

# Spark cluster not found

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

---

## Contents

- [Issue](#)
- [Resolution](#)
- [Investigation](#)

## Issue

Dataflow job failing with the following error:

Execution failed. Error message:

```
'Type=Microsoft.DataTransfer.Execution.Core.ExecutionException,Message=Spark cluster not found,Source=Microsoft.DataTransfer.Executions.RuntimeTask,StackTrace= at Microsoft.DataTransfer.Execution.RuntimeTask.ActivityExecutorWorkerAdapter.<ExecuteAsync>d__1.MoveNext() --- End of stack trace from previous location where exception was thrown --- at System.Runtime.ExceptionServices.ExceptionDispatchInfo.Throw() at System.Runtime.CompilerServices.TaskAwaiter.HandleNonSuccessAndDebuggerNotification(Task task) at Microsoft.DataTransfer.Execution.RuntimeTask.MappingDataflowExecutionWorker.<ExecuteAsync>d__1.MoveNext(),'
```

## Resolution

This error is encountered when delete debug session is called by the user while the job is still being executed.

---

## Investigation

You can use the below query to find the cluster, activity details that have been utilized for the specific activity/pipeline.

```

let starttime = datetime(2020-09-01);
let endtime = datetime(2020-09-18);
cluster('adfcus.kusto.windows.net').database('AzureDataFactory').ActivityRuns
| union cluster('adfneu.kusto.windows.net').database('AzureDataFactory').ActivityRuns
| where category == "SandboxActivityRuns"
| where PreciseTimeStamp >= starttime and PreciseTimeStamp <= endtime
| where activityRunId == 'XXXXXXXXXXXXXXXXXXXXXXX'
//| where pipelineRunId == 'XXXXXXXXXXXXXXXXXXXXXXX'// contains 'XXXXXXXXXXXXXXXXXXXXXXX'
//| where pipelineName contains 'cosmostoparquetpipeline'
| where activityType contains 'dataflow'
| where status != 'Queued' and status != 'InProgress'
| join kind=leftouter (
    cluster('azuredmprod.kusto.windows.net').database('AzureDataMovement').TelemetryEvent
    | union cluster('azuredm.kusto.windows.net').database('AzureDataMovement').TelemetryEvent
    | where TelemetryId == "DataflowActivityRunInfo"
    | extend payload = parse_json(Payload)
    | extend runtimeMetrics = payload['RuntimeMetrics']
    | where isnotnull(runtimeMetrics)
    | project ActivityId, clusterid = tostring(runtimeMetrics['ClusterId'])
) on $left.activityRunId == $right.ActivityId
| distinct PreciseTimeStamp,start,end,pipelineRunId, activityRunId, status,clusterid, dataFactoryId,failureTyp

```

PreciseTimeStamp	start	end	pipelineRunId	activityRunId	status	clusterid	ComputeServiceType	datafactoryid	failureType	managedVN...	activityType	category
2020-09-16 18:10:19.1872986	2020-09-16 18:10:00.7901264	2020-09-16 18:10:19.0396678	d3f1a7af-45a4-41b0-ae49-196c29ef9461	c55252df-103b-4918-9f5b-e14e76745d29	Failed	0916-100608-pigs743	DataflowAzureDataBricksInteractiveCluster	0e13fcc7-9fca-4716-82c1-ae8454272c92	SystemError		ExecuteDataFlow	SandboxActivityRuns

Once we get the cluster id from the above query, can use below query to find the cluster alive time.

```

let starttime = datetime(2020-09-14);
let endtime = datetime(2020-09-18);
cluster('adfcus.kusto.windows.net').database('AzureDataFactory').AdfTraceEvent
| union cluster('adfneu.kusto.windows.net').database('AzureDataFactory').AdfTraceEvent
| where env_time >= starttime and env_time <= endtime
//| where DataFactoryId == "XXXXXXXXXXXXXXXXXXXXXXX"
| where Message contains 'XXXXXXXXXXXXXXXXXXXX-pigs743'
| summarize clusterstarttime = min(env_time), clusterendtime = max(env_time), Clusteralivetime = max(env_time)-

```

clusterstarttime	clusterendtime	Clusteralivetime
2020-09-16 10:06:08.3343347	2020-09-16 18:15:26.2115525	08:09:17.8772178

### Additional Information:

- Icm References: [Icm Link](#) [Icm Link](#)

How good have you found this content?

