

# Emerging Distributed Computing and Challenges for Services Engineering

Hong-Linh Truong  
Distributed Systems Group, TU Wien

[truong@dsg.tuwien.ac.at](mailto:truong@dsg.tuwien.ac.at)  
[@linhsolar](http://dsg.tuwien.ac.at/staff/truong)

# Goals

- See emerging trends in distributed systems and computing
- Have a critical look at use cases and analyze use cases
- See the service engineering technologies needed for such use cases



# Outline

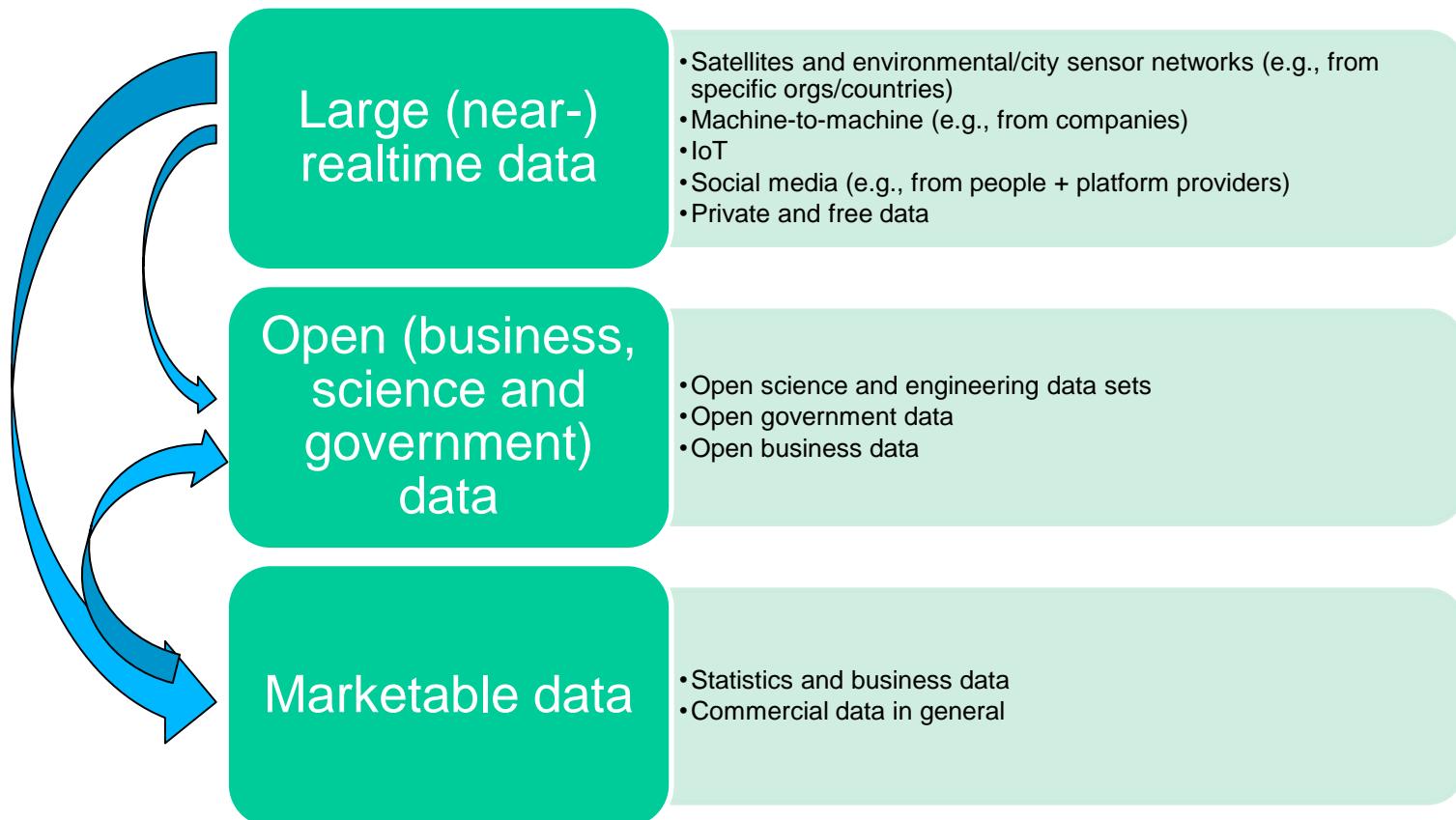
- Some emerging models – properties and issues
  - IoT resources
  - Data provisioning models
  - Computational infrastructures/frameworks provisioning
  - Human computation provisioning
  - Software-defined \*
- Use cases
- Advanced services engineering
  - Single service/platform engineering
  - Multi-platform services engineering



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



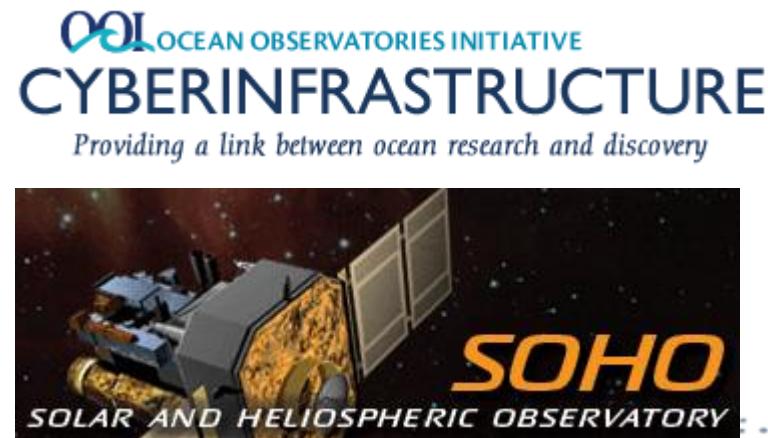
# Emerging data provisioning models



Data are assets



# Examples of large-scale (near-) realtime data



# Large-scale (near-)realtime data: properties and issues

## Some properties

- Having massive data
- Requiring large-scale, big (near-) real time processing and storing capabilities
- Enabling predictive and realtime data analytics

## Some issues

- Timely analytics
  - Performance and scalability
- Quality of data control
- Handle of unknown data patterns
- Benefit/cost versus quality tradeoffs



# 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 [+]

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, BC 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
 

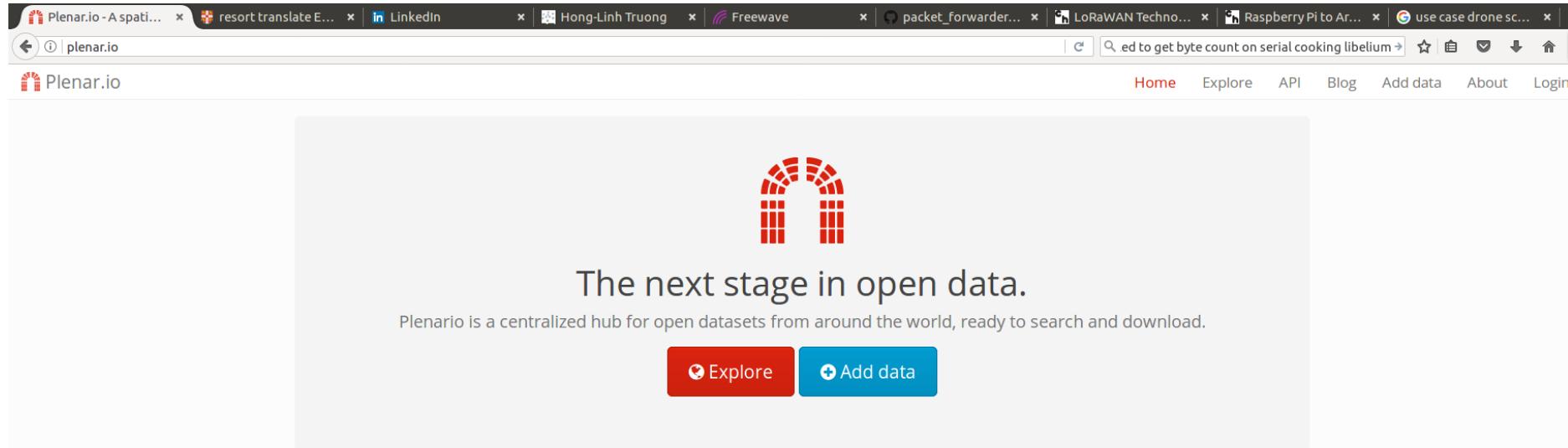
- 1000 Genomes Project
- Common Crawl Corpus
- Google Books Ngrams

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



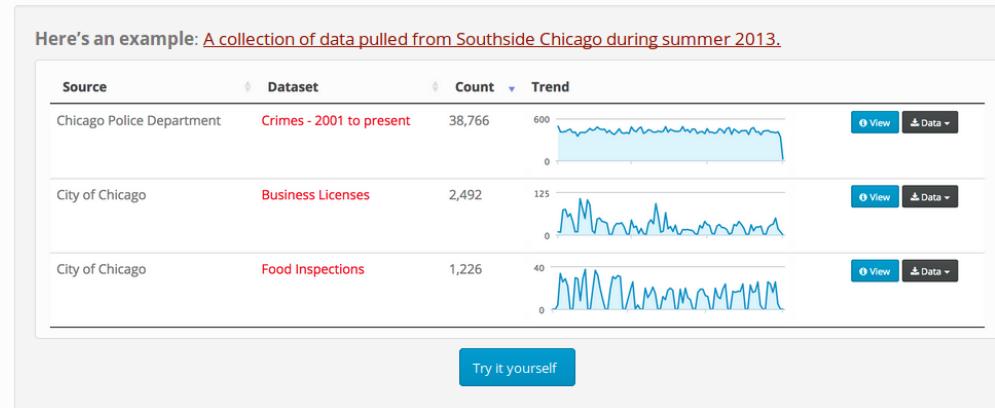
The next stage in open data.

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](#)

# Open data: properties and issues

## Some properties

- Having large, multiple data sources but mainly static data
  - Real-time, open data is growing
- Having good quality control in many cases
- Usually providing the data as a whole set

## Some issues

- Fine-grained content search
- Balance between processing cost and performance
- Correlation/combination with real-time/private data



# 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

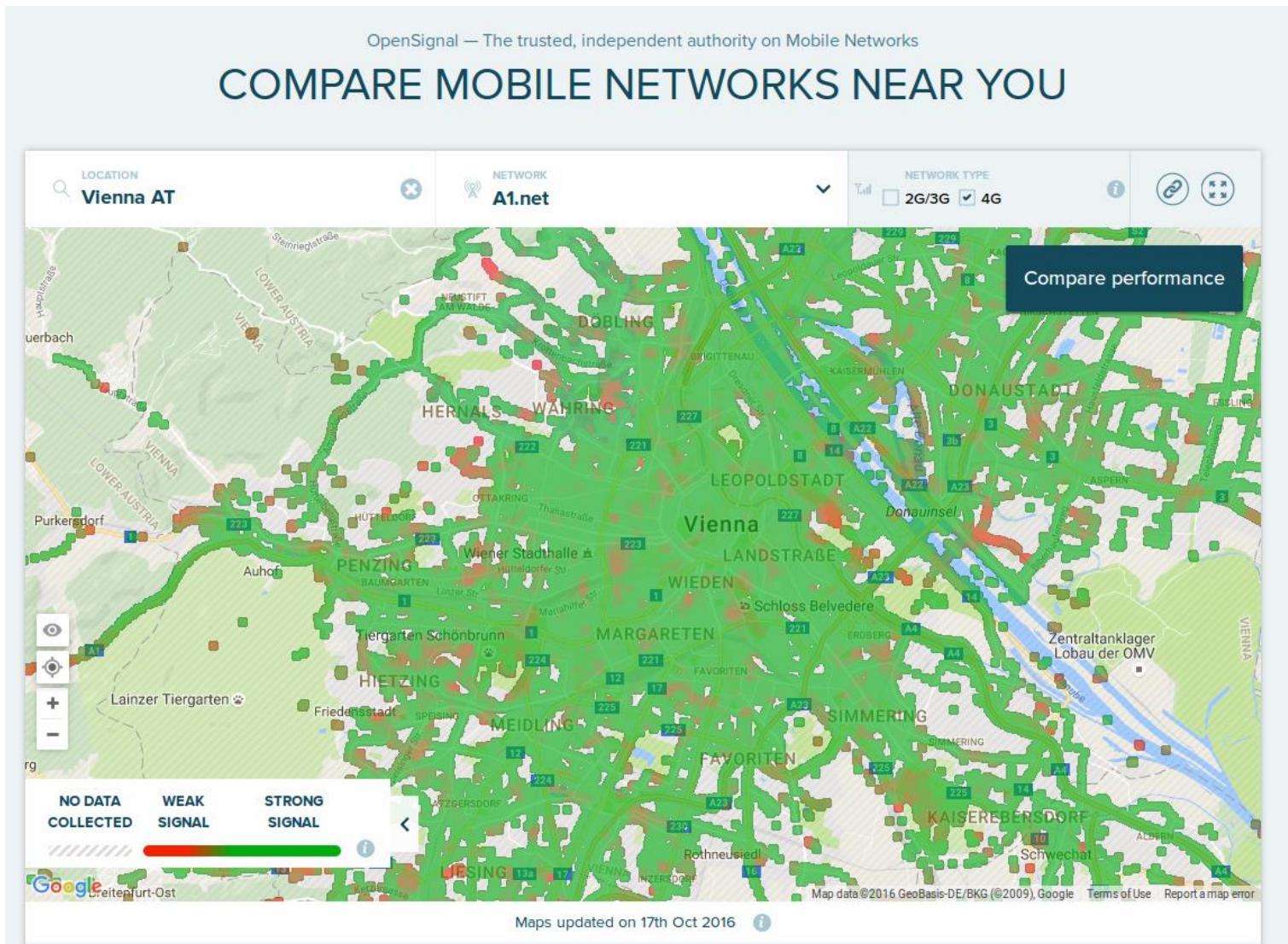
## Real-Time Targeting

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 >](#)



# Marketable data: properties and issues

## Some properties

- Can be large, multiple data sources but mainly static data
- Having good quality control
- Have strong data contract terms
- Some do not offer the whole dataset

## Some issues

- Multiple levels of service/data contracts
- Compatible with other data sources w.r.t. contract
- Cost w.r.t. up-to-date data
- Near-realtime data marketplaces



# 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)



# 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
  - Etc.



# Examples

Data Processing Framework

**Amazon Elastic MapReduce**

**Cloudera Impala**

**Apache Kylin**

**Google Cloud Dataflow**

**Summingbird**

**Azure Stream Analytics**

Data  
Transfer/Messaging  
Middleware

**StormMQ**

**Apache Nifi**

**Kafka**

**MQTT**

**AmazonSQS**

**CloudAMQP**

**Google BigQuery**

**Elastic Search**

**Influx DB**

Data Storages

**MongoLab**

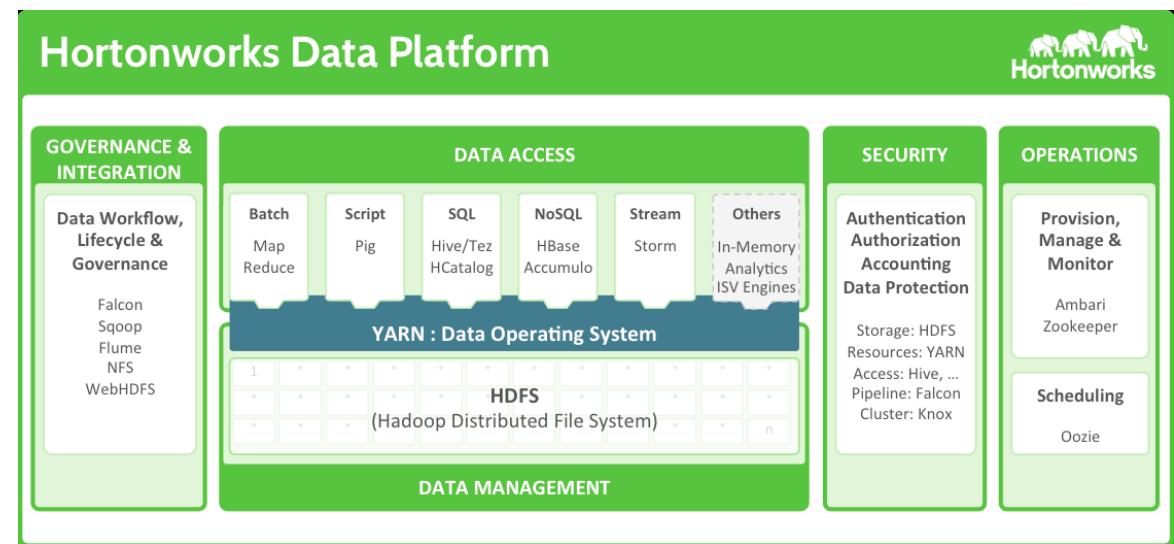
**Amazon S3**

**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

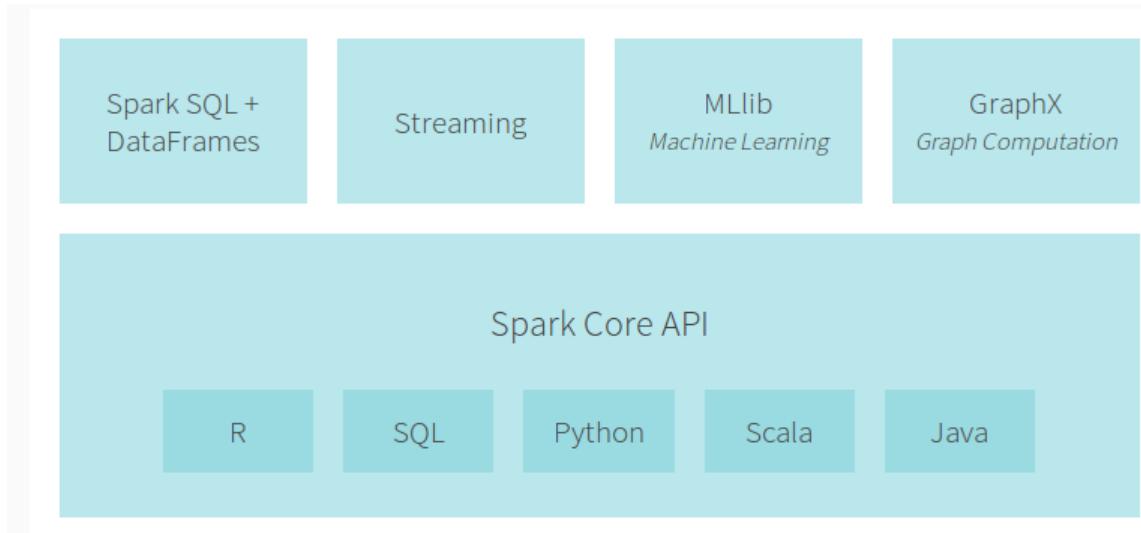
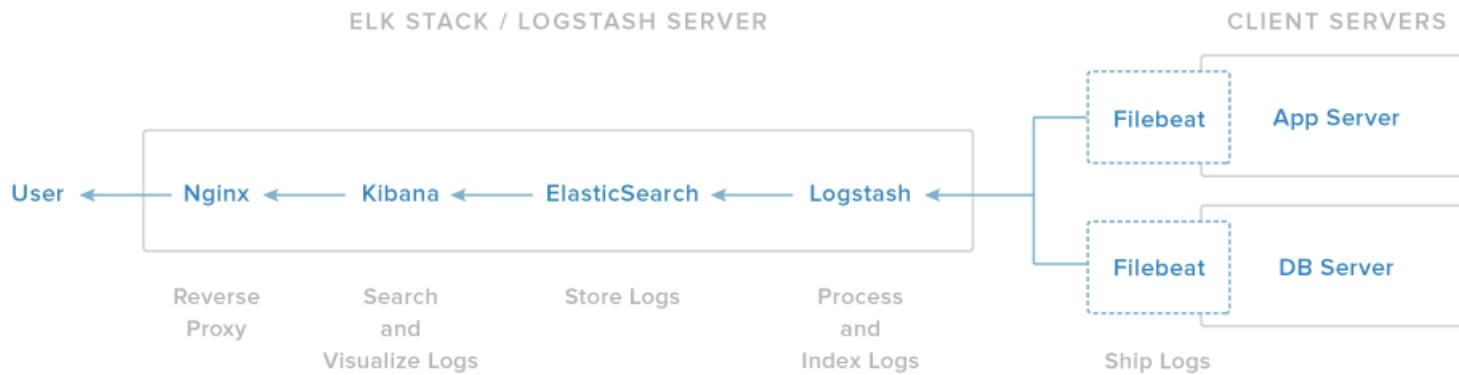


Figure source:  
<https://databricks.com/spark/about>

# 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

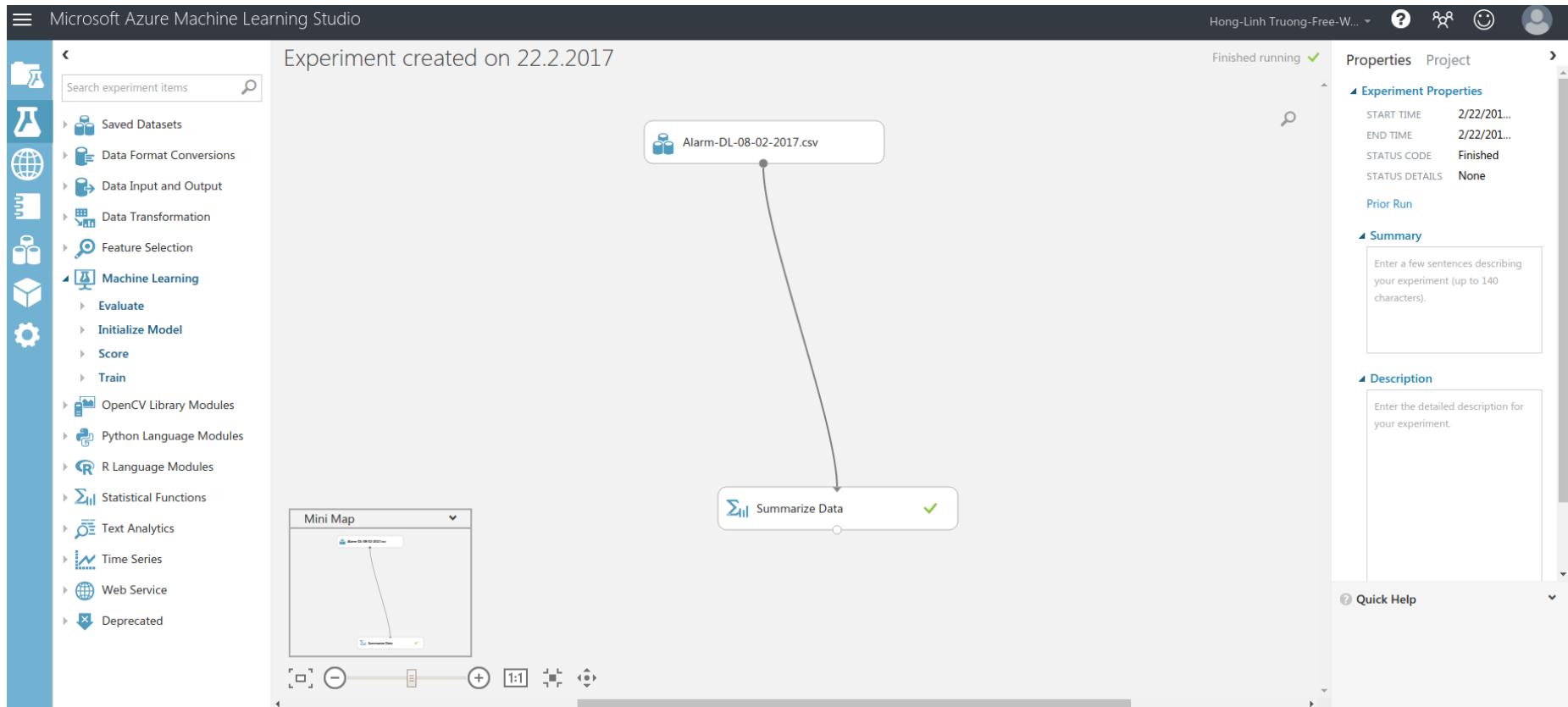
- Time-Series Data Visualization

## Kapacitor

- Time-Series Data Processing



# Machine Learning Stack



# 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



# 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



# Human computation models – properties and issues

## Some properties

- Huge number of people
- Capabilities might not know in advance
- Unpredictable behavior
- Simple coordination models

## Some issues

- Reliability
- Quality control
- Reliability assurance
- Proactive, on-demand acquisition
- Incentive strategies
- Collectives



# Emerging Software-defined \*

- Software-defined concepts
  - To have better way to manage dynamic changes in computation, network and data
  - Capabilities to manage and control computation, data, and network features at runtime using software
  - Management and control are performed via open APIs
- Software-defined techniques
  - Software-defined networking, Software-defined storage, Software-defined services

- K IRKPATRICK , K. Software-defined networking. *Commun. ACM* 56, 9 (Sept. 2013), 16–19.
- L ANGO , J. Toward software-defined slas. *Commun. ACM* 57, 1 (Jan. 2014), 54–60.
- S UGIKI , A., AND K ATO , K. Elements and composition of software-defined data centers. In *Proceedings of the Posters and Demo Track* (New York, NY, USA, 2012), Middleware '12, ACM, pp. 3:1–3:2.



Discussion time:

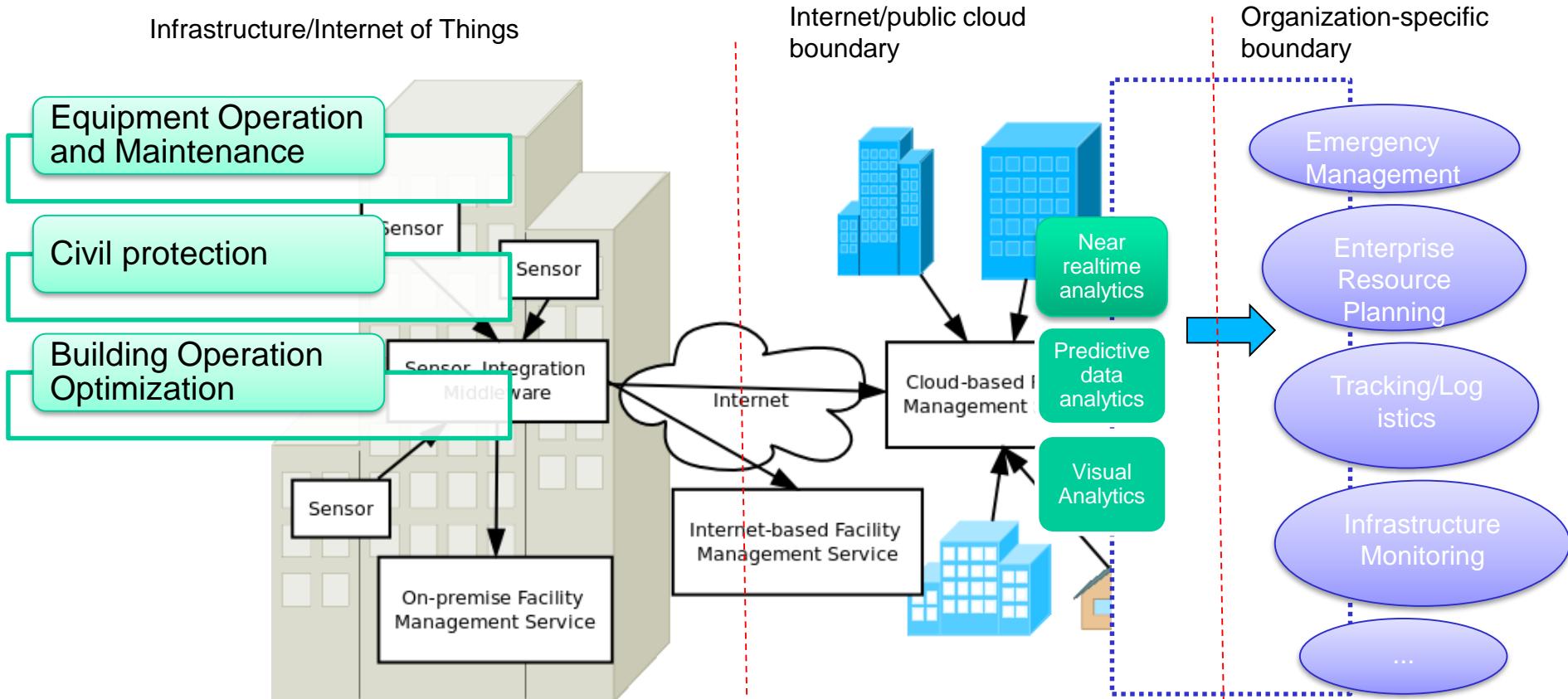
# **DO I NEED TO STUDY THEM ALL? WHY?**



# USE CASES/SCENARIOS



# Smart building management



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

Can we combine open government data  
with building monitoring data?



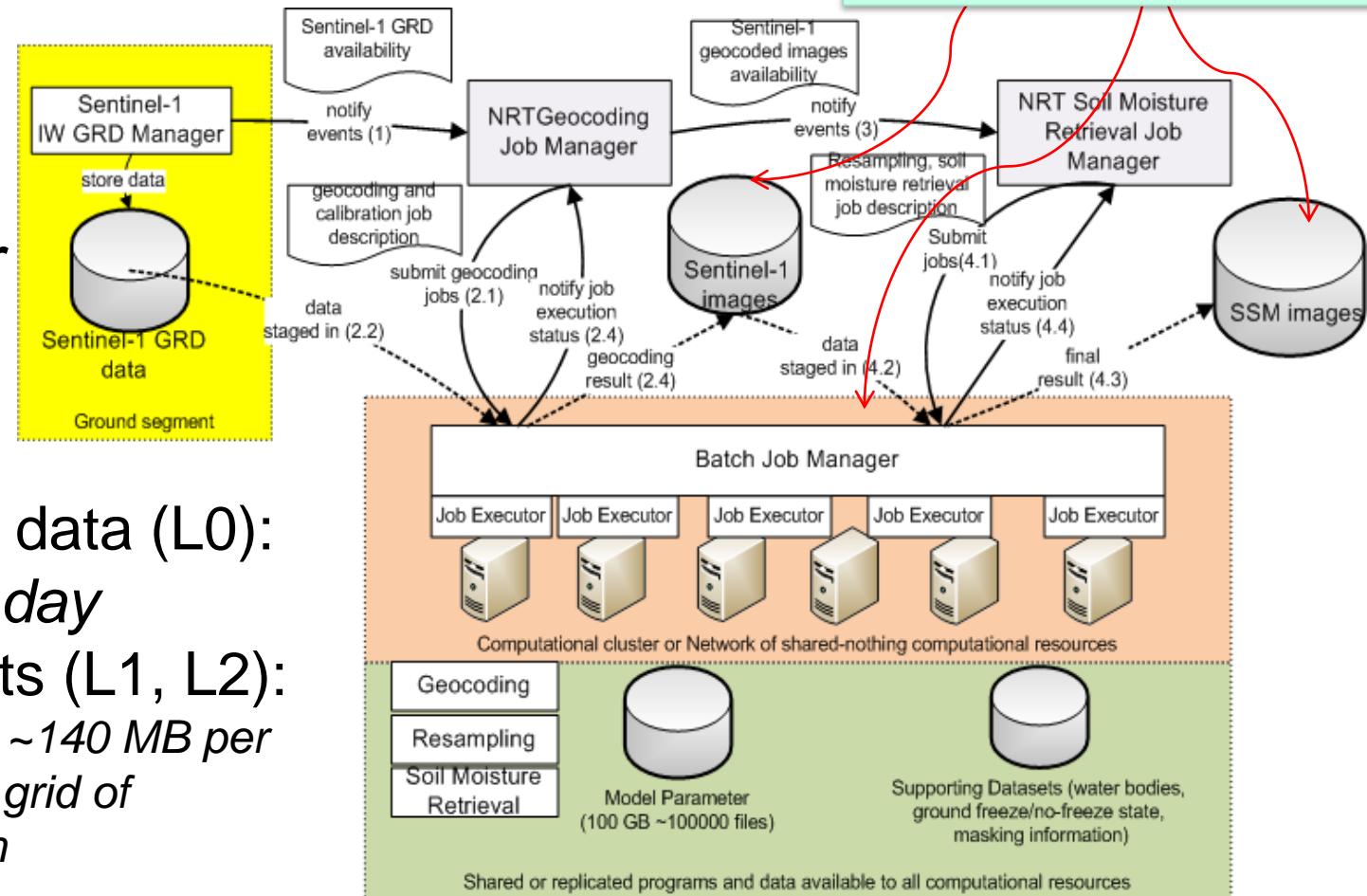
# Earth Observation

## Soil moisture analysis for Sentinel-1

A lot of input data (L0):  
 ~2.7 TB per day

A lot of results (L1, L2):  
 e.g., L1 has ~140 MB per  
 day for a grid of  
 1kmx1km

Can we combine them  
 with open government  
 data?



Michael Hornacek, Wolfgang Wagner, Daniel Sabel, Hong-Linh Truong, Paul Snoeij, Thomas Hahmann, Erhard Diedrich, Marcela Doubkova,  
**Potential for High Resolution Systematic Global Surface Soil Moisture Retrieval Via Change Detection Using Sentinel-1**, IEEE  
 Journal of Selected Topics in Applied Earth Observations and Remote Sensing, April, 2012



# Open data and social data

Source: <http://www.udata-api.org/>

<dataset organisation="IMF" category="International Financial statistics" name="BOP: capital account credit, US\$"/>  
<dataset organisation="IMF" category="International Financial statistics" name="BOP: capital account debit, US\$"/>  
<dataset organisation="IMF" category="International Financial statistics" name="BOP: current transfers credit, US\$"/>  
<dataset organisation="IMF" category="International Financial statistics" name="BOP: current transfers debit, US\$"/>  
<dataset organisation="IMF" category="International Financial statistics" name="BOP: exports of goods, f.o.b., US\$"/>  
<dataset organisation="IMF" category="International Financial statistics" name="BOP: financial account, US\$"/>  
<dataset organisation="IMF" category="International Financial statistics" name="BOP: imports of goods, f.o.b., US\$"/>  
<dataset organisation="IMF" category="International Financial statistics" name="BOP: income credit, US\$"/>  
<dataset organisation="IMF" category="International Financial statistics" name="BOP: income debit, US\$"/>  
<dataset organisation="IMF" category="International Financial statistics" name="BOP: net errors and omissions, US\$"/>  
<dataset organisation="IMF" category="International Financial statistics" name="BOP: services credit, US\$"/>  
<dataset organisation="IMF" category="International Financial statistics" name="BOP: services debit, US\$"/>  
<dataset organisation="IMF" category="International Financial statistics" name="BOP: trade balance, goods and services, US\$"/>  
<dataset organisation="IMF" category="International Financial statistics" name="BOP: trade balance, goods, US\$"/>  
<dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Adolescent fertility rate"/>  
<dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Gross national income per capita (PPP international \$)"/>  
<dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Infant mortality rate (per 1,000 live births)"/>  
<dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Inhaler use (percent)"/>  
<dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Internet users (percent)"/>  
<dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Life expectancy at birth (years)"/>  
<dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Long-term poverty (percent) below the poverty line (percent living on less than US\$1.90 a day)"/>  
<dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Malaria incidence (per 1,000 population)"/>  
<dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Measles cases (per 1,000 population)"/>  
<dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Net enrolment ratio female (percent)"/>  
<dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Net enrolment ratio male (percent)"/>  
<dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Number of health facilities (thousands) total"/>  
<dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Population growth rate (percent)"/>  
<dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Proportion of ban areas (percent)"/>  
<dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Proportion of people below the poverty line (percent living on less than US\$1.90 a day)"/>

DATA.GOV  
EMPPOWERING PEOPLE

HOME DATA ▾ TOOLS COMMUNITY METRICS DIALOGUE GALLERY WHAT'S NEW [SIGN IN]

## INTERNATIONAL OPEN GOVERNMENT DATA CONFERENCE

Sponsored by:  
U.S. General Services Administration

Hosted by:  
U.S. Department of Commerce

November 15 - 17, 2010  
Washington, DC

[VIEW MORE ▶](#)

### Most Popular Datasets

- 1 U.S. Overseas Loans and Grants (Greenbook)
- 2 Worldwide M1+ Earthquakes, Past 7 Days
- 3 Food and Drug Administration—Recalls
- 4 FDIC Failed Bank List
- 5 Clean Air Market Clean Air Interstate Rule..

### SEARCH OUR CATALOGS

[SEARCH ▶](#)

### APPS

With so much government data to work with, developers are creating a wide variety of applications, mashups, and visualizations. From crime statistics by neighborhood to the best towns to find a job to seeing the environmental health of your community—these applications arm citizens with the information they need to make decisions every day. Enjoy these highlights of the hundreds of applications available.

[READ MORE ▶](#)

### COMMUNITY

Data.gov is leading the way in democratizing public sector data and driving innovation. The data is being surfaced from many locations making the Government data stores available to researchers to perform their own analysis. Developers are finding good uses for the datasets, providing interesting and useful applications that allow for new views and public analysis. This is a work in progress, but this movement is spreading to cities, states, and other countries. After just one year a community is born around open government data.

**Just look at the numbers:**

- 7 Other nations establishing open data
- 16 States now offering data sites
- 9 Cities in America with open data
- 236 New applications from Data.gov datasets
- 253 Data contacts in Federal Agencies
- 305,709 Datasets available on Data.gov

[READ MORE ▶](#)

### SEMANTIC WEB

As the Web of linked documents evolves to include the Web of linked data, we're working to maximize the potential of Semantic Web technologies to realize the promise of Linked Open Government Data.

Thanks to our collaboration with the **Tetherless World Constellation at the Rensselaer Polytechnic Institute**, Data.gov is now hosting one of the largest open collections of RDF datasets in the world! Check out some of their semantic mashups [we're featuring](#) and read our blog entry to learn more about where we are, where we're going, and why we think this platform will add tremendous value to democratized data.

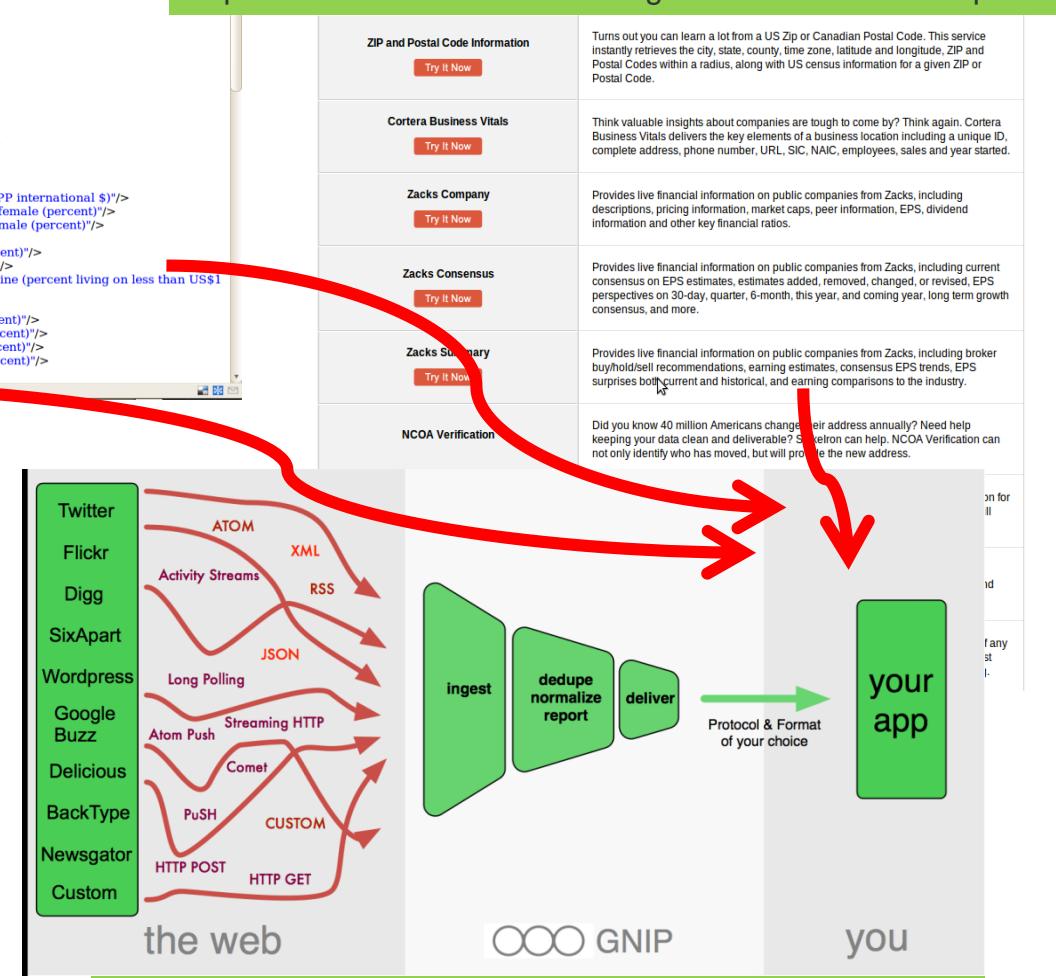
[READ MORE ▶](#)

Twitter  
Flickr  
Digg  
SixApart  
Wordpress  
Google Buzz  
Delicious  
BackType  
Newsgator  
Custom

Activity Stream  
Long Poll  
Atom Push  
Push  
HTTP POST

the v

Source:  
<http://www.strikeiron.com/Catalog/StrikeIronServices.aspx>



Source: <http://docs.gnip.com/w/page/23722723/Introduction-to-Gnip>

# Video analytics + business applications/public security

## *Use Case 3: Video Analytics*

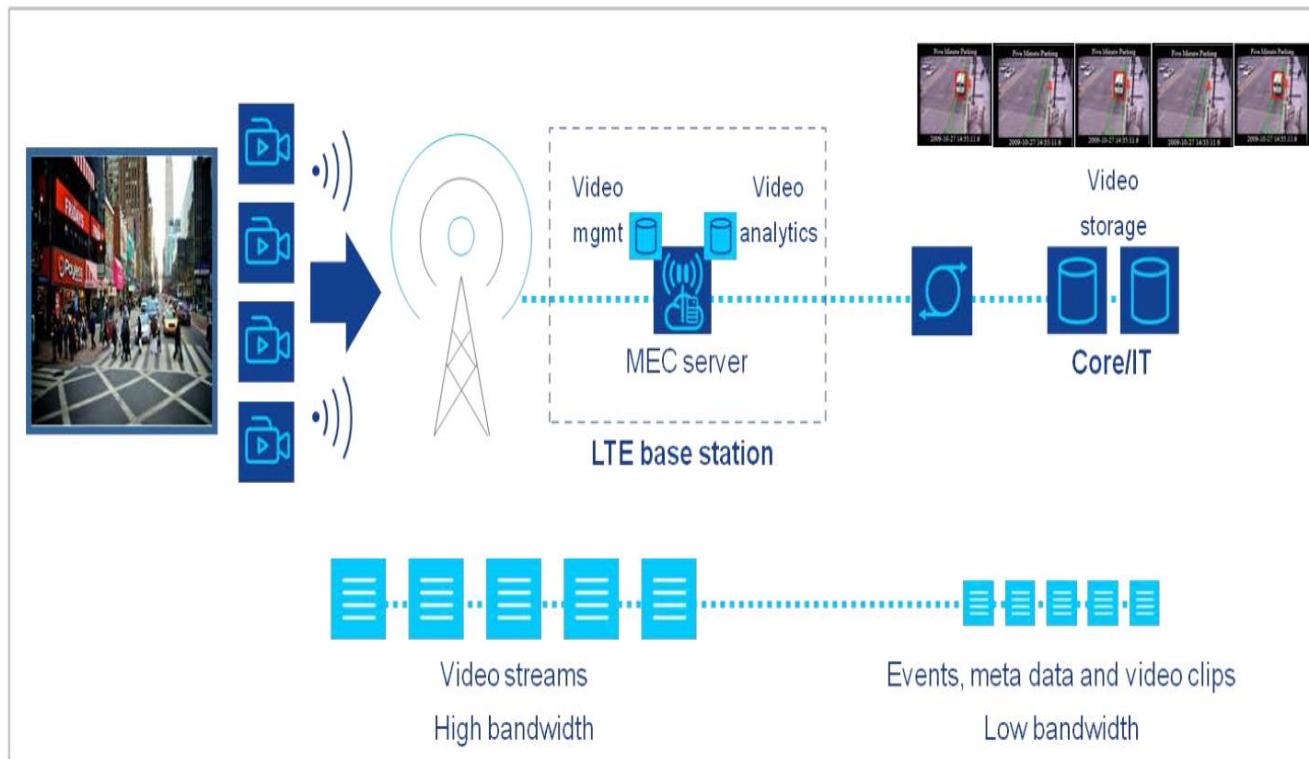


Figure source:

[https://portal.etsi.org/portals/0/tbpages/mec/docs/mobile-edge\\_computing\\_introductory\\_technical\\_white\\_paper\\_v1%2018-09-14.pdf](https://portal.etsi.org/portals/0/tbpages/mec/docs/mobile-edge_computing_introductory_technical_white_paper_v1%2018-09-14.pdf)



# Critical infrastructures/services for citizens and business

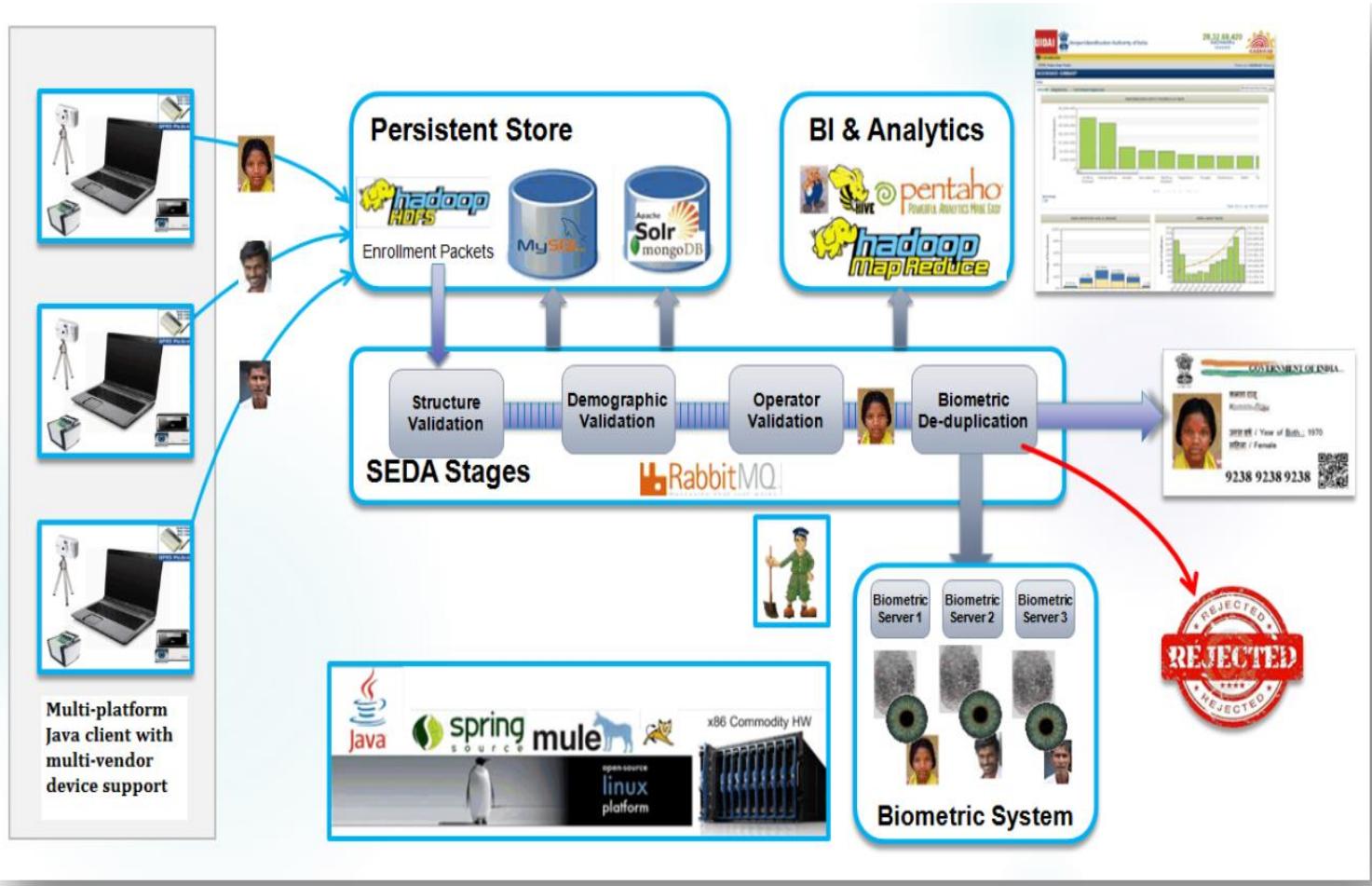
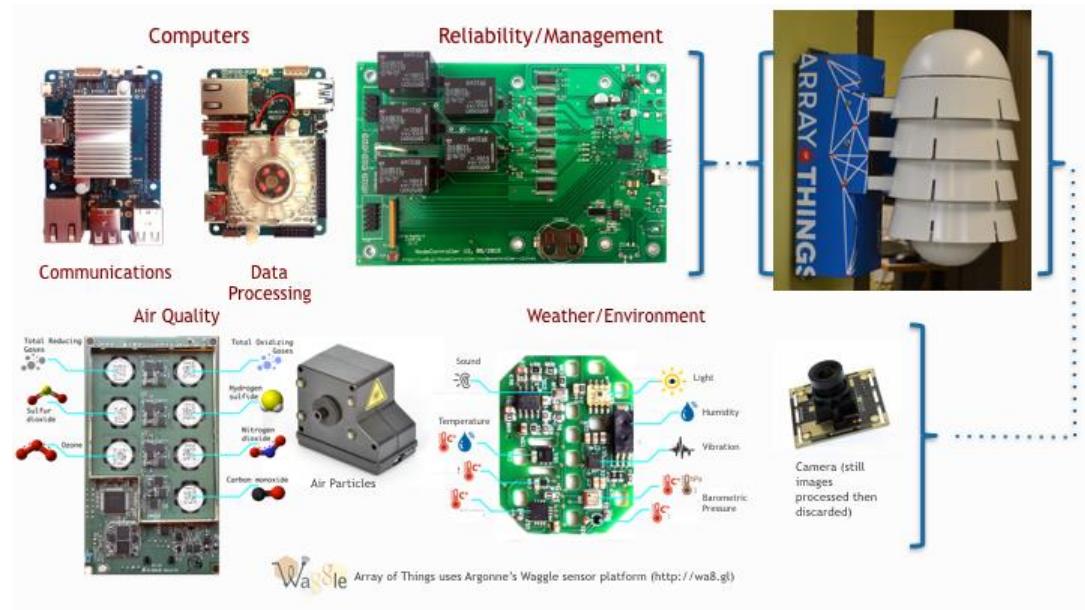
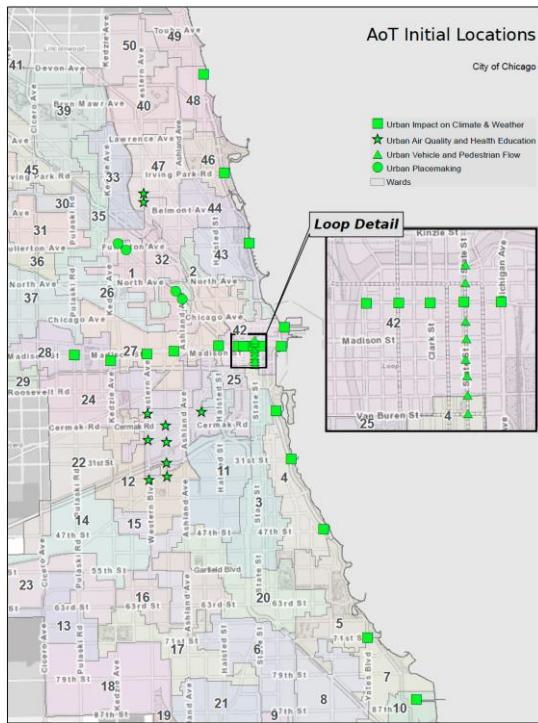


Figure source: [http://uidai.gov.in/images/AadhaarTechnologyArchitecture\\_March2014.pdf](http://uidai.gov.in/images/AadhaarTechnologyArchitecture_March2014.pdf)

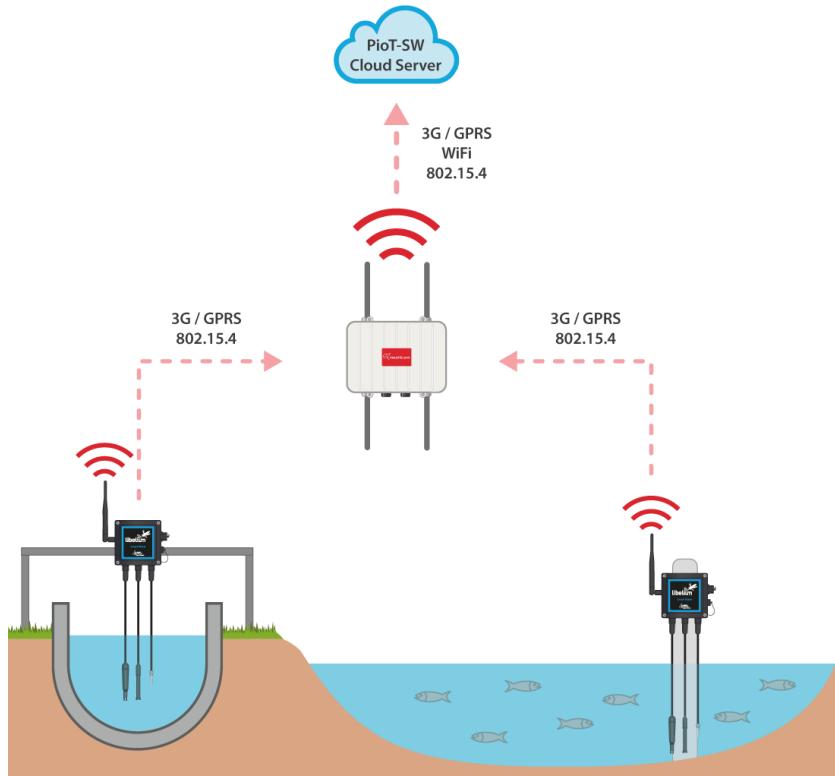
# IoT data in Cities



<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: Picture courtesy  
Future Position X, Sweden



# Drones for logistics



Figure 16: Urban First and Last Mile

Source: DHL Trend Report “Unmanned Aerial Vehicles”

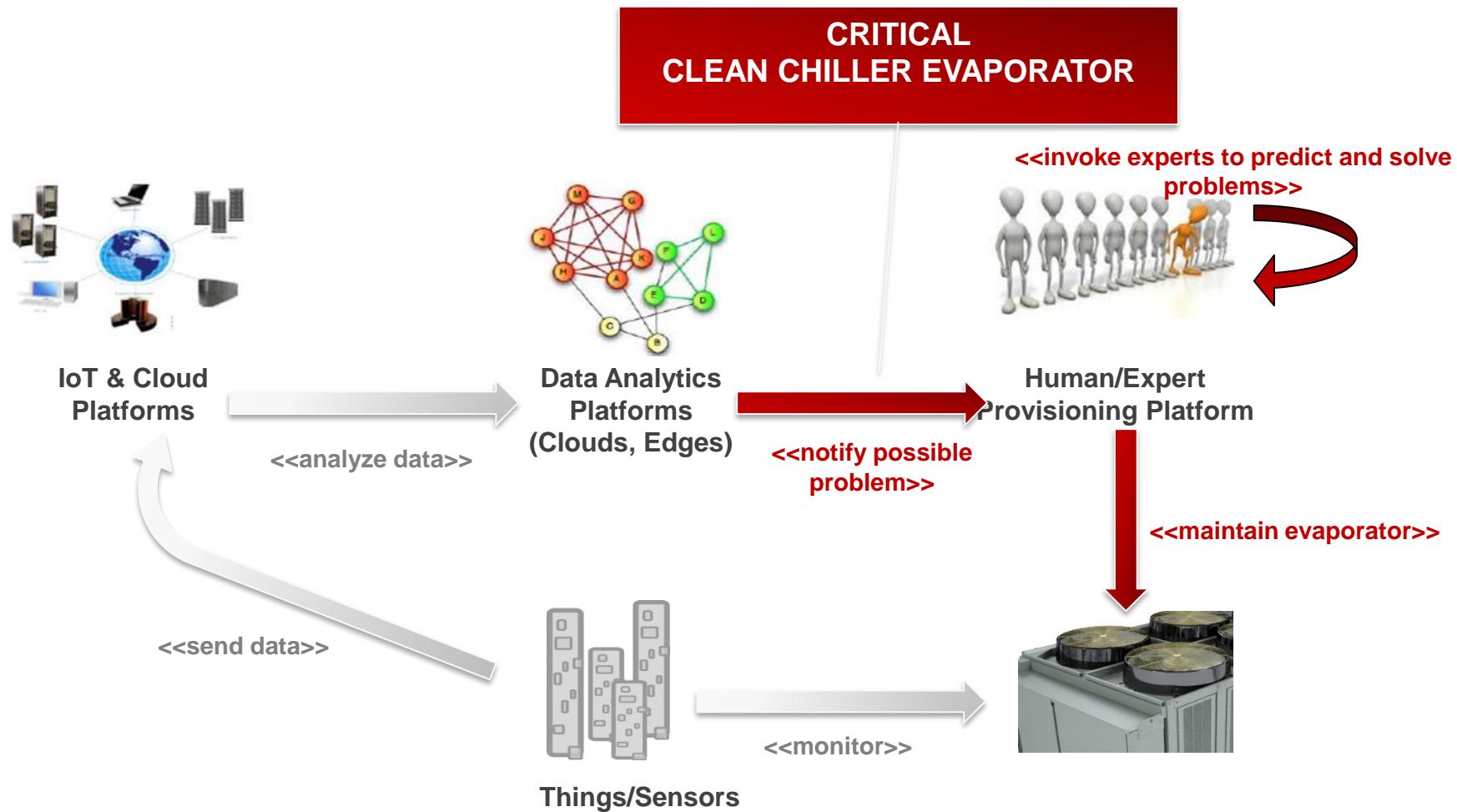
[http://www.dhl.com/content/dam/downloads/g0/about\\_us/logistics\\_insights/dhl\\_trend\\_report\\_uav.pdf](http://www.dhl.com/content/dam/downloads/g0/about_us/logistics_insights/dhl_trend_report_uav.pdf)



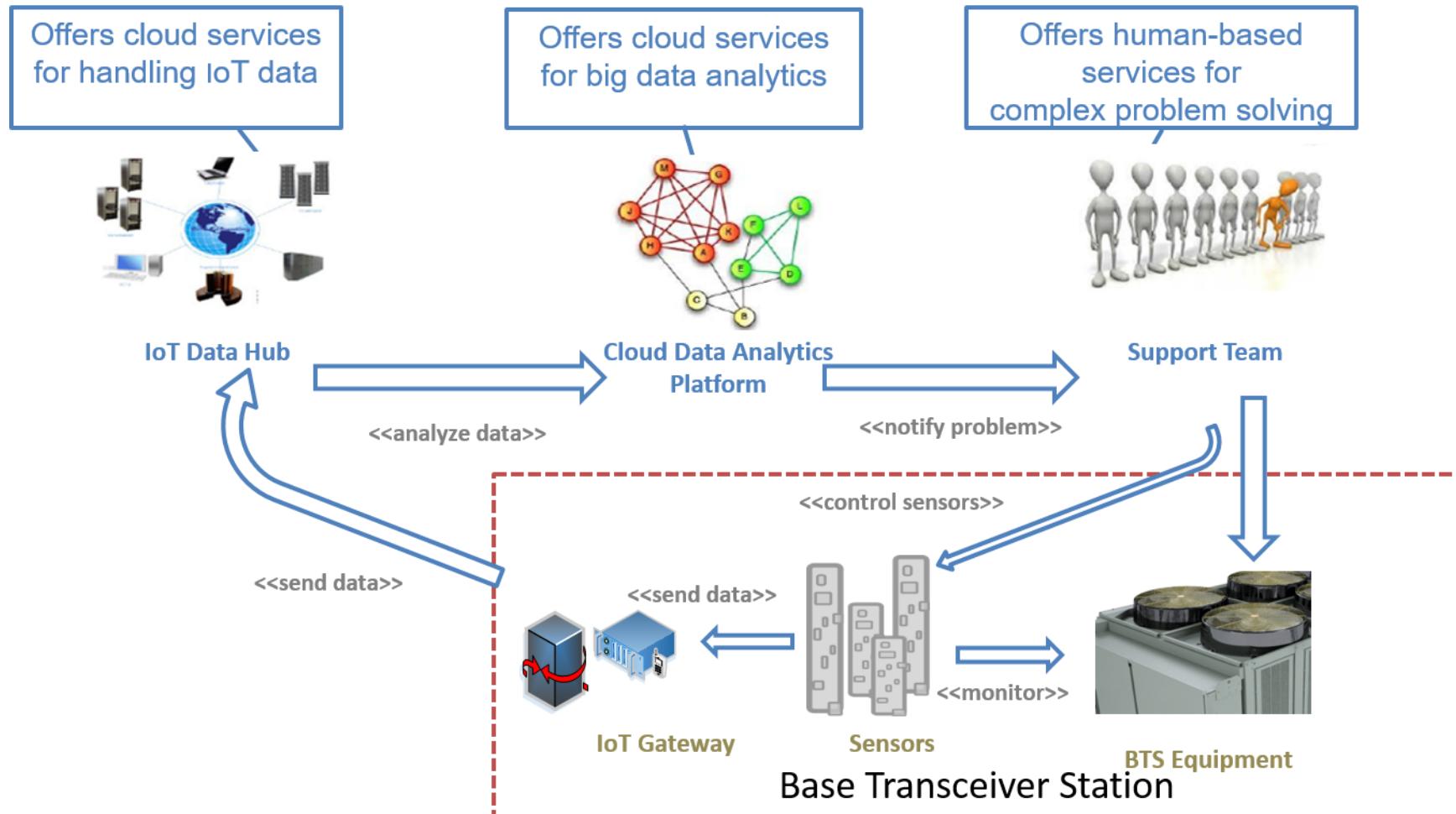
# CONVERGENCE OF MULTIPLE COMPUTING MODELS



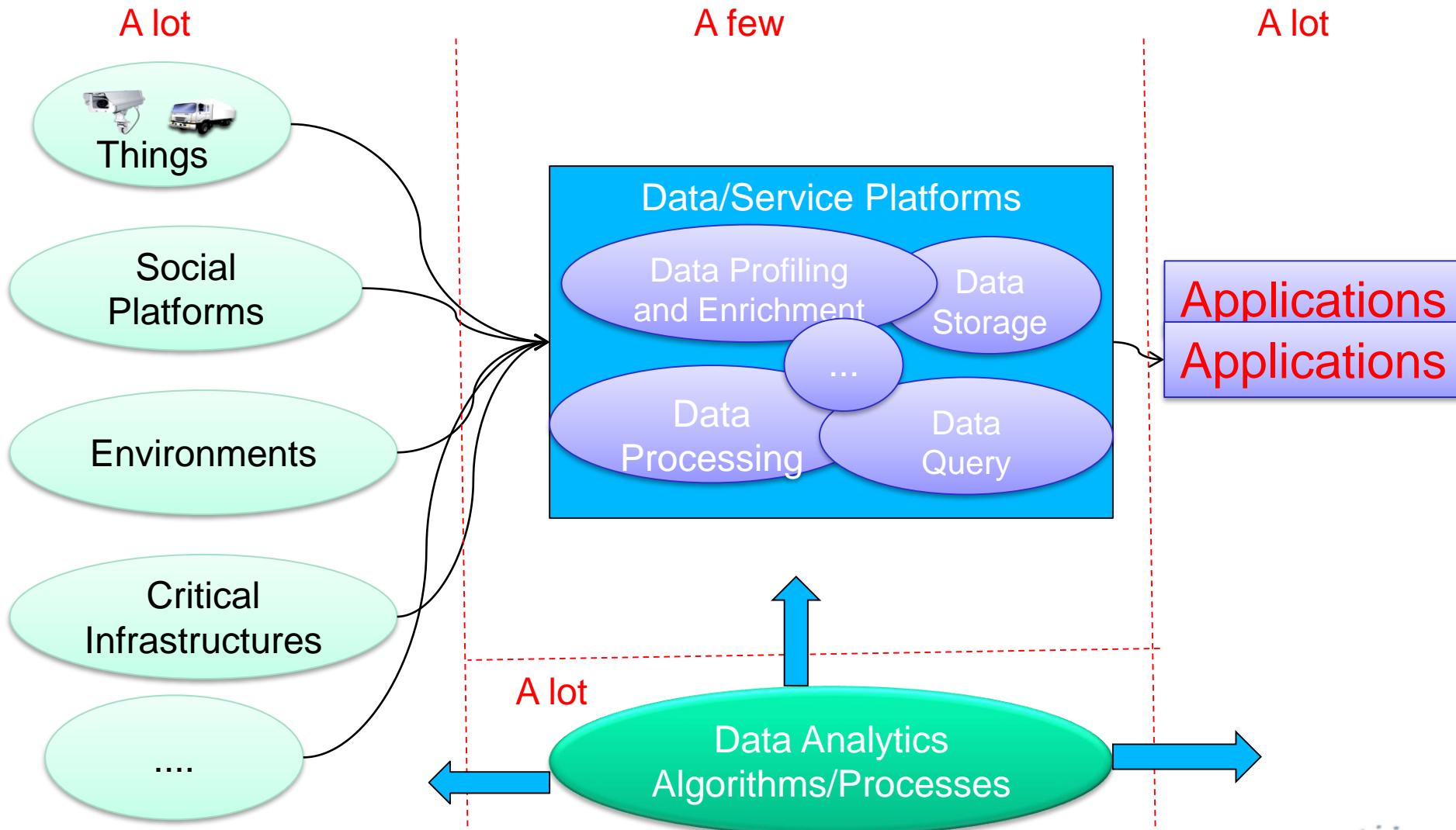
# Predictive Maintenance in Cities



# Predictive Maintenance in Telcos



# Cloud-based Analytics



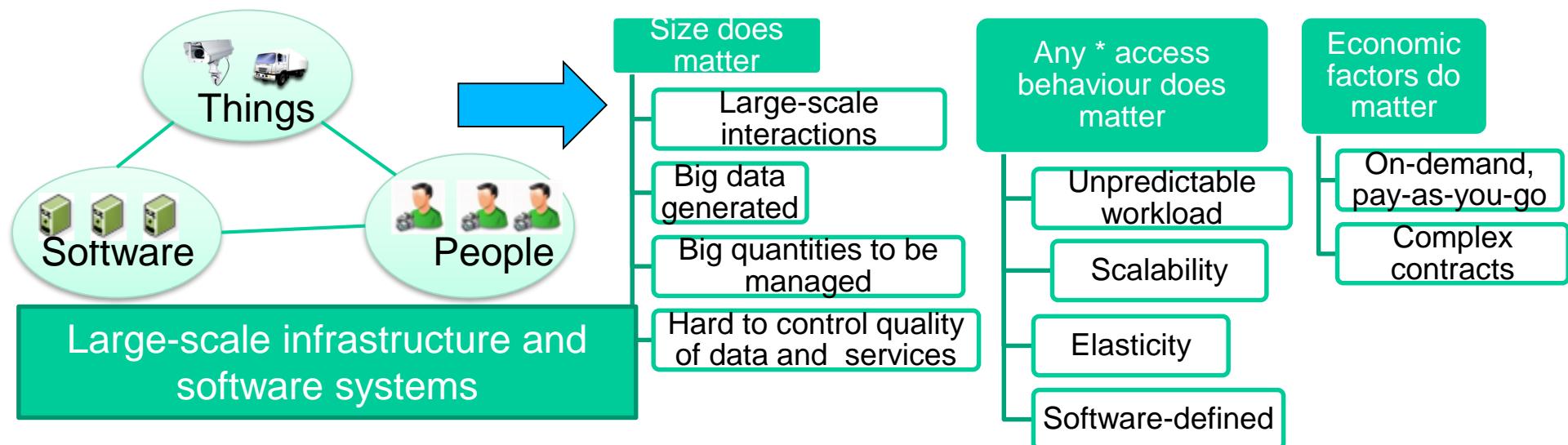
# BRING YOUR OWN EXPERIENCE: CLOUD-BASED ANALYTICS

See also <http://www.allthingsdistributed.com/2015/03/the-importance-of-cloud-based-analytics.html>

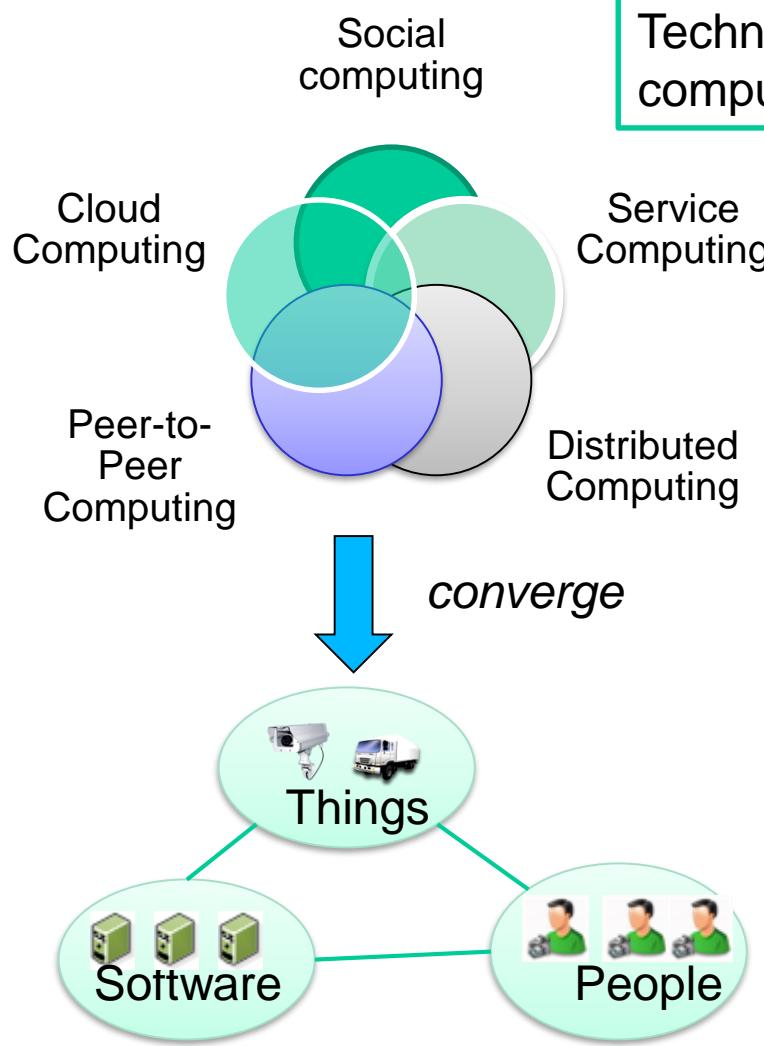


# Today's Computing Models

- 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

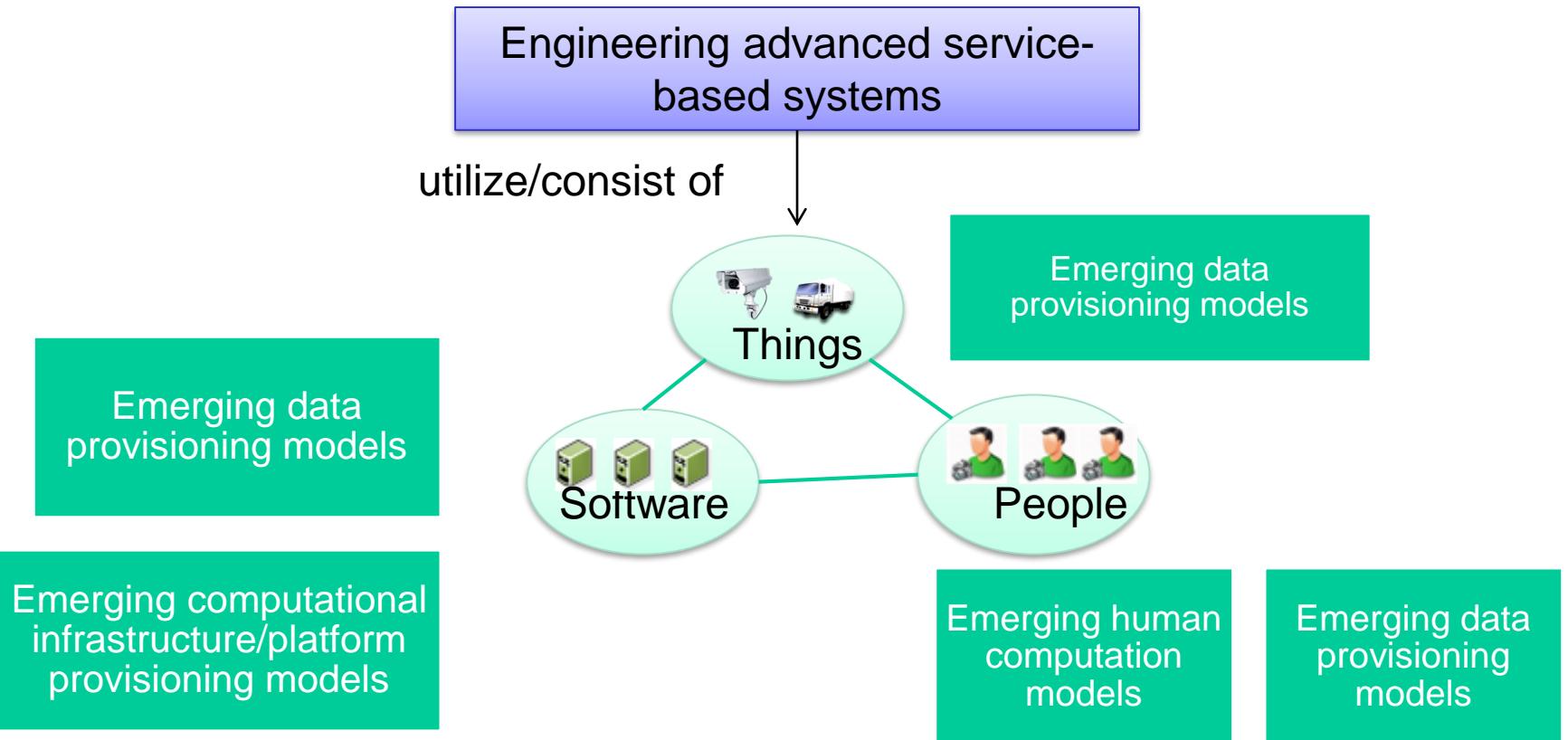


*introduces*

Emerging forms of computing models, systems and applications

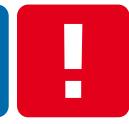


# Summary of emerging models wrt advanced service-based systems



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





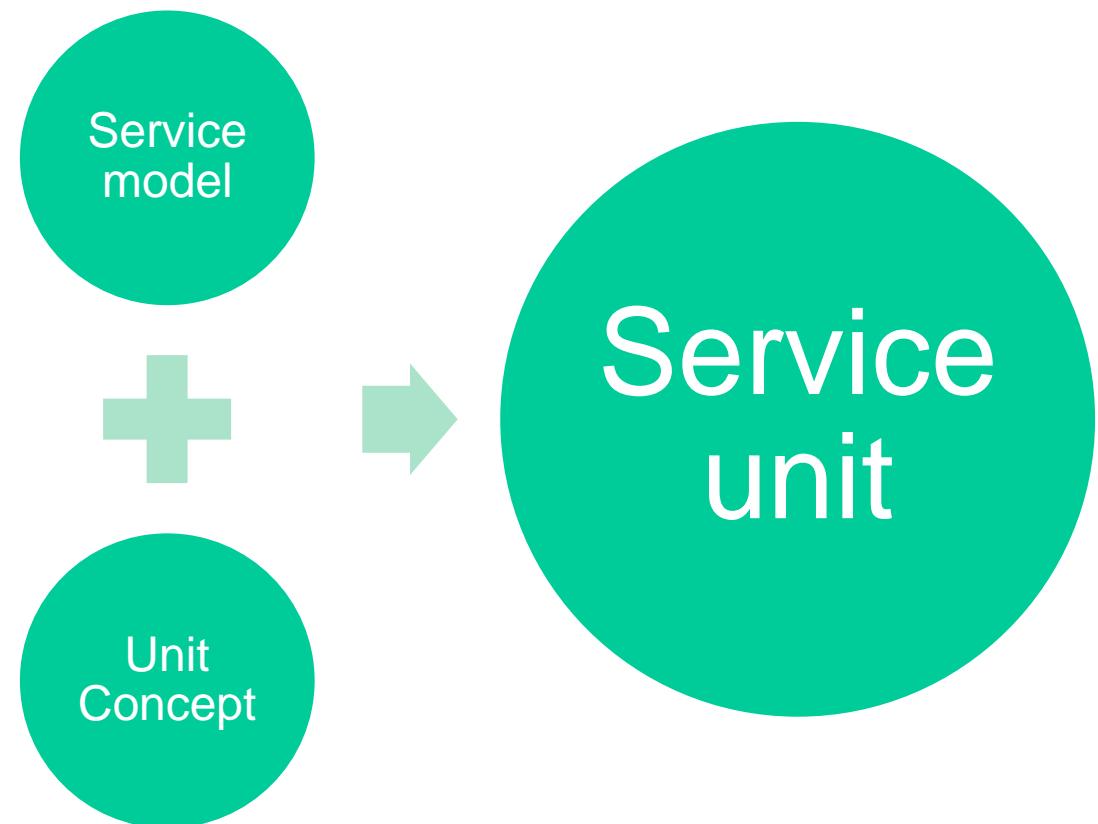
# ADVANCED SERVICES ENGINEERING'S FOCUS

# Single service/platform engineering(1)

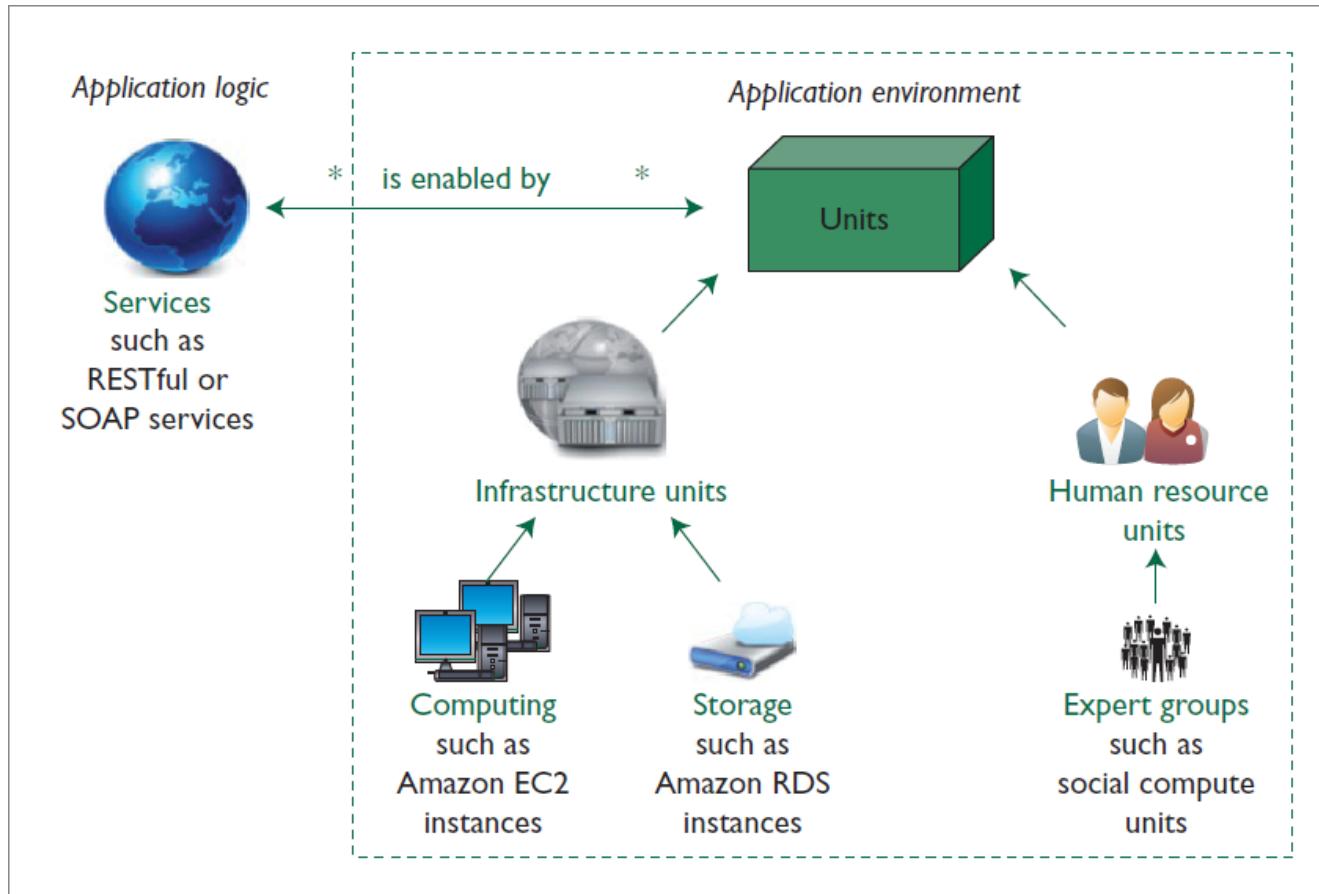
- The service model can be applied to things, people and software

Consumption,  
ownership,  
provisioning, price, etc.

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



# Single service/platform engineering – service units (2)



Source: Stefan Tai, Philipp Leitner, Schahram Dustdar: Design by Units: Abstractions for Human and Compute Resources for Elastic Systems. IEEE Internet Computing 16(4): 84-88 (2012)



# Single service/platform engineering – service unit provisioning

- Provisioning software under services
- Provisioning things under services
- Provisioning human under services
  - Crowd platforms of massive numbers of individuals
  - Collectives of humans

1. Mark Turner, David Budgen, and Pearl Brereton. 2003. **Turning Software into a Service**. *Computer* 36, 10 (October 2003), 38-44. DOI=10.1109/MC.2003.1236470 <http://dx.doi.org/10.1109/MC.2003.1236470>
2. Luigi Atzori, Antonio Iera, and Giacomo Morabito. 2010. **The Internet of Things: A survey**. *Comput. Netw.* 54, 15 (October 2010), 2787-2805. DOI=10.1016/j.comnet.2010.05.010 <http://dx.doi.org/10.1016/j.comnet.2010.05.010>
3. Dominique Guinard, Vlad Trifa, Stamatis Karnouskos, Patrik Spiess, Dominic Savio: **Interacting with the SOA-Based Internet of Things: Discovery, Query, Selection, and On-Demand Provisioning of Web Services**. IEEE T. Services Computing 3(3): 223-235 (2010)
4. Schahram Dustdar, Kamal Bhattacharya: **The Social Compute Unit**. IEEE Internet Computing 15(3): 64-69 (2011)
5. Hong-Linh Truong, Schahram Dustdar, Kamal Bhattacharya "**Programming Hybrid Services in the Cloud**", Springer-Verlag, 10th International Conference on Service-oriented Computing (ICSOC 2012), November 12-16, 2012, Shanghai, China



# Single service/platform engineering – examples (1)

- Service engineering with a single system/platform
  - Using Excel to access Azure datamarket places
  - Using Boto to access data in Amazon S3
  - Using Hadoop/Spark within a cluster to analyze data
  - Using workflows to process data (e.g., Trident/Taverna/Airflow)
  - Using StormMQ/Kafka to store messages

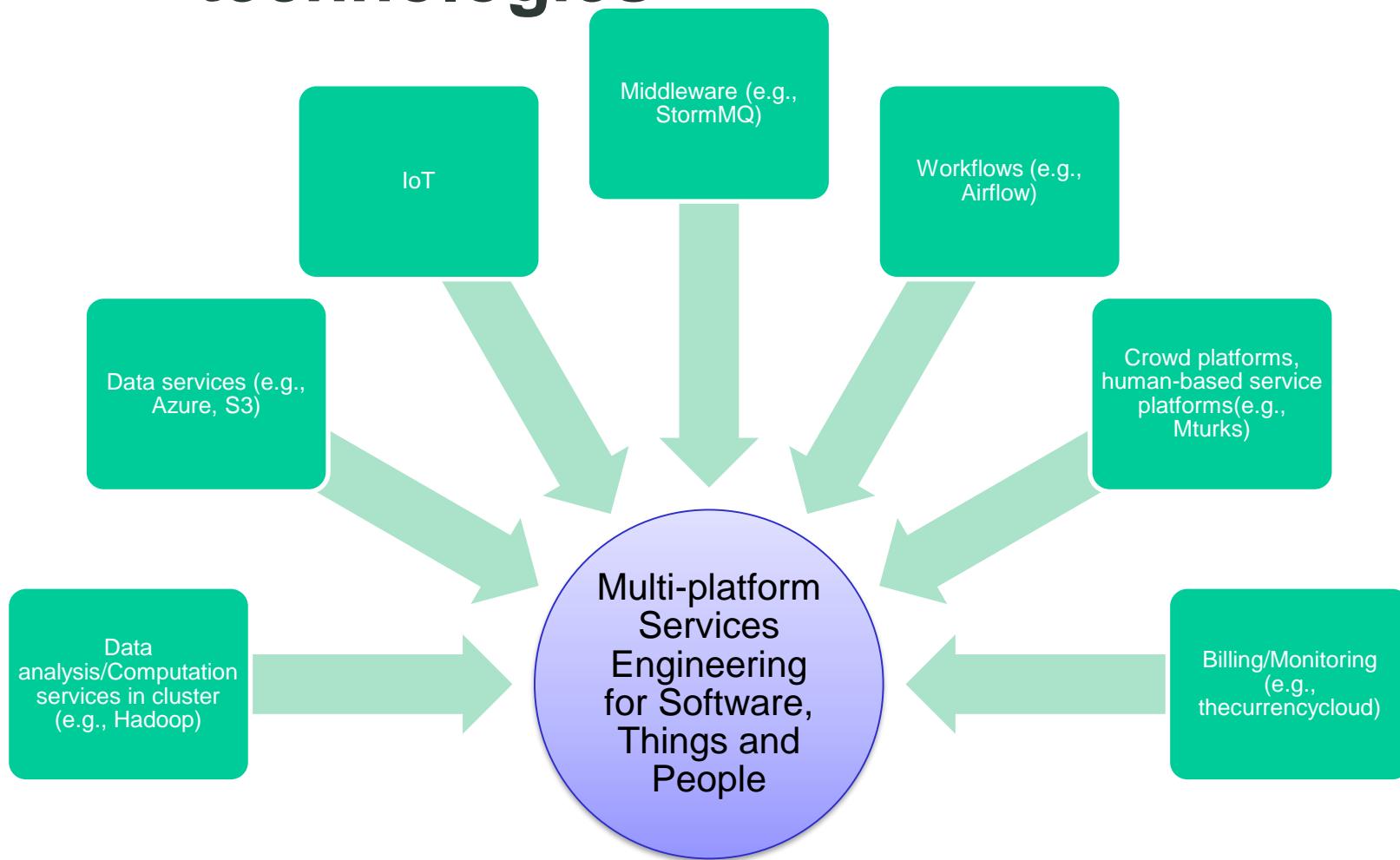


# Single service/platform engineering – examples (2)

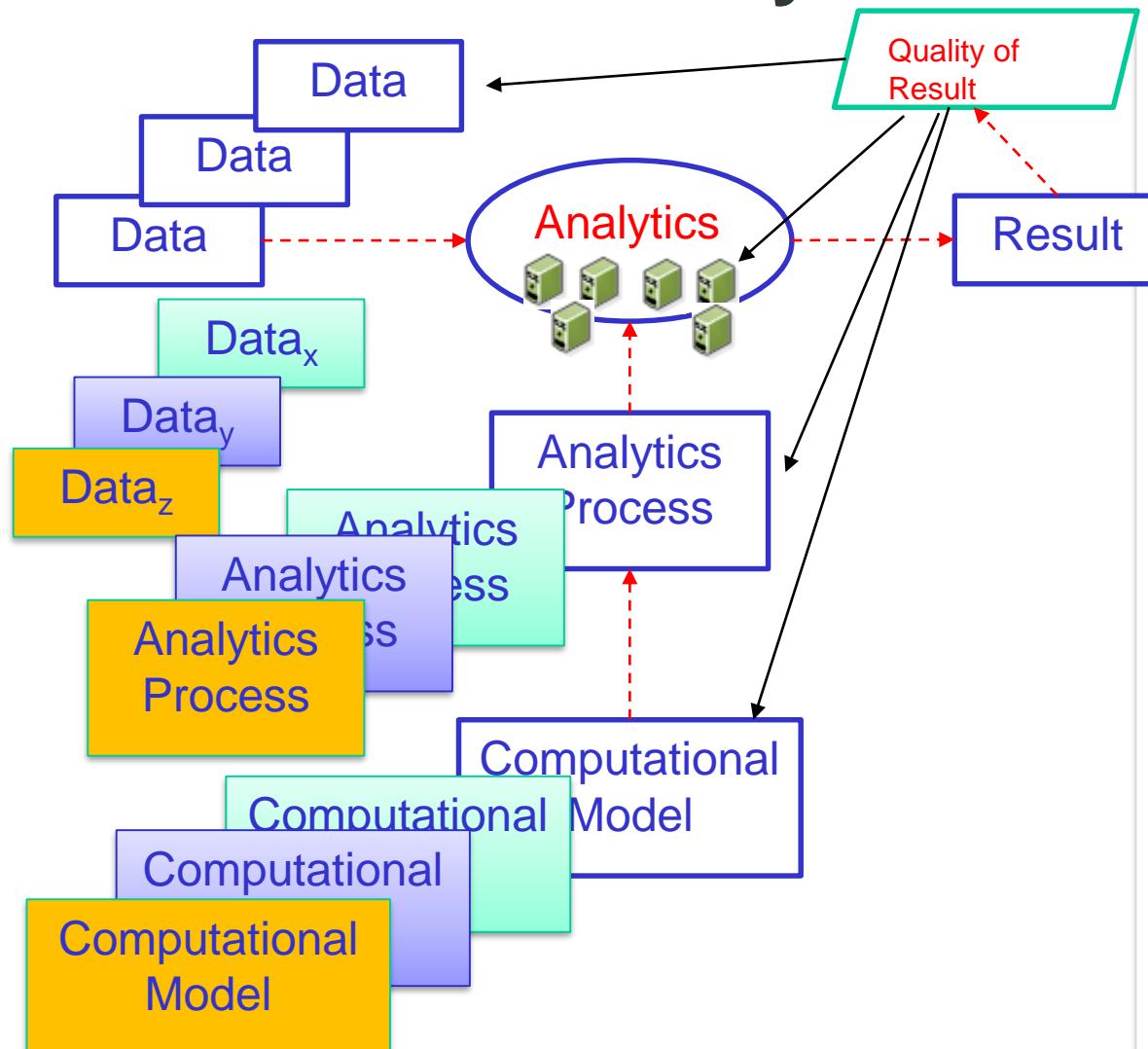
The screenshot illustrates a dual-platform engineering approach. On the left, a Microsoft Excel spreadsheet displays flight delay data for January 19, 2012, with columns including ArrDelay, Carrier, DayofMonth, DepDelay, Dest, FlightDate, Month, Origin, RowId, and Year. On the right, a Firefox browser window shows the Windows Azure Marketplace, specifically the search results for "US Air Carrier Flight Delays". The marketplace listing includes the title "US Air Carrier Flight Delays", the publisher "OakLeaf Systems", and a brief description of the dataset.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	ArrDelay	Carrier	DayofMonth	DepDelay	Dest	FlightDate	Month	Origin	RowId	Year						
2	0 US		19	0 PHX	2012-01-19T	1	LAS		1	2012						
3	42 US		19	64 DCA	2012-01-19T	1	PHX		2	2012						
4	3 US		19	0 DCA	2012-01-19T	1	PHX		3	2012						
5	0 US		19	23 KOA	2012-01-19T	1	PHX		4	2012						
6	0 US		19	0 PHX	2012-01-19T	1	BUR		5	2012						
7	0 US		19	2 DCA	2012-01-19T	1	PHX		6	2012						
8	0 US		19	0 PHX	2012-01-19T	1	LIH		7	2012						
9	10 US		19	0 LAS	2012-01-19T	1	DCA									
10	0 US		19	0 DCA	2012-01-19T	1	LAS									
11	0 US		19	1 MSP	2012-01-19T	1	PHX									
12	227 US		19	72 PHX	2012-01-19T	1	SEA									
13	29 US		19	0 PHX	2012-01-19T	1	MSP									
14	1 US		19	7 PHX	2012-01-19T	1	LAS									
15	0 US		19	0 MSP	2012-01-19T	1	PHX									
16	0 US		19	13 PHX	2012-01-19T	1	MSP									
17	0 US		19	3 MSP	2012-01-19T	1	PHX									
18	5 US		19	0 PHX	2012-01-19T	1	ONT									
19	0 US		19	0 MSP	2012-01-19T	1	PHX									
20	0 US		19	0 CLT	2012-01-19T	1	BOS									
21	0 US		19	0 BOS	2012-01-19T	1	CLT									
22	89 US		19	101 PHX	2012-01-19T	1	DEN									
23	0 US		19	0 SEA	2012-01-19T	1	PHX									
24	0 US		19	0 PHX	2012-01-19T	1	DFW									
25	3 US		19	8 SEA	2012-01-19T	1	PHX									
26	0 US		19	0 PHX	2012-01-19T	1	JFK									
27	0 US		19	0 SEA	2012-01-19T	1	PHX									
28	6 US		19	1 PHX	2012-01-19T	1	DCA									
29	0 US		19	0 SEA	2012-01-19T	1	PHX									

# Internet-scale multi-platform services engineering – required technologies



# Internet-scale service engineering – the elasticity



- **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.)



# Internet-scale service engineering - - big/near-real time data impact

- Which are data concerns that impact the data processing?
- How to use data concerns to optimize data analytics and service provisioning?
- How to use available data assets for advanced services in an elastic manner?
- What are the role of human-based servies in dealing with complex data analytics?



# Internet-scale 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



# Thanks for your attention

Hong-Linh Truong  
Distributed Systems Group, TU Wien  
[truong@dsg.tuwien.ac.at](mailto:truong@dsg.tuwien.ac.at)  
<http://dsg.tuwien.ac.at/staff/truong>  
@linhsolar