Troubleshooting execution or failures of Data flow Activities

Last updated by | Ranjith Katukojwala | Jun 21, 2022 at 8:11 AM PDT

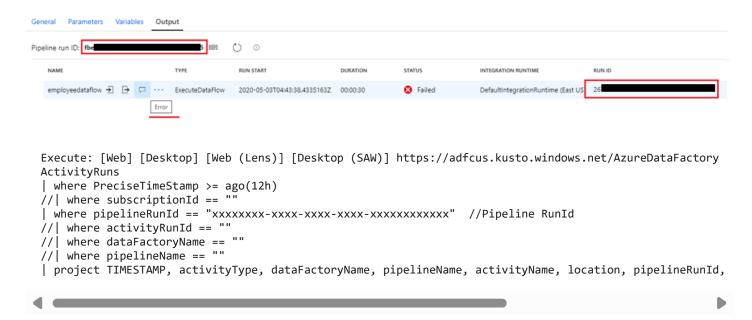## Contents

A detailed description of tables:

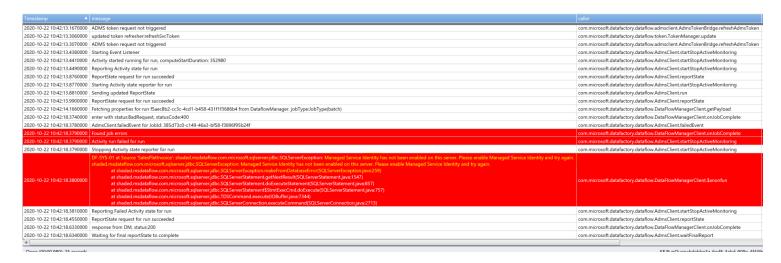| Table	What kind of errors we can expect
CustomLogEvents	Not able to create cluster/Databricks error
DatabricksLoggingTest	Error happens while running job on spark including transforming/transformed event - which tells about the progress, how many rows are processed till what stage. (stage is a combination of multiple transformations)

Investigation

- To troubleshoot issues in executing Dataflow activities, you will need the Pipeline RunId or the Activity RunId. The error details can provide the reason for the failure. In case you have to investigate the issue from the backend, ask customer to share the pipelineRunId or ActivityRunId from the portal, or use the **ActivityRuns** table to get the runIds and status of the pipeline execution:


```
Execute: [Web] [Desktop] [Web (Lens)] [Desktop (SAW)] https://adfcus.kusto.windows.net/AzureDataFactory
ActivityRuns
| where PreciseTimeStamp >= ago(12h)
//| where subscriptionId == ""
| where pipelineRunId == "xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx"  //Pipeline RunId
//| where activityRunId == ""
//| where dataFactoryName == ""
//| where pipelineName == ""
| project TIMESTAMP, activityType, dataFactoryName, pipelineName, activityName, location, pipelineRunId,
```

- You can notice the activityName and status column to check the status

For more specific errors, you may leverage the DatabrickloggingTest table. (Using another example below as old runId expired)

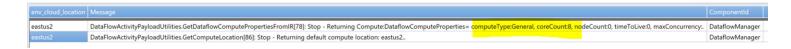For instance for this specific activity run id

```
DataflowClusterLogs
| where ActivityRunId == "xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx"
```


In this case, the DatabrickloggingTest table points to error that MSI auth is not enabled from DW to storage. We can ask the user to check if DW can access the storage account.

To check **Compute Type and Core Count**.

```
cluster('adfcus.kusto.windows.net').database('AzureDataFactory').AdfTraceEvent
| union cluster('adfneu.kusto.windows.net').database('AzureDataFactory').AdfTraceEvent
  | where TraceCorrelationId  == "XXXXXXXXXXXXXXXXXXXXXXXX" //<PipelineRunId>
  | where Message contains "DataFlowActivityPayloadUtilities"
  | distinct env_cloud_location, Message,ComponentId
```
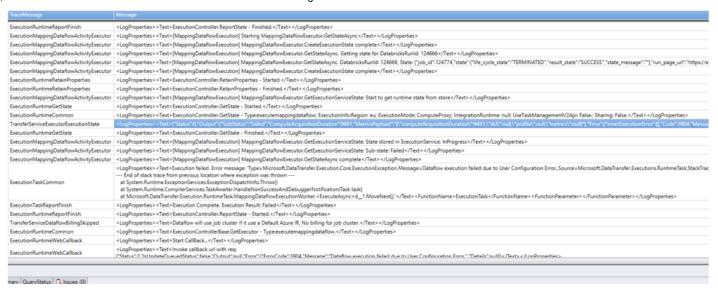

The below query helps us to get the DSL skeleton. Please use PipelineRunId in the below query. Note: May not be available for some Debug runs. PG team is working on this.

```
cluster('adfcus.kusto.windows.net').database('AzureDataFactory').DataFlowTelemetryEvent
| union cluster('adfneu.kusto.windows.net').database('AzureDataFactory').DataFlowTelemetryEvent
| where EventSource == "Broker"
| where TraceCorrelationId == ""
```

```
Event
        ],
        "sinks": [
          {
            "datasetName": "sink1",
            "datasetStore": "adlsgen2",
            "connectorDatasetType": "INLINE",
            "datasetOptions": {
              "format": "'delta'",
              "transactionCommit": null,
              "quoteChar": null,
              "errorHandlingOption": null,
              "outputRejectedData": null,
              "rowDelimiter": null,
              "nullValue": null,
              "charToEscapeQuoteEscaping": null,
              "encodingName": null,
              "compressionCodec": null,
              "escapeChar": null,
              "reportSuccessOnError": null,
              "range": null,
              "commentChar": null,
              "sheetName": null,
              "columnNamesAsHeader": null,
              "sheetIndex": null,
              "columnDelimiter": null,
              "firstRowAsHeader": null
            },
            "authenticationType": "AccountKey",
            "stagingStorageType": "",
            "stagingAuthenticationType": ""
          }
        ],
        "transformations": {
          "aggregate": 1,
          "join": 2,
          "source": 3,
          "derive": 1,
          "foldDown": 1,
          "select": 1,
          "alterRow": 1,
          "sink": 1
        }
      },
      "runtimeDetails": {
        "dataflowName": null,
        "sinkRuntimeMetrics": null
      }
    }
```

- You may further query the **CustomLogEvent** table to find the cause of the error:

```
CustomLogEvent
| where PreciseTimeStamp >= ago(12h)
| where ActivityId == "xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx" //Activity RunId
| where Message has "Dataflow execution failed"
| project PreciseTimeStamp, Role, TraceMessage, Message, EventName, TraceComponent, StatusCode
| sort by PreciseTimeStamp asc
```

Data Collection

If you've to engage a SME/EEE/PG, kindly request the following data from the customer:

- Failed Pipeline RunId or Activity Run Id for investigation
- Pipeline Support Files

NOTE: To retrieve Dataflow backend logs, please reach out to EEEs as it would need access via SAW machine.

Additional Information:

- **Icm References:** 210506861 ☒
- **Keywords:**

How good have you found this content?