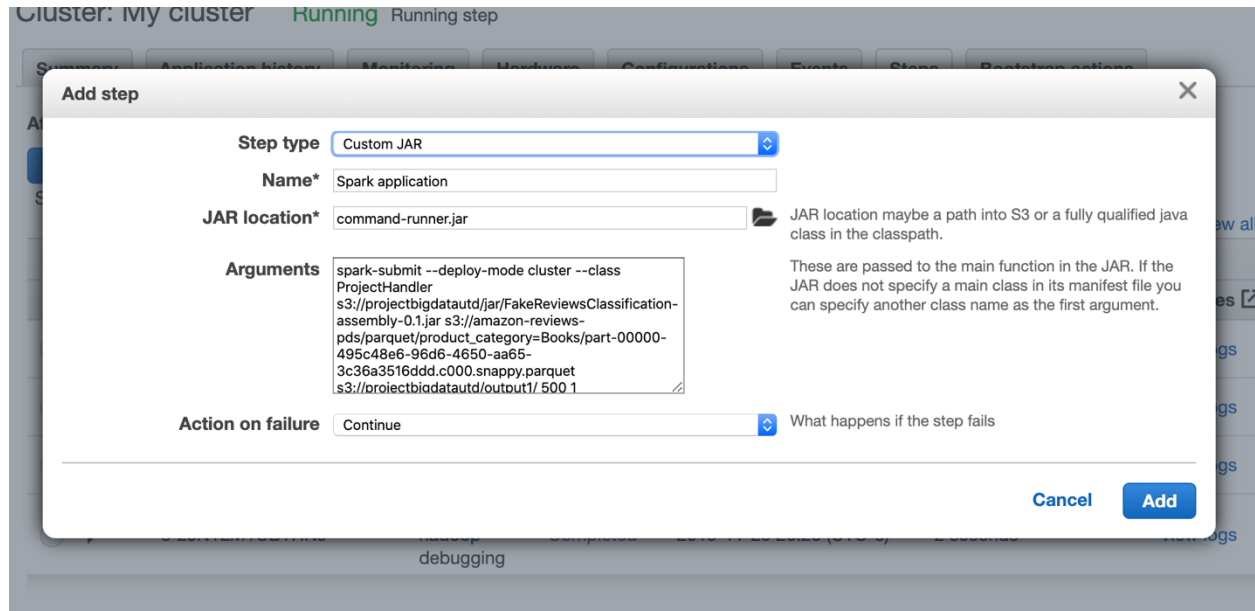# //Anish Hegde, Shobhit Jain, Arihant Chhajed, Pawan Vaidya

Readme

1. Import the project in IntelliJ IDEA
2. Go to SBT and click on sbt assembly to generate the JAR
3. Open AWS console and navigate to EMR
4. Create a cluster with spark 2.4.0
5. Add a spark step for spark submit

In our case the parameters are

*spark-submit --deploy-mode cluster --class ProjectHandler s3://projectbigdatautd/jar/FakeReviewsClassification-assembly-0.1.jar s3://amazon-reviews-pds/parquet/product_category=Books/part-00000-495c48e6-96d6-4650-aa65-3c36a3516ddd.c000.snappy.parquet s3://projectbigdatautd/output1/ **10000 1***