How to query ADF Pipeline Run ID in Databricks Job submitted using Databricks Notebook Activity

Last updated by | Ranjith Katukojwala | Mar 7, 2023 at 11:35 AM PST

Contents

Scenario

Databricks activities created by the ADF pipeline has job name in the format of: 'ADF_(data factory name) *(pipeline name)*(activity name)_(activity run id)'. It doesn't give information on Pipeline Run ID. Customer wanted to find the Pipeline RunID in Databricks itself without having to go to Monitor screen in ADF and check for RunID of the Pipeline that submitted a particular Databricks Job.
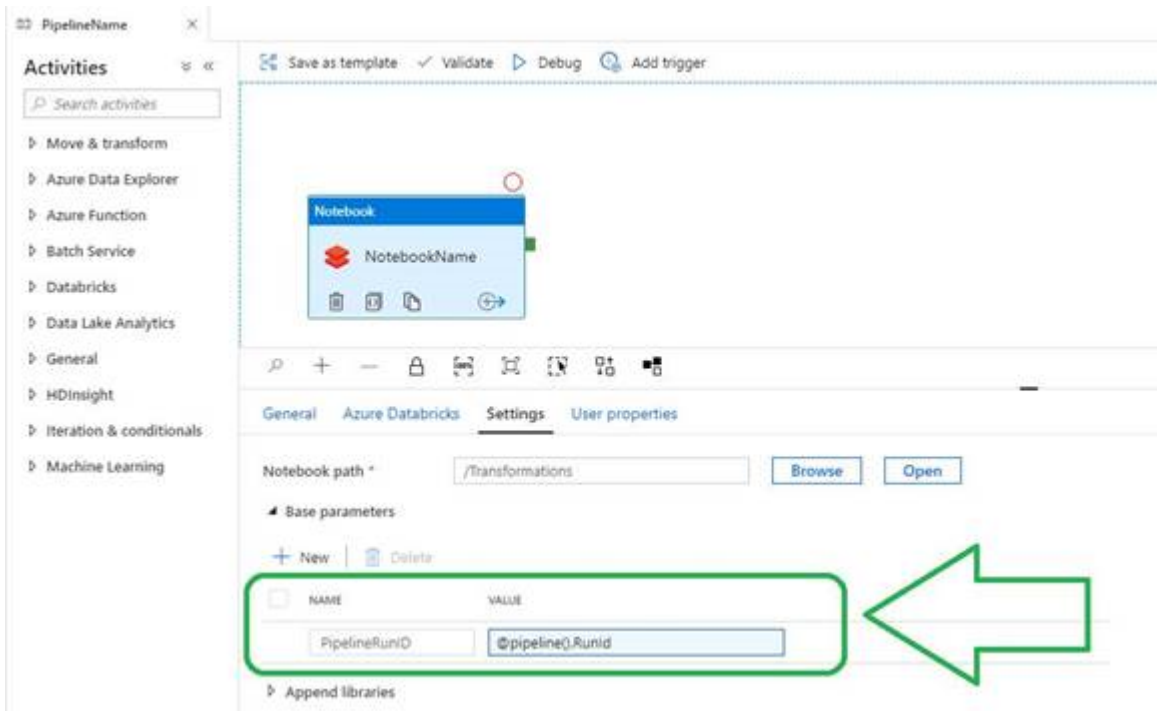
Research

In the process of providing resolution to the customer, we observed:

- We can't find the Pipeline RunIDs by Activity RunIDs using REST API
- We can only find details of a Pipeline Run using Pipeline RunID or DataFactory name.
- We cannot modify the name of the Job in Databricks. ADF_(data factory name)*(pipeline name)*(activity name)_(activity run id) . This format is by design.
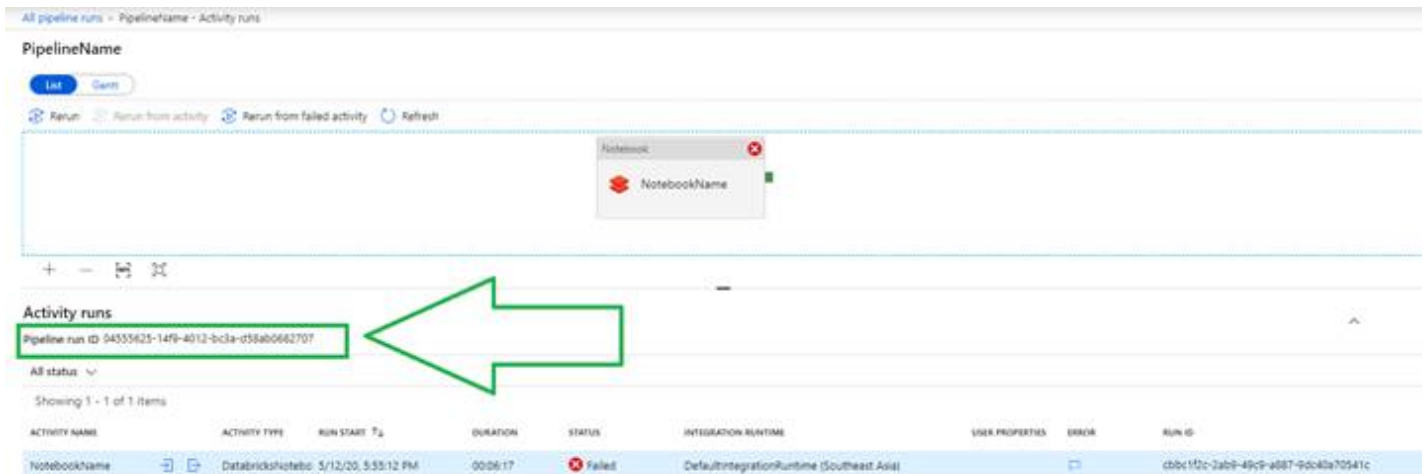- It is not possible to change the format of Databricks Job Name to include Pipeline Run Id

Resolution

Steps to get the ADF Pipeline Run ID in Databricks Job submitted using Databricks Activity

1. You can get the Pipeline RunID in Databricks Job by defining a Base Parameter called PipelineRunID using Dynamic content value as **@pipeline().RunId** in the Databricks Notebook activity as in the screenshot below:

2. Pipeline RunID is 04555625-14f9-4012-bc3a-d58ab0682707 for the pipeline run:
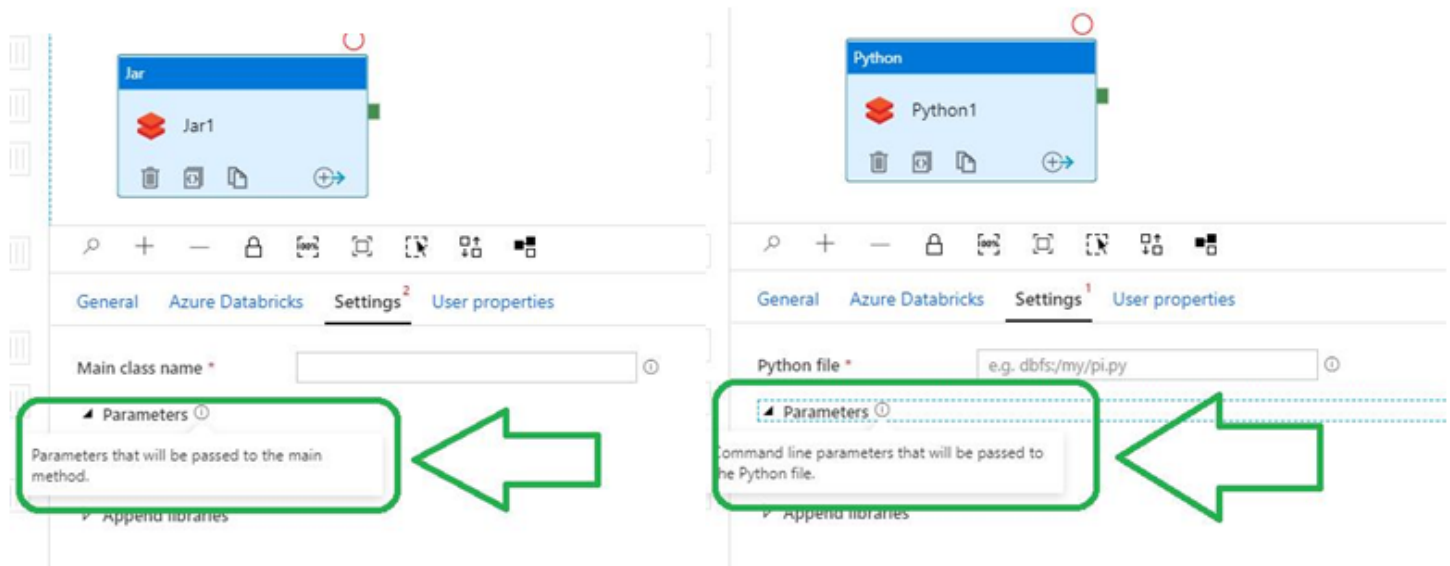

3. We can get the same Pipeline RunID 04555625-14f9-4012-bc3a-d58ab0682707 in the parameters of the Job in Databricks without checking it in ADF monitoring page:

NOTE:

- IMPORTANT!!! -- The Solution above works only for Notebook activity. It doesn't work for Jar and Python Activities for Databricks.
- In case of Jar activity, the Parameters are "Parameters that will be passed to the Main method." These parameters will be passed to Code, not to the Databricks Job. So we can't pass the RunID to the Databricks Job in case of Jar activity.
- In case of Python activity, the Parameters are "Command line Parameters that will be passed to the Python file." These parameters will be passed to Python file, not to the Databricks Job. So we can't pass the RunID to Databricks Job in case of Python activity.

ADDITIONAL INFORMATION:

Author: v-goman

Reviewer: vimals

Keywords:

How good have you found this content?