KNIME Analytics Platform Installation Guide

Table of Contents

Installing KNIME Analytics Platform

ℹ️ 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.

   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.

    ℹ Also check the KNIME Quickstart 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.

i   **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.
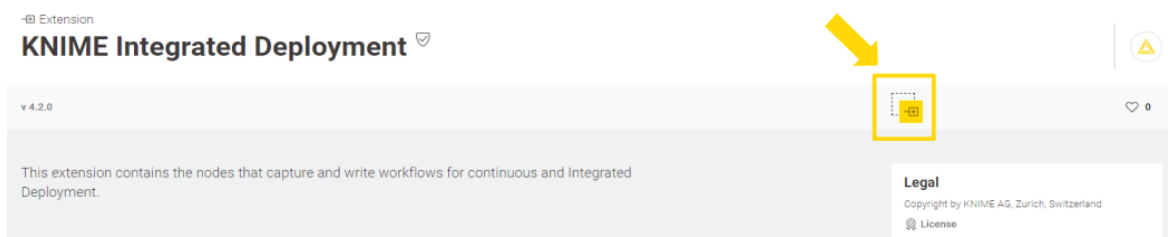

Figure 2. Install the KNIME Integrated Deployment Extension from KNIME Hub

  ◦ 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.
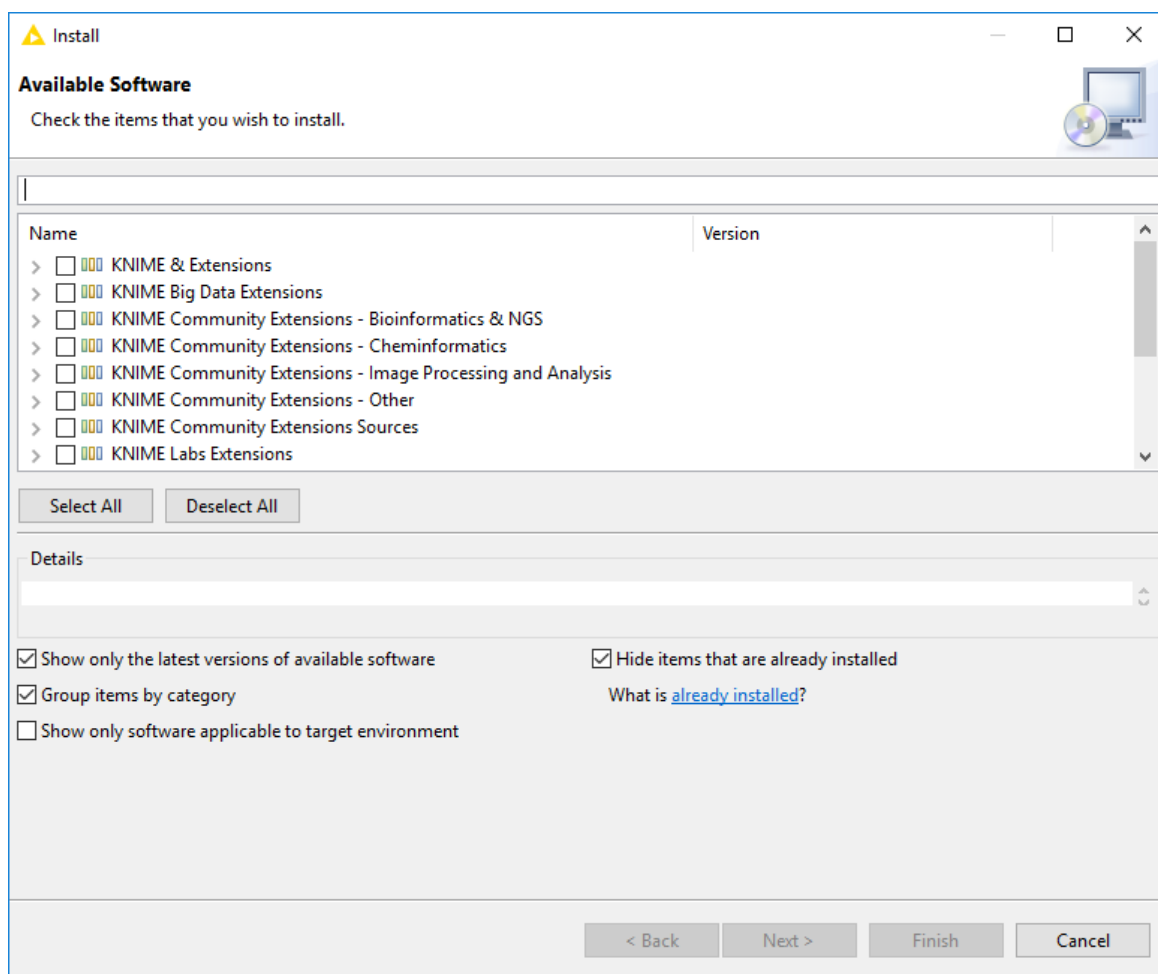
Figure 3. Installing Extensions and Integrations from KNIME Analytics Platform

- 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.

> ℹ️ 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*….
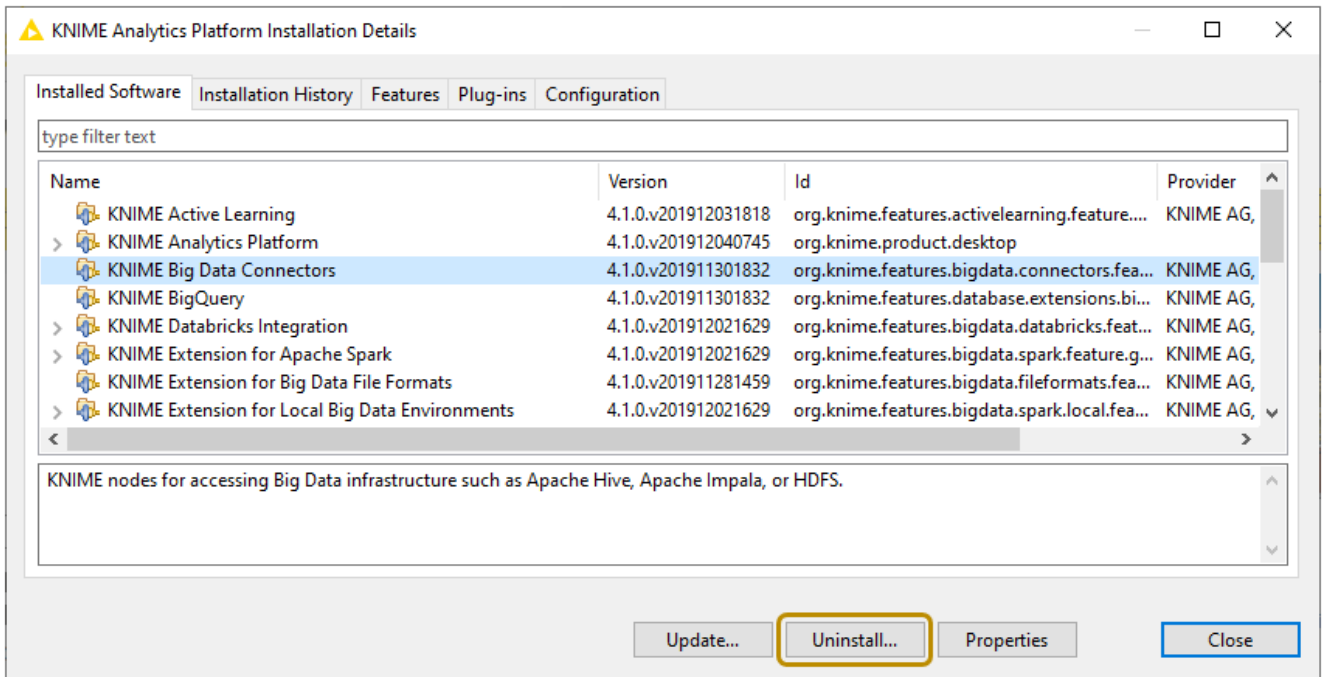
Figure 4. Uninstalling Extensions and Integrations

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:
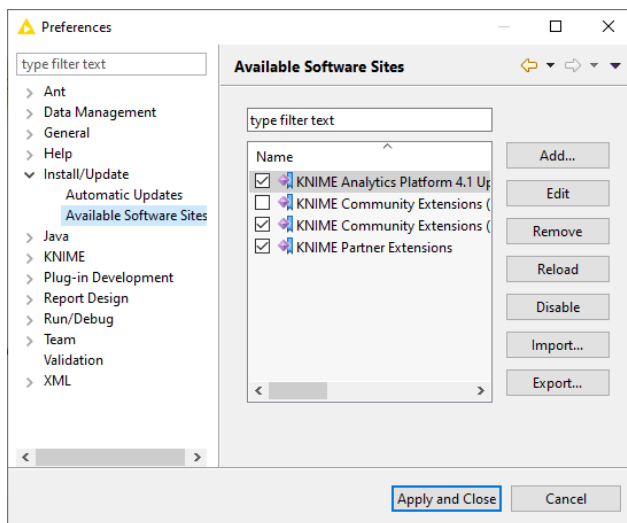
Figure 5. Available Update Sites

KNIME Analytics Platform 4.2 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 4.2 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.
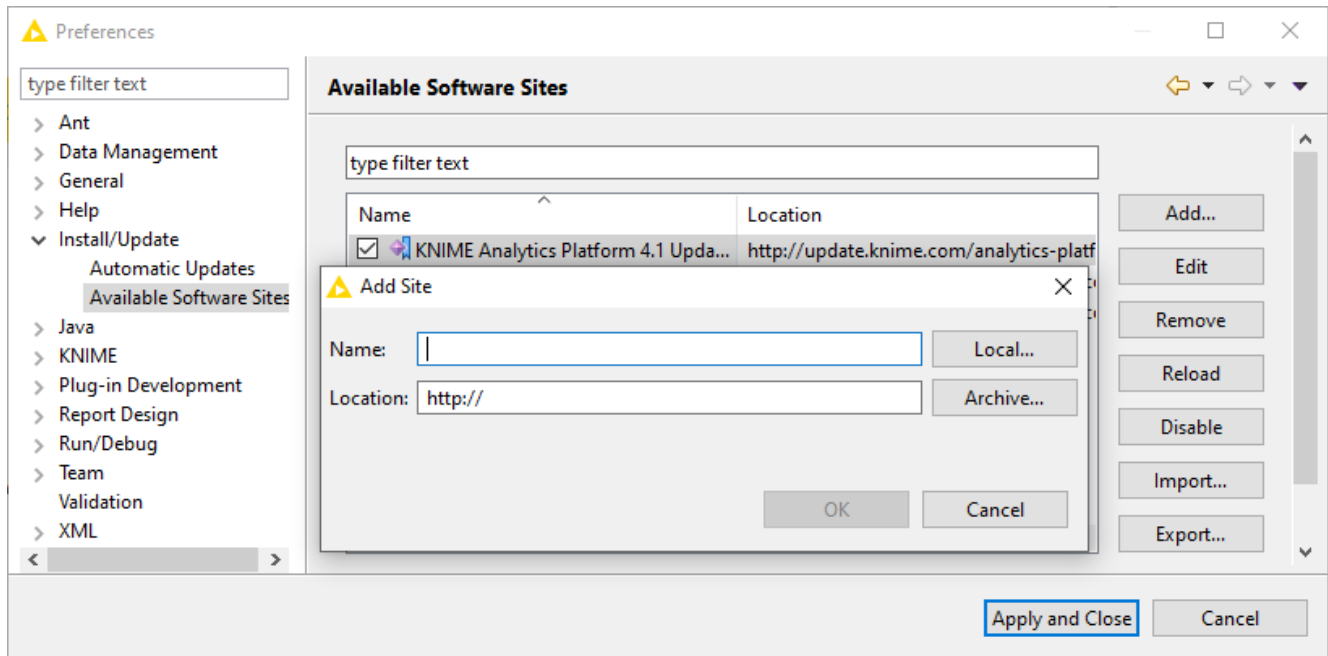
Figure 6. Add Update Sites

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. You can download KNIME update sites as zip files here.

- Save the zip file containing the extensions to your local system
- 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.
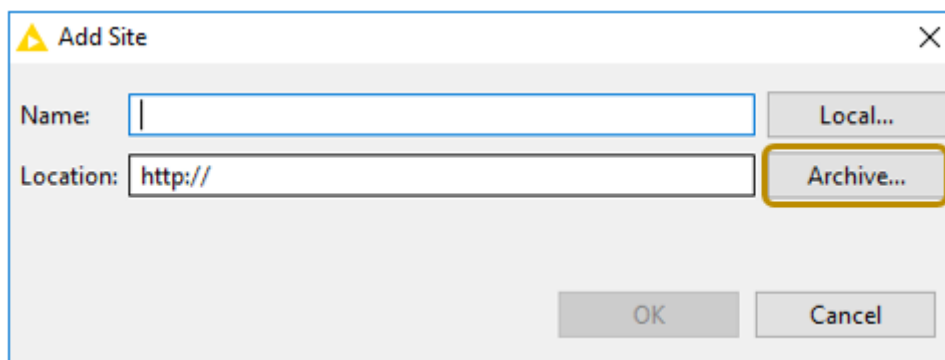
Figure 7. Adding Update Sites from Zip Archive

> **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.

- 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 4.2)

Detailed changelog for v4.2.x releases

KNIME Analytics Platform 4.2.0

(see highlight summary)
Release date: July 13, 2020

New node

- AP-14762: Integrated Deployment Nodes — Capture Workflow Start/End, Workflow Combiner, Workflow Writer, …

- AP-12359: Nodes for Amazon DynamoDB

- AP-14761: Various new reader and writer nodes supporting bulk reading and remote file system connection (Labs)

- AP-14679: Microsoft Authentication

- AP-14509: SAP Reader (based on Theobald Xtract Universal)

- AP-14447: SharePoint Online Connector

- AP-11268: Salesforce Connection (multiple nodes, e.g. Salesforce (OAuth2) Connector, SOQL Query, …)

- AP-14294: Simple File Reader — simplified usage, faster execution time

- AP-14279: TensorFlow 2 Keras Reader

- AP-14138: Google Cloud Storage Connector (Labs)

- AP-13886: String Manipulation (Multiple Columns)

- AP-13727: Create Temp Dir node

- AP-14763: Joiner (Labs) — multiple outputs, performance, UX, …

- AP-13477: Row Splitter (Labs)

- AP-13431: List Files/Folders (Labs) (new file handling)

- AP-11447: Table Difference Finder

Enhancements

- AP-14409: Update Java to JRE to version "8u252-b09"

- AP-13377: Update KNIME Analytics Platform to use the latest Eclipse (Eclipse 2020-03, i.e. 4.15)

- AP-14766: Widget nodes: New design in component views (old mode available via 'legacy' flag in layout editor)

- AP-14500: Add workflow data area to the "Relative To" category (new file handling)

- AP-14486: "Spotfire (Labs)" nodes deprecated — fully replaced by partner contributed nodes

- AP-14431: Textprocessing Performance Improvements

- AP-14421: Enable file text field in new file handling dialog in remote workflow editor

- AP-14400: Update Jackson to 2.11.0

- AP-14286: LineReader - Add BOM and gz support

- AP-14227: File menu entry to export a workflow summary

- AP-14178: Add duplicate checker to RowKey column

- AP-14161: Preference Page for Off-Heap Column Store activation

- AP-14090: urlDecode String Manipulator

- AP-14051: Register UrlEncode String Manipulator with Column Expressions Node

- AP-13990: Next functionality for node search

- AP-13943: String Manipulation node(s): new 'urlencode' method

- AP-13909: Support component drag'n'drop from private spaces (hub) into AP

- AP-13833: Colorize the connection lines of a selected node

- AP-13779: Metanode fails with "Error in sub flow"

- AP-13740: Support exchangeability of port types

- AP-13703: Create FSLocation Variable Type

- AP-13584: Table View (output port) with new option to copy data with header (to paste into Excel)

- AP-13548: Make 'Number to String' node streamable

- AP-13505: Disable dyn. port adaptations during execution and adopt warning message

- AP-13503: Make R environment configurable on a node level

- AP-13381: Add dynamic port support for Merge Variables node

- AP-13360: Add option to export summary / provenance trace from executed workflow

- AP-13344: Display column headers options should be added to the labeling view

- AP-13305: Make Column Appender with flexible Number of Input ports

- AP-13295: Add 'Configuration', 'Widget' and 'Container' node types

- AP-13272: Provide Ungroup node to skip rows with empty collection

- AP-13006: Top K Selector to give all occurrences of n biggest/smallest values

- AP-12999: Add portobject to possible porttypes of components

- AP-12806: Update Sorter UI for Sorter and top-k Node

- AP-12730: Add Node Monitor as default view

- AP-12726: Add float64 (double) datatype in Keras Input Layer

- AP-12666: Change default search mode in Table view to be case insensitive and to include column names

- AP-12620: Workflow Test Framework to Test for Errors in Components

- AP-12530: Add option to replace included columns in Column Combiner node

- AP-12469: Add select best option to the Feature Selection Filter node

- AP-12393: Allow to open dialog of DB nodes with missing/invalid input connection

- AP-12392: Support reconnecting to database if connection is invalid

- AP-11595: Improve formatting of long flow variable names in flow variable tab

- AP-11440: Implement SWT.OpenUrl event to support opening KNIME URLs for all OSs

- AP-11348: DB Loader: Support Microsoft SQL Server bulk copy API

- AP-11050: Support boolean columns in DB Row Filter

- AP-9898: Inform user about absence of flow variables in nodes with no settings

- AP-9198: Hide memory policy tab in nodes with no data table output ports

- AP-8121: Missing Value: Support fixed boolean value

- AP-7258: CSV Reader: Reduce number of opened input streams

- AP-6180: New 'hidden'-flag for node(set) extensions

- BD-1024: (Big Data Extensions): Add note about cluster-level permissions to node description of Create Databricks Environment

- BD-1006: (Big Data Extensions): Support Timezone option in Create Spark Context nodes

Bug Fixes

- AP-13853: Perl integration does not work with Perl >=5.30 any more

- AP-14550: Provide fallback for non-standard SQL types when reading from databases

- AP-14523: Flow Variables don't work on Salesforce Node

- AP-14492: H2O Integration: Local context issues on certain IPsec VPN setups

- AP-14429: Amazon S3 Connection (Labs) sometimes does not list all subdirectories of a directory

- AP-14420: Java Snippet does not clear warning messages on reset

- AP-14417: Tableau Hyper API: Node fail when data volume increases.

- AP-14406: Text Mining: Add invertible flag to StanfordNLP Tokenizers to ensure tokens not to be normalized

- AP-14114: Support non-standard JDBC Oracle types such as BINARY_FLOAT and BINARY_DOUBLE

- AP-14075: Parallel Chunk Start dialogue has bugs: missing flow variable and always greyed out checkbox

- AP-13947: Scorer node output column named specifity instead of specificity

- AP-13925: PortTypeRegistry fails ungracefully when encountering erroneous port types among all available port types

- AP-13923: REST nodes don't work with empty passwords in BASIC authentication

- AP-13763: Amazon Athena Connector node could not login without role switching

- AP-13760: Temporary table and file stores in loops not deleted on clear

- AP-13700: Table Creator: Copying of cells with cmd+c doesn't work on Mac

- AP-13691: Null pointer exception in Table Column to Variable node

- AP-13689: Wrong workflow svg stored when saving workflow while in annotation edit mode

- AP-13560: OutputStreams can remain open after workflow was closed

- AP-13538: Empty folders in workflow directory are missing from export

- AP-13528: Workflow editor locks up when saving large workflow using temporary workflow editor

- AP-13369: GET Request node: inconsistency with retrieved binary objects

- AP-13220: Math Formula: Check for column names containing dollar signs

- AP-13181: Fix in-workflow buffers not being cleared when workflow is closed and cleaned up

- AP-13176: Deadlock when using flow variable buttons with JTextFields

- AP-13104: Inner nested components duplicate error messages

- AP-12656: Word metanode still used when checking if update available

- AP-12111: Configuration of Call Workflow (Table Based) node fails, if wf on server contains File Download node in wrapped metanode

- AP-11772: Make dropping node on connections and nodes a uniform experience

- AP-11639: Executing follow up nodes block reset of non-related content of metanode

- AP-11451: Unused Metanode input causes warnings upon workflow opening

- AP-11328: FTP node's open() always calls changeDirectoryUp at least one time

- AP-9868: Make "Lift Chart (JavaScript)" node responsive

- AP-9721: Execute option to work on all nodes

- BD-1010: (Big Data Extensions): Parquet Reader and ORC Reader fail when reading directories with hidden files

- BD-817: (Big Data Extensions): Create Local Big Data Environment node fails to create root scratch directory on Windows if domain server is not accessible

KNIME Analytics Platform 4.2.1

Release date: August 26, 2020

New Nodes

- AP-14806: Salesforce Simple Query

- AP-13721: String to Path node

- AP-13434: Create directory (new file handling)

- AP-13432: Delete files no table input (new file handling)

Enhancements

- AP-14686: Container nodes with new option "Append unique ID to parameter name"

- AP-13549: Improve performance of Column Expression

- WEBP-417: Minor improvements on JS-view warning message pop-up

Bug Fixes

- AP-14849: Table View: Sorting a column will only display first subset of rows (scrolling possible but scroll bar height is misleading)

- AP-14809: Amazon Nodes (Athena/Redshift/S3/ML nodes) and Google Big Query nodes not visible on MacOS

- AP-14771: dropins functionality not working any more

- WEBP-463: Custom CSS is appended to the wrong iFrame in single views

- AP-14935: DB Loader throws exception for Date & Time columns without milliseconds for Microsoft SQL Server

- AP-14915: Execution of one metanode blocks editing of other metanodes (both contained in super metanode)

- AP-14912: Custom URL does not disable mode selection

- AP-14892: "Bit vector" cell factory available twice — causing problems in file reader dialog & call workflow nodes

- AP-14876: Connection errors in I/O nodes when reading/writing from/to server in high frequency

- AP-14875: Fix FilterOptions for folders in new DialogComponentFileChooser

- AP-14852: Workflow Writer not working using relative paths when using TeamSpace (client side extension when using KNIME Server)

- AP-14779: OSX/Linux: Always starts with KNIME theme for new workspaces – to handle dark theme user interface

- AP-14759: File Download renders empty link when no link title is provided

- AP-14322: Integrated Deployment: Workflow Executor may hang if if runs into thread exhaustion

- AP-13748: Error when saving workflow with failure in Recursive Loop End

- AP-10830: JS Pie/donut chart mixes up colors when there are missing values

- BD-1037: (Big Data Extensions): Create Spark Context (Livy) does not set Spark session owner when connecting to non-kerberized cluster

- BD-1011: (Big Data Extensions): ML based Spark Model Learner and Predictor nodes must not save/load model from local Spark driver filesystem

- WEBP-503: New style widget in single view should not load any css

- WEBP-473: File Upload should not use colons in file names

- WEBP-432: Legacy flag is not checked for components with nested components in the layout

- WEBP-380: Improve design of custom JS alert

- WEBP-365: Credential Widget not working with 'Use server account information'

KNIME Analytics Platform 4.2.2

Release date: September 28, 2020

Enhancements

- AP-15041: Extract Context Properties node should provide job ID on server

- AP-15007: Column Expression node to offer new (string) functions: "contains", "left", "right" (analogous to Excel)

- WEBP-535: File upload (master): use upload path with server host & increase security when accessing the file system on the server

Bug Fixes

- AP-15093: On Linux, editing workflow annotation with a background colour throws error

- AP-15058: Workflow Writer node dialog does not open

- AP-15035: Replacing a node after selecting multiple nodes causes workflow corruption

- AP-15196: JavaScript View framework: setValidationError method creates invalid JS

- AP-15155: Google Drive Connection node omits "cause" in case of connection issues

- AP-15149: Amazon Athena Connector assume role duration configurable via knime.ini parameter

- AP-15086: DB Table Creator node fails to create temporary tables for Redshift

- AP-15060: SimpleFileReader does not work on WindowsServer

- AP-15053: Create missing folders and overwrite settings not updated when opening dialogs

- AP-15051: Tagged Document Viewer error when selection disabled

- AP-15021: Tagged Document Viewer execution initially fails if default inputs are used

- AP-14858: Connection port missing from deploy to server node

- AP-14773: Extract Context Properties node has empty relativ workflow path and wrong absolut in queue-based executor

- AP-13997: Heatmap node sorting Row IDs if they contain int values

- BD-1056: (Big Data Extensions): DB Loader fails to load data on CDP 7.1 because LOAD DATA is not supported on transactional tables

KNIME Analytics Platform 4.2.3

Release date: November 4, 2020

Enhancement

- AP-15232: "-profileLocation" paths should be relative to installation folder (unless they are absolute)

Bug Fixes

- AP-15396: File Chooser throws error if default path is '/'

- AP-15286: "workflow.svg" looks slightly different (labels cropped) on MacOSX

- AP-15482: Visual Layout editor doesn't work, if apostrophe (') is used in node annotation

- AP-15454: Fix Path sorting for PathAccessor and FSFiles

- AP-15384: Athena Connector does not report proper exception if S3 authentication fails during dialog opening

- AP-15371: Reading from Custom URL fails when URL contains percent-encoded characters in query or fragment

- AP-15362: SharePoint Online Connector directory creation has race condition that causes wrong error to be thrown

- AP-15313: Resetting connector node together with Create Temp Dir node sometimes results in NPE

- AP-15307: Labeling View Widget buttons order is wrong

- AP-15278: AbstractDialogComponentFileChooser doesn't always update the status message properly

- AP-15207: Windows Defender handling fails with recent Windows updates

- AP-15151: Workflow Writer "Deployment Options" are gone

- AP-15076: Conda env specification via executor.epf seemingly ignored on distributed executors

- AP-13814: Conditional Box Plot has a strange sorting for the nominal values when containing 00, 01, 10 …

- WEBP-565: File Upload Widget receives status 403 when used on KNIME Server Small

- WEBP-555: Large file uploads crash Chrome and Edge

- WEBP-547: Padding of button on File Upload Widget

- WEBP-524: JavaScript views are not working in legacy Edge

KNIME Analytics Platform 4.2.4

Release date: December 14, 2020

Bug Fixes

- AP-15809: NTLM (Labs) authentication in KNIME REST nodes fails with NoClassDefFoundError

- AP-15883: Workflow Writer crashes AP (caused by xgboost port object)

- AP-15868: Thread leakage in Logfile appender

- AP-15723: DB Connection Closer node should log a warning instead of an error if connection is already closed

- AP-15557: NPE when capturing workflow which is part of a Parallel Chunk Loop

- AP-15551: NullPointerException when persisting workflow summary with job

- AP-15550: NullPointerException in Call Workflow (Table based) node

- AP-15542: Workflow summary generation fails for nodes in an inactive branch with absent flow variable

- AP-15476: Writing workflow summary locks up executor

- AP-15249: XGBoost Linear Ensemble Learner (Regression) Thrifty error

- AP-10548: Workflow manager can only load one workflow at a time (possible temporary freeze when used via KNIME Server)

KNIME Analytics Platform 4.2.5

Release date: March 4, 2021

Enhancements

- BD-1120: (Big Data Extensions): Create Databricks ENV Node should allow user to overwrite the AuthMech JDBC parameter

Bug Fixes

- BD-1121: (Big Data Extensions): Handle Databricks File System REST API changes

21