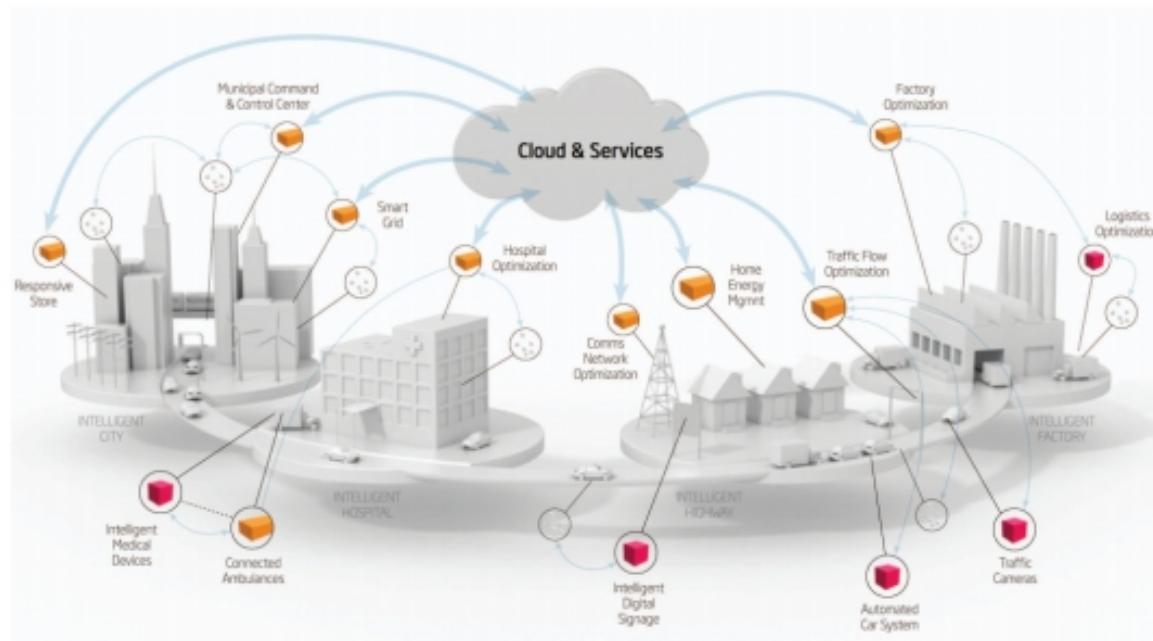


Cloud/Fog/Edge Computing  
18-738 Sports Technology

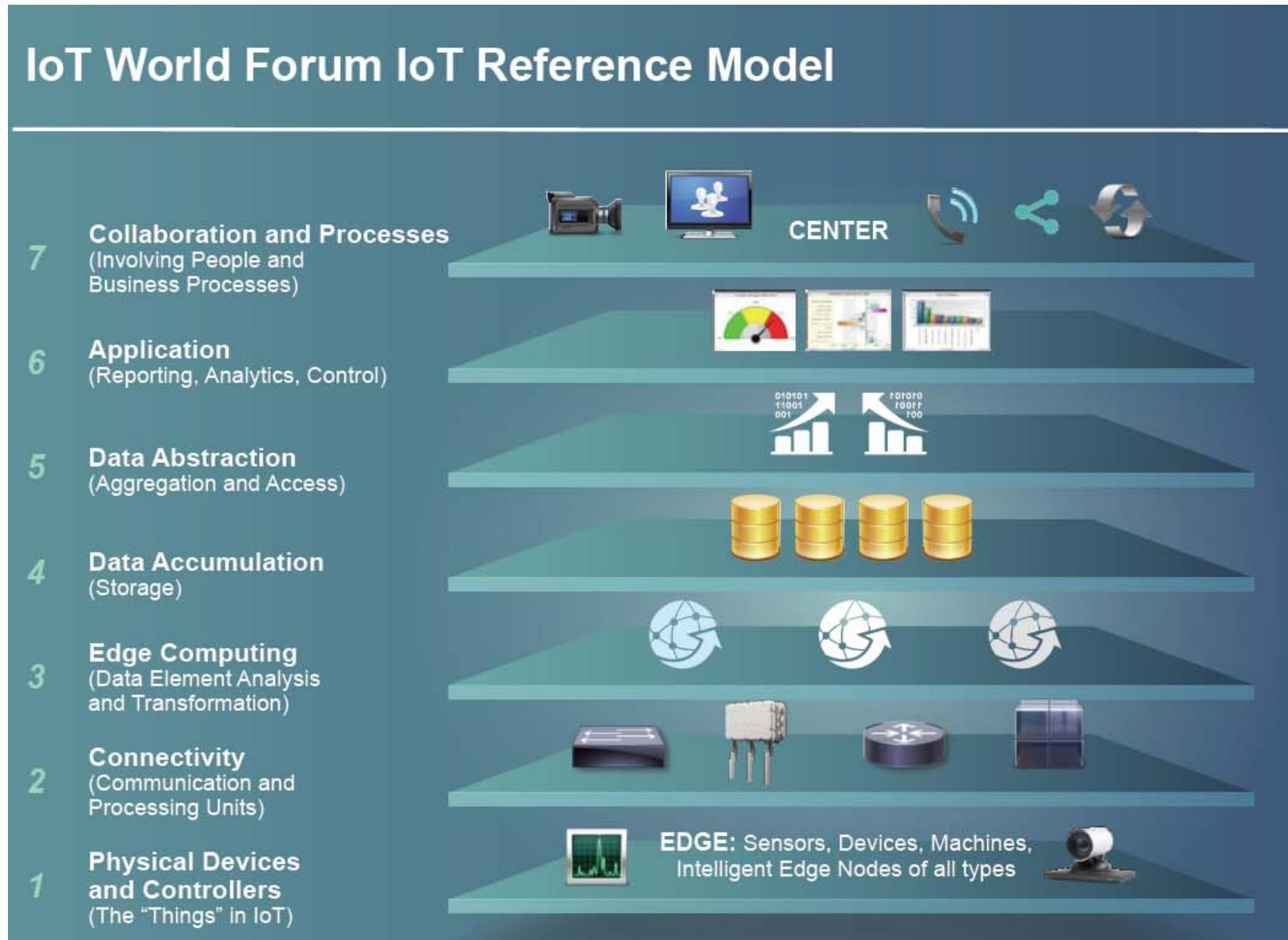
Priya Narasimhan  
ECE Department  
Carnegie Mellon University  
@yinzcampriya

# IoT and the Cloud

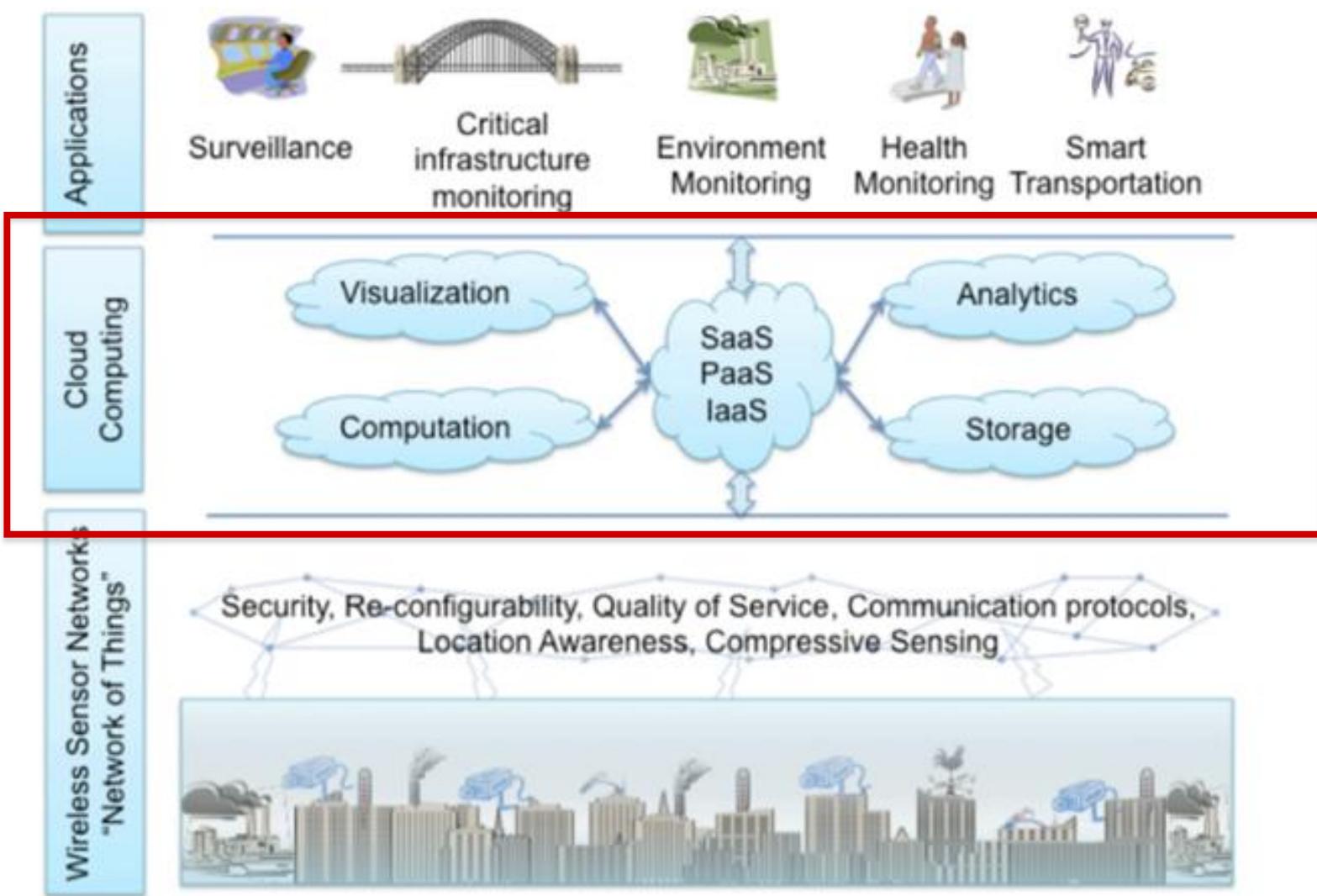
- Cloud computing is a fundamental enabler
- Allows for scalable, on-demand resources
- Allows for big-data analytics
- Allows for offloading compute-intensive tasks to beefier resources
- Allows for anytime-anywhere access to data and results



# IoT Architecture/Model



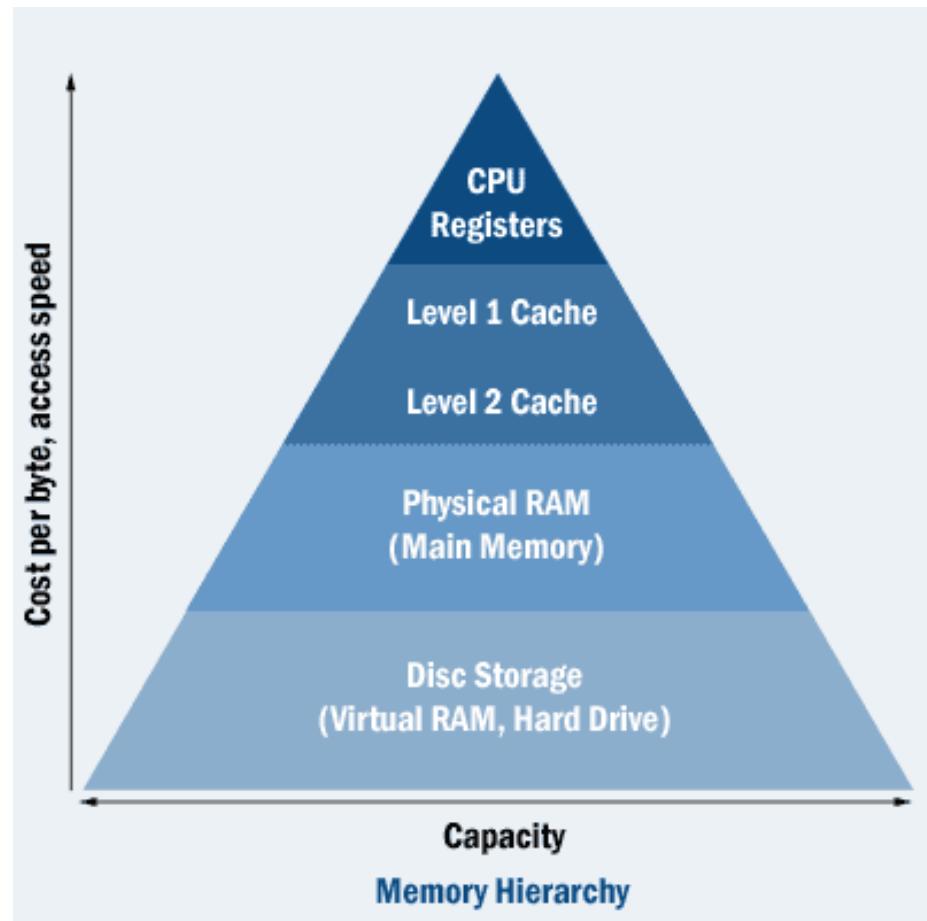
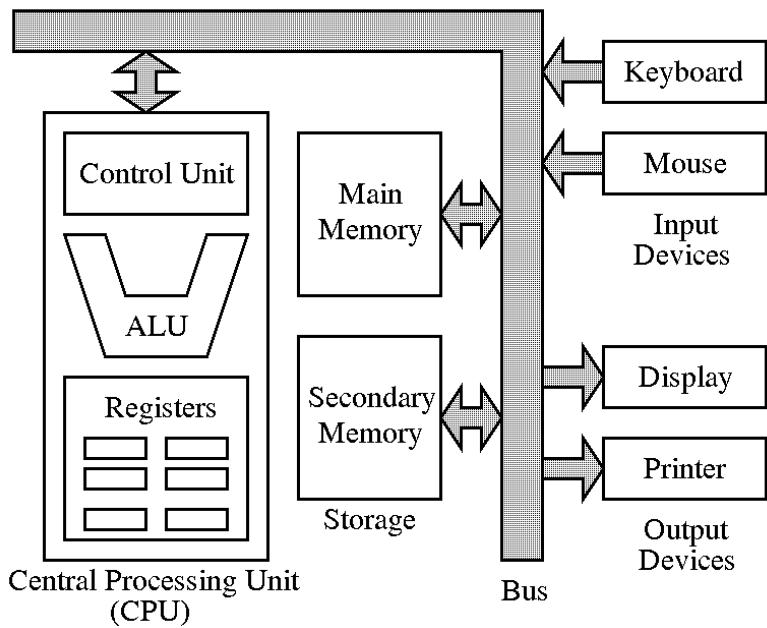
# IoT and the Cloud



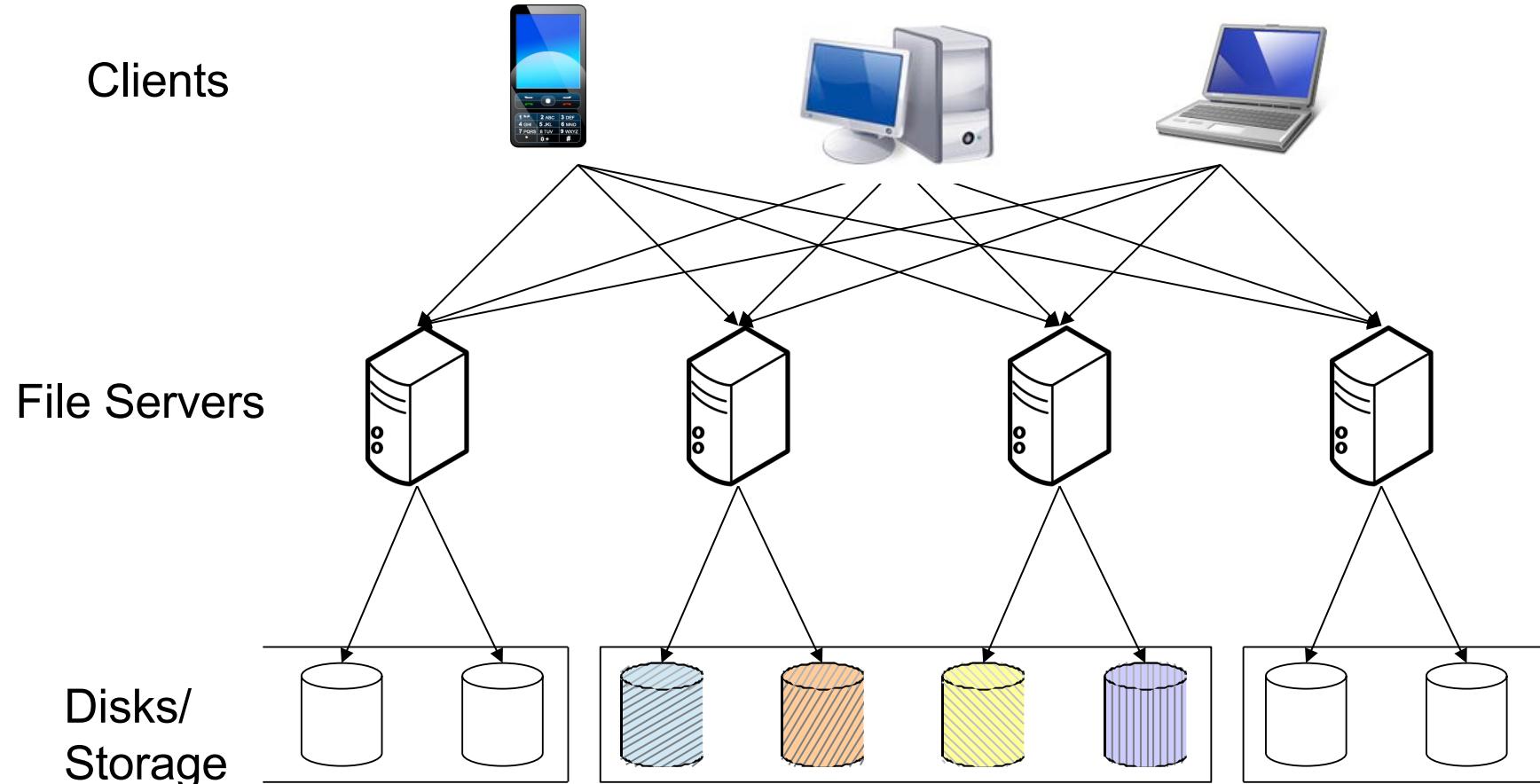
# Questions to Answer

- 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?
- What is a private vs. public cloud?

# Basic Computer Architecture



# Basic System Architecture



# Virtualization

There are multiple processes on the hardware. This is like a reservation. You reserved part of the hardware for the virtual machine.

e.g. Amazon Instances is virtual machine -> they are not full os.

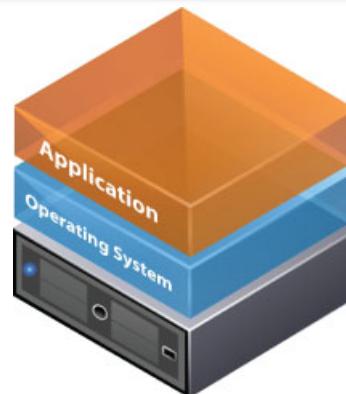
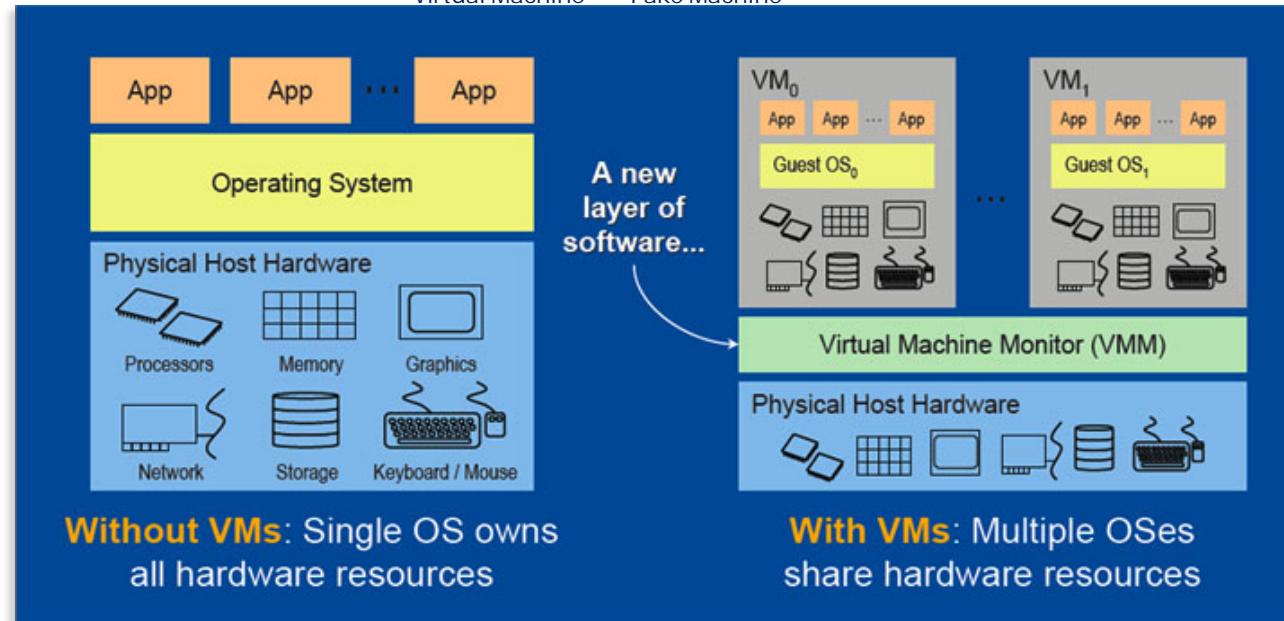
Virtual Machine = Slice of OS + Slice of Hardware = guaranteed reserved resources -> mini os.

1. guaranteed resource reservation

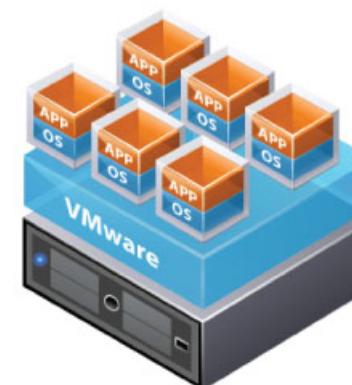
2. isolation from other hardware ... when fails, it's not going to fail the entire system

Bare Metal = Hardware

"Virtual Machine" = "Fake Machine"



Traditional Architecture



Virtual Architecture



# Cloud Computing

# Outline

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

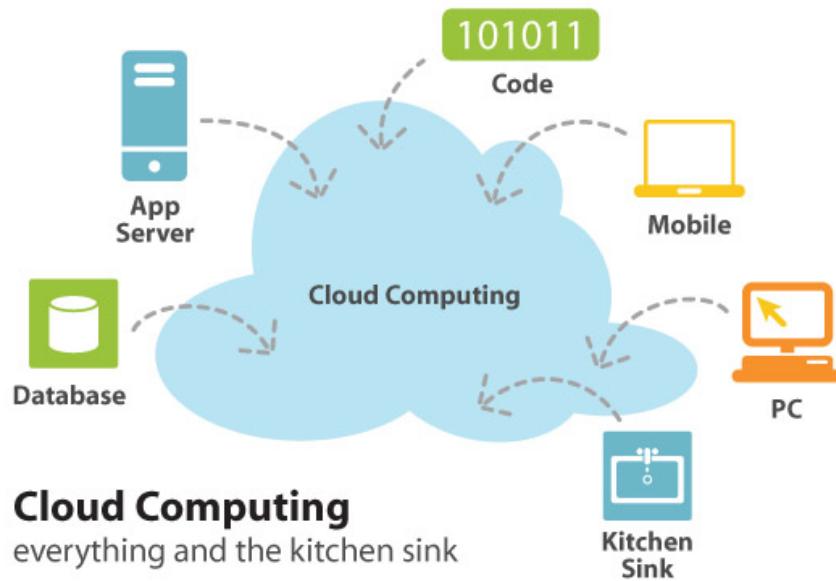
Act of Leasing vs Owning infrastructure:

Leasing: leave problems to the owner

Owning: every problems you need to fix by yourself

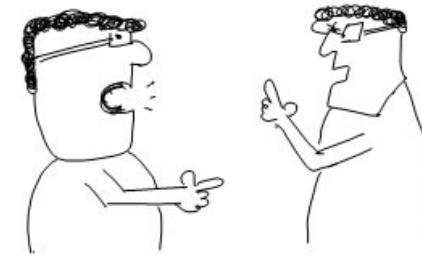
Total Cost Ownership: everything is on you versus the owner

# What is Cloud Computing?



WHERE THE HECK  
IS MY DATA?

ITS THERE, UP  
IN THE CLOUDS.



Brainstuck.com

Renting an apartment = managed infrastructure

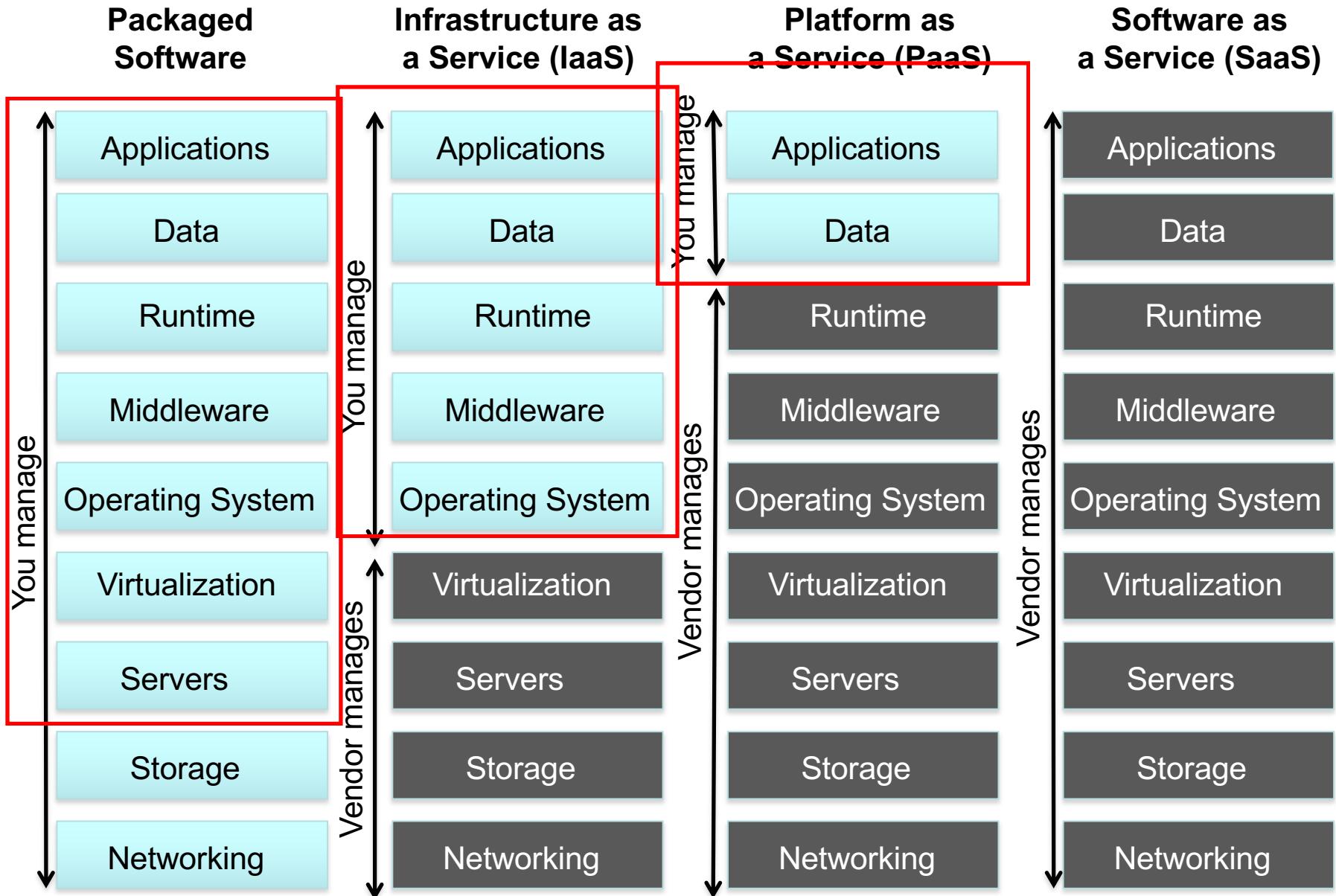
Renting an apartment = furnished apartment, unfurnished, utilities paid/unpaid

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

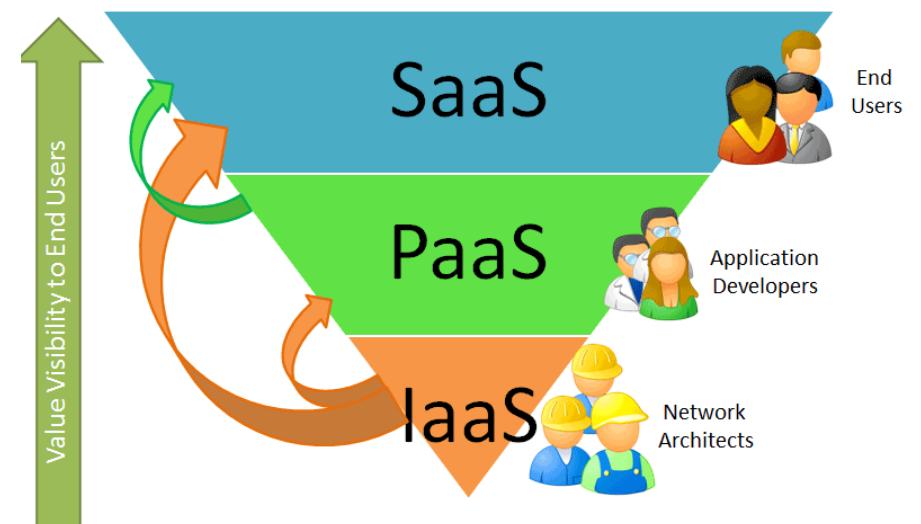
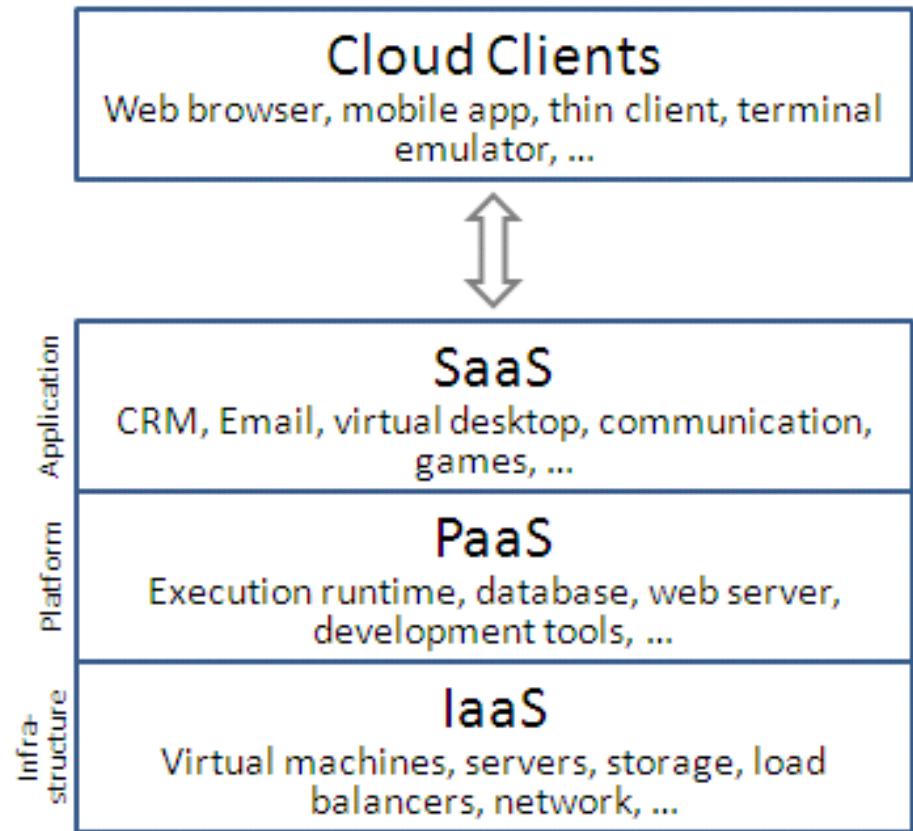
Renting an apartment = flavors of cloud computing

Owning your own house = building your own data center

"Pay as you go" also called "utility computing" -- just like electricity.



# 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.

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

Power of numbers

New business opportunities due to flexibility and scale



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

# 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

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

Lack of performance guarantees

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

Responsiveness

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

Regulations



# Cloud Computing

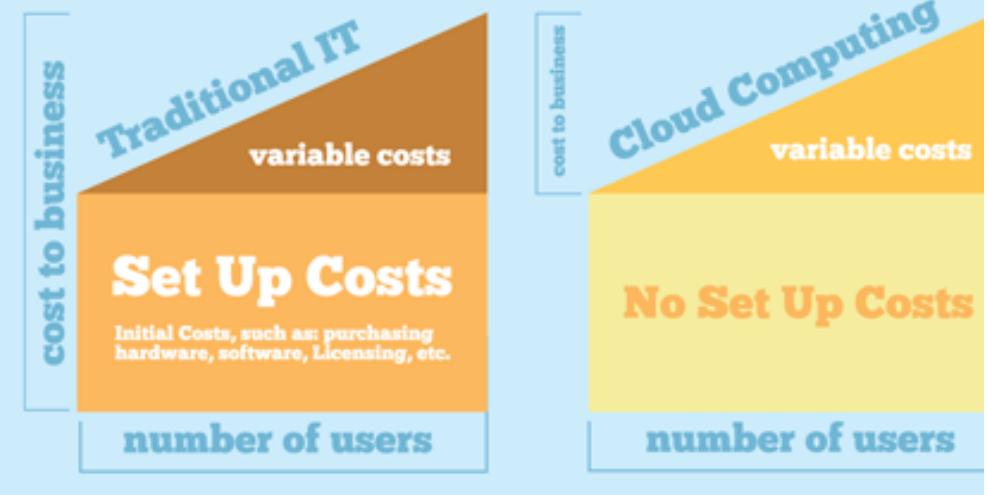
what is that, and why should i care?



One reason Amazon.com started their journey into the clouds is they were only using about 10% of their available server resources, with about 90% more or less just “going to waste” waiting for the occasional service spikes.

## What are the benefits?

- Anywhere/Anytime Access to Data
- Improved Reliability and Security of Data
- Wider Access to Free Software
  - Software that's up-to-date with less malware



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
- 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

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

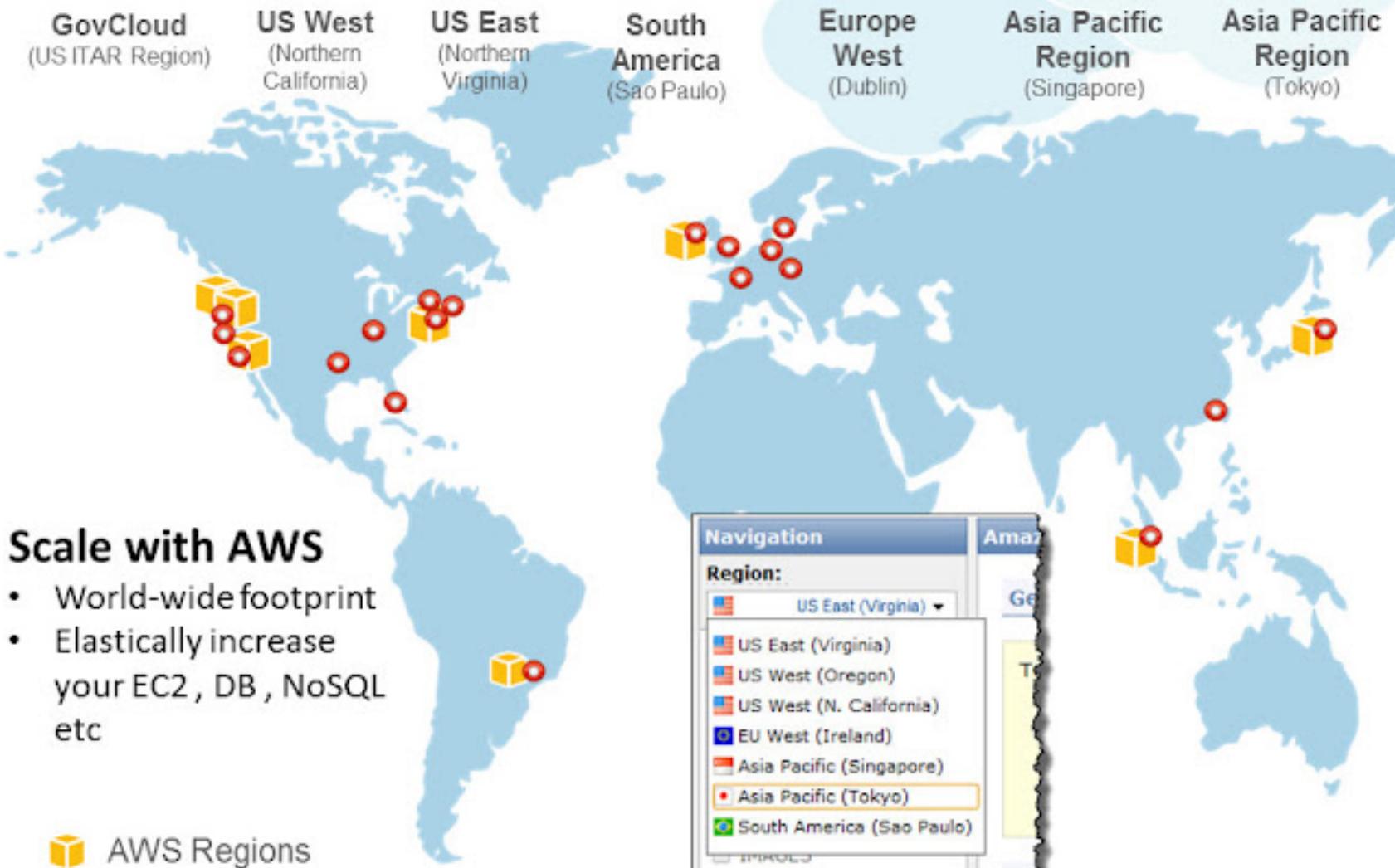
Overall objective  
Use just as much resources as you need

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



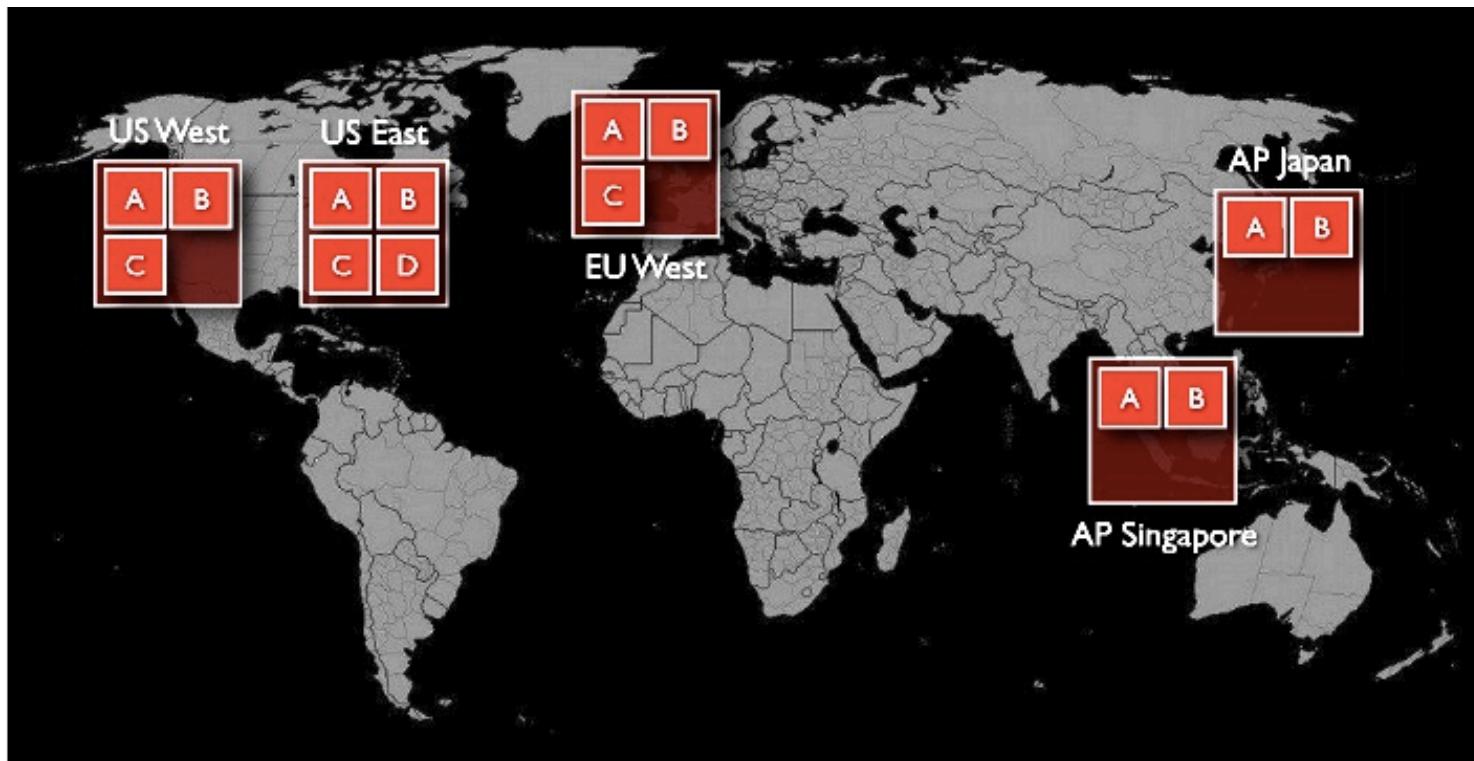
## Scale with AWS

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

Image Source: AWS

# Availability Zones (AZ)

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



Source: Simone Brunozzi, 2011

# AWS Ecosystem



# Alternative PaaS Platforms

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

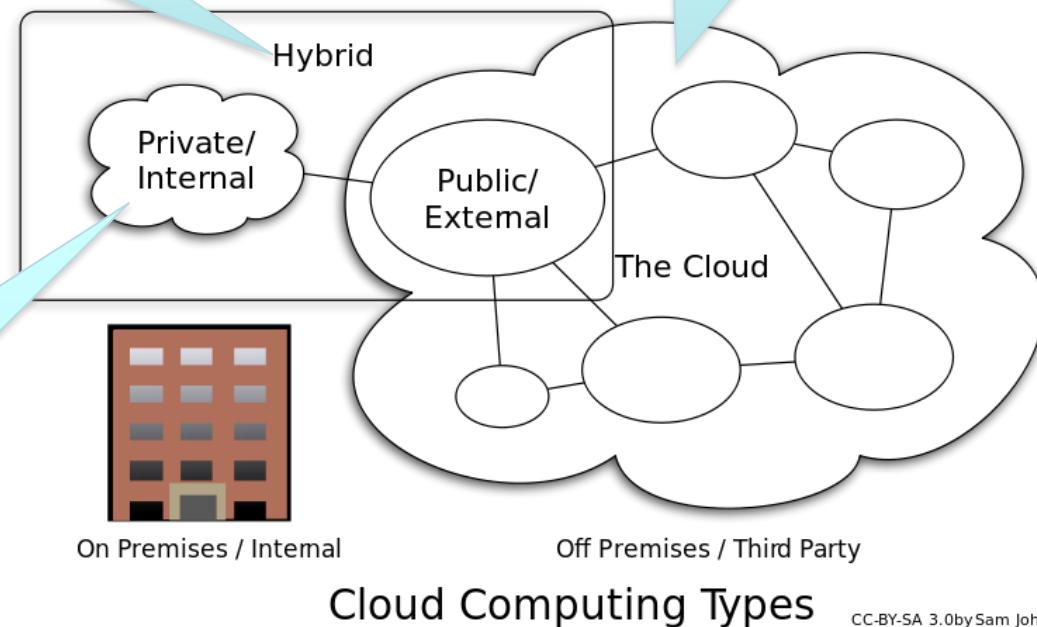
# Public, Private, Hybrid Clouds

Private cloud = cloud computing inside organization

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



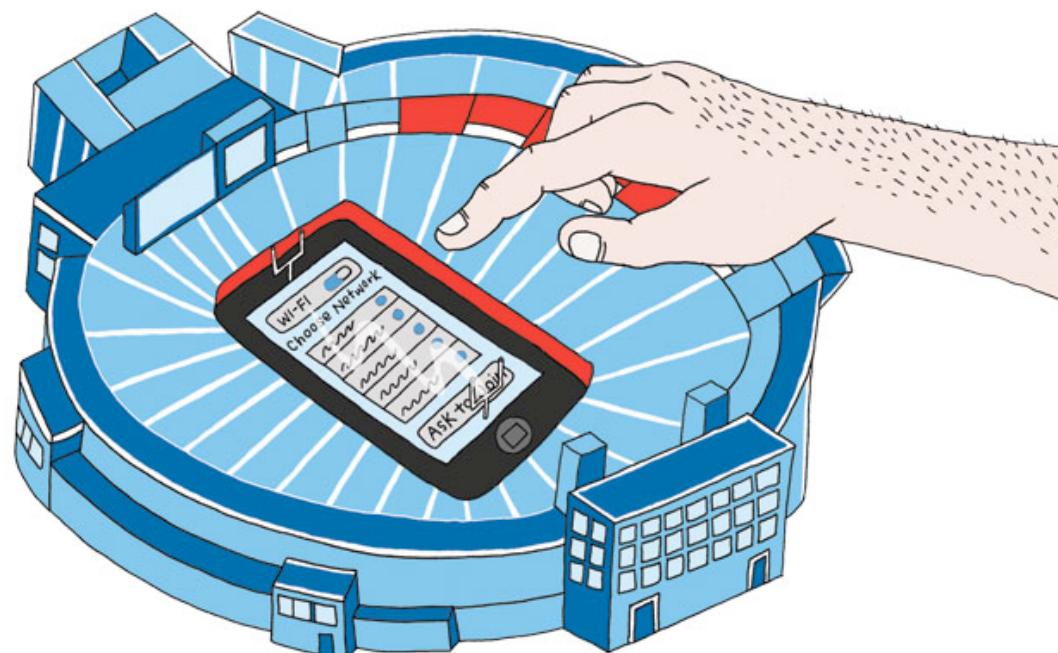
CC-BY-SA 3.0 by Sam Johnston

# 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

# 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

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

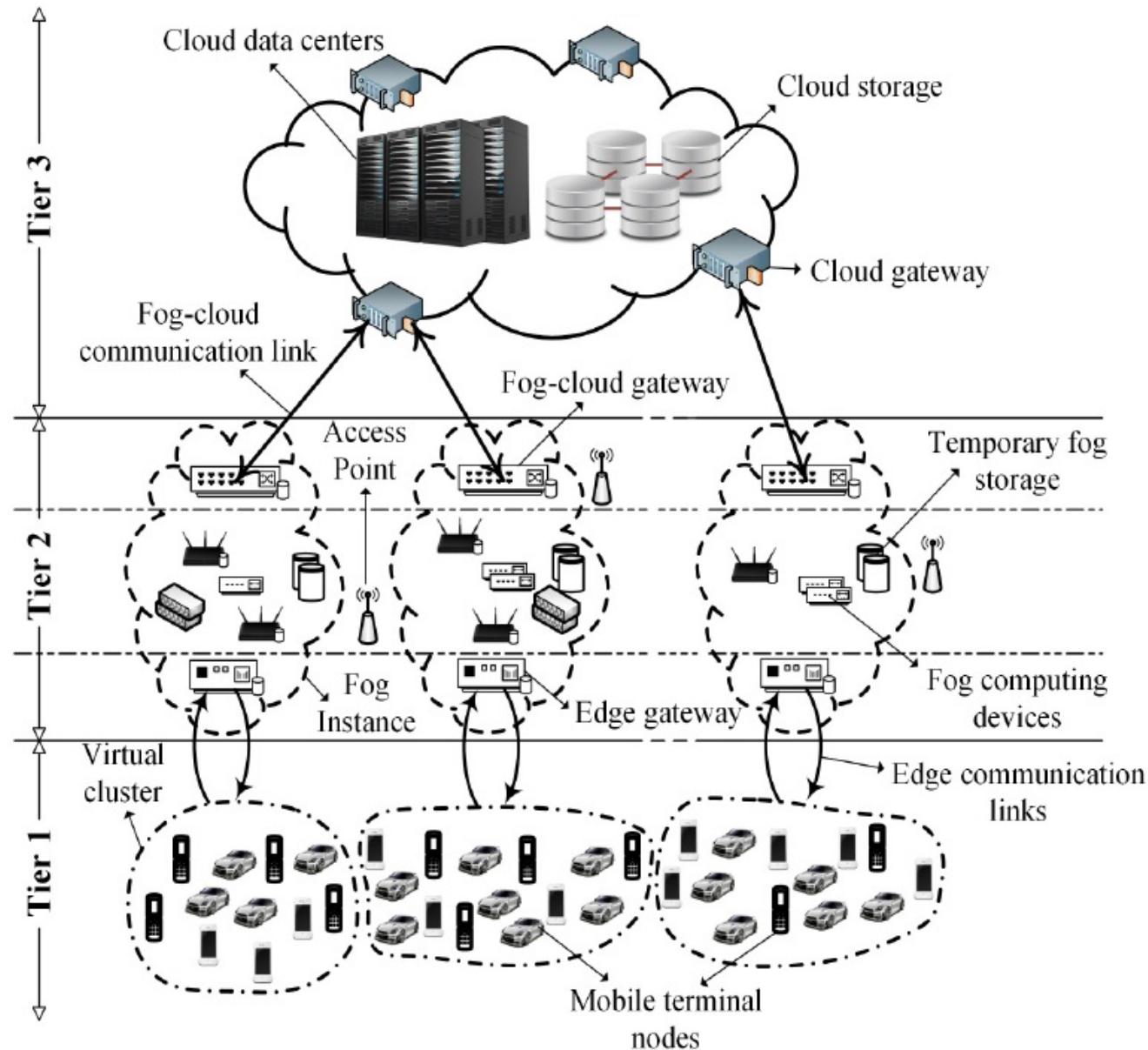
The argument is that "*if you can do computation locally and save bandwidth, why ship data to a remote computation server?*"

- 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

- 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



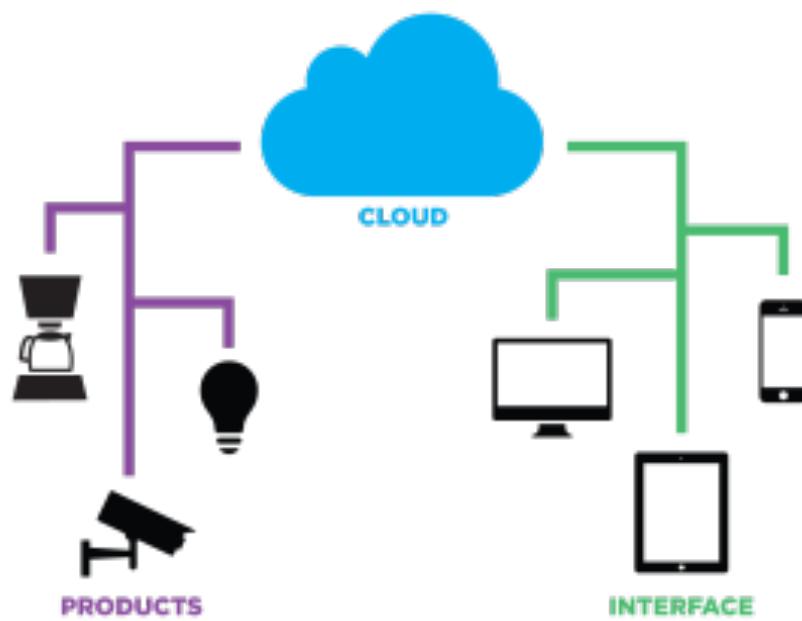
# Digging a little deeper

- Virtualized platform that
  - 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

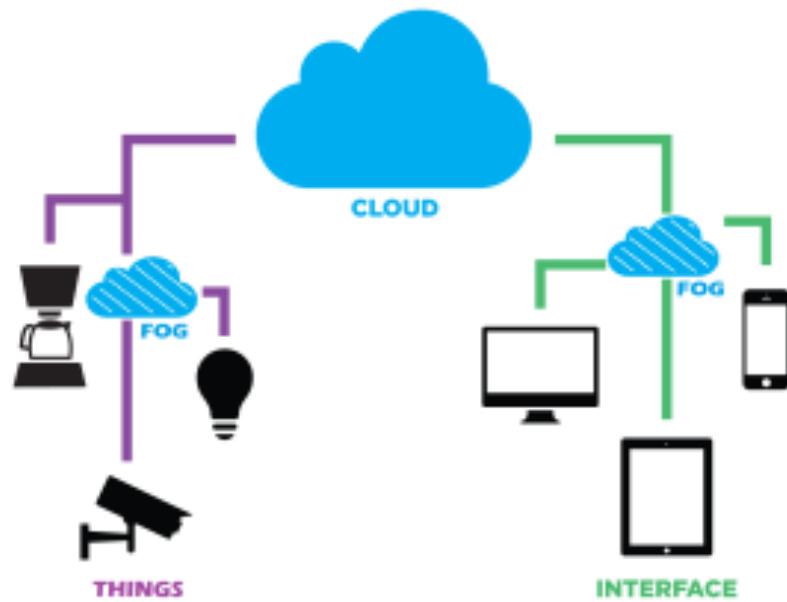
The stores connects to local servers, they do local computing..

The local servers then in turns connects to an Amazon server.

# 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 compressed 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

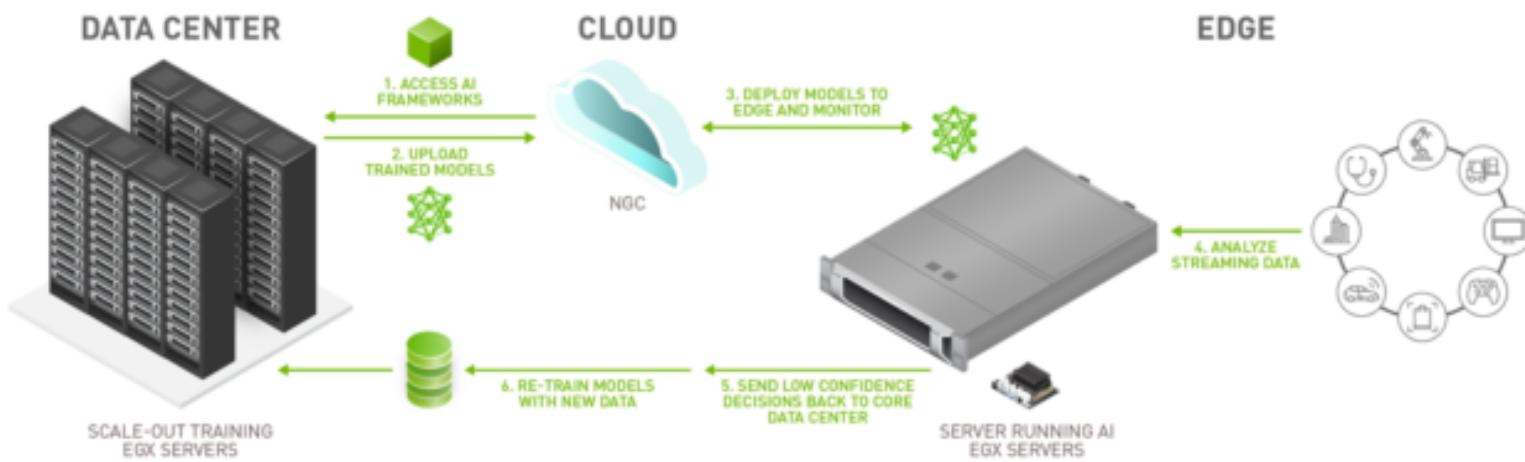
Another layer:  
Cloud  
Fog  
Edge - Groups of sensors.

# Edge Computing

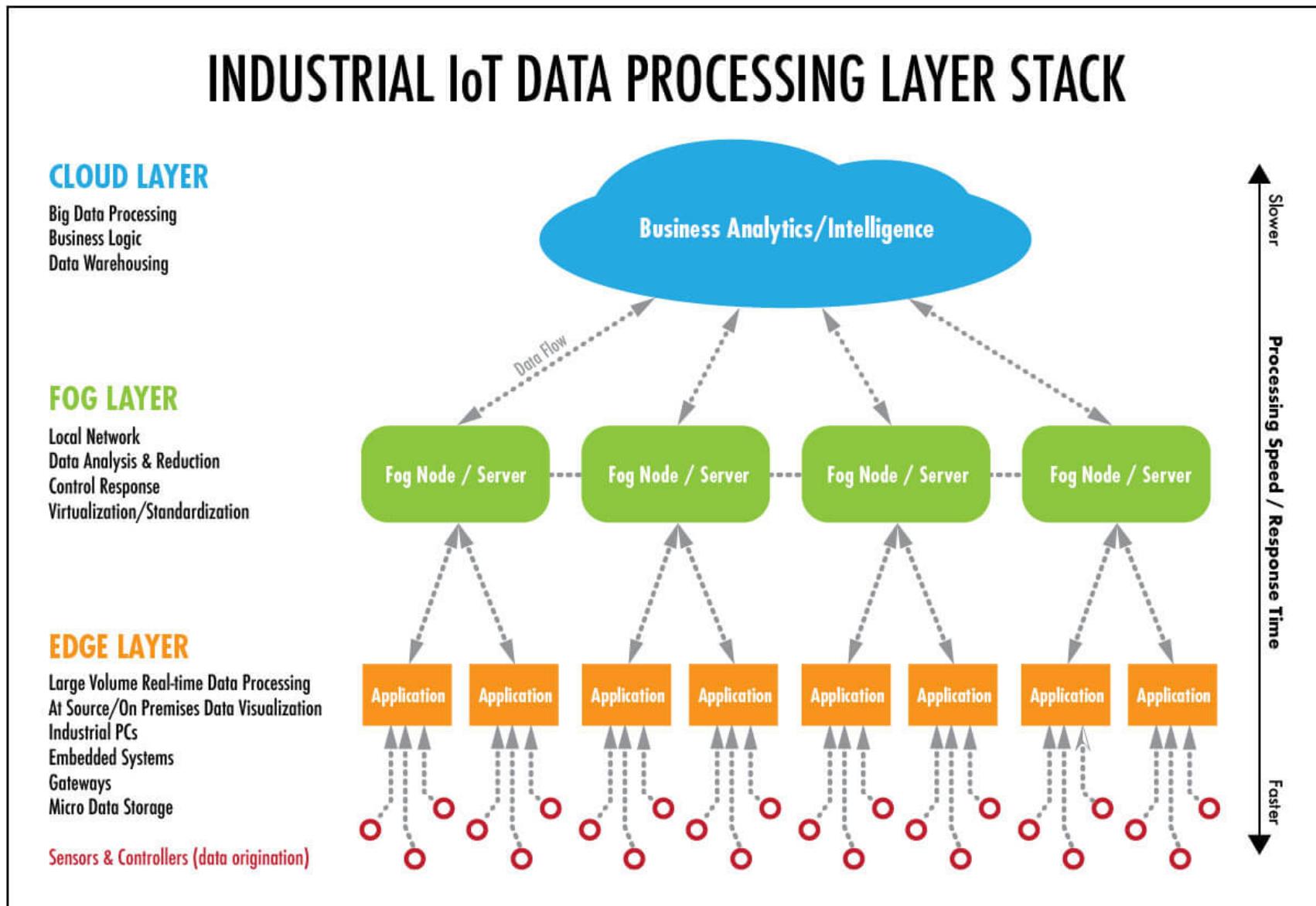
Edge computing is the practice of moving compute power physically closer to where data is generated, usually an IoT device or sensor.

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.



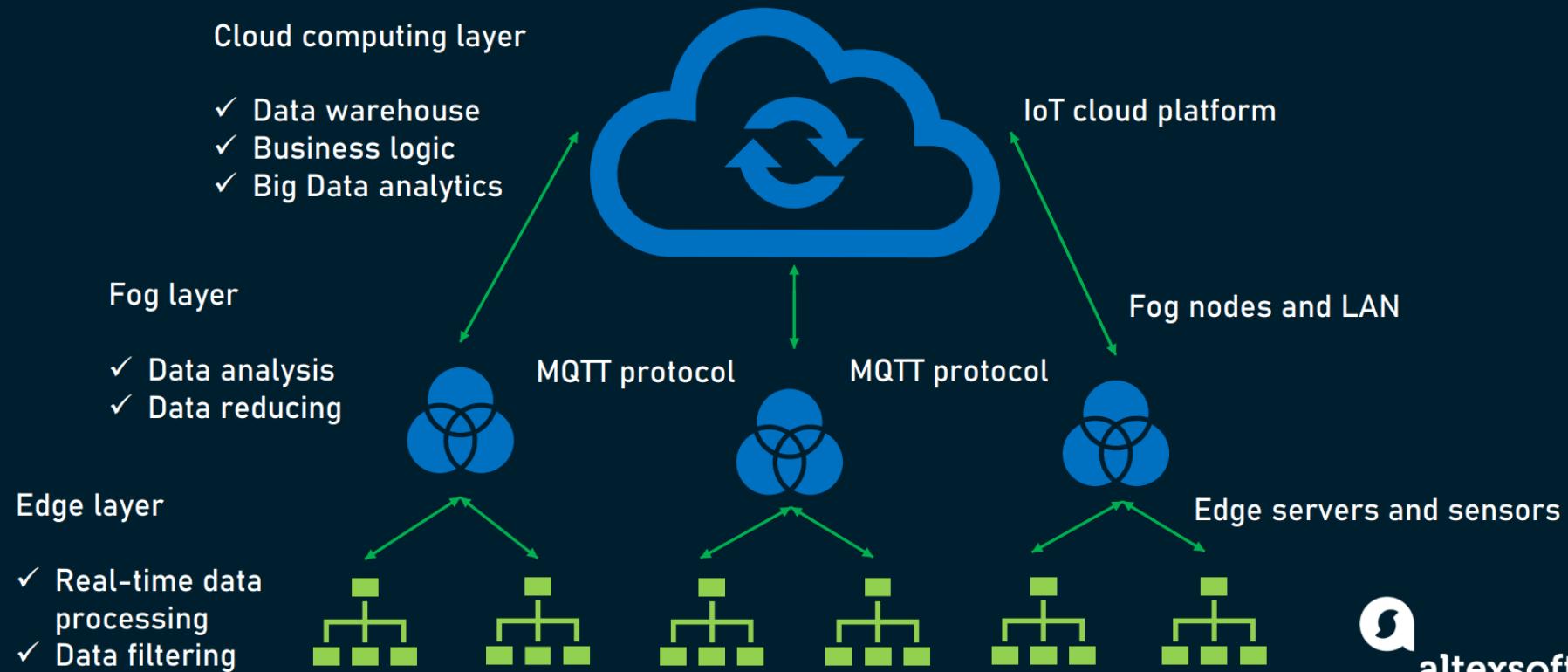
# Edge Computing: Industry View



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

# Edge Computing: Another Industry View

## EDGE COMPUTING ARCHITECTURE

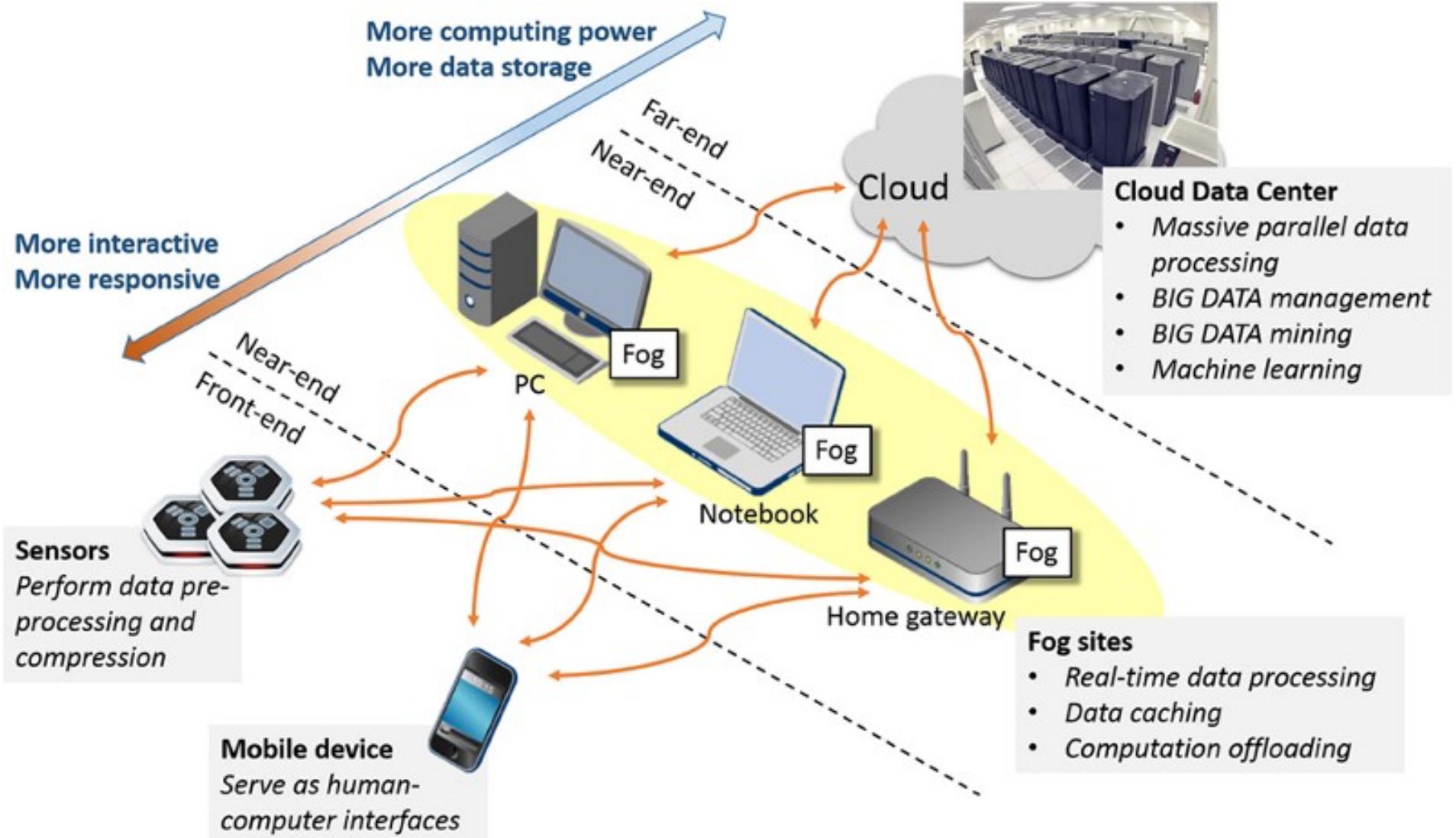


# Edge Computing: Industry View (UPS)

- 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.

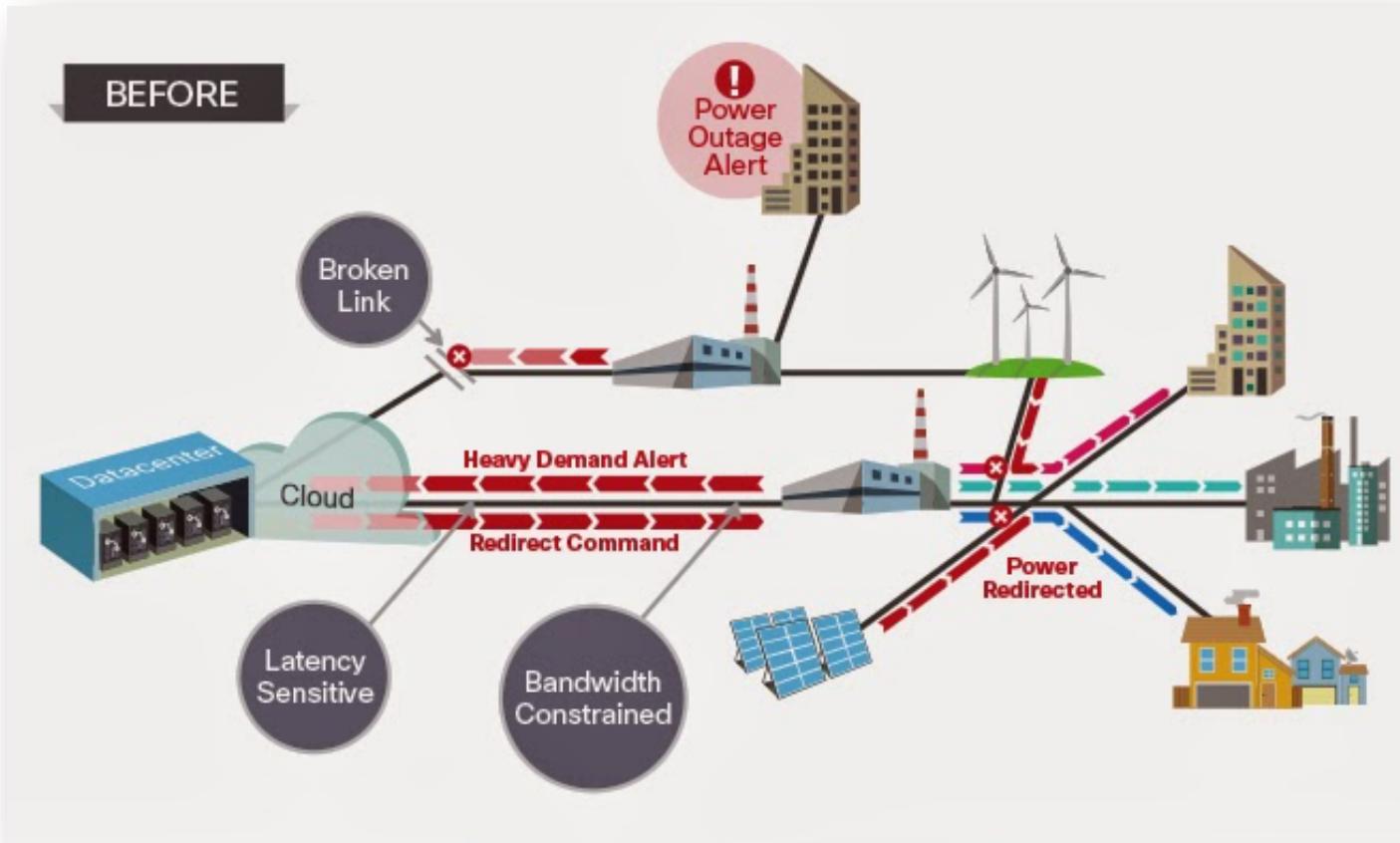


# Cloud+Fog+Edge



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

# Smart Grid: Before Fog



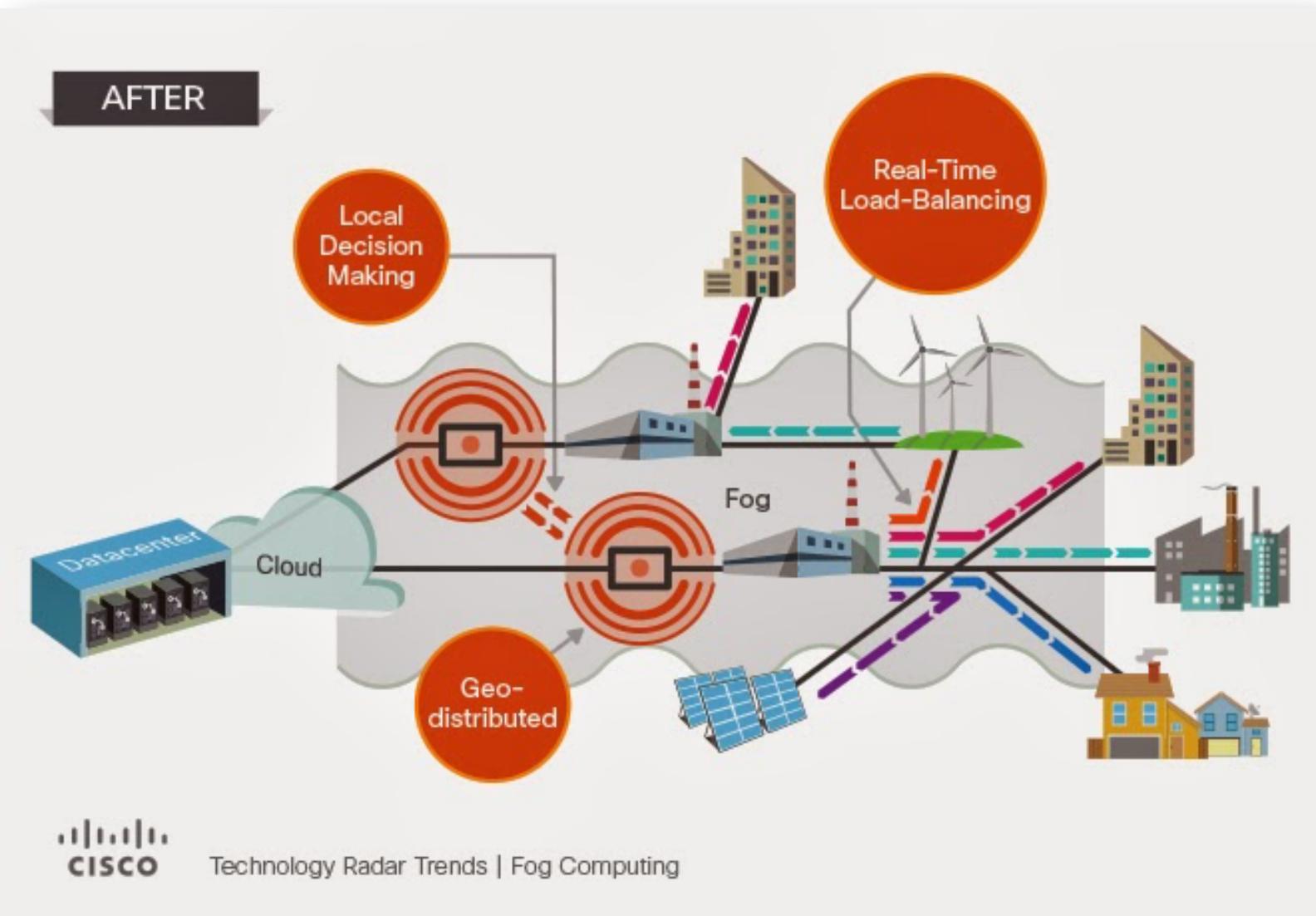
If you don't have fog computing then you have data conjetion on cecullar networks.

Colossal  
Amounts  
of Data

|  |                    |                       |
|--|--------------------|-----------------------|
|  | Energy Utility Co. | .5 TB/day             |
|  | Offshore Oil Field | .75 TB/week           |
|  | Large Refinery     | 1TB/day               |
|  | Airplane           | 10TB/30 min of flight |

From  
Remote  
Heterogeneous  
Sources

# Smart Grid: After Fog+Edge

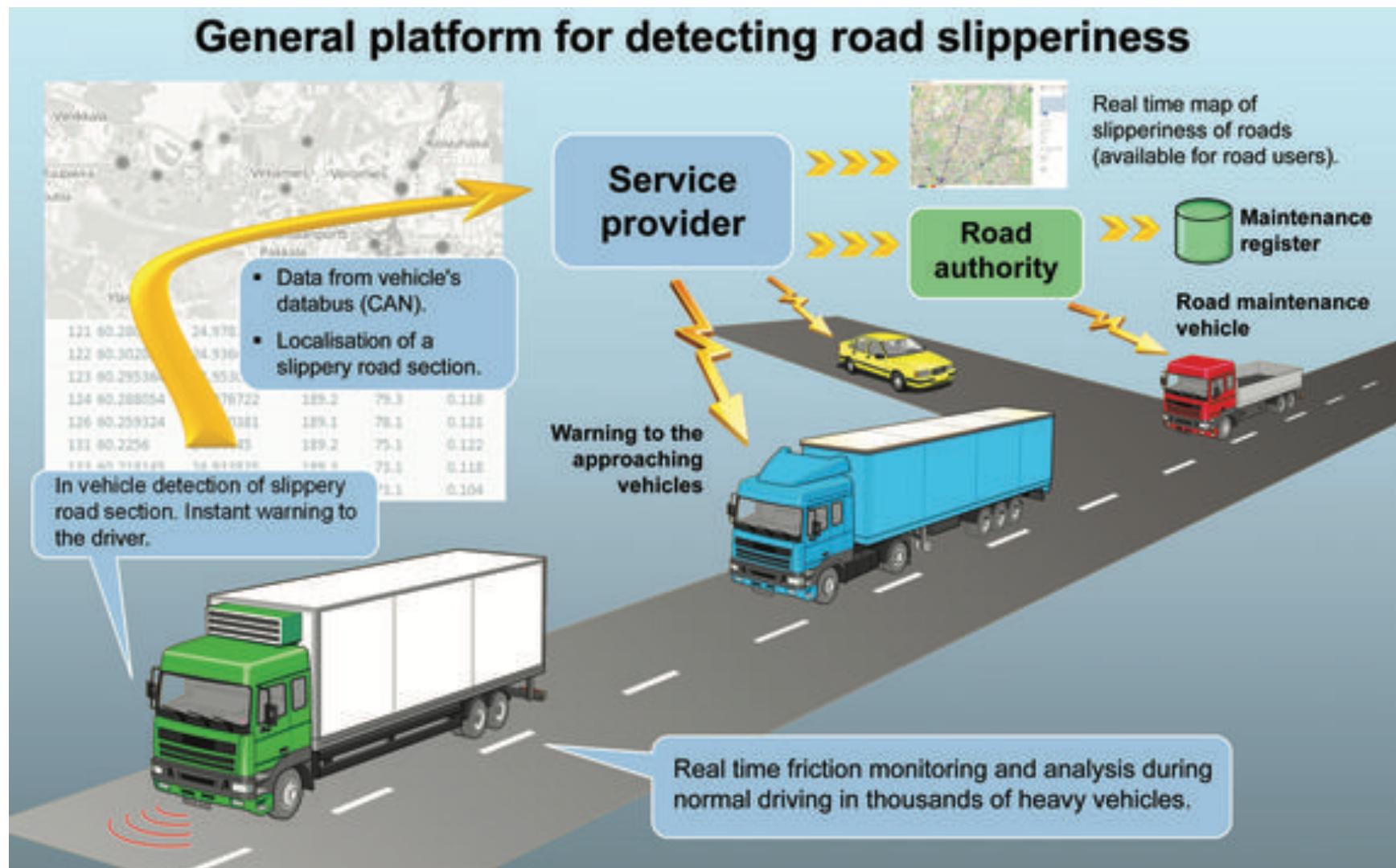


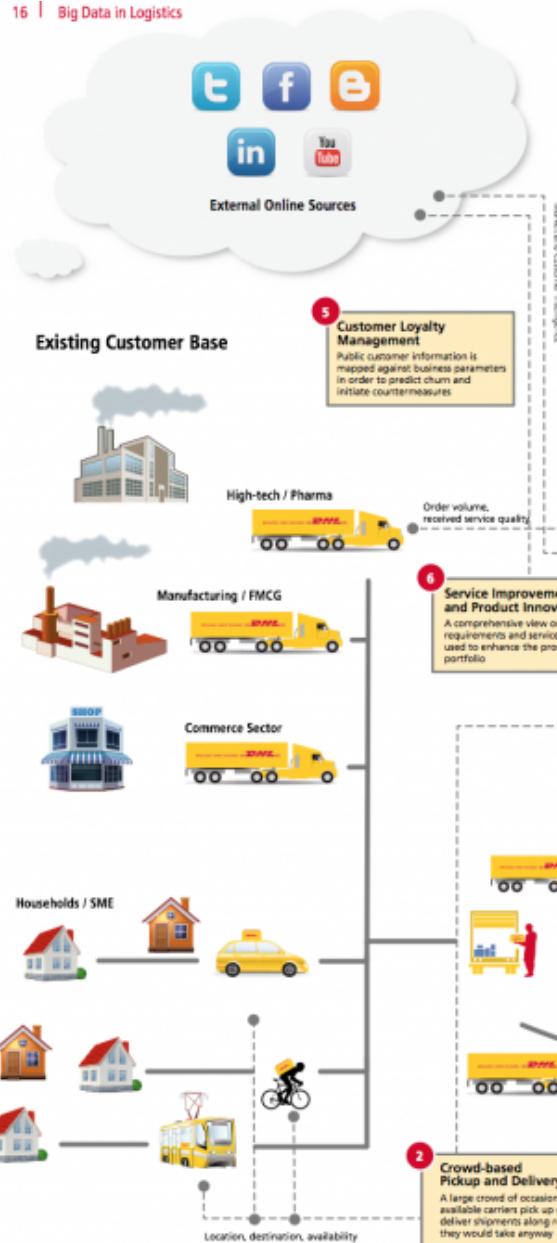
# Smart Healthcare: With Fog+Edge



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.

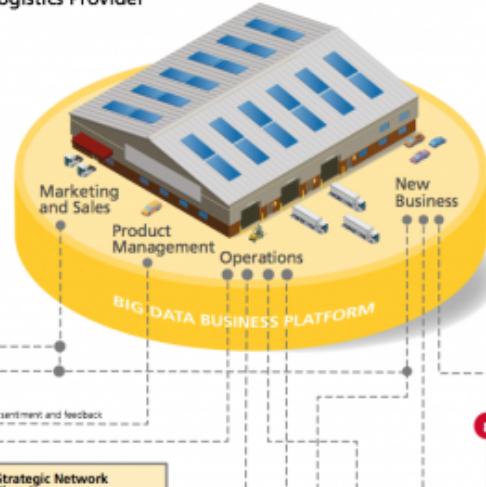
# Smart Road-Monitoring: Fog+Edge





## Big Data in Logistics

The Data-driven Logistics Provider



New Customer Base



Flow of data  
Flow of physical goods

# Edge Computing: Industry View (DHL)

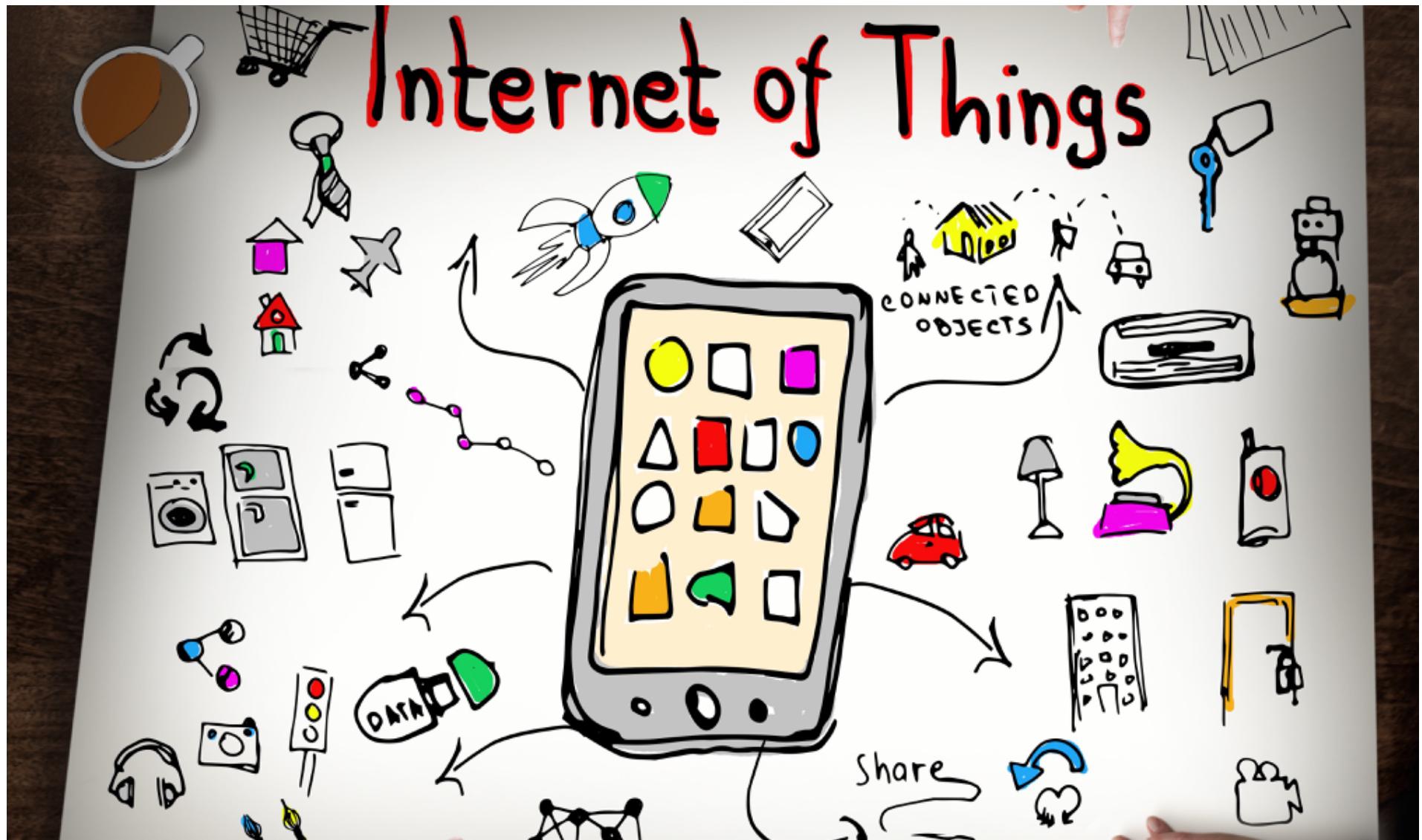
- DHL smart warehouse
- Monitor operational activities within their warehouses in real-time using IoT technologies
- Visualize the operations within a warehouse through the use of heat-maps
- Indoor positioning of various connected assets such as handheld scanners and monitoring utilization of Materials 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](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