Advanced Cloud Computing
Cloud Concepts and Foundations

Wei Wang
CSE@HKUST
Spring 2022

THE DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Outline

- Real-world examples of the cloud
- Definitions of cloud computing
- Key cloud concepts and characteristics
- Deployment scenarios and cloud benefits
- Introduction to AWS

Cloud: Massive Scale

- Facebook [GigaOM, 2012]
  - 30K in 2009 -> 60K in 2010 -> 180K in 2012
- Microsoft [DC knowledge]
  - > 1 million, 2013
- AWS EC2 [Randy Bias, 2009]
  - 40K, 8 cores per machine
- Google [DC knowledge]
  - > 900 K, 2013

Datacenter: outside

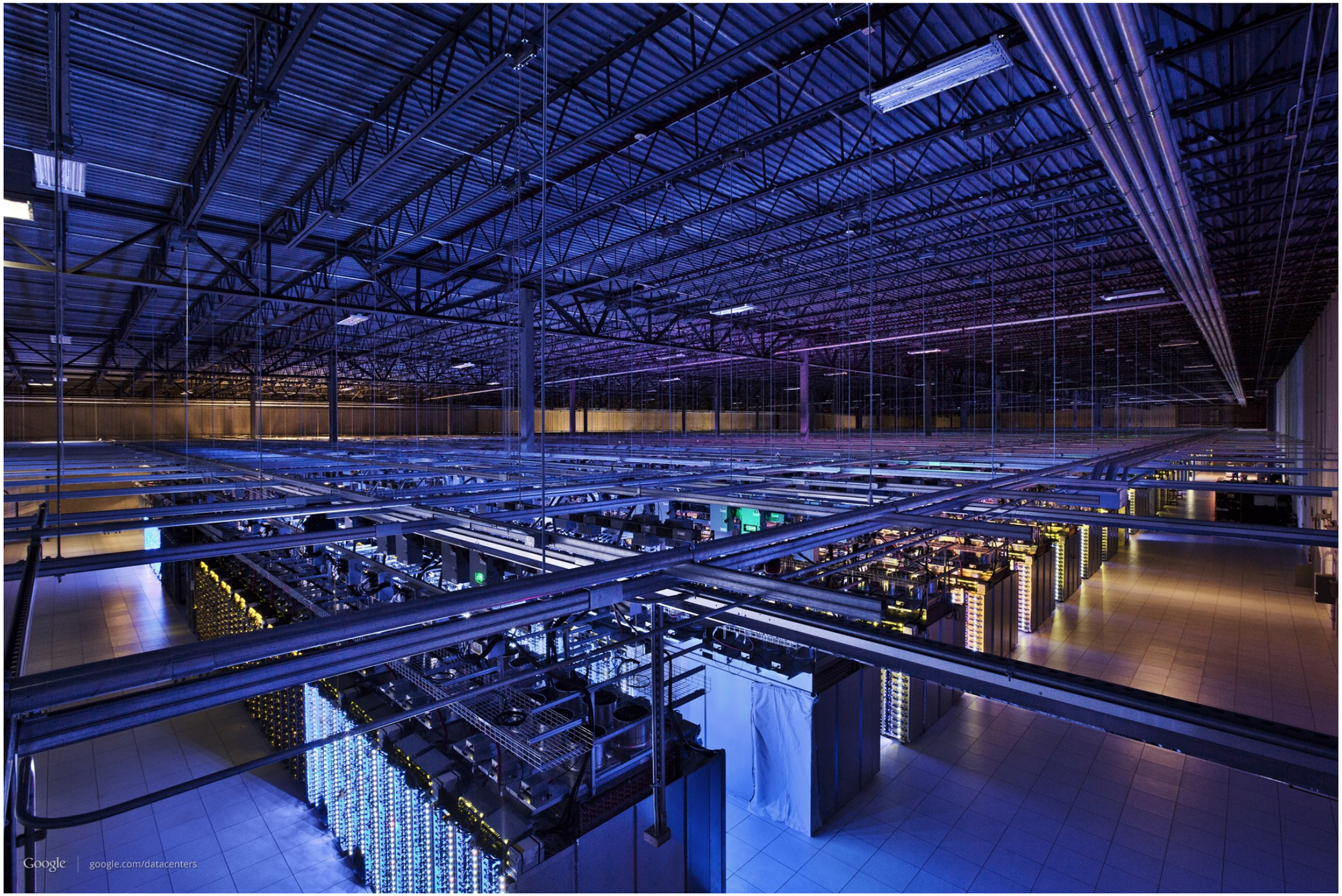
Copyright: Google

Datacenter: outside

Copyright: Google

A bird’s-eye view of DC

Datacenter: inside



Copyright: Google

Server racks

Photo credit: Google

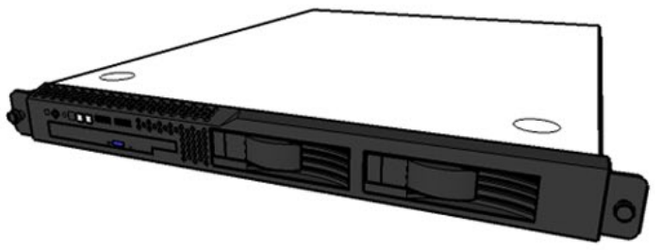
When the nights come…

Copyright: Google

Server: inside

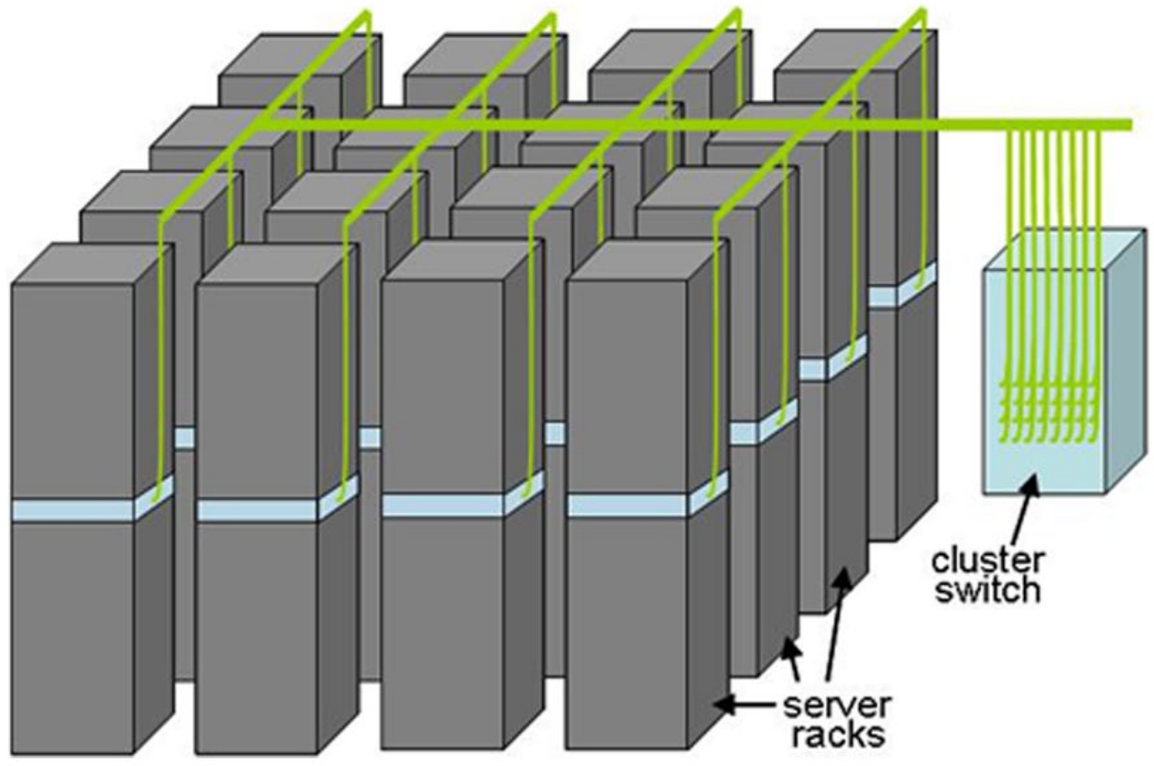
Server cage

A look into the datacenter



Commodity Server

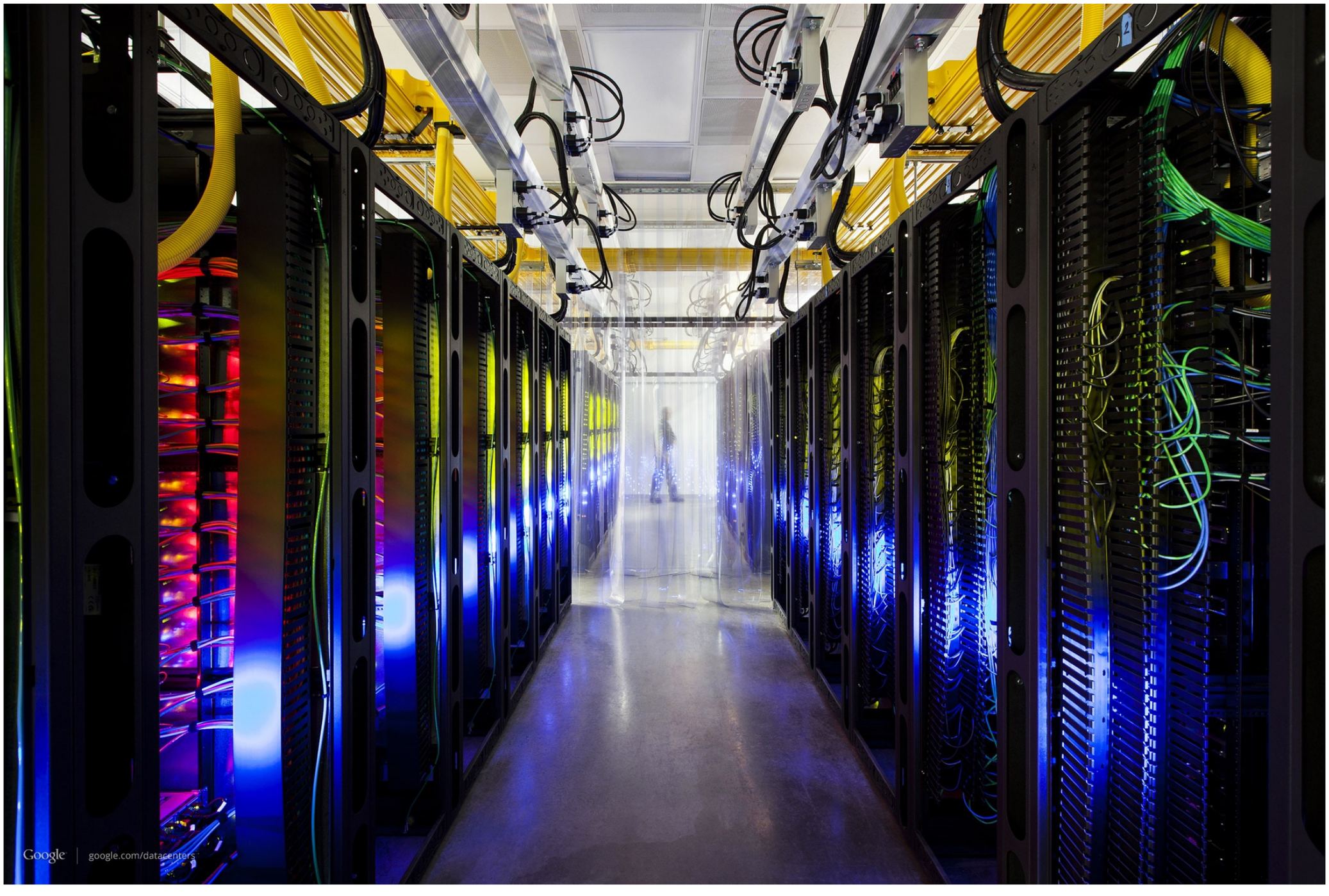
Rack



Cell

Source: L. Barroso et al., “The datacenter as a computer: An introduction to the design of warehouse-scale machines.”

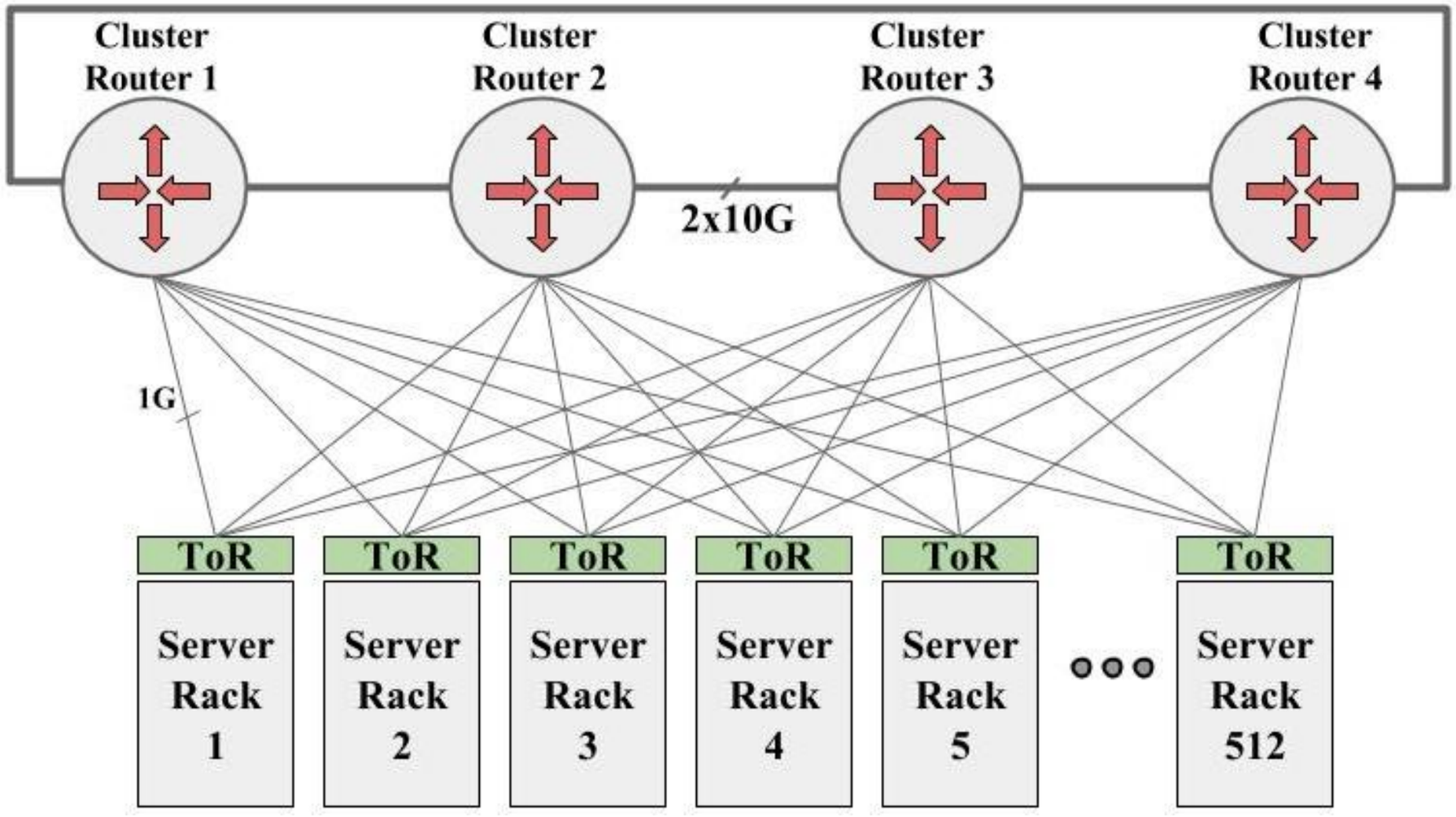
Network room



Copyright: Google

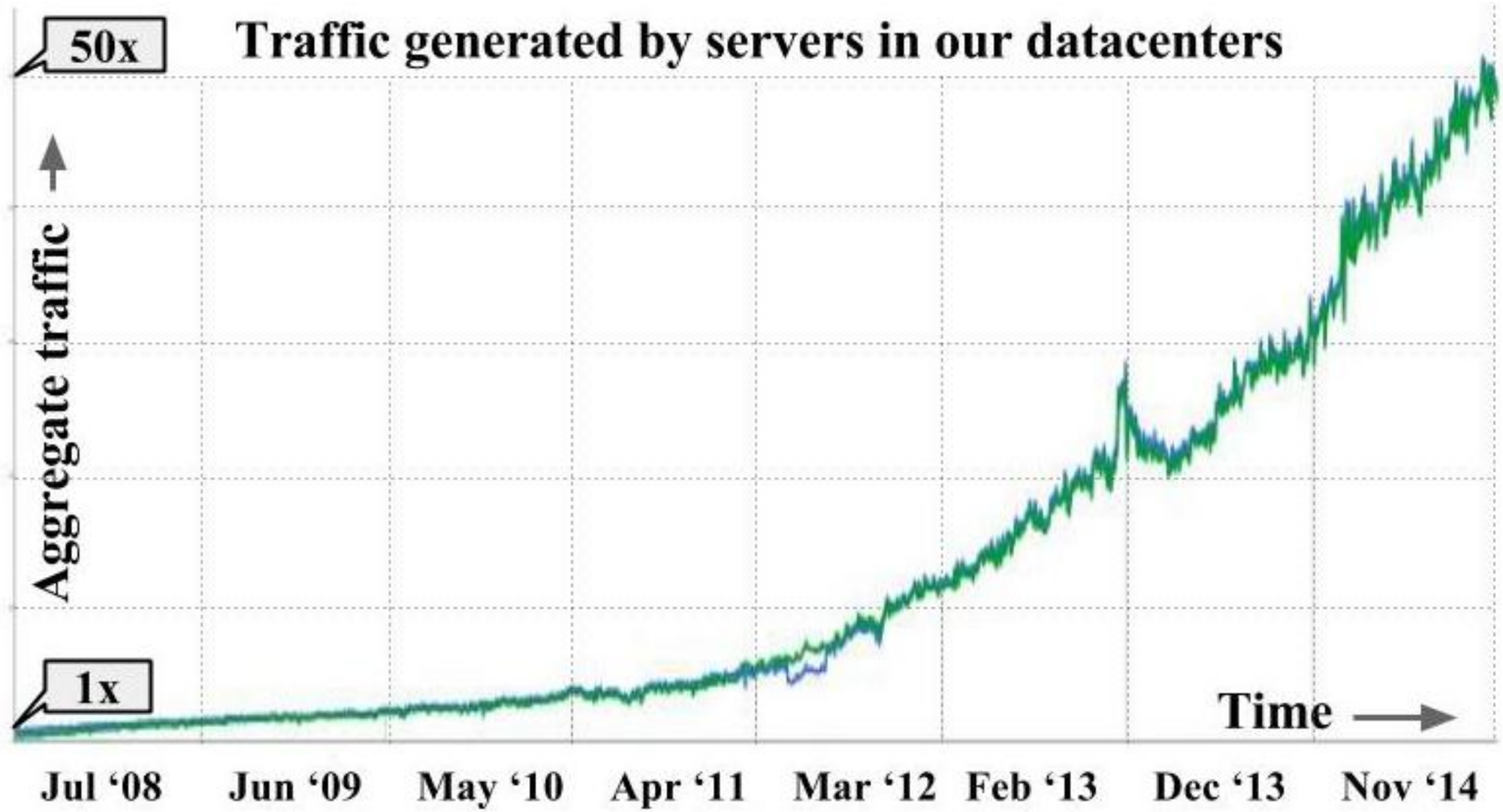
Network for a small-sized cluster

- Back to 2004 when Google has only 20k servers in a datacenter



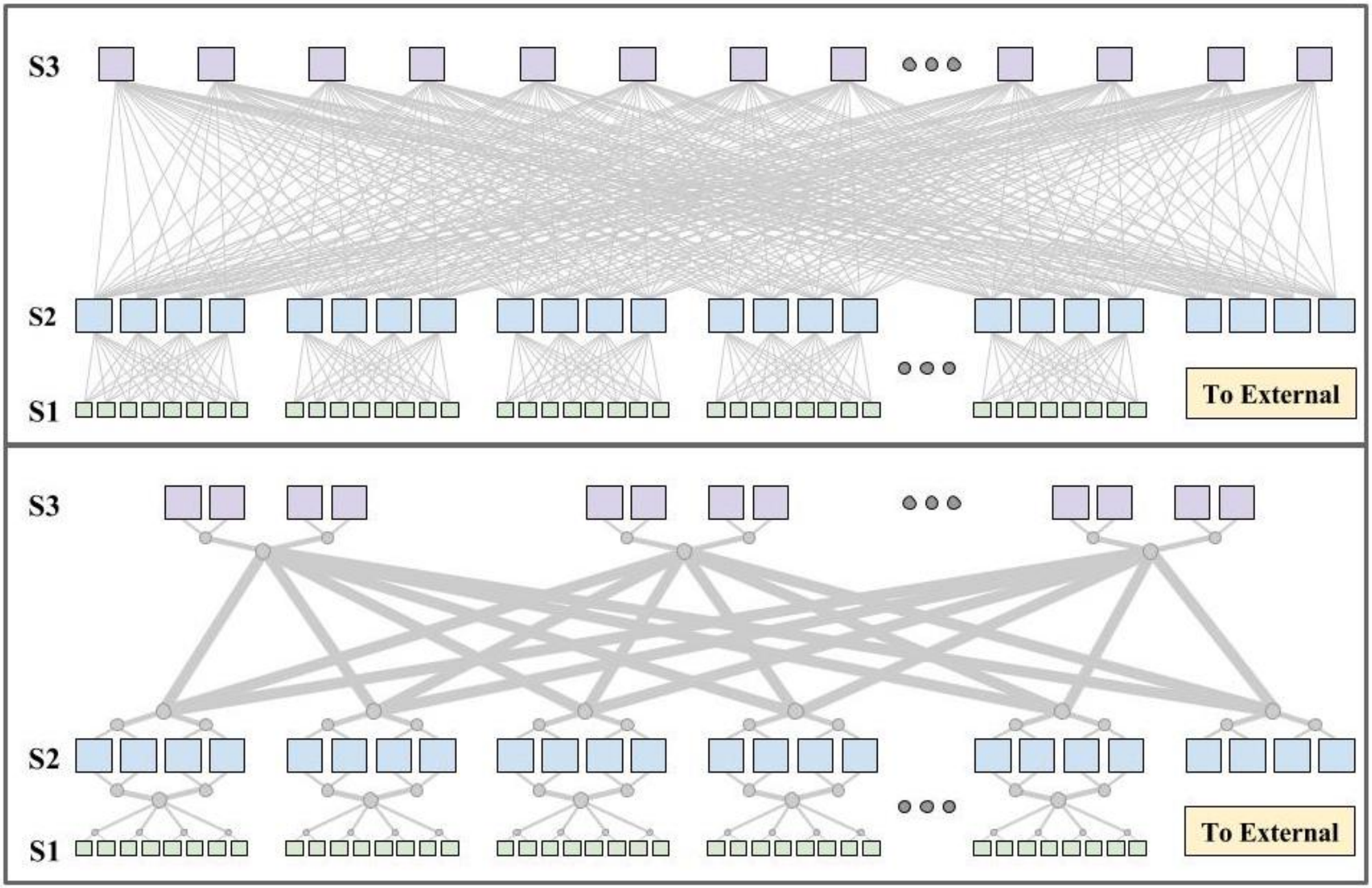
Source: A. Singh et al., “Jupiter rising: A decade of Clos topologies and centralized control in Google’s datacenter network,” ACM SIGCOMM’15.

Things have changed quite a lot

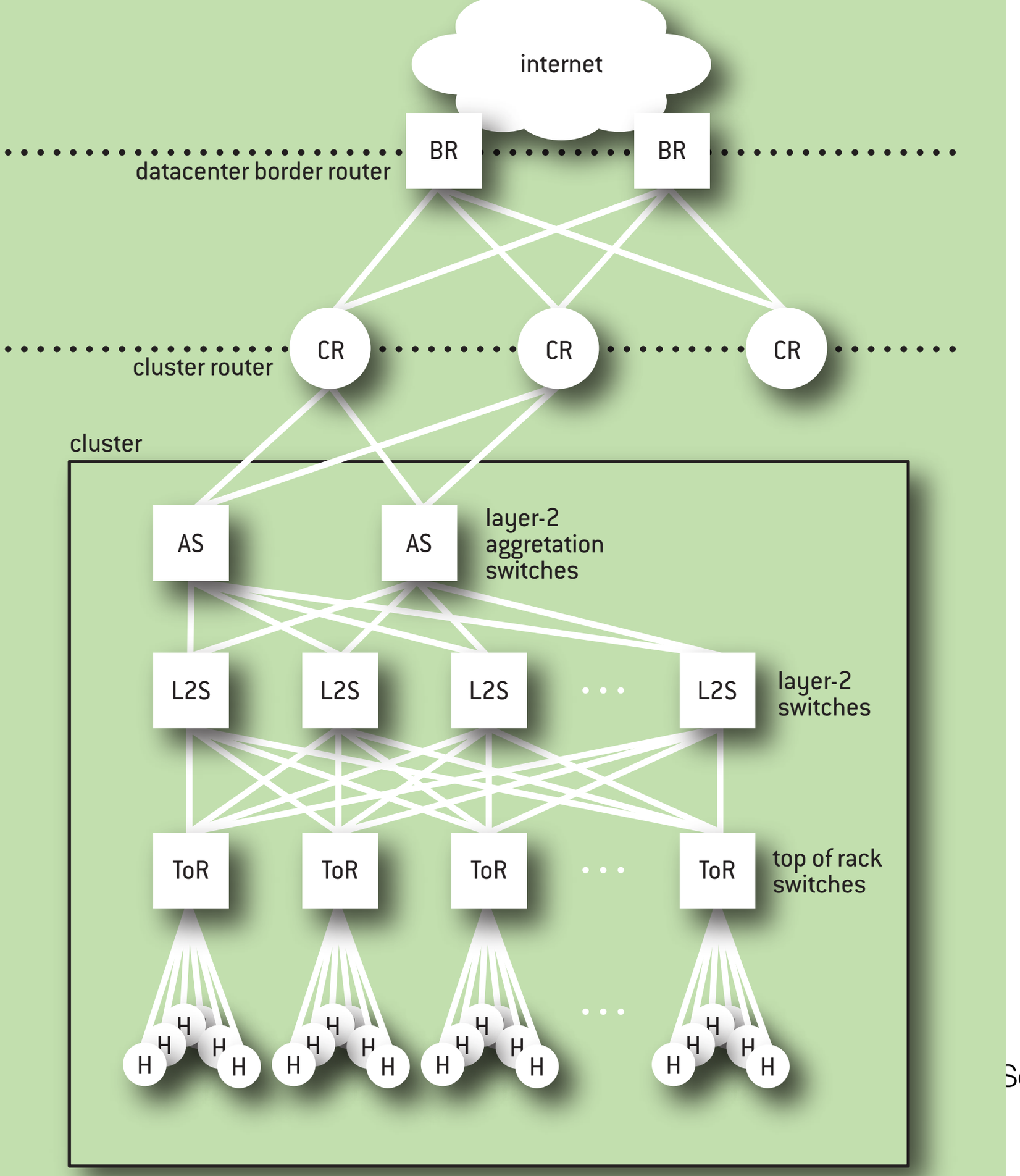


Source: A. Singh et al., "Jupiter rising: A decade of Clos topologies and centralized control in Google's datacenter network," ACM SIGCOMM'15.

When scaled up



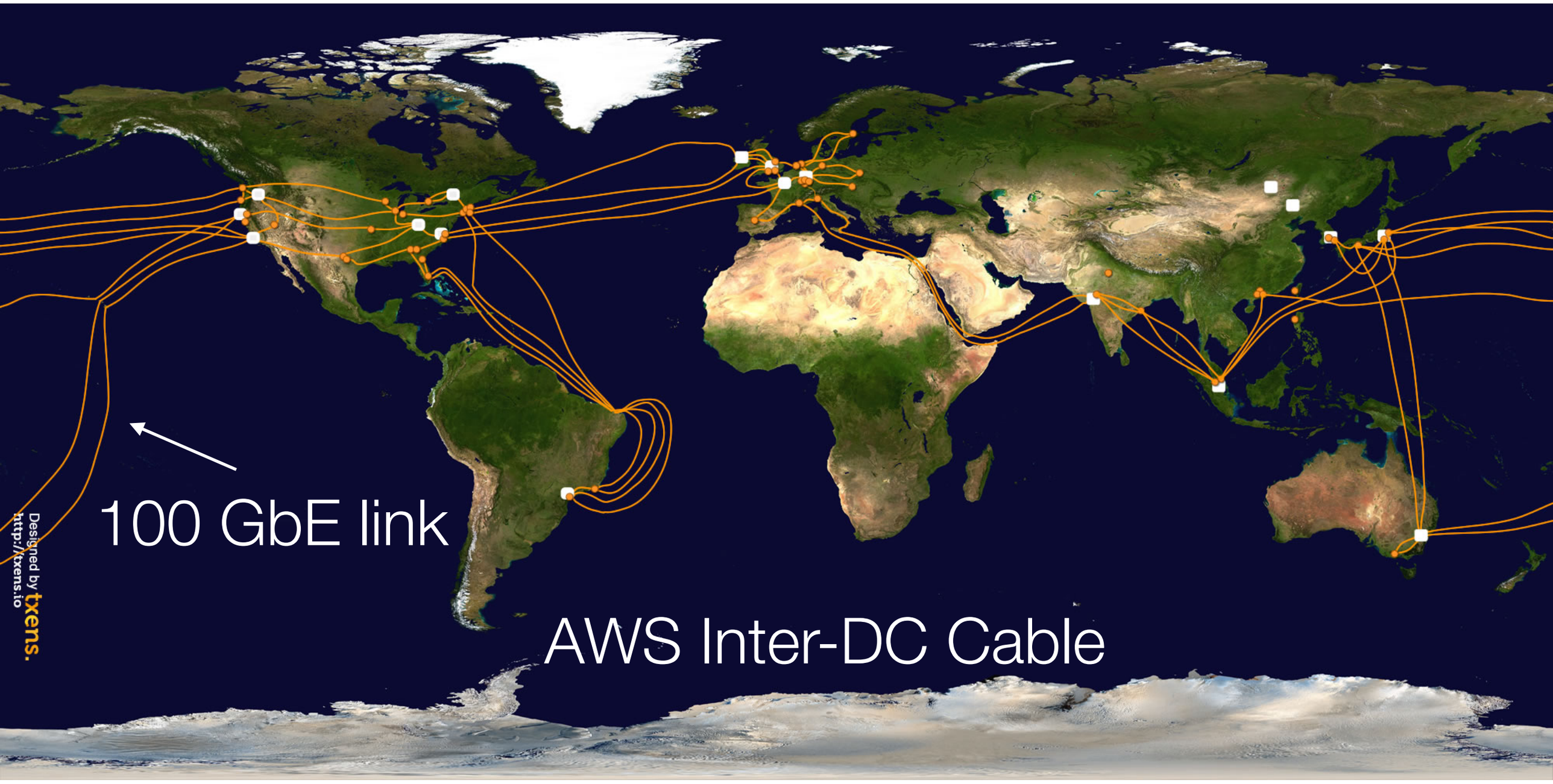
Source: A. Singh et al., “Jupiter rising: A decade of Clos topologies and centralized control in Google’s datacenter network,” ACM SIGCOMM’15.



Tree-like DC Network

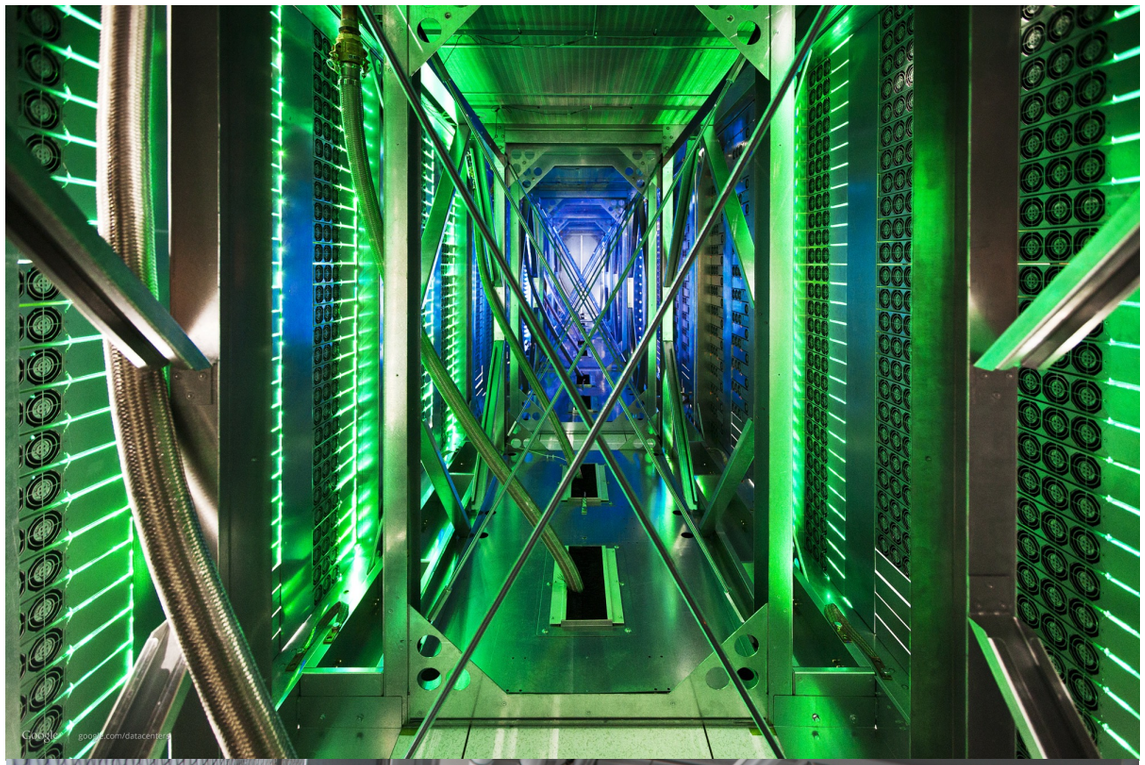
Source: http://queue.acm.org/detail.cfm?id=2208919

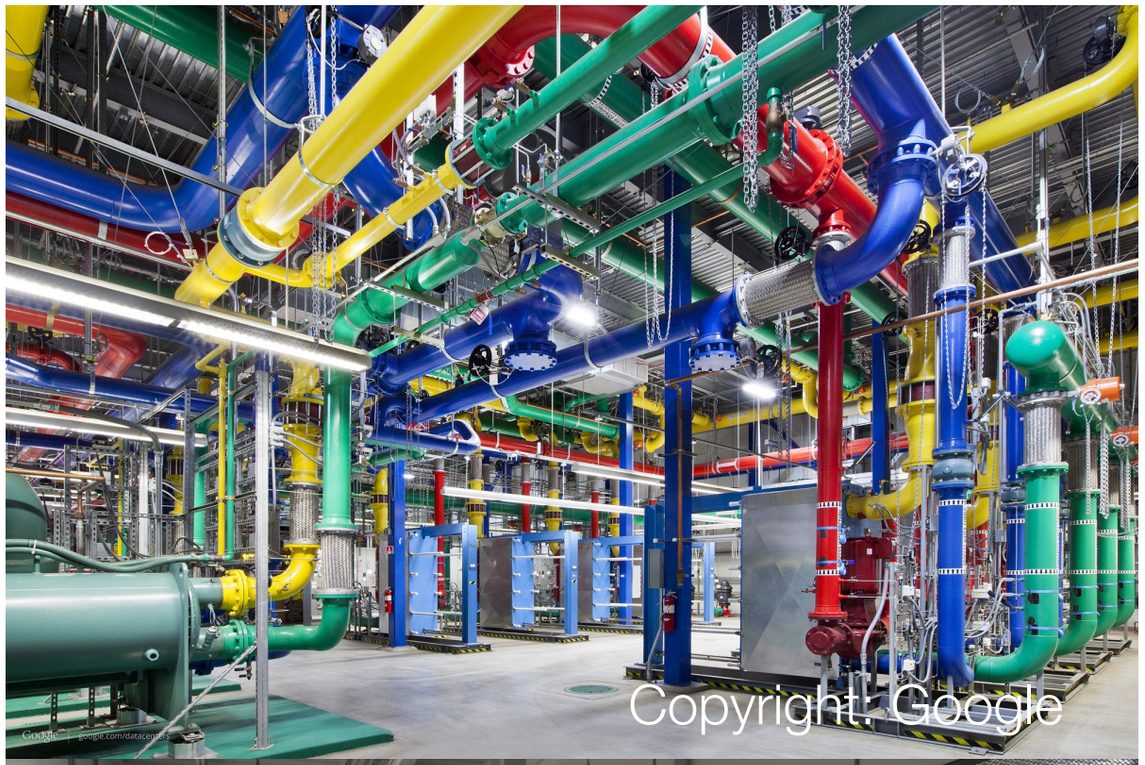
Inter-DC WAN



Source: txens, http://txens.io/index.php/2016/12/16/a-deep-dive-into-the-aws-network/

Cooling



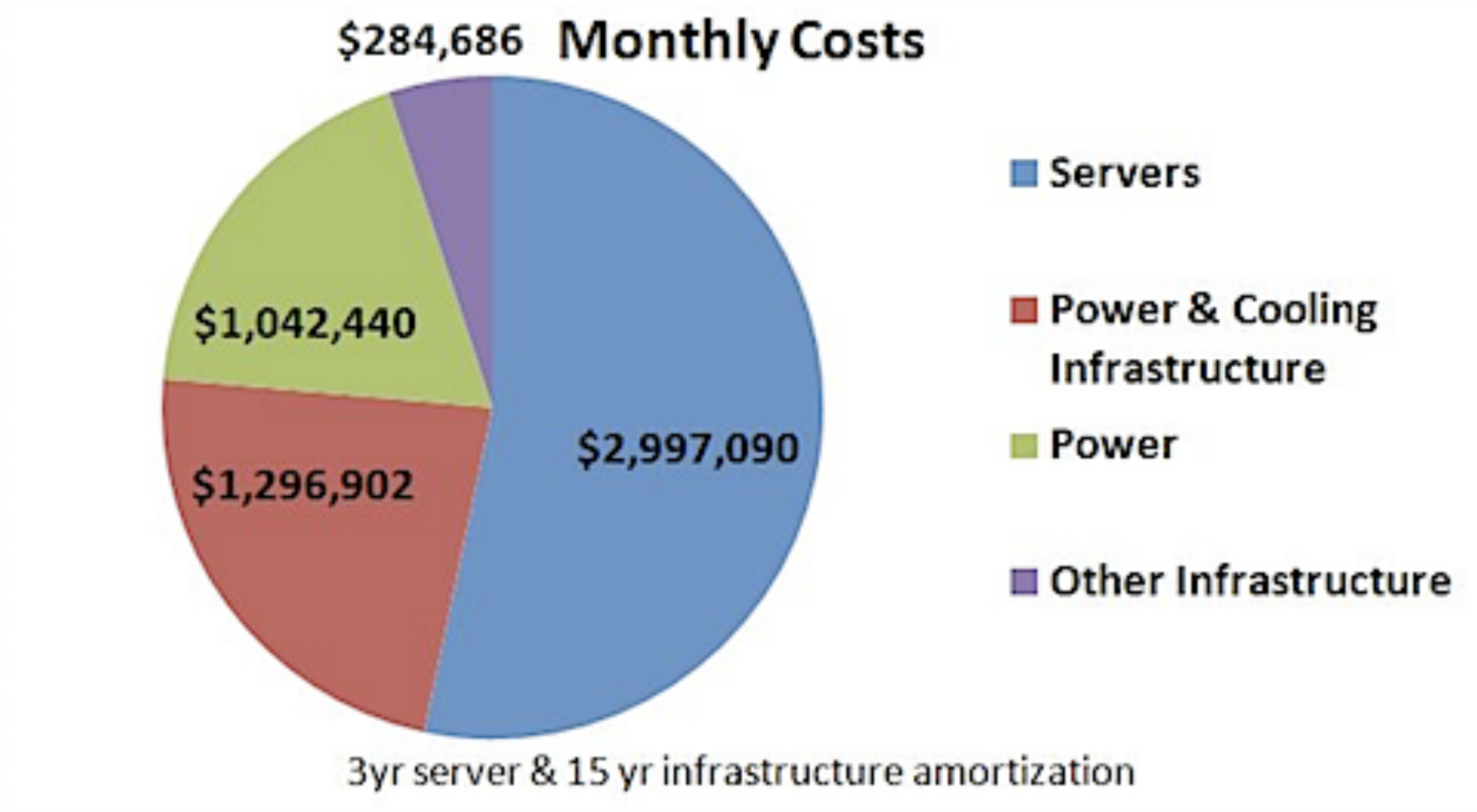


Power

Copyright: GigaOM

Copyright: Nation of Change

Cost Structure



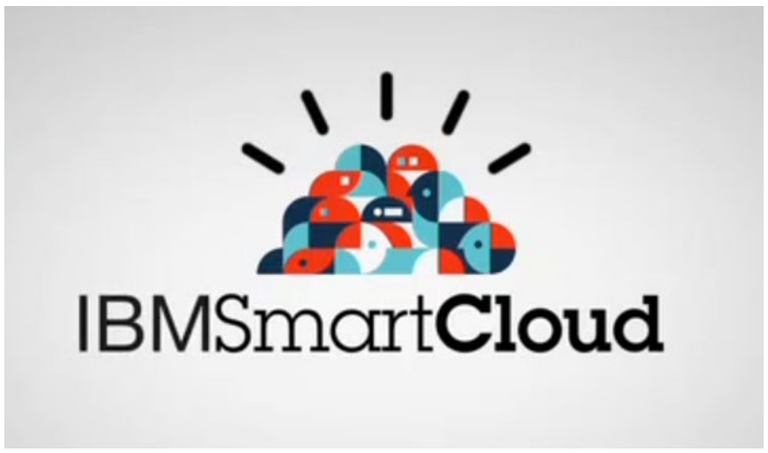
Source: James Hamilton, http://perspectives.mvdirona.com/2008/11/cost-of-power-in-large-scale-data-centers/

Explore Google Datacenter

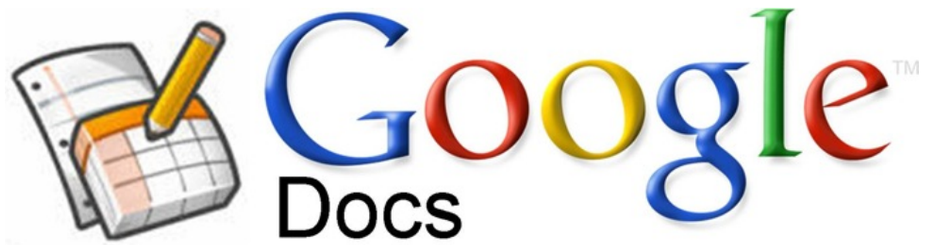
- https://www.youtube.com/watch?v=avP5d16wEp0

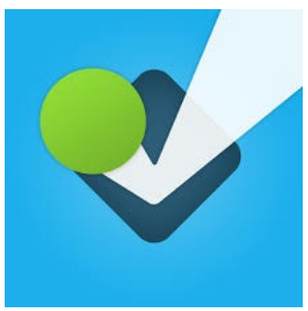
Cloud providers

- 



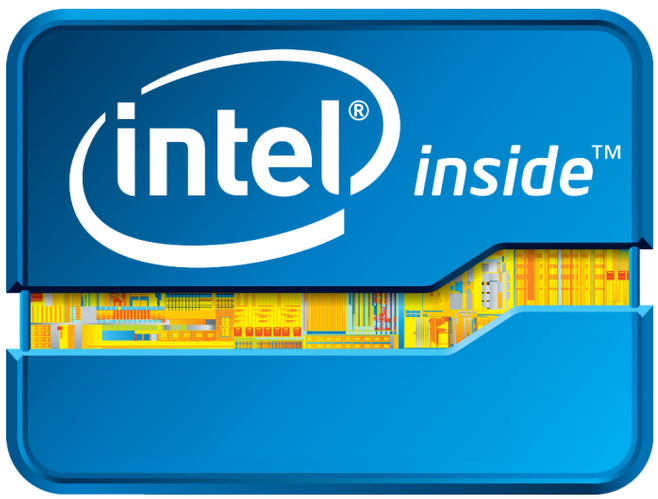
Cloud-based services



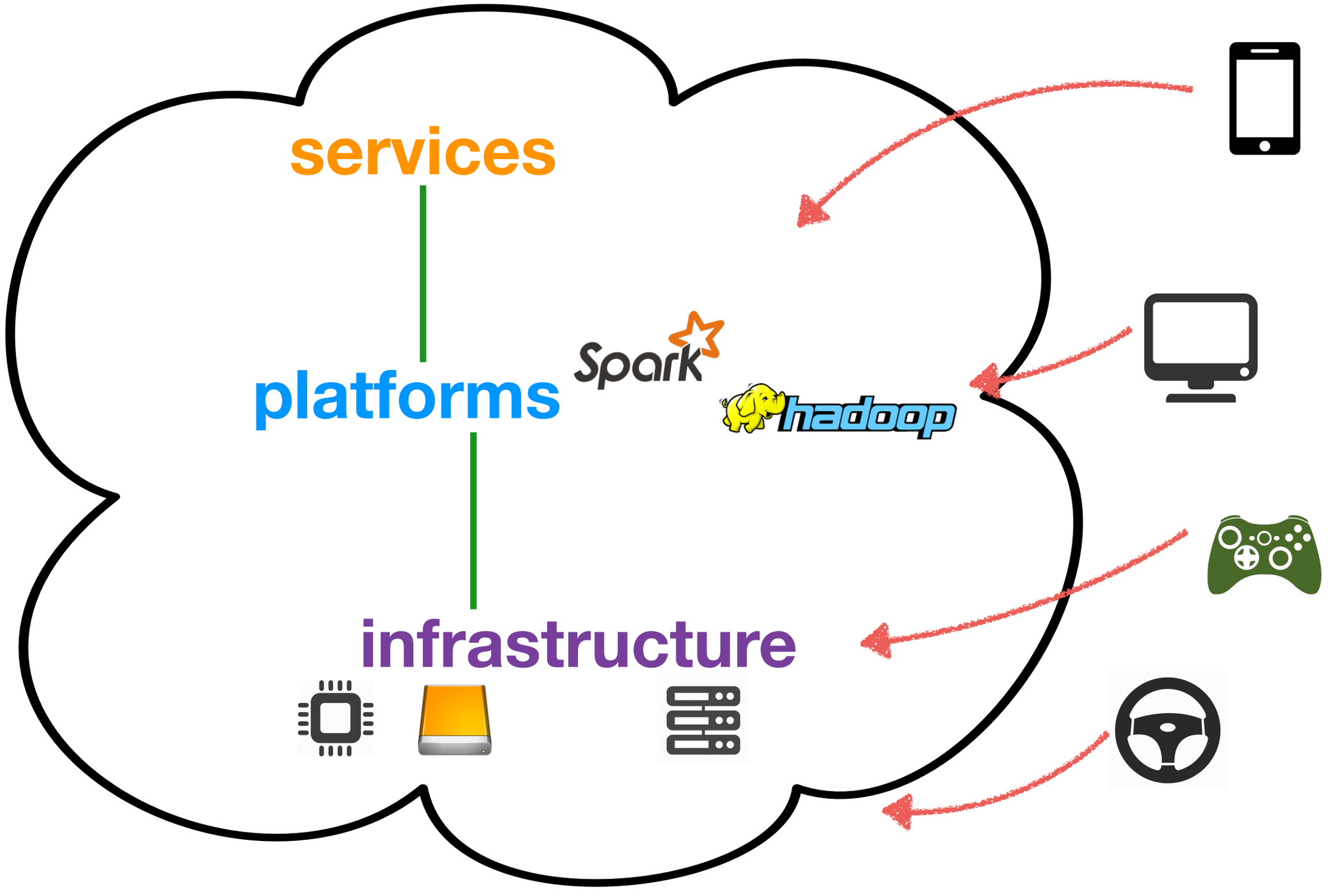


Social networking

Cloud vendors



So what is a cloud?



A long definition

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.

National Institute of Standards and Technology (NIST), U.S. Department of Commerce

On-demand computing delivered over the **Internet** with **pay-as-you-go** pricing

Utility Computing

- Computing as the 5th utility (after water, electricity, gas, and telephony)
  - Applications and computing resources delivered as a service over the Internet
  - Pay-as-you-go
- Provided by the hardwares and system softwares hosted in the datacenters

Infrastructure as software

- Cloud computing enables users to **stop thinking of infrastructure as hardware**

- Instead, **think of (and use) it as software**

Amazon EC2

https://youtu.be/TsRBftzZsQo

1 instance runs 1000 h =
1000 instances run 1 h

Revolutionary!

Suppose you open a startup and need 100 servers

What would you do traditionally, in a pre-cloud era?

Traditional computing model

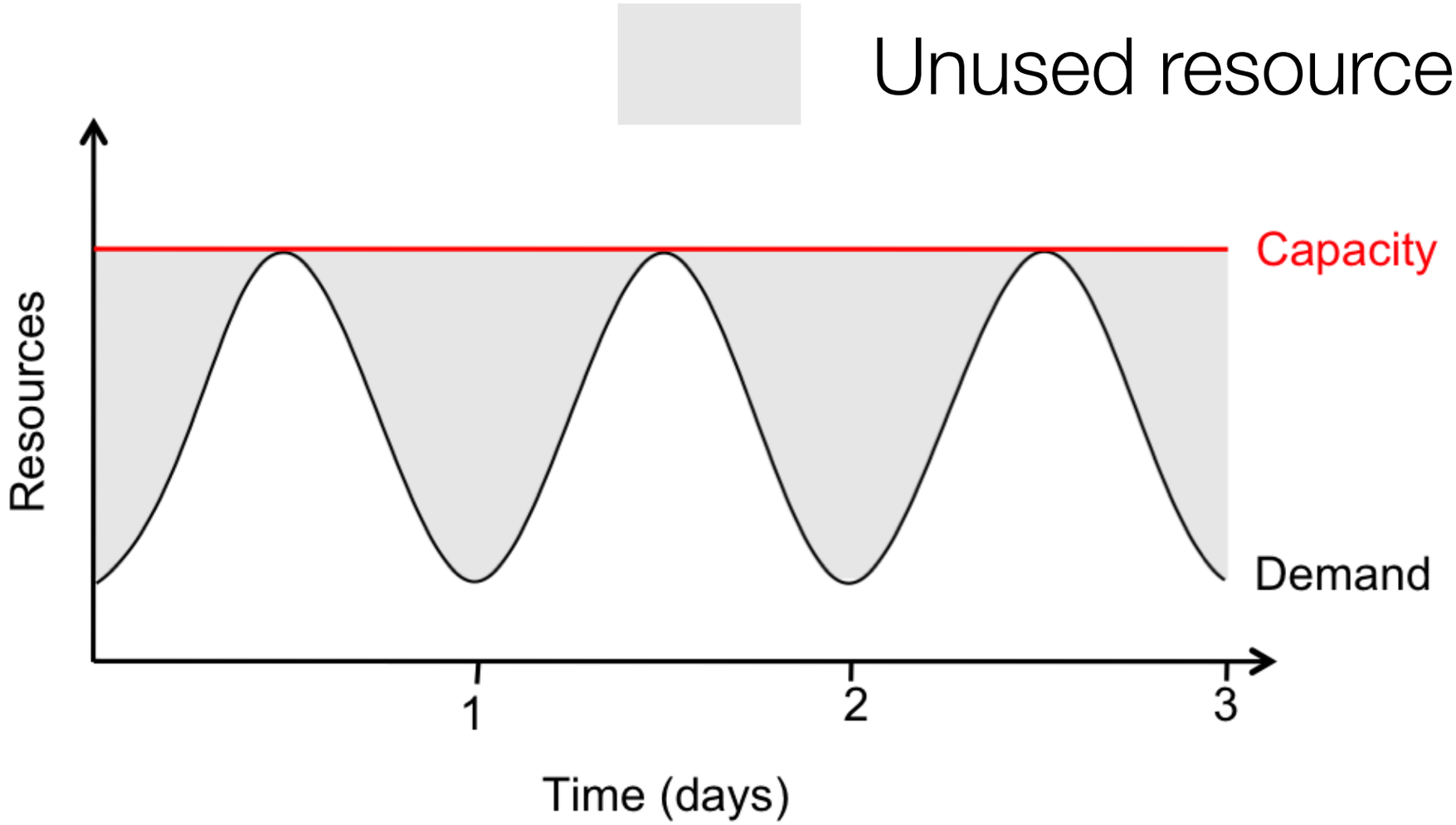
- Infrastructure as hardware
- Hardware solutions:
  - require space, staff, physical security, planning, capital expenditure
  - have a long hardware procurement cycle
  - need to provision capacity by guessing theoretical maximum peaks

Cloud computing model

- Infrastructure as software
- Flexible software solutions:
  - no upfront infrastructure investment
  - can change more quickly, easily, and cost-effectively than hardware solutions
  - eliminate the undifferentiated heavy-lifting tasks and labor works
  - Always-on services

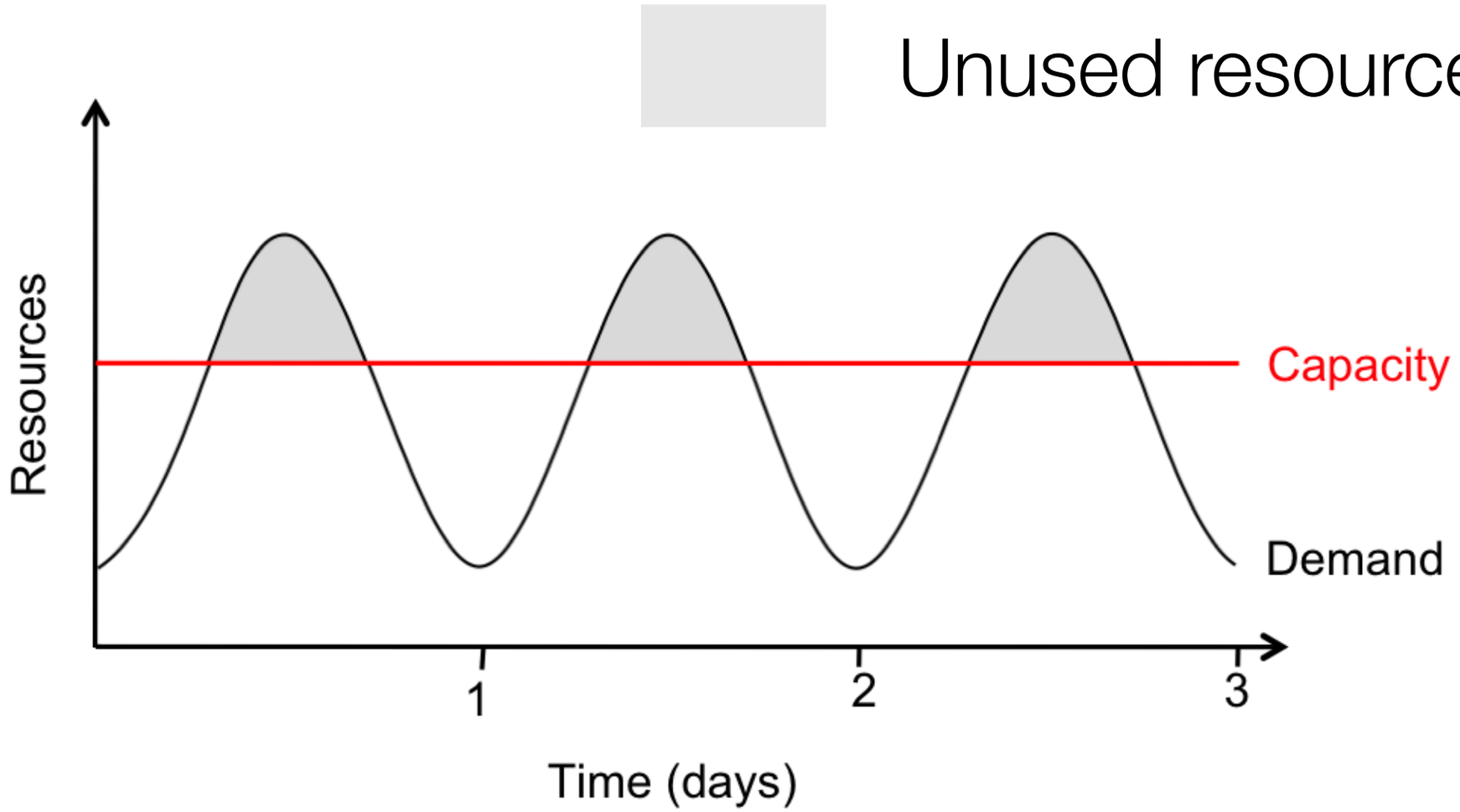
The demands are *elastic* — the 100 servers are only needed in peak time

Provisioning for peak load



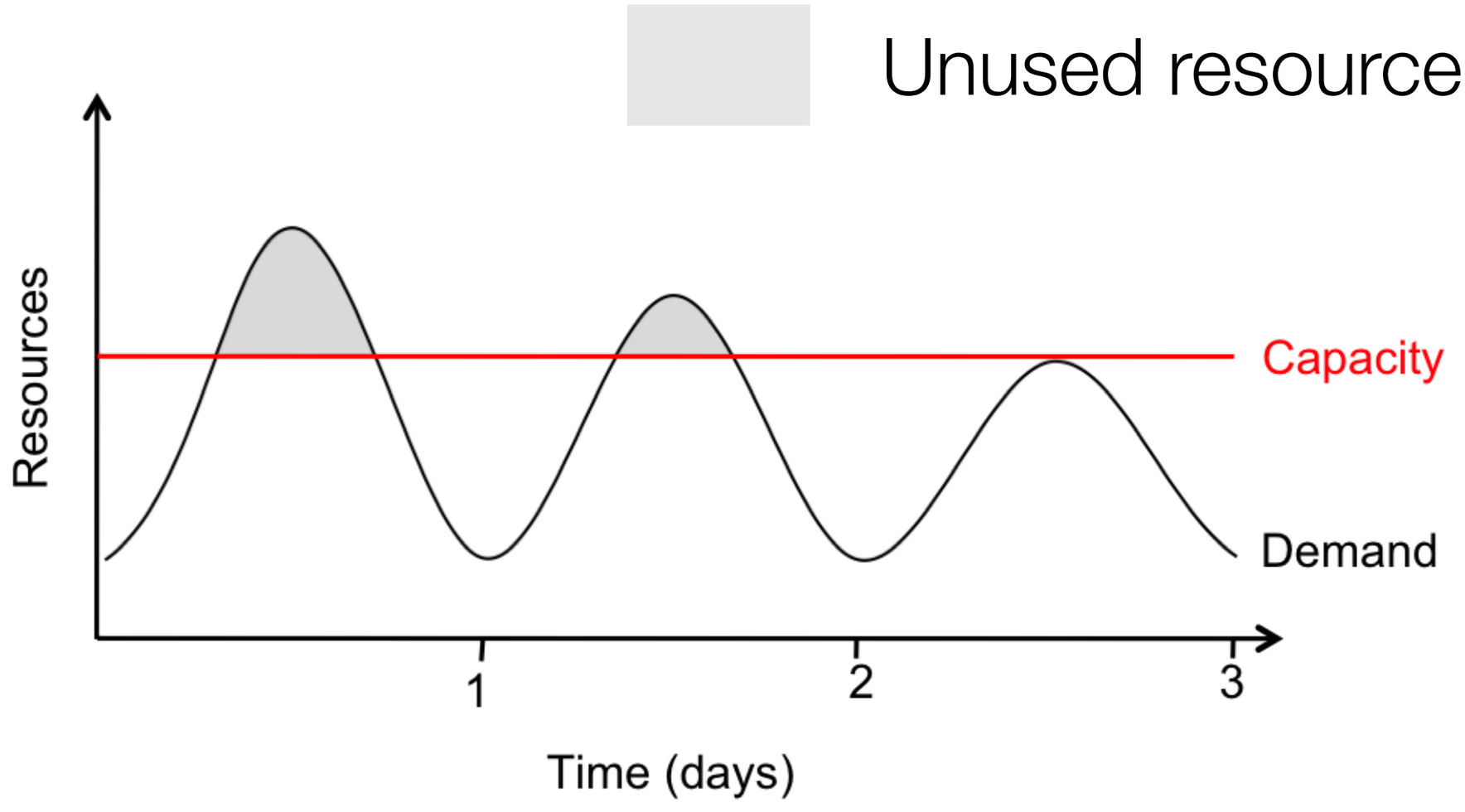
Source: Ambrust et al., “Above the clouds: A Berkeley’s view of Cloud Computing.”

Underprovisioning



Source: Ambrust et al., “Above the clouds: A Berkeley’s view of Cloud Computing.”

Underprovisioning

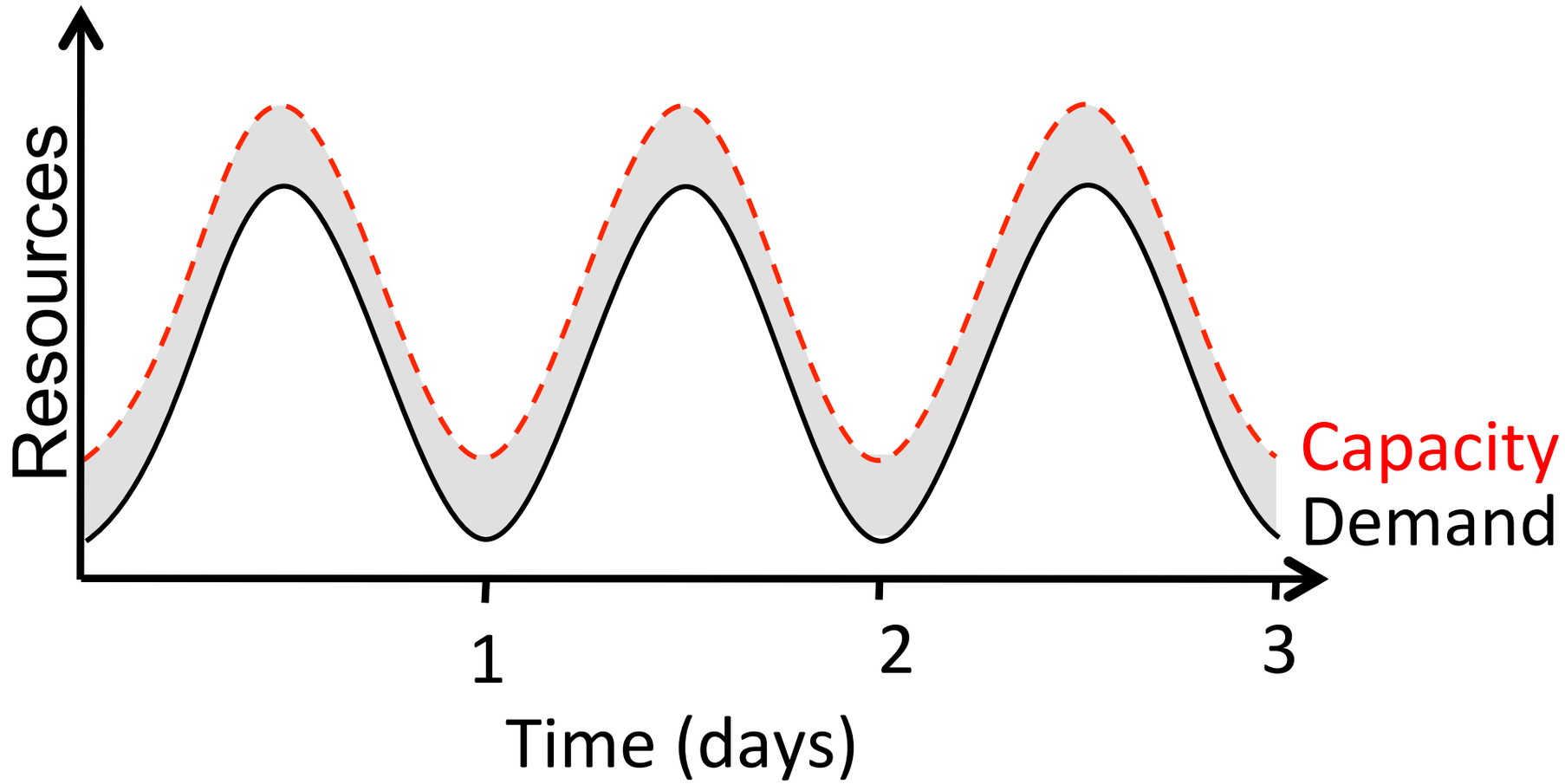


Source: Ambrust et al., “Above the clouds: A Berkeley’s view of Cloud Computing.”

Let’s do it in cloud!

Cloud provisioning on demand

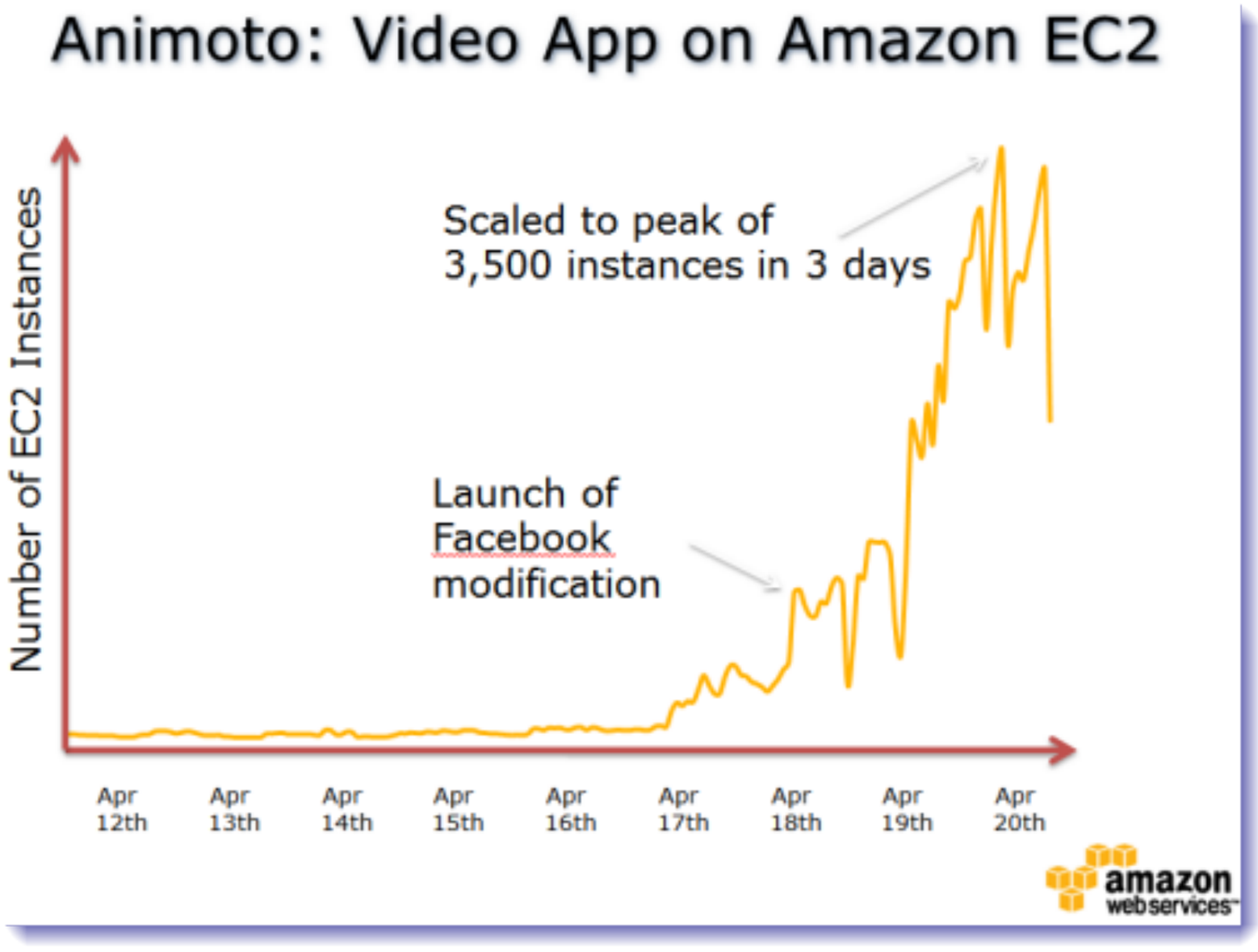
- Pay for what you used



An Animoto Case Study

https://youtu.be/VwDS6MexKEo

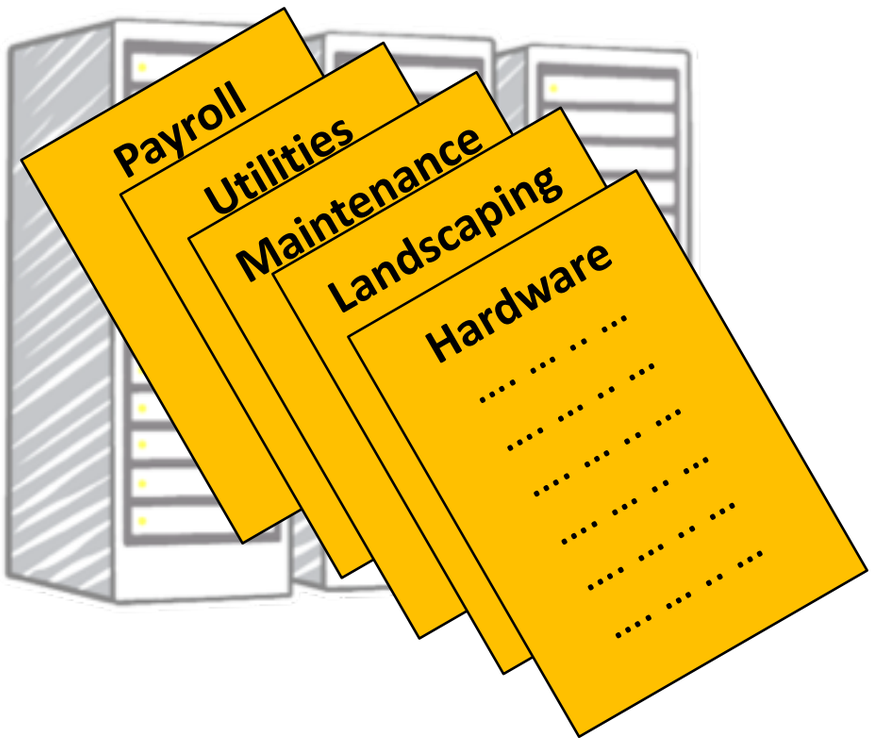
Animoto: Video App on Amazon EC2



Copyright: AWS & Animoto

Other benefits enabled by cloud

Stop spending money on running and maintaining datacenters

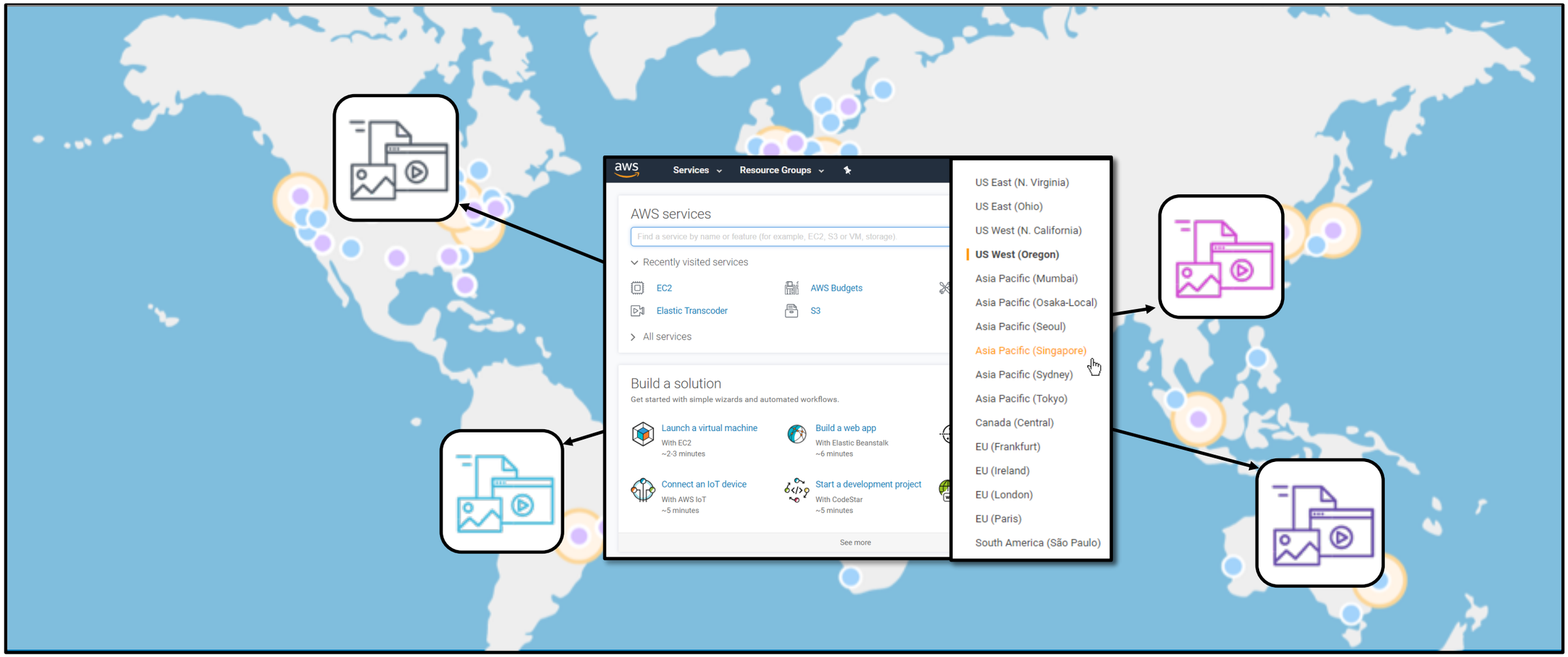


Running data centers

Business and customers

Other benefits enabled by cloud

Go global in minutes



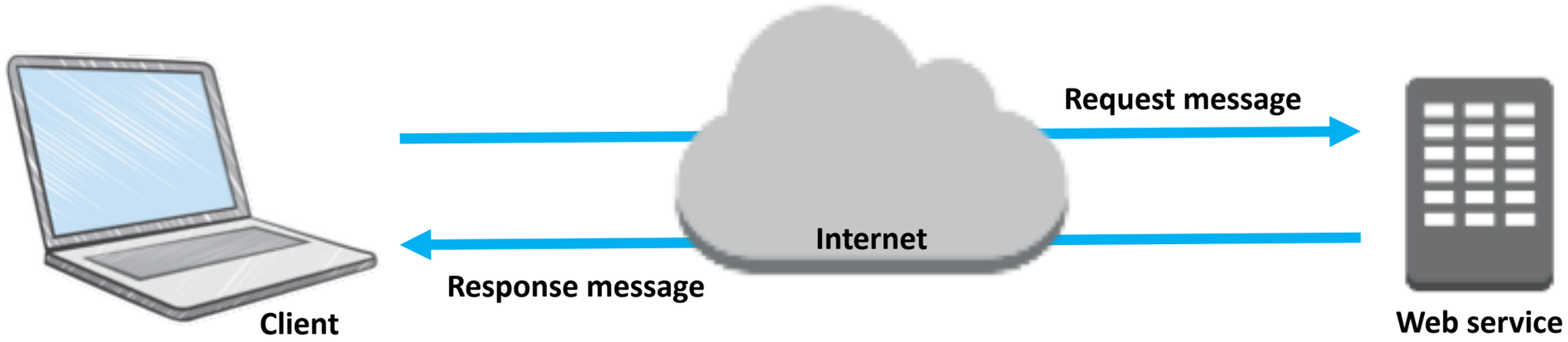
Summary: Why cloud?

- Better capital utilization
- The unit cost of on-demand capacity may be higher than the unit cost of fixed capacity; offset by no charge when capacity is not being used
- Elasticity, easy to scale up and down
- Access to complex infrastructure and resources without internal resources

Amazon Web Services

What are web services?

- A piece of software that makes itself available over the internet and uses a **standardized format**—such as XML or JSON—for the request and the response of an **application programming interface (API) interaction**.



What is AWS?

- AWS is **a secure cloud platform** that offers **a broad set of global cloud-based products**

- provides **on-demand access** to compute, storage, network, database, and other IT resources and management tools

What is AWS?

- AWS is **a secure cloud platform** that offers **a broad set of global cloud-based products**
- AWS provides **on-demand access** to compute, storage, network, database, and other IT resources and management tools
- AWS offers **flexibility** and **pay-as-you-go pricing**
- AWS services work together like building blocks

Categories of AWS services

Compute services –

- Amazon EC2
- AWS Lambda
- AWS Elastic Beanstalk
- Amazon EC2 Auto Scaling
- Amazon ECS
- Amazon EKS
- Amazon ECR
- AWS Fargate

Storage services –

- Amazon S3
- Amazon S3 Glacier
- Amazon EFS
- Amazon EBS

Management and Governance services –

- AWS Trusted Advisor
- AWS CloudWatch
- AWS CloudTrail
- AWS Well-Architected Tool
- AWS Auto Scaling
- AWS Command Line Interface
- AWS Config
- AWS Management Console
- AWS Organizations

Database services –

- Amazon RDS
- Amazon DynamoDB
- Amazon Redshift
- Amazon Aurora

Security, Identity, and Compliance services –

- AWS IAM
- Amazon Cognito
- AWS Shield
- AWS Artifact
- AWS KMS

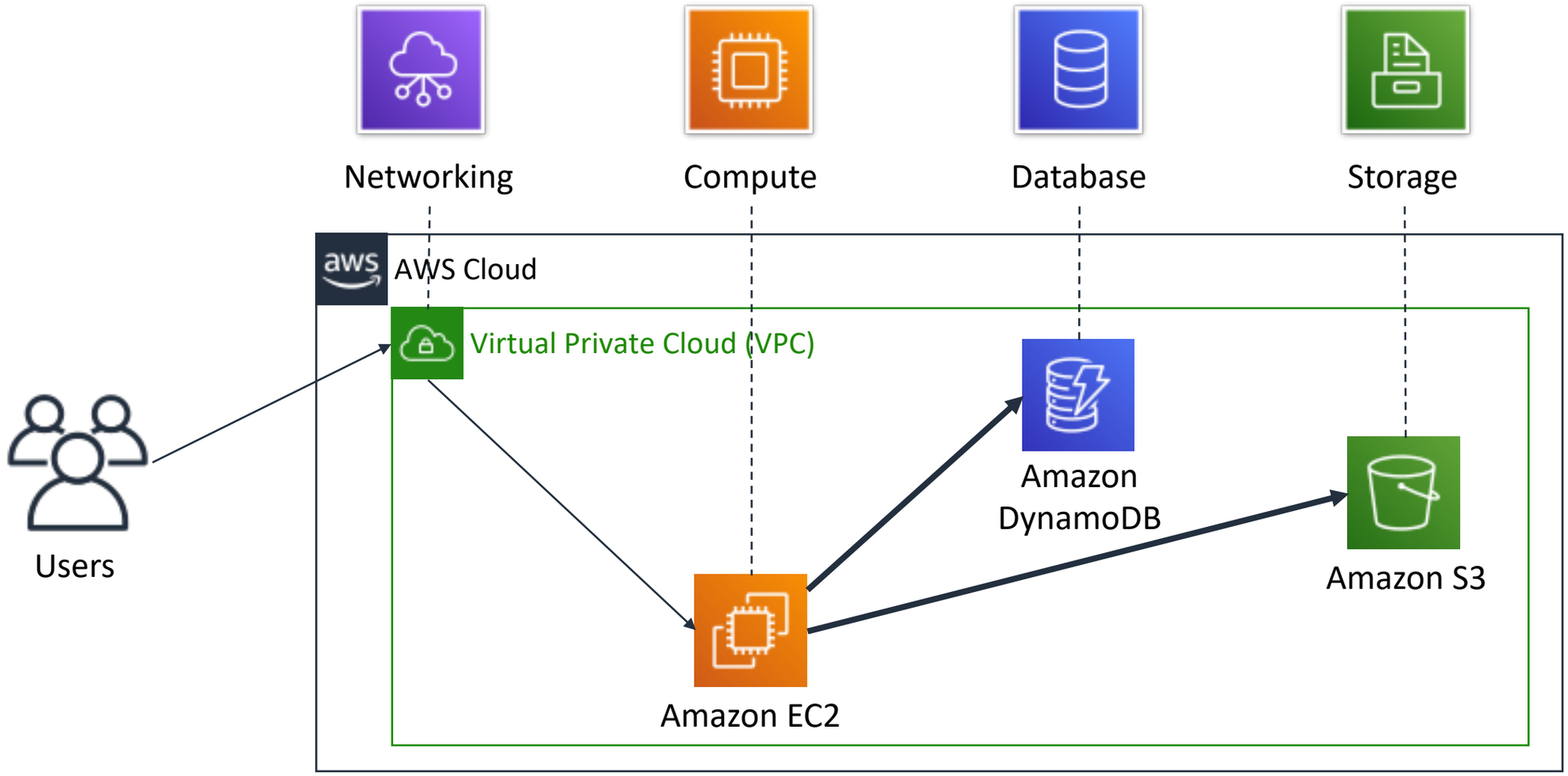
Networking and Content Delivery services –

- Amazon VPC
- Amazon Route 53
- Amazon CloudFront
- Elastic Load Balancing

AWS Cost Management services –

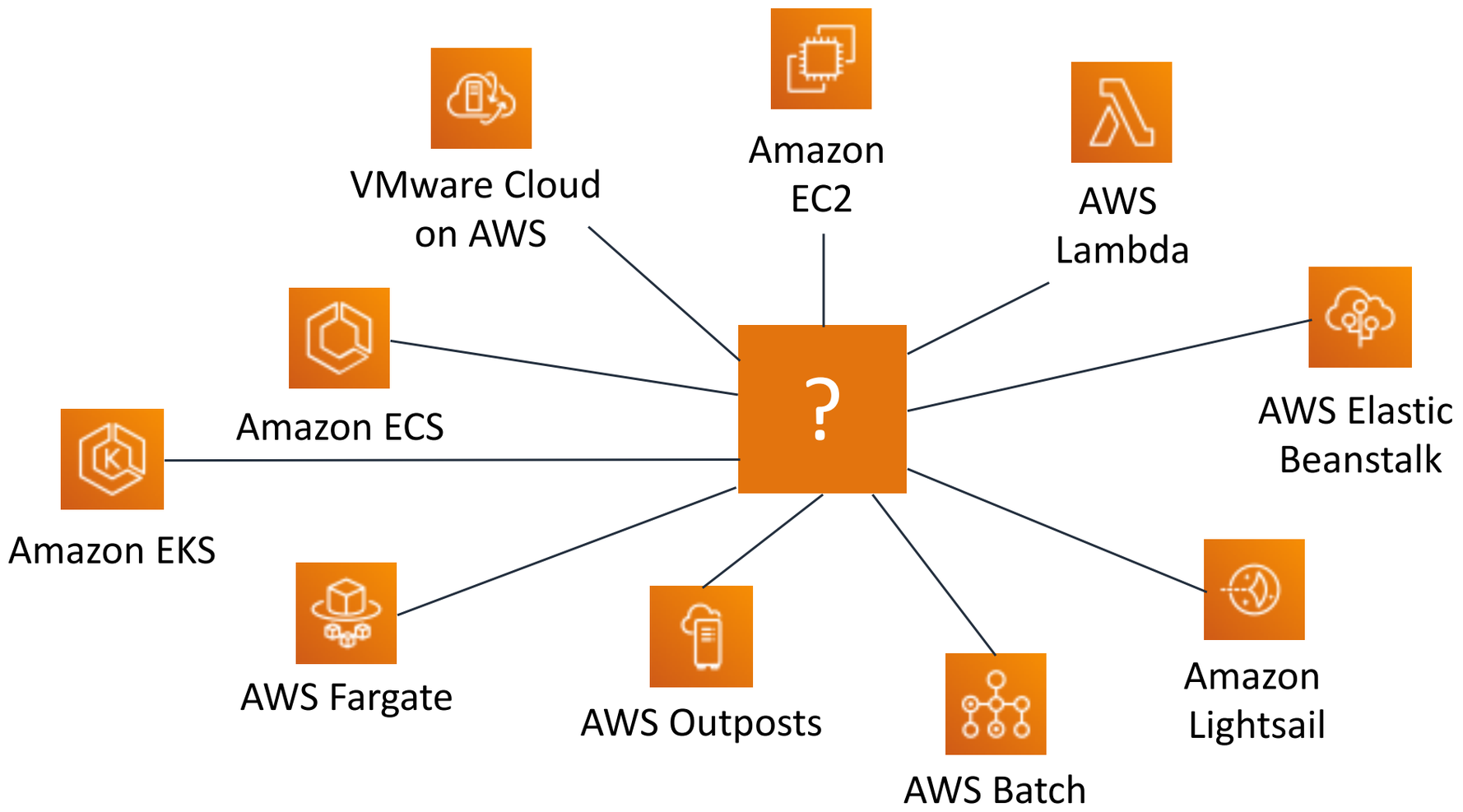
- AWS Cost & Usage Report
- AWS Budgets
- AWS Cost Explorer

Simple solution example



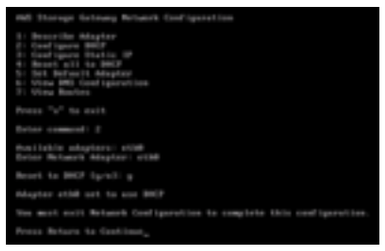
Choosing a service

- The service you choose depends on your business goals and technology requirements



3 ways to interact w/ AWS

- AWS Management Console
  - easy-to-use graphical interface



- Command Line Interface (AWS CLI)
  - access to services by discrete commands or scripts

- Software Development Kits (SDKs)
  - access to services directly from your code (e.g., Java, Python, etc.)

Credit

- Some slides are adapted from Prof. Hong Xu's slides for CS 4296/5296 in CityU
- Some slides are adapted from AWS Academy Class (Cloud foundations)