

KNIME Big Data Extensions

KNIME[®] Big Data Extensions integrate the power of Apache Hadoop and Spark with the ease-of-use of KNIME Analytics Platform & KNIME Server. Our software takes the confusion out of big data by making it accessible within our familiar analytics environment.

KNIME Big Data Extensions consist of two complementary node libraries:

- KNIME Big Data Connectors enable you to import/export HDFS data and perform SQL analytics within Hive and Impala.
- KNIME Spark Executor enables you to create and run Spark applications from within KNIME Analytics Platform or KNIME Server, unleashing the power of scalable analytics. Read/write data in HDFS, Hive, and Impala from within Spark.

Unleash the Power of Hadoop

Migrating your analytics to big data has now been reduced to swapping a few nodes in existing workflows.

KNIME Big Data Extensions bring you into the Hadoop ecosystem with support for enterprise-grade, industry-leading Hadoop distributions.

Query Hive data and apply advanced analytics in Apache Spark within a single, visual KNIME workflow, making Hadoop accessible without coding.

A Powerful Combination

KNIME Big Data Extensions bring a familiar, easy-to-use graphical approach to big data problems.

These libraries blend the power of KNIME Analytics Platform with Hadoop to expand the advantages of both:

- SQL-style big data querying
- Sophisticated data mining
- Advanced predictive analytics
- In-memory processing
- Extensive additional functionality

Advanced Features

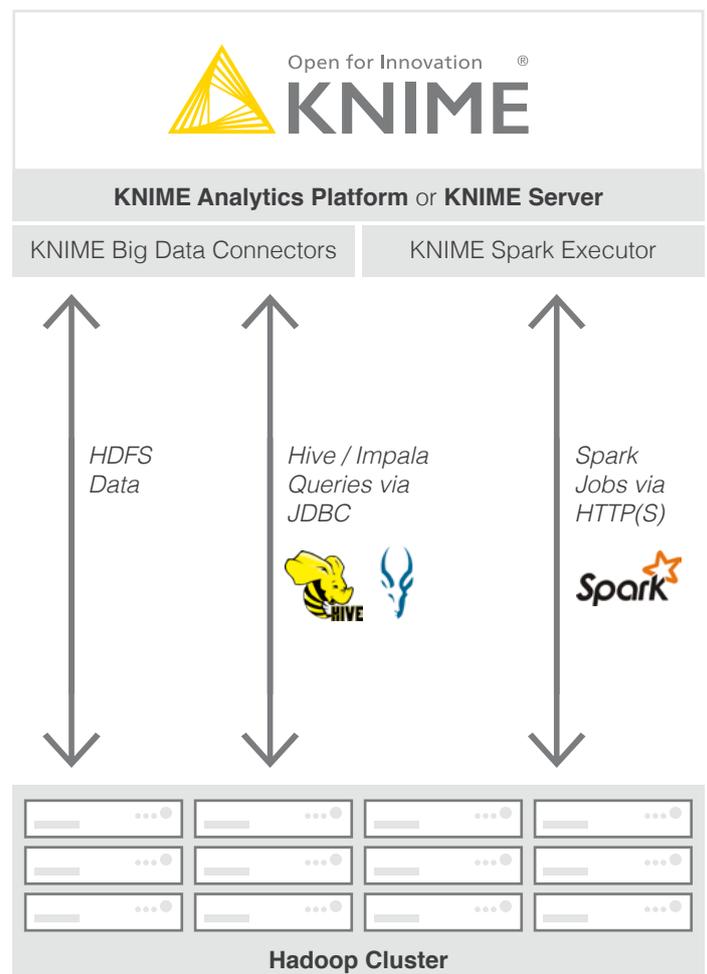
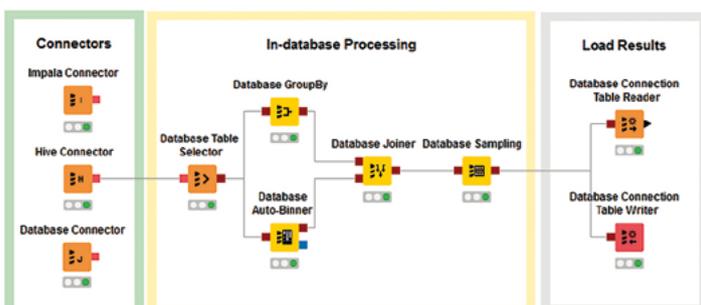
- Effortlessly connect to popular Hadoop distributions
- Seamlessly integrate Apache Spark with >1000 native KNIME nodes using familiar KNIME workflows
- Mix & match remote and distributed computing as needed
- MLib integration enables a popular suite of machine learning algorithms
- Import predictive models into Spark with PMML models generated from KNIME workflows

KNIME Big Data Connectors

KNIME Big Data Connectors allow easy access to Apache Hadoop data from within KNIME Analytics Platform and KNIME Server. This extension offers a set of KNIME nodes for accessing Hadoop/HDFS via Hive or Impala and ships with all required libraries.

Nodes Included: HDFS Connection, HDFS File Permission, Hive Connector, Hive Loader, Impala Connector, Impala Loader

- Move data between KNIME Analytics Platform or KNIME Server and Hive/Impala
- Write Hive/Impala SQL queries using the standard KNIME Database Query node
- Process SQL queries directly in Hive and Impala using standard KNIME database nodes

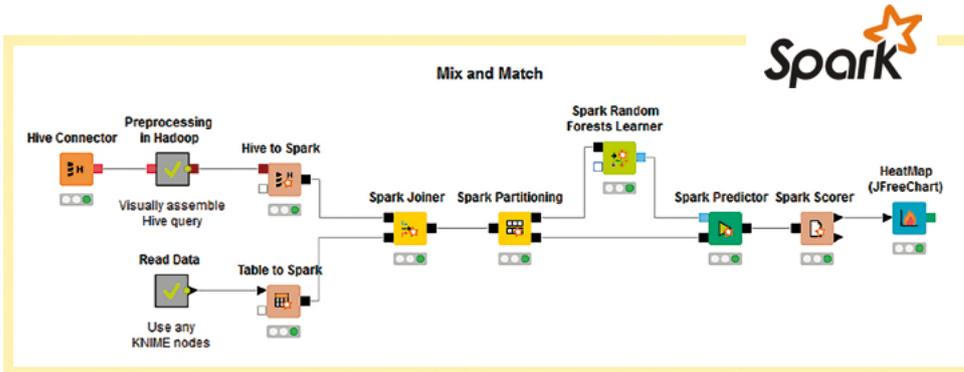


KNIME Spark Executor

KNIME Spark Executor is a set of nodes used to create and execute Apache Spark applications with the familiar KNIME Analytics Platform. Visual programming allows code-free big-data science, while scripting nodes allow detailed control when desired.

This KNIME Spark Executor library of nodes enables you to:

- Effortlessly connect to popular Hadoop distributions
- Seamlessly integrate Apache Spark with >1000 native KNIME nodes using familiar KNIME workflows
- Mix & match remote and distributed computing as needed
- MLlib integration enables a popular suite of machine learning algorithms
- Import predictive models into Spark with PMML models generated from KNIME workflows



Spark Node Categories

This library includes nodes to perform the following functions on Spark:

- I/O
- Manipulation
- Dimensionality Reduction
- Machine Learning (Extensive List)
- Statistics
- Scoring

MLlib Integration

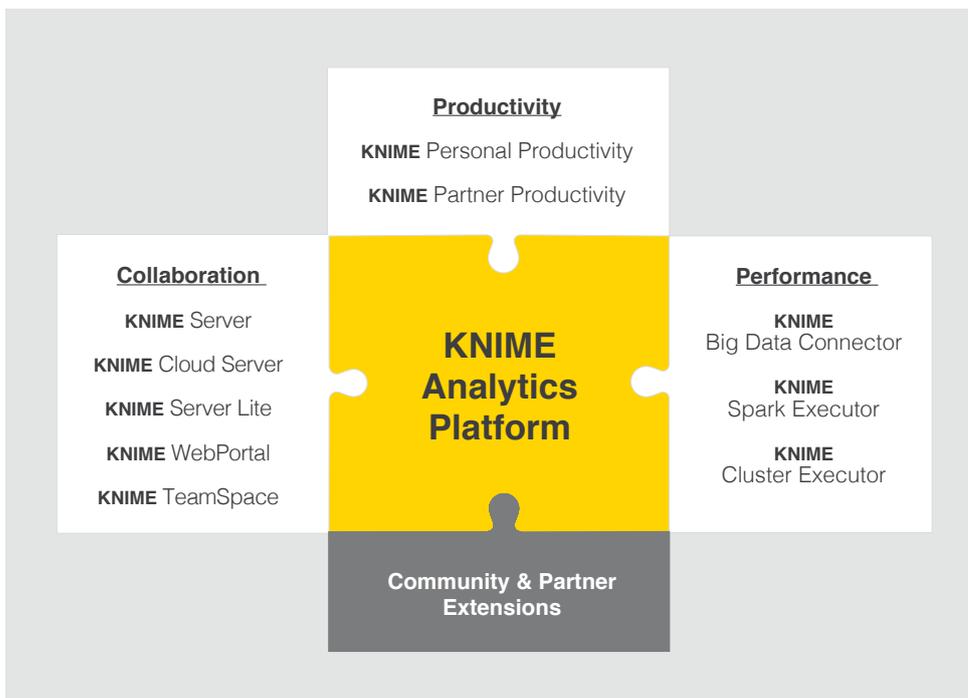
Integrate Spark's scalable machine learning library into your workflows to perform:

- Classification
- Regression
- Clustering
- Collaborative Filtering
- Dimensionality Reduction

Extending KNIME Analytics Platform

The full featured, unrestricted, open source, and free KNIME Analytics Platform is the perfect environment for unleashing the potential of a single data scientist. When you are ready to take your analytics to the next level, KNIME software extends those capabilities with commercial extensions to enhance collaboration, performance, and productivity.

Whatever your data needs, KNIME Commercial Software can take you there.



About KNIME

At KNIME®, we build software for fast, easy, and intuitive access to advanced data science, helping organizations drive innovation.

Our KNIME Analytics Platform is the leading open solution for data-driven innovation, designed for discovering the potential hidden in data, mining for fresh insights, or predicting new futures. Organizations can take their collaboration, productivity and performance to the next level with a robust range of commercial extensions to our open source platform.

For over a decade, a thriving community of data scientists in over 60 countries has been applying our platform to every kind of data: from numbers to images, molecules to humans, signals to complex networks, and simple statistics to big data analytics.

KNIME's headquarters are in Zurich, with additional offices in Konstanz, Berlin, and Silicon Valley. We're open for innovation®, so visit us at

KNIME.com