Improving Manufacturing Plants Through Big Data Analytics

SDS2019

14 June 2019

Prof. Dr. Kurt Stockinger

zhaw School of Engineering

Martin Weber

NOSER ENGINEERING
WE KNOW HOW

Marc Schöni

Microsoft

Agenda

- Use Case
- Goals
- Architecture Blueprint
- Experiment
- Conclusions
- Evaluation
- Q&A / Discussion

Agenda

- Use Case
- Goals
- Solution Architecture
- Experiment
- Conclusions
- Evaluation
- Q&A / Discussion

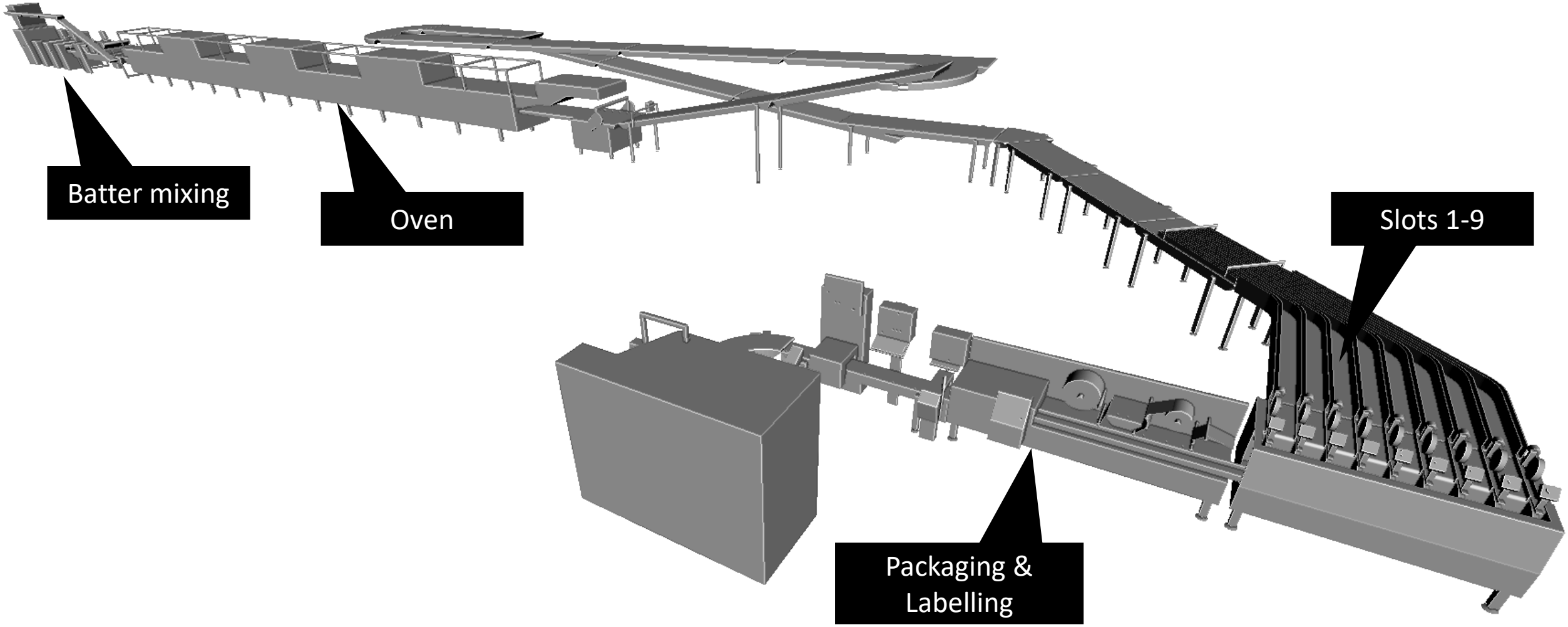
About Midor

- Founded in 1928
- Located in Meilen ZH
- 600 Employees
- Part of M-Industry
- Produce 250’000 items daily for Migros und others
- 32 production lines, 940 different products (different biscuits, ice cream, snacks, dessert powder, etc.)

Introduction

- Production line 16 produces the Blévita, one of Midors sales hits
- Short production stops are reducing the output
- What causes these disturbances?
- Can they be predicted?

Line 16



Agenda

- Use Case
- Goals
- Solution Architecture
- Experiment
- Conclusions
- Evaluation
- Q&A / Discussion

Goals

Qualitative Goals

1. Improve efficiency
2. Flexible and scalable architecture
3. Allows processing of various data formats

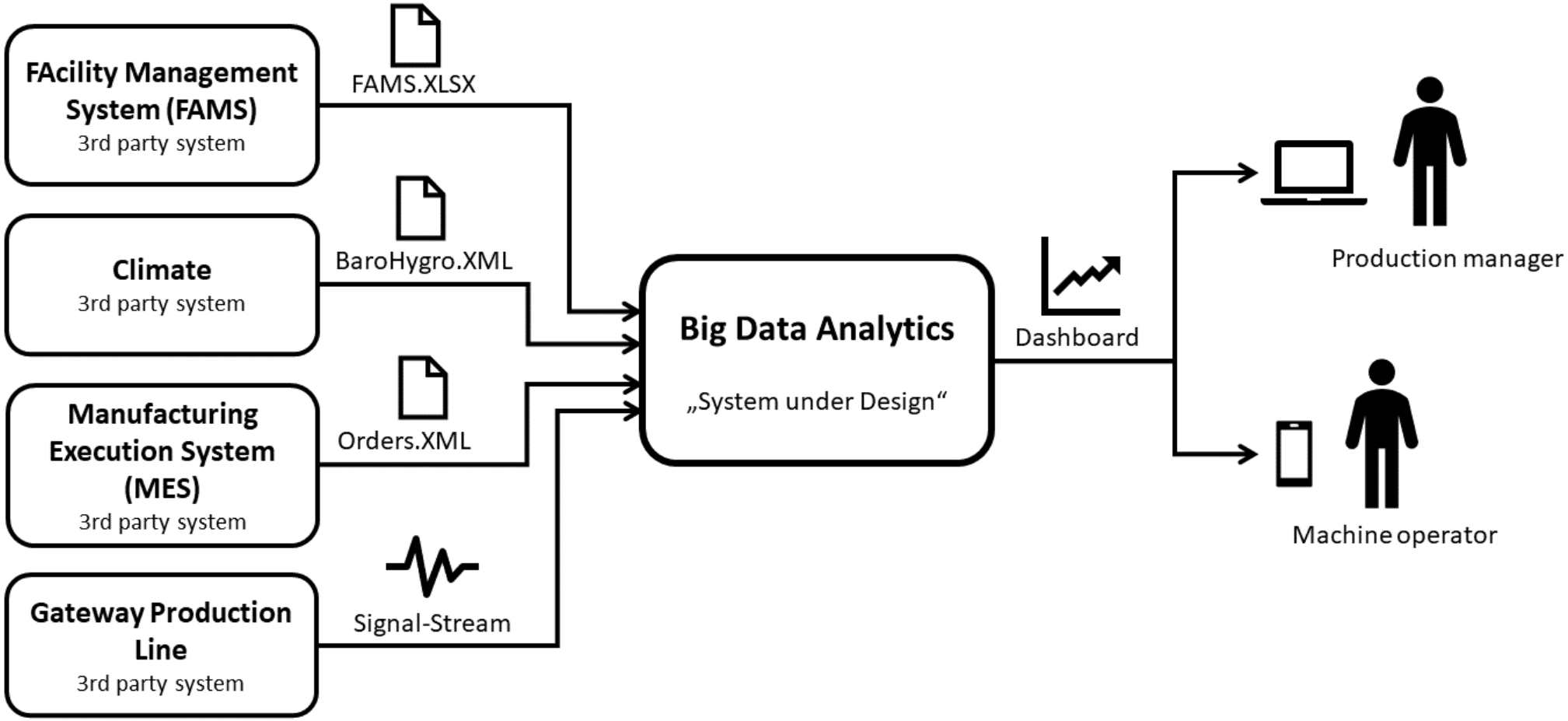
Quantitative Goals

1. Find the most relevant features causing the disturbances
2. Latency for Inference of < 5 seconds
3. Cost should be at worst proportional to amount of processed data

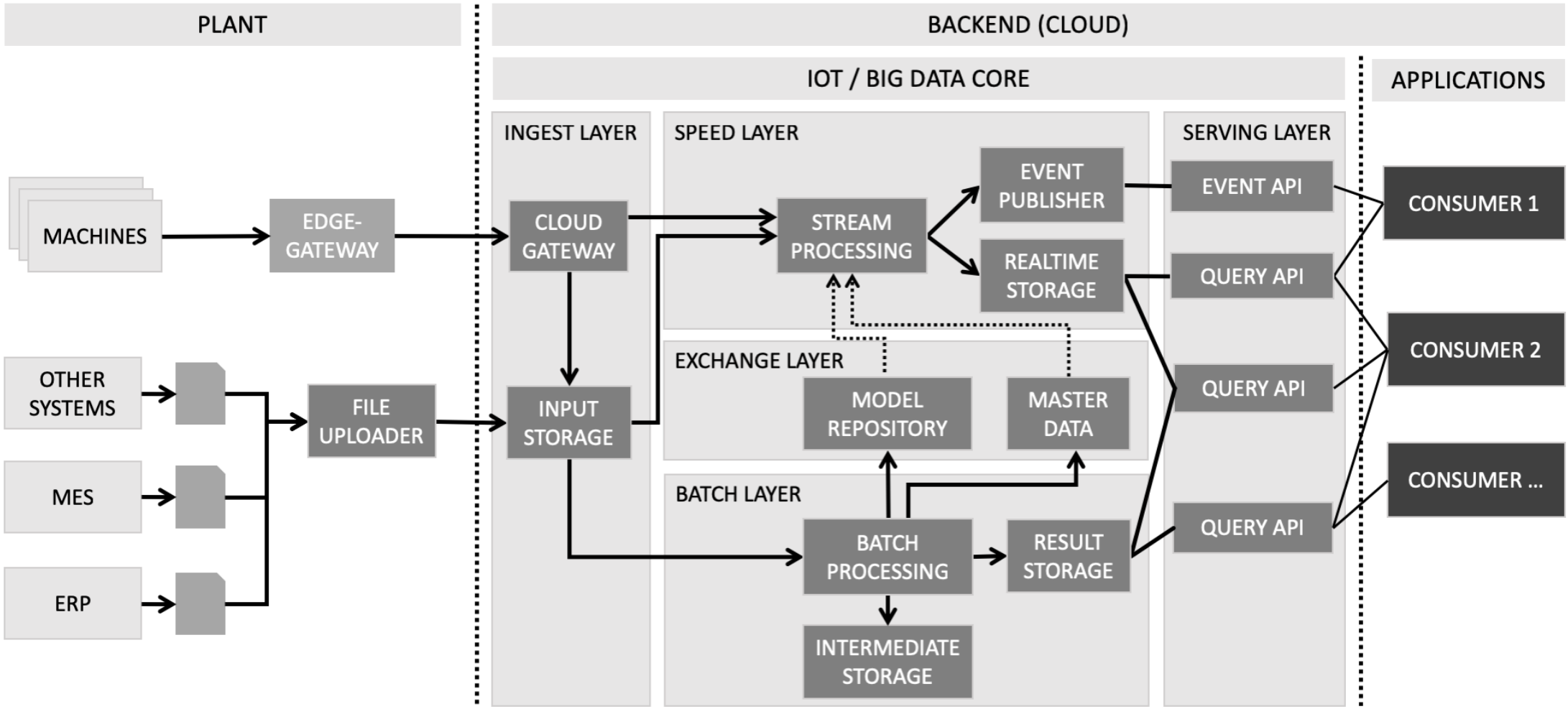
Agenda

- Use Case
- Goals
- Solution Architecture
- Experiment
- Conclusions
- Evaluation
- Q&A / Discussion

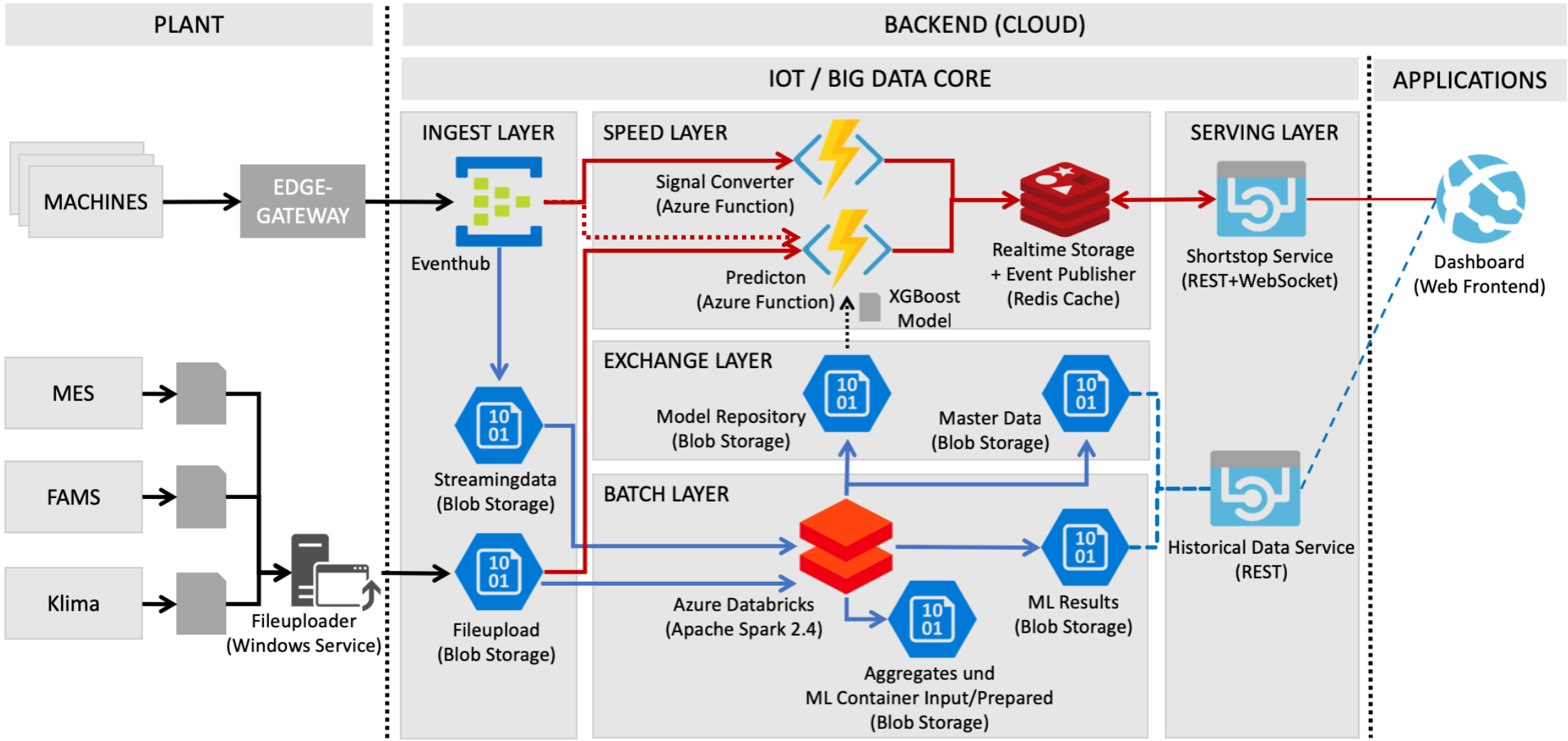
System Context



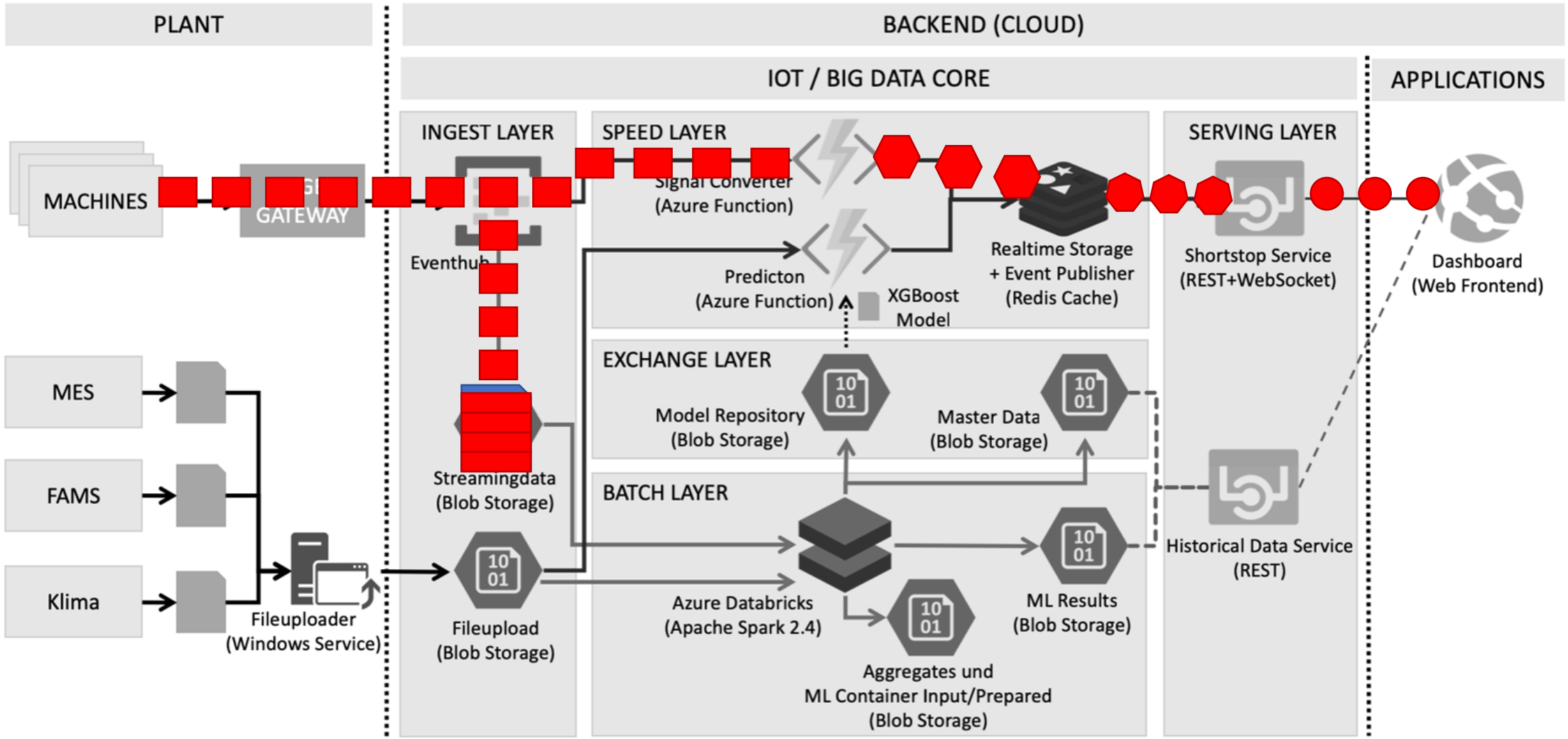
Conceptual Architecture



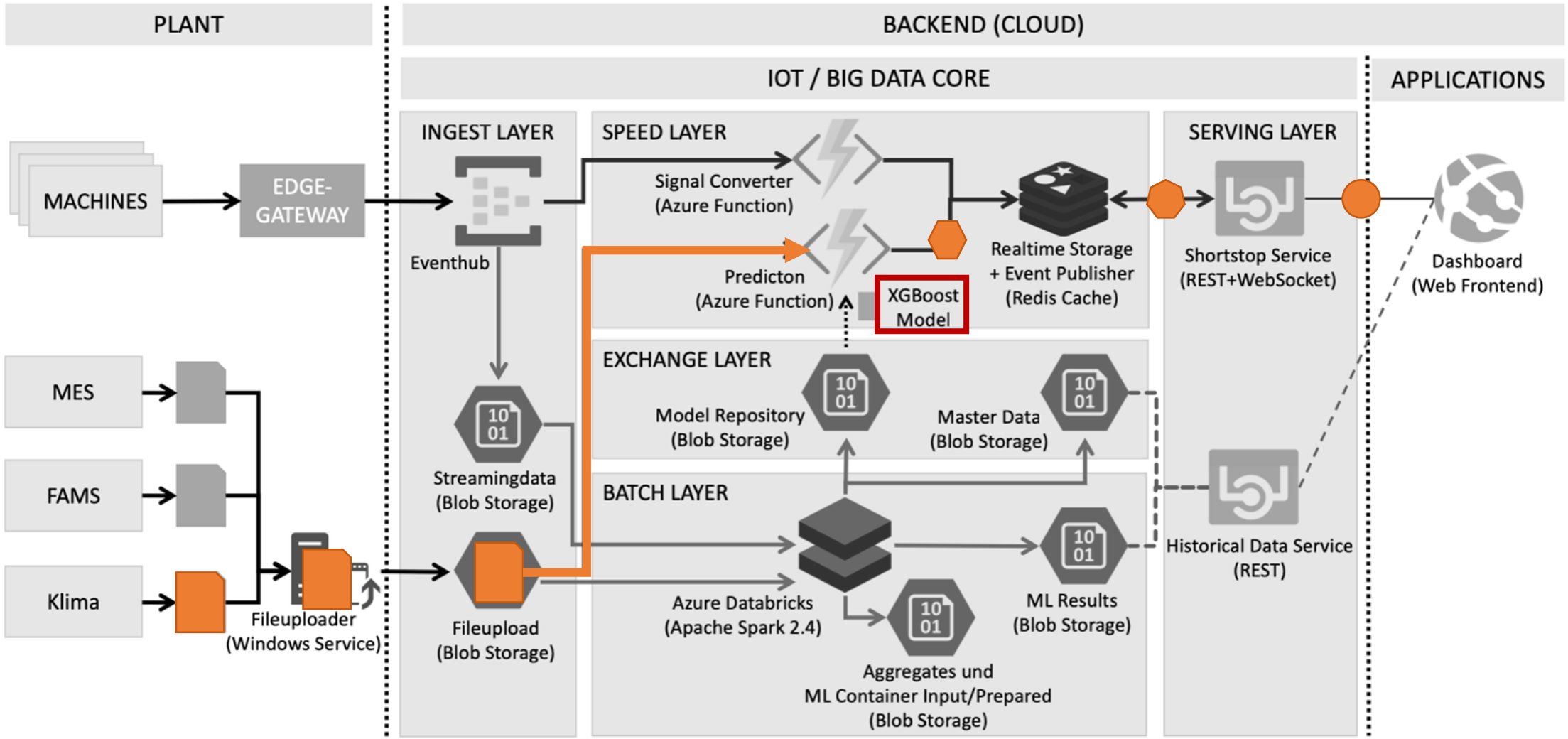
Solution Architecture



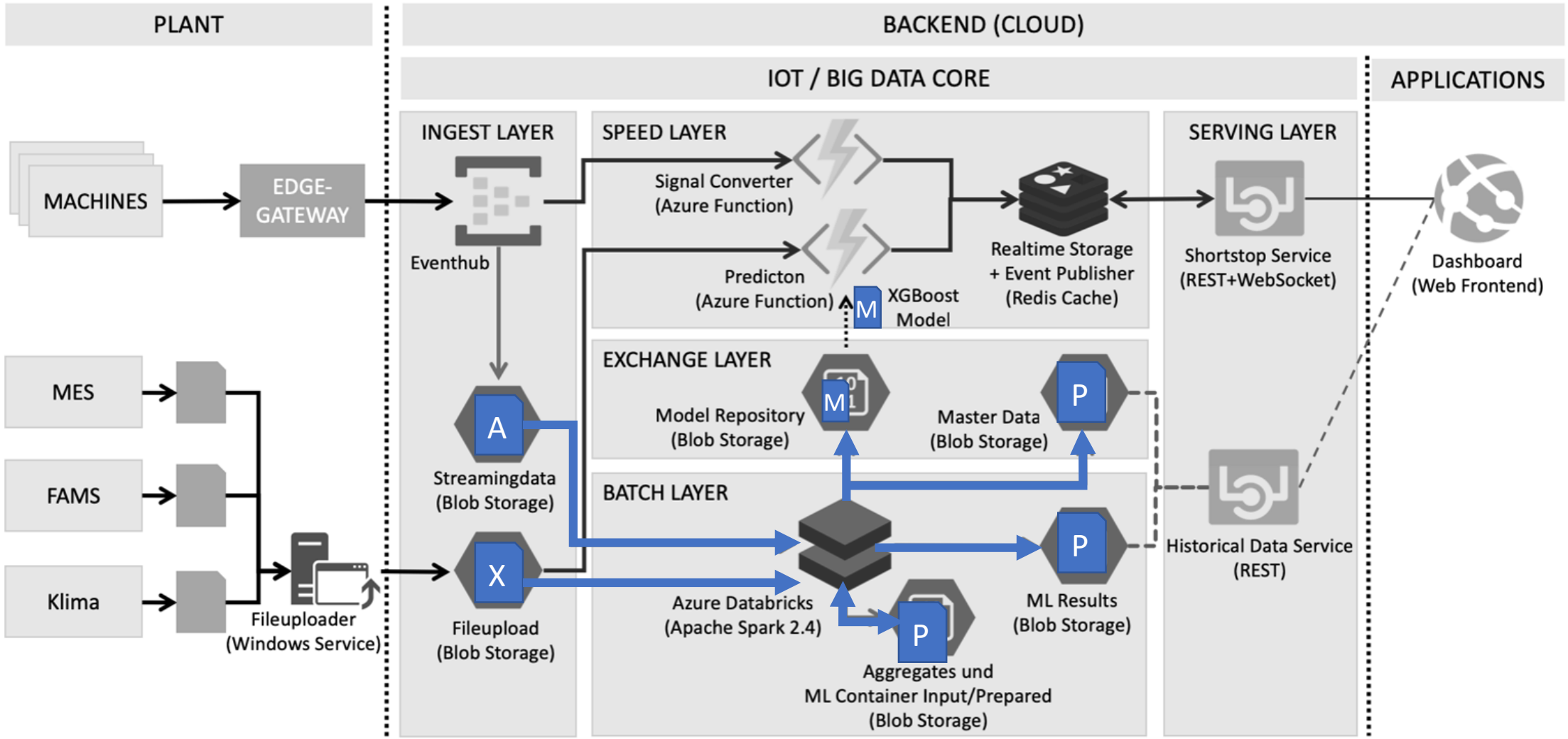
Solution Architecture



Solution Architecture



Solution Architecture



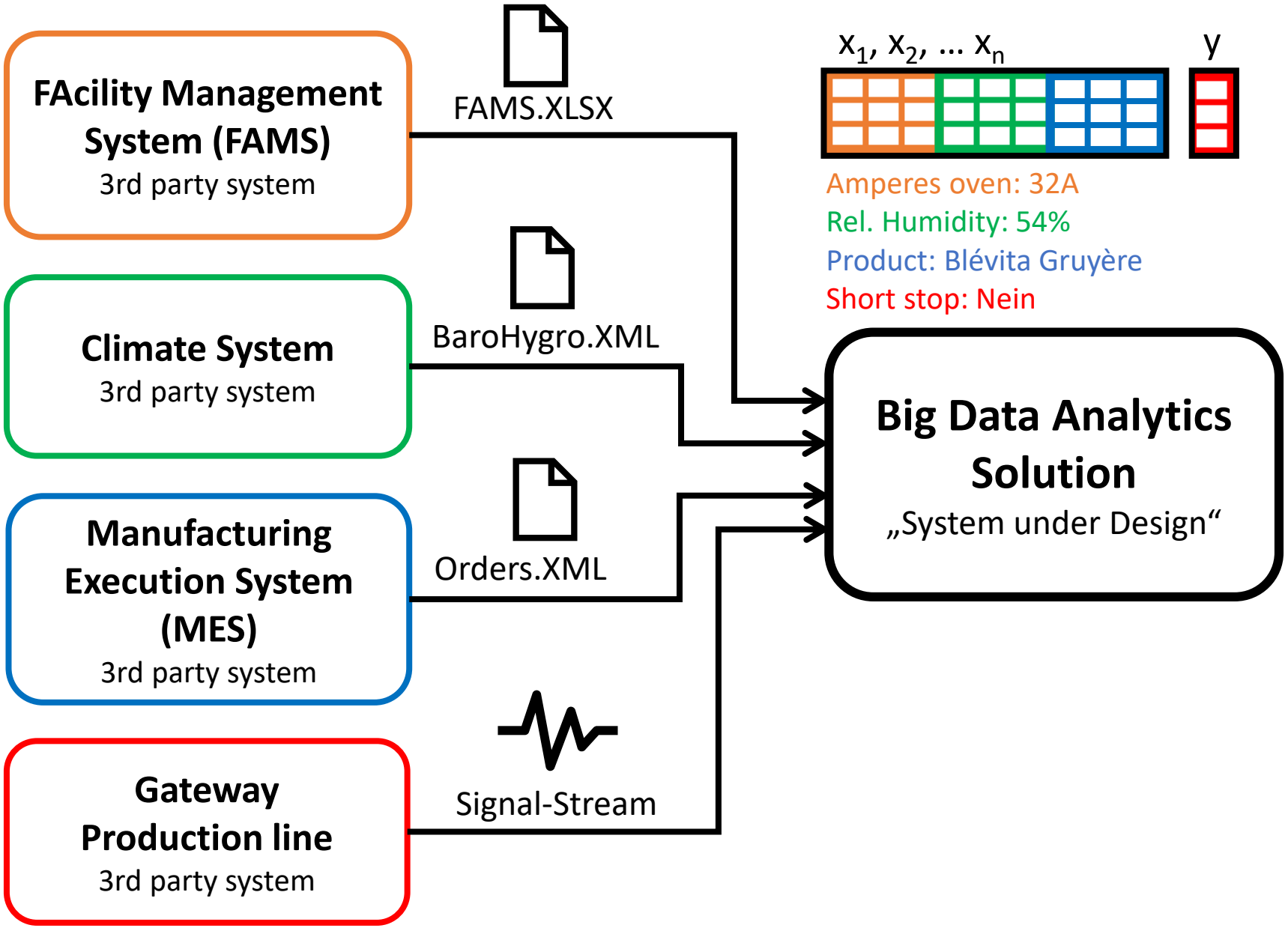
Agenda

- Use Case
- Goals
- Solution Architecture
- Experiment
- Conclusions
- Evaluation
- Q&A / Discussion

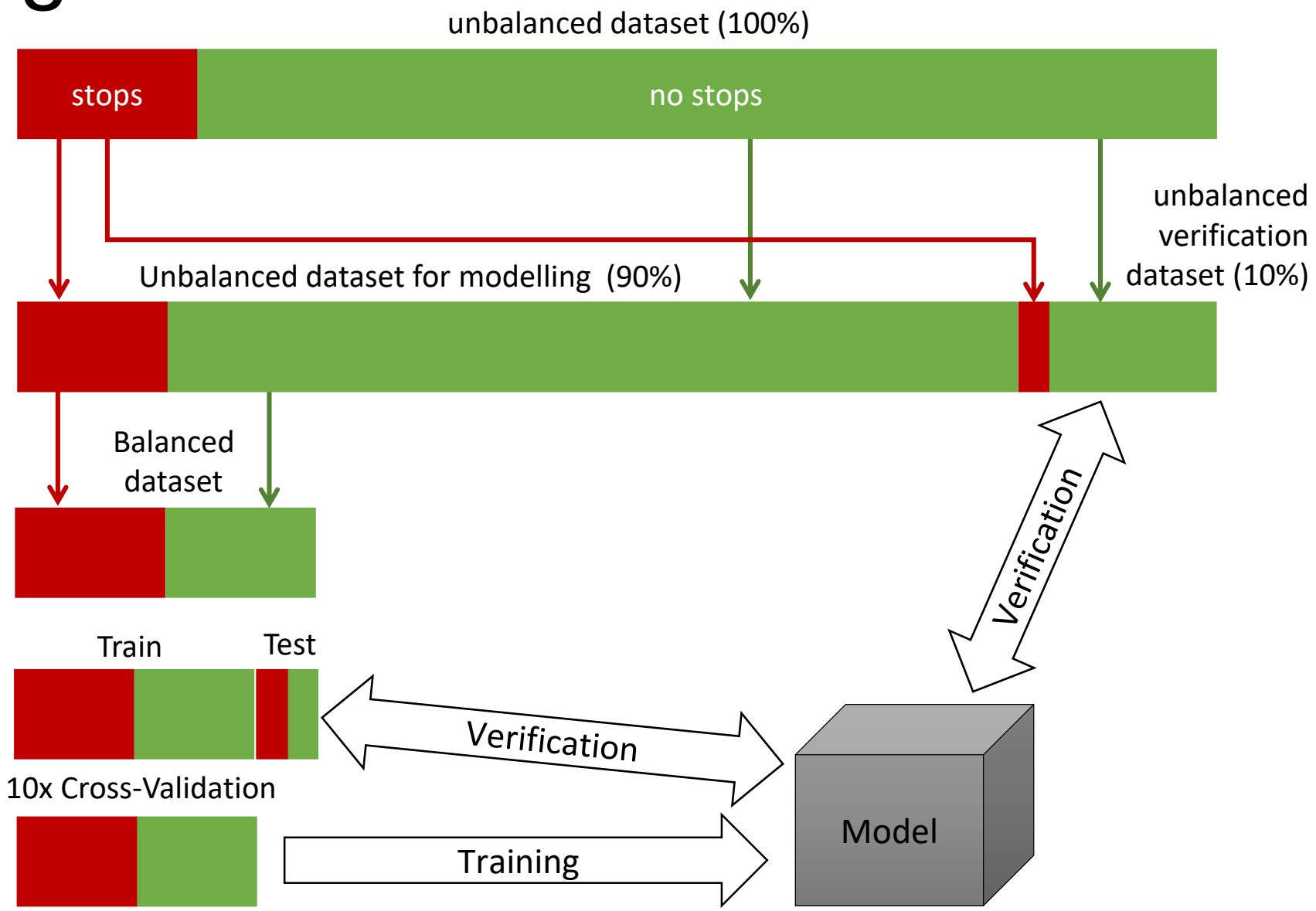
Objective

- Over a period of ~1 year data about short production stoppages was collected (→ Label)
- Over the same period additional data about orders, climate conditions etc. were captured (→ Features)
- Is it possible to find a pattern in these datasets regarding the occurrence of short stops?

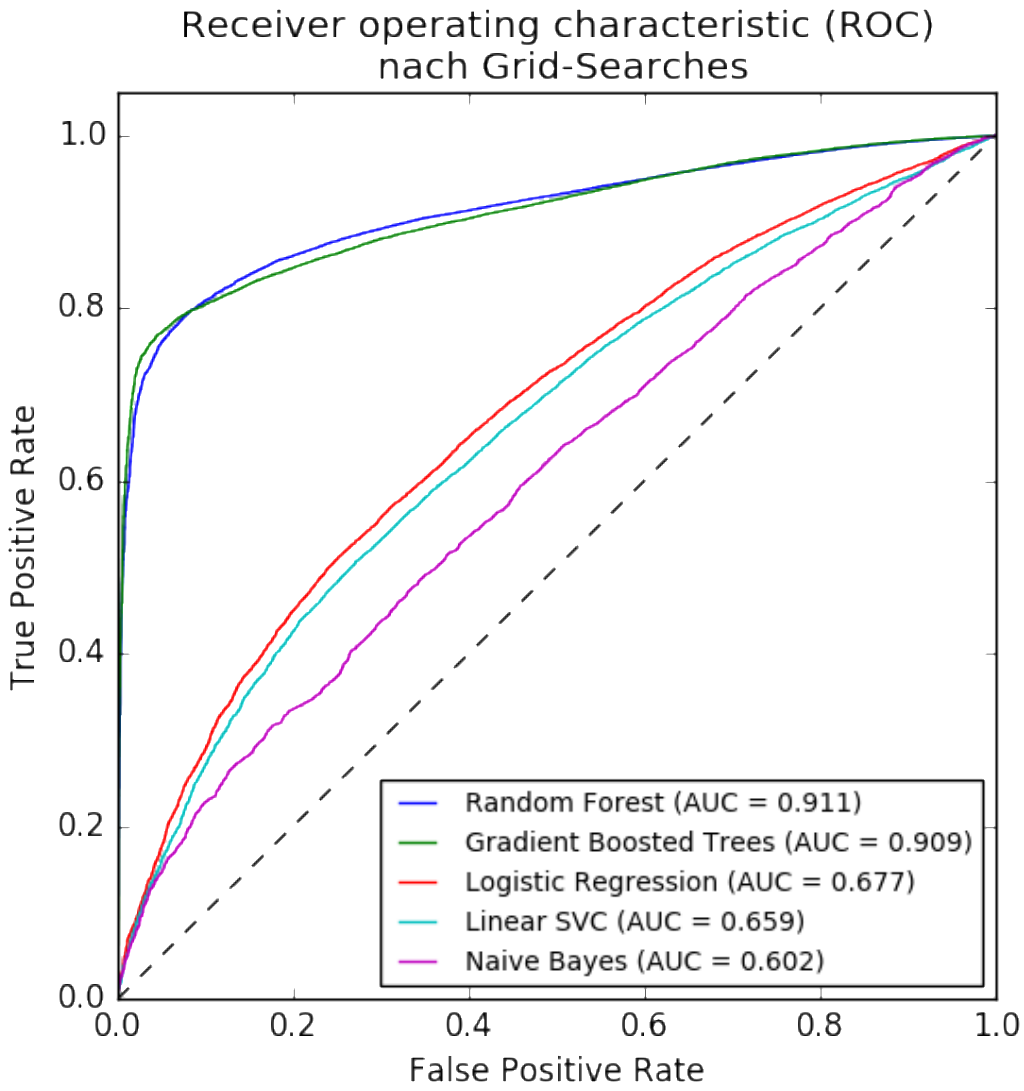
Building the dataset



Splitting the dataset

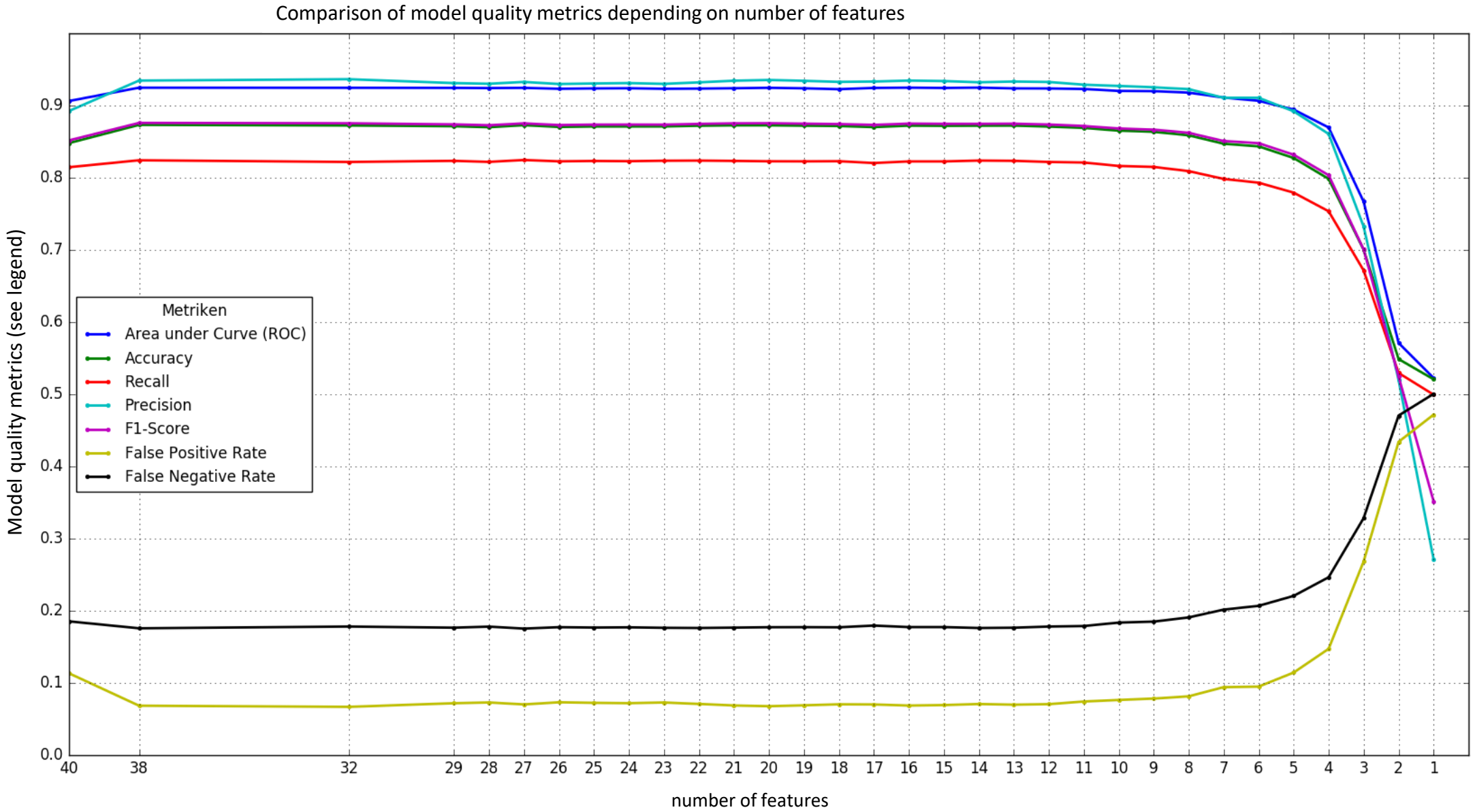


Modelquality

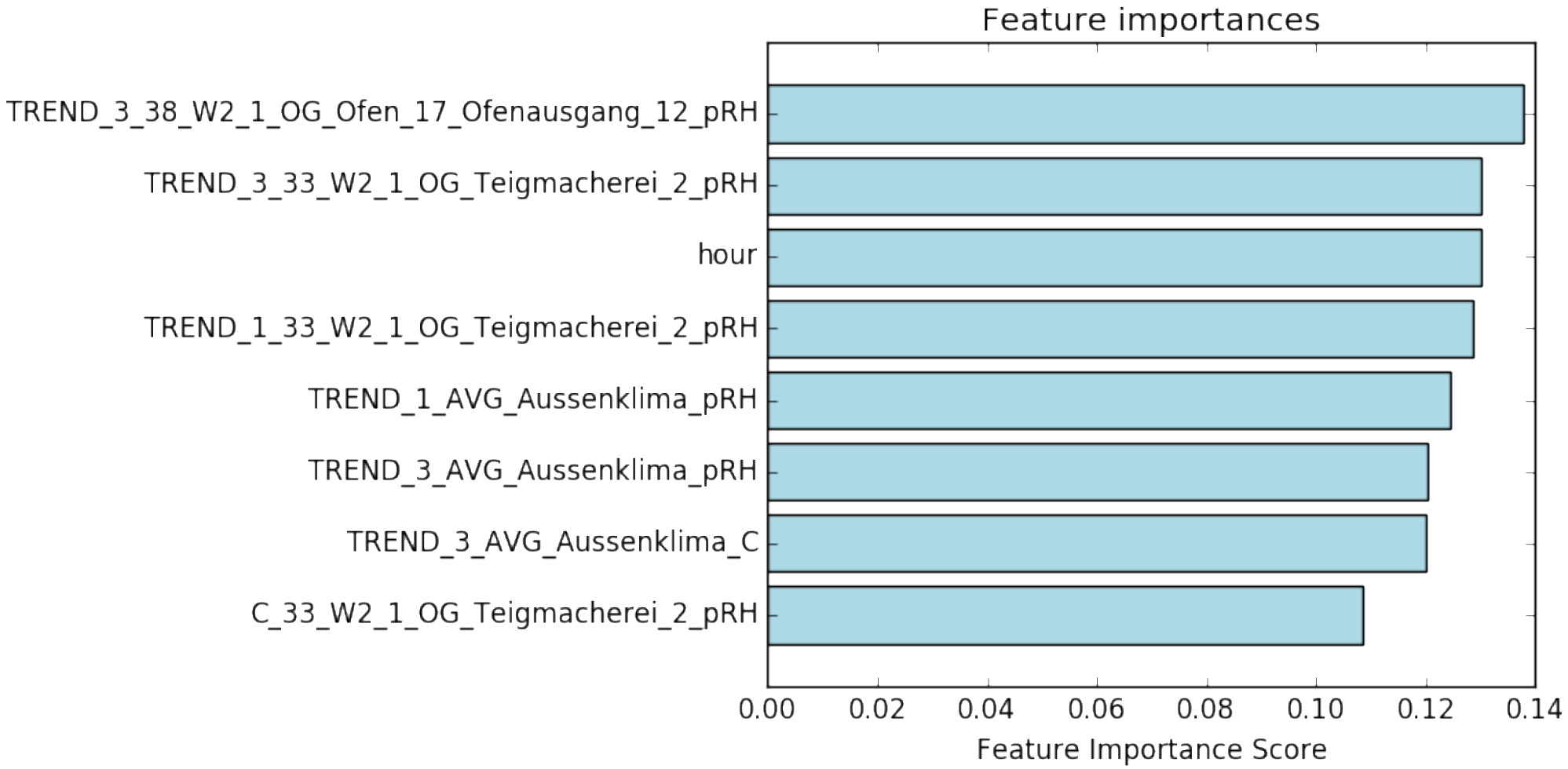


	Accuracy	Precision	Recall	F1 Score	AUC
Random Forest	.852	.818	.897	**.856**	.911
Gradient Boosted Tree	.847	.814	.892	**.852**	.909
Logistic Regression	.629	.616	.644	.630	.677
Support Vector Machine	.612	.692	.588	.636	.659
Naïve Bayes	.572	.680	.552	.610	.602

Feature Reduction



Important Features



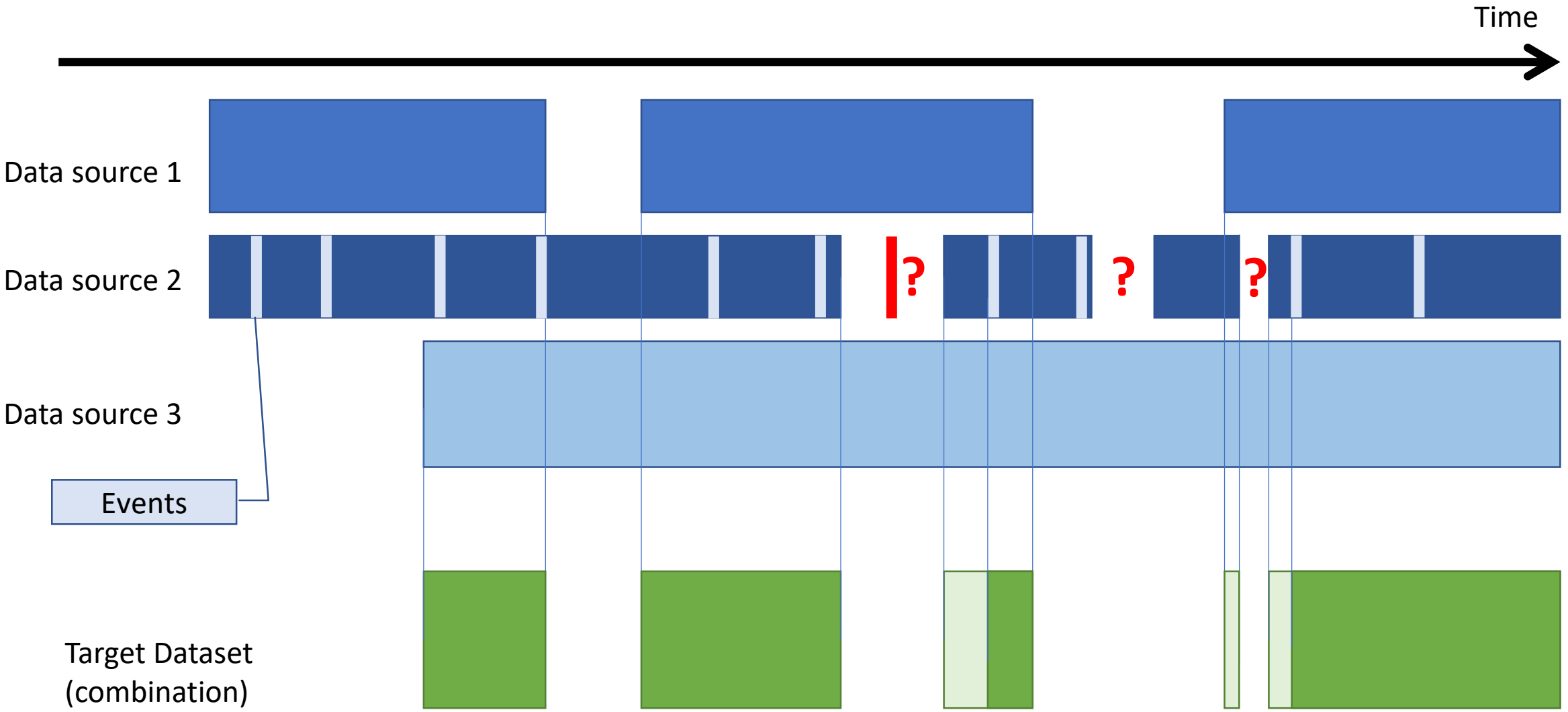
Agenda

- Use Case
- Goals
- Solution Architecture
- Experiment
- Conclusions
- Evaluation
- Q&A / Discussion

Findings: Machine Learning

- We can predict shorts stops with a F1-Score of about 85%
- The integration of different data sources took the most time
- It is not possible to a priori estimate the importance of predictors/features per data source → Integrate all data sources
- The prediction itself does not provide a business value without additional steps (operationalization)

Conclusion: Data import



Conclusions: Architecture

- Findings from modelling provide new boundary conditions for the big data architecture:
  - Number and kind of data sources
  - Amount of data
  - Requirements for inference service (Compute, Memory)
- Principles of Lambda-Architecture have proven their effectiveness
- Benefits of Kapa architecture (single code base) using libraries
- Tools for ML-Pipeline export & operationalization are in early stages
- Monitoring of data quality is a crucial success factor

Agenda

- Use Case
- Goals
- Solution Architecture
- Experiment
- Conclusions
- Evaluation
- Q&A / Discussion

Evaluation

Quality Goals

1. Improve efficiency ✓
2. Flexible and scalable architecture ✓
3. Allows processing of various data formats ✓

Quantity Goals

1. Find the most relevant Features causing the disturbances ✓
2. Latency for Inference of < 5 seconds ✓*
3. Cost should be at worst proportional to amount of processed data ✓

Agenda

- Use Case
- Goals
- Solution Architecture
- Experiment
- Conclusions
- Evaluation
- Q&A / Discussion