KNIME Analytics Platform Installation Guide

KNIME AG, Zurich, Switzerland

Version 4.6 (last updated on 2023-09-14)

Table of Contents

Installing KNIME Analytics Platform

i For step-by-step **videos** on how to install KNIME Analytics Platform, please take a look at our KNIMETV YouTube channel.

1. Go to the download page on the KNIME.com website to start installing KNIME Analytics Platform.

2. The download page shows three tabs which can be opened individually:

   - *Register for Help and Updates*: here you can optionally provide some personal information and sign up to our mailing list to receive the latest KNIME news

   - *Download KNIME*: this is where you can download the software

   - *Getting Started*: this tab gives you information and links about what you can do after you have installed KNIME Analytics Platform

3. Now open the *Download KNIME* tab and click the installation option that fits your operating system. KNIME Analytics Platform can be installed on Windows, Linux, or macOS.

   Notes on the different options for Windows:

   - The Windows installer extracts the compressed installation folder, adds an icon to your desktop, and suggests suitable memory settings.

   - The self-extracting archive simply creates a folder containing the KNIME installation files. You don't need any software to manage archiving.

   - The zip archive can be downloaded, saved, and extracted in your preferred location on a system to which you have full access rights.

Figure 1. KNIME Analytics Platform versions

4. Read and accept the privacy policy and terms and conditions. Then click *Download*.

5. Once downloaded, proceed with installing KNIME Analytics Platform:

   ◦ *Windows:* Run the downloaded installer or self-extracting archive. If you have chosen to download the zip archive instead, unpack it to a location of your choice. Run `knime.exe` to start KNIME Analytics Platform.

   ◦ *Linux:* Extract the downloaded tarball to a location of your choice. Run the `knime` executable to start KNIME Analytics Platform.

   ◦ *Mac:* Double click the downloaded dmg file and wait for the verification to finish. Then move the KNIME icon to *Applications*. Double click the KNIME icon in the list of applications to launch KNIME Analytics Platform.

   ℹ️ Also check the KNIME Getting Started Guide and the KNIME Workbench Guide.

Configuration settings and knime.ini file

When installing KNIME Analytics Platform, configuration settings are set to their defaults, and they can later be changed in the *knime.ini* file. The configuration settings, i.e. options used by the Java Virtual Machine when KNIME Analytics Platform is launched, range from memory settings to system properties required by some extensions.

You can find `knime.ini` in the installation folder of KNIME Analytics Platform.

ℹ️ **On MacOS**: To locate `knime.ini` on MacOS, open Finder and navigate to your installed Applications. Next, right click the KNIME application, select *Show Package Contents* in the menu, and navigate to `Contents` → `Eclipse`.

The `knime.ini` file can be edited with any plaintext editor, such as Notepad (Windows), TextEdit (MacOS) or gedit (Linux).

Allocating memory in knime.ini file

The entry `-Xmx1024m` in the `knime.ini` file specifies how much memory KNIME Analytics Platform is allowed to use. The setting for this value will depend on how much memory is available in your machine. KNIME recommends setting it to approximately one half of your available memory, but you can modify the value based on your needs. For example, if your computer has 16 GB of memory, you might set the entry to `-Xmx8192m`.

Installing Extensions and Integrations

If you want to add capabilities to KNIME Analytics Platform, you can install extensions and integrations. The available extensions range from free open source extensions and integrations provided by KNIME to free extensions contributed by the community and commercial extensions including novel technology nodes provided by our partners.

The KNIME extensions and integrations developed and maintained by KNIME contain deep learning algorithms provided by Keras, high performance machine learning provided by H2O, big data processing provided by Apache Spark, and scripting provided by Python and R, just to mention a few.

Install extensions from:

- KNIME Hub:

  ◦ Search for the Extension or Integration you want to install in the search bar

  ◦ Click Extensions on the results page

  ◦ Click the extension you want to install, and from the extension page and drag and drop the squared yellow icon, shown in Figure 2, to the workbench of KNIME Analytics Platform. A window will open asking if you want to search and install the extension or integration. Click *Yes* and follow the instructions.
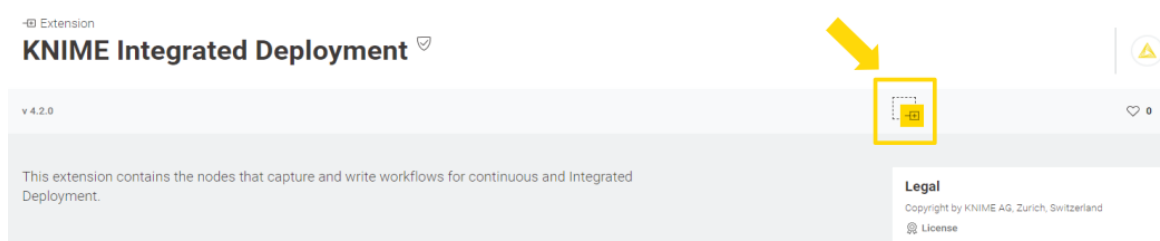

Figure 2. Install the KNIME Integrated Deployment Extension from KNIME Hub

  ◦ Restart KNIME Analytics Platform.

- KNIME Analytics Platform:

  ◦ Click *File* on the menu bar and then *Install KNIME Extensions*…. The dialog shown in Figure 3 opens.
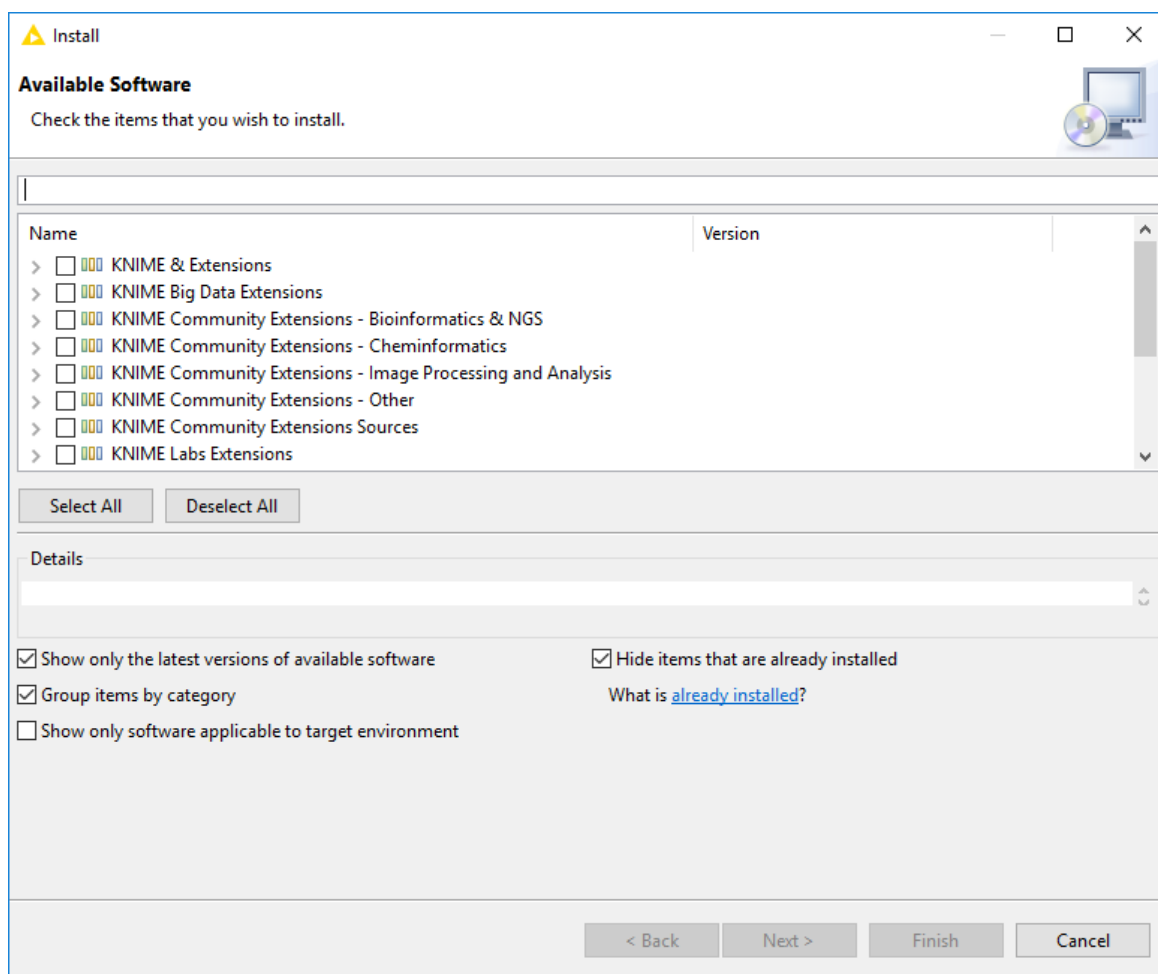
Figure 3. Installing Extensions and Integrations from KNIME Analytics Platform

- Select the extensions you want to install

- Click *Next* and follow the instructions

- Restart KNIME Analytics Platform.

The *Install KNIME Extensions* menu provides the extensions that are available via the update sites you have enabled.

> For more information, take a look at our video on How to Install Extensions in KNIME Analytics Platform. Also see the Extensions and Integrations Guide.

To uninstall an extension, click *Help*, *About KNIME Analytics Platform*, and then *Installation Details*. A dialog shown in Figure 4 opens. Now, select the extension that you want to uninstall, and click *Uninstall*….
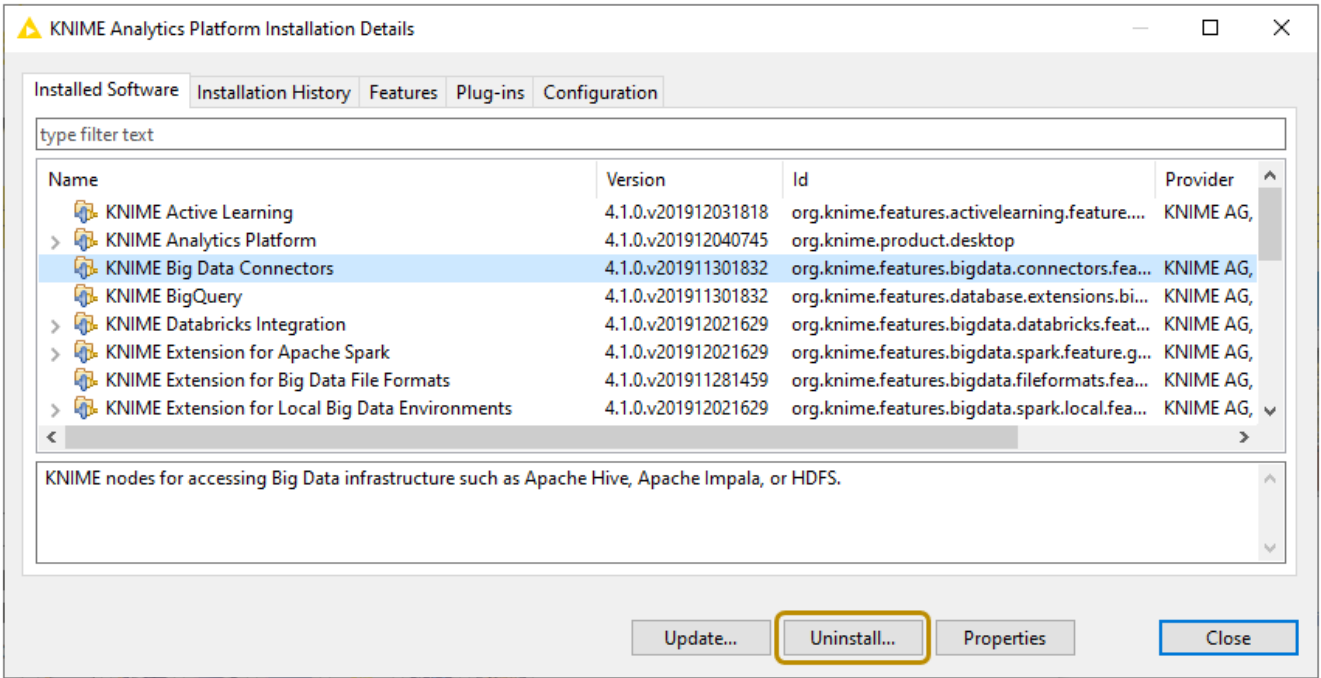
Figure 4. Uninstalling Extensions and Integrations

Updating KNIME Analytics Platform and Extensions

It is good to make sure that you always use the latest version of KNIME Analytics Platform and its extensions.

Do this by:

1. Clicking *File → Update KNIME…*. In the dialog that opens, select the available updates you want to install and then click *Next*.

2. Proceed by following the instructions. KNIME Analytics Platform has to be restarted in order to apply the updates.

Update Sites

The Update Sites are where KNIME retrieves additional software in the form of extensions as well as updates. To see or edit the available update sites, select *File → Preferences → Install/Update → Available Software Sites*.

Default Update Sites

These four updates sites are provided by KNIME and are always available:
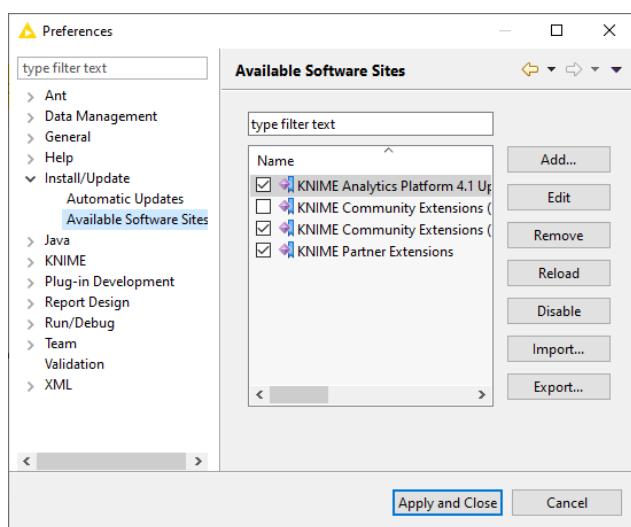
Figure 5. Available Update Sites

KNIME Analytics Platform 4.6 Update Site: Provides all extensions and integrations maintained by KNIME: R, Python, H2O Machine Learning, Apache Spark for big data, and many more. Contains KNIME Labs Extensions, which are extensions that are not yet part of the set of stable KNIME extensions because their functionality may not yet be finalized.

KNIME Community Extensions (Trusted): Provides trusted community extensions, i.e. extensions created by the KNIME community, which have been tested for backward compatibility and compliance with KNIME quality standards.

KNIME Partner Extensions: Provides extensions created by KNIME partners.

Community Extensions (Experimental): Provides additional extensions created by the KNIME community.

KNIME Analytics Platform 4.6 Update Site and *KNIME Community Extensions (Trusted)* are enabled by default.

Adding External Update Sites

To install extensions that are not part of the above update sites, click *Add* to manually add the relevant update site, inserting the Name and Location as shown in Figure 6.
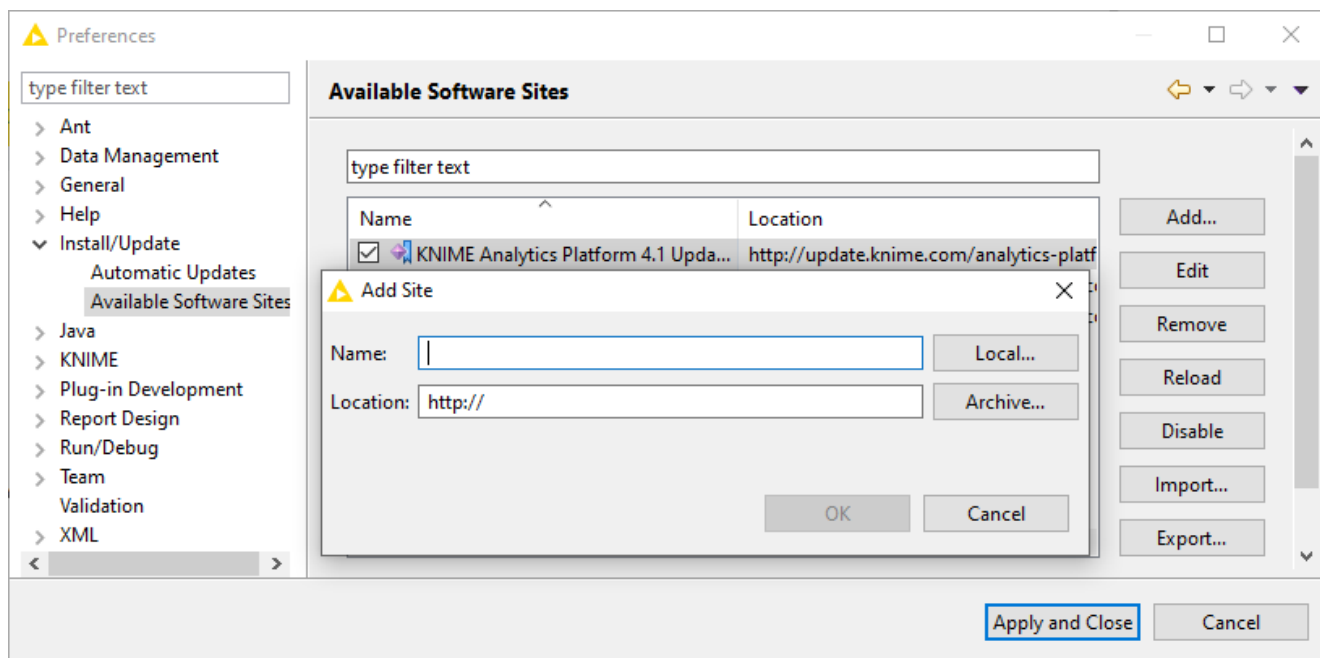
Figure 6. Add Update Sites

After adding a new update site you will see it listed in the *Available Software Sites*. You must now enable it by selecting it from the list.

Adding Local Update Sites

If your working environment has limited internet access or you receive an error message "Proxy Authentication Required" when connecting to a remote update site (provided by a URL), you can install extensions from a local zip file.

1. Download KNIME update sites as zip files at the following links:

    ◦ KNIME Analytics Platform Update Site

    ◦ KNIME Community Extensions

    ◦ KNIME Partner Extensions

2. Save the zip file containing the extensions to your local system

3. Select *File → Preferences → Install/Update → Available Software Sites* and enter the path to the zip file by clicking *Add → Archive…* as shown in Figure 7.
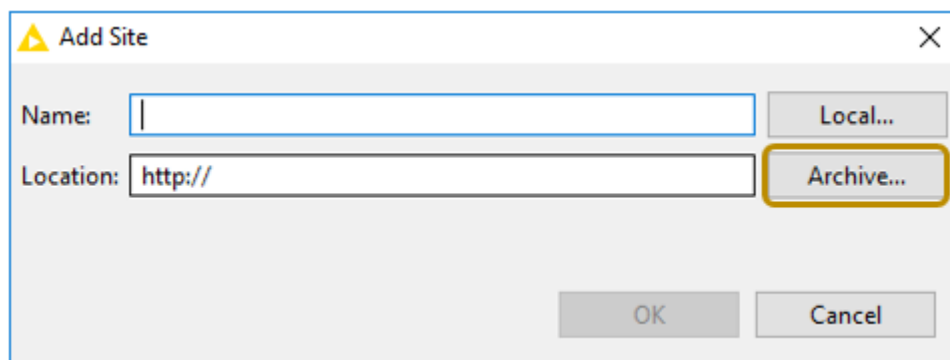
Figure 7. Adding Update Sites from Zip Archive

> **i** If the same extensions are provided by a URL, you will first have to disable the update site by disabling it in the list.

4. Now click *Apply and Close*

> **i** If the same extensions are also provided by a remote update site, you will first have to disable that update site by deselecting its entry in the *Available Software Sites* dialog and confirming via *Apply and Close*.

Working with the Nightly Builds

Once a night, a new version of KNIME Analytics Platform is created directly from our development branch. The Nightly Build versions available here provide insight into what's coming up in the next regular release. However, for real work, always use a version of a standard KNIME release. Also read the following disclaimer before proceeding:

Really, really, *really* important disclaimer

This is most definitely not production quality code. These nightly builds are what we use internally to validate and test recent developments, so they are not tested as thoroughly as standard KNIME releases. Furthermore new nodes or functionality may change substantially (or disappear entirely) from one build to the next. It's even possible that workflows you edit or create with nightly builds stop being readable by future (or past) versions…

These nightlies are a great way to get a sneak peek at what may be coming in the next version of KNIME and provide feedback and suggestions. They are not a particularly safe way to do real work.

Changelog (KNIME Analytics Platform 4.6)

Detailed changelog for v4.6.x releases

KNIME Analytics Platform 4.6.0

Release date: June 15, 2022

New nodes

- AP-18932: Delete SharePoint Online List
- AP-18869: SharePoint Online List Writer
- AP-18401: DB Looping
- AP-18398: DB Delete (Filter)
- AP-18075: DB Data Spec Extractor
- AP-11366: DB Concatenate
- UIEXT-327: Echarts "Scatter Plot"
- UIEXT-190: Table View
- UIEXT-6: ECharts "Bar chart"
- UIEXT-5: ECharts "Line plot"

Enhancements

- NXT-1144: KNIME Modern UI Preview - preview of the new user interface
- AP-19056: Linux: Add warning popup in case Wayland display server is detected (standard in Ubuntu 22.04. and currently not properly supported)
- AP-19032: Provide Python API to develop KNIME nodes
- AP-19031: JSON To XML node with option to remove invalid characters
- AP-18965: Regex Splitter Node: Doesn't handle multiline strings
- AP-18937: Call Workflow Service to use the Mountpoint Connector
- AP-18892: Update Apache CXF from 3.2.14 to 3.5.2
- AP-18776: Help menu with additional entry redirecting to KNIME Forum
- AP-18731: "Transpose" node to be memory-aware, auto-determining batch size

- AP-18717: Allow to extend Python path programmatically in nodes that use the Python3KernelBackend

- AP-18625: Columnar Backend: Reorder virtual operations to evaluate "slice" before "append"

- AP-18617: Case Switch End Node: Execution time significantly increased when only one input is active (pass on input table)

- AP-18598: Add support for using a custom Azure AD application with the Microsoft Authentication node

- AP-18473: Structure XGBoost Learner settings according to tabs

- AP-18431: Update of log4j-1.2.15 to 1.2.21 (reload4j)

- AP-18396: Support Path type in Variable Expressions node

- AP-18395: H2 driver update to version 2.1.212

- AP-18392: Amazon S3 Connector should use multi-part upload for large files

- AP-18386: Support path flow variables in Column Expressions node

- AP-18362: Add option to specify base score in XGBoost Learners

- AP-18354: Improve Python Script (Labs) node's Python process connection startup time

- AP-18306: Add row key to error messages in XGBoost nodes

- AP-18283: Update H2O to 3.36.0.3

- AP-18276: Store tag sets as part of DataColumnSpec

- AP-18272: Add colsample_bynode option in XGBoost Tree Ensemble Learner

- AP-18228: Allow to set global thread pool size via environment variable

- AP-18202: Microsoft SQL Server node to support transaction disabling to be compatible with Serverless SQL Pool in Azure Synapse Analytics

- AP-18135: Allow setting+appending values of extension types in Pandas Dataframe

- AP-18036: Polling configuration (advanced settings) for Call Workflow Service node

- AP-17944: MySQL driver updated to version 8.0.29

- AP-17943: PostgreSQL driver updated to version 42.3.5

- AP-17815: Amazon Redshift driver updated to version 2.1.0.3

- AP-17703: Provide Oracle Database driver version 19.14.0 as new feature for installation

- AP-17553: Allow keeping RowIDs when joining

- AP-16675: Show hub host name in tab

- AP-16434: SQLite driver updated to version 3.36.0.3

- AP-15924: Support "NOT LIKE" in DB Row Filter

- AP-15639: Allow to select from all available dialects for default db type

- AP-15500: Microsoft Access driver updated to version 5.0.1

- AP-14913: Static bar values in Bar Chart

- AP-14756: Use bundled Python environment in Python Script (Labs) node

- AP-14142: Let Missing Cell directly override isMissing() of DataCell

- AP-12399: Column Expression Node: Ability to access previous and next rows

- AP-11529: XGBoost: Update to version 1.5.1

- AP-11401: Support row weights in XGBoost

- AP-10987: Add linear model parameters as additional output for the XGBoost Linear Ensemble Learner

- AP-10986: Feature Importance Output for XGBoost Tree Ensemble Learner

- AP-10657: XGBoost: Support for vector (bit/byte/double) features

- AP-5490: Generic Web Service Client node (SOAP) to use recent versions of CXF client library (3.5.2)

- BD-1194: (Big Data Extensions): Update H2O Sparkling Water to 3.36.1.2

- BD-1184: (Big Data Extensions): Add support for H2O Sparkling Water to 3.32.1.2-1-3.1

- BD-1182: (Big Data Extensions): Add Spark 3.1 support to Create Databricks Environment node

- BD-1181: (Big Data Extensions): Add Spark 3.1 support to Create Spark Context (Livy) node

- BD-1180: (Big Data Extensions): Add support for Spark 3.1 to KNIME Extension for Apache Spark

- BD-1168: (Big Data Extensions): Update H2O Sparkling Water to 3.36.0.3-1-3.2

- BD-1164: (Big Data Extensions): Add Spark 3.2 support to Create Databricks Environment node

- BD-1161: (Big Data Extensions): Update Create Local Big Data Environment to Spark 3.2.1

- BD-1160: (Big Data Extensions): Add support for Spark 3.2 to KNIME Extension for Apache Spark

- BD-1147: (Big Data Extensions): Add support for new type annotations to Parquet Writer (new file handling)

- NXT-1072: Remote workflow editor: enable configuration nodes in component dialog

Bug Fixes

- AP-18931: R integration does not work with R 4.2.0 any more (on Ubuntu)

- AP-18911: Fixed Width File Reader: Column settings are deleted when controlling the location via flow variable on opening the dialog

- AP-18801: Excel Writer generates error if overwriting a sheet that is used as source in pivot table

- AP-18728: Variable Condition Loop End error is missing information (only log output)

- AP-19046: Table Reader has problems with BlobDataCells when reading multiple tables

- AP-19034: REST Nodes don't work with encoded query parameter names

- AP-19013: "Call Workflow" nodes do not hide any system credentials when invoking remote service (causing error)

- AP-18994: BufferedRowWrite#setMissing is off by one

- AP-18989: The MDF Reader feature doesn't properly install Python

- AP-18969: Read Images Node: Fails on SVG input

- AP-18935: Component is not ignoring error message caught by Try Catch

- AP-18929: Installation directory on Windows must not be writable by everyone

- AP-18906: Copying table cells out of table view node is not working in CEF

- AP-18895: Image Port Type: Unable to clone input data invalid character

- AP-18879: Webpage Retriever fails at cookie and error cause extraction

- AP-18865: Check for correct conda package versions by Python (Labs) in the Pref Page

- AP-18863: Conda Environment Propagation blocks Executor on Configure

- AP-18857: PythonGateway creation gets stuck because std out and error are not consumed

- AP-18854: Inconsistent Behaviour between Match All and Match Any when comparing Row IDs in Joiner

- AP-18851: Parquet/ORC Writer: Mode File/Folder configuration not displayed correctly

- AP-18850: Breakpoint node description is not showing available settings

- AP-18797: File Download Widget: Download Link not working if Component ID is 0

- AP-18792: Cache size heuristics do not make sense on MacOS

- AP-18785: Fix missing value handling in Lists of Pandas ExtensionTypes

- AP-18774: JDBC type REAL looses precision due to double conversion

- AP-18733: Keyboard shortcut for updating linked component does not work on top-level workflow

- AP-18723: Save as workflow doesn't show all spaces on MacOs

- AP-18700: Make sure we do not install jedi=0.18.1 in conda envs because it is broken

- AP-18684: Hard JVM crashes on Windows caused by XGBoost nodes run in parallel

- AP-18683: Writing to the root folder in SharepointOnline causes a NullPointerException

- AP-18682: HTTP(S) Connector always fails to trust bad SSL certificates

- AP-18673: KNIME UI sometimes not coming up in Linux Dev environment

- AP-18659: reg:linear is deprecated in new XGBoost version

- AP-18650: Case Switch End node changes its initial configuration if dialog is opened

- AP-18648: DB Writer fails in streaming mode with enabled DB Data outport

- AP-18630: Integrated Deployment: Variables in "static" input are injected in reverse order

- AP-18601: DB Transaction End/Catch node with inactive input ports does not return inactive spec in component

- AP-18594: KNIME AP freezes when several metanodes have the same input connection

- AP-18320: Workflow Executor does not work when workflow has no inputs/outputs

- AP-18299: XGBoost: Linear Booster Updater 'Shotgun' cannot handle some selectable feature selectors

- AP-18250: Nullpointer after closing KNIME AP due to unfinished wf coach loading

- AP-18248: "Reset" workflow action resets nodes with executing successors

- AP-18134: Documentation of JSON Processing extension talks about XML

- AP-18123: Improve error message when empty tables are sent to XGBoost

- AP-18047: Workflow Summary Extractor extracts and stores passwords

- AP-18020: Python Script node prohibits deletion of jobs folder on executor

- AP-18019: Webpage Retriever: NPE on certain webpages

- AP-18008: Inconsistent capitalization of items in Help menu

- AP-17953: Reader/Writer nodes with Encoding tab should not be initialized to use the OS default encoding option

- AP-17951: DataCells are not cached in the heap cache

- AP-17712: Update smbj library in SMB Connector to potentially fix errors (NullPointerException, PipeShare error) when running parallel operations against DFS

- AP-17551: Mean calculated with Column Aggregator node is inconsistent with other methods

- AP-16818: Node Dialog too wide with long Flow Variable list

- AP-16705: Send Email node does not encode Chinese characters in attachment file name

- AP-15855: Feature buffer empty exceptions in XGBoost Learner and Predictor

- AP-15254: Sunburst chart has error after selection, close and reopen

- AP-15099: XGBoost can lead to memory problems and KNIME crashes

- AP-14815: XGBoost Predictor fails on unknown categorical values

- AP-11514: Itemset Miner (Borgelt) uses too much memory by default

- AP-11061: Rename XGBoost Linear Ensemble Learner

KNIME Analytics Platform 4.6.1

Release date: July 27, 2022

Enhancements

- AP-19164: Allow to append list of columns to Schema

- AP-18920: Allow to specify connection/read timeouts KNIME Mountpoint Connector when connecting to Server/Hub mountpoints

- AP-7465: Add AWS Session Token support to Amazon Authentication node

Bug Fixes

- AP-19193: FTP connector: SSL session reuse fails with Java reflection error

- AP-19129: Copy/Pasting any "Credentials" nodes don't work

- AP-19268: PyArrow version 8.0.0 in bundled env does not support extension types

- AP-19233: Python Node Extension: Dialog validation results in unhelpful console output

- AP-19241: Mac : Mouse events delayed or not registered at all when CEF (Hub View or node view) was launched

- AP-19191: Python Node Extension: Use node name if no short description is provided

- AP-19163: Document Data Extractor adds whitespace after last term

- AP-19183: Python Node Extension: Dialog of unconnected node doesn't open

- AP-19153: Python Node Extension: Can't pass single column to Schema.from_columns

- AP-19152: Python Node Extension: Uninformative error message if binary port object is not bytes

- AP-19149: Python Node Extension: Can't drag'n'drop in APNext

- AP-19147: Python Node Extension: None is not an allowed output of configure

- AP-19139: Python Node Extension: Dialog doesn't open when a new setting is added

- AP-19138: Python Node Extension: Uninformative error on duplicate RowKeys

- AP-19127: KNIME Analytics Platform does not start on systems without Visual Studio C++ Redistributable

- AP-19112: Python Node Extension: Logging in Parameter Settings Validators raises error

- AP-19111: Python Node Dialog does not show error of dialog creation

- AP-19104: Cannot put extension_module into a subdirectory

- AP-19093: Dialog is empty if (Multi)ColumnSelection is pointing to non-table input port

- AP-19091: Empty Docstring raises none exception

- AP-19089: error raising in Python execute method raises other error

- AP-19088: Python Node Extension: Spacing in port description off

- AP-19084: Interactive range slider filter widget throws error when column contains only negative double values

- AP-19074: Python Node Extension: Cannot parse Python node without any inputs or outputs if not extending PythonNode

- AP-18963: Column Parameters are not displayed

- AP-18579: Read Excel Sheet Names node fails for large files

- AP-17130: Bar chart visualization cuts off Y-axis label when X-axis is included

- BD-1200: (Big Data Extensions): Spark Column Filter and Spark Sort nodes throw AnalysisException for columns containing a dot

- UIEXT-324: Table View pagination doesn't work in the WebPortal

KNIME Analytics Platform 4.6.2

Release date: September 21, 2022

Enhancements

- AP-19401: Dynamic tag set support in Tags to String node
- AP-19385: Add option to SMB Connector to set SMB version (defaults to 2.x)
- AP-16352: Create Temp Folder should expose flow variable in configure

Bug Fixes

- AP-19501: Node dialog not available in remote job view for various DB nodes
- AP-19484: Kerberos login with keytab sometimes fails when keytab contains weak key type
- AP-19413: Bypass Proxy Settings not working in 4.6.x versions of AP or Server
- AP-19399: XGBoost Classification Learner dialogs switch to soft prob objective
- AP-19380: Tag Filter Dialog doesn't load the selected tags
- AP-19374: Horizontal bar chart truncates bar labels
- AP-19308: Copy Component inside of itself breaks Workflow
- AP-19304: Pasted linked components cannot be edited after unlinking
- BD-1204: (Big Data Extensions): Hive to Spark on Local BDE can't access sun.util.calendar.ZoneInfo on date columns
- UIEXT-440: Filtering jobs in webportal does not use the filtered
- AP-19395: Uninformative exception message if the process queue fails to create a process
- AP-19372: Default for new Windows Installation is not "Bundled" but "Conda" for Python Script (Labs)
- AP-19303: Component update fails when the shared component is edited and component was not previously configured
- AP-19293: Google Drive Connector does not work properly on MacOS
- AP-19251: Fix static bar values in horizontal Bar Chart
- AP-19229: CSV Writer does not always escape correctly
- AP-19221: Excel Reader skips empty rows and columns, although the option is

unchecked in the "Advanced Settings" tab

- AP-19063: CSV Writer not placing quotes correctly if | is the delimiter

- AP-18904: Google Drive Connector should give better error message for files which are Google Sheets/Docs/Slides

- AP-18451: Workflow relative paths do no work in temporary workflow copy

- BD-1207: (Big Data Extensions): Spark DataFrame Java Snippet: Update Java version used for compiling code

- UIEXT-338: New Table View does not update on widget re-execution composite view

- UIEXT-422: Fast re-execution shouldn't result in an error

KNIME Analytics Platform 4.6.3

Release date: September 30, 2022

Bug Fixes

- UIEXT-519: Re-execution blocks local AP page in component views

- AP-19567: AP Installation: Redfield BERT nodes break KNIME update process on Windows

KNIME Analytics Platform 4.6.4

Release date: November 23, 2022

Enhancements

- AP-19585: Suggest/Force a full KNIME restart after a Java version bump

- AP-19115: Improved SAP Reader (Theobald Software) with support for KNIME destination

Bug Fixes

- AP-19610: Copy and paste should not add default node annotation to a node with empty annotation

- AP-19525: Call Remote Workflow Nodes failing with an Unsupported Media Type error

- UIEXT-536: Table View does not use full space when collapsed in the WebPortal

- AP-19699: CVE-2022-44749: Zip Slip vulnerability when opening workflows from untrusted sources

- AP-19697: Azure Blob Storage/Data Lake Storage Gen2 Connectors should not ignore proxy settings

- AP-19624: REST Proxy Detection fails to work correctly if System property `proxyHost` is blank

- AP-19608: SQLite Connector fails to read columns of type BLOB

- AP-19565: Enable long paths as part of the Windows installer

- AP-19530: Py4J logs are not suppressed on Windows

- AP-19389: Microsoft Authentication/SharePoint Online nodes should not ignore proxy settings

- AP-19084: Interactive range slider filter widget throws error when column contains only negative double values

- UIEXT-544: Sunburst chart selection is not working

KNIME Analytics Platform 4.6.5

Release date: February 1, 2023

Bug Fixes

- AP-19936: Microsoft Authentication node fails with NullPointerException when during Username/Password Authentication on a federated Azure AD Account

Open for Innovation

KNIME

KNIME AG
Talacker 50
8001 Zurich, Switzerland
www.knime.com
info@knime.com