UNIVERSIDADE
BEIRA INTERIOR

Data Center Design and Interconnection Networks

Tecnologias de Virtualização e Centros de Dados

Mestrado em Engenharia Informática

Mário M. Freire
Departamento de Informática

Ano Letivo 2023/2024

Credits

- These slides are partly based on the book:

  Distributed and Cloud Computing: From Parallel Processing to the Internet of Things, Kai Hwang, Jack Dongarra, Geoffrey C. Fox (Authors), Morgan Kaufmann, 1st edition, 2011, ISBN-13: 978-0123858801, 672 pages.

Agenda

- Warehouse-Scale Data Center Design
- Data-Center Interconnection Networks
- Modular Data Center in Shipping Containers
- Interconnection of Modular Data Centers
- Data-Center Management Issues

Warehouse-Scale Data Center Design

- Most **data centers** are **built** with **commercially** available **components**.
- An **off-the-shelf server** consists of a number of **processor sockets**, each with a **multicore CPU** and its **internal cache hierarchy**, local shared and coherent **DRAM**, and a number of directly attached **disk drives**.
- The **DRAM** and **disk resources** within the rack are accessible through **first-level rack switches** and all resources in all **Racks** are accessible via a **cluster-level switch**.
- Consider a **datacenter** built with **2,000 servers**, each with **8-32 GB** of **DRAM** and four **1-4 TB** disk drives. Each group of **40 servers** is connected through a **1-10 Gbps link** to a rack-level switch that has an additional eight **1-10 Gbps** ports used for connecting the rack to the cluster-level switch.

Warehouse-Scale Data Center Design

Architecture of a small server cluster (~ 1000 servers), interconnected by an Ethernet switch and housed in a warehouse or in a container environment

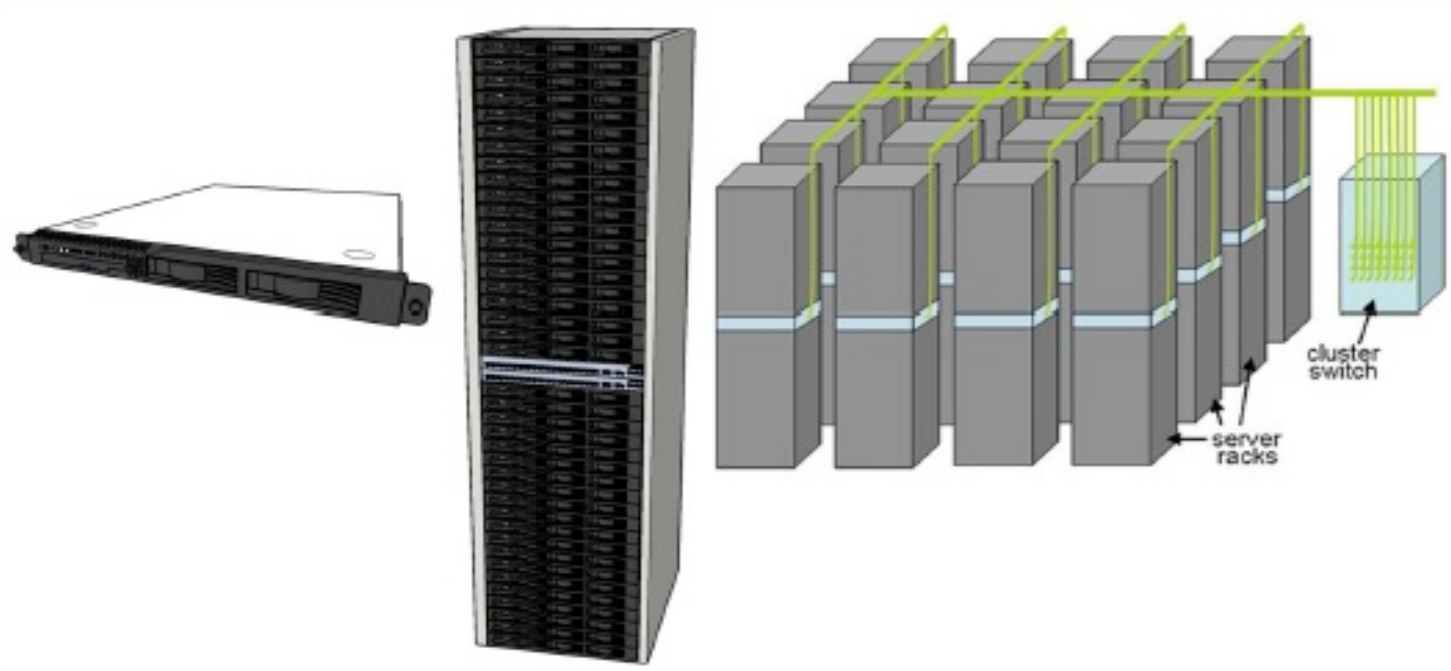


FIGURE 1.1: Typical elements in warehouse-scale systems: 1U server (left), 7′ rack with Ethernet switch (middle), and diagram of a small cluster with a cluster-level Ethernet switch/router (right).

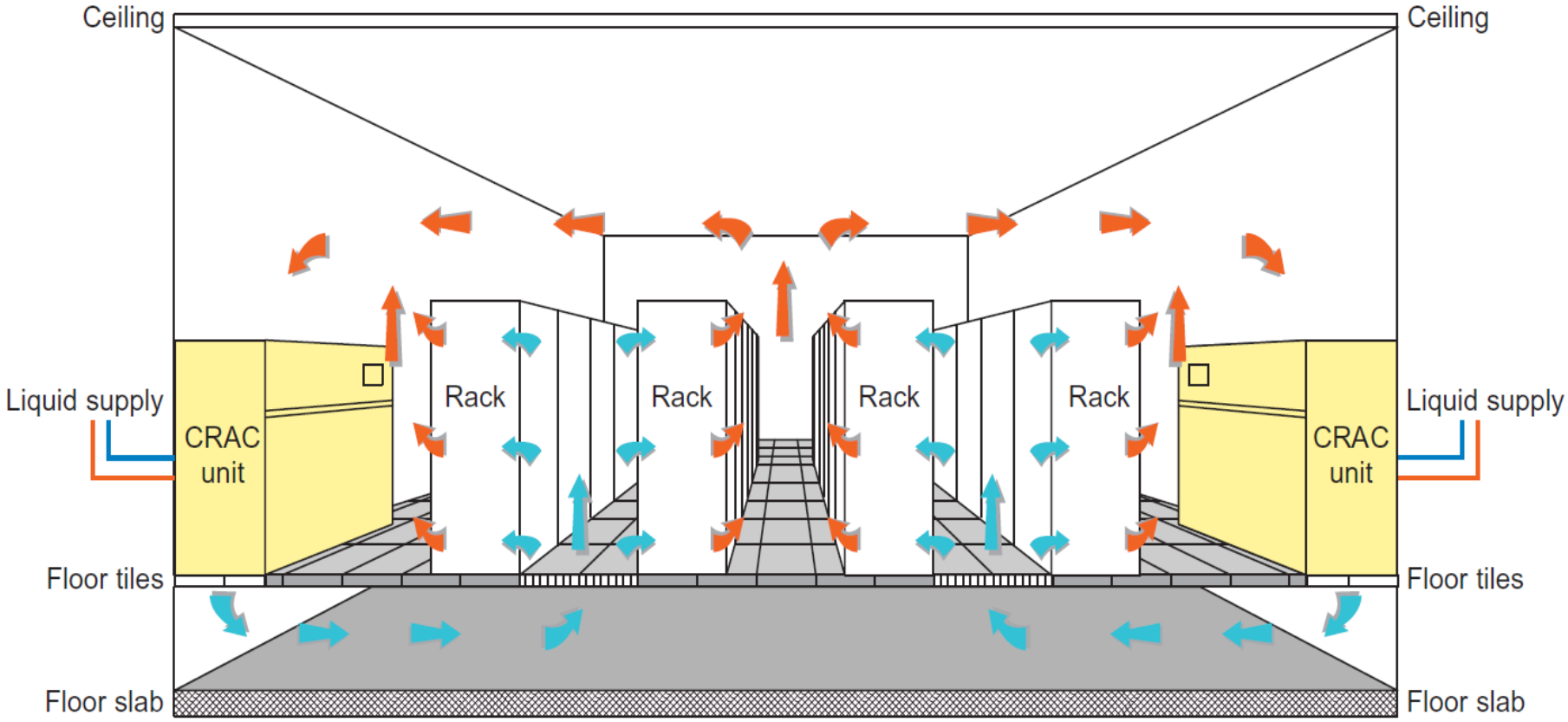
In: Distributed and Cloud Computing: From Parallel Processing to the Internet of Things, Kai Hwang, Jack Dongarra, Geoffrey C. Fox, Morgan Kaufmann, 1st edition, 2011, ISBN-13: 978-0123858801, 672 pages.

Warehouse-Scale Data Center Design

Video Ilustrativo organização e acessórios de um Rack:
https://www.youtube.com/watch?v=aniem53phRU&t=82s

Warehouse-Scale Data Center Design

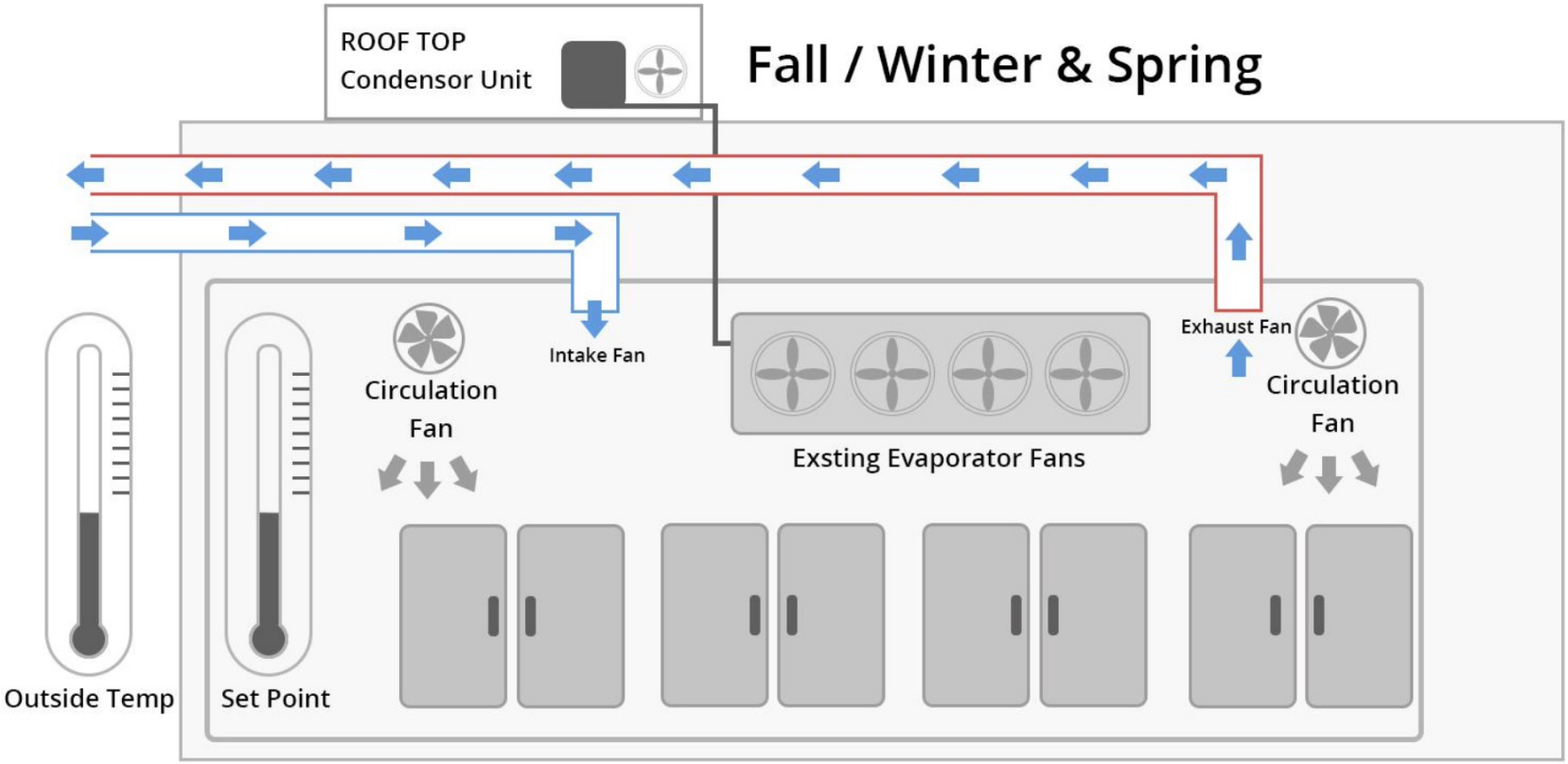
The cooling system in a raised-floor data center with hot-cold air circulation supporting water heat exchange facilities



In: Distributed and Cloud Computing: From Parallel Processing to the Internet of Things, Kai Hwang, Jack Dongarra, Geoffrey C. Fox, Morgan Kaufmann, 1st edition, 2011, ISBN-13: 978-0123858801, 672 pages.

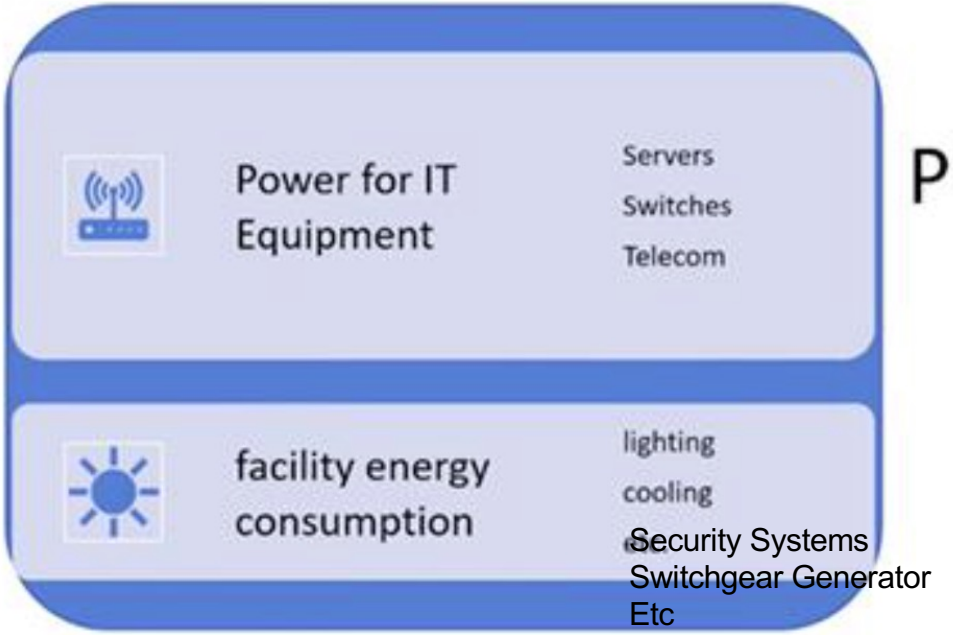
Warehouse-Scale Data Center Design

Free-cooling alternative



Warehouse-Scale Data Center Design

Data Center Power Usage Efficiency (PUE)



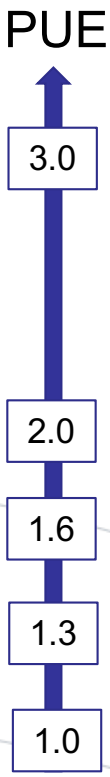
$$PUE = \frac{\text{Total Facility Power}}{\text{IT Equipment Power}}$$

Exemplo: Altice Datacenter Covilhã
tem um PUE de 1.25

Fonte: https://www.altice-empresas.pt/solucoes/cloud-datacenter/data-centers/covilha

Warehouse-Scale Data Center Design

Data Center Power Usage Efficiency (PUE)



Para fornecer 500W a um servidor são necessário 1500 W

Atíngivel com o desenho correcto

Muito evoluído tecnicamente
(referência Uptime Institute)
100% Eficiente

Altice Datacenter Covilha
PUE=1.25

Para fornecer 500 Watt a um servidor precisa **apenas de 625 Watt!**

Data-Center Interconnection Networks

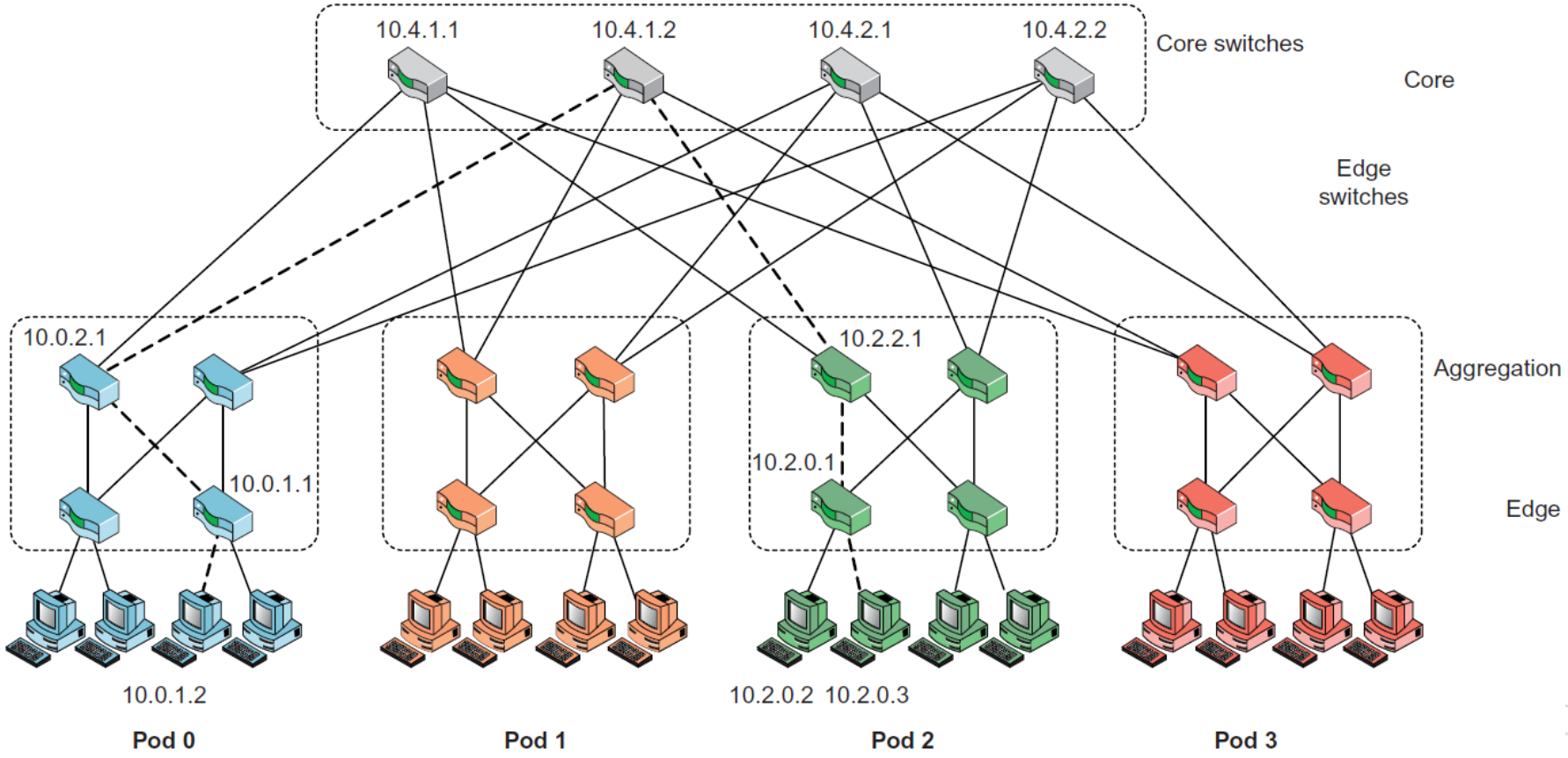
- A critical core design of a **data center** is the **interconnection network** among all **servers** in the datacenter cluster.
- This network design must meet **five special requirements**: **low latency**, **high bandwidth**, **low cost**, **message-passing interface** (MPI) **communication support**, and **fault tolerance**.
- The design of an inter-server network must satisfy both point-to-point and collective communication patterns among all server nodes.

Data-Center Interconnection Networks

- There are **two approaches** for building **data-center-scale networks**: One is **switch-centric** and the other is **server-centric**.
- In a **switch-centric network**, the **switches** are used to **connect** the **server nodes**. The switch-centric design does not affect the server side. No modifications to the servers are needed.
- The **server-centric design** does **modify** the **operating system** running on the **servers**. **Special drivers** are designed for **relaying** the **traffic**. **Switches** still have to be organized to achieve the connections.

Data-Center Interconnection Networks

A fat-tree interconnection topology for scalable data-center construction

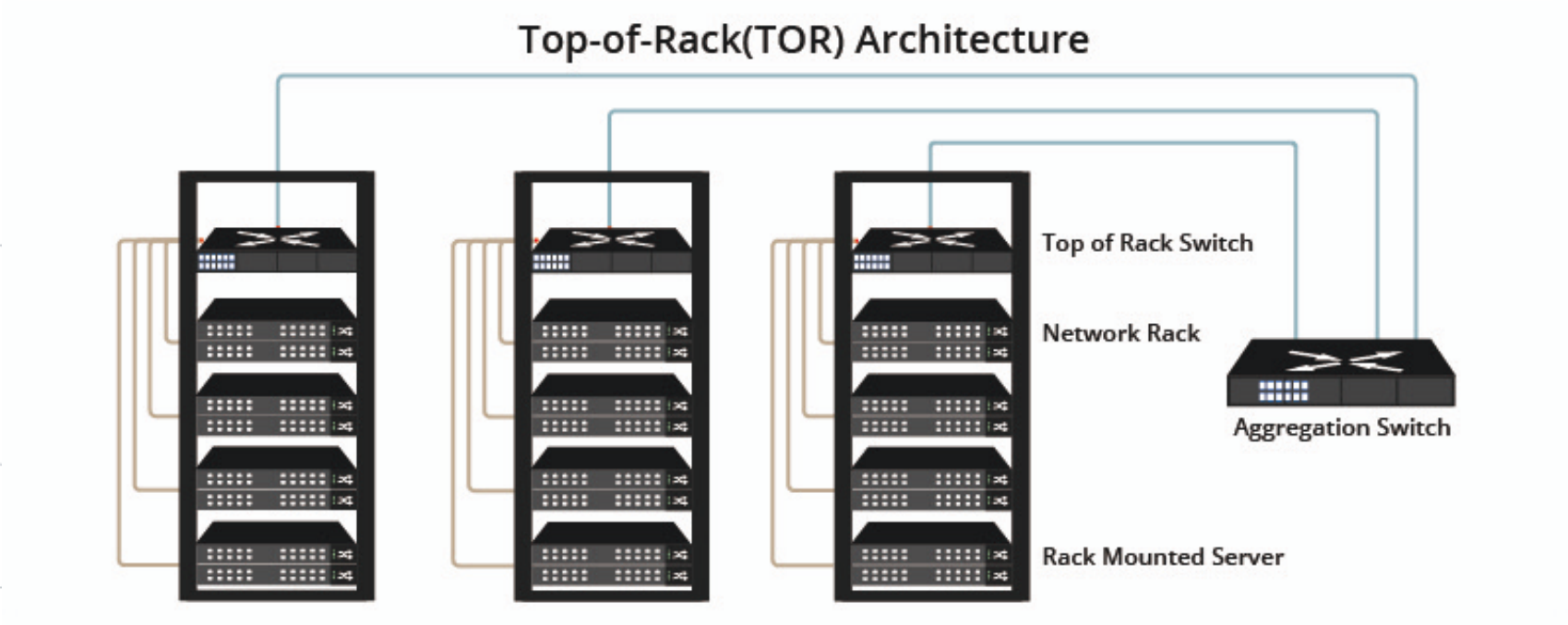


In: Distributed and Cloud Computing: From Parallel Processing to the Internet of Things, Kai Hwang, Jack Dongarra, Geoffrey C. Fox, Morgan Kaufmann, 1st edition, 2011, ISBN-13: 978-0123858801, 672 pages.

Data-Center Interconnection Networks

Arquitecturas ToR (Top of Rack) e EoR (End of Row)

(1) "***TOR architecture*** *refers to that the fiber patch cables in the server cabinets are connected directly to the switches placed on top of severs cabinets, and then directly to the core switch via the uplink port of the copper or fiber switches*".

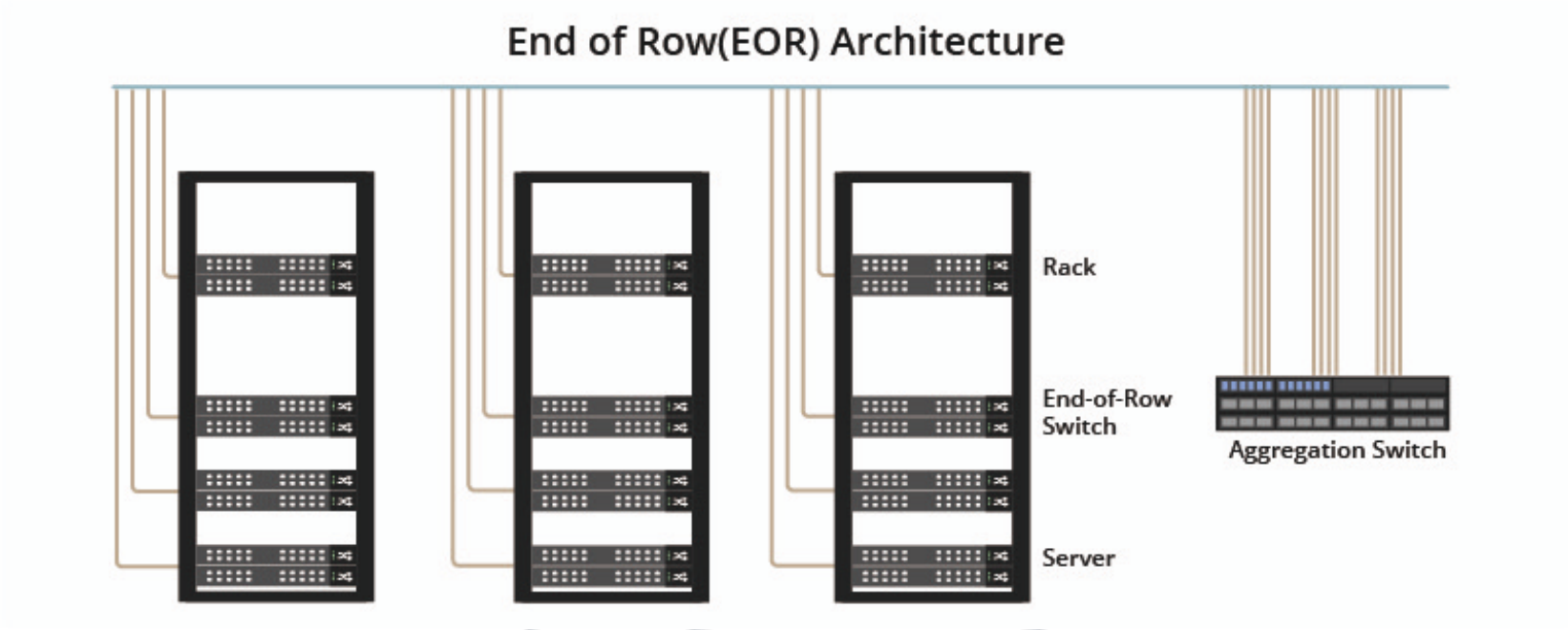


In: https://community.fs.com/blog/tor-vs-eor-data-center-architecture-design.html

Data-Center Interconnection Networks

Arquitecturas ToR (Top of Rack) e EoR (End of Row)

(2) "***The EOR architecture*** *means that the access layer switch installed at the end of the server cabinet completes a permanent connection to other facilities in the cabinet through horizontal cables. The EOR architecture requires a large number of horizontal cables to connect the devices and switch (…)*".



In: https://community.fs.com/blog/tor-vs-eor-data-center-architecture-design.html

Data-Center Interconnection Networks

ToR vs EoR

	ToR Design	EoR Design
Network Deployment	Minimum 1 switch per rack	Switches centrally residing in 1-2 racks of the same row
	Each rack is a separate	Module racks work as a group
Required Devices	Switch count is higher	Less number of switches
	Less number of Cables	Higher number of Cables
Power & Cooling	Underutilization of switch	Effective utilization of switches
	High power consumption	Less power consumption
	Greater need for cooling	Lesser need of cooling
Network Expansion	Greater layer 2 data traffic	Lesser layer 2 data traffic
	Network expansion is easy	Network expansion is difficult

Data-Center Interconnection Networks

3n Torus topology for scalable data-center construction

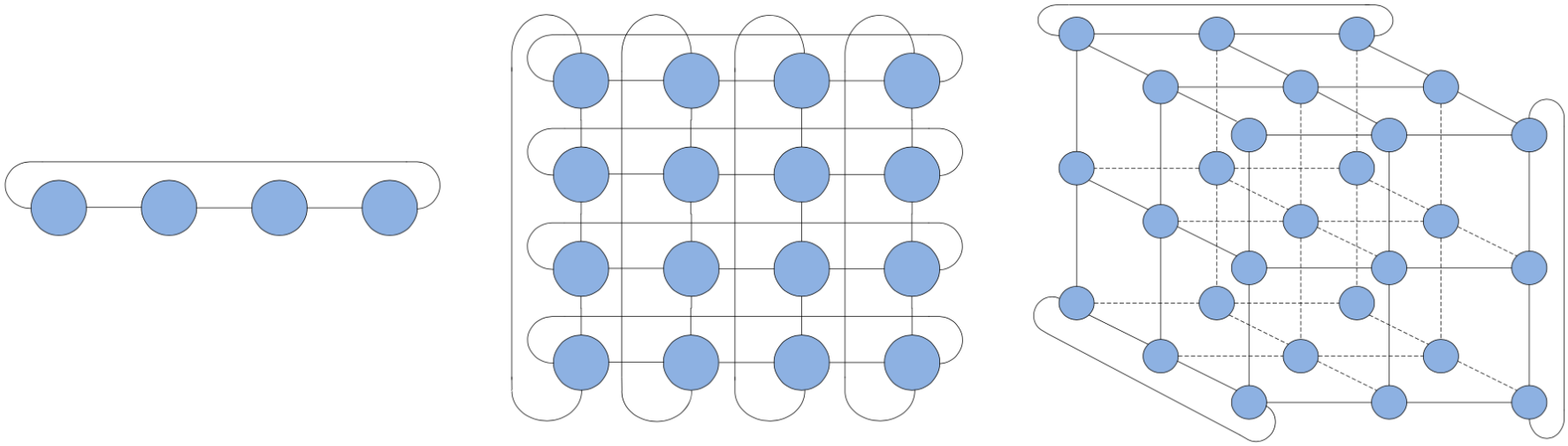


Fig. 1. Examples of 1D, 2D, 3D Torus topologies.

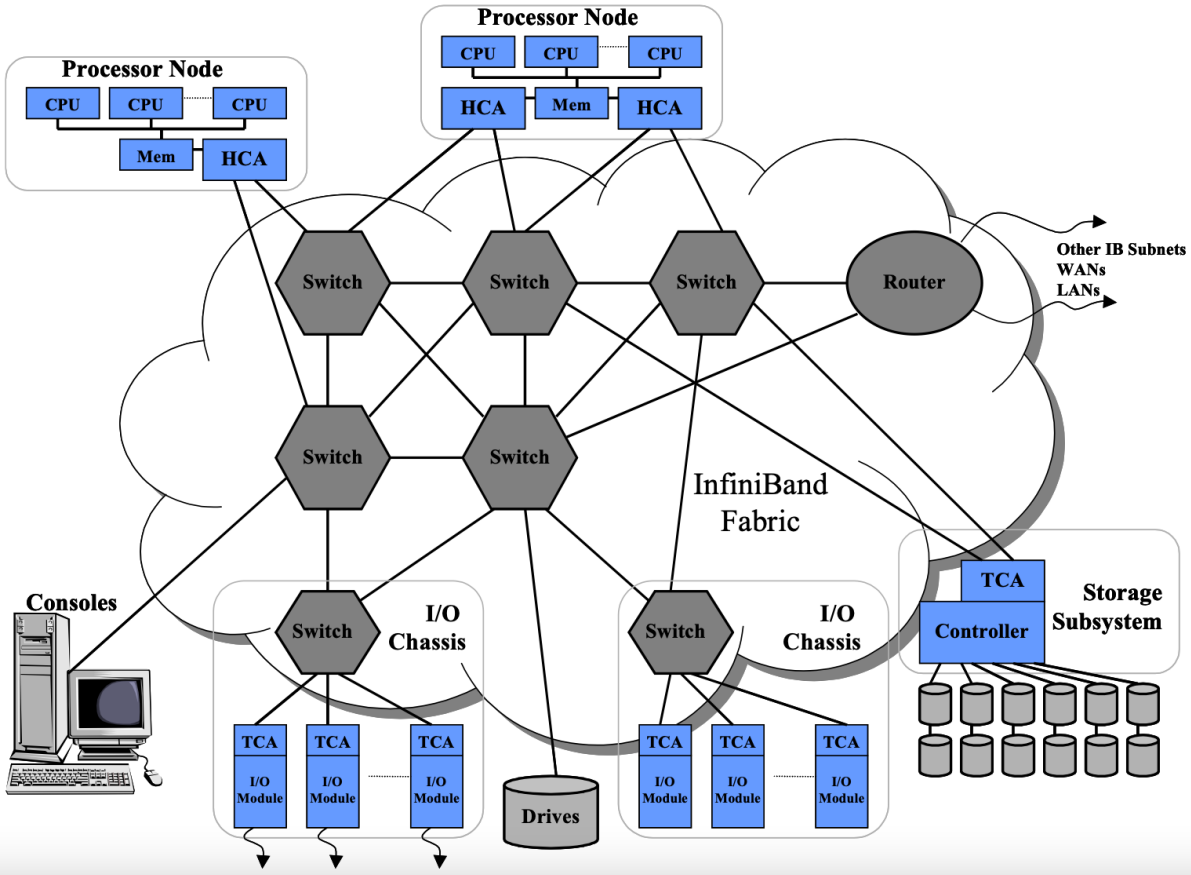
NOVO SLIDE 2021/2022
(Alexandre Fonte)

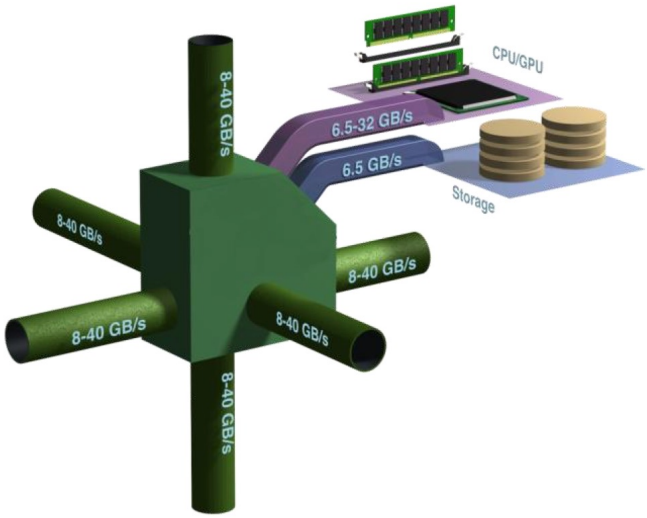
https://www.researchgate.net/publication/283100278_NovaCube_A_low_latency_Torus-based_network_architecture_for_data_centers/link/5a3e6a7d458515f6b03ba584/download

Data-Center Interconnection Networks

3n Torus topology for scalable data-center construction

Figure 1. InfiniBand System Fabric





NOVO SLIDE 2021/2022
(Alexandre Fonte)

In:https://www.hpcadvisorycouncil.com/events/2011/switzerland_workshop/pdf/Presentations/Day%201/10_HPCAC.pdf.

In: https://network.nvidia.com/related-docs/whitepapers/IB_Intro_WP_190.pdf

Modular Data Center in Shipping Containers

- A **modern data center** is **structured** as a **shipyard** of **server clusters** housed in **truck-towed** containers.
- Inside the container, hundreds of blade servers are housed in racks surrounding the container walls. An array of fans forces the heated air generated by the server racks to go through a heat exchanger, which cools the air for the next rack (detail in callout) on a continuous loop.
- The SGI ICE Cube container can house 46,080 processing cores or 30-120 PB of storage per container.
- Data centers usually are built at a site where leases and utilities for electricity are cheaper, and cooling is more efficient.

Modular Data Center in Shipping Containers

A modular data center built in a truck-towed ICE Cube container, that can be cooled by chilled air circulation with cold-water heat exchanges

In: Distributed and Cloud Computing: From Parallel Processing to the Internet of Things, Kai Hwang, Jack Dongarra, Geoffrey C. Fox, Morgan Kaufmann, 1st edition, 2011, ISBN-13: 978-0123858801, 672 pages.

Modular Data Center in Shipping Containers

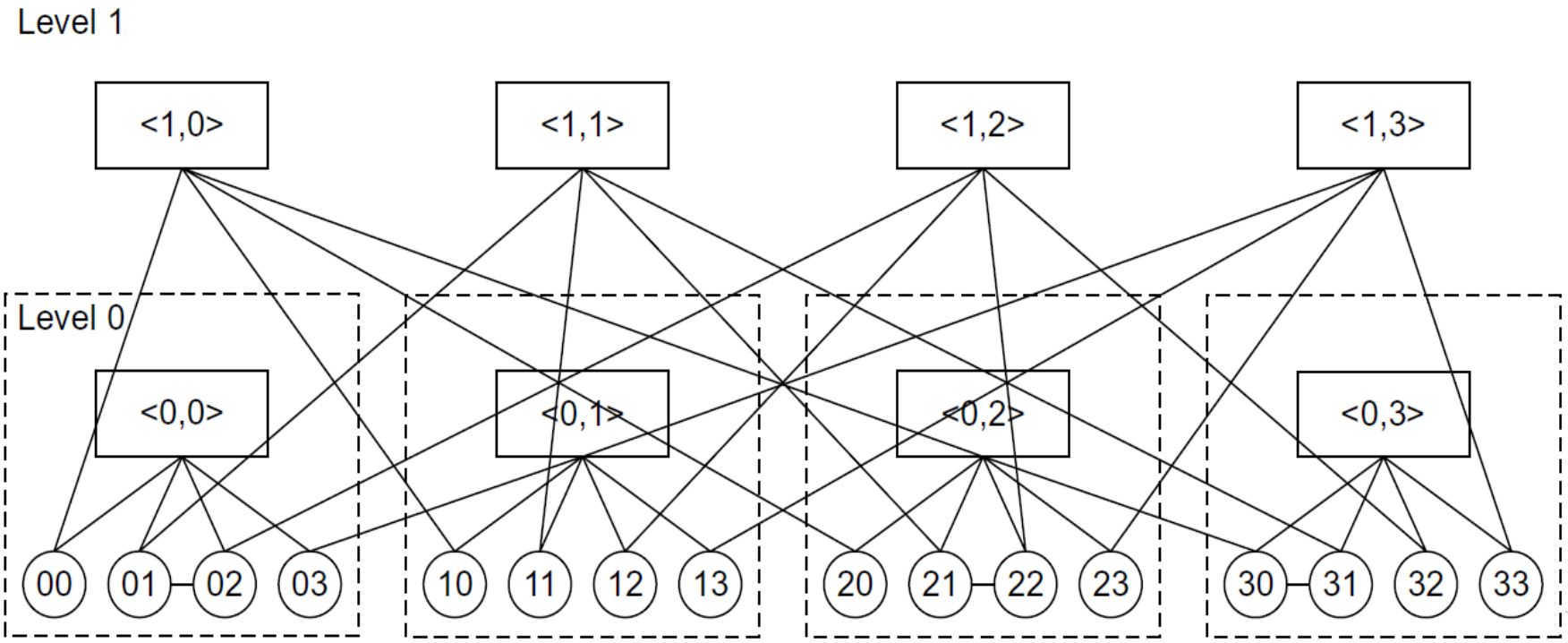
Video Ilustrativo:
https://www.youtube.com/watch?v=-U1ZVOcviL8&t=17s

Interconnection of Modular Data Centers

- Container-based data-center modules are meant for construction of even larger data centers using a farm of container modules.
- The **containers** are considered the **building blocks** for **datacenters**. Thus, despite the design of the inner container network, one needs another level of networking among multiple containers.
- The **architecture** builds a **virtual hypercube** at the **container level**, in addition to the cube structure inside the container (BCube). With the server container built with the BCube network, the MDCube is used to build a large-scale data center for supporting cloud application communication patterns.

Interconnection of Modular Data Centers

BCube, a high-performance, server-centric network for building modular data centers



In: Distributed and Cloud Computing: From Parallel Processing to the Internet of Things, Kai Hwang, Jack Dongarra, Geoffrey C. Fox, Morgan Kaufmann, 1st edition, 2011, ISBN-13: 978-0123858801, 672 pages.

Data-Center Management Issues

- Basic requirements for managing the resources of a data center:
- **Making common users happy**: The data center should be designed to provide quality of service to the majority of users for at least 30 years.
- **Controlled information flow**: Information flow should be streamlined. Sustained services and high availability (HA) are the primary goals.
- **Multiuser manageability**: The system must be managed to support all functions of a data center, including traffic flow, database updating, and server maintenance.
- **Scalability to prepare for database growth**: The system should allow growth as workload increases. The storage, processing, I/O, power, and cooling subsystems should be scalable.

Data-Center Management Issues

- **Reliability in virtualized infrastructure**: Failover, fault tolerance, and VM live migration should be integrated to enable recovery of critical applications from failures or disasters.
- **Low cost to both users and providers**: The cost to users and providers of the cloud system built over the data centers should be reduced, including all operational costs.
- **Security enforcement and data protection**: Data privacy and security defense mechanisms must be deployed to protect the data center against network attacks and system interrupts and to maintain data integrity from user abuses or network attacks.
- **Green information technology**: Saving power consumption and upgrading energy efficiency are in high demand when designing and operating current and future data centers.