BITS Pilani

BITS Pilani Presentation

Dr. Shwetha Vittal
Computer Science & Information Systems
shwetha.vittal@pilani.bits-pilani.ac.in

innovate achieve lead

BITS Pilani

Pilani Campus

CCZG527 Cloud Computing
Session 1 - Introduction to Cloud Computing

Dr. Shwetha Vittal

- **Assistant Professor at BITS Pilani – Off Campus**
- **Education:** PhD Computer Science & Engineering from **Indian Institute of Technology Hyderabad (IITH)**
- **Industry**
  - About 14 years of Industry experience: Majorly in Radisys Corporation, Juniper Networks, Mavenir Systems.
  - **Key Executions** Protocol stack, application developments, and performance optimizations for VoIP, 3G, 4G (LTE), and 5G products.
- **Teaching**
  - Cloud Computing, Telecom Network Management, Computer Networks
- **Research Areas**
  - 5G Core, Network slicing, High Availability, Resilience, Orchestration
  - IEEE/ACM Publications

Course Objectives

The course aims at:	
CO1	Students will learn the fundamental ideas behind Cloud Computing, the evolution of the paradigm, its applicability; benefits, as well as current and future challenges;
CO2	Students will learn the basic idea and principles in Cloud infrastructure management
CO3	Students will learn about cloud components Compute, Storage and Networking technologies
CO4	Students will learn a variety of programming models and develop working experience

Textbook(s)

T1	Dinkar Sitaram and Geetha Manjunath. Moving to the Cloud. Syngress (Elsevier) Pub, 2011
T2	Marinescu, ***Cloud computing theory and practice***, Morgan Kaufmann Publisher

Modular Structure

Module #	Module name
1	Introduction to Cloud Computing
2	Virtualization Techniques and Types
3	Infrastructure as a Service
4	Platform as a Service and SaaS
5	Managing Virtual Resources on the Cloud: Provisioning and Migration
6	Capacity management and Scheduling in cloud computing
7	Issues and Challenges : Availability, Multi-Tenancy, Security and SLA
8	Application Development and Deployment on Cloud

Evaluation Scheme

Item	Name	Type	Duration	Weight	Day, Date, Session, Time
EC-1	Quiz – 1 & 2	Online		10%	TBA
	Experiential learning Assignment-I (Presentation)	Take Home	15-20 days	20%	TBA
EC-2	Mid-Semester Test (Topics in Session Nos. 1 to 8)	Closed Book	~2 hours	30%	Per Programme schedule
EC-3	Comprehensive Exam (All topics (Session Nos. 1 to 16))	Open Book	~3 hours	40%	Per Programme schedule

Pedagogy

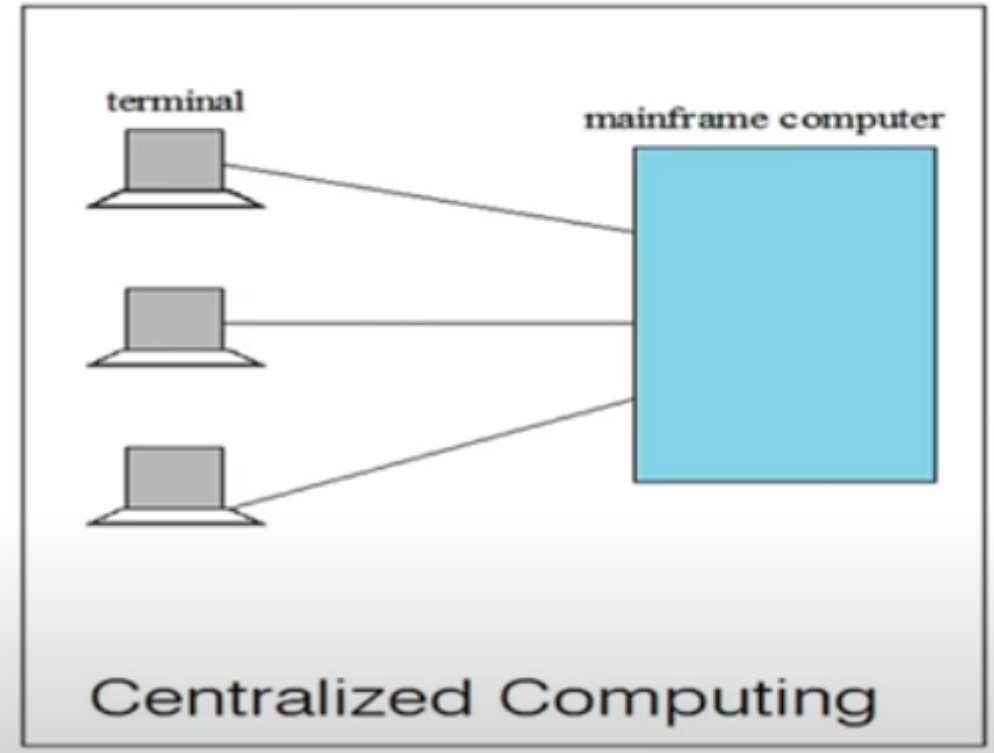
Sl No	Name	Type	Duration	No. of Sessions
1	Class lectures	Microsoft Teams (Sun, 3:40 PM to 5:40 PM)	2.00 hr	16 (as per calendar)

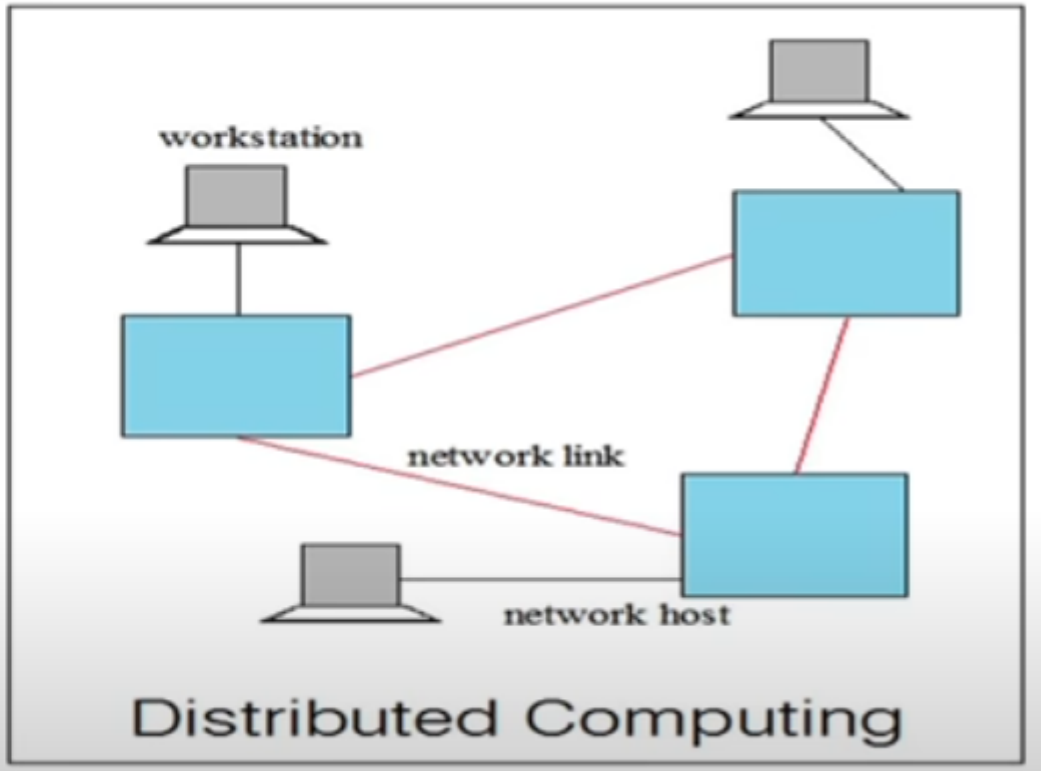
Introduction to Cloud Computing, Services, and Deployment models

Agenda

- Introduction to Cloud Computing – Origins and Motivation
- 3-4-5 rule of Cloud Computing
- Types of Clouds and Services
- Cloud Infrastructure and Deployment

Centralized Vs Distributed Computing

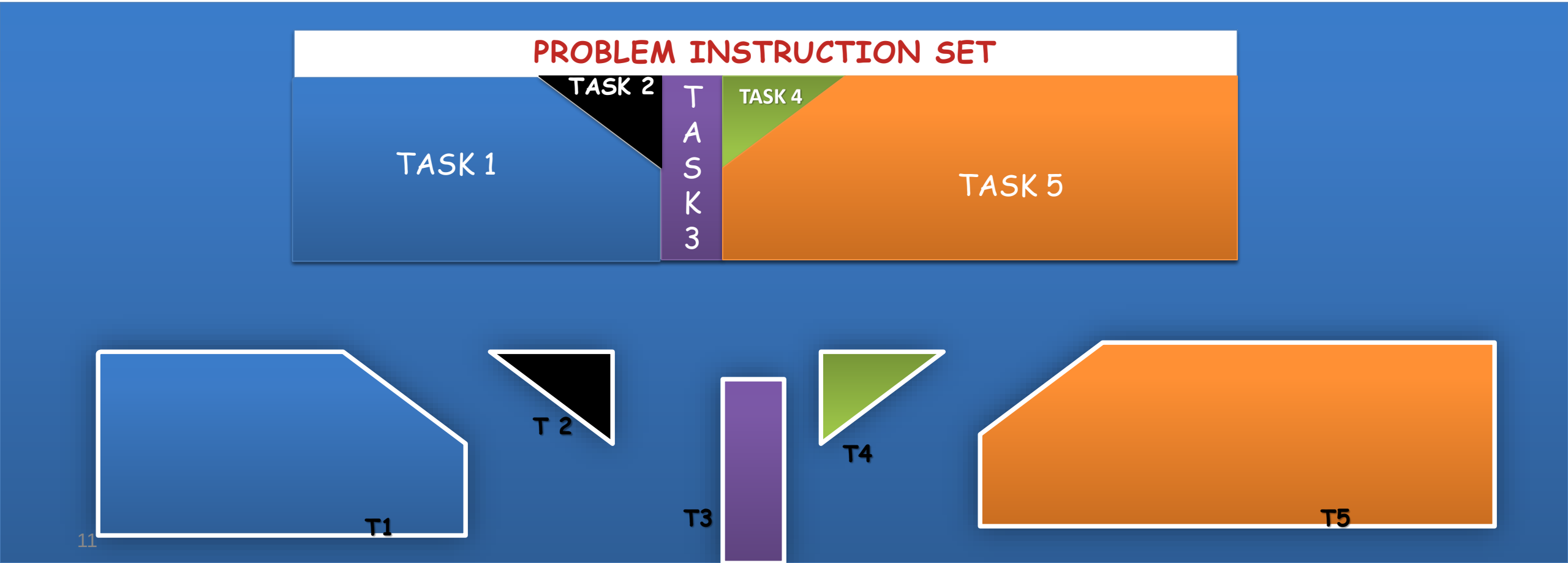




- Early computing – was a uniprocessor – Centralized computing
- Distributed Computing ?

Distributed Computing

In distributed computing a program is split up into parts that run simultaneously on multiple computers communicating over a network



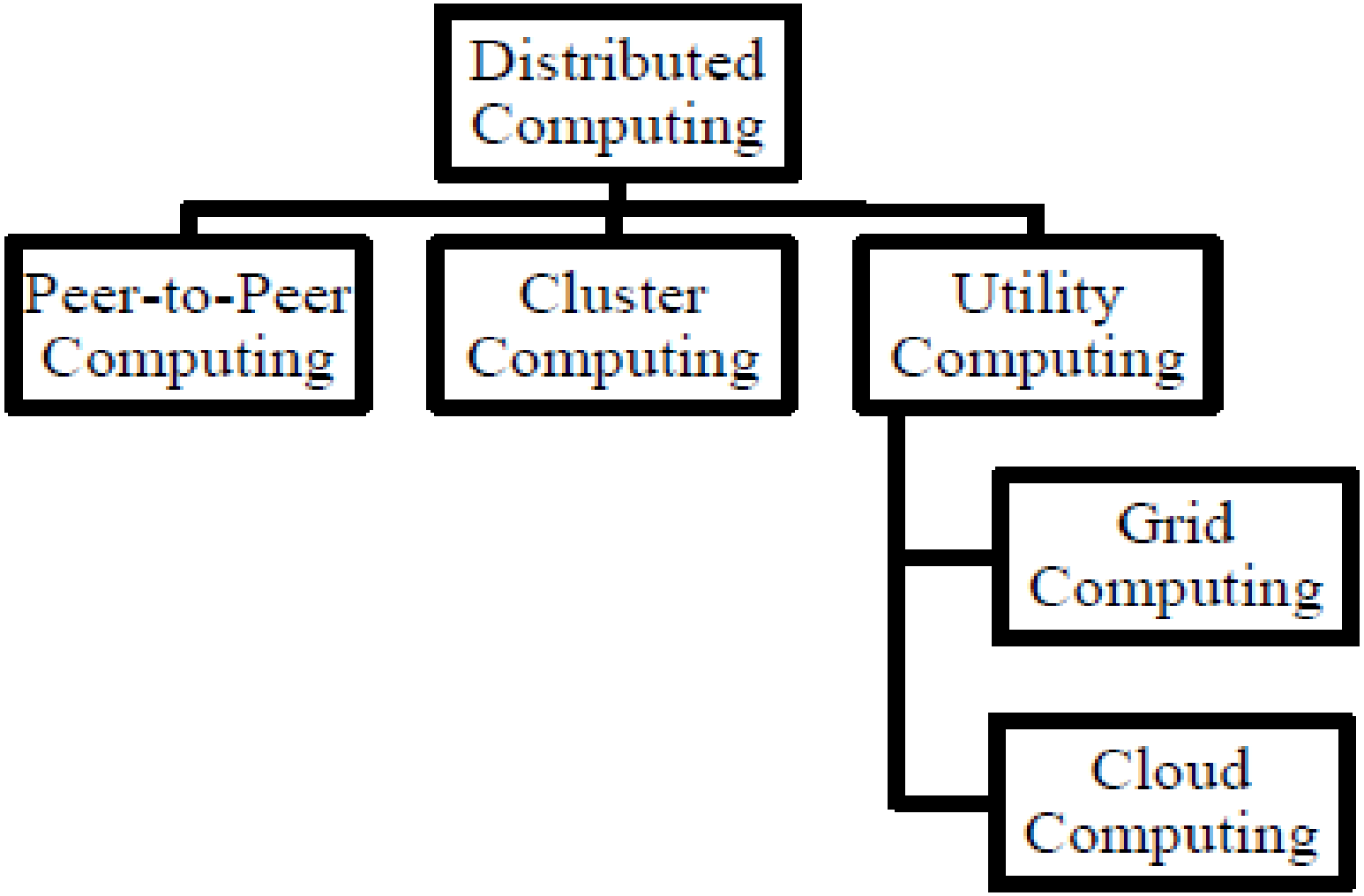
Distributed Computing

- Consider if there are ***n*** systems connected in a network
- Then we can split one program into ***n*** different tasks and
- Compute them **Concurrently.**

Why do we need Distributed Computing?

- Computation requirements are ever increasing
- Silicon based (sequential) architectures reaching their limits in processing capabilities (clock speed) as they are constrained by.
- Significant development in networking technology is paving a way for network-based cost-effective parallel computing.
- The parallel processing technology is mature and is being exploited commercially.

Distributed Computing Models



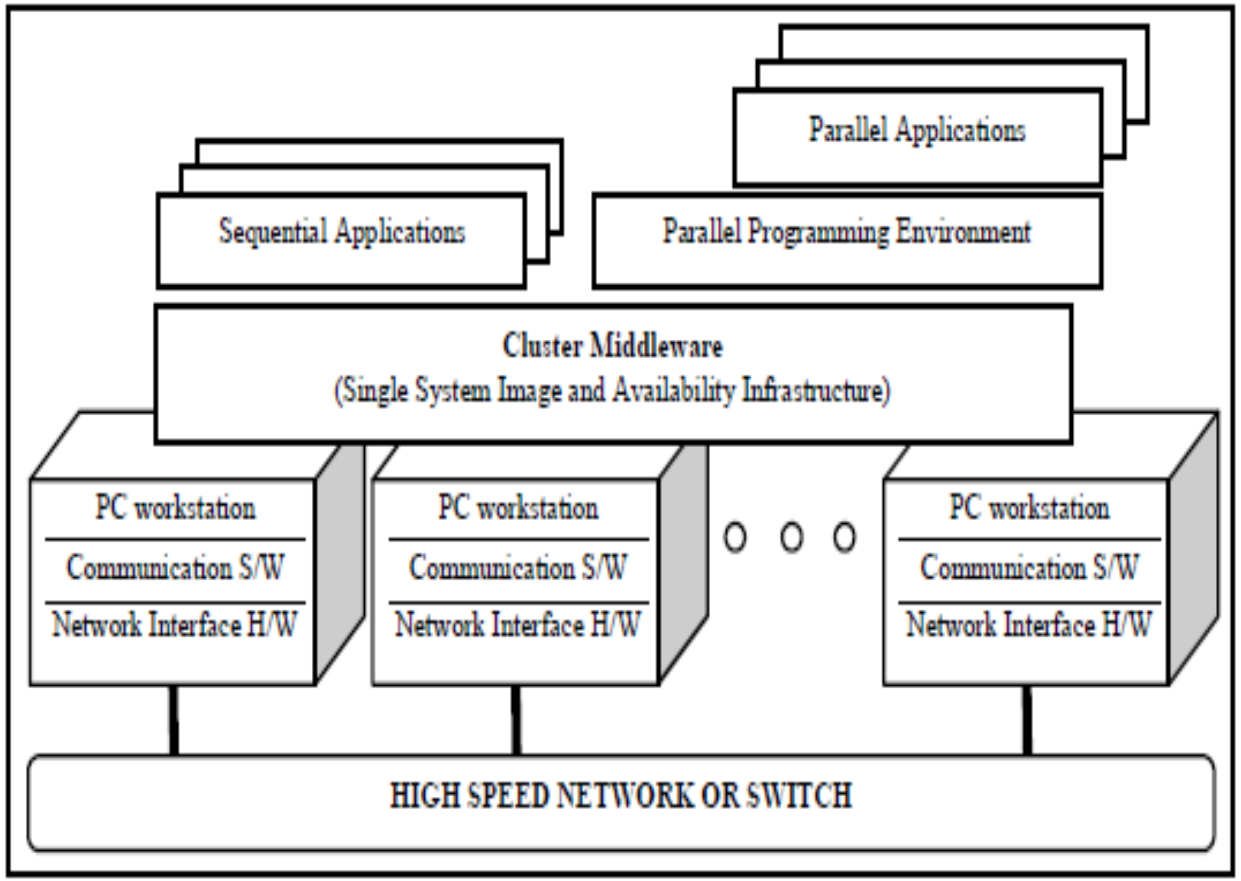
Peer-to-peer Computing

In a P2P system,

- Every node acts as both a client and a server, providing part of the system resources.
- Peer machines are simply client computers connected to the Internet.
- All client machines act autonomously to join or leave the system freely.
- This implies that no master-slave relationship exists among the peers.
- No central coordination or no central database is needed.

Cluster Computing

- Comprises a set of independent or stand-alone computers and a network interconnecting them.
- Co-ordinated Use: It works cooperatively together as a single integrated computing resource.
- A cluster is local in that all of its component subsystems are supervised within a single administrative domain, usually residing in a single room and managed as a single computer system
- Single System Image (SSI): Makes a cluster appear like a single machine to the user.



Cluster Types

- High Availability/Failover Cluster
  - System Availability, Reliability, Fault Tolerance
- Load balancing Cluster
  - Scalability
- Parallel / Distributing Processing Cluster
  - High Performance in combination with load balancing - scalability

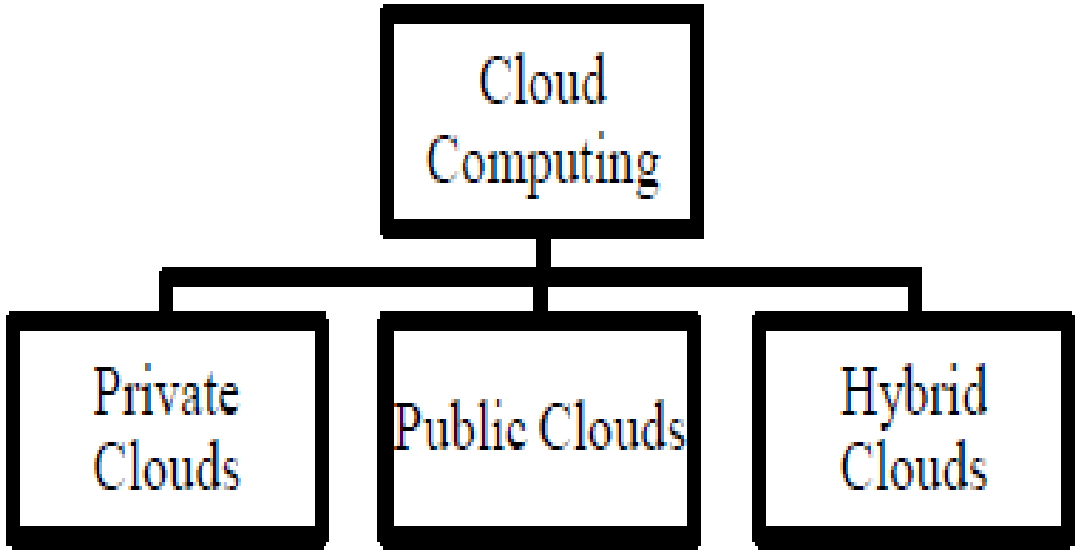
Grid Computing

- Making computer power as easy to access as power grid.
- Grid computing **enables coordinated resource sharing and problem solving in dynamic**, multi-institutional virtual organizations.
- Leverages **hardware/software virtualization, distributed sharing of the virtualized resources.** – Hardware, software, network services, computing power.
- Grid is often constructed across LAN, WAN, or Internet backbone networks at regional, national, or global scales.
- Enterprises or organizations present grids as integrated computing resources.
- Provides a unified view – target solving problem - scientific research, business logic

Utility Computing

- Based on a service provisioning model, where users (consumers) pay providers for using computing power only when they need to.
- Utility computing focuses on a business model, by which customers receive computing resources from a service provider on pay basis.
- All grid/cloud platforms are regarded as utility service providers

Cloud Computing



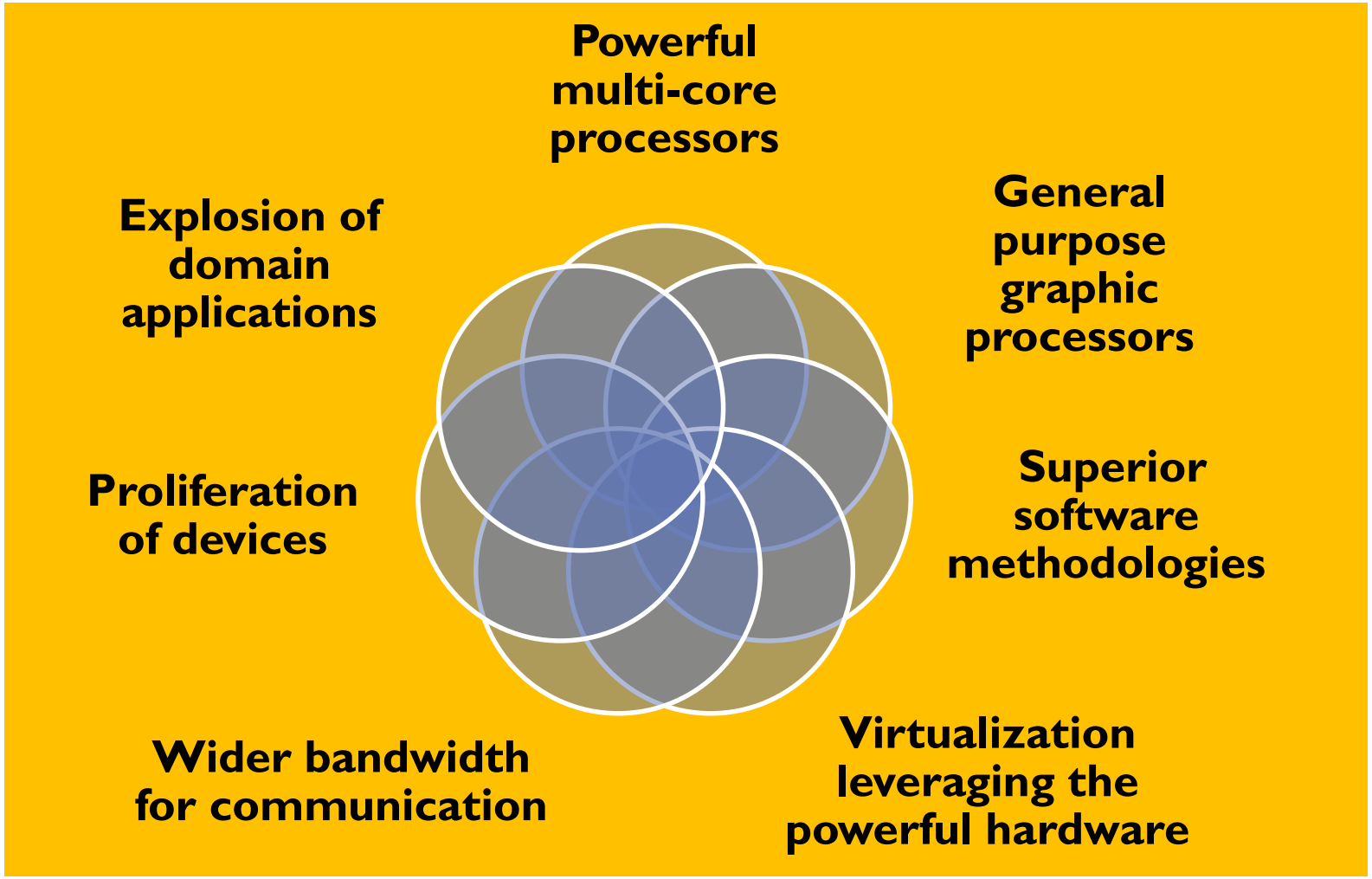
Cloud computing is a computing paradigm that involves outsourcing of computing resources with the capabilities of

* resource scalability,
* on-demand provisioning
* with little or no up-front IT infrastructure investment costs.

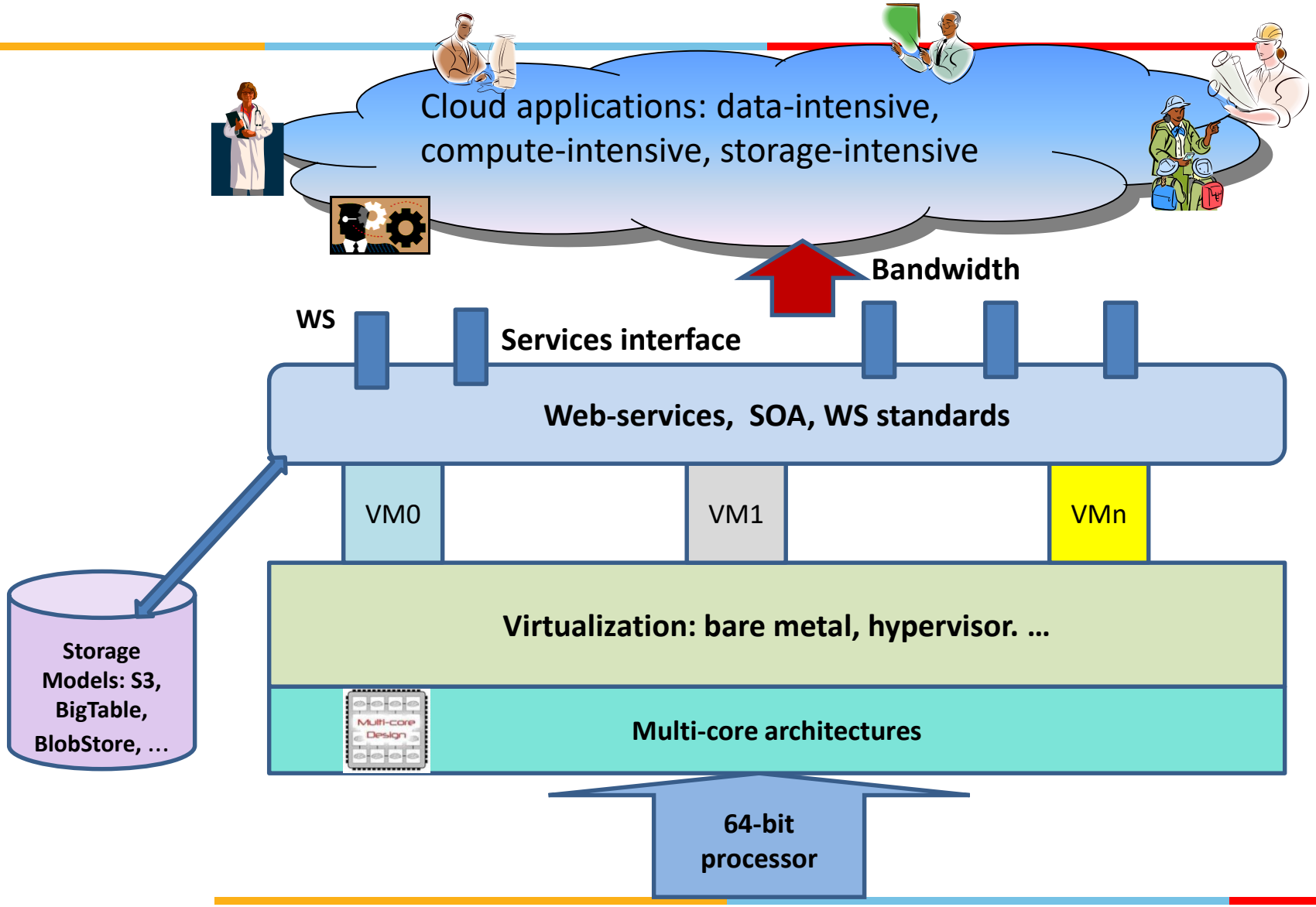
Evolution of Web

1. Web 1.0 – 1990s, One way service from enterprises /institutes to users
2. Web 2.0 -2000s
   - Allowed users to upload information to the web
   - Rapid growth in user generated content – blogs, reviews, tags, annotations
   - Enabled social networking
3. Information Explosion – news, photographs, ~5.5B users of Internet (2024, statista.com)
4. Mobile Web – huge growth in mobile devices, mobile internet access, location-based services
5. Web 3.0: Personal data may be tracked – Affects data privacy – typically termed as "**Cyberspace looks at you**"

Motivation


Powerful multi-core processors
General purpose graphic processors
Superior software methodologies
Virtualization leveraging the powerful hardware
Wider bandwidth for communication
Proliferation of devices
Explosion of domain applications

Technology Advances



What is Cloud Computing?

Cloud Computing is a general term used to describe a new class of network based computing that takes place over the Internet,

- basically a step on from Utility Computing
- a collection/group of integrated and networked hardware, software and Internet infrastructure (called a platform).
- Using the Internet for communication and transport provides hardware, software and networking services to clients

These platforms hide the complexity and details of the underlying infrastructure from users and applications by providing very simple graphical interface or API (Applications Programming Interface).

What is Cloud Computing cont….

- Self-service
- Commodity pricing
- Transparent scalability
- Shared infrastructure

Cloud Summary

- Shared pool of configurable computing resources
- On-demand network access
- Provisioned by the Service Provider

Cloud computing is the on-demand delivery of IT resources over the Internet with pay-as-you-go pricing. Instead of buying, owning, and maintaining physical data centers and servers, you can access technology services, such as computing power, storage, and databases, on an as-needed basis from a cloud provider like Amazon Web Services (AWS). – **AMAZON**

Cloud computing is on-demand access, via the internet, to computing resources—applications, servers (physical servers and virtual servers), data storage, development tools, networking capabilities, and more—hosted at a remote data center managed by a cloud services provider (or CSP). The CSP makes these resources available for a monthly subscription fee or bills them according to usage. – **IBM**

Simply put, cloud computing is the delivery of computing services—including servers, storage, databases, networking, software, analytics, and intelligence—over the Internet (“the cloud”) to offer faster innovation, flexible resources, and economies of scale. You typically pay only for cloud services you use, helping you lower your operating costs, run your infrastructure more efficiently, and scale as your business needs change. – **MICROSOFT**

Cloud computing is the act of running workloads within clouds—which are IT environments that abstract, pool, and share scalable resources across a network. Neither cloud computing nor clouds are technologies unto themselves.

Cloud computing is an act—the function of running a workload in a cloud.

Clouds are environments—places where applications run.

Technologies are things—software and hardware used to build and use clouds. **- REDHAT**

Cloud Computing: Definition

The US National Institute of Standards (NIST) defines cloud computing as follows:

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

3-4-5 rule of Cloud Computing

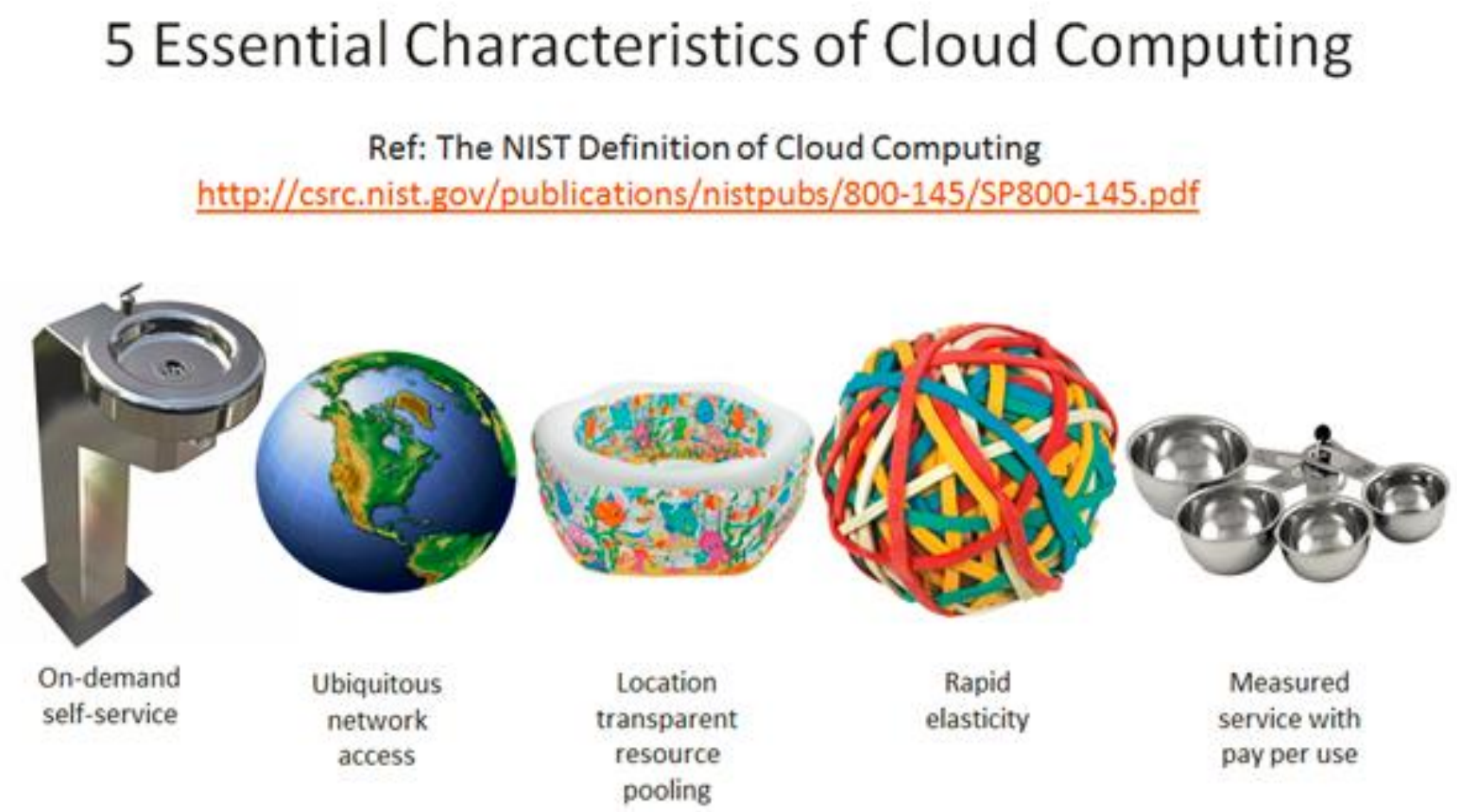
NIST specifies 3-4-5 rule of Cloud Computing

3 Cloud service models or service types for any cloud platform

4 Deployment models

5 Essential characteristics of cloud computing infrastructure

Characteristics of Cloud Computing



- On demand self-service
- Broad network access
- Resource pooling
- Rapid elasticity
- Measured service

1. On Demand Self Service

- No Manual intervention by service provider
  - Compute, storage, platform resources **self/auto provisioned**
  - E.g: Register to aws.amazon.com, login, select the pay scheme and start using

Image Courtesy: 5 Key Cloud Computing Characteristics Explained (cloudwards.net)

2. Broad Network Access

- Service available over **network connectivity**
  - Public cloud : Internet connection
  - Private Cloud: Local Network
- **Physical location independence**
- Allowing for maximum flexibility for the user for choosing middleware clients and software.

Image Courtesy: 5 Key Cloud Computing Characteristics Explained (cloudwards.net)

3&4: Resource Pooling & Rapid Elasticity

3. Resource Pooling

- Huge number of **concurrent** users/clients supported
- Cloud services provided using **shared pool of resources** among users.

4. Rapid Elasticity

- Rapidly increase /decrease resources as needed

- Specify minimum and maximum
- Dynamically change – based on the load at any time
- Faster time to provision & deploy

All Image Courtesy: 5 Key Cloud Computing Characteristics Explained (cloudwards.net)

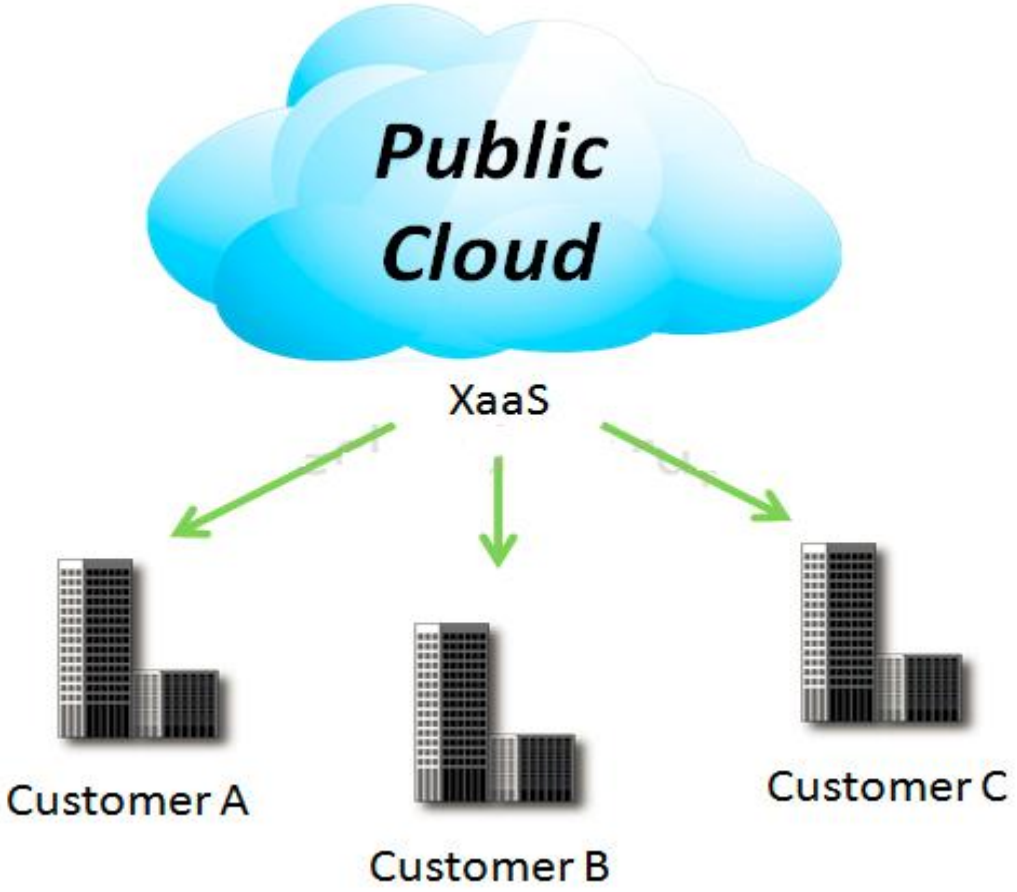
5. Measured Service

- Pay as you go model
- Pay only for resources used by user's own application/software depending on the service model leveraged

Image Courtesy: 5 Key Cloud Computing Characteristics Explained (cloudwards.net)

4 Deployment Models

1. Public Cloud



Mega-scale cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services.

Public Cloud

Key characteristics:

- Scalability
- Cost effectiveness
- Reliability
- Flexibility
- Location Independence
- On demand computing
- Pay per use pricing
- Broad network access
- Resource pooling

Examples:

- AWS
- Microsoft Azure
- Google Cloud Platform (GCP)
- Alibaba

Public Cloud

Advantage of public cloud:

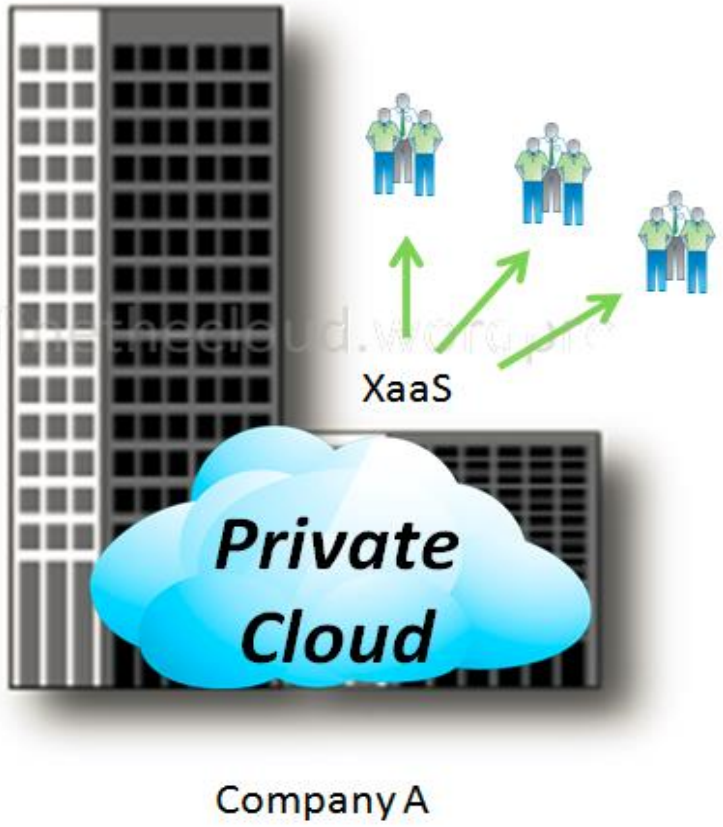
- It helps organizations to have less investment and maintenance costs
- User demands can be easily met with scalability
- Less resource wastage
- High reliability

Disadvantages of public cloud:

- Limited control on the infrastructure configurations
- Security: There may be data privacy issues
- Compliance: Unsure about the compliance of the security rules related to data storage
- Vendor Lock-In: Not easy to migrate away due to tailored software and operating procedures

4 Deployment Models

2. Private Cloud



The cloud infrastructure is operated solely for an organization.
It may be managed by the organization or a third party and may exist on premise or off premise.

Private Cloud

Key characteristics/Benefits of Private Cloud:

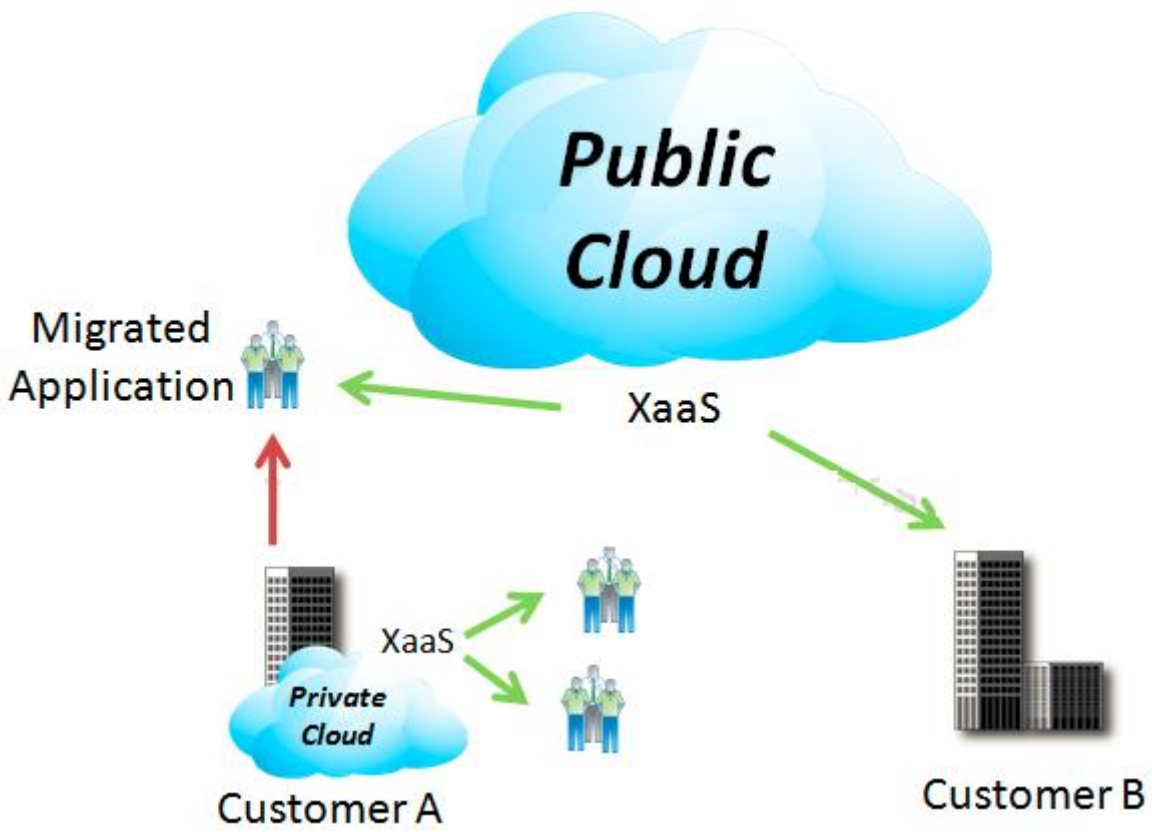
- Predictable server usage
- Improved resource utilization
- Reduced costs
- Increased security
- Regulatory compliance
- Customization

Drawback: Cost and accountability of managing the private cloud - Company's IT department

Need the same staffing, management, and maintenance expenses as traditional datacenter ownership

4 Deployment Models

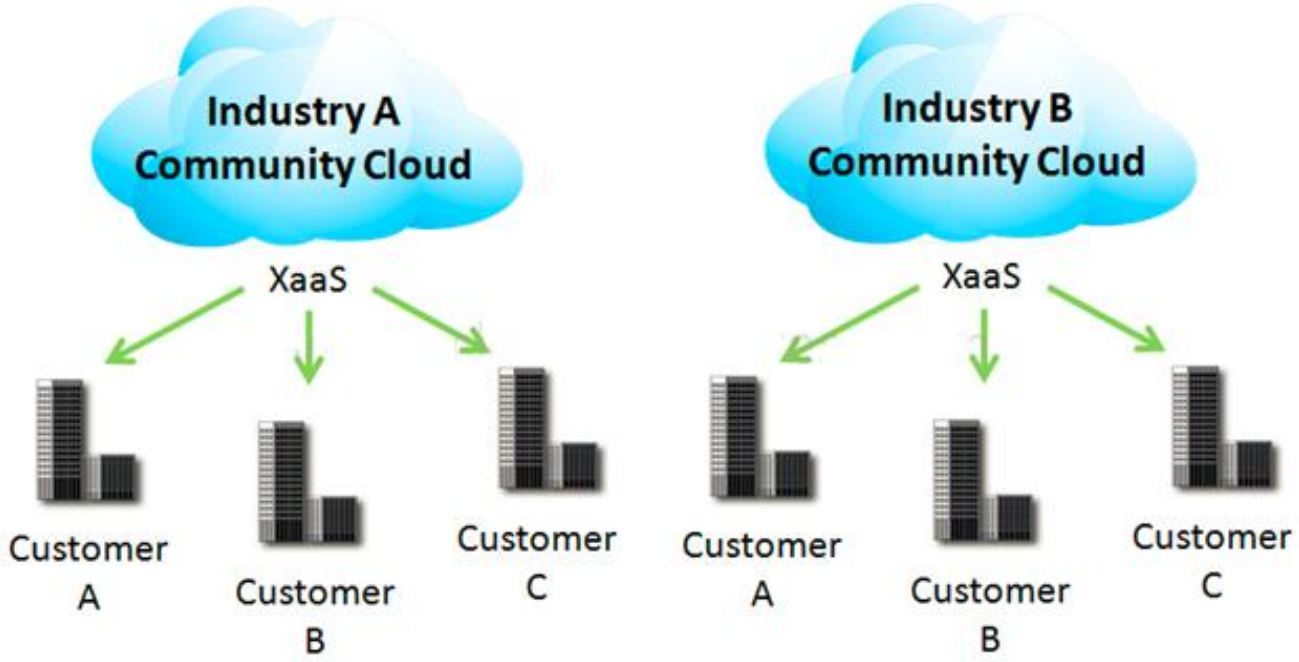
3. Hybrid Cloud



The cloud infrastructure is a composition of two or more clouds (private or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability

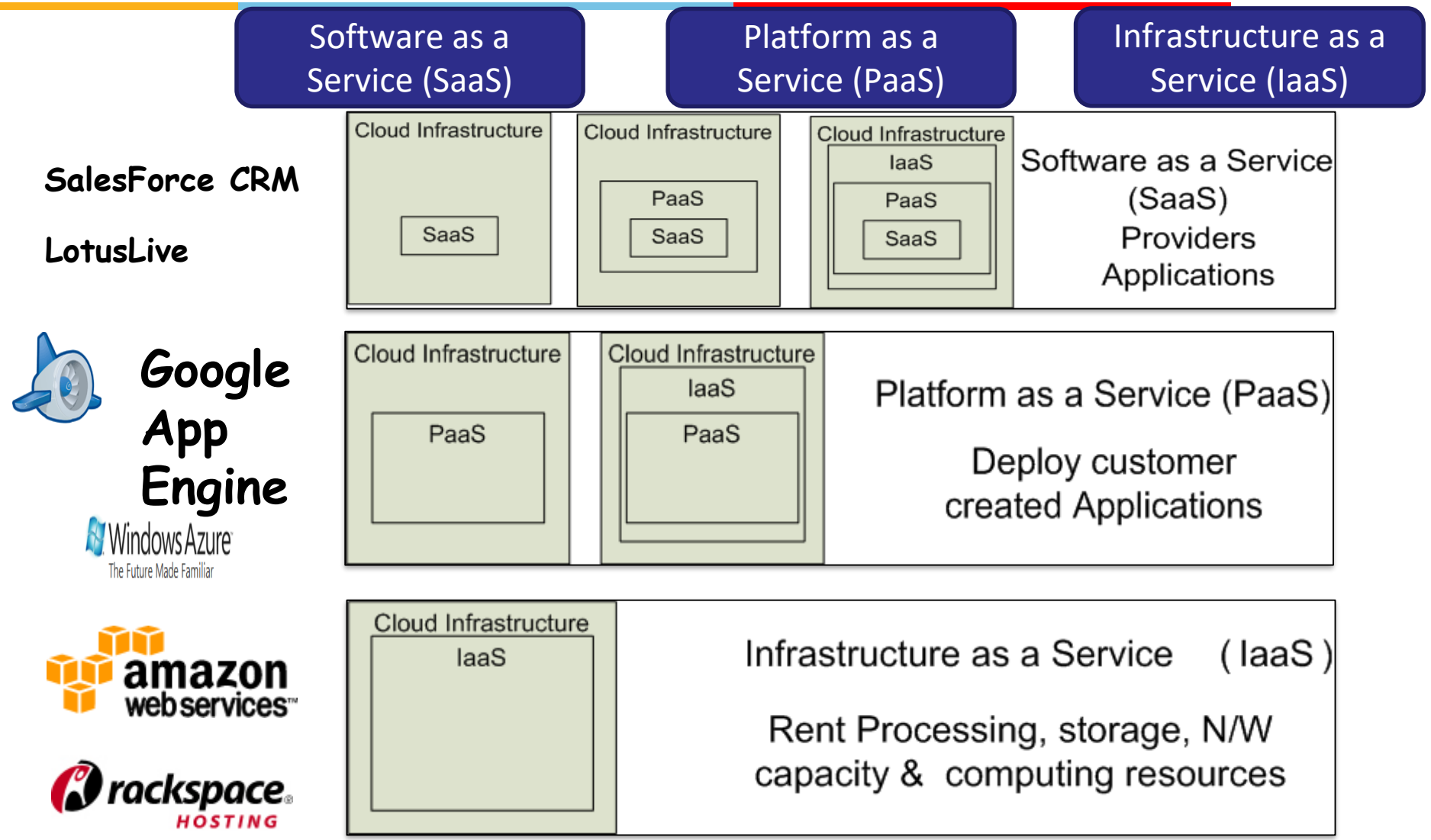
4 Deployment Models

4. Community Cloud



- Community Clouds are when an infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations).
- It may be managed by the organizations or a third party and may exist on premise or off premise' according to NIST.

3 Cloud Service Models



Infrastructure as a Service

Infrastructure as a service delivers:

- Basic storage and computing capabilities as standardized services over the network.
- Servers, storage systems, switches, routers , and other systems are pooled and made available to handle workloads
- Workloads include application components to high-performance computing applications.
- E.g: Renting VMs to put a web server

Platform as a Service

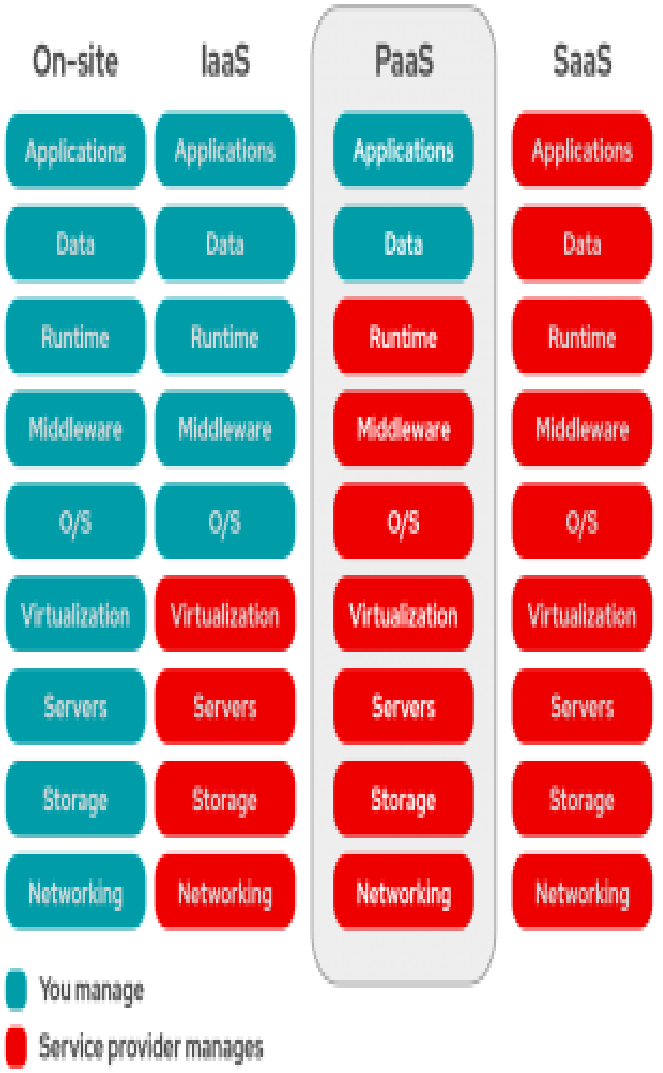
Platform as a service encapsulates a layer of software and provides it as a service that can be used to build higher-level services.

2 Perspectives for PaaS :-

Producer:- Someone producing PaaS might produce a platform by integrating an OS, middleware, application software, and even a development environment that is then provided to a customer as a service.

Consumer:-Someone using PaaS would see an encapsulated service that is presented to them through an API. The customer interacts with the platform through the API, and the platform does what is necessary to manage and scale itself to provide a given level of service.

E.g: Deploy and run a web application using Google app engine



Software as a Service (SaaS)

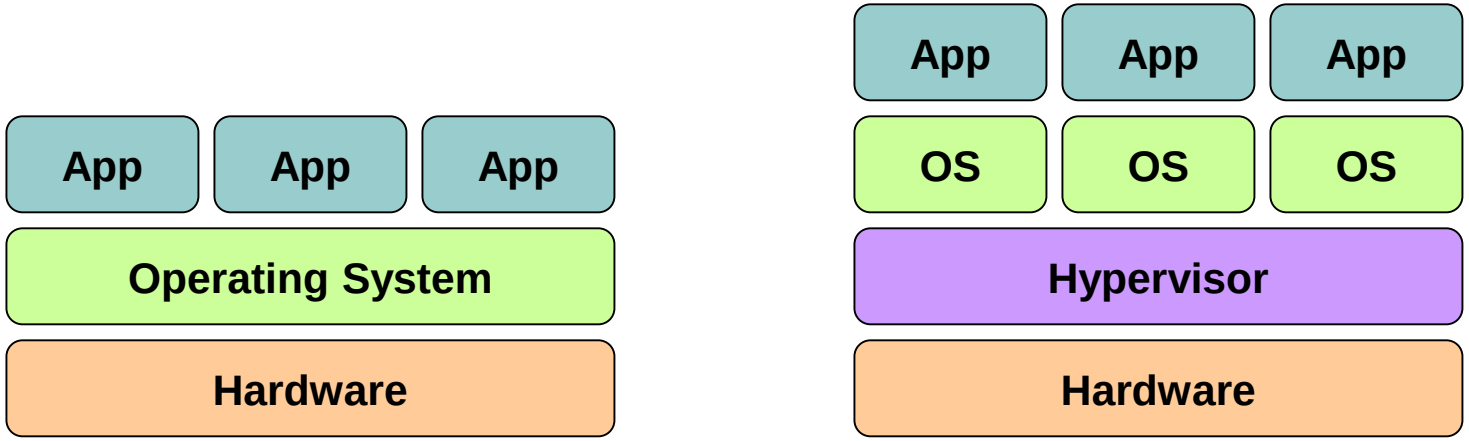
Software as a service features a complete application offered as a service on demand.

- A single instance of the software runs on the cloud and services multiple end users or client organizations.

E.g. salesforce.com , Google Apps,
Offload mail server online to Gmail/Outlook

Cloud Infrastructure

Key Technology is Virtualization



Virtualization plays an important role as an enabling technology for datacentre implementation by abstracting compute, network, and storage service platforms from the underlying physical hardware

Characteristics Of Cloud Providers

- Provide on-demand provisioning of computational resources
- Use virtualization technologies to lease these resources
- Provide public and simple remote interfaces to manage those resources
- Use a pay-as-you-go cost model, typically charging by the hour
- Operate data centers large enough to provide a seemingly unlimited amount of resources to their clients

Management of Virtualized Resources

- Distributed Management of Virtual Machines
- Reservation-Based Provisioning of Virtualized Resources
- Provisioning to Meet SLA Commitments

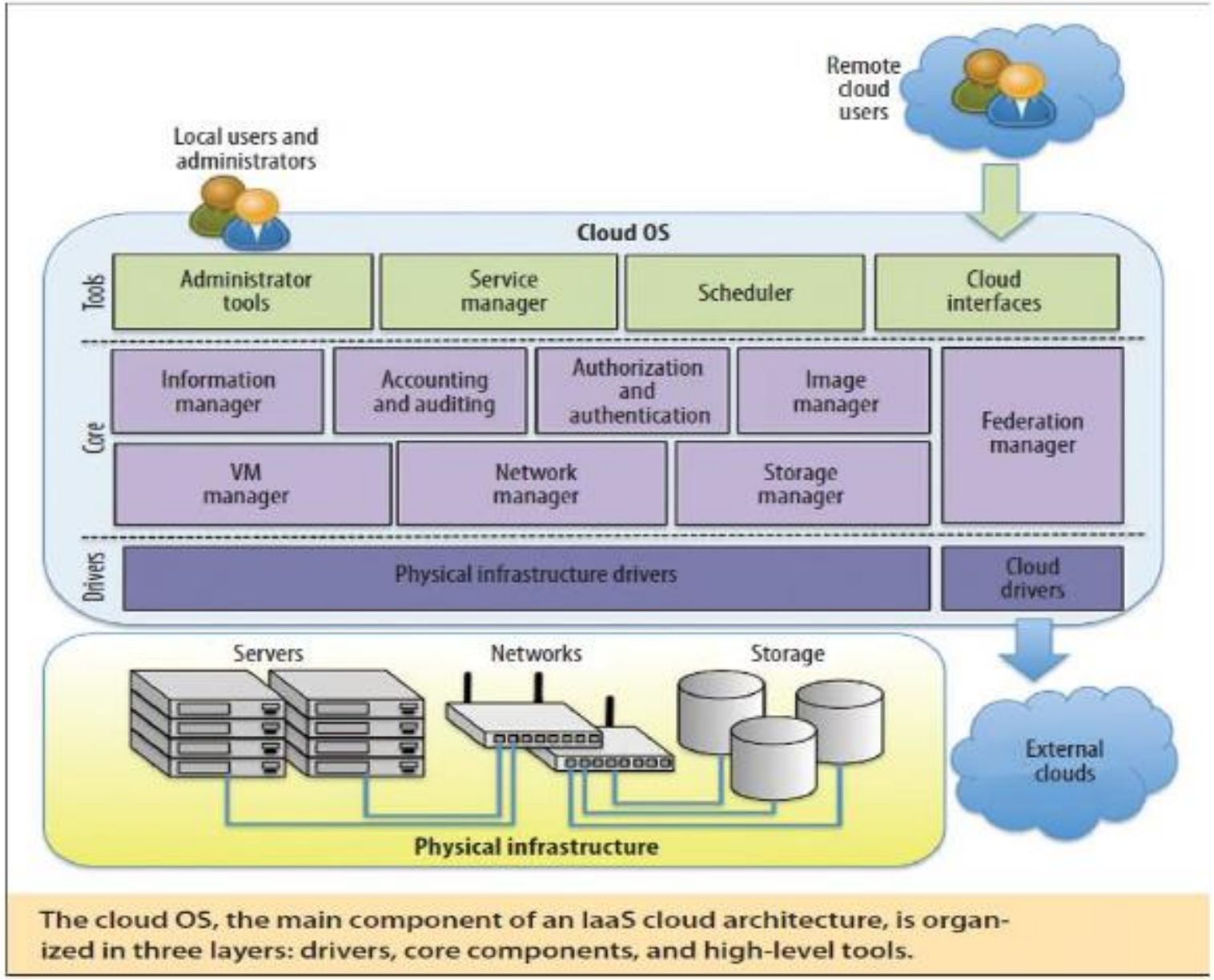
Cloud Infrastructure Anatomy

The key component of an IaaS cloud architecture is the cloud OS, which manages the physical and virtual infrastructures and controls the provisioning of virtual resources according to the needs of the user services

A cloud OS's role is to efficiently manage datacenter resources to deliver a flexible, secure, and isolated multitenant execution environment for user services that abstracts the underlying physical infrastructure and offers different interfaces and APIs for interacting with the cloud

While local users and administrators can interact with the cloud using local interfaces and administrative tools that offer rich functionality for managing, controlling, and monitoring the virtual and physical infrastructure, remote cloud users employ public cloud interfaces that usually provide more limited functionality

The Cloud OS



The cloud OS, the main component of an IaaS cloud architecture, is organized in three layers: drivers, core components, and high-level tools.

The cloud operating system is responsible for:

1. managing the physical and virtual infrastructure,
2. orchestrating and commanding service provisioning and deployment
3. providing federation capabilities for accessing and deploying virtual resources in remote cloud infrastructures

Value of Cloud

Value Delivered	From Traditional	From Cloud
Design and Release Application	Months	Weeks/Days
Test Provisioning	Weeks	20 Minutes
Change Management	Months	Days or Hours
Install Database	1 Day	12 Minutes
Install Operating System	1 Day	30 Seconds
Service Provisioning	Weeks/Days	Hours/Minutes

Summary

- Introduction to Cloud Computing
- Evolution of Cloud Computing
- 3-4-5 Rule
  - 5 Characteristics
  - 4 Deployment Models
  - 3 Service Models