How to check if Dataflow Activity source has wildcard path

Last updated by | Anudeep Sharma | Feb 23, 2023 at 12:17 PM PST

Contents

- Issue
- Reason
- Recommendation

Issue

How to determine if the customer is using a wild card path in the source and processed file count. Basics to learn for this
https://supportability.visualstudio.com/AzureDataFactory/_wiki/wikis/AzureDataFactory/857242/Introduction-to-different-concepts-in-Mapping-Dataflows

Reason

The customer might mention that the Dataflow pipeline is running more than expected time. Although increased core counts and compute type to Memory-optimized, still with less volume, the pipeline is running longer than expected.

Recommendation

In this situation, you might need to check whether the customer is using any wildcard path or not and the no. of files processed.

1. You may use the below query to identify the type of source.

```
cluster('adfcus.kusto.windows.net').database("AzureDataFactory").DataFlowTelemetryEvent
| union  cluster('adfneu.kusto.windows.net').database("AzureDataFactory").DataFlowTelemetryEvent
| where TraceCorrelationId == "<Dataflow Pipeline Run Id>"
| where EventSource == 'Broker'
|project Event
```

2. You may use the below query to identify whether wildcard is enabled and if yes, the FilePathcount, FileCount and the time to discover the files/folders.

```
let runid = <Dataflow Pipeline or Activity Run Id>;
 cluster('adfcus.kusto.windows.net').database("AzureDataFactory").DataflowClusterLogs
 |union  cluster('adfneu.kusto.windows.net').database("AzureDataFactory").DataflowClusterLogs
      | where ActivityRunId == runid or TraceCorrelationId == runid
   | where Message contains "new event"
       or Message contains "Duration to initialize file system"
       or Message contains "Duration to discover files/folder configured by wildcards or paths"
       or Message contains "FileCount"
```

Message
PathsTimed at transformation: source1. FileCount: 23785 FilePathCount: 1
AdmsClient.processingEvent for JobId: 866d53ed-c374-4212-92b7-6e6ec957d35b, metric: BatchProcessingMetric(~> source1, pathResolutionDuration(long) - Duration to discover files/folder configured by wildcards or paths { },18576)

How good have you found this content?