**STEP to Merge and save output from S3 bucket to S3 bucket using EMR**

inpath="s3://projectdatagroup1/DatafromEC2/"

>>> df = spark.read.option("header", "true").csv(s3\_path)

>>>df.show()

>>> outpath='s3://projectdatagroup1/MergedOutput/'

>>> df.write.mode("overwrite").csv(outpath, header=True)

# Count NULLs in each column

null\_counts = df.select([sum(when(col(c).isNull(), 1).otherwise(0)).alias(c) for c in df.columns]) # Show result null\_counts.show()

>>> df = df.dropna()

>>> df.show()

>>> df.coalesce(1).write.mode("overwrite").parquet(outpath)