Week 1 Notes

Friday, April 17, 2020 1:31 PM

- clouds get alot of hype ~ even some governments now use

Popular Cloud Providers

- AWS
  - EC2 : Elastic Computing Cloud
  - S3 : Simple Storage Service
  - EBS : Elastic Block Storage
- Microsoft Azure
- Google compute Engine
- Rightscale, Salesforce, EMC, Gigaspaces, 10gen, Datastax, Oracle, VMWare, Yahoo, Cloudera
- Clouds can be public or private
  - Private clouds restrict access, usually at company level
    - ↳ Individuals can also purchase private clouds for computational needs.
  - Public clouds are generally available to anyone with a credit card.
    - AWS, Google Compute Engine, Azure

AWS

- Amazon S3 (simple storage service - store arbitrary datasets), pay by GB-monthly stored
- Amazon EC2 (Elastic Compute Cloud - upload and run arbitrary OS images, pay per CPU hour used.

Google App Engine

- develop applications within their app engine
- you can then upload data in a specified format

Why is Cloud Computing Popular???

- Customers save time and money by avoiding the time and costs of setting up servers.

What is a Cloud???

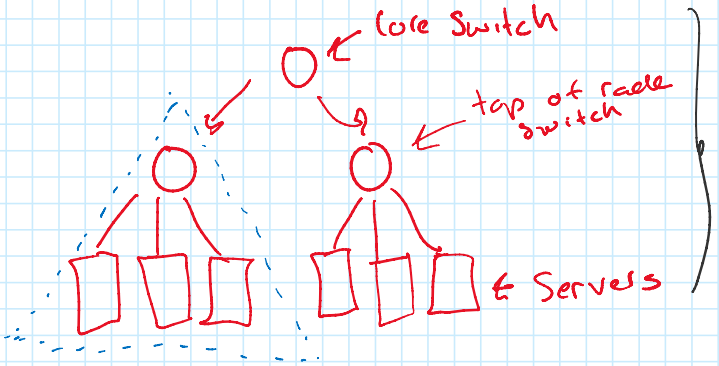
- there is no single definition that works in all scenarios
- for this course, we will define a cloud as:

  Cloud = Lots of storage + compute cycles nearby

- A single-site cloud center (aka "Data Center") consists of the following:
  - Computer Nodes (grouped into racks)
    - a rack is a unit of several servers that typically share the same power
    - they often share a "top of the rack switch"
  - switches are usually connected by a network topology (for instance, a tree)
  - storage (backend) nodes connected to the network, usually with more SSD's or larger hard disks.
  - Front end for submitting jobs and receiving requests

- Storage (backend) nodes connected to the network, usually with more SSD's or larger hard disks.
- Front end for submitting jobs and receiving requests
- Software services (OS systems for example)

- A geographically distributed data-center is a sight with multiple single data centers.



- So then what is a cluster?
- How are clusters different?

Technological Trends:

- Doubling periods
  - storage - 12 months
  - bandwidth - 9 months
  - CPU compute capacity - 18 months

} Moore's Law

- In 1965 MIT's Fernando Corbató envisioned a world where computing is a utility - ie; you could just "plug in" and run your favorite applications.

Clouds Today:

- There are 4 major features which distinguish the clouds of today

1. Massive Scale
2. On-demand Access: Pay as you go, no upfront commitment (anyone can access it)
3. Data-intensive Nature: What was MBs has now become TBs, PBs, and XBs.
   - Data logs, forensics, web data, etc.
   - Humans have data numbness
4. New Cloud Programming Paradigms: MapReduce / Hadoop, NoSQL / Cassandra / MongoDB
   - highly accessible
   - Lots of open source

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

Power:

- WUE = annual water usage / IT Equipment Energy (L/KWh) (low is good) (water usage)
- PUE = Total Facility Power / IT Equipment Power (low is good - eg google = 1.11) (power usage)

Cooling:

- so much power requires efficient cooling
- air is often pulled in, water is filtered and sprayed into the air

ON Demand Access; *AAS Classification:

- On-demand services such as AWS Elastic Compute Cloud (EC2) charge only a few cents to a few dollars per CPU hour.
  - AWS Simple Storage Service (S3) charges only a few cents to a few dollars per GB per month.
- Cloud services are often classified based on the nature of service
  - HaaS: Hardware as a service
    - barebones hardware machines, complete autonomy
    - not a good idea because of security risks
  - IaaS: Infrastructure as a service
    - access to machines to store your own instances and operating systems (Ex. AWS, Azure, etc)
  - PaaS: Platform as a service (easier but less flexible than IaaS)
    - flexible computing and storage infrastructure, coupled with a software platform (Ex. Google AppEngine)
  - SaaS
    - you get access when you need them (Ex. Google Docs)

Data Intensive Computing:

- Computation Intensive Computing
  - Example: MPI-based, High Performance Computing, Grids
  - Typically run on super computers
- 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 is no longer the most important resource.
  - I/O (disk / network) is now most important

New Cloud Computing Paradigms:

- Easy to write and run highly parallel programs in the new cloud computing paradigms.
  - Google: MapReduce and Sawzall
  - Amazon MapReduce service (pay as you go)
  - Google (Map Reduce)
  - Yahoo! (Hadoop + Pig)
  - Facebook (Hadoop + Hive)
  - NoSQL: MySQL (industry standard)

Rule of thumb

- If you're going to need cloud storage for more than 3 or 6 months, it's cheaper to own

↑ many start ups

Rule of thumb

- If you're going to need cloud storage for more than 5.55 months, it's cheaper to own.
- If you're going to need cloud storage and computing for more than 12 months, it's cheaper to own.

} many start ups => use cloud computing