**Read Hive table data using Spark-submit and livy**

**Pre-requisites:**

* Install Git
* Install Maven
* Need EMR Cluster with following connections

Hive, Livy, Spark, Scala

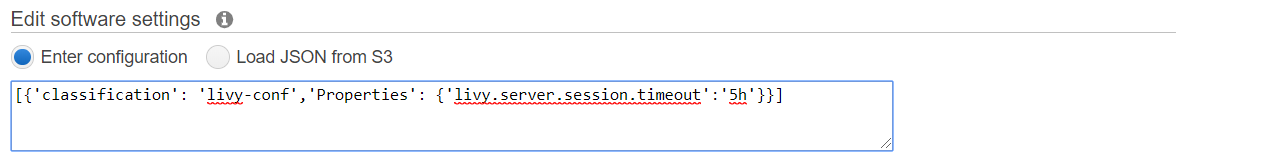
**Step:1**

Create EMR Cluster with following connections

Hive, Livy, Spark, Scala

While creating EMR Cluster we need add below configuration in Edit settings.

[{'classification': 'livy-conf','Properties': {'livy.server.session.timeout':'5h'}}]



**Step:2**

git clone <https://github.com/Naresh240/ReadHiveData--SparkSubmit-Livy.git>

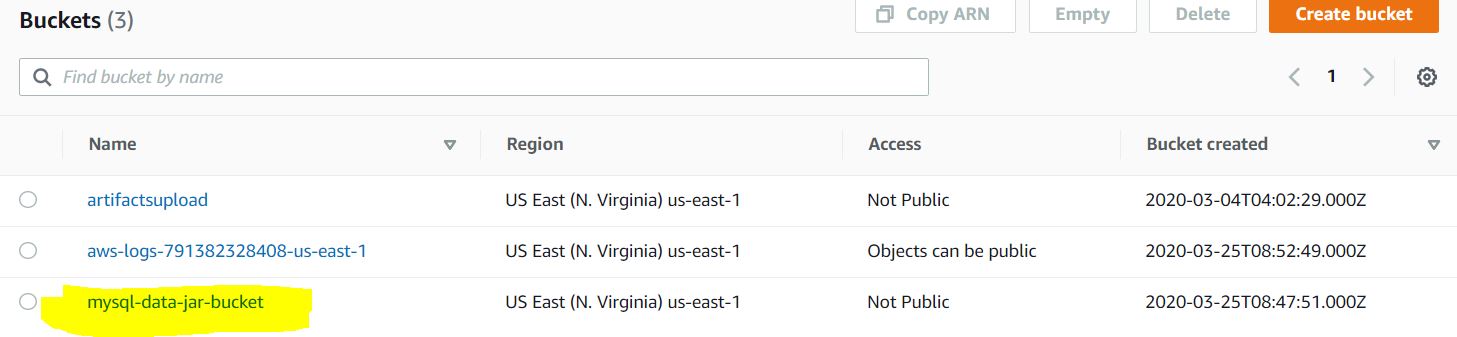
**Step:3**

cd ReadHiveData--SparkSubmit-Livy

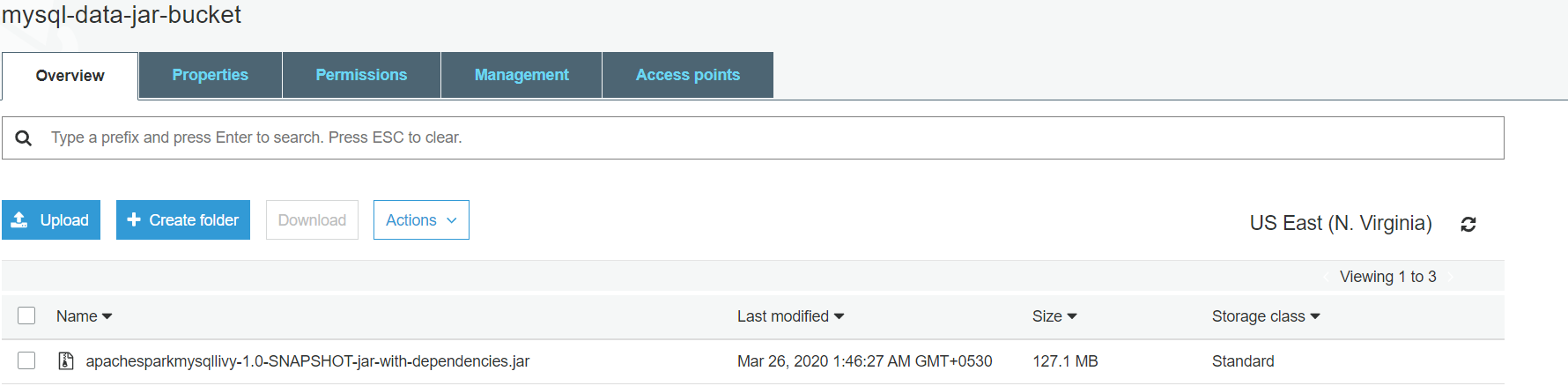
mvn clean install

**Step:4**

Create S3 bucket

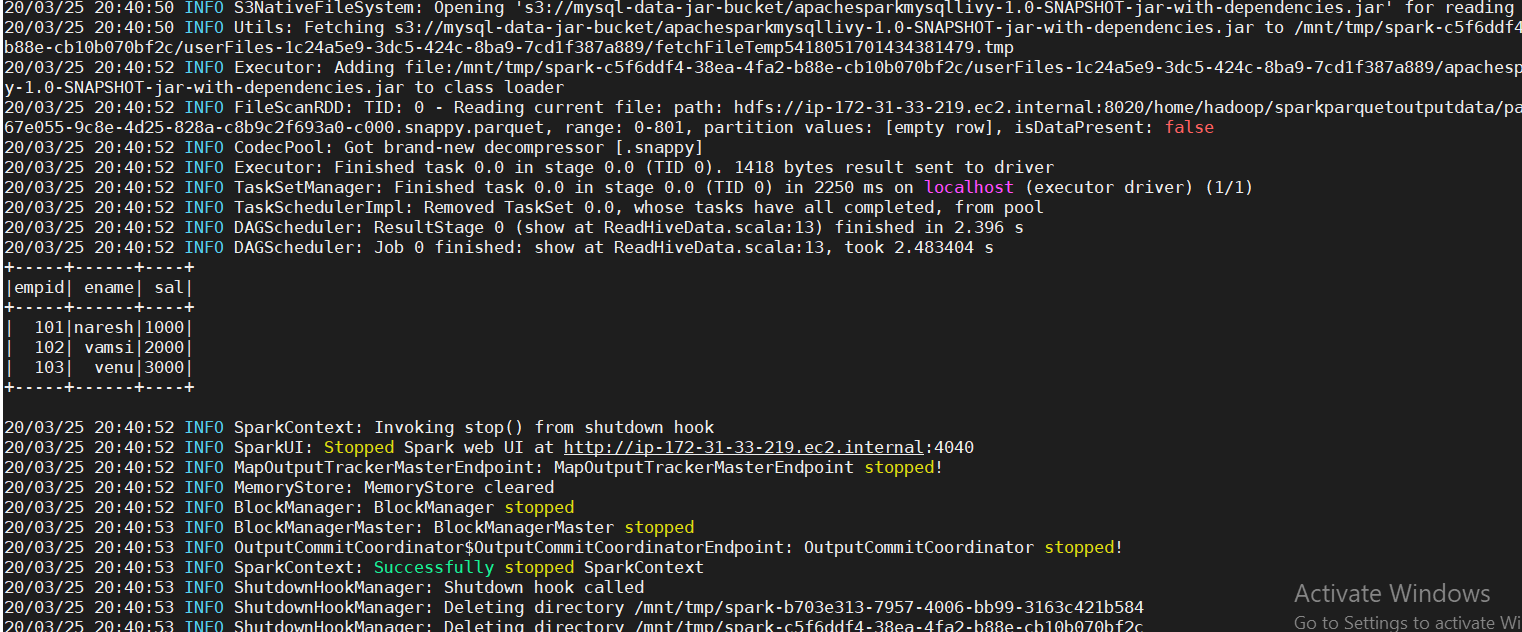


Copy fat jar to S3:bucket



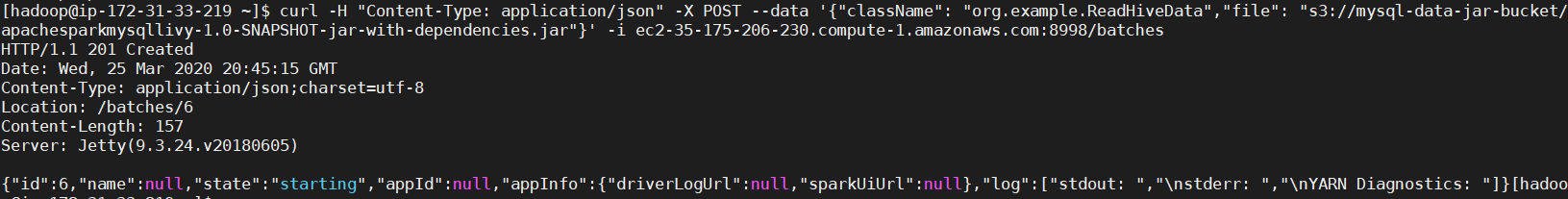
**Step:5** - Sumbit spark job using below command

spark-submit --master yarn --jars /usr/lib/hadoop-lzo/lib/\*:/usr/lib/hadoop/hadoop-aws.jar:/usr/share/aws/aws-java-sdk/\*:/usr/share/aws/emr/emrfs/conf:/usr/share/aws/emr/emrfs/lib/\*:/usr/share/aws/emr/emrfs/auxlib/\*:/usr/share/aws/emr/security/lib/\* --conf spark.executor.memory=4g --conf spark.executor.cores=2 --conf spark.executor.instances=12 --conf spark.sql.warehouse.dir=/home/hadoop --class org.example.ReadHiveData s3://mysql-data-jar-bucket/apachesparkmysqllivy-1.0-SNAPSHOT-jar-with-dependencies.jar



**Step:6**

curl -H "Content-Type: application/json" -X POST --data '{"className": "org.example.ReadHiveData","file": "s3://mysql-data-jar-bucket/apachesparkmysqllivy-1.0-SNAPSHOT-jar-with-dependencies.jar"}' -i ec2-35-175-206-230.compute-1.amazonaws.com:8998/batches



**Checking for livy state Starting:**

curl http://ec2-35-175-206-230.compute-1.amazonaws.com:8998/batches/6



**check logs for livy:**

curl http://ec2-35-175-206-230.compute-1.amazonaws.com:8998/batches/6/log

