Cloud Computing

CSC 375, Fall 2016

"When you move to the cloud, companies don't expect a multi-hundred-million-dollar project to make their CRM from Salesforce work with ERP from Oracle. We have to make that implementation work out of the box."

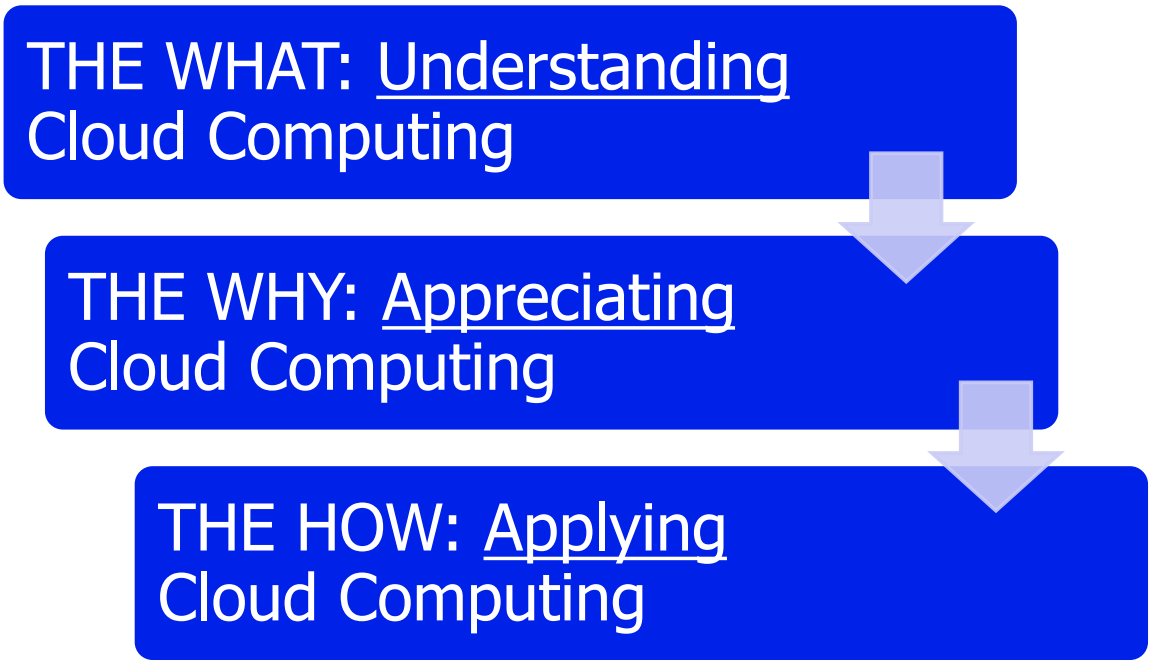
-Larry Ellison, CEO, Oracle

Why Cloud Computing?

- In a recent survey, just 22% of Americans understood the concept of "the cloud"
- There is still a wide gap between the perceptions and realities of cloud computing

http://mashable.com/2012/08/29/clueless-cloud/

Outline



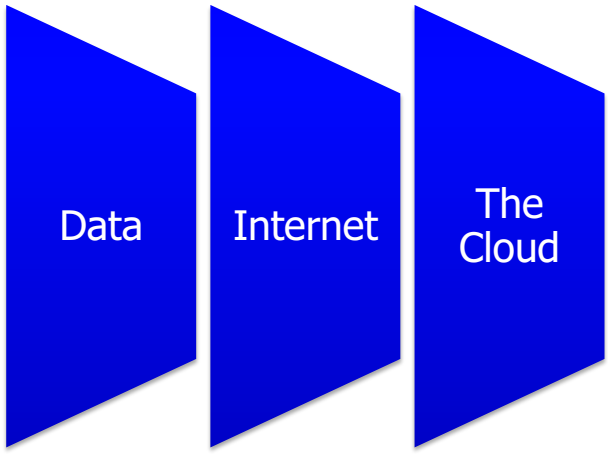
Cloud Computing: Definition

- "The practice of using a network of remote servers hosted on the Internet to store, manage, and process data, rather than a local server or a personal computer"

http://www.us-cert.gov/reading_room/USCERT-CloudComputingHuthCebula.pdf
http://www.blippitt.com/2010-new-oxford-american-dictionary-new-words/

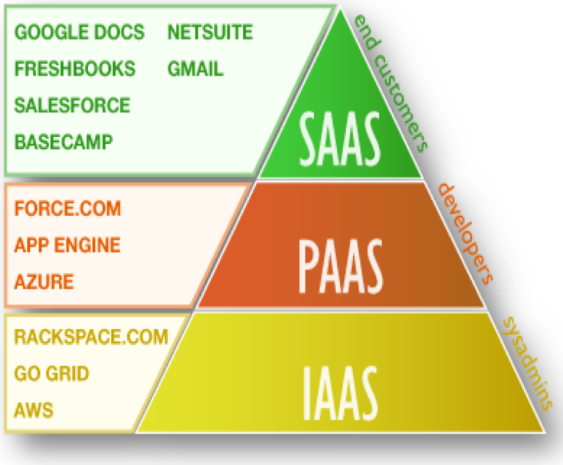
Data, Internet, and the Cloud

- Instead of storing data on individual physical machines, data is stored on the Internet
- Cloud can be Public, Private, Community, or Hybrid



Types of Cloud Computing

- Software as a Service (SaaS)
  - Access to resources and applications
- Platform as a Service (PaaS)
  - Access to development and operational components
- Infrastructure as a Service (IaaS)
  - Completely outsources needed storage and resources



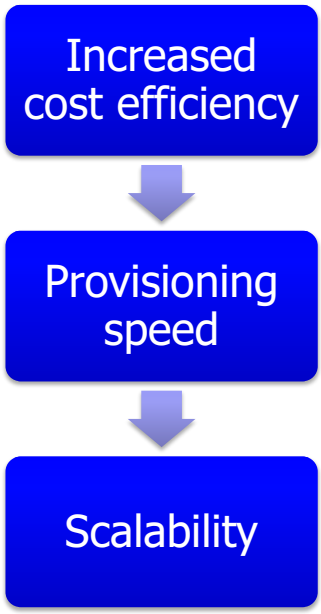
The WHY: Why are so many companies moving to the cloud?

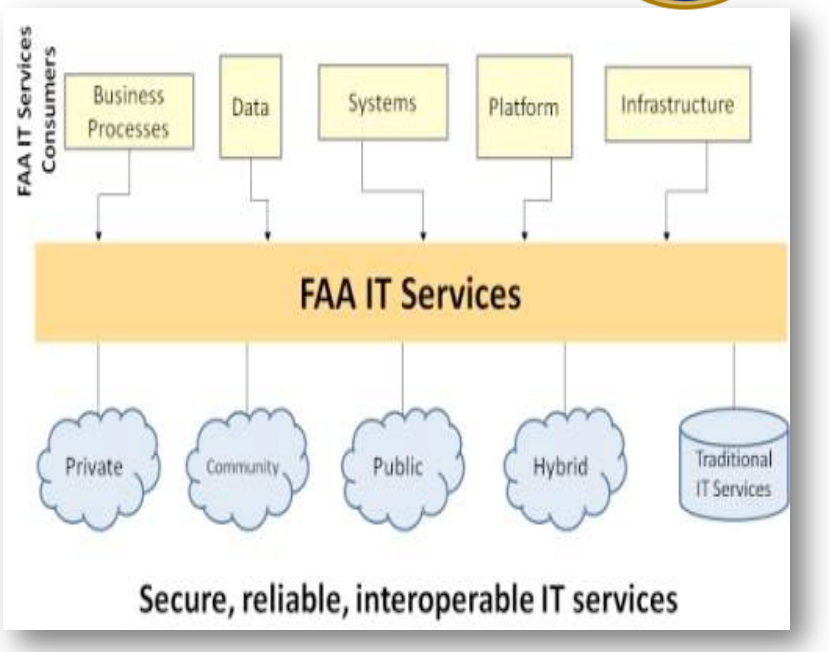
Benefits of Cloud Computing

Availability	• Access data anywhere, anytime through standard internet connection
Cost reductions	• Pay as you go model • Savings on maintenance, repairs, and upgrades
Mobility	• Can access data and application from various devices • Connecting employees, partners, and suppliers globally
Scalability	• Users have access to resources that scale quickly based on their demand

http://www.businessweek.com/magazine/toc/09_24/B4135cloud_computing.htm
http://www.sei.cmu.edu/library/assets/whitepapers/Cloudcomputingbasics.pdf

Putting the Cloud to Work: FAA





http://www.fastcompany.com/1843699/air-traffic-cloud

Putting the Cloud to Work:

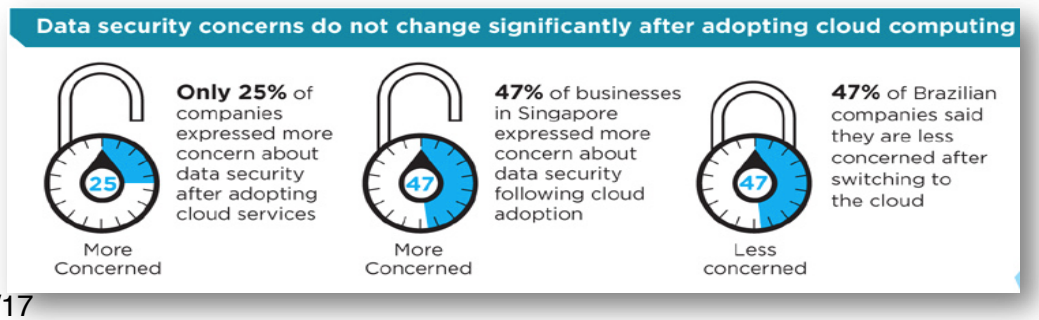
- IHG: 9 hotel brands, 660,000 rooms, 153 Million guests a year
- Believer in major cloud vendors:
  - Amazon Web Services
  - Salesforce CRM
- Operating their own private cloud:
  - "Camelot" supported by cloud services from HP & Verizon
  - Runs IHG's: Revenue Mgmt System, Loyalty program, promotions
- Continuing to Reach for the Cloud

http://www.informationweek.com/hardware/data-centers/4-companies-getting-real-results-from-cl/229000706

The How: Addressing Risks and Implementing a Strategy

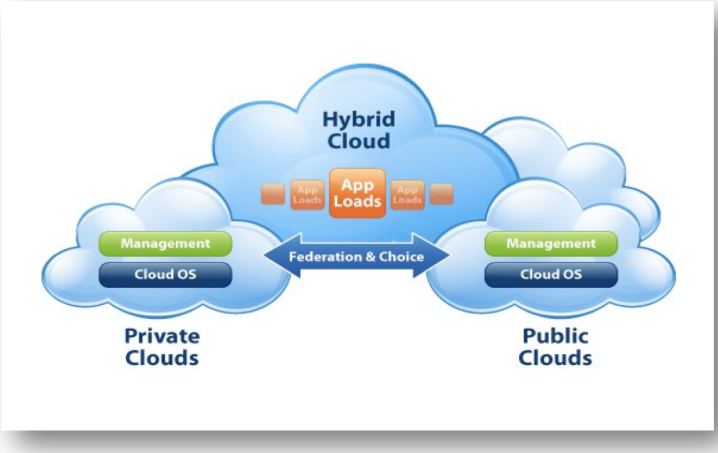
Issues with Cloud Computing: Security and Privacy

- Does the cloud back up your data?
- Is your data always safe?
- Can you conduct business abroad?
- Who is given access to your data?
- Who else is on your server?



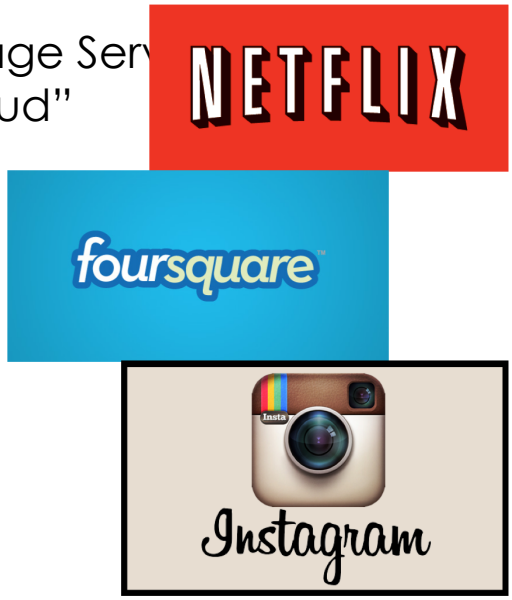
Issues with Cloud Computing: Security and Privacy Solutions

- Triage, Secure, and Back Up
- Authentication & Authorization
- "Private Cloud" vs. "Public Cloud"
- Hybrid Approach



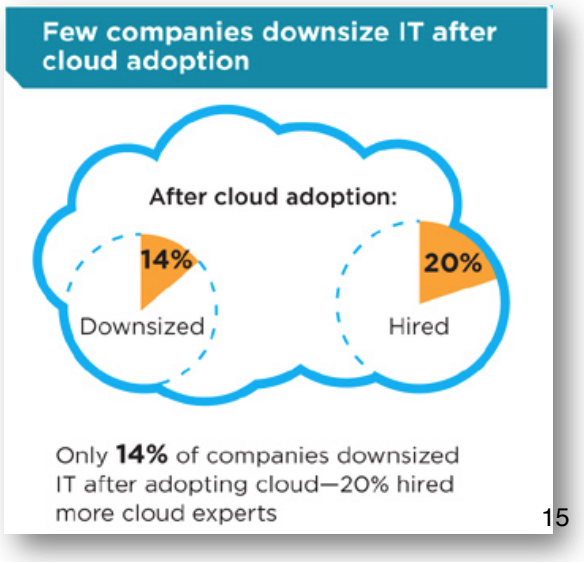
Issues with Cloud Computing: Reliability

- Past Outages
  - Amazon's "Simple Storage Ser
    "Elastic Computing Cloud"
  - Microsoft & Google
- Prevention & Strategic
  - Choice of server
  - Think ahead
  - Hybrid Approach



Issues with Cloud Computing: Downsizing Workforces?

- Increasing need for cloud experts and third-party companies
- Increase in workforce leads to increase in efficiency



Appendix

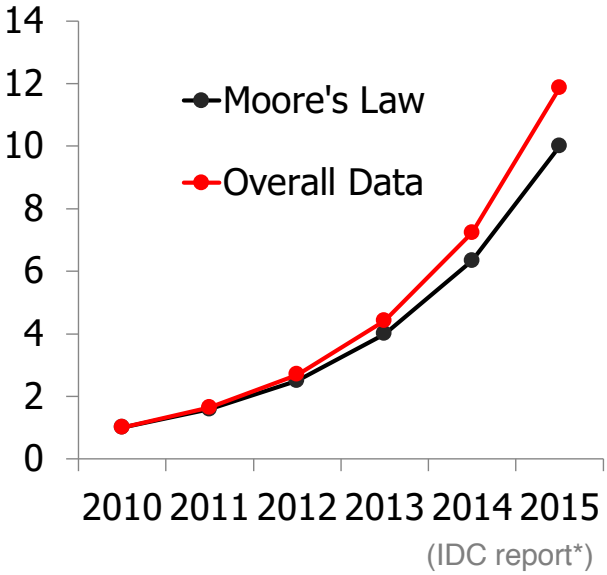
Data Centers

Datacenter evolution

Facebook's daily logs: 60 TB

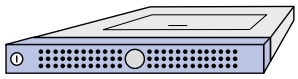
1000 genomes project: 200 TB

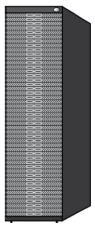
Google web index: 10+ PB

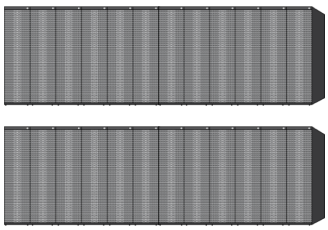


Slide from Ion Stoica

Datacenter Organization







Single server:

- 8-24 cores
- DRAM: 16-64GB @ 100ns
- Disk: 2 TB @10ms

Rack:

- 50 machines
- DRAM: 800-3200GB @ 300 µs
- Disk: 100TB @ 10ms

Row/cluster:

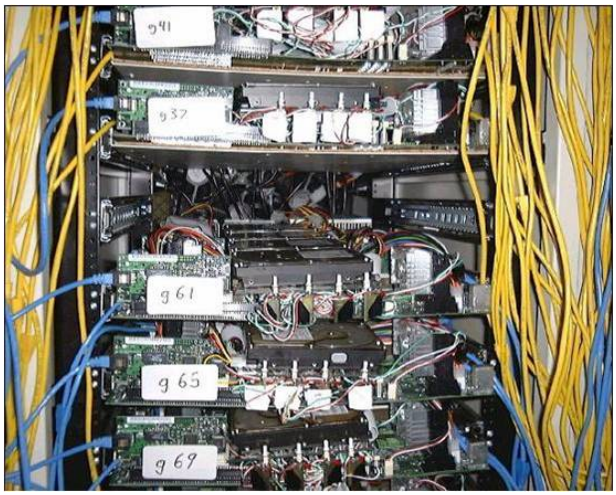
- **30+ racks**
- **DRAM: 24-96TB @ 500 µs**
- **Disk: 3 PB @ 10ms**

Sun Containers

Sun Containers, cont'd

Google 1997

Google 2001



Commodity CPUs

Lots of disks

Low bandwidth network

Cheap !

Google Data Center 2015

Google Modular Data Center

- In 2009, Google revealed the hardware at the core of its operation
  - each server has its own 12-volt battery
  - Data centers are composed of standard shipping containers--each with 1,160 servers and a power consumption that can reach 250 kilowatts.

Google Modular Data Center

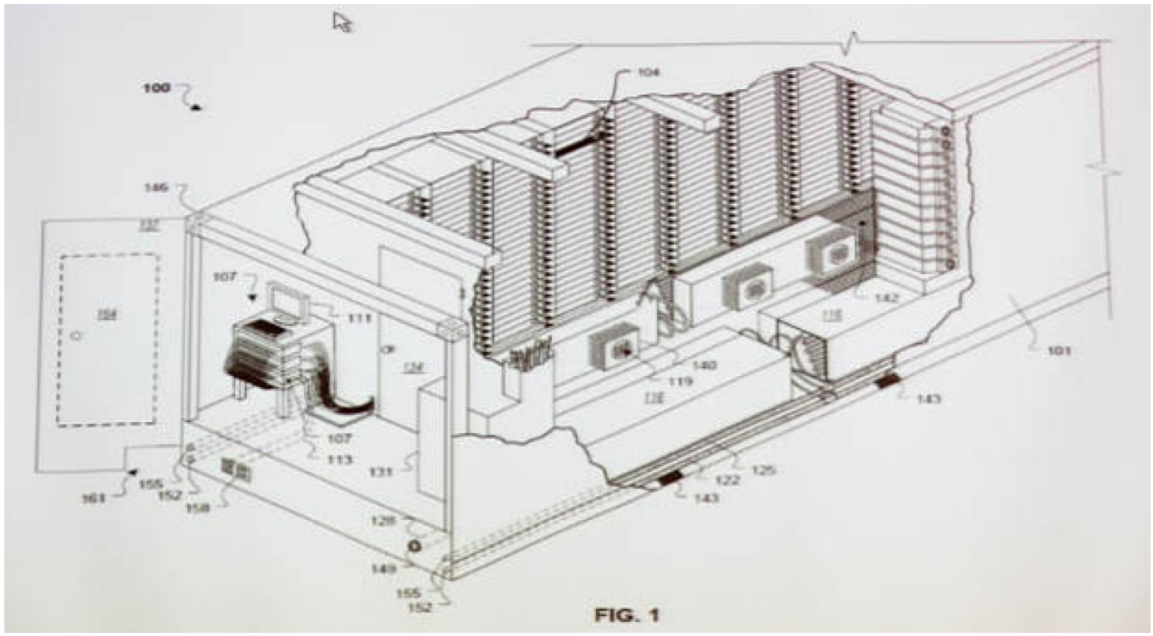
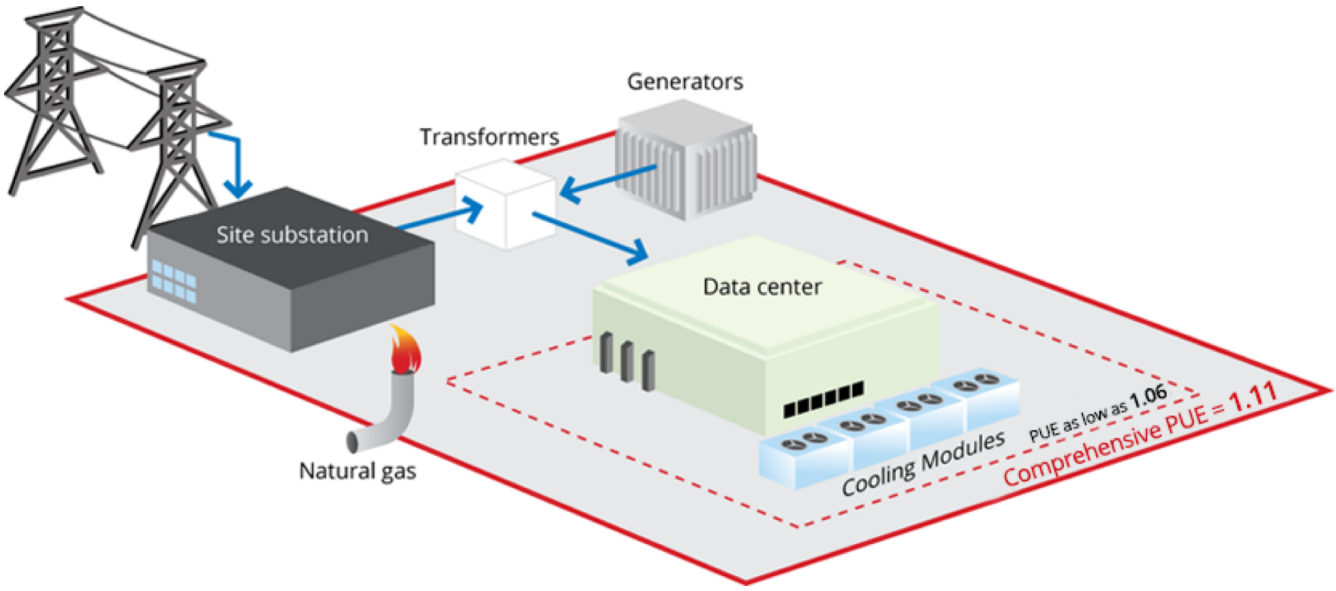


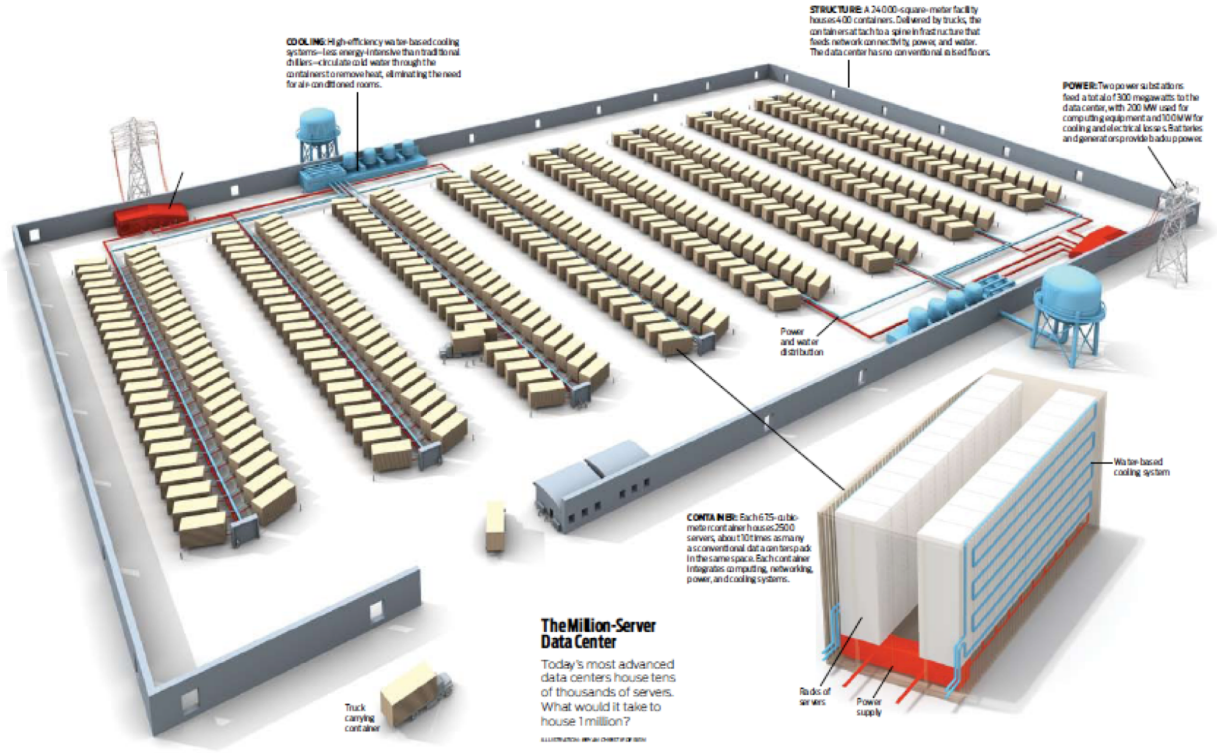
FIG. 1

Google Containers

Google Server Room



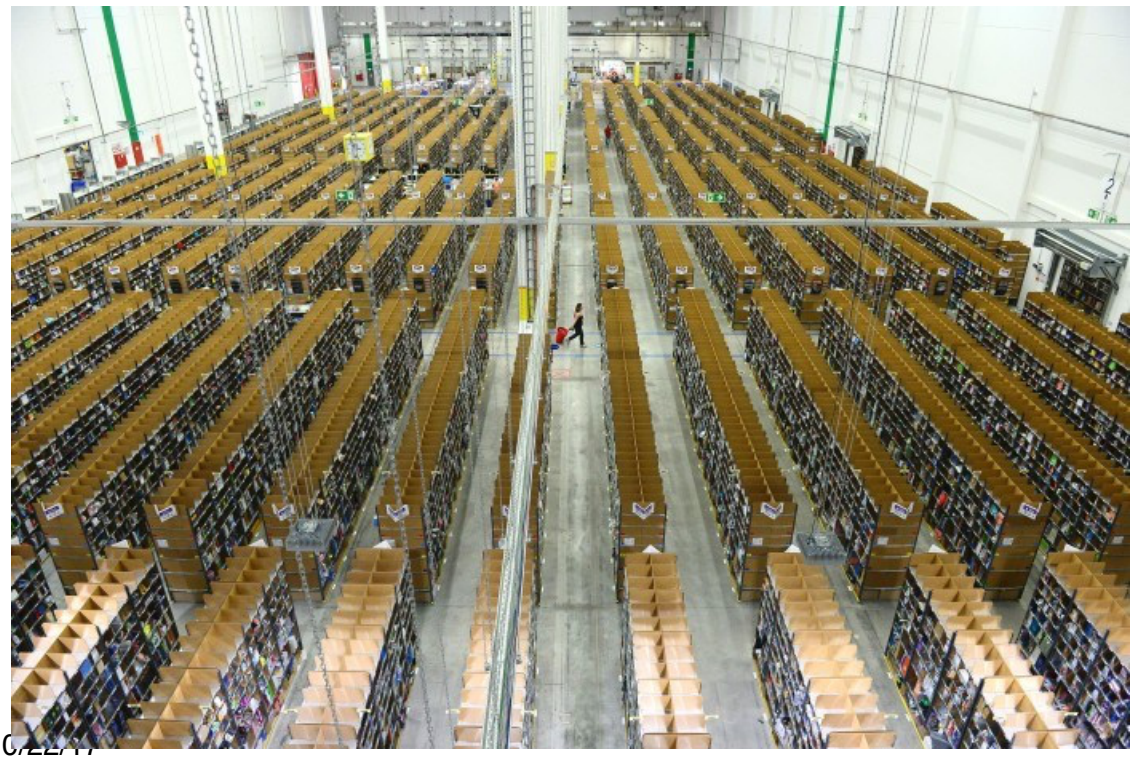
Microsoft Containers



Microsoft Containers, cont'd

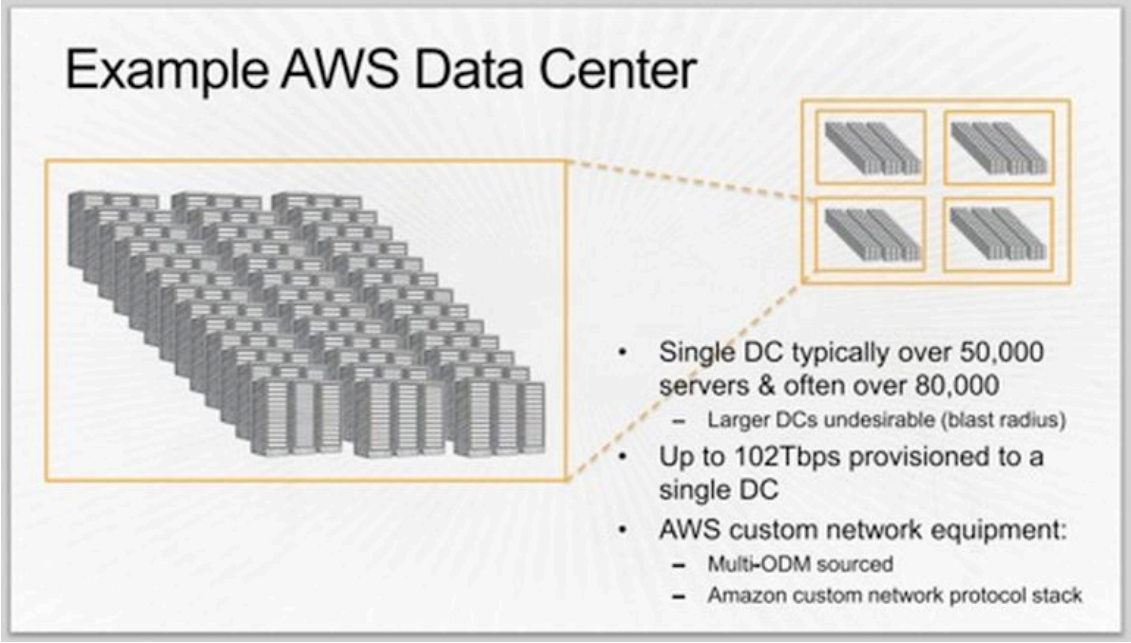
Apple Data Center

Amazon Data Center



Amazon Data Center [Dublin, 28335 m²]

Amazon Data Center



Amazon Data Centers

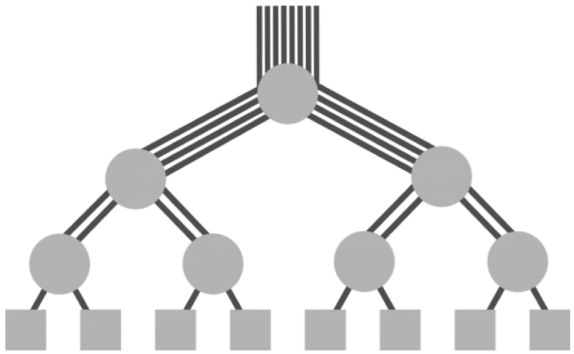
Amazon EC2

Machine	Memory (GB)	Compute Units (ECU)	Local Storage (GB)	Cost / hour
t1.micro	0.615	2	0	$0.02
m1.xlarge	15	8	1680	$0.48
cc2.8xlarge	60.5	88 (Xeon 2670)	3360	$2.40

1 ECU = CPU capacity of a 1.0-1.2 GHz 2007 Opteron or 2007 Xeon processor

Datacenter Networking

Initially tree topology
Over subscribed links



Fat tree, Bcube, VL2 etc.

Lots of research to get full bisection bandwidth

Datacenter Design

Goals

Power usage effectiveness (PUE)

Cost-efficiency

Custom machine design

Open Compute Project (Facebook)

Failures are Frequent

Typical first year for a new cluster (Jeff Dean, Google):

- ~0.5 overheating (power down most machines in <5 mins, ~1-2 days to recover)
- ~1 PDU failure (~500-1000 machines suddenly disappear, ~6 hours to come back)
- ~1 rack-move (plenty of warning, ~500-1000 machines powered down, ~6 hours)
- ~1 network rewiring (rolling ~5% of machines down over 2-day span)
- ~20 rack failures (40-80 machines instantly disappear, 1-6 hours to get back)
- ~5 racks go wonky (40-80 machines see 50% packet loss)
- ~8 network maintenances (4 might cause ~30-minute random connectivity losses)
- ~12 router reloads (takes out DNS and external vips for a couple minutes)
- ~3 router failures (have to immediately pull traffic for an hour)
- ~dozens of minor 30-second blips for DNS
- ~1000 individual machine failures
- ~thousands of hard drive failures
- Slow disks, bad memory, misconfigured machines, flaky machines, etc.

- Long distance links: wild dogs, sharks, dead horses, drunken hunters, etc.

How Many Datacenters?

- 1-10 datacenter servers/human?
- 100,000 servers/datacenter

	U.S.	World
Servers	**0.3-3B**	**7-70B**
Datacenters	**3000-30,000**	**70,000-700,000**

- 80-90% of general-purpose computing will soon be in datacenters?

Hopper vs. Datacenter

	Hopper	Datacenter[2]
Nodes	6384	1000s to 10000s
CPUs (per node)	2x12 cores	~2x6 cores
Memory (per node)	32-64GB	~48-128GB
Storage (overall)	~4 PB	120-480 PB
Interconnect	~ 66.4 Gbps	~10Gbps

[2]http://blog.cloudera.com/blog/2013/08/how-to-select-the-right-hardware-for-your-new-hadoop-cluster/

Part II: Virtualization

What is virtualization?

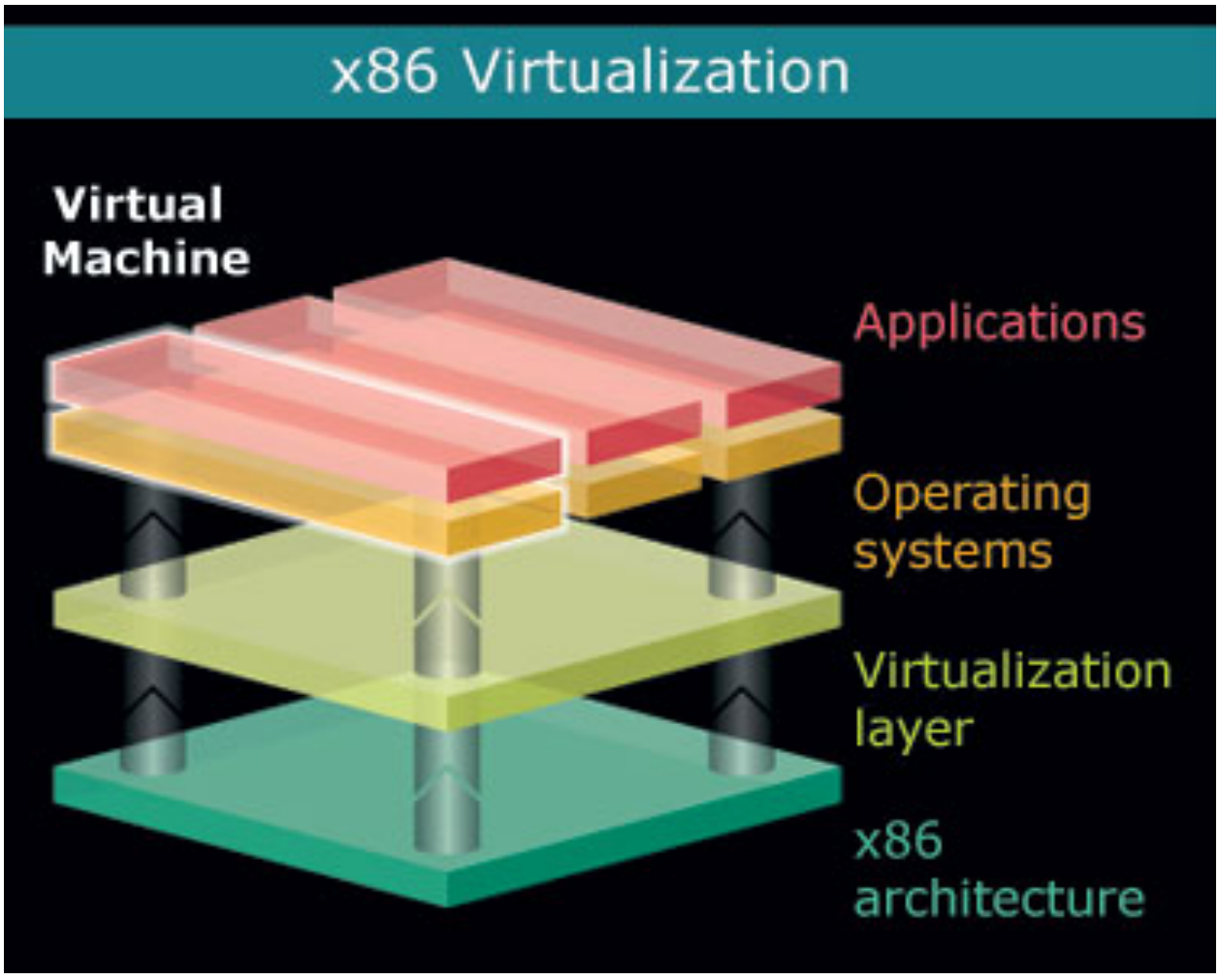
- Virtualization allows one computer to do the job of multiple computers.
- Virtual environments let one computer host multiple operating systems at the same time

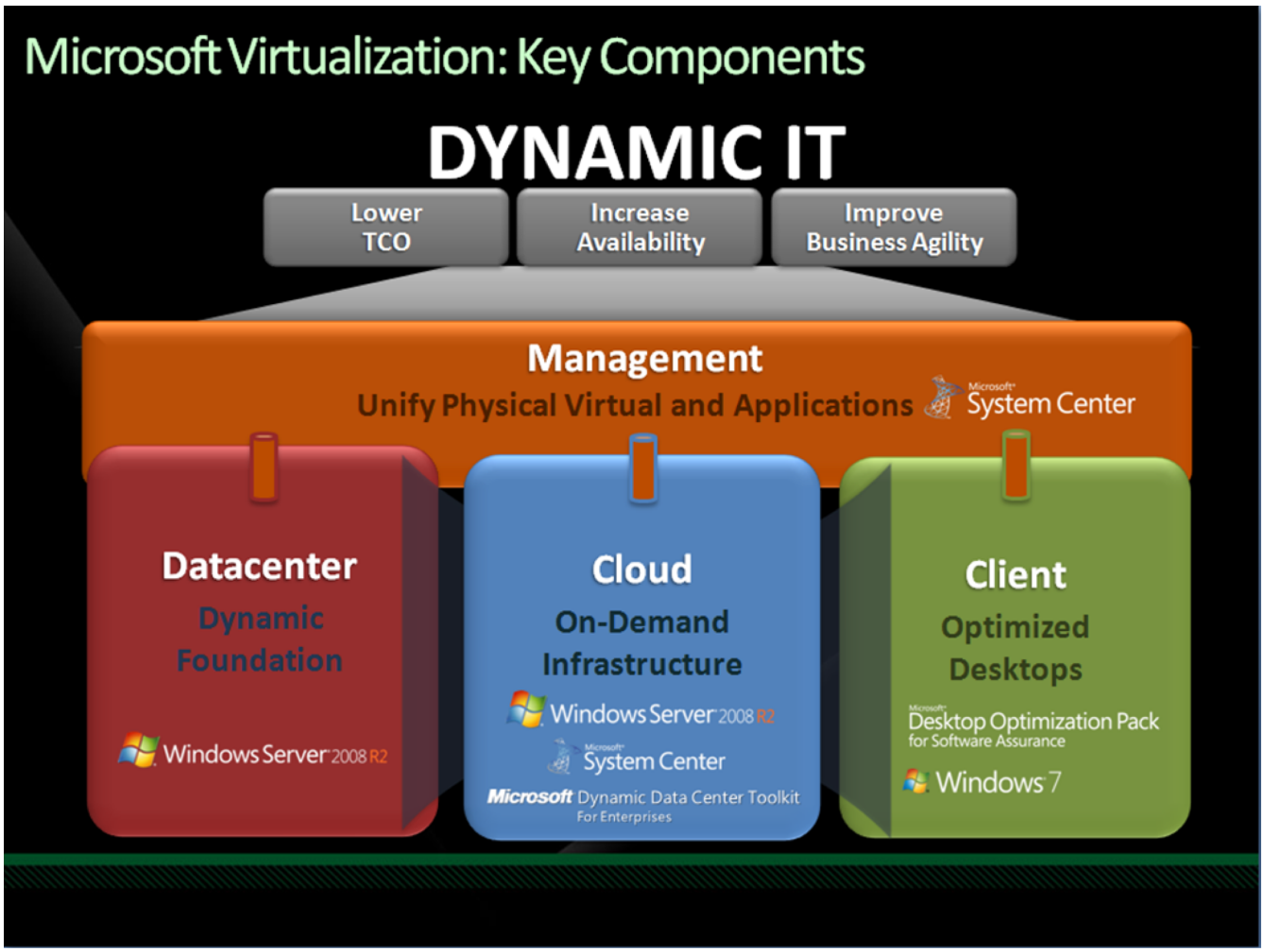
How does it work?

- Virtualization transforms hardware into software.
- It is the creation of a fully functional virtual computer that can run its own applications and operating system.
- Creates virtual elements of the CPU, RAM, and hard disk.

Background and Evolution

- Virtualization arose from a need in the 1960's to partition large mainframe hardware.
- Improved in the 1990s to allow mainframes to multitask.
- First implemented by IBM more than 30 years ago.





Pros

- Benefits include freedom in choice of operating system.
- It saves time and money.
- Consolidates server and infrastructure.
- Makes it easier to manage and secure desktop environments.

Cons

- Only powerful computers can successfully create virtual environment.
- Requires training to operate.

Conclusion

- "Companies that don't understand the risk just shouldn't use cloud computing, the potential for a security breach or a compliance violation can be high" –IDC Analyst Phil Hochmuth
- Understanding cloud computing will allow you to understand both the risks and the rewards
- Virtualization is a cost-effective way to run multiple operating systems and software.
- It allows for greater access to basic system resources.
- It is safer and easier to manage than physical hardware.