

# How to Guide

## Data Writers (Data Pipeline)

**Version: Release 1.1**

## Contents

|                             |    |
|-----------------------------|----|
| Component Description-..... | 3  |
| 1. S3 Writer.....           | 3  |
| 2. JDBC Writer.....         | 5  |
| 3. HDFS Writer.....         | 7  |
| 4. Cassandra Writer.....    | 9  |
| 5. ES Writer.....           | 12 |

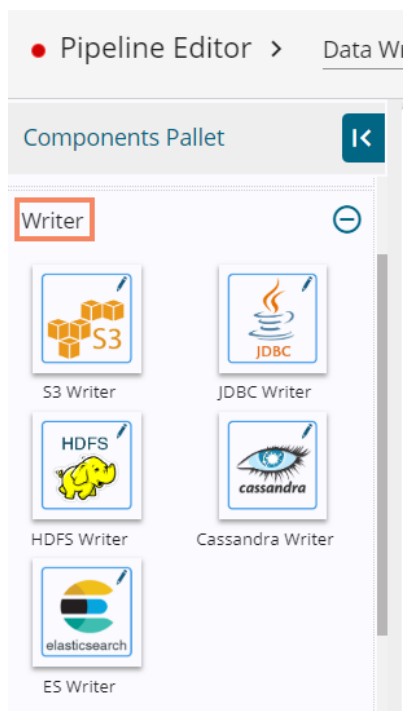
# Component Description-

BDB Data Pipeline provides Writer components to write input data to the required place.

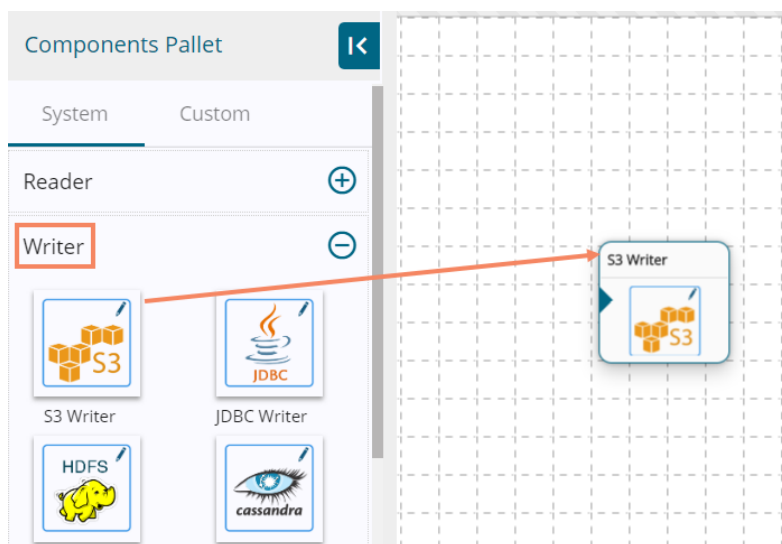
## 1. S3 Writer

The S3 writer component writes data to AWS S3 location.

- 1) Navigate to the Pipeline Workflow Editor.
- 2) Expand the Writer section using the Components Pallet.



- 3) Drag the S3 Writer component to the workspace.



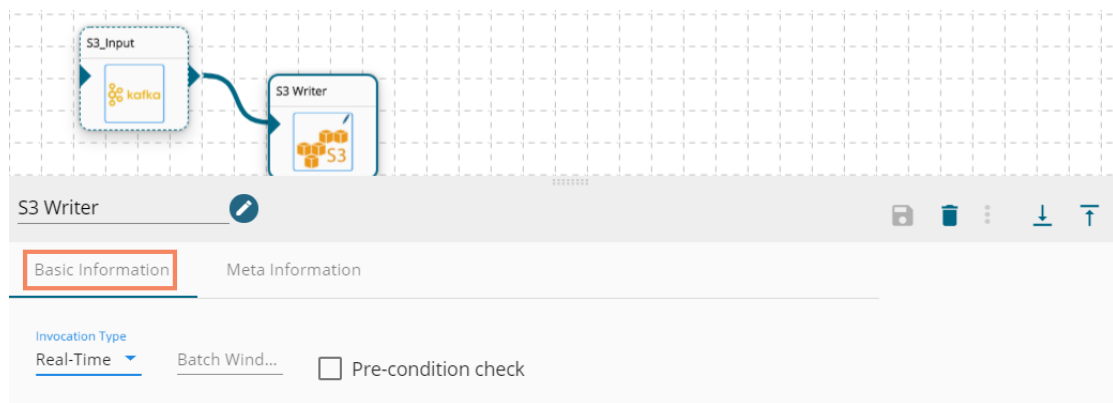
- 4) S3 Writer requires data input from an event and writes the data to S3 location.
- 5) Create one event and drag it to the workspace.
- 6) Connect the created input Event to the dragged S3 writer component.

Note: The data in the input event can come from any Ingestion, Readers or shared events.



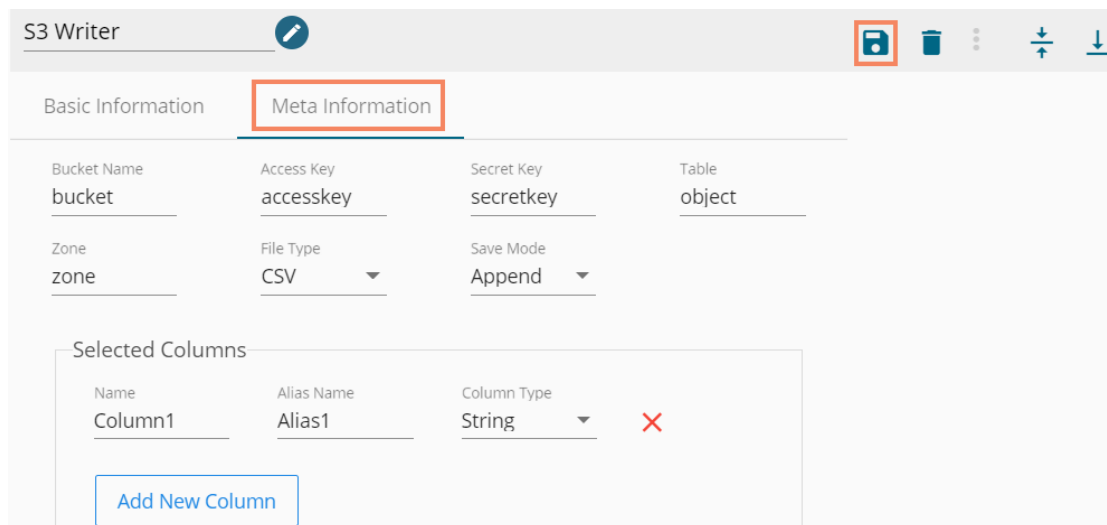
- 7) Click the dragged S3 Writer component to get the configuration tabs.
- 8) The Basic Information tab opens by default.
  - a. Select the invocation type ( Real-Time/Batch)

Note: Currently, Pipeline only supports the Real-time option.

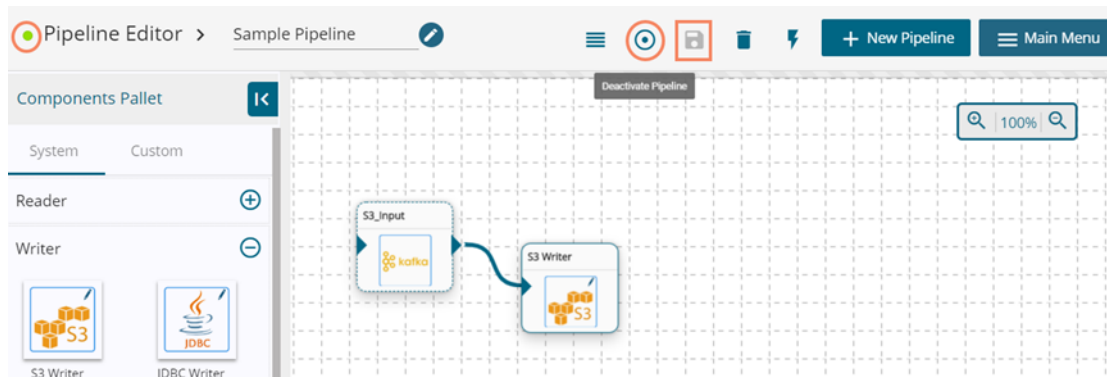


- 9) Select Meta Information tab for the S3 Writer component and provide all the required information for the S3 location:
  - a. Bucket Name
  - b. Table Name
  - c. Zone
  - d. Access Key
  - e. Secret Key
  - f. Selected Columns- Select the columns that you wish to store in the S3 location. Provide Column Name, Alias Name, Column Type for the selected column(s).  
(Add multiple columns to the Selected Columns section using the 'Add New Column' option.)
  - g. File Type-Select a file type (out of CSV/AVRO/JSON/Parquet) using the drop-down menu.
  - h. Save Mode- Select a save mode (out of Append/Overwrite) using the drop-down menu.

10) Save the S3 Writer component.



11) Save and activate the pipeline (The Pipeline Editor displays a green color dot for the activated pipelines).

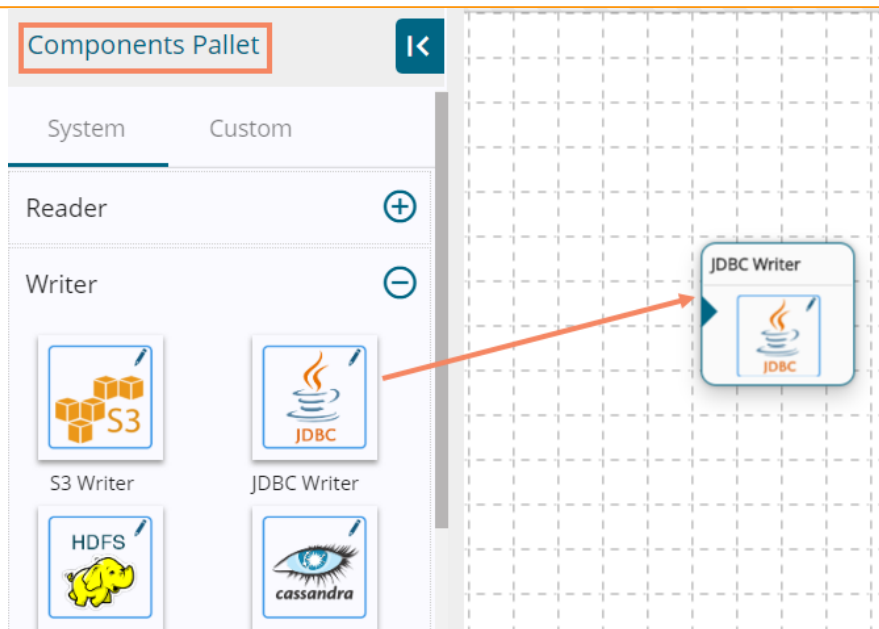


12) The data comes through the input event and writes to the S3 location.

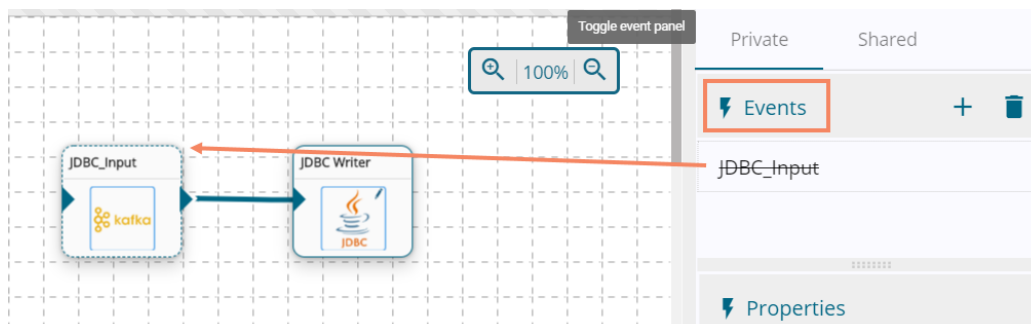
## 2. JDBC Writer

The JDBC writer helps to write data to the JDBC supported relational databases

- 1) Navigate to the Pipeline Workflow Editor
- 2) Expand the Writer section using the Components Pallet
- 3) Drag the JDBC Writer component to the workflow

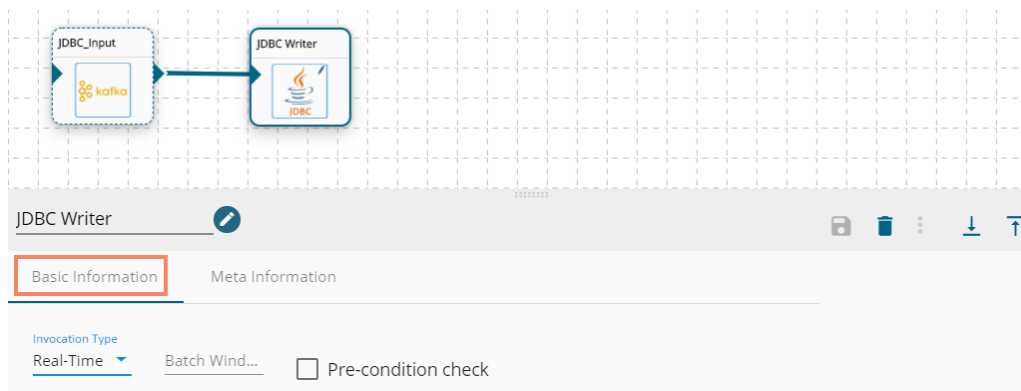


- 4) JDBC Writer requires input from an event and writes data to DB location.
- 5) Create and drag an Event to the workspace.
- 6) Connect the Event to the JDBC writer component (The data in input event can come from any Ingestion, Readers or shared events).

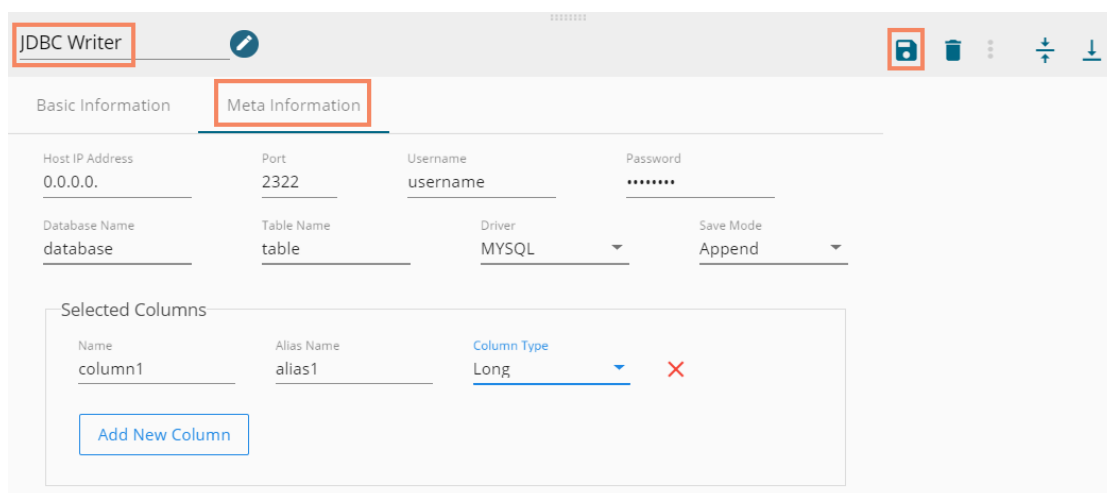


- 7) Select the dragged JDBC Writer component to get the configuration tabs.
- 8) The Basic information tab opens by default.
  - a. Select the invocation type ( Real Time/Batch)

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



- 9) Select the Meta Information tab.
- 10) Provide the required fields for the JDBC DB Location:
  - a. Host
  - b. Port
  - c. Username
  - d. Password
  - e. Driver (MySQL/MSSQL/Oracle/Postgres)
  - f. Database Name
  - g. Table Name
  - h. Save Mode (Append/Overwrite)
  - i. Selected Columns – Only the selected columns get stored to the configured Database location. Provide the Column name, Alias Name for the selected column, and select a column type. (Add multiple selected columns by clicking the ‘Add New Column’ option).
- 11) Save the JDBC writer component.

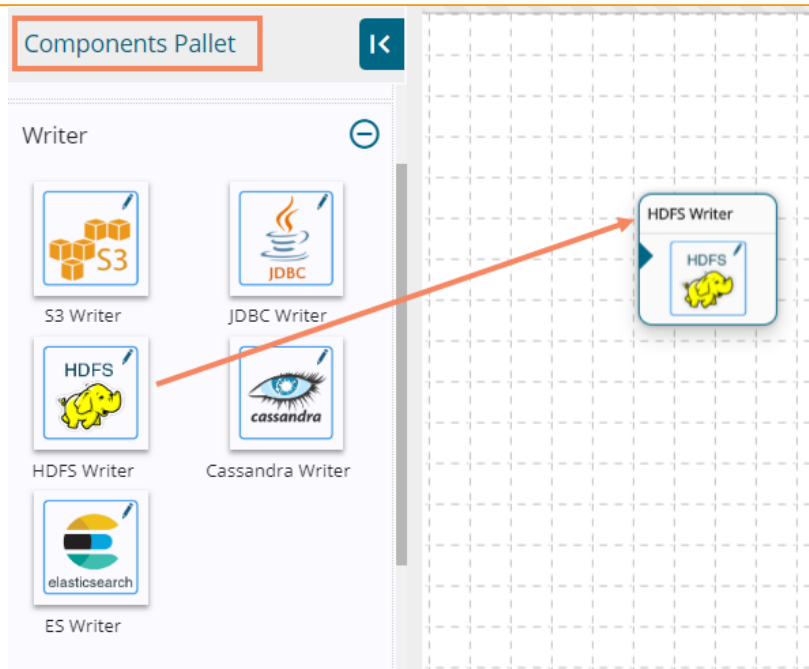


- 12) Save the pipeline and activate it. The input event reads the data and writes to the configured JDBC database location.

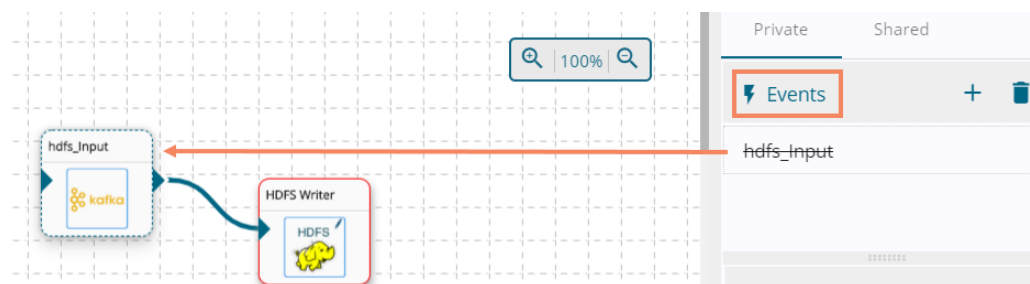
### 3. HDFS Writer

The HDFS writer writes the data to the HDFS location.

- 1) Navigate to the Pipeline Workflow Editor
- 2) Expand the Writer section using the Components Pallet
- 3) Drag the HDFS Writer component to the workspace

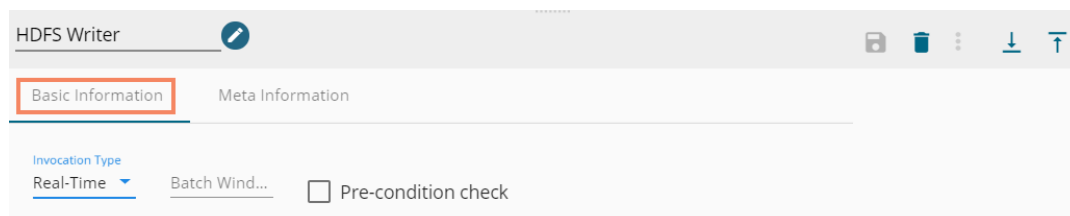


- 4) HDFS Writer requires input from an event and writes data to HDFS location.
- 5) Create and drag the input Event to the workspace.
- 6) Connect the input Event to the HDFS Writer component (The data in the input event can come from any Ingestion, Readers, or shared events.)



- 7) Select the dragged HDFS Writer 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.

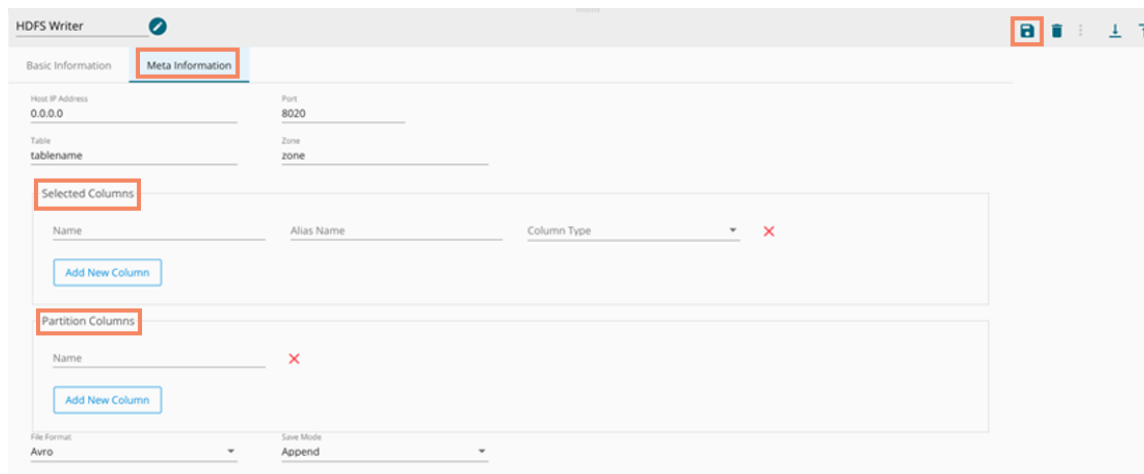




- 9) Select the Meta Information tab and provide the mandatory fields:
  - a. Host IP address
  - b. Port number
  - c. Table,
  - d. Zone,
  - e. File Format – Select any one file format (out of CSV/AVRO/JSON/Parquet) from the drop-down menu,
  - f. Save Mode- Select any one mode to save the data (out of Append/Overwrite) from the drop-down menu.
  - g. Selected Columns- Users can select the columns which they wish to store in the data writer. Provide a column name, alias name, and column type for the selected columns.
  - h. Users can also provide details of the Partition Columns.

Note:

  - i. Users can add multiple Selected Columns and Partition Columns by clicking the ‘Add New Column’ option.
  - ii. Selected Columns and Partition Columns are optional.
- 10) Save the HDFS Writer component.

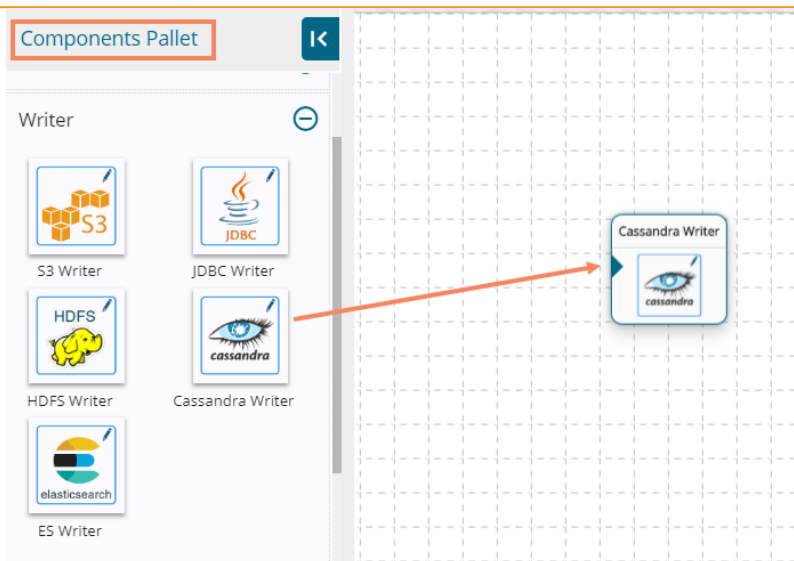


- 11) Save the pipeline and activate it. The input event reads the data and writes to the configured HDFS location.

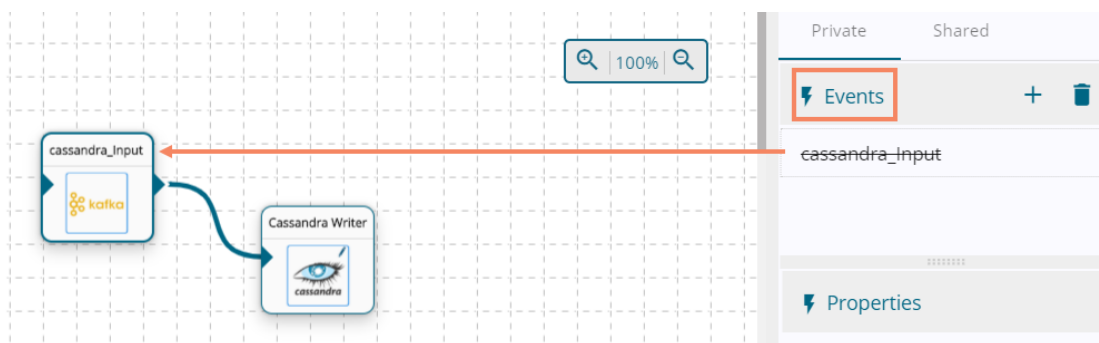
## 4. Cassandra Writer

The Cassandra Writer writes data to the Cassandra Database.

- 1) Navigate to the Pipeline Workflow Editor.
- 2) Expand the Writer section using the Components Pallet.
- 3) Drag the Cassandra Writer component to the workspace.

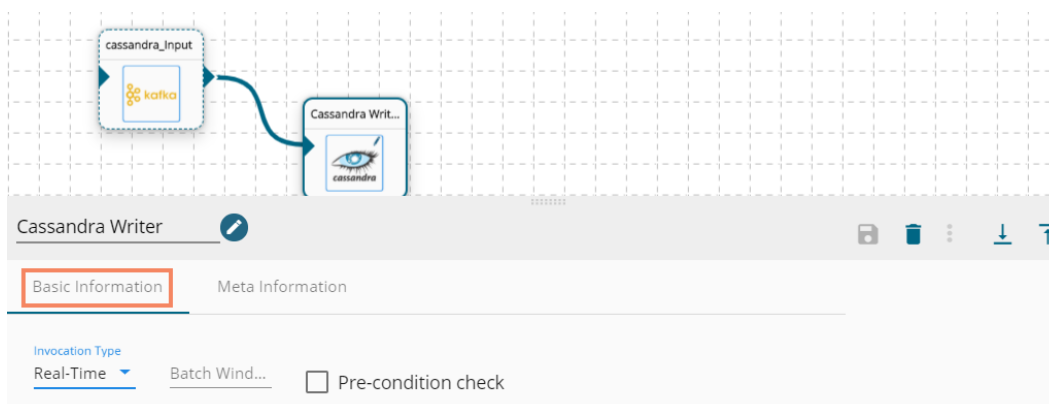


- 4) Cassandra Writer requires input from an event and writes data to DB location.
- 5) Create and drag the input Event to the workspace.
- 6) Connect the input Event to the Cassandra Writer component (The data in the input event can come from any Ingestion, Readers, or shared events.)



- 7) Select the dragged Cassandra Writer 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 the Meta Information tab and provide the mandatory fields:
  - a. Host IP address
  - b. Port number
  - c. Keyspace
  - d. Table
  - e. Cluster
  - f. Username
  - g. Password
  - h. Compression Method
  - i. Consistency
  - j. No. of Rows per Batch
  - k. Save Mode- Select any one mode to save the data (out of Append/Overwrite) from the drop-down menu.
  - l. UUID Column Name- Enter the UUID column name (Optional)
  - m. Selected Columns- Users can select the columns which they wish to store in the data writer. Provide a column name, alias name, and column type for the selected columns.
  - n. Users can also provide details of the Partition Columns.

Note:

  - i. Users can add multiple Selected Columns and Partition Columns by clicking the 'Add New Column' option.
  - ii. UUID Column Name, Selected Columns and Partition Columns are optional.
- 10) Save the Cassandra Writer component.

The screenshot shows the 'Cassandra Writer' configuration window. The 'Meta Information' tab is active. The configuration includes the following fields:

- Host IP Address: 0.0.0.0
- Port: 9042
- Keyspace: keyspace
- Table: table
- Cluster: cluster
- Username: username
- Password: [masked]
- Compression Method: SNAPPY
- Consistency: 1
- No. of Rows Per Batch: [empty]

Under 'Selected Columns', there is one entry:

| Name | Alias Name | Column Type |
|------|------------|-------------|
| name | firstname  | String      |

Under 'Partition Columns', there is one entry:

| Name |
|------|
| name |

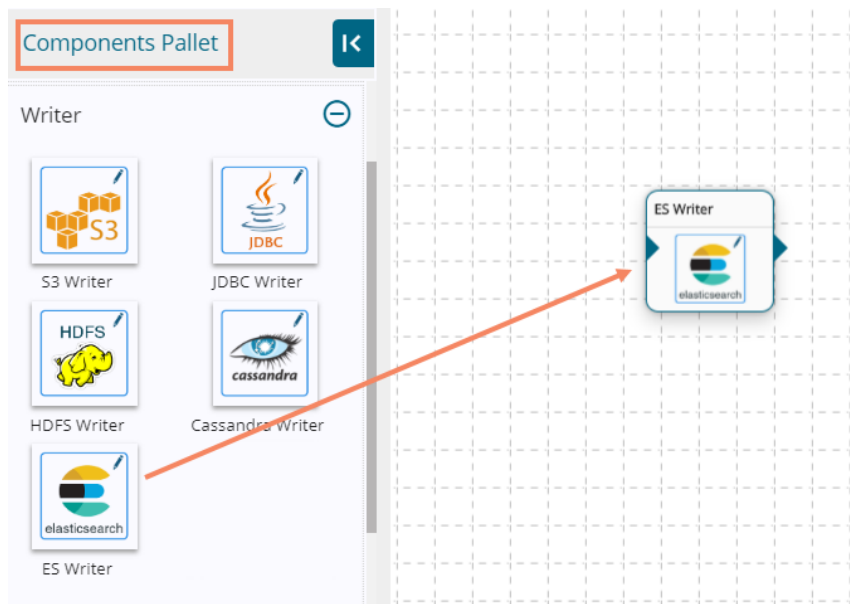
At the bottom, 'Save Mode' is set to 'Append' and 'UUID Column Name' is 'name'.

- 11) Save the pipeline and activate it. The component will read the data coming to the input event and write to the configured Cassandra database location.

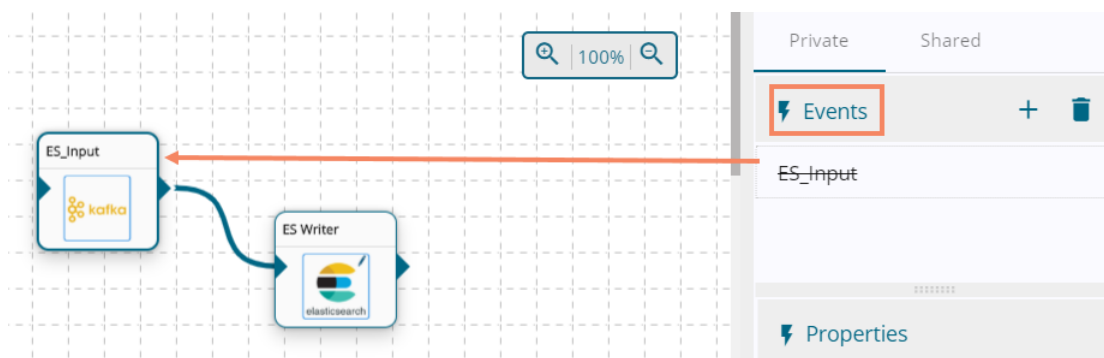
## 5. ES Writer

The ES writer writes the data to the Elastic Search engine.

- 1) Navigate to workflow editor.
- 2) Expand the Writer section using the Components Pallet.
- 3) Drag the ES Writer component to the Workspace.

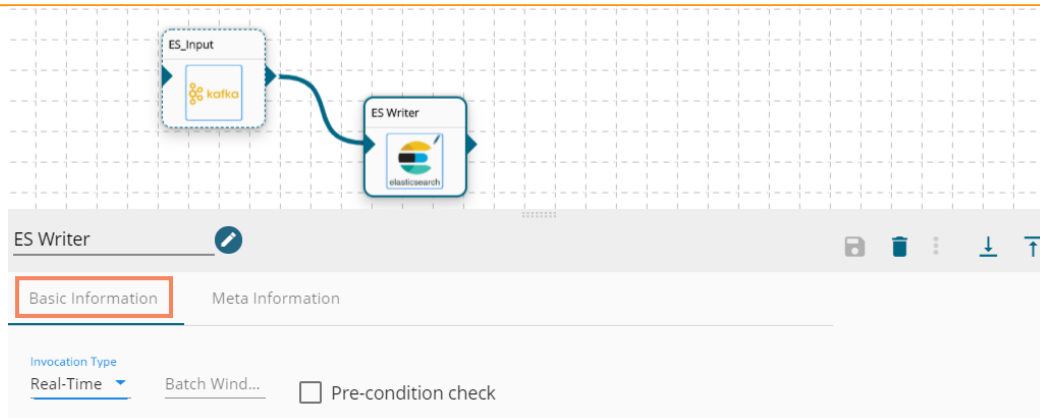


- 4) ES Writer requires input from an event and writes data to ES location.
- 5) Create and drag the input Event to the workspace.
- 6) Connect the input Event to the ES Writer component (The data in the input event can come from any Ingestion, Readers, or shared events.)

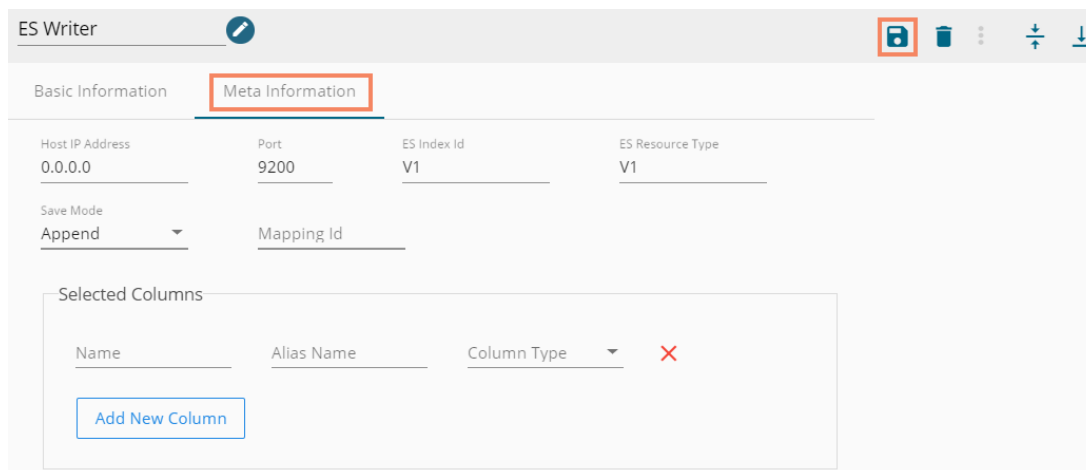


- 7) Select the dragged ES Writer component to get the configuration tabs.
- 8) The Basic Information tab opens by default.
  - a. Select the invocation type ( Real Time/Batch)

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



- 9) Select the Meta Information tab and provide the mandatory fields:
  - a. Host IP Address
  - b. Port number
  - c. ES Index Id
  - d. ES Resource Type,
  - e. Save mode – it supports only 'Append' option
  - f. Selected Columns- Data from the selected columns only stores in the ES location. Provide a column name, alias name, and column type. (Optional)
  - g. Mapping Id- Provide mapping id (Optional)
- 10) Save the ES Writer component from the metadata window first before saving the pipeline.



- 11) Save the pipeline and activate it. The input event reads the data and writes to the configured ES location.