User Guide

Data Preparation R-1.2

Contents

1. About this Guide

1.1. Document History

Product Version	Date (Release date)	Description
BizViz Data Preparation 1.0	August 31st, 2017	First Release of the document
BizViz Data Preparation 1.1	December 11th, 2017	Updated document
BizViz Data Preparation 1.2	April 15th, 2018	Updated document

1.2. Overview

This guide covers:

- Introduction and steps to use the Big Data BizViz ETL plugin
- Configuration details for the Data Preparation components

1.3. Target Audience

This guide is aimed at business users of all skill levels who deal with vast amounts of data and requires data preparation to be attempted before getting informative insights from the collated business datasets.

2. Introduction

2.1. Introducing the Big Data BizViz Data Preparation

The BDB Data Preparation is a self-service data preparation tool that empowers data-driven Business users with powerful capabilities to extract, transform, and merge new data sources. The tool offers a range of components to transform and merge the selected dataset. Users can get analytics-ready data faster to generate valuable insights in less time.

2.2. Prerequisites and Supported Devices

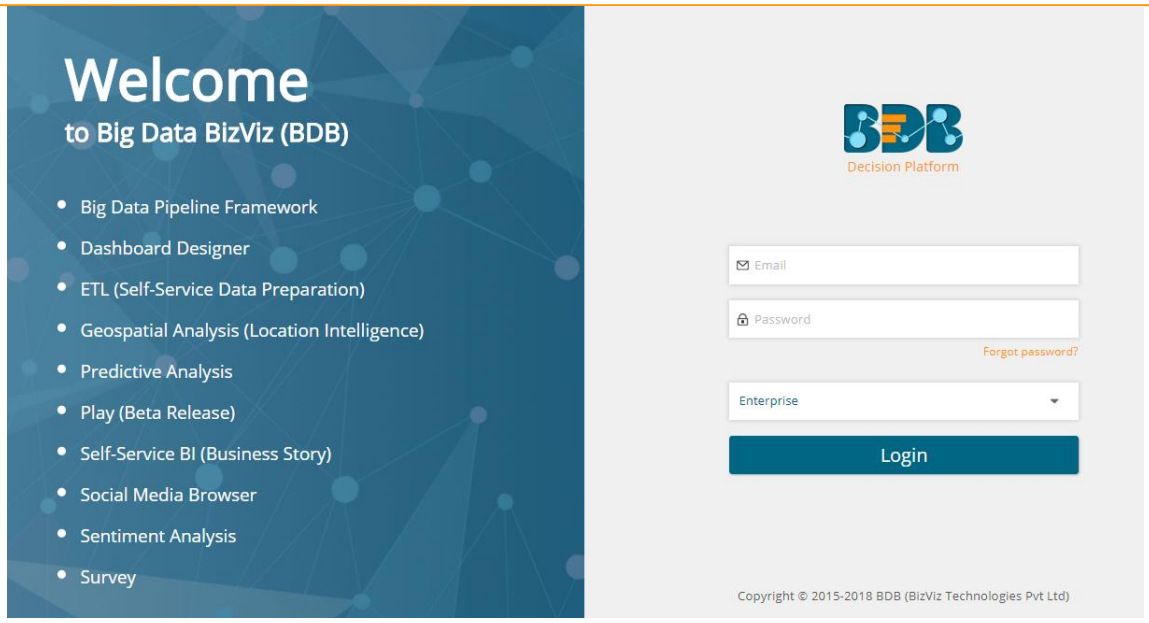
- A browser that supports HTML5
- Operating System: Windows 7
- Basic understanding of the BizViz Server

3. Getting Started with the BDB Data Preparation

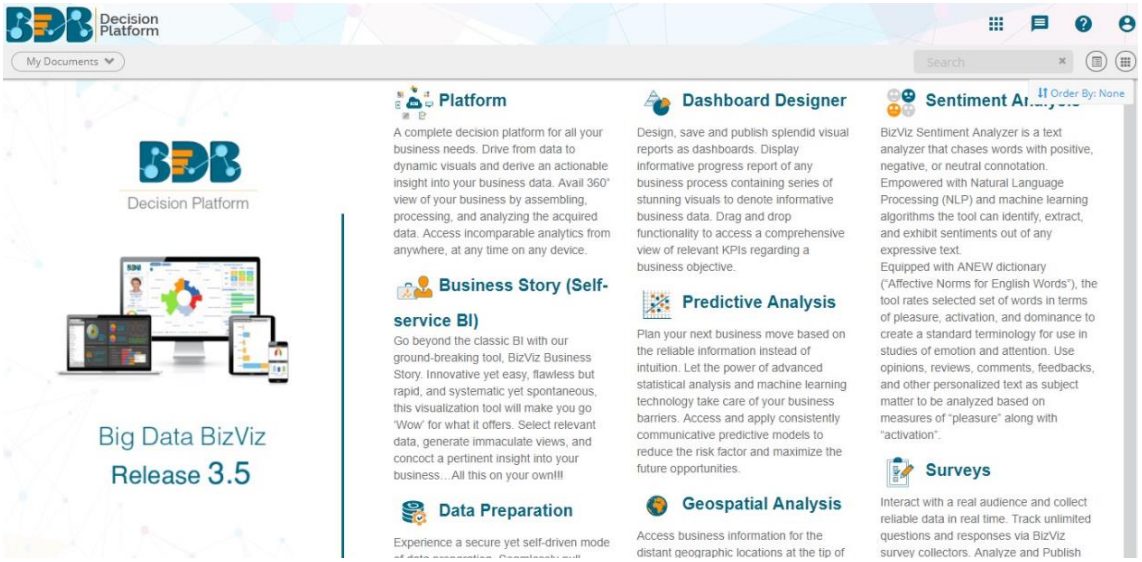
3.1. Accessing the BDB Data Preparation

This section explains how to access the BizViz Platform and a variety of plugins that it offers:

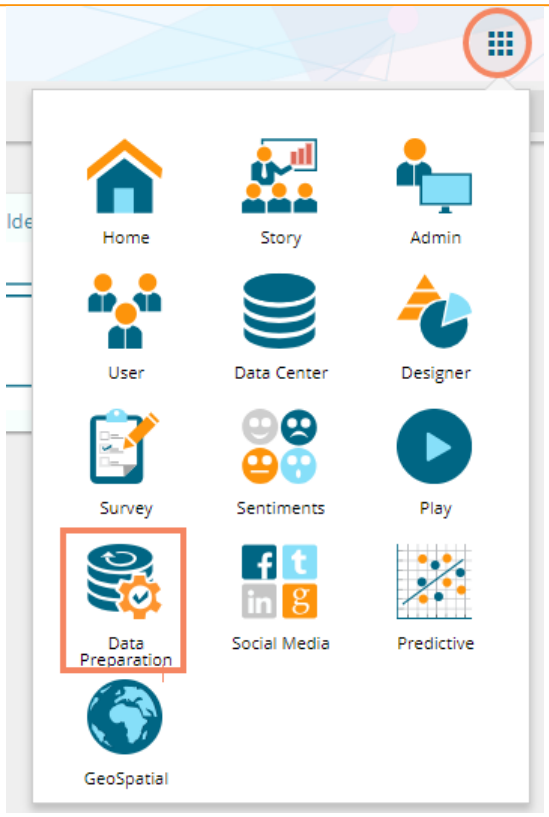
i) Open BizViz Enterprise Platform Link: http://apps.bdbizviz.com/app/
ii) Enter your credentials to log in to the platform.
iii) Click **'Login'**



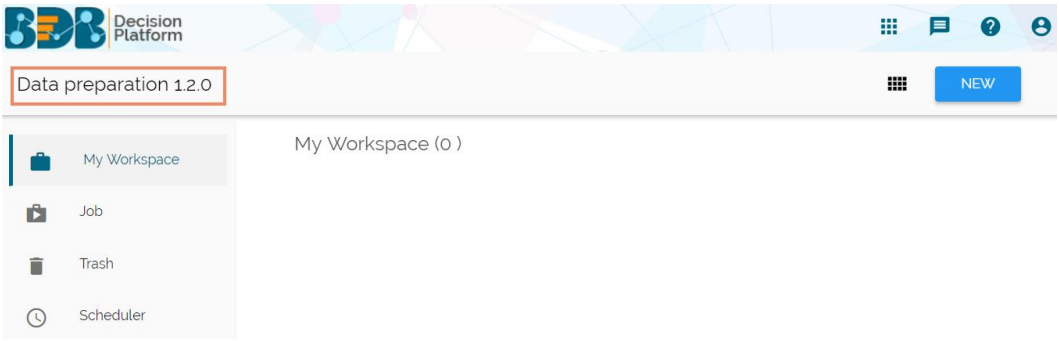
iv) BizViz Platform home page will open.



v) Click on the '**App**' menu option
vi) All the available plugins will be listed in the displayed window
vii) Select the '**Data Preparation**' plugin



i) Users will be redirected to the Data preparation landing page.
ii) Users will find four major modules on the Data Preparation landing page:
   a. My Workspace (Default Component)
   b. Job
   c. Trash
   d. Scheduler

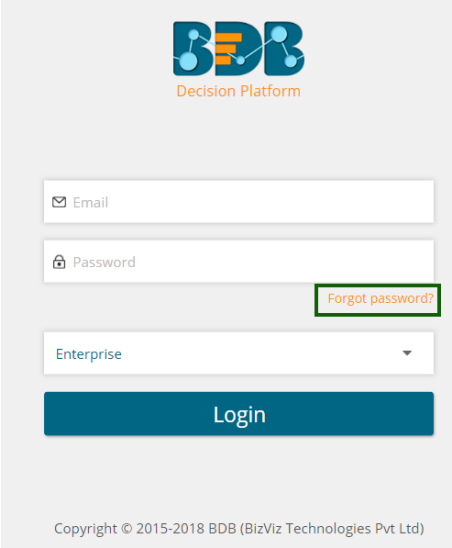


This document will describe all the major components and the related workflows at details.

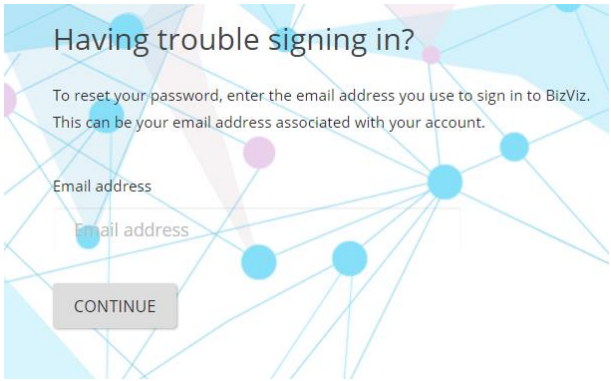
3.2. Forgot Password Option

Users are provided with a choice to change the password on the Login page of the platform.

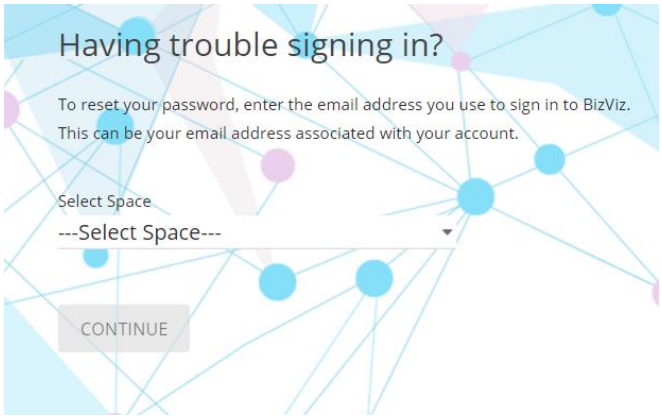
i) Navigate to the Login page.
ii) Click '**Forgot Password?**' option.



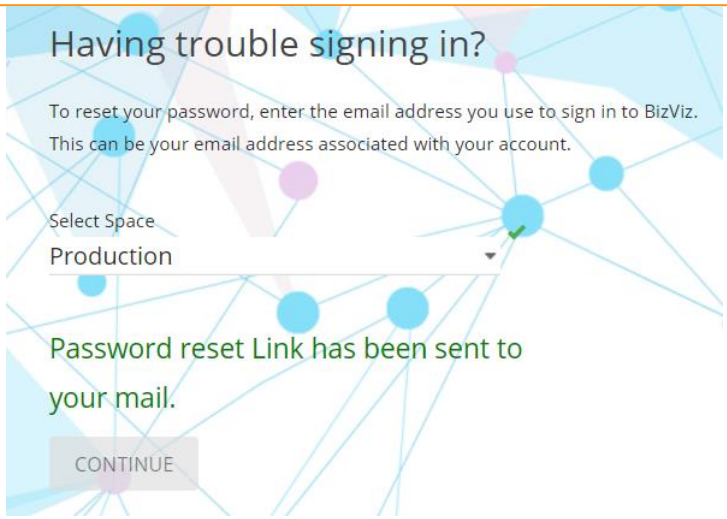
iii) Users will be redirected to a new window.
iv) Provide the email id that is registered with BDB to send the reset password link.
v) Click the '**Continue**' option.



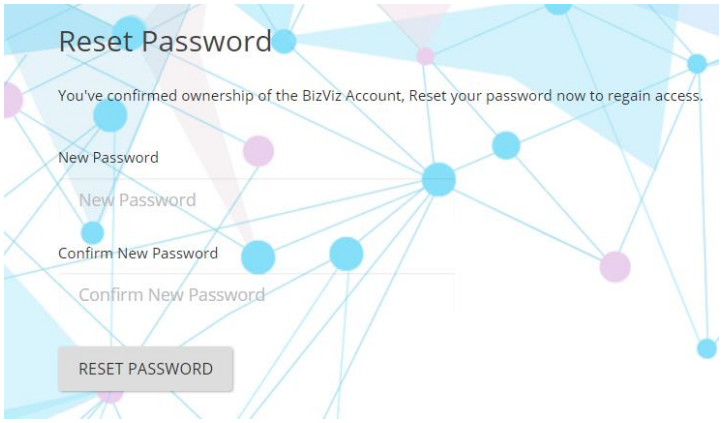
vi) Users will be redirected to select a space if needed and click the 'Continue' option.



vii) A notification will appear stating that the reset password link has been sent to the registered email.



viii) Click the link from your registered email
ix) Users will be redirected to the 'Reset Password' page to set a new password
x) Set a new password
xi) Confirm the newly set password
xii) Click 'RESET PASSWORD' option



xiii) The password will be successfully reset for the selected BDB account.

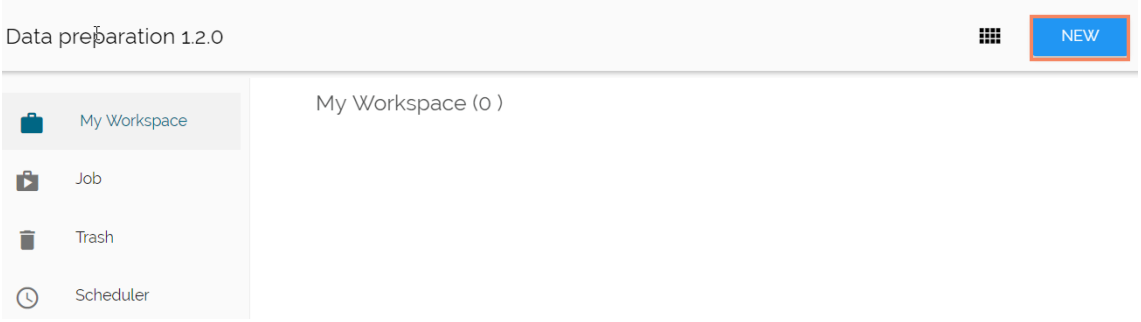
4. Basic Features

The landing page of Data Preparation launches workspace view. '**My Workspace**' will be displayed by default.

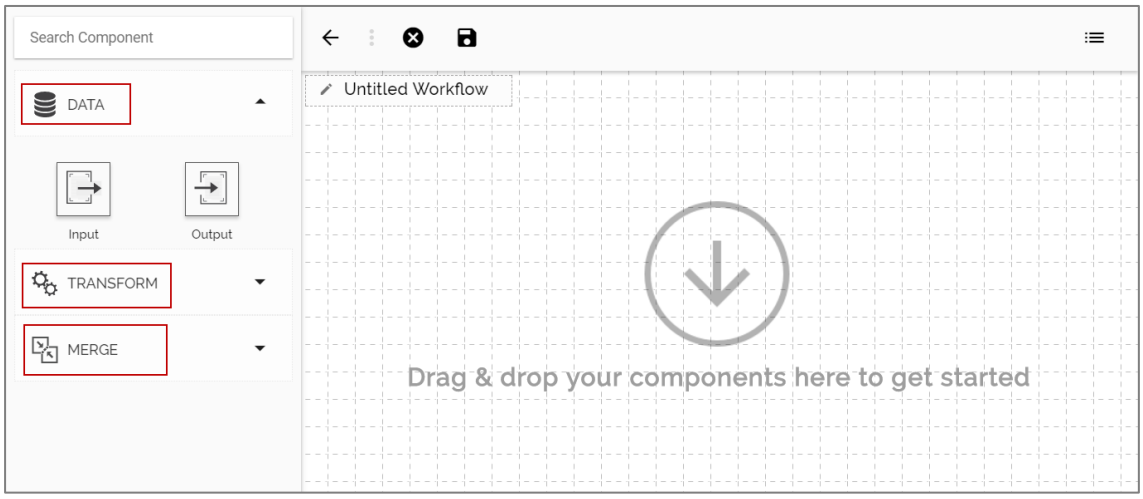
4.1. Workflow Editor

'**My Workspace**' is a placeholder for the workflows which are created using various data preparation components. Users can create the workflows using the workflow editor.

i) Navigate to the '**Workspace**' page.
ii) Click '**New**'

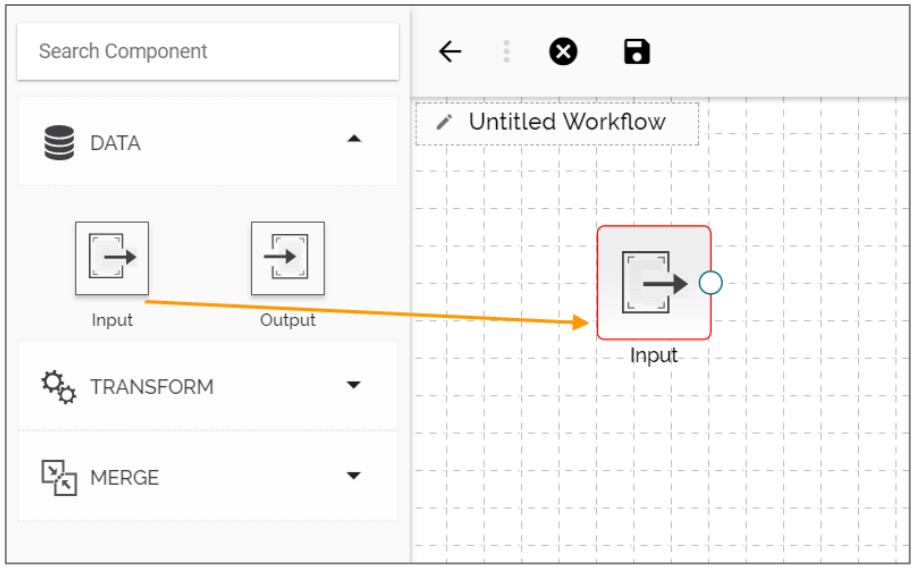


iii) Users will be redirected to the **'Workflow Editor.'**
iv) The Workflow editor exposes users to 3 main aspects to autonomously prepare data:
   a. Data
   b. Transform
   c. Merge

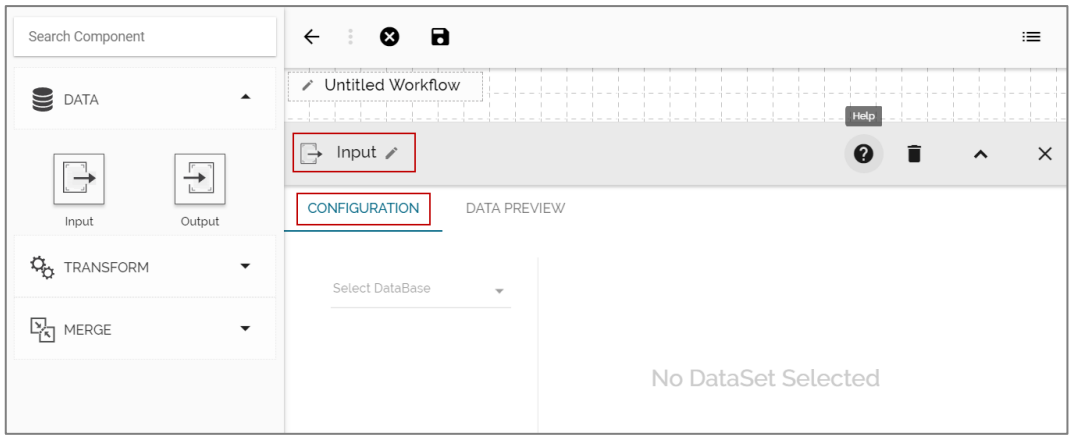


4.2. Extracting Data: Full and Incremental

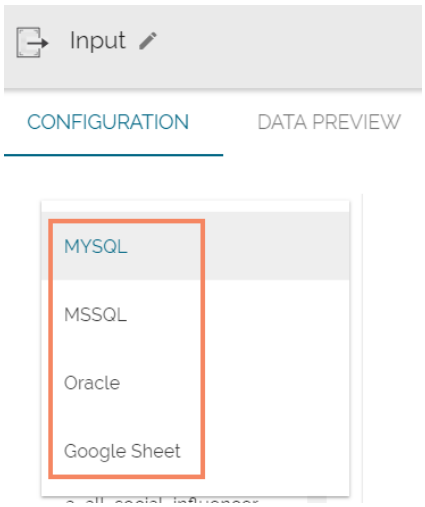
i) Navigate to the Workflow Editor.
ii) The **'Data'** option will be selected by default.
iii) Drag and Drop the **'Input'** component onto the workflow editor.



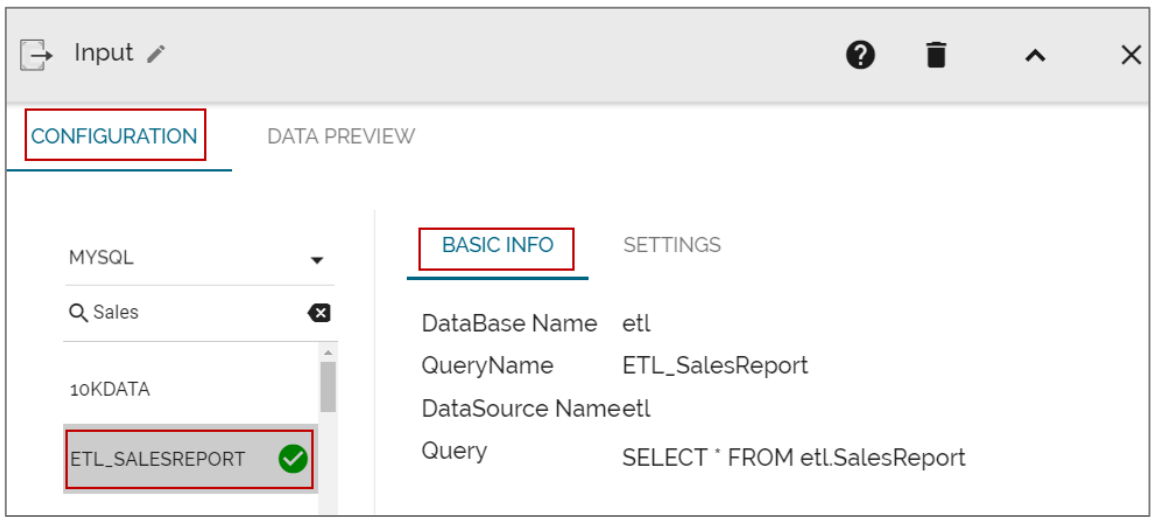
iv) Use right-click on the dragged input component
v) A new window will be displayed to configure the input data.



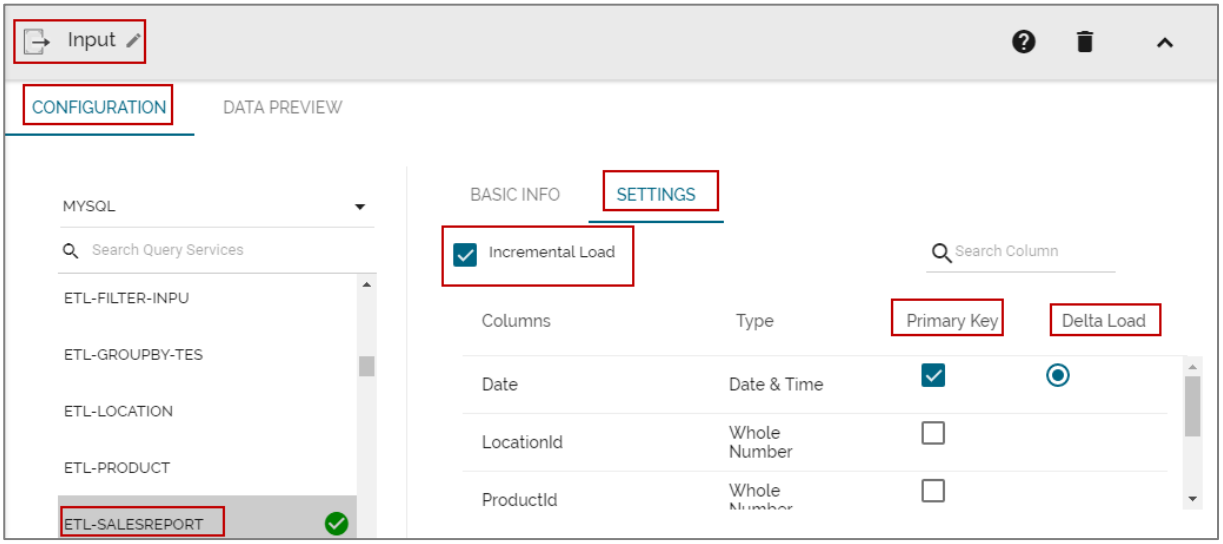
vi) Select a database type using the drop-down menu (At present only MYSQL, MSSQL, Oracle, and Google Sheet are supported).



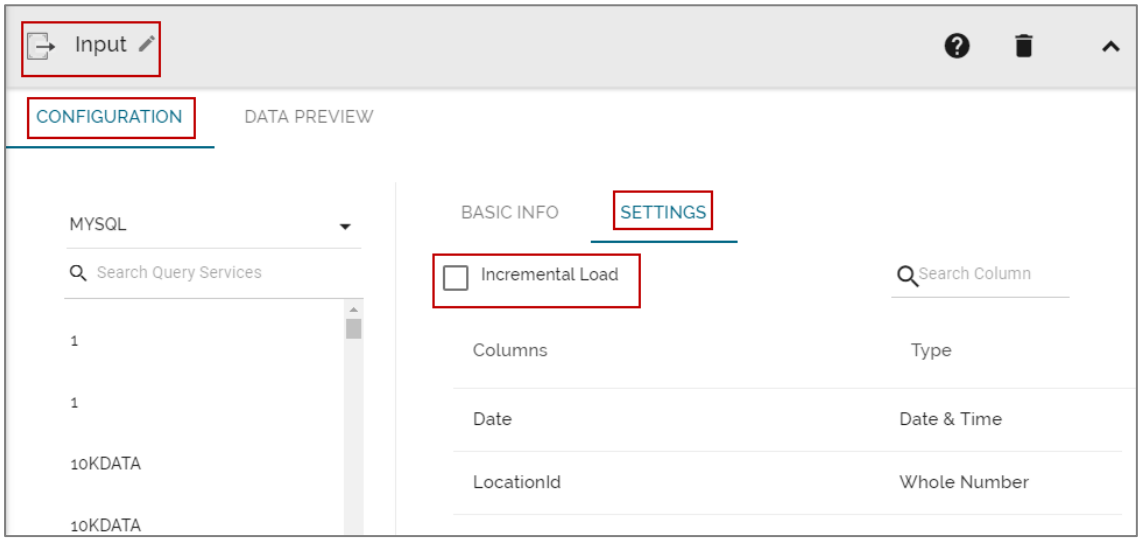
vii) Selecting a database type will redirect users to the list of data sets based on the selected database.
viii) Select a query service from the list.
ix) The basic information of the database and query service will be displayed (By Default).



x) Click the **'Settings'** tab.
xi) Users will be redirected to enable **'Increment Load'** to access the recently updated data.
xii) By enabling the **'Increment Load**,' Users need to configure the following options:
   a. **'Primary Key'**- Select a primary key of the data source.
   b. **'Delta Load'**-Select a column of type timestamp or date or long which is updated whenever a new row is inserted or updated in the data source. This column will be used to perform the **'Incremented Load'**



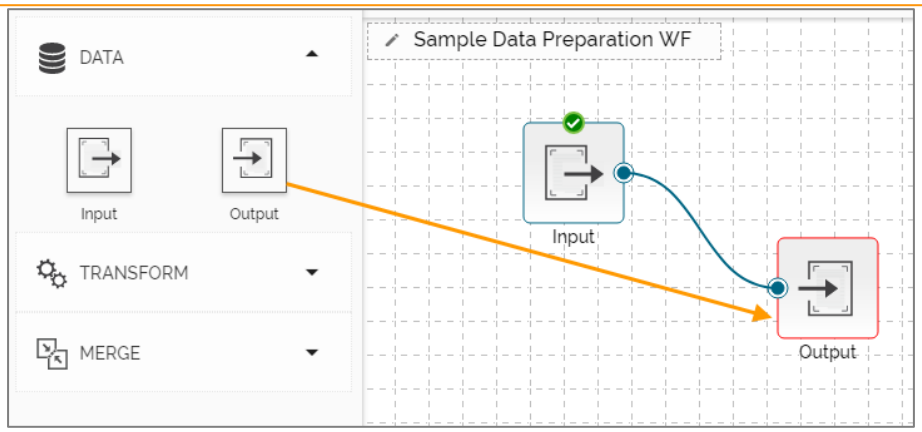
Note: Users can choose not to enable the increment load. In this case, the following details will be displayed, and the full data will be extracted.



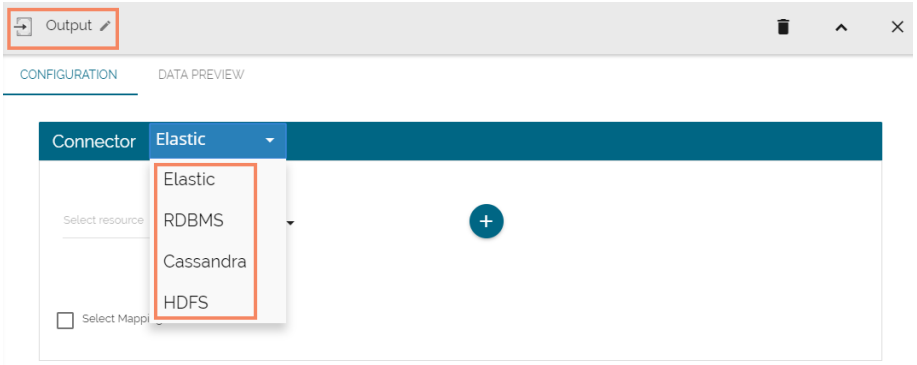
4.3. Loading Data

Users can load the extracted data into an elastic for visualization via the output component.

i) Drag and drop the **'Output'** component on the Workflow editor.
ii) Connect it with the configured **'Input'** component.

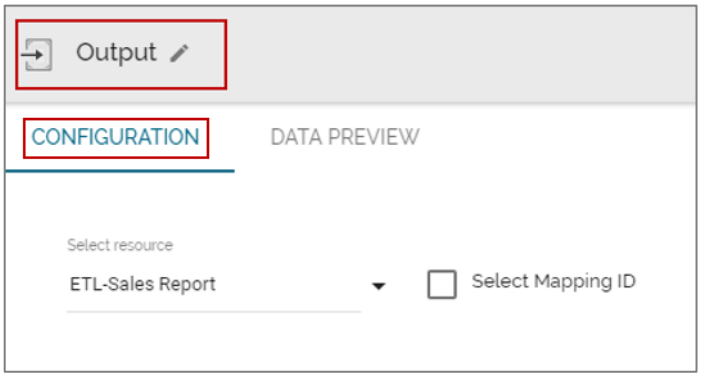


iii) Click on the '**Output**' component to display the 'CONFIGURATION' option.
iv) Users will get the following options:
   a. Elastic
   b. RDBMS
   c. Cassandra
   d. HDFS
v) Select an option and configure it

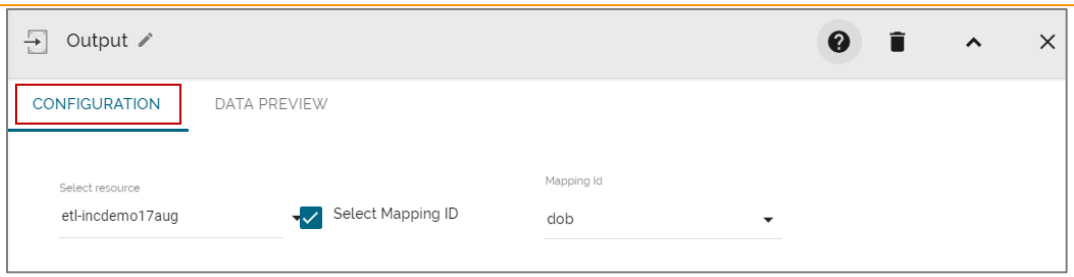


a. Configuring Elastic

i. Select a resource using the drop-down menu (for the Elastic writer)
ii. Enable '**Select Mapping ID**' option-By enabling this choice users will be redirected to select a mapping id from the '**Mapping id**' drop-down menu.

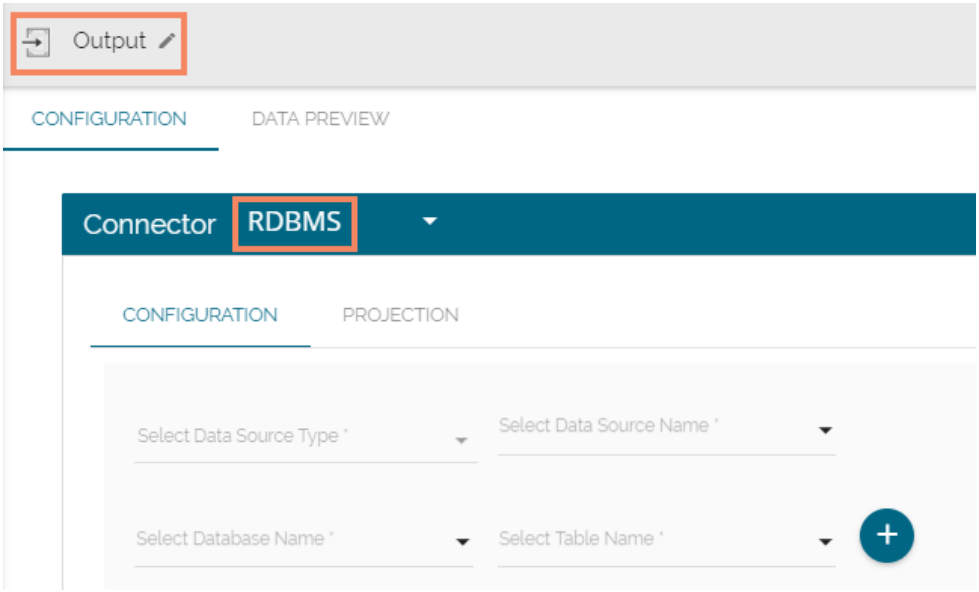


Note: If the '**Select Mapping Id**' option is enabled, users will be asked to configure the mapping id using the drop-down menu:



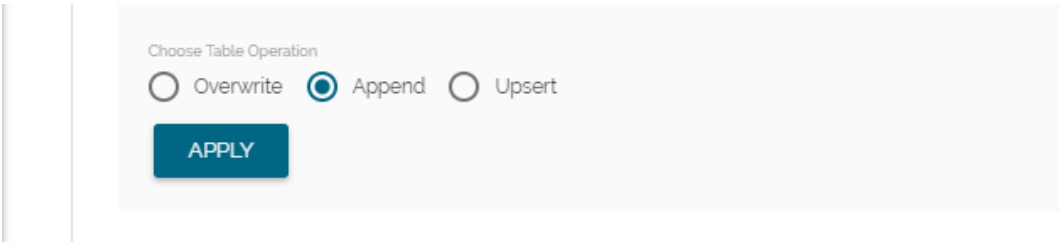
b. Configuring RDBMS

i. Select a Data Source Type from the drop-down menu
ii. Select a Data Source Name from the drop-down menu
iii. Select a Database Name from the drop-down menu
iv. Select a Table Name from the drop-down menu
v. Select '**ADD**' option to Create a New Table



vi. Choose Table Operation
1. Overwrite: Using this function, the existing records will be overwritten in
2. Append: Using this function, the records get added at the end of the elements
3. Upsert: Using this function only the new records will be added to the file

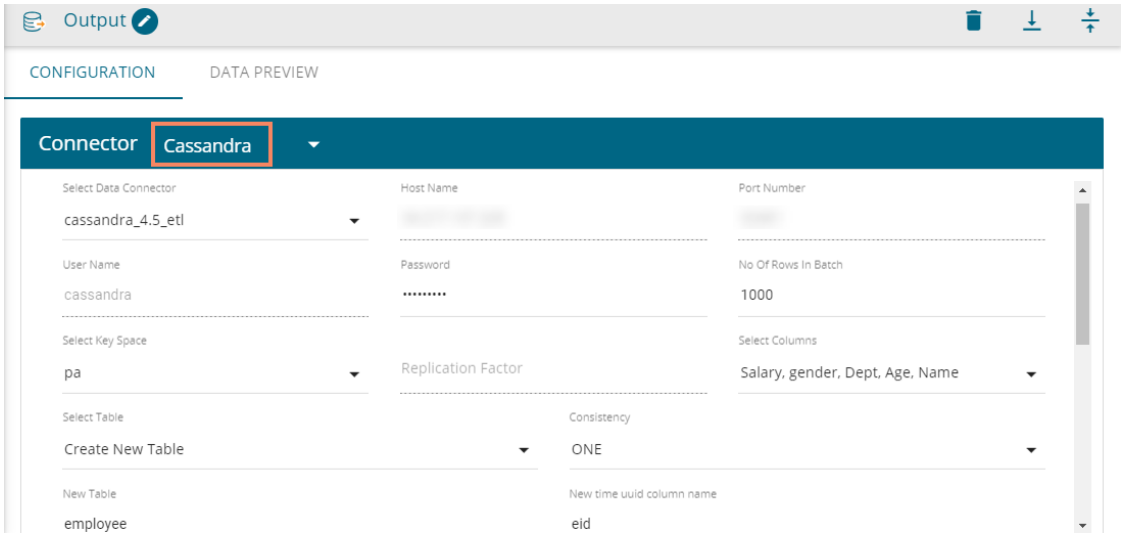
vii. Click '**APPLY**'



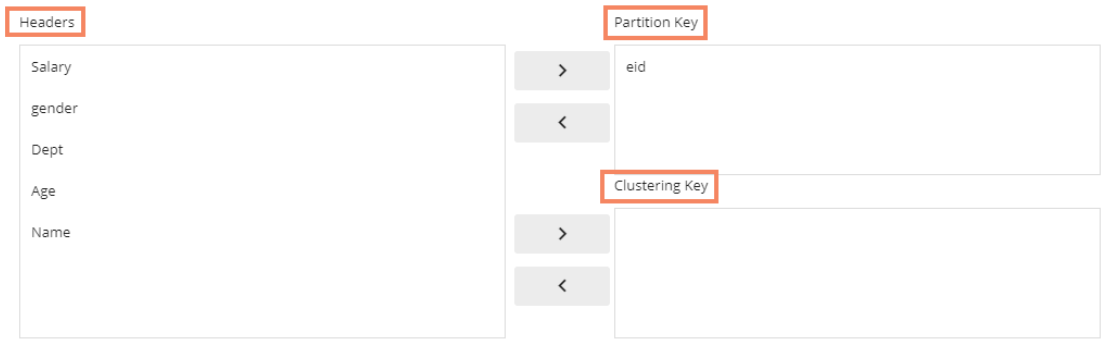
c. Configuring Cassandra

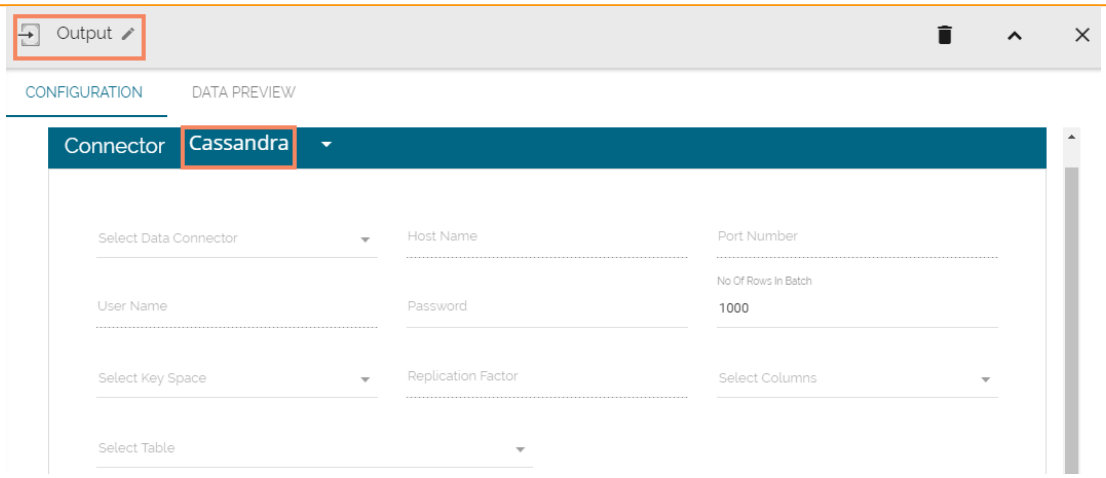
i. Select a Data Connector from the drop-down menu
ii. Enter the Host Name
iii. Enter the Port Number
iv. Enter the User Name
v. Enter the Password
vi. Enter No. of Rows in Batch

vii. Select Key Space from the drop-down menu
viii. Enter the Replication Factor
ix. Select Columns from the drop-down menu
x. Select a table from the drop-down menu
xi. Consistency: Select a Consistency option from the drop-down menu
xii. New Table: Provide a title to the newly created table using the '**New Table**' field
xiii. New time uuid column name: Provide a name for the new Time UUID Column



xiv. **Headers**: All the columns from the data set will be listed.
xv. **Partition Key**: The Partition Key determines which node stores the data. It is responsible for data distribution across the nodes.
   1. The UUID Column name gets displayed under the '**Partition Key**' window.
   2. The user can select and move any column from '**Header**' (Select Column) to '**Partition Key**' space.
   3. The sequence of the columns listed under Partition Key can be arranged by using '**Up**' or '**Down**' options.
xvi. **Clustering Key**: The Clustering Key is a storage engine process that sorts data within the partition. It determines per-partition clustering.
   1. The items listed under the Clustering Key box can be arranged by using '**Up**' or '**Down**' options.
   2. Users can select any column from '**Headers**' (Select Column) to '**Clustering Key**' space.





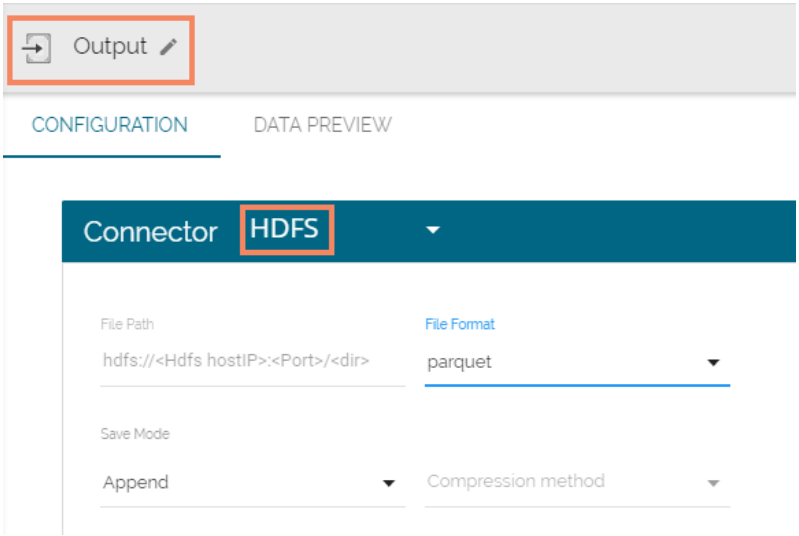
d. Configuring HDFS

i. Provide file path

ii. Select a File Format from the below given choices in the drop-down menu
   1. Parquet
   2. Json
   3. Avro
   4. CSV

iii. Select a Save Mode from the below given options in the drop-down menu
   1. Append
   2. Overwrite
   3. Error
   4. Ignore

iv. Select a Compression Method from the below given options in the drop-down menu
   1. Gzip
   2. Snappy
   3. None



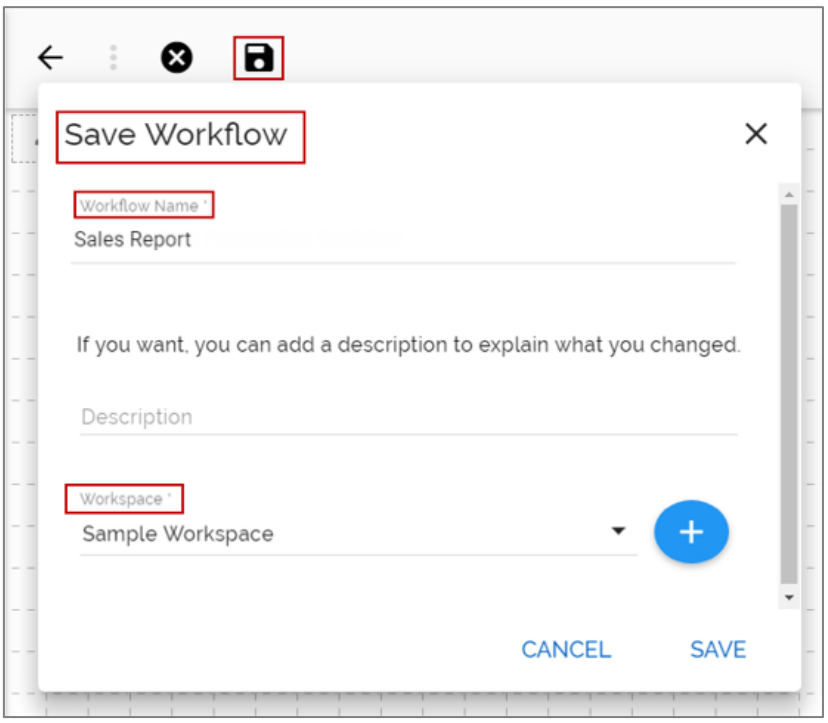
4.4. Saving a Workflow

Users are provided with two options to save a workflow.

i) Click the '**Save**' option 

ii) A new window pops-up to redirect the user to save the workflow.
   a. Enter a Workflow name
   b. Enter Description (Optional)
   c. Select or Add a Workspace

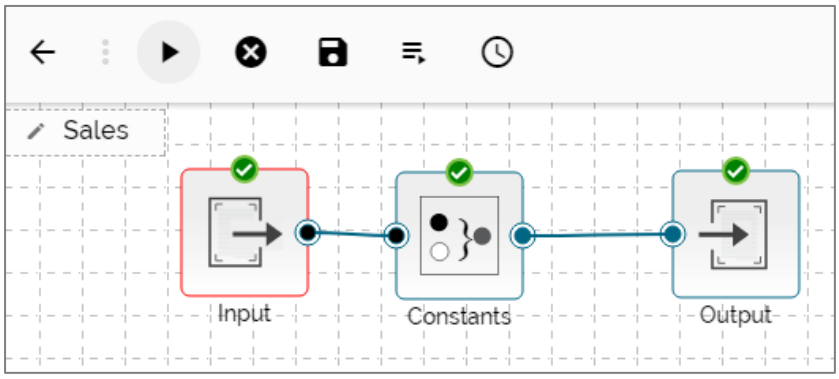
iii) Click **'Save'**



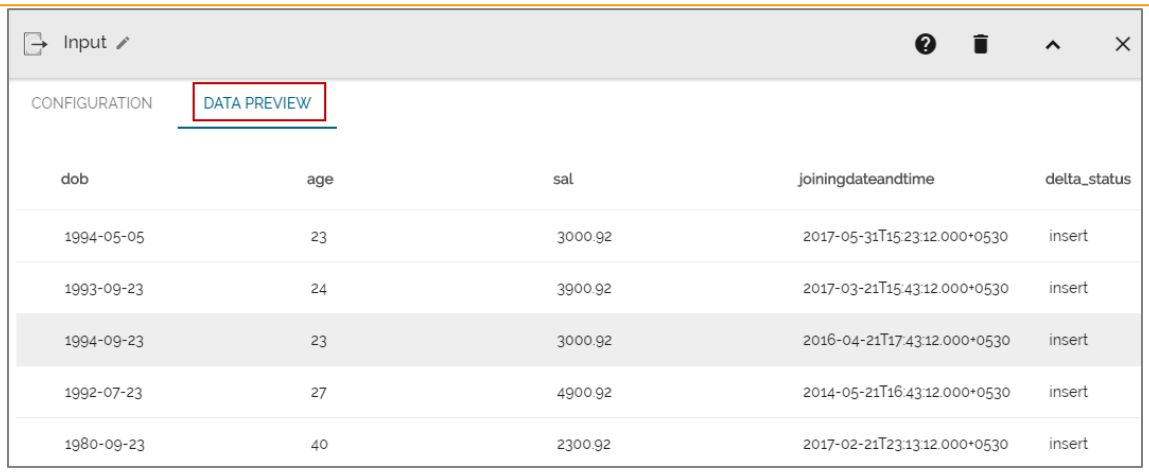
4.5. Run Preview

Users can run the created workflow without affecting their production system through **'Run Preview'** option. Users need to save the workflows to get the 'Run Preview' option.

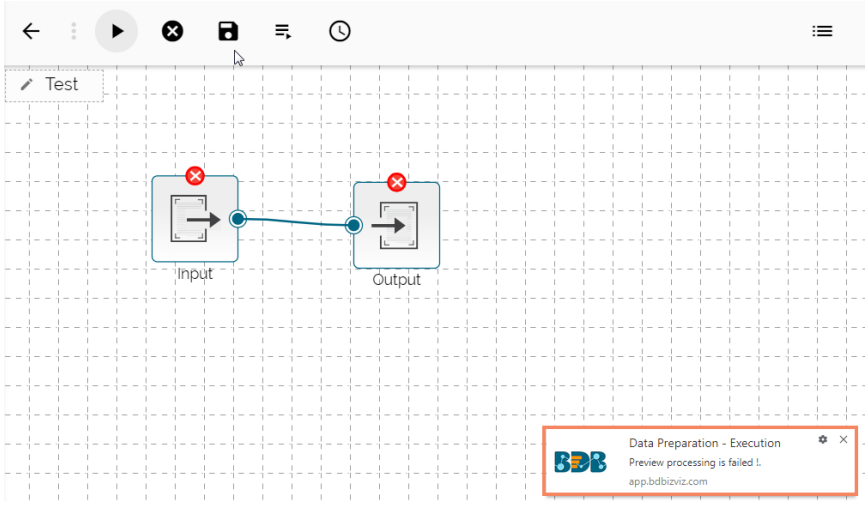
i) After saving a workflow, Users will be able to access more options on the workflow editor toolbar.
ii) Click **'Run Preview'** option ▶
iii) The ongoing execution process will be displayed through a continuous blue line.
iv) Users will get notified about the beginning and end of the execution process by pop-up messages.
v) After the execution gets completed a green tick mark will be displayed. The input data with a green checkmark is ready to preview.



vi) Open **'Data Preview'** by clicking the input component to view the preview of the extracted data.



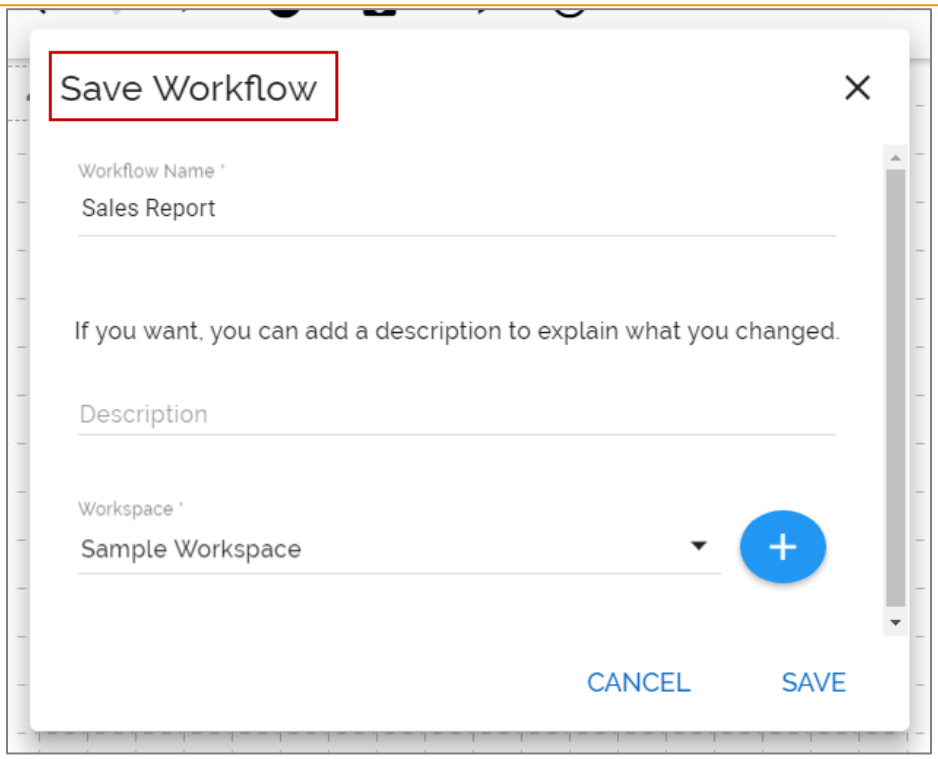
Note: users will get notifications on the screen for success or failure of the preview processing.



4.6. Save and Execute

By using the '**Save and Execute**' option  users can save and write a workflow in the metadata to create a datastore out of it.

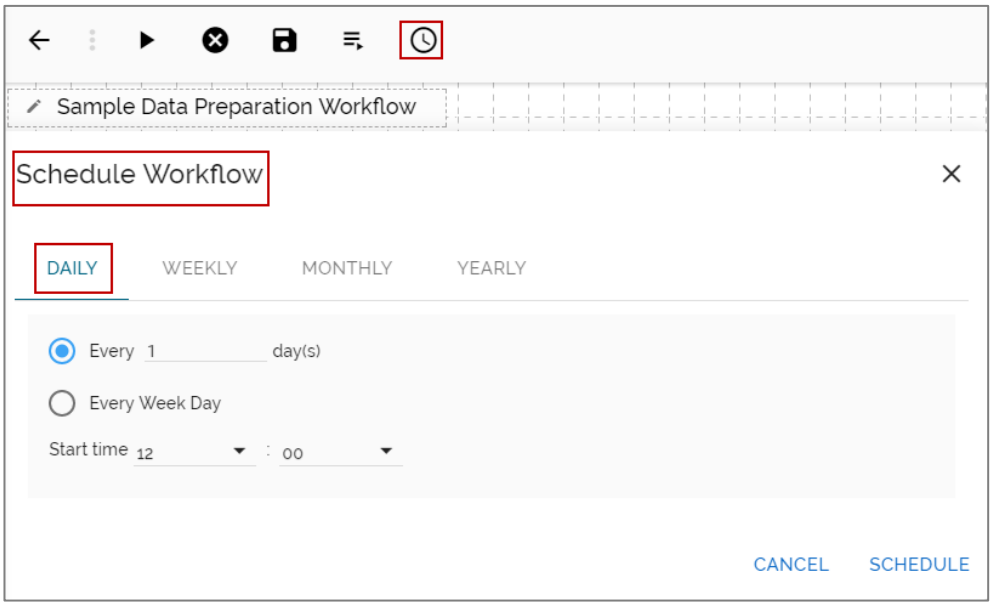
i) Click the '**Save**' option.
ii) A new window pops-up is redirecting the user to save the workflow.
   a. Enter a Workflow name
   b. Enter Description (Optional)
   c. Select or Add a Workspace
iii) Click '**Save.**'



4.7. Schedule a Workflow

Users can schedule a created workflow for data refresh.

i) Create a workflow
ii) Save and run the workflow
iii) Click the '**Scheduler**' icon
iv) Click a range of time
v) Fill in the required information for the selected time range. E.g., The below-given image displays scheduler configuration details for the '**DAILY**' option.
vi) Click '**SCHEDULE**'.

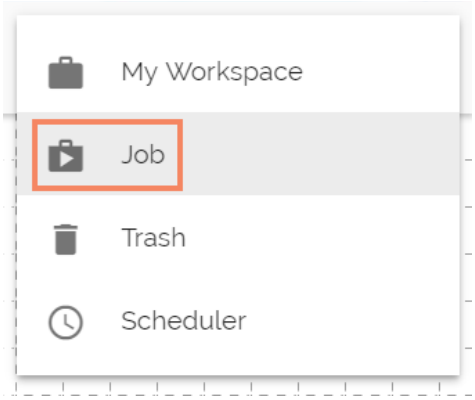


vii) The selected workflow will be scheduled.

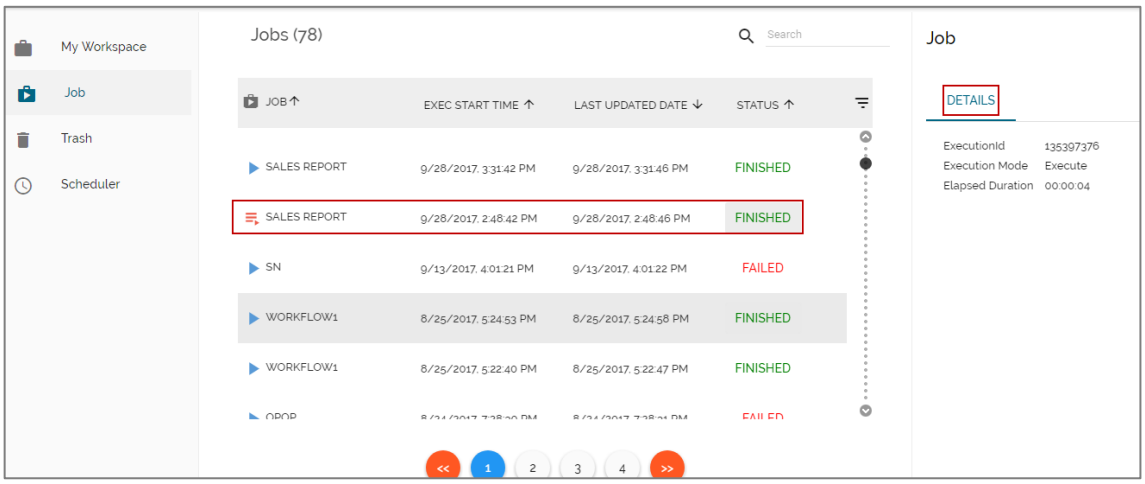
4.8. Job

Users can see the job status for the saved workflows.

i) Navigate to the Data Preparation landing page
ii) Click icon from the workflow editor
iii) Select the '**Job**' option from the menu list



iv) Users will be displayed the job details in a table



Note: The execution details will be displayed on the right-hand side of the '**Job**' page. Users need to click on the 'STATUS' of a job using the list of the jobs.

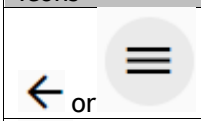
4.9. Trash

The '**Trash**' folder is provided to store all the deleted workflows and workspaces. Users can restore the deleted workflows and workspaces using this folder.

i) Click on the '**Trash**' option.
ii) Users will be redirected to see all the deleted files and folders under the trash folder.
iii) Click '**Restore**' to restore the selected workflow/workspace.
iv) Click '**Delete**' to permanently delete the selected workflow/workspace.

Note:

a. Users can check out all the essential features of the Data Preparation module on a relevant input dataset.
b. Other options provided on the workflow editor are as described below:

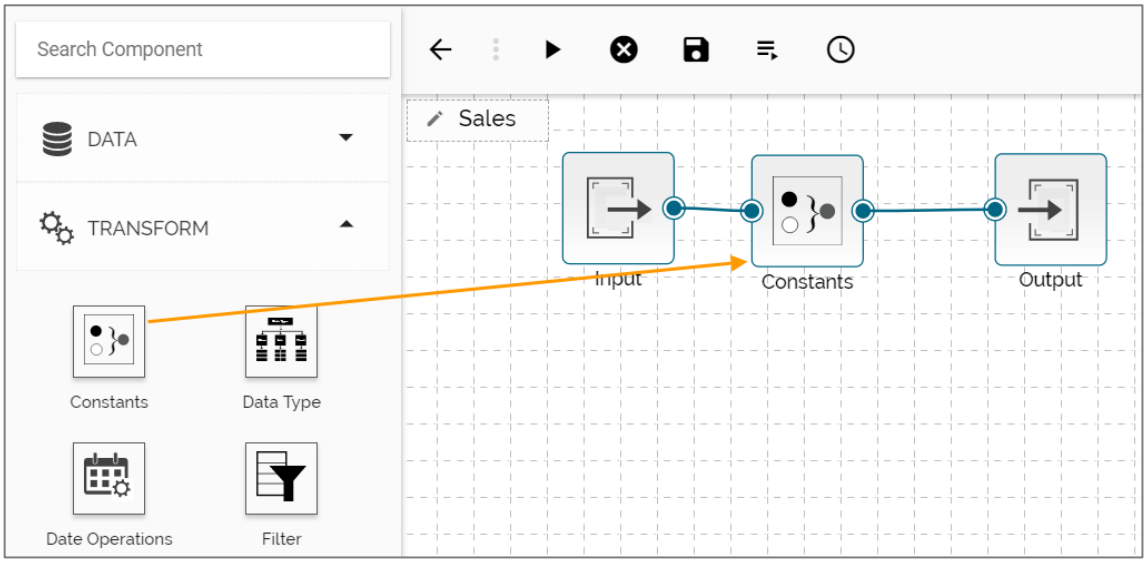
Icons	Name	Description
	Hide and Show Components	Hides or shows the components on the left-hand side.
	Clear Workflow	Clears the current workflow from the workflow editor.
	Save	Saves a workflow
	Navigator	Redirects Users to the following hyperlinks: 1. Workspace 2. Job 3. Trash 4. Scheduler

5. Transform

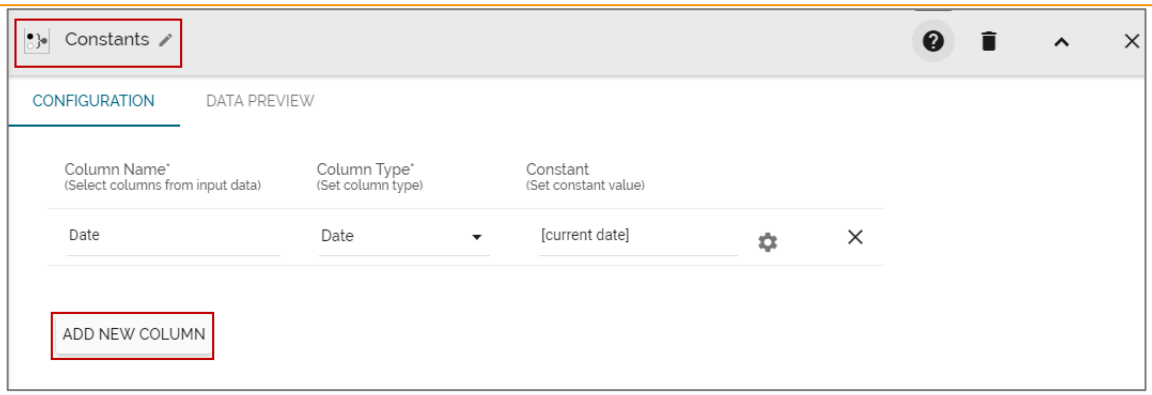
5.1. Constants

Users can give a corresponding valid constant value for each type of column.

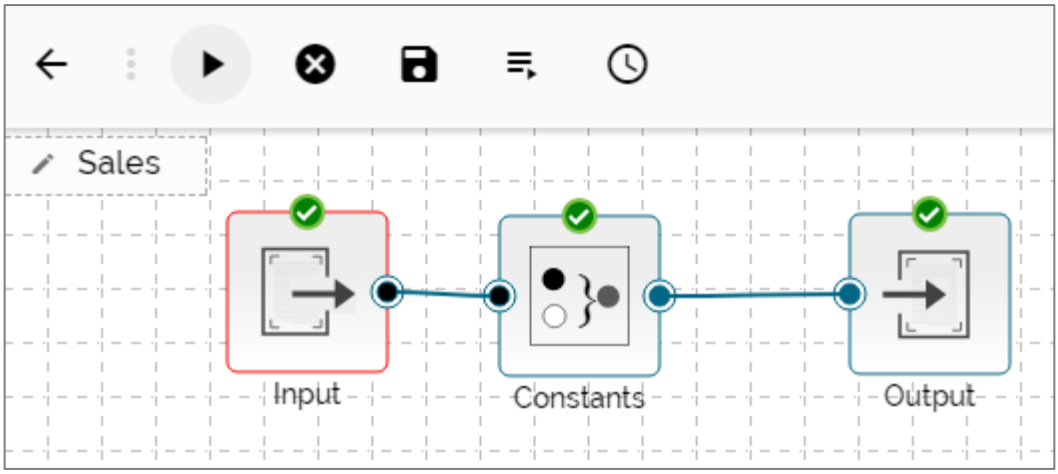
i) Navigate to the Workflow editor.
ii) Connect the '**Constants**' component to the configured input dataset.



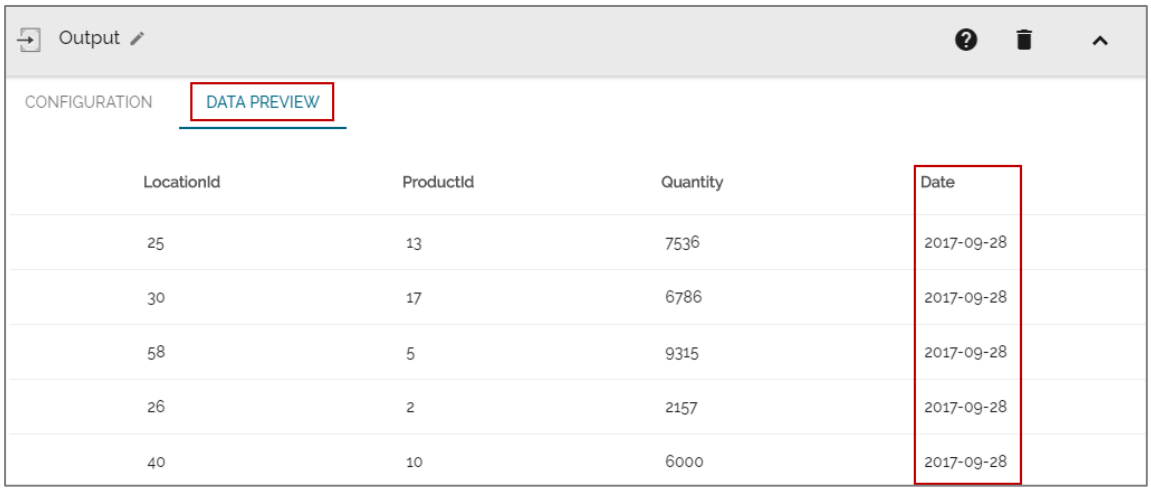
iii) Configure the required details for the '**Constants**' component:
a. Column Name: Select columns from input data
b. Column Type: Set column type using the drop-down menu
c. Constant: Set a constant value
d. Remove: Click the '**Remove**' icon to remove the added constant information.



iv) Save the workflow.
v) Run/Execute the workflow.



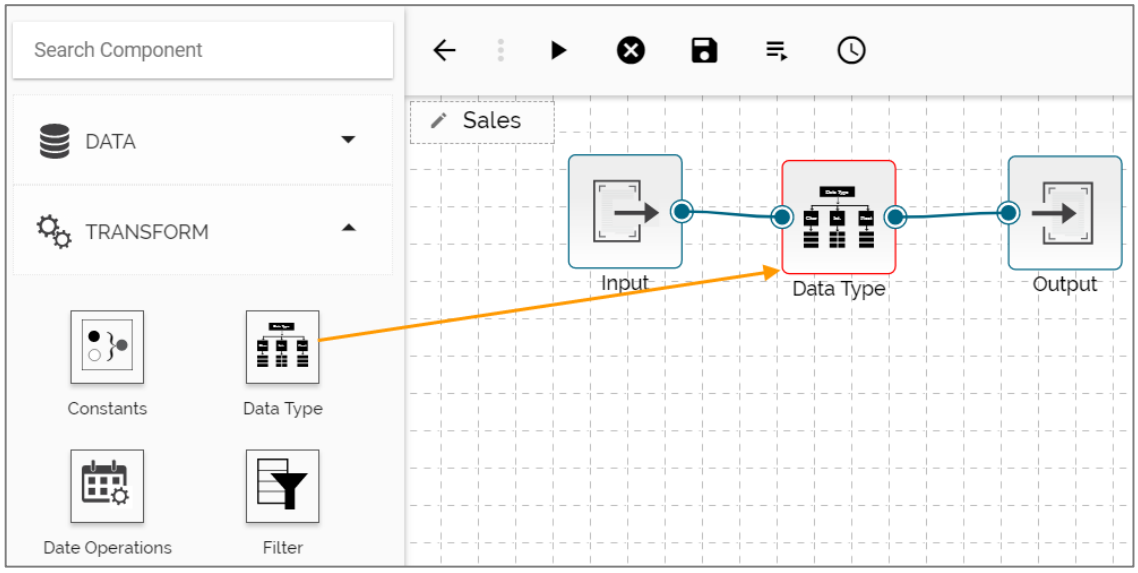
vi) The set constant value will be applied to the selected column in the output dataset.



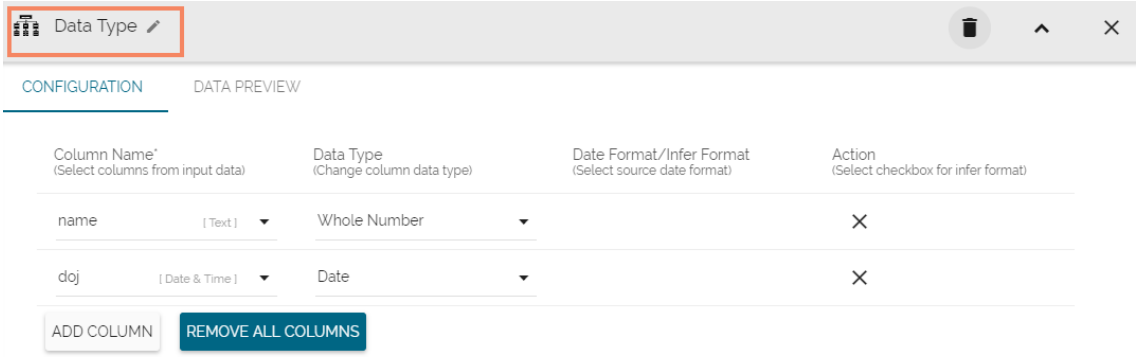
5.2. Data Type

Users can change the data type of the selected columns by using the '**Date Type.**'

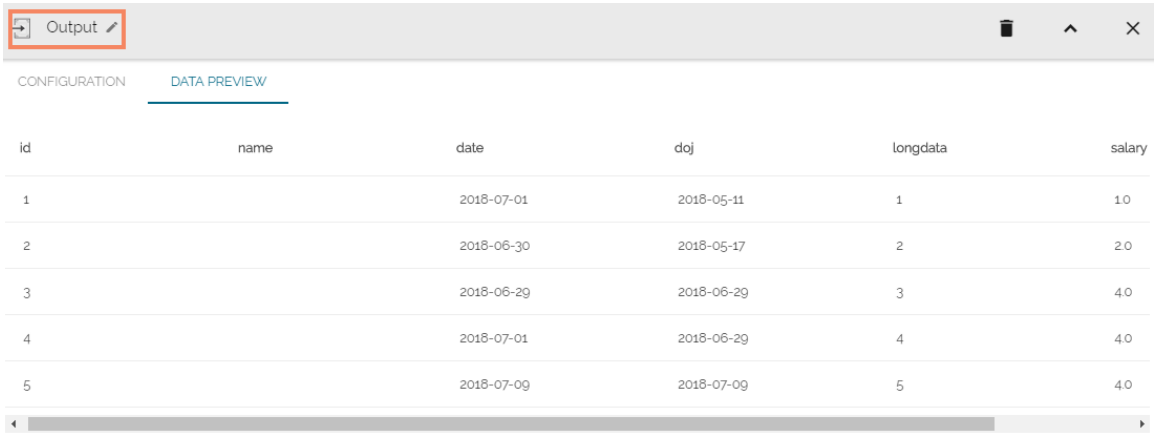
i) Navigate to the Workflow editor.
ii) Connect the '**Data Type**' component to the configured input dataset and output component.



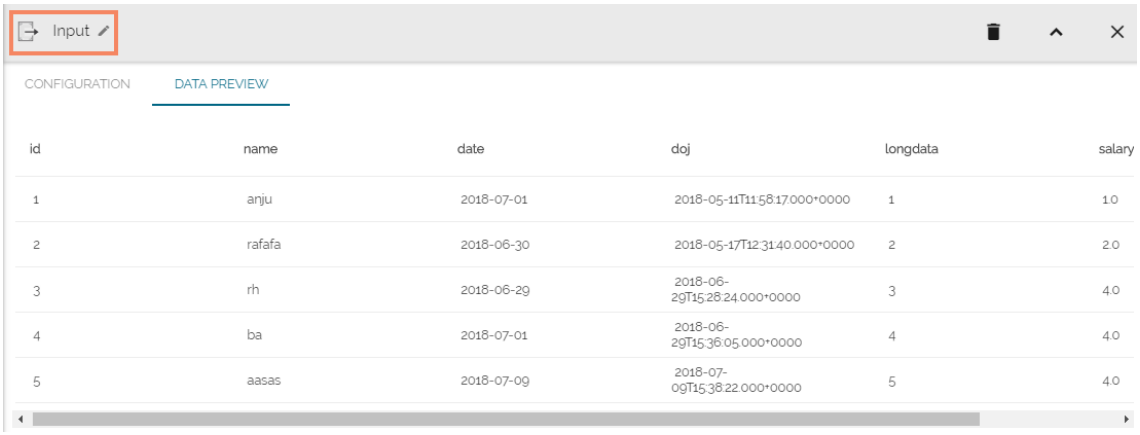
iii) Select the columns and change the column data type using the drop-down menu.
a. Column Name: Select columns from input data
b. Data Type: Change column data type
c. Date Format: Select source date format
E.g. In this case, the column data type has been changed from '**Date & Time**' to '**Date.**'



iv) Save the workflow.
v) Run/Execute the workflow.
vi) Click the 'DATA PREVIEW' tab for the Output component to see the transform result

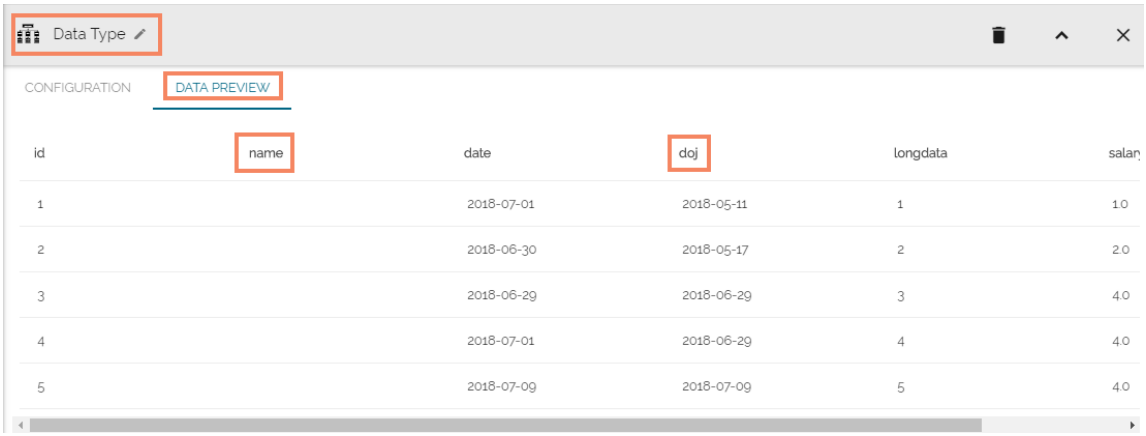


vii) Users can compare the data previews of the Input and Data Type modules (E.g., the selected input, in this case, contains the following column types)



Note:

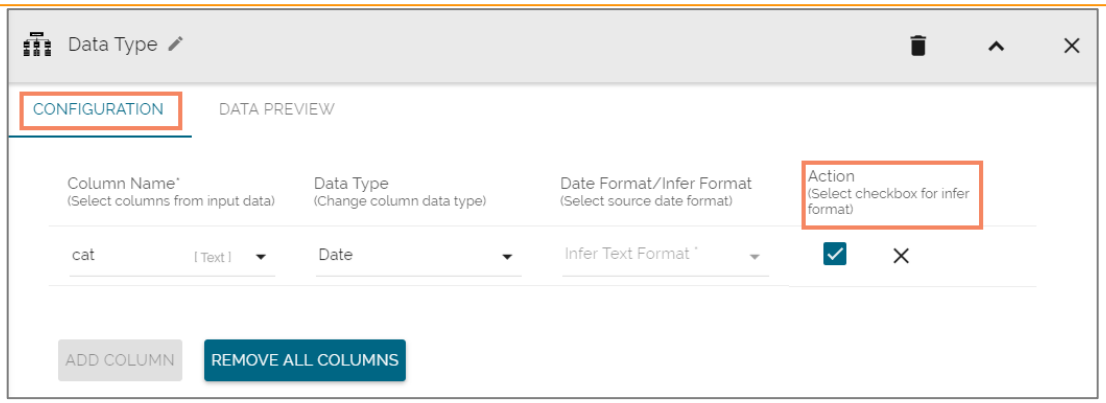
a. Users can get the same Data Preview as Output dataset while opening the 'DATA PREVIEW' tab from any selected transform component. E.g., The 'DATA PREVIEW' tab for the 'DATA TYPE' Transform component is as displayed below.



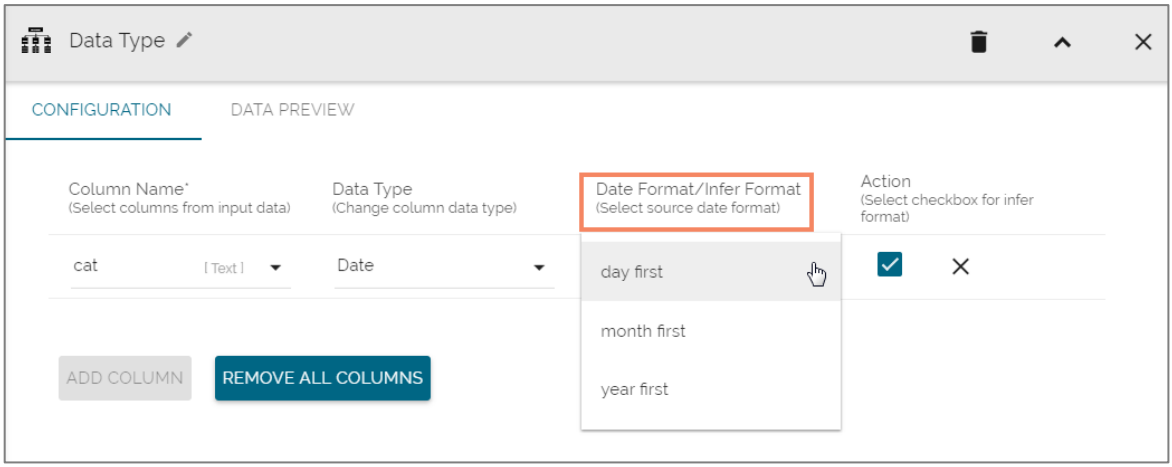
5.2.1. Inferring Date & Date Time Formats

The Infer Date/Data Time functionality is provided for users to include various Date/Date Time formats which are not provided by the application.

i) Users need to create a workflow using the '**Data Type**' transform and select a '**Text**' type column from the '**Column Name.**'
ii) Select '**Date**' as Data Type
iii) Enable the inferring using the '**Action**' option



iv) Select the '**Date Format/Infer Format**' from the given choices



v) All the dates as per the selected infer date format from the source dataset will be listed in the output

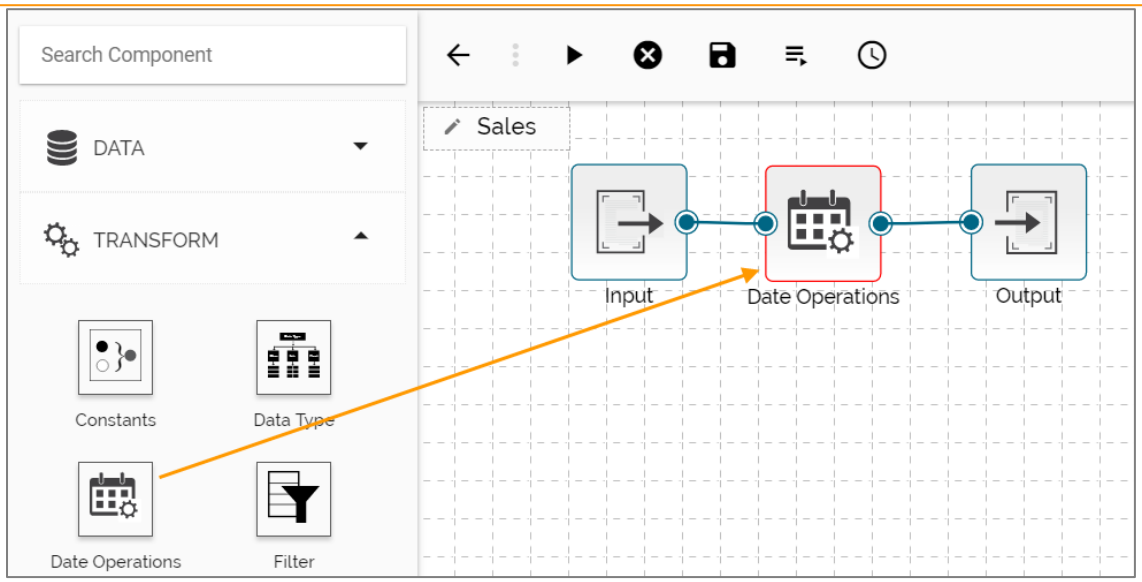
Note:

a. The functionality only works for the '**Text**' type of column.
b. If the source data format does not befit in the selected infer format, then those entries will not be listed in the output.

5.3. Date Operations

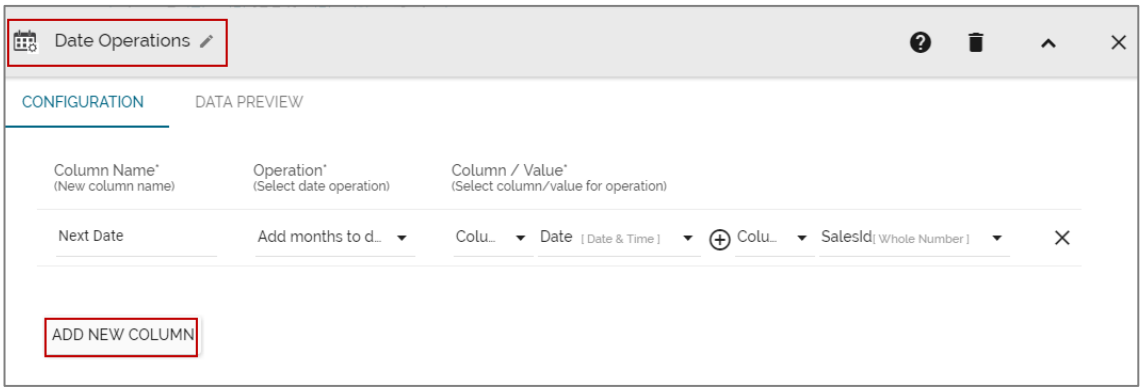
Users can perform various operations of dates addition/subtraction with integers or other dates. It also allows extraction of parts of dates like day-part, month part, etc.

i) Navigate to the Workflow editor.
ii) Connect the '**Date Operations**' component to the configured input dataset and output component.

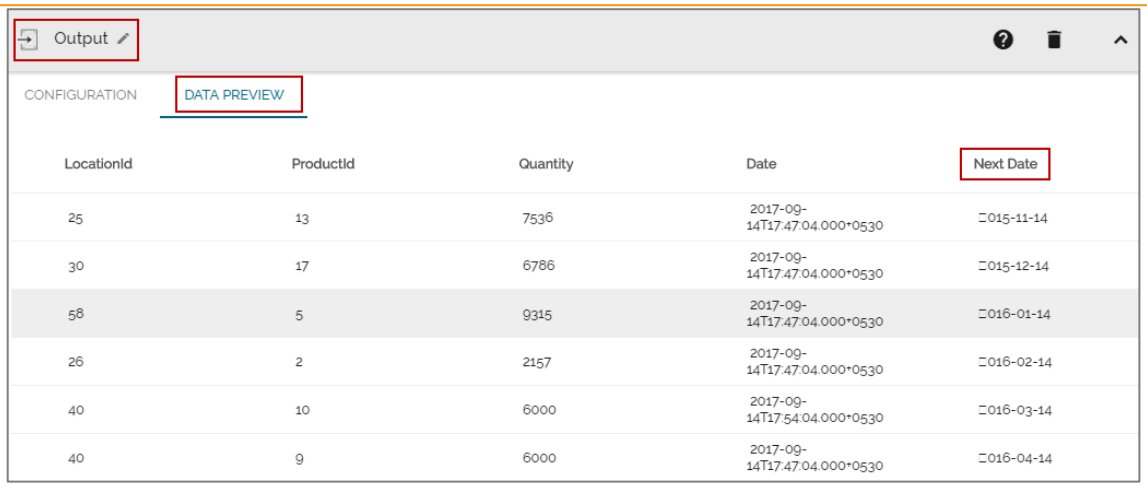


iii) Configure the **'Date Operations'** component as described below:

a. Column Name: Enter the New Column Name
b. Operations: Select one operation using the drop-down menu.
c. Column/Value: Select a column or value for operations.
   i. By selecting '**column**' option, the column drop-down menu will be displayed.
   ii. By selecting the '**value**' option, users will be redirected to enter a value.



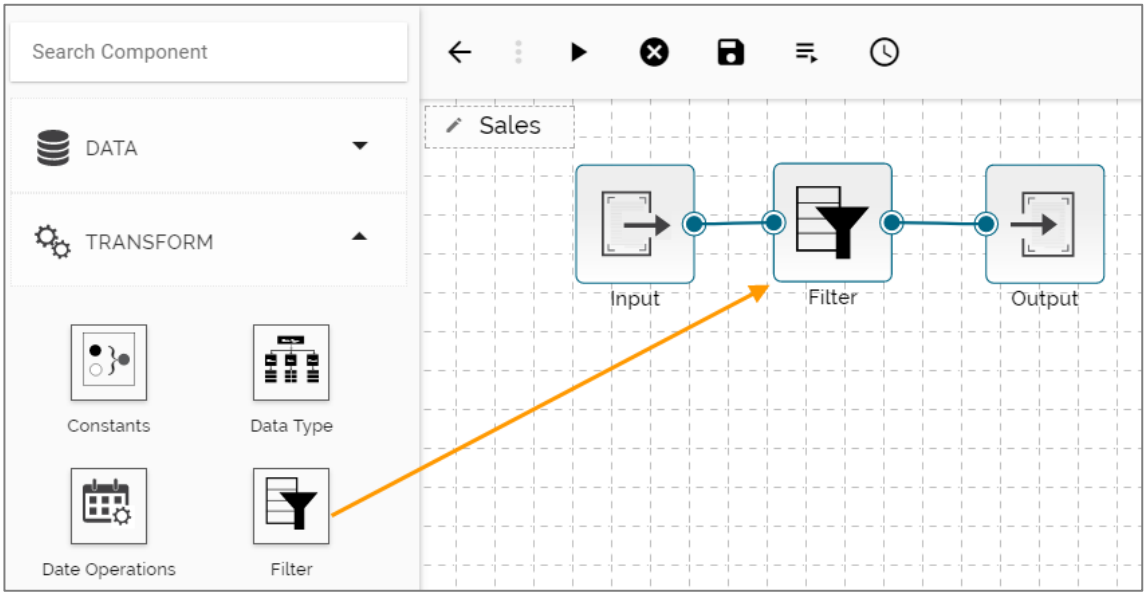
iv) Save the workflow.
v) Run/Execute the workflow.
vi) The new column, '**Next Date**' will be added in the output dataset. Users can view it in the output data preview.



5.4. Filter

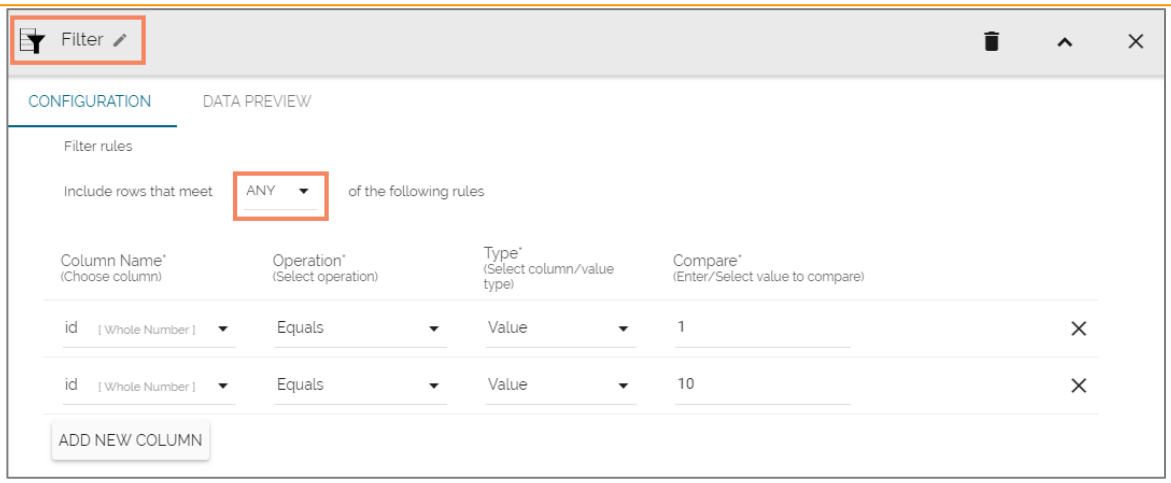
Users can filter the input dataset by specifying conditional expressions using the '**Filter**' transform. Multiple filter conditions can be imposed in the same transform. The following table lists the map of data types and permissible filter conditions.

i) Navigate to the Workflow editor.
ii) Connect the '**Filter**' component to the configured input dataset and output component.

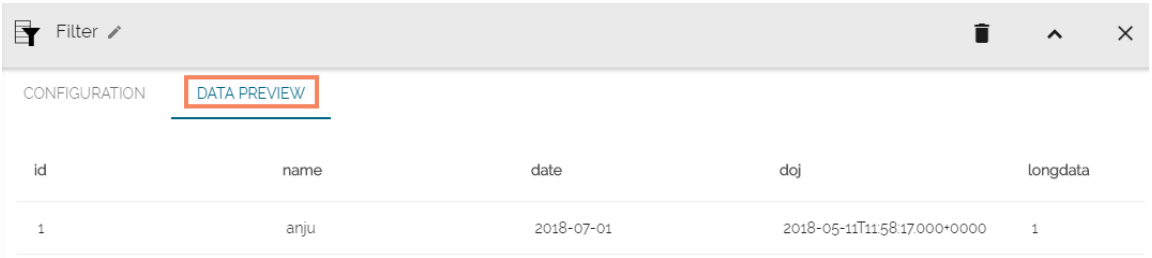


iii) Configure the '**Filter**' Component as described below:

a. Select a filter rule from the drop-down
   i. ALL: By selecting this option filter will be applied only if all the added conditions are true
   ii. ANY: By selecting this option filter will be applied even if any one condition is true
b. Column Name: Choose a column from the drop-down menu
c. Operation: Select an operation from the drop-down menu
d. Type: Select one option out of '**Column**' or '**Value.**'
e. Compare: Enter a value/Select a column from the list to compare with



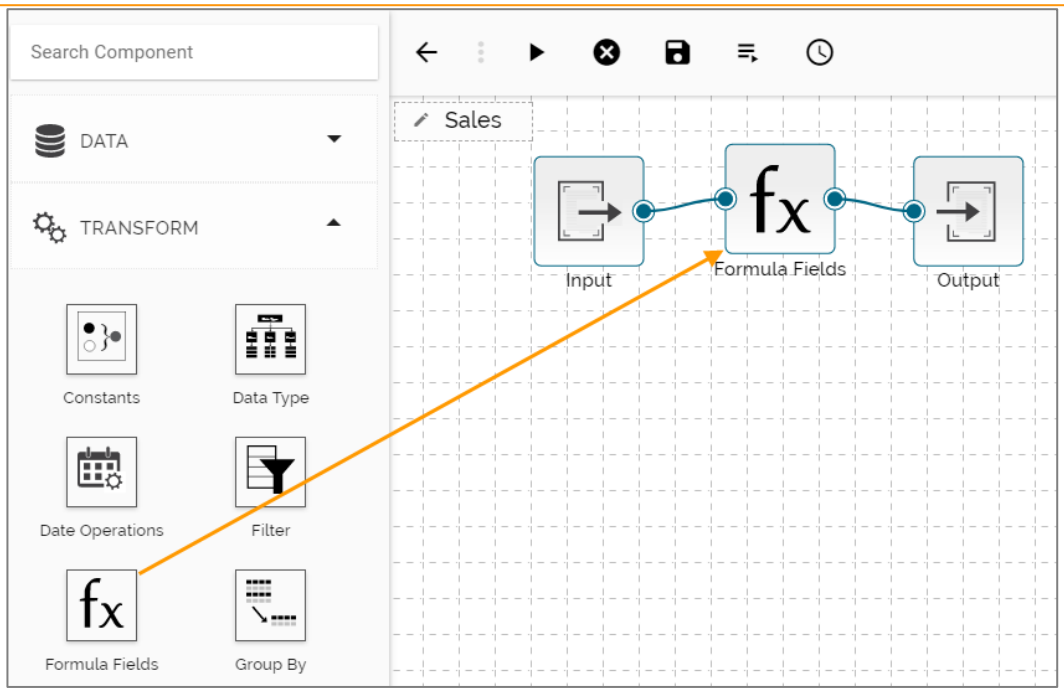
iv) Save the workflow
v) Run the workflow
vi) The input data will be filtered as per the applied conditions



5.5. Formula Fields

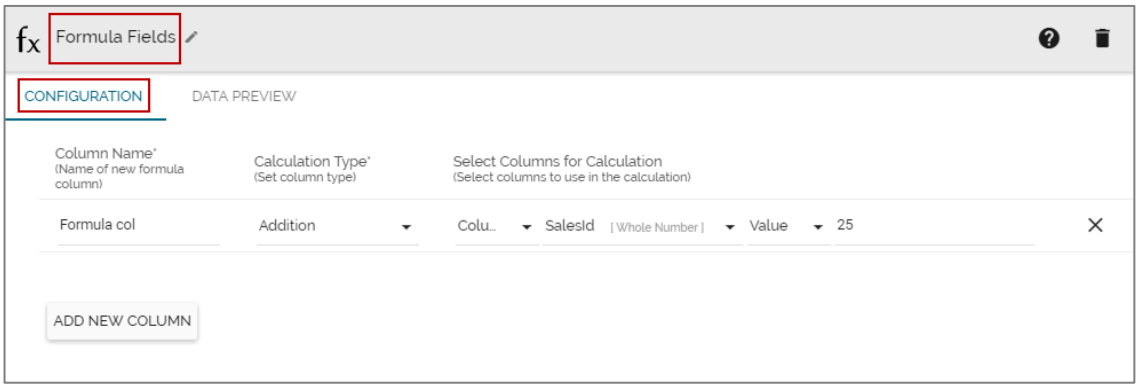
Users can perform most common arithmetic operations (add, subtract, multiply and divide) on constants and columns.

i) Navigate to the Workflow editor.
ii) Connect the '**Formula Fields**' to the configured input dataset and output component.

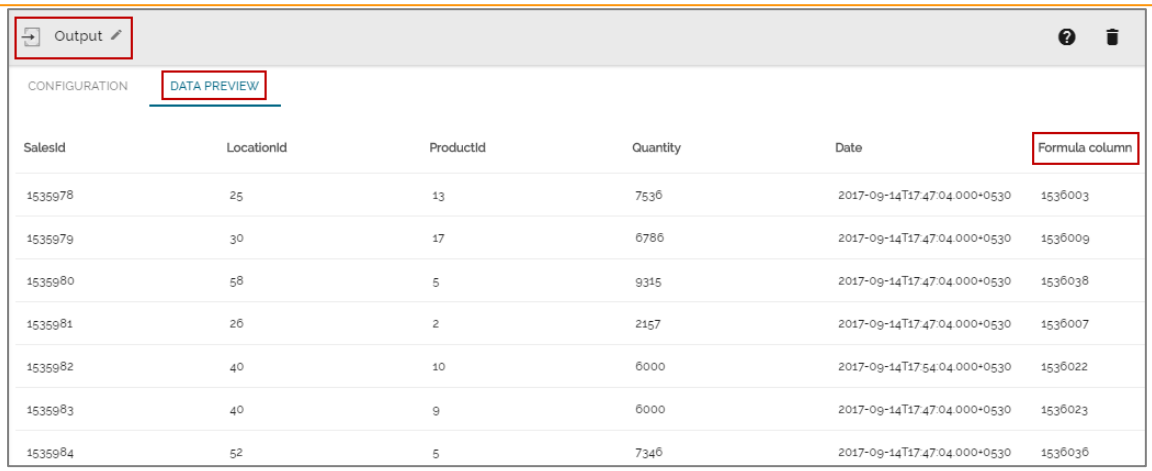


iii) Configure the '**Formula**' component as described below:

a. Column Name: Enter a name for the formula column
b. Calculation Type: Select a calculation type using the drop-down menu
c. Select Columns for Calculation: Select columns to be used in the calculation. Users can choose either a column or enter a value to complete the calculation process.
E.g. In this case, the value option is chosen.



iv) Save the workflow.
v) Run the workflow.
vi) The calculated column will be added in the output dataset.

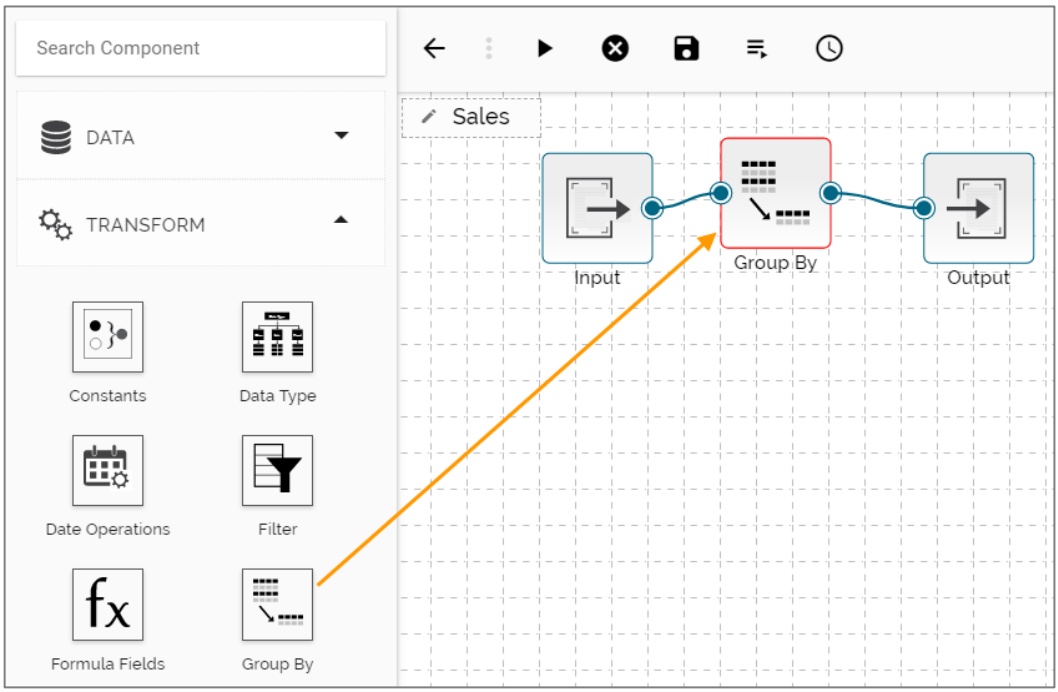


SalesId	LocationId	ProductId	Quantity	Date	Formula column
1535978	25	13	7536	2017-09-14T17:47:04.000+0530	1536003
1535979	30	17	6786	2017-09-14T17:47:04.000+0530	1536009
1535980	58	5	9315	2017-09-14T17:47:04.000+0530	1536038
1535981	26	2	2157	2017-09-14T17:47:04.000+0530	1536007
1535982	40	10	6000	2017-09-14T17:54:04.000+0530	1536022
1535983	40	9	6000	2017-09-14T17:47:04.000+0530	1536023
1535984	52	5	7346	2017-09-14T17:47:04.000+0530	1536036

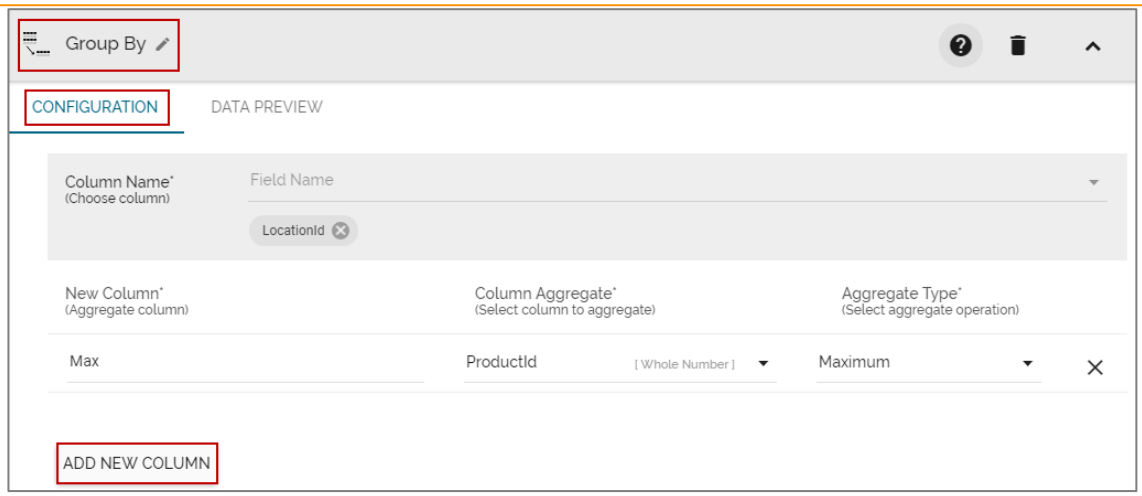
5.6. Group By

The '**Group By**' feature allows multiple aggregations on the same or different columns. Users can obtain numerous aggregations in the same transform. The aggregated values are added to a new column.

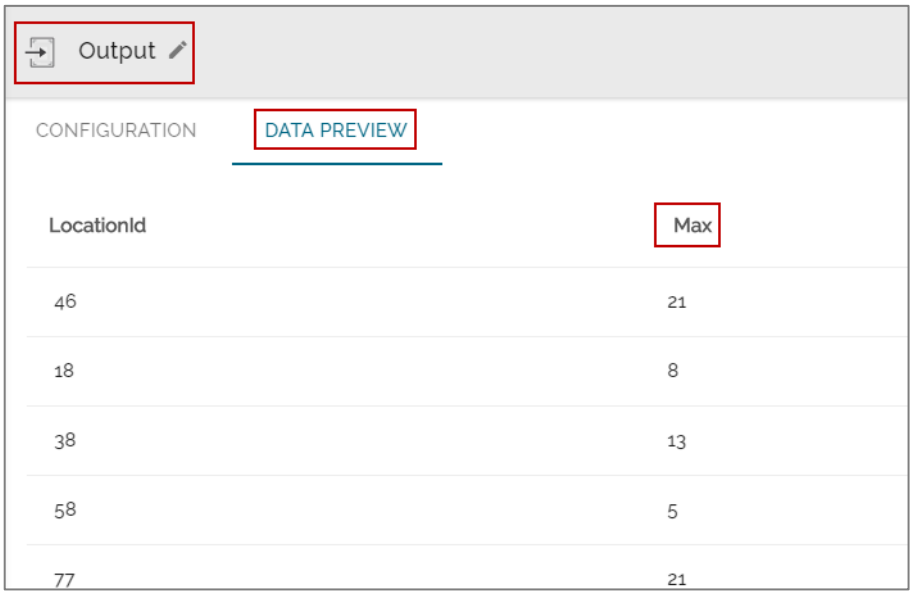
i) Navigate to the Workflow editor.
ii) Connect the '**Group By**' component to the configured input dataset and output component.



iii) Configure the '**Group By**' component as described below:
a. Column Name: Select a column from the drop-down menu
b. New Column: Enter a title for the aggregate column
c. Column Aggregate: Select a column from the drop-down menu to apply aggregation
d. Aggregate Type: Select an aggregation operation from the drop-down menu



iv) Save the workflow
v) Run the workflow
vi) The aggregated column will be displayed in the output data preview



Note: The supported data types and aggregate operations are displayed in the following table:

Data Type	Aggregate
Text	Count Count Including NULLs Count Distinct Values First Non-Null Value Last Non-Null Value First Value Last Value Combine Strings Separated by Comma
Date Date Time	Minimum Maximum Count Count Including Nulls Count Distinct Values First Non-Null Value Last Non-Null Value First Value

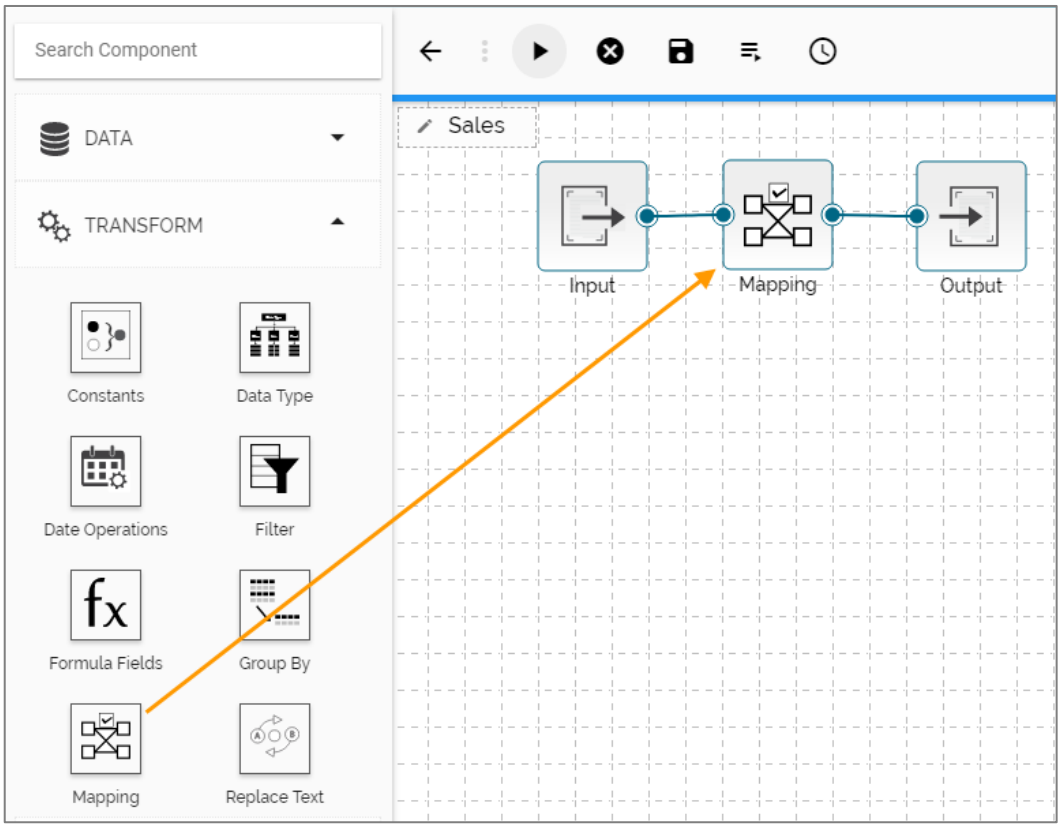
	Last Value
Whole Number Decimal Decimal (Fixed)	Sum Average Minimum Maximum Standard Deviation Count Count Including NULLs

5.7. Mapping

Users should be able to select, remove or rename columns in the input dataset to fit the structure of the sink.

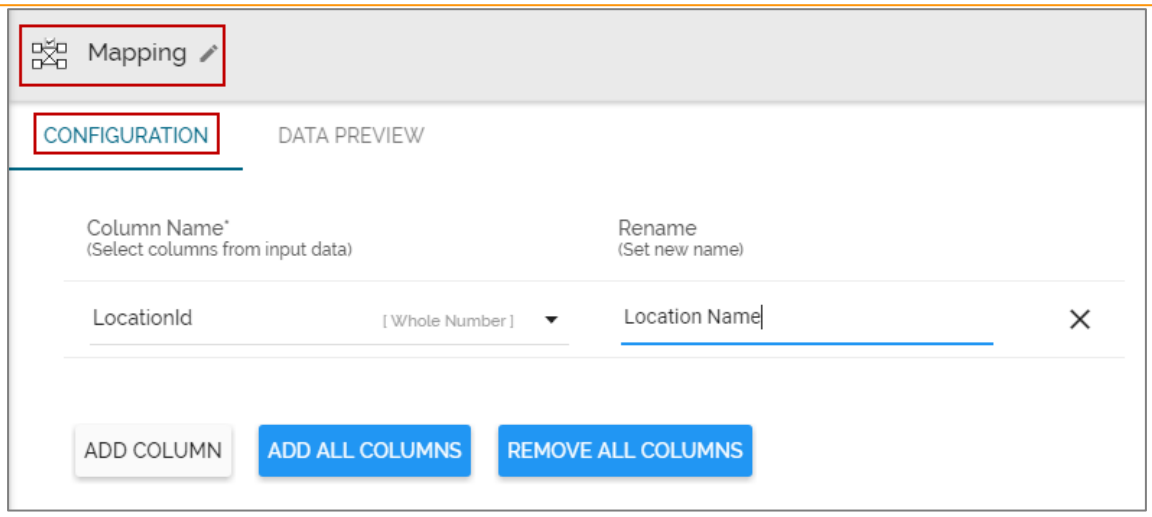
i) Navigate to the Workflow editor

ii) Connect the '**Mapping**' component to the configured input dataset and output component

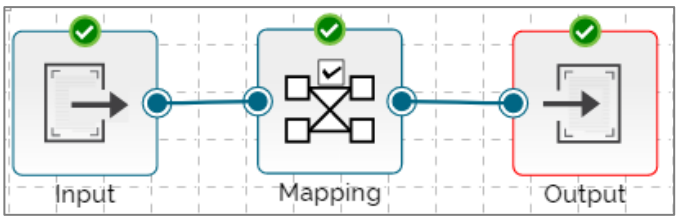


iii) Configure the '**Mapping**' component:

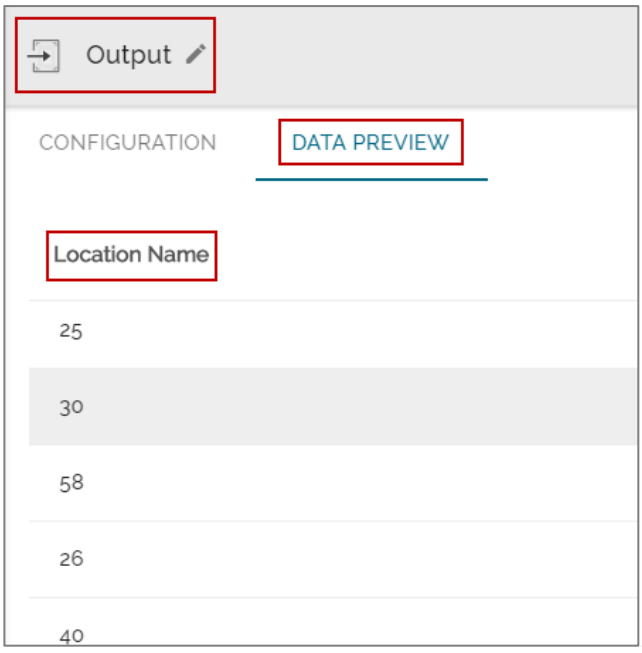
a. Column Name: Select a Column from the input data using the drop-down menu
b. Rename: Rename the selected column of the input data
c. ADD Column: Click this option to add one more column from the input dataset
d. ADD ALL COLUMNS: Click this option to map all the columns from the input dataset
e. REMOVE ALL COLUMNS: Click this option to remove all the added columns for mapping



iv) Save the workflow
v) Run the workflow
vi) The aggregated column will be displayed in the output data preview



vii) The aggregated column will be displayed in the output data preview

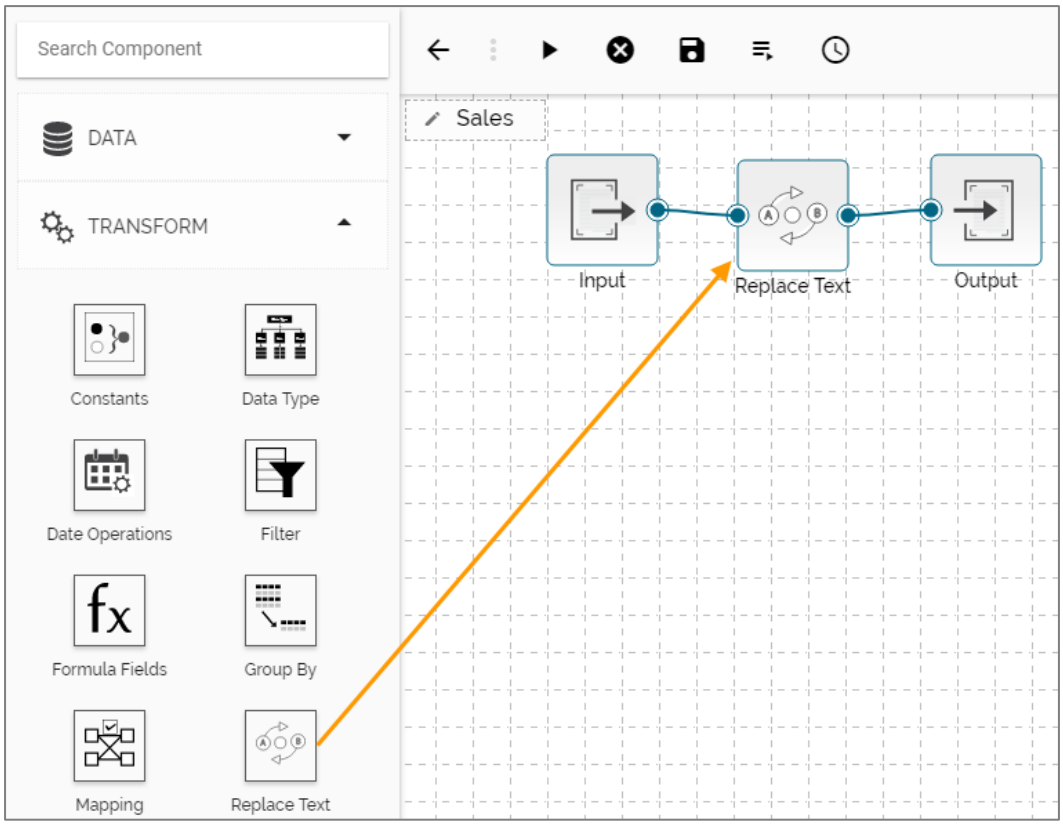


s

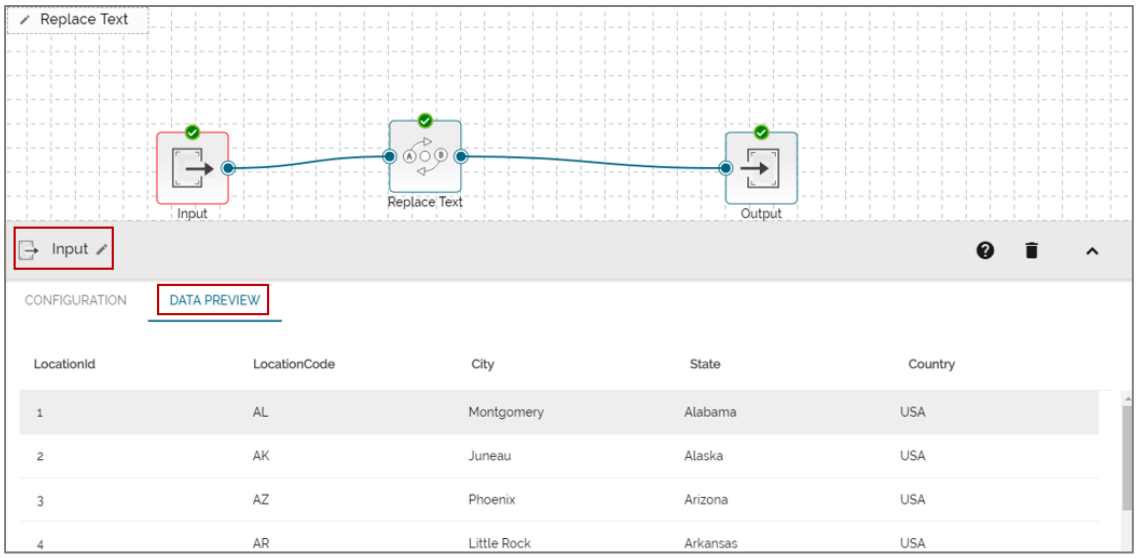
5.8. Replace Text

Users can search by whole word, sensitive to case, search for special values like NULL or empty strings, or use regular expressions, and then replace with any given constant values or even empty strings. Only text columns can be transformed using this component. Users can replace text for the multiple text columns.

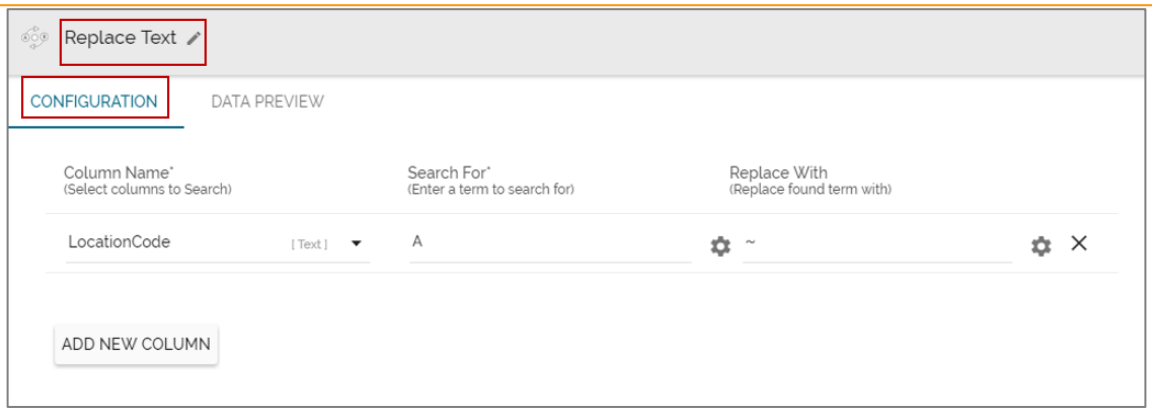
i) Navigate to the Workflow editor.
ii) Connect the '**Replace Text**' component with the configured Input dataset and Output component.



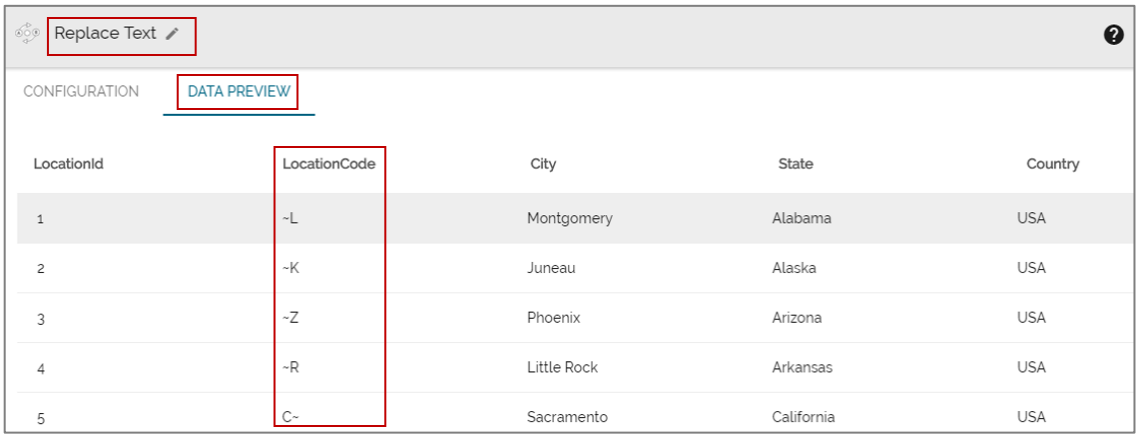
iii) Run the workflow to preview the input data.



iv) Configure the '**Replace Text**' component as described below:
a. Column Name: Select a column from the input data set.
b. Search for: Enter a term from the selected column to search for.
c. Replace with: Enter a term to replace the searched term in the input data.



v) Run the workflow.
vi) Save the workflow.
vii) Open the Output data preview to see the replacement of the selected text in the column.



LocationId	LocationCode	City	State	Country
1	~L	Montgomery	Alabama	USA
2	~K	Juneau	Alaska	USA
3	~Z	Phoenix	Arizona	USA
4	~R	Little Rock	Arkansas	USA
5	C~	Sacramento	California	USA

Note:

a. Users can click on the **'ADD NEW COLUMN'** option to configure the multiple columns for any transform component.
b. Users can also see data preview of the various transform components.

6. Merge

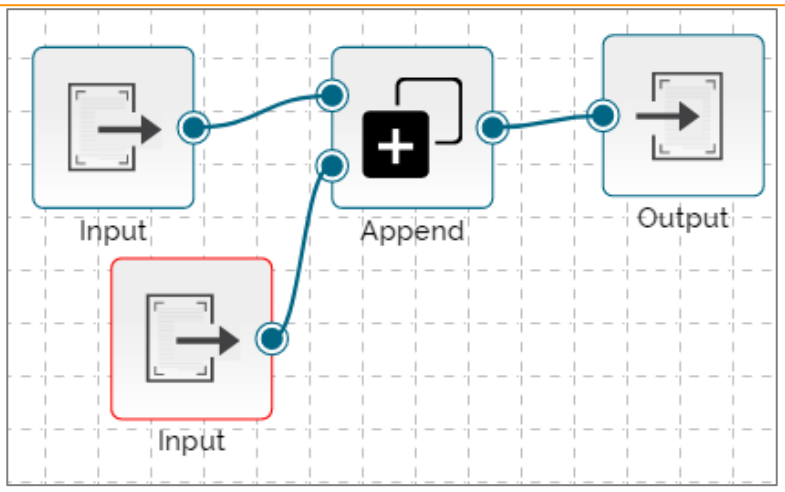
Users can use the '**Merge**' components to combine input data sets and get the required output.

6.1. Append

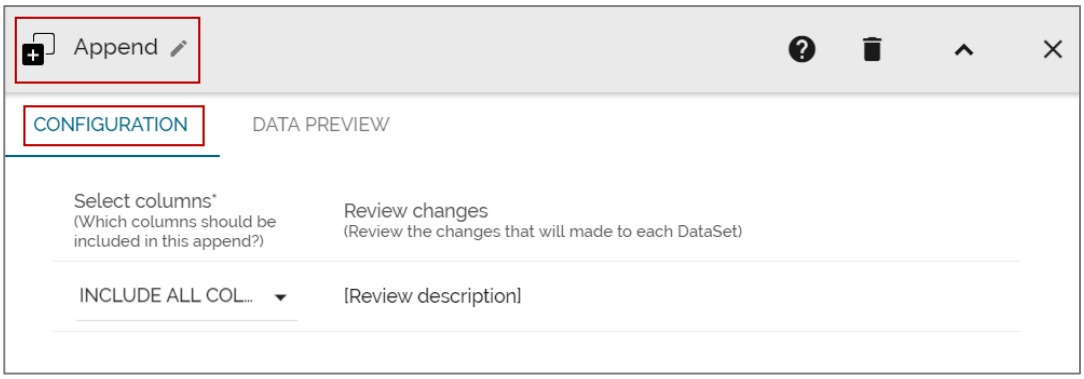
The '**Append**' feature combines one dataset on top of another. If the datasets are of different structures, still the union is possible, and the output will be a unified more massive structure with NULL values populated wherever data is missing. Users can choose whether to include only shared columns or all columns to append.

6.1.1. Append All Columns

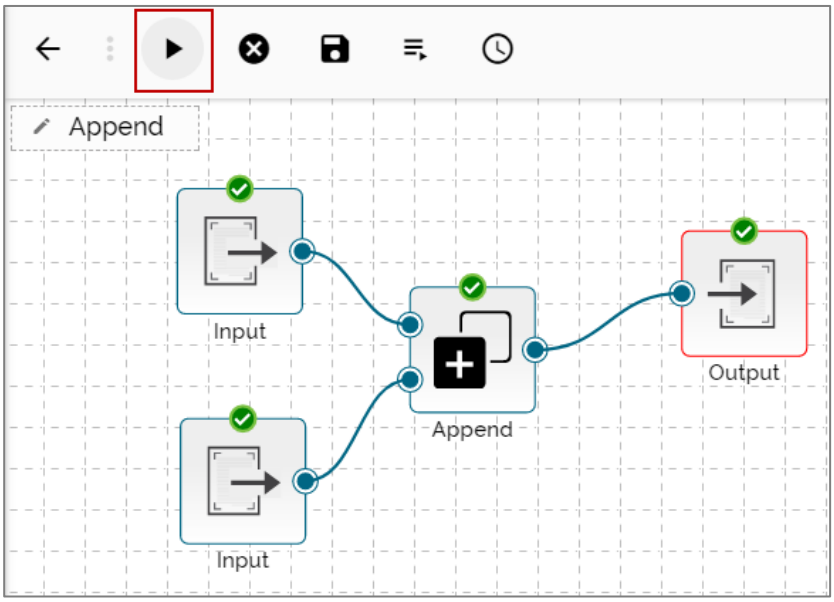
i) Navigate to the Workflow editor.
ii) Configure two input datasets.
iii) Connect the '**Append**' component with the configured Input datasets and an Output component.



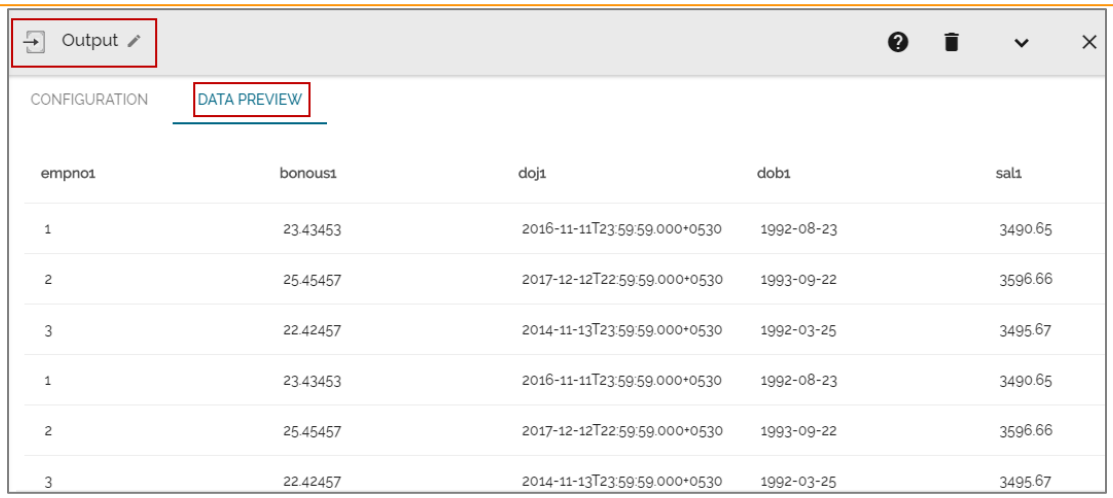
iv) Select '**Include All Columns**' option using the '**Select Columns**' drop-down menu.



v) Save the workflow.
vi) Run the workflow.

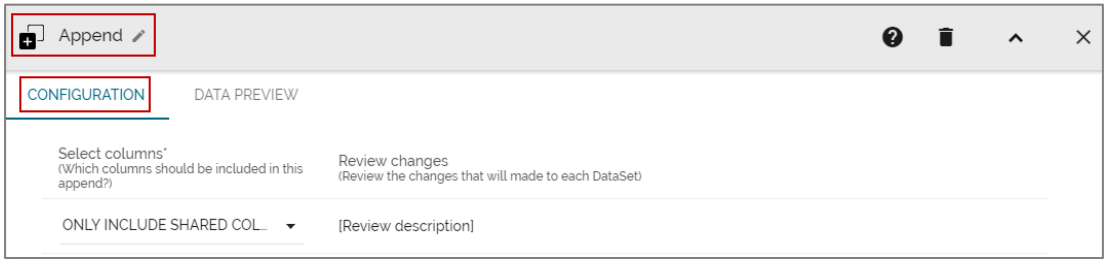


vii) The entire data of both the input data sets will be appended in the output data preview.

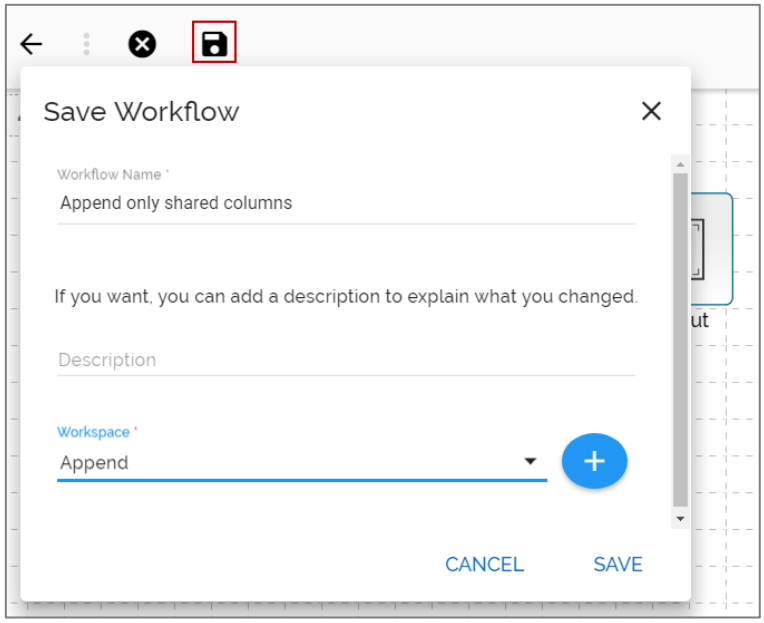


Append Only Shared Columns

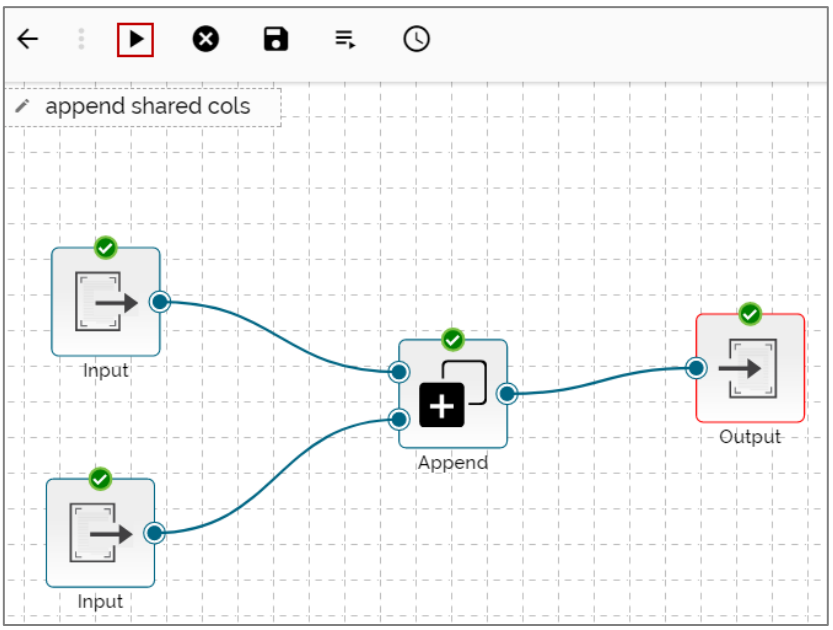
i) Connect the '**Append**' component to the configures input datasets and an output component.
ii) Choose '**ONLY INCLUDE SHARED COLUMNS**' as an option to append the datasets.
iii) The entire data of both the input data sets will be appended in the output data preview.



iv) Save the Workflow.



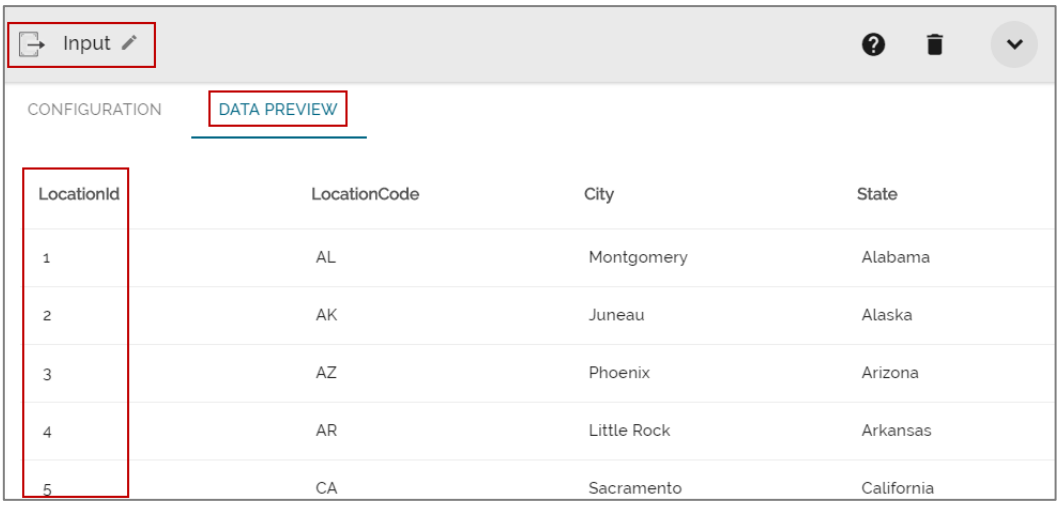
v) Run the Workflow.



vi) The shared column(s) will be appended in the output data set.

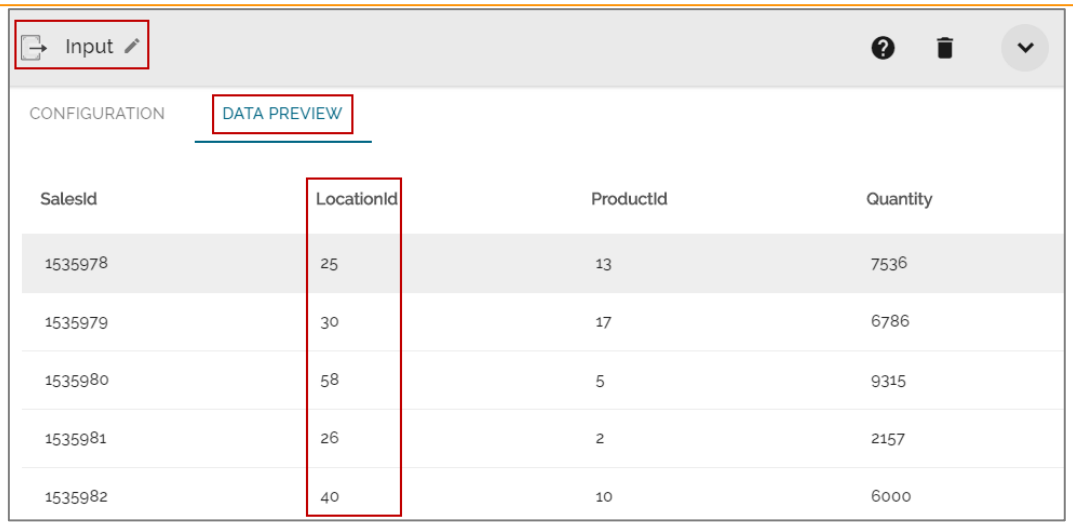
E.g. The following images illustrate that the shared column '**Location**' has been displayed under the data preview of Append and Output components.

a. Input Dataset-1

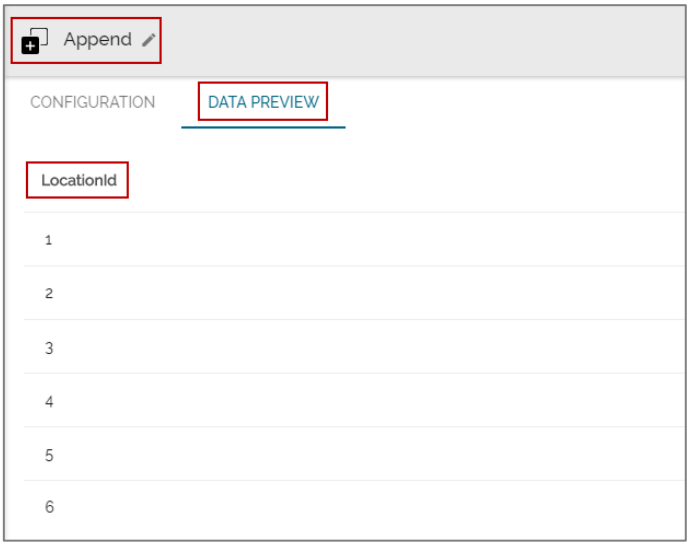


LocationId	LocationCode	City	State
1	AL	Montgomery	Alabama
2	AK	Juneau	Alaska
3	AZ	Phoenix	Arizona
4	AR	Little Rock	Arkansas
5	CA	Sacramento	California

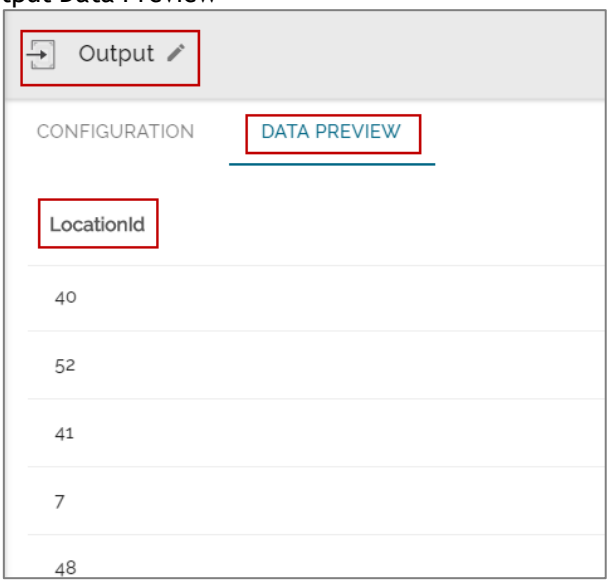
b. Input Dataset-2



c. Append Data Preview



d. Output Data Preview

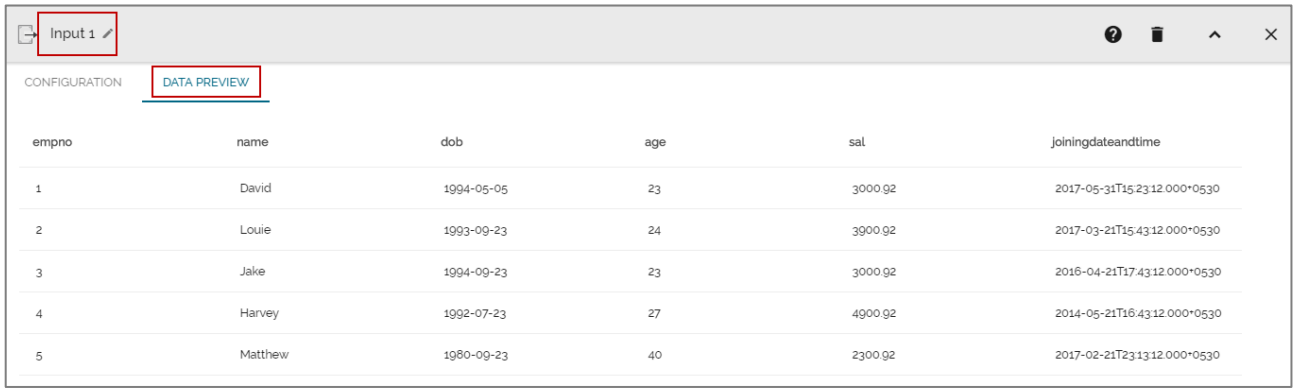


6.2. Join

Users can join two datasets and use the merged output to write the workflow in the selected metadata.

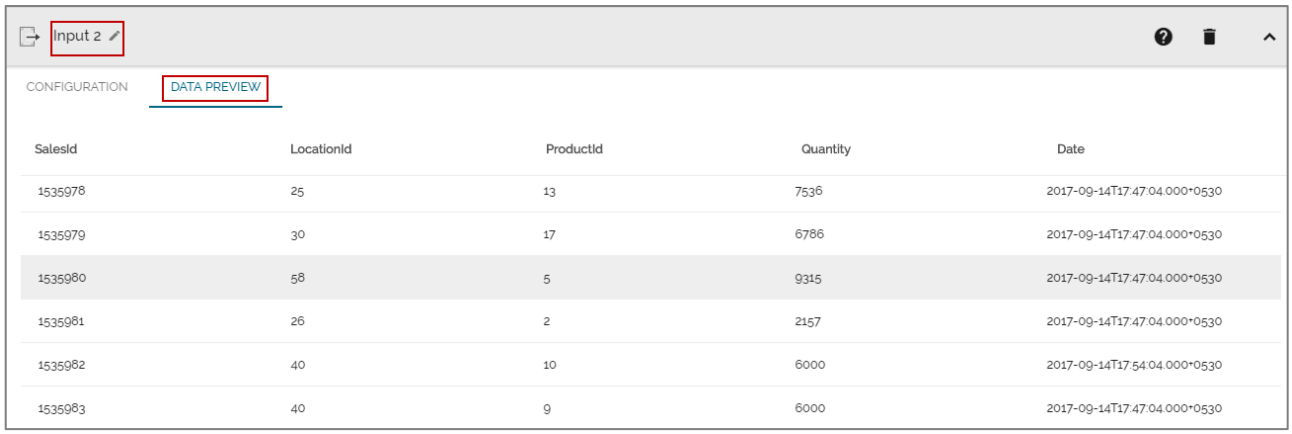
i) Drag two input datasets and configure them to see the dataset preview.

Input Data Set 1



empno	name	dob	age	sal	joiningdateandtime
1	David	1994-05-05	23	3000.92	2017-05-31T15:23:12.000+0530
2	Louie	1993-09-23	24	3900.92	2017-03-21T15:43:12.000+0530
3	Jake	1994-09-23	23	3000.92	2016-04-21T17:43:12.000+0530
4	Harvey	1992-07-23	27	4900.92	2014-05-21T16:43:12.000+0530
5	Matthew	1980-09-23	40	2300.92	2017-02-21T23:13:12.000+0530

Input Data Set 2

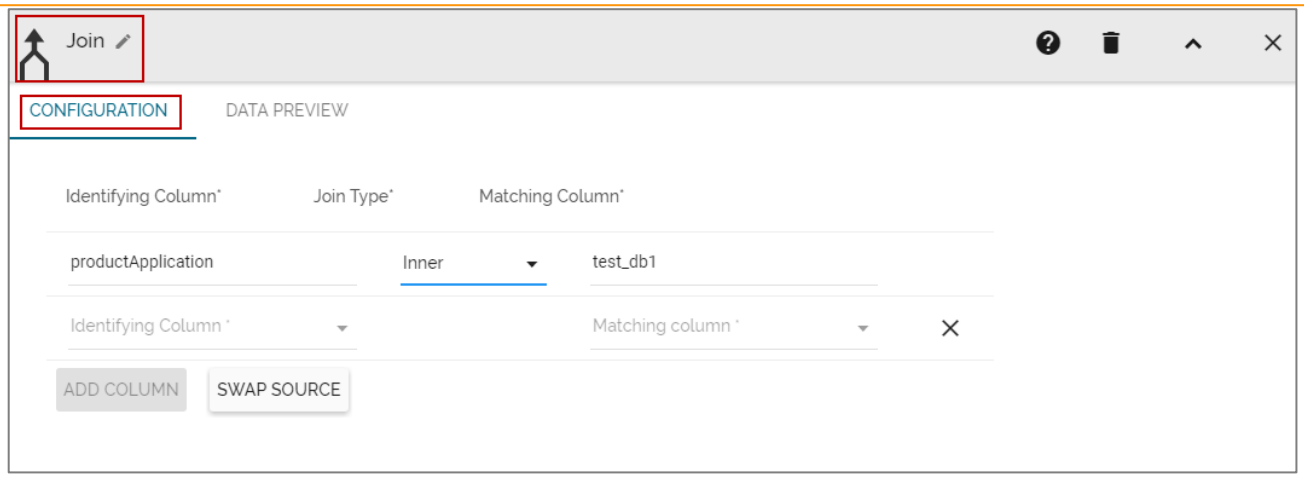


SalesId	LocationId	ProductId	Quantity	Date
1535978	25	13	7536	2017-09-14T17:47:04.000+0530
1535979	30	17	6786	2017-09-14T17:47:04.000+0530
1535980	58	5	9315	2017-09-14T17:47:04.000+0530
1535981	26	2	2157	2017-09-14T17:47:04.000+0530
1535982	40	10	6000	2017-09-14T17:54:04.000+0530
1535983	40	9	6000	2017-09-14T17:47:04.000+0530

ii) Connect the '**Join**' component with the above-given input datasets and one output component to complete the workflow.

iii) Configure the '**Join**' component as described below:

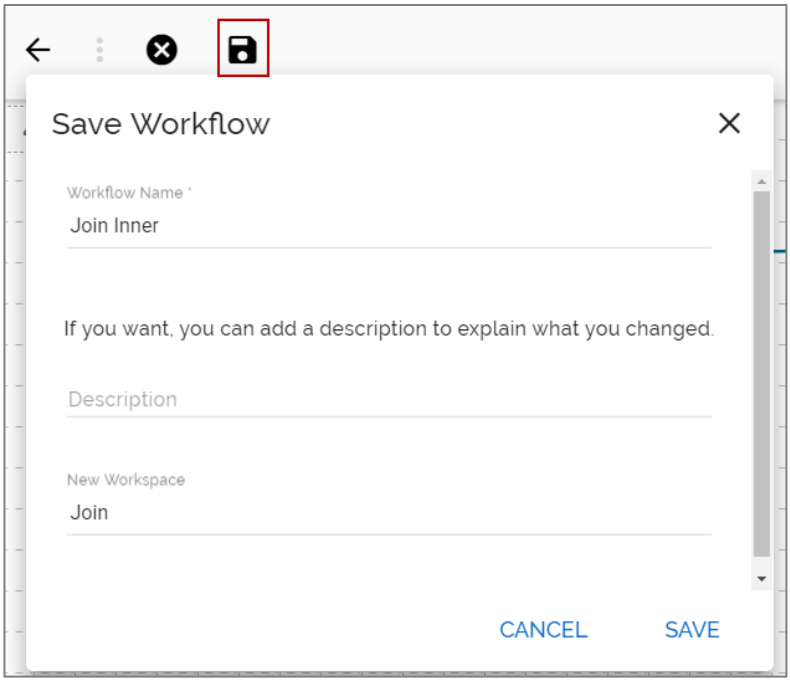
a. Identify Column: Identify a column from the input dataset 1
b. Join Type: Choose a join type to merge the selected datasets out of the given choices
    i. Inner
    ii. Left Outer
    iii. Right Outer
    iv. Full Outer
c. Matching Column: Select a column from the input dataset 2



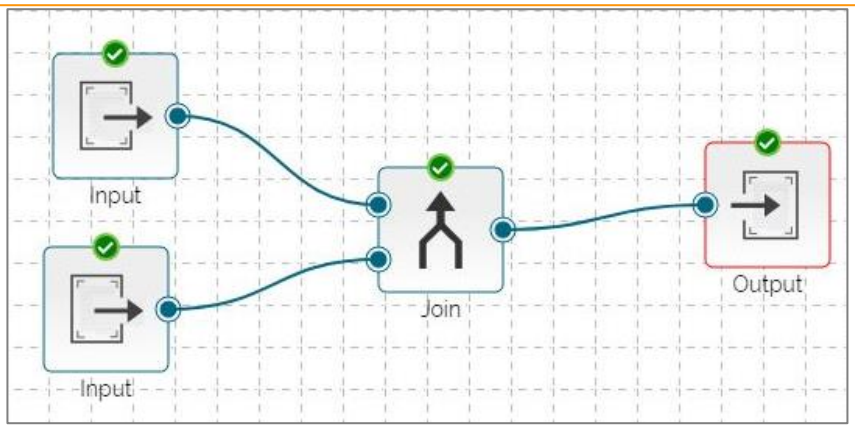
Note:

a. By default, the '**Inner**' join type will be selected. Users can apply multiple inner joins by using the '**ADD COLUMN**' tab.
b. Click '**SWAP SOURCE**' to interchange the input datasets and the selected columns from the data sets.

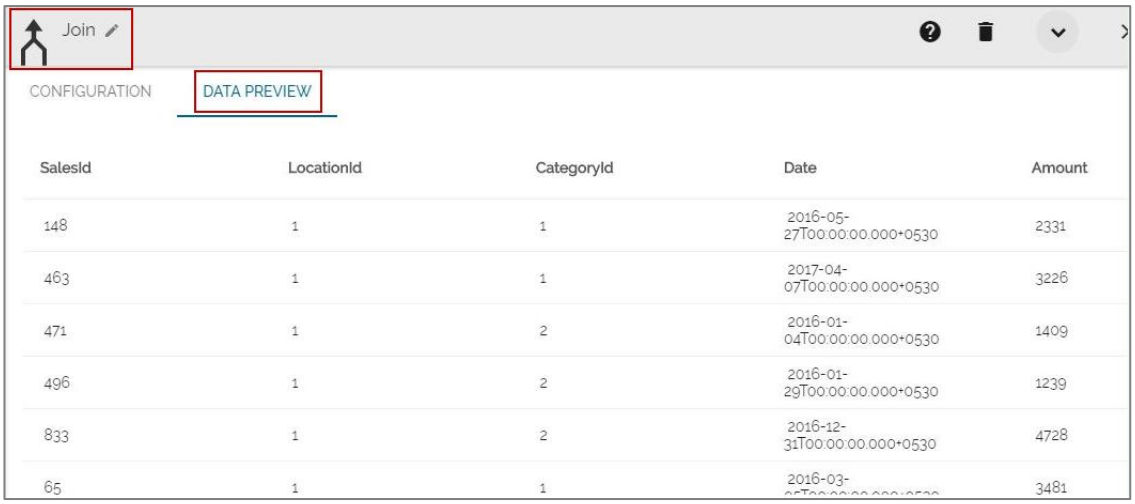
iv) Save the workflow.



v) Run the workflow.

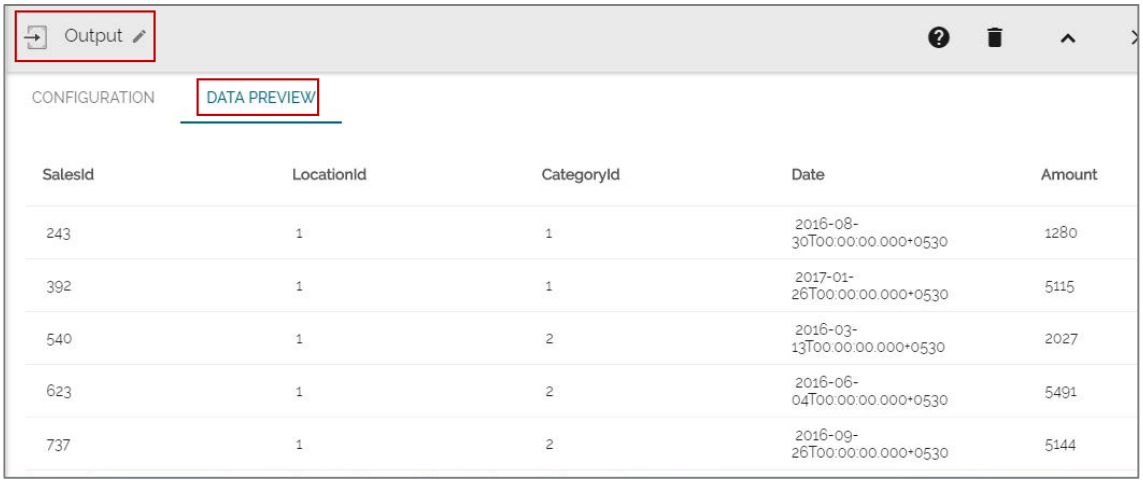


vi) Click the '**Data Preview**' tab from the Join component to view data preview of the merged data.



SalesId	LocationId	CategoryId	Date	Amount
148	1	1	2016-05-27T00:00:00.000+0530	2331
463	1	1	2017-04-07T00:00:00.000+0530	3226
471	1	2	2016-01-04T00:00:00.000+0530	1409
496	1	2	2016-01-29T00:00:00.000+0530	1239
833	1	2	2016-12-31T00:00:00.000+0530	4728
65	1	1	2016-03-	3481

vii) Users can preview data under the '**Data Preview**' tab of the selected output component.



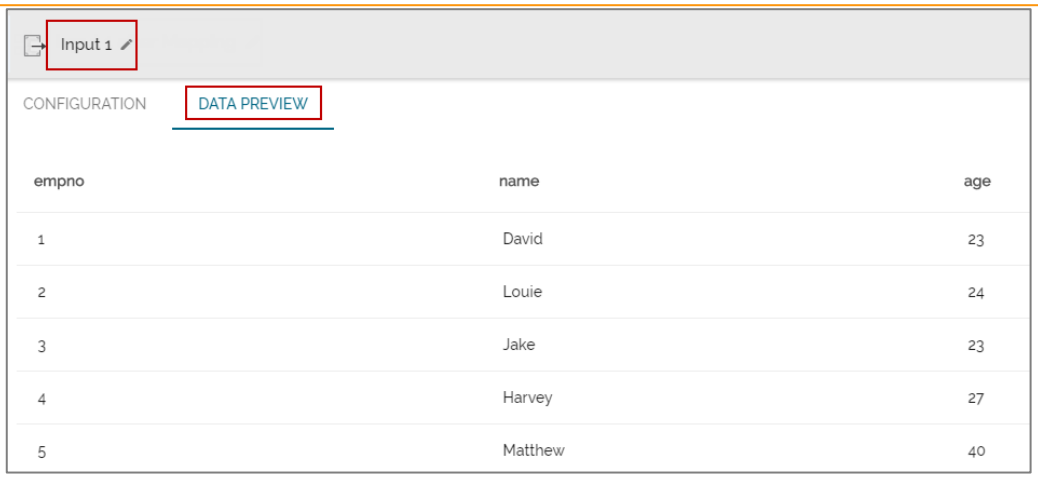
SalesId	LocationId	CategoryId	Date	Amount
243	1	1	2016-08-30T00:00:00.000+0530	1280
392	1	1	2017-01-26T00:00:00.000+0530	5115
540	1	2	2016-03-13T00:00:00.000+0530	2027
623	1	2	2016-06-04T00:00:00.000+0530	5491
737	1	2	2016-09-26T00:00:00.000+0530	5144

6.2.1. Join Types:

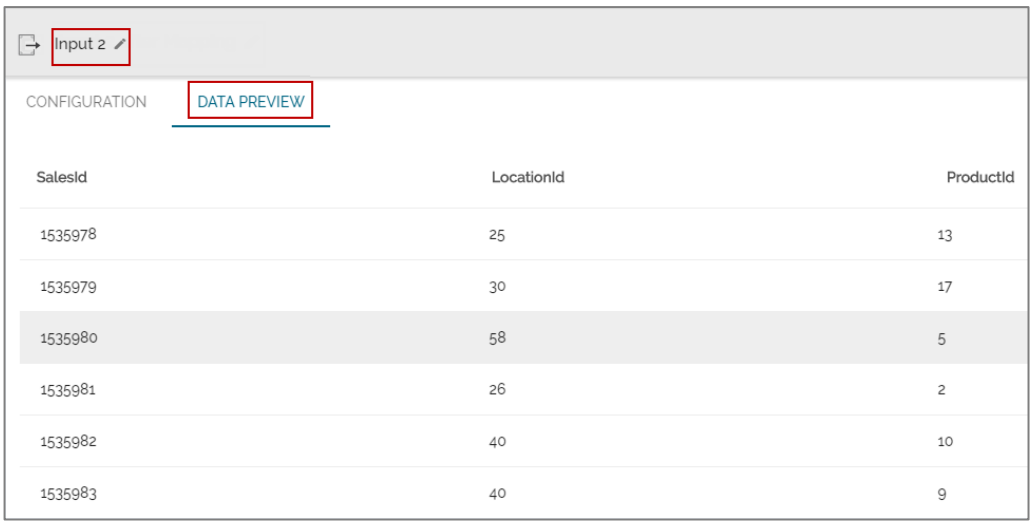
The '**Join**' feature offers four types of join to merge datasets.

The sample data sets used to describe the supported join types are:

1. **Input Dataset 1**



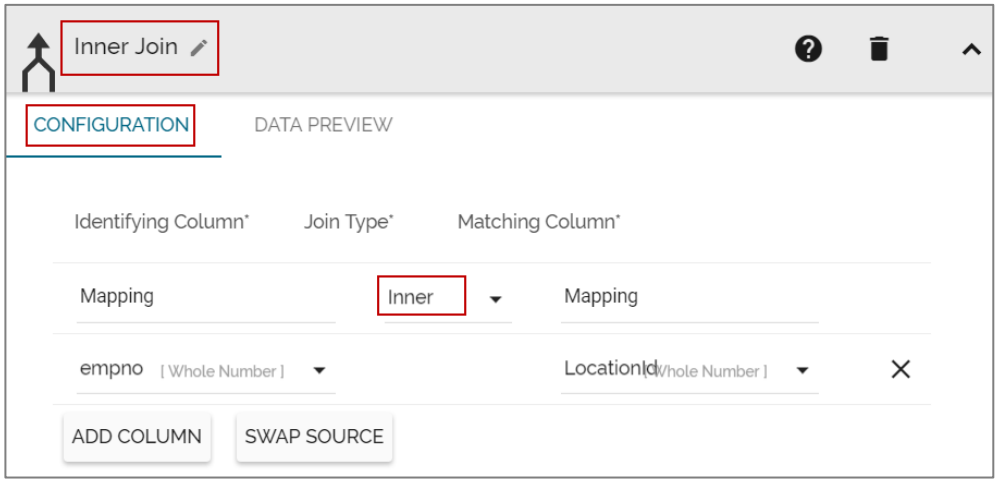
2. **Input Dataset 2**



a) Inner Join

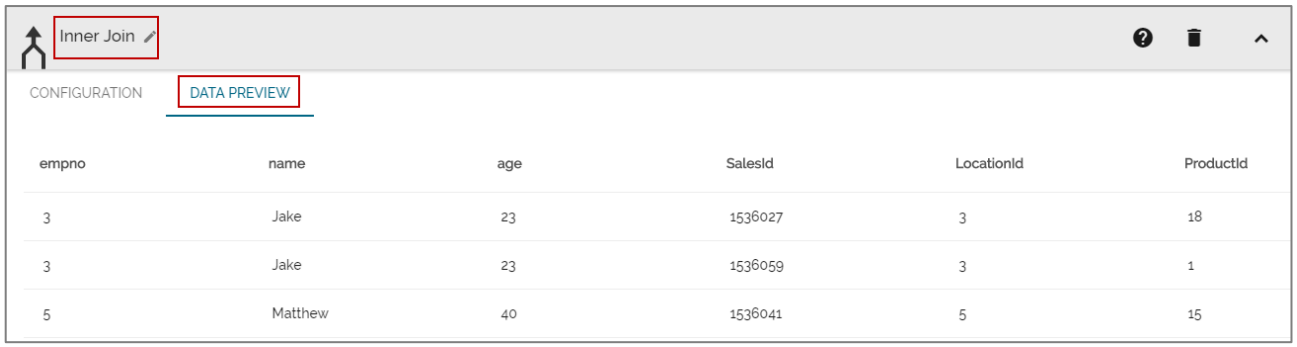
i. Connect the join component to the configured input datasets and output component to create a workflow.

ii. Specify a join type from the '**Configuration**' tab of the join component.



iii. Save and run the workflow.

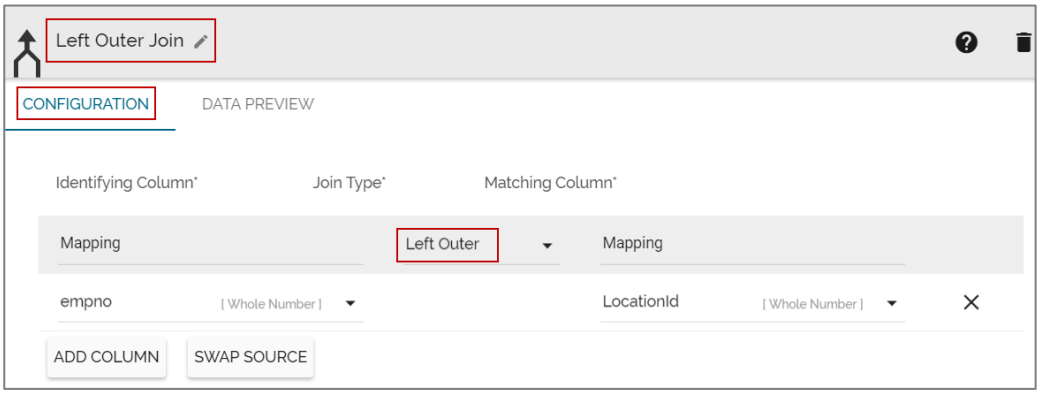
iv. Click the '**Data Preview**' tab using the join component to view the merged datasets.



b) Left Outer Join

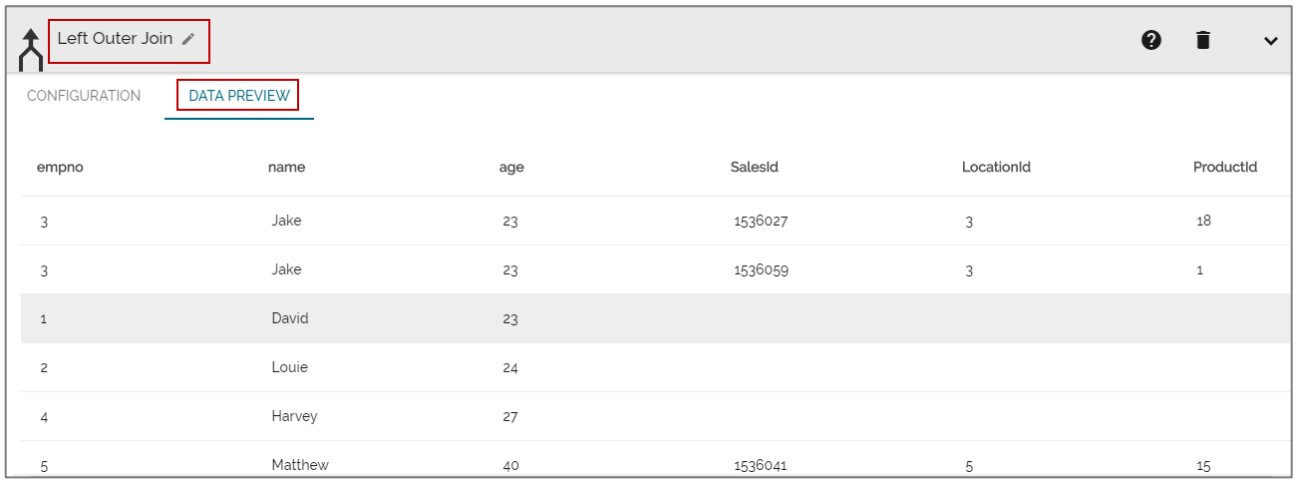
i. Connect the join component to the configured input datasets and output component to create a workflow.

ii. Specify a join type from the '**Configuration**' tab of the join component.



iii. Save and run the workflow.

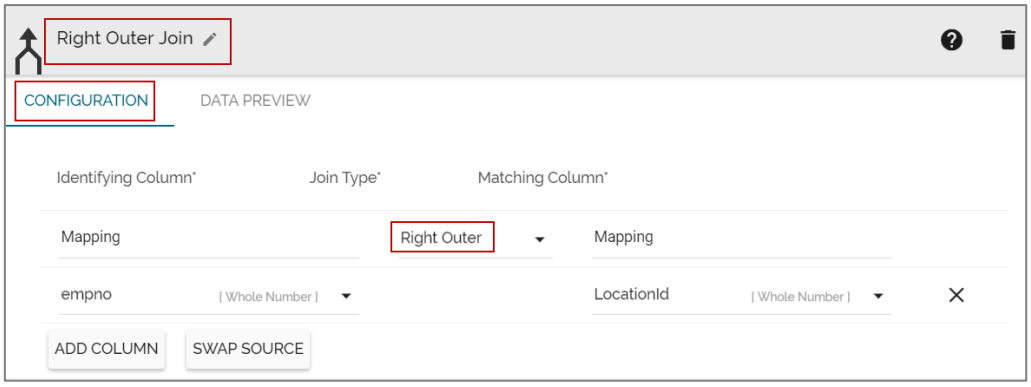
iv. Click the '**Data Preview**' tab using the join component to view the merged datasets.



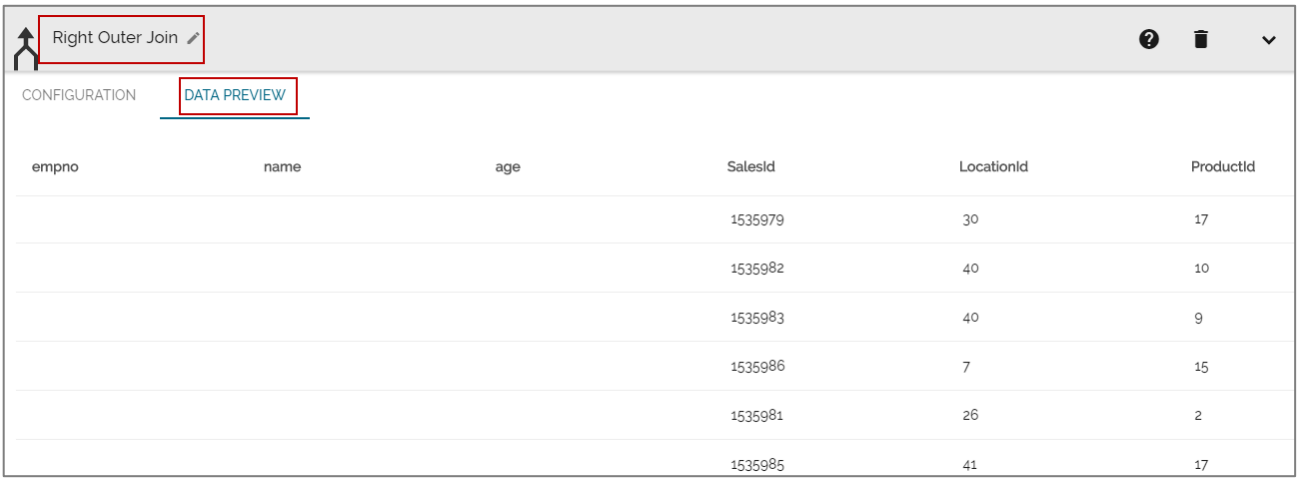
Note: The output data preview will be aligned with the selected left input dataset.

c) Right Outer Join

i. Connect the join component to the configured input datasets and output component to create a workflow.
ii. Specify a join type from the '**Configuration**' tab of the join component.

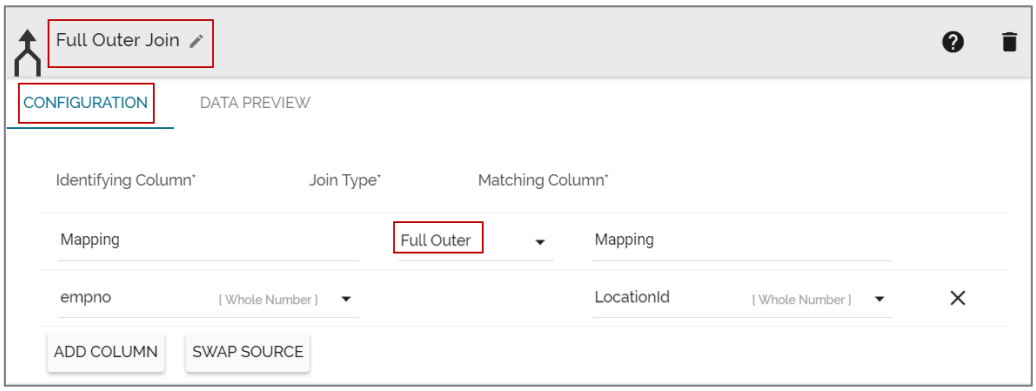


iii. Save and run the workflow.
iv. Click the '**Data Preview**' tab using the join component to view the merged datasets.

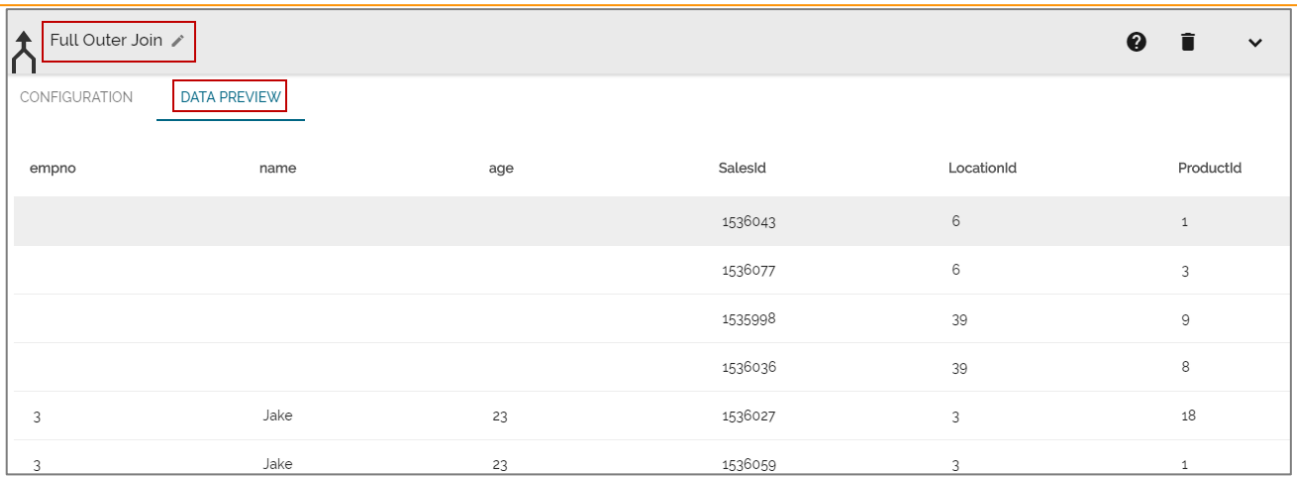


d) Full Outer

i. Connect the join component to the configured input datasets and output component to create a workflow.
ii. Specify a join type from the '**Configuration**' tab of the join component.



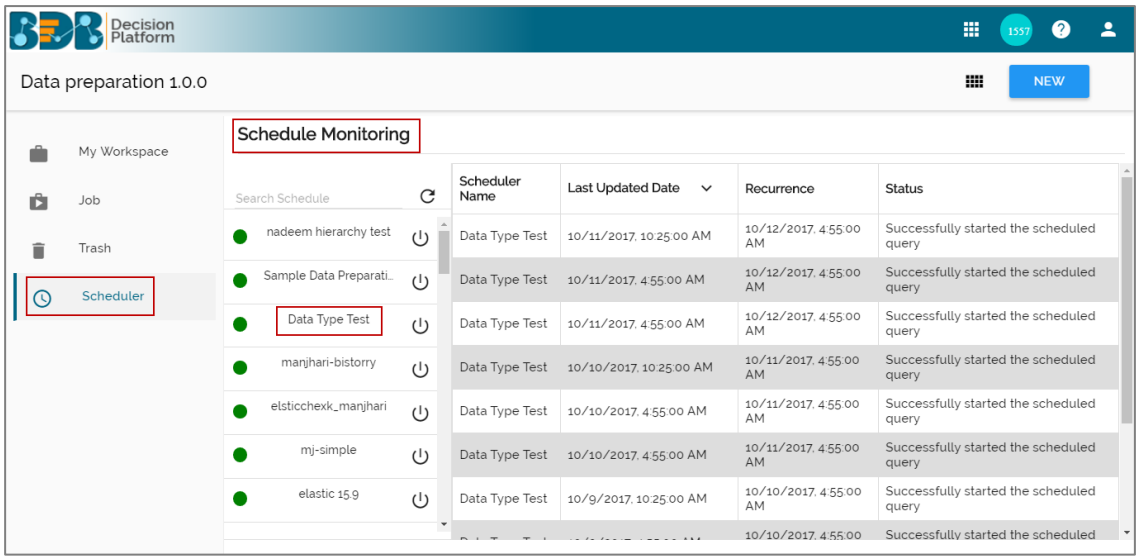
iii. Save and run the workflow.
iv. Click the '**Data Preview**' tab using the join component to view the merged datasets.



7. Scheduler

The 'Scheduler' section displays the schedule monitoring details. Users can see a list containing all the scheduled workflows.

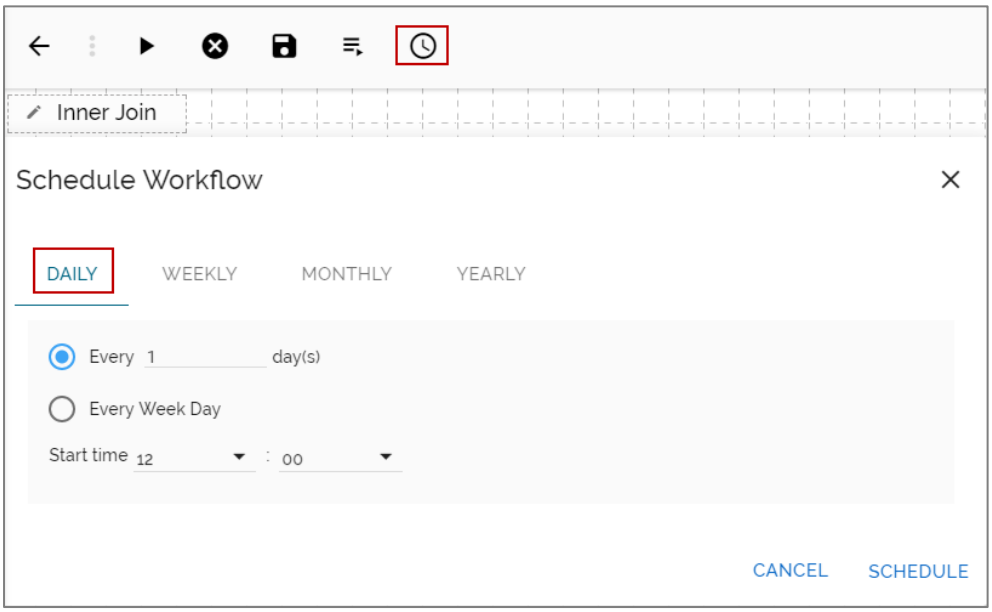
i) Click the '**Navigator**' icon
ii) Select '**Scheduler**' from the menu panel.
iii) Users will be redirected to the '**Schedule Monitoring**' page.
iv) The scheduled workflow will be added to the list of all the schedules.
v) Click on a scheduled workflow will display the following schedule details:
   a. Scheduler Name
   b. Last Updated Date
   c. Recurrence date and time
   d. Status



7.1. Schedule Configuration Options

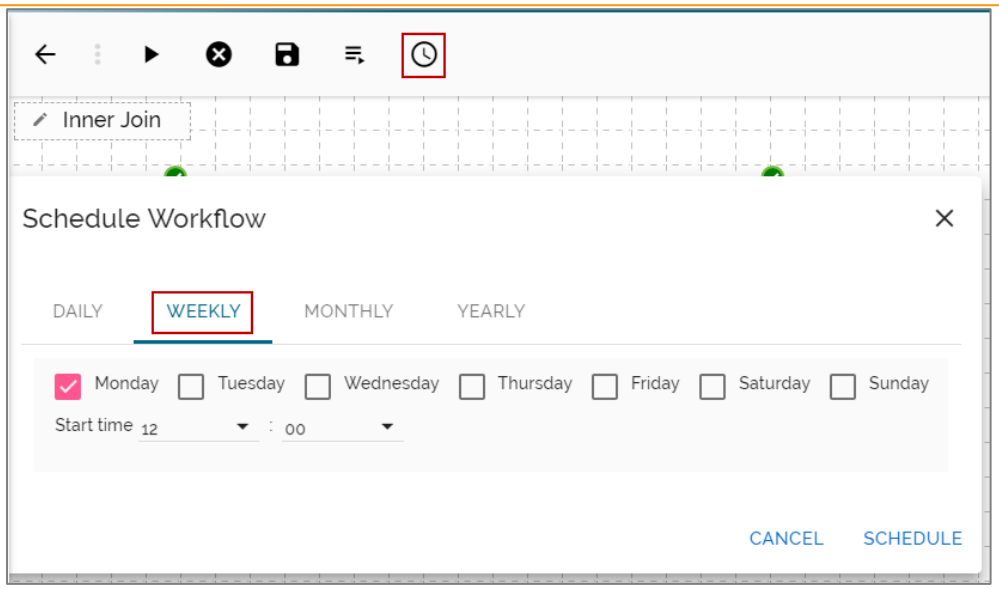
These options are provided to configure a range of time for a scheduled workflow. The user can select only one option at a time from the given menu.

1. **Daily:** User can schedule the job daily by using this option.
    a. Click the '**Scheduler**' icon on the workflow editor.
    b. Choose the '**DAILY**' option from the '**Schedule Workflow**' window (It is a default option).
        i. Select an option out of the given choices.
            1. Every __ day(s)
            2. Every Weekday
            3. Set the start time using the drop-down.
    c. Click the '**SCHEDULE**' option.



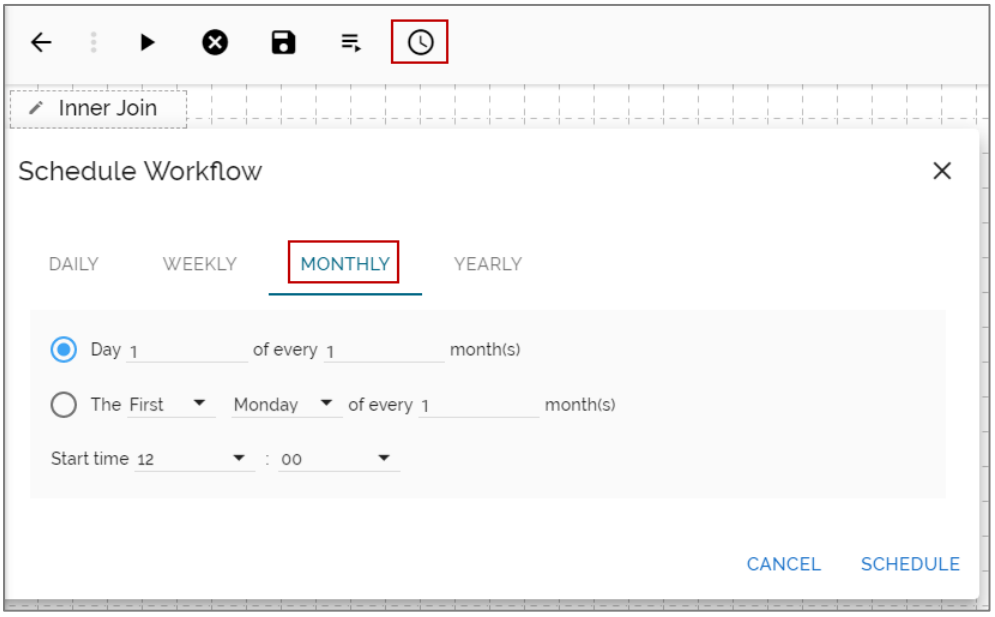
2. **Weekly:** User can schedule the job weekly by using this option.

    a. Click the '**Scheduler**' icon on the workflow editor.
    b. Choose the '**WEEKLY**' option from the '**Schedule Workflow**' window.
        i. Select an option out of the given choices.
            1. Choose the days of the week by check marking in the box.
            2. Set the start time using the drop-down.
    c. Click the '**SCHEDULE**' option.



3. Monthly: User can schedule the job Monthly by using this option.

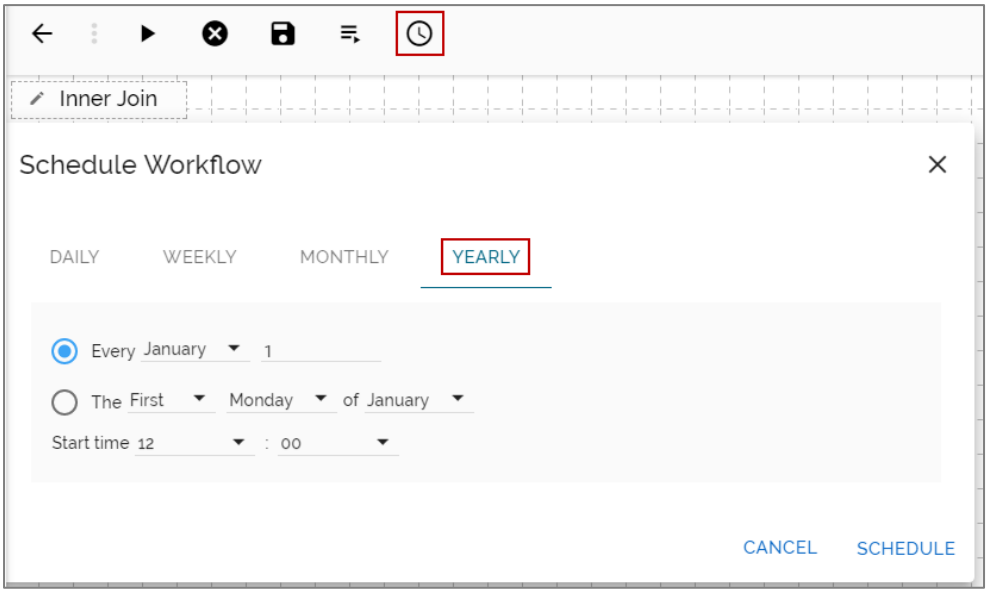
a. Click the '**Scheduler**' icon on the workflow editor.
b. Choose the '**MONTHLY**' option from the '**Schedule Workflow**' window.
    i. Select an option out of the given choices to choose a day for each month.
    ii. Set the start time using the drop-down.
c. Click the '**SCHEDULE**' option.



4. Yearly: User can schedule the job yearly by using this option.

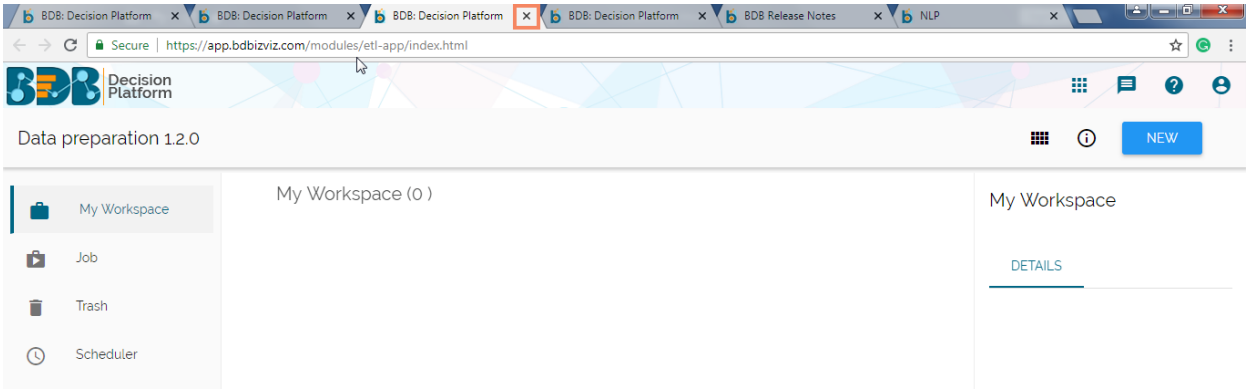
a. Click the '**Scheduler**' icon on the workflow editor.
b. Choose the '**YEARLY**' option from the '**Schedule Workflow**' window.
    i. Select an option out of the given choices.
        1. Specify either a day or date of a specific month in a year.
        2. Set the start time using the drop-down.

c. Click the '**SCHEDULE**' option.



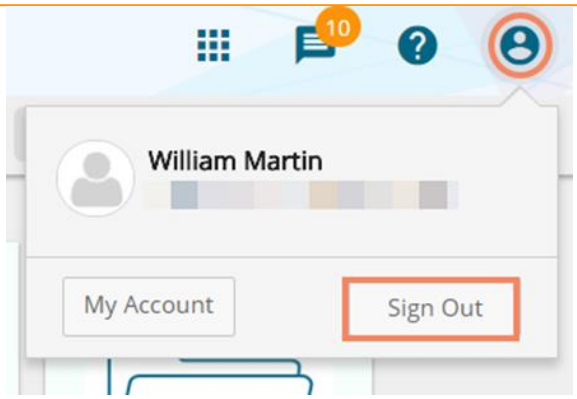
8. Signing Out

It is possible for a user to log out from the BDB Data Preparation plugin at any given stage. Users need to click on the '**Close**' ✕ option to close the Data Preparation page.



Follow the below given steps to log out from the BDB Platform.

i) Click the '**User**' icon on the Platform homepage.
ii) A menu appears with the logged in user details (User's name and email id).
iii) Click '**Sign Out.**'



iv) Users can successfully log out of the **BDB** Platform.

Note: Clicking on '**Sign Out**' redirects the user back to the login page of the BDB platform.