Cloud Computing (CS60118)

(Spring 2020-2021)

Introduction

Dr. Sudip Misra
Professor
Department of Computer Science and Engineering
Indian Institute of Technology Kharagpur
Email: smisra@sit.iitkgp.ernet.in
Website: http://cse.iitkgp.ac.in/~smisra/
Research Lab: cse.iitkgp.ac.in/~smisra/swan/

Outline

- Introduction to Cloud computing
- Cloud computing architecture and its components
- Service and data management in the cloud
- Federation, Presence, Identity, and Privacy in the Cloud
- Case Studies
- Future directions: Where are we heading?
- Demos

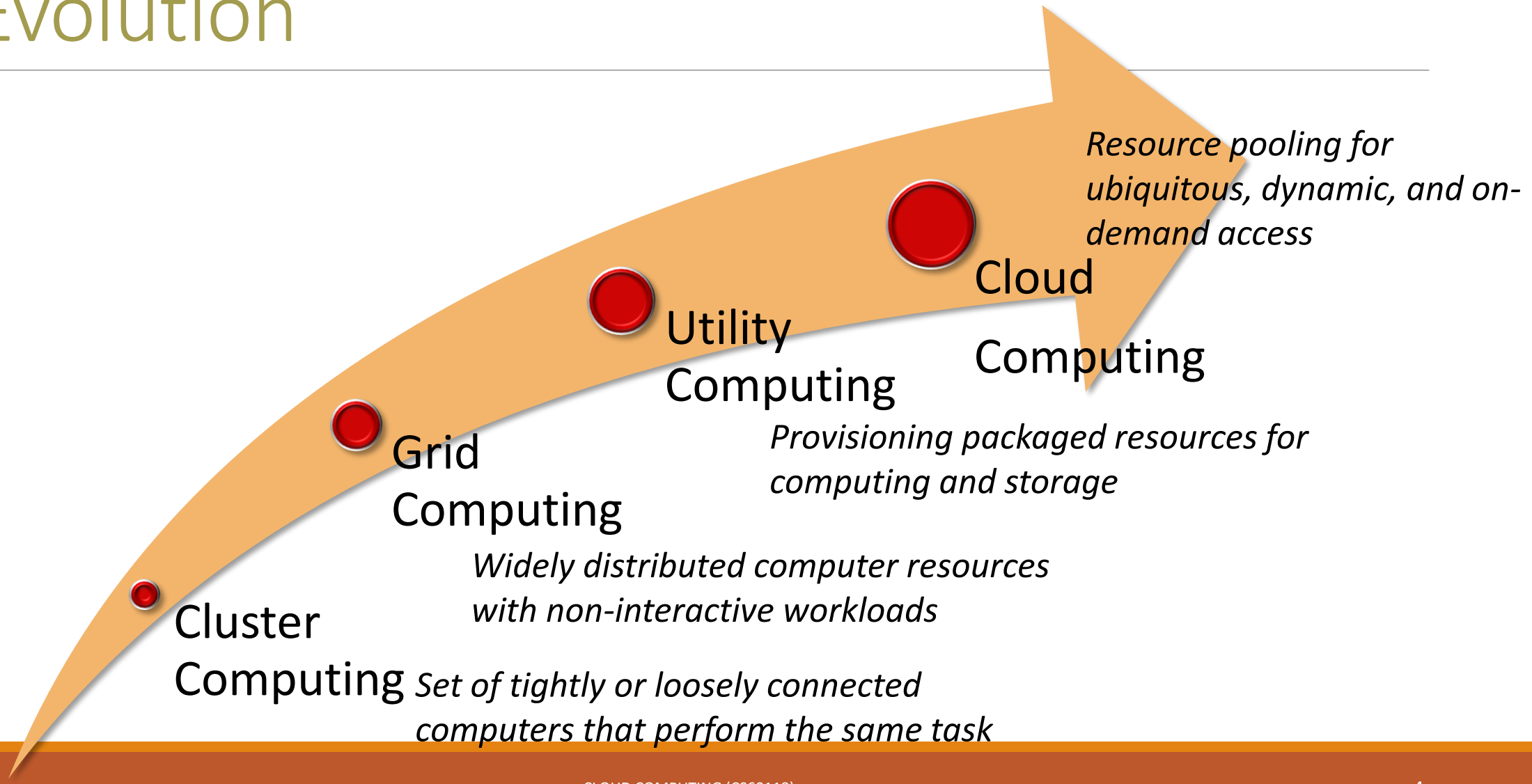
What is Cloud Computing?

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. This cloud model is composed of five essential characteristics, three service models, and four deployment models.

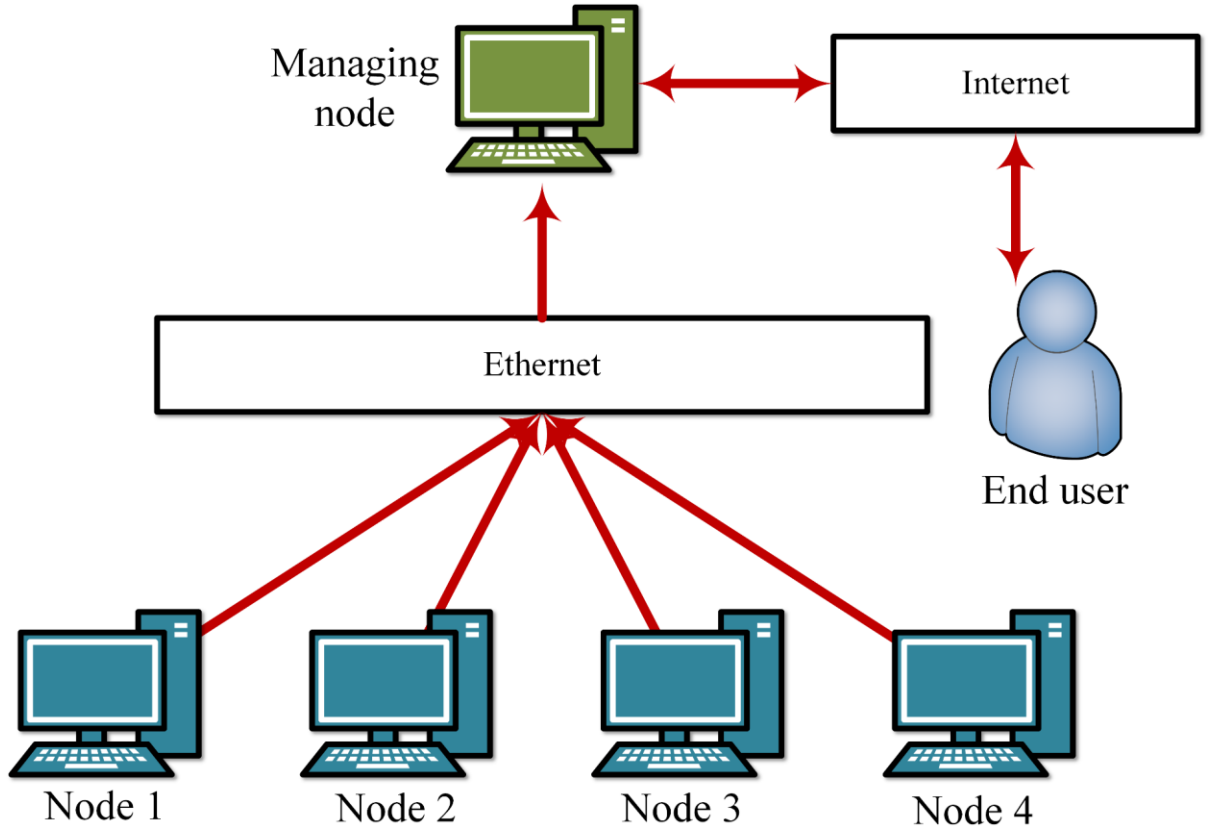
-NIST

Source: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf

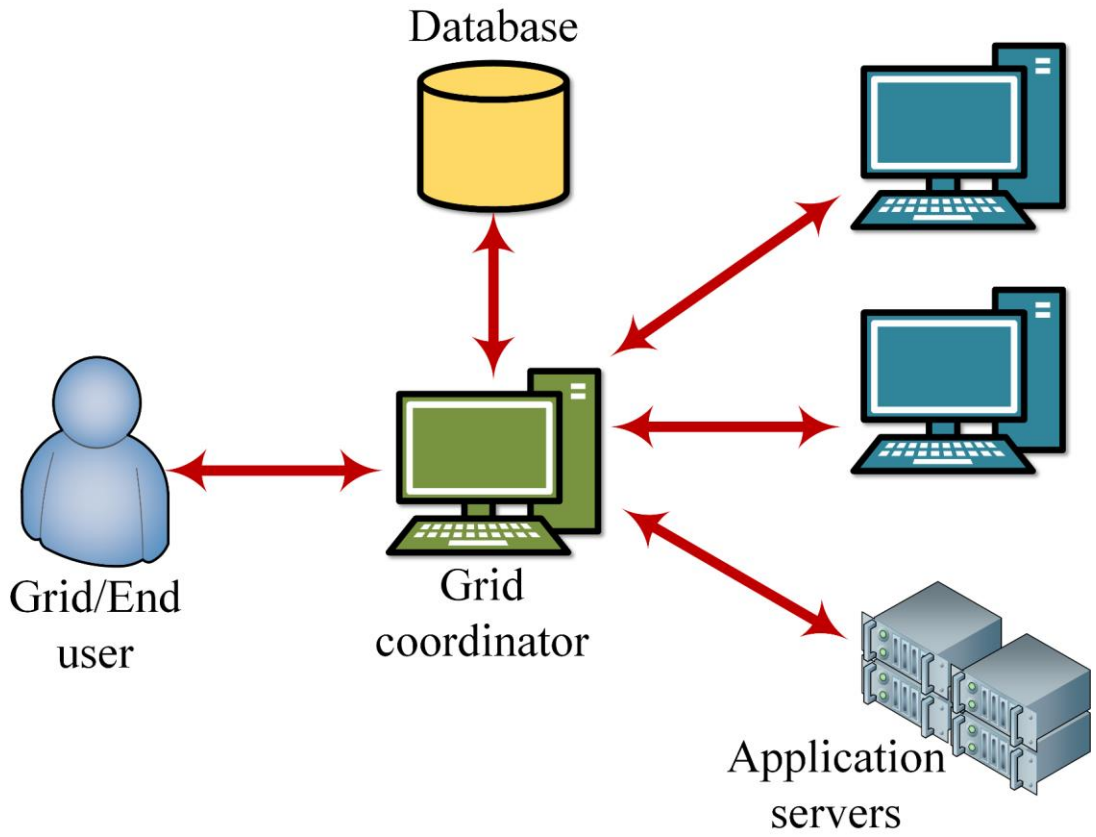
Evolution


Cluster Computing
Set of tightly or loosely connected computers that perform the same task
Grid Computing
Widely distributed computer resources with non-interactive workloads
Utility Computing
Provisioning packaged resources for computing and storage
Cloud Computing
Resource pooling for ubiquitous, dynamic, and on-demand access

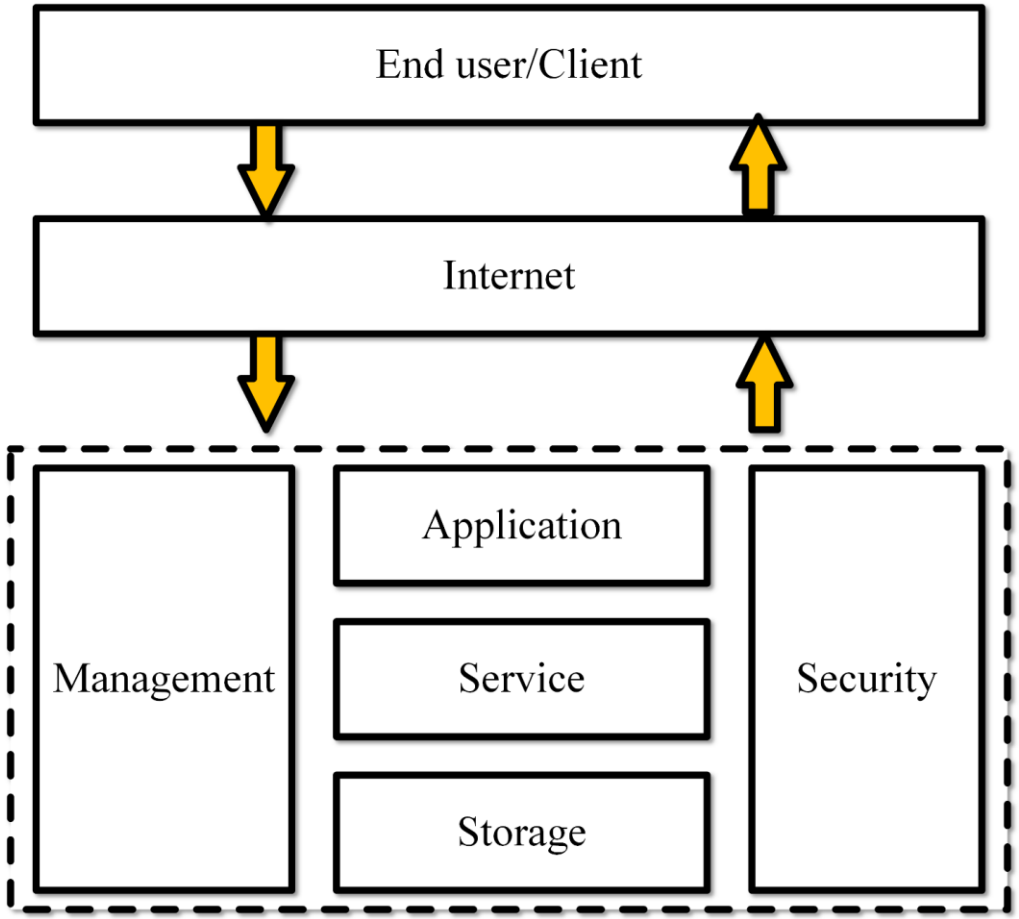
Cluster Computing



Grid Computing



Cloud Computing



Participants

Consumers

- End-users that use the cloud services for personal/business activities.

Service providers

- Cloud providers offering a variety of functions such as infrastructure, applications, tools, and others.

Designers

- Applications and tools builders in the cloud.

General Characteristics

On-demand self-service
- Consumers can provision resources without interacting with service providers.

Broad network access
- Independent (with respect to thin and thick clients) provisioning of resources.

Resource pooling
- grouping together resources for maximizing advantage and minimizing risk to the users.

Rapid elasticity
- Real-time scaling of resources commensurate with demand.

Measured service
- Metering the resource usage for billing.

Source: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf

Essential Characteristics

Broad network access

- Cloud resources should be available over the network
- Should support standard mechanisms for information retrieval using traditional interfaces
- Supported clients: Heterogeneous Thin and Thick clients

Thin Client

Features

- Stateless
- Fanless desktop terminal
- No hard drive

Example: Virtual desktops

Highly dependent on central server (due to lack of localized hard drive)

Reliable connections are important

Advantages:

- Reduced cost
- Increased security
- Scalable and easily manageable

Thick Client

Opposite of thin clients

Full-featured computers connected to a network

They are functional whether connected to the network or not

Servers provide program/codes which are not present in the memory

Example: Personal computers, laptops, etc

Disadvantages in comparison to thin clients:

- Difficult to secure and manage
- Costlier to deploy
- High energy consumption

Essential Characteristics Contd.

Rapid Elasticity

- Rapid, elastic, and automated allocation of cloud resources
- Dynamic allocation/release for scale-out and scale-in
- Customers should feel infinite resources

Measured service

- Resource usage should be recorded and monitored
- Facility to dynamically control and optimize the resource usage
- Transparency between the service provider and consumer

Essential Characteristics Contd.

On-demand self-service

- Automatic provision of server time and network storage
- Self-service

Resource pooling

- Automated pooling of all available resources
- Serve multiple end users using a multi-tenant model
- Allocation of resources according to user demand

Difference between Cloud and Cluster Computing

Cloud Computing	Cluster Computing
Heterogeneous resources	Homogeneous resources
Support for virtualization	No support for virtualization
Very low capex	Very high capex
Low security requirement	High security requirement
Low maintenance	Relatively high maintenance
Support for multiple OS simultaneously	Support for singular OS
Centralized and decentralized management	Centralized management
High scalability	Low scalability
Dynamic resource allotment	Tightly coupled resources
Application domain independent software provisioning	Application domain dependent software provisioning

Difference between Cloud and Grid Computing

Cloud Computing	Grid Computing
Client-server architecture	Distributed architecture
Centralized resource usage	Distributed resource usage
Highly flexible	Flexibility is low
Payment according to usage	Does not support pay-per-use
High accessibility	Low accessibility
Support for multiple OS simultaneously	Support for singular OS
Centralized and decentralized management	Centralized management
High scalability	Low scalability
Middleware independent	Requires grid computing middleware

Mainframe Computing

First came into existence in 1951

Highly powerful and reliable computing machines

Massive input-output operations

High fault tolerance

Cluster computing is a replacement for mainframe computing

Components of the Cloud

Clients/end-users: Thin and thick clients (mobile and stationary)

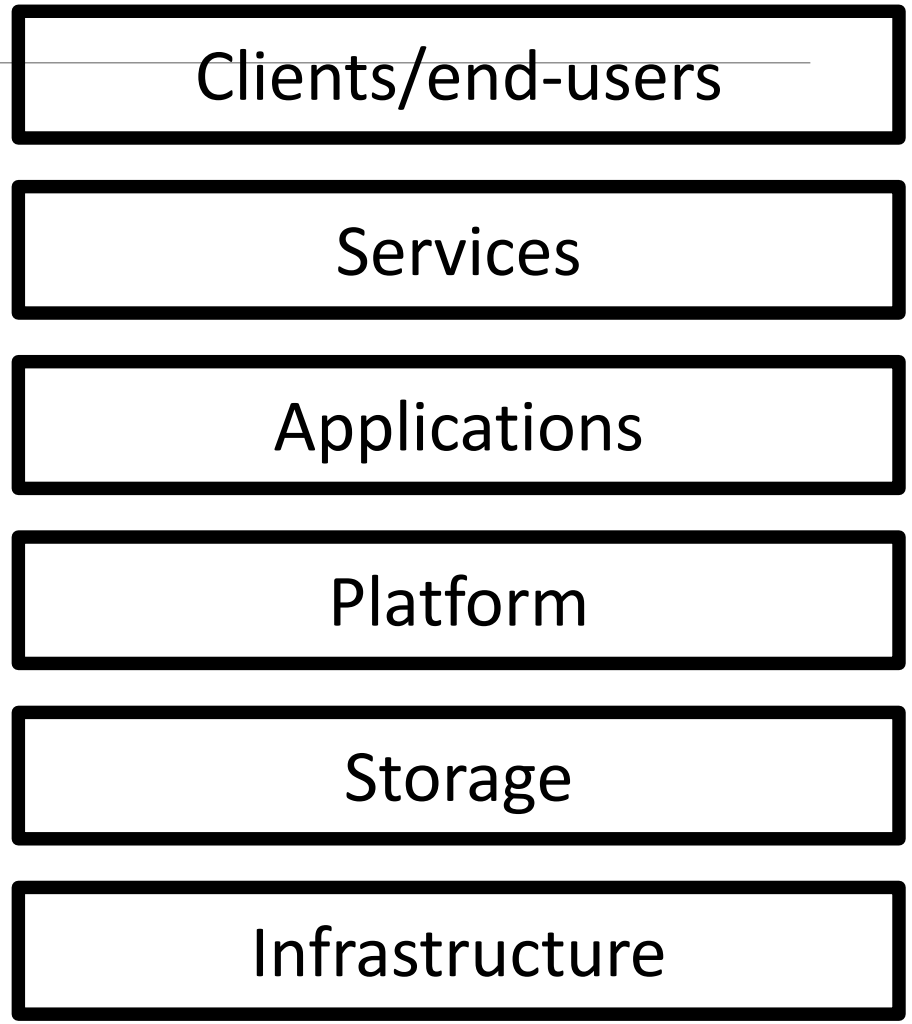
Services: Products and solutions

Applications: Web Apps, SaaS, etc

Platform: Apps/Web hosting using PaaS

Storage: Database, Data Storage-as-a-Service (DSaaS)

Infrastructure: Virtualization, IaaS, EC2



Everyone Uses Cloud

For instance, ***Google's Gmail service***, we are all using a cloud email service.

Anything that can be served as a service-

- Compute power to computing infrastructure
- Applications and business processes

Only two constraints:

- Common standardization
- Automation.

Standardization

Consistent service delivery using consistent interfaces

Necessary both before and after deployment

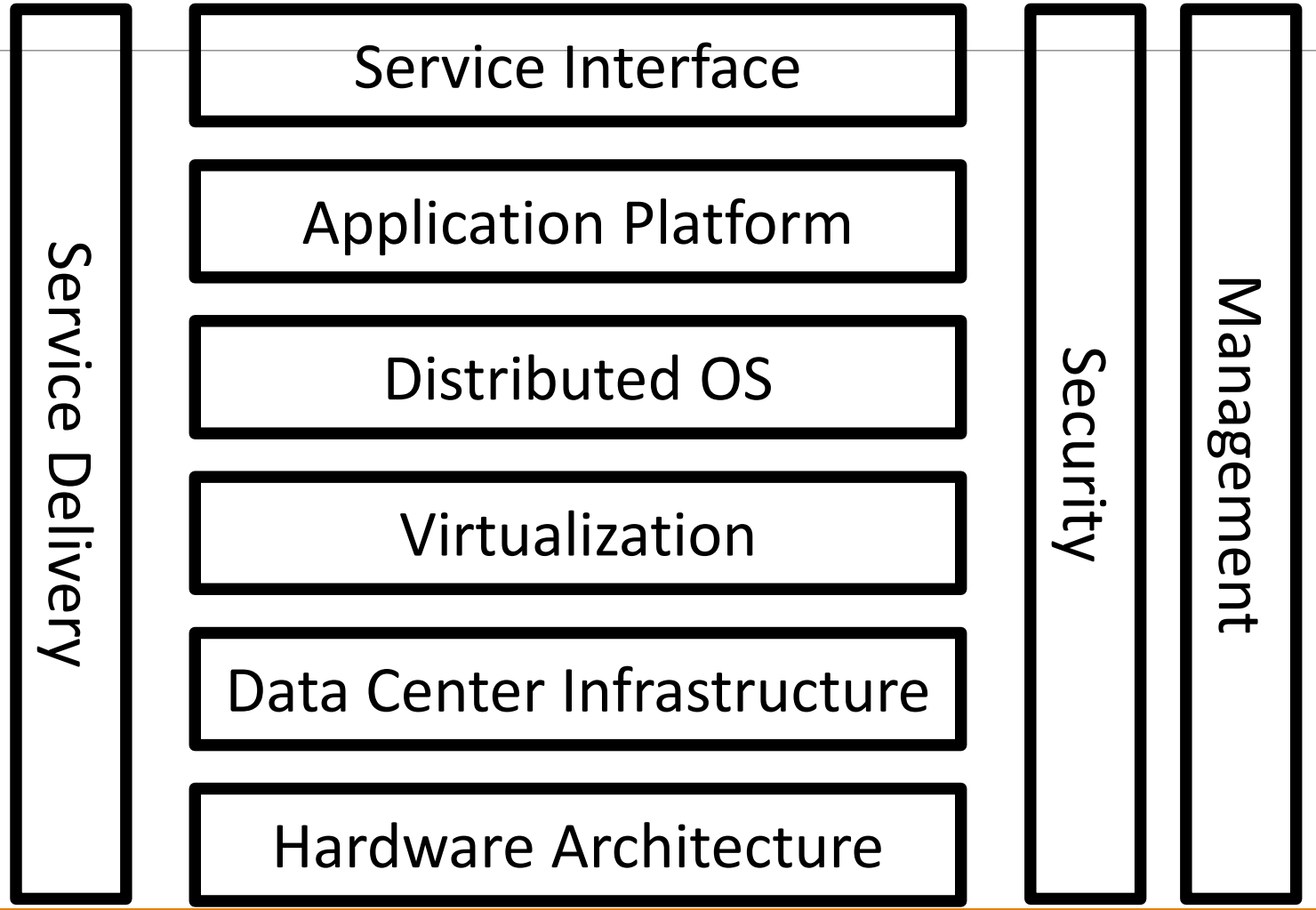
Main areas for standardization:

- Service calling
- Service interfaces
- Resource and network management

Typically, 6 layers need standardization

Motive: Scalability, interoperability, and security in the cloud

Standardization Layers



Source: Cloud Computing Standardization and Research, ZTE Technologies

Automation

Self-service provisioning model

Reduce manual intervention for:

- Provisioning
- Configuring
- Managing cloud environments

For instance, automation ensures returning the resources to the resource pool after a provisioned service is exhausted

Facilitates capacity planning and overall workload management

Advantages of Automation

Improved security and resilience

- Reduction of human errors and malicious insiders

Improved backup processes

- Improves an organization's resilience to disaster

Improved governance

- Improved control over infrastructure

Note: Automation and orchestration are different concepts. ***Orchestration*** *deals with the scheduling and integration of automated tasks across heterogeneous systems. It is the step after achieving automation.*

Deployment Models

Public cloud
- Resources owned and operated by third parties.
- Allow usage other companies/customers.

Private cloud
- Resources owned and operated by the company/organization.
- Used behind a firewall.

Hybrid
- Combination of private and public clouds.

Multi-cloud
- Organization using two or more public clouds.

Public Cloud

Resources owned and operated by third parties

Usage by organizations/customers

Customers unaware od underlying infrastructure

Third parties offer rich set of services in addition to the resources

- Security
- Specialized infrastructures such as GPU

On-demand delivery of services and resources

Support multi-tenancy

Multi-Tenancy

Multiple applications operate in a shared environment

Logical isolation

A single instance of a software application serves multiple tenants

Tenants = Customers

Tenants have the option of customizing the user interface and operating rules (cannot manipulate application code)

Applicable for both private and public clouds

Improves scalability

Cheaper than single-tenancy

Disadvantages of Multi-Tenancy

Less flexible

Complex architecture compared to single-tenancy

Strict authentication rules (high physical integration among tenants)

May lead to slow response time (if some tenant consumes all resources)

Note: *Some public cloud services offer single-tenancy at higher costs. However, networking services may need to be shared.*

Private Cloud

Resources owned and operated by an organization

Explicit use by its employees, partners, and customers

Can be owned and operated by third parties too

Private cloud hosted behind firewalls

Automation higher than public clouds

Public cloud vendors often install data centers in the organization’s premises

Third parties may own the on premise resources and bill clients

Public Cloud vs Private Cloud

	Public Cloud	Private Cloud
Virtualized Resources	Publicly shared	Privately Shared
Customer Types	Multiple	Limited
Connectivity	Over Internet	Over Internet/private network
Security	Low	High

Hybrid Cloud

Combination of private and public cloud models

Collaborative use of the cloud models to achieve common goals

Combine services

- Computing environment with the following features:
  - Unified
  - Automated
  - Well-managed
  - Transparent to end users

Superior orchestration methods required for operations and deployment

Multicloud

Two or more public clouds are involved

Multicloud may or may not involve private clouds

- Hybrid cloud must have both private and public cloud models

Multicloud helps in better abstraction of activities and services

Corporate Computing: *Combination of multiple public services with private clouds and data centers*

Note: *Not all combinations of public and private clouds mean hybrid cloud computing.*

Common Misconceptions

The following does not mean hybrid computing:

- Architectures with public cloud service disconnected with the private cloud and data center
- No movement of data from a company's deployed software to its data center
- Each division/activity is run on mutually exclusive public clouds

The following qualifies as hybrid computing:

- The public deployments sends data to a private cloud or data center
- Each division/activity is run on public clouds and can collaborate among each other
- An organization can move workloads from one public cloud to another on requirement

Other Types of Cloud

Community Cloud

- Shared set-up between several organizations having common concerns (security, compliance, jurisdiction, etc)
- Managed internally or by third party

Distributed Cloud

- Collection of scattered set of computing devices in different locations, however, connected to a single network
- Two types: Public-resource computing and Volunteer cloud

Other Types of Cloud Contd.

Multi-cloud

- Multiple cloud computing services offered via single heterogeneous architecture
- Increases fault-tolerance and flexibility

Inter-cloud

- Unified global ‘cloud of clouds’ based on the Internet
- Supports interoperability between cloud service providers

Comparison of the Deployment Models

	On-premise	Off-premise
Dedicated Access	Private cloud	Hosted private cloud
Shared Access	Community cloud	Public cloud

Business Advantages

Nearly zero cost for upfront infrastructure investment

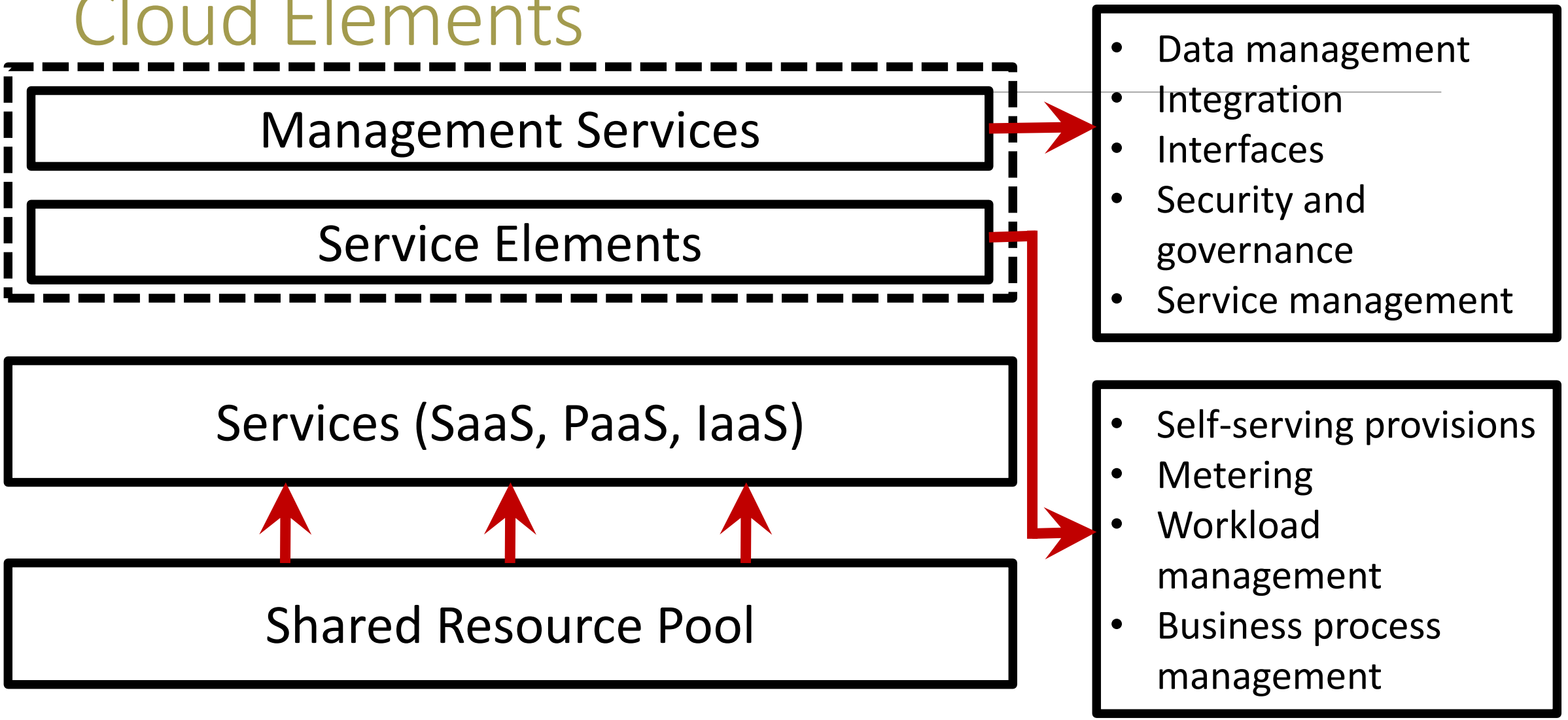
Real-time infrastructure availability

More efficient resource utilization

Usage-based costing

Reduced time to market

Cloud Elements



Service Models

Software-as-a-Service (SaaS)

Platform-as-a-Service (PaaS)

Infrastructure-as-a-Service (IaaS)

Software-as-a-Service (SaaS)

Facility to execute service provider's applications at user's end

Applications are available as 'services'

Services can be accessed via different types of client devices (e.g. web browser, app)

End-users do not posses the control of the cloud infrastructure

Examples: Google Apps, Salesforce, Learn.com.

Platform-as-a-Service (PaaS)

Facility for the consumer to execute consumer-created or acquired applications onto cloud infrastructure

Support for deployment of such applications

The user does not control the cloud infrastructure

User can control the deployed applications using given configurations

Examples: Windows Azure, Google App Engine

Infrastructure-as-a-Service (IaaS)

Facility to access computing resources such as network, storage, and operating system

User can deploy, execute and control any software (Operating systems and other applications)

In some case, the user can control selected networking components (e.g., host firewalls).

Examples: Amazon EC2, GoGrid, iland, Rackspace Cloud Servers.

Difference Between the Service Models

IaaS	Paas	SaaS
Resources/Infrastructure for storage and development	Platforms and tools to create, test, and deploy applications and software.	Web applications
Virtual machines, virtual storage, etc.	Integrated Development Environments (IDEs)	Developed Softwares
It is used by network architects.	It is used by developers.	It is used by end users.

Resource Life Cycle in the Cloud

Intuition: Customers/end users consume resources only when needed

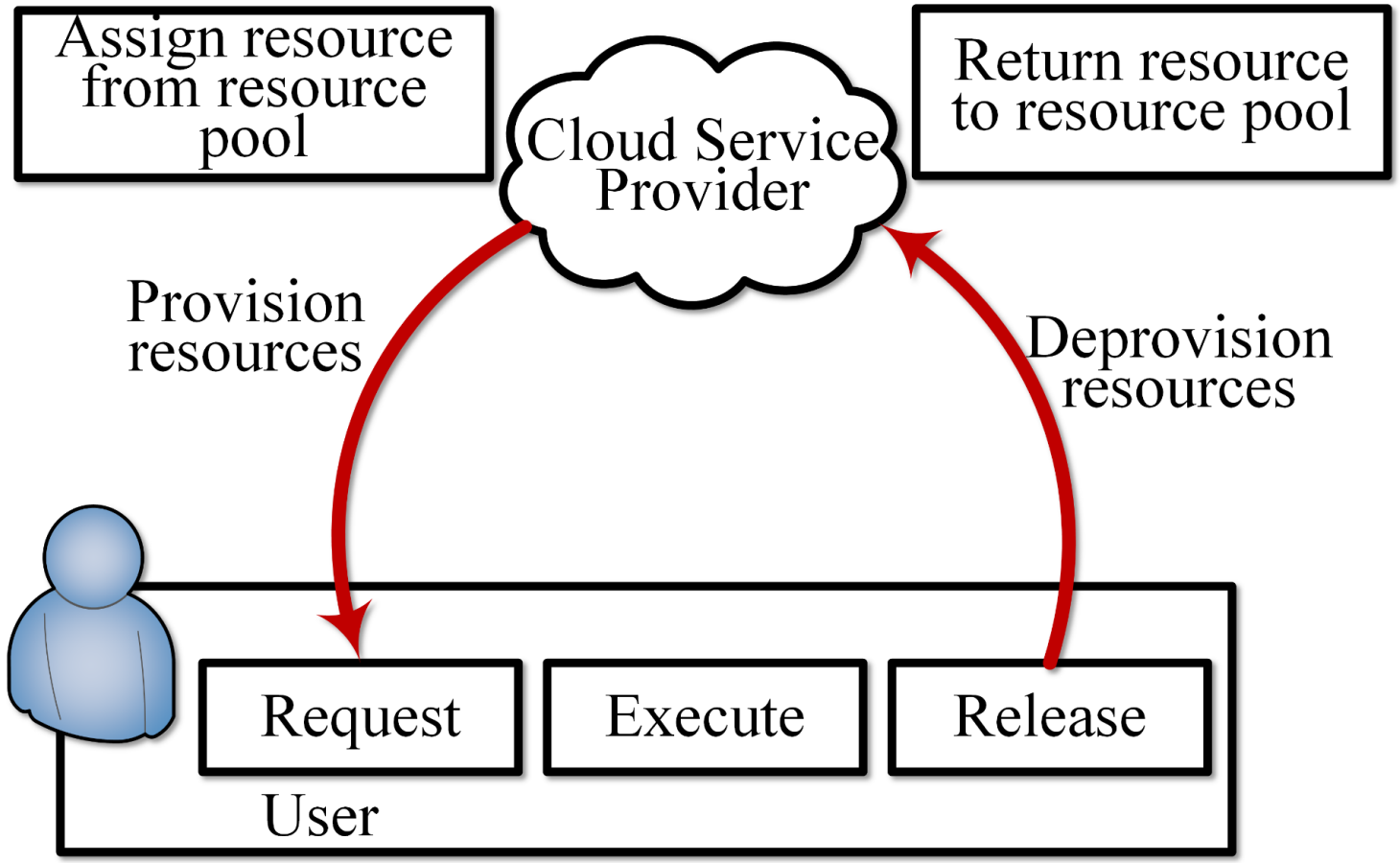
They are charged only when they are using the resources (measured)

Analogous to ***renting***

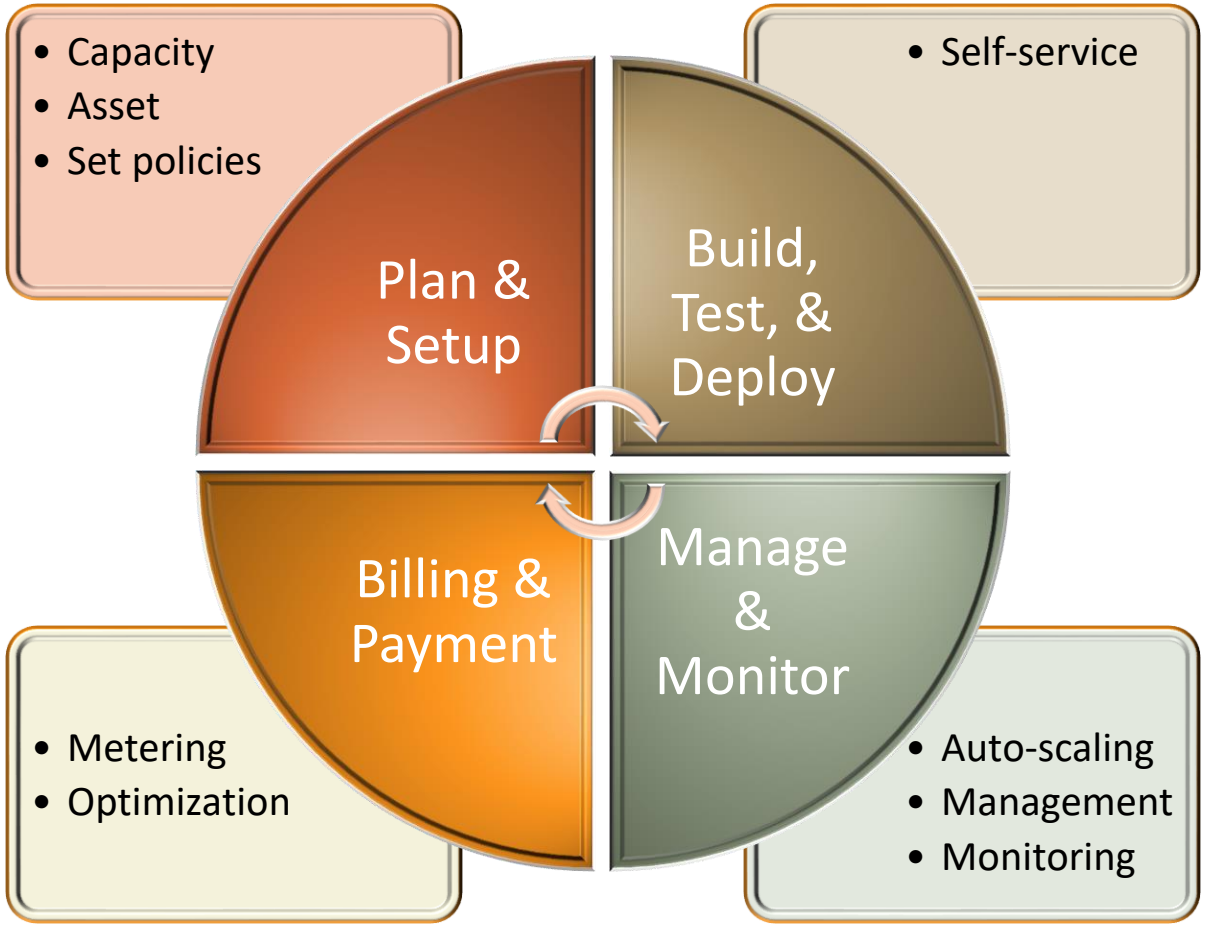
Cloud design considerations:

- Respond immediately on receiving customer requests
- Ensure resource availability at all times
- Return assigned resources back to the resource pool after usage by the customers

Resource Life Cycle Management



Cloud Life Cycle Management



Self-Service

Important feature of cloud computing

Website acts as an interface for the customers for:

- Select and purchase cloud services
- Modify service configurations
- Launch the service

Start using the purchased resource and service within seconds of deployment

No Self-Service in Traditional Data Centers

Customer needs to:

- File an application for the resources
- Concerned personnel verify and approve the application
- Concerned personnel install and launch the requested configurations
- No elasticity
- Handover of the deployed system to the customer
- Customer waits until approval (days/weeks)

Elasticity

Cloud resources change their original/provisioned configurations

Automatic increase and decrease in resource allotment

Example:

- Storage service
- Customer may start with low space
- Increase capacity on reaching limit
- Shrink capacity on deleting files
- Pay-per-use

Elasticity benefits both customers and service providers

Workload

Atomic service/program routine

Executable

Cloud is a collection of workloads

Organization between the workloads is important

Issues:

- Localizing workloads and data near each other
- Optimize performance
- Cloud strategy

Management Services

Maintain consistent quality of service

Irrespective of cloud deployment and service models

Resistance to outages and slow networks

Ensure overall security (to all participants)

Data management in hybrid and multicloud environments

THANK YOU