Cloud Computing

Dr. Bibhudatta Sahoo

Communication & Computing Group

Department of CSE, NIT Rourkela

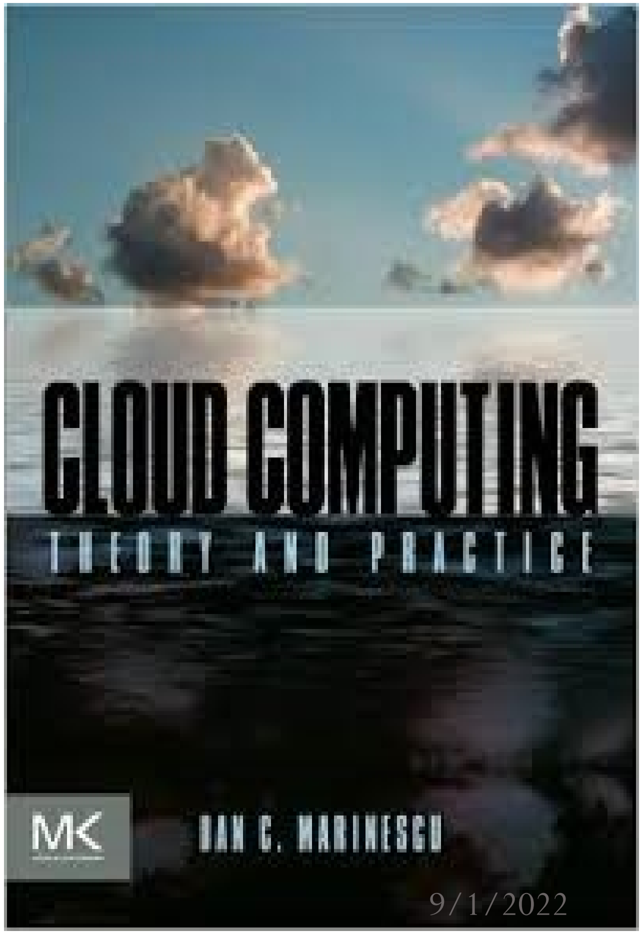
Email: bdsahu@nitrkl.ac.in, 9937324437, 2462358

Cloud Computing: Theory and Practice

Dan C. Marinescu
Computer Science Division
Department of Electrical Engineering & Computer Science
University of Central Florida, Orlando, FL 32816, USA
Email: dcm@cs.ucf.edu

November 10, 2012

Dan C. Marinescu *was a professor of computer science at Purdue University from 1984 to 2001. Then he joined the Computer Science Department at the University of Central Florida. He has held visiting faculty positions at the IBM T. J. Watson Research Center, the Institute of Information Sciences in Beijing, the Scalable Systems Division of Intel Corp., Deutsche Telecom AG and INRIA Rocquancourt in France. His research interests cover parallel and distributed systems, cloud computing, scientific computing, quantum computing, and quantum information theory.*



Chapter 1

- 1.1 Network-centric computing and network-centric content
- 1.2 Peer-to-peer systems
- 1.3 Cloud computing
- 1.4 Cloud computing delivery models and services
- 1.5 Ethical issues in cloud computing
- 1.6 Cloud vulnerabilities
- 1.7 Major challenges faced by cloud computing .

Network-centric Computing (NCC)

Network-centric content

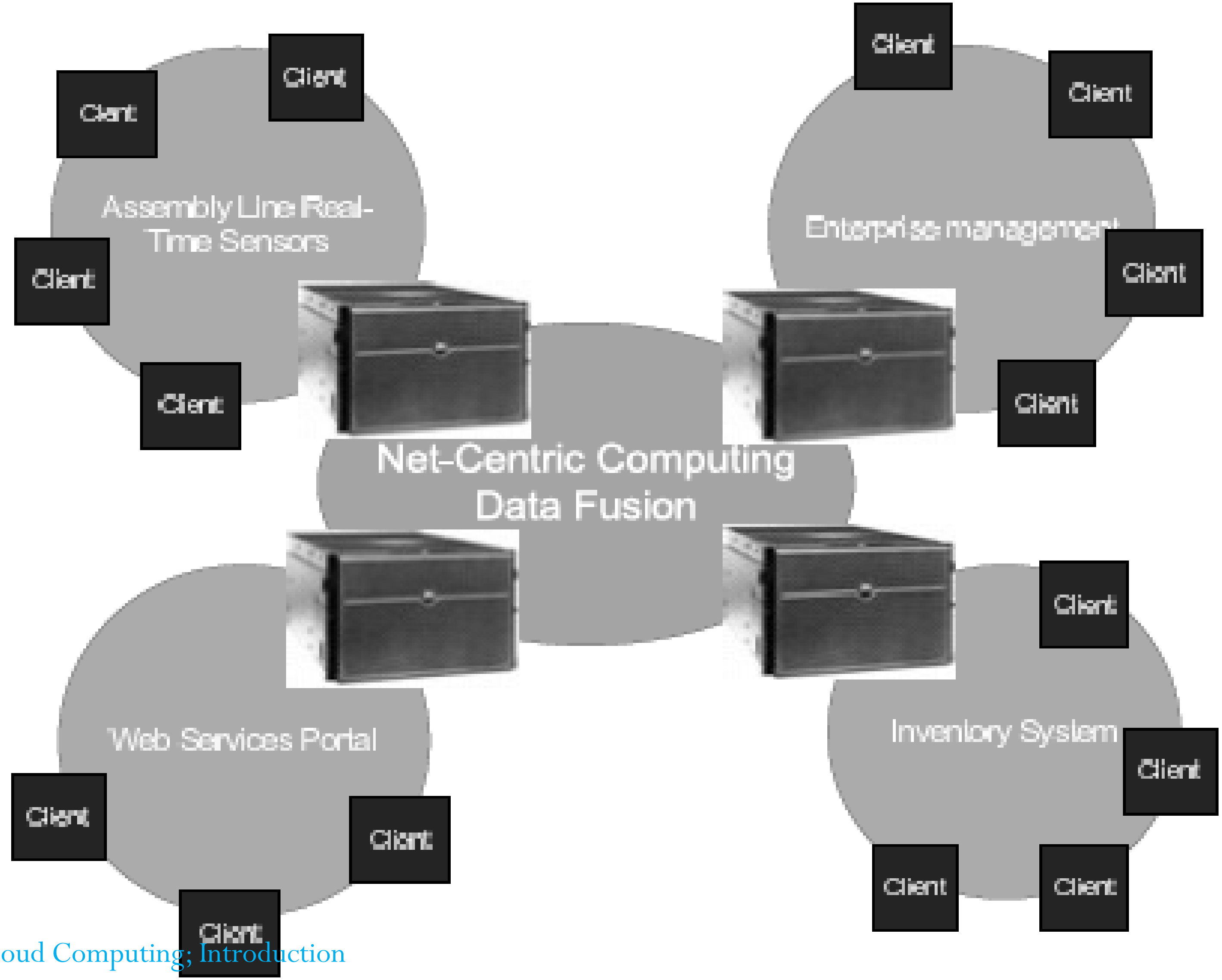
Network-centric Computing (NCC)

- The underlying principle of Net-Centric Computing (NCC) is a distributed environment where applications and data are downloaded from servers and exchanged with peers across a network on as as-needed basis.
- The web browser is an **example** of networked **application**.
- A networked **application** uses the **application** layer protocols such as HTTP, SMTP, and FTP to communicate with servers and other **applications**.
- Many times a protocol and its **application** is so closely tied with each other that it is difficult to separate them.
- NCC covers wide-variety of software engineering researchers and practitioners, in part because it is an enabling technology for modern distributed systems (e.g., Web applications) .

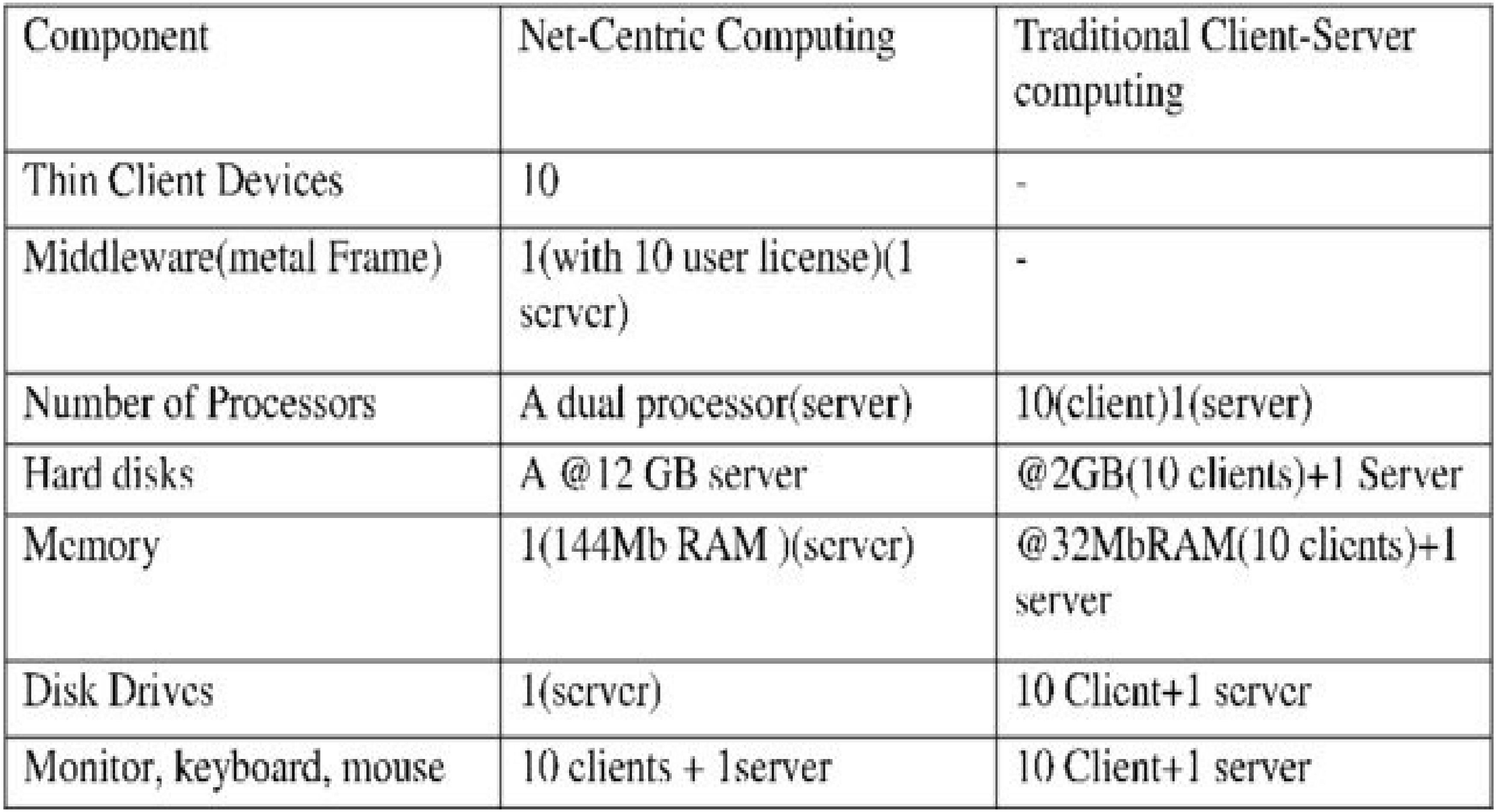
Three layers of Net-Centric Technology

- The **Information Service Layer** pertains to the abstraction of objects. The focus here is on the quality, security, auditability and control.
- The **Network Service Layer** pertains to all aspects of communications, particularly configuration, fault management, security, performance and accounting.
- The **Component Control Layer** pertains to the development, acquisition and implementation of components that form the infrastructure for distributed computing.

Net-centric Computing and Data Fusion



Hardware requirements in Net-Centric Computing



Component	Net-Centric Computing	Traditional Client-Server computing
Thin Client Devices	10	-
Middleware(metal Frame)	1(with 10 user license)(1 server)	-
Number of Processors	A dual processor(server)	10(client)1(server)
Hard disks	A @12 GB server	@2GB(10 clients)+1 Server
Memory	1(144Mb RAM)(server)	@32MbRAM(10 clients)+1 server
Disk Drives	1(server)	10 Client+1 server
Monitor, keyboard, mouse	10 clients + 1server	10 Client+1 server

Net-centric Computing and Data Fusion

- The convergence of computing power, communications capability and content the information, data or knowledge that forms the ”stuff” of the solution with the heart of the solution is network-centric, or ”netcentric,” solutions.
- Net-centric computing refers to an emerging technology architecture and an evolutionary stage of client/server computing.
- The common architecture of netcentric computing supports access to information through multiple electronic channels (personal and network computers, cell phones, kiosks, telephones, etc.).
- This information is made accessible to many more users not just an organization’s workforce but also its customers, suppliers and business partners through technologies that employ open, commonly accepted standards (Internet, Java, Web browsers and so forth).

Network-centric content

- The term ***content*** *refers to any type or volume of media,* be it static or dynamic, monolithic or modular, live or stored, produced by aggregation, or mixed.
- *Information is the result of functions applied to content. The creation and* consumption of audio and visual content is likely to transform the Internet to support increased quality in terms of resolution, frame rate, color depth, stereoscopic information and it seems reasonable to assume that the Future Internet will be content-centric.

Network-centric content

- The content should be treated as having meaningful semantic connotations rather than a string of bytes; the focus will be the information that can be extracted by content mining when users request named data and content providers publish data objects.
- Content-centric routing will allow users to fetch the desired data from the most suitable location in terms of network latency or download time.
- There are also some challenges, such as providing secure services for content manipulation, ensuring global rights-management, control over unsuitable content, and reputation management.

Common characteristics Network-centric computing and network-centric contents

- Most applications are data intensive. Computer simulation becomes a powerful tool for scientific research in virtually all areas of science from physics, biology, and chemistry, to archeology. The widespread use of sensors contribute to the increase of the volume of data. **Multimedia applications** are increasingly more popular; the ever larger media increase the load placed on storage, networking, and processing systems.
- *Virtually all applications are network-intensive. Indeed, transferring large volumes of* data requires high bandwidth networks; parallel computing, computation steering, and data streaming are examples of applications that can only run efficiently on low latency networks.

Common characteristics Network-centric computing and network-centric contents

- *The systems are accessed using thin clients running on systems with limited resources.*
- The infrastructure supports some form of workflow management. The complex computational tasks require coordination of several applications; composition of services is a basic tenet of Web 2.0.

Challenges from network-centric computing and network-centric content paradigms

Shared resource management of Computing and communication resources (CPU cycles, storage, network bandwidth) to support data-intensive applications.

1. Multiplexing leads to a higher resource utilization; indeed, when multiple applications share a system their peak demands for resources are not **synchronized** and the average system utilization increases.
2. The **management of large pools of resources** poses new challenges as complex systems are subject to phase transitions.
3. New resource management strategies, such as self-organization, and decisions based on approximate knowledge of the state of the system must be considered.
4. Ensuring **Quality of Service (QoS)** guarantees is extremely challenging in such environments, as total performance isolation is elusive.

Challenges from network-centric computing and network-centric content paradigms

Data sharing facilitates collaborative activities.

1. Many applications in science, engineering, as well as, industrial, financial, governmental applications require multiple types of analysis of shared data sets and multiple decisions carried out by groups scattered around the globe.
2. Open software development sites are another example of such collaborative activities. Data sharing poses not only security and privacy challenges but also requires mechanisms for access control for authorized users and for detailed logs of the history of data changes.

Challenges from network-centric computing and network-centric content paradigms

- **Cost reduction**. Concentration of resources creates the opportunity to pay-as-you-go for computing and thus, eliminates the initial investment and reduces significantly the maintenance and operation costs of the local computing infrastructure.
- **User convenience and elasticity**, *the ability to accommodate workloads with very large* peak-to-average ratios.

Peer-to-peer systems (P2P)

Peer-to-peer systems (P2P)

- In a P2P network, the "peers" are computer systems which are connected to each other via the Internet.
- Files can be shared directly between systems on the network without the need of a central server. In other words, each computer on a P2P network becomes a file server as well as a client.
- A computer can join a peer-to-peer network with an Internet connection and P2P software.
- Once connected to the network, P2P software allows you to search for files on other people's computers. Meanwhile, other users on the network can search for files on your computer, but typically only within a single folder that you have designated to share.
- While P2P networking makes file sharing easy and convenient, is also has led to a lot of software piracy and illegal music downloads. Therefore, it is best to be on the safe side and only download software and music from legitimate websites.

Peer-to-peer systems (P2P)

- Peer-to-peer architecture (P2P architecture) is a commonly used computer networking architecture in which each workstation, or node, has the same capabilities and responsibilities. It is often compared and contrasted to the classic client/server architecture, in which some computers are dedicated to serving others.

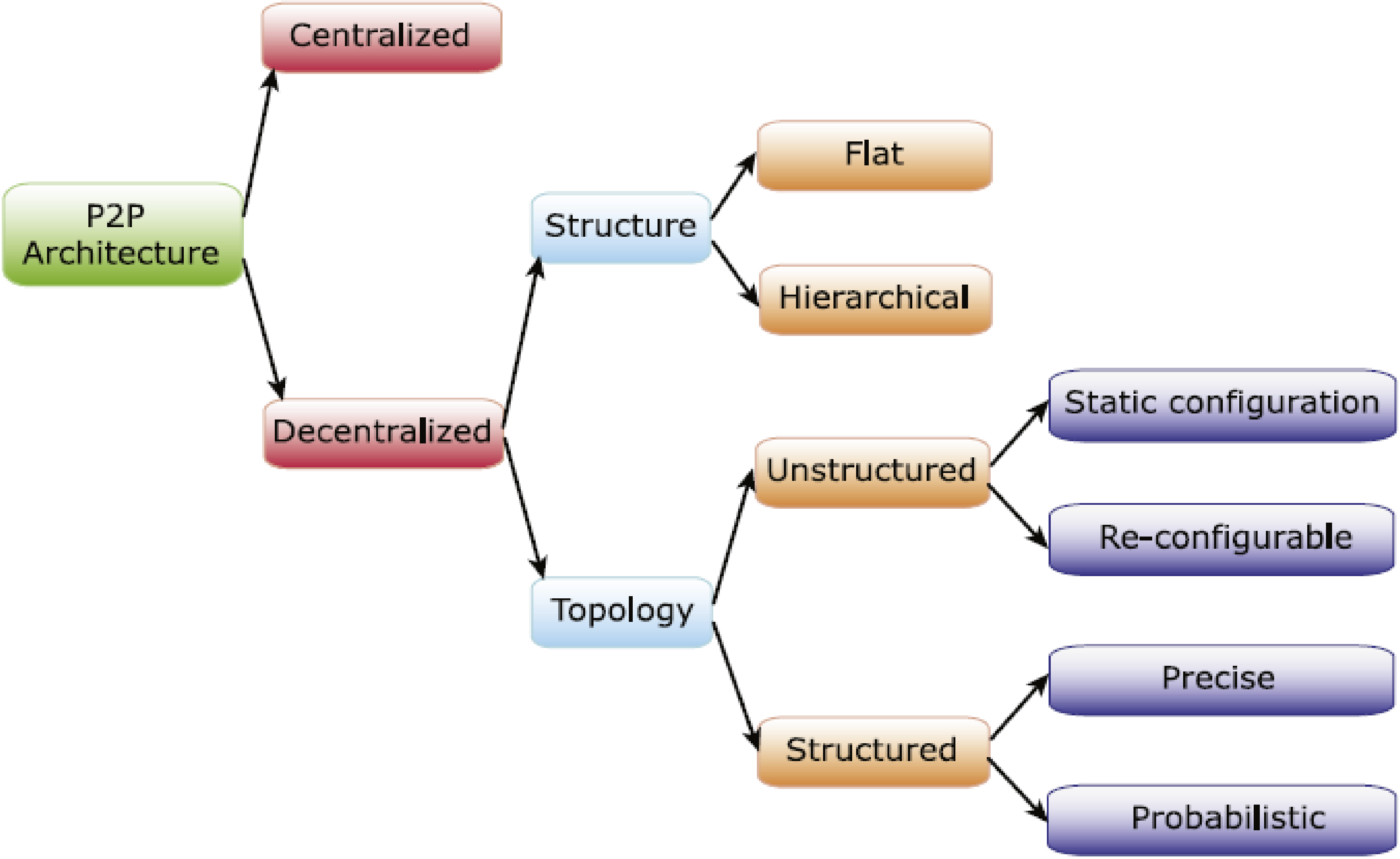
- P2P may also be used to refer to a single software program designed so that each instance of the program may act as both client and server, with the same responsibilities and status.

Peer-to-peer systems (P2P)

- P2P networks have many applications, but the most common is for content distribution. This includes software publication and distribution, content delivery networks, streaming media and peercasting for multicasting streams, which facilitates on-demand content delivery. Other applications involve science, networking, search and communication networks.

- P2P architecture is often referred to as a peer-to-peer network.

A taxonomy of P2P systems

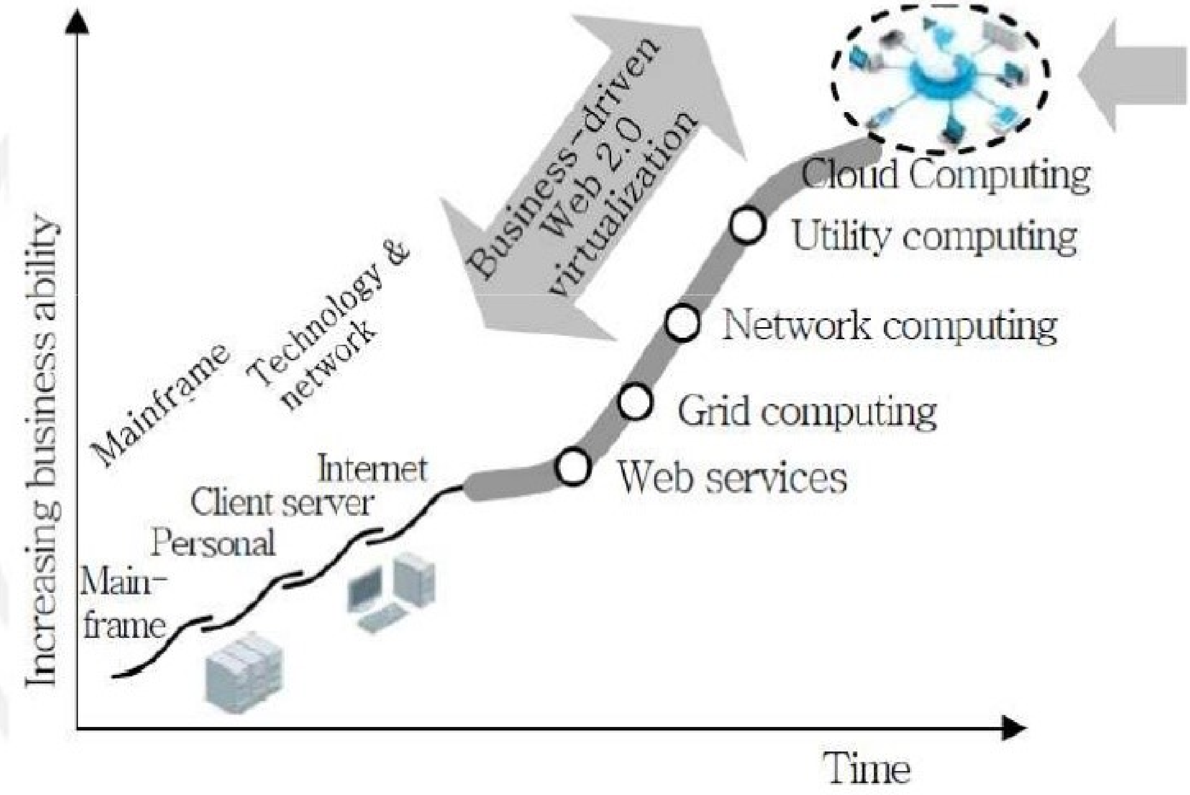


Desirable properties of P2P systems

- P2P systems require a minimally dedicated infrastructure, as resources are contributed by the participating systems;
- P2P systems *are highly decentralized;*
- P2P systems are scalable, the individual nodes are not required to be aware of the global state;
- P2P systems *are resilient to faults and attacks, as few of their elements are critical for the delivery* of service and the abundance of resources can support a high degree of replication;
- P2P systems *individual nodes do not require excessive network bandwidth as servers used in case* of the client-server model do; and last but not least,
- *the systems are shielded from censorship due to the dynamic and often unstructured* system architecture.

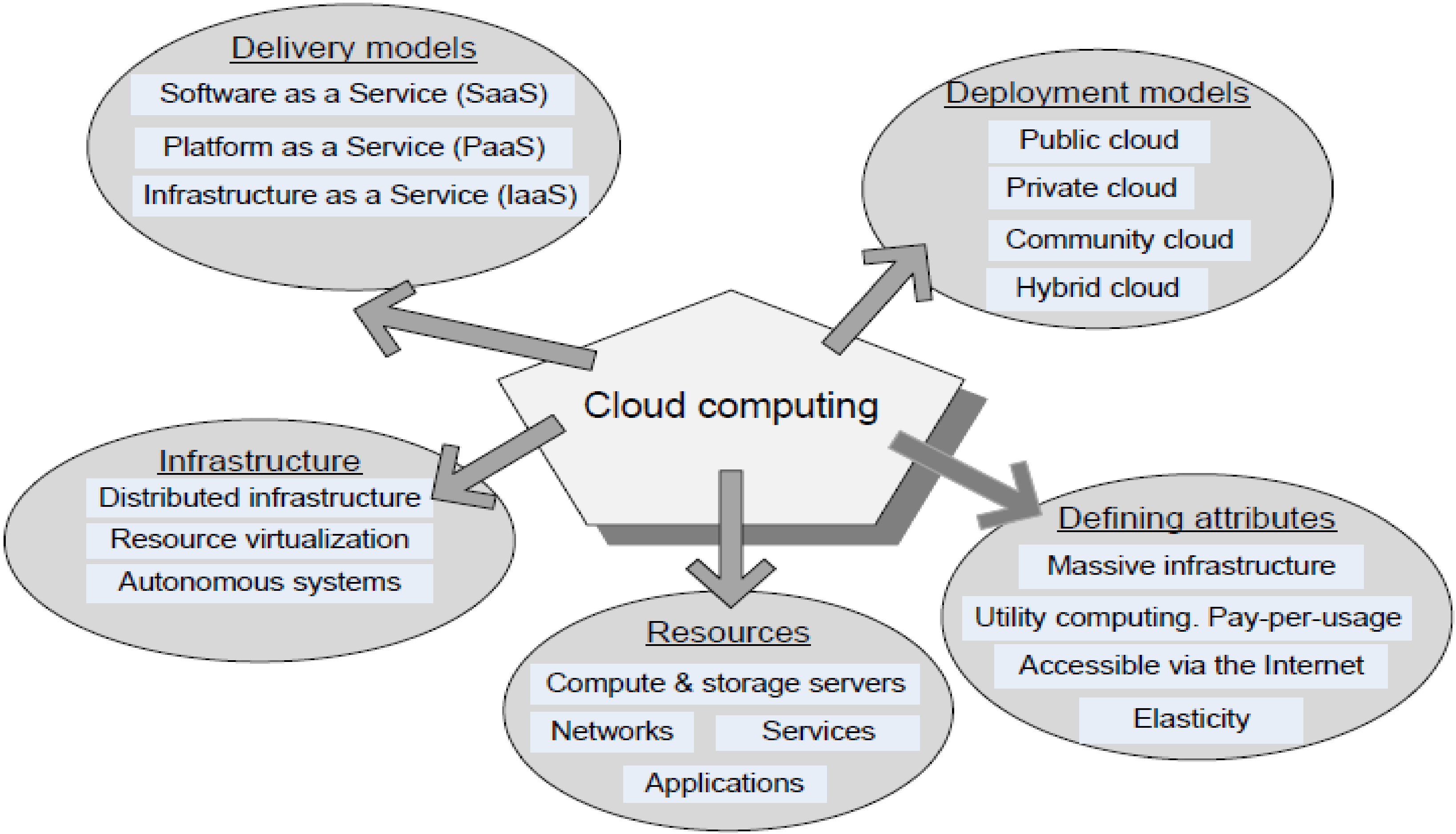
Cloud Computing : Introduction

History of Computer



IEEE – Cloud Computing IT Telkom

The new philosophy for delivering computing services: Cloud computing



The new philosophy for delivering computing services

- Cloud computing uses Internet technologies to offer elastic services.
- The term "elastic computing" refers to the ability of dynamically acquiring computing resources and supporting a variable workload. A cloud service provider maintains a massive infrastructure to support elastic services.
- *The resources used for these services can be metered and the users can be charged* only for the resources they used.
- *The maintenance and security are ensured by service providers.*

The new philosophy for delivering computing services

- *Economy of scale allows service providers to operate more efficiently due to specialization* and centralization.
- *Cloud computing is cost-effective due to resource multiplexing; lower costs for the* service provider are passed on to the cloud users.
- *The application data is stored closer to the site where it is used in a device and location independent* manner; potentially, this data storage strategy increases reliability and security and, at the same time, it lowers communication costs

Cloud Computing

Cloud computing

- A model for enabling service users to have ubiquitous, convenient and 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. Cloud computing enables cloud services.
- NOTE - It is considered from a telecommunication perspective that users are not buying resources but cloud services that are enabled by cloud computing environments.
- The cloud computing model promotes availability and is composed of five essential characteristics, five cloud service categories and four deployment models.

Cloud service

- **Cloud service**: A service that is delivered and consumed on demand at any time, through any access network, using any connected devices using cloud computing technologies.
- **Cloud services** means **services** made available to users on demand via the Internet from a **cloud computing** provider's servers as opposed to being provided from a company's own on-premises servers.
- A cloud service is any resource that is provided over the Internet. The most common cloud service resources are Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS).

The Open Group Architecture Framework

- **The Open Group Architecture Framework (TOGAF)** is a framework for enterprise architecture that provides an approach for designing, planning, implementing, and governing an enterprise information technology architecture. TOGAF has been a registered trademark of **The Open Group** in the United States and other countries since 2011.
- TOGAF is a high level approach to design. It is typically modeled at four levels: Business, Application, Data, and Technology. It relies heavily on modularization, standardization, and already existing, proven technologies and products.

http://www.opengroup.org

The Open Group Cloud Ecosystem Reference Model

- To develop, manage, and govern an Enterprise Architecture of a fictitious organization named "CloudEcoSource" with use of the Cloud Ecosystem Reference Model and the TOGAF standard.
- It describes an approach for each phase of the TOGAF Architecture Development Method (ADM) and what the architect should consider when looking to apply the TOGAF ADM to an enterprise Cloud Ecosystem.
- Tis informative case study describes a generic approach to develop an Enterprise Architecture by utilizing Architecture Building Blocks (ABBs) identified in the Cloud Ecosystem Reference Model.
- The standard defines an approach to identify Solution Building Blocks (SBBs) to address the architectural capabilities of ABBs.

http://www.opengroup.org/cloud/cloud/cloud_ecosystem_rm/togaf.htm

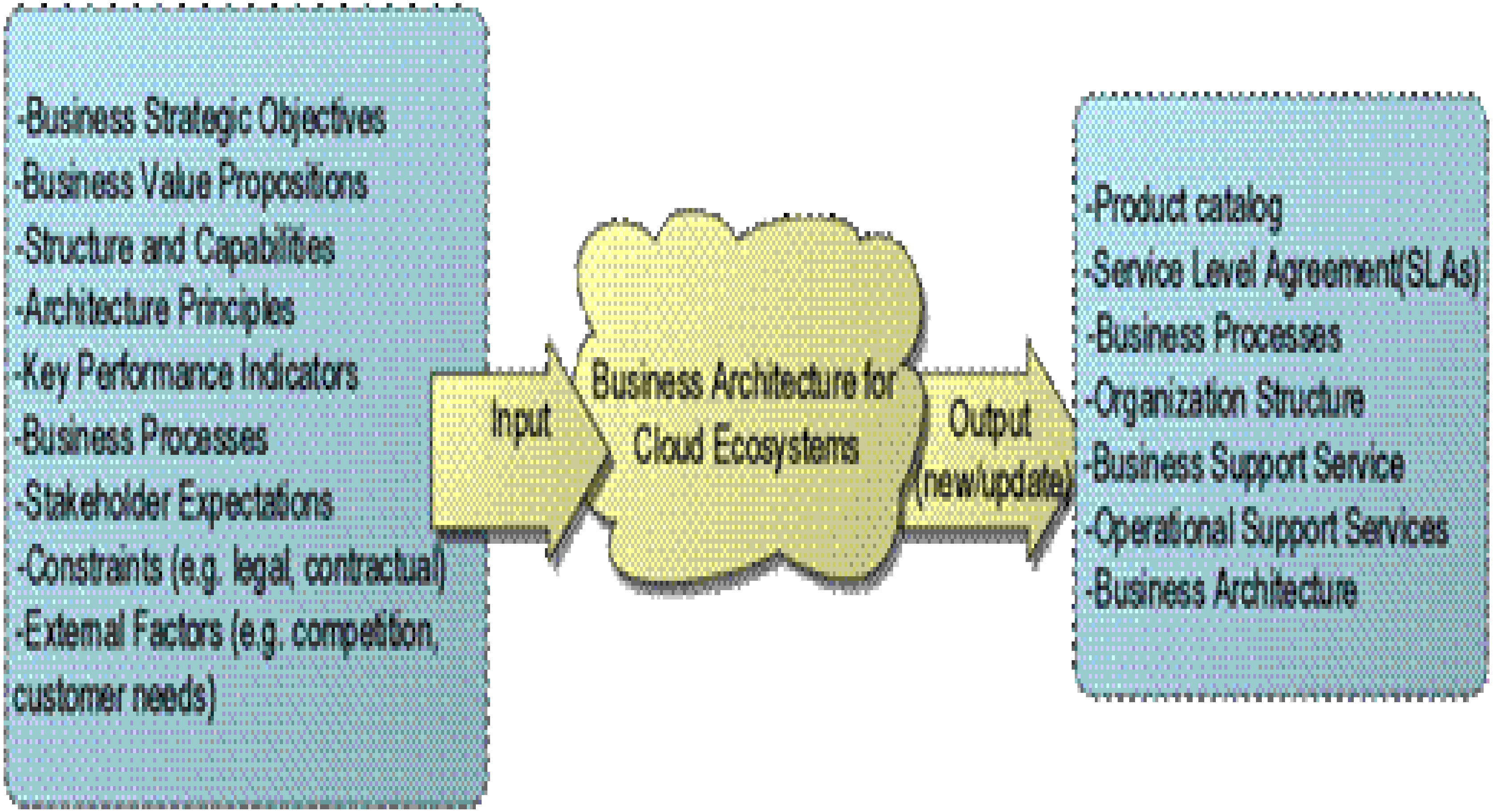
A brief description of the cloud-specific initiatives of CloudEcoSource

- CloudEcoSource plans to use the TOGAF standard to create and evolve an Enterprise Architecture for the Cloud Ecosystem. :
- **IaaS initiative**: *Infrastructure modernization and consolidation* – An IaaS-focused initiative of CloudEcoSource with the expectations on how to transform and regulate dynamic resources consumption in a multi-tenant infrastructure environment with effective management of privacy of its tenants (i.e., enterprise customers).
- **PaaS initiative**: *Rapid application development platform* – A PaaS-focused initiative to identify and describe architectural capabilities of a platform for CloudEcoSource business solutions. The instances of the platforms could be deployed and operated either solely by an enterprise or by participating entities of the Cloud Ecosystem.
- **SaaS initiative**: *Enhanced collaborations among multiple service providers* – A SaaS-focused initiative where CloudEcoSource is assembling business capabilities for business collaborations that extend the organization's traditional enterprise application boundaries with extended users (both internal and external to the organization).

Cloud ecosystem actors

- **Cloud service user (CSU):** A person or organization that consumes delivered cloud services.
- NOTE – A CSU can include intermediate users that will deliver cloud services provided by a cloud service provider (CSP) to actual users of the cloud service, i.e. end users. End users can be persons, machines, or applications.
- **Cloud service provider (CSP):** An organization that provides and maintains delivered cloud services.
- **Cloud service partner (CSN):** A person or organization that provides support to the building of the service offer of a cloud service provider (e.g. service integration).

The expected input and output of the Business Architecture phase from a Cloud Ecosystem perspective



-Business Strategic Objectives
-Business Value Propositions
-Structure and Capabilities
-Architecture Principles
-Key Performance Indicators
-Business Processes
-Stakeholder Expectations
-Constraints (e.g. legal, contractual)
-External Factors (e.g. competition, customer needs)

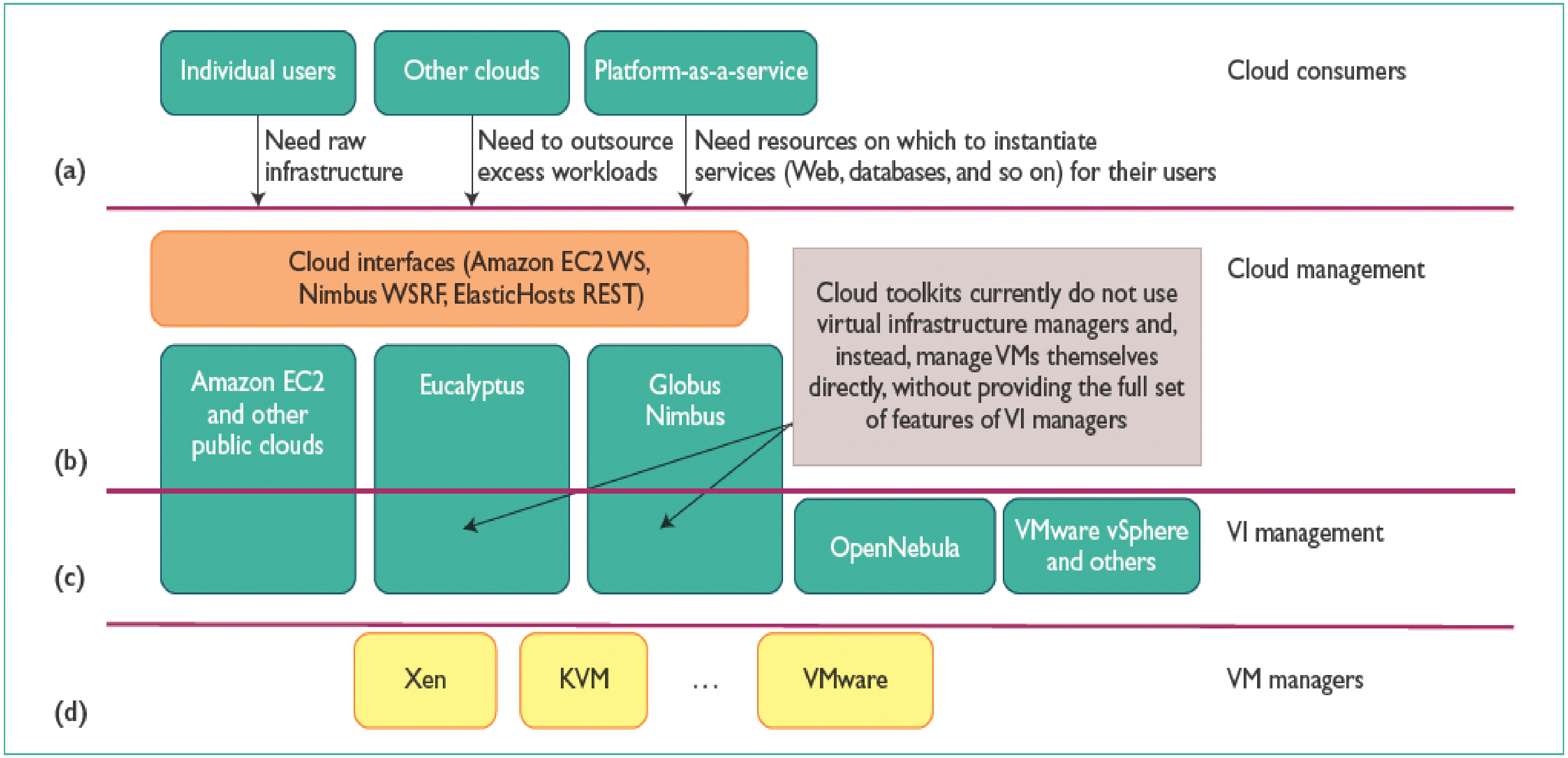
Input

Business Architecture for Cloud Ecosystems

Output (new/update)

-Product catalog
-Service Level Agreement(SLAs)
-Business Processes
-Organization Structure
-Business Support Service
-Operational Support Services
-Business Architecture

Cloud Ecosystem



Individual users
Other clouds
Platform-as-a-service
Cloud consumers
(a)
Need raw infrastructure
Need to outsource excess workloads
Need resources on which to instantiate services (Web, databases, and so on) for their users
Cloud interfaces (Amazon EC2WS, Nimbus WSRF, ElasticHosts REST)
Cloud management
Cloud toolkits currently do not use virtual infrastructure managers and, instead, manage VMs themselves directly, without providing the full set of features of VI managers
Amazon EC2 and other public clouds
Eucalyptus
Globus Nimbus
(b)
OpenNebula
VMware vSphere and others
VI management
(c)
Xen
KVM
...
VMware
VM managers
(d)

The cloud ecosystem for building private clouds. (a) Cloud consumers need flexible infrastructure on demand. (b) Cloud management provides remote and secure interfaces for creating, controlling, and monitoring virtualized resources on an infrastructure-as-a-service cloud. (c) Virtual infrastructure (VI) management provides primitives to schedule and manage VMs across multiple physical hosts. (d) VM managers provide simple primitives (start, stop, suspend) to manage VMs on a single host.

Cloud Ecosystem

- The public cloud ecosystem has evolved around providers, users, and technologies.
- The previous figure suggests one possible ecosystem for private clouds. There are 4 levels of development of ecosystem development: cloud users/consumers, cloud management, VI management, and VM managers.
- At the cloud management level, the cloud manager provides virtualized resources over an IaaS platform.
- At the virtual infrastructure (VI) management level, the manager allocates VMs over multiple server clusters. Examples: OpenNebula, VMWare vSphere. These can manage VM managers like Xen, KVM etc. These support dynamic placement and VM management on a pool of physical resources, automatic load balancing, server consolidation, and dynamic infrastructure resizing and partitioning.

Cloud Ecosystem

- At the VM management level the VM managers handles VMs installed on individual host machines. Examples: Xen, VMWare, KVM.
- An ecosystem of cloud tools attempts to span both cloud management and VI management. Besides public clouds such as Amazon EC2, open source cloud tools for virtualization of cloud infrastructure include Eucalyptus and Globus Nimbus.
- To access these cloud tools, one can use the Amazon EC2WS interface among others.

.

Cloud computing essential characteristics

Cloud computing essential characteristics

- **On-demand self-service**: A cloud service user can unilaterally provision computing capabilities, such as server time, network storage and communication and collaboration services, as needed automatically without requiring human interaction with each service's cloud service provider.
- **Broad network access:** Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, laptops, and PDAs).

Cloud computing essential characteristics

- **Resource pooling:** The cloud service provider's computing resources are pooled to serve multiple users using a multi-tenant model, with different physical and virtual resources that are dynamically assigned and reassigned according to user demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources, but may be able to specify the location at a higher level of abstraction (e.g., country, state, data centre). Examples of resources include storage (typically on hard or optical disc drives), processing, memory (typically on DRAM), network bandwidth, and virtual machines.

Cloud computing essential characteristics

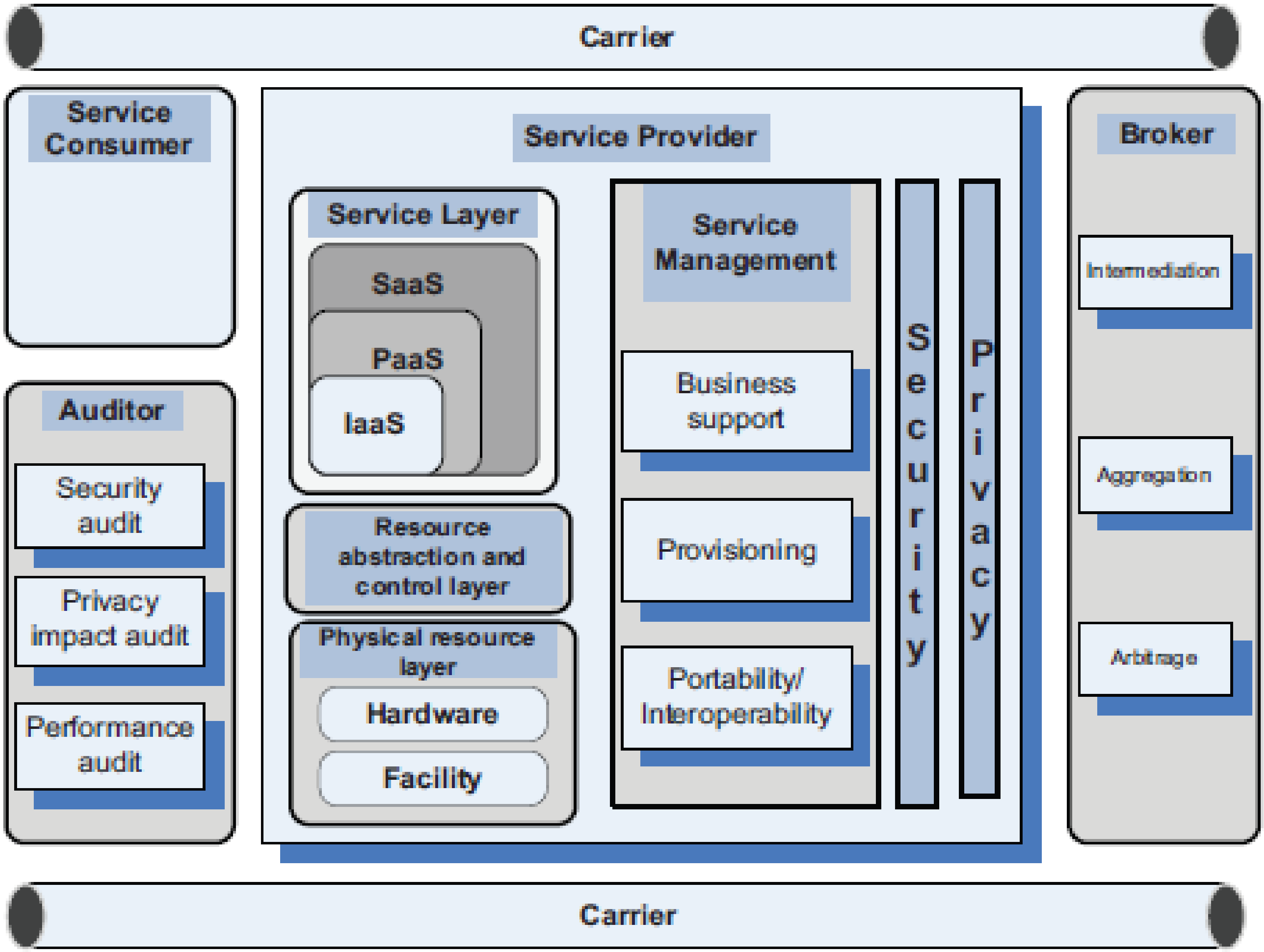
- **Rapid elasticity:** Capabilities can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out, and rapidly released to quickly scale in. To the cloud service user, the capabilities available for provisioning often appear to be unlimited and can be purchased in any quantity at any time.
- **Measured service:** Cloud systems automatically control and optimize resource use (e.g., storage, processing and bandwidth) by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., the number of active user accounts). Resource usage can be monitored, controlled, and reported, providing transparency for both the cloud service provider and cloud service user of the utilized service.

Cloud computing delivery models and services

The entities involved in Cloud computing delivery models

- ***service consumer*** *- entity that maintains a business relationship with, and* uses service from, service providers;
- ***service provider*** *- entity responsible for making a service* available to service consumers; *carrier - the intermediary that provides connectivity* and transport of cloud services between providers and consumers;
- ***broker*** *- an entity that* manages the use, performance and delivery of cloud services, and negotiates relationships between providers and consumers;
- ***auditor*** *- a party that can conduct independent assessment* of cloud services, information system operations, performance and security of the cloud implementation.

Cloud computing delivery models and services



Cloud computing delivery models and services

- An *audit is a systematic evaluation of a cloud system by measuring how* well it conforms to a set of established criteria.
  - *Example, a security audit evaluates cloud security, a privacy-impact audit evaluates the cloud privacy assurance, while a performance audit evaluates the cloud performance*

- The carrier provides connectivity between service providers, service consumers, brokers, and auditors.
  - *Note that it is difficult to distinguish the services associated with cloud computing from those that any computer operations center. While many of the services discussed in this section could be provided by a cloud architecture, they are available in non-cloud architectures as well.*

Cloud service categories

Defining Cloud (IBM)

- **Computing as a service** over the Internet
- Cloud computing, often referred to as simply “the cloud,” is the delivery of on-demand computing resources, everything from applications to data centres, over the Internet on a pay-for-use basis.

Cloud software as a service (SaaS):

- **A category of cloud services where the capability provided to the cloud service user is to use the cloud service provider's applications running on a cloud infrastructure.**
- *NOTE - All applications have the common characteristic to be non-real-time and may be of different kinds, including IT and business applications, and may be accessible from different user devices. The cloud service user does not manage or control the underlying cloud infrastructure, with the possible exception of limited user-specific application configuration settings.*

Communications as a service (CaaS):

- A category of cloud services where the capability provided to the cloud service user is to use real-time communication and collaboration services.
- *NOTE - Communication and collaboration services include voice over IP, instant messaging, and video conferencing, for different user devices.*

Cloud platform as a service (PaaS):

- A category of cloud services where the capability provided to the cloud service user is to deploy user-created or acquired applications onto the cloud infrastructure using platform tools supported by the cloud service provider.
- *NOTE - platform tools may include programming languages and tools for application development, interface development, database development, storage and testing. The cloud service user does not manage or control the underlying cloud infrastructure, but has control over the deployed applications and, possibly, over the application hosting environment configurations.*

Cloud infrastructure as a service (IaaS):

- A category of cloud services where the capability provided by the cloud service provider to the cloud service user is to provision processing, storage, intra-cloud network connectivity services (e.g. VLAN, firewall, load balancer, and application acceleration), and other fundamental computing resources of the cloud infrastructure where the cloud service user is able to deploy and run arbitrary application.
- *NOTE - The cloud service user does not manage or control the resources of the underlying cloud infrastructure but has control over operating systems, deployed applications, and possibly limited control of select networking components (e.g., host firewalls).*

Network as a service (NaaS):

- A category of cloud services where the capability provided to the cloud service user is to use transport connectivity services and/or inter-cloud network connectivity services.
- *NOTE - NaaS services include flexible and extended VPN, bandwidth on demand, etc.*

Sensing as a service (S^{2}aaS),

- Sensing as a service (S^{2}aaS), i.e., providing sensing services using mobile phones via a cloud computing system. An S^{2}aaS cloud needs to meet the following requirements:
  - *1) it must be able to support various mobile phone sensing applications on different smartphone platforms;*
  - *2) it must be energy-efficient; and*
  - *3) it must have effective incentive mechanisms that can be used to attract mobile users to participate in sensing activities.*
- *NOTE: Sensors on (or attached to) mobile phones can enable attractive sensing applications in different domains, such as environmental monitoring, social networking, healthcare, transportation, etc.*

Delivery models in Cloud

- A *cloud delivery model* represents a specific, pre-packaged combination of IT resources offered by a cloud provider. Three common cloud delivery models have become widely established and formalized:

  Infrastructure-as-a-Service (IaaS)

  Platform-as-a-Service (PaaS)

  Software-as-a-Service (SaaS)

Other delivery models in Cloud

- Many specialized variations of the three base cloud delivery models have emerged, each comprised of a distinct combination of IT resources. Some examples include:

  Storage-as-a-Service

  Database-as-a-Service

  Security-as-a-Service

  Communication-as-a-Service

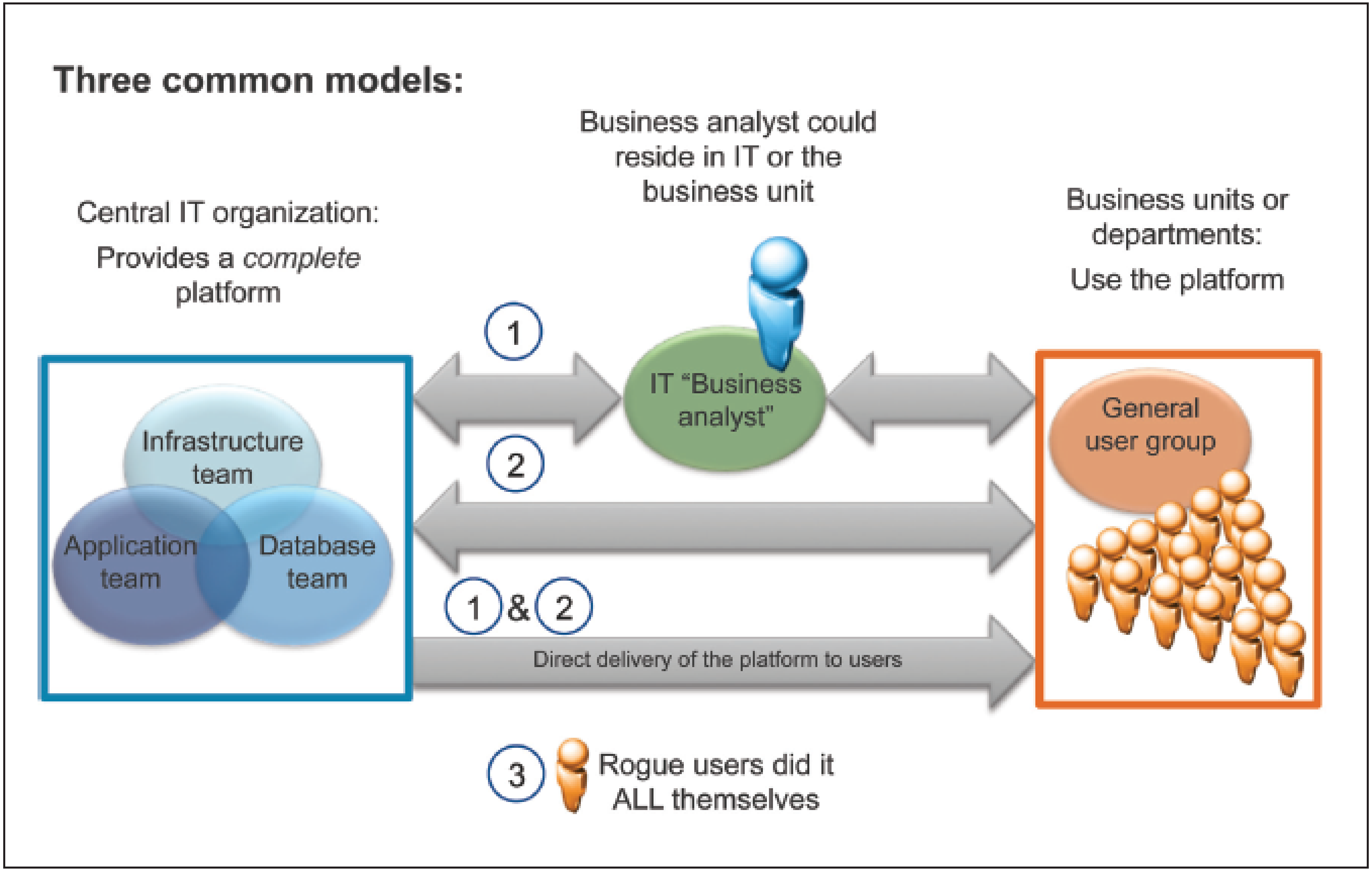
  Integration-as-a-Service

  Testing-as-a-Service

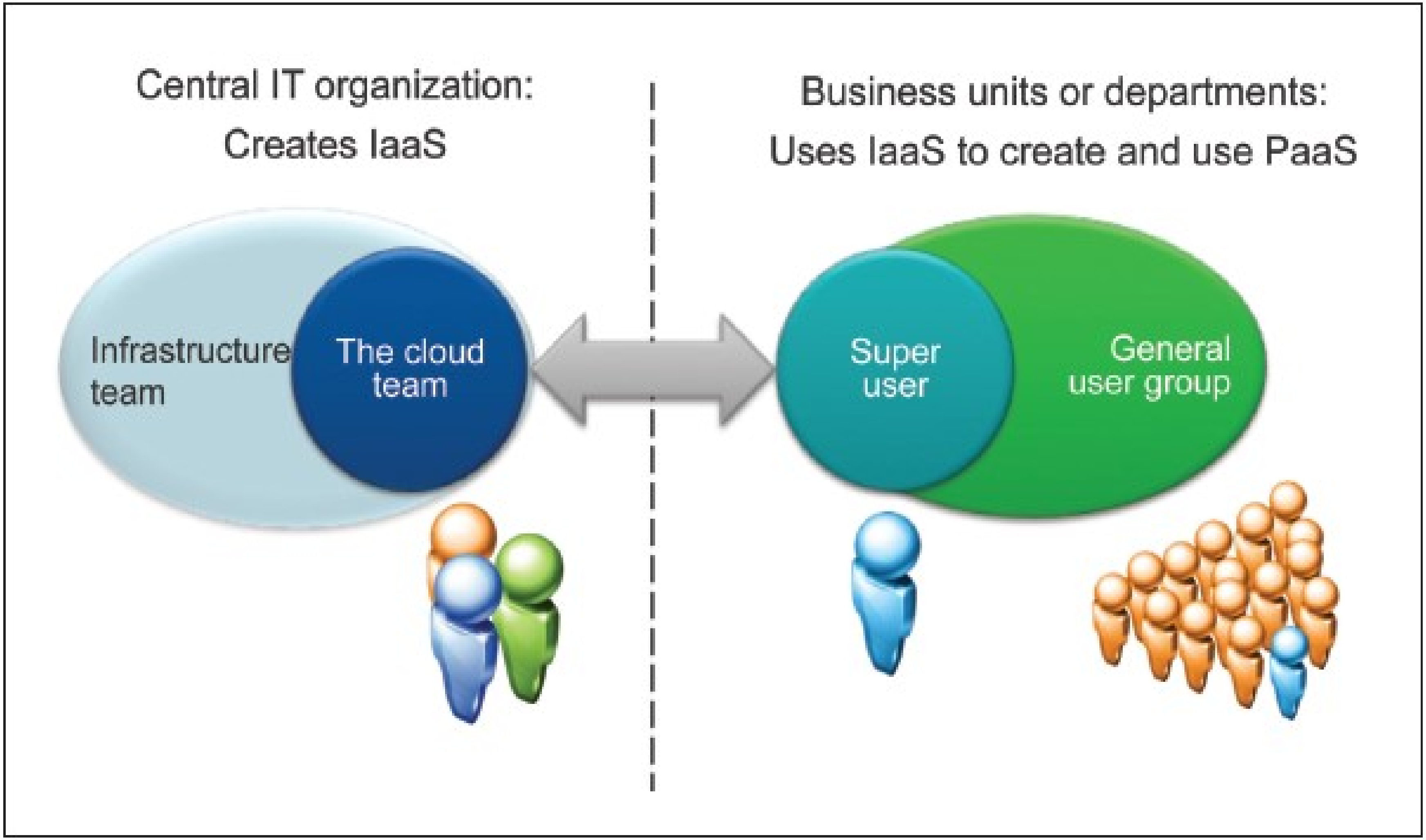
  Process-as-a-Service

- *NOTE also that a cloud delivery model can be referred to as a cloud service delivery model because each model is classified as a different type of cloud service offering.*

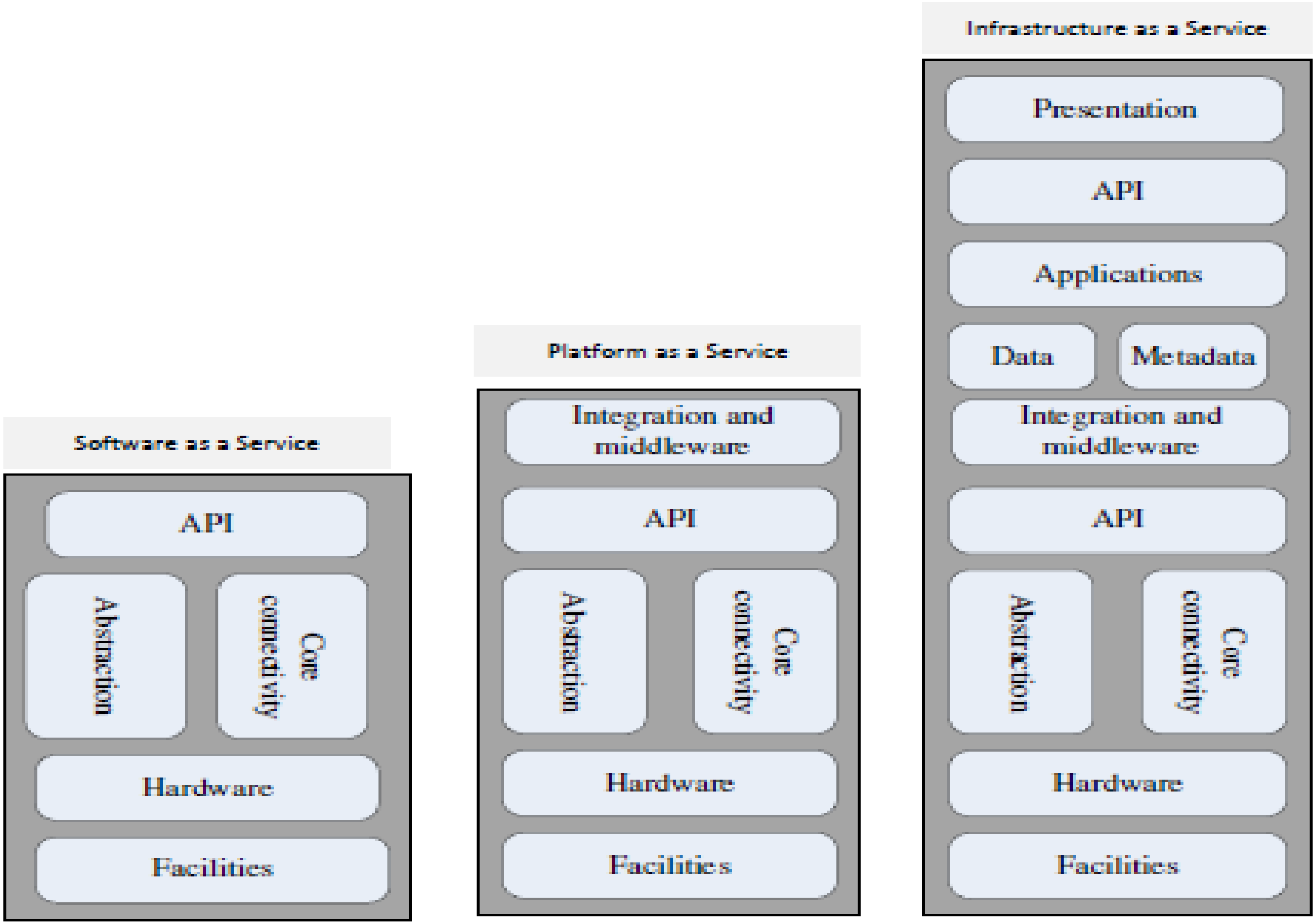
The traditional service delivery model before the cloud


Three common models:
Business analyst could reside in IT or the business unit
Central IT organization:
Provides a *complete* platform
Business units or departments:
Use the platform
1
IT "Business analyst"
2
Infrastructure team
Application team
Database team
General user group
1 & 2
Direct delivery of the platform to users
3
Rogue users did it ALL themselves

The cloud service delivery model



The structure of the three delivery models



Cloud computing services (IBM)

- **Software as a service (SaaS):** Cloud-based applications, or software as a service (SaaS) run on distant computers "in the cloud" that are owned and operated by others and that connect to users' computers via the Internet and, usually, a web browser.
- **Platform as a service (PaaS):** Platform as a service provides a cloud-based environment with everything required to support the complete lifecycle of building and delivering web-based (cloud) applications, without the cost and complexity of buying and managing the underlying hardware, software, provisioning and hosting.
- **Infrastructure as a service (IaaS):** Infrastructure as a service provides companies with computing resources including servers, networking, storage, and data centre space on a pay-per-use basis.

The activities are necessary to support the three delivery models.

1. Service management and provisioning including: virtualization, service provisioning, call center, operations management, systems management, QoS management, billing and accounting, asset management, SLA management, technical support, and backups.
2. Security management including: ID and authentication, certification and accreditation, intrusion prevention, intrusion detection, virus protection, cryptography, physical security, incident response, access control, audit and trails, and firewalls.
3. Customer services such as: customer assistance and on-line help, subscriptions, business intelligence, reporting, customer preferences, and personalization.
4. Integration services including data management and development.

Ethical issues in cloud computing

Ethical issues in cloud computing : Why ?

Cloud Computing:

- Users **outsource** their computing needs to third parties, over the Internet
- The control shifts from users to third parties
- Multiple services can be **interconnected** to provide a specific service
- Data in **multiple physical locations** around the world, possibly owned and administered by different organizations
- All this **complexity is hidden** from the user

Control and Responsibility

- Users relinquish **control** over computation and data
- Unauthorized access, data corruption, who is **responsible**?
- **Deperimeterisation**: disappearing of boundaries between systems and organizations
- The border between organization + infrastructure blurs, but also the **accountability**
- Problem of **many hands**, service oriented architecture

Function creep

- Data collected or a specific purpose can be used for **other purposes**
- A database with biometric data for authentication can be used for crime investigation
- **Unimplementing** might become difficult because of wide scale use

Privacy

- There is a consensus that it is important, the concept is hard to **explicate**
- Aim to constraint access to **certain types** of personal data. Which types? Conception differs per **type of data/context**
- Different opinions of privacy by the service providers
- Different layers/service providers with different policies

Privacy across borders and diversity

- Legislation differences, i.e.: Facebook case in Germany, Google Maps in China
- Cultural differences: emphasis on the concept of community and negative concept of privacy in Eastern cultures (Capurro, 2005)
- A minimal sense of privacy is shared, but an internationally accepted rich sense is lacking (Moor, 2004)
- Convergences of values and norms do take place, i.e. incorporation of traditional Chinese values and Western values (Yao-Huai, 2005)
- An opportunity to take pluralistic ground and avoid relativism. Globalization can play a role
- But.. Risk of cultural imperialism.
- Do not impose values, but bridge cultures. Ethics can play an important role in reaching the middle ground (Moor, 2005)

Precautionary Principle

- Precautionary principle: refrain from actions in the face of scientific uncertainties about serious or irreversible harm
- In software engineering ethics: do not abort the development of the technology, but anticipate consequences that are not foreseeable (Pieters 2009)
- Uncertainty is no excuse not to do this
- Technical standardizations, professional, national and international law and regulations must follow

Future for Cloud Computing Ethics: Value Sensitive Design

- Include moral values of ethical importance in design
- Uses empirical studies, interviews with stakeholders to include their views into the design
- Conceptualize the values
- Translate it into technical design
- Why?
  - *Design is about changing the world, inherently normative*
  - *Designers have been doing it all the time: but make it more explicit, transparent and systematic*
  - *Design for X: Design for maintainability, Design for reliability, etc.*

Future for Cloud Computing Ethics: Value Sensitive Design

- **Use is important** : Same technology in different contexts realizes different values
- **Design is important as well** : Differently designed technologies (with same function) in same user context realize different values.
- Deal with **value trade-offs** (privacy vs accountability, trust vs security), and include **user views** into the design by means of empirical investigations
- **Discover** values, **translate** them into design, **verify**

Cloud vulnerabilities

Cloud vulnerabilities

- Clouds are affected by malicious attacks and failures of the infrastructure, e.g., power failures. Such events can affect the Internet domain name servers and prevent access to a cloud or can directly affect the clouds.

Cloud vulnerabilities

- **Session Riding:** Session riding occurs when an online attacker steals an internet user's cookie to use the application later as the real user. The attackers might also use the Cross-Site Request Forgery (CSRF) attacks for them to trick the user to send authentic requests to random websites to accomplish various missions.
- **Virtual Machine Escape:** Within virtualized settings, the physical servers operate multiple virtual apparatuses on top of the hypervisors. An online attacker can remotely exploit a hypervisor by using a weakness present in that particular hypervisor. However, such vulnerabilities are pretty rare, but they are real. Also, a virtual machine can avoid the virtualized sandbox setting to gain access to the hypervisor. Consequently, all the virtual machines ultimately run on the virtual machine.

Cloud vulnerabilities

- **Unsafe Cryptography:** Cryptography algorithms normally use random number generators. They use unpredictable information sources to produce actual random numbers that are needed to get a large entropy pool. When the random number generators provide only a limited entropy pool, the numbers can be forced. In a client's computer, the major source of randomization is user mouse operations and the key presses. Servers however normally operate without user interaction. That consequently means that there will be a lower number of sources for randomization. Hence, the virtual machines usually rely on the sources that are available to them. That could lead to easily guessable numbers that do not give much uncertainty in cryptographic algorithms.
- **CSP Lock-in:** You have to choose a provider that has guarantee cloud security will enable you to shift easily to another provider when necessary. You do not want to choose a CSP that will force you to use its services. That is because sometimes you would prefer to use a CSP in one thing and another CSP for something different

Cloud vulnerabilities

- **Cloud computing threats:** Before you decide to shift to the cloud computing, you have to put into consideration the platform's security vulnerabilities. You also need to assess the possible threats to determine whether the cloud platform is worth the risk due to the numerous advantages it has to offer. The following are the major security threats experienced regarding cloud security.
- **Ease of Use:** It is a reality that cloud computing services can easily be exploited by malicious attackers since its registration process is pretty simple. You are only required to have valid credit card to get started on this platform. In some cases, you can even pay for the cloud computing charges by through PayPal, Payza, Bitcoin, Western Union or Litecoin. By using the payment methods, you can stay completely anonymous. The cloud platform can be used maliciously for various ill purposes like malware distribution, botnet C&C servers, spamming, DDoS, hash cracking and password cracking.

Cloud vulnerabilities

- **Secure Data Transmission:** When sending the data from clients to a cloud computing platform, the data can be transferred by using a secure, encrypted communication channel such as SSL/TLS. That prevents various attacks like the dreaded MITM. During these attacks, your online data could be stolen by an attacker intercepting your communication.
- **Insecure APIs:** Most cloud services are exposed by their application programming interfaces. Since the APIs are easily accessible from any location on the Internet, malicious attackers can exploit them to compromise the integrity and confidentiality of the internet users. An attacker has access to a token used by a legit user to access the service through cloud computing. The API can apply the same token to interfere with the customer data. Hence, it is imperative that all cloud services provide a safe API to prevent such attacks.

Cloud vulnerabilities

- **Malicious Insiders:** It is possible for a staff member at a cloud service provider to have complete access to your confidential resources. Therefore, cloud service providers should set proper security measures to track their employee actions. Normally, cloud service providers never follow the best security procedures and fail to implement security policies. Hence, their employees can collect confidential information from customers without getting detected.

Minimize the Risks of the Cloud

- Carefully consider the type of data to trust to the cloud
- If client data or information is involved, make sure you have exercised reasonable diligence in assessing the security and privacy risks of your chosen cloud provider(s)
- Consider public, private or hybrid cloud options
- Think about obtaining insurance coverage to help protect against the risks
- When possible negotiate key contractual terms with the cloud provider, including:
  - *responsibility for employees and subcontractors*
  - *compliance with applicable laws and regulations*
  - *minimum security standards, including authentication methods*
  - *warranties and service levels with specified remedies*
  - *Indemnifications*
  - *governing law and jurisdiction*

Cloud challenges

Cloud challenges

- Cloud computing inherits some of the challenges of parallel and distributed computing.
- It also faces many major challenges of its own. The specific challenges differ for the three cloud delivery models, but in all cases the difficulties are created by the **very nature of utility** computing, which is based on **resource sharing** and **resource virtualization** and requires a **different trust model** than the ubiquitous user-centric model that has been the standard for a long time.

Cloud challenges

- **Security:** Gaining the trust of a large user base is critical for the future of cloud computing. It's unrealistic to expect that a public cloud will provide a suitable environment for all applications. Highly sensitive applications related to critical infrastructure management, health-care applications and others will most likely be hosted by private clouds.
- Many real-time applications will probably still be confined to private clouds. Some applications may be best served by a hybrid cloud setup. Such applications could keep sensitive data on a private cloud and use a public cloud for some of the processing.

Cloud challenges

- The Software as a Service (SaaS) model faces similar challenges as other online services required to protect private information, such as financial or health-care services. In this case, a user interacts with cloud services through a well-defined interface. In principle, therefore, it's less challenging for the services provider to close some of the attack channels.
- Still, such services are vulnerable to DoS attacks and malicious insiders. Data in storage is most vulnerable to attack, so devote special attention to protecting storage servers. The data replication necessary to ensure continuity of service in case of storage system failure increases vulnerability. Data encryption may protect data in storage, but eventually data must be decrypted for processing. Then it's exposed to attack.

Cloud challenges

- The Infrastructure as a Service (IaaS) model is by far the most challenging to defend against attacks. Indeed, an IaaS user has much more freedom than the other two cloud delivery models. An additional source of concern is that the considerable cloud resources could be used to initiate attacks against the network and the computing infrastructure.
- Virtualization is a critical design option for this model, but it exposes the system to new sources of attack. The trusted computing base (TCB) of a virtual environment includes not only the hardware and the hypervisor but also the management OS. You can save the entire state of a virtual machine (VM) to a file to allow migration and recovery, both highly desirable operations.

Cloud challenges

- Yet this possibility challenges the strategies to bring the servers belonging to an organization to a desirable and stable state. Indeed, an infected VM can be inactive when the systems are cleaned up. Then it can wake up later and infect other systems. This is another example of the deep intertwining of desirable and undesirable effects of basic cloud computing technologies.
- The next major challenge is related to resource management on a cloud. Any systematic (rather than ad hoc) resource management strategy requires the existence of controllers tasked to implement several classes of policies: admission control, capacity allocation, load balancing, energy optimization and, last but not least, the provision of quality of service (QoS) guarantees.

Cloud challenges

- To implement these policies, the controllers need accurate information about the global state of the system. Determining the state of a complex system with 106 servers or more, distributed over a large geographic area, isn't feasible. Indeed, the external load, as well as the state of individual resources, changes very rapidly. Thus, controllers must be able to function with incomplete or approximate knowledge of the system state.
- It seems reasonable to expect that such a complex system can only function based on self-management principles. But self-management and self-organization raise the bar for the implementation of logging and auditing procedures critical to the security and trust in a provider of cloud computing services.

Cloud challenges

- Under self-management it becomes next to impossible to identify the reasons that a certain action that resulted in a security breach was taken.
- The last major challenge I'll address is related to interoperability and standardization. Vendor lock-in—the fact that a user is tied to a particular cloud services provider—is a major concern for cloud users. Standardization would support interoperability and thus alleviate some of the fears that a service critical for a large organization may not be available for an extended period of time.

- Imposing standards at a time when a technology is still evolving is challenging, and it can be counterproductive because it may stifle innovation.
- It’s important to realize the complexity of the problems posed by cloud computing and to understand the wide range of technical and social problems cloud computing raises.
- The effort to migrate IT activities to public and private clouds will have a lasting effect.

Dan C. Marinescu

The data centers

A large group of networked computer servers typically used by organizations for the remote storage, processing, or distribution of large amounts of data.

The data centers

- A data center (sometimes spelled *datacenter*) is a centralized repository, either physical or virtual, for the storage, management, and dissemination of data and information organized around a particular body of knowledge or pertaining to a particular business.
- The terms "cloud" and "data center" may sound like interchangeable technical jargon or trendy buzz words referring to the same infrastructure, but the two computing systems have less in common than the fact that they both store data.

A cloud and a data center

- The main difference between a cloud and a data center is that a cloud is an off-premise form of computing that stores data on the Internet, whereas a data center refers to on-premise hardware that stores data within an organization's local network.
- While cloud services are outsourced to third-party cloud providers who perform all updates and ongoing maintenance, data centers are typically run by an in-house IT department.

Need a cloud or a data center 1/2

- A data center is ideal for companies that need a customized, dedicated system that gives them full control over their data and equipment. Since only the company will be using the infrastructure's power, a data center is also more suitable for organizations that run many different types of applications and complex workloads.
- A data center, however, has limited capacity -- once you build a data center, you will not be able to change the amount of storage and workload it can withstand without purchasing and installing more equipment.

Need a cloud or a data center 2/2

- A cloud system is scalable to your business needs. It has potentially unlimited capacity, based on your vendor's offerings and service plans.
- One disadvantage of the cloud is that you will not have as much control as you would a data center, since a third party is managing the system.
- Furthermore, unless you have a private cloud within the company network, you will be sharing resources with other cloud users in your provider's public cloud.

Cloud security vs. data center security

- The cloud is an external form of computing, it may be less secure or take more work to secure than a data center. Unlike data centers, where you are responsible for your own security, you will be entrusting your data to a third-party provider that may or may not have the most up-to-date security certifications.
- If your cloud resides on several data centers in different locations, each location will also need the proper security measures.
- A data center is also physically connected to a local network, which makes it easier to ensure that only those with company-approved credentials and equipment can access stored apps and information.
- The cloud, however, is accessible by anyone with the proper credentials anywhere that there is an Internet connection. This opens a wide array of entry and exit points, all of which need to be protected to make sure that data transmitted to and from these points are secure.

Cloud vs. data center costs

- For most small businesses, the cloud is a more cost-effective option than a data center. Because you will be building an infrastructure from the ground up and will be responsible for your own maintenance and administration, a data center takes much longer to get started and can cost businesses $10 million to $25 million per year to operate.
- Unlike a data center, cloud computing does not require time or capital to get up and running. Instead, most cloud providers offer a range of affordable subscription plans to meet your budget and scale the service to your performance needs. Whereas data centers take time to build, depending on your provider, cloud services are available for use almost immediately after registration.

Categorizing the data centers

https://uptimeinstitute.com/

The Uptime Institute categorizes the **data centers** by four levels: **Tier** I, **II**, **III** and IV. These levels correspond to a certain number of guarantees on the type of hardware deployed in the **data center** to ensure redundancy. Availability: 99.67%

The Uptime Institute has categorized data centers into four hosting tiers:

- **Tier I Data Centers:** Data centers with Tier I topology offers single uplink and servers, with 99.671% uptime. However, these data centers are non-redundant, catering to basic business demands. As a result, any unwarranted failure in the capacity system thwarts the ongoing performance.
- **Tier II Data Centers:** These modern data centers have single, non-redundant path for power source. Data centers listed with this topology offers redundant capacity components to ascertain smooth access, with 99.741% network uptime.
- **Tier III Data Centers:** Equipped with redundant components with manifold power and cooling options, these data centers can efficiently and expeditiously switch to maintain backup paths ensuring 99.982% network availability.
- **Tier IV data Centers:** These are fault -tolerant **data center** having multiple power and environment control channels with activedata backup options, providing 9.995% network availability.

Categorization of data centers

- **Tier 1**: composed of a single path for power and cooling distribution, without redundant components, providing 99.671% availability.
- **Tier II**: composed of a single path for power and cooling distribution, with redundant components, providing 99.741% availability
- **Tier III**: composed of multiple active power and cooling distribution paths, but only one path active, has redundant components, and is concurrently maintainable, providing 99.982% availability
- **Tier IV**: composed of multiple active power and cooling distribution paths, has redundant components, and is fault tolerant, providing 99.995% availability.

Categorizing the data centers

- **Tier 1** = Non-redundant capacity components (single uplink and servers).
- **Tier 2** = Tier 1 + Redundant capacity components.
- **Tier 3** = Tier 1 + Tier 2 + Dual-powered equipments and multiple uplinks.
- **Tier 4** = Tier 1 + Tier 2 + Tier 3 + all components are fully fault-tolerant including uplinks, storage, chillers, HVAC systems, servers etc. Everything is dual-powered.

Categorizing the data centers

- Tier 1 to 4 data center is a standardized methodology used to define uptime of data center and useful for measuring: Data center performance, Investment, and ROI (return on investment)
- Tier 4 data center considered as **most robust and less prone** to failures. Tier 4 is designed to host mission critical servers and computer systems, with fully redundant subsystems (cooling, power, network links, storage etc) and compartmentalized security zones controlled by biometric access controls methods.
- Naturally, the simplest is a Tier 1 data center used by small business or shops.

References

- http://www.slideshare.net/hamdani2/cloud-ecosystem
- http://whatiscloud.com/cloud_delivery_models/index
- http://www.bsnlcloud.com/pages/Compute_as_a_Service.asp
- http://www.businessnewsdaily.com/4982-cloud-vs-data-center.html
- https://**uptimeinstitute**.com/
- http://www.datasciencecentral.com/profiles/blogs/the-top-cloud-computing-vulnerabilities-and-threats
-

Thanks for Your Attention!

Exercises

Exercises and Problems

Problem 1. Mobile devices could benefit from cloud computing; explain the reasons you think that this statement is true or provide arguments supporting the contrary. Discuss several cloud applications for mobile devices; explain which one of the three cloud computing delivery models, SaaS, PaaS, or IaaS, would be used by each one of the applications and why.

Problem 2. Do you believe that the homogeneity of a large-scale distributed systems is an advantage? Discuss the reasons for your answer. What aspects of hardware homogeneity are the most relevant in your view and why? What aspects of software homogeneity do you believe are the most relevant and why?

Exercises and Problems

Problem 3. Peer-to-peer systems and clouds share a few goals, but not the means to accomplish them. Compare the two classes of systems in terms of architecture, resource management, scope, and security.

Problem 4. Compare the three cloud computing delivery models, SaaS, PaaS, and IaaS, from the point of view of the application developers and users. Discuss the security and the reliability of each one of them. Analyze the differences between the PaaS and the IaaS.

Problem 5. Overprovisioning is the reliance on extra capacity to satisfy the needs of a large community of users when the average-to-peak resource demand ratio is very high. Give an example of a large-scale system using overprovisioning and discuss if overprovisioning is sustainable in that case and what are the limitations of it. Is cloud elasticity based on overprovisioning sustainable? Give the arguments to support your answer.

Exercises and Problems

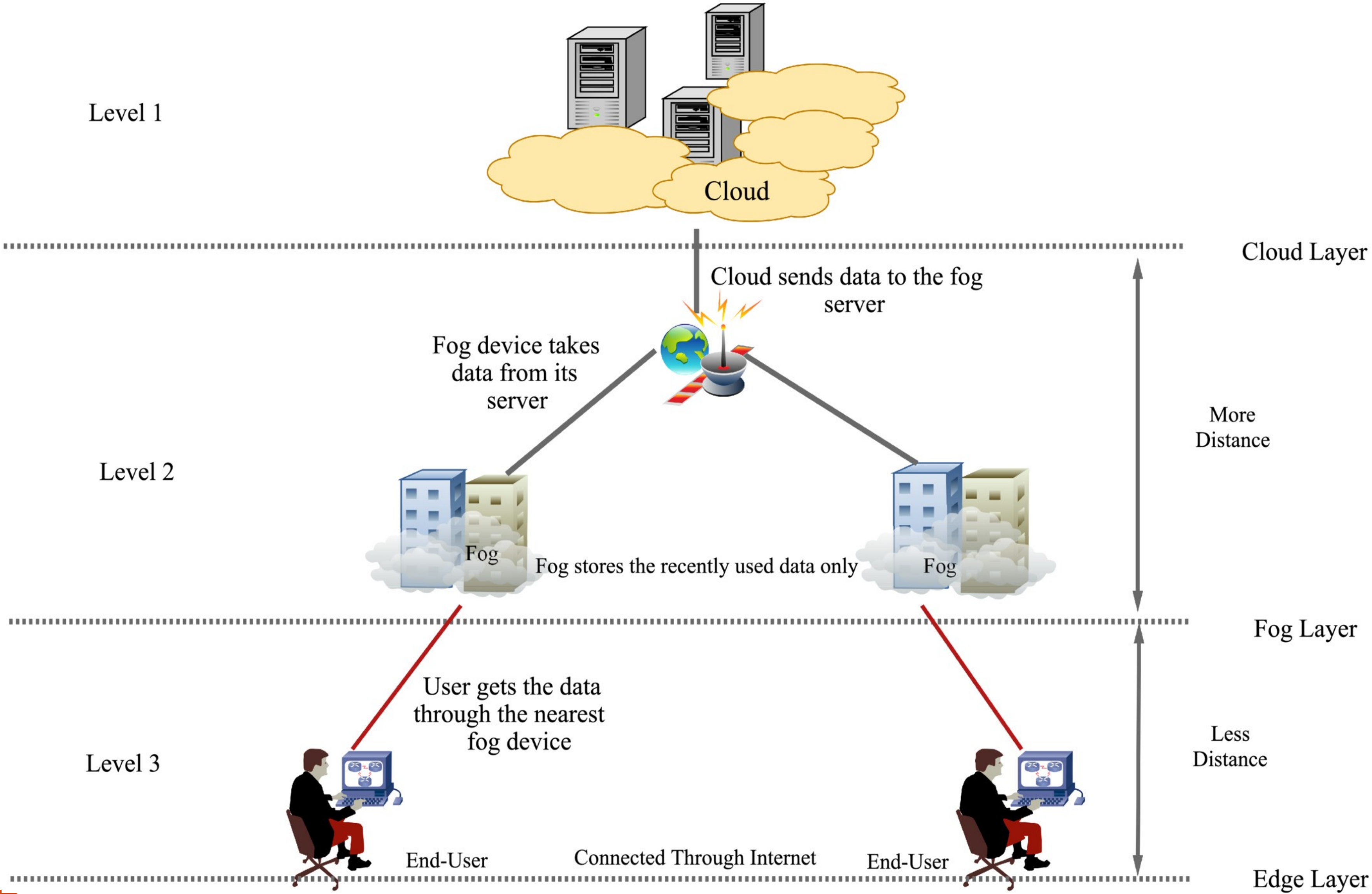
Problem 7. An organization debating whether to install a private cloud or to use a public cloud, e.g., the AWS, for its computational and storage needs, asks your advice. What information will you require to base your recommendation on, and how will you use each one of the following items: (a) the description of the algorithms and the type of the applications the organization will run; (b) the system software used by these applications; (c) the resources needed by each application; (d) the size of the user population; (e) the relative experience of the user population; (d) the costs involved.

Problem 8. A university is debating the question in Problem 7. What will be your advice and why? Should software licensing be an important element of the decision?

Exercises and Problems

- **Problem 9.** An IT company decides to provide free access to a public cloud dedicated to higher education. Which one of the three cloud computing delivery models, SaaS, PaaS, or IaaS should it embrace and why? What applications would be most beneficial for the students? Will this solution have an impact on distance learning? Why

Fog Computing Model


Level 1
Cloud
Cloud Layer
Cloud sends data to the fog server
Fog device takes data from its server
Level 2
More Distance
Fog
Fog stores the recently used data only
Fog
Fog Layer
User gets the data through the nearest fog device
Level 3
Less Distance
End-User
Connected Through Internet
End-User
Edge Layer