Investigate OOM issues

Last updated by | Vitor Tomaz | Nov 22, 2021 at 10:30 AM PST

Contents

Issue

During the public preview, R resources are set to a maximum of 20% of the SQL Managed Instance resources, and depend on which service tier you choose.

In General, in most of the oom scenarios, if the process indeed runs out of memory, the customer would get the following error message:

If there is insufficient memory available for R, you will get an error message. Common error messages are:

*Unable to communicate with the runtime for 'R' script for request id: *******. Please check the requirements of 'R' runtime*

_'R' script error occurred during execution of 'sp_execute_external_script' with HRESULT 0x80004004. ...an external script error occurred: "..could not allocate memory (0 Mb) in C function 'R_AllocStringBuffer'"

An external script error occurred: Error: cannot allocate vector of size._

Case 1: External script execution for 'R' script ran out of resources

Server: Msg 39127, Level 16, State 20, Line 0 External script execution for 'R' script ran out of resources. External script request id is: 25D82ED9-3B30-456B-8DA1-898A4978B3B0.

A typical example of this would be, A user runs a query, this would result in new R/Python processes getting spawned. If there is not enough memory, these processes would crash and fail to come up and query would fail. In such a case windows job objects are notified that OOM occurred.

This is currently done with the help of Windows JobObjects. So if we allocate or reach the memory limits, we would get a Job object notification that we hit memory limits.

Case 2: cannot allocate vector of size X

A **'R' script error occurred during execution of 'sp_execute_external_script' with HRESULT 0x80004004. External script request id is 25D82ED9-3B30-456B-8DA1-898A4978B3B0. Server: Msg 39019, Level 16, State 1, Line 0 An external script error occurred:**
Error: cannot allocate vector of size 500.0 Mb

In these kind of cases where R/Python processes come up, but the script tries to allocate memory. Notice the " cannot allocate vector of size 500.0 Mb"

R/Python processes usually have error handling logic where checks are in place to make sure memory is not allocated if not enough memory is available. In such a case, R/Python process would return errors as part of their script itself and wouldn't allocate the memory. Since checks are done before allocating memory, the windows job objects would not get notified that we hit Out of memory. In such cases the above error message would not be returned to customer. But, the script output would show the error that there is not enough memory.

Investigation/Analysis

Check if the root cause is OOM

First get the session_id from the failure and run a query against the problematic server and database

```
MonExtensibilityTrace
| where LogicalServerName contains "januspremiumserver" and AppName contains "ebea7f0b4a82"
| where session_id contains "Input sessionId here"
| project TIMESTAMP, NodeName, session_id , error_code ,trace_message
| order by TIMESTAMP asc
```

How to view the OOM hits vs memory available for the sandboxes

```
MonExtensibilityJobObjectTrace
| where SubscriptionId == "Input SubscriptionId here"
| where trace_message contains "OOM" or trace_message contains "memory limit"
| extend OOM = case(trace_message contains "OOM", "true", "false")
| project TIMESTAMP, OOM, memory_limit_mb, AppName
| render scatterchart
```

Find the memory_usage and cpu_usage:

```
MonExtensibilityResourceUsage
| where AppName == "f23584189693"
| order by TIMESTAMP asc
```

Check if the OOM issue is due to Connection Object Leak OOM due to Connection Object Leak

Mitigation

Ask the customer to upgrade to higher Vcore SLO.
Example: if the customer is using Business Critical Gen5 (256 GB, 16 vCores), ask them to upgrade to 24 vcores (Higher than current vcores)

In case of a Service Level Objective (SLO) update, please create and ICM to re-enable the dedicated resource limits for R/Python. (Public Preview Limitation).

Queue Name : Extensibility (ML Services & Language Extensions on SQL Azure)

More Information

sys.dm_external_script_resource_usage_stats DMV is available to check if issues is live

You can run the following script for UsageStats

```
SET NOCOUNT ON;

DROP TABLE IF EXISTS UsageStats
GO

DECLARE @startTime datetime2(0) = GETDATE();
DECLARE @secondsToCapture INT = 600;  ---- 10 minutes capture window

CREATE TABLE UsageStats (
        timestamp DATETIME2,
        package_name NVARCHAR(100),
        memory_usage FLOAT,
        cpu_usage FLOAT
)

WHILE (GETDATE() < DATEADD(SECOND, @secondsToCapture, @startTime))
BEGIN
        INSERT INTO UsageStats (timestamp, package_name, memory_usage, cpu_usage)
        SELECT CURRENT_TIMESTAMP as timestamp, dmv.* FROM sys.dm_external_script_resource_usage_stats as dmv
        WAITFOR DELAY '00:00:00.500';
END

SELECT * FROM UsageStats
```

Public Doc Reference

It is not possible to limit R resources through Resource Governor and external resource pools.

During the public preview, R resources are set to a maximum of 20% of the SQL Managed Instance resources, and depend on which service tier you choose. For more information, see Azure SQL Database purchasing models.

Memory usage depends on how much is used in your R scripts and the number of parallel queries being executed. If you receive the errors above, you can scale your database to a higher service tier to resolve this.

https://docs.microsoft.com/en-us//azure/azure-sql/managed-instance/machine-learning-services-differences?view=sql-server-ver15#resource-governance

How good have you found this content?