Big Data@SIEMENS a KNIME Story
- Backstage -

Tobias Kötter
KNIME
KNIME Hadoop Integration (some important components)

- HBase
- Phoenix
- Drill
- YARN
- Spark

KNIME Big Data Connectors
YARN
KNIME Batch Executors on Hadoop

- **REST Client** submits workflow parameters via REST.
- **KNIME Workflow Designer** designs and uploads workflow.
- **Oozie Server** coordinates workflow execution.
- **YARN Container** executes workflow.
- **Hadoop Master Node** supervises the execution.
- **Hadoop Worker Node** performs the actual computation.
- **HDFS** provides distributed file storage.
- **YARN** manages resource allocation for tasks.
KNIME Executors on Hadoop

KNIME Workflow Users

REST Client

KNIME Workflow Designer

Execute workflow

Execute workflow

Uploads

Submits

Starts

KNIME Server

Hadoop Worker Node

YARN Container

Executes

YARN

KNIME Executes workflow

YARN Client

KNIME Workflow Users

KNIME Server
Spark
Spark Streaming
KNIME on Spark
Acknowledgement

• Jan Pospisil  Siemens
• Arnulf Betz  Siemens
• Aila Kleemann  Siemens
• Björn Lohrmann  KNIME
• Budi Yanto  KNIME
The KNIME® trademark and logo and OPEN FOR INNOVATION® trademark are used by KNIME.com AG under license from KNIME GmbH, and are registered in the United States.