Paxata Release Notes

Technical Release Notes 2019.1 Release

This document is produced by Paxata as a reference. It makes no warranties or guarantees. Information contained within may not be the most up-to-date available. Please refer to the Paxata website for the most current information.

All trademarks within this document are property of their respective companies.

Apache Hadoop® is a registered trademark of The Apache Software Foundation.

Avro™, HDFS™, and Apache Hive™ are trademarks of The Apache Software Foundation.

Cloudera™ and Cloudera CDH™ are trademarks of Cloudera, Inc.

CentOS™ is a trademark of the CentOS Project.

Chrome™ is a trademark of Google Inc.

Excel® and Windows® are registered trademarks of Microsoft Corporation in the United States and/or other countries.

Firefox™ is a trademark of the Mozilla Software Foundation.

Hortonworks™ and Hortonworks HDP™ are trademarks of Hortonworks, Inc.

Java™ and JDK™ are trademarks of Oracle Corporation.

Linux® is the registered trademark of Linus Torvalds in the U.S. and other countries.

Mac OS X® is a trademark of Apple Inc., registered in the U.S. and other countries.

MongoDB™ is a trademark of MongoDB, Inc.

Red Hat® is a registered trademark of Red Hat, Inc.

SUSE® is a registered trademark of SUSE LLC in the United States and other countries.

© 2019 Paxata, Inc.

The information in this document is subject to change without notice. Paxata shall not be liable for any damages resulting from technical errors or omissions that may be present in this document, or from use of this document.

Note: Technical documentation and support materials include details based on the full set of capabilities and features of a specific release. Please note that individual access to specific functionality may vary based on deployment and license types.
# TABLE OF CONTENTS

- NEW IN THIS VERSION  INTELLIGENT AUTOMATION .................................................. 4
- CHANGED IN THIS VERSION ........................................................................................ 13
- ENHANCEMENTS FOR ON PREM CUSTOMERS ...................................................... 13
- RESOLVED IN THIS UPDATE ..................................................................................... 13
- APPLICATION AND CONNECTOR SUPPORT ............................................................ 15
NEW IN THIS VERSION

INTELLIGENT AUTOMATION

Intelligent Automation now extends automation capabilities to curated data flows through the Automatic Project Flow (APF) feature, which allows you to intelligently operationalize curated data flows. With a single click, APF computes the entire sequence of data prep Steps across Paxata Projects, datasets and AnswerSets required to produce an end-to-end, automated output Flow for your data. You can run the Flow just once to produce an end-result AnswerSet or set the Flow to run on a recurring time-based schedule. All subsequent runs can then be easily managed through the Monitoring Interface.

The following example demonstrates how the Intelligent Automation engine calculates all of the Flow dependencies required to produce a single AnswerSet from various other Paxata Projects and dataset inputs:

Using the APF feature, a single click in the "Sales Variance Report" Project creates the Flow. The APF Configuration Interface then allows you to set triggers for automatically running the Flow, and the APF Monitoring Interface provides dashboard controls to manage it.

With APF, Business Analysts and Data Engineers can simplify complex data flows by breaking them into smaller groups of Paxata Projects that can be operationalized—with each Project focused on performing a related or cohesive set of Steps for improved readability and limited complexity. When the Projects are finalized, simply select the final Project in the sequence as the "target" Project. APF takes care of the rest—sequencing, preparing and automating the entire end-to-end flow without any manual stitching required.

For teams that require input from both Business and the IT Leaders, the data prep process is simplified because everyone completes their data prep work in their own Paxata Project, and then the entire sequence is operationalized from a single "target" Project. Again, APF takes care of the rest with no manual stitching required, regardless of who created or owns the Projects and AnswerSets.
Paxata Release Notes

After building a Flow, an APF graph allows you to quickly and easily visualize your datasets and AnswerSets, and how they flow into the individual Projects that are used to generate the Flow’s final output:

Hovering over any dataset or Project in the Flow also displays the corresponding downstream lineage (in pink) and upstream dependencies (in blue). For example, when hovering over the dataset for March 2016 Transactions:

When hovering over an intermediate Project in the Flow—in this example Customer Loyalty-Women Members—the upstream dependencies display through the blue lines while the downstream lineage displays through the pink lines.

Notice in both examples above that if datasets and Projects do not participate in the portion of the Flow that you’ve selected, then they are grayed out in the graph.

See the application’s inline help and the **Automatic Project Flows Help Shelf** article for complete details.
NEW CONNECTORS, ENHANCEMENTS AND FILE TYPE SUPPORT

The following new connectors are now available:

- JIRA
- NetSuite
- Oracle Marketing Cloud (Eloqua)
- Salesforce Marketing Cloud
- ADLS Gen2
- Google Analytics
- MongoDB
- Amazon Athena for querying data in S3 buckets
- Tableau version 2 with support for Hyper extracts

Contact Paxata Customer success to have these connectors installed. Refer to the Data Source Support matrix for the details of each connector.

**SharePoint Connector enhancements:**

- New support for column type: "Multiselect Person or Group". Values from this column are imported as type text. The separate values within a "cell" are delimited by commas and then imported into Paxata.
- Enhanced user messaging: a warning is now displayed when there are columns that the SharePoint connector cannot import from a custom list because those column types are not supported.

**JDBC Connector enhancements:**

- General error handling enhancements to provide more useful information when an export fails.
- Provide an error to user when a dataset includes a column that exceeds a database's max length for a column name.

**New file type support:**

- Import from SAS BDAT
- Export Paxata AnswerSets to Excel XLSX
COMPUTED COLUMNS ADDITIONS AND ENHANCEMENTS

Computed Columns functions have been expanded to include 33 new functions:

- **String**
  - TRIM
  - TRIMLEFT
  - TRIMRIGHT
  - PADLEFT
  - PADRIGHT
  - REPEAT
  - REVERSE
  - CHAR

- **Search, Extract, Replace functions**
  - REPLACE()
  - REGEXPTRACT()
  - REGEXPPLACE()
  - SUBSTITUTE()
  - SEARCH()

- **Numeric functions**
  - CEILING()
  - FLOOR()
  - ROUNDPERC()
  - EXP()
  - FACTORIAL()
  - LN()
  - LOG()
  - LOG10()
  - MOD()
  - POWER()
  - QUARTER()
  - SIGN()
  - SQRT()

- **Date functions**
  - DATE()
  - DATEDIFF()
  - ENDOFMONTH()
  - FROMUNIXTIME()
  - NETWORKDAYS()
  - WEEKOFYEAR()
  - WORKDAY()

- Trim can now be in expressions (in addition to column menu option)
- Repeat strings N times

Address use cases involving replacement of text. Example: Regexp replace can look for digits in credit card field and replace with a n asterisk.

More parity with Excel functions:
- CEILING
- FLOOR
- ROUNDDOWN
- ROUNUP

More log functions

Referential Date function, for example:
- endofmonth()
- weekofyear()

See the application's **Computed Column Overview and Function Reference** Help Shelf article (in the application) for details.

Additionally, the Computed Columns panel provides an option to disable automatic update when you prefer to manually update the grid with the results of your formula. By default, auto-update is enabled and the timer indicates the count-down until auto-refresh of the grid occurs. To disable auto-update, de-select the option and a green update button appears in place of the timer. Click the Update button to manually refresh the grid any time while you're building a formula.
Paxata Release Notes

**DATE CONVERSION FUNCTIONALITY IN COLUMN MENU**
Date Conversion functionality is now built directly into the column menu. This allows you to easily convert a column to a date type instead of needing to create a computed column Step in your Project with an expression to convert it.

1. Select the "into date" option now available from the column drop-down menu:

   ![Column Menu Screenshot]

   2. The Change menu opens. Select the format of your date data as it currently exists in the column—for example, *yyyy-MM-dd*:

   ![Change Menu Screenshot]

   Note that you can also select "Custom" to provide a custom format.

3. When you save your selection, notice the column's type is converted to date type as indicated by the timer symbol.

   ![Column after Save Screenshot]

   The Change menu also allows you to convert multiple columns to date type in single Step:

   ![Multi-Column Change Menu Screenshot]

   Refer to the Change Menu Help Shelf article (in the application) for details.
FILTERGRAMS ENHANCEMENTS
Data exploration with Filtergrams has been accelerated in newly created Projects. The moment you bring your base dataset into a new Project, you can immediately begin exploring your data using Filtergrams—without additional clicks or saves to your Project.

Filtergrams also now include a new Dynamic Percentiles feature to specify percentiles of the selected Filtergram values. For example, if you have an inventory dataset that has a column for products sold per week per region, you can use Dynamic Percentiles to filter select the top %5 highest selling products for each region.

Your percentile selections are also dynamically applied to newer versions of datasets that are automatically updated in the Library through Automatic Project Flows. For example, you can use the APF feature with Dynamic Percentiles to automatically produce an AnswerSet each week to identify the top %5 highest selling products for each region per week.

PROJECT STEPS COPY AND REUSE IN OTHER PROJECTS
The new Project Steps reuse feature enables Project collaboration with team members and enhances productivity through Steps reuse. This feature allows you to easily copy Steps from a Paxata Project and then reuse those Steps elsewhere in the same Project or copy them into another Paxata Project. When you copy the Steps, you have the option to copy them to your computer's clipboard for a one-click paste. Alternatively, you can copy the Steps to a file, which can be saved for later use or shared with others.

Refer to the Steps Reuse Help Shelf article (in the application) for details.

IMPORTANT: due to limitations in the current ESR version of the FireFox browser, this feature is only supported for use in the Google Chrome enterprise browser.
Paxata Release Notes

SHAPING OPERATIONS ENHANCEMENT

Group by operation Deduplicate introduces a new Fuzzy option to deduplicate similar values. When this feature is enabled, you see the "Fuzzy" option for Deduplicate as an alternate selection to "Exact" match:

Fuzzy deduplication finds matching rows using a fuzzy algorithm. This means similar values, that are not exact matches, are grouped together and deduplicated. For example:

Similarly, Fuzzy Deduplication will group similar items even if they have blank values:

To enable this feature, contact Paxata Customer Success and request that Fuzzy Deduplication be enabled.
Paxata Release Notes

LIBRARY ENHANCEMENTS
One-click Project creation from the Library allows you to create a Project from a dataset. Hover over a dataset in the Library, click the Create Project button, and a new Paxata Project is automatically created and saved with that dataset as your base dataset.

In-line Library functionality is also now available when selecting a Lookup or Append operation in your Project—available datasets for selection are then displayed within the context of your Project:

Additionally, from the Project, you can explore data sources and then import a dataset to use for the Lookup or Append operation:
Paxata Release Notes

Library: data source listings now alphabetical. When viewing the directories and files on any data source, all results are sorted alphabetically, beginning at the directory level, subdirectory, and then again alphabetically within each directory. Note that sorting is case insensitive.

Library: enhanced navigation through Library pages using breadcrumbs displayed at the top of each page.

Library: import of fixed width files with custom lines separators /r and /n are now supported.

Library: CSV files with mismatched quotes can now be easily remediated during import. For ill-formed CSV files that do not have matching close quotes, the parse panel now provides an option to treat a delimiter as the implicit close quote when a close quote is missing.

BULK COLUMN RENAMING
Bulk column rename allows you to rename all columns using a single, comma-separated string. Simply begin typing the new column names in the field below the Columns List panel and separate each name with a comma. Notice that column names displayed in the Columns List panel update accordingly:

Note that you can also paste in new column names from header files, separated by commas, to quickly rename all of the columns in your dataset.

JOIN DETECTION ENHANCEMENTS
Join detection algorithm enhancements to provide more accurate join suggestions when a Project has a majority of columns with single values.

ERROR MESSAGE ENHANCEMENTS
Enhanced error messages with remediation details when there are network connectivity issues.
Paxata Release Notes

**CHANGED IN THIS VERSION**

- For customers who are currently using the 2018.2 Automation feature to automate datasets and Projects: Automatic Project Flows (APF), with its enhanced capabilities to handle complex flows, will replace the Automation feature. By default, the new APF feature is *not* enabled until you contact Paxata’s Customer Success team to enable it and assist you in migrating your current automation jobs over to the new APF environment.

- Project Step behavior enhancement: when a Step is in edit mode and no changes have been applied, the Save button is disabled and a newly introduced Close button is used to exit the edit mode.

**ENHANCEMENTS FOR ON PREM CUSTOMERS**

- New `px.properties` options:
  - Specify an email address for automation results: a new property for configuring account name, that is different from the `px.email.from.address` property, is now supported. When set, the new property—`px.email.account`—is used as the account 'username' for authorization. If this new property is not set, the `px.email.from.address` is used as both the username of the account for authorization and the 'from' address for automation results emails.
  - Specify locale to JP by default: `px.default.locale` is a new property that can be set to values of `en` or `jp`. This setting assigns a default user interface locale to all new users that are created after the setting is enabled. By default, the setting is `en`.

**RESOLVED IN THIS UPDATE**

- Computed Column Expressions: the `@Esc` keystroke to locate matching column names does not locate matching column names that begin with double-byte characters. *Original issue reported under Paxata JIRA ticket: PROJ-4122*

- Role management in UI contains a "Database" Group and "Access JDBC/ODBC" permission though these have been deprecated. Now removed from UI. *Original issue reported under Paxata JIRA ticket: PLAT-1698*

- Library import option 'Row Separator' only accepts one character when 'Use Hexadecimal' is unchecked. Now fixed and accepts multiple characters. *Original issue reported under Paxata JIRA ticket: LIB-3584*

- Deleting multiple versions of datasets results in server out of memory error. Now fixed. *Original issue reported under Paxata JIRA ticket: PLAT-1616*

- Computed Columns: functions require additional time to complete because, by default, the cursor is not placed inside the parenthesis when a function is selected. Now fixed. *Original issue reported under Paxata JIRA ticket: PROJ-3831*

- SAML users: Import tutorial video displays every time users login, even after the option to not display again is selected. *Original issue reported under Paxata JIRA ticket: UI-30*

- SAML users: after successfully logging into the IDP page, users are directed to login to Paxata. However, Paxata rejects users and redirects them back to IDP login page. Now fixed. *Original issue reported under Paxata JIRA ticket: PLAT-1968*
• Library: when deleting a dataset, an erroneous warning displays to indicate the dataset is currently being used by a Project—even when that Project no longer exists. Now fixed.  
  *Original issue reported under Paxata JIRA ticket: PROJ-2316*

• When a new Project Step that introduces a new column is added, multiple refreshes occur on the data grid before the user can continue working in the Project. Now fixed.  
  *Original issue reported under Paxata JIRA ticket: PROJ-2989*

• Library: import fails, but storage directory for the import is still created, and with an incorrect path. Now fixed and the failure gracefully errors out with a message to user that import did not succeed.  
  *Original issue reported under Paxata JIRA ticket: LIB-3382*

• Automation: the UI for editing an existing automation job does not correctly persist the newly saved date/time schedule and instead persists the initial (older) date, thus resulting in an error that you cannot save the newly scheduled time because it is a date that has already passed on the server. Now fixed.  
  *Original issue reported under Paxata JIRA ticket: AUTO-762*

• Local export fails when an exported dataset contains a comma in the name or when user provides a name containing a comma. Now fixed.  
  *Original issue reported under Paxata JIRA ticket: LIB-3274*

• Teradata BYTEINT columns are imported as Paxata Boolean. Now fixed and BYTEINT columns imported as numeric. Note: if you need to convert the numeric column type to Boolean, use a computed column IF-THEN expression to convert the values in the column.  
  *Original issue reported under Paxata JIRA ticket: CONN-1565*

• Security updates to enhance blocking of cross origin web socket requests.  
  *Original issue reported under Paxata JIRA ticket: PLAT-2030*

• HDFS connector: now provides a more specific error message to indicate that user cannot connect to a data source because user does not have required permissions to read a file.  
  *Original issue reported under Paxata JIRA ticket: CONN-88*

• BundleFinder errors spamming log files: messages indicating that BundleFinder has found duplicate keys or LocalizerImpl is missing localization messages. Now fixed.  
  *Original issue reported under Paxata JIRA ticket: PROJ-4419*

• HornetQ messages too verbose. Now fixed and message sizes are reduced.  
  *Original issue reported under Paxata JIRA ticket: PLAT-1325*

• Failure in Project view with CrudServiceMessages error: now fixed with improvements in the exceptions that are logged for the queue management system so that log files can be more efficiently managed.  
  *Original issue reported under Paxata JIRA ticket: PLAT-2155*

• Frontend and pipeline logs HTTP 400 error when no dataset is selected for an append or lookup Step. Now fixed and a step level warning has been added for append and lookup steps when no dataset is selected.  
  *Original issue reported under Paxata JIRA ticket: PROJ-4434*

• Shaping operations Group By and Pivot: when utilizing the statistical functions var, varp, stdev, stdevp over a large numbers of rows, >100k, the resulting new column may be incorrectly calculated. Now fixed.
APPLICATION AND CONNECTOR SUPPORT

- Browser support:
  - Google Chrome Version 78.0.3887.7 (Official Build) (64-bit) for Mac and Windows
  - Mozilla Firefox: Extended Support Release (ESR) 68.0.2 (64-bit) for Mac and Windows

  Note: the minimum resolution for the Paxata application is 1024x768

- Please refer to the Data Source Support matrix for the details on the latest and currently supported connectors.

Note: please refer to the Known Issues and Unsupported Functionality list for details of open issues and unsupported functions.