Cloud Computing

Introduction to Cloud Computing

Agenda

- What is Cloud Computing ?
  - Different perspectives
  - Properties and characteristics
  - Benefits from cloud computing
- Service and deployment models
  - Three service models
  - Four deployment models

Cloud Definitions

- Definition from ***NIST*** ***(National Institute of Standards and Technology)***
  - Cloud computing is a model for enabling 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 promotes availability and is composed of five essential characteristics, three service models, and four deployment models.

NIST

National Institute of Standards and Technology

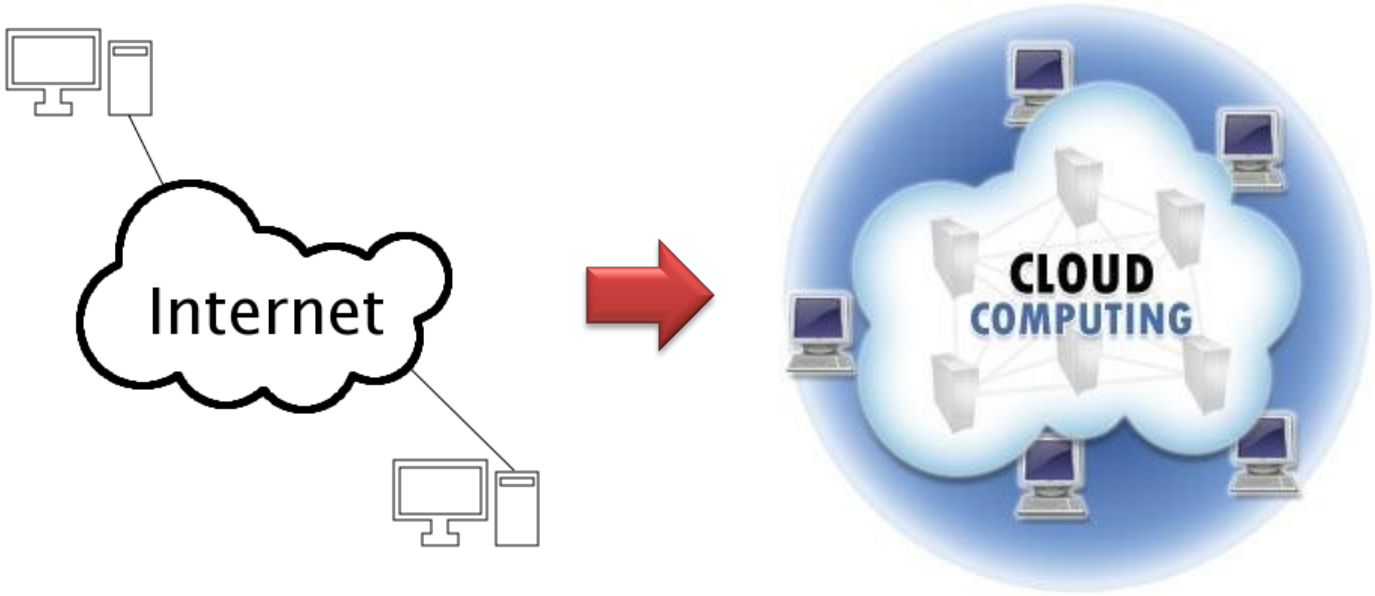
Technology Administration, U.S. Department of Commerce

Cloud Definitions

- Definition from ***Wikipedia***
  - Cloud computing is Internet-based computing, whereby shared resources, software, and information are provided to computers and other devices on demand, like the electricity grid.
  - Cloud computing is a style of computing in which dynamically scalable and often virtualized resources are provided as a service over the Internet.

Cloud Definitions

- Definition from ***Whatis.com***
  - The name cloud computing was inspired by the cloud symbol that's often used to represent the Internet in flowcharts and diagrams. Cloud computing is a general term for anything that involves delivering hosted services over the Internet.



Properties and characteristics

WHAT IS CLOUD COMPUTING ?

In Our Humble Opinion

- Cloud computing is a paradigm of computing, a new way of thinking about IT industry but not any specific technology.
  - Central ideas
    - ***Utility Computing***
    - ***SOA*** - Service Oriented Architecture
    - ***SLA*** - Service Level Agreement
  - Properties and characteristics
    - High ***scalability*** and ***elasticity***
    - High ***availability*** and ***reliability***
    - High ***manageability*** and ***interoperability***
    - High ***accessibility*** and ***portability***
    - High ***performance*** and ***optimization***
  - Enabling techniques
    - Hardware virtualization
    - Parallelized and distributed computing
    - Web service

Properties and Characteristics

Utility Computing
SOA + SLA

Central Ideas

- Perspective from user :
  - Users do not care about how the works are done
    - Instead, they only concern about what they can get
  - Users do not care about what the provider actually did
    - Instead, they only concern about their quality of service
  - Users do not want to own the physical infrastructure
    - Instead, they only want to pay as many as they used

- What dose user really care ?
  - They only care about their “Service”

Utility Computing
SOA + SLA

Utility Computing

- One service provisioning model
  - Service provider makes computing resources and infrastructure management available to the customer as needed, and charges them for specific usage rather than a flat rate.
  - Like other types of on-demand computing , the utility model seeks to maximize the efficient use of resources and/or minimize associated costs.

Utility Computing
SOA + SLA

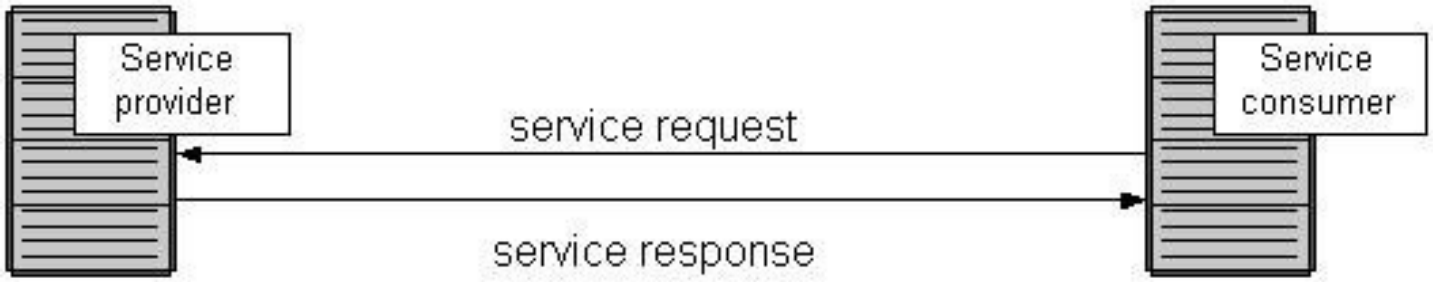
What Is Service?

- Service is what you connect together using Web Services.
- Service is the endpoint of a connection.
- Functionalities of service :
  - A service should be well-defined
  - A service should be self-contained
  - A service should not depend on the context or state of other services.

Utility Computing
SOA + SLA

Service Oriented Architecture

- Definition
  - Service Oriented Architecture (SOA) is essentially a collection of services which communicate with each other
  - Contain a flexible set of design principles used during the phases of systems development and integration
  - Provide a loosely-integrated suite of services that can be used within multiple business domains
- Approach
  - Usually implemented by Web Service model



Utility Computing
SOA + SLA

What Is Web Service?

- Definition :
  - Web service is self-describing and stateless modules that perform discrete units of work and are available over the network
  - Web service providers offer APIs that enable developers to exploit functionality over the Internet, rather than delivering full-blown applications

- Web Services Description Language (WSDL) :
  - Expressed in XML which include both data type and messages
  - Four types of operations :
    - One-way - Messages sent without a reply required
    - Request & response - Sending and replying messages
    - Solicit response - A request for a response
    - Notification - Messages sent to multiple receivers

Utility Computing
SOA + SLA

Service Level Agreement

- Definition
  - A service-level agreement (SLA) is a contract between a network service provider and a customer that specifies, usually in measurable terms (QoS), what services the network service provider will furnish

- Common content in contract
  - Performance guarantee metrics
    - Up-time and down-time ratio
    - System throughput
    - Response time
  - Problem management detail
  - Penalties for non-performance
  - Documented security capabilities

Scalability & Elasticity

- What is scalability ?
  - A desirable property of a system, a network, or a process, which indicates its ability to either handle growing amounts of work in a graceful manner or to be readily enlarged.
- What is elasticity ?
  - The ability to apply a quantifiable methodology that allows for the basis of an adaptive introspection with in a real time infrastructure.
- But how to achieve these properties ?
  - Dynamic provisioning
  - Multi-tenant design

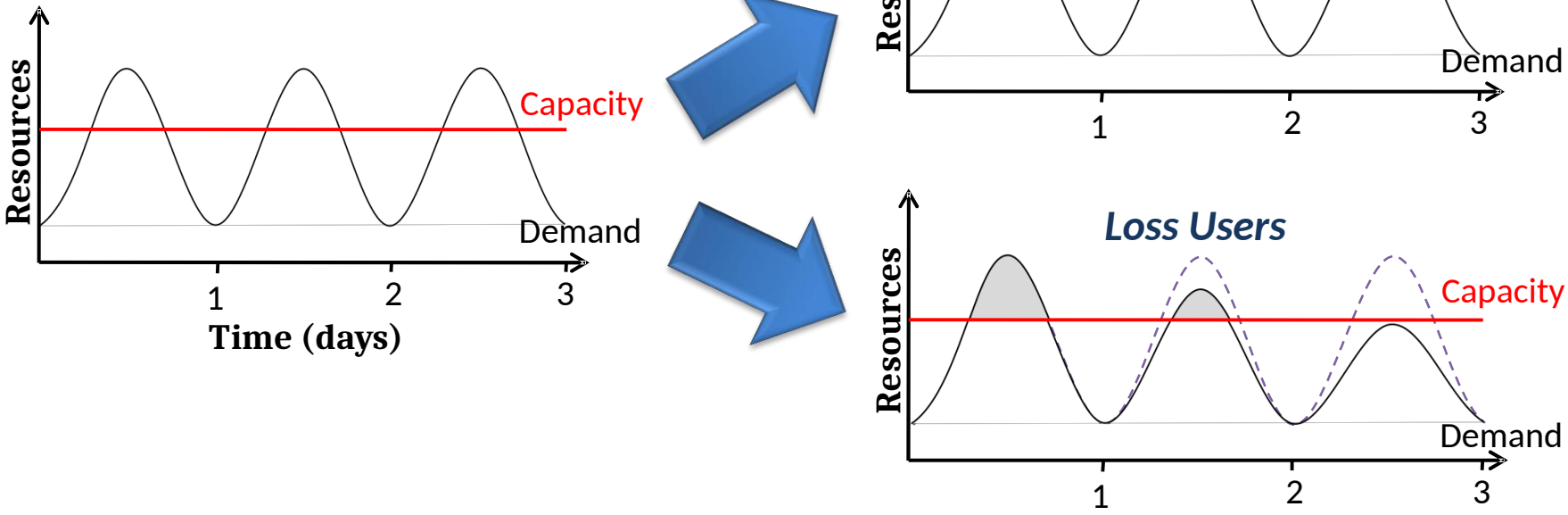
Dynamic Provisioning

- What is dynamic provisioning ?
  - Dynamic Provisioning is a simplified way to explain a complex networked server computing environment where server computing instances are provisioned or deployed from a administrative console or client application by the server administrator, network administrator, or any other enabled user.

Scalability
Elasticity
• Dynamic provision
• Multi-tenant design

Dynamic Provisioning

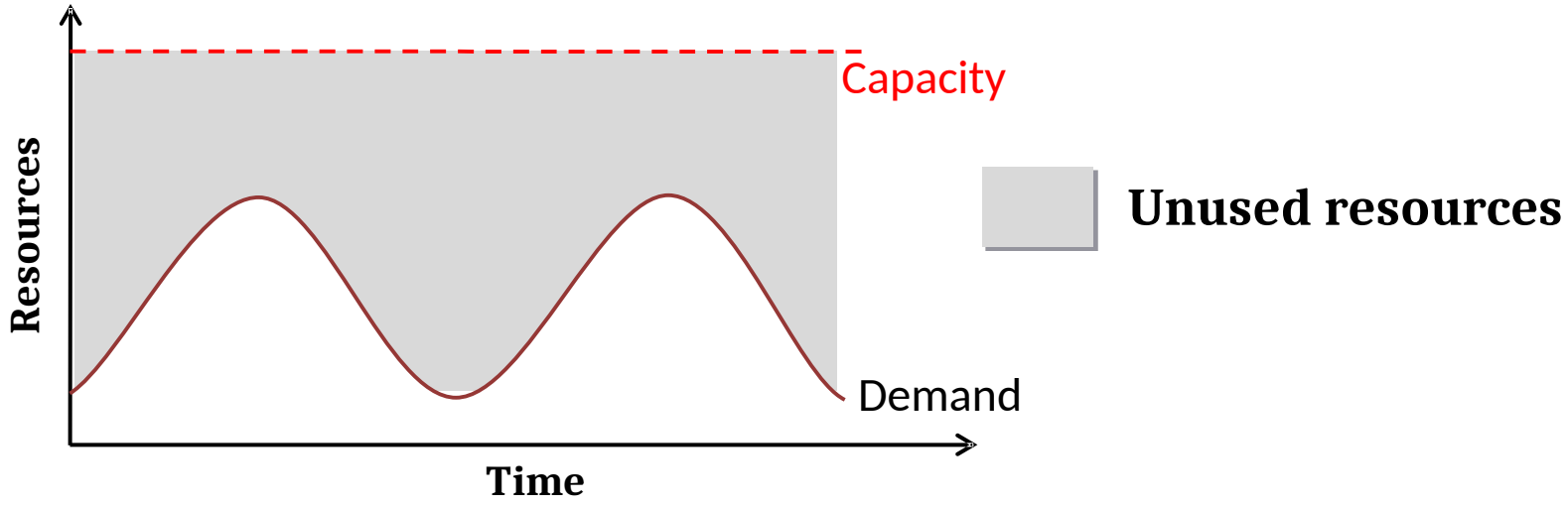
- In traditional computing model, two common problems :
  - Underestimate system utilization which result in under provision



Scalability
Elasticity
• Dynamic provision
• Multi-tenant design

Dynamic Provisioning

- Overestimate system utilization which result in low utilization

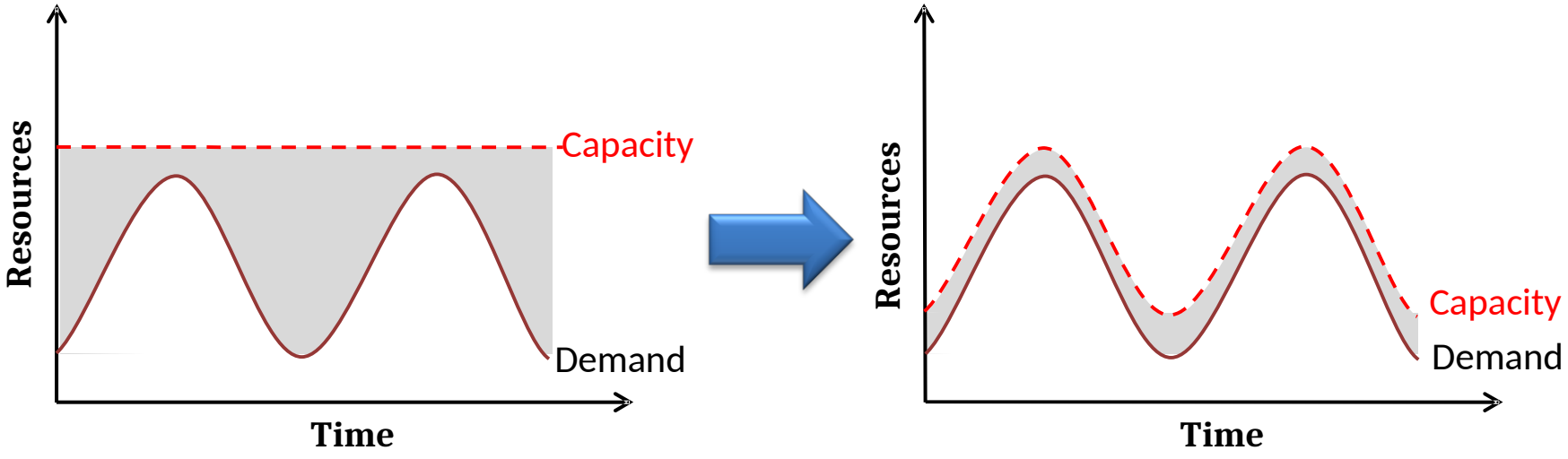


- How to solve this problem ??
  - Dynamically provision resources

Scalability
Elasticity
• Dynamic provision
• Multi-tenant design

Dynamic Provisioning

- Cloud resources should be provisioned dynamically
  - Meet seasonal demand variations
  - Meet demand variations between different industries
  - Meet burst demand for some extraordinary events



Multi-tenant Design

- What is multi-tenant design ?
  - Multi-tenant refers to a principle in software architecture where a single instance of the software runs on a server, serving multiple client organizations.
  - With a multi-tenant architecture, a software application is designed to virtually partition its data and configuration thus each client organization works with a customized virtual application instance.

- Client oriented requirements :
  - Customization
    - Multi-tenant applications are typically required to provide a high degree of customization to support each target organization's needs.
  - Quality of service
    - Multi-tenant applications are expected to provide adequate levels of security and robustness.

Availability & Reliability

- What is availability ?
  - The degree to which a system, subsystem, or equipment is in a specified operable and committable state at the start of a mission, when the mission is called for at an unknown time.
  - Cloud system usually require high availability
    - Ex. “Five Nines” system would statistically provide 99.999% availability
- What is reliability ?
  - The ability of a system or component to perform its required functions under stated conditions for a specified period of time.
- But how to achieve these properties ?
  - Fault tolerance system
  - Require system resilience
  - Reliable system security

Fault Tolerance

- What is fault tolerant system ?
  - Fault-tolerance is the property that enables a system to continue operating properly in the event of the failure of some of its components.
  - If its operating quality decreases at all, the decrease is proportional to the severity of the failure, as compared to a naively-designed system in which even a small failure can cause total breakdown.

- Four basic characteristics :
  - No single point of failure
  - Fault detection and isolation to the failing component
  - Fault containment to prevent propagation of the failure
  - Availability of reversion modes

Availability
Reliability
- Fault tolerance
- System resilience
- System security

Fault Tolerance

- Single Point Of Failure (SPOF)
  - A part of a system which, if it fails, will stop the entire system from working.
  - The assessment of a potentially single location of failure identifies the critical components of a complex system that would provoke a total systems failure in case of malfunction.
- Preventing single point of failure
  - If a system experiences a failure, it must continue to operate without interruption during the repair process.

Availability
Reliability
- Fault tolerance
- System resilience
- System security

Fault Tolerance

- Fault Detection and Isolation (FDI)
  - A subfield of control engineering which concerns itself with monitoring a system, identifying when a fault has occurred and pinpoint the type of fault and its location.

- Isolate failing component
  - When a failure occurs, the system must be able to isolate the failure to the offending component.

Fault Tolerance

- Fault Containment
  - Some failure mechanisms can cause a system to fail by propagating the failure to the rest of the system.
  - Mechanisms that isolate a rogue transmitter or failing component to protect the system are required.
- Available of reversion modes
  - System should be able to maintain some check points which can be used in managing the state changes.

System Resilience

- What is resilience ?
  - Resilience is the ability to provide and maintain an acceptable level of service in the face of faults and challenges to normal operation.
  - Resiliency pertains to the system's ability to return to its original state after encountering trouble. In other words, if a risk event knocks a system offline, a highly resilient system will return back to work and function as planned as soon as possible.

- Some risk events
  - If power is lost at a plant for two days, can our system recover ?
  - If a key service is lost because a database becomes corrupt, can the business recover ?

System Resilience

- Disaster Recovery
  - Disaster recovery is the process, policies and procedures related to preparing for recovery or continuation of technology infrastructure critical to an organization after a natural or human-induced disaster.

- Some common strategies :
  - Backup
    - Make data off-site at regular interval
    - Replicate data to an off-site location
    - Replicate whole system
  - Preparing
    - Local mirror systems
    - Surge protector
    - Uninterruptible Power Supply (UPS)

Availability
Reliability
- Fault tolerance
- System resilience
- System security

System Security

- Security issue in Cloud Computing :
  - Cloud security is an evolving sub-domain of computer security, network security, and, more broadly, information security.
  - It refers to a broad set of policies, technologies, and controls deployed to protect data, applications, and the associated infrastructure of cloud computing.

System Security

- Important security and privacy issues :
  - Data Protection
    - To be considered protected, data from one customer must be properly segregated from that of another.
  - Identity Management
    - Every enterprise will have its own identity management system to control access to information and computing resources.
  - Application Security
    - Cloud providers should ensure that applications available as a service via the cloud are secure.
  - Privacy
    - Providers ensure that all critical data are masked and that only authorized users have access to data in its entirety.

Manageability & Interoperability

- What is manageability ?
  - Enterprise-wide administration of cloud computing systems. Systems manageability is strongly influenced by network management initiatives in telecommunications.
- What is interoperability ?
  - Interoperability is a property of a product or system, whose interfaces are completely understood, to work with other products or systems, present or future, without any restricted access or implementation.
- But how to achieve these properties ?
  - System control automation
  - System state monitoring

Control Automation

- What is Autonomic Computing ?
  - Its ultimate aim is to develop computer systems capable of self-management, to overcome the rapidly growing complexity of computing systems management, and to reduce the barrier that complexity poses to further growth.

- Architectural framework :
  - Composed by Autonomic Components (AC) which will interact with each other.
  - An AC can be modeled in terms of two main control loops (local and global) with sensors (for self-monitoring), effectors (for self-adjustment), knowledge and planer/adapter for exploiting policies based on self- and environment awareness.

Control Automation

- Four functional areas :
  - Self-Configuration
    - Automatic configuration of components.
  - Self-Healing
    - Automatic discovery, and correction of faults.
  - Self-Optimization
    - Automatic monitoring and control of resources to ensure the optimal functioning with respect to the defined requirements.
  - Self-Protection
    - Proactive identification and protection from arbitrary attacks.

Manageability
Interoperability
- Control automation
- System monitoring
- Billing system

System Monitoring

- What is system monitor ?
  - A System Monitor in systems engineering is a process within a distributed system for collecting and storing state data.
- What should be monitored in the Cloud ?
  - Physical and virtual hardware state
  - Resource performance metrics
  - Network access patterns
  - System logs
  - … etc
- Anything more ?
  - Billing system

Manageability
Interoperability
- Control automation
- System monitoring
- Billing system

Billing System

- Billing System in Cloud
  - Users pay as many as they used.
  - Cloud provider must first determine the list of service usage price.
  - Cloud provider have to record the resource or service usage of each user, and then charge users by these records.
- How can cloud provider know users' usage ?
  - Get those information by means of monitoring system.
  - Automatically calculate the total amount of money which user should pay. And automatically request money from use's banking account.

Performance & Optimization

- Performance guarantees ??
  - As the great computing power in cloud, application performance should be guaranteed.
  - Cloud providers make use of powerful infrastructure or other underlining resources to build up a highly performed and highly optimized environment, and then deliver the complete services to cloud users.
- But how to achieve this property ?
  - Parallel computing
  - Load balancing
  - Job scheduling

Parallel Processing

- Parallel Processing
  - Parallel processing is a form of computation in which many calculations are carried out simultaneously, operating on the principle that large problems can often be divided into smaller ones, which are then solved concurrently.

- Parallelism in different levels :
  - Bit level parallelism
  - Instruction level parallelism
  - Data level parallelism
  - Task level parallelism

Performance Optimization
- Parallel processing
- Load balancing
- Job scheduling

Parallel Processing

- Hardware approaches
  - Multi-core computer
  - Symmetric multi-processor
  - General purpose graphic processing unit
  - Vector processor
  - Distributed computing
    - Cluster computing
    - Grid computing
- Software approaches
  - Parallel programming language
  - Automatic parallelization

Load Balancing

- What is load balancing ?
  - Load balancing is a technique to distribute workload evenly across two or more computers, network links, CPUs, hard drives, or other resources, in order to get optimal resource utilization, maximize throughput, minimize response time, and avoid overload.

- Why should be load balanced ?
  - Improve resource utilization
  - Improve system performance
  - Improve energy efficiency

Unbalanced 

Job Scheduling

- What is job scheduler ?
  - A job scheduler is a software application that is in charge of unattended background executions, commonly known for historical reasons as batch processing.
- What should be scheduled in Cloud ?
  - Computation intensive tasks
  - Dynamic growing and shrinking tasks
  - Tasks with complex processing dependency
- How to approach ?
  - Use pre-defined workflow
  - System automatic configuration

Accessibility & Portability

- What is accessibility ?
  - Accessibility is a general term used to describe the degree to which a product, device, service, or environment is accessible by as many people as possible.
- What is service portability ?
  - Service portability is the ability to access services using any devices, anywhere, continuously with mobility support and dynamic adaptation to resource variations.
- But how to achieve these properties ?
  - Uniform access
  - Thin client

Accessibility
Portability
• Uniform access
• Thin client

Uniform Access

- How do users access cloud services ?
  - Cloud providers should provide their cloud services by means of widespread accessing media. In other word, users from different operating systems or other accessing platforms should be able to directly be served.
  - Nowadays, web browser technique is one of the most widespread platform in almost any intelligent electronic devices. Cloud services take this into concern, and delivery their services with web-based interface through the Internet.

Thin Client

- What is thin client ?
  - Thin client is a computer or a computer program which depends heavily on some other computer to fulfill its traditional computational roles. This stands in contrast to the traditional fat client, a computer designed to take on these roles by itself.
- Characteristics :
  - Cheap client hardware
    - While the cloud providers handle several client sessions at once, the clients can be made out of much cheaper hardware.
  - Diversity of end devices
    - End user can access cloud service via plenty of various electronic devices, which include mobile phones and smart TV.
  - Client simplicity
    - Client local system do not need complete operational functionalities.

What can we gain from cloud ?

WHAT IS CLOUD COMPUTING ?

Benefits From Cloud

- Cloud computing brings many benefits :
  - For the market and enterprises
    - Reduce initial investment
    - Reduce capital expenditure
    - Improve industrial specialization
    - Improve resource utilization
  - For the end user and individuals
    - Reduce local computing power
    - Reduce local storage power
    - Variety of thin client devices in daily life

Reduce Initial Investment

- Traditional process of enterprises to initiate business :
  - Survey and analysis the industry and market
  - Estimate the quantity of supply and demand
  - Purchase and deploy IT infrastructure
  - Install and test the software system
  - Design and develop enterprise specific business service
  - Announce the business service to clients

- Some drawbacks :
  - The survey, analysis and estimation may not 100% correct
  - Infrastructure deployment is time consuming
  - Enterprises should take the risk of wrong investment

Reduce Initial Investment

- Initiate business with Cloud Computing services :
  - Survey and analysis the industry and market
  - Chose one cloud provider for enterprise deployment
  - Design and develop business service upon cloud environment
  - Announce the business service to clients
- Some benefits :
  - Enterprise do not need to own the infrastructure
  - Enterprise can develop and deploy business service in short time
  - Enterprise can reduce the business loss of wrong investment

Reduce Initial Investment

- What dose cloud computing achieve ?

	Traditional	With Cloud Computing
Investment Risk	*Enterprise takes the risk*	*Cloud reduces the risk*
Infrastructure	*Enterprise owns the infrastructure*	*Cloud provider owns the infrastructure*
Time duration	*Long deployment time*	*Fast to business ready*

Reduce Capital Expenditure

- Traditional capital expenditure of enterprises :
  - Each enterprise should establish its own IT department
  - IT department should handle the listing jobs
    - Manage and administrate hardware and software
    - Apply regular data backup and check point process
    - Purchase new infrastructure and eliminate outdated one
    - Always standby for any unexpected IT problems

- Some drawbacks :
  - Enterprise pays for IT investment which is not its business focus
  - Enterprise should take the risk of hardware/software malfunction
  - Replacing and updating infrastructure is time consuming and risky

Reduce Capital Expenditure

- Capital expenditure with Cloud Computing service :
  - Enterprise can almost dismiss its IT department
  - The jobs of IT department can be achieved by cloud provider
    - Dynamically update and upgrade hardware or software
    - Dynamically provision and deploy infrastructure for enterprise
    - Automatically backup data and check consistency
    - Self-recover from disaster or system malfunction
- Some benefits :
  - Enterprise can shift effort to its business focus
  - Enterprise can reconfigure its IT services in short time
  - Enterprise pays to cloud provider as many as the service used

Reduce Capital Expenditure

- What dose cloud computing achieve ?

	Traditional	With Cloud Computing
Business focus	*Need to own its IT department*	*Cloud provider takes care everything*
Payment	*Pay for all investment and human resource*	*Enterprise pays as the service used*
Time duration	*Long establish time*	*Fast to business ready*

Improve Industrial Specialization

- Traditional industry and market :
  - Every enterprise has to own its IT department
  - IT resource is managed by enterprise themselves
  - IT complexity should be well taken care by enterprise themselves
- Some drawbacks :
  - IT department is not the business focus of enterprise
  - Most of enterprises do not well maintain their IT resources
  - Enterprises seldom optimize their IT resource usage

Improve Industrial Specialization

- Collaboration with Cloud providers :
  - Cloud providers centrally maintain IT infrastructure for clients
  - Cloud providers employ experts for management and administration
  - Cloud providers focus on providing reliable IT services
  - Enterprises only rent the service they need and care
- Some benefits :
  - Industrial specialization will be improved
  - IT service performance will be optimized
  - Enterprise business focus will be enhanced
  - IT resource waste will be reduced

Improve Industrial Specialization

- What dose cloud computing achieve ?

	Traditional	With Cloud Computing
Collaboration	*Enterprise needs to take care everything*	*Enterprise focuses on its own business*
Management	*Enterprise works with poor manageability*	*Cloud provider applies professional control*
Relationship	*Stand alone enterprise*	*Win-Win partnership*

Improve Resource Utilization

- Traditional industry and market :
  - Enterprise seldom takes care about IT resource utilization
  - Enterprise owns its IT resource without well management
  - IT resource usually over invested for peak demand
- Some drawbacks :
  - Power and space utilization among enterprises are wasted
  - IT resources across enterprises cannot be shared

Improve Resource Utilization

- Collaboration with Cloud providers :
  - IT resources are centrically managed and optimized
    - Cloud provider builds performance optimized hardware
    - Cloud provider builds consolidated cooling system
    - Cloud provider will consider the geographic issues
    - Cloud provider will consider legal policy issues
- Some benefits :
  - IT infrastructure can be shared among enterprises
  - IT infrastructure performance and utilization can be optimized
  - Large-scale integrated optimization can be applied

Improve Resource Utilization

- What dose cloud computing achieve ?

	Traditional	With Cloud Computing
IT Resource Utilization	*IT resource under utilized most of time*	*Share to improve utilization of IT resource*
Power Consumption	*Waste power and cooling system*	*Cloud system should be global optimized*

Reduce Local Computing Power

- Traditional local computing power requirement :
  - One need to buy your own personal computer
  - Buy powerful processor if you need intensive computing
  - Buy large memory to meet application requirement
  - Install plenty of applications in need
- Some drawbacks :
  - One can hardly replicate the same system environment
  - One needs to regularly update or upgrade software and hardware
  - One needs to reinstall all applications if you reinstall the OS

Reduce Local Computing Power

- Using Cloud Computing services :
  - One can utilize the remote computing power in the cloud
  - One needs only basic computing power to connect to internet
  - Application in the cloud will automatically upgrade
- Some benefits :
  - One can access his/her applications anywhere through the Internet
  - One can dynamically request for computing power on demand
  - Application may need not to be reinstalled even reinstall the OS

Reduce Local Computing Power

- What dose cloud computing achieve ?

	Traditional	With Cloud Computing
Hardware Requirement	*User needs to buy powerful hardware*	*Only basic hardware to connect to internet*
Software Requirement	*Install application in local computer*	*No local installation requirement*
Portability	*Hard to be portable*	*Natively portable*

Reduce Local Storage Power

- Traditional local storage power requirement :
  - User programs and data files are stored in local devices
  - User has to backup data regularly preventing hardware damage
- Some drawbacks :
  - Storage space may not enough for burst data requirement
  - Storage space may be over needed which result in resource waste
  - Data consistency is hard to maintain between computers
  - Need to sacrifice part of storage space for data backup

Reduce Local Storage Power

- Using Cloud Computing services :
  - User programs and data files are stored in the cloud
  - Cloud service provider will guarantee the data availability
- Some benefits :
  - One can dynamically allocate storage space on demand
  - One can access data anywhere through the Internet
  - No need to care about data consistency between computers
  - No need to care about data loss due to hardware damage

Reduce Local Storage Power

- What dose cloud computing achieve ?

	Traditional	With Cloud Computing
Storage Space	*Limited to local disk, may be under utilized*	*Dynamically allocated on demand*
Storage Data Consistency	*Difficult to maintain data consistency*	*Data consistency maintained by cloud*
Availability	*Regular user backup*	*Cloud service guarantee*

Variety of End Devices

- Traditional computing resource :
  - One can connect to the Internet by personal computer
  - Only personal computer can deliver reasonable computing power
  - Small device cannot perform incentive computation due to its power and hardware limitation
- Some drawbacks :
  - Computing power is not portable
  - Small devices can only perform simplified works

Variety of End Devices

- Devices collaborate with Cloud services :
  - Device connects to the Internet through wireless network
  - Device accesses cloud services through web service interface
  - Device sends computing incentive jobs into cloud and wait for results
- Some benefits :
  - User can easily access cloud service through small devices
  - User can access almost unlimited computing power anywhere
  - Small devices can be intelligently managed through cloud

Variety of End Devices

- What dose cloud computing achieve ?

	Traditional	With Cloud Computing
Computing Power	*Only accessed through desktop computer*	*Accessed through small smart devices*
Small Device Intelligence	*Functionalities was limited due to their power consumption*	*Shift computing incentive jobs into cloud, and then wait for results*

Agenda

- What is Cloud Computing ?
  - Different perspectives
  - Properties and characteristics
  - Benefits from cloud computing

- Service and deployment models
  - Three service models
  - Four deployment models

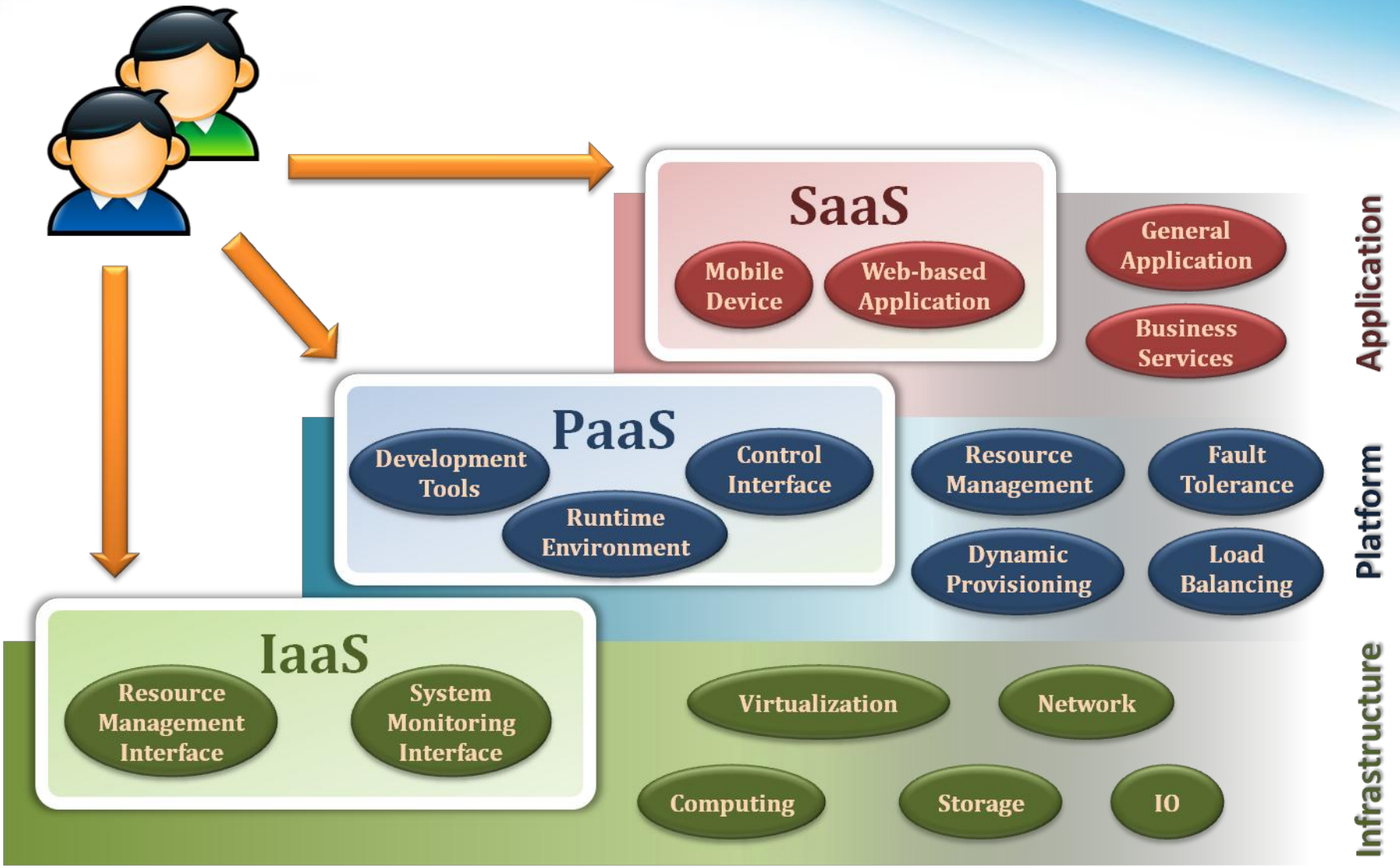
Choose the service you need.

SERVICE MODELS

Service Models Overview

- What if you want to have an IT department ?
  - Similar to ***build a new house*** in previous analogy
    - You can rent some virtualized infrastructure and build up your own IT system among those resources, which may be fully controlled.
    - Technical speaking, use the ***Infrastructure as a Service (IaaS)*** solution.
  - Similar to ***buy an empty house*** in previous analogy
    - You can directly develop your IT system through one cloud platform, and do not care about any lower level resource management.
    - Technical speaking, use the ***Platform as a Service (PaaS)*** solution.
  - Similar to ***live in a hotel*** in previous analogy
    - You can directly use some existed IT system solutions, which were provided by some cloud application service provider, without knowing any detail technique about how these service was achieved.
    - Technical speaking, use the ***Software as a Service (SaaS)*** solution.

Service Model Overview



Infrastructure as a Service

Platform as a Service

Software as a Service

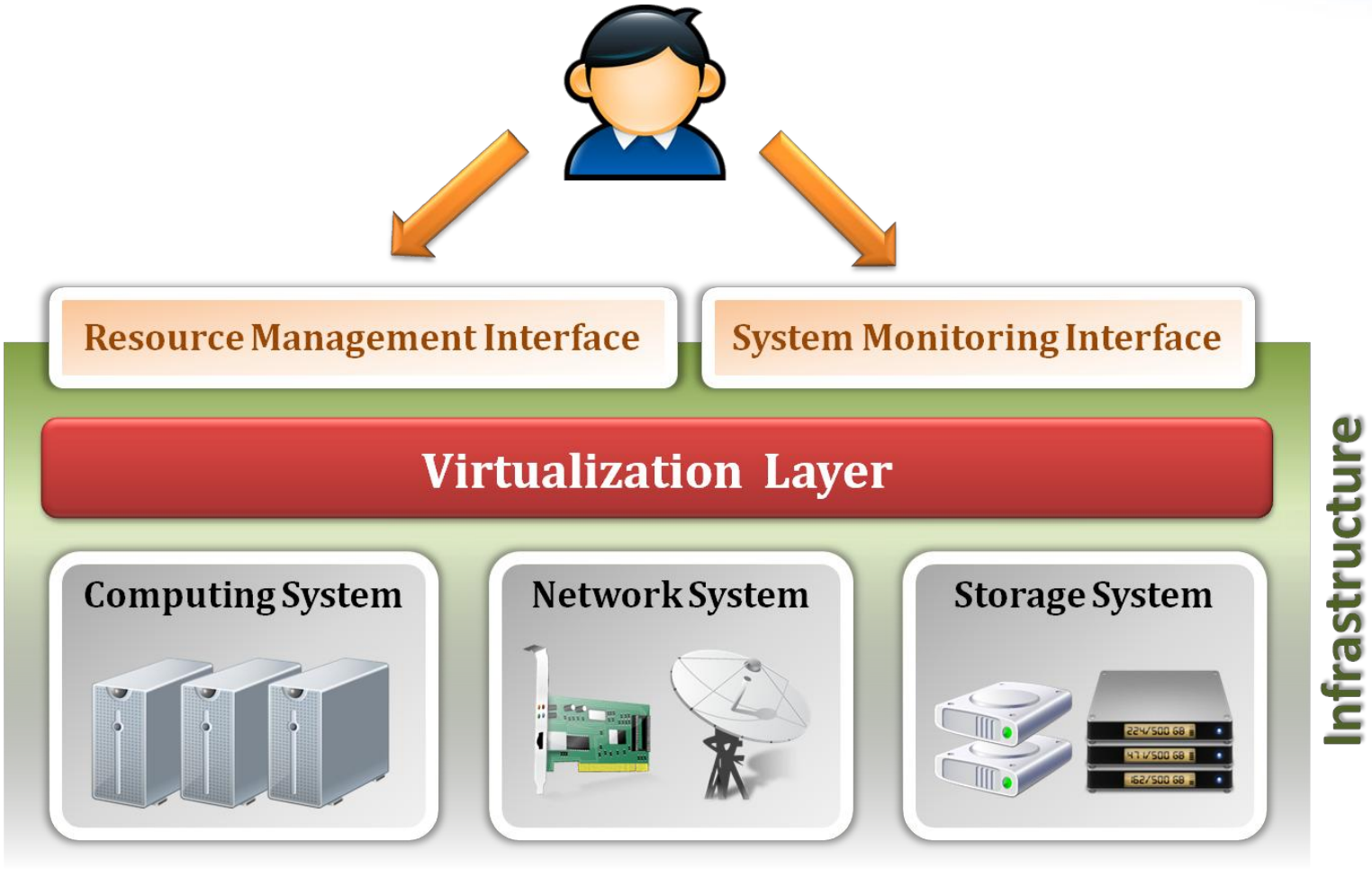
SERVICE MODELS

Infrastructure as a Service

- Infrastructure as a Service - IaaS
  - The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications.
  - The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of select networking components .
- Examples :
  - Amazon EC2
  - Eucalyputs
  - OpenNebula
  - … etc

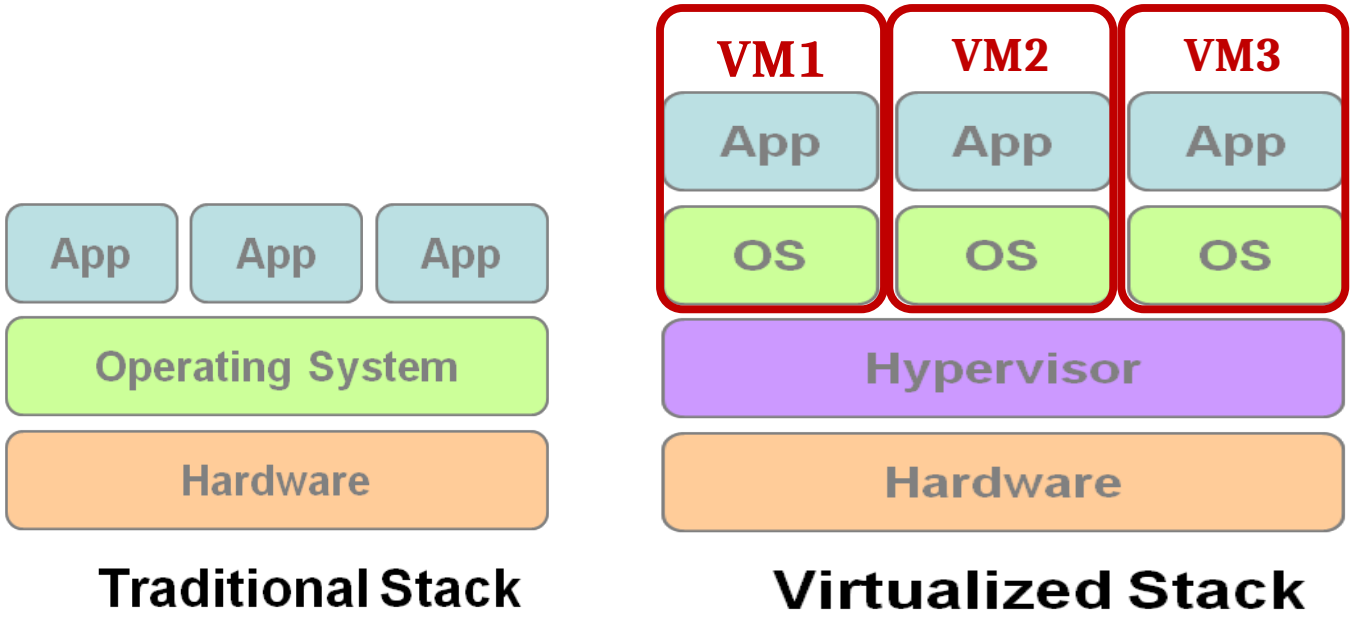
Infrastructure as a Service

- System architecture :



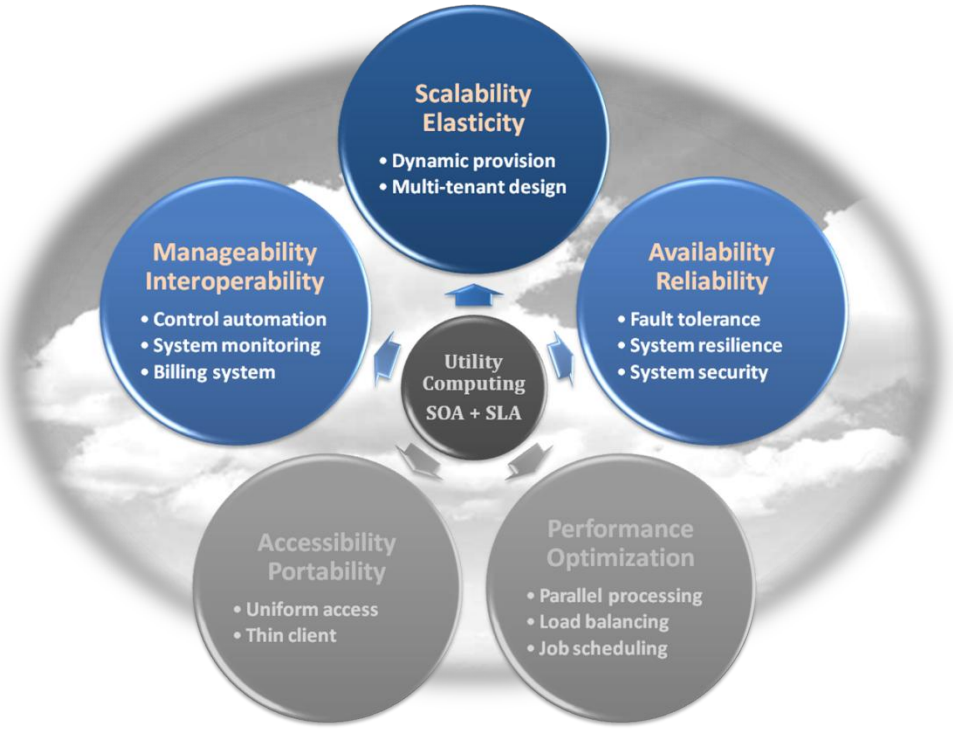
Infrastructure as a Service

- Enabling technique - ***Virtualization***
  - Virtualization is an abstraction of logical resources away from underlying physical resources.
    - Virtualization technique shift OS onto hypervisor.
    - Multiple OS share the physical hardware and provide different services.
    - Improve utilization, availability, security and convenience.



Infrastructure as a Service

- Properties supported by virtualization technique :
  - Manageability and Interoperability
  - Availability and Reliability
  - Scalability and Elasticity



Infrastructure as a Service

- Provide service –**Resource Management Interface**
  - Several types of virtualized resource :
    - ***Virtual Machine*** – As an IaaS provider, we should be able to provide the basic virtual machine operations, such as *creation*, *suspension*, *resumption* and *termination*, …etc.
    - ***Virtual Storage*** – As an IaaS provider, we should be able to provide the basic virtual storage operations, such as *space allocation*, *space release*, *data writing* and *data reading*, …etc.
    - ***Virtual Network*** – As an IaaS provider, we should be able to provide the basic virtual network operations, such as *IP address allocation*, *domain name register*, *connection establishment* and *bandwidth provision*, …etc.

Infrastructure as a Service

- Provide service – **System Monitoring Interface**
  - Several types of monitoring metrics :
    - ***Virtual Machine*** – As an IaaS provider, we should be able to monitor some system states of each virtual machine, such as *CPU loading*, *memory utilization*, *IO loading* and *internal network loading*, …etc.
    - ***Virtual Storage*** – As an IaaS provider, we should be able to monitor some storage states of each virtual storage, such as *virtual space utilization*, *data duplication* and *storage device access bandwidth*, …etc.
    - ***Virtual Network*** – As an IaaS provider, we should be able to monitor some network states of each virtual network, such as *virtual network bandwidth*, *network connectivity* and *network load balancing*, …etc.

IaaS - Summary

- **IaaS is the deployment platform that abstract the infrastructure.**

- IaaS enabling technique
  - Virtualization
    - Server Virtualization
    - Storage Virtualization
    - Network Virtualization

- IaaS provided services
  - Resource Management Interface
  - System Monitoring Interface

Infrastructure as a Service

Platform as a Service

Software as a Service

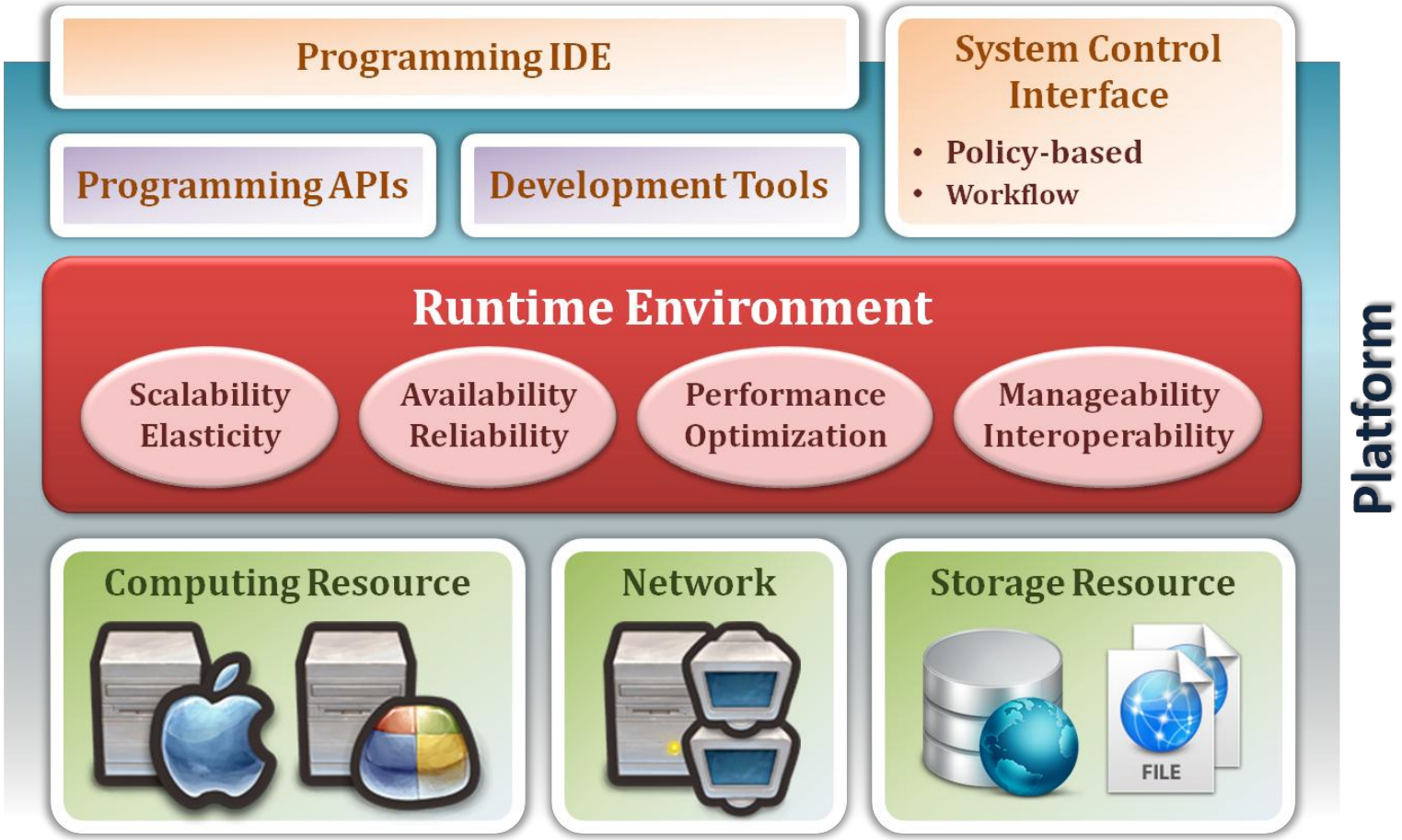
SERVICE MODELS

Platform as a Service

- Platform as a Service - PaaS
  - The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages and tools supported by the provider.
  - The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly application hosting environment configurations.
- Examples :
  - Microsoft Windows Azure
  - Google App Engine
  - Hadoop
  - … etc

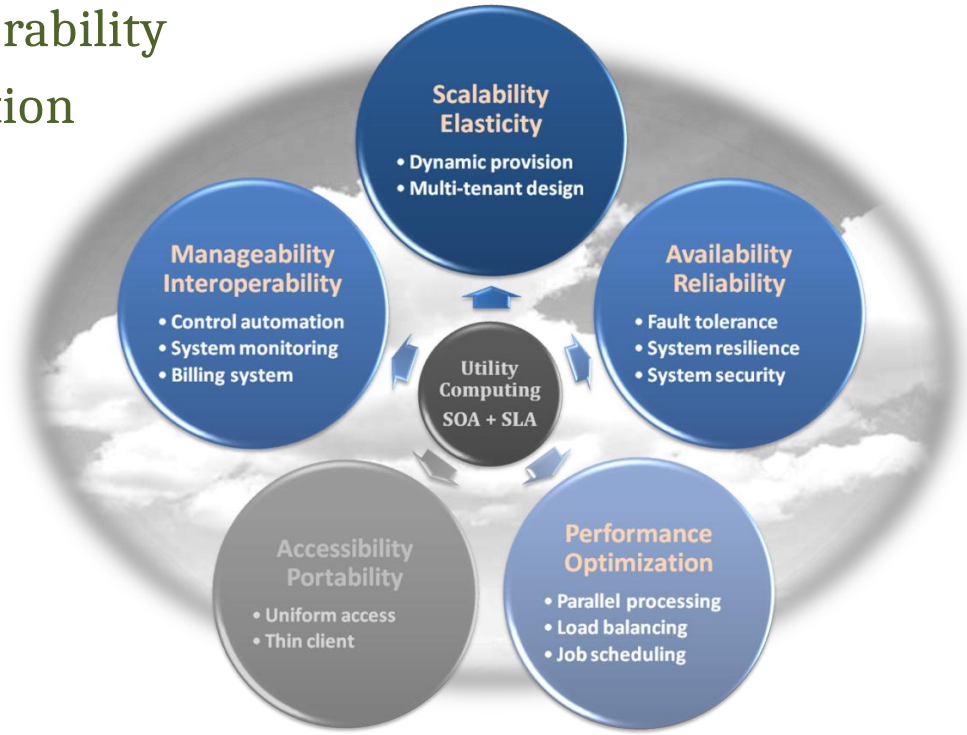
Platform as a Service

- System architecture :



Platform as a Service

- Enabling technique – **Runtime Environment Design**
  - Runtime environment refers to collection of software services available. Usually implemented by a collection of program libraries.
- Common properties in Runtime Environment :
  - Manageability and Interoperability
  - Performance and Optimization
  - Availability and Reliability
  - Scalability and Elasticity



Platform as a Service

- Provide service – **Programming IDE**
  - Users make use of programming IDE to develop their services among PaaS.
    - This IDE should integrate the full functionalities which supported from the underling runtime environment.
    - This IDE should also provide some development tools, such as profiler, debugger and testing environment.
  - The programming APIs supported from runtime environment may be various between different cloud providers, but there are still some common operating functions.
    - Computation, storage and communication resource operation

Platform as a Service

- Provide service – **System Control Interface**
  - Policy-Based Control
    - Typically described as a principle or rule to guide decisions and achieve rational outcome(s)
    - Make the decision according to some requirements
  - Workflow Control
    - Describe the flow of installation and configuration of resources
    - Workflow processing daemon delivers speedy and efficient construction and management of cloud resources

PaaS - Summary

- **PaaS is the development platform that abstract the infrastructure, OS, and middleware to drive developer productivity.**

- PaaS enabling technique
  - Runtime Environment

- PaaS provide services
  - Programming IDE
    - Programming APIs
    - Development tools
  - System Control Interface
    - Policy based approach
    - Workflow based approach

Infrastructure as a Service

Platform as a Service

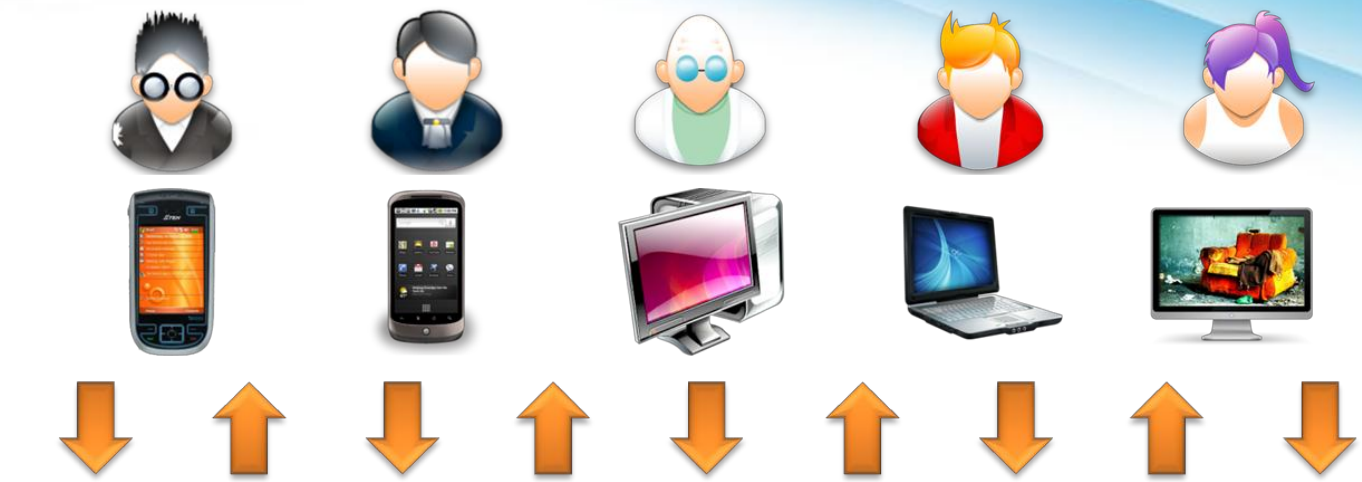
Software as a Service

SERVICE MODELS

Software as a Service

- Software as a Service - SaaS
  - The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through a thin client interface such as a web browser (e.g., web-based email).
  - The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.
- Examples :
  - Google Apps (e.g., Gmail, Google Docs, Google sites, …etc)
  - SalesForce.com
  - EyeOS
  - … etc

Software as a Service



Web-base Applications

Web Portal

General Applications

Business Applications

Scientific Applications

Government Applications

Commercial Platform

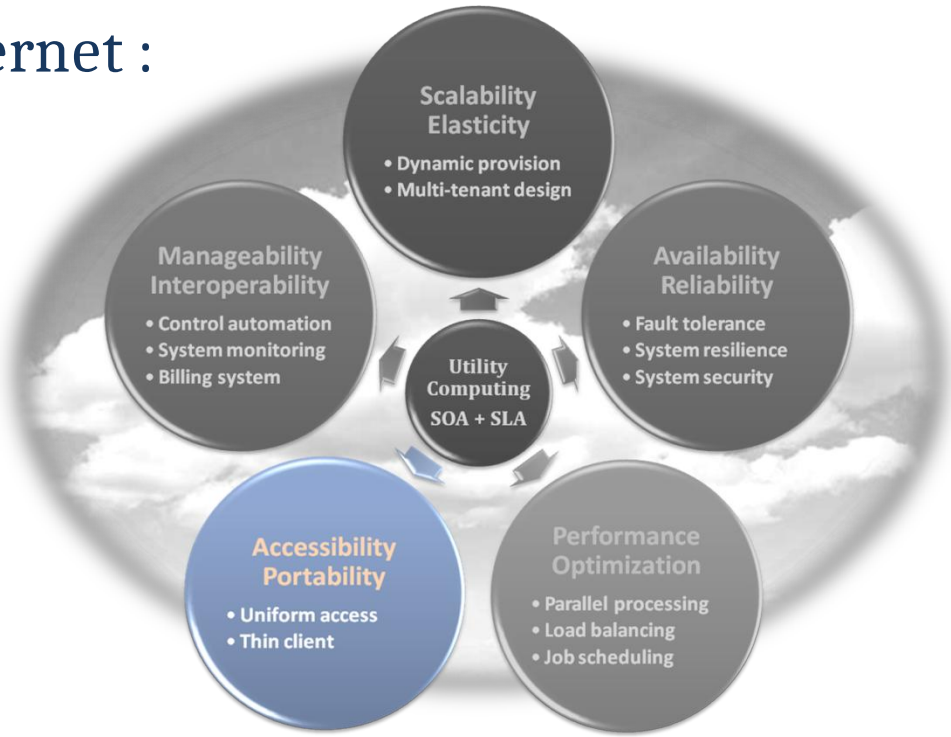
Open Source Platform

Other

Application

Software as a Service

- Enabling Technique – **Web Service**
  - Web 2.0 is the trend of using the full potential of the web
    - Viewing the Internet as a computing platform
    - Running interactive applications through a web browser
    - Leveraging interconnectivity and mobility of devices
    - Enhanced effectiveness with greater human participation
- Properties provided by Internet :
  - Accessibility and Portability



Software as a Service

- Provide service – **Web-based Applications**
  - Conventional applications should translate their access interfaces onto web-based platform.
  - Applications in different domains
    - ***General Applications*** – Applications which are designed for general propose, such as office suit, *multimedia* and *instant message*, …etc.
    - ***Business Applications*** – Application which are designed for business propose, such as *ERP*, *CRM* and *market trading system*, …etc.
    - ***Scientific Applications*** – Application which are designed for scientific propose, such as *aerospace simulation* and *biochemistry simulation*, …etc.
    - ***Government Applications*** – Applications which are designed for government propose, such as *national medical system* and *public transportation system service*, …etc.

Software as a Service

- Provide service – **Web Portal**
  - Apart from the standard search engine feature, web portals offer other services such as e-mail, news, stock prices, information, databases and entertainment.
  - Portals provide a way for enterprises to provide a consistent look and feel with access control and procedures for multiple applications and databases, which otherwise would have been different entities altogether.
  - Some examples :
    - iGoogle
    - MSNBC
    - Netvibes
    - Yahoo!

SaaS - Summary

- **SaaS is the finished applications that you rent and customize.**
- SaaS enabling technique
  - Web Service
- SaaS provide services
  - Web-based Applications
    - General applications
    - Business applications
    - Scientific applications
    - Government applications
  - Web Portal

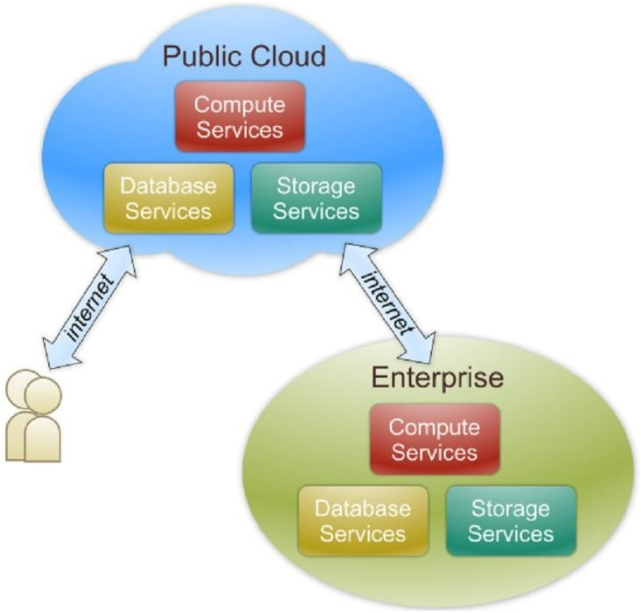
Deployment Model

- There are four primary cloud deployment models :
  - Public Cloud
  - Private Cloud
  - Community Cloud
  - Hybrid Cloud

- Each can exhibit the previously discussed characteristics; their differences lie primarily in the scope and access of published cloud services, as they are made available to service consumers.

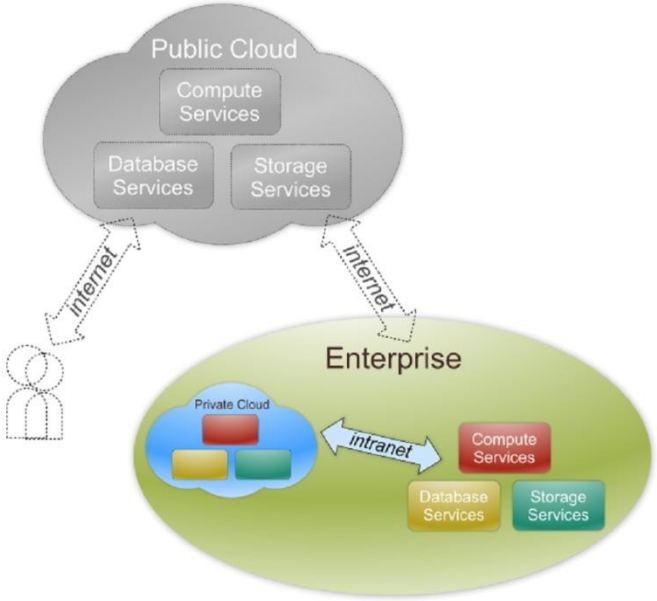
Public Cloud

- Public cloud definition
  - The cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services.
  - Also known as external cloud or multi-tenant cloud, this model essentially represents a cloud environment that is openly accessible.
  - Basic characteristics :
    - Homogeneous infrastructure
    - Common policies
    - Shared resources and multi-tenant
    - Leased or rented infrastructure
    - Economies of scale



Private Cloud

- Private cloud definition
  - 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.
  - Also referred to as internal cloud or on-premise cloud, a private cloud intentionally limits access to its resources to service consumers that belong to the same organization that owns the cloud.
  - Basic characteristics :
    - Heterogeneous infrastructure
    - Customized and tailored policies
    - Dedicated resources
    - In-house infrastructure
    - End-to-end control



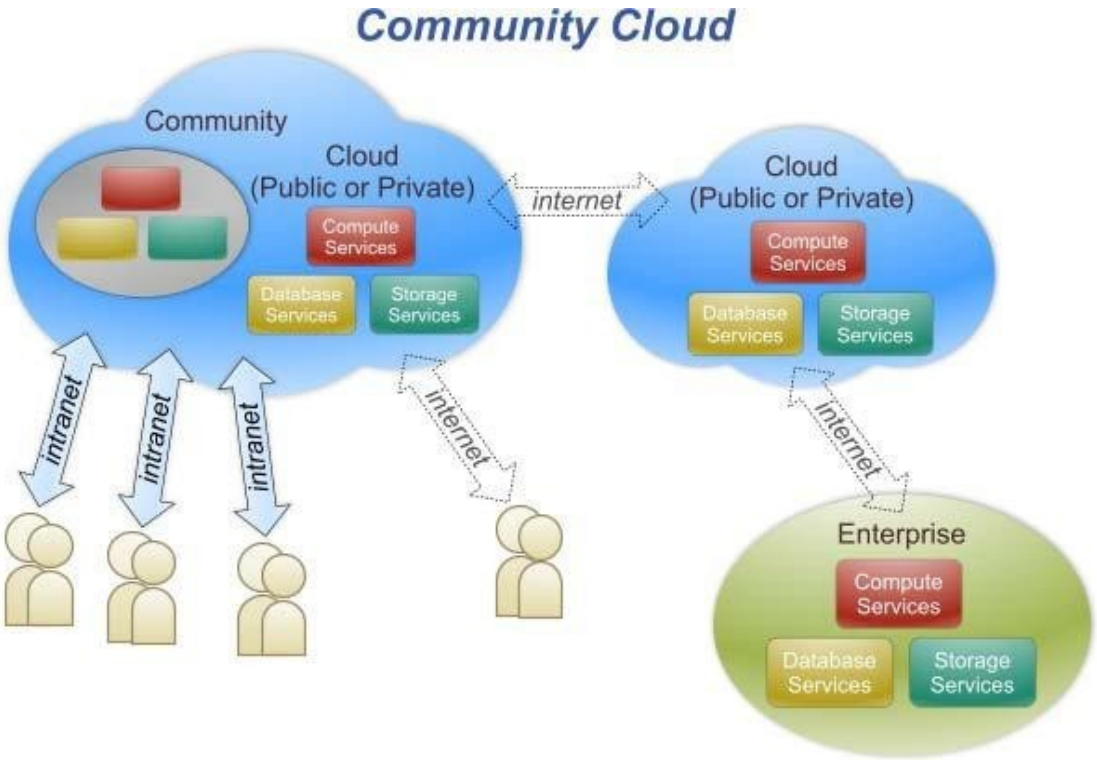
Public vs. Private

- Comparison :

	Public Cloud	Private Cloud
Infrastructure	*Homogeneous*	*Heterogeneous*
Policy Model	*Common defined*	*Customized & Tailored*
Resource Model	*Shared & Multi-tenant*	*Dedicated*
Cost Model	*Operational expenditure*	*Capital expenditure*
Economy Model	*Large economy of scale*	*End-to-end control*

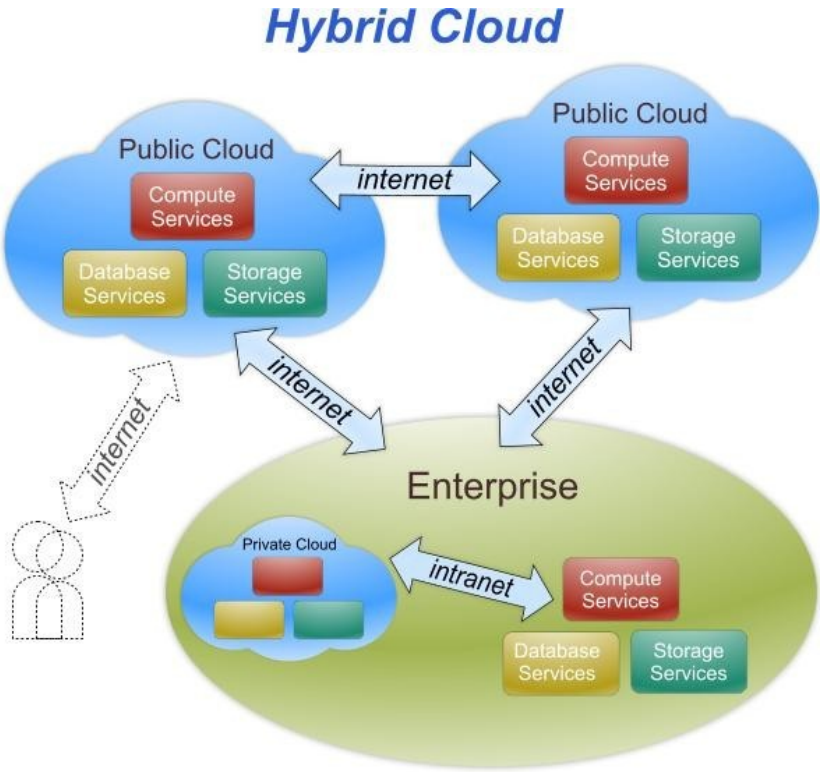
Community Cloud

- Community cloud definition
  - The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations).

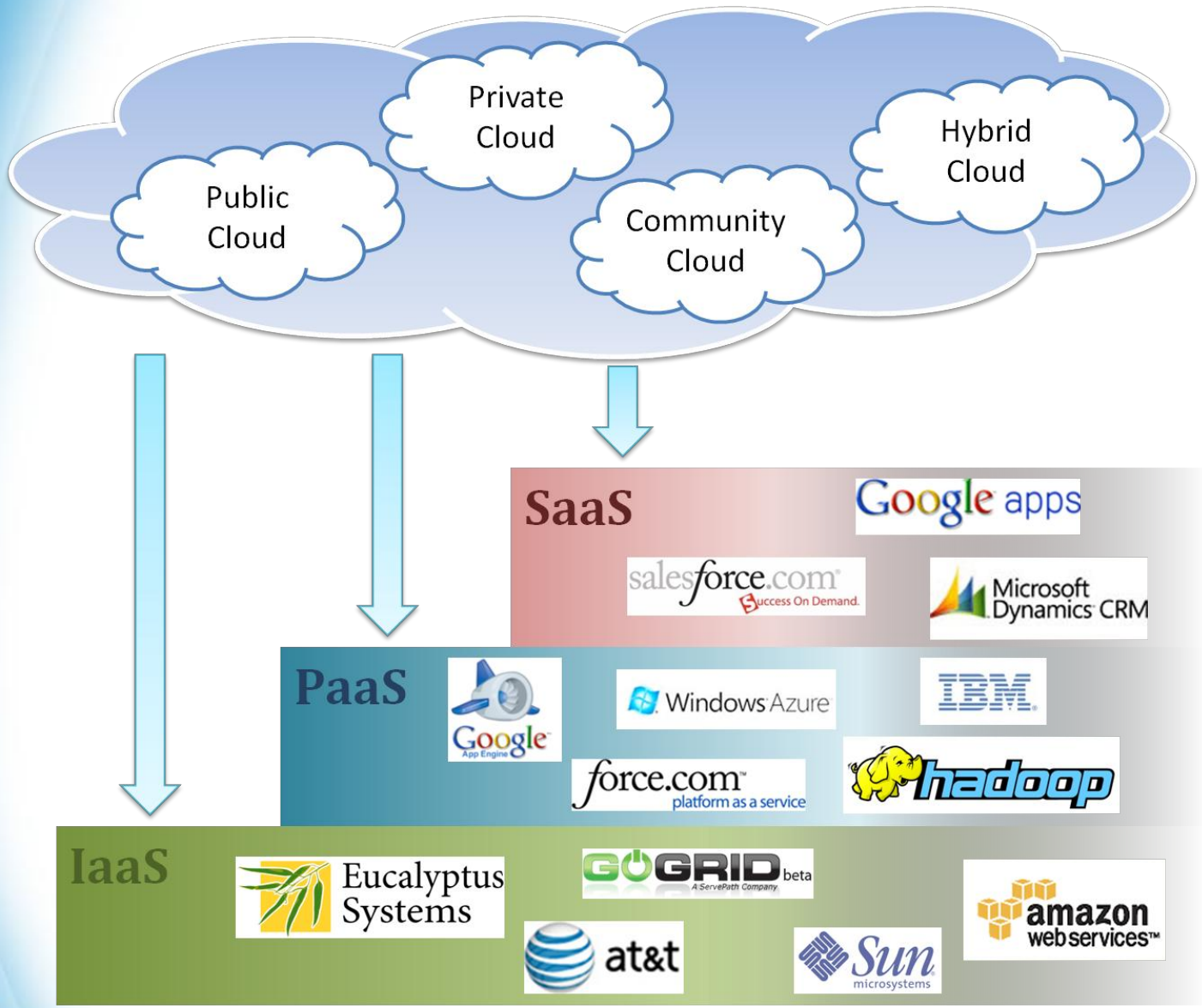


Hybrid Cloud

- Hybrid cloud definition
  - The cloud infrastructure is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load-balancing between clouds).



Cloud Ecosystem



Summary

- What is cloud computing in your mind
  - Clear or Cloudy?

- Cloud computing is a new paradigm shift of computing
- Cloud computing can provide high quality of properties and characteristics based on essentially central ideas

- Service models and deployment models provide services that can be used to
  - Rent fundamental computing resources
  - Deploy and develop customer-created applications on clouds
  - Access provider's applications over network (wired or wireless)

References

- NIST (National Institute of Standards and Technology). http://csrc.nist.gov/groups/SNS/cloud-computing/
- M. Armbrust et. al., “Above the Clouds: A Berkeley View of Cloud Computing,” Technical Report No. UCB/EECS-2009-28, University of California at Berkeley, 2009.
- R. Buyya et. al., “Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility,” Future Generation Computer Systems, 2009.
- Cloud Computing Use Cases. http://groups.google.com/group/cloud-computing-use-cases
- Cloud Computing Explained. http://www.andyharjanto.com/2009/11/wanted-cloud-computing-explained-in.html
- From Wikipedia, the free encyclopedia
- All resources of the materials and pictures were partially retrieved from the Internet.