Digital.ON

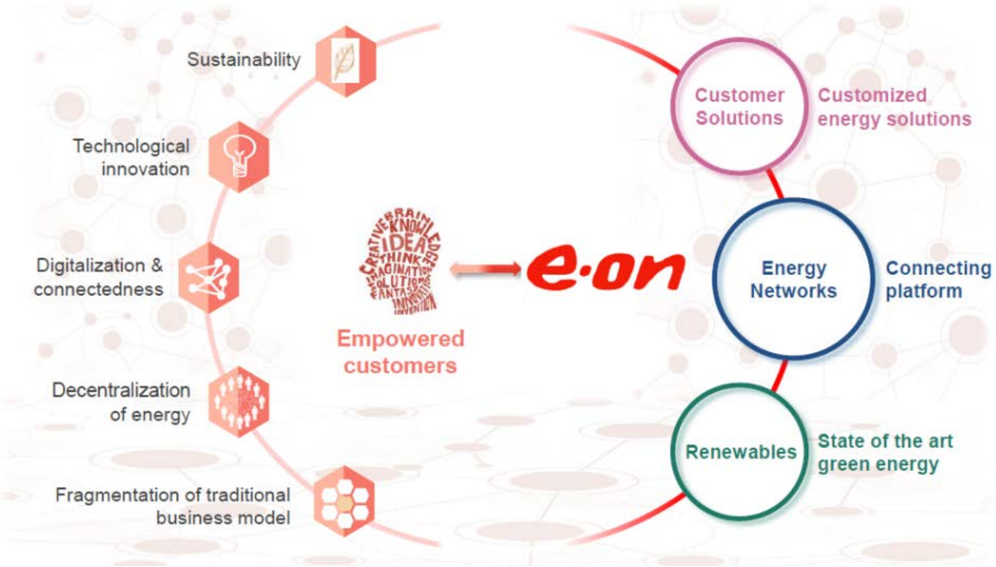
Scaling Predictive Maintenance

Christian Schroeder

Distribution is by far the largest profit contributor within E.ON

Digital.ON

Activities of future E.ON



Objective to develop E.ON's identity in a "new energy world"

- Customer-centric
- Sustainability
- Local proximity
- Small scale, distributed
- Clean technologies

Source: E.ON Roadshow and FY 2017 Presentation

E.ON's initial bet was to implemented a central data science capability

Digital.ON

"Data Science is an interdisciplinary field about processes and systems to extract knowledge or insights from data in various forms."

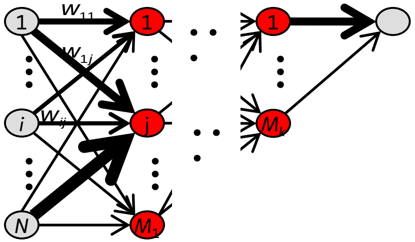
WIKIPEDIA
The Free Encyclopedia

001001010101
001110110101
010101001001

Data

(Maybe) huge amounts of it!
Data Quality/Reliability?!?

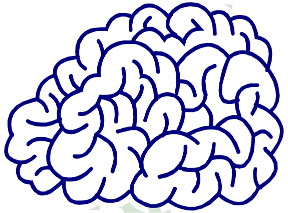
+



Algorithms

Machine Learning
Artificial Intelligence

+



Expertise

Domain & Statistical
Knowledge

TERADATA ASTER

Implement a central
on –premise analytical environment

Staff an data science team, sponsored
by the innovation department

Fund analytical projects,
inviting the organization to explore
their ideas

With Predictive Maintenance we aim for improving on costs, revenues, and quality

Digital.ON

Growing replacement need – allocate CAPEX properly

Serve customers better

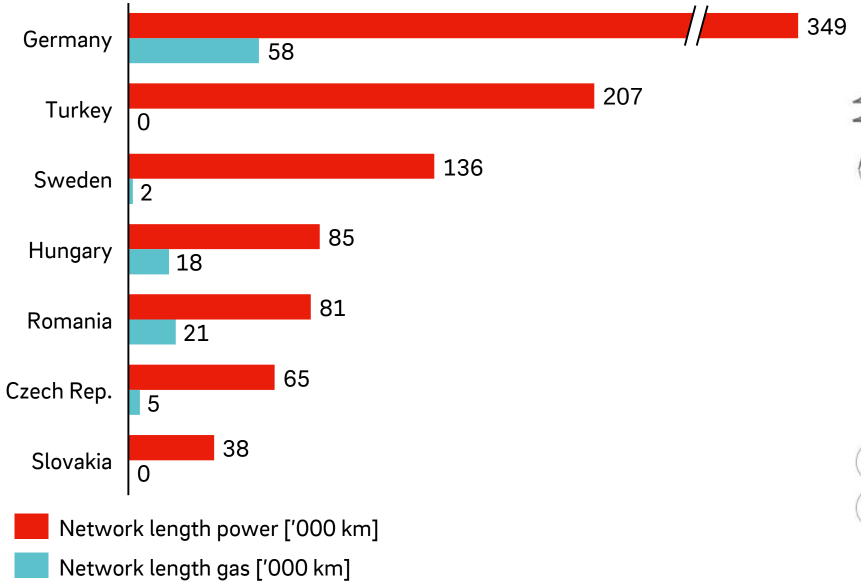
Become more efficient

Grid size does not correlate with data abundance, due to the heterogeneity of the information space

Digital.ON

Overview of grid lengths

Grid activities of E.ON



Source: E.ON Intranet / Eurelectric: Power Distribution in Europe – Facts & Figures

E.ON had its „customer journey“ into data science

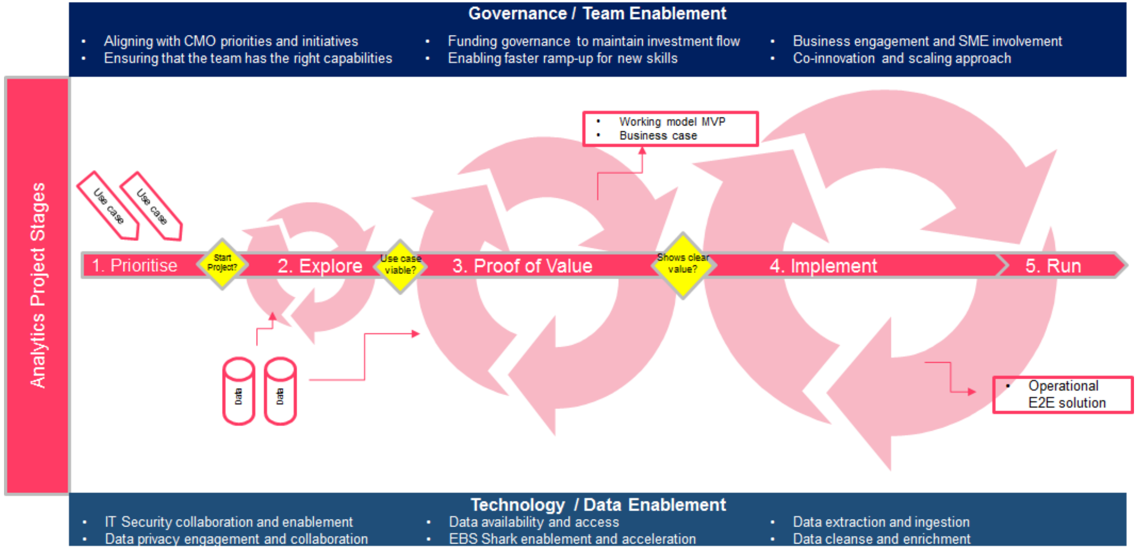
Incremental approach towards successful solution implementation

- Main statement
  - The organization is on a journey, much as your usual customers.
    You need to guide them through it as a consultant
- Steps
  - Stages, similar to a customer journey
    - Information, create interest
      - Explain what data science is, how it works, and which common examples the users might refer to, when think about the possibilities
      - Eventually, after a while the organization will have ist own success stories in data science, and can use them as a showcase
    - Clarify needs, and scope
      - This includes clearly defining the initial hypotheses
      - The assumed value should be stated, and a assumed baseline should be included, too
    - Check feasibility, with data quality assessment and EDA
      - Use sample data to prove that the hypotheses are correct
    - Develop analytical models, and align with SME

- Notes
  - Similar to design thinking process
  - Use the following parameters to select you path
    - Statistics on outages, and impact analysis
    - Feature engineering along conventional physics and electrical engineering wisdom
    - Statistics on data quality, and explore imputing dat

In the log run, we had to evaluate our odds of success for each and every idea...

Digital.ON



- **More ideas than capacity to serve them**, forces you to quickly assess your odds of success
- **Ideas are sometimes only speculations**, so you need to validate them continuously throughout the process
- **Data is only theoretically available**, so checking the quality, and adequacy early on
- **Business value is not always quantifiable**, and therefore difficult to argue for

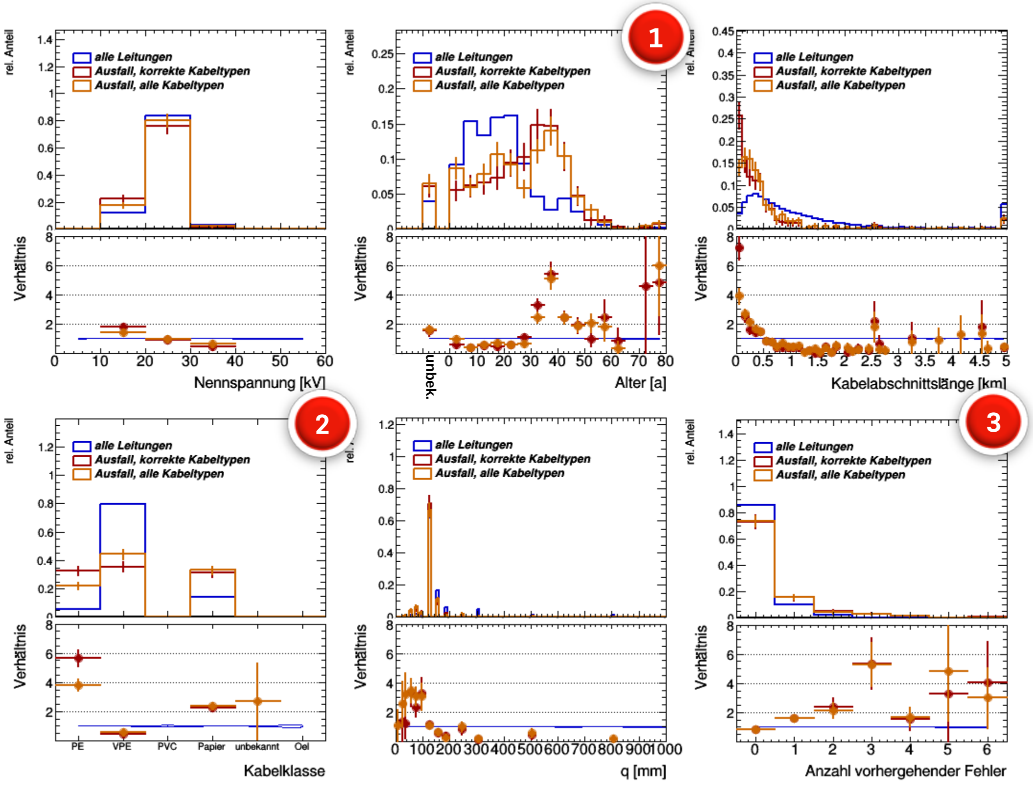
Three-step logic to move ahead with different approaches

Digital.ON

Approach	Explanation	Steps to take
Exploration	• Validation of a new or extension of an existing hypothesis • Usage of data extractions • Focusing the analytical effort	• Data gathering • Explorative data analysis • Modelling
Industria-lization	• Translation of the exploration into a real solution • Technical connection of (new) data sources • Piloting of the solution in the real IT-surrounding	• Platform development • Development of data integration • Tool development
Roll-out	• Pre-study with focus on data assessment • Transition of the solution into the Service Management • Briefing and training of employees / users	• Transition into the line • Service Management

Explorative Data Analysis: Joint work between data science and electrical engineering

Digital.ON

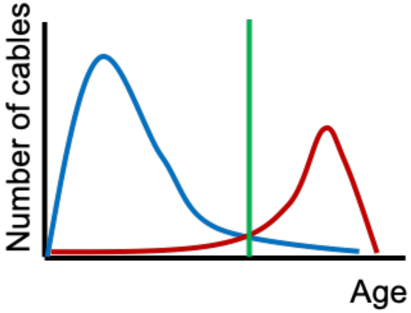


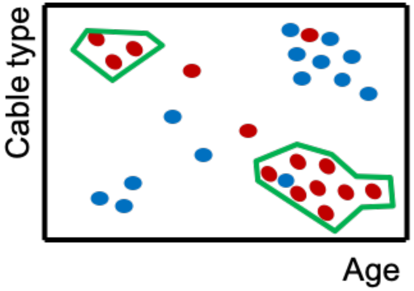
1. **Cable age:**
   cables of an age about
   30 – 40 year show increase outage patterns
2. **Cable type**: PE cables show an increased outage frequency
3. **Outage sequence on MV-line**:
   Increase of outage rate for previously damaged lines.

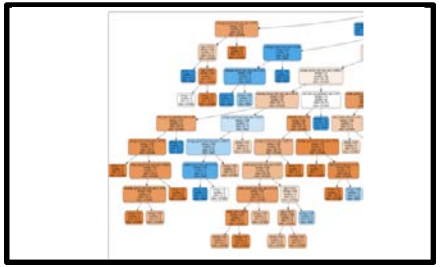
“Supervised Machine Learning” is our current option, based on the amount & quality of data available

Digital.ON

Self-learning algorithm





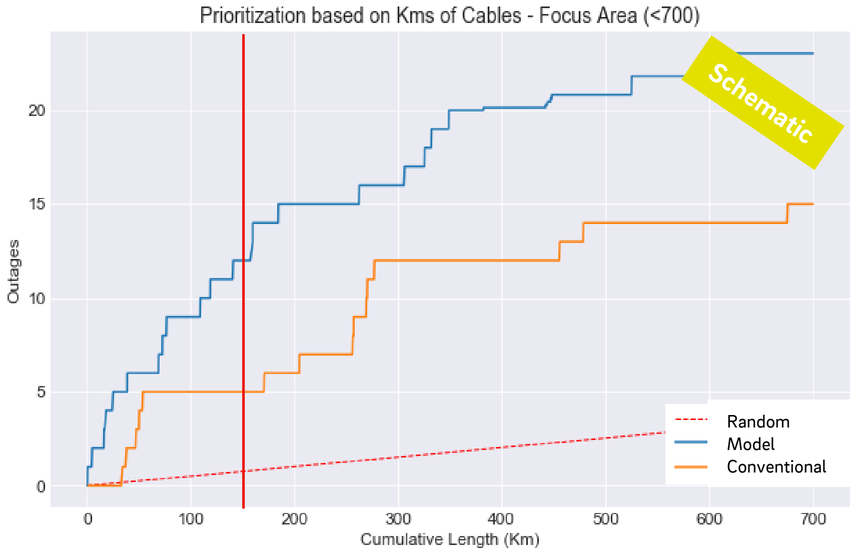


Requirements: Substantial number of training records.

Goal: find optimal classification.

In essence, we use a data-based time machine to understand model performance

Digital.ON



Test of model performance

Method:

1. Pick past network state at **1.1.2015**
2. **Train** model with **outages prior to 2015**
3. **Predict outages** for given network
4. Compare prediction with **outage history in 2015**

How many outages could have been prevented by replacing the first 150 km that are predicted to most likely fail?

The typical result of our solution is a replacement priority list

Input

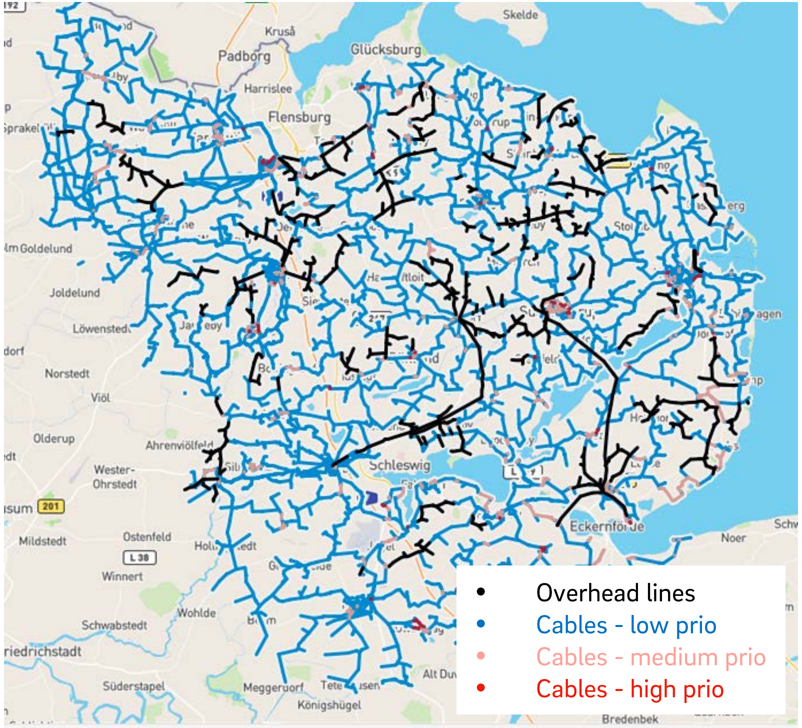
- MV-cable data and outages
- **Inner cable faults** without known reason

Modelling

- Yearly prediction
- Classification algorithm

Output

- Ranked list
- Employed by DSO by considering also soft criteria and ongoing projects



MV-cables in SH-Netz with PdM priorities

The project technically had a "free of charge" entry point, but gradually was moved out to fly on its own

Digital.ON

TERADATA ASTER

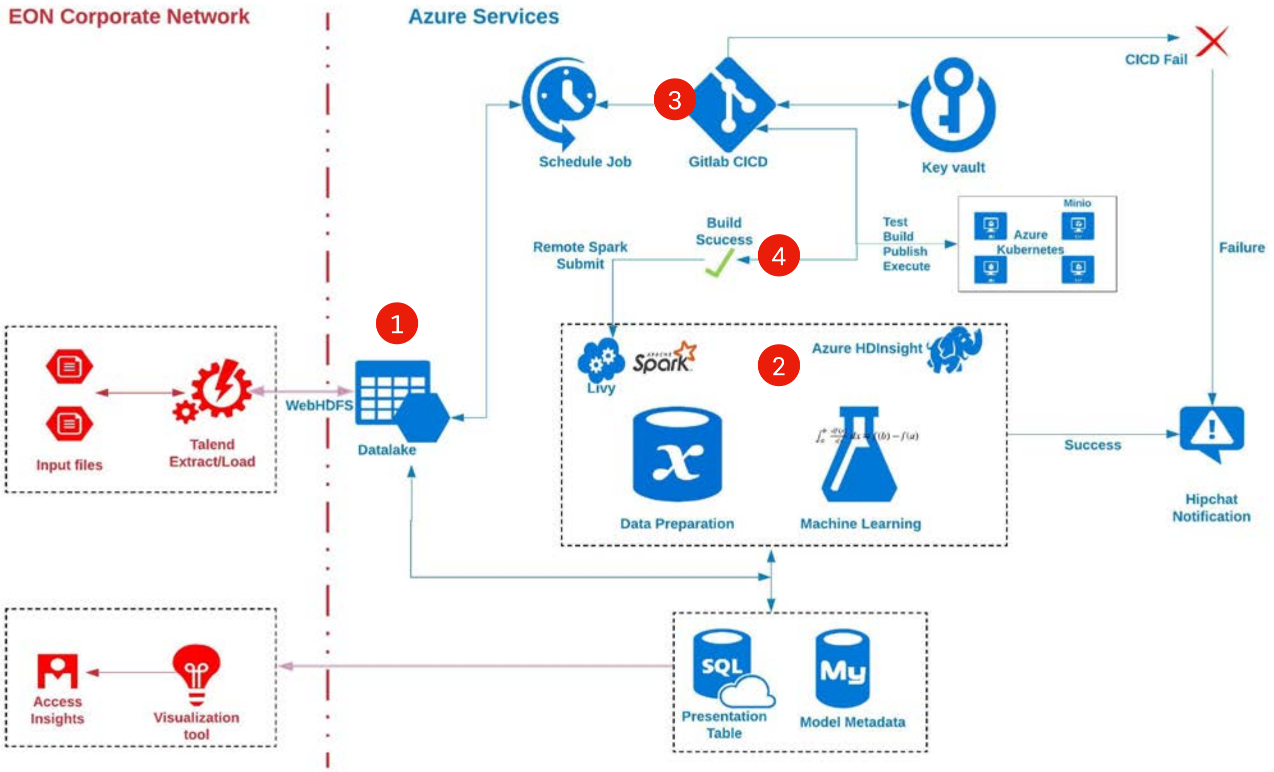
	Teradata On-premise Big Data Appliance	Azure Private Cloud-based Central Platform	Azure Private Cloud Based Group of Solutions
Pro	• Clear, centralized funding • Dedicated high performance hardware • Tight information security control • Skill requirements are predictable	• Self-maintenance allows pick-and-chose platform integration • Easy start for new users and projects	• Projects need to organize their cloud funding • Highest flexibility in solution design
Contra	• Highest cost • Confined to the capabilities of the provider	• Funding • Skill ramp-up grows with the number of tools	• Governance on tools, architecture, data stores
Costs	• High	• Moderate, pay as you go	• Higher on a per project perspective
Decision	• Abandoned in favor of a cloud approach	• Partly decommissioned, in favor of dedicated resources per project	• Increase architecture governance • Use templates, to foster common best practices

Azure Azure

The combination of these is perfect in our case

We aim for industrialization, to improve efficiency and gain quality

Digital.ON



General features:

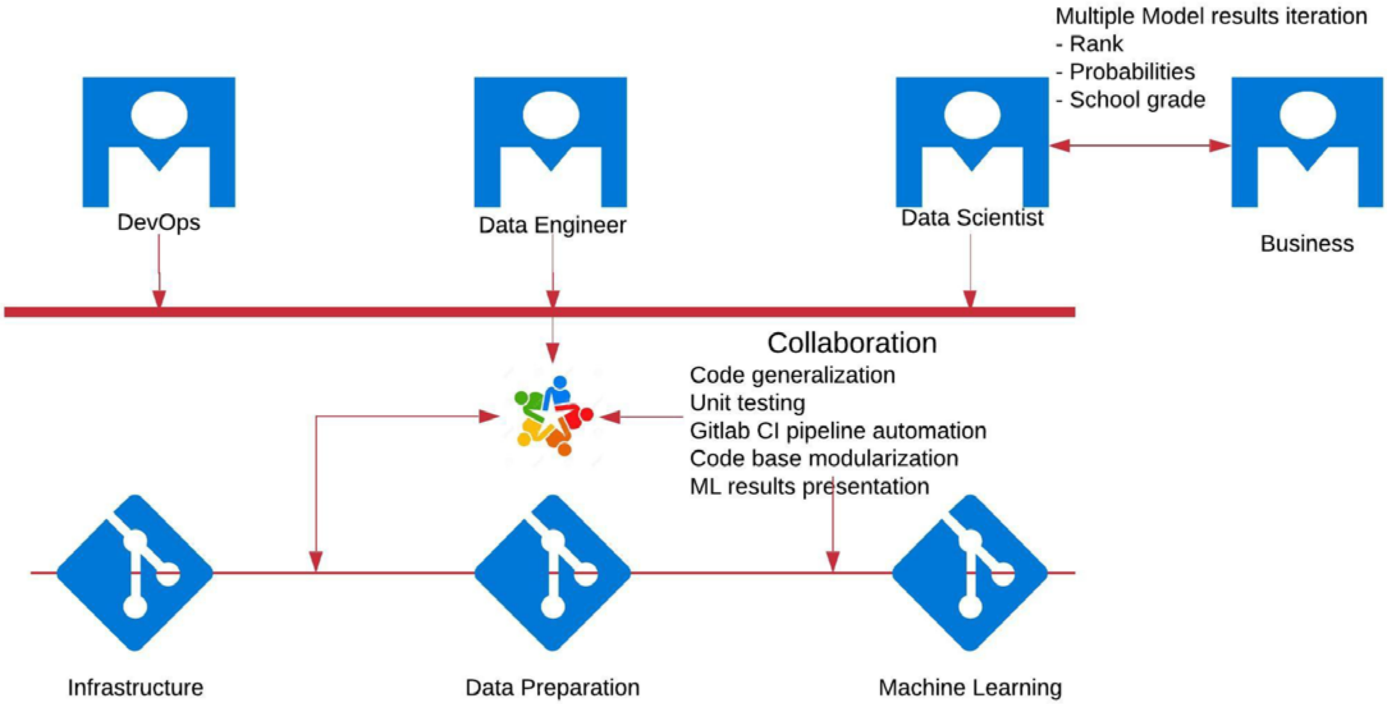
- Cloud based
- Heavy use of open source big data tech.
- Scalable

Special features:

1. **Datalake**, to historize data
2. **Common libraries** for common DSO data processing
3. **Job control** for "everything", incl. infrastructure setup (IaaS)
4. Automated **E2E test framework**

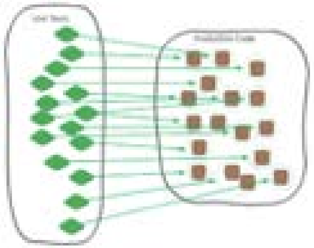
Over time, we transpose the group of distinct roles, in to a project team, enacting the capability

Digital.ON

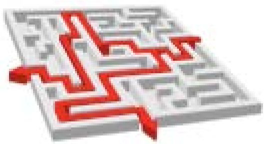


Moving best practices of software development into analytics is our approach to solution quality

Digital.ON



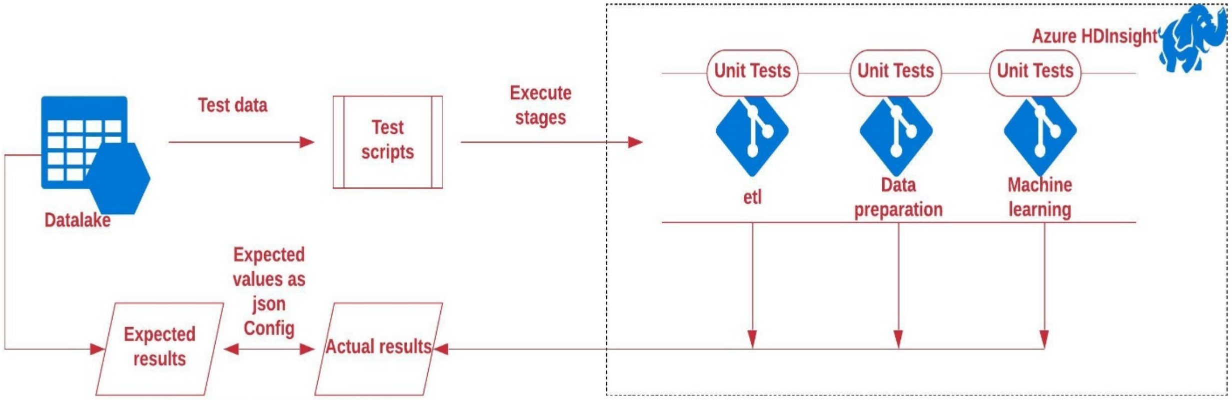
Unit test



End to end test

Testing Framework

E2E EDI Senario

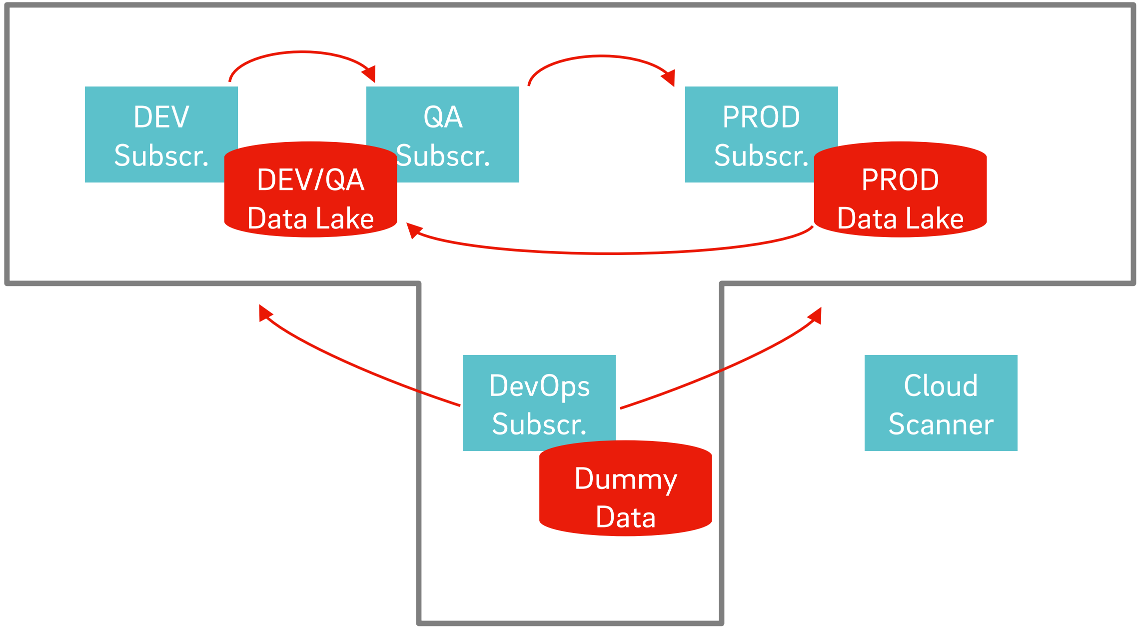


Compare & Checks Conditions

expected no of rows
expected no of columns
unique value
column range
duplicate
prediction data unique value check

A „T-shaped" infrastructure is our approach to keep flexibility, and not sacrifice security in the cloud

Digital.ON



- **Subscriptions**, to allow separating the role based access control
- **Production and development Data Lake syncs**, to keep the data environment usable
- **Infrastructure coding**, with a dedicated subscription and dummy data, to reduce the information security risk
- **Cloud scanner technology**, to systematically and continuously govern the cloud configuration
- **Security reviews and pen-tests,** to find more complex deficiencies in the setup
- **Four-eye principle**, when changing infrastructure code

Rollout of our a single model to an arbitrary DSO is executed along a dedicated work plan, with confined timeline

Digital.ON

Phase 1

Planning & Data Alignment

- First evaluation of data basis & IT landscape
- Conduct 1-day workshop on-site

1 day

Phase 2

Pre-study

- Analysis of data quality (by data engineering)
- High-level relevance analysis
- WS partly on-site

2 weeks

Phase 3

Proof of concept

- Data preparation
- Modelling of one asset class[1]
- Calculate anticipated financial upsides

~ 8 weeks

Phase 4

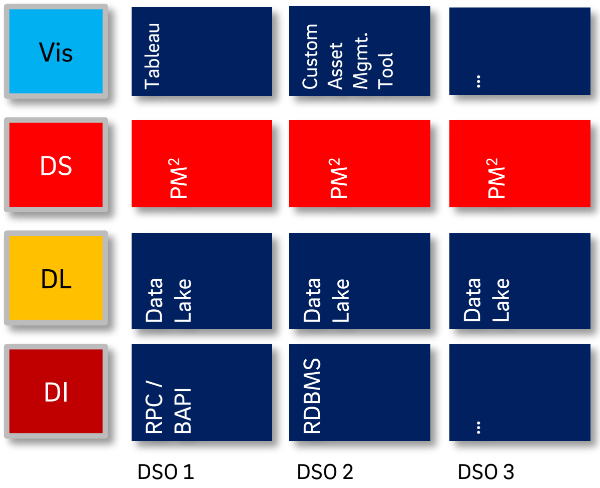
Automation/integration

- Industrialization of first asset class
- Fully implement into asset management

~ 6 weeks[2]

[1] No complete integration of ETL processes, etc.
[2] Dependent on IT landscape

In the Predictive Maintenance product, we try to focus on our core capabilities, and delegate or reuse the rest



1. ESP = E.ON Solution Platform
2. PM = Predictive Maintenance core logic

- **Presentation part** with a multitude of tools and requirements from the users. We try to descope them from the core service
- **Data science and engineering is the core**, and we try to harmonize on that, and keep it centralized
- **Data Lake as a shared component**, so DSO can re-use our solutions data also in other contexts
- **Data sources are diverse, and heterogeneous**, and therefore will need to be delegated to the DSOs