



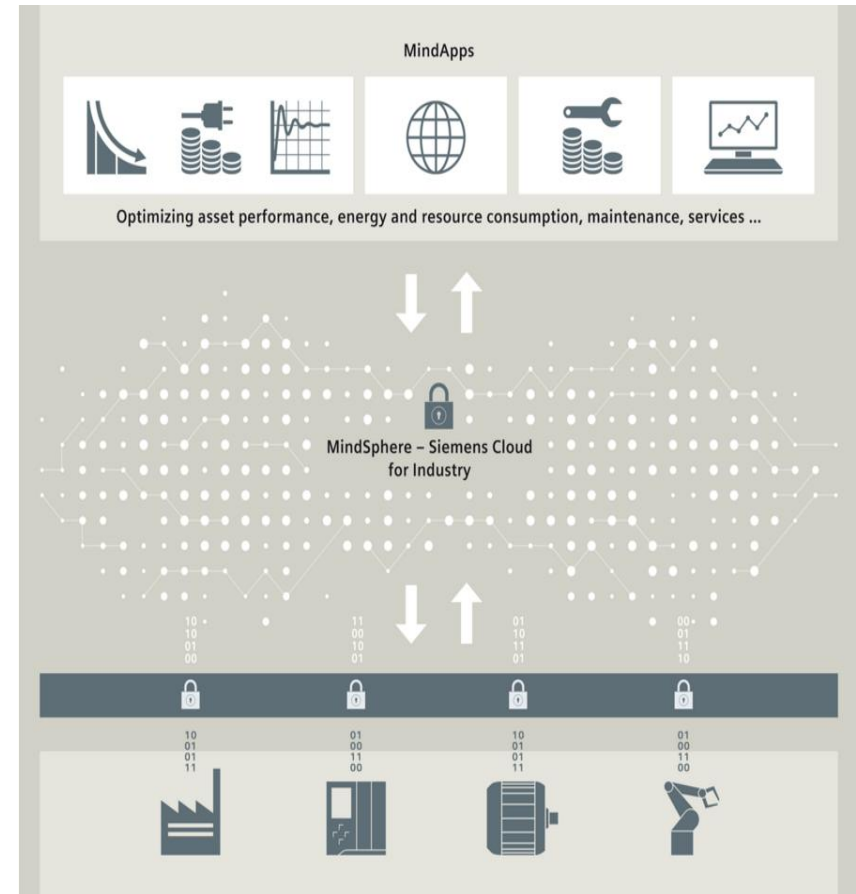
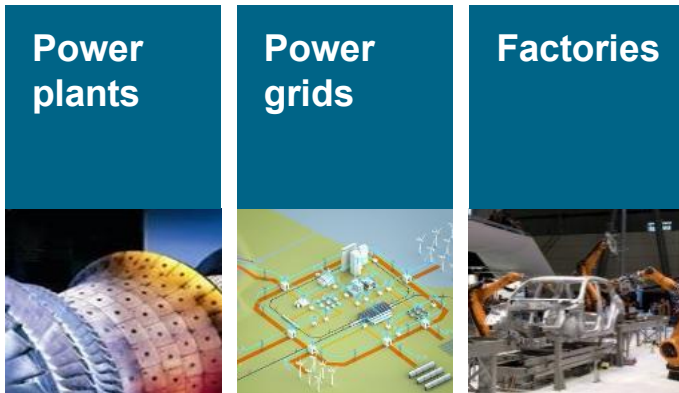
SIEMENS



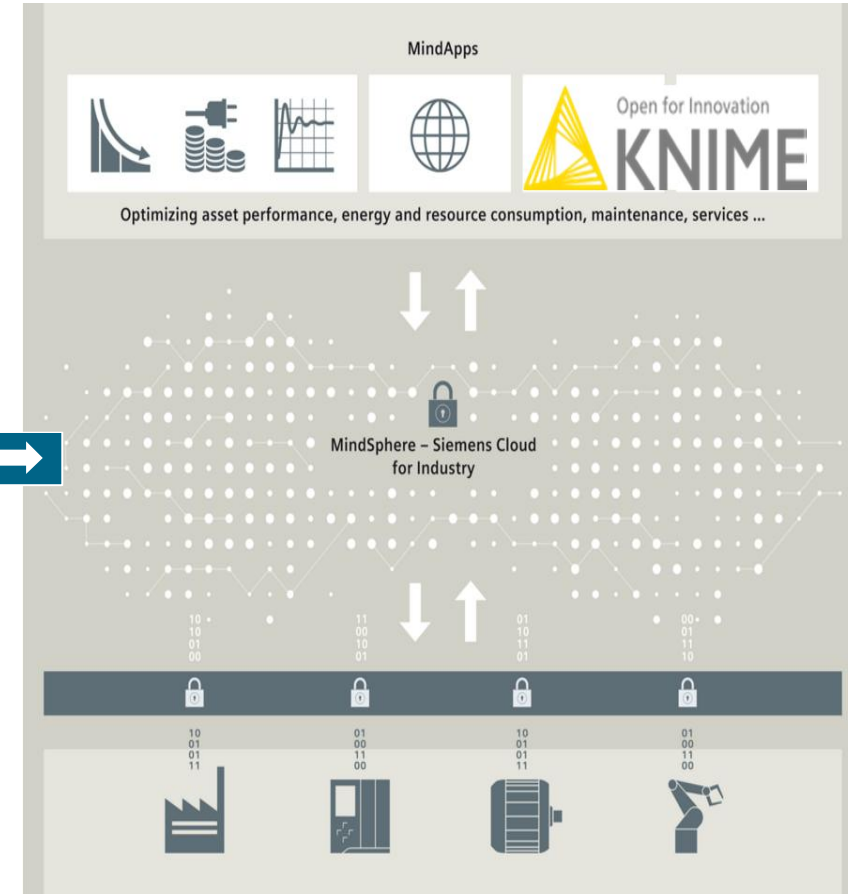
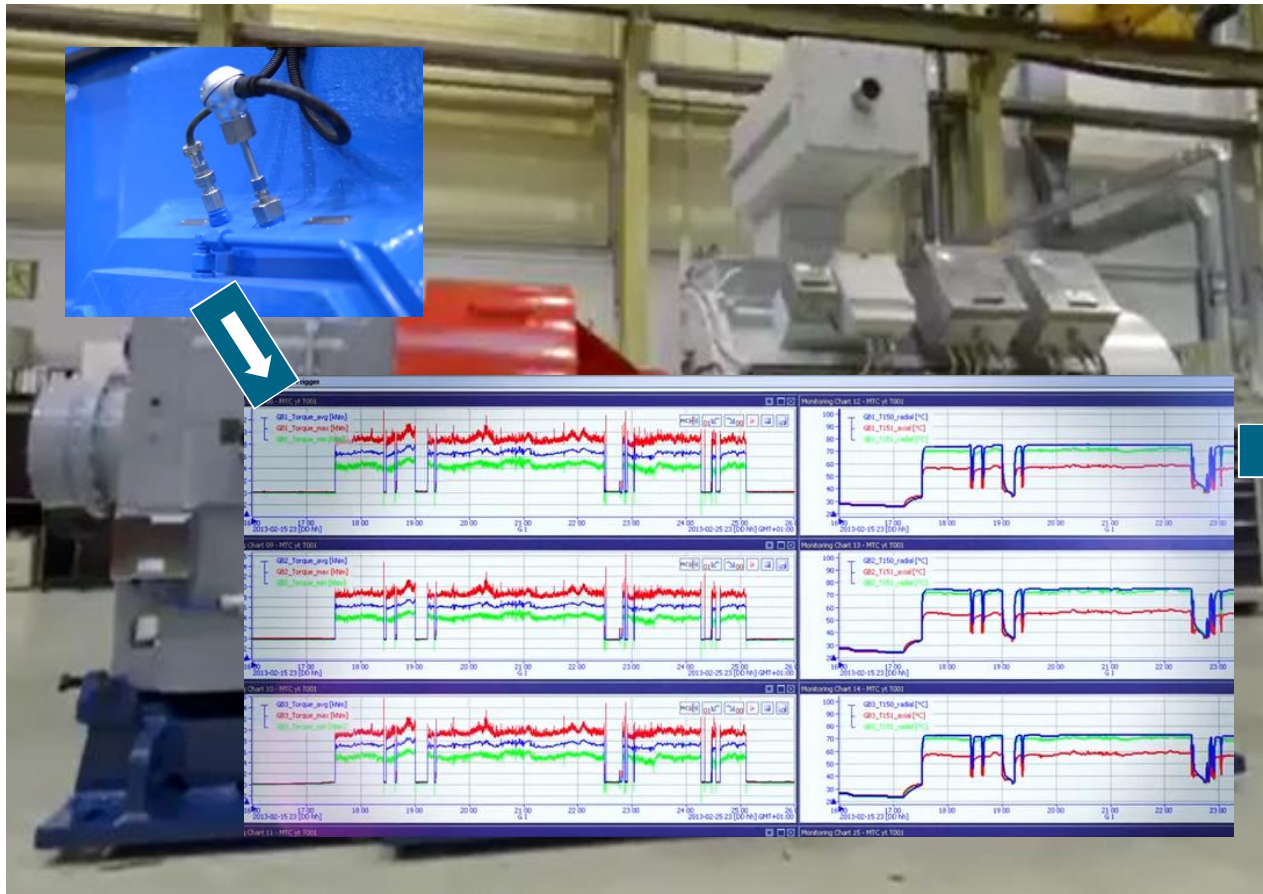
Big Data@SIEMENS a KNIME Story

Jan Pospisil – Head of Data Analytics, Digital Factory

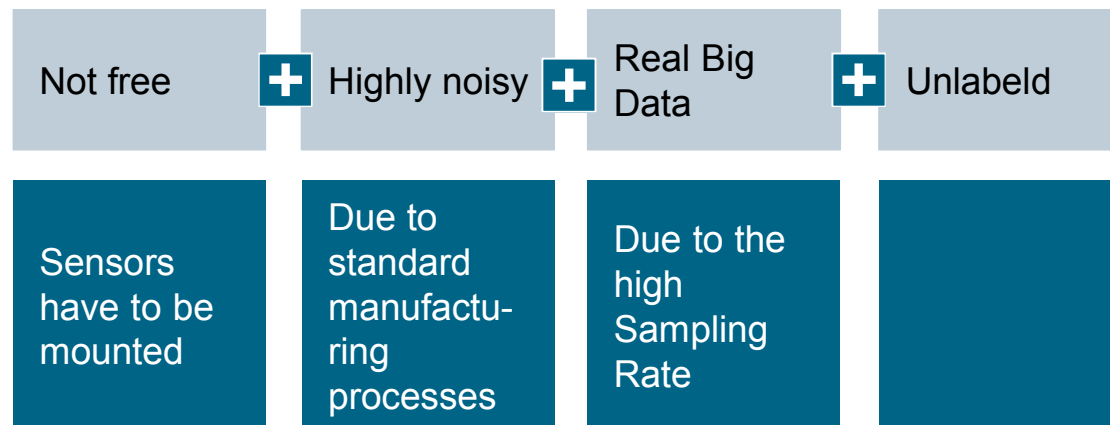
How to Monitor Thousands of Devices through the Industrial Internet? How to deal with Big Data?



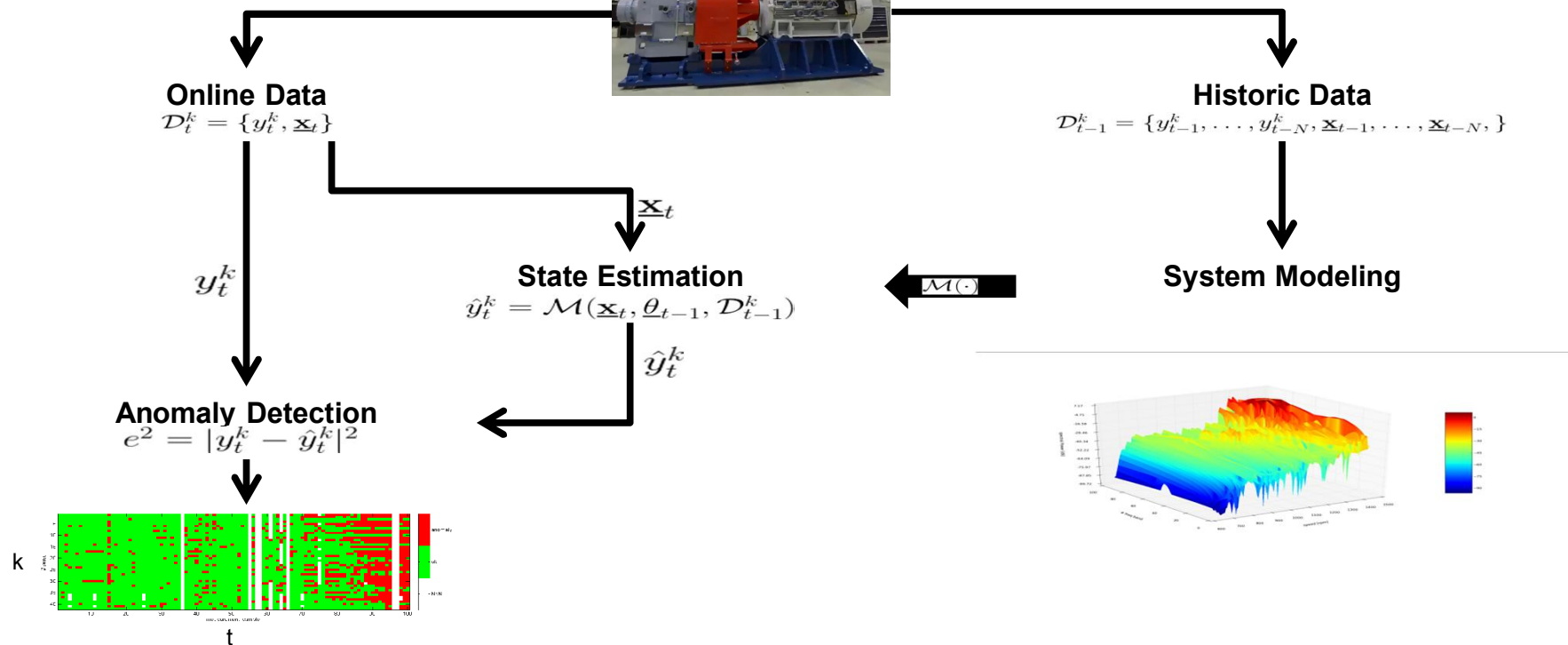
Drive Train Anomaly Detection in Siemens Cloud for Industry - MindSphere



Second Differentiator: The Data and Data Processing



Drive Train Anomaly Detection on Unlabeled Data in MindSphere



Calculation of Amount of Data for 1,000 Rotating Devices

Number of Sensors	10
Signal Derivatives per Sensor	5
FFT Points	16384
Samples per Day	288
Total amount of Data per Day	2GB



0.736 PByte
per Year
per 1,000 Assets



Need for a nearly
linear scaling Data
Storage combined
with nearly linear
scaling Compute
Power
Data Science



Consequence:
Integrate Knime
Hadoop!
and other NoSQL
Technologies

KNIME Hadoop Integration (some important components)

Module	Functionality
YARN	<ul style="list-style-type: none">▪ Locate available compute resources▪ Locate available IO resources▪ Leave Workflow control at KNIME▪ Hide resource control from KNIME user
HBase	<ul style="list-style-type: none">▪ Transform huge/sparse matrixes which do not fit into RAM▪ Abstract/hide HBase Layer from KNIME user
Phoenix	<ul style="list-style-type: none">▪ Variable schema SQL on HBase▪ Abstract/hide Phoenix Layer from KNIME user
Drill	<ul style="list-style-type: none">▪ Schema- free SQL on Hadoop▪ Abstract/hide Phoenix Layer from KNIME user
Spark	<ul style="list-style-type: none">▪ Distributed in memory computing▪ Micro batch streaming

Contributors

Tobias Kötter

Chief KNIME Architect

Arnulf Betz

Chief Hadoop Architect

Henning Ochsenfeld

Chief Data Scientist

... and many more ...

The image features a Siemens logo in the top left corner. The background is a blurred scene of a trade show or exhibition with people and various displays. A large digital screen in the center shows a complex industrial interface with several text prompts: "How much performance and energy efficiency is hidden in your plant?", "How secure is your data and your intellectual property?", "Have you tapped all sources of additional revenue?", "Do you have the highest required level of plant availability?", "Do you know exactly where you are expending material and energy cost?", "Improve quality and throughput and reduce costs - how does that work?", and "Do the typical energy guzzlers really offer the best savings potential?".

SIEMENS

Thank you for your attention