Lecture 1 : Introducing Cloud and Edge Computing

Dr. Bibhas Ghoshal

Assistant Professor

Department of Information Technology

Indian Institute of Information Technology Allahabad

Resources

Course website : https://profile.iiita.ac.in/bibhas.ghoshal/teaching_cloud.html

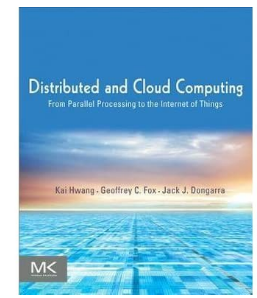
- Slides, Important links, research articles, homework and lab assignments will be provided here

<u>Recommended Books</u>

1. **Distributed and Cloud Computing : From Parallel processing to Internet of Things**

   Kai Hwang, Jack Dongarra, Geoffrey.C.Fox

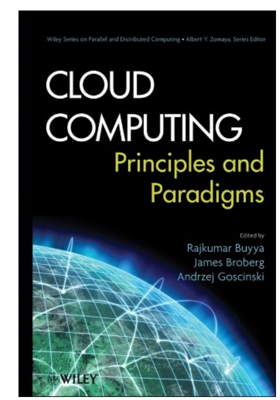
   Morgan Kauffman, 2011



2. **Cloud Computing : Principles and Paradigms**

   Edited by *Raj Kumar Buyya, James Broberg, Andrzej Goscinski*

   Wiley, 2011

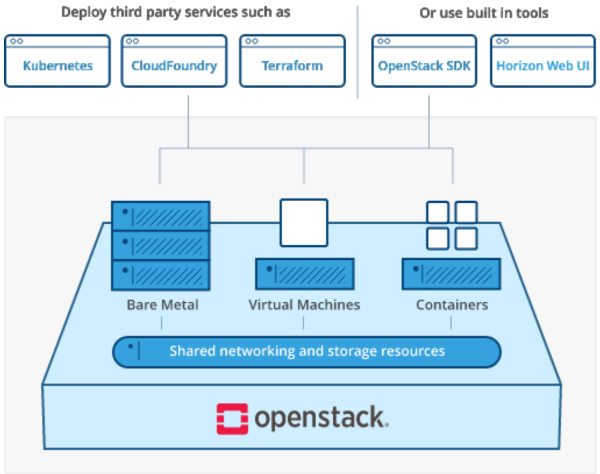


3. Research Papers

   (will be provided on the course website time to time)

- Slides provide key concepts, books provide details
- Use lecture notes supplemented with textbook and internet

Resources for Tut and Lab



CloudLab

Cloudsim Project

docker

hadoop MapReduce

Software Environments

XML

WSDL
WEB SERVICES DESCRIPTION LANGUAGE

SOAP Web Service

{REST}
RESTful Web Services

Web Services and languages

What is the course all about ???

- **Cloud Computing Technology**
  - Cloud computing basics – overview, architecture, deployment
  - Virtualization
  - How to use cloud tools, APIs, SDKs
  - How to access cloud – cloud simulator
- **Cloud Computing as Distributed Systems Environment**
  - How to design Applications for cloud
  - Web services, Service oriented architecture
  - XML, SOAP, WSDL and REST
  - Managing Data, Economics
  - Open Stack
  - Map Reduce
  - Case study with commercial clouds - Microsoft Azure and Google Cloud Platform
- **Cloud Computing as a Research Topic**
  - Cloud Security, Federated
  - Fog Computing
  - DOCKER Container
  - Cloud Computing Challenges
- **Edge Computing**
  - Edge Computing Basics – Architectures, Applications, Edge Intelligence
  - IoT Cloud
  - Mobile Computing

What is a computer ?

Computer

From Wikipedia, the free encyclopedia

For other uses, see Computer (disambiguation).

A **computer** is a machine that can be programmed to carry out sequences of arithmetic or logical operations automatically. Modern computers can perform generic sets of operations known as programs. These programs enable computers to perform a wide range of tasks. A **computer system** is a "complete" computer that includes the hardware, operating system (main software), and peripheral equipment needed and used for "full" operation. This term may also refer to a group of computers that are linked and function together, such as a computer network or computer cluster.

A broad range of industrial and consumer products use computers as control systems. Simple special-purpose devices like microwave ovens and remote controls are included, as are factory devices like industrial robots and computer-aided design, as well as general-purpose devices like personal computers and mobile devices like smartphones. Computers power the Internet, which links hundreds of millions of other computers and users.

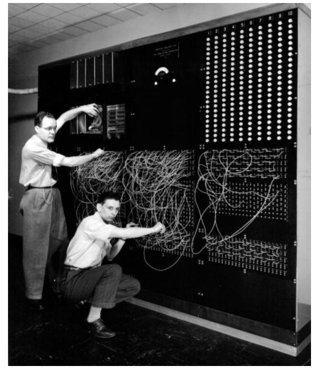
Early computers were only meant to be used for calculations. Simple manual instruments like the abacus have aided people in doing calculations since ancient times. Early in the Industrial Revolution, some mechanical devices were built to automate long tedious tasks, such as guiding patterns for looms. More sophisticated electrical machines did specialized analog calculations in the early 20th century. The first digital electronic calculating machines were developed during World War II. The first semiconductor transistors in the late 1940s were followed by the silicon-based MOSFET (MOS transistor) and monolithic integrated circuit (IC) chip technologies in the late 1950s, leading to the microprocessor and the microcomputer revolution in the 1970s. The speed, power and versatility of computers have been increasing dramatically ever since then, with transistor counts increasing at a rapid pace (as predicted by Moore's law), leading to the Digital Revolution during the late 20th to early 21st centuries.

Conventionally, a modern computer consists of at least one processing element, typically a central processing unit (CPU) in the form of a microprocessor, along with some type of computer memory, typically semiconductor memory chips. The processing element carries out arithmetic and logical operations, and a sequencing and control unit can change the order of operations in response to stored information. Peripheral devices include input devices (keyboards, mice, joystick, etc.), output devices (monitor screens, printers, etc.), and input/output devices that perform both functions (e.g., the 2000s-era touchscreen). Peripheral devices allow information to be retrieved from an external source and they enable the result of operations to be saved and retrieved.

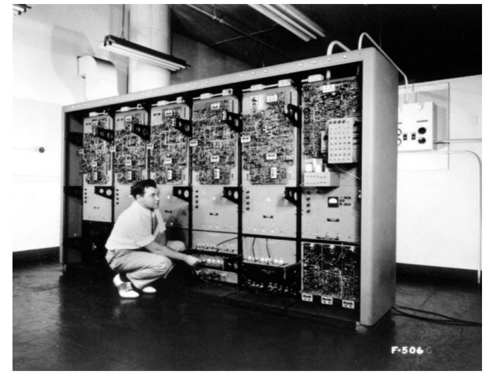
Computers and computing devices from different eras – clockwise from top left:
Early vacuum tube computer (ENIAC)
Mainframe computer (IBM System 360)
Desktop computer (IBM ThinkCentre S50 with monitor)
Supercomputer (IBM Summit)
Video game console (Nintendo GameCube)

History of Computing

Babbage Difference Engine, 1879

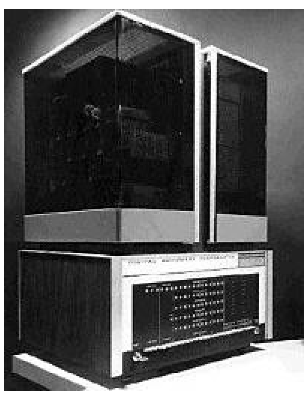


Harvard Mark-I, 1944



MIT Whirlwind, 1951

IBM 360 Mainframe, 1964
$2.5-3 million



DEC PDP-8 **mini**computer, 1965, $18,000

ATARI **micro** computers Gaming + home computing, 1979

Slide Adapted from : Virtualization andCloud Computing, Sourav Bansal, http://www.cse.iitd.ernet.in/~sbansal/csl862

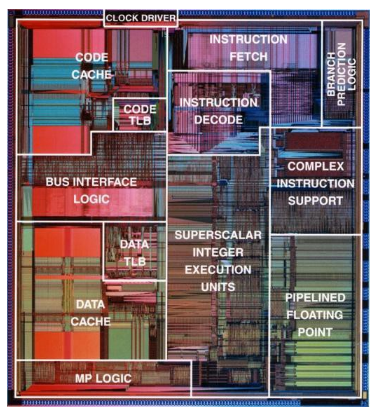
History of Computing

Apple Machintosh, 1984

IBM PC/AT, 1984
Intel 80286 Microprocessor

Microsoft Windows, 1985

Linux, 1991



Intel Pentium Processor Diagram,1993

Google
Internet and Search, 1998

IPAD, 2010

Slide Adapted from : Virtualization andCloud Computing, Sourav Bansal, http://www.cse.iitd.ernet.in/~sbansal/csl862

Trends in Computing

1. **Distributed Computing**
2. **Grid Computing**
3. **Cluster Computing**
4. **Utility Computing**
5. **Cloud Computing**

Distributed Computing

Distributed Computing is a method in which different parts of a computer program run simultaneously on two or more computers that are communicating with each other over a network.

<u>Traditional Computing Components:</u>

Data Storage	Core Logic	User Interface
1. Database	1. Application Specific	1. GUI
2.File based	2. Algorithm Implementation	2. Command Line

All Components sit on the same computer

Distributed Computing :

Data Storage
1. Database
2.File based

Core Logic
1. Application Specific
2. Algorithm Implementation

User Interface
1. GUI
2. Command Line

Characteristics of Distributed Computing

1. Each node plays a partial role

2. Fault Tolerance

3. Transparency

4. Resource sharing

5. Load Sharing

6. Easy to expand

7. Performance

Cluster Computing

Cluster Computing : Type of parallel or distributed computer systems which consists of stand-alone computers working together as a single integrated computing resource

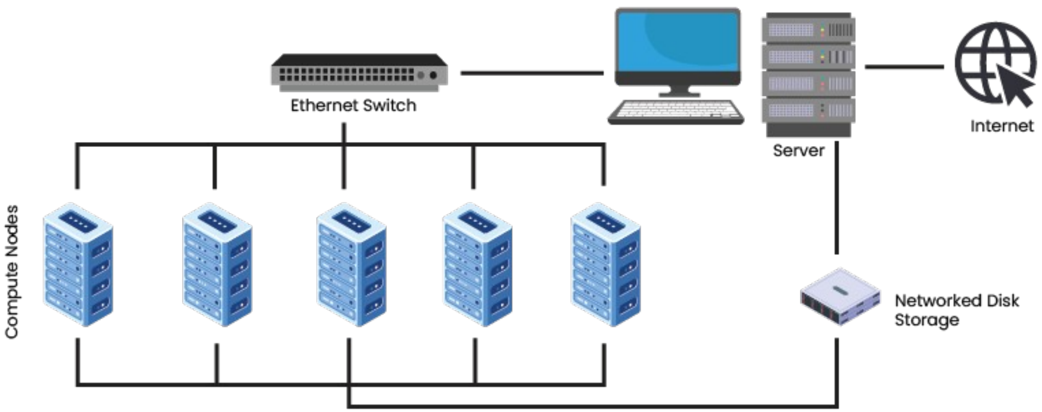
Clusters are deployed to improve speed and/or reliability over that provided by a single computer while typically being much more cost effective than single computer of comparable speed or reliability

Characteristics :

i. Network is faster than typical LAN
ii. Low latency Communication protocols
iii. Loosely coupled than Shared Multiprocessors

Types of Cluster :

i. High availability or Fail over cluster
ii. Load Balancing cluster
iii. Parallel/Distributed Processing cluster



Big Computers

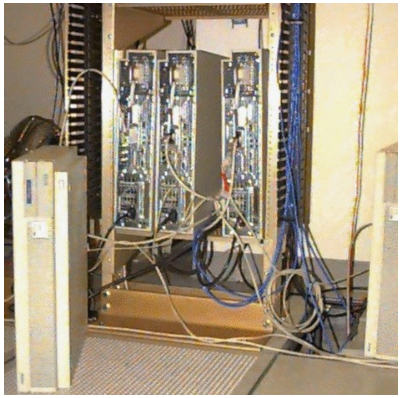
Sun E-10000 "supermini"

i. Up to 64 processors @250MHz
ii. Up to 64 GB RAM
iii. Up to 20 TB Disk
iv. Used by eBay, among others

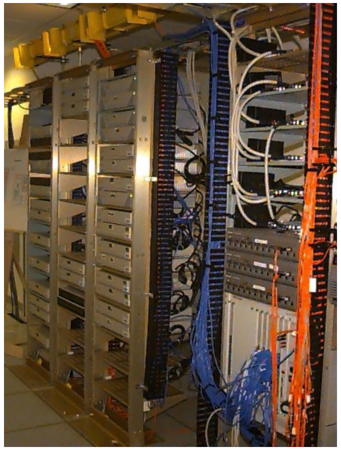
PC

200 MHz CPU, 32MB RAM, 4 GB disk

UC Berkeley Networks Of Workstations (1994-1999)



NOW0,
1994
4 HP-735's



NOW1,
1995
32 SUN
Spark
Stations

NOW2,
1997
60 SUN
Spark-2

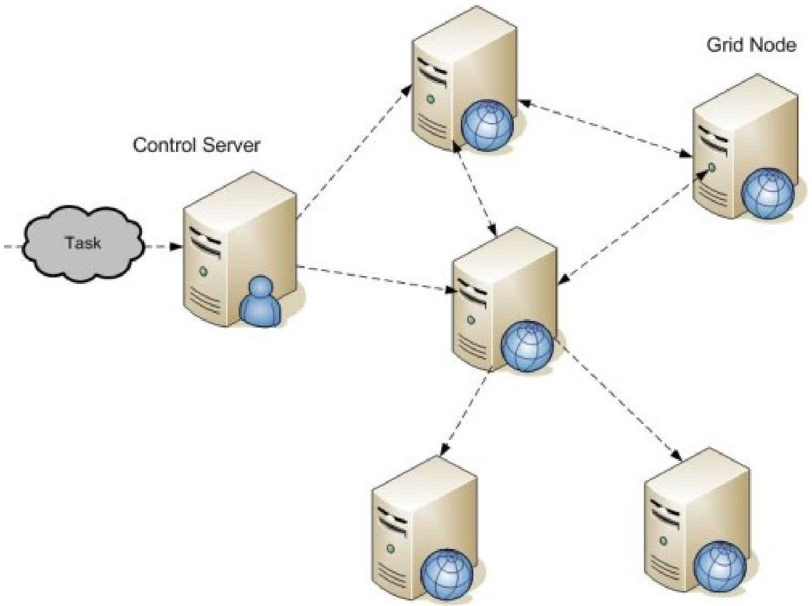
Grid Computing

Grid Computing : combination of computer resources from multiple administrative domains applied to a common task *

Core idea
- distributed parallel computation
- super virtual computer

- Electrical Grid Analogy

Electrical Power Grid	Grid
Users get access to electricity through wall sockets with no care of where or how the electricity is generated	Users get access to computing resources (processor, storage, data applications) as needed with no knowledge of where the resources are located
The power grid links power plants of different kinds	The grid links computing resources and provides mechanism to access them



* Ian Foster's and Carl Kesselman's seminal work, "The Grid: Blueprint for a new computing infrastructure" (2004)

Grid Computing

Characteristics:

1. Share more than information - data, computing power, applications in dynamic environment, multi-institutional and virtual organization
2. Efficient use of resources – People from many institutions work to solve a problem
3. Join local communities
4. Interactions with layer underneath must be transparent

Need for Grid Computing :

1. Exploiting under utilized resources
2. Data Visualization is important
3. Scientific problems are becoming complex and users need more accurate and precise solutions to their problems in minimum time – need scientific collaboration

Uses of Grid Computing :

1. Weather forecast
2. Detection and Modelling of natural disasters
3. Physics applications
4. Crystallography

Type of Grids :

1. Computational Grid
2. Data Grid
3. Collaboration grid

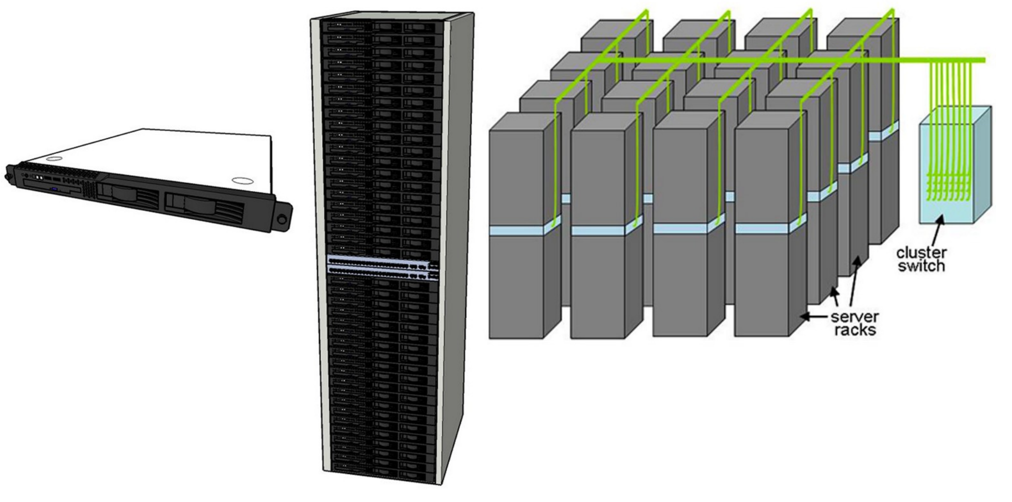
Data Centre Computing

When computing needs exceed computing resources, there are 2 options (not mutually exclusive)

1. Scale-up (vertical scaling): Increase resources of individual server

2. Scale-out (horizontal scaling): Increase the number of servers

Modern Data Centres lean towards Scale-out



Data Center Architecture

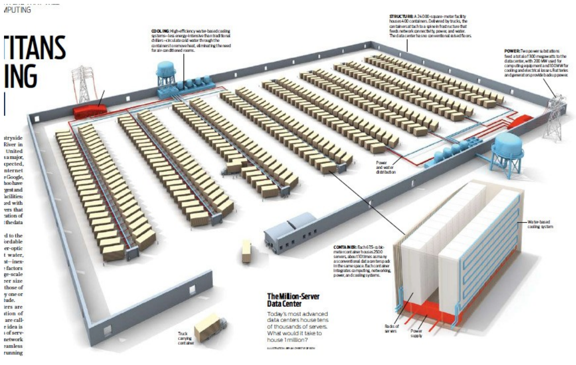
Data Centre

2005: Datacenter is new "server"

"Program" => Web search, email, map/GIS, …

"Computer" => 1000's computers, storage, network

Warehouse-sized facilities and workloads



Figures : Sun Microsystems, CNET, & datacenterknowledge.com

Data Centre as a Computer?

Facebook Datacenter, Clonee, Ireland

Google Datacenter, USc

Inside the Google Datacenter

Figure Source : Internet

Utility Computing Arrives

- Amazon Elastic Compute Cloud (EC2)
- Compute unit rental: $0.08-0.80/hr.

1 CU ≈ 1.0-1.2 GHz 2007 AMD Opteron/Xeon core

Instances	Platform	Cores	Memory	Disk
Small - $0.10/hr	32-bit	1	1.7GB	160 GB
Large - $0.40/hr	64-bit	4	7.5GB	850 GB
XLarge - $0.80/hr	64-bit	8	15.0GB	1690 GB

Billing rounded to nearest hour; pay-as-you-go storage also available

A new paradigm for deploying services

"Computing as Utility" – A vision from MULTICS c. 1969

Computing Today

SaaSComputing

The Million Server Datacenter
Providing IaaS

Slide Adapted from : Virtualization andCloud Computing, Sourav Bansal, ***http://www.cse.iitd.ernet.in/~sbansal/csl862***

What exactly is Cloud Computing?

“ It is Nothing New “

“ ...we’ve redefined Cloud Computing to include everything that we already do... I don’t understand what we would do differently ... other than change the wording of some of our ads.”

– Larry Ellison, CEO, Oracle
(Wall Street Journal, Sept. 26, 2008)

Must Read !!!

Above the Clouds:
A Berkeley View of Cloud Computing

(February 10, 2009)

abovetheclouds.cs.berkeley.edu

Cloud Computing

- Old Idea : Software as a Service (SaaS)

  Software hosted on local servers or desktops (dumb terminals)

- New : pay-as-you go utility computing

  Illusion of infinite resources on demand

  Fine grained billing : release == don't pay; Services – longer commitment, more $$$/hour

  Public vs. Private cloud

- Why Now?

  i. The Web "Space Race": Build-out of extremely large datacenters (10,000's of commodity PCs)

  ii. Driven by growth in demand (more users) -

  Infrastructure software : e.g., Google File System

  Operational expertise

  Discovered economy of scale: 5-7x cheaper than provisioning a medium-sized (100's machines) facility

  iii. More pervasive broadband internet

  iv. Open source software

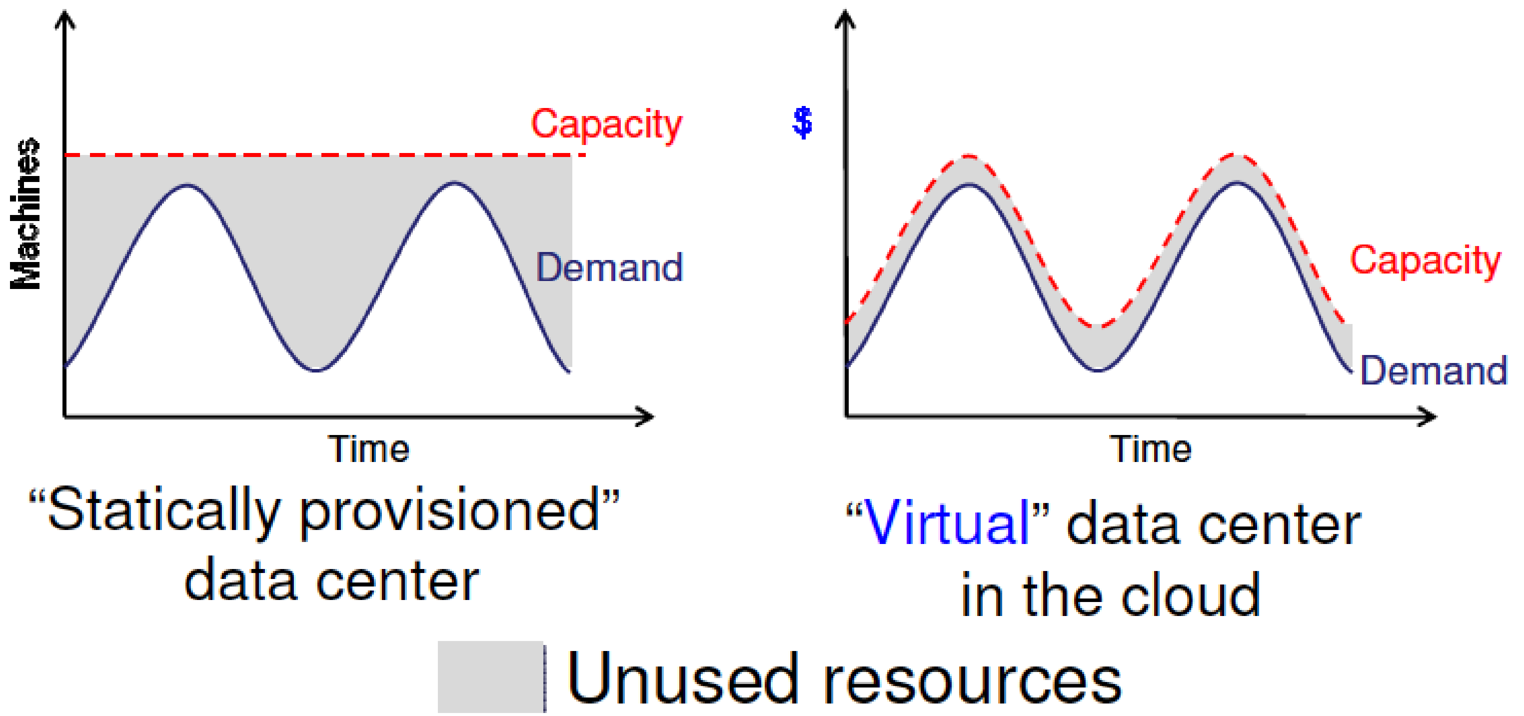
Cloud Computing - NIST

Cloud Computing is a model for enabling convenient, on demand network access to a pool of configurable computing resources (eg. Network, server, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models

http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc

Cloud Economics

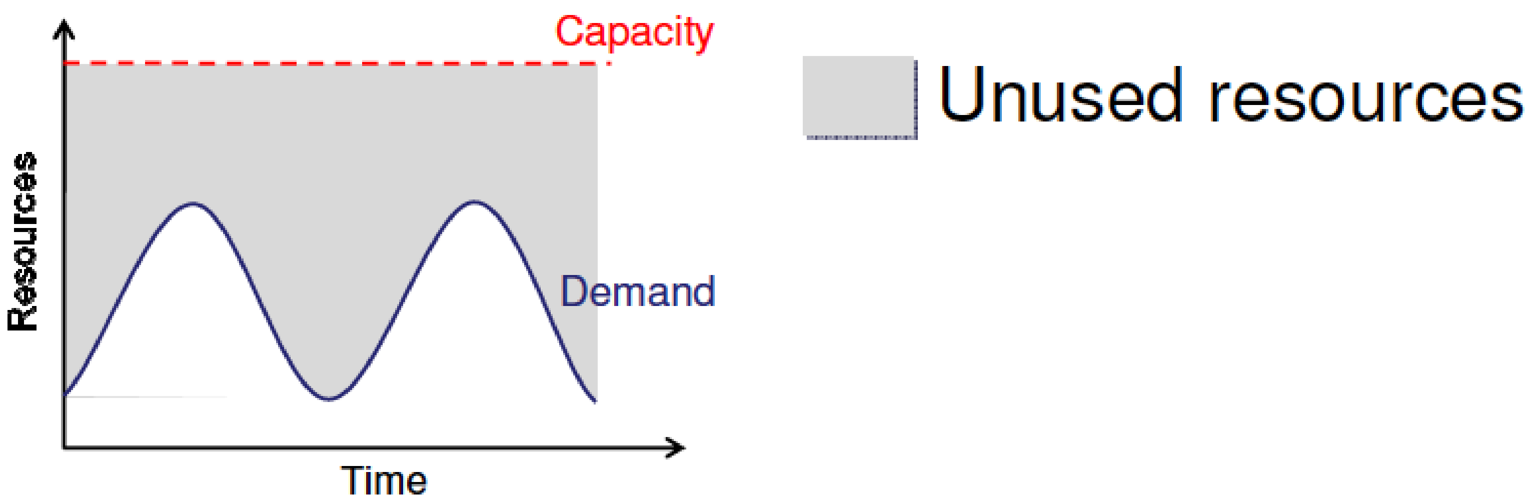
- Static Provisioning for peak – wasteful but necessary



Slide Adapted from : ***Above the Clouds:A Berkeley View of CloudComputing*** *(February 10, 2009)*

Risk of UnderUtilization

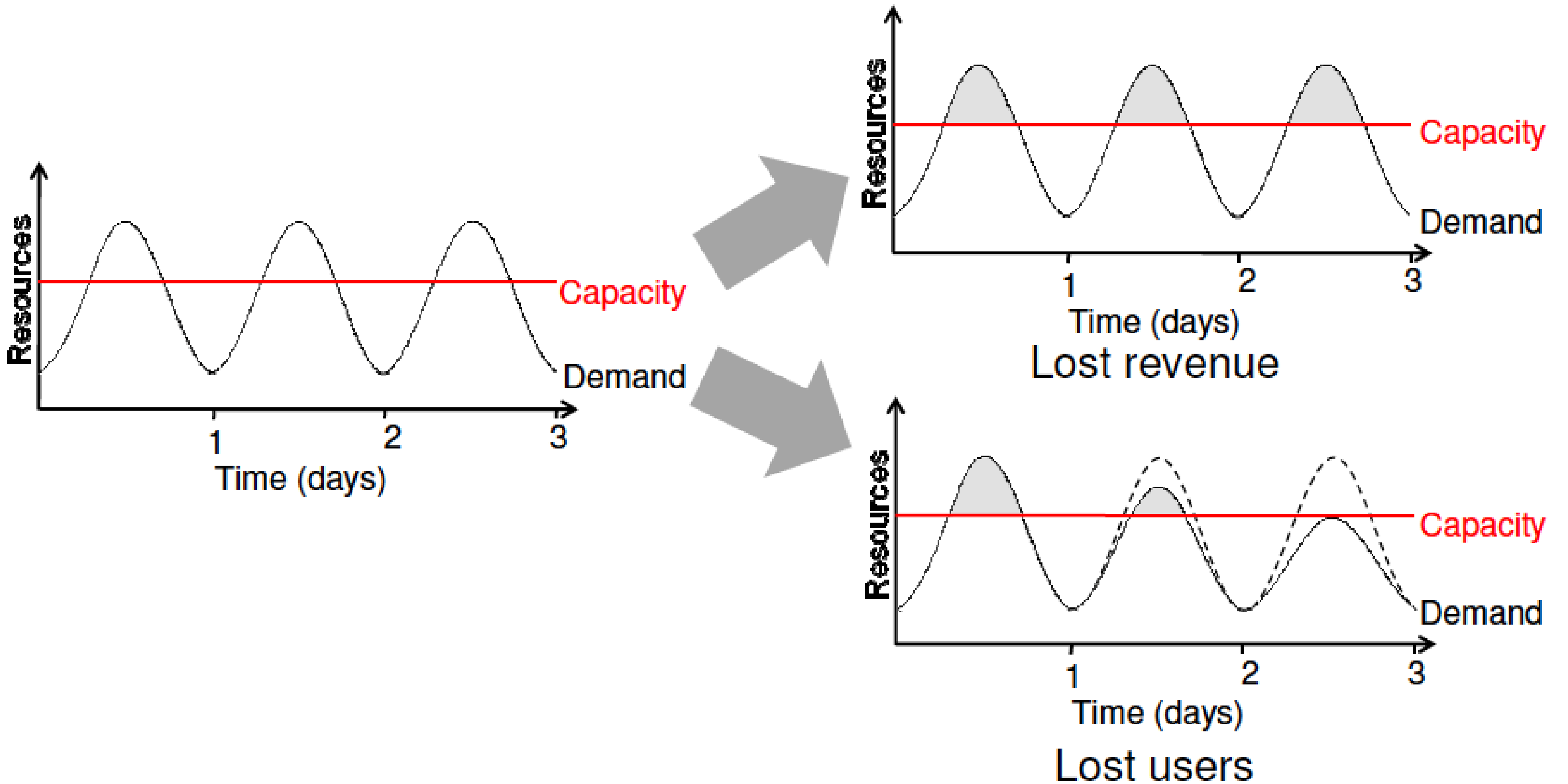
- Underutilization results if peak predictions are too optimistic



Static data center

Slide Adapted from : ***Above the Clouds:A Berkeley View of CloudComputing*** ***(February 10, 2009)***

Risk of Under Provisioning



***Slide Adapted from : Above the Clouds:A Berkeley View of CloudComputing** (February 10, 2009)*

Cloud Computing Vs. Grid Computing

- Distinctions not clear because Cloud and grid share similar visions
  - Reducing computing costs
  - Increasing flexibility and reliability by using third-party operated hardware
- Resource Sharing
  - Grid : collaboration
  - Cloud : assigned resources not shared
- Virtualization
  - Grid : Virtualization of data and computing resources
  - Cloud : virtualization of hardware and software platforms
- Security
  - Grid : security through credential delegation
  - Cloud : security through isolation
- Self Management
  - Grid : reconfigurability
  - Cloud : reconfigurability and self healing
- Payment mode
  - Grid : rigid
  - Cloud : flexible

Cloud Computing Characteristics

- Cost Associativity
  - 1,000 CPUs for 1 hour same price as 1 CPU for 1,000 hours
- On-demand self service
  Consumer can unilaterally provision computing capabilities without human interaction with service provider
- Broad network access
  Capabilities are available through network and suing standard mechanisms
- Multi-tenancy / Resource pooling
  Provider's resources are pooled to serve multiple customers
- Rapid elasticity
  capabilities can be rapidly and elastically provisioned
  Unlimited virtual resources
  Predicting a ceiling is difficult
- Measured service

Relevant Technologies in Cloud Computing

- Access
  - heterogenous set of thick and thin clients
  - high speed broad band access
  - Data centres (large computing capacity)
- Virtualization (Decoupling from physical computing resources)
  - Types : Hardware , Memory, Storage, Network
  - Hardware : emulation (QEMU) , para-virtualization (Xen), full virtualization(VMware)
  - Memory : Decouples RAM resources from individual systems
    Aggregates these resources into a virtualized memory pool available to any computer
  - Storage : abstracting logical storage from physical storage
  - Data Virtualization : Data as an abstract system independent of underlying data base systems, data structure
    And storage
  - Network : Virtualized Network addressing space with or within or across network subnets; VPN

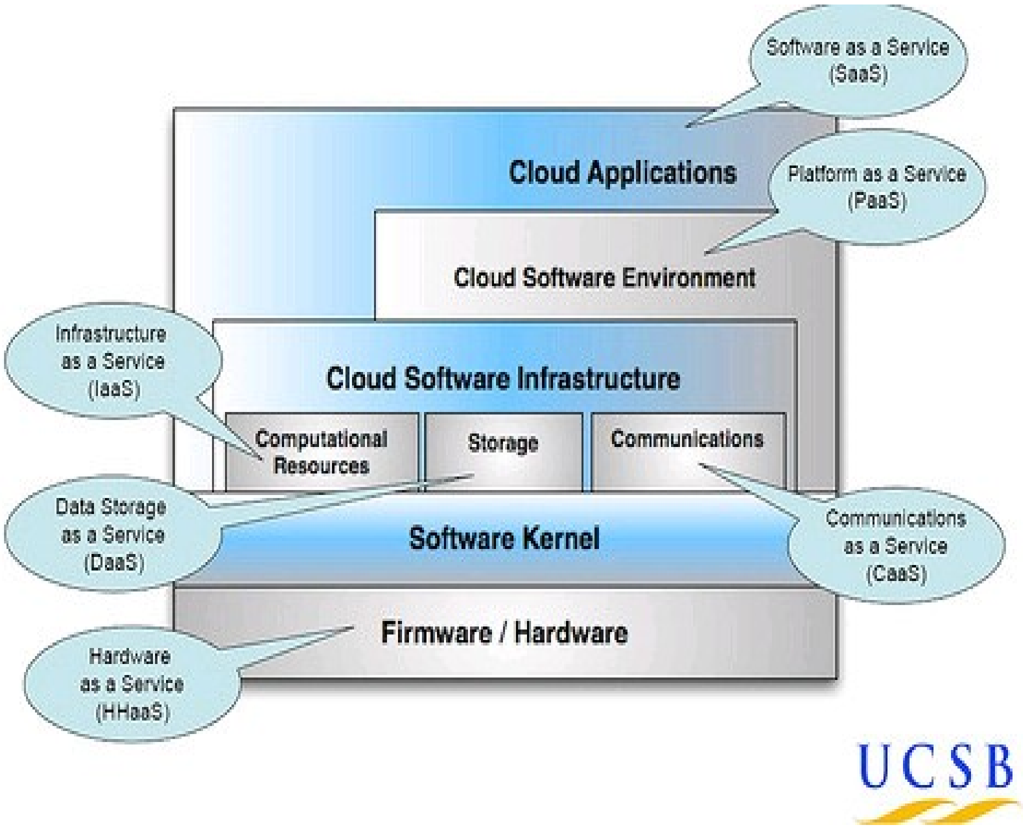
  Q : How do you measure Virtual Resources

  Amazon ECU (Elastic Compute Unit) : EC2 Compute unit equals 1.0 – 1.2 GHz 2007 Opteron or 2007 Xeon processor
- APIs

  Required for various operations and applications; No standards

Cloud Service Taxonomy

- Layer
  - Hardware – as - a Service (Haas)
  - Infrastructure -as -a Service (IaaS)
  - Data Storage – as- a- Service (DaaS)
  - Communication-as-a Service (CaaS)
  - Platform-as-a Service (PaaS)
  - Software-as-a Service (SaaS)



- Type
  - Public Cloud
  - Private Cloud
  - Hybrid Cloud

Software as a Service (SaaS)

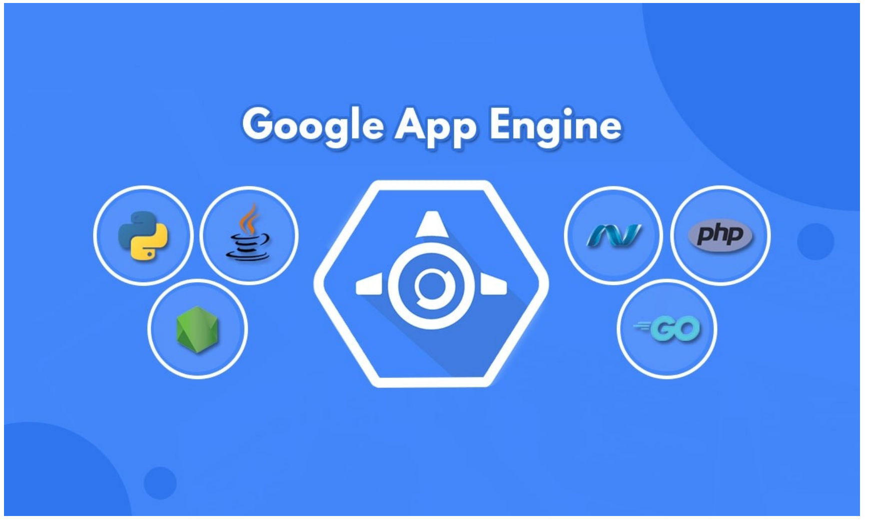
- Vendor / Provider controlled application accessed over the network

- Characteristics :
  - Network based access
  - Multi-tenancy
  - Single software release for all
  - browser based client and client tools

- Disadvantages :
  - dependency on network and cloud provider
  - limited client bandwidth
  - cloud service provider in charge of the data

Platform as a Service (PaaS)

- Vendor provided deployment environment tools and technology selected by the vendor (control over data life cycle)

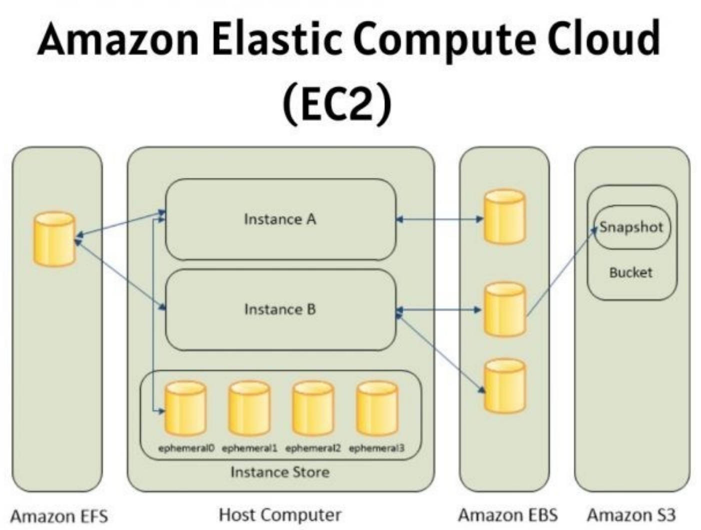


- Characteristics :
  - Rapid development and deployment
  - Small startup cost
  - multi-tenancy
  - load balancing and fail over
  - native integrated environment – performance, load

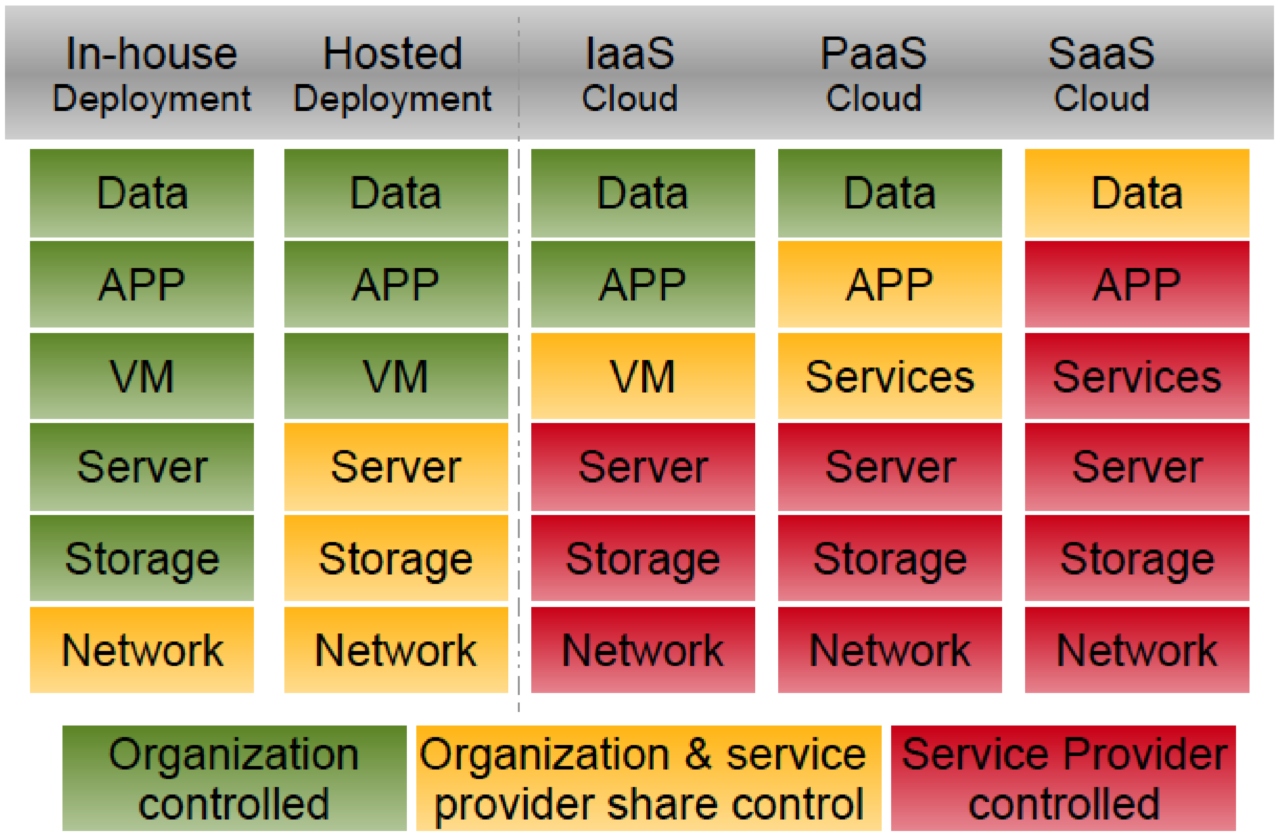
- Disadvantages :
  - inherits all disadvantages from SaaS
  - choice of technology limited to vendor provided tools and services

Infrastructure as a Service (IaaS)

- Vendor provided and consumer provisioned computing resources
- Characteristics :
  - Consumer is provided customized virtual machines
  - Consumer has control over OS, memory, storage, servers and deployment configurations
  - limited control over network resources
  - infrastructure scalability
  - economical cost

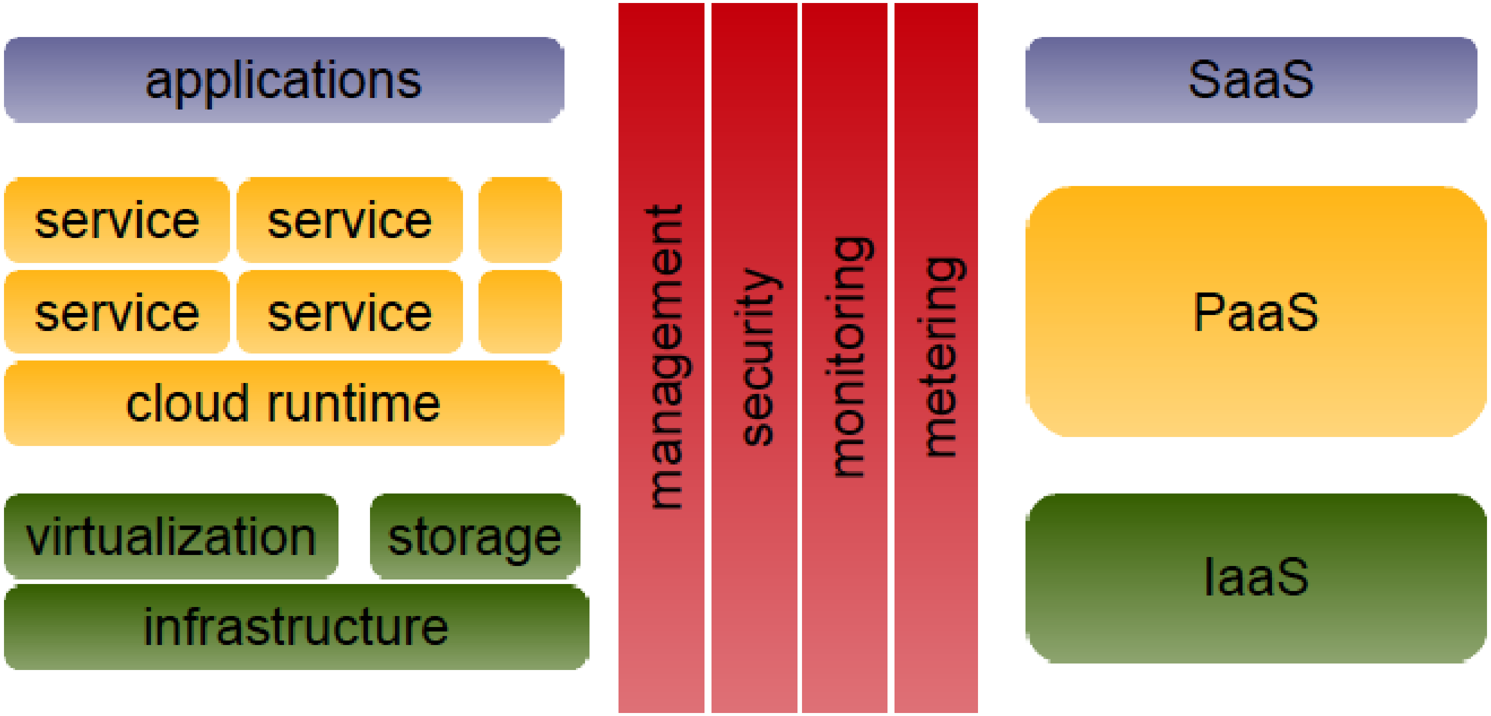


Cloud Services and Control



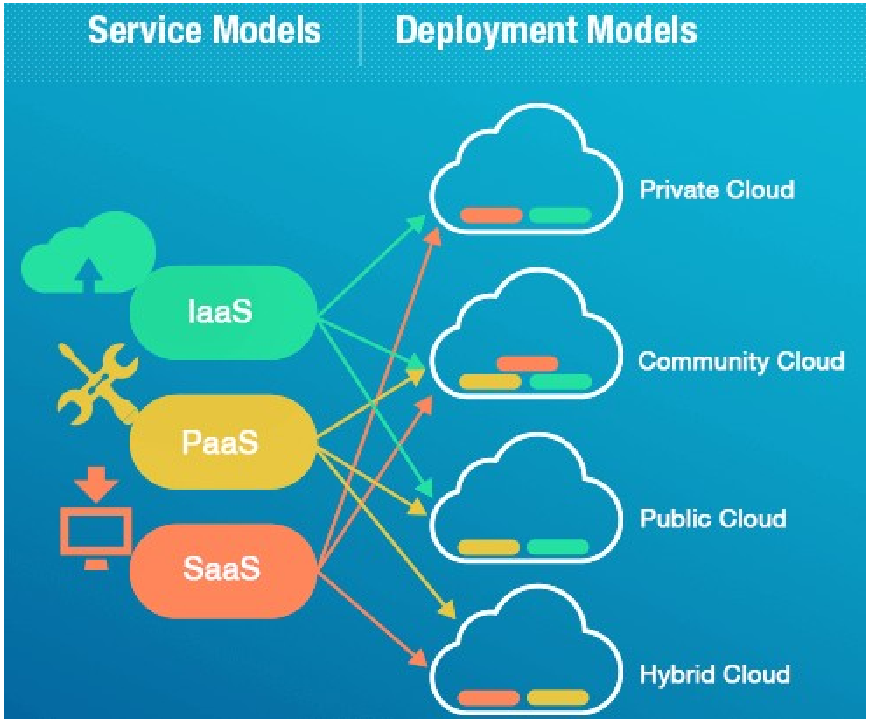
[1] **Visualizing the Boundaries of Control in the Cloud. Dec 2009.**
http://kscottmorrison.com/2009/12/01/visualizing-the-boundaries-of-control-in-the-cloud/

Cloud Services Reference Model



Cloud Deployment Models

- Private Cloud
  - infrastructure operated by an organization
  - managed by organization and third party



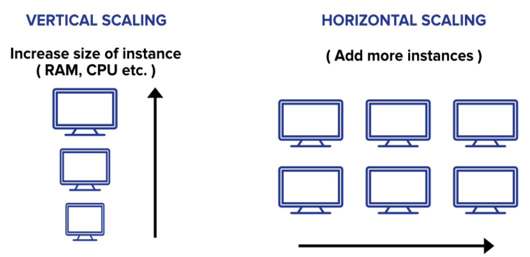
- Community cloud
  - supports a specific community
  - infrastructure shared by several organizations

- Public cloud
  - infrastructure is made available to public
  - owned by an organization selling cloud services

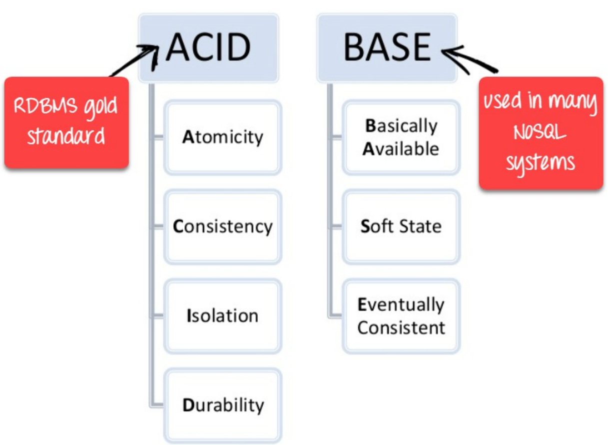
- Hybrid cloud
  - infrastructure is a composition of two or more cloud deployment models
  - enables data and application portability

Cloud Distributed Storage

- Distributed Storage : Two approaches of scaling
  - Vertical - bigger hardware
  - Horizontal – more hardware
    (horizontal partitioning, functional partitioning)



- CAP Theorem *
  - web services cannot ensure all the three of the following properties at once :
    - consistency : set of operations has occurred all at once
    - availability : an operation must terminate in an intended response
    - partition tolerance : operations will complete even if individual components are unavailable

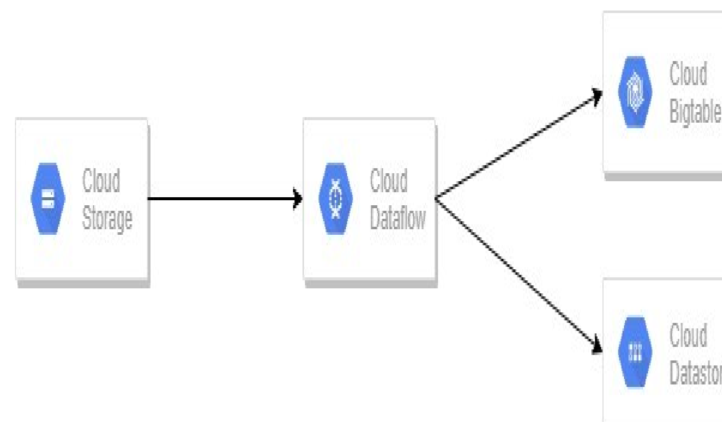


* Eric Brewer, University of California, Berkeley

Cloud Distributed Storage

- General Characteristics :
  - Simplified data model
  - Built on Distributed File System
    GFS – Google File System
    HDFS – Hadoop Distributed File System

- Highly Available – Relaxed Consistency
- Fault-Tolerant – Replication
- Eventual Consistency – all replicas will be updated in different time and in different order



Cloud Distributed Computation

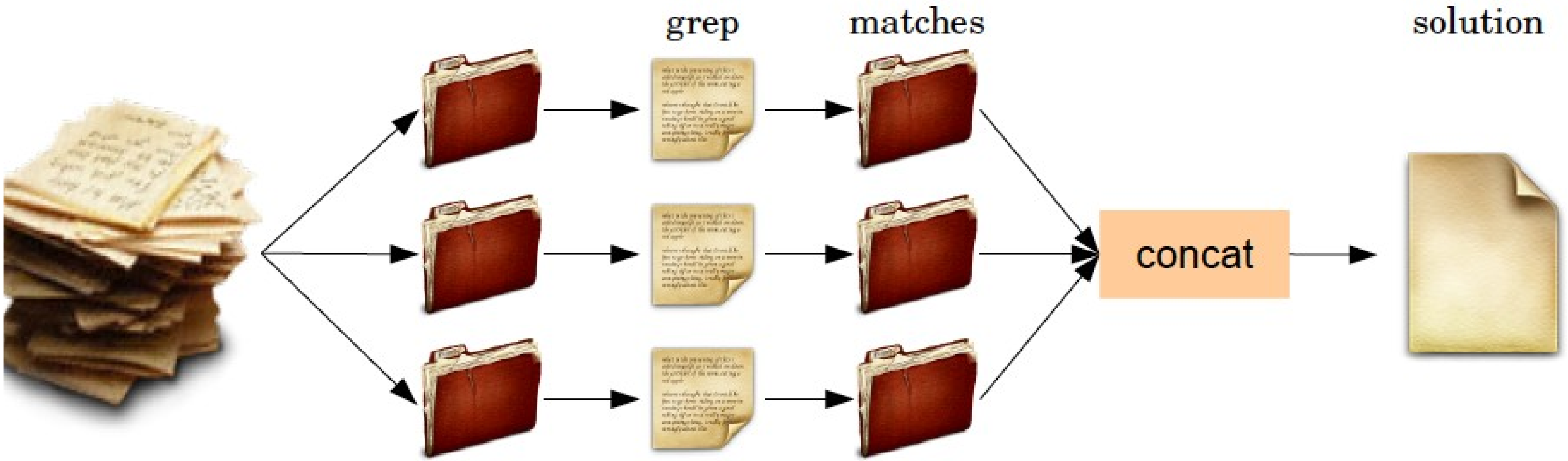
·Motivation
- distributed computing – thousands of computers
- large data sets
- fault tolerant
- easy to configure and manage

• Basic idea : functional programming
- large problem broken into simpler problems (functional decomposition)
- each small problem can be solved by a functional transformation of input data
- each problem can be executed in complete isolation

• Server to solve big problem

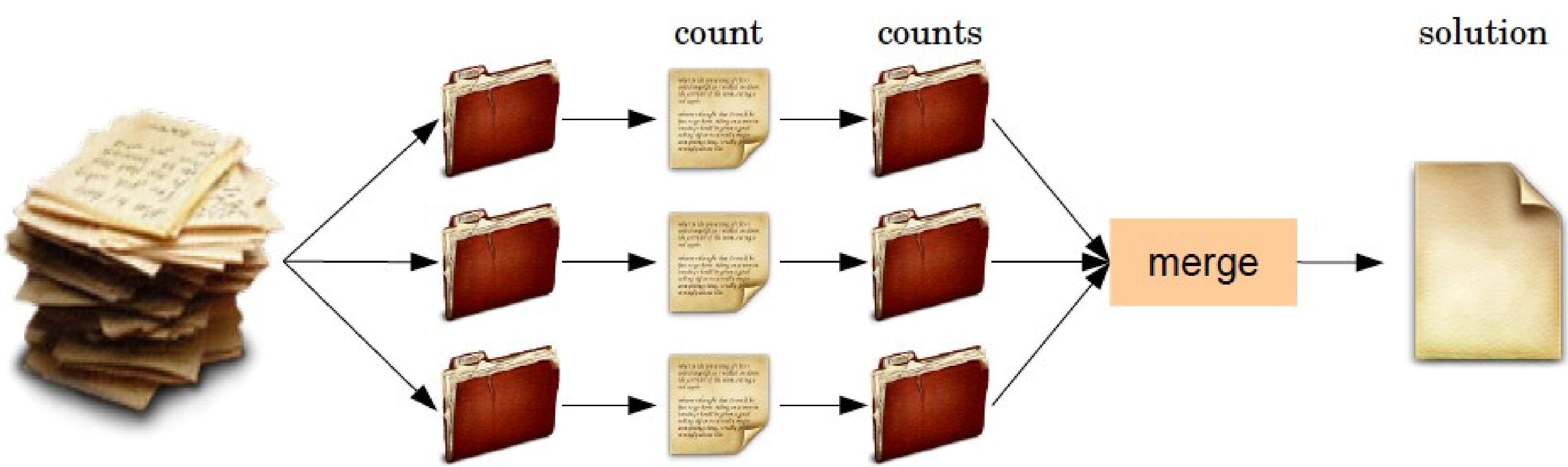
Cloud Distributed Computation

Distributed *grep*

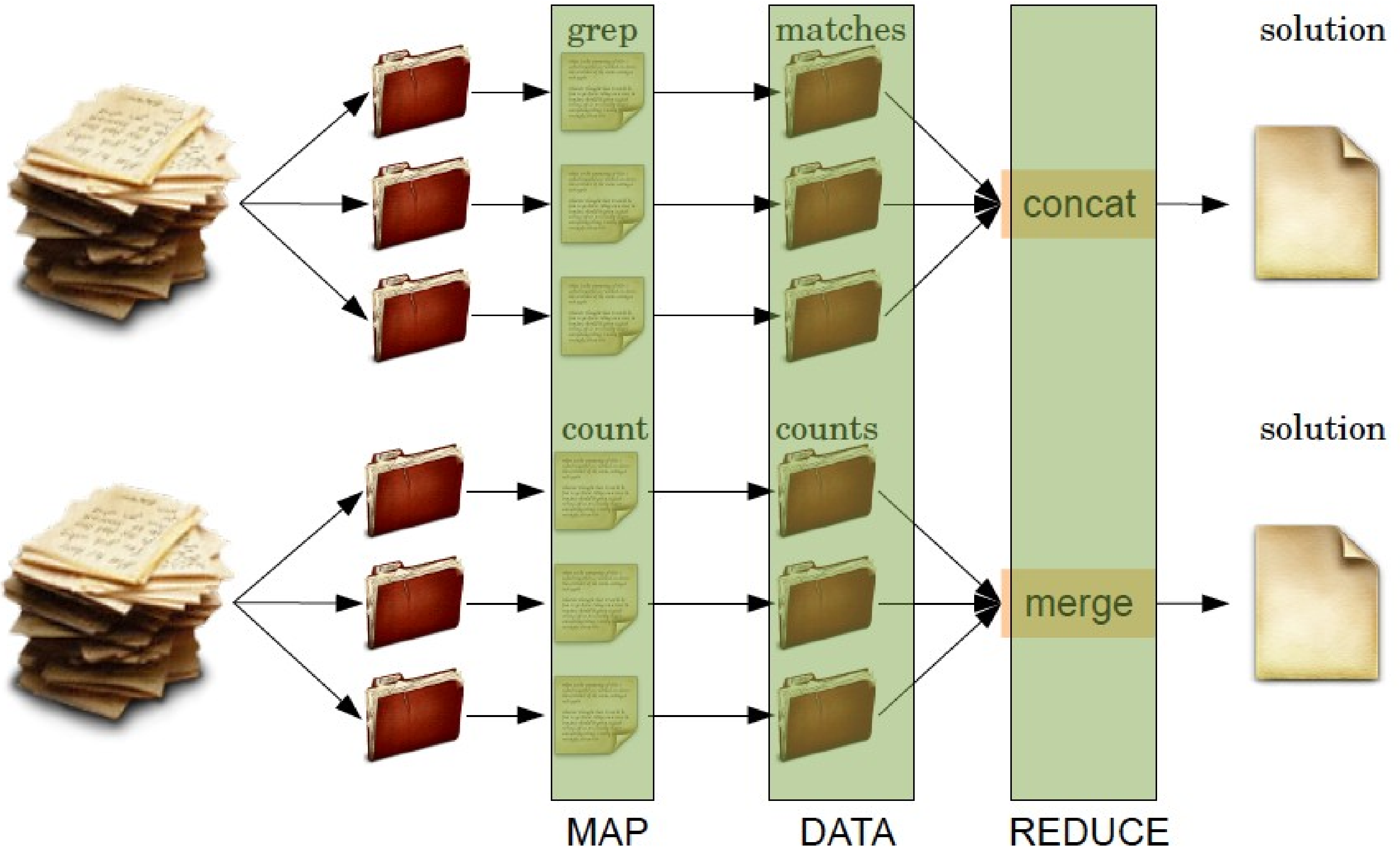


Cloud Distributed Computation

Distributed *word count (wc)*



Cloud Distributed Computation : Map Reduce



Obstacles and Opportunities for Cloud Computing

	Obstacle	Opportunity
1	Availability of Service	Use Multiple Cloud Providers to provide Business Continuity; Use Elasticity to Defend Against DDOS attacks
2	Data Lock-In	Standardize APIs; Make compatible software available to enable Surge Computing
3	Data Confidentiality and Auditability	Deploy Encryption, VLANs, and Firewalls; Accommodate National Laws via Geographical Data Storage
4	Data Transfer Bottlenecks	FedExing Disks; Data Backup/Archival; Lower WAN Router Costs; Higher Bandwidth LAN Switches
5	Performance Unpredictability	Improved Virtual Machine Support; Flash Memory; Gang Scheduling VMs for HPC apps
6	Scalable Storage	Invent Scalable Store
7	Bugs in Large-Scale Distributed Systems	Invent Debugger that relies on Distributed VMs
8	Scaling Quickly	Invent Auto-Scaler that relies on Machine Learning; Snapshots to encourage Cloud Computing Conservationism
9	Reputation Fate Sharing	Offer reputation-guarding services like those for email
10	Software Licensing	Pay-for-use licenses; Bulk use sales

Slide Adapted from : ***Above the Clouds:A Berkeley View of CloudComputing*** ***(February 10, 2009)***

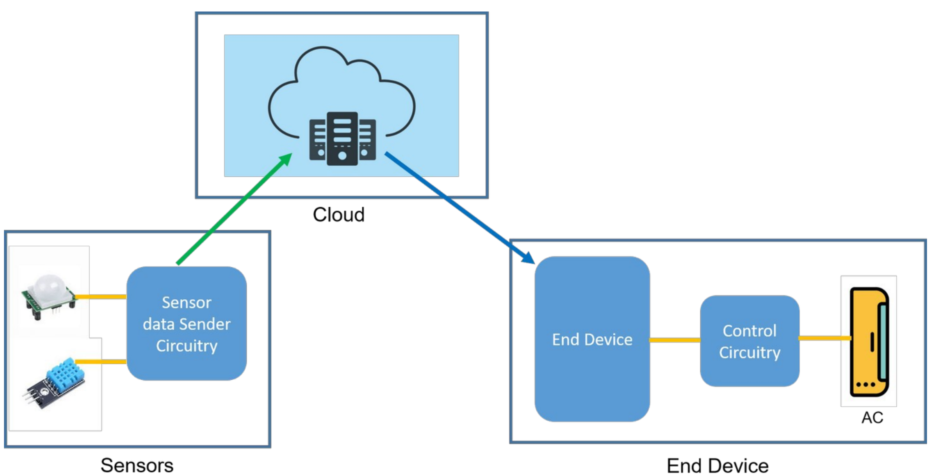
Cloud Computing for IoT

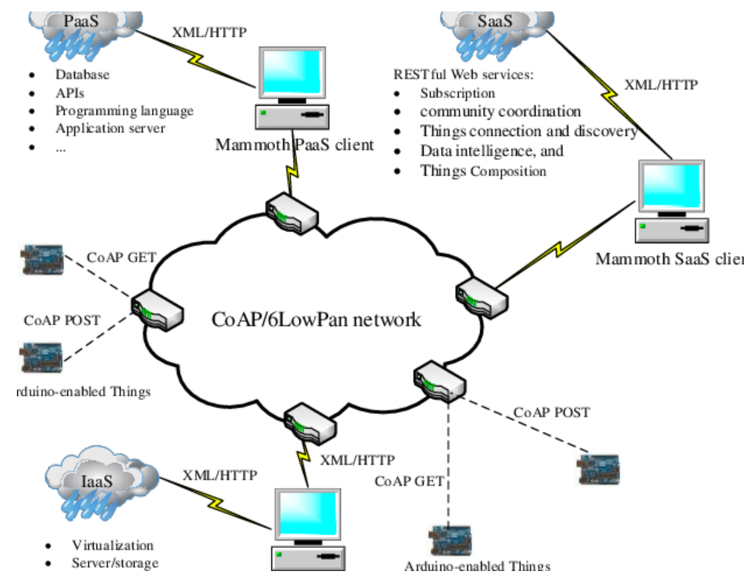
Cloud is an IoT enabler :

Huge amount of IoT Data needs storage, retrieval and management (sustained services)
Fast Analytics for Prediction and Critical Decision making

Benefits of Cloud in IoT :

Scalability ; Data Mobility; Time to Market; Security; Cost effectiveness

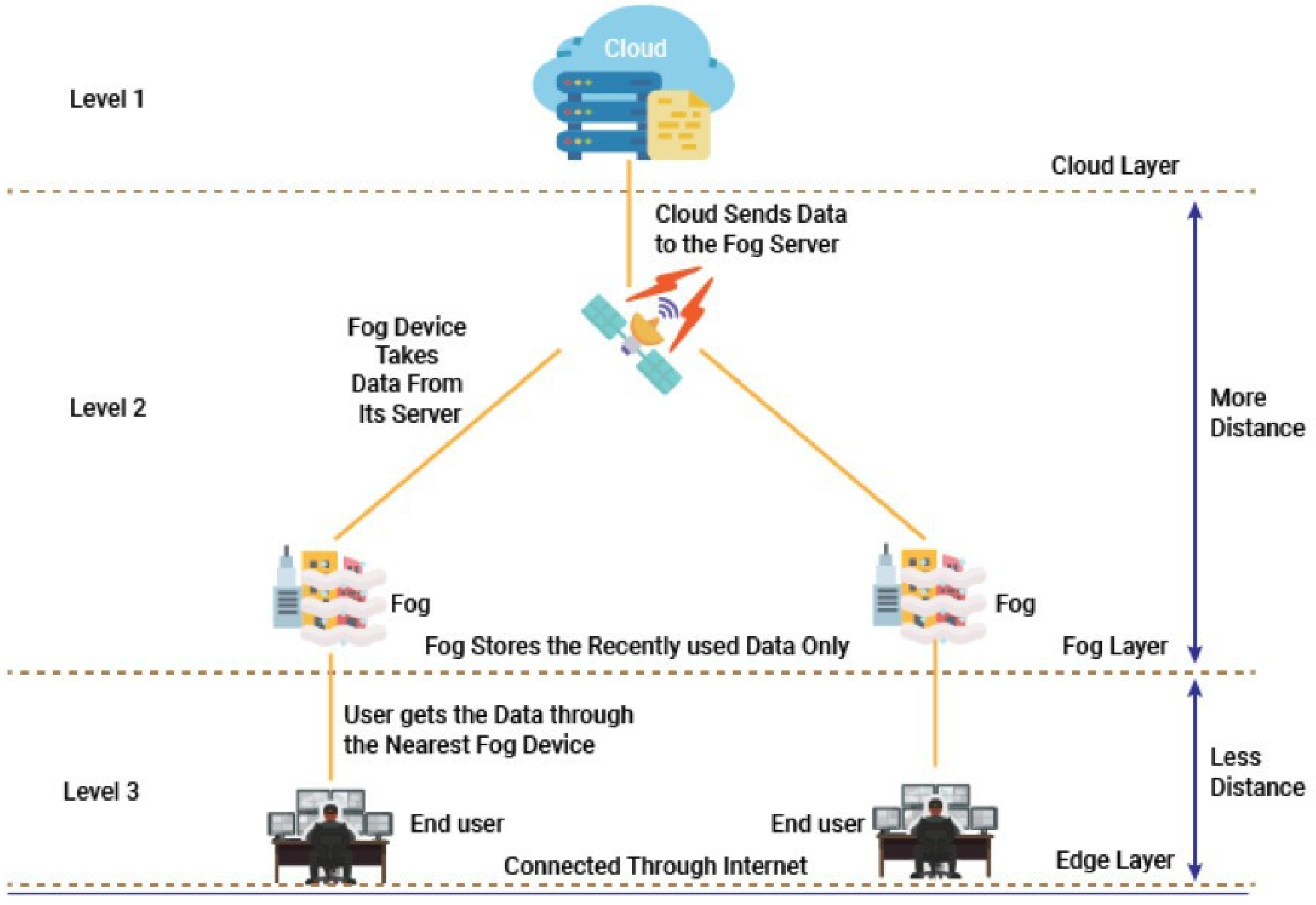




Challenges in Using Cloud Platform for IoT

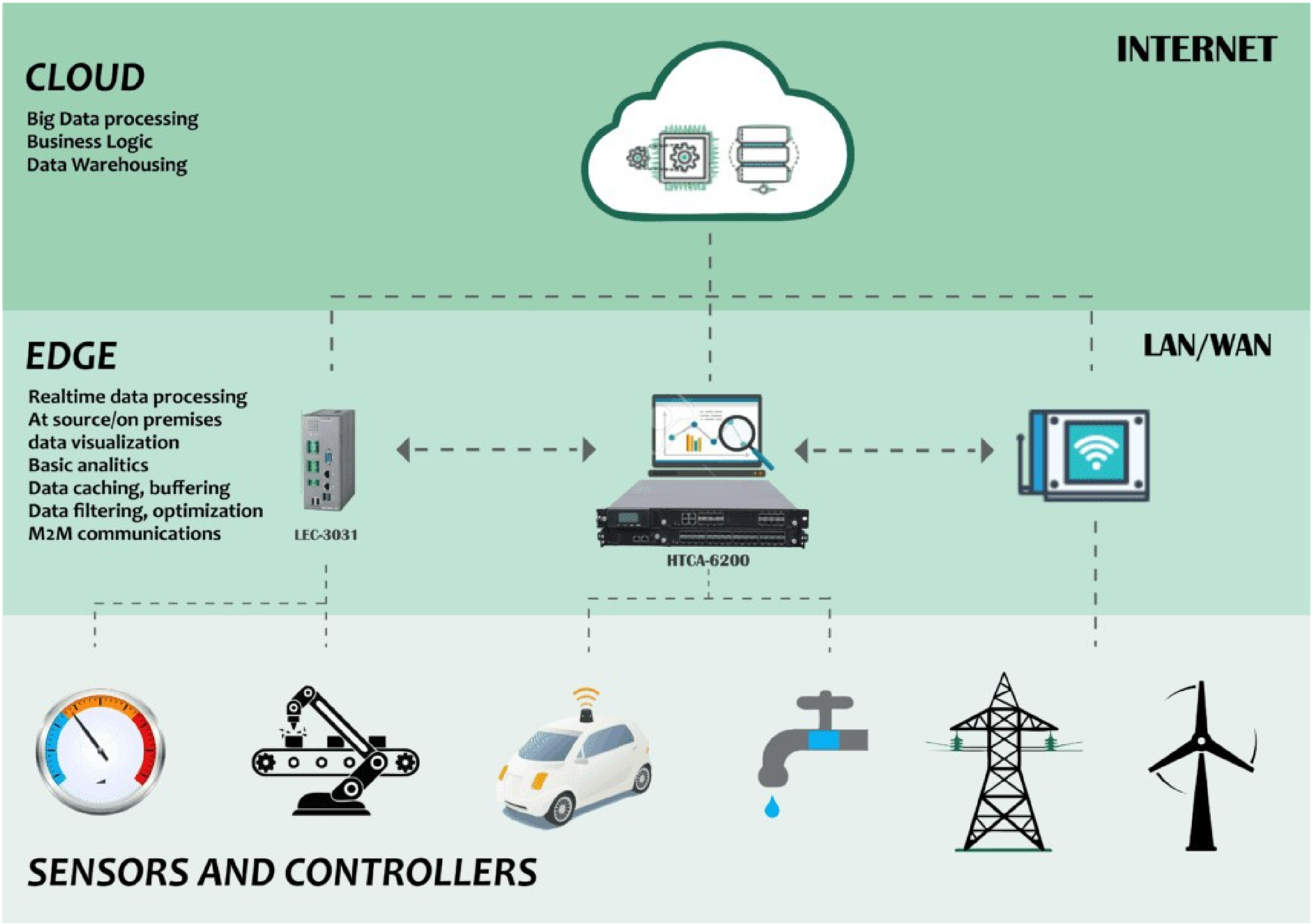
Source : Internet

Fog Computing



Source : Internet

Edge Computing

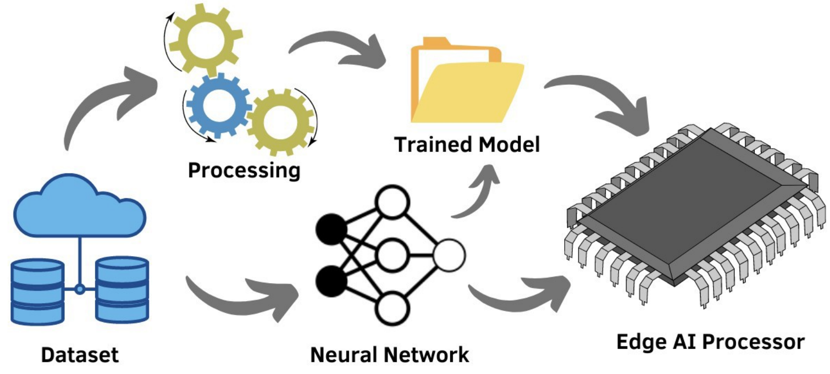


Benefits of Edge Computing :

1. Faster Response Time
2. Reliable Operations with intermittent connectivity
3. Security and Compliance
4. Cost-Effective Solutions
5. Interoperability between Legacy and Modern Devices

Source : Internet

Edge AI



Edge Computing + Artificial Intelligence = Edge AI

AI algorithms are executed locally on a hardware device using the data collected from Edge computing.

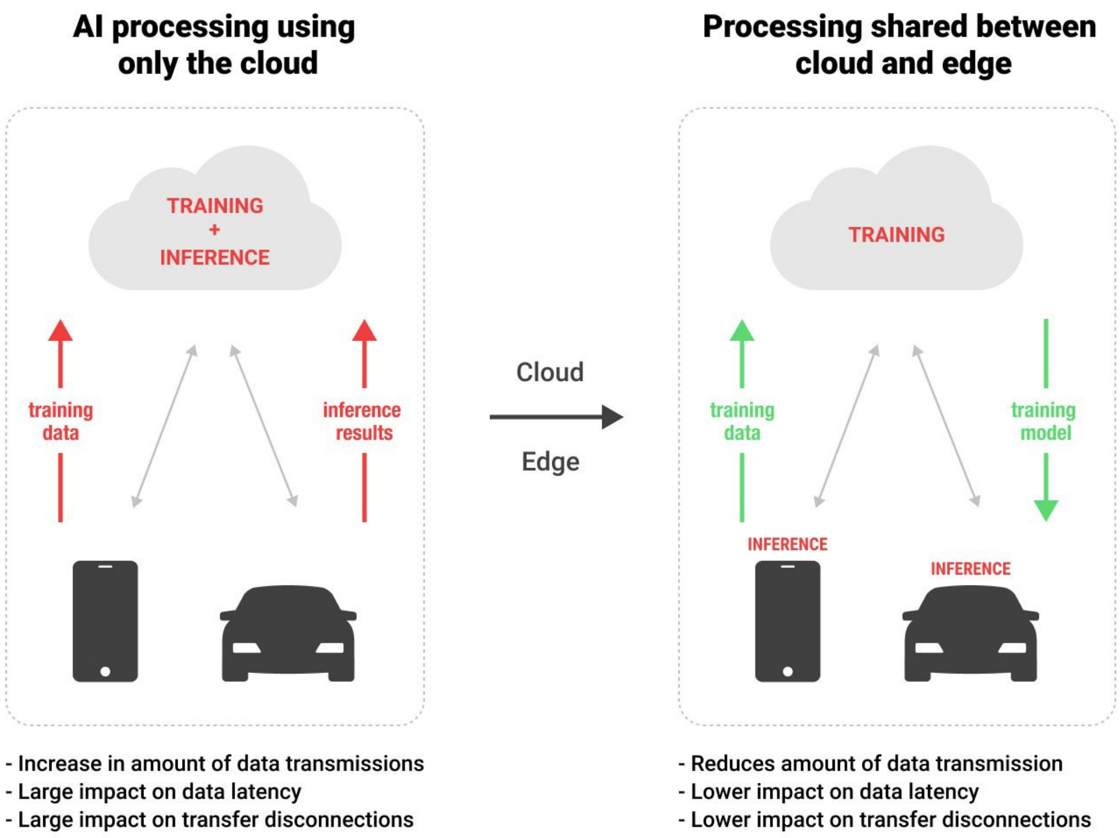
AI computation is done near the user at the edge of the network, close to where the data is located

Benefits :

1. Reduces latency
2. Reduces power consumption
3. Reduces Data Cost

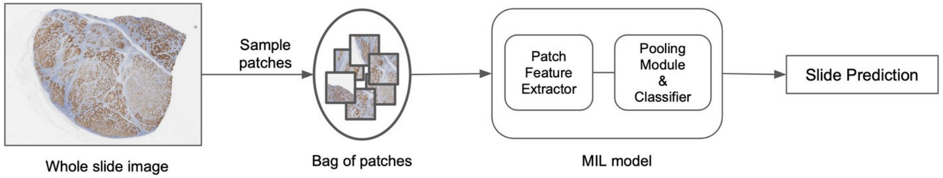
Source : Internet

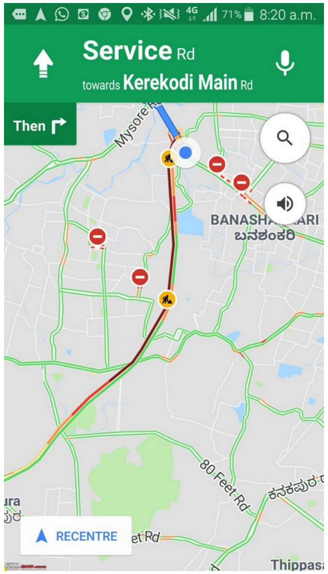
Edge AI Vs. Cloud AI

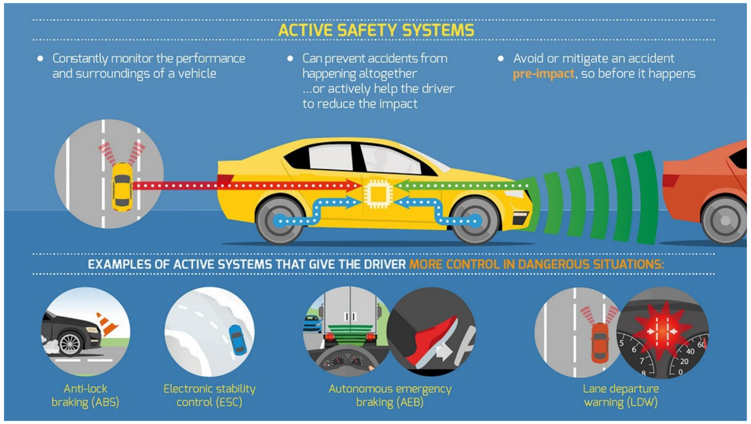


Source : Internet

Edge AI in Use

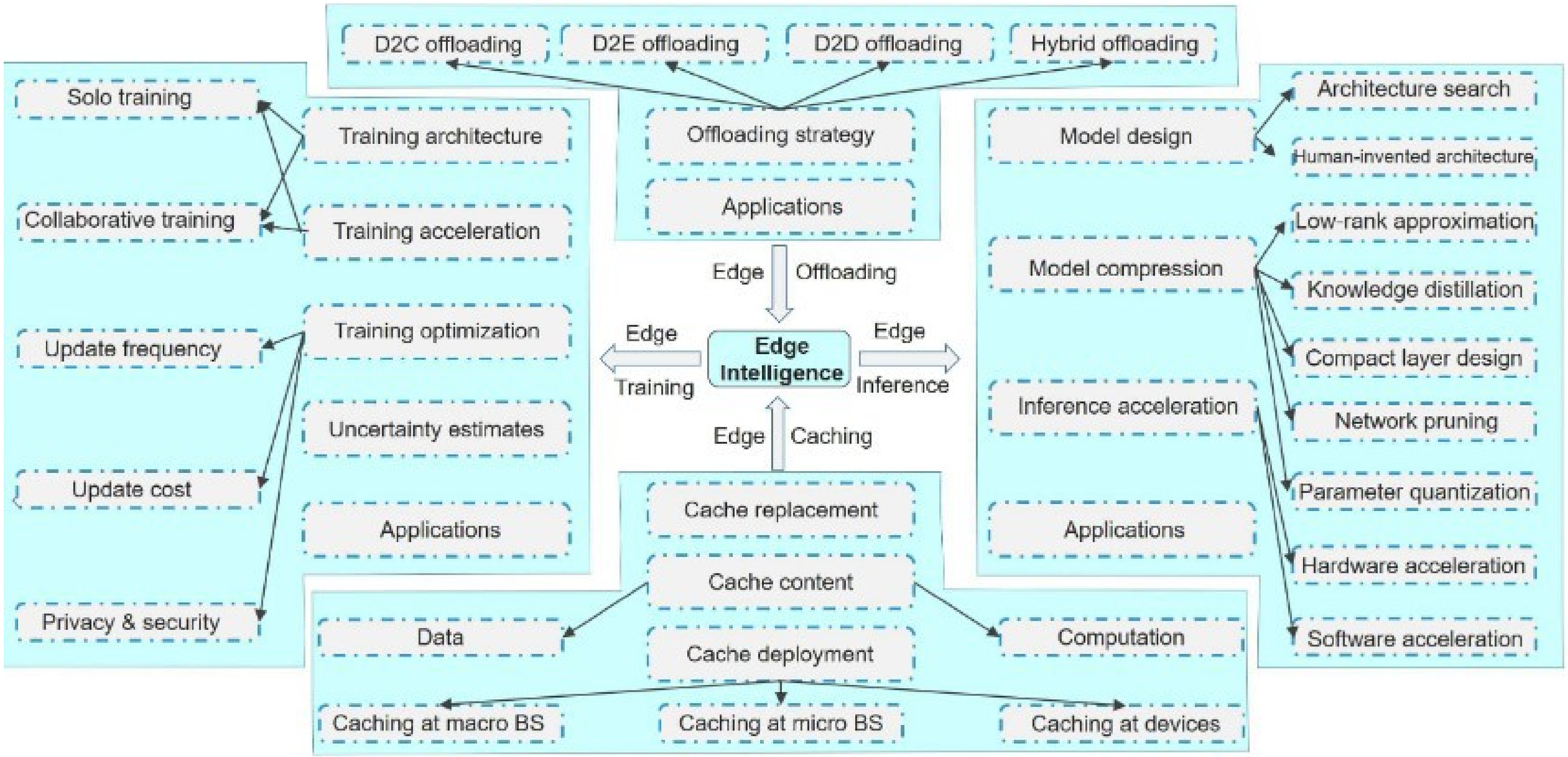






Source : Internet

Research Challenges in Edge AI



Source : Edge Intelligence: Architectures, Challenges, and Applications; Dianlei Xu et al.

Register yourself on Google Classroom

For any queries : bibhas.ghoshal@iiita.ac.in

Web : profile.iiita.ac.in/bibhas.ghoshal