

# Case Study - Data factory cost is high

Last updated by | Jackie Huang | Jan 4, 2022 at 12:24 AM PST

## Contents

- Issue # 1
  - Root Cause
  - Investigation
- Issue # 2
  - Problem Statement
  - Root Cause

## Issue # 1

Customer complained that the Data Factory cost are really high. He's running a simple pipeline to move data with one activity move data from A to B. The pipelines didn't have any trigger & yet it seems to have been running every few seconds for over 2 weeks !.

Used the below query to check the usage report:

```
ExecutionApiCall
| where TIMESTAMP >= ago(15d)
| where DataFactoryName contains "<dataFactoryName>" and SubscriptionId has "xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx"
| project TIMESTAMP, ActivityId, ExecutionType, PipelineJobId, DataFactoryName, PipelineName, IntegrationRuntime
| join
(
BillingEvent
| where ResourceUri contains "dataFactoryName" and SubscriptionId has "xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx"
| project ActivityId, Quantity, MeterId, ResourceUri
) on $left.ActivityId == $right.ActivityId
```

We found that there are about 204150 runs. All the runs have occurred in the same pipeline "MoveFiles" and all had only GetMetadata Activity.

Also used below query to check if customer had any triggers

```
TriggerRuns
| where TIMESTAMP >= ago(15d)
| where subscriptionId contains "xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx"
```

Found that there were only 7 trigger runs. Customer has deleted the Data Factory now.

Customer would like to know the root cause for this error

## Root Cause

Looked into logs in the past 20 days, and found that there were 12 pipeline runs either failed or cancelled. Most of them ran for a long time before final state, e.g. 7 days. They are all for pipeline "MoveFiles", triggered by schedule trigger "triggerCopy". Before get to final state (failed or cancelled), all of them have validation activity executed for multiple times, each validation attempt results in a GetMetadata activity on ADMS side. Below logs shows all of the validation activity had lots of validation attempts and then timed out or cancelled.

Factory has been deleted so we cannot check pipeline definition. But the behavior shows customer has a validation activity with a bad config resulted in running validation frequently but never meet the criteria until timeout, and there is no timeout configured so default timeout (7 days) is used. Validation activity is for "to ensure the pipeline only continues execution once it has validated the attached dataset reference exists, that it meets the specified criteria, or timeout has been reached." Here is validation activity doc:











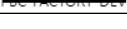

<https://docs.microsoft.com/en-us/azure/data-factory/control-flow-validation-activity> 

Customer has a pipeline triggered by schedule trigger. The pipeline has validation activity. For the investigated runs, validation activity run validation frequently until timed out or cancelled, the large amount of validation attempt caused high cost.

## Investigation

Pipeline runs in past 20days:

```
PipelineRuns
| where subscriptionId =~ "xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxxx"
| where status in ("Succeeded", "Failed", "TimedOut", "Cancelled")
| extend duration = end - start
| project TIMESTAMP, dataFactoryName, pipelineName, runId, start, end, duration, status, predecessors
```

TIMESTAMP	dataFactoryName	pipelineName	runId	start	end	duration	status	predecessors
2020-08-01 23:01:27.417723		MoveFiles	4464b65b-5ce4-43b7-8fca-1cc94cd82069	2020-07-25 23:00:01.3002781	2020-08-01 23:01:27.3761585	7.00:01:26.0758804	Failed	[{"Name": "trig
2020-08-02 23:01:18.1485829		MoveFiles	21b02c92-36a0-4f91-9b16-5f7d133ad443	2020-07-26 23:00:00.6029684	2020-08-02 23:01:18.1043156	7.00:01:17.5013472	Failed	[{"Name": "trig
2020-08-03 23:01:03.4888661		MoveFiles	f24e9805-89d3-4f96-88fe-b40bf6e21809	2020-07-27 23:00:01.1611211	2020-08-03 23:01:03.4346669	7.00:01:02.2735458	Failed	[{"Name": "trig
2020-08-04 23:00:58.7133565		MoveFiles	294e3afe-e32b-46dd-b2d7-e15b82759267	2020-07-28 23:00:00.3356061	2020-08-04 23:00:58.6753025	7.00:00:58.3396964	Failed	[{"Name": "trig
2020-08-05 23:01:12.1070344		MoveFiles	9ba8f7a7-e707-465c-9b45-601faffae624	2020-07-29 23:00:00.8017682	2020-08-05 23:01:12.0556423	7.00:01:11.2538741	Failed	[{"Name": "trig
2020-08-06 10:46:44.8443788		MoveFiles	744ee824-836e-4716-951d-e2bc44889f61	2020-08-04 23:00:01.1995359	2020-08-06 10:46:44.8062179	1.11:46:43.6066820	Cancelled	[{"Name": "trig
2020-08-06 10:46:44.8558902		MoveFiles	76e41705-54e8-47c3-9f31-cef1b05b16c5	2020-08-03 23:00:00.6608366	2020-08-06 10:46:44.8117885	2.11:46:44.1509519	Cancelled	[{"Name": "trig
2020-08-06 10:46:44.8571926		MoveFiles	a18c62ad-b244-4374-a55f-9ec1c5685045	2020-08-01 22:59:59.9495257	2020-08-06 10:46:44.8229560	4.11:46:44.8734303	Cancelled	[{"Name": "trig
2020-08-06 10:46:44.8599386		MoveFiles	39ef43e1-9926-4e9b-8b3b-bd435702518a	2020-07-30 23:00:00.1361399	2020-08-06 10:46:44.8208176	6.11:46:44.6846777	Cancelled	[{"Name": "trig
2020-08-06 10:46:44.8649771		MoveFiles	c1079140-1ab0-4cea-90d1-a7d8c7b17e1e	2020-08-02 23:00:00.7379436	2020-08-06 10:46:44.8230472	3.11:46:44.0851036	Cancelled	[{"Name": "trig
2020-08-06 10:46:44.8672533		MoveFiles	4825622e-ed9c-4fdd-803c-d7009064d770	2020-07-31 23:00:00.8341638	2020-08-06 10:46:44.8364807	5.11:46:44.0023169	Cancelled	[{"Name": "trig
2020-08-06 10:46:44.8931445		MoveFiles	8597b15d-d8f6-44a0-b652-e67b90add750	2020-08-05 23:00:00.8848181	2020-08-06 10:46:44.8510625	11:46:43.9662444	Cancelled	[{"Name": "trig

Validation activities status change:

```
ActivityRuns
| where subscriptionId =~ "xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxxx"
| summarize count() by dataFactoryName, pipelineName, pipelineRunId, activityRunId, activityType, status
```

dataFactoryName	pipelineName	pipelineRunId	activityRunId	activityType	status	count_
PROD-FACTORY-001	MoveFiles	c1079140-1ab0-4cea-90d1-a7d8c7b17e1e	f6e8dc74-8bb0-43f1-b9e6-7b902a05532b	Validation	InProgress	22661
PROD-FACTORY-001	MoveFiles	76e41705-54e8-47c3-9f31-cef1b05b16c5	3cf43a01-b19b-498e-a11e-b727e8973e4e	Validation	InProgress	16290
PROD-FACTORY-001	MoveFiles	a18c62ad-b244-4374-a55f-9ec1c5685045	2c36d431-0999-4eda-8b52-ffeffb58f7b5	Validation	InProgress	29782
PROD-FACTORY-001	MoveFiles	744ee824-836e-4716-951d-e2bc44889f61	0514241a-d1ac-4a22-b249-e7370c88b611	Validation	InProgress	9701
PROD-FACTORY-001	MoveFiles	39ef43e1-9926-4e9b-8b3b-bd435702518a	a7d98330-bc3d-40d9-8daf-ab9ed029ad63	Validation	InProgress	35842
PROD-FACTORY-001	MoveFiles	4825622e-ed9c-4fdd-803c-d7009064d770	9dbbca2c-d2b1-4a30-aad9-f0de5bbf9e47	Validation	InProgress	35591
PROD-FACTORY-001	MoveFiles	9ba8f7a7-e707-465c-9b45-601affae624	f891e25a-ae6e-482a-b749-7de3788ea3d7	Validation	InProgress	33330
PROD-FACTORY-001	MoveFiles	8597b15d-d8f6-44a0-b652-e67b90add750	b7b7be8a-867a-484e-9ac1-f79f56b34e14	Validation	InProgress	3224
PROD-FACTORY-001	MoveFiles	294e3afe-e32b-46dd-b2d7-e15b82759267	6bcb6ddc-202a-494e-b6ea-77c835de5e4e	Validation	InProgress	25725
PROD-FACTORY-001	MoveFiles	4464b65b-5ce4-43b7-8fca-1cc94cd82069	be8d7ef2-5172-47bc-9c26-5eaab7d76bcc	Validation	InProgress	6388
PROD-FACTORY-001	MoveFiles	f24e9805-89d3-4f96-88fe-b40bf6e21809	596469af-f541-468f-8dd2-525f6523aecd	Validation	InProgress	20129
PROD-FACTORY-001	MoveFiles	21b02c92-36a0-4f91-9b16-5f7d133ad443	4f8ac108-a56e-423b-b56e-7abd9cb8f093	Validation	InProgress	13477
PROD-FACTORY-001	MoveFiles	587b7154-6575-48fe-b9b7-daad002284a7	7256d923-a5c2-44cb-b525-7100fe0f5d10	Validation	InProgress	138
PROD-FACTORY-001	MoveFiles	294e3afe-e32b-46dd-b2d7-e15b82759267	6bcb6ddc-202a-494e-b6ea-77c835de5e4e	Validation	TimedOut	1
PROD-FACTORY-001	MoveFiles	9ba8f7a7-e707-465c-9b45-601affae624	f891e25a-ae6e-482a-b749-7de3788ea3d7	Validation	TimedOut	1
PROD-FACTORY-001	MoveFiles	f24e9805-89d3-4f96-88fe-b40bf6e21809	596469af-f541-468f-8dd2-525f6523aecd	Validation	TimedOut	1
PROD-FACTORY-001	MoveFiles	21b02c92-36a0-4f91-9b16-5f7d133ad443	4f8ac108-a56e-423b-b56e-7abd9cb8f093	Validation	TimedOut	1
PROD-FACTORY-001	MoveFiles	744ee824-836e-4716-951d-e2bc44889f61	0514241a-d1ac-4a22-b249-e7370c88b611	Validation	Cancelled	1
PROD-FACTORY-001	MoveFiles	a18c62ad-b244-4374-a55f-9ec1c5685045	2c36d431-0999-4eda-8b52-ffeffb58f7b5	Validation	Cancelled	1
PROD-FACTORY-001	MoveFiles	8597b15d-d8f6-44a0-b652-e67b90add750	b7b7be8a-867a-484e-9ac1-f79f56b34e14	Validation	Cancelled	1
PROD-FACTORY-001	MoveFiles	4825622e-ed9c-4fdd-803c-d7009064d770	9dbbca2c-d2b1-4a30-aad9-f0de5bbf9e47	Validation	Cancelled	1
PROD-FACTORY-001	MoveFiles	c1079140-1ab0-4cea-90d1-a7d8c7b17e1e	f6e8dc74-8bb0-43f1-b9e6-7b902a05532b	Validation	Cancelled	1
PROD-FACTORY-001	MoveFiles	76e41705-54e8-47c3-9f31-cef1b05b16c5	3cf43a01-b19b-498e-a11e-b727e8973e4e	Validation	Cancelled	1
PROD-FACTORY-001	MoveFiles	39ef43e1-9926-4e9b-8b3b-bd435702518a	a7d98330-bc3d-40d9-8daf-ab9ed029ad63	Validation	Cancelled	1
PROD-FACTORY-001	MoveFiles	4464b65b-5ce4-43b7-8fca-1cc94cd82069	be8d7ef2-5172-47bc-9c26-5eaab7d76bcc	Validation	TimedOut	1

#### ActivityRuns

```

| where subscriptionId =~ "xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx"
| where status == "TimedOut"
| extend duration = end - start
| project TIMESTAMP, dataFactoryName, activityName, activityType, start, end, duration, status

```

TIMESTAMP	dataFactoryName	activityName	activityType	start	end	duration	status
2020-08-01 23:01:24.4010208	PROD-FACTORY-001	Validation1	Validation	2020-07-25 23:00:03.2818690	2020-08-01 23:00:18.5777838	7.00:00:15.2959148	TimedOut
2020-08-02 23:01:15.6374036	PROD-FACTORY-001	Validation1	Validation	2020-07-26 23:00:01.4239336	2020-08-02 23:00:11.5979503	7.00:00:10.1740167	TimedOut
2020-08-03 23:01:01.3684919	PROD-FACTORY-001	Validation1	Validation	2020-07-27 23:00:04.8349235	2020-08-03 23:00:07.9590603	7.00:00:03.1241368	TimedOut
2020-08-04 23:00:56.2884329	PROD-FACTORY-001	Validation1	Validation	2020-07-28 23:00:01.7139682	2020-08-04 23:00:04.7978815	7.00:00:03.0839133	TimedOut
2020-08-05 23:01:09.6230407	PROD-FACTORY-001	Validation1	Validation	2020-07-29 23:00:03.6058679	2020-08-05 23:00:08.9719445	7.00:00:05.3660766	TimedOut

## Issue # 2

### Problem Statement

We got another [lcm](#)  where Product Group identified the high cost was due to the validation activity.

### Root Cause


Based on our logs, the customer is using Validation activities that ran for days. Validation activities will block pipeline execution until the conditions configured are met, as such every time the conditions are checked, the customer is billed 1 activity run.

Since these activities were configured to check every 10 seconds and timeout after 7 days (which are the defaults provided, but customer is free to configure as they wish), the customer was billed 1 activity run every 10 seconds for 7 days per validation activity.

The recommendation by product team is for customer to configure a small timeout for validation activities and to adjust the frequency of the checks as needed.

logs: <https://jarvis-west.dc.ad.msft.net/3F1E85E6> 

#### Additional Information:

- **Icm References:** [Icm Link](#) 
- **Author:** vimals
- **Reviewer:** vimals
- **Keywords:**

How good have you found this content?

