Databricks job running for more than an hour failing

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

Contents

- Issue
- Workaround
- Resolution

Issue

If a CRI comes in with task failure after running for a long time (> 1 hr), it may be due to the Databricks cluster running out of disk space.

This is a regression introduced on 1/13, and fixed with the PR on 1/17 (so should be rolled out by end of Jan).

Workaround

Add the newClusterEnableElasticDisk property to LinkedService, as follows:(for new cluster)

```
"name": "...databricks",
"type": "Microsoft.DataFactory/factories/linkedservices",
"properties": {
    "type": "AzureDatabricks",
    "typeProperties": {
        "domain": "https://westus.azuredatabricks.net",
        "newClusterNodeType": "Standard_DS3_v2",
        "newClusterNumOfWorker": "1",
        "newClusterVersion": "4.2.x-scala2.11",
        "newClusterEnableElasticDisk": true,
        "encryptedCredential": "..."
    }
}
```

Resolution

Permanent solution will be available once the fix is deployed in production(end of jan) until then please follow above workaround

Additional Information:

Icm References:

• Author: muralird **Reviewer**: vimals

• Keywords:

How good have you found this content?



