

# Spark Activity Spark job failed, batch id

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

---

## Contents

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

## Issue

Spark activity run failed with error like:

"Activity SparkActivity failed: Spark job failed, batch id:10611"

And if you go to the activity run page, you will see error like:

```
{
  "errorCode": "2312",
  "message": "Spark job failed, batch id:10611",
  "failureType": "UserError",
  "target": "SparkActivity"
}
```

## Cause

The error message with spark batch Id, means the job was submitted to the cluster successfully by ADF, but the job run failed. Usually there might be following possible issues:

1. Customer's spark program has bugs, it can't be run in the cluster. This is applicable to most of the customers, especially for those who are in development phase of their pipelines.
2. The cluster is in bad state. This is only true for a few customers. If a customer already deployed their pipeline, and it works well but suddenly it doesn't work, then the possibility of this issue is relatively higher, but it still can't rule out customer's spark program run failed due to other reasons like data changed, etc.

## Resolution

We can find out the real cause (one of the two above) by checking the spark job logs, please follow the steps below to do so:

- Enable debug info.
- We enable debug info by setting "getDebugInfo" to "Always" or "Failure" in the spark activity. Please refer to document [here](#).

The screenshot displays the Azure Data Factory (ADF) Spark Activity configuration interface. The left sidebar shows the 'Factory Resources' tree with 'Pipelines' and 'Datasets'. The main pane shows the 'SparkJarPipeline' selected, with a 'Batch Service' activity. The 'Script / Jar' tab is active, showing fields for 'File Path', 'Class Name', 'Proxy User', and 'Debug Information'. The 'Debug Information' dropdown is set to 'Always' and is circled in blue. A 'Spark' tooltip is visible in the top right corner.

- Run the activity again and check both livy & yarn log of the spark activity under the log directory.
- The log directory is in the storage linked to the HDI cluster linked service, and we can find the path by checking the "logLocation" property of the activity output.

shawnxiadf2 | Monitor Pipeline Runs ▾

Pipelines / SparkJarJob

Refresh

Activity Runs

Pipeline Run ID a8b5571f

All Succeeded In Progress

Activity Name

SparkActivity

HDInsightSpark

04/11/2018, 2:04:10 PM

00:02:26

Failed

DefaultIntegrationRuntime (East US)

d003a317-3389-43d4-b118-60db98d33100

Output

```
{
  "computeInformation": {
    "https://sparkfunctionaltest.azurehdinsight.net/",
    "batchid": "10609",
    "ExecutionProgress": "Failed",
    "logLocation": "https://hdinsightsa.blob.core.windows.net/shawnxiadf2/SparkJarJob/log/d003a317-3389-43d4-b118-60db98d33100/"
  },
  "jobid": "application_1521152740094_4321",
  "effectiveIntegrationRuntime": "DefaultIntegrationRuntime"
}
```

Activity Name	Duration	Status	Integration Runtime	RunID
SparkActivity	00:02:26	Failed	DefaultIntegrationRuntime (East US)	d003a317-3389-43d4-b118-60db98d33100

- Alternatively, once you see the jobid in the Output, you can open the Yarn UI from Ambari, find the application there and check the logs. If the is in VNet, you might have to do it from a VM which is inside the same vnet/subnet since the public URI for HDI cluster might not be accessible from public domain.
- Given the log, if it is the issue of your spark program, you need to fix it. If it is the issue of the cluster, please assign/file incident to HDI team.

### Additional Information:

- ADF uses Livy API to manage spark applications on HDI cluster, we can also help customer manually do this to debug possible application issues before use ADF to orchestrate the real runs.
- Document on how to use Livy API on HDI cluster: <https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-livy-rest-interface> ☑
- Livy API spec: <https://docs.microsoft.com/en-us/rest/api/hdinsightspark/hdinsight-spark-batch-job> ☑
- Usually you can call the Livy API to submit your spark job, and check in yarn resource manager UI of the job details.
- If you don't see corresponding yarn application, it should fail in the Livy layer, you call the get log API in Livy to understand the reason: <https://docs.microsoft.com/en-us/rest/api/hdinsightspark/hdinsight-spark-batch-job#get-the-full-log-of-a-batch-job> ☑

### Icm References:

- Author:** hvermis
- Reviewer:** vimals
- Keywords:**

### How good have you found this content?

