**Performance Comparison of AWS S3 EMR and AWS EMR HADOOP Analysis:**

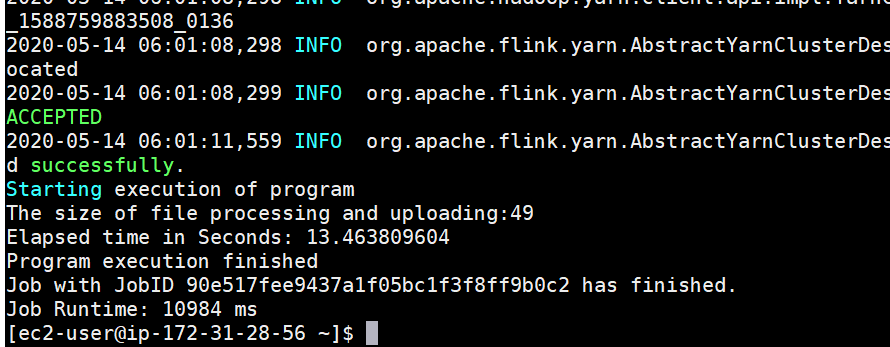
**Analysis Method used for implementation:**

**1)Using AWS EMR CLI – to access the HDFS INPUT and load into HDFS OUTPUT with handling 1 MB size of data**

**AWS CLI command used:**

**HADOOP\_CONF\_DIR=/etc/hadoop/conf /usr/bin/flink run -m yarn-cluster -yn 3 -yjm 1024 -ytm 4096 /usr/lib/flink/examples/streaming/flink\_tutorial\_test-1.0-SNAPSHOT.jar --input s3://emrsource.olp.han.com/compare/sample\_txt.txt --output hdfs:///flink-output/data\_514\_0**

**Output Result:**

****

**Total Elapsed Time: 13.4680 seconds.**

**2) Using AWS EMR CLI to ACCESS AWS S3 Bucket to get the Input file and apply program to find the aggregated values to store a output again into AWS S3 Bucket.**

**AWS CLI command I used:**

**aws --region cn-northwest-1 emr add-steps --cluster-id j-1BPSV6Q9AIYIS \**

**--steps Type=CUSTOM\_JAR,Name=Test\_5\_14,Jar=s3://emrsource.olp.han.com/compare/command-runner.jar,\**

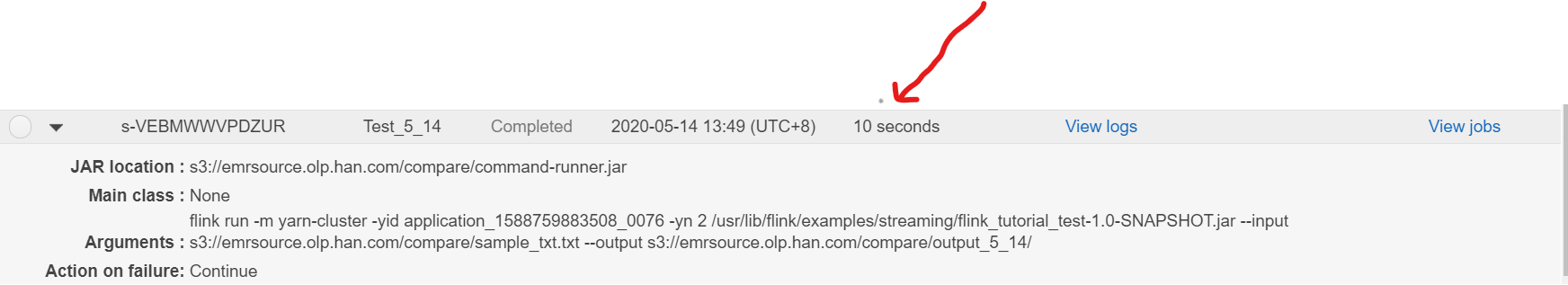
**Args="flink","run","-m","yarn-cluster","-yid","application\_1588759883508\_0076","-yn","2",\**

**"/usr/lib/flink/examples/streaming/flink\_tutorial\_test-1.0-SNAPSHOT.jar",\**

**"--input","s3://emrsource.olp.han.com/compare/sample\_txt.txt","--output","s3://emrsource.olp.han.com/compare/output\_5\_14/"**

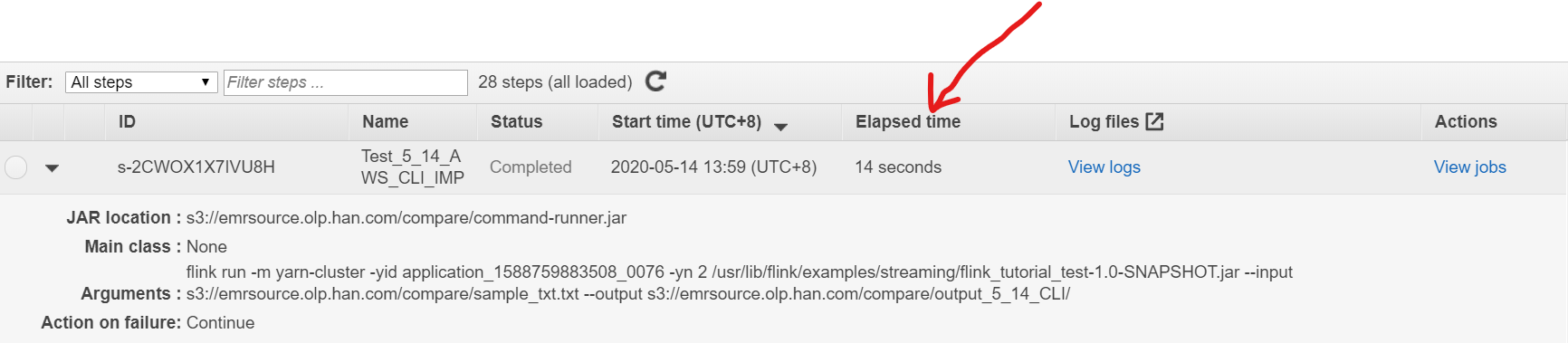
**Output Result for Serval Times Running:**

**Ran the Application first time:**

****

**Total Elapsed Time : 10 Seconds**

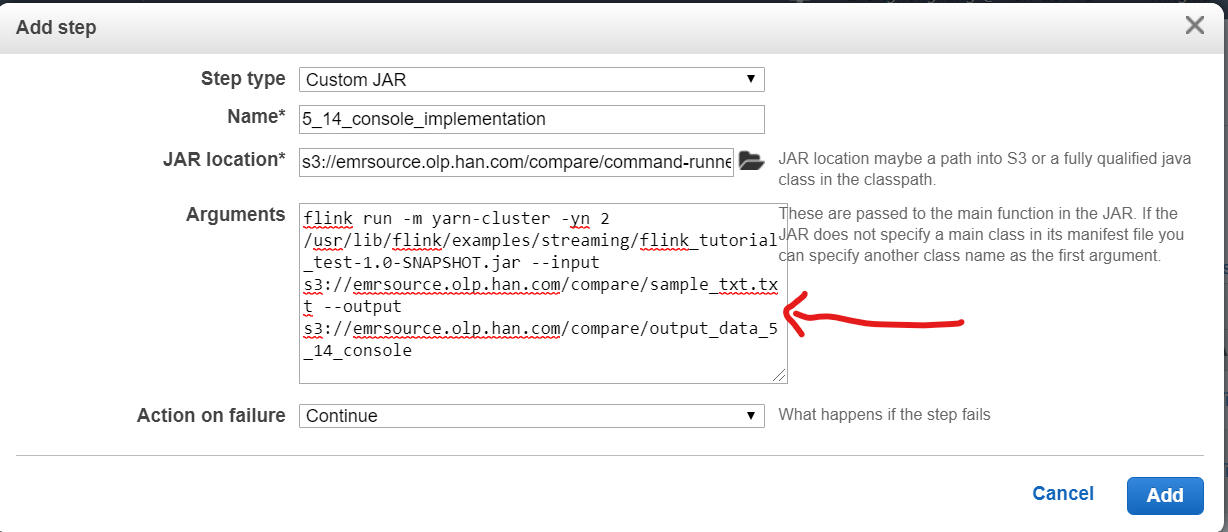
**Ran the Application Second time using AWS EMR CLI and check the elapsed time:**

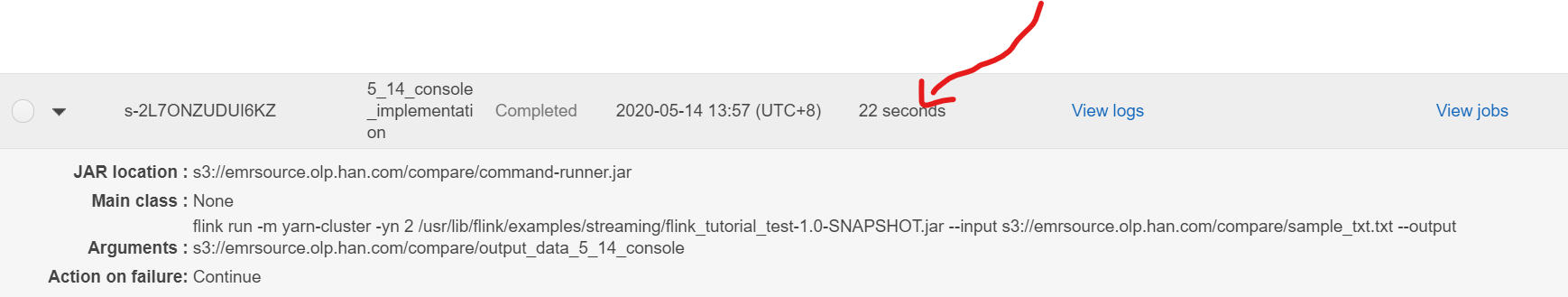
****

**Total Elapsed Time : 14 Seconds:**

**3) Using AWS EMR Console way to run a Application:**

**AWS S3 Bucket to get the Input file and apply program to find the aggregated values to store a output again into AWS S3 Bucket.**

****

****

**Total Elapsed Time : 22 seconds:**

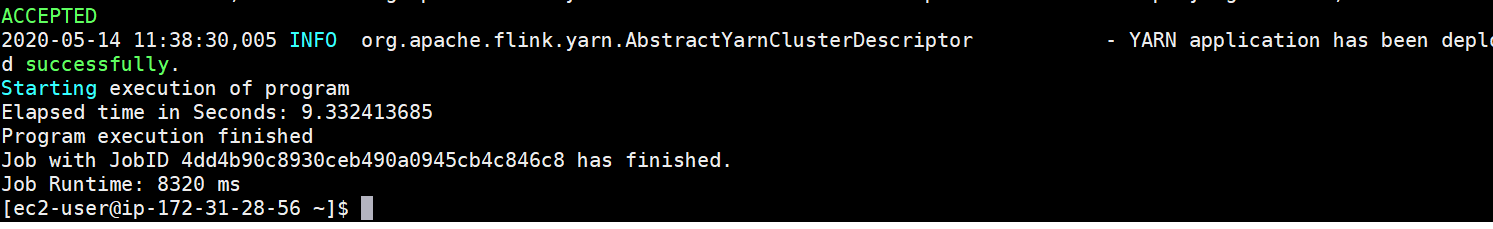
**Comparison Result Analysis using Hadoop,AWS EMR CLI, AWS EMR Console**

|  |  |  |  |
| --- | --- | --- | --- |
| **Implementation-Method** | **AWS Hadoop** | **AWS EMR CLI** | **AWS EMR CONSOLE** |
| ELAPSED TIME | 13.46 Seconds | 1st run – 10 sec / 2nd run -14 sec | 22 seconds |
|  |  |  |  |
|  |  |  |  |

**The Best Performance way to Run a application using AWS EMR CLI :**

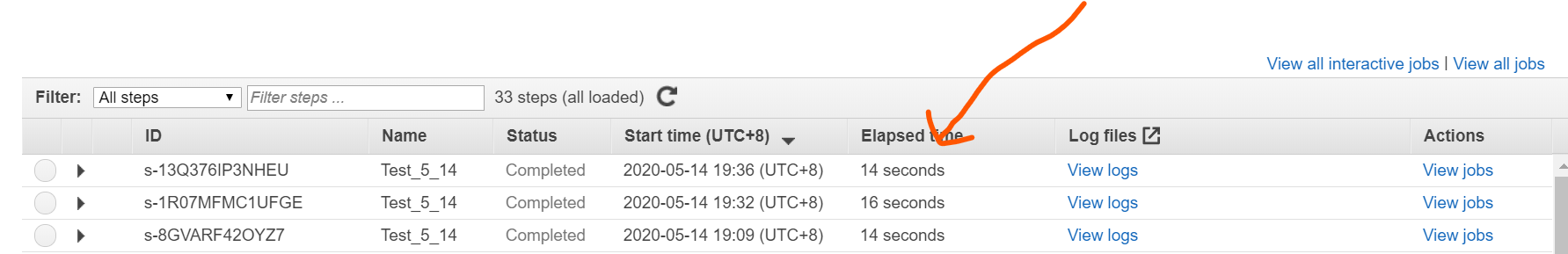
**Total Elapsed Time is : 10 sec 1st run**

**Total Elapsed Time is : 14 sec 2nd run**

****

**Aws Hadoop IO:**

**Elapsed Time is around : 9.33 seconds.**

****

**Aws EMR S3 IO:**

**Elapsed Time is around : 14 seconds.**