Selected Topics in CN

Introduction to Clouds

This slide is based on CS 525: Advanced Distributed Systems 2015, UIUC

The Hype

- Forrester in 2010 – Cloud computing will go from $40.7 billion in 2010 to $241 billion in 2020.
- Gartner in 2009 - Cloud computing revenue will soar faster than expected and will exceed $150 billion by 2013. It will represent 19% of IT spending by 2015.
- IDC in 2009: “Spending on IT cloud services will triple in the next 5 years, reaching $42 billion.”
- Companies and even Federal/state governments using cloud computing now: fbo.gov

Many cloud providers

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

Two categories of clouds

- Can be either a (i) public cloud, or (ii) private 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
  - Amazon EC2 (Elastic Compute Cloud): upload and run arbitrary OS images, pay per CPU hour used
  - Google AppEngine/Compute Engine: develop applications within their appengine framework, upload data that will be imported into their format, and run

Customers save time and $$$

- 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.

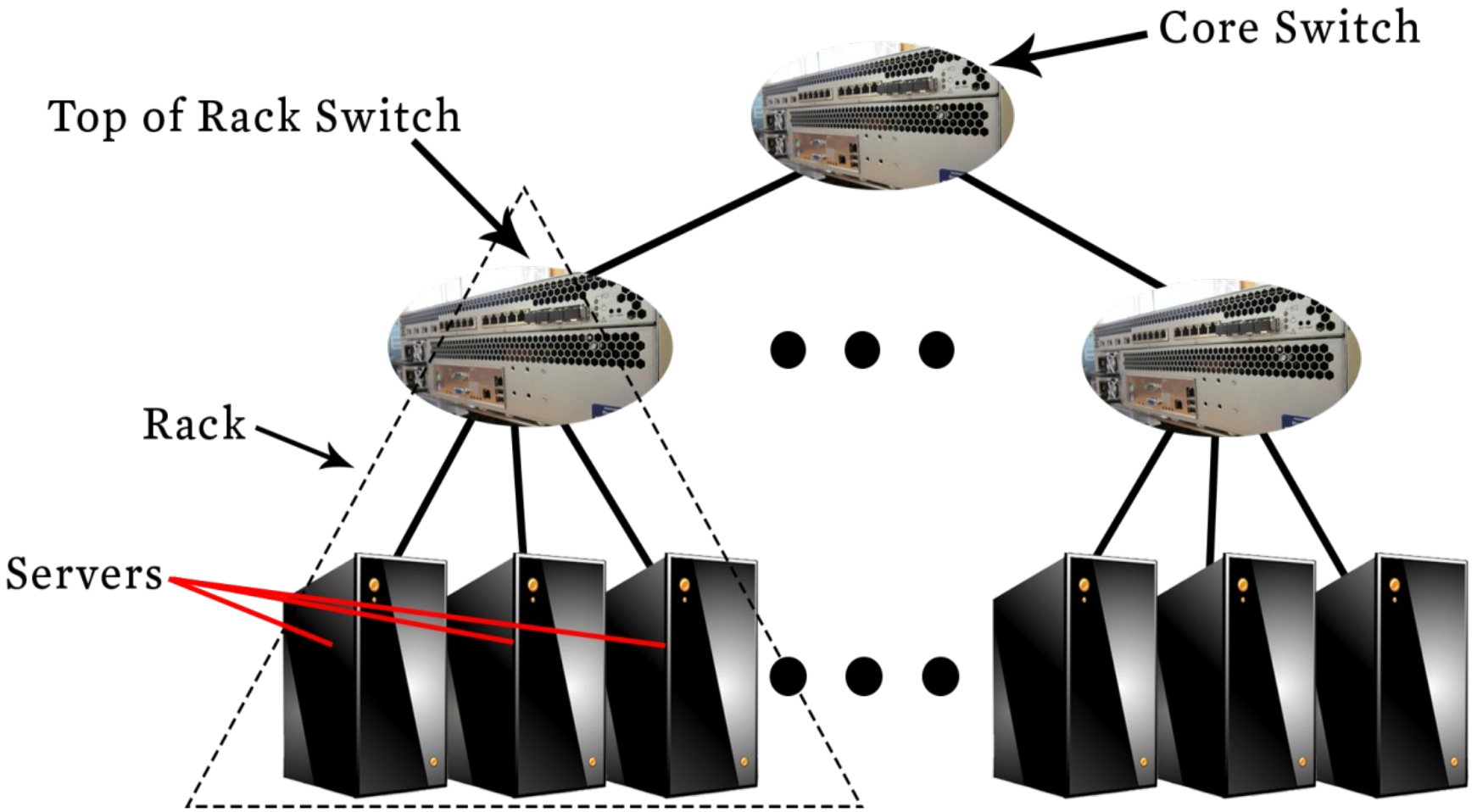
What is a cloud?

Cloud = Lots of storage + compute cycles nearby

What is a cloud?

- A single-site cloud (aka “Datacenter”) consists of
  - Compute nodes (grouped into racks)
  - Switches, connecting the racks
  - A network topology, e.g., hierarchical
  - Storage (backend) nodes connected to the network
  - Front-end for submitting jobs and receiving client requests
  - (Often called 3-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



Four Features New in Today's Clouds

- Massive scale.
- On-demand access: Pay-as-you-go, no upfront commitment.
  - And anyone can access it
- 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!
- New Cloud Programming Paradigms: MapReduce/Hadoop, NoSQL/Cassandra/MongoDB and many others.
  - High in accessibility and ease of programmability
  - Lots of open-source

(I) Massive Scale

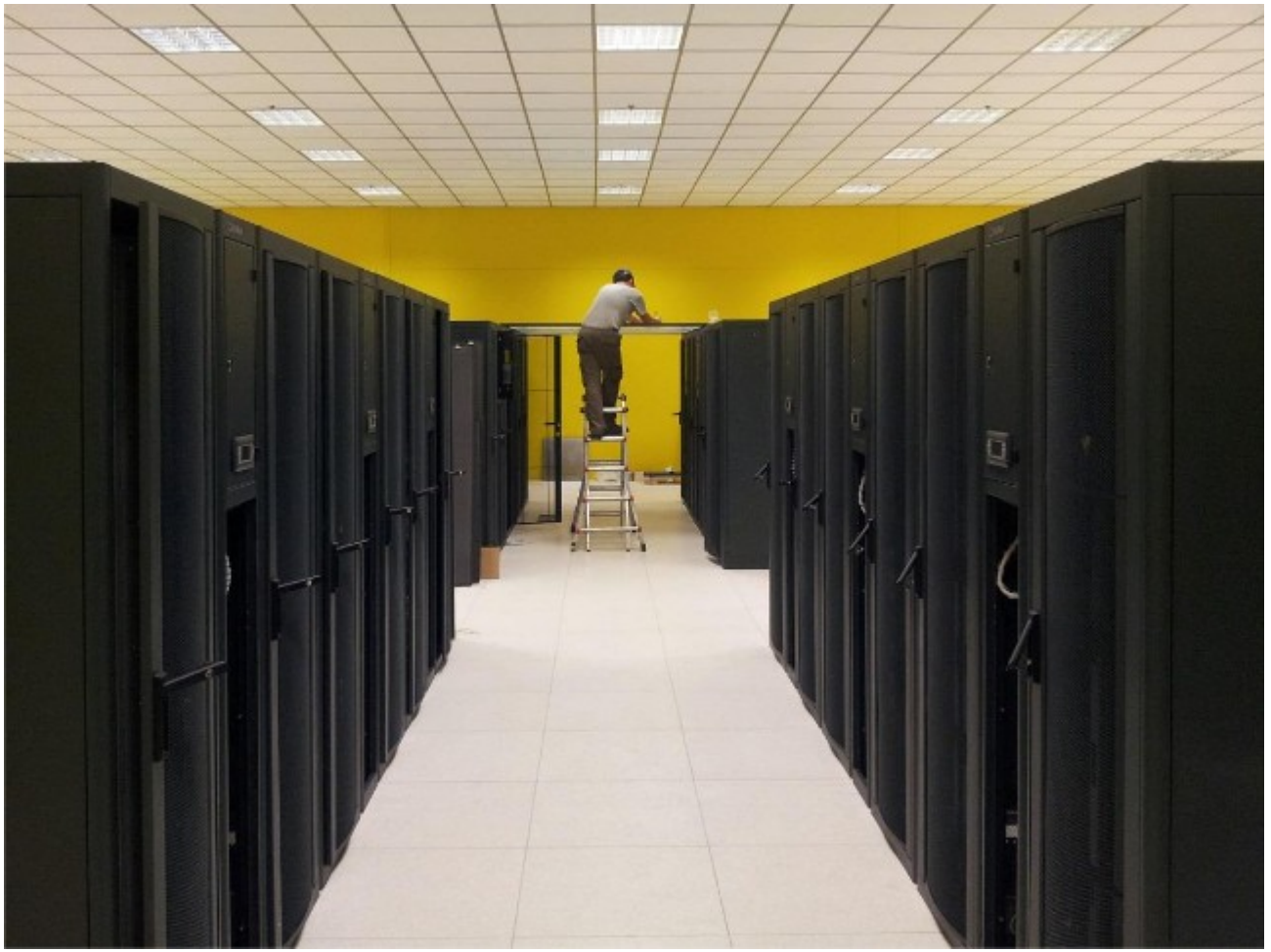
- Facebook [GigaOm, 2012]
  - 30K in 2009 -> 60K in 2010 -> 180K in 2012
- Microsoft [NYTimes, 2008]
  - 150K machines
  - Growth rate of 10K per month
  - 80K total running Bing
- Yahoo! [2009]:
  - 100K
  - Split into clusters of 4000
- AWS EC2 [Randy Bias, 2009]
  - 40K machines
  - 8 cores/machine
- eBay [2012]: 50K machines
- HP [2012]: 380K in 180 DCs
- Google: A lot

What does a datacenter look like from inside?

Reference:

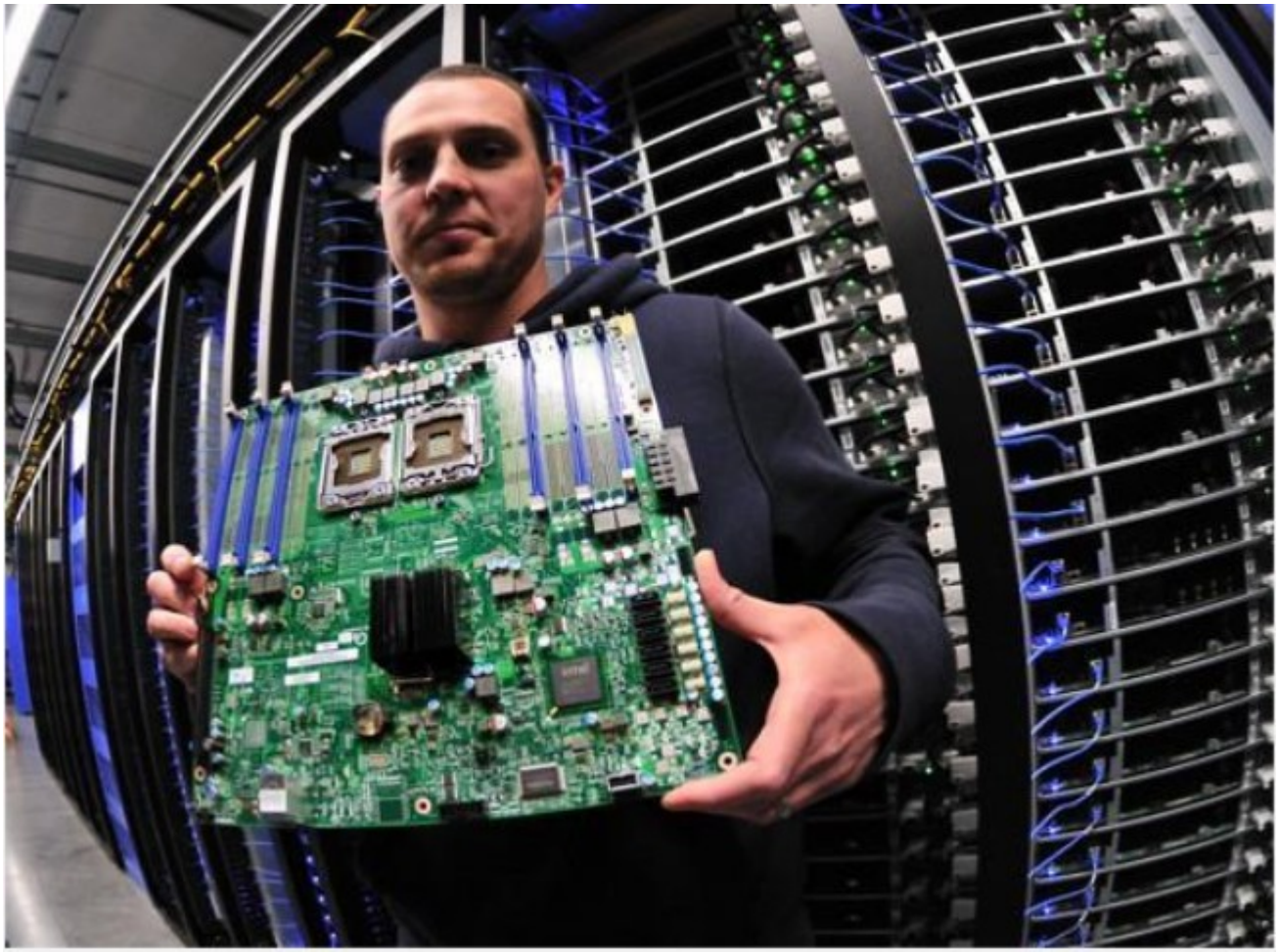
http://gigaom.com/cleantech/a-rare-look-inside-facebooks-oregon-data-center-photos-video/

Front



Back

Inside



Some highly secure (e.g. finance info)

Power

Off-site

On-site

Cooling

Air sucked in
Combined with purified water
Moves cool air through system

Fun Videos

- Microsoft GFS Datacenter Tour (Youtube)

  http://www.youtube.com/watch?v=hOxA1l1pQIw

- 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

- 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 to a few $ 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 (what's another way, e.g., using Linux). Often said to subsume HaaS.
  - Ex: Amazon Web Services (AWS: EC2 and S3), Eucalyptus, Rightscale, Microsoft Azure, Google Compute Engine, OpenStack.

(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 on demand

(III) Data-intensive Computing

- Computation-Intensive Computing
  - Example areas: MPI-based, High-performance computing, Grids
  - Typically run on supercomputers (e.g., NCSA Blue Waters)
- Data-Intensive
  - Typically store data at datacenters
  - Use compute nodes nearby
  - Compute nodes run computation services
- In data-intensive computing, the focus shifts from computation to the data: CPU utilization no longer the most important resource metric, instead I/O is (disk and/or network)

(IV) New Cloud Programming Paradigms

Easy to write and run highly parallel programs in new cloud programming paradigms:

- Google: MapReduce and Sawzall
- Amazon: Elastic MapReduce service (pay-as-you-go)
- Google (MapReduce)
  - Indexing: a chain of 24 MapReduce jobs
  - ~200K jobs processing 50PB/month (in 2006)
- Yahoo! (Hadoop + Pig)
  - WebMap: a chain of several MapReduce jobs
  - 300 TB of data, 10K cores, many tens of hours
- Facebook (Hadoop + Hive)
  - ~300TB total, adding 2TB/day (in 2008)
  - 3K jobs processing 55TB/day
- Similar numbers from other companies, e.g., Yieldex, eharmony.com, etc.
- NoSQL: MySQL is an industry standard, but Cassandra is 2400 times faster!

Two Categories of Clouds

- Can be either a (i) public cloud, or (ii) private cloud
  - Private clouds are accessible only to company employees
  - Public clouds provide service to any paying customer

- You’re starting a new service/company: should you use a public cloud or purchase your own private cloud?

Single site Cloud: to Outsource or Own?

- Medium-sized organization: wishes to run a service for M months
  - Service requires 128 servers (1024 cores) and 524 TB
  - Same as UIUC CCT (Cloud Computing Testbed) cloud site
- Outsource (e.g., via AWS): monthly cost
  - S3 costs: $0.12 per GB month. EC2 costs: $0.10 per CPU hour (costs from 2009)
  - Storage = $ 0.12 X 524 X 1000 ~ $62 K
  - Total = Storage + CPUs = $62 K + $0.10 X 1024 X 24 X 30 ~ $136 K
- Own: monthly cost
  - Storage ~ $349 K / M
  - Total ~ $ 1555 K / M + 7.5 K (includes 1 sysadmin / 100 nodes)
  - using 0.45:0.4:0.15 split for hardware:power:network and 3 year lifetime of hardware

Single site Cloud: to Outsource or Own?

- Breakeven analysis: more preferable to own if:

  - $349 K / M < $62 K (storage)

  - $ 1555 K / M + 7.5 K < $136 K (overall)

- Breakeven points
  - M > 5.55 months (storage)
  - M > 12 months (overall)

- As a result
  - Startups use clouds a lot
  - Cloud providers benefit monetarily most from storage

Sum up

- Clouds build on many previous generations of distributed systems
- Especially the timesharing and data processing industry of the 1960-70s.
- Need to identify unique aspects of a problem to classify it as a new cloud computing problem
- Scale, On-demand access, data-intensive, new programming
- Otherwise, the solutions to your problem may already exist!