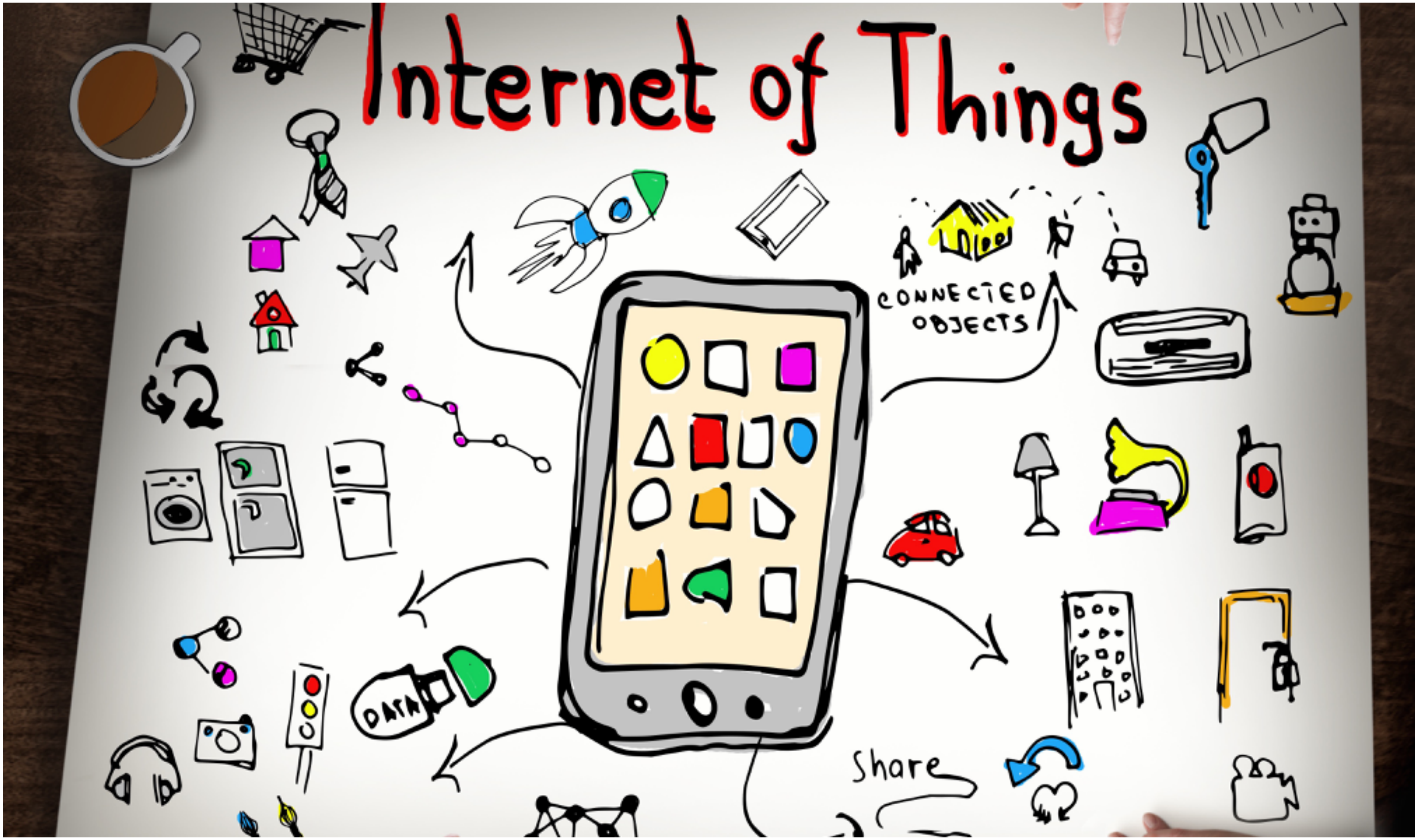


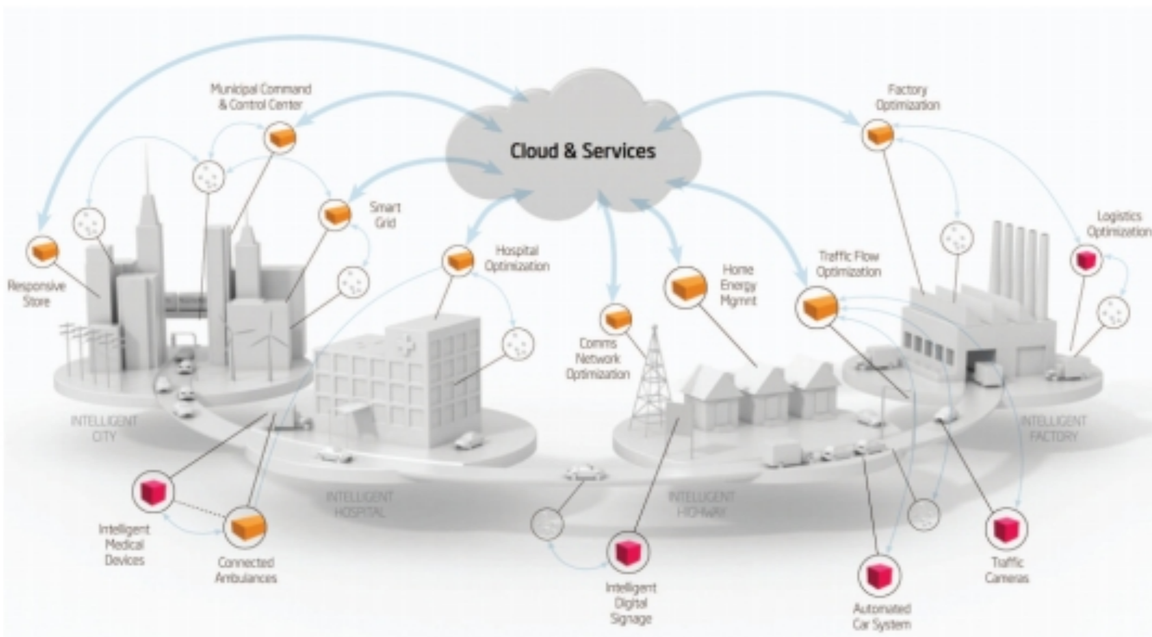
Cloud/Fog/Edge Computing

18-738 Sports Technology

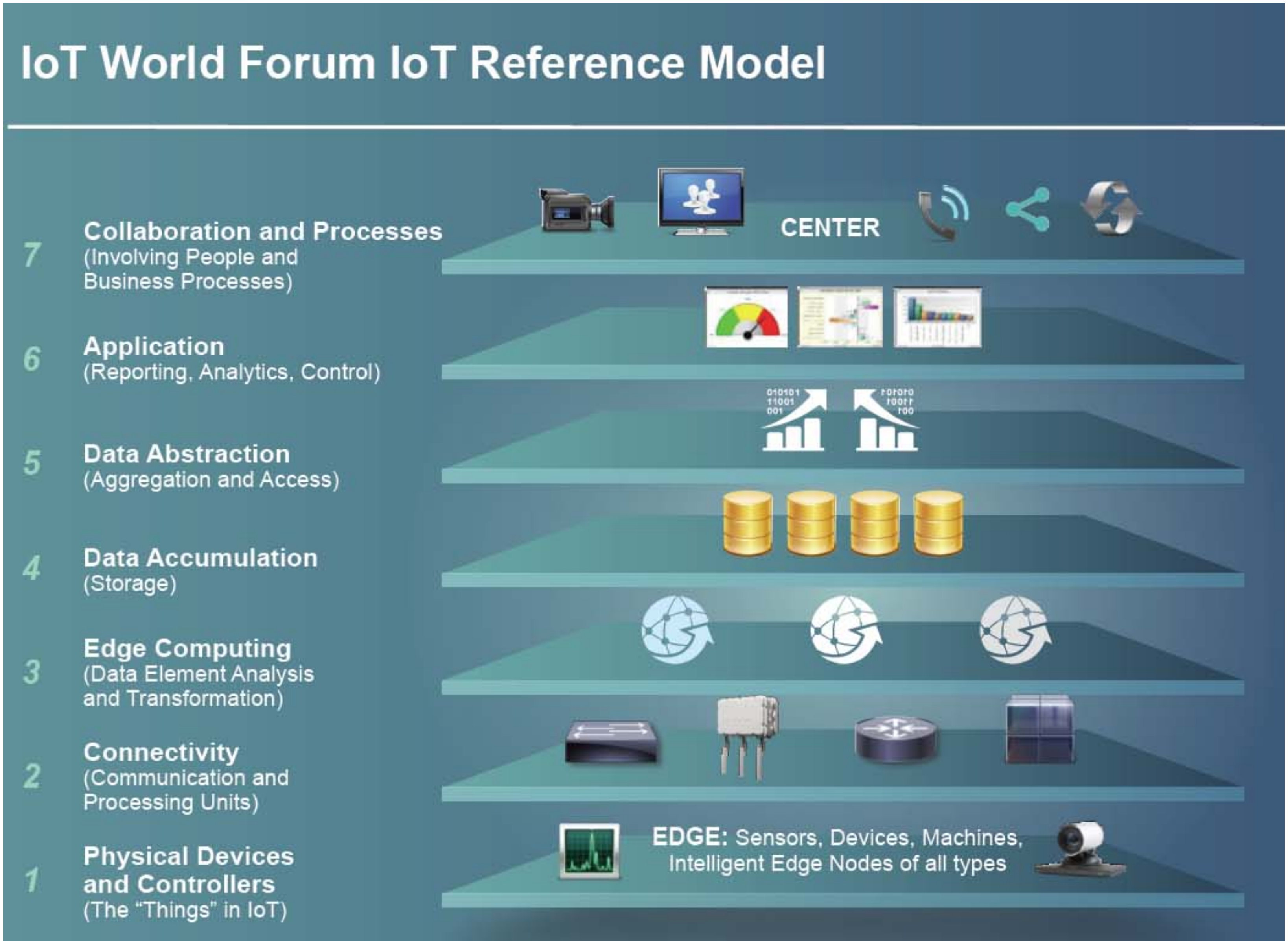
Priya Narasimhan
ECE Department
Carnegie Mellon University
@yinzcampriya

IoT and the Cloud

- Cloud computing is a fundamental enabler
- Allows for scalable, on-demand resources
- Allows for big-data analytics
- Allows for offloading compute-intensive tasks to beefier resources
- Allows for anytime-anywhere access to data and results



IoT Architecture/Model

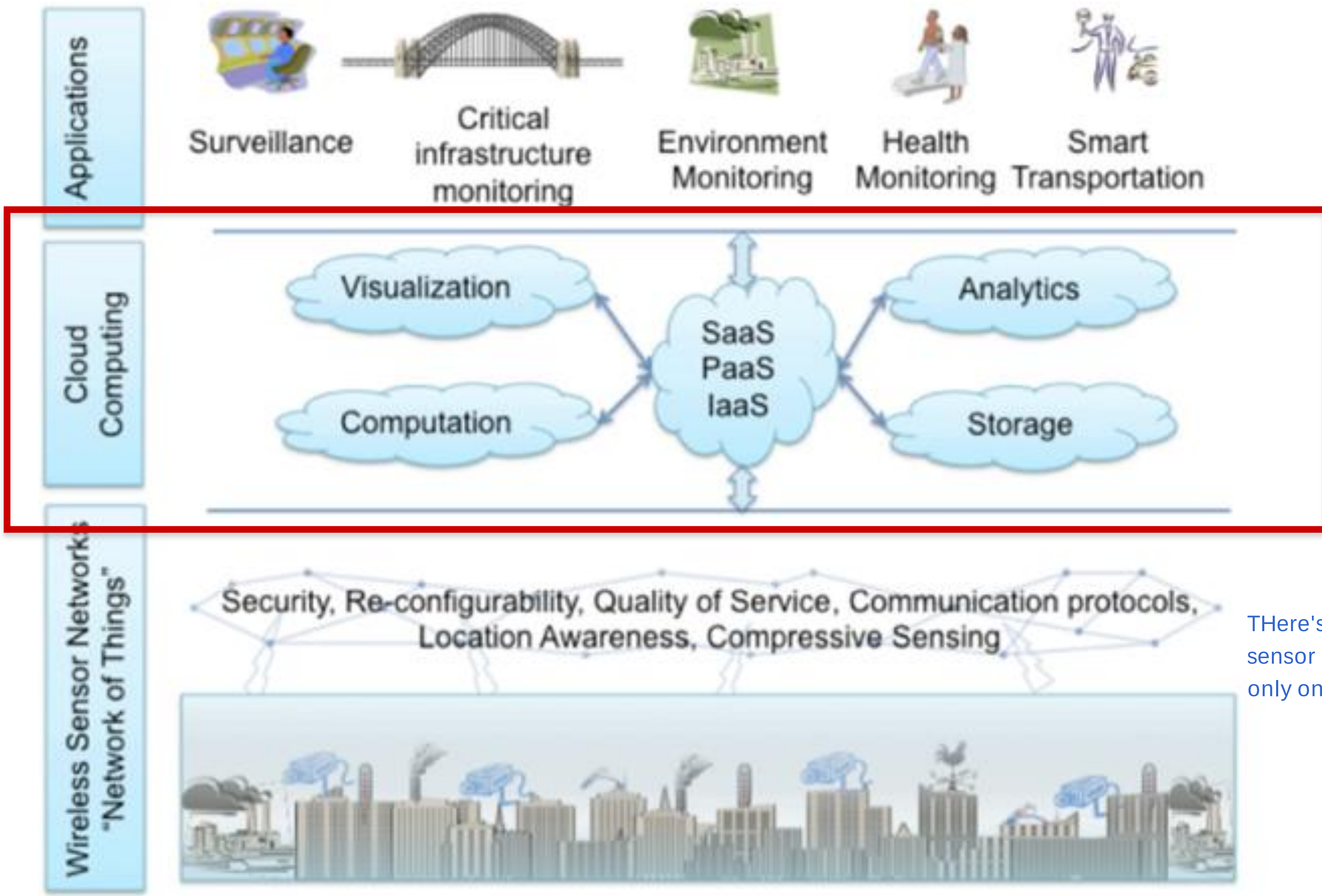


IoT and the Cloud

Application

Cloud

Sensors.



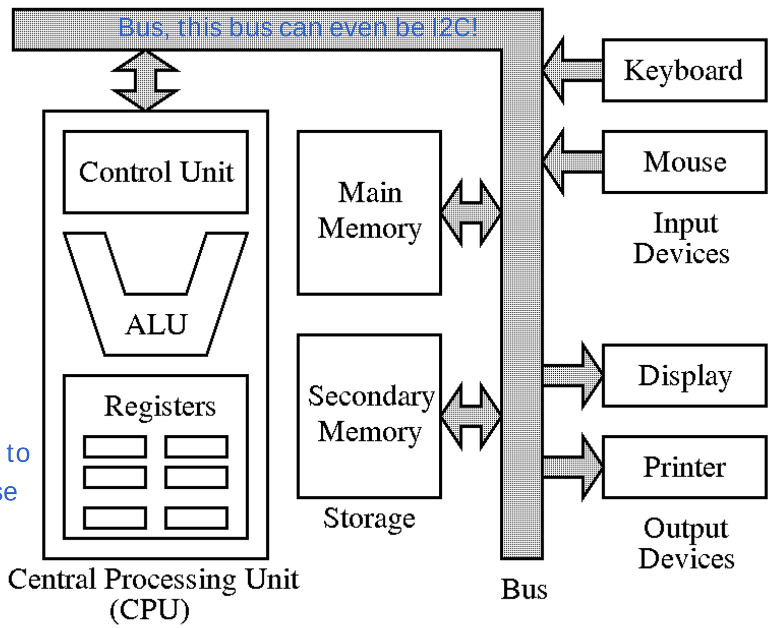
THere's more layers to the sensor layers, not only one.

Questions to Answer

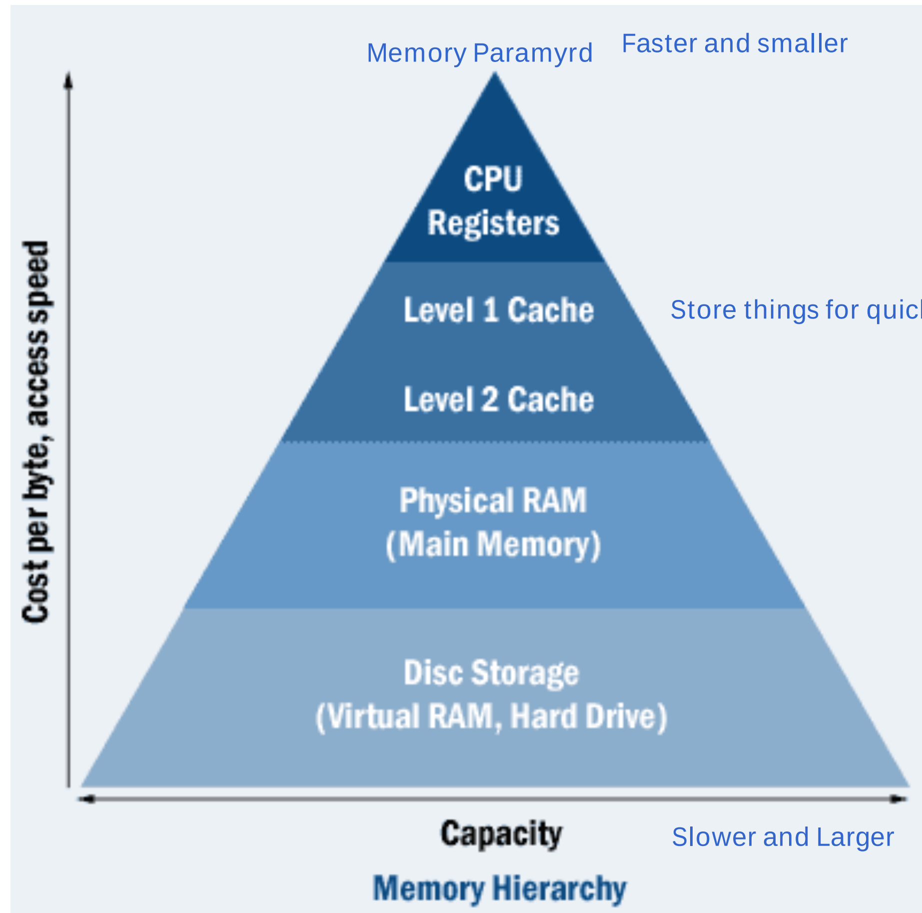
- What types of hardware do I need to be aware of and where do they fit in?
  - Enterprise servers, department servers, file servers, print servers, routers, workstations, PCs, handheld devices, embedded devices
- What's inside the box?
  - CPU, memory, interfaces, internal bus, …
- What are strengths/weaknesses of key platforms?
- What are the key SW-HW interactions and dependencies?
- What is (best) done in HW and what is (best) done in SW?
- What is cloud computing? There's no one type of cloud computing,
- What is a private vs. public cloud?

Basic Computer Architecture

registers =
quick access to
memory close
to CPU

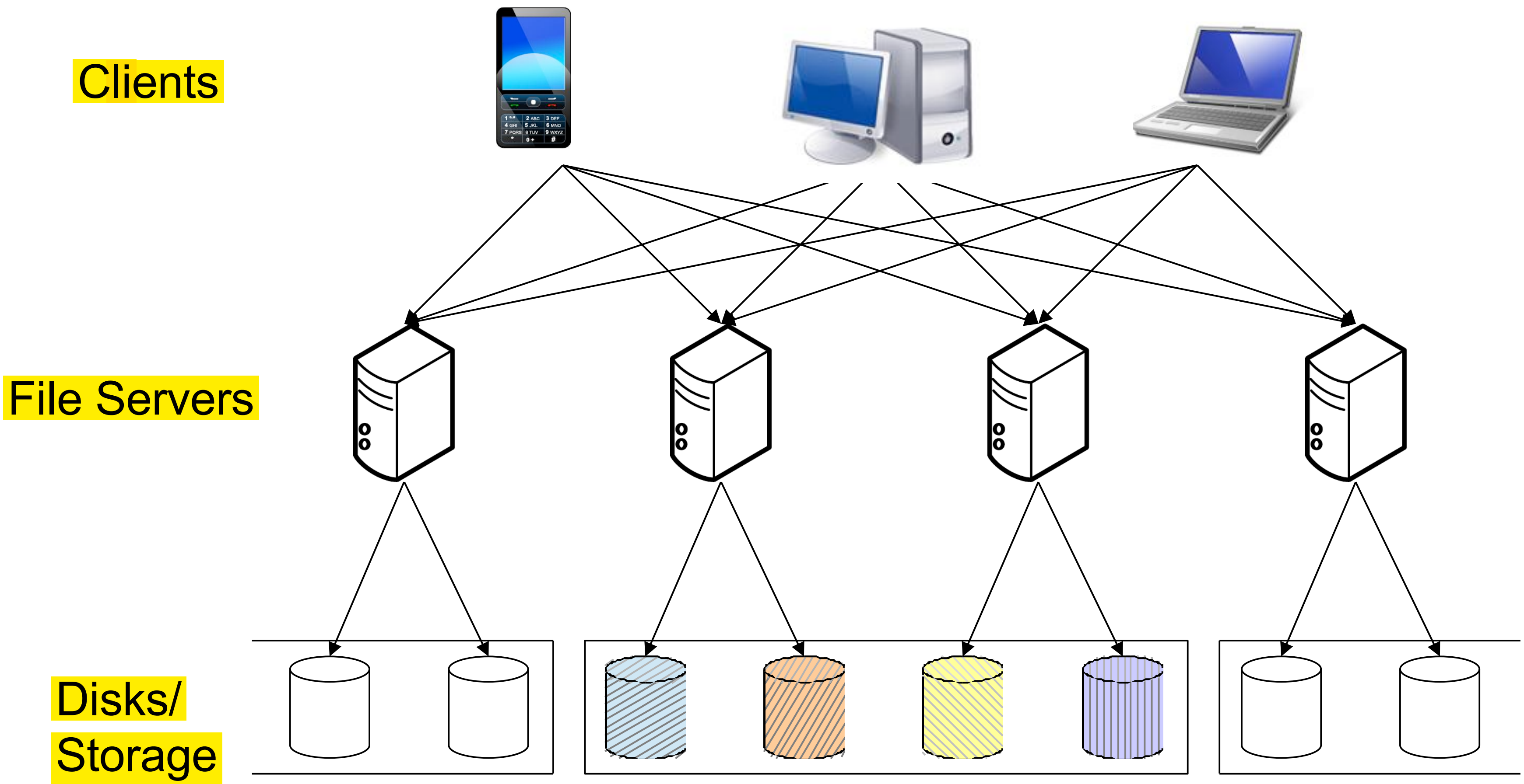


$$$ / Byte is larger higher in the pyramid.



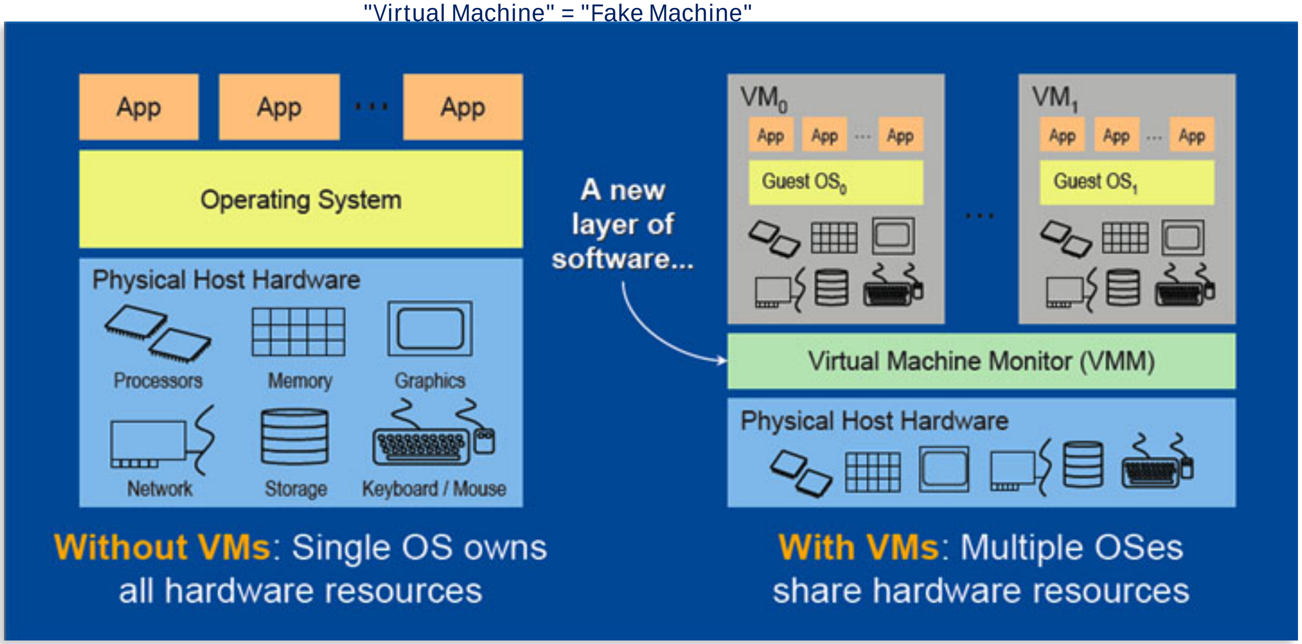
Store things for quick access.

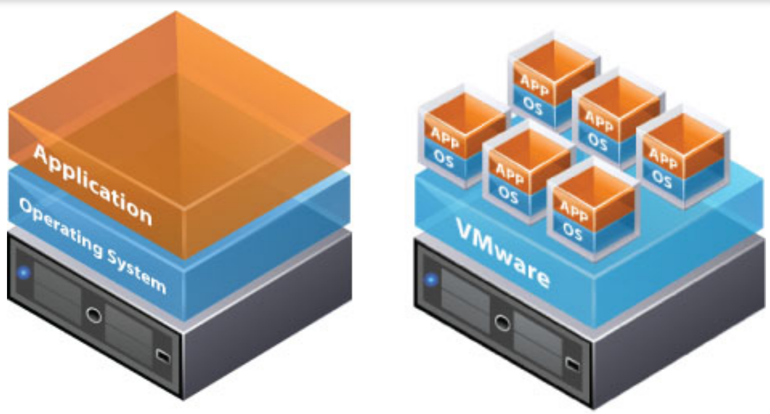
Basic System Architecture



Virtualization

There are multiple processes on the hardware. Virtualization = reservation systems. You reserved part of the hardware for the
virtual machine.
e.g. Amazon Instances == virtual machine -> they are not full os.
Virtual Machine = Slice of OS + Slice of Hardware = guaranteed reserved resources -> An OS running on another real OS.
1. guaranteed resource reservation
2. isolation from other application running on the same hardware ... when fails, it's not going to fail the entire system.
Each machien is not aware of other virutal machines.
Bare Metal = Hardware
"Virtual Machine" = "Fake Machine"





Traditional Architecture   Virtual Architecture

Bare Metal = Hardware that something runs on.
OS = Linux, Windows, etc...
A bunch of tasks run in form of threads and processes.


Cloud Computing

Outline

- What is cloud computing?
- What are its benefits?
- What are the risks?
- What are the economics?
- What are common platforms?

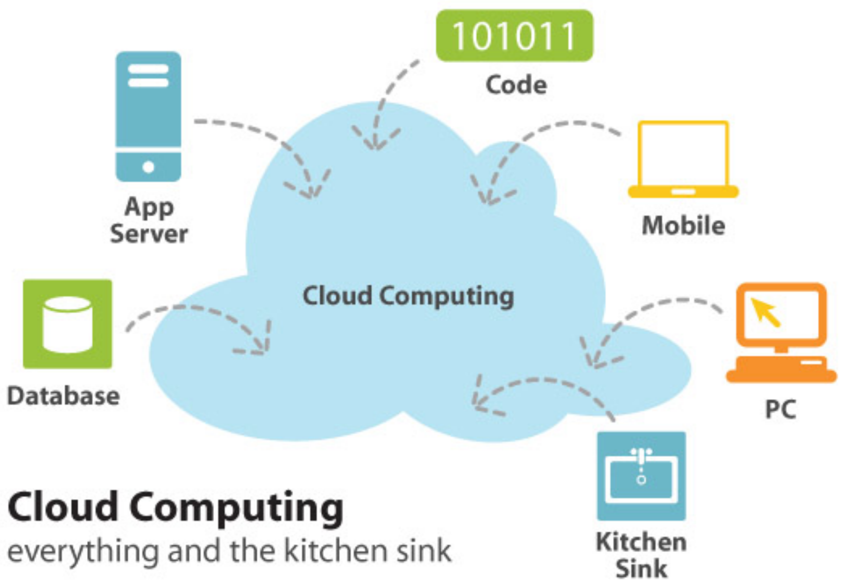
Act of Leasing vs Owning infrastructure:
Leasing: leave problems to the owner

Owning: every problems you need to fix by yourself

Total Cost Ownership: everythingis on you versus the owner

What is Cloud Computing?

CLoud Computing == the act of leasing vs owning infrastructure.



Renting an apartment = managed infrastructure
Renting an apartment = furnished apartment, unfurnished, utilities paid/unpaid

Renting an apartment = flavors of cloud computing
TCO is transferred to the owner.

Owning your own house = building your own data center

TCO = Total Cost of Ownship,
any furniture breaks, it's your problem. But you also have all freedom.

- Broad term that covers
  - Software as a Service (SaaS)
  - Platform as a Service (PaaS)
  - Infrastructure as a Service (IaaS)
- Also called "utility computing"

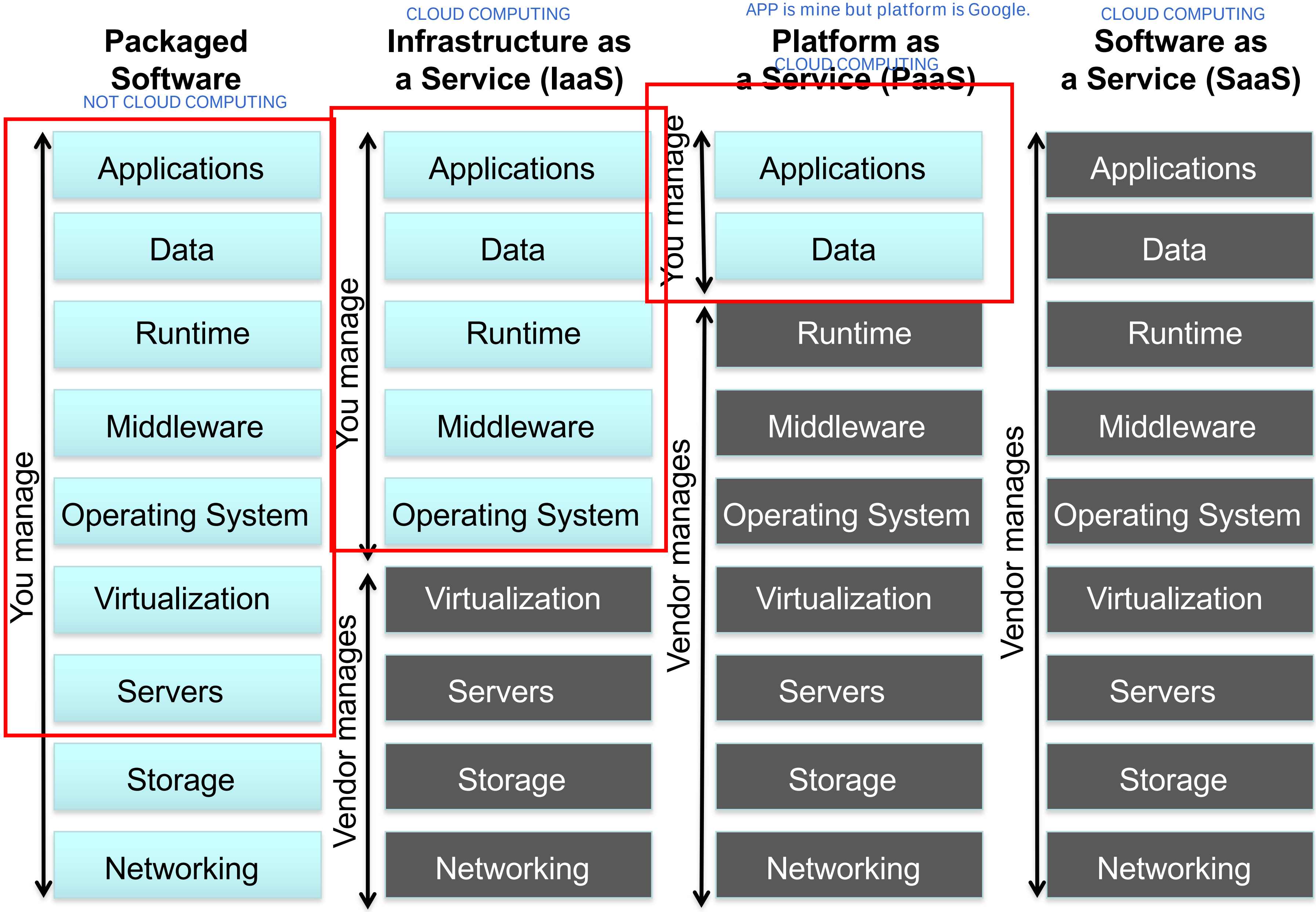
"Pay as you go" also called "utility computing" -- just like electricity.
Cloud computing is bought like utility such as water and electricity. You pay for it as much as you use.

AMAZON EC2 rent by the minutes, by the hours.

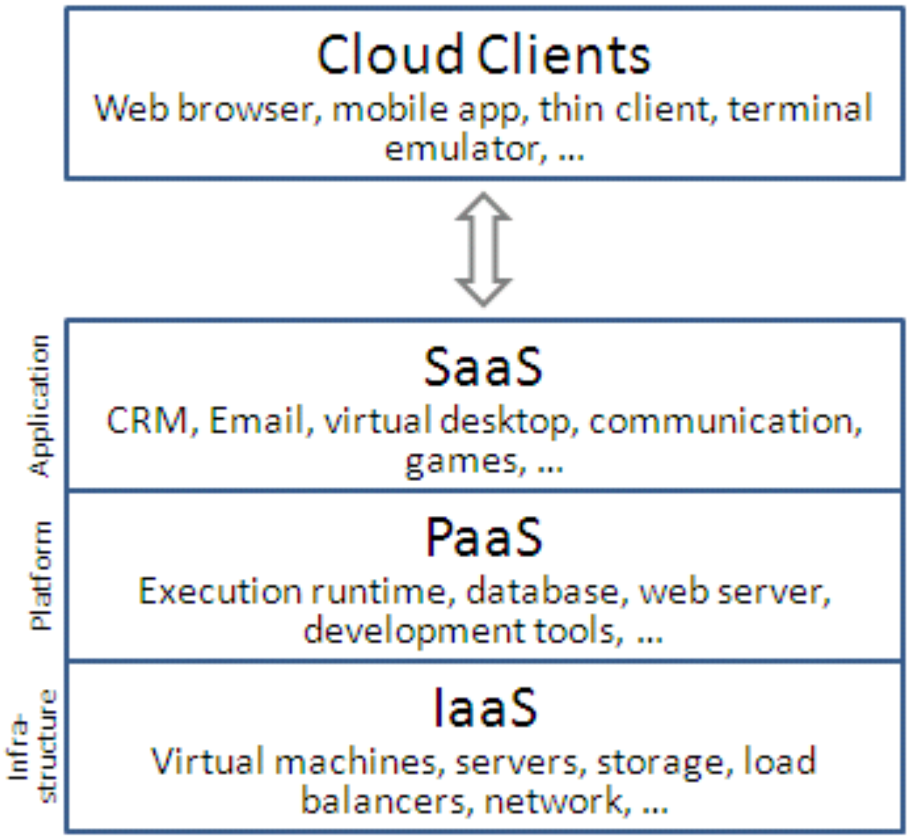
You have the choice of choosing the OS / Hardware. Exmple: Amazon Virtual Machines.

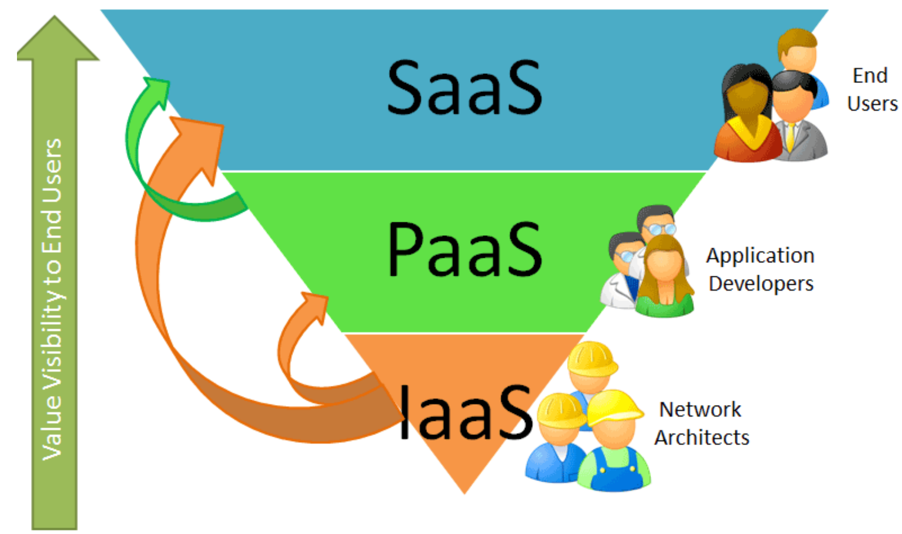
Example: *Service* on top of Amazon EC2.
You can build services based on Google Map's API.
APP is mine but platform is Google.

If software as a service, everything is contolled by someone else.
All you get is a Login, you don't know where the data is. Example: Sales Force.



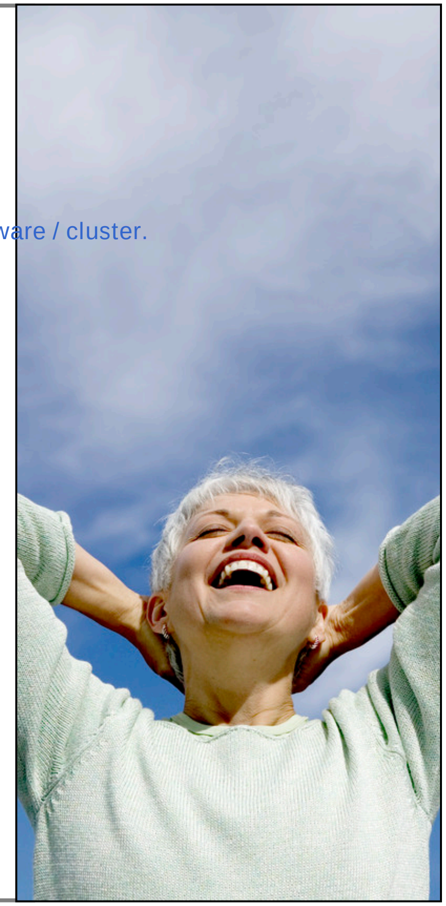
Layers of Cloud Computing






To cloudify or
not to cloudify?

Why Cloud Computing?



Reduced hardware costs (rent instead of buy)

No need to purchase storage/servers

No need to "refresh" cluster

refresh = how often do you need to replace home repairs,

If you have a data center, %5 will likely fry up every year. You don't need to refresh / replace the hardware / cluster.

Reduced software costs

Simply pay a "metering fee" for usage

Reduced physical space and administrative costs

No added costs for electricity, cooling, UPS, firewalls

No need for large in-house sysadmin team

Agility and on-demand payment model

Quickly add (and pay for) computational resources

If you need quickly add some service someday, you can have them quickly.

Power of numbers

New business opportunities due to flexibility and scale

Rent for hours, get work done, and finish.

Sports is pretty much in-demand. Situation on Monday and on Tuesday are differernt. Thus renting server is more flexible.

Top five Users of Cloud Computing, ESPN, Yahoo, and Netflix. Sports is On-Demand, on Friday maybe Saturday is most busy hours.

Why Not Cloud Computing?

Reliability might be out of your hands

- Cloud infrastructure SLAs are often not enough
- Must live with provided reliability mechanisms

Vendor lock-in

- Hard to migrate from a cloud provider

Lack of performance guarantees

- Instances vs. servers, virtualized environments
- Reserved instances are more expensive

Guaranteed Resource != Guaranteed Performance --> large delays.

Maybe you have high latency!

Responsiveness

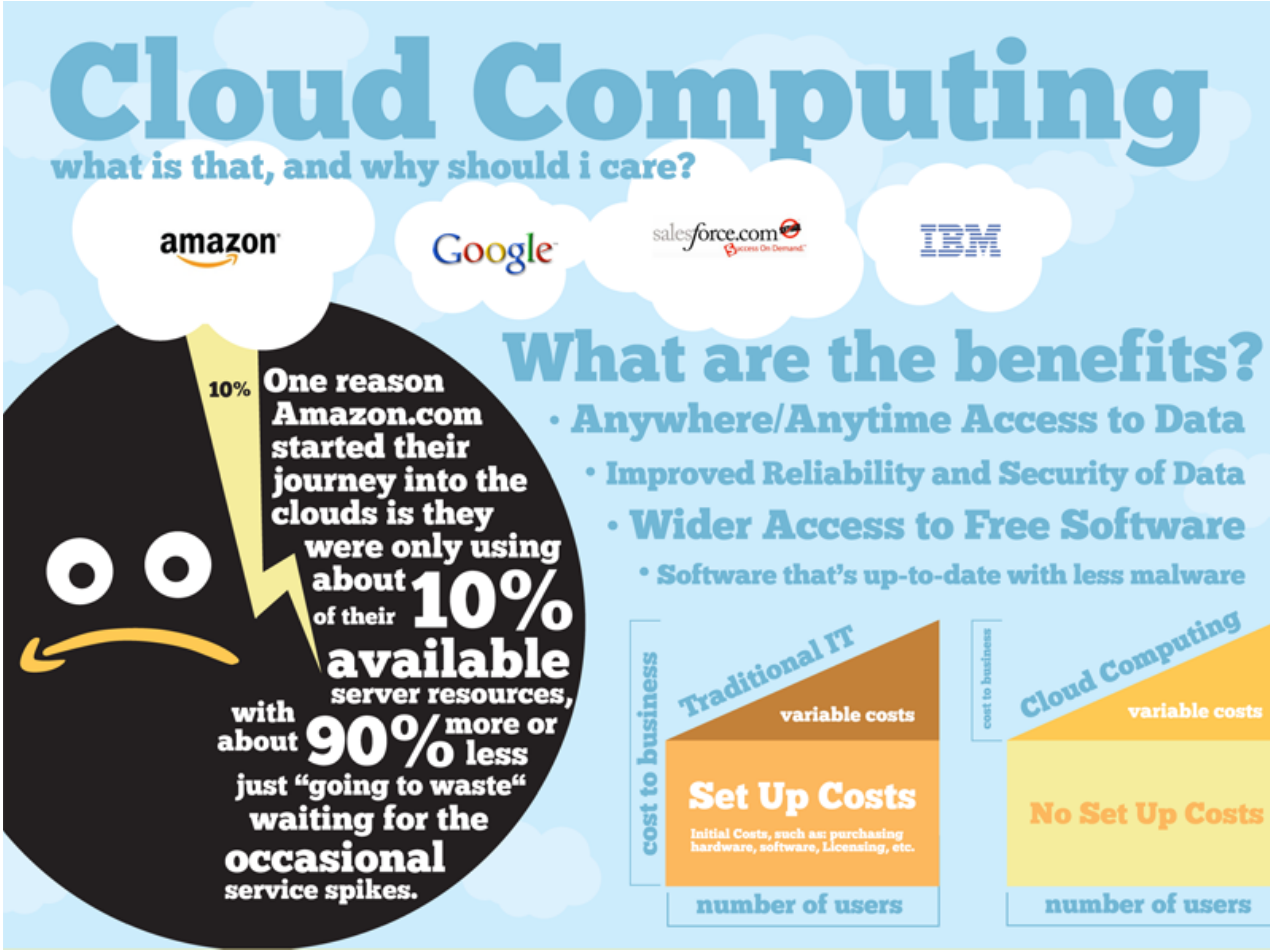
- Not ideal for low-latency edge services
- Spiky, rapidly changing workloads difficult to handle

If you have something not working, you need to call the landlord, the service provider to fix the problem.

Regulations

Health care data can be sensitive. They require that some is never placed on the cloud.

Amazon came up with EC2 because of Mother's Day. They bought servers for a peak load on the mother's day.



Amazon started on Mother's Day because they used to sell books. On mothers day the sales peaked and they bought a lot of servers for that day.

Amazon Web Services (AWS)

- Amazon Elastic Compute Cloud (EC2)
- On-demand, resizable computational capacity
  - As many servers as you need, when you need them
- Complete control
  - APIs for creating, scaling and managing instances
- Instance configurations
  - Variation by CPU, cores, memory, storage
- Pre-built Amazon Machine Images (AMIs)
  - Entire OS and software stack that can be loaded
  - Linux, Windows, Open Solaris

Amazon EC2 Instances

- User can create, launch, terminate instances that vary in storage/compute
  - Micro-instance
  - Small instance
  - Large instance
  - Extra-large instance
- User can decide geographic location (region) of instance
- Auto-scaling: Dynamic adaptation to workload
- User pays hourly charge per running VM + data-transfer charge + storage charge
  - Example: $0.02/hour for the smallest on-demand Linux VM
- User can reserve instances for multiple years
- Pricing: http://aws.amazon.com/ec2/pricing/

Amazon Elastic Block Storage

- Persistent storage
  - Independent of any computation/EC2 instance

Volumns you can access any where in the world.
Raw-unformatted devices sitting on EC2.

- Platform-agnostic
  - Raw, unformatted block device accessible from Linux/Solaris/Windows
- Scalability
  - From 1GB-1TB
  - Easy to attach, store, backup and restore volumes
- Reliability
  - Built-in data redundancy

Auto Scaling

ESPN -- sport has busy hours
Netflix -- Movie viewer have busy hours.

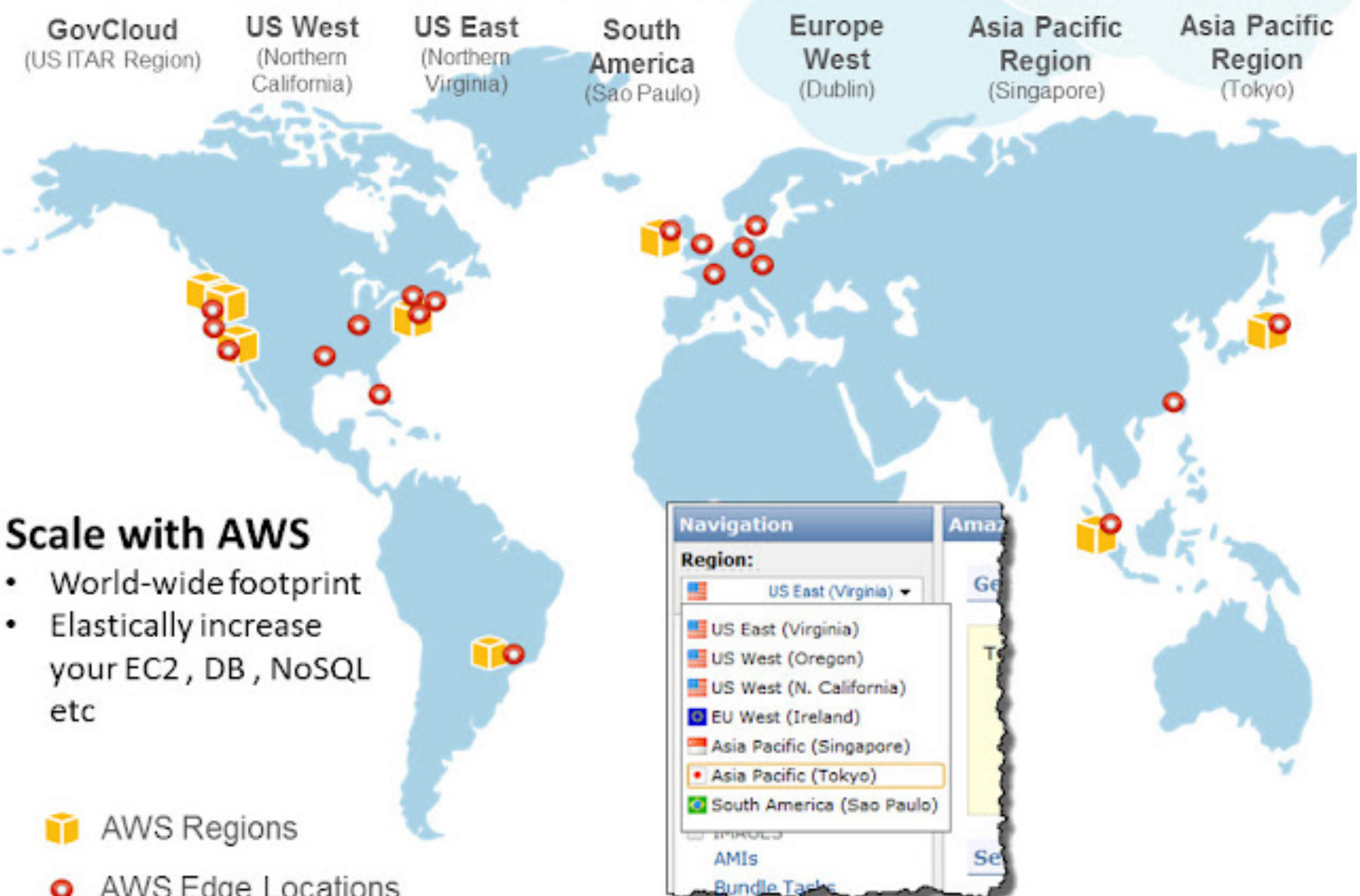
Scale AWS capacity up or down automatically
- According to conditions you specify

Overall objective
- Use just as much resources as you need

Number of EC2 instances
- Increases transparently during high demand
- Drops transparently during workload lulls

Triggered by Amazon CloudWatch
- Monitoring of AWS cloud resources and applications
- Track and collect standard metrics (CPU usage, latency, utilization, request counts)
- Track and collect custom metrics
- Set alarms when metrics cross thresholds

AWS Global Infrastructure



Scale with AWS

- World-wide footprint
- Elastically increase your EC2 , DB , NoSQL etc

Image Source: AWS

Availability Zones (AZ)

Amazon in addition to edge computing, now has the notion of MEC (mobile edge computing)

- Distinct locations, insulated from failures
- Low-latency connectivity within same region

Source: Simone Brunozzi, 2011

AWS Ecosystem

Alternative PaaS Platforms

Microsoft Asure ~= EC2 ~= Google App Engine

- **Microsoft Azure**
  - Rich cloud platform for .NET developers

    Microsoft Azure is mainly used by some sport teams such as Lalita. The entire backend is based on microsoft Azure.
  - Supports .NET technologies, Java and web technologies
  - Ecosystem of services, e.g., load balancers, scalable distributed database, business analytics, content delivery network
  - No free version
- **Google App Engine (GAE)**
  - Similar in spirit to EC2, pay-as-you-use model
  - Infrastructure similar to that underlying Gmail and Google Docs
  - Ecosystem, e.g., load balancer, SQL, MapReduce
  - Time-driven quotas for bandwidth/CPU (with free starter quota)
  - Support for specific languages and Google APIs
    - Primarily Python and Java, no support for C# and .NET

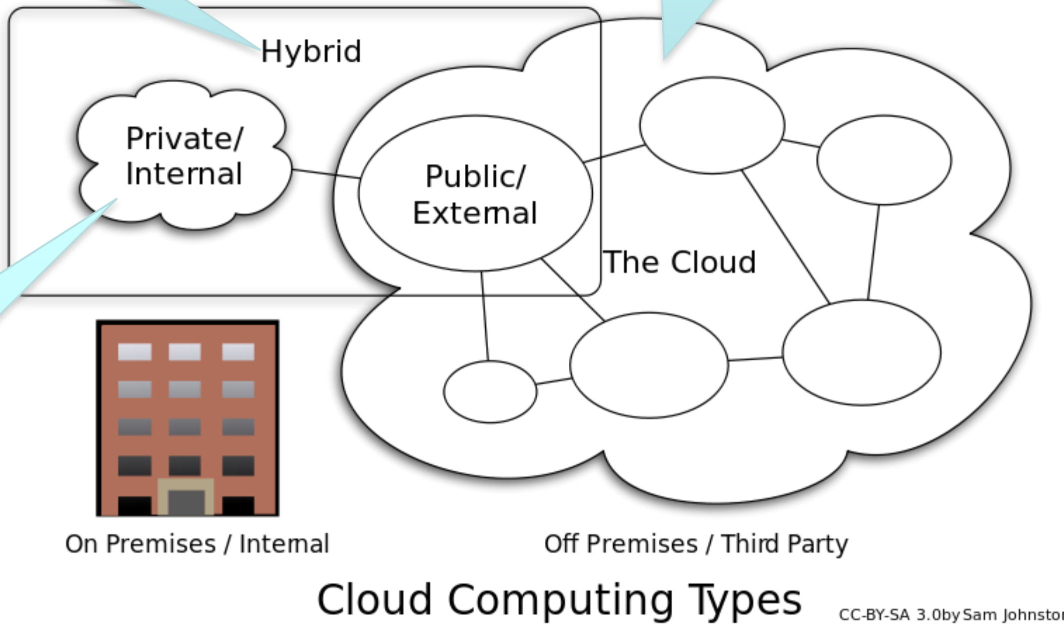
Public, Private, Hybrid Clouds

Private cloud = cloud computing inside organization.
It's similar to EC2 in that it also has virtual machine and hardwares. But not everyone out there can access.

Composition of two or more clouds (private, public) that are bound together, offering the benefits of both. Local immediate use/QoS independent of Internet connectivity/performance, without losing external scalability and rapid deployment.

Applications, storage, and other resources are made available to the general public by a service provider, e.g., Amazon AWS, Microsoft, Google. Services are free or pay-per-use.

Cloud operated solely for a single org, managed internally or by a third-party. Orgs "still have to buy, build, and manage them." Virtualization of the org's business environment



Private Cloud: HealthCare + Hospital -- Pretty obssessed. They buy, build and manage all cloud centers.

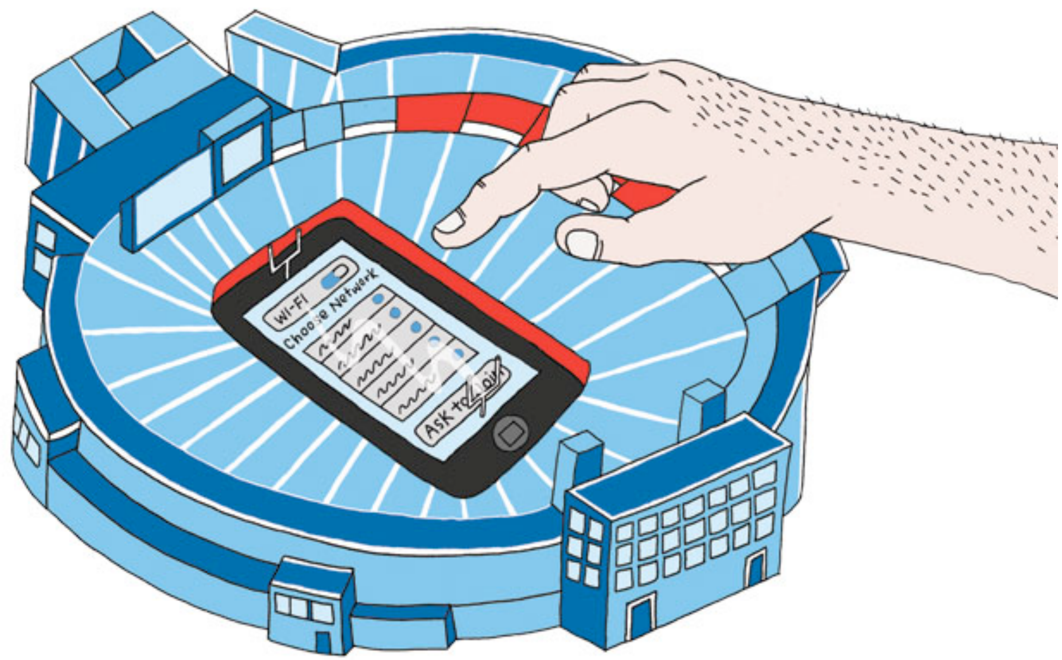
Challenges with Cloud Computing

- Large volumes of data needing to be transmitted from an IoT system
  - Half a terabyte of data from a Boeing airplane during a single flight, including all the sensors
  - Bandwidth of available networks may be insufficient for real-time responsiveness
- Cellular networks are typically provisioned for
  - Email, simple apps
  - Not necessarily for high-bandwidth usage
- Wi-Fi networks often supplement cellular systems
  - Video streaming, audio streaming
  - Higher-bandwidth services

Wifi = offload for cecullar so WiFi make Cecullar's life easier.

Recap: Wireless inside stadiums

- Wireless inside stadiums
  - Increasingly prevalent topic in the news
  - The need for having connectivity for fans
  - Necessary for basic equipment to work, for content publishing



Recap: In-stadium wireless usage

They have to bring cloud closer to the stadium.

Driven by what fans' consumption patterns

Let's look at **Super Bowl XLVIII** in New Jersey, 2014

- Peak usage: 13,500 (of 85,529 fans possible) connected over Wi-Fi
- 3.2 terabytes of data was transferred over the Wi-Fi system
- Facebook, Instagram and Twitter responsible for 10% of bandwidth
- 60% of connected fans shared on Facebook, 18% Twitter, 17% Instagram
- > 90,000 photos were uploaded to Instagram over Wi-Fi by devices
- 18% of devices were running software updates
- Peak usage of Wi-Fi was during the halftime show
- Increased move towards high-density Wi-Fi networks inside stadiums
- Increased discussion of multicast vs. unicast

Fog/edge computing

Fog computing is don't put everything data onto the cloud. We keep the data close to where the data is generated so that we can deliver with lower latency.
Cloud is higher, fog is lower -- thus the name.
Bandwidth computation -- sending the data over the entire cecullar network will drain battery. If you send it to a local wifi network you can have better .

Fog Computing is an architecture that

- Uses a collaborative multitude of end-user clients, or near-user, **edge** devices
- Carries out data storage, communication, control, and management tasks **locally** to improve responsiveness
- Is suited to **localized** lightweight interactions around sensing and responding to an event that IoT introduces

The argument is that *"if you can do computation locally and save bandwidth, why ship data to a remote computation server?"*

- Trade-off bandwidth vs. computation
- Maybe the actual data does not matter remotely, only the processed version of it
- Local decision-making vs. remote computation

Coined by Cisco in a research paper

Cloud Computing is not enough to handle all the bandwidth from sensor data. Why don't we push the computing closer to the data?

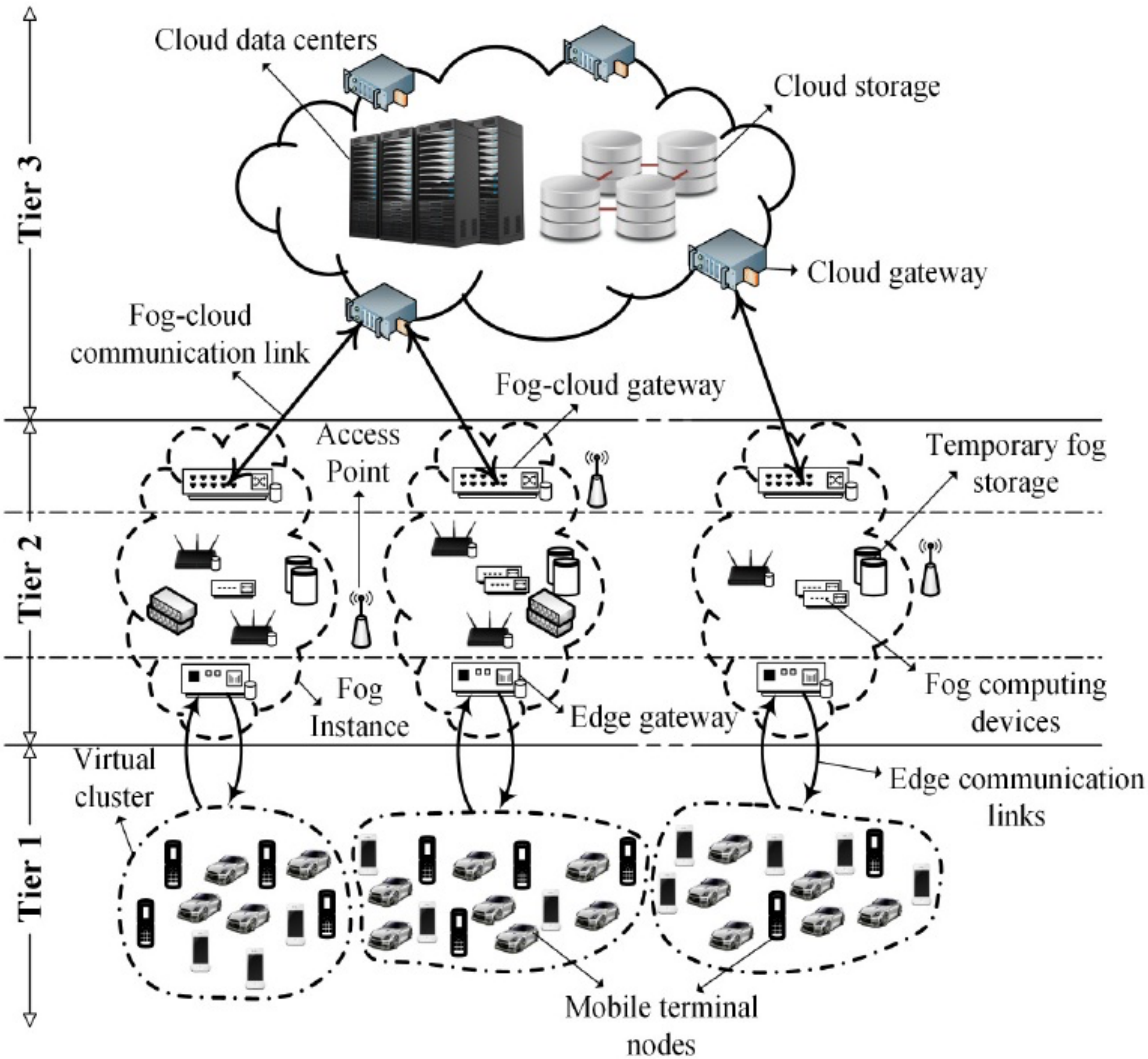
- Argued that cloud computing was not sufficient for IoT system that require
  - Mobility support
  - Geo-distribution
  - Location-awareness
  - Low latency
  - Examples in smart grid, logistics, road-monitoring (next few slides)
- Why “fog”?
  - It’s like a cloud close to the ground ☺

Coined by Cisco in a research paper

Tier 3 Cloud.

Tier 2 , close to sensors.

Tier 1, the sensors.



Digging a little deeper

- Virtualized platform that

  Can we have virtual machine in fog?

  – Provides compute/storage/networking services between end devices and "traditional" cloud computing
- Located typically at the edge of the network
- Can connect a number of edge devices together for improved responsiveness and improved latency
- Creating "islands" of connectivity and collaboration
- The "islands" may communicate via a back-end cloud
- Think of this in different verticals
  – A connected car

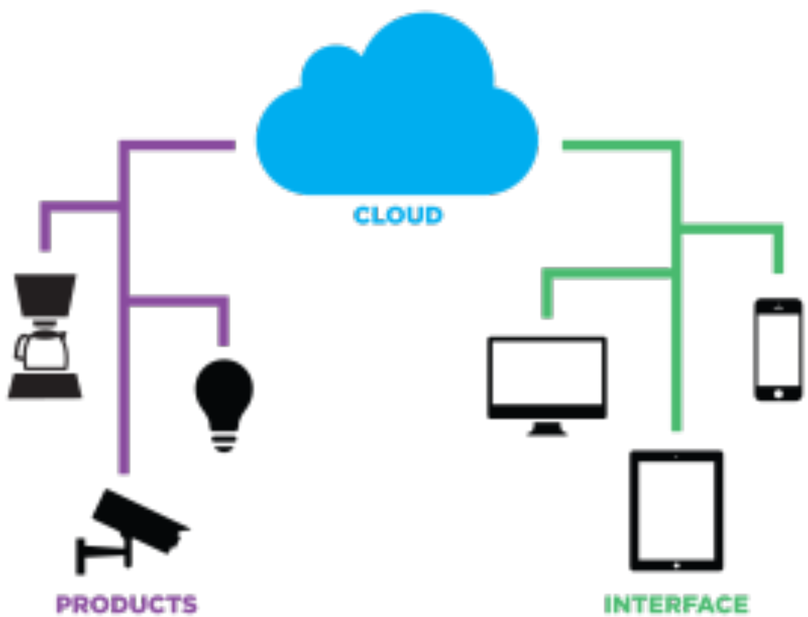
    The 100+ sensors connects to that single entity server locally, that entity connects to the Amazon sensor.

  – A connected store
  – A connected neighborhood

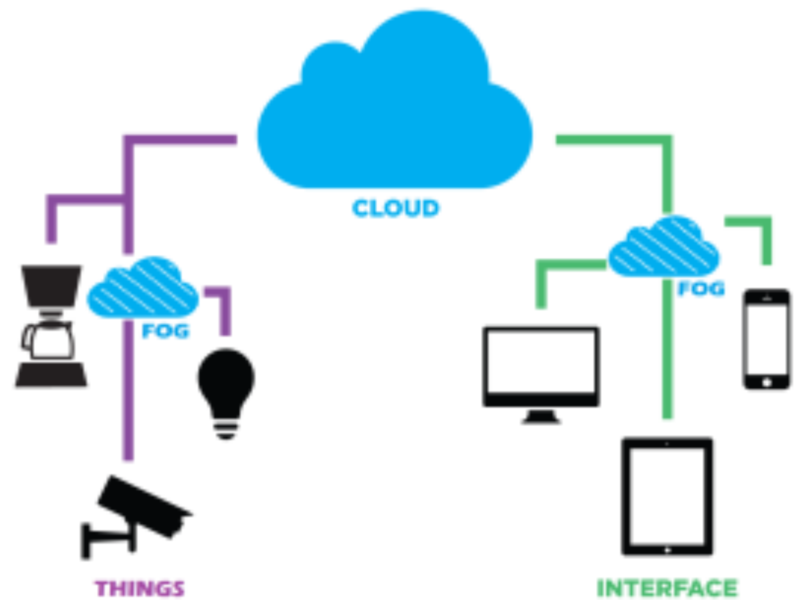
The stores connects to local servers, they do local computing..
The local serves then in turns coonnects to an Amazon server.

A store can be a fog computing node, all the snaks, drinks, the floors have sensor. These sensor data communicates with the store fog node. In stadium, the store is connected to the cloud, but not the small items.
The control room in stadium is connected to the cloud. But cameras and sensors are connected to local sensors. Very common noew

Architecturally



100% cloud



Fog + cloud

Source: http://iot-labs.com.my/2016/03/foggy-about-fog-computing/

Key Advantages of Fog Over Cloud

- Fault-tolerance: Can handle intermittent loss of connectivity because local processing can work off previous data If no cloud available, you can still compute.
- Data Compression: Raw data from local sensors can be compress before sending to the cloud to reduce bandwidth Be selective on what to send to the cloud.
- Lower latency: Local computing resources provide a faster response
- Offline configuration: Re-programming or re-configuration of edge nodes can be done without impacting the overall system

If you want to add fog nodes, you can add really quickly.

- *All of the advantages of a decentralized, decoupled architecture*
- *Here's the thing: Most of the data is generated at the edge, anyway*

The stadiums today are basically the fog serves.

Cloud vs. Fog

Big takeaway, big advantages and improvement in moving less data across the network.
Keeping it contained, actualy cecullar data is not overloaded.

Cloud Computing	Fog Computing
Data is processed in the cloud, which is time-consuming for large volumes of data	Data is processed at the edge, which can be optimal in terms of processing data and can involve smaller data-sets
Challenge with bandwidth, in order to send every bit of data over to a central cloud	Less of a challenge with bandwidth, as data is processed and aggregated at every layer, instead of sending every bit to a cloud
Possibly slower response times, depending on where the cloud servers are physically located	Possibly improved response times because “servers” can be located closer to the edge

Big take-away: Significant reduction in data movement

Edge Computing

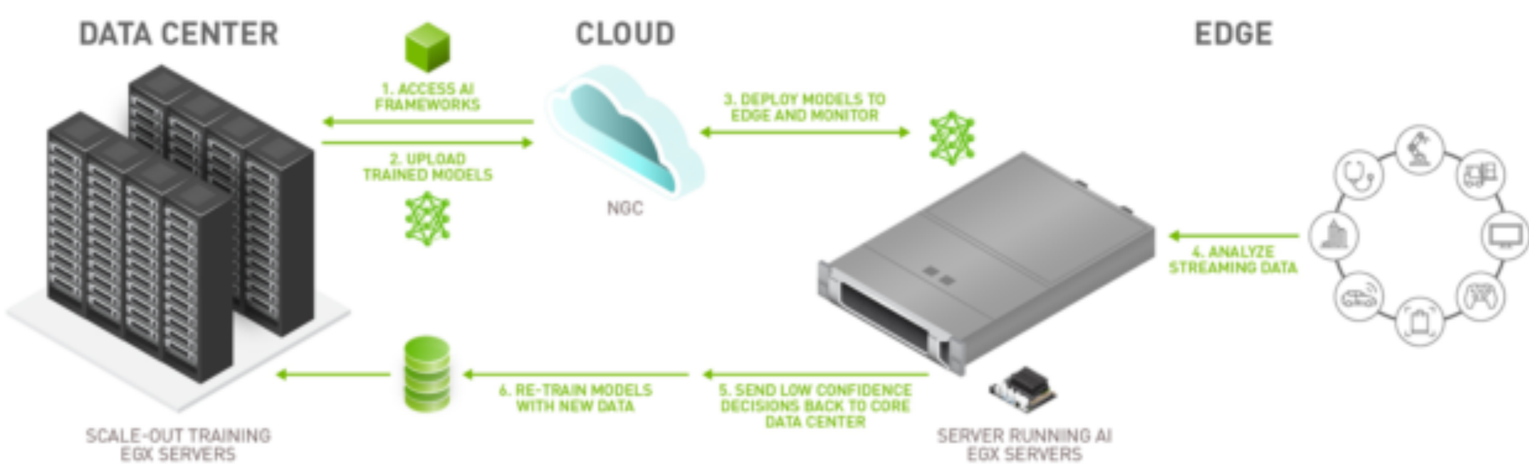
Another layer:
Cloud
Fog
Edge - Groups of sensors.

Edge Computing is to push computing even further to the sensors them selves.

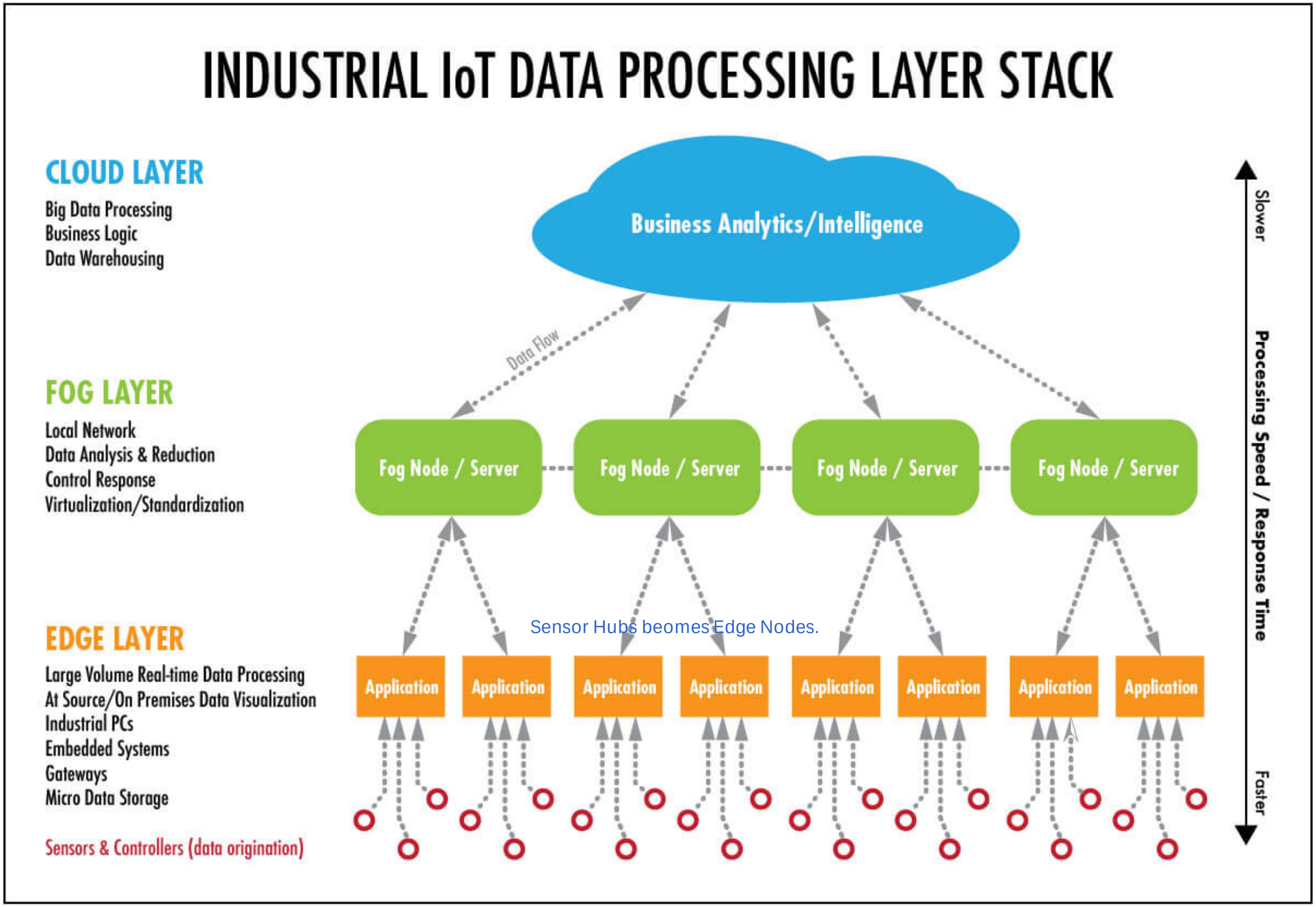
Edge computing is the practice of moving compute power physically closer to where data is generated, usually an IoT device or sensor.

It is named for the way compute power is brought to the "edge" of a device or network.

Edge computing is used to process data faster, increase bandwidth.



Edge Computing: Industry View

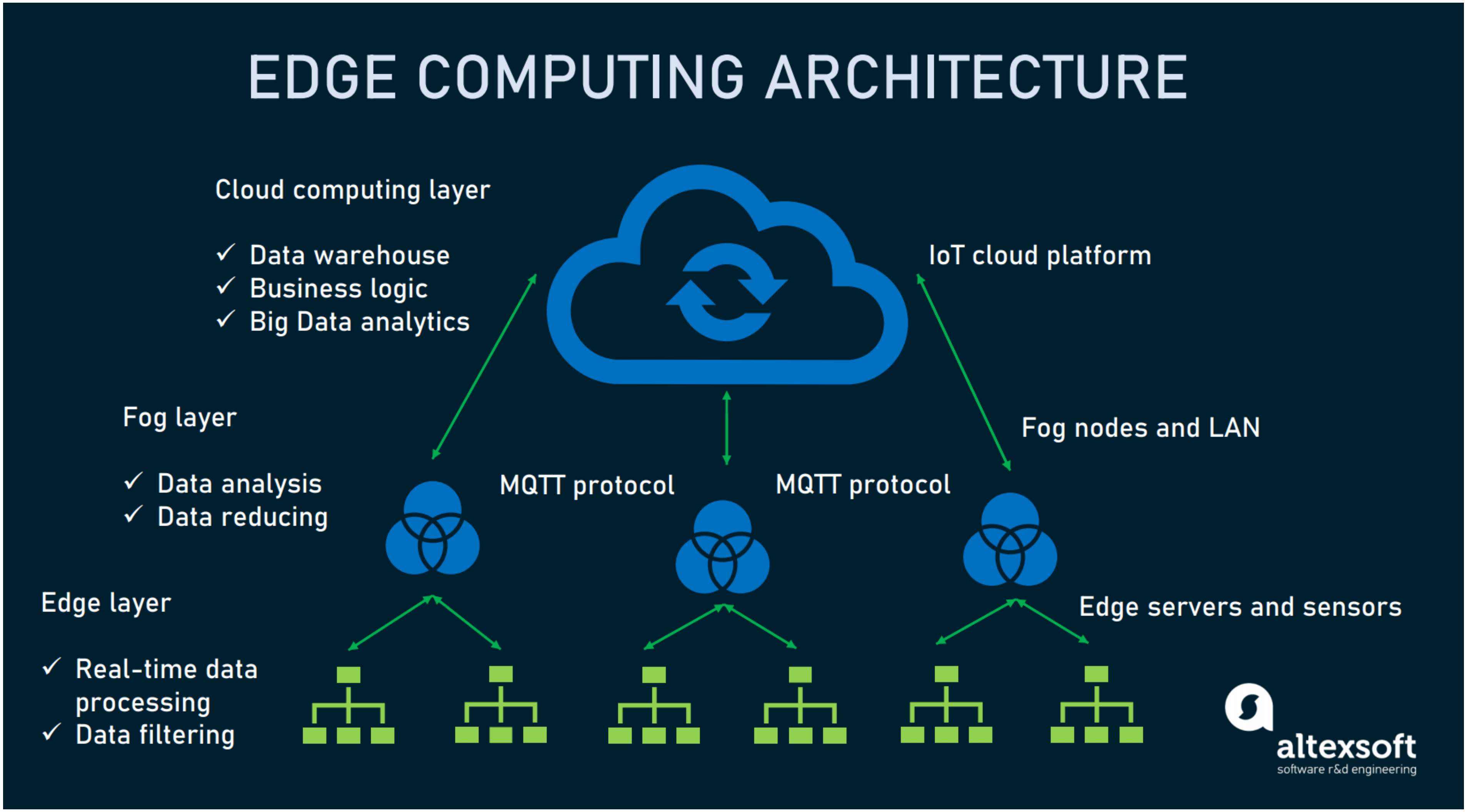


Sensor Hubs beomes Edge Nodes.

Edge Computing can be used to study individual hocky players in real time.

Edge Computing: Another Industry View

Edge layer protocals = I2C SPI RS232...
Fog Layer = MQTT COAP....

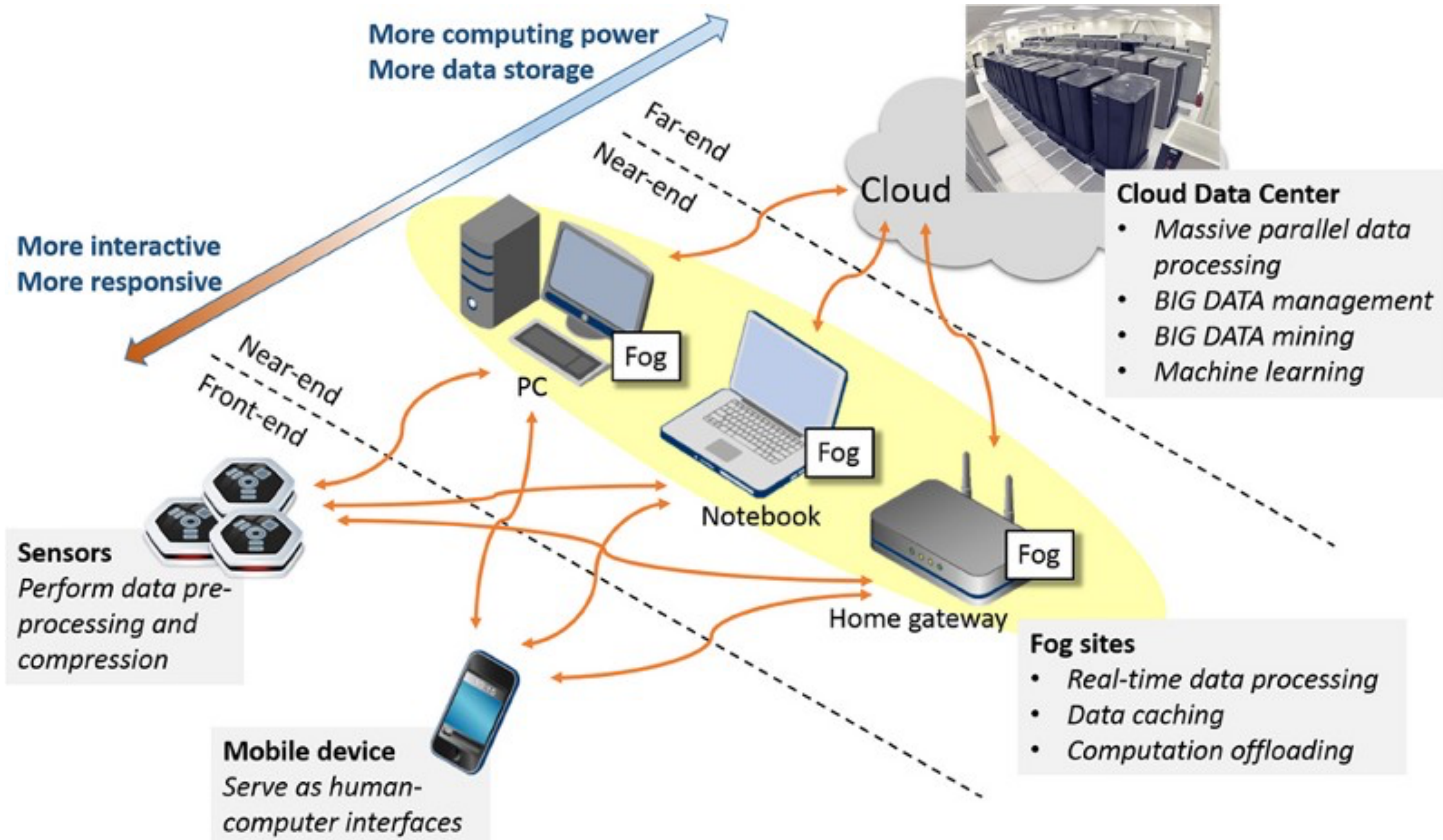


Edge Computing: Industry View (UPS)

UPS uses edge computing to make parcel selection decision locally to make decisions fast.

- USPS delivers 7.3 billion packages a year or 231 per second.
- To cope with this enormous load, the company has deployed AI algorithms on its edge servers located across 195 sites.
- Each server has a built-in optical character recognition (OCR) functionality to analyze images from more than 1,000 mail sorting machines daily.
- Running locally, deep learning models categorize packages, check if the postage matches a parcel’s size, weight, and destination, and decipher barcodes, even the damaged ones.
- Edge intelligence also helps locate missing parcels: With AI, this takes a couple of people and just a few hours — instead of the previous several days and 8 to 10 people.

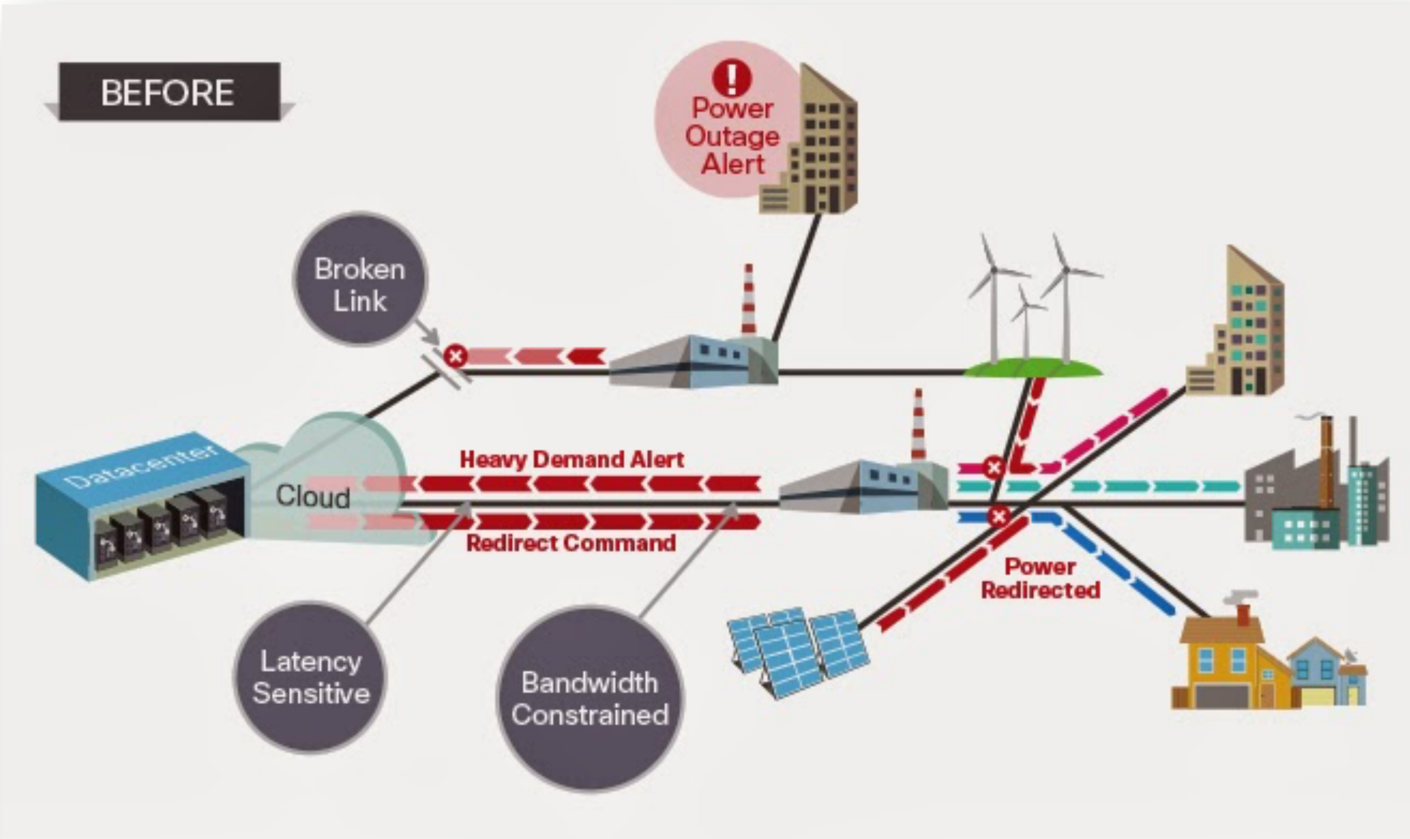
Cloud+Fog+Edge



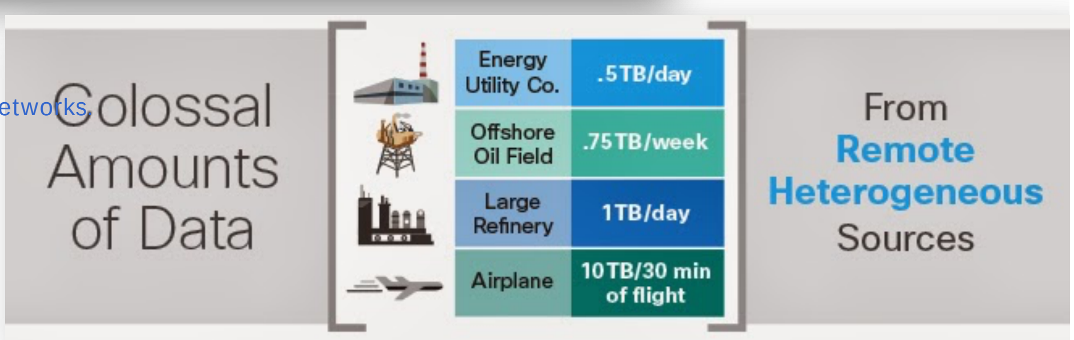
Source: https://www.energomonitor.com/insight/edge-computing-fog-computing-benefits-of-fogging/

Smart Grid: Before Fog

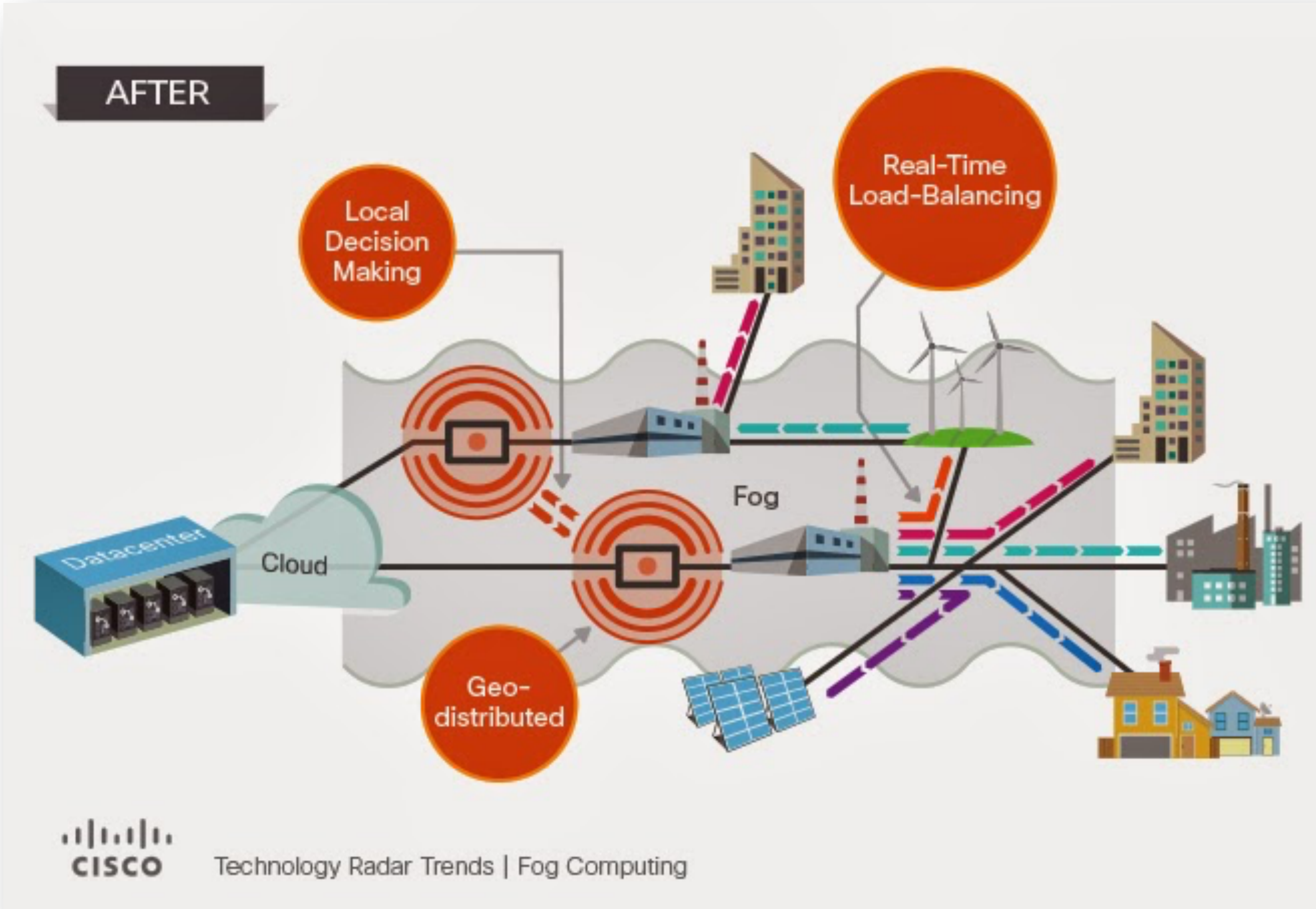
The Grid if everything is using high bandwidth with cecullar that's going to be problem.



If you don't have fog computing then you have data conjestion on cecullar networks.



Smart Grid: After Fog+Edge

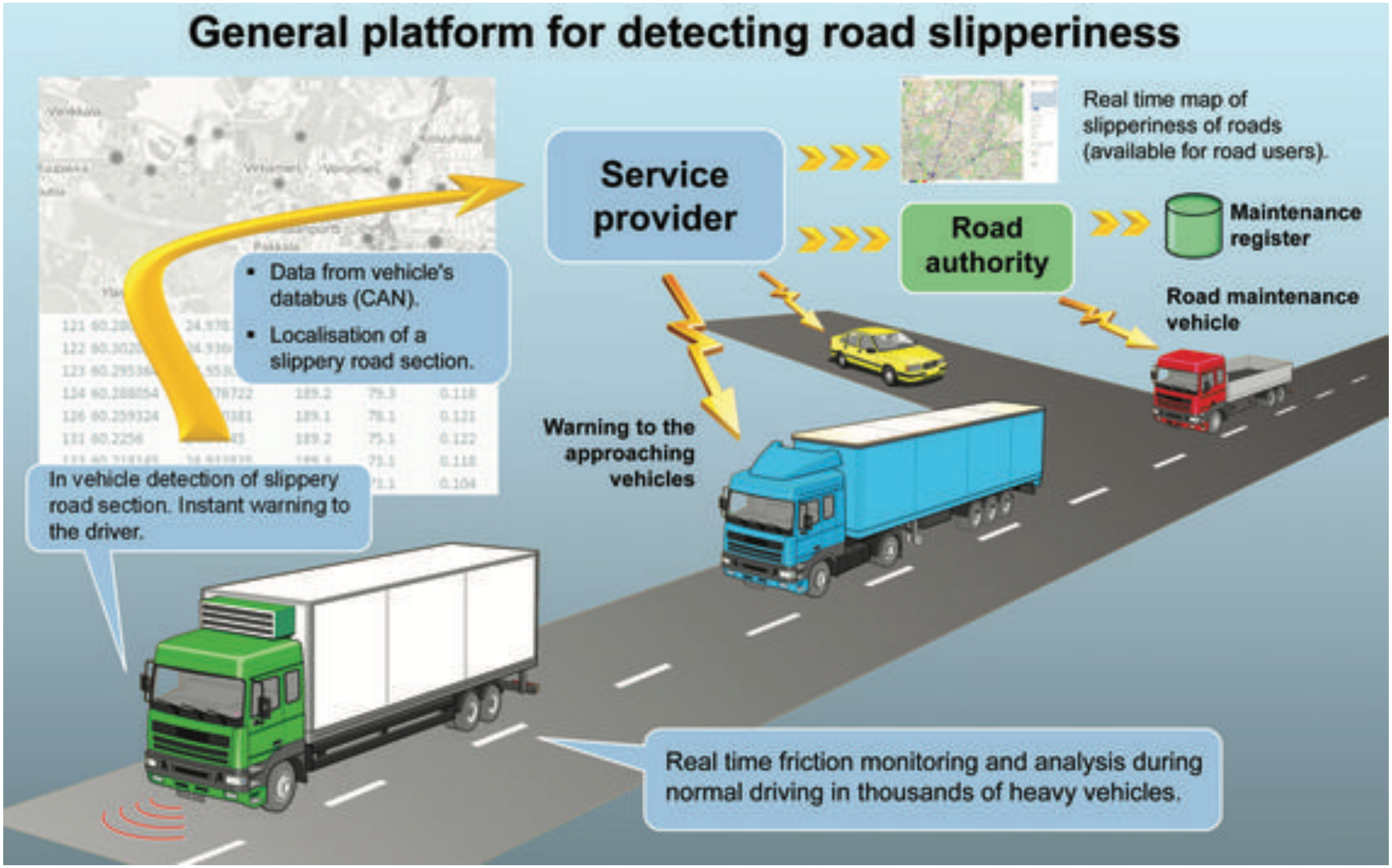


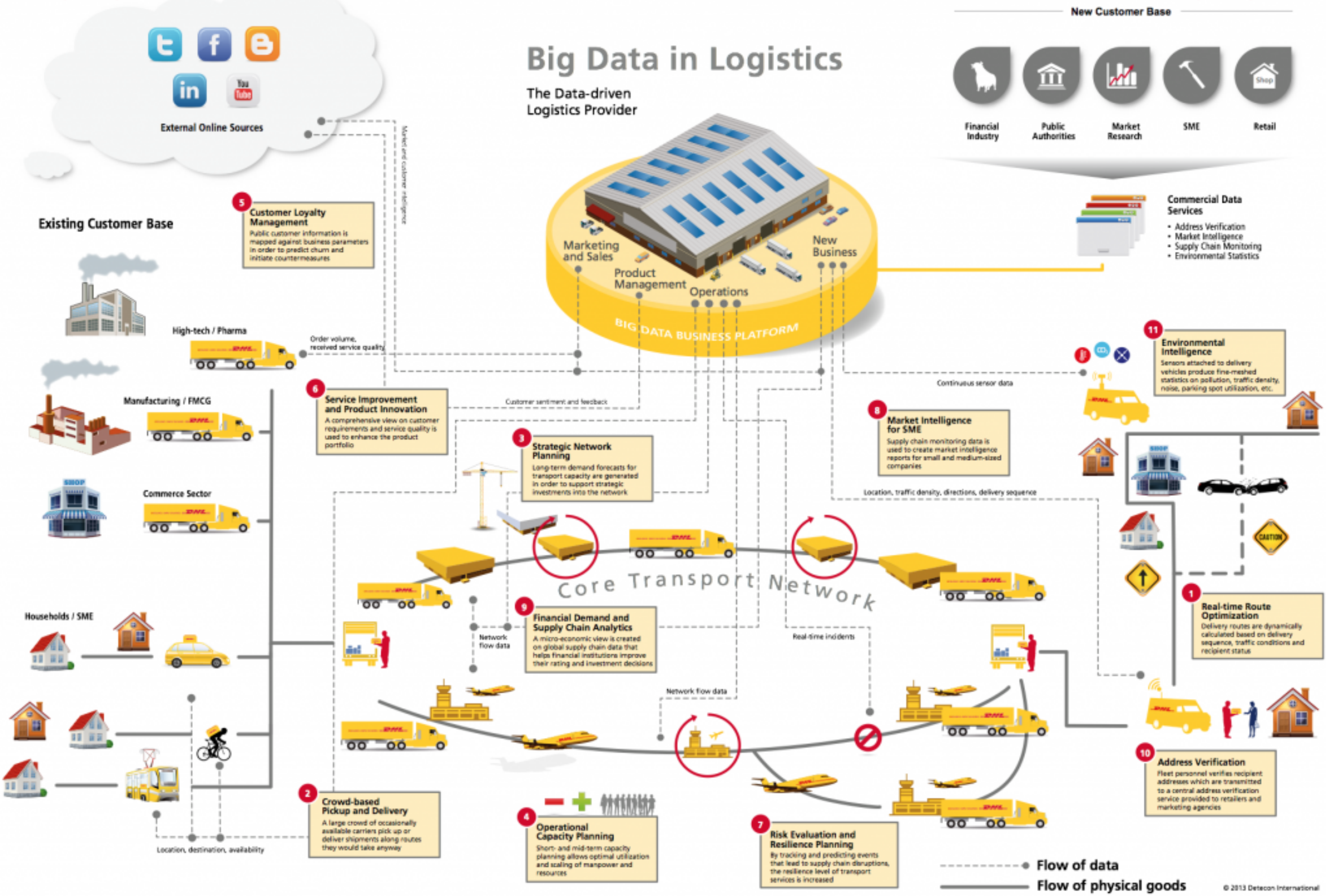
Smart Healthcare: With Fog+Edge

Source: http://www.sciencedirect.com/science/article/pii/S0167739X17302121

This is done by the network of trucks. If you have a fleet of trucks that collectively sending as fog note. They don't need to send all the data to the cloud.

Smart Road-Monitoring: Fog+Edge





Big Data in Logistics
The Data-driven Logistics Provider
New Customer Base
Financial Industry
Public Authorities
Market Research
SME
Retail
Commercial Data Services
• Address Verification
• Market Intelligence
• Supply Chain Monitoring
• Environmental Statistics
External Online Sources
Existing Customer Base
High-tech / Pharma
Manufacturing / FMCG
Commerce Sector
Households / SME
Marketing and Sales
Product Management
Operations
New Business
BIG DATA BUSINESS PLATFORM
Market and customer intelligence
Order volume, received service quality
Customer sentiment and feedback
Continuous sensor data
Location, traffic density, directions, delivery sequence
Network flow data
Real-time incidents
Location, destination, availability
Core Transport Network
1 Real-time Route Optimization
Delivery routes are dynamically calculated based on delivery sequence, traffic conditions and recipient status
2 Crowd-based Pickup and Delivery
A large crowd of occasionally available carriers pick up or deliver shipments along routes they would take anyway
3 Strategic Network Planning
Long-term demand forecasts for transport capacity are generated in order to support strategic investments into the network
4 Operational Capacity Planning
Short- and mid-term capacity planning allows optimal utilization and scaling of manpower and resources
5 Customer Loyalty Management
Public customer information is mapped against business parameters in order to predict churn and initiate countermeasures
6 Service Improvement and Product Innovation
A comprehensive view on customer requirements and service quality is used to enhance the product portfolio
7 Risk Evaluation and Resilience Planning
By tracking and predicting events that lead to supply chain disruptions, the resilience level of transport services is increased
8 Market Intelligence for SME
Supply chain monitoring data is used to create market intelligence reports for small and medium-sized companies
9 Financial Demand and Supply Chain Analytics
A micro-economic view is created on global supply chain data that helps financial institutions improve their rating and investment decisions
10 Address Verification
Fleet personnel verifies recipient addresses which are transmitted to a central address verification service provided to retailers and marketing agencies
11 Environmental Intelligence
Sensors attached to delivery vehicles produce fine-meshed statistics on pollution, traffic density, noise, parking spot utilization, etc.
Flow of data
Flow of physical goods

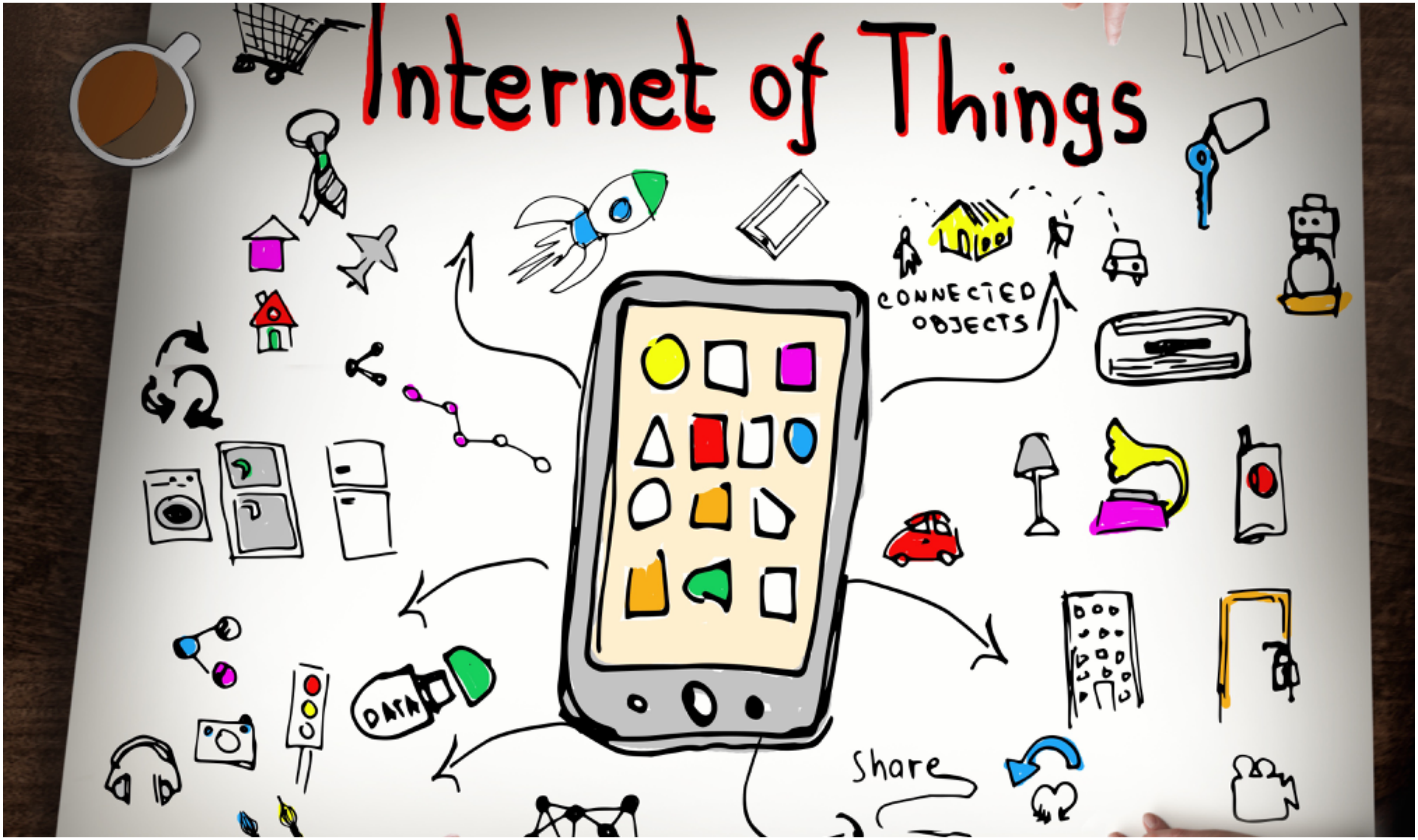
Edge Computing: Industry View (DHL)

- DHL smart warehouse
- Monitor operational activities within their warehouses in real-time using IoT technologies
- Visualize the operations within a warehouse through the use of heat-maps
- Indoor positioning of various connected assets such as handheld scanners and monitoring utilization of Materials Handling Equipment (MHE), combined with data from the Warehouse Management System to understand what tasks are being allocated to whom
- Improve efficiency and health & safety by giving accurate data on response rates, record any incidents and measure overall utilization of equipment and space.

The FOG notes fits to DHL, UPS just fits to stadium. The players are now on flg notes, and the stadium are on fog nodes as well.

Additional References

- Comparing different cloud platforms (EC2, Azure, Google AE)
  http://forwardthinking.pcmag.com/software/282947-cloud-thinking-amazon-microsoft-and-google
- Tech Debate: Cloud: Public or private?
  http://www.networkworld.com/community/tech-debate-private-public-cloud
- Economics of cloud computing
  http://www.boozallen.com/media/file/Economics-of-Cloud-Computing.pdf
- Is cloud computing really cheaper?
  http://www.forbes.com/sites/reuvencohen/2012/08/03/is-cloud-computing-really-cheaper/
- Picking cloud platforms: What is right for your business?
  http://www.eweek.com/cloud/picking-cloud-platforms-what-is-right-for-your-business/
- Edge Computing or Fog Computing, or, on the benefits of fogging
  https://www.energomonitor.com/insight/edge-computing-fog-computing-benefits-of-fogging/
- Fog Computing and the Internet of Things: Extend the cloud to where the things are
  https://www.cisco.com/c/dam/en_us/solutions/trends/iot/docs/computing-overview.pdf



Cloud/Fog/Edge Computing

18-738 Sports Technology

Priya Narasimhan
ECE Department
Carnegie Mellon University
@yinzcampriya