

Advanced Services Engineering, Summer 2018

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



- 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



- Some emerging models
 - IoT resources
 - (Big/realtime) 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



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

Emerging data provisioning models







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

ASE Summer 2018

available here.

Read more

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



Static + Realtime Open 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.

Source	Dataset	🕴 Count 🤻	Trend	
Chicago Police Department	Crimes - 2001 to present	38,766	600 0 View 2	i Data -
City of Chicago	Business Licenses	2,492	125 0 Yew 2	Data v
City of Chicago	Food Inspections	1,226		Data -



Telecommunication

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 \ast

Download data

Download data

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 »

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 »

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 $\ensuremath{\mathsf{w}}$

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 »



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





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 >



OpenSignal - The trusted, independent authority on Mobile Networks

COMPARE MOBILE NETWORKS NEAR YOU







And trend in monetizing data



Source: https://databrokerdao.com/

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

in EU call for proposals

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

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

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.







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

Hortonworks Data Platform



Hortonworks

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



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



Figure source: https://databricks.co m/spark/about



Building using elastic components: Elasticsearch, Elasticsearch Hadoop, Kibana, and Logstash https://www.elastic.co/



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



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

- <u>
 Time-Series Data Storage</u>
 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)

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



Discussion time:

DO I NEED TO STUDY THEM ALL? WHY?



USE CASES/SCENARIOS

ASE Summer 2018

32



Critical infrastructures/services for citizens and business



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

Smart building management







Figure source: https://eodatacube.eu/ ASE 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-gettingsmart-glasses/



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





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

IoT data in City-scale





https://arrayofthings.github.io/node.html



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

DECENTRALIZED NETWORK audur augur geth geth geth Laugur augur augur geth geth geth 1 1 1 P2P ETHEREUM WIRE PROTOCOL augur node augur-core USER'S prediction COMPUTER markets SQLite/PostareSQL DB geth ethereum blockchain HTTP Websocket RPC RPC BROWSER calls ethrpc augur.js calls networking augur-ui javascript API JSON RPC accounts data web front-end in-browser transactions data react/flux UI comments USER **MIDDLEWARE** calls data data calls keythereum augur-ABI key generation serialization encryption formatting import/export . .





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



Predictive Maintenance in Smart Buildings

CRITICAL CLEAN CHILLER EVAPORATOR



Predictive Maintenance in Telcos





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



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



Single service/platform engineering – service unit provisioning

- Provisioning software, things and human capabilities under services
- E.gg., video analytics, machine learning service, Spark cluster, BigQuery, HDFS, Ethereum, etc.

- 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, Domnic 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)



Internet-scale multi-platform services engineering – required technologies



Service engineering – the elasticity



More data → more computational resources (e.g. more VMs)

More types of data \rightarrow more computational models \rightarrow 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.)

A 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



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?

Advanced service engineering --Steps





- 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

ASE Summer 2018 59



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