

Case Study: Financial Regulation

Domain: Banking

All financial institutions have been introduced to a new financial regulation that will consider the impact of possible future events when calculating their risk exposure.

For this change, all financial trading services will have to source data from far more data sources including both trade and risk data, which means large amounts of data that have not historically been required for accounting.

The data is of trading index listings for each trading day.

Tasks:

To adhere to the regulation a leading trading service provider has decided to create a new repository (Big Data) and a multi-facet platform that can cater to these models.

As part of the R&D team, you are required to execute a POC exercise to maintain a master data model.

1. Test the spark environment by executing the spark's HdfsTest.scala example.
2. Try to implement the same example in pyspark and perform spark-submit.
3. Analyze the behavior of spark application on Spark web UI
4. Edit the application and add custom logs. Once executed check the Spark logs.
5. Transfer the sample dataset from RDBMS to HDFS
6. Validate the loaded data by comparing the statistics of data both in source and HDFS
7. Create a new directory EQ in HDFS and transfer the data where series is EQ
8. Set total trades which are less than 500 to 0 and transfer only updated rows.

Location of HdfsTest.scala

`/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/spark/examples/jars/spark-examples*.jar use_cases/HdfsTest.scala`

Location of the Spark shell and the spark Submit Binary Files

```
[raghavendr48edu@ip-10-1-1-204 ~]$ cd /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/bin/
[raghavendr48edu@ip-10-1-1-204 bin]$ ls
avro-tools      impala-collect-diagnostics  kafka-preferred-replica-election  oozie          sqoop-import-all-tables
beeline         impala-collect-minidumps    kafka-producer-perf-test          oozie-setup    sqoop-job
bigtop-detect-javahome  impalad                    kafka-reassign-partitions         parquet-tools   sqoop-list-databases
catalogd        impala-shell               kafka-replica-verification       pig            sqoop-list-tables
cli_mt          kafka-broker-api-versions   kafka-run-class                   pyspark        sqoop-merge
cli_st          kafka-acls                  kafka-sentry                      sentry         sqoop-metastore
flume-ng        kafka-configs               kafka-server-start                solrctl        sqoop-version
hadoop          kafka-console-consumer      kafka-server-stop                 spark-shell    statestored
hadoop-fuse-dfs  kafka-console-producer      kafka-streams-application-reset   spark-submit   yarn
hbase           kafka-consumer-groups       kafka-topics                      sqoop          zookeeper-client
hbase-indexer   kafka-consumer-perf-test    kafka-verifiable-consumer        sqoop          zookeeper-security-migration
hbase-indexer-sentry  kafka-delegation-tokens     kafka-verifiable-producer        sqoop-codegen  zookeeper-server
hcat            kafka-delete-records        kite-dataset                      sqoop-create-hive-table  zookeeper-server
hdfs            kafka-dump-log              kudu                              sqoop-eval     zookeeper-server-cleanup
hive            kafka-log-dirs              load_gen                          sqoop-export   zookeeper-server-initialize
hiveserver2     kafka-mirror-maker          mapred                           sqoop-help     sqoop-import
[raghavendr48edu@ip-10-1-1-204 bin]$
```

```
[raghavendr48edu@ip-10-1-1-204 opt]$ ls
anaconda3      cloudera      conda      mahout-0.12.2  sbt          scripts
apache-mahout-distribution-0.12.2.tar.gz  cm_cdp_cdh_log4j_jndi_removal.sh  dbscript  mongobusercreation.sh  scala-2.11.12  zz
[raghavendr48edu@ip-10-1-1-204 opt]$
```

```
[raghavendr48edu@ip-10-1-1-204 cloudera]$ ls
cm  cm-agent  cm_cdp_cdh_log4j_jndi_removal.sh  csd  installer  log4shell-backup  parcel-cache  parcel-repo  parcels
[raghavendr48edu@ip-10-1-1-204 cloudera]$
```

```
[raghavendr48edu@ip-10-1-1-204 parcels]$ ls
CDH  CDH-6.2.1-1.cdh6.2.1.p0.1425774  CFM  CFM-1.0.0.0  FLINK  FLINK-1.9.0-csa1.0.0.0-cdh6.3.0
[raghavendr48edu@ip-10-1-1-204 parcels]$
```

```
[raghavendr48edu@ip-10-1-1-204 CDH-6.2.1-1.cdh6.2.1.p0.1425774]$ ls
bin  etc  include  jars  lib  lib64  libexec  meta  share
[raghavendr48edu@ip-10-1-1-204 CDH-6.2.1-1.cdh6.2.1.p0.1425774]$
```

```
[raghavendr48edu@ip-10-1-1-204 ~]$ cd /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/
[raghavendr48edu@ip-10-1-1-204 lib]$ ls
avro      hadoop      hadoop-kms      hbase      hive-hcatalog  impala-shell  kudu      pig      solr      zookeeper
bigtop-utils  hadoop-hdfs  hadoop-mapreduce  hbase-solr  hue           kafka         oozie     search  spark  zookeeper-native
flume-ng     hadoop-httpfs  hadoop-yarn      hive        impala        kite          parquet   sentry   sqoop
[raghavendr48edu@ip-10-1-1-204 lib]$
```

```
[raghavendr48edu@ip-10-1-1-204 opt]$ cd /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/spark
[raghavendr48edu@ip-10-1-1-204 spark]$ ls
bin  cloudera  conf  examples  hive  jars  kubernetes  LICENSE  NOTICE  python  python.tar.gz  RELEASE  sbin  work
[raghavendr48edu@ip-10-1-1-204 spark]$
```

```
[raghavendr48edu@ip-10-1-1-204 spark]$ cd /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/spark/examples
[raghavendr48edu@ip-10-1-1-204 examples]$ ls
jars
[raghavendr48edu@ip-10-1-1-204 examples]$
```

```
[raghavendr48edu@ip-10-1-1-204 examples]$ cd /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/spark/examples/jars
[raghavendr48edu@ip-10-1-1-204 jars]$ ls
scopt_2.11-3.7.0.jar  spark-examples_2.11-2.4.0-cdh6.2.1.jar
[raghavendr48edu@ip-10-1-1-204 jars]$
```

1. Test the spark environment by executing the spark's HdfsTest.scala

example.

HDFSTest.scala

```
package org.apache.spark.examples
import org.apache.spark.sql.SparkSession
object HdfsTest {
  /** Usage: HdfsTest [file] */
  def main(args: Array[String]) {
    if (args.length < 1) {
      System.err.println("Usage: HdfsTest <file>")
      System.exit(1)
    }
    val spark = SparkSession.builder.appName("HdfsTest") .getOrCreate()
    val file = spark.read.text(args(0)).rdd
    val mapped = file.map(s => s.length).cache()
    for (iter <- 1 to 10) {
      val start = System.currentTimeMillis()
      for (x <- mapped) { x + 2 }
      val end = System.currentTimeMillis()
      println("Iteration " + iter + " took " + (end-start) + " ms")
    }
    spark.stop()
  }
}
```

From

http://bigdata.cloudlabs.edureka.co:8888/hue/filebrowser/view=/user/raghavendr48edu/use_cases/HdfsTest.scala

SparkSubmit

```
[raghavendr48edu@ip-10-1-1-204 jars]$ spark-submit --class
org.apache.spark.examples.HdfsTest --master yarn --deploy-mode
cluster /opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/spark/examples/jars/spark-examples*.jar
use_cases/HdfsTest.scala
23/05/16 15:25:12 INFO client.RMProxy: Connecting to ResourceManager at ip-
10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/16 15:25:12 INFO yarn.Client: Requesting a new application from cluster
with 3 NodeManagers
23/05/16 15:25:12 INFO conf.Configuration: resource-types.xml not found
23/05/16 15:25:12 INFO resource.ResourceUtils: Unable to find 'resource-
types.xml'.
23/05/16 15:25:12 INFO yarn.Client: Verifying our application has not requested
more than the maximum memory capability of the cluster (20480 MB per
containe
r)
23/05/16 15:25:12 INFO yarn.Client: Will allocate AM container, with 1408 MB
memory including 384 MB overhead
```

23/05/16 15:25:12 INFO yarn.Client: Setting up container launch context for our AM
23/05/16 15:25:12 INFO yarn.Client: Setting up the launch environment for our AM container
23/05/16 15:25:12 INFO yarn.Client: Preparing resources for our AM container
23/05/16 15:25:12 INFO yarn.Client: Uploading resource file:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/spark/examples/jars/spark-examples_2.11-2.4.0-cdh6.2.1.jar ->
hdfs://nameservice1/user/raghavendr48edu/.sparkStaging/application_1682019329825_39245/spark-examples_2.11-2.4.0-cdh6.2.1.jar
23/05/16 15:25:12 INFO yarn.Client: Uploading resource file:/tmp/spark-db27f191-6b27-4149-b73c-5071e533f5d8/_spark_conf_5370573057872246418.zip -> hdfs://nameservice1/user/raghavendr48edu/.sparkStaging/application_1682019329825_39245/_spark_conf_.zip
23/05/16 15:25:12 INFO spark.SecurityManager: Changing view acls to: raghavendr48edu
23/05/16 15:25:12 INFO spark.SecurityManager: Changing modify acls to: raghavendr48edu
23/05/16 15:25:12 INFO spark.SecurityManager: Changing view acls groups to:
23/05/16 15:25:12 INFO spark.SecurityManager: Changing modify acls groups to:
23/05/16 15:25:12 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(raghavendr48edu);
groups with view permissions: Set(); users with modify permissions: Set(raghavendr48edu); groups with modify permissions: Set()
23/05/16 15:25:12 INFO conf.HiveConf: Found configuration file <file:/etc/hive/conf.cloudera.hive/hive-site.xml>
23/05/16 15:25:13 INFO security.YARNHadoopDelegationTokenManager: Attempting to load user's ticket cache.
23/05/16 15:25:13 INFO yarn.Client: Submitting application application_1682019329825_39245 to ResourceManager
23/05/16 15:25:13 INFO impl.YarnClientImpl: Submitted application application_1682019329825_39245
23/05/16 15:25:14 INFO yarn.Client: Application report for application_1682019329825_39245 (state: ACCEPTED)
23/05/16 15:25:14 INFO yarn.Client:
client token: N/A
diagnostics: AM container is launched, waiting for AM container to Register with RM
ApplicationMaster host: N/A
ApplicationMaster RPC port: -1
queue: root.default
start time: 1684250713073
final status: UNDEFINED
tracking URL: http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1682019329825_39245/
user: raghavendr48edu

23/05/16 15:25:15 INFO yarn.Client: Application report for application_1682019329825_39245 (state: ACCEPTED)
23/05/16 15:25:16 INFO yarn.Client: Application report for application_1682019329825_39245 (state: ACCEPTED)
23/05/16 15:25:17 INFO yarn.Client: Application report for application_1682019329825_39245 (state: ACCEPTED)
23/05/16 15:25:18 INFO yarn.Client: Application report for application_1682019329825_39245 (state: RUNNING)
23/05/16 15:25:18 INFO yarn.Client:
 client token: N/A
 diagnostics: N/A
 ApplicationMaster host: ip-10-1-2-103.ap-south-1.compute.internal
 ApplicationMaster RPC port: 43158
 queue: root.default
 start time: 1684250713073
 final status: UNDEFINED
 tracking URL: http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1682019329825_39245/

user: raghavendr48edu

23/05/16 15:25:19 INFO yarn.Client: Application report for application_1682019329825_39245 (state: RUNNING)
23/05/16 15:25:20 INFO yarn.Client: Application report for application_1682019329825_39245 (state: RUNNING)
23/05/16 15:25:21 INFO yarn.Client: Application report for application_1682019329825_39245 (state: RUNNING)
23/05/16 15:25:22 INFO yarn.Client: Application report for application_1682019329825_39245 (state: RUNNING)
23/05/16 15:25:23 INFO yarn.Client: Application report for application_1682019329825_39245 (state: RUNNING)
23/05/16 15:25:24 INFO yarn.Client: Application report for application_1682019329825_39245 (state: RUNNING)
23/05/16 15:25:25 INFO yarn.Client: Application report for application_1682019329825_39245 (state: RUNNING)
23/05/16 15:25:26 INFO yarn.Client: Application report for application_1682019329825_39245 (state: RUNNING)
23/05/16 15:25:27 INFO yarn.Client: Application report for application_1682019329825_39245 (state: RUNNING)
23/05/16 15:25:28 INFO yarn.Client: Application report for application_1682019329825_39245 (state: RUNNING)
23/05/16 15:25:29 INFO yarn.Client: Application report for application_1682019329825_39245 (state: RUNNING)
23/05/16 15:25:30 INFO yarn.Client: Application report for application_1682019329825_39245 (state: RUNNING)
23/05/16 15:25:31 INFO yarn.Client: Application report for application_1682019329825_39245 (state: RUNNING)
23/05/16 15:25:32 INFO yarn.Client: Application report for application_1682019329825_39245 (state: RUNNING)
23/05/16 15:25:33 INFO yarn.Client: Application report for application_1682019329825_39245 (state: RUNNING)
23/05/16 15:25:34 INFO yarn.Client: Application report for application_1682019329825_39245 (state: RUNNING)

23/05/16 15:25:35 INFO yarn.Client: Application report for application_1682019329825_39245 (state: FINISHED)
23/05/16 15:25:35 INFO yarn.Client:
client token: N/A
diagnostics: N/A
ApplicationMaster host: ip-10-1-2-103.ap-south-1.compute.internal
ApplicationMaster RPC port: 43158
queue: root.default
start time: 1684250713073
final status: SUCCEEDED
tracking URL: http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1682019329825_39245/

user: raghavendr48edu
23/05/16 15:25:35 INFO util.ShutdownHookManager: Shutdown hook called
23/05/16 15:25:35 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-db27f191-6b27-4149-b73c-5071e533f5d8
23/05/16 15:25:35 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-ef180879-1b83-4424-b017-3c4c1e0458fd

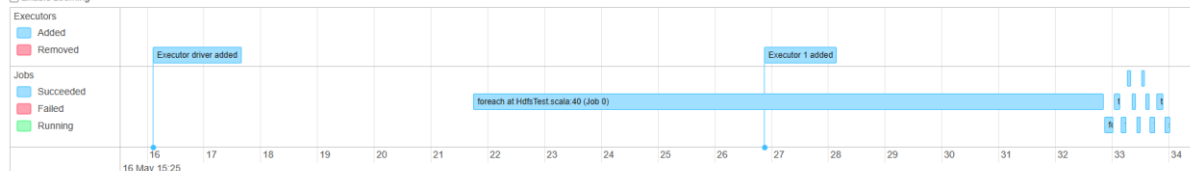
From <<https://bigdata.cloudlabs.edureka.co/shell/>>

3. Analyze the behavior of spark application on Spark web UI

Spark Jobs (?)

User: raghavendr48edu
Total Uptime: 18 s
Scheduling Mode: FIFO
Completed Jobs: 12

Event Timeline
☐ Enable zooming



Completed Jobs (12)

Job Id	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
11	sum at HdfsTest.scala:45 sum at HdfsTest.scala:45	2023/05/16 15:25:33	97 ms	1/1	1/1
10	take at HdfsTest.scala:44 take at HdfsTest.scala:44	2023/05/16 15:25:33	0.1 s	1/1	1/1
9	foreach at HdfsTest.scala:40 foreach at HdfsTest.scala:40	2023/05/16 15:25:33	96 ms	1/1	1/1
8	foreach at HdfsTest.scala:40 foreach at HdfsTest.scala:40	2023/05/16 15:25:33	65 ms	1/1	1/1
7	foreach at HdfsTest.scala:40 foreach at HdfsTest.scala:40	2023/05/16 15:25:33	71 ms	1/1	1/1
6	foreach at HdfsTest.scala:40 foreach at HdfsTest.scala:40	2023/05/16 15:25:33	70 ms	1/1	1/1
5	foreach at HdfsTest.scala:40 foreach at HdfsTest.scala:40	2023/05/16 15:25:33	77 ms	1/1	1/1
4	foreach at HdfsTest.scala:40 foreach at HdfsTest.scala:40	2023/05/16 15:25:33	76 ms	1/1	1/1
3	foreach at HdfsTest.scala:40 foreach at HdfsTest.scala:40	2023/05/16 15:25:33	97 ms	1/1	1/1
2	foreach at HdfsTest.scala:40 foreach at HdfsTest.scala:40	2023/05/16 15:25:33	0.1 s	1/1	1/1
1	foreach at HdfsTest.scala:40 foreach at HdfsTest.scala:40	2023/05/16 15:25:32	0.2 s	1/1	1/1
0	foreach at HdfsTest.scala:40 foreach at HdfsTest.scala:40	2023/05/16 15:25:21	11 s	1/1	1/1

```
scala> object HdfsTest {
|
|  /** Usage: HdfsTest [file] */
|  def main(args: Array[String]) {
|    if (args.length < 1) {
|      System.err.println("Usage: HdfsTest <file>")
|      System.exit(1)
|    }
|    val spark = SparkSession
|      .builder
|      .appName("HdfsTest")
|      .getOrCreate()
|    val file = spark.read.text(args(0)).rdd
|    val mapped = file.map(s => s.length).cache()
|    for (iter <- 1 to 10) {
|      val start = System.currentTimeMillis()
```

```

|     for (x <- mapped) { x + 2 }
|     val end = System.currentTimeMillis()
|     println("Iteration " + iter + " took " + (end-start) + " ms")
|   }
|   spark.stop()
| }
| }
defined object HdfsTest

```

```

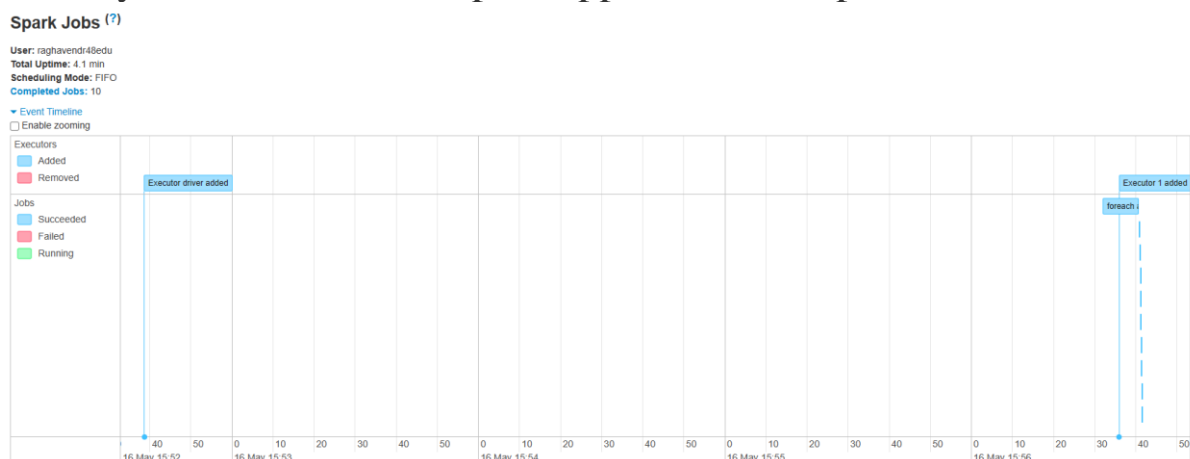
scala> HdfsTest.main(Array("AppleStore.csv"))
23/05/16 15:56:30 WARN sql.SparkSession$Builder: Using an existing
SparkSession; some configuration may not take effect.
Iteration 1 took 8800 ms
Iteration 2 took 160 ms
Iteration 3 took 76 ms
Iteration 4 took 82 ms
Iteration 5 took 62 ms
Iteration 6 took 59 ms
Iteration 7 took 58 ms
Iteration 8 took 84 ms
Iteration 9 took 73 ms
Iteration 10 took 54 ms

scala>

```

From <<https://bigdata.cloudlabs.edureka.co/shell/>>

3. Analyze the behavior of spark application on Spark web UI



Completed Jobs (10)

Job Id	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
9	foreach at <console>.44 foreach at <console>.44	2023/05/16 15:56:41	47 ms	1/1	1/1
8	foreach at <console>.44 foreach at <console>.44	2023/05/16 15:56:41	63 ms	1/1	1/1
7	foreach at <console>.44 foreach at <console>.44	2023/05/16 15:56:41	77 ms	1/1	1/1
6	foreach at <console>.44 foreach at <console>.44	2023/05/16 15:56:41	51 ms	1/1	1/1
5	foreach at <console>.44 foreach at <console>.44	2023/05/16 15:56:41	51 ms	1/1	1/1
4	foreach at <console>.44 foreach at <console>.44	2023/05/16 15:56:41	53 ms	1/1	1/1
3	foreach at <console>.44 foreach at <console>.44	2023/05/16 15:56:41	74 ms	1/1	1/1
2	foreach at <console>.44 foreach at <console>.44	2023/05/16 15:56:40	67 ms	1/1	1/1
1	foreach at <console>.44 foreach at <console>.44	2023/05/16 15:56:40	0.1 s	1/1	1/1
0	foreach at <console>.44 foreach at <console>.44	2023/05/16 15:56:32	9 s	1/1	1/1

OR

Step 1: Create a folder named training_project
mkdir training_project

Step 2: Create build.sbt in the training_project folder with the following content (using vi build.sbt)

```
name := "SparkMe Project"
version := "1.0"
organization := "edureka"
scalaVersion := "2.11.8"
val sparkVersion = "2.1.0"
libraryDependencies += "org.apache.spark" %% "spark-core" % sparkVersion % "provided"
libraryDependencies += "org.apache.spark" %% "spark-sql" % sparkVersion % "provided"
resolvers += Resolver.mavenLocal
[raghavendr48edu@ip-10-1-1-204 ~]$ cd training_project
[raghavendr48edu@ip-10-1-1-204 training_project]$ vi build.sbt
```

Step 3: Verify build.sbt by executing the command given below
sbt compile

The above command will create directory "project" and "target".

```
[raghavendr48edu@ip-10-1-1-204 training_project]$ sbt compile
[info] welcome to sbt 1.5.5 (Oracle Corporation Java 1.8.0_181)
[info] loading project definition from
/home/raghavendr48edu/training_project/project
[info] loading settings for project training_project from build.sbt ...
[info] set current project to SparkMe Project (in build
file:/home/raghavendr48edu/training_project/)
[info] Executing in batch mode. For better performance use sbt's shell
[info] Updating
https://repo1.maven.org/maven2/org/scala-lang/scala-library/2.11.8/scala-
library-2.11.8.pom
100.0% [#####] 1.4 KiB (2.0 KiB / s)
[info] Resolved dependencies
[info] Updating
```

https://repo1.maven.org/maven2/org/apache/spark/spark-sql_2.11/2.1.0/spark-sql_2.11-2.1.0.pom
100.0% [#####] 9.4 KiB (56.9 KiB / s)

https://repo1.maven.org/maven2/org/apache/spark/spark-core_2.11/2.1.0/spark-core_2.11-2.1.0.pom
100.0% [#####] 20.4 KiB (49.2 KiB / s)

https://repo1.maven.org/maven2/org/apache/spark/spark-parent_2.11/2.1.0/spark-parent_2.11-2.1.0.pom
100.0% [#####] 98.2 KiB (318.9 KiB / s)

<https://repo1.maven.org/maven2/org/apache/apache/14/apache-14.pom>
100.0% [#####] 14.3 KiB (103.2 KiB / s)

<https://repo1.maven.org/maven2/net/razorvine/pyrolite/4.13/pyrolite-4.13.pom>
100.0% [#####] 3.5 KiB (25.0 KiB / s)

<https://repo1.maven.org/maven2/log4j/log4j/1.2.17/log4j-1.2.17.pom>
100.0% [#####] 21.2 KiB (133.6 KiB / s)

https://repo1.maven.org/maven2/org/apache/spark/spark-tags_2.11/2.1.0/spark-tags_2.11-2.1.0.pom
100.0% [#####] 1.8 KiB (13.0 KiB / s)

https://repo1.maven.org/maven2/org/json4s/json4s-jackson_2.11/3.2.11/json4s-jackson_2.11-3.2.11.pom
100.0% [#####] 2.3 KiB (16.5 KiB / s)

https://repo1.maven.org/maven2/org/apache/spark/spark-sketch_2.11/2.1.0/spark-sketch_2.11-2.1.0.pom
100.0% [#####] 3.0 KiB (7.6 KiB / s)

https://repo1.maven.org/maven2/org/apache/spark/spark-unsafe_2.11/2.1.0/spark-unsafe_2.11-2.1.0.pom
100.0% [#####] 4.5 KiB (11.3 KiB / s)

<https://repo1.maven.org/maven2/org/glassfish/jersey/core/jersey-client/2.22.2/jersey-client-2.22.2.pom>
100.0% [#####] 7.4 KiB (18.3 KiB / s)

<https://repo1.maven.org/maven2/com/twitter/chill-java/0.8.0/chill-java-0.8.0.pom>
100.0% [#####] 2.0 KiB (13.9 KiB / s)

<https://repo1.maven.org/maven2/org/spark-project/spark/unused/1.0.0/unused-1.0.0.pom>
100.0% [#####] 2.3 KiB (15.4 KiB / s)

<https://repo1.maven.org/maven2/com/google/code/findbugs/jsr305/1.3.9/jsr305-1.3.9.pom>
100.0% [#####] 965B (7.1 KiB / s)

<https://repo1.maven.org/maven2/org/apache/ivy/ivy/2.4.0/ivy-2.4.0.pom>
100.0% [#####] 6.1 KiB (41.4 KiB / s)

<https://repo1.maven.org/maven2/org/apache/parquet/parquet-hadoop/1.8.1/parquet-hadoop-1.8.1.pom>
100.0% [#####] 3.4 KiB (24.4 KiB / s)

https://repo1.maven.org/maven2/org/apache/spark/spark-network-shuffle_2.11/2.1.0/spark-network-shuffle_2.11-2.1.0.pom
100.0% [#####] 4.1 KiB (26.6 KiB / s)

<https://repo1.maven.org/maven2/com/univocity/univocity-parsers/2.2.1/univocity-parsers-2.2.1.pom>
100.0% [#####] 8.0 KiB (14.3 KiB / s)

<https://repo1.maven.org/maven2/io/dropwizard/metrics/metrics-graphite/3.1.2/metrics-graphite-3.1.2.pom>

100.0% [#####] 1.3 KiB (9.0 KiB / s)
<https://repo1.maven.org/maven2/org/apache/curator/curator-recipes/2.4.0/curator-recipes-2.4.0.pom>

100.0% [#####] 2.4 KiB (17.4 KiB / s)
<https://repo1.maven.org/maven2/com/clearspring/analytics/stream/2.7.0/stream-2.7.0.pom>

100.0% [#####] 6.5 KiB (44.2 KiB / s)
https://repo1.maven.org/maven2/com/twitter/chill_2.11/0.8.0/chill_2.11-0.8.0.pom

100.0% [#####] 2.3 KiB (15.4 KiB / s)
<https://repo1.maven.org/maven2/org/slf4j/jcl-over-slf4j/1.7.16/jcl-over-slf4j-1.7.16.pom>

100.0% [#####] 963B (7.1 KiB / s)
<https://repo1.maven.org/maven2/org/slf4j/slf4j-log4j12/1.7.16/slf4j-log4j12-1.7.16.pom>

100.0% [#####] 909B (6.2 KiB / s)
<https://repo1.maven.org/maven2/oro/oro/2.0.8/oro-2.0.8.pom>

100.0% [#####] 140B (1.0 KiB / s)
<https://repo1.maven.org/maven2/io/dropwizard/metrics/metrics-jvm/3.1.2/metrics-jvm-3.1.2.pom>

100.0% [#####] 967B (7.0 KiB / s)
<https://repo1.maven.org/maven2/org/apache/xbean/xbean-asm5-shaded/4.4/xbean-asm5-shaded-4.4.pom>

100.0% [#####] 3.3 KiB (24.8 KiB / s)
https://repo1.maven.org/maven2/com/fasterxml/jackson/module/jackson-module-scala_2.11/2.6.5/jackson-module-scala_2.11-2.6.5.pom

100.0% [#####] 4.0 KiB (27.3 KiB / s)
<https://repo1.maven.org/maven2/commons-net/commons-net/2.2/commons-net-2.2.pom>

100.0% [#####] 12.4 KiB (77.4 KiB / s)
<https://repo1.maven.org/maven2/javax/servlet/javax.servlet-api/3.1.0/javax.servlet-api-3.1.0.pom>

100.0% [#####] 13.2 KiB (81.0 KiB / s)
<https://repo1.maven.org/maven2/org/apache/commons/commons-crypto/1.0.0/commons-crypto-1.0.0.pom>

100.0% [#####] 21.3 KiB (130.8 KiB / s)
<https://repo1.maven.org/maven2/org/glassfish/jersey/containers/jersey-container-servlet/2.22.2/jersey-container-servlet-2.22.2.pom>

100.0% [#####] 4.5 KiB (30.1 KiB / s)
<https://repo1.maven.org/maven2/com/ning/compress-lzf/1.0.3/compress-lzf-1.0.3.pom>

100.0% [#####] 7.7 KiB (47.9 KiB / s)
<https://repo1.maven.org/maven2/org/glassfish/jersey/containers/jersey-container-servlet-core/2.22.2/jersey-container-servlet-core-2.22.2.pom>

100.0% [#####] 4.6 KiB (28.1 KiB / s)
<https://repo1.maven.org/maven2/net/java/dev/jets3t/jets3t/0.7.1/jets3t-0.7.1.pom>

100.0% [#####] 1.5 KiB (9.7 KiB / s)
<https://repo1.maven.org/maven2/org/apache/commons/commons-math3/3.4.1/commons-math3-3.4.1.pom>

100.0% [#####] 26.6 KiB (145.1 KiB / s)

<https://repo1.maven.org/maven2/org/apache/commons/commons-lang3/3.5/commons-lang3-3.5.pom>
100.0% [#####] 22.8 KiB (122.6 KiB / s)
<https://repo1.maven.org/maven2/com/fasterxml/jackson/core/jackson-databind/2.6.5/jackson-databind-2.6.5.pom>
100.0% [#####] 5.9 KiB (43.2 KiB / s)
<https://repo1.maven.org/maven2/io/dropwizard/metrics/metrics-json/3.1.2/metrics-json-3.1.2.pom>
100.0% [#####] 1.4 KiB (9.7 KiB / s)
<https://repo1.maven.org/maven2/io/dropwizard/metrics/metrics-core/3.1.2/metrics-core-3.1.2.pom>
100.0% [#####] 846B (5.5 KiB / s)
<https://repo1.maven.org/maven2/org/glassfish/jersey/core/jersey-common/2.22.2/jersey-common-2.22.2.pom>
100.0% [#####] 10.2 KiB (61.9 KiB / s)
<https://repo1.maven.org/maven2/io/netty/netty/3.8.0.Final/netty-3.8.0.Final.pom>
100.0% [#####] 25.6 KiB (151.4 KiB / s)
<https://repo1.maven.org/maven2/org/xerial/snappy/snappy-java/1.1.2.6/snappy-java-1.1.2.6.pom>
100.0% [#####] 3.2 KiB (21.0 KiB / s)
<https://repo1.maven.org/maven2/org/apache/parquet/parquet-column/1.8.1/parquet-column-1.8.1.pom>
100.0% [#####] 4.9 KiB (32.9 KiB / s)
<https://repo1.maven.org/maven2/net/sf/py4j/py4j/0.10.4/py4j-0.10.4.pom>
100.0% [#####] 2.1 KiB (15.0 KiB / s)
<https://repo1.maven.org/maven2/org/apache/avro/avro-mapred/1.7.7/avro-mapred-1.7.7.pom>
100.0% [#####] 6.6 KiB (43.9 KiB / s)
https://repo1.maven.org/maven2/org/apache/spark/spark-network-common_2.11/2.1.0/spark-network-common_2.11-2.1.0.pom
100.0% [#####] 4.9 KiB (37.3 KiB / s)
<https://repo1.maven.org/maven2/org/slf4j/slf4j-api/1.7.16/slf4j-api-1.7.16.pom>
100.0% [#####] 2.7 KiB (19.8 KiB / s)
<https://repo1.maven.org/maven2/io/netty/netty-all/4.0.42.Final/netty-all-4.0.42.Final.pom>
100.0% [#####] 17.2 KiB (113.8 KiB / s)
https://repo1.maven.org/maven2/org/apache/spark/spark-catalyst_2.11/2.1.0/spark-catalyst_2.11-2.1.0.pom
100.0% [#####] 7.6 KiB (54.9 KiB / s)
https://repo1.maven.org/maven2/org/apache/spark/spark-launcher_2.11/2.1.0/spark-launcher_2.11-2.1.0.pom
100.0% [#####] 6.2 KiB (45.9 KiB / s)
<https://repo1.maven.org/maven2/net/jpountz/lz4/lz4/1.3.0/lz4-1.3.0.pom>
100.0% [#####] 1.9 KiB (12.5 KiB / s)
<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-client/2.2.0/hadoop-client-2.2.0.pom>
100.0% [#####] 10.8 KiB (76.1 KiB / s)
<https://repo1.maven.org/maven2/org/roaringbitmap/RoaringBitmap/0.5.11/RoaringBitmap-0.5.11.pom>
100.0% [#####] 6.2 KiB (42.9 KiB / s)

<https://repo1.maven.org/maven2/org/slf4j/jul-to-slf4j/1.7.16/jul-to-slf4j-1.7.16.pom>

100.0% [#####] 986B (6.6 KiB / s)

<https://repo1.maven.org/maven2/org/glassfish/jersey/core/jersey-server/2.22.2/jersey-server-2.22.2.pom>

100.0% [#####] 12.0 KiB (81.4 KiB / s)

<https://repo1.maven.org/maven2/org/apache/apache/7/apache-7.pom>

100.0% [#####] 14.1 KiB (95.2 KiB / s)

<https://repo1.maven.org/maven2/org/glassfish/jersey/containers/project/2.22.2/project-2.22.2.pom>

100.0% [#####] 3.7 KiB (25.4 KiB / s)

<https://repo1.maven.org/maven2/io/dropwizard/metrics/metrics-parent/3.1.2/metrics-parent-3.1.2.pom>

100.0% [#####] 11.7 KiB (72.3 KiB / s)

<https://repo1.maven.org/maven2/org/apache/curator/apache-curator/2.4.0/apache-curator-2.4.0.pom>

100.0% [#####] 20.1 KiB (119.4 KiB / s)

<https://repo1.maven.org/maven2/io/netty/netty-parent/4.0.42.Final/netty-parent-4.0.42.Final.pom>

100.0% [#####] 43.8 KiB (213.4 KiB / s)

<https://repo1.maven.org/maven2/com/fasterxml/jackson/jackson-parent/2.6.2/jackson-parent-2.6.2.pom>

100.0% [#####] 7.6 KiB (51.1 KiB / s)

<https://repo1.maven.org/maven2/org/apache/avro/avro-parent/1.7.7/avro-parent-1.7.7.pom>

100.0% [#####] 18.6 KiB (114.7 KiB / s)

<https://repo1.maven.org/maven2/org/apache/parquet/parquet/1.8.1/parquet-1.8.1.pom>

100.0% [#####] 17.0 KiB (102.5 KiB / s)

<https://repo1.maven.org/maven2/org/apache/commons/commons-parent/41/commons-parent-41.pom>

100.0% [#####] 63.3 KiB (342.3 KiB / s)

<https://repo1.maven.org/maven2/org/slf4j/slf4j-parent/1.7.16/slf4j-parent-1.7.16.pom>

100.0% [#####] 13.2 KiB (32.6 KiB / s)

<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-project-dist/2.2.0/hadoop-project-dist-2.2.0.pom>

100.0% [#####] 17.3 KiB (112.2 KiB / s)

<https://repo1.maven.org/maven2/net/java/jvnet-parent/3/jvnet-parent-3.pom>

100.0% [#####] 4.7 KiB (30.8 KiB / s)

<https://repo1.maven.org/maven2/org/apache/xbean/xbean/4.4/xbean-4.4.pom>

100.0% [#####] 17.3 KiB (115.2 KiB / s)

<https://repo1.maven.org/maven2/org/glassfish/jersey/project/2.22.2/project-2.22.2.pom>

100.0% [#####] 82.3 KiB (261.3 KiB / s)

<https://repo1.maven.org/maven2/org/apache/commons/commons-parent/40/commons-parent-40.pom>

100.0% [#####] 62.6 KiB (393.8 KiB / s)

<https://repo1.maven.org/maven2/org/sonatype/oss/oss-parent/5/oss-parent-5.pom>

100.0% [#####] 4.0 KiB (30.0 KiB / s)

<https://repo1.maven.org/maven2/org/apache/commons/commons-parent/34/commons-parent-34.pom>
100.0% [#####] 54.6 KiB (199.4 KiB / s)

<https://repo1.maven.org/maven2/org/apache/commons/commons-parent/17/commons-parent-17.pom>
100.0% [#####] 30.5 KiB (180.2 KiB / s)

<https://repo1.maven.org/maven2/com/fasterxml/oss-parent/24/oss-parent-24.pom>
100.0% [#####] 19.0 KiB (129.1 KiB / s)

<https://repo1.maven.org/maven2/net/java/jvnet-parent/4/jvnet-parent-4.pom>
100.0% [#####] 7.6 KiB (52.2 KiB / s)

<https://repo1.maven.org/maven2/org/apache/apache/18/apache-18.pom>
100.0% [#####] 15.3 KiB (105.5 KiB / s)

<https://repo1.maven.org/maven2/org/apache/geronimo/genesis/genesis-java5-flava/2.1/genesis-java5-flava-2.1.pom>
100.0% [#####] 5.4 KiB (36.9 KiB / s)

<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-project/2.2.0/hadoop-project-2.2.0.pom>
100.0% [#####] 36.3 KiB (219.9 KiB / s)

<https://repo1.maven.org/maven2/org/apache/avro/avro-toplevel/1.7.7/avro-toplevel-1.7.7.pom>
100.0% [#####] 9.4 KiB (23.7 KiB / s)

<https://repo1.maven.org/maven2/org/apache/apache/17/apache-17.pom>
100.0% [#####] 15.7 KiB (111.2 KiB / s)

<https://repo1.maven.org/maven2/org/apache/apache/16/apache-16.pom>
100.0% [#####] 15.0 KiB (105.1 KiB / s)

<https://repo1.maven.org/maven2/org/apache/apache/13/apache-13.pom>
100.0% [#####] 13.6 KiB (94.8 KiB / s)

<https://repo1.maven.org/maven2/org/apache/apache/10/apache-10.pom>
100.0% [#####] 14.5 KiB (98.3 KiB / s)

<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-main/2.2.0/hadoop-main-2.2.0.pom>
100.0% [#####] 16.2 KiB (113.2 KiB / s)

<https://repo1.maven.org/maven2/org/apache/geronimo/genesis/genesis-default-flava/2.1/genesis-default-flava-2.1.pom>
100.0% [#####] 15.7 KiB (96.2 KiB / s)

<https://repo1.maven.org/maven2/org/apache/geronimo/genesis/genesis/2.1/genesis-2.1.pom>
100.0% [#####] 13.9 KiB (84.4 KiB / s)

<https://repo1.maven.org/maven2/org/glassfish/hk2/hk2-bom/2.4.0-b34/hk2-bom-2.4.0-b34.pom>
100.0% [#####] 20.4 KiB (121.5 KiB / s)

<https://repo1.maven.org/maven2/net/java/jvnet-parent/5/jvnet-parent-5.pom>
100.0% [#####] 8.7 KiB (59.3 KiB / s)

<https://repo1.maven.org/maven2/com/fasterxml/jackson/core/jackson-annotations/2.6.5/jackson-annotations-2.6.5.pom>
100.0% [#####] 1.2 KiB (9.2 KiB / s)

<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-mapreduce-client-jobclient/2.2.0/hadoop-mapreduce-client-jobclient-2.2.0.pom>
100.0% [#####] 5.3 KiB (40.2 KiB / s)

https://repo1.maven.org/maven2/org/json4s/json4s-core_2.11/3.2.11/json4s-core_2.11-3.2.11.pom

100.0% [#####] 2.5 KiB (18.7 KiB / s)
<https://repo1.maven.org/maven2/org/glassfish/hk2/hk2-locator/2.4.0-b34/hk2-locator-2.4.0-b34.pom>

100.0% [#####] 5.4 KiB (35.5 KiB / s)
<https://repo1.maven.org/maven2/org/glassfish/jersey/media/jersey-media-jaxb/2.22.2/jersey-media-jaxb-2.22.2.pom>

100.0% [#####] 7.0 KiB (45.6 KiB / s)
<https://repo1.maven.org/maven2/com/fasterxml/jackson/module/jackson-module-paranamer/2.6.5/jackson-module-paranamer-2.6.5.pom>

100.0% [#####] 4.4 KiB (31.4 KiB / s)
<https://repo1.maven.org/maven2/org/antlr/antlr4-runtime/4.5.3/antlr4-runtime-4.5.3.pom>

100.0% [#####] 2.3 KiB (16.0 KiB / s)
<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-annotations/2.2.0/hadoop-annotations-2.2.0.pom>

100.0% [#####] 2.3 KiB (16.2 KiB / s)
<https://repo1.maven.org/maven2/org/apache/zookeeper/zookeeper/3.4.5/zookeeper-3.4.5.pom>

100.0% [#####] 2.9 KiB (20.4 KiB / s)
<https://repo1.maven.org/maven2/javax/ws/rs/javax.ws.rs-api/2.0.1/javax.ws.rs-api-2.0.1.pom>

100.0% [#####] 20.9 KiB (120.2 KiB / s)
<https://repo1.maven.org/maven2/org/apache/curator/curator-framework/2.4.0/curator-framework-2.4.0.pom>

100.0% [#####] 2.2 KiB (5.5 KiB / s)
<https://repo1.maven.org/maven2/org/glassfish/hk2/hk2-api/2.4.0-b34/hk2-api-2.4.0-b34.pom>

100.0% [#####] 4.5 KiB (33.0 KiB / s)
https://repo1.maven.org/maven2/org/scalatest/scalatest_2.11/2.2.6/scalatest_2.11-2.2.6.pom

100.0% [#####] 5.5 KiB (40.7 KiB / s)
<https://repo1.maven.org/maven2/com/fasterxml/jackson/core/jackson-core/2.6.5/jackson-core-2.6.5.pom>

100.0% [#####] 4.8 KiB (31.9 KiB / s)
<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-mapreduce-client-app/2.2.0/hadoop-mapreduce-client-app-2.2.0.pom>

100.0% [#####] 4.2 KiB (28.6 KiB / s)
<https://repo1.maven.org/maven2/org/apache/parquet/parquet-format/2.3.0-incubating/parquet-format-2.3.0-incubating.pom>

100.0% [#####] 6.8 KiB (49.5 KiB / s)
<https://repo1.maven.org/maven2/commons-httpclient/commons-httpclient/3.1/commons-httpclient-3.1.pom>

100.0% [#####] 7.6 KiB (51.6 KiB / s)
<https://repo1.maven.org/maven2/org/apache/avro/avro-ipc/1.7.7/avro-ipc-1.7.7.pom>

100.0% [#####] 5.0 KiB (36.5 KiB / s)
<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-common/2.2.0/hadoop-common-2.2.0.pom>

100.0% [#####] 26.2 KiB (181.0 KiB / s)
<https://repo1.maven.org/maven2/org/codehaus/jackson/jackson-mapper-asl/1.9.13/jackson-mapper-asl-1.9.13.pom>

100.0% [#####] 1.5 KiB (10.4 KiB / s)

<https://repo1.maven.org/maven2/org/glassfish/hk2/external/javax.inject/2.4.0-b34/javax.inject-2.4.0-b34.pom>

100.0% [#####] 6.3 KiB (46.0 KiB / s)

<https://repo1.maven.org/maven2/org/apache/parquet/parquet-common/1.8.1/parquet-common-1.8.1.pom>

100.0% [#####] 2.8 KiB (19.6 KiB / s)

<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-mapreduce-client-core/2.2.0/hadoop-mapreduce-client-core-2.2.0.pom>

100.0% [#####] 3.0 KiB (22.0 KiB / s)

<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-yarn-api/2.2.0/hadoop-yarn-api-2.2.0.pom>

100.0% [#####] 2.9 KiB (22.0 KiB / s)

<https://repo1.maven.org/maven2/com/google/guava/guava/14.0.1/guava-14.0.1.pom>

100.0% [#####] 5.3 KiB (36.5 KiB / s)

<https://repo1.maven.org/maven2/org/glassfish/hk2/osgi-resource-locator/1.0.1/osgi-resource-locator-1.0.1.pom>

100.0% [#####] 6.0 KiB (41.4 KiB / s)

<https://repo1.maven.org/maven2/org/glassfish/jersey/bundles/repackaged/jersey-guava/2.22.2/jersey-guava-2.22.2.pom>

100.0% [#####] 13.0 KiB (87.3 KiB / s)

<https://repo1.maven.org/maven2/org/codehaus/jackson/jackson-core-asl/1.9.13/jackson-core-asl-1.9.13.pom>

100.0% [#####] 1.2 KiB (9.3 KiB / s)

<https://repo1.maven.org/maven2/javax/annotation/javax.annotation-api/1.2/javax.annotation-api-1.2.pom>

100.0% [#####] 13.1 KiB (95.7 KiB / s)

<https://repo1.maven.org/maven2/javax/validation/validation-api/1.1.0.Final/validation-api-1.1.0.Final.pom>

100.0% [#####] 7.7 KiB (53.7 KiB / s)

<https://repo1.maven.org/maven2/org/codehaus/janino/commons-compiler/3.0.0/commons-compiler-3.0.0.pom>

100.0% [#####] 666B (4.4 KiB / s)

<https://repo1.maven.org/maven2/org/codehaus/janino/janino/3.0.0/janino-3.0.0.pom>

100.0% [#####] 925B (6.5 KiB / s)

<https://repo1.maven.org/maven2/org/scala-lang/scala-reflect/2.11.8/scala-reflect-2.11.8.pom>

100.0% [#####] 1.5 KiB (10.5 KiB / s)

<https://repo1.maven.org/maven2/commons-codec/commons-codec/1.10/commons-codec-1.10.pom>

100.0% [#####] 11.3 KiB (55.0 KiB / s)

<https://repo1.maven.org/maven2/com/esotericsoftware/kryo-shaded/3.0.3/kryo-shaded-3.0.3.pom>

100.0% [#####] 3.1 KiB (22.3 KiB / s)

<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-hdfs/2.2.0/hadoop-hdfs-2.2.0.pom>

100.0% [#####] 23.5 KiB (162.9 KiB / s)

<https://repo1.maven.org/maven2/org/apache/parquet/parquet-jackson/1.8.1/parquet-jackson-1.8.1.pom>

100.0% [#####] 3.5 KiB (24.0 KiB / s)

<https://repo1.maven.org/maven2/org/fusesource/leveldbjni/leveldbjni-all/1.8/leveldbjni-all-1.8.pom>

100.0% [#####] 7.1 KiB (49.7 KiB / s)

<https://repo1.maven.org/maven2/org/apache/parquet/parquet-encoding/1.8.1/parquet-encoding-1.8.1.pom>

100.0% [#####] 3.0 KiB (20.9 KiB / s)

<https://repo1.maven.org/maven2/org/antlr/antlr4-master/4.5.3/antlr4-master-4.5.3.pom>

100.0% [#####] 3.6 KiB (25.5 KiB / s)

<https://repo1.maven.org/maven2/org/codehaus/janino/janino-parent/3.0.0/janino-parent-3.0.0.pom>

100.0% [#####] 8.2 KiB (58.2 KiB / s)

<https://repo1.maven.org/maven2/org/glassfish/jersey/bundles/repackaged/project/2.22.2/project-2.22.2.pom>

100.0% [#####] 2.7 KiB (19.3 KiB / s)

<https://repo1.maven.org/maven2/org/glassfish/jersey/media/project/2.22.2/project-2.22.2.pom>

100.0% [#####] 2.9 KiB (20.8 KiB / s)

<https://repo1.maven.org/maven2/org/apache/commons/commons-parent/35/commons-parent-35.pom>

100.0% [#####] 56.4 KiB (311.7 KiB / s)

<https://repo1.maven.org/maven2/com/esotericsoftware/kryo-parent/3.0.3/kryo-parent-3.0.3.pom>

100.0% [#####] 4.9 KiB (37.6 KiB / s)

<https://repo1.maven.org/maven2/com/fasterxml/jackson/jackson-parent/2.6.1/jackson-parent-2.6.1.pom>

100.0% [#####] 7.6 KiB (50.4 KiB / s)

<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-yarn/2.2.0/hadoop-yarn-2.2.0.pom>

100.0% [#####] 6.4 KiB (42.4 KiB / s)

<https://repo1.maven.org/maven2/org/fusesource/leveldbjni/leveldbjni-project/1.8/leveldbjni-project-1.8.pom>

100.0% [#####] 9.4 KiB (62.9 KiB / s)

<https://repo1.maven.org/maven2/org/glassfish/hk2/hk2-parent/2.4.0-b34/hk2-parent-2.4.0-b34.pom>

100.0% [#####] 46.7 KiB (270.0 KiB / s)

<https://repo1.maven.org/maven2/com/google/guava/guava-parent/14.0.1/guava-parent-14.0.1.pom>

100.0% [#####] 2.5 KiB (17.6 KiB / s)

<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-mapreduce-client/2.2.0/hadoop-mapreduce-client-2.2.0.pom>

100.0% [#####] 6.2 KiB (15.5 KiB / s)

<https://repo1.maven.org/maven2/org/glassfish/hk2/external/2.4.0-b34/external-2.4.0-b34.pom>

100.0% [#####] 2.9 KiB (21.3 KiB / s)

<https://repo1.maven.org/maven2/org/glassfish/pom/8/pom-8.pom>

100.0% [#####] 10.3 KiB (65.2 KiB / s)

<https://repo1.maven.org/maven2/org/glassfish/jersey/bundles/project/2.22.2/project-2.22.2.pom>

100.0% [#####] 3.0 KiB (22.9 KiB / s)

<https://repo1.maven.org/maven2/org/apache/apache/15/apache-15.pom>

100.0% [#####] 14.9 KiB (97.2 KiB / s)

<https://repo1.maven.org/maven2/org/codehaus/codehaus-parent/4/codehaus-parent-4.pom>

100.0% [#####] 4.7 KiB (31.0 KiB / s)

<https://repo1.maven.org/maven2/org/fusesource/fusesource-pom/1.9/fusesource-pom-1.9.pom>

100.0% [#####] 14.4 KiB (95.0 KiB / s)

<https://repo1.maven.org/maven2/com/fasterxml/oss-parent/23/oss-parent-23.pom>

100.0% [#####] 18.8 KiB (108.9 KiB / s)

<https://repo1.maven.org/maven2/org/apache/curator/curator-client/2.4.0/curator-client-2.4.0.pom>

100.0% [#####] 2.2 KiB (15.5 KiB / s)

<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-mapreduce-client-common/2.2.0/hadoop-mapreduce-client-common-2.2.0.pom>

100.0% [#####] 3.2 KiB (22.0 KiB / s)

<https://repo1.maven.org/maven2/org/javassist/javassist/3.18.1-GA/javassist-3.18.1-GA.pom>

100.0% [#####] 9.3 KiB (64.1 KiB / s)

https://repo1.maven.org/maven2/org/scala-lang/modules/scala-xml_2.11/1.0.2/scala-xml_2.11-1.0.2.pom

100.0% [#####] 2.5 KiB (17.0 KiB / s)

<https://repo1.maven.org/maven2/org/xmlenc/xmlenc/0.52/xmlenc-0.52.pom>

100.0% [#####] 623B (4.2 KiB / s)

<https://repo1.maven.org/maven2/com/google/protobuf/protobuf-java/2.5.0/protobuf-java-2.5.0.pom>

100.0% [#####] 8.2 KiB (60.6 KiB / s)

<https://repo1.maven.org/maven2/org/commons-lang/commons-lang/2.5/commons-lang-2.5.pom>

100.0% [#####] 16.7 KiB (123.0 KiB / s)

<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-yarn-common/2.2.0/hadoop-yarn-common-2.2.0.pom>

100.0% [#####] 4.9 KiB (36.7 KiB / s)

<https://repo1.maven.org/maven2/org/mortbay/jetty/jetty-util/6.1.26/jetty-util-6.1.26.pom>

100.0% [#####] 3.8 KiB (26.5 KiB / s)

<https://repo1.maven.org/maven2/org/commons-net/commons-net/3.1/commons-net-3.1.pom>

100.0% [#####] 18.0 KiB (104.0 KiB / s)

<https://repo1.maven.org/maven2/com/thoughtworks/paranamer/paranamer/2.6/paranamer-2.6.pom>

100.0% [#####] 4.7 KiB (32.2 KiB / s)

<https://repo1.maven.org/maven2/org/commons-cli/commons-cli/1.2/commons-cli-1.2.pom>

100.0% [#####] 7.8 KiB (54.1 KiB / s)

https://repo1.maven.org/maven2/org/json4s/json4s-ast_2.11/3.2.11/json4s-ast_2.11-3.2.11.pom

100.0% [#####] 2.0 KiB (4.9 KiB / s)

<https://repo1.maven.org/maven2/org/objenesis/objenesis/2.1/objenesis-2.1.pom>

100.0% [#####] 2.5 KiB (17.4 KiB / s)

<https://repo1.maven.org/maven2/org/commons-configuration/commons-configuration/1.6/commons-configuration-1.6.pom>

100.0% [#####] 13.0 KiB (87.9 KiB / s)
<https://repo1.maven.org/maven2/commons-io/commons-io/2.1/commons-io-2.1.pom>

100.0% [#####] 10.4 KiB (66.6 KiB / s)
<https://repo1.maven.org/maven2/org/apache/commons/commons-compress/1.4.1/commons-compress-1.4.1.pom>

100.0% [#####] 10.5 KiB (79.0 KiB / s)
<https://repo1.maven.org/maven2/org/glassfish/hk2/external/aopalliance-repackaged/2.4.0-b34/aopalliance-repackaged-2.4.0-b34.pom>

100.0% [#####] 6.6 KiB (49.3 KiB / s)
<https://repo1.maven.org/maven2/org/scala-lang/scalap/2.11.8/scalap-2.11.8.pom>

100.0% [#####] 1.4 KiB (10.5 KiB / s)
<https://repo1.maven.org/maven2/com/esotericsoftware/minlog/1.3.0/minlog-1.3.0.pom>

100.0% [#####] 3.2 KiB (22.1 KiB / s)
<https://repo1.maven.org/maven2/org/apache/avro/avro/1.7.7/avro-1.7.7.pom>

100.0% [#####] 5.5 KiB (38.7 KiB / s)
<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-mapreduce-client-shuffle/2.2.0/hadoop-mapreduce-client-shuffle-2.2.0.pom>

100.0% [#####] 1.6 KiB (11.9 KiB / s)
<https://repo1.maven.org/maven2/org/apache/commons/commons-math/2.1/commons-math-2.1.pom>

100.0% [#####] 10.1 KiB (69.5 KiB / s)
<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-auth/2.2.0/hadoop-auth-2.2.0.pom>

100.0% [#####] 6.9 KiB (48.0 KiB / s)
<https://repo1.maven.org/maven2/org/glassfish/hk2/hk2-utils/2.4.0-b34/hk2-utils-2.4.0-b34.pom>

100.0% [#####] 5.0 KiB (28.1 KiB / s)
<https://repo1.maven.org/maven2/com/google/google/1/google-1.pom>

100.0% [#####] 1.5 KiB (11.5 KiB / s)
<https://repo1.maven.org/maven2/com/thoughtworks/paranamer/paranamer-parent/2.6/paranamer-parent-2.6.pom>

100.0% [#####] 11.0 KiB (72.4 KiB / s)
<https://repo1.maven.org/maven2/org/mortbay/jetty/project/6.1.26/project-6.1.26.pom>

100.0% [#####] 9.0 KiB (59.1 KiB / s)
<https://repo1.maven.org/maven2/org/objenesis/objenesis-parent/2.1/objenesis-parent-2.1.pom>

100.0% [#####] 16.3 KiB (107.3 KiB / s)
<https://repo1.maven.org/maven2/org/apache/commons/commons-parent/24/commons-parent-24.pom>

100.0% [#####] 46.2 KiB (169.7 KiB / s)
<https://repo1.maven.org/maven2/org/apache/commons/commons-parent/23/commons-parent-23.pom>

100.0% [#####] 43.4 KiB (256.6 KiB / s)
<https://repo1.maven.org/maven2/org/apache/commons/commons-parent/22/commons-parent-22.pom>

100.0% [#####] 40.9 KiB (248.1 KiB / s)
<https://repo1.maven.org/maven2/org/apache/commons/commons-parent/14/commons-parent-14.pom>

100.0% [#####] 30.7 KiB (189.2 KiB / s)
<https://repo1.maven.org/maven2/org/apache/commons/commons-parent/12/commons-parent-12.pom>

100.0% [#####] 26.3 KiB (175.3 KiB / s)
<https://repo1.maven.org/maven2/org/apache/commons/commons-parent/11/commons-parent-11.pom>

100.0% [#####] 24.1 KiB (156.3 KiB / s)
<https://repo1.maven.org/maven2/org/apache/apache/9/apache-9.pom>

100.0% [#####] 14.8 KiB (105.7 KiB / s)
<https://repo1.maven.org/maven2/org/mortbay/jetty/jetty-parent/10/jetty-parent-10.pom>

100.0% [#####] 3.2 KiB (22.9 KiB / s)
<https://repo1.maven.org/maven2/org/apache/apache/4/apache-4.pom>

100.0% [#####] 4.4 KiB (31.8 KiB / s)
<https://repo1.maven.org/maven2/org/eclipse/jetty/jetty-parent/14/jetty-parent-14.pom>

100.0% [#####] 16.1 KiB (111.8 KiB / s)
<https://repo1.maven.org/maven2/javax/inject/javax.inject/1/javax.inject-1.pom>

100.0% [#####] 612B (4.6 KiB / s)
<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-yarn-server-common/2.2.0/hadoop-yarn-server-common-2.2.0.pom>

100.0% [#####] 3.5 KiB (26.3 KiB / s)
<https://repo1.maven.org/maven2/org/tukaani/xz/1.0/xz-1.0.pom>

100.0% [#####] 1.9 KiB (14.2 KiB / s)
<https://repo1.maven.org/maven2/commons-beanutils/commons-beanutils-core/1.8.0/commons-beanutils-core-1.8.0.pom>

100.0% [#####] 1.6 KiB (10.6 KiB / s)
<https://repo1.maven.org/maven2/commons-digester/commons-digester/1.8/commons-digester-1.8.pom>

100.0% [#####] 6.8 KiB (45.3 KiB / s)
<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-yarn-client/2.2.0/hadoop-yarn-client-2.2.0.pom>

100.0% [#####] 1.9 KiB (13.7 KiB / s)
<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-yarn-server-nodemanager/2.2.0/hadoop-yarn-server-nodemanager-2.2.0.pom>

100.0% [#####] 7.2 KiB (51.1 KiB / s)
<https://repo1.maven.org/maven2/org/scala-lang/scala-compiler/2.11.8/scala-compiler-2.11.8.pom>

100.0% [#####] 2.3 KiB (16.3 KiB / s)
<https://repo1.maven.org/maven2/commons-collections/commons-collections/3.2.1/commons-collections-3.2.1.pom>

100.0% [#####] 12.2 KiB (31.2 KiB / s)
<https://repo1.maven.org/maven2/org/apache/commons/commons-parent/9/commons-parent-9.pom>

100.0% [#####] 21.4 KiB (137.3 KiB / s)
<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-yarn-server/2.2.0/hadoop-yarn-server-2.2.0.pom>

100.0% [#####] 1.8 KiB (11.4 KiB / s)
<https://repo1.maven.org/maven2/com/google/inject/guice/3.0/guice-3.0.pom>

100.0% [#####] 7.1 KiB (48.0 KiB / s)
<https://repo1.maven.org/maven2/commons-beanutils/commons-beanutils/1.7.0/commons-beanutils-1.7.0.pom>

100.0% [#####] 357B (2.4 KiB / s)
https://repo1.maven.org/maven2/org/scala-lang/modules/scala-parser-combinators_2.11/1.0.4/scala-parser-combinators_2.11-1.0.4.pom

100.0% [#####] 2.1 KiB (14.2 KiB / s)
https://repo1.maven.org/maven2/org/scala-lang/modules/scala-xml_2.11/1.0.4/scala-xml_2.11-1.0.4.pom

100.0% [#####] 2.5 KiB (16.6 KiB / s)
<https://repo1.maven.org/maven2/com/google/inject/guice-parent/3.0/guice-parent-3.0.pom>

100.0% [#####] 12.7 KiB (82.8 KiB / s)
<https://repo1.maven.org/maven2/com/google/google/5/google-5.pom>

100.0% [#####] 2.4 KiB (16.7 KiB / s)
<https://repo1.maven.org/maven2/aopalliance/aopalliance/1.0/aopalliance-1.0.pom>

100.0% [#####] 363B (2.5 KiB / s)
<https://repo1.maven.org/maven2/jline/jline/0.9.94/jline-0.9.94.pom>

100.0% [#####] 6.2 KiB (43.1 KiB / s)
[info] Resolved dependencies
[info] Updating
[info] Resolved dependencies
[info] Fetching artifacts of
<https://repo1.maven.org/maven2/org/glassfish/hk2/external/aopalliance-repackaged/2.4.0-b34/aopalliance-repackaged-2.4.0-b34.jar>

100.0% [#####] 14.4 KiB (98.8 KiB / s)
<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-auth/2.2.0/hadoop-auth-2.2.0.jar>

100.0% [#####] 48.6 KiB (292.7 KiB / s)
https://repo1.maven.org/maven2/org/apache/spark/spark-sketch_2.11/2.1.0/spark-sketch_2.11-2.1.0.jar

100.0% [#####] 28.8 KiB (173.8 KiB / s)
<https://repo1.maven.org/maven2/org/apache/parquet/parquet-hadoop/1.8.1/parquet-hadoop-1.8.1.jar>

100.0% [#####] 213.0 KiB (696.0 KiB / s)
<https://repo1.maven.org/maven2/javax/servlet/javax.servlet-api/3.1.0/javax.servlet-api-3.1.0.jar>

100.0% [#####] 93.6 KiB (525.6 KiB / s)
<https://repo1.maven.org/maven2/com/ning/compress-lzf/1.0.3/compress-lzf-1.0.3.jar>

100.0% [#####] 78.0 KiB (428.4 KiB / s)
<https://repo1.maven.org/maven2/commons-digester/commons-digester/1.8/commons-digester-1.8.jar>

100.0% [#####] 140.2 KiB (345.4 KiB / s)
<https://repo1.maven.org/maven2/io/dropwizard/metrics/metrics-graphite/3.1.2/metrics-graphite-3.1.2.jar>

100.0% [#####] 20.4 KiB (137.6 KiB / s)
https://repo1.maven.org/maven2/org/apache/spark/spark-tags_2.11/2.1.0/spark-tags_2.11-2.1.0.jar

100.0% [#####] 16.7 KiB (119.3 KiB / s)
<https://repo1.maven.org/maven2/org/glassfish/jersey/containers/jersey-container-servlet/2.22.2/jersey-container-servlet-2.22.2.jar>

100.0% [#####] 17.7 KiB (130.9 KiB / s)

https://repo1.maven.org/maven2/com/twitter/chill_2.11/0.8.0/chill_2.11-0.8.0.jar
100.0% [#####] 218.3 KiB (766.1 KiB / s)

<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-mapreduce-client-core/2.2.0/hadoop-mapreduce-client-core-2.2.0.jar>
100.0% [#####] 1.4 MiB (2.4 MiB / s)

<https://repo1.maven.org/maven2/commons-net/commons-net/3.1/commons-net-3.1.jar>
100.0% [#####] 267.0 KiB (339.6 KiB / s)

<https://repo1.maven.org/maven2/io/dropwizard/metrics/metrics-core/3.1.2/metrics-core-3.1.2.jar>
100.0% [#####] 109.9 KiB (654.3 KiB / s)

<https://repo1.maven.org/maven2/org/glassfish/hk2/hk2-locator/2.4.0-b34/hk2-locator-2.4.0-b34.jar>
100.0% [#####] 177.0 KiB (867.8 KiB / s)

<https://repo1.maven.org/maven2/org/spark-project/spark/unused/1.0.0/unused-1.0.0.jar>
100.0% [#####] 2.7 KiB (18.6 KiB / s)

<https://repo1.maven.org/maven2/org/slf4j/slf4j-api/1.7.16/slf4j-api-1.7.16.jar>
100.0% [#####] 39.6 KiB (295.2 KiB / s)

<https://repo1.maven.org/maven2/org/glassfish/hk2/osgi-resource-locator/1.0.1/osgi-resource-locator-1.0.1.jar>
100.0% [#####] 19.8 KiB (130.9 KiB / s)

<https://repo1.maven.org/maven2/javax/annotation/javax.annotation-api/1.2/javax.annotation-api-1.2.jar>
100.0% [#####] 25.7 KiB (190.7 KiB / s)

<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-hdfs/2.2.0/hadoop-hdfs-2.2.0.jar>
100.0% [#####] 5.0 MiB (7.0 MiB / s)

<https://repo1.maven.org/maven2/io/dropwizard/metrics/metrics-json/3.1.2/metrics-json-3.1.2.jar>
100.0% [#####] 15.5 KiB (110.4 KiB / s)

<https://repo1.maven.org/maven2/commons-collections/commons-collections/3.2.1/commons-collections-3.2.1.jar>
100.0% [#####] 561.9 KiB (1.8 MiB / s)

https://repo1.maven.org/maven2/org/apache/spark/spark-unsafe_2.11/2.1.0/spark-unsafe_2.11-2.1.0.jar
100.0% [#####] 43.1 KiB (297.6 KiB / s)

<https://repo1.maven.org/maven2/org/objenesis/objenesis/2.1/objenesis-2.1.jar>
100.0% [#####] 40.8 KiB (293.4 KiB / s)

<https://repo1.maven.org/maven2/xmlenc/xmlenc/0.52/xmlenc-0.52.jar>
100.0% [#####] 14.7 KiB (99.7 KiB / s)

<https://repo1.maven.org/maven2/io/netty/netty/3.8.0.Final/netty-3.8.0.Final.jar>
100.0% [#####] 1.2 MiB (5.6 MiB / s)

<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-yarn-client/2.2.0/hadoop-yarn-client-2.2.0.jar>
100.0% [#####] 92.5 KiB (690.4 KiB / s)

<https://repo1.maven.org/maven2/com/twitter/chill-java/0.8.0/chill-java-0.8.0.jar>
100.0% [#####] 49.4 KiB (355.6 KiB / s)

<https://repo1.maven.org/maven2/org/apache/ivy/ivy/2.4.0/ivy-2.4.0.jar>
100.0% [#####] 1.2 MiB (3.1 MiB / s)

<https://repo1.maven.org/maven2/org/glassfish/hk2/hk2-utils/2.4.0-b34/hk2-utils-2.4.0-b34.jar>

100.0% [#####] 116.2 KiB (769.4 KiB / s)

<https://repo1.maven.org/maven2/org/apache/commons/commons-crypto/1.0.0/commons-crypto-1.0.0.jar>

100.0% [#####] 131.4 KiB (932.2 KiB / s)

https://repo1.maven.org/maven2/org/json4s/json4s-jackson_2.11/3.2.11/json4s-jackson_2.11-3.2.11.jar

100.0% [#####] 39.4 KiB (264.4 KiB / s)

<https://repo1.maven.org/maven2/com/thoughtworks/paranamer/paranamer/2.6/paranamer-2.6.jar>

100.0% [#####] 32.0 KiB (240.9 KiB / s)

<https://repo1.maven.org/maven2/org/mortbay/jetty/jetty-util/6.1.26/jetty-util-6.1.26.jar>

100.0% [#####] 173.0 KiB (1.0 MiB / s)

https://repo1.maven.org/maven2/org/apache/spark/spark-catalyst_2.11/2.1.0/spark-catalyst_2.11-2.1.0.jar

100.0% [#####] 7.0 MiB (7.3 MiB / s)

<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-yarn-server-nodemanager/2.2.0/hadoop-yarn-server-nodemanager-2.2.0.jar>

100.0% [#####] 456.7 KiB (2.7 MiB / s)

<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-common/2.2.0/hadoop-common-2.2.0.jar>

100.0% [#####] 2.6 MiB (6.4 MiB / s)

<https://repo1.maven.org/maven2/commons-cli/commons-cli/1.2/commons-cli-1.2.jar>

100.0% [#####] 40.2 KiB (282.8 KiB / s)

<https://repo1.maven.org/maven2/org/apache/curator/curator-framework/2.4.0/curator-framework-2.4.0.jar>

100.0% [#####] 175.5 KiB (1.2 MiB / s)

<https://repo1.maven.org/maven2/org/glassfish/hk2/hk2-api/2.4.0-b34/hk2-api-2.4.0-b34.jar>

100.0% [#####] 174.8 KiB (1.2 MiB / s)

<https://repo1.maven.org/maven2/org/xerial/snappy/snappy-java/1.1.2.6/snappy-java-1.1.2.6.jar>

100.0% [#####] 1.0 MiB (5.7 MiB / s)

<https://repo1.maven.org/maven2/com/esotericsoftware/kryo-shaded/3.0.3/kryo-shaded-3.0.3.jar>

100.0% [#####] 350.0 KiB (2.3 MiB / s)

<https://repo1.maven.org/maven2/javax/inject/javax.inject/1/javax.inject-1.jar>

100.0% [#####] 2.4 KiB (18.3 KiB / s)

<https://repo1.maven.org/maven2/org/oro/oro/2.0.8/oro-2.0.8.jar>

100.0% [#####] 63.7 KiB (486.5 KiB / s)

<https://repo1.maven.org/maven2/com/esotericsoftware/minlog/1.3.0/minlog-1.3.0.jar>

100.0% [#####] 5.6 KiB (37.4 KiB / s)

<https://repo1.maven.org/maven2/org/scala-lang/scalap/2.11.8/scalap-2.11.8.jar>

100.0% [#####] 784.0 KiB (4.6 MiB / s)

<https://repo1.maven.org/maven2/commons-httpclient/commons-httpclient/3.1/commons-httpclient-3.1.jar>

100.0% [#####] 297.9 KiB (2.0 MiB / s)

<https://repo1.maven.org/maven2/com/univocity/univocity-parsers/2.2.1/univocity-parsers-2.2.1.jar>
100.0% [#####] 317.9 KiB (2.2 MiB / s)

<https://repo1.maven.org/maven2/org/apache/parquet/parquet-column/1.8.1/parquet-column-1.8.1.jar>
100.0% [#####] 923.7 KiB (5.1 MiB / s)

https://repo1.maven.org/maven2/org/scala-lang/modules/scala-parser-combinators_2.11/1.0.4/scala-parser-combinators_2.11-1.0.4.jar
100.0% [#####] 413.8 KiB (2.6 MiB / s)

<https://repo1.maven.org/maven2/org/apache/commons/commons-math/2.1/commons-math-2.1.jar>
100.0% [#####] 812.9 KiB (2.4 MiB / s)

https://repo1.maven.org/maven2/org/apache/spark/spark-network-common_2.11/2.1.0/spark-network-common_2.11-2.1.0.jar
100.0% [#####] 2.2 MiB (7.9 MiB / s)

https://repo1.maven.org/maven2/org/json4s/json4s-ast_2.11/3.2.11/json4s-ast_2.11-3.2.11.jar
100.0% [#####] 80.5 KiB (490.8 KiB / s)

<https://repo1.maven.org/maven2/com/fasterxml/jackson/module/jackson-module-paranamer/2.6.5/jackson-module-paranamer-2.6.5.jar>
100.0% [#####] 40.3 KiB (281.8 KiB / s)

<https://repo1.maven.org/maven2/com/fasterxml/jackson/core/jackson-annotations/2.6.5/jackson-annotations-2.6.5.jar>
100.0% [#####] 45.9 KiB (332.5 KiB / s)

<https://repo1.maven.org/maven2/org/apache/curator/curator-recipes/2.4.0/curator-recipes-2.4.0.jar>
100.0% [#####] 235.7 KiB (1.4 MiB / s)

<https://repo1.maven.org/maven2/org/codehaus/jackson/jackson-mapper-asl/1.9.13/jackson-mapper-asl-1.9.13.jar>
100.0% [#####] 762.4 KiB (4.8 MiB / s)

<https://repo1.maven.org/maven2/io/dropwizard/metrics/metrics-jvm/3.1.2/metrics-jvm-3.1.2.jar>
100.0% [#####] 38.4 KiB (290.6 KiB / s)

<https://repo1.maven.org/maven2/org/apache/curator/curator-client/2.4.0/curator-client-2.4.0.jar>
100.0% [#####] 65.7 KiB (490.1 KiB / s)

<https://repo1.maven.org/maven2/org/apache/avro/avro-ipc/1.7.7/avro-ipc-1.7.7.jar>
100.0% [#####] 188.5 KiB (1.2 MiB / s)

<https://repo1.maven.org/maven2/com/fasterxml/jackson/core/jackson-databind/2.6.5/jackson-databind-2.6.5.jar>
100.0% [#####] 1.1 MiB (5.9 MiB / s)

<https://repo1.maven.org/maven2/org/fusesource/leveldbjni/leveldbjni-all/1.8/leveldbjni-all-1.8.jar>
100.0% [#####] 1021.2 KiB (5.3 MiB / s)

<https://repo1.maven.org/maven2/aopalliance/aopalliance/1.0/aopalliance-1.0.jar>
100.0% [#####] 4.4 KiB (30.3 KiB / s)

<https://repo1.maven.org/maven2/com/google/code/findbugs/jsr305/1.3.9/jsr305-1.3.9.jar>
100.0% [#####] 32.2 KiB (228.7 KiB / s)

<https://repo1.maven.org/maven2/org/apache/commons/commons-math3/3.4.1/commons-math3-3.4.1.jar>

100.0% [#####] 1.9 MiB (6.8 MiB / s)

<https://repo1.maven.org/maven2/org/apache/parquet/parquet-format/2.3.0-incubating/parquet-format-2.3.0-incubating.jar>

100.0% [#####] 378.1 KiB (2.6 MiB / s)

<https://repo1.maven.org/maven2/org/glassfish/hk2/external/javax.inject/2.4.0-b34/javax.inject-2.4.0-b34.jar>

100.0% [#####] 5.8 KiB (43.4 KiB / s)

<https://repo1.maven.org/maven2/net/jpountz/lz4/lz4/1.3.0/lz4-1.3.0.jar>

100.0% [#####] 231.3 KiB (1.6 MiB / s)

<https://repo1.maven.org/maven2/org/codehaus/janino/janino/3.0.0/janino-3.0.0.jar>

100.0% [#####] 718.3 KiB (4.2 MiB / s)

<https://repo1.maven.org/maven2/commons-lang/commons-lang/2.5/commons-lang-2.5.jar>

100.0% [#####] 272.6 KiB (1.8 MiB / s)

<https://repo1.maven.org/maven2/net/java/dev/jets3t/jets3t/0.7.1/jets3t-0.7.1.jar>

100.0% [#####] 368.9 KiB (2.4 MiB / s)

<https://repo1.maven.org/maven2/commons-beanutils/commons-beanutils/1.7.0/commons-beanutils-1.7.0.jar>

100.0% [#####] 184.2 KiB (1.3 MiB / s)

<https://repo1.maven.org/maven2/org/apache/xbean/xbean-asm5-shaded/4.4/xbean-asm5-shaded-4.4.jar>

100.0% [#####] 141.3 KiB (967.6 KiB / s)

<https://repo1.maven.org/maven2/org/apache/parquet/parquet-encoding/1.8.1/parquet-encoding-1.8.1.jar>

100.0% [#####] 278.8 KiB (2.0 MiB / s)

<https://repo1.maven.org/maven2/org/glassfish/jersey/containers/jersey-container-servlet-core/2.22.2/jersey-container-servlet-core-2.22.2.jar>

100.0% [#####] 64.7 KiB (483.0 KiB / s)

<https://repo1.maven.org/maven2/org/roaringbitmap/RoaringBitmap/0.5.11/RoaringBitmap-0.5.11.jar>

100.0% [#####] 197.2 KiB (1011.3 KiB / s)

<https://repo1.maven.org/maven2/jline/jline/0.9.94/jline-0.9.94.jar>

100.0% [#####] 85.3 KiB (631.7 KiB / s)

<https://repo1.maven.org/maven2/org/javassist/javassist/3.18.1-GA/javassist-3.18.1-GA.jar>

100.0% [#####] 697.5 KiB (4.1 MiB / s)

<https://repo1.maven.org/maven2/io/netty/netty-all/4.0.42.Final/netty-all-4.0.42.Final.jar>

100.0% [#####] 2.2 MiB (7.8 MiB / s)

https://repo1.maven.org/maven2/org/apache/spark/spark-network-shuffle_2.11/2.1.0/spark-network-shuffle_2.11-2.1.0.jar

100.0% [#####] 60.1 KiB (420.4 KiB / s)

<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-mapreduce-client-jobclient/2.2.0/hadoop-mapreduce-client-jobclient-2.2.0.jar>

100.0% [#####] 34.4 KiB (233.9 KiB / s)

<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-yarn-api/2.2.0/hadoop-yarn-api-2.2.0.jar>

100.0% [#####] 1.1 MiB (5.4 MiB / s)

<https://repo1.maven.org/maven2/com/google/protobuf/protobuf-java/2.5.0/protobuf-java-2.5.0.jar>

100.0% [#####] 521.0 KiB (3.5 MiB / s)

<https://repo1.maven.org/maven2/org/slf4j/slf4j-log4j12/1.7.16/slf4j-log4j12-1.7.16.jar>

100.0% [#####] 9.7 KiB (70.8 KiB / s)

<https://repo1.maven.org/maven2/commons-io/commons-io/2.1/commons-io-2.1.jar>

100.0% [#####] 159.3 KiB (1.2 MiB / s)

<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-mapreduce-client-app/2.2.0/hadoop-mapreduce-client-app-2.2.0.jar>

100.0% [#####] 470.7 KiB (3.0 MiB / s)

<https://repo1.maven.org/maven2/org/tukaani/xz/1.0/xz-1.0.jar>

100.0% [#####] 92.5 KiB (679.8 KiB / s)

<https://repo1.maven.org/maven2/commons-configuration/commons-configuration/1.6/commons-configuration-1.6.jar>

100.0% [#####] 291.8 KiB (2.0 MiB / s)

<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-yarn-server-common/2.2.0/hadoop-yarn-server-common-2.2.0.jar>

100.0% [#####] 171.4 KiB (1.2 MiB / s)

<https://repo1.maven.org/maven2/org/apache/parquet/parquet-common/1.8.1/parquet-common-1.8.1.jar>

100.0% [#####] 37.7 KiB (256.5 KiB / s)

<https://repo1.maven.org/maven2/org/slf4j/jul-to-slf4j/1.7.16/jul-to-slf4j-1.7.16.jar>

100.0% [#####] 4.5 KiB (31.6 KiB / s)

<https://repo1.maven.org/maven2/org/glassfish/jersey/core/jersey-common/2.22.2/jersey-common-2.22.2.jar>

100.0% [#####] 682.0 KiB (4.1 MiB / s)

<https://repo1.maven.org/maven2/org/glassfish/jersey/media/jersey-media-jaxb/2.22.2/jersey-media-jaxb-2.22.2.jar>

100.0% [#####] 71.0 KiB (473.5 KiB / s)

<https://repo1.maven.org/maven2/org/apache/commons/commons-compress/1.4.1/commons-compress-1.4.1.jar>

100.0% [#####] 235.7 KiB (1.6 MiB / s)

<https://repo1.maven.org/maven2/org/glassfish/jersey/core/jersey-server/2.22.2/jersey-server-2.22.2.jar>

100.0% [#####] 929.4 KiB (5.2 MiB / s)

<https://repo1.maven.org/maven2/org/glassfish/jersey/core/jersey-client/2.22.2/jersey-client-2.22.2.jar>

100.0% [#####] 163.5 KiB (1.1 MiB / s)

<https://repo1.maven.org/maven2/net/razorvine/pyrolite/4.13/pyrolite-4.13.jar>

100.0% [#####] 92.6 KiB (666.0 KiB / s)

<https://repo1.maven.org/maven2/org/scala-lang/scala-reflect/2.11.8/scala-reflect-2.11.8.jar>

100.0% [#####] 4.4 MiB (10.0 MiB / s)

<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-mapreduce-client-common/2.2.0/hadoop-mapreduce-client-common-2.2.0.jar>

100.0% [#####] 641.0 KiB (4.1 MiB / s)

<https://repo1.maven.org/maven2/commons-codec/commons-codec/1.10/commons-codec-1.10.jar>

100.0% [#####] 277.5 KiB (1.9 MiB / s)

<https://repo1.maven.org/maven2/org/scala-lang/scala-library/2.11.8/scala-library-2.11.8.jar>

100.0% [#####] 5.5 MiB (9.9 MiB / s)

https://repo1.maven.org/maven2/org/apache/spark/spark-sql_2.11/2.1.0/spark-sql_2.11-2.1.0.jar

100.0% [#####] 6.4 MiB (9.8 MiB / s)

<https://repo1.maven.org/maven2/com/google/inject/guice/3.0/guice-3.0.jar>

100.0% [#####] 693.8 KiB (4.0 MiB / s)

<https://repo1.maven.org/maven2/javax/validation/validation-api/1.1.0.Final/validation-api-1.1.0.Final.jar>

100.0% [#####] 62.3 KiB (404.4 KiB / s)

<https://repo1.maven.org/maven2/org/apache/commons/commons-lang3/3.5/commons-lang3-3.5.jar>

100.0% [#####] 468.6 KiB (2.8 MiB / s)

<https://repo1.maven.org/maven2/org/codehaus/janino/commons-compiler/3.0.0/commons-compiler-3.0.0.jar>

100.0% [#####] 35.8 KiB (271.0 KiB / s)

https://repo1.maven.org/maven2/com/fasterxml/jackson/module/jackson-module-scala_2.11/2.6.5/jackson-module-scala_2.11-2.6.5.jar

100.0% [#####] 503.5 KiB (3.1 MiB / s)

https://repo1.maven.org/maven2/org/scala-lang/modules/scala-xml_2.11/1.0.4/scala-xml_2.11-1.0.4.jar

100.0% [#####] 632.7 KiB (3.9 MiB / s)

<https://repo1.maven.org/maven2/org/antlr/antlr4-runtime/4.5.3/antlr4-runtime-4.5.3.jar>

100.0% [#####] 295.2 KiB (1.7 MiB / s)

<https://repo1.maven.org/maven2/org/apache/avro/avro-mapred/1.7.7/avro-mapred-1.7.7-hadoop2.jar>

100.0% [#####] 176.5 KiB (1.2 MiB / s)

<https://repo1.maven.org/maven2/com/clearspring/analytics/stream/2.7.0/stream-2.7.0.jar>

100.0% [#####] 170.3 KiB (1.1 MiB / s)

https://repo1.maven.org/maven2/org/json4s/json4s-core_2.11/3.2.11/json4s-core_2.11-3.2.11.jar

100.0% [#####] 575.6 KiB (3.7 MiB / s)

<https://repo1.maven.org/maven2/org/slf4j/jcl-over-slf4j/1.7.16/jcl-over-slf4j-1.7.16.jar>

100.0% [#####] 16.0 KiB (108.4 KiB / s)

https://repo1.maven.org/maven2/org/scalatest/scalatest_2.11/2.2.6/scalatest_2.11-2.2.6.jar

100.0% [#####] 7.1 MiB (10.6 MiB / s)

https://repo1.maven.org/maven2/org/apache/spark/spark-launcher_2.11/2.1.0/spark-launcher_2.11-2.1.0.jar

100.0% [#####] 65.0 KiB (409.1 KiB / s)

<https://repo1.maven.org/maven2/org/apache/avro/avro-ipc/1.7.7/avro-ipc-1.7.7-tests.jar>

100.0% [#####] 338.5 KiB (2.2 MiB / s)

<https://repo1.maven.org/maven2/org/scala-lang/scala-compiler/2.11.8/scala-compiler-2.11.8.jar>

100.0% [#####] 14.8 MiB (11.4 MiB / s)

<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-mapreduce-client-shuffle/2.2.0/hadoop-mapreduce-client-shuffle-2.2.0.jar>

100.0% [#####] 21.0 KiB (146.1 KiB / s)
<https://repo1.maven.org/maven2/org/apache/parquet/parquet-jackson/1.8.1/parquet-jackson-1.8.1.jar>
100.0% [#####] 1023.6 KiB (5.6 MiB / s)
<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-annotations/2.2.0/hadoop-annotations-2.2.0.jar>
100.0% [#####] 16.4 KiB (124.1 KiB / s)
<https://repo1.maven.org/maven2/org/apache/zookeeper/zookeeper/3.4.5/zookeeper-3.4.5.jar>
100.0% [#####] 761.7 KiB (4.3 MiB / s)
<https://repo1.maven.org/maven2/javax/ws/rs/javax.ws.rs-api/2.0.1/javax.ws.rs-api-2.0.1.jar>
100.0% [#####] 112.8 KiB (778.1 KiB / s)
<https://repo1.maven.org/maven2/org/glassfish/jersey/bundles/repackaged/jersey-guava/2.22.2/jersey-guava-2.22.2.jar>
100.0% [#####] 948.5 KiB (5.8 MiB / s)
<https://repo1.maven.org/maven2/com/fasterxml/jackson/core/jackson-core/2.6.5/jackson-core-2.6.5.jar>
100.0% [#####] 252.8 KiB (1.8 MiB / s)
<https://repo1.maven.org/maven2/org/apache/avro/avro/1.7.7/avro-1.7.7.jar>
100.0% [#####] 426.1 KiB (2.9 MiB / s)
<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-yarn-common/2.2.0/hadoop-yarn-common-2.2.0.jar>
100.0% [#####] 1.2 MiB (7.2 MiB / s)
<https://repo1.maven.org/maven2/commons-beanutils/commons-beanutils-core/1.8.0/commons-beanutils-core-1.8.0.jar>
100.0% [#####] 201.2 KiB (1.5 MiB / s)
<https://repo1.maven.org/maven2/net/sf/py4j/py4j/0.10.4/py4j-0.10.4.jar>
100.0% [#####] 112.8 KiB (761.9 KiB / s)
<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-client/2.2.0/hadoop-client-2.2.0.jar>
100.0% [#####] 2.5 KiB (17.5 KiB / s)
<https://repo1.maven.org/maven2/org/codehaus/jackson/jackson-core-asl/1.9.13/jackson-core-asl-1.9.13.jar>
100.0% [#####] 226.8 KiB (1.5 MiB / s)
<https://repo1.maven.org/maven2/jline/jline/2.12.1/jline-2.12.1.jar>
100.0% [#####] 208.9 KiB (1.4 MiB / s)
<https://repo1.maven.org/maven2/log4j/log4j/1.2.17/log4j-1.2.17.jar>
100.0% [#####] 478.4 KiB (3.0 MiB / s)
<https://repo1.maven.org/maven2/com/google/guava/guava/14.0.1/guava-14.0.1.jar>
100.0% [#####] 2.1 MiB (7.2 MiB / s)
https://repo1.maven.org/maven2/org/apache/spark/spark-core_2.11/2.1.0/spark-core_2.11-2.1.0.jar
100.0% [#####] 11.3 MiB (11.0 MiB / s)
[info] Fetched artifacts of
[success] Total time: 19 s, completed May 25, 2023 8:03:20 AM

From <<https://bigdata.cloudlabs.edureka.co/shell/>>

Step 4:

Now create source folder directory `mkdir -p src/main/scala/com/edureka/training`

This is where scala file will be placed.

d. Create scala file with code: vi SparkSort.scala

```
import org.apache.spark.sql.Session

object HdfsTest {
  /** Usage: HdfsTest [file] */
  def main(args: Array[String]) {
    if (args.length < 1) {
      System.err.println("Usage: HdfsTest")
      System.exit(1)
    }

    val spark = SparkSession.builder.appName("CustomHdfsTest").getOrCreate()
    val file = spark.read.text(args(0)).rdd
    val mapped = file.map(s => s.length).cache()

    for (iter <- 1 to 10) {
      val start = System.currentTimeMillis()
      for (x <- mapped) {
        x + 2
      }
      val end = System.currentTimeMillis()
      println("-->Iteration " + iter + " took " + (end - start) + " ms")
    }

    spark.stop()
  }
}
```

e. Compile the scala package using the command sbt package

```
[raghavendr48edu@ip-10-1-1-204 training_project2]$ sbt package
[info] welcome to sbt 1.5.5 (Oracle Corporation Java 1.8.0_181)
[info] loading project definition from
/home/raghavendr48edu/training_project2/project
[info] loading settings for project training_project2 from build.sbt ...
[info] set current project to SparkMe Project (in build
file:/home/raghavendr48edu/training\_project2/)
[info] compiling 1 Scala source to
/home/raghavendr48edu/training_project2/target/scala-2.11/classes ...
[success] Total time: 5 s, completed May 25, 2023 10:18:14 AM
[raghavendr48edu@ip-10-1-1-204 training_project2]$
```

From <<https://bigdata.cloudlabs.edureka.co/shell/>>

f. Copy the input_sort_py.txt file to HDFS from local FTP

hdfs dfs -copyFromLocal /home/raghavendr48edu/input_sort_py.txt

g. Submit the spark job with the scala code

spark-submit --class com.edureka.training.m4.SparkSort --deploy-mode client target/scala-2.11/sparkme-project_2.11-1.0.jar input_sort_py.txt

[raghavendr48edu@ip-10-1-1-204 training_project3]\$ **spark-submit --class HdfsTest --deploy-mode client target/scala-2.11/sparkme-project_2.11-1.0.jar input_sort_py.txt**

23/05/25 12:27:09 INFO spark.SparkContext: Running Spark version 2.4.0-cdh6.2.1

23/05/25 12:27:09 INFO logging.DriverLogger: Added a local log appender at: /tmp/spark-149fb59c-5b64-4b50-a117-6212d417e2c9/___driver_logs___/driver.log

23/05/25 12:27:09 INFO spark.SparkContext: Submitted application: CustomHdfsTest

23/05/25 12:27:09 INFO spark.SecurityManager: Changing view acls to: raghavendr48edu

23/05/25 12:27:09 INFO spark.SecurityManager: Changing modify acls to: raghavendr48edu

23/05/25 12:27:09 INFO spark.SecurityManager: Changing view acls groups to:

23/05/25 12:27:09 INFO spark.SecurityManager: Changing modify acls groups to:

23/05/25 12:27:09 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(raghavendr48edu);

groups with view permissions: Set(); users with modify permissions: Set(raghavendr48edu); groups with modify permissions: Set()

23/05/25 12:27:09 INFO util.Utils: Successfully started service 'sparkDriver' on port 40066.

23/05/25 12:27:09 INFO spark.SparkEnv: Registering MapOutputTracker

23/05/25 12:27:09 INFO spark.SparkEnv: Registering BlockManagerMaster

23/05/25 12:27:09 INFO storage.BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information

23/05/25 12:27:09 INFO storage.BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up

23/05/25 12:27:09 INFO storage.DiskBlockManager: Created local directory at /tmp/blockmgr-7f3f18dd-6ed8-4d58-b499-fd04eef80a15

23/05/25 12:27:09 INFO memory.MemoryStore: MemoryStore started with capacity 366.3 MB

23/05/25 12:27:09 INFO spark.SparkEnv: Registering OutputCommitCoordinator

23/05/25 12:27:09 INFO spark.SparkContext: Added JAR
file:/home/raghavendr48edu/training_project3/target/scala-2.11/sparkme-project_2.11-1.0.jar at spark://ip-10-1-1-204.ap-south-1.compute.internal:40066/jars/sparkme-project_2.11-1.0.jar with timestamp 1685017629956

23/05/25 12:27:10 INFO util.Utils: Using initial executors = 0, max of spark.dynamicAllocation.initialExecutors, spark.dynamicAllocation.minExecutors and spark.executor.instances

23/05/25 12:27:10 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032

23/05/25 12:27:10 INFO yarn.Client: Requesting a new application from cluster with 3 NodeManagers

23/05/25 12:27:10 INFO conf.Configuration: resource-types.xml not found

23/05/25 12:27:10 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.

23/05/25 12:27:10 INFO yarn.Client: Verifying our application has not requested more than the maximum memory capability of the cluster (20480 MB per container)

23/05/25 12:27:10 INFO yarn.Client: Will allocate AM container, with 896 MB memory including 384 MB overhead

23/05/25 12:27:10 INFO yarn.Client: Setting up container launch context for our AM

23/05/25 12:27:10 INFO yarn.Client: Setting up the launch environment for our AM container

23/05/25 12:27:10 INFO yarn.Client: Preparing resources for our AM container

23/05/25 12:27:11 INFO yarn.Client: Uploading resource file:/tmp/spark-149fb59c-5b64-4b50-a117-6212d417e2c9/spark_conf_4441404531900358458.zip -> hdfs://nameservice1/user/raghavendr48edu/.sparkStaging/application_1684866872278_1272/___spark_conf__.zip

23/05/25 12:27:11 INFO spark.SecurityManager: Changing view acls to: raghavendr48edu

23/05/25 12:27:11 INFO spark.SecurityManager: Changing modify acls to: raghavendr48edu

23/05/25 12:27:11 INFO spark.SecurityManager: Changing view acls groups to:

23/05/25 12:27:11 INFO spark.SecurityManager: Changing modify acls groups to:

23/05/25 12:27:11 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(raghavendr48edu); groups with view permissions: Set(); users with modify permissions: Set(raghavendr48edu); groups with modify permissions: Set()

23/05/25 12:27:11 INFO yarn.Client: Submitting application application_1684866872278_1272 to ResourceManager

23/05/25 12:27:11 INFO impl.YarnClientImpl: Submitted application application_1684866872278_1272

23/05/25 12:27:12 INFO yarn.Client: Application report for application_1684866872278_1272 (state: ACCEPTED)

23/05/25 12:27:12 INFO yarn.Client: client token: N/A

diagnostics: AM container is launched, waiting for AM container to Register with RM

ApplicationMaster host: N/A

ApplicationMaster RPC port: -1

queue: root.default

start time: 1685017631490

final status: UNDEFINED

tracking URL: http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_1272/

user: raghavendr48edu

23/05/25 12:27:13 INFO yarn.Client: Application report for application_1684866872278_1272 (state: ACCEPTED)

23/05/25 12:27:14 INFO yarn.Client: Application report for application_1684866872278_1272 (state: ACCEPTED)

23/05/25 12:27:15 INFO yarn.Client: Application report for application_1684866872278_1272 (state: ACCEPTED)

23/05/25 12:27:16 INFO yarn.Client: Application report for application_1684866872278_1272 (state: ACCEPTED)

23/05/25 12:27:17 INFO yarn.Client: Application report for application_1684866872278_1272 (state: ACCEPTED)

23/05/25 12:27:18 INFO yarn.Client: Application report for application_1684866872278_1272 (state: ACCEPTED)

23/05/25 12:27:19 INFO yarn.Client: Application report for application_1684866872278_1272 (state: ACCEPTED)

23/05/25 12:27:20 INFO yarn.Client: Application report for application_1684866872278_1272 (state: RUNNING)

23/05/25 12:27:20 INFO yarn.Client:

client token: N/A

diagnostics: N/A

ApplicationMaster host: 10.1.2.40

ApplicationMaster RPC port: -1

queue: root.default

start time: 1685017631490

final status: UNDEFINED

tracking URL: http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_1272/

user: raghavendr48edu

23/05/25 12:27:20 INFO cluster.YarnClientSchedulerBackend: Application application_1684866872278_1272 has started running.

23/05/25 12:27:20 INFO cluster.SchedulerExtensionServices: Starting Yarn extension services with app application_1684866872278_1272 and attemptId None

23/05/25 12:27:20 INFO util.Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 37637.

23/05/25 12:27:20 INFO netty.NettyBlockTransferService: Server created on ip-10-1-1-204.ap-south-1.compute.internal:37637

23/05/25 12:27:20 INFO storage.BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy

23/05/25 12:27:20 INFO cluster.YarnClientSchedulerBackend: Add WebUI Filter.
org.apache.hadoop.yarn.server.webproxy.amfilter.AmIpFilter,
Map(PROXY_HOSTS -> ip-10-1-1-204.ap-south-1.compute.internal, PROXY_URI_BASES -> http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_1272), /pr
oxy/application_1684866872278_1272
23/05/25 12:27:20 INFO storage.BlockManagerMaster: Registering
BlockManager BlockManagerId(driver, ip-10-1-1-204.ap-south-1.compute.internal, 37637, None)
23/05/25 12:27:20 INFO storage.BlockManagerMasterEndpoint: Registering
block manager ip-10-1-1-204.ap-south-1.compute.internal:37637 with 366.3 MB
RAM, Block
ManagerId(driver, ip-10-1-1-204.ap-south-1.compute.internal, 37637, None)
23/05/25 12:27:20 INFO storage.BlockManagerMaster: Registered
BlockManager BlockManagerId(driver, ip-10-1-1-204.ap-south-1.compute.internal, 37637, None)
23/05/25 12:27:20 INFO storage.BlockManager: external shuffle service port =
7337
23/05/25 12:27:20 INFO storage.BlockManager: Initialized BlockManager:
BlockManagerId(driver, ip-10-1-1-204.ap-south-1.compute.internal, 37637,
None)
23/05/25 12:27:20 INFO util.log: Logging initialized @12715ms
23/05/25 12:27:21 INFO scheduler.EventLoggingListener: Logging events to
hdfs://nameservice1/user/spark/applicationHistory/application_1684866872278_1272
23/05/25 12:27:21 INFO util.Utils: Using initial executors = 0, max of
spark.dynamicAllocation.initialExecutors, spark.dynamicAllocation.minExecutors
and spa
rk.executor.instances
23/05/25 12:27:21 WARN
cluster.YarnSchedulerBackend\$YarnSchedulerEndpoint: Attempted to request
executors before the AM has registered!
23/05/25 12:27:21 INFO spark.SparkContext: Registered listener
com.cloudera.spark.lineage.NavigatorAppListener
23/05/25 12:27:21 INFO logging.DriverLogger\$DfsAsyncWriter: Started driver
log file sync to:
/user/spark/driverLogs/application_1684866872278_1272_driver.log
23/05/25 12:27:21 INFO cluster.YarnClientSchedulerBackend:
SchedulerBackend is ready for scheduling beginning after reached
minRegisteredResourcesRatio: 0.8
23/05/25 12:27:21 INFO internal.SharedState: loading hive config file:
<file:/etc/hive/conf.cloudera.hive/hive-site.xml>
23/05/25 12:27:21 INFO internal.SharedState: spark.sql.warehouse.dir is not
set, but hive.metastore.warehouse.dir is set. Setting spark.sql.warehouse.dir to
the value of hive.metastore.warehouse.dir ('/user/hive/warehouse').
23/05/25 12:27:21 INFO internal.SharedState: Warehouse path is
'/user/hive/warehouse'.
23/05/25 12:27:21 INFO state.StateStoreCoordinatorRef: Registered
StateStoreCoordinator endpoint
23/05/25 12:27:23 INFO datasources.FileSourceStrategy: Pruning directories
with:

23/05/25 12:27:23 INFO datasources.FileSourceStrategy: Post-Scan Filters:
23/05/25 12:27:23 INFO datasources.FileSourceStrategy: Output Data Schema:
struct<value: string>
23/05/25 12:27:23 INFO execution.FileSourceScanExec: Pushed Filters:
23/05/25 12:27:23 INFO
cluster.YarnSchedulerBackend\$YarnSchedulerEndpoint: ApplicationMaster
registered as NettyRpcEndpointRef(spark-client://YarnAM)
23/05/25 12:27:24 INFO codegen.CodeGenerator: Code generated in
183.770987 ms
23/05/25 12:27:24 INFO memory.MemoryStore: Block broadcast_0 stored as
values in memory (estimated size 336.4 KB, free 366.0 MB)
23/05/25 12:27:24 INFO memory.MemoryStore: Block broadcast_0_piece0
stored as bytes in memory (estimated size 32.2 KB, free 365.9 MB)
23/05/25 12:27:24 INFO storage.BlockManagerInfo: Added broadcast_0_piece0
in memory on ip-10-1-1-204.ap-south-1.compute.internal:37637 (size: 32.2 KB,
free:
366.3 MB)
23/05/25 12:27:24 INFO spark.SparkContext: Created broadcast 0 from rdd at
SparkSort.scala:23
23/05/25 12:27:24 INFO execution.FileSourceScanExec: Planning scan with bin
packing, max size: 4194304 bytes, open cost is considered as scanning 4194304
byt
es.
23/05/25 12:27:24 INFO spark.SparkContext: Starting job: foreach at
SparkSort.scala:33
23/05/25 12:27:24 INFO scheduler.DAGScheduler: Got job 0 (foreach at
SparkSort.scala:33) with 1 output partitions
23/05/25 12:27:24 INFO scheduler.DAGScheduler: Final stage: ResultStage 0
(foreach at SparkSort.scala:33)
23/05/25 12:27:24 INFO scheduler.DAGScheduler: Parents of final stage: List()
23/05/25 12:27:24 INFO scheduler.DAGScheduler: Missing parents: List()
23/05/25 12:27:24 INFO scheduler.DAGScheduler: Submitting ResultStage 0
(MapPartitionsRDD[4] at map at SparkSort.scala:25), which has no missing
parents
23/05/25 12:27:24 INFO memory.MemoryStore: Block broadcast_1 stored as
values in memory (estimated size 10.6 KB, free 365.9 MB)
23/05/25 12:27:24 INFO memory.MemoryStore: Block broadcast_1_piece0
stored as bytes in memory (estimated size 5.6 KB, free 365.9 MB)
23/05/25 12:27:24 INFO storage.BlockManagerInfo: Added broadcast_1_piece0
in memory on ip-10-1-1-204.ap-south-1.compute.internal:37637 (size: 5.6 KB,
free: 3
66.3 MB)
23/05/25 12:27:24 INFO spark.SparkContext: Created broadcast 1 from
broadcast at DAGScheduler.scala:1164
23/05/25 12:27:24 INFO scheduler.DAGScheduler: Submitting 1 missing tasks
from ResultStage 0 (MapPartitionsRDD[4] at map at SparkSort.scala:25) (first
15 tas
ks are for partitions Vector(0))
23/05/25 12:27:24 INFO cluster.YarnScheduler: Adding task set 0.0 with 1
tasks
23/05/25 12:27:24 INFO spark.ExecutorAllocationManager: Requesting 1 new
executor because tasks are backlogged (new desired total will be 1)

23/05/25 12:27:35 INFO cluster.YarnSchedulerBackend\$YarnDriverEndpoint:
Registered executor NettyRpcEndpointRef(spark-client://Executor)
(10.1.2.40:54824) with ID 1
23/05/25 12:27:35 INFO spark.ExecutorAllocationManager: New executor 1 has
registered (new total is 1)
23/05/25 12:27:35 INFO scheduler.TaskSetManager: Starting task 0.0 in stage
0.0 (TID 0, ip-10-1-2-40.ap-south-1.compute.internal, executor 1, partition 0,
NO
DE_LOCAL, 8347 bytes)
23/05/25 12:27:35 INFO storage.BlockManagerMasterEndpoint: Registering
block manager ip-10-1-2-40.ap-south-1.compute.internal:33447 with 366.3 MB
RAM, BlockManagerId(1, ip-10-1-2-40.ap-south-1.compute.internal, 33447, None)
23/05/25 12:27:36 INFO storage.BlockManagerInfo: Added broadcast_1_piece0
in memory on ip-10-1-2-40.ap-south-1.compute.internal:33447 (size: 5.6 KB,
free: 366.3 MB)
23/05/25 12:27:46 INFO storage.BlockManagerInfo: Added broadcast_0_piece0
in memory on ip-10-1-2-40.ap-south-1.compute.internal:33447 (size: 32.2 KB,
free: 366.3 MB)
23/05/25 12:27:50 INFO storage.BlockManagerInfo: Added rdd_4_0 in memory
on ip-10-1-2-40.ap-south-1.compute.internal:33447 (size: 32.0 B, free: 366.3
MB)
23/05/25 12:27:50 INFO scheduler.TaskSetManager: Finished task 0.0 in stage
0.0 (TID 0) in 15041 ms on ip-10-1-2-40.ap-south-1.compute.internal (executor
1)
(1/1)
23/05/25 12:27:50 INFO cluster.YarnScheduler: Removed TaskSet 0.0, whose
tasks have all completed, from pool
23/05/25 12:27:50 INFO scheduler.DAGScheduler: ResultStage 0 (foreach at
SparkSort.scala:33) finished in 25.723 s
23/05/25 12:27:50 INFO scheduler.DAGScheduler: Job 0 finished: foreach at
SparkSort.scala:33, took 25.772812 s
--> Iteration 1 took 25786 ms
23/05/25 12:27:50 INFO spark.SparkContext: Starting job: foreach at
SparkSort.scala:33
23/05/25 12:27:50 INFO scheduler.DAGScheduler: Got job 1 (foreach at
SparkSort.scala:33) with 1 output partitions
23/05/25 12:27:50 INFO scheduler.DAGScheduler: Final stage: ResultStage 1
(foreach at SparkSort.scala:33)
23/05/25 12:27:50 INFO scheduler.DAGScheduler: Parents of final stage: List()
23/05/25 12:27:50 INFO scheduler.DAGScheduler: Missing parents: List()
23/05/25 12:27:50 INFO scheduler.DAGScheduler: Submitting ResultStage 1
(MapPartitionsRDD[4] at map at SparkSort.scala:25), which has no missing
parents
23/05/25 12:27:50 INFO memory.MemoryStore: Block broadcast_2 stored as
values in memory (estimated size 10.6 KB, free 365.9 MB)
23/05/25 12:27:50 INFO memory.MemoryStore: Block broadcast_2_piece0
stored as bytes in memory (estimated size 5.6 KB, free 365.9 MB)

23/05/25 12:27:50 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on ip-10-1-1-204.ap-south-1.compute.internal:37637 (size: 5.6 KB, free: 366.3 MB)

23/05/25 12:27:50 INFO spark.SparkContext: Created broadcast 2 from broadcast at DAGScheduler.scala:1164

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 1 (MapPartitionsRDD[4] at map at SparkSort.scala:25) (first 15 tasks are for partitions Vector(0))

23/05/25 12:27:50 INFO cluster.YarnScheduler: Adding task set 1.0 with 1 tasks

23/05/25 12:27:50 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 1.0 (TID 1, ip-10-1-2-40.ap-south-1.compute.internal, executor 1, partition 0, PROCESS_LOCAL, 8347 bytes)

23/05/25 12:27:50 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on ip-10-1-2-40.ap-south-1.compute.internal:33447 (size: 5.6 KB, free: 366.3 MB)

23/05/25 12:27:50 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 1.0 (TID 1) in 244 ms on ip-10-1-2-40.ap-south-1.compute.internal (executor 1) (1/1)

23/05/25 12:27:50 INFO cluster.YarnScheduler: Removed TaskSet 1.0, whose tasks have all completed, from pool

23/05/25 12:27:50 INFO scheduler.DAGScheduler: ResultStage 1 (foreach at SparkSort.scala:33) finished in 0.276 s

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Job 1 finished: foreach at SparkSort.scala:33, took 0.282008 s

-->Iteration 2 took 288 ms

23/05/25 12:27:50 INFO spark.SparkContext: Starting job: foreach at SparkSort.scala:33

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Got job 2 (foreach at SparkSort.scala:33) with 1 output partitions

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Final stage: ResultStage 2 (foreach at SparkSort.scala:33)

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Parents of final stage: List()

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Missing parents: List()

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Submitting ResultStage 2 (MapPartitionsRDD[4] at map at SparkSort.scala:25), which has no missing parents

23/05/25 12:27:50 INFO memory.MemoryStore: Block broadcast_3 stored as values in memory (estimated size 10.6 KB, free 365.9 MB)

23/05/25 12:27:50 INFO memory.MemoryStore: Block broadcast_3_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.9 MB)

23/05/25 12:27:50 INFO storage.BlockManagerInfo: Added broadcast_3_piece0 in memory on ip-10-1-1-204.ap-south-1.compute.internal:37637 (size: 5.6 KB, free: 366.3 MB)

23/05/25 12:27:50 INFO spark.SparkContext: Created broadcast 3 from broadcast at DAGScheduler.scala:1164

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 2 (MapPartitionsRDD[4] at map at SparkSort.scala:25) (first 15 tasks are for partitions Vector(0))

23/05/25 12:27:50 INFO cluster.YarnScheduler: Adding task set 2.0 with 1 tasks

23/05/25 12:27:50 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 2.0 (TID 2, ip-10-1-2-40.ap-south-1.compute.internal, executor 1, partition 0, PROCESS_LOCAL, 8347 bytes)

23/05/25 12:27:50 INFO storage.BlockManagerInfo: Added broadcast_3_piece0 in memory on ip-10-1-2-40.ap-south-1.compute.internal:33447 (size: 5.6 KB, free: 366.3 MB)

23/05/25 12:27:50 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 2.0 (TID 2) in 95 ms on ip-10-1-2-40.ap-south-1.compute.internal (executor 1) (1/1)

23/05/25 12:27:50 INFO cluster.YarnScheduler: Removed TaskSet 2.0, whose tasks have all completed, from pool

23/05/25 12:27:50 INFO scheduler.DAGScheduler: ResultStage 2 (foreach at SparkSort.scala:33) finished in 0.111 s

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Job 2 finished: foreach at SparkSort.scala:33, took 0.118225 s

-->Iteration 3 took 125 ms

23/05/25 12:27:50 INFO spark.SparkContext: Starting job: foreach at SparkSort.scala:33

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Got job 3 (foreach at SparkSort.scala:33) with 1 output partitions

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Final stage: ResultStage 3 (foreach at SparkSort.scala:33)

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Parents of final stage: List()

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Missing parents: List()

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Submitting ResultStage 3 (MapPartitionsRDD[4] at map at SparkSort.scala:25), which has no missing parents

23/05/25 12:27:50 INFO memory.MemoryStore: Block broadcast_4 stored as values in memory (estimated size 10.6 KB, free 365.9 MB)

23/05/25 12:27:50 INFO memory.MemoryStore: Block broadcast_4_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.9 MB)

23/05/25 12:27:50 INFO storage.BlockManagerInfo: Added broadcast_4_piece0 in memory on ip-10-1-1-204.ap-south-1.compute.internal:37637 (size: 5.6 KB, free: 366.2 MB)

23/05/25 12:27:50 INFO spark.SparkContext: Created broadcast 4 from broadcast at DAGScheduler.scala:1164

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 3 (MapPartitionsRDD[4] at map at SparkSort.scala:25) (first 15 tasks are for partitions Vector(0))

23/05/25 12:27:50 INFO cluster.YarnScheduler: Adding task set 3.0 with 1 tasks

23/05/25 12:27:50 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 3.0 (TID 3, ip-10-1-2-40.ap-south-1.compute.internal, executor 1, partition 0, PROCESS_LOCAL, 8347 bytes)

23/05/25 12:27:50 INFO storage.BlockManagerInfo: Added broadcast_4_piece0 in memory on ip-10-1-2-40.ap-south-1.compute.internal:33447 (size: 5.6 KB, free: 36 6.2 MB)

23/05/25 12:27:50 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 3.0 (TID 3) in 70 ms on ip-10-1-2-40.ap-south-1.compute.internal (executor 1) (1/1)

23/05/25 12:27:50 INFO cluster.YarnScheduler: Removed TaskSet 3.0, whose tasks have all completed, from pool

23/05/25 12:27:50 INFO scheduler.DAGScheduler: ResultStage 3 (foreach at SparkSort.scala:33) finished in 0.083 s

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Job 3 finished: foreach at SparkSort.scala:33, took 0.088030 s

-->Iteration 4 took 93 ms

23/05/25 12:27:50 INFO spark.SparkContext: Starting job: foreach at SparkSort.scala:33

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Got job 4 (foreach at SparkSort.scala:33) with 1 output partitions

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Final stage: ResultStage 4 (foreach at SparkSort.scala:33)

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Parents of final stage: List()

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Missing parents: List()

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Submitting ResultStage 4 (MapPartitionsRDD[4] at map at SparkSort.scala:25), which has no missing parents

23/05/25 12:27:50 INFO memory.MemoryStore: Block broadcast_5 stored as values in memory (estimated size 10.6 KB, free 365.9 MB)

23/05/25 12:27:50 INFO memory.MemoryStore: Block broadcast_5_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.9 MB)

23/05/25 12:27:50 INFO storage.BlockManagerInfo: Added broadcast_5_piece0 in memory on ip-10-1-1-204.ap-south-1.compute.internal:37637 (size: 5.6 KB, free: 3 66.2 MB)

23/05/25 12:27:50 INFO spark.SparkContext: Created broadcast 5 from broadcast at DAGScheduler.scala:1164

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 4 (MapPartitionsRDD[4] at map at SparkSort.scala:25) (first 15 tasks are for partitions Vector(0))

23/05/25 12:27:50 INFO cluster.YarnScheduler: Adding task set 4.0 with 1 tasks

23/05/25 12:27:50 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 4.0 (TID 4, ip-10-1-2-40.ap-south-1.compute.internal, executor 1, partition 0, PROCESS_LOCAL, 8347 bytes)

23/05/25 12:27:50 INFO storage.BlockManagerInfo: Added broadcast_5_piece0 in memory on ip-10-1-2-40.ap-south-1.compute.internal:33447 (size: 5.6 KB, free: 36 6.2 MB)

23/05/25 12:27:50 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 4.0 (TID 4) in 104 ms on ip-10-1-2-40.ap-south-1.compute.internal (executor 1) (1 /1)

23/05/25 12:27:50 INFO cluster.YarnScheduler: Removed TaskSet 4.0, whose tasks have all completed, from pool

23/05/25 12:27:50 INFO scheduler.DAGScheduler: ResultStage 4 (foreach at SparkSort.scala:33) finished in 0.115 s

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Job 4 finished: foreach at SparkSort.scala:33, took 0.117970 s

-->Iteration 5 took 123 ms

23/05/25 12:27:50 INFO spark.SparkContext: Starting job: foreach at SparkSort.scala:33

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Got job 5 (foreach at SparkSort.scala:33) with 1 output partitions

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Final stage: ResultStage 5 (foreach at SparkSort.scala:33)

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Parents of final stage: List()

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Missing parents: List()

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Submitting ResultStage 5 (MapPartitionsRDD[4] at map at SparkSort.scala:25), which has no missing parents

23/05/25 12:27:50 INFO memory.MemoryStore: Block broadcast_6 stored as values in memory (estimated size 10.6 KB, free 365.9 MB)

23/05/25 12:27:50 INFO memory.MemoryStore: Block broadcast_6_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.8 MB)

23/05/25 12:27:50 INFO storage.BlockManagerInfo: Added broadcast_6_piece0 in memory on ip-10-1-1-204.ap-south-1.compute.internal:37637 (size: 5.6 KB, free: 3 66.2 MB)

23/05/25 12:27:50 INFO spark.SparkContext: Created broadcast 6 from broadcast at DAGScheduler.scala:1164

23/05/25 12:27:50 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 5 (MapPartitionsRDD[4] at map at SparkSort.scala:25) (first 15 tasks are for partitions Vector(0))

23/05/25 12:27:50 INFO cluster.YarnScheduler: Adding task set 5.0 with 1 tasks

23/05/25 12:27:50 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 5.0 (TID 5, ip-10-1-2-40.ap-south-1.compute.internal, executor 1, partition 0, PROCESS_LOCAL, 8347 bytes)

23/05/25 12:27:51 INFO storage.BlockManagerInfo: Added broadcast_6_piece0 in memory on ip-10-1-2-40.ap-south-1.compute.internal:33447 (size: 5.6 KB, free: 36 6.2 MB)

23/05/25 12:27:51 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 5.0 (TID 5) in 493 ms on ip-10-1-2-40.ap-south-1.compute.internal (executor 1) (1/1)

23/05/25 12:27:51 INFO cluster.YarnScheduler: Removed TaskSet 5.0, whose tasks have all completed, from pool

23/05/25 12:27:51 INFO scheduler.DAGScheduler: ResultStage 5 (foreach at SparkSort.scala:33) finished in 0.506 s

23/05/25 12:27:51 INFO scheduler.DAGScheduler: Job 5 finished: foreach at SparkSort.scala:33, took 0.510316 s

-->Iteration 6 took 516 ms

23/05/25 12:27:51 INFO spark.SparkContext: Starting job: foreach at SparkSort.scala:33

23/05/25 12:27:51 INFO scheduler.DAGScheduler: Got job 6 (foreach at SparkSort.scala:33) with 1 output partitions

23/05/25 12:27:51 INFO scheduler.DAGScheduler: Final stage: ResultStage 6 (foreach at SparkSort.scala:33)

23/05/25 12:27:51 INFO scheduler.DAGScheduler: Parents of final stage: List()

23/05/25 12:27:51 INFO scheduler.DAGScheduler: Missing parents: List()

23/05/25 12:27:51 INFO scheduler.DAGScheduler: Submitting ResultStage 6 (MapPartitionsRDD[4] at map at SparkSort.scala:25), which has no missing parents

23/05/25 12:27:51 INFO memory.MemoryStore: Block broadcast_7 stored as values in memory (estimated size 10.6 KB, free 365.8 MB)

23/05/25 12:27:51 INFO memory.MemoryStore: Block broadcast_7_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.8 MB)

23/05/25 12:27:51 INFO storage.BlockManagerInfo: Added broadcast_7_piece0 in memory on ip-10-1-1-204.ap-south-1.compute.internal:37637 (size: 5.6 KB, free: 366.2 MB)

23/05/25 12:27:51 INFO spark.SparkContext: Created broadcast 7 from broadcast at DAGScheduler.scala:1164

23/05/25 12:27:51 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 6 (MapPartitionsRDD[4] at map at SparkSort.scala:25) (first 15 tasks are for partitions Vector(0))

23/05/25 12:27:51 INFO cluster.YarnScheduler: Adding task set 6.0 with 1 tasks

23/05/25 12:27:51 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 6.0 (TID 6, ip-10-1-2-40.ap-south-1.compute.internal, executor 1, partition 0, PROCESS_LOCAL, 8347 bytes)

23/05/25 12:27:51 INFO storage.BlockManagerInfo: Added broadcast_7_piece0 in memory on ip-10-1-2-40.ap-south-1.compute.internal:33447 (size: 5.6 KB, free: 366.2 MB)

23/05/25 12:27:51 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 6.0 (TID 6) in 94 ms on ip-10-1-2-40.ap-south-1.compute.internal (executor 1) (1/1)

23/05/25 12:27:51 INFO cluster.YarnScheduler: Removed TaskSet 6.0, whose tasks have all completed, from pool

23/05/25 12:27:51 INFO scheduler.DAGScheduler: ResultStage 6 (foreach at SparkSort.scala:33) finished in 0.108 s
23/05/25 12:27:51 INFO scheduler.DAGScheduler: Job 6 finished: foreach at SparkSort.scala:33, took 0.110870 s
-->Iteration 7 took 115 ms
23/05/25 12:27:51 INFO spark.SparkContext: Starting job: foreach at SparkSort.scala:33
23/05/25 12:27:51 INFO scheduler.DAGScheduler: Got job 7 (foreach at SparkSort.scala:33) with 1 output partitions
23/05/25 12:27:51 INFO scheduler.DAGScheduler: Final stage: ResultStage 7 (foreach at SparkSort.scala:33)
23/05/25 12:27:51 INFO scheduler.DAGScheduler: Parents of final stage: List()
23/05/25 12:27:51 INFO scheduler.DAGScheduler: Missing parents: List()
23/05/25 12:27:51 INFO scheduler.DAGScheduler: Submitting ResultStage 7 (MapPartitionsRDD[4] at map at SparkSort.scala:25), which has no missing parents
23/05/25 12:27:51 INFO memory.MemoryStore: Block broadcast_8 stored as values in memory (estimated size 10.6 KB, free 365.8 MB)
23/05/25 12:27:51 INFO memory.MemoryStore: Block broadcast_8_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.8 MB)
23/05/25 12:27:51 INFO storage.BlockManagerInfo: Added broadcast_8_piece0 in memory on ip-10-1-1-204.ap-south-1.compute.internal:37637 (size: 5.6 KB, free: 366.2 MB)
23/05/25 12:27:51 INFO spark.SparkContext: Created broadcast 8 from broadcast at DAGScheduler.scala:1164
23/05/25 12:27:51 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 7 (MapPartitionsRDD[4] at map at SparkSort.scala:25) (first 15 tasks are for partitions Vector(0))
23/05/25 12:27:51 INFO cluster.YarnScheduler: Adding task set 7.0 with 1 tasks
23/05/25 12:27:51 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 7.0 (TID 7, ip-10-1-2-40.ap-south-1.compute.internal, executor 1, partition 0, PROCESS_LOCAL, 8347 bytes)
23/05/25 12:27:51 INFO storage.BlockManagerInfo: Added broadcast_8_piece0 in memory on ip-10-1-2-40.ap-south-1.compute.internal:33447 (size: 5.6 KB, free: 366.2 MB)
23/05/25 12:27:51 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 7.0 (TID 7) in 70 ms on ip-10-1-2-40.ap-south-1.compute.internal (executor 1) (1/1)
23/05/25 12:27:51 INFO cluster.YarnScheduler: Removed TaskSet 7.0, whose tasks have all completed, from pool
23/05/25 12:27:51 INFO scheduler.DAGScheduler: ResultStage 7 (foreach at SparkSort.scala:33) finished in 0.081 s
23/05/25 12:27:51 INFO scheduler.DAGScheduler: Job 7 finished: foreach at SparkSort.scala:33, took 0.084157 s
-->Iteration 8 took 89 ms

23/05/25 12:27:51 INFO spark.SparkContext: Starting job: foreach at SparkSort.scala:33
23/05/25 12:27:51 INFO scheduler.DAGScheduler: Got job 8 (foreach at SparkSort.scala:33) with 1 output partitions
23/05/25 12:27:51 INFO scheduler.DAGScheduler: Final stage: ResultStage 8 (foreach at SparkSort.scala:33)
23/05/25 12:27:51 INFO scheduler.DAGScheduler: Parents of final stage: List()
23/05/25 12:27:51 INFO scheduler.DAGScheduler: Missing parents: List()
23/05/25 12:27:51 INFO scheduler.DAGScheduler: Submitting ResultStage 8 (MapPartitionsRDD[4] at map at SparkSort.scala:25), which has no missing parents
23/05/25 12:27:51 INFO memory.MemoryStore: Block broadcast_9 stored as values in memory (estimated size 10.6 KB, free 365.8 MB)
23/05/25 12:27:51 INFO memory.MemoryStore: Block broadcast_9_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.8 MB)
23/05/25 12:27:51 INFO storage.BlockManagerInfo: Added broadcast_9_piece0 in memory on ip-10-1-1-204.ap-south-1.compute.internal:37637 (size: 5.6 KB, free: 366.2 MB)
23/05/25 12:27:51 INFO spark.SparkContext: Created broadcast 9 from broadcast at DAGScheduler.scala:1164
23/05/25 12:27:51 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 8 (MapPartitionsRDD[4] at map at SparkSort.scala:25) (first 15 tasks are for partitions Vector(0))
23/05/25 12:27:51 INFO cluster.YarnScheduler: Adding task set 8.0 with 1 tasks
23/05/25 12:27:51 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 8.0 (TID 8, ip-10-1-2-40.ap-south-1.compute.internal, executor 1, partition 0, PROCESS_LOCAL, 8347 bytes)
23/05/25 12:27:51 INFO storage.BlockManagerInfo: Added broadcast_9_piece0 in memory on ip-10-1-2-40.ap-south-1.compute.internal:33447 (size: 5.6 KB, free: 366.2 MB)
23/05/25 12:27:51 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 8.0 (TID 8) in 60 ms on ip-10-1-2-40.ap-south-1.compute.internal (executor 1) (1/1)
23/05/25 12:27:51 INFO cluster.YarnScheduler: Removed TaskSet 8.0, whose tasks have all completed, from pool
23/05/25 12:27:51 INFO scheduler.DAGScheduler: ResultStage 8 (foreach at SparkSort.scala:33) finished in 0.072 s
23/05/25 12:27:51 INFO scheduler.DAGScheduler: Job 8 finished: foreach at SparkSort.scala:33, took 0.075601 s
--> Iteration 9 took 80 ms
23/05/25 12:27:51 INFO spark.SparkContext: Starting job: foreach at SparkSort.scala:33
23/05/25 12:27:51 INFO scheduler.DAGScheduler: Got job 9 (foreach at SparkSort.scala:33) with 1 output partitions
23/05/25 12:27:51 INFO scheduler.DAGScheduler: Final stage: ResultStage 9 (foreach at SparkSort.scala:33)

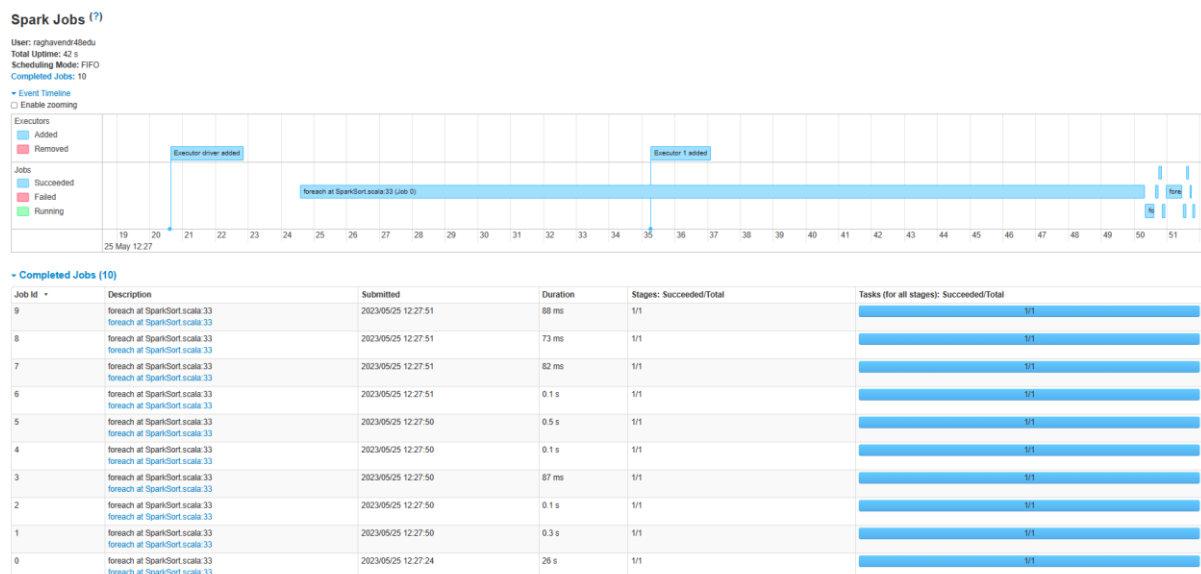
23/05/25 12:27:51 INFO scheduler.DAGScheduler: Parents of final stage: List()
23/05/25 12:27:51 INFO scheduler.DAGScheduler: Missing parents: List()
23/05/25 12:27:51 INFO scheduler.DAGScheduler: Submitting ResultStage 9
(MapPartitionsRDD[4] at map at SparkSort.scala:25), which has no missing
parents
23/05/25 12:27:51 INFO memory.MemoryStore: Block broadcast_10 stored as
values in memory (estimated size 10.6 KB, free 365.8 MB)
23/05/25 12:27:51 INFO memory.MemoryStore: Block broadcast_10_piece0
stored as bytes in memory (estimated size 5.6 KB, free 365.8 MB)
23/05/25 12:27:51 INFO storage.BlockManagerInfo: Added
broadcast_10_piece0 in memory on ip-10-1-1-204.ap-south-
1.compute.internal:37637 (size: 5.6 KB, free:
366.2 MB)
23/05/25 12:27:51 INFO spark.SparkContext: Created broadcast 10 from
broadcast at DAGScheduler.scala:1164
23/05/25 12:27:51 INFO scheduler.DAGScheduler: Submitting 1 missing tasks
from ResultStage 9 (MapPartitionsRDD[4] at map at SparkSort.scala:25) (first
15 tas
ks are for partitions Vector(0))
23/05/25 12:27:51 INFO cluster.YarnScheduler: Adding task set 9.0 with 1
tasks
23/05/25 12:27:51 INFO scheduler.TaskSetManager: Starting task 0.0 in stage
9.0 (TID 9, ip-10-1-2-40.ap-south-1.compute.internal, executor 1, partition 0,
PR
OCESS_LOCAL, 8347 bytes)
23/05/25 12:27:51 INFO storage.BlockManagerInfo: Added
broadcast_10_piece0 in memory on ip-10-1-2-40.ap-south-
1.compute.internal:33447 (size: 5.6 KB, free: 3
66.2 MB)
23/05/25 12:27:51 INFO scheduler.TaskSetManager: Finished task 0.0 in stage
9.0 (TID 9) in 73 ms on ip-10-1-2-40.ap-south-1.compute.internal (executor 1)
(1/
1)
23/05/25 12:27:51 INFO cluster.YarnScheduler: Removed TaskSet 9.0, whose
tasks have all completed, from pool
23/05/25 12:27:51 INFO scheduler.DAGScheduler: ResultStage 9 (foreach at
SparkSort.scala:33) finished in 0.085 s
23/05/25 12:27:51 INFO scheduler.DAGScheduler: Job 9 finished: foreach at
SparkSort.scala:33, took 0.088958 s
-->Iteration 10 took 94 ms
23/05/25 12:27:51 INFO cluster.YarnClientSchedulerBackend: Interrupting
monitor thread
23/05/25 12:27:51 INFO cluster.YarnClientSchedulerBackend: Shutting down all
executors
23/05/25 12:27:51 INFO cluster.YarnSchedulerBackend\$YarnDriverEndpoint:
Asking each executor to shut down
23/05/25 12:27:51 INFO cluster.SchedulerExtensionServices: Stopping
SchedulerExtensionServices
(serviceOption=None,
services=List(),
started=false)
23/05/25 12:27:51 INFO cluster.YarnClientSchedulerBackend: Stopped

```

23/05/25 12:27:51 INFO spark.MapOutputTrackerMasterEndpoint:
MapOutputTrackerMasterEndpoint stopped!
23/05/25 12:27:51 INFO memory.MemoryStore: MemoryStore cleared
23/05/25 12:27:51 INFO storage.BlockManager: BlockManager stopped
23/05/25 12:27:51 INFO storage.BlockManagerMaster: BlockManagerMaster
stopped
23/05/25 12:27:51 INFO
scheduler.OutputCommitCoordinator$OutputCommitCoordinatorEndpoint:
OutputCommitCoordinator stopped!
23/05/25 12:27:51 INFO spark.SparkContext: Successfully stopped
SparkContext
23/05/25 12:27:51 INFO util.ShutdownHookManager: Shutdown hook called
23/05/25 12:27:51 INFO util.ShutdownHookManager: Deleting directory
/tmp/spark-149fb59c-5b64-4b50-a117-6212d417e2c9
23/05/25 12:27:51 INFO util.ShutdownHookManager: Deleting directory
/tmp/spark-aeca6266-9a43-4039-b421-4d6e67bad587
[raghavendr48edu@ip-10-1-1-204 training_project3]$

```

From <<https://bigdata.cloudlabs.edureka.co/shell/>>



2. Try to implement the same example in pyspark and perform spark-submit.

/input_sort_py.txt

```

1 2 3 0 15 8 7
10 2 24 14 1 8 90 101 2 3 0 15 8 7
10 2 24 14 1 8 90 101 2 3 0 15 8 7

```

From

http://bigdata.cloudlabs.edureka.co:8888/hue/filebrowser/view=/user/raghavendr48edu/use_cases/input_sort_py.txt

```
spark-submit --class org.apache.spark.examples.HdfsTest --master  
yarn --deploy-mode cluster /opt/cloudera/parcels/CDH-6.2.1-  
1.cdh6.2.1.p0.1425774/lib/spark/examples/jars/spark-examples*.jar  
use_cases/input_sort_py.txt
```

23/05/18 06:27:30 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032

23/05/18 06:27:30 INFO yarn.Client: Requesting a new application from cluster with 3 NodeManagers

23/05/18 06:27:30 INFO conf.Configuration: resource-types.xml not found

23/05/18 06:27:30 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.

23/05/18 06:27:30 INFO yarn.Client: Verifying our application has not requested more than the maximum memory capability of the cluster (20480 MB per container)

23/05/18 06:27:30 INFO yarn.Client: Will allocate AM container, with 1408 MB memory including 384 MB overhead

23/05/18 06:27:30 INFO yarn.Client: Setting up container launch context for our AM

23/05/18 06:27:30 INFO yarn.Client: Setting up the launch environment for our AM container

23/05/18 06:27:30 INFO yarn.Client: Preparing resources for our AM container

23/05/18 06:27:30 INFO yarn.Client: Uploading resource

file:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/spark/examples/jars/spark-examples_2.11-2.4.0-cdh6.2.1.jar -> hdfs://nameservi

ce1/user/raghavendr48edu/.sparkStaging/application_1684298513961_0025/spark-examples_2.11-2.4.0-cdh6.2.1.jar

23/05/18 06:27:31 INFO yarn.Client: Uploading resource file:/tmp/spark-cfe28f48-65a9-4f7f-8ff8-b6cb5fd22b8f/_spark_conf_4970982468786708687.zip ->

hdfs://nameservice1/user/raghavendr48edu/.sparkStaging/application_1684298513961_0025/___spark_conf__.zip

23/05/18 06:27:31 INFO spark.SecurityManager: Changing view acls to: raghavendr48edu

23/05/18 06:27:31 INFO spark.SecurityManager: Changing modify acls to: raghavendr48edu

23/05/18 06:27:31 INFO spark.SecurityManager: Changing view acls groups to:

23/05/18 06:27:31 INFO spark.SecurityManager: Changing modify acls groups to:

23/05/18 06:27:31 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions:

Set(raghavendr48edu); groups with view permissions: Set(); u

sers with modify permissions: Set(raghavendr48edu); groups with modify permissions: Set()

23/05/18 06:27:31 INFO conf.HiveConf: Found configuration file

<file:/etc/hive/conf.cloudera.hive/hive-site.xml>

23/05/18 06:27:31 INFO security.YARNHadoopDelegationTokenManager:
Attempting to load user's ticket cache.
23/05/18 06:27:31 INFO yarn.Client: Submitting application
application_1684298513961_0025 to ResourceManager
23/05/18 06:27:31 INFO impl.YarnClientImpl: Submitted application
application_1684298513961_0025
23/05/18 06:27:32 INFO yarn.Client: Application report for
application_1684298513961_0025 (state: ACCEPTED)
23/05/18 06:27:32 INFO yarn.Client:
client token: N/A
diagnostics: AM container is launched, waiting for AM container to Register
with RM
ApplicationMaster host: N/A
ApplicationMaster RPC port: -1
queue: root.default
start time: 1684391251521
final status: UNDEFINED
tracking URL: http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684298513961_0025/

user: raghavendr48edu
23/05/18 06:27:33 INFO yarn.Client: Application report for
application_1684298513961_0025 (state: ACCEPTED)
23/05/18 06:27:34 INFO yarn.Client: Application report for
application_1684298513961_0025 (state: ACCEPTED)
23/05/18 06:27:35 INFO yarn.Client: Application report for
application_1684298513961_0025 (state: ACCEPTED)
23/05/18 06:27:36 INFO yarn.Client: Application report for
application_1684298513961_0025 (state: ACCEPTED)
23/05/18 06:27:37 INFO yarn.Client: Application report for
application_1684298513961_0025 (state: ACCEPTED)
23/05/18 06:27:38 INFO yarn.Client: Application report for
application_1684298513961_0025 (state: ACCEPTED)
23/05/18 06:27:39 INFO yarn.Client: Application report for
application_1684298513961_0025 (state: ACCEPTED)
23/05/18 06:27:40 INFO yarn.Client: Application report for
application_1684298513961_0025 (state: ACCEPTED)
23/05/18 06:27:41 INFO yarn.Client: Application report for
application_1684298513961_0025 (state: ACCEPTED)
23/05/18 06:27:42 INFO yarn.Client: Application report for
application_1684298513961_0025 (state: ACCEPTED)
23/05/18 06:27:43 INFO yarn.Client: Application report for
application_1684298513961_0025 (state: ACCEPTED)
23/05/18 06:27:44 INFO yarn.Client: Application report for
application_1684298513961_0025 (state: ACCEPTED)
23/05/18 06:27:45 INFO yarn.Client: Application report for
application_1684298513961_0025 (state: ACCEPTED)
23/05/18 06:27:46 INFO yarn.Client: Application report for
application_1684298513961_0025 (state: RUNNING)
23/05/18 06:27:46 INFO yarn.Client:
client token: N/A
diagnostics: N/A

ApplicationMaster host: ip-10-1-2-40.ap-south-1.compute.internal
ApplicationMaster RPC port: 33438
queue: root.default
start time: 1684391251521
final status: UNDEFINED
tracking URL: http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684298513961_0025/

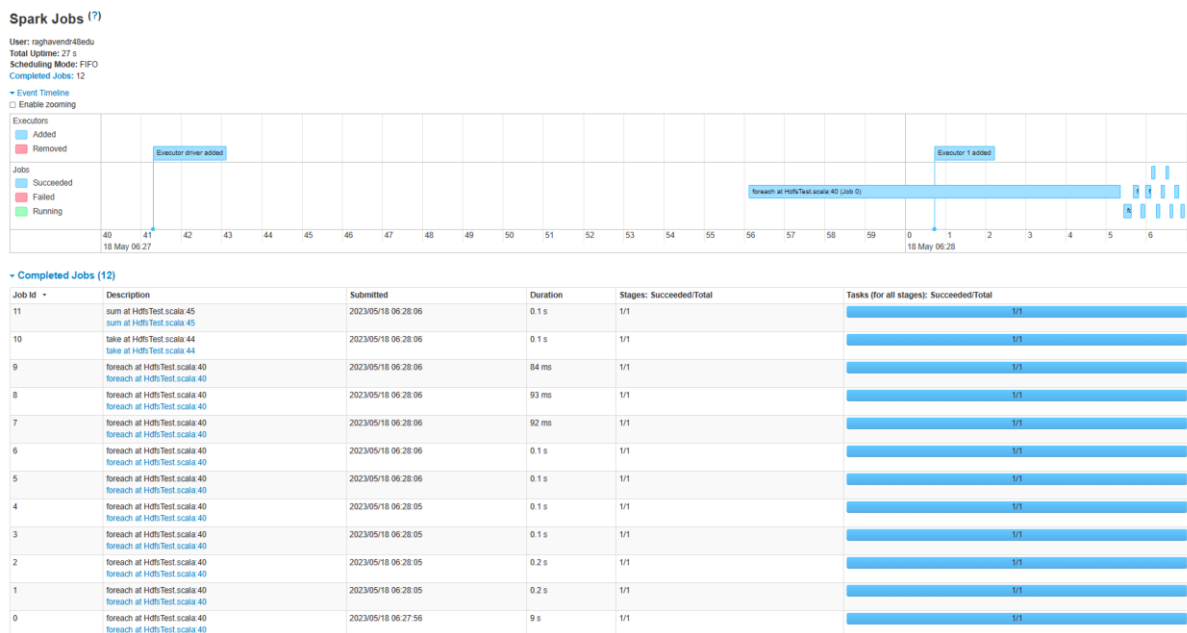
user: raghavendr48edu

23/05/18 06:27:47 INFO yarn.Client: Application report for application_1684298513961_0025 (state: RUNNING)
23/05/18 06:27:48 INFO yarn.Client: Application report for application_1684298513961_0025 (state: RUNNING)
23/05/18 06:27:49 INFO yarn.Client: Application report for application_1684298513961_0025 (state: RUNNING)
23/05/18 06:27:50 INFO yarn.Client: Application report for application_1684298513961_0025 (state: RUNNING)
23/05/18 06:27:51 INFO yarn.Client: Application report for application_1684298513961_0025 (state: RUNNING)
23/05/18 06:27:52 INFO yarn.Client: Application report for application_1684298513961_0025 (state: RUNNING)
23/05/18 06:27:53 INFO yarn.Client: Application report for application_1684298513961_0025 (state: RUNNING)
23/05/18 06:27:54 INFO yarn.Client: Application report for application_1684298513961_0025 (state: RUNNING)
23/05/18 06:27:55 INFO yarn.Client: Application report for application_1684298513961_0025 (state: RUNNING)
23/05/18 06:27:56 INFO yarn.Client: Application report for application_1684298513961_0025 (state: RUNNING)
23/05/18 06:27:57 INFO yarn.Client: Application report for application_1684298513961_0025 (state: RUNNING)
23/05/18 06:27:58 INFO yarn.Client: Application report for application_1684298513961_0025 (state: RUNNING)
23/05/18 06:27:59 INFO yarn.Client: Application report for application_1684298513961_0025 (state: RUNNING)
23/05/18 06:28:00 INFO yarn.Client: Application report for application_1684298513961_0025 (state: RUNNING)
23/05/18 06:28:01 INFO yarn.Client: Application report for application_1684298513961_0025 (state: RUNNING)
23/05/18 06:28:02 INFO yarn.Client: Application report for application_1684298513961_0025 (state: RUNNING)
23/05/18 06:28:03 INFO yarn.Client: Application report for application_1684298513961_0025 (state: RUNNING)
23/05/18 06:28:04 INFO yarn.Client: Application report for application_1684298513961_0025 (state: RUNNING)
23/05/18 06:28:05 INFO yarn.Client: Application report for application_1684298513961_0025 (state: RUNNING)
23/05/18 06:28:06 INFO yarn.Client: Application report for application_1684298513961_0025 (state: RUNNING)
23/05/18 06:28:07 INFO yarn.Client: Application report for application_1684298513961_0025 (state: FINISHED)
23/05/18 06:28:07 INFO yarn.Client:

client token: N/A
 diagnostics: N/A
 ApplicationMaster host: ip-10-1-2-40.ap-south-1.compute.internal
 ApplicationMaster RPC port: 33438
 queue: root.default
 start time: 1684391251521
 final status: SUCCEEDED
 tracking URL: http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684298513961_0025/
 user: raghavendr48edu
 23/05/18 06:28:07 INFO util.ShutdownHookManager: Shutdown hook called
 23/05/18 06:28:07 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-f0f0186d-24ee-4775-b22c-558a17f3b8d0
 23/05/18 06:28:07 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-cfe28f48-65a9-4f7f-8ff8-b6cb5fd22b8f

From <<https://bigdata.cloudlabs.edureka.co/shell/>>

3. Analyze the behavior of spark application on Spark web UI



sort.py

```

from __future__ import print_function
import sys,time
from pyspark.sql import SparkSession
if __name__ == "__main__":
    if len(sys.argv) != 2:
        print("Usage: hdfstest.py <file>", file=sys.stderr)
        exit(-1)
    spark = SparkSession\
        .builder\

```



```

        .appName("HdfsTest_py")\
        .getOrCreate()
file_ = spark.read.text(sys.argv[1]).rdd
mapped = file_.map(lambda s:len(s)).cache()
for i in range(10):
    start_time = time.time()
    mapped.map(lambda x: x+2)
    end_time = time.time()
    print('----> Iteration took:', end_time - start_time, 'ms')
spark.stop()

```

From <http://bigdata.cloudlabs.edureka.co:8888/hue/filebrowser/view=/user/raghavendr48edu/use_cases/sort.py>

```

[raghavendr48edu@ip-10-1-1-204 ~]$ spark-submit --class
org.apache.spark.examples.HdfsTest --master yarn --deploy-mode
cluster /opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/spark/examples/jars/spark-examples*.jar use_cases/sort.py
23/05/18 06:32:25 INFO client.RMProxy: Connecting to ResourceManager at ip-
10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/18 06:32:25 INFO yarn.Client: Requesting a new application from cluster
with 3 NodeManagers
23/05/18 06:32:25 INFO conf.Configuration: resource-types.xml not found
23/05/18 06:32:25 INFO resource.ResourceUtils: Unable to find 'resource-
types.xml'.
23/05/18 06:32:25 INFO yarn.Client: Verifying our application has not requested
more than the maximum memory capability of the cluster (20480 MB per
containe
23/05/18 06:32:25 INFO yarn.Client: Will allocate AM container, with 1408 MB
memory including 384 MB overhead
23/05/18 06:32:25 INFO yarn.Client: Setting up container launch context for
our AM
23/05/18 06:32:25 INFO yarn.Client: Setting up the launch environment for our
AM container
23/05/18 06:32:25 INFO yarn.Client: Preparing resources for our AM container
23/05/18 06:32:25 INFO yarn.Client: Uploading resource
file:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/spark/examples/jars/spark-examples_2.11
application_1684298513961_0026/spark-examples_2.11-2.4.0-cdh6.2.1.jar
23/05/18 06:32:25 INFO yarn.Client: Uploading resource file:/tmp/spark-
254c41ee-8256-452e-ae25-
d9efd412afdd/ spark_conf 2681805809164057652.zip -> hdfs://n
__spark_conf__.zip
23/05/18 06:32:25 INFO spark.SecurityManager: Changing view acls to:
raghavendr48edu
23/05/18 06:32:25 INFO spark.SecurityManager: Changing modify acls to:
raghavendr48edu
23/05/18 06:32:25 INFO spark.SecurityManager: Changing view acls groups to:
23/05/18 06:32:25 INFO spark.SecurityManager: Changing modify acls groups
to:

```

23/05/18 06:32:25 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(raghavendr48edu); avendr48edu); groups with modify permissions: Set()
23/05/18 06:32:25 INFO conf.HiveConf: Found configuration file <file:/etc/hive/conf.cloudera.hive/hive-site.xml>
23/05/18 06:32:26 INFO security.YARNHadoopDelegationTokenManager: Attempting to load user's ticket cache.
23/05/18 06:32:26 INFO yarn.Client: Submitting application application_1684298513961_0026 to ResourceManager
23/05/18 06:32:26 INFO impl.YarnClientImpl: Submitted application application_1684298513961_0026
23/05/18 06:32:27 INFO yarn.Client: Application report for application_1684298513961_0026 (state: ACCEPTED)
23/05/18 06:32:27 INFO yarn.Client:
 client token: N/A
 diagnostics: AM container is launched, waiting for AM container to Register with RM
 ApplicationMaster host: N/A
 ApplicationMaster RPC port: -1
 queue: root.default
 start time: 1684391546066
 final status: UNDEFINED
 tracking URL: http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684298513961_0026/
 user: raghavendr48edu
23/05/18 06:32:28 INFO yarn.Client: Application report for application_1684298513961_0026 (state: ACCEPTED)
23/05/18 06:32:29 INFO yarn.Client: Application report for application_1684298513961_0026 (state: ACCEPTED)
23/05/18 06:32:30 INFO yarn.Client: Application report for application_1684298513961_0026 (state: ACCEPTED)
23/05/18 06:32:31 INFO yarn.Client: Application report for application_1684298513961_0026 (state: RUNNING)
23/05/18 06:32:31 INFO yarn.Client:
 client token: N/A
 diagnostics: N/A
 ApplicationMaster host: ip-10-1-2-103.ap-south-1.compute.internal
 ApplicationMaster RPC port: 38195
 queue: root.default
 start time: 1684391546066
 final status: UNDEFINED
 tracking URL: http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684298513961_0026/
 user: raghavendr48edu
23/05/18 06:32:32 INFO yarn.Client: Application report for application_1684298513961_0026 (state: RUNNING)
23/05/18 06:32:33 INFO yarn.Client: Application report for application_1684298513961_0026 (state: RUNNING)

```

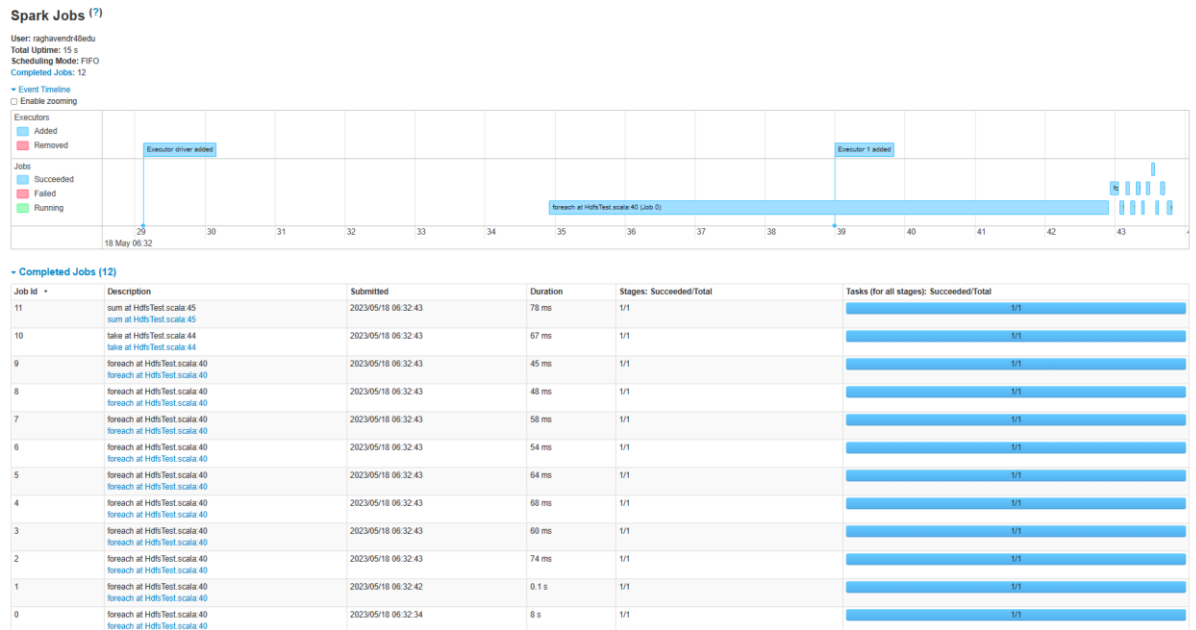
23/05/18 06:32:34 INFO yarn.Client: Application report for
application_1684298513961_0026 (state: RUNNING)
23/05/18 06:32:35 INFO yarn.Client: Application report for
application_1684298513961_0026 (state: RUNNING)
23/05/18 06:32:36 INFO yarn.Client: Application report for
application_1684298513961_0026 (state: RUNNING)
23/05/18 06:32:37 INFO yarn.Client: Application report for
application_1684298513961_0026 (state: RUNNING)
23/05/18 06:32:38 INFO yarn.Client: Application report for
application_1684298513961_0026 (state: RUNNING)
23/05/18 06:32:39 INFO yarn.Client: Application report for
application_1684298513961_0026 (state: RUNNING)
23/05/18 06:32:40 INFO yarn.Client: Application report for
application_1684298513961_0026 (state: RUNNING)
23/05/18 06:32:41 INFO yarn.Client: Application report for
application_1684298513961_0026 (state: RUNNING)
23/05/18 06:32:42 INFO yarn.Client: Application report for
application_1684298513961_0026 (state: RUNNING)
23/05/18 06:32:43 INFO yarn.Client: Application report for
application_1684298513961_0026 (state: RUNNING)
23/05/18 06:32:44 INFO yarn.Client: Application report for
application_1684298513961_0026 (state: FINISHED)
23/05/18 06:32:44 INFO yarn.Client:
    client token: N/A
    diagnostics: N/A
    ApplicationMaster host: ip-10-1-2-103.ap-south-1.compute.internal
    ApplicationMaster RPC port: 38195
    queue: root.default
    start time: 1684391546066
    final status: SUCCEEDED
    tracking URL: http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application\_1684298513961\_0026/

    user: raghavendr48edu
23/05/18 06:32:44 INFO util.ShutdownHookManager: Shutdown hook called
23/05/18 06:32:44 INFO util.ShutdownHookManager: Deleting directory
/tmp/spark-254c41ee-8256-452e-ae25-d9efd412afdd
23/05/18 06:32:44 INFO util.ShutdownHookManager: Deleting directory
/tmp/spark-5991d061-6a93-4f8e-9e6b-5d591f6f0661
[raghavendr48edu@ip-10-1-1-204 ~]$

```

From <<https://bigdata.cloudlabs.edureka.co/shell/>>

3. Analyze the behavior of spark application on Spark web UI



OR

```
[raghavendr48edu@ip-10-1-1-204 training_project3]$ vi sort.py
[raghavendr48edu@ip-10-1-1-204 training_project3]$ spark-submit sort.py
input_sort_py.txt
23/05/25 12:38:29 INFO spark.SparkContext: Running Spark version 2.4.0-cdh6.2.1
23/05/25 12:38:29 INFO logging.DriverLogger: Added a local log appender at:
/tmp/spark-23ddc119-4019-420a-bbaa-28c5a9565eb8/___driver_logs___/driver.log
23/05/25 12:38:29 INFO spark.SparkContext: Submitted application:
HdfsTest_py
23/05/25 12:38:29 INFO spark.SecurityManager: Changing view acls to:
raghavendr48edu
23/05/25 12:38:29 INFO spark.SecurityManager: Changing modify acls to:
raghavendr48edu
23/05/25 12:38:29 INFO spark.SecurityManager: Changing view acls groups to:
23/05/25 12:38:29 INFO spark.SecurityManager: Changing modify acls groups
to:
23/05/25 12:38:29 INFO spark.SecurityManager: SecurityManager:
authentication disabled; ui acls disabled; users  with view permissions:
Set(raghavendr48edu); groups with view permissions: Set(); users  with modify
permissions: Set(ragh
avendr48edu); groups with modify permissions: Set()
23/05/25 12:38:29 INFO util.Utils: Successfully started service 'sparkDriver' on
port 39364.
23/05/25 12:38:29 INFO spark.SparkEnv: Registering MapOutputTracker
23/05/25 12:38:29 INFO spark.SparkEnv: Registering BlockManagerMaster
23/05/25 12:38:29 INFO storage.BlockManagerMasterEndpoint: Using
org.apache.spark.storage.DefaultTopologyMapper for getting topology
information
```

23/05/25 12:38:29 INFO storage.BlockManagerMasterEndpoint:
BlockManagerMasterEndpoint up
23/05/25 12:38:29 INFO storage.DiskBlockManager: Created local directory at
/tmp/blockmgr-8ec9ecdc-ab16-4e14-b50d-a20dc7347e38
23/05/25 12:38:29 INFO memory.MemoryStore: MemoryStore started with
capacity 366.3 MB
23/05/25 12:38:29 INFO spark.SparkEnv: Registering
OutputCommitCoordinator
23/05/25 12:38:30 INFO util.Utils: Using initial executors = 0, max of
spark.dynamicAllocation.initialExecutors, spark.dynamicAllocation.minExecutors
and spark.executor.instances
23/05/25 12:38:30 INFO client.RMProxy: Connecting to ResourceManager at ip-
10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/25 12:38:30 INFO yarn.Client: Requesting a new application from cluster
with 3 NodeManagers
23/05/25 12:38:30 INFO conf.Configuration: resource-types.xml not found
23/05/25 12:38:30 INFO resource.ResourceUtils: Unable to find 'resource-
types.xml'.
23/05/25 12:38:30 INFO yarn.Client: Verifying our application has not requested
more than the maximum memory capability of the cluster (20480 MB per
container)
23/05/25 12:38:30 INFO yarn.Client: Will allocate AM container, with 896 MB
memory including 384 MB overhead
23/05/25 12:38:30 INFO yarn.Client: Setting up container launch context for
our AM
23/05/25 12:38:30 INFO yarn.Client: Setting up the launch environment for our
AM container
23/05/25 12:38:30 INFO yarn.Client: Preparing resources for our AM container
23/05/25 12:38:31 INFO yarn.Client: Uploading resource [file:/tmp/spark-
23ddc119-4019-420a-bbaa-
28c5a9565eb8/ spark_conf 7864336031031315848.zip](file:/tmp/spark-23ddc119-4019-420a-bbaa-28c5a9565eb8/spark_conf_7864336031031315848.zip) ->
hdfs://nameservice1/user/raghavendr48edu/.sparkStaging/application_1684866
872278_1276/
__spark_conf__.zip
23/05/25 12:38:31 INFO spark.SecurityManager: Changing view acls to:
raghavendr48edu
23/05/25 12:38:31 INFO spark.SecurityManager: Changing modify acls to:
raghavendr48edu
23/05/25 12:38:31 INFO spark.SecurityManager: Changing view acls groups to:
23/05/25 12:38:31 INFO spark.SecurityManager: Changing modify acls groups
to:
23/05/25 12:38:31 INFO spark.SecurityManager: SecurityManager:
authentication disabled; ui acls disabled; users with view permissions:
Set(raghavendr48edu); groups with view permissions: Set(); users with modify
permissions: Set(ragh
avendr48edu); groups with modify permissions: Set()
23/05/25 12:38:31 INFO yarn.Client: Submitting application
application_1684866872278_1276 to ResourceManager
23/05/25 12:38:31 INFO impl.YarnClientImpl: Submitted application
application_1684866872278_1276
23/05/25 12:38:32 INFO yarn.Client: Application report for
application_1684866872278_1276 (state: ACCEPTED)

23/05/25 12:38:32 INFO yarn.Client:
client token: N/A
diagnostics: AM container is launched, waiting for AM container to Register with RM
ApplicationMaster host: N/A
ApplicationMaster RPC port: -1
queue: root.default
start time: 1685018311519
final status: UNDEFINED
tracking URL: http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_1276/

user: raghavendr48edu

23/05/25 12:38:33 INFO yarn.Client: Application report for application_1684866872278_1276 (state: ACCEPTED)

23/05/25 12:38:34 INFO yarn.Client: Application report for application_1684866872278_1276 (state: RUNNING)

23/05/25 12:38:34 INFO yarn.Client:
client token: N/A
diagnostics: N/A
ApplicationMaster host: 10.1.2.103
ApplicationMaster RPC port: -1
queue: root.default
start time: 1685018311519
final status: UNDEFINED
tracking URL: http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_1276/

user: raghavendr48edu

23/05/25 12:38:34 INFO cluster.YarnClientSchedulerBackend: Application application_1684866872278_1276 has started running.

23/05/25 12:38:34 INFO cluster.SchedulerExtensionServices: Starting Yarn extension services with app application_1684866872278_1276 and attemptId None

23/05/25 12:38:34 INFO util.Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 37610.

23/05/25 12:38:34 INFO netty.NettyBlockTransferService: Server created on ip-10-1-1-204.ap-south-1.compute.internal:37610

23/05/25 12:38:34 INFO storage.BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy

23/05/25 12:38:34 INFO storage.BlockManagerMaster: Registering BlockManager BlockManagerId(driver, ip-10-1-1-204.ap-south-1.compute.internal, 37610, None)

23/05/25 12:38:34 INFO storage.BlockManagerMasterEndpoint: Registering block manager ip-10-1-1-204.ap-south-1.compute.internal:37610 with 366.3 MB RAM, BlockManagerId(driver, ip-10-1-1-204.ap-south-1.compute.internal, 37610, None)

23/05/25 12:38:34 INFO storage.BlockManagerMaster: Registered BlockManager BlockManagerId(driver, ip-10-1-1-204.ap-south-1.compute.internal, 37610, None)

23/05/25 12:38:34 INFO storage.BlockManager: external shuffle service port = 7337

23/05/25 12:38:34 INFO storage.BlockManager: Initialized BlockManager: BlockManagerId(driver, ip-10-1-1-204.ap-south-1.compute.internal, 37610, None)

23/05/25 12:38:34 INFO cluster.YarnClientSchedulerBackend: Add WebUI Filter. org.apache.hadoop.yarn.server.webproxy.amfilter.AmIpFilter, Map(PROXY_HOST -> ip-10-1-1-204.ap-south-1.compute.internal, PROXY_URI_BASES -> http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_1276), /proxy/application_1684866872278_1276

23/05/25 12:38:34 INFO util.log: Logging initialized @7043ms

23/05/25 12:38:35 INFO scheduler.EventLoggingListener: Logging events to hdfs://nameservice1/user/spark/applicationHistory/application_1684866872278_1276

23/05/25 12:38:35 INFO util.Utils: Using initial executors = 0, max of spark.dynamicAllocation.initialExecutors, spark.dynamicAllocation.minExecutors and spark.executor.instances

23/05/25 12:38:35 WARN cluster.YarnSchedulerBackend\$YarnSchedulerEndpoint: Attempted to request executors before the AM has registered!

23/05/25 12:38:35 INFO spark.SparkContext: Registered listener com.cloudera.spark.lineage.NavigatorAppListener

23/05/25 12:38:35 INFO logging.DriverLogger\$DfsAsyncWriter: Started driver log file sync to: /user/spark/driverLogs/application_1684866872278_1276_driver.log

23/05/25 12:38:35 INFO cluster.YarnClientSchedulerBackend: SchedulerBackend is ready for scheduling beginning after reached minRegisteredResourcesRatio: 0.8

23/05/25 12:38:35 INFO internal.SharedState: loading hive config file: <file:/etc/hive/conf.cloudera.hive/hive-site.xml>

23/05/25 12:38:35 INFO internal.SharedState: spark.sql.warehouse.dir is not set, but hive.metastore.warehouse.dir is set. Setting spark.sql.warehouse.dir to the value of hive.metastore.warehouse.dir ('/user/hive/warehouse').

23/05/25 12:38:35 INFO internal.SharedState: Warehouse path is '/user/hive/warehouse'.

23/05/25 12:38:35 INFO cluster.YarnSchedulerBackend\$YarnSchedulerEndpoint: ApplicationMaster registered as NettyRpcEndpointRef(spark-client://YarnAM)

23/05/25 12:38:35 INFO state.StateStoreCoordinatorRef: Registered StateStoreCoordinator endpoint

23/05/25 12:38:37 INFO datasources.FileSourceStrategy: Pruning directories with:

23/05/25 12:38:37 INFO datasources.FileSourceStrategy: Post-Scan Filters:

23/05/25 12:38:37 INFO datasources.FileSourceStrategy: Output Data Schema: struct<value: string>

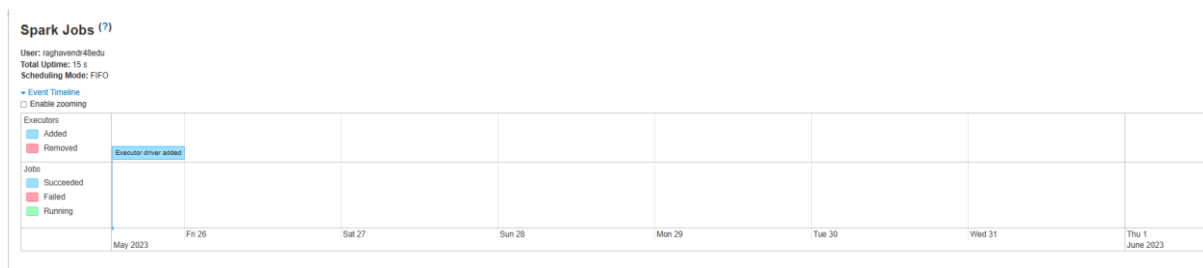
23/05/25 12:38:37 INFO execution.FileSourceScanExec: Pushed Filters:

23/05/25 12:38:37 INFO codegen.CodeGenerator: Code generated in 181.767905 ms

23/05/25 12:38:38 INFO memory.MemoryStore: Block broadcast_0 stored as values in memory (estimated size 337.3 KB, free 366.0 MB)

23/05/25 12:38:38 INFO memory.MemoryStore: Block broadcast_0_piece0 stored as bytes in memory (estimated size 32.3 KB, free 365.9 MB)
23/05/25 12:38:38 INFO storage.BlockManagerInfo: Added broadcast_0_piece0 in memory on ip-10-1-1-204.ap-south-1.compute.internal:37610 (size: 32.3 KB, free: 366.3 MB)
23/05/25 12:38:38 INFO spark.SparkContext: Created broadcast 0 from javaToPython at NativeMethodAccessorImpl.java:0
23/05/25 12:38:38 INFO execution.FileSourceScanExec: Planning scan with bin packing, max size: 4194304 bytes, open cost is considered as scanning 4194304 bytes.
----> **Iteration took: 1.621246337890625e-05 ms**
----> **Iteration took: 6.9141387939453125e-06 ms**
----> **Iteration took: 4.76837158203125e-06 ms**
----> **Iteration took: 4.0531158447265625e-06 ms**
----> **Iteration took: 4.0531158447265625e-06 ms**
----> **Iteration took: 3.814697265625e-06 ms**
----> **Iteration took: 4.76837158203125e-06 ms**
----> **Iteration took: 4.0531158447265625e-06 ms**
----> **Iteration took: 3.5762786865234375e-06 ms**
----> **Iteration took: 3.814697265625e-06 ms**
23/05/25 12:38:38 INFO cluster.YarnClientSchedulerBackend: Interrupting monitor thread
23/05/25 12:38:38 INFO cluster.YarnClientSchedulerBackend: Shutting down all executors
23/05/25 12:38:38 INFO cluster.YarnSchedulerBackend\$YarnDriverEndpoint: Asking each executor to shut down
23/05/25 12:38:38 INFO cluster.SchedulerExtensionServices: Stopping SchedulerExtensionServices
(serviceOption=None,
services=List(),
started=false)
23/05/25 12:38:38 INFO cluster.YarnClientSchedulerBackend: Stopped
23/05/25 12:38:38 INFO spark.MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
23/05/25 12:38:38 INFO memory.MemoryStore: MemoryStore cleared
23/05/25 12:38:38 INFO storage.BlockManager: BlockManager stopped
23/05/25 12:38:38 INFO storage.BlockManagerMaster: BlockManagerMaster stopped
23/05/25 12:38:38 INFO scheduler.OutputCommitCoordinator\$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
23/05/25 12:38:38 INFO spark.SparkContext: Successfully stopped SparkContext
23/05/25 12:38:38 INFO util.ShutdownHookManager: Shutdown hook called
23/05/25 12:38:38 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-23ddc119-4019-420a-bbaa-28c5a9565eb8/pyspark-a0d714ae-d1a1-4947-88c4-2d4078b982b5
23/05/25 12:38:38 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-23ddc119-4019-420a-bbaa-28c5a9565eb8
23/05/25 12:38:38 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-f378884c-0204-440a-9c08-54c4e52c9c92
[raghavendr48edu@ip-10-1-1-204 training_project3]\$

From <<https://bigdata.cloudlabs.edureka.co/shell/>>



4. Edit the application and add custom logs. Once executed check the Spark logs.

a. Edit the scala file to add custom logger

```
vi SparkSort.scala import org.apache.spark.sql.SparkSession
object HdfsTest
{ /** Usage: HdfsTest [file] */
  def main(args: Array[String])
  {
    if (args.length < 1)
    {
      System.err.println("Usage: HdfsTest ") System.exit(1)
    }
    val spark=SparkSession.builder.appName("Custom
HdfsTest").getOrCreate()
    println("Hello from Raghav")
    val file = spark.read.text(args(0)).rdd
    val mapped = file.map(s => s.length).cache()
    for (iter <- 1 to 10)
    {
      val start = System.currentTimeMillis()
      for (x <- mapped) { x + 2 }
      val end = System.currentTimeMillis()
      println("-->Iteration " + iter + " took " + (end-start) + "
ms")
    }
  }
}
```

```

    }
    spark.stop()
  }
}

```

```

[raghavendr48edu@ip-10-1-1-204 training_project3]$ cd
src/main/scala/com/edureka/training/m4
[raghavendr48edu@ip-10-1-1-204 m4]$ vi SparkSort.scala
[raghavendr48edu@ip-10-1-1-204 m4]$ cd
/home/raghavendr48edu/training_project3

```

b. Compile and package the updated scala code sbt compile sbt package c. Re-submit the spark job with the scala code spark-submit --class com.edureka.training.m4.SparkSort --deploy-mode client target/scala-2.11/sparkme-project_2.11-1.0.jar input_sort_py.txt

```

[raghavendr48edu@ip-10-1-1-204 training_project3]$ sbt compile
[info] welcome to sbt 1.5.5 (Oracle Corporation Java 1.8.0_181)
[info] loading project definition from
/home/raghavendr48edu/training_project3/project
[info] loading settings for project training_project3 from build.sbt ...
[info] set current project to SparkMe Project (in build
file:/home/raghavendr48edu/training_project3/)
[info] Executing in batch mode. For better performance use sbt's shell
[info] compiling 2 Scala sources to
/home/raghavendr48edu/training_project3/target/scala-2.11/classes ...
[success] Total time: 5 s, completed May 25, 2023 1:05:35 PM
[raghavendr48edu@ip-10-1-1-204 training_project3]$ sbt package
[info] welcome to sbt 1.5.5 (Oracle Corporation Java 1.8.0_181)
[info] loading project definition from
/home/raghavendr48edu/training_project3/project
[info] loading settings for project training_project3 from build.sbt ...
[info] set current project to SparkMe Project (in build
file:/home/raghavendr48edu/training_project3/)
[success] Total time: 1 s, completed May 25, 2023 1:05:47 PM
[raghavendr48edu@ip-10-1-1-204 training_project3]$ spark-submit --class
HdfsTest --deploy-mode client target/scala-2.11/sparkme-project_2.11-
1.0.jar input_sort_py.txt
23/05/25 13:06:15 INFO spark.SparkContext: Running Spark version 2.4.0-
cdh6.2.1
23/05/25 13:06:15 INFO logging.DriverLogger: Added a local log appender at:
/tmp/spark-f12e6dab-06b6-44ff-af99-981e5c09d7a8/___driver_logs___/driver.log
23/05/25 13:06:15 INFO spark.SparkContext: Submitted application:
CustomHdfsTest

```

23/05/25 13:06:15 INFO spark.SecurityManager: Changing view acls to: raghavendr48edu
23/05/25 13:06:15 INFO spark.SecurityManager: Changing modify acls to: raghavendr48edu
23/05/25 13:06:15 INFO spark.SecurityManager: Changing view acls groups to:
23/05/25 13:06:15 INFO spark.SecurityManager: Changing modify acls groups to:
23/05/25 13:06:15 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(raghavendr48edu); groups with view permissions: Set(); users with modify permissions: Set(raghavendr48edu); groups with modify permissions: Set()
23/05/25 13:06:15 INFO util.Utils: Successfully started service 'sparkDriver' on port 34843.
23/05/25 13:06:15 INFO spark.SparkEnv: Registering MapOutputTracker
23/05/25 13:06:15 INFO spark.SparkEnv: Registering BlockManagerMaster
23/05/25 13:06:15 INFO storage.BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information
23/05/25 13:06:15 INFO storage.BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up
23/05/25 13:06:15 INFO storage.DiskBlockManager: Created local directory at /tmp/blockmgr-04a5a7dc-4c57-4a7b-b095-f5f7ec6764ac
23/05/25 13:06:15 INFO memory.MemoryStore: MemoryStore started with capacity 366.3 MB
23/05/25 13:06:15 INFO spark.SparkEnv: Registering OutputCommitCoordinator
23/05/25 13:06:15 INFO spark.SparkContext: Added JAR file:/home/raghavendr48edu/training_project3/target/scala-2.11/sparkme-project_2.11-1.0.jar at spark://ip-10-1-1-204.ap-south-1.compute.internal:34843/jars/sparkme-project_2.11-1.0.jar with timestamp 1685019975886
23/05/25 13:06:15 INFO util.Utils: Using initial executors = 0, max of spark.dynamicAllocation.initialExecutors, spark.dynamicAllocation.minExecutors and spark.executor.instances
23/05/25 13:06:16 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/25 13:06:16 INFO yarn.Client: Requesting a new application from cluster with 3 NodeManagers
23/05/25 13:06:16 INFO conf.Configuration: resource-types.xml not found
23/05/25 13:06:16 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
23/05/25 13:06:16 INFO yarn.Client: Verifying our application has not requested more than the maximum memory capability of the cluster (20480 MB per container)
23/05/25 13:06:16 INFO yarn.Client: Will allocate AM container, with 896 MB memory including 384 MB overhead
23/05/25 13:06:16 INFO yarn.Client: Setting up container launch context for our AM
23/05/25 13:06:16 INFO yarn.Client: Setting up the launch environment for our AM container
23/05/25 13:06:16 INFO yarn.Client: Preparing resources for our AM container

23/05/25 13:06:17 INFO yarn.Client: Uploading resource file:/tmp/spark-f12e6dab-06b6-44ff-af99-981e5c09d7a8/spark_conf_4661786933308266813.zip ->
hdfs://nameservice1/user/raghavendr48edu/.sparkStaging/application_1684866872278_1287/
__spark_conf__.zip
23/05/25 13:06:17 INFO spark.SecurityManager: Changing view acls to:
raghavendr48edu
23/05/25 13:06:17 INFO spark.SecurityManager: Changing modify acls to:
raghavendr48edu
23/05/25 13:06:17 INFO spark.SecurityManager: Changing view acls groups to:
23/05/25 13:06:17 INFO spark.SecurityManager: Changing modify acls groups
to:
23/05/25 13:06:17 INFO spark.SecurityManager: SecurityManager:
authentication disabled; ui acls disabled; users with view permissions:
Set(raghavendr48edu); groups with view permissions: Set(); users with modify
permissions: Set(ragh
avendr48edu); groups with modify permissions: Set()
23/05/25 13:06:17 INFO yarn.Client: Submitting application
application_1684866872278_1287 to ResourceManager
23/05/25 13:06:17 INFO impl.YarnClientImpl: Submitted application
application_1684866872278_1287
23/05/25 13:06:18 INFO yarn.Client: Application report for
application_1684866872278_1287 (state: ACCEPTED)
23/05/25 13:06:18 INFO yarn.Client:
client token: N/A
diagnostics: AM container is launched, waiting for AM container to Register
with RM
ApplicationMaster host: N/A
ApplicationMaster RPC port: -1
queue: root.default
start time: 1685019977354
final status: UNDEFINED
tracking URL: [http://ip-10-1-1-204.ap-south-
1.compute.internal:6066/proxy/application_1684866872278_1287/](http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_1287/)

user: raghavendr48edu
23/05/25 13:06:19 INFO yarn.Client: Application report for
application_1684866872278_1287 (state: ACCEPTED)
23/05/25 13:06:20 INFO yarn.Client: Application report for
application_1684866872278_1287 (state: ACCEPTED)
23/05/25 13:06:21 INFO yarn.Client: Application report for
application_1684866872278_1287 (state: ACCEPTED)
23/05/25 13:06:22 INFO yarn.Client: Application report for
application_1684866872278_1287 (state: RUNNING)
23/05/25 13:06:22 INFO yarn.Client:
client token: N/A
diagnostics: N/A
ApplicationMaster host: 10.1.2.175
ApplicationMaster RPC port: -1
queue: root.default
start time: 1685019977354

final status: UNDEFINED

tracking URL: http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_1287/

user: raghavendr48edu

23/05/25 13:06:22 INFO cluster.YarnClientSchedulerBackend: Application application_1684866872278_1287 has started running.

23/05/25 13:06:22 INFO cluster.SchedulerExtensionServices: Starting Yarn extension services with app application_1684866872278_1287 and attemptId None

23/05/25 13:06:22 INFO util.Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 42947.

23/05/25 13:06:22 INFO netty.NettyBlockTransferService: Server created on ip-10-1-1-204.ap-south-1.compute.internal:42947

23/05/25 13:06:22 INFO storage.BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy

23/05/25 13:06:22 INFO storage.BlockManagerMaster: Registering BlockManager BlockManagerId(driver, ip-10-1-1-204.ap-south-1.compute.internal, 42947, None)

23/05/25 13:06:22 INFO storage.BlockManagerMasterEndpoint: Registering block manager ip-10-1-1-204.ap-south-1.compute.internal:42947 with 366.3 MB RAM, BlockManagerId(driver, ip-10-1-1-204.ap-south-1.compute.internal, 42947, None)

23/05/25 13:06:22 INFO storage.BlockManagerMaster: Registered BlockManager BlockManagerId(driver, ip-10-1-1-204.ap-south-1.compute.internal, 42947, None)

23/05/25 13:06:22 INFO storage.BlockManager: external shuffle service port = 7337

23/05/25 13:06:22 INFO storage.BlockManager: Initialized BlockManager: BlockManagerId(driver, ip-10-1-1-204.ap-south-1.compute.internal, 42947, None)

23/05/25 13:06:22 INFO cluster.YarnClientSchedulerBackend: Add WebUI Filter. org.apache.hadoop.yarn.server.webproxy.amfilter.AmIpFilter, Map(PROXY_HOS TS -> ip-10-1-1-204.ap-south-1.compute.internal, PROXY_URI_BASES -

> [http://ip-10-1-1-204.ap-south-](http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_1287/)

[1.compute.internal:6066/proxy/application_1684866872278_1287\),](http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_1287/)
[/proxy/application_1684866872278_1287](http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_1287/)

23/05/25 13:06:22 INFO util.log: Logging initialized @8510ms

23/05/25 13:06:22 INFO scheduler.EventLoggingListener: Logging events to hdfs://nameservice1/user/spark/applicationHistory/application_1684866872278_1287

23/05/25 13:06:22 INFO util.Utils: Using initial executors = 0, max of spark.dynamicAllocation.initialExecutors, spark.dynamicAllocation.minExecutors and spark.executor.instances

23/05/25 13:06:22 WARN

cluster.YarnSchedulerBackend\$YarnSchedulerEndpoint: Attempted to request executors before the AM has registered!

23/05/25 13:06:22 INFO spark.SparkContext: Registered listener com.cloudera.spark.lineage.NavigatorAppListener

23/05/25 13:06:22 INFO logging.DriverLogger\$DfsAsyncWriter: Started driver log file sync to:
/user/spark/driverLogs/application_1684866872278_1287_driver.log
23/05/25 13:06:22 INFO cluster.YarnClientSchedulerBackend:
SchedulerBackend is ready for scheduling beginning after reached
minRegisteredResourcesRatio: 0.8

Hello From Raghav

23/05/25 13:06:23 INFO internal.SharedState: loading hive config file:
<file:/etc/hive/conf.cloudera.hive/hive-site.xml>
23/05/25 13:06:23 INFO internal.SharedState: spark.sql.warehouse.dir is not set, but hive.metastore.warehouse.dir is set. Setting spark.sql.warehouse.dir to the value of hive.metastore.warehouse.dir ('/user/hive/warehouse').
23/05/25 13:06:23 INFO internal.SharedState: Warehouse path is '/user/hive/warehouse'.
23/05/25 13:06:23 INFO state.StateStoreCoordinatorRef: Registered StateStoreCoordinator endpoint
23/05/25 13:06:23 INFO
cluster.YarnSchedulerBackend\$YarnSchedulerEndpoint: ApplicationMaster registered as NettyRpcEndpointRef(spark-client://YarnAM)
23/05/25 13:06:25 INFO datasources.FileSourceStrategy: Pruning directories with:
23/05/25 13:06:25 INFO datasources.FileSourceStrategy: Post-Scan Filters:
23/05/25 13:06:25 INFO datasources.FileSourceStrategy: Output Data Schema: struct<value: string>
23/05/25 13:06:25 INFO execution.FileSourceScanExec: Pushed Filters:
23/05/25 13:06:26 INFO codegen.CodeGenerator: Code generated in 193.529709 ms
23/05/25 13:06:26 INFO memory.MemoryStore: Block broadcast_0 stored as values in memory (estimated size 336.4 KB, free 366.0 MB)
23/05/25 13:06:26 INFO memory.MemoryStore: Block broadcast_0_piece0 stored as bytes in memory (estimated size 32.2 KB, free 365.9 MB)
23/05/25 13:06:26 INFO storage.BlockManagerInfo: Added broadcast_0_piece0 in memory on ip-10-1-1-204.ap-south-1.compute.internal:42947 (size: 32.2 KB, free: 366.3 MB)
23/05/25 13:06:26 INFO spark.SparkContext: Created broadcast 0 from rdd at SparkSort.scala:23
23/05/25 13:06:26 INFO execution.FileSourceScanExec: Planning scan with bin packing, max size: 4194304 bytes, open cost is considered as scanning 4194304 bytes.
23/05/25 13:06:26 INFO spark.SparkContext: Starting job: foreach at SparkSort.scala:33
23/05/25 13:06:26 INFO scheduler.DAGScheduler: Got job 0 (foreach at SparkSort.scala:33) with 1 output partitions
23/05/25 13:06:26 INFO scheduler.DAGScheduler: Final stage: ResultStage 0 (foreach at SparkSort.scala:33)
23/05/25 13:06:26 INFO scheduler.DAGScheduler: Parents of final stage: List()
23/05/25 13:06:26 INFO scheduler.DAGScheduler: Missing parents: List()
23/05/25 13:06:26 INFO scheduler.DAGScheduler: Submitting ResultStage 0 (MapPartitionsRDD[4] at map at SparkSort.scala:25), which has no missing parents

23/05/25 13:06:26 INFO memory.MemoryStore: Block broadcast_1 stored as values in memory (estimated size 10.6 KB, free 365.9 MB)
23/05/25 13:06:26 INFO memory.MemoryStore: Block broadcast_1_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.9 MB)
23/05/25 13:06:26 INFO storage.BlockManagerInfo: Added broadcast_1_piece0 in memory on ip-10-1-1-204.ap-south-1.compute.internal:42947 (size: 5.6 KB, free: 366.3 MB)
23/05/25 13:06:26 INFO spark.SparkContext: Created broadcast 1 from broadcast at DAGScheduler.scala:1164
23/05/25 13:06:26 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 0 (MapPartitionsRDD[4] at map at SparkSort.scala:25) (first 15 tasks are for partitions Vector(0))
23/05/25 13:06:26 INFO cluster.YarnScheduler: Adding task set 0.0 with 1 tasks
23/05/25 13:06:26 INFO spark.ExecutorAllocationManager: Requesting 1 new executor because tasks are backlogged (new desired total will be 1)
23/05/25 13:06:38 INFO cluster.YarnSchedulerBackend\$YarnDriverEndpoint: Registered executor NettyRpcEndpointRef(spark-client://Executor) (10.1.2.40:51290) with ID 1
23/05/25 13:06:38 INFO spark.ExecutorAllocationManager: New executor 1 has registered (new total is 1)
23/05/25 13:06:38 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 0.0 (TID 0, ip-10-1-2-40.ap-south-1.compute.internal, executor 1, partition 0, NODE_LOCAL, 8347 bytes)
23/05/25 13:06:38 INFO storage.BlockManagerMasterEndpoint: Registering block manager ip-10-1-2-40.ap-south-1.compute.internal:41263 with 366.3 MB RAM, BlockManagerId(1, ip-10-1-2-40.ap-south-1.compute.internal, 41263, None)
23/05/25 13:06:39 INFO storage.BlockManagerInfo: Added broadcast_1_piece0 in memory on ip-10-1-2-40.ap-south-1.compute.internal:41263 (size: 5.6 KB, free: 366.3 MB)
23/05/25 13:06:48 INFO storage.BlockManagerInfo: Added broadcast_0_piece0 in memory on ip-10-1-2-40.ap-south-1.compute.internal:41263 (size: 32.2 KB, free: 366.3 MB)
23/05/25 13:06:51 INFO storage.BlockManagerInfo: Added rdd_4_0 in memory on ip-10-1-2-40.ap-south-1.compute.internal:41263 (size: 32.0 B, free: 366.3 MB)
23/05/25 13:06:51 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 0.0 (TID 0) in 13505 ms on ip-10-1-2-40.ap-south-1.compute.internal (executor 1) (1/1)
23/05/25 13:06:51 INFO cluster.YarnScheduler: Removed TaskSet 0.0, whose tasks have all completed, from pool
23/05/25 13:06:51 INFO scheduler.DAGScheduler: ResultStage 0 (foreach at SparkSort.scala:33) finished in 25.105 s
23/05/25 13:06:51 INFO scheduler.DAGScheduler: Job 0 finished: foreach at SparkSort.scala:33, took 25.158750 s
-->Iteration 1 took 25171 ms
23/05/25 13:06:51 INFO spark.SparkContext: Starting job: foreach at SparkSort.scala:33
23/05/25 13:06:51 INFO scheduler.DAGScheduler: Got job 1 (foreach at SparkSort.scala:33) with 1 output partitions

23/05/25 13:06:51 INFO scheduler.DAGScheduler: Final stage: ResultStage 1 (foreach at SparkSort.scala:33)
23/05/25 13:06:51 INFO scheduler.DAGScheduler: Parents of final stage: List()
23/05/25 13:06:51 INFO scheduler.DAGScheduler: Missing parents: List()
23/05/25 13:06:51 INFO scheduler.DAGScheduler: Submitting ResultStage 1 (MapPartitionsRDD[4] at map at SparkSort.scala:25), which has no missing parents
23/05/25 13:06:51 INFO memory.MemoryStore: Block broadcast_2 stored as values in memory (estimated size 10.6 KB, free 365.9 MB)
23/05/25 13:06:51 INFO memory.MemoryStore: Block broadcast_2_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.9 MB)
23/05/25 13:06:51 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on ip-10-1-1-204.ap-south-1.compute.internal:42947 (size: 5.6 KB, free: 366.3 MB)
23/05/25 13:06:51 INFO spark.SparkContext: Created broadcast 2 from broadcast at DAGScheduler.scala:1164
23/05/25 13:06:51 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 1 (MapPartitionsRDD[4] at map at SparkSort.scala:25) (first 15 tasks are for partitions Vector(0))
23/05/25 13:06:51 INFO cluster.YarnScheduler: Adding task set 1.0 with 1 tasks
23/05/25 13:06:51 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 1.0 (TID 1, ip-10-1-2-40.ap-south-1.compute.internal, executor 1, partition 0, PROCESS_LOCAL, 8347 bytes)
23/05/25 13:06:51 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on ip-10-1-2-40.ap-south-1.compute.internal:41263 (size: 5.6 KB, free: 366.3 MB)
23/05/25 13:06:51 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 1.0 (TID 1) in 263 ms on ip-10-1-2-40.ap-south-1.compute.internal (executor 1) (1/1)
23/05/25 13:06:51 INFO cluster.YarnScheduler: Removed TaskSet 1.0, whose tasks have all completed, from pool
23/05/25 13:06:51 INFO scheduler.DAGScheduler: ResultStage 1 (foreach at SparkSort.scala:33) finished in 0.294 s
23/05/25 13:06:51 INFO scheduler.DAGScheduler: Job 1 finished: foreach at SparkSort.scala:33, took 0.302399 s
-->Iteration 2 took 307 ms
23/05/25 13:06:51 INFO spark.SparkContext: Starting job: foreach at SparkSort.scala:33
23/05/25 13:06:51 INFO scheduler.DAGScheduler: Got job 2 (foreach at SparkSort.scala:33) with 1 output partitions
23/05/25 13:06:51 INFO scheduler.DAGScheduler: Final stage: ResultStage 2 (foreach at SparkSort.scala:33)
23/05/25 13:06:51 INFO scheduler.DAGScheduler: Parents of final stage: List()
23/05/25 13:06:51 INFO scheduler.DAGScheduler: Missing parents: List()
23/05/25 13:06:51 INFO scheduler.DAGScheduler: Submitting ResultStage 2 (MapPartitionsRDD[4] at map at SparkSort.scala:25), which has no missing parents
23/05/25 13:06:51 INFO memory.MemoryStore: Block broadcast_3 stored as values in memory (estimated size 10.6 KB, free 365.9 MB)
23/05/25 13:06:51 INFO memory.MemoryStore: Block broadcast_3_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.9 MB)

23/05/25 13:06:51 INFO storage.BlockManagerInfo: Added broadcast_3_piece0 in memory on ip-10-1-1-204.ap-south-1.compute.internal:42947 (size: 5.6 KB, free: 366.3 MB)

23/05/25 13:06:51 INFO spark.SparkContext: Created broadcast 3 from broadcast at DAGScheduler.scala:1164

23/05/25 13:06:51 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 2 (MapPartitionsRDD[4] at map at SparkSort.scala:25) (first 15 tasks are for partitions Vector(0))

23/05/25 13:06:51 INFO cluster.YarnScheduler: Adding task set 2.0 with 1 tasks

23/05/25 13:06:51 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 2.0 (TID 2, ip-10-1-2-40.ap-south-1.compute.internal, executor 1, partition 0, PROCESS_LOCAL, 8347 bytes)

23/05/25 13:06:52 INFO storage.BlockManagerInfo: Added broadcast_3_piece0 in memory on ip-10-1-2-40.ap-south-1.compute.internal:41263 (size: 5.6 KB, free: 366.3 MB)

23/05/25 13:06:52 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 2.0 (TID 2) in 116 ms on ip-10-1-2-40.ap-south-1.compute.internal (executor 1) (1/1)

23/05/25 13:06:52 INFO cluster.YarnScheduler: Removed TaskSet 2.0, whose tasks have all completed, from pool

23/05/25 13:06:52 INFO scheduler.DAGScheduler: ResultStage 2 (foreach at SparkSort.scala:33) finished in 0.131 s

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Job 2 finished: foreach at SparkSort.scala:33, took 0.138163 s

-->Iteration 3 took 144 ms

23/05/25 13:06:52 INFO spark.SparkContext: Starting job: foreach at SparkSort.scala:33

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Got job 3 (foreach at SparkSort.scala:33) with 1 output partitions

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Final stage: ResultStage 3 (foreach at SparkSort.scala:33)

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Parents of final stage: List()

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Missing parents: List()

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Submitting ResultStage 3 (MapPartitionsRDD[4] at map at SparkSort.scala:25), which has no missing parents

23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_4 stored as values in memory (estimated size 10.6 KB, free 365.9 MB)

23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_4_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.9 MB)

23/05/25 13:06:52 INFO storage.BlockManagerInfo: Added broadcast_4_piece0 in memory on ip-10-1-1-204.ap-south-1.compute.internal:42947 (size: 5.6 KB, free: 366.2 MB)

23/05/25 13:06:52 INFO spark.SparkContext: Created broadcast 4 from broadcast at DAGScheduler.scala:1164

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 3 (MapPartitionsRDD[4] at map at SparkSort.scala:25) (first 15 tasks are for partitions Vector(0))

23/05/25 13:06:52 INFO cluster.YarnScheduler: Adding task set 3.0 with 1 tasks

23/05/25 13:06:52 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 3.0 (TID 3, ip-10-1-2-40.ap-south-1.compute.internal, executor 1, partition 0, PROCESS_LOCAL, 8347 bytes)

23/05/25 13:06:52 INFO storage.BlockManagerInfo: Added broadcast_4_piece0 in memory on ip-10-1-2-40.ap-south-1.compute.internal:41263 (size: 5.6 KB, free: 366.2 MB)

23/05/25 13:06:52 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 3.0 (TID 3) in 91 ms on ip-10-1-2-40.ap-south-1.compute.internal (executor 1) (1/1)

23/05/25 13:06:52 INFO cluster.YarnScheduler: Removed TaskSet 3.0, whose tasks have all completed, from pool

23/05/25 13:06:52 INFO scheduler.DAGScheduler: ResultStage 3 (foreach at SparkSort.scala:33) finished in 0.106 s

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Job 3 finished: foreach at SparkSort.scala:33, took 0.112156 s

-->Iteration 4 took 118 ms

23/05/25 13:06:52 INFO spark.SparkContext: Starting job: foreach at SparkSort.scala:33

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Got job 4 (foreach at SparkSort.scala:33) with 1 output partitions

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Final stage: ResultStage 4 (foreach at SparkSort.scala:33)

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Parents of final stage: List()

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Missing parents: List()

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Submitting ResultStage 4 (MapPartitionsRDD[4] at map at SparkSort.scala:25), which has no missing parents

23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_5 stored as values in memory (estimated size 10.6 KB, free 365.9 MB)

23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_5_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.9 MB)

23/05/25 13:06:52 INFO storage.BlockManagerInfo: Added broadcast_5_piece0 in memory on ip-10-1-1-204.ap-south-1.compute.internal:42947 (size: 5.6 KB, free: 366.2 MB)

23/05/25 13:06:52 INFO spark.SparkContext: Created broadcast 5 from broadcast at DAGScheduler.scala:1164

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 4 (MapPartitionsRDD[4] at map at SparkSort.scala:25) (first 15 tasks are for partitions Vector(0))

23/05/25 13:06:52 INFO cluster.YarnScheduler: Adding task set 4.0 with 1 tasks

23/05/25 13:06:52 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 4.0 (TID 4, ip-10-1-2-40.ap-south-1.compute.internal, executor 1, partition 0, PROCESS_LOCAL, 8347 bytes)

23/05/25 13:06:52 INFO storage.BlockManagerInfo: Added broadcast_5_piece0 in memory on ip-10-1-2-40.ap-south-1.compute.internal:41263 (size: 5.6 KB, free: 366.2 MB)

23/05/25 13:06:52 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 4.0 (TID 4) in 90 ms on ip-10-1-2-40.ap-south-1.compute.internal (executor 1) (1/1)

23/05/25 13:06:52 INFO cluster.YarnScheduler: Removed TaskSet 4.0, whose tasks have all completed, from pool

23/05/25 13:06:52 INFO scheduler.DAGScheduler: ResultStage 4 (foreach at SparkSort.scala:33) finished in 0.101 s
23/05/25 13:06:52 INFO scheduler.DAGScheduler: Job 4 finished: foreach at SparkSort.scala:33, took 0.105685 s
-->Iteration 5 took 110 ms
23/05/25 13:06:52 INFO spark.SparkContext: Starting job: foreach at SparkSort.scala:33
23/05/25 13:06:52 INFO scheduler.DAGScheduler: Got job 5 (foreach at SparkSort.scala:33) with 1 output partitions
23/05/25 13:06:52 INFO scheduler.DAGScheduler: Final stage: ResultStage 5 (foreach at SparkSort.scala:33)
23/05/25 13:06:52 INFO scheduler.DAGScheduler: Parents of final stage: List()
23/05/25 13:06:52 INFO scheduler.DAGScheduler: Missing parents: List()
23/05/25 13:06:52 INFO scheduler.DAGScheduler: Submitting ResultStage 5 (MapPartitionsRDD[4] at map at SparkSort.scala:25), which has no missing parents
23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_6 stored as values in memory (estimated size 10.6 KB, free 365.9 MB)
23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_6_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.8 MB)
23/05/25 13:06:52 INFO storage.BlockManagerInfo: Added broadcast_6_piece0 in memory on ip-10-1-1-204.ap-south-1.compute.internal:42947 (size: 5.6 KB, free: 366.2 MB)
23/05/25 13:06:52 INFO spark.SparkContext: Created broadcast 6 from broadcast at DAGScheduler.scala:1164
23/05/25 13:06:52 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 5 (MapPartitionsRDD[4] at map at SparkSort.scala:25) (first 15 tasks are for partitions Vector(0))
23/05/25 13:06:52 INFO cluster.YarnScheduler: Adding task set 5.0 with 1 tasks
23/05/25 13:06:52 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 5.0 (TID 5, ip-10-1-2-40.ap-south-1.compute.internal, executor 1, partition 0, PROCESS_LOCAL, 8347 bytes)
23/05/25 13:06:52 INFO storage.BlockManagerInfo: Added broadcast_6_piece0 in memory on ip-10-1-2-40.ap-south-1.compute.internal:41263 (size: 5.6 KB, free: 366.2 MB)
23/05/25 13:06:52 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 5.0 (TID 5) in 94 ms on ip-10-1-2-40.ap-south-1.compute.internal (executor 1) (1/1)
23/05/25 13:06:52 INFO cluster.YarnScheduler: Removed TaskSet 5.0, whose tasks have all completed, from pool
23/05/25 13:06:52 INFO scheduler.DAGScheduler: ResultStage 5 (foreach at SparkSort.scala:33) finished in 0.105 s
23/05/25 13:06:52 INFO scheduler.DAGScheduler: Job 5 finished: foreach at SparkSort.scala:33, took 0.109608 s
-->Iteration 6 took 114 ms
23/05/25 13:06:52 INFO spark.SparkContext: Starting job: foreach at SparkSort.scala:33
23/05/25 13:06:52 INFO scheduler.DAGScheduler: Got job 6 (foreach at SparkSort.scala:33) with 1 output partitions
23/05/25 13:06:52 INFO scheduler.DAGScheduler: Final stage: ResultStage 6 (foreach at SparkSort.scala:33)

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Parents of final stage: List()
23/05/25 13:06:52 INFO scheduler.DAGScheduler: Missing parents: List()
23/05/25 13:06:52 INFO scheduler.DAGScheduler: Submitting ResultStage 6 (MapPartitionsRDD[4] at map at SparkSort.scala:25), which has no missing parents
23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_7 stored as values in memory (estimated size 10.6 KB, free 365.8 MB)
23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_7_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.8 MB)
23/05/25 13:06:52 INFO storage.BlockManagerInfo: Added broadcast_7_piece0 in memory on ip-10-1-1-204.ap-south-1.compute.internal:42947 (size: 5.6 KB, free: 366.2 MB)
23/05/25 13:06:52 INFO spark.SparkContext: Created broadcast 7 from broadcast at DAGScheduler.scala:1164
23/05/25 13:06:52 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 6 (MapPartitionsRDD[4] at map at SparkSort.scala:25) (first 15 tasks are for partitions Vector(0))
23/05/25 13:06:52 INFO cluster.YarnScheduler: Adding task set 6.0 with 1 tasks
23/05/25 13:06:52 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 6.0 (TID 6, ip-10-1-2-40.ap-south-1.compute.internal, executor 1, partition 0, PROCESS_LOCAL, 8347 bytes)
23/05/25 13:06:52 INFO storage.BlockManagerInfo: Added broadcast_7_piece0 in memory on ip-10-1-2-40.ap-south-1.compute.internal:41263 (size: 5.6 KB, free: 366.2 MB)
23/05/25 13:06:52 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 6.0 (TID 6) in 65 ms on ip-10-1-2-40.ap-south-1.compute.internal (executor 1) (1/1)
23/05/25 13:06:52 INFO cluster.YarnScheduler: Removed TaskSet 6.0, whose tasks have all completed, from pool
23/05/25 13:06:52 INFO scheduler.DAGScheduler: ResultStage 6 (foreach at SparkSort.scala:33) finished in 0.077 s
23/05/25 13:06:52 INFO scheduler.DAGScheduler: Job 6 finished: foreach at SparkSort.scala:33, took 0.080482 s
-->Iteration 7 took 85 ms
23/05/25 13:06:52 INFO spark.SparkContext: Starting job: foreach at SparkSort.scala:33
23/05/25 13:06:52 INFO scheduler.DAGScheduler: Got job 7 (foreach at SparkSort.scala:33) with 1 output partitions
23/05/25 13:06:52 INFO scheduler.DAGScheduler: Final stage: ResultStage 7 (foreach at SparkSort.scala:33)
23/05/25 13:06:52 INFO scheduler.DAGScheduler: Parents of final stage: List()
23/05/25 13:06:52 INFO scheduler.DAGScheduler: Missing parents: List()
23/05/25 13:06:52 INFO scheduler.DAGScheduler: Submitting ResultStage 7 (MapPartitionsRDD[4] at map at SparkSort.scala:25), which has no missing parents
23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_8 stored as values in memory (estimated size 10.6 KB, free 365.8 MB)
23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_8_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.8 MB)

23/05/25 13:06:52 INFO storage.BlockManagerInfo: Added broadcast_8_piece0 in memory on ip-10-1-1-204.ap-south-1.compute.internal:42947 (size: 5.6 KB, free: 366.2 MB)

23/05/25 13:06:52 INFO spark.SparkContext: Created broadcast 8 from broadcast at DAGScheduler.scala:1164

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 7 (MapPartitionsRDD[4] at map at SparkSort.scala:25) (first 15 tasks are for partitions Vector(0))

23/05/25 13:06:52 INFO cluster.YarnScheduler: Adding task set 7.0 with 1 tasks

23/05/25 13:06:52 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 7.0 (TID 7, ip-10-1-2-40.ap-south-1.compute.internal, executor 1, partition 0, PROCESS_LOCAL, 8347 bytes)

23/05/25 13:06:52 INFO storage.BlockManagerInfo: Added broadcast_8_piece0 in memory on ip-10-1-2-40.ap-south-1.compute.internal:41263 (size: 5.6 KB, free: 366.2 MB)

23/05/25 13:06:52 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 7.0 (TID 7) in 100 ms on ip-10-1-2-40.ap-south-1.compute.internal (executor 1) (1/1)

23/05/25 13:06:52 INFO cluster.YarnScheduler: Removed TaskSet 7.0, whose tasks have all completed, from pool

23/05/25 13:06:52 INFO scheduler.DAGScheduler: ResultStage 7 (foreach at SparkSort.scala:33) finished in 0.111 s

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Job 7 finished: foreach at SparkSort.scala:33, took 0.114674 s

-->Iteration 8 took 119 ms

23/05/25 13:06:52 INFO spark.SparkContext: Starting job: foreach at SparkSort.scala:33

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Got job 8 (foreach at SparkSort.scala:33) with 1 output partitions

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Final stage: ResultStage 8 (foreach at SparkSort.scala:33)

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Parents of final stage: List()

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Missing parents: List()

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Submitting ResultStage 8 (MapPartitionsRDD[4] at map at SparkSort.scala:25), which has no missing parents

23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_9 stored as values in memory (estimated size 10.6 KB, free 365.8 MB)

23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_9_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.8 MB)

23/05/25 13:06:52 INFO storage.BlockManagerInfo: Added broadcast_9_piece0 in memory on ip-10-1-1-204.ap-south-1.compute.internal:42947 (size: 5.6 KB, free: 366.2 MB)

23/05/25 13:06:52 INFO spark.SparkContext: Created broadcast 9 from broadcast at DAGScheduler.scala:1164

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 8 (MapPartitionsRDD[4] at map at SparkSort.scala:25) (first 15 tasks are for partitions Vector(0))

23/05/25 13:06:52 INFO cluster.YarnScheduler: Adding task set 8.0 with 1 tasks

23/05/25 13:06:52 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 8.0 (TID 8, ip-10-1-2-40.ap-south-1.compute.internal, executor 1, partition 0, PROCESS_LOCAL, 8347 bytes)

23/05/25 13:06:52 INFO storage.BlockManagerInfo: Added broadcast_9_piece0 in memory on ip-10-1-2-40.ap-south-1.compute.internal:41263 (size: 5.6 KB, free: 366.2 MB)

23/05/25 13:06:52 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 8.0 (TID 8) in 70 ms on ip-10-1-2-40.ap-south-1.compute.internal (executor 1) (1/1)

23/05/25 13:06:52 INFO cluster.YarnScheduler: Removed TaskSet 8.0, whose tasks have all completed, from pool

23/05/25 13:06:52 INFO scheduler.DAGScheduler: ResultStage 8 (foreach at SparkSort.scala:33) finished in 0.081 s

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Job 8 finished: foreach at SparkSort.scala:33, took 0.084467 s

-->Iteration 9 took 89 ms

23/05/25 13:06:52 INFO spark.SparkContext: Starting job: foreach at SparkSort.scala:33

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Got job 9 (foreach at SparkSort.scala:33) with 1 output partitions

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Final stage: ResultStage 9 (foreach at SparkSort.scala:33)

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Parents of final stage: List()

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Missing parents: List()

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Submitting ResultStage 9 (MapPartitionsRDD[4] at map at SparkSort.scala:25), which has no missing parents

23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_10 stored as values in memory (estimated size 10.6 KB, free 365.8 MB)

23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_10_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.8 MB)

23/05/25 13:06:52 INFO storage.BlockManagerInfo: Added broadcast_10_piece0 in memory on ip-10-1-1-204.ap-south-1.compute.internal:42947 (size: 5.6 KB, free: 366.2 MB)

23/05/25 13:06:52 INFO spark.SparkContext: Created broadcast 10 from broadcast at DAGScheduler.scala:1164

23/05/25 13:06:52 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 9 (MapPartitionsRDD[4] at map at SparkSort.scala:25) (first 15 tasks are for partitions Vector(0))

23/05/25 13:06:52 INFO cluster.YarnScheduler: Adding task set 9.0 with 1 tasks

23/05/25 13:06:52 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 9.0 (TID 9, ip-10-1-2-40.ap-south-1.compute.internal, executor 1, partition 0, PROCESS_LOCAL, 8347 bytes)

23/05/25 13:06:52 INFO storage.BlockManagerInfo: Added broadcast_10_piece0 in memory on ip-10-1-2-40.ap-south-1.compute.internal:41263 (size: 5.6 KB, free: 366.2 MB)

23/05/25 13:06:52 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 9.0 (TID 9) in 56 ms on ip-10-1-2-40.ap-south-1.compute.internal (executor 1) (1/1)

23/05/25 13:06:52 INFO cluster.YarnScheduler: Removed TaskSet 9.0, whose tasks have all completed, from pool

```

23/05/25 13:06:52 INFO scheduler.DAGScheduler: ResultStage 9 (foreach at
SparkSort.scala:33) finished in 0.069 s
23/05/25 13:06:52 INFO scheduler.DAGScheduler: Job 9 finished: foreach at
SparkSort.scala:33, took 0.074613 s
-->Iteration 10 took 79 ms
23/05/25 13:06:52 INFO cluster.YarnClientSchedulerBackend: Interrupting
monitor thread
23/05/25 13:06:52 INFO cluster.YarnClientSchedulerBackend: Shutting down all
executors
23/05/25 13:06:52 INFO cluster.YarnSchedulerBackend$YarnDriverEndpoint:
Asking each executor to shut down
23/05/25 13:06:52 INFO cluster.SchedulerExtensionServices: Stopping
SchedulerExtensionServices
(serviceOption=None,
services=List(),
started=false)
23/05/25 13:06:52 INFO cluster.YarnClientSchedulerBackend: Stopped
23/05/25 13:06:52 INFO spark.MapOutputTrackerMasterEndpoint:
MapOutputTrackerMasterEndpoint stopped!
23/05/25 13:06:52 INFO memory.MemoryStore: MemoryStore cleared
23/05/25 13:06:52 INFO storage.BlockManager: BlockManager stopped
23/05/25 13:06:52 INFO storage.BlockManagerMaster: BlockManagerMaster
stopped
23/05/25 13:06:52 INFO
scheduler.OutputCommitCoordinator$OutputCommitCoordinatorEndpoint:
OutputCommitCoordinator stopped!
23/05/25 13:06:52 INFO spark.SparkContext: Successfully stopped
SparkContext
23/05/25 13:06:52 INFO util.ShutdownHookManager: Shutdown hook called
23/05/25 13:06:52 INFO util.ShutdownHookManager: Deleting directory
/tmp/spark-8cca7a44-6b6a-4454-93d1-cdc4d2a3f436
23/05/25 13:06:52 INFO util.ShutdownHookManager: Deleting directory
/tmp/spark-f12e6dab-06b6-44ff-af99-981e5c09d7a8
[raghavendr48edu@ip-10-1-1-204 training_project3]$

```

From <<https://bigdata.cloudlabs.edureka.co/shell/>>

If you see Hello From Raghav is logged during execution

```

[raghavendr48edu@ip-10-1-1-204 training_project3]$ yarn logs --
applicationId application_1684866872278_1287
WARNING: YARN_OPTS has been replaced by HADOOP_OPTS. Using value of
YARN_OPTS.
23/05/25 13:16:00 INFO client.RMProxy: Connecting to ResourceManager at ip-
10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032

```

End of LogType:prelaunch.err

Container: container_1684866872278_1287_01_000001 on ip-10-1-2-175.ap-south-1.compute.internal_8041

LogAggregationType: AGGREGATED

=====
=====

LogType:prelaunch.out

LogLastModifiedTime:Thu May 25 13:06:54 +0000 2023

LogLength:70

LogContents:

Setting up env variables

Setting up job resources

Launching container

End of LogType:prelaunch.out

Container: container_1684866872278_1287_01_000001 on ip-10-1-2-175.ap-south-1.compute.internal_8041

LogAggregationType: AGGREGATED

=====
=====

LogType:stderr

LogLastModifiedTime:Thu May 25 13:06:54 +0000 2023

LogLength:34735

LogContents:

23/05/25 13:06:19 INFO util.SignalUtils: Registered signal handler for TERM

23/05/25 13:06:19 INFO util.SignalUtils: Registered signal handler for HUP

23/05/25 13:06:19 INFO util.SignalUtils: Registered signal handler for INT

23/05/25 13:06:20 INFO spark.SecurityManager: Changing view acls to:

yarn,raghavendr48edu

23/05/25 13:06:20 INFO spark.SecurityManager: Changing modify acls to:

yarn,raghavendr48edu

23/05/25 13:06:20 INFO spark.SecurityManager: Changing view acls groups to:

23/05/25 13:06:20 INFO spark.SecurityManager: Changing modify acls groups

to:

23/05/25 13:06:20 INFO spark.SecurityManager: SecurityManager:

authentication disabled; ui acls disabled; users with view permissions: Set(yarn,

raghavendr48edu); groups with view permissions: Set(); users with modify

permissions: Se

t(yarn, raghavendr48edu); groups with modify permissions: Set()

23/05/25 13:06:21 INFO yarn.ApplicationMaster: ApplicationAttemptId:

appattempt_1684866872278_1287_000001

23/05/25 13:06:21 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8030

23/05/25 13:06:21 INFO yarn.YarnRMClient: Registering the ApplicationMaster

23/05/25 13:06:22 INFO client.TransportClientFactory: Successfully created

connection to ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:34843
after 162 ms (0 ms spent in bootstraps)

23/05/25 13:06:22 INFO yarn.ApplicationMaster: Preparing Local resources

23/05/25 13:06:23 INFO yarn.ApplicationMaster:

=====
=====

YARN executor launch context:

env:

CLASSPATH ->

```
{{PWD}}<CPS>{{PWD}}/__spark_conf__<CPS>{{PWD}}/__spark_libs__/*<
CPS>{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/spark/jars/*<CPS>{{HADOOP_COMMON_HOME}}/..
/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/
ib/spark/hive/*<CPS>$HADOOP_CLIENT_CONF_DIR<CPS>$HADOOP_COMMON
_HOME/*<CPS>$HADOOP_COMMON_HOME/lib/*<CPS>$HADOOP_HDFS_HOME
/*<CPS>$HADOOP_HDFS_HOME/lib/*<CPS>$HADOOP_YARN_HOME/*<CPS>$
HADOOP_YARN_HOME/lib/*<CPS>$HADOOP_CLIENT_CONF_DIR<CPS
>$PWD/mr-
framework/*<CPS>$MR2_CLASSPATH<CPS>{{HADOOP_COMMON_HOME}}/..
../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/accessors-smart-
1.2.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/acc
essors-smart.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/asm-
5.0.4.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/asm.jar:{{HADOOP_COMMON_HOME}}
/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/avro.jar:{{HADOOP_COMMON_HOME}}
/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/aws-java-sdk-
bundle-1.11.271.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.14257
74/lib/hadoop/client/aws-java-sdk-
bundle.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/azure-data-lake-store-sdk-
2.2.9.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hado
op/client/azure-data-lake-store-
sdk.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-beanutils-
1.9.3.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/com
mons-beanutils.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-cli-
1.2.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-
cli.jar:{{HADOOP_COMMON
HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-
codec-1.11.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-
codec.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cd
h6.2.1.p0.1425774/lib/hadoop/client/commons-collections-
3.2.2.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-
```

collections.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774
/lib/hadoop/client/commons-compress-1.18.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-compress.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-configuration2-2.1.1.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-configuration2.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-io-2.6.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-io.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-lang-2.6.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-lang.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-lang3-3.7.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-lang3.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-logging-1.1.3.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-logging.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-math3-3.1.1.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-math3.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-net-3.1.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-net.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/curator-client-2.12.0.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/curator-client.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/curator-framework-2.12.0.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/curator-framework.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/curator-recipes-2.12.0.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/curator-recipes.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/gson-2.2.4.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/gson.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/guava-

11.0.2.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/had
oop/client/guava.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-annotations-3.0.0-
cdh6.2.1.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-annot
ations.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-auth-3.0.0-
cdh6.2.1.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-auth.jar:{{HADOOP_COMM
ON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-aws-3.0.0-
cdh6.2.1.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-
aws.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2
.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-azure-3.0.0-
cdh6.2.1.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-azure-datalake-3.0.0-
cdh6.2.1.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6
.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-azure-
datalake.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-
azure.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/l
ib/hadoop/client/hadoop-common-3.0.0-
cdh6.2.1.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-
common.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/had
oop-hdfs-client-3.0.0-cdh6.2.1.jar:{{HADOOP_COMMON_HOME}}/../../CDH-
6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-hdfs-
client.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-mapred
uce-client-common-3.0.0-
cdh6.2.1.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-mapreduce-client-
common.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client
/hadoop-mapreduce-client-core-3.0.0-
cdh6.2.1.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-mapreduce-client-
core.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/had
oop/client/hadoop-mapreduce-client-jobclient-3.0.0-
cdh6.2.1.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-mapreduce-client-
jobclient.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.
1.p0.1425774/lib/hadoop/client/hadoop-yarn-api-3.0.0-
cdh6.2.1.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-yarn-
api.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib

```

/hadoop/client/hadoop-yarn-client-3.0.0-
cdh6.2.1.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-yarn-
client.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/cl
ient/hadoop-yarn-common-3.0.0-
cdh6.2.1.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-yarn-
common.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/htrac
e-core4-4.1.0-incubating.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/htrace-
core4.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/httpclient-4.5.3.jar:{{HA
DOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/httpclient.jar:{{HADOOP_COMMON_H
OME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/httpcore-
4.4.6.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1
-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/httpcore.jar:{{HADOOP_COMMON_HO
ME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/jaxb-api-
2.2.11.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/clie
nt/jaxb-api.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/jcip-annotations-1.0-
1.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/jcip-annotations.jar:{{HADOO
P_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/json-smart-
2.3.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/json-
smart.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.
cdh6.2.1.p0.1425774/lib/hadoop/client/jsp-api-
2.1.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/jsp-
api.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/jsr30
5-3.0.0.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/jsr305.jar:{{HADOOP_COMMON_HOM
E}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/jsr311-api-
1.1.1.jar:{{HADOOP_COMMON_HOME}}/../../
./CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/jsr311-
api.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-admin-
1.0.0.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.142577
4/lib/hadoop/client/kerb-admin.jar:{{HADOOP_COMMON_HOME}}/../../CDH-
6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-client-
1.0.0.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-client.ja
r:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-common-

```

1.0.0.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-common.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-core-1.0.0.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-core.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-crypto-1.0.0.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-crypto.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-identity-1.0.0.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-identity.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-server-1.0.0.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-server.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-simplekdc-1.0.0.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-simplekdc.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-util-1.0.0.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-util.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerby-asn1-1.0.0.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerby-asn1.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerby-config-1.0.0.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerby-config.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerby-pkix-1.0.0.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerby-pkix.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerby-util-1.0.0.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerby-util.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerby-xdr-1.0.0.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerby-xdr.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/log4j-1.2.17.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/log4j.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/nimbus-jose-jwt-

4.41.1.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
 1.cdh6.2.1.p0.1425774/lib/hadoop/client/nimbus-jose-jwt.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
 1.cdh6.2.1.p0.1425774/lib/hadoop/client/okhttp-
 2.7.5.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
 1.cdh6.2.1.p0.1425774/lib/hadoop/client/okhttp.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1
 -1.cdh6.2.1.p0.1425774/lib/hadoop/client/okio-
 1.6.0.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
 1.cdh6.2.1.p0.1425774/lib/hadoop/client/okio.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/paranamer-2.8.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
 1.cdh6.2.1.p0.1425774/lib/hadoop/client/paranamer.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/protobuf-java-2.5.0.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/protobuf-java.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
 1.cdh6.2.1.p0.1425774/lib/hadoop/client/re2j-
 1.1.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/re2j.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
 1.cdh6.2.1.p0.1425774/lib/hadoop/client/snappy-java-
 1.1.4.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
 1.cdh6.2.1.p0.1425774/lib/hadoop/client/snappy-java.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
 1.cdh6.2.1.p0.1425774/lib/hadoop/client/stax2-api-
 3.1.4.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
 1.cdh6.2.1.p0.1425774/lib/hadoop/client/stax2-api.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/wildfly-openssl-
 1.0.4.Final.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
 1.cdh6.2.1.p0.1425774/lib/hadoop/client/wildfly-openssl.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/woodstox-core-
 5.0.3.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
 1.cdh6.2.1.p0.1425774/lib/hadoop/client/woodstox-core.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
 1.cdh6.2.1.p0.1425774/lib/hadoop/client/xz-1.6
 .jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
 1.cdh6.2.1.p0.1425774/lib/hadoop/client/xz.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hbase/bin/./lib/client-facing-thirdparty/audience-annotations-0.5.0.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
 1.cdh6.2.1.p0.1425774/lib/hbase/bin/./lib/client-facing-thirdparty/commons-logging-1.2.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
 1.cdh6.2.1.p0.1425774/lib/hbase/bin/./lib/client-facing-thirdparty/findbugs-annotations-1.3.9-
 1.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
 1.cdh6.2.1.p0.1425774/lib/hbase/bin/./lib/client-facing-thirdparty/htrace-core4-4.2.0-incubating.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-
 1.cdh6.2.1.p0.1425774/lib/hbase/bin/./lib/shaded-clients/hbase-shaded-mapreduce-2.1.0-cdh6.2.1.jar:{{HADOOP_COMMON_HOME}}/../../CDH-6.2.1-

```

1.cdh6.2.1.p0.1425774/lib/hbase/hbase-
spark.jar<CPS>{{PWD}}/__spark_conf__/_hadoop_conf__
MKL_NUM_THREADS -> 1
SPARK_DIST_CLASSPATH -> /opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/accessors-smart-
1.2.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/accessors-
smart.jar:/opt/cloudera/parce
ls/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/asm-
5.0.4.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/asm.jar:/opt/cloudera/parcels/CDH-
6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/avro.jar:/opt/cl
oudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/aws-java-
sdk-bundle-1.11.271.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/aws-java-sdk-
bundle.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.
1.p0.1425774/lib/hadoop/client/azure-data-lake-store-sdk-
2.2.9.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/azure-data-lake-store-
sdk.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/c
lient/commons-beanutils-1.9.3.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-
beanutils.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-cli-1.2.jar:/opt/cloudera/p
arcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-
cli.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-codec-
1.11.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/cl
ient/commons-codec.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-collections-
3.2.2.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-
collections.jar:/opt/cloudera/
parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-
compress-1.18.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-
compress.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/
hadoop/client/commons-configuration2-2.1.1.jar:/opt/cloudera/parcels/CDH-
6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-
configuration2.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-io-2.6.ja
r:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-
io.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-lang-
2.6.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774
/lib/hadoop/client/commons-lang.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-lang3-
3.7.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-lang3.jar:/opt/cloudera/p
arcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-logging-
1.1.3.jar:/opt/cloudera/parcels/CDH-6.2.1-

```

1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-logging.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-math3-3.1.1.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-math3.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-net-3.1.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/commons-net.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/curator-client-2.12.0.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/curator-client.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/curator-framework-2.12.0.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/curator-framework.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/curator-recipes-2.12.0.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/curator-recipes.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/gson-2.2.4.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/gson.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/guava-11.0.2.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/guava.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-annotations-3.0.0-cdh6.2.1.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-annotations.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-auth-3.0.0-cdh6.2.1.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-auth.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-aws-3.0.0-cdh6.2.1.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-aws.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-azure-3.0.0-cdh6.2.1.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-azure-datalake-3.0.0-cdh6.2.1.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-azure-datalake.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-azure.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-common-3.0.0-cdh6.2.1.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-common.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-hdfs-client-3.0.0-cdh6.2.1.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client

/hadoop-hdfs-client.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-mapreduce-client-common-
3.0.0-cdh6.2.1.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-mapreduce-
client-common.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-mapreduce-client-core-3.0.0-
cdh6.2.1.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-mapreduce-client-c
ore.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-mapreduce-client-jobclient-
3.0.0-cdh6.2.1.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-mapreduce-client-jobcli
ent.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-yarn-api-3.0.0-
cdh6.2.1.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-yarn-
api.jar:/opt/cloudera/parcels/CDH-6.
2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-yarn-client-3.0.0-
cdh6.2.1.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-yarn-
client.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/
hadoop/client/hadoop-yarn-common-3.0.0-
cdh6.2.1.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/hadoop-yarn-
common.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/htrace-core4-4.1
.0-incubating.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/htrace-
core4.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/httpclient-
4.5.3.jar:/opt/cloudera/parcels/CDH-6.2.1-1.c
dh6.2.1.p0.1425774/lib/hadoop/client/httpclient.jar:/opt/cloudera/parcels/CDH-
6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/httpcore-
4.4.6.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/httpcore.jar:/opt/cl
oudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/jaxb-api-
2.2.11.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/jaxb-
api.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/
client/jcip-annotations-1.0-1.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/jcip-
annotations.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/json-smart-2.3.jar:/opt/cloudera/par
cels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/json-
smart.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/jsp-api-
2.1.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/jsp-a
pi.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/jsr305-
3.0.0.jar:/opt/cloudera/parcels/CDH-6.2.1-

1.cdh6.2.1.p0.1425774/lib/hadoop/client/jsr305.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/jsr311-api-1.1.1.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/jsr311-api.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-admin-1.0.0.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-admin.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-client-1.0.0.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-client.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-common-1.0.0.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-common.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-core-1.0.0.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-core.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-crypto-1.0.0.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-crypto.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-identity-1.0.0.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-identity.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-server-1.0.0.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-server.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-simplekdc-1.0.0.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-simplekdc.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-util-1.0.0.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerb-util.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerby-asn1-1.0.0.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerby-asn1.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerby-config-1.0.0.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerby-config.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerby-pkix-1.0.0.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerby-pkix.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerby-util-1.0.0.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerby-

util.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerby-xdr-1.0.0.jar:/opt/cl
oudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/kerby-
xdr.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/log4j-
1.2.17.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/cl
ient/log4j.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/nimbus-jose-jwt-
4.41.1.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/nimbus-jose-
jwt.jar:/opt/cloudera/parcels/CDH-6.2
.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/okhttp-
2.7.5.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/okhttp.jar:/opt/cloudera/parcels/CDH-
6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/okio-1.6.0.jar:/opt/
cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/okio.jar:/opt/cloudera/parcels/CDH-
6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/paranamer-
2.8.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/clie
nt/paranamer.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/protobuf-java-
2.5.0.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/protobuf-
java.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/re2j-
1.1.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/re2j.jar:/opt/cloudera/parcels/CDH-
6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/snappy-java-1.1.4.jar:/opt/cl
oudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/snappy-
java.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/stax2-api-
3.1.4.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hado
op/client/stax2-api.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/wildfly-openssl-
1.0.4.Final.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/wildfly-openssl.jar:/opt/cloudera/p
arcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/client/woodstox-core-
5.0.3.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/woodstox-
core.jar:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop
/client/xz-1.6.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/client/xz.jar:/opt/cloudera/parcels/CDH-
6.2.1-1.cdh6.2.1.p0.1425774/lib/hbase/bin/./lib/client-facing-
thirdparty/audience-annotations-0.5.0.jar:/opt/c
loudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hbase/bin/./lib/client-
facing-thirdparty/commons-logging-1.2.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hbase/bin/./lib/client-facing-thirdparty/findbugs-
annotati
ons-1.3.9-1.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hbase/bin/./lib/client-facing-thirdparty/htrace-

```

core4-4.2.0-incubating.jar:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hbase/bin/../lib/shaded-cli
ents/hbase-shaded-mapreduce-2.1.0-cdh6.2.1.jar:/opt/cloudera/parcels/CDH-
6.2.1-1.cdh6.2.1.p0.1425774/lib/hbase/hbase-spark.jar
SPARK_YARN_STAGING_DIR ->
hdfs://nameservice1/user/raghavendr48edu/.sparkStaging/application_1684866
872278_1287
SPARK_USER -> raghavendr48edu
OPENBLAS_NUM_THREADS -> 1

```

command:

```

LD_LIBRARY_PATH="\{\{HADOOP_COMMON_HOME\}\}/../../CDH-6.2.1-
1.cdh6.2.1.p0.1425774/lib/hadoop/lib/native:$LD_LIBRARY_PATH\" \
\{\{JAVA_HOME\}\}/bin/java \
-server \
-Xmx1024m \
-Djava.io.tmpdir=\{\{PWD\}\}/tmp \
'-Dspark.driver.port=34843' \
'-Dspark.authenticate=false' \
'-Dspark.network.crypto.enabled=false' \
'-Dspark.shuffle.service.port=7337' \
-Dspark.yarn.app.container.log.dir=<LOG_DIR> \
-XX:OnOutOfMemoryError='kill %p' \
org.apache.spark.executor.CoarseGrainedExecutorBackend \
--driver-url \
spark://CoarseGrainedScheduler@ip-10-1-1-204.ap-south-
1.compute.internal:34843 \
--executor-id \
<executorId> \
--hostname \
<hostname> \
--cores \
1 \
--app-id \
application_1684866872278_1287 \
--user-class-path \
file:$PWD/_app_.jar \
1><LOG_DIR>/stdout \
2><LOG_DIR>/stderr

```

resources:

```

__spark_conf__ -> resource { scheme: "hdfs" host: "nameservice1" port: -1
file:
"/user/raghavendr48edu/.sparkStaging/application_1684866872278_1287/__sp
ark_conf__.zip" } size: 169453 timestamp: 1685019977258 type: ARCHIVE
visibilit
y: PRIVATE

```

```

=====
=====

```

23/05/25 13:06:23 INFO util.Utils: Using initial executors = 0, max of spark.dynamicAllocation.initialExecutors, spark.dynamicAllocation.minExecutors and spark.executor.instances
23/05/25 13:06:23 INFO conf.Configuration: resource-types.xml not found
23/05/25 13:06:23 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
23/05/25 13:06:24 INFO yarn.ApplicationMaster: Started progress reporter thread with (heartbeat : 3000, initial allocation : 200) intervals
23/05/25 13:06:26 INFO yarn.YarnAllocator: Driver requested a total number of 1 executor(s).
23/05/25 13:06:26 INFO yarn.YarnAllocator: Will request 1 executor container(s), each with 1 core(s) and 1408 MB memory (including 384 MB of overhead)
23/05/25 13:06:26 INFO yarn.YarnAllocator: Submitted container request for host ip-10-1-2-40.ap-south-1.compute.internal,ip-10-1-2-175.ap-south-1.compute.internal,ip-10-1-2-103.ap-south-1.compute.internal.
23/05/25 13:06:28 INFO yarn.YarnAllocator: Launching container container_1684866872278_1287_01_000002 on host ip-10-1-2-40.ap-south-1.compute.internal for executor with ID 1
23/05/25 13:06:28 INFO yarn.YarnAllocator: Received 1 containers from YARN, launching executors on 1 of them.
23/05/25 13:06:52 INFO yarn.YarnAllocator: Driver requested a total number of 0 executor(s).
23/05/25 13:06:52 INFO yarn.ApplicationMaster\$AMEndpoint: Driver terminated or disconnected! Shutting down. ip-10-1-1-204.ap-south-1.compute.internal:34843
23/05/25 13:06:52 INFO yarn.ApplicationMaster\$AMEndpoint: Driver terminated or disconnected! Shutting down. ip-10-1-1-204.ap-south-1.compute.internal:34843
23/05/25 13:06:52 INFO yarn.ApplicationMaster: Final app status: SUCCEEDED, exitCode: 0
23/05/25 13:06:52 INFO yarn.ApplicationMaster: Unregistering ApplicationMaster with SUCCEEDED
23/05/25 13:06:52 INFO impl.AMRMClientImpl: Waiting for application to be successfully unregistered.
23/05/25 13:06:53 INFO yarn.ApplicationMaster: Deleting staging directory hdfs://nameservice1/user/raghavendr48edu/.sparkStaging/application_1684866872278_1287
23/05/25 13:06:53 INFO util.ShutdownHookManager: Shutdown hook called

End of LogType:stderr

End of LogType:stdout

End of LogType:prelaunch.err

Container: container_1684866872278_1287_01_000002 on ip-10-1-2-40.ap-south-1.compute.internal_8041

LogAggregationType: AGGREGATED

=====
=====

LogType:prelaunch.out

LogLastModifiedTime:Thu May 25 13:06:53 +0000 2023

LogLength:70

LogContents:

Setting up env variables

Setting up job resources

Launching container

End of LogType:prelaunch.out

Container: container_1684866872278_1287_01_000002 on ip-10-1-2-40.ap-south-1.compute.internal_8041

LogAggregationType: AGGREGATED

=====
=====

LogType:stderr

LogLastModifiedTime:Thu May 25 13:06:53 +0000 2023

LogLength:14898

LogContents:

23/05/25 13:06:31 INFO executor.CoarseGrainedExecutorBackend: Started daemon with process name: 20157@ip-10-1-2-40.ap-south-1.compute.internal

23/05/25 13:06:31 INFO util.SignalUtils: Registered signal handler for TERM

23/05/25 13:06:31 INFO util.SignalUtils: Registered signal handler for HUP

23/05/25 13:06:31 INFO util.SignalUtils: Registered signal handler for INT

23/05/25 13:06:34 INFO spark.SecurityManager: Changing view acls to: yarn,raghavendr48edu

23/05/25 13:06:34 INFO spark.SecurityManager: Changing modify acls to: yarn,raghavendr48edu

23/05/25 13:06:34 INFO spark.SecurityManager: Changing view acls groups to:

23/05/25 13:06:34 INFO spark.SecurityManager: Changing modify acls groups to:

23/05/25 13:06:34 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(yarn, raghavendr48edu); groups with view permissions: Set(); users with modify permissions: Set

t(yarn, raghavendr48edu); groups with modify permissions: Set()

23/05/25 13:06:35 INFO client.TransportClientFactory: Successfully created connection to ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:34843 after 366 ms (0 ms spent in bootstraps)

23/05/25 13:06:36 INFO spark.SecurityManager: Changing view acls to: yarn,raghavendr48edu

23/05/25 13:06:36 INFO spark.SecurityManager: Changing modify acls to: yarn,raghavendr48edu
23/05/25 13:06:36 INFO spark.SecurityManager: Changing view acls groups to:
23/05/25 13:06:36 INFO spark.SecurityManager: Changing modify acls groups to:
23/05/25 13:06:36 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(yarn, raghavendr48edu); groups with view permissions: Set(); users with modify permissions: Set(yarn, raghavendr48edu); groups with modify permissions: Set()
23/05/25 13:06:36 INFO client.TransportClientFactory: Successfully created connection to ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:34843 after 40 ms (0 ms spent in bootstraps)
23/05/25 13:06:36 INFO storage.DiskBlockManager: Created local directory at /mnt/data/yarn/nm/usercache/raghavendr48edu/appcache/application_1684866872278_1287/blockmgr-c7497527-690a-445b-9ce9-1e4c41ff4676
23/05/25 13:06:36 INFO memory.MemoryStore: MemoryStore started with capacity 366.3 MB
23/05/25 13:06:38 INFO executor.CoarseGrainedExecutorBackend: Connecting to driver: spark://CoarseGrainedScheduler@ip-10-1-1-204.ap-south-1.compute.internal:34843
23/05/25 13:06:38 INFO executor.CoarseGrainedExecutorBackend: Successfully registered with driver
23/05/25 13:06:38 INFO executor.Executor: Starting executor ID 1 on host ip-10-1-2-40.ap-south-1.compute.internal
23/05/25 13:06:38 INFO util.Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 41263.
23/05/25 13:06:38 INFO netty.NettyBlockTransferService: Server created on ip-10-1-2-40.ap-south-1.compute.internal:41263
23/05/25 13:06:38 INFO storage.BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy
23/05/25 13:06:38 INFO storage.BlockManagerMaster: Registering BlockManager BlockManagerId(1, ip-10-1-2-40.ap-south-1.compute.internal, 41263, None)
23/05/25 13:06:38 INFO storage.BlockManagerMaster: Registered BlockManager BlockManagerId(1, ip-10-1-2-40.ap-south-1.compute.internal, 41263, None)
23/05/25 13:06:38 INFO storage.BlockManager: external shuffle service port = 7337
23/05/25 13:06:38 INFO storage.BlockManager: Registering executor with local external shuffle service.
23/05/25 13:06:38 INFO client.TransportClientFactory: Successfully created connection to ip-10-1-2-40.ap-south-1.compute.internal/10.1.2.40:7337 after 44 ms (0 ms spent in bootstraps)
23/05/25 13:06:38 INFO storage.BlockManager: Initialized BlockManager: BlockManagerId(1, ip-10-1-2-40.ap-south-1.compute.internal, 41263, None)
23/05/25 13:06:38 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 0
23/05/25 13:06:38 INFO executor.Executor: Running task 0.0 in stage 0.0 (TID 0)

23/05/25 13:06:38 INFO executor.Executor: Fetching spark://ip-10-1-1-204.ap-south-1.compute.internal:34843/jars/sparkme-project_2.11-1.0.jar with timestamp 1685019975886

23/05/25 13:06:38 INFO client.TransportClientFactory: Successfully created connection to ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:34843 after 11 ms (0 ms spent in bootstraps)

23/05/25 13:06:38 INFO util.Utils: Fetching spark://ip-10-1-1-204.ap-south-1.compute.internal:34843/jars/sparkme-project_2.11-1.0.jar to /mnt/data/yarn/nm/usercache/raghavendr48edu/appcache/application_1684866872278_1287/spark-cec639e8-f970-406d-8627-b2651deaeb04/fetchFileTemp7968286197553735895.tmp

23/05/25 13:06:38 INFO util.Utils: Copying /mnt/data/yarn/nm/usercache/raghavendr48edu/appcache/application_1684866872278_1287/spark-cec639e8-f970-406d-8627-b2651deaeb04/2257928111685019975886_cache to /mnt/data/yarn/nm/usercache/raghavendr48edu/appcache/application_1684866872278_1287/container_1684866872278_1287_01_000002/./sparkme-project_2.11-1.0.jar

23/05/25 13:06:38 INFO executor.Executor: Adding file:/mnt/data/yarn/nm/usercache/raghavendr48edu/appcache/application_1684866872278_1287/container_1684866872278_1287_01_000002/./sparkme-project_2.11-1.0.jar to class loader

23/05/25 13:06:39 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 1

23/05/25 13:06:39 INFO client.TransportClientFactory: Successfully created connection to ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:42947 after 3 ms (0 ms spent in bootstraps)

23/05/25 13:06:39 INFO memory.MemoryStore: Block broadcast_1_piece0 stored as bytes in memory (estimated size 5.6 KB, free 366.3 MB)

23/05/25 13:06:39 INFO broadcast.TorrentBroadcast: Reading broadcast variable 1 took 273 ms

23/05/25 13:06:40 INFO memory.MemoryStore: Block broadcast_1 stored as values in memory (estimated size 10.6 KB, free 366.3 MB)

23/05/25 13:06:43 INFO codegen.CodeGenerator: Code generated in 1023.979536 ms

23/05/25 13:06:47 INFO codegen.CodeGenerator: Code generated in 51.370738 ms

23/05/25 13:06:47 INFO datasources.FileScanRDD: Reading File path: hdfs://nameservice1/user/raghavendr48edu/input_sort_py.txt, range: 0-85, partition values: [empty row]

23/05/25 13:06:48 INFO codegen.CodeGenerator: Code generated in 48.007236 ms

23/05/25 13:06:48 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 0

23/05/25 13:06:48 INFO memory.MemoryStore: Block broadcast_0_piece0 stored as bytes in memory (estimated size 32.2 KB, free 366.3 MB)

23/05/25 13:06:48 INFO broadcast.TorrentBroadcast: Reading broadcast variable 0 took 76 ms

23/05/25 13:06:48 INFO memory.MemoryStore: Block broadcast_0 stored as values in memory (estimated size 443.0 KB, free 365.8 MB)

23/05/25 13:06:51 INFO memory.MemoryStore: Block rdd_4_0 stored as values in memory (estimated size 32.0 B, free 365.8 MB)

23/05/25 13:06:51 INFO executor.Executor: Finished task 0.0 in stage 0.0 (TID 0). 1438 bytes result sent to driver
23/05/25 13:06:51 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 1
23/05/25 13:06:51 INFO executor.Executor: Running task 0.0 in stage 1.0 (TID 1)
23/05/25 13:06:51 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 2
23/05/25 13:06:51 INFO memory.MemoryStore: Block broadcast_2_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.8 MB)
23/05/25 13:06:51 INFO broadcast.TorrentBroadcast: Reading broadcast variable 2 took 71 ms
23/05/25 13:06:51 INFO memory.MemoryStore: Block broadcast_2 stored as values in memory (estimated size 10.6 KB, free 365.8 MB)
23/05/25 13:06:51 INFO storage.BlockManager: Found block rdd_4_0 locally
23/05/25 13:06:51 INFO executor.Executor: Finished task 0.0 in stage 1.0 (TID 1). 1352 bytes result sent to driver
23/05/25 13:06:51 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 2
23/05/25 13:06:51 INFO executor.Executor: Running task 0.0 in stage 2.0 (TID 2)
23/05/25 13:06:51 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 3
23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_3_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.8 MB)
23/05/25 13:06:52 INFO broadcast.TorrentBroadcast: Reading broadcast variable 3 took 60 ms
23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_3 stored as values in memory (estimated size 10.6 KB, free 365.8 MB)
23/05/25 13:06:52 INFO storage.BlockManager: Found block rdd_4_0 locally
23/05/25 13:06:52 INFO executor.Executor: Finished task 0.0 in stage 2.0 (TID 2). 1352 bytes result sent to driver
23/05/25 13:06:52 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 3
23/05/25 13:06:52 INFO executor.Executor: Running task 0.0 in stage 3.0 (TID 3)
23/05/25 13:06:52 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 4
23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_4_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.8 MB)
23/05/25 13:06:52 INFO broadcast.TorrentBroadcast: Reading broadcast variable 4 took 37 ms
23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_4 stored as values in memory (estimated size 10.6 KB, free 365.8 MB)
23/05/25 13:06:52 INFO storage.BlockManager: Found block rdd_4_0 locally
23/05/25 13:06:52 INFO executor.Executor: Finished task 0.0 in stage 3.0 (TID 3). 1352 bytes result sent to driver
23/05/25 13:06:52 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 4
23/05/25 13:06:52 INFO executor.Executor: Running task 0.0 in stage 4.0 (TID 4)

23/05/25 13:06:52 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 5
23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_5_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.8 MB)
23/05/25 13:06:52 INFO broadcast.TorrentBroadcast: Reading broadcast variable 5 took 37 ms
23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_5 stored as values in memory (estimated size 10.6 KB, free 365.8 MB)
23/05/25 13:06:52 INFO storage.BlockManager: Found block rdd_4_0 locally
23/05/25 13:06:52 INFO executor.Executor: Finished task 0.0 in stage 4.0 (TID 4). 1352 bytes result sent to driver
23/05/25 13:06:52 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 5
23/05/25 13:06:52 INFO executor.Executor: Running task 0.0 in stage 5.0 (TID 5)
23/05/25 13:06:52 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 6
23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_6_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.8 MB)
23/05/25 13:06:52 INFO broadcast.TorrentBroadcast: Reading broadcast variable 6 took 34 ms
23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_6 stored as values in memory (estimated size 10.6 KB, free 365.7 MB)
23/05/25 13:06:52 INFO storage.BlockManager: Found block rdd_4_0 locally
23/05/25 13:06:52 INFO executor.Executor: Finished task 0.0 in stage 5.0 (TID 5). 1352 bytes result sent to driver
23/05/25 13:06:52 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 6
23/05/25 13:06:52 INFO executor.Executor: Running task 0.0 in stage 6.0 (TID 6)
23/05/25 13:06:52 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 7
23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_7_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.7 MB)
23/05/25 13:06:52 INFO broadcast.TorrentBroadcast: Reading broadcast variable 7 took 28 ms
23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_7 stored as values in memory (estimated size 10.6 KB, free 365.7 MB)
23/05/25 13:06:52 INFO storage.BlockManager: Found block rdd_4_0 locally
23/05/25 13:06:52 INFO executor.Executor: Finished task 0.0 in stage 6.0 (TID 6). 1352 bytes result sent to driver
23/05/25 13:06:52 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 7
23/05/25 13:06:52 INFO executor.Executor: Running task 0.0 in stage 7.0 (TID 7)
23/05/25 13:06:52 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 8
23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_8_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.7 MB)
23/05/25 13:06:52 INFO broadcast.TorrentBroadcast: Reading broadcast variable 8 took 28 ms

23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_8 stored as values in memory (estimated size 10.6 KB, free 365.7 MB)
23/05/25 13:06:52 INFO storage.BlockManager: Found block rdd_4_0 locally
23/05/25 13:06:52 INFO executor.Executor: Finished task 0.0 in stage 7.0 (TID 7). 1352 bytes result sent to driver
23/05/25 13:06:52 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 8
23/05/25 13:06:52 INFO executor.Executor: Running task 0.0 in stage 8.0 (TID 8)
23/05/25 13:06:52 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 9
23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_9_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.7 MB)
23/05/25 13:06:52 INFO broadcast.TorrentBroadcast: Reading broadcast variable 9 took 28 ms
23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_9 stored as values in memory (estimated size 10.6 KB, free 365.7 MB)
23/05/25 13:06:52 INFO storage.BlockManager: Found block rdd_4_0 locally
23/05/25 13:06:52 INFO executor.Executor: Finished task 0.0 in stage 8.0 (TID 8). 1352 bytes result sent to driver
23/05/25 13:06:52 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 9
23/05/25 13:06:52 INFO executor.Executor: Running task 0.0 in stage 9.0 (TID 9)
23/05/25 13:06:52 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 10
23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_10_piece0 stored as bytes in memory (estimated size 5.6 KB, free 365.7 MB)
23/05/25 13:06:52 INFO broadcast.TorrentBroadcast: Reading broadcast variable 10 took 27 ms
23/05/25 13:06:52 INFO memory.MemoryStore: Block broadcast_10 stored as values in memory (estimated size 10.6 KB, free 365.7 MB)
23/05/25 13:06:52 INFO storage.BlockManager: Found block rdd_4_0 locally
23/05/25 13:06:52 INFO executor.Executor: Finished task 0.0 in stage 9.0 (TID 9). 1352 bytes result sent to driver
23/05/25 13:06:52 INFO executor.CoarseGrainedExecutorBackend: Driver commanded a shutdown
23/05/25 13:06:52 INFO executor.CoarseGrainedExecutorBackend: Driver from ip-10-1-1-204.ap-south-1.compute.internal:34843 disconnected during shutdown
23/05/25 13:06:52 INFO executor.CoarseGrainedExecutorBackend: Driver from ip-10-1-1-204.ap-south-1.compute.internal:34843 disconnected during shutdown
23/05/25 13:06:52 INFO memory.MemoryStore: MemoryStore cleared
23/05/25 13:06:52 INFO storage.BlockManager: BlockManager stopped
23/05/25 13:06:52 INFO util.ShutdownHookManager: Shutdown hook called
23/05/25 13:06:52 INFO util.ShutdownHookManager: Deleting directory /mnt/data/yarn/nm/usercache/raghavendr48edu/appcache/application_1684866872278_1287/spark-cec639e8-f970-406d-8627-b2651deaeb04

End of LogType:stderr

End of LogType:stdout

From <<https://bigdata.cloudlabs.edureka.co/shell/>>

5.Transfer the sample dataset from RDBMS to HDFS

Step 9.1: Download Dataset -Download dataset and transfer to cloud lab using FTP

Step 9.2: Load data in mysql - Login into mysql - Create schema in mysql

show databases;

use database

Show tables;

127 login: raghavendr48edu

raghavendr48edu@127.0.0.1's password:

Last login: Thu May 18 14:47:35 2023 from localhost

[raghavendr48edu@ip-10-1-1-204 ~]\$ **mysql -u raghavendr48edu -pRedKangaroo99\$**

mysql: [Warning] Using a password on the command line interface can be insecure.

Welcome to the MySQL monitor. Commands end with ; or \g.

Your MySQL connection id is 3441

Server version: 5.7.33 MySQL Community Server (GPL)

Copyright (c) 2000, 2021, Oracle and/or its affiliates.

Oracle is a registered trademark of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

Or

[raghavendr48edu@ip-10-1-1-204 ~]\$ **mysql -u raghavendr48edu -p**

Enter password:

Welcome to the MySQL monitor. Commands end with ; or \g.

Your MySQL connection id is 3538

Server version: 5.7.33 MySQL Community Server (GPL)

Copyright (c) 2000, 2021, Oracle and/or its affiliates.

Oracle is a registered trademark of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql>

From <<https://bigdata.cloudlabs.edureka.co/shell/>>

```
mysql> select @@hostname;
+-----+
| @@hostname |
+-----+
| ip-10-1-1-204.ap-south-1.compute.internal |
+-----+
1 row in set (0.00 sec)
```

From <<https://bigdata.cloudlabs.edureka.co/shell/>>

```
mysql> create table financial_regulation(SYMBOL varchar(100),SERIES
varchar(50), OPEN double, HIGH double, LOW double, CLOSE double,
LAST double, PREVCLOSE d
ouble,TOTTRDQTY double,TOITRDVAL double,TIMESTAMP date,
TOTALTRADES int,ISIN varchar(50));
ERROR 1046 (3D000): No database selected
```

```
mysql> show databases
```

```
-> show databases;
```

```
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual
that corresponds to your MySQL server version for the right syntax to use near
'sh
```

```
ow databases' at line 2
```

```
mysql> show databases;
```

```
+-----+
| Database |
+-----+
| information_schema |
| raghavendr48edu |
+-----+
2 rows in set (0.03 sec)
```

```
mysql> use raghavendr48edu
```

```
Database changed
```

```
mysql> show tables;
```

```
Empty set (0.00 sec)
```

```
mysql> create table financial_regulation(SYMBOL varchar(100),SERIES
varchar(50), OPEN double, HIGH double, LOW double, CLOSE double,
LAST double, PREVCLOSE double,TOTTRDQTY double,TOITRDVAL double,TIMESTAMP date,
TOTALTRADES int,ISIN varchar(50));
Query OK, 0 rows affected (0.01 sec)
```

```
mysql> show tables;
```

```
+-----+
| Tables_in_raghavendr48edu |
+-----+
| financial_regulation      |
+-----+
1 row in set (0.00 sec)
```

From <<https://bigdata.cloudlabs.edureka.co/shell/>>

- Load csv in table

```
mysql> load data local infile
```

```
'/home/raghavendr48edu/FINAL_FROM_DF.csv' into table
financial_regulation FIELDS TERMINATED BY ',';
```

```
Query OK, 846405 rows affected, 10 warnings (5.98 sec)
```

```
Records: 846405 Deleted: 0 Skipped: 0 Warnings: 10
```

```
mysql> select * from financial_regulation limit 5;
```

```
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
| SYMBOL   | SERIES | OPEN   | HIGH   | LOW    | CLOSE  | LAST   |
PREVCLOSE | TOTTRDQTY | TOITRDVAL | TIMESTAMP | TOTALTRADES | ISIN
|
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
| SYMBOL   | SERIES |      0 |      0 |      0 |      0 |      0 |      0 |
0 | 0000-00-00 |      0 | ISIN    |
| 20MICRONS | EQ    |    37.8 |    37.8 | 36.15 | 36.85 | 37.4 |    37.05 |
27130 |    994657.9 | 2017-06-28 |    202 | INE144J01027 |
| 3IINFOTECH | EQ    |     4.1 |     4.85 | 4 | 4.55 | 4.65 |     4.05 |
20157058 | 92148517.65 | 2017-06-28 |    7353 | INE748C01020 |
| 3MINDIA   | EQ    | 13425.15 | 13469.55 | 12920 | 13266.7 | 13300 |
13460.55 |    2290 | 30304923.35 | 2017-06-28 |    748 | INE470A01017 |
| 63MOONS   | EQ    |     61 |     61.9 | 60.35 | 61 | 61.1 |     60.65 |
27701 |    1689421 | 2017-06-28 |    437 | INE111B01023 |
```

```
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
5 rows in set (0.01 sec)
```

From <<https://bigdata.cloudlabs.edureka.co/shell/>>

```
mysql> exit;
Bye
```

```
mysql> load data local infile
'/home/raghavendr48edu/FINAL_FROM_DF.csv' into table
financial_regulation FIELDS TERMINATED BY ',' IGNORE 1 ROWS;
Query OK, 846404 rows affected (6.07 sec)
Records: 846404 Deleted: 0 Skipped: 0 Warnings: 0
```

```
mysql> select * from financial_regulation limit 5;
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
| SYMBOL | SERIES | OPEN | HIGH | LOW | CLOSE | LAST |
PREVCLOSE | TOTTRDQTY | TOITRDVAL | TIMESTAMP | TOTALTRADES | ISIN |
|
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
| 20MICRONS | EQ | 37.8 | 37.8 | 36.15 | 36.85 | 37.4 | 37.05 |
27130 | 994657.9 | 2017-06-28 | 202 | INE144J01027 |
| 3IINFOTECH | EQ | 4.1 | 4.85 | 4 | 4.55 | 4.65 | 4.05 |
20157058 | 92148517.65 | 2017-06-28 | 7353 | INE748C01020 |
| 3MINDIA | EQ | 13425.15 | 13469.55 | 12920 | 13266.7 | 13300 |
13460.55 | 2290 | 30304923.35 | 2017-06-28 | 748 | INE470A01017 |
| 63MOONS | EQ | 61 | 61.9 | 60.35 | 61 | 61.1 | 60.65 |
27701 | 1689421 | 2017-06-28 | 437 | INE111B01023 |
| 8KMILES | EQ | 546.1 | 548 | 535 | 537.45 | 535.2 | 547.45 |
79722 | 43208620.05 | 2017-06-28 | 1866 | INE650K01021 |
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
5 rows in set (0.00 sec)
```

```
mysql> exit;
Bye
```

From <<https://bigdata.cloudlabs.edureka.co/shell/>>

Step 9.3: Transferring data from Mysql to HDFS - List Databases

```
sqoop list-databases --connect jdbc:mysql://sqoopdb.edu.cloudlab.com/userdb -- username  
raghavendr48edu --password RedKangaroo99$
```

```
sqoop list-databases --connect jdbc:mysql://localhost --username  
raghavendr48edu --password RedKangaroo99$
```

```
Warning: /opt/cloudera/parcels/CDH-6.2.1-  
1.cdh6.2.1.p0.1425774/bin/./lib/sqoop/./accumulo does not exist! Accumulo  
imports will fail.  
Please set $ACCUMULO_HOME to the root of your Accumulo installation.  
SLF4J: Class path contains multiple SLF4J bindings.  
SLF4J: Found binding in [jar:file:/opt/cloudera/parcels/CDH-6.2.1-  
1.cdh6.2.1.p0.1425774/jars/slf4j-log4j12-  
1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.clas  
s]  
SLF4J: Found binding in [jar:file:/opt/cloudera/parcels/CDH-6.2.1-  
1.cdh6.2.1.p0.1425774/jars/log4j-slf4j-impl-  
2.8.2.jar!/org/slf4j/impl/StaticLoggerBinder.cl  
ass]  
SLF4J: See http://www.slf4j.org/codes.html#multiple\_bindings for an explanati  
on.  
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]  
23/05/19 10:10:48 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7-cdh6.2.1  
23/05/19 10:10:49 WARN tool.BaseSqoopTool: Setting your password on the  
command-line is insecure. Consider using -P instead.  
23/05/19 10:10:49 INFO manager.MySQLManager: Preparing to use a MySQL  
streaming resultset.  
information_schema  
raghavendr48edu
```

From <<https://bigdata.cloudlabs.edureka.co/shell/>>

```
[raghavendr48edu@ip-10-1-1-204 ~]$ sqoop list-databases --connect  
jdbc:mysql://ip-10-1-1-204.ap-south-1.compute.internal --username  
raghavendr48edu --password RedKangaroo99$
```

```
Warning: /opt/cloudera/parcels/CDH-6.2.1-  
1.cdh6.2.1.p0.1425774/bin/./lib/sqoop/./accumulo does not exist! Accumulo  
imports will fail.  
Please set $ACCUMULO_HOME to the root of your Accumulo installation.  
SLF4J: Class path contains multiple SLF4J bindings.  
SLF4J: Found binding in [jar:file:/opt/cloudera/parcels/CDH-6.2.1-  
1.cdh6.2.1.p0.1425774/jars/slf4j-log4j12-  
1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.clas  
s]  
SLF4J: Found binding in [jar:file:/opt/cloudera/parcels/CDH-6.2.1-  
1.cdh6.2.1.p0.1425774/jars/log4j-slf4j-impl-  
2.8.2.jar!/org/slf4j/impl/StaticLoggerBinder.cl  
ass]
```


SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]

23/05/19 10:14:17 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7-cdh6.2.1

23/05/19 10:14:17 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.

23/05/19 10:14:17 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.

information_schema

raghavendr48edu

[raghavendr48edu@ip-10-1-1-204 ~]\$

From <<https://bigdata.cloudlabs.edureka.co/shell/>>

```
[raghavendr48edu@ip-10-1-1-204 ~]$ sqoop list-databases --connect  
jdbc:mysql://ip-10-1-1-204.ap-south-  
1.compute.internal:3306/raghavendr48edu --username  
raghavendr48edu --password RedKangaroo99$
```

Warning: /opt/cloudera/parcels/CDH-6.2.1-

1.cdh6.2.1.p0.1425774/bin/./lib/sqoop/./accumulo does not exist! Accumulo imports will fail.

Please set \$ACCUMULO_HOME to the root of your Accumulo installation.

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/opt/cloudera/parcels/CDH-6.2.1-

1.cdh6.2.1.p0.1425774/jars/slf4j-log4j12-

1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/opt/cloudera/parcels/CDH-6.2.1-

1.cdh6.2.1.p0.1425774/jars/log4j-slf4j-impl-

2.8.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]

23/05/19 14:30:29 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7-cdh6.2.1

23/05/19 14:30:29 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.

23/05/19 14:30:30 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.

information_schema

raghavendr48edu

From <<https://bigdata.cloudlabs.edureka.co/shell/>>

```
[raghavendr48edu@ip-10-1-1-204 ~]$ sqoop import --connect
jdbc:mysql://ip-10-1-1-204.ap-south-
1.compute.internal:3306/raghavendr48edu --username raghavendr48
edu --password RedKangaroo99$ --table financial_regulation --m 1 --
driver com.mysql.jdbc.Driver --target-dir
/user/raghavendr48edu/financial_regulation_data/
Warning: /opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/bin/./lib/sqoop/./accumulo does not exist! Accumulo
imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/jars/slf4j-log4j12-
1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.clas
s]
SLF4J: Found binding in [jar:file:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/jars/log4j-slf4j-impl-
2.8.2.jar!/org/slf4j/impl/StaticLoggerBinder.cl
ass]
SLF4J: See http://www.slf4j.org/codes.html#multiple\_bindings for an explanati
on.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
23/05/19 17:01:33 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7-cdh6.2.1
23/05/19 17:01:33 WARN tool.BaseSqoopTool: Setting your password on the
command-line is insecure. Consider using -P instead.
23/05/19 17:01:33 WARN sqoop.ConnFactory: Parameter --driver is set to an
explicit driver however appropriate connection manager is not being set (via --
conn
ection-manager). Sqoop is going to fall back to
org.apache.sqoop.manager.GenericJdbcManager. Please specify explicitly which
connection manager should be use
d next time.
23/05/19 17:01:33 INFO manager.SqlManager: Using default fetchSize of 1000
23/05/19 17:01:33 INFO tool.CodeGenTool: Beginning code generation
23/05/19 17:01:34 INFO manager.SqlManager: Executing SQL statement:
SELECT t.* FROM financial_regulation AS t WHERE 1=0
23/05/19 17:01:34 INFO manager.SqlManager: Executing SQL statement:
SELECT t.* FROM financial_regulation AS t WHERE 1=0
23/05/19 17:01:34 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is
/opt/cloudera/parcels/CDH/lib/hadoop-mapreduce
23/05/19 17:01:35 ERROR orm.CompilationManager: Could not rename
/tmp/sqoop-
raghavendr48edu/compile/09fd834432c724d694319b2fb5866004/financial_regu
lation.jav
a to /home/raghavendr48edu/./financial_regulation.java. Error: Destination
'/home/raghavendr48edu/./financial_regulation.java' already exists
23/05/19 17:01:35 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-
raghavendr48edu/compile/09fd834432c724d694319b2fb5866004/financial_regu
lation.jar
23/05/19 17:01:35 INFO mapreduce.ImportJobBase: Beginning import of
financial_regulation
```

23/05/19 17:01:36 INFO Configuration.deprecation: mapred.jar is deprecated.
Instead, use mapreduce.job.jar
23/05/19 17:01:36 INFO manager.SqlManager: Executing SQL statement:
SELECT t.* FROM financial_regulation AS t WHERE 1=0
23/05/19 17:01:36 INFO Configuration.deprecation: mapred.map.tasks is
deprecated. Instead, use mapreduce.job.maps
23/05/19 17:01:36 INFO client.RMProxy: Connecting to ResourceManager at ip-
10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/19 17:01:37 INFO mapreduce.JobResourceUploader: Disabling Erasure
Coding for path: /user/raghavendr48edu/.staging/job_1684298513961_0686
23/05/19 17:01:39 INFO db.DBInputFormat: Using read committed transaction
isolation
23/05/19 17:01:39 INFO mapreduce.JobSubmitter: number of splits:1
23/05/19 17:01:39 INFO Configuration.deprecation:
yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead,
use yarn.system-metrics-publi
sher.enabled
23/05/19 17:01:39 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1684298513961_0686
23/05/19 17:01:39 INFO mapreduce.JobSubmitter: Executing with tokens: []
23/05/19 17:01:40 INFO conf.Configuration: resource-types.xml not found
23/05/19 17:01:40 INFO resource.ResourceUtils: Unable to find 'resource-
types.xml'.
23/05/19 17:01:40 INFO impl.YarnClientImpl: Submitted application
application_1684298513961_0686
23/05/19 17:01:40 INFO mapreduce.Job: The url to track the job: [http://ip-10-
1-1-204.ap-south-
1.compute.internal:6066/proxy/application_1684298513961_0686/](http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684298513961_0686/)
23/05/19 17:01:40 INFO mapreduce.Job: Running job:
job_1684298513961_0686
23/05/19 17:01:47 INFO mapreduce.Job: Job job_1684298513961_0686
running in uber mode : false
23/05/19 17:01:47 INFO mapreduce.Job: map 0% reduce 0%
23/05/19 17:02:02 INFO mapreduce.Job: map 100% reduce 0%
23/05/19 17:02:03 INFO mapreduce.Job: Job job_1684298513961_0686
completed successfully
23/05/19 17:02:03 INFO mapreduce.Job: Counters: 33
File System Counters
FILE: Number of bytes read=0
FILE: Number of bytes written=246524
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=85
HDFS: Number of bytes written=80474953
HDFS: Number of read operations=6
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0
Job Counters
Launched map tasks=1
Other local map tasks=1

```
Total time spent by all maps in occupied slots (ms)=13055
Total time spent by all reduces in occupied slots (ms)=0
Total time spent by all map tasks (ms)=13055
Total vcore-milliseconds taken by all map tasks=13055
Total megabyte-milliseconds taken by all map tasks=13368320
Map-Reduce Framework
  Map input records=846404
  Map output records=846404
  Input split bytes=85
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=1286
  CPU time spent (ms)=13760
  Physical memory (bytes) snapshot=759418880
  Virtual memory (bytes) snapshot=2604331008
  Total committed heap usage (bytes)=791150592
  Peak Map Physical memory (bytes)=759418880
  Peak Map Virtual memory (bytes)=2604331008
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=80474953
23/05/19 17:02:03 INFO mapreduce.ImportJobBase: Transferred 76.7469 MB in
26.7332 seconds (2.8708 MB/sec)
23/05/19 17:02:03 INFO mapreduce.ImportJobBase: Retrieved 846404 records.
[raghavendr48edu@ip-10-1-1-204 ~]$
```

From <<https://bigdata.cloudlabs.edureka.co/shell/>>

6. Validate the loaded data by comparing the statistics of data both in source and HDFS

```
[raghavendr48edu@ip-10-1-1-204 ~]$ hdfs dfs -cat
/user/raghavendr48edu/financial_regulation_data/*|wc -l
846404
```

From <<https://chat.openai.com/c/969e01fc-3524-47bc-a24f-101e1128de61>>

```
[raghavendr48edu@ip-10-1-1-204 ~]$ hdfs dfs -cat
/user/raghavendr48edu/financial_regulation_data/*|head -5
```

```

20MICRONS,EQ,37.8,37.8,36.15,36.85,37.4,37.05,27130.0,994657.9,2017-06-
28,202,INE144J01027
3IINFOTECH,EQ,4.1,4.85,4.0,4.55,4.65,4.05,2.0157058E7,9.214851765E7,2017
-06-28,7353,INE748C01020
3MINDIA,EQ,13425.15,13469.55,12920.0,13266.7,13300.0,13460.55,2290.0,3.
030492335E7,2017-06-28,748,INE470A01017
63MOONS,EQ,61.0,61.9,60.35,61.0,61.1,60.65,27701.0,1689421.0,2017-06-
28,437,INE111B01023
8KMILES,EQ,546.1,548.0,535.0,537.45,535.2,547.45,79722.0,4.320862005E7,
2017-06-28,1866,INE650K01021
cat: Unable to write to output stream.

```

From <<https://bigdata.cloudlabs.edureka.co/shell/>>

7.Create a new directory EQ in HDFS and transfer the data where series is EQ

```
hdfs dfs -mkdir /user/raghavendr48edu/financial_regulation_data/eq
```

```

[raghavendr48edu@ip-10-1-1-204 ~]$ hdfs dfs -rm -r
/user/raghavendr48edu/financial_regulation_data/eq
23/05/19 11:35:18 INFO fs.TrashPolicyDefault: Moved:
'hdfs://nameservice1/user/raghavendr48edu/financial_regulation_data/eq' to
trash at: hdfs://nameservice1
/user/raghavendr48edu/.Trash/Current/user/raghavendr48edu/financial_regulati
on_data/eq

```

```

[raghavendr48edu@ip-10-1-1-204 ~]$ sqoop import --connect
jdbc:mysql://ip-10-1-1-204.ap-south-
1.compute.internal:3306/raghavendr48edu --username raghavendr4
8edu --password RedKangaroo99$ -m 1 --table financial_regulation --
where "SERIES='EQ'" --target-dir
/user/raghavendr48edu/financial_regulation_data/eq --driver
com.mysql.jdbc.Driver

```

```

Warning: /opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/bin/./lib/sqoop/./accumulo does not exist! Accumulo
imports will fail.

```

```
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
```

```
SLF4J: Class path contains multiple SLF4J bindings.
```

```

SLF4J: Found binding in [jar:file:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/jars/slf4j-log4j12-
1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.clas
s]

```

```

SLF4J: Found binding in [jar:file:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/jars/log4j-slf4j-impl-
2.8.2.jar!/org/slf4j/impl/StaticLoggerBinder.cl
ass]

```

SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]

23/05/19 17:09:52 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7-cdh6.2.1

23/05/19 17:09:52 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.

23/05/19 17:09:52 WARN sqoop.ConnFactory: Parameter --driver is set to an explicit driver however appropriate connection manager is not being set (via --conn

ection-manager). Sqoop is going to fall back to

org.apache.sqoop.manager.GenericJdbcManager. Please specify explicitly which connection manager should be used next time.

23/05/19 17:09:52 INFO manager.SqlManager: Using default fetchSize of 1000

23/05/19 17:09:52 INFO tool.CodeGenTool: Beginning code generation

23/05/19 17:09:53 INFO manager.SqlManager: Executing SQL statement:

SELECT t.* FROM financial_regulation AS t WHERE 1=0

23/05/19 17:09:53 INFO manager.SqlManager: Executing SQL statement:

SELECT t.* FROM financial_regulation AS t WHERE 1=0

23/05/19 17:09:53 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /opt/cloudera/parcels/CDH/lib/hadoop-mapreduce

23/05/19 17:09:55 ERROR orm.CompilationManager: Could not rename

/tmp/sqoop-

raghavendr48edu/compile/0edf9693a832d2505650f6de1279c071/financial_regulation.jav

a to /home/raghavendr48edu/./financial_regulation.java. Error: Destination '/home/raghavendr48edu/./financial_regulation.java' already exists

23/05/19 17:09:55 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-raghavendr48edu/compile/0edf9693a832d2505650f6de1279c071/financial_regulation.jar

23/05/19 17:09:55 INFO mapreduce.ImportJobBase: Beginning import of financial_regulation

23/05/19 17:09:55 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar

23/05/19 17:09:55 INFO manager.SqlManager: Executing SQL statement:

SELECT t.* FROM financial_regulation AS t WHERE 1=0

23/05/19 17:09:56 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps

23/05/19 17:09:56 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032

23/05/19 17:09:56 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /user/raghavendr48edu/.staging/job_1684298513961_0687

23/05/19 17:09:58 INFO db.DBInputFormat: Using read committed transaction isolation

23/05/19 17:09:59 INFO mapreduce.JobSubmitter: number of splits:1

23/05/19 17:09:59 INFO Configuration.deprecation:

yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled

23/05/19 17:09:59 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1684298513961_0687

23/05/19 17:09:59 INFO mapreduce.JobSubmitter: Executing with tokens: []

23/05/19 17:09:59 INFO conf.Configuration: resource-types.xml not found
23/05/19 17:09:59 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
23/05/19 17:09:59 INFO impl.YarnClientImpl: Submitted application application_1684298513961_0687
23/05/19 17:09:59 INFO mapreduce.Job: The url to track the job: http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684298513961_0687/
23/05/19 17:09:59 INFO mapreduce.Job: Running job: job_1684298513961_0687
23/05/19 17:10:08 INFO mapreduce.Job: Job job_1684298513961_0687 running in uber mode : false
23/05/19 17:10:08 INFO mapreduce.Job: map 0% reduce 0%
23/05/19 17:10:27 INFO mapreduce.Job: map 100% reduce 0%
23/05/19 17:10:27 INFO mapreduce.Job: Job job_1684298513961_0687 completed successfully
23/05/19 17:10:27 INFO mapreduce.Job: Counters: 33
File System Counters
FILE: Number of bytes read=0
FILE: Number of bytes written=246865
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=85
HDFS: Number of bytes written=70804177
HDFS: Number of read operations=6
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0
Job Counters
Launched map tasks=1
Other local map tasks=1
Total time spent by all maps in occupied slots (ms)=17585
Total time spent by all reduces in occupied slots (ms)=0
Total time spent by all map tasks (ms)=17585
Total vcore-milliseconds taken by all map tasks=17585
Total megabyte-milliseconds taken by all map tasks=18007040
Map-Reduce Framework
Map input records=739199
Map output records=739199
Input split bytes=85
Spilled Records=0
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=4161
CPU time spent (ms)=16810
Physical memory (bytes) snapshot=695107584
Virtual memory (bytes) snapshot=2596241408
Total committed heap usage (bytes)=769130496
Peak Map Physical memory (bytes)=695107584
Peak Map Virtual memory (bytes)=2596241408
File Input Format Counters

```

Bytes Read=0
File Output Format Counters
Bytes Written=70804177
23/05/19 17:10:27 INFO mapreduce.ImportJobBase: Transferred 67.5241 MB in
31.6986 seconds (2.1302 MB/sec)
23/05/19 17:10:27 INFO mapreduce.ImportJobBase: Retrieved 739199 records.
[raghavendr48edu@ip-10-1-1-204 ~]$

```

From <<https://bigdata.cloudlabs.edureka.co/shell/>>

8.Set total trades which are less than 500 to 0 and transfer only updated rows.

```

mysql> alter table financial_regulation add column timestamp timestamp;
ERROR 1060 (42S21): Duplicate column name 'timestamp'
mysql> alter table financial_regulation add column time_stamp timestamp;
Query OK, 0 rows affected (8.37 sec)
Records: 0 Duplicates: 0 Warnings: 0

```

```
mysql> select * from financial_regulation limit 5;
```

```

+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+
| SYMBOL    | SERIES | OPEN    | HIGH    | LOW    | CLOSE   | LAST    |
PREVCLOSE | TOTTRDQTY | TOITRDVAL | TIMESTAMP | TOTALTRADES | ISIN
| time_stamp      |
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+
| 20MICRONS | EQ     | 37.8 | 37.8 | 36.15 | 36.85 | 37.4 | 37.05 |
27130 | 994657.9 | 2017-06-28 | 202 | INE144J01027 | 2023-05-21
06:29:03 |
| 3IINFOTECH | EQ     | 4.1 | 4.85 | 4 | 4.55 | 4.65 | 4.05 |
20157058 | 92148517.65 | 2017-06-28 | 7353 | INE748C01020 | 2023-05-21
06:29:03 |
| 3MINDIA    | EQ     | 13425.15 | 13469.55 | 12920 | 13266.7 | 13300 |
13460.55 | 2290 | 30304923.35 | 2017-06-28 | 748 | INE470A01017 |
2023-05-21 06:29:03 |
| 63MOONS    | EQ     | 61 | 61.9 | 60.35 | 61 | 61.1 | 60.65 |
27701 | 1689421 | 2017-06-28 | 437 | INE111B01023 | 2023-05-21
06:29:03 |
| 8KMILES    | EQ     | 546.1 | 548 | 535 | 537.45 | 535.2 | 547.45 |
79722 | 43208620.05 | 2017-06-28 | 1866 | INE650K01021 | 2023-05-21
06:29:03 |

```



```
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
-----+
5 rows in set (0.00 sec)
```

```
mysql> alter table financial_regulation add column currentdate DATE;
(you get null values)
Query OK, 0 rows affected (7.66 sec)
Records: 0 Duplicates: 0 Warnings: 0
```

```
mysql> update financial_regulation SET currentdate=CURRENT_DATE;
Query OK, 846404 rows affected (16.18 sec)
Rows matched: 846404 Changed: 846404 Warnings: 0
```

```
mysql> alter table financial_regulation add updatetrades INT(11);
Query OK, 0 rows affected (7.02 sec)
Records: 0 Duplicates: 0 Warnings: 0
```

```
mysql> update financial_regulation set updatetrades=0 where
TOTALTRADES<500;
Query OK, 399261 rows affected (8.48 sec)
Rows matched: 399261 Changed: 399261 Warnings: 0
```

```
[raghavendr48edu@ip-10-1-1-204 ~]$ lastvalue=$(mysql -u
raghavendr48edu -p -h ip-10-1-1-204.ap-south-1.compute.internal -D
raghavendr48edu -e "SELECT max(time_stamp) from
financial_regulation" -sN)
Enter password:
```

```
[raghavendr48edu@ip-10-1-1-204 ~]$ sqoop import --connect
jdbc:mysql://ip-10-1-1-204.ap-south-
1.compute.internal/raghavendr48edu --username raghavendr48edu --
password RedKangaroo99$ --table financial_regulation --where
"SERIES='EQ'" --incremental lastmodified --check-column time_stamp -
-last-value "$lastvalue" --target-dir
/user/raghavendr48edu/financial_regulation_updated/ --driver
com.mysql.jdbc.Driver -m 1
Warning: /opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/bin/../lib/sqoop/./accumulo does not exist! Accumulo
imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/opt/cloudera/parcels/CDH-6.2.1-
1.cdh6.2.1.p0.1425774/jars/slf4j-log4j12-
1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
```

SLF4J: Found binding in [jar:file:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/jars/log4j-slf4j-impl-2.8.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
23/05/21 06:51:32 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7-cdh6.2.1
23/05/21 06:51:32 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
23/05/21 06:51:32 WARN sqoop.ConnFactory: Parameter --driver is set to an explicit driver however appropriate connection manager is not being set (via --connection-manager). Sqoop is going to fall back to org.apache.sqoop.manager.GenericJdbcManager. Please specify explicitly which connection manager should be used next time.
23/05/21 06:51:32 INFO manager.SqlManager: Using default fetchSize of 1000
23/05/21 06:51:32 INFO tool.CodeGenTool: Beginning code generation
23/05/21 06:51:33 INFO manager.SqlManager: Executing SQL statement:
SELECT t.* FROM financial_regulation AS t WHERE 1=0
23/05/21 06:51:33 INFO manager.SqlManager: Executing SQL statement:
SELECT t.* FROM financial_regulation AS t WHERE 1=0
23/05/21 06:51:33 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /opt/cloudera/parcels/CDH/lib/hadoop-mapreduce
23/05/21 06:51:34 ERROR orm.CompilationManager: Could not rename /tmp/sqoop-
raghavendr48edu/compile/c82967a11cb29494d2a150e8a6f06ecc/financial_regulation.java to /home/raghavendr48edu/./financial_regulation.java. Error: Destination '/home/raghavendr48edu/./financial_regulation.java' already exists
23/05/21 06:51:34 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-
raghavendr48edu/compile/c82967a11cb29494d2a150e8a6f06ecc/financial_regulation.jar
23/05/21 06:51:35 INFO manager.SqlManager: Executing SQL statement:
SELECT t.* FROM financial_regulation AS t WHERE 1=0
23/05/21 06:51:35 INFO tool.ImportTool: Incremental import based on column time_stamp
23/05/21 06:51:35 INFO tool.ImportTool: Lower bound value: '2023-05-21 06:32:38'
23/05/21 06:51:35 INFO tool.ImportTool: Upper bound value: '2023-05-21 06:51:35.0'
23/05/21 06:51:35 INFO mapreduce.ImportJobBase: Beginning import of financial_regulation
23/05/21 06:51:36 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
23/05/21 06:51:36 INFO manager.SqlManager: Executing SQL statement:
SELECT t.* FROM financial_regulation AS t WHERE 1=0
23/05/21 06:51:36 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
23/05/21 06:51:36 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/21 06:51:36 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /user/raghavendr48edu/.staging/job_1684298513961_0928

23/05/21 06:51:39 INFO db.DBInputFormat: Using read committed transaction isolation
23/05/21 06:51:39 INFO mapreduce.JobSubmitter: number of splits:1
23/05/21 06:51:39 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
23/05/21 06:51:39 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1684298513961_0928
23/05/21 06:51:39 INFO mapreduce.JobSubmitter: Executing with tokens: []
23/05/21 06:51:39 INFO conf.Configuration: resource-types.xml not found
23/05/21 06:51:39 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
23/05/21 06:51:39 INFO impl.YarnClientImpl: Submitted application application_1684298513961_0928
23/05/21 06:51:39 INFO mapreduce.Job: The url to track the job: http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684298513961_0928/

23/05/21 06:51:39 INFO mapreduce.Job: Running job: job_1684298513961_0928
23/05/21 06:51:47 INFO mapreduce.Job: Job job_1684298513961_0928 running in uber mode : false
23/05/21 06:51:47 INFO mapreduce.Job: map 0% reduce 0%
23/05/21 06:52:02 INFO mapreduce.Job: map 100% reduce 0%
23/05/21 06:52:03 INFO mapreduce.Job: Job job_1684298513961_0928 completed successfully
23/05/21 06:52:03 INFO mapreduce.Job: Counters: 33
 File System Counters
 FILE: Number of bytes read=0
 FILE: Number of bytes written=247448
 FILE: Number of read operations=0
 FILE: Number of large read operations=0
 FILE: Number of write operations=0
 HDFS: Number of bytes read=85
 HDFS: Number of bytes written=33605184
 HDFS: Number of read operations=6
 HDFS: Number of large read operations=0
 HDFS: Number of write operations=2
 HDFS: Number of bytes read erasure-coded=0
 Job Counters
 Launched map tasks=1
 Other local map tasks=1
 Total time spent by all maps in occupied slots (ms)=12072
 Total time spent by all reduces in occupied slots (ms)=0
 Total time spent by all map tasks (ms)=12072
 Total vcore-milliseconds taken by all map tasks=12072
 Total megabyte-milliseconds taken by all map tasks=12361728
 Map-Reduce Framework
 Map input records=293796
 Map output records=293796
 Input split bytes=85
 Spilled Records=0

```

Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=1497
CPU time spent (ms)=10130
Physical memory (bytes) snapshot=498610176
Virtual memory (bytes) snapshot=2588659712
Total committed heap usage (bytes)=569901056
Peak Map Physical memory (bytes)=498610176
Peak Map Virtual memory (bytes)=2588659712
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=33605184
23/05/21 06:52:03 INFO mapreduce.ImportJobBase: Transferred 32.0484 MB in
27.0969 seconds (1.1827 MB/sec)
23/05/21 06:52:03 INFO mapreduce.ImportJobBase: Retrieved 293796 records.
23/05/21 06:52:03 INFO tool.ImportTool: Final destination exists, will run
merge job.
23/05/21 06:52:03 INFO tool.ImportTool: Moving data from temporary directory
_sqoop/972ebccb5f1442768d2441d6dc44fef7_financial_regulation to final
destination /user/raghavendr48edu/financial_regulation_updated
23/05/21 06:52:03 INFO tool.ImportTool: Incremental import complete! To run
another incremental import of all data following this import, supply the following
arguments:
23/05/21 06:52:03 INFO tool.ImportTool: --incremental lastmodified
23/05/21 06:52:03 INFO tool.ImportTool: --check-column time_stamp
23/05/21 06:52:03 INFO tool.ImportTool: --last-value 2023-05-21 06:51:35.0
23/05/21 06:52:03 INFO tool.ImportTool: (Consider saving this with 'sqoop job
--create')
[raghavendr48edu@ip-10-1-1-204 ~]$

```

From <<https://bigdata.cloudlabs.edureka.co/shell/>>

```

mysql> select count(*) from financial_regulation where SERIES="EQ" and updatetrades=0;
+-----+
| count(*) |
+-----+
|    293796    |
+-----+
1 row in set (0.58 sec)

```

Or

```

[raghavendr48edu@ip-10-1-1-204 ~]$ sqoop import --connect
jdbc:mysql://ip-10-1-1-204.ap-south-

```

1.compute.internal/raghavendr48edu --username raghavendr48edu --password RedKangaroo99\$ --table financial_regulation --where "SERIES='EQ'" --incremental append --check-column updatetrades --target-dir /user/raghavendr48edu/incremental_append/ --driver com.mysql.jdbc.Driver -m 1

Warning: /opt/cloudera/parcels/CDH-6.2.1-

1.cdh6.2.1.p0.1425774/bin/./lib/sqoop/./accumulo does not exist! Accumulo imports will fail.

Please set \$ACCUMULO_HOME to the root of your Accumulo installation.

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/opt/cloudera/parcels/CDH-6.2.1-

1.cdh6.2.1.p0.1425774/jars/slf4j-log4j12-

1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/opt/cloudera/parcels/CDH-6.2.1-

1.cdh6.2.1.p0.1425774/jars/log4j-slf4j-impl-

2.8.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]

23/05/21 11:36:00 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7-cdh6.2.1

23/05/21 11:36:00 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.

23/05/21 11:36:00 WARN sqoop.ConnFactory: Parameter --driver is set to an explicit driver however appropriate connection manager is not being set (via --connection-manager). Sqoop is going to fall back to org.

apache.sqoop.manager.GenericJdbcManager. Please specify explicitly which connection manager should be used next time.

23/05/21 11:36:00 INFO manager.SqlManager: Using default fetchSize of 1000

23/05/21 11:36:00 INFO tool.CodeGenTool: Beginning code generation

23/05/21 11:36:00 INFO manager.SqlManager: Executing SQL statement:

SELECT t.* FROM financial_regulation AS t WHERE 1=0

23/05/21 11:36:00 INFO manager.SqlManager: Executing SQL statement:

SELECT t.* FROM financial_regulation AS t WHERE 1=0

23/05/21 11:36:00 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /opt/cloudera/parcels/CDH/lib/hadoop-mapreduce

23/05/21 11:36:02 ERROR orm.CompilationManager: Could not rename

/tmp/sqoop-

raghavendr48edu/compile/41827c3d892387105201a4e66603ddd3/financial_regulation.java to /home/raghavendr48edu/./financial_regulation.java

va. Error: Destination '/home/raghavendr48edu/./financial_regulation.java' already exists

23/05/21 11:36:02 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-raghavendr48edu/compile/41827c3d892387105201a4e66603ddd3/financial_regulation.jar

23/05/21 11:36:02 INFO tool.ImportTool: Maximal id query for free form incremental import: SELECT MAX(updatetrades) FROM financial_regulation WHERE SERIES='EQ'

23/05/21 11:36:03 INFO tool.ImportTool: Incremental import based on column updatetrades

23/05/21 11:36:03 INFO tool.ImportTool: Upper bound value: 0

23/05/21 11:36:03 INFO mapreduce.ImportJobBase: Beginning import of financial_regulation
23/05/21 11:36:03 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
23/05/21 11:36:03 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM financial_regulation AS t WHERE 1=0
23/05/21 11:36:04 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
23/05/21 11:36:04 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/21 11:36:04 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /user/raghavendr48edu/.staging/job_1684298513961_0933
23/05/21 11:36:06 INFO db.DBInputFormat: Using read committed transaction isolation
23/05/21 11:36:06 INFO mapreduce.JobSubmitter: number of splits:1
23/05/21 11:36:06 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
23/05/21 11:36:06 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1684298513961_0933
23/05/21 11:36:06 INFO mapreduce.JobSubmitter: Executing with tokens: []
23/05/21 11:36:06 INFO conf.Configuration: resource-types.xml not found
23/05/21 11:36:06 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
23/05/21 11:36:06 INFO impl.YarnClientImpl: Submitted application application_1684298513961_0933
23/05/21 11:36:07 INFO mapreduce.Job: The url to track the job: http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684298513961_0933/

23/05/21 11:36:07 INFO mapreduce.Job: Running job: job_1684298513961_0933
23/05/21 11:36:15 INFO mapreduce.Job: Job job_1684298513961_0933 running in uber mode : false
23/05/21 11:36:15 INFO mapreduce.Job: map 0% reduce 0%
23/05/21 11:36:25 INFO mapreduce.Job: map 100% reduce 0%
23/05/21 11:36:25 INFO mapreduce.Job: Job job_1684298513961_0933 completed successfully
23/05/21 11:36:25 INFO mapreduce.Job: Counters: 33
 File System Counters
 FILE: Number of bytes read=0
 FILE: Number of bytes written=247289
 FILE: Number of read operations=0
 FILE: Number of large read operations=0
 FILE: Number of write operations=0
 HDFS: Number of bytes read=85
 HDFS: Number of bytes written=33605184
 HDFS: Number of read operations=6
 HDFS: Number of large read operations=0
 HDFS: Number of write operations=2
 HDFS: Number of bytes read erasure-coded=0
 Job Counters

```
Launched map tasks=1
Other local map tasks=1
Total time spent by all maps in occupied slots (ms)=8272
Total time spent by all reduces in occupied slots (ms)=0
Total time spent by all map tasks (ms)=8272
Total vcore-milliseconds taken by all map tasks=8272
Total megabyte-milliseconds taken by all map tasks=8470528
Map-Reduce Framework
  Map input records=293796
  Map output records=293796
  Input split bytes=85
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=795
  CPU time spent (ms)=9130
  Physical memory (bytes) snapshot=522072064
  Virtual memory (bytes) snapshot=2602045440
  Total committed heap usage (bytes)=780664832
  Peak Map Physical memory (bytes)=522072064
  Peak Map Virtual memory (bytes)=2602045440
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=33605184
23/05/21 11:36:25 INFO mapreduce.ImportJobBase: Transferred 32.0484 MB in
21.1136 seconds (1.5179 MB/sec)
23/05/21 11:36:25 INFO mapreduce.ImportJobBase: Retrieved 293796 records.
23/05/21 11:36:25 INFO util.AppendUtils: Creating missing output directory -
incremental_append
23/05/21 11:36:25 INFO tool.ImportTool: Incremental import complete! To run
another incremental import of all data following this import, supply the following
arguments:
23/05/21 11:36:25 INFO tool.ImportTool: --incremental append
23/05/21 11:36:25 INFO tool.ImportTool: --check-column updatetrades
23/05/21 11:36:25 INFO tool.ImportTool: --last-value 0
23/05/21 11:36:25 INFO tool.ImportTool: (Consider saving this with 'sqoop job
--create')
[raghavendr48edu@ip-10-1-1-204 ~]$
```

From <<https://bigdata.cloudlabs.edureka.co/shell/>>