SAP ODP: Slow performance

Last updated by | Tina Hu | Jun 29, 2022 at 2:32 AM PDT

Identify the Bottleneck

Execute below query to identify the bottleneck:

```
cluster('azuredmprod.kusto.windows.net').database('AzureDataMovement').CustomLogEvent
| where * contains <activityId> and TraceMessage == "TranferServiceJobTelemetry"
| extend text =parse_json(telemetry_trim_CustomLogEventMessage(Message))
| project text["SapOdpListPackages.WorkingDuration"], text["SapOdpListPackages.PeakWorkerCount"],text["SapOdpF
```

```
10
11  ⊟ CustomLogEvent
12      | where * contains "f3629f3b-1da3-4ef1-adf9-6b574d398a5e" and TraceMessage == "TranferServiceJobTelemetry"
13      | extend text =parse_json(telemetry_trim_CustomLogEventMessage(Message))
14      | project text["SapOdpListPackages.WorkingDuration"],
15      text["SapOdpListPackages.PeakWorkerCount"],
16      text["SapOdpFetchData.WorkingDuration"],
17      text["SapOdpFetchData.PeakWorkerCount"],
18      text["SapOdpParseData.WorkingDuration"],
19      text["SapOdpParseData.PeakWorkerCount"]
20
21  ⊟ DataflowClusterLogs
```

text_SapOdpListPackages.WorkingDuration	text_SapOdpListPackages.PeakWorkerCount	text_SapOdpFetchData.WorkingDuration	text_SapOdpFetchData.PeakWorkerCount	text_SapOdpParseData.WorkingDuration	text_SapOdpParseData.PeakWorkerCount
1.161447	1	0	0	0	0

If multiple rows exists, please check the last rows.

- If **SapOdpListPackages** duration is longer than others, it means the bottleneck is on SAP ODQ task side, please suggest customer tune the performance in their SAP system.
- If **SapOdpFetchData** duration is longer than others, it means the bottleneck is on network, try to follow the steps in **Basic Settings**, **Manual Data Partitioning** and suggest customer tune the network between SHIR installed machine and SAP server.
- If **SapOdpParseData** duration is longer than others, it means the bottleneck is on machine(CPU or Memory), then try to follow the steps in **Basic Settings**, **CPU**, **Manual Data Partitioning**

Basic Settings

1. Make sure we enable the parallel copy, make sure the sink side is a folder instead of a file, we can have 4~5 parallel count in a single copy activity after enabling it.
2. Make the source SAP server, SHIR installed machine and sink storage/database in the same region, or as close as possible. If source and sink are not in the same region, detect the bottleneck according to the copy activity output (reading from source v.s. writing to the sink), and make the SHIR installed machine closer to the bottleneck side.

CPU

CPU is important to improve the performance, 16 core CPU can have 12.8MB/s throughput, while 8 core CPU can have 8.4MB/s throughput. Even CPU is not up to 100%(like around 70%~80%), use a more powerful machine can still help improve the throughput.

Manual Data Partitioning

Try to split the data evenly with different query conditions in multiple copy activities, we can use foreach activity to trigger multiple copy activities with different query conditions at the same time.