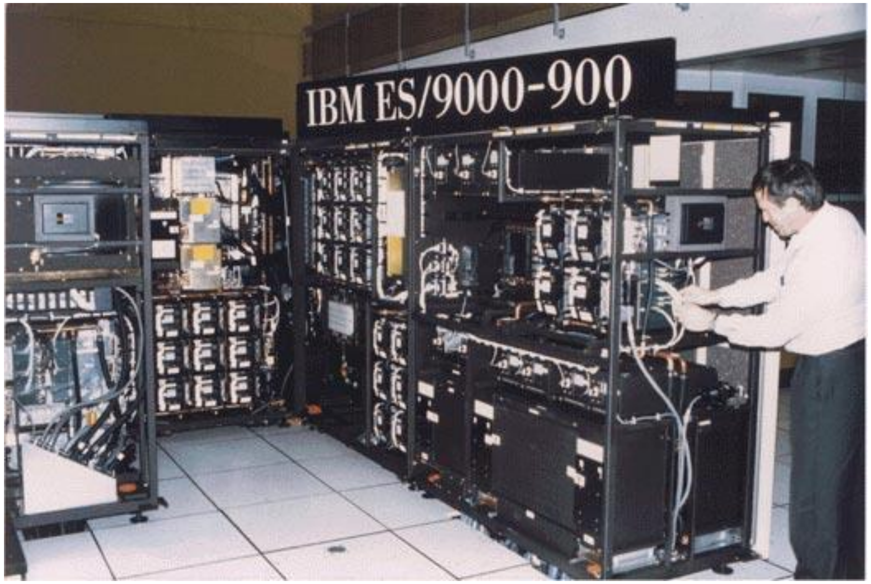
# Computing returns from Cloud
# *Fog, Edge, and Mist Computing*

# It all started with the mainframe




Computing is centralized in the data center

- Systems designed for sharing computing resources
- More or less “dumb” remote terminals

*“I think there is a world market for maybe five computers.”*
*Thomas Watson, president of IBM, 1943*

*"There is no reason anyone would want a computer in their home.”*
*Ken Olsen, founder of Digital Equipment Corporation, 1977*

# It all started with the mainframe
## *Even virtualization*

**Still remember talking about the IBM VM/370?**

- Concurrent execution of multiple production operating systems
- Testing and development of experimental systems
- Adoption of new systems with continued use of legacy systems
- Ability to accommodate applications requiring special-purpose OS


- Introduced notions of *handshake* and *virtual-equals-real mode* to allow the sharing of resource control information with the hypervisor
- Leveraged ability to co-design hardware, VMM, and guest OS

*Want to know more about VM/370 and its predecessors?*
*https://share.confex.com/share/119/webprogram/Handout/Session11255/45%20Years%20of%20Mainframe%20Virtualization%202012-08-05.pdf*

# Computing goes outside the DC…
## *The age of the PC (the 80's and the 90's)*

- In the 80's and the 90's computing power went to the x86 PC.
- Hardware was apparently cheap and accessible to everyone.
- Computing, applications, services and data become fragmented across several nodes (PCs), often poorly coupled (or not coupled at all) and poorly synchronized.

# …and again back to the DC
## *The age of Cloud Computing*

- In the 2000’s the number of servers in the data center was increasing at unsustainable rates, resulting in the urging need for more efficient management of resources (energy, space, O&M, hardware costs…)
- The server hardware becomes more mature and virtualization-friendly.
- Faster networks and distributed computing technologies allowed higher centralization of resources.
- And the right conditions were set for Cloud Computing paradigms and the concentration of resources in the data center.

# Where do we stand today?
## *Time for a new spin?*

**Yes and No!**

- Applications, services and data massively went back to the datacenter
- Cloud computing is widely accepted and disseminated, with clear advantages
- But...

  in some cases, the "Cloud" is a place too far away!

  - Too much latency
  - Too slow network connection
  - Too expensive network connection
  - Too much data to move around
    – Need for more scalable approaches

# A simple example: C-RAN
***C-RAN: Centralized/Cloud Radio Access Network***

**Currently each Base Station includes:**

- The poles and radio antennas
- The Remote Radio Head (RRH) processing
- The Baseband Unit (BBU) which processes low-level communications with the User Equipment (UE)
- This means each Base Station needs a master **and a large container at the base**, full of dedicated computing equipment:
  - Expensive (dedicated, ruggedized to survive harsh environment, need to provide local power, local air conditioning, local security, local maintenance…)
  - Inelastic and inefficient (dimensioned for maximum capacity of that specific cell, which is rarely necessary)

# A simple example: C-RAN

## *C-RAN: Centralized/Cloud Radio Access Network*



*https://www.extremetech.com/extreme/158342-a-rare-look-inside-an-lte-cell-site-operated-by-sprint-in-san-francisco*

# A simple example: C-RAN

## *C-RAN: Centralized/Cloud Radio Access Network*

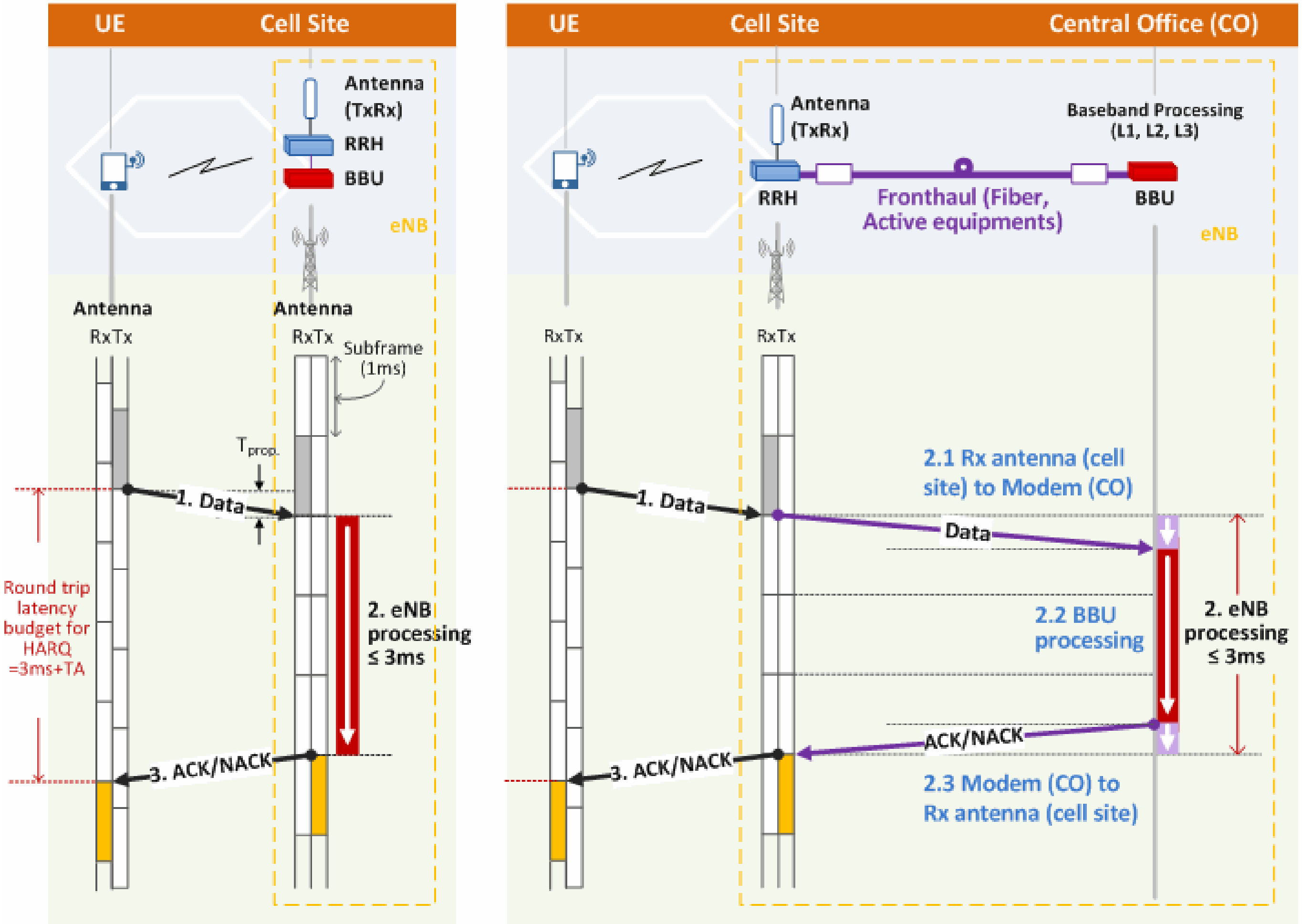
**Why not to "Cloudify" the BBU?**

- It is basically software, which may even run in a VM
- A poll of BBUs shared by multiple Base Stations is more efficient at handling peak loads
- Non-necessary BBUs may be temporarily shutdown, to save energy
- Moving BBUs to data centres enables more efficient cooling and power solutions
- Large container in the base of the pole becomes no longer necessary

- **But…**
  - Let's have a look at LTE BBU's latency requirements

# A simple example: C-RAN

## *C-RAN: Centralized/Cloud Radio Access Network*




(a) Timing requirement of uplink synchronous HARQ processing

(b) Impact of fronthaul delay (expedited BBU processing needed) in C-RAN

$T_p$: Propagation Delay TA: Timing Advance TA=$2T_p$

To secure UL/DL time alignment at a cell site (TX, Rx antenna), BBU allocates a TA value to UE after subtracting the fiber latency from the TA value.

LTE preamble format: 0

Source: https://www.netmanias.com/en/?m=view&id=blog&no=6276&xtag=c-ran-fronthaul-lte&xref=fronthaul-size-calculation-of-maximum-distance-between-rrh-and-bbu

# A simple example: C-RAN

## *C-RAN: Centralized/Cloud Radio Access Network*

| 2.1 Rx antenna (cell site) to Modem (CO) | 2.2 BBU processing (CO) | 2.3 Modem (CO) to Rx antenna(cell site) |
|---|---|---|
| a. RRH/RF processing (UL)<br>b. RRH/CPRI processing (UL)<br>c. Fiber latency (RRH to BBU)<br>d. Active equipment processing | e. BBU/CPRI processing<br>f. PHY: UL frame decoding<br>g. MAC: ACK/NACK creation<br>h. PHY: DL frame creation<br>i. BBU/CPRI processing | j. Fiber latency (BBU to RRH)<br>k. Active equipment processing<br>l. RRH/CPRI processing (DL)<br>m. RRH/RF processing (DL) |

| Delay Components | Related NEs | Description | Typical values |
|---|---|---|---|
| 1. Round trip RF processing time | RRH | a+m | ~ 25-40μsec |
| 2. Round trip CPRI processing time | RRH, BBU | b+e+i+l | ~ 10 μsec |
| 3. BBU round trip baseband processing time | BBU | f+g+h | ~ 2,700 μsec |
| 4. Fronthaul equipments round trip processing delay | Fronthaul equipments | d+k | ~40 μsec (OTN encapsulation)<br>~ few μsec (Non-OTN encapsulation) |

**Maximum Fiber RTT**
= 3msec - (1. round-trip RF processing delay + 2. round-trip CPRI processing delay + 3. BBU round-trip processing time + 4. Fronthaul equipments round trip processing delay)

**Maximum Fiber distance [Km]**
= Maximum Fiber RTT / 10μsec/Km

Example)
Maximum Fiber RTT (Non-OTN encapsulation) = 3msec - (40μsec + 10 μsec + 2,700 μsec + 4 μsec) = 246 μsec
Maximum Fiber distance = 246 μsec / (10μsec/Km) = **24.6 Km**

*Slightly larger distances may be possible with faster BBU processors and adjustments to other parameters, but anyway never above a few tenths of km*
*(see for instance http://www.eurecom.fr/en/publication/4640/download/cm-publi-4640.pdf)*

Source: https://www.netmanias.com/en/?m=view&id=blog&no=6276&xtag=c-ran-fronthaul-lte&xref=fronthaul-size-calculation-of-maximum-distance-between-rrh-and-bbu

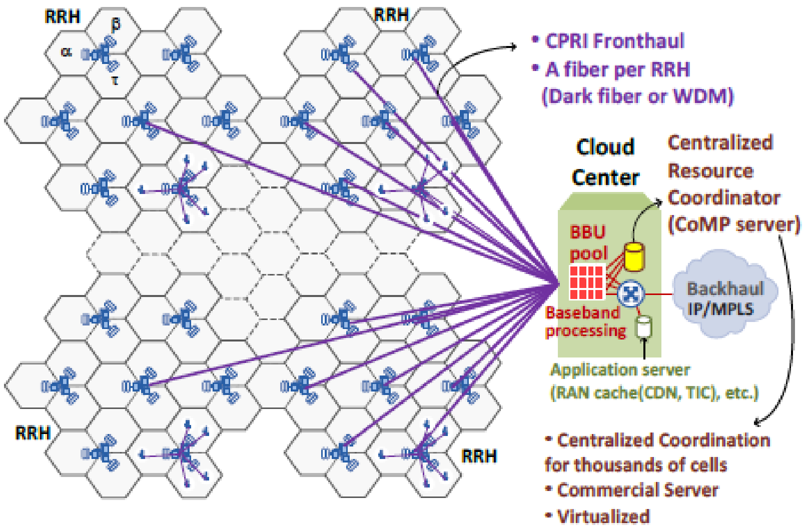
# A simple example: C-RAN

***C-RAN: Centralized/Cloud Radio Access Network***

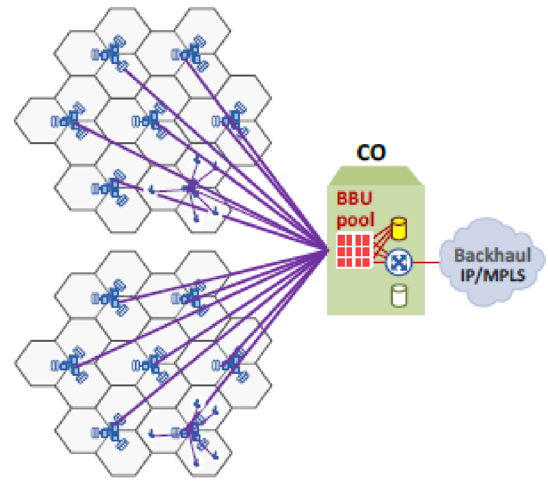
**Solution:**

- Instead of a classic datacenter, consider a smaller regional datacenter able to still bring the advantages of cloud computing but physically closer to the base stations (at the *Edge*).
- BBUs may be either just a shared pool of legacy boxes or some sort of virtual machines (as long as they perform fast enough, for instance using lightweight containers and DPDK for data plane processing).
- In some situations, the RRH may also be concentrated, but with much shorter distances involved and optical fiber connecting the RRH to the antennas.

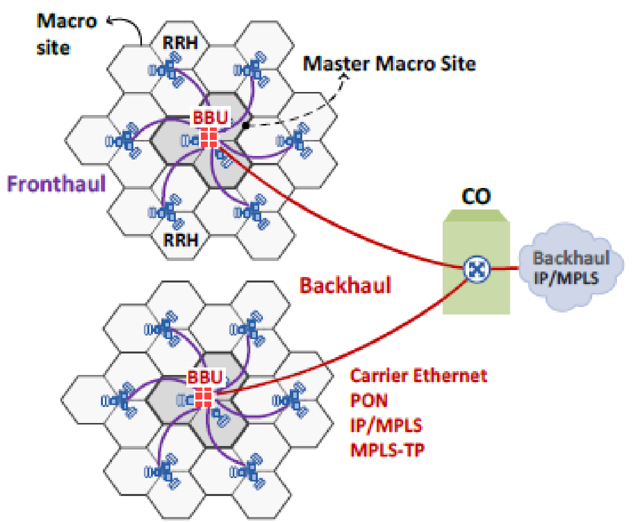
## 1. City-wide C-RAN: Thousands of RRHs (cell)



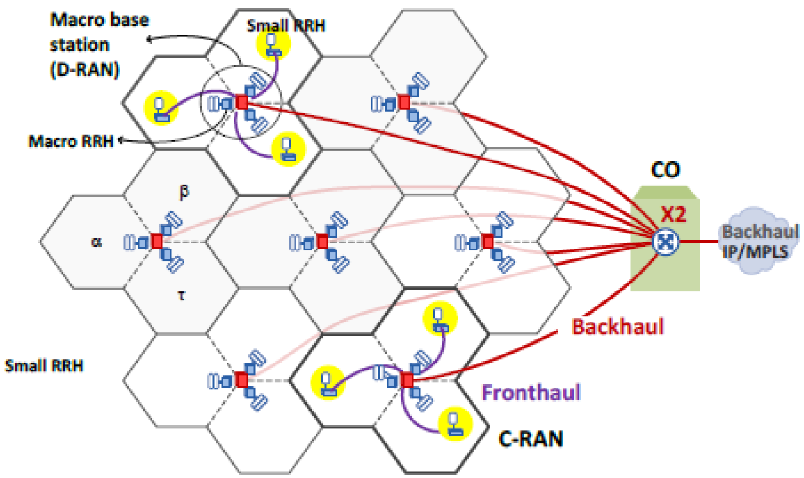

## 2. CO-based C-RAN: Hundreds of RRHs (cell)



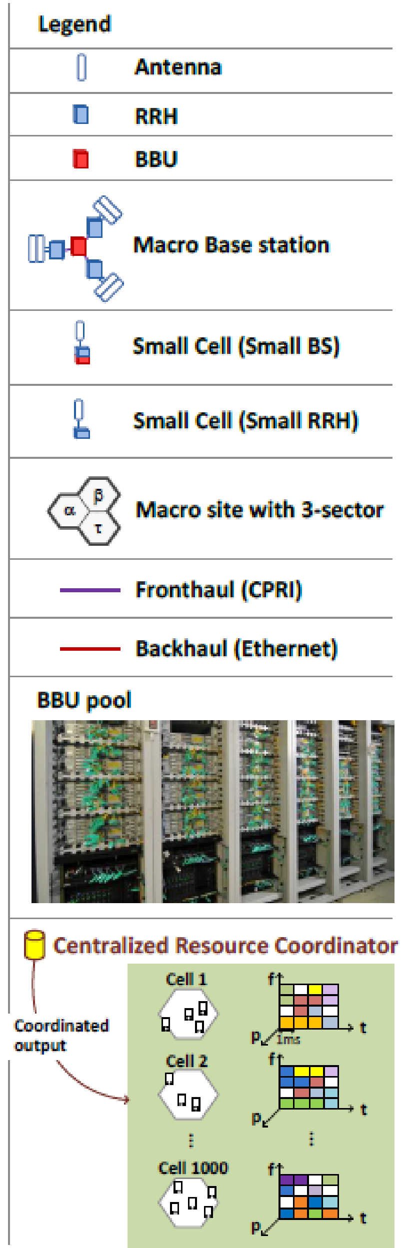

## 3. Master Macro site-based C-RAN: Tens of RRHs (cell)




## 4. Local C-RAN (C-RAN over D-RAN): A few RRHs (cell)

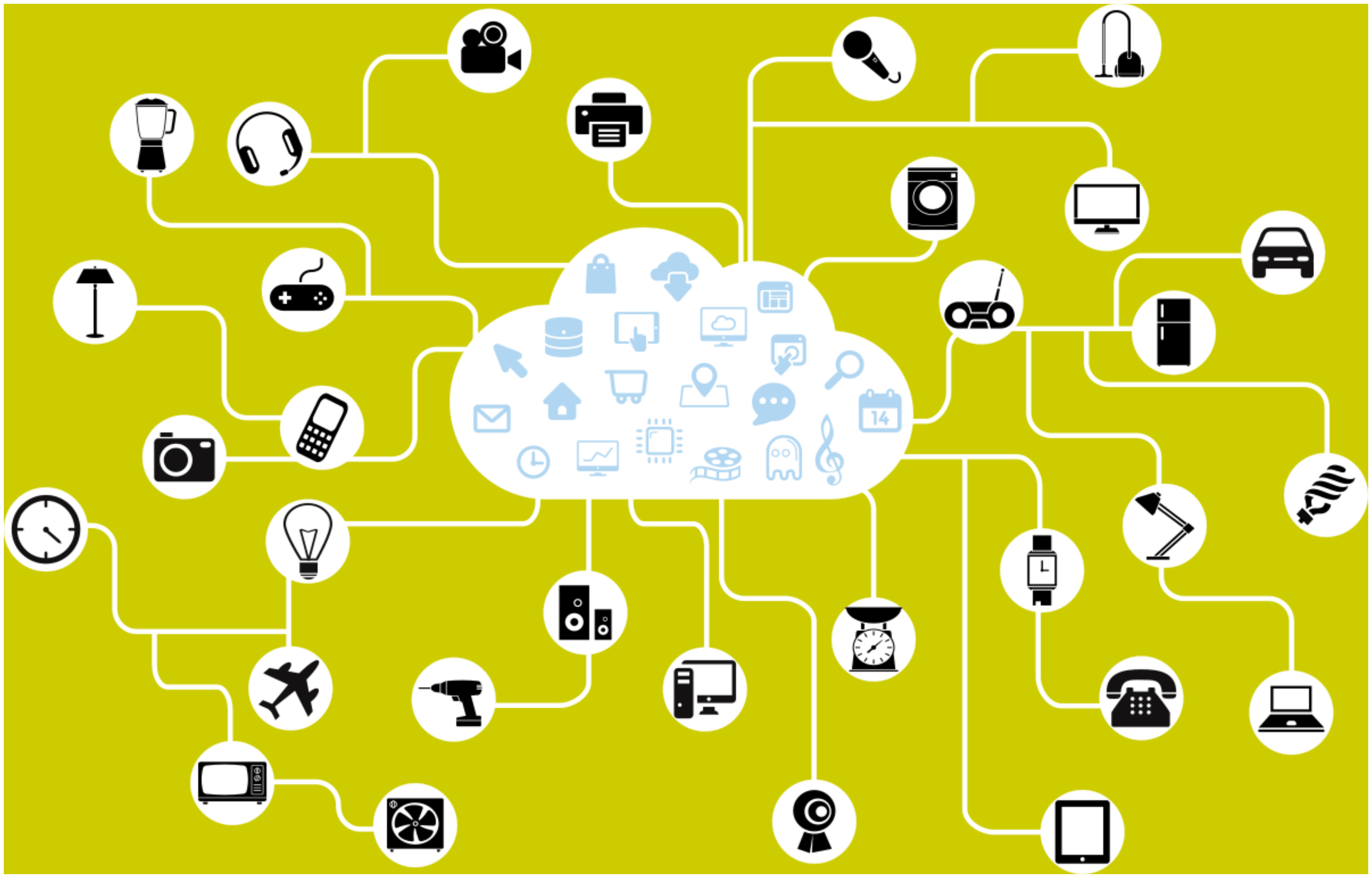



Small RRHs are aggregated via CPRI fronthaul while macro cells are still in D-RAN architecture.



# Another example: IoT

***IoT: Internet of Things*** *(in case you live in Mars!)*



Source: Mozilla

# Another example: IoT

***IoT: Internet of Things*** *(in case you live in Mars!)*

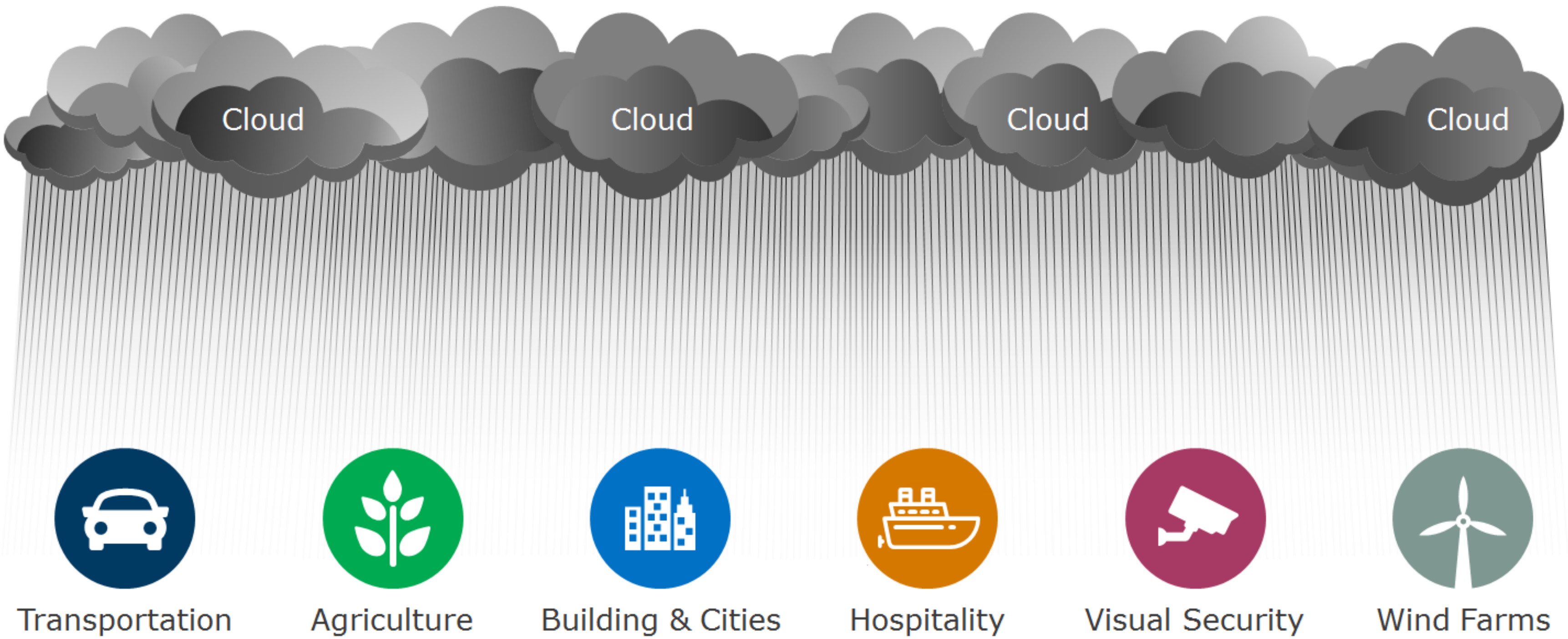
**The rise of the IoT drastically changes various balances:**

- Many IoT devices may capture enormous amounts of sensing data, but processing this data at the devices themselves may be uninteresting or unfeasible (e.g., need to correlate collected data with data from other devices, low capability).
- Moving so much data to centralized data centers may be technically or economically unfeasible.
- Even communication capabilities of these devices are reduced, often requiring local “gateways” for internet access.
- These devices are difficult to structure from an administrative point of view (in which network are they? who owns and manages the device? Who owns the data? …)
  Way more fragmented and dispersed than the old plain internet.
- Some of these devices have strong latency requirements .

# Another example: IoT

***IoT: Internet of Things*** *(in case you live in Mars!)*

We can't afford to have the "Cloud" and the "Thing" without anything in the middle (AKA: *unfettered cloud computing*, according to the OpenFog Consortium).




Source: OpenFog Consortium

# Another example: IoT

***Example: the scenario for smart car and traffic control***

- In 2016 the average person created around 650MB of data every day and by 2020, some project that to more than double.

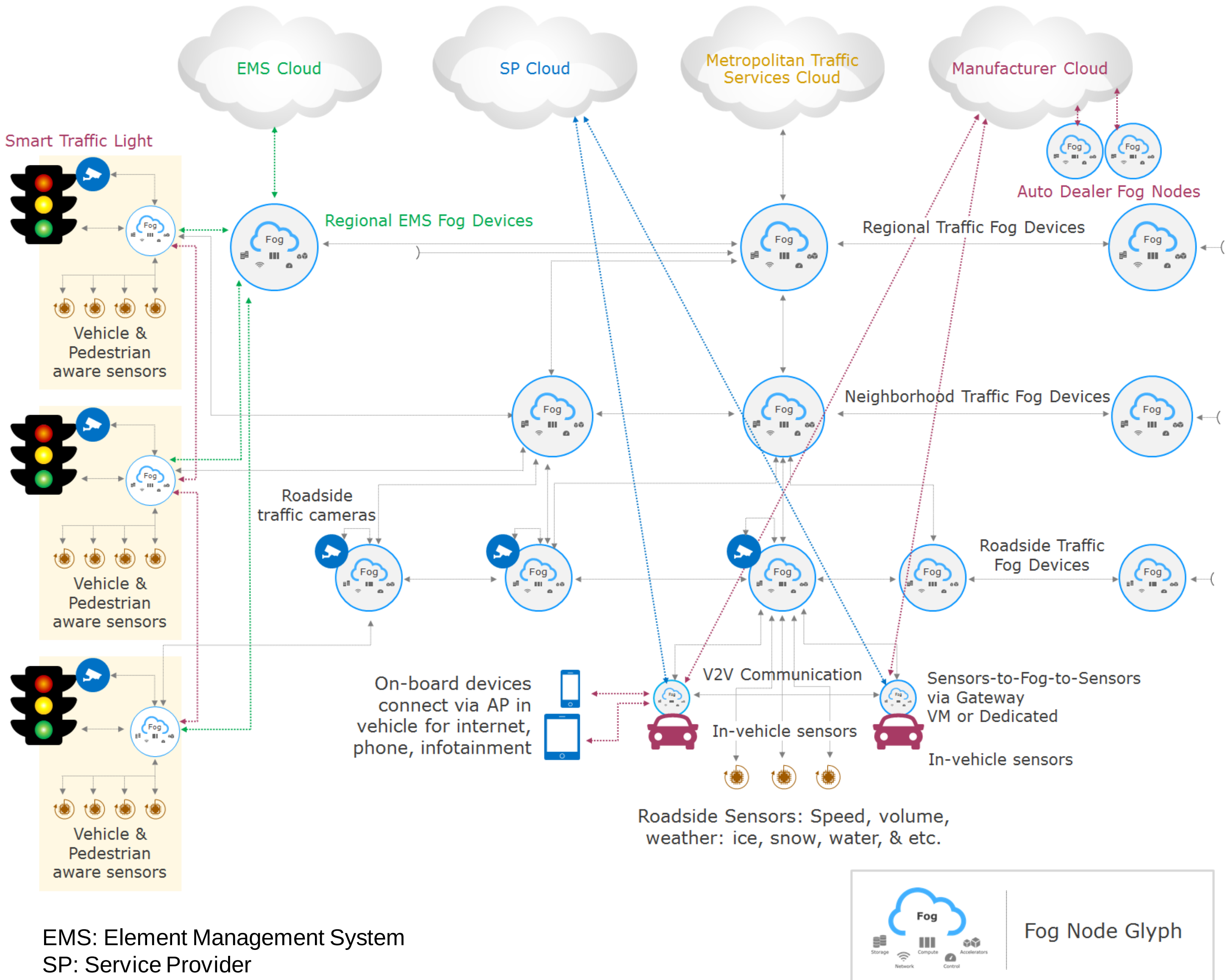
- Smart autonomous cars will further stress the scale, generating multiple terabytes of data every day from the combinations of light detection and ranging (LIDAR), global positioning systems (GPS), cameras, etc.

- When the smart car is coupled with intelligent infrastructure, it is clear that a cloud-only model will not work for autonomous transportation, and that another approach is required

Source: OpenFog Consortium

# Another example: IoT

***Example: the scenario for smart car and traffic control***



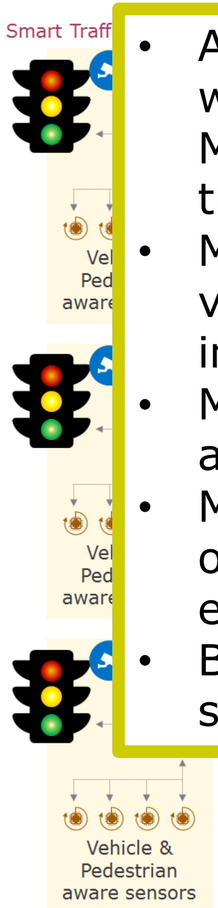

Source: OpenFog Consortium

EMS: Element Management System
SP: Service Provider

# Another example: IoT

***Example: the scenario for smart car and traffic control***


- A rich set of interactions among multiple fog domains, as well as multiple cloud domains, including Element Management Systems (EMS), service provider (SP), metro traffic services, and system manufacturer clouds.
- Mobile fog nodes supporting vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), and vehicle-to-x (V2X) interactions.
- Multiple fog networks owned and operated by different authorities providing similar (and different) functionality.
- Multi-tenancy across fog nodes will also be a burgeoning opportunity to consolidate multiple fog networks, improving efficiency.
- Both private and public fog and cloud networks used by a single end-point device.






EMS: Element Management System
SP: Service Provider

Source: OpenFog Consortium

# Can we bring the cloud closer?
## *Fog Computing*

- Originally coined by Cisco, around 2005,
- Several different (but much essentially similar) definitions

From NIST:

“Fog computing is a horizontal, **physical or virtual resource paradigm that resides between smart end-devices and traditional cloud or data centers**.

This paradigm supports **vertically-isolated, latency-sensitive applications by providing ubiquitous, scalable, layered, federated, and distributed** computing, storage, and network connectivity.”

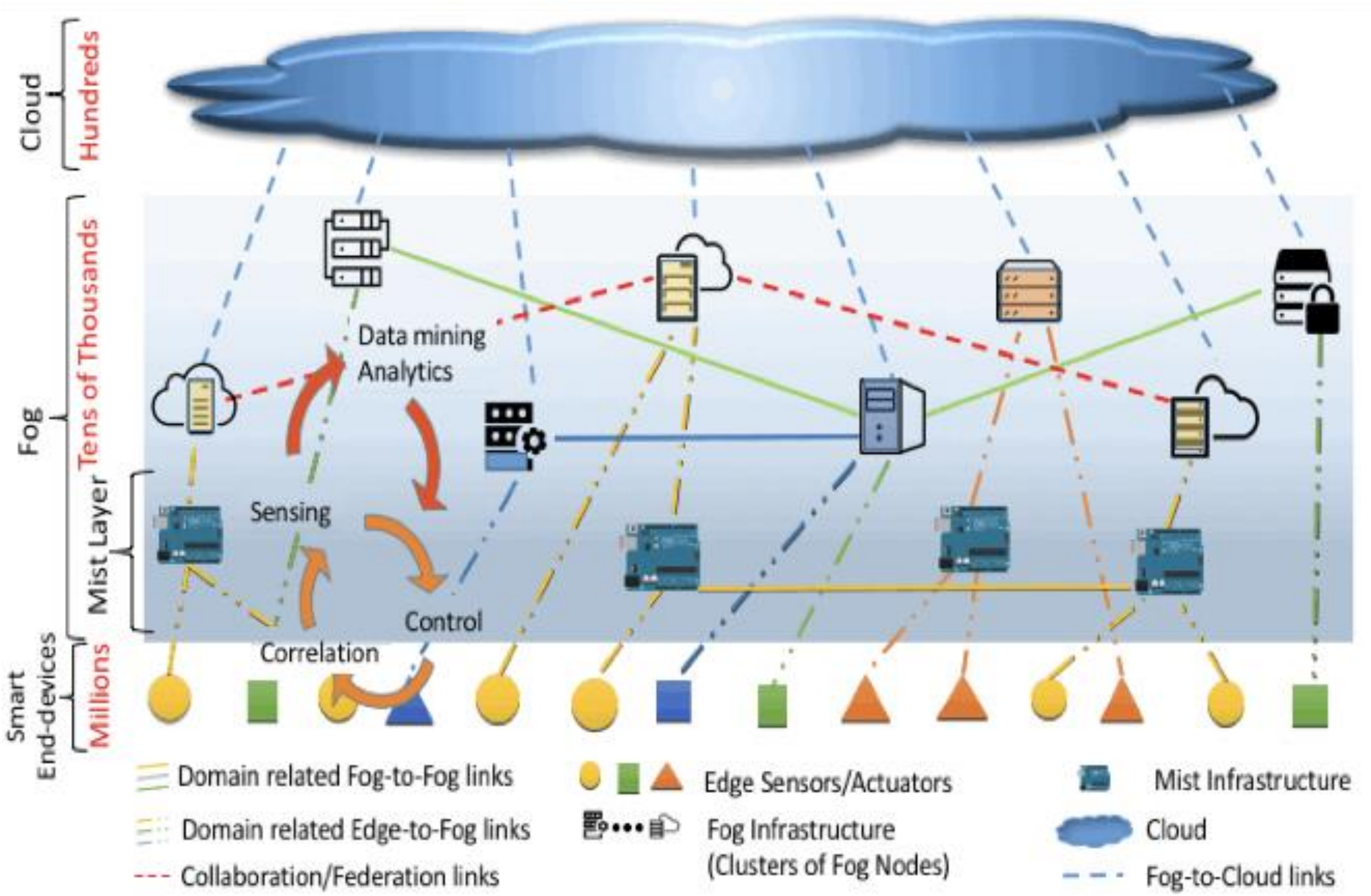
https://csrc.nist.gov/csrc/media/publications/sp/800-191/draft/documents/sp800-191-draft.pdf

From the OpenFog Consortium:

“A horizontal, system-level architecture that **distributes computing, storage, control and networking functions closer to the users along a cloud-to-thing continuum**.”

https://www.openfogconsortium.org

# Can we bring the cloud closer?
## *Fog Computing*




https://csrc.nist.gov/csrc/media/publications/sp/800-191/draft/documents/sp800-191-draft.pdf

# Key characteristics of Fog Computing *(according to NIST)*

- Contextual location awareness, and low latency
- Wide geographical distribution
- Large-scale sensor networks (e.g., IoT)
- Very large number of nodes
- Support for mobility (e.g., C-RAN)
- Real-time interactions
- Predominance of wireless access (e.g., IoT devices)
- Heterogeneity
- Interoperability and federation (multiple devices and service providers)
- Support for real-time big data analytics and interplay with the Cloud

*(not necessarily all coexist at each scenario)*

In general Fog Computing is expected to preserve the benefits of Cloud, including containerization, virtualization, orchestration, manageability and efficiency – although at some points of the ecosystem some concessions may be acceptable.

# The Fog Node
## *(according to NIST)*

“Fog nodes are **intermediary compute elements** of the smart end-devices access network that are situated between the Cloud and the smart end-devices. **Fog nodes may be either *physical* or *virtual* elements and are tightly coupled with the smart end-devices or access networks**. Fog nodes typically provide some form of data management and communication service between the peripheral layer where smart end-devices reside and the Cloud.
Fog nodes, especially virtual ones, also referred as *cloudlets*, can be federated to provide horizontal expansion of the functionality over disperse geolocations.”

- Fog nodes may range from a full-blown “mini-data center” with physical servers and virtual machines similar to those of the cloud (e.g., C-RAN) to an industrial PC running in a smartgrid power substation or even a lightweight execution space inside a home router.

- Fog nodes in general have the same type of service types (PaaS, SaaS, IaaS) and deployment models (public, private, hybrid) of the Cloud.

# Fog vs. Mist Computing

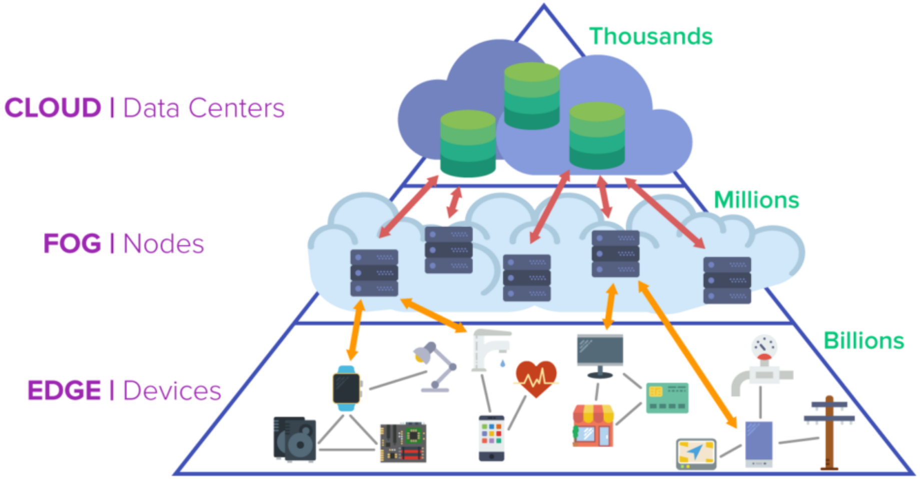
- When moving from the Cloud to the end devices the optimal density of computing units decreases…
  - Full-blown virtual machines
  - Containers
  - Unikernels, serverless architectures, and other experimental paths (with virtualization), or OSGi (no virtualization).

- Fog computing is still not enough for some scenarios, and we need lighter approaches, closer to the Edge. Mist computing generally refers to these lighter approaches, and is still very far from standardization.

- Density is probably the main differentiator between Fog and Mist computing, but very often the timescale of Mist “instances” is also shorter-lived and/or requires much faster migration/instantiation processes.

# Fog vs. Edge Computing

- Nowadays many people use the term Fog computing and Edge computing without much differentiation.
- Actually, there is no commonly agreed differentiation between both, but we may say that many people consider Edge computing as the inclusion of the edge devices themselves in the ecosystem, as hosts of programmable computing and storage resources (with or without support for virtualization).
- As we move to the Edge some of the Cloud characteristics tend to fade: Edge is not a seamless extension of the Cloud, there is less control and manageability, and much less scalability (per node).



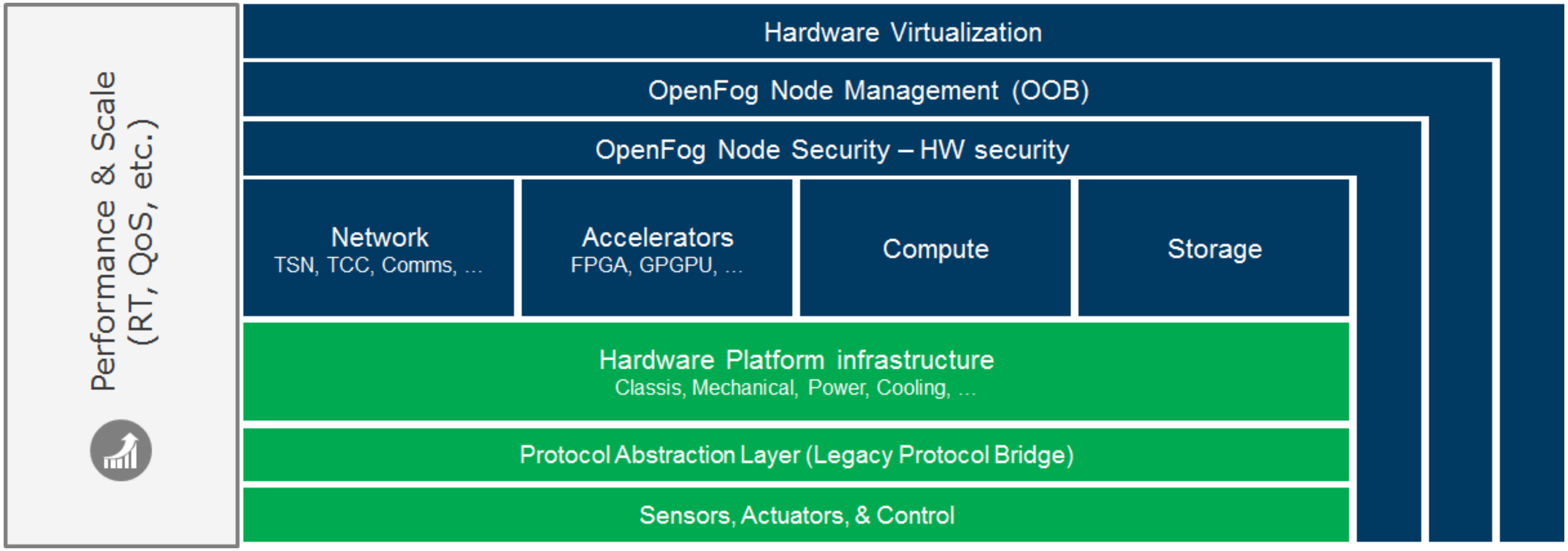

Source: https://www.pubnub.com/blog/moving-the-cloud-to-the-edge-computing/

# Reference Architectures

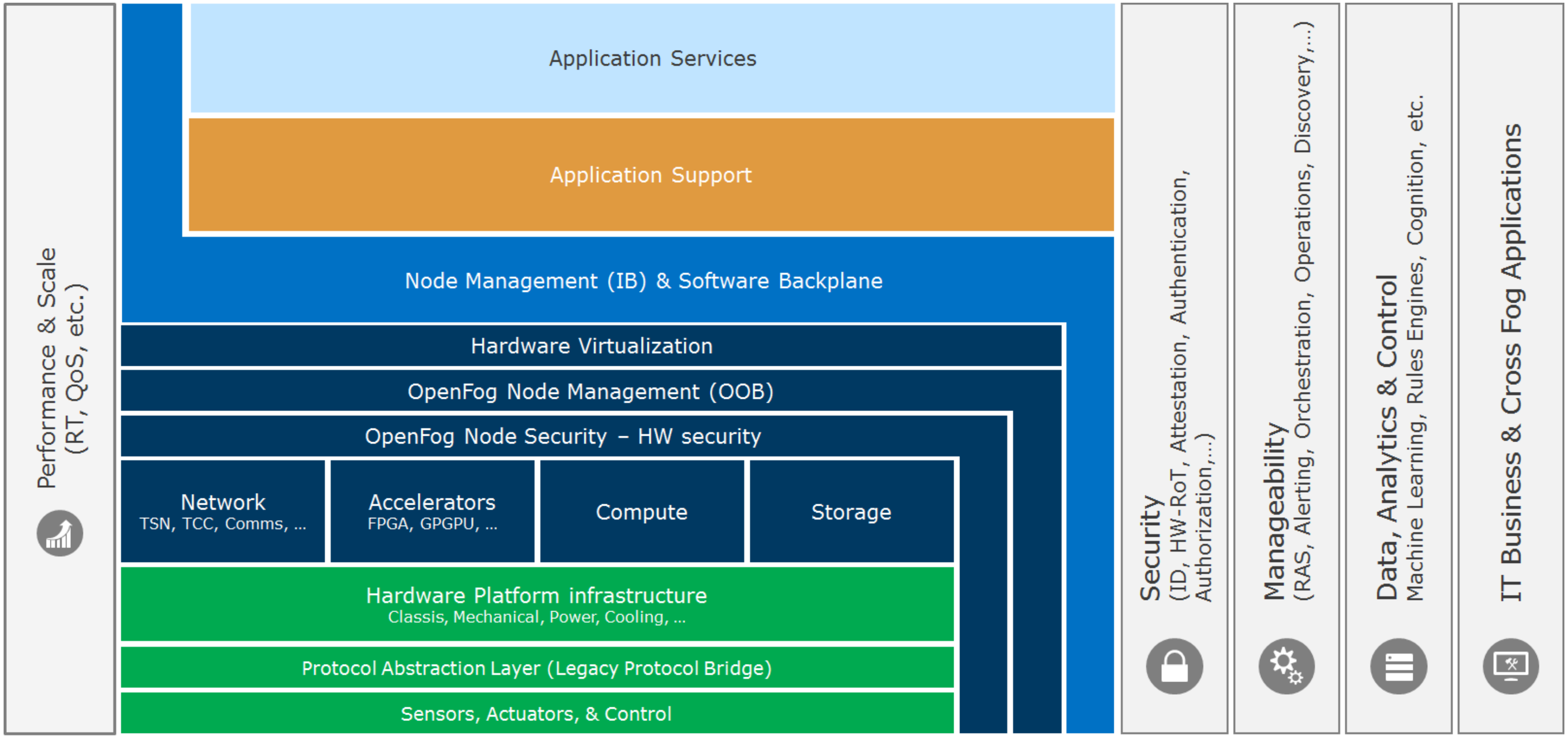
**OpenFog - [https://opcfoundation.org](https://opcfoundation.org)**

- The OpenFog Consortium (OpenFog) is a public-private ecosystem formed to accelerate the adoption of fog computing in order to solve the bandwidth, latency and communications challenges associated with the Internet of Things (IoT), Artificial Intelligence, Robotics, the Tactile Internet and other advanced concepts in the digitized world. It was founded by ARM, Cisco, Dell, Intel, Microsoft and Princeton University Edge Computing Laboratory in November 2015.

# OpenFog Reference Architecture
## *(basic focus on the fog node)*



# OpenFog Reference Architecture
## *(a more complete perspective)*



# How does this link with telcos?

**CORD – Central Office Re-Architected as a Datacenter**

**https://www.opennetworking.org/cord**

*Central Office in telecommunications jargon is the place where subscriber lines (e.g., DSL, fiber) are concentrated to enter the telecommunications operator network – a medium-sized city like Coimbra typically has 1 to 3 Central Offices*

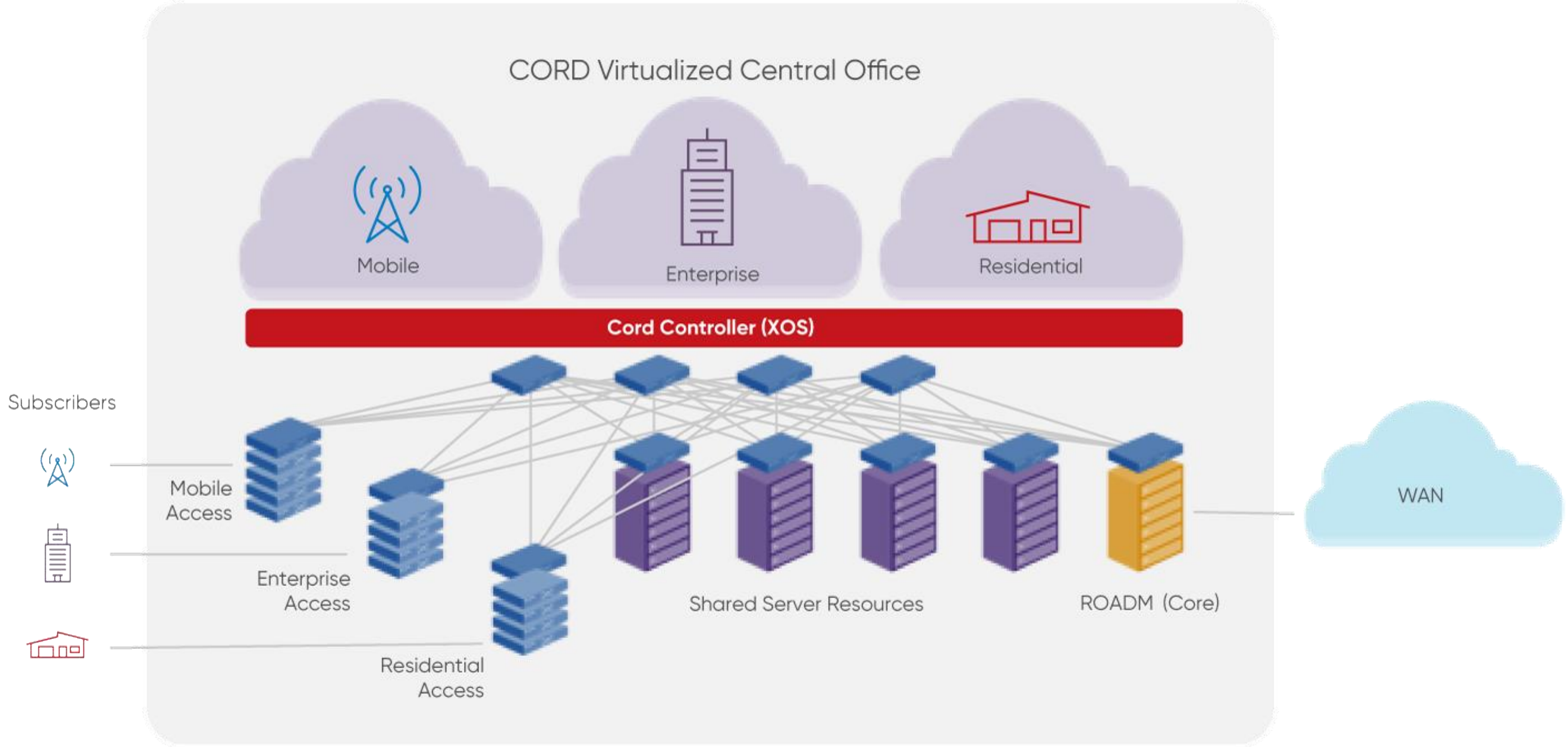
- Commodity servers interconnected by a fabric of *white-box* switches
- Switching fabric in a spine-leaf topology for optimized East-to-West traffic
- Specialized access hardware for connecting subscribers (residential, mobile and/or enterprise)

# How does this link with telcos?

**CORD – Central Office Re-Architected as a Datacenter**

**https://www.opennetworking.org/cord**

Very good video: https://www.youtube.com/watch?v=h1EzCBuY53Q



# Further Reading

**OpenFog Consortium Materials:**

**https://opcfoundation.org**

**NIST Definition of Fog Computing:**

**https://csrc.nist.gov/csrc/media/publications/sp/800-191/draft/documents/sp800-191-draft.pdf**

**OSGi Alliance (a bit out of topic, but anyway…):**

**https://www.osgi.org**