

Introduction to Cloud Computing



List of Cloud Providers

- AWS: Amazon Web Services
 - EC2: Elastic Compute Cloud
 - S3: Simple Storage Service
 - EBS: Elastic Block Storage
- Microsoft Azure
- Google Cloud/Compute Engine/AppEngine
- Rightscale, Salesforce, EMC, Gigaspaces, 10gen, Datastax, Oracle, VMWare, Yahoo, Cloudera
- And many many more!

Categories of Clouds

- Can be either a (i) public cloud, or (ii) private cloud (iii) hybrid cloud
- Private clouds are accessible only to company employees
- Public clouds provide service to any paying customer:
 - Amazon S3 (Simple Storage Service): store arbitrary datasets, pay per GB-month stored
 - As of 2019: 0.4c-3 c per GB month
 - Amazon EC2 (Elastic Compute Cloud): upload and run arbitrary OS images, pay per CPU hour used
 - As of 2019: 0.2 c per CPU hr to \$7.2 per CPU hr (depending on strength)
 - Google cloud: similar pricing as above
 - Google AppEngine/Compute Engine: develop applications within their appengine framework, upload data that will be imported into their format, and run

Customers Save Time & Money

- Dave Power, Associate Information Consultant at Eli Lilly and Company: “With AWS, Powers said, a new server can be up and running in **three minutes** (it used to take Eli Lilly **seven and a half weeks** to deploy a server internally) and a **64-node Linux cluster** can be online in five minutes (compared with three months internally). ... It's just shy of instantaneous.”
- Ingo Elfering, Vice President of Information Technology Strategy, GlaxoSmithKline: “With Online Services, we are able to reduce our IT **operational costs** by roughly **30%** of what we're spending”
- Jim Swartz, CIO, Sybase: “At Sybase, a private cloud of virtual servers inside its datacenter has saved nearly **\$US2 million annually** since 2006, Swartz says, because the company can share computing power and storage resources across servers.”
- 100s of startups in Silicon Valley can harness large computing resources without buying their own machines.

What is a Cloud?

- It's a cluster!
- It's a supercomputer!
- It's a datastore!
- It's superman!



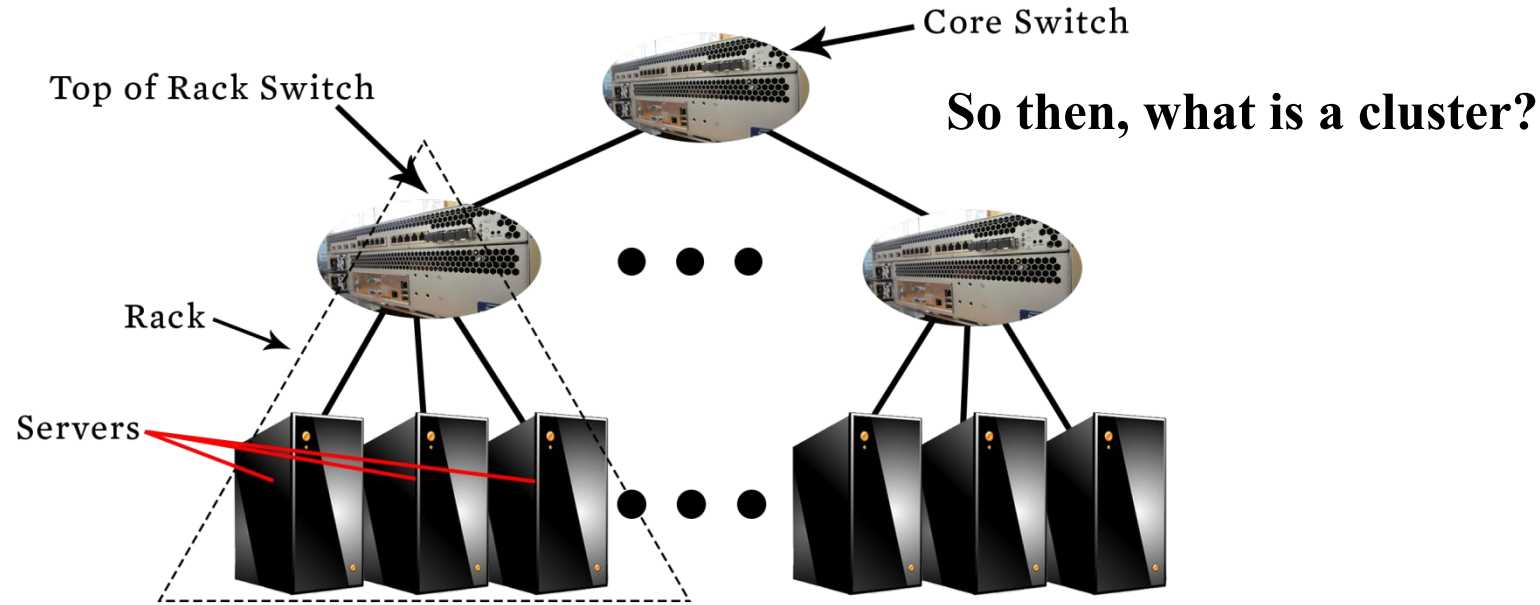
- None of the above
- All of the above

- Cloud = Lots of storage + compute cycles nearby

What is a Cloud?

- A single-site cloud (aka “Datacenter”) consists of
 - Compute nodes (grouped into racks) (2)
 - Switches, connecting the racks
 - A network topology, e.g., hierarchical
 - Storage (backend) nodes connected to the network (3)
 - Front-end for submitting jobs and receiving client requests (1)
 - (1-3: Often called “three-tier architecture”)
 - Software Services
- A geographically distributed cloud consists of
 - Multiple such sites
 - Each site perhaps with a different structure and services

A Sample Cloud Topology



Trends: Technology

- Doubling Periods – storage: 12 mos, bandwidth: 9 mos, and (what law is this?) cpu compute capacity: 18 mos
- Then and Now
 - Bandwidth
 - 1985: mostly 56Kbps links nationwide
 - 2015: Tbps links widespread
 - Disk capacity
 - Today's PCs have TBs, far more than a 1990 supercomputer

Trends: Users

- Then and Now

- Biologists:

- 1990: were running small single-molecule simulations
 - Today: CERN's Large Hadron Collider producing many PB/year

Four Features New in Today's Clouds

I. Massive scale.

II. On-demand access: Pay-as-you-go, no upfront commitment.

- And anyone can access it

III. Data-intensive Nature: What was MBs has now become TBs, PBs and XBs.

- Daily logs, forensics, Web data, etc.
- Humans have data numbness: Wikipedia (large) compressed is only about 10 GB!

IV. New Cloud Programming Paradigms: MapReduce/Hadoop, NoSQL/Cassandra/MongoDB and many others.

- High in accessibility and ease of programmability
- Lots of open-source

Combination of one or more of these gives rise to novel and unsolved distributed computing problems in cloud computing.

I. Massive Scale

- Facebook [GigaOm, 2012]
 - 30K in 2009 -> 60K in 2010 -> 180K in 2012
- Microsoft [NYTimes, 2008]
 - 150K machines
 - Growth rate of 10K per month
 - 80K total running Bing
 - In 2013, Microsoft Cosmos had 110K machines (4 sites)
- Yahoo! [2009]:
 - 100K
 - Split into clusters of 4000
- AWS EC2 [Randy Bias, 2009]
 - 40K machines
 - 8 cores/machine
- eBay [2012]: 50K machines
- HP [2012]: 380K in 180 DCs
- Google [2011, Data Center Knowledge] : 900K

Quiz: Where is the World's Largest Datacenter?

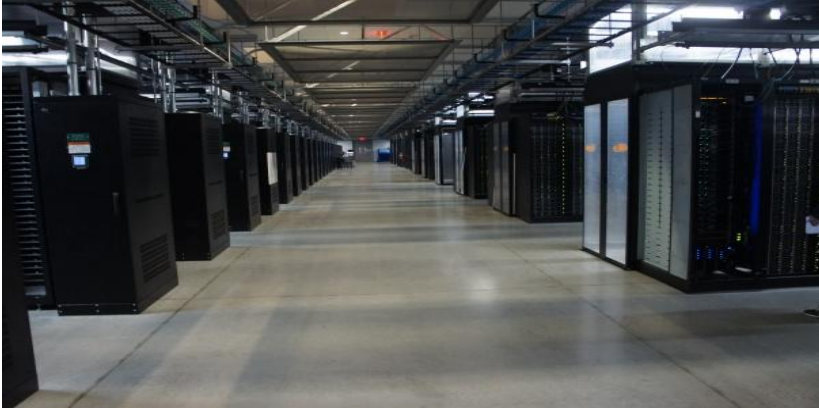
- (2018) China Telecom. 10.7 Million sq. ft.
- (2017) “The Citadel” Nevada. 7.2 Million sq. ft.
- (2015) In Chicago!
 - 350 East Cermak, Chicago, 1.1 MILLION sq. ft.
 - Shared by many different “carriers”
 - Critical to Chicago Mercantile Exchange
- See:
 - <https://www.gigabitmagazine.com/top10/top-10-biggest-data-centres-world>
 - <https://www.racksolutions.com/news/data-center-news/top-10-largest-data-centers-world/>

What does a datacenter look like from inside?

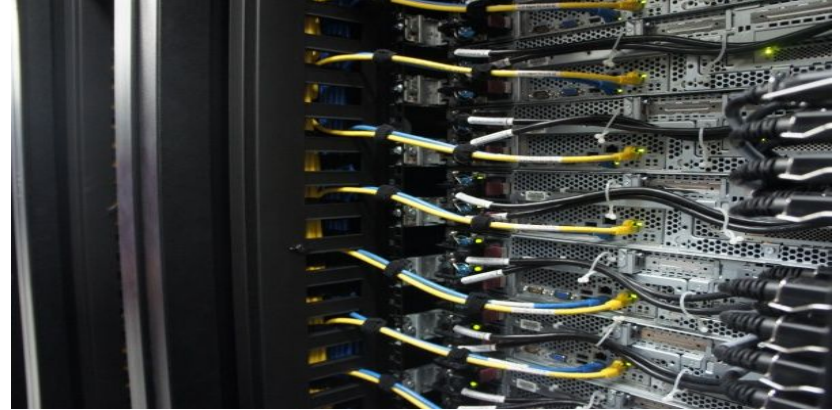
- A virtual walk through a datacenter
- Reference:

<http://gigaom.com/cleantech/a-rare-look-inside-facebooks-oregon-data-center-photos-video/>

Servers



Front



Back



In



Some highly secure (e.g., financial info) ¹⁵

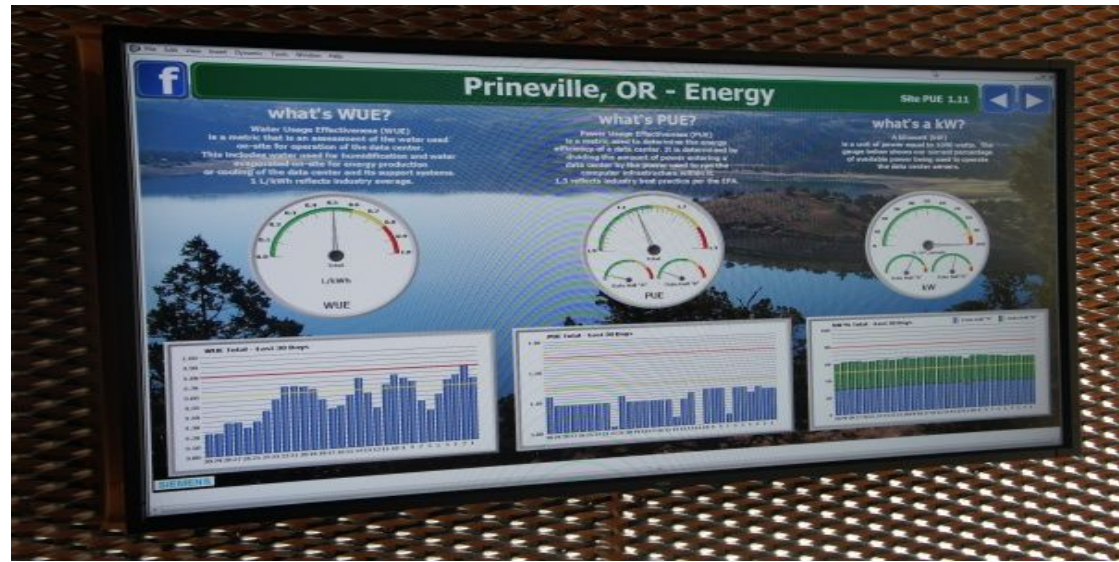
Power



Off-site

On-site

- $WUE = \text{Annual Water Usage} / \text{IT Equipment Energy (L/kWh)}$ – low is good
- $PUE = \text{Total facility Power} / \text{IT Equipment Power}$ – low is good
(e.g., Google~1.1)



Cooling



Air sucked in from top (also, Bugzappers)



Water purified



Water sprayed into air



15 motors per server bank

Extra - Data center Videos to Watch

- Microsoft GFS Datacenter Tour (Youtube)
 - <http://www.youtube.com/watch?v=hOxA111pQIw>
- Timelapse of a Datacenter Construction on the Inside (Fortune 500 company)
 - <http://www.youtube.com/watch?v=ujO-xNvXj3g>

II. On-demand access: *aaS Classification

- PaaS: Platform as a Service
 - You get access to flexible computing and storage infrastructure, coupled with a software platform (often tightly coupled)
 - Ex: Google's AppEngine (Python, Java, Go)
- SaaS: Software as a Service
 - You get access to software services, when you need them. Often said to subsume SOA (Service Oriented Architectures).
 - Ex: Google docs, MS Office 365 Online

II. On-demand access: *aaS Classification

On-demand: renting a cab vs. (previously) renting a car, or buying one. E.g.:

- AWS Elastic Compute Cloud (EC2): a few cents to a few \$ per CPU hour
- AWS Simple Storage Service (S3): a few cents per GB-month
- HaaS: Hardware as a Service
 - You get access to barebones hardware machines, do whatever you want with them, Ex: Your own cluster
 - Not always a good idea because of security risks
- IaaS: Infrastructure as a Service
 - You get access to flexible computing and storage infrastructure. Virtualization is one way of achieving this (cgroups, Kubernetes, Dockers, VMs,...). Often said to subsume HaaS.
 - Ex: Amazon Web Services (AWS: EC2 and S3), OpenStack, Eucalyptus, Rightscale, Microsoft Azure, Google Cloud.