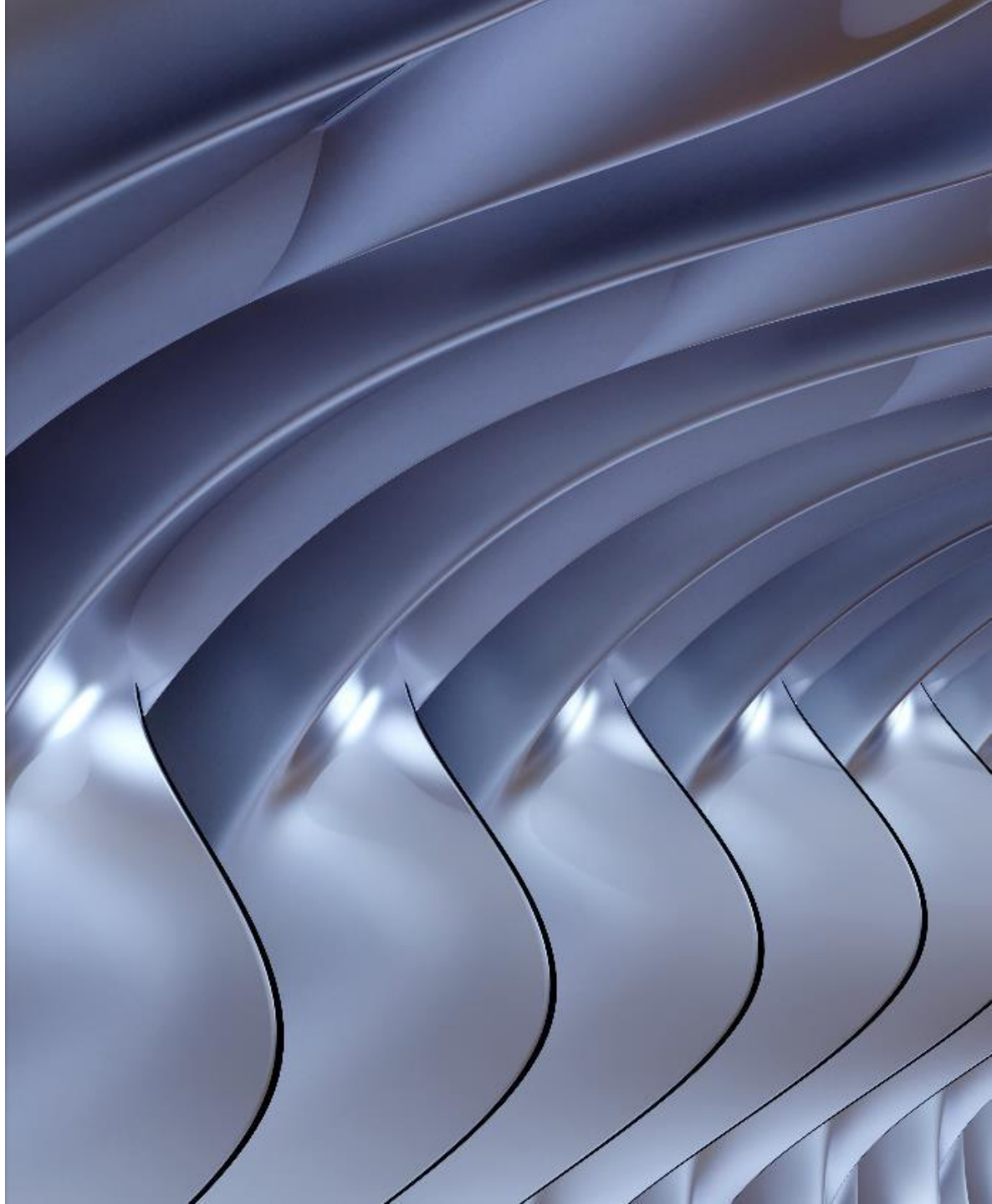
CONTROL IN STEEL

Big Data Technologies in Downstream Steel Process

Raffaella Grieco
raffaella.grieco@rina.org



CONTROL IN STEEL

NewTech4Steel European RFCS project

Cloud and Big Data concepts (approaches, architectures and methods) applied in steel industry in particular at the fast-rolling processes (Lambda Architecture)

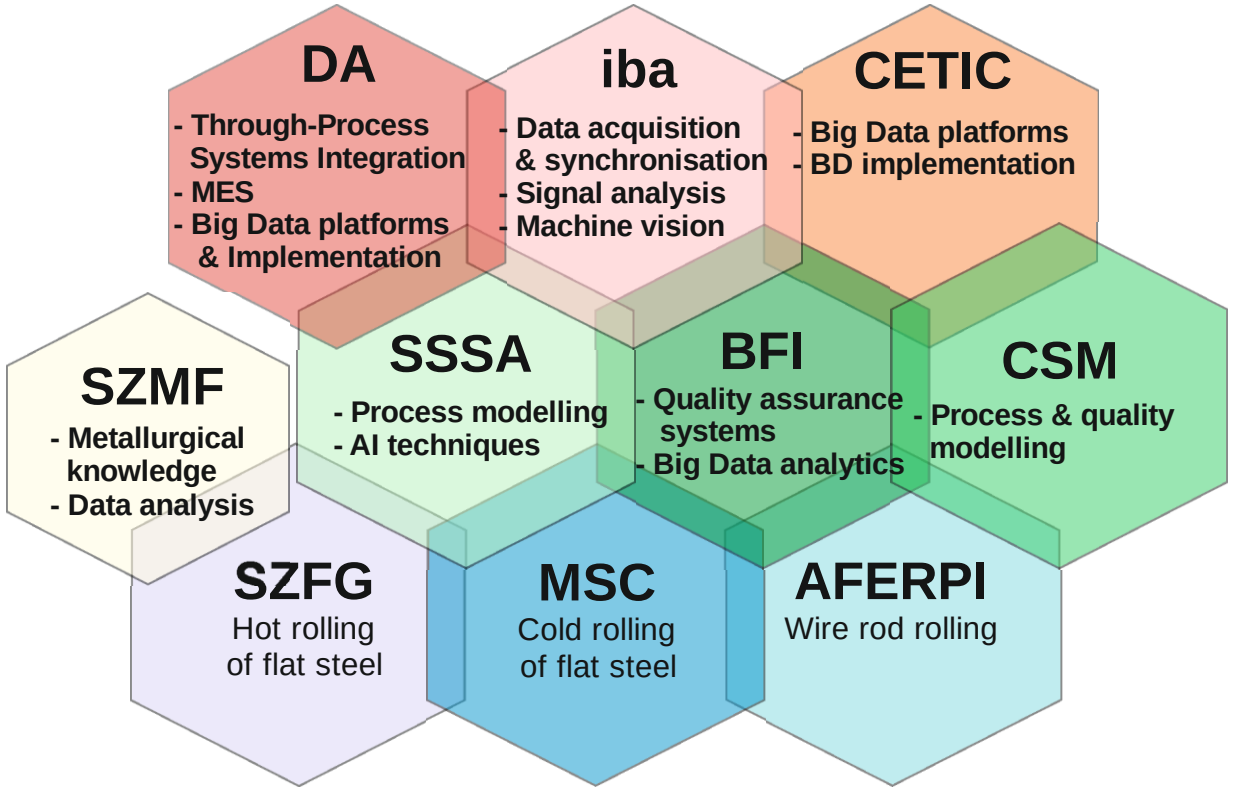
Project duration: 01.06.2018 – 31.05.2022

Providing and adapting infrastructures and architectures

Application of knowledge, methodologies and techniques to use cases

Industrial use cases

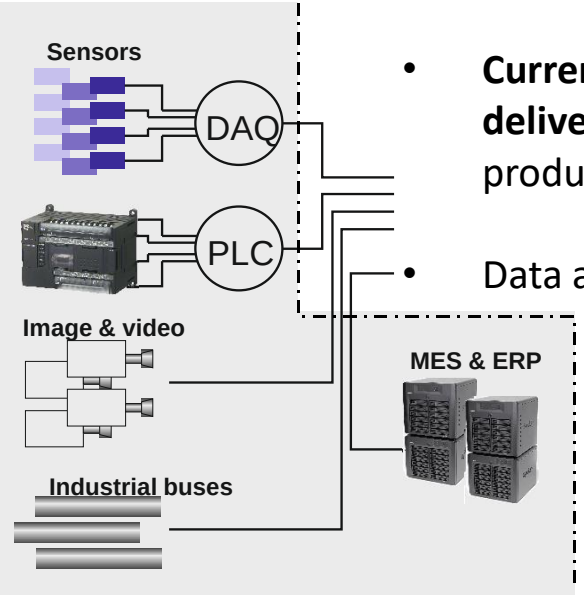
- **Hot-rolled wide strip**
- **Cold rolling area**
- **Wire rod rolling**



Motivation

There is the demand to **increase the stability of production processes** and **the quality of the products** in the production of rolled steel.

- 1D/2D geometry (sensor, PLC) event – 10Hz
- Surface inspection (image)
- Process data (PLC, sensor, bus) event – 500Hz
- …

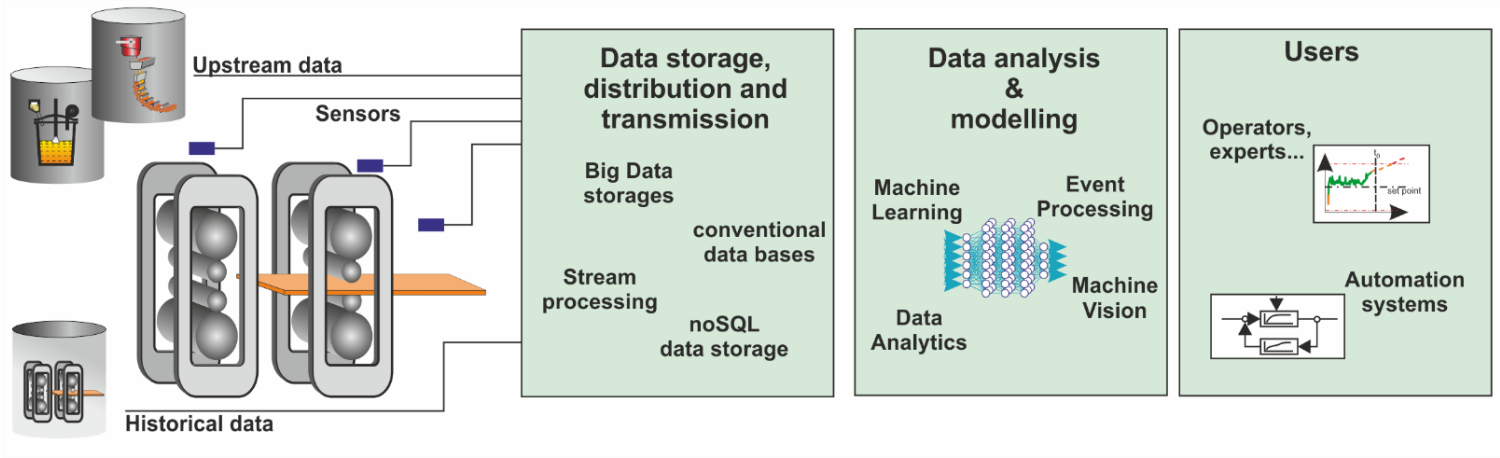


- **Current sensor equipment** at rolling facilities in steel production **delivers masses of data and information** about the process, the product and its quality
- Data are supplied **at high sample rates** and **high spatial resolutions**

Problem: process supervision **lacks** an **online or near-online exploitation** of the available data

Idea and Objective

Applicable methodologies to meet the requirements of online/near-online supervision of processes can be found in knowledge domains like **Data Analytics**, **Machine Vision** and in **Big Data techniques.**



- **To toughen up the data storage systems** at the industrial sites by enhancing the existing solutions and integrating innovative tools from Big Data technology.

- **To establish a methodologies for data transmission** to achieve highest possible data throughput rates, necessary for the aimed online or near online processing.

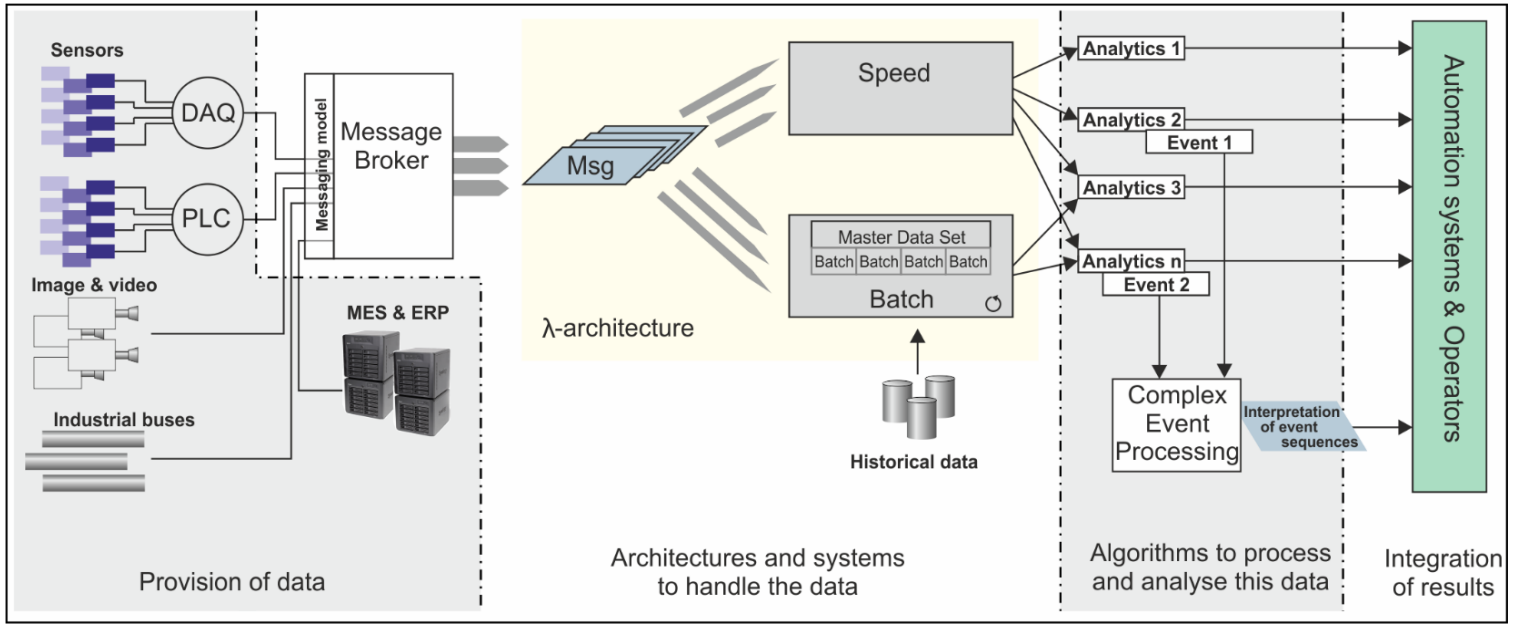
- **To develop and apply newest technologies of Data Analytics** for analysis and modelling, because the existing data acquisition systems supply masses of data of various structures and contents, which have to be analysed and processed together.

CONTROL IN STEEL

Technical approach

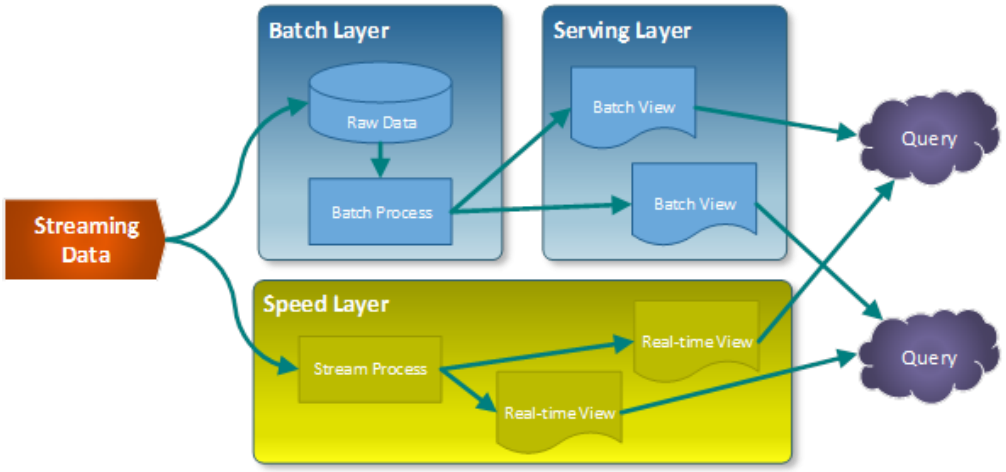
The technical solutions based on

- the state-of-the-art equipment of the investigated production facilities (**brown-field implementation**),
- the experts' knowledge about the underlying processes and their physical relations (**model definition**),
- the application of innovative methodologies of modern data processing (**Big Data methodologies**).



Lambda Architecture

The platform which merges the transmission of data for **online processing** (streaming) by a speed layer and for the **batch processing** where actual data are combined with those coming from historical data bases is the place where the Lambda architecture is applied.



According to **Marz**, the essential requirements of the LA:

- Fault-tolerant and robust enough to withstand code failures and human errors
- The layers must be horizontally scalable.
- Low latency in order to achieve real-time results
- Easy to maintain and debug

CONTROL IN STEEL

Cold Rolling Use Case

Aim:

Development of a platform software solution based on Lambda Architecture to strengthen and optimize the processes of coils galvanization

a. **Forecast of top defects at HDG line due to flatness problem**
b. **Prediction of specific setup parameters to avoid or limit process problems and product defects during the galvanization**

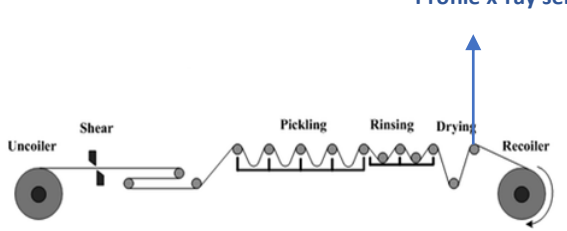
Platform solution requirement:

- **Ingestion** and elaboration of high frequency variables (process parameters at Galvanizing line)
- **Storing** of length-based data (coil profile, mechanical charactetistics, stressometer matrix at CRM for each pass, Zinc profile top/bottom), time series and large data per coil (images) in datalake
- Heavy **data processing** for data preparation in the training cycle of ML
- Implementation, deployment and update of **ML models**
- Storage of results on relational-db
- HMI accessible via web

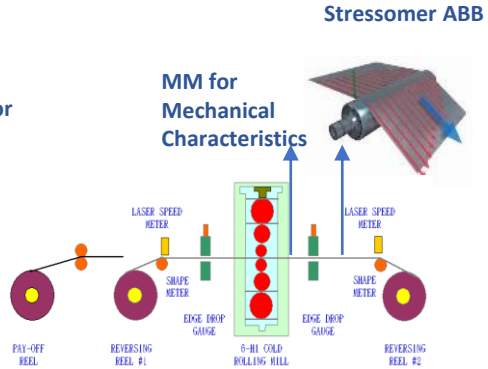
CONTROL IN STEEL

Plant Layout

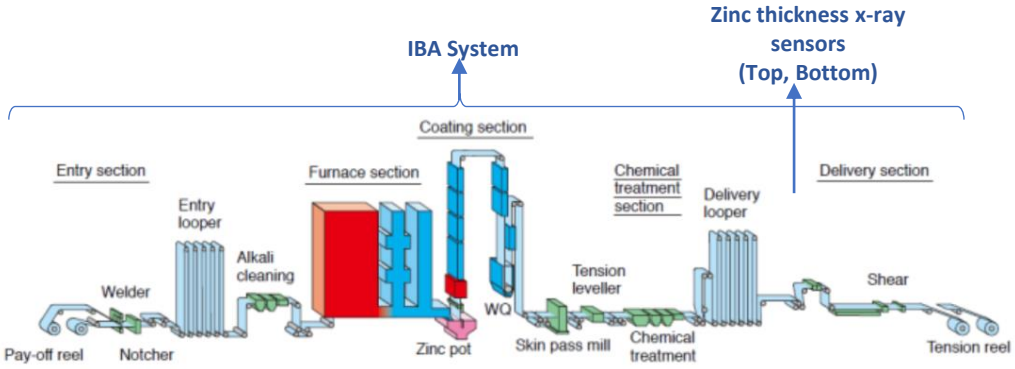
Sensors and additional equipments position of the examined use case



Pickling line



CRM



HDG line

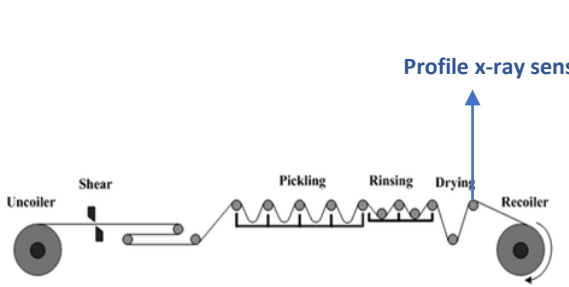
Identified Data Source

- **Pickling data**: Thickness along coil length (old system)
  - stored in files on server file system
- **Stressometer data at CRM for each pass**: Tension stress along coil length
  - stored in files on server file system
- **Mechanical characteristics**: Values along coil length
  - stored in files in proprietary format on server file system
- **Galvanization depth**: Zinc thickness on both sides
  - stored in files on server file system
- **Process parameters at HDG line from IBA System**: a set of process parameters
  - exported file from IBA QDR stored on server filesystem/streamed to Apache KafKa
- **Production Data**: Black coil data (dimension, mechanical characteristics…), production data, quality report, laboratory tests
  - per coil
  - stored on Sql database

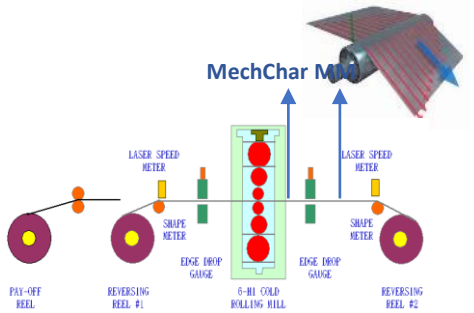
Methodological approach

Development of a platform to strengthen and optimize the processes of coils galvanization

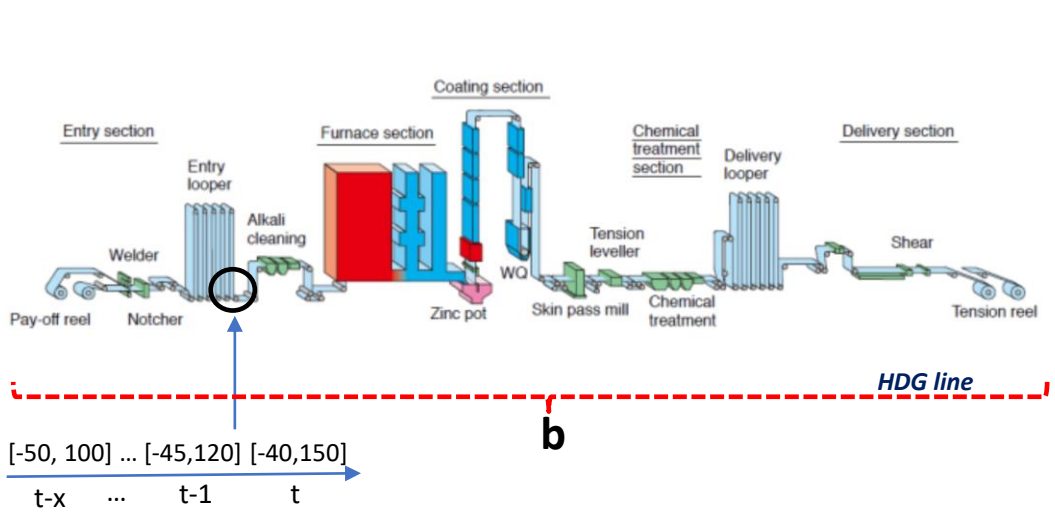
a. **Prediction of top defects at HDG line due to flatness problem**
b. **Prediction of specific parameters setup to avoid or limit problems during the galvanization**



Pickling line



CRM



Machine Learning/Process Parameter forecast

✓ **Forecast of specific process parameter**
- **centering device position**
- **guide position**

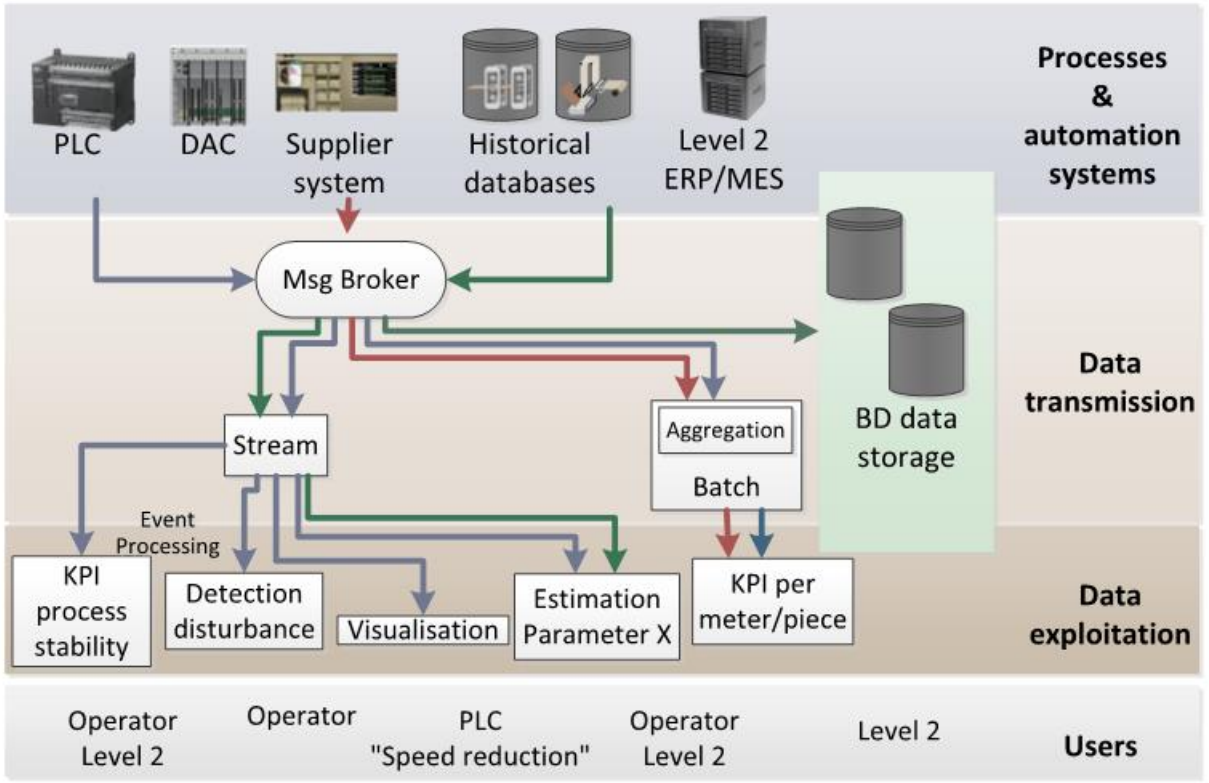
Online monitoring of guide/centering device

Ci
Ci+1
Ci+2
...
Cn

List of next coil to process (Coil Classified by means of feature at point **a** in lambda batch layer)

CONTROL IN STEEL

Suitable tools/platforms for BD framework



Mechanical Char MM
Scanning Profile x-ray sensors,
Stressometer ABB,
IBA Pda,
Production Data

Nifi
Kafka

Spark
Spark Streaming

Tensor Flow
C++

php/HTML

The main modules involves:

Collection/Ingestion: collect, preprocess, validate and format the raw data, stores into the Data Lake and sends to the broker

- Apache NIFI
- IBA to Kafka

Storage:

Data Lake: to gather all data coming from the ingestion layer into one system to perform data analysis and integration

- HDFS

Database: to store current calculated and historical data in one single place

- MySql

Analysis:

Offline: Spark, C++, TensorFlow

Data Transmission: Kafka

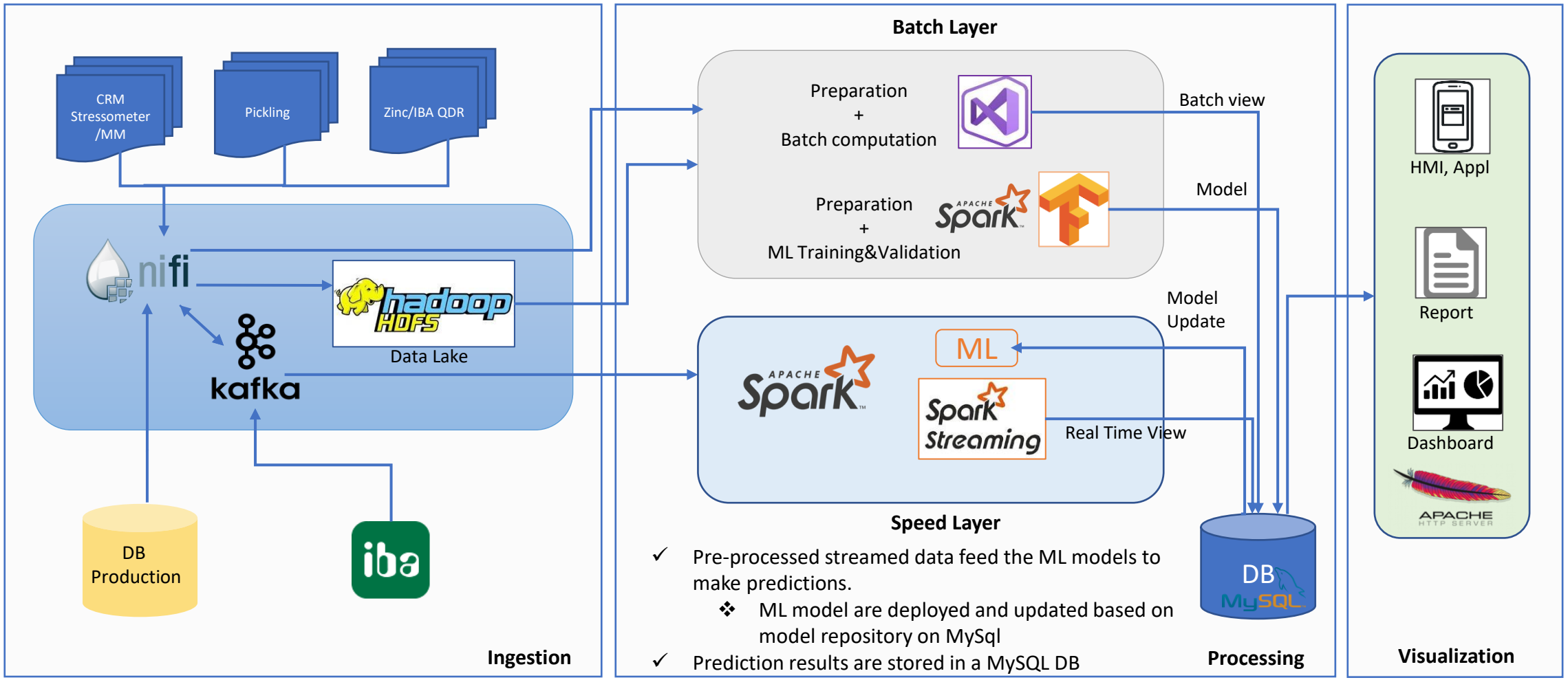
DataProcessing: NIFI/C++(batch), Spark Streaming (realtime)

Data Visualization: php/HTML

CONTROL IN STEEL

Project Lambda Architecture Schema

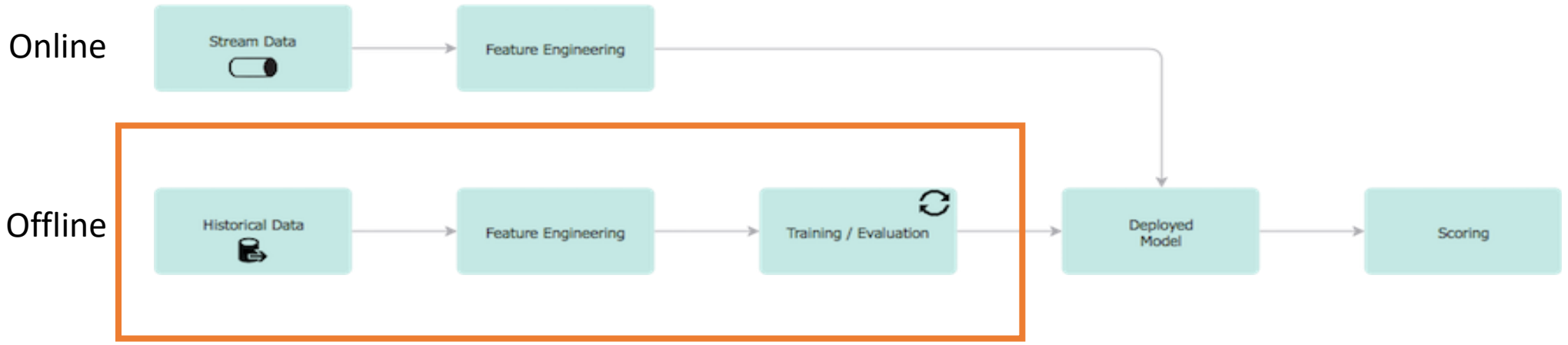
- ✓ Model are created and retrained
  - ❖ The result (as code format or parameters values is stored in the dedicated MySQL server)
- ✓ Batch view on current production are loaded on MySQL server



- ✓ Pre-processed streamed data feed the ML models to make predictions.
  - ❖ ML model are deployed and updated based on model repository on MySql
- ✓ Prediction results are stored in a MySQL DB

CONTROL IN STEEL

Data Analytics for analysis and modelling

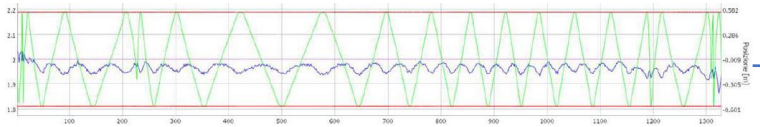


Data preparation for analysis and modelling

Pickling line

Pickling

Profile



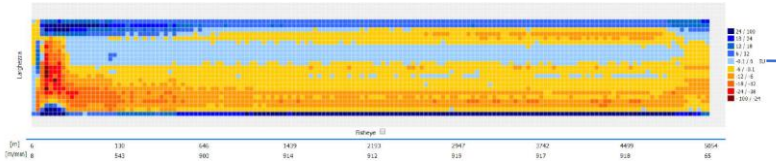
Lpos

```
4.525    1.076    -0.096
4.51     2.091    -0.22
4.498    3.173    -0.344
4.482    4.232    -0.489
4.474    5.246    -0.281
4.501    6.28      0.194
4.417    7.296     0.579
4.439    8.37      0.423
4.472    9.535     0.278
4.466   10.601     0.175
4.474   11.748     0.072
4.477   12.768    -0.011
4.491   13.896    -0.094
```

CRM

CRM Stressometer/ MM

Stressometer tensions



Width sections

Lpos

```
125.69  125.69  125.69  125.69  125.69      125.69  125.69  125.69   6.558
125.75  125.75  125.75  125.75  125.75      125.75  125.75  125.75   6.558
125.58  125.58  125.58  125.58  125.58      125.58  125.58  125.58   6.559
125.80  125.80  125.80  125.80  125.80 ...  125.80  125.80  125.80   6.564
125.76  125.76  125.76  125.76  125.76      125.76  125.76  125.76   6.579
125.88  125.88  125.88  125.88  125.88      125.88  125.88  125.88   6.612
125.87  125.87  125.87  125.87  125.87      125.87  125.87  125.87   6.658
```

Mechanical characteristics

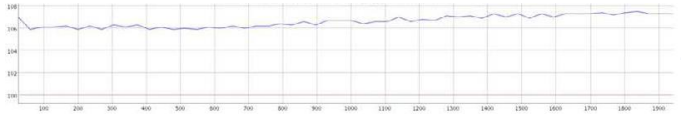
Lpos

```
[[Position]]    [[O100]]        [[O101]]       [[O102]]
79      216.177 272.089 100.000 1.846   22.651  0.004
83      238.930 307.547 100.000 2.092   16.960  0.004
88      288.526 390.699 100.000 2.366   10.135  0.004
93      260.453 342.601 100.000 1.751   13.343  0.005
97      245.189 317.585 100.000 1.643   15.770  0.005
102     239.763 308.876 100.000 1.654   16.794  0.005
106     245.284 317.738 100.000 1.705   15.753  0.005
111     252.121 328.851 100.000 1.733   14.592  0.005
116     253.004 330.296 100.000 1.762   14.452  0.005
```

HDG line

Zinc/IBA QDR

Zinc Profile Top/Bottom



Width sections

Lpos

```
LPos; Zone 01;    02;    03;    04;    05;    06;

 35.9;     0.0;   0.0;   0.0;   0.0;   0.0;  83.0;
 93.8;     0.0;   0.0;   0.0;   0.0;   0.0;  84.5;
134.3;     0.0;   0.0;   0.0;   0.0;   0.0;  83.4;  ...
173.5;     0.0;   0.0;   0.0;   0.0;   0.0;  73.3;
212.3;     0.0;   0.0;   0.0;   0.0;   0.0;  95.4;
251.3;     0.0;   0.0;   0.0;   0.0;   0.0;  92.6;
```

Iba QDR process parameters export

Lpos

```
"0,00","201,44","180,96","37,46","-74,00","0,30","-295,00","0,30","-377,00","4,90","14,30","1,40","
"1,00","200,72","181,05","37,46","-74,00","-0,10","-295,00","0,30","-427,00","11,20","14,30","5,30"
"2,00","200,87","180,92","37,48","-74,00","-0,43","-294,00","0,30","-507,00","11,10","13,90","9,70"
"3,00","200,78","180,99","37,50","-74,00","-0,50","-293,00","0,50","-534,38","10,63","14,56","10,76
"4,00","200,78","180,99","37,50","-74,00","-0,50","-293,00","0,50","-534,38","10,63","14,56","10,76
"5,00","200,87","180,99","37,53","-74,00","-1,03","-292,00","0,87","-670,00","7,40","17,40","15,50"
"6,00","201,24","181,07","37,54","-74,00","-0,60","-291,53","1,00","-719,00","5,90","20,50","19,10"
"7,00","201,36","181,03","37,56","-74,00","-0,70","-287,86","1,00","-758,00","5,10","23,20","21,20"
"8,00","200,64","180,96","37,58","-74,00","-0,50","-281,51","1,20","-790,00","4,50","25,10","23,90"
```

DB Production

- Production report
- Scrap report (head/tail cutting at the entry of the HDG line)
- Elongations report

Linea	Rotolo	Rotolo_Padre	Ora Fine	Difetto1 (Descr_)	Difetto2 (Descr_)	Difetto3 (Descr_)	Note	Spess Rea	Larghezza	Spess Rea	Larghezza
Zincatura3	20RW008705	20RX003909#	3/24/2020 6:06					0,42	1253	0,42	1253
Zincatura3	20RW008706	20RX003909#	3/24/2020 6:15					0,42	1253	0,42	1253
Zincatura3	20RW008707	20RX003909#20RX003910#	3/24/2020 6:23					0,42	1254	0,42	1254
Zincatura3	20RW008708	20RX003910#	3/24/2020 6:32				lieve bozzetta a circa 20cm lato o	0,42	1254	0,42	1254
Zincatura3	20RW008709	20RX003910#	3/24/2020 6:40				lieve bozzetta a circa 20cm lato o	0,42	1254	0,42	1254
Zincatura3	20RW008710	20RX003910#	3/24/2020 6:46				lieve bozzetta a circa 20cm lato o	0,42	1254	0,42	1254
Zincatura3	20RW008711	20RX003910#	3/24/2020 6:51				lieve bozzetta a circa 20cm lato o	0,42	1254	0,42	1254
Zincatura3	20RW008712	19RF137230#	3/24/2020 7:09				a tratti lievi macchie d ossido di	0,49	1255	0,49	1255

Data preparation, analysis and modelling

Preparation of dataset based on historical data for the training/test of Machine Learning models involves:

- Rescale/Transform
  - Stressometer tension data are transformed to obtain IUnit values
- Data Alignment
  - Data on different sampling on length position are interpolated to obtain the maximum sample rate
  - Data sampled on different width section are interpolated based on the Stressometer sample rate
  - At HDG line coils are head/tail cut, a procedure realign the correct position with the IBA QDR and Zinc profile data based on two reported data:
    - the elongation at pickling line
    - the weight of the cut material at HDG line
- Missing values handling & Filter on noise
  - Row with missing values are dropped
- Outliers detection and treatment
  - CRM Stressometer external sections can contain suspicious values, they are excluded from the case

Analysis for feature/variable extraction:

- Feature engineering & Dimensionality reduction

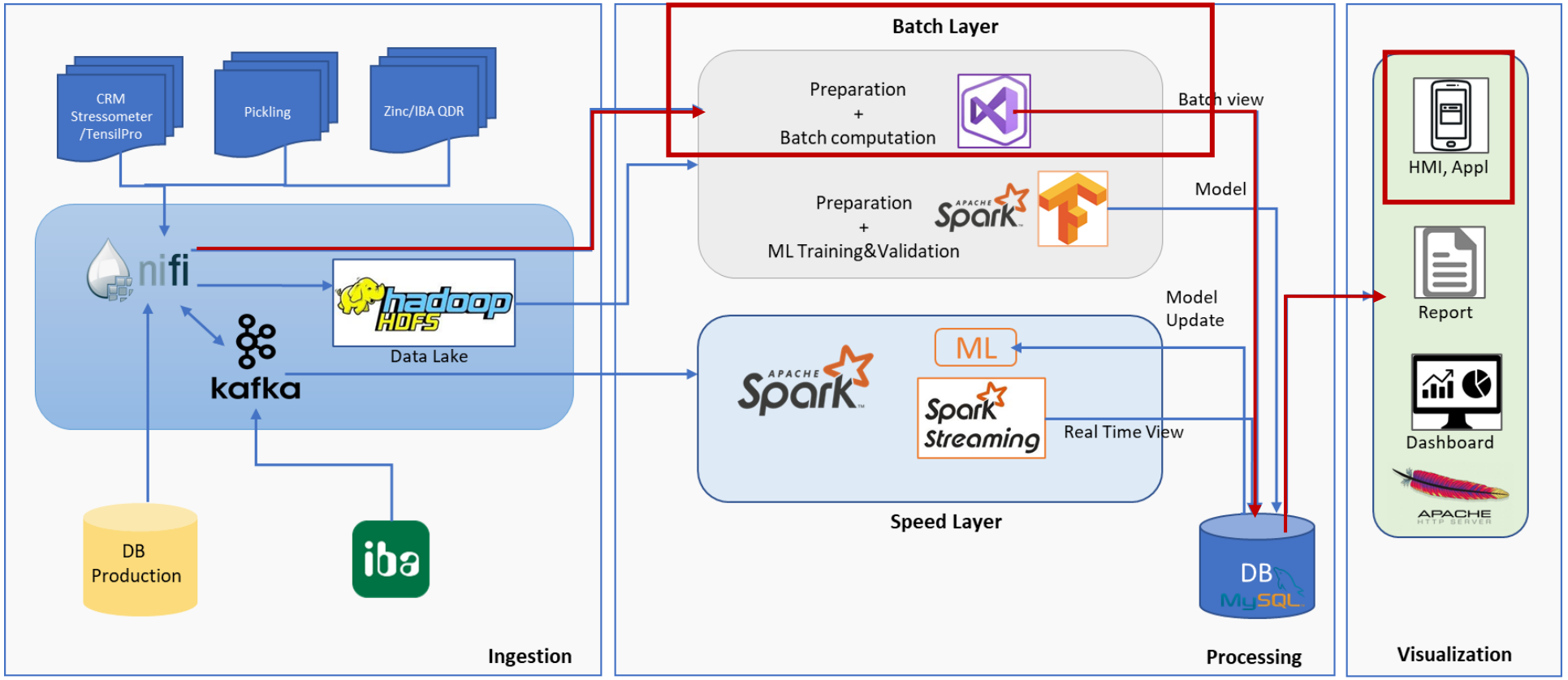
  Correlations, PCA

Modelling

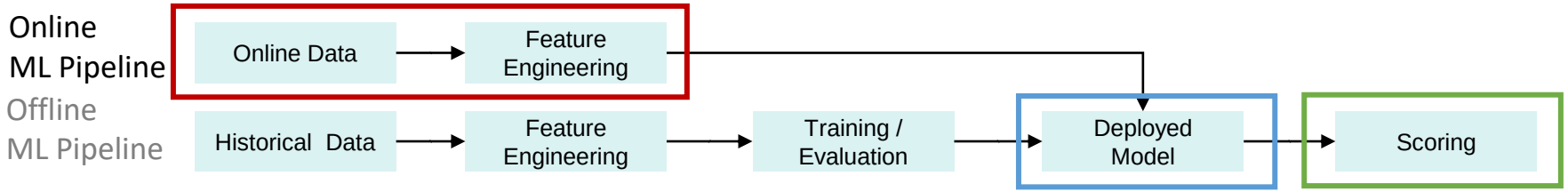
- Data Segregation
  - Split subsets of data to train the model and further validate how it performs against new data
- Training
  - Different schemes of Back Propagation Neural Network

CONTROL IN STEEL

Batch Layer Deploy



ML flow of batch process (a)



CRM Stressometer/ MM

- **Rescale/Transform**
  - Stressometer tension to IUnit values
- **Data Alignment**
  - Interpolation on length/width

DB Production

→Profile

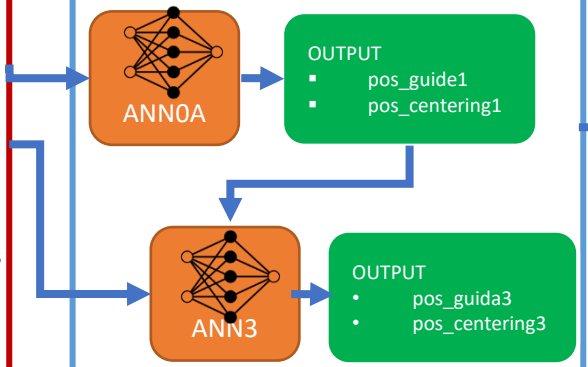
- Crown
- Wedge
- Build_up
- Thickness

→Line, Coil_Qlt_In, Coil_Supplier

- Vel_Process_Before_Acc Min/Max
- Vel_Process_HF_After_Acc_Min/Max
- Perc_Acc_Min/Max
- Vel_Process_Before_Acc_SET/Vel_Process_HF_After_Acc_SET/ Perc_Acc_SET

List of configuration for setup parameters

50_150_170
63_150_170
77_150_170
90_150_170
80_150_170



❖ Calculation of index on each position of coil length based on the value of guide and centering device predicted by ML model

Index	Position	Guide	Note
1	>=-20 <=20	>=-80 <=80	**Position**: Centered **Guide**: Low Correction
2	>=-20 <=20	>-200 <-80; >80 <200	**Position**: Centered **Guide**: Medium Correction
3	>=-50 <-20; >20 <=50	<=-80 >=80	**Position**: Slight Heel **Guide**: Low Correction
4	>=-50 <-20; >20 <=50	>-200 <-80; >80 <200	**Position**: Slight Heel **Guide**: Medium Correction
5	>=-180 <-50; >50 <=180	<=-200 >=200	**Position**: Severe Heel **Guide**: Max Correction

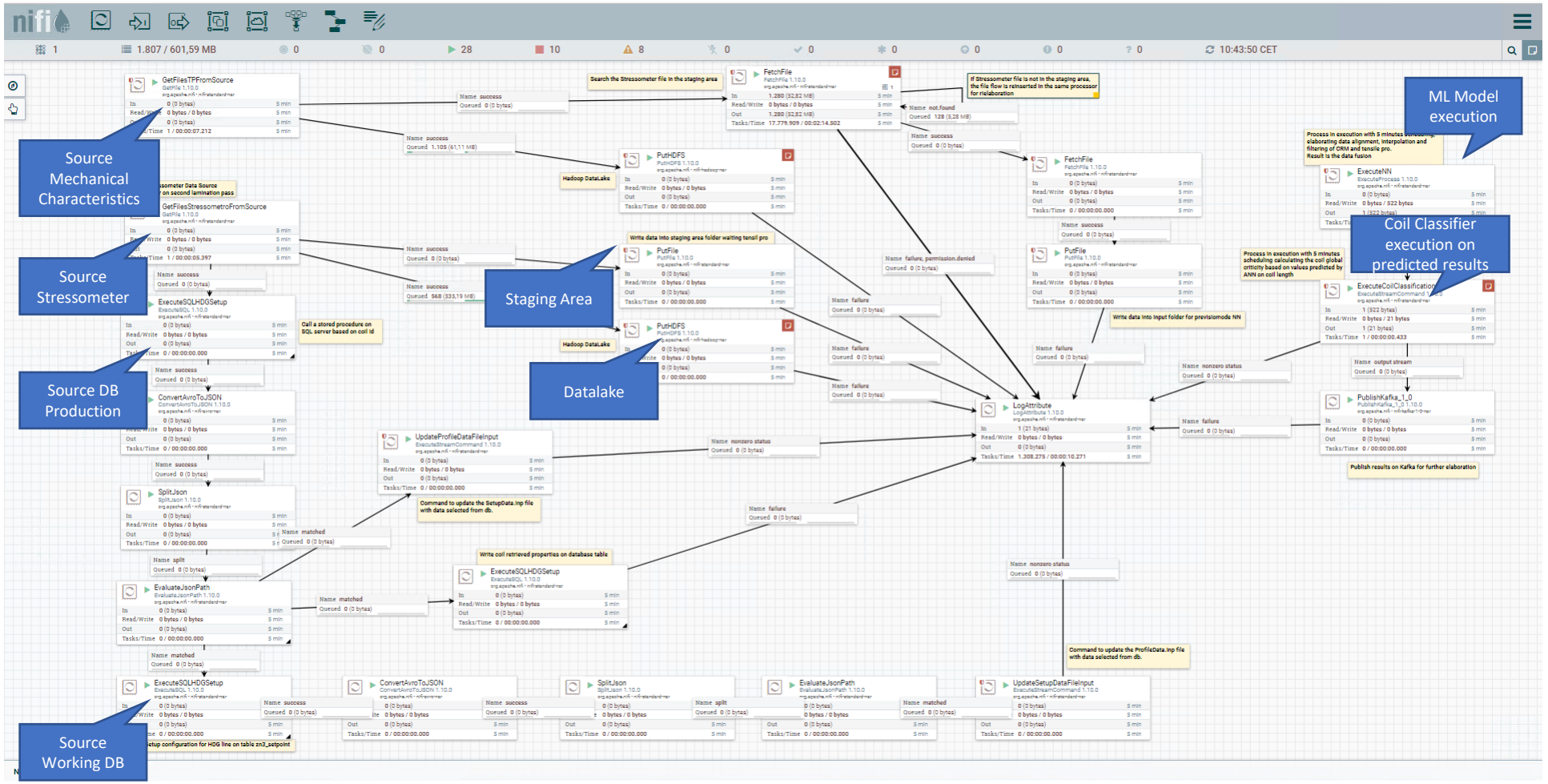
❖ Calculation of global coil criticality (red, yellow, green) based on the percentages of positional indexes

Criticality classification logic			
Index	% Coil Length		
1	0-100		
2	0-100		
3	80 -100	50 - 79.9	0-49.9
4	15 -100	2 -14.9	0-1.9
5	5 -100	1 - 4.9	0-0.9
0	0-100		

CONTROL IN STEEL

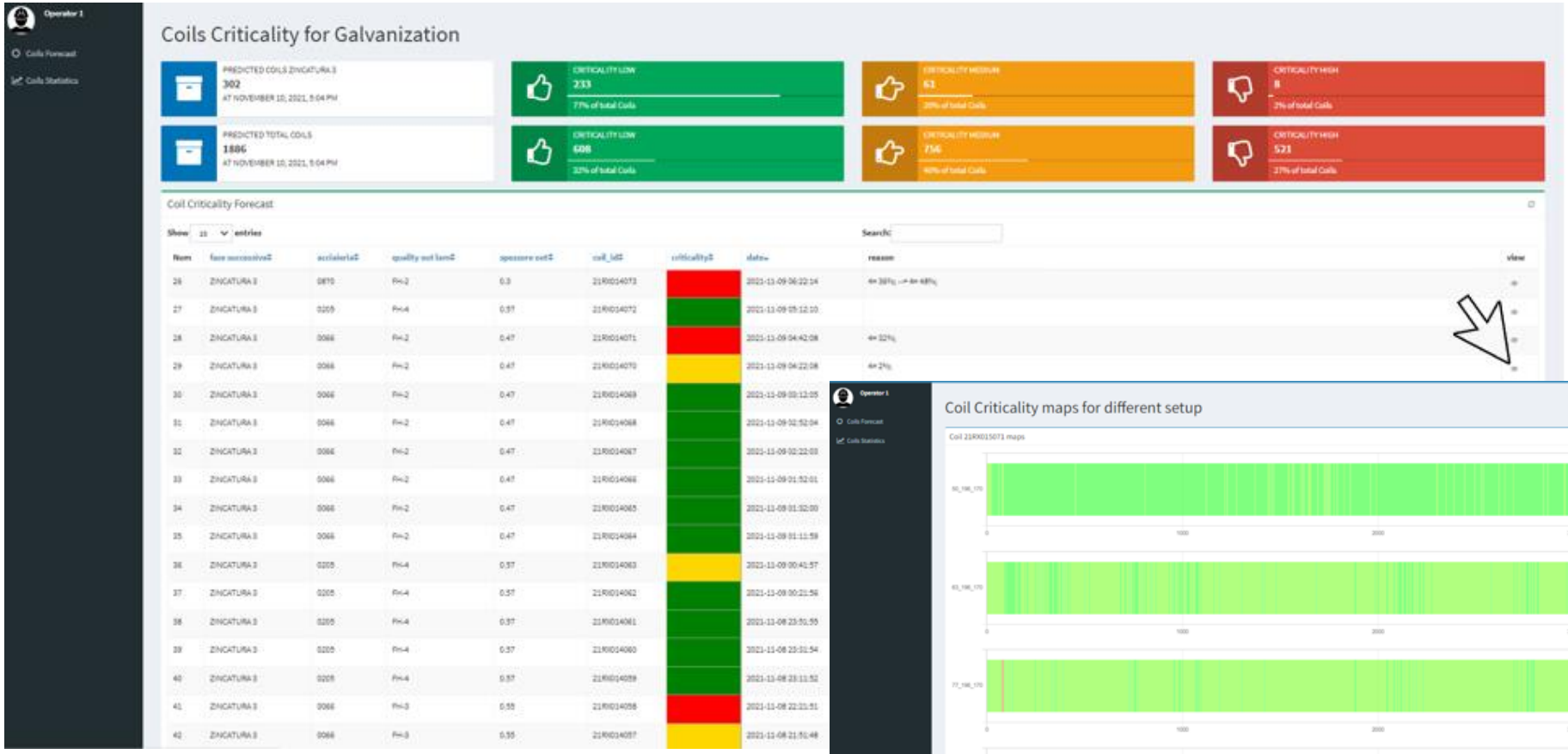
Flow implementation of batch process (a)

Nifi execution flow with processors interactions starting from data sources until the ML model execution



CONTROL IN STEEL

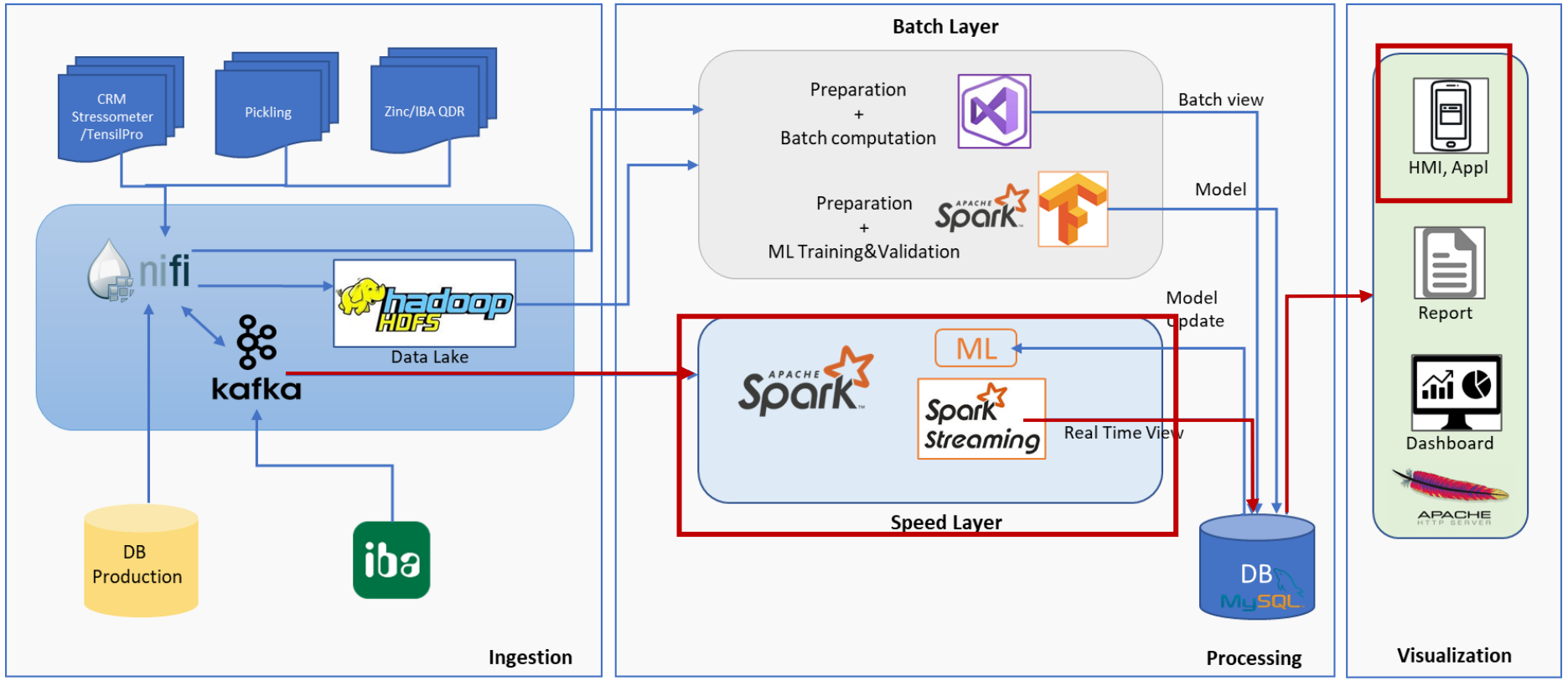
Visualization of batch process (a)



Web Interface based on Apache Http Server

CONTROL IN STEEL

Speed Layer Deploy

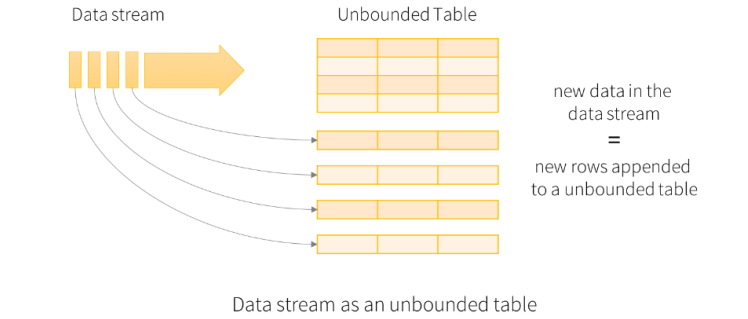


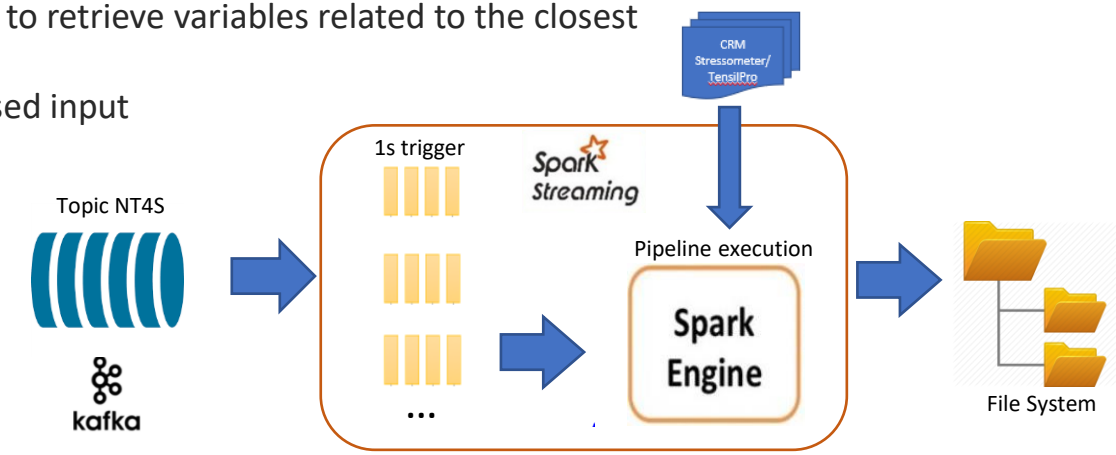
Implementation details of real time process (b)

- **Spark Structured Streaming** (introduced with Spark 2.x)
  - Model of streaming built on the Spark SQL library, based on Dataframe and Dataset APIs (Since Spark 2.0 Dataframe and Dataset can represent static, bounded data, as well as streaming, unbounded data)
  - Easily apply any SQL query
  - Spark SQL engine will take care of streaming, running in incrementally and continuously way and updating the final result as streaming data continues to arrive
- **Using ForeachBatch** (Operations that allow you to apply arbitrary operations and writing logic on the output of micro-batch)
  - **Pyspark ML pipeline**

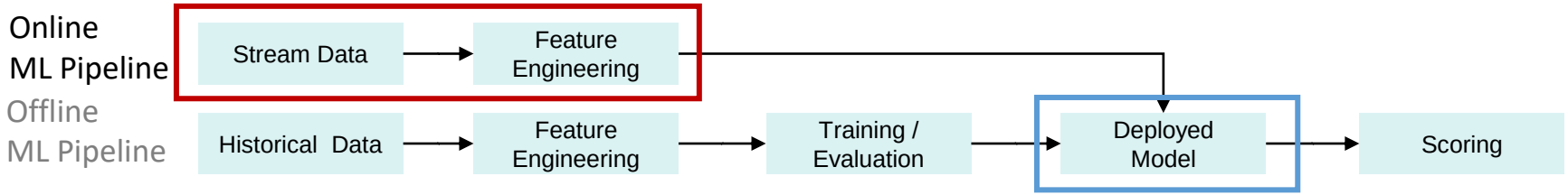
    Foreach micro-batch received by topic
    - Get CoilId and entering length
    - Access Mechanical Characteristics, CRM file to retrieve variables related to the closest position to the entering length
    - Call the ANN procedure passing the composed input
    - Write results on file system

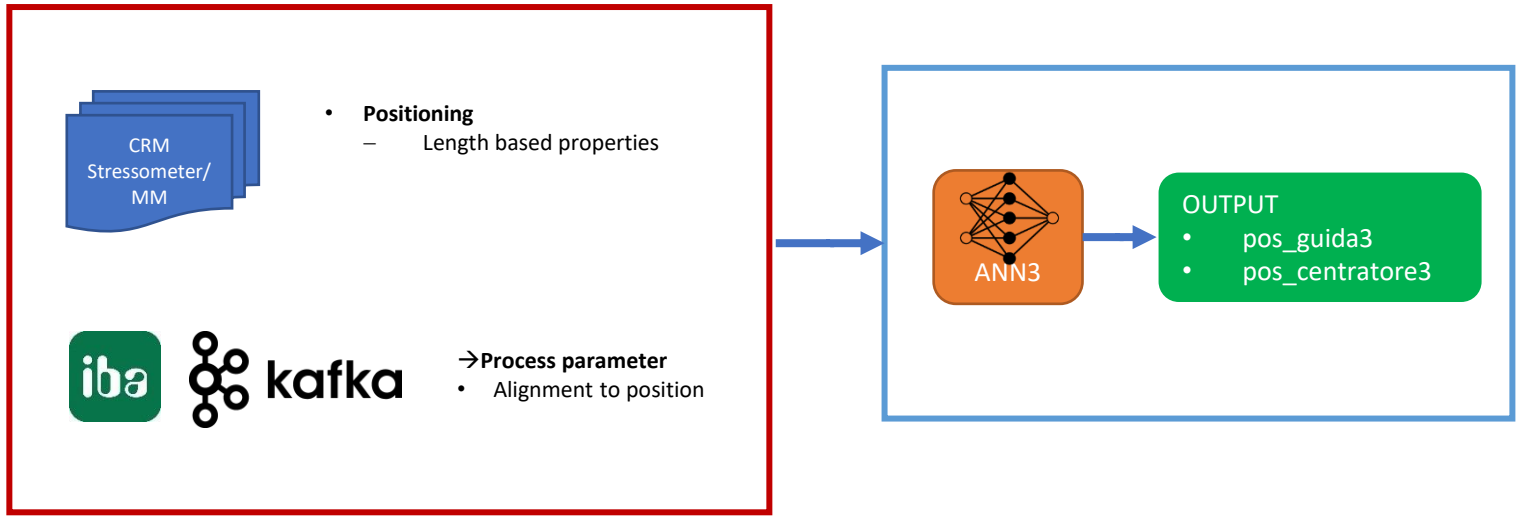
Structured Streaming Programming Guide - Spark 3.2.0 Documentation (apache.org)





ML Flow of real time process (b)





Summary of results

- ✓ Survey about necessary/available data was generated including information about location, origin, content and amount
- ✓ Definition of requirements about integration of new tools and architectures into the existing plant IT systems.
- ✓ Definition of HMI under different aspects (security, types of users, visualizations etc)
- ✓ Benchmarks and selection of tools and platforms for data storage, transfer and analysis
- ✓ Setup of IT system with necessary measures (additional computation server, expanded storage, elongation of historical data retention, proprietary acquisition system)
- ✓ Installation and tests for assessing tools and platforms for data handling, storage and analysis
- ✓ Development of auxiliary tools for the integration and **exploitation** of data into the new system
- ✓ Deployment of implemented features
- ❖ Validation of results

CONTROL IN STEEL

Thank you for your attention!

RINA Consulting - CSM S.p.A.
Via Di Castel Romano 100 - 00128 ROMA - ITALIA

www.rina.org