Cheat Sheet: Connectors with KNIME Analytics Platform

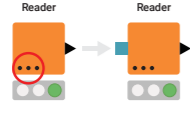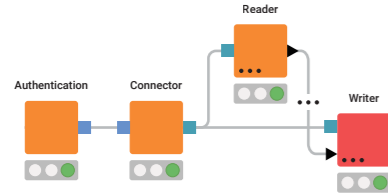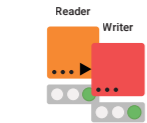**Open for Innovation** KNIME

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 three dots in the node lower left corner.

Reader — **Reader**

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.

Reader — **Writer**

Authentication — **Connector** — **Reader** — **Writer**

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).

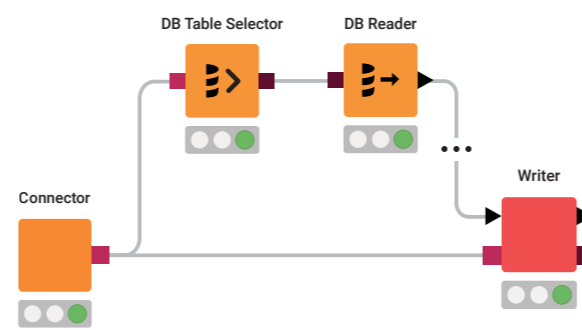#### Integrations

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.

H2O Mojo Reader — Reads H2O's Generated MOJO models.

Index Reader — Reads Lucene table index.

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

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

Oracle Connector | **Snowflake Connector** | **PostgreSQL Connector** | **MySQL Connector**

H2 Connector | **SQLite Connector** | **Microsoft Access Connector** | **Microsoft SQL Server Connector**

NoSQL

MongoDB Connector | **OrientDB Connection** | **Neo4j Connection**

Big Data

Vertica Connector | **Impala Connector** | **Hive Connector**

DB Table Selector — **DB Reader**

Connector — **Writer**

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 Connector Extension — 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.

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

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

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.

HTTP(S) Connector | **SSH Connector** | **FTP Connector**

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 — 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 cached model can be then used with the BERT Classification Learner node.

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

ON THE CLOUD

Authentication

Google Authentication — Authenticates against Google API services, via the "Authenticate" button's pop-up window. The Google Authentication (API Key) node performs the same authentication via a P12 key.

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

Amazon Authentication — 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.

Azure Data Lake Storage Gen2 Connector | **Google Drive Connector** | **Amazon S3 Connector**

Google Cloud Storage Connector | **SharePoint Online Connector** | **Azure Blob Storage Connector**

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.

HDFS Connector | **HDFS Connector (KNOX)** | **Databricks File System Connector**

CLOUD

Google BigQuery Connector | **KNIME Amazon DynamoDB Nodes** | **Amazon Redshift Connector** | **Amazon Athena Connector**

Cloud Services

Google Sheets Connection — 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 Connection — 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.

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.

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.

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.

KNIME® BEGINNER'S LUCK

A Guide to KNIME Analytics Platform for Beginners
Authors: Satoru Hayasaka and Rosaria Silipo

KNIME® ADVANCED LUCK

A Guide to KNIME Analytics Platform for Advanced Users
Authors: Rosaria Silipo and Jeanette Prinz

FROM WORDS TO WISDOM

An Introduction to Text Mining with KNIME®
Vincenzo Tursi and Rosaria Silipo

Open for Innovation
KNIME
Editor: Rosaria Silipo

A Collection of Case Studies

Practicing Data Science

3rd Edition

WILL THEY BLEND

SECOND EDITION

Data Blending with KNIME

KNIME®
Rosaria Silipo & Lada Rudnitckaia

Maarit Widmann, Alfredo Roccato

From Modeling to Model Evaluation

Wheeling like a hamster in the data science creation cycle?

Blend & Transform | Validate & Deploy
Model & Visualize | Creation | Production Process | Production | Consume & Interact
Optimize & Capture | Monitor & Update

Open for Innovation
KNIME

KNIME for Life Sciences

A Collection of Use Cases

Guided Labeling

Human-in-the-Loop Label Generation with Active Learning & Weak Supervision

Written by
Paolo Tamagnini

from **Excel®** to **KNIME®**

Author: Kathrin Melcher

from **sas®** to **KNIME®**

Authors: Rosaria Silipo
Satoru Hayasaka

from **Alteryx** to **KNIME®**

Author: Corey Weisinger