Cloud Computing

COSC 2639

Lecture 1

Introduction to Cloud Computing and System Architectures

Course Overview

- ❑ Offering Coordinator and Tutor: **Prabath Abeysekara**
- ❑ Email: prabath.abeysekara@rmit.edu.au
- ❑ Tutor: **Tim Lo**
- ❑ Email: tim.lo@rmit.edu.au
- ❑ Course code: **COSC2639/2697**
- ❑ 1 hr course webinar per week
- ❑ 1 hr class webinar per week

Pre-requisites, Assumed knowledge and Capabilities

- ❑ Good knowledge in programming (This is a MUST!)
- ❑ Good knowledge in Java/PHP/Python or any other programming language supported by GCP and AWS
- ❑ Basic knowledge in Net-centric Communication
- ❑ Basic knowledge in NodeJS
- ❑ Basic knowledge in Data Analytics

Important:

- ❑ Be able to access all Google (Cloud) Services (otherwise you CANNOT do Assessment 1 (30%)

Assessments

Assessment 01: Timed Practical Google Cloud Application Implementation (30%)

❑ Individual
❑ Specification released in early Week 2
❑ Due in Week 5
❑ Demonstration needed for assessment (Week 6)

Assessment 02: Timed Practical AWS Cloud Database Application Development (20%)

❑ Individual
❑ Specification released in early Week 6
❑ Due in Week 9
❑ Demonstration needed for assessment (Week 10)

Assessment 03: Timed Practical AWS Cloud System Development (50%)

❑ Individual
❑ Specification released in early Week 4
❑ Due in Week 12
❑ Demonstration needed for assessment (Week 13)

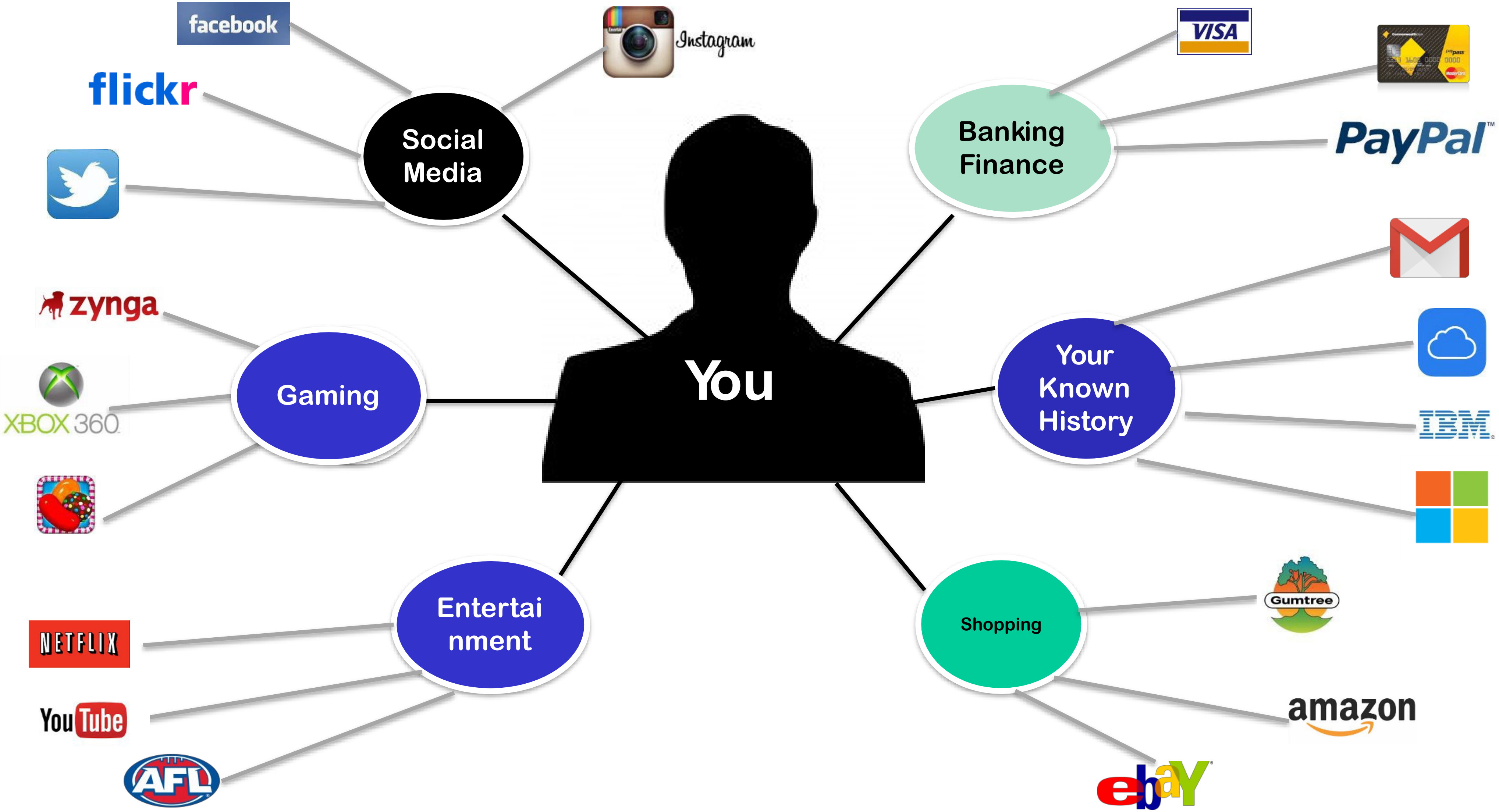
Use of Generative AI tools

- □ GenAI tools (e.g. ChatGPT) have become a part of everyone's learning journey, nowadays.

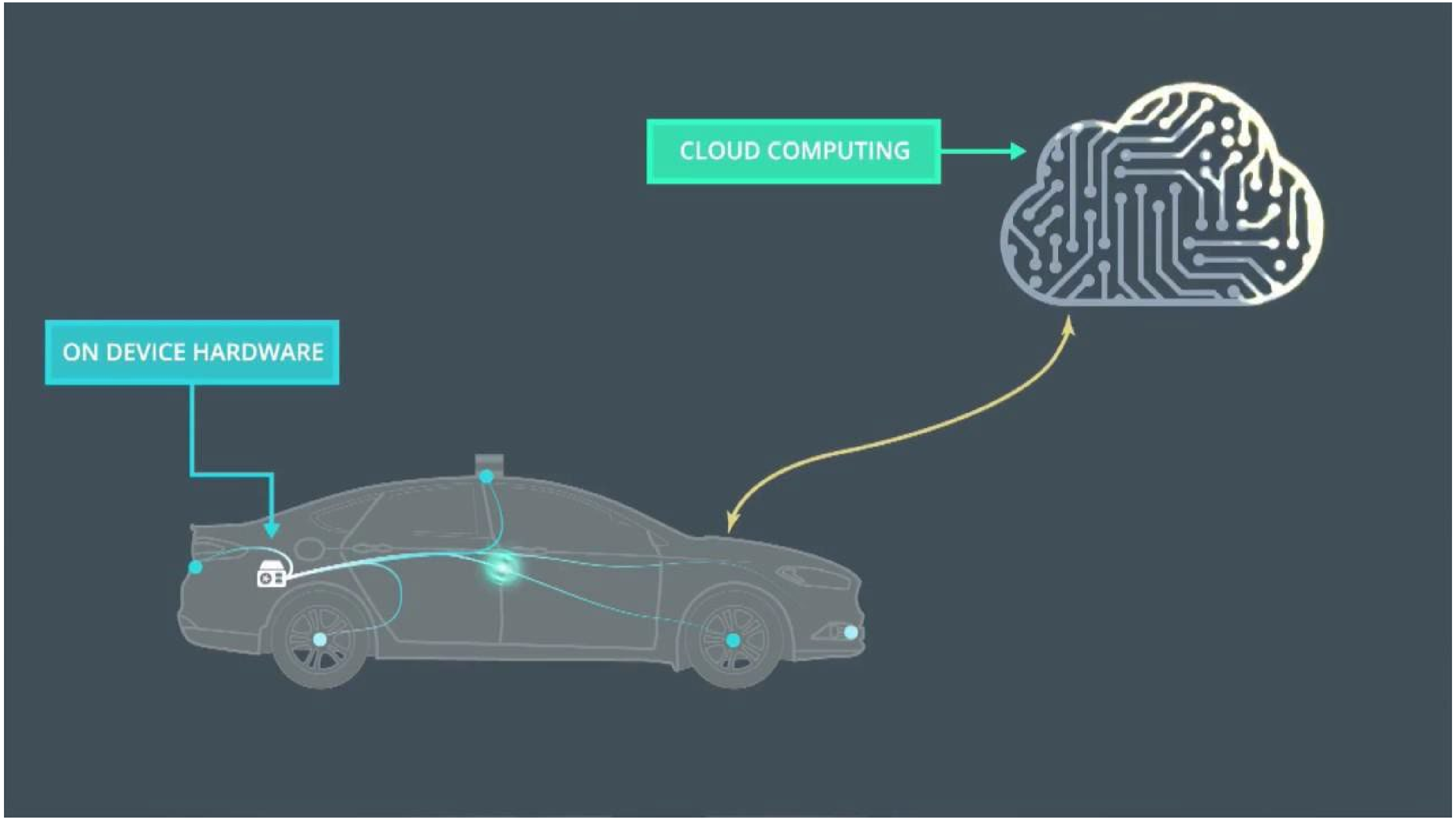
- □ Encourage using them as "assistive technologies" to make you more productive.

- □ Discourage using them to do the "assignments" for you! ☺

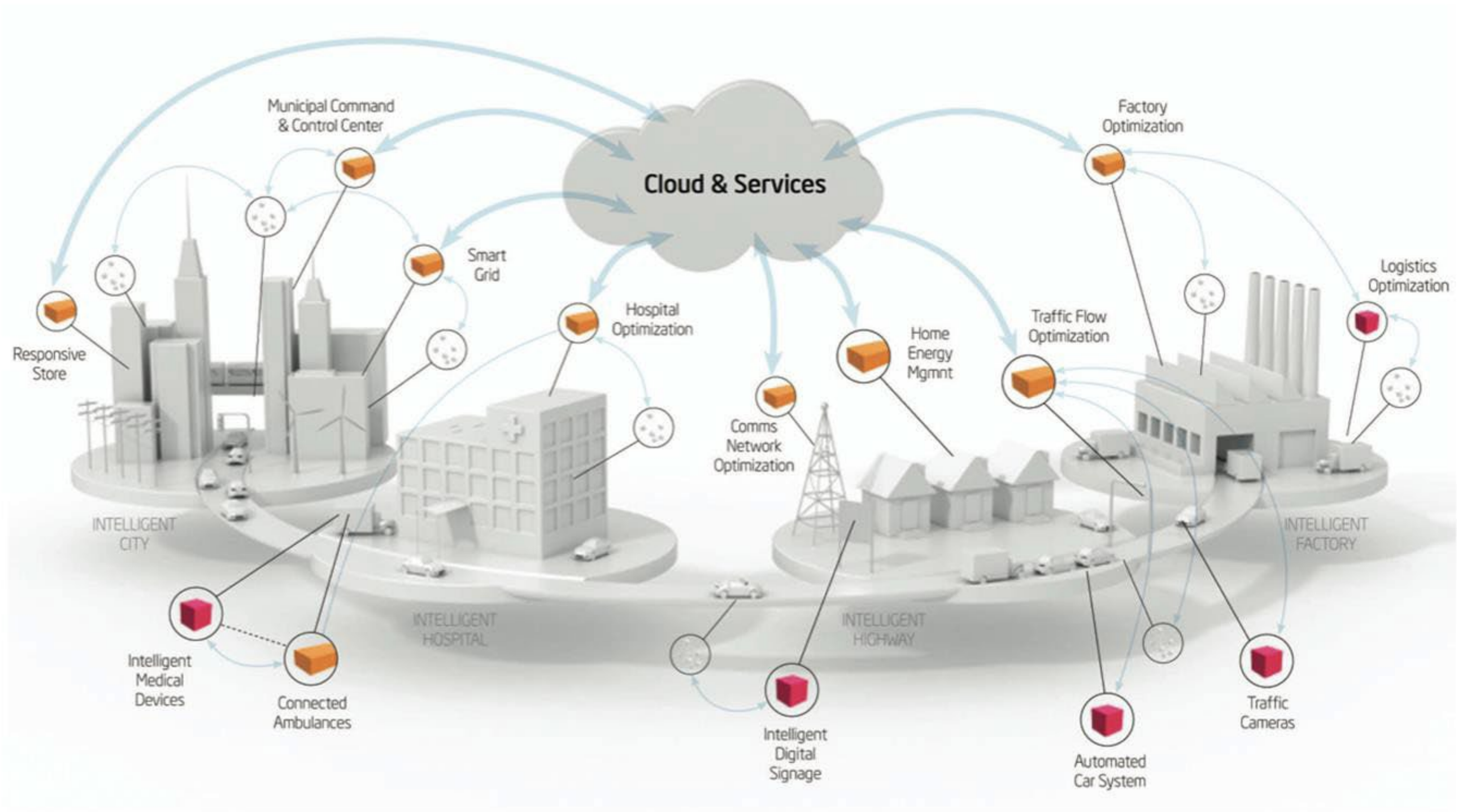
1 Cloud computing use cases



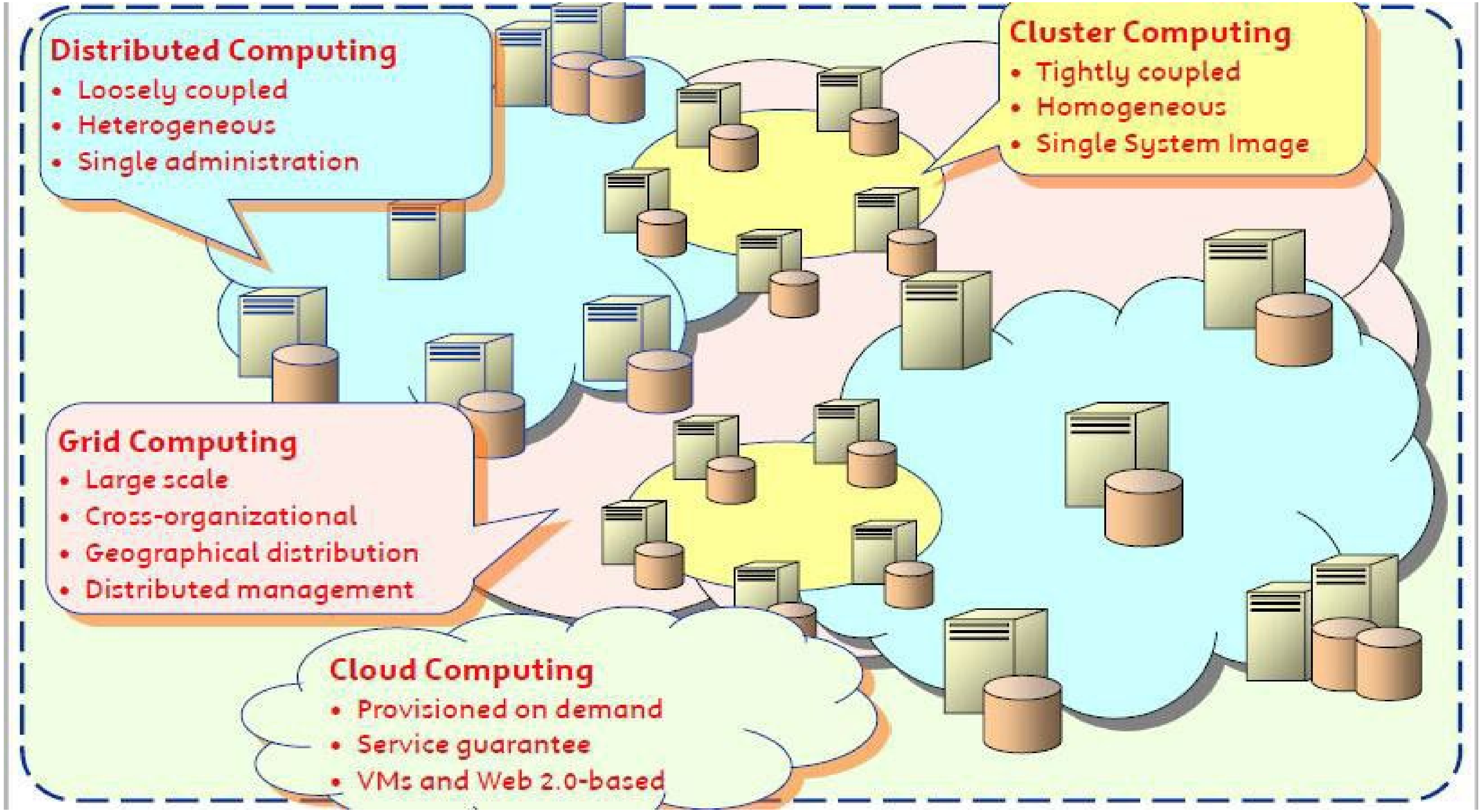
Use Case – Cloud accelerates AI: Self-driving car



Use Case – Cloud accelerates AI: Smart city



2 Cloud computing evolution



Distributed Computing
- Loosely coupled
- Heterogeneous
- Single administration

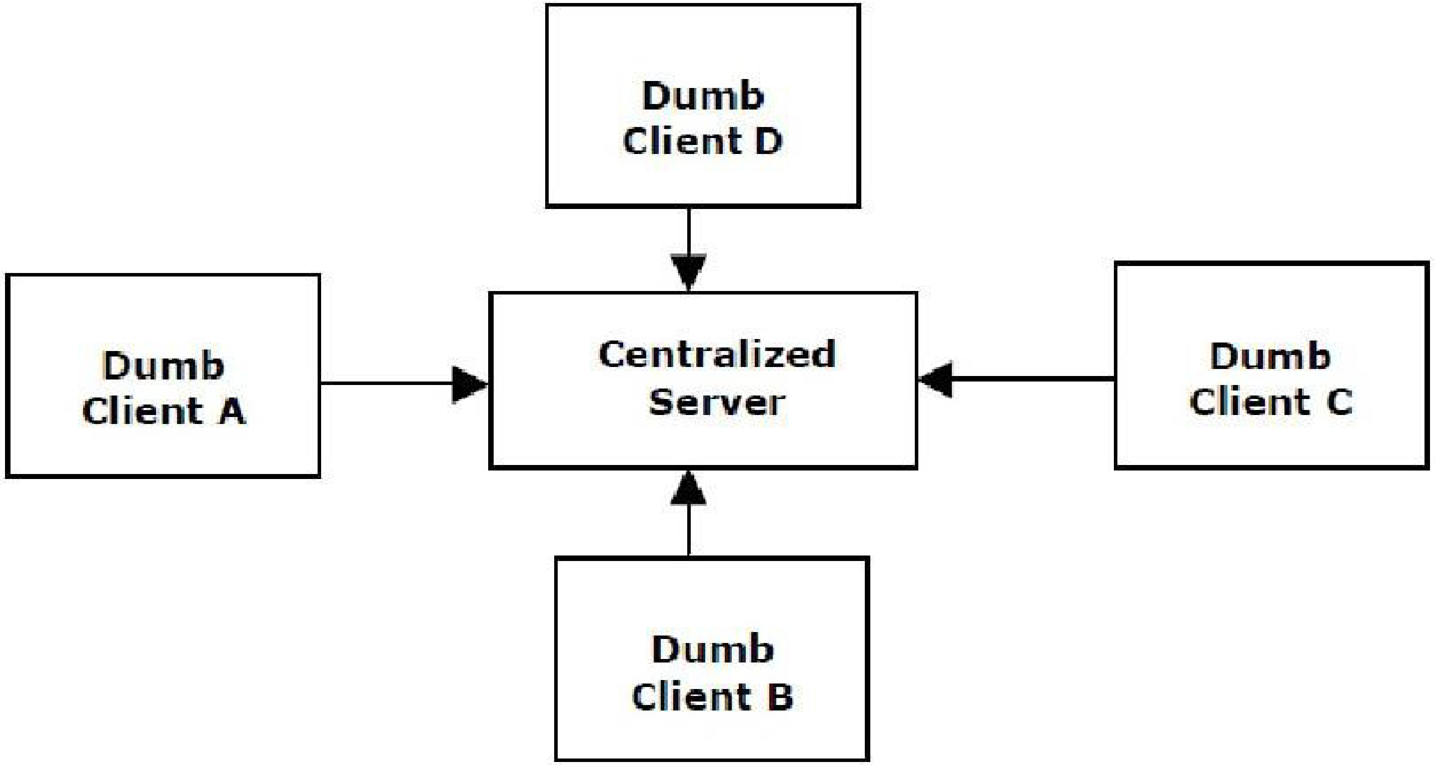
Cluster Computing
- Tightly coupled
- Homogeneous
- Single System Image

Grid Computing
- Large scale
- Cross-organizational
- Geographical distribution
- Distributed management

Cloud Computing
- Provisioned on demand
- Service guarantee
- VMs and Web 2.0-based

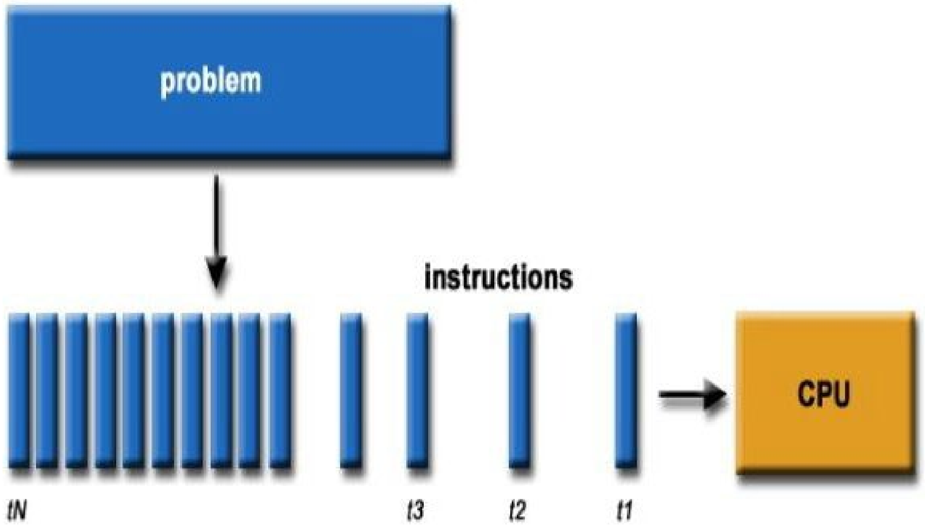
2.1 Centralized computing

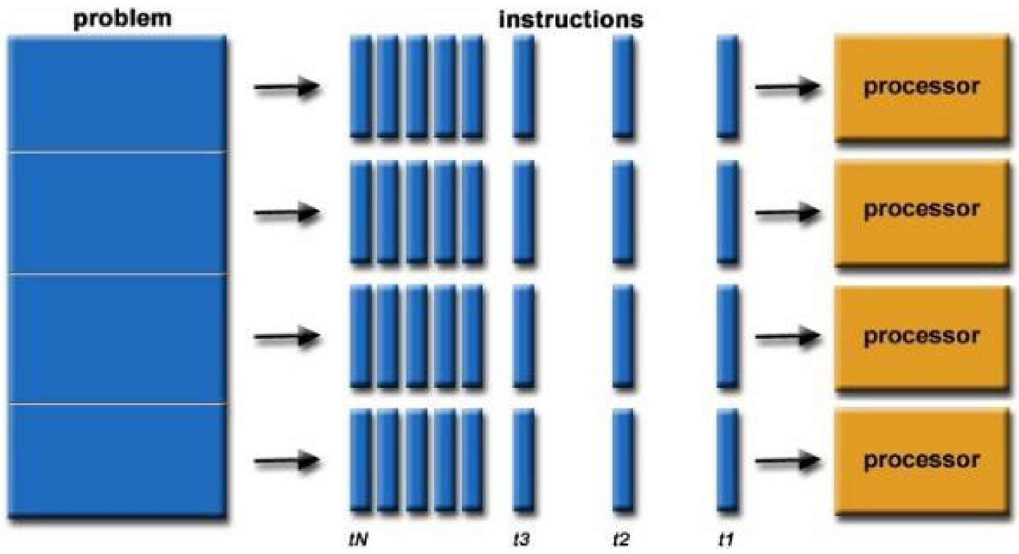
- All computer resources are centralized in one physical system.
- All resources (processors, memory, and storage) are tightly coupled within one integrated OS.
- Example: A single printer in a network of computers.



2.2 Parallel computing

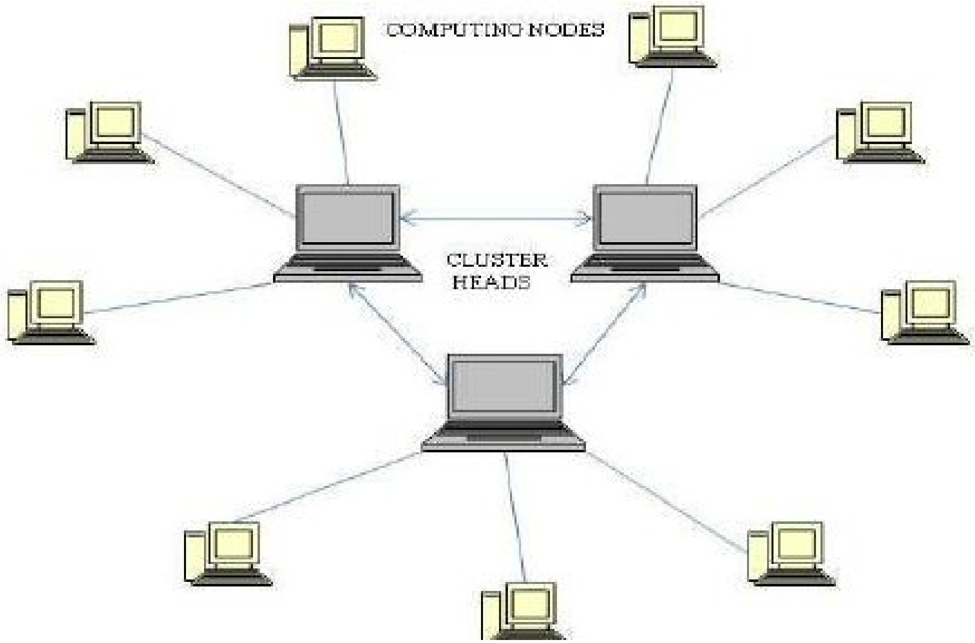
- Multiple processes can run in parallel using a shared memory
- all processors are tightly coupled with centralized memory.
- Inter-processor communication is accomplished through shared memory or via message passing.
- A computer system capable of parallel computing (has multiple processors) is commonly known as a parallel computer.



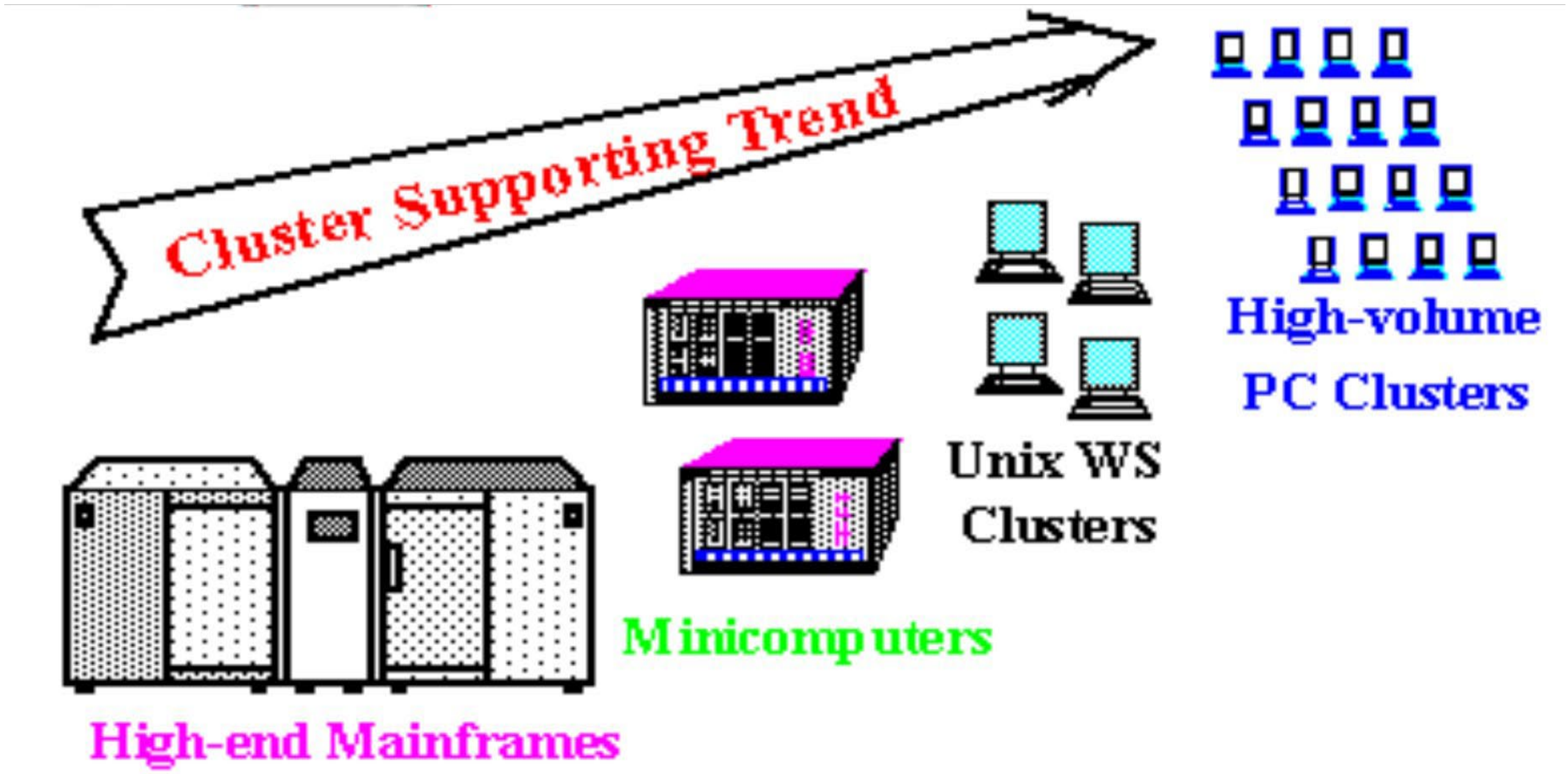


2.3 Cluster Computing

- ❑ Cluster is tightly-coupled
- ❑ Resources are homogenous or identical
- ❑ Work together for a specific job
- ❑ Each node performs the same task
- ❑ Clustering explores massive parallelism at the job level and achieves high availability (HA) through stand-alone operations.

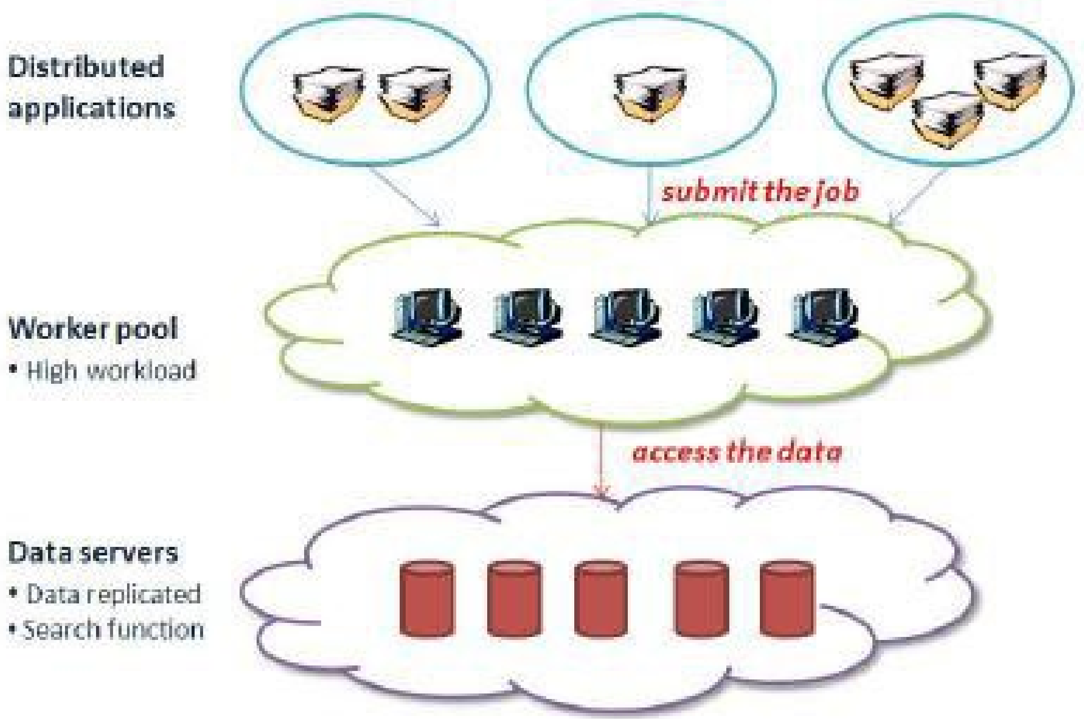


Evolution of cluster computing



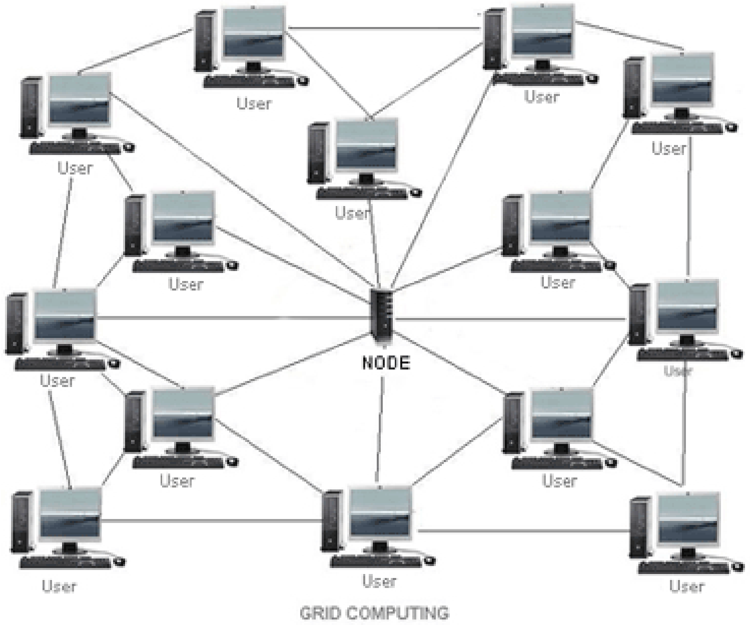
2.4 Distributed computing

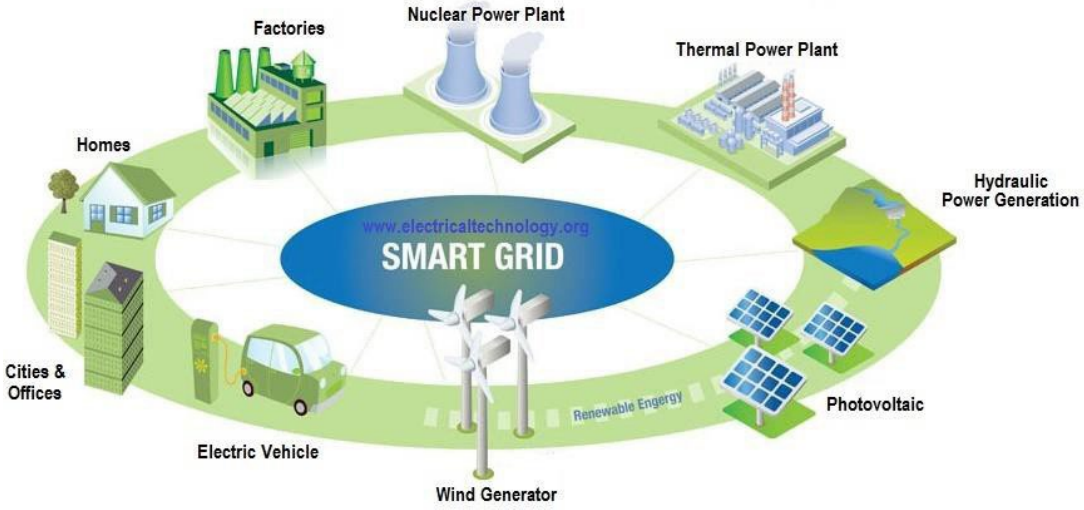
- Consists of multiple autonomous computers.
- Each computer has its own private memory.
- Communicate through a computer network.
- Information exchange in a distributed system is accomplished through message passing.
- Originated from distributed system is – Distributed programming



2.5 Grid computing

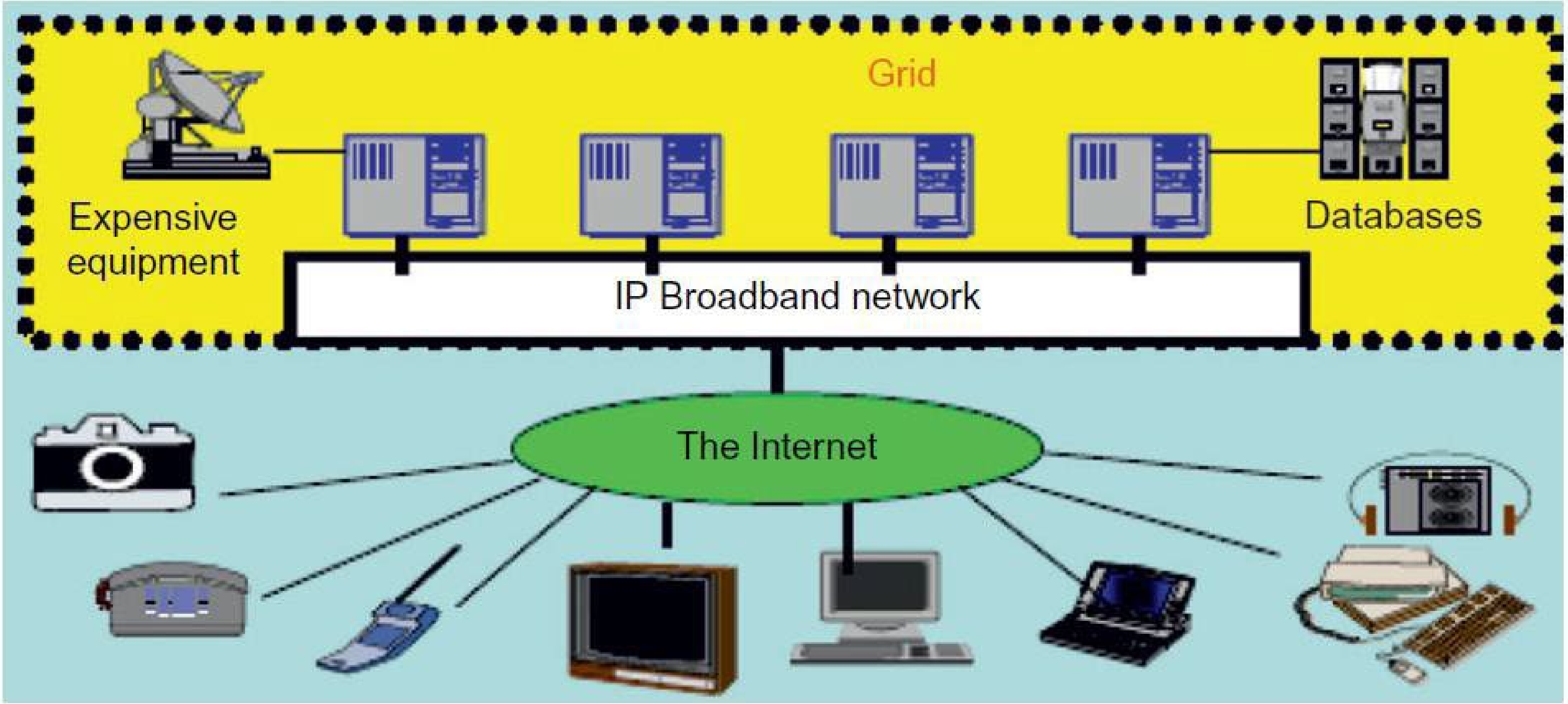
- Grid Computing
  - A variation of distributed computing
  - Operates at a much larger scale
  - Cross-organizational
  - Geographically distributed
  - Distributed management





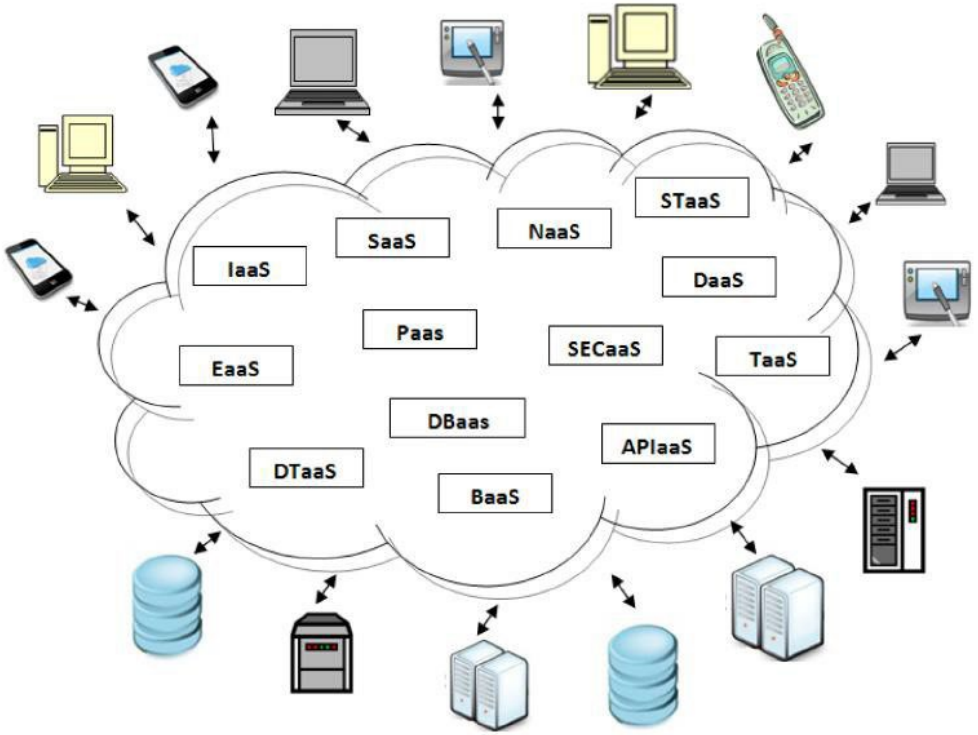
Grid Computing

- ❑ Grid systems are classified in two categories: computational or data grids and P2P grids.



2.6 Cloud Computing

- It can be centralized or distributed. How?
  - Cloud can be built with physical or virtualized resources over large data centers that are centralized or distributed.
- Applies parallel or distributed computing or combination of both
- Provisioned on-demand
- Service guarantees
- Three layer-ed services



2.7 Mobile Computing

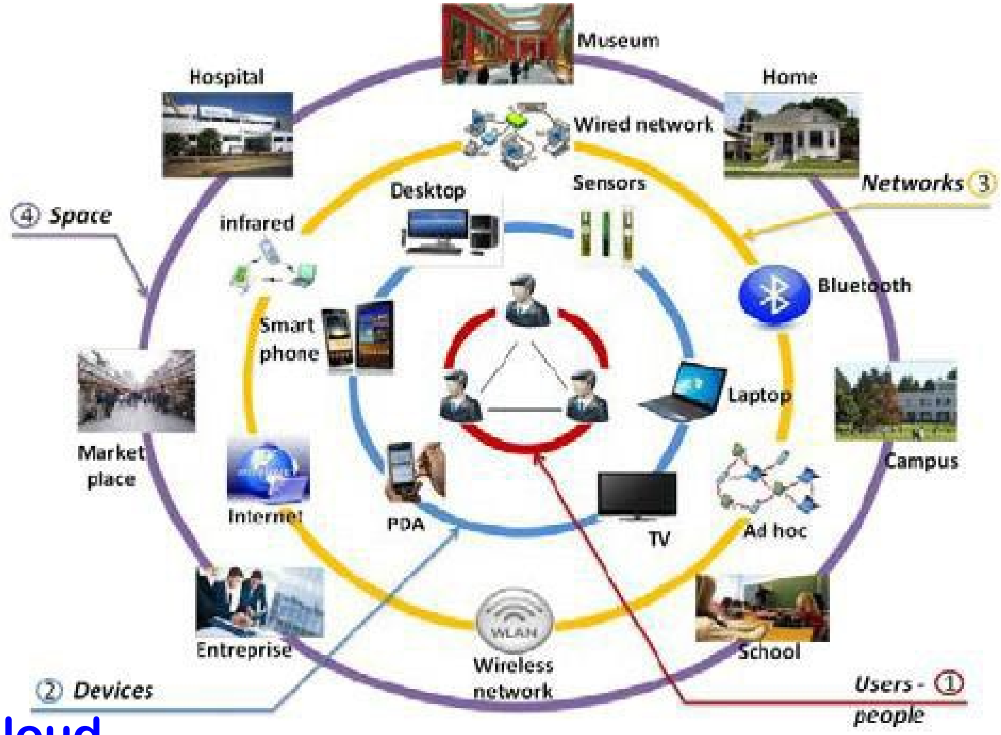
- The computing environment is mobile and moves with users.
- Example?

2.8 Ubiquitous computing

- ❑ Wireless communication between the components
- ❑ Refers to computing with pervasive

devices at any place and time using wired or wireless communication.

- ❑ Also known as Pervasive computing.
- ❑ Example?
  - ○ Smart Traffic Light
  - ○ Home automation system
  - ○ Fitbit!



Bing Huang, Athman Bouguettaya, Hai Dong: Enabling Edge Cloud Intelligence for Activity Learning in Smart Home (2020)

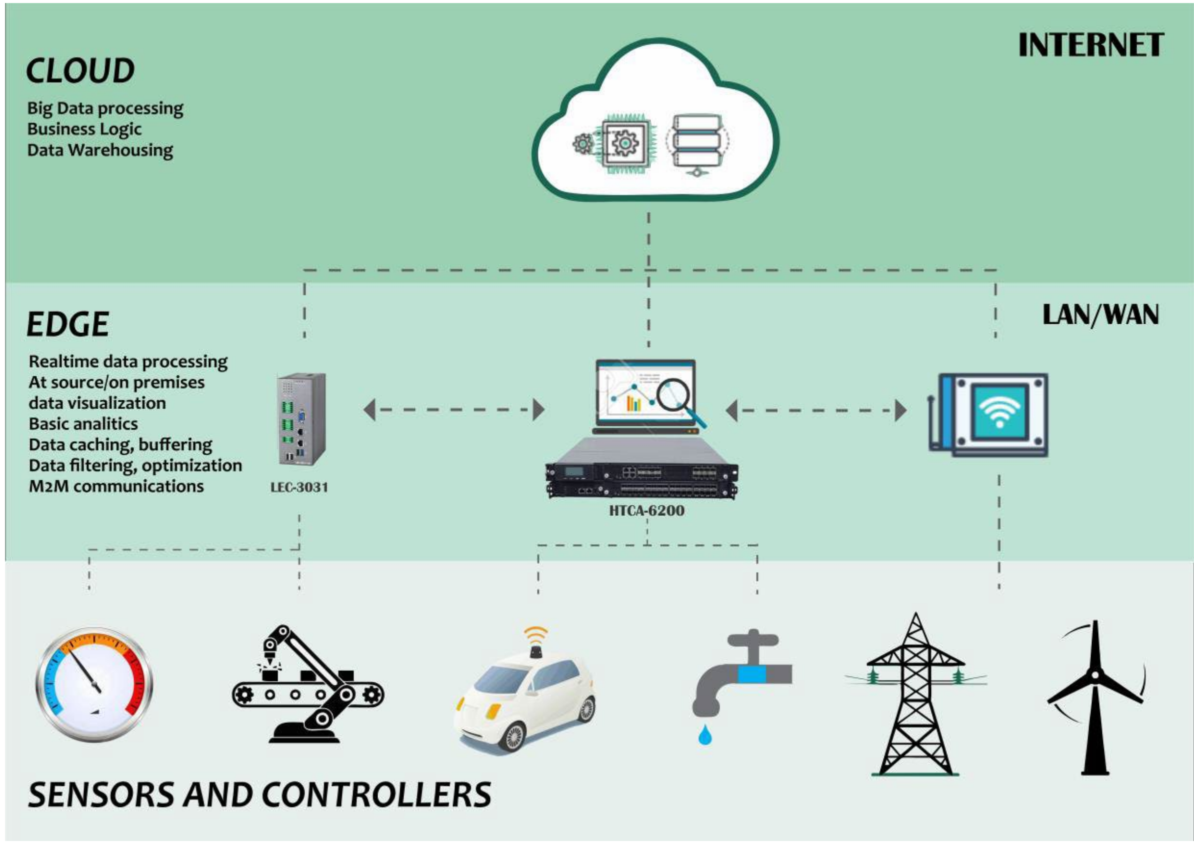
2.9 Edge computing

❑Local processing before reaching cloud

❑Complements the concept of IoT

❑Data from IoT devices are analyzed at the edge before reaching the cloud.

❑ Sensors, controllers, and other connected devices collect and analyze IoT data themselves, or transmit it to a near by computing computing device (server, laptop, Raspberry Pi).



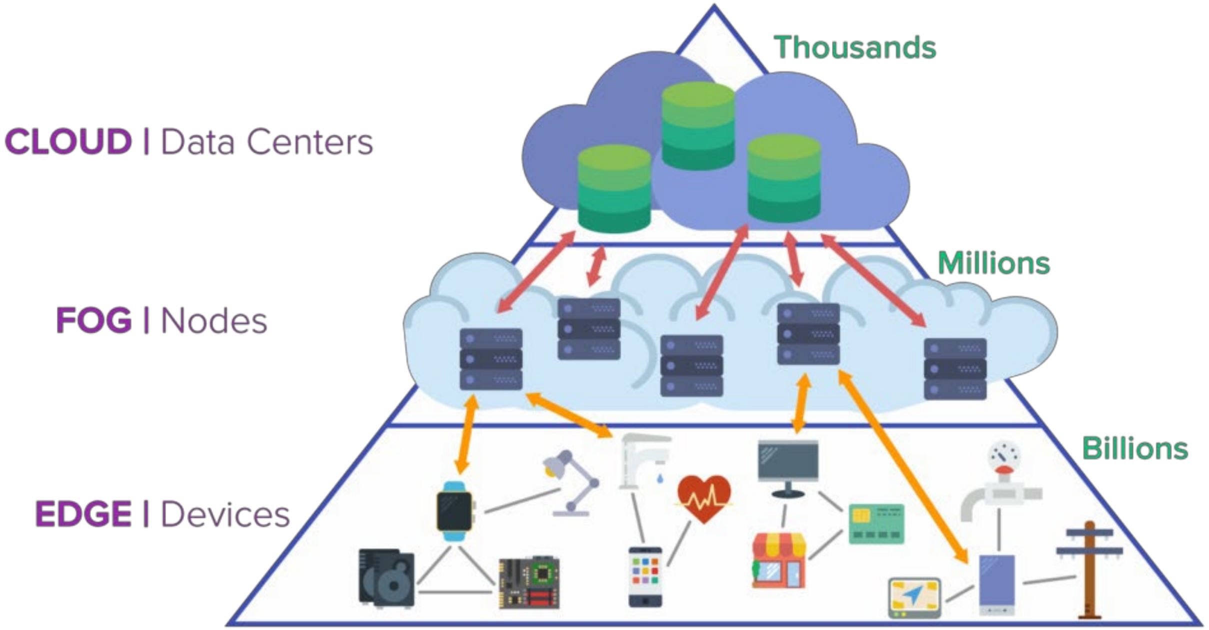
[Abeysekara, P., Dong, H., Qin, A. K.: Distributed Machine Learning for Predictive Analytics in Mobile Edge Computing Based IoT Environments. The 2020 International Joint Conference on Neural Networks (IJCNN 2020) (July 2020)]()

2.10 Fog computing

❑ A small cloud before data go the actual cloud

❑ Enterprise network

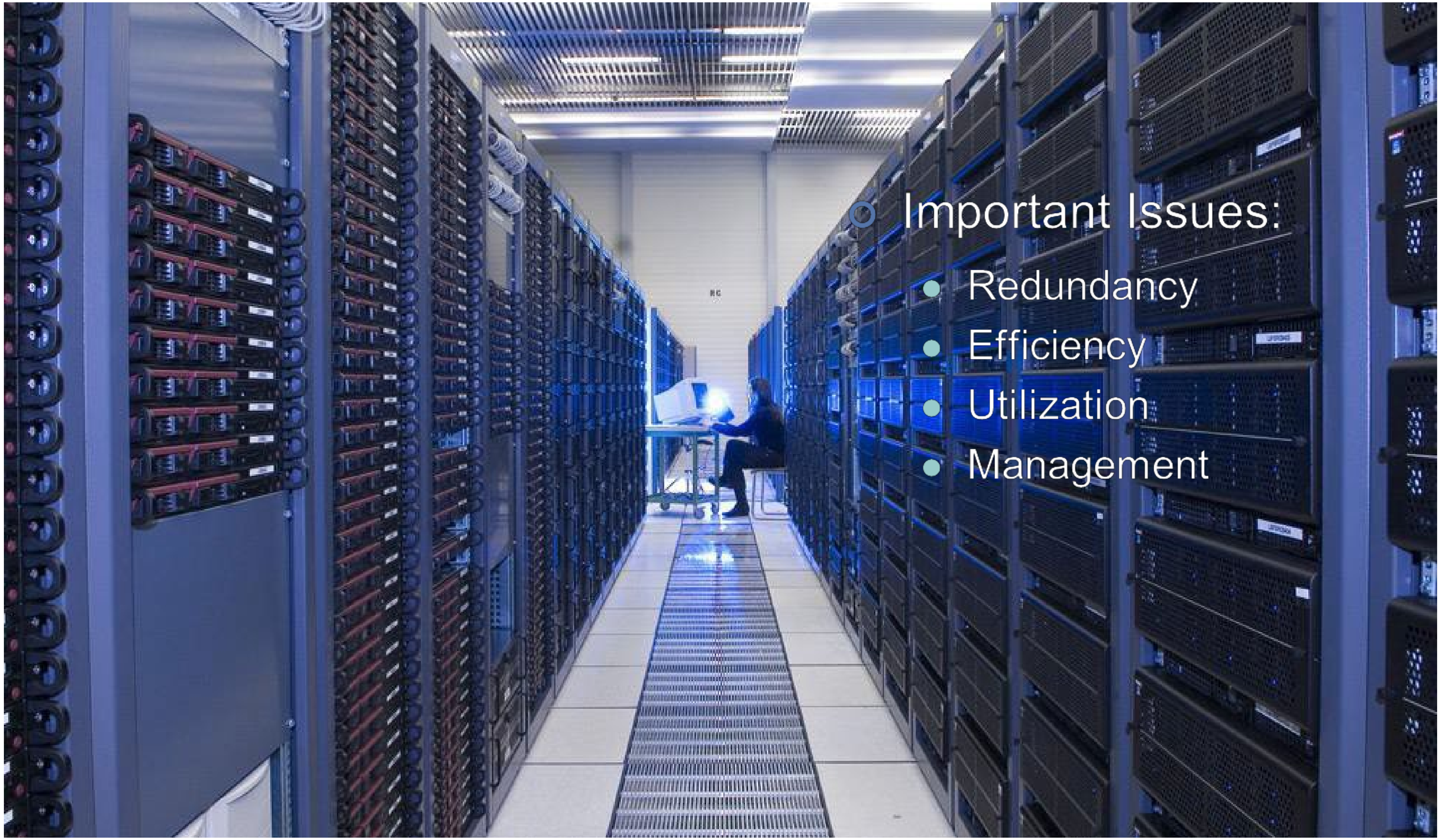
❑ Provide computing, storage and networking services for data processing.



3 Essential Cloud Characteristics

- □ Large data centers
- □ Virtualization
- □ Pay-as-you go (reduced cost)
- □ Scalable and on-demand services
- □ Broad network access
- □ Resource pooling
  - ○ Location independence
- □ Rapid elasticity
- □ Measured service

3.1 Large Data Centers

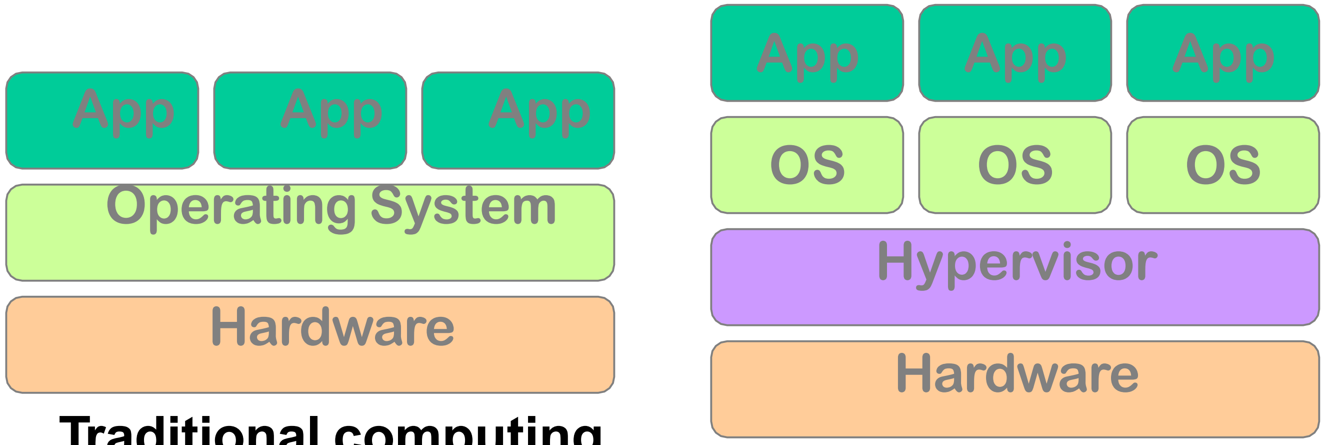


- Important Issues:
  - Redundancy
  - Efficiency
  - Utilization
  - Management

Large Data Centers

- Data centers have grown rapidly in recent years.
- Google, Yahoo!, Amazon, Microsoft, HP, Apple, and IBM are all companies have invested billions of dollars in datacentre construction and automation.
- Huge volumes of hardware, software, and database resources in these data centers can be allocated dynamically to millions of Internet users simultaneously.
- Guaranteed QoS and cost-effectiveness.

3.2 Virtualization



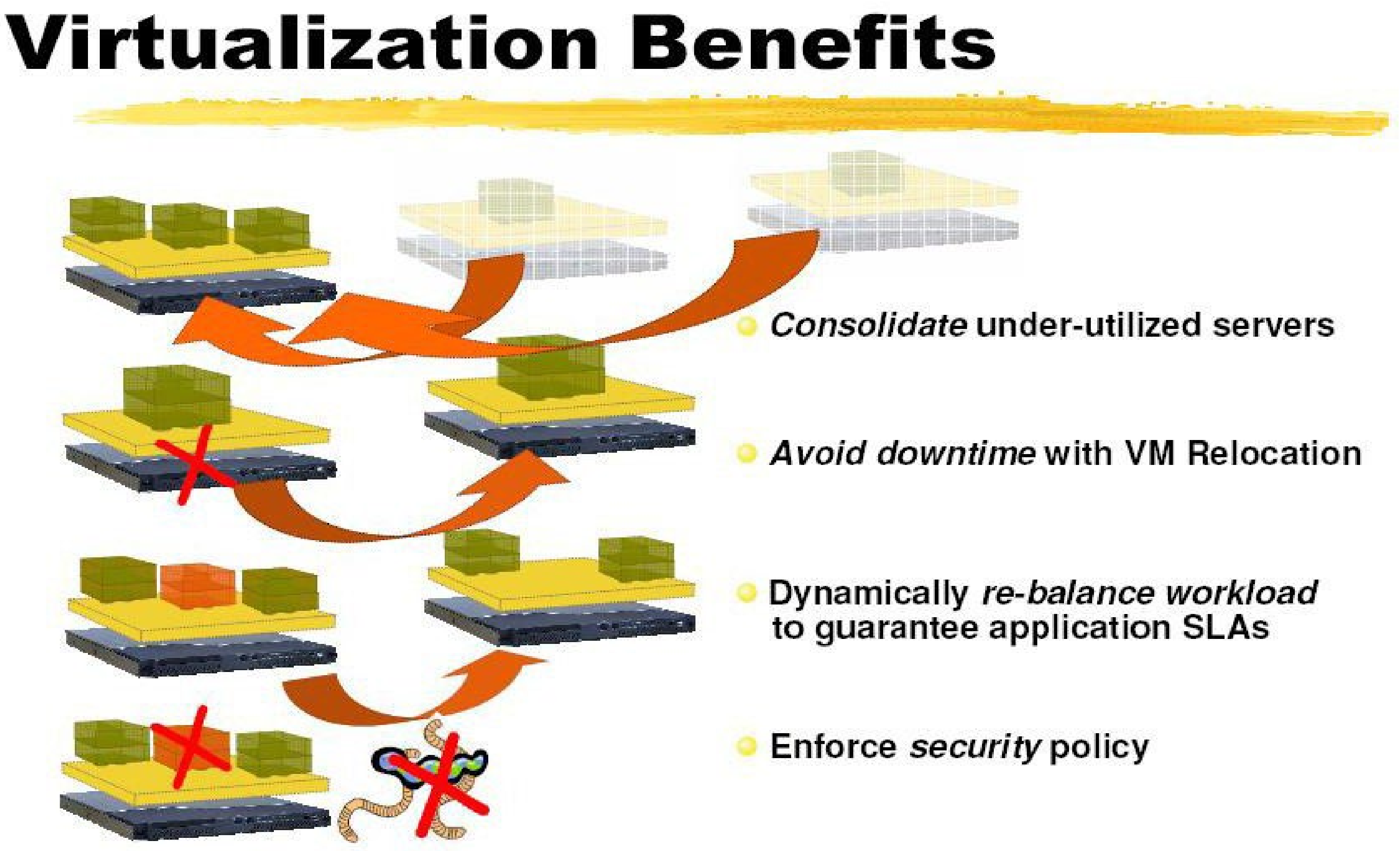
Traditional computing environment

Virtualized computing environment

A hypervisor or virtual machine monitor (VMM) is a piece of computer software, firmware or hardware that creates and runs virtual machines.

- ❑ **Hypervisor** manages multiple operating systems (or multiple instances of the same operating system) on a single computer system.
- ❑ The hypervisor manages the system's processor, memory, and other resources to allocate what each operating system requires.
- ❑ Hypervisors are designed for a particular processor architecture and may also be called **virtualization managers**.

Benefits of Virtualization

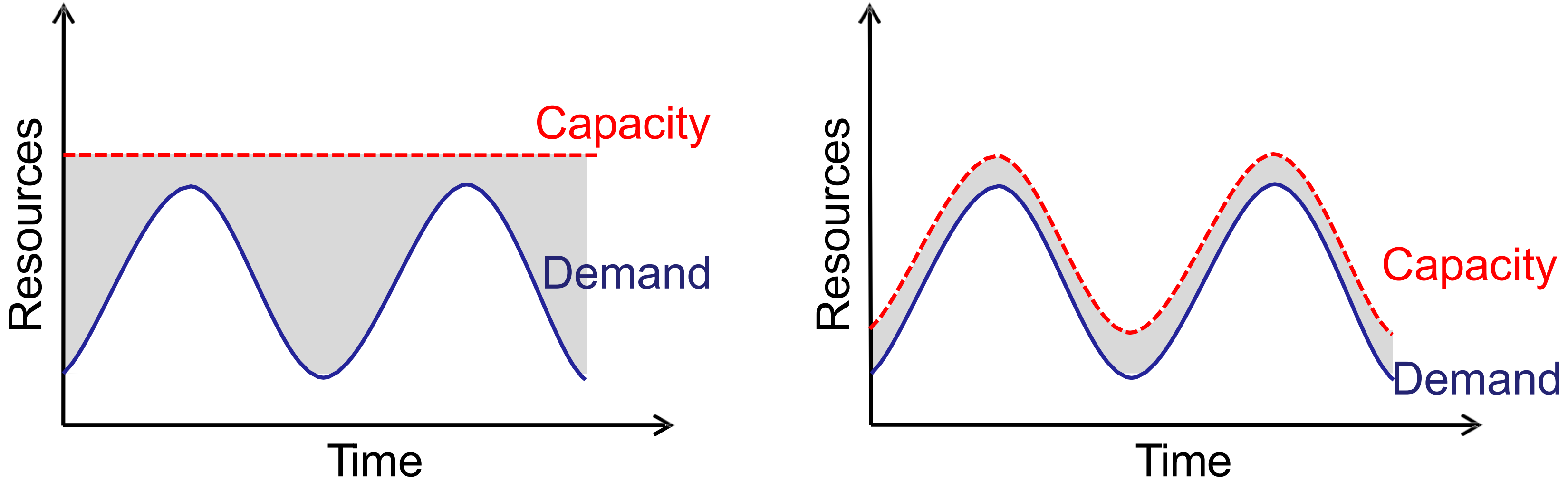


Virtualization Benefits

- *Consolidate* under-utilized servers
- *Avoid downtime* with VM Relocation
- Dynamically *re-balance workload* to guarantee application SLAs
- Enforce *security* policy

Benefits of Virtualization & Cloud Economics

Pay by use instead of provisioning for peak



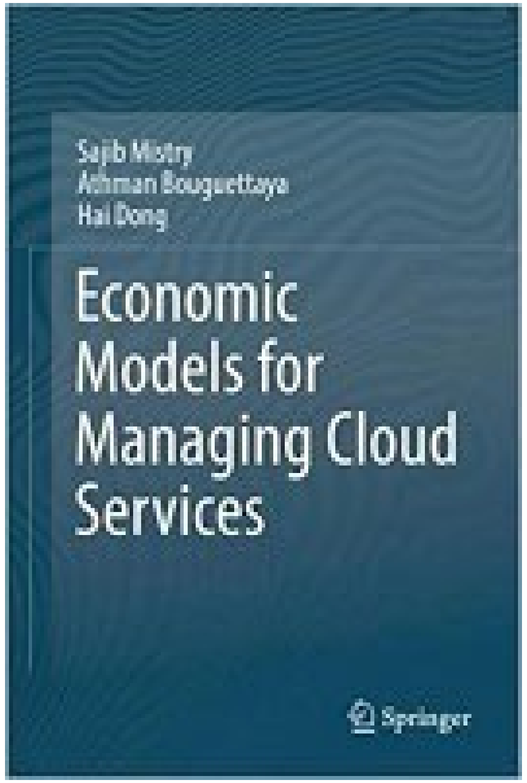
Static data center

Data center in the cloud

▨ Unused resources

Benefits of Virtualization & Cloud Economics

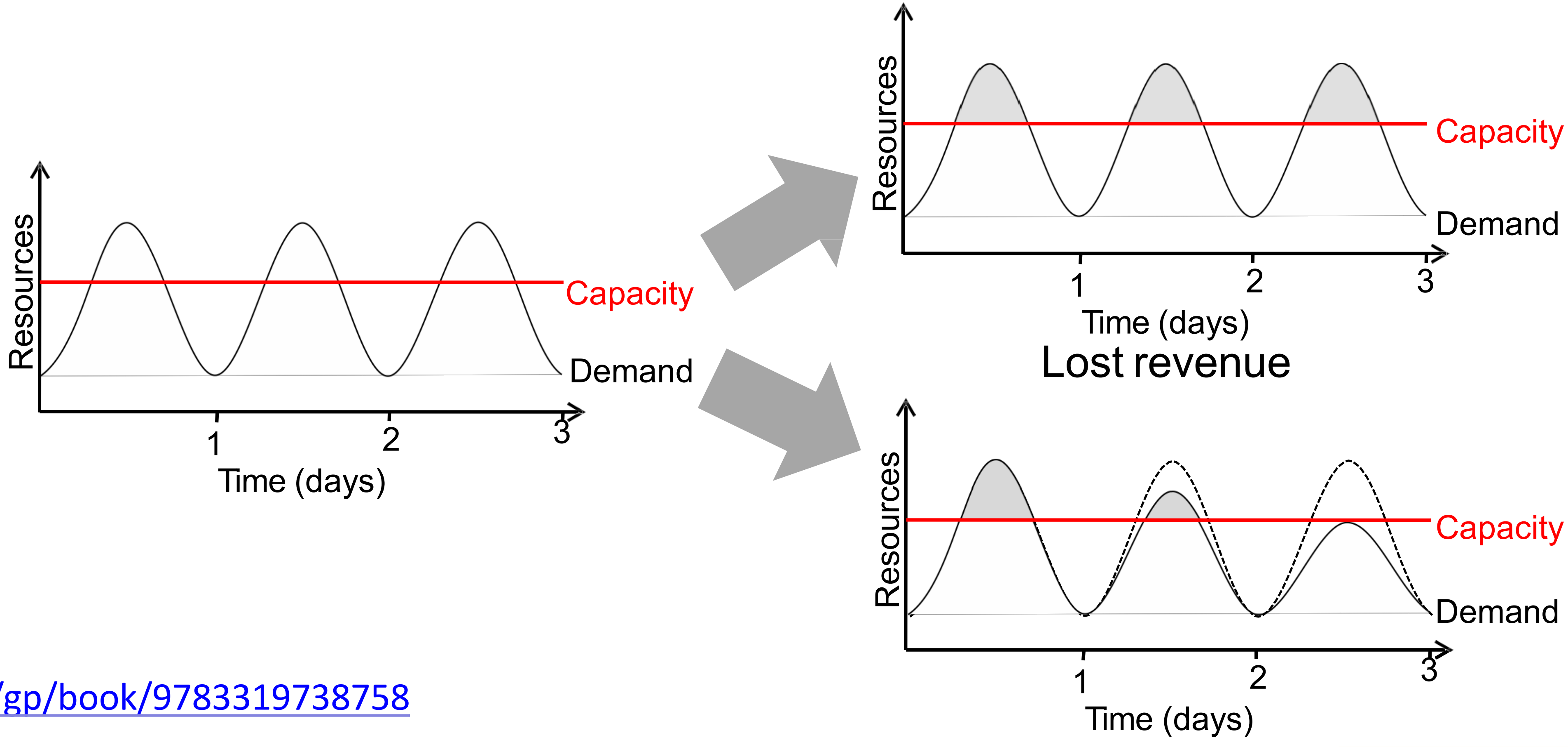
Heavy penalty for under-provisioning



Sajib Mistry
Athman Bouguettaya
Hai Dong

Economic Models for Managing Cloud Services

Springer



Lost revenue

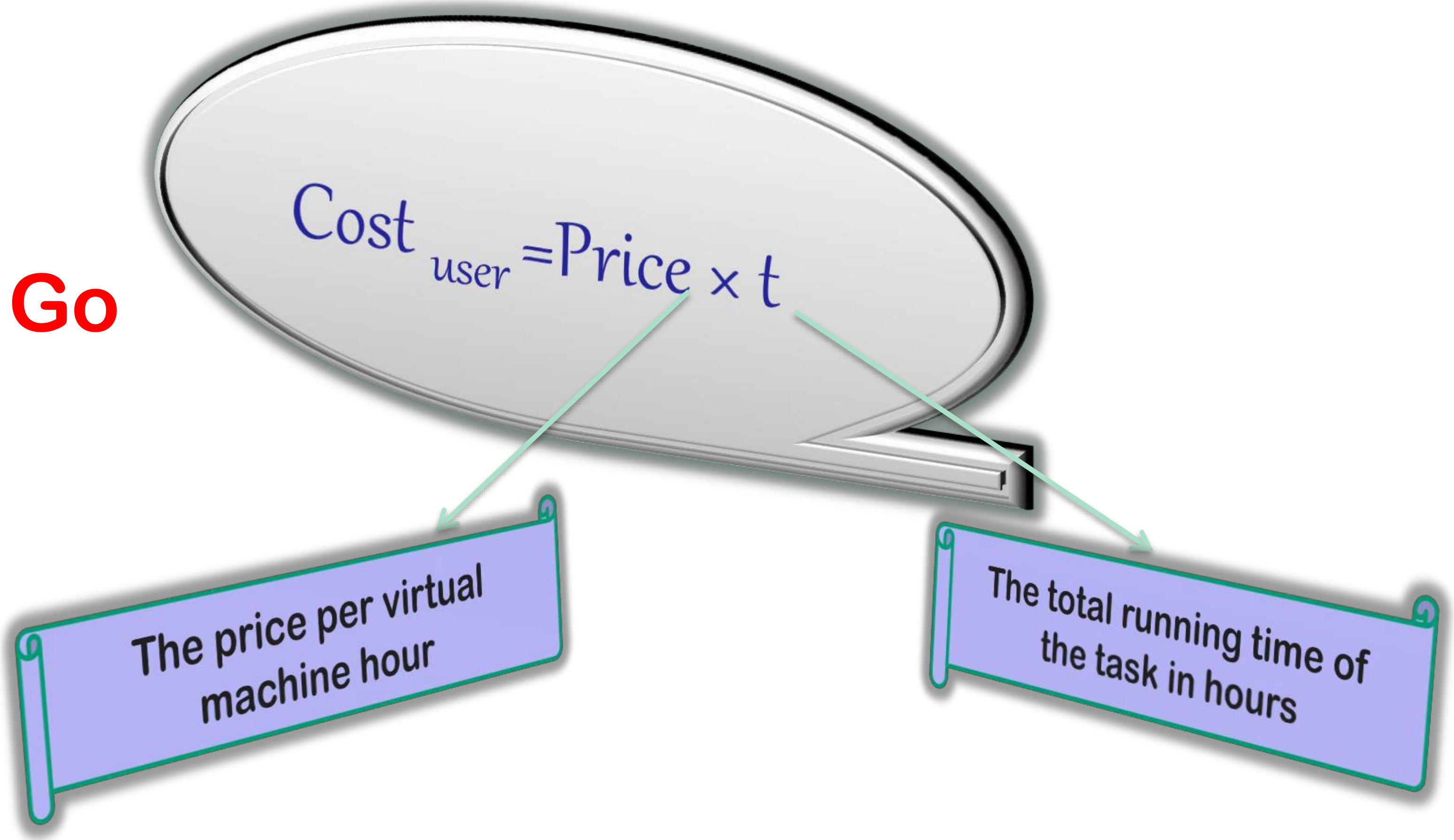
Lost users

https://www.springer.com/gp/book/9783319738758

3.3 Pay as you go

We calculate users' expenses when they execute a task in Amazon:

Pay as you Go



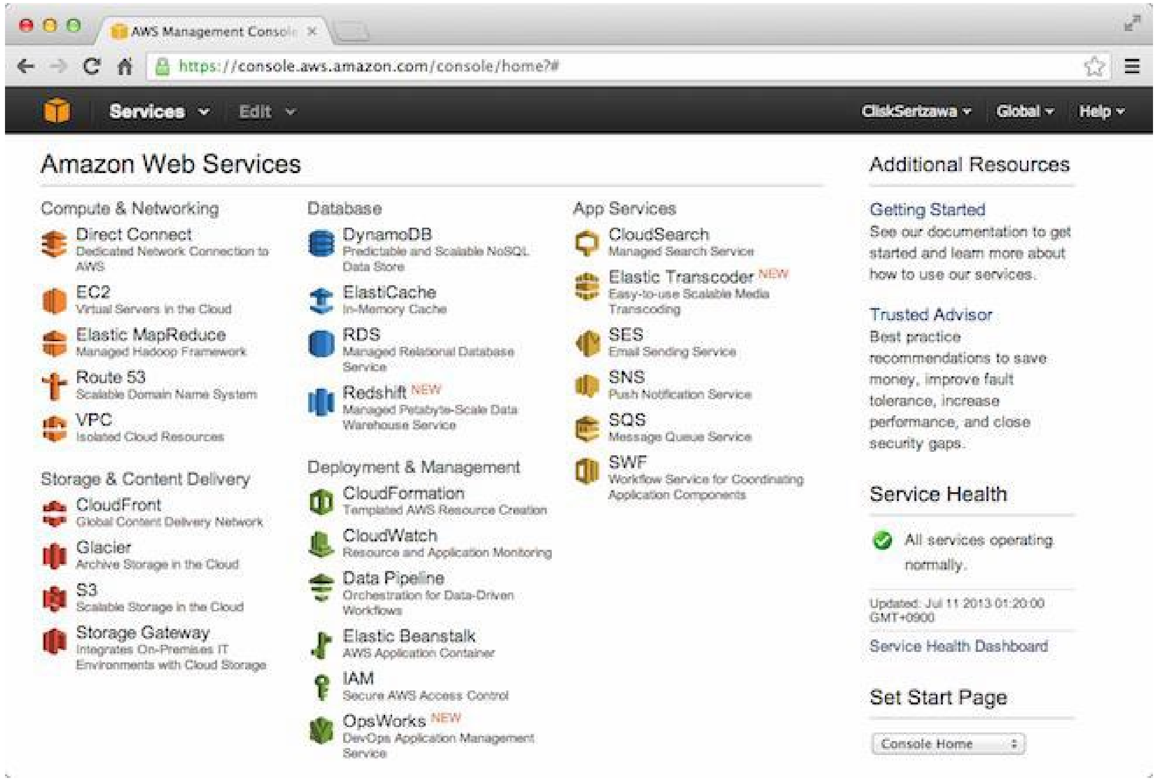
$$Cost_{user} = Price \times t$$

The price per virtual machine hour

The total running time of the task in hours

3.4 On Demand Self-Service

- ❑ A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider.
- ❑ Example: AWS Management Console



Amazon Web Services

Compute & Networking

- Direct Connect — Dedicated Network Connection to AWS
- EC2 — Virtual Servers in the Cloud
- Elastic MapReduce — Managed Hadoop Framework
- Route 53 — Scalable Domain Name System
- VPC — Isolated Cloud Resources

Storage & Content Delivery

- CloudFront — Global Content Delivery Network
- Glacier — Archive Storage in the Cloud
- S3 — Scalable Storage in the Cloud
- Storage Gateway — Integrates On-Premises IT Environments with Cloud Storage

Database

- DynamoDB — Predictable and Scalable NoSQL Data Store
- ElastiCache — In-Memory Cache
- RDS — Managed Relational Database Service
- Redshift NEW — Managed Petabyte-Scale Data Warehouse Service

Deployment & Management

- CloudFormation — Templated AWS Resource Creation
- CloudWatch — Resource and Application Monitoring
- Data Pipeline — Orchestration for Data-Driven Workflows
- Elastic Beanstalk — AWS Application Container
- IAM — Secure AWS Access Control
- OpsWorks NEW — DevOps Application Management Service

App Services

- CloudSearch — Managed Search Service
- Elastic Transcoder NEW — Easy-to-use Scalable Media Transcoding
- SES — Email Sending Service
- SNS — Push Notification Service
- SQS — Message Queue Service
- SWF — Workflow Service for Coordinating Application Components

Additional Resources

Getting Started
See our documentation to get started and learn more about how to use our services.

Trusted Advisor
Best practice recommendations to save money, improve fault tolerance, increase performance, and close security gaps.

Service Health

✓ All services operating normally.

Updated: Jul 11 2013 01:20:00 GMT+0900

Service Health Dashboard

Set Start Page

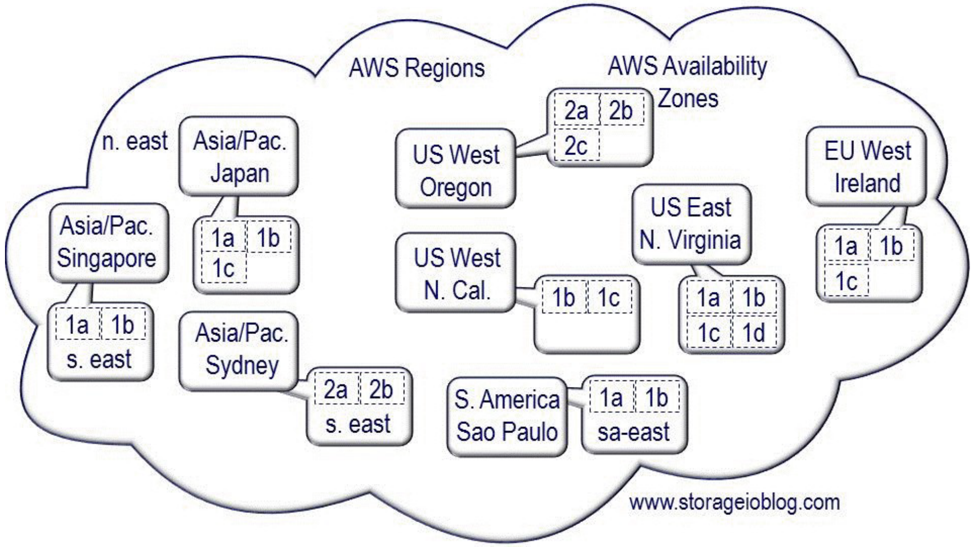
Console Home

3.5 Broad Network Access

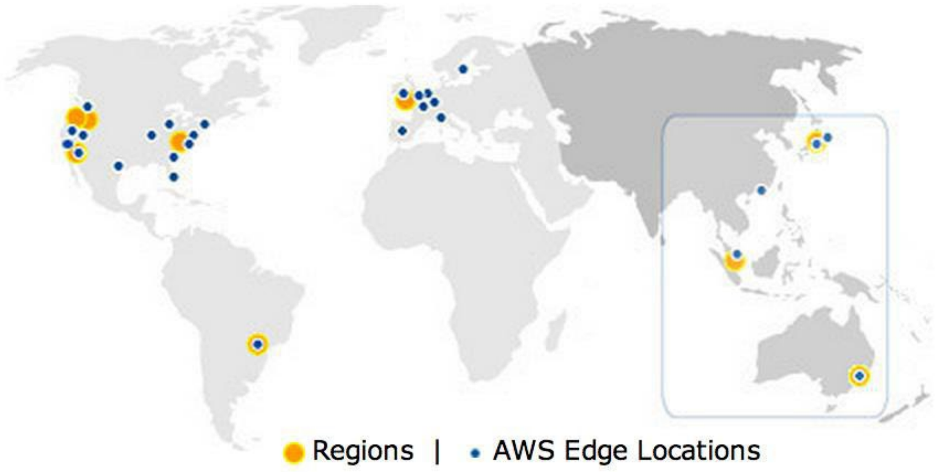
- Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms.
- Example: Mobile phones, tablets, laptops, computers

3.6 Resource Pooling

- ❑ The provider’s computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand
- ❑ Examples: AWS availability zones

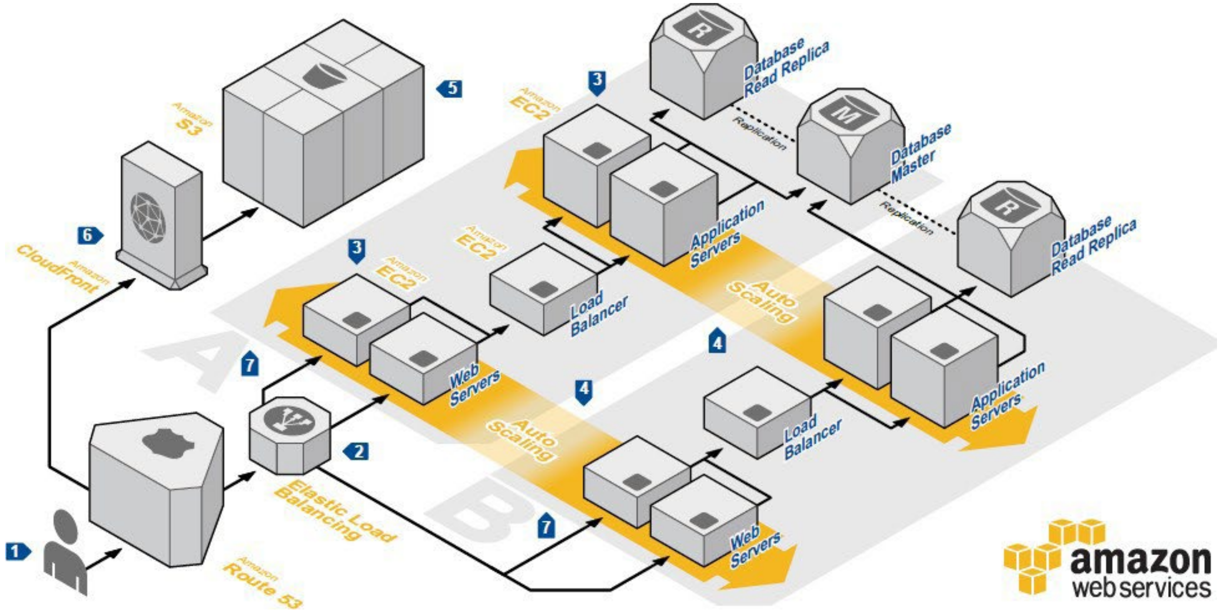


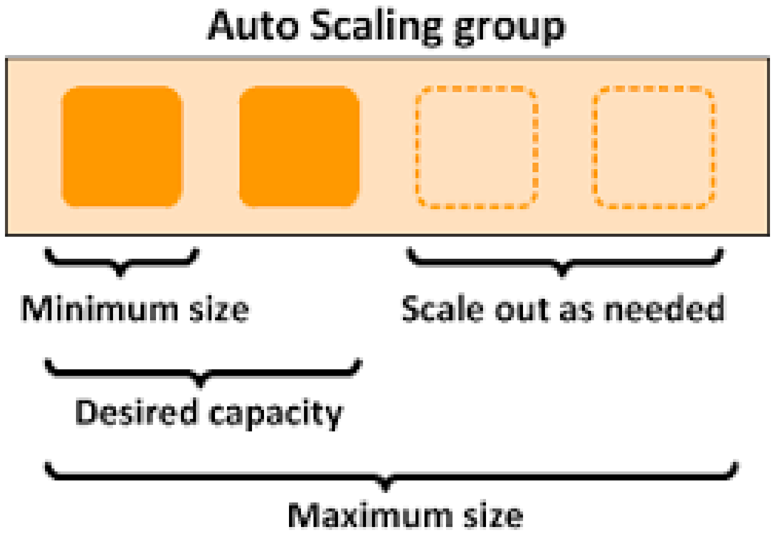
www.storageioblog.com



3.7 Rapid Elasticity

- Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand
- Examples: AWS auto-scaling (will learn more in future lecture)





RMIT UNIVERSITY

3.8 Measured Service

- ❑ Cloud systems automatically control and optimize resource by leveraging a metering capability
- ❑ Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilized service.
- ❑ Example: Pay-as-you-Go model

4 Cloud Service Models: IaaS, PaaS, SaaS

- ❑ Cloud Infrastructure as a Service (IaaS)
  - ❍ Rent processing, storage, network capacity, and other fundamental computing resources
- ❑ Cloud Platform as a Service (PaaS)
  - ❍ Deploy customer-created applications to a cloud
- ❑ Cloud Software as a Service (SaaS)
  - ❍ Use provider's applications over a network
- ❑ To be considered "cloud" they must be deployed on top of cloud infrastructure that has the key characteristics

- ❑ Utility computing
  - ❍ Why buy machines when you can rent cycles?
  - ❍ Examples: Amazon's EC2, GoGrid, AppNexus
- ❑ Platform as a Service (PaaS)
  - ❍ Give me nice API and take care of the implementation
  - ❍ Example: Google App Engine
- ❑ Software as a Service (SaaS)
  - ❍ Just run it for me!
  - ❍ Example: Gmail, Facebook

Cloud Service Model Hierarchy

SalesForce CRM

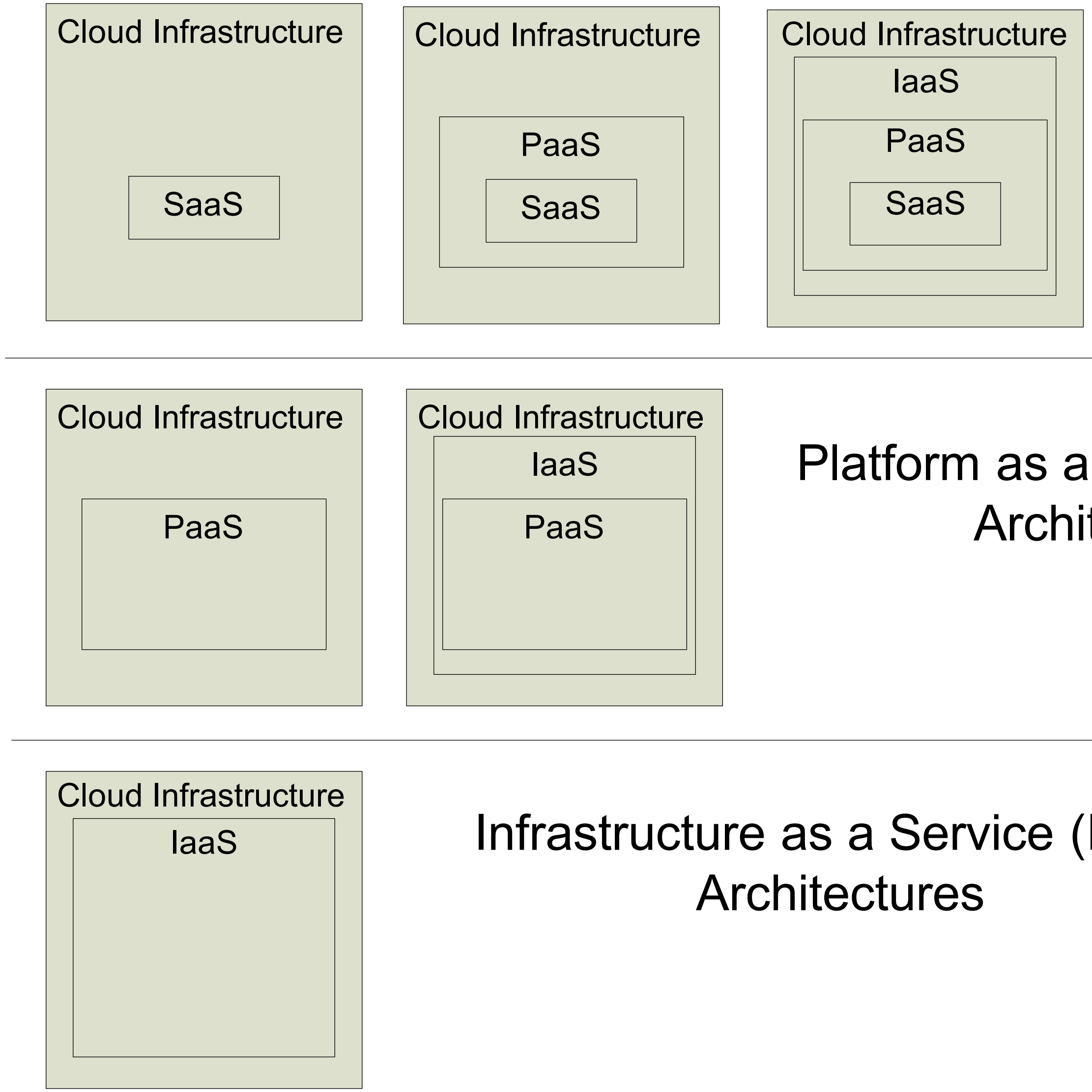
LotusLive

Google App

Windows Azure The Future Made Familiar

amazon web services

rackspace HOSTING



Software as a Service (SaaS) Architectures

Platform as a Service (PaaS) Architectures

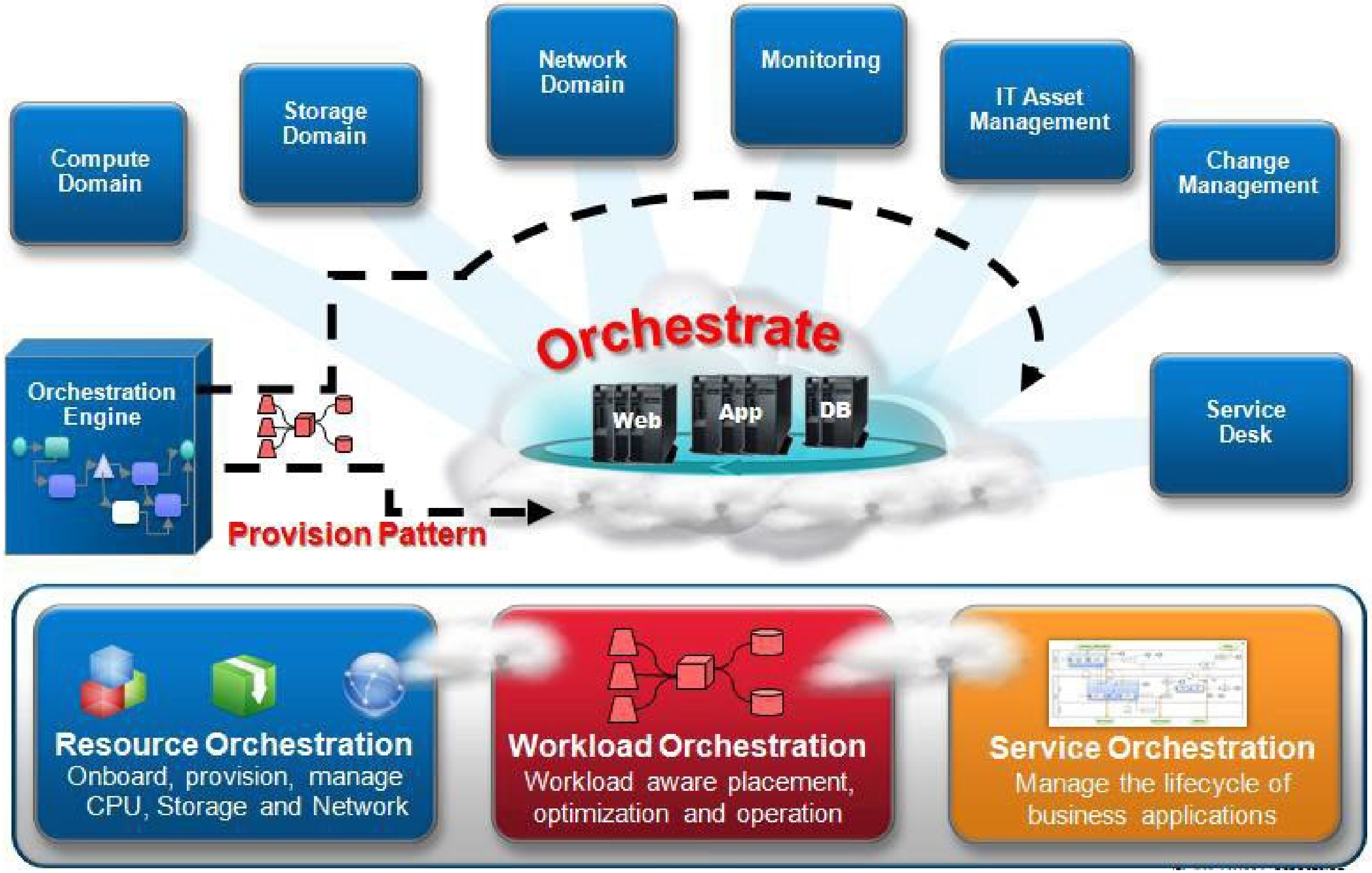
Infrastructure as a Service (IaaS) Architectures

5 Cloud Deployment Models

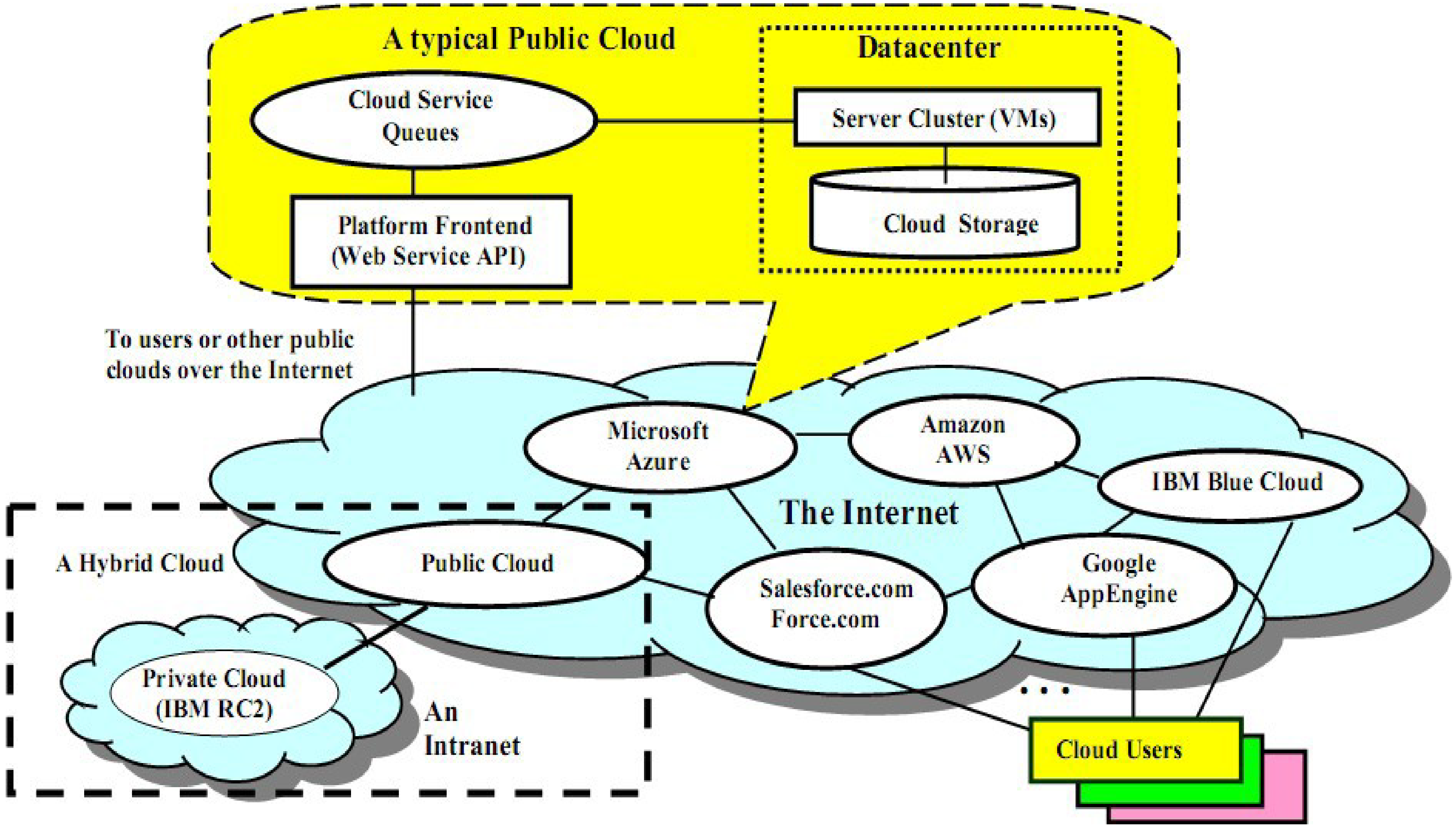
- ❑ Private cloud
  - ○ enterprise owned or leased. Cloud infrastructure for single org only, may be managed by the org or a 3rd party, on- or off-premise
- ❑ Community cloud
  - ○ shared infrastructure for specific community. Shared by several orgs that have shared concerns, managed by org or 3rd party
- ❑ Public cloud
  - ○ Sold to the public, mega-scale infrastructure. Cloud infrastructure is available to the general public, owned by org selling cloud services
- ❑ Hybrid cloud
  - ○ composition of two or more cloud deployment models

Cloud Orchestration

- Managing and controlling all your cloud environments



Orchestrated Cloud Environments


A typical Public Cloud
Datacenter
Cloud Service Queues
Server Cluster (VMs)
Cloud Storage
Platform Frontend (Web Service API)
To users or other public clouds over the Internet
Microsoft Azure
Amazon AWS
IBM Blue Cloud
The Internet
A Hybrid Cloud
Public Cloud
Salesforce.com Force.com
Google AppEngine
Private Cloud (IBM RC2)
An Intranet
. . .
Cloud Users