SAMI 2021

Data engineering case-study in digitalized manufacturing

**István Pölöskei**

istvan.poeloeskei@adesso.eu

Introduction

- ► István Pölöskei – Lead AI Engineer (istvan.poeloeskei@adesso.eu)
  - > Natural language processing, deep learning, data pipelines, cloud-native
- ► adesso Hungary Kft.
  - > location of adesso SE
  - > Consulting, cloud-native, microservices, devops, AI, big data

Presentation summary

- ► Introduction of data flows
- ► Cloud and machine learning
- ► Data engineering in production
- ► Example case-study
- ► Integration of machines
- ► Cloud connection
- ► Processing plant data in the cloud
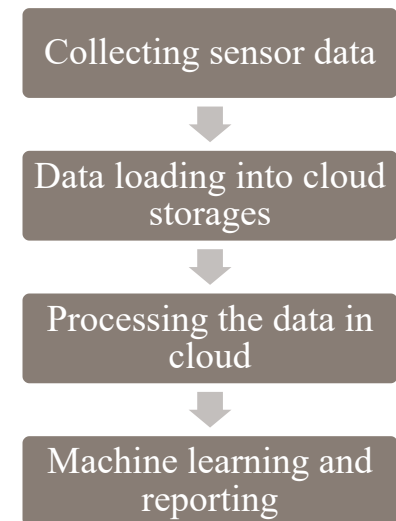- ► Providing analytics
- ► Evaluation

adesso

► IT solutions and sensors had already been established on the shop floor, accumulating data for reporting objectives.

► The digitalization in the manufacturing sector is strong; modern IoT technologies (Internet of Things) have already been utilized.

► The managed data-streams are unstructured.

► Processing measures are unavoidable for mining its information content.

► Cloud design gives an architecture for it.

► The integration is the core of the manufacturing use-case.

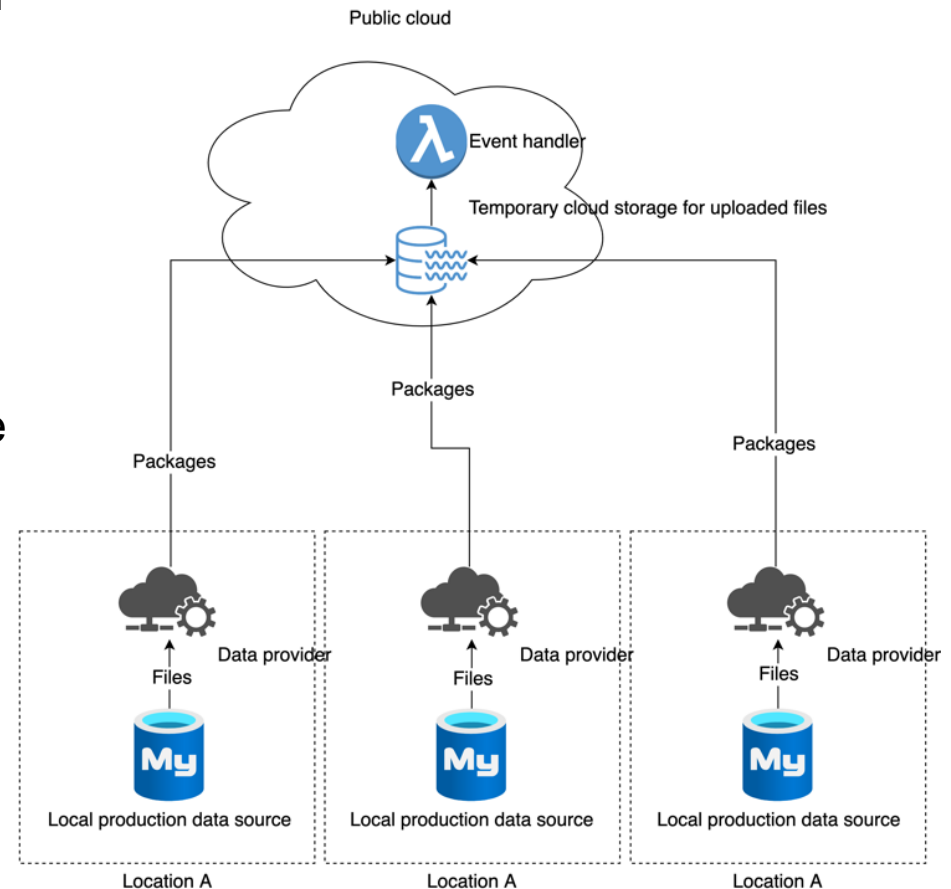► Uniformity is a strict element on the sensor level as well; each sensor apparently has its own schema.

Manufacturing execution software

My

Production raw data

Plant connector

Shop floor

adesso
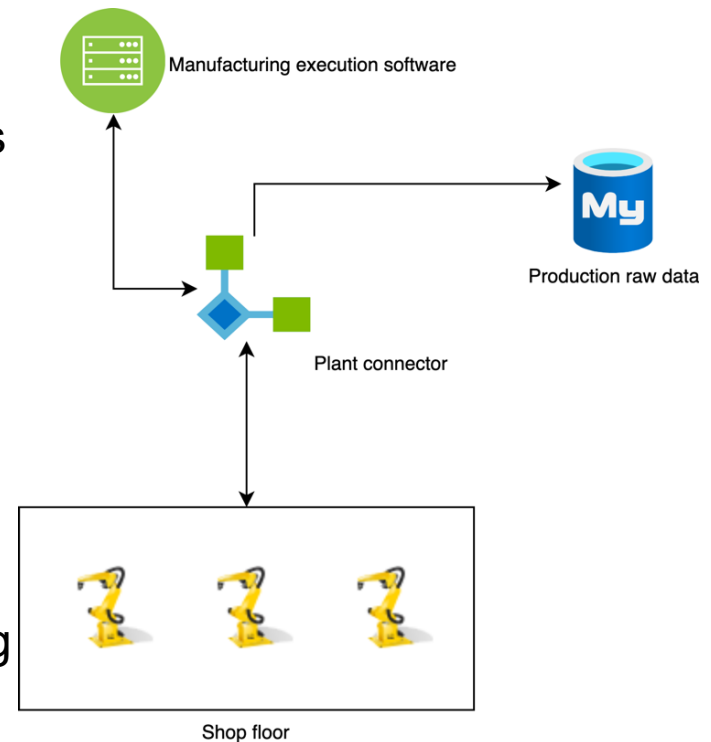
► Lack of business perception of data science might catch the wrong problem. Having an engineer-mindset is a must-have in the factory.

► Data engineer: constructing a bridge between the data-science and sensors, producing consistent input for algorithms.

► In a pipeline, the company can clear its data for performing the claimed input of machine-learning algorithms.

► The data engineer sets up the data pipeline by practicing some DevOps and software engineering competencies.

adesso

▶ Scope: manufacturing in the multi-national environment

▶ Strategy: from the general – to more specific use-cases.

▶ Standardization! Integration engineers set up the schemas on the machine level.

▶ Multi-national: each plant where the data-based solution is injected should be integrated into a cloud-based landscape.

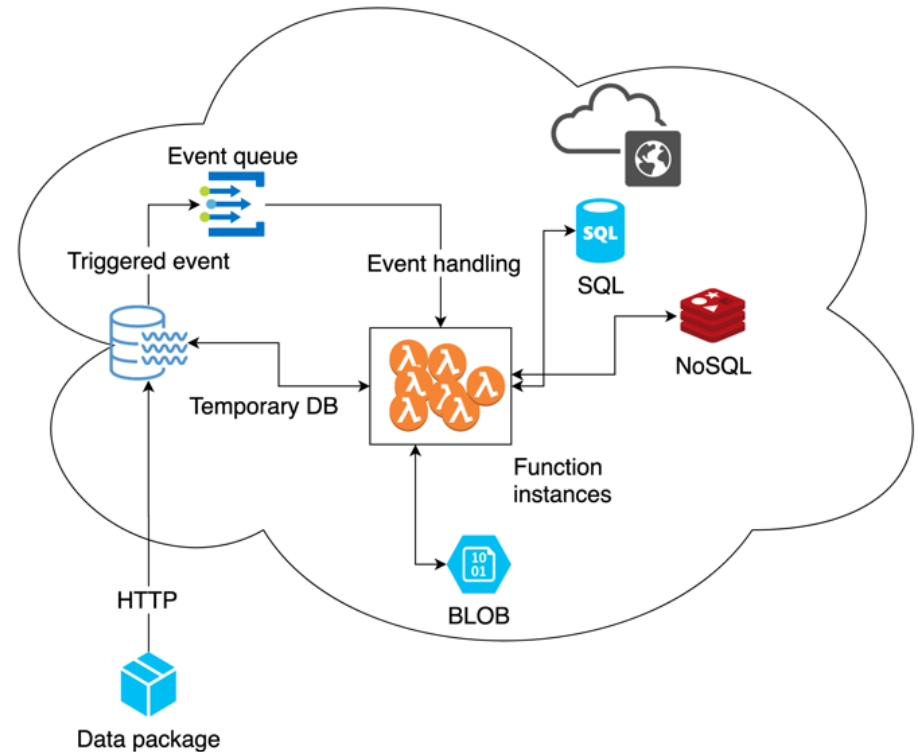▶ The quality of the model is strongly dependent on the condition of the training dataset.

Collecting sensor data

↓

Data loading into cloud storages

↓

Processing the data in cloud

↓

Machine learning and reporting

adesso

Cloud connection

- ► The local sources should be joined into cloud-based storage
- ► The pre-processing actions are for filtering and highlighting the business value in unstructured data
- ► The uploaded package triggers the event-handler
- ► By event-handling, the post-processing segment is started

Public cloud

Event handler

Temporary cloud storage for uploaded files

Packages

Packages

Packages

Data provider

Data provider

Data provider

Files

Files

Files

Local production data source

Local production data source

Local production data source

Location A

Location A

Location A

► An event handler or scheduler can manage the processing of incoming data

► Queue-based solution (retry mechanism)

► By cloud functions (like AWS Lambda), it is possible to set up a robust application for handling the incoming data (scale-out)

► For storing data, the data enginee has several options

Event queue

Triggered event

Event handling

SQL

SQL

NoSQL

Temporary DB

Function instances

HTTP

BLOB

Data package

adesso

Cloud analytics integration

- ▶ Verified input for machine learning algorithms
- ▶ A microservice-based architecture is responsible for the reporting scenarios
- ▶ UI application

adesso

Summary
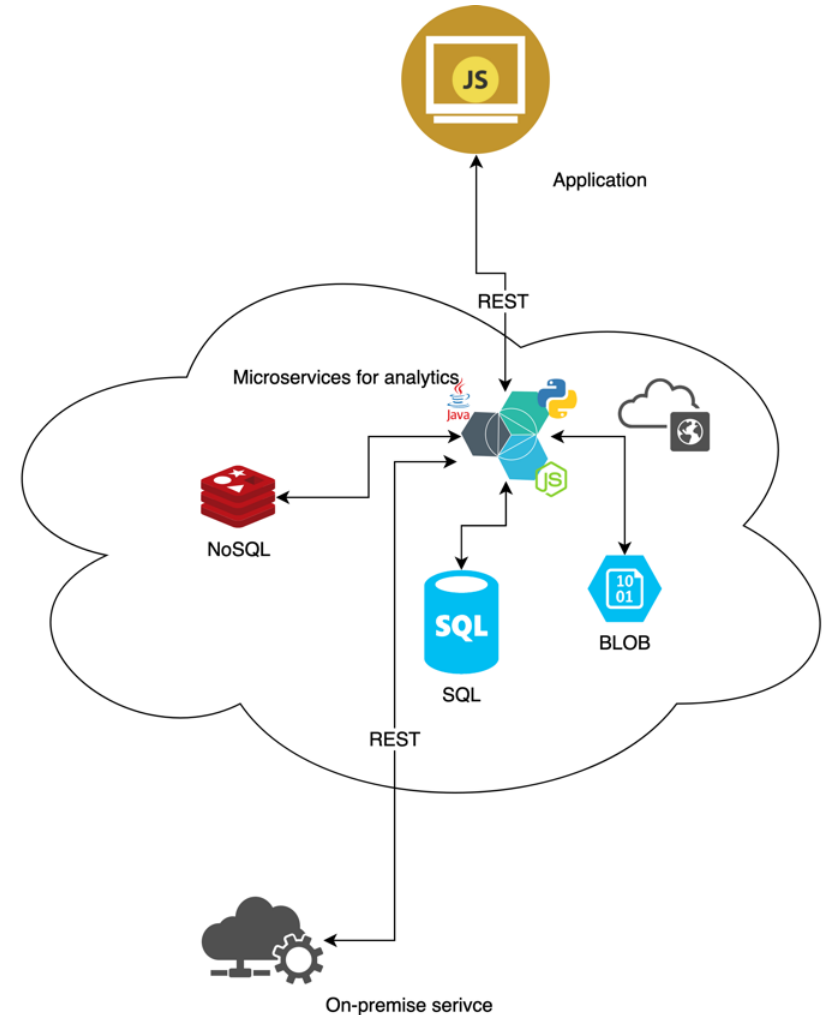
- ► Data pipeline: a combination of big data and machine learning
- ► Challenge: processing the existing data volume
- ► Data engineer: provides the context as infrastructure
- ► Solution: data pipeline, it uses the cloud native's features

adesso