Quantifying and Detecting Incidents in IoT Big Data Analytics

Hong-Linh Truong

Faculty of Informatics, TU Wien, Austria
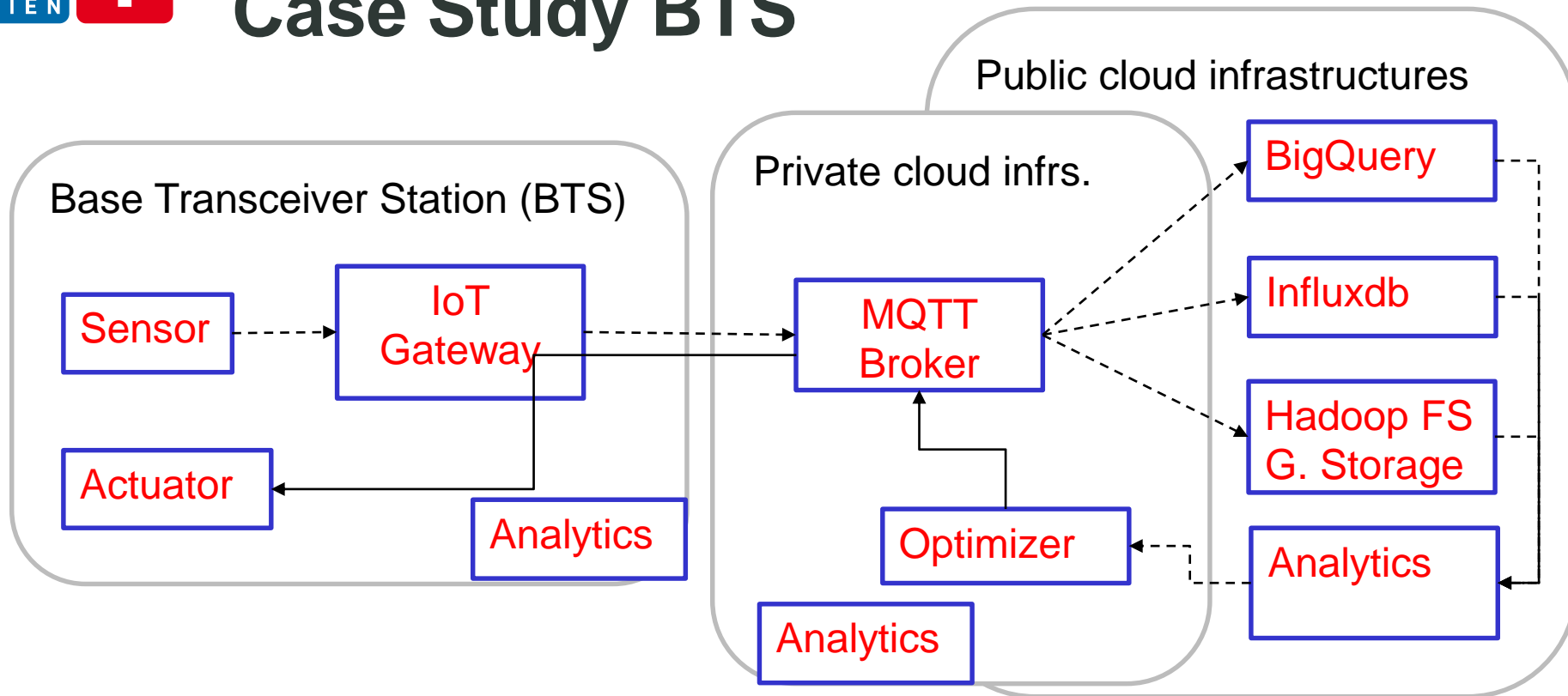
hong-linh.truong@tuwien.ac.at

http://rdsea.github.io

Note: Ongoing work

Case Study BTS

- Large-scale systems (1K+ BTS)
- Flexible back-end clouds
  - Generic enough for other applications (e.g., in smart agriculture)
- With bad infrastructures for IoT and connectivity

Challenges

The ultimate goal of the (domain) data scientist is to meet

Quality of Analytics (QoA)

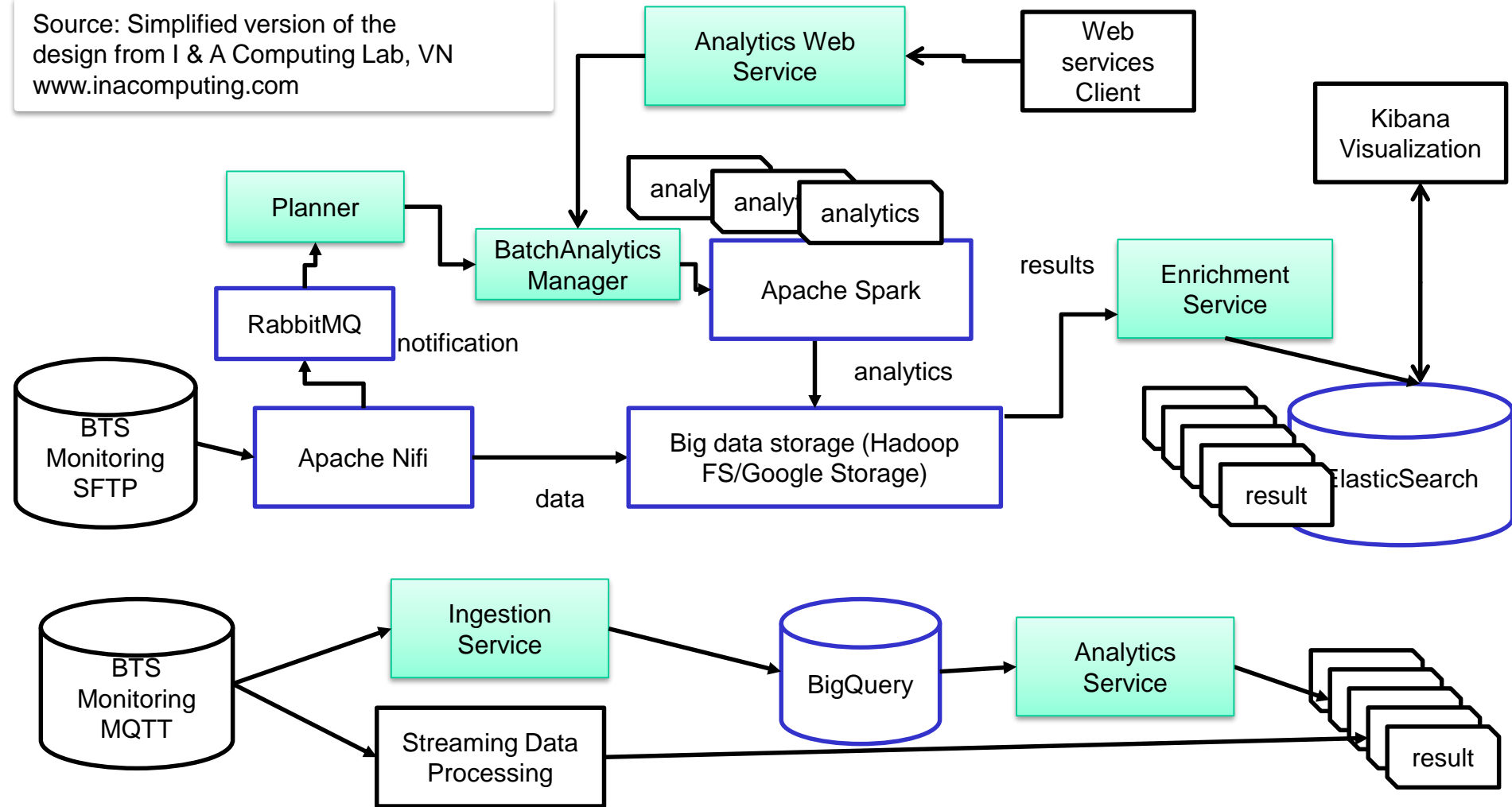QoA: cost, performance (response time), quality of data (up-to-date ness, accuracy)

(Remember Christoph Quix's talk about quality)

But there are many interactions that might cause incidents

Hong-Linh Truong , Aitor Murguzur, Erica Yang, Challenges in Enabling Quality of Analytics in the Cloud, ACM JDIQ Challenge paper, 2017.
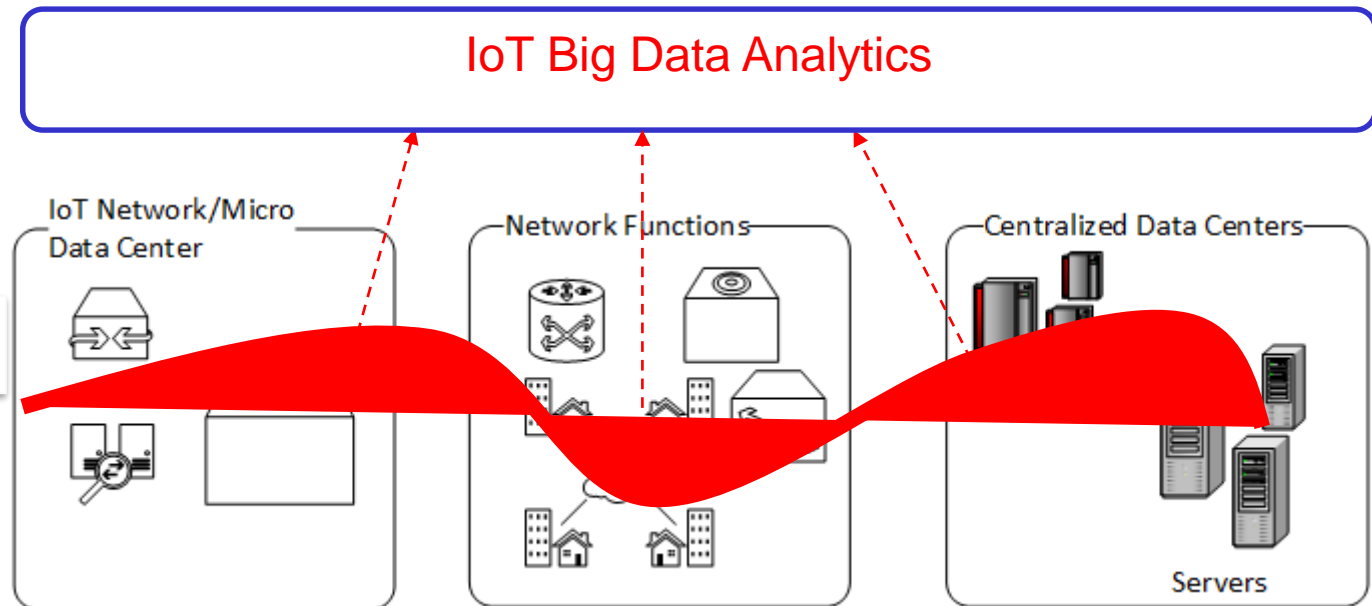
Problem 1: the complexity of software stacks and subsystems

Source: Simplified version of the design from I & A Computing Lab, VN www.inacomputing.com

Analytics Web Service

Web services Client

Kibana Visualization

Planner

analy / analy / analytics

BatchAnalytics Manager

Apache Spark

results

Enrichment Service

RabbitMQ

notification

analytics

BTS Monitoring SFTP

Apache Nifi

data

Big data storage (Hadoop FS/Google Storage)

result

ElasticSearch

BTS Monitoring MQTT

Ingestion Service
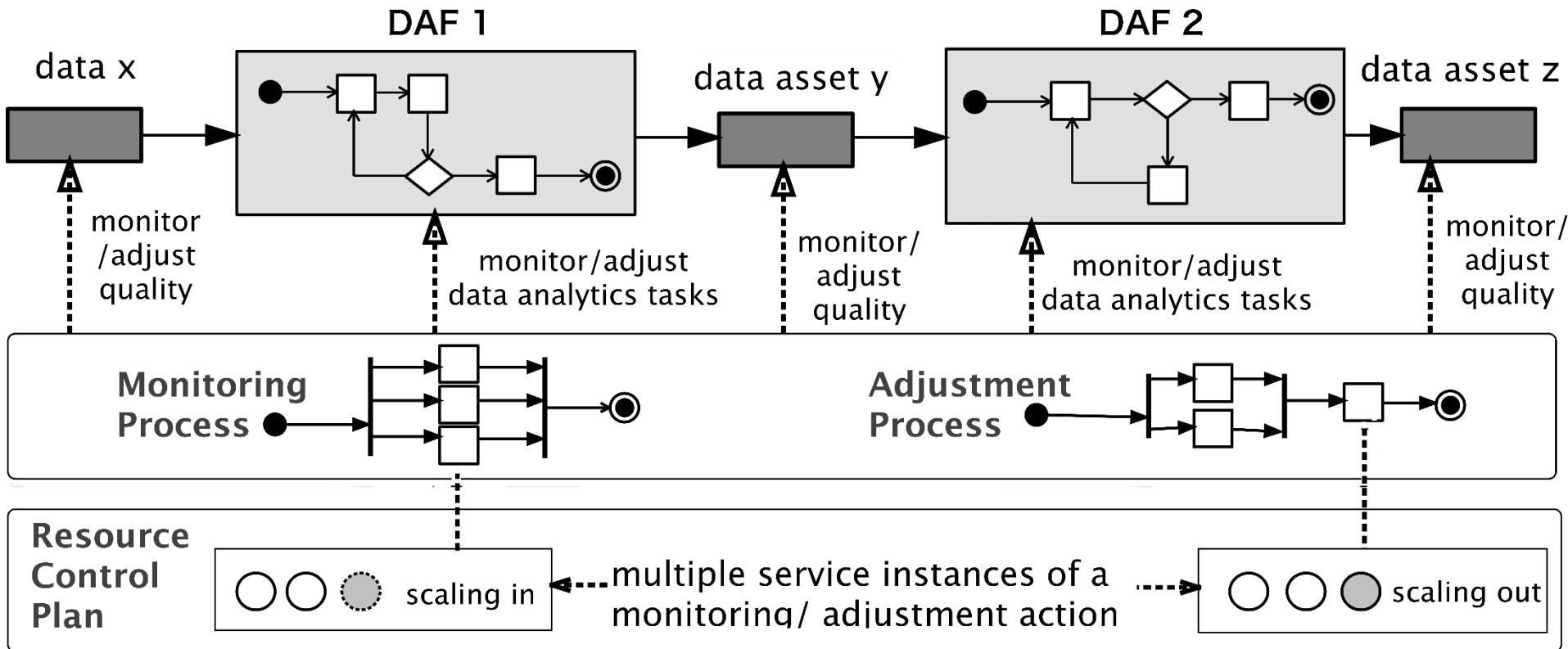
BigQuery

Analytics Service

Streaming Data Processing

result

Porblem 2: Complexity of the underlying virtual computing and network infrastructures

Heavily based on virtual resources

- IoT, Network functions and Clouds
  - (Remember Manfred Hauswirth's talk yesterday about fog/edge computing and NFV/5G networks)

DAF 1

data x

monitor /adjust quality

monitor/adjust data analytics tasks

data asset y

monitor/ adjust quality

DAF 2

monitor/adjust data analytics tasks

data asset z

monitor/ adjust quality

Monitoring Process

Adjustment Process

Resource Control Plan

scaling in

multiple service instances of a monitoring/ adjustment action

scaling out

Tien-Dung Nguyen, Hong Linh Truong, Georgiana Copil, Duc-Hung Le, Daniel Moldovan, Schahram Dustdar: On Developing and Operating of Data Elasticity Management Process. ICSOC 2015: 105-119
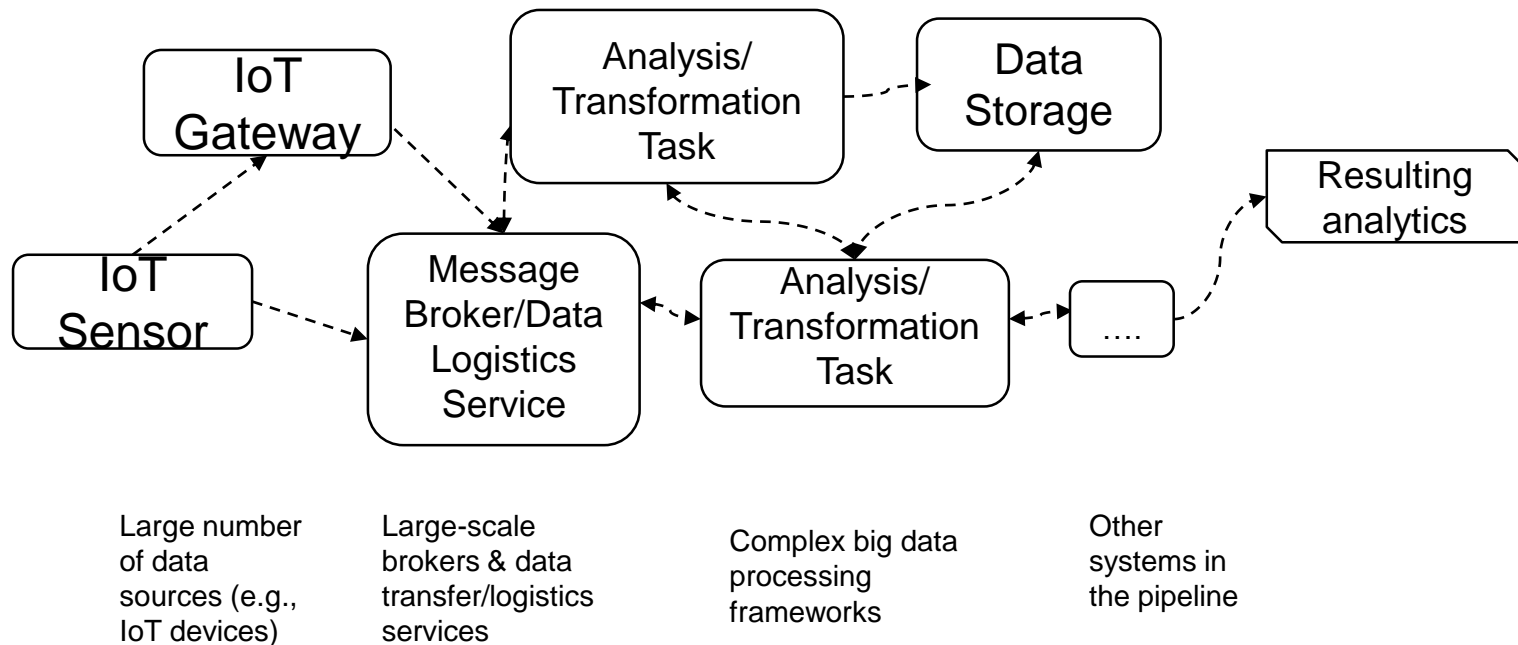
Our ideas for incident monitoring and analytics

- Classification of incidents:
  - to quantify incidents and identify possible data sources, monitoring techniques and analytics.

- Measurement/Instrumentation:
  - to provide mechanisms for measurement and data collection for incidents.

- Incident analytics:
  - to find out the root cause and dependencies of incidents.

Hong Linh Truong, Manfred Halper: Classifying Incidents in Cloud-based IoT Big Data Analytics, Working paper, 2017.

W3H: what, when, where and how for incidents

Too complex with many types of software. Can we have a simplified taxonomy for mapping incidents?

Large number of data sources (e.g., IoT devices)

Large-scale brokers & data transfer/logistics services

Complex big data processing frameworks

Other systems in the pipeline

Hong Linh Truong, Manfred Halper: Classifying Incidents in Cloud-based IoT Big Data Analytics, Working paper, 2017.

Monitoring and Analytics

Not just fast, distributed and cross layer monitoring

→ Hard to collect some incident related data for quality of data
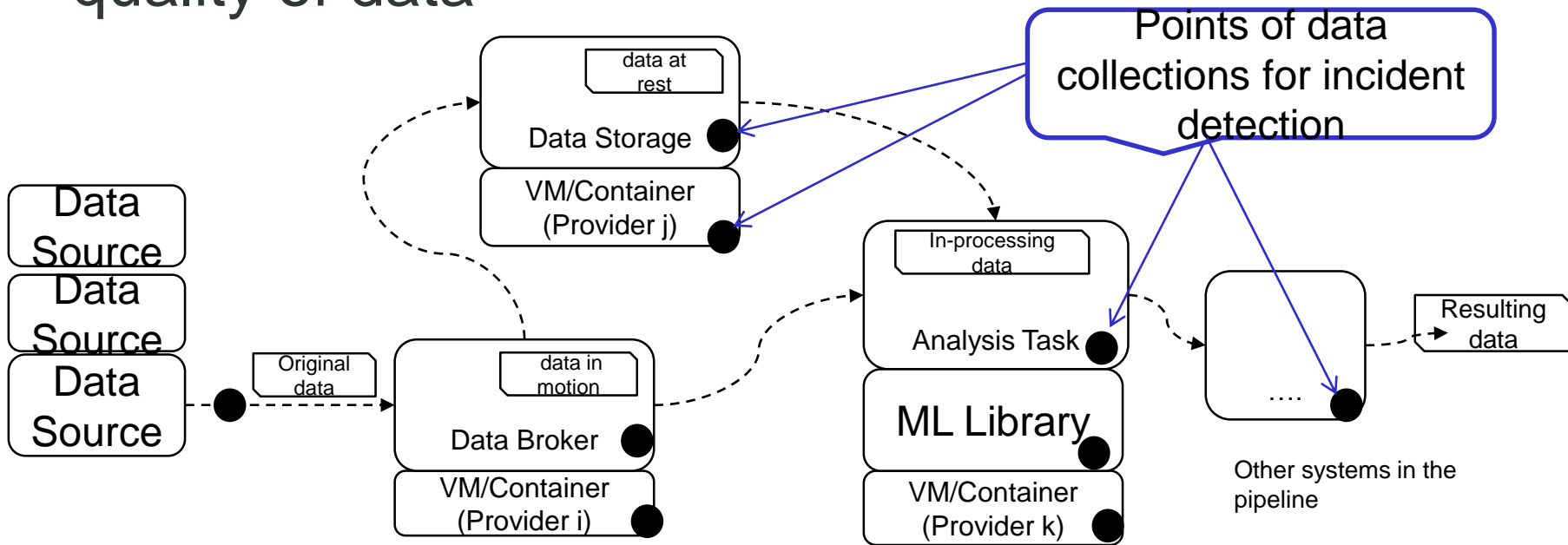
Analytics: will be based on big data principles with ML but dependency analysis is not trivial

Thanks for your attention!

Hong-Linh Truong

Faculty of Informatics
TU Wien

rdsea.github.io