

Troubleshoot OOM issue on Self-hosted IR

Last updated by | Ruoyu Li | Apr 3, 2023 at 6:38 PM PDT

Contents

- [Issue Symptom](#)
 - [Error Message](#)
- [Troubleshooting and mitigation](#)
 - [General guidance](#)
 - [Review SHIR memory utilization](#)
 - [Review diawp process level memory limitation](#)
 - [Check for known patterns](#)

Issue Symptom

Copy activity failed with below error message and with initial troubleshooting you noticed the copy runs on Self-hosted IR.

Error Message

```
ErrorCode=SystemErrorOutOfMemory, 'Type=Microsoft.DataTransfer.Common.Shared.HybridDeliveryException, Message=A task failed with out of memory., Source=Microsoft.DataTransfer.TransferTask, 'Type=System.OutOfMemoryException, Message=Exception of type 'System.OutOfMemoryException' was thrown., Source=mscorlib,'
```

Troubleshooting and mitigation

General guidance

- check what is the copy source and sink pair
- check whether the OOM error is consistent or intermittent
- check whether it's related with data size - generally we recommend to **split data into smaller files instead of processing one single large file**
- review SHIR heartbeats telemetry and collect SHIR report ID

Review SHIR memory utilization

You can check heartbeats table to review SHIR memory utilization. Please check **MemoryUtilization** column of the result. Sometimes, you also need to check **CommittedBytesInUseRatio** column.

```
Heartbeats
| where TIMESTAMP between (datetime(2023-03-15 11:10:00)..datetime(2023-03-15 12:00:00))
| where AgentGroupName == "<SHIR_name>" and AgentInstanceName == '<SHIR_node_name>'/'VM-ADF-PRD-ETQ1' //
| where SubscriptionId =~ 'subscription_ID'
| extend TotalMemory = tolong(extract(@"TotalMemory\": (\d+)", 1, Telemetries))
| extend AvailableMemory = tolong(extract(@"AvaliableMemory\": (\d+)", 1, Telemetries))
| extend CpuUtilization = tolong(extract(@"CpuUtilization\": (\d+)", 1, Telemetries))
| extend MaxCapacity = tolong(extract(@"MaxCapacity\": (\d+)", 1, Telemetries))
| extend UsedCapacity = tolong(extract(@"UsedCapacity\": (\d+)", 1, Telemetries))
| extend queueLength=tolong(extract(@"QueueLength\": (\d+)", 1, Telemetries))
| extend CommittedBytesInUseRatio=todouble(CommittedBytesInUseRatio=parse_json(Telemetries)["CommittedBytesInUseRatio"])
| project TIMESTAMP, AgentGroupId, AgentGroupName, AgentInstanceId, AgentInstanceName, IsUpgradeGranted, Telemetries, MemoryUtilization = (TotalMemory-Availab
```

- If SHIR node level memory utilization is low, please suggest customer to **reduce concurrent jobs count** so each diawp process is likely to be allocated more memory.
- If SHIR VM level memory utilization is high, please suggest customer to **scale up SHIR VM**.

Review diawp process level memory limitation

We can collect perfmon trace to review process-level memory utilization. See [this TSG](#) for details.

In SHIR we do have a setting to limit how much memory a process can use. It is calculated by machine cores number and some other settings, usually in SHIR this number is **1GB** for one single diawp.exe process.

We can check its value in telemetry by using query below.

```
CustomLogEvent
| where ActivityId == "<activity_run_id>"
and Message has "MaxMemoryLimit"
```

Check for known patterns

Some connectors (JSON, Excel, XML, Parquet) are more likely to hit OOM issues, please check if it matches any known patterns.

- [JSON, Excel, XML-OOM Issue](#)
- [Parquet/ORC OOM](#)