
Red Hat Enterprise Linux 7.5 (HVM), SSD Volume Type - ami-6871a115	<input checked="" type="checkbox"/>	64-bit	Select
Red Hat Enterprise Linux version 7.5 (HVM), EBS General Purpose (SSD) Volume Type			
Root device type: ebs Virtualization type: hvm ENA Enabled: Yes			

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

Operating System	Free tier eligible	64-bit	Select
CentOS 7 (x86_64) - with Updates HVM ★★★★★ (58) Version 1805_01 Sold by Centos.org	<input checked="" type="checkbox"/>	64-bit	Select
This is the Official CentOS 7 x86_64 HVM image that has been built with a minimal profile, suitable for use in HVM instance types only. The image contains just enough packages...			
Linux/Unix, CentOS 7 - 64-bit Amazon Machine Image (AMI)			
CentOS 6 (x86_64) - with Updates HVM ★★★★★ (33) Version 1805_01 Sold by Centos.org	<input checked="" type="checkbox"/>	64-bit	Select
This is the Official CentOS 6 x86_64 HVM image that has been built with a minimal profile. The image contains just enough packages to run within AWS, bring up an SSH Server...			
Linux/Unix, CentOS 6 - 64-bit Amazon Machine Image (AMI)			
Debian GNU/Linux 8 (Jessie) ★★★★★ (86) Version 8.7 Sold by Debian	<input checked="" type="checkbox"/>	64-bit	Select
Debian is a computer operating system composed of software packages released as free and open source software primarily under the GNU General Public License along with other...			
Linux/Unix, Debian 8.6+1 - 64-bit Amazon Machine Image (AMI)			
CentOS 6.5 (x86_64) - Release Media ★★★★★ (55) Version 6.5 - 2013-12-01 Sold by CentOS.org	<input checked="" type="checkbox"/>	64-bit	Select
This is the Official CentOS 6.5 x86_64 image that has been built with a minimal profile. The image contains just enough packages to run within AWS, bring up an SSH Server...			

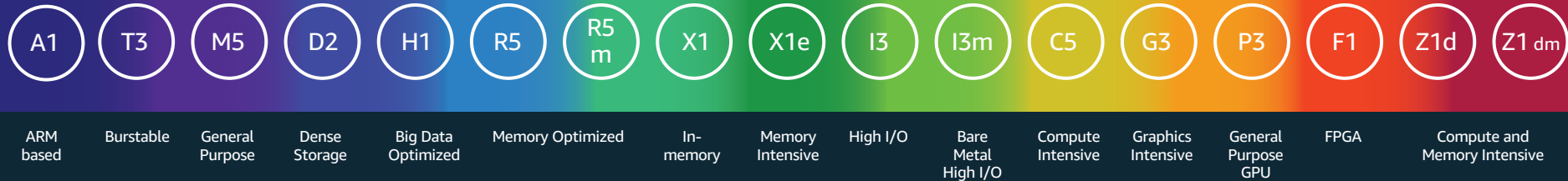
Use the AMI ID to launch through the API or AWS Command Line Interface (AWS CLI)

```
aws ec2 run-instances --image-id ami-04681a1dbd79675a5 --instance-type c4.8xlarge --count 10 --key-name MyKey
```





Instance Types



EC2 Elastic GPUs

- Graphics acceleration for EC2 instances



EC2 Fleet

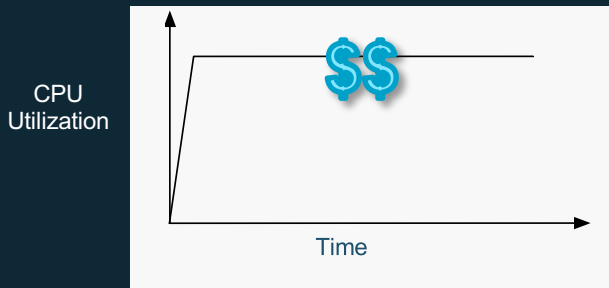
- Simplified provisioning
- Massive scale
- Flexible capacity allocation



What are Amazon Elastic Compute Cloud (Amazon EC2) burstable instances?

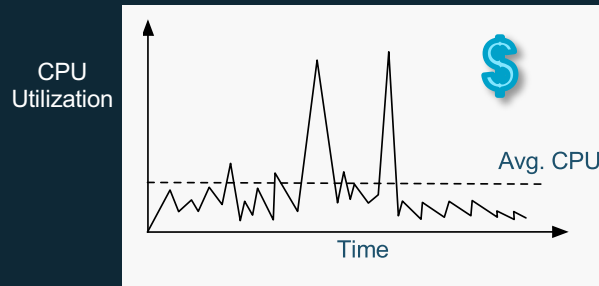
General Purpose Applications...

Continuous Fixed CPU
usage



Why pay more?

Variable Peaks CPU usage



or

Why pay less?
When you can pay less for
the same performance

Opportunity: Most instances aren't very busy



Burstable instances



Optimized for most general purpose workloads with occasional high CPU use



Ability to sustain high levels of CPU use indefinitely, if needed

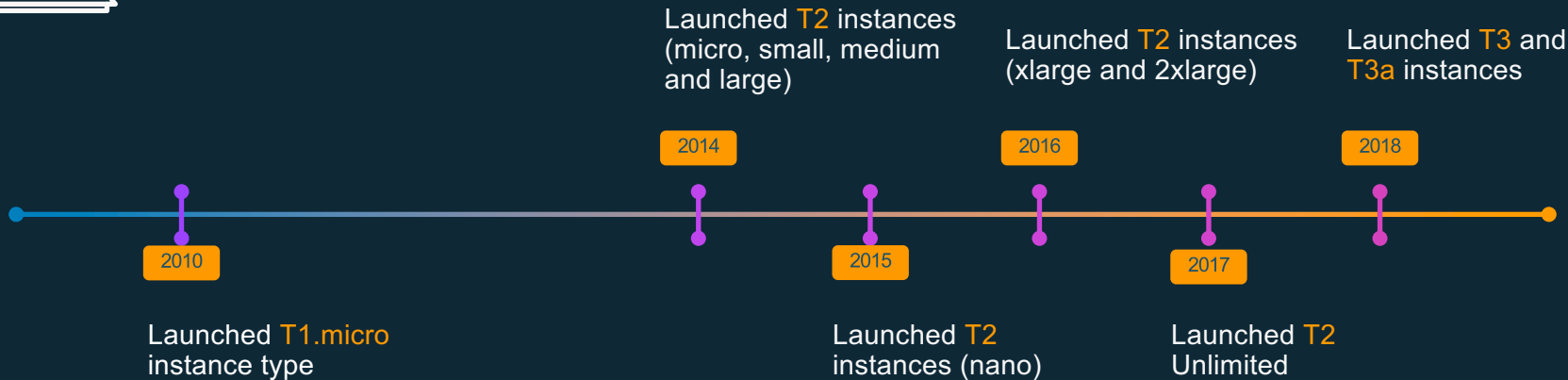


Moderate CPU usage is included in the offering



For sustained high CPU usage, additional CPU usage is “pay as you go”

Burstable instances family



Use burstable instances for...



Most general purpose instances that don't need fixed CPU resources



Applications that occasionally need quick access to high CPU

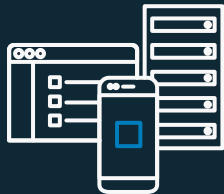


Idle or need moderate CPU for majority of the time



Need to burst CPU whenever and for as long as needed

... for a variety of general purpose workloads



-
- Microservices
 - Interactive Applications
 - Small and Medium Databases
 - Code Repositories
 - Virtual Desktops
 - Dev & Test Environment

CPU credits by instance

Instance size	vCPU count	Optimized for CPU usage	CPU credits per hour	Bucket size (Max CPU credits)
t3.nano	2	$\leq 5\%$	6	144
t3.micro	2	$\leq 10\%$	12	288
t3.small	2	$\leq 20\%$	24	576
t3.medium	2	$\leq 20\%$	24	576
t3.large	2	$\leq 30\%$	36	864
t3.xlarge	4	$\leq 40\%$	96	2,304
t3.2xlarge	8	$\leq 40\%$	192	4,608

Token bucket examples



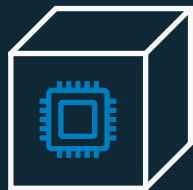
If an instance uses CPU below the **baseline** rate, the bucket will fill up to the **bucket size**.

If an instance uses CPU above the **baseline** rate, the bucket will start emptying.

If the instance CPU use drops back below the **baseline** rate before the bucket is empty, the bucket starts filling back up.

If the bucket is empty and the instance is still using CPU above the **baseline** rate, additional CPU usage above the **baseline** rate is billed at the end of the month

Standard mode



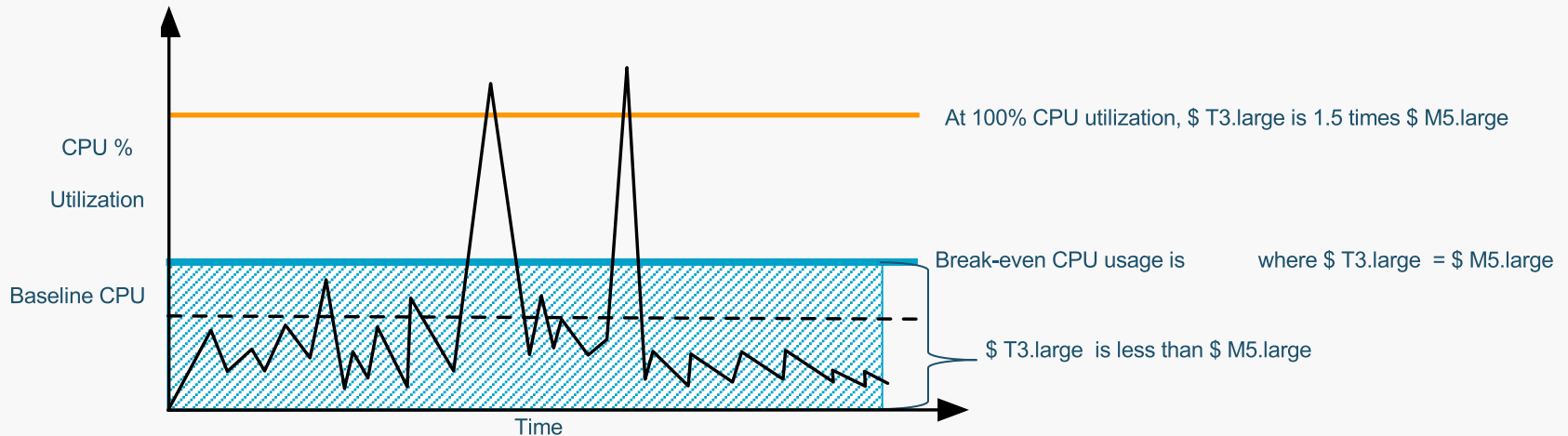
Standard mode is still available for T3 and T2, and is still the default on T2

Guarantees not paying for additional CPU usage

When token bucket is empty, instance cannot exceed baseline level of CPU usage.

Can flip between standard and unlimited mode at any time

For sizes where T3 and M5 overlap, when is T3 cheaper?



For sizes where T3 and M5 overlap, when is T3 cheaper?

instance size	vCPU	T3 price*/hour	M5 price /hour	Diff	T3 Base CPU %	Charge per vCPU hour	Charge per vCPU minute	Add'l mins avail.	Add'l CPU % avail.	Break-even CPU %
A	B	C	D	E = D - C	F	G	H = G / 60	I = E / H	J = (I / 60) / A	K = F + J
t3.large	2	\$0.0835	\$0.096	\$0.0125	30%	\$0.05	\$0.000833	15	12.5%	42.5%

t3 instance size	Break-even CPU %
t3.large	42.5%
t3.xlarge	52.5%
t3.2xlarge	52.5%

* prices based on us-east-1 and Linux OS



Purchasing options at a glance

On-Demand Instances

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

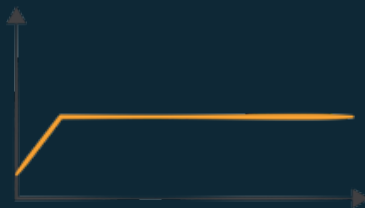
For Spiky workloads or to define needs



Reserved Instances

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

For committed utilization



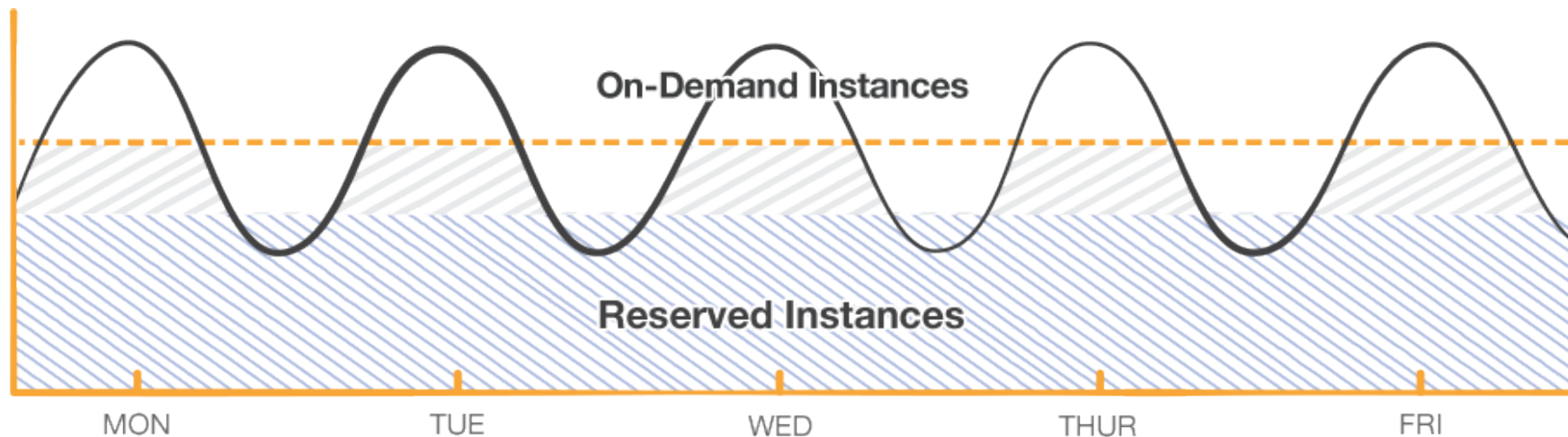
Spot Instances

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 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

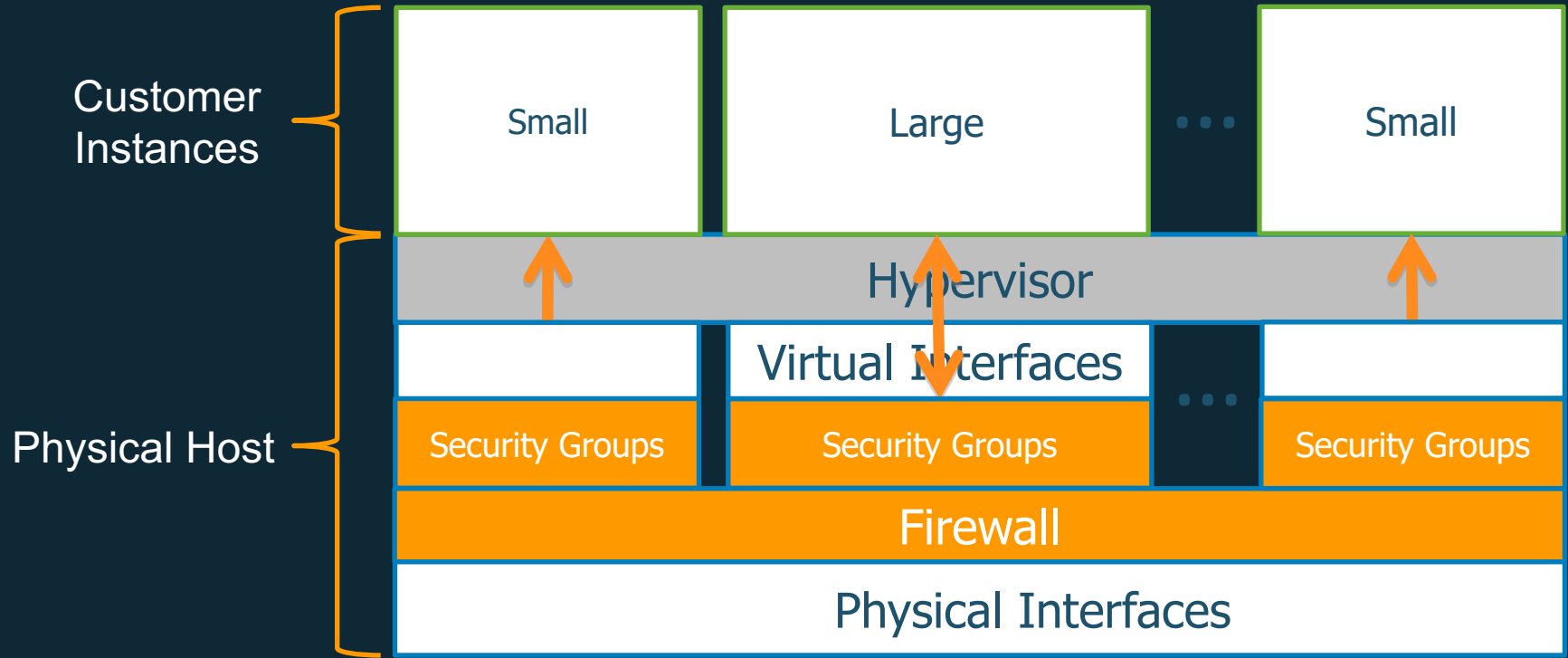


175
instance types
for virtually
every workload
and business need



EC2 Design

EC2 Host Virtualization





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**



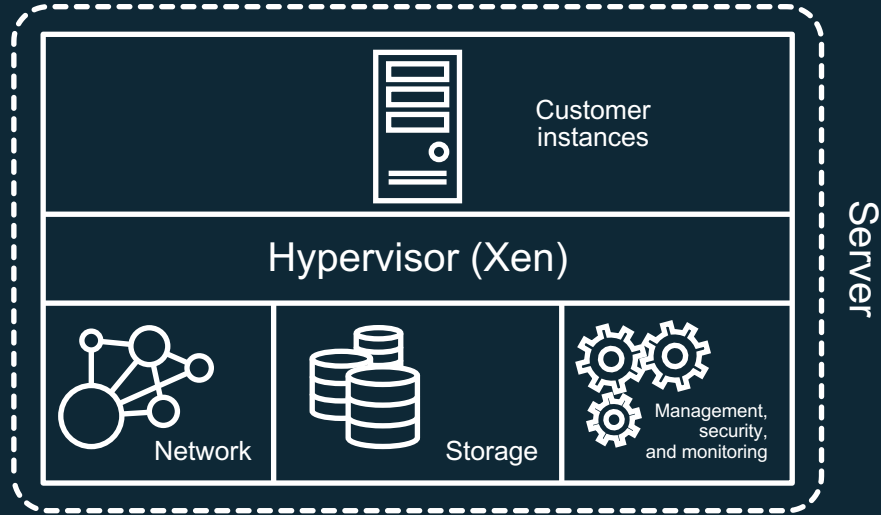
Hypervisor update

Original EC2 host architecture

All resources were on the server

Instance goals

- Security
- Performance
- Familiarity

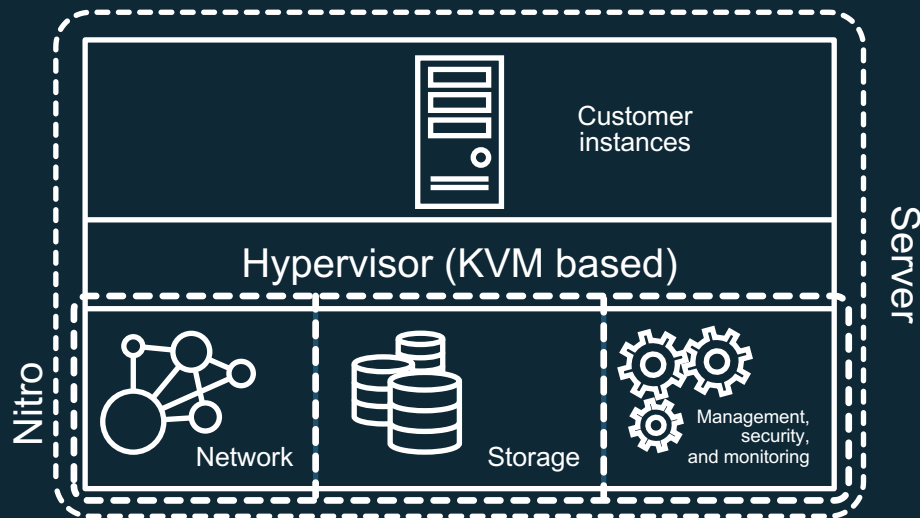




EC2 instance built on AWS Nitro System

Nearly 100% of available compute resources available to customers' workload

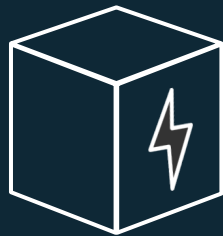
Improved security





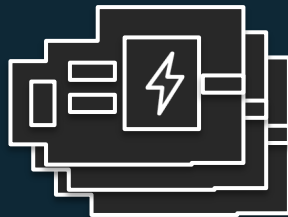
AWS Nitro System

Nitro Hypervisor



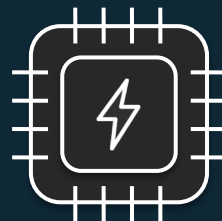
- Lightweight hypervisor
- Memory and CPU allocation
- Bare Metal-like performance

Nitro Card



- VPC Networking
- Amazon EBS
- Local Instance
- System Controller

Nitro Security Chip



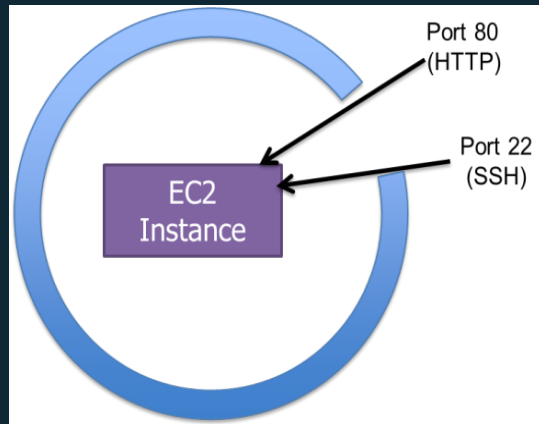
- Integrated into motherboard
- Protects hardware resources
- Hardware Root of Trust

Modular Building Blocks for rapid design and delivery of **EC2** instances

EC2 Security Groups

Security Group Rules

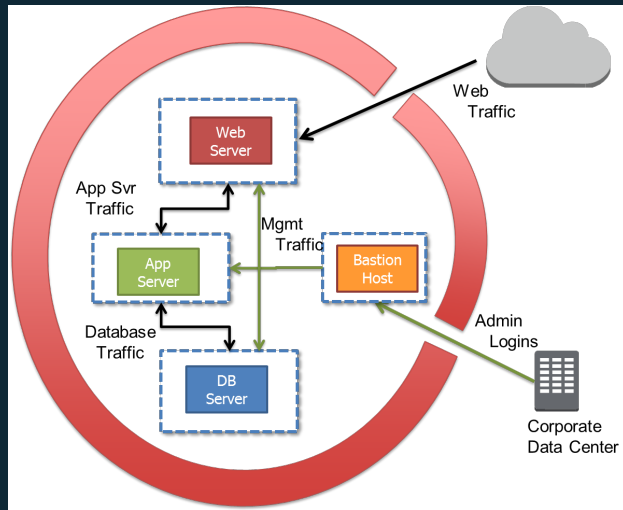
- Name
- Description
- Protocol
- Port range
- IP address, IP range, Security Group name



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”

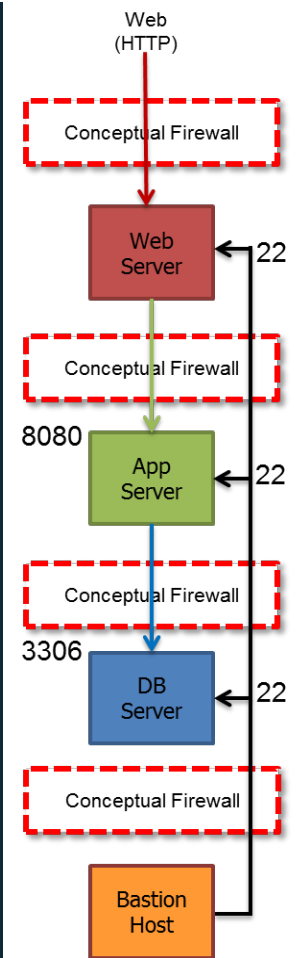
“DB” Security Group:

TCP 3306 “App”

TCP 22 “Mgmt”

“Mgmt” Security Group:

TCP 22 163.128.25.32/32



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?

