Service-oriented and Cloud Computing – Recap, Trends and Focus Points

Hong-Linh Truong
Faculty of Informatics, TU Wien

truong@dsg.tuwien.ac.at
http://dsg.tuwien.ac.at/staff/truong
@linhsolar

Goals

- Understand current advanced development and trends in services computing and cloud computing
  - With IoT, bigdata/data science, and network functions
- Capture key questions for this course in the following 4 sub-areas
  - Scalable data, services and systems management
  - Elasticity and control
  - Big data analytics
  - End-to-end engineering analytics

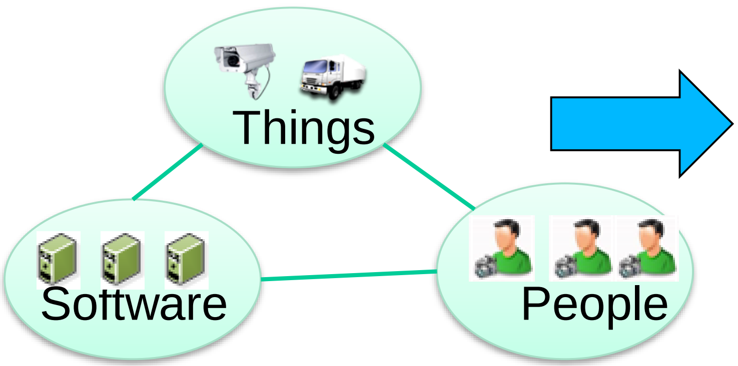
Recap

SERVICES AND CLOUD COMPUTING

Services Computing

- Services offer well-defined interfaces for consumers to
  - access resources: *contents, things, machines, and people*
  - provide functions*: computation, sensing, actuating, analytics, etc.*
  - offer diverse types of business models: *pay-per-use, and subscription*
- Service‘s technical interfaces: SOAP, REST, etc.
- Services are associated with and characterized by scalability, reliability, elasticity, etc.
- Services are provisioned in distributed systems

Services computing



Large-scale, distributed computational infrastructure and software systems

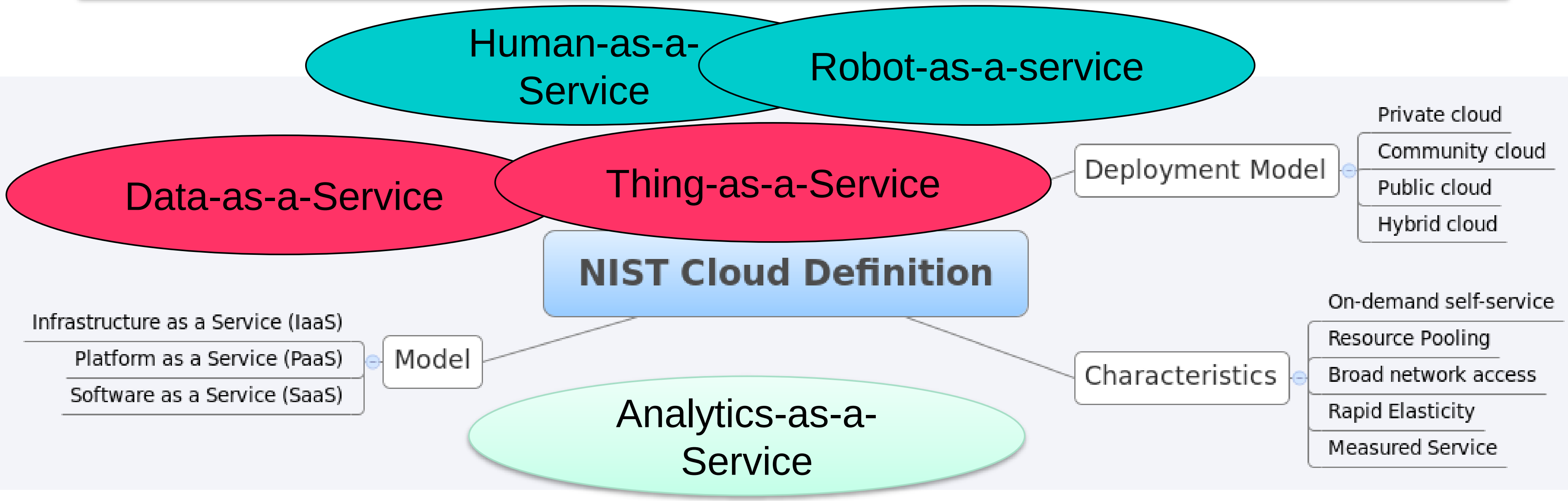
- **Size does matter**
  - Large-scale interactions
  - Big data generated and consumed
  - Big quantities to be managed
  - Hard to control quality of data and services
- **Any * access behaviour does matter**
  - Unpredictable workload
  - Scalability
  - Elasticity
  - Software-defined
- **Economic factors do matter**
  - On-demand, pay-as-you-g
  - Complex contracts
  - Blockchain?

Cloud definitions

Original definition from NIST

"This cloud model promotes availability and is composed of five essential **characteristics,** three **service models**, and four **deployment models**."

Source: NIST Definition of Cloud Computing v15, http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc



Internet of Things (IoT)

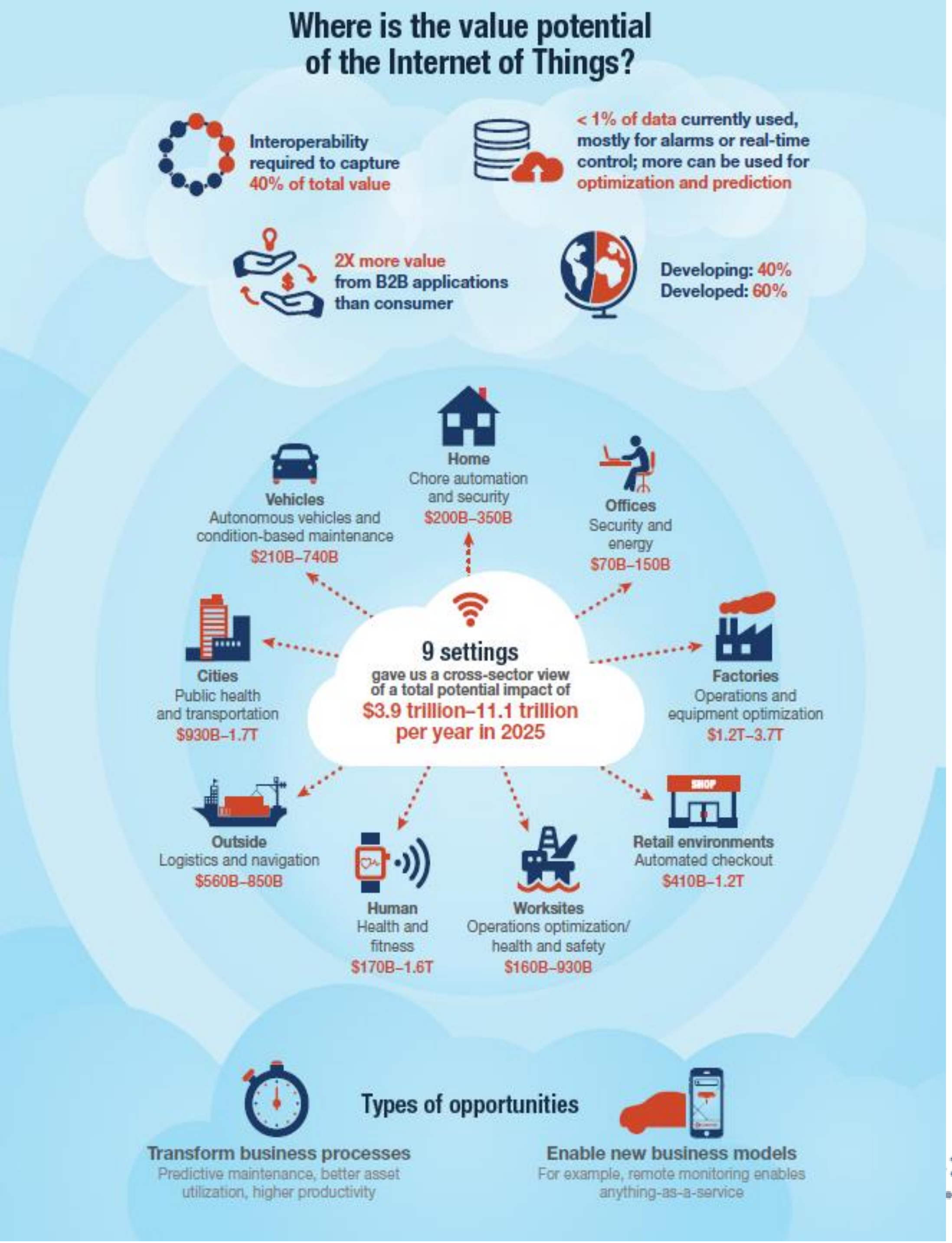
- Things and Objects
  - Home
  - Shops
  - Official Business
  - Hospital
  - Factory
  - Infrastructure

http://www.control4.com/blog/2014/03/the-internet-of-things-and-the-connected-home

- How to make such things and objects being connected and interacting each other?

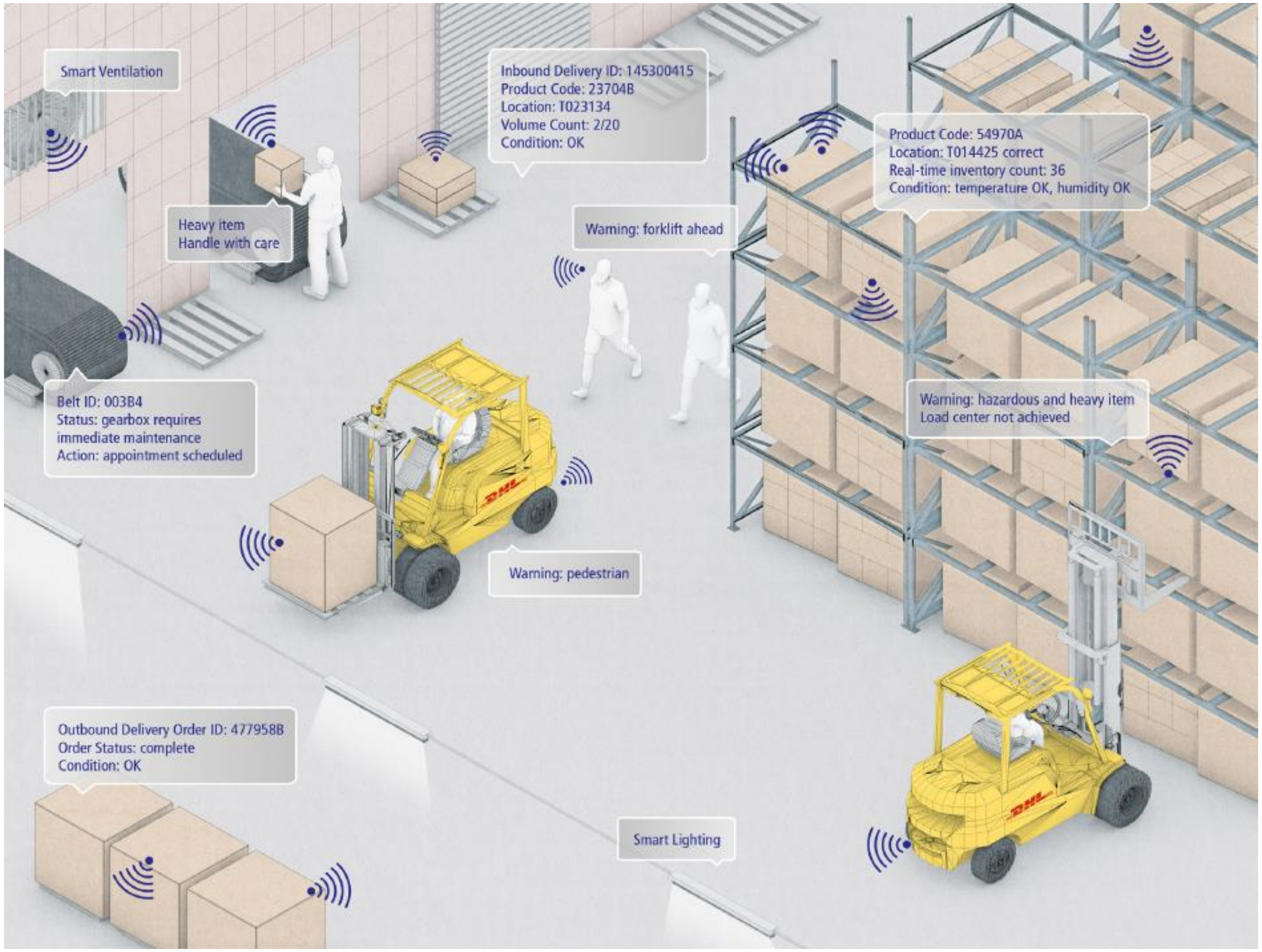
IoT

Figure source: McKinsey Global Institute: THE INTERNET OF THINGS: MAPPING THE VALUE BEYOND THE HYPE JUNE 2015 HIGHLIGHTS

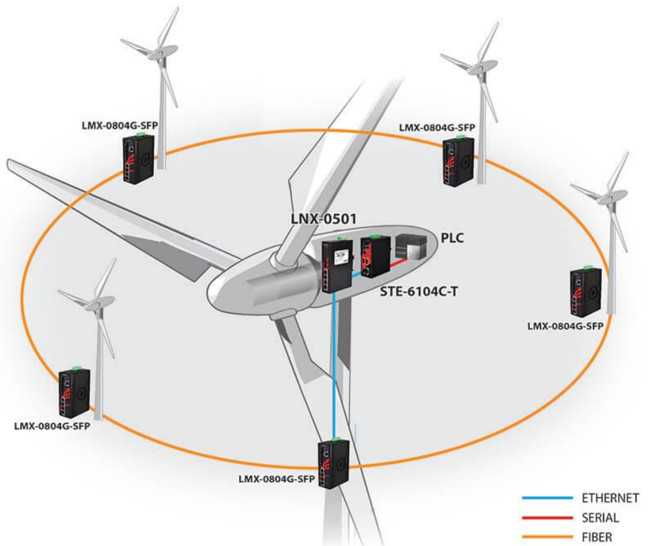


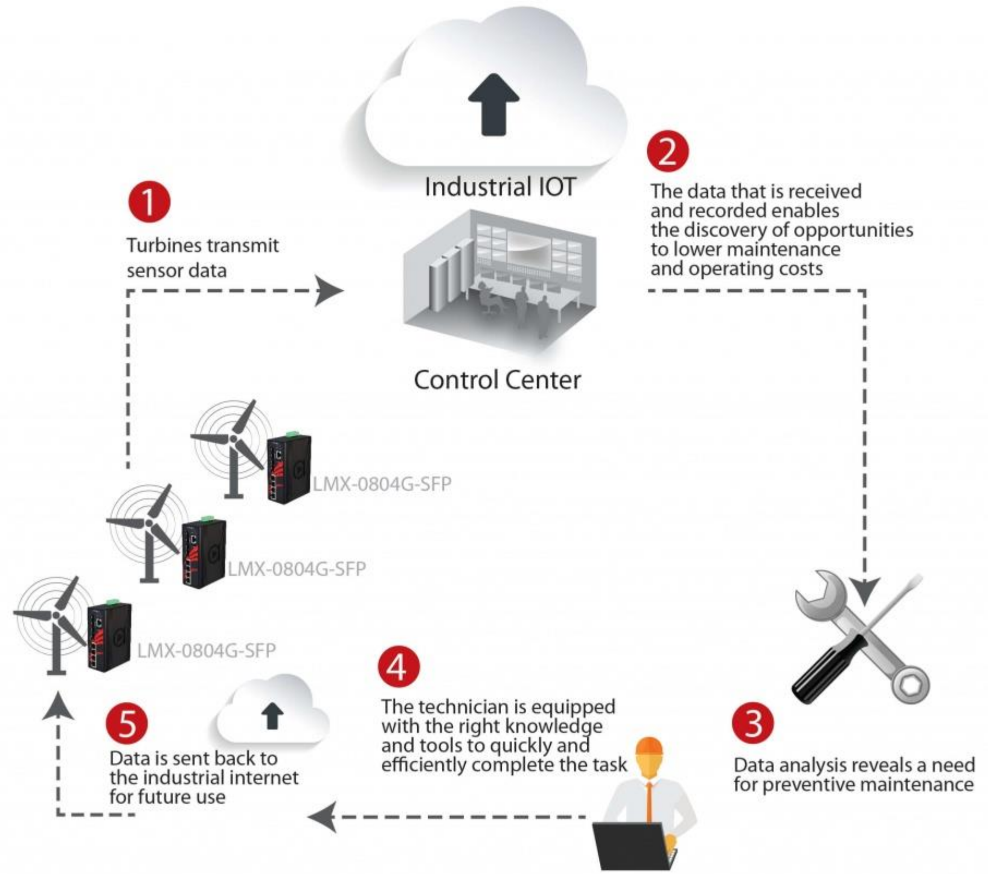
Logistics scenario from DHL

Figure source: DHL Trend Research & Cisco Consulting Services, INTERNET OF THINGS IN LOGISTICS, 2015



Industrial internet





Figures source: http://www.windpowerengineering.com/design/electrical/controls/wind-farm-networks/talking-turbines-internet-things/

Video analytics + business applications/public security

Use Case 3: Video Analytics

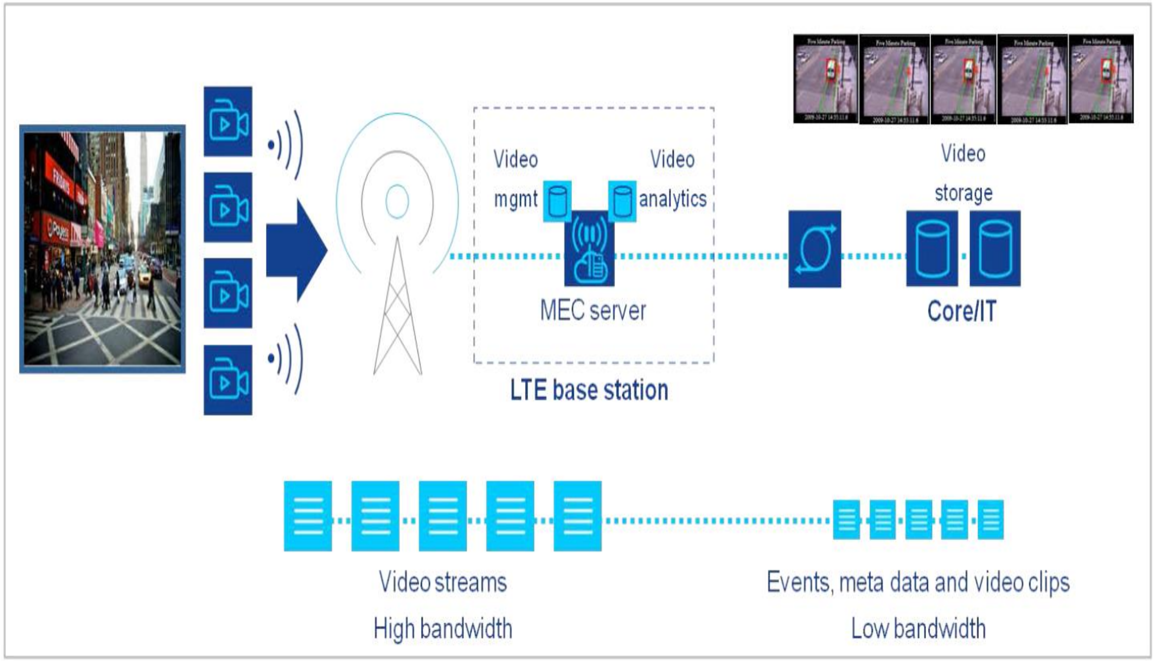


Figure 4: Example of video analytics

Figure source:
https://portal.etsi.org/portals/0/tbpages/mec/docs/mobile-edge_computing_-_introductory_technical_white_paper_v1%2018-09-14.pdf

Identifying key focuses in this course

NEW TRENDS

Co-located Services in an ecosystem

- Big data, IoT Events/datahub, IoT device management, Serverless function, AI as a service, etc. are collocated in the same place
  - Google, Amazon, Azure, etc

- Multiple types of services provided by different providers in the same cloud infrastructures

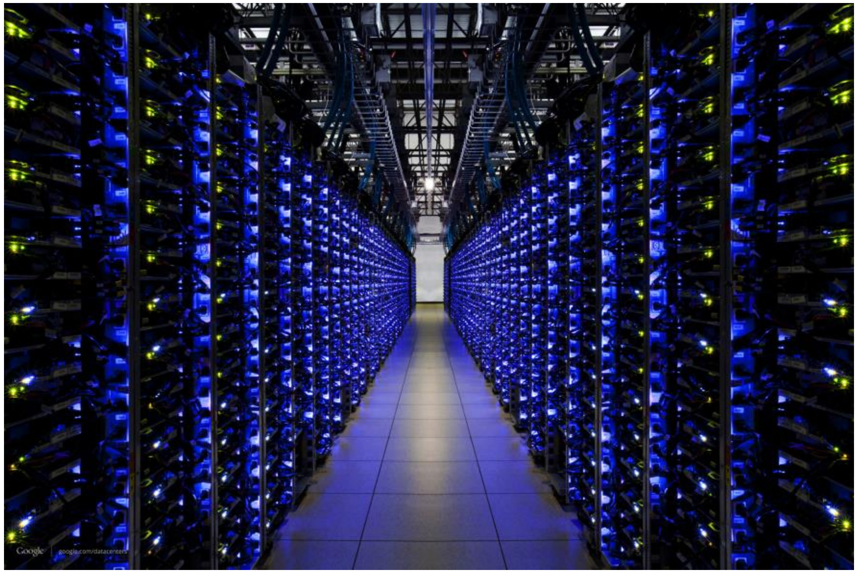
Multiclouds

- Multi aspects
  - Distributed, small and big, diverse
- Mini clouds and micro data centers
  - Used in edge/fog computing model to deploy resources close to data and analytics
- Cloud technologies: no longer associated with “big data centers”
  - They refer to “cloud” methodologies, models and techniques for distributed computing and services

Take a read: https://www.rightscale.com/blog/cloud-industry-insights/cloud-computing-trends-2017-state-cloud-survey

Connecting data centers to IoT

Data Center: Processing, Storage, Networking, Management, Distribution



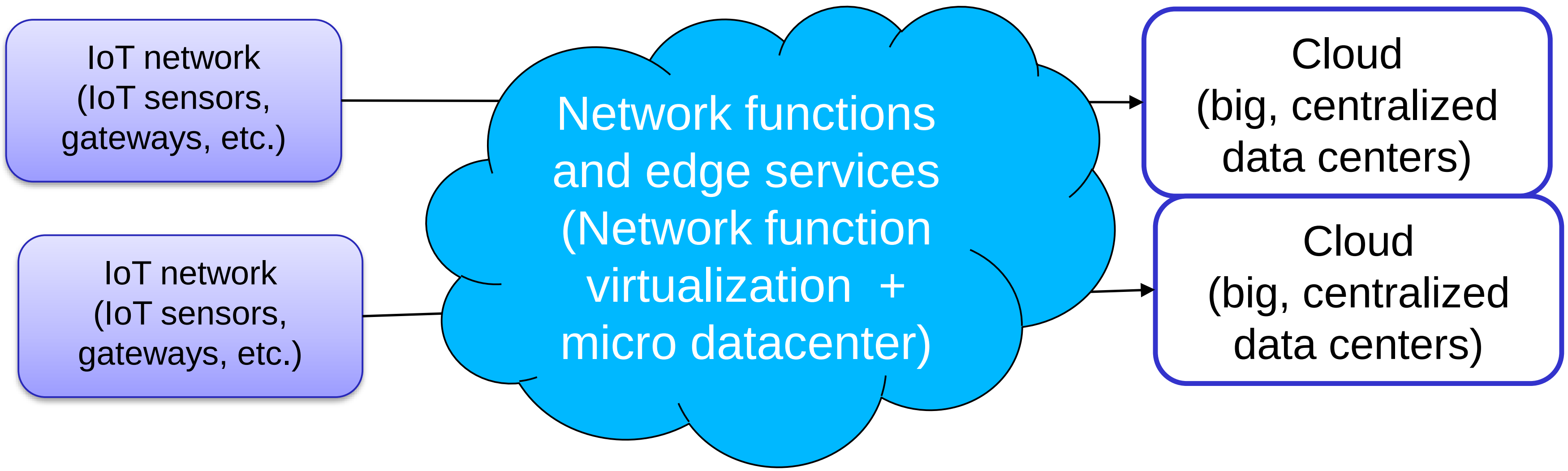
http://www.infoescola.com/wp-content/uploads/2013/01/datacenter-google.jpg

IoT devices: Gateways, Sensors, Actuators, Topologies of Gateways

http://www.control4.com/blog/2014/03/the-internet-of-things-and-the-connected-home

Combining IoT + Network functions + Clouds

- Resources as services: IoT networks (sensors, gateways), edge/fog systems (micro data centers and network functions), big data centers (cloud VM, storage)



Services and Cloud for AI/IA & VR/AR

- Artificial Intelligence and Intelligence Amplification provided as services
- Connected *
  - Car, Health, and Assets
- Cloud robotics
- Manufacturing/Industry 4.0
- Virtual reality (VR) and augmented reality (AR) utilizing IoT data and services

Take a read: https://conficio.design/mixed-reality-and-iot

Technical trends

- Service interfaces
  - Not just REST, SOAP but also well-defined interfaces using messages via brokers and gRPC
- Service implementation
  - Java, Python, Nodejs, Go, etc.
- Underlying distributed resources
  - Raspberry PI, Virtual machines, OS Containers, virtual clusters, CPU cluster + GPU, Cloud TPUs
- “Microservices” mindset
  - Container, Container and Container!
  - Serverless Computing: Moving from pay-per-use “infrastructures” to pay-per-use function calls

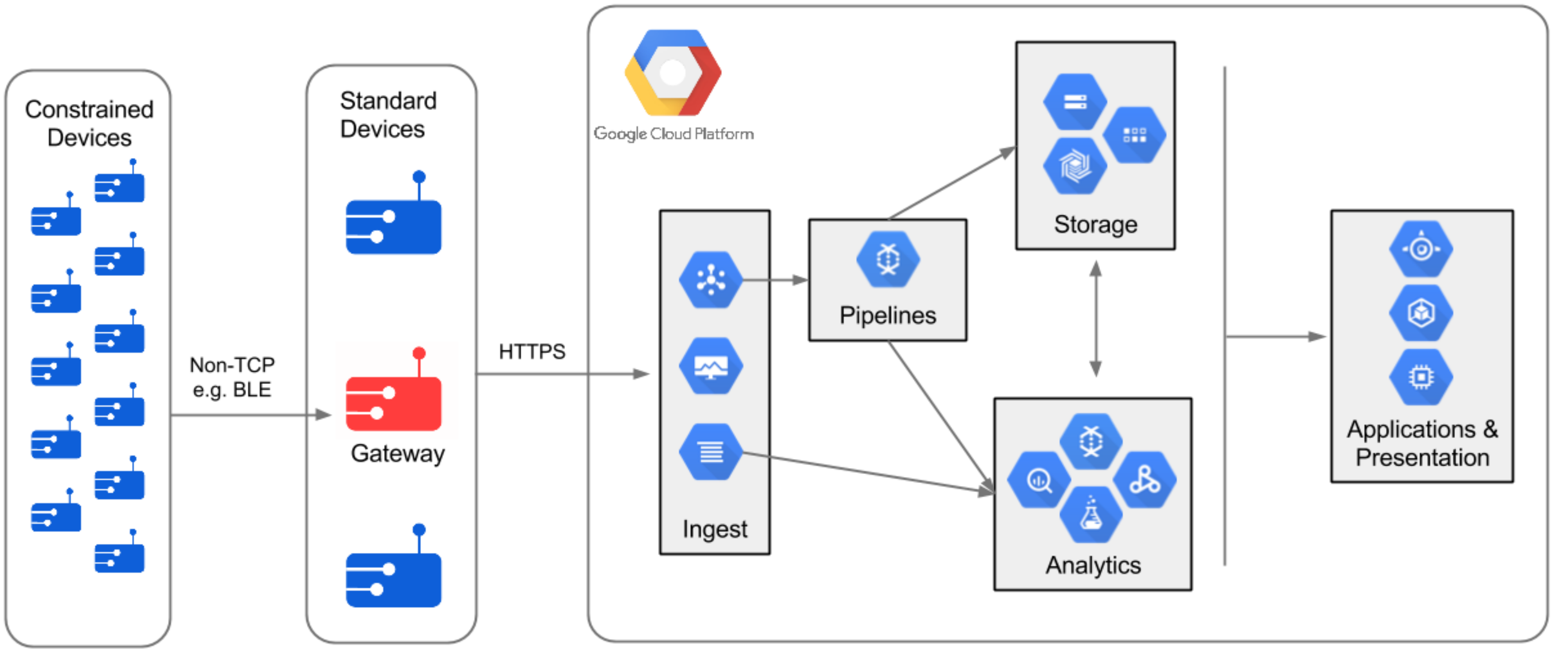
#1 key focus in this course

SCALABLE DATA, SERVICES AND SYSTEMS MANAGEMENT

Main types of virtualization of infrastructures for distributed apps

- Compute resource virtualization
  - Compute resources: CPU, memory, I/O, etc.
  - To provide virtual resources for „virtual machines“
- Storage virtualization
  - Resources: storage devices, harddisk, etc.
  - To optimize the usage and management of data storage
- Network Function Virtualization
  - Network resources: network equipment
  - To consolidate network equipment and dynamically provision and manage network functions

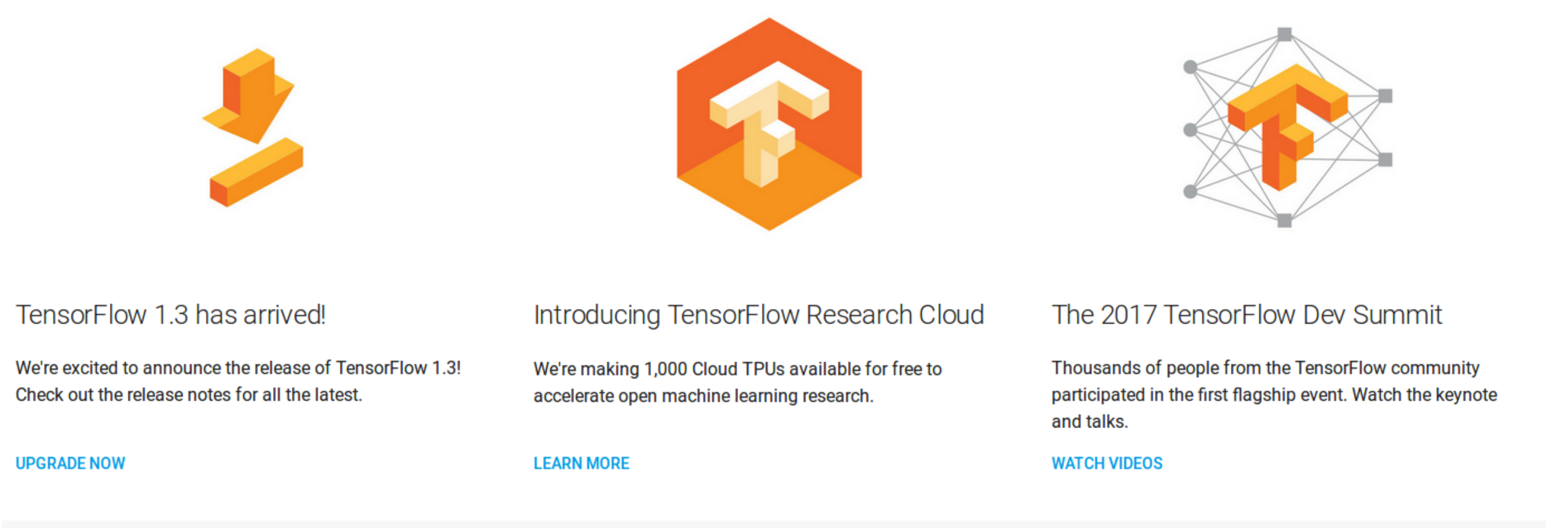
Example: IoT scenario



Source: https://cloud.google.com/solutions/architecture/streamprocessing

Example: AI with Tensorflow

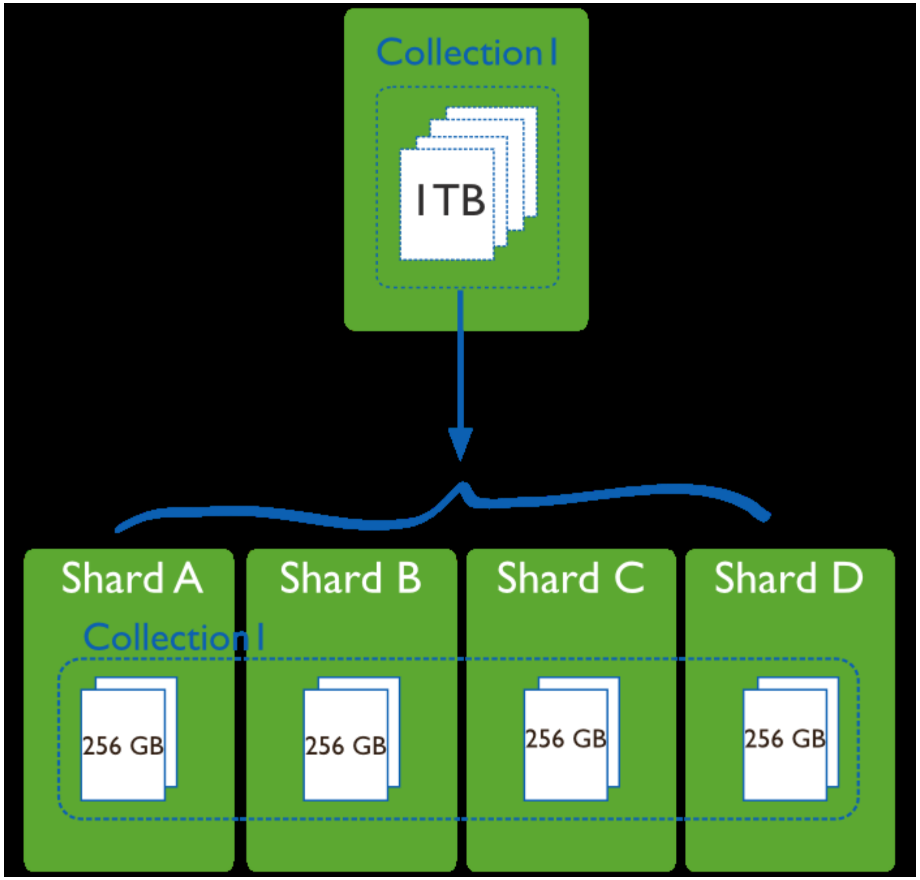
Key issue: algorithms (require a lot of computing power) + data (a lot of data)



Source: https://www.tensorflow.org/

Data Sharding

Need also Routing, Metadata Service, etc.



Soure: https://docs.mongodb.org/manual/core/sharding-introduction/

Load balancing of data services

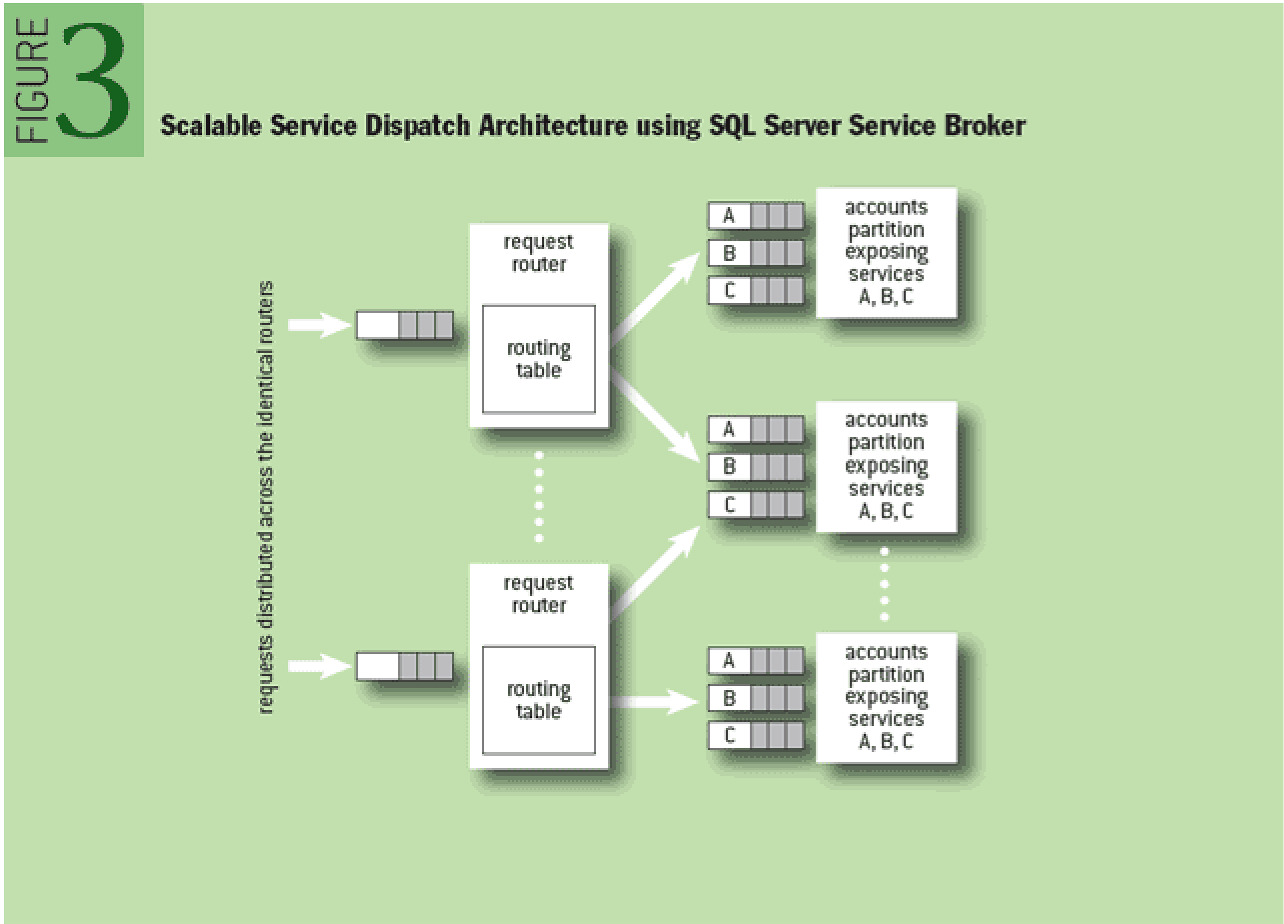
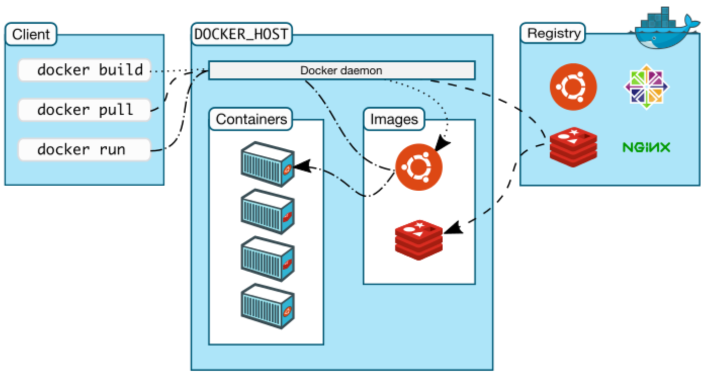
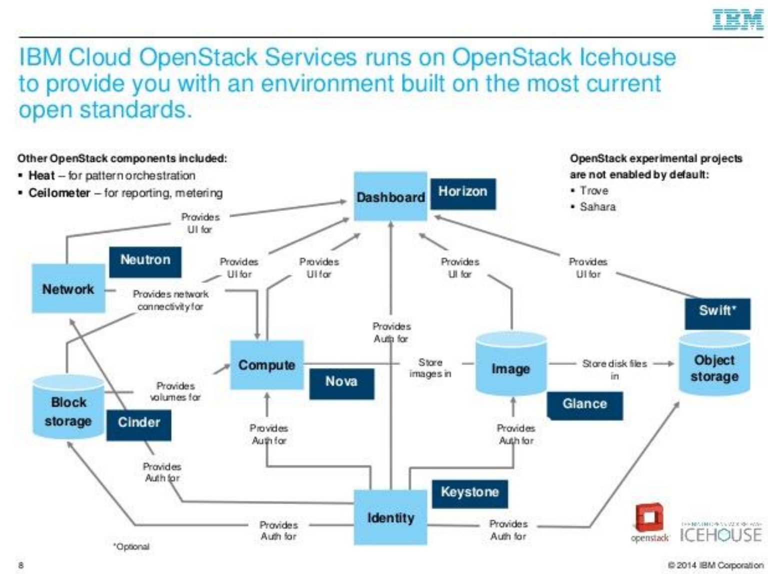


Figure source: http://queue.acm.org/detail.cfm?id=1971597

Examples



Source: https://docs.docker.com/engine/understanding-docker/



Source: http://www.slideshare.net/OpenStack_Online/ibm-cloud-open-stack-services

Kubernetes is an open-source system for automating deployment, scaling, and management of containerized applications.

It groups containers that make up an application into logical units for easy management and discovery. Kubernetes builds upon 15 years of experience of running production workloads at Google, combined with best-of-breed ideas and practices from the community.

Source: https://kubernetes.io/

Key research questions

- Management of complex types of resources (VM/containers, data, IoT devices, etc.)
  - Which are important advanced algorithms and design for scalable data/services/system management?
  - How to deal with high availability of data and data sharding?
  - How to deal with geographical multi-cloud load balancing?
  - Data services and container technologies

#2 key focus in this course

ELASTICITY AND CONTROL

Cloud Service Elasticity

A common problem in Cloud controls

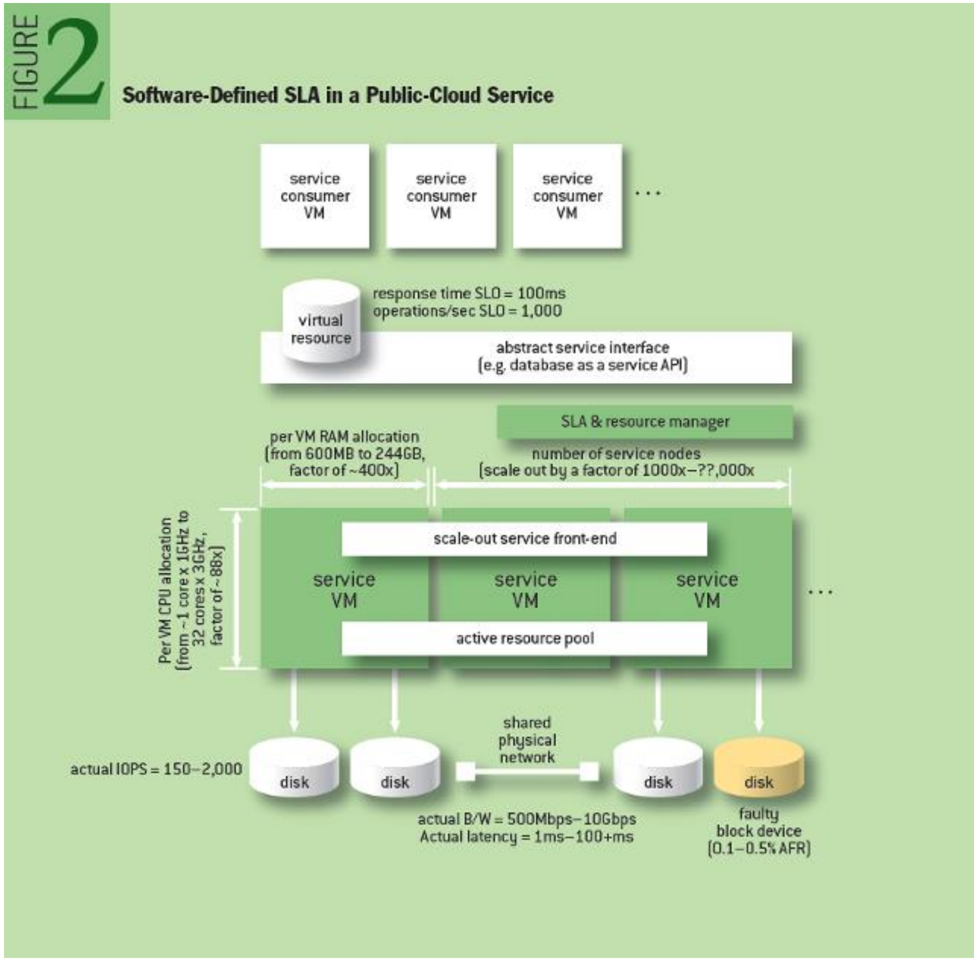


Figure source: http://queue.acm.org/detail.cfm?id=2560948

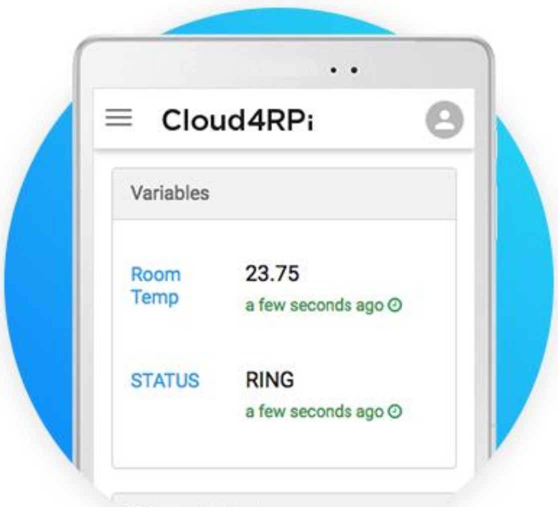
But we have much more complex situations

- Services in the car + services in the cloud
- Services in the robot + services in the cloud
- Interactions/communications
  - within cars/robots versus interactions/communications
  - among cars/robots and clouds
- Different data flows
  - Within the edge
  - Between the edge and the cloud

Controls between clouds, edge/fog and IoT

Source: https://en.wikipedia.org/wiki/File:Waymo_self-driving_car_front_view.gk.jpg

Source: https://spectrum.ieee.org/automaton/robotics/robotics-software/cloud-robotics



Source: https://cloud4rpi.io

How to control 1000+ assets or a network of Base Transceiver Stations?

Key research questions

- Which algorithms and techniques can be used for elasticity controls in IoT, Cloud and fog computing?
- How to coordinate tasks among entities in IoT, edge systems and cloud centers?
- How does NFV leverage resources virtualization and elasticity techniques?

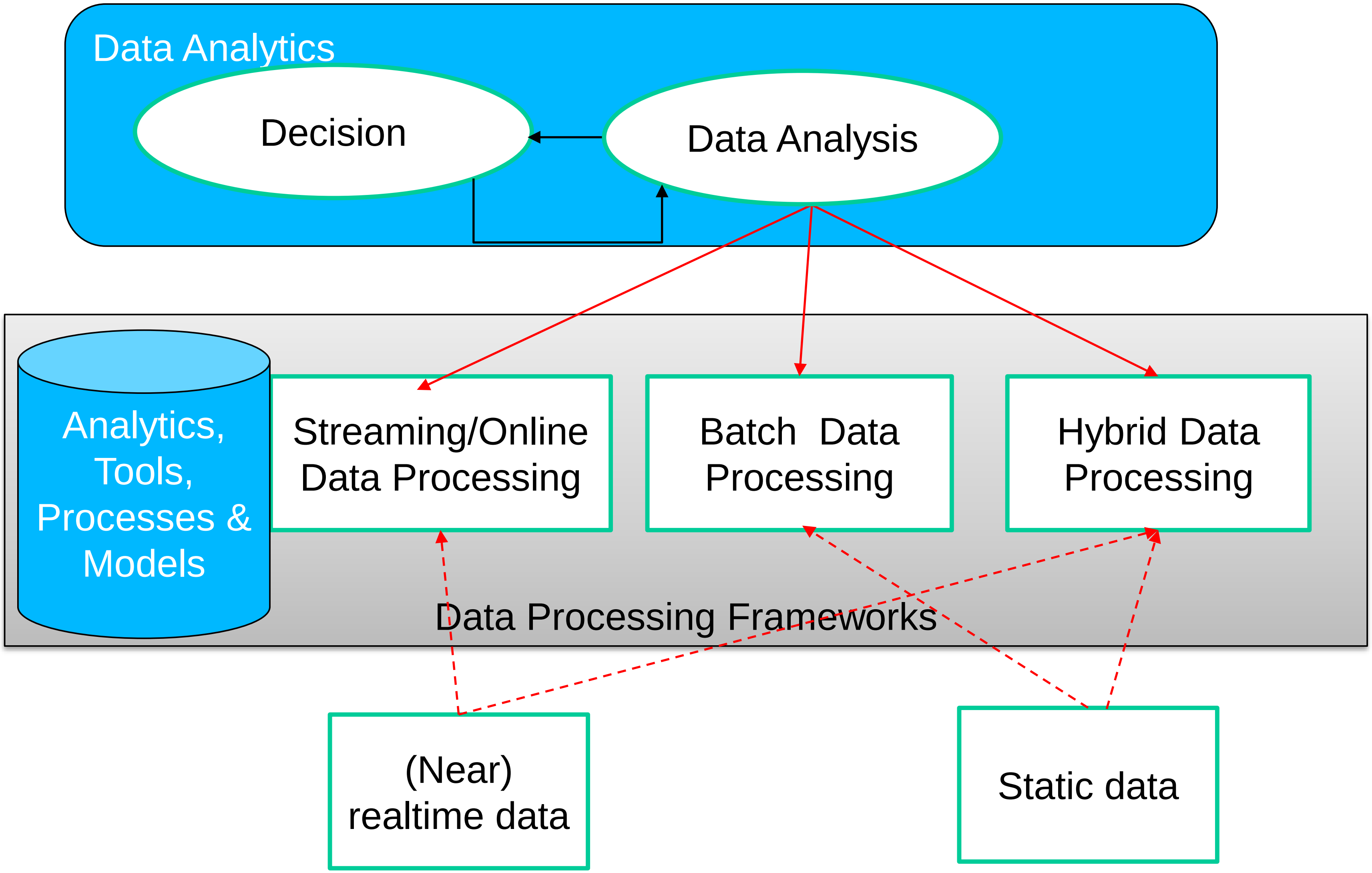
#3 key focus in this course

BIG **SYSTEM** FOR DATA ANALYTICS

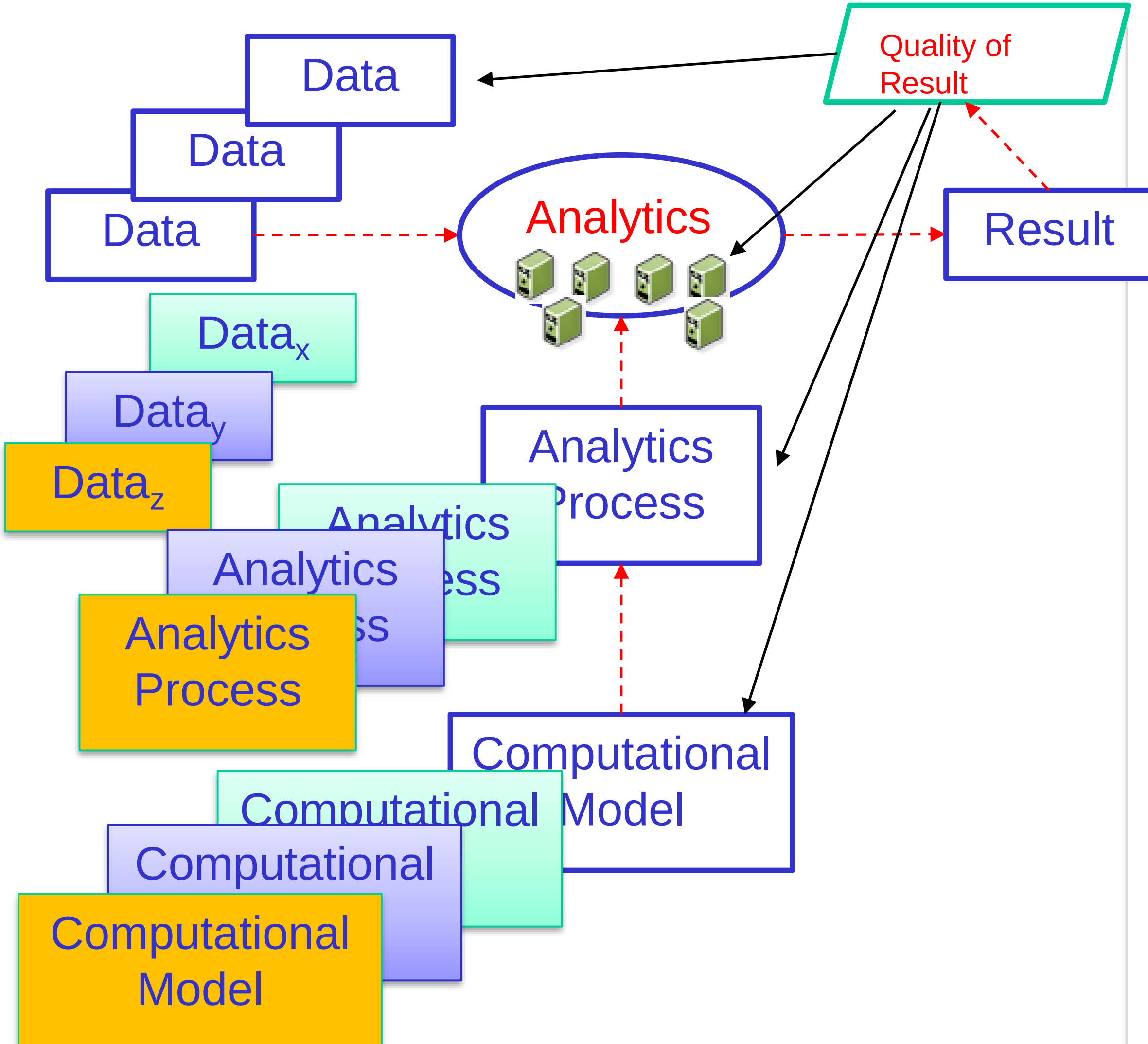
Data Processing Framework

- Batch processing
  - Mapreduce/Hadoop, Apache Spark
  - Scientific workflows
- (Near) realtime streaming data and complex event processing
  - Flint, Apex, Storm, etc.
- Hybrid data processing
  - Summingbird, Apache Kylin
  - Impala, Storm-YARN
  - Apache Spark

Conceptual View



Quality of analytics



- **More data** → more computational resources (e.g. more VMs)
- **More types of data** → more computational models → more analytics processes
- Change **quality of analytics**
  - Change quality of data
  - Change response time
  - Change cost
  - Change types of result (form of the data output, e.g. tree, visual, story, etc.)

S

Hong-Linh Truong, Schahram Dustdar, "Principles of Software-defined Elastic Systems for Big Data Analytics", (c) IEEE Computer Society, IEEE International Workshop on Software Defined Systems, 2014 IEEE International Conference on Cloud Engineering (IC2E 2014), Boston, Massachusetts, USA, 10-14 March 2014

Human + Machine for Data Analytics?

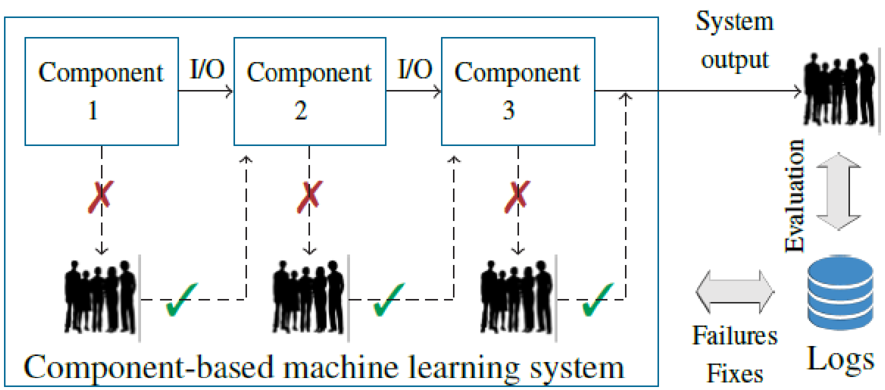


Figure 1: Troubleshooting with humans in the loop

Figure source: Besmira Nushi, Ece Kamar, Donald Kossmann and Eric Horvitz. On Human Intellect and Machine Failures: Troubleshooting Integrative Machine Learning Systems, AAAI 2017.

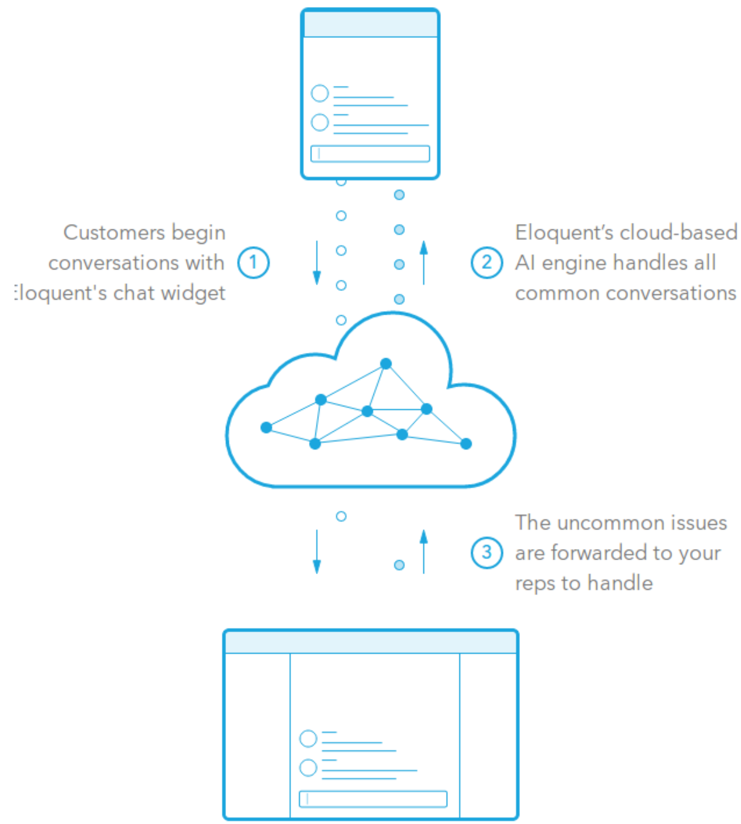


Figure source: https://www.eloquent.ai/

Key questions

- Which techniques and algorithms are important for big data ingest and analytics?
- How does data analytics leverage elasticity?
- How do big data analytics systems support machine learning?
- How to achieve quality-aware analytics?

#4 key focus in this course

Key focus in this course

END-TO-END SERVICE SYSTEM ENGINEERING

Common goals for IoT Cloud service engineering analytics

- Type 1
  - **Mainly focus** on IoT networks: sensors, IoT gateways, IoT-to-cloud connectivity (e.g., connect to predix.io, IBM Bluemix, Amazon IoT, etc.)
- Type 2
  - **Mainly focus** on (public/private) services in data centers: e.g., load balancer, NoSQL databases, and big data ingest systems
  - Using both open sources and cloud-provided services
- Type 3
  - **Equally focus** on both IoT and cloud sides and have the need to control at both sides
  - Highly interactions between the two sides, not just data flows from IoT to clouds

End-to-End resource management

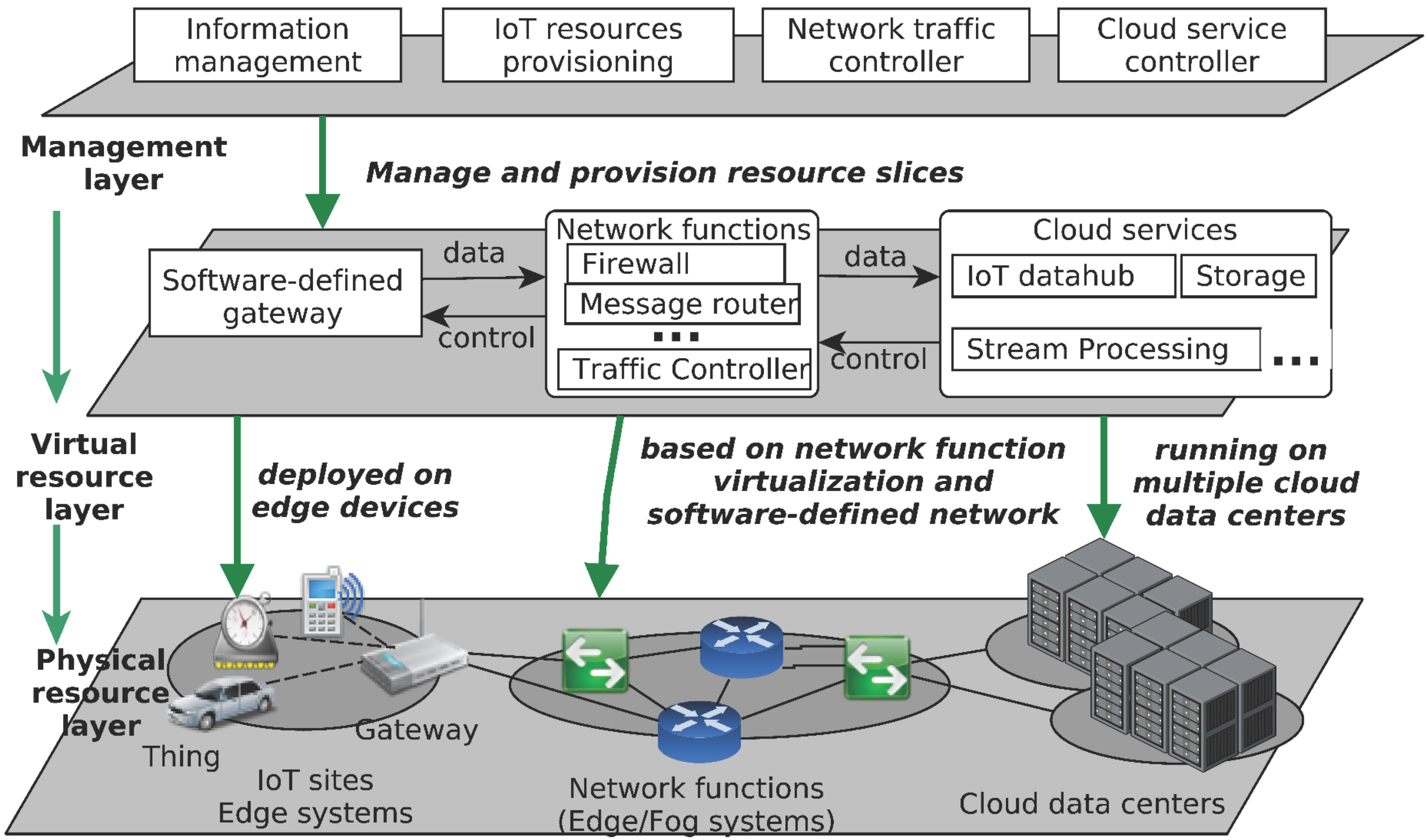
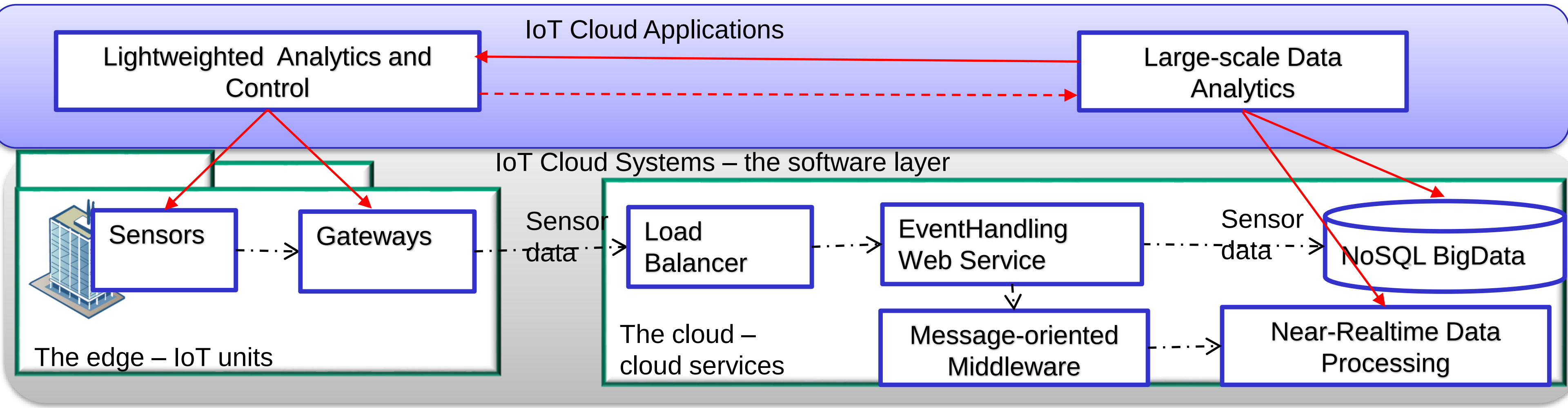
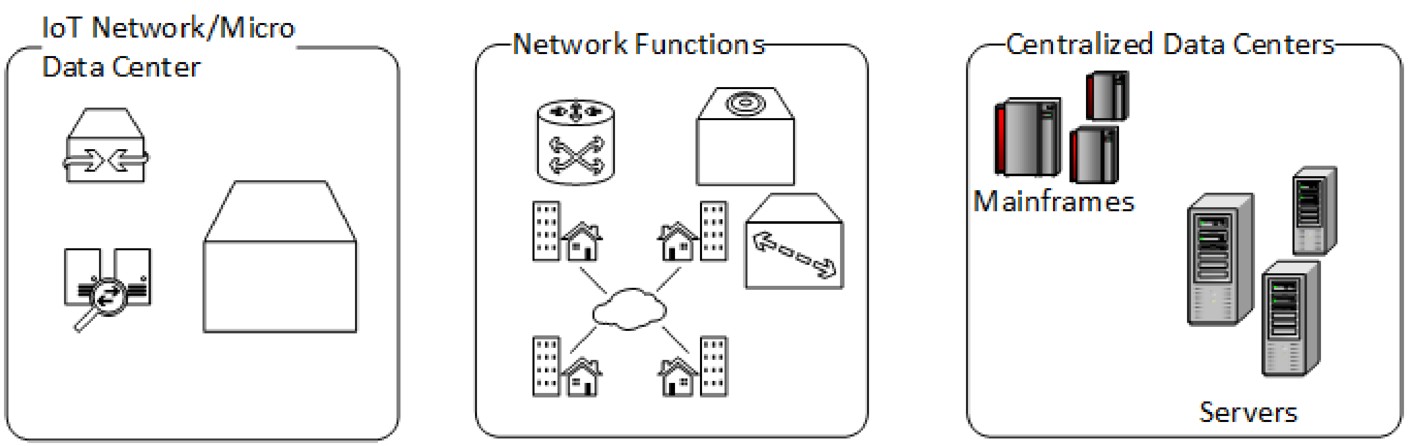


Figure source: Duc-Hung Le, Nanjangud C. Narendra, Hong Linh Truong:
HINC - Harmonizing Diverse Resource Information across IoT, Network Functions, and Clouds. FiCloud 2016: 317-324

End-to-End View on Applications and Systems





Uncertainty & testing

- Many possible uncertainties associated with interactions among data, services and systems
- Supporting testing uncertainties and uncertainties analytics
  - Conventional aspects, e.g., infrastructural physical resources and typical system operations
  - Emerging novel aspects: data uncertainties (data/data-centric), elasticity of resources (w.r.t function and composition), and governance (related to business/trustworthiness)

Key questions

- How do we monitor and analyze IoT, cloud and fog/edge systems?
- Which are important techniques and tools for instrumenting and monitoring IoT, cloud and fog systems?
- How do we determine metrics for end-to-end view?
- What types of uncertainties? How to test them

Summary

- Different types of services built atop IoT, fog/edge and cloud systems
  - The underlying infrastructures and systems are very complex
  - Virtualization and elasticity are important techniques
  - Big data problems: dealing with a lot of near real time data
  - Performance monitoring and testing
- The course focus points
  - Advanced system design techniques and algorithms
  - Not covered in Distributed Systems Technologies or Advanced Service Engineering

Thanks for your attention

Hong-Linh Truong
Faculty of Informatics
TU Wien
truong@dsg.tuwien.ac.at
http://dsg.tuwien.ac.at/staff/truong