How to Guide

Create and Update a Data Pipeline

Version: Release 2.2

Contents

1. Introducing the BDB Data Pipeline

Discover the real power of Big Data Analytics with BDB Pipeline. Gain seamless insights into a massive amount of structured, semi-structured, and unstructured data. Support your decisions with advanced Machine Learning Algorithms and Visualization techniques, all in one go.

2. Getting Started with the BDB Data Pipeline

1. Browse the BDB Enterprise URL link: https://app.bdb.ai
2. The BDB Login page opens.
3. Login to the BDB Platform.
   a. Enter user-specific credentials (Email ID with valid Password).
   b. Select an authentication type using the drop-down.
   c. Click the 'Continue' option.
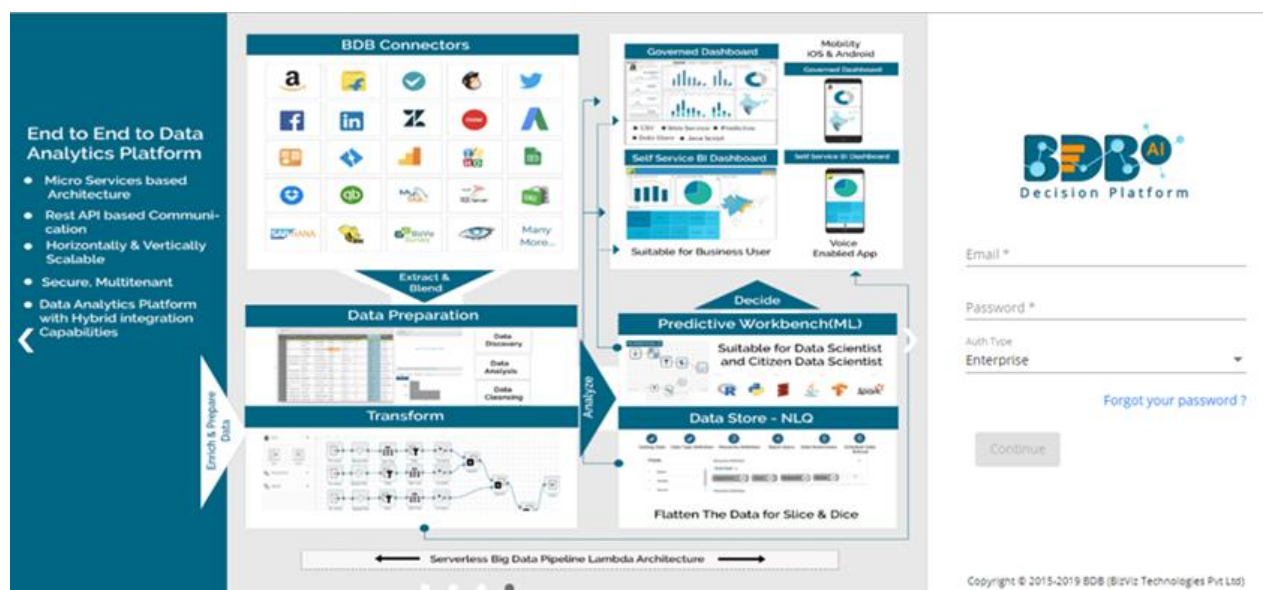
4. The BDB Platform homepage opens.

   Note: Make sure that the user has permission to access the Data Pipeline plugin.

5. Click the Apps menu.
6. Select the Data Pipeline Plugin.
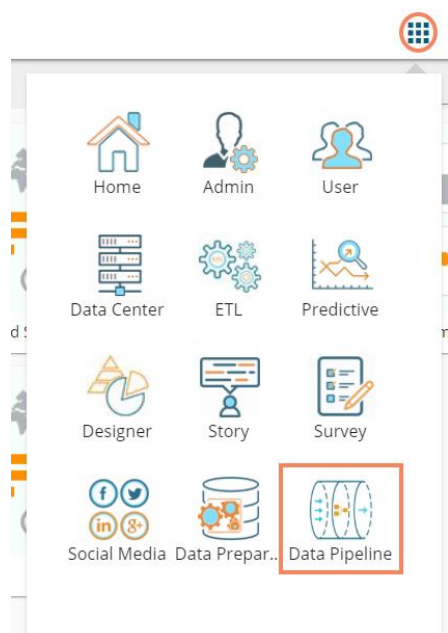
7. Data Pipeline landing page opens displaying the Pipeline list in a new window.

# 3. Creating a New Data Pipeline

1. Navigate to the Data Pipeline landing page.

2. Click the 'Create New Pipeline' icon provided on the top right side of the page.
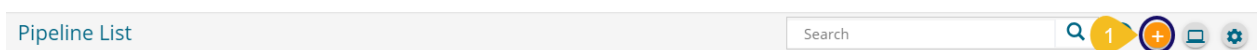
3. Pipeline Editor page opens with a pop-up window to create a new Pipeline.
4. Provide the following information to create a new pipeline:
   a. Enter Name: Enter a name for the new Pipeline
   b. Description: Describe the Pipeline (Optional)

c. Select a resource allocation option using the radio button- the given choices are:
   i. Low
   ii. Medium
   iii. High
d. Click the '**Save**' option.
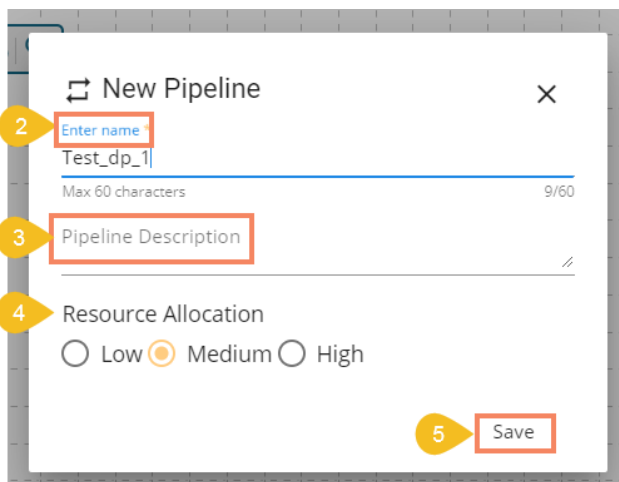
5. The Pipeline Editor page opens for the newly created pipeline.
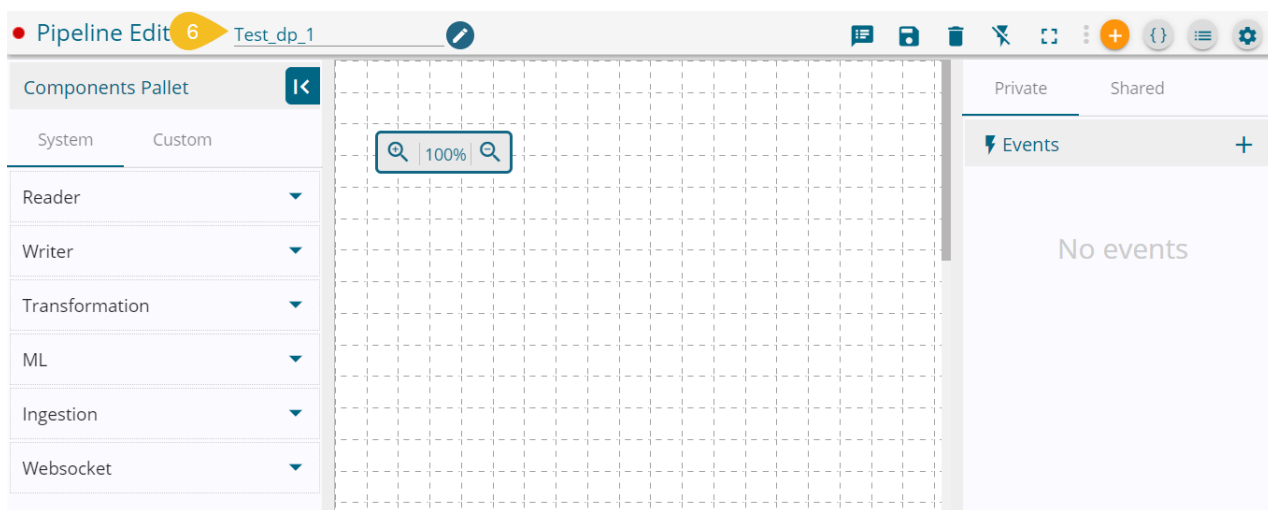
The icons and options provided on the Pipeline Editor page

| Icon/Option	Name	Function
	Open/ Close the Components Pallet	Opens or Closes the Component Pallet
	Edit the Pipeline Name	Allows to modify or change the given Pipeline name
	Toggle Log Panel	Displays the background activities
	Update Pipeline	Saves and updates the pipeline
	Delete Pipeline	Deletes the pipeline
&	Toggle Event Panel	Opens the event panel to create and display events

	Full Screen	Displays the present page in full screen
	Create New Pipeline	Redirects the users to create a new pipeline
	Pipeline JSON Editor	Redirects the users to the JSON Editor Page
	List Pipeline	Redirects the users to the page listing all the existing pipelines
	Pipeline Workflow	Redirects the users to the Pipeline Workflow
	Settings	Redirects the users to the Pipeline Settings page

Note: Since in the above images the user is already on the Pipeline Workflow page, so the Pipeline Workflow icon is displayed in the header.

6. Drag and drop the required components to the workspace. (E.g., JDBC Reader is dragged to the workspace in the below given image)
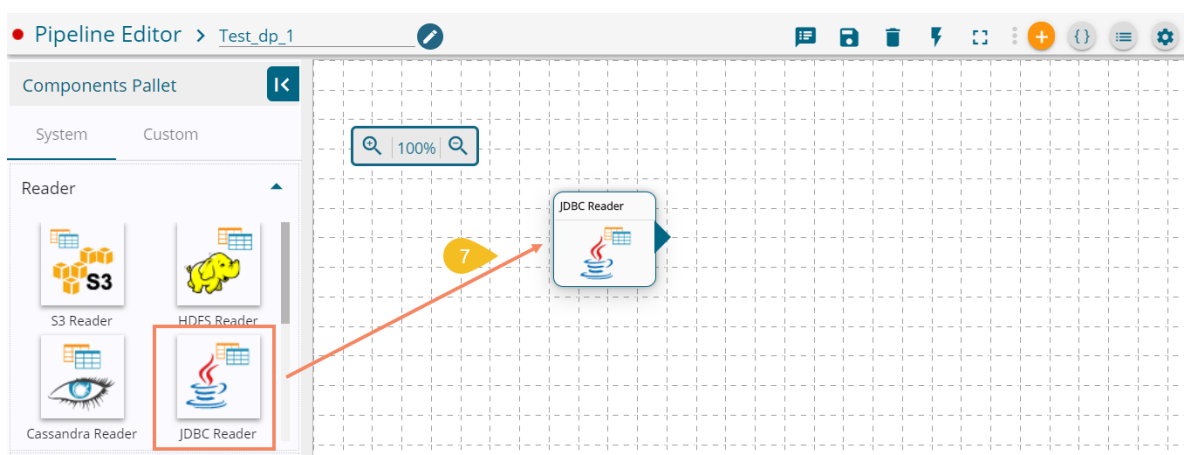
7. Click on the component
8. Provide the Basic Information about the dragged component:
   a. Select an Invocation Type using the drop-down menu.
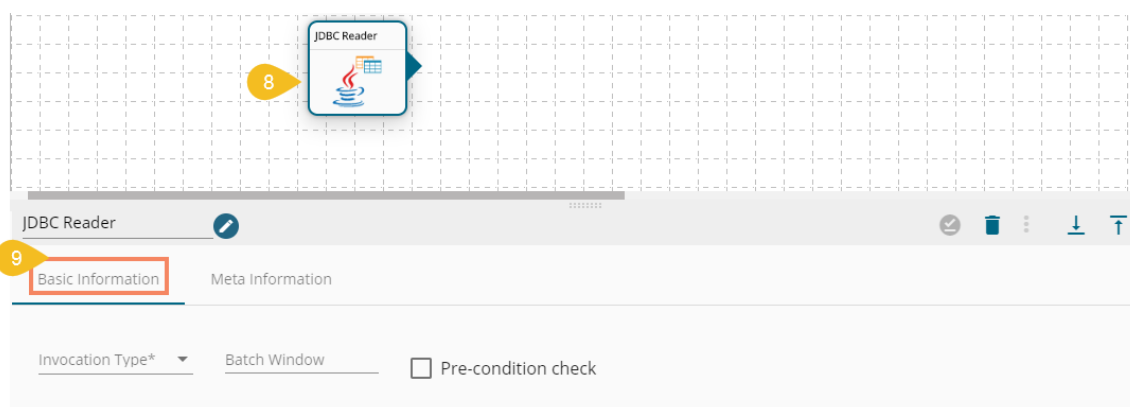
9. Open the Meta Information tab and configure the following information:
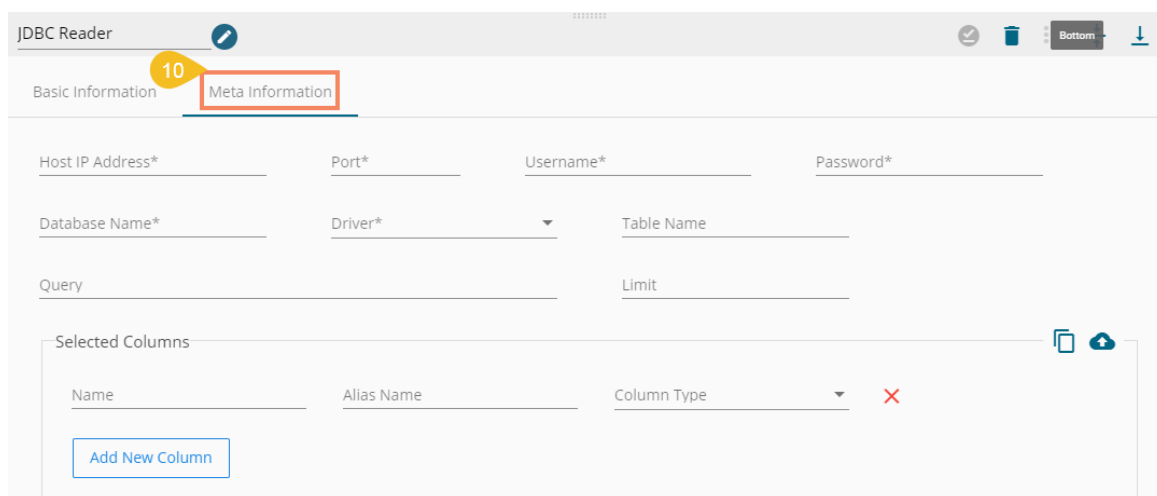   i.   Host IP Address
   ii. Port number

     iii.   Username

     iv.   Password

     v.   Database Name

     vi.   Driver- Select from the drop-down menu

     vii.   Table Name

     viii.   Query

     ix.   Limit

     x.   Selected Columns: Fill the following details manually to select the required columns

          1.   Name

          2.   Alias Name

   3.   Column Type

          Or

         Use '**Download Data**' and '**Upload File**' options to select the desired columns.

         1.   Upload File: Users can upload the existing system files (CSV, EXCEL) using the '**Upload File**' ☁ icon.

         2.   Download Data (Schema): Users can download the schema structure in JSON format by using the '**Download Data**' ⧉ icon.

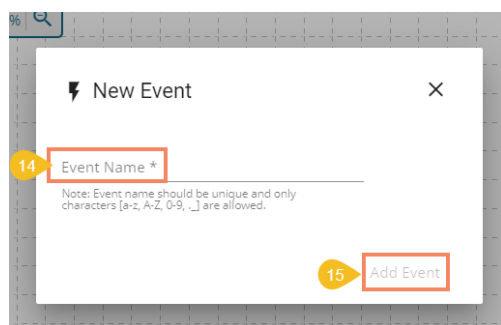   Note: The users can also upload the downloaded JSON file to another component using the '**Upload File**' icon.

10. Click the '**Toggle Event Panel**' icon
11. The Event Panel opens on the right side of the page
12. Click the '**Add New Event**' icon
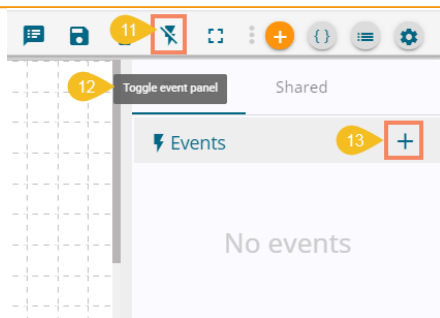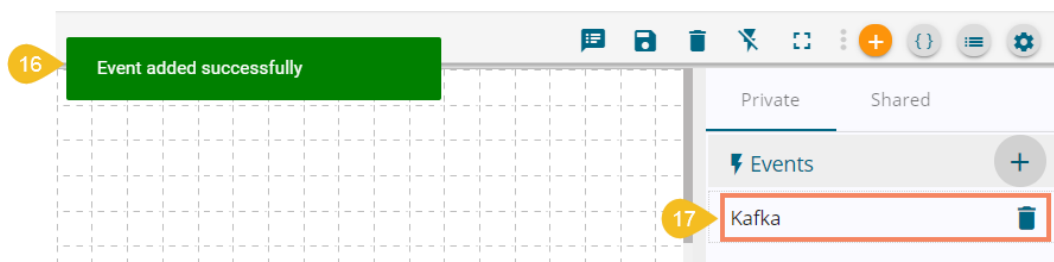
13. The '**New Event**' window opens
   a.   Provide Event Name
    b.   Click the 'Add Event' option

Note: Event name should be unique, and only characters [a-z, A-Z, 0-9, _] are allowed.

14. A success message appears, and the new event gets added to the event panel

15. Drag the event onto the workspace and Connect it with the dragged component
    (To connect clicking output node of the dragged component and input node of the event)

16. Click the '**Update Pipeline**' 🖫 icon to save the created pipeline
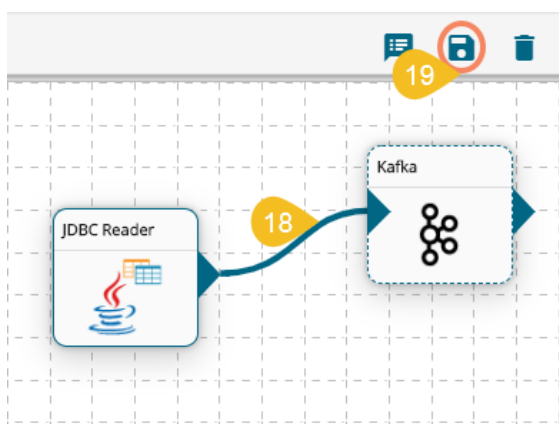
17. A success message appears to assure that the new pipeline has been successfully updated.
18. Click the '**Activate Pipeline**' ⊙ icon to activate the pipeline (It appears only after the newly created pipeline gets successfully updated).
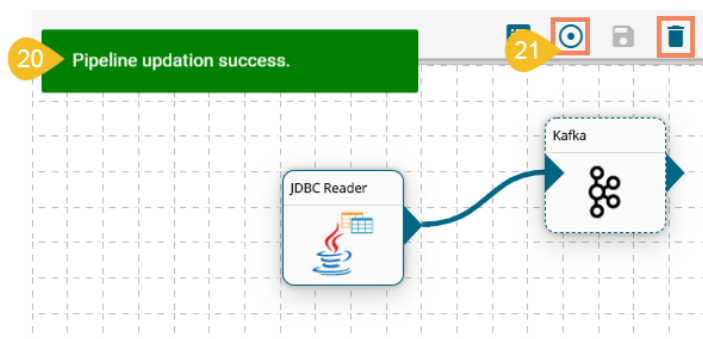
19. The newly created pipeline appears at the top of the Pipeline List.
20. The status of the pipeline is indicated through color coding under the 'Status' column.
    a. Green-Activated Pipeline
    b. Red- Deactivated Pipeline
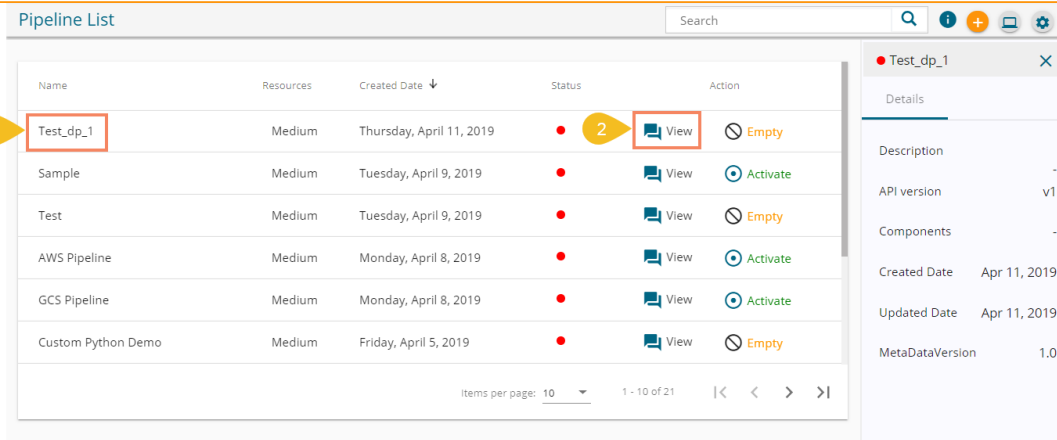

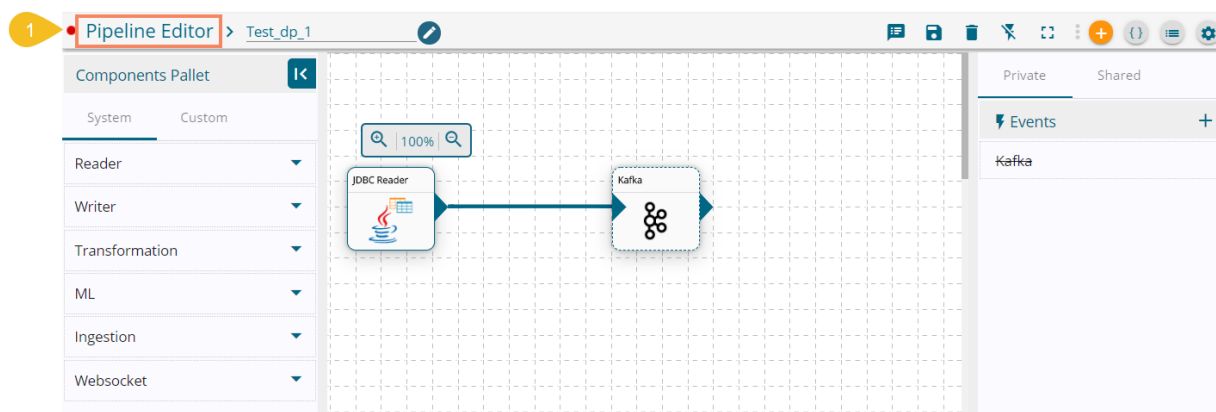
Note: The details of the selected pipeline from the Pipeline List appears on the right side of the page.

4. Updating a Data Pipeline

1. Navigate to Pipeline List using the Main Menu option.
2. Select the pipeline you want to update from the list and click '**View**' icon.
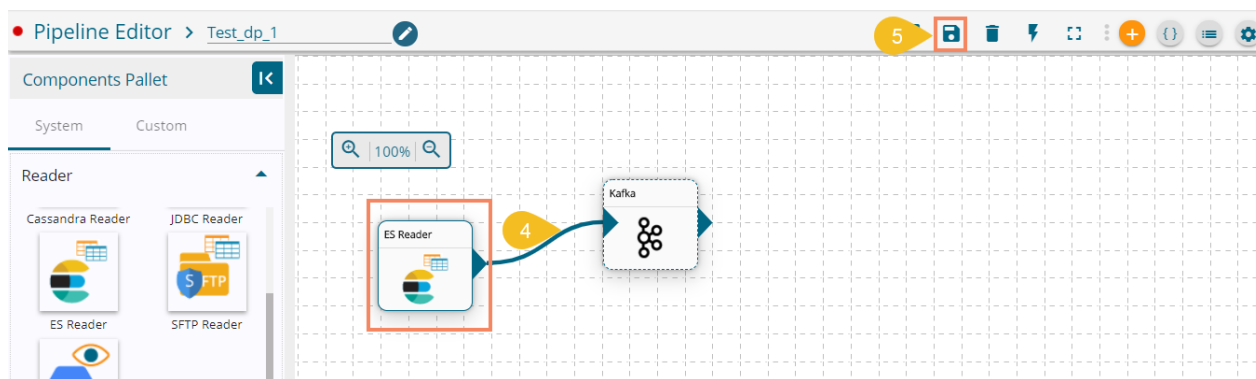
3. The workflow screen (Pipeline Editor) re-opens for the selected pipeline.

4. The user can modify or remove the existing components, drag new components to the selected pipeline.
   E.g., In the below given image, the ES Reader component is connected to the Kafka event, instead of the HDFS Reader.

5. Click the '**Update Pipeline'** icon to update the pipeline.

6. The recent change gets saved for the selected Pipeline.

Note: Click the '**Delete**' icon from the Pipeline Editor page to delete the selected pipeline. The deleted Pipeline gets removed from the Pipeline list.

5. Signing Out

The users can Sign-out from the Data Pipeline tab at any given stage, but preferable is that the users should complete all the tasks they wish to perform and save it before closing the tab or singing out from the Platform.

The Signing Out process for the Data Preparation has two steps:

1. Closing the BDB Data Pipeline

Once you have completed the Data Pipeline tasks, save your work and close the Data Pipeline tab.

i)   Click the **'Close'** button (the 'X' on the right edge) from the Data Pipeline tab.
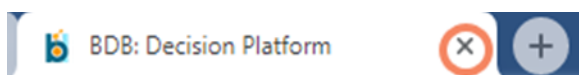

ii) A window appears to confirm the action of closing the BDB Pipeline.
iii) Click the '**Leave**' option to close the selected pipeline.
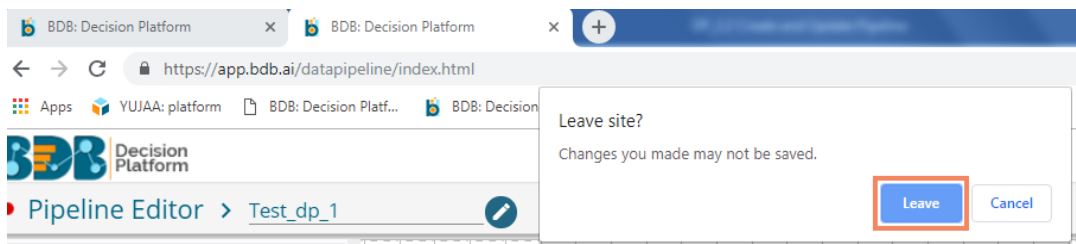
2. Sign Out from the BDB Platform

i)   After closing the Data Pipeline plugin, the users get redirected to the Platform homepage.

ii) Click the '**User**' icon on the Platform homepage.
iii) A menu appears with the logged in user details (User's name and email id).
iv) Click '**Sign Out**.'
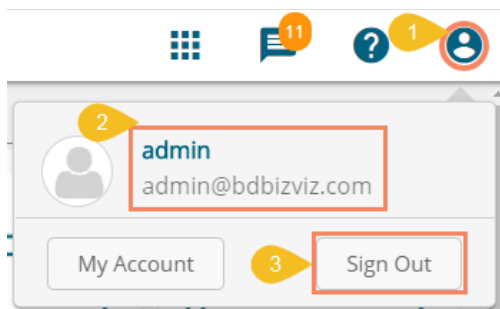
v)   The user successfully signs off from the **BDB Platform**.

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

Strictly Confidential