

Dataflow activity – get spark cluster settings

Last updated by | Nelson Gomes | Dec 28, 2022 at 12:27 PM PST


Contents

- [Subject](#)
- [Steps](#)

Subject

Getting Dataflow Spark Cluster properties/settings

Steps

When facing dataflow long runs and failures due to OOM it is important to understand the cluster settings. You can use the following query to get spark cluster setting for a specific dataflow run. If you want to understand the meaning of a spark setting you can check [spark org link](https://spark.apache.org/docs/latest/configuration.html#application-properties) 

```
let dataFlowRunId = '5df6f5fa-3b23-4643-b111-115470983f7f';
cluster('adfneu.kusto.windows.net').database('AzureDataFactory').DataflowClusterLogs | where ActivityRunId == d
| extend sparkconfig=replace_string(replace_string(Message,'Initial spark configuration: ',''), '('','')
| extend sparkconfig=split(sparkconfig,',')
| mv-expand streams=sparkconfig
| extend streams=toString(streams)
| extend streams=replace_string(streams,'"','')
| extend countComma=countof(streams,',')
| where countComma==1
| extend sparkProperty=replace_string(extract("(?(s).*)", 0, streams),',','')
| extend sparkPropertyValue=replace_string(extract("(,(?s).*)", 1, streams),',','')
| extend link="https://spark.apache.org/docs/latest/configuration.html#application-properties"
| project sparkProperty,sparkPropertyValue,link
```

You can for instances check properties such as: **spark.driver.cores**, **spark.executor.cores**, **spark.driver.maxResultSize**, **spark.driver.memory**...

sparkProperty	sparkPropertyValue
spark.advise.nameToClass.TaskError	com.microsoft.impulse.insights.TaskErrorInsight
spark.org.apache.hadoop.yarn.server.webproxy.amfilter...	http://vm-2a675086:8088/proxy/application_1670855540958_0001
spark.executor.cores	4
spark.arcadia.session.token	eyJhbGciOiJodHRwOi8vd3d3LnczLm9yZy8yMDAxLzA0L3htbGRzaWctbW9yZS5nc2Etc2hhMjU2liwia2lkjoiNzI3MTJDMzJBQUYxMkMyNjUzRjc
spark.history.fs.cleaner.interval	30d
spark.synapse.dagnostic.builtinEmitters	ShoeboxEmitter
spark.yarn.preserve.staging.files	false
spark.sql.crossJoin.enabled	true
spark.driver.maxResultSize	4096m
	6R70r3e9-ff4R-4r1e-2df7-4qR9Q5270Rr0

Additional information

IcM reference:

- **Author:** negome
- **Reviewer:**

How good have you found this content?

