Find Cluster version of an activity run in Monitoring output

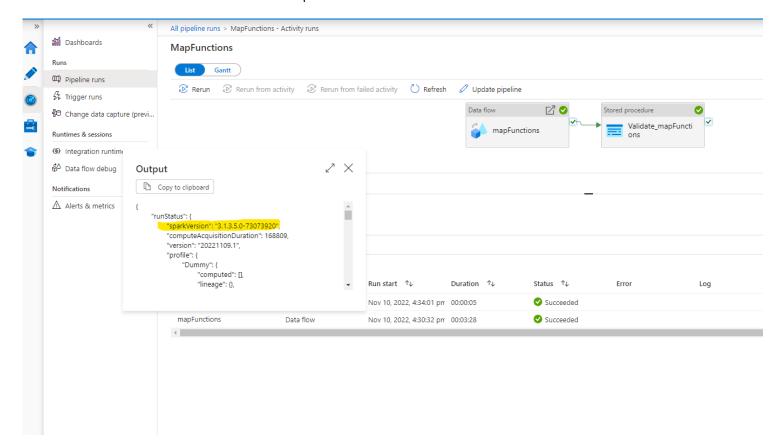Last updated by | Ranjith Katukojwala | Nov 23, 2022 at 1:44 PM PST

Contents

Issue

How Users can now figure out which version of spark was used to run the dataflow.

Recommendation:

The version is now present as part of the monitoring output.

The version is currently displayed for all the activities that used Synapse Spark 3.x version. The absence of the version could be deemed to have run on Spark version 2.4.4.

However, in the future, PG team could improve on this to provide spark versions irrespective of old or newer spark versions

Kusto Queries :

```
cluster('adfcus.kusto.windows.net').database('AzureDataFactory').DataflowClusterLogs
| union cluster('adfneu.kusto.windows.net').database('AzureDataFactory').DataflowClusterLogs
| where ActivityRunId == actId
| extend SparkConf = iff(Message contains "Initial spark configuration: ", Message, "")
| extend SparkVersion = extract("/var/lib/(.*)/shs_db", 1, SparkConf)
```

```
cluster('adfneu.kusto.windows.net').database("AzureDataFactory").DataFlowTelemetryEvent
| union cluster('adfcus.kusto.windows.net').database("AzureDataFactory").DataFlowTelemetryEvent
| extend ActivityId = extract("runId\":\"([0-9a-z-]+)", 1, Event)
| where ActivityId == actId
| extend SparkVersion = iff(extract("sparkVersion\":\"([0-9a-z-.]+)", 1, Event) == "3.1.2", "3.1.2", "2.4.4")
```

◀ ▶

How good have you found this content?