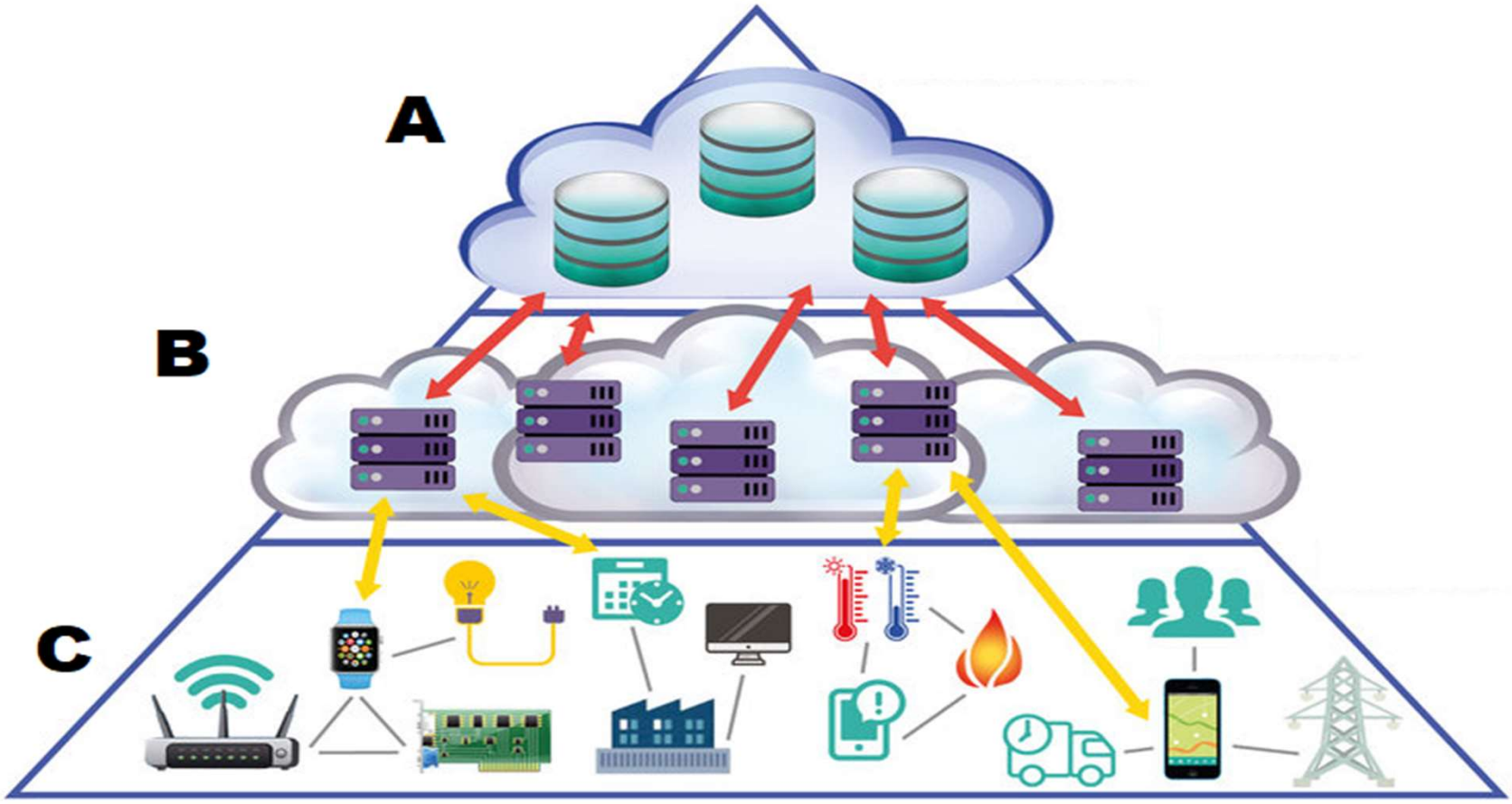
Cloud Computing

Module 1

Review



1. Mist Computing 2. Fog Computing 3. Cloud Computing
4. Edge Computing 5. Dew Computing

A Simple Example

- ABC Corporation is an e-commerce company
- They have traditionally hosted their website on an on-premise server
- Recently, they have been getting a lot of business
- Their IT head, Bob, observes that their server sometimes operates very near its capacity
- What options does Bob have?

Two Options

Horizontal Scaling

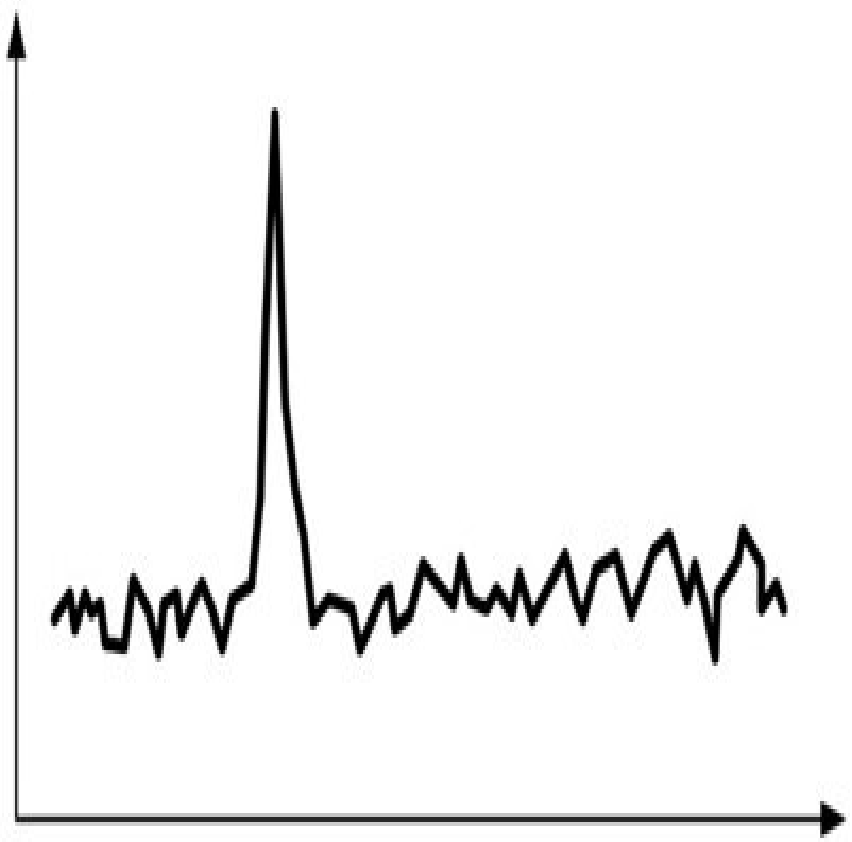
- Add more systems to handle the extra load
- Usually less expensive in the long run
- No downtime
- Potentially unlimited scaling

Vertical Scaling

- Upgrade the existing system
- Usually expensive in the long run
- Possible downtime
- There is a limit to which scaling can be done

But what if…

- But what if the website only shows a spike in usage during holiday season?
- If you upgrade, your new hardware remains unused most of the year
- If you don't upgrade, the company loses out on important revenue



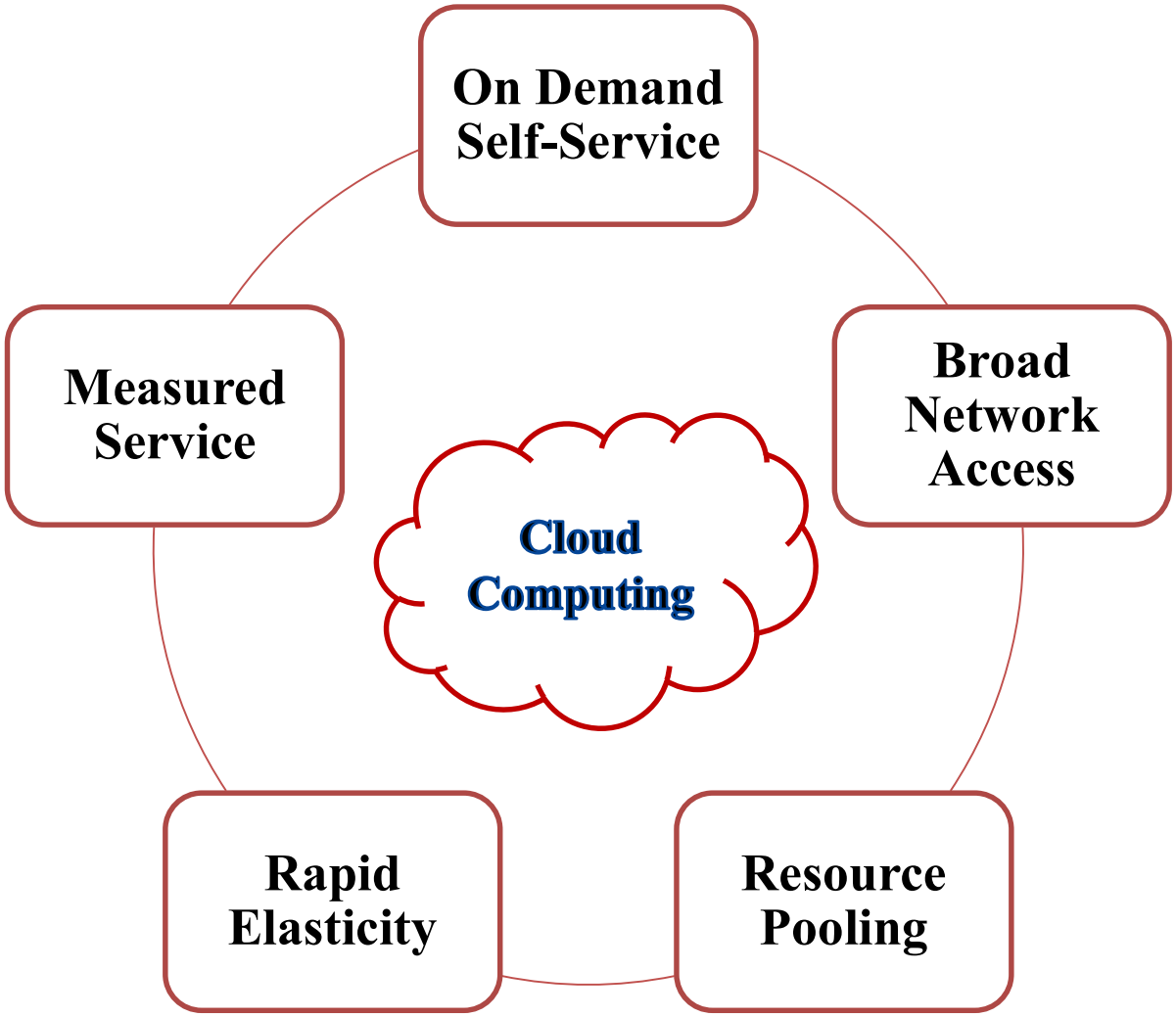
Secret Option

- Bob decides to move their business to the cloud
  - Automatic scaling when usage increases
  - When usage decreases, the server resources are scaled back
  - No upfront capital cost
  - Pay only for the resources that are used

NIST Definition of Cloud Computing

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

Essential Features of Cloud Computing



Contd…

1. **On Demand Self – Service**
   - Resources can be provisioned and released when needed
   - No need for human interaction
2. **Broad Network Access**
   - Resources can be accessed over the network
   - Any device – smartphone, tablet, laptop, thin client, PC etc.

Contd…

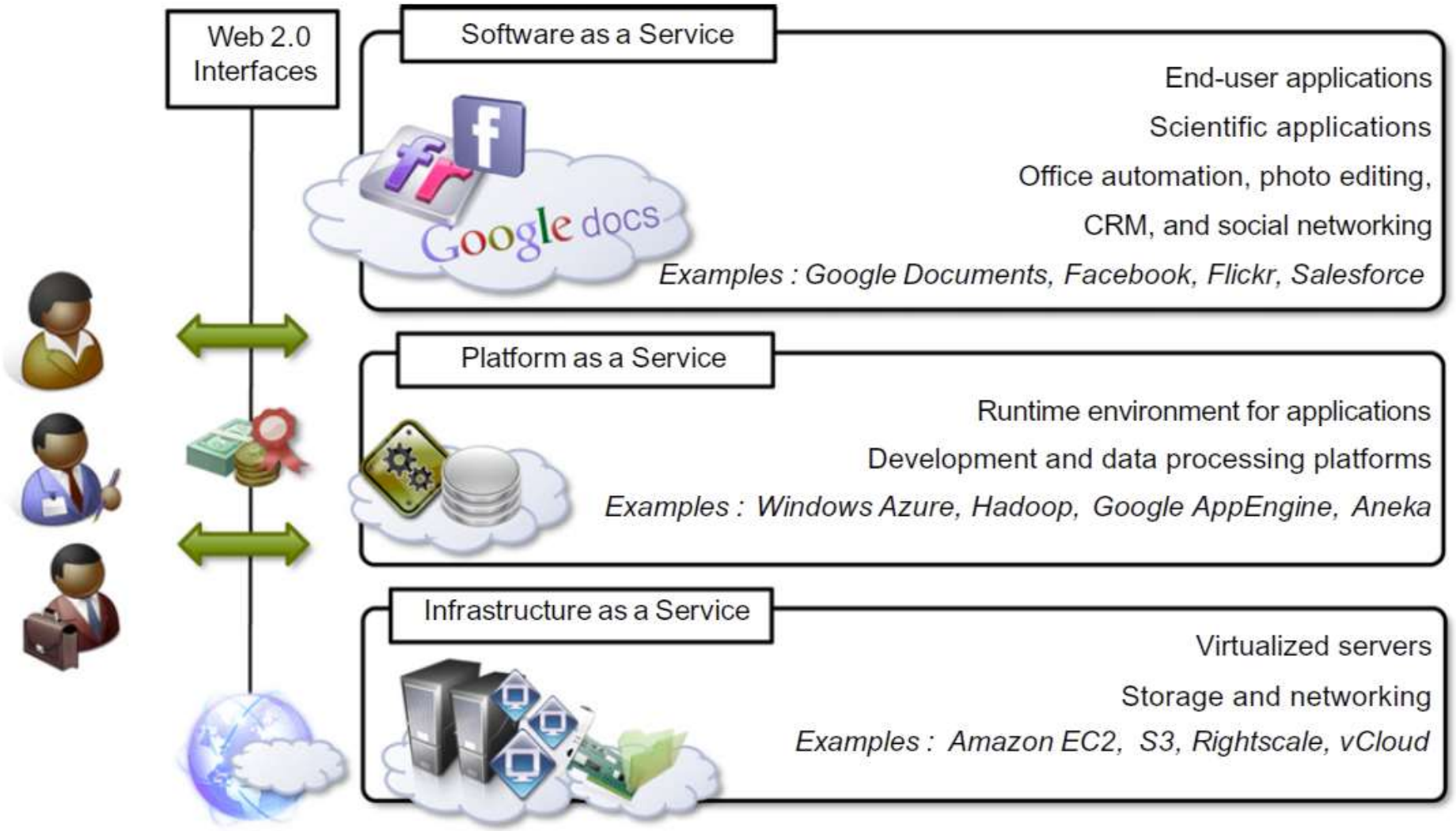
3. **Resource Pooling**
   - Service provider has a pool of resources that can be allocated and deallocated to customers
4. **Rapid Elasticity**
   - Resources can be allocated on the fly
   - Customers get an illusion of unlimited resources
5. **Measured Service**
   - Pay per use

SLA – The Warranty for Cloud Services

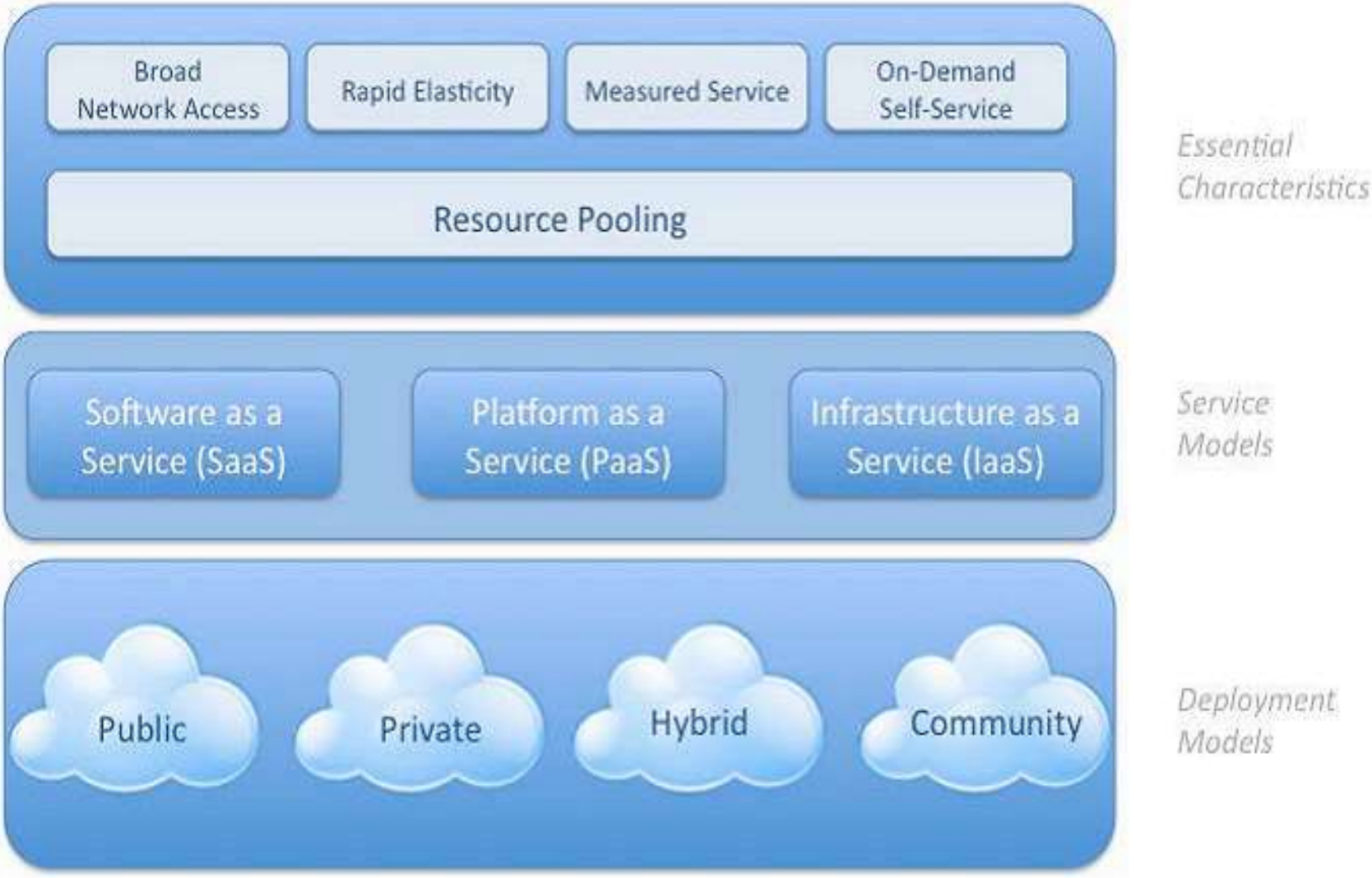
- SLA (Service Level Agreement) is a bond for performance negotiated between the cloud services provider and the client
  - Availability and Performance
  - Security / privacy of the data
  - Disaster Recovery expectations
  - Location of the data
  - Access and portability to the data
  - Process to identify problems and resolution expectations
  - Dispute mediation process (e.g. escalation process, consequences)
  - Exit Strategy with expectations on the provider to ensure smooth transition

Cloud Computing Vision

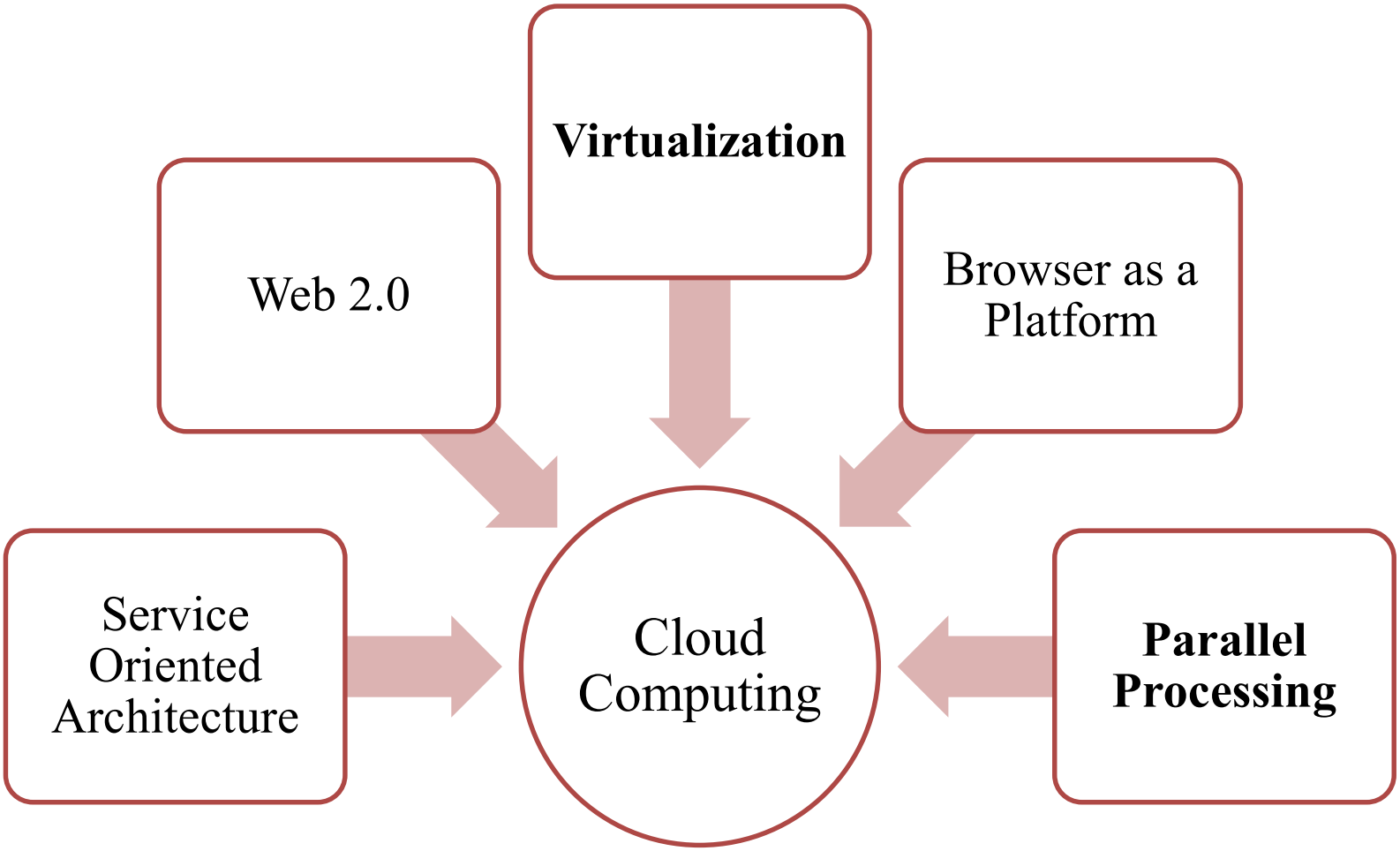
Cloud Computing Reference Model



NIST Visual Model



Enabling Technologies for Cloud Computing



Virtualization

- Virtualization is a core technology for cloud computing.
- Virtualization is essentially a technology that allows creation of different computing environments.
- These environments are called virtual because they simulate the interface that is expected by a guest.
- The most common example of virtualization is hardware virtualization.
- This technology allows simulating the hardware interface expected by an operating system.
- Hardware virtualization allows the coexistence of different software stacks on top of the same hardware.
- These stacks are contained inside virtual machine instances, which operate in complete isolation from each other.

Web 2.0

- The Web is the primary interface through which cloud computing delivers its services.
- Web 2.0 enables developers to architect applications and deliver services through the Internet.
- Web 2.0 brings interactivity and flexibility into Webpages.
- These capabilities are obtained by integrating a collection of standards and technologies such as XML, Asynchronous JavaScript and XML (AJAX), Web Services, and others.
- Examples of Web 2.0 applications are Google Documents, Google Maps, Flickr, Facebook, Twitter, YouTube, de.li.cious, Blogger, and Wikipedia.

Service Oriented Computing

- Service orientation is the core reference model for cloud computing systems.
- This approach adopts the concept of services as the main building blocks of application and system development.
- It supports the development of rapid, low-cost, flexible, interoperable, and evolvable application and systems.
- A service is supposed to be loosely coupled, reusable, programming language independent, and location transparent.
- Service-oriented computing introduces and diffuses two fundamental concepts to cloud computing: Quality of Service (QoS) and Software-as-a-Service (SaaS).
- Quality of service (QoS) identifies a set of functional and non-functional attributes that can be used to evaluate the behavior of a service from different perspectives