

IT Management with a Digital Twin

Advancing Big Data Analytics through Embedding of a Data Hub

Dr. Sebastian Appelhans | thyssenkrupp AG  
March 26<sup>th</sup> / 27<sup>th</sup> 2019 | Big Data Minds 2019 | Brühl

engineering.tomorrow.together.



thyssenkrupp

thyssenkrupp is a globally diversified industrial company within a portfolio of very heterogeneous business models operating within 12 key industries

Automotive



Civil Works & Infrastructure



Minerals & Mining



Chemicals



Power Generation & Distribution



Household Appliances



Food & Beverages



Aerospace



Systems Engineering



Oil and Gas



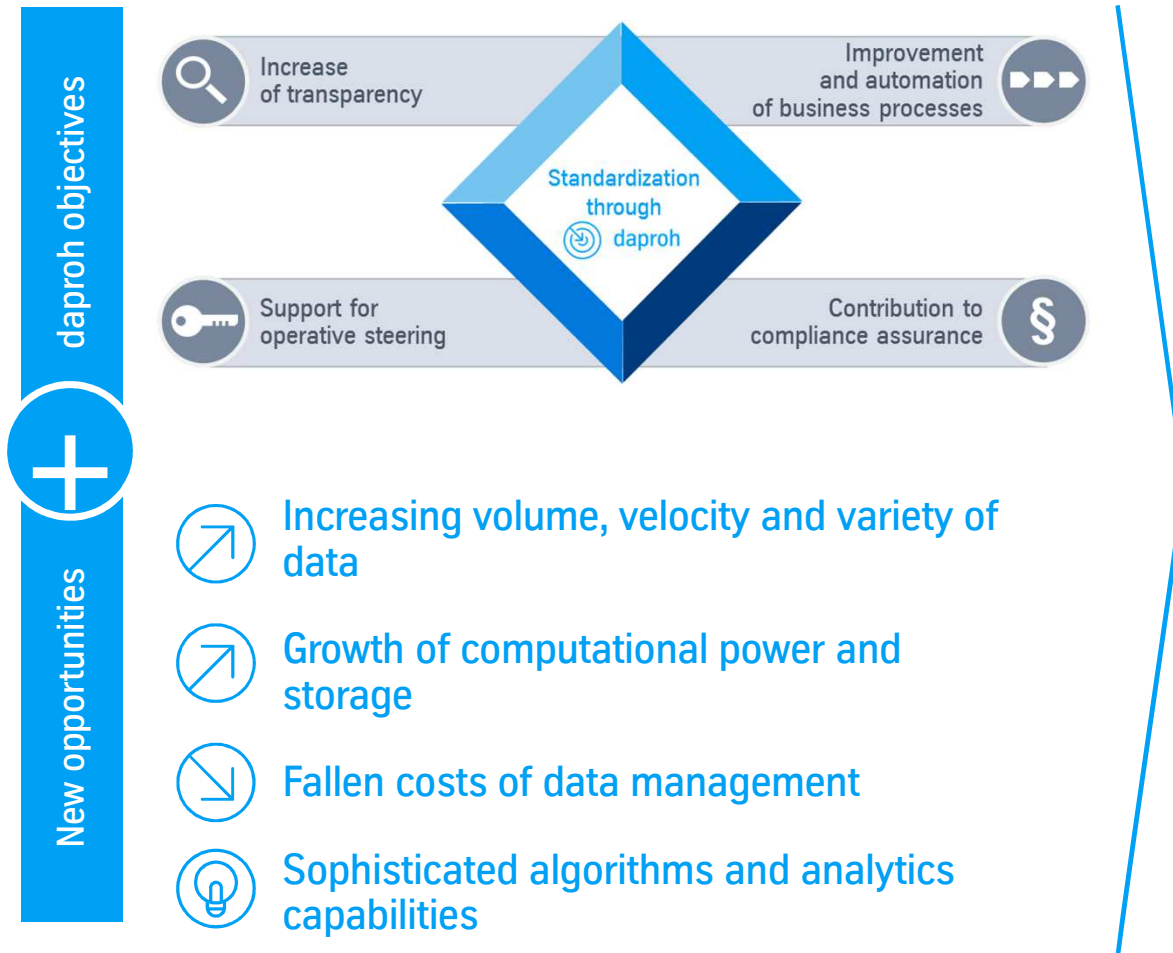
Ship Building



Special-purpose Vehicles



The deep integration into the corporate-wide data and process harmonization program provides the administrative and infrastructural framework necessary for the operationalization of industry 4.0



### Cost reduction

Improvement of the efficiency of operations

### Improved decision making

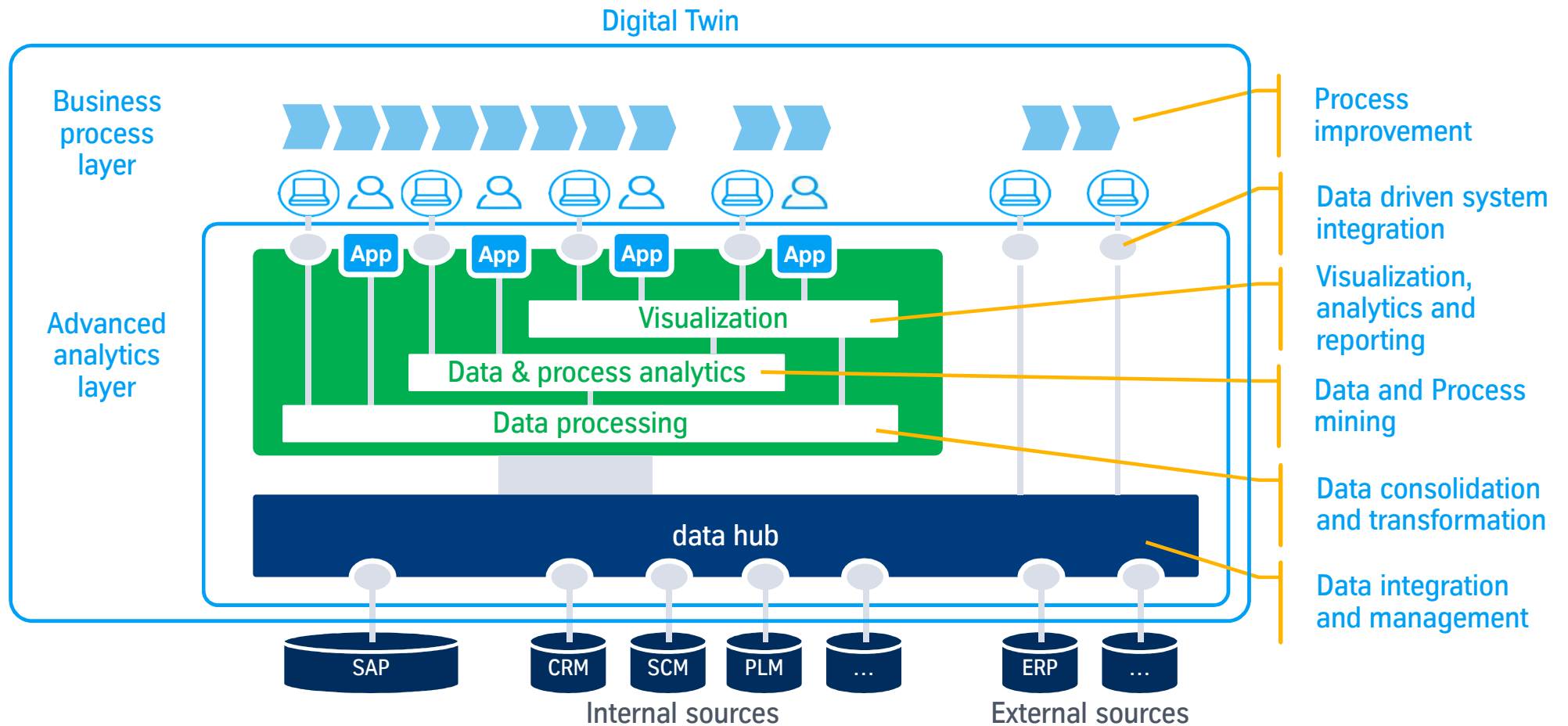
Providing insights on the impact of different variables in enterprise processes

### New products and services

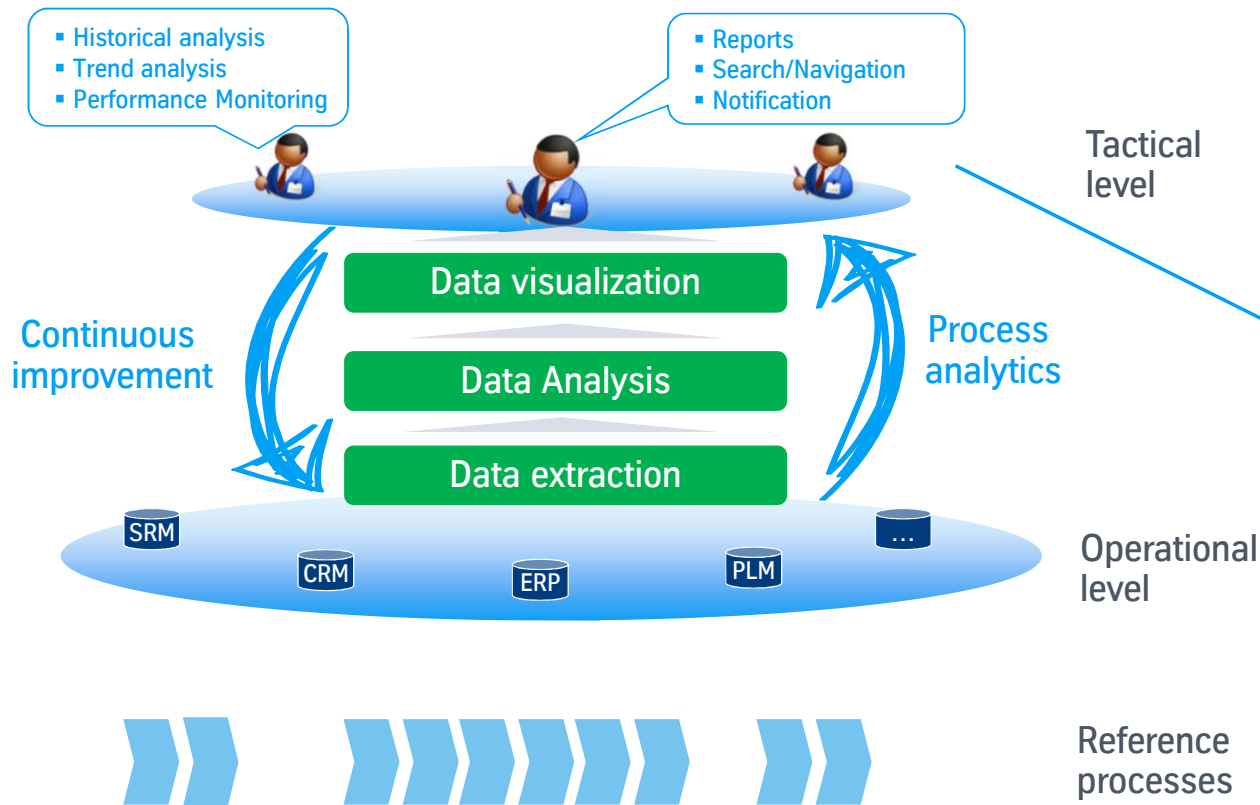
Enabling data driven data intensive business models



# A virtual suite serves as integrated service platform combining data extraction and managing with extensive data analytics and visualization capabilities

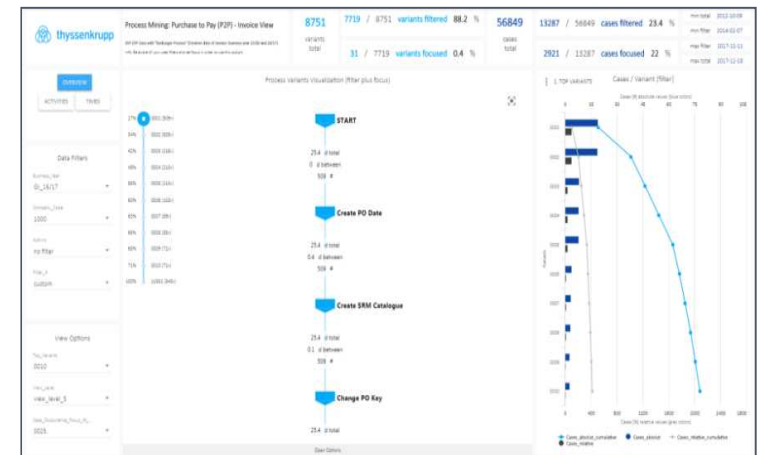


# Decision makers on tactical level need appropriate data provision from different sources and support by the analysis of business processes and simulation of measure implementation



## daproh big data cockpit blueprints

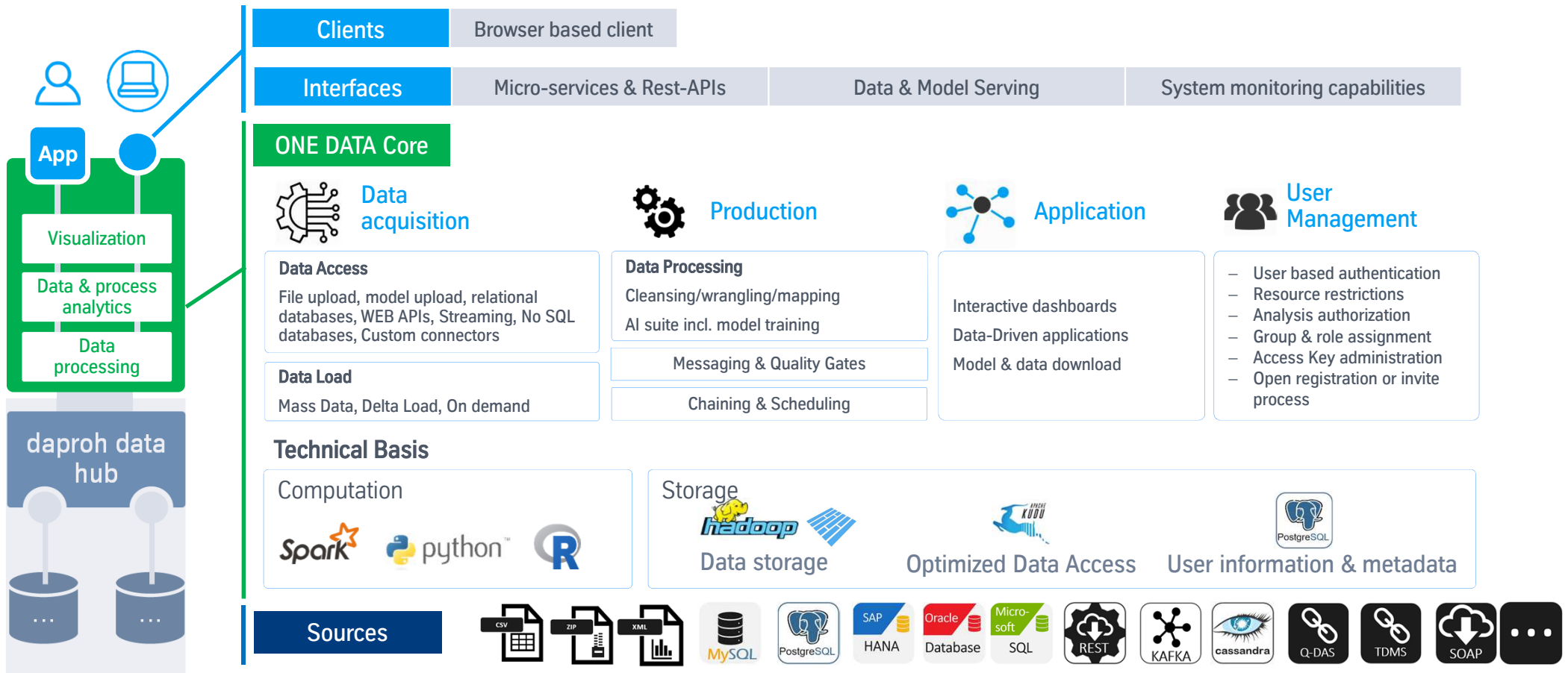
Intuitive and user centric user interface which avoids the complex process of data acquisition and allows different roles to access, to integrate, and to analyze enterprise data within a comprehensive cockpit.



e.g. Purchase-2-pay cockpit



Based on a modern Apache Hadoop and Spark architecture, the allows the processing and analysis of all thyssenkrupp's data formats and works within as well as non-daproh environments



# Central element of our architecture is a multi-client capable Data Hub that centralizes data access, data processing and quality assurance and provides role-specific access to published clean data

## Areas with high necessity for action

<b>Data</b>	<ul style="list-style-type: none"> <li>▪ Access to different data sources via a central instance</li> <li>▪ Transparency of data sources and their quality</li> <li>▪ Quality assured and documented data</li> </ul>
<b>Models</b>	<ul style="list-style-type: none"> <li>▪ Flexibility and efficiency in building models</li> <li>▪ Reproducibility and reusability</li> <li>▪ Transparent stock of models and workflows</li> </ul>
<b>Flexibility &amp; Quality</b>	<ul style="list-style-type: none"> <li>▪ Balance between flexibility and efficiency in conducting data science projects</li> <li>▪ Clearly defined quality criteria</li> </ul>
<b>Accountability</b>	<ul style="list-style-type: none"> <li>▪ Clear roles and responsibilities in the field of data governance</li> </ul>

## Solution

Core elements of the architecture

	<b>Data Hub</b>	A virtual centralized access- and management-layer for data sources with defined quality properties, including documentation
	<b>Processing Library &amp; Model Hub</b>	A centralized access and management layer for the integration of different models and workflows
	<b>Data quality assurance</b>	Defined and documented release processes that ensure data and model quality
	<b>Governance</b>	Roles for defining, checking and assuring quality standards and ensuring efficient and logical data flows and storage



Thank you

for your attention!

Dr. Sebastian Appelhans

engineering.tomorrow.together.



thyssenkrupp