Cloud Computing and Distributed Systems

Introduction

Raja Appuswamy

Eurecom

Cloud computing: The disruption

"In 2020, the global cloud computing market was valued at $371.4 billion, and it is estimated that by 2025 it will rise to a staggering $832.1 billion." – Marketsandmarkets

"74% of Amazon's operating profit comes from AWS" – Amazon

"80% of organizations will migrate toward the cloud by 2025." – Gartner

"50% of all data will be held in the cloud by 2020. Cloud data centers will process 94% of workloads in 2021." – IDC & Cisco

"Global data centers used roughly 416 terawatts (3% of the total electricity) last year, nearly 40% more than the entire United Kingdom." - Forbes

"Big data solutions via cloud subscriptions will increase about 7.5 times faster than on-premise options." - Forrester

"AI without the cloud is tough" – Information Age

This Course

- **What you will learn (roadmap)**
  - **Economic foundations**
    - Service models
  - **Infrastructure foundations**
    - Virtualization, containerization, serverless functions
  - **Systems foundations**
    - Hadoop, Apache Spark
    - Relational databases, Distributed file systems, maybe bitcoin & blockchain
  - **Programming foundations**
    - Map—reduce and functional programming
    - SQL and NoSQL
  - **Algorithmic foundations**
    - Consistency, Serializability, Transactions
    - Atomic commitment, two-phase, three-phase commit
    - Consensus, PAXOS, CAP theorem

Who is this course for?

- **Cloud developers, architects, data engineers, data analysts, …**
  - Cloud computing sector is expected to grow 14 percent annually and create **one million new jobs** in 2022
  - Right now there are an estimated **5.6 million cloud-related jobs worldwide**
  - $80,000 to $200,000 per year
  - You will ace system design interview with this course
  - You will be on your way to getting certified

- **Requirements**
  - Some familiarity with operating systems concepts
  - Some familiarity with computer architecture
  - Knowledge of c++, python, git, web frameworks

Grading

- **Exams**
  - Final - 30% of the overall grade
  - Midterm - 20% of the overall grade
  - Entirely based on course lectures
  - Design questions, problems, …
- **In-class Quizzes**
  - 10% of grade
  - Based on previous lecture, after mid-class break
- **Labs**
  - 40% of grade
  - Work with public cloud platform(s)
  - Develop an awesome dist. system (more later)

How to make the most of this course?

- **Attend classes**
  - Many discussions in live classes
  - Lecture notes in Moodle
  - Last year’s lectures on mediaserver
- **PLAN AHEAD FOR LABS & PROJECT**
  - Labs/project will be hard if you have non-CS background
  - Everything can be done remotely
  - We can hold additional recitation sessions if necessary
- **Books**
  - Designing data intensive applications
  - Spark: The Definitive Guide

Tentative timeline

Date	Tentative lecture topic
Oct 5	Introduction to cloud computing
Oct 12	infrastructure fundamentals
Oct 19	Intra-node parallelism
Oct 26	Shared-nothing parallelism, MPI, Map-Reduce
Nov 09	Memory Hierarchy & Spark
Nov 16	Memory Hierarchy & Spark
Nov 23	**Midterm**
Nov 30	Cloud storage: Databases & Transactions
Dec 07	Cloud storage: File systems & Consistency
Dec 14	Atomic commitment in distributed systems
Jan 04	Atomic commitment in distributed systems
Jan 11	Consensus in distributed systems
Jan 18	Consensus in distributed systems
Jan 25	Consensus in distributed systems

Introduction to the Cloud Computing

We live in a world of data

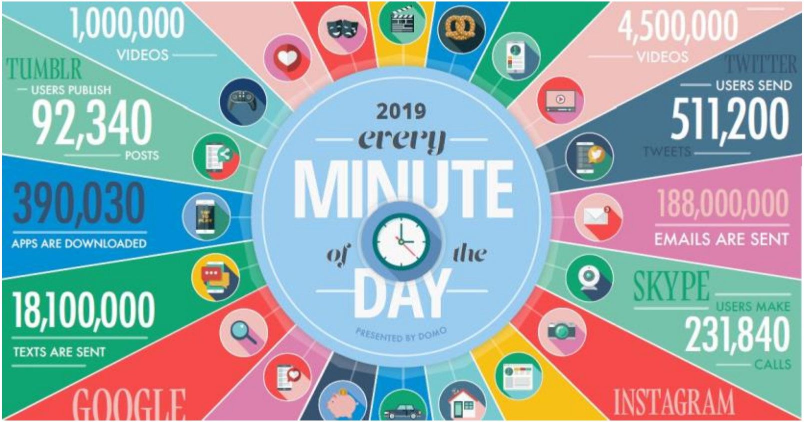


Figure: Data deluge.

Big Data

- **Big data is defined as large pools of data that can be captured, communicated, aggregated, stored, and analyzed.**
- **Data continues to grow**

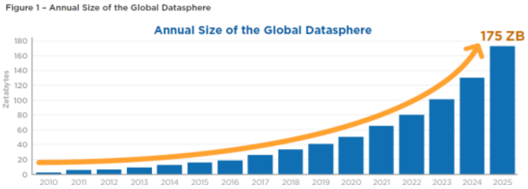


Figure: Global datasphere

- **Applications are becoming data intensive**
  - More data leads to better accuracy
  - With more data, accuracy of different algorithms converges

Let’s look at your data.

You want to access, shared, process your data from all your devices, anytime, anywhere.

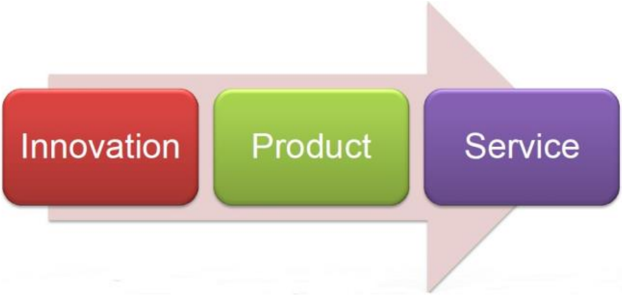
How will we manage all this data?

- **Manage it ourselves?**
  - How do we store it?
  - How do we share it?
  - How can we enable access to it from any place?
  - How do we process all of it?
  - How do we secure it?
  - ....
- **What if it is managed by someone else?**
  - Someone provides a management “service”
  - You pay a subscription for this “service”

Utility–Product–Service lifecycle: Water

Utility–Product–Service lifecycle: Electricity

Generalizing the lifecycle



Cloud computing: The prophecy

•In 1965, MIT's Fernando Corbató and the other designers of the Multics operating system envisioned a **computer facility operating “like a power company or water company”**.

• Plug your thin client into the computing Utility and Play your favorite Intensive Compute & Communicate Application

Cloud Computing

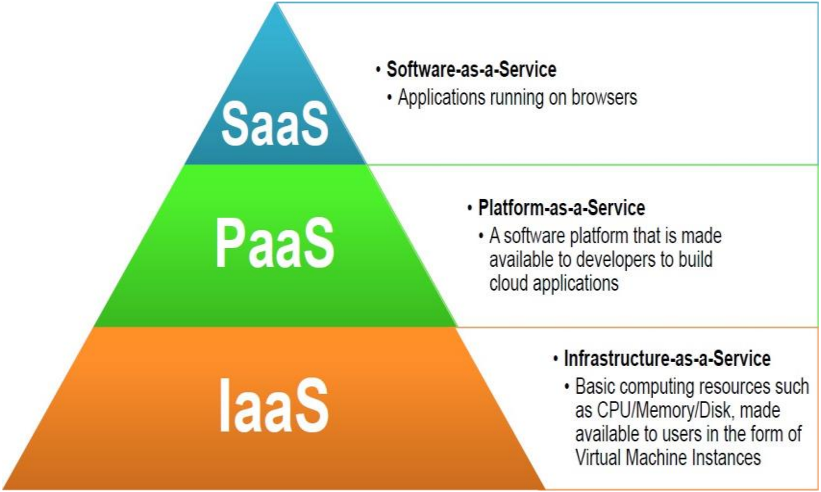
Transformation of IT from a product to a service

Formal definition

IT as a service

- How do we offer IT as a service?
- Different users have different needs
    - Average end user
    - Mobile app developer
    - Enterprise systems architect
- Let us look at some service models

Basic cloud service models



SaaS

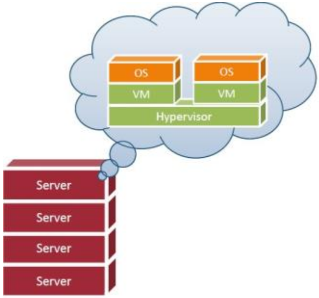
- Software is delivered as a service over the Internet, eliminating the need to install and run the application on the customer's own computer
- Simplifies maintenance and support
- You use SaaS products everyday
    - Gmail, Google docs, Youtube, ...
- Salesforce.com is a popular commercial pioneer (ERP, CRM, ...)

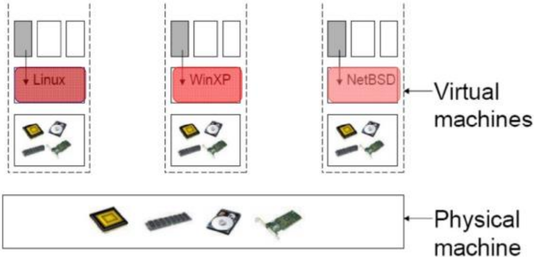
PaaS

- The Cloud provider exposes a set of tools (a platform) and APIs which allows users to create SaaS applications
- The SaaS application runs on the provider's infrastructure
- The cloud provider manages the underlying hardware and requirements
- Examples: Google App Engine, Windows Azure Web App service

IaaS

- The cloud provider leases to users Virtual Machine Instances (i.e., computer infrastructure) using the virtualization technology
- The user has access to a standard Operating System environment and can install and configure all the layers above it
- Ex: AWS EC2, Rackspace, Google Compute Engine





Other services models

- Hardware-as-a-service (HaaS)
    - You get access to barebones hardware machines, do whatever you want with them, Ex: Your own cluster
    - https://www.youtube.com/watch?v=pqfd4t9ISHY
- X-as-a-service, where X can be
    - Backend (BaaS), Desktop (DaaS), ...

Cloud Computing

Cloud Computing is the delivery of computing as a **service** rather than a **product**,

whereby **shared resources, software**, and **information** are provided to computers and other devices,

as a **metered service** over a **network**.

Cloud Infrastructure

What is a server?

- **Servers are computers that provide "services" to "clients"**
  - Typically designed for reliability and to service a large number of requests
  - Dual-socket servers are the fundamental building block of cloud infrastructure
- **Organizations typically require many physical servers to provide various services**
  - Web server, database server, mail server, ...
- **Server hardware is becoming more compact**
  - conserving floor space
  - improving manageability
  - power and cooling

What is a rack?

- Servers are grouped, placed, and organized in racks
- Equipment are designed in a modular fashion to fit into rack units (1RU = 4.45cm)
- A single rack (6 ft or 180cms) can hold up to 42 1U servers

Figure: Global datasphere

What is a data center?

- Facility used to house a large number of computer systems and associated components
    - Air conditioning
    - Power supply
    - Hazard protection
    - Security and monitoring systems
    - Networking and connectivity

- Let’s take a look at two datacenters
    - https://www.youtube.com/watch?v=zDAYZU4A3w0&t=416s
    - https://www.youtube.com/watch?v=L2oJw1a_qEM

Problems with privately owned data centers

- **Expensive to setup (High capital expenses or CAPEX)**
  - Real estate, server and peripherals, ...
- **Expensive to operate (High operational expenses or OPEX)**
  - Energy costs (Good data centers have efficiency of 1.7, 0.7 Watts lost for each 1W delivered to the servers)
  - Administration costs
- **Difficult for applications to grow/shrink**
  - How do we map applications to servers?
  - What if we over/under provision?
- **Low utilization (30% server usage considered good)**
  - Throw money at the performance problem (peak provisioning)
  - Uneven application fit: each server has CPU, memory, and disk: most applications exhaust one resource, stranding the others
  - Uncertainty in demand: Demand for a new service can spike quickly

What if

- **Turn the servers into a single large resource pool and let services dynamically expand and contract their footprint as needed?**
- **Two main requirements:**
  - Means for rapidly and dynamically satisfying application fluctuating resource needs
    - Provided by virtualization
  - Means for servers to quickly and reliably access shared and persistent data
    - Provided by programming models and distributed file/storage/database systems

What is a cloud then?

- **Single-site cloud**
  - A data center hardware and software that the vendors use to offer the computing resources and services
- **Geographically distributed cloud**
  - Multiple such sites, with each site perhaps having different structure and services
  - https://www.youtube.com/watch?v=47e_3WBCe-Q

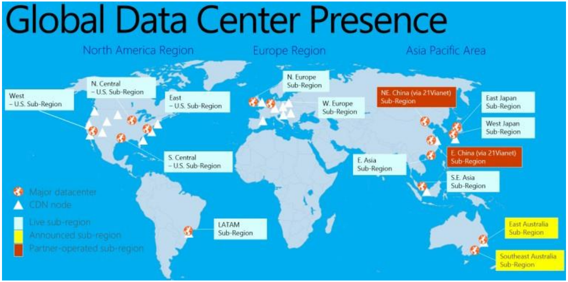


Figure: Azure: 1 million servers, 100 data centers across 90 countries.

Cloud h/w-s/w stack

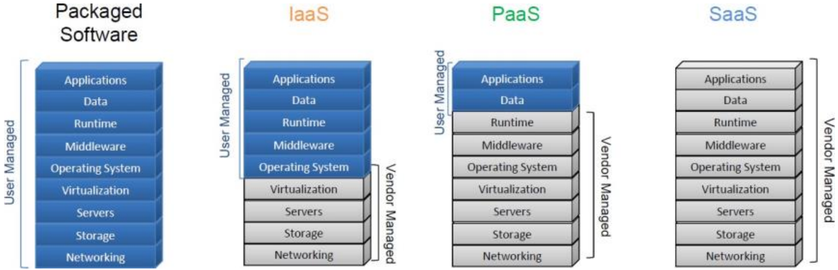
The Cloud Stack

- **Applications**
  - Cloud applications can range from Web applications to scientific computational jobs
- **Data**
  - Old SQL systems (Oracle, SQLServer)
  - NoSQL systems (MongoDB, Cassandra)
  - NewSQL systems (TimesTen, Impala, Hekaton)
- **Runtime environment**
  - Runtime platforms to support cloud programming models
  - Example: Hadoop, Spark

The Cloud Stack

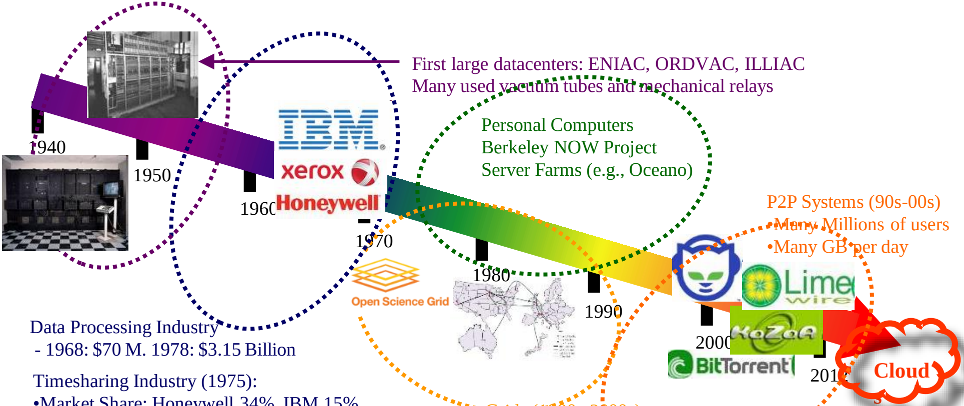
- **Middleware**
  - Platforms for Resource Management, Monitoring, Provisioning, Identity Management and Security
- **Operating systems**
  - Standard Operating Systems used in Personal Computing
  - Packaged with libraries and software for quick deployment and provisioning
  - E.g., Amazon Machine Images (AMI) contain OS as well as required software packages as a "snapshot" for instant deployment
- **Virtualization (serverse, storage, networking)**
  - Key enabler of cloud computing
  - Providers resource virtualization, multitenancy
  - Ex: Amazon EC2 is based on the Xen virtualization platform, Azure based on HyperV

Cloud service models and the cloud stack



Cloud Computing

“A Cloudy History of Time”



Data Processing Industry
- 1968: $70 M. 1978: $3.15 Billion

Timesharing Industry (1975):
•Market Share: Honeywell 34%, IBM 15%,
Xerox 10%, CDC 10%, DEC 10%, UNIVAC 10%

Grids (1980s-2000s):
•GriPhyN (1970s-80s)
•Open Science Grid and Lambda Rail (2000s)
•Globus & other standards (1990s-2000s)

Cloud computing: Full circle back to time sharing

Supporting technologies

- Cloud computing is a combination of technologies
    - Connectivity to move data => **Networked systems**
    - Interactivity for seamless interface => **Web 2.0 and HCI**
    - Reliability against failures => **Dependable systems**
    - Acceptable performance => **Parallel and distributed systems**
    - Ease of programmability for developing new services => **Programming languages**
    - Manageability for Big Data => **Storage systems**
    - Pay-as-you-go to avoid capital investment => **Utility computing & economics**
    - Scalability and elasticity for changing needs => **Virtualization**

Why Cloud Computing?

Applications enabled by cloud computing

- **High-growth applications**
  - When you startup gains traction, can you keep up?
  - Friendster(2001): Could not keep up with user growth
  - Facebook (2006): $Billion company today
  - Airbnb, Uber, Expedia, ...
- **Aperiodic applications**
  - How do you deal with sudden load peaks?
    - https://aws.amazon.com/blogs/aws/amazon-prime-day-2022-aws-for-the-win/
    - Flipkart website crashed on their “Big Billion Day” sale due to DDOS attack: https://review.firstround.com/navigating-the-leap-from-big-tech-to-startups-advice-from-a-former-google-and-flipkart-exec
  - If you design for peak, how do you deal with low loads?
    - Amazon normal day: 1.3 billion transactions

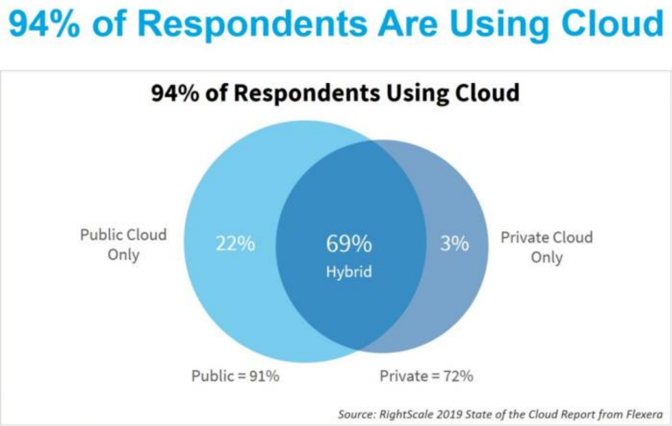
Applications enabled by cloud computing(2)

- **On-off applications**
  - Scientific simulation using 1000s of computers
    - DNA Nexus and Baylor college of medicine analyzed DNA of more than 14,000 individuals
    - 2.4 million core-hours of computational time, 440 TB of results, 1PB of storage
  - Why not rent computing time to run such one-off experiments?
- **Periodic applications**
  - Stock market analysis
    - Mine market data during day
    - Analyze data during night
    - Different computational requirements at different times
  - Dynamic, flexible infrastructure can reduce costs, improve performance

Types of clouds

- **Public (external) cloud**
  - Open market for on demand computing and IT resources
  - Concerns: Limited SLA, reliability, availability, security, and trust
- **Private (internal) cloud**
  - For large enterprises with the budget and large-scale IT
- **Hybrid cloud**
  - Extend private cloud by connecting it to public cloud
  - Use the local cloud, and when you need more resources, burst into the public cloud
  - Dropbox use case: https://www.wired.com/2016/03/epic-story-dropboxs-exodus-amazon-cloud-empire/
- **Multi cloud**
  - Replicate/partition microservices across public cloud vendors
  - Prevent vendor lock in, resilient to cloud outage

Cloud adoption



- All major cloud providers are extending their offering to private and hybrid markets
  - Example: Google Anthos, Microsoft AzureStack

Know the leaders