ADF concurrent Mapping dataflow limit issue

Last updated by | Anil K B | Dec 6, 2022 at 1:28 AM PST

Contents

Symptom

- Customer is facing below error on Mapping Dataflow execution, throttling error:

```
{"Status":2,"IsUpdateQueuedStatus":false,"Output":null,"Error":{"ErrorCode":4502,"Message":"There are substant
```

Limitations on concurrent Mapping Dataflow execution:

There are two limits depending on which IR customer is using, **50 concurrent dataflows for normal Azure IR**, and **20 concurrent dataflows for Managed VNet IR**: https://docs.microsoft.com/en-us/azure/azure-resource-manager/management/azure-subscription-service-limits#azure-data-factory-limits

Request cx to follow Best practices:

Best practice is to spread out the executions to use multiple IR's so that limit is not exceeded, mentioned in this doc: https://docs.microsoft.com/en-us/azure/data-factory/data-flow-troubleshoot-errors#error-code-4502. Increasing limits can lead to other issues, so should first try recommending cx to follow below.

- If customer is running different dataflows in parallel, use a different IR for each dataflow.
- If customer is running single dataflow in foreach (and hitting this limit), they can run this sequentially instead or limit the batch count on the Foreach activity. Batch count is set to 50 by default.

Customer still wants to increase limit, collect below info and create ava post, get approval from dataflow product group before raising ICM.

1. First check from ActivityRuns table if they are really crossing the limit for concurrent runs or this is some bug they are facing.
2. Why do they want to increase the limit to? Do they have business justification for not following our recommended best practices?
3. Are they using normal Azure IR or Managed VNet IR in the dataflow?

4. How many cores is the IR in the dataflow using? https://supportability.visualstudio.com/AzureDataFactory/_wiki/wikis/AzureDataFactory/440659/How-to-get-number-of-cores-for-Dataflow-from-the-pipeline

Below is how to find the above information

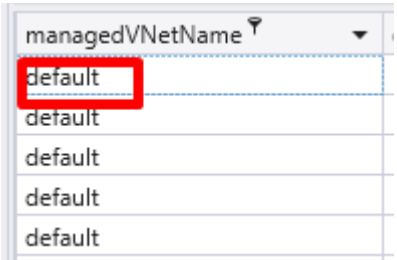
- First run below query (for next query), to get the timestamp when the issue started and pipeline detailed information like subscriptiond id, data factory name and integration runtime name which we can use for further investigation

```
ActivityRuns
| where pipelineRunId == "xxxxxxxxxxxx"
| where activityType == "ExecuteDataFlow"
```

- Then we can run the below query to see if they are using Managed VNet IR or normal Azure IR in their dataflow

```
ActivityRuns
| where subscriptionId == "xxxxxxxxxxxxxxxxxxxx"
| where dataFactoryName =~ "xxxxxxxxxxxxxxxxxxxx"
| where activityType == "ExecuteDataFlow"
| where effectiveIntegrationRuntime == "xxxxxxxxxxxxxxxxxxxxx"
| project-reorder managedVNetName, errorCode
| where TIMESTAMP >= datetime(xxxxxxxxx) and TIMESTAMP <= datetime(xxxxxxxxx)
| summarize count() by status
```

- Go through the managedVNetName column in the results, if the value is **default** that means it's managed vnet IR

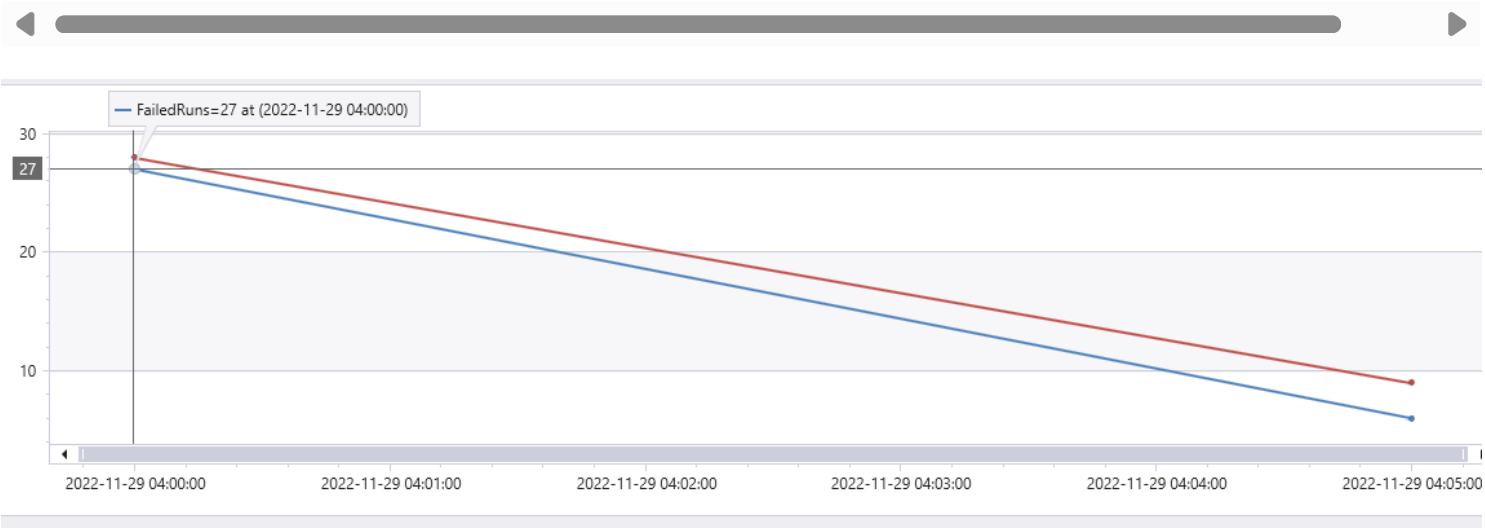


- All the info can be found from above query. Find the startTime and endTime of the dataflow activity above to use in below query (this is checking if customer had 50 or 20 dataflows in-progress or not at the time they got their error).

- If the total number of dataflows runs is less than the limit this would be a bug, if exceeding the limit then it’s valid error.

- Finally, we can use the time chart for validating whether the activities failing when the concurrent run number is out of the IR's limitation.

```
ActivityRuns
| where subscriptionId =~ "*****"
| where dataFactoryName =~ "*****"
| where activityType == "ExecuteDataFlow"
| where effectiveIntegrationRuntime contains "*****"
| where PreciseTimeStamp between(datetime(*****)..datetime(*****))
| summarize TotalRuns=dcount(activityRunId),FailedRuns=countif(status == "Failed") by bin(PreciseTimeStamp,2m)
//you can adjust the time range according to the failed activity start&end time for calculating the concurrent
| render timechart
```



To get number of cores cx is using with the IR:
https://supportability.visualstudio.com/AzureDataFactory/_wiki/wikis/AzureDataFactory/440659/How-to-get-number-of-cores-for-Dataflow-from-the-pipeline

How good have you found this content?