

# EMuDig4.0

Effizienzschub in der Massivumformung durch Entwicklung und Integration digitaler Technologien im Engineering der gesamten Wertschöpfungskette

## Aluminium



Cloud-Infrastruktur zur Datenanalyse von Produktionsdaten aus der Massivumformungsbranche



# 3. Ergebnisse des Forschungsprojektes

**EMuDig4.0**

DEUTSCHE  
MASSIV  
UMFORMUNG  
NEUE  
IDEEEN  
SCHMIEDEN



## **Dr. Christoph Lehmann**

Wissenschaftlicher Mitarbeiter  
Center for Scalable Data Analytics and Artificial Intelligence  
ZIH - Zentrum für Informationsdienste und Hochleistungsrechnen  
Technische Universität Dresden

[Christoph.Lehmann@tu-dresden.de](mailto:Christoph.Lehmann@tu-dresden.de)  
Tel. +49 351 463 42489

## Cloud-Infrastruktur zur Datenanalyse von Produktionsdaten aus der Massivumformung

# Ausgangslage

**EMuDig4.0**

DEUTSCHE  
MASSIV  
UMFORMUNG  
NEUE  
IDEEN  
SCHMIEDEN

- Jeder Prozess erzeugt eine große Datenmenge.
- Daten jedes Prozesses werden separat aufgezeichnet.
- Viele verschiedene Datenbanksysteme, Dateiformate im Einsatz
- Keine Verknüpfung zwischen Prozessen, Werkzeugen und Produkten vorhanden
- Rückverfolgung von Artikeln, wenn überhaupt, nur manuell möglich

# Ziele des ZIH

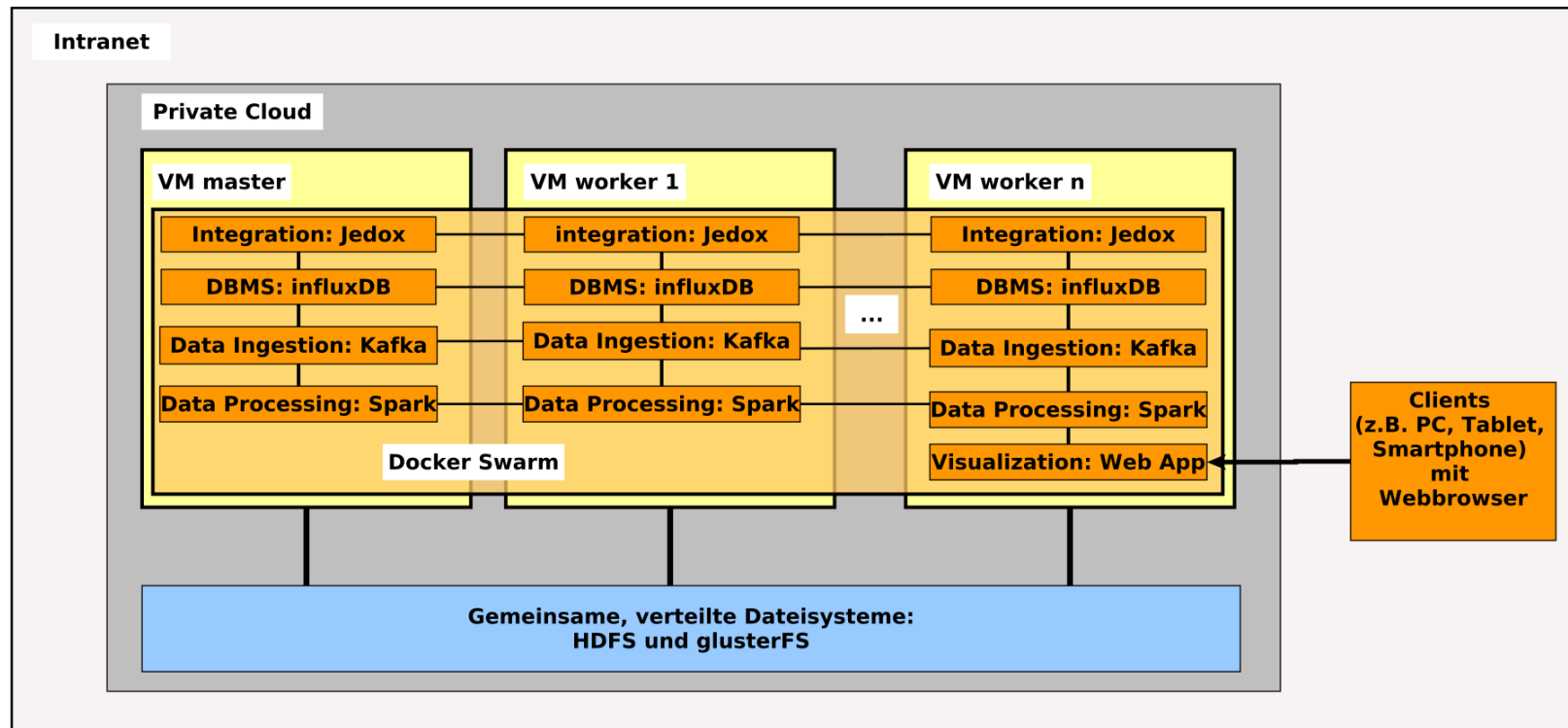
*EMuDig4.0*

DEUTSCHE  
MASSIV  
UMFORMUNG  
NEUE  
IDEEN  
SCHMIEDEN

- 1. Datenanalyse-Infrastruktur** Entwurf einer skalierbaren Datenanalyse-Infrastruktur und exemplarische Umsetzung in Test-Umgebung
- 2. Datenanalysen / Vorhersagen** Auswahl, Anpassung und Anwendung von Analyse- und Vorhersagemethoden basierend auf Produktionsdaten und Fragestellungen der Projektpartner OTTO-FUCHS und Hirschvogel

# Ergebnisse des Forschungsprojektes

## Skalierbare Datenanalyse-Infrastruktur



1. Zusammenführung der Sensordaten, Verwaltung, Prüfung und Speicherung ankommender Datenströme
2. Vorbereitung und Speicherung von (Langzeit-) Analysen (Gemeinsames Dateisystem)
3. Interaktive Offline-Datenanalyse: Benutzer können über Webbrowser auf Analysen zugreifen und selbst Analysen durchführen.

Architekturkonzept für skalierbare Datenanalyse-Infrastruktur entworfen und exemplarisch in Test-Cloud umgesetzt  
-> als Vorlage mindestens in der Branche (aber auch darüber hinaus) einsetzbar

## Ergebnisse des Forschungsprojektes

Interaktive Web-Applikationen zur Arbeit mit vorhandenen Prozessdaten

### Datenvorverarbeitung

**OF Preprocessing App (dev.)**

**Select data type:** forging\_windows

**Select data version:** 20190724

**Select variable names:** [empty]

**Statistics:**

- Number of available observations: 264223
- Number of selected observations: 264223
- Number of available variables: 20
- Number of selected variables: 20
- Sample size: 100 %
- Sampling with replacement?: NO
- 1. Data folder is selected?: YES
- 2. Data structure is read in?: YES
- 3. Data are preprocessed?: YES
- 4. Data are written to file?: NO

**Table With Data Checks**

Check of the data quality for all selected variables within the selected time interval. A data quality rate  $r$  will be computed. The rate is defined as follows:  $r := 1 - nErrorsTotal / (nRules * nObservations)$ . A rate of 1 = 100 % means that all data have been successfully checked without any errors. Per default, the table only show those rules of the selected variables which have been failed. This behavior can be changed by activating the option 'Show results of all rules!'. The rules are contained inside a text file and will be evaluated for all selected data. The result table consists of the following columns: \* Variable: Name of variable, \* Rule: Rule which has been evaluated, \* Execution: Indicator if rule could be successfully evaluated ('ok' or 'failed'), \* Error\_sum: Total number of rule violations, \* Error\_rel: Relative number of rule violations wrt. total number of observations.

100 %

### Datenanalyse

**Data Analysis App**

**Select plots / tables:**

- Table Current Dataset
- Table Bad Items
- Table Data Checks
- Table Absolute Frequencies
- Table Basic Stat. Measures
- Plot Abs. Frequencies
- Scatter Plots Over Time
- Time Series Plots
- 2D Scatter Plots
- Matrix Correlation Coeff.
- Plots Auto Correlations
- Plots Cross Correlations
- Parallel Coordinate Plots
- PCA
- K-Means
- KNN
- MINE
- Plots grouped by quantiles
- Hierarchical clustering algorithms
- Select all

**Statistics:**

- Number of observations: [empty]
- Number of variables: 2
- Time interval: days
- Sampling with replacement?: NO
- Merged with quality data?: NO

**Select process:** [empty]

**Select time interval:** [empty]

**Select charge ids:** Nothing selected

**Select customer ids / items:** Nothing selected

**Analysis Options:**

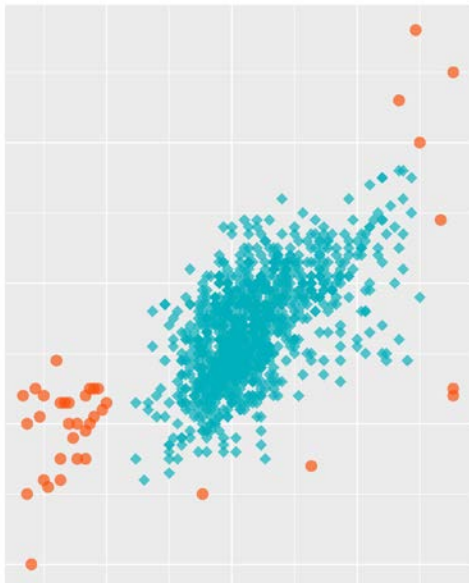
- Absolute frequencies
- Scatter plots over time
- 2D scatter plots

**Buttons:** Select good items, Select bad items

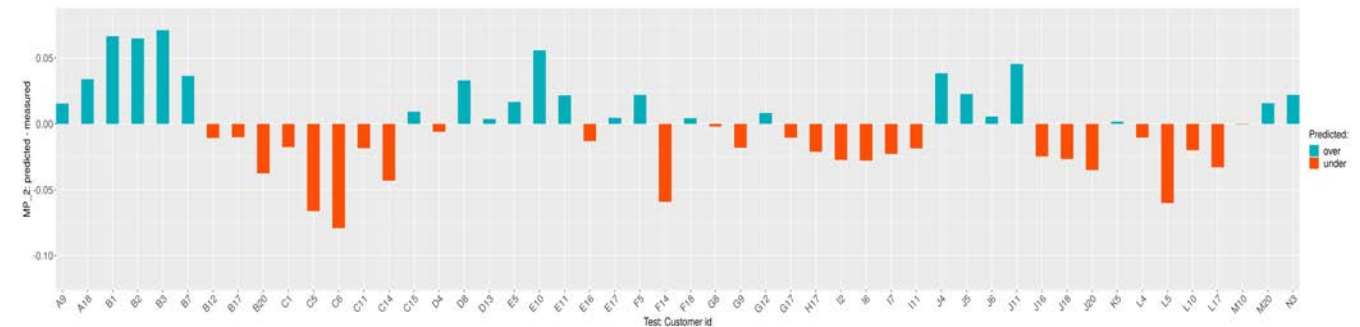
# Ergebnisse des Forschungsprojektes

Ausgewählte Fragestellungen der Datenanalyse

a) Einteilung von Bauteilen in Ausreißerklassen (Gut- bzw. Schlechtclassen) auf Basis mehrerer Gütemerkmale. (Bagplots, Clusteranalyse)



b) Vorhersage von Bauteil-Gütemaßen durch Prozessgrößen. (Regression, künstliche neuronale Netze)



c) Prognose von Störungen oder Werkzeugwechseln in den nächsten x Minuten. (Entscheidungsbaum)

Entwicklung eines generischen R-Paketes mit Open-Source-Lizenz zur Anwendung auf vergleichbare Datenstrukturen für Datenvorverarbeitung und Datenanalyse in Web-Applikationen -> verwendbar innerhalb der Branche, aber auch darüber hinaus

# Wesentliche Erkenntnisse

1. erfolgreiche prototypische Umsetzung der Cloud-Infrastruktur war einfacher als erwartet
2. prozessübergreifende Datensicht vs. Einzelbetrachtung Prozess-Schritt
  - sinnvollerweise Start mit Einzelbetrachtung, später auf gesamte Prozesskette ausrollen
3. Erstellung auswertbarer Datensätze u.U. sehr aufwändig
  - frühzeitig Voraussetzungen im Unternehmen schaffen: technische Voraussetzungen + zeitliche Ressourcen + Kommunikation + individuelle Bereitschaft
4. wesentliche Rahmenaspekte: frühzeitige Berücksichtigung Datenschutz und Datensicherheit
  - Flexibilität vs. Sicherheit der technischen Voraussetzungen bei Aggregation und Analyse (Datentransfer, Internetanbindung, Verschlüsselung, ...) sowie Kooperationen
5. Bearbeitung inhaltlicher Fragestellungen als Kooperation: interne Experten + Datenanalysten
  - geeigneter Datengrundlage: nicht wichtig, möglichst viele Daten zu erheben, sondern die passenden



# Ausblick und Anschlussprojekte

**EMuDig4.0**

DEUTSCHE  
MASSIV  
UMFORMUNG

NEUE  
IDEEEN  
SCHMIEDEN

- Cloud-Infrastruktur auf Unternehmensebene integrieren
- weitere Fragestellungen der Datenanalyse bearbeiten
- Identifikation geeigneter Datengrundlage zur Analyse weiterer Fragestellungen

# Förderhinweis/Danksagung

Das Verbundvorhaben „EMuDig 4.0 – Effizienzschub in der Massivumformung durch Entwicklung und Integration digitaler Technologien im Engineering der gesamten Wertschöpfungskette“ wurde im Rahmen des Technologiewettbewerbs Digitale Technologien für die Wirtschaft (PAiCE) vom Bundesministerium für Wirtschaft und Energie (BMWi) aufgrund eines Beschlusses des Deutschen Bundestages gefördert.

Der Konsortialschlussbericht kann über das Internet unter <https://www.massivumformung.de/forschung/emudig-40/projekt/> angefordert werden.

Gefördert durch:



aufgrund eines Beschlusses  
des Deutschen Bundestages

# EMuDig4.0

DEUTSCHE  
MASSIV  
UMFORMUNG  
NEUE  
IDEEN  
SCHMIEDEN

PAiCE

