Amazon EC2

John Amaral, Jacob Komissar, Daniel McDonough,
Connor McNamara, Christopher Myers,
James Taylor, MaryAnn VanValkenburg

This Presentation

- Warehouse-Scale Computing
- Amazon EC2
- Virtualization
- Instances

What is Warehouse-Scale Computing?

- “The foundation of Internet services billions of people use every day: search, social networking, online maps, video sharing, online shopping, email services ...” - p466 Hennessy & Patterson 6th ed.

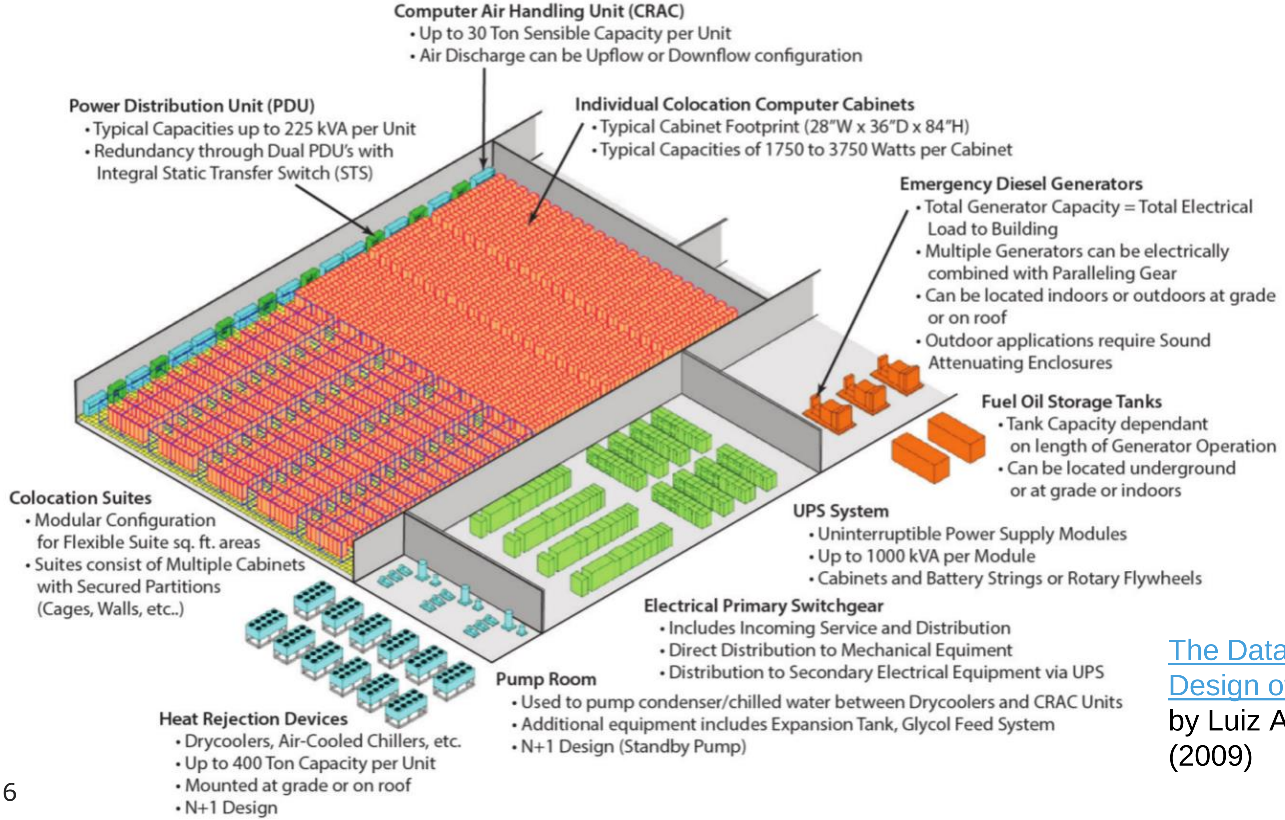
What is Warehouse-Scale Computing?

Data Center “Tier List”

- Tier I data centers have a single path for power distribution, UPS, and cooling distribution, without redundant components.
- Tier II adds redundant components to this design, improving availability.
- Tier III data centers have one active and one alternate distribution path for utilities. Each path has redundant components and are concurrently maintainable, that is, they provide redundancy even during maintenance.
- Tier IV data centers have two simultaneously active power and cooling distribution paths, redundant components in each path, and are supposed to tolerate any single equipment failure without impacting the load.

[The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines]
by Luiz André Barroso and Urs Hölzle, Google Inc. (2009)

What is Warehouse-Scale Computing?



[The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines]
by Luiz André Barroso and Urs Hölzle, Google Inc. (2009)

What is Warehouse-Scale Computing?

- Warehouse-Scale Computer (WSC):
  - A single huge machine running many processes.
  - A machine has memory, CPUs, and storage and buses that connect them all.
  - A warehouse-scale machine has thousands of machines all with a few, specific, configurations.
  - You treat the machines as CPUs and/or storage;
  - The network is the bus that connects them all.

The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines
by Luiz André Barroso and Urs Hölzle, Google Inc. (2009)

What is Warehouse-Scale Computing?

- Belong to a single organization
- The application, middle-ware, and system software are built in-house
- WSCs run a smaller number of very large applications (or Internet services), and the common resource management infrastructure allows significant deployment flexibility.

Principal Issues of WSC

- Energy Efficiency
- Cost-performance
- Dependability
- Security
- Architecture

Power Usage Effectiveness

- Energy efficiency
  - Primary concern in the design of WSC
  - Important component of the total cost of ownership
- Power Usage Effectiveness (PUE)
  - Total building power / IT equipment power
  - Not considering efficiency of server networking
  - Perfection = 1.0
  - Google WSC = 1.2

Slide taken from Professor Wawrzynek and Professor Weaver from Berkeley EECS

Cost Performance

- Maximum uptime
- Minimize the cost per kWh
- Minimize downtime
- Minimize PUE

Cost Performance

Table 6.1: Range of datacenter construction costs expressed in U.S. dollars per watt of critical power

Cost/W	Source
$12–25	Uptime Institute estimates for small- to medium-sized datacenters; the lower value is for "Tier 1" designs that are rarely used in practice [155]
$9–13	Dupont Fabros 2011 10 K report [43] contains financial information suggesting the following cost for its most recent facilities (built in 2010 and 2011 - see p 39 for critical load and p 76 for cost): $204 M for 18.2 MW (NJ1 Phase I) => $11.23/W $116 M for 13 MW (ACC6 Phase I) => $8.94/W $229 M for 18.2 MW (SC1 Phase 1) => $12.56/W
$8-10	Microsoft's investment of $130 M for 13.2 MW ($9.85/W) capacity expansion to its data center in Dublin, Ireland [105] Facebook is reported to have spent $210 M for 28 MW ($7.50/W) at its Prineville data center [49]

Critical power is defined as the peak power level that can be provisioned to IT equipment.

The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines
by Luiz André Barroso and Urs Hölzle, Google Inc. (2009)

Cost Performance

- Cluster Failure
  - Cluster ~ 1,800 servers
  - In each cluster's first year it is estimated:
    - 1,000 individual machine failures will occur
    - Thousands of hard drive failures will occur
    - 1 power distribution unit will fail
    - The cluster will have to be rewired once
    - 50% chance of overheating

https://www.datacenterknowledge.com/archives/2008/05/30/failure-rates-in-google-data-centers Jeff Dean

Dependability

- Users expect data to be correct
- Users expect availability & good performance
- Replication
- Sharding / Partitioning
- Load-balancing
- Health checking and watchdog timers

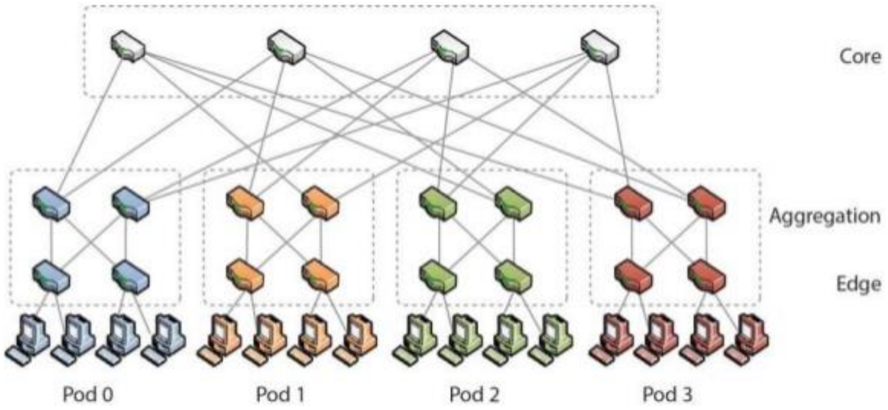
Security

- Security concerns in cloud computing are very similar to ones in conventional computing with added threats of:
    - Shared data and resources
    - Malicious internal actors having the potential to monitor traffic into and out of the service
    - Low barrier to entry being leveraged by malicious actors

Architecture

- Ample Parallelism
- Ever Changing Workload
- Scale
- Upgradability

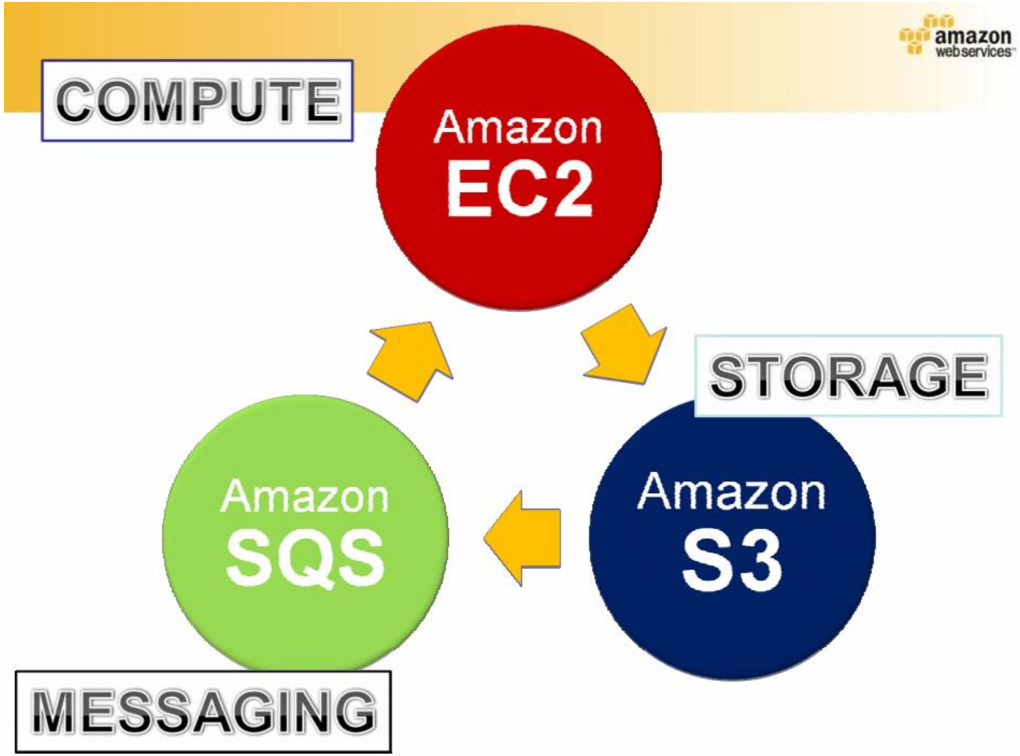
ARCHITECTURAL OVERVIEW OF WSCS



Sample three-stage fat tree topology.

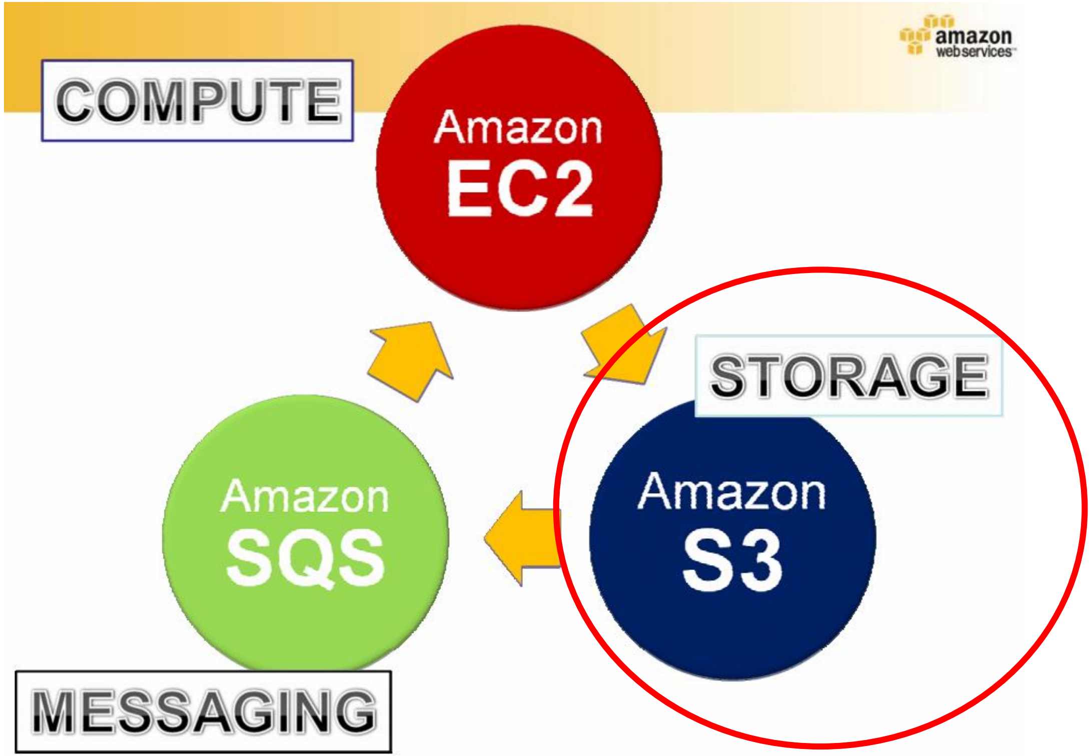
Overview of Amazon Web Services (AWS)

- Storage (S3)
  - Store data
  - Data protection
- Messaging (SQS)
  - Message queuing
  - Sending “jobs” between server components
- Compute (EC2)
  - Computational space



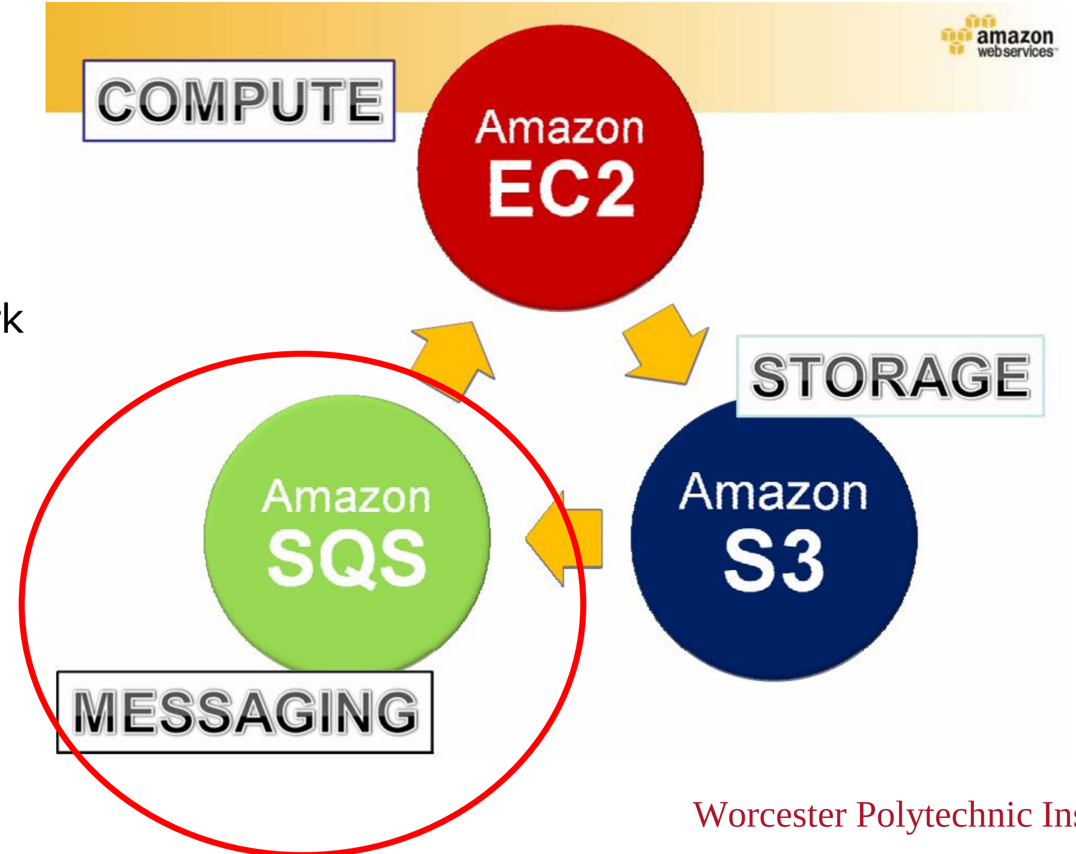
Overview of Amazon Web Services (AWS)

- Simple Storage Service (S3)
  - Store and protect any amount of data
  - Scalability
  - Private and public storage
  - Mobile apps, back-ups, big data analytics, etc.
  - Uses RAID at datacenter level



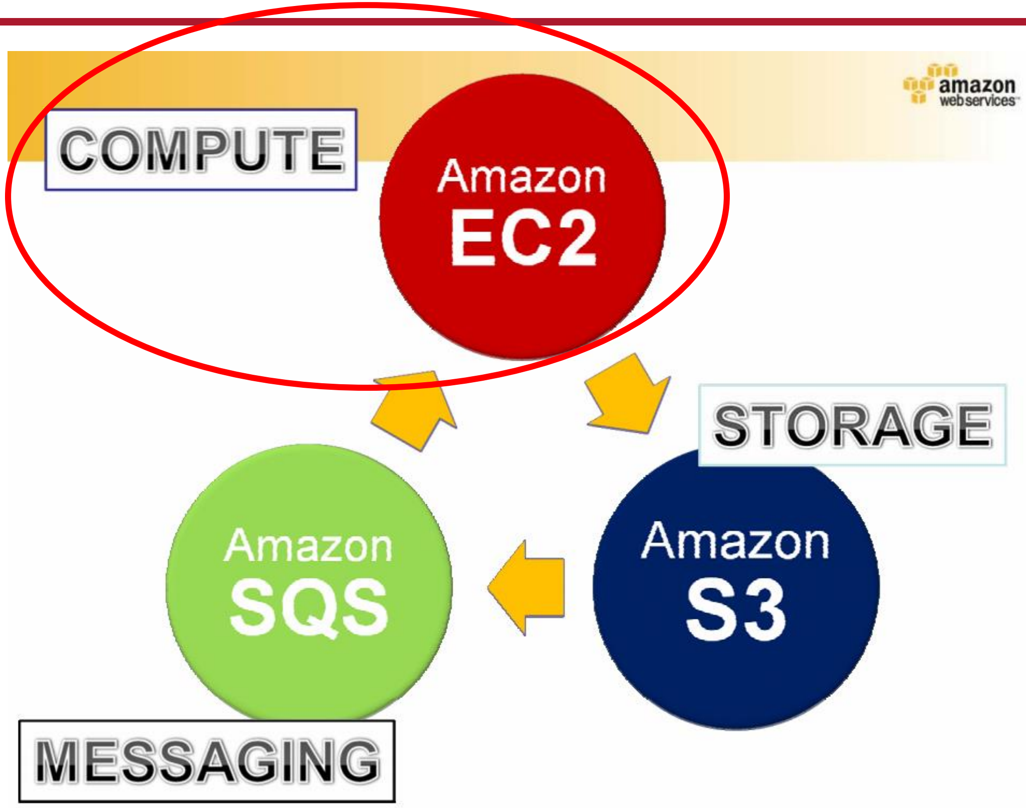
Overview of Amazon Web Services (AWS)

- Simple Queue Service (SQS)
  - Message queuing service
  - Differentiating work
  - Send, store, and receive messages in any volume
  - Standard vs. FIFO queues



Overview of Amazon Web Services (AWS)

- Elastic Compute Cloud (EC2)

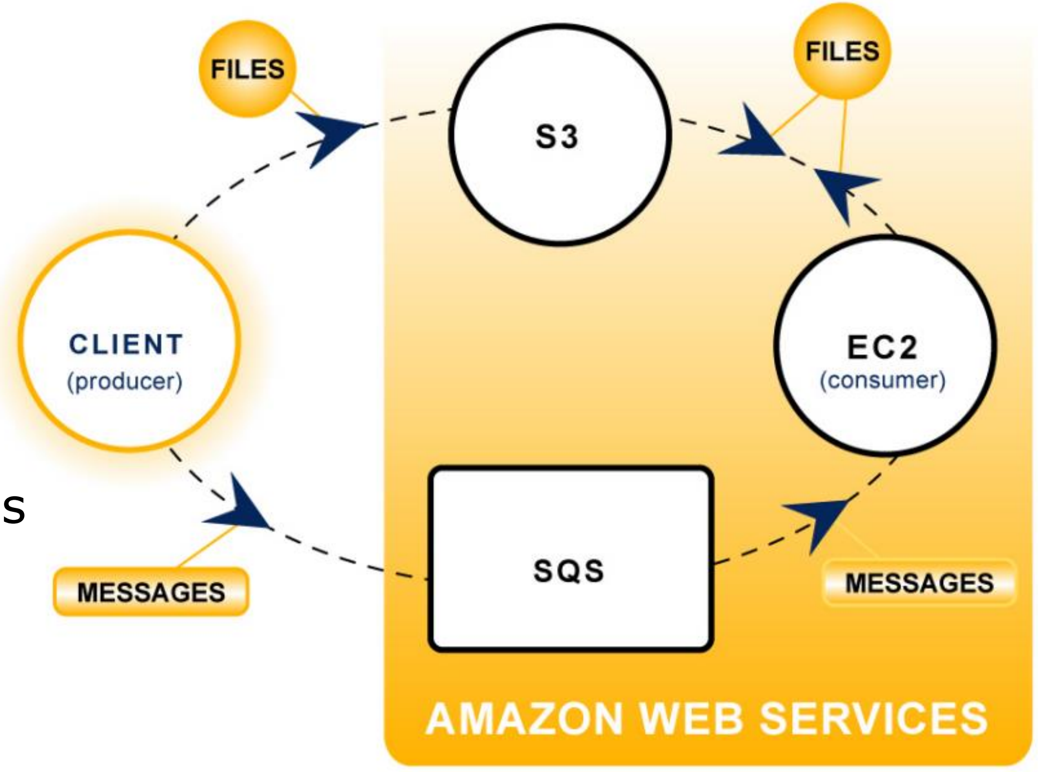


Amazon Elastic Compute Cloud (EC2)

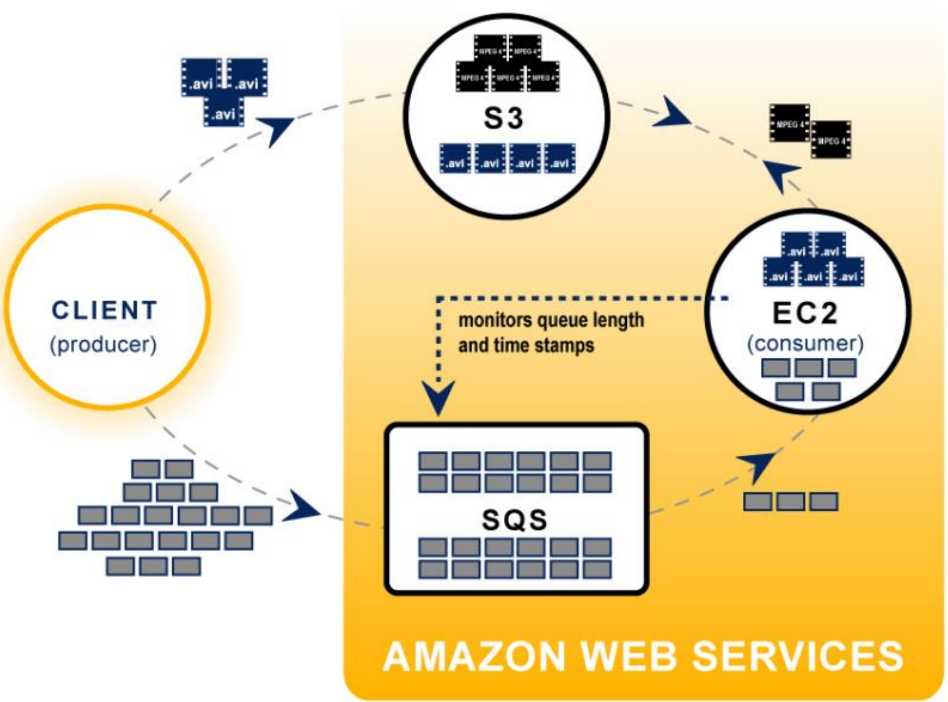
- Secure, resizeable computational space
- Quickly obtain and boot server instances
- Pay only for capacity used
- Auto scaling
  - Automatically add/remove instances
  - Predictive - based on predicted demand
  - Dynamic - respond to changing demand

How AWS Works

- Client sends request (message) to SQS or requests file from S3
- SQS queues and sends messages to EC2 for computations
- EC2 computes instructions and updates / receives files from S3
- Client obtains needed data from AWS



How AWS Works (Example)



History of EC2

- Beta introduced by Amazon in 2006
- Mostly developed by Amazon team in Cape Town, South Africa
- Continuously adding new instance types
- 2008 - availability zones, no more beta
    - Availability zones - 1 or more data centers, each at separate facility
- 2009 - auto scaling, cloud monitoring

Availability Zones (2017)

Design Issue

How to provide *efficient*, *cheap*, *dependable*, and *secure* resources to the **most** users and maximize profit?

Virtualization

What is virtualization?

Definitions from different sources…

- *OSTEP:* The illusion that you exclusively possess all the resources
- *VMWare*[1]*:* Decoupling of hardware and software
- *Redhat*[2]*:* Using the physical machine’s full capacity by distributing its capabilities among many users and environments

[1] https://pubs.vmware.com/vsphere-51/index.jsp?topic=%2Fcom.vmware.vsphere.vcenterhost.doc%2FGUID-ED375B12-7D08-4B7E-81EE-DCE83E51B1AF.html
[2] https://www.redhat.com/en/topics/virtualization/what-is-virtualization

Why virtualization?

- Can support multiple users on the same hardware
- Intelligently balance loads
- Pay for what you need
- *Users don't need to know how it works*

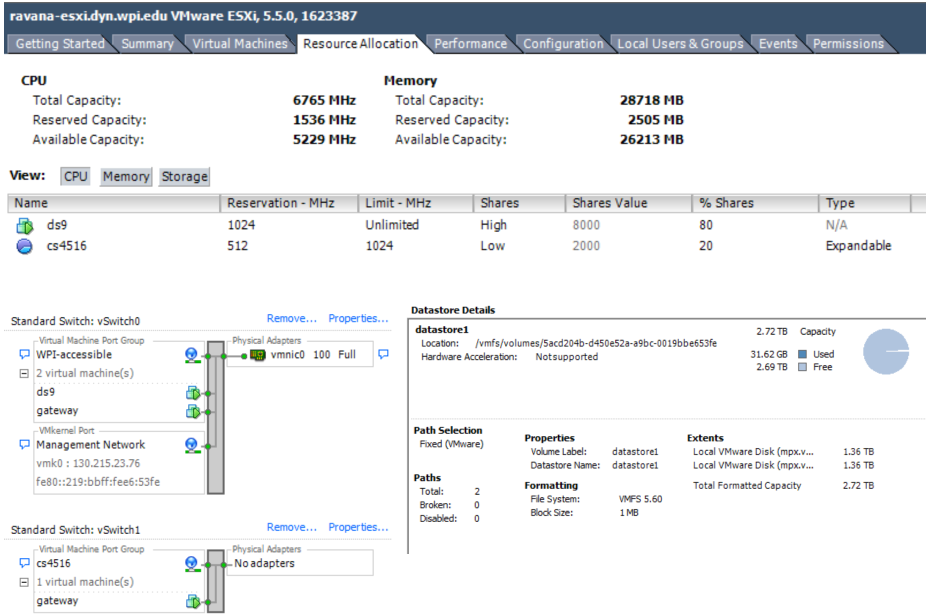
Baum, L. F. (1939). **The wizard of Oz**. Hollywood, Calif.: Metro Goldwyn Mayer

Virtual Machines — Overview

- Just as virtual memory provides “fake” memory, full virtualization provides “fake” hardware that runs virtual machines (VMs)
- Managed by a **hypervisor**
  - Examples: Oracle VirtualBox, VMWare ESXi
  - Hypervisors are to virtual machines & operating systems as operating systems are to processes
- Can use emulated hardware, hardware passthroughs, or virtual devices (e.g. virtual network switches), etc. to assist VMs
- VMs cannot access each other despite running on the same physical hardware

Hypervisors & VM Management

- Can assign resources to VMs
  - Memory
  - CPU time (allocated in MHz)
  - Disk space
  - Network bandwidth, requests
- Resource pools for groups of machines
- Virtual network devices and topology
- Redundancy, load balancing



Hypervisor Types

- **Hosted** hypervisors run as a process under another OS
  - e.g. VirtualBox, VMWare Workstation, QEMU
  - Not used in EC2
- **Bare-metal** hypervisors run directly on hardware, without a separate OS
  - e.g. Xen, VMWare ESXi
- Kernel-based Virtual Machine (KVM) is somewhere in between: it allows the Linux Kernel to be used as a bare-metal hypervisor
  - VMs run as Linux processes, using the Kernel’s resource management
  - KVM is now included in the Linux Kernel

Virtualization Techniques

- **Emulation:** implement the entire system in software
  - Very slow; not practical for large-scale computing
  - Required at some level for running executables from other architectures
- **Hardware-Assisted Virtualization:** hardware interrupts are handled by the hypervisor
  - Requires hardware support
- **Paravirtualization:** replace hardware accesses in software
  - Requires changes to software

Hardware-Assisted Virtualization

- Intel VT, AMD-V
    - New instructions allow CPUs to enter a mode for virtual machines
    - VMMs decide which events will exit each VM (and return control to the VMM)
- The VMM creates a separate interrupt table to handle interrupts that occur in the guest OS
- Historically slowed down by hardware accesses
    - Hypervisors needed to emulate any hardware the guest uses
- Driver paravirtualization has allowed hardware-assisted virtualization to surpass full paravirtualization

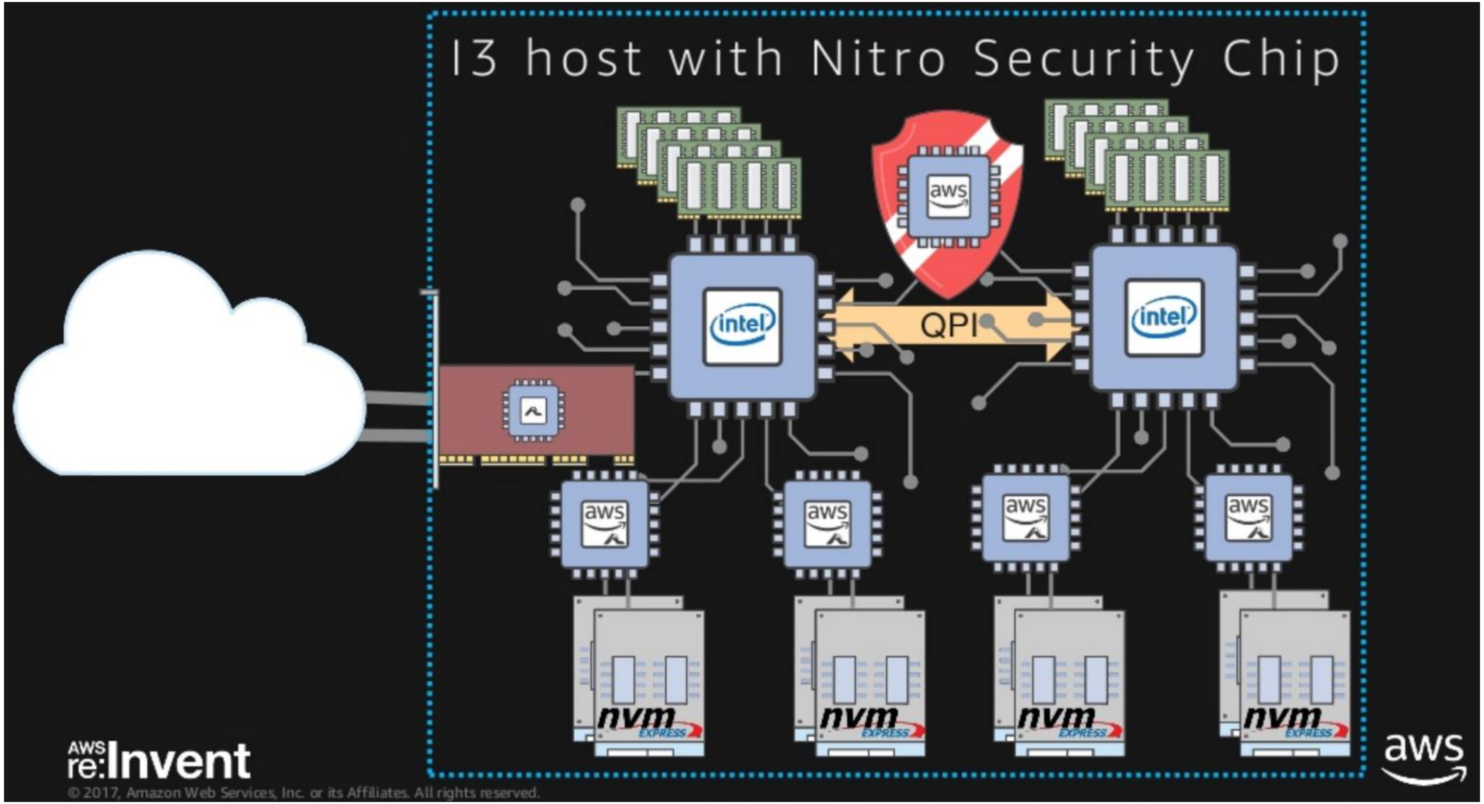
Paravirtualization (PV)

- Technique for virtualizing operations in software
  - Replace privileged/hardware calls in software with calls to the hypervisor
  - See Hennessy & Patterson (6th ed.) p. 126
- Allows virtualization on systems without any hardware support
  - Especially important pre-2006
- Main approach to x86 virtualization before 2006
  - Achieved full virtualization by using modified operating systems
- “Paravirtual drivers”: special drivers that call to the hypervisor
  - Used to improve hardware-assisted virtualization

Virtualization in EC2

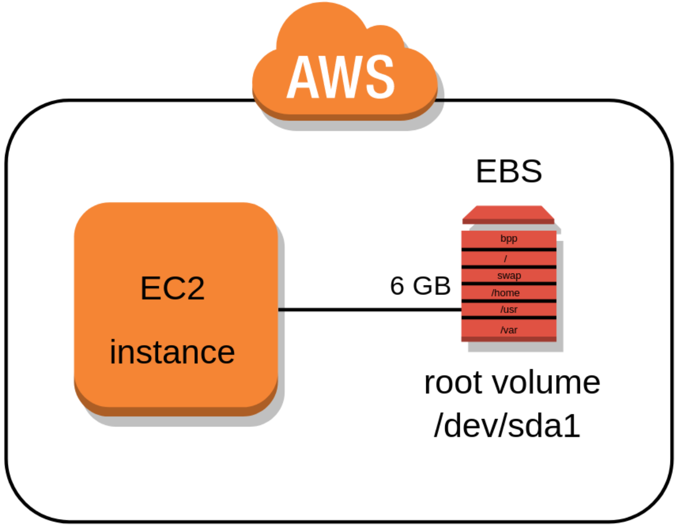
- Amazon both PV and hardware-assisted virtualization, on different types of EC2 instance
- Used Xen exclusively until late 2017
- **Nitro**
  - Amazon-developed system, built on KVM (Kernel Virtual Machine)
    - Uses Linux as a bare-metal hypervisor—VMs run as Linux processes
    - Requires hardware-assisted virtualization
  - Nitro removes unnecessary OS components
  - Some elements of the hypervisor are implemented in hardware as ASICs
    - Nitro Cards control access to different types of hardware
    - Nitro Security Chip traps I/O to non-volatile storage (e.g. firmware)
  - Used in every new type of EC2 instance

Nitro Architecture



Amazon EC2 Instances

- AWS server instances created upon booting
- Essentially, virtual machines that function as private servers for customers
- Currently 5 major types:
  - General purpose
  - Computer optimized
  - Memory optimized
  - Accelerated computing
  - Storage optimized



General Purpose Instances

- Ideal for scale-out workloads
  - Web servers, dev environments, etc.
- Much cheaper than other instance types
- Custom AWS Graviton processors with 64-bit ARM Neoverse cores and custom AWS silicon
- Use AWS Nitro System - hardware and lightweight hypervisor

General Purpose Instances

Model	vCPU	Mem (GiB)	Storage	Network Performance (Gbps)
a1.medium	1	2	EBS-Only	Up to 10
a1.large	2	4	EBS-Only	Up to 10
a1.xlarge	4	8	EBS-Only	Up to 10
a1.2xlarge	8	16	EBS-Only	Up to 10
a1.4xlarge	16	32	EBS-Only	Up to 10

Compute Optimized Instances

- Ideal for intense, computational workloads
  - Machine/deep learning, multiplayer gaming, etc.
- High performance, low price per computation
- 3.0 GHz Intel Xeon Platinum processors with AVX-512 instruction set
- Intel Turbo Boost - each core runs up to 3.5 GHz
- Enhanced networking with Elastic Network Adapter (ENA)
- SSDs using Non-volatile Memory Express (NVMe) connected to host server

Compute Optimized Instances

Model	vCPU*	Mem (GiB)	Storage (GiB)	Dedicated EBS Bandwidth (Mbps)	Network Performance (Gbps)
c5.large	2	4	EBS-Only	Up to 3,500	Up to 10
c5.xlarge	4	8	EBS-Only	Up to 3,500	Up to 10
c5.2xlarge	8	16	EBS-Only	Up to 3,500	Up to 10
c5.4xlarge	16	32	EBS-Only	3,500	Up to 10
c5.9xlarge	36	72	EBS-Only	7,000	10
c5.18xlarge	72	144	EBS-Only	14,000	25
c5d.large	2	4	1 x 50 NVMe SSD	Up to 3,500	Up to 10
c5d.xlarge	4	8	1 x 100 NVMe SSD	Up to 3,500	Up to 10
c5d.2xlarge	8	16	1 x 200 NVMe SSD	Up to 3,500	Up to 10
c5d.4xlarge	16	32	1 x 400 NVMe SSD	3,500	Up to 10

Memory Optimized Instances

- Ideal for memory-intensive workloads
  - High performance databases, real-time big data analytics, etc.
- 3.1 GHz Intel Xeon Platinum 8175 processors with AVX-512 instruction set
- AWS Nitro System
- SSDs with NVME connected to host server

Memory Optimized Instances

Model	vCPU	Mem (GiB)	Storage (GiB)	Dedicated EBS Bandwidth (Mbps)	Networking Performance (Gbps)
r5.large	2	16	EBS-Only	up to 3,500	Up to 10
r5.xlarge	4	32	EBS-Only	up to 3,500	Up to 10
r5.2xlarge	8	64	EBS-Only	up to 3,500	Up to 10
r5.4xlarge	16	128	EBS-Only	3,500	Up to 10
r5.12xlarge	48	384	EBS-Only	7,000	10
r5.24xlarge	96	768	EBS-Only	14,000	25
r5.metal	96*	768	EBS-Only	14,000	25
r5d.large	2	16	1 x 75 NVMe SSD	up to 3,500	Up to 10

Accelerated Computing Instances

- General purpose GPU instances
  - Computational fluid dynamics, speech recognition, autonomous vehicles, etc.
- 2x, 8x, and 16xlarge have Intel Xeon E5-2686 v4 processors
- 24xlarge has 2.5 GHz Intel Xeon P-8175M processors, AVX-512 instruction set
- NVLink for inter-GPU communication

Accelerated Computing Instances

Model	GPUs	vCPU	Mem (GiB)	GPU Mem (GiB)	GPU P2P	Storage (GiB)	Dedicated EBS Bandwidth	Networking Performance
p3.2xlarge	1	8	61	16	-	EBS-Only	1.5 Gbps	Up to 10 Gigabit
p3.8xlarge	4	32	244	64	NVLink	EBS-Only	7 Gbps	10 Gigabit
p3.16xlarge	8	64	488	128	NVLink	EBS-Only	14 Gbps	25 Gigabit
p3dn.24xlarge	8	96	768	256	NVLink	2 x 900 NVMe SSD	14 Gbps	100 Gigabit

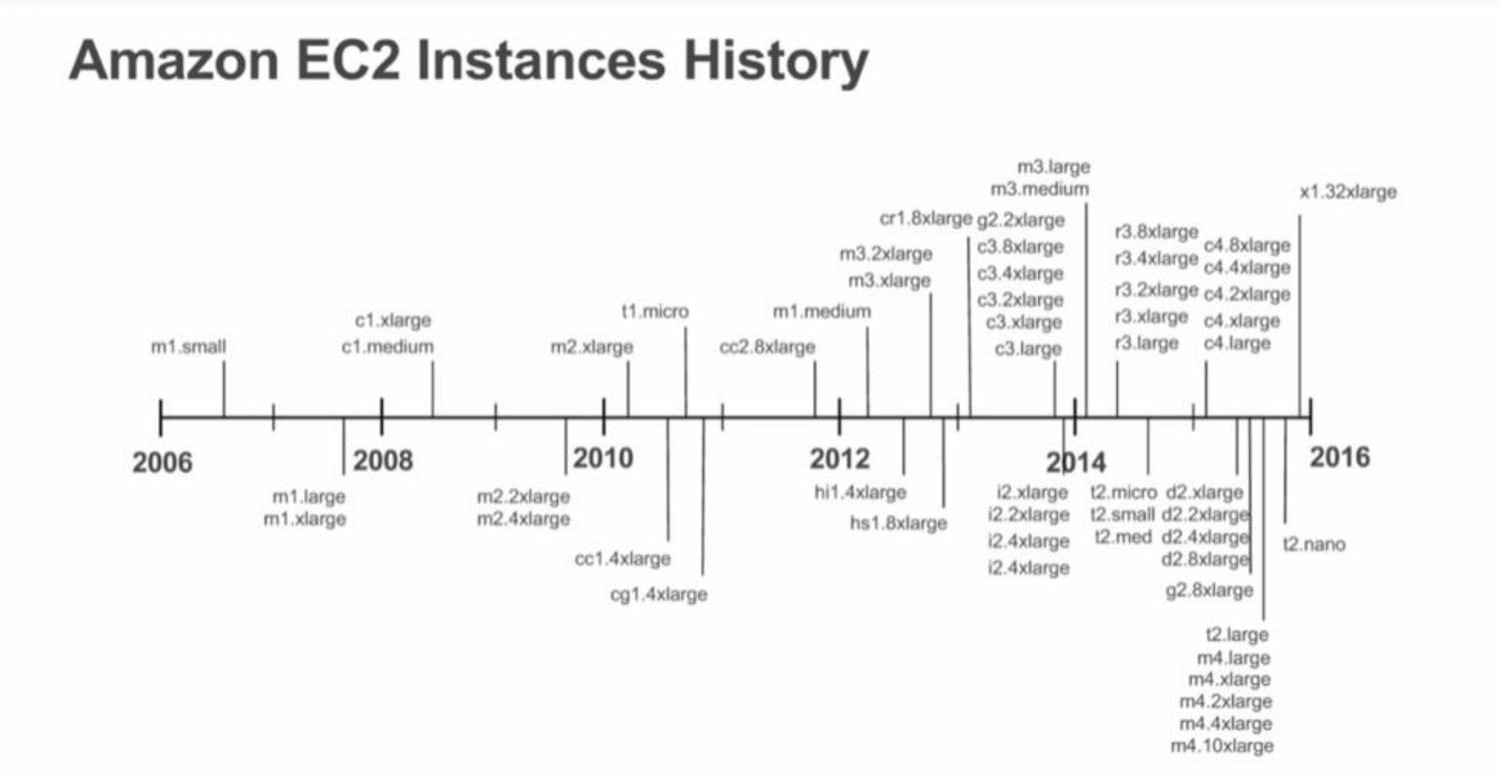
Storage Optimized Instances

- Ideal for workloads needing large amounts of storage space
  - MapReduce workloads, network file systems, etc.
- High disk throughput
- Balance of computation and memory
- 2.3 GHz Intel Xeon E5 2686 v4 processors
- Enhanced networking with ENA

Storage Optimized Instances

Model	vCPU*	Mem (GiB)	Networking Performance	Storage (GB)
h1.2xlarge	8	32	Up to 10 Gigabit	1 x 2,000 HDD
h1.4xlarge	16	64	Up to 10 Gigabit	2 x 2,000 HDD
h1.8xlarge	32	128	10 Gigabit	4 x 2,000 HDD
h1.16xlarge	64	256	25 Gigabit	8 x 2,000 HDD

History of EC2 Instances

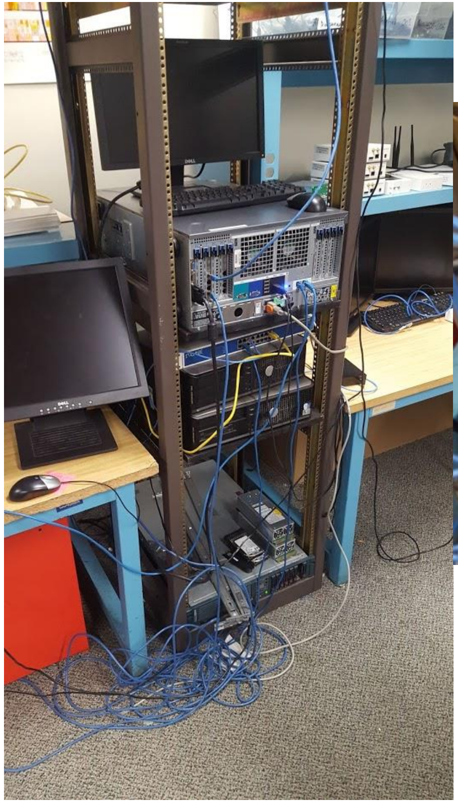


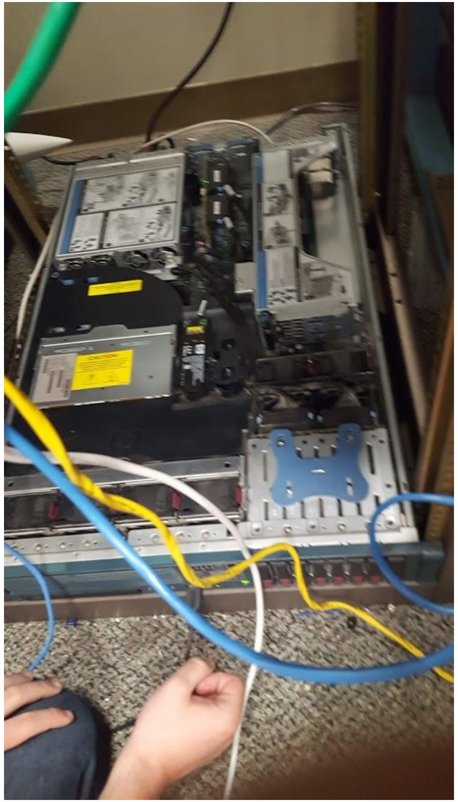
Instance Payment Types

- On-demand
  - Pay for each hour of capacity use
  - Increase/decrease capacity based on computational demand
- Reserved
  - Up to 75% less in price than On-demand
  - Can change OS types or ownerships
- Spot
  - Bid on extra capacity
  - Reduce cost of running applications while increasing capacity and throughput
  - Supports new types of cloud computing

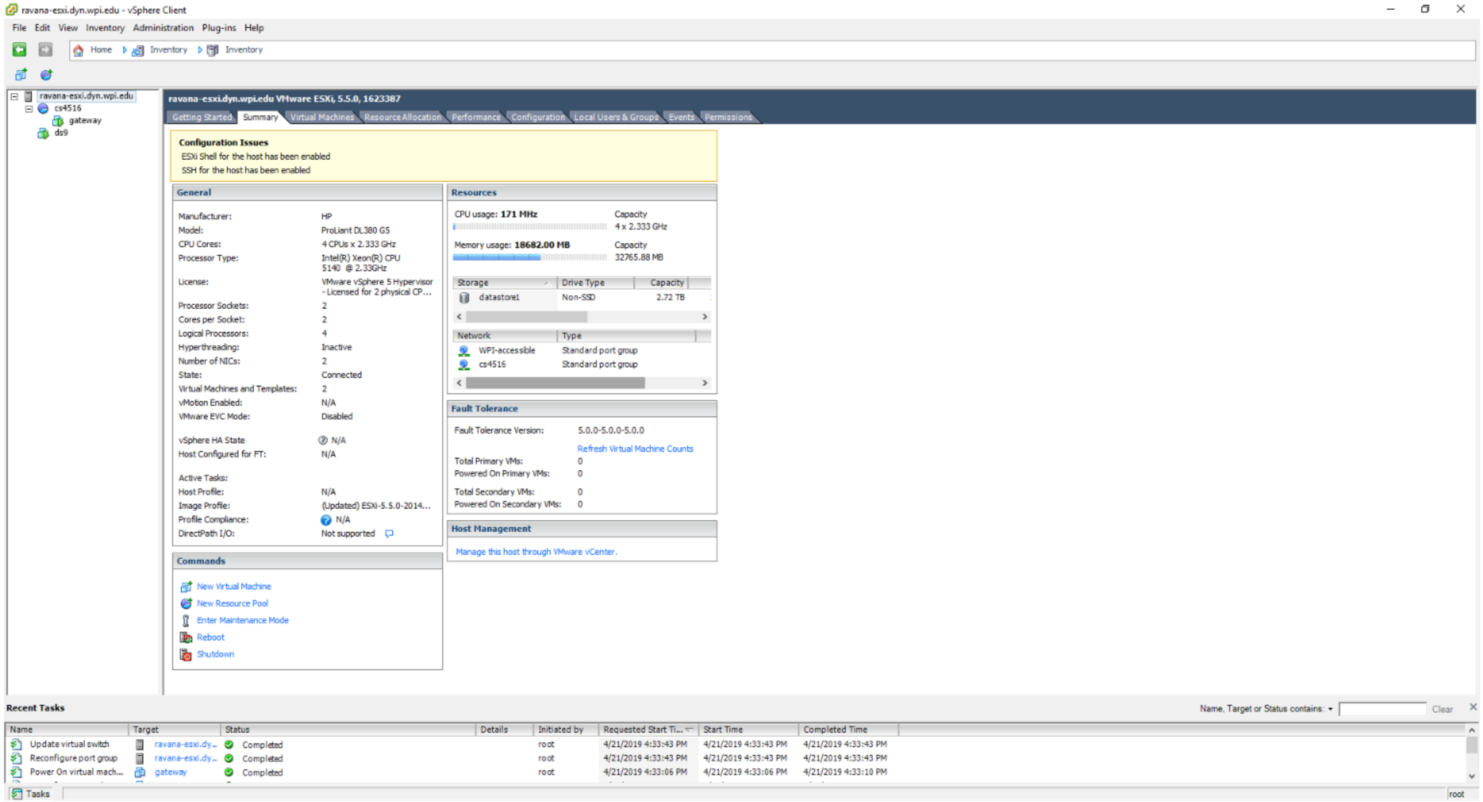
Only one of many services...

Demo



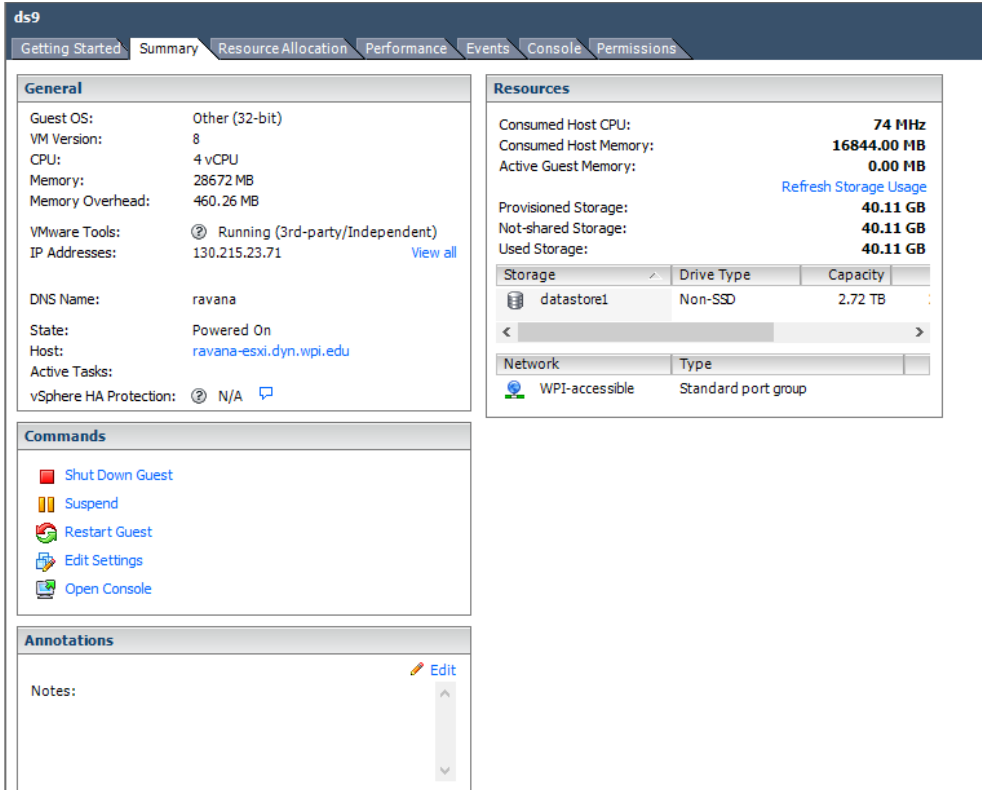


Demo



ravana-esxi.dyn.wpi.edu - vSphere Client
File Edit View Inventory Administration Plug-ins Help
Home Inventory Inventory
ravana-esxi.dyn.wpi.edu
cs4516
gateway
ds9
ravana-esxi.dyn.wpi.edu VMware ESXi, 5.5.0, 1623387
Getting Started Summary Virtual Machines Resource Allocation Performance Configuration Local Users & Groups Events Permissions
Configuration Issues
ESXi Shell for the host has been enabled
SSH for the host has been enabled
General
Manufacturer: HP
Model: ProLiant DL380 G5
CPU Cores: 4 CPUs x 2.333 GHz
Processor Type: Intel(R) Xeon(R) CPU 5140 @ 2.33GHz
License: VMware vSphere 5 Hypervisor - Licensed for 2 physical CP...
Processor Sockets: 2
Cores per Socket: 2
Logical Processors: 4
Hyperthreading: Inactive
Number of NICs: 2
State: Connected
Virtual Machines and Templates: 2
vMotion Enabled: N/A
VMware EVC Mode: Disabled
vSphere HA State N/A
Host Configured for FT: N/A
Active Tasks:
Host Profile: N/A
Image Profile: (Updated) ESXi-5.5.0-2014...
Profile Compliance: N/A
DirectPath I/O: Not supported
Commands
New Virtual Machine
New Resource Pool
Enter Maintenance Mode
Reboot
Shutdown
Resources
CPU usage: 171 MHz Capacity 4 x 2.333 GHz
Memory usage: 18682.00 MB Capacity 32765.88 MB
Storage Drive Type Capacity
datastore1 Non-SSD 2.72 TB
Network Type
WPI-accessible Standard port group
cs4516 Standard port group
Fault Tolerance
Fault Tolerance Version: 5.0.0-5.0.0-5.0.0
Refresh Virtual Machine Counts
Total Primary VMs: 0
Powered On Primary VMs: 0
Total Secondary VMs: 0
Powered On Secondary VMs: 0
Host Management
Manage this host through VMware vCenter.
Recent Tasks
Name, Target or Status contains: Clear
Name Target Status Details Initiated by Requested Start Ti... Start Time Completed Time
Update virtual switch ravana-esxi.dy... Completed root 4/21/2019 4:33:43 PM 4/21/2019 4:33:43 PM 4/21/2019 4:33:43 PM
Reconfigure port group ravana-esxi.dy... Completed root 4/21/2019 4:33:43 PM 4/21/2019 4:33:43 PM 4/21/2019 4:33:43 PM
Power On virtual mach... gateway Completed root 4/21/2019 4:33:06 PM 4/21/2019 4:33:06 PM 4/21/2019 4:33:10 PM
Tasks
root

Demo



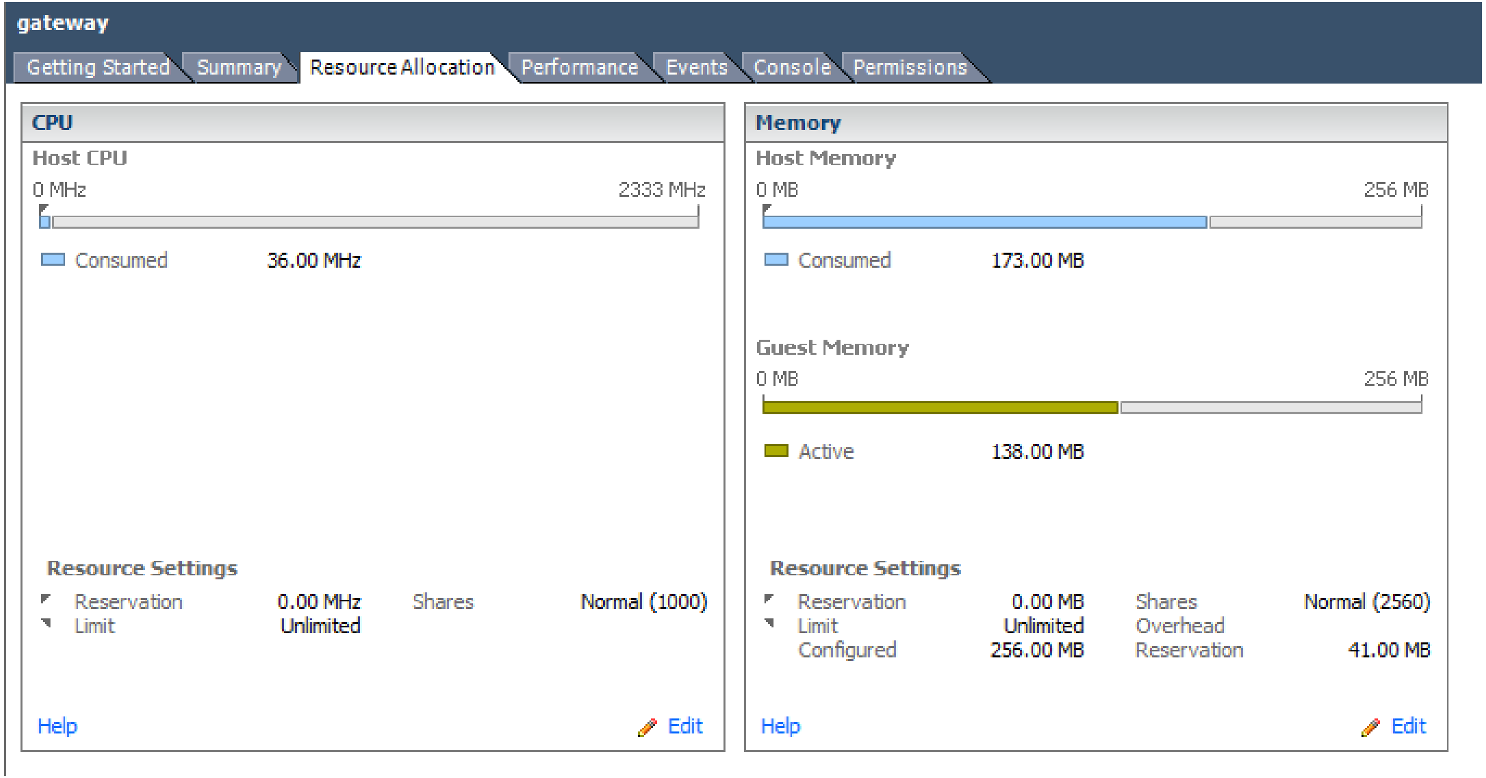
ds9
Getting Started
Summary
Resource Allocation
Performance
Events
Console
Permissions
General
Guest OS: Other (32-bit)
VM Version: 8
CPU: 4 vCPU
Memory: 28672 MB
Memory Overhead: 460.26 MB
VMware Tools: Running (3rd-party/Independent)
IP Addresses: 130.215.23.71
View all
DNS Name: ravana
State: Powered On
Host: ravana-esxi.dyn.wpi.edu
Active Tasks:
vSphere HA Protection: N/A
Commands
Shut Down Guest
Suspend
Restart Guest
Edit Settings
Open Console
Annotations
Edit
Notes:
Resources
Consumed Host CPU: 74 MHz
Consumed Host Memory: 16844.00 MB
Active Guest Memory: 0.00 MB
Refresh Storage Usage
Provisioned Storage: 40.11 GB
Not-shared Storage: 40.11 GB
Used Storage: 40.11 GB
Storage
Drive Type
Capacity
datastore1
Non-SSD
2.72 TB
Network
Type
WPI-accessible
Standard port group

Demo

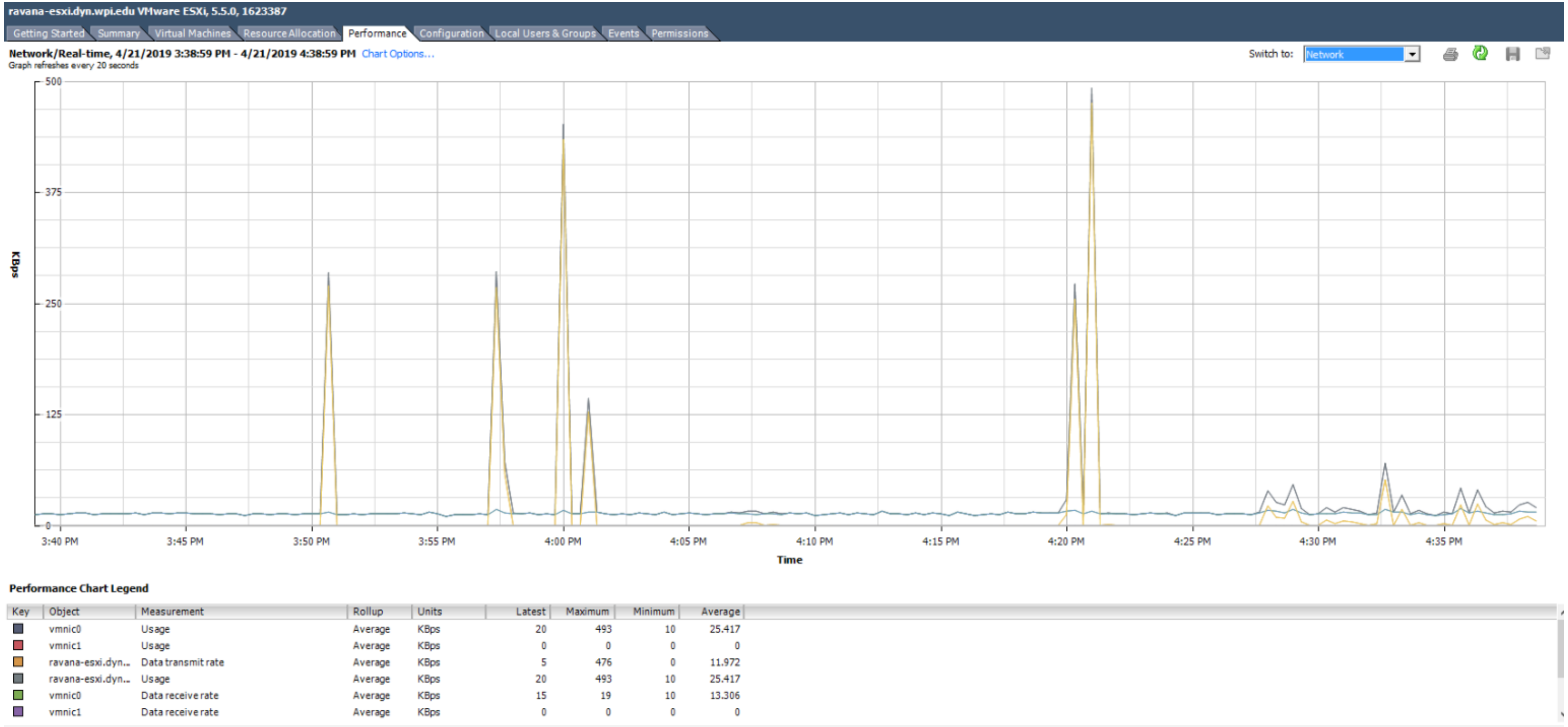
```
 sisko@ravana  ~  echo Hello, world!                                          16:30:05
;echo Hello, world!;echoHello, world!
 sisko@ravana  ~  df                                                          16:30:13
;df;dfFilesystem              1K-blocks     Used Available Use% Mounted on
/dev/mapper/centos-root  10258432  6319360   3939072  62% /
devtmpfs                 14314076        0  14314076   0% /dev
tmpfs                    14325996        0  14325996   0% /dev/shm
tmpfs                    14325996    33632  14292364   1% /run
tmpfs                    14325996        0  14325996   0% /sys/fs/cgroup
/dev/sda1                 1038336   309200    729136  30% /boot
tmpfs                     2865200        0   2865200   0% /run/user/1000
tmpfs                    25165824 10867608  14298216  44% /home/neonnarwhal/public_html
 sisko@ravana  ~  free                                                        16:30:18
;free;free              total        used        free      shared  buff/cache   available
Mem:       28651992      585292    16235012    10902656    11831688    16766952
Swap:       1257468           0     1257468
 sisko@ravana  ~  ifconfig ens160                                             16:30:20
;ifconfig ens160;ifconfigens160: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
        inet 130.215.23.71  netmask 255.255.248.0  broadcast 130.215.23.255
        inet6 fe80::4cb5:ed3:68c8:8930  prefixlen 64  scopeid 0x20<link>
        ether 00:0c:29:a3:1e:7c  txqueuelen 1000  (Ethernet)
        RX packets 156509800  bytes 18937826139 (17.6 GiB)
        RX errors 0  dropped 9  overruns 0  frame 0
        TX packets 3683512  bytes 2443131421 (2.2 GiB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0

 sisko@ravana  ~                                                              16:30:44
```

Demo


gateway
Getting Started
Summary
Resource Allocation
Performance
Events
Console
Permissions
CPU
Host CPU
0 MHz
2333 MHz
Consumed
36.00 MHz
Resource Settings
Reservation
0.00 MHz
Limit
Unlimited
Shares
Normal (1000)
Help
Edit
Memory
Host Memory
0 MB
256 MB
Consumed
173.00 MB
Guest Memory
0 MB
256 MB
Active
138.00 MB
Resource Settings
Reservation
0.00 MB
Limit
Unlimited
Configured
256.00 MB
Shares
Normal (2560)
Overhead
Reservation
41.00 MB
Help
Edit

Demo



Key	Object	Measurement	Rollup	Units	Latest	Maximum	Minimum	Average
	vmnic0	Usage	Average	KBps	20	493	10	25.417
	vmnic1	Usage	Average	KBps	0	0	0	0
	ravana-esxi.dyn...	Data transmit rate	Average	KBps	5	476	0	11.972
	ravana-esxi.dyn...	Usage	Average	KBps	20	493	10	25.417
	vmnic0	Data receive rate	Average	KBps	15	19	10	13.306
	vmnic1	Data receive rate	Average	KBps	0	0	0	0

Thank you!