How to Guide

Transformation Components

(Data Pipeline)

Version: Release 1.1

# Contents

Component Description-

Transformation components allow the users to transform the data. BDB Data Pipeline provides the following Transformations components to transform a variety of data.

1. Aggregation

The Aggregation transform component allows the users to perform the following actions on the selected input columns:

a. If the input data type is number (E.g., Integer or double) then, the output column takes the data after performing the addition aggregation function on the selected columns.
b. If the input data type is String, then after applying the Aggregation transform the data gets concatenated from the selected columns.
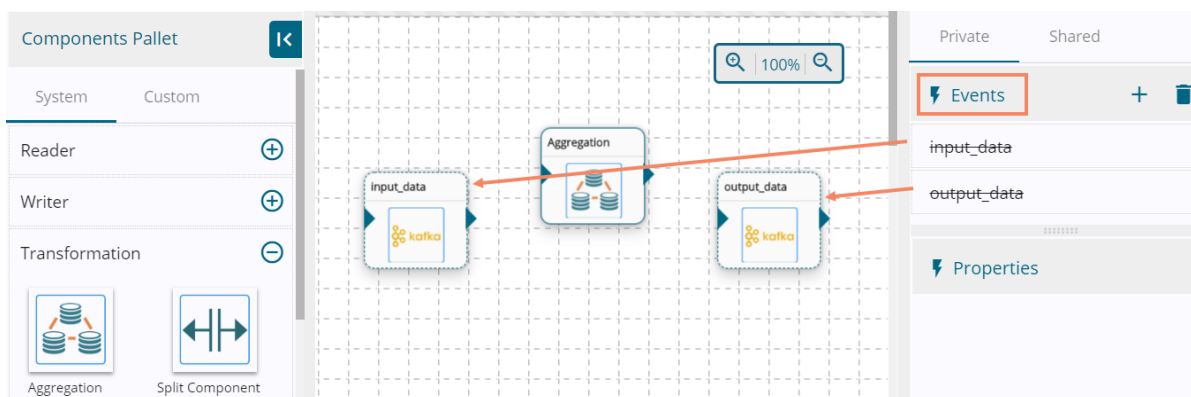
2) Navigate to the Pipeline Workflow Editor.
3) Expand the Transformation section using the Components Pallet.
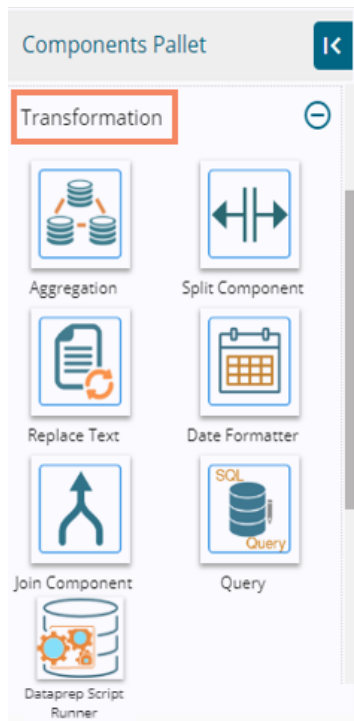
4) Drag and drop the Aggregation component to the workspace.
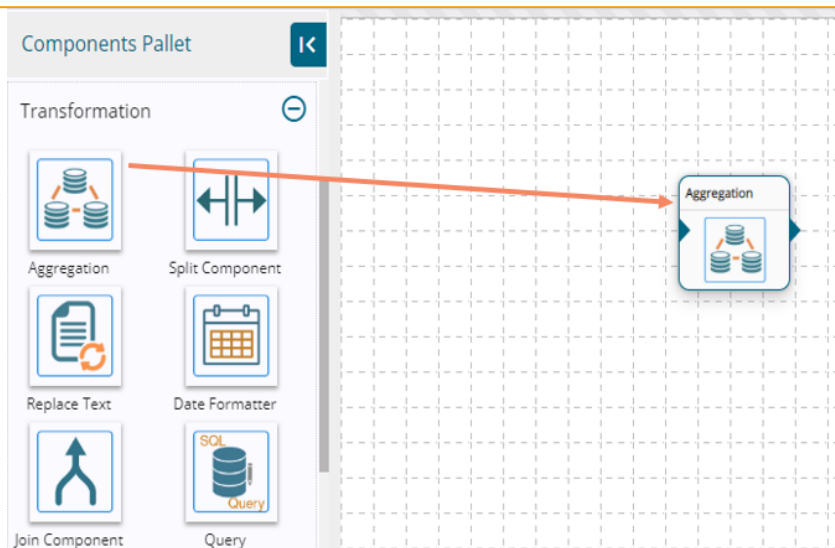
5) Transformation component requires an input event (to get the data) and sends the data to an output event.

6) Create two Events and drag them to the Workspace.

7) Connect the input event, and the output event to the dragged Transform component.
(The data in input event can come from any Ingestion, Readers or shared events.)

 www.bdb.ai
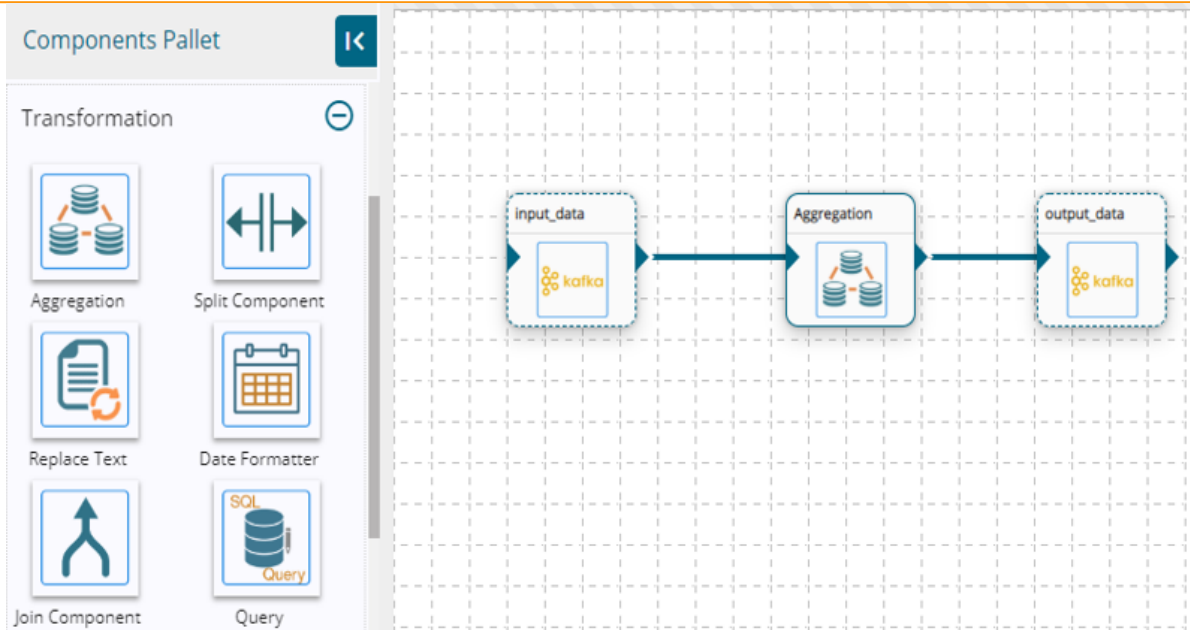
8) Click the Aggregation component to get the configuration tabs.
9) The Basic Information tab opens by default.
    a. Select the invocation type (Real Time/Batch).

    Note: Currently, Pipeline supports only Real-Time option.
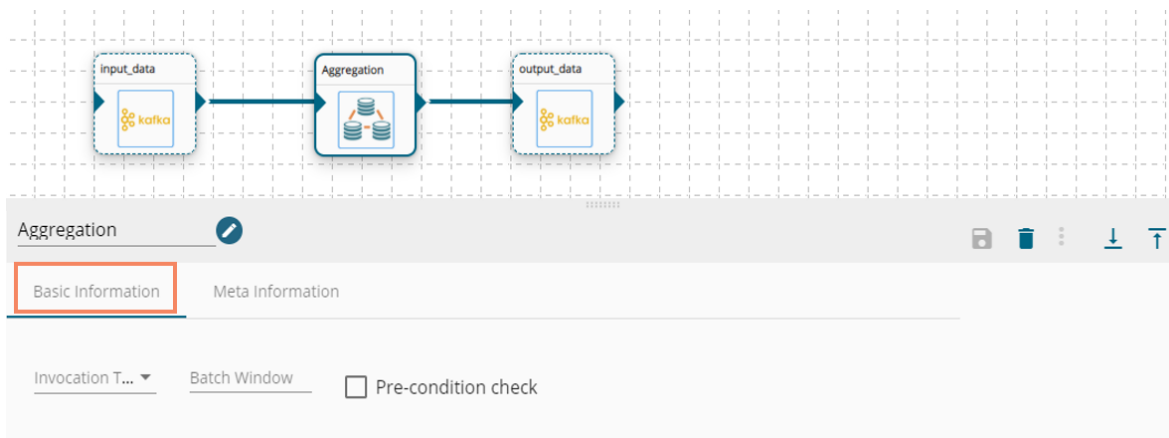
10) Select Meta Information and click on column type field.
    a. A drop-down menu appears with options to select a data type for the input columns.
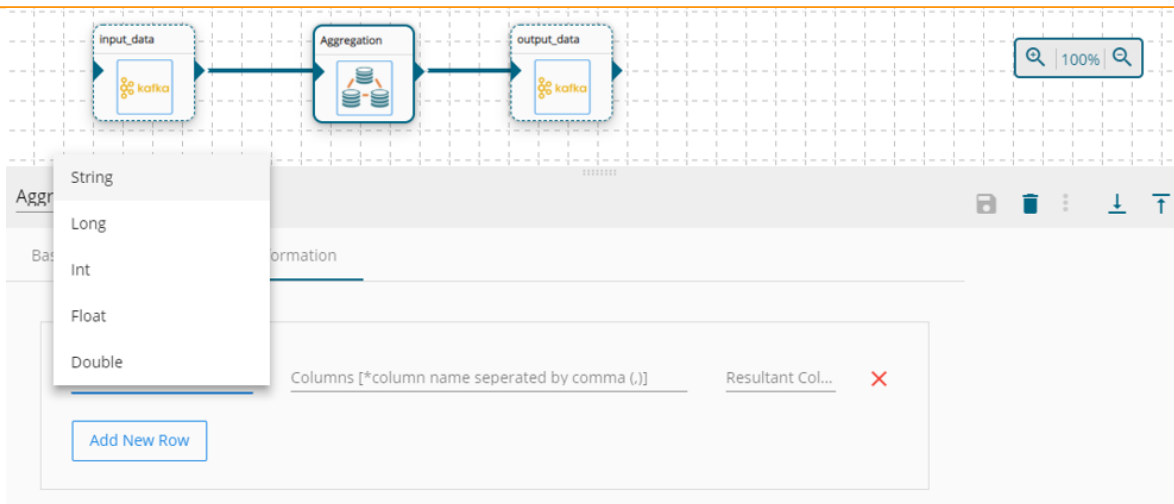
b. Provide input Columns Name (separated by comma).

c. Give a name for the Resultant Column (the output column) which gets created after transformation.

d. To perform multiple aggregations, click the 'Add New Row' option and follow the steps as mentioned above.

e. Save the Aggregation component.
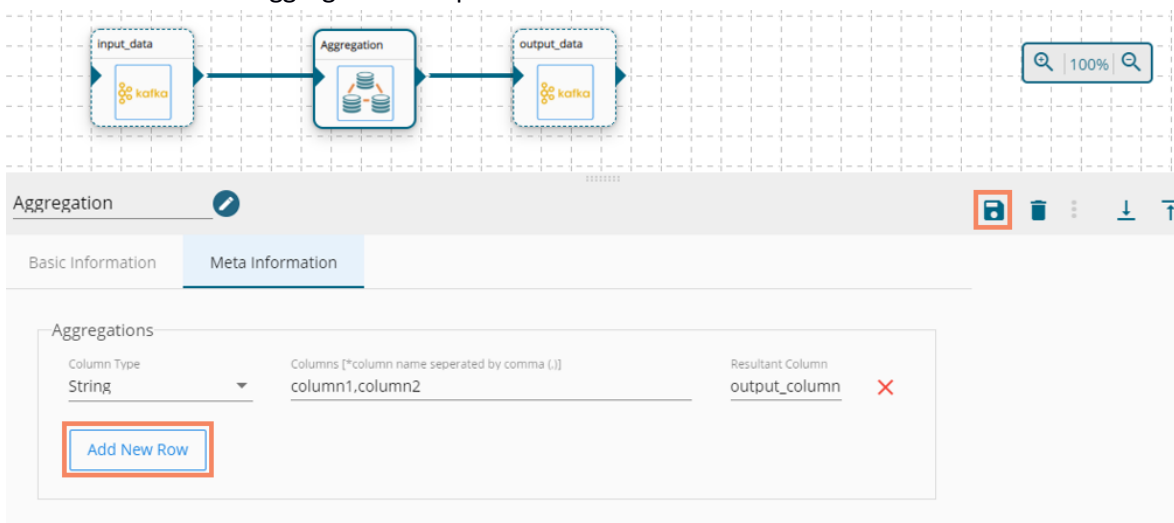
11) Save the pipeline and activate it.

12) The component reads the data coming from the input event, transforms the data and sends the output data with newly created column(s) to the output event.
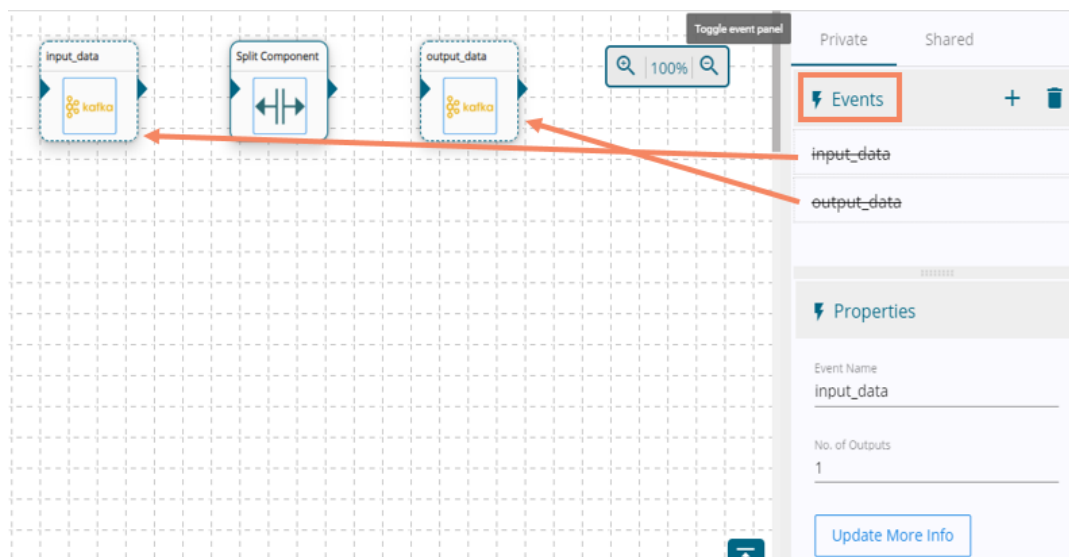
2. Split Component

The Split component helps users to split the selected column(s) from the input data set based on the given regular expression.

1) Navigate to the Pipeline Workflow Editor.
2) Expand the Transformation section using the Components Pallet.
3) Drag the Split Component to the Workspace.
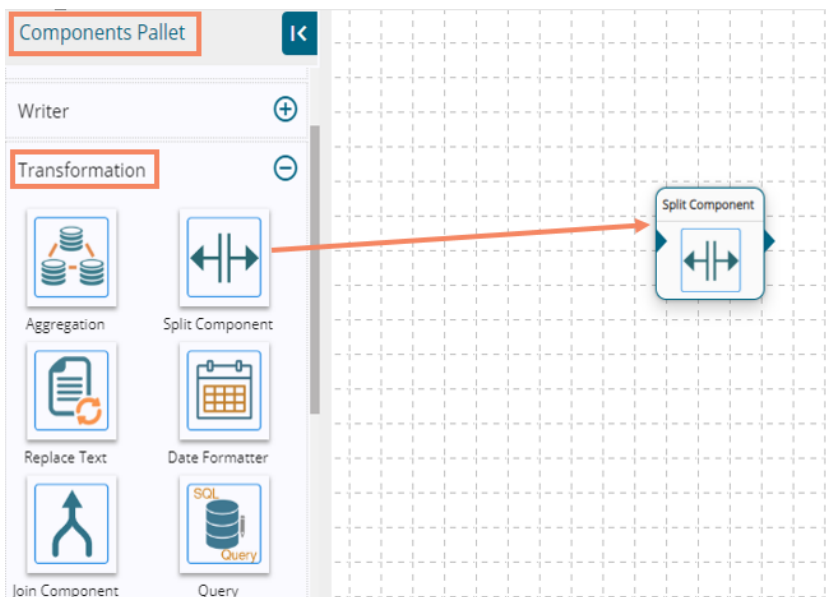
4) Transformation component requires an input event (to get the data) and sends the data to an output event.
5) Create two Events and drag them to the Workspace.

6) Connect the input event and the output event (The data in the input event can come from any Ingestion, Readers or shared events).
7) Click on the Split component to get the configuration tabs.
8) The Basic Information tab opens by default.
   a. Select the invocation type (Real-Time/Batch).

   Note: Currently, Pipeline supports only Real-Time option.
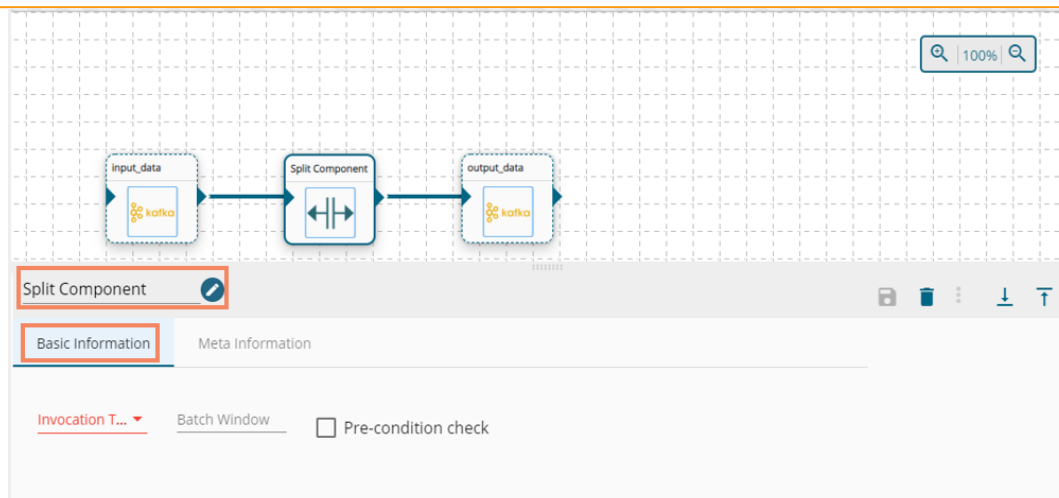
9) Select Meta Information and fill all the required inputs fields.
   a. Fill the Regular Expression by which the data gets split (E.g., Comma (,) is used as the Regular Expression in the below given image).
   b. Column Name- Provide the input column name.
   c. New Columns- Provide Output columns name separated by a comma. These output columns get created after transformation.
   d. Provide Default Value in the given field.
   e. To perform multiple Split transformation actions, click the '**Add New Row**' option and follow the above steps.
   f. Save the Split component.
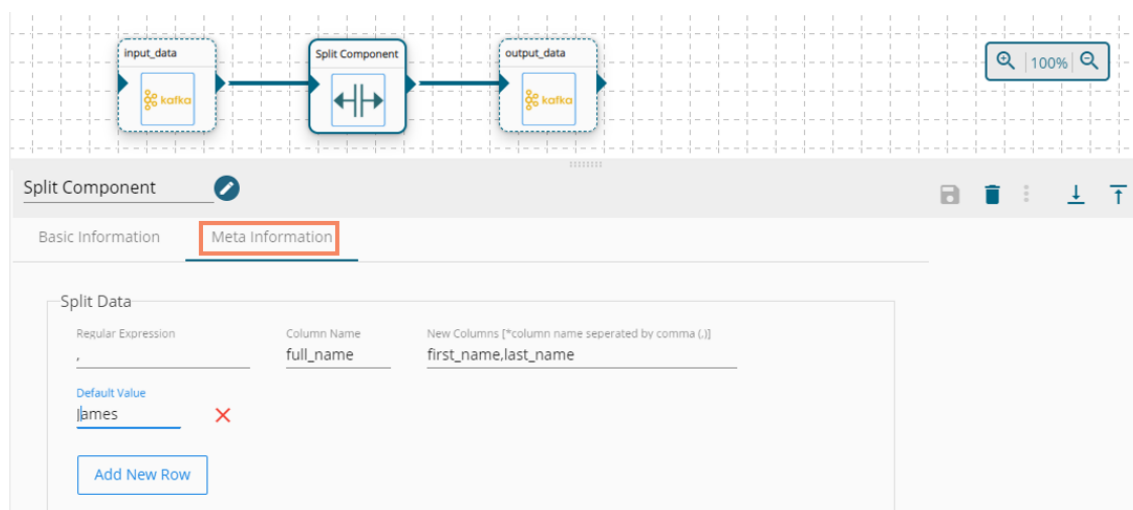
10) Save the pipeline and activate it. The component reads the data coming from input event, transform the data and give the output data with newly created column(s).

3. Replace Text Component

The Replace Text transform component allows users to replace the searched data in the selected columns with the user-defined replacement text.

1) Navigate to the Pipeline Workflow Editor.
2) Expand the Transformation section using the Components Pallet.
3) Drag the Replace Text component to the workspace.
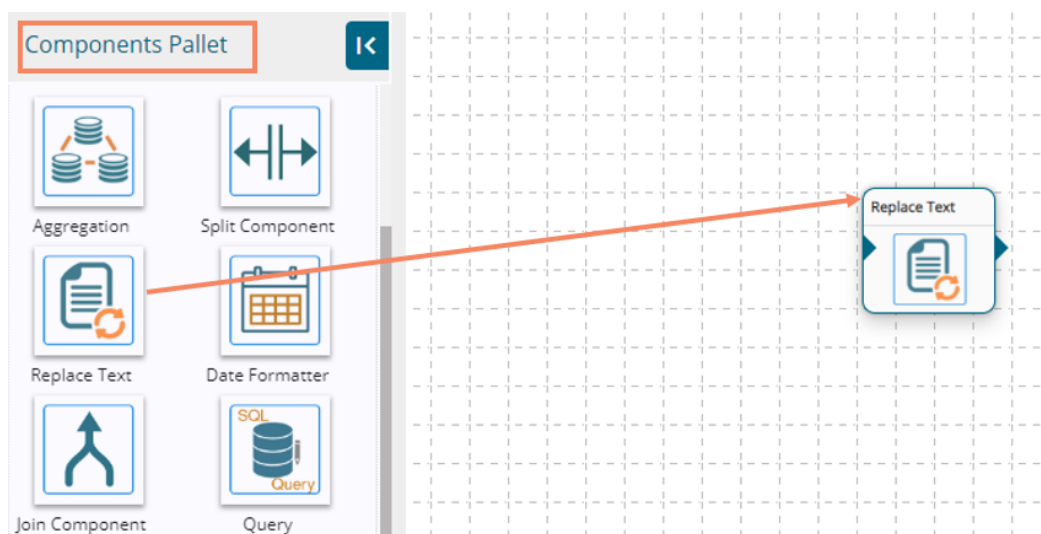
4) Transformation component requires an input event (to get the data) and sends the data to an output event.
5) Create two Events and drag them to the Workspace.
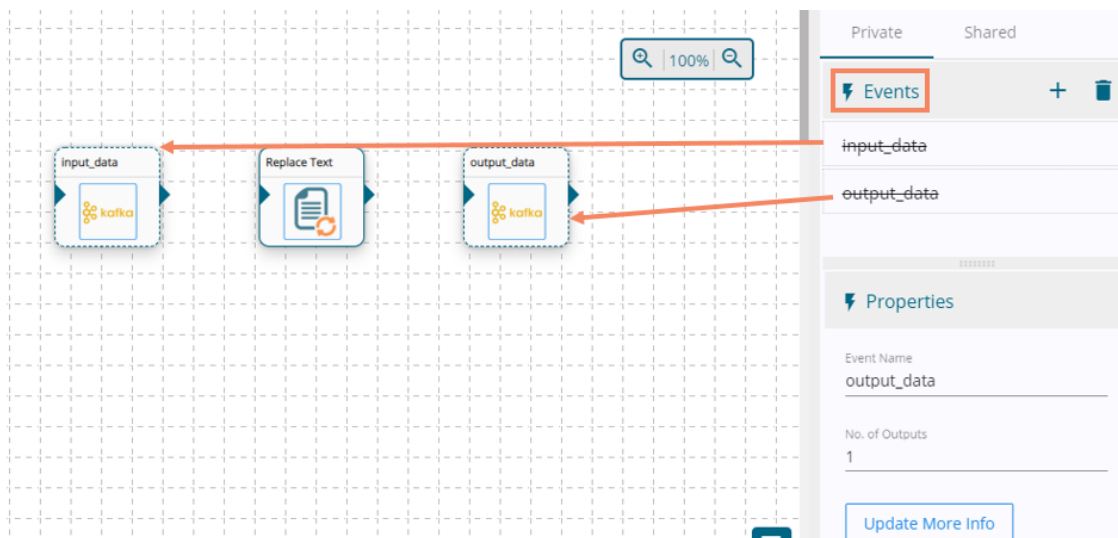
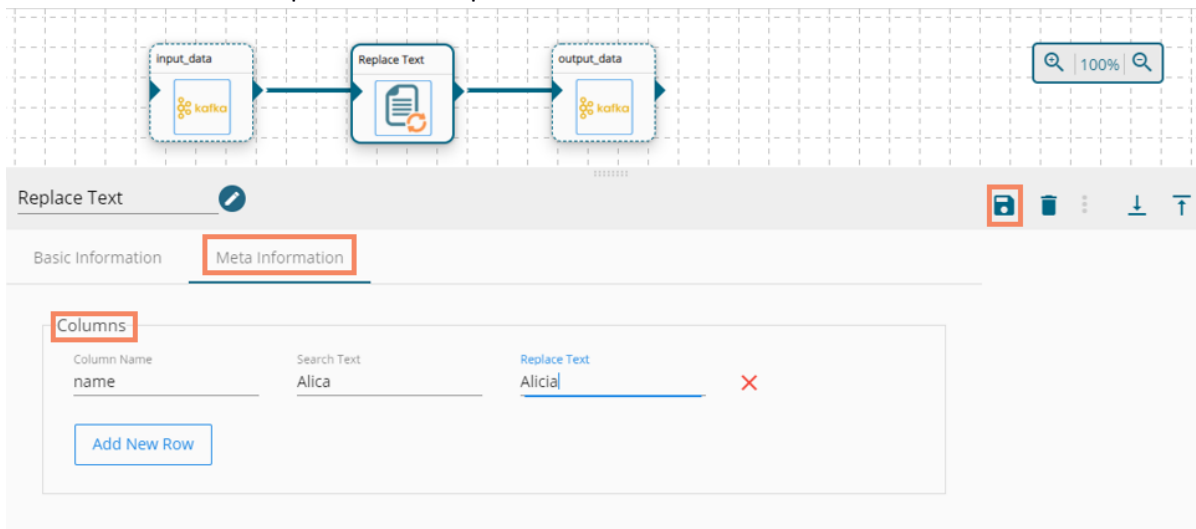
6) Connect the input event and the output event (The data in the input event can come from any Ingestion, Readers or shared events).
7) Click the Replace component to get the configuration tabs.
8) The Basic Information tab opens by default.

a. Select the invocation type (Real-Time/Batch).

Note: Currently, Pipeline supports only Real-Time option.
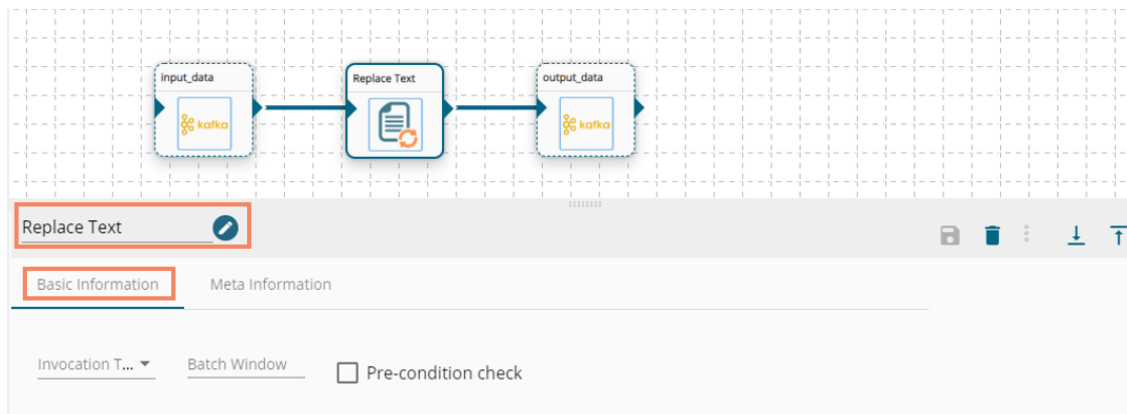
9) Select Meta Information and fill all the required inputs fields.
   a. Column Name-Provide a column name from the input data.
   b. Search Text- The data that you wish to search from the selected column.
   c. Replace Text- The replace text replaces the searched data in the selected column.
   d. To perform multiple Replace Text transformations, click the '**Add New Row**' option and follow the above steps.
   e. Save the Replace Text component.

10) Save the pipeline and activate it. The component reads the data coming from the input event, transform the data and give the output data.
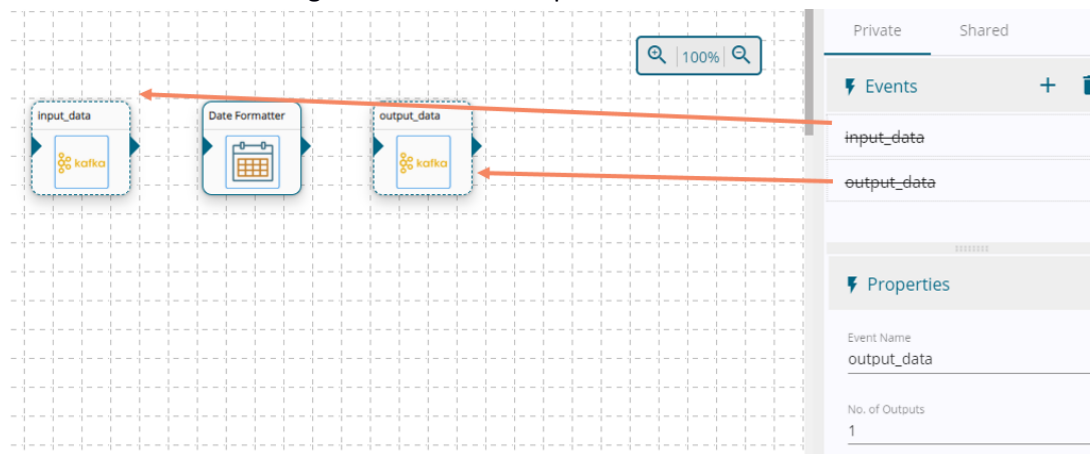
4.   Date Formatter Component

Users can alter the Date format by using this transform component.

1) Navigate to the Pipeline Workflow Editor.

2) Expand the Transformation section using the Components Pallet.

3) Drag and Drop Date Formatter component to the workspace.
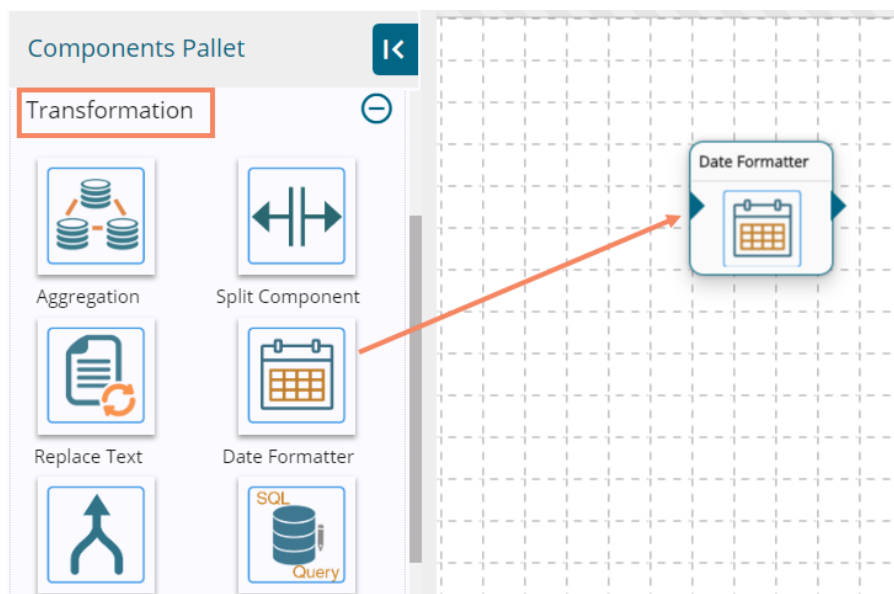
4) Transformation component requires an input event (to get the data) and sends the data to an output event.

5) Create two Events and drag them to the Workspace.

6) Connect the input event and the output event (The data in the input event can come from any Ingestion, Readers or shared events).

7) Click the Date Formatter component to get the configuration tabs.

8) The Basic Information tab opens by default.

   a. Select the invocation type (Real-Time/Batch).

   Note: Currently, Pipeline supports only Real-Time option.
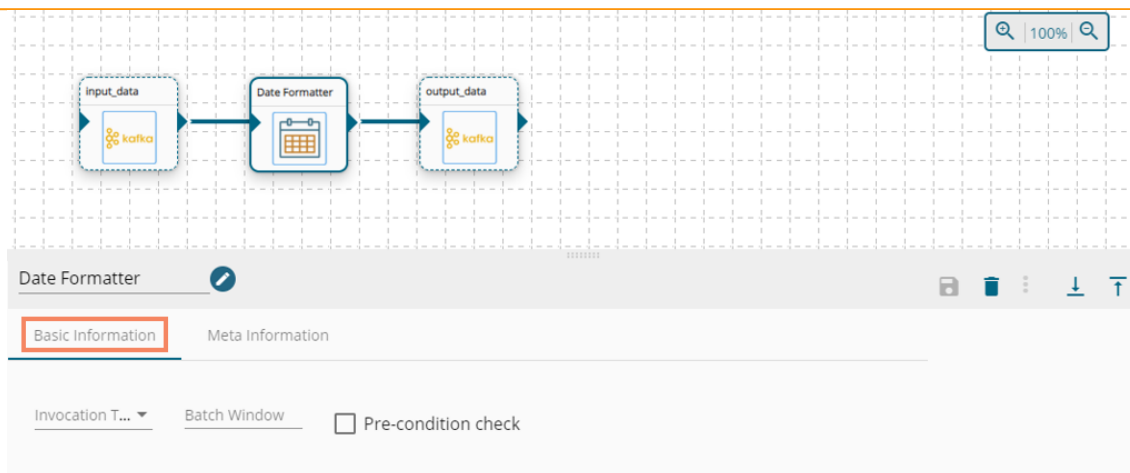
9) Select Meta Information and fill all the required inputs fields.
    a. Fill in the Column Name.
    b. Select the input format.
    c. Select the output format.
        (This component transforms date type for selected input column to provided output date type.)
    d. To perform multiple Date Formatter transformations, click the 'Add New Row' option and follow the above steps.
   e. Save the Date Formatter component.
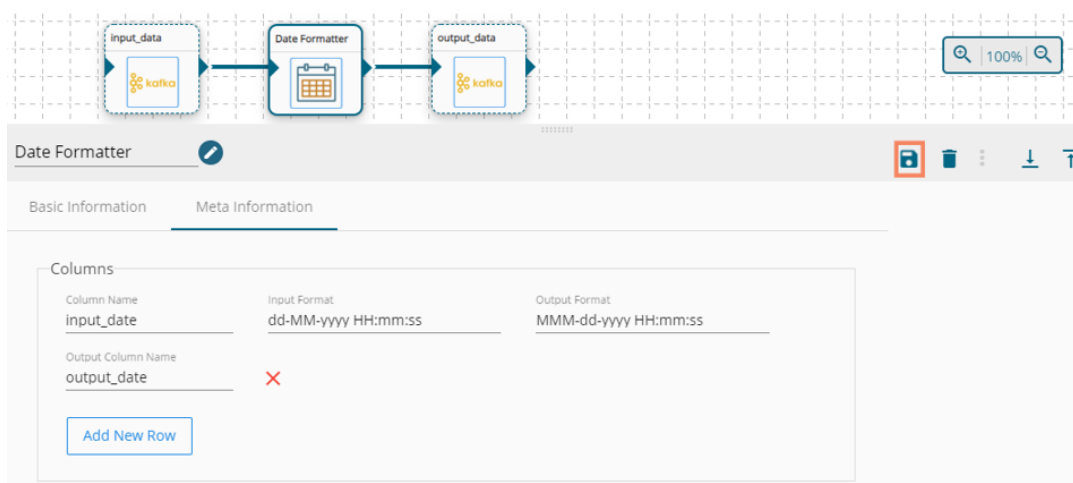
10) Save the pipeline and activate it. The component reads the data coming from input event, transform the data and give the output data with newly created column(s).

5. Join Component

The Join transform allows users to join two or more input data sets as per the user-defined join conditions.

1) Navigate to the Pipeline Workflow Editor.

2) Expand the Transformation section using the Components Pallet.

3) Drag and Drop Join component to the workspace.
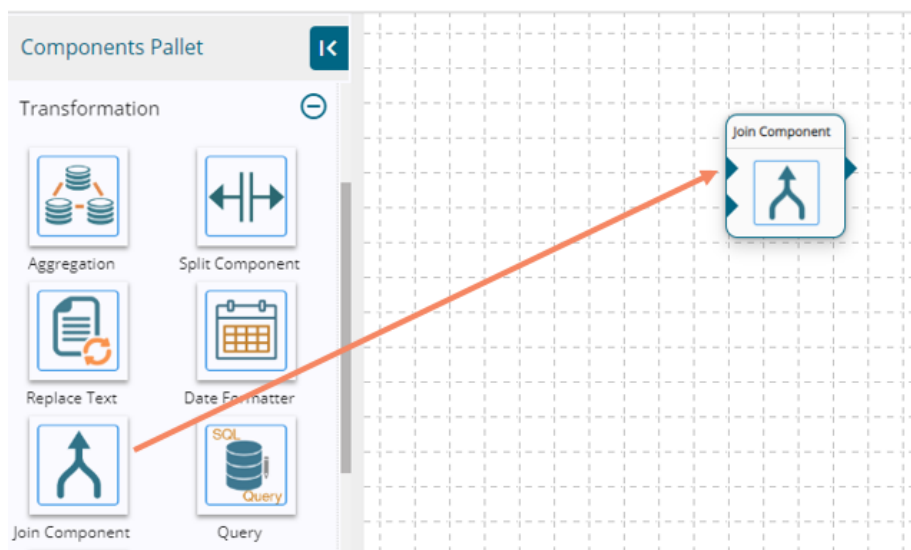
4) This transformation component requires two input data sources and sends the data to an output event.

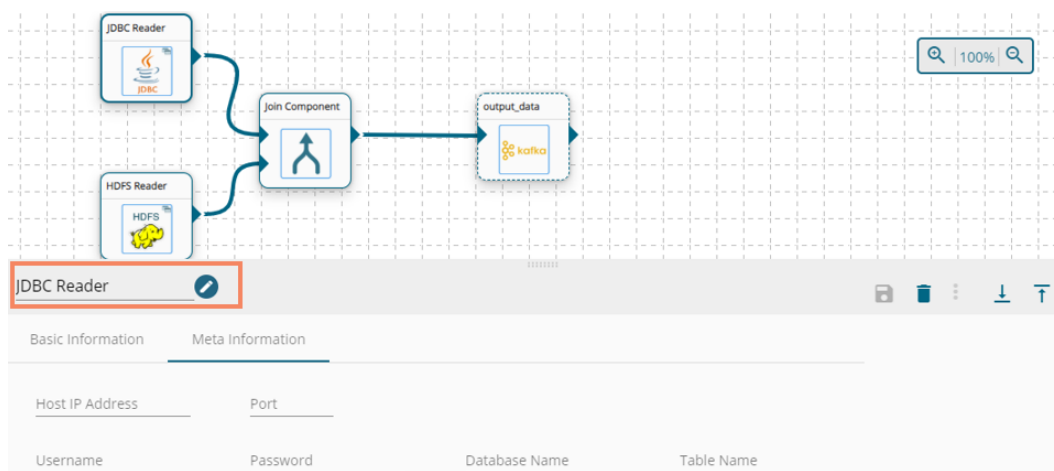5) Create two input data sources and one output Event and drag them to the workspace.

6) Connect the input data sources and the output event. The data in input data sources can select only from Readers.
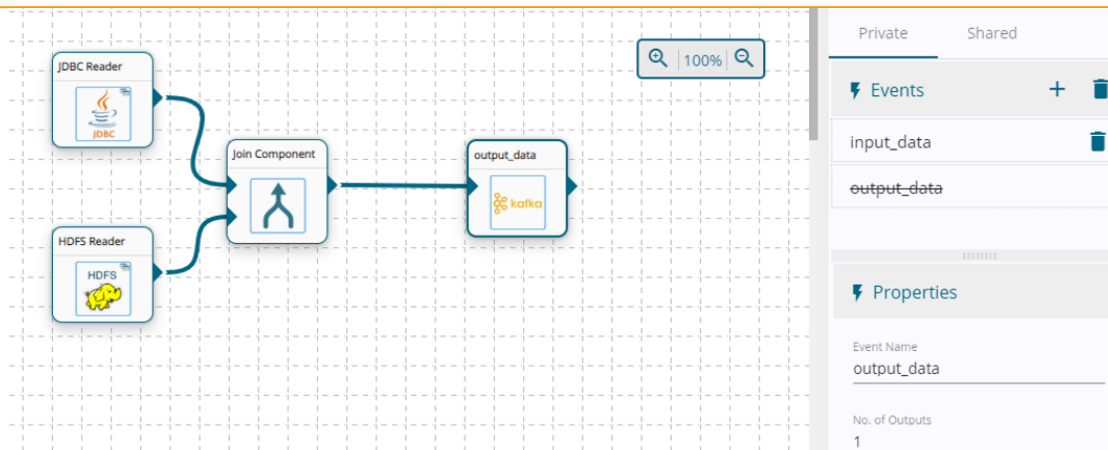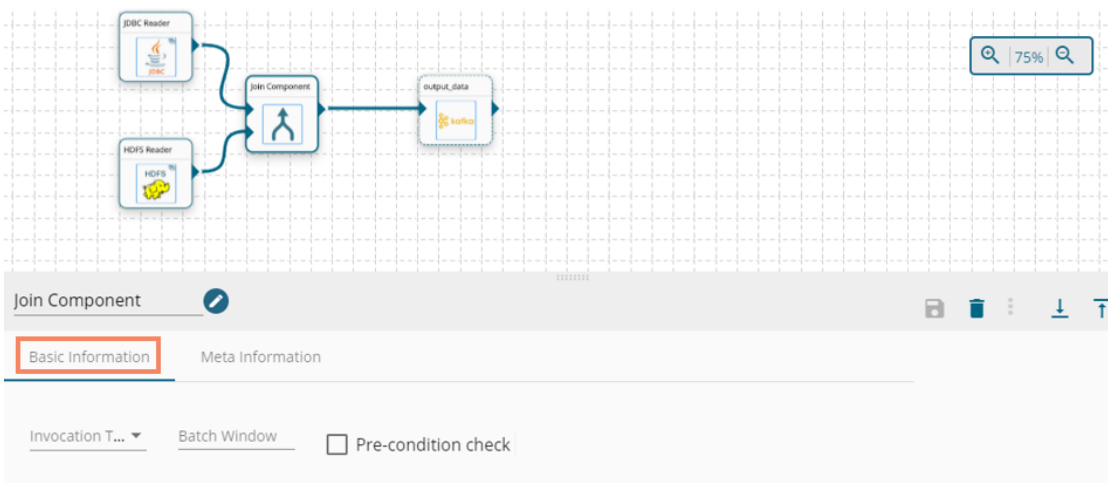
7) Configure the Input Data sources.

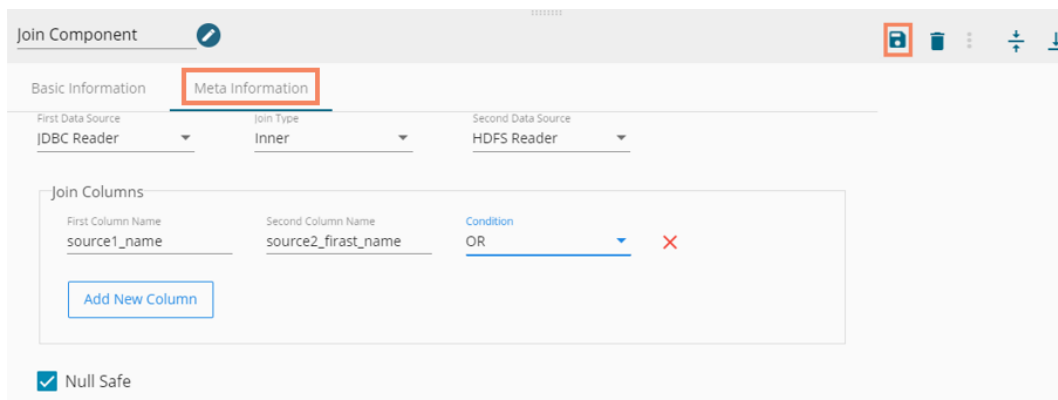Note: Refer to the Reader Component document for more information.

8) Click the Join component to get the configuration fields.
9) The Basic Information tab opens by default.
   a. Select the invocation type (Real-Time/Batch)

Note: Currently, Pipeline supports only Real-Time option.

10) Select Meta Information and fill all the required inputs fields.
   a. Select the first data source.
   b. Select a Join Type from the drop-down menu.
   c. Select the Second Data Source.
   d. Configure the Join Columns- Enter the First Column Name, Second Column Name, and select a Condition from the drop-down menu.
      Note: The Condition field can be ignored for the single join.
   e. To use multiple joins, click on the '**Add New Column**' option.
   f. Use a check mark in the box beside the 'Null Safe' option to include the 'null' values for join columns.
   g. Save the Join component.

11) Save the pipeline and activate it. The component reads the data coming from input data sources and returns the output data.
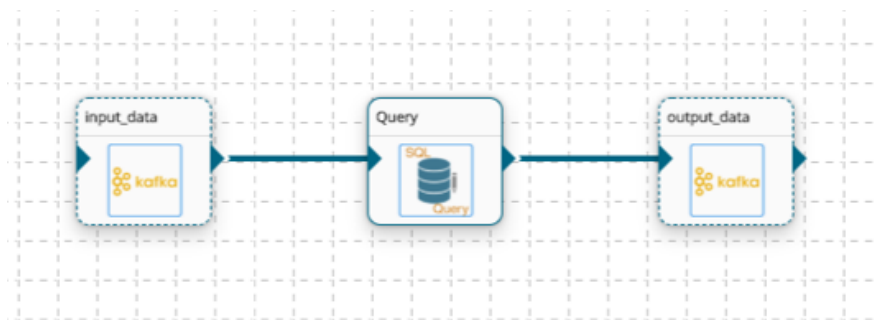
6. Query Component

This component helps users to get data as per entered query.

1) Navigate to the Pipeline Workflow Editor.
2) Expand the Transformation section using the Components Pallet.
3) Drag and Drop Query component to the workspace.
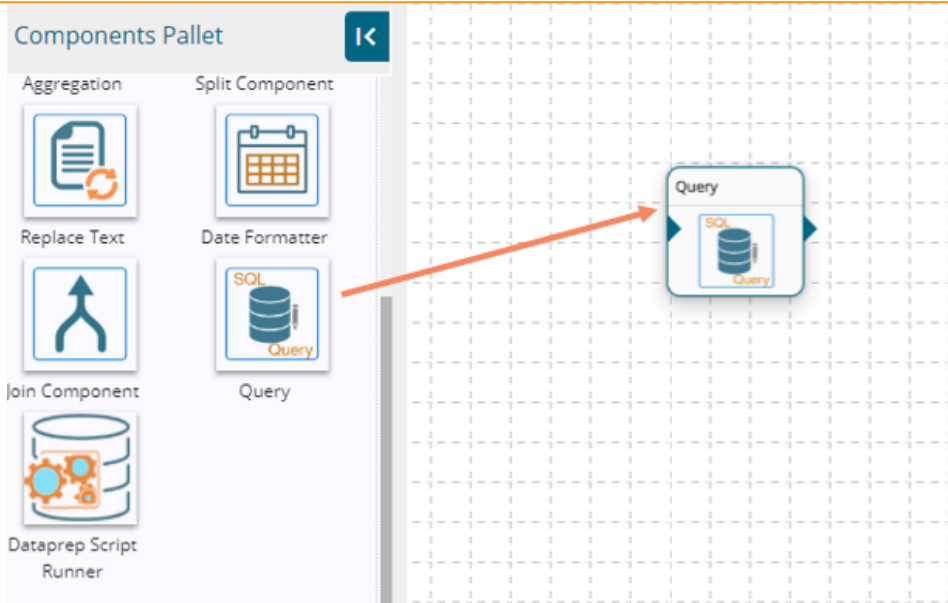
4) Transformation component requires an input event (to get the data) and sends the data to an output event.

5) Create two Events and drag them to the Workspace.

6) Connect the input event and the output event (The data in the input event can come from any Ingestion, Readers or shared events).


7) Click the Query component to get the configuration tabs.

8) The Basic Information tab opens by default.

   a. Select the invocation type (Real-Time/Batch).

   Note: Currently, Pipeline supports only Real-Time option.
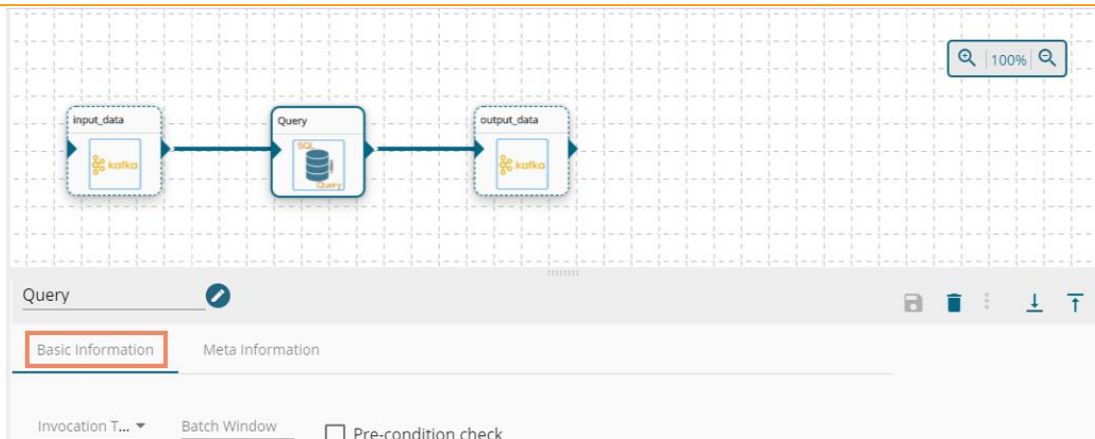
9) Select Meta Information and fill all the required inputs fields.
   a. Enter a valid data query to fetch data.
   b. Provide the Table Name.
   c. Selected Columns- This option helps for schema conversion. Provide Name, Alias name and Column type for the column from which you wish to select the data.
   d. Save the Query component.
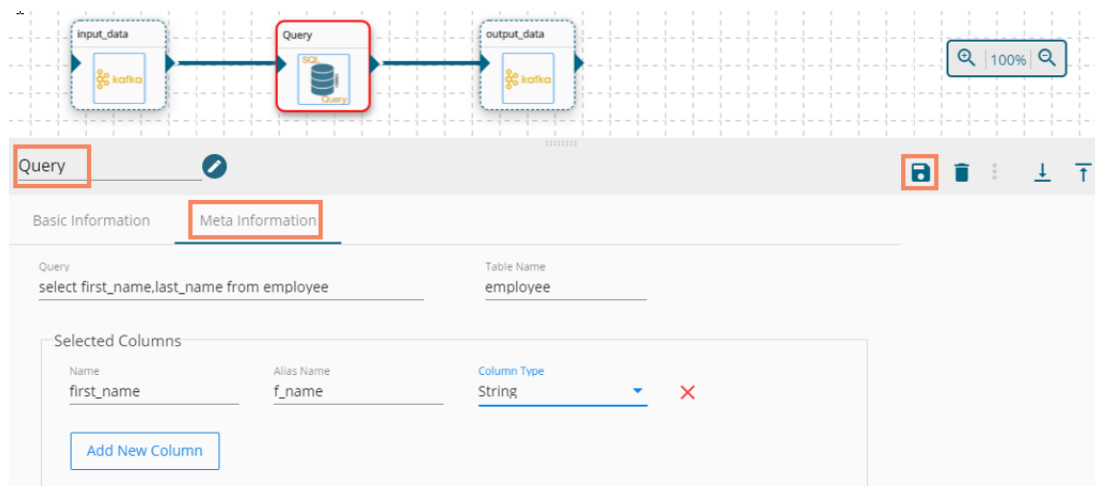
10) Save the pipeline and activate it. The component reads the data coming from the input event, transforms the data and returns output data.
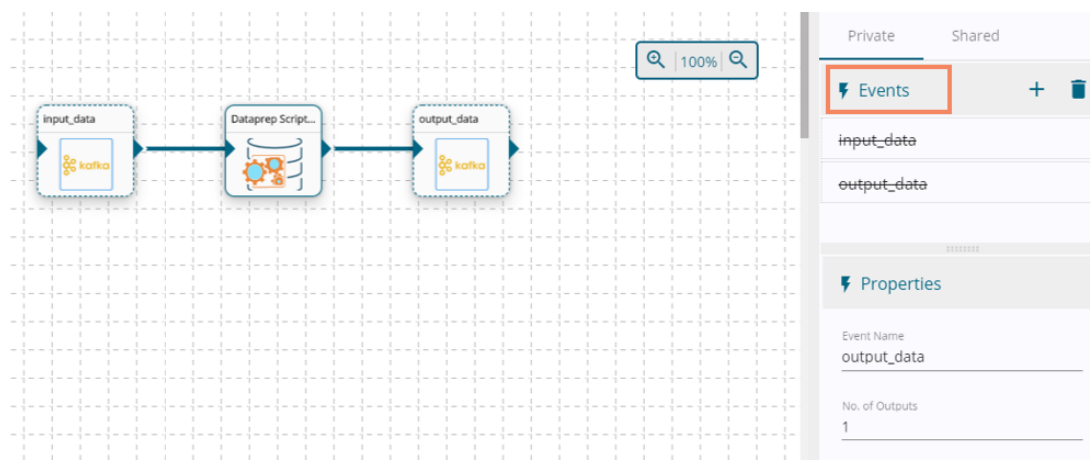
7. Dataprep Script Runner Component

This transform component enables the users to use the exported Dataprep scripts in the Data Pipeline.

1) Navigate to the Pipeline Workflow Editor.
2) Expand the Transformation section using the Components Pallet.
3) Drag and Drop Dataprep Script Runner component to the workspace.
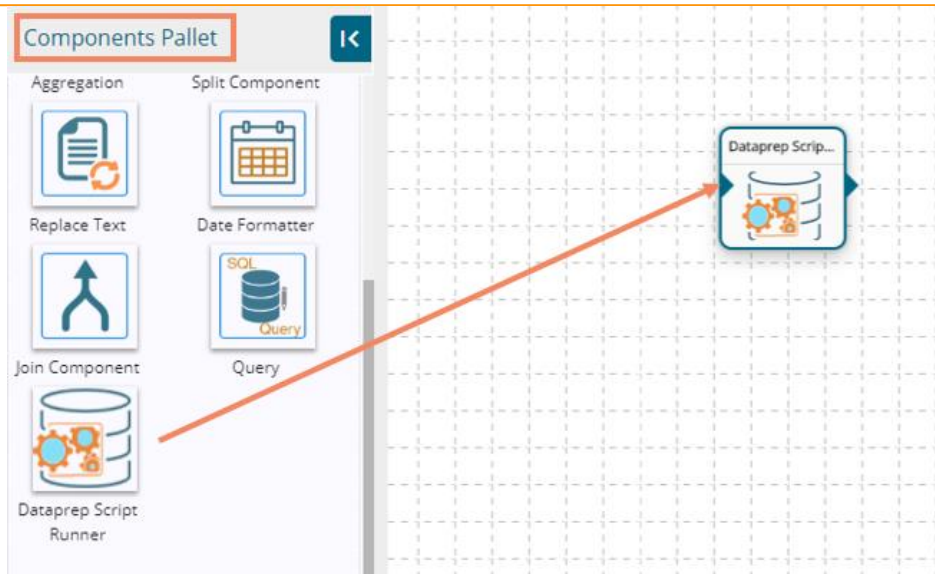
4) Transformation component requires an input event (to get the data) and sends the processed data to an output event. So, create two events and drag and drop to workflow

5) Create two Events and drag them to the Workspace.

6) Connect the input event and the output event. (The data in input event can come from any Ingestion, Readers or shared events)
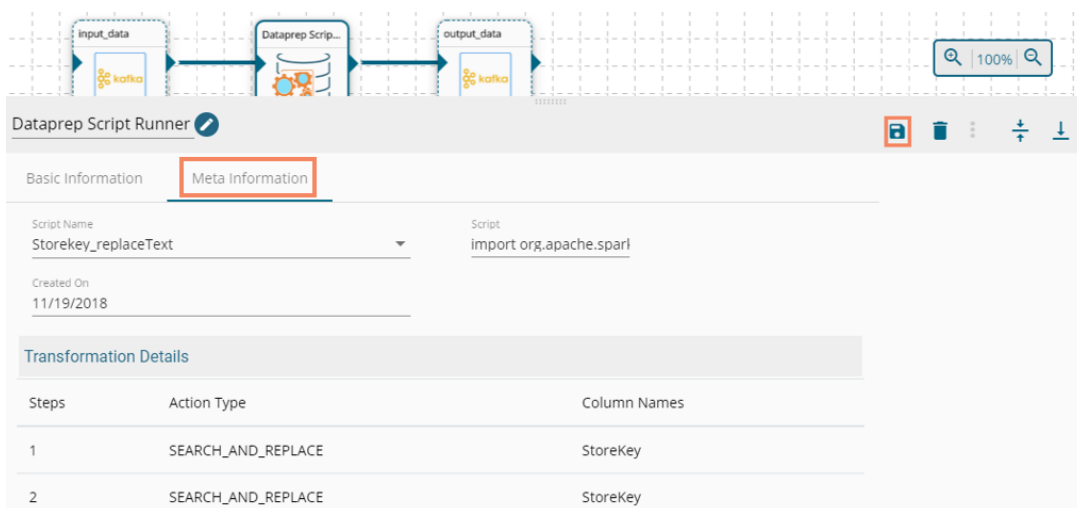

7) Click the Dataprep Script Runner component to get the configuration tabs.

8) The Basic Information tab opens by default.

    a. Select the invocation type (Real Time/Batch)

Note: Currently, Pipeline supports only Real-time option.
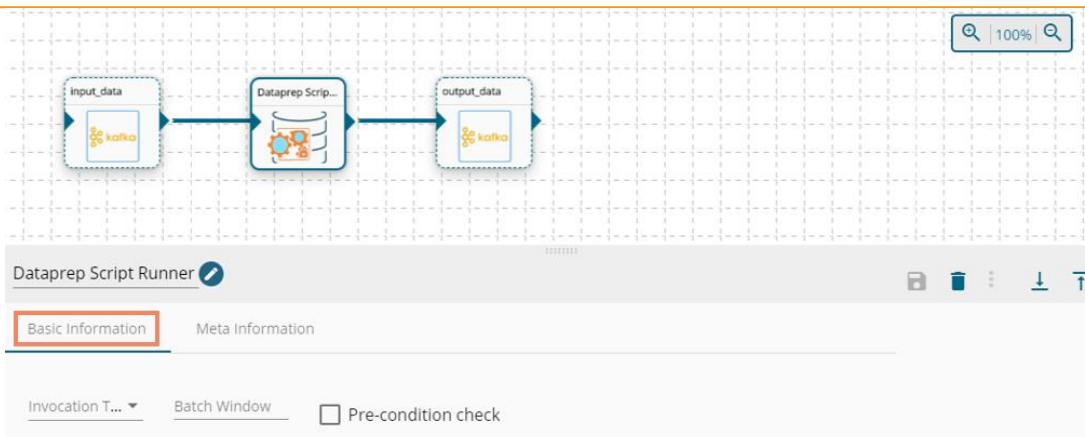
9) Select Meta Information and fill all the required inputs fields.
   a. Select a script which is exported from Dataprep Application.
   b. Save the component.

10) Save the pipeline and activate it. The component reads the data coming from the input event, transform the data and returns output data.