Activity queue long time on MVNet IR

Last updated by | Veena Pachauri | Mar 8, 2023 at 10:31 PM PST

Contents

Problem Description

- Customer's external or pipeline activities queued for a long time (over 4 minutes) on Managed VNet IR
- Customer's copy activities queued for a long time on Managed VNet IR that has Copy TTL enabled.

Triage

Get the life cycle of the activity.

```
TaskHostingEvent
| where ActivityId == "{activityId}"
| where JobId != "00000000-0000-0000-0000-000000000000"
| distinct JobId
| take 1
| join TaskHostingEvent on JobId
| project-away JobId1
| project-reorder TIMESTAMP, LogId, Message, JobId, TaskId, QueueId
```

If there is a long-time gap between the log with LogId being `JobCreated` and its next log, go through the following cases to figure out the root case.

Case 1: Race condition of TTL

Get the non-zero queue id from the previous query. Then use it to run query:

```
let activityStartTime = "{activityStartTime}";
TaskManagementServiceActivity
| where TIMESTAMP between ((activityStartTime-70m)..75m)
| where PathAndQuery has "{queueId}"
| where RequestRoute == "tasks/next"
| where StatusCode != "0"
| summarize LastOK = maxif(TIMESTAMP, StatusCode== "OK"), LastNoContent = maxif(TIMESTAMP, StatusCode == "NoContent")
| extend IdleTime = LastNoContent - LastOK, NoLogTime = (activityStartTime+5m) - LastNoContent
```

If the value of `IdleTime` is larger than or equal to 1 hour, AND the value of `NoLogTime` larger than 3 minutes, this is customer's case.

Root cause

There is a race condition that a customer starts an activity just when the TTL is about to reach. We failed to start a new Batch node to handle the activity. We have a work item to improve it: [MVNet activity could queue long time due to race condition in TTL](#) ⧉

Mitigation

This is a transient issue. To prevent the same issue from happening again, please let customers set timeout and retry for activities. Retry can work for this case.

Case 2: Capacity issue

In MVNET IR, we have concurrency limit on pipeline activities (special limit on script activity) and external activities:

https://azure.microsoft.com/en-us/pricing/details/data-factory/data-pipeline/ ⧉

Type	Concurrency limit
Pipeline activities	50
External activities	800
Script activities	4 ~ 6

Type	Azure Integration Runtime Price	Azure Managed VNET Integration Runtime Price	Self-Hosted Integration Runtime Price
Orchestration[1]	$1 per 1,000 runs	$1 per 1,000 runs	$1.50 per 1,000 runs
Data movement Activity[2]	$0.25/DIU-hour	$0.25/DIU-hour	$0.10/hour
Pipeline Activity[3]	$0.005/hour	$1/hour (Up to 50 concurrent pipeline activities)	$0.002/hour
External Pipeline Activity[4]	$0.00025/hour	$1/hour (Up to 800 concurrent pipeline activities)	$0.0001/hour

Check the concurrency from below query or Orchestration.

```
TaskHostingEvent
| where Tenant == "AKS"
| where QueueId == "{queueId}"
| where TIMESTAMP between ((activityStartTime-1d)..2d)
| where LogId in ("JobCreated", "PulledOffNewTask", "MessageDeleted")
| summarize count() by bin(TIMESTAMP, 5m),LogId
| render timechart

CustomLogEvent
| where TIMESTAMP between ((activityStartTime-3h)..6h)
| where RoleInstance == "<tvm name>"
| where TraceMessage == "ExecutionTaskHostTaskCommon"
```

Case 3: Script activity stuck in queue when running on MVNET IR

Starting from 2023-01-04 12:00:00, we saw some machines dedicated for script activities (one machine per MVNET IR) entered unhealthy state after running for some days, causing the worker cannot send out any network requests or communicate with ADF services. This results in no workers available and scripts activities stuck in queue.

ICM: https://portal.microsofticm.com/imp/v3/incidents/details/361801476/home ⧉

We have taken the action to replace all the machines with a new set of machines and upgrade the OS version to latest version at 2023-01-19. This can mitigate all the impacted customers' issues. But not sure whether issue will reoccur as RCA is still unclear.

Symptom

See below error in runtime:

```
TraceMessage    Message
ExecutionTaskGetExecutionStatus <LogProperties><Text>Get execution status failed. Exception: StatusCode: 500. ErrorMessage: Time out to retry.. ActivityId:  e
```

```
CustomLogEvent
| where ActivityId == '<Activity RunId>'
| where Tenant != "AKS"
```

Sample run:

```
CustomLogEvent
| where ActivityId == 'e0e92dac-ebb3-48c8-95cc-c6829eb2f841'
| where Tenant != "AKS"
```

	Message	S
b2f841	<LogProperties><Text>Get execution status failed. RetryCount: 0, Exception: StatusCode: 500. ErrorMessage: Time out to retry.. ActivityId：e0e92dac-ebb3-48c8-95cc-c6829eb2f841.</Text></LogProperti...	0
b2f841	<LogProperties><Text>Get execution status failed. RetryCount: 1, Exception: StatusCode: 500. ErrorMessage: Time out to retry.. ActivityId：e0e92dac-ebb3-48c8-95cc-c6829eb2f841.</Text></LogProperti...	0
b2f841	<LogProperties><Text>Get execution status failed. RetryCount: 2, Exception: StatusCode: 500. ErrorMessage: Time out to retry.. ActivityId：e0e92dac-ebb3-48c8-95cc-c6829eb2f841.</Text></LogProperti...	0
b2f841	<LogProperties><Text>Get execution status failed. RetryCount: 3, Exception: StatusCode: 500. ErrorMessage: Time out to retry.. ActivityId：e0e92dac-ebb3-48c8-95cc-c6829eb2f841.</Text></LogProperti...	0
b2f841	<LogProperties><Text>Get execution status failed. Exception: StatusCode: 500. ErrorMessage: Time out to retry.. ActivityId：e0e92dac-ebb3-48c8-95cc-c6829eb2f841.</Text></LogProperties>	

Mitigation

It is due to the machine reserved for this MVNET IR entering unhealthy state.

1. We can ask customer to stop all the pipeline activities (Lookup, GetMetadata, Delete, Script, etc.). After TTL reached, he can restart these pipeline activities and a new machine will be assigned to his MVNET IR. Pipeline activity TTL explanation: https://learn.microsoft.com/en-us/azure/data-factory/managed-

virtual-network-private-endpoint#pipeline-and-external-activity ⧉

Edit integration runtime

Settings   **Virtual network**   Data flow runtime

Virtual network configuration ⓘ

◯ Disable   ⦿ Enable

Interactive authoring ⓘ

⦿ Disable   ◯ Enable

Copy compute scale ⓘ

◯ Disable   ⦿ Enable

Data integration unit ⓘ

| 4 DIU | ⌄ |

Time to live ⓘ

| 15 minutes | ⌄ |

Pipeline and external compute scale ⓘ

Time to live ⓘ

| 60 minutes | ⌄ |

2. If customer disagree mitigation #1 , you can wait for PG to drain the bad node automatically. Because RCA is still unknown and ADMS PG will have China new year holiday (2023-01-21 ~ 2023-01-27), we will scan the production **daily** to see whether the issue re-occurs and drain the bad nodes if finding anything. No worries, **we guarantee your customer's issue will be resolved in 1 day**.

New customer impact query:

```
CustomLogEvent
| where Message contains 'Get execution status failed. RetryCount: 3, Exception: StatusCode: 500. ErrorMessage: Time out to retry'
| where Tenant != "AKS"
| where TIMESTAMP >= ago(7d)
| summarize count() by bin(TIMESTAMP, 1h), SourceMoniker
```