

CLOUD COMPUTING **(Undergraduate Course)**

Chapter 1 - Overview

Presenter: **Dr. Nguyen Dinh Long**

Email: dinhlonghcmut@gmail.com

Oct. 2022

Our story ...

Airbnb, Inc. is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. Based in San Francisco, California, the platform is accessible via website and mobile app.



CEO: Brian Chesky (Aug 2008–)

Headquarters: San Francisco, California, United States

Revenue: 5.99 billion USD (2021)

Number of employees: 6,132 (Dec 2021)

No build any building, home, office ...

No buy/own any build, home, ...



Our story ...

Meta Platforms, Inc., doing business as Meta and formerly named Facebook, Inc., and The Facebook, Inc., is an American multinational technology conglomerate based in Menlo Park, California. The company owns Facebook, Instagram, and WhatsApp, among other products and services.

No directly contact with anybody
No buy any devices/hardware for development



facebook

CEO: Mark Zuckerberg (Feb 4, 2004—)

Headquarters: Menlo Park, California, United States

Revenue: 85.96 billion USD (2020)



Our story ...

Uber Technologies, Inc. is an American mobility as a service provider, allowing users to book a car and driver to transport them in a way similar to a taxi. It is based in San Francisco with operations in approximately 72 countries and 10,500 cities in 2021.

No buy any car for serving customer
No hire/use any driver for working



Uber

Founded: March 2009, San Francisco, California, US
CEO: Dara Khosrowshahi (Aug 30, 2017–)
Number of employees: 32,500 (2022)
Revenue: 17.46 billion USD (2021)

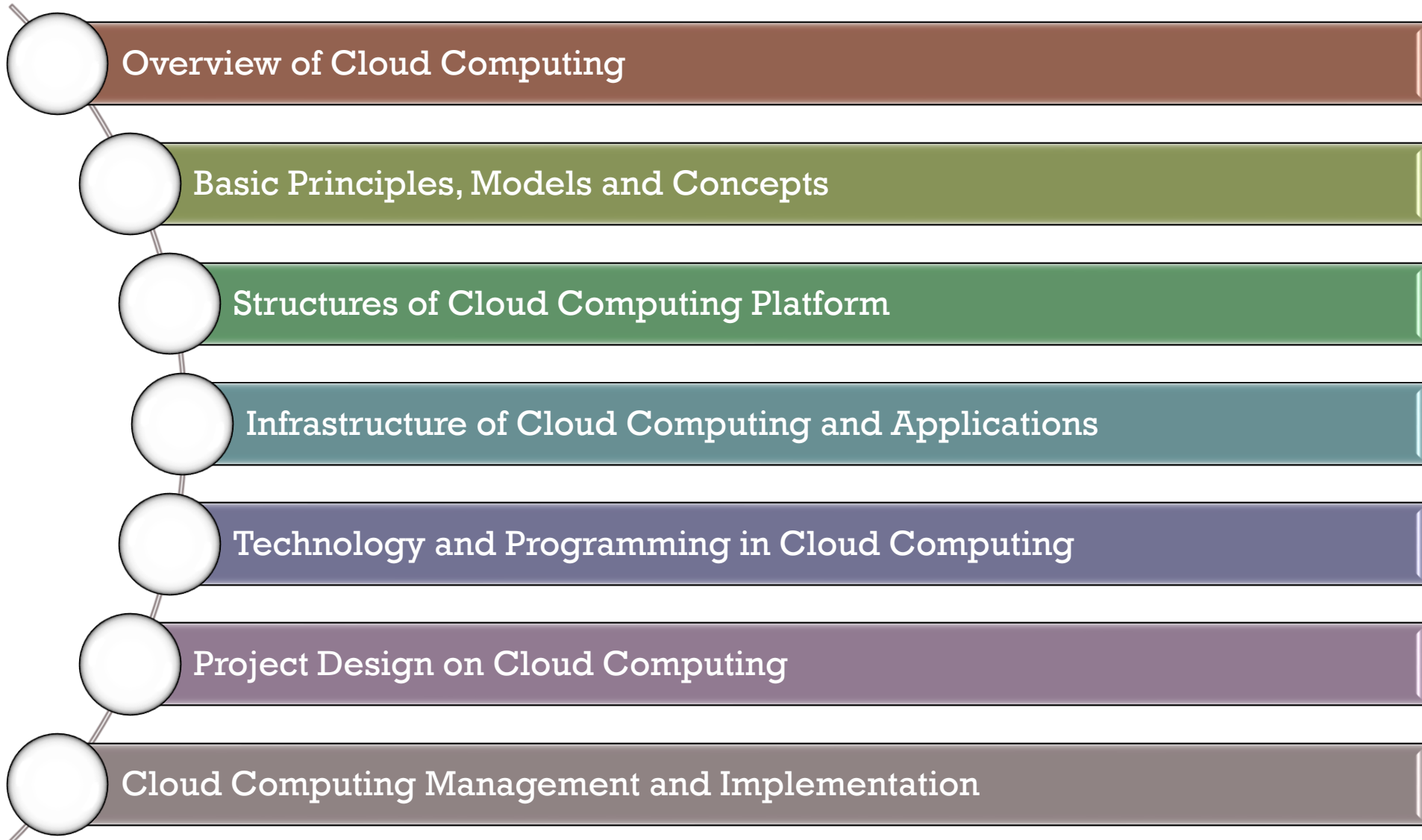


Grab in Vietnam

- ❑ How many people in Southern Vietnam???
- ❑ 1/10 là user of Grab
- ❑ $\frac{1}{4}$ user using Grab services/day \rightarrow 1 service take 1 route for Grab
- ❑ 1 route assuming 20K vnd/route, Grab take 15% cost in 1 route
- ❑ Determine the revenue of Grab per day, month, year ...



Outline



References

Main:

- Thomas Erl, Zaigham Mahmood, and Ricardo Puttini. 2013. *Cloud Computing Concepts, Technology & Architecture*. Prentice Hall.
- Michael J. Kavis. 2014. *Architecting the Cloud: Design Decisions for Cloud Computing Service Models*. Wiley
- Arshdeep Bahga, and Vijay Madisetti. 2013. *Cloud Computing: A Hands-On Approach*. CreateSpace Independent Publishing Platform

More:

- Rajkuma Buyya, Jame Broberg and Andrzej Goscinski. 2011. *Cloud Computing –Principles and paradigms*, Wiley
- Nick Antonopoulos, and Lee Gillam. 2010. *Cloud Computing - Principles, Systems and Applications*, Springer-Verlag London Limited.
- Slides here are modified from several sources in Universities and Internet.

Content of Chapter 1

1. The road of Cloud computing
2. Understanding Cloud Computing
 - Concepts, principles, ...
3. Fundamental, technology in Cloud computing
4. Applications of cloud computing
5. Challenges, Issues ...

The road of Cloud computing

Our Data Now...



Documents and Media



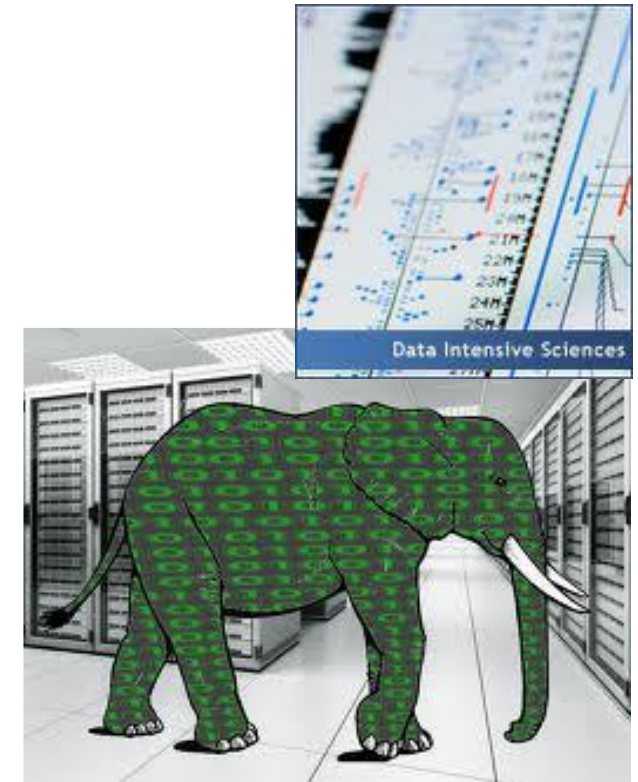
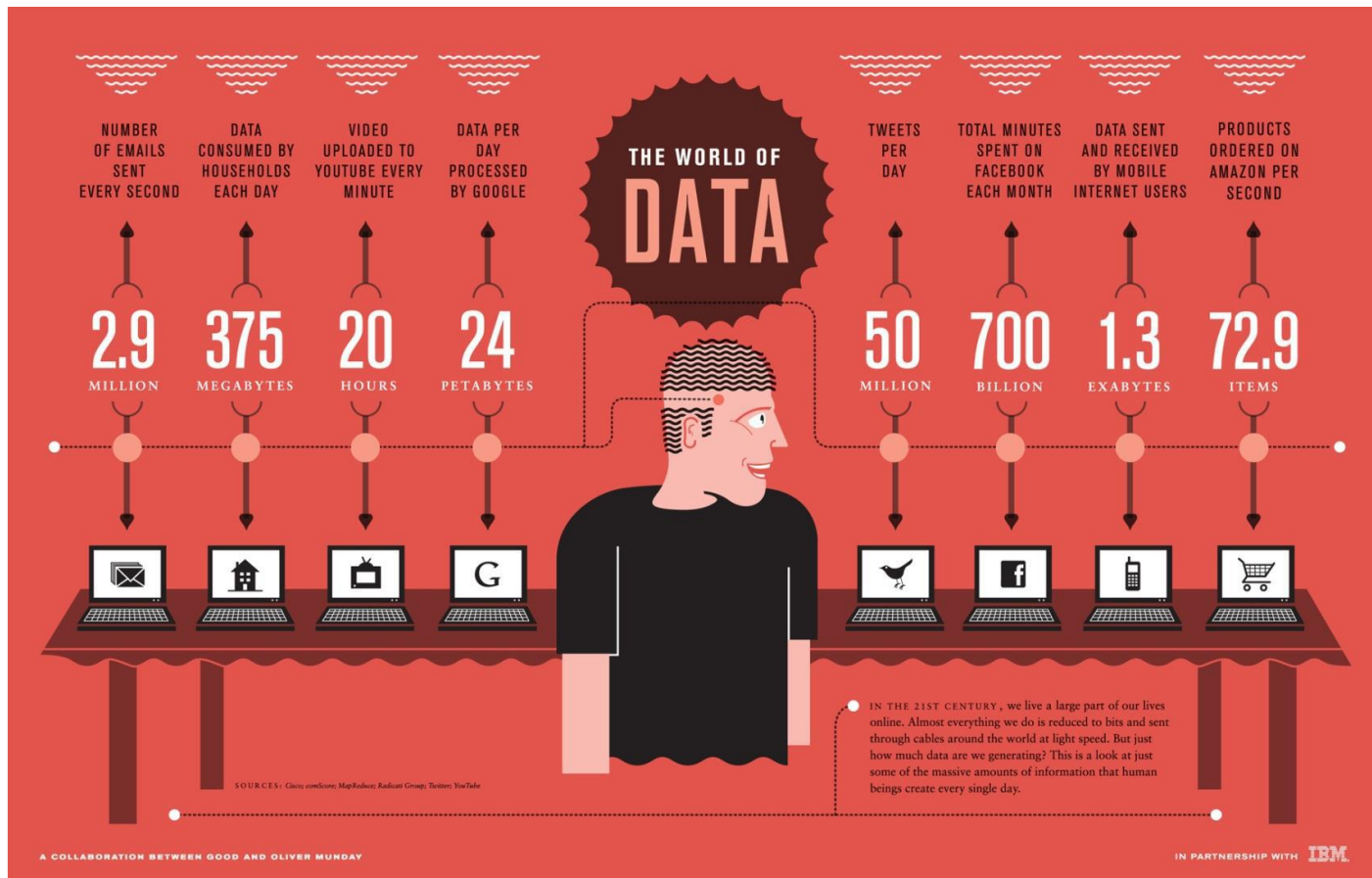
Personal Data



Emails, Calendars, Contacts,
Location Information, etc...

The road of Cloud computing

- ❑ We Live in a World of Data ... and Big Data ...
- Big data is defined as large pools of data that can be captured, communicated, aggregated, stored, and analyzed.
- Data continues to grow: In mid-2010, the information universe carried 1.2 zettabytes and 2020 expect nearly 44 times more at 35 zettabytes coming our way.



The road of Cloud computing

- ❑ What Do We Do With Data?
- ❑ Using Diverse Interfaces & Devices



Desktops



Store



Share



Access



Process



Encrypt



.... and more!



Mobile Devices



Consumer Electronics



...and even appliances

The road of Cloud computing

❑ We Live in a World of Data...

We also want to access, share and process our data from all of our devices, **anytime, anywhere!**

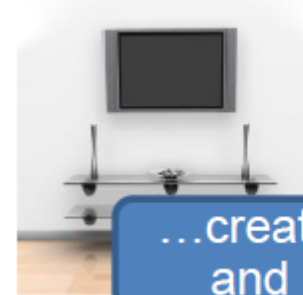
❑ What About the Future?



...work on documents?



...get your news & info?



...create, access, store and share media?



...communicate with friends and family?



...navigate?



...live in an intelligent home?

The road of Cloud computing

□ We Live in a World of Data...



The road of Cloud computing

❑ ... How Will We Manage Our Data?

Manage it
ourselves?

- Personal, but time consuming.

How would you get
access to your data
wherever you are?

- Would you keep it on your devices?
- or would you keep it online?

What if it's managed
by someone else?

- and you can get this “**service**” for free or with a subscription?

The road of Cloud computing

□ Think of it this Way ...



The road of Cloud computing

□ Think of it this Way ...



The road of Cloud computing

□ Think of it this Way ...



The road of Cloud computing

- ❑ So What is Cloud Computing?
- ❑ Where is it come from?

“Cloud Computing is the transformation of IT from a product to a service”



The road of Cloud computing

□ Why Cloud Computing?



Pay-as-You-Go economic model

- Reduce capital expenditure
- No upfront cost
- Reduced Time to Market



Simplified IT management

- All you need is access to the internet.
- It's the providers responsibility to manage the details.



Scale quickly and effortlessly

- Resources can be rented and released as required
- Software Controlled
- Instant scalability



Flexible options

- Configure software packages, instance types operating systems.
- Any software platform
- Access from any machine connected to the Internet



Resource Utilization is improved

- Reduce Idle resources by sharing and consolidation
- Better utilization of CPU / Storage and Bandwidth.



Carbon Footprint decreased

- Sharing of resources means less servers, less power and less emissions.



The road of Cloud computing



Understanding Cloud computing

❑ A Brief History

The idea of computing in a “cloud” traces back to the origins of utility computing, a concept that computer scientist John McCarthy publicly proposed in 1961:

“If computers of the kind I have advocated become the computers of the future, then computing may someday be organized as a public utility just as the telephone system is a public utility. ... The computer utility could become the basis of a new and important industry.”

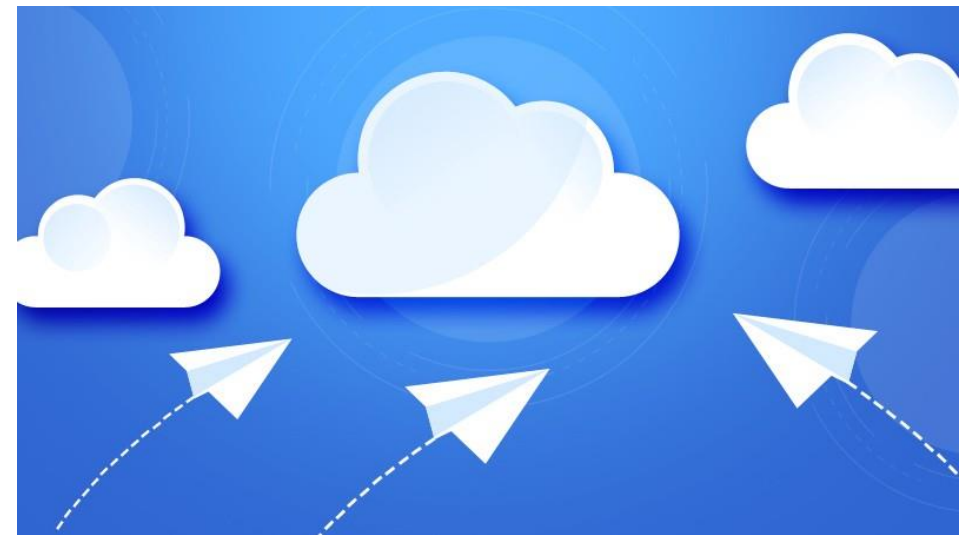
In the late 1990s, [Salesforce.com](https://www.salesforce.com) pioneered the notion of bringing remotely provisioned services into the enterprise. In 2002, [Amazon.com](https://aws.amazon.com) launched the Amazon Web Services (AWS) platform, a suite of enterprise oriented services that provide remotely provisioned storage, computing resources, and business functionality.

It wasn't until 2006 that the term “cloud computing” emerged in the commercial arena. It was during this time that Amazon launched its Elastic Compute Cloud (EC2) services that enabled organizations to “lease” computing capacity and processing power to run their enterprise applications. Google Apps also began providing browser-based enterprise applications in the same year, and three years later, the Google App Engine became another historic milestone.

Understanding Cloud computing

❑ Cloud

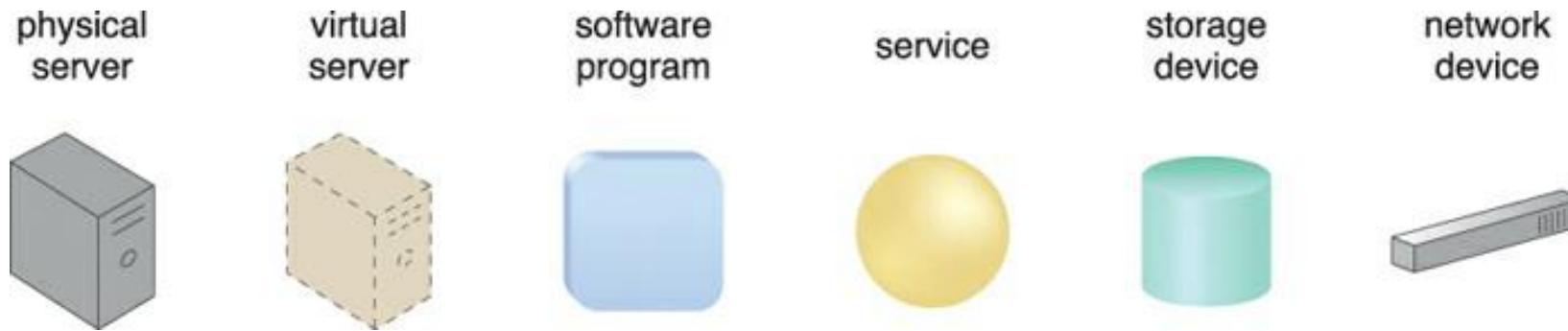
- A cloud refers to a distinct IT environment that is designed for the purpose of remotely provisioning scalable and measured IT resources. The term originated as a metaphor for the Internet which is, in essence, a network of networks providing remote access to a set of decentralized IT resources. Prior to cloud computing becoming its own formalized IT industry segment, the symbol of a cloud was commonly used to represent the Internet in a variety of specifications and mainstream documentation of Web-based architectures.
- The term “cloud” and the cloud symbol from the Internet. As a specific environment used to remotely provision IT resources, a cloud has a finite boundary.
- IT resources provided by cloud environments, on the other hand, are dedicated to supplying back-end processing capabilities and user-based access to these capabilities.
- Protocols refer to standards and methods that allow computers to communicate with each other in a pre-defined and structured manner.
A cloud can be based on the use of any protocols that allow for the remote access to its IT resources.



Understanding Cloud computing

IT Resource

- An IT resource is a physical or virtual IT-related artifact that can be either software-based, such as a virtual server or a custom software program, or hardware-based, such as a physical server or a network device



- Technology architectures and various interaction scenarios involving IT resources
 - The IT resources shown within the boundary of a given cloud symbol usually do not represent all of the available IT resources hosted by that cloud. Subsets of IT resources are generally highlighted to demonstrate a particular topic.
 - Focusing on the relevant aspects of a topic requires many of these diagrams to intentionally provide abstracted views of the underlying technology architectures. This means that only a portion of the actual technical details are shown.

Understanding Cloud computing

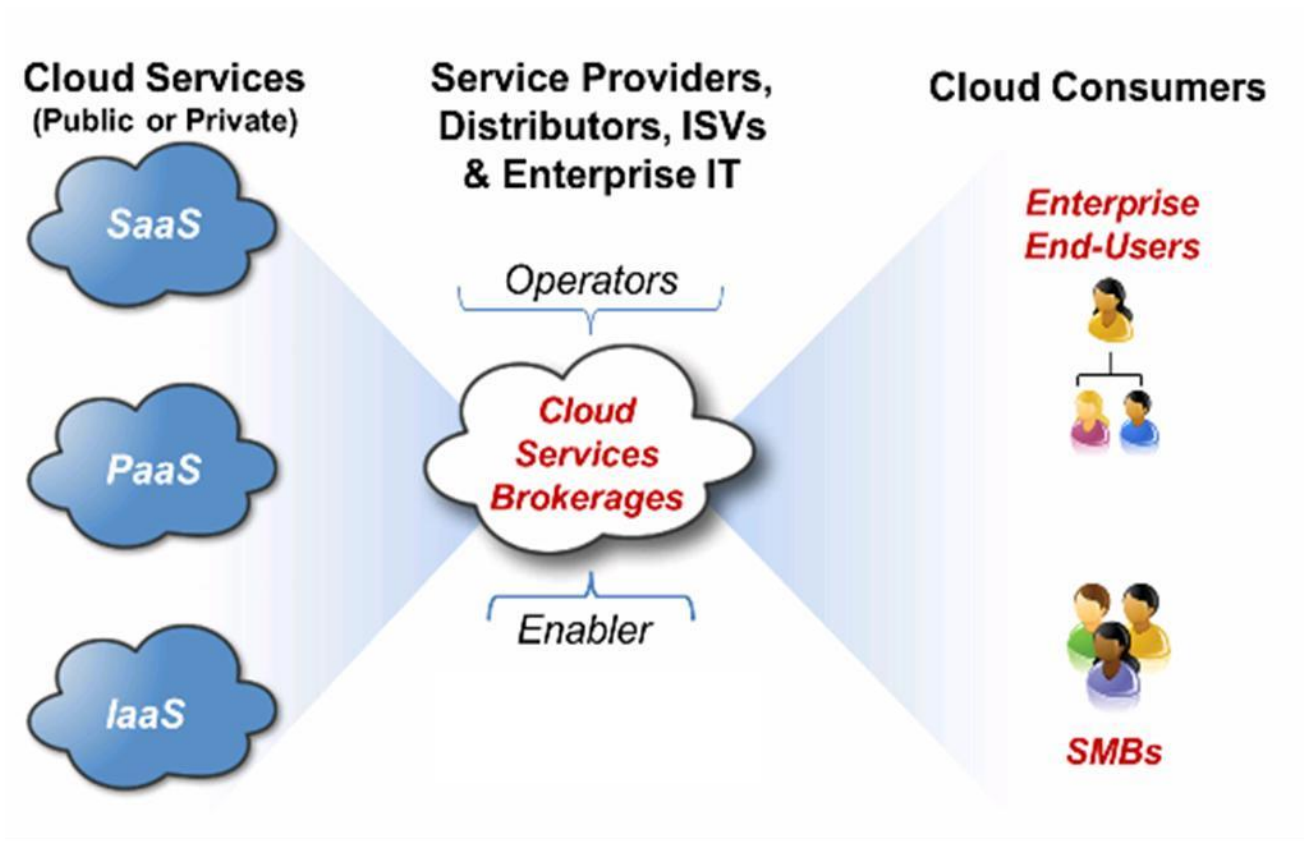
❑ On-Premise

- A cloud represents an option for the deployment of IT resources. An IT resource that is hosted in a conventional IT enterprise within an organizational boundary
→ is considered to be located on the premises of the IT enterprise, or **on-premise** for short.
- Key points:
 - An on-premise IT resource can access and interact with a cloud-based IT resource.
 - An on-premise IT resource can be moved to a cloud, thereby changing it to a cloud-based IT resource.
 - Redundant deployments of an IT resource can exist in both on-premise and cloud-based environments.
- In private clouds, on-premise and cloud-based IT resources can be confused

Understanding Cloud computing

❑ Cloud Consumers and Cloud Providers

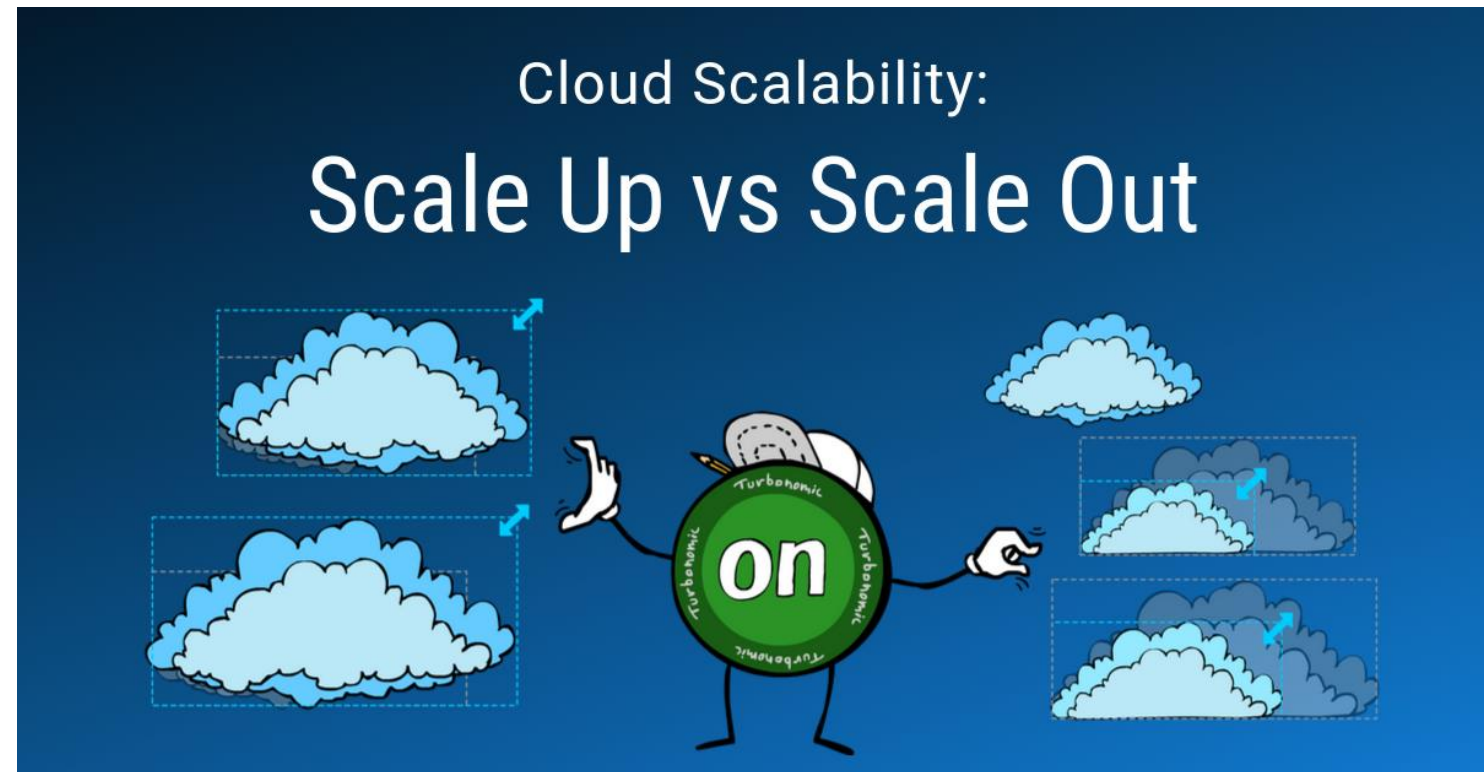
- The party that provides cloud-based IT resources is the cloud provider.
- The party that uses cloud-based IT resources is the cloud consumer.
- These terms represent roles usually assumed by organizations in relation to clouds and corresponding cloud provisioning contracts.



Understanding Cloud computing

Scaling

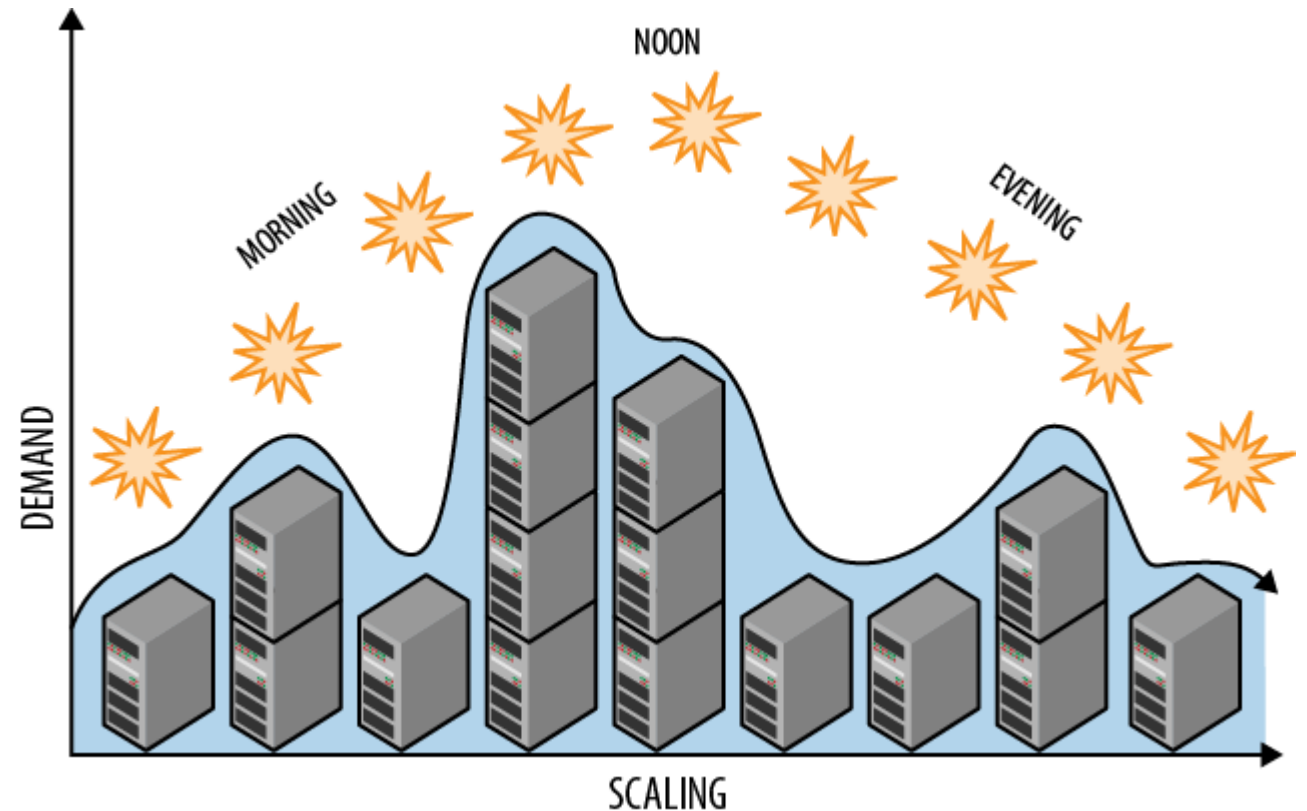
- Scaling, from an IT resource perspective, represents the ability of the IT resource to handle increased or decreased usage demands.
- The following are types of scaling:
 - Horizontal Scaling – scaling out and scaling in
 - Vertical Scaling – scaling up and scaling down



Understanding Cloud computing

Scaling

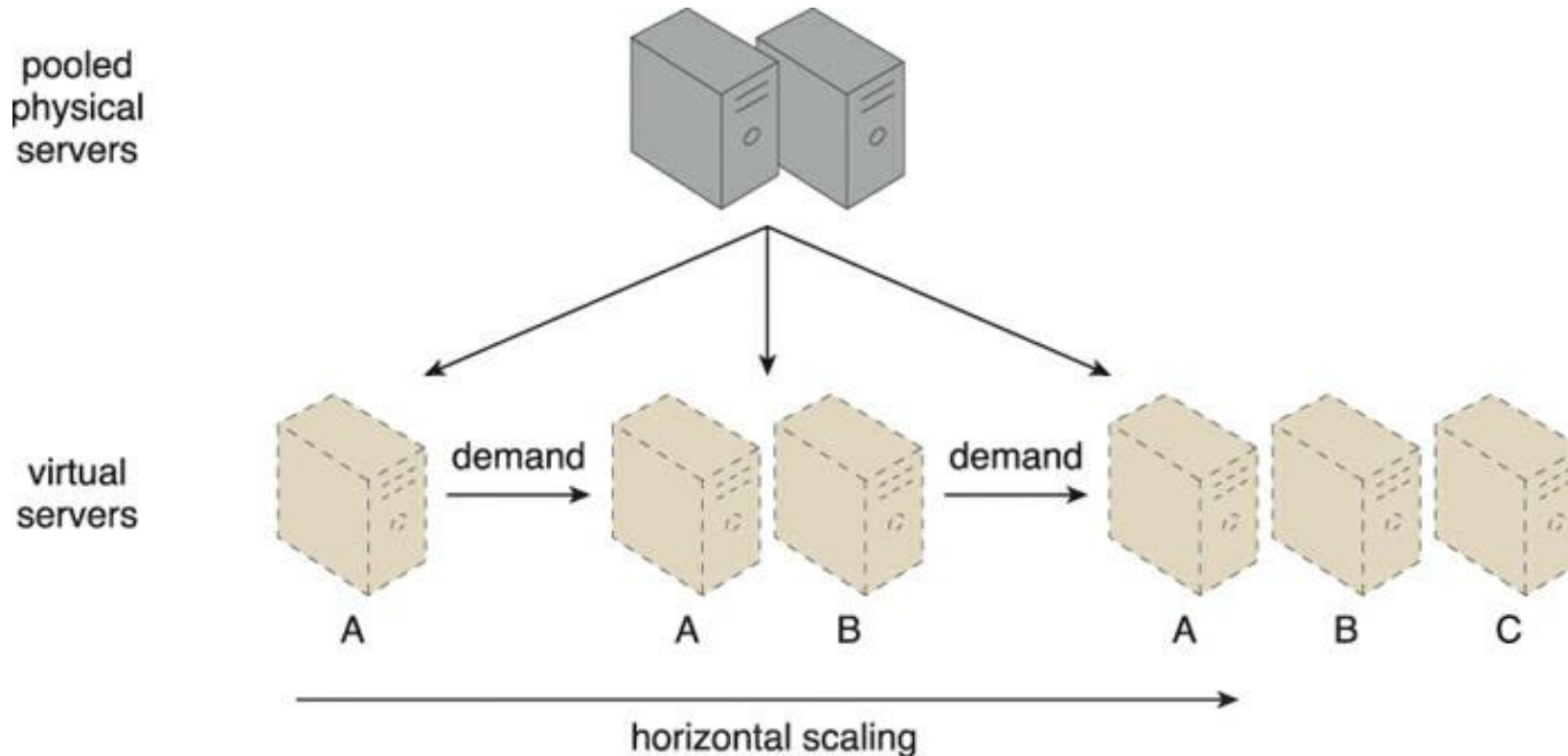
- Scaling, from an IT resource perspective, represents the ability of the IT resource to handle increased or decreased usage demands.
- The following are types of scaling:
 - Horizontal Scaling – scaling out and scaling in
 - Vertical Scaling – scaling up and scaling down



Understanding Cloud computing

Scaling - Horizontal Scaling:

- The allocating or releasing of IT resources that are of the same type is referred to as horizontal scaling.
- The horizontal allocation of resources is referred to as scaling out and the horizontal releasing of resources is referred to as scaling in.
- Horizontal scaling is a common form of scaling within cloud environments.

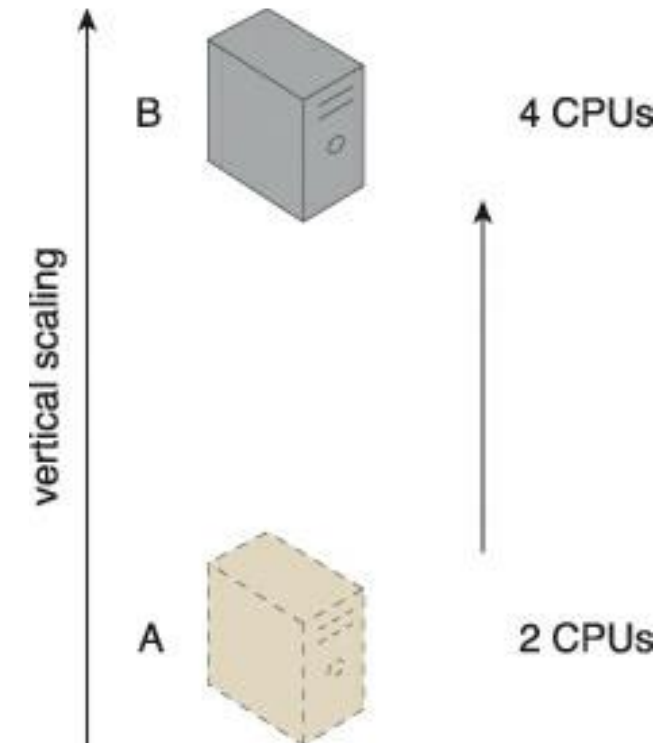


Understanding Cloud computing

❑ Scaling - Vertical Scaling:

- When an existing IT resource is replaced by another with higher or lower capacity, vertical scaling is considered to have occurred.
- Specifically, the replacing of an IT resource with another that has a higher capacity is referred to as scaling up and the replacing an IT resource with another that has a lower capacity is considered scaling down.
- Vertical scaling is less common in cloud environments due to the downtime required while the replacement is taking place..

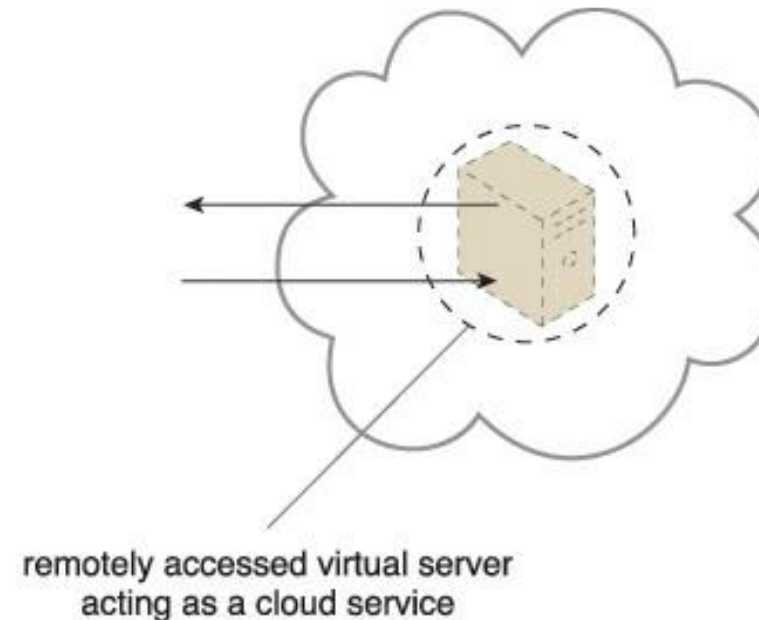
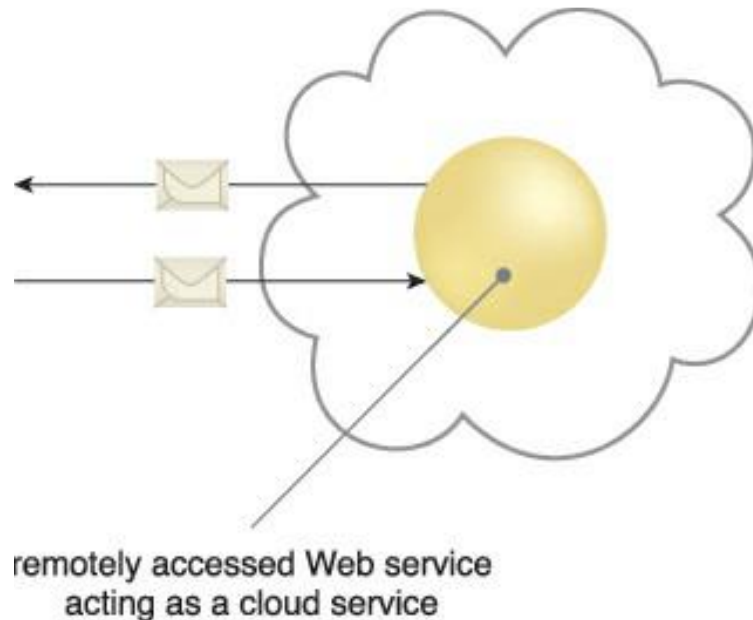
Horizontal Scaling	Vertical Scaling
less expensive (through commodity hardware components)	more expensive (specialized servers)
IT resources instantly available	IT resources normally instantly available
resource replication and automated scaling	additional setup is normally needed
additional IT resources needed	no additional IT resources needed
not limited by hardware capacity	limited by maximum hardware capacity



Understanding Cloud computing

❑ Cloud Service

- A database or a physical server deployed within a cloud may only be accessible by other IT resources that are within the same cloud. A software program with a published API may be deployed specifically to enable access by remote clients.
- A **cloud service** is any IT resource that is made remotely accessible via a cloud. A cloud service can exist as a simple Web-based software program with a technical interface invoked via the use of a messaging protocol, or as a remote access point for administrative tools or larger environments and other IT resources.



Understanding Cloud computing

❑ Cloud Service Consumer

- The cloud service consumer is a temporary runtime role assumed by a software program when it accesses a cloud service.
- Common types of cloud service consumers can include software programs and services capable of remotely accessing cloud services with published service contracts, as well as workstations, laptops and mobile devices running software capable of remotely accessing other IT resources positioned as cloud services.



Understanding Cloud computing

❑ Goals and Benefits

▪ Reduced Investments and Proportional Costs

a company with sizable batch-centric tasks can complete them as quickly as their application software can scale. Using 100 servers for one hour costs the same as using one server for 100 hours. This “elasticity” of IT resources, achieved without requiring steep initial investments to create a large-scale computing infrastructure, can be extremely compelling

▪ Increased Scalability

Cloud consumers can scale their cloud-based IT resources to accommodate processing fluctuations and peaks automatically or manually. Similarly, cloud-based IT resources can be released (automatically or manually) as processing demands decrease.

▪ Increased Availability and Reliability

An IT resource with increased availability is accessible for longer periods of time (for example, 22 hours out of a 24 hour day). Cloud providers generally offer “resilient” IT resources for which they are able to guarantee high levels of availability

Challenges in Cloud computing

❑ Technical Challenges

- Programming is tricky but improving
- Tools are continuously evolving
- Moving large data is still expensive
- Security
- Quality of Service
- Green computing
- Internet Dependence

❑ Non-Technical Challenges

- Vendor Lock-In
- Non-standardized
- Security Risks
- Privacy
- Legal
- Service Level Agreements

Technology in Cloud computing

❑ Requirements to Transform IT to a Service:

- Connectivity
 - For moving data around
- Interactivity
 - Seamless interfaces
- Reliability
 - Failure will affect many people, not just one
- Performance
 - It should not be slower or less efficient than what people already have
- Pay-as-you-Go
 - Should not pay an upfront fee for the service



Technology in Cloud computing

❑ Requirements to Transform IT to a Service:

- Ease of Programmability
 - Ease of development of complex services to users
- Manage Large Amounts of Data
 - Big Data
- Efficiency
 - Cost
 - Power
- Scalability & Elasticity
 - Flexible and rapid response to changing user needs

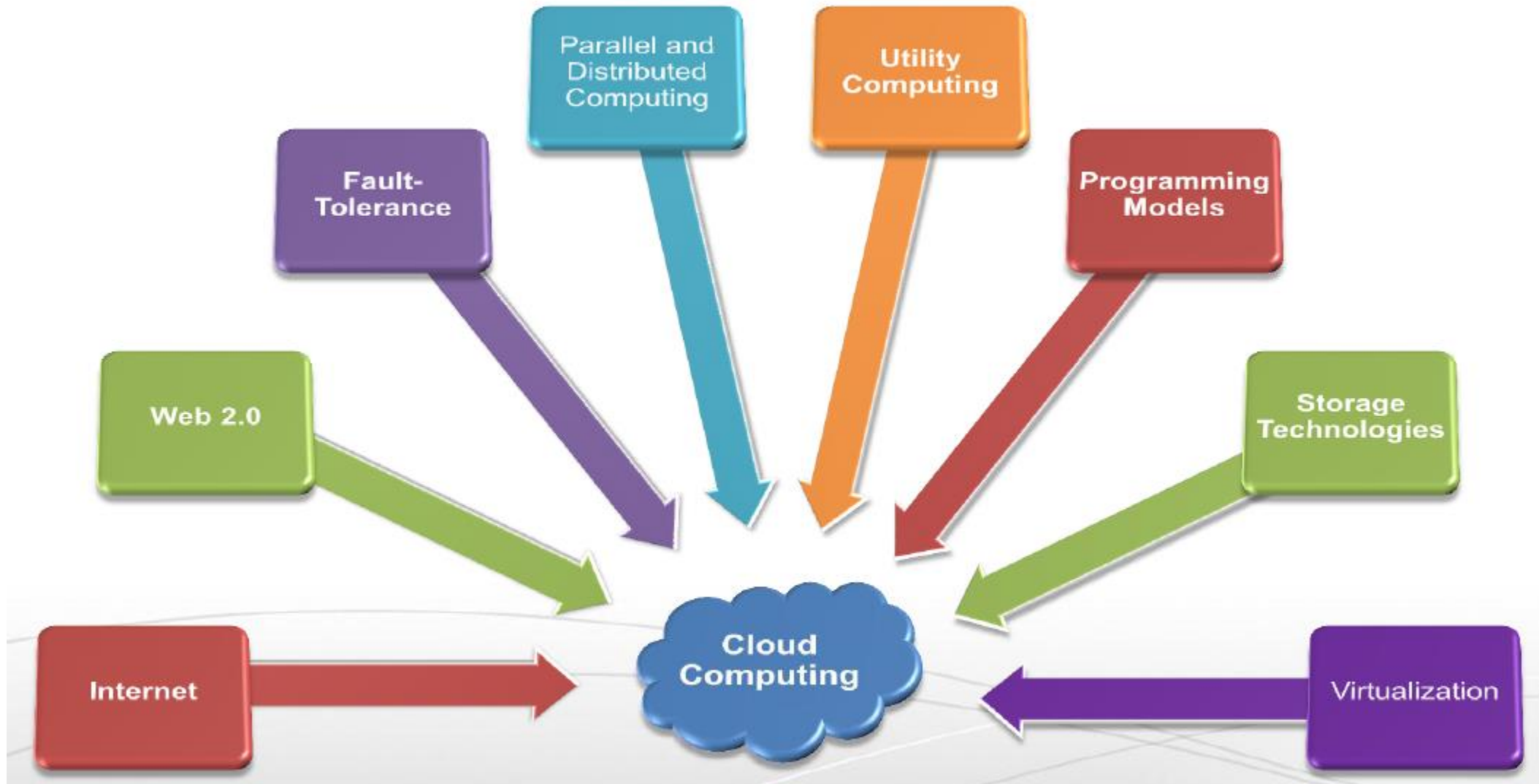
Ease of Programmability
+ Ease of development of complex services to users
Programming Model

Manage Large Amounts of Data
+ Big Data
Storage Technologies

Efficiency
+ Cost
+ Power
Scalability & Elasticity
+ Flexible and rapid response to changing user needs
Virtualization Technologies

Technology in Cloud computing

□ Combine the Enabling Technologies...



Applications of Cloud computing

Stream

Classwork

People


Grades

Cloud_Computing

DH20DM

 Customize

Class code

mlocqxr 



Announce something to your class



Stream

Classwork

People

Grades

Upcoming

No work due soon

[View all](#)



This is where

Use the stream to s
student questions

All topics

Overview of Cloud c...

Introduction and Ou...

 Create



Google Calendar



Class Drive folder

Overview of Cloud computing



Students will see this topic once work is added to it

Introduction and Outline



Students will see this topic once work is added to it

Applications of Cloud computing

High Growth Applications



Startup
Businesses

2001



friendster®

Could not keep up with the growth of
their number of users.

2006

facebook

Growing exponentially

vs.



What do you do when your startup
gains traction?



Can you grow quick
enough?

Applications of Cloud computing



Website went down on 9/11/2001 due to traffic



February 14th – Busiest Day of the Year



US Holiday Season



Website crashed within 10 minutes of the free trouser promotion during Superbowl 2010

Even if you design your website infrastructure to handle peak loads, wont it be idle during other times?

Aperiodic Bursting Applications



Seasonal
Businesses

Applications of Cloud computing

Researchers running large-scale scientific simulation using 1000s of computers.

Modern Drug Discovery



- Data-intensive simulation and tests to discover new compounds
- Large compute power required for simulation jobs
- Time to market is crucial

Why not rent computer time to run these simulations?

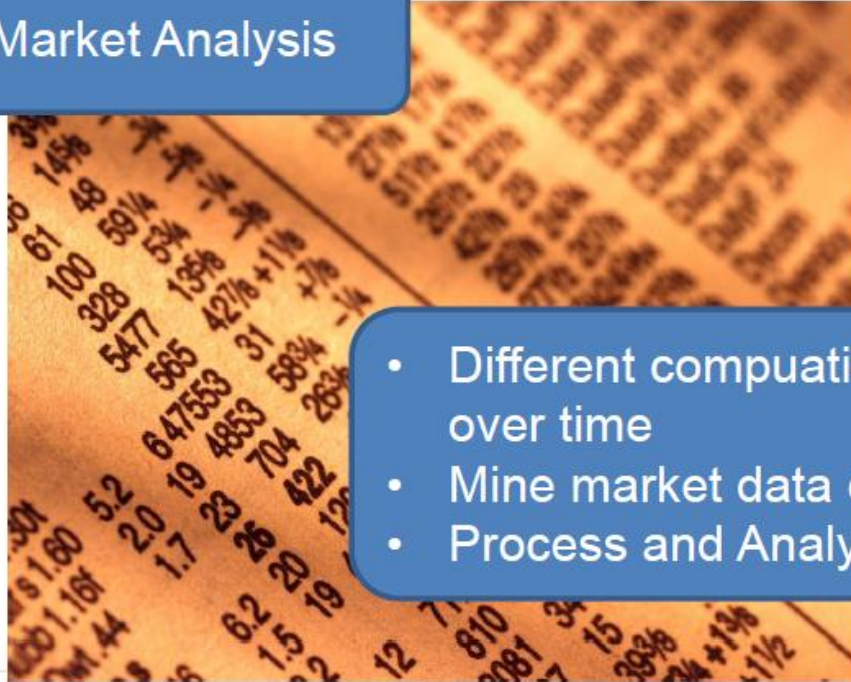
On-Off Applications



Research
Computing

Applications of Cloud computing

Sock Market Analysis



- Different computational requirements over time
- Mine market data during the day.
- Process and Analyze at night.

Periodic Applications



Dynamic and Flexible infrastructure can reduce costs and improve performance.