Cheat sheet: Connectors with KNIME Analytics Platform

Open for Innovation
KNIME

On premise

Access

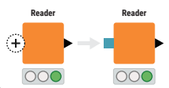
Environments

Create Local Big Data Environment: Creates a fully functional local big data environment including Apache Hive, Apache Spark, and HDFS. Allows opening Spark WebUI and sharing Spark contexts between KNIME workflows. Ideal for test runs.

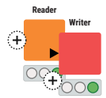
Create Databricks Environment: Creates a Databricks Environment connected to an existing Databricks cluster.

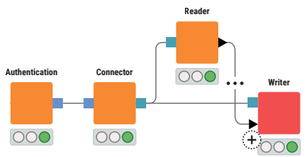
Create Spark Context (Livy): Creates a new Spark context via Apache Livy. Requires access to a remote file system in order to exchange temporary files between KNIME and the Spark context running on the cluster.

Most Reader nodes can read both local and remote data. They connect to remote data sources via dynamic ports. These can be activated by clicking the plus on the left side of the node.



Most Reader nodes have their corresponding Writer nodes. Similar to Reader nodes, most Writer nodes support writing the data directly to remote locations via dynamic ports.





File systems

Files

Table Reader: Reads data from a .table file. .table files are saved using a KNIME proprietary format, include the file structure, and are optimized for space and speed. Other nodes are available to read tabular formatted files, e.g., Parquet or ORC files.

File Reader: Reads all text files, particularly character separated files, such as CSV files. Other similar reader nodes are dedicated to reading special file formats, like Excel or CSV files.

Tika Parser: Parses textual content and metadata and extracts embedded files and attachments from more than 280 file formats. Also provides an authentication option for encrypted files.

Tess4J: Reads textual data straight out of document copies or photos using the Tesseract OCR library.

JSON Reader: Reads either the whole JSON document or the selected part of the document, specified with a JSONPath query. The XML Reader node reads XML documents.

Web Log Reader: Reads Apache log files.

Image Reader: Reads PNG and SVG images, as well as ZIP files containing images, by browsing over the file system. Similar reader nodes read images from URLs, URIs, or Paths in the input table.

List Audio Files: Reads audio files into a data cell. It is often used together with the Audio Viewer node to play audio files.

Network Reader: Reads and creates a network saved to a graph file with the Network Writer node.

Viz Input Connector: Reads networks from visone and Cytoscape.

MDF Reader: Reads the measurement data of one or more channels of an ASAM MDF file, either fully or in part.

SDF Reader: Loads molecules from MDL Structure-Data Files (SDF).

Integrations

Python Source: Executes a Python script in a local Python environment. Supports Python 2 and 3 and Jupyter notebooks import.

Python Object Reader: Reads Python pickle objects. Supports Python 2 and 3 and Jupyter notebooks import.

R Source (Table): Reads diverse data sources from R into a KNIME table.

Index Reader: Reads Lucene table index.

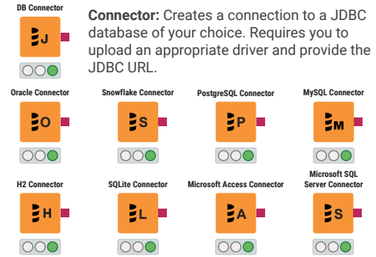
H2O MOJO Reader: Reads H2O's Generated MOJO models.

SAS7BDAT Reader: Reads data from sas7bdat files.

Databases

Dedicated Connector nodes connect to a specific SQL, noSQL, or big data platform, and require a limited number of settings e.g., hostname and credentials.

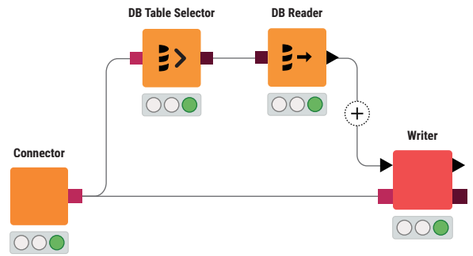
SQL



Connector: Creates a connection to a JDBC database of your choice. Requires you to upload an appropriate driver and provide the JDBC URL.

NoSQL

Big Data



Web services

Services

KNIME REST Client Extension: Calls a REST service in the GET, POST, PUT, DELETE, or PATCH mode. Can send one single service request set in the configuration window, or multiple service requests stored in a column of the input table. Options to set authentication, request header, & response header are available.

KNIME Twitter Connectors: Connect to Twitter's API, retrieve tweets, users, or post new tweets. Require credentials for the Twitter's Developer account.

SPARQL Endpoint: Connects to a SPARQL endpoint. Can be then used with Semantic Web nodes.

Triple File Reader: Reads triples stored in a file (.ttl, .rdf. .rj, .nt, .trig, .trix).

Memory Endpoint: Provides an in-memory Semantic Web endpoint. Supports default and named graphs and works with all offered Semantic Web nodes.

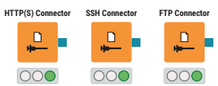
KNIME Salesforce Integration: Interact with Salesforce's REST API performing authentication and SOQL queries execution.

Web

Webpage Retriever: Retrieves web pages by issuing HTTP GET requests and parsing the requested HTML webpage from one or more URLs. The output can be returned in an XHTML or String format.

RSS Feed Reader: Connects to an RSS Feed URL, parses the RSS feeds, and retrieves the metadata. The results can be saved as String, Document, XML, or HTTP response code columns.

These nodes connect to web servers and specify a working directory with a UNIX-like syntax. The downstream nodes can then access the files on the server (FTP, SSH protocols) or read single files from a server (HTTP(S) protocol). The connection is closed when the Connector node is reset or the workflow is closed.



Models

Model Reader: Reads KNIME formatted models generated with any of the Learner nodes. The PMML Reader node reads PMML formatted models.

Keras Network Reader: Reads a Keras deep learning network. A pre-trained network can be read from an HDF5 (.h5) file. A network specification without weights can be read from JSON or YAML files.

TensorFlow Network Reader: Reads a TensorFlow deep learning network of the SavedModel format from a directory or zip file.

TensorFlow 2 Network Reader: Reads a TensorFlow 2 deep learning network from a file or folder. The model should be saved as an HDF5 (.h5) file, a SavedModel file, or zip file of a SavedModel.

Word Vector Model Reader (legacy): Reads word vector models saved by the Word Vector Writer Node, models in .txt, .csv, or .bin.gz formats.

BERT Model Selector: Downloads BERT models from TensorFlow Hub and HuggingFace to the disk. The cashed model can be then used with the BERT Classification Learner node.

OpenNLP NER Model Reader: Reads OpenNLP named entity tagging models.

Resources

- **KNIME Press:** Access various data science books and other cheat sheets at ***knime.com/knimepress***, including beginner and advanced topics.
- **KNIME blog:** Engaging topics, challenges, industry news, & knowledge nuggets at ***knime.com/blog***.
- **Self-paced courses:** Take our free online self-paced courses to learn about data analysis, data engineering, or data science with KNIME (with hands-on exercises) at ***knime.com/learning***.
- **KNIME Community Hub:** Browse and share workflows, nodes, and components or access collection pages for dedicated topics at ***hub.knime.com***.
- **KNIME Forum:** Join our global community & engage in conversations at ***forum.knime.com***.
- **KNIME Business Hub:** For team-based collaboration, automation, management, & deployment check out KNIME Business Hub at ***knime.com/knime-business-hub***.

On the cloud

Authentication

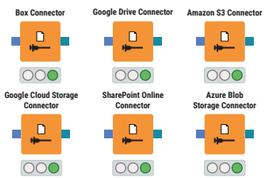
Google Authenticator: Authenticates to Google services via the specified in the node settings. In the configuration window, choose the scopes and authenticate using the "Login" button.

Microsoft Authenticator: Authenticates against Microsoft Azure and Office 365 cloud services via a number of interactive authentication options.

Amazon Authenticator: Authenticates against Amazon services.

Cloud storage systems

Dedicated Connector nodes connect to remote file systems, specify the working directory with a UNIX-like syntax, and allow downstream nodes to access the remote file system just as a local one, e.g., to read or write files and folders, browse, list files, copy, move, etc. The connection is closed when the Connector node is reset or the workflow is closed.

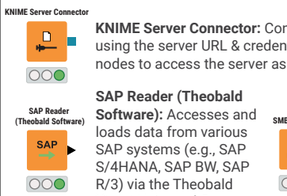


Distributed file systems

Dedicated Connector nodes connect to a specific distributed file system (HDFS, WebHDFS, HTTPFS, Databricks, ...), and require a limited number of settings e.g., hostname and credentials.

Cloud

Servers



KNIME Server Connector: Connects to a KNIME Server using the server URL & credentials. Allows downstream nodes to access the server as a file system.

SAP Reader (Theobald Software): Accesses and loads data from various SAP systems (e.g., SAP S/4HANA, SAP BW, SAP R/3) via the Theobald Xtract Universal Server.

SMB Connector: Connects to a remote SMB server. Allows downstream nodes to access the server as a file system.

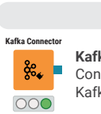
Cloud services

Google Sheets Connector: Connects to Google Sheets. Depending on the authentication method, the sheet should be either opened with a Google account or shared with a service account.

Google Analytics Connector: Connects to Google Analytics API.

KNIME Amazon Machine Learning Integration: Interacts with AWS AI/ML-Services like AWS Comprehend, AWS Translate, and AWS Personalize. The authentication via Amazon Authentication is required.

Messaging systems



Kafka Connector: Connects to a Kafka cluster.

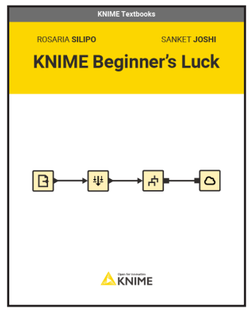
Kafka Consumer: Consumes Kafka cluster messages on real time data feeds from sensor devices and stores them in a table.

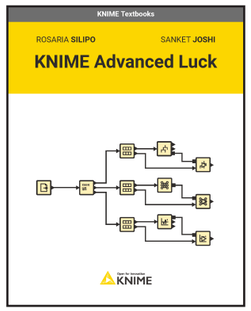
Note: Missing your favorite source? This list is just an extract of the whole set of the connector nodes currently available within KNIME Analytics Platform. Besides, new connector nodes are being created as we speak.

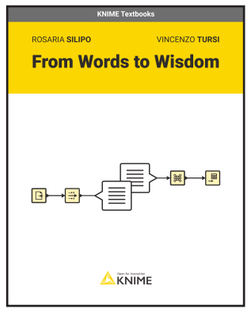
KNIME Press

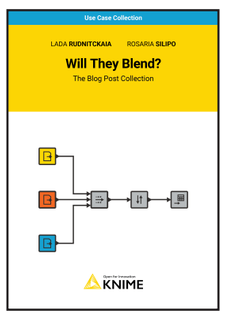
Open for Innovation
KNIME

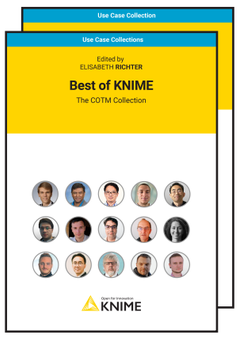
Extend your KNIME knowledge with our collection of books from KNIME Press. For beginner and advanced users, through to those interested in specialty topics such as topic detection, data blending, and classic solutions to common use cases using KNIME Analytics Platform - there's something for everyone. Available for download at www.knime.com/knimepress.

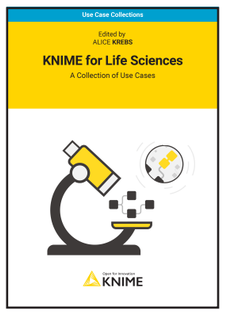


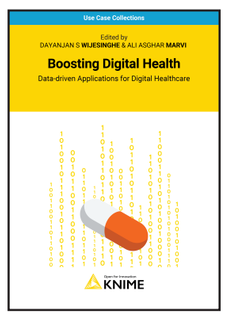


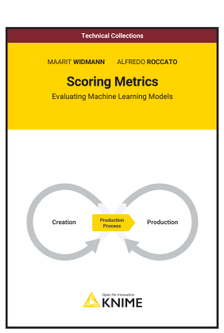


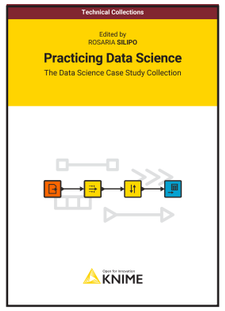


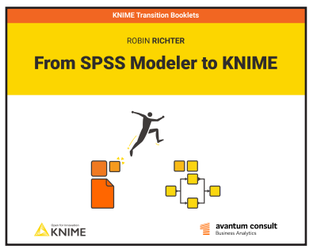


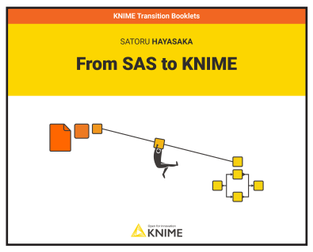


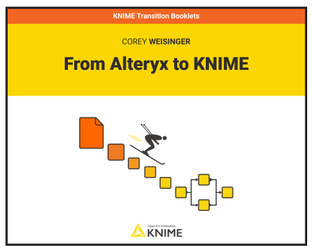












Need help?
Contact us!

© 2024 KNIME AG. All rights reserved. The KNIME® trademark and logo and OPEN FOR INNOVATION® trademark are used by KNIME AG under license from KNIME GmbH, and are registered in the United States. KNIME® is also registered in Germany.