Could not execute broadcast in 60 secs

Last updated by | Jackie Huang | Jan 4, 2022 at 12:24 AM PST

Contents

- Issue
- Root Cause
- Workaround

Issue

Customer seeing the following error while executing Data flow jobs:

Caused by: org.apache.spark.SparkException: **Could not execute broadcast in 60 secs.** You can increase the timeout for broadcasts via spark.sql.broadcastTimeout or disable broadcast join by setting spark.sql.autoBroadcastJoinThreshold to -1 at org.apache.spark.sql.execution.exchange.BroadcastExchangeExec.doExecuteBroadcast(BroadcastExchangeExec.scala:190)

Root Cause

This error is caused due to broadcast join timeout issue. Customer was using debug session to execute their Data flow job, the timeout will be 60 secs instead of 300 secs for job cluster. It could be either because of the data increased on customer's source or the network communication got slow for the sink.

Workaround

Choosing "use activity runtime" option when starting a debug run or use "trigger now", it will create a cluster with 300 seconds timeout setting instead of 60 seconds timeout setting of broadcast join which should help on this case.

Additional Information:

- **Icm References:** Icm Link ⤢
- **Author:** vimals
- **Reviewer:** vimals
- **Keywords:**

How good have you found this content?