

CSCI 381/780

Cloud Computing

Introduction

Jun Li
Queens College



What is cloud?

- ▶ An emerging computing paradigm where **data and services** reside in **massively scalable data centers** and can be ubiquitously accessed from any connected devices **over the Internet**.



Example: Google's services

- ▶ Gmail, Google Calendar, Google Docs, Google Sheet, Google Slides, Google Drive, Youtube, Google Cloud



Gmail



Calendar



Sheets



Docs



Slides



YouTube



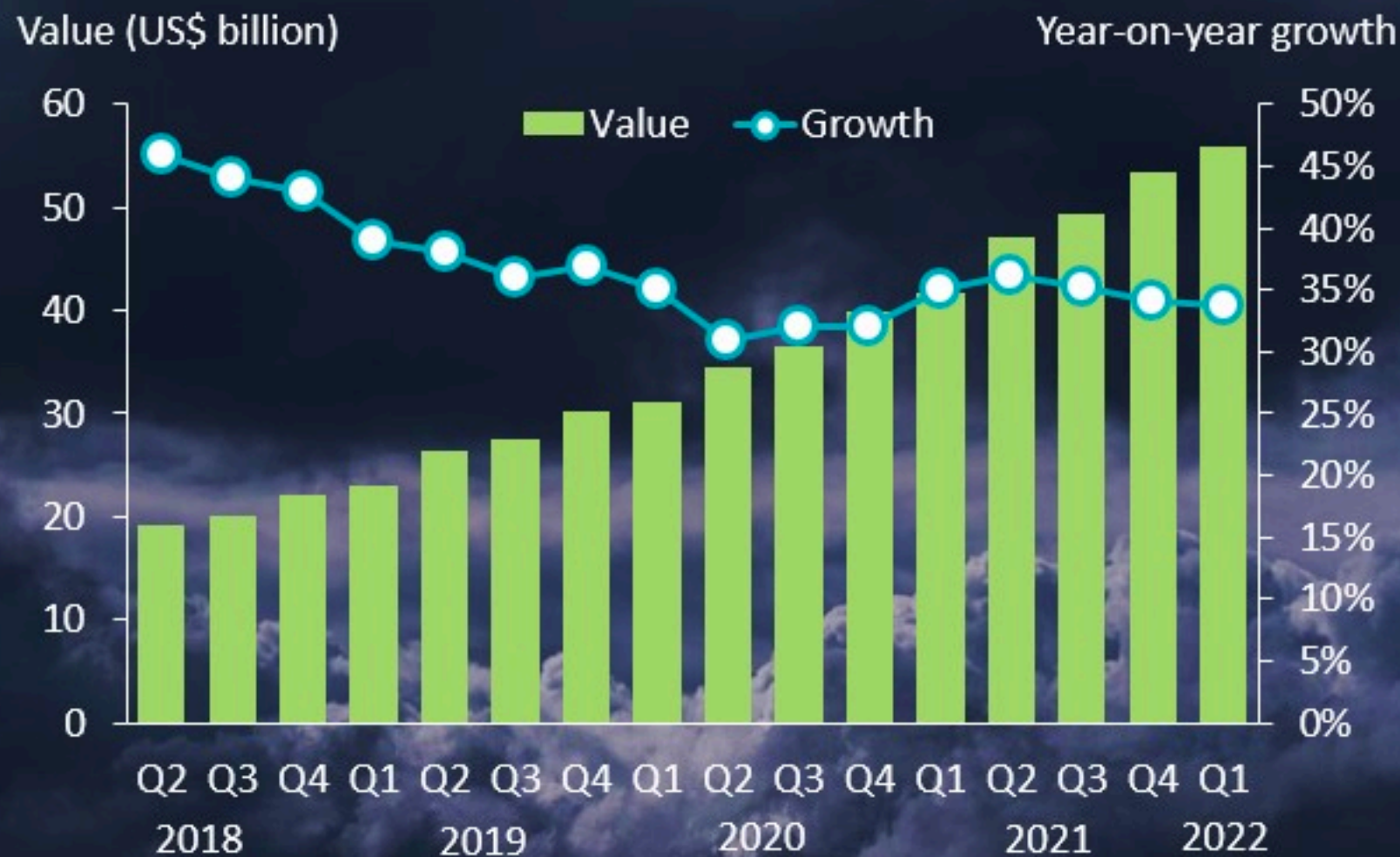
Google Cloud

Example: cloud storage

- ▶ Dropbox, Google Drive, iCloud storage, Microsoft OneDrive, Box, ...



Worldwide cloud infrastructure services spend,
Q2 2018 to Q1 2022



Worldwide cloud
infrastructure
spend grew 34%
year on year to
US\$55.9 billion in
Q1 2022

Source: Canalys estimates, April 2022

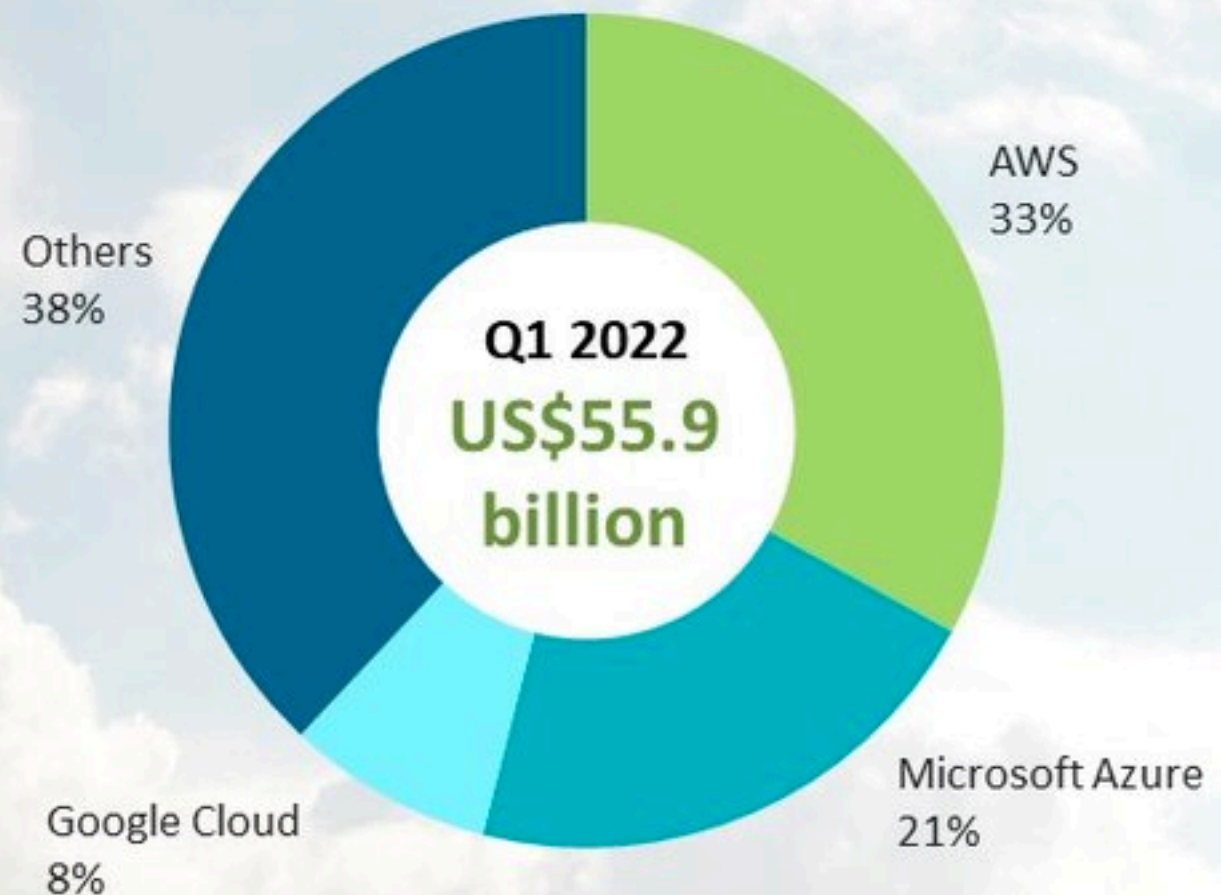
Major players of cloud platforms

- ▶ Amazon : Amazon Web Services (AWS)
- ▶ Google : Google Cloud Platform (GCP)
- ▶ Microsoft: Microsoft Azure
- ▶ Alibaba, salesforce, IBM,



Worldwide cloud infrastructure
services spend, Q1 2022

The top three
cloud service providers
accounted for 62% of
total cloud spend in
Q1 2022



Source: Canalys estimates, April 2022



More definitions

- ▶ Narrow/Brief: An updated version of utility computing: basically virtual servers (running VMWare, Xen, KVM, VirtualBox) available over the Internet.

Utility computing

- ▶ “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.”

John McCarthy, 1961



More definitions

- ▶ More Technical :
 - ▶ Utility Computing: A pool of virtualized computer resources that IT can access on demand (Google App Engine, Amazon EC2...)
 - ▶ Software as a Service (SaaS): Deliver a single application through the browser to thousands of customers (Gmail, Google Drive, DropBox, ...)
 - ▶ Also, PaaS, IaaS, CaaS, FaaS, AISSA, MLSSA....

Everything-as-a-service

- ▶ AaaS: Architecture as a service
- ▶ AlaaS: AI as a service
- ▶ BaaS: Blockchain as a service
- ▶ CaaS: Container as a service
- ▶ DaaS: Data as a service
- ▶ DBaaS: database as a service
- ▶ EaaS: Ethernet as a service
- ▶ FaaS: Function as a service
- ▶ GaaS: Globalization or Governance as a service
- ▶ HaaS: Hardware as a service
- ▶ IoTaaS: IoT as a service
- ▶ **IaaS: Infrastructure as a Service**
- ▶ IDaaS: Identity as a Service
- ▶ LaaS: Lending as a Service
- ▶ MLaaS: Machine learning as a Service
- ▶ NaaS: Networking as a Service (Software-Defined WAN)
- ▶ **SaaS: Software as a Service**
- ▶ **PaaS: Platform as a Service**
- ▶ SECaaS: Security as a Service

More definitions

- ▶ More Practical:
 - ▶ Cloud computing describes the use of software, storage, or processing services delivered over the Web from massive data centers

More definitions

- ▶ News / Blogs:
- ▶ Cloud computing, the notion of outsourcing hardware and software to Internet service providers,..... (Hint: "Don't do it by yourself !")
- ▶ A style of computing where massively scalable IT-enabled capabilities are provided "as a service" over the network.

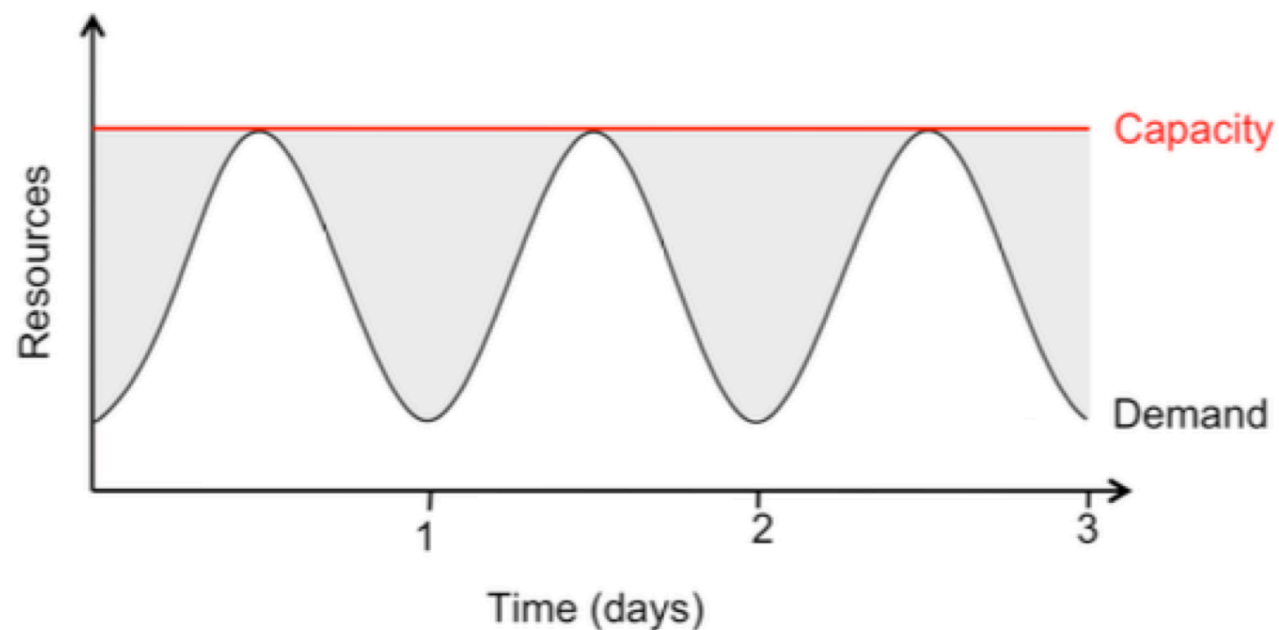
Essential characteristics

- ▶ **On-demand self-service.** Unilateral provisioning of computing resources, such as server time, storage or network bandwidth, without requiring human interaction with service providers.
- ▶ **Broad network access.** Capabilities are available over the network and accessed through heterogeneous thin or thick client platforms (e.g., mobile phones, tablets, laptops, and workstations).
- ▶ **Resource pooling.** The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model.
- ▶ **Rapid elasticity.** Quick scale up or scale down of resources through elastic provisioning or the release of capabilities in near real time.
- ▶ **Measured service.** Resource usage can be monitored, controlled, reported (thus being charged “pay-as-you-go”).

Why cloud?

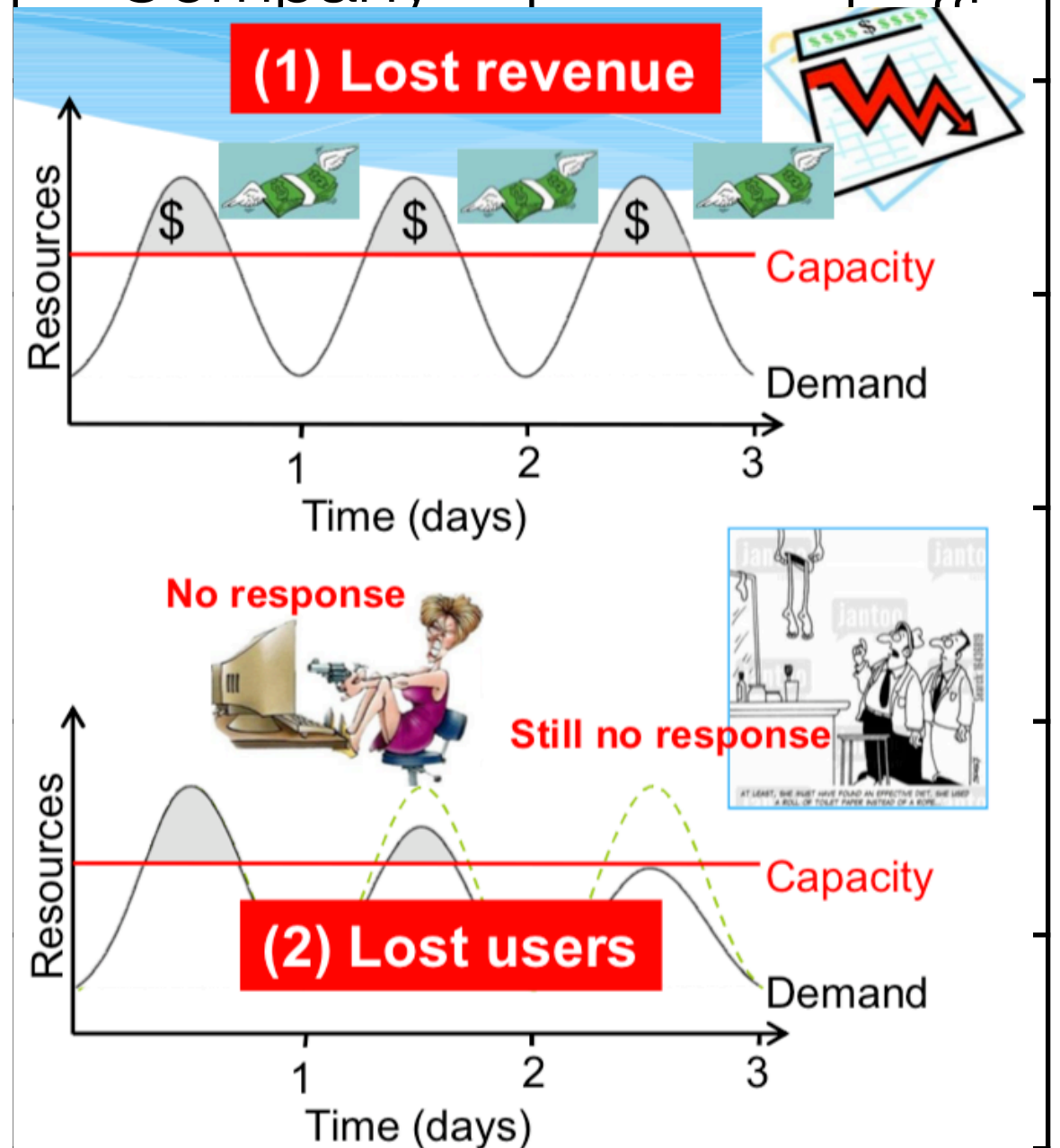
Resource planning at traditional data center

► Resource planning



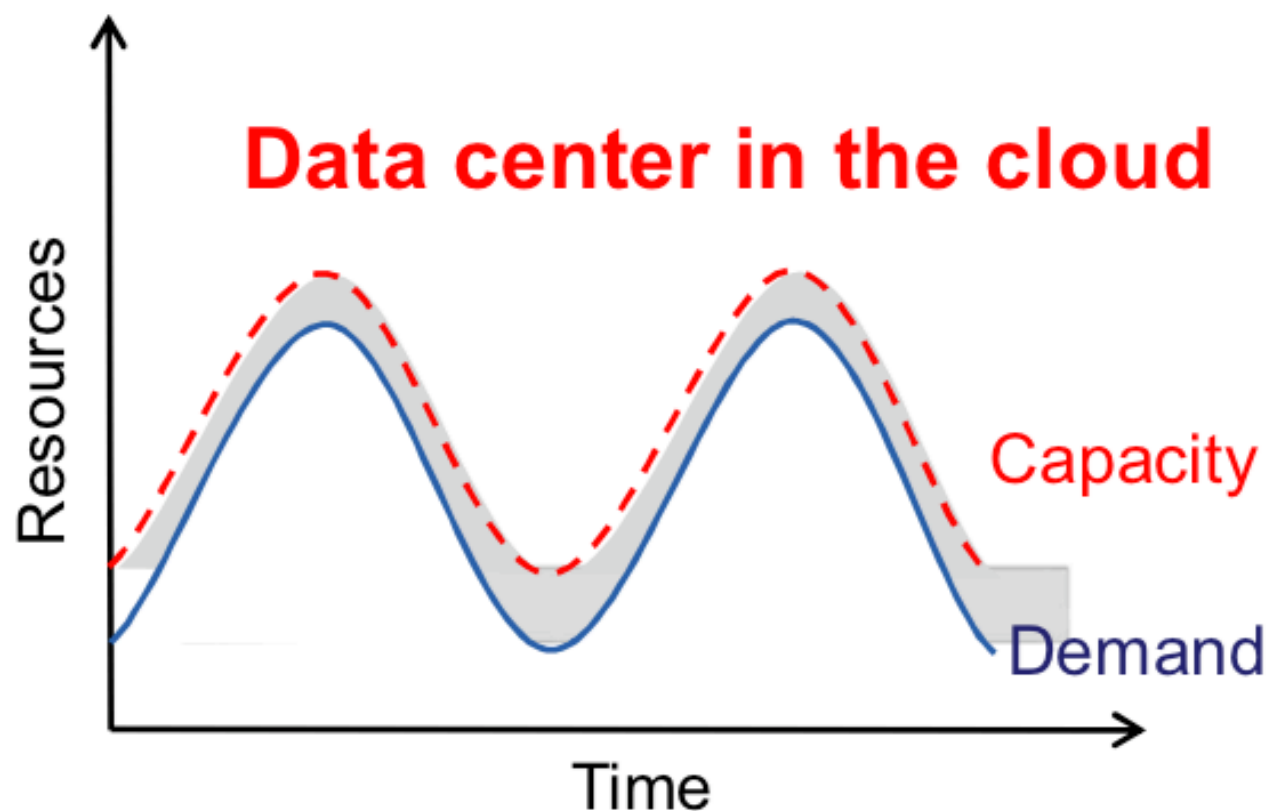
► Over-provisioning: low utilization

► Under-provisioning: Peak traffic / Company



Cloud economics

- ▶ Pay-as-you-go (usage-based) pricing:
 - ▶ Most services charge per minute, per byte, etc
 - ▶ No minimum or up-front fee
 - ▶ Helpful when apps have variable utilization



“Pay by use” (like electricity/water/gas)