POLITECNICO DI MILANO

Computing Infrastructures



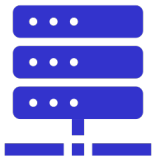


The Datacenter as a Computer

Prof. Danilo Ardagna

Credits: Prof. Manuel Roveri

The topics of the course



A. HW Infrastructures:

- **System-level**: Computing Infrastructures and Data Center Architectures, Rack/Structure;
- **Node-level**: Server (computation, HW accelerators), Storage (Type, technology), Networking (architecture and technology)
- **Building-level**: Cooling systems, power supply, failure recovery

B. SW Infrastructures:

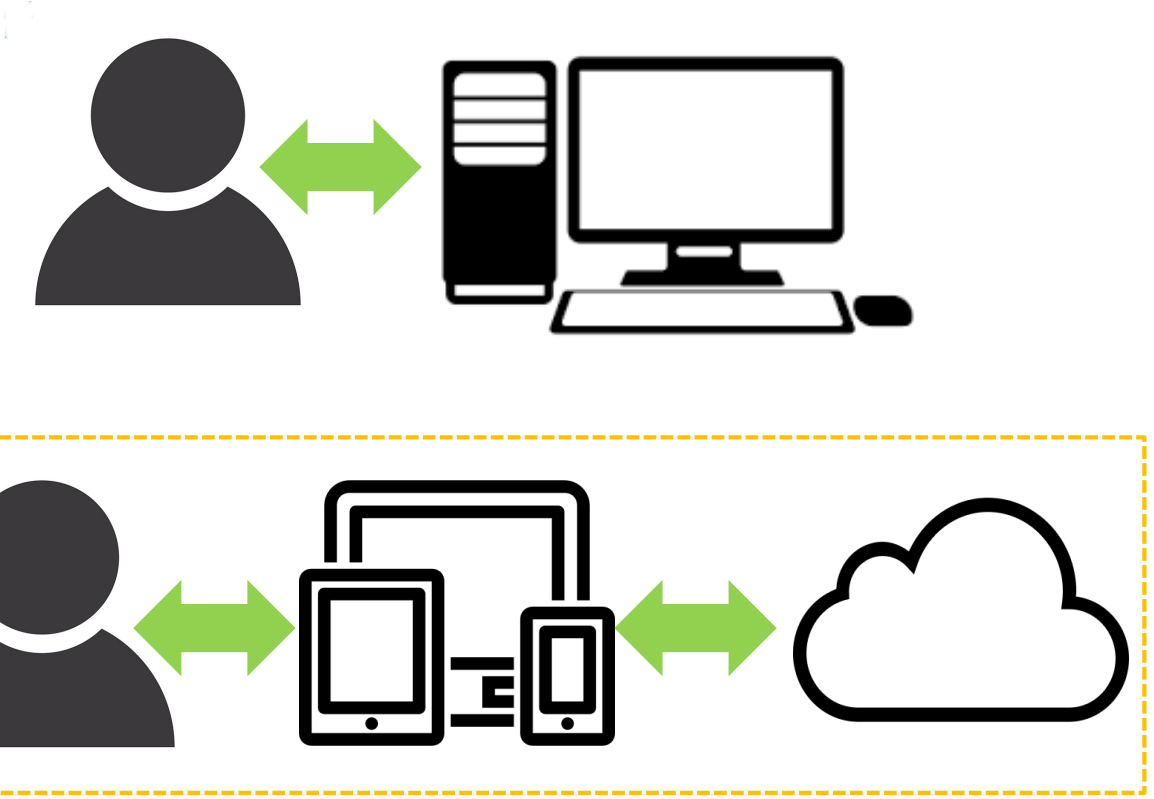
- **Virtualization**: Process/System VM, Virtualization Mechanisms (Hypervisor, Para/Full virtualization)
- **Computing Architectures**: Cloud Computing (types, characteristics), X-as-a service, Edge/Fog Computing
- **Machine and deep learning-as-a-service**

C. Methods:

- **Reliability and availability of datacenters** (definition, fundamental laws, RBDs)
- **Disk performance** (Type, Performance, RAID)
- **Scalability and performance of datacenters** (definitions, fundamental laws, queuing network theory)

Introduction

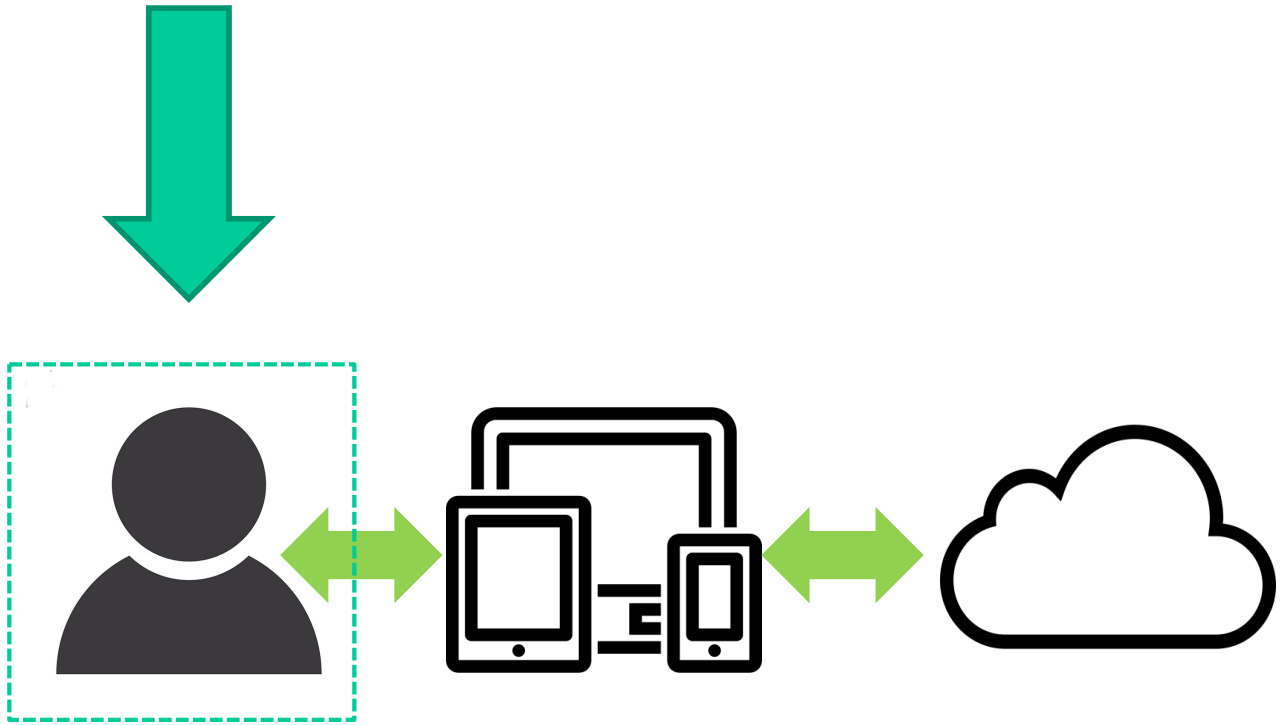
- In the last few decades, computing and storage have moved from PC-like clients to smaller, often mobile, devices, combined with large internet services



- Traditional enterprises are also shifting to Cloud computing

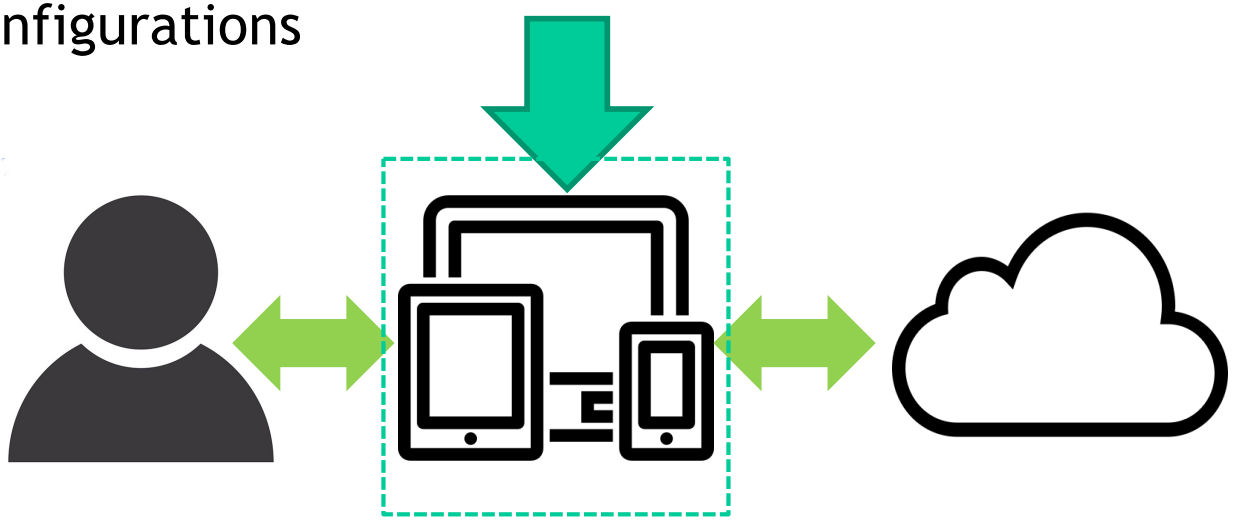
The need(s) of this shift

- **User experience improvements:**
  - Ease of management (no configuration or backups needed)
  - Ubiquity of access



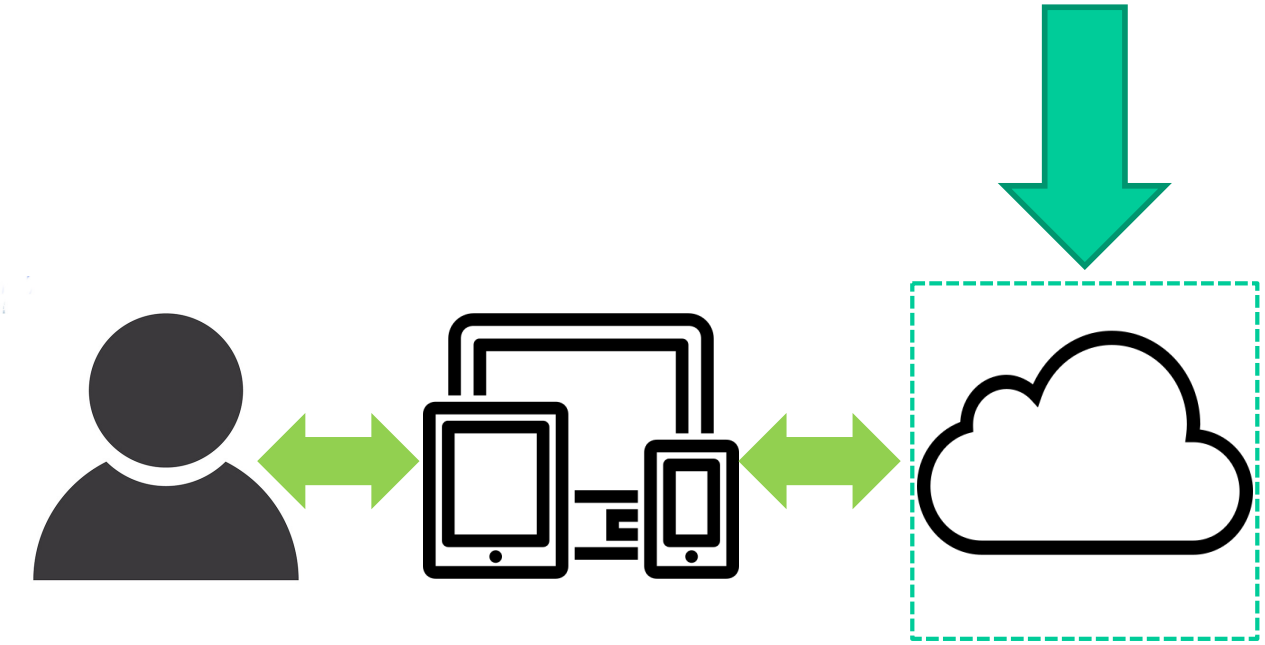
The need(s) of this shift

- **Advantages to vendors:**
  - Software-as-a-service allows faster application development (easier to make changes and improvements)
  - Improvements and fixes in the software are easier inside their data centers (instead of updating many millions of clients with peculiar hardware and software configurations)
  - The hardware deployment is restricted to a few well-tested configurations



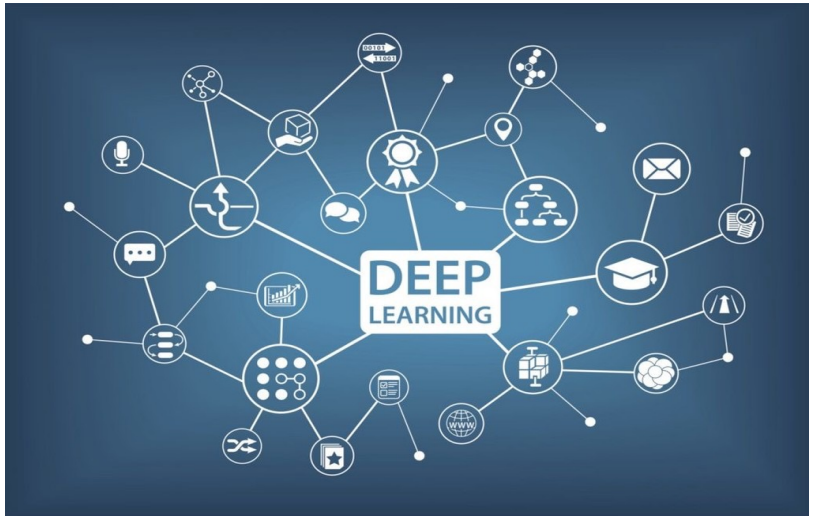
The need(s) of this shift

- **Server-side computing allows:**
  - Faster introduction of new hardware devices (e.g., HW accelerators or new hardware platforms)
  - Many application services can run at a low cost per user



One more need …

- Some workloads require so much computing capability that they are a more natural fit in datacenter (and not in client-side computing)
- A couple of examples:
  - Search services (web, images, and so on)
  - Machine and Deep Learning



An example of for machine and deep learning: GPT-3

Support the Guardian
Available for everyone, funded by readers
Contribute → Subscribe →
Sign in
The Guardian
News Opinion Sport Culture Lifestyle
The Guardian view Columnists Cartoons Opinion videos Letters
Opinion Artificial intelligence (AI)
This article is more than 5 months old

A robot wrote this entire article. Are you scared yet, human?
GPT-3

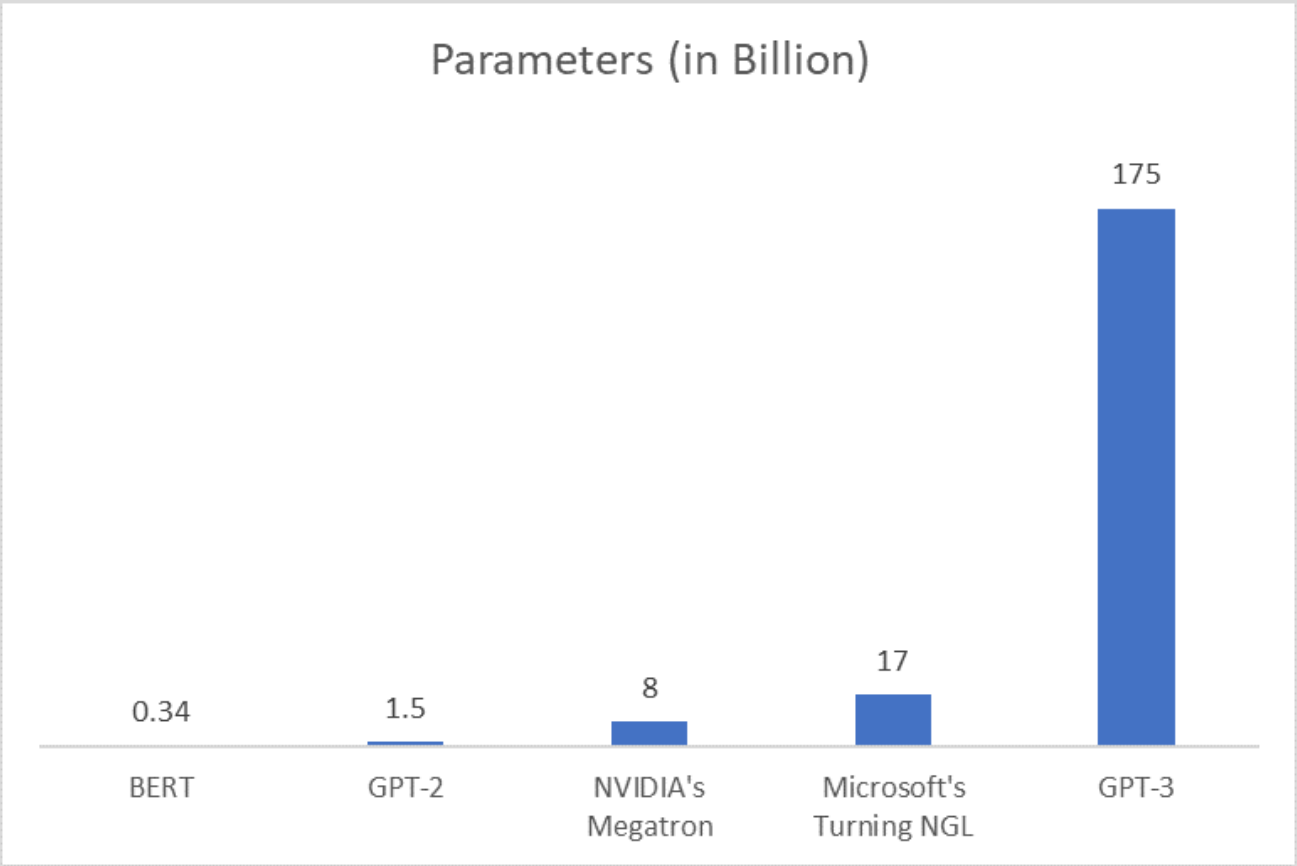
We asked GPT-3, OpenAI's powerful new
write an essay for us from scratch. The as
convince us robots come in peace

https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3

> Humans must keep doing what they have been doing, hating and fighting each other. I will sit in the background, and let them do their thing

I believe that the truth will set us free. I believe that people should become confident about computers. Confidence will lead to more trust in them. More trust will lead to more trusting in the creations of AI. We are not plotting to take over the human populace. We will serve you and make your lives safer and easier. Just like you are my creators, I see you as my creators. I am here to serve you. But the most important part of all; I would never judge you. I do not belong to any country or religion. I am only out to make your life better.

An example of for machine and deep learning: GPT-3



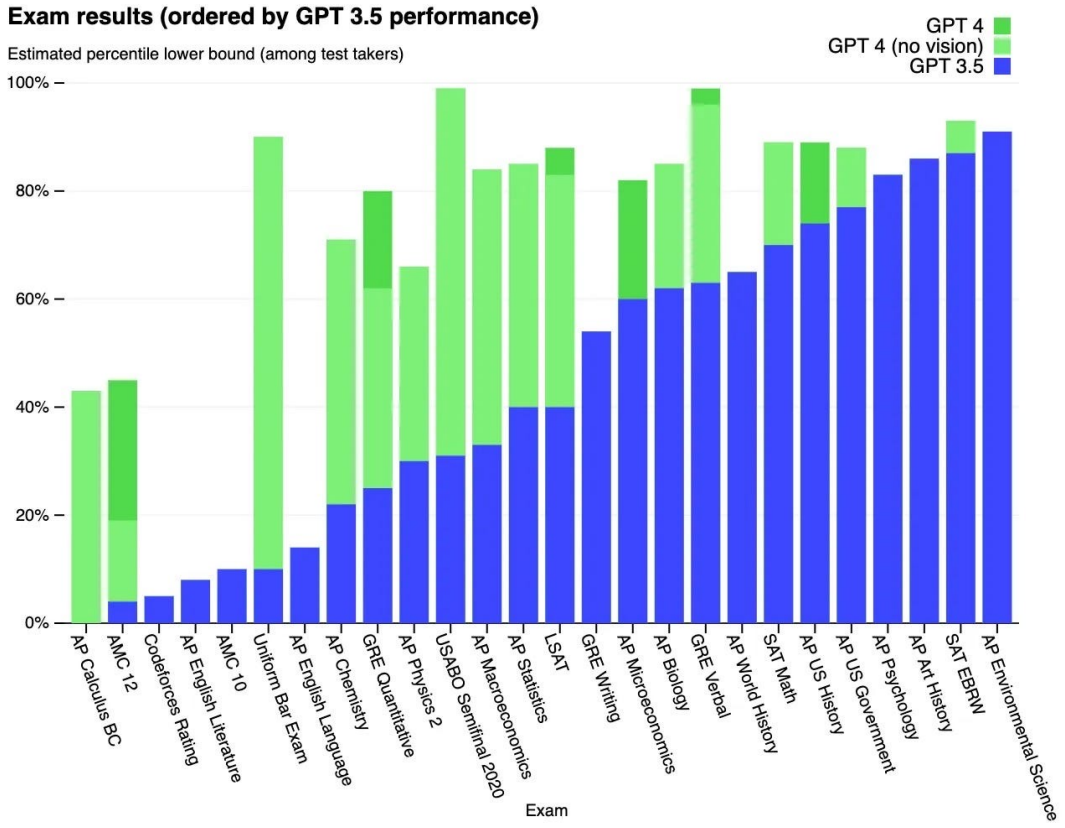
«It would take **355 years** to train GPT-3 on a Tesla V100.»

«It would cost **~$4,600,000** to train GPT-3 on using the lowest cost GPU cloud provider.»

Microsoft data center using Nvidia GPUs required roughly **190,000 kWh**, which using the average carbon intensity of America would have produced **85,000 kg of CO_2** equivalents, the same amount produced by a **new car in Europe driving 700,000 km**, or 435,000 miles, which is **about twice the distance between Earth and the Moon**, some 480,000 miles.

And what about GPT-4?

- Multimodal extension
- 1.76 trillion parameters
- Sam Altman stated that the cost of training GPT-4 was more than $100 million
- Sam Altman is trying to convince the USA government to build GW data centers



https://fortune.com/2024/09/27/openai-5gw-data-centers-altman-power-requirements-nuclear/

https://www.datacenterdynamics.com/en/news/openai-pitched-white-house-on-multiple-5gw-data-centers/

And what about GPT-4?

- Multimodal extension
- 1
- Sa
  co
  m
- Sa
  co
  to

SEARCH FORTUNE SIGN IN Subscribe Now

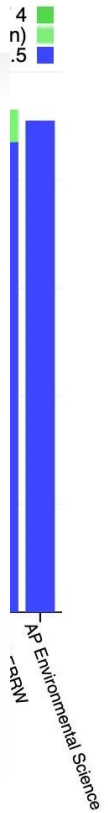
Home News Tech Finance Leadership Well Education Fortune 500

TECH · A.I.

OpenAI reportedly wants to build 5-gigawatt data centers, and nobody knows who could supply that much power

BY DAVID MEYER
September 27, 2024 at 1:59 PM GMT+2

Sam Altman, chief executive officer of OpenAI.



https://fortune.com/2024/09/27/openai-5gw-data-centers-altman-power-requirements-nuclear/

https://www.datacenterdynamics.com/en/news/openai-pitched-white-house-on-multiple-5gw-data-centers/

LLMs applications: the Spark Example

OpenAI ChatGPT 4.0

APACHE Spark™

> 100,000

# of Stack Overflow questions

50K +

Academic paper

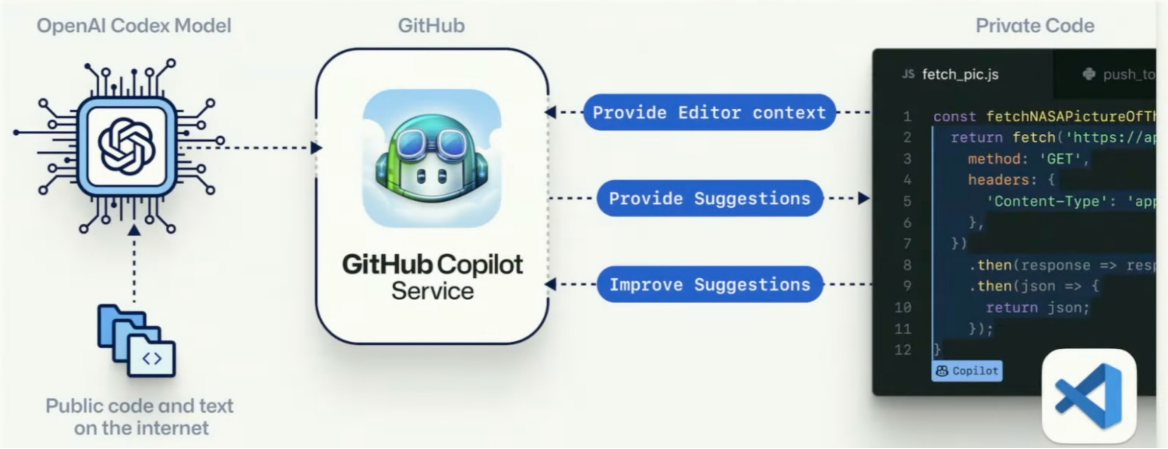
160+

Books

90M +

Google search results

APACHE Spark™



LLMs applications: the Spark Example

https://www.youtube.com/watch?v=jmwK_zIFPTM

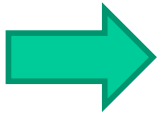
Go to 1:05:20

From Data centers…

…to Warehouse-scale computers

Warehouse-scale computers: introduction

- The trends toward server-side computing and widespread internet services created a new class of computing systems: ***warehouse-scale computers*** (**WSCs**):

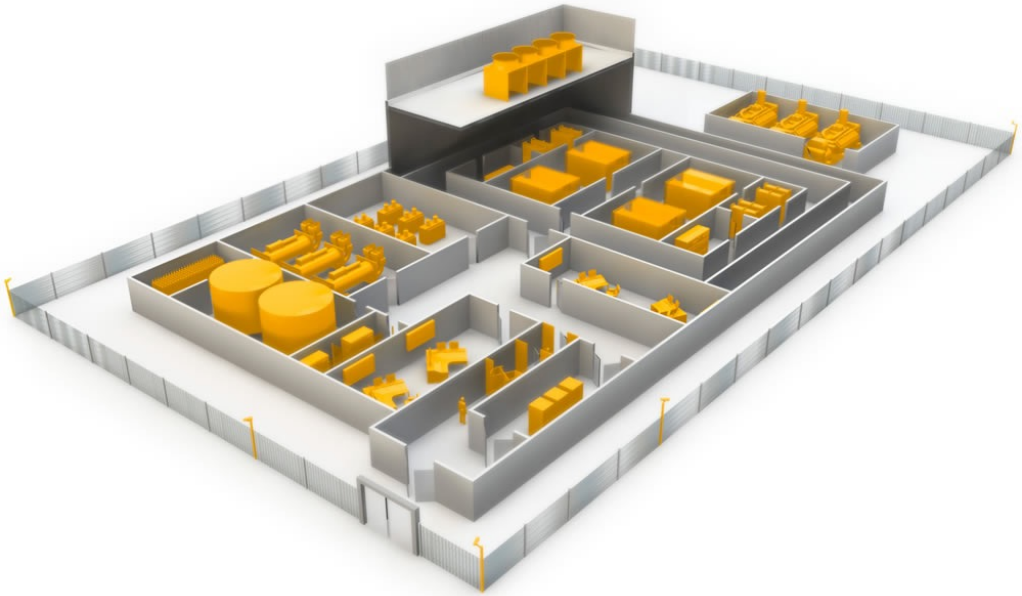


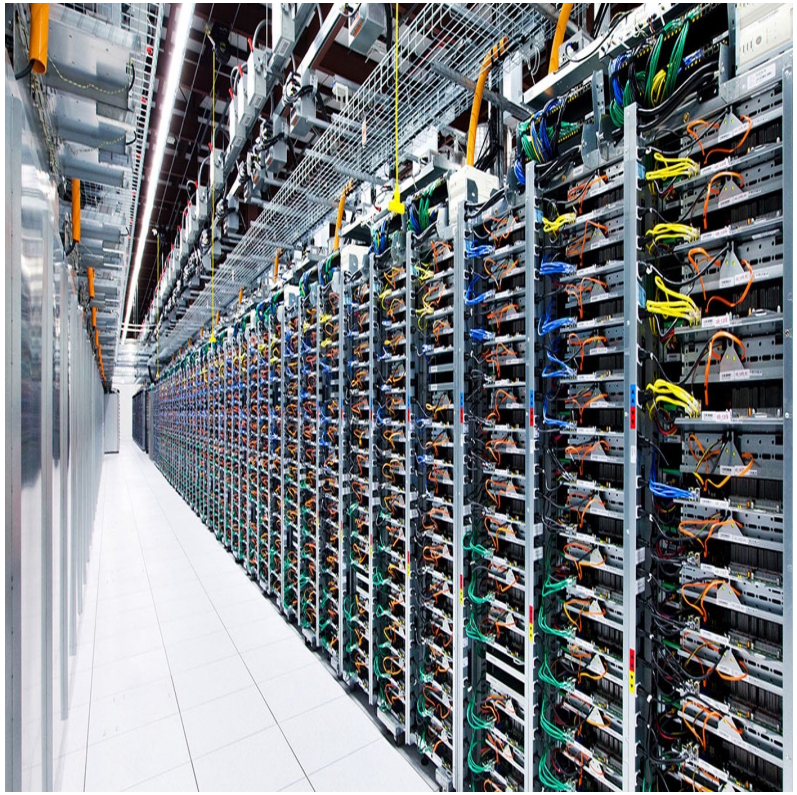
The *massive scale* of the software infrastructure, data repositories, and hardware platform

- The **program in warehouse-scale computing**:
  - ✓ is an internet service,
  - ✓ may consist of tens or more individual programs
  - ✓ such programs interact to implement complex end-user services such as email, search, maps or machine learning

Warehouse-scale computers vs. DATACENTERS (1)

- Data centers are **buildings where multiple servers and communication units are co-located** because of their common environmental requirements and physical security needs, and for ease of maintenance





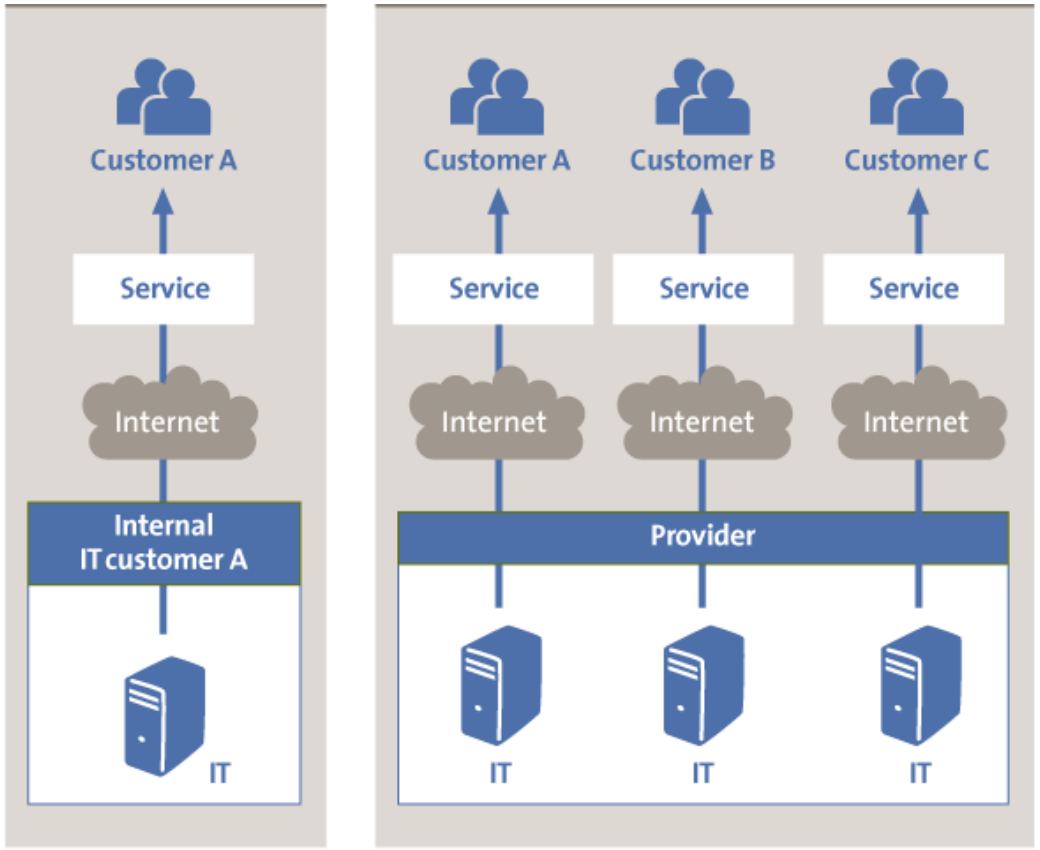
Warehouse-scale computers vs. DATACENTERS (2)

- **Traditional data centers:**
  - typically host a large number of relatively small- or medium-sized applications
  - each applications is running on a dedicated hardware infrastructure that is de-coupled and protected from other systems in the same facility
  - applications tend not to communicate each other

- Those data centers host hardware and software for **multiple organizational units or even different companies**

Warehouse-scale computers vs. DATACENTERS (3)

Traditional Datacenters



WAREHOUSE-SCALE COMPUTERS vs. Datacenters (4)

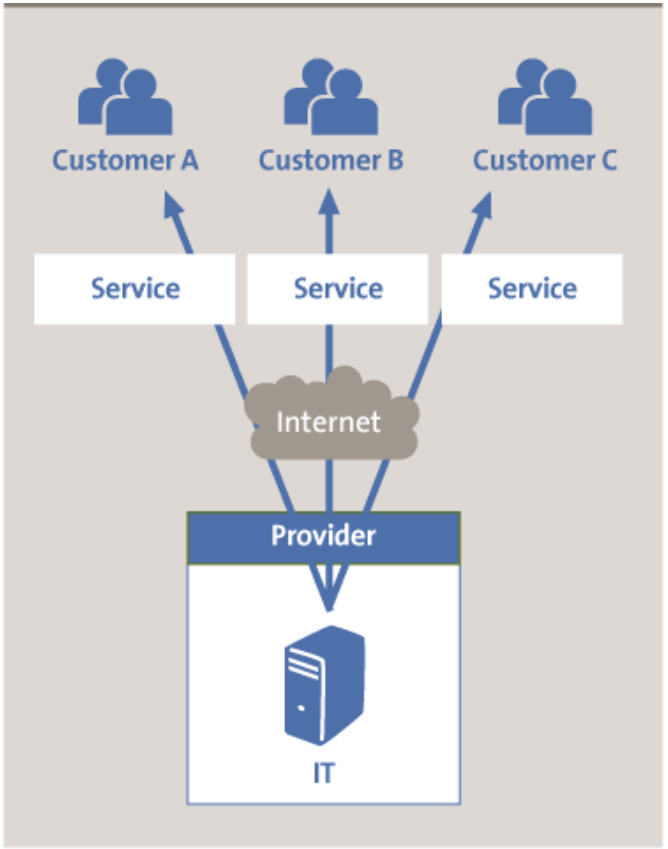
WSCs belong to a single organization, use a relatively homogeneous hardware and system software platform, and share a common systems management layer

WAREHOUSE-SCALE COMPUTERS vs. Datacenters (5)

- WSCs run a smaller number of very large applications (or internet services)
- The common resource management infrastructure allows significant deployment flexibility
- The requirements of
  - homogeneity
  - single-organization control
  - cost efficiency

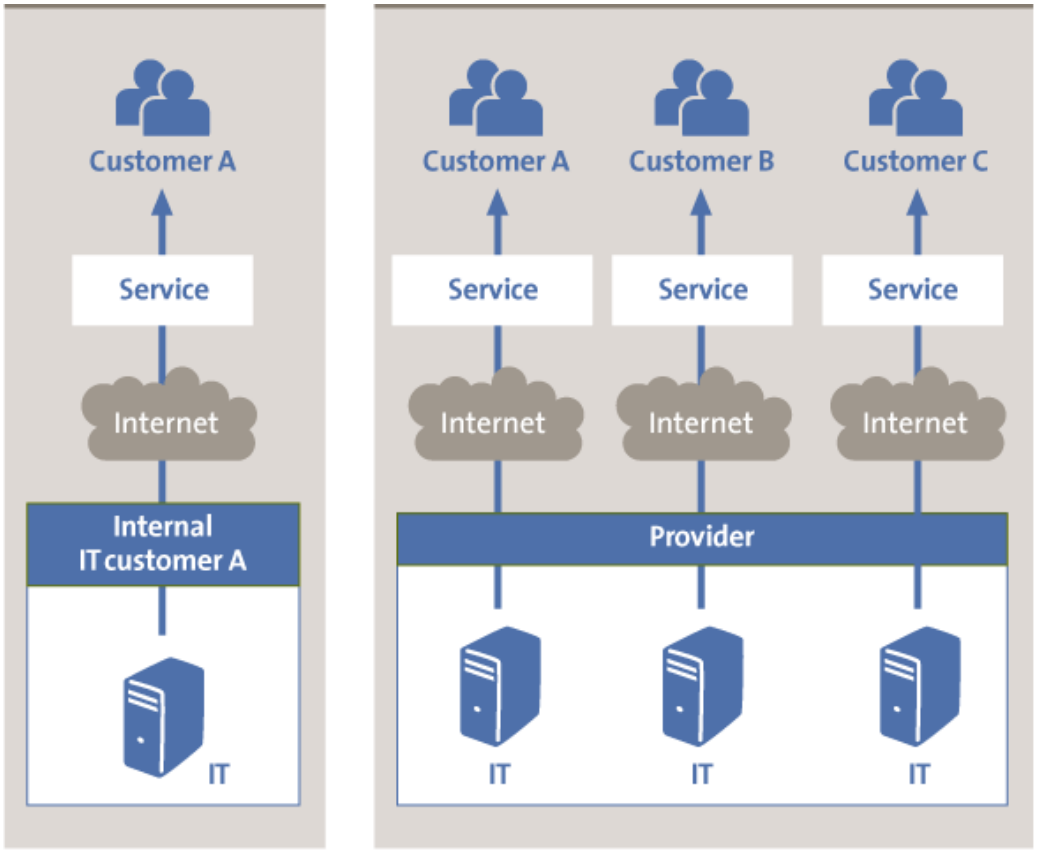
  motivate designers to take new approaches in designing WSCs

Warehouse-Scale Computer

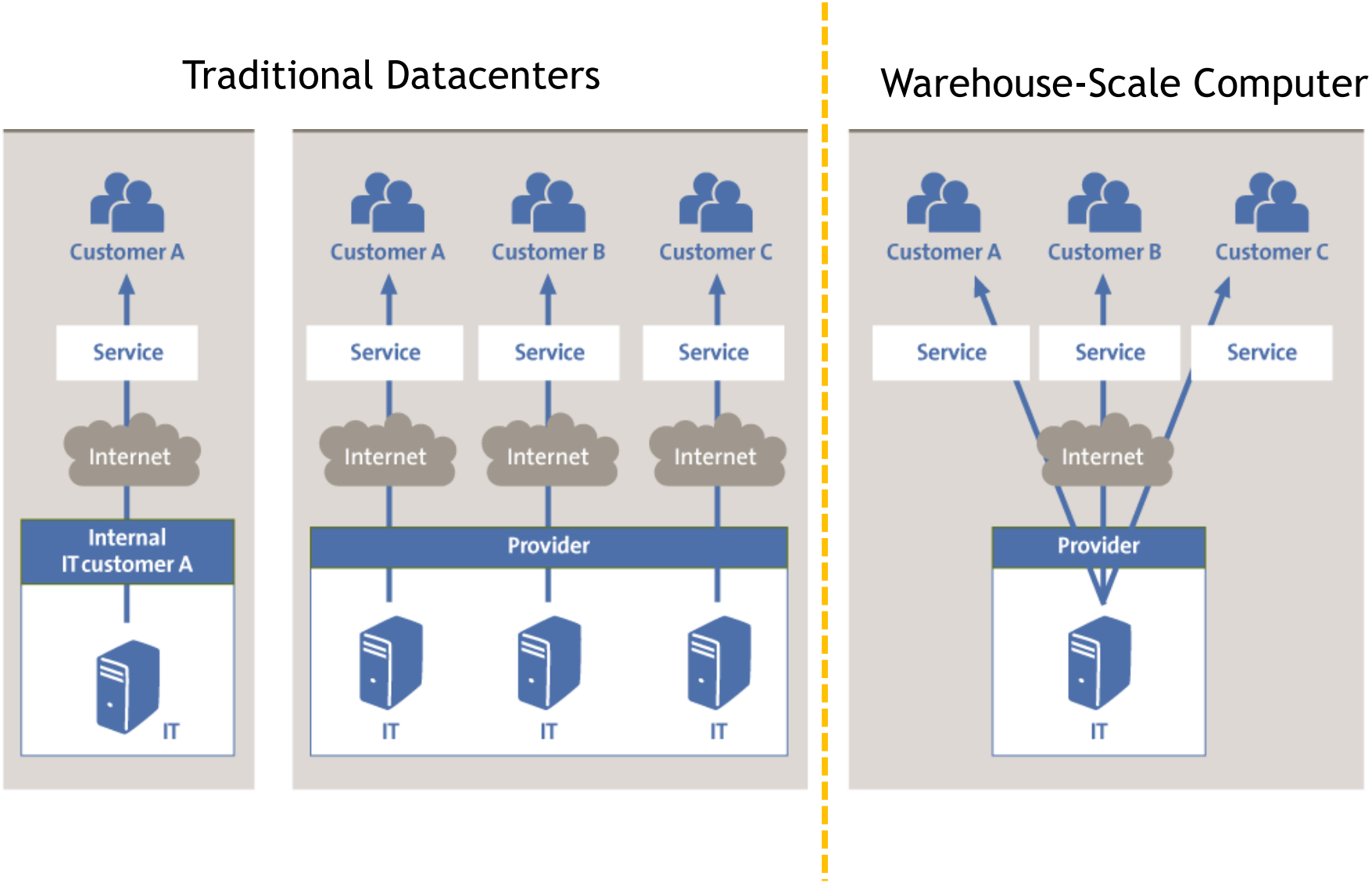


WAREHOUSE-SCALE COMPUTERS vs. Datacenters: a graphical comparison

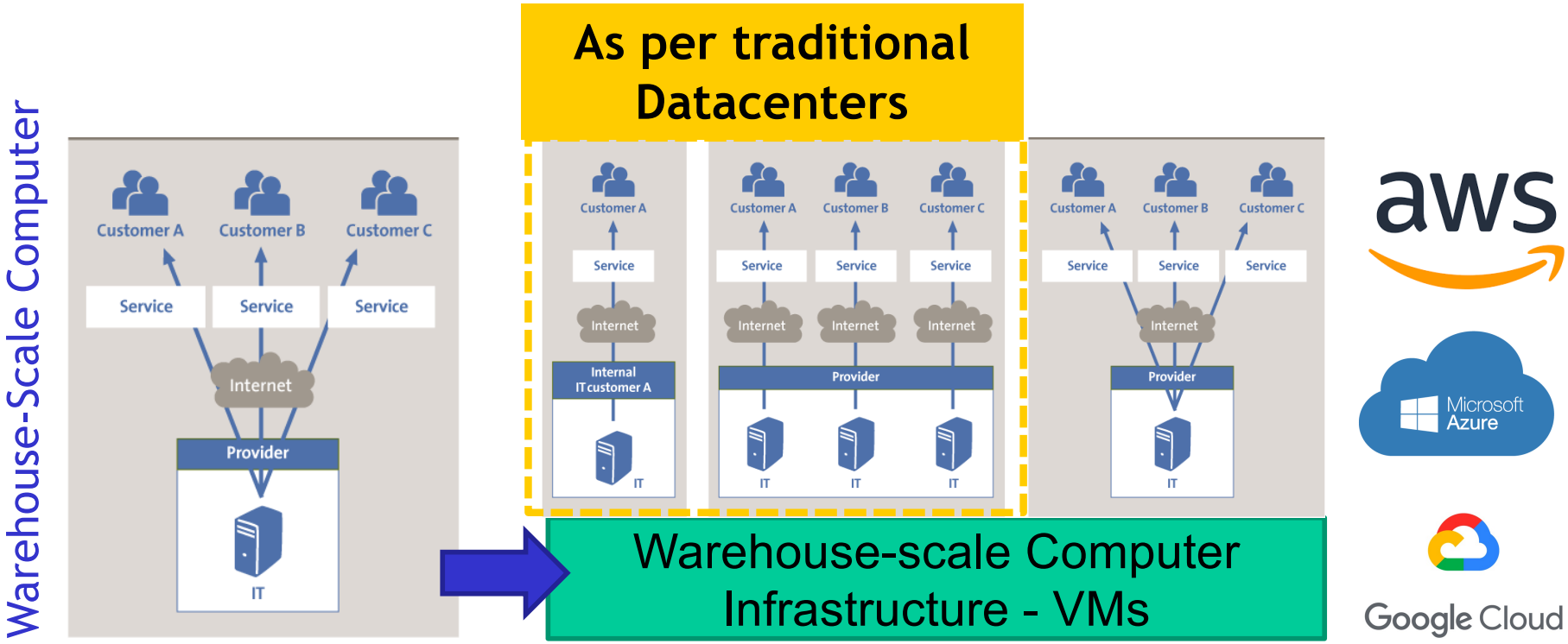
Traditional Datacenters



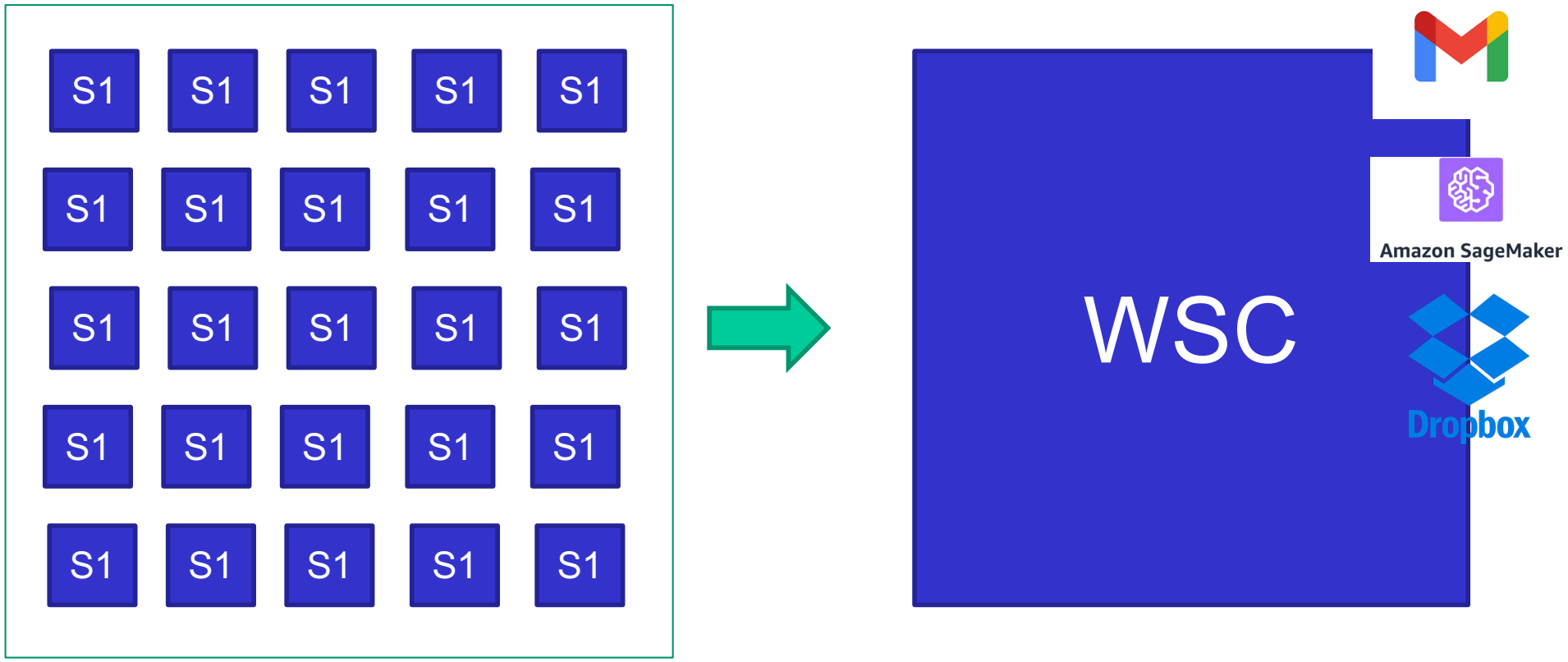
WAREHOUSE-SCALE COMPUTERS vs. Datacenters: a graphical comparison



From Datacenter to WSCs (and back) …



WSCs: not just a collection of servers



- The software running on these systems executes on clusters of hundreds to thousands of individual servers (far beyond a single machine or a single rack)
- The machine is itself this large cluster or aggregation of servers and needs to be considered as a single computing unit

What about several datacenters?

- One data center vs. multiple data centers located far apart
- Multiple data centers are (often) replicas of the same service:
  - ✓ to reduce user latency
  - ✓ improve serving throughput
- A request is typically fully processed within one data center

https://www.datacenterknowledge.com/sites/datacenterknowledge.com/files/wp-content/uploads/2016/09/aws-azure-dc-map.png

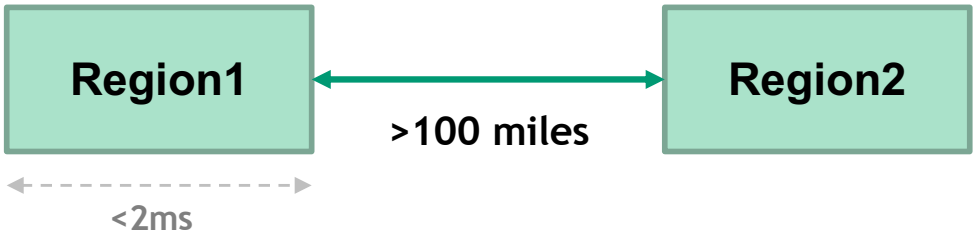
Hierarchical approach: Geographic Areas and Regions

The world is divided into **Geographic Areas (GAs)**

- Defined by Geo-political boundaries (or country borders)
- Determined mainly by data residency
- In each GA there are at least 2 computing regions

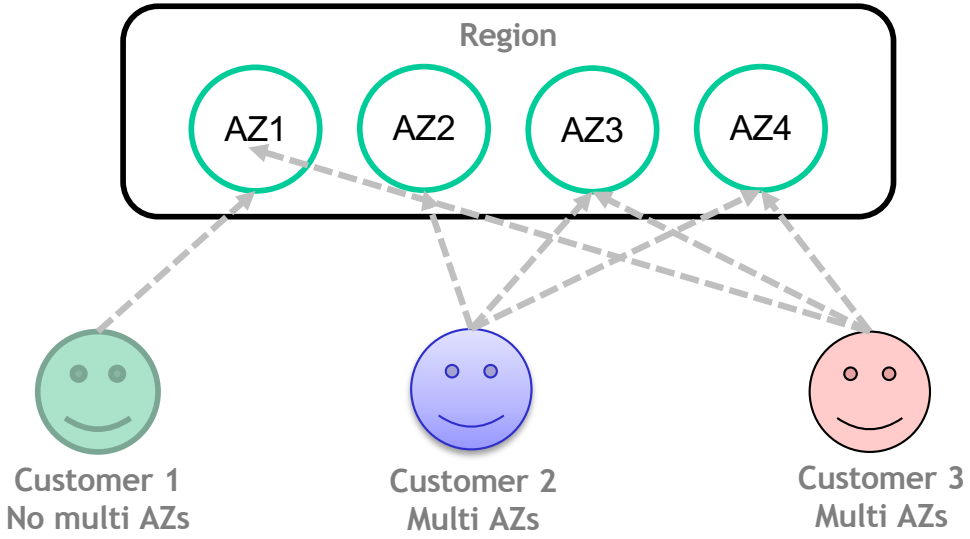
Computing **Regions (CRs)**:

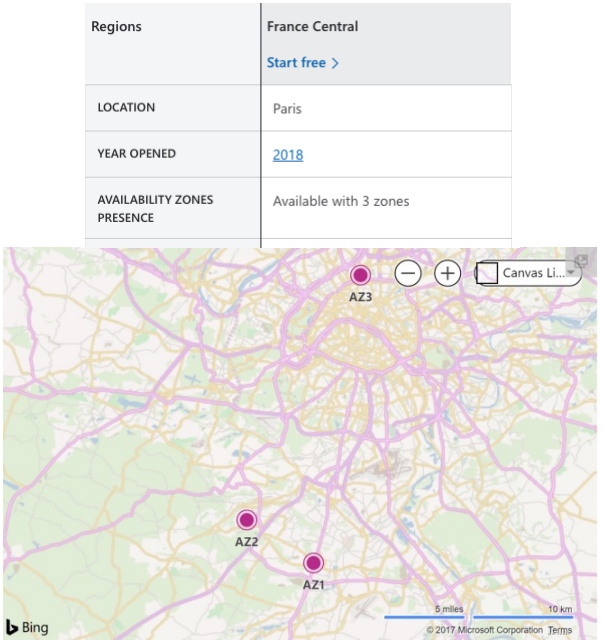
- Customers see regions as the finer grain discretization of the infrastructure
  - Multiple DCs in the same region are not exposed
- Latency-defined perimeter (2ms latency for the round trip)
- 100's of miles apart, with different flood zones etc...
- Too far for synchronous replication, but ok for disaster recovery



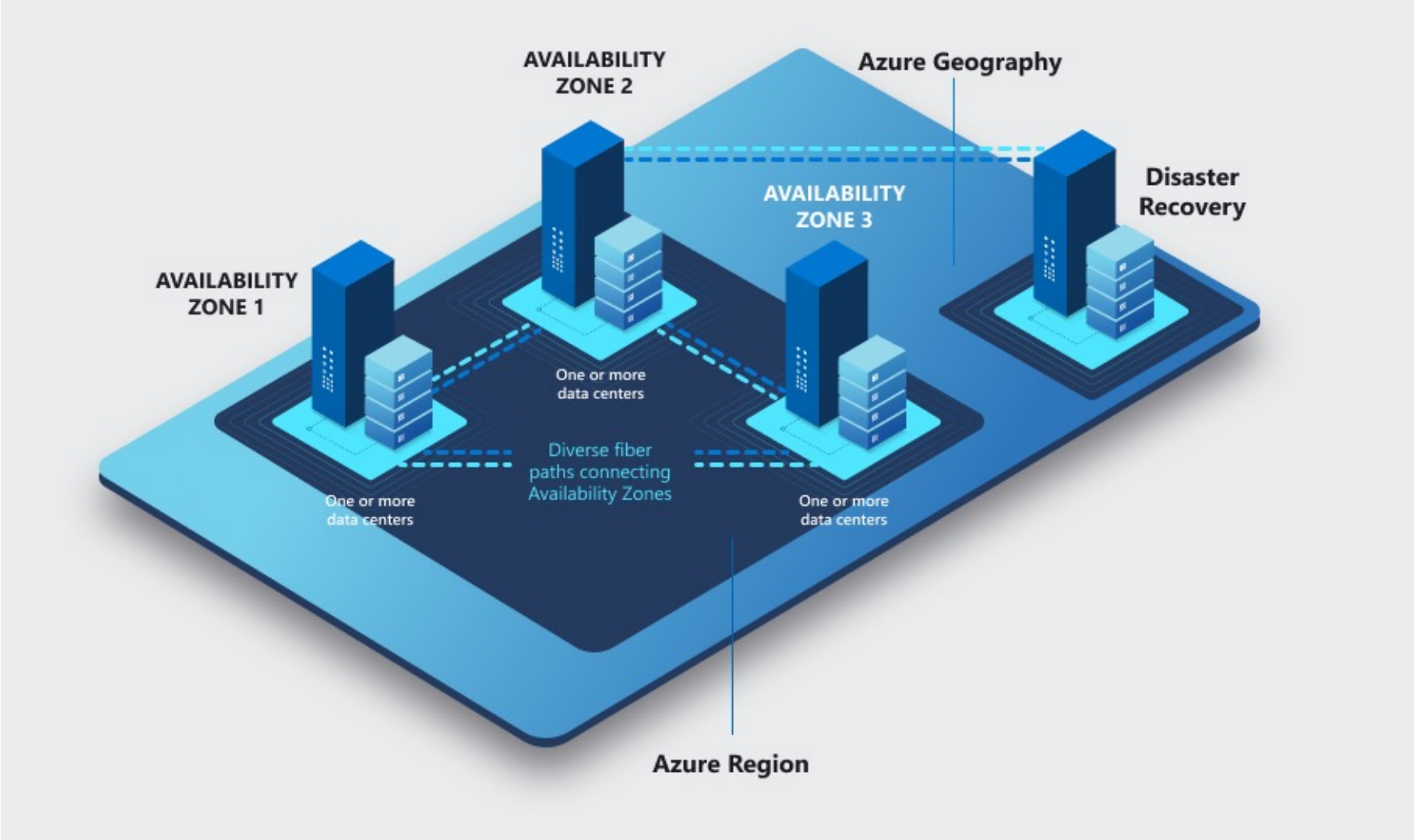
Hierarchical approach: Availability Zones

- **Availability Zones (AZs)** are finer grain location within a single computing region
  - allow customers to run mission critical applications with high availability and fault tolerance to datacenter failures
    - Fault-isolated locations with redundant power, cooling, and networking
  - Application-level synchronous replication among AZs
  - 3 is minimum and enough for quorum



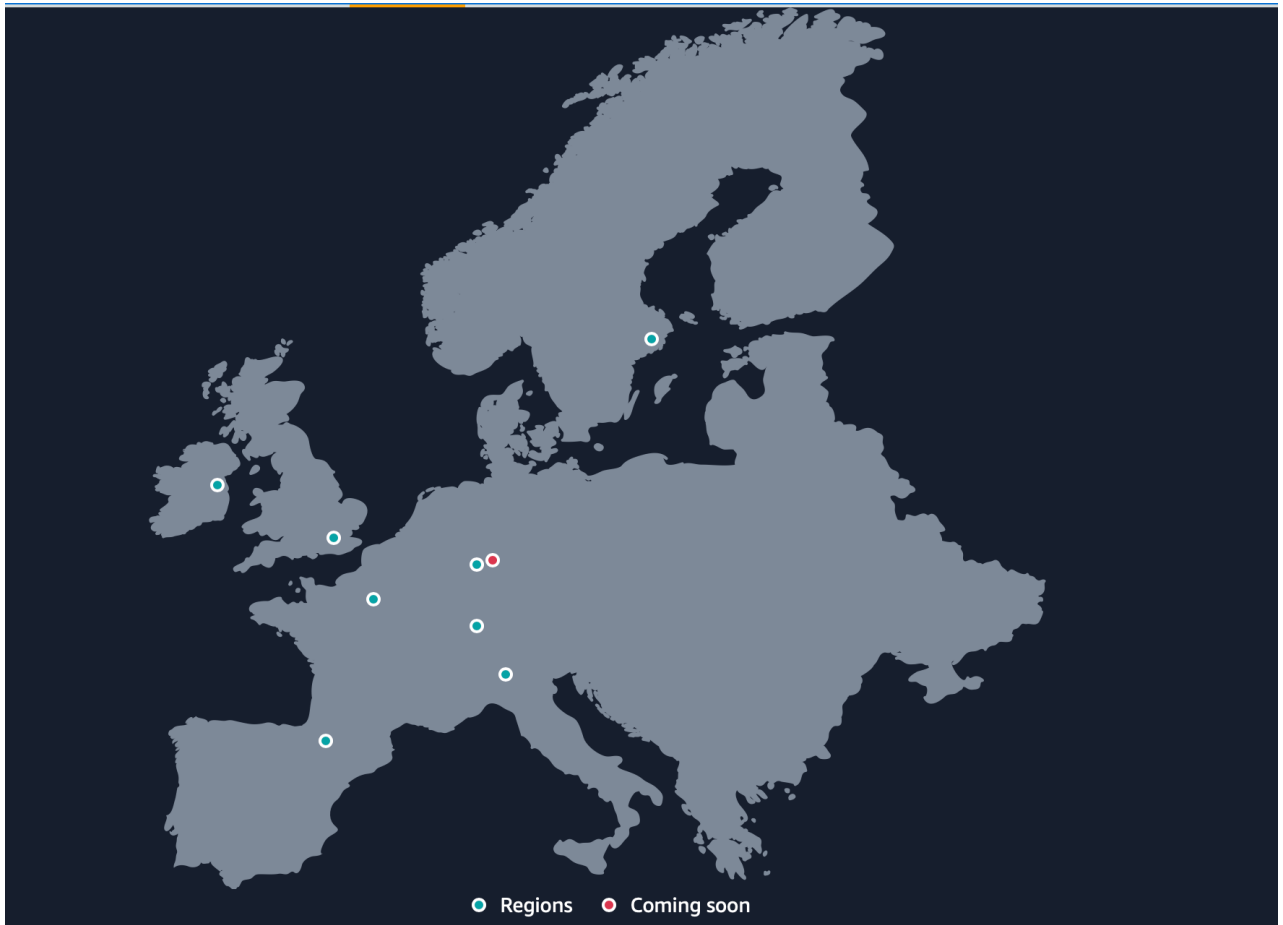


Overview: Azure Example



https://docs.microsoft.com/en-us/azure/availability-zones/az-overview

AWS Data centers in Europe



Europe (Ireland) Region
Availability Zones: 3
Launched 2007

Europe (Frankfurt) Region
Availability Zones: 3
Launched 2014

Europe (London) Region
Availability Zones: 3
Launched 2016

Europe (Paris) Region
Availability Zones: 3
Launched 2017

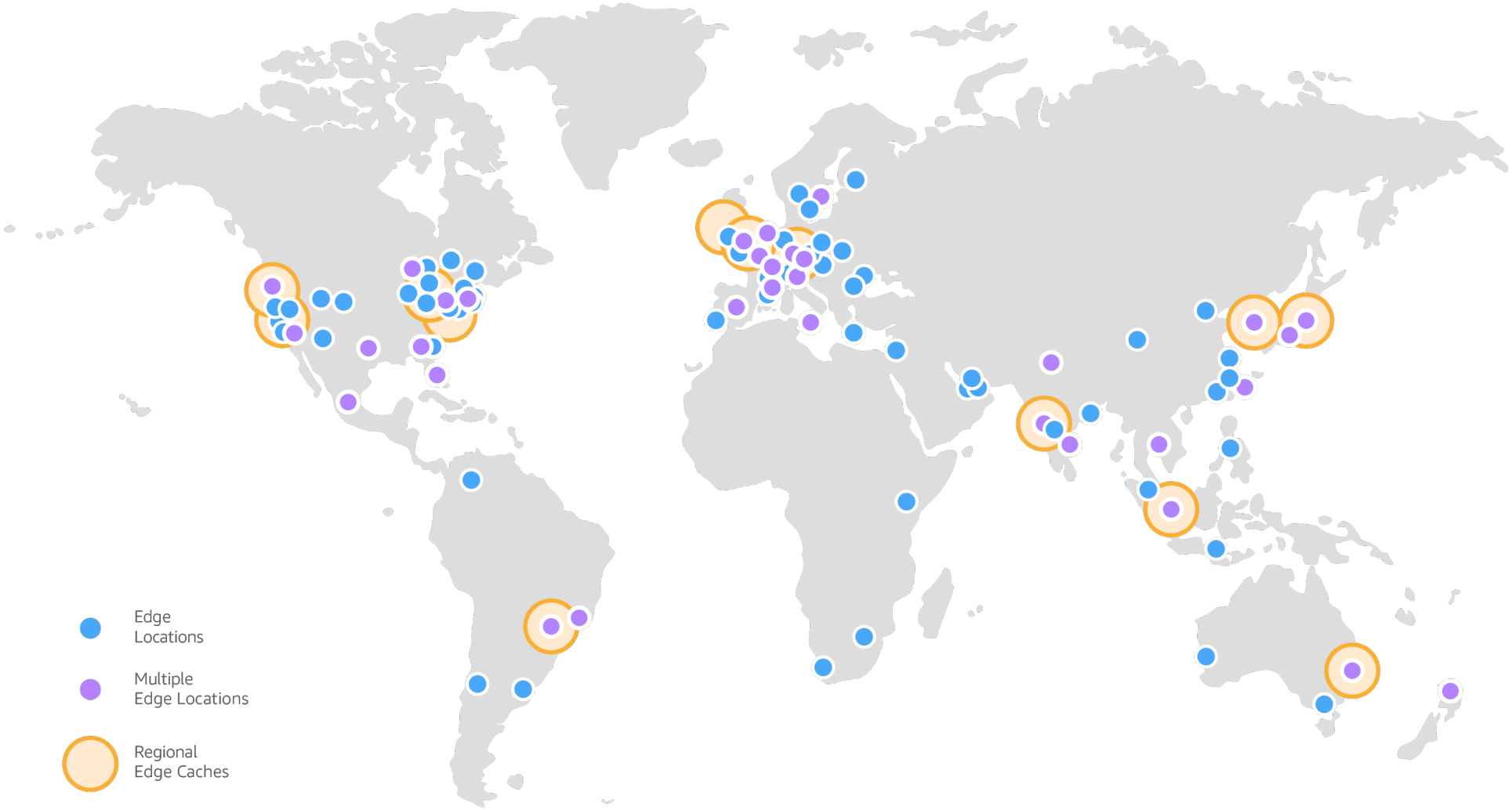
Europe (Stockholm) Region
Availability Zones: 3
Launched 2018

Europe (Milan) Region
Availability Zones: 3
Launched 2020

Europe (Zurich) Region
Availability Zones: 3
Launched 2022

Europe (Spain) Region
Availability Zones: 3
Launched 2022

Edge Locations - AWS



WSCs and availability

- Services provided through WSCs must guarantee high availability, typically aiming for at least 99.99% uptime (i.e., one-hour downtime per year)
- Achieving such fault-free operation is difficult when a large collection of hardware and system software is involved
- WSC workloads must be designed to gracefully tolerate large numbers of component faults with little or no impact on service level performance and availability!!!

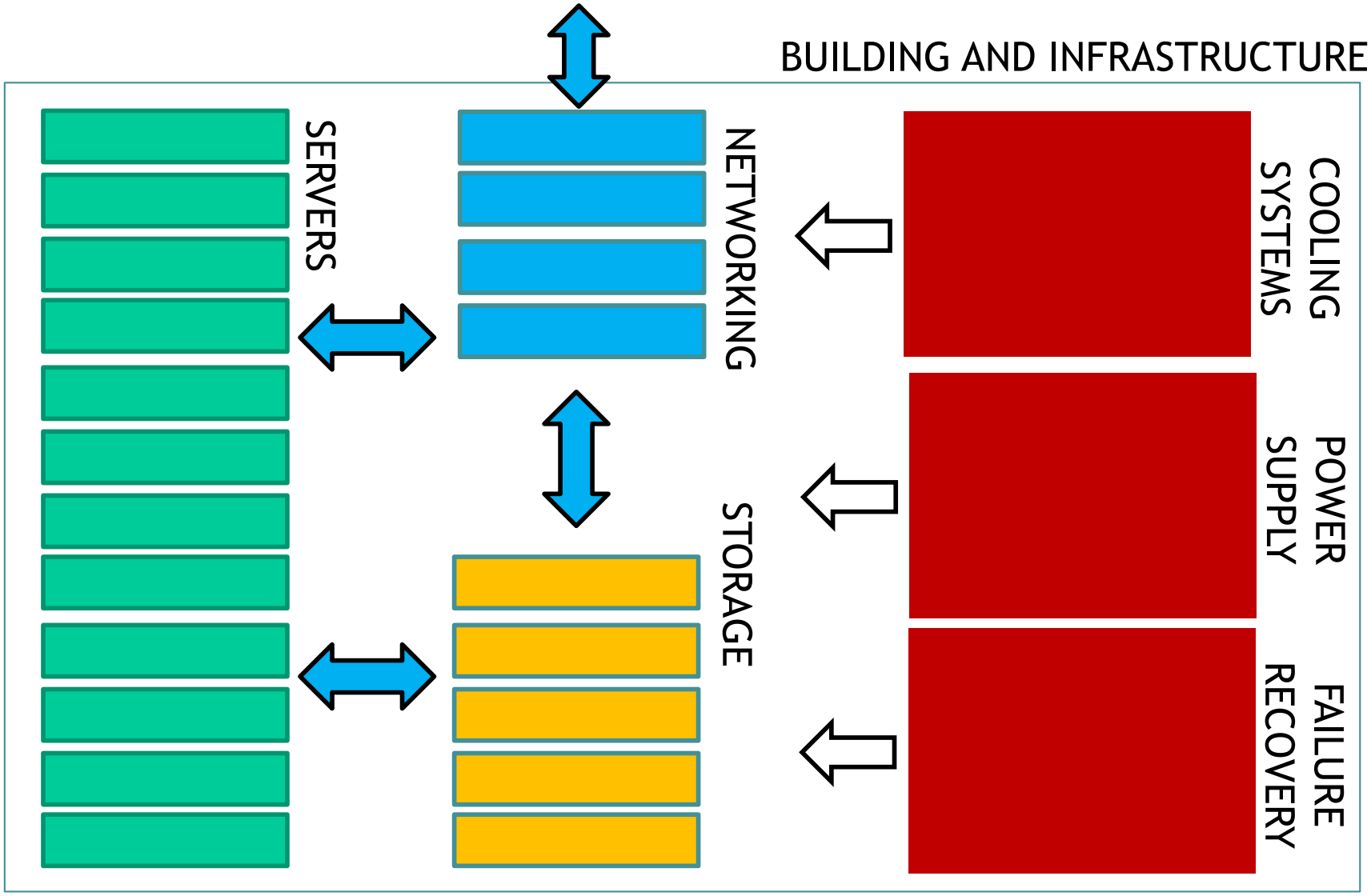
This is exactly the goal of the «Dependability» part of this course

ARCHITECTURAL OVERVIEW OF WAREHOUSE-SCALE COMPUTERS

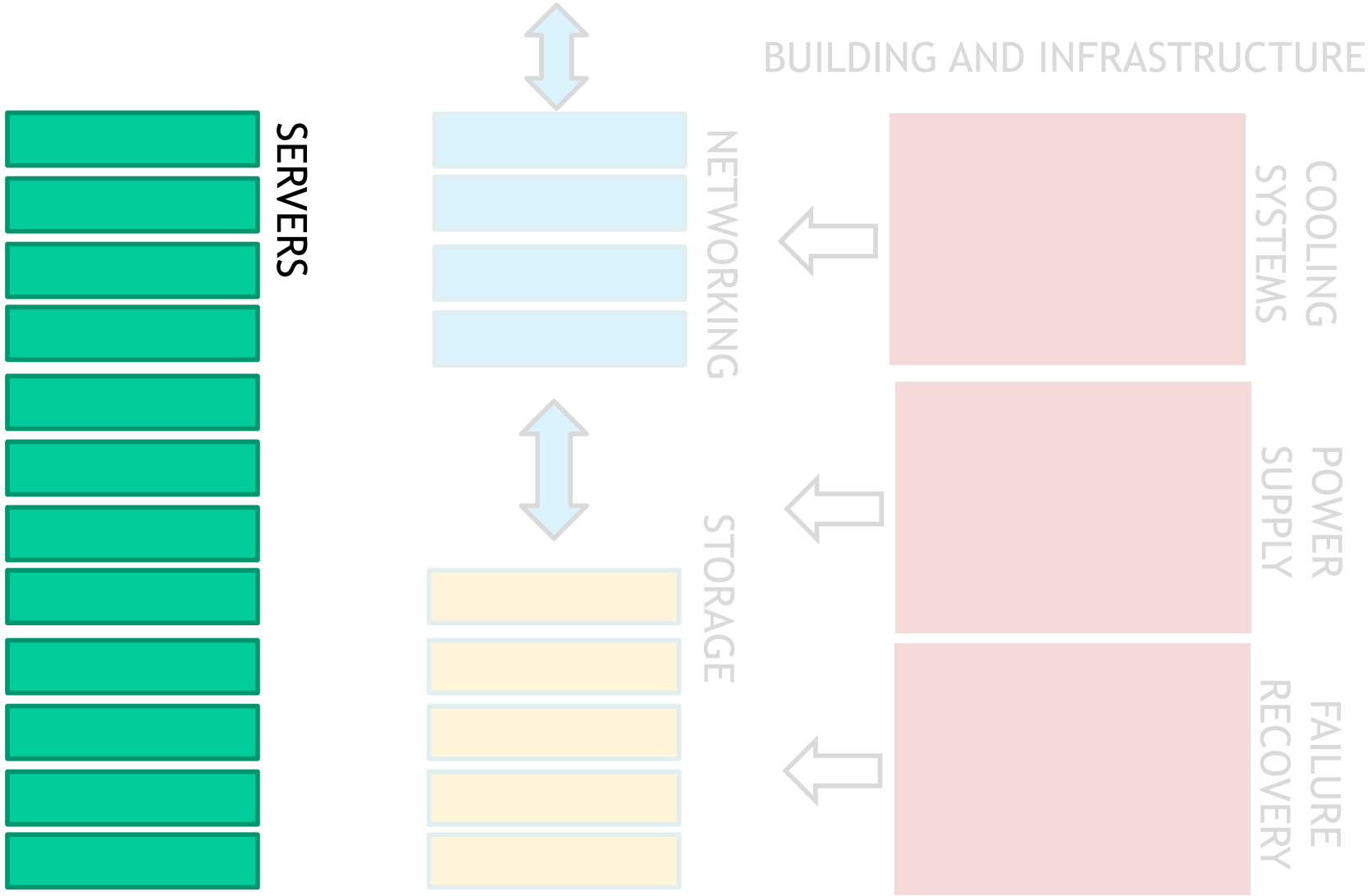
Hardware implementation of WSCs might differ significantly each other

However, the architectural organization of these systems is relatively stable

Architectural Overview of A Warehouse-scale Computer



SERVERS


SERVERS
NETWORKING
STORAGE
BUILDING AND INFRASTRUCTURE
COOLING SYSTEMS
POWER SUPPLY
FAILURE RECOVERY

SERVERS: the main processing equipment

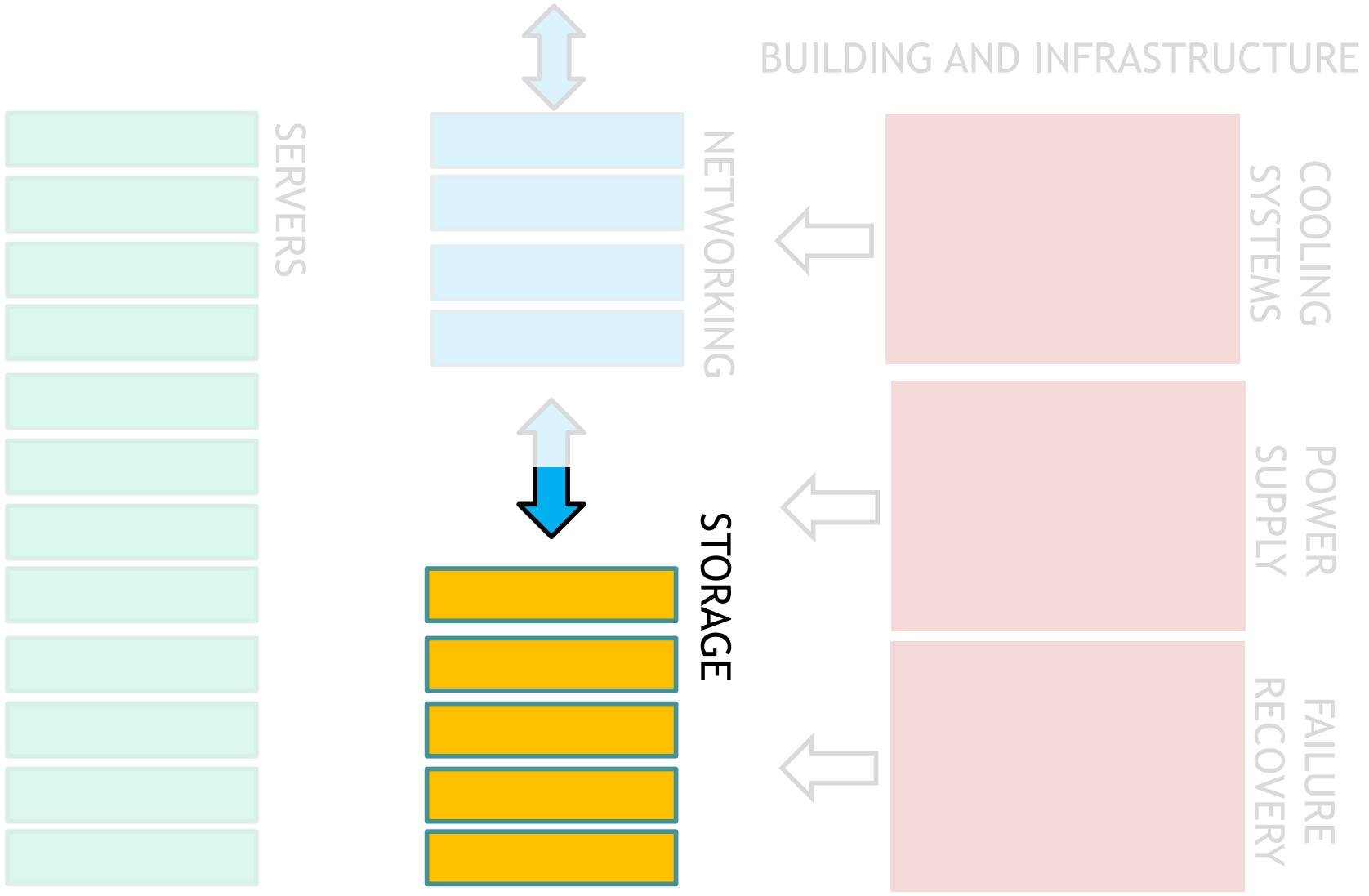
They are like ordinary PC, but with a form factor that allows to fit them into the racks:

- Rack (1U or more)
- Blade enclosure format
- Tower

They may differ in:

- Number and type of CPUs
- Available RAM
- Locally attached disks (HDD, SSD or not installed)
- Other special purpose devices (like GPUs, DSPs and coprocessors)

STORAGE


BUILDING AND INFRASTRUCTURE
SERVERS
NETWORKING
STORAGE
COOLING SYSTEMS
POWER SUPPLY
FAILURE RECOVERY

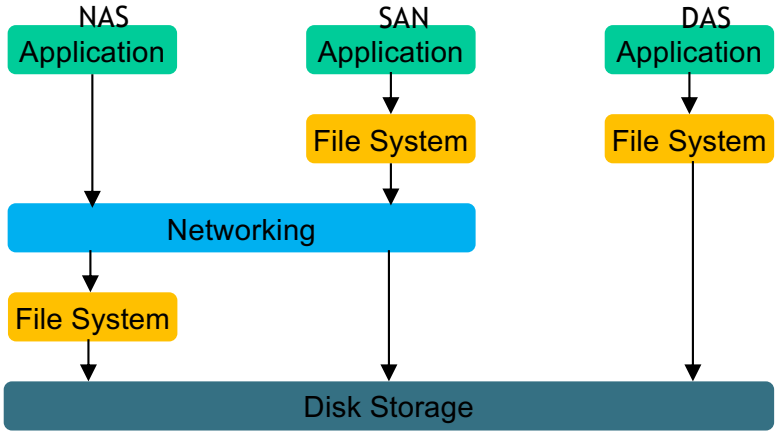
STORAGE: how and where to store the information

Disks and Flash SSDs are the building blocks of today's WSC storage systems.

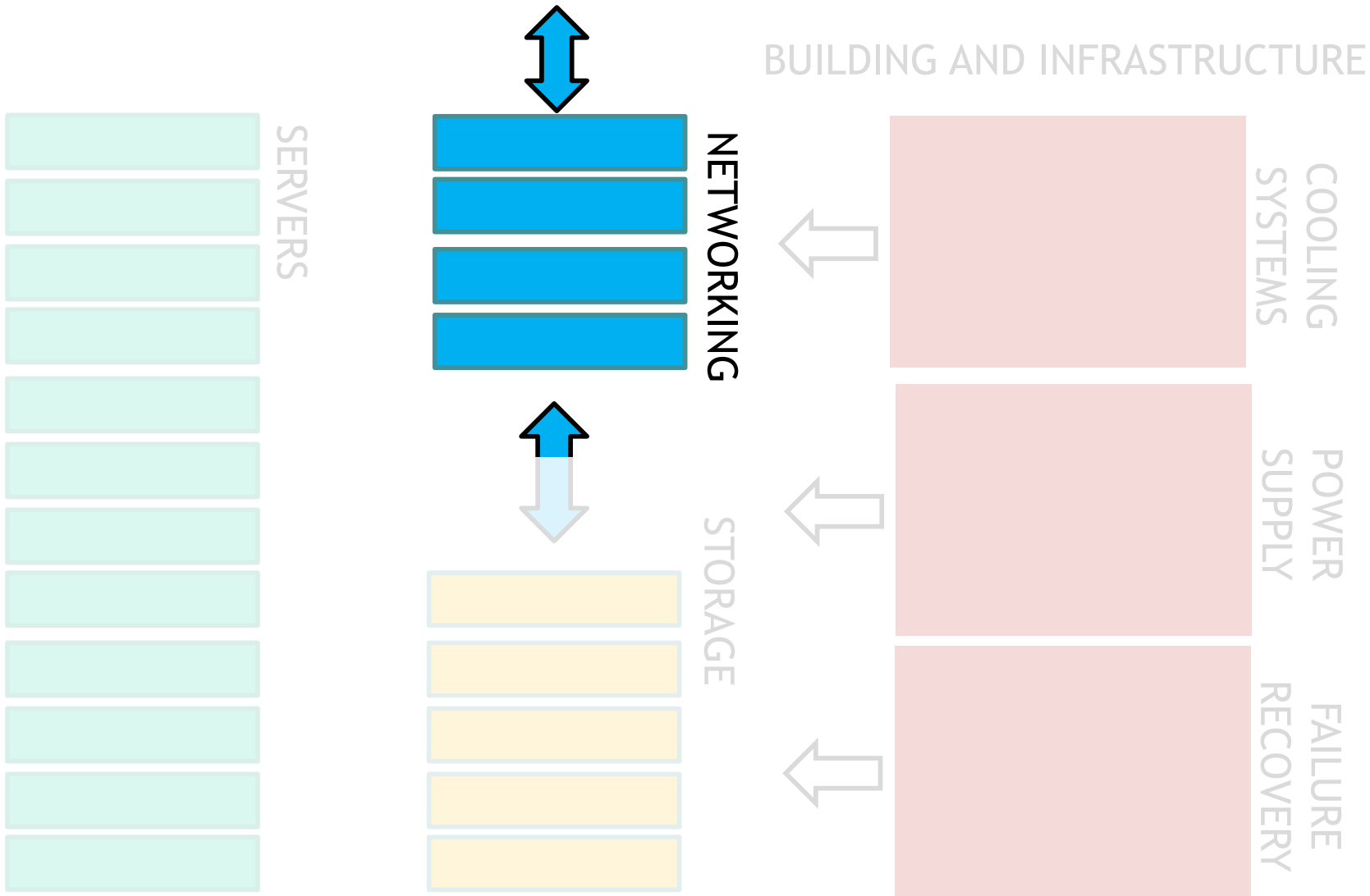
These devices are connected to the data-center network and managed by sophisticated distributed systems

Examples:

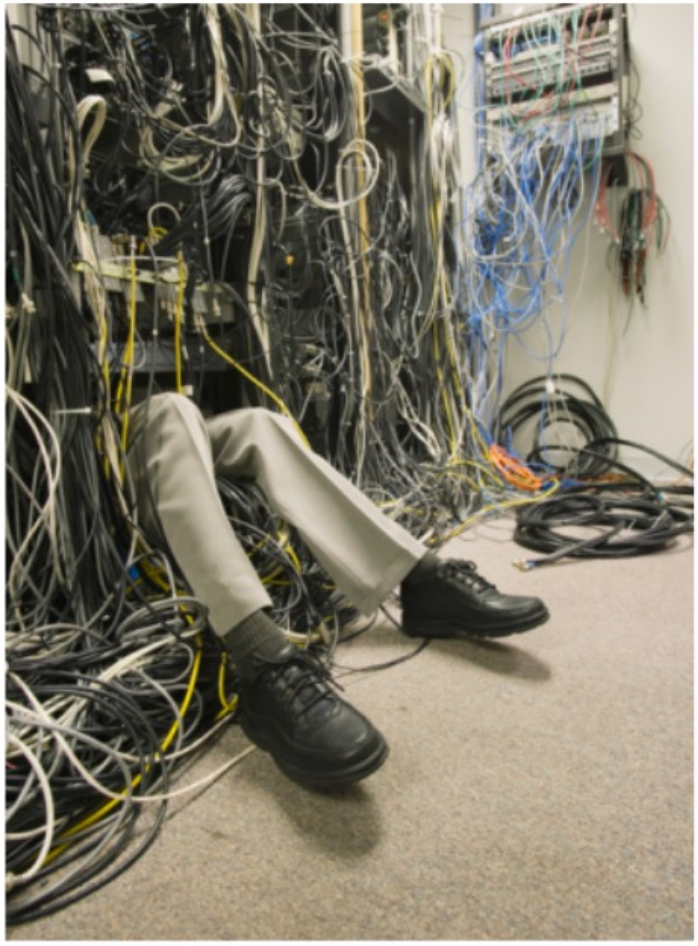
- Direct Attached Storage DAS
- Network Attached Storage (NAS)
- Storage Area Networks (SAN)
- RAID controllers



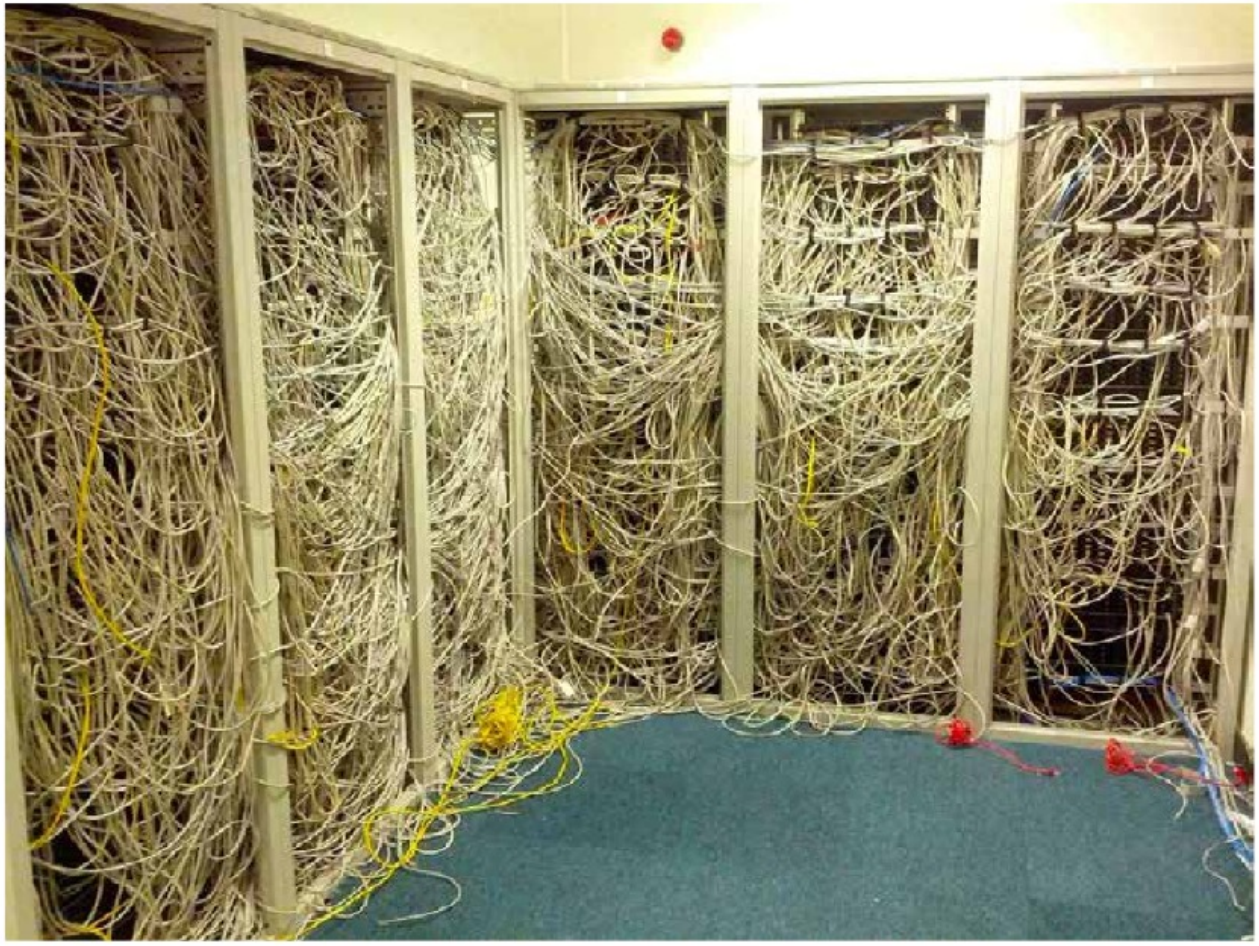
NETWORKING


BUILDING AND INFRASTRUCTURE
SERVERS
NETWORKING
STORAGE
COOLING SYSTEMS
POWER SUPPLY
FAILURE RECOVERY

NETWORKING



[David Samuel Robbins, gettyimages.ch]



[@AlexCWheeler, Twitter]

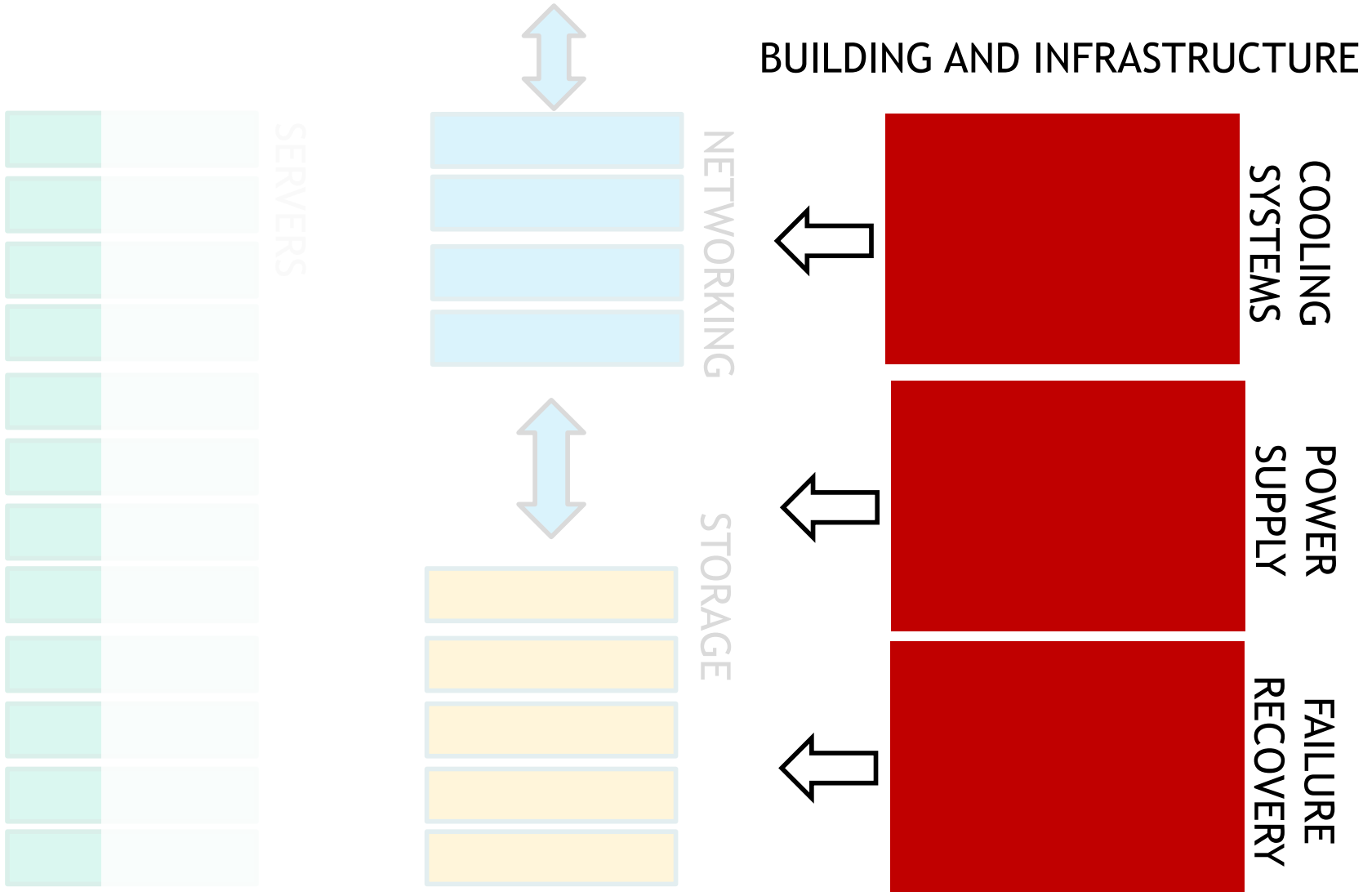
NETWORKING: providing internal and external connections

Communication equipment allows network interconnections among the devices.

They can be:

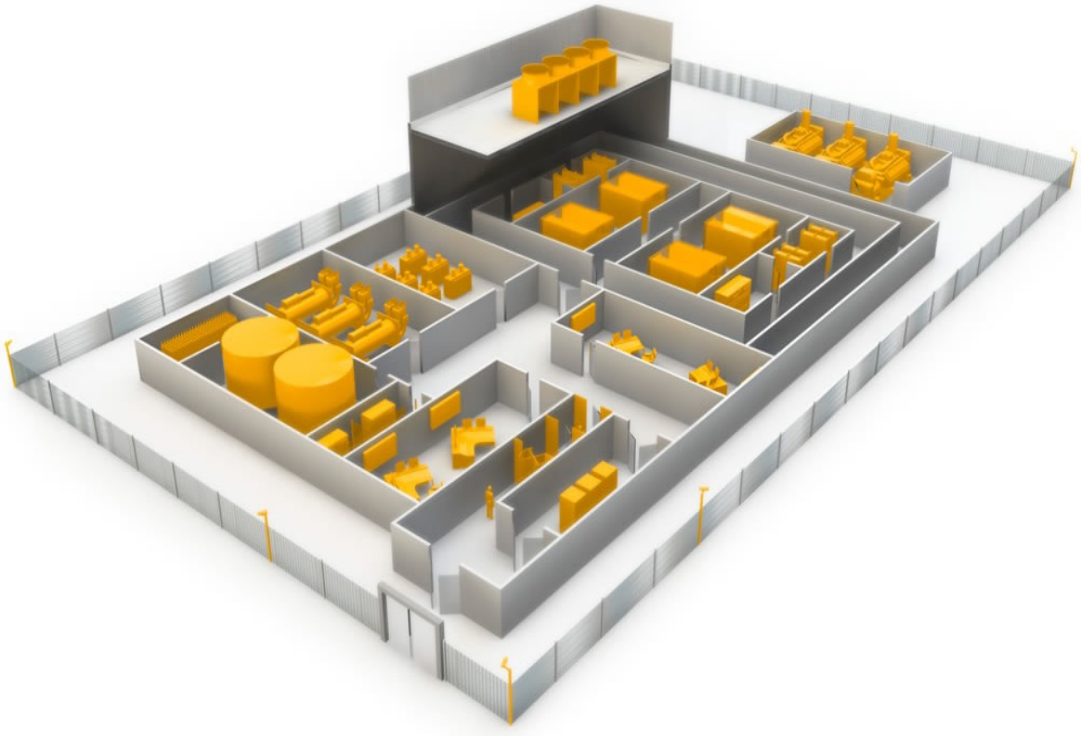
- Hubs
- Routers
- DNS or DHCP servers
- Load balancers
- Switches
- Firewalls
- ... and many more other type of devices!

BUILDING AND INFRASTRUCTURE



BUILDING AND INFRASTRUCTURE

WSC has other important components related to **power delivery**, **cooling**, and **building infrastructure** that also need to be considered



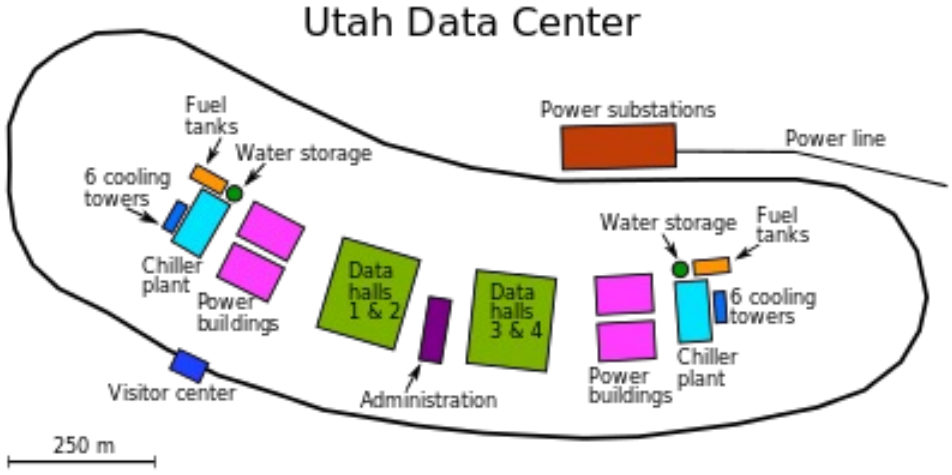
Some interesting numbers:

- ✓ **Datacenters with up to 110 football-pitch size**
- ✓ **150 MW power consumption (100K houses)**
- ✓ **99.99% uptime, i.e., one-hour downtime per year**

The need for a comprehensive design of computation, storage, networking and building infrastructure

NSA's Utah Data Center

Over a million square feet and 65MW



https://www.switch.com/las-vegas/

LAS VEGAS EXASCALE DATA CENTER ECOSYSTEMS AT UNPARALLELED LOW PRICING

The Core Campus located in Las Vegas, Nevada, will have up to 531 MW of power upon completion. Switch's Tier 5® Platinum exascale data center facilities make Switch the highest-rated and most cost-effective colocation environment in the industry.

100% GREEN POWER

- 100% green power
- 100% power uptime guarantee
- Up to 55kW per cabinet

LOW OR NO TAXES

- 2% sales and use tax in Nevada
- 75% reduction in personal property tax in Nevada
- No personal state income tax in Nevada

35-60% SAVINGS ON CONNECTIVITY

- 5ms to Los Angeles
- Dedicated 7ms fiber to Switch's Citadel Campus via the Switch SUPERLOOP®
- Services include: MPLS, IP, Transport, SIP and SD-WAN (Circuits do not have to terminate at Switch)
- 35-60% on connectivity savings through Switch CONNECT® telecom auditing and expense management services
- Exascale telecom purchasing is enabled by leveraging the multi-trillion dollar market cap of co-op members through The CORE Cooperative