

How to Guide

Stream Component (DP)

Version: R-4.5

Contents

1. Component Overview	3
2. Step by Step Process to Stream Component	3
3. Windowing	4

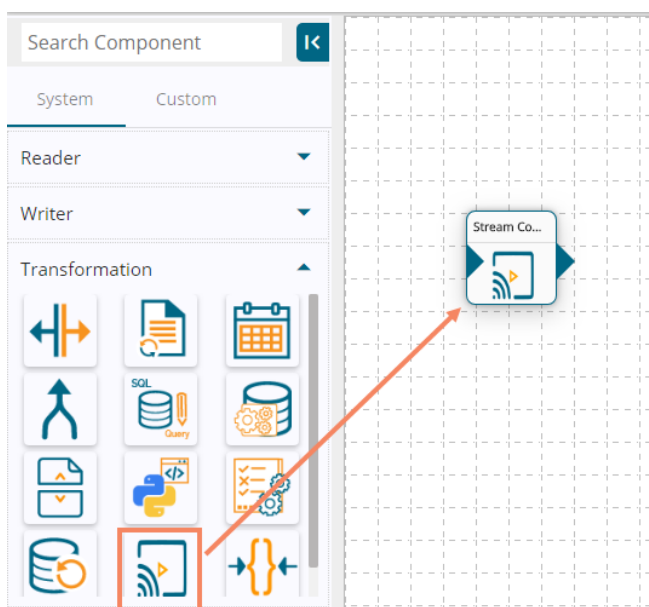
1. Component Overview

Stream Component is a transformation component, which is specifically designed to process IoT device-generated data. We can use this component along with any streaming data input source and perform aggregation on a time window. Currently, we are supporting (min, max, avg, last) operations.

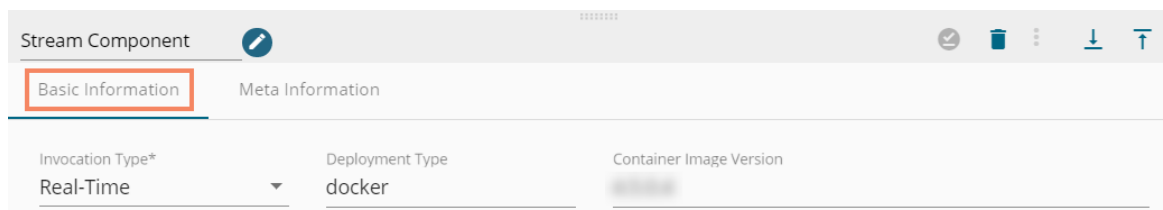
2. Step by Step Process to Stream Component

The steps to configure the Stream component is described in this section.

- i) Navigate to the Pipeline Editor.
- ii) Drag the Stream Component from the Transformation section provided under the Components Pallet and drop to the Pipeline Workspace.

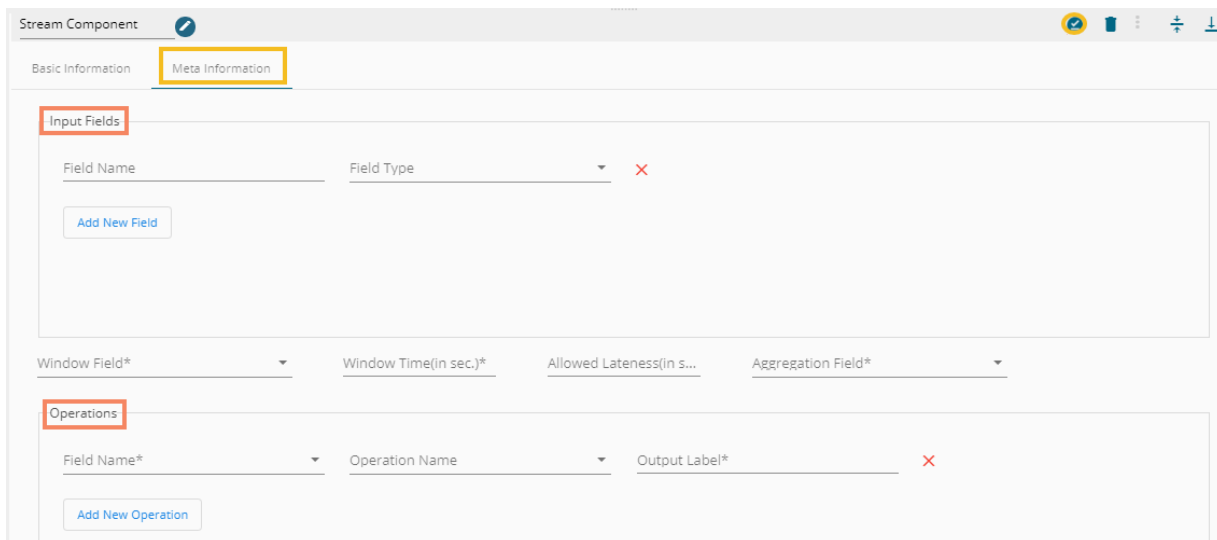


- iii) Click the dragged Stream component to get the Configuration fields:
- iv) The Basic Information tab opens by default.
 - a. Select the invocation type (Real-Time/Batch)
 - b. Deployment Type: It comes preselected based on the component.
 - c. Container Image Version: It comes preselected based on the component.



- v) Click on the Meta Information tab to open the configuration fields for the selected columns.

- a. Input Fields: User needs to provide all/selected columns that IoT device is producing along with their Data Type.
- b. Window Field: The user needs to provide the field using which we want to do bucketing of data. This field must be a timestamp value.
- c. Window Time: Time duration of Window (Bucket).
- d. Allowed Lateness: Allowed time up to which the component accepts late data.
- e. Aggregation Field: User needs to provide field using which component performs aggregation operations on operation fields.
- f. Operations: User needs to select the field name, operation name, and provide the output field label. Multiple operations are also supported
- g. Click the 'Save' icon.



Note: the user can add new field and new operation by clicking the 'Add New Field' and 'Add New Operation' options respectively.

3. Windowing

Windowing is an operation in spark structured streaming where the stream of data gets divided into small amount of window and then performs aggregation operations on the data.

For detailed information, please visit this link: -

<https://spark.apache.org/docs/2.3.0/structured-streaming-programming-guide.html#window-operations-on-event-time>