

Emerging Distributed Computing and Challenges for Services Engineering

Hong-Linh Truong

Faculty of Informatics, TU Wien

hong-linh.truong@tuwien.ac.at
<http://www.infosys.tuwien.ac.at/staff/truong>
Twitter: @linhsolar

Goals

- See emerging trends in complex cloud native software service systems
- Have a critical look at use cases and analyze use cases
- See the service engineering technologies needed for such use cases

Outline

- Some emerging models
 - IoT resources
 - (Big/realtme) data provisioning models
 - Computational infrastructures/frameworks provisioning
 - Human computation provisioning
 - Machine Learning as a service
 - Blockchain
- Use cases
- Advanced services engineering
 - Single service/platform engineering
 - Multi-platform services engineering

Current trends: emerging systems

- Internet of Things (IoT)/Cyber-physical systems
 - Integration and virtualization of sensors/actuators and edge networks
- IoT and cloud integration → IoT cloud systems
 - Dealing with sensors/actuators and gateways integration with cloud data centers
- Fog/edge computing and mobile-edge computing
 - Nano/micro data centers + cloud-based data centers
 - Incorporation of Network Function Virtualization (NFV)/5G
- Blockchain protocols and systems
 - decentralized distributed ledger

Current trends: data, software, and services

- „Big“ and „small“ data
 - Data from Things (Internet of Things)
 - Human-sensing data, science data
- Big data analytics
 - Streaming processing and Machine Learning as a Service
- Human-Machine interaction
 - Individuals, crowds and collectives augmenting machine intelligence
 - Intelligence Amplification
- Elastic Services
 - Dynamic, flexible data, computation and analytics provisioning and integration models

ASE – complex requirements

- **Data** - big and near real-time data must be handled in a timely manner to extract insightful information
- **Systems** - complex applications/systems executed atop multiple, diverse distributed computing environments
 - Data centers/cloud infrastructures, IoT systems, human computation environments, blockchain, etc.
- **Quality** - cross-boundary, multiple concerns w.r.t trustworthiness, quality, regulation and cost/benefits must be assured.
- **Intelligence** - machine learning techniques & human knowledge
- **Elasticity** - flexible and dynamic management
 - Multi-dimensional elastic capabilities

Questions

**ARE YOU WORKING ON SUCH
SYSTEMS? IS THIS COURSE
SUITABLE FOR YOU?**

ASE – complex requirements (2)

For complex functions offered across distributed cloud and edge computing environments

- We want to have **a coherent, uniform view** of diverse types of resources and platforms
- We want to **coordinate** capabilities of these resources and platforms
- We want to support **elastic quality expectation** (from the users) for specific contexts!

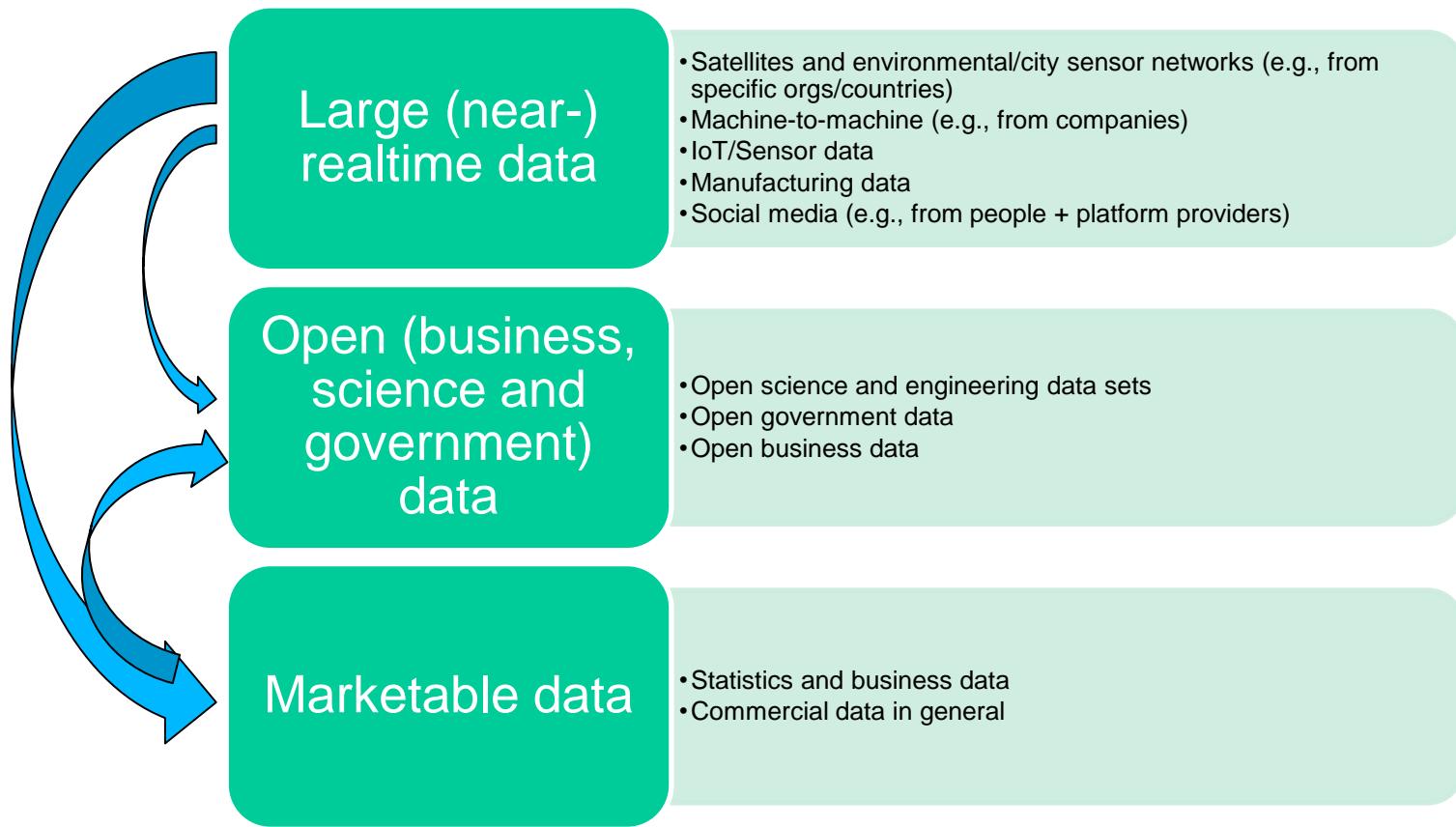
→ Engineering systems for such requirements is very challenging

We **research** and **explore** emerging techniques for **interesting scenarios** by utilizing **existing, advanced technologies!**

Deliver your understanding through the development of **your own open-source mini project!**

**WHICH ARE EMERGING
FORMS OF DISTRIBUTED
COMPUTING MODELS,
SYSTEMS AND APPLICATIONS
THAT YOU SEE?**

Emerging data provisioning models



Data are assets

Satellite data



The screenshot shows the Sentinel Online website interface. At the top, there is a navigation bar with links for Need Help?, FAQ, Contact Us, About Sentinel Online, and a search bar with options for Google Custom Search. Below the navigation bar, there are dropdown menus for Missions, User Guides, Technical Guides, Thematic Areas, Data Access (which is currently selected), and Toolboxes. A "You are here" breadcrumb trail indicates the user is at Home > Data Access. On the right side, there is a "Share" button and social media links for Twitter, Google+, Facebook, and LinkedIn. The main content area features a large banner for the Copernicus Sentinel Open Access Hub, which includes an icon of a globe with a network of points and two satellites, and a blue button labeled "sentinel open access hub >". Below the banner, a text box states: "The free, full and open data policy adopted for the Copernicus programme foresees access available to all users for the Sentinel data products, via a simple [self-registration](#). News and further information about the service is available [here](#)." To the right of the main content, there is a sidebar titled "Data Access" with links to Data Access Home, Use Typologies and available Services, User Support and Registration, and Access to Sentinel Data. Below this, there is a "Latest News" section with a link to "Read more".

Source: <https://sentinel.esa.int/web/sentinel/sentinel-data-access>

Example of open data



Browse Groups About



a comprehensive list of open data catalogs curated by experts from around the world.

268 registered data catalogs available.

publicdata.eu beta — Europe's Public Data

Search Groups About Apps Ideas

Login Register

europe's public data

Find datasets

Finance and Budgeting (436) Social Questions (229)
Environment (227) Transportation (199)
Education and Communication (194) Agriculture, Fisheries, Forestry (181)
Population (145) Economy and Industry (118)
Health (83)

View larger map [+]

ASE Fudan FIST, Summer 2018

DATA.GOV.UK Beta
Opening up Government

Home Data Participate Apps Location Linked Data Library Lab About

Search Map Search Publishers Tags Public Roles & Salaries Spend Browser Spend Reports

Recently Changed ParisData Mississauga, Ont Prince George, Br Data Publica Calgary, Alberta View revision log

Search Datasets 8729 Datasets Search... Search

Tags View all tags national-indicators Health health Spending Data care spend-transactions communities school NERC_DDC local-government transparency nhs children health-well-being-and-care population finance child health-and-social-care education disclosure

Publishers View all publishers

- Office for National Statistics (847)
- Department for Communities and Local Government (739)
- NHS Information Centre for Health and Social Care (514)
- British Geological Survey (364)
- Centre for Ecology & Hydrology (326)
- Department for Environment, Food and Rural Affairs (322)
- Welsh Government (241)
- Department of Health (239)
- Department for Children, Schools and Families (227)
- Home Office (221)

UK Location Conduct Map Based Search

The UK Location Programme has introduced over 1000 location data records into data.gov.uk and tools to support their use. To find which of these datasets cover a particular location, you can use Map Based Search.

Many of these datasets provide a Web Map Service too, and for some a preview of the data is available. Click to find out more about Map Based Search and about Preview on Map.

amazon web services

Sign Up My Account / Console English

AWS Products & Solutions Public Data Sets Developers Support

Browse By Category

- Astronomy
- Biology
- Chemistry
- Climate
- Economics
- Encyclopedic
- Geographic
- Mathematics

Public Data Sets

Public Data Sets on AWS provides a centralized repository of public data sets that can be seamlessly integrated into AWS cloud-based applications. AWS is hosting the public data sets at no charge for the community, and like all AWS services, users pay only for the compute and storage they use for their own applications. Learn more about [Public Data Sets on AWS](#) and visit the [Public Data Sets forum](#).

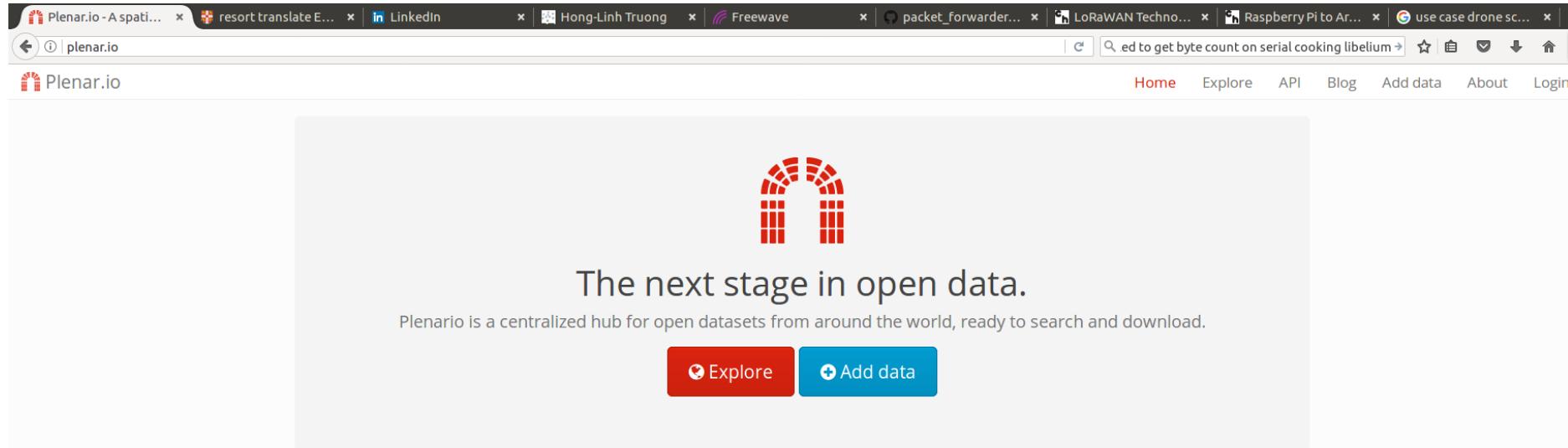
Featured Public Data Sets

- 1000 Genomes Project
- Common Crawl Corpus
- Google Books Ngrams

A corpus of web crawl data composed of 5 billion web pages. This data set is freely available on Amazon S3 and formatted in the ARC (.arc) file format.

A data set containing Google Books n-gram corpuses. This data set is freely available on Amazon S3 in a Hadoop friendly file format and is licensed under a Creative Commons Attribution 3.0 Unported License. The original dataset is available from <http://books.google.com/ngrams/>.

Static + Realtime Open Data



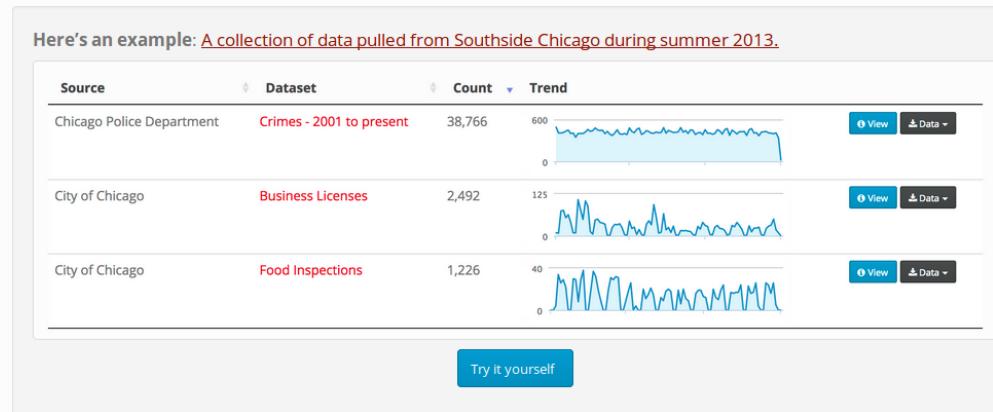
The next stage in open data.

Plenario is a centralized hub for open datasets from around the world, ready to search and download.

[Explore](#) [Add data](#)

One database. One map.

All data in Plenario exists on a single map and a single timeline, making it incredibly easy to access multiple datasets at once—especially those originally housed at different data portals.



Telecommunication

[https://dandelion.eu/datamine/
open-big-data/](https://dandelion.eu/datamine/open-big-data/)

Telecommunications - SMS, Call, Internet - MI TELCO

This dataset provides information about the telecommunication activity over the city of Milano. [read more »](#)

[Download data](#)

Telecommunications - MI to Provinces TELCO

This dataset provides information regarding the level of interaction between the areas of the city of Milan and the Italian provinces. [read more »](#)

[Download data](#)

Telecommunications - MI to MI TELCO

This dataset provides information regarding the directional interaction strength between the city of Milan different areas based on the calls exchanged between Telecom Italia Mobile users. [read more »](#)

[Download data](#)

Milano Weather Station Data WEATHER

The dataset describes various meteorological phenomena type and intensity of Milan city using sensors located within the city limits. [read more »](#)

[Download data](#)

Precipitation - Milano WEATHER

The dataset describes precipitation intensity and type over the city of Milan. [read more »](#)

[Download data](#)

Air Quality - MI ENVIRONMENT

The dataset describes the pollution type and intensity of Milan city using various types of sensors located within the city limits. [read more »](#)

[Download data](#)

MilanoToday NEWS

This dataset contains all the articles published on the website milanotoday.it from 01/11/2013 to 31/12/2013. [read more »](#)

[Get data via API](#)

Social Pulse - Milano SOCIAL

This dataset contains data derived from an analysis of geolocalized tweets originated from Milan during the months of November and December. [read more »](#)

[Get data via API](#)

Marketable data examples



HOME

DATA MARKETPLACE ▾

DATA SCORING

PARTNERSHIP PROGRAM

COMPANY INFO ▾

BLOG

BUY DATA

 • Data Marketplace • Buy Data

View data available on BDEX with this Data Visualization Tool

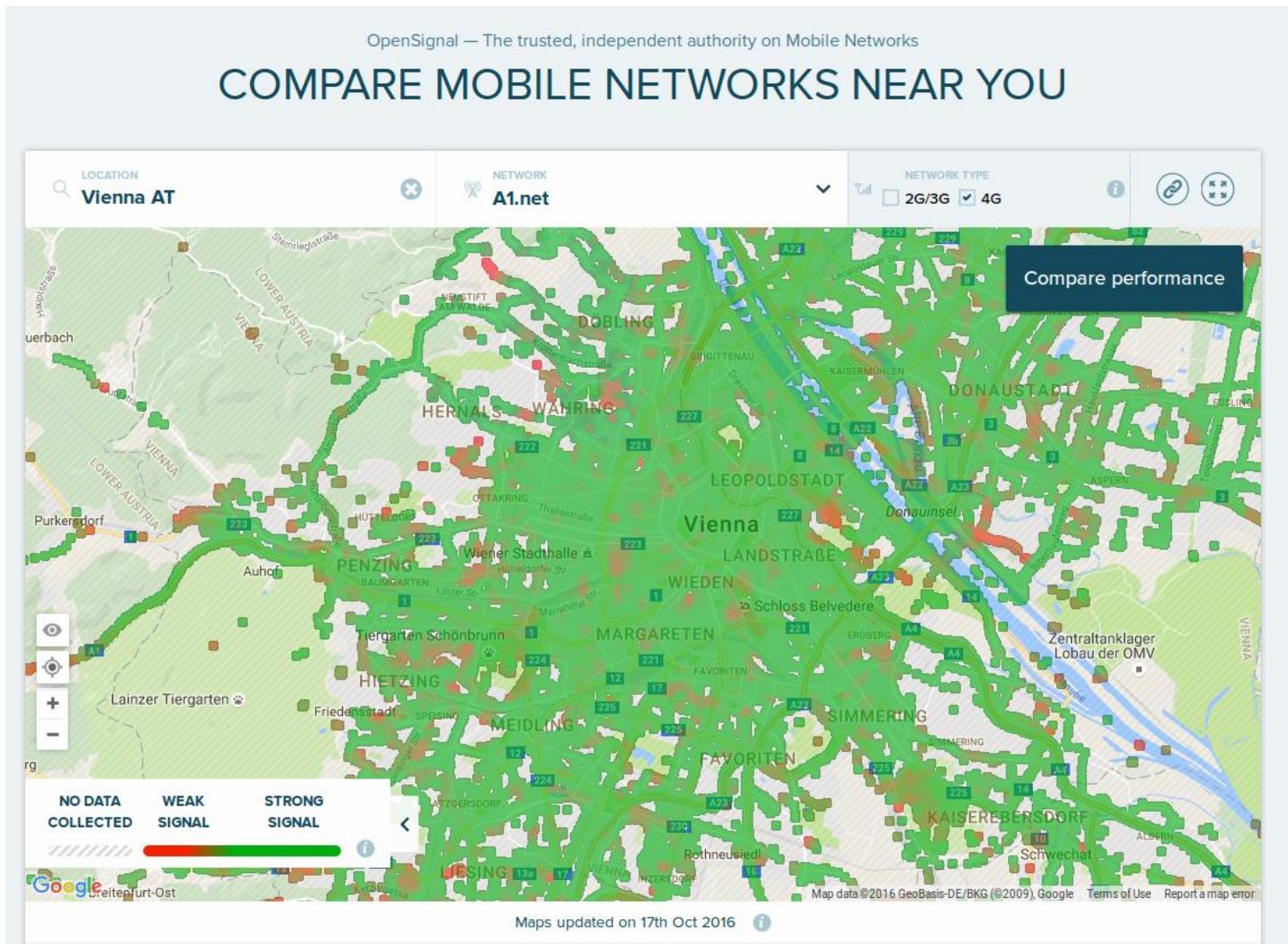
Real-Time Targeting

Data Buying and Targeting through the BDEX suite of tools truly takes the industry to the next level. Through BDEX's unique tool set Advertisers, Publishers and Retailers alike have the ability to target with a level of granularity that was never before possible.

- Buy Data That is Only Seconds Old
- Filter Based on Data Quality (conversions)
- Create Custom Audience Groups
- Combine an Unlimited Number of Data Points
- Set Budgets by Data Point
- Manage Campaigns in Real-Time

[Buy Data >](#)[Sell Data >](#)[Advertisers / Publishers >](#)[DMP / DSP Solutions >](#)[BDEX Retail >](#)[Data Downloads >](#)[Data Scoring >](#)

Marketable data examples



Marketable data examples

OpenSignal

[PRODUCTS](#)

[SUPPORT](#)

[BLOG](#)

[ABOUT](#)



Data Products



For Mobile Operators



For Telcos Regulators



For Industry Analysts



For Large Companies

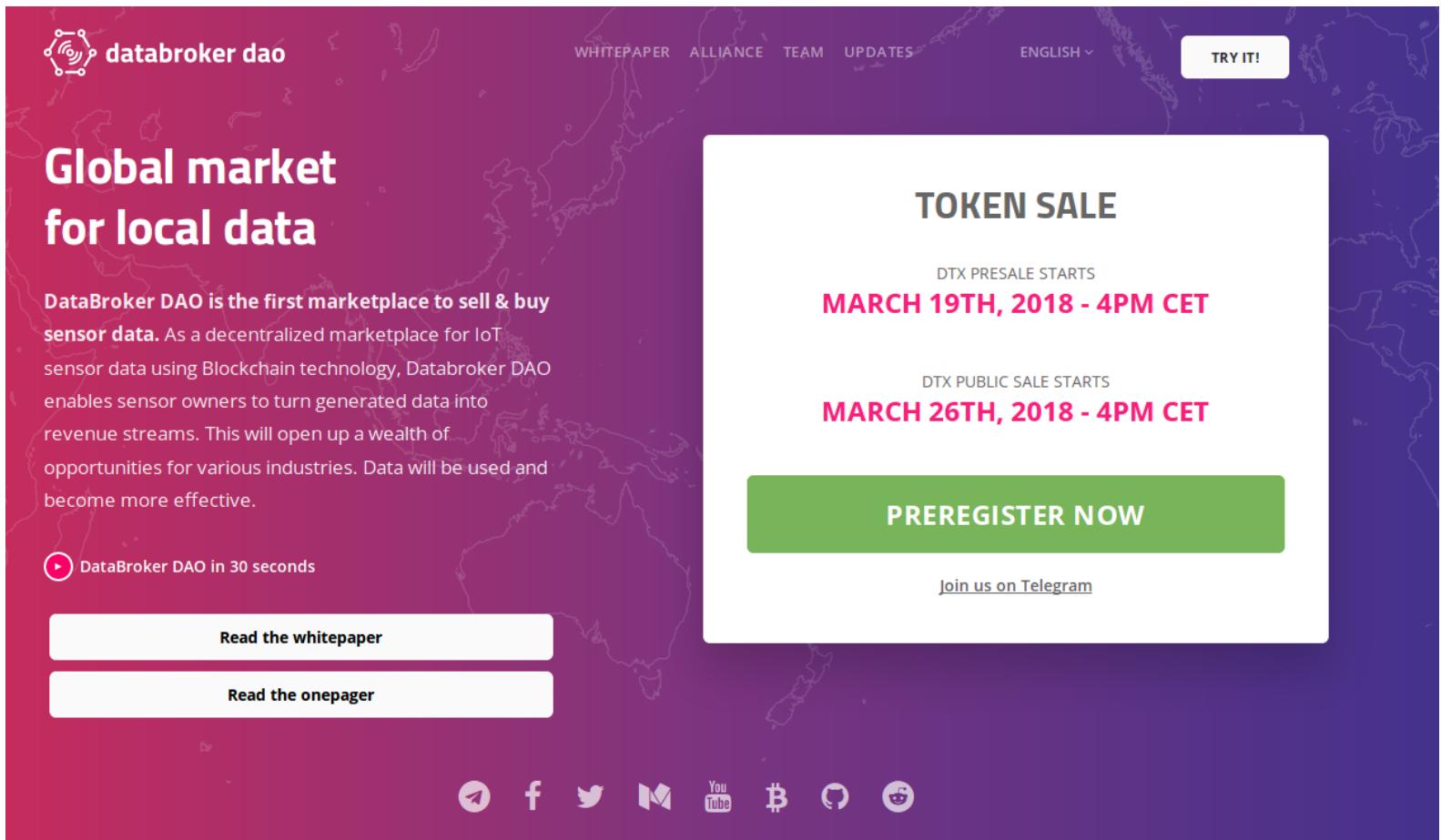


Our solutions

See how our data and analytics can help your business

[Visit the solutions site >](#)

And trend in monetizing data



The screenshot shows the homepage of the DataBroker DAO website. At the top, there is a navigation bar with links for WHITEPAPER, ALLIANCE, TEAM, UPDATES, ENGLISH, and a TRY IT! button. The main heading is "Global market for local data". Below it, a text block explains that DataBroker DAO is the first marketplace to sell & buy sensor data, using Blockchain technology to enable sensor owners to turn generated data into revenue streams. It features a video thumbnail titled "DataBroker DAO in 30 seconds", and buttons to "Read the whitepaper" and "Read the onepager". On the right side, a large white box contains the text "TOKEN SALE" and "DTX PRESALE STARTS MARCH 19TH, 2018 - 4PM CET". Another section below it says "DTX PUBLIC SALE STARTS MARCH 26TH, 2018 - 4PM CET" and has a green "PREREGISTER NOW" button. A link to "Join us on Telegram" is also present. At the bottom, there are social media icons for Telegram, Facebook, Twitter, LinkedIn, YouTube, Bitcoin, Ethereum, and Golem.

Global market for local data

DataBroker DAO is the first marketplace to sell & buy sensor data. As a decentralized marketplace for IoT sensor data using Blockchain technology, Databroker DAO enables sensor owners to turn generated data into revenue streams. This will open up a wealth of opportunities for various industries. Data will be used and become more effective.

► DataBroker DAO in 30 seconds

[Read the whitepaper](#)

[Read the onepager](#)

TOKEN SALE

DTX PRESALE STARTS
MARCH 19TH, 2018 - 4PM CET

DTX PUBLIC SALE STARTS
MARCH 26TH, 2018 - 4PM CET

[PREREGISTER NOW](#)

[Join us on Telegram](#)

Source: <https://databrokerdao.com/>

in EU call for proposals

TOPIC : Supporting the emergence of data markets and the data economy

Topic identifier:	ICT-13-2018-2019		
Publication date:	27 October 2017		
Types of action:	IA Innovation action		
DeadlineModel:	single-stage	Deadline:	28 March 2019 17:00:00
Opening date:	16 October 2018		
Types of action:	RIA Research and Innovation action CSA Coordination and support action		
DeadlineModel:	single-stage	Deadline:	17 April 2018 17:00:00
Opening date:	31 October 2017		

Time Zone : (Brussels time)

Source: <http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/ict-13-2018-2019.html>

Emerging computing infrastructure and platform provisioning models

- Infrastructure-as-a-Service
 - Machine as a service
 - Storage as a Service
 - Database as a Service
 - Network as a Service (think about Network Function Virtualization with 5G)
- Edge/Fog computing
 - Distributed edge/fog systems
 - analytics at the edge
 - Network functions and other system operations at the edge/fog systems
 - IoT-as-a-service

Emerging computing infrastructure and platform provisioning models

- Platform-as-a-Service
 - Application middleware
 - Computational frameworks
 - Data processing frameworks
 - Management middleware (e.g., monitoring, control, deployment)
- Technologies
 - Virtualization
 - Microservice architectures
 - Serverless computing
 - Machine learning/deep learning
 - Blockchain
 - Etc.

Examples

Data Processing Framework

Amazon Elastic MapReduce

Apache Flink

Apache Apex

Apache Spark

Google Cloud Dataflow

Kafka SQL

Azure Stream Analytics

Data
Transfer/Messaging
Middleware

StormMQ

Apache Nifi

Apache Kafka

MQTT

AmazonSQS

CloudAMQP

Google BigQuery

Elastic Search

Influx DB

Data Storages

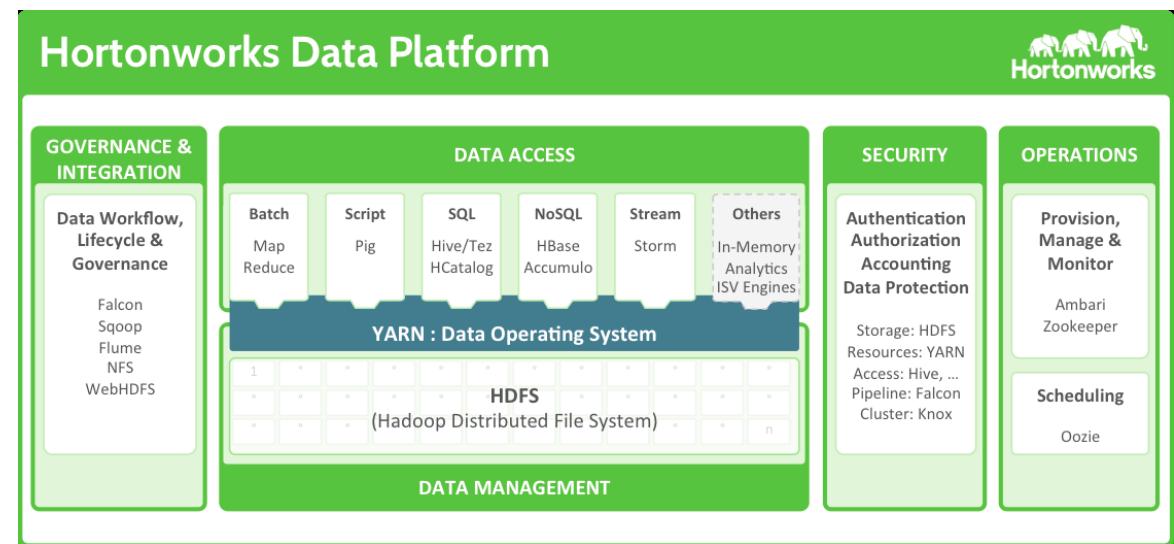
MongoLab

Amazon S3

Cassandra

Hadoop ecosystem

- Built around Mapreduce programming models and Hadoop software ecosystems
 - <http://hadoop.apache.org/>
- From “The Forrester Wave™: Big Data Hadoop Distributions, Q1 2016”: Top Hadoop solution providers are Cloudera, Hortonworks, IBM, MapR Technologies, and Pivotal Software



Source: <http://hortonworks.com/blog/defining-enterprise-hadoop/>

Spark ecosystem

Programming with Java, Scala, Python, R
We can have a separate modules

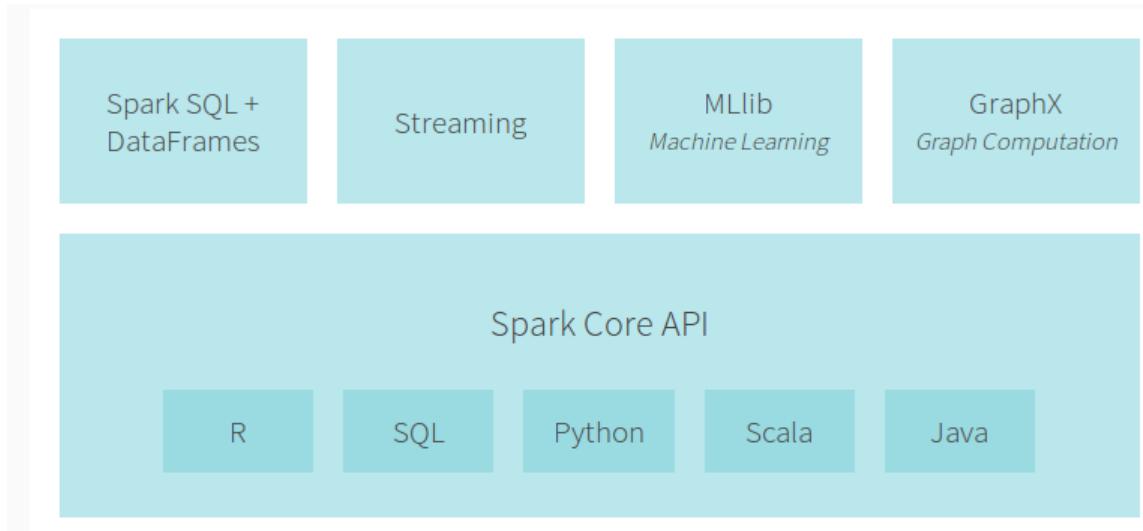
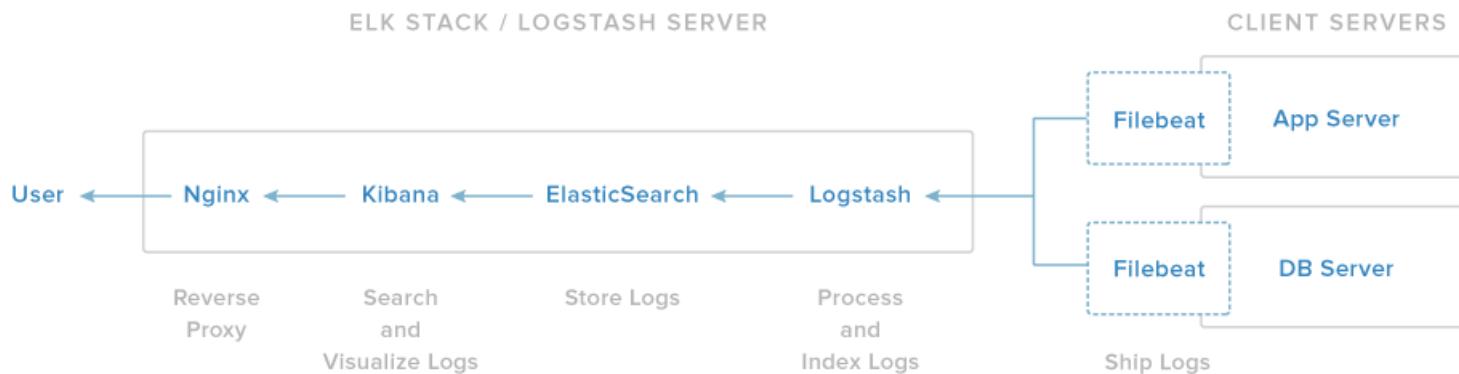


Figure source:
<https://databricks.com/spark/about>

ELK Stack

Building using elastic components: Elasticsearch, Elasticsearch Hadoop, Kibana, and Logstash

<https://www.elastic.co/>



Source: <https://www.digitalocean.com/community/tutorials/how-to-install-elasticsearch-logstash-and-kibana-elk-stack-on-ubuntu-14-04>

TICK Stack

Main from services of
Influx

<https://www.influxdata.com>

Focus on time series data

- Collect
- Storage
- Visualize
- ETL

TICK

Telegraf

- Time-Series Data Collector

InfluxDB

- Time-Series Data Storage

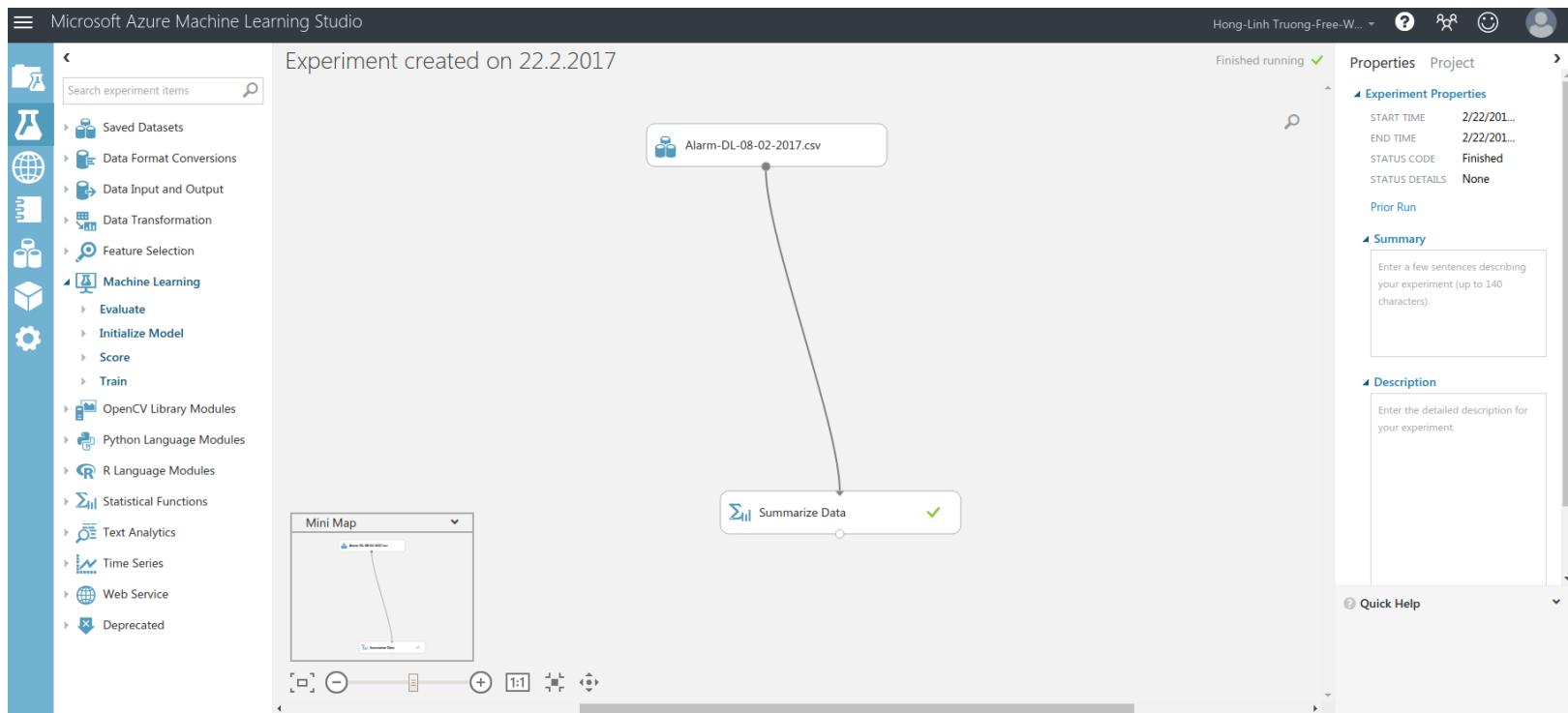
Chronograf

- Time-Series Data Visualization

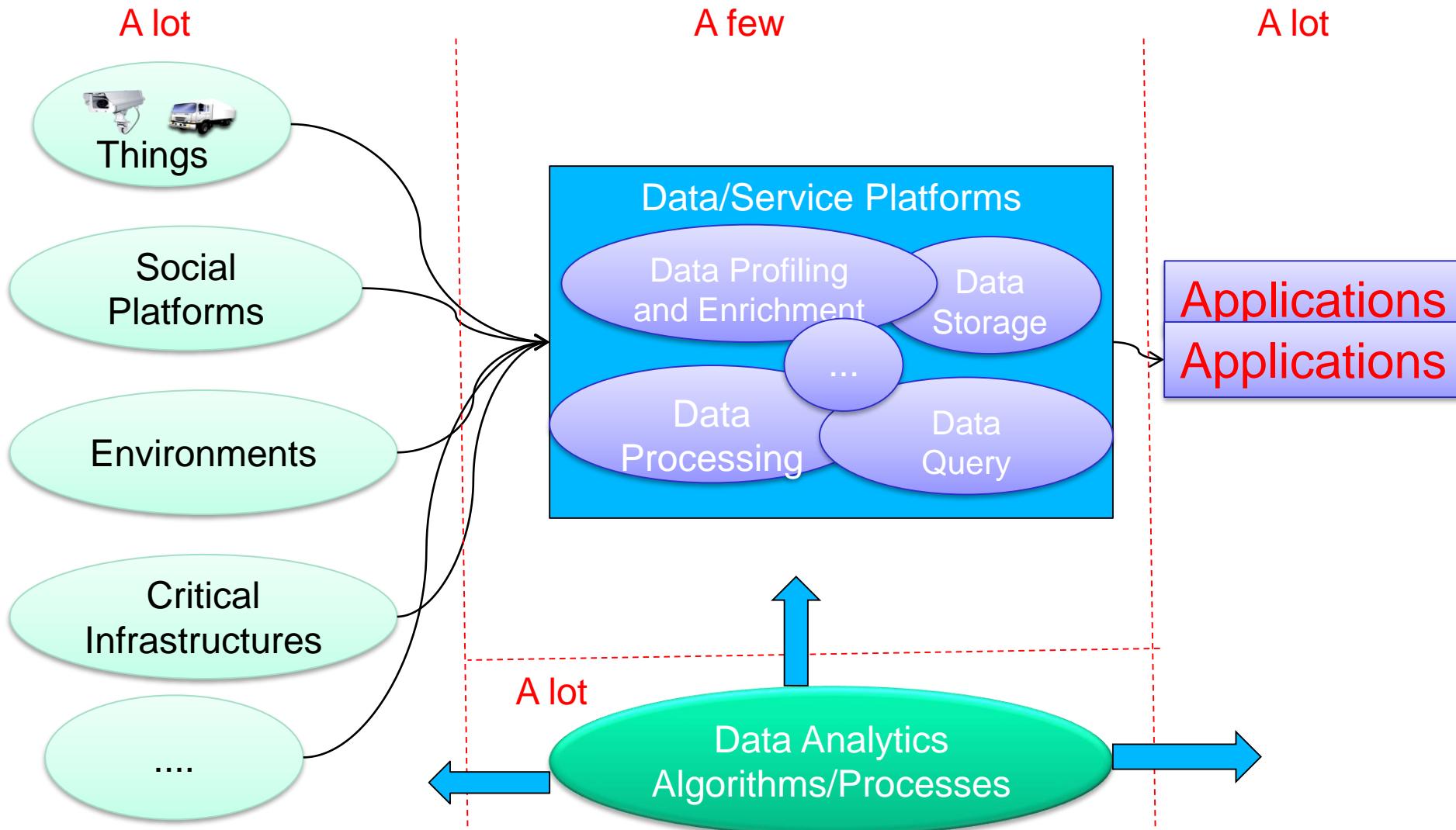
Kapacitor

- Time-Series Data Processing

Machine Learning Stack



Cloud-based Analytics



Emerging computing infrastructure/platform provisioning models— properties and issues

Some properties

- Rich types of services from multiple providers
 - Better choices in terms of functions and costs
- Concepts are similar but diverse APIs
- Strong dependencies/tight ecosystems

Some issues

- On-demand information management from multiple sources
- APIs complexity and API management
- Cross-vendor integration
- Execution in Multi-cloud environments
- Data locality
- Service mess/discovery

Emerging human computation models

- Crowdsourcing platforms
 - (Anonymous) people computing capabilities exploited via task bids
- Expert as Individual Compute Unit
 - An individual is treated like „a processor“ or “functional unit“. A service can wrap human capabilities to support the communication and coordination of tasks
- A set of individuals as *collectives*
 - A set of people and software that are initiated and provisioned as a service for solving tasks

The main point: humans are a computing element

Examples of human computation (2)

```

1 import edu.umass.cs.automan.adapters.MTurk._
2
3 object SimpleProgram extends App {
4   val a = MTurkAdapter { mt =>
5     mt.access_key_id = "XXXX"
6     mt.secret_access_key = "XXXX"
7   }
8
9   def which_one() = a.RadioButtonQuestion { q =>
10    q.budget = 8.00
11    q.text = "Which one of these does not belong?"
12    q.options = List(
13      a.Option('oscar, "Oscar the Grouch"),
14      a.Option('kermit, "Kermit the Frog"),
15      a.Option('spongebob, "Spongebob Squarepants"),
16      a.Option('cookie, "Cookie Monster"),
17      a.Option('count, "The Count")
18    )
19  }
20
21  println("The answer is " + which_one())
22 }
```



Source: Daniel W. Barowy, Charlie Curtsinger, Emery D. Berger, Andrew McGregor: **AutoMan: a platform for integrating human-based and digital computation**. OOPSLA 2012: 639-654

Discussion time:

**DO I NEED TO STUDY THEM
ALL? WHY?**

USE CASES/SCENARIOS

Critical infrastructures/services for citizens and business

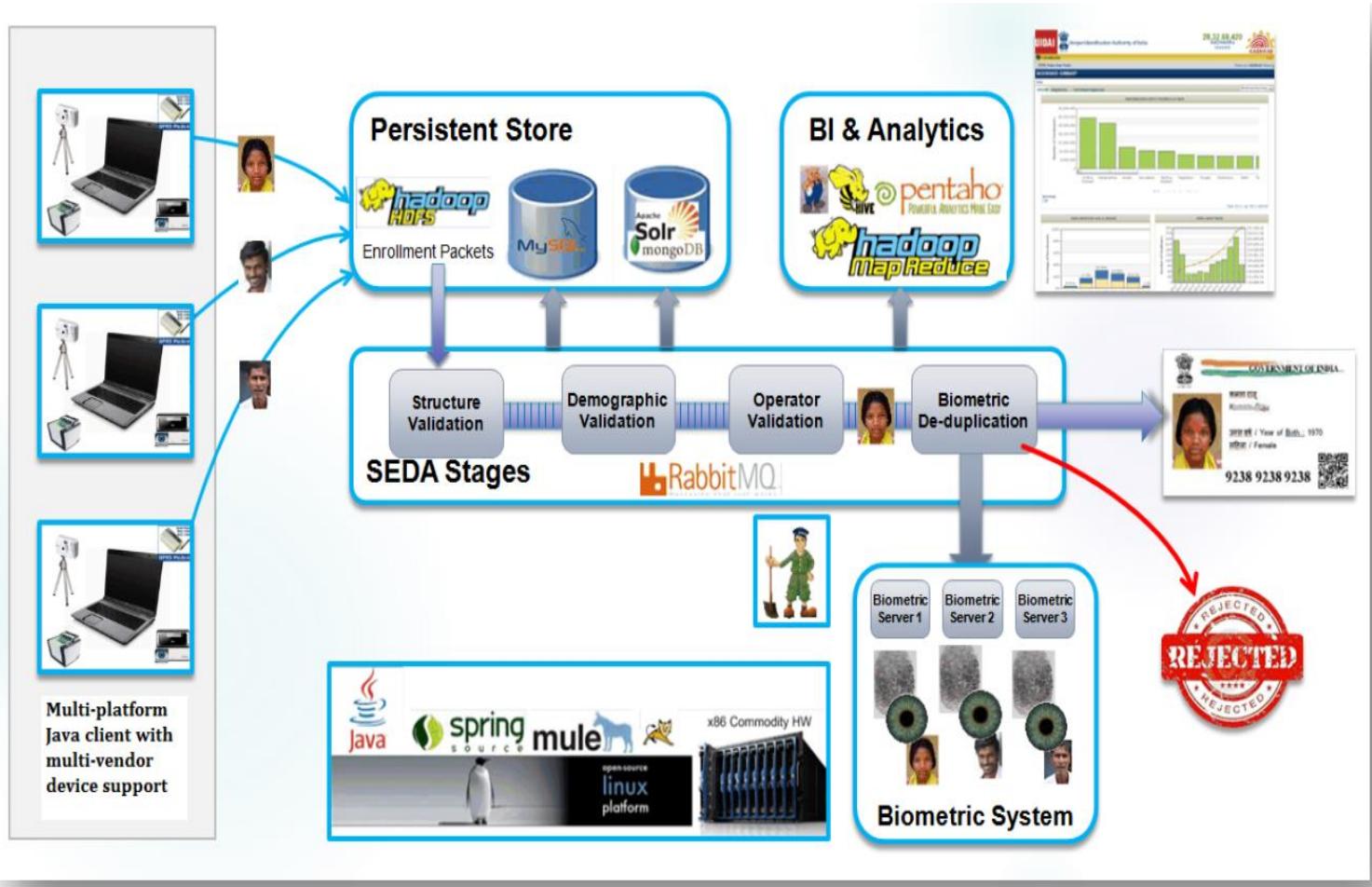
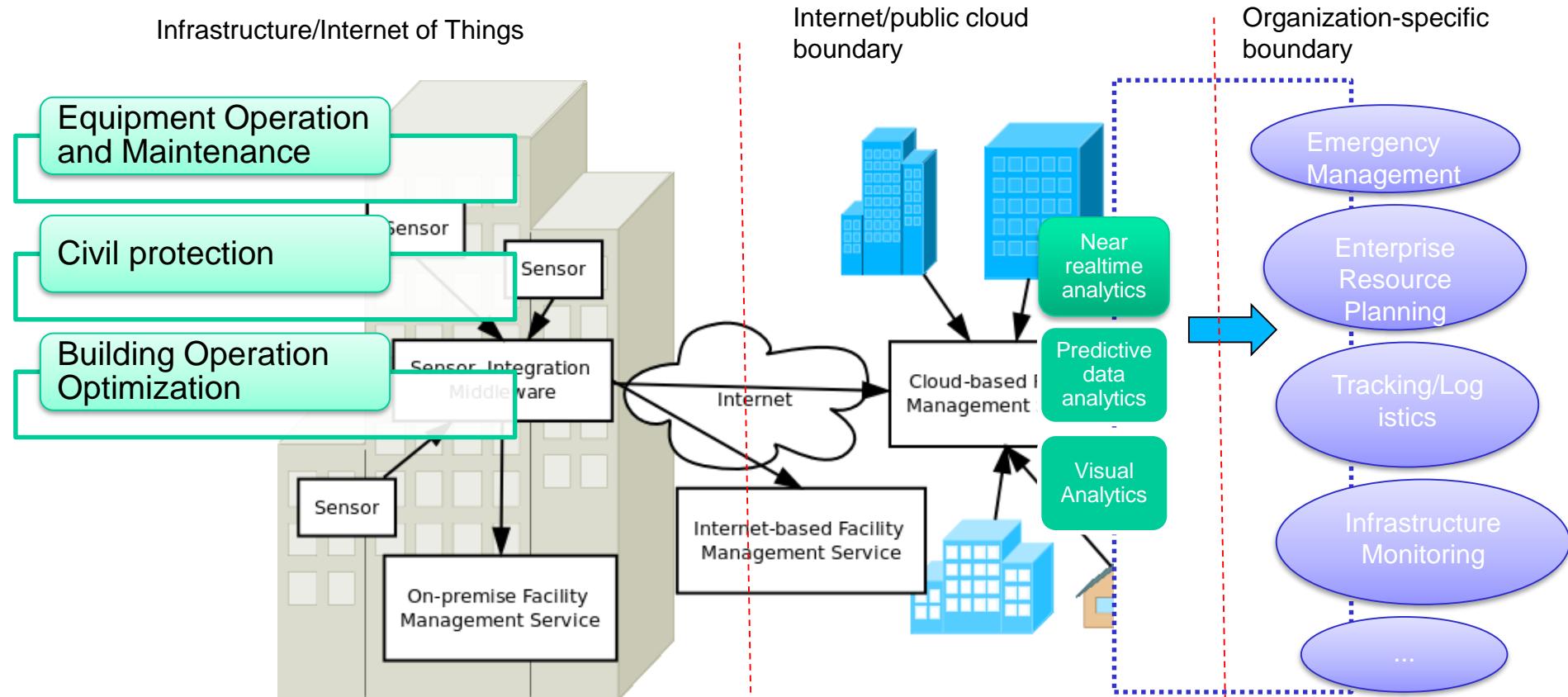


Figure source: http://uidai.gov.in/images/AadhaarTechnologyArchitecture_March2014.pdf

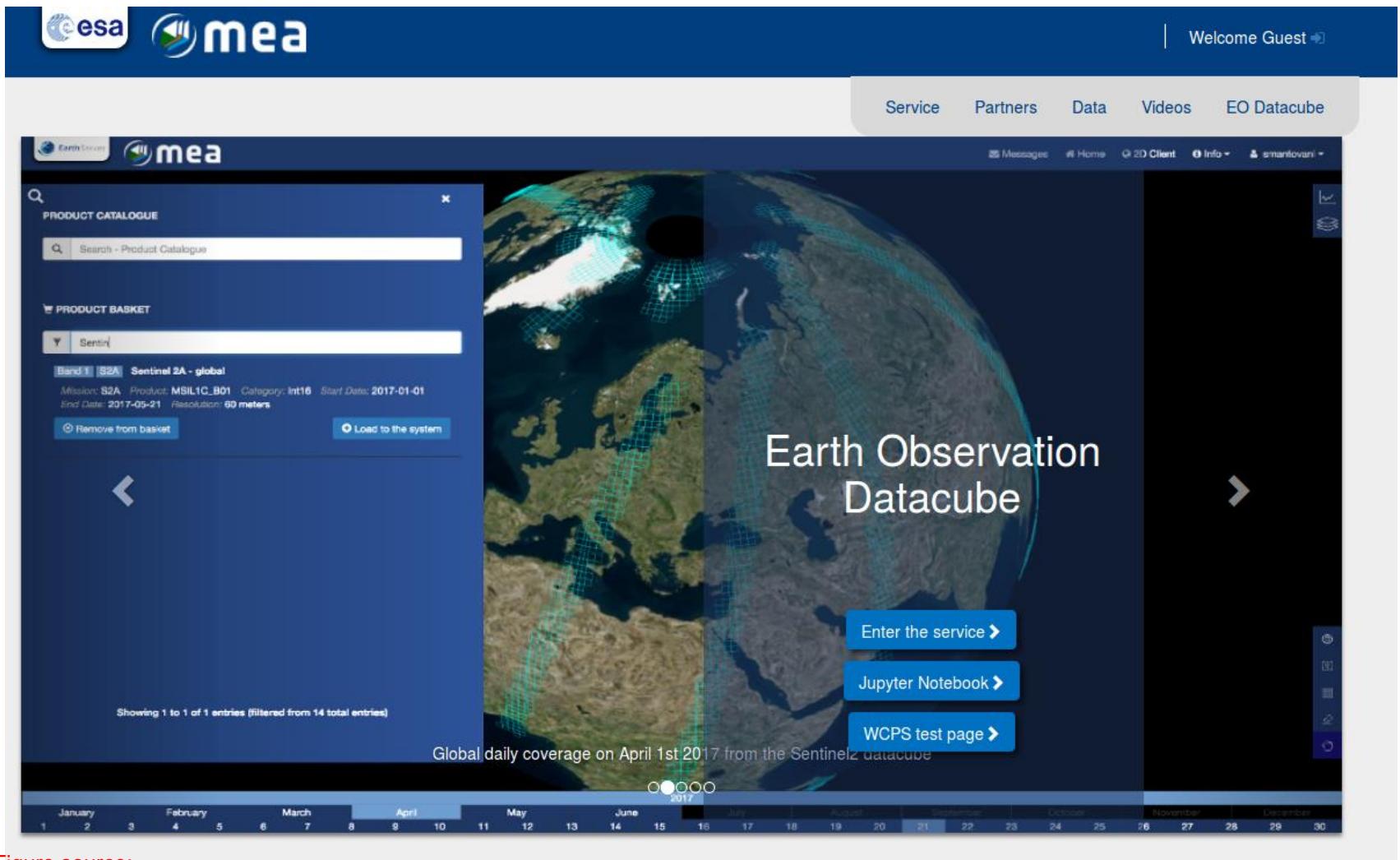
Smart building management



Cities, e.g. including:
10000+ buildings
1000000+ sensors

Can we combine open government data
with building monitoring data?

Earth Observation



The screenshot shows the homepage of the Earth Observation Datacube. At the top, there are logos for ESA and mea, and a welcome message "Welcome Guest". Below the header, there are navigation links for Service, Partners, Data, Videos, and EO Datacube. On the left, there is a "PRODUCT CATALOGUE" search bar and a "PRODUCT BASKET" section showing a selection for "Sentinel 2A - global". The main area features a large globe with a green grid overlay, representing global coverage. Overlaid on the globe is the text "Earth Observation Datacube". At the bottom of the globe, it says "Global daily coverage on April 1st 2017 from the Sentinel-2 datacube". There are three blue buttons with white text: "Enter the service >", "Jupyter Notebook >", and "WCPS test page >". A horizontal calendar at the very bottom shows the months from January to December, with the date April 1st highlighted.

Figure source:

<https://eodatacube.eu/>

ASE Fudan FIST, Summer 2018

Video analytics + business applications/public security

Use Case 3: Video Analytics

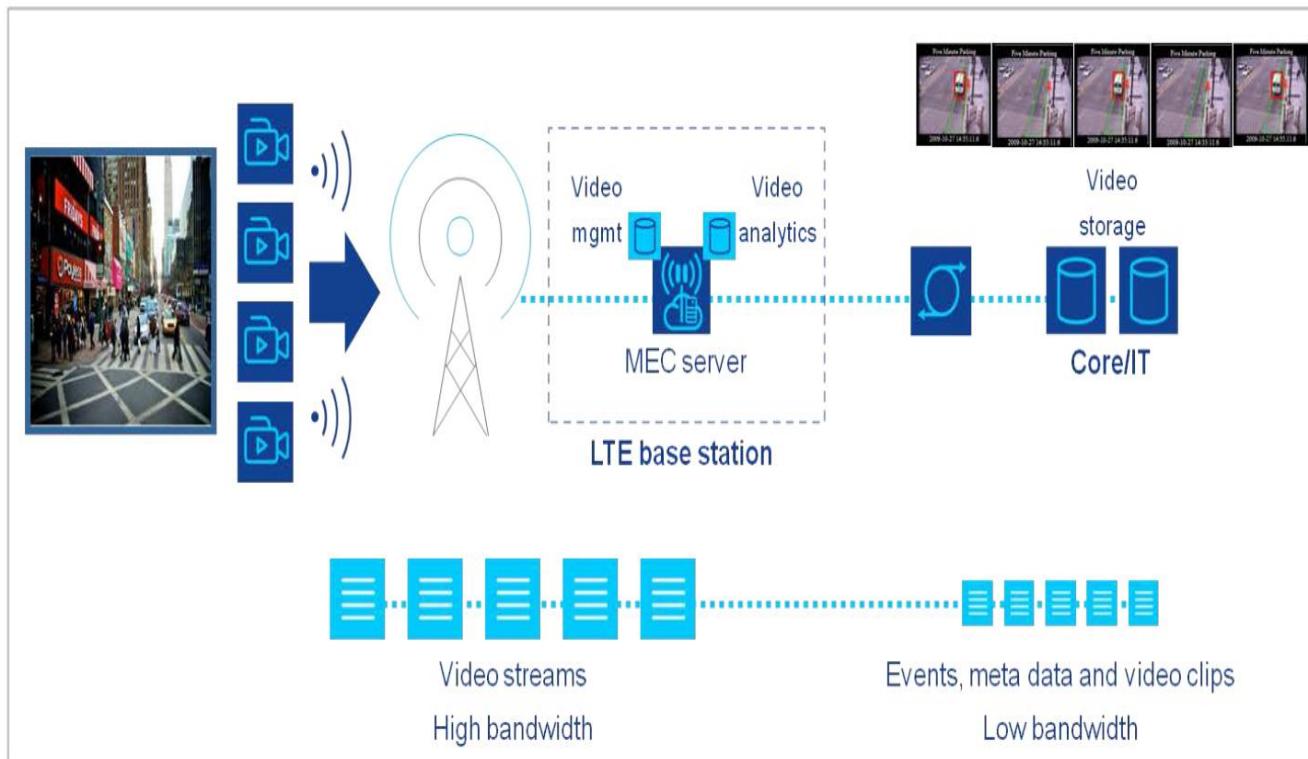


Figure 4: Example of video analytics

Figure source:

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

Edge/Cloud ML-based Video analytics

Chinese police are using smart glasses to identify potential suspects

Posted Feb 8, 2018 by Jon Russell (@jonrussell)



Figure source:
<https://techcrunch.com/2018/02/08/chinese-police-are-getting-smart-glasses/>

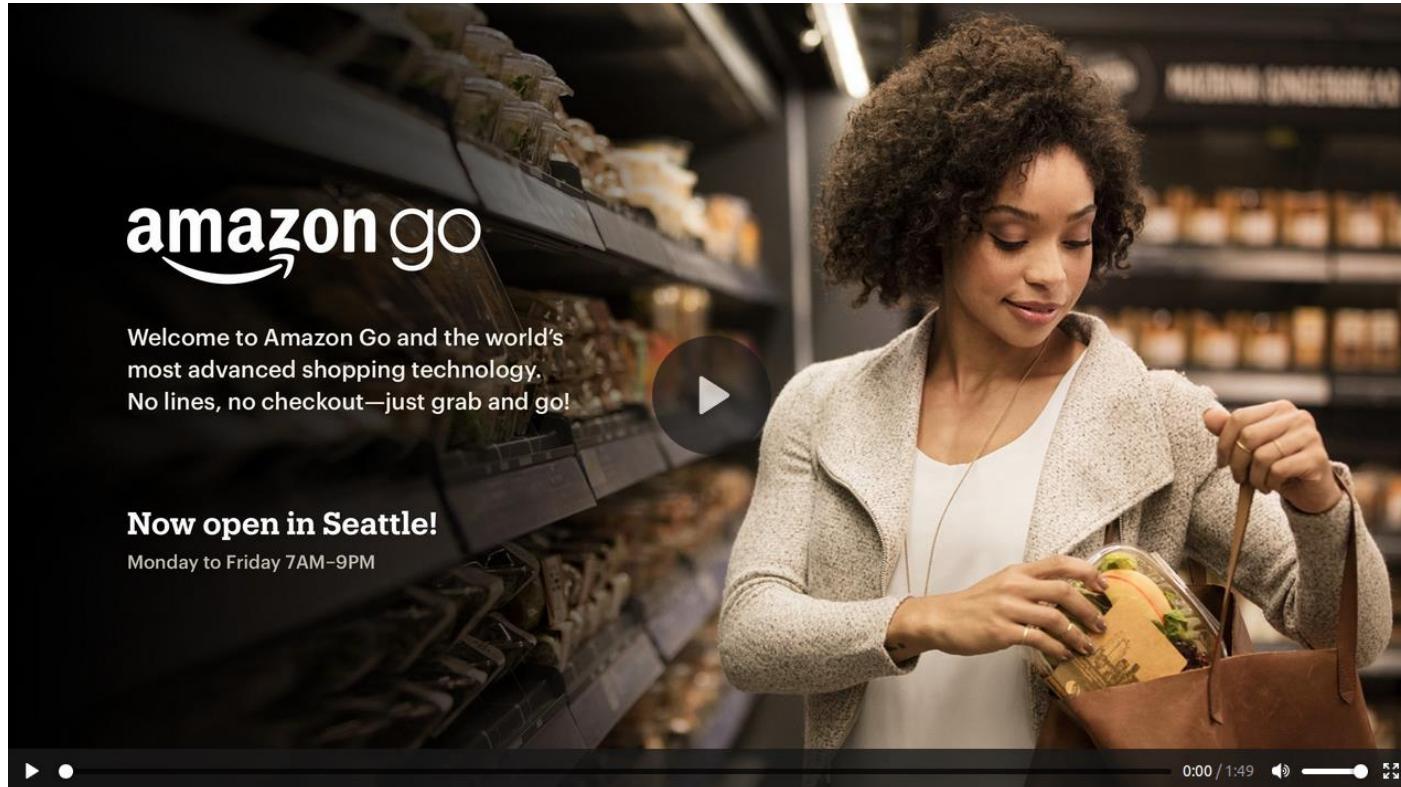


© SRI International

Figure caption 2: Video data include cabin and roadway views. This view illustrates possible relationships between observed driver behavior and the roadway context outside the vehicle, which includes other vehicles, cyclists, pedestrians, and lane markings.

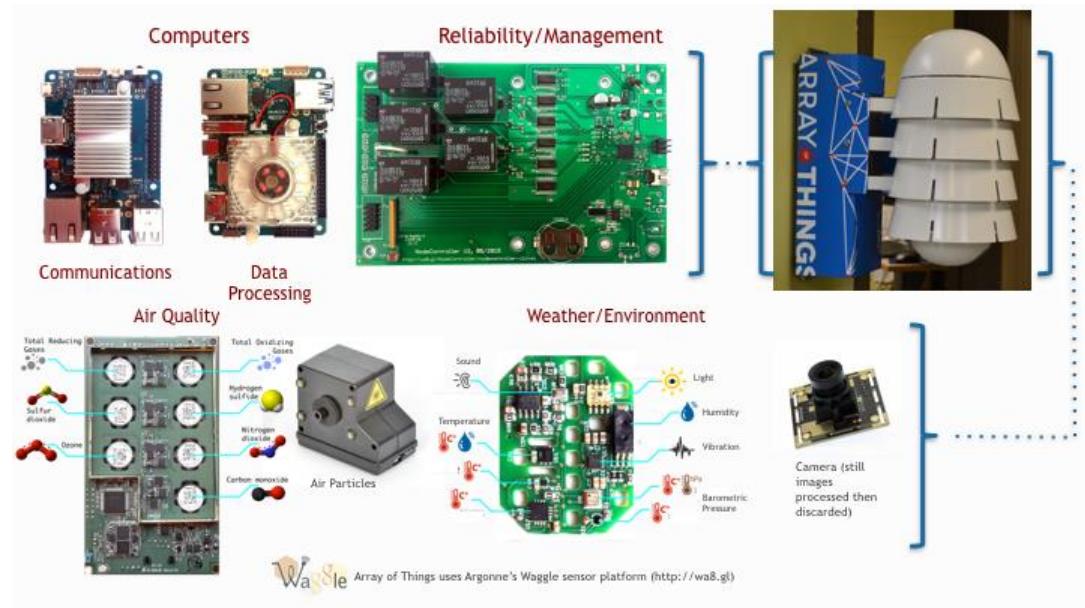
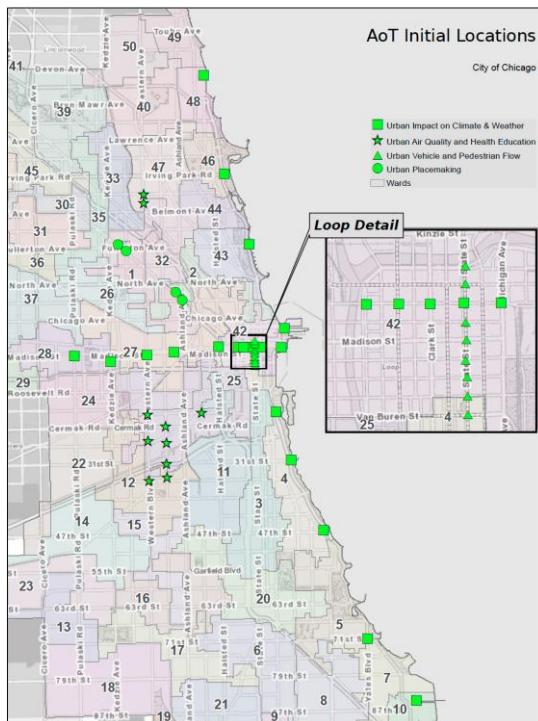
Figure source:
https://www.fhwa.dot.gov/research/resources/computervision_breakthrough.cfm

Shop scale



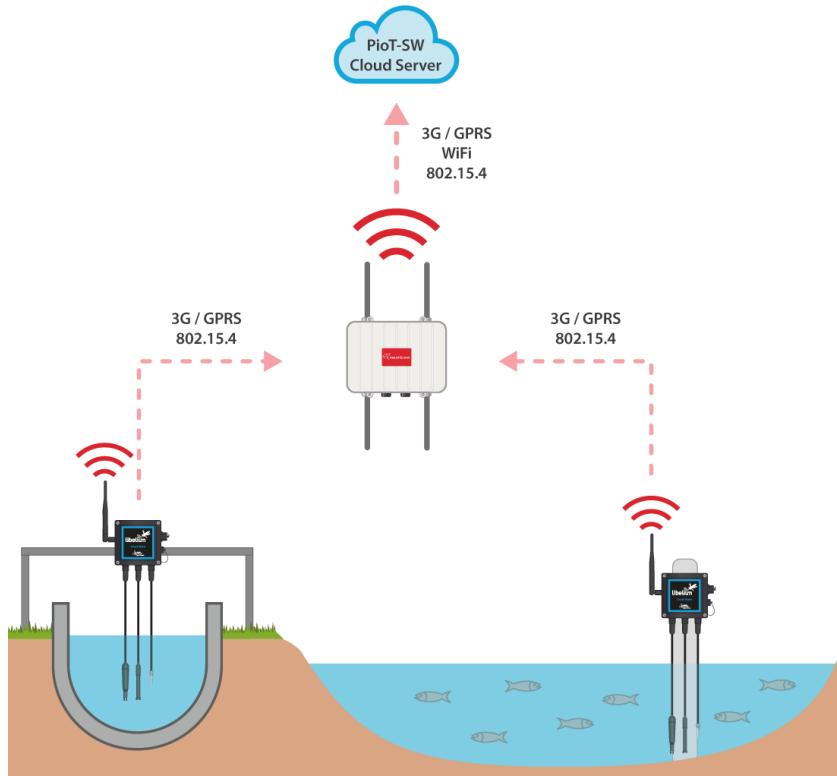
<https://www.amazon.com/b?ie=UTF8&node=16008589011>

IoT data in City-scale



Source: <https://arrayofthings.github.io/node.html>

Smart Farming



Source: <http://www.libelium.com/fish-farm-monitoring-in-vietnam-by-controlling-water-quality-in-ponds-and-tanks/>

Geo Sports in the stadium scale



Geo Sports: Picture courtesy
Future Position X, Sweden



Connected Cars

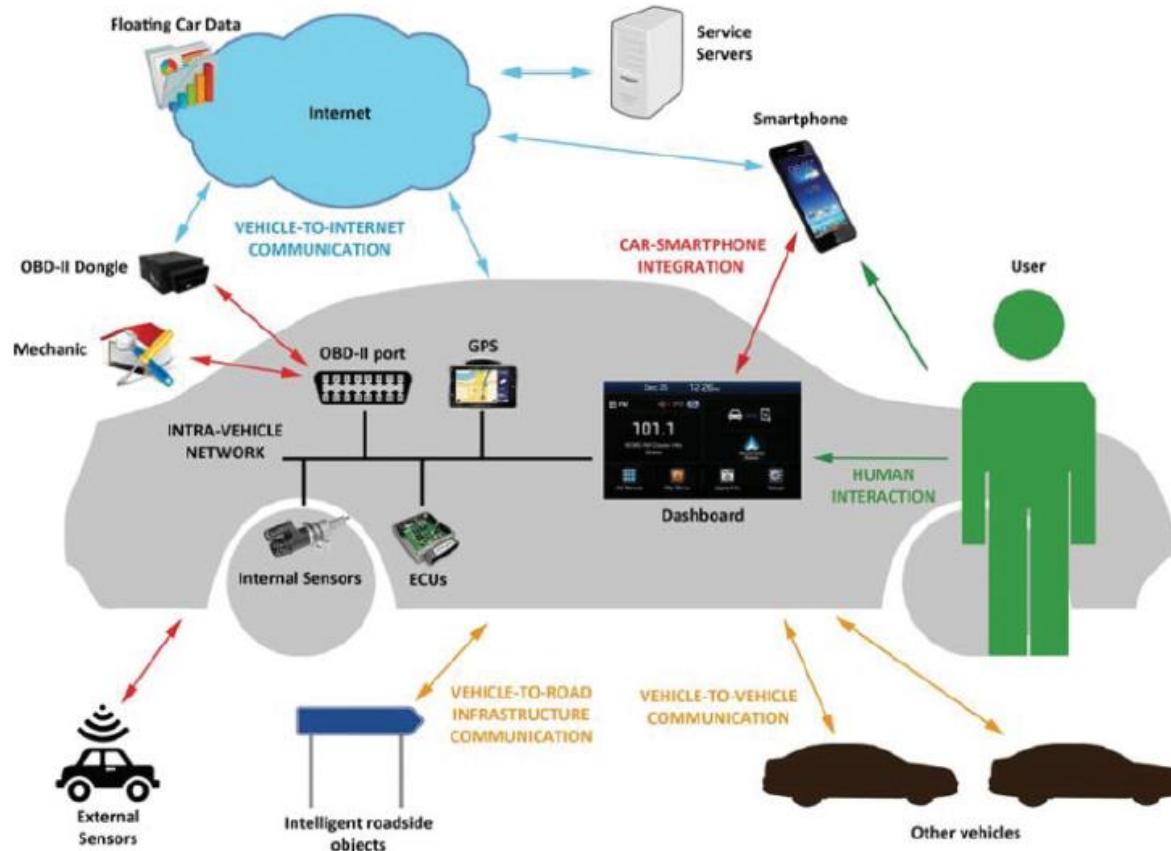


Fig. 3. Overview of the connected car system.

Source: Riccardo Coppola and Maurizio Morisio. 2016. Connected Car: Technologies, Issues, Future Trends. ACM Comput. Surv. 49, 3, Article 46 (October 2016), 36 pages. DOI: <https://doi.org/10.1145/2971482>

Drones for logistics

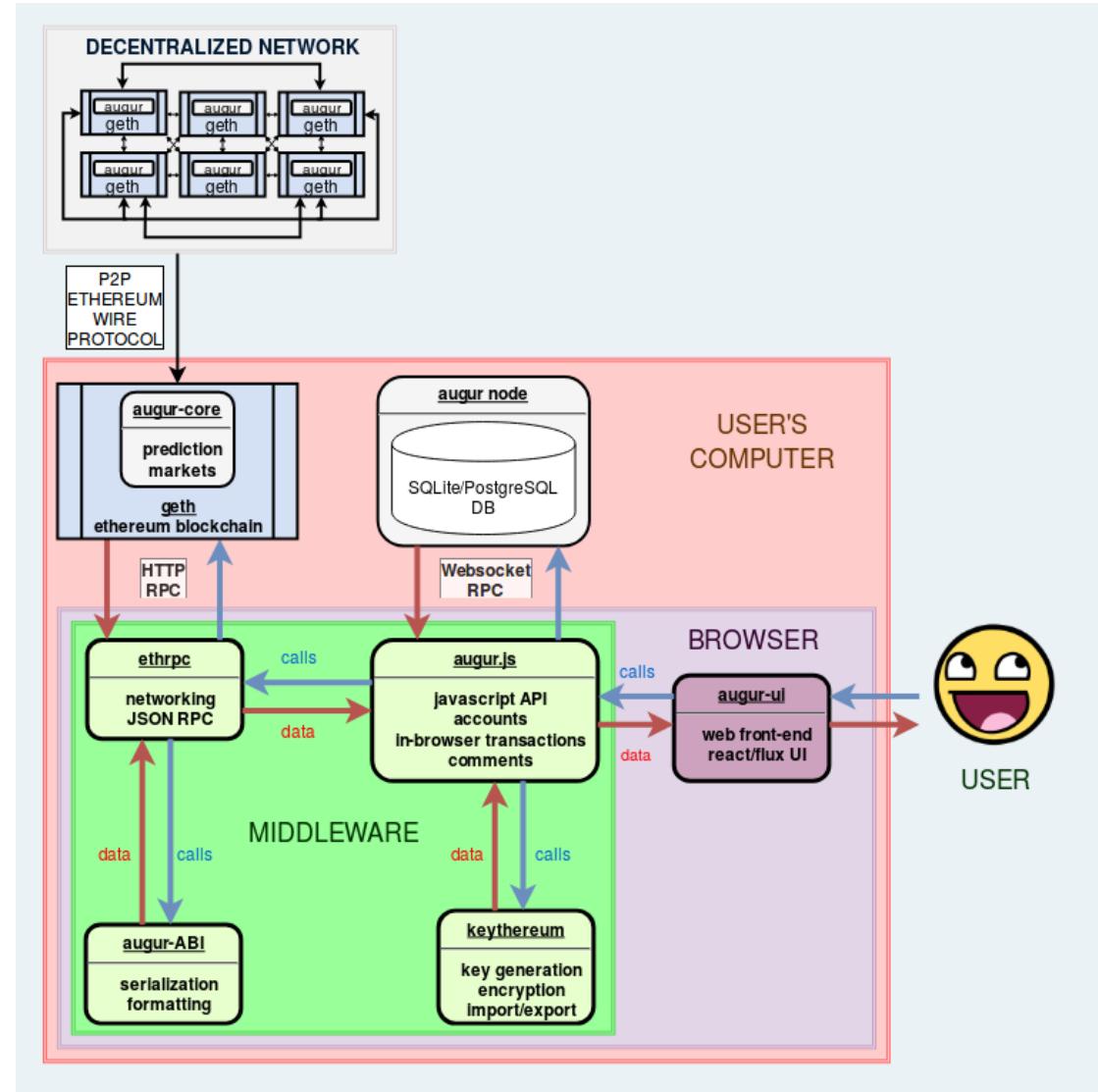


Figure 16: Urban First and Last Mile

Source: DHL Trend Report "Unmanned Aerial Vehicles"

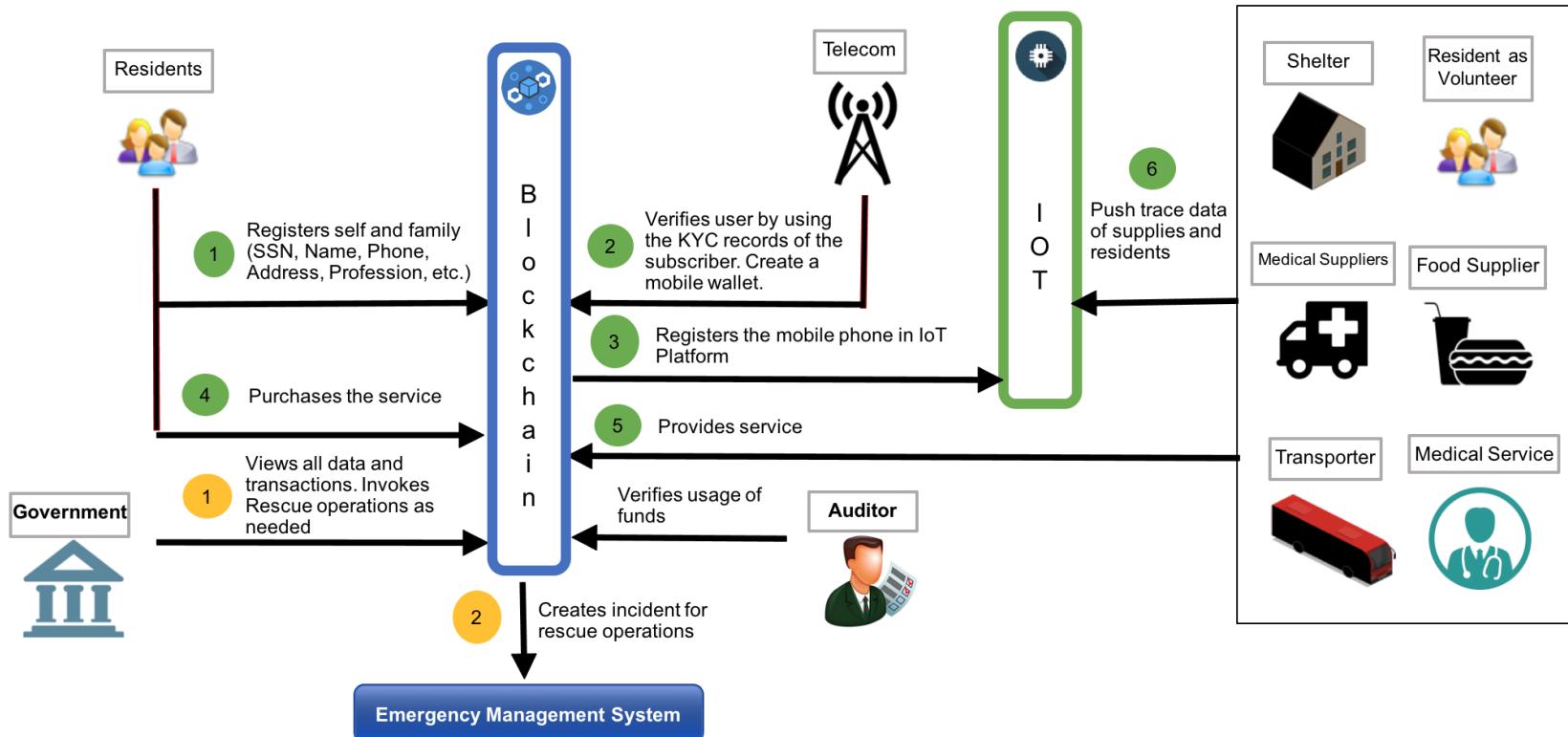
http://www.dhl.com/content/dam/downloads/g0/about_us/logistics_insights/dhl_trend_report_uav.pdf

Blockchain for prediction markets



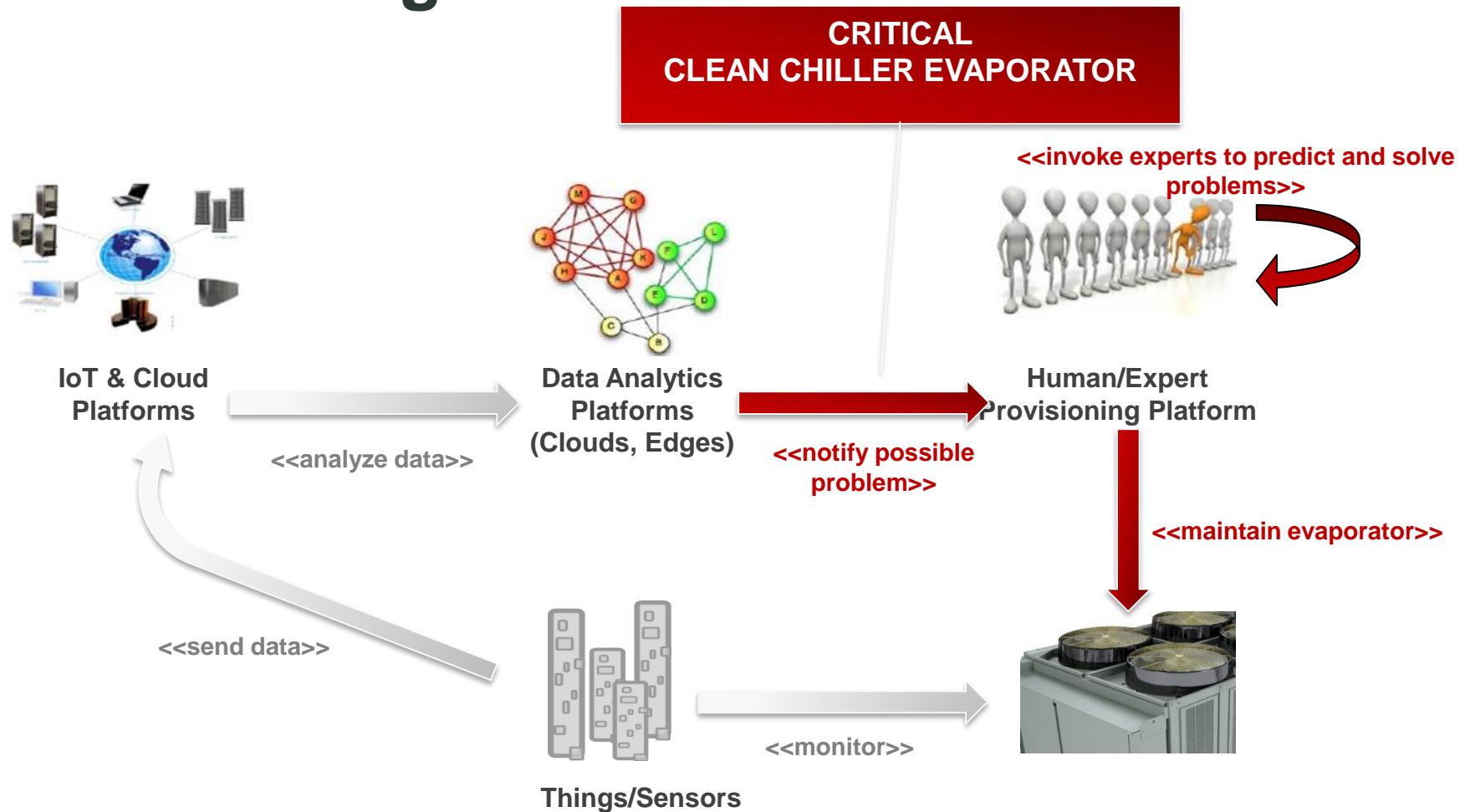
Source: <http://docs.augur.net>

Blockchain and IoT for Disaster Management?

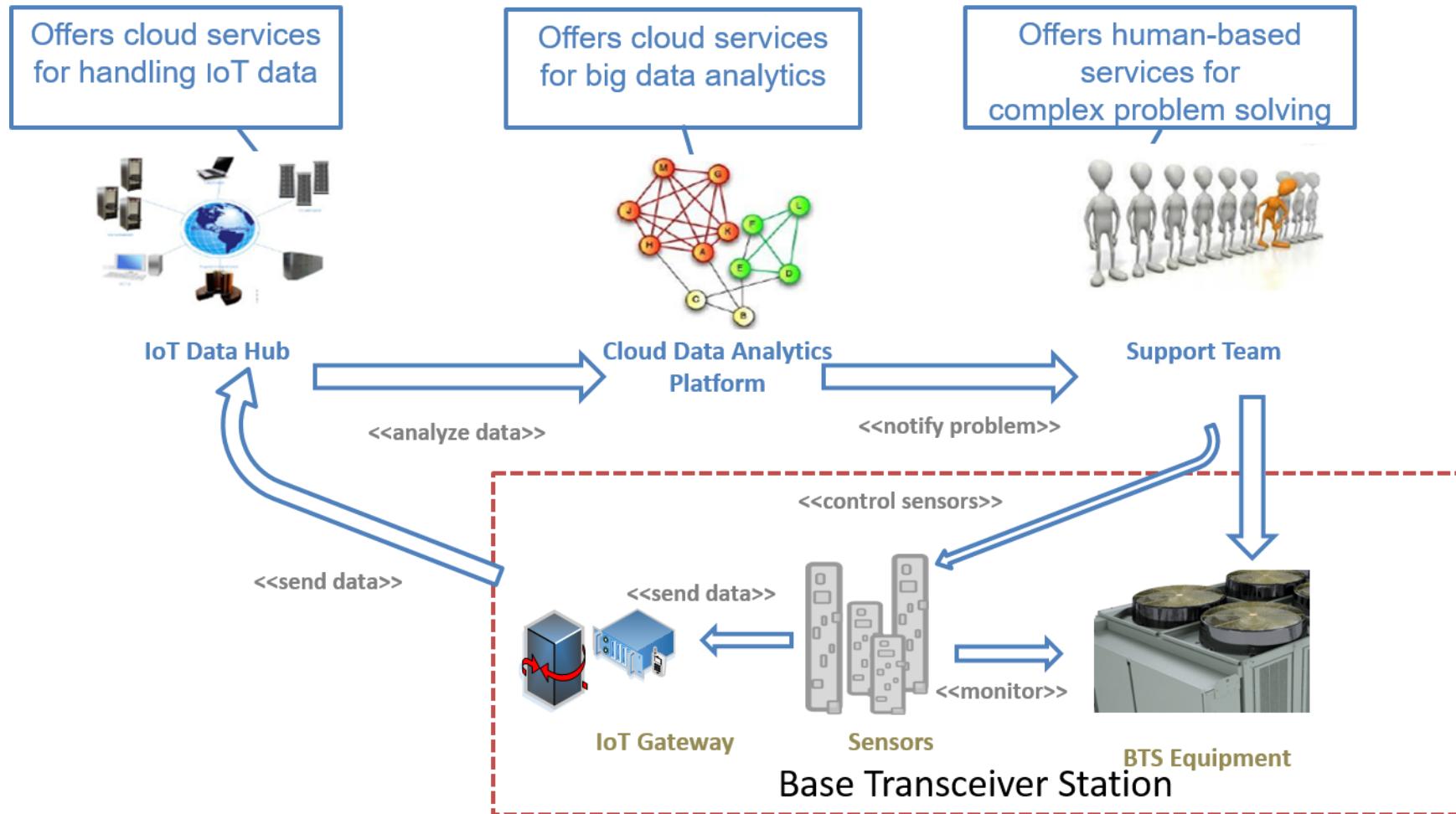


Source: <https://developer.ibm.com/blockchain/2017/12/09/disaster-management-using-blockchain-iot/>

Predictive Maintenance in Smart Buildings



Predictive Maintenance in Telcos



Do you see the big picture?

From ASE viewpoint: maybe you can have your own reasons for

October 31, 2017

3 Reasons Why Data
Engineering Is More
Important Than Data
Science for Companies
Today

Written By:

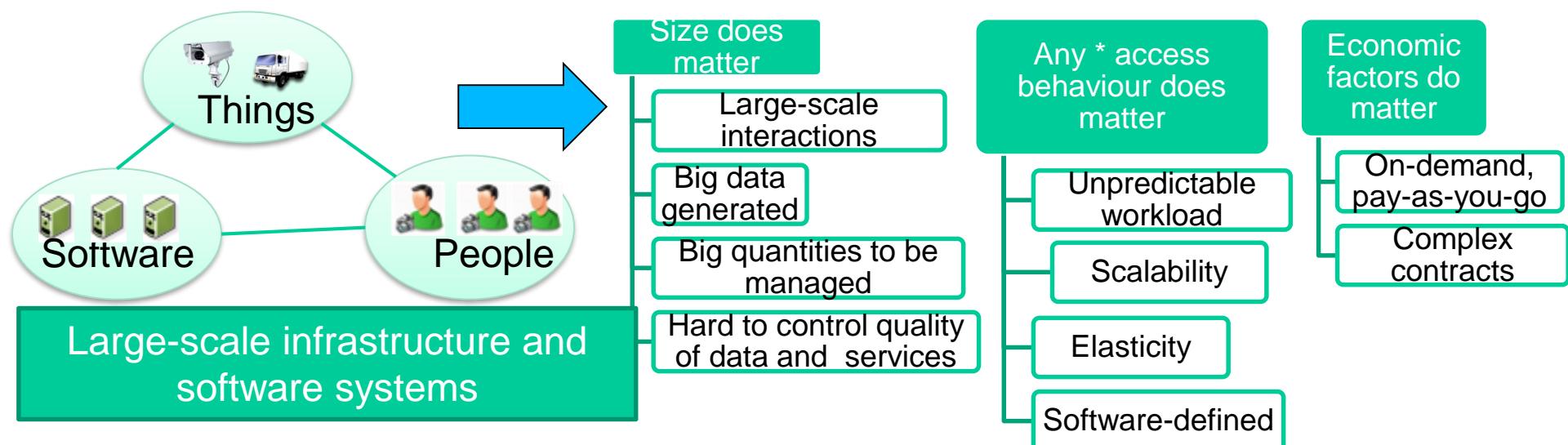
BEN HARDEN

Source: <https://www.captechconsulting.com/blogs/3-reasons-why-data-engineering-is-more-important-than-data-science-for-companies-today>

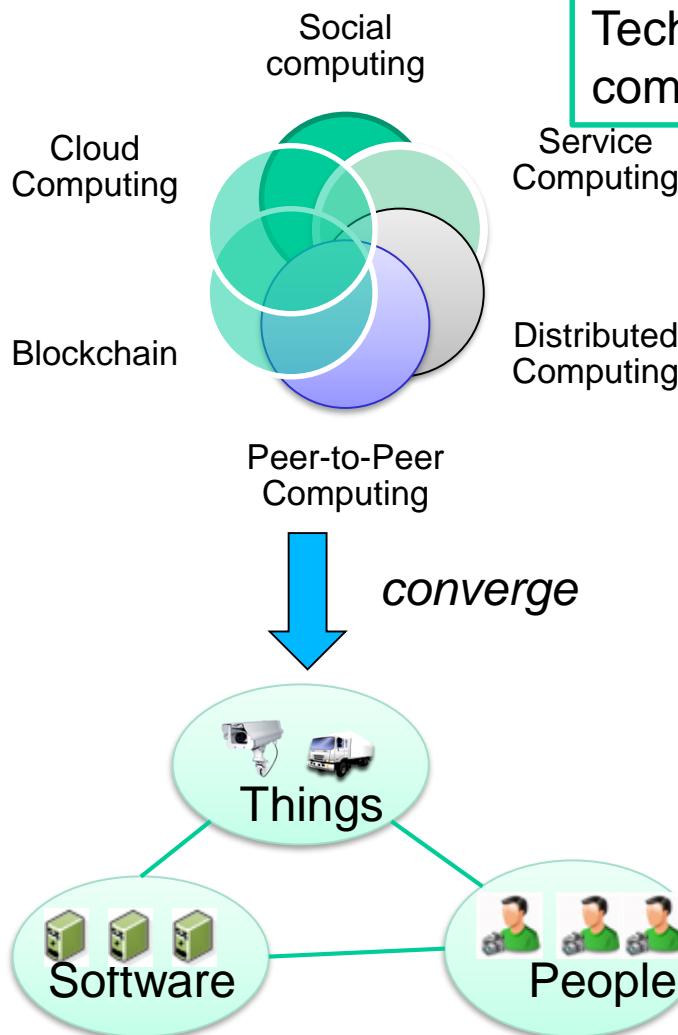
CONVERGENCE OF MULTIPLE COMPUTING MODELS

Today's Computing Models

- Internet infrastructure and software connect *contents, things, and people*, each has different roles (*computation, sensing, analytics, etc.*)



Today's Computing Models



Technologies and computing models

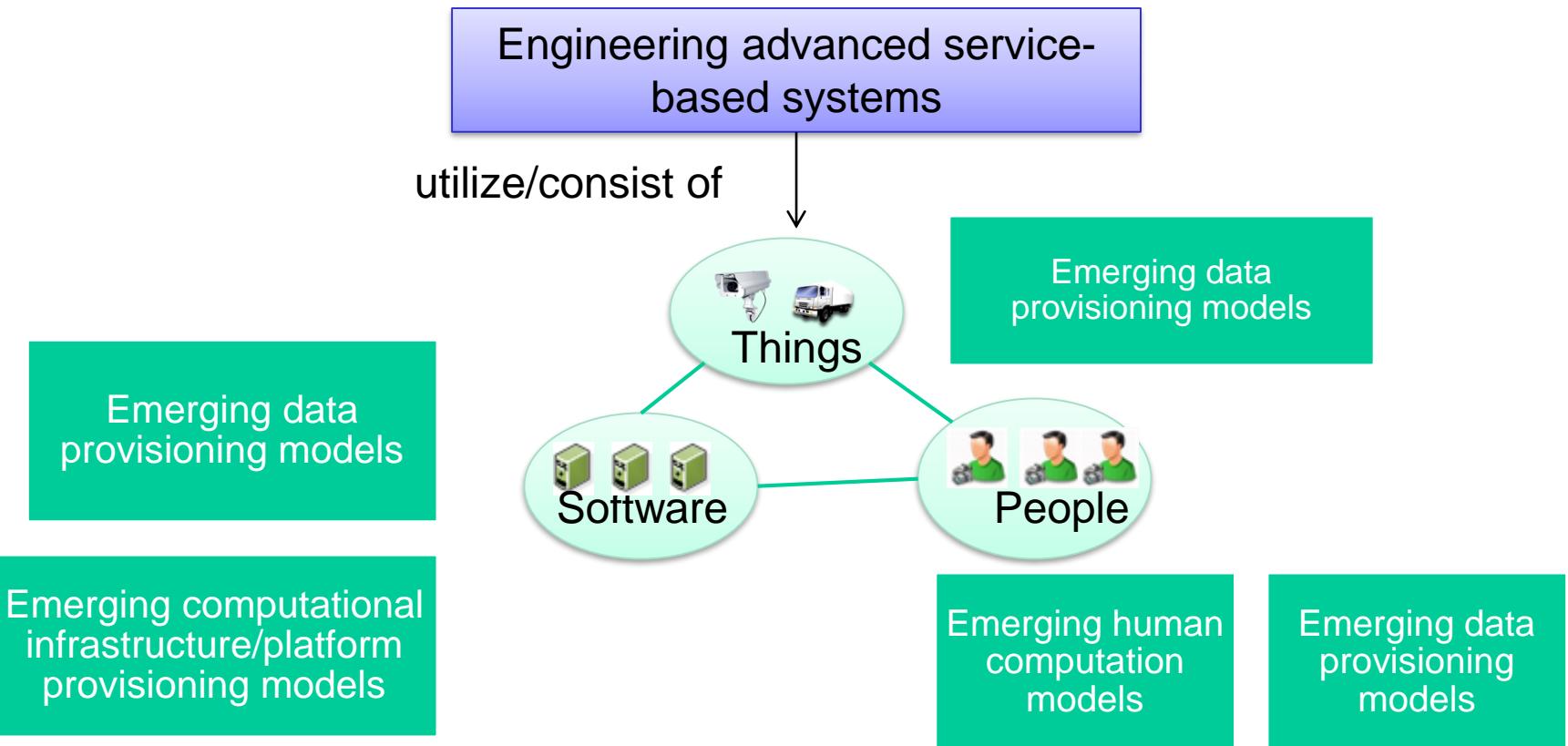
- Big and high performance centralized data analytics
- IoT data streaming analytics
- Large-scale applications spanning data centers and edge servers/gateways
- Adaptive collective systems of humans and machines



introduces

Emerging forms of computing models, systems and applications

Summary of emerging models wrt advanced service-based systems



Challenges in Virtualization, Programming, Communication, and Coordination, etc.

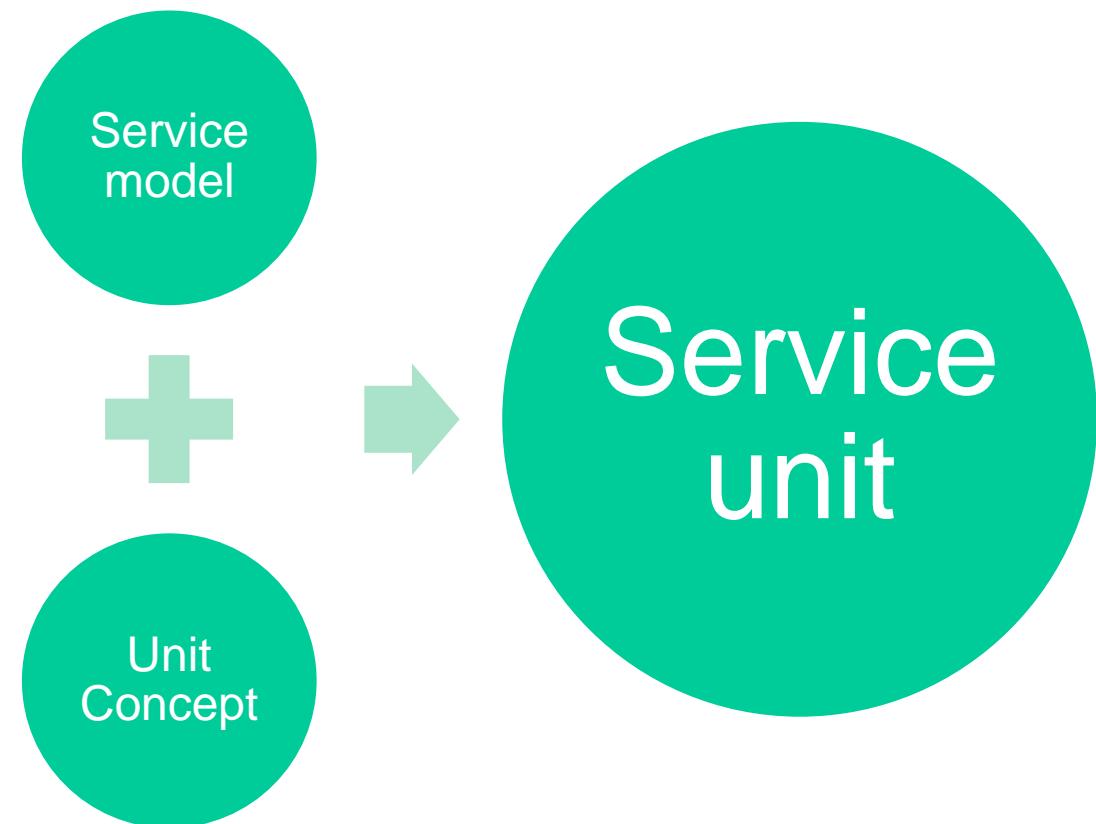
ADVANCED SERVICES ENGINEERING'S FOCUS

Single service/platform engineering(1)

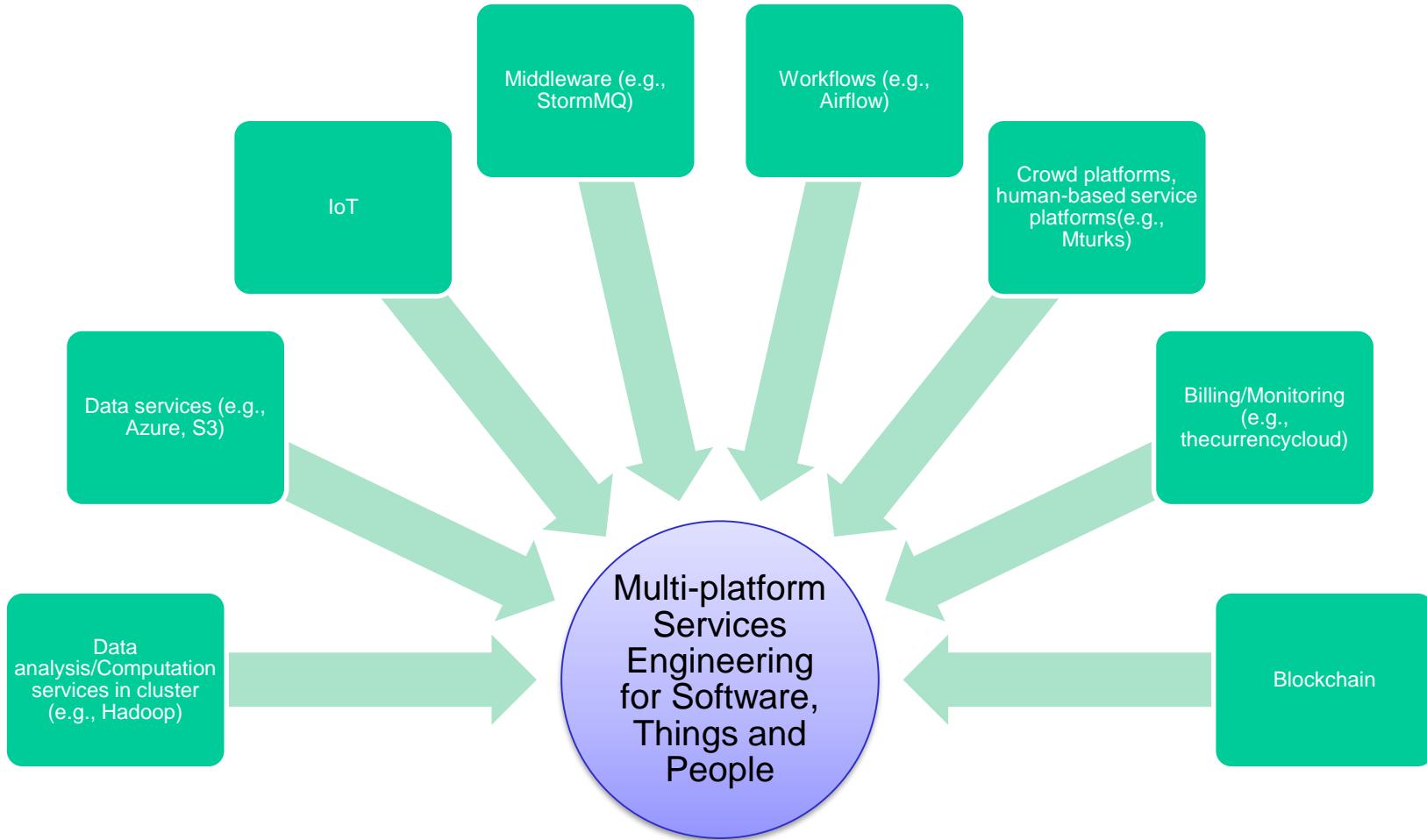
- The service model can be applied to things, people and software

Consumption,
ownership,
provisioning, price, etc.

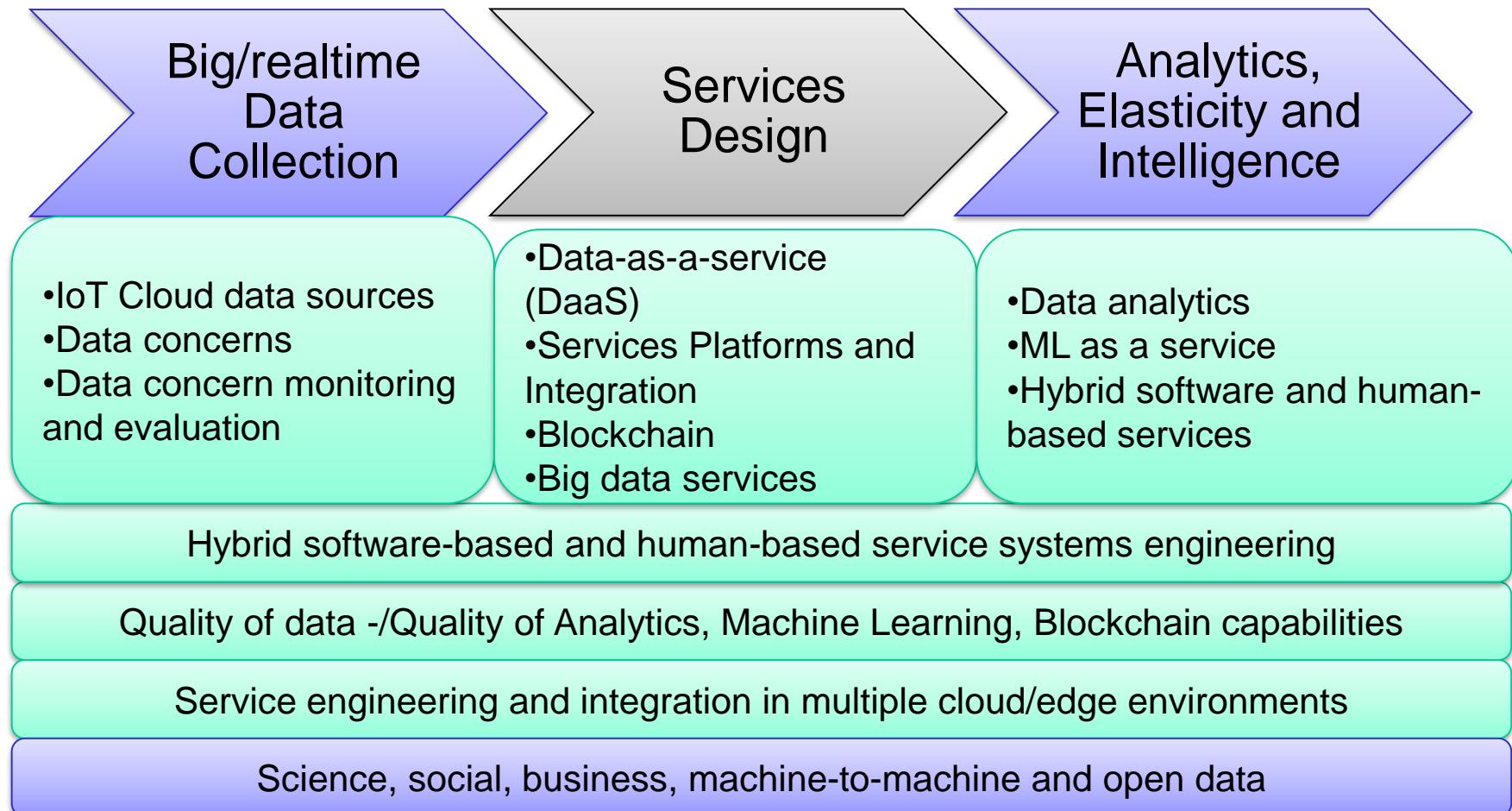
„basic
component“/“basic
function“ modeling
and description
/microservices



Internet-scale multi-platform services engineering – required technologies



Engineering methodology



Advanced service engineering -- Steps

Single service/platform engineering

Service units for representing fundamental things, people and software

Provisioning of fundamental service units

Engineering with single service units



Understanding Properties/Concerns

Data /Service/Application concerns; their dependencies

Monitoring, evaluation and provisioning of concerns

Utilization of data/service concerns



Large-scale, multi-platform services engineering

Identify platform/application problems

Identify the scale, complexity and *city

Design units, selection of existing service units;

Development and integration, optimization

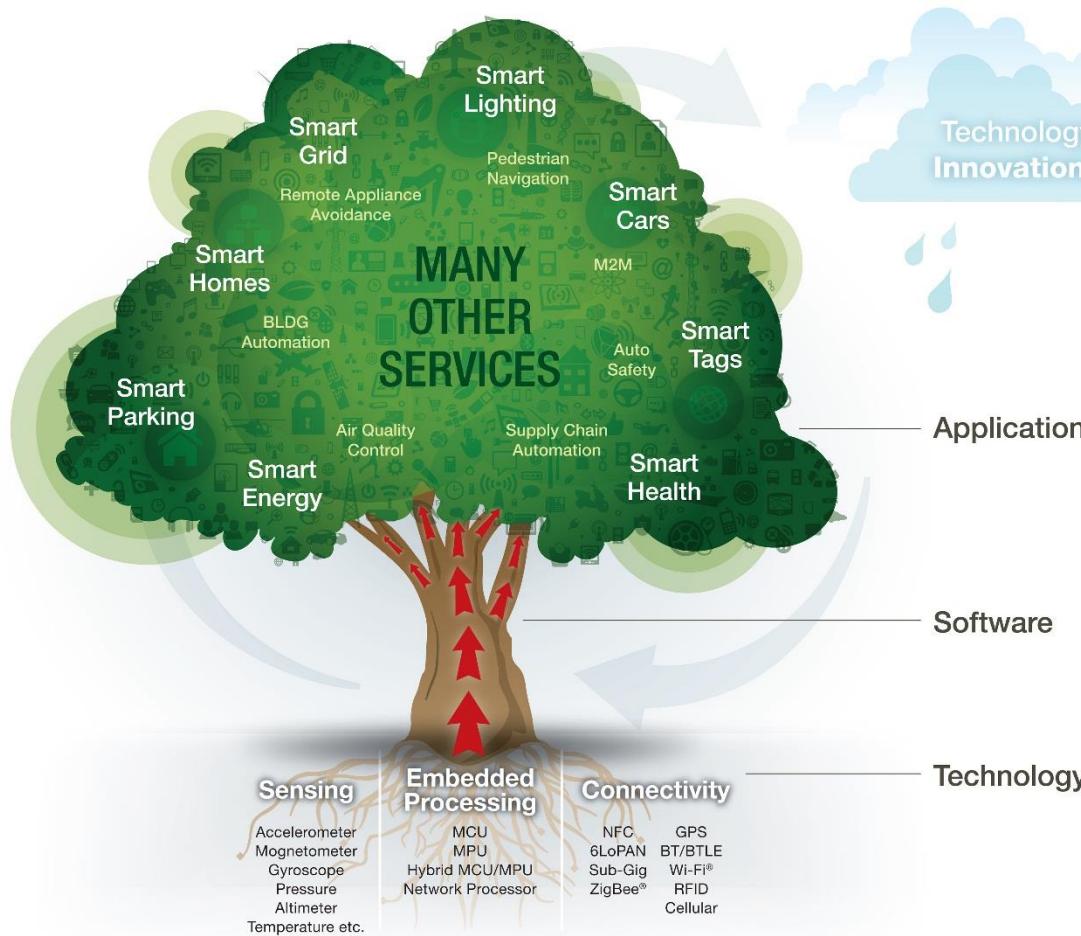
Exercises

- Read papers mentioned in slides
 - Get their main ideas
- Check services mentioned in examples
 - Examine capabilities of the mentioned services
 - Including price models and underlying technologies
 - Examine their size and scale
 - Examine their ecosystems and dependencies
- Work on possible categories of single service units that are useful for your work
 - Some common service units with capabilities and providers

EXAMPLES FROM PREVIOUS YEARS

<http://www.infosys.tuwien.ac.at/teaching/courses/ase/#examples>

Define your scenario



Remember we
can also
suggest the
topic!

Source: <http://eecatalog.com/IoT/files/2014/04/Freescale-Internet-of-Things-Tree.jpg?file=2014/04/Freescale-Internet-of-Things-Tree.jpg>

Thanks for your attention

Hong-Linh Truong
Faculty of Informatics, TU Wien
hong-linh.truong@tuwien.ac.at
<http://www.infosys.tuwien.ac.at/staff/truong>