

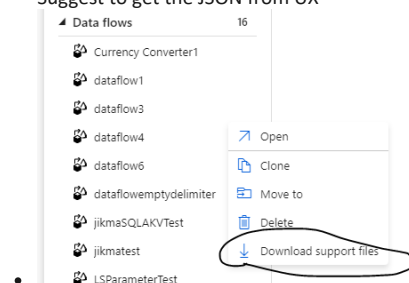
Dataflow Troubleshooter

Last updated by | Jackie Huang | Jan 4, 2022 at 12:24 AM PST

Dataflow Troubleshooter

Wednesday, February 5, 2020
08:41 PM

1. ADF (data flow) Overview
 - General ADF
 - Data Flow activity run
 - Data Flow sandbox run
 - Data Flow debug session
 - Wrangling data flow
2. One sample E2E for data flow
 - Data flow parameter vs pipeline parameter -- <https://docs.microsoft.com/en-us/azure/data-factory/parameters-data-flow#set-mapping-data-flow-parameters-from-pipeline>
3. Kusto table for data flow + other external activity
 - a. ActivityRuns/PipelineRuns (ADF) -- Check all activity/pipeline run info
 - b. ApiOperationEvent (ADF) -- Track API and get the cluster ID especially for data flow debug
 - c. AdfTraceEvent (ADF) -- All ADF logs before the job lands to ADMS side
 - d. CustomLogEvent (ADMS) -- All ADMS logs for job execution
 - e. DatabrickLoggingTest (Data flow only) -- All traces for parsing payload + some execution log
 - f. DatabrickApiEvent (Data flow only) -- minor table, used to track the api trace in databricks side.
 - g. HDI table if it's external provider issue: <https://hdinsight.kusto.windows.net> (covered by other TSG)
4. Issue category:
 - Prerequisite
 - Determine which job it is (activity run/sandbox run/dataflow debug)
 - Use AdfTraceEvent + CustomLogEvent to categorize the issue
 - Payload parsing
 - DatabrickLoggingTest table
 - Source/Sink connection
 - Google for common error (most likely SQL)
 - Transformation + execution
 - Suggest to get the JSON from UX



- Get the cluster ID/job run url (CustomLogEvent)

5. Questions

Wildcard path sample:

```
wildcardPaths:[$tempPath+'/append/**/*.*.csv'],
wildcardPaths:['./src/test/resources/data/\\(re*ases\\|\\{a\\}\\).csv'],
```

Additional Information:

Icm References:

Keywords: Data flow

Authors: jikma

Reviewers: vimals

Created with Microsoft OneNote 2016.

How good have you found this content?

