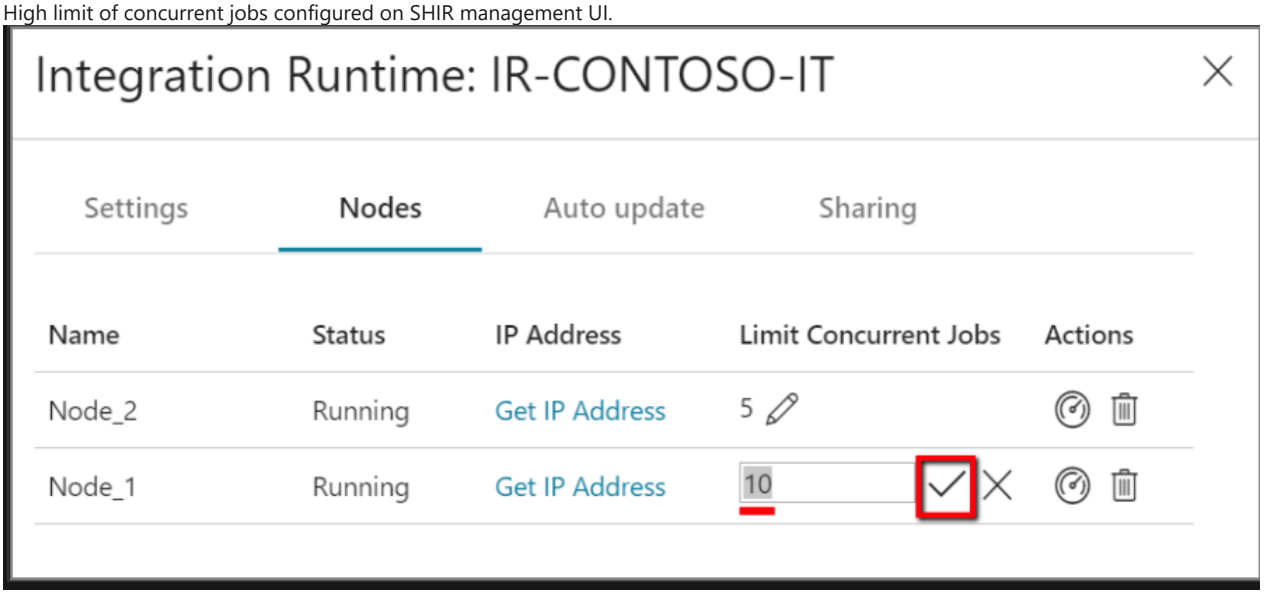
Troubleshoot OutOfMemory error caused by copy activity executed on SHIR

Last updated by | Veena Pachauri | Mar 8, 2023 at 11:23 PM PST

Issue description

If you receive below error message during the execution of copy activity run against SHIR, **"ErrorCode=UserErrorOutOfMemory,'Type=Microsoft.DataTransfer.Common.Shared.HybridDeliveryException,Message=Runtime get failed because the memory usage of self-hosted integration runtime machine is too high, please follow https://docs.microsoft.com/en-us/azure/data-factory/self-hosted-integration-runtime-troubleshoot-guide to address the high memory issue.,Source=,''Type=System.OutOfMemoryException,Message=Exception of type 'System.OutOfMemoryException' was thrown.,Source=Microsoft.DataTransfer.ClientLibrary,'"**, this indicates that the SHIR node is currently experiencing heavy workloads due to high concurrency, possibly contributed by below factors,

- High limit of concurrent jobs configured on SHIR management UI.



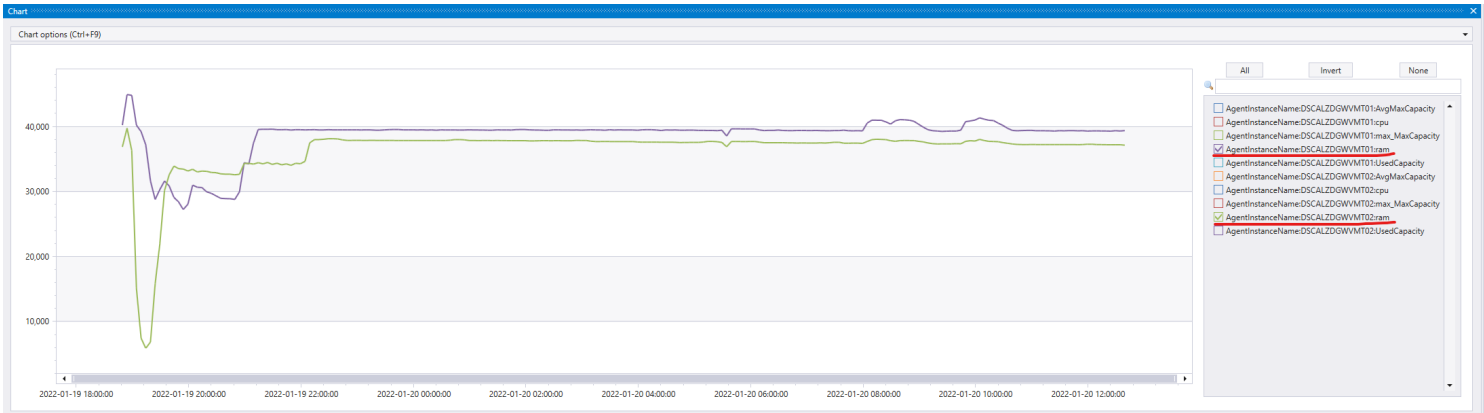
- ForEach activity configured in the pipeline where parallel execution is enabled. (isSequential set to false)
- For copy where format-based serialization/deserialization (eg. write data to blob stage in the format of parquet files) is involved, which is CPU intensive by nature, hence, the memory consumption may increase significantly during the process.
- High degree of copy parallelism, as controlled by the setting of "parallelCopies"

Diagnosis

- Run Kusto function to render time-series graph of the available memory on SHIR machines that are involved in the copy.

```
cluster('azuredmprod.kusto.windows.net').database('AzureDataMovement').query_SHIRHeartBeatsByActivityRunId(<CopyActivityRunId>, <FromDateInDatetimeFormat>, <T
```

- Check the concerned RAM metrics (naming pattern AgentInstanceName:[SHIRMachineName]:ram) in the right panel to focus on RAM usage only. Note that the vertical axis is measured by the available memory in MB.



- Run Kusto query to get the memory usage data of the concerned copy activity run. Check the "**Telemetry**" column in the query result.

```
cluster('azuredmprod.kusto.windows.net').database('AzureDataMovement').Telemetry_Prod_LongLatency_Copy_CompleteActivityRuns
| where ActivityId == <CopyActivityRunId>
```

If the value of "**Memory.CommittedBytesInUseRatio**" is greater than 90%, there's a high chance that copy activity would fail with out of memory. Usually the percentage should stay below 80% so that SHIR machine is not overloaded within copy duration.

The metric "**GatewayNodeName**" also indicates the exact machine that causes the out of memory error.

Mitigation

The following actions can be taken to alleviate SHIR nodes from heavy workloads and further resolve the OutOfMemory issue.

- Reduce the # of concurrent jobs that a SHIR node can run. This is controlled by the setting of "**Limit Concurrent Jobs**" under manage tab of integration runtimes.
- Reduce the concurrency of **ForEach** activity (aka. the number of iterations) or simply switch to sequential execution.
- Reduce the degree of copy parallelism, as controlled by "**parallelCopies**" under the settings tab of a single copy activity.
- Scale up the SHIR node with more physical memory.
- Scale out SHIR by adding more nodes up to 4.