# Data Movement Engine – Part 3 - DEMO

## Setting up param file for S3 as source and MYSQL DB as target:

As I have used TMDB movie.json file from their official website, I have placed the json file in S3 bucket.

1. Create a table in MYSQL :

CREATE TABLE IF NOT EXISTS movie(

adult BOOLEAN NOT NULL,

id INT NOT NULL,

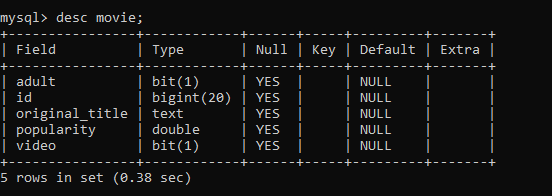
original\_title VARCHAR(300) NOT NULL,

popularity DOUBLE(10,3),

video BOOLEAN NOT NULL,

PRIMARY KEY (id)

);



1. prepare param file :

|  |
| --- |
| {  "sourceType": "S3",  "sourceAccessKey": "AKIAIOLTF3EUFVNI6PWA",  "sourceFileName": "movie\_ids\_11\_17\_2019",  "sourceSecretKey": "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",  "sourceBucketPath": "dataanalysis100",  "sourceFileFormat": "json",  "sourceMultiLineFlag": "false",  "targetType": "MYSQL",  "targetIP": "172.22.21.104",  "targetPort": "3306",  "targetSchemaName": "mydb",  "targetTableName": "movie",  "targetUsername": "root",  "targetPassword": "password",  "targetSaveMode": "overwrite"  } |

1. Download jar files:

slf4j-api-1.7.21.jar

commons-csv-1.3.jar

log4j-1.2.17.jar

mysql-connector-java-8.0.18.jar

hadoop-aws-2.6.0.jar

aws-java-sdk-1.7.4.jar

1. Create a log file Directory
2. Prepare a log4j.properties file :

To log the details

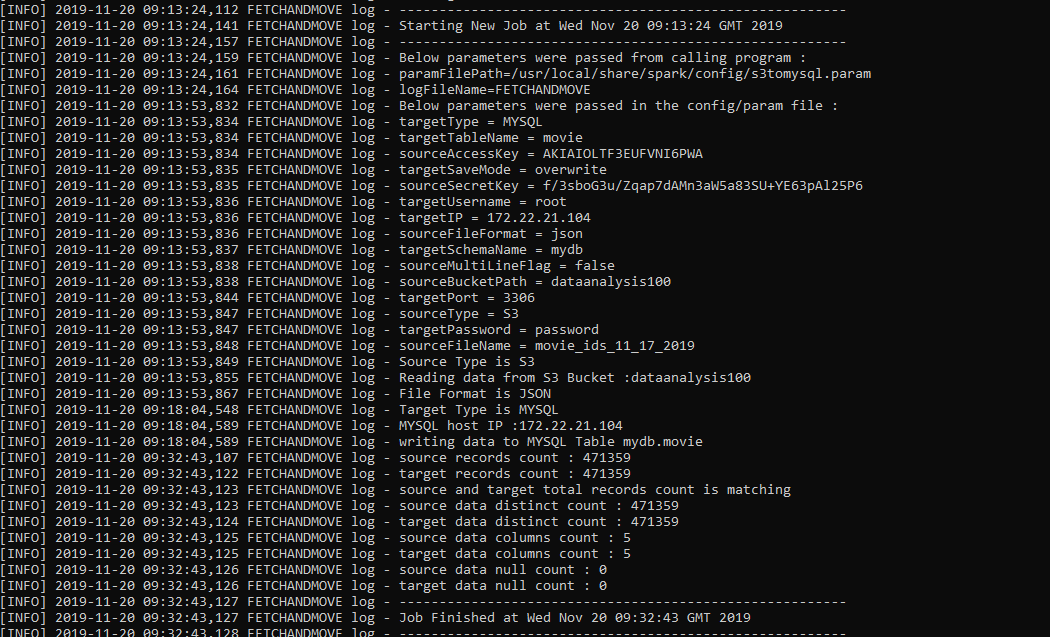
My log4j.properties looks like this

|  |
| --- |
| log4j.rootCategory=INFO, console  log4j.appender.console=org.apache.log4j.ConsoleAppender  log4j.appender.console.target=System.err  log4j.appender.console.layout=org.apache.log4j.PatternLayout  log4j.appender.console.layout.ConversionPattern=%d{yy/MM/dd HH:mm:ss} %p %c{1}: %m%n  #Settings to quiet third party logs that are too verbose  log4j.logger.org.spark-project.jetty=WARN  log4j.logger.org.spark-project.jetty.util.component.AbstractLifeCycle=ERROR  log4j.logger.org.apache.spark.repl.SparkIMain$exprTyper=INFO  log4j.logger.org.apache.spark.repl.SparkILoop$SparkILoopInterpreter=INFO  # ============== FETCHANDMOVE rows file appender ==============  log4j.appender.FETCHANDMOVE\_Log\_Appender=org.apache.log4j.DailyRollingFileAppender  log4j.appender.FETCHANDMOVE\_Log\_Appender.File=/usr/local/share/spark/log/FETCHANDMOVE/FETCHANDMOVE.log  log4j.appender.FETCHANDMOVE\_Log\_Appender.DatePattern='.'yyyy-MM-dd  log4j.appender.FETCHANDMOVE\_Log\_Appender.layout=org.apache.log4j.PatternLayout  log4j.appender.FETCHANDMOVE\_Log\_Appender.layout.ConversionPattern=[%p] %d %c %M - %m%n  log4j.logger.FETCHANDMOVE=INFO, FETCHANDMOVE\_Log\_Appender  #Disable additivity for all class loggers  log4j.additivity.FETCHANDMOVE\_Log\_Appender=false |

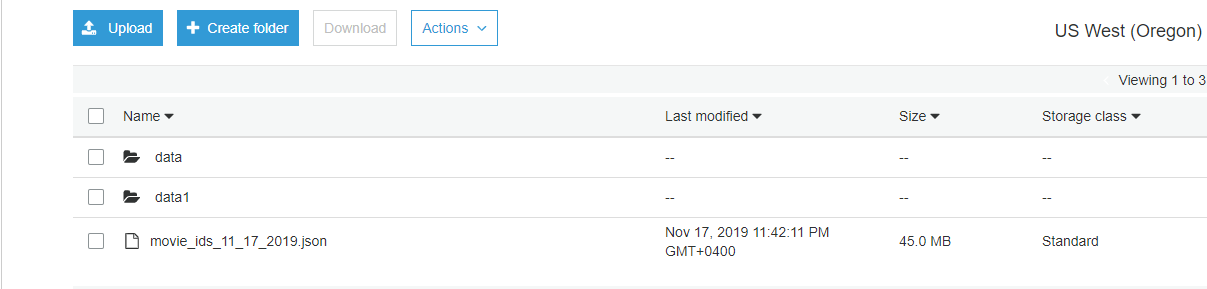
Run Command:

|  |
| --- |
| spark-submit --jars /usr/local/share/spark/JARS/mysql-connector-java-8.0.18.jar,/usr/local/share/spark/JARS/hadoop-aws-2.6.0.jar,/usr/local/share/spark/JARS/aws-java-sdk-1.7.4.jar,/usr/local/share/spark/JARS/log4j-1.2.17.jar,/usr/local/share/spark/JARS/commons-csv-1.3.jar,/usr/local/share/spark/JARS/slf4j-api-1.7.21.jar --class com.spark.movedata.FetchAndMoveData --files /usr/local/share/spark/log4j.properties --conf spark.storage.memoryFraction=0.6 --conf spark.shuffle.memoryFraction=0.2 --conf spark.yarn.executor.memoryOverhead=4048 --conf spark.yarn.am.memory=5g --conf spark.driver.extraJavaOptions='-Dlog4j.configuration=file:/usr/local/share/spark/log4j.properties' --executor-memory 10g --driver-memory 10g /usr/local/share/spark/JARS/FetchAndMoveData-0.0.1-SNAPSHOT.jar /usr/local/share/spark/config/mysqltos3.param FETCHANDMOVE |

Outputs:



Source:



Target:

