Big Data Analytics

Saeed Sharifian

Course Objective

- This course has a system-based focus
- Learn the theory of machine learning and deep learning
- Learn the practical aspects of building machine learning and deep learning algorithms using data parallel programming platforms, such as Spark and TensorFlow

large scale machine learning

- Spark ML

large scale deep learning

- TensorFlow
- Deep feedforward networks
- Convolutional neural networks (CNNs)
- Recurrent neural networks (RNNs)
- Autoencoders

Scala/Python programming

Artificial Intelligence Challenge

- Artificial intelligence (AI) can solve problems that can be described by a list of formal mathematical rules.
- The challenge is to solve the tasks that are hard for people to describe formally.
- Let computers to learn from experience.

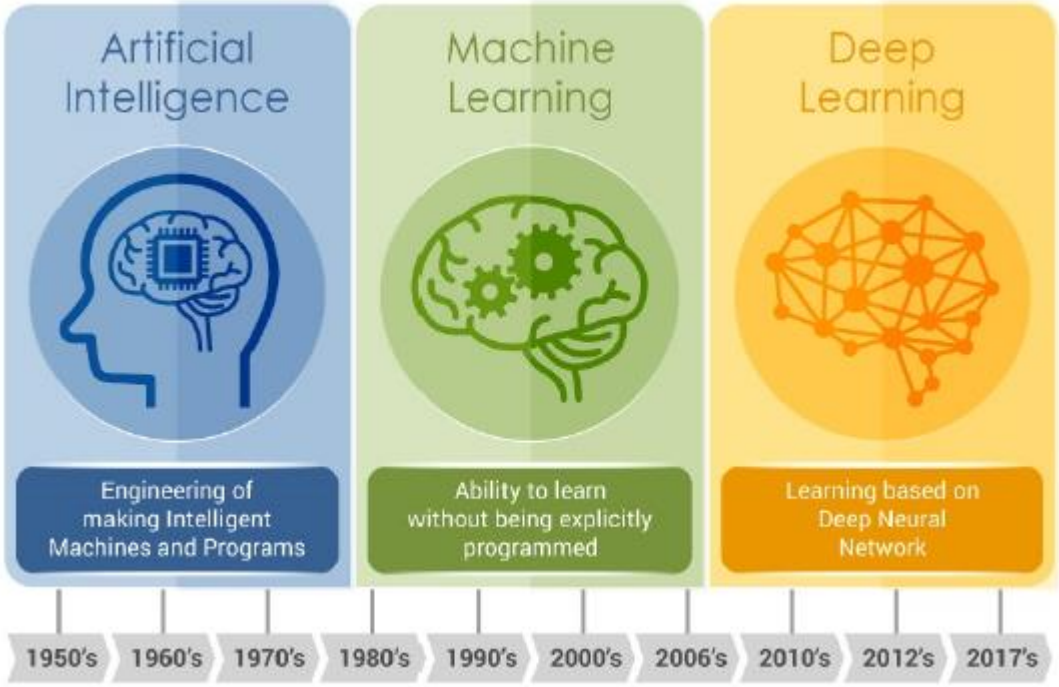
Sheepdog or Mop

Chihuahua or Muffin

Barn Owl or Apple

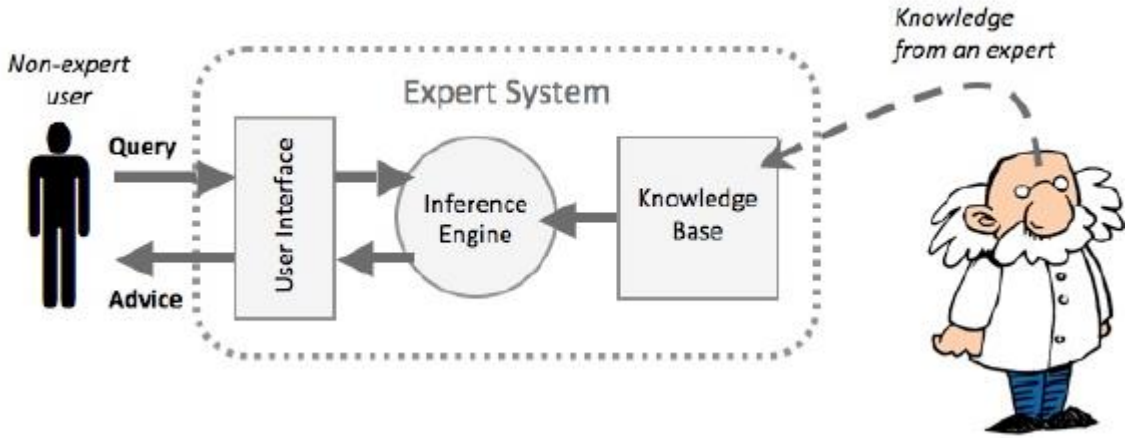
AI Generations

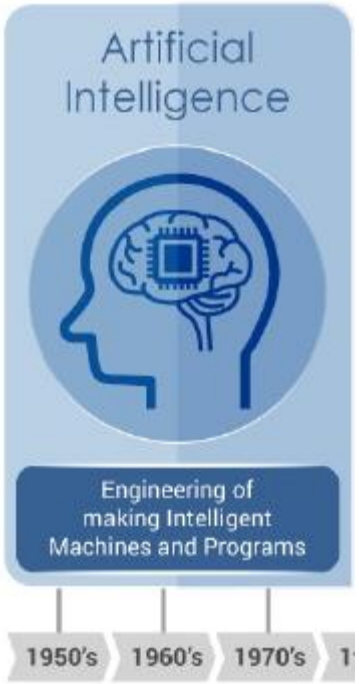
- Rule-based AI
- Machine learning
- Deep learning



AI Generations - Rule-based AI

- Hard-code knowledge
- Computers reason using logical inference rules





- The programs that solve problems in a specific domain.
- Two engines:
  - Knowledge engine: represents the facts and rules about a specific topic.
  - Inference engine: applies the facts and rules from the knowledge engine to new facts.

AI Generations - Machine Learning

- If AI systems acquire their own knowledge
- Learn from data without being explicitly programmed
  - A ML algorithm is an algorithm that is able to learn from data.
  - What is learning?
  - A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

AI Generations - Deep Learning

- For many tasks, it is difficult to know what features should be extracted
- Use machine learning to discover the mapping from representation to output

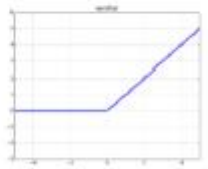
Why Does Deep Learning Work Now?

- Huge quantity of data
- Tremendous increase in computing power
- Better training algorithms

Data

GPUs

Weight Initialization



Non-Linearity

Big Data

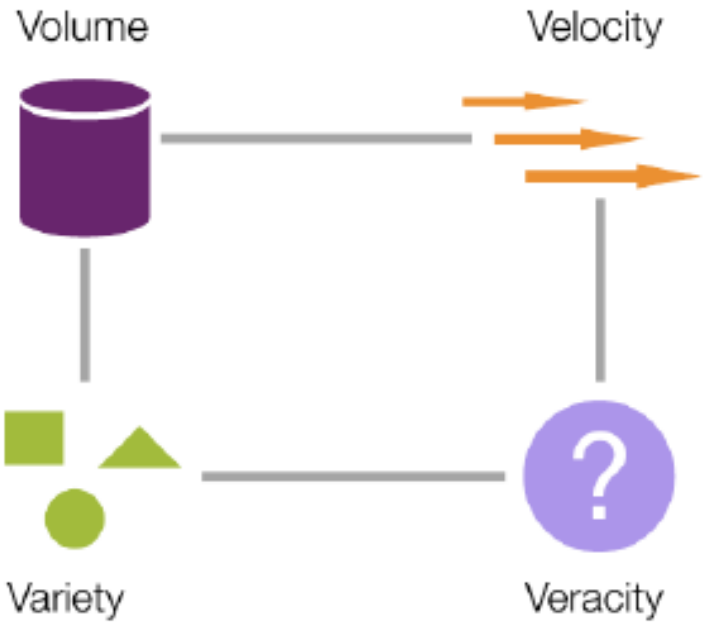
DevOps Borat
@DEVOPS_BORAT

Small Data is when is fit in RAM.
Big Data is when is crash because is not fit in RAM.

2/6/13, 8:22 AM

The Four Dimensions of Big Data

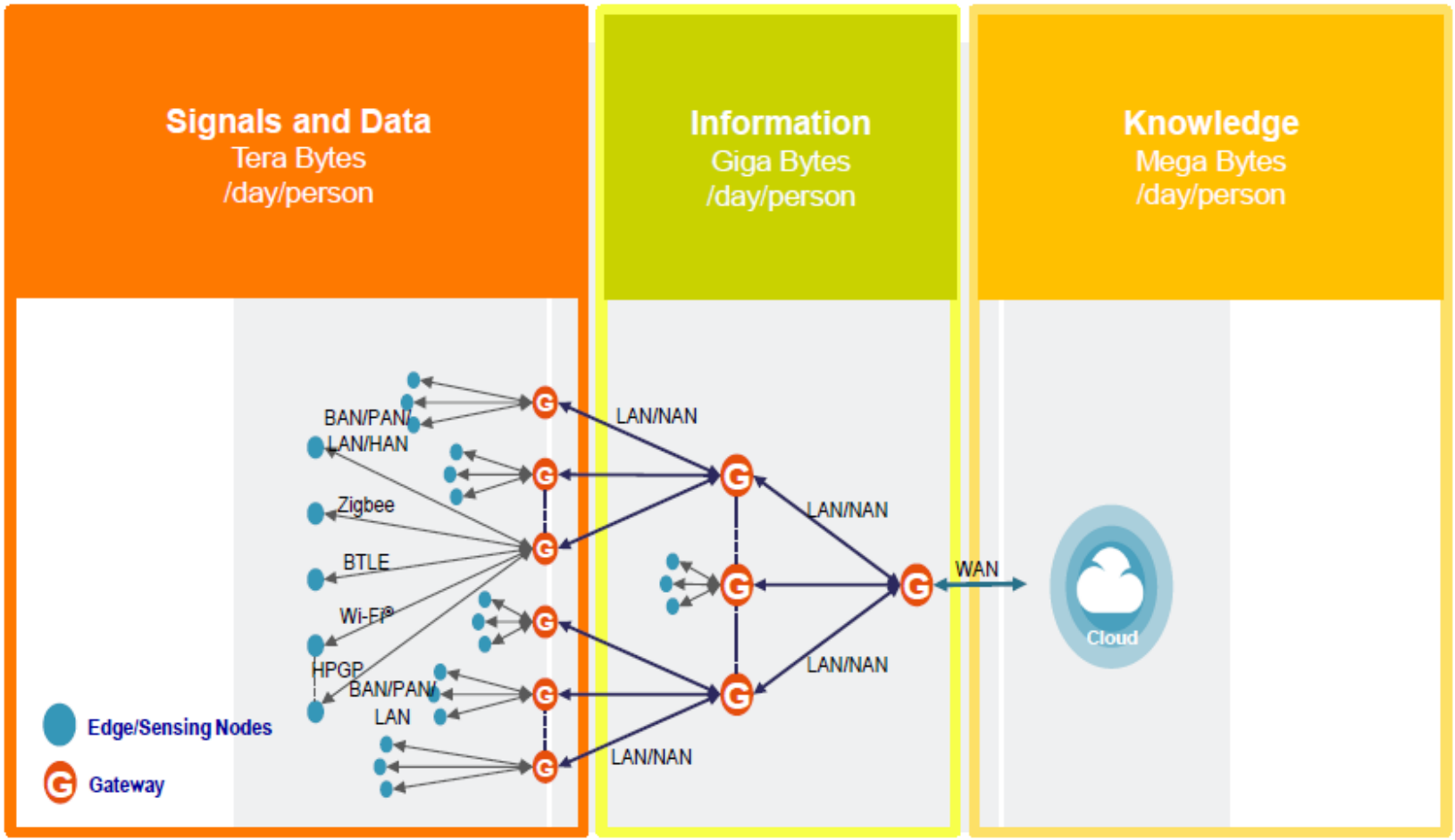
- Volume: data size
- Velocity: data generation rate
- Variety: data heterogeneity
- This 4th V is for Vacillation: Veracity/Variability/Value

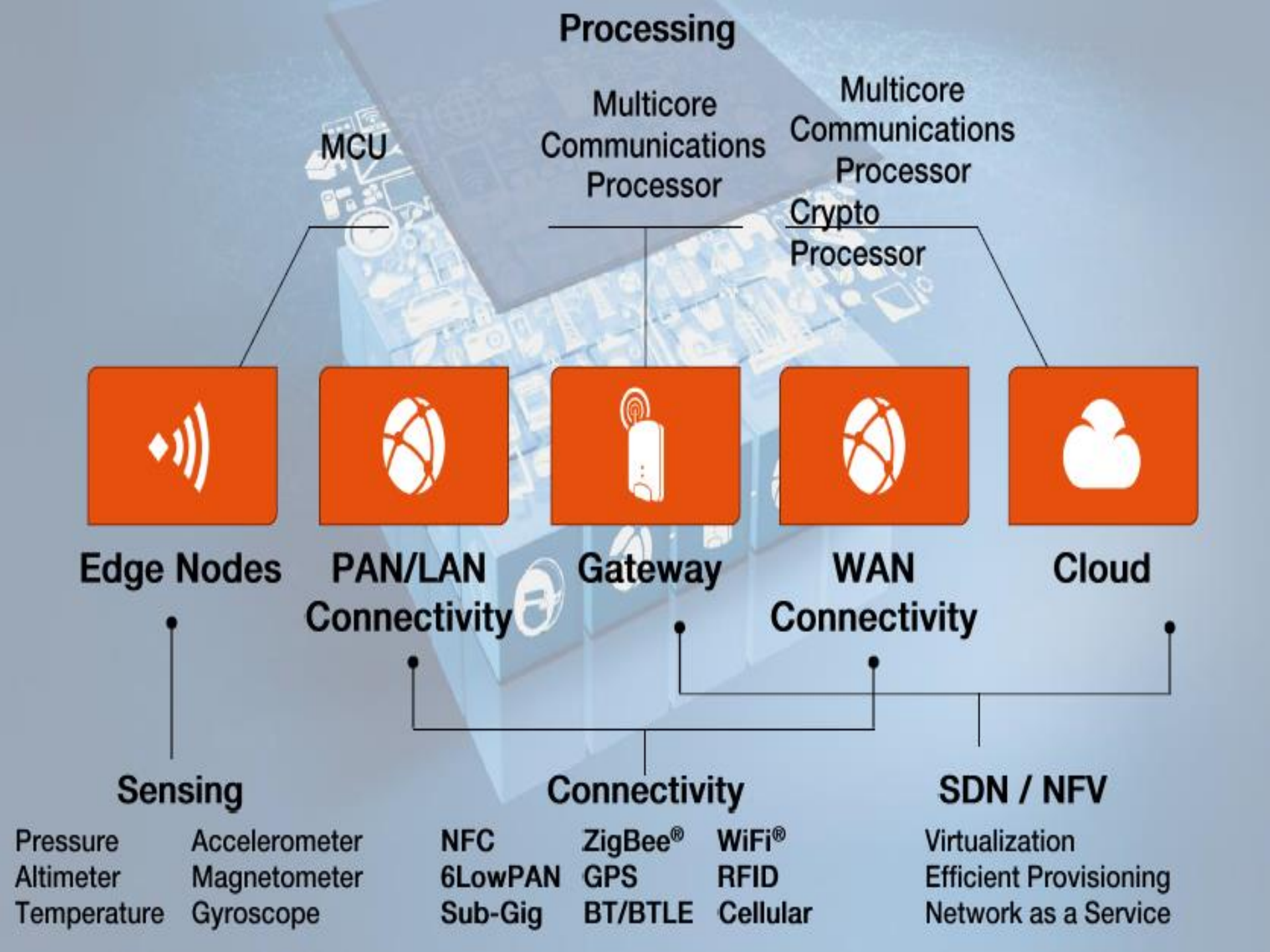


How Much Data?

2018 *This Is What Happens In An Internet Minute*

IOT flow Data > Information > Knowledge and Business Value




Processing
MCU
Multicore Communications Processor
Multicore Communications Processor
Crypto Processor
Edge Nodes
PAN/LAN Connectivity
Gateway
WAN Connectivity
Cloud
Sensing
Pressure
Altimeter
Temperature
Accelerometer
Magnetometer
Gyroscope
Connectivity
NFC
6LowPAN
Sub-Gig
ZigBee®
GPS
BT/BTLE
WiFi®
RFID
Cellular
SDN / NFV
Virtualization
Efficient Provisioning
Network as a Service

Cloud Computing and Big Data

- The main trends:
  - Computers not getting any faster
  - Internet connections getting faster
  - More people connected to the Internet

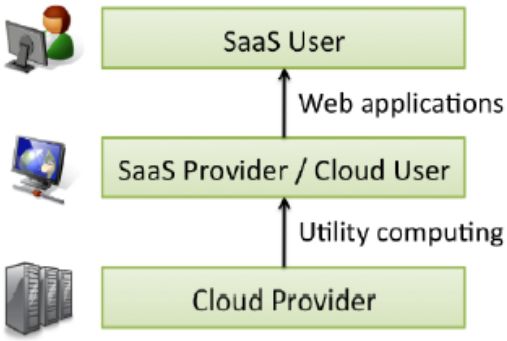
Conclusion

Move the computation and storage of big data to the cloud!

Cisco predicts that by 2020, 92% of IT market workloads will be processed by cloud data centers, while only 8% will be processed by traditional data centers.

Cloud Computing Definition

- Cloud Computing refers to both:
  1. The applications delivered as services over the Internet
  2. The hardware and systems software in the datacenters that provide those services
- The services: called Software as a Service (SaaS)
- The datacenter hardware and software is called cloud



Cloud Characteristics

On-demand self-service

Ubiquitous network access

Location transparent resource pooling

Rapid elasticity

Measured service with pay per use

- A consumer can independently provision computing capabilities without human interaction with the service provider.
  - Available over the network
  - Accessed through mobile phones, laptops, ...

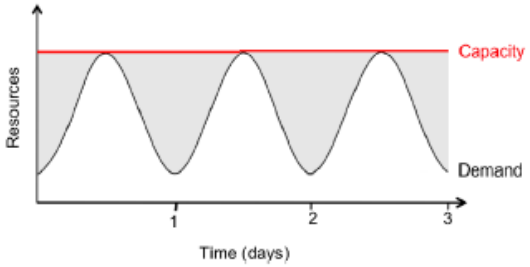
Cloud Characteristics

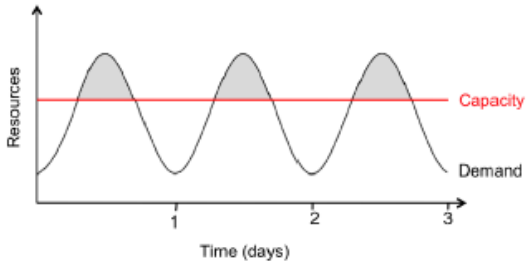
On-demand self-service | Ubiquitous network access | Location transparent resource pooling | Rapid elasticity | Measured service with pay per use

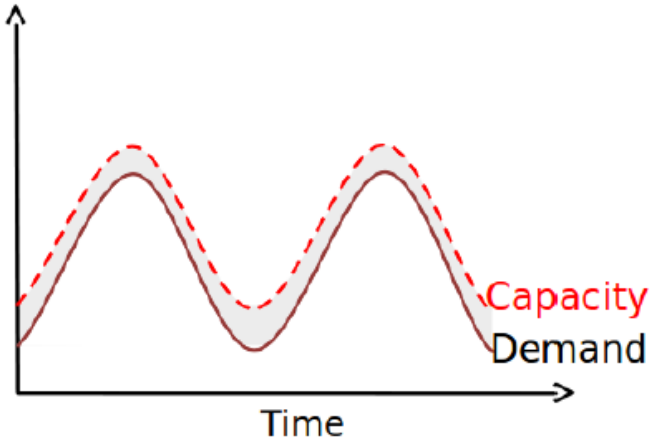
- Provider's computing resources are pooled to serve consumers
- Resource usage can be monitored, controlled, and reported providing transparency for both the provider and consumer.

Cloud Characteristics

- Capabilities can be rapidly and elastically provisioned, in some cases automatically.



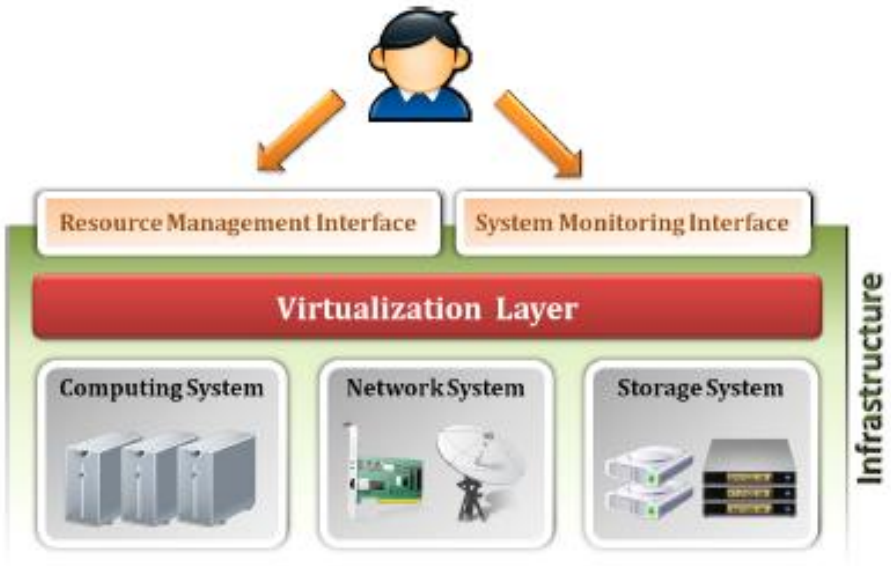




Rapid elasticity

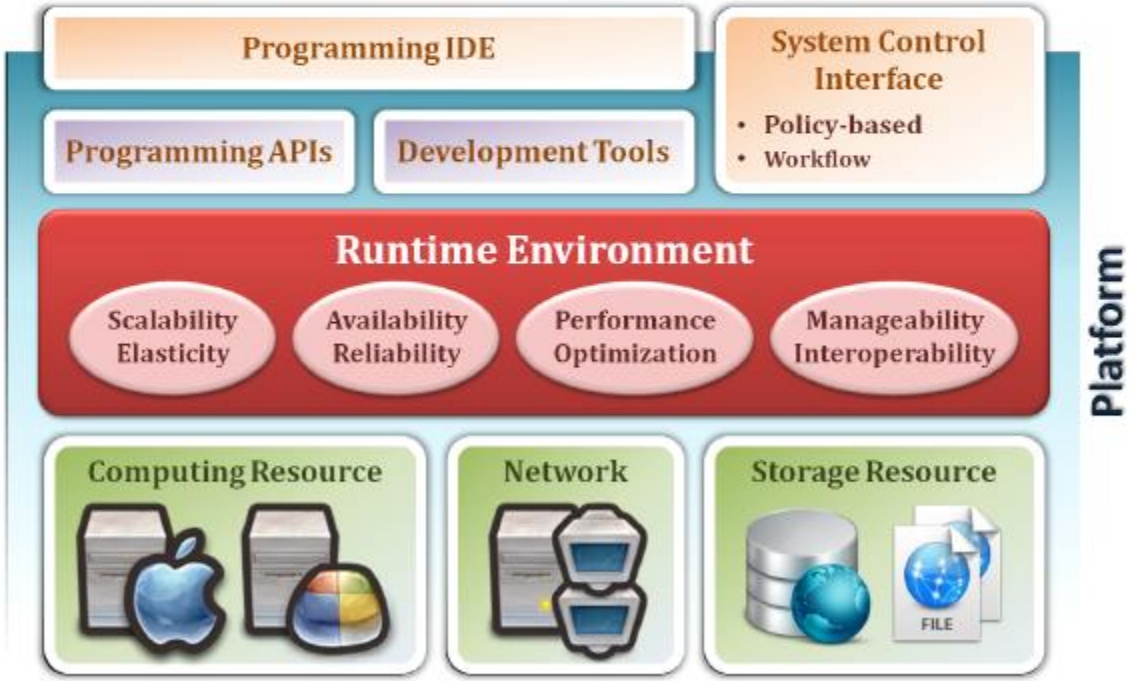
Infrastructure as a Service (IaaS)

- Vendor provides resources, e.g., processing, storage, network, ...
- Consumer is provided customized virtual machines.
- Example: Amazon Web Services (EC2 instances and S3 storage)



Platform as a Service (PaaS)

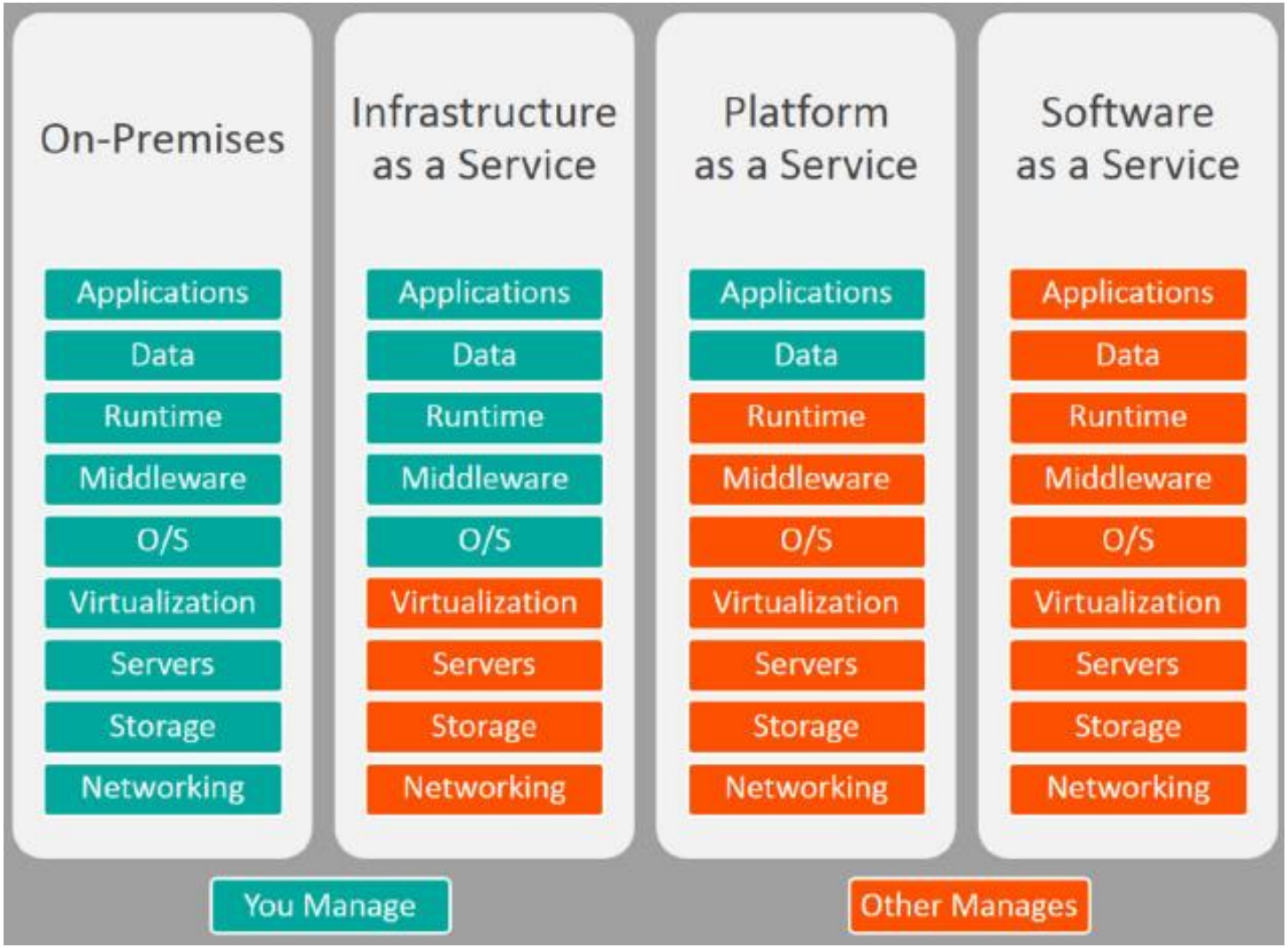
- Vendor provides hardware and development environment.
- Example: Google app engine



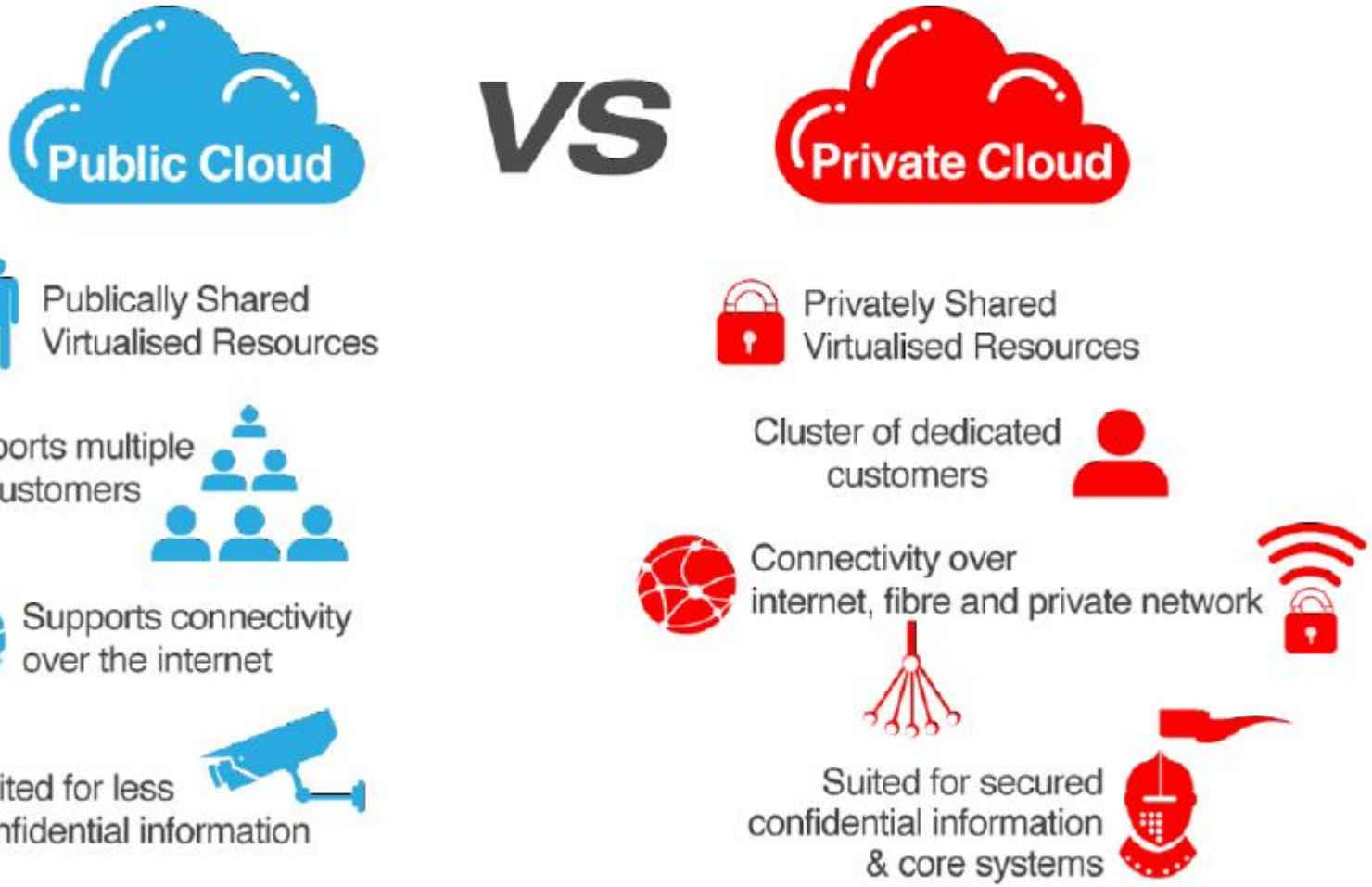
Software as a Service (SaaS)

- Vendor provides applications accessed over the network.
- Example: Gmail, Github

IaaS - PaaS - SaaS



Deployment Models

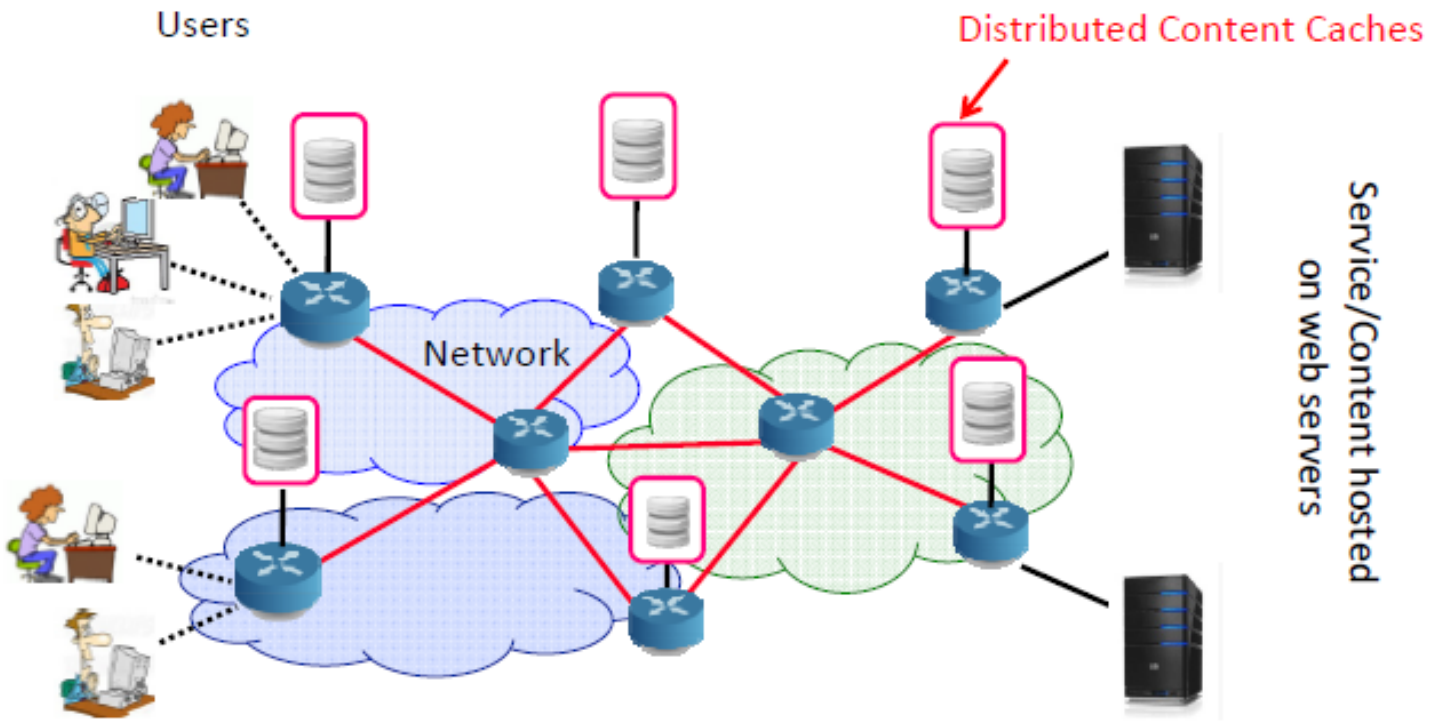


Public Cloud Infrastructure Vendors

- Amazon Web Services (AWS)
- Microsoft Azure
- Google Cloud Platform
- IBM Bluemix
- ...

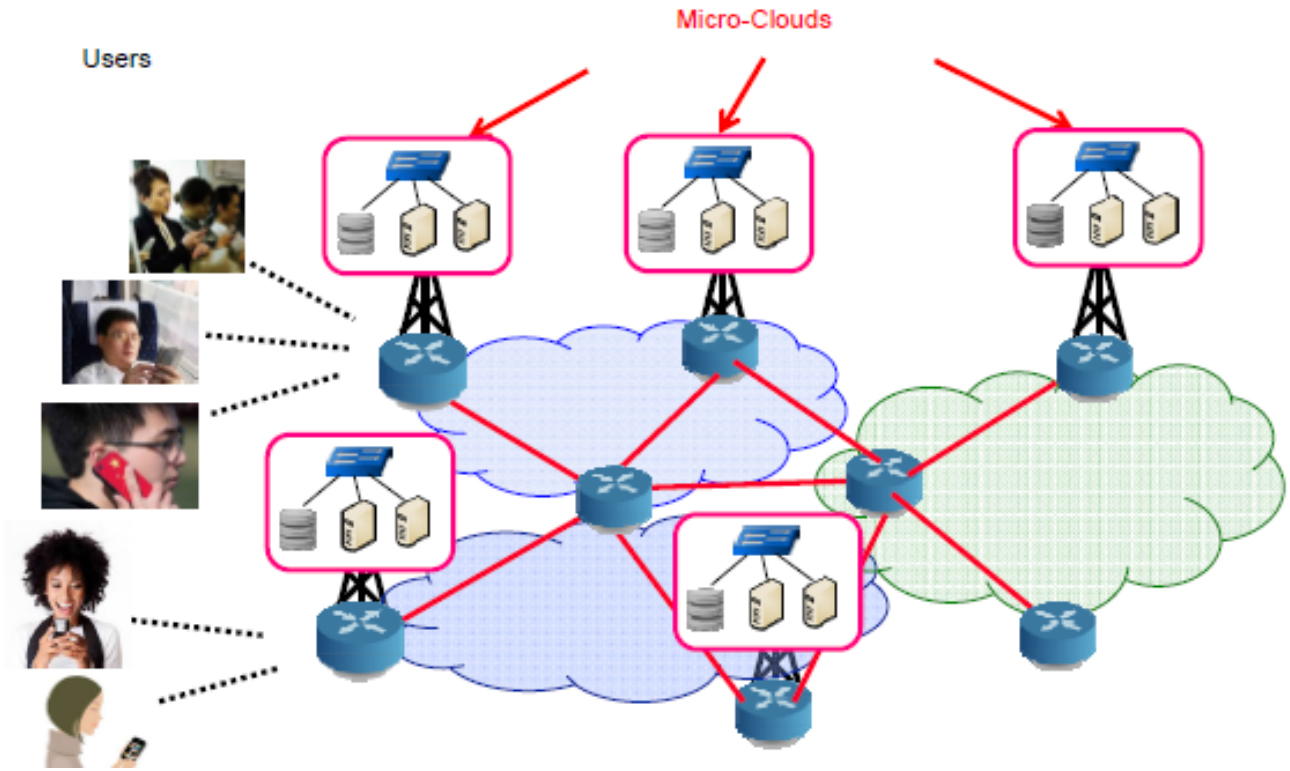
Past: Data in the Edge

- To serve world-wide users, latency was critical and so the data was replicated and brought to edge



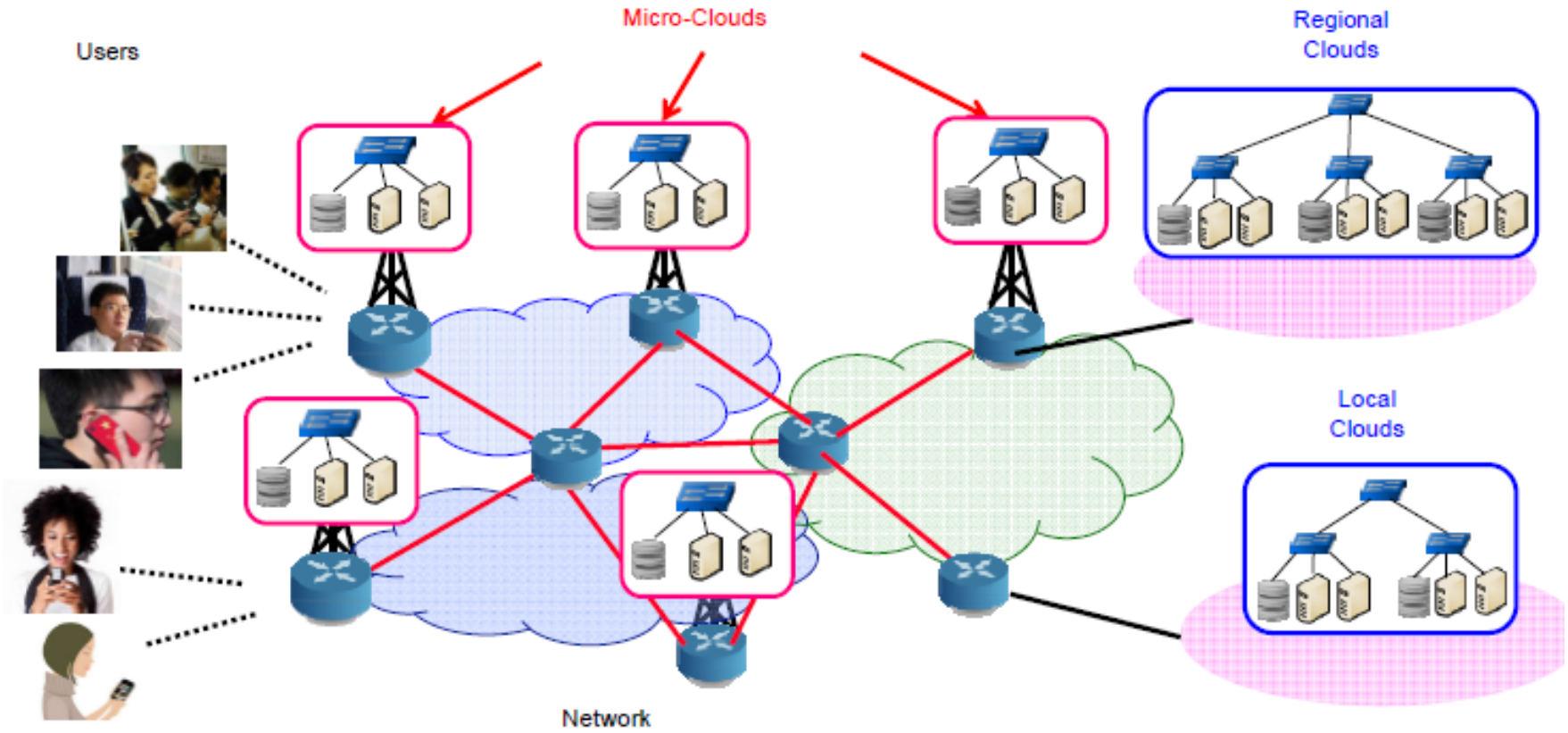
Trend 2: Computation in the Edge

- To service mobile users/IoT, the computation needs to come to edge ⇒ Mobile Edge Computing, Fog Computing



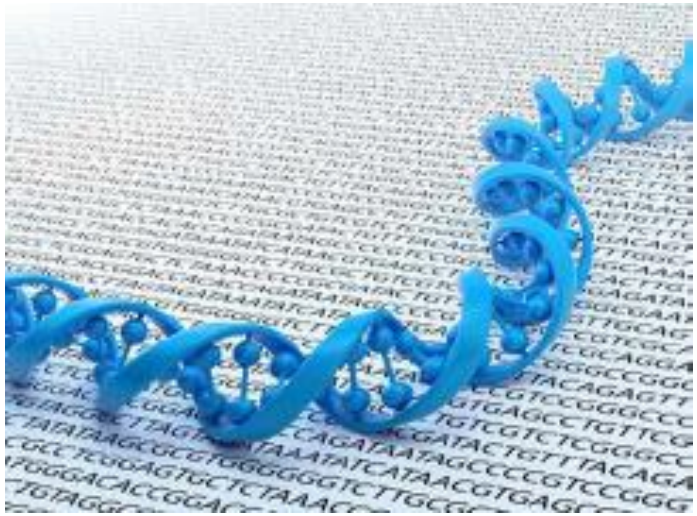
Trend 3: Multi-Cloud

- Larger and infrequent jobs serviced by local and regional clouds ⇒ Fog Computing



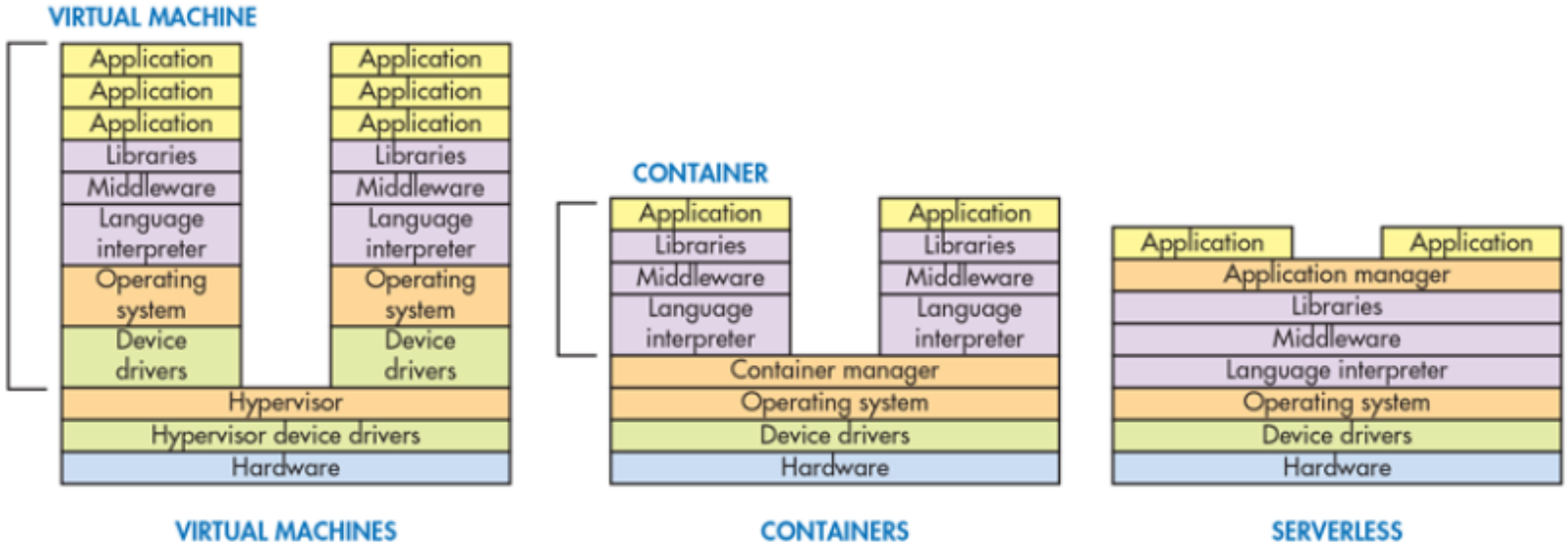
Cloud Computing Main Services

- Computing
- Storage
- Database
- Big data analytics
- ...



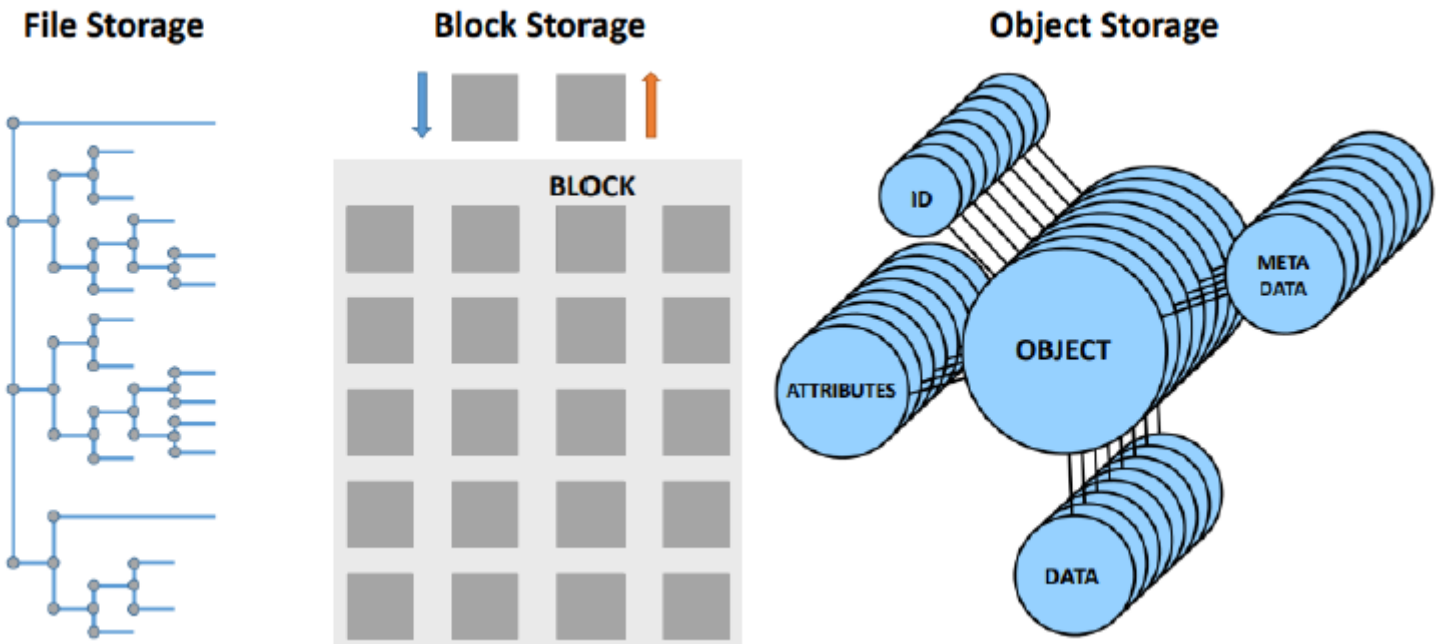
Computing Services

- Virtual machines
- Container services
- Serverless compute



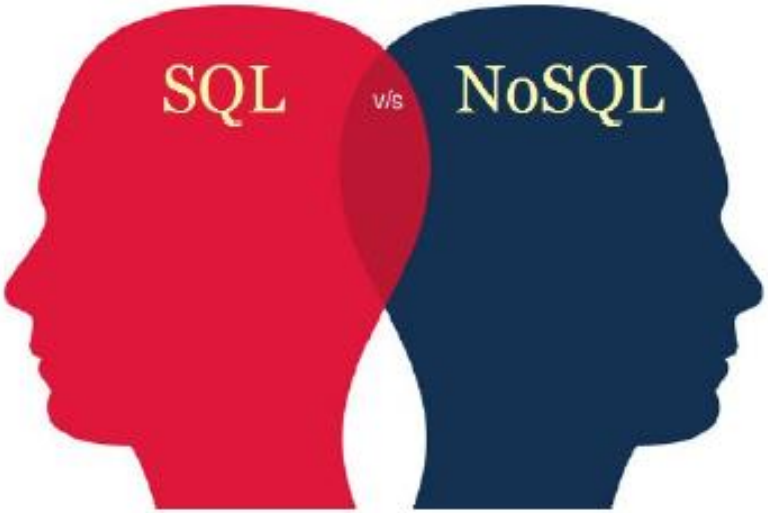
Storage Services

- File storage
- Block storage
- Object storage



Database Services

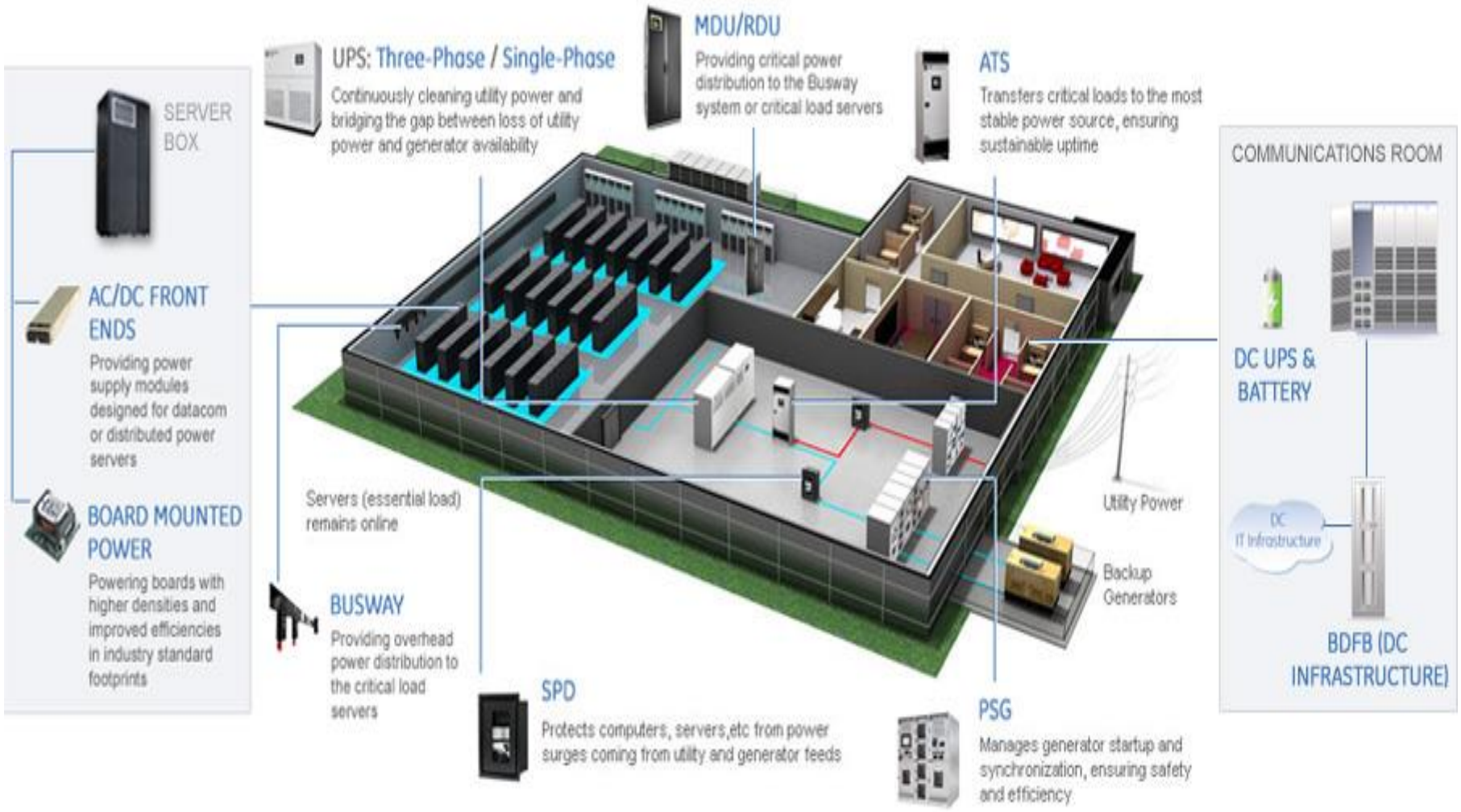
- Relational Database Management Services (RDBMS)
- NoSQL databases
- In-Memory data services



Big Data Analytics

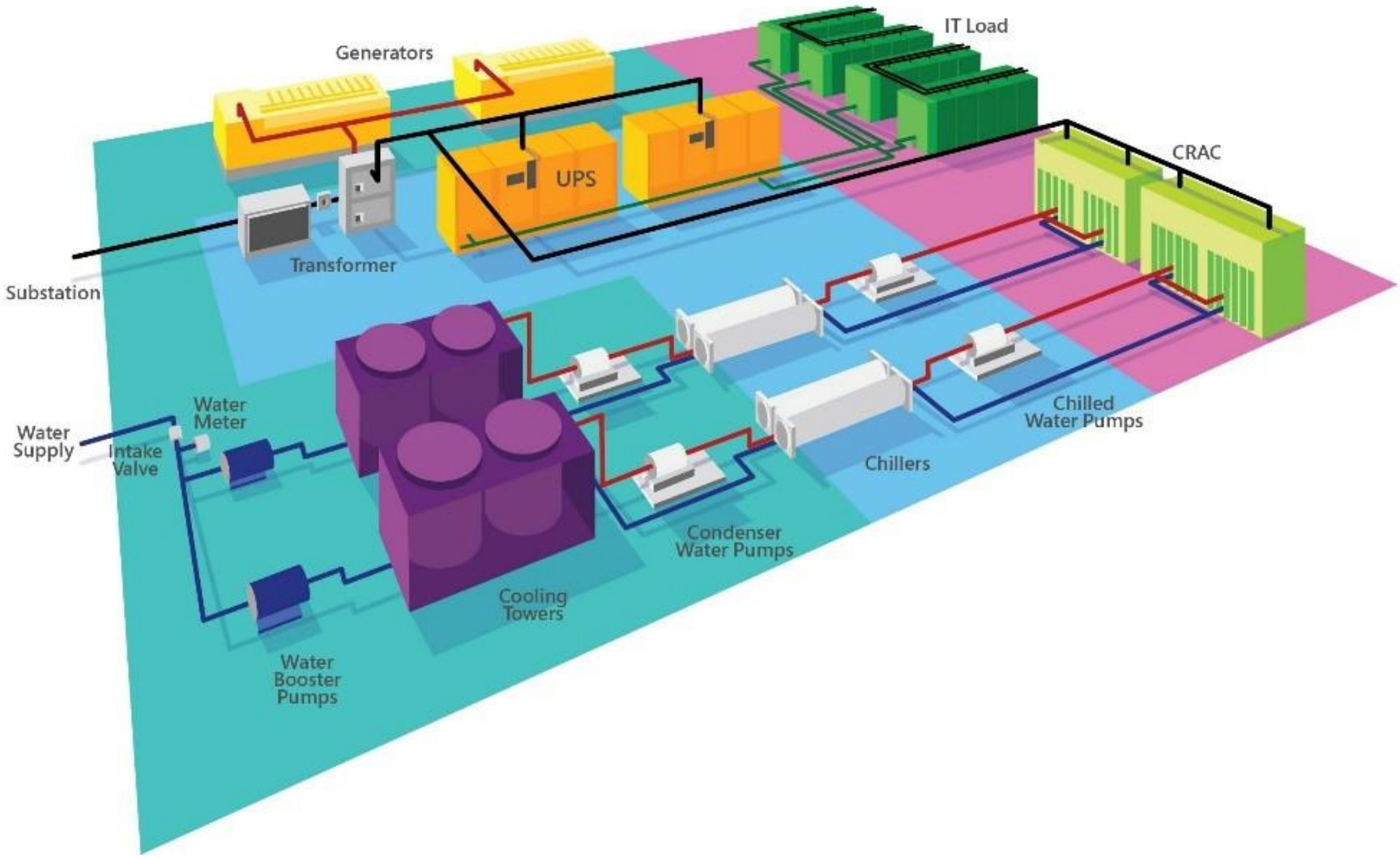
- Big Data Managed Cluster-as-a-Service
- Data warehouse
- Data streaming
- Data queuing

DATA CENTER


SERVER BOX
AC/DC FRONT ENDS
Providing power supply modules designed for datacom or distributed power servers
BOARD MOUNTED POWER
Powering boards with higher densities and improved efficiencies in industry standard footprints
UPS: Three-Phase / Single-Phase
Continuously cleaning utility power and bridging the gap between loss of utility power and generator availability
MDU/RDU
Providing critical power distribution to the Busway system or critical load servers
ATS
Transfers critical loads to the most stable power source, ensuring sustainable uptime
COMMUNICATIONS ROOM
DC UPS & BATTERY
DC IT Infrastructure
BDFB (DC INFRASTRUCTURE)
Servers (essential load) remains online
Utility Power
Backup Generators
BUSWAY
Providing overhead power distribution to the critical load servers
SPD
Protects computers, servers,etc from power surges coming from utility and generator feeds
PSG
Manages generator startup and synchronization, ensuring safety and efficiency

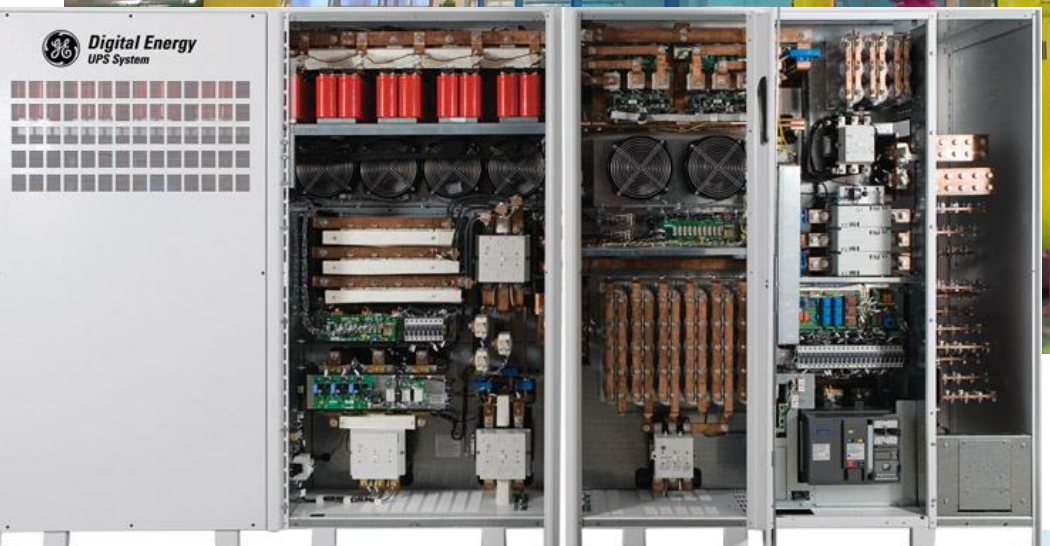
DATA CENTER

UPS& HVAC

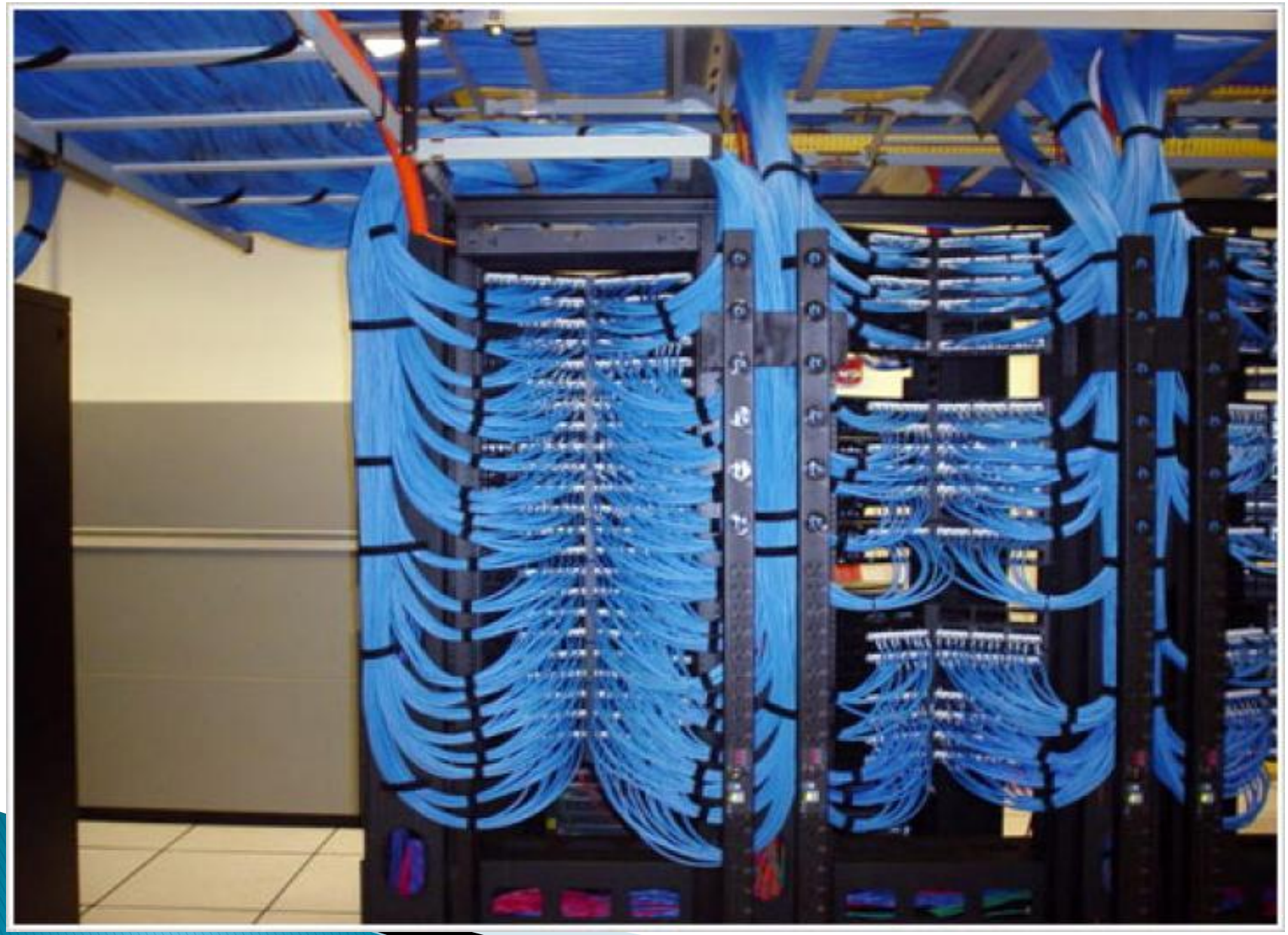

Generators
IT Load
UPS
CRAC
Transformer
Substation
Water Meter
Water Supply
Intake Valve
Chilled Water Pumps
Chillers
Condenser Water Pumps
Cooling Towers
Water Booster Pumps

Computers

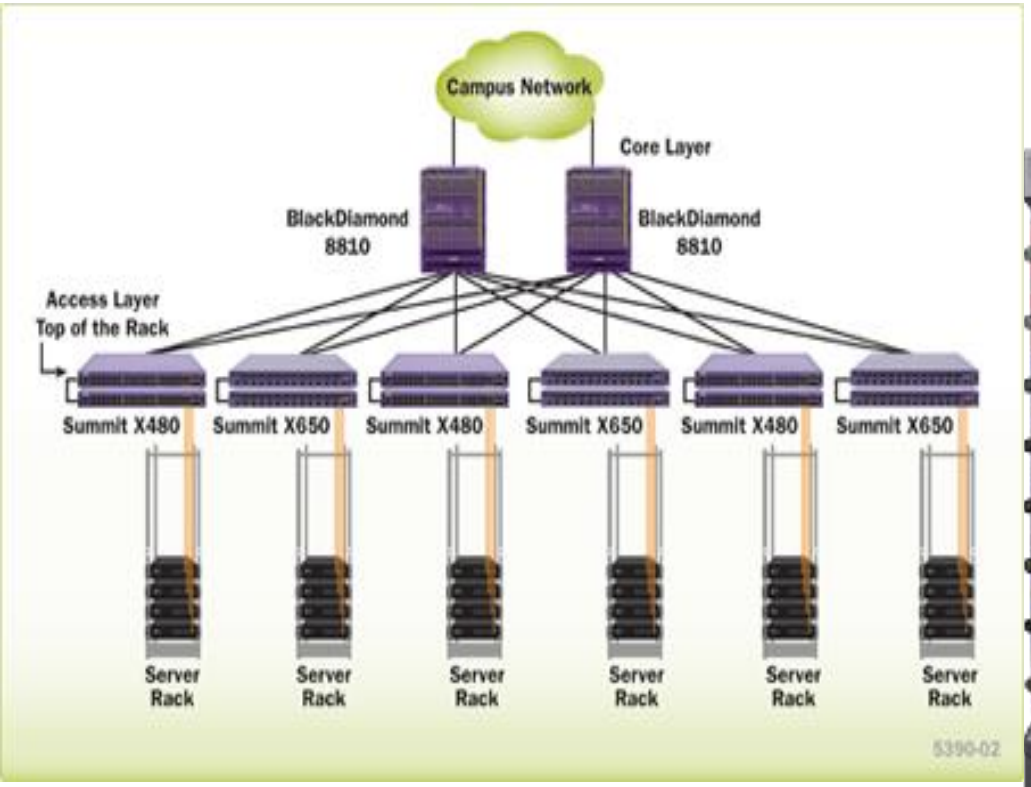
UPS& HVAC



Network



Network


Campus Network
Core Layer
BlackDiamond
8810
BlackDiamond
8810
Access Layer
Top of the Rack
Summit X480
Summit X650
Summit X480
Summit X650
Summit X480
Summit X650
Server
Rack
Server
Rack
Server
Rack
Server
Rack
Server
Rack
Server
Rack
5390-02

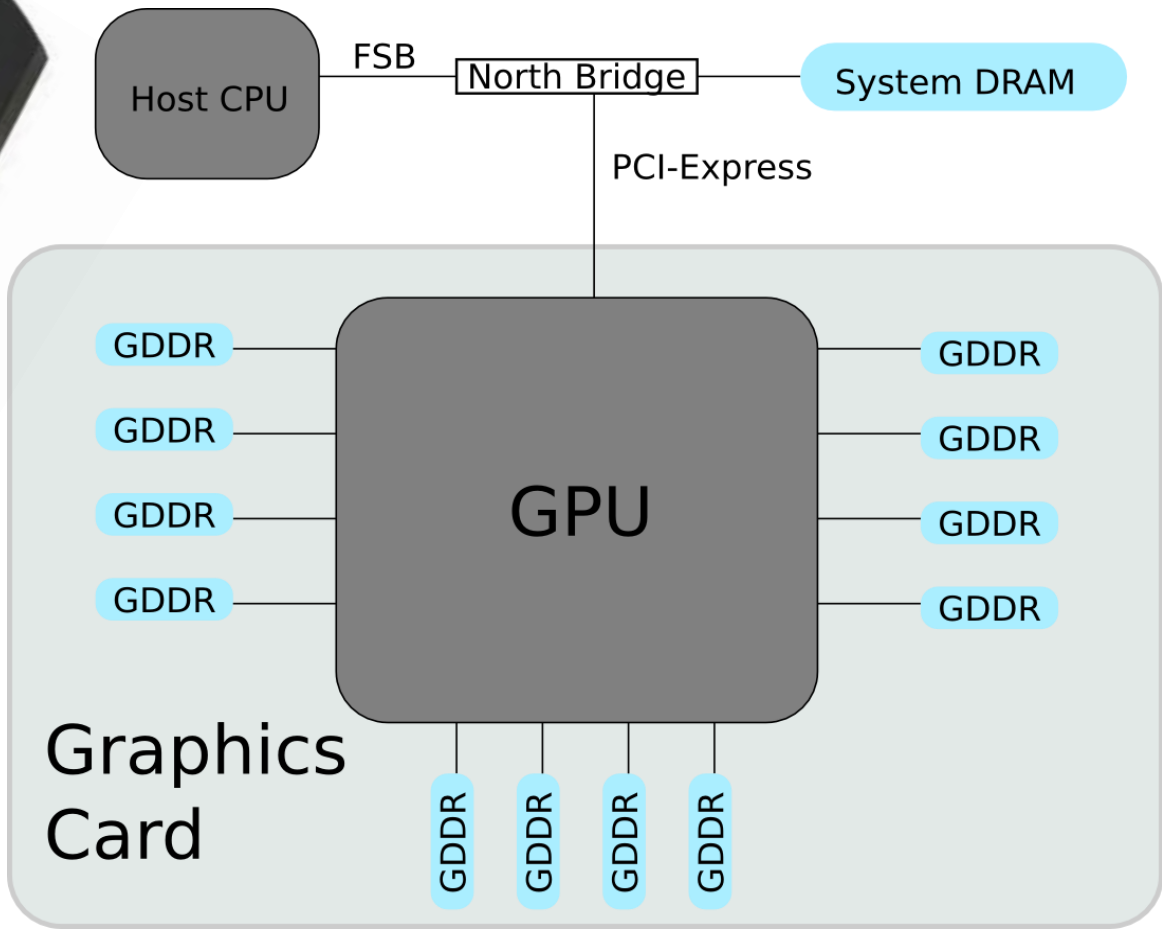
Storage

- ✓ ISCSI
- ✓ FC
- ✓ FCOE

Storage

GPU

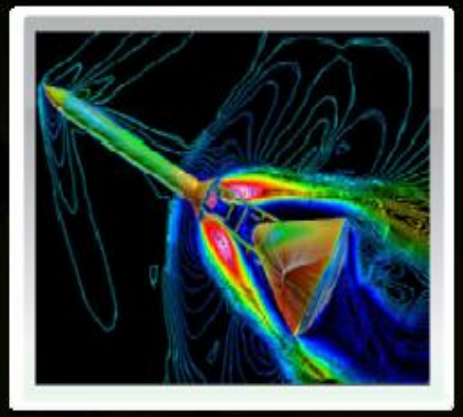

NVIDIA
TESLA


Host CPU
FSB
North Bridge
System DRAM
PCI-Express
GDDR
GDDR
GDDR
GDDR
GPU
GDDR
GDDR
GDDR
GDDR
Graphics Card
GDDR
GDDR
GDDR
GDDR

GPU Family

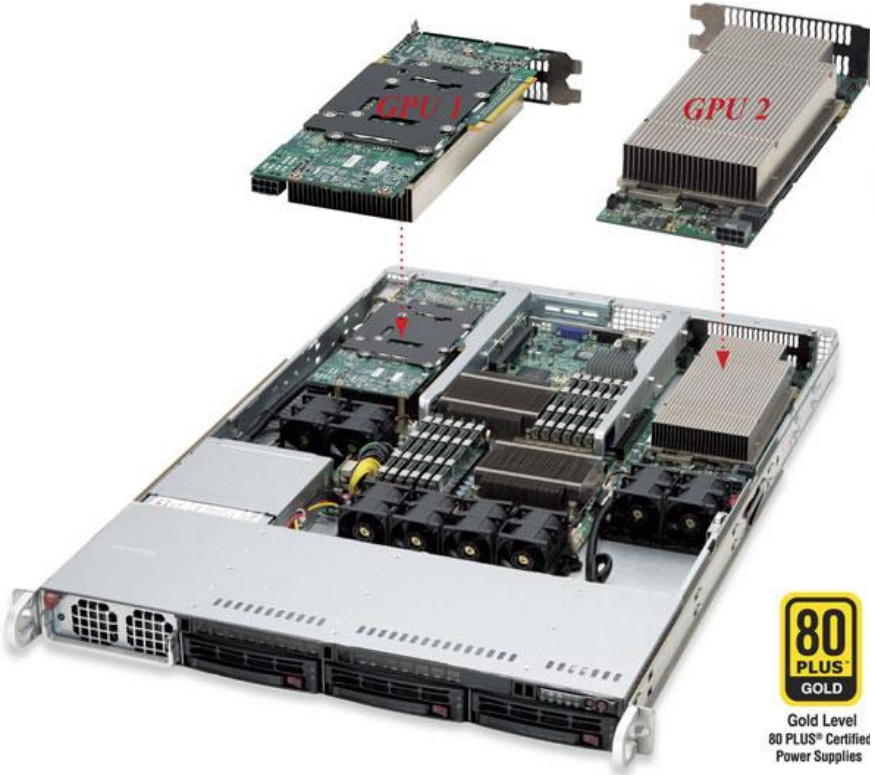
GeForce®
Entertainment

Tesla™
High-Performance Computing

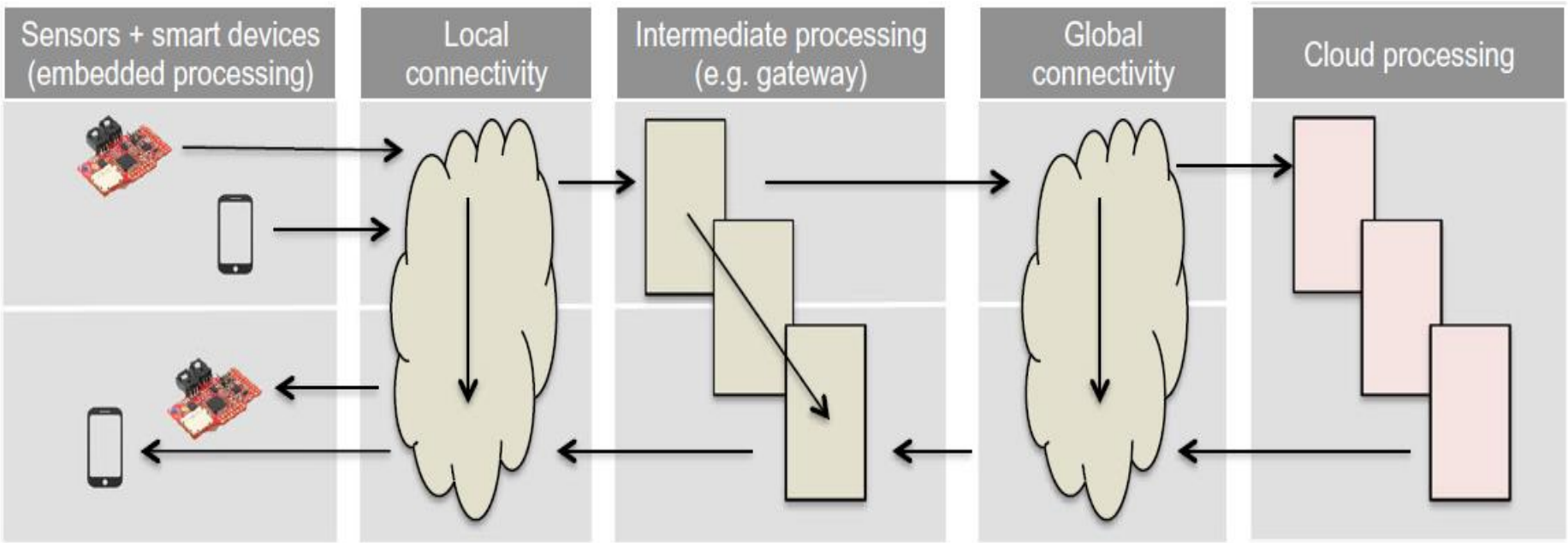


Quadro®
Design & Creation

GPU HW

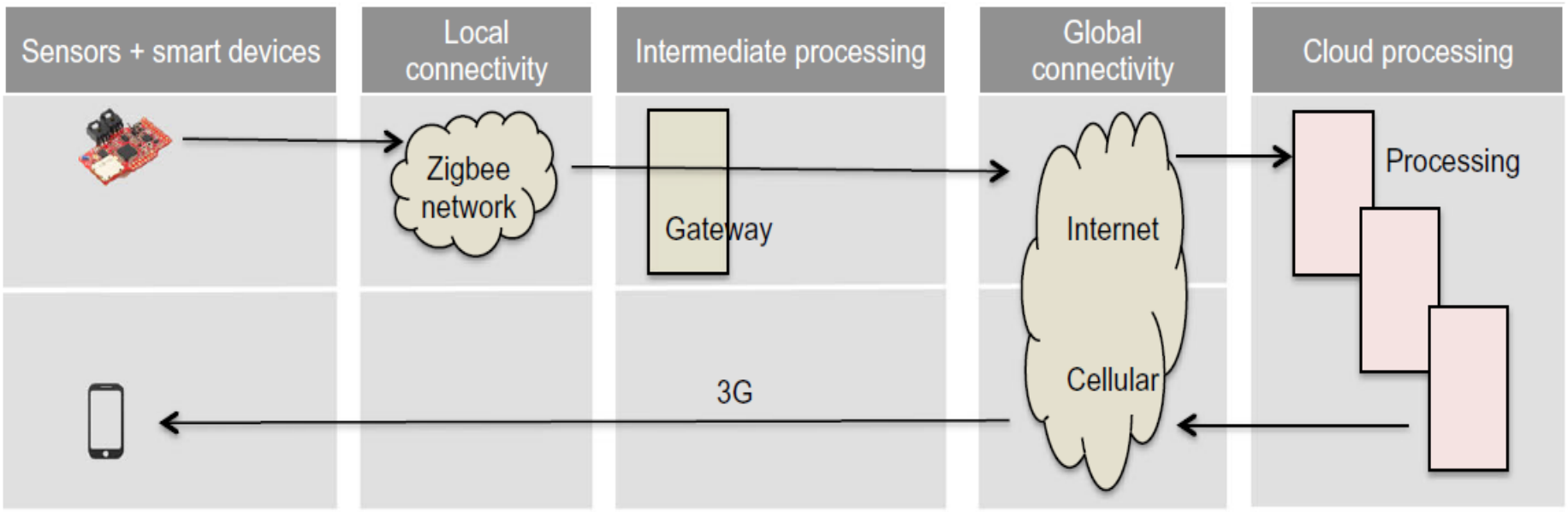

GPU 1
GPU 2
80
PLUS
GOLD
Gold Level
80 PLUS® Certified
Power Supplies

A generic Internet of Things system



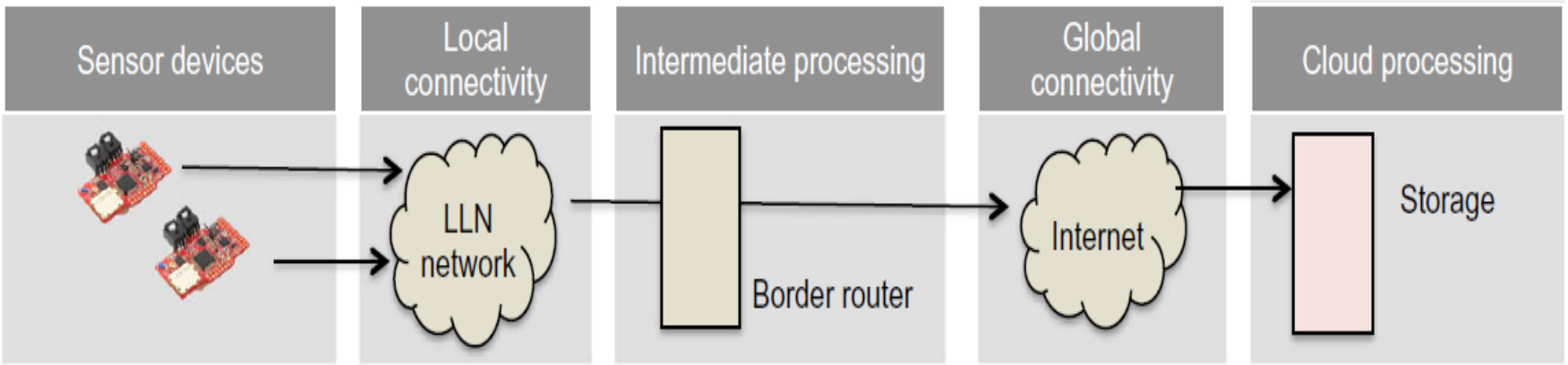
oT scenarios mapped to the generic IoT

- ✓ Zigbee sensor generates an alarm
- ✓ upon which a user needs to be informed.
- ✓ The intelligence resides in the cloud.



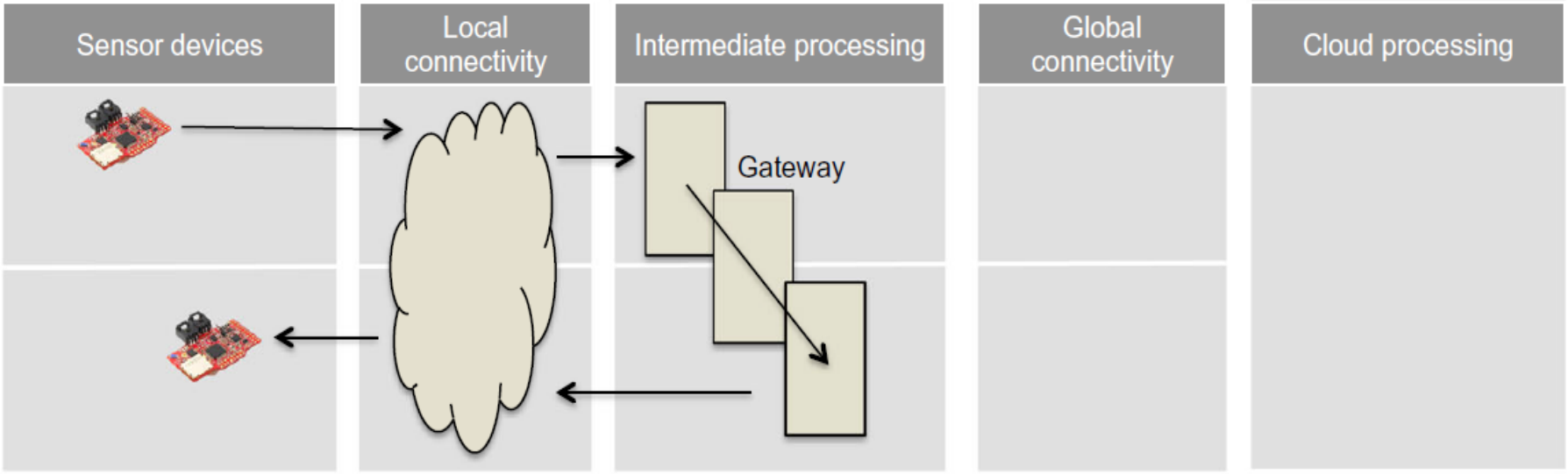
oT scenarios mapped to the generic IoT

- ✓ Sensors in a 6LoWPAN sensor network monitor their environment
- ✓ All measurements are sent to the cloud where they are stored.



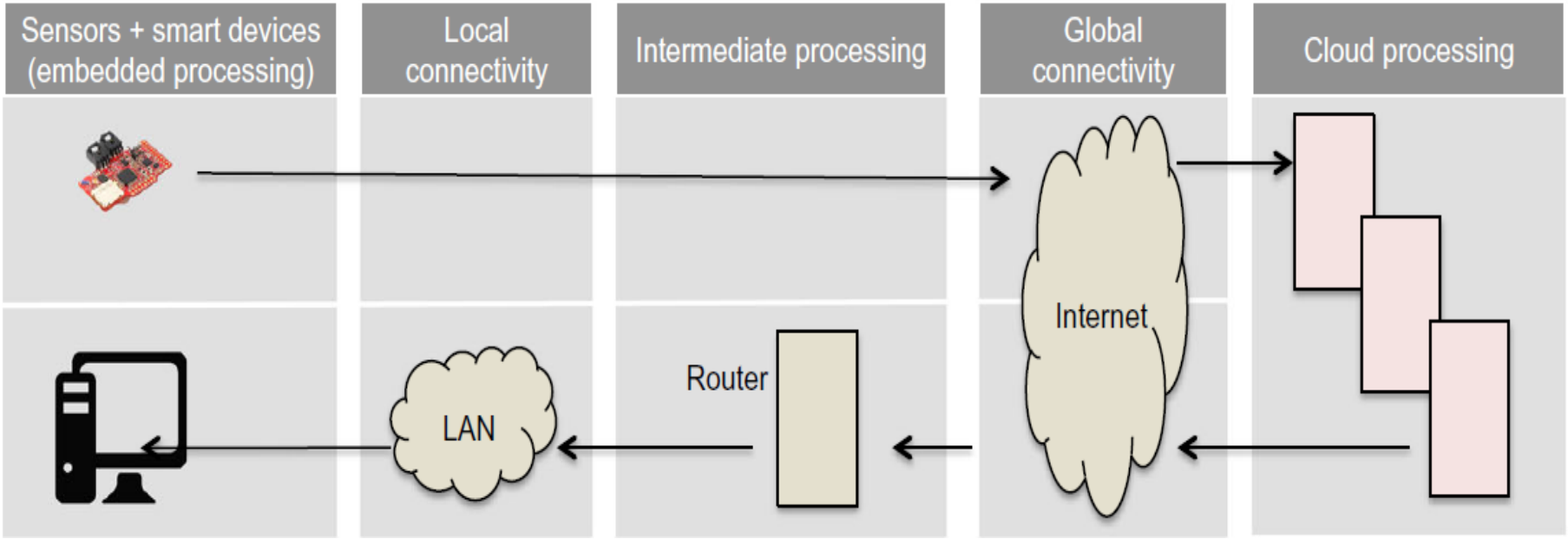
oT scenarios mapped to the generic IoT

- ✓ Temperature sensors in a home automation network measure temperature
- ✓ Based on the measurements, the HVAC system is being triggered.
- ✓ All intelligence resides in the gateway.



oT scenarios mapped to the generic IoT

- ✓ A tracking device with a SIM card communicates its position to a cloud service over a GPRS connection.
- ✓ The information is stored in the cloud and can be consulted by customers.



oT scenarios mapped to the generic IoT

- ✓ User 1 (residing in the same network).
- ✓ user 2 (residing in a remote network).
- ✓ are interacting with a sensor device.