VDEh-Betriebsforschungsinstitut GmbH

Big Data Storage and Steel Industry

Jens Brandenburger

Norbert Holzknecht

BFI Angewandte Spitzenforschung | VDEh-Betriebsforschungsinstitut GmbH

NewTech4Steel

Big Data Storage and Steel Industry

Definition and description of technologies

Multiple definitions of „Big Data“

- „Big Data describes data stocks which, due to their size, diversity or fast-paced nature, can only be processed to a limited extent by current databases and data management tools.”
  - Prof. Dr. Hasso Plattner: „Enzyklopädie der Wirtschaftsinformatik“

GITO

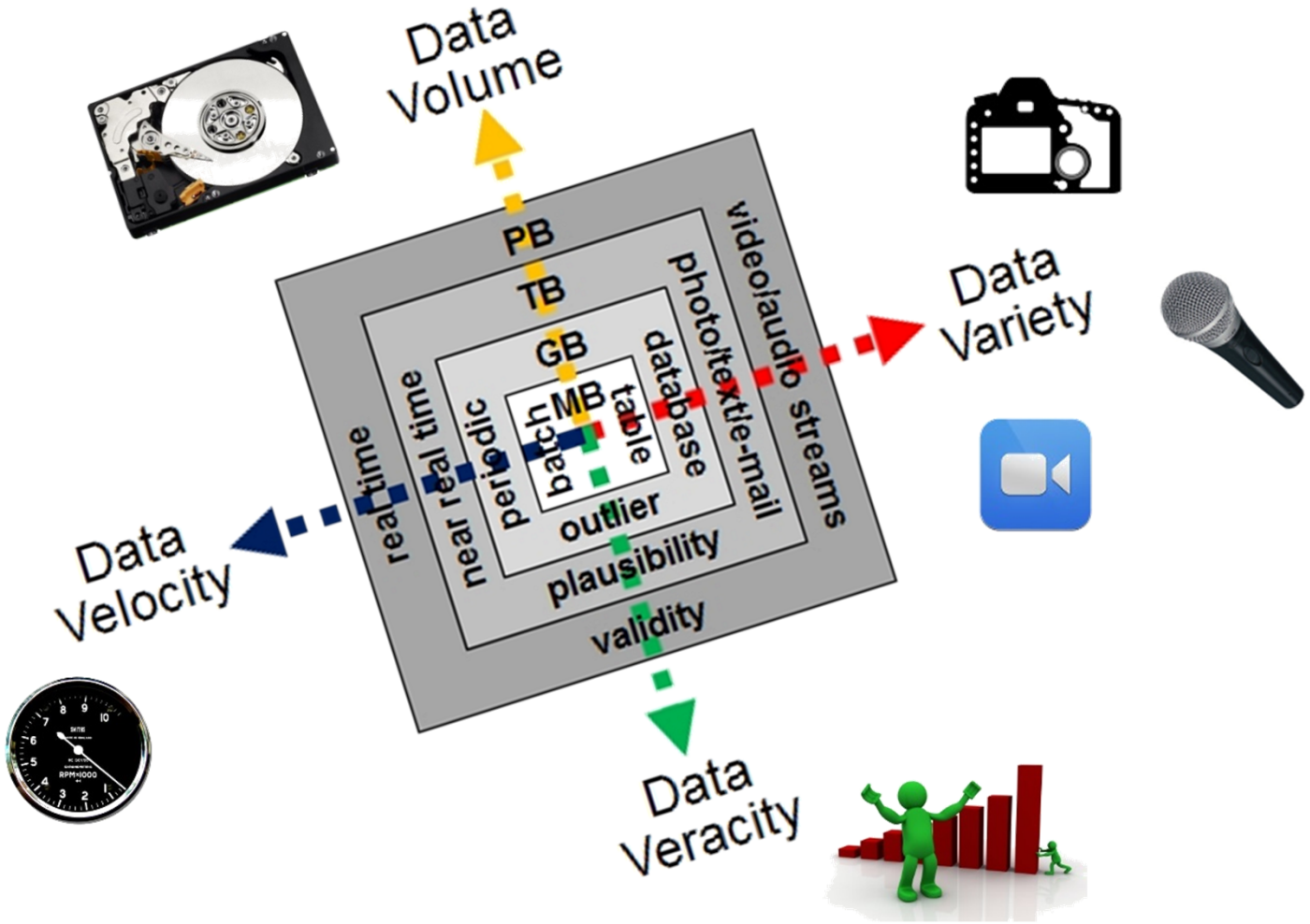
- „Big data is a term used to refer to the study and applications of data sets that are too complex for traditional data-processing application software to adequately deal with. Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualisation, querying, updating, information privacy and data source.”
  - Wikipedia: https://en.wikipedia.org/wiki/Big_data


WIKIPEDIA

- „Big Data refers to the analysis of large amounts of data from multiple sources at high speed with the aim of generating economic benefits.“
  - BITKOM-Leitfaden: „Big Data im Praxiseinsatz: Szenarien, Beispiele, Effekte.“

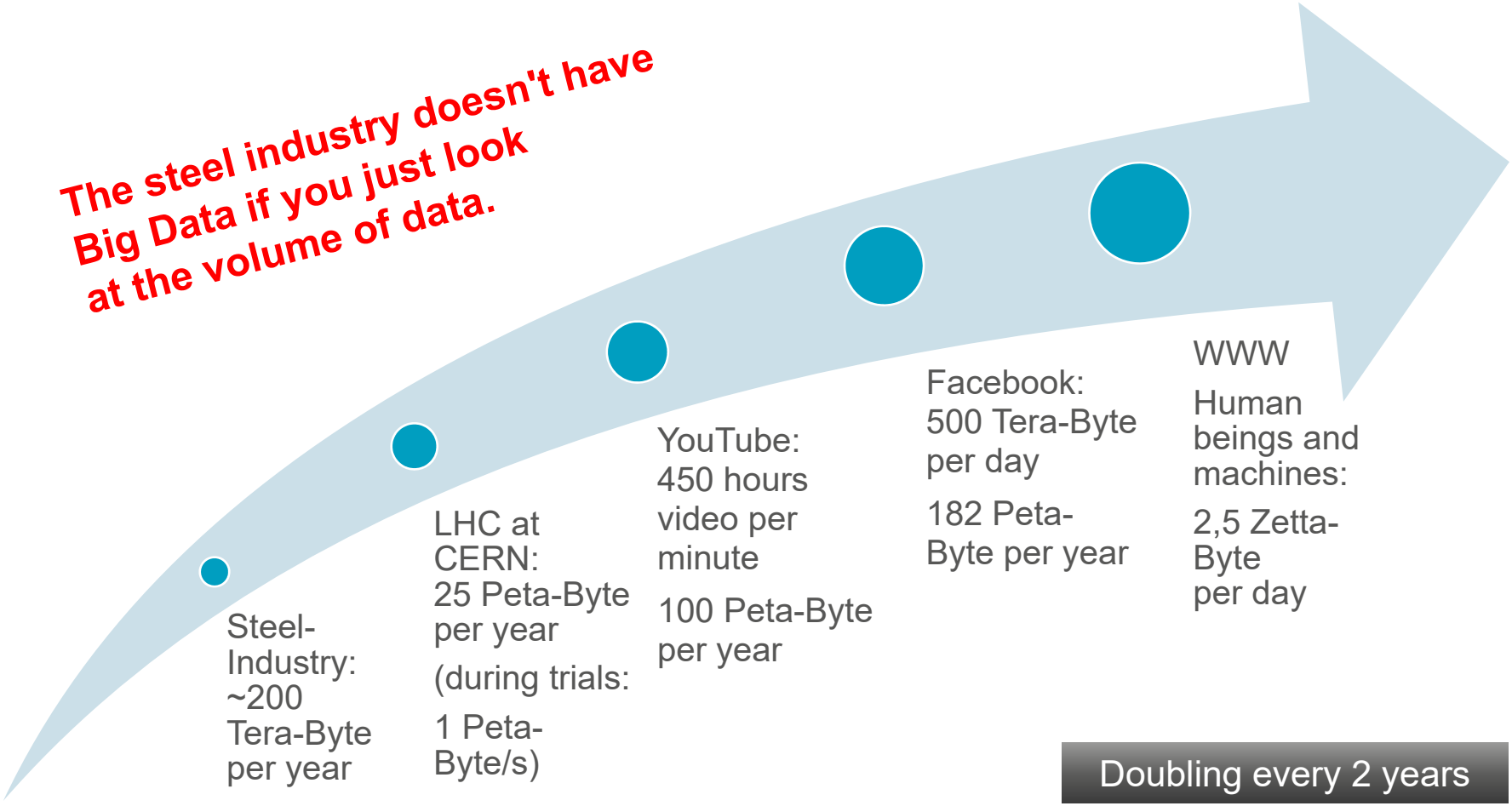
bitkom

„Big Data“, the four V‘s



Volume: Amount of information

The steel industry doesn't have Big Data if you just look at the volume of data.

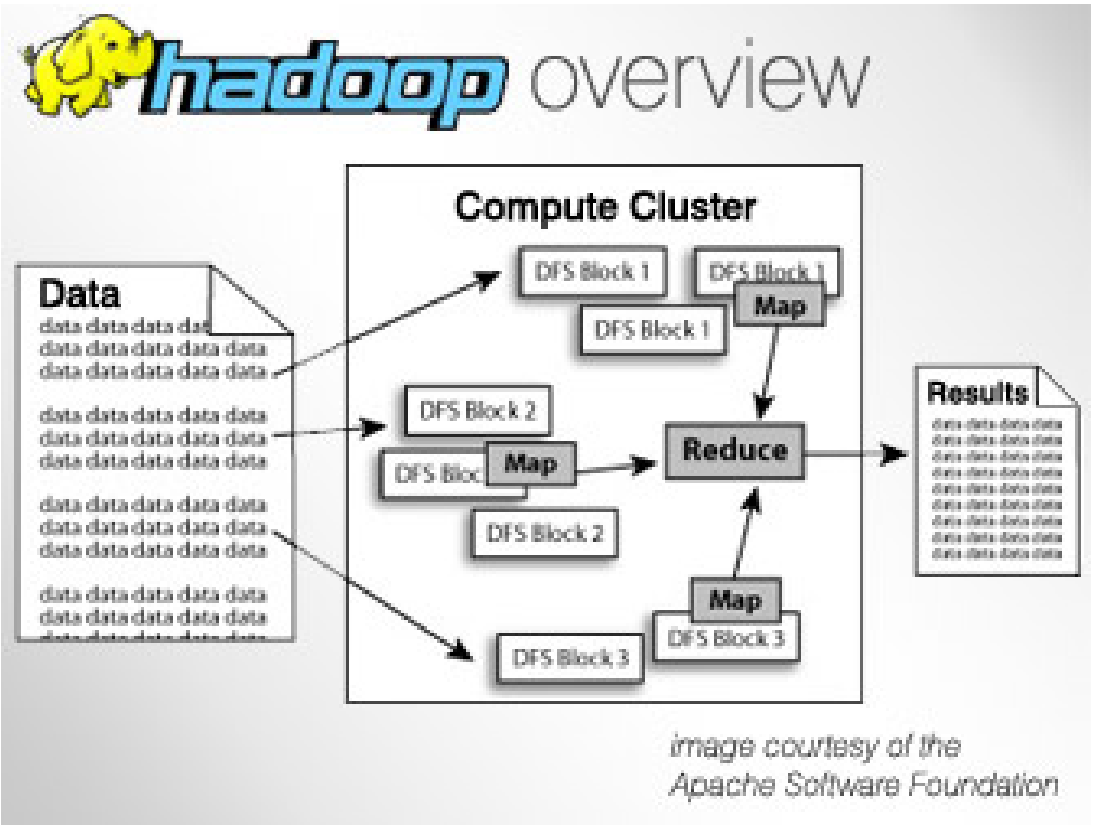


Doubling every 2 years

Tera: 10^{12} Peta: 10^{15} Exa: 10^{18} Zetta: 10^{21} Yota: 10^{24}

State: 2016

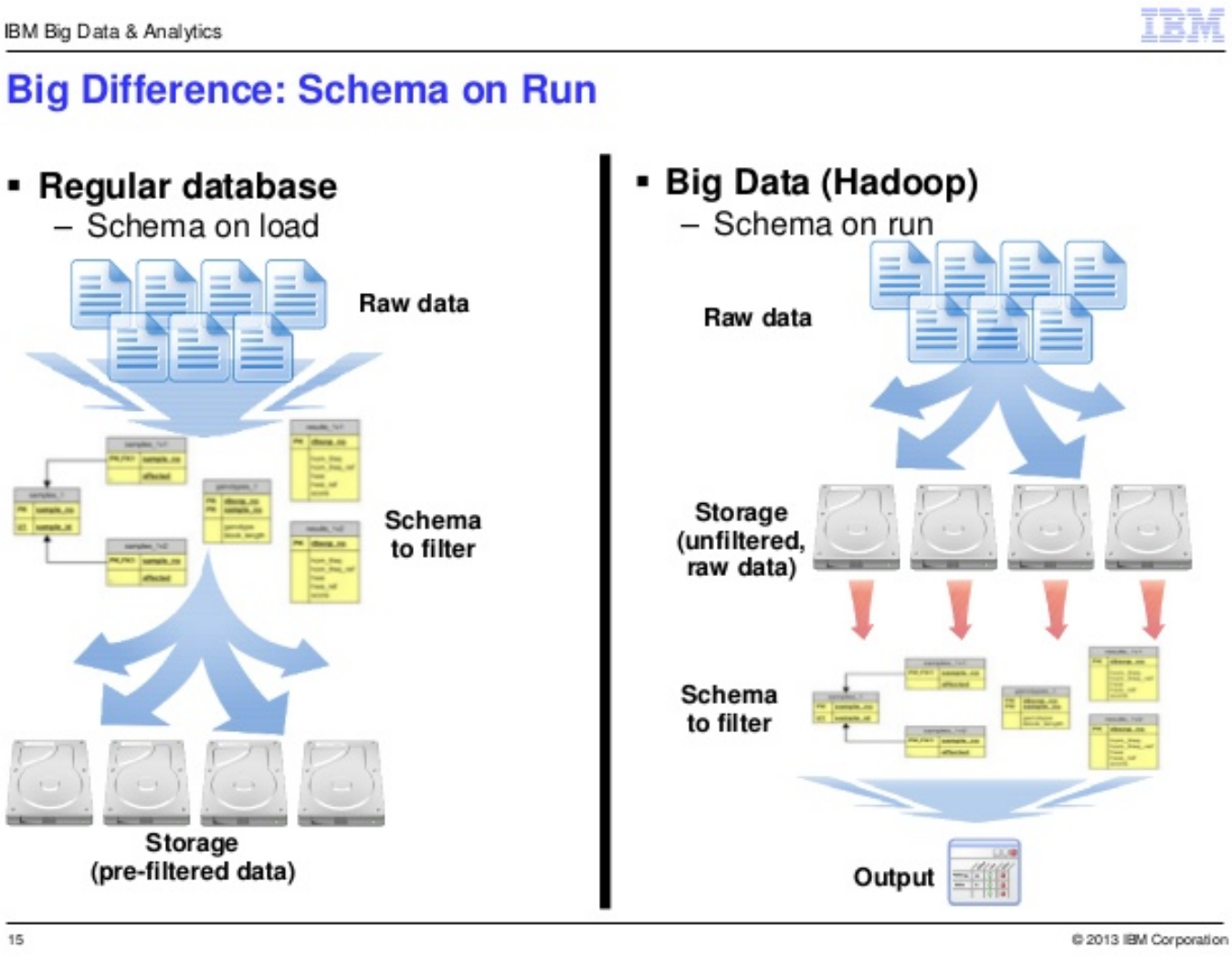
Volume: Storage of large amounts of data: Hadoop



- Framework for data storage and processing
- File system HDFS
- Redundant storage on several nodes
- Scalable
- Queries as jobs on several nodes („Map“)
- Consolidation of data in one „Reduce“ step

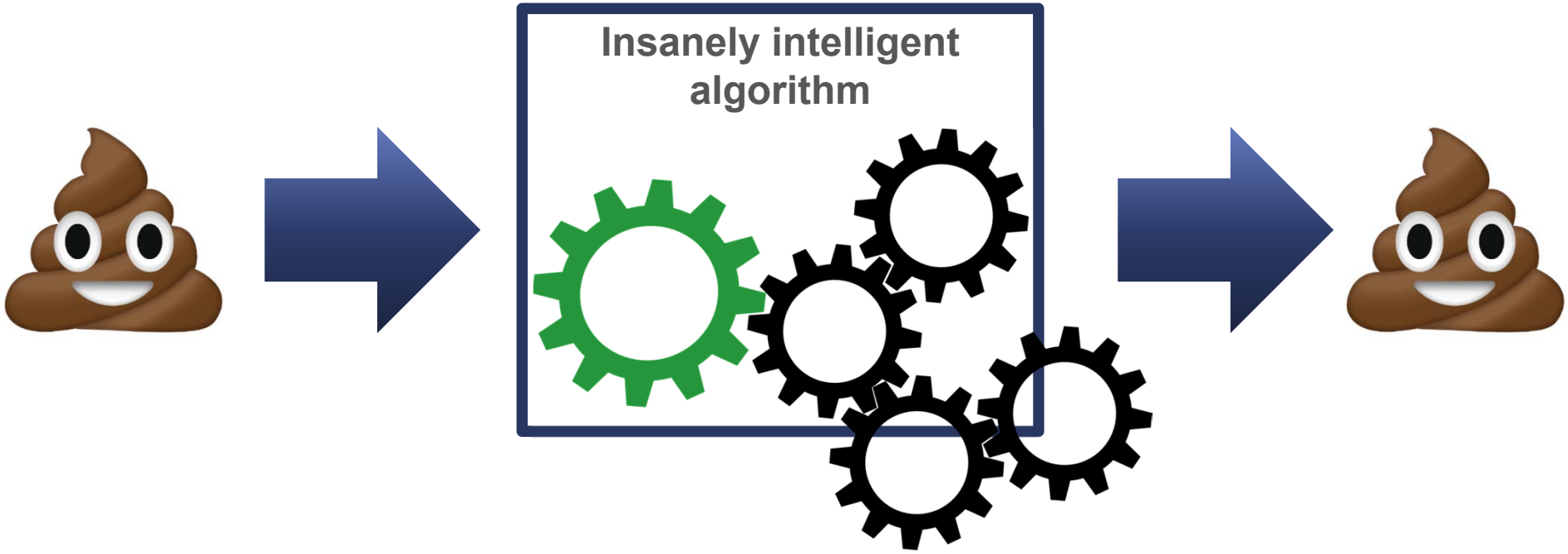
Source: http://www.learncomputer.com/intro-to-hadoop/

Volume: Difference relational DB ←→ Hadoop



Quelle: http://flatblack.co/big-data-database-architecture/big-data-database-architecture-lovely-on-architecture-intended-for-overview-4/

The garbage in / garbage out principle

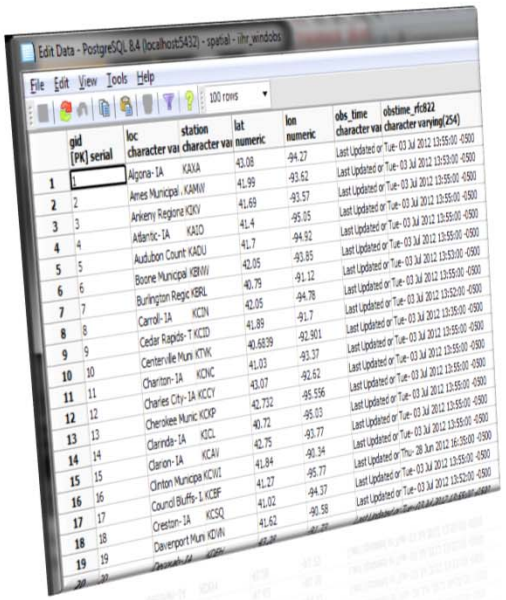


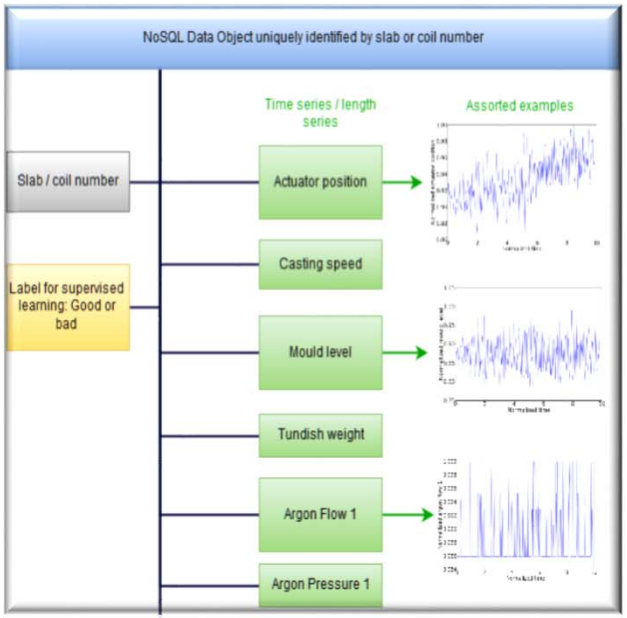
Variety: Information diversity

- Scalar values (product-related data)
  - Ident, Weight, Chemical analysis, Material class, Customer, Order, …
- 1D-Data
  - Length-/time-based measurements: Strip tension, Width, Speed, …
- 2D-Data
  - Thickness, Flatness, Temperature, Coating layer, …
- Events
  - Production defects (automatic/manual), Production faults, Maintenance actions, …
- Images
  - Surface defects, Ultrasonic testing, 3D scans, Mechanical testing, …
- Audio data
  - Gear noises, Plant vibrations, …
- Video data
  - Plant surveillance, Process monitoring,…

Steel Industry

Variety: Structured and unstructured data





structured semi-structured unstructured

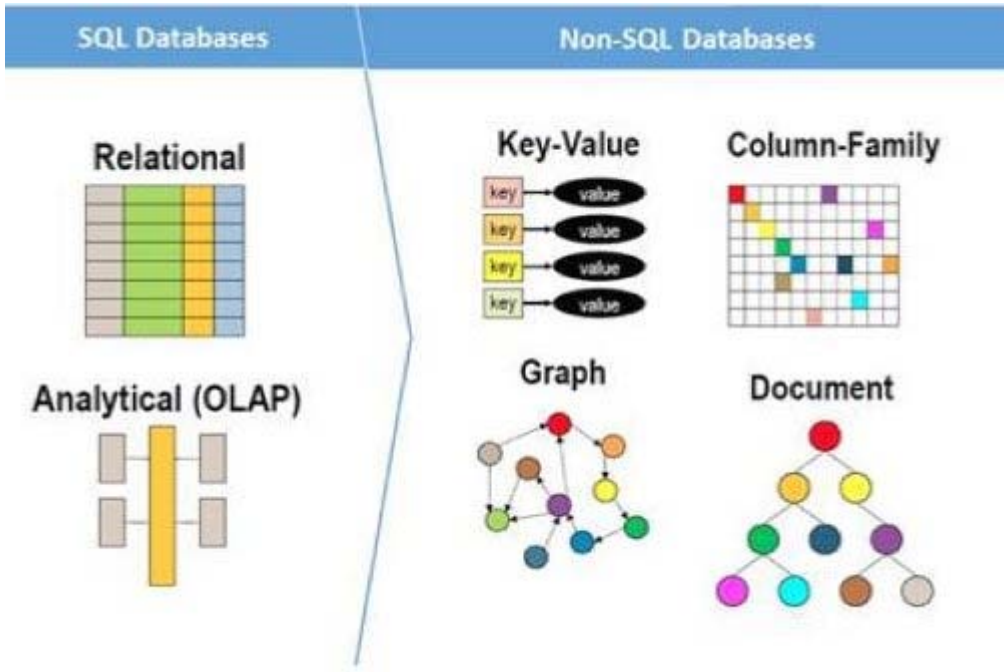
Variety: Storage of large amounts of data: NoSQL

- Object-oriented data storage
- No predefined data model necessary
- Storage of different data types
- Avoid unnecessary complexity (e.g. no transactions)
- Very easy to scale by adding nodes
- Storage of methods (e.g. models, visualisations)

Variety: NoSQL types

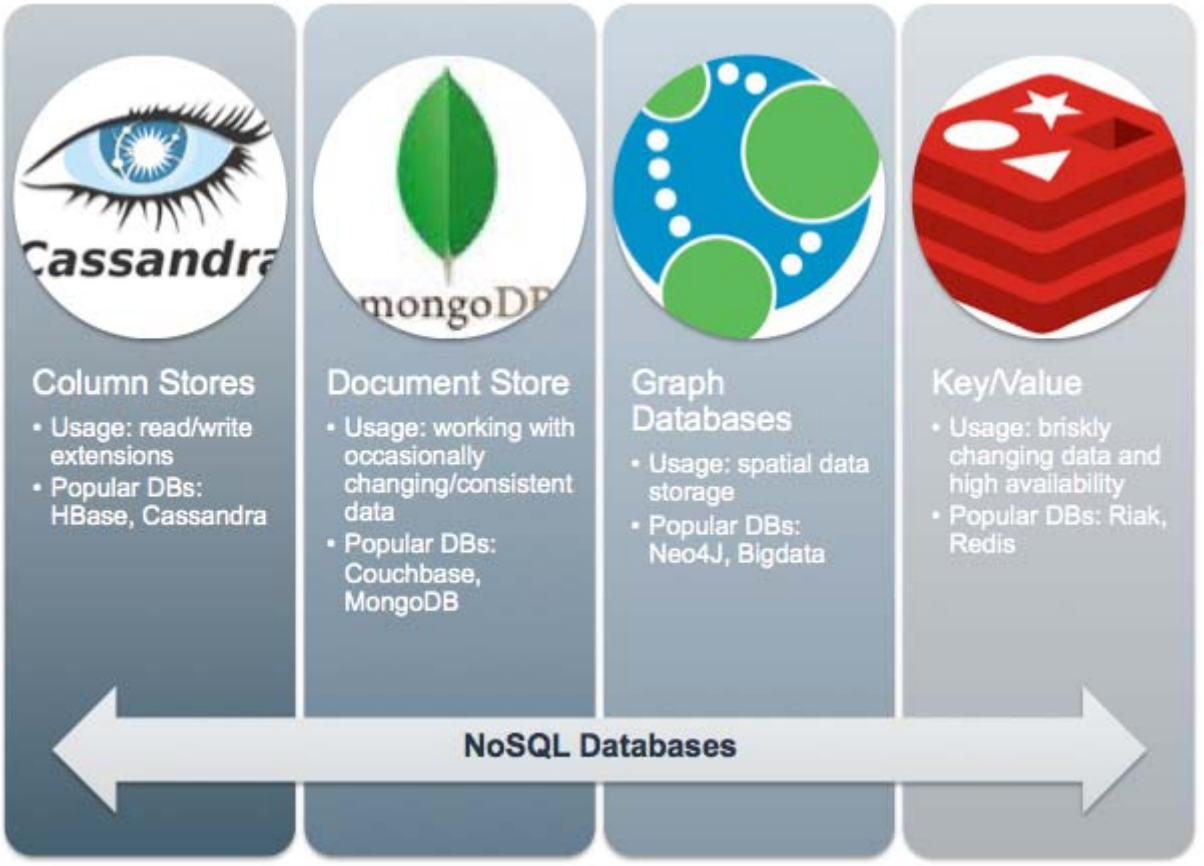
Types of NoSQL Databases

- Key-values stores
- Wide-column stores
- Graph databases
- Multi-model databases
- Document databases



Source: https://www.kdnuggets.com/2016/07/seven-steps-understanding-nosql-databases.html

Variety: NoSQL implementations



Source: https://blog.appdynamics.com/engineering/a-newbie-guide-to-databases/

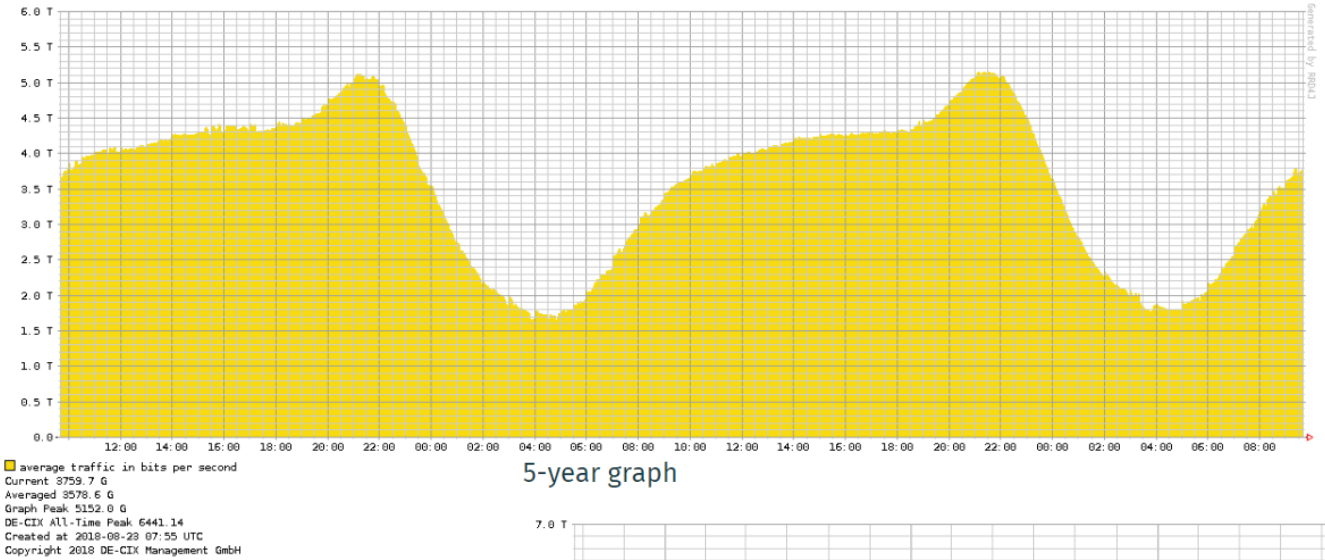
Variety: Application example: Digital Twin

(8) ObjectId("58cba1a2c9a1ba2ad093871f")	{ 12 fields }
_id	ObjectId("58cba1a2c9a1ba2ad093871f")
HotRollingMill	{ 13 fields }
T_Target_Coiler	610
Material	
T_Target_Coiler_Tolerance	10
T_Target_RollingEnd	840
T_Target_RollingEnd_Tolerance	10
t_Coiler_String	
t_Weigher_String	
NextStoragePosition	
Quality	
FileID	
HasDBEntryTimeStamps	False
Series	
TRG	0
PicklingLine	{ 3 fields }
T_Measured_DB	[461 elements]
FileID_DB	
t_Pickling_Measurement	2017-03-20 09:30:56
HasCoilerTemperatureData	False
STR	0
SeriesExtension	3
HasDBTemperatureData	True
Ring	
ID	

- Twins for products
- Twins for machines
- Twins for orders
- …

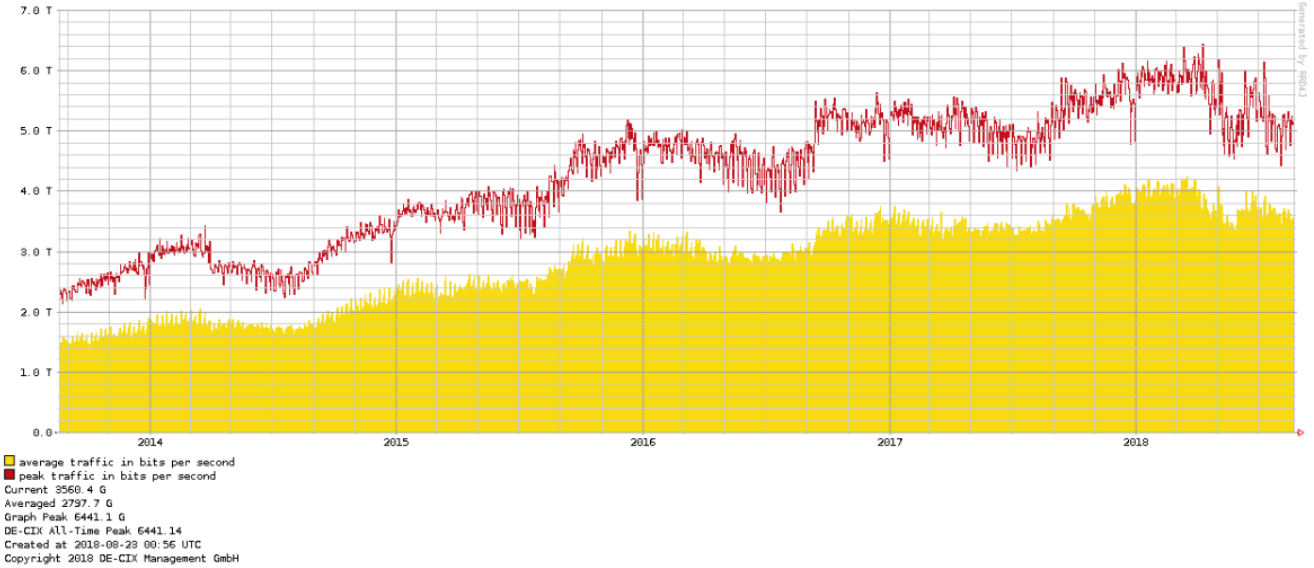
Velocity: Excursus: Data transfer in the Internet

2-day graph



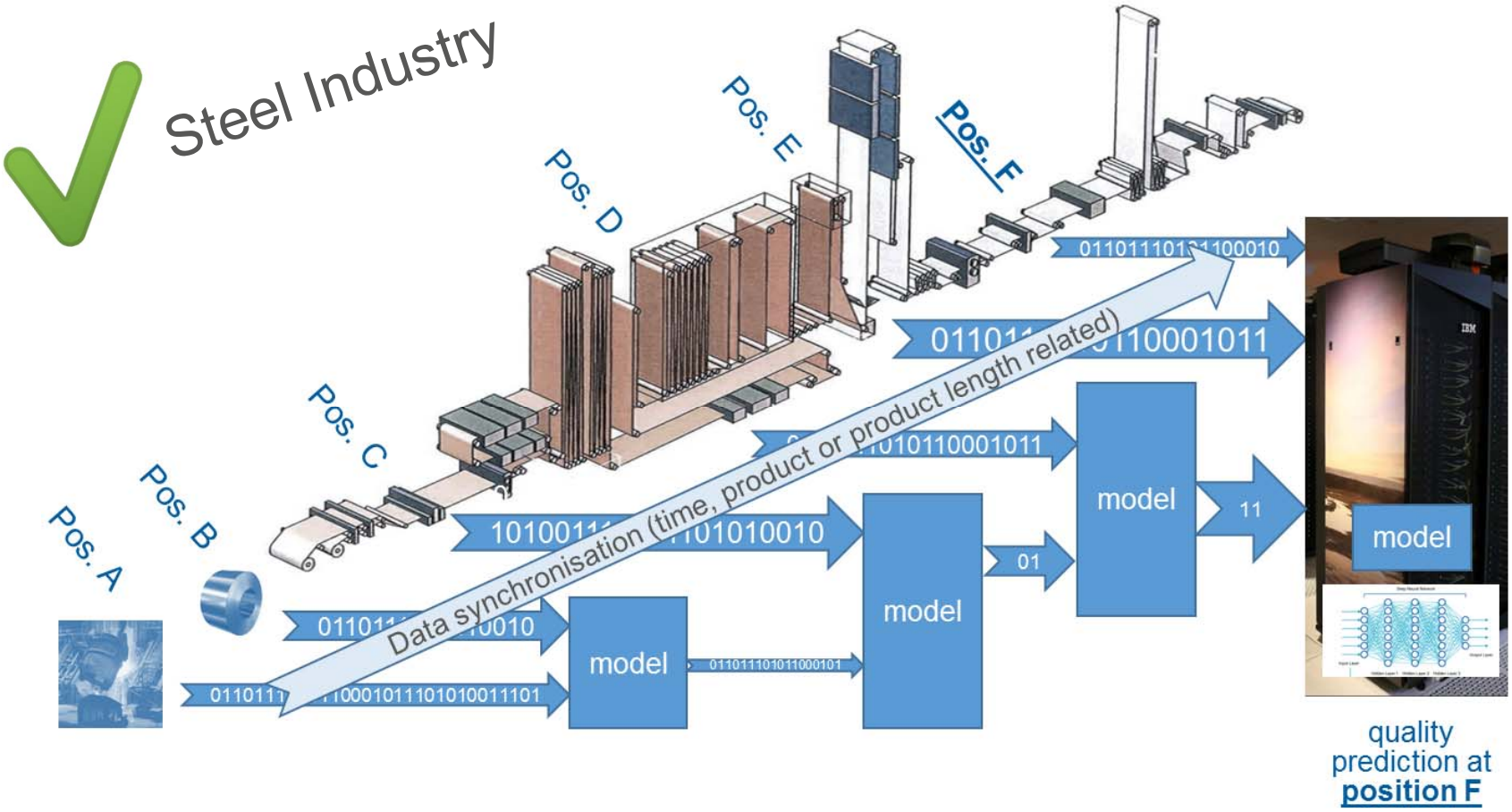
All Time high:
6,44 Tera-Bits/s

5-year graph



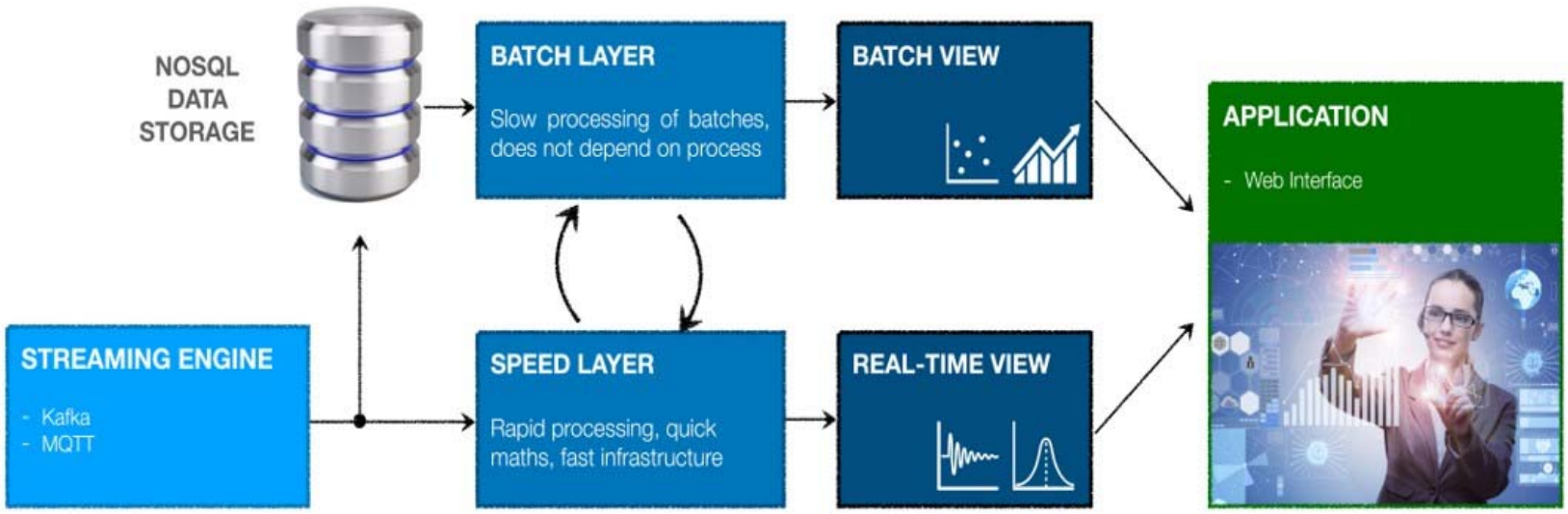
Source:
German Internet exchange node DE-CIX Frankfurt

Velocity: Processing speed



› Data must be processed process-accompanying.

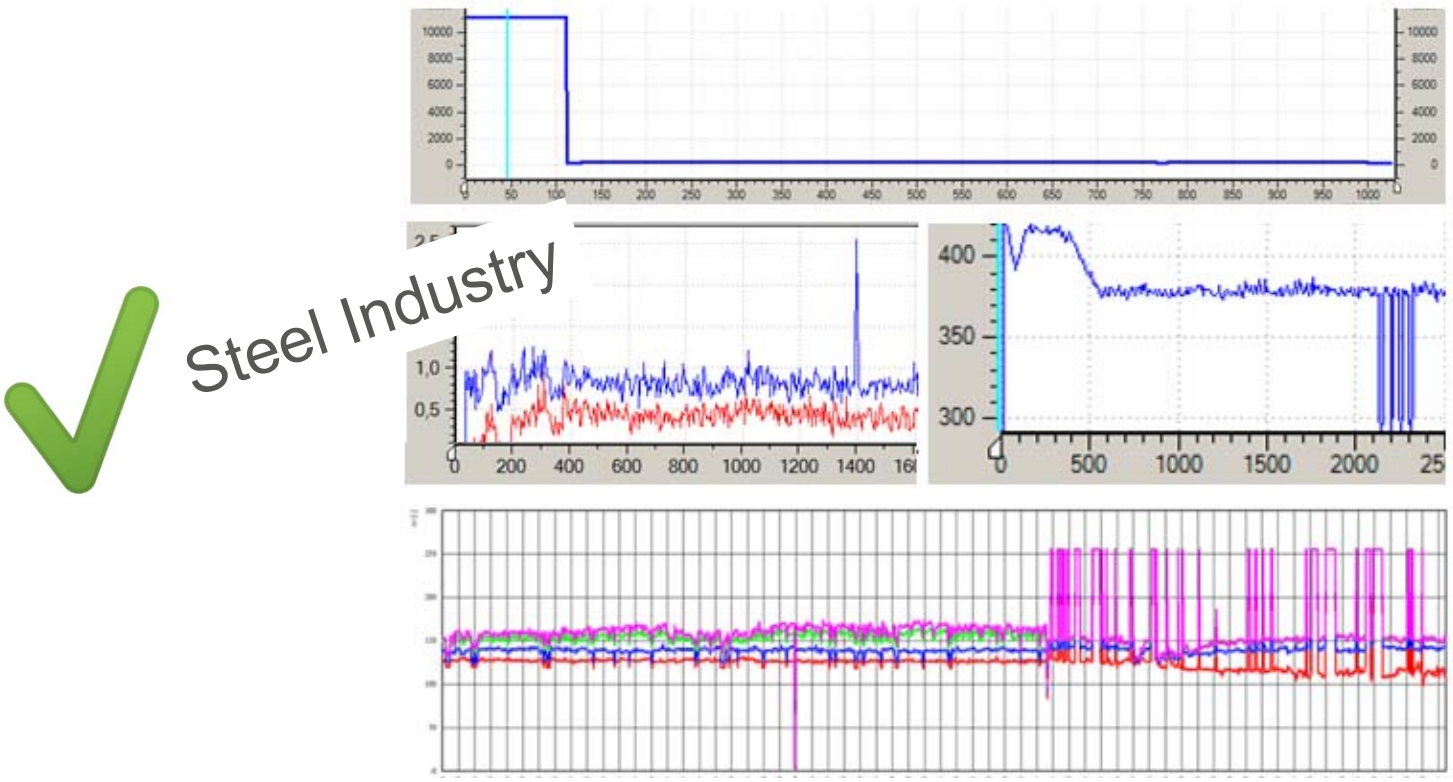
Velocity: Lambda architecture makes Big Data manageable



- **Prototypical infrastructure** for scalable Big Data applications (N. Marz et al. 2016)
- **Lambda architecture** enables dual approach
  - (a) slow evaluation of batches, mathematical/numerical intensive
  - (b) fast evaluation of real time data, reduced processing time

Veracity: Data quality

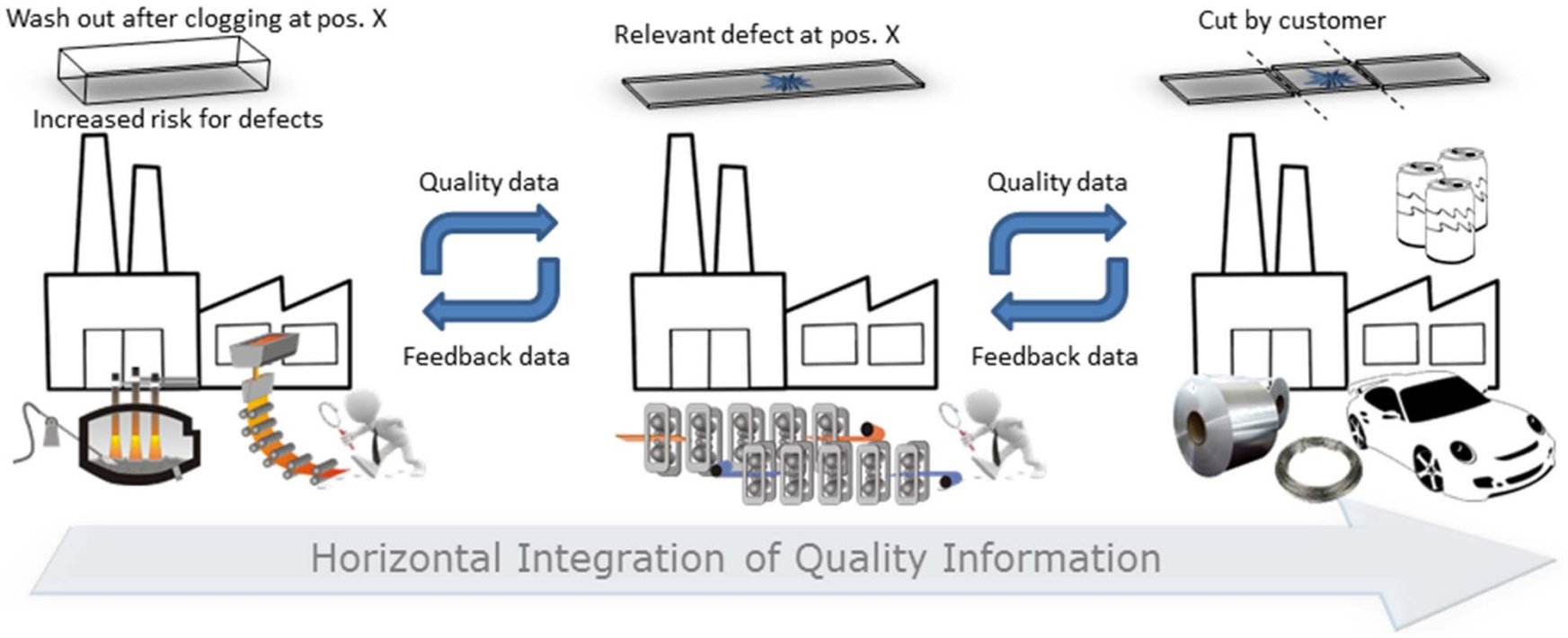
Examples of real process signals:



Data quality must be ensured before data usage

Veracity: Quality 4.0

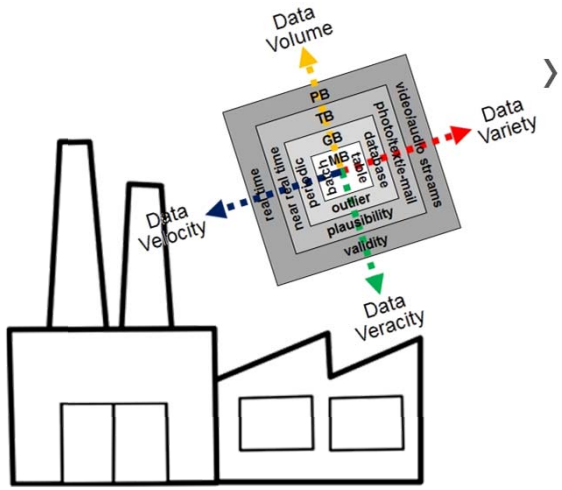
› Running RFCS-Project



Big Data Storage and Steel Industry

Potentials and applications

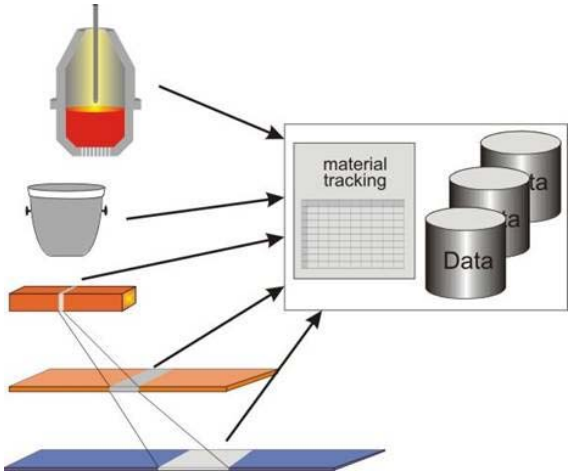
Requirements / Challenges of Big Data in Steel Industry

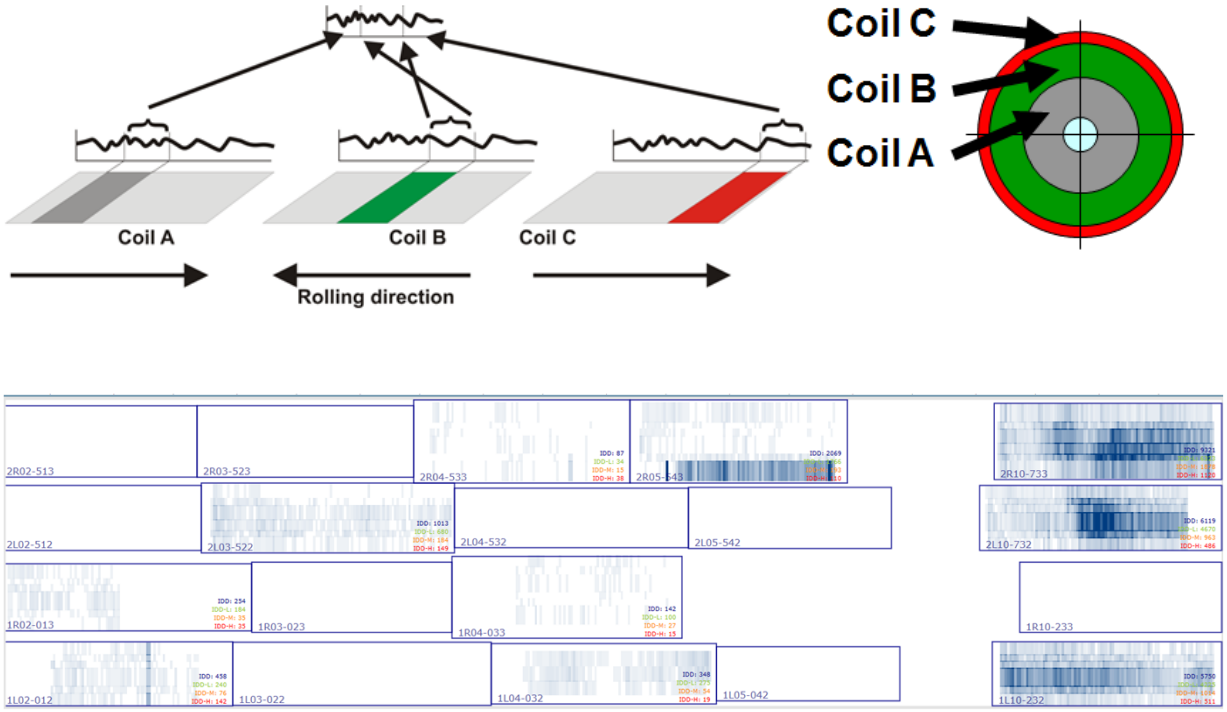


- Digitalisation
  - Heterogeneous age structure of plants in the steel industry
  - Investment requirements for data acquisition at older plants
  - Continuous data acquisition along the process chain necessary
- Data access
  - Data acquisition systems with proprietary data storage
  - Use of standardised interfaces
- Gapless material and position tracking
  - Mandatory for the correct assignment of data to the correct position

Material tracking

Gapless material tracking is mandatory





BFI has developed tracking solution starting from heat

Industrial Big Data

Data Sources

Volume

Variety

Velocity

Veracity

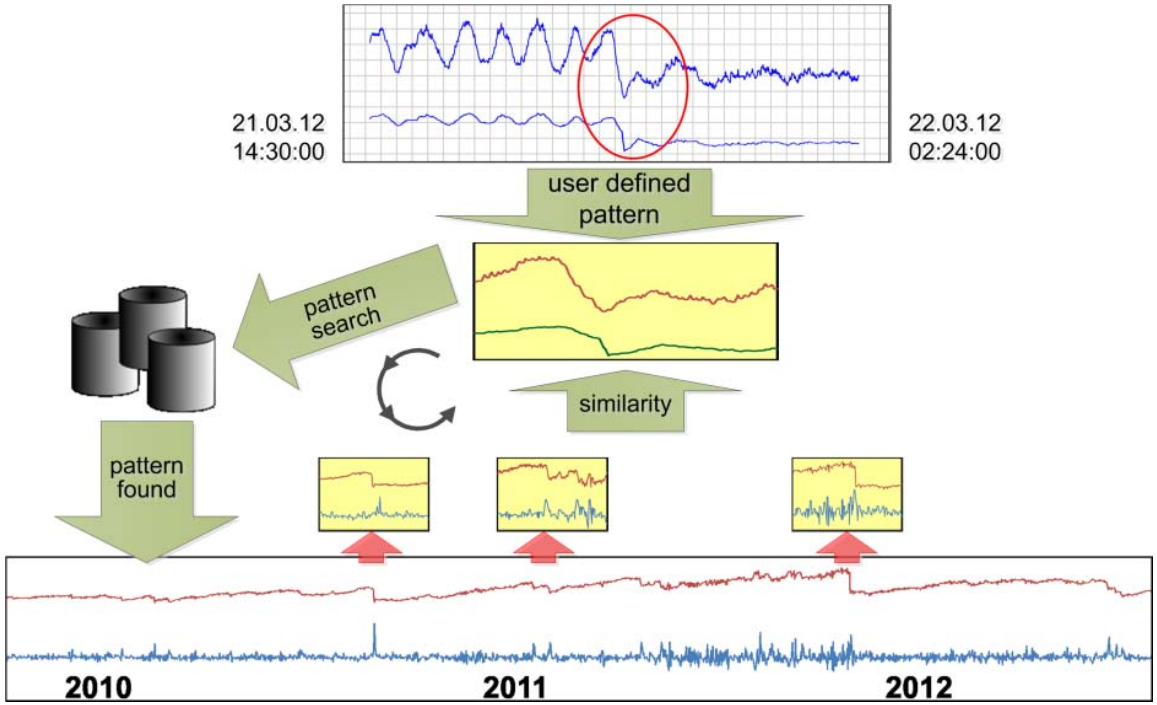
Fast

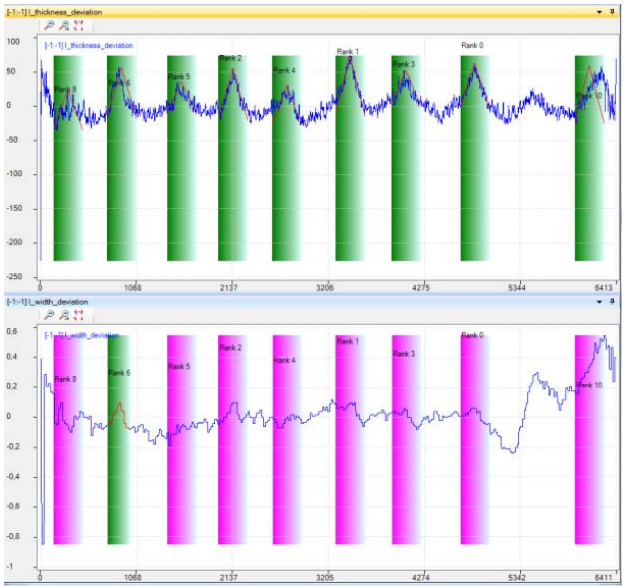
Flexible

Focused

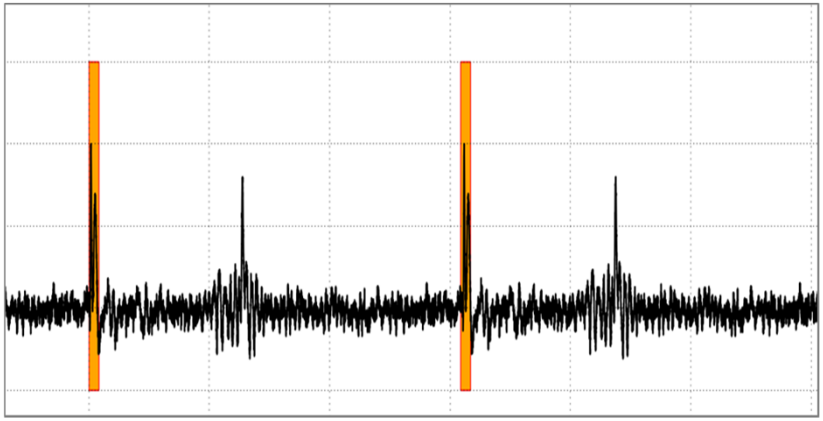
Data Access

Situation detection in large amounts of data

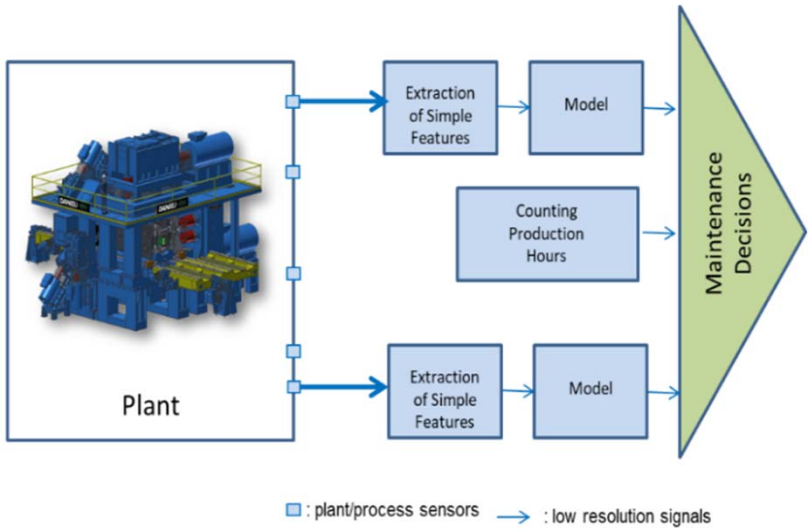


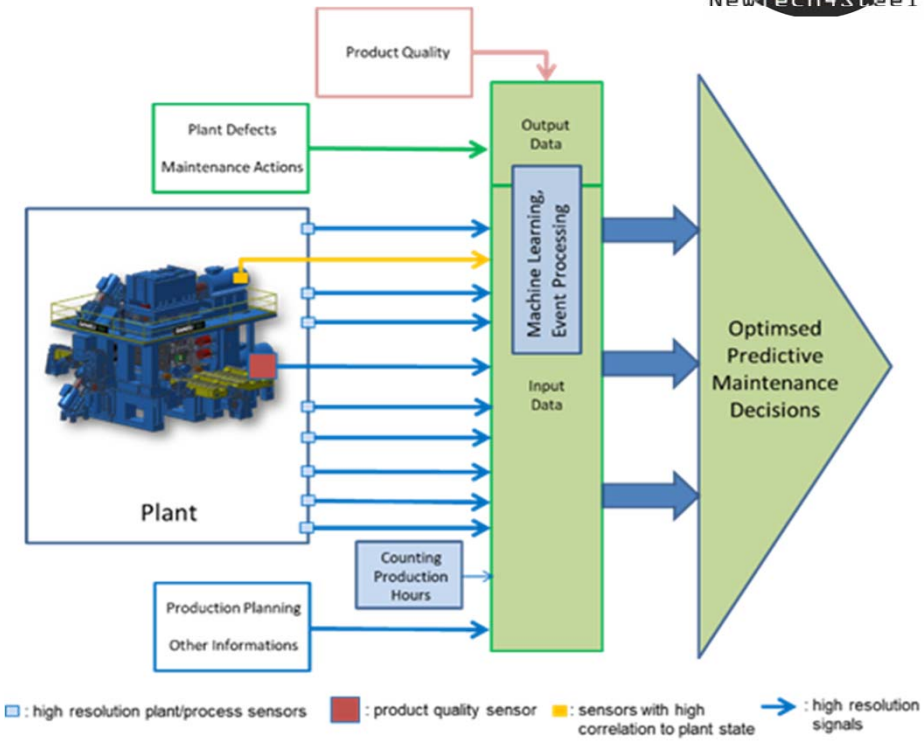


Multi-dimensional, very fast search in large amounts of data



New approaches to predictive, data-driven maintenance

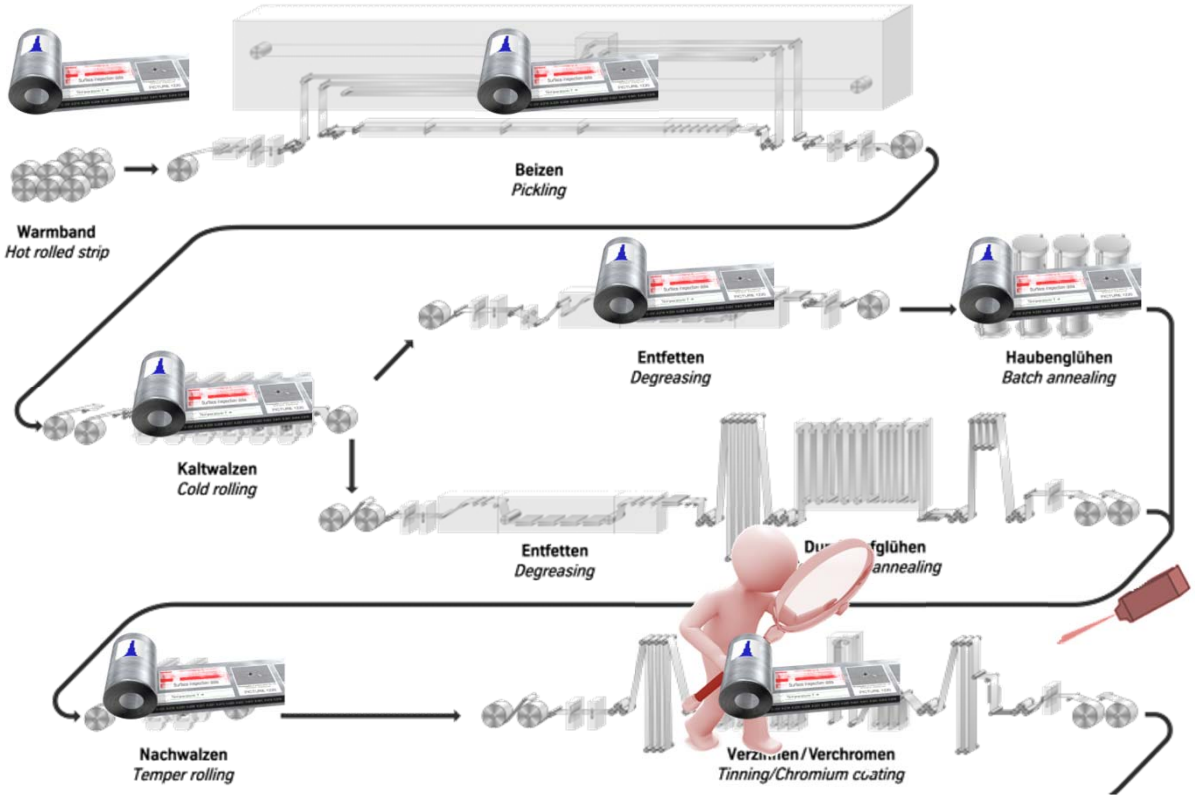




- New Asset Management Strategies
- Technology: Machine Learning, Event Handling, Intelligent Sensor Technology
- Dependencies between product quality and plant condition

Example: fast visualisation of process and product data

Complex production flow



Coil changes during processing

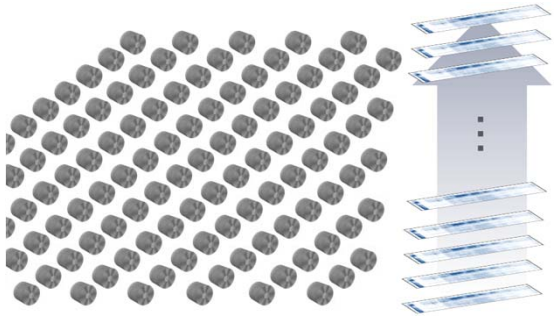
› Different viewpoints require different data

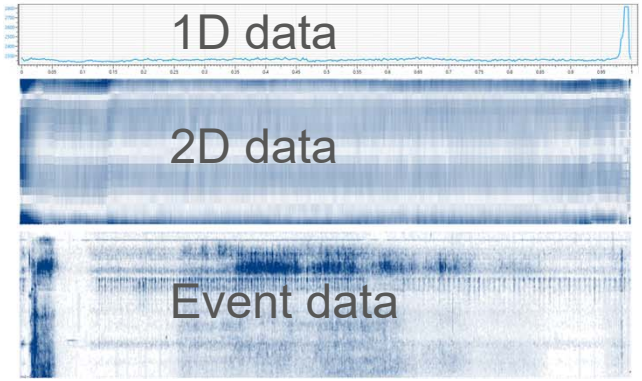
For full production analysis

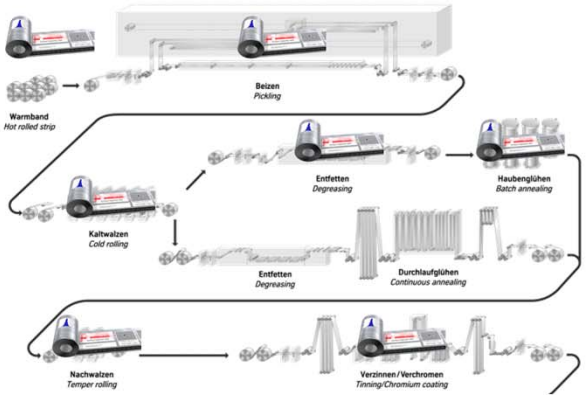
› Material tracking must be considered

Example: fast visualisation of process and product data

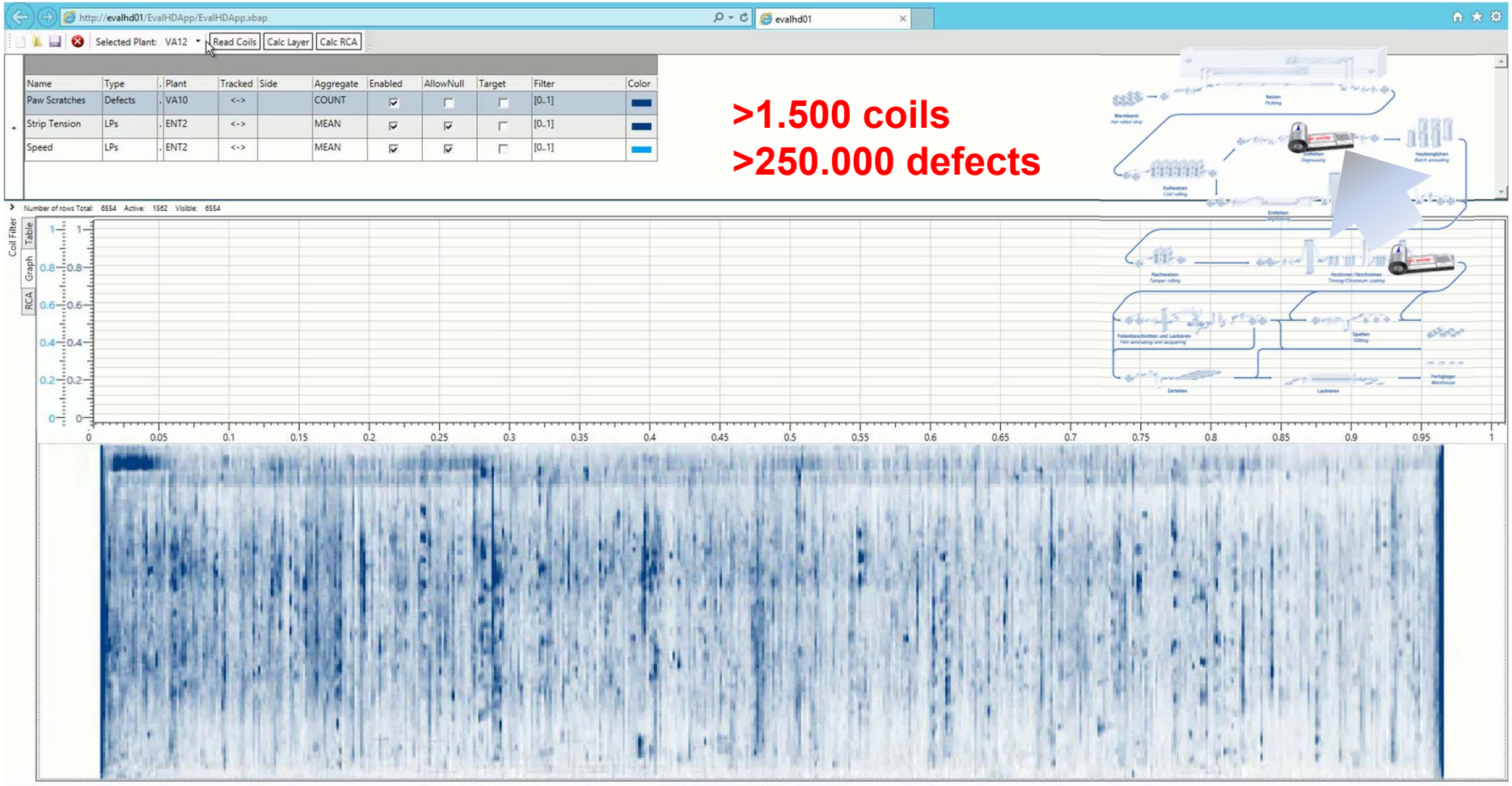
- Multi-dimensional problem
- Concept specialised for manufacturing industry required







Example: fast visualisation of process and product data



BFI HR-Server: Realised with standard technologies

Production lines

Hot strip mills (2)
Coating Line
Pickling line (2)
Batch annealing (2)
Stretcher Leveler (2)
Cold rolling mills (2)
Continuous annealing lines (3)
Cleaning lines (2)
Temper mills (3)
Chromium line (1)
Tinning lines (4)

2D-Data

Thickness
Flatness
Coating Layer
Temperature

Event-based

Hole- /
Edge crack detection
Internal defect detection
ASIS

1D-Data

Bending, Center shift
Charge density
Coating layer
Conductivity
Current density
Flatness, Flow,
Internal defect detection
Oil layer, Other, pH-Value,
Pivoting, Positioning roll
Pressure, Reduction ratio
Residual current
Rolling pressure
Rotation Frequency
Skin pass, Speed
Stretcher Leveler
Strip tension
Technology (IMPOC)
Temperature
Thickness, Width

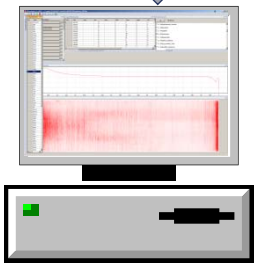
BFI HR Data Server

DataStudio Import

Material tracking information

Order information

Visualisation



~1.200 data sources; ~1,5 years production in direct access

Thank you for your attention