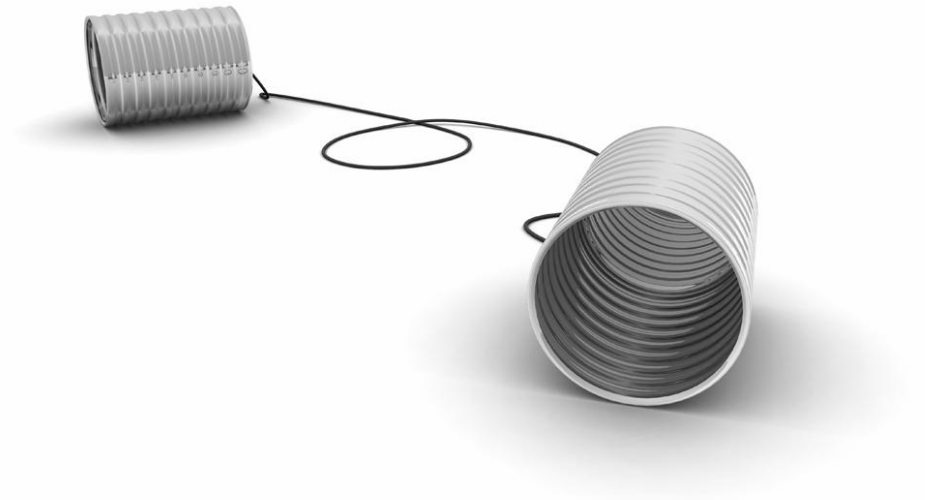




Cloud Computing

**Introduction to Cloud
Computing**



WHAT IS CLOUD COMPUTING ?

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.



National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

Cloud Definitions

- Definition from *Wikipedia*

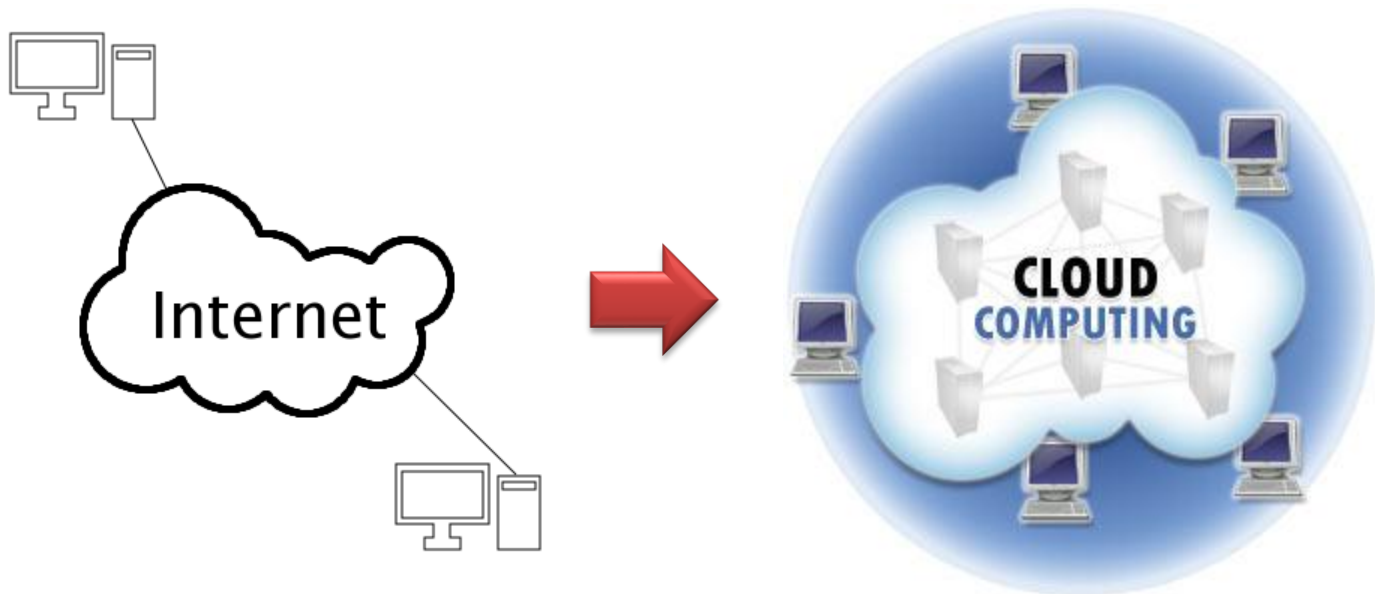
- Cloud computing is **Internet-based computing**, whereby **shared resources, software, and information** are provided to computers and other devices **on demand**, like the electricity grid.
- Cloud computing is a style of computing in which **dynamically scalable** and often **virtualized resources**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



WIKIPEDIA
The Free Encyclopedia

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.**



Cloud Definitions

- Definition from *Berkeley*

- Cloud Computing refers to both the applications **delivered as services over the Internet** and the hardware and systems software **in the data centers that provide those services**.
- The services themselves have long been referred to as **Software as a Service (SaaS)**, so we use that term. The hardware and software **what will call a Cloud**.
- When a Cloud is made available in a **pay-as-you-go** manner to the public..... The service being sold is **Utility Computing**.

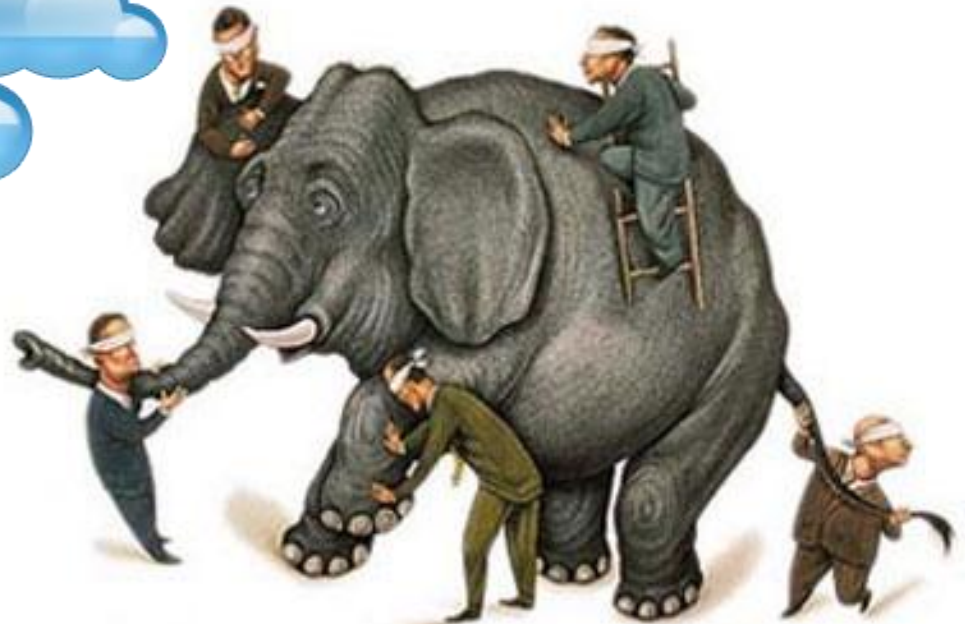


Cloud Definitions

- Definition from *Buyya*

- A Cloud is a type of **parallel and distributed system** consisting of a collection of **interconnected and virtualized computers** that are **dynamically provisioned** and presented as **more or more unified computing resources** based on **service-level agreements** established through negotiation between the service provider and consumers.





Properties and characteristics

Properties and Characteristics

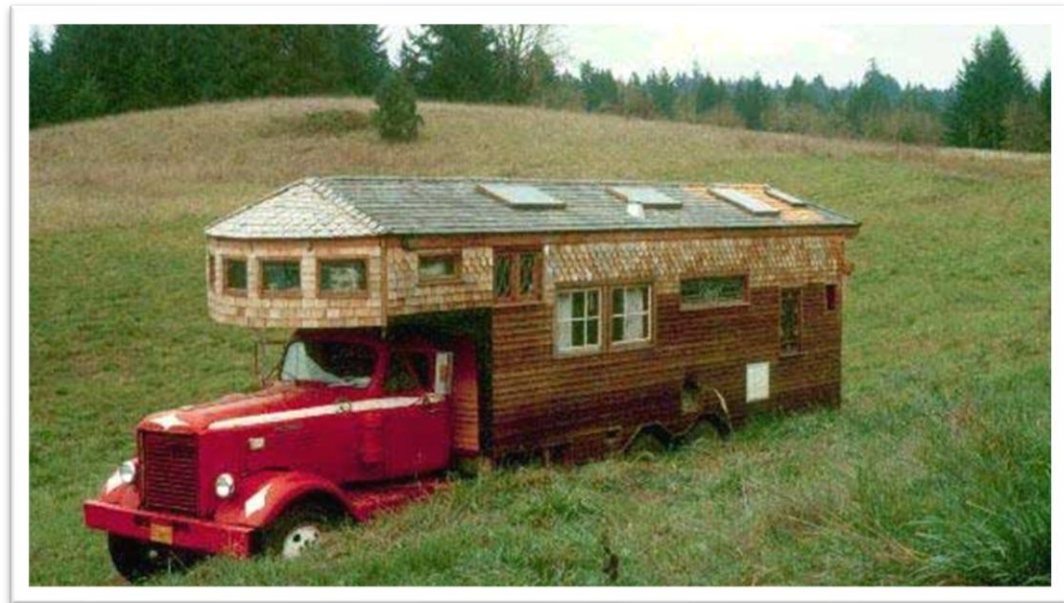


Scalability & Elasticity

- 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

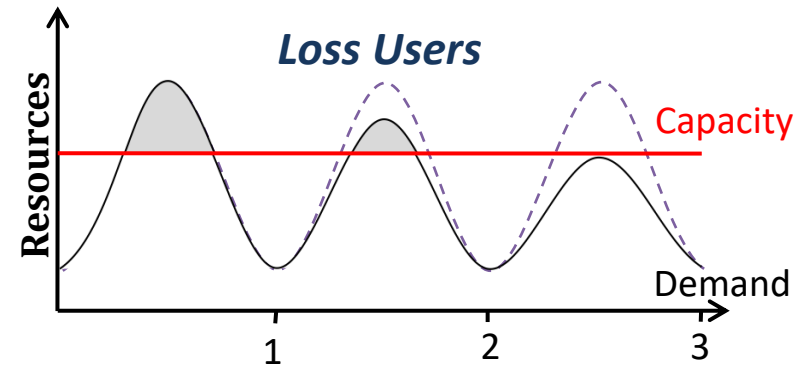
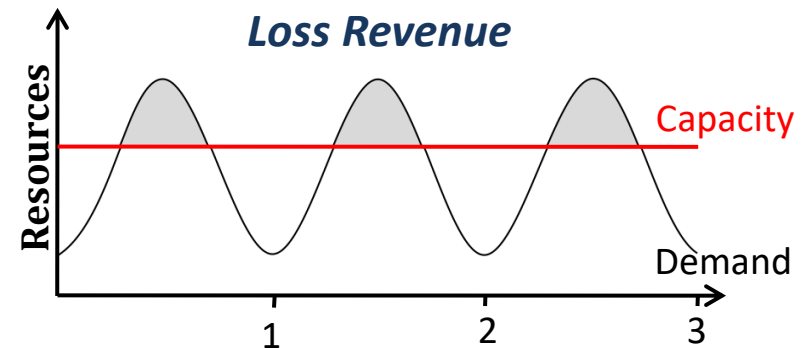
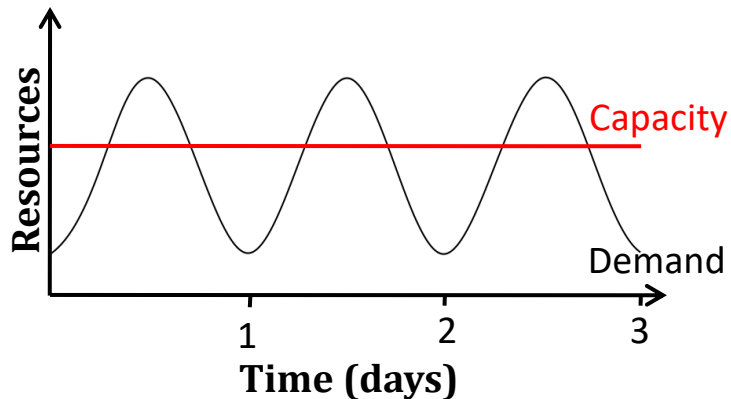
Dynamic Provisioning

- 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.



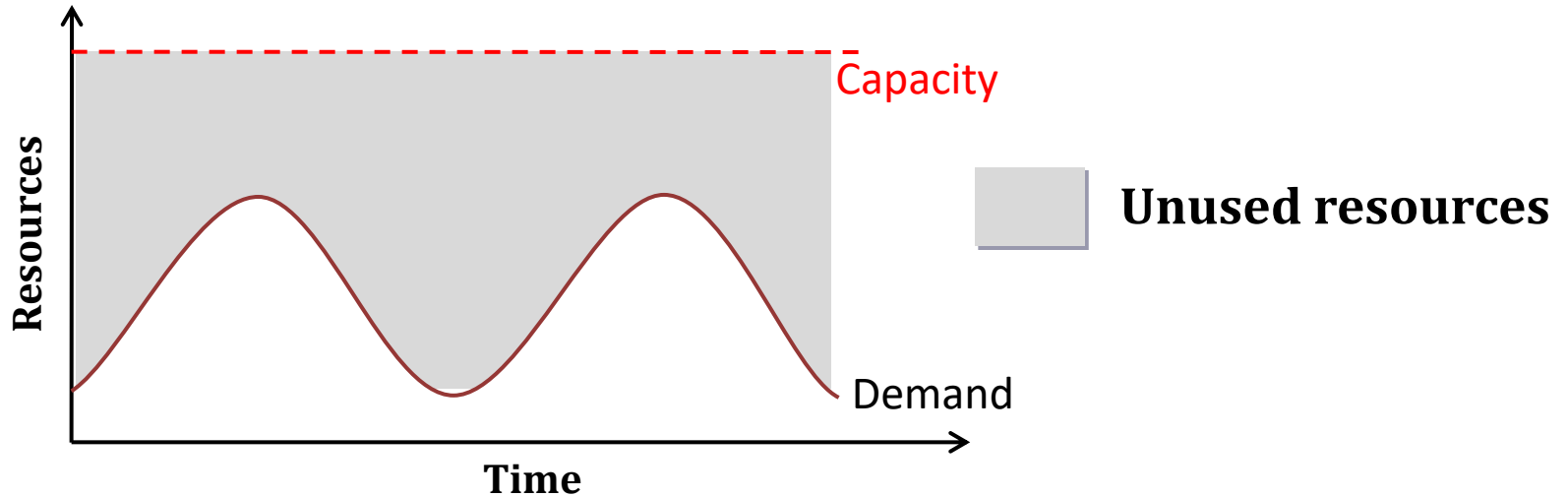
Dynamic Provisioning

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



Dynamic Provisioning

- Overestimate system utilization which result in low utilization



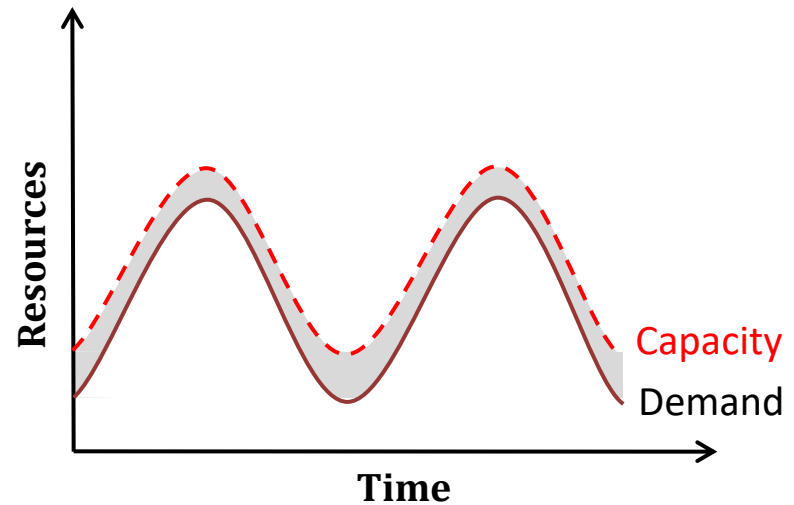
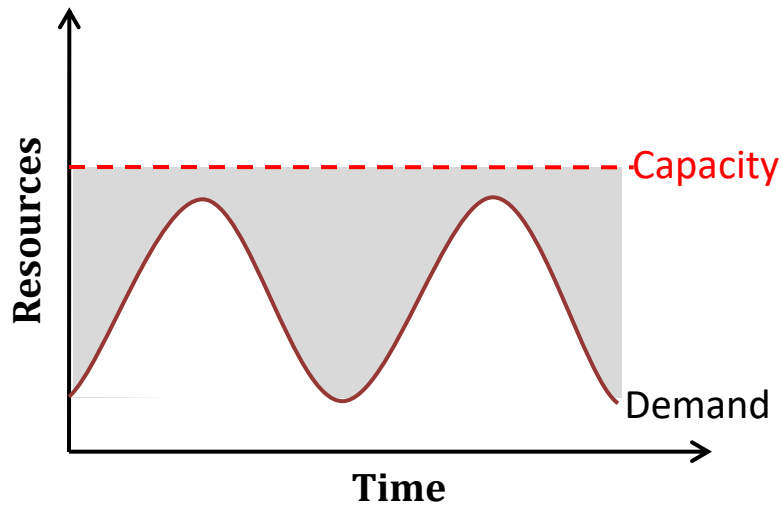
- How to solve this problem ??
 - Dynamically provision resources

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

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.



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 transmitter or failing component to protect the system are required.

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 corruption, 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

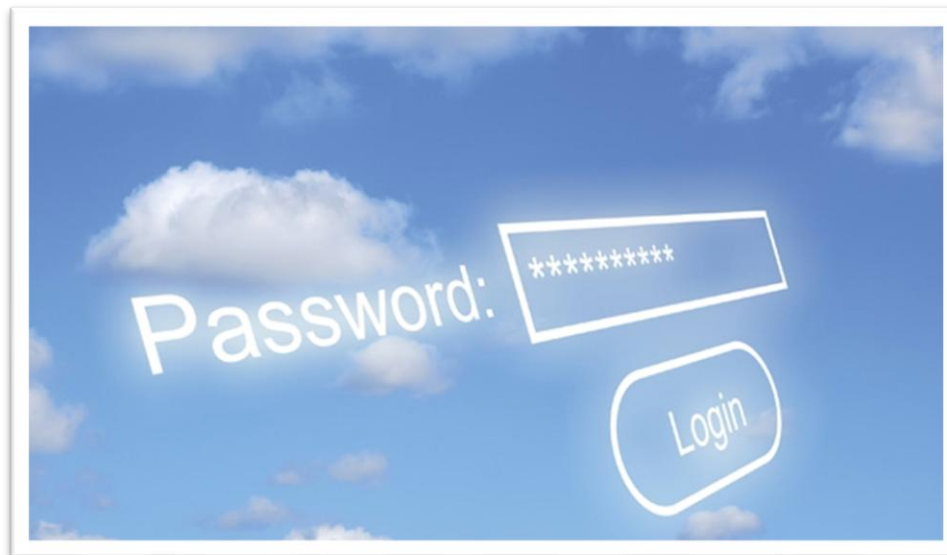
- Preparing

- Local mirror systems
- Uninterruptible Power Supply (UPS)



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.

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.