2022 Fall : Cloud Computing

Instructor: Young Tae Noh (ytnoh@kentech.ac.kr)

Structure

Lectures driven by reading List

Updated reading list in our courseweb (or recent papers (’20 ~ present) from Sigcomm, NSDI, OSDI, SOSP, EuroSys, ATC, Infocom, CoNEXT [,HotCloud, HotNet, HotSDN]).

+ Class participation

Assignments/Presentation

Each student will present FOUR paper in the reading list

~~Mini-project~~

~~Focused project that deals with specific research issues on cloud stack and processing framework~~

Grading

Final Term: 40%
Presentation: 40% (from the reading list on courseweb)
Class participation: 10%
Paper critics: 10%

Presentation RSVP:
https://docs.google.com/spreadsheets/d/1Ys0OMEBCKnJtBZgjVuZr5PM9R1LR2HLGJD28TmDzdBQ/edit#gid=0

Regarding Paper Read and Write Critics

How to read research papers:

https://www.eecs.harvard.edu/~michaelm/postscripts/ReadPaper.pdf
https://cseweb.ucsd.edu/~wgg/CSE210/howtoread.html
http://www.owlnet.rice.edu/~cainproj/courses/HowToReadSciArticle.pdf

A sample critic:

http://web.mit.edu/6.897/www/essays/cloudcmp-lacurts.pdf

Lecture Motivation…

- **General overview on cloud computing**
  - What is cloud computing
    - Services
    - Types
  - Advantages and disadvantages
  - Enabling technologies
  - An example infrastructure

Lecture Outline

- What is Cloud?
- What is Cloud Computing?
- Cloud Computing Services
- History of Cloud Computing
- Why Cloud Computing
- Drawbacks of Cloud Computing
- Types of Clouds

(Based on slides from Aditya Akella and Majid F. Sakr)

Cloud Computing in Numbers

Datacenter instance:

Costs in billion$ range

> 100,000 servers

Number of servers (estimates*)

Google: ~1 mil servers

Microsoft, Yahoo!, IBM, HP: several 100,000s each

Amazon, Ebay, GoDaddy, Facebook, Akamai: > 50,000

(*http://www.datacenterknowledge.com/archives/2009/05/14/whos-got-the-most-web-servers/)

A Cloud is …

- **Datacenter hardware and software** that the vendors use to offer the computing resources and services

Cloud Computing

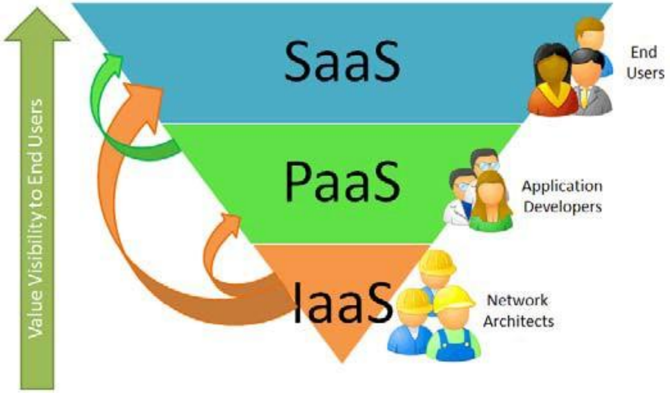
- **Represents both the cloud & the provided services**
- **Why call it “cloud computing”?**
  - Some say because the computing happens out there "*in the clouds*"

    Wikipedia: "*the term derives from the fact that most technology diagrams depict the Internet or IP availability by using a drawing of a cloud."*

NIST Definition

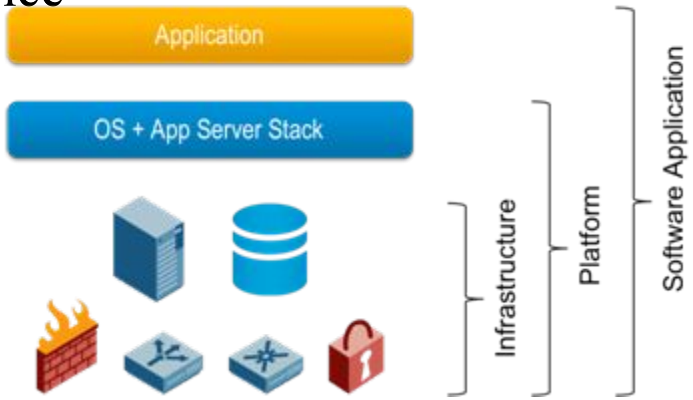
A model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

Cloud Computing Services

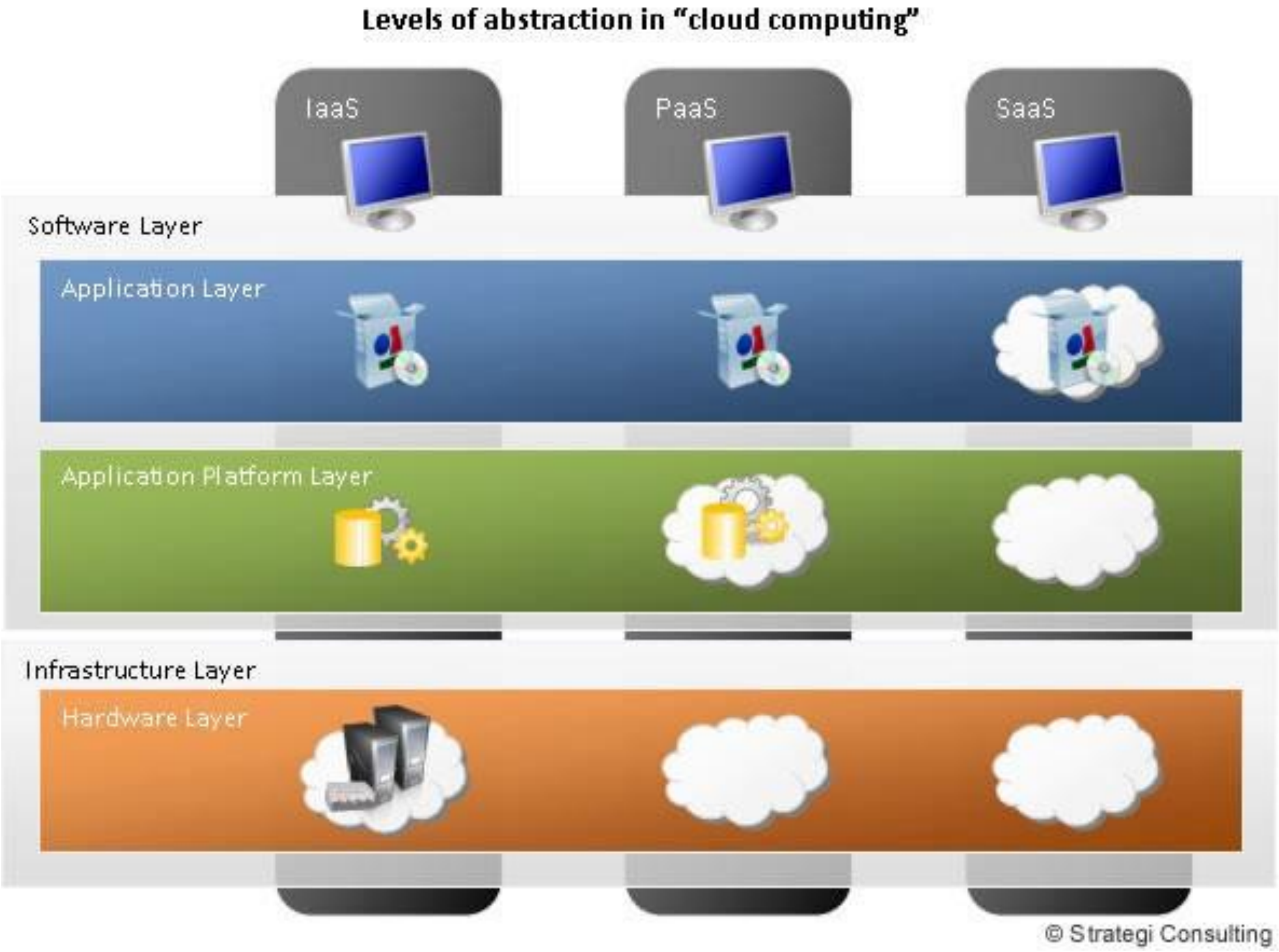


Three basic services:

- **Software as a Service (SaaS) model**
  - Apps through browser
- **Platform as a Service (PaaS) model**
  - Delivery of a computing platform for custom software development as a service
- **Infrastructure as a Service (IaaS) model**
  - Deliver of computer infrastructure as a service
- **XaaS, the list continues to grow…**



Cloud Services (XaaS)

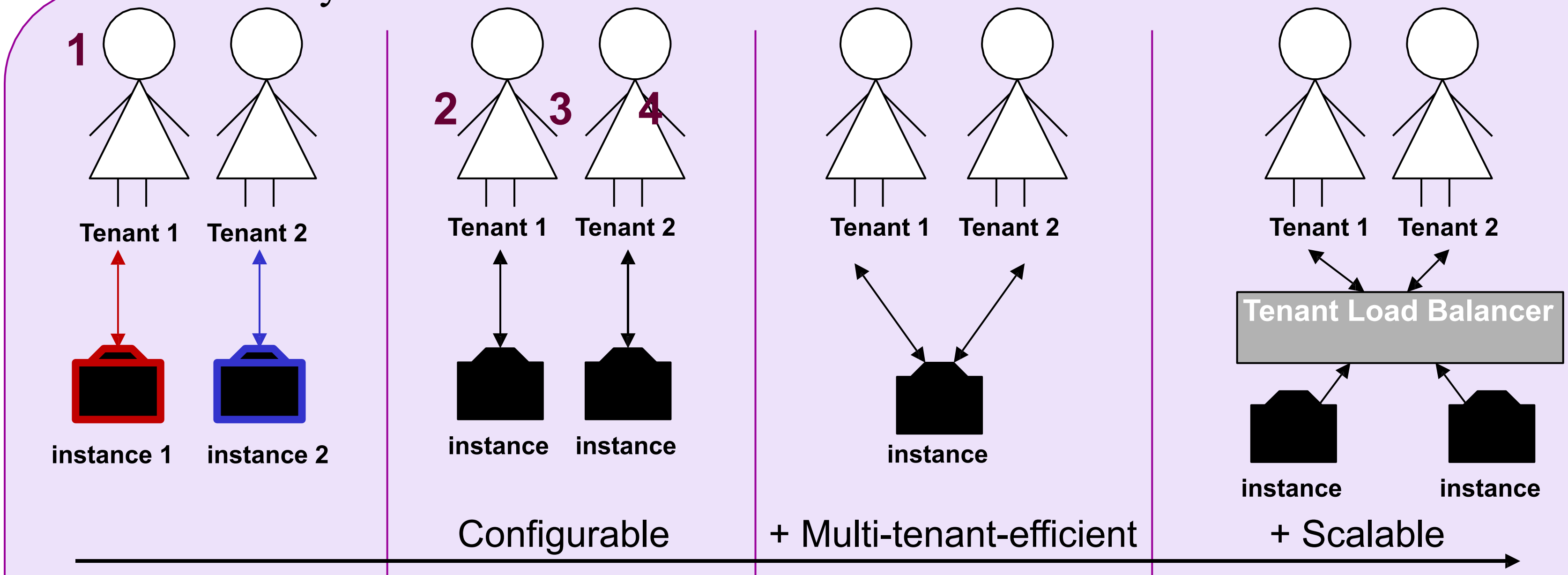


SaaS (1/3)

- **Started around 1999**
- **Application is licensed to a customer as a service on demand**
- **Software Delivery Model:**
  - Hosted on the vendor’s web servers
  - Downloaded at the consumer’s device and disabled when on-demand contract is over

SaaS (2/3)

- **SaaS architecture/ Maturity levels:**
  - **Distinguishing attributes:** configurability, multi-tenant efficiency, scalability



- Each has its own customized version of the application and run its own instance

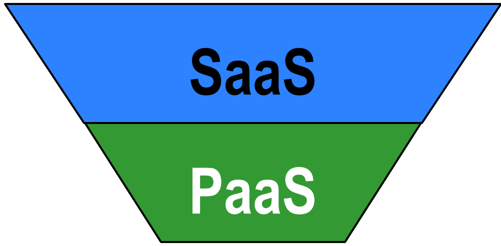
- Same application but distinct instance/customer

- (+):Efficient use of server resources without apparent differences to end users
- (-): scalability limits

SaaS (3/3)

- **Examples**

PaaS (1/2)

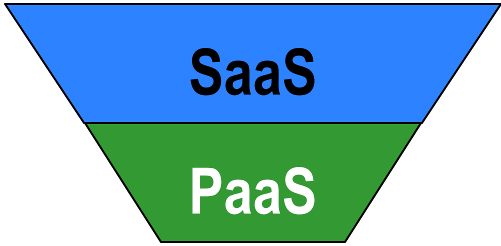


- Delivery of an integrated computing platform (to build/test/deploy custom apps) & solution stack as a service.

- Deploy your applications & don’t worry about buying & managing the underlying hardware and software layers

- The provider provides the networks, servers, storage, operating system (OS), middleware (e.g. Java runtime, .NET runtime, integration, etc.), database and other services to host the consumer's application.

PaaS (2/2)



- **Examples**

Google Apps

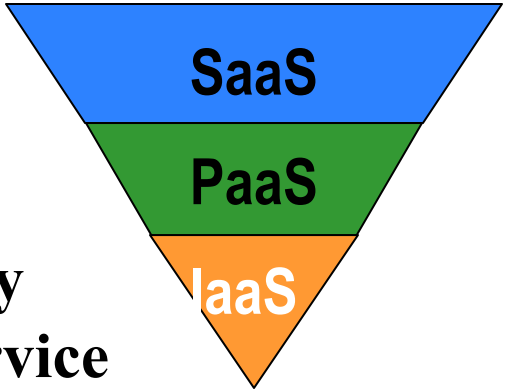
Google Cloud Platform

AWS Elastic Beanstalk

APPRENDA

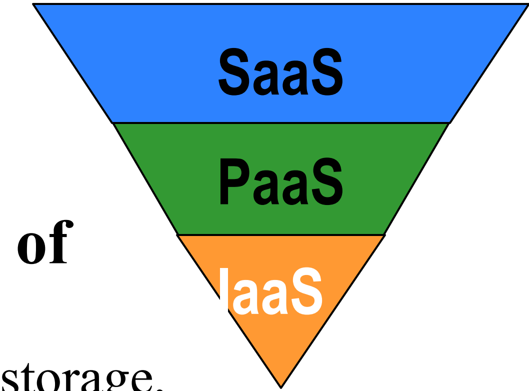
apache Stratos

IaaS (1/5)

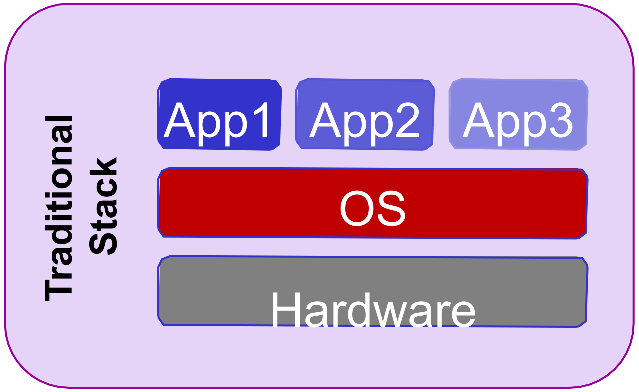


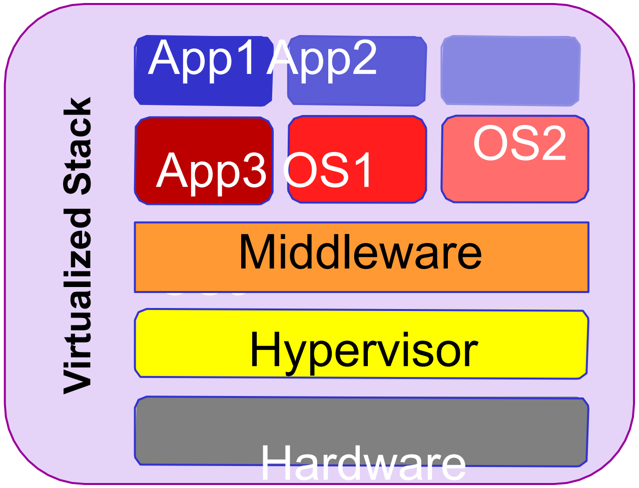
- Delivery of computer infrastructure (typically platform virtualization environment) as a service
- Buy resources
  - Servers
  - Software
  - Data center space
  - Network equipment as fully outsourced services
- Example:

IaaS (2/5)

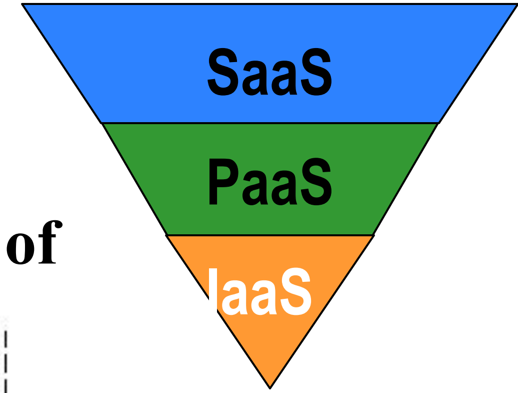


- **Virtualization Technology is a major enabler of IaaS**
  - It's a path to share IT resource pools: Web servers, storage, data, network, software and databases.
  - Higher utilization rates

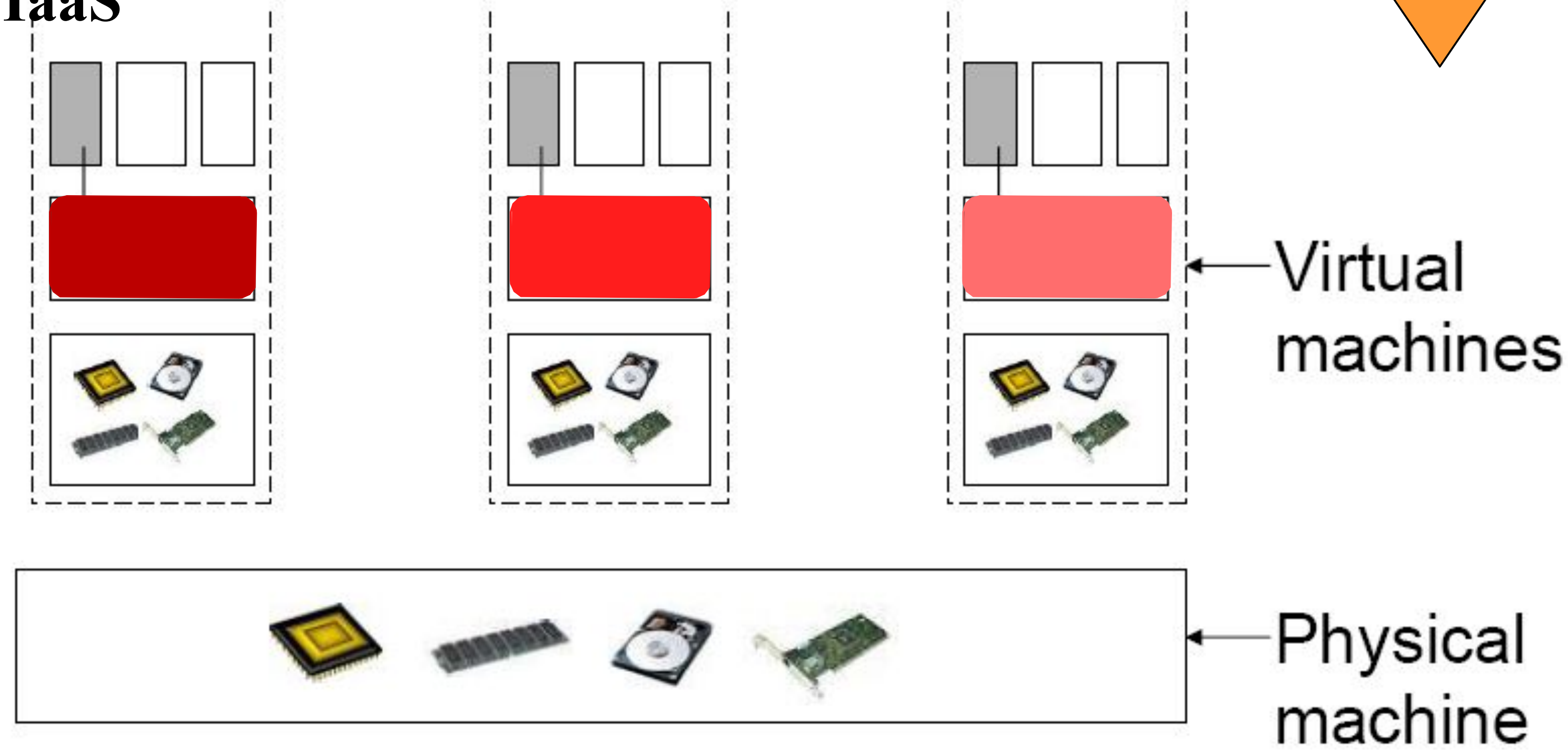




IaaS (3/5)

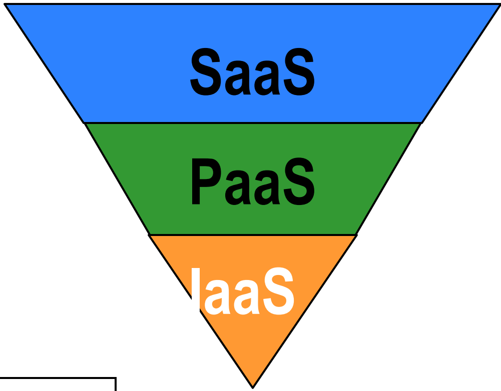


- **Virtualization Technology is a major enabler of IaaS**



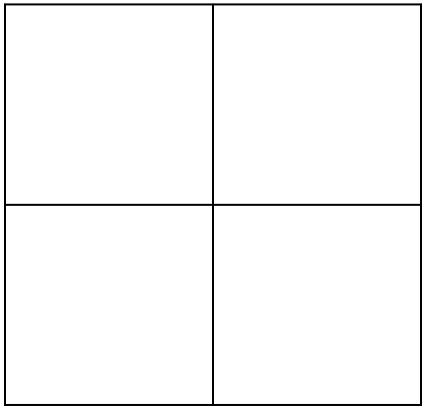
HARDWARE

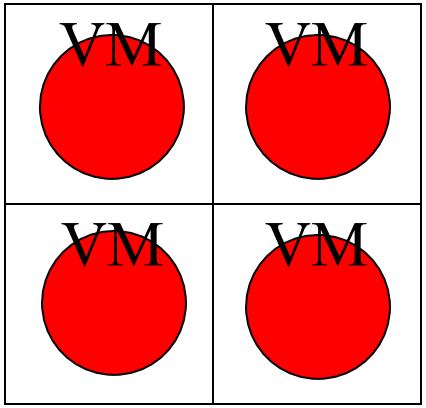
IaaS (4/5)

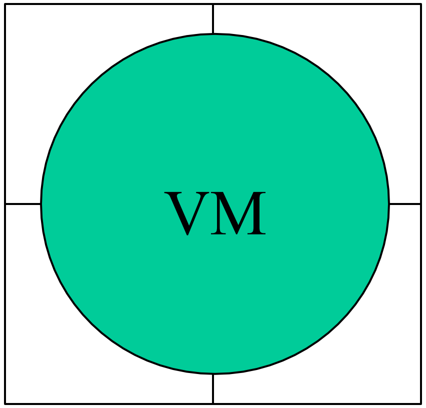


- **Granularity of VMs**
  - Multi-core processors

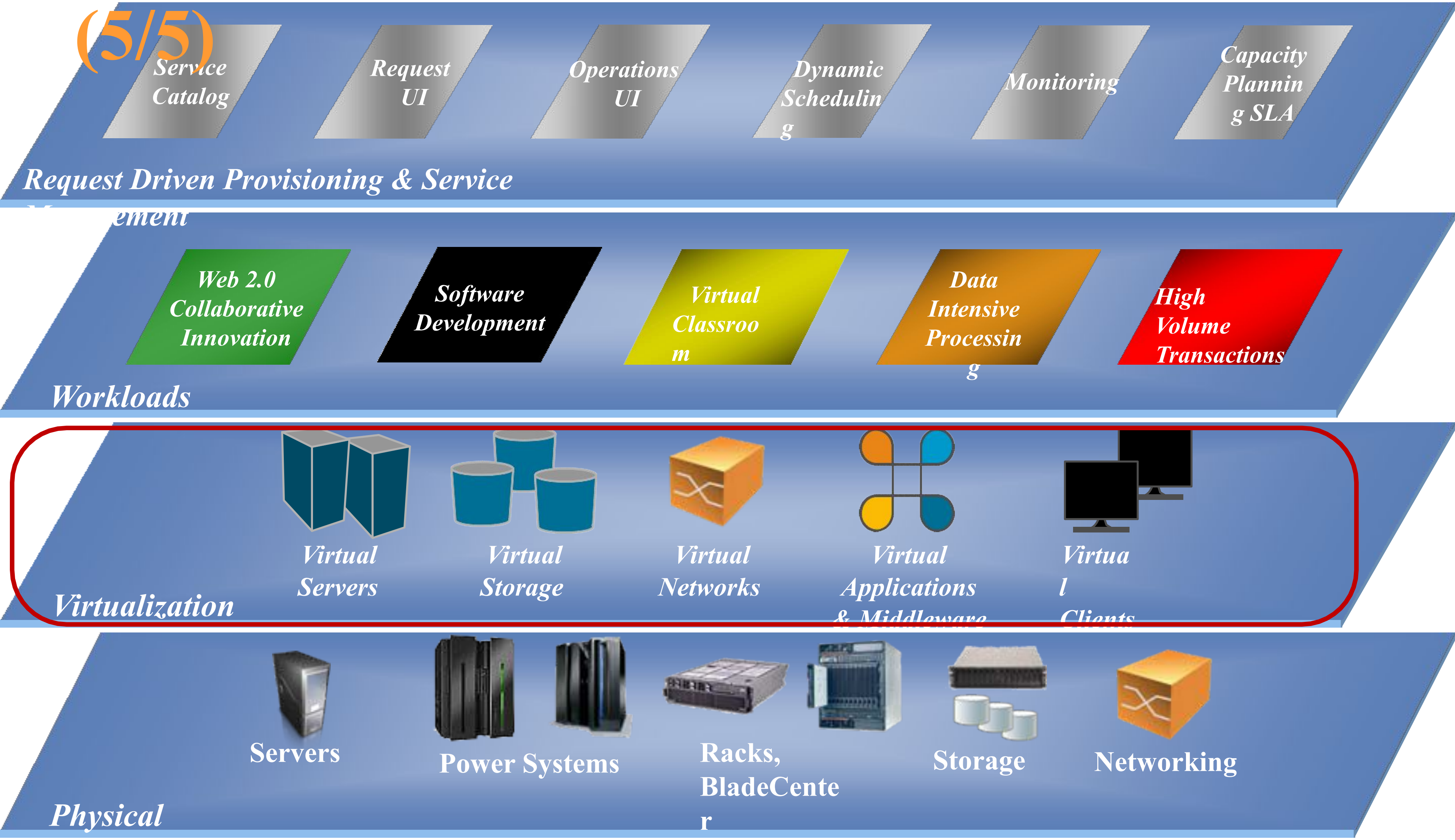
Quad Core:







IaaS (5/5)



Service Catalog

Request UI

Operations UI

Dynamic Scheduling

Monitoring

Capacity Planning SLA

Request Driven Provisioning & Service Management

Web 2.0 Collaborative Innovation

Software Development

Virtual Classroom

Data Intensive Processing

High Volume Transactions

Workloads

Virtual Servers

Virtual Storage

Virtual Networks

Virtual Applications & Middleware

Virtual Clients

Virtualization

Servers

Power Systems

Racks, BladeCenter

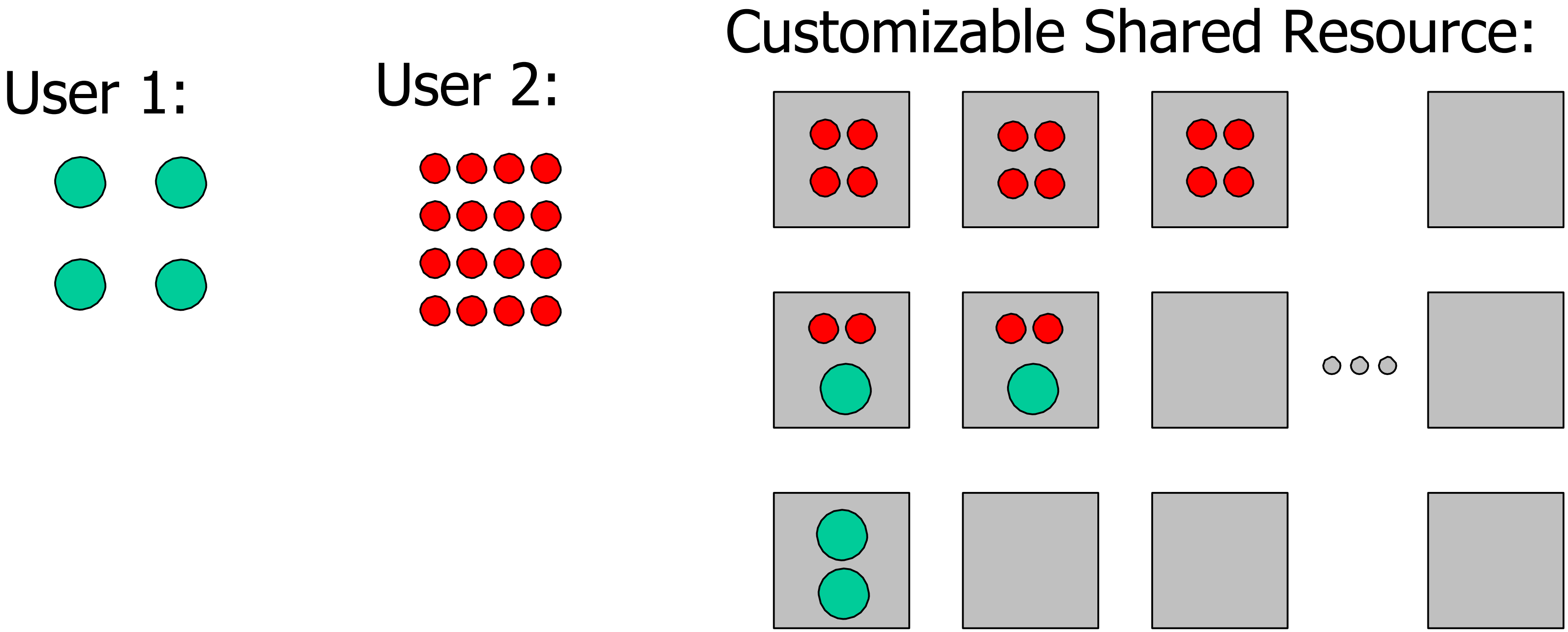
Storage

Networking

Physical

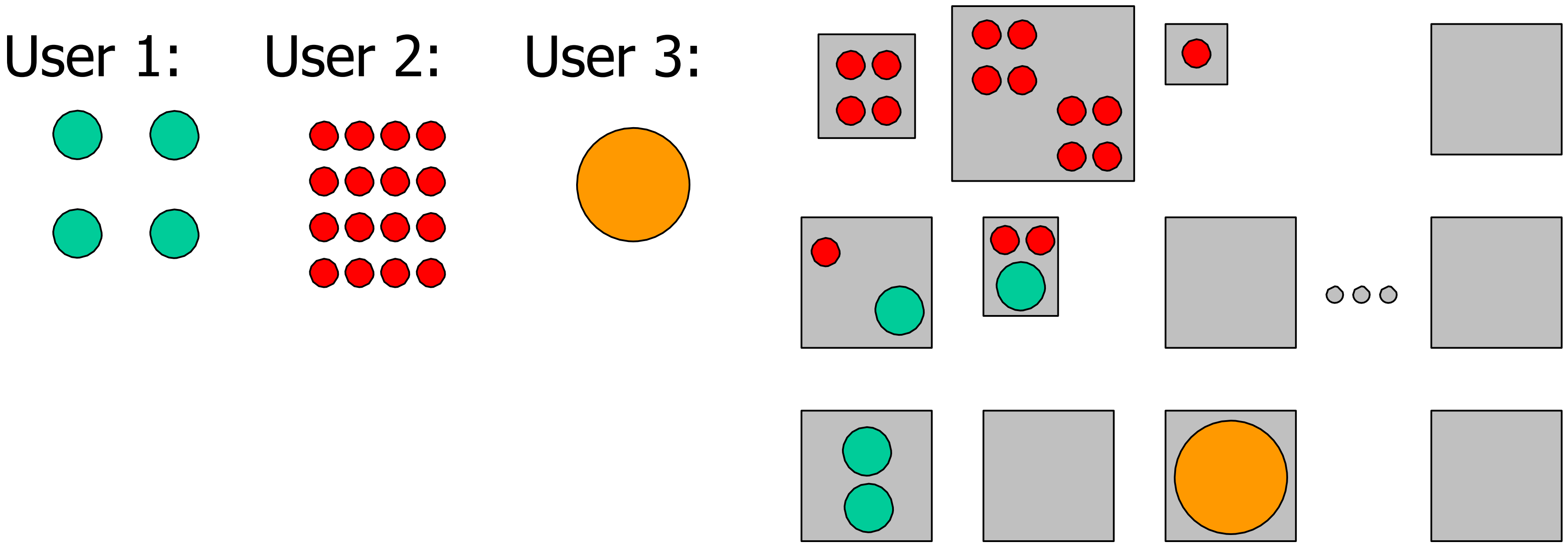
Resource sharing and consolidation

- **Offering computing resources as a service or utility through:**
  - Virtualization
  - Dynamic provisioning



Heterogeneous Physical Resources

Customizable Shared **Heterogeneous** Resource:

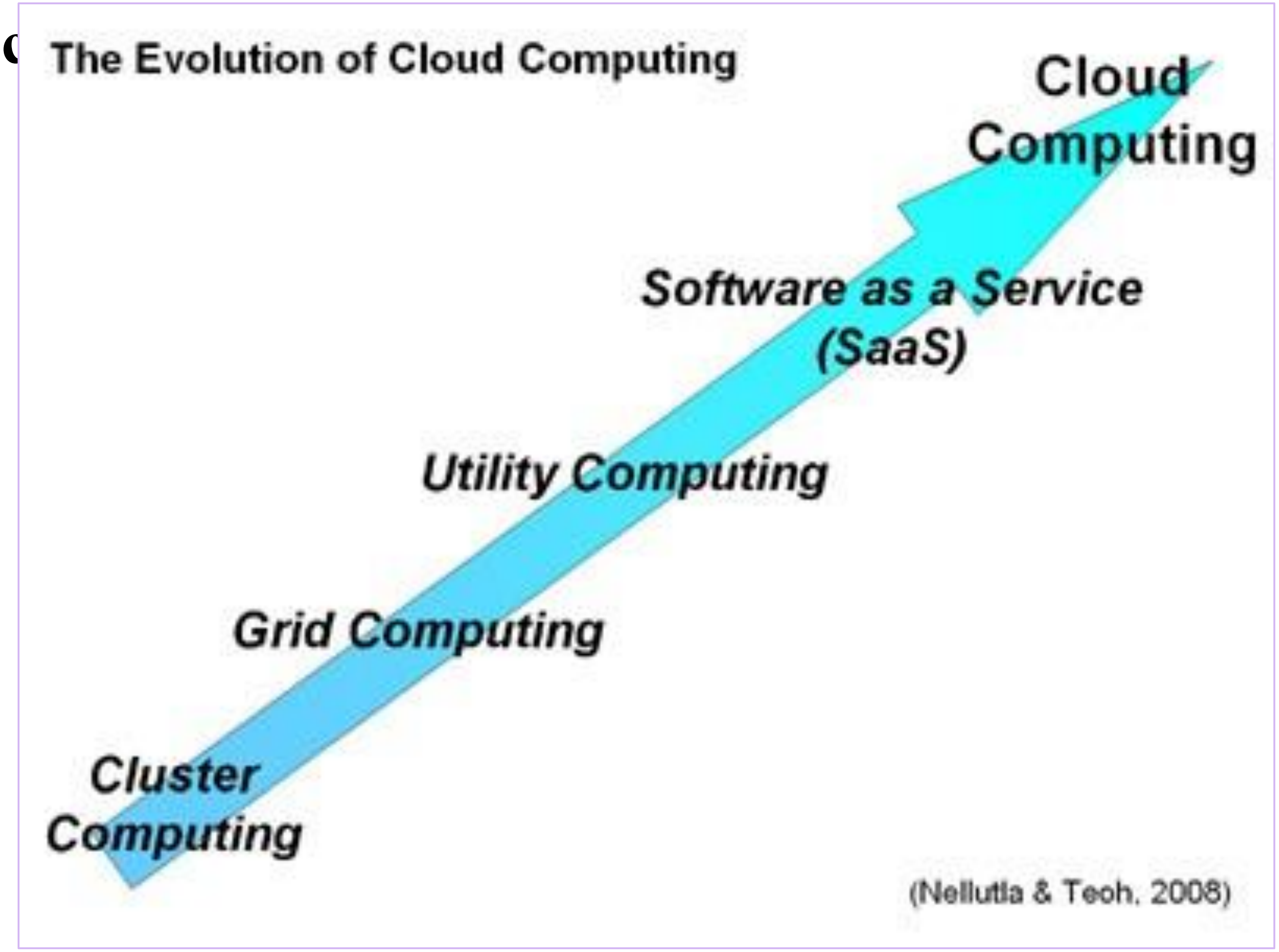


More (XaaS): Everything as a Service EaaS

- Desktop: DaaS
  - Use your desktop virtually from anywhere
- Communication: CaaS
- Virtualization: VaaS
- Hardware: HaaS
- …etc

Evolution

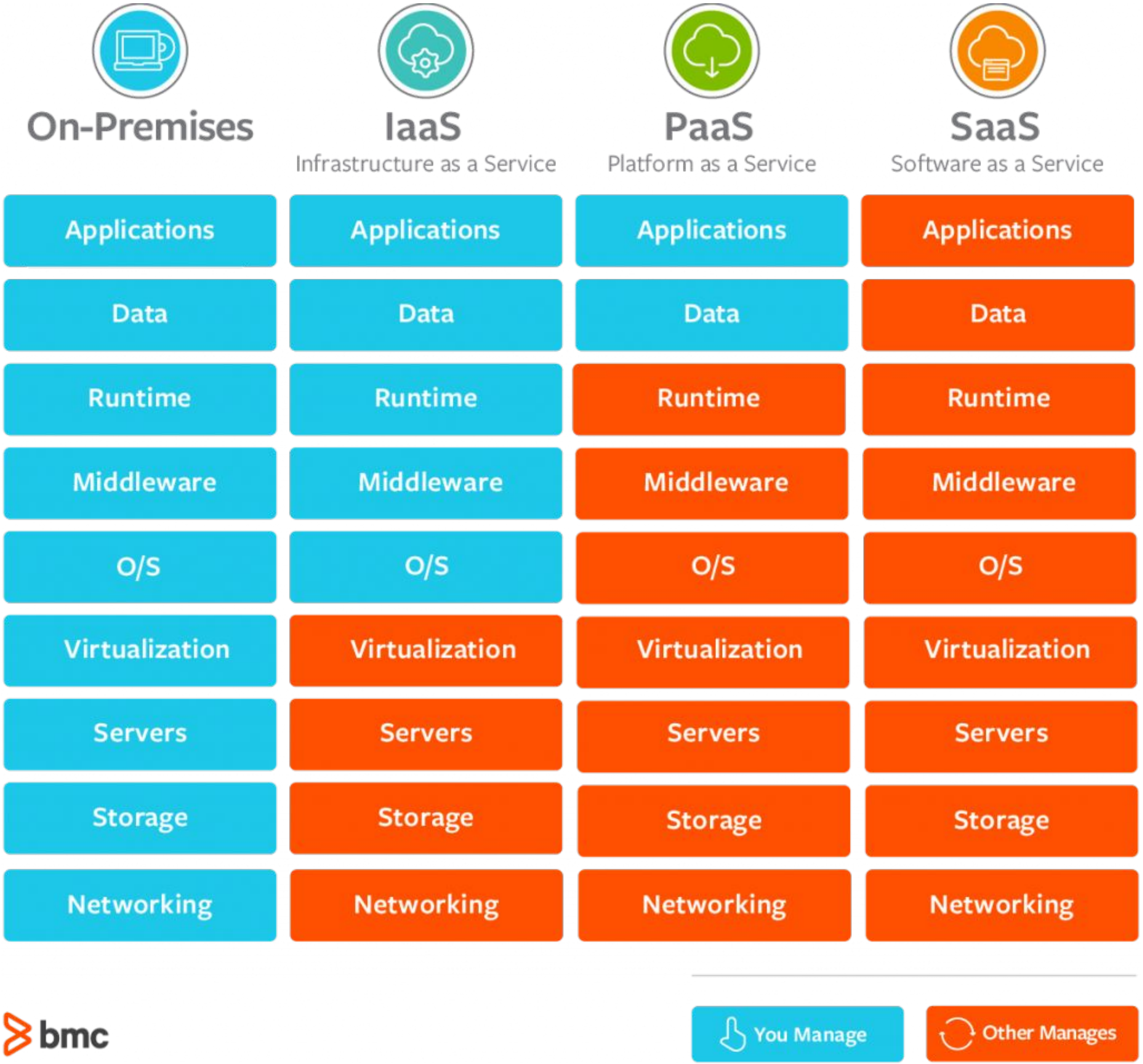
- **Discussed in lec**



The Evolution of Cloud Computing

(Nellutla & Teoh, 2008)

Summary of key differences



On-Premises	IaaS Infrastructure as a Service	PaaS Platform as a Service	SaaS Software as a Service
🟦 Applications	🟦 Applications	🟦 Applications	🟧 Applications
🟦 Data	🟦 Data	🟦 Data	🟧 Data
🟦 Runtime	🟦 Runtime	🟧 Runtime	🟧 Runtime
🟦 Middleware	🟦 Middleware	🟧 Middleware	🟧 Middleware
🟦 O/S	🟦 O/S	🟧 O/S	🟧 O/S
🟦 Virtualization	🟧 Virtualization	🟧 Virtualization	🟧 Virtualization
🟦 Servers	🟧 Servers	🟧 Servers	🟧 Servers
🟦 Storage	🟧 Storage	🟧 Storage	🟧 Storage
🟦 Networking	🟧 Networking	🟧 Networking	🟧 Networking

🟦 You Manage

🟧 Other Manages

Enabling Technologies

- Virtualization
- Web 2.0
- Distributed Storage
- Distributed Computing
- Utility Computing
- Network Bandwidth & Latency
- Fault-Tolerant Systems

Why Cloud Computing?

- Large-Scale Data-Intensive Applications
- Flexibility
- Scalability
- Customized to your current needs:
  - Hardware
  - Software
- Effect:
  - Reduce Cost
  - Reduce Maintenance
  - High Utilization
  - High Availability

Why Cloud Computing?

- **Flexibility**
  - Software: Any software platform
  - Access: access resources from any machine connected to the Internet
  - Deploy infrastructure from anywhere at anytime
    - Software controls infrastructure

Why Cloud Computing?

- **Scalability**
  - Instant
  - Control via software
    - Add/cancel/rebuild resources instantly
  - Start small, then scale your resources up/down as you need
  - Illusion of infinite resources available on demand

Why Cloud Computing?

- **Customization**
  - Everything in your wish list
    - Software platforms
    - Storage
    - Network bandwidth

Why Cloud Computing?

- Cost
  - Pay-as-you-go model
  - Small/medium size companies can tap the infrastructure of corporate giants.
    - Time to service/market
    - No upfront cost

Example: EC2

- **Amazon Elastic Compute Cloud (EC2)**
- **“Compute unit” rental: $0.10-0.80/hr.**
- **1 CU ≈ 1.0-1.2 GHz 2007 AMD Opteron/Xeon core**

“Instances”	Platform	Cores	Memory	Disk
Small - $0.10 / hr	32-bit	1	1.7 GB	160 GB
Large - $0.40 / hr	64-bit	4	7.5 GB	850 GB – 2 spindles
XLarge - $0.80 / hr	64-bit	8	15.0 GB	1690 GB – 3 spindles

- **No up-front cost, no contract, no minimum**

- **Billing rounded to nearest hour; pay-as-you-go storage also available**

Why Cloud Computing?

- **Maintenance**
  - Reduce the size of a client's IT department
  - Is the responsibility of the cloud vendor
  - This Includes:
    - Software updates
    - Security patches
    - Monitoring system's health
    - System backup
    - …etc

Why Cloud Computing?

- **Utilization**
  - Consolidation of a large number of resources
    - CPU cycles
    - Storage
    - Network Bandwidth

Why Cloud Computing?

- **Availability**
  - Having access to software, platform, infrastructure from anywhere at any time
  - All you need is a device connected to the internet

- **Reliability**

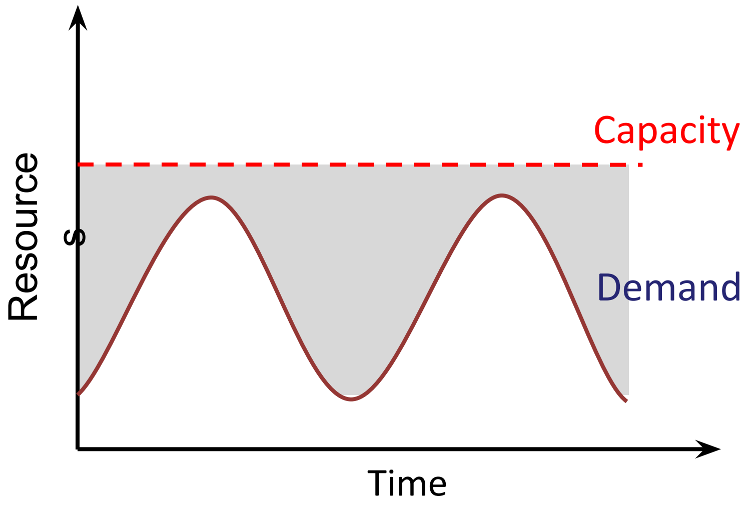
The system's fault tolerance is managed by the cloud providers and users no longer need to worry about it.

Why Cloud Computing?

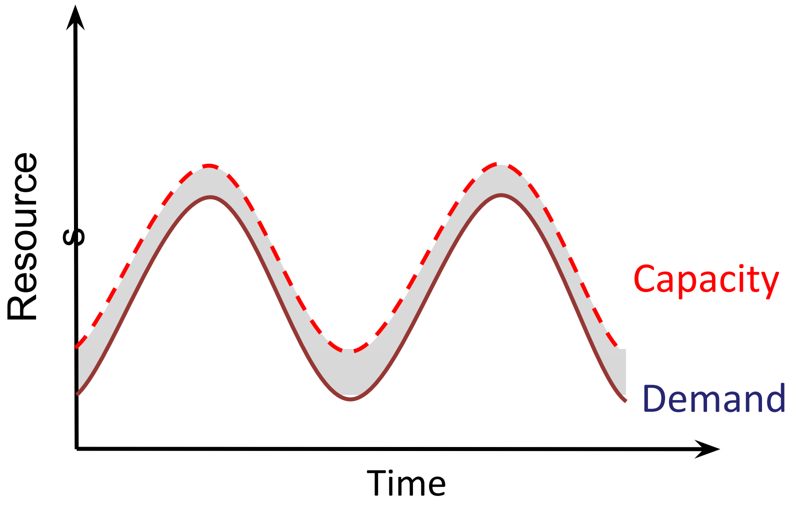
- CO2 Footprint
  - Consolidation of servers
  - Higher utilization
  - Reduced power usage

Cloud Economics

- Static provisioning for peak: wasteful, but necessary for SLA



“Statically provisioned” data center

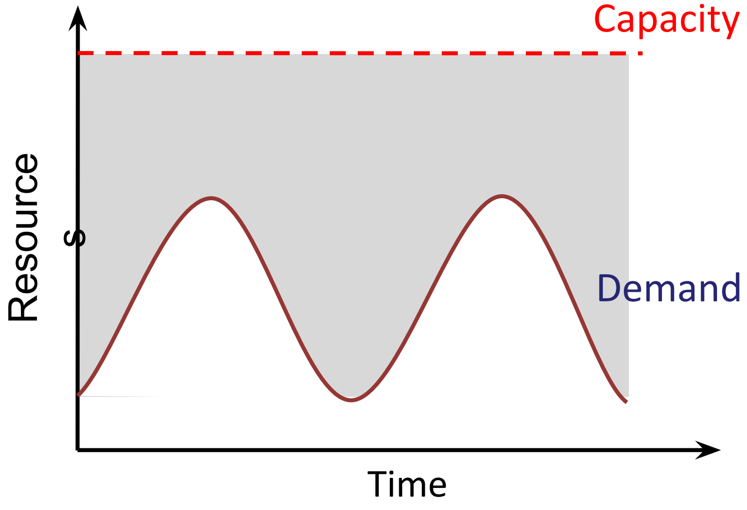


“Virtual” data center in the cloud

▨ Unused resources

Risk of underutilization

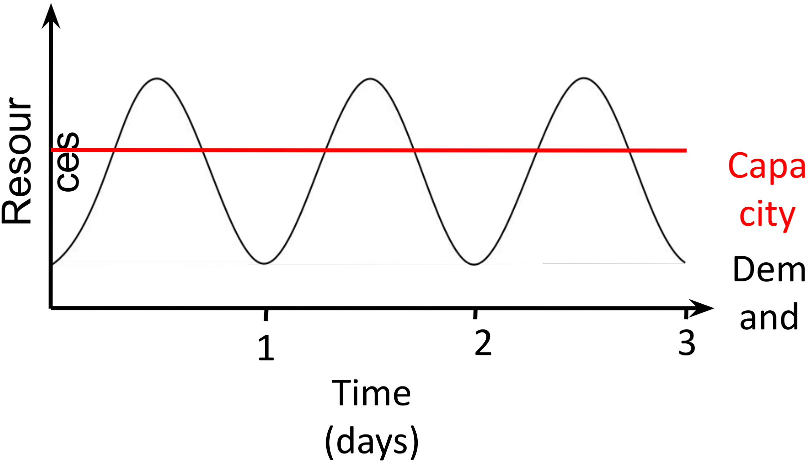
- Underutilization results if “peak” predictions are too optimistic

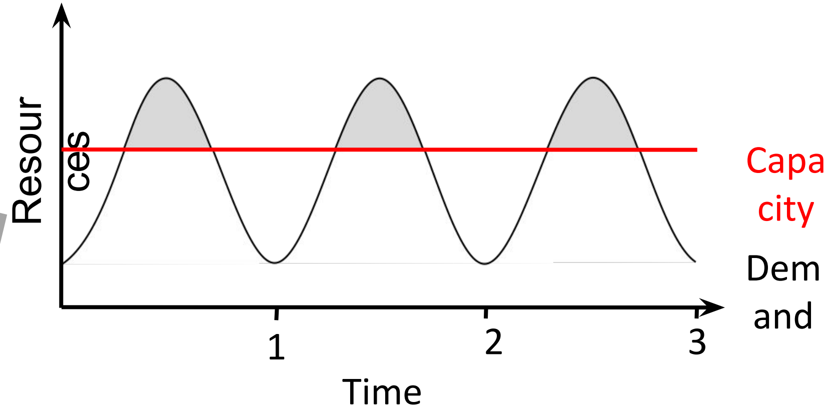


Static data center

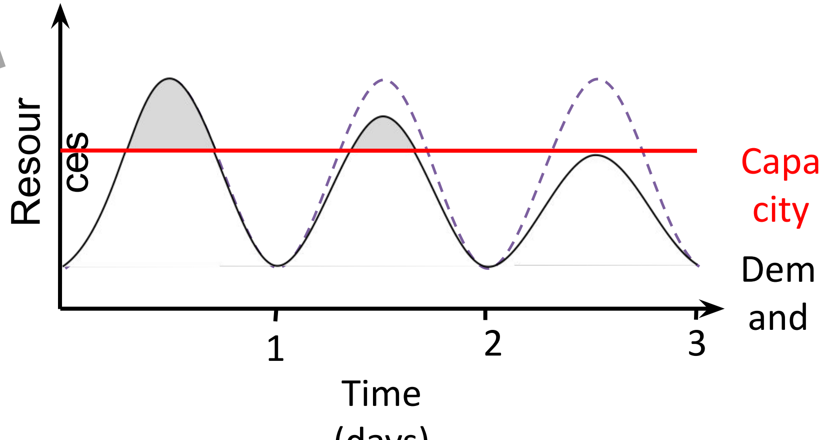
▨ Unused resources

Risks of underprovisioning





Lost revenue



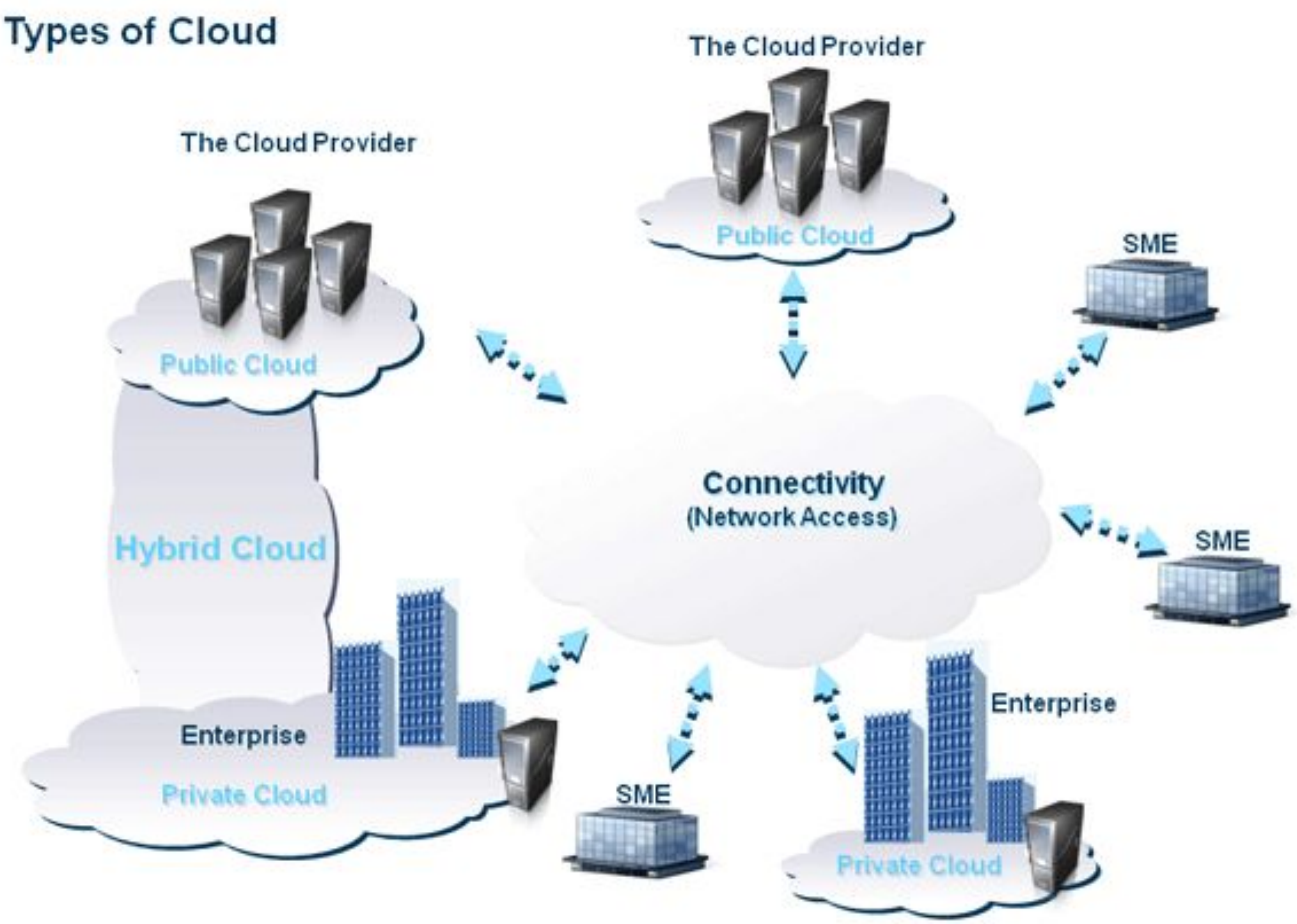
Lost users

Drawbacks

- Security
- Privacy
- Vendor lock-in
- Network-dependent
- Migration

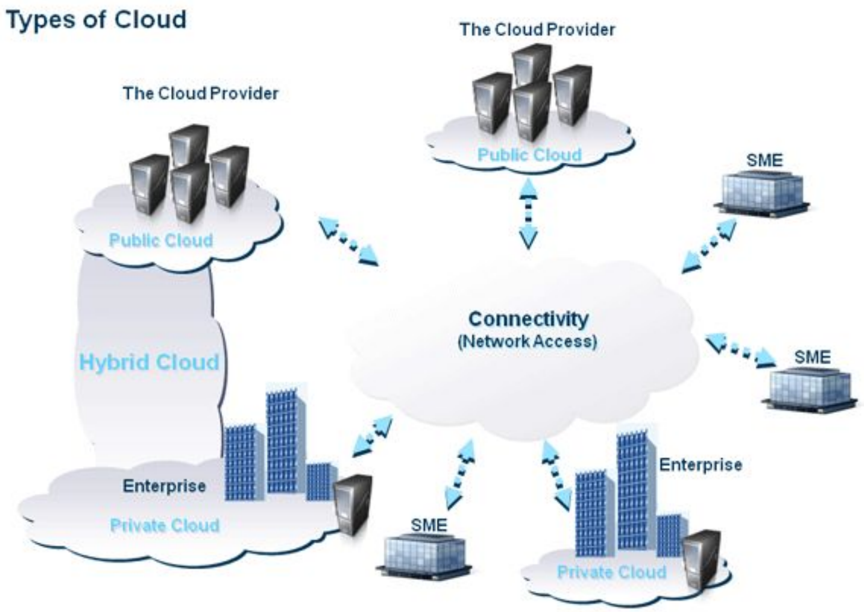
Types of Clouds (1/4)

- Public
- Privat
e
- Hybri
d



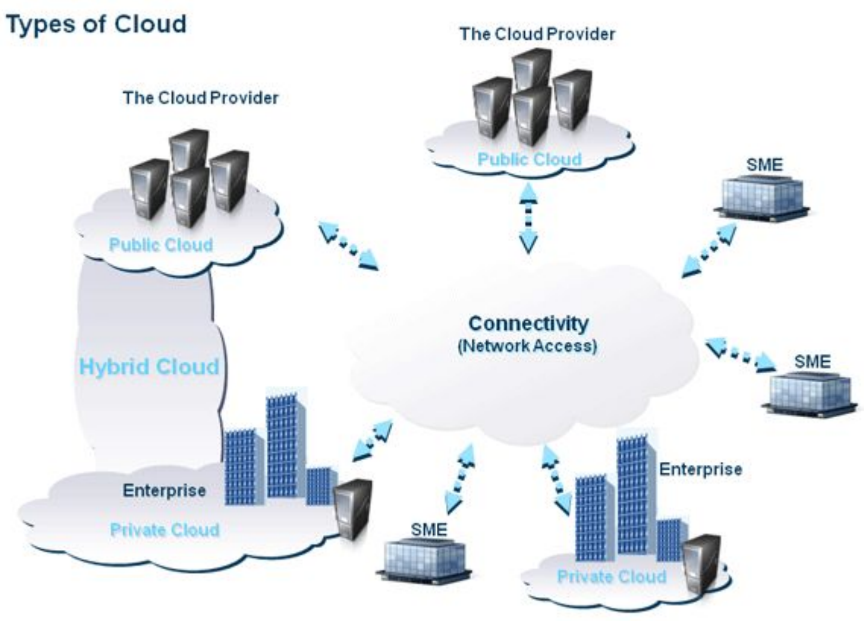
Types of Clouds (2/4)

- **Public (external) cloud**
  - Open Market for on demand computing and IT resources
  - Concerns: Limited SLA, Reliability, Availability, Security, Trust and Confidence
  - Examples: IBM, Google, Amazon, …



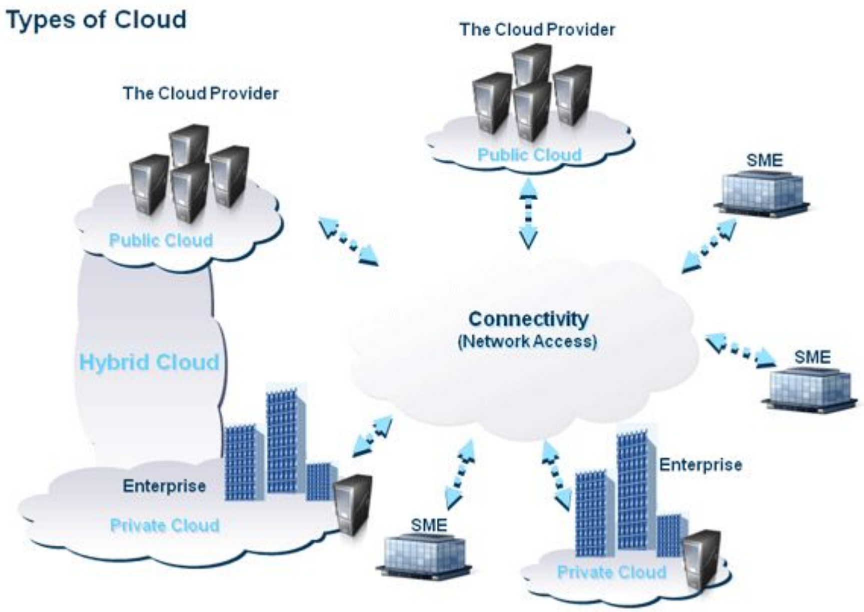
Types of Clouds (3/4)

- **Private (Internal) cloud**
  - For Enterprises/Corporations with large scale IT



Types of Clouds (4/4)

- **Hybrid cloud**
  - Extend the Private Cloud(s) by connecting it to other external cloud vendors to make use of available cloud services from external vendors
- **Cloud Burst**
  - Use the local cloud,
    when needing more resources,
    burst into the public cloud



MapReduce and Apache Hadoop

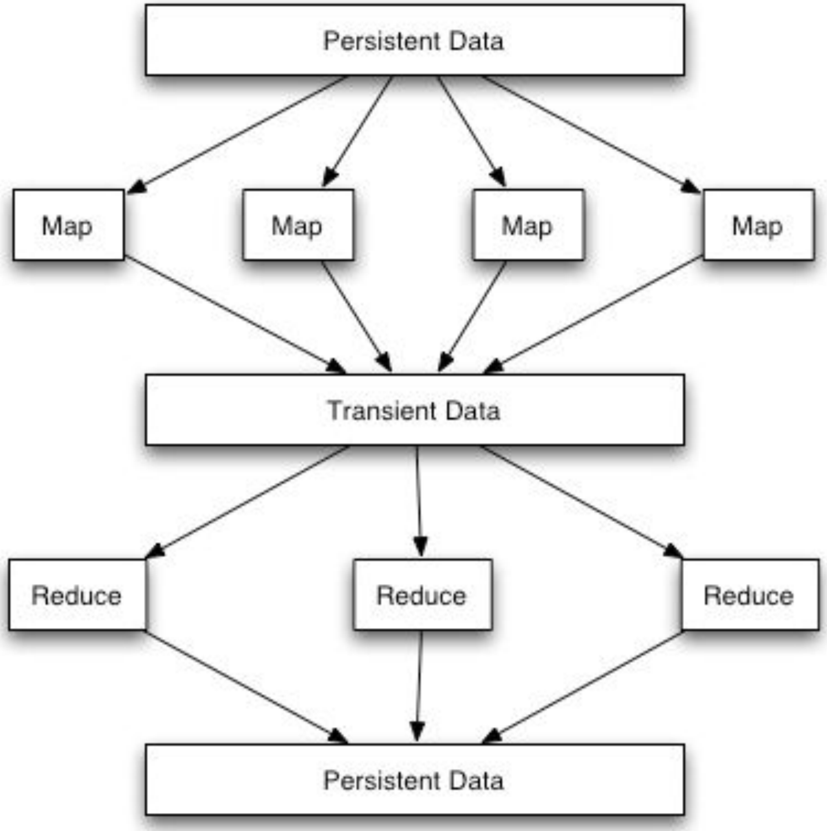
- **MapReduce: Abstraction that simplifies writing applications that access massively distributed data**

- **Hadoop: Open source MapReduce software platform**

- **Distributes data and processing across many nodes**
- **Processes the data locally at each node**
- **Transparent fault tolerance through**
  - Automatic data duplication

MapReduce Programming Model

- **Functional programming that is easily parallelizable**
- **Split into two phases:**
  - Map – Perform custom function on all items in an array
  - Reduce – Collate map results using custom function
- **Scales well – computation separated from processing dataflow**
- **Illustrative example:**
  - Map that squares the value of numbers in an array
    {1, 2, 3, 4} -> {1, 4, 9, 16}
  - Reduce that sums the squares : 30



Hadoop Map/Reduce

- **The Map-Reduce programming model**
  - **Framework for distributed processing of large data sets**
  - **Pluggable user code runs in generic framework**
- **Example:**
  - **cat * | grep | sort | unique -c | cat > file**
  - **input** | map | **shuffle** | reduce | **output**
- **Natural for unstructured data:**
  - **Log processing**
  - **Web search indexing**
  - **Ad-hoc queries**

Apache Hadoop

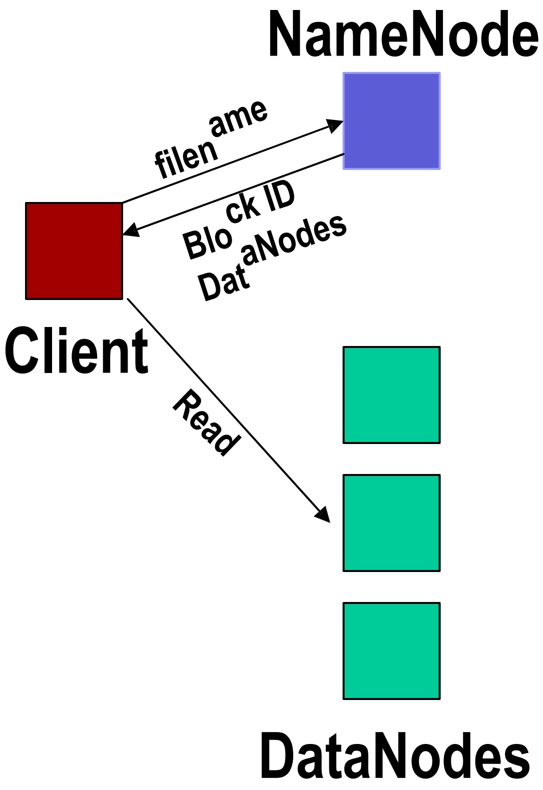
- **Open source MapReduce software platform**
- **Automatically provides framework for developing MapReduce applications**
  - Handles mapping and reducing logistics
  - Programmer just provides custom functionality
- **Currently takes custom functionality in Java and Python**
- **Uses an open source Eclipse plug-in to interface with Hadoop**

HDFS

- Very Large Distributed File System
  - 10K nodes, 100 million files, 10 PB
- Assumes Commodity Hardware
  - Files are replication in order to handle hardware failure
  - System detects failures and recovers from them
- Optimized for Batch Processing
  - Data locations exposed so that computations can move to where data resides
  - Provides very high aggregate bandwidth

Distributed File System

- Single Namespace for entire cluster
- Data Coherency
  - Write-once-read-many access model
  - Client can only append to existing files
- Files are broken up into blocks
  - Typically 128 MB block size
  - Each block replicated on multiple DataNodes
- Intelligent Client
  - Client can find location of blocks
  - Client accesses data directly from DataNode


NameNode
filename
Block ID
DataNodes
Client
Read
DataNodes