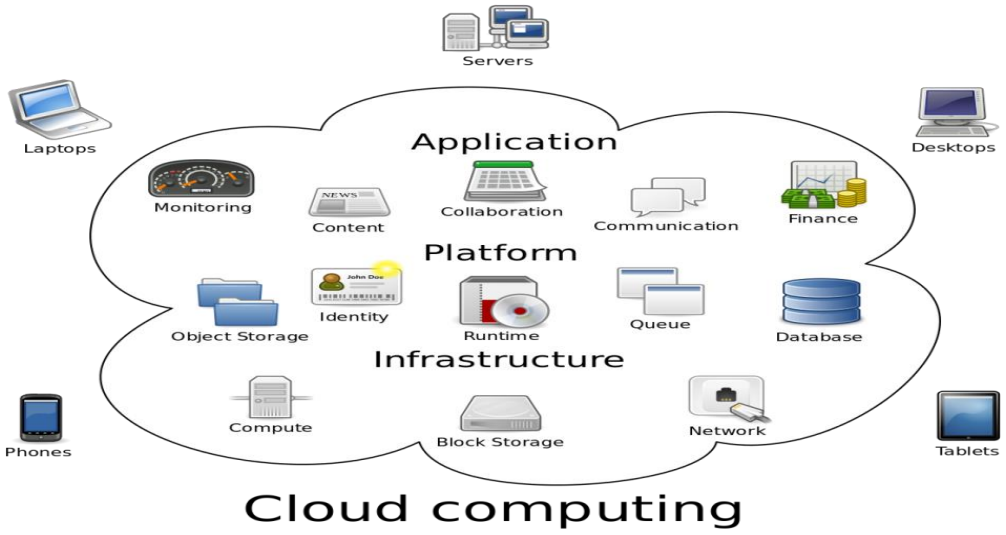
Introduction to Cloud Computing

Dr. Rajiv Misra
Associate Professor
Dept. of Computer Science & Engg.
Indian Institute of Technology Patna
rajivm@iitp.ac.in

Preface

Content of this Lecture:

- In this lecture, we will discuss a brief introduction to Cloud Computing and also focus on the aspects *i.e*. Why Clouds, What is a Cloud, Whats new in todays Clouds and also distinguish Cloud Computing from the previous generation of distributed systems.



Scalable Computing Over the Internet

- Evolutionary changes that have occurred in **distributed and cloud computing** over the past 30 years, **driven by applications with variable workloads and large data sets .**
- Evolutionary changes in machine architecture, operating system platform, network connectivity, and application workload.
- **Distributed computing** system uses multiple computers to solve large-scale problems over the Internet. Thus, **distributed computing becomes data-intensive and network-centric.**
- **The emergence of computing clouds** instead demands high-throughput computing (HTC) systems built with distributed computing technologies.
- **High-throughput computing (HTC)** appearing as computer clusters, service-oriented architecture, computational grids, peer-to-peer networks, Internet clouds, and the future Internet of Things.

The Hype of Cloud: Forecasting

- 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**."
- Forrester in 2010 – Cloud computing will go from **$40.7 billion** in 2010 to **$241 billion** in 2020.
- 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 Compute Engine/AppEngine
- Rightscale, Salesforce, EMC, Gigaspaces, 10gen, Datastax, Oracle, VMWare, Yahoo, Cloudera
- And 100s more…

service cloud

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 App Engine/Compute Engine:** develop applications within their App Engine framework, upload data that will be imported into their format, and run

Customers Save: Time and Money

- "With AWS, a new server can be up and running in **three minutes** compared to **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."
- "With Online Services, reduce the IT **operational costs** by roughly **30%** of spending"
- "A private cloud of virtual servers inside its datacenter has saved nearly **crores of rupees annually**, because the company can share computing power and storage resources across servers."
- 100s of startups can harness large computing resources without buying their own machines.

What is a Cloud?

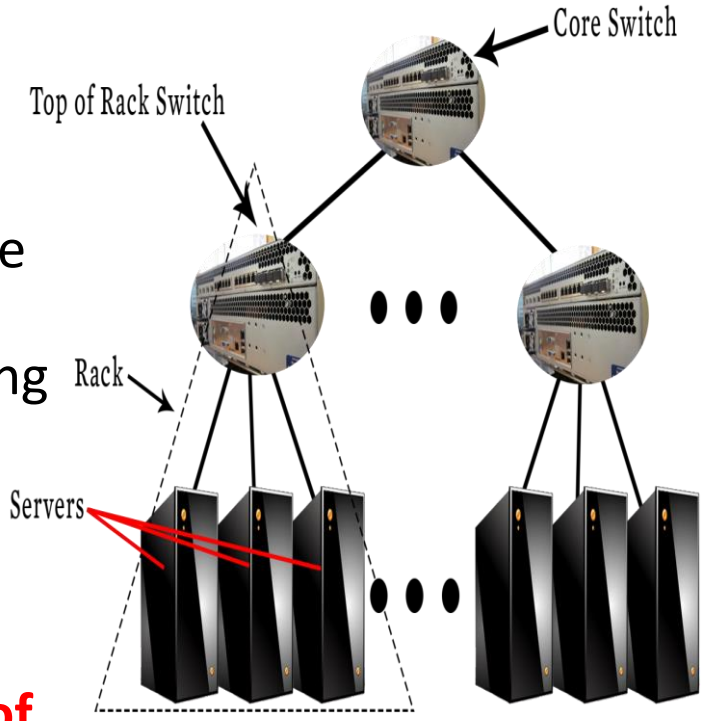
- Advances in virtualization make it possible to see the growth of Internet clouds **as a new computing paradigm**.
- i.e. dramatic differences between developing software for millions to use **as a service** versus distributing software to run on their PCs."

History:

- **In 1984, John Gage** Sun Microsystems gave the slogan, **"The network is the computer."**
- **In 2008, David Patterson** UC Berkeley said, **"The data center is the computer."**
- Recently, **Rajkumar Buyya** of Melbourne University simply said: **"The cloud is the computer."**
- Some people view **clouds as grids** or **clusters** with changes through virtualization, since clouds are anticipated to process huge data sets generated by the traditional Internet, social networks, and the future IoT.

What is a Cloud?

- **A single-site cloud (as known as "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 "three-tier architecture")
  - Software Services

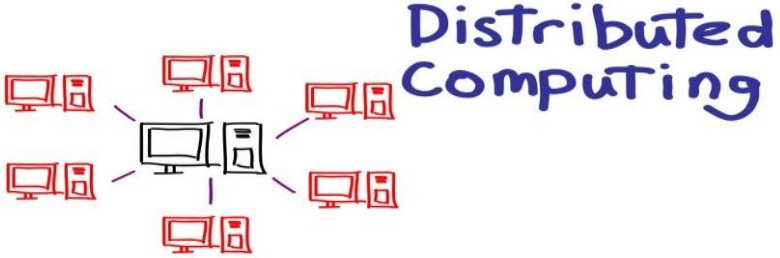


- **A geographically distributed cloud consists of**
  - Multiple such sites
  - Each site perhaps with a different structure and services

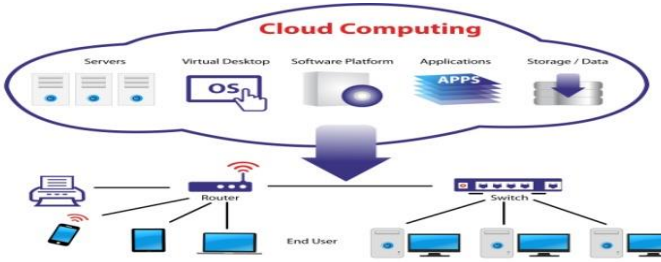
Computing Paradigm Distinctions

- Cloud computing overlaps with distributed computing.

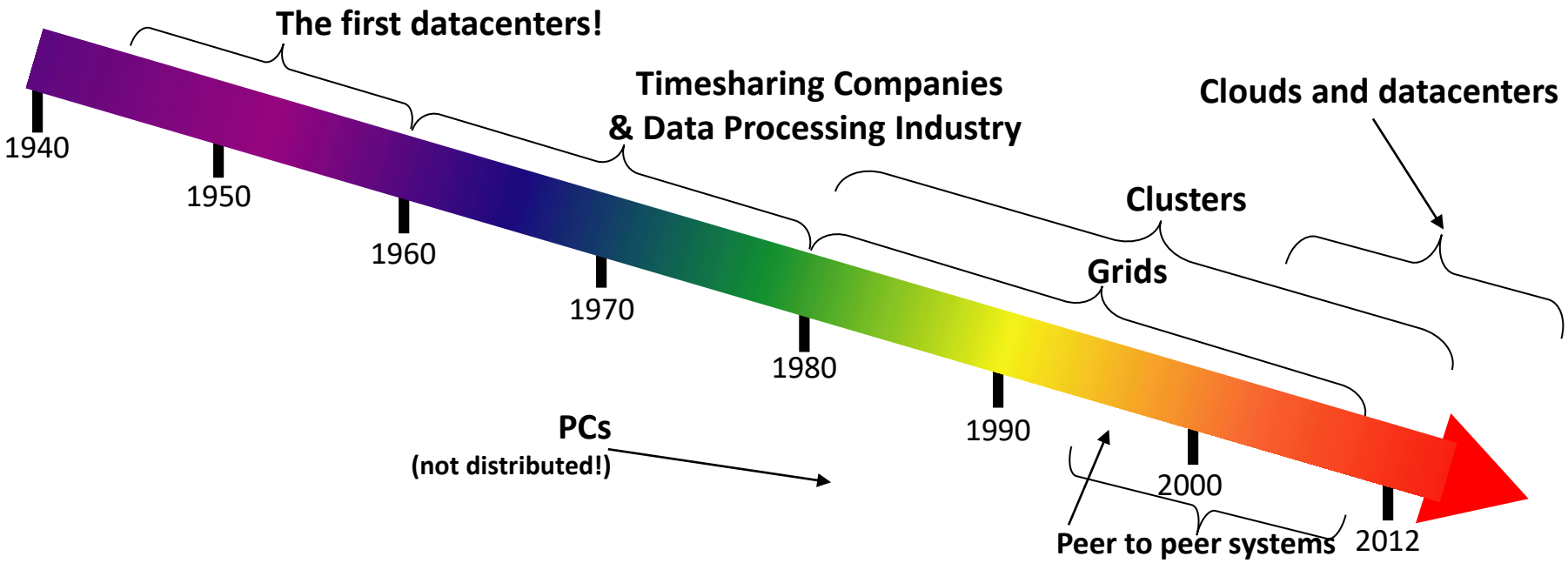
- **Distributed computing:** A ***distributed system*** consists of multiple autonomous computers, having its own memory, communicating through ***message passing***.



- **Cloud computing:** Clouds can be built with physical or virtualized resources over large data centers that are distributed systems. Cloud computing is also considered to be a form of ***utility computing or service computing.***



"A Cloudy History of Time"



"A Cloudy History of Time"

First large datacenters: ENIAC, ORDVAC, ILLIAC
Many used vacuum tubes and mechanical relays

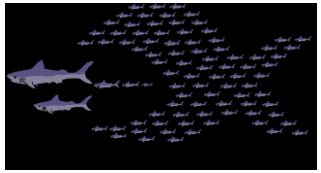
1940

1950

1960

IBM
xerox
Honeywell

1970



CRAY
THE SUPERCOMPUTER COMPANY

Berkeley NOW Project
Supercomputers
Server Farms (e.g., Oceano)

P2P Systems (90s-00s)
- Many Millions of users
- Many GB per day

1980

1990

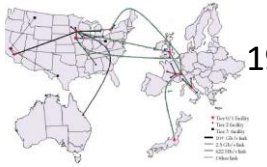
2000

2012

Lime wire
KaZaA
BitTorrent

Clouds

Open Science Grid



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%
- Honeywell 6000 & 635, IBM 370/168,
  Xerox 940 & Sigma 9, DEC PDP-10, UNIVAC 1108

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

Scalable Computing Trends: Technology

- **Doubling Periods** – storage: 12 months, bandwidth: 9 months, and CPU compute capacity: 18 months (what law is this?)
- **Moore's law** indicates that processor speed doubles every 18 months.
- **Gilder's law** indicates that network bandwidth has doubled each year in the past.

- 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

The Trend toward Utility Computing

- Aiming towards autonomic operations that can be self-organized to support dynamic discovery. Major computing paradigms are composable with ***QoS and SLAs (service-level agreements).***
- In 1965, MIT's Fernando Corbató 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** Intensive Compute & Communicate Application
- **Utility computing** focuses on a business model in which customers receive computing resources from a paid service provider.
- All **grid/cloud platforms are regarded as utility service providers**.

Features of Today's Clouds

I. **Massive scale:** Very large data centers, contain tens of thousands sometimes hundreds of thousands of servers and you can run your computation across as many servers as you want and as many servers as your application will 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.

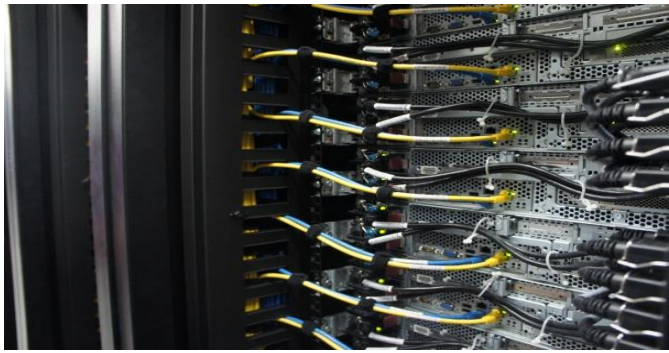
IV. **New Cloud Programming Paradigms:** MapReduce/Hadoop, NoSQL/Cassandra/MongoDB and many others.

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

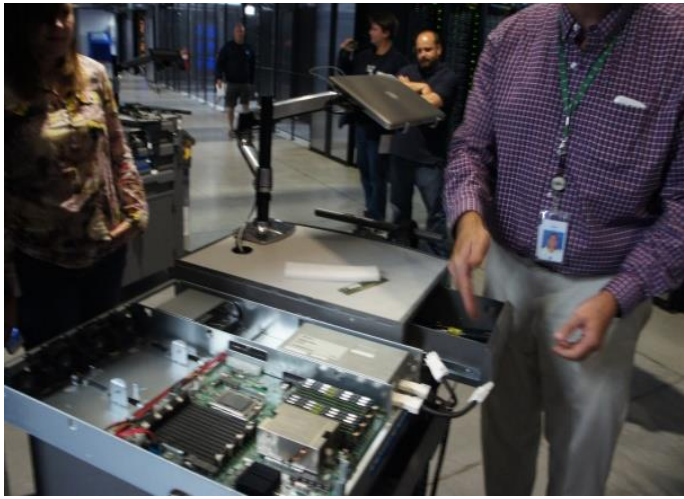
I. Massive Scale

- **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: A lot**

What does a datacenter look like from inside?



Lots of Servers



Power and Energy

- WUE = Annual Water Usage / IT Equipment Energy (L/kWh)
  - low is good
- PUE = Total facility Power / IT Equipment Power
  - low is good (e.g., Google~1.11)

Off-site

On-site

Cooling

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

II. On-demand access: *AAS Classification

- **On-demand:** renting vs. 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**
  - 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**
  - Get access to flexible computing and storage infrastructure. **Virtualization** is one way of achieving this. subsume HaaS.
  - Ex: Amazon Web Services (AWS: EC2 and S3), OpenStack, Eucalyptus, Rightscale, Microsoft Azure, Google Cloud.

II. On-demand access: *AAS Classification

- **PaaS: Platform as a Service**
  - 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**
  - Get access to software services, when you need them. 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 (~2008)
  - **Facebook** (Hadoop + Hive)
    - ~300TB total, adding 2TB/day (in 2008)
    - 3K jobs processing 55TB/day
  - NoSQL: MySQL is an industry standard, but Cassandra is 2400 times faster

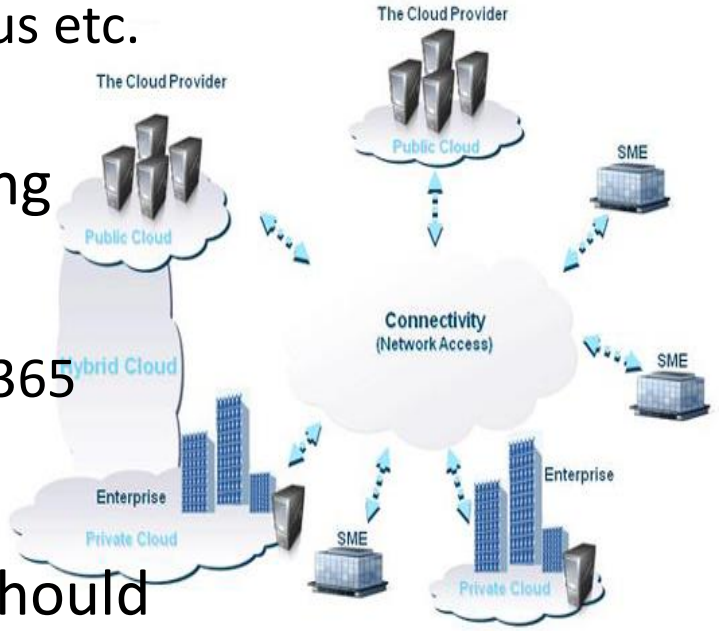
Two Categories of Clouds

- Can be either a (i) public cloud, or (ii) private cloud

- **Private clouds** are accessible only to company employees
- Example of popular vendors for creating private clouds are VMware, Microsoft Azure, Eucalyptus etc.

- **Public clouds** provide service to any paying customer
- Examples of large public cloud services include Amazon EC2, Google AppEngine, Gmail, Office365 and Dropbox etc.

- 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
- **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

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)

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

Conclusion

- Clouds build on many previous generations of distributed systems

- Characteristics of cloud computing problem

  - **Scale, On-demand access, data-intensive, new programming**