POLITECNICO DI MILANO

Computing Infrastructures

1) Computing Infrastructures

Prof. Danilo Ardagna

Credits: Prof. Manuel Roveri

The topics of the course: what are we going to see today?

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 re

Feb to mid March

B. SW Infrastructures:

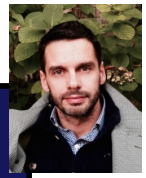
- **Virtualization**: Process/System VM, Virtualization Mechanisms (Hypervisor, Para/Full virtualization)
- **Computing Architectures**: Cloud Computing (types, characteristics), Edge/Fog Computing, X-as-a service
- **Machine and deep learning-as-a-service**

End of May to June

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)

Mid March to April



May

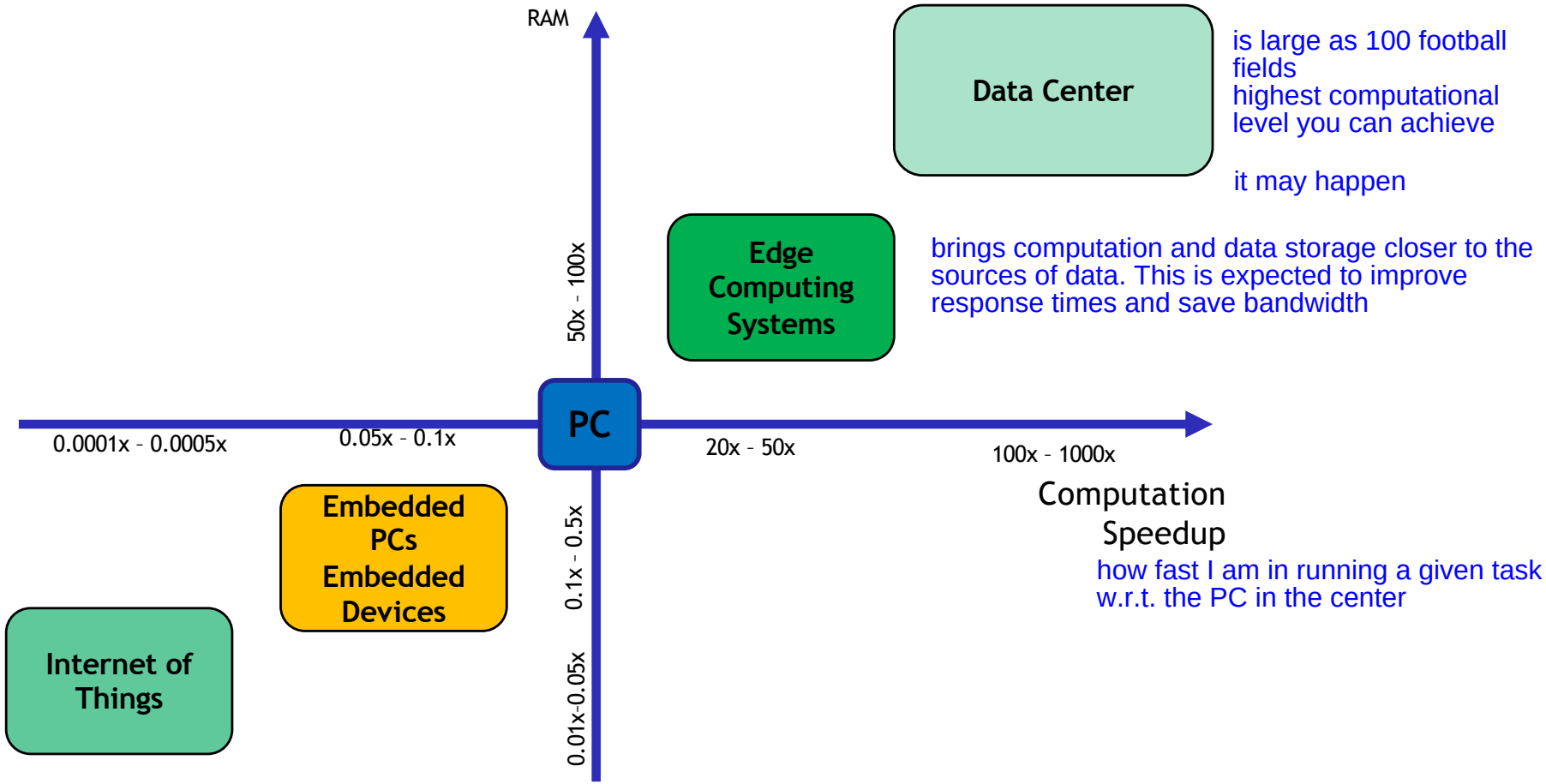
What is a computing infrastructure?

Technological infrastructure that provides hardware and software for computation to other systems and services

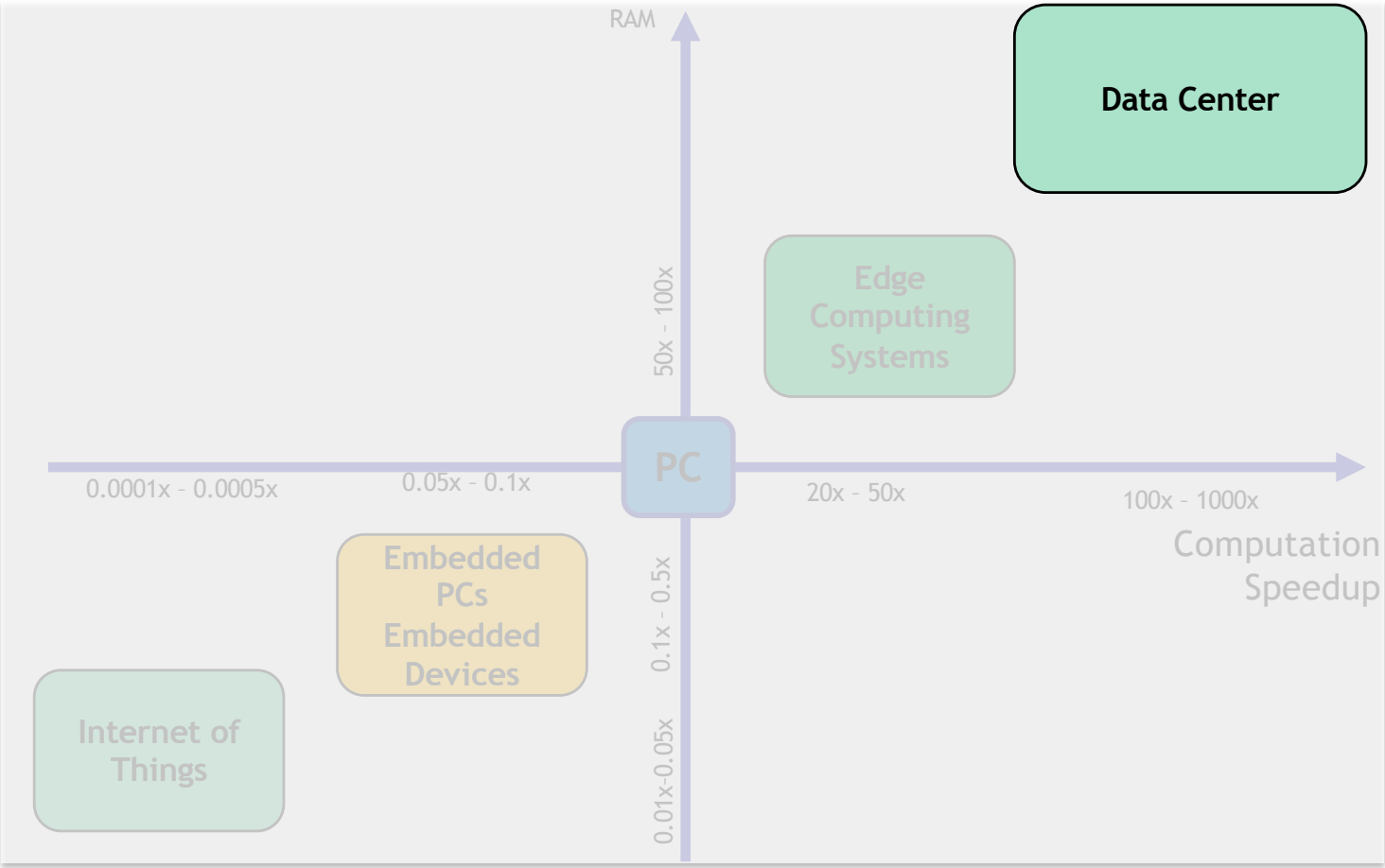
HW and SW work together to optimize services to multiple users
examples of computing infrastructure:
-super computers
-cloud data center
-AWS is a service that run on top of Amazon computer infrastructure
-internet is not a computing infrastructure, it is a network infrastructure
-politecnico system is a computing infrastructure

Examples of Computing Infrastructures

PC is not a computing infrastructure because is personal, doesn't provide services to third party



Examples of Computing Infrastructures



Data Centers: a technological perspective

The Pionen White Mountains is a Swedish data center. This center is located in Stockholm and is one of the largest data centers in the world.

Server for Processing

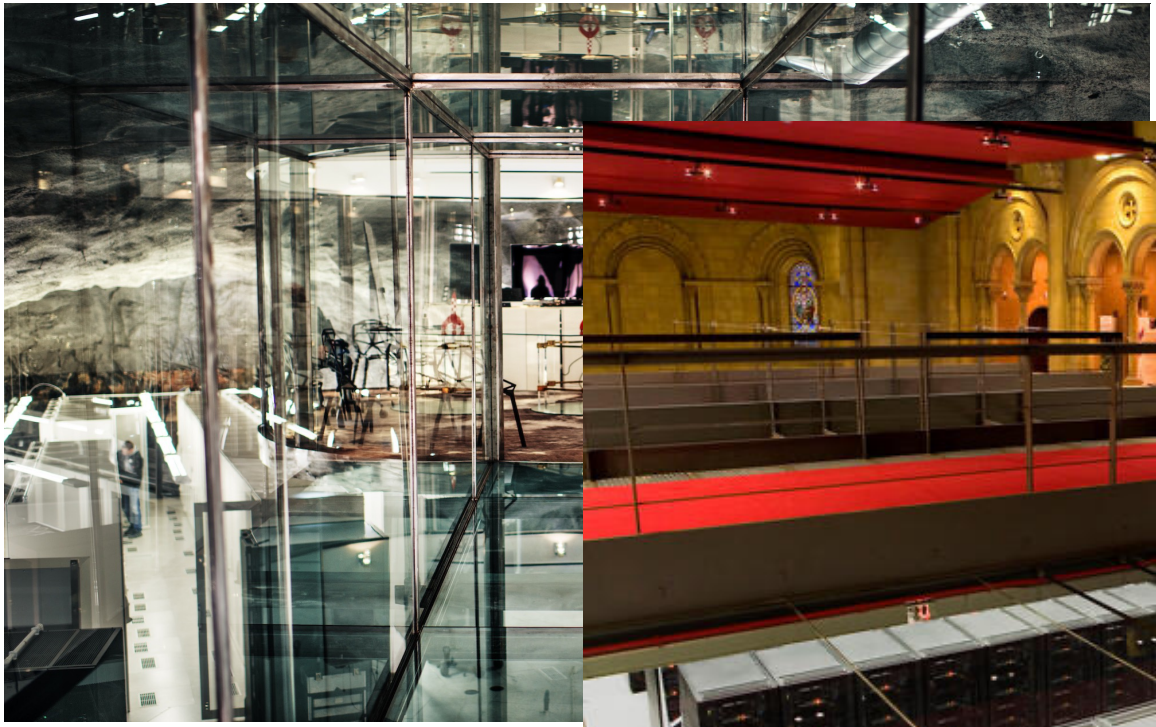
Server for Storage

Server for Communication

three pilas

these servers are in wrecks and there is a network switch that connects all the servers in a wreck then all connected together to provide external connectivity

Data Centers: a technological perspective



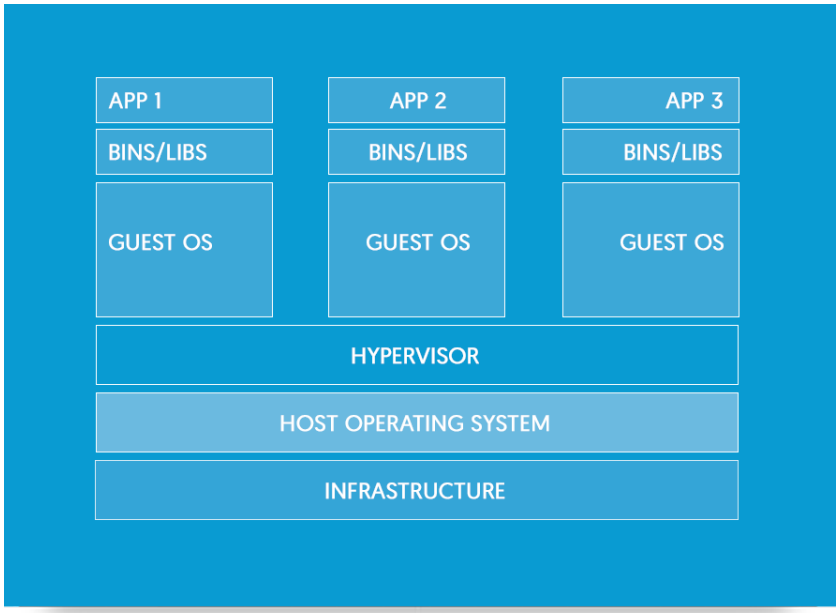
The Pionen White Mountains is a Swedish c
in Stockholm and is one of the larges

Barcelona Super Computing Center. Marenostrum used to be a church
good for the temperature

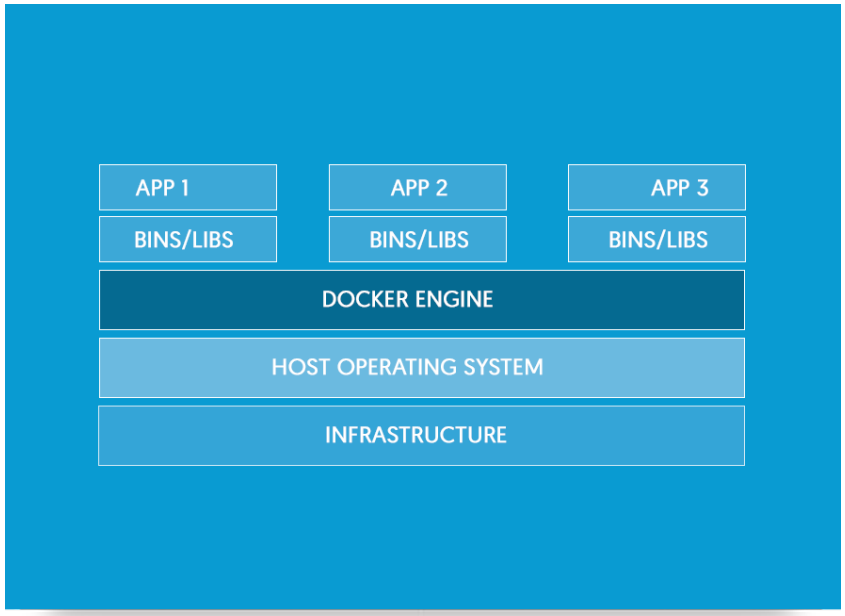
Virtual Machines and Containers

the data centers are based on virtualization
why virtualization?
utilization= fraction of time in which the server is running
in the past one application per server and 10/15% utilization consuming a lot of energy
=> virtualization
I can move applications in virtual machines -> if I need two applications running at 15% each, I can replace the two systems with a single server where the two application can run isolated

VMMs
Provide the full stack (OS, LIB, APP).
Applications depend on guest OS.



Containers
Applications are packaged with all their dependencies into a standardized unit for software development/deployment.



virtualization -> reduced the costs of data center
-> it allows me to manage effectively the infrastructures

Data Centers: advantages and disadvantages

advantages of cloud data centers

- ✓ Lower IT costs
- ✓ High performance
- ✓ Instant software updates
- ✓ "Unlimited" storage capacity
- ✓ Increased data reliability
- ✓ Universal document access
- ✓ Device Independence

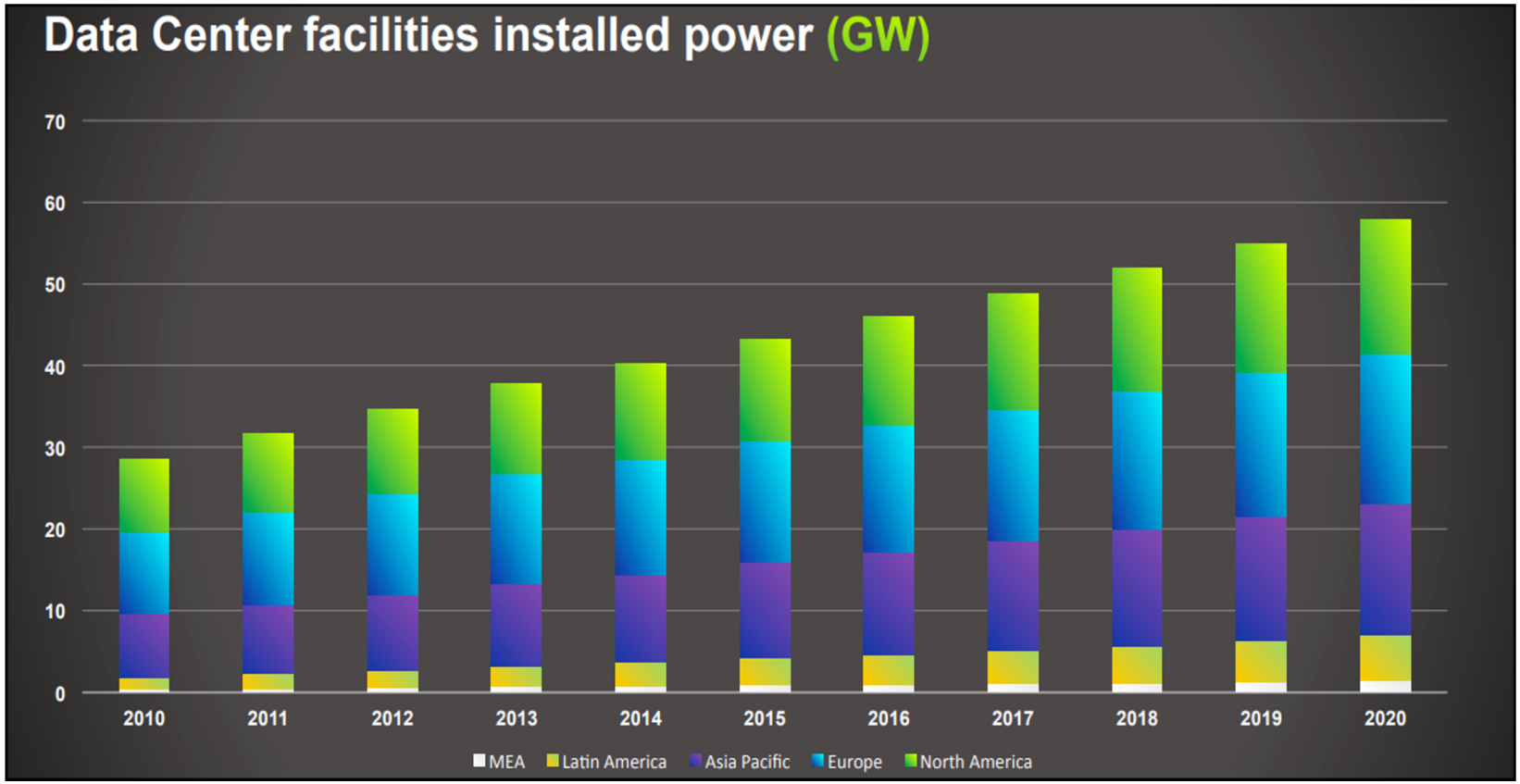
access same data from different devices

- ➢ Require a constant Internet connection
- ➢ Do not work well with low-speed connections
- ➢ Hardware Features might be limited
- ➢ Privacy and security issues
- ➢ High Power Consumption
- ➢ <u>Latency in making decision</u>

usually the data centers have an homogeneous structure

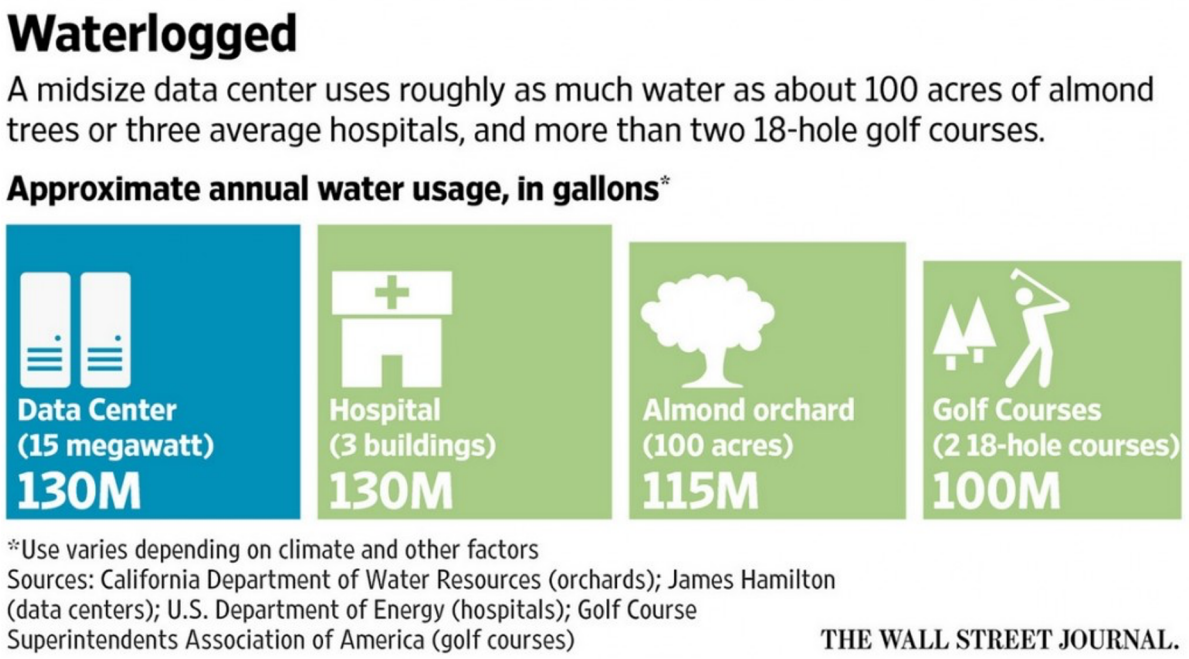
Data Centers: advantages and disadvantages

INSTALLED POWER electricity energy consumed by a data center --> consume the same amount of power of small cities

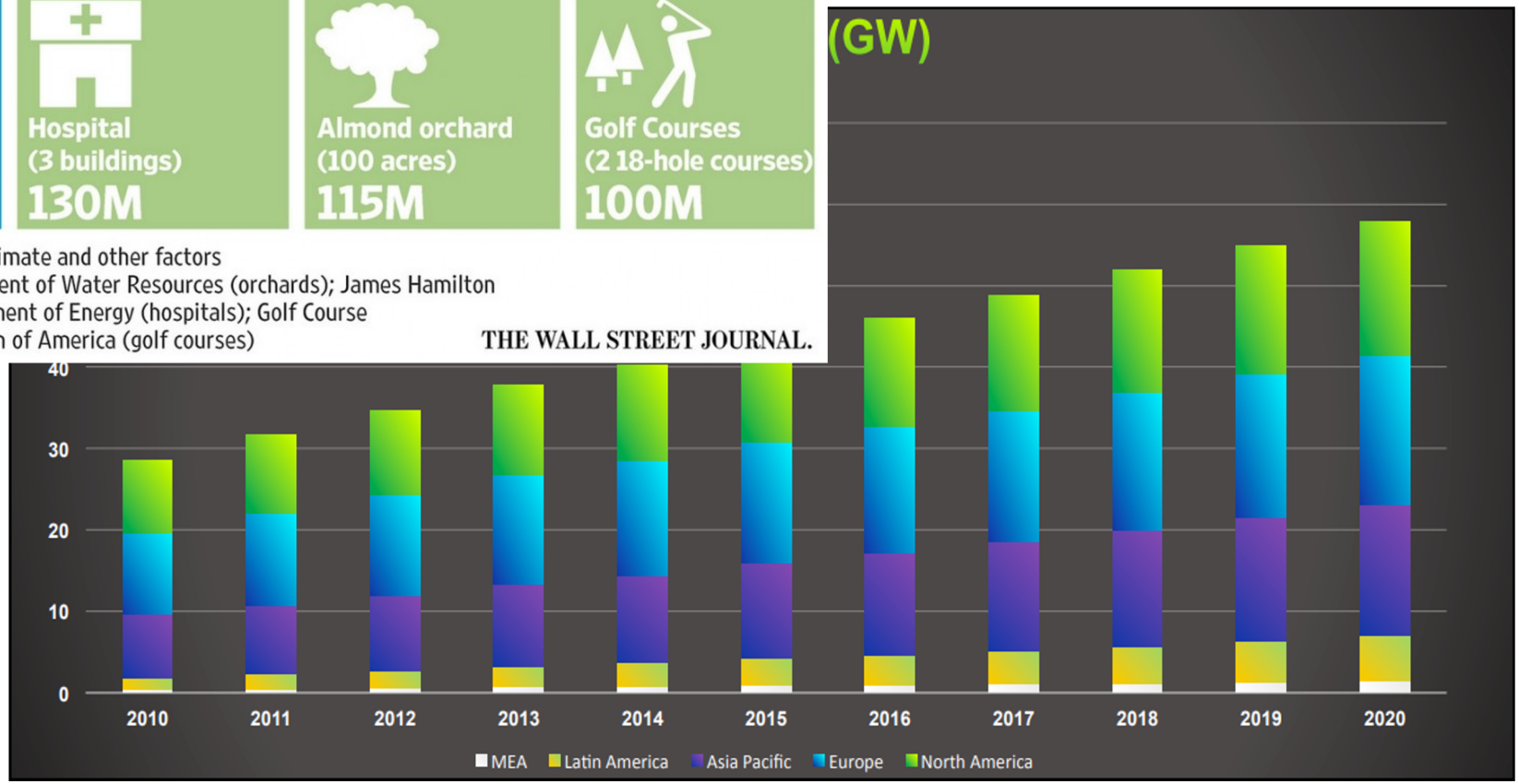


Data Centers: advantages and disadvantages

WATER

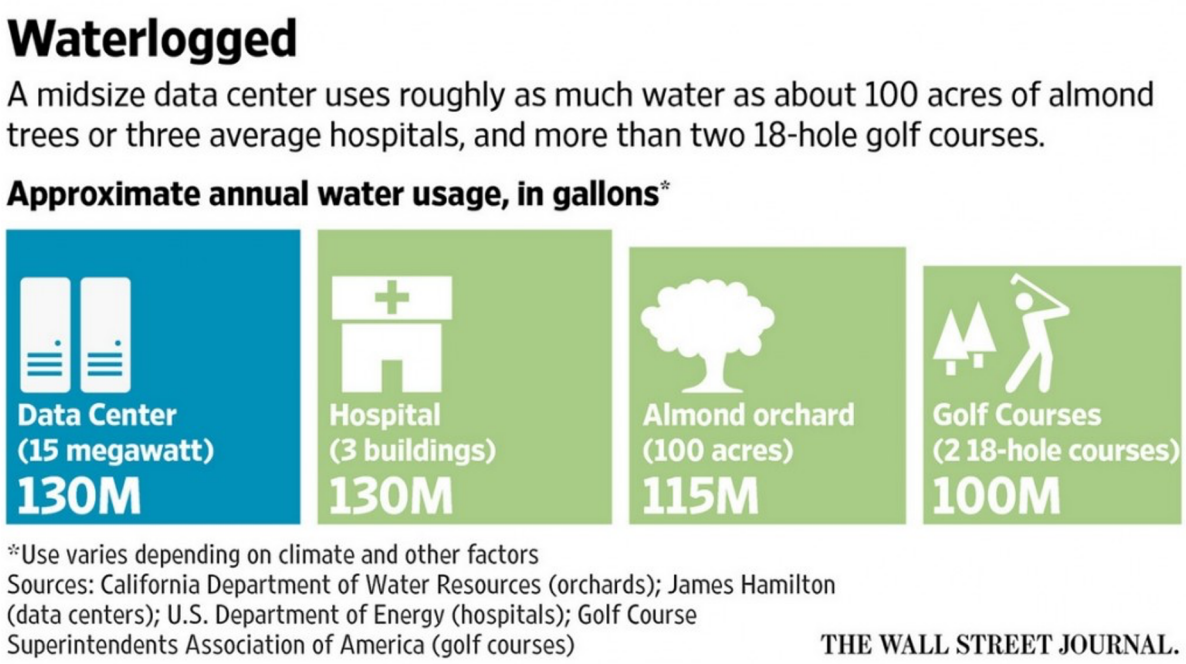


ED POWER

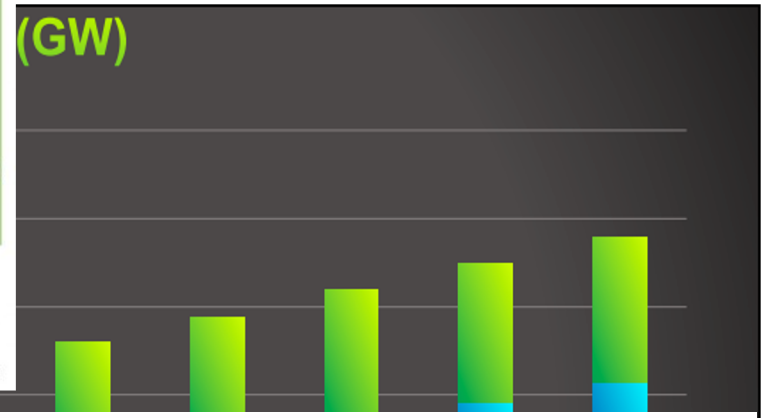


Data Centers: advantages and disadvantages

WATER

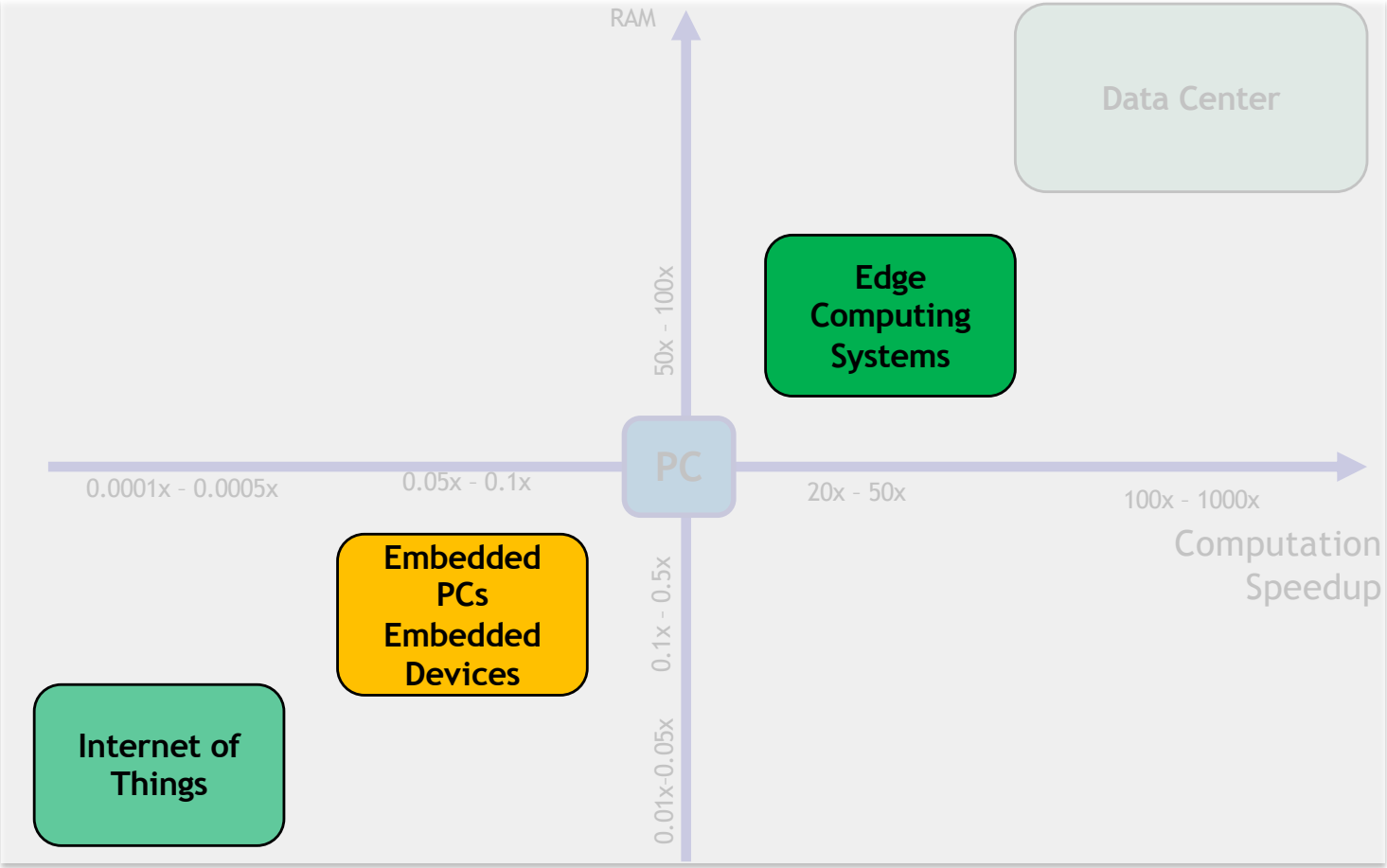


ED POWER



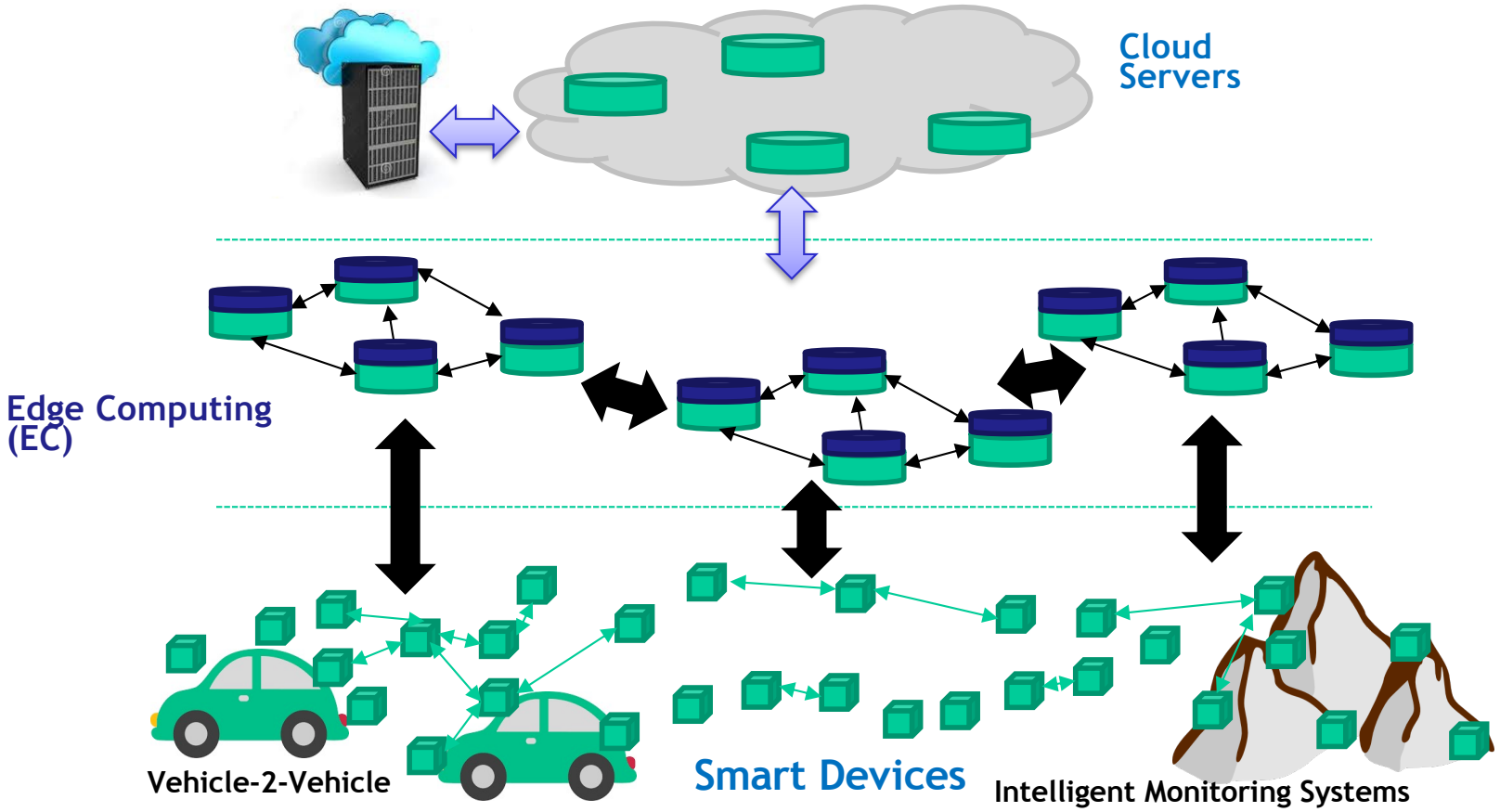
Amortized Cost	Component	Sub-Components
~45%	Servers	CPU, memory, disk
~25%	Infrastructure	UPS, cooling, power distribution
~15%	Power draw	Electrical utility costs
~15%	Network	Switches, links, transit

Edge Computing, PC Embedded and IoT



Edge/Fog Computing Systems

whenever the latency is too much
put a computational layer between the cloud servers
and where the data is produced



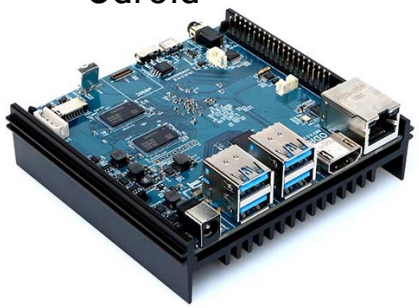
IoT Gateways	SYS-E50-9AP-WIFI	SYS-E100-9S-E	SYS-E300-8D	SYS-5018D-FN8T
Processor/Cache				
CPU	**Intel® Atom® processor E3940** Single socket FCBGA 1296 9.5W, 4C	**7th Generation Intel® Core i5-7300U Processor** Single Socket FCBGA 1356 System-on-Chip CPU TDP support 15W	**Intel® processor D-1518,** 2.2GHz; CPU TDP support 35W FCBGA 1667: 4 Cores, 8 Threads / 6MB	**Intel® Xeon® processor D-1518** 2.2GHz; CPU TDP support 35W FCBGA 1667: 4 Cores, 8 Threads / 6MB
System Memory				
Memory Capacity	Up to 8GB Unbuffered non-ECC SO-DIMM DDR3L-1866MHz, in 1 DIMM socket	Up to 32GB Unbuffered non-ECC SO-DIMM, DDR4-2133MHz, in 2 DIMM slots	4x DDR4 DIMM sockets Supports up to 128GB DDR4 ECC RDIMM Supports up to 64GB DDR4 ECC/non-ECC UDIMM	4x DDR4 DIMM sockets Supports up to 128GB DDR4 ECC RDIMM Supports up to 64GB DDR4 ECC/non-ECC UDIMM
Memory Type	DDR3L up to 1866MHz	DDR4 up to 2133MHz	2133/1866/1600MHz ECC DR4 ECC RDIMM and ECC/Non-ECC UDIMM	2133/1866/1600MHz ECC DDR4 ECC RDIMM and ECC/Non-ECC UDIMM
DIMM Sizes	8GB, 4GB, 2GB	16GB, 8GB, 4GB	32GB, 16GB, 8GB, 4GB	32GB, 16GB, 8GB, 4GB
Memory Voltage	1.35 V	1.2 V	1.2 V	1.2 V

- ✓ High computational capacity
- ✓ Distributed computing
- ✓ Privacy and security
- ✓ Reduced Latency in making a decision

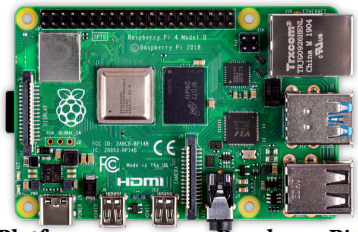
- ➢ Require a power connection
- ➢ Require connection with the Cloud

Embedded PCs

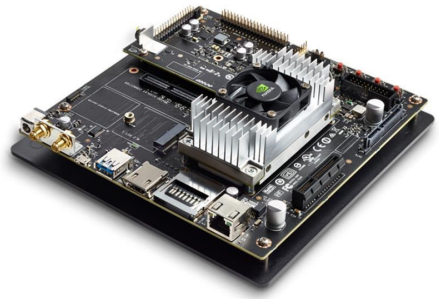
Odroid



Raspberry



Platform	Raspberry Pi 3B
CPU	ARM11 @ 1.2 GHz
RAM	1024 MB



Jetson TX2

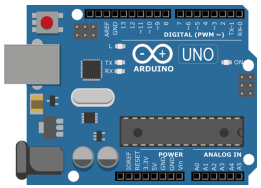
Google Coral

- ✓ Pervasive computing
- ✓ High performance unit
- ✓ Availability of development boards
- ✓ Programmed as PC
- ✓ Large community

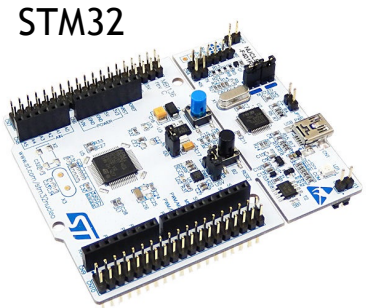
- ➢ Pretty high power consumption
- ➢ (Some) HW design has to be done

Internet-of-Things

Arduino



STM32

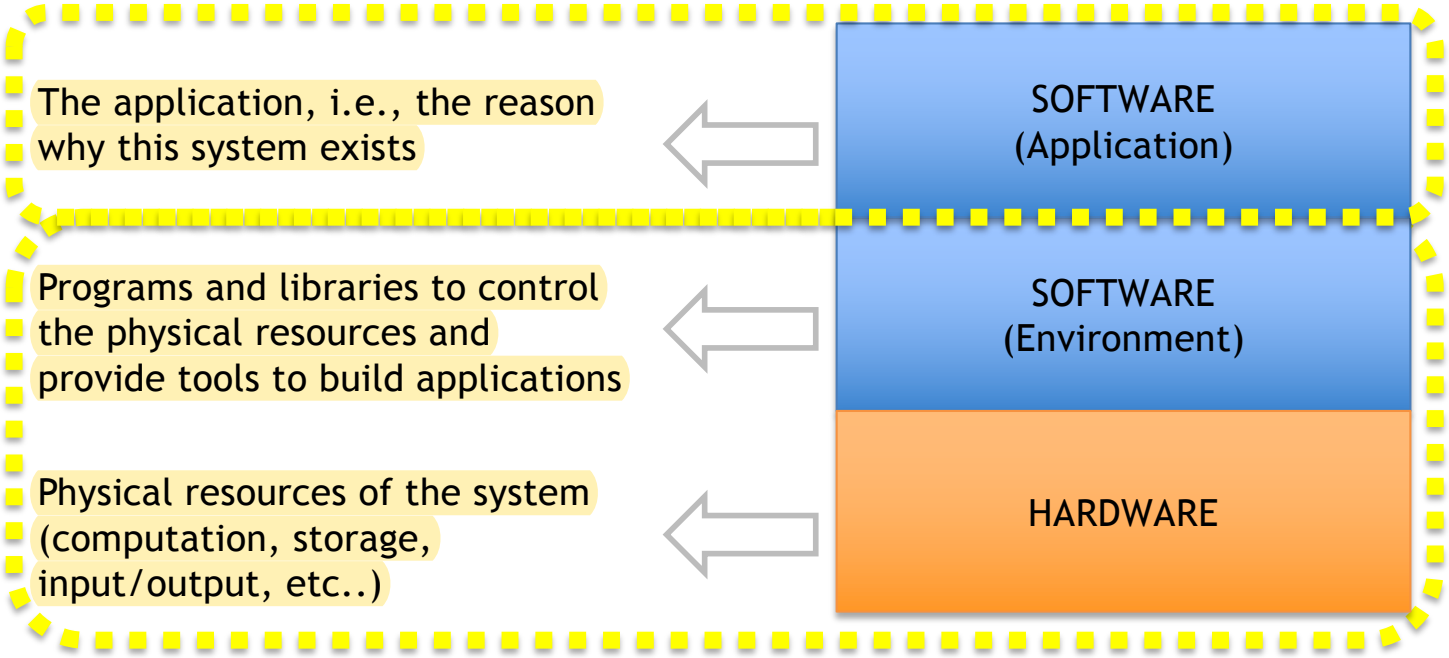
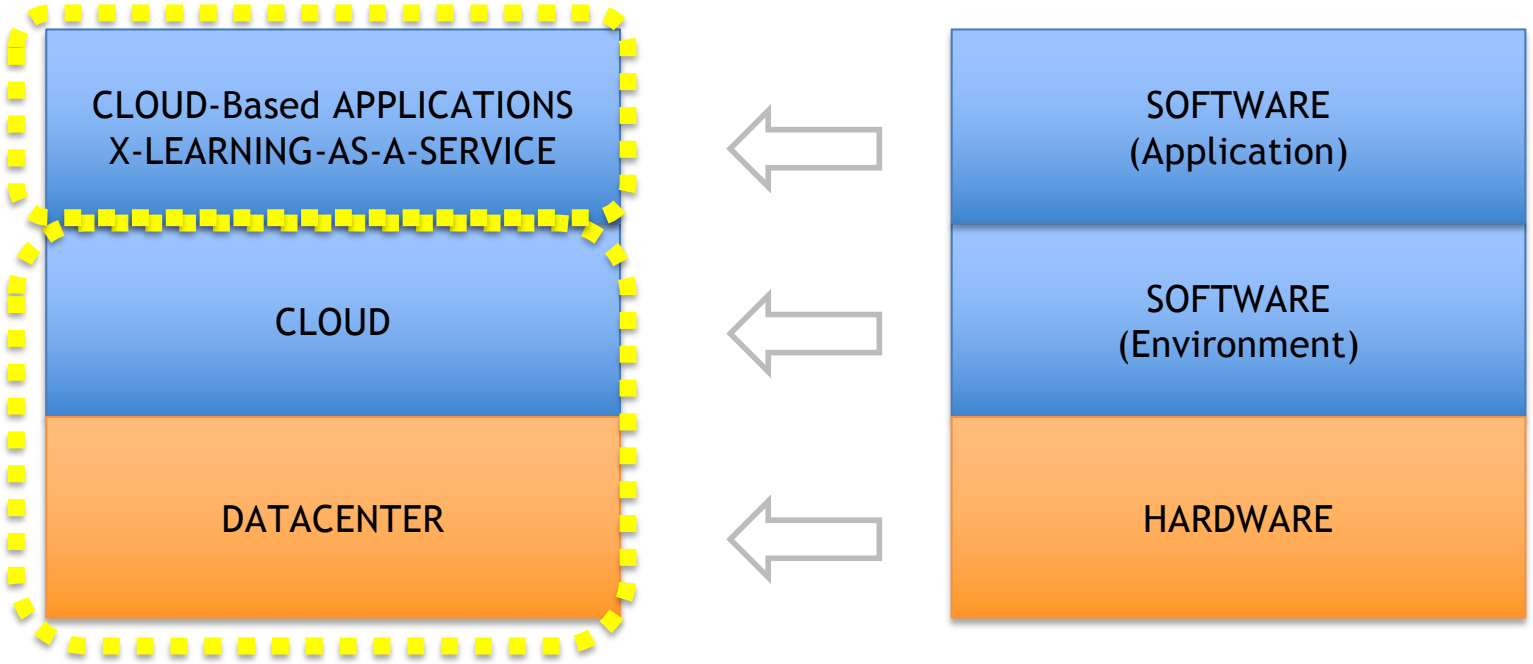


	STM32 L1 Series	STM32F4 Series
Domain	Ultra Low-Power	High-Performance
Flash Memory (kB)	32 to 512	64 to 2048
RAM Memory (kB)	4 to 80	32 to 320
CPU	ARM® Cortex®-M3	ARM® Cortex®-M4
Frequency (MHz)	32	84 to 180
Supply Voltage (V)	1.65 to 3.6	1.71 to 3.6
Supply Current (μA)	0.28 (0.28) to 230	1.1 (140) to 282

- ✓ Highly Pervasive
- ✓ Wireless connection
- ✓ Battery Powered
- ✓ Low costs
- ✓ Sensing and actuating

- ➢ Low computing ability
- ➢ Constraints on energy
- ➢ Constraints on memory (RAM/FLASH)
- ➢ Difficulties in programming

An IT perspective for Computing Infrastructures



An IT perspective for Computing Infrastructures