

Troubleshooting execution or failures of Data flow Activities

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

Contents

- [Investigation](#)
- [Data Collection](#)

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:

General

Parameters

Variables

Output

Pipeline run ID: fbe[REDACTED]

NAME	TYPE	RUN START	DURATION	STATUS	INTEGRATION RUNTIME	RUN ID
employeedataflow	ExecuteDataFlow	2020-05-03T04:43:38.4335163Z	00:00:30	Failed	DefaultIntegrationRuntime (East US)	26[REDACTED]

Error

```
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-xxxxxxxxxxxxx" //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

Timestamp	activityType	dataFactoryName	pipelineName	activityName	location	pipelineRunId	activityRunId	status	failureType	start
2020-05-03 04:43:38.4078465	ExecuteDataFlow	DATAFLOWLAB	employeeDataflow	employeeDataflow	eastus	7b-2b5d-41-3301-410f-102910281-8029	20200503-0443-384078465-010-010-010	Queued		2020
2020-05-03 04:43:39.1883619	ExecuteDataFlow	DATAFLOWLAB	employeeDataflow	employeeDataflow	eastus	7b-2b5d-41-3301-410f-102910281-8029	20200503-0443-384078465-010-010-010	InProgress		2020

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-xxxxxxxxxxxxxx"
```

Timestamp	message	caller
2020-10-22 10:42:13.1670000	ADMS token request not triggered	com.microsoft.datafactory.dataflow.admsclient.AdmsTokenBridge.refreshAdmsToken
2020-10-22 10:42:13.3060000	updated token refresher:refreshSvcToken	com.microsoft.datafactory.dataflow.token.TokenManager.update
2020-10-22 10:42:13.3070000	ADMS token request not triggered	com.microsoft.datafactory.dataflow.admsclient.AdmsTokenBridge.refreshAdmsToken
2020-10-22 10:42:13.4380000	Starting Event Listener	com.microsoft.datafactory.dataflow.AdmsClient.startStopActiveMonitoring
2020-10-22 10:42:13.4410000	Activity started running for run, computeStartDuration: 352980	com.microsoft.datafactory.dataflow.AdmsClient.startStopActiveMonitoring
2020-10-22 10:42:13.4490000	Reporting Activity state for run	com.microsoft.datafactory.dataflow.AdmsClient.startStopActiveMonitoring
2020-10-22 10:42:13.8760000	ReportState request for run succeeded	com.microsoft.datafactory.dataflow.AdmsClient.reportState
2020-10-22 10:42:13.8770000	Starting Activity state reporter for run	com.microsoft.datafactory.dataflow.AdmsClient.startStopActiveMonitoring
2020-10-22 10:42:13.8810000	Sending updated ReportState	com.microsoft.datafactory.dataflow.AdmsClient.run
2020-10-22 10:42:13.9900000	ReportState request for run succeeded	com.microsoft.datafactory.dataflow.AdmsClient.reportState
2020-10-22 10:42:14.1660000	Fetching properties for run f5aec8b2-cc3c-4cd1-b458-431f1f3686b4 from DataflowManager: jobType:jobType(batch)	com.microsoft.datafactory.dataflow.DataFlowManagerClient.getPayload
2020-10-22 10:42:18.3740000	enter with status:BadRequest, statusCode:400	com.microsoft.datafactory.dataflow.DataFlowManagerClient.onJobComplete
2020-10-22 10:42:18.3780000	AdmsClient.failedEvent for JobId: 385d73c0-c149-46a3-bf58-f3896f95b24f	com.microsoft.datafactory.dataflow.AdmsClient.failedEvent
2020-10-22 10:42:18.3790000	Found job errors	com.microsoft.datafactory.dataflow.DataFlowManagerClient.onJobComplete
2020-10-22 10:42:18.3790000	Activity run failed for run	com.microsoft.datafactory.dataflow.AdmsClient.startStopActiveMonitoring
2020-10-22 10:42:18.3790000	Stopping Activity state reporter for run	com.microsoft.datafactory.dataflow.AdmsClient.startStopActiveMonitoring
2020-10-22 10:42:18.3800000	DF-SYS-01 at Source 'SalesFlatInvoice': shaded.msdataflow.com.microsoft.sqlserver.jdbc.SQLServerException: Managed Service Identity has not been enabled on this server. Please enable Managed Service Identity and try again. at shaded.msdataflow.com.microsoft.sqlserver.jdbc.SQLServerException.makeFromDatabaseError(SQLServerException.java:259) at shaded.msdataflow.com.microsoft.sqlserver.jdbc.SQLServerStatement.getNextResult(SQLServerStatement.java:1547) at shaded.msdataflow.com.microsoft.sqlserver.jdbc.SQLServerStatement.doExecuteStatement(SQLServerStatement.java:857) at shaded.msdataflow.com.microsoft.sqlserver.jdbc.SQLServerStatement\$StmntExecCmd.doExecute(SQLServerStatement.java:757) at shaded.msdataflow.com.microsoft.sqlserver.jdbc.TDSCommand.execute(IOBuffer.java:7344) at shaded.msdataflow.com.microsoft.sqlserver.jdbc.SQLServerConnection.executeCommand(SQLServerConnection.java:2713)	com.microsoft.datafactory.dataflow.DataFlowManagerClient.\$anonfun
2020-10-22 10:42:18.3810000	Reporting Failed Activity state for run	com.microsoft.datafactory.dataflow.AdmsClient.startStopActiveMonitoring
2020-10-22 10:42:18.4550000	ReportState request for run succeeded	com.microsoft.datafactory.dataflow.AdmsClient.reportState
2020-10-22 10:42:18.6330000	response from DM, status:200	com.microsoft.datafactory.dataflow.DataFlowManagerClient.onJobComplete
2020-10-22 10:42:18.6340000	Waiting for final reportState to complete	com.microsoft.datafactory.dataflow.AdmsClient.waitFinalReport

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 == "XXXXXXXXXXXXXXXXXXXXXXXXXX" //<PipelineRunId>
| where Message contains "DataFlowActivityPayloadUtilities"
| distinct env_cloud_location, Message, ComponentId
```

env_cloud_location	Message	ComponentId
eastus2	DataFlowActivityPayloadUtilities.GetDataflowComputePropertiesFromIR[78]: Stop - Returning Compute:DataflowComputeProperties= computeType:General, coreCount:8, nodeCount:0, timeToLive:0, maxConcurrency:..	DataflowManager
eastus2	DataFlowActivityPayloadUtilities.GetComputeLocation[86]: Stop - Returning default compute location: eastus2..	DataflowManager

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

```

TraceMessage	Message
ExecutionRuntimeReportFinish	<LogProperties> <Text>ExecutionController.ReportState - Finished.</Text> </LogProperties>
ExecutionMappingDataflowActivityExecutor	<LogProperties> <Text>[MappingDataflowExecution] Starting MappingDataflowExecutor.GetStateAsync.</Text> </LogProperties>
ExecutionMappingDataflowActivityExecutor	<LogProperties> <Text>[MappingDataflowExecution] MappingDataflowExecutor.CreateExecutionState complete.</Text> </LogProperties>
ExecutionMappingDataflowActivityExecutor	<LogProperties> <Text>[MappingDataflowExecution] MappingDataflowExecutor.GetStateAsync. Getting state for DatabricksRunId: 124666.</Text> </LogProperties>
ExecutionMappingDataflowActivityExecutor	<LogProperties> <Text>[MappingDataflowExecution] MappingDataflowExecutor.GetStateAsync. DatabricksRunId: 124666, State: { "job_id":124774,"state":{"life_cycle_state":"TERMINATED"},"result_state":"SUCCESS","state_message":""},"run_page_url":"https://e
ExecutionMappingDataflowActivityExecutor	<LogProperties> <Text>[MappingDataflowExecution] MappingDataflowExecutor.CreateExecutionState complete.</Text> </LogProperties>
ExecutionRuntimeRetainProperties	<LogProperties> <Text>ExecutionController.RetainProperties - Started.</Text> </LogProperties>
ExecutionMappingDataflowActivityExecutor	<LogProperties> <Text>[MappingDataflowExecution] MappingDataflowExecutor.GetExecutionServiceState: Start to get runtime state from store.</Text> </LogProperties>
ExecutionRuntimeGetState	<LogProperties> <Text>ExecutionController.GetState - Started.</Text> </LogProperties>
ExecutionRuntimeCommon	<LogProperties> <Text>ExecutionController.GetState - Type:executemappingdataflow, ExecutionInfoRegion: eu, ExecutionMode: ComputeProxy, IntegrationRuntime: null, UseTaskManagementV2Api: False, Sharing: False.</Text> </LogProperties>
TransferServiceExecutorExecutionState	<LogProperties> <Text>{"Status":"Output","SubStatus":"Failed","ComputeAcquisitionDuration":"9491","MetricsPayload":{"computeAcquisitionDuration":"9491","diagnostics":{"diagnostics":{"error":{"innerExecutionError":{"Code":"3904","Message":
ExecutionRuntimeGetState	<LogProperties> <Text>ExecutionController.GetState - Finished.</Text> </LogProperties>
ExecutionMappingDataflowActivityExecutor	<LogProperties> <Text>[MappingDataflowExecution] MappingDataflowExecutor.GetExecutionServiceState: State stored in ExecutionService: InProgress.</Text> </LogProperties>
ExecutionMappingDataflowActivityExecutor	<LogProperties> <Text>[MappingDataflowExecution] MappingDataflowExecutor.GetExecutionServiceState: Sub-state: Failed.</Text> </LogProperties>
ExecutionMappingDataflowActivityExecutor	<LogProperties> <Text>[MappingDataflowExecution] MappingDataflowExecutor.GetStateAsync complete.</Text> </LogProperties>
ExecutionTaskCommon	<LogProperties> <Text>Execution failed. Error message: Type=Microsoft.DataTransfer.Execution.Core.ExecutionException,Message=Dataflow execution failed due to User Configuration Error,Source=Microsoft.DataTransfer.Executions.RuntimeTaskStackTrac --- End of stack trace from previous location where exception was thrown --- at System.Runtime.ExceptionServices.ExceptionDispatchInfo.Throw() at System.Runtime.CompilerServices.TaskAwaiter.HandleNonSuccessAndDebuggerNotification(Task task) at Microsoft.DataTransfer.Execution.RuntimeTask.MappingDataflowExecutionWorker.<ExecuteAsync>d__1.MoveNext() at Microsoft.DataTransfer.Execution.RuntimeTask.MappingDataflowExecutionWorker.<ExecuteAsync>d__1.MoveNext() </Text> <FunctionName>ExecutionTask</FunctionName> <FunctionParameter></FunctionParameter> </LogProperties>
ExecutionTaskReportFinish	<LogProperties> <Text>Execution Complete. Execution Result: Failed.</Text> </LogProperties>
ExecutionRuntimeReportFinish	<LogProperties> <Text>ExecutionController.ReportState - Started.</Text> </LogProperties>
TransferServiceDataflowBillingSkipped	<LogProperties> <Text>Dataflow will use job cluster if it use a Default Azure IR, No billing for job cluster.</Text> </LogProperties>
ExecutionRuntimeCommon	<LogProperties> <Text>ExecutionControllerBase.GetExecutor - Type:executemappingdataflow.</Text> </LogProperties>
ExecutionRuntimeWebCallback	<LogProperties> <Text>Start CallBack.</Text> </LogProperties>
ExecutionRuntimeWebCallback	<LogProperties> <Text>Invoke callback url with req: {"Status":"Idle","UpdateQueueStatus":"false","Output":null,"Error":{"Errorcode":"3904","Message":"Dataflow execution failed due to User Configuration Error","Details":null}}.</Text> </LogProperties>

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?

