How to Guide

Data Ingestion(Data Pipeline)

Version: Release 1.1
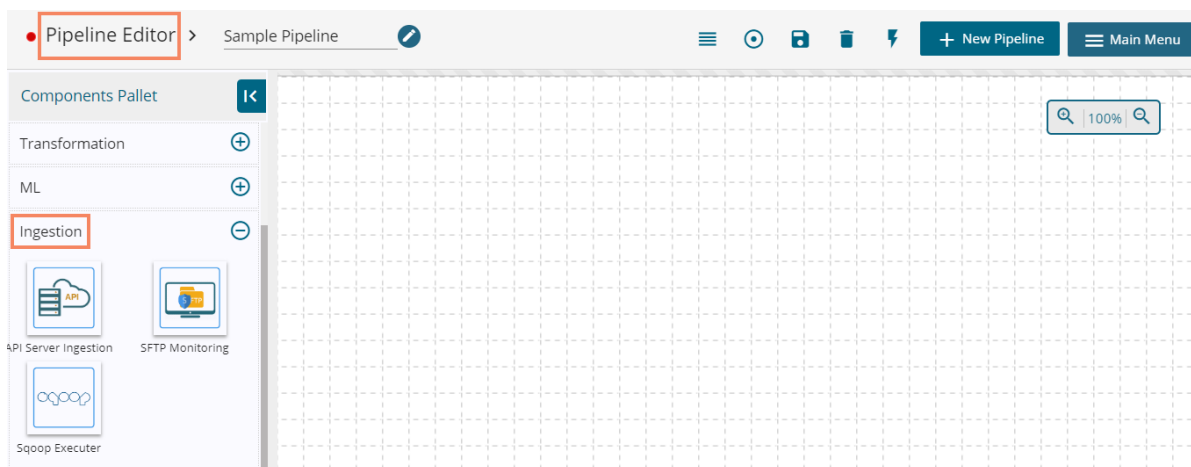
Contents

Component Description-

Ingestion components allow the users to ingest data in the pipeline from outside the pipeline as in files from some SFTP location or manual ingestion of data using a service in real-time.

1. API Server Ingestion

1) Navigate to the **Pipeline (Workflow) Editor** and expand the **Ingestion** section using the **Components Pallet**

2) Drag and Drop the **API Server Ingestion** component to the workflow
3) Click on the dragged ingestion component to get the configuration tabs
4) Configure the **Basic Information** tab
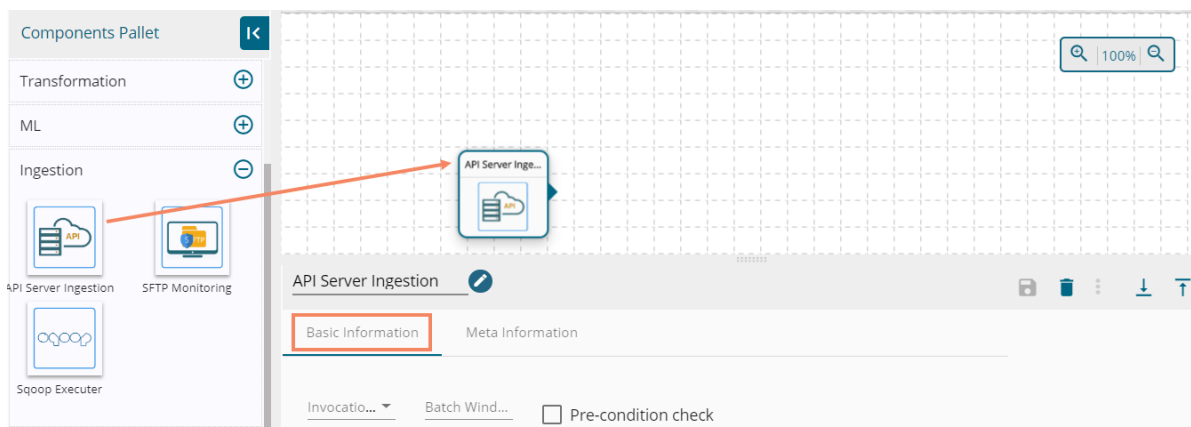
5) Navigate to the **Settings** page of the Data Pipeline plugin using the Main Menu
6) Click the **List Ingestions** option to open the list of ingestions
7) Select an Ingestion from the list
8) The **Ingestion Details** opens on the top-right side of the page with **Ingestion ID** and **Ingestion Secret** (Key)
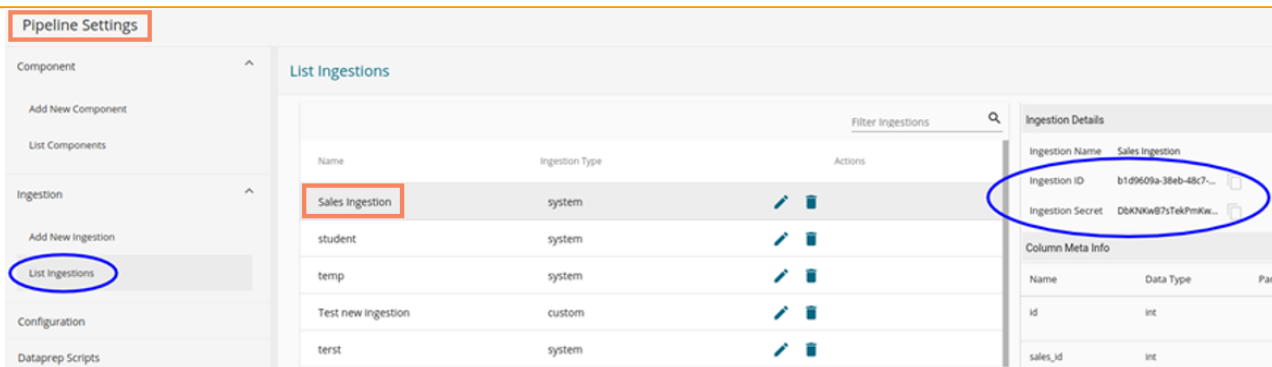
9) Fill the **Ingestion ID** and **Ingestion Secret** (Key) the **Meta Information** tab of the API Ingestion component

10) Save the API Ingestion component

11) Connect the API Ingestion component to an Event and save the Pipeline.

12) As we save the pipeline, an auto-generated **Component Ingestion URL** appears inside the Meta Information tab to ingest the data.

13) The users can use the Component Ingestion URL with the following format in the program or add anywhere in the third-party portal.
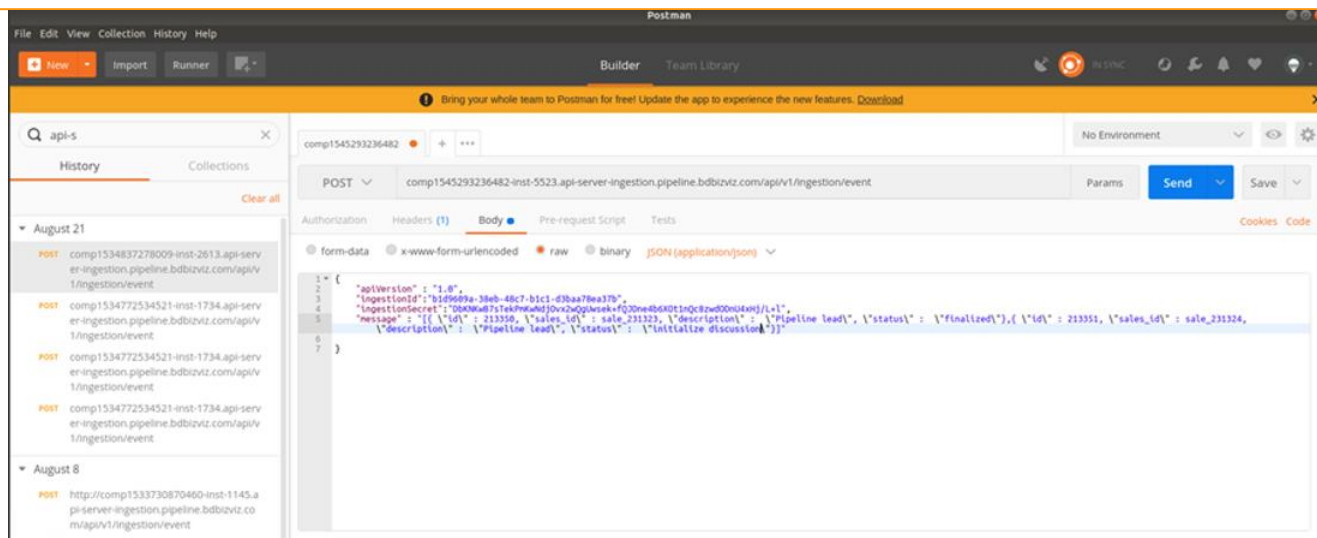
Service Info : -

- ingestionID and ingestionSecret must same as fill in a pipeline component
- message accepting data in Array as Stringify format
- In the message, all the object must follow the same column info as declared in settings page at the time ingestion create.

URL → comp1545293236482-inst-5523.api-server-ingestion.pipeline.bdbizviz.com/api/v1/ingestion/event (will be Auto-generated Component Ingestion URL)

Type → POST

Content-Type → application/json

Body →

```
{
 "apiVersion": "1.0",
 "ingestionId": "b1d9609a-38eb-48c7-b1c1-d3baa78ea37b",
 "ingestionSecret": "DbKNKwB7sTekPmKwNdjOvx2wQgUwsek+fQJDne4b6XOt1nQc8zwdODnU4xHj/L+l",
 "message": "[{ \"id\" : 213350, \"sales_id\" : sale_231323, \"description\" :  \"Pipeline lead\", \"status\" :  \"finalized\"},{ \"id\" : 213351, \"sales_id\" : sale_231324, \"description\" :  \"Pipeline lead\", \"status\" :  \"initialize discussion\"}]"
}
```

2. SFTP Monitor

1. Navigate to the Pipeline Workflow Editor and expand the Ingestion section provided in the Components Pallet
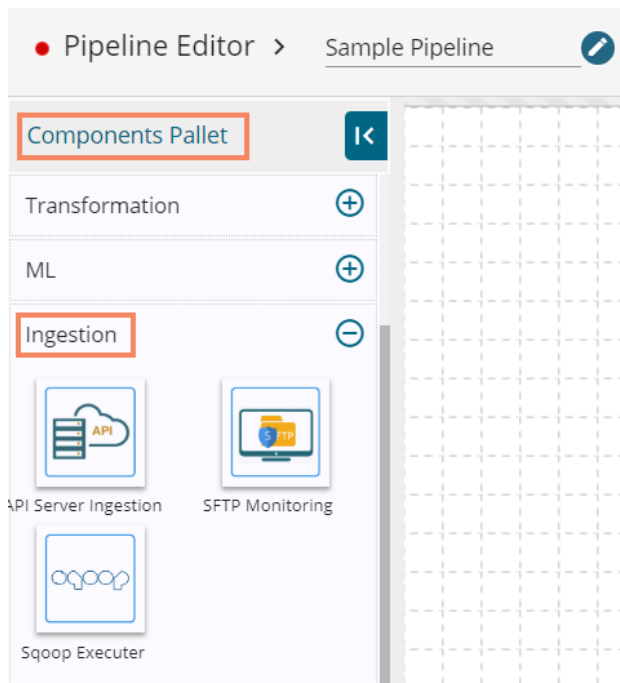

2. Drag and Drop the SFTP Monitoring ingestion component to the workspace
3. Click the dragged ingestion component to get the configuration tabs
4. Configure the Basic Information Tab
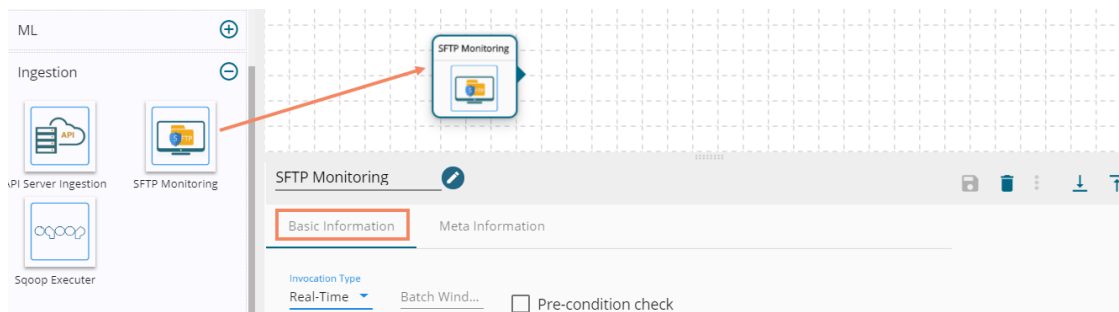
5. Configure the **Meta Information** tab for the dragged SFTP Monitoring component
6. Authentication type can be PEM/PPK or Password if the password needs to fill password else need to upload required PEM/PPK file (E.g., the following example displays configuration using password authentication.)
7. Save the SFTP Monitoring component

8. SFTP Monitoring component requires an Event to send output.
9. Create and connect to an output Event
10. Save and activate the Pipeline.
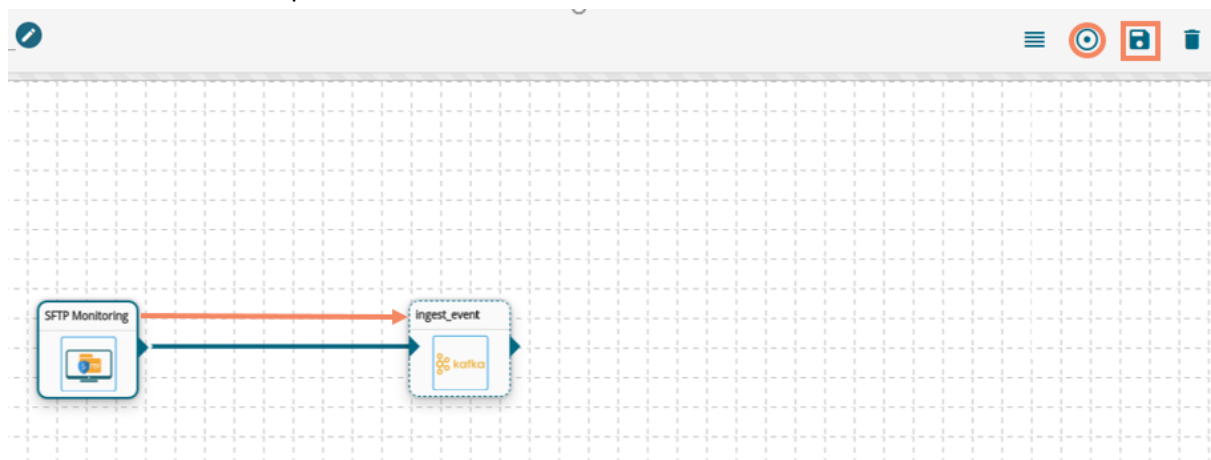
11. The SFTP Monitoring component monitors the file coming to the monitored path and copies the file in Copy Path location for SFTP Reader to read.

3. Sqoop Executor

Sqoop component in BDB Pipeline helps users to transfer tables and databases from RDBMS (supported by Sqoop) to HDFS (Hadoop System Distributed File).

1) Navigate to the Pipeline Workflow Editor page.
2) Expand the Ingestion section using the Components Pallet.
3) Drag the Sqoop component to the workspace.
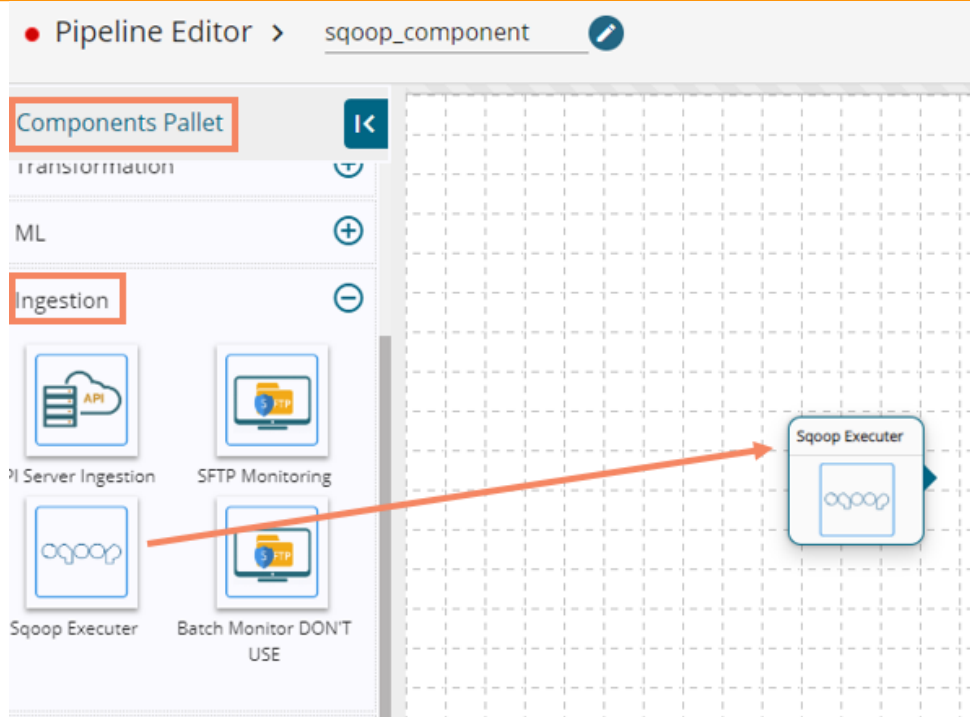
4) Configure the Sqoop component by providing the required Basic and Meta information.
5) The Basic Information tab opens by default by clicking the Sqoop Executer component.
   a. Select an Invocation Type from the drop-down menu (Real-Time/Batch).
   b. Use a check mark in the 'Is Schedule?' option, if it is scheduled.

   Note: Currently, Pipeline supports only Real-Time invocation option.
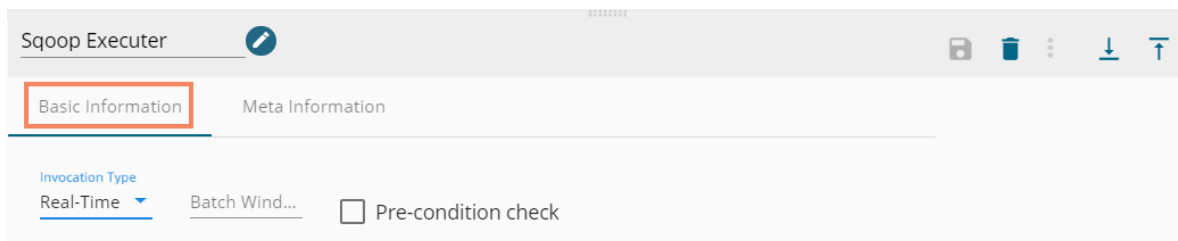
6) Open the Meta Information tab and provide the following information:
   a. Username
   b. Host IP address
   c. Port number (the default number for this fields is 22)
   d. Authentication- Select an authentication type from the drop-down menu
      i. Password- provide correct password for this authentication option
      ii. PEM/PPK file-choose a file and provide the file name if the user selects this authentication option
   e. Command-Enter the relevant Sqoop command
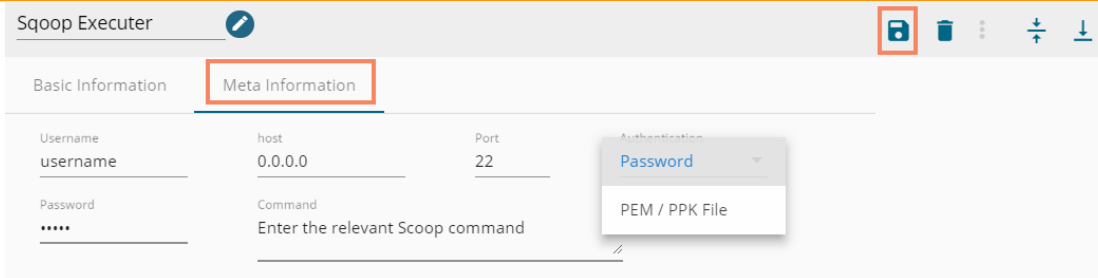
7) Create one Event and connect it to the Sqoop component for receiving the success message.
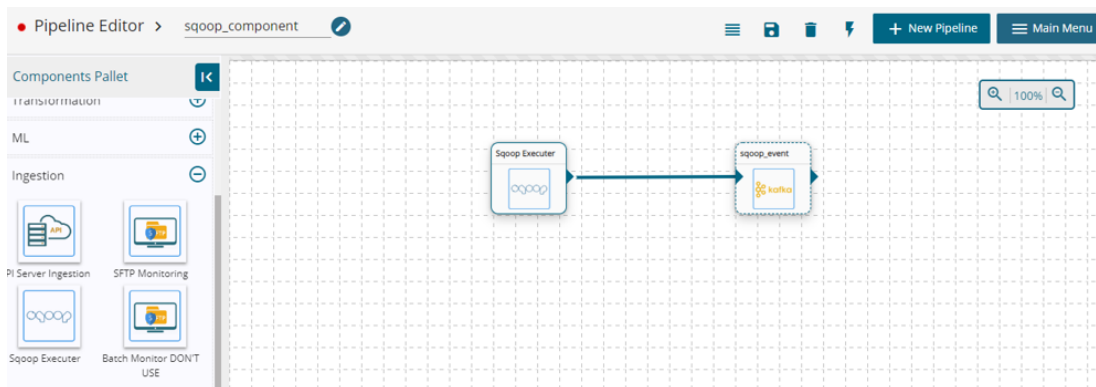
8) Once the data gets transferred to HDFS, the user can use it for further processing in the pipeline.