STU MTF

eit Manufacturing

Funded by the European Union

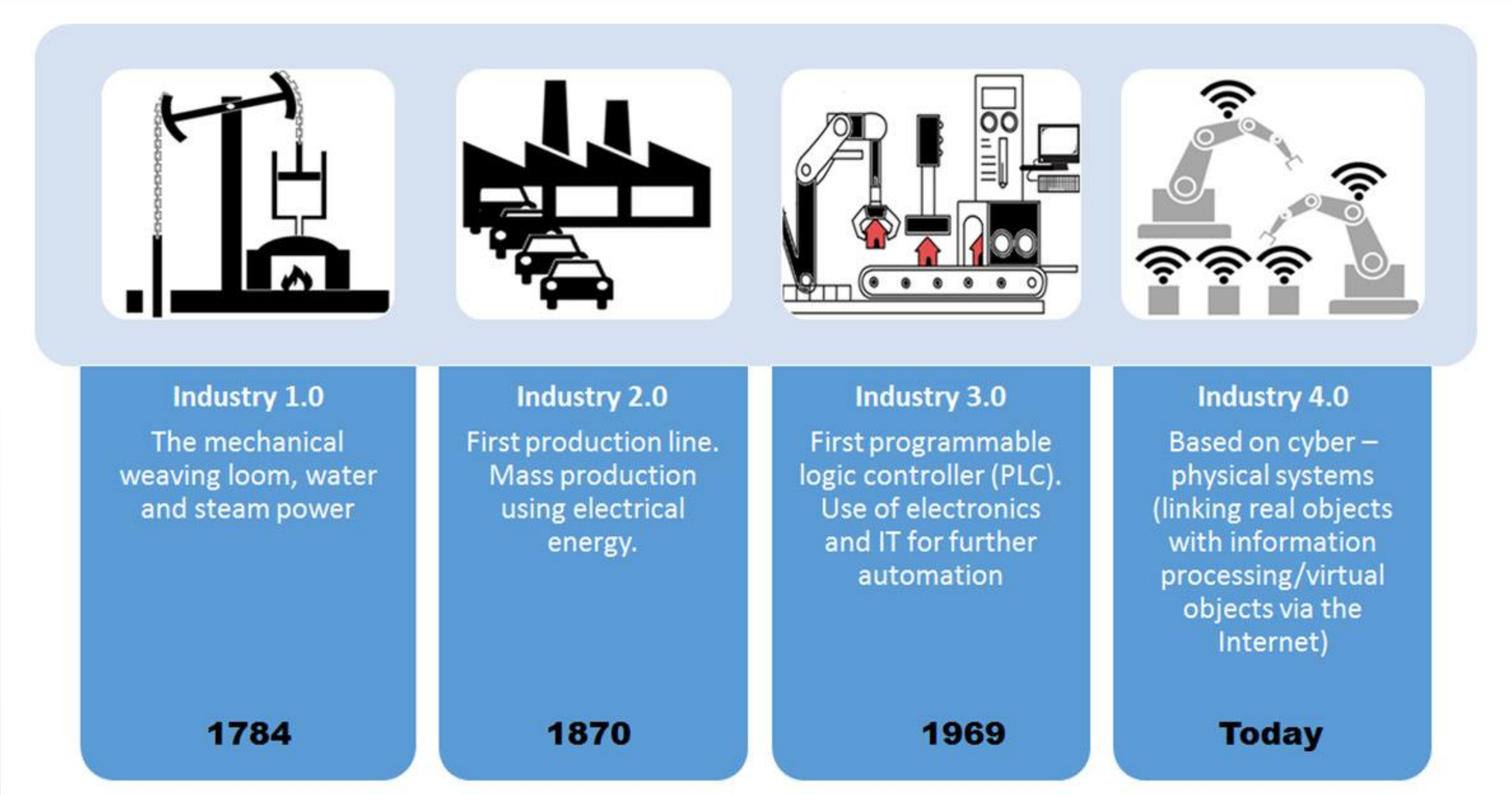
EITM RIS I40H

Big Data in the Industry 4.0 Concept

Lukas Spendla, Peter Schreiber, Pavel Vazan

Slovak University of Technology in Bratislava | Faculty of Materials Science and Technology in Trnava
Institute of Applied Informatics, Automation and Mechatronics

Industrial revolutions



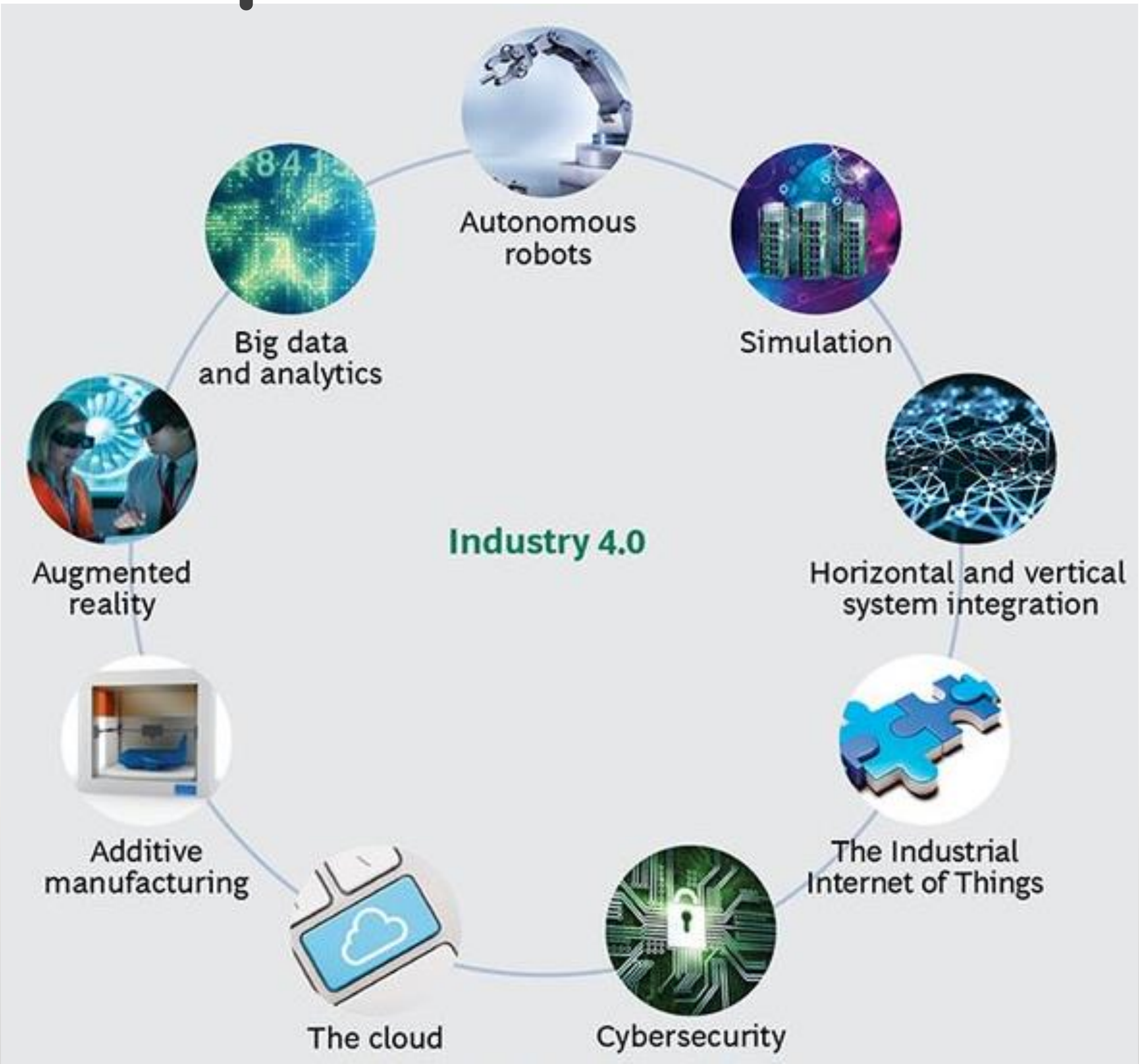
Industry 4.0

The underlying concept of Industry 4.0 is to connect embedded systems and smart production facilities to generate a digital convergence between industry, business and internal functions and processes.

It includes many technologies and concepts such as Internet of Things, Big Data, Cloud, Cyber-physical systems, Augmented reality, Integration…

(Gartner 2015)

Main components

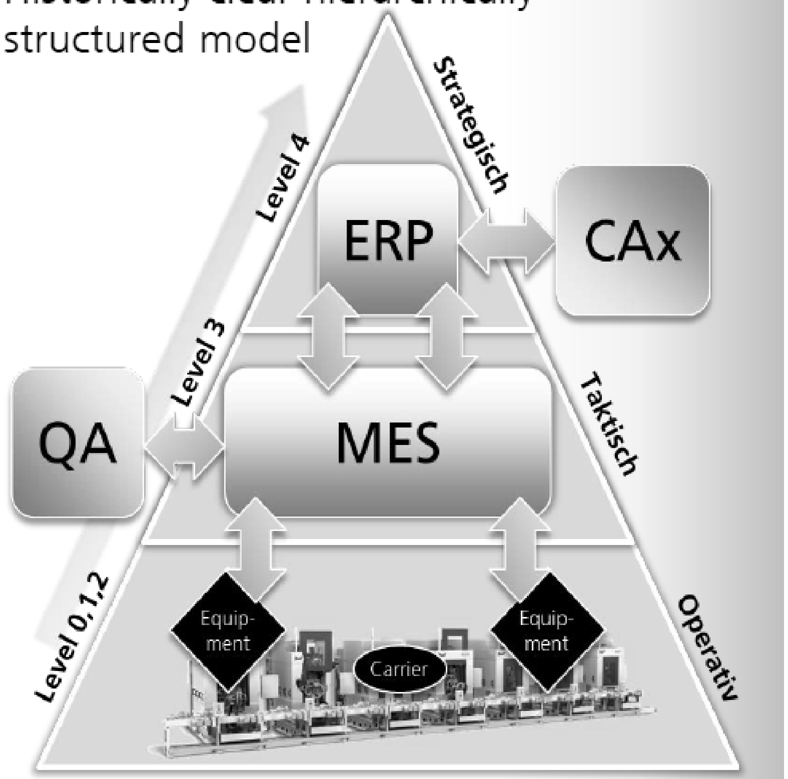


Industry 4.0

The pyramid becomes a net in the cloud

Today

Historically clear hierarchically structured model



Tomorrow

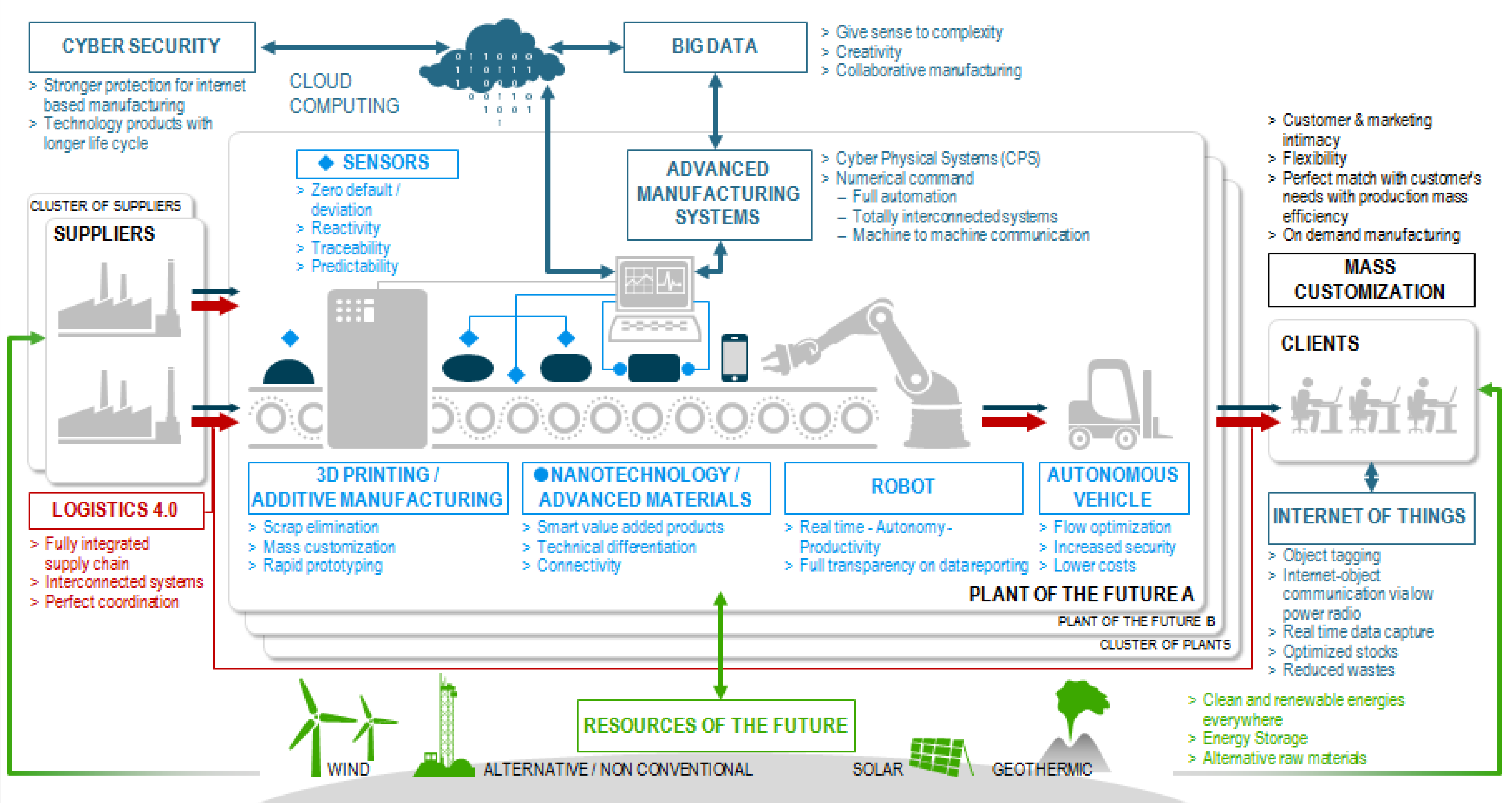
- **Service-orientation**
  - Service-orientation (XaaS) in all areas
  - Service-oriented IT architecture (SoA)
- **De-hierarchization**
  - Dispersal of hierarchical structure
  - New applications based on services
- **App-ization**
  - App development by process owner
  - Simulation in real time
- **Open standardization**
  - Efficiency advantages of IT clouds
  - Focus on information / semantics

Industry 4.0

Change of paradigms in information and communication technology

Today:	Tomorrow:
Centralized	Decentralized (CPS, Cloud)
Software suite	Apps (SaaS)
Integration	Communication
Monolith	Open internet standard
Delayed information	Real-time information
License fee	Pay-per-use

Vision of Industry 4.0 implementation



Big Data in the Industry 4.0 Concept

Big Data

Data growth

- Every day, we create 2.5 quintillion (10^{18}) bytes of data — so much that **90% of the data** in the world today has been created in the **last two years** alone.

  (IBM)

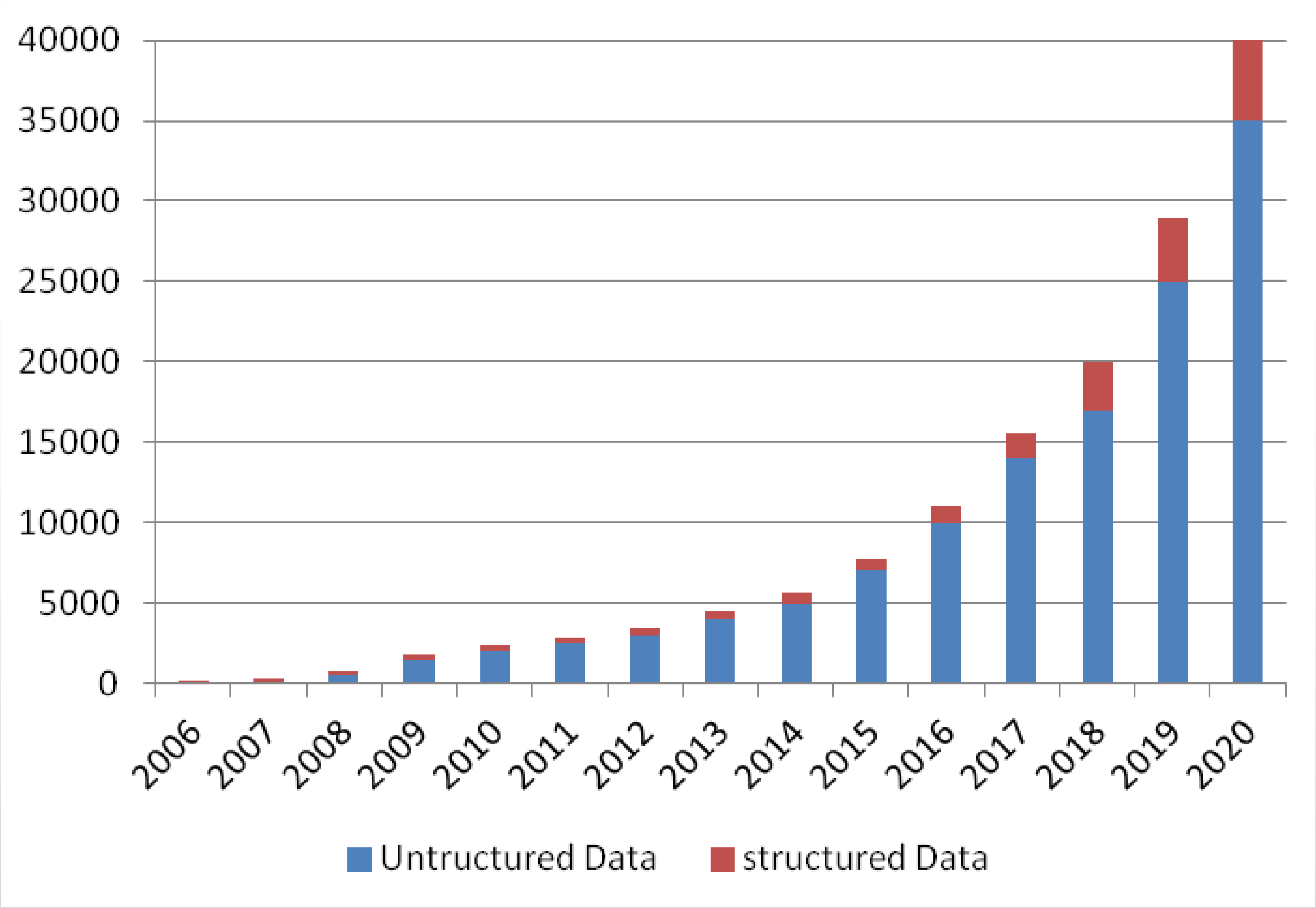
- Data production will be **44 times** greater in 2020 than in 2009.

  (Wikibon Blog)

- The volume of **business data** worldwide is expected to **double** every **14 months**.

  (eBay)

Data growth



40000
35000
30000
25000
20000
15000
10000
5000
0
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Untructured Data
structured Data

Data types

Structured data

- Relational databases, tables, ...

Semi structured data

- XML, JSON, ...

Unstructured data

- Word, PDF, Text, Logs, ...

Big Data

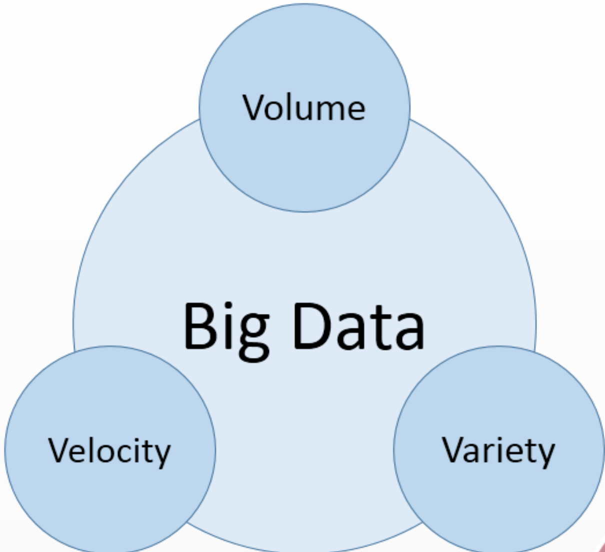
Big Data are high-volume, high-velocity, and/or high-variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization.

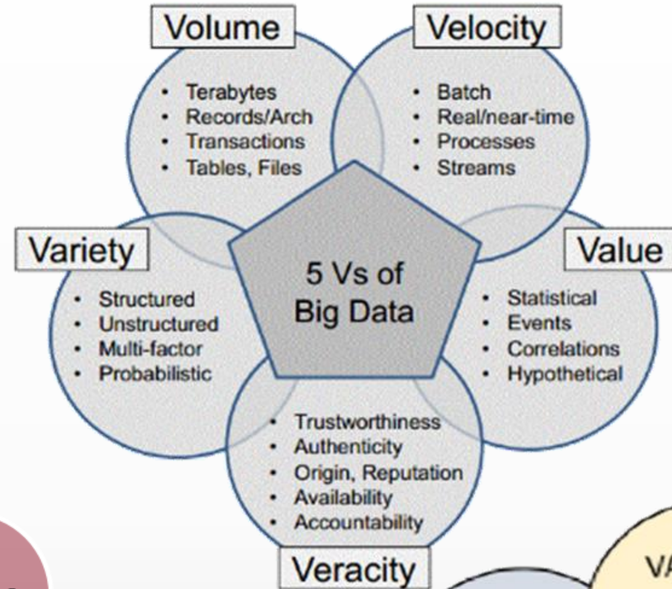
(Gartner 2012)

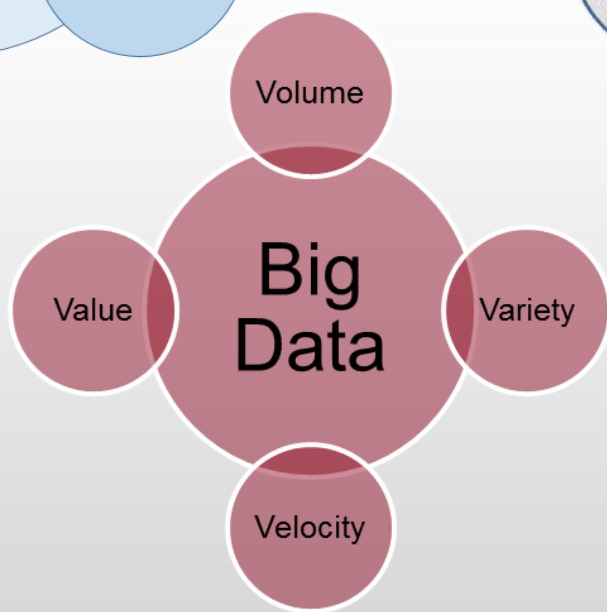
Big Data – 3V

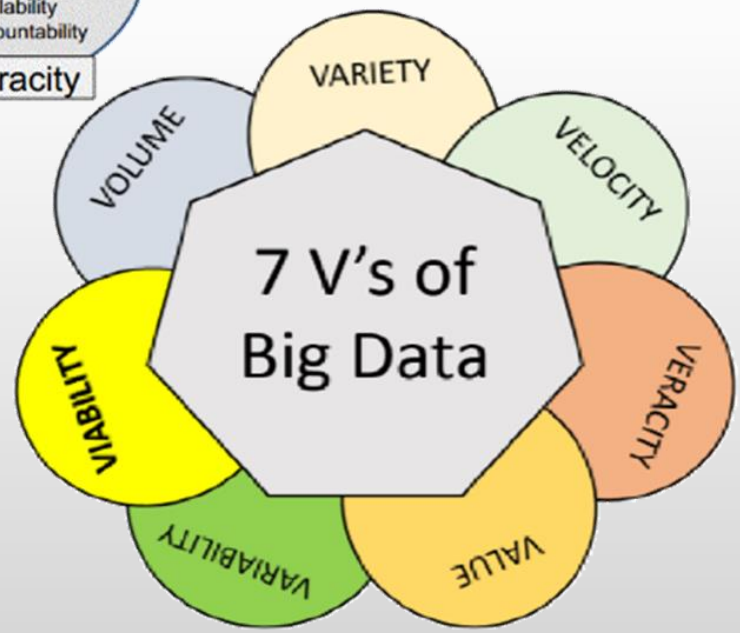
- **Volume –** machine generated data is produced in larger quantities than non traditional data.
- **Velocity –** refers to the speed of data processing.
- **Variety –** refers to large variety of input data which in turn generates large amount of data as output.

Big Data – 3, 4, 5, 7V

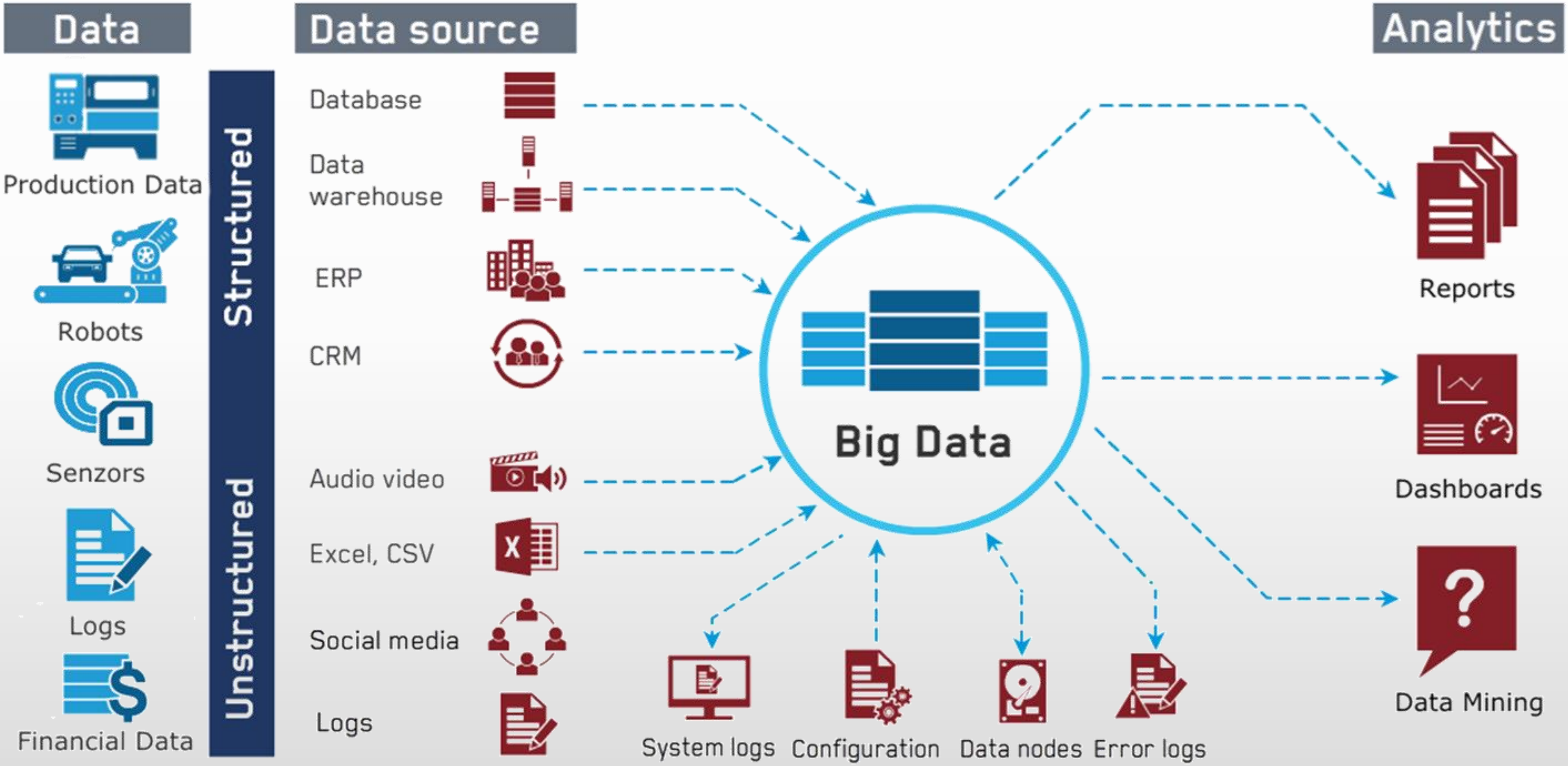








Big Data

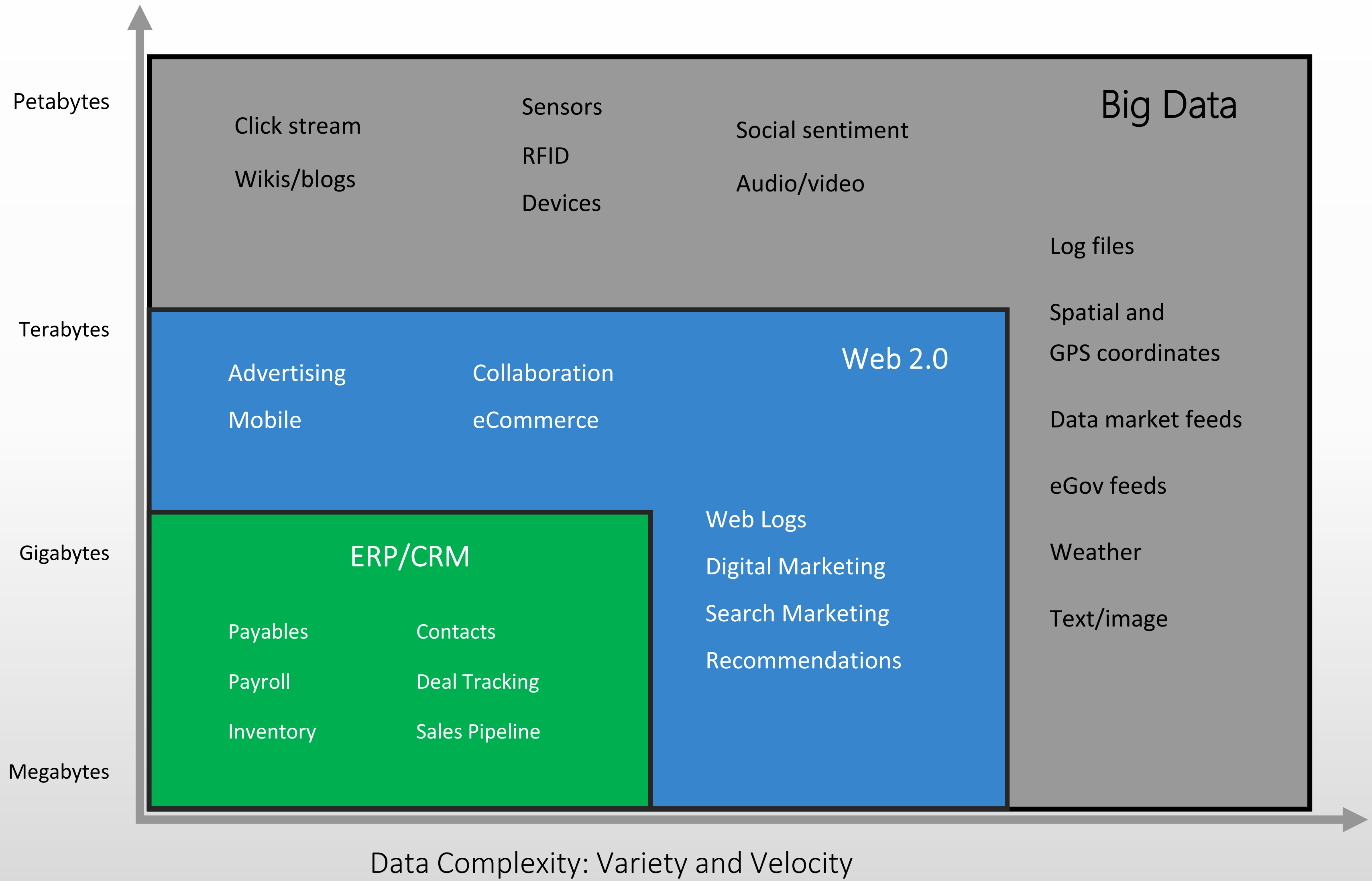


Data
Data source
Analytics
Production Data
Robots
Senzors
Logs
Financial Data
Structured
Unstructured
Database
Data warehouse
ERP
CRM
Audio video
Excel, CSV
Social media
Logs
Big Data
Reports
Dashboards
Data Mining
System logs
Configuration
Data nodes
Error logs

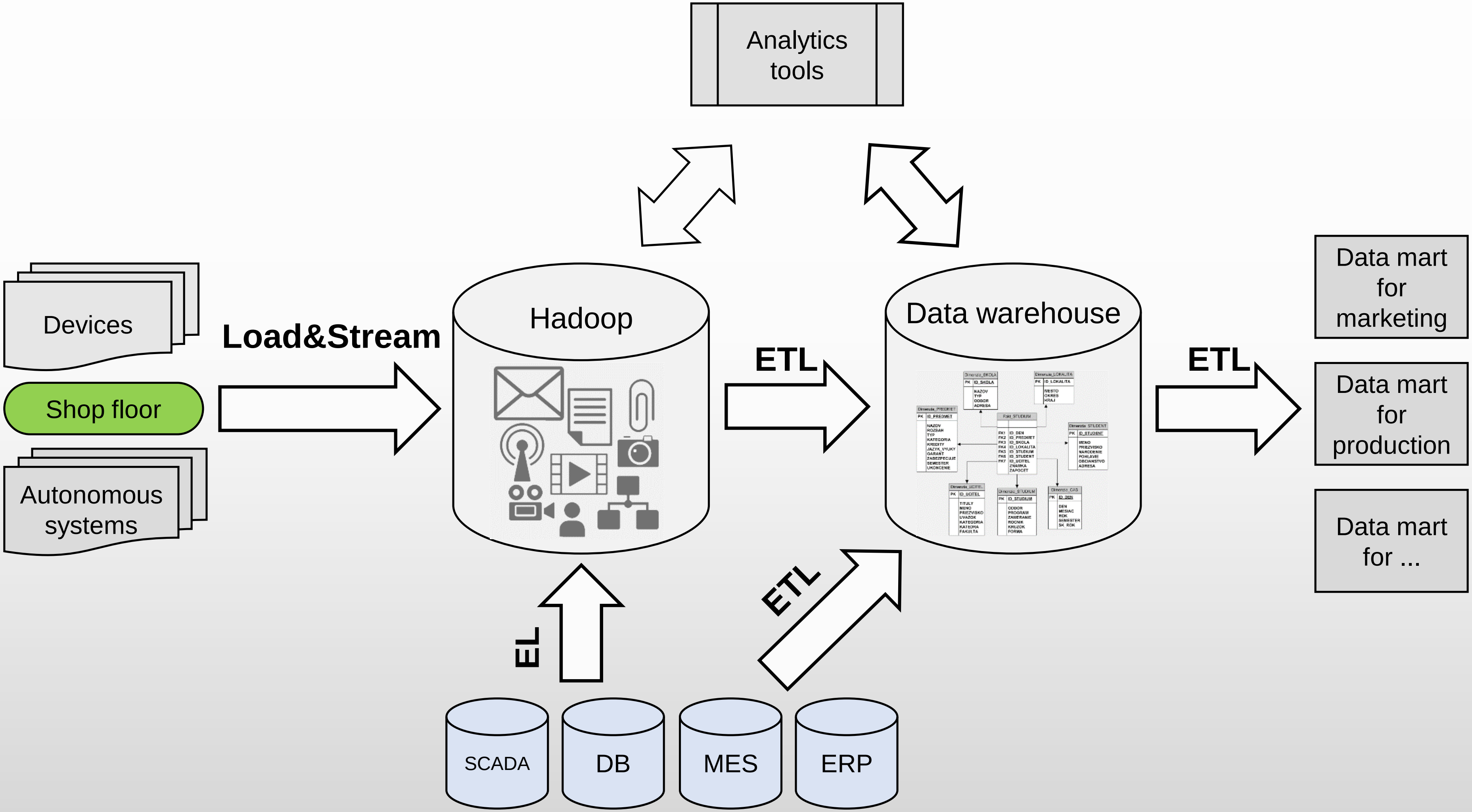
Big Data Analytics

	Traditional Analytics (BI)	vs Big Data Analytics
Focus on	• Descriptive analytics • Diagnosis analytics	• **Predictive analytics** • **Data Science**
Data Sets	• Limited data sets • Cleansed data • Simple models	• Large scale data sets • More types of data • Raw data • Complex data models
Supports	**Causation:** what happened, and why?	**Correlation**: new insight More accurate answers

From ERP/CRM to Big Data



Data processing in the industrial area



Technologies in Big Data

Predictive analysis
Data security
Data mining
Data streams analysis
In-memory computing
Distributed file system
Data virtualization
Data integration
Data preparation and processing
Data quality

Technologies in Big Data

Operational Big Data

- Real-time data processing and storage
- Interactive data processing
- NoSQL technology

Analytical Big Data

- Parallel data processing
- Comprehensive analysis of all data
- MPP and MapReduce technologies

Technologies in Big Data

	Operational	**Analytical**
Response	1ms – 100ms	1min – 100min
Parallel processing	1000 – 100 000	1 – 10
Access	Write and read	Read
Query	Selective	Unselective
Data	Operational	Historical
Users	Customers	Data analysts
Technologies	NoSQL	MPP and MapReduce

Analytic technologies in Big Data

Hadoop

- Uses the MapReduce algorithm

Apache Spark

- In-memory data streams processing

Cluster Map Reduce

- Uses Gluster file system

High Performance Computing Cluster

- Platform for parallel data processing using ECL language

Hydra

- Enables more efficient data search, uses a modified MapReduce algorithm

Current State in Manufacturing

Business Challenges:

- Complexity and rapid growth of machine data.
- Difficult to capture small fraction of machine for better decision.
- In-ability to analyze machine data and combine it with enterprise data for a full view analysis.

Big Data in Manufacturing

Benefits:

- Gain real-time visibility into operations, customer experience, transactions and behavior.
- Proactively plan to increase operational efficiency.
- Identify and investigate anomalies.
- Monitor end-to-end infrastructure to proactively avoid service degradation or outages.

Using Big Data in practice

- Predictive maintenance planning
- Analysis and planning of warehousing
- Proactive plan to increase operational efficiency
- Identification of anomalies in the production system
- Overview of the current state of the production system, customer segment and transactions
- Data analysis to prevent outage and planning of predictive controls of compressors and turbines for aircraft engines of Boeing B777, B747, Airbus A380 ......

Top 5 Myth About Big Data

- Big Data is only about massive data volume
- Big Data means Hadoop
- Big Data means unstructured data
- Big Data is for social media feeds and sentiment analysis
- NoSQL means No SQL

Big Data in the Industry 4.0 Concept

Hadoop

Hadoop

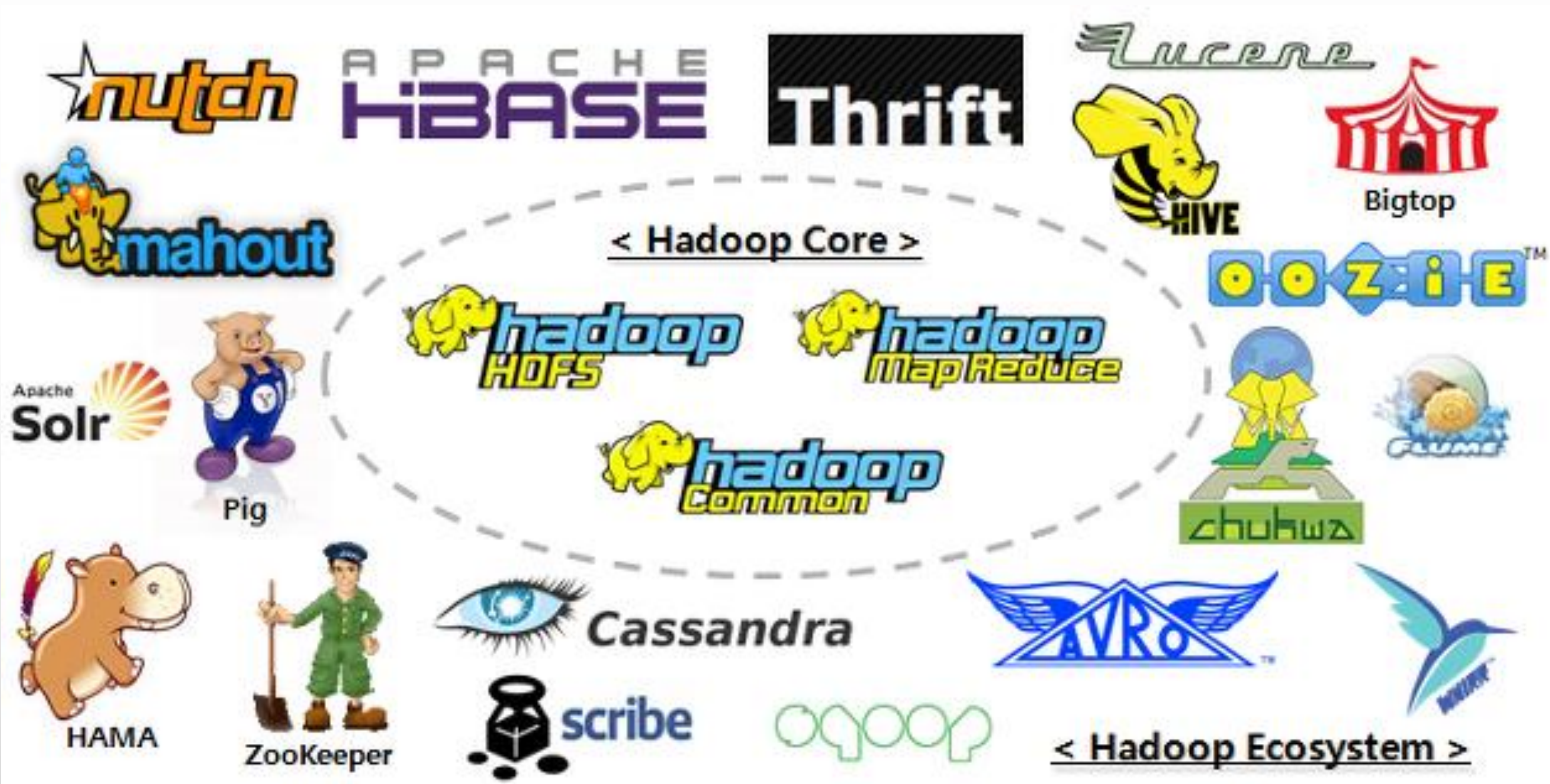
- Hadoop is a software framework for storing, processing and analyzing Big Data
  - Parallel processing
  - Scalability
  - Resistance to failures
  - Open source

- It allows to store huge amounts of data very efficiently
- It covers the distributed file system and allows developers to focus on tools

Why Hadoop is needed

- The amount of data sources
  - Internet of Things
  - Sensor data
  - Data streams
- More data means more questions
- More data means better answers
- Hadoop enables simple scalability
  - The price for TB is usually lower than for traditional enterprise systems
- Hadoop can be integrated with existing components in data centers

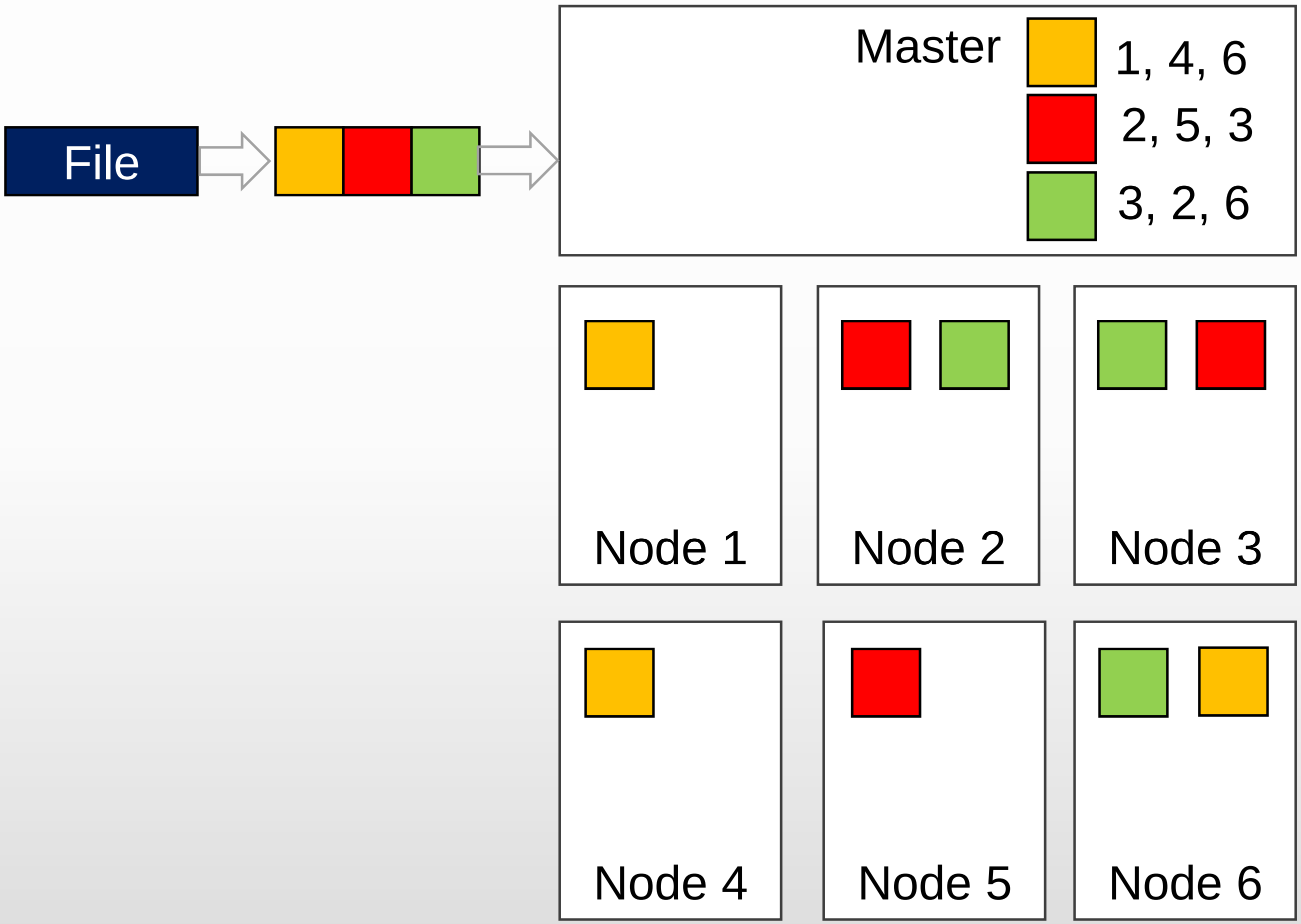
Hadoop ecosystem



Hadoop Distributed File System

- HDFS is a storage for Hadoop
- File system that can hold any data types
- Provides cheap and robust storage for large volumes of data
  - Data is replicated across multiple computers
- HDFS performance is optimal with a “fairly small” amount of large files
  - Better are millions, like trillions of files
  - A typical file size is 100MB or more
- Files in HDFS are write only once
  - It is not possible to add additional data to the files
  - Random writing is not possible

HDFS and file storing



Big Data in the Industry 4.0 Concept

Application of Big Data in manufacturing

Application of Big Data in manufacturing

- Assembly process – data collecting, integration and predictive maintenance
- Paint shop – data collecting, integration and predictions
- Screwing robots – data collecting, integration and monitoring

Big Data in the Industry 4.0 Concept

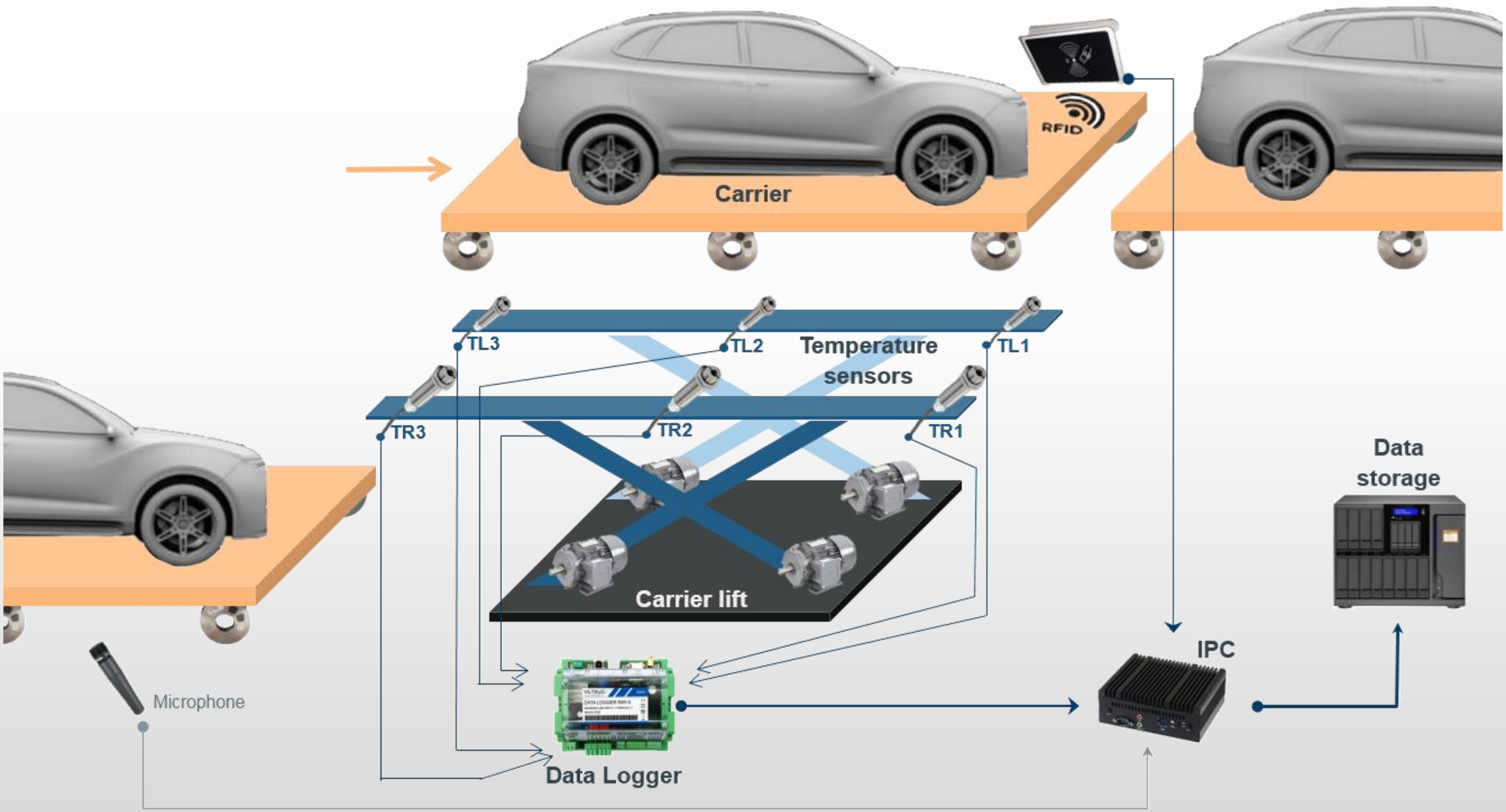
Assembly process

Funded by the European Union

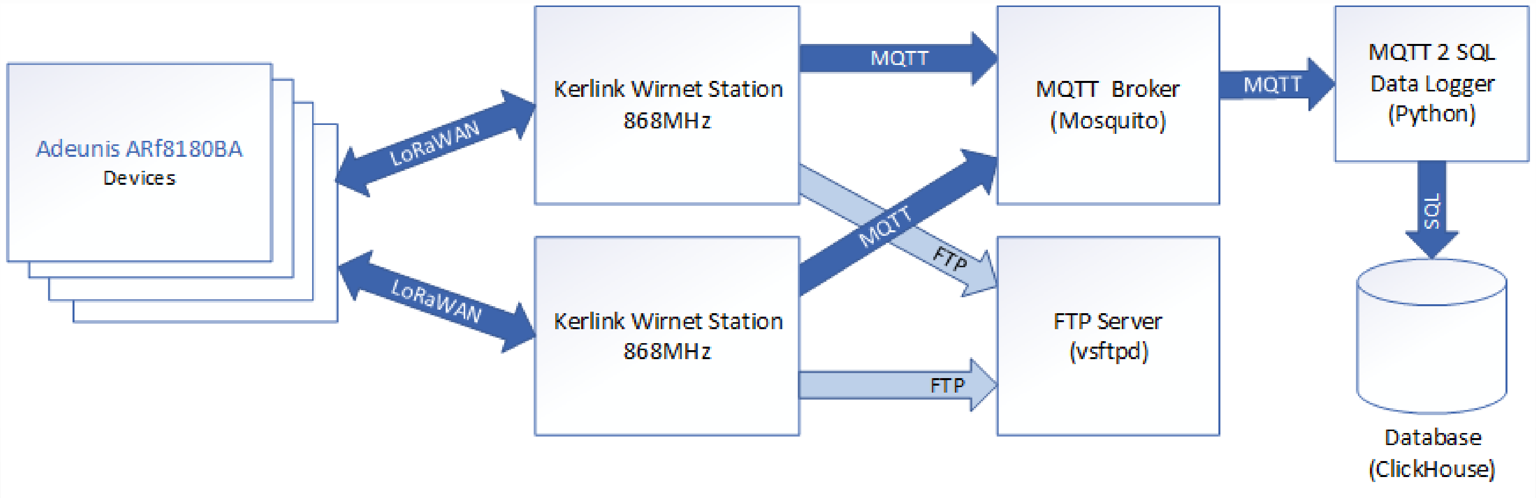
Assembly process

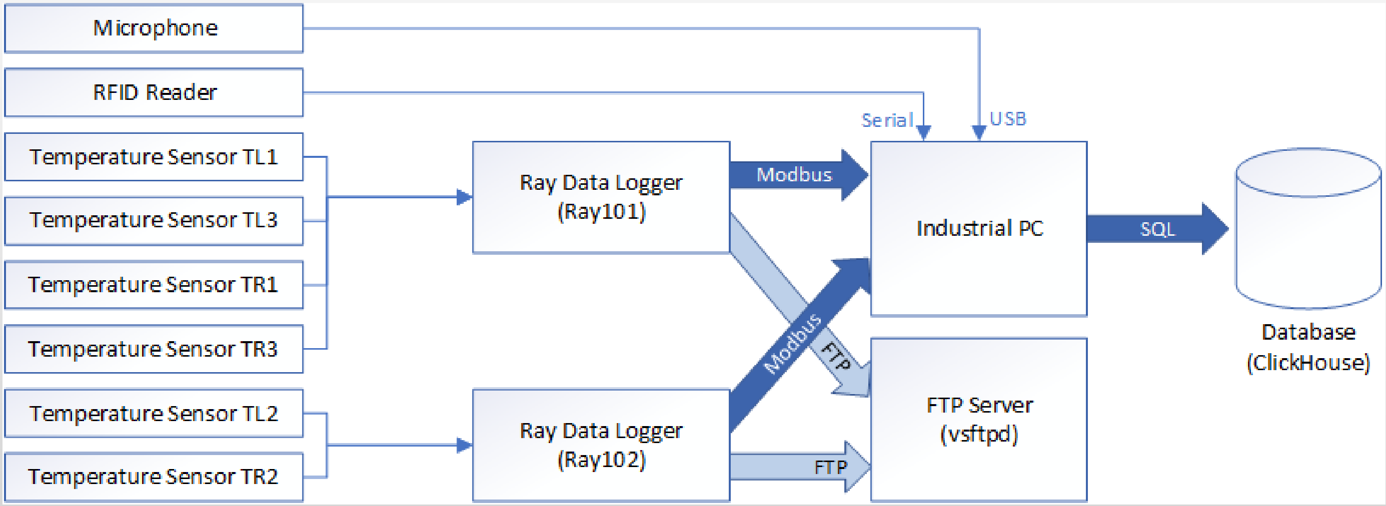
- The assembly process is one of the most important processes in car production.
- Assembly conveyor consists of 59 carriers that move along a closed track with a length of approximately 1250m.
- Each carrier, that is used to transport car bodies during the entire assembly process, consists 6 wheels with the bearings.
- From time to time, the bearing of one of the carrier wheels seize and thus the conveyor and, of course, whole assembly process is stopped.

Stationary temperature and sound measurement infrastructure

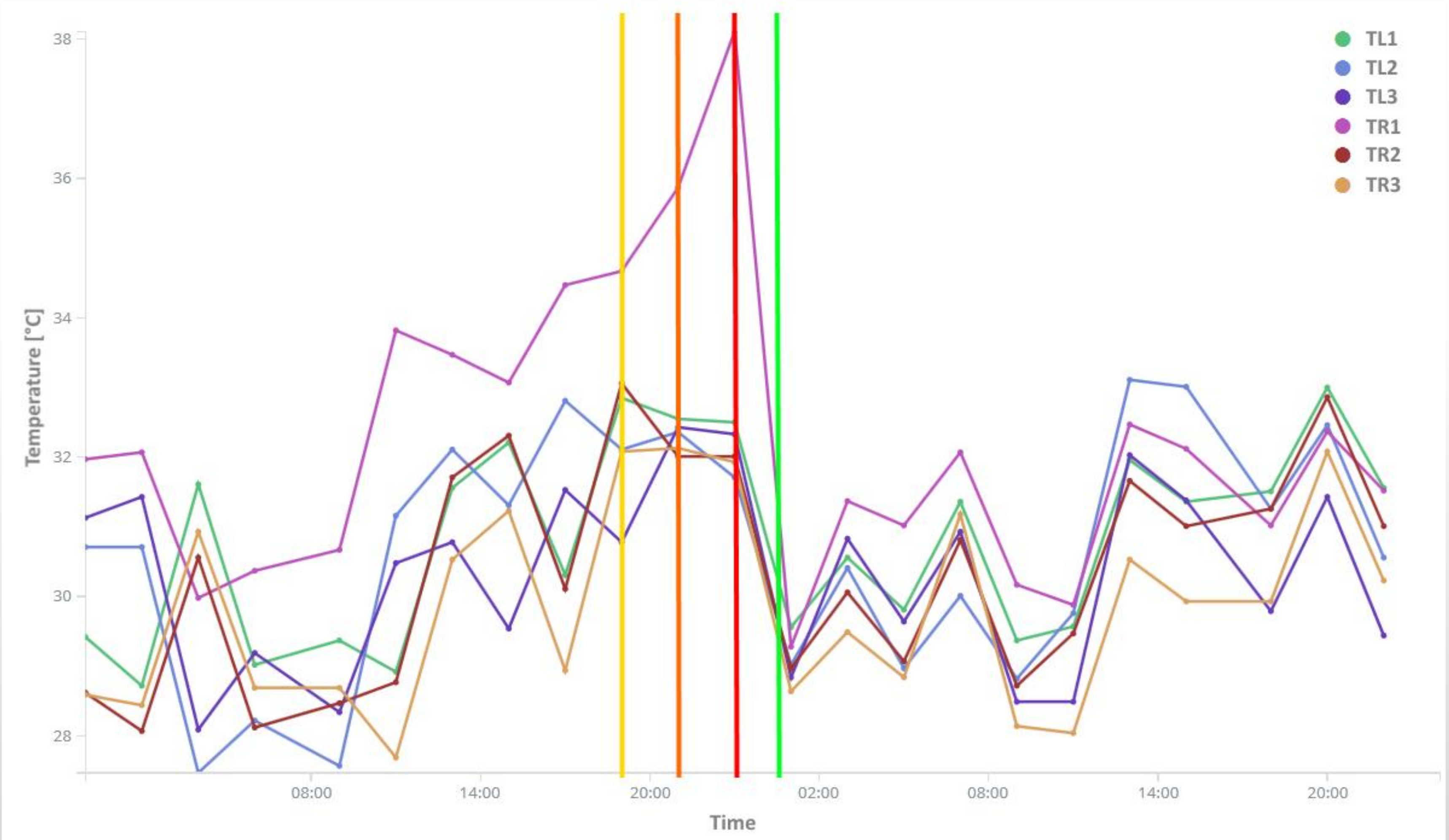


Data collection and integration archit. for IIoT measurement

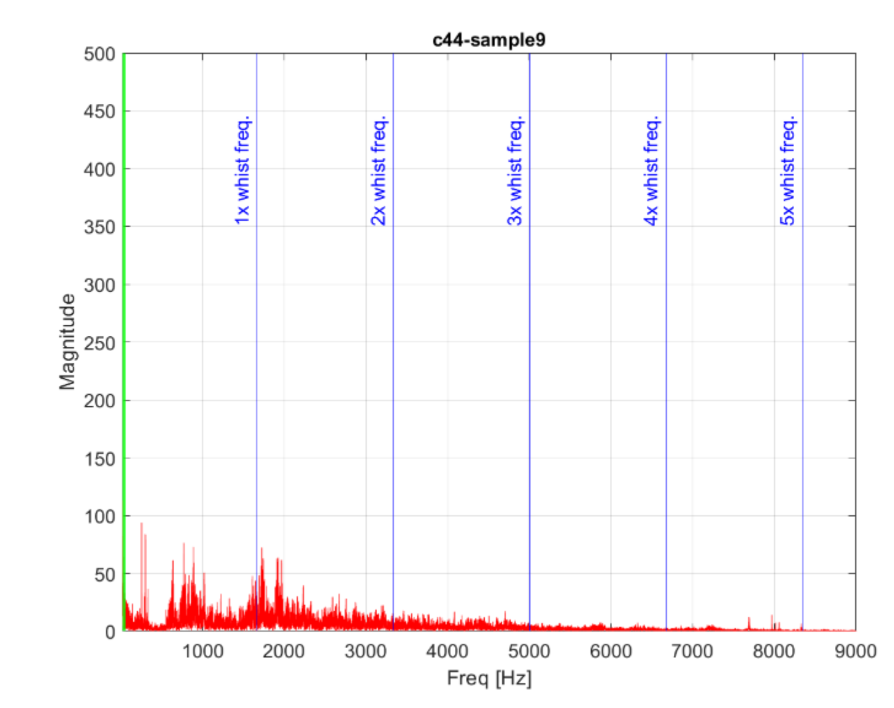


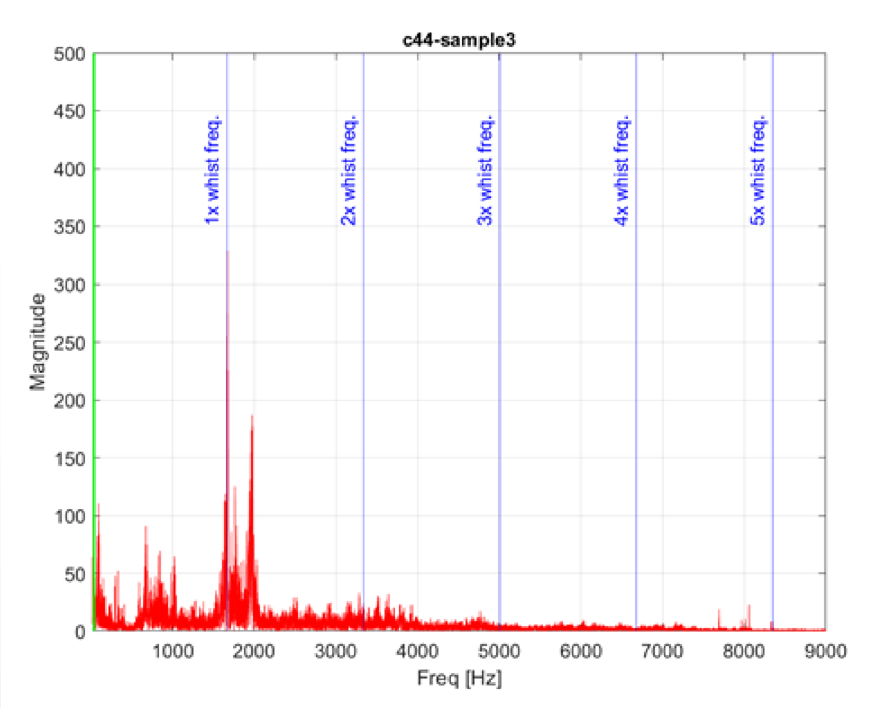


Carrier bearing temp. before and after predictive maintenance



Frequency spectrum of healthy and fault bearing





Assembly process – results

- Implementation of IIoT sensors for temperature and sound measurement.
- Big data collection and integration architecture for the whole assembly process.
- Implementation of trigger-based anomaly detection.
- Implementation of bearing failure detection using the artificial intelligence.
- Implementation of sound measurement system for the bearing condition detection.

Big Data in the Industry 4.0 Concept

Paint shop

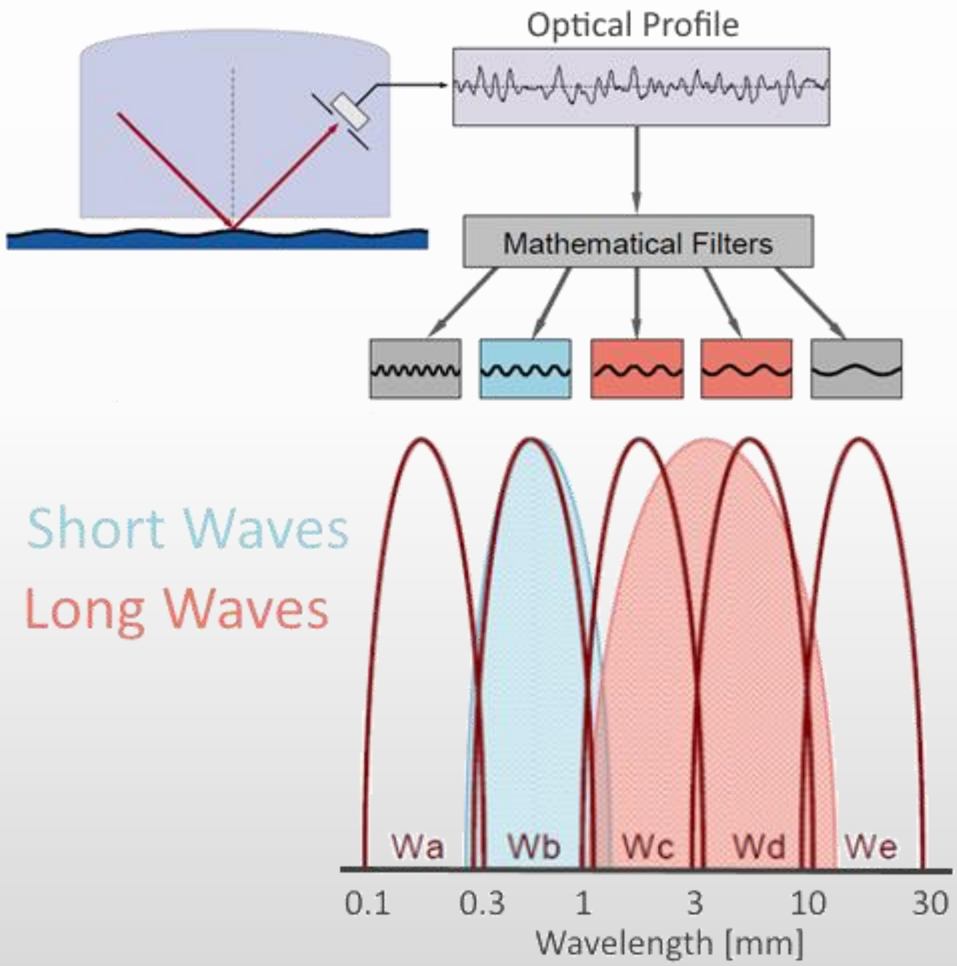
Paint shop

- Paint process is one of the most complex process from the manufacturing point of view.

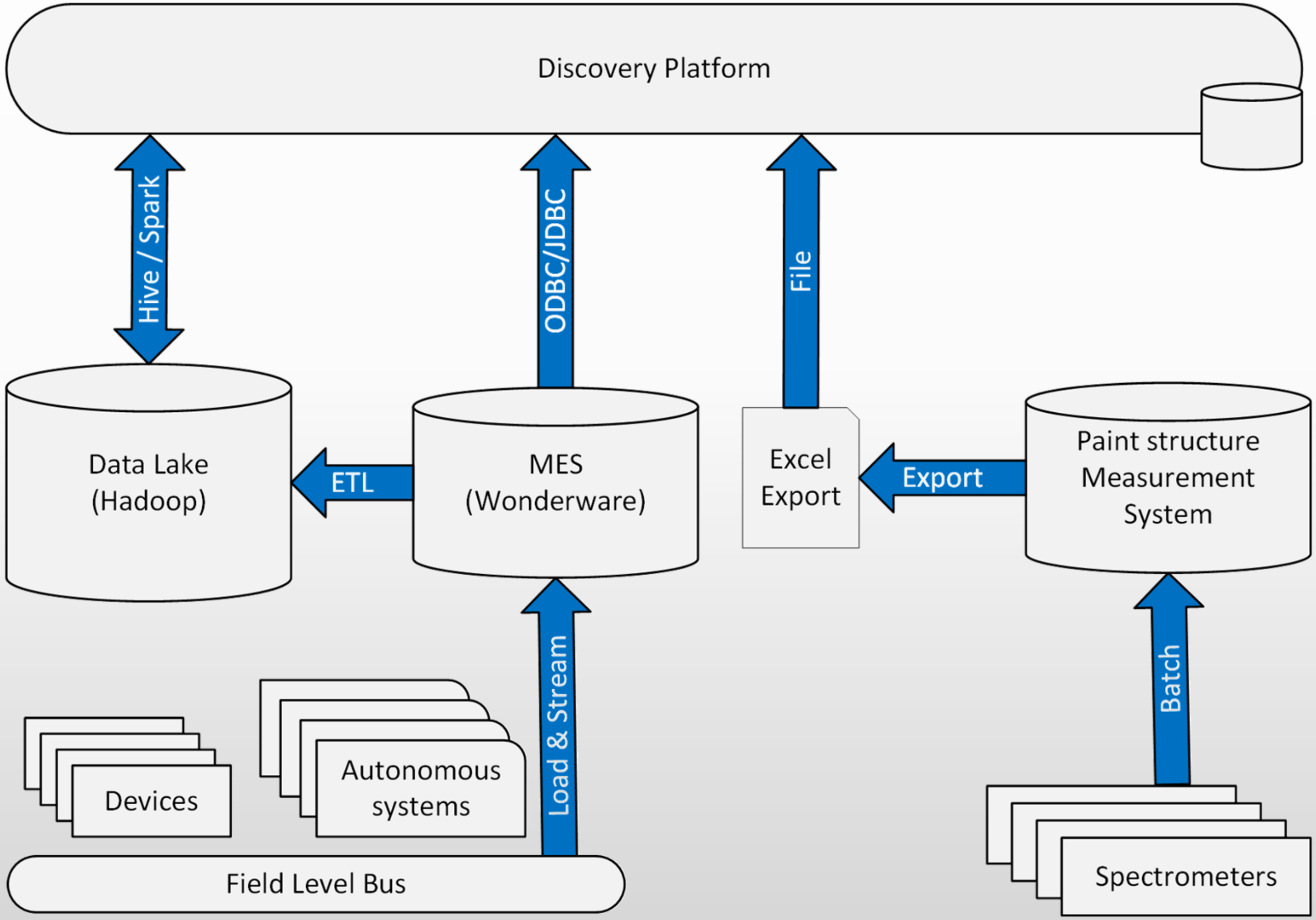
- The system collects data from approximately 700 process tags, i.e. sensors, in intervals between 100 milliseconds and up to 10 minutes, based on the process value requirements.

- Multiple identified correlations between the paint shop process parameters, weather and paint structure errors, but there were no clear relationships.

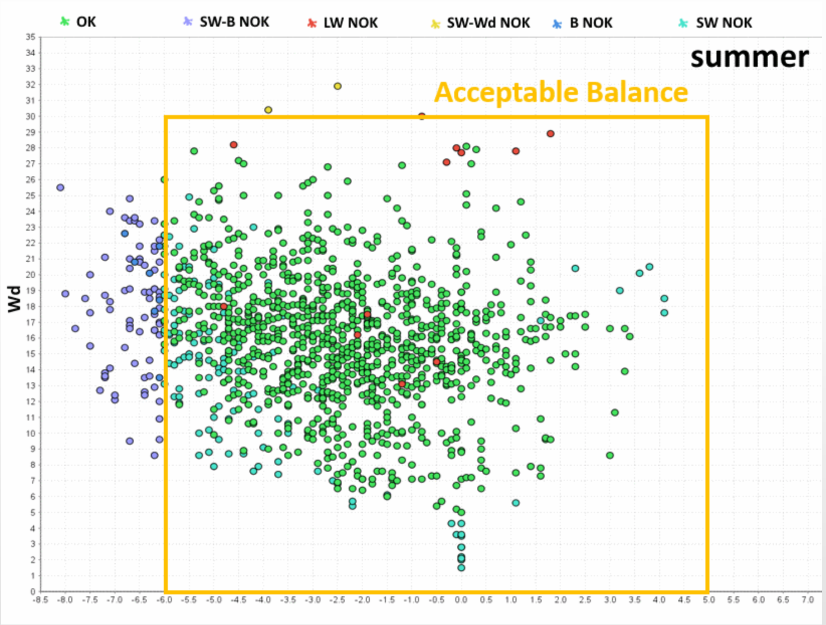
Measurement principle

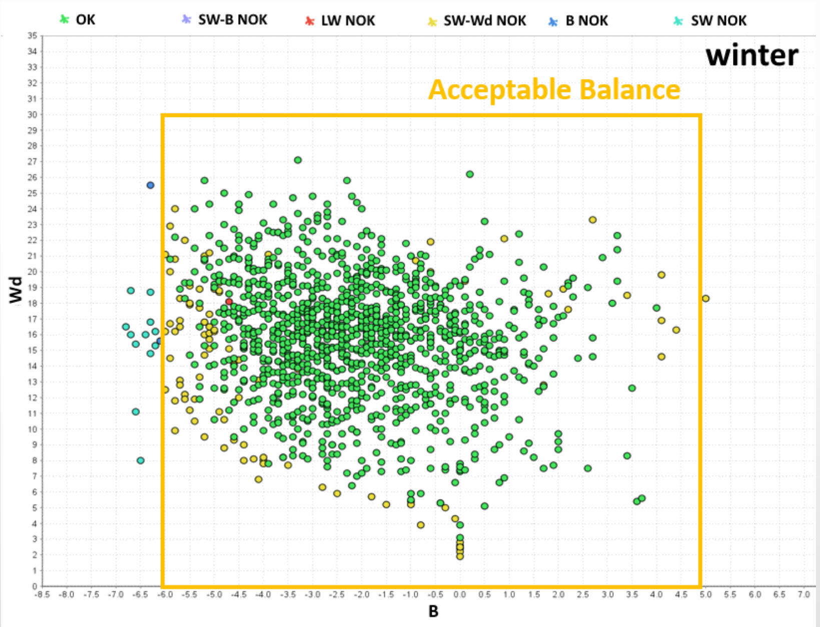


Data collection and integration architecture

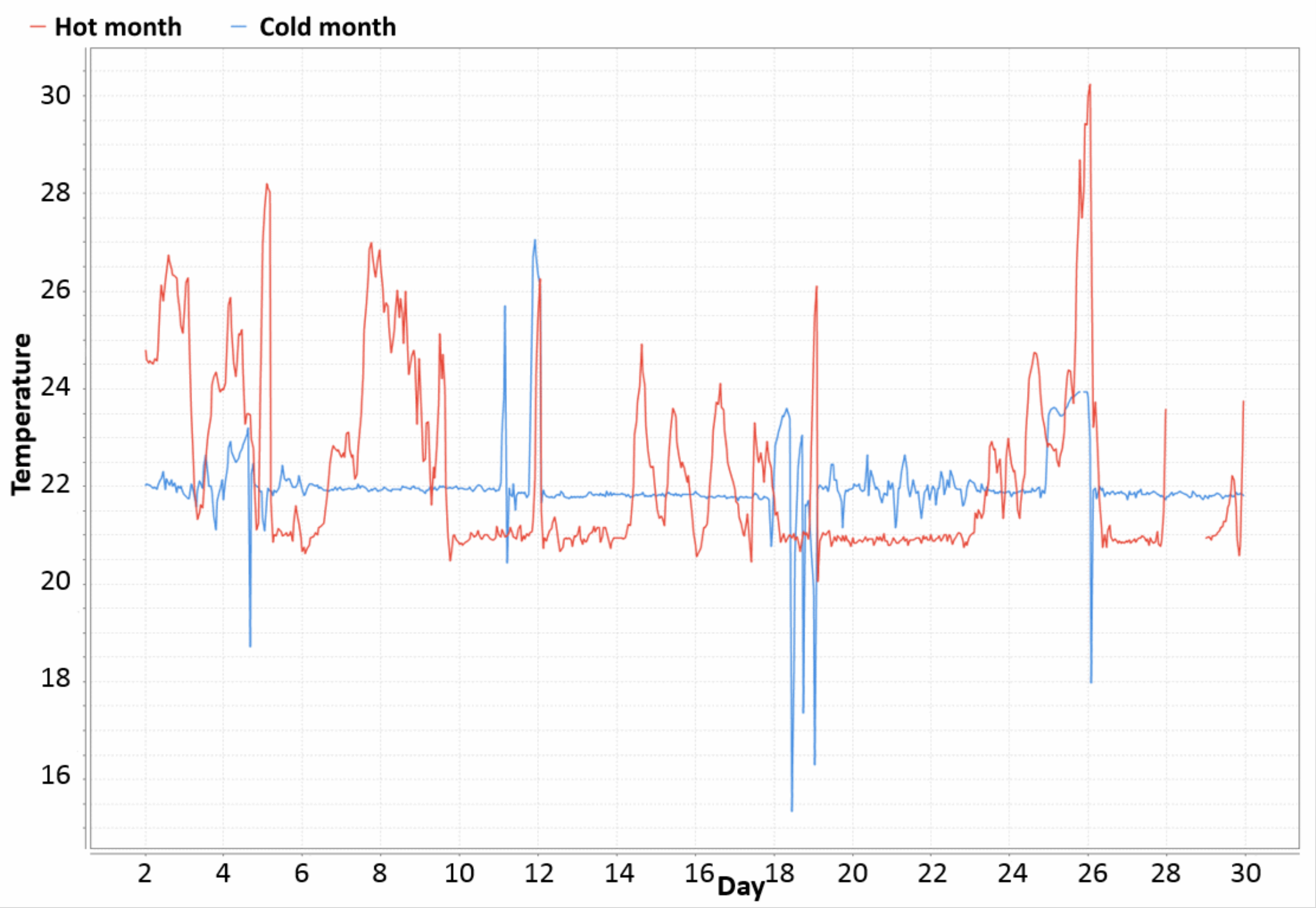


Results of the paint structure evaluation





Process parameter disproportion between cold and hot months



Paint shop – results

- Big data collection and integration architecture for bodyworks paint process.
- Identified process parameters with the most significant impact on paint structure quality confirmed by experts and operators from practice.
- Identified disproportions between several process parameters during the summer and winter operation of the paint shop.

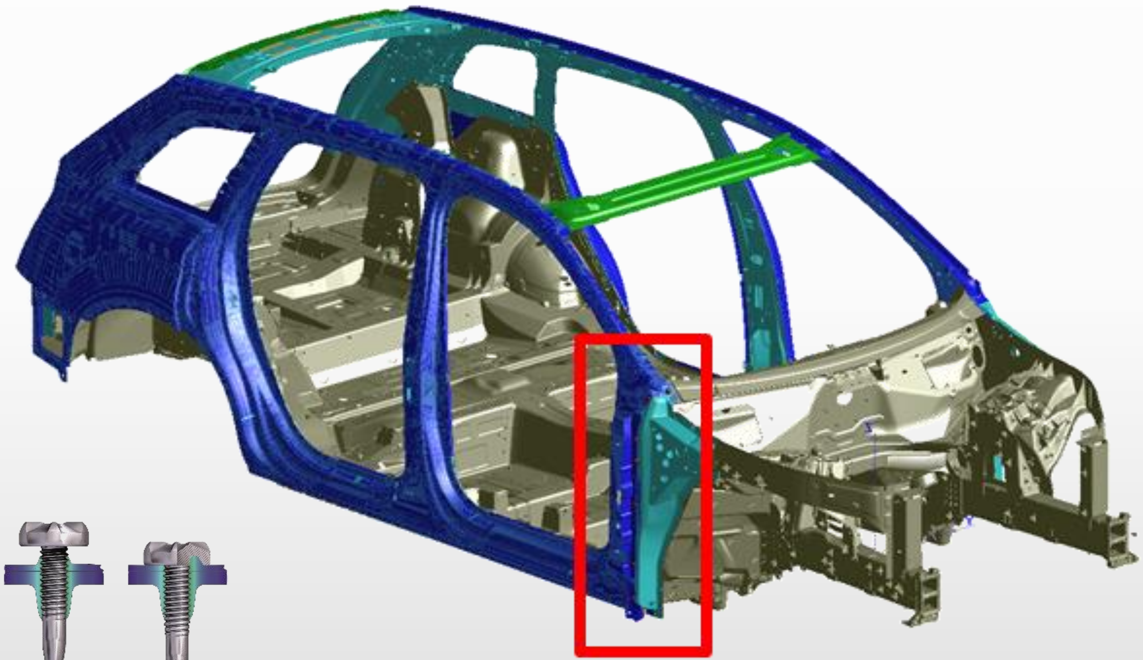
Big Data in the Industry 4.0 Concept

Screwing robots

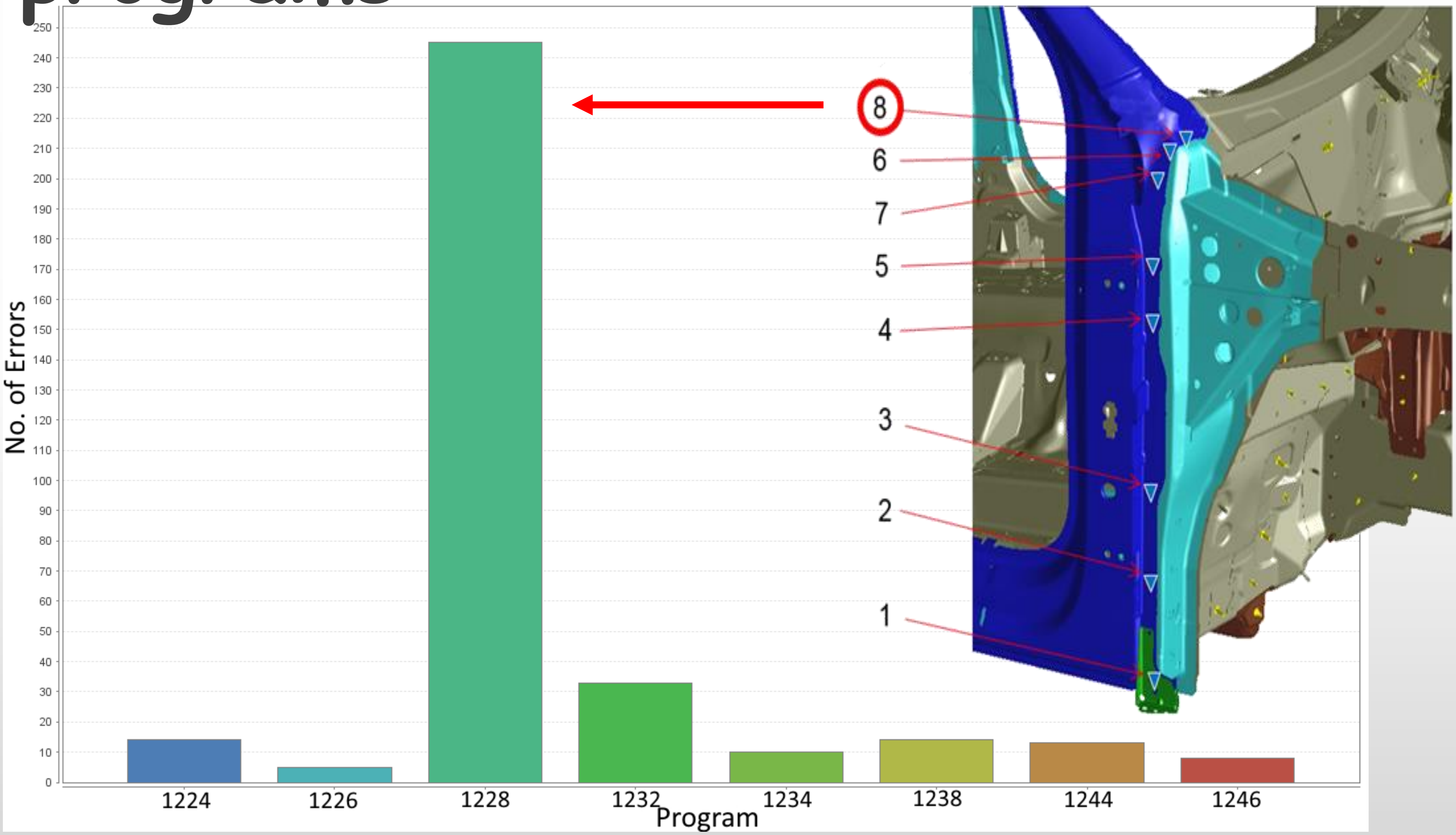
Screwing robots

- Screwing process affects the quality and durability of the bodywork.
- Screwing process performed using more than 90 screwing robots.
- Data from the screwing process are stored only temporarily – depending on the size of the screwing robot memory.

Screwing process



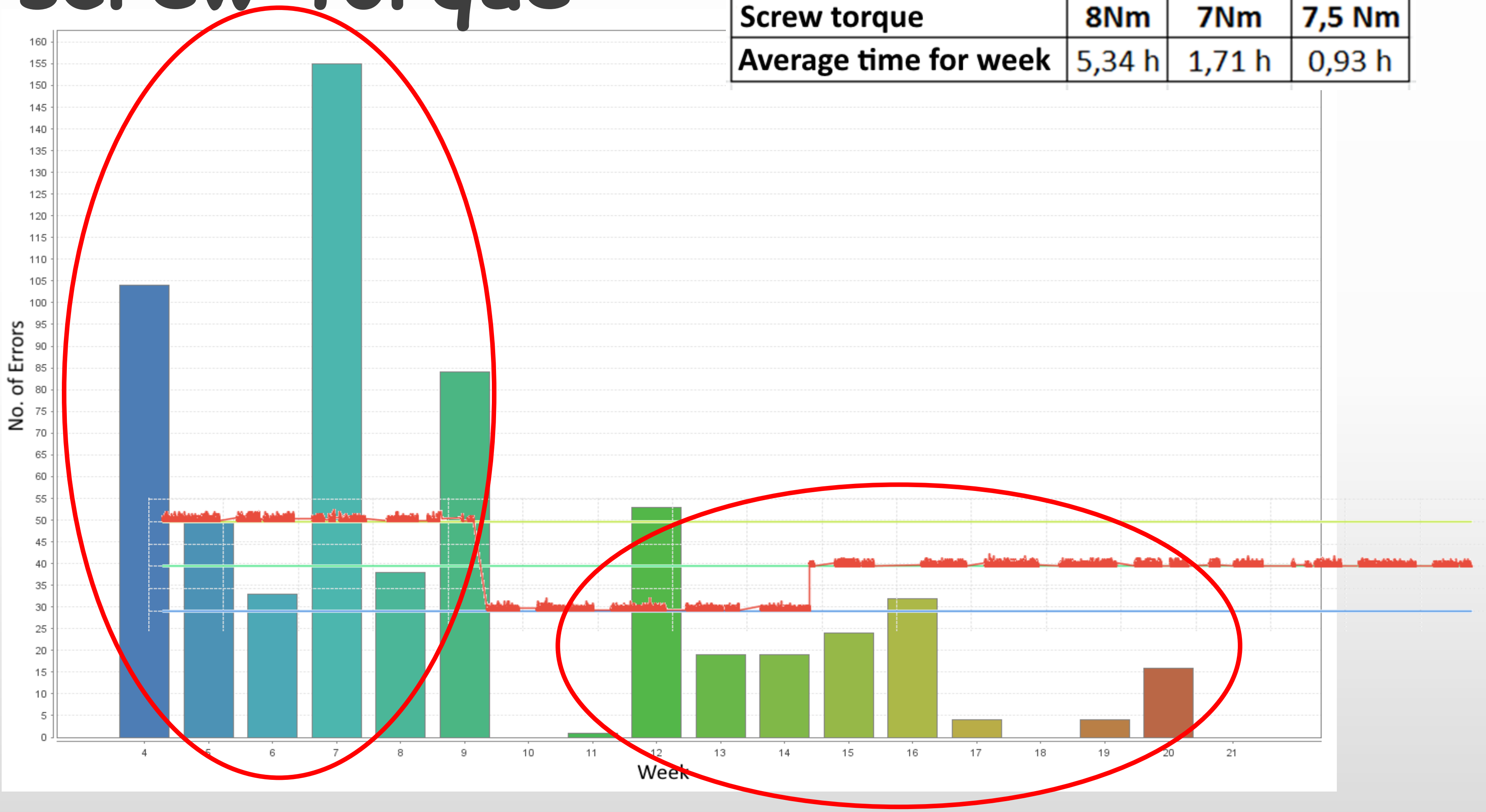
Number of errors in screwing programs



Number of errors based on the screw torque

Amount of error time

Screw torque	8Nm	7Nm	7,5 Nm
Average time for week	5,34 h	1,71 h	0,93 h



Screwing robots – results

- Big data collection and integration architecture for screwing process.
- Identified the most critical screwing programs.
- Identified the most common screw program errors.
- Monitoring and identification of the impact of changes in predefined values of screwing process parameters.
- Monitoring the influence of the screwing point and the robot on the life of the screwing robot.

Big Data in the Industry 4.0 Concept

Lukas Spendla, Peter Schreiber, Pavel Vazan

Slovak University of Technology in Bratislava | Faculty of Materials Science and Technology in Trnava
Institute of Applied Informatics, Automation and Mechatronics