

Progetto BD4M

La Piattaforma BD4M



BD₄M

Big Data For Manufacturing

Prof. Mauro Tortonesi
Università di Ferrara
mauro.tortonesi@unife.it
Referente scientifico progetto

Big Data: caratteristiche fondamentali

John Mashey ('90s) conia il termine Big Data identificando 3 aspetti fondamentali:

- **Volume**: enormi quantità di dati, che vanno processati in maniera efficiente, distribuita, affidabile e cost-effective.
- **Velocity**: i dati sono generati sempre più velocemente, ed è necessario analizzarli in real-time
- **Variety**: dati di formati molto diversi, fino all'80% non-strutturati

Più avanti alla definizione si aggiungono:

- **Veracity**: dati non sempre affidabili, corretti o utilizzabili
- **Value**: enfasi sul valore del singolo dato e su come ottenere valore processandone grandi quantità
- **Variability**: fonti e caratteristiche dei dati possono variare significativamente nel tempo

Big Data e Manufacturing

Big Data è una delle *Key Enabling Technologies* del paradigma Industria 4.0

Diversi possibili obiettivi:

- **migliorare funzionamento di impianti e processi produttivi**
- **ottimizzare manutenzione (predictive maintenance, black-box diagnostics)**
- **ottimizzare supply chain**
- **identificare nuovi modelli di business, ecc.**

L'applicazione delle soluzioni Big Data all'ambito manufacturing è molto recente e lo scenario ancora relativamente inesplorato

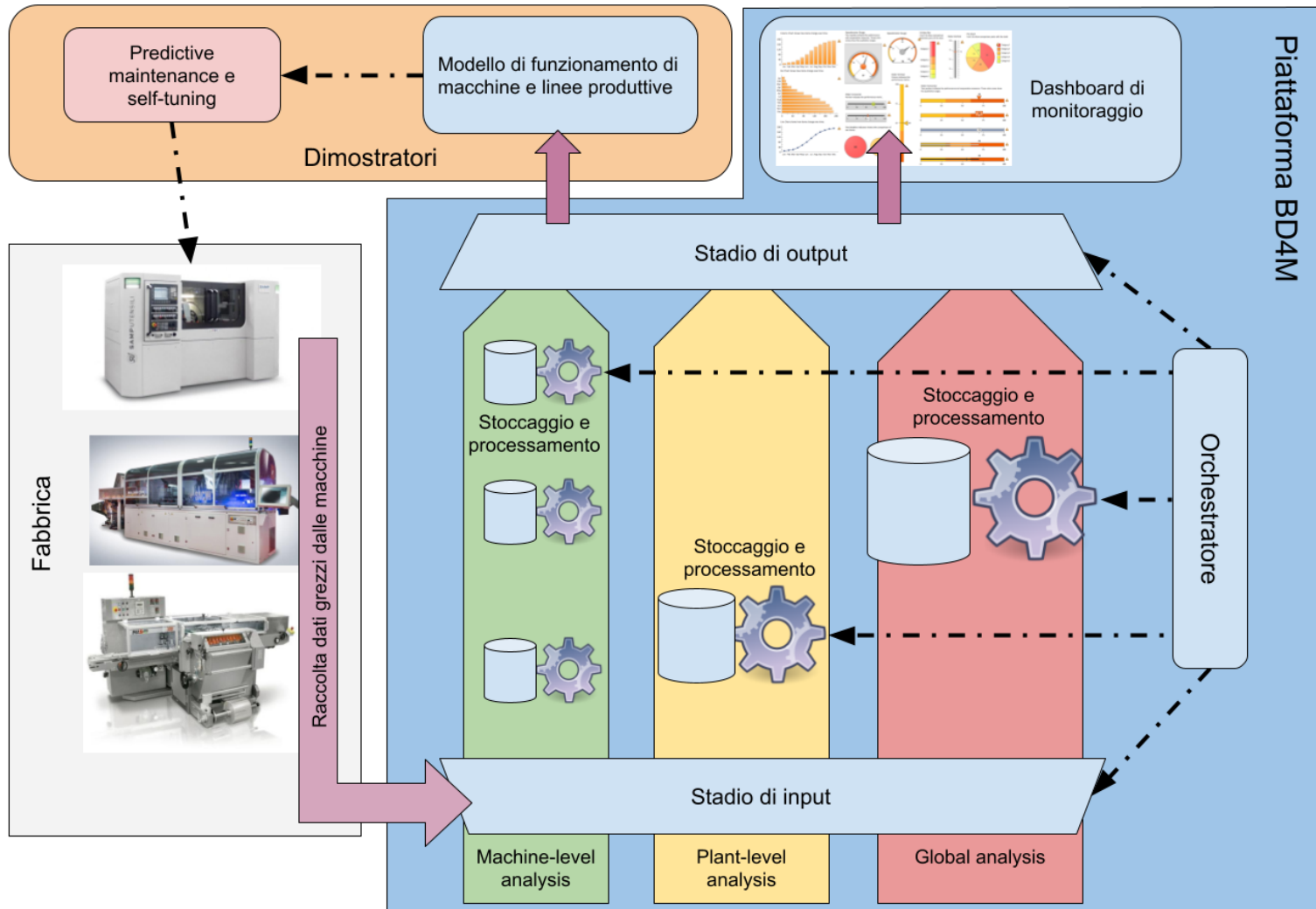
Motivazione del progetto

Realizzazione di BD4M: una piattaforma Big Data di larga applicabilità in contesti Industria 4.0

Diversi obiettivi per BD4M:

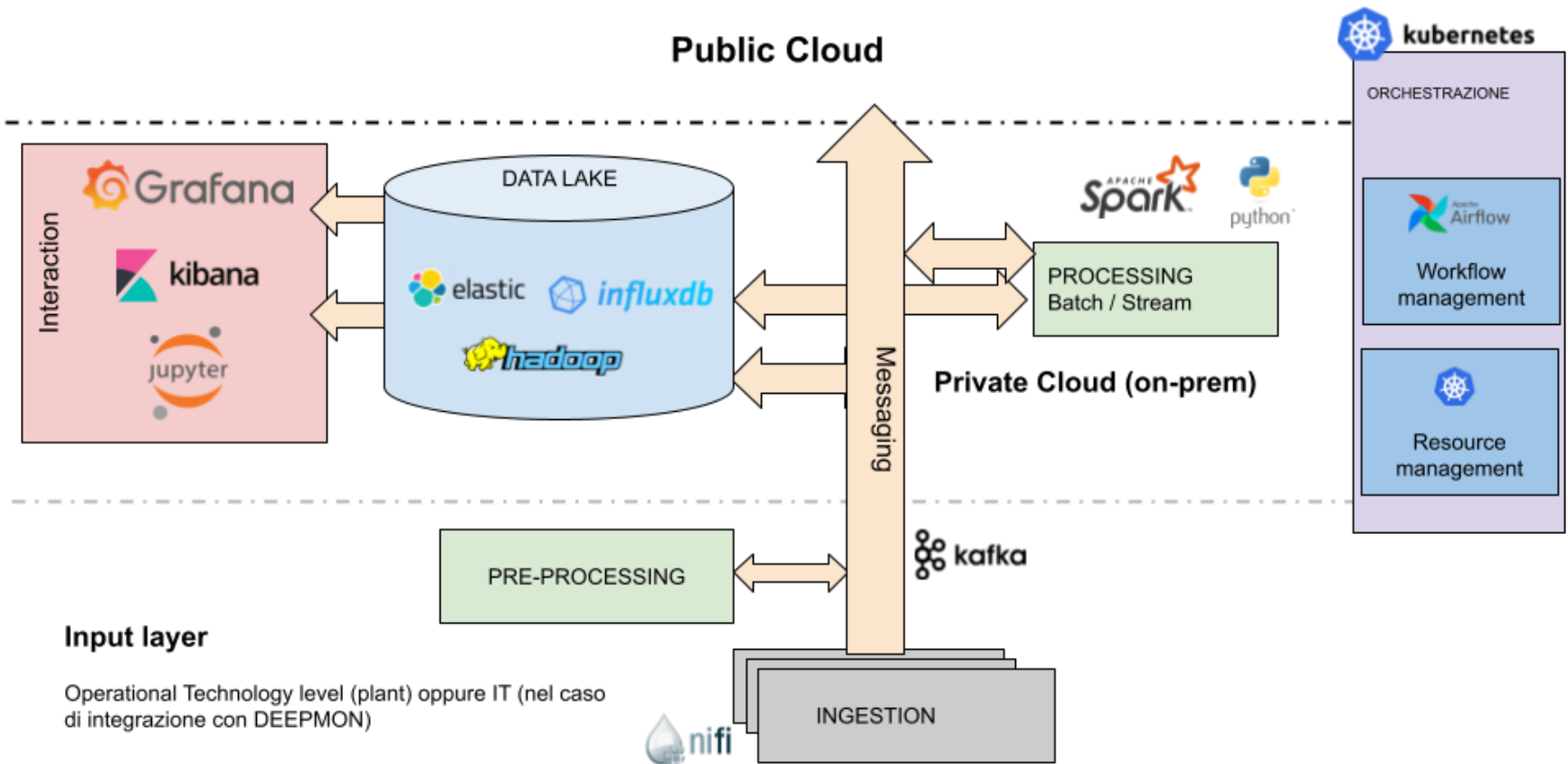
- Piattaforma modulare e installabile in praticamente qualsiasi ambiente computazionale (dall'edge al Cloud)
- Adozione tecnologie open source
- Supporto ad analisi a diversi contesti basate su soluzioni di Machine Learning (ML) avanzate
- Supporto a workload dinamici

Piattaforma BD4M: Concept



Piattaforma BD4M

Piattaforma BD4M: Scelte Tecnologiche



Alcuni aspetti fondamentali della piattaforma

Le tecnologie saranno discusse nell'intervento dell'Ing. Marco Govoni sugli use case industriali affrontati nel progetto

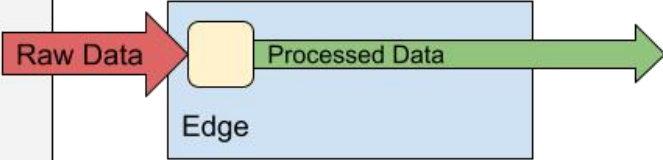
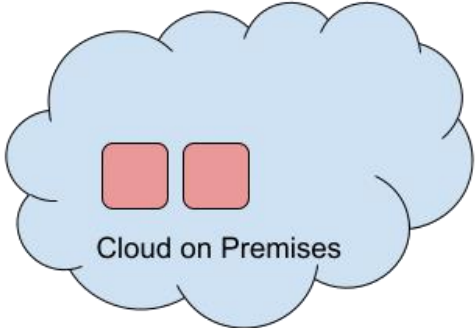
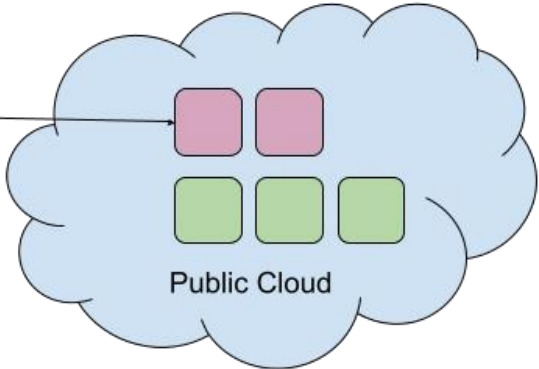
Nel seguito vorrei approfondire due aspetti fondamentali della piattaforma BD4M:

- 1. Orchestrazione e capacità di gestione workload dinamici**
- 2. Supporto ad applicazioni avanzate di machine learning**

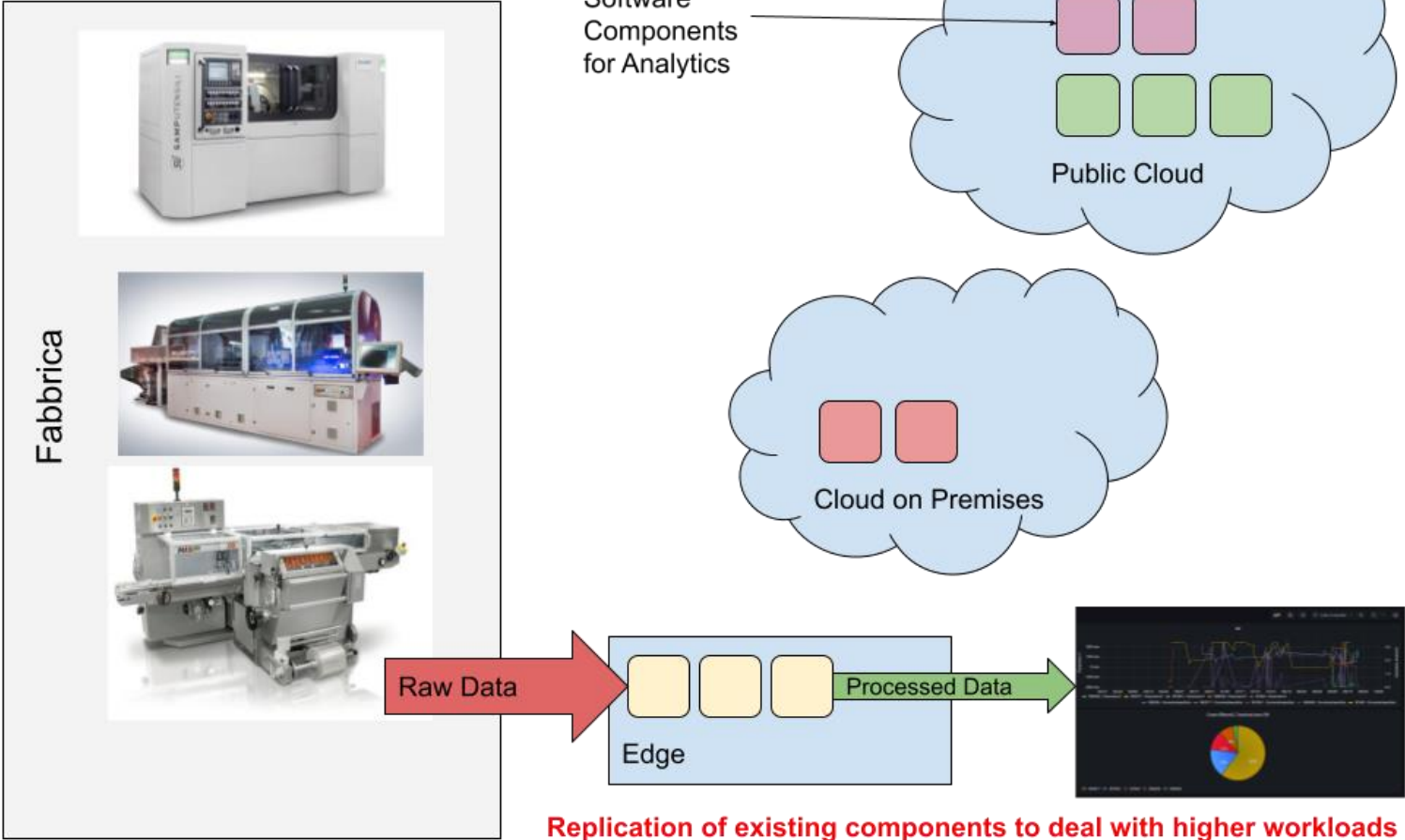
Analytics: Baseline



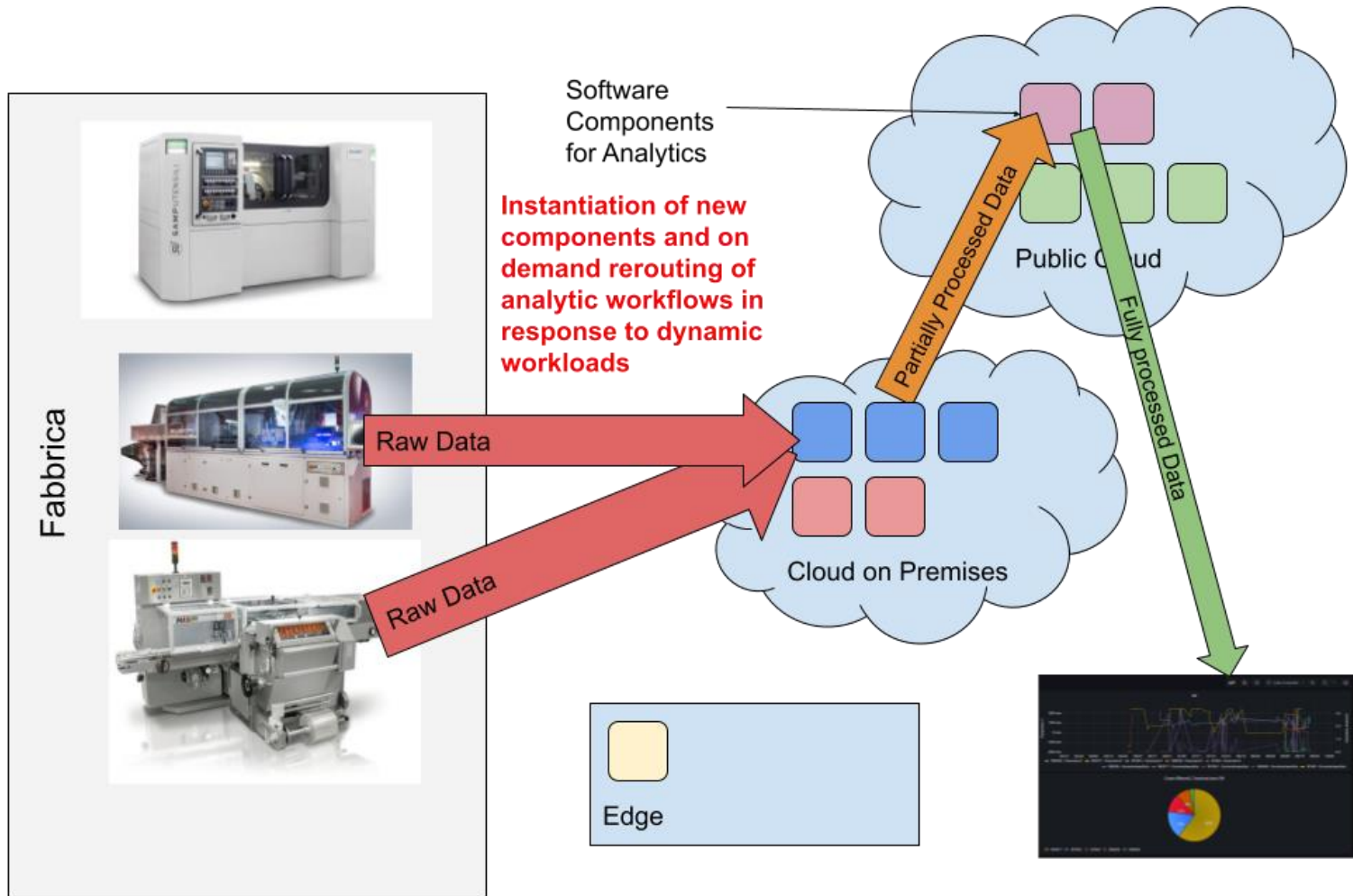
Software Components for Analytics



Analytics: Higher Workload



Analytics: Higher Workload



Orchestrazione

Diverse scelte di progetto importanti:

- Containerizzazione dei componenti della piattaforma
- Adozione come orchestratore di riferimento di Kubernetes
- Adozione di soluzioni di IT automation state-of-the-art, come Ansible e Helm
- Sperimentazione di soluzioni per dynamic analytics workflow management come Apache Airflow

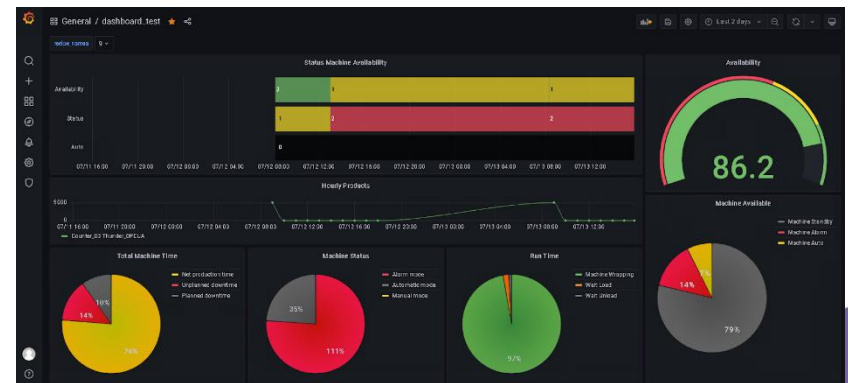
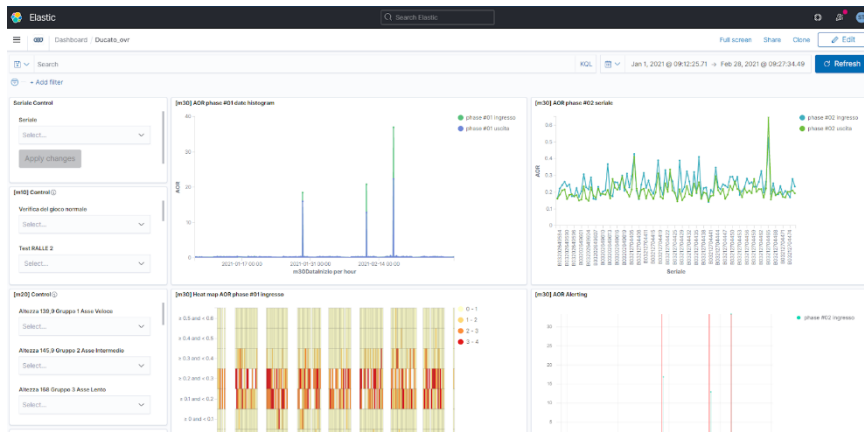
Questo rappresenta **uno degli aspetti più forward-looking della piattaforma BD4M**

Analytics: from Visual Dashboards...

BD4M adotta soluzioni per la realizzazione di dashboard in real-time customizzabili sulle esigenze dello use case

- Molto apprezzate dagli use case provider

Tuttavia uno degli use case industriali ci ha spinto ad approfondire le soluzioni di analytics adottate e a esplorare soluzioni avanzate di Machine Learning (ML)

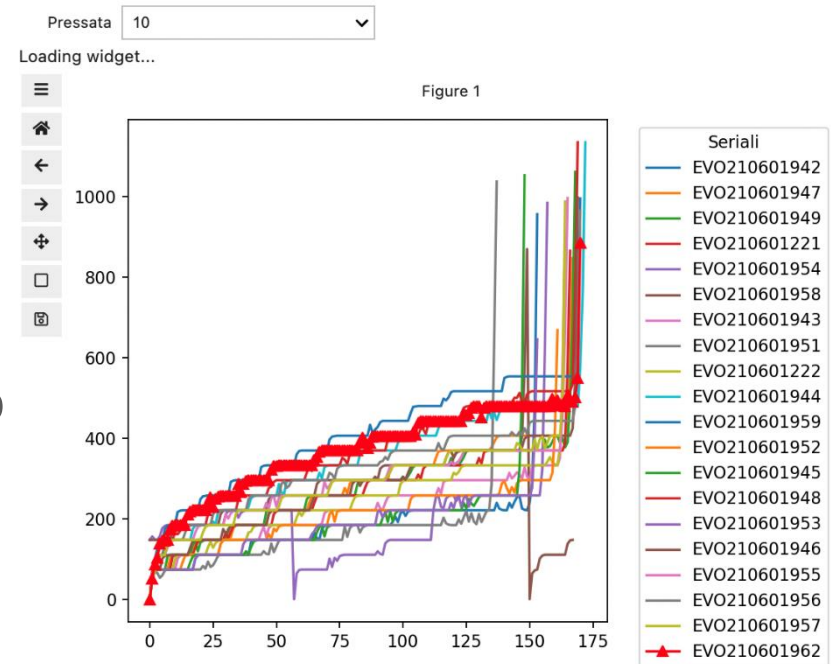


Analytics: ... to state-of-the-art ML

Nello use case Bonfiglioli Riduttori ci siamo scontrati con un problema di identificazione di componenti difettosi reso particolarmente difficile da diversi aspetti

- High dimensionality data
- Imbalanced data set
- Uncertainty

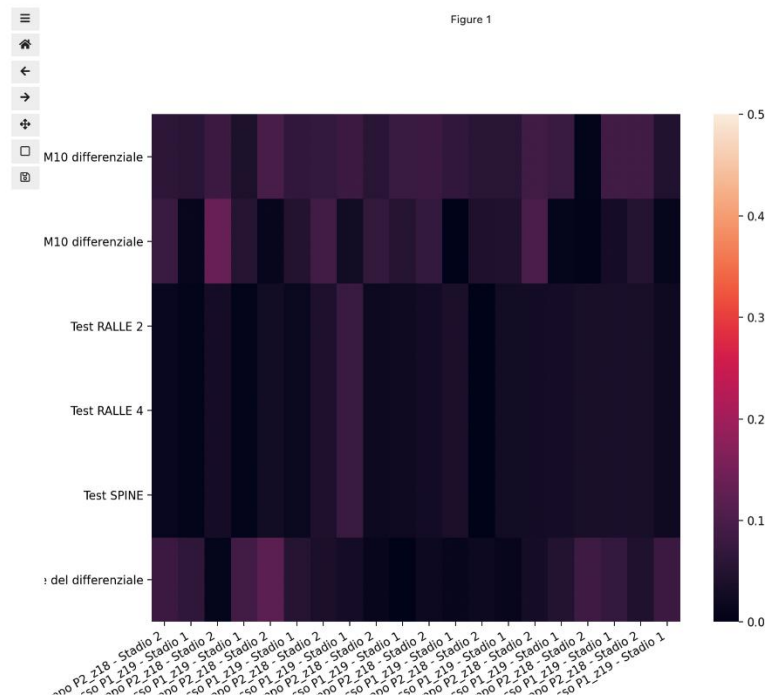
Per affrontare il problema abbiamo realizzato dashboard ad hoc



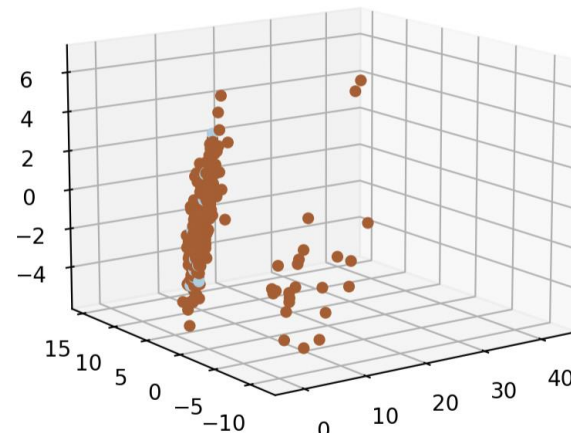
Linearity and curse of dimensionality

Analisi iniziali basate su soluzioni ML mainstream (PCA e logistic regression) hanno fornito informazioni molto importanti ma dato risultati deludenti dal punto di vista della classificazione

Per una migliore visualizzazione abbiamo suddiviso la heat map in 4 parti

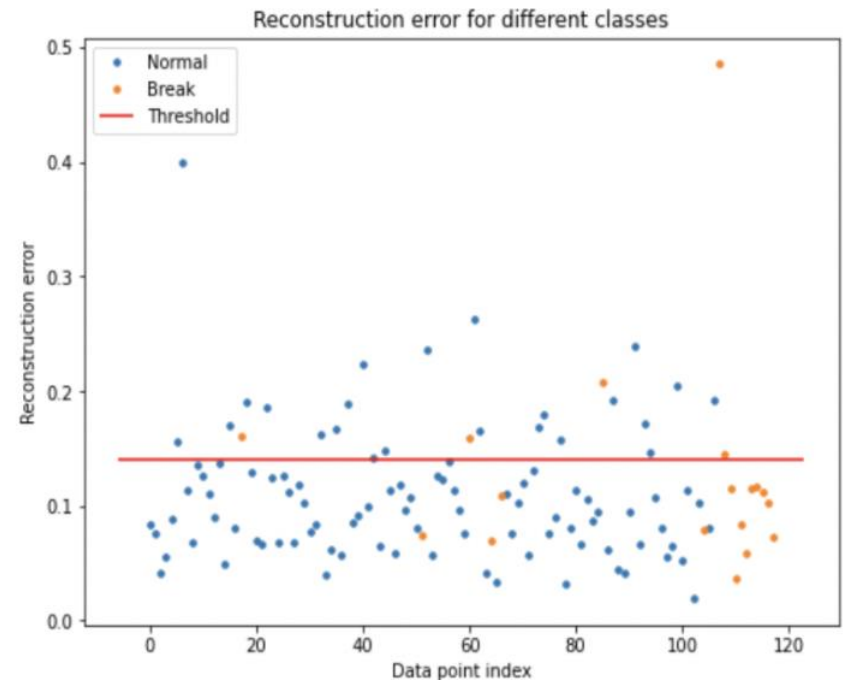
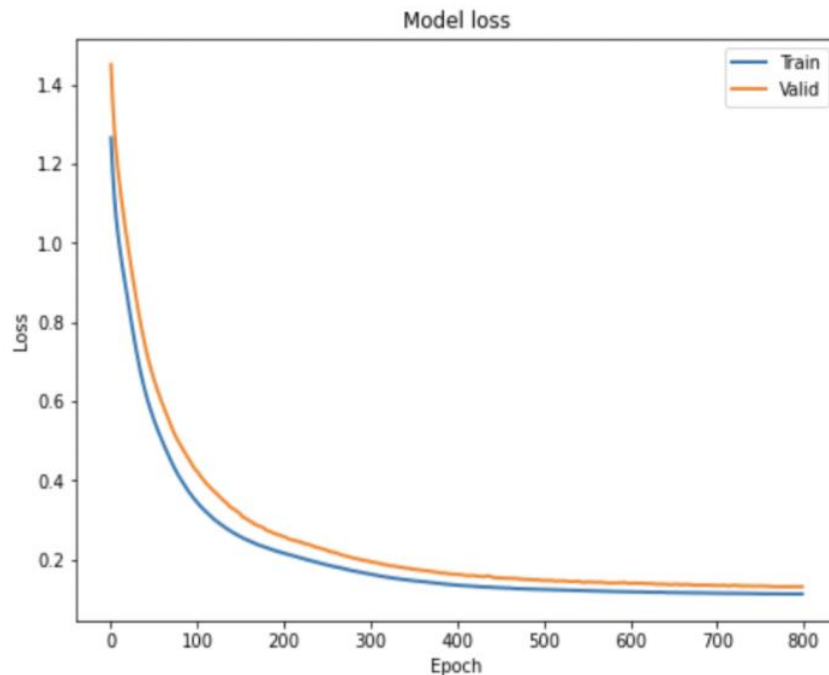


Separation of Observations using PCA



Semi-supervised learning w/ autoencoder

Abbiamo quindi deciso di esplorare **soluzioni avanzate e state-of-the-art di ML**, basate su reti neurali (in particolare, autoencoder) e semi-supervised learning, ottenendo eccellenti risultati



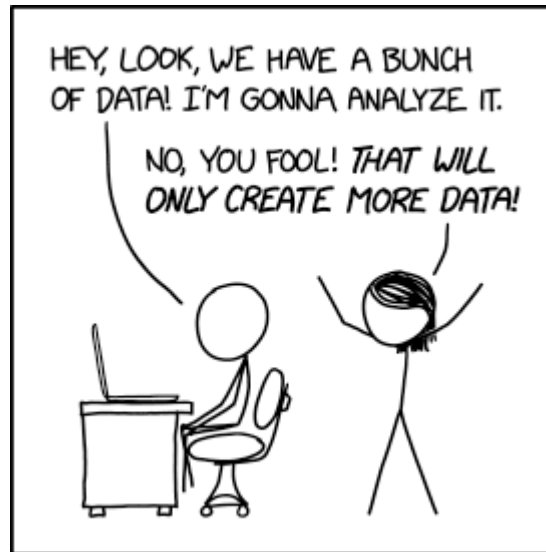
Risultati raggiunti

La piattaforma BD4M è stata realizzata e **validata in 4 use case industriali in condizioni di produzione** o simili

Supporto alla computazione dinamica rende la **piattaforma applicabile anche a use case addirittura più ambiziosi di quelli considerati nel progetto**

Piena soddisfazione per **maturità e potenzialità delle soluzioni tecnologiche adottate**

THANK YOU!



<https://xkcd.com/2582/>