

Submitted by: Brittny Hopwood

## Lab 5

**NOTE: THIS NOTEBOOK WILL TAKE 5-10 MINUTES TO COMPLETE.**

**PLEASE BE PATIENT.**

# Analyze Data Quality with SageMaker Processing Jobs and Spark

Typically a machine learning (ML) process consists of few steps. First, gathering data with various ETL jobs, then pre-processing the data, featurizing the dataset by incorporating standard techniques or prior knowledge, and finally training an ML model using an algorithm.

Often, distributed data processing frameworks such as Spark are used to process and analyze data sets in order to detect data quality issues and prepare them for model training.

In this notebook we'll use Amazon SageMaker Processing with a library called **Deequ**, and leverage the power of Spark with a managed SageMaker Processing Job to run our data processing workloads.

Here are some great resources on Deequ:

- Blog Post: <https://aws.amazon.com/blogs/big-data/test-data-quality-at-scale-with-deequ/>
- Research Paper: <https://assets.amazon.science/4a/75/57047bd343fab46ec14b34cdb3b/towards-automated-data-quality-management-for-machine-learning.pdf>



## Amazon Customer Reviews Dataset

<https://s3.amazonaws.com/dsoaws/amazon-reviews-pds/readme.html>

## Dataset Columns:

- `marketplace` : 2-letter country code (in this case all "US").
- `customer_id` : Random identifier that can be used to aggregate reviews written by a single author.
- `review_id` : A unique ID for the review.
- `product_id` : The Amazon Standard Identification Number (ASIN).  
`http://www.amazon.com/dp/<ASIN>` links to the product's detail page.
- `product_parent` : The parent of that ASIN. Multiple ASINs (color or format variations of the same product) can roll up into a single parent.
- `product_title` : Title description of the product.
- `product_category` : Broad product category that can be used to group reviews (in this case digital videos).
- `star_rating` : The review's rating (1 to 5 stars).
- `helpful_votes` : Number of helpful votes for the review.
- `total_votes` : Number of total votes the review received.
- `vine` : Was the review written as part of the [Vine](#) program?
- `verified_purchase` : Was the review from a verified purchase?
- `review_headline` : The title of the review itself.
- `review_body` : The text of the review.
- `review_date` : The date the review was written.

```
In [ ]: %store -r ingest_create_athena_table_tsv_passed
```

```
In [ ]: try:
        ingest_create_athena_table_tsv_passed
    except NameError:
        print("+++++")
        print("[ERROR] YOU HAVE TO RUN THE NOTEBOOKS IN THE INGEST FOLDER FIRST. You di
        print("+++++")
```

```
In [ ]: print(ingest_create_athena_table_tsv_passed)
```

True

```
In [ ]: if not ingest_create_athena_table_tsv_passed:
        print("+++++")
        print("[ERROR] YOU HAVE TO RUN THE NOTEBOOKS IN THE INGEST FOLDER FIRST. You di
        print("+++++")
    else:
        print("[OK]")
```

[OK]

# Run the Analysis Job using a SageMaker Processing Job with Spark

Next, use the Amazon SageMaker Python SDK to submit a processing job. Use the Spark container that was just built with our Spark script.

```
In [ ]: import sagemaker
import boto3

sess = sagemaker.Session()
bucket = sess.default_bucket()
role = sagemaker.get_execution_role()
region = boto3.Session().region_name
```

```
sagemaker.config INFO - Not applying SDK defaults from location: /etc/xdg/sagemaker/
config.yaml
sagemaker.config INFO - Not applying SDK defaults from location: /home/sagemaker-use
r/.config/sagemaker/config.yaml
sagemaker.config INFO - Not applying SDK defaults from location: /home/sagemaker-use
r/.config/sagemaker/config.yaml
```

## Review the Spark preprocessing script.

```
In [ ]: !pygmentize preprocess-deequ-pyspark.py
```

```

from __future__ import print_function
from __future__ import unicode_literals

import time
import sys
import os
import shutil
import csv
import subprocess

subprocess.check_call([sys.executable, "-m", "pip", "install", "--no-deps", "pydeequ
==0.1.5"])
subprocess.check_call([sys.executable, "-m", "pip", "install", "pandas==1.1.4"])

import pyspark
from pyspark.sql import SparkSession
from pyspark.sql.types import StructField, StructType, StringType, IntegerType, DoubleType
from pyspark.sql.functions import *

from pydeequ.analyzers import *
from pydeequ.checks import *
from pydeequ.verifications import *
from pydeequ.suggestions import *

# PySpark Deequ GitHub Repo: https://github.com/aws-labs/python-deequ

def main():
    args_iter = iter(sys.argv[1:])
    args = dict(zip(args_iter, args_iter))

    # Retrieve the args and replace 's3://' with 's3a://' (used by Spark)
    s3_input_data = args["s3_input_data"].replace("s3://", "s3a://")
    print(s3_input_data)
    s3_output_analyze_data = args["s3_output_analyze_data"].replace("s3://", "s3
a://")
    print(s3_output_analyze_data)

    spark = SparkSession.builder.appName("PySparkAmazonReviewsAnalyzer").getOrCreate()

    schema = StructType(
        [
            StructField("marketplace", StringType(), True),
            StructField("customer_id", StringType(), True),
            StructField("review_id", StringType(), True),
            StructField("product_id", StringType(), True),
            StructField("product_parent", StringType(), True),
            StructField("product_title", StringType(), True),
            StructField("product_category", StringType(), True),
            StructField("star_rating", IntegerType(), True),
            StructField("helpful_votes", IntegerType(), True),
            StructField("total_votes", IntegerType(), True),
            StructField("vine", StringType(), True),
            StructField("verified_purchase", StringType(), True),
        ]
    )

```

```

        StructField("review_headline", StringType(), True),
        StructField("review_body", StringType(), True),
        StructField("review_date", StringType(), True),
    ]
)

dataset = spark.read.csv(s3_input_data, header=True, schema=schema, sep="\t", quote="")

# Calculate statistics on the dataset
analysisResult = (
    AnalysisRunner(spark)
    .onData(dataset)
    .addAnalyzer(Size())
    .addAnalyzer(Completeness("review_id"))
    .addAnalyzer(ApproxCountDistinct("review_id"))
    .addAnalyzer(Mean("star_rating"))
    .addAnalyzer(Compliance("top_star_rating", "star_rating >= 4.0"))
    .addAnalyzer(Correlation("total_votes", "star_rating"))
    .addAnalyzer(Correlation("total_votes", "helpful_votes"))
    .run()
)

metrics = AnalyzerContext.successMetricsAsDataFrame(spark, analysisResult)
metrics.show(truncate=False)
metrics.repartition(1).write.format("csv").mode("overwrite").option("header", True).option("sep", "\t").save(
    "{}/{dataset-metrics}".format(s3_output_analyze_data)
)

# Check data quality
verificationResult = (
    VerificationSuite(spark)
    .onData(dataset)
    .addCheck(
        Check(spark, CheckLevel.Error, "Review Check")
        .hasSize(lambda x: x >= 200000)
        .hasMin("star_rating", lambda x: x == 1.0)
        .hasMax("star_rating", lambda x: x == 5.0)
        .isComplete("review_id")
        .isUnique("review_id")
        .isComplete("marketplace")
        .isContainedIn("marketplace", ["US", "UK", "DE", "JP", "FR"])
    )
    .run()
)

print(f"Verification Run Status: {verificationResult.status}")
resultsDataFrame = VerificationResult.checkResultsAsDataFrame(spark, verificationResult)
resultsDataFrame.show(truncate=False)
resultsDataFrame.repartition(1).write.format("csv").mode("overwrite").option("header", True).option(
    "sep", "\t"
).save("{}/{constraint-checks}".format(s3_output_analyze_data))

```

```

verificationSuccessMetricsDataFrame = VerificationResult.successMetricsAsDataFra
me(spark, verificationResult)
verificationSuccessMetricsDataFrame.show(truncate=False)
verificationSuccessMetricsDataFrame.repartition(1).write.format("csv").mode("ove
rwrite").option(
    "header", True
).option("sep", "\t").save("{}success-metrics".format(s3_output_analyze_data))

# Suggest new checks and constraints
suggestionsResult = ConstraintSuggestionRunner(spark).onData(dataset).addConstra
intRule(DEFAULT()).run()

suggestions = suggestionsResult["constraint_suggestions"]
parallelizedSuggestions = spark.sparkContext.parallelize(suggestions)

suggestionsResultsDataFrame = spark.createDataFrame(parallelizedSuggestions)
suggestionsResultsDataFrame.show(truncate=False)
suggestionsResultsDataFrame.repartition(1).write.format("csv").mode("overwrit
e").option("header", True).option(
    "sep", "\t"
).save("{}constraint-suggestions".format(s3_output_analyze_data))

# spark.stop()

if __name__ == "__main__":
    main()

```

```
In [ ]: from sagemaker.spark.processing import PySparkProcessor
```

```

processor = PySparkProcessor(
    base_job_name="spark-amazon-reviews-analyzer",
    role=role,
    framework_version="2.4",
    instance_count=1,
    instance_type="ml.m5.2xlarge",
    max_runtime_in_seconds=300,
)

```

```
In [ ]: s3_input_data = "s3://{}/amazon-reviews-pds/tsv/".format(bucket)
print(s3_input_data)
```

```
s3://sagemaker-us-east-1-211125778552/amazon-reviews-pds/tsv/
```

```
In [ ]: !aws s3 ls $s3_input_data
```

```

2024-04-27 16:37:43  18997559 amazon_reviews_us_Digital_Software_v1_00.tsv.gz
2024-04-27 16:37:45  27442648 amazon_reviews_us_Digital_Video_Games_v1_00.tsv.gz
2024-04-27 16:37:47  12134676 amazon_reviews_us_Gift_Card_v1_00.tsv.gz

```

## Setup Output Data

```
In [ ]: from time import gmtime, strftime
```

```
timestamp_prefix = strftime("%Y-%m-%d-%H-%M-%S", gmtime())

output_prefix = "amazon-reviews-spark-analyzer-{}".format(timestamp_prefix)
processing_job_name = "amazon-reviews-spark-analyzer-{}".format(timestamp_prefix)

print("Processing job name: {}".format(processing_job_name))
```

Processing job name: amazon-reviews-spark-analyzer-2024-04-28-21-15-58

```
In [ ]: s3_output_analyze_data = "s3://{}/{}/output".format(bucket, output_prefix)

print(s3_output_analyze_data)
```

s3://sagemaker-us-east-1-211125778552/amazon-reviews-spark-analyzer-2024-04-28-21-15-58/output

## Start the Spark Processing Job

*Notes on Invoking from Lambda:*

- However, if we use the boto3 SDK (ie. with a Lambda), we need to copy the `preprocess.py` file to S3 and specify the everything include `--py-files`, etc.
- We would need to do the following before invoking the Lambda: `!aws s3 cp preprocess.py s3:///sagemaker/spark-preprocess-reviews-demo/code/preprocess.py`  
`!aws s3 cp preprocess.py s3:///sagemaker/spark-preprocess-reviews-demo/py_files/preprocess.py`
- Then reference the `s3://` above in the `--py-files`, etc.
- See Lambda example code in this same project for more details.

*Notes on not using ProcessingInput and Output:*

- Since Spark natively reads/writes from/to S3 using `s3a://`, we can avoid the copy required by `ProcessingInput` and `ProcessingOutput` (`FullyReplicated` or `ShardedByS3Key`) and just specify the S3 input and output buckets/prefixes."
- See <https://github.com/aws-labs/amazon-sagemaker-examples/issues/994> for issues related to using `/opt/ml/processing/input/` and `output/`
- If we use `ProcessingInput`, the data will be copied to each node (which we don't want in this case since Spark already handles this)

```
In [ ]: from sagemaker.processing import ProcessingOutput

processor.run(
    submit_app="preprocess-deequ-pyspark.py",
    submit_jars=["deequ-1.0.3-rc2.jar"],
    arguments=[
        "s3_input_data",
        s3_input_data,
        "s3_output_analyze_data",
        s3_output_analyze_data,
    ],
    logs=True,
```

```
wait=False,
)
```

INFO:sagemaker:Creating processing-job with name spark-amazon-reviews-analyzer-2024-04-28-21-15-58-216

```
In [ ]: from IPython.core.display import display, HTML

processing_job_name = processor.jobs[-1].describe()["ProcessingJobName"]

display(
    HTML(
        '<b>Review <a target="blank" href="https://console.aws.amazon.com/sagemaker'
        region, processing_job_name
    )
)
```

/tmp/ipykernel\_23549/1518584736.py:1: DeprecationWarning: Importing display from IPython.core.display is deprecated since IPython 7.14, please import from IPython display  
 from IPython.core.display import display, HTML

### Review Processing Job

```
In [ ]: from IPython.core.display import display, HTML

processing_job_name = processor.jobs[-1].describe()["ProcessingJobName"]

display(
    HTML(
        '<b>Review <a target="blank" href="https://console.aws.amazon.com/cloudwatch'
        region, processing_job_name
    )
)
```

/tmp/ipykernel\_23549/3361610660.py:1: DeprecationWarning: Importing display from IPython.core.display is deprecated since IPython 7.14, please import from IPython display  
 from IPython.core.display import display, HTML

### Review CloudWatch Logs After a Few Minutes

```
In [ ]: from IPython.core.display import display, HTML

s3_job_output_prefix = output_prefix

display(
    HTML(
        '<b>Review <a target="blank" href="https://s3.console.aws.amazon.com/s3/buc'
        bucket, s3_job_output_prefix, region
    )
)
```



```
/tmp/ipykernel_23549/2348046881.py:1: DeprecationWarning: Importing display from IPython.core.display is deprecated since IPython 7.14, please import from IPython display
from IPython.core.display import display, HTML
```

Review [S3 Output Data After The Spark Job Has Completed](#)

## Monitor the Processing Job

```
In [ ]: running_processor = sagemaker.processing.ProcessingJob.from_processing_name(
        processing_job_name=processing_job_name, sagemaker_session=sess
    )

processing_job_description = running_processor.describe()

print(processing_job_description)
```

```
{'ProcessingInputs': [{'InputName': 'jars', 'AppManaged': False, 'S3Input': {'S3Uri': 's3://sagemaker-us-east-1-211125778552/spark-amazon-reviews-analyzer-2024-04-28-21-15-58-216/input/jars', 'LocalPath': '/opt/ml/processing/input/jars', 'S3DataType': 'S3Prefix', 'S3InputMode': 'File', 'S3DataDistributionType': 'FullyReplicated', 'S3CompressionType': 'None'}}, {'InputName': 'code', 'AppManaged': False, 'S3Input': {'S3Uri': 's3://sagemaker-us-east-1-211125778552/spark-amazon-reviews-analyzer-2024-04-28-21-15-58-216/input/code/preprocess-deequ-pyspark.py', 'LocalPath': '/opt/ml/processing/input/code', 'S3DataType': 'S3Prefix', 'S3InputMode': 'File', 'S3DataDistributionType': 'FullyReplicated', 'S3CompressionType': 'None'}}], 'ProcessingJobName': 'spark-amazon-reviews-analyzer-2024-04-28-21-15-58-216', 'ProcessingResources': {'ClusterConfig': {'InstanceCount': 1, 'InstanceType': 'ml.m5.2xlarge', 'VolumeSizeInGB': 30}}, 'StoppingCondition': {'MaxRuntimeInSeconds': 300}, 'AppSpecification': {'ImageUri': '173754725891.dkr.ecr.us-east-1.amazonaws.com/sagemaker-spark-processing:2.4-cpu', 'ContainerEntrypoint': ['smspark-submit', '--jars', '/opt/ml/processing/input/jars', '/opt/ml/processing/input/code/preprocess-deequ-pyspark.py'], 'ContainerArguments': ['s3_input_data', 's3://sagemaker-us-east-1-211125778552/amazon-reviews-pds/tsv/', 's3_output_analyze_data', 's3://sagemaker-us-east-1-211125778552/amazon-reviews-spark-analyzer-2024-04-28-21-15-58/output']}, 'Environment': {}, 'RoleArn': 'arn:aws:iam::211125778552:role/service-role/AmazonSageMaker-ExecutionRole-20240215T152311', 'ProcessingJobArn': 'arn:aws:sagemaker:us-east-1:211125778552:processing-job/spark-amazon-reviews-analyzer-2024-04-28-21-15-58-216', 'ProcessingJobStatus': 'InProgress', 'LastModifiedTime': datetime.datetime(2024, 4, 28, 21, 15, 59, 354000, tzinfo=tzlocal()), 'CreationTime': datetime.datetime(2024, 4, 28, 21, 15, 58, 982000, tzinfo=tzlocal()), 'ResponseMetadata': {'RequestId': 'a34290ff-9931-47eb-9088-58a4117ec59d', 'HTTPStatusCode': 200, 'HTTPHeaders': {'x-amzn-requestid': 'a34290ff-9931-47eb-9088-58a4117ec59d', 'content-type': 'application/x-amz-json-1.1', 'content-length': '1943', 'date': 'Sun, 28 Apr 2024 21:15:59 GMT'}, 'RetryAttempts': 0}}
```

```
In [ ]: running_processor.wait()
```

```

.....04-28 21:20 smspark.cli INFO Parsing arguments. ar
gv: ['/usr/local/bin/smspark-submit', '--jars', '/opt/ml/processing/input/jars', '/o
pt/ml/processing/input/code/preprocess-deequ-pyspark.py', 's3_input_data', 's3://sag
emaker-us-east-1-211125778552/amazon-reviews-pds/tsv/', 's3_output_analyze_data', 's
3://sagemaker-us-east-1-211125778552/amazon-reviews-spark-analyzer-2024-04-28-21-15-
58/output']
04-28 21:20 smspark.cli INFO Raw spark options before processing: {'jars': '/op
t/ml/processing/input/jars', 'class_': None, 'py_files': None, 'files': None, 'verbo
se': False}
04-28 21:20 smspark.cli INFO App and app arguments: ['/opt/ml/processing/input/
code/preprocess-deequ-pyspark.py', 's3_input_data', 's3://sagemaker-us-east-1-211125
778552/amazon-reviews-pds/tsv/', 's3_output_analyze_data', 's3://sagemaker-us-east-1
-211125778552/amazon-reviews-spark-analyzer-2024-04-28-21-15-58/output']
04-28 21:20 smspark.cli INFO Rendered spark options: {'jars': '/opt/ml/processi
ng/input/jars/deequ-1.0.3-rc2.jar', 'class_': None, 'py_files': None, 'files': None,
'verbos': False}
04-28 21:20 smspark.cli INFO Initializing processing job.
04-28 21:20 smspark-submit INFO {'current_host': 'algo-1', 'hosts': ['algo-1']}
04-28 21:20 smspark-submit INFO {'ProcessingJobArn': 'arn:aws:sagemaker:us-east-
1:211125778552:processing-job/spark-amazon-reviews-analyzer-2024-04-28-21-15-58-21
6', 'ProcessingJobName': 'spark-amazon-reviews-analyzer-2024-04-28-21-15-58-216', 'A
ppSpecification': {'ImageUri': '173754725891.dkr.ecr.us-east-1.amazonaws.com/sagemak
er-spark-processing:2.4-cpu', 'ContainerEntrypoint': ['smspark-submit', '--jars', '/
opt/ml/processing/input/jars', '/opt/ml/processing/input/code/preprocess-deequ-pyspa
rk.py'], 'ContainerArguments': ['s3_input_data', 's3://sagemaker-us-east-1-211125778
552/amazon-reviews-pds/tsv/', 's3_output_analyze_data', 's3://sagemaker-us-east-1-21
1125778552/amazon-reviews-spark-analyzer-2024-04-28-21-15-58/output']}, 'ProcessingI
nputs': [{'InputName': 'jars', 'AppManaged': False, 'S3Input': {'LocalPath': '/opt/m
l/processing/input/jars', 'S3Uri': 's3://sagemaker-us-east-1-211125778552/spark-amaz
on-reviews-analyzer-2024-04-28-21-15-58-216/input/jars', 'S3DataDistributionType':
'FullyReplicated', 'S3DataType': 'S3Prefix', 'S3InputMode': 'File', 'S3CompressionTy
pe': 'None', 'S3DownloadMode': 'StartOfJob'}, 'DatasetDefinition': None}, {'InputNam
e': 'code', 'AppManaged': False, 'S3Input': {'LocalPath': '/opt/ml/processing/input/
code', 'S3Uri': 's3://sagemaker-us-east-1-211125778552/spark-amazon-reviews-analyzer
-2024-04-28-21-15-58-216/input/code/preprocess-deequ-pyspark.py', 'S3DataDistributio
nType': 'FullyReplicated', 'S3DataType': 'S3Prefix', 'S3InputMode': 'File', 'S3Compr
essionType': 'None', 'S3DownloadMode': 'StartOfJob'}, 'DatasetDefinition': None}],
'ProcessingOutputConfig': {'Outputs': [], 'KmsKeyId': None}, 'ProcessingResources':
{'ClusterConfig': {'InstanceCount': 1, 'InstanceType': 'ml.m5.2xlarge', 'VolumeSizeI
nGB': 30, 'VolumeKmsKeyId': None}}, 'RoleArn': 'arn:aws:iam::211125778552:role/servi
ce-role/AmazonSageMaker-ExecutionRole-20240215T152311', 'StoppingCondition': {'MaxRu
ntimeInSeconds': 300}}
04-28 21:20 smspark.cli INFO running spark submit command: spark-submit --maste
r yarn --deploy-mode client --jars /opt/ml/processing/input/jars/deequ-1.0.3-rc2.jar
/opt/ml/processing/input/code/preprocess-deequ-pyspark.py s3_input_data s3://sagemak
er-us-east-1-211125778552/amazon-reviews-pds/tsv/ s3_output_analyze_data s3://sagema
ker-us-east-1-211125778552/amazon-reviews-spark-analyzer-2024-04-28-21-15-58/output
04-28 21:20 smspark-submit INFO waiting for hosts
04-28 21:20 smspark-submit INFO starting status server
04-28 21:20 smspark-submit INFO Status server listening on algo-1:5555
04-28 21:20 smspark-submit INFO bootstrapping cluster
04-28 21:20 smspark-submit INFO transitioning from status INITIALIZING to BOOTST
RAPPING
04-28 21:20 smspark-submit INFO copying aws jars
Serving on http://algo-1:5555
04-28 21:20 smspark-submit INFO Found hadoop jar hadoop-aws.jar

```

```

04-28 21:20 smspark-submit INFO      Copying optional jar jets3t-0.9.0.jar from /usr/
lib/hadoop/lib to /usr/lib/spark/jars
04-28 21:20 smspark-submit INFO      copying cluster config
04-28 21:20 smspark-submit INFO      copying /opt/hadoop-config/hdfs-site.xml to /us
r/lib/hadoop/etc/hadoop/hdfs-site.xml
04-28 21:20 smspark-submit INFO      copying /opt/hadoop-config/core-site.xml to /us
r/lib/hadoop/etc/hadoop/core-site.xml
04-28 21:20 smspark-submit INFO      copying /opt/hadoop-config/yarn-site.xml to /us
r/lib/hadoop/etc/hadoop/yarn-site.xml
04-28 21:20 smspark-submit INFO      copying /opt/hadoop-config/spark-defaults.conf t
o /usr/lib/spark/conf/spark-defaults.conf
04-28 21:20 smspark-submit INFO      copying /opt/hadoop-config/spark-env.sh to /usr/
lib/spark/conf/spark-env.sh
04-28 21:20 root          INFO      Detected instance type: m5.2xlarge with total memo
ry: 32768M and total cores: 8
04-28 21:20 root          INFO      Writing default config to /usr/lib/hadoop/etc/hado
op/yarn-site.xml
04-28 21:20 root          INFO      Configuration at /usr/lib/hadoop/etc/hadoop/yarn-s
ite.xml is:
<?xml version="1.0"?>
<!-- Site specific YARN configuration properties -->
<configuration>
  <property>
    <name>yarn.resourcemanager.hostname</name>
    <value>10.0.222.169</value>
    <description>The hostname of the RM.</description>
  </property>
  <property>
    <name>yarn.nodemanager.hostname</name>
    <value>algo-1</value>
    <description>The hostname of the NM.</description>
  </property>
  <property>
    <name>yarn.nodemanager.webapp.address</name>
    <value>algo-1:8042</value>
  </property>
  <property>
    <name>yarn.nodemanager.vmem-pmem-ratio</name>
    <value>5</value>
    <description>Ratio between virtual memory to physical memory.</description>
  </property>
  <property>
    <name>yarn.resourcemanager.am.max-attempts</name>
    <value>1</value>
    <description>The maximum number of application attempts.</description>
  </property>
  <property>
    <name>yarn.nodemanager.env-whitelist</name>
    <value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,YARN_H
OME,AWS_CONTAINER_CREDENTIALS_RELATIVE_URI</value>
    <description>Environment variable whitelist</description>
  </property>

  <property>
    <name>yarn.scheduler.minimum-allocation-mb</name>
    <value>1</value>

```

```

</property>
<property>
  <name>yarn.scheduler.maximum-allocation-mb</name>
  <value>31784</value>
</property>
<property>
  <name>yarn.scheduler.minimum-allocation-vcores</name>
  <value>1</value>
</property>
<property>
  <name>yarn.scheduler.maximum-allocation-vcores</name>
  <value>8</value>
</property>
<property>
  <name>yarn.nodemanager.resource.memory-mb</name>
  <value>31784</value>
</property>
<property>
  <name>yarn.nodemanager.resource.cpu-vcores</name>
  <value>8</value>
</property>
</configuration>
04-28 21:20 root      INFO      Writing default config to /usr/lib/spark/conf/spark-
k-defaults.conf
04-28 21:20 root      INFO      Configuration at /usr/lib/spark/conf/spark-default
s.conf is:
spark.driver.extraClassPath    /usr/lib/hadoop-lzo/lib/*:/usr/lib/hadoop/hadoop-aw
s.jar:/usr/share/aws/aws-java-sdk/*:/usr/share/aws/emr/emrfs/conf:/usr/share/aws/em
r/emrfs/lib/*:/usr/share/aws/emr/emrfs/auxlib/*:/usr/share/aws/emr/goodies/lib/emr-s
park-goodies.jar:/usr/share/aws/emr/security/conf:/usr/share/aws/emr/security/lib/
*:/usr/share/aws/hmclient/lib/aws-glue-datacatalog-spark-client.jar:/usr/share/java/
Hive-JSON-Serde/hive-openx-serde.jar:/usr/share/aws/sagemaker-spark-sdk/lib/sagemake
r-spark-sdk.jar:/usr/share/aws/emr/s3select/lib/emr-s3-select-spark-connector.jar
spark.driver.extraLibraryPath  /usr/lib/hadoop/lib/native:/usr/lib/hadoop-lzo/lib/
native
spark.executor.extraClassPath  /usr/lib/hadoop-lzo/lib/*:/usr/lib/hadoop/hadoop-aw
s.jar:/usr/share/aws/aws-java-sdk/*:/usr/share/aws/emr/emrfs/conf:/usr/share/aws/em
r/emrfs/lib/*:/usr/share/aws/emr/emrfs/auxlib/*:/usr/share/aws/emr/goodies/lib/emr-s
park-goodies.jar:/usr/share/aws/emr/security/conf:/usr/share/aws/emr/security/lib/
*:/usr/share/aws/hmclient/lib/aws-glue-datacatalog-spark-client.jar:/usr/share/java/
Hive-JSON-Serde/hive-openx-serde.jar:/usr/share/aws/sagemaker-spark-sdk/lib/sagemake
r-spark-sdk.jar:/usr/share/aws/emr/s3select/lib/emr-s3-select-spark-connector.jar
spark.executor.extraLibraryPath /usr/lib/hadoop/lib/native:/usr/lib/hadoop-lzo/lib/
native
spark.driver.host=10.0.222.169
spark.hadoop.mapreduce.fileoutputcommitter.algorithm.version=2
spark.driver.memory 2048m
spark.driver.memoryOverhead 204m
spark.driver.defaultJavaOptions -XX:OnOutOfMemoryError='kill -9 %p' -XX:+UseConcMark
SweepGC -XX:CMSInitiatingOccupancyFraction=70 -XX:MaxHeapFreeRatio=70 -XX:+CMSClassU
nloadingEnabled
spark.executor.memory 26847m
spark.executor.memoryOverhead 2684m
spark.executor.cores 8
spark.executor.defaultJavaOptions -verbose:gc -XX:OnOutOfMemoryError='kill -9 %p' -X
X:+PrintGCDetails -XX:+PrintGCDateStamps -XX:+UseParallelGC -XX:InitiatingHeapOccupa

```

```

ncyPercent=70 -XX:ConcGCTThreads=2 -XX:ParallelGCTThreads=6
spark.executor.instances 1
spark.default.parallelism 16
04-28 21:20 root          INFO      Finished Yarn configuration files setup.
04-28 21:20 root          INFO      No file at /opt/ml/processing/input/conf/configura
tion.json exists, skipping user configuration
24/04/28 21:20:46 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG:  host = algo-1/10.0.222.169
STARTUP_MSG:  args = [-format, -force]
STARTUP_MSG:  version = 2.10.0-amzn-0
STARTUP_MSG:  classpath = /usr/lib/hadoop/etc/hadoop:/usr/lib/hadoop/lib/audience-a
nnotations-0.5.0.jar:/usr/lib/hadoop/lib/jackson-jaxrs-1.9.13.jar:/usr/lib/hadoop/li
b/commons-collections-3.2.2.jar:/usr/lib/hadoop/lib/mockito-all-1.8.5.jar:/usr/lib/h
adoop/lib/spotbugs-annotations-3.1.9.jar:/usr/lib/hadoop/lib/stax2-api-3.1.4.jar:/us
r/lib/hadoop/lib/curator-client-2.7.1.jar:/usr/lib/hadoop/lib/commons-lang3-3.4.ja
r:/usr/lib/hadoop/lib/jcip-annotations-1.0-1.jar:/usr/lib/hadoop/lib/jetty-util-6.1.
26-emr.jar:/usr/lib/hadoop/lib/jaxb-api-2.2.2.jar:/usr/lib/hadoop/lib/asm-3.2.jar:/u
sr/lib/hadoop/lib/jettison-1.1.jar:/usr/lib/hadoop/lib/paranamer-2.3.jar:/usr/lib/ha
dooop/lib/java-xmlbuilder-0.4.jar:/usr/lib/hadoop/lib/jaxb-impl-2.2.3-1.jar:/usr/lib/
hadoop/lib/avro-1.7.7.jar:/usr/lib/hadoop/lib/curator-framework-2.7.1.jar:/usr/lib/h
adoop/lib/jersey-json-1.9.jar:/usr/lib/hadoop/lib/curator-recipes-2.7.1.jar:/usr/li
b/hadoop/lib/log4j-1.2.17.jar:/usr/lib/hadoop/lib/apacheds-kerberos-codec-2.0.0-M15.
jar:/usr/lib/hadoop/lib/jets3t-0.9.0.jar:/usr/lib/hadoop/lib/jackson-xc-1.9.13.jar:/
usr/lib/hadoop/lib/httpcore-4.4.11.jar:/usr/lib/hadoop/lib/guava-11.0.2.jar:/usr/li
b/hadoop/lib/jsp-api-2.1.jar:/usr/lib/hadoop/lib/xmlenc-0.52.jar:/usr/lib/hadoop/li
b/servlet-api-2.5.jar:/usr/lib/hadoop/lib/commons-codec-1.4.jar:/usr/lib/hadoop/lib/
slf4j-api-1.7.25.jar:/usr/lib/hadoop/lib/api-util-1.0.0-M20.jar:/usr/lib/hadoop/lib/
commons-math3-3.1.1.jar:/usr/lib/hadoop/lib/activation-1.1.jar:/usr/lib/hadoop/lib/z
ookeeper-3.4.14.jar:/usr/lib/hadoop/lib/stax-api-1.0-2.jar:/usr/lib/hadoop/lib/commo
ns-configuration-1.6.jar:/usr/lib/hadoop/lib/jersey-core-1.9.jar:/usr/lib/hadoop/li
b/commons-beanutils-1.9.4.jar:/usr/lib/hadoop/lib/hamcrest-core-1.3.jar:/usr/lib/had
oop/lib/json-smart-1.3.1.jar:/usr/lib/hadoop/lib/commons-logging-1.1.3.jar:/usr/lib/
hadoop/lib/jetty-sslengine-6.1.26-emr.jar:/usr/lib/hadoop/lib/commons-compress-1.19.
jar:/usr/lib/hadoop/lib/commons-io-2.4.jar:/usr/lib/hadoop/lib/jackson-core-asl-1.9.
13.jar:/usr/lib/hadoop/lib/httpclient-4.5.9.jar:/usr/lib/hadoop/lib/netty-3.10.6.Fin
al.jar:/usr/lib/hadoop/lib/commons-lang-2.6.jar:/usr/lib/hadoop/lib/jetty-6.1.26-em
r.jar:/usr/lib/hadoop/lib/jackson-mapper-asl-1.9.13.jar:/usr/lib/hadoop/lib/commons-
digester-1.8.jar:/usr/lib/hadoop/lib/protobuf-java-2.5.0.jar:/usr/lib/hadoop/lib/jsc
h-0.1.54.jar:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.25.jar:/usr/lib/hadoop/lib/jersey
-server-1.9.jar:/usr/lib/hadoop/lib/commons-cli-1.2.jar:/usr/lib/hadoop/lib/apacheds
-i18n-2.0.0-M15.jar:/usr/lib/hadoop/lib/woodstox-core-5.0.3.jar:/usr/lib/hadoop/lib/
htrace-core4-4.1.0-incubating.jar:/usr/lib/hadoop/lib/commons-net-3.1.jar:/usr/lib/h
adoop/lib/snappy-java-1.1.7.3.jar:/usr/lib/hadoop/lib/api-asn1-api-1.0.0-M20.jar:/us
r/lib/hadoop/lib/gson-2.2.4.jar:/usr/lib/hadoop/lib/nimbus-jose-jwt-4.41.1.jar:/usr/
lib/hadoop/lib/jsr305-3.0.0.jar:/usr/lib/hadoop/lib/junit-4.11.jar:/usr/lib/hadoo
p/./hadoop-openstack-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-azure-datalake.ja
r:/usr/lib/hadoop/./hadoop-distcp-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-aliyu
n.jar:/usr/lib/hadoop/./hadoop-ant-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-azur
e.jar:/usr/lib/hadoop/./hadoop-nfs-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-aut
h.jar:/usr/lib/hadoop/./hadoop-yarn-common-2.10.0-amzn-0.jar:/usr/lib/hadoop/./had
oop-aliyun-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-yarn-server-resourcemanager.j
ar:/usr/lib/hadoop/./hadoop-aws.jar:/usr/lib/hadoop/./hadoop-azure-datalake-2.10.0
-amzn-0.jar:/usr/lib/hadoop/./hadoop-rumen.jar:/usr/lib/hadoop/./hadoop-extras.ja
r:/usr/lib/hadoop/./hadoop-yarn-server-applicationhistoryservice-2.10.0-amzn-0.ja

```

r:/usr/lib/hadoop/./hadoop-annotations-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-archives.jar:/usr/lib/hadoop/./hadoop-gridmix-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-auth-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-yarn-server-resourcemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-yarn-server-applicationhistoryservice.jar:/usr/lib/hadoop/./hadoop-annotations.jar:/usr/lib/hadoop/./hadoop-yarn-registry-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-extras-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-yarn-common.jar:/usr/lib/hadoop/./hadoop-gridmix.jar:/usr/lib/hadoop/./hadoop-streaming-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-distcp.jar:/usr/lib/hadoop/./hadoop-yarn-server-common-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-yarn-server-web-proxy.jar:/usr/lib/hadoop/./hadoop-aws-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-common-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop/./hadoop-resourceestimator.jar:/usr/lib/hadoop/./hadoop-openstack.jar:/usr/lib/hadoop/./hadoop-yarn-server-web-proxy-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-archive-logs.jar:/usr/lib/hadoop/./hadoop-nfs.jar:/usr/lib/hadoop/./hadoop-yarn-registry.jar:/usr/lib/hadoop/./hadoop-common-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-streaming.jar:/usr/lib/hadoop/./hadoop-yarn-api-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-yarn-server-common.jar:/usr/lib/hadoop/./hadoop-datajoin-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-common.jar:/usr/lib/hadoop/./hadoop-sls.jar:/usr/lib/hadoop/./hadoop-rumen-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-azure-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-archives-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-ant.jar:/usr/lib/hadoop/./hadoop-sls-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-yarn-api.jar:/usr/lib/hadoop/./hadoop-datajoin.jar:/usr/lib/hadoop/./hadoop-archive-logs-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-resourceestimator-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./usr/lib/hadoop-hdfs/lib/jetty-util-6.1.26-emr.jar:/usr/lib/hadoop-hdfs/lib/asm-3.2.jar:/usr/lib/hadoop-hdfs/lib/okio-1.6.0.jar:/usr/lib/hadoop-hdfs/lib/commons-daemon-1.0.13.jar:/usr/lib/hadoop-hdfs/lib/log4j-1.2.17.jar:/usr/lib/hadoop-hdfs/lib/guava-11.0.2.jar:/usr/lib/hadoop-hdfs/lib/xmlenc-0.52.jar:/usr/lib/hadoop-hdfs/lib/xercesImpl-2.12.0.jar:/usr/lib/hadoop-hdfs/lib/servlet-api-2.5.jar:/usr/lib/hadoop-hdfs/lib/commons-codec-1.4.jar:/usr/lib/hadoop-hdfs/lib/xml-apis-1.4.01.jar:/usr/lib/hadoop-hdfs/lib/jersey-core-1.9.jar:/usr/lib/hadoop-hdfs/lib/jackson-core-2.6.7.jar:/usr/lib/hadoop-hdfs/lib/commons-logging-1.1.3.jar:/usr/lib/hadoop-hdfs/lib/commons-io-2.4.jar:/usr/lib/hadoop-hdfs/lib/jackson-core-asl-1.9.13.jar:/usr/lib/hadoop-hdfs/lib/netty-3.10.6.Final.jar:/usr/lib/hadoop-hdfs/lib/commons-lang-2.6.jar:/usr/lib/hadoop-hdfs/lib/jetty-6.1.26-emr.jar:/usr/lib/hadoop-hdfs/lib/jackson-annotations-2.6.7.jar:/usr/lib/hadoop-hdfs/lib/jackson-mapper-asl-1.9.13.jar:/usr/lib/hadoop-hdfs/lib/protobuf-java-2.5.0.jar:/usr/lib/hadoop-hdfs/lib/jersey-server-1.9.jar:/usr/lib/hadoop-hdfs/lib/commons-cli-1.2.jar:/usr/lib/hadoop-hdfs/lib/htrace-core4-4.1.0-incubating.jar:/usr/lib/hadoop-hdfs/lib/leveldbjni-all-1.8.jar:/usr/lib/hadoop-hdfs/lib/jackson-databind-2.6.7.jar:/usr/lib/hadoop-hdfs/lib/okhttp-2.7.5.jar:/usr/lib/hadoop-hdfs/lib/jsr305-3.0.0.jar:/usr/lib/hadoop-hdfs/lib/netty-all-4.0.23.Final.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-client-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-native-client.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-rbf-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-rbf-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-client-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-native-client-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-nfs.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-client.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-nfs-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-rbf.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-native-client-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop-yarn/lib/audience-annotations-0.5.0.jar:/usr/lib/hadoop-yarn/lib/javax.inject-1.jar:/usr/lib/hadoop-yarn/lib/jackson-jaxrs-1.9.13.jar:/usr/lib/hadoop-yarn/lib/commons-collections-3.2.2.jar:/usr/lib/hadoop-yarn/lib/spotbugs-annotations-3.1.9.jar:/usr/lib/hadoop-yarn/lib/stax2-api-3.1.4.jar:/usr/lib/hadoop-yarn/lib/curator-client-2.7.1.jar:/usr/lib/hadoop-yarn/lib/commons-lang3-3.4.jar:/usr/lib/hadoop-yarn/lib/jcip-annotations-1.0-1.jar:/usr/lib/hadoop-yarn/lib/jetty-util-6.1.26-emr.jar:/usr/lib/hadoop-yarn/lib/jaxb-api-2.2.2.jar:/u

sr/lib/hadoop-yarn/lib/jersey-client-1.9.jar:/usr/lib/hadoop-yarn/lib/asm-3.2.jar:/usr/lib/hadoop-yarn/lib/jettison-1.1.jar:/usr/lib/hadoop-yarn/lib/paranamer-2.3.jar:/usr/lib/hadoop-yarn/lib/java-xmlbuilder-0.4.jar:/usr/lib/hadoop-yarn/lib/jaxb-impl-2.2.3-1.jar:/usr/lib/hadoop-yarn/lib/avro-1.7.7.jar:/usr/lib/hadoop-yarn/lib/ehcache-3.3.1.jar:/usr/lib/hadoop-yarn/lib/aopalliance-1.0.jar:/usr/lib/hadoop-yarn/lib/curator-framework-2.7.1.jar:/usr/lib/hadoop-yarn/lib/jersey-json-1.9.jar:/usr/lib/hadoop-yarn/lib/curator-recipes-2.7.1.jar:/usr/lib/hadoop-yarn/lib/log4j-1.2.17.jar:/usr/lib/hadoop-yarn/lib/apacheds-kerberos-codec-2.0.0-M15.jar:/usr/lib/hadoop-yarn/lib/jets3t-0.9.0.jar:/usr/lib/hadoop-yarn/lib/jackson-xc-1.9.13.jar:/usr/lib/hadoop-yarn/lib/httpcore-4.4.11.jar:/usr/lib/hadoop-yarn/lib/guava-11.0.2.jar:/usr/lib/hadoop-yarn/lib/jsp-api-2.1.jar:/usr/lib/hadoop-yarn/lib/HikariCP-java7-2.4.12.jar:/usr/lib/hadoop-yarn/lib/xmlenc-0.52.jar:/usr/lib/hadoop-yarn/lib/servlet-api-2.5.jar:/usr/lib/hadoop-yarn/lib/commons-codec-1.4.jar:/usr/lib/hadoop-yarn/lib/api-util-1.0.0-M20.jar:/usr/lib/hadoop-yarn/lib/commons-math3-3.1.1.jar:/usr/lib/hadoop-yarn/lib/guice-servlet-3.0.jar:/usr/lib/hadoop-yarn/lib/activation-1.1.jar:/usr/lib/hadoop-yarn/lib/zookeeper-3.4.14.jar:/usr/lib/hadoop-yarn/lib/stax-api-1.0-2.jar:/usr/lib/hadoop-yarn/lib/commons-configuration-1.6.jar:/usr/lib/hadoop-yarn/lib/jersey-core-1.9.jar:/usr/lib/hadoop-yarn/lib/fst-2.50.jar:/usr/lib/hadoop-yarn/lib/commons-beanutils-1.9.4.jar:/usr/lib/hadoop-yarn/lib/json-smart-1.3.1.jar:/usr/lib/hadoop-yarn/lib/commons-logging-1.1.3.jar:/usr/lib/hadoop-yarn/lib/jetty-sslengine-6.1.26-emr.jar:/usr/lib/hadoop-yarn/lib/commons-compress-1.19.jar:/usr/lib/hadoop-yarn/lib/commons-io-2.4.jar:/usr/lib/hadoop-yarn/lib/jackson-core-asl-1.9.13.jar:/usr/lib/hadoop-yarn/lib/httpclient-4.5.9.jar:/usr/lib/hadoop-yarn/lib/netty-3.10.6.Final.jar:/usr/lib/hadoop-yarn/lib/commons-lang-2.6.jar:/usr/lib/hadoop-yarn/lib/jetty-6.1.26-emr.jar:/usr/lib/hadoop-yarn/lib/jackson-mapper-asl-1.9.13.jar:/usr/lib/hadoop-yarn/lib/geronimo-jcache-1.0\_spec-1.0-alpha-1.jar:/usr/lib/hadoop-yarn/lib/mssql-jdbc-6.2.1.jre7.jar:/usr/lib/hadoop-yarn/lib/commons-digester-1.8.jar:/usr/lib/hadoop-yarn/lib/protobuf-java-2.5.0.jar:/usr/lib/hadoop-yarn/lib/jsch-0.1.54.jar:/usr/lib/hadoop-yarn/lib/jersey-server-1.9.jar:/usr/lib/hadoop-yarn/lib/commons-cli-1.2.jar:/usr/lib/hadoop-yarn/lib/apacheds-i18n-2.0.0-M15.jar:/usr/lib/hadoop-yarn/lib/woodstox-core-5.0.3.jar:/usr/lib/hadoop-yarn/lib/htrace-core4-4.1.0-incubating.jar:/usr/lib/hadoop-yarn/lib/commons-net-3.1.jar:/usr/lib/hadoop-yarn/lib/json-io-2.5.1.jar:/usr/lib/hadoop-yarn/lib/jersey-guice-1.9.jar:/usr/lib/hadoop-yarn/lib/snappy-java-1.1.7.3.jar:/usr/lib/hadoop-yarn/lib/leveldbjni-all-1.8.jar:/usr/lib/hadoop-yarn/lib/api-asn1-api-1.0.0-M20.jar:/usr/lib/hadoop-yarn/lib/gson-2.2.4.jar:/usr/lib/hadoop-yarn/lib/nimbus-jose-jwt-4.41.1.jar:/usr/lib/hadoop-yarn/lib/jsr305-3.0.0.jar:/usr/lib/hadoop-yarn/lib/metrics-core-3.0.1.jar:/usr/lib/hadoop-yarn/lib/guice-3.0.jar:/usr/lib/hadoop-yarn/lib/java-util-1.9.0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-tests-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-nodemanager.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-client-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-timeline-pluginstorage-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-tests.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-applications-unmanaged-am-launcher.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-router-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-resourcemanager.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-applicationhistoryservice-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-resourcemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-applicationhistoryservice.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-applications-distributedshell-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-applications-unmanaged-am-launcher-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-registry-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-common.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-sharedcachemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-client.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-web-proxy.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-sharedcachemanager.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-web-proxy-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-router.jar:/usr/lib/hadoop-yarn

n/./hadoop-yarn-registry.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-api-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-timeline-pluginstorage.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-common.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-applications-distributedshell.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-api.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-nodemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/lib/javax.inject-1.jar:/usr/lib/hadoop-mapreduce/lib/asm-3.2.jar:/usr/lib/hadoop-mapreduce/lib/paranamer-2.3.jar:/usr/lib/hadoop-mapreduce/lib/avro-1.7.7.jar:/usr/lib/hadoop-mapreduce/lib/aopalliance-1.0.jar:/usr/lib/hadoop-mapreduce/lib/log4j-1.2.17.jar:/usr/lib/hadoop-mapreduce/lib/guice-servlet-3.0.jar:/usr/lib/hadoop-mapreduce/lib/jersey-core-1.9.jar:/usr/lib/hadoop-mapreduce/lib/hamcrest-core-1.3.jar:/usr/lib/hadoop-mapreduce/lib/commons-compress-1.19.jar:/usr/lib/hadoop-mapreduce/lib/commons-io-2.4.jar:/usr/lib/hadoop-mapreduce/lib/jackson-core-asl-1.9.13.jar:/usr/lib/hadoop-mapreduce/lib/netty-3.10.6.Final.jar:/usr/lib/hadoop-mapreduce/lib/jackson-mapper-asl-1.9.13.jar:/usr/lib/hadoop-mapreduce/lib/protobuf-java-2.5.0.jar:/usr/lib/hadoop-mapreduce/lib/jersey-server-1.9.jar:/usr/lib/hadoop-mapreduce/lib/jersey-guice-1.9.jar:/usr/lib/hadoop-mapreduce/lib/snappy-java-1.1.7.3.jar:/usr/lib/hadoop-mapreduce/lib/leveldbjni-all-1.8.jar:/usr/lib/hadoop-mapreduce/lib/junit-4.11.jar:/usr/lib/hadoop-mapreduce/lib/guice-3.0.jar:/usr/lib/hadoop-mapreduce/./audience-annotations-0.5.0.jar:/usr/lib/hadoop-mapreduce/./javax.inject-1.jar:/usr/lib/hadoop-mapreduce/./hadoop-openstack-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-azure-datalake.jar:/usr/lib/hadoop-mapreduce/./hadoop-distcp-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jackson-jaxrs-1.9.13.jar:/usr/lib/hadoop-mapreduce/./commons-collections-3.2.2.jar:/usr/lib/hadoop-mapreduce/./hadoop-aliyun.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-hs-plugins-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-ant-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./azure-storage-5.4.0.jar:/usr/lib/hadoop-mapreduce/./spotbugs-annotations-3.1.9.jar:/usr/lib/hadoop-mapreduce/./stax2-api-3.1.4.jar:/usr/lib/hadoop-mapreduce/./hadoop-azure.jar:/usr/lib/hadoop-mapreduce/./curator-client-2.7.1.jar:/usr/lib/hadoop-mapreduce/./commons-lang3-3.4.jar:/usr/lib/hadoop-mapreduce/./jcip-annotations-1.0-1.jar:/usr/lib/hadoop-mapreduce/./jetty-util-6.1.26-emr.jar:/usr/lib/hadoop-mapreduce/./jaxb-api-2.2.2.jar:/usr/lib/hadoop-mapreduce/./jersey-client-1.9.jar:/usr/lib/hadoop-mapreduce/./azure-keyvault-core-0.8.0.jar:/usr/lib/hadoop-mapreduce/./asm-3.2.jar:/usr/lib/hadoop-mapreduce/./jettison-1.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-auth.jar:/usr/lib/hadoop-mapreduce/./ojalgo-43.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-common.jar:/usr/lib/hadoop-mapreduce/./paranamer-2.3.jar:/usr/lib/hadoop-mapreduce/./java-xmlbuilder-0.4.jar:/usr/lib/hadoop-mapreduce/./hadoop-aliyun-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-resourcemanager.jar:/usr/lib/hadoop-mapreduce/./jaxb-impl-2.2.3-1.jar:/usr/lib/hadoop-mapreduce/./avro-1.7.7.jar:/usr/lib/hadoop-mapreduce/./aliyun-java-sdk-sts-3.0.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-aws.jar:/usr/lib/hadoop-mapreduce/./hadoop-azure-datalake-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-rumen.jar:/usr/lib/hadoop-mapreduce/./ehcache-3.3.1.jar:/usr/lib/hadoop-mapreduce/./aopalliance-1.0.jar:/usr/lib/hadoop-mapreduce/./aliyun-java-sdk-ram-3.0.0.jar:/usr/lib/hadoop-mapreduce/./curator-framework-2.7.1.jar:/usr/lib/hadoop-mapreduce/./jersey-json-1.9.jar:/usr/lib/hadoop-mapreduce/./hadoop-extras.jar:/usr/lib/hadoop-mapreduce/./curator-recipes-2.7.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-app.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-applicationhistoryservice-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./log4j-1.2.17.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-jobclient.jar:/usr/lib/hadoop-mapreduce/./apacheds-kerberos-codec-2.0.0-M15.jar:/usr/lib/hadoop-mapreduce/./jets3t-0.9.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-archives.jar:/usr/lib/hadoop-mapreduce/./jackson-xc-1.9.13.jar:/usr/lib/hadoop-mapreduce/./hadoop-gridmix-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./httpcore-4.4.11.jar:/usr/lib/hadoop-mapreduce/./guava-11.0.2.jar:/usr/lib/hadoop-mapreduce/./hadoop-auth-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jsp-api-2.1.jar:/usr/lib/hadoop-mapreduce/./aliyun-sdk-oss-3.4.1.jar:/usr/lib/hadoop-mapreduce



e/./HikariCP-java7-2.4.12.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-examples.jar:/usr/lib/hadoop-mapreduce/./xmlenc-0.52.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-resourcemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-applicationhistoryservice.jar:/usr/lib/hadoop-mapreduce/./servlet-api-2.5.jar:/usr/lib/hadoop-mapreduce/./commons-codec-1.4.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-jobclient-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./api-util-1.0.0-M20.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-registry-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-extras-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./commons-math3-3.1.1.jar:/usr/lib/hadoop-mapreduce/./guice-servlet-3.0.jar:/usr/lib/hadoop-mapreduce/./activation-1.1.jar:/usr/lib/hadoop-mapreduce/./aws-java-sdk-bundle-1.11.852.jar:/usr/lib/hadoop-mapreduce/./zookeeper-3.4.14.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-core.jar:/usr/lib/hadoop-mapreduce/./stax-api-1.0-2.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-common.jar:/usr/lib/hadoop-mapreduce/./commons-configuration-1.6.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-app-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jersey-core-1.9.jar:/usr/lib/hadoop-mapreduce/./fst-2.50.jar:/usr/lib/hadoop-mapreduce/./commons-beanutils-1.9.4.jar:/usr/lib/hadoop-mapreduce/./json-smart-1.3.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-gridmix.jar:/usr/lib/hadoop-mapreduce/./jackson-core-2.6.7.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-hs-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./commons-logging-1.1.3.jar:/usr/lib/hadoop-mapreduce/./hadoop-streaming-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-distcp.jar:/usr/lib/hadoop-mapreduce/./jetty-sslengine-6.1.26-emr.jar:/usr/lib/hadoop-mapreduce/./commons-compress-1.19.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-web-proxy.jar:/usr/lib/hadoop-mapreduce/./hadoop-aws-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./commons-io-2.4.jar:/usr/lib/hadoop-mapreduce/./hadoop-resourceestimator.jar:/usr/lib/hadoop-mapreduce/./jackson-core-asl-1.9.13.jar:/usr/lib/hadoop-mapreduce/./httpclient-4.5.9.jar:/usr/lib/hadoop-mapreduce/./netty-3.10.6.Final.jar:/usr/lib/hadoop-mapreduce/./commons-lang-2.6.jar:/usr/lib/hadoop-mapreduce/./hadoop-openstack.jar:/usr/lib/hadoop-mapreduce/./jetty-6.1.26-emr.jar:/usr/lib/hadoop-mapreduce/./jackson-annotations-2.6.7.jar:/usr/lib/hadoop-mapreduce/./jackson-mapper-asl-1.9.13.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-web-proxy-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./aliyun-java-sdk-ecs-4.2.0.jar:/usr/lib/hadoop-mapreduce/./geronimo-jcache\_1.0\_spec-1.0-alpha-1.jar:/usr/lib/hadoop-mapreduce/./hadoop-archive-logs.jar:/usr/lib/hadoop-mapreduce/./mssql-jdbc-6.2.1.jre7.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-registry.jar:/usr/lib/hadoop-mapreduce/./commons-digester-1.8.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-hs-plugins.jar:/usr/lib/hadoop-mapreduce/./protobuf-java-2.5.0.jar:/usr/lib/hadoop-mapreduce/./jsch-0.1.54.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-hs.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-core-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jdom-1.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jersey-server-1.9.jar:/usr/lib/hadoop-mapreduce/./commons-cli-1.2.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-shuffle.jar:/usr/lib/hadoop-mapreduce/./hadoop-streaming.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-api-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./commons-httpclient-3.1.jar:/usr/lib/hadoop-mapreduce/./aliyun-java-sdk-core-3.4.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-common.jar:/usr/lib/hadoop-mapreduce/./apache-ds-i18n-2.0.0-M15.jar:/usr/lib/hadoop-mapreduce/./woodstox-core-5.0.3.jar:/usr/lib/hadoop-mapreduce/./htrace-core4-4.1.0-incubating.jar:/usr/lib/hadoop-mapreduce/./hadoop-datajoin-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-examples-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./commons-net-3.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-sls.jar:/usr/lib/hadoop-mapreduce/./json-io-2.5.1.jar:/usr/lib/hadoop-mapreduce/./jersey-guice-1.9.jar:/usr/lib/hadoop-mapreduce/./hadoop-rumen-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./snappy-java-1.1.7.3.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-jobclient-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop-mapreduce/./leveldbjni-all-1.8.jar:/usr/lib/hadoop-mapreduce/./hadoop-azure

```

-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./api-asn1-api-1.0.0-M20.jar:/usr/lib/
hadoop-mapreduce/./jackson-databind-2.6.7.jar:/usr/lib/hadoop-mapreduce/./hadoop-m
apreduce-client-shuffle-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./gson-2.2.4.ja
r:/usr/lib/hadoop-mapreduce/./nimbus-jose-jwt-4.41.1.jar:/usr/lib/hadoop-mapreduc
e/./hadoop-archives-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jsr305-3.0.0.ja
r:/usr/lib/hadoop-mapreduce/./hadoop-ant.jar:/usr/lib/hadoop-mapreduce/./hadoop-sl
s-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-api.jar:/usr/lib/hadoop
-mapreduce/./hadoop-datajoin.jar:/usr/lib/hadoop-mapreduce/./metrics-core-3.0.1.ja
r:/usr/lib/hadoop-mapreduce/./azure-data-lake-store-sdk-2.2.3.jar:/usr/lib/hadoop-m
apreduce/./hadoop-archive-logs-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./guice
-3.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-resourceestimator-2.10.0-amzn-0.jar:/us
r/lib/hadoop-mapreduce/./java-util-1.9.0.jar
STARTUP_MSG:   build = git@aws157git.com:/pkg/Aws157BigTop -r d1e860a34cc1aea3d600c5
7c5c0270ea41579e8c; compiled by 'ec2-user' on 2020-09-19T02:05Z
STARTUP_MSG:   java = 1.8.0_312
*****/
24/04/28 21:20:46 INFO namenode.NameNode: registered UNIX signal handlers for [TERM,
HUP, INT]
24/04/28 21:20:46 INFO namenode.NameNode: createNameNode [-format, -force]
Formatting using clusterid: CID-fa2135ef-5d1b-4954-be9a-b407f6b3a20a
24/04/28 21:20:46 INFO namenode.FSEditLog: Edit logging is async:true
24/04/28 21:20:46 INFO namenode.FSNamesystem: KeyProvider: null
24/04/28 21:20:46 INFO namenode.FSNamesystem: fsLock is fair: true
24/04/28 21:20:46 INFO namenode.FSNamesystem: Detailed lock hold time metrics enable
d: false
24/04/28 21:20:46 INFO namenode.FSNamesystem: fsOwner                = root (auth:SIMPL
E)
24/04/28 21:20:46 INFO namenode.FSNamesystem: supergroup            = supergroup
24/04/28 21:20:46 INFO namenode.FSNamesystem: isPermissionEnabled = true
24/04/28 21:20:46 INFO namenode.FSNamesystem: HA Enabled: false
24/04/28 21:20:46 INFO common.Util: dfs.datanode.fileio.profiling.sampling.percentag
e set to 0. Disabling file IO profiling
24/04/28 21:20:46 INFO blockmanagement.DatanodeManager: dfs.block.invalidate.limit:
configured=1000, counted=60, effected=1000
24/04/28 21:20:46 INFO blockmanagement.DatanodeManager: dfs.namenode.datanode.regist
ration.ip-hostname-check=true
24/04/28 21:20:46 INFO blockmanagement.BlockManager: dfs.namenode.startup.delay.bloc
k.deletion.sec is set to 000:00:00:00.000
24/04/28 21:20:46 INFO blockmanagement.BlockManager: The block deletion will start a
round 2024 Apr 28 21:20:46
24/04/28 21:20:46 INFO util.GSet: Computing capacity for map BlocksMap
24/04/28 21:20:46 INFO util.GSet: VM type                = 64-bit
24/04/28 21:20:46 INFO util.GSet: 2.0% max memory 889 MB = 17.8 MB
24/04/28 21:20:46 INFO util.GSet: capacity              = 2^21 = 2097152 entries
24/04/28 21:20:46 INFO blockmanagement.BlockManager: dfs.block.access.token.enable=f
alse
24/04/28 21:20:46 WARN conf.Configuration: No unit for dfs.heartbeat.interval(3) ass
uming SECONDS
24/04/28 21:20:46 WARN conf.Configuration: No unit for dfs.namenode.safemode.extensi
on(30000) assuming MILLISECONDS
24/04/28 21:20:46 INFO blockmanagement.BlockManagerSafeMode: dfs.namenode.safemode.t
hreshold-pct = 0.9990000128746033
24/04/28 21:20:46 INFO blockmanagement.BlockManagerSafeMode: dfs.namenode.safemode.m
in.datanodes = 0
24/04/28 21:20:46 INFO blockmanagement.BlockManagerSafeMode: dfs.namenode.safemode.e
xtension = 30000

```

```

24/04/28 21:20:46 INFO blockmanagement.BlockManager: defaultReplication      = 3
24/04/28 21:20:46 INFO blockmanagement.BlockManager: maxReplication          = 51
2
24/04/28 21:20:46 INFO blockmanagement.BlockManager: minReplication          = 1
24/04/28 21:20:46 INFO blockmanagement.BlockManager: maxReplicationStreams    = 2
24/04/28 21:20:46 INFO blockmanagement.BlockManager: replicationRecheckInterval = 30
00
24/04/28 21:20:46 INFO blockmanagement.BlockManager: encryptDataTransfer      = fa
lse
24/04/28 21:20:46 INFO blockmanagement.BlockManager: maxNumBlocksToLog        = 10
00
24/04/28 21:20:46 INFO namenode.FSNamesystem: Append Enabled: true
24/04/28 21:20:46 INFO namenode.FSDirectory: GLOBAL serial map: bits=24 maxEntries=1
6777215
24/04/28 21:20:46 INFO util.GSet: Computing capacity for map INodeMap
24/04/28 21:20:46 INFO util.GSet: VM type          = 64-bit
24/04/28 21:20:46 INFO util.GSet: 1.0% max memory 889 MB = 8.9 MB
24/04/28 21:20:46 INFO util.GSet: capacity          = 2^20 = 1048576 entries
24/04/28 21:20:46 INFO namenode.FSDirectory: ACLs enabled? false
24/04/28 21:20:46 INFO namenode.FSDirectory: XAttrs enabled? true
24/04/28 21:20:46 INFO namenode.NameNode: Caching file names occurring more than 10
times
24/04/28 21:20:46 INFO snapshot.SnapshotManager: Loaded config captureOpenFiles: fal
seskipCaptureAccessTimeOnlyChange: false
24/04/28 21:20:46 INFO util.GSet: Computing capacity for map cachedBlocks
24/04/28 21:20:46 INFO util.GSet: VM type          = 64-bit
24/04/28 21:20:46 INFO util.GSet: 0.25% max memory 889 MB = 2.2 MB
24/04/28 21:20:46 INFO util.GSet: capacity          = 2^18 = 262144 entries
24/04/28 21:20:46 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.b
uckets = 10
24/04/28 21:20:46 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users =
10
24/04/28 21:20:46 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minu
tes = 1,5,25
24/04/28 21:20:46 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
24/04/28 21:20:46 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total hea
p and retry cache entry expiry time is 600000 millis
24/04/28 21:20:46 INFO util.GSet: Computing capacity for map NameNodeRetryCache
24/04/28 21:20:46 INFO util.GSet: VM type          = 64-bit
24/04/28 21:20:46 INFO util.GSet: 0.029999999329447746% max memory 889 MB = 273.1 KB
24/04/28 21:20:46 INFO util.GSet: capacity          = 2^15 = 32768 entries
24/04/28 21:20:46 INFO namenode.FSImage: Allocated new BlockPoolId: BP-393350832-10.
0.222.169-1714339246784
24/04/28 21:20:46 INFO common.Storage: Storage directory /opt/amazon/hadoop/hdfs/nam
enode has been successfully formatted.
24/04/28 21:20:46 INFO namenode.FSImageFormatProtobuf: Saving image file /opt/amazo
n/hadoop/hdfs/namenode/current/fsimage.ckpt_000000000000000000 using no compression
24/04/28 21:20:46 INFO namenode.FSImageFormatProtobuf: Image file /opt/amazon/hadoo
p/hdfs/namenode/current/fsimage.ckpt_000000000000000000 of size 323 bytes saved in
0 seconds .
24/04/28 21:20:46 INFO namenode.NNStorageRetentionManager: Going to retain 1 images
with txid >= 0
24/04/28 21:20:46 INFO namenode.FSImage: FSImageSaver clean checkpoint: txid = 0 whe
n meet shutdown.
24/04/28 21:20:46 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****

```

```

SHUTDOWN_MSG: Shutting down NameNode at algo-1/10.0.222.169
*****/
04-28 21:20 smspark-submit INFO      waiting for cluster to be up
24/04/28 21:20:47 INFO resourcemanager.ResourceManager: STARTUP_MSG:
/*****
STARTUP_MSG: Starting ResourceManager
STARTUP_MSG:  host = algo-1/10.0.222.169
STARTUP_MSG:  args = []
STARTUP_MSG:  version = 2.10.0-amzn-0
STARTUP_MSG:  classpath = /usr/lib/hadoop/etc/hadoop:/usr/lib/hadoop/etc/hadoop:/usr/lib/hadoop/etc/hadoop:/usr/lib/hadoop/lib/audience-annotations-0.5.0.jar:/usr/lib/hadoop/lib/jackson-jaxrs-1.9.13.jar:/usr/lib/hadoop/lib/commons-collections-3.2.2.jar:/usr/lib/hadoop/lib/mockito-all-1.8.5.jar:/usr/lib/hadoop/lib/spotbugs-annotations-3.1.9.jar:/usr/lib/hadoop/lib/stax2-api-3.1.4.jar:/usr/lib/hadoop/lib/curator-client-2.7.1.jar:/usr/lib/hadoop/lib/commons-lang3-3.4.jar:/usr/lib/hadoop/lib/jcip-annotations-1.0-1.jar:/usr/lib/hadoop/lib/jetty-util-6.1.26-emr.jar:/usr/lib/hadoop/lib/jaxb-api-2.2.2.jar:/usr/lib/hadoop/lib/asm-3.2.jar:/usr/lib/hadoop/lib/jettison-1.1.jar:/usr/lib/hadoop/lib/paranamer-2.3.jar:/usr/lib/hadoop/lib/java-xmlbuilder-0.4.jar:/usr/lib/hadoop/lib/jaxb-impl-2.2.3-1.jar:/usr/lib/hadoop/lib/avro-1.7.7.jar:/usr/lib/hadoop/lib/curator-framework-2.7.1.jar:/usr/lib/hadoop/lib/jersey-json-1.9.jar:/usr/lib/hadoop/lib/curator-recipes-2.7.1.jar:/usr/lib/hadoop/lib/log4j-1.2.17.jar:/usr/lib/hadoop/lib/apacheds-kerberos-codec-2.0.0-M15.jar:/usr/lib/hadoop/lib/jets3t-0.9.0.jar:/usr/lib/hadoop/lib/jackson-xc-1.9.13.jar:/usr/lib/hadoop/lib/httpcore-4.4.11.jar:/usr/lib/hadoop/lib/guava-11.0.2.jar:/usr/lib/hadoop/lib/jsp-api-2.1.jar:/usr/lib/hadoop/lib/xmlenc-0.52.jar:/usr/lib/hadoop/lib/servlet-api-2.5.jar:/usr/lib/hadoop/lib/commons-codec-1.4.jar:/usr/lib/hadoop/lib/slf4j-api-1.7.25.jar:/usr/lib/hadoop/lib/api-util-1.0.0-M20.jar:/usr/lib/hadoop/lib/commons-math3-3.1.1.jar:/usr/lib/hadoop/lib/activation-1.1.jar:/usr/lib/hadoop/lib/zookeeper-3.4.14.jar:/usr/lib/hadoop/lib/stax-api-1.0-2.jar:/usr/lib/hadoop/lib/commons-configuration-1.6.jar:/usr/lib/hadoop/lib/jersey-core-1.9.jar:/usr/lib/hadoop/lib/commons-beanutils-1.9.4.jar:/usr/lib/hadoop/lib/hamcrest-core-1.3.jar:/usr/lib/hadoop/lib/json-smart-1.3.1.jar:/usr/lib/hadoop/lib/commons-logging-1.1.3.jar:/usr/lib/hadoop/lib/jetty-sslengine-6.1.26-emr.jar:/usr/lib/hadoop/lib/commons-compress-1.19.jar:/usr/lib/hadoop/lib/commons-io-2.4.jar:/usr/lib/hadoop/lib/jackson-core-asl-1.9.13.jar:/usr/lib/hadoop/lib/httpclient-4.5.9.jar:/usr/lib/hadoop/lib/netty-3.10.6.Final.jar:/usr/lib/hadoop/lib/commons-lang-2.6.jar:/usr/lib/hadoop/lib/jetty-6.1.26-emr.jar:/usr/lib/hadoop/lib/jackson-mapper-asl-1.9.13.jar:/usr/lib/hadoop/lib/commons-digester-1.8.jar:/usr/lib/hadoop/lib/protobuf-java-2.5.0.jar:/usr/lib/hadoop/lib/jsch-0.1.54.jar:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.25.jar:/usr/lib/hadoop/lib/jersey-server-1.9.jar:/usr/lib/hadoop/lib/commons-cli-1.2.jar:/usr/lib/hadoop/lib/apacheds-i18n-2.0.0-M15.jar:/usr/lib/hadoop/lib/woodstox-core-5.0.3.jar:/usr/lib/hadoop/lib/htrace-core4-4.1.0-incubating.jar:/usr/lib/hadoop/lib/commons-net-3.1.jar:/usr/lib/hadoop/lib/snappy-java-1.1.7.3.jar:/usr/lib/hadoop/lib/api-asn1-api-1.0.0-M20.jar:/usr/lib/hadoop/lib/gson-2.2.4.jar:/usr/lib/hadoop/lib/nimbus-jose-jwt-4.41.1.jar:/usr/lib/hadoop/lib/jsr305-3.0.0.jar:/usr/lib/hadoop/lib/junit-4.11.jar:/usr/lib/hadoop/./hadoop-openstack-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-azure-datalake.jar:/usr/lib/hadoop/./hadoop-distcp-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-aliyun.jar:/usr/lib/hadoop/./hadoop-ant-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-azure.jar:/usr/lib/hadoop/./hadoop-nfs-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-auth.jar:/usr/lib/hadoop/./hadoop-yarn-common-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-aliyun-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-yarn-server-resourcemanager.jar:/usr/lib/hadoop/./hadoop-aws.jar:/usr/lib/hadoop/./hadoop-azure-datalake-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-rumen.jar:/usr/lib/hadoop/./hadoop-extras.jar:/usr/lib/hadoop/./hadoop-yarn-server-applicationhistoryservice-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-annotation-s-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-archives.jar:/usr/lib/hadoop/./hadoop-gridmix-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-auth-2.10.0-amzn-0.jar:/usr/li

```

b/hadoop/./hadoop-yarn-server-resourcemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-yarn-server-applicationhistoryservice.jar:/usr/lib/hadoop/./hadoop-annotations.jar:/usr/lib/hadoop/./hadoop-yarn-registry-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-extras-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-yarn-common.jar:/usr/lib/hadoop/./hadoop-gridmix.jar:/usr/lib/hadoop/./hadoop-streaming-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-distcp.jar:/usr/lib/hadoop/./hadoop-yarn-server-common-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-yarn-server-web-proxy.jar:/usr/lib/hadoop/./hadoop-aws-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-common-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop/./hadoop-resourceestimator.jar:/usr/lib/hadoop/./hadoop-opendstack.jar:/usr/lib/hadoop/./hadoop-yarn-server-web-proxy-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-archive-logs.jar:/usr/lib/hadoop/./hadoop-nfs.jar:/usr/lib/hadoop/./hadoop-yarn-registry.jar:/usr/lib/hadoop/./hadoop-common-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-streaming.jar:/usr/lib/hadoop/./hadoop-yarn-api-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-yarn-server-common.jar:/usr/lib/hadoop/./hadoop-datajoin-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-common.jar:/usr/lib/hadoop/./hadoop-sls.jar:/usr/lib/hadoop/./hadoop-rumen-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-azure-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-archives-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-ant.jar:/usr/lib/hadoop/./hadoop-sls-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-yarn-api.jar:/usr/lib/hadoop/./hadoop-datajoin.jar:/usr/lib/hadoop/./hadoop-archive-logs-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-resourceestimator-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./usr/lib/hadoop-hdfs/lib/jetty-util-6.1.26-emr.jar:/usr/lib/hadoop-hdfs/lib/asm-3.2.jar:/usr/lib/hadoop-hdfs/lib/okio-1.6.0.jar:/usr/lib/hadoop-hdfs/lib/commons-daemon-1.0.13.jar:/usr/lib/hadoop-hdfs/lib/log4j-1.2.17.jar:/usr/lib/hadoop-hdfs/lib/guava-11.0.2.jar:/usr/lib/hadoop-hdfs/lib/xmlenc-0.52.jar:/usr/lib/hadoop-hdfs/lib/xercesImpl-2.12.0.jar:/usr/lib/hadoop-hdfs/lib/servlet-api-2.5.jar:/usr/lib/hadoop-hdfs/lib/commons-codec-1.4.jar:/usr/lib/hadoop-hdfs/lib/xml-apis-1.4.01.jar:/usr/lib/hadoop-hdfs/lib/jersey-core-1.9.jar:/usr/lib/hadoop-hdfs/lib/jackson-core-2.6.7.jar:/usr/lib/hadoop-hdfs/lib/commons-logging-1.1.3.jar:/usr/lib/hadoop-hdfs/lib/commons-io-2.4.jar:/usr/lib/hadoop-hdfs/lib/jackson-core-asl-1.9.13.jar:/usr/lib/hadoop-hdfs/lib/netty-3.10.6.Final.jar:/usr/lib/hadoop-hdfs/lib/commons-lang-2.6.jar:/usr/lib/hadoop-hdfs/lib/jetty-6.1.26-emr.jar:/usr/lib/hadoop-hdfs/lib/jackson-annotations-2.6.7.jar:/usr/lib/hadoop-hdfs/lib/jackson-mapper-asl-1.9.13.jar:/usr/lib/hadoop-hdfs/lib/protobuf-java-2.5.0.jar:/usr/lib/hadoop-hdfs/lib/jersey-server-1.9.jar:/usr/lib/hadoop-hdfs/lib/commons-cli-1.2.jar:/usr/lib/hadoop-hdfs/lib/htrace-core4-4.1.0-incubating.jar:/usr/lib/hadoop-hdfs/lib/leveldb-jni-all-1.8.jar:/usr/lib/hadoop-hdfs/lib/jackson-databind-2.6.7.jar:/usr/lib/hadoop-hdfs/lib/okhttp-2.7.5.jar:/usr/lib/hadoop-hdfs/lib/jsr305-3.0.0.jar:/usr/lib/hadoop-hdfs/lib/netty-all-4.0.23.Final.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-client-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-native-client.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-rbf-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-rbf-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-client-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-native-client-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-nfs.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-client.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-nfs-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-rbf.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-native-client-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop-yarn/lib/audience-annotations-0.5.0.jar:/usr/lib/hadoop-yarn/lib/javax.inject-1.jar:/usr/lib/hadoop-yarn/lib/jackson-jaxrs-1.9.13.jar:/usr/lib/hadoop-yarn/lib/commons-collections-3.2.2.jar:/usr/lib/hadoop-yarn/lib/spotbugs-annotations-3.1.9.jar:/usr/lib/hadoop-yarn/lib/stax2-api-3.1.4.jar:/usr/lib/hadoop-yarn/lib/curator-client-2.7.1.jar:/usr/lib/hadoop-yarn/lib/commons-lang3-3.4.jar:/usr/lib/hadoop-yarn/lib/jcip-annotations-1.0-1.jar:/usr/lib/hadoop-yarn/lib/jetty-util-6.1.26-emr.jar:/usr/lib/hadoop-yarn/lib/jaxb-api-2.2.2.jar:/usr/lib/hadoop-yarn/lib/jersey-client-1.9.jar:/usr/lib/hadoop-yarn/lib/asm-3.2.jar:/usr/lib/hadoop-yarn/lib/jettison-1.1.jar:/usr/lib/hadoop-yarn/lib/paranamer-2.3.jar:/usr/lib/hadoop-yarn/lib/java-xmlbuilder-

0.4.jar:/usr/lib/hadoop-yarn/lib/jaxb-impl-2.2.3-1.jar:/usr/lib/hadoop-yarn/lib/avro-1.7.7.jar:/usr/lib/hadoop-yarn/lib/ehcache-3.3.1.jar:/usr/lib/hadoop-yarn/lib/aopalliance-1.0.jar:/usr/lib/hadoop-yarn/lib/curator-framework-2.7.1.jar:/usr/lib/hadoop-yarn/lib/jersey-json-1.9.jar:/usr/lib/hadoop-yarn/lib/curator-recipes-2.7.1.jar:/usr/lib/hadoop-yarn/lib/log4j-1.2.17.jar:/usr/lib/hadoop-yarn/lib/apacheds-kerberos-codec-2.0.0-M15.jar:/usr/lib/hadoop-yarn/lib/jets3t-0.9.0.jar:/usr/lib/hadoop-yarn/lib/jackson-xc-1.9.13.jar:/usr/lib/hadoop-yarn/lib/httpcore-4.4.11.jar:/usr/lib/hadoop-yarn/lib/guava-11.0.2.jar:/usr/lib/hadoop-yarn/lib/jsp-api-2.1.jar:/usr/lib/hadoop-yarn/lib/HikariCP-java7-2.4.12.jar:/usr/lib/hadoop-yarn/lib/xmlenc-0.52.jar:/usr/lib/hadoop-yarn/lib/servlet-api-2.5.jar:/usr/lib/hadoop-yarn/lib/commons-codec-1.4.jar:/usr/lib/hadoop-yarn/lib/api-util-1.0.0-M20.jar:/usr/lib/hadoop-yarn/lib/commons-math3-3.1.1.jar:/usr/lib/hadoop-yarn/lib/guice-servlet-3.0.jar:/usr/lib/hadoop-yarn/lib/activation-1.1.jar:/usr/lib/hadoop-yarn/lib/zookeeper-3.4.14.jar:/usr/lib/hadoop-yarn/lib/stax-api-1.0-2.jar:/usr/lib/hadoop-yarn/lib/commons-configuration-1.6.jar:/usr/lib/hadoop-yarn/lib/jersey-core-1.9.jar:/usr/lib/hadoop-yarn/lib/fst-2.50.jar:/usr/lib/hadoop-yarn/lib/commons-beanutils-1.9.4.jar:/usr/lib/hadoop-yarn/lib/json-smart-1.3.1.jar:/usr/lib/hadoop-yarn/lib/commons-logging-1.1.3.jar:/usr/lib/hadoop-yarn/lib/jetty-sslengine-6.1.26-emr.jar:/usr/lib/hadoop-yarn/lib/commons-compress-1.19.jar:/usr/lib/hadoop-yarn/lib/commons-io-2.4.jar:/usr/lib/hadoop-yarn/lib/jackson-core-asl-1.9.13.jar:/usr/lib/hadoop-yarn/lib/httpclient-4.5.9.jar:/usr/lib/hadoop-yarn/lib/netty-3.10.6.Final.jar:/usr/lib/hadoop-yarn/lib/commons-lang-2.6.jar:/usr/lib/hadoop-yarn/lib/jetty-6.1.26-emr.jar:/usr/lib/hadoop-yarn/lib/jackson-mapper-asl-1.9.13.jar:/usr/lib/hadoop-yarn/lib/geronimo-jcache\_1.0\_spec-1.0-alpha-1.jar:/usr/lib/hadoop-yarn/lib/mssql-jdbc-6.2.1.jre7.jar:/usr/lib/hadoop-yarn/lib/commons-digester-1.8.jar:/usr/lib/hadoop-yarn/lib/protobuf-java-2.5.0.jar:/usr/lib/hadoop-yarn/lib/jsch-0.1.54.jar:/usr/lib/hadoop-yarn/lib/jersey-server-1.9.jar:/usr/lib/hadoop-yarn/lib/commons-cli-1.2.jar:/usr/lib/hadoop-yarn/lib/apacheds-i18n-2.0.0-M15.jar:/usr/lib/hadoop-yarn/lib/woodstox-core-5.0.3.jar:/usr/lib/hadoop-yarn/lib/htrace-core4-4.1.0-incubating.jar:/usr/lib/hadoop-yarn/lib/commons-net-3.1.jar:/usr/lib/hadoop-yarn/lib/json-io-2.5.1.jar:/usr/lib/hadoop-yarn/lib/jersey-guice-1.9.jar:/usr/lib/hadoop-yarn/lib/snappy-java-1.1.7.3.jar:/usr/lib/hadoop-yarn/lib/leveldbjni-all-1.8.jar:/usr/lib/hadoop-yarn/lib/api-asn1-api-1.0.0-M20.jar:/usr/lib/hadoop-yarn/lib/gson-2.2.4.jar:/usr/lib/hadoop-yarn/lib/nimbus-jose-jwt-4.41.1.jar:/usr/lib/hadoop-yarn/lib/jsr305-3.0.0.jar:/usr/lib/hadoop-yarn/lib/metrics-core-3.0.1.jar:/usr/lib/hadoop-yarn/lib/guice-3.0.jar:/usr/lib/hadoop-yarn/lib/java-util-1.9.0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-tests-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-nodemanager.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-client-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-timeline-pluginstorage-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-tests.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-applications-unmanaged-am-launcher.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-router-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-resourcemanager.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-applicationhistoryservice-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-resourcemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-applicationhistoryservice.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-applications-distributedshell-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-applications-unmanaged-am-launcher-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-registry-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-common.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-sharedcachemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-client.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-web-proxy.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-sharedcachemanager.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-web-proxy-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-router.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-registry.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-api-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-timeline-pluginstorage.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-common.

```

jar:/usr/lib/hadoop-yarn/./hadoop-yarn-applications-distributedshell.jar:/usr/lib/h
adoop-yarn/./hadoop-yarn-api.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-nodemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/lib/javax.inject-1.jar:/usr/lib/hadoop-mapreduce/lib/asm-3.2.jar:/usr/lib/hadoop-mapreduce/lib/paranamer-2.3.jar:/usr/lib/hadoop-mapreduce/lib/avro-1.7.7.jar:/usr/lib/hadoop-mapreduce/lib/aopalliance-1.0.jar:/usr/lib/hadoop-mapreduce/lib/log4j-1.2.17.jar:/usr/lib/hadoop-mapreduce/lib/guice-servlet-3.0.jar:/usr/lib/hadoop-mapreduce/lib/jersey-core-1.9.jar:/usr/lib/hadoop-mapreduce/lib/hamcrest-core-1.3.jar:/usr/lib/hadoop-mapreduce/lib/commons-compress-1.19.jar:/usr/lib/hadoop-mapreduce/lib/commons-io-2.4.jar:/usr/lib/hadoop-mapreduce/lib/jackson-core-asl-1.9.13.jar:/usr/lib/hadoop-mapreduce/lib/netty-3.10.6.Final.jar:/usr/lib/hadoop-mapreduce/lib/jackson-mapper-asl-1.9.13.jar:/usr/lib/hadoop-mapreduce/lib/protobuf-java-2.5.0.jar:/usr/lib/hadoop-mapreduce/lib/jersey-server-1.9.jar:/usr/lib/hadoop-mapreduce/lib/jersey-guice-1.9.jar:/usr/lib/hadoop-mapreduce/lib/snappy-java-1.1.7.3.jar:/usr/lib/hadoop-mapreduce/lib/leveldbjni-all-1.8.jar:/usr/lib/hadoop-mapreduce/lib/junit-4.11.jar:/usr/lib/hadoop-mapreduce/lib/guice-3.0.jar:/usr/lib/hadoop-mapreduce/./audience-annotations-0.5.0.jar:/usr/lib/hadoop-mapreduce/./javax.inject-1.jar:/usr/lib/hadoop-mapreduce/./hadoop-openstack-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-azure-datalake.jar:/usr/lib/hadoop-mapreduce/./hadoop-distcp-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jackson-jaxrs-1.9.13.jar:/usr/lib/hadoop-mapreduce/./commons-collections-3.2.2.jar:/usr/lib/hadoop-mapreduce/./hadoop-aliyun.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-hs-plugins-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-ant-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./azure-storage-5.4.0.jar:/usr/lib/hadoop-mapreduce/./spotbugs-annotations-3.1.9.jar:/usr/lib/hadoop-mapreduce/./stax2-api-3.1.4.jar:/usr/lib/hadoop-mapreduce/./hadoop-azure.jar:/usr/lib/hadoop-mapreduce/./curator-client-2.7.1.jar:/usr/lib/hadoop-mapreduce/./commons-lang3-3.4.jar:/usr/lib/hadoop-mapreduce/./jcip-annotations-1.0-1.jar:/usr/lib/hadoop-mapreduce/./jetty-util-6.1.26-emr.jar:/usr/lib/hadoop-mapreduce/./jaxb-api-2.2.2.jar:/usr/lib/hadoop-mapreduce/./jersey-client-1.9.jar:/usr/lib/hadoop-mapreduce/./azure-keyvault-core-0.8.0.jar:/usr/lib/hadoop-mapreduce/./asm-3.2.jar:/usr/lib/hadoop-mapreduce/./jettison-1.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-auth.jar:/usr/lib/hadoop-mapreduce/./ojslgo-4.3.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-common.jar:/usr/lib/hadoop-mapreduce/./paranamer-2.3.jar:/usr/lib/hadoop-mapreduce/./java-xmlbuilder-0.4.jar:/usr/lib/hadoop-mapreduce/./hadoop-aliyun-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-resourcemanager.jar:/usr/lib/hadoop-mapreduce/./jaxb-impl-2.2.3-1.jar:/usr/lib/hadoop-mapreduce/./avro-1.7.7.jar:/usr/lib/hadoop-mapreduce/./aliyun-java-sdk-sts-3.0.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-aws.jar:/usr/lib/hadoop-mapreduce/./hadoop-azure-datalake-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-rumen.jar:/usr/lib/hadoop-mapreduce/./ehcache-3.3.1.jar:/usr/lib/hadoop-mapreduce/./aopalliance-1.0.jar:/usr/lib/hadoop-mapreduce/./aliyun-java-sdk-ram-3.0.0.jar:/usr/lib/hadoop-mapreduce/./curator-framework-2.7.1.jar:/usr/lib/hadoop-mapreduce/./jersey-json-1.9.jar:/usr/lib/hadoop-mapreduce/./hadoop-extras.jar:/usr/lib/hadoop-mapreduce/./curator-recipes-2.7.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-app.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-applicationhistoryservice-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./log4j-1.2.17.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-jobclient.jar:/usr/lib/hadoop-mapreduce/./apacheds-kerberos-codec-2.0.0-M15.jar:/usr/lib/hadoop-mapreduce/./jets3t-0.9.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-archives.jar:/usr/lib/hadoop-mapreduce/./jackson-xc-1.9.13.jar:/usr/lib/hadoop-mapreduce/./hadoop-gridmix-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./httpcore-4.4.11.jar:/usr/lib/hadoop-mapreduce/./guava-11.0.2.jar:/usr/lib/hadoop-mapreduce/./hadoop-auth-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jsp-api-2.1.jar:/usr/lib/hadoop-mapreduce/./aliyun-sdk-oss-3.4.1.jar:/usr/lib/hadoop-mapreduce/./HikariCP-java7-2.4.12.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-examples.jar:/usr/lib/hadoop-mapreduce/./xmlenc-0.52.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-resourcemanager

```

r-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-applicationhisto  
 ryservice.jar:/usr/lib/hadoop-mapreduce/./servlet-api-2.5.jar:/usr/lib/hadoop-mapre  
 duce/./commons-codec-1.4.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-j  
 obclient-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./api-util-1.0.0-M20.jar:/usr/  
 lib/hadoop-mapreduce/./hadoop-yarn-registry-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapre  
 duce/./hadoop-extras-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./commons-math3-  
 3.1.1.jar:/usr/lib/hadoop-mapreduce/./guice-servlet-3.0.jar:/usr/lib/hadoop-mapredu  
 ce/./activation-1.1.jar:/usr/lib/hadoop-mapreduce/./aws-java-sdk-bundle-1.11.852.j  
 ar:/usr/lib/hadoop-mapreduce/./zookeeper-3.4.14.jar:/usr/lib/hadoop-mapreduce/./ha  
 doop-mapreduce-client-core.jar:/usr/lib/hadoop-mapreduce/./stax-api-1.0-2.jar:/usr/  
 lib/hadoop-mapreduce/./hadoop-yarn-common.jar:/usr/lib/hadoop-mapreduce/./commons-  
 configuration-1.6.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-app-2.10.  
 0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jersey-core-1.9.jar:/usr/lib/hadoop-mapred  
 uce/./fst-2.50.jar:/usr/lib/hadoop-mapreduce/./commons-beanutils-1.9.4.jar:/usr/li  
 b/hadoop-mapreduce/./json-smart-1.3.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-gridm  
 ix.jar:/usr/lib/hadoop-mapreduce/./jackson-core-2.6.7.jar:/usr/lib/hadoop-mapreduc  
 e/./hadoop-mapreduce-client-hs-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./commo  
 ns-logging-1.1.3.jar:/usr/lib/hadoop-mapreduce/./hadoop-streaming-2.10.0-amzn-0.ja  
 r:/usr/lib/hadoop-mapreduce/./hadoop-distcp.jar:/usr/lib/hadoop-mapreduce/./jetty-  
 sslengine-6.1.26-emr.jar:/usr/lib/hadoop-mapreduce/./commons-compress-1.19.jar:/us  
 r/lib/hadoop-mapreduce/./hadoop-yarn-server-common-2.10.0-amzn-0.jar:/usr/lib/hadoo  
 p-mapreduce/./hadoop-yarn-server-web-proxy.jar:/usr/lib/hadoop-mapreduce/./hadoop-  
 aws-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./commons-io-2.4.jar:/usr/lib/hadoo  
 p-mapreduce/./hadoop-resourceestimator.jar:/usr/lib/hadoop-mapreduce/./jackson-cor  
 e-asl-1.9.13.jar:/usr/lib/hadoop-mapreduce/./httpclient-4.5.9.jar:/usr/lib/hadoop-m  
 apreduce/./netty-3.10.6.Final.jar:/usr/lib/hadoop-mapreduce/./commons-lang-2.6.ja  
 r:/usr/lib/hadoop-mapreduce/./hadoop-openstack.jar:/usr/lib/hadoop-mapreduce/./jet  
 ty-6.1.26-emr.jar:/usr/lib/hadoop-mapreduce/./jackson-annotations-2.6.7.jar:/usr/li  
 b/hadoop-mapreduce/./jackson-mapper-asl-1.9.13.jar:/usr/lib/hadoop-mapreduce/./had  
 oop-yarn-server-web-proxy-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./aliyun-java  
 -sdk-ecs-4.2.0.jar:/usr/lib/hadoop-mapreduce/./geronimo-jcache\_1.0\_spec-1.0-alpha-  
 1.jar:/usr/lib/hadoop-mapreduce/./hadoop-archive-logs.jar:/usr/lib/hadoop-mapreduc  
 e/./mssql-jdbc-6.2.1.jre7.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-registry.ja  
 r:/usr/lib/hadoop-mapreduce/./commons-digester-1.8.jar:/usr/lib/hadoop-mapreduc  
 e/./hadoop-mapreduce-client-hs-plugins.jar:/usr/lib/hadoop-mapreduce/./protobuf-ja  
 va-2.5.0.jar:/usr/lib/hadoop-mapreduce/./jsch-0.1.54.jar:/usr/lib/hadoop-mapreduc  
 e/./hadoop-mapreduce-client-hs.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-cl  
 ient-core-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jdom-1.1.jar:/usr/lib/hadoo  
 p-mapreduce/./hadoop-mapreduce-client-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapr  
 educe/./jersey-server-1.9.jar:/usr/lib/hadoop-mapreduce/./commons-cli-1.2.jar:/us  
 r/lib/hadoop-mapreduce/./hadoop-mapreduce-client-shuffle.jar:/usr/lib/hadoop-mapred  
 uce/./hadoop-streaming.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-api-2.10.0-amzn  
 -0.jar:/usr/lib/hadoop-mapreduce/./commons-httpclient-3.1.jar:/usr/lib/hadoop-mapre  
 duce/./aliyun-java-sdk-core-3.4.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-serv  
 er-common.jar:/usr/lib/hadoop-mapreduce/./apacheds-i18n-2.0.0-M15.jar:/usr/lib/hadoo  
 p-mapreduce/./woodstox-core-5.0.3.jar:/usr/lib/hadoop-mapreduce/./htrace-core4-4.  
 1.0-incubating.jar:/usr/lib/hadoop-mapreduce/./hadoop-datajoin-2.10.0-amzn-0.jar:/u  
 sr/lib/hadoop-mapreduce/./hadoop-mapreduce-examples-2.10.0-amzn-0.jar:/usr/lib/hado  
 op-mapreduce/./commons-net-3.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-sls.jar:/us  
 r/lib/hadoop-mapreduce/./json-io-2.5.1.jar:/usr/lib/hadoop-mapreduce/./jersey-guic  
 e-1.9.jar:/usr/lib/hadoop-mapreduce/./hadoop-rumen-2.10.0-amzn-0.jar:/usr/lib/hadoo  
 p-mapreduce/./snappy-java-1.1.7.3.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce  
 -client-jobclient-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop-mapreduce/./leveldbjni-al  
 l-1.8.jar:/usr/lib/hadoop-mapreduce/./hadoop-azure-2.10.0-amzn-0.jar:/usr/lib/hadoo  
 p-mapreduce/./api-asn1-api-1.0.0-M20.jar:/usr/lib/hadoop-mapreduce/./jackson-datab  
 ind-2.6.7.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-shuffle-2.10.0-am



zn-0.jar:/usr/lib/hadoop-mapreduce/./gson-2.2.4.jar:/usr/lib/hadoop-mapreduce/./nibus-jose-jwt-4.41.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-archives-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jsr305-3.0.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-ant.jar:/usr/lib/hadoop-mapreduce/./hadoop-sls-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-api.jar:/usr/lib/hadoop-mapreduce/./hadoop-datajoin.jar:/usr/lib/hadoop-mapreduce/./metrics-core-3.0.1.jar:/usr/lib/hadoop-mapreduce/./azure-data-lake-store-sdk-2.2.3.jar:/usr/lib/hadoop-mapreduce/./hadoop-archive-logs-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./guice-3.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-resourceestimator-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./java-util-1.9.0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-tests-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-nodemanager.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-client-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-timeline-pluginstorage-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-tests.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-applications-unmanaged-am-launcher.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-router-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-resourcemanager.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-applicationhistoryservice-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-resourcemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-applicationhistoryservice.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-applications-distributedshell-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-applications-unmanaged-am-launcher-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-registry-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-common.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-sharedcachemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-client.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-web-proxy.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-sharedcachemanager.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-web-proxy-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-router.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-registry.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-api-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-timeline-pluginstorage.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-common.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-applications-distributedshell.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-api.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-nodemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/lib/audience-annotations-0.5.0.jar:/usr/lib/hadoop-yarn/lib/javax.inject-1.jar:/usr/lib/hadoop-yarn/lib/jackson-jaxrs-1.9.13.jar:/usr/lib/hadoop-yarn/lib/commons-collections-3.2.2.jar:/usr/lib/hadoop-yarn/lib/spotbugs-annotations-3.1.9.jar:/usr/lib/hadoop-yarn/lib/stax2-api-3.1.4.jar:/usr/lib/hadoop-yarn/lib/curator-client-2.7.1.jar:/usr/lib/hadoop-yarn/lib/commons-lang3-3.4.jar:/usr/lib/hadoop-yarn/lib/jcip-annotations-1.0-1.jar:/usr/lib/hadoop-yarn/lib/jetty-util-6.1.26-emr.jar:/usr/lib/hadoop-yarn/lib/jaxb-api-2.2.2.jar:/usr/lib/hadoop-yarn/lib/jersey-client-1.9.jar:/usr/lib/hadoop-yarn/lib/asm-3.2.jar:/usr/lib/hadoop-yarn/lib/jettison-1.1.jar:/usr/lib/hadoop-yarn/lib/paranamer-2.3.jar:/usr/lib/hadoop-yarn/lib/java-xmlbuilder-0.4.jar:/usr/lib/hadoop-yarn/lib/jaxb-impl-2.2.3-1.jar:/usr/lib/hadoop-yarn/lib/avro-1.7.7.jar:/usr/lib/hadoop-yarn/lib/ehcache-3.3.1.jar:/usr/lib/hadoop-yarn/lib/aopalliance-1.0.jar:/usr/lib/hadoop-yarn/lib/curator-framework-2.7.1.jar:/usr/lib/hadoop-yarn/lib/jersey-json-1.9.jar:/usr/lib/hadoop-yarn/lib/curator-recipes-2.7.1.jar:/usr/lib/hadoop-yarn/lib/log4j-1.2.17.jar:/usr/lib/hadoop-yarn/lib/apacheds-kerberos-codec-2.0.0-M15.jar:/usr/lib/hadoop-yarn/lib/jets3t-0.9.0.jar:/usr/lib/hadoop-yarn/lib/jackson-xc-1.9.13.jar:/usr/lib/hadoop-yarn/lib/httpcore-4.4.11.jar:/usr/lib/hadoop-yarn/lib/guava-11.0.2.jar:/usr/lib/hadoop-yarn/lib/jsp-api-2.1.jar:/usr/lib/hadoop-yarn/lib/HikariCP-java7-2.4.12.jar:/usr/lib/hadoop-yarn/lib/xmlenc-0.52.jar:/usr/lib/hadoop-yarn/lib/servlet-api-2.5.jar:/usr/lib/hadoop-yarn/lib/commons-codec-1.4.jar:/usr/lib/hadoop-yarn/lib/api-util-1.0.0-M20.jar:/usr/lib/hadoop-yarn/lib/commons-math3-3.1.1.jar:/usr/lib/hadoop-yarn/lib/guice-servlet-3.0.jar:/usr/lib/hadoop-yarn/lib/activation-1.1.jar:/usr/lib/hadoop-yarn/lib/zookeeper-3.4.14.jar:/usr/lib/hadoop-yarn/lib/stax-api-1.0-2.jar:/usr/lib/hadoop-yarn/lib/commons-configuration-1.6.jar:/usr/lib/hadoop-yarn

```

n/lib/jersey-core-1.9.jar:/usr/lib/hadoop-yarn/lib/fst-2.50.jar:/usr/lib/hadoop-yarn
n/lib/commons-beanutils-1.9.4.jar:/usr/lib/hadoop-yarn/lib/json-smart-1.3.1.jar:/us
r/lib/hadoop-yarn/lib/commons-logging-1.1.3.jar:/usr/lib/hadoop-yarn/lib/jetty-sslen
gine-6.1.26-emr.jar:/usr/lib/hadoop-yarn/lib/commons-compress-1.19.jar:/usr/lib/hado
op-yarn/lib/commons-io-2.4.jar:/usr/lib/hadoop-yarn/lib/jackson-core-asl-1.9.13.ja
r:/usr/lib/hadoop-yarn/lib/httpclient-4.5.9.jar:/usr/lib/hadoop-yarn/lib/netty-3.10.
6.Final.jar:/usr/lib/hadoop-yarn/lib/commons-lang-2.6.jar:/usr/lib/hadoop-yarn/lib/j
etty-6.1.26-emr.jar:/usr/lib/hadoop-yarn/lib/jackson-mapper-asl-1.9.13.jar:/usr/lib/
hadoop-yarn/lib/geronimo-jcache_1.0_spec-1.0-alpha-1.jar:/usr/lib/hadoop-yarn/lib/ms
sql-jdbc-6.2.1.jre7.jar:/usr/lib/hadoop-yarn/lib/commons-digester-1.8.jar:/usr/lib/h
adoop-yarn/lib/protobuf-java-2.5.0.jar:/usr/lib/hadoop-yarn/lib/jsch-0.1.54.jar:/us
r/lib/hadoop-yarn/lib/jersey-server-1.9.jar:/usr/lib/hadoop-yarn/lib/commons-cli-1.
2.jar:/usr/lib/hadoop-yarn/lib/apacheds-i18n-2.0.0-M15.jar:/usr/lib/hadoop-yarn/lib/
woodstox-core-5.0.3.jar:/usr/lib/hadoop-yarn/lib/htrace-core4-4.1.0-incubating.jar:/
usr/lib/hadoop-yarn/lib/commons-net-3.1.jar:/usr/lib/hadoop-yarn/lib/json-io-2.5.1.j
ar:/usr/lib/hadoop-yarn/lib/jersey-guice-1.9.jar:/usr/lib/hadoop-yarn/lib/snappy-jav
a-1.1.7.3.jar:/usr/lib/hadoop-yarn/lib/leveldbjni-all-1.8.jar:/usr/lib/hadoop-yarn/l
ib/api-asn1-api-1.0.0-M20.jar:/usr/lib/hadoop-yarn/lib/gson-2.2.4.jar:/usr/lib/hadoo
p-yarn/lib/nimbus-jose-jwt-4.41.1.jar:/usr/lib/hadoop-yarn/lib/jsr305-3.0.0.jar:/us
r/lib/hadoop-yarn/lib/metrics-core-3.0.1.jar:/usr/lib/hadoop-yarn/lib/guice-3.0.ja
r:/usr/lib/hadoop-yarn/lib/java-util-1.9.0.jar:/usr/lib/hadoop/etc/hadoop/rm-config/
log4j.properties:/usr/lib/hadoop-yarn/./timelineservice/hadoop-yarn-server-timeline
service-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./timelineservice/hadoop-yarn-server
-timelineservice-hbase-coprocessor-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./timelin
eservice/hadoop-yarn-server-timelineservice-hbase-client-2.10.0-amzn-0.jar:/usr/lib/
hadoop-yarn/./timelineservice/hadoop-yarn-server-timelineservice-hbase-common-2.10.
0-amzn-0.jar:/usr/lib/hadoop-yarn/./timelineservice/lib/jsr311-api-1.1.1.jar:/usr/l
ib/hadoop-yarn/./timelineservice/lib/hbase-common-1.2.6.jar:/usr/lib/hadoop-yarn
/./timelineservice/lib/jcodings-1.0.8.jar:/usr/lib/hadoop-yarn/./timelineservice/
lib/metrics-core-2.2.0.jar:/usr/lib/hadoop-yarn/./timelineservice/lib/commons-csv-
1.0.jar:/usr/lib/hadoop-yarn/./timelineservice/lib/jackson-core-2.6.7.jar:/usr/lib/
hadoop-yarn/./timelineservice/lib/hbase-client-1.2.6.jar:/usr/lib/hadoop-yarn/./ti
melineservice/lib/joni-2.1.2.jar:/usr/lib/hadoop-yarn/./timelineservice/lib/htrace-
core-3.1.0-incubating.jar:/usr/lib/hadoop-yarn/./timelineservice/lib/hbase-protocol
-1.2.6.jar:/usr/lib/hadoop-yarn/./timelineservice/lib/netty-all-4.0.23.Final.jar:/u
sr/lib/hadoop-yarn/./timelineservice/lib/hbase-annotations-1.2.6.jar
STARTUP_MSG:   build = git@aws157git.com:/pkg/Aws157BigTop -r d1e860a34cc1aea3d600c5
7c5c0270ea41579e8c; compiled by 'ec2-user' on 2020-09-19T02:05Z
STARTUP_MSG:   java = 1.8.0_312
*****/
24/04/28 21:20:47 INFO resourcemanager.ResourceManager: registered UNIX signal handl
ers for [TERM, HUP, INT]
24/04/28 21:20:47 INFO nodemanager.NodeManager: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NodeManager
STARTUP_MSG:   host = algo-1/10.0.222.169
STARTUP_MSG:   args = []
STARTUP_MSG:   version = 2.10.0-amzn-0
STARTUP_MSG:   classpath = /usr/lib/hadoop/etc/hadoop:/usr/lib/hadoop/etc/hadoop:/us
r/lib/hadoop/etc/hadoop:/usr/lib/hadoop/lib/audience-annotations-0.5.0.jar:/usr/lib/
hadoop/lib/jackson-jaxrs-1.9.13.jar:/usr/lib/hadoop/lib/commons-collections-3.2.2.ja
r:/usr/lib/hadoop/lib/mockito-all-1.8.5.jar:/usr/lib/hadoop/lib/spotbugs-annotations
-3.1.9.jar:/usr/lib/hadoop/lib/stax2-api-3.1.4.jar:/usr/lib/hadoop/lib/curator-clien
t-2.7.1.jar:/usr/lib/hadoop/lib/commons-lang3-3.4.jar:/usr/lib/hadoop/lib/jcip-annot
ations-1.0-1.jar:/usr/lib/hadoop/lib/jetty-util-6.1.26-emr.jar:/usr/lib/hadoop/lib/j
axb-api-2.2.2.jar:/usr/lib/hadoop/lib/asm-3.2.jar:/usr/lib/hadoop/lib/jettison-1.1.j

```

ar:/usr/lib/hadoop/lib/paranamer-2.3.jar:/usr/lib/hadoop/lib/java-xmlbuilder-0.4.jar:/usr/lib/hadoop/lib/jaxb-impl-2.2.3-1.jar:/usr/lib/hadoop/lib/avro-1.7.7.jar:/usr/lib/hadoop/lib/curator-framework-2.7.1.jar:/usr/lib/hadoop/lib/jersey-json-1.9.jar:/usr/lib/hadoop/lib/curator-recipes-2.7.1.jar:/usr/lib/hadoop/lib/log4j-1.2.17.jar:/usr/lib/hadoop/lib/apacheds-kerberos-codec-2.0.0-M15.jar:/usr/lib/hadoop/lib/jets3t-0.9.0.jar:/usr/lib/hadoop/lib/jackson-xc-1.9.13.jar:/usr/lib/hadoop/lib/httpcore-4.4.11.jar:/usr/lib/hadoop/lib/guava-11.0.2.jar:/usr/lib/hadoop/lib/jsp-api-2.1.jar:/usr/lib/hadoop/lib/xmlenc-0.52.jar:/usr/lib/hadoop/lib/servlet-api-2.5.jar:/usr/lib/hadoop/lib/commons-codec-1.4.jar:/usr/lib/hadoop/lib/slf4j-api-1.7.25.jar:/usr/lib/hadoop/lib/api-util-1.0.0-M20.jar:/usr/lib/hadoop/lib/commons-math3-3.1.1.jar:/usr/lib/hadoop/lib/activation-1.1.jar:/usr/lib/hadoop/lib/zookeeper-3.4.14.jar:/usr/lib/hadoop/lib/stax-api-1.0-2.jar:/usr/lib/hadoop/lib/commons-configuration-1.6.jar:/usr/lib/hadoop/lib/jersey-core-1.9.jar:/usr/lib/hadoop/lib/commons-beanutils-1.9.4.jar:/usr/lib/hadoop/lib/hamcrest-core-1.3.jar:/usr/lib/hadoop/lib/json-smart-1.3.1.jar:/usr/lib/hadoop/lib/commons-logging-1.1.3.jar:/usr/lib/hadoop/lib/jetty-sslengine-6.1.26-emr.jar:/usr/lib/hadoop/lib/commons-compress-1.19.jar:/usr/lib/hadoop/lib/commons-io-2.4.jar:/usr/lib/hadoop/lib/jackson-core-asl-1.9.13.jar:/usr/lib/hadoop/lib/httpclient-4.5.9.jar:/usr/lib/hadoop/lib/netty-3.10.6.Final.jar:/usr/lib/hadoop/lib/commons-lang-2.6.jar:/usr/lib/hadoop/lib/jetty-6.1.26-emr.jar:/usr/lib/hadoop/lib/jackson-mapper-asl-1.9.13.jar:/usr/lib/hadoop/lib/commons-digester-1.8.jar:/usr/lib/hadoop/lib/protobuf-java-2.5.0.jar:/usr/lib/hadoop/lib/jsch-0.1.54.jar:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.25.jar:/usr/lib/hadoop/lib/jersey-server-1.9.jar:/usr/lib/hadoop/lib/commons-cli-1.2.jar:/usr/lib/hadoop/lib/apacheds-i18n-2.0.0-M15.jar:/usr/lib/hadoop/lib/woodstox-core-5.0.3.jar:/usr/lib/hadoop/lib/htrace-core4-4.1.0-incubating.jar:/usr/lib/hadoop/lib/commons-net-3.1.jar:/usr/lib/hadoop/lib/snappy-java-1.1.7.3.jar:/usr/lib/hadoop/lib/api-asn1-api-1.0.0-M20.jar:/usr/lib/hadoop/lib/gson-2.2.4.jar:/usr/lib/hadoop/lib/nimbus-jose-jwt-4.41.1.jar:/usr/lib/hadoop/lib/jsr305-3.0.0.jar:/usr/lib/hadoop/lib/junit-4.11.jar:/usr/lib/hadoop/./hadoop-openstack-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-azure-datalake.jar:/usr/lib/hadoop/./hadoop-distcp-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-aliyun.jar:/usr/lib/hadoop/./hadoop-ant-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-azure.jar:/usr/lib/hadoop/./hadoop-nfs-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-auth.jar:/usr/lib/hadoop/./hadoop-yarn-common-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-aliyun-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-yarn-server-resourcemanager.jar:/usr/lib/hadoop/./hadoop-aws.jar:/usr/lib/hadoop/./hadoop-azure-datalake-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-rumen.jar:/usr/lib/hadoop/./hadoop-extras.jar:/usr/lib/hadoop/./hadoop-yarn-server-applicationhistoryservice-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-annotations-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-archives.jar:/usr/lib/hadoop/./hadoop-gridmix-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-auth-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-yarn-server-resourcemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-yarn-server-applicationhistoryservice.jar:/usr/lib/hadoop/./hadoop-annotations.jar:/usr/lib/hadoop/./hadoop-yarn-registry-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-extras-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-yarn-common.jar:/usr/lib/hadoop/./hadoop-gridmix.jar:/usr/lib/hadoop/./hadoop-streaming-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-distcp.jar:/usr/lib/hadoop/./hadoop-yarn-server-common-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-yarn-server-web-proxy.jar:/usr/lib/hadoop/./hadoop-aws-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-common-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop/./hadoop-resourceestimator.jar:/usr/lib/hadoop/./hadoop-openstack.jar:/usr/lib/hadoop/./hadoop-yarn-server-web-proxy-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-archive-logs.jar:/usr/lib/hadoop/./hadoop-nfs.jar:/usr/lib/hadoop/./hadoop-yarn-registry.jar:/usr/lib/hadoop/./hadoop-common-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-streaming.jar:/usr/lib/hadoop/./hadoop-yarn-api-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-yarn-server-common.jar:/usr/lib/hadoop/./hadoop-datajoin-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-common.jar:/usr/lib/hadoop/./hadoop-op-sls.jar:/usr/lib/hadoop/./hadoop-rumen-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-azure-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-archives-2.10.0-amzn-0.jar:/usr

r/lib/hadoop/./hadoop-ant.jar:/usr/lib/hadoop/./hadoop-sls-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-yarn-api.jar:/usr/lib/hadoop/./hadoop-datajoin.jar:/usr/lib/hadoop/./hadoop-archive-logs-2.10.0-amzn-0.jar:/usr/lib/hadoop/./hadoop-resourceestimator-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./:/usr/lib/hadoop-hdfs/lib/jetty-util-6.1.26-emr.jar:/usr/lib/hadoop-hdfs/lib/asm-3.2.jar:/usr/lib/hadoop-hdfs/lib/okio-1.6.0.jar:/usr/lib/hadoop-hdfs/lib/commons-daemon-1.0.13.jar:/usr/lib/hadoop-hdfs/lib/log4j-1.2.17.jar:/usr/lib/hadoop-hdfs/lib/guava-11.0.2.jar:/usr/lib/hadoop-hdfs/lib/xmlenc-0.52.jar:/usr/lib/hadoop-hdfs/lib/xercesImpl-2.12.0.jar:/usr/lib/hadoop-hdfs/lib/servlet-api-2.5.jar:/usr/lib/hadoop-hdfs/lib/commons-codec-1.4.jar:/usr/lib/hadoop-hdfs/lib/xml-apis-1.4.01.jar:/usr/lib/hadoop-hdfs/lib/jersey-core-1.9.jar:/usr/lib/hadoop-hdfs/lib/jackson-core-2.6.7.jar:/usr/lib/hadoop-hdfs/lib/commons-logging-1.1.3.jar:/usr/lib/hadoop-hdfs/lib/commons-io-2.4.jar:/usr/lib/hadoop-hdfs/lib/jackson-core-asl-1.9.13.jar:/usr/lib/hadoop-hdfs/lib/netty-3.10.6.Final.jar:/usr/lib/hadoop-hdfs/lib/commons-lang-2.6.jar:/usr/lib/hadoop-hdfs/lib/jetty-6.1.26-emr.jar:/usr/lib/hadoop-hdfs/lib/jackson-annotations-2.6.7.jar:/usr/lib/hadoop-hdfs/lib/jackson-mapper-asl-1.9.13.jar:/usr/lib/hadoop-hdfs/lib/protobuf-java-2.5.0.jar:/usr/lib/hadoop-hdfs/lib/jersey-server-1.9.jar:/usr/lib/hadoop-hdfs/lib/commons-cli-1.2.jar:/usr/lib/hadoop-hdfs/lib/htrace-core4-4.1.0-incubating.jar:/usr/lib/hadoop-hdfs/lib/leveldb-jni-all-1.8.jar:/usr/lib/hadoop-hdfs/lib/jackson-databind-2.6.7.jar:/usr/lib/hadoop-hdfs/lib/okhttp-2.7.5.jar:/usr/lib/hadoop-hdfs/lib/jsr305-3.0.0.jar:/usr/lib/hadoop-hdfs/lib/netty-all-4.0.23.Final.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-client-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-native-client.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-rbf-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-rbf-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-client-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-native-client-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-nfs.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-client.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-nfs-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-rbf.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-native-client-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop-yarn/lib/audience-annotations-0.5.0.jar:/usr/lib/hadoop-yarn/lib/javax.inject-1.jar:/usr/lib/hadoop-yarn/lib/jackson-jaxrs-1.9.13.jar:/usr/lib/hadoop-yarn/lib/commons-collections-3.2.2.jar:/usr/lib/hadoop-yarn/lib/spotbugs-annotations-3.1.9.jar:/usr/lib/hadoop-yarn/lib/stax2-api-3.1.4.jar:/usr/lib/hadoop-yarn/lib/curator-client-2.7.1.jar:/usr/lib/hadoop-yarn/lib/commons-lang3-3.4.jar:/usr/lib/hadoop-yarn/lib/jcip-annotations-1.0-1.jar:/usr/lib/hadoop-yarn/lib/jetty-util-6.1.26-emr.jar:/usr/lib/hadoop-yarn/lib/jaxb-api-2.2.2.jar:/usr/lib/hadoop-yarn/lib/jersey-client-1.9.jar:/usr/lib/hadoop-yarn/lib/asm-3.2.jar:/usr/lib/hadoop-yarn/lib/jettison-1.1.jar:/usr/lib/hadoop-yarn/lib/paranamer-2.3.jar:/usr/lib/hadoop-yarn/lib/java-xmlbuilder-0.4.jar:/usr/lib/hadoop-yarn/lib/jaxb-impl-2.2.3-1.jar:/usr/lib/hadoop-yarn/lib/avro-1.7.7.jar:/usr/lib/hadoop-yarn/lib/ehcache-3.3.1.jar:/usr/lib/hadoop-yarn/lib/aopalliance-1.0.jar:/usr/lib/hadoop-yarn/lib/curator-framework-2.7.1.jar:/usr/lib/hadoop-yarn/lib/jersey-json-1.9.jar:/usr/lib/hadoop-yarn/lib/curator-recipes-2.7.1.jar:/usr/lib/hadoop-yarn/lib/log4j-1.2.17.jar:/usr/lib/hadoop-yarn/lib/apacheds-kerberos-codec-2.0.0-M15.jar:/usr/lib/hadoop-yarn/lib/jets3t-0.9.0.jar:/usr/lib/hadoop-yarn/lib/jackson-xc-1.9.13.jar:/usr/lib/hadoop-yarn/lib/httpcore-4.4.11.jar:/usr/lib/hadoop-yarn/lib/guava-11.0.2.jar:/usr/lib/hadoop-yarn/lib/jsp-api-2.1.jar:/usr/lib/hadoop-yarn/lib/HikariCP-java7-2.4.12.jar:/usr/lib/hadoop-yarn/lib/xmlenc-0.52.jar:/usr/lib/hadoop-yarn/lib/servlet-api-2.5.jar:/usr/lib/hadoop-yarn/lib/commons-codec-1.4.jar:/usr/lib/hadoop-yarn/lib/api-util-1.0.0-M20.jar:/usr/lib/hadoop-yarn/lib/commons-math3-3.1.1.jar:/usr/lib/hadoop-yarn/lib/guice-servlet-3.0.jar:/usr/lib/hadoop-yarn/lib/activation-1.1.jar:/usr/lib/hadoop-yarn/lib/zookeeper-3.4.14.jar:/usr/lib/hadoop-yarn/lib/stax-api-1.0-2.jar:/usr/lib/hadoop-yarn/lib/commons-configuration-1.6.jar:/usr/lib/hadoop-yarn/lib/jersey-core-1.9.jar:/usr/lib/hadoop-yarn/lib/fst-2.50.jar:/usr/lib/hadoop-yarn/lib/commons-beanutils-1.9.4.jar:/usr/lib/hadoop-yarn/lib/json-smart-1.3.1.jar:/usr/lib/hadoop-yarn/lib/commons-logging-1.1.3.jar:/usr/lib/hadoop-yarn

n/lib/jetty-sslengine-6.1.26-emr.jar:/usr/lib/hadoop-yarn/lib/commons-compress-1.19.  
 jar:/usr/lib/hadoop-yarn/lib/commons-io-2.4.jar:/usr/lib/hadoop-yarn/lib/jackson-core-  
 asl-1.9.13.jar:/usr/lib/hadoop-yarn/lib/httpclient-4.5.9.jar:/usr/lib/hadoop-yarn/  
 lib/netty-3.10.6.Final.jar:/usr/lib/hadoop-yarn/lib/commons-lang-2.6.jar:/usr/lib/hadoop-  
 yarn/lib/jetty-6.1.26-emr.jar:/usr/lib/hadoop-yarn/lib/jackson-mapper-asl-1.9.1  
 3.jar:/usr/lib/hadoop-yarn/lib/geronimo-jcache\_1.0\_spec-1.0-alpha-1.jar:/usr/lib/hadoop-  
 yarn/lib/mssql-jdbc-6.2.1.jre7.jar:/usr/lib/hadoop-yarn/lib/commons-digester-1.  
 8.jar:/usr/lib/hadoop-yarn/lib/protobuf-java-2.5.0.jar:/usr/lib/hadoop-yarn/lib/jsch-  
 0.1.54.jar:/usr/lib/hadoop-yarn/lib/jersey-server-1.9.jar:/usr/lib/hadoop-yarn/lib/  
 commons-cli-1.2.jar:/usr/lib/hadoop-yarn/lib/apacheds-i18n-2.0.0-M15.jar:/usr/lib/hadoop-  
 yarn/lib/woodstox-core-5.0.3.jar:/usr/lib/hadoop-yarn/lib/htrace-core4-4.1.0-incubating.jar:/usr/lib/hadoop-yarn/lib/commons-net-3.1.jar:/usr/lib/hadoop-yarn/lib/json-io-2.5.1.jar:/usr/lib/hadoop-yarn/lib/jersey-guice-1.9.jar:/usr/lib/hadoop-yarn/lib/snappy-java-1.1.7.3.jar:/usr/lib/hadoop-yarn/lib/leveldbjni-all-1.8.jar:/usr/lib/hadoop-yarn/lib/api-asn1-api-1.0.0-M20.jar:/usr/lib/hadoop-yarn/lib/gson-2.2.4.jar:/usr/lib/hadoop-yarn/lib/nimbus-jose-jwt-4.41.1.jar:/usr/lib/hadoop-yarn/lib/jsr305-3.0.0.jar:/usr/lib/hadoop-yarn/lib/metrics-core-3.0.1.jar:/usr/lib/hadoop-yarn/lib/guice-3.0.jar:/usr/lib/hadoop-yarn/lib/java-util-1.9.0.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-server-tests-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-server-nodemanager.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-client-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-server-timeline-pluginstorage-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-server-tests.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-applications-unmanaged-am-launcher.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-server-router-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-server-resourcemanager.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-server-applicationhistoryservice-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-server-resourcemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-server-applicationhistoryservice.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-applications-distributedshell-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-applications-unmanaged-am-launcher-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-registry-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-common.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-server-sharedcachemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-client.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-server-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-server-web-proxy.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-server-sharedcachemanager.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-server-web-proxy-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-server-router.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-registry.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-api-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-server-timeline-pluginstorage.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-server-common.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-applications-distributedshell.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-api.jar:/usr/lib/hadoop-yarn/lib/hadoop-yarn-server-nodemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/lib/javax.inject-1.jar:/usr/lib/hadoop-mapreduce/lib/asm-3.2.jar:/usr/lib/hadoop-mapreduce/lib/paranamer-2.3.jar:/usr/lib/hadoop-mapreduce/lib/avro-1.7.7.jar:/usr/lib/hadoop-mapreduce/lib/aopalliance-1.0.jar:/usr/lib/hadoop-mapreduce/lib/log4j-1.2.17.jar:/usr/lib/hadoop-mapreduce/lib/guice-servlet-3.0.jar:/usr/lib/hadoop-mapreduce/lib/jersey-core-1.9.jar:/usr/lib/hadoop-mapreduce/lib/hamcrest-core-1.3.jar:/usr/lib/hadoop-mapreduce/lib/commons-compress-1.19.jar:/usr/lib/hadoop-mapreduce/lib/commons-io-2.4.jar:/usr/lib/hadoop-mapreduce/lib/jackson-core-asl-1.9.13.jar:/usr/lib/hadoop-mapreduce/lib/netty-3.10.6.Final.jar:/usr/lib/hadoop-mapreduce/lib/jackson-mapper-asl-1.9.13.jar:/usr/lib/hadoop-mapreduce/lib/protobuf-java-2.5.0.jar:/usr/lib/hadoop-mapreduce/lib/jersey-server-1.9.jar:/usr/lib/hadoop-mapreduce/lib/jersey-guice-1.9.jar:/usr/lib/hadoop-mapreduce/lib/snappy-java-1.1.7.3.jar:/usr/lib/hadoop-mapreduce/lib/leveldbjni-all-1.8.jar:/usr/lib/hadoop-mapreduce/lib/junit-4.11.jar:/usr/lib/hadoop-mapreduce/lib/guice-3.0.jar:/usr/lib/hadoop-mapreduce/lib/audience-annotations-0.5.0.jar:/usr/lib/hadoop-mapreduce/lib/javax.inject-1.jar:/usr/lib/hadoop-mapreduce/lib/hadoop-openstack-2.10.0-amzn-0.

jar:/usr/lib/hadoop-mapreduce/./hadoop-azure-datalake.jar:/usr/lib/hadoop-mapreduce/./hadoop-distcp-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jackson-jaxrs-1.9.13.jar:/usr/lib/hadoop-mapreduce/./commons-collections-3.2.2.jar:/usr/lib/hadoop-mapreduce/./hadoop-aliyun.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-hs-plugins-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-ant-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./azure-storage-5.4.0.jar:/usr/lib/hadoop-mapreduce/./spotbugs-annotations-3.1.9.jar:/usr/lib/hadoop-mapreduce/./stax2-api-3.1.4.jar:/usr/lib/hadoop-mapreduce/./hadoop-azure.jar:/usr/lib/hadoop-mapreduce/./curator-client-2.7.1.jar:/usr/lib/hadoop-mapreduce/./commons-lang3-3.4.jar:/usr/lib/hadoop-mapreduce/./jcip-annotations-1.0-1.jar:/usr/lib/hadoop-mapreduce/./jetty-util-6.1.26-emr.jar:/usr/lib/hadoop-mapreduce/./jaxb-api-2.2.2.jar:/usr/lib/hadoop-mapreduce/./jersey-client-1.9.jar:/usr/lib/hadoop-mapreduce/./azure-keyvault-core-0.8.0.jar:/usr/lib/hadoop-mapreduce/./asm-3.2.jar:/usr/lib/hadoop-mapreduce/./jettison-1.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-auth.jar:/usr/lib/hadoop-mapreduce/./ojslgo-4.3.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-common.jar:/usr/lib/hadoop-mapreduce/./paranamer-2.3.jar:/usr/lib/hadoop-mapreduce/./java-xmlbuilder-0.4.jar:/usr/lib/hadoop-mapreduce/./hadoop-aliyun-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-resourcemanager.jar:/usr/lib/hadoop-mapreduce/./jaxb-impl-2.2.3-1.jar:/usr/lib/hadoop-mapreduce/./avro-1.7.7.jar:/usr/lib/hadoop-mapreduce/./aliyun-java-sdk-sts-3.0.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-aws.jar:/usr/lib/hadoop-mapreduce/./hadoop-azure-datalake-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-rumen.jar:/usr/lib/hadoop-mapreduce/./ehcache-3.3.1.jar:/usr/lib/hadoop-mapreduce/./aopalliance-1.0.jar:/usr/lib/hadoop-mapreduce/./aliyun-java-sdk-ram-3.0.0.jar:/usr/lib/hadoop-mapreduce/./curator-framework-2.7.1.jar:/usr/lib/hadoop-mapreduce/./jersey-json-1.9.jar:/usr/lib/hadoop-mapreduce/./hadoop-extras.jar:/usr/lib/hadoop-mapreduce/./curator-recipes-2.7.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-app.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-applicationhistoryservice-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./log4j-1.2.17.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-jobclient.jar:/usr/lib/hadoop-mapreduce/./apacheds-kerberos-codec-2.0.0-M15.jar:/usr/lib/hadoop-mapreduce/./jets3t-0.9.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-archives.jar:/usr/lib/hadoop-mapreduce/./jackson-xc-1.9.13.jar:/usr/lib/hadoop-mapreduce/./hadoop-gridmix-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./httpcore-4.4.11.jar:/usr/lib/hadoop-mapreduce/./guava-11.0.2.jar:/usr/lib/hadoop-mapreduce/./hadoop-auth-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jsp-api-2.1.jar:/usr/lib/hadoop-mapreduce/./aliyun-sdk-oss-3.4.1.jar:/usr/lib/hadoop-mapreduce/./HikariCP-java7-2.4.12.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-examples.jar:/usr/lib/hadoop-mapreduce/./xmlenc-0.52.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-resourcemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-applicationhistoryservice.jar:/usr/lib/hadoop-mapreduce/./servlet-api-2.5.jar:/usr/lib/hadoop-mapreduce/./commons-codec-1.4.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-jobclient-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./api-util-1.0.0-M20.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-registry-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-extras-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./commons-math3-3.1.1.jar:/usr/lib/hadoop-mapreduce/./guice-servlet-3.0.jar:/usr/lib/hadoop-mapreduce/./activation-1.1.jar:/usr/lib/hadoop-mapreduce/./aws-java-sdk-bundle-1.11.852.jar:/usr/lib/hadoop-mapreduce/./zookeeper-3.4.14.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-core.jar:/usr/lib/hadoop-mapreduce/./stax-api-1.0-2.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-common.jar:/usr/lib/hadoop-mapreduce/./commons-configuration-1.6.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-app-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jersey-core-1.9.jar:/usr/lib/hadoop-mapreduce/./fst-2.50.jar:/usr/lib/hadoop-mapreduce/./commons-beanutils-1.9.4.jar:/usr/lib/hadoop-mapreduce/./json-smart-1.3.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-gridmix.jar:/usr/lib/hadoop-mapreduce/./jackson-core-2.6.7.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-hs-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./commo

ns-logging-1.1.3.jar:/usr/lib/hadoop-mapreduce/./hadoop-streaming-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-distcp.jar:/usr/lib/hadoop-mapreduce/./jetty-sslengine-6.1.26-emr.jar:/usr/lib/hadoop-mapreduce/./commons-compress-1.19.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-web-proxy.jar:/usr/lib/hadoop-mapreduce/./hadoop-aws-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./commons-io-2.4.jar:/usr/lib/hadoop-mapreduce/./hadoop-resourceestimator.jar:/usr/lib/hadoop-mapreduce/./jackson-core-asl-1.9.13.jar:/usr/lib/hadoop-mapreduce/./httpclient-4.5.9.jar:/usr/lib/hadoop-mapreduce/./netty-3.10.6.Final.jar:/usr/lib/hadoop-mapreduce/./commons-lang-2.6.jar:/usr/lib/hadoop-mapreduce/./hadoop-openstack.jar:/usr/lib/hadoop-mapreduce/./jetty-6.1.26-emr.jar:/usr/lib/hadoop-mapreduce/./jackson-annotations-2.6.7.jar:/usr/lib/hadoop-mapreduce/./jackson-mapper-asl-1.9.13.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-web-proxy-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./aliyun-java-sdk-ecs-4.2.0.jar:/usr/lib/hadoop-mapreduce/./geronimo-jcache\_1.0\_spec-1.0-alpha-1.jar:/usr/lib/hadoop-mapreduce/./hadoop-archive-logs.jar:/usr/lib/hadoop-mapreduce/./mssql-jdbc-6.2.1.jre7.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-registry.jar:/usr/lib/hadoop-mapreduce/./commons-digester-1.8.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-hs-plugins.jar:/usr/lib/hadoop-mapreduce/./protobuf-java-2.5.0.jar:/usr/lib/hadoop-mapreduce/./jsch-0.1.54.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-hs.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-core-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jdom-1.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jersey-server-1.9.jar:/usr/lib/hadoop-mapreduce/./commons-cli-1.2.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-shuffle.jar:/usr/lib/hadoop-mapreduce/./hadoop-streaming.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-api-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./commons-httpclient-3.1.jar:/usr/lib/hadoop-mapreduce/./aliyun-java-sdk-core-3.4.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-common.jar:/usr/lib/hadoop-mapreduce/./apacheds-i18n-2.0.0-M15.jar:/usr/lib/hadoop-mapreduce/./woodstox-core-5.0.3.jar:/usr/lib/hadoop-mapreduce/./htrace-core4-4.1.0-incubating.jar:/usr/lib/hadoop-mapreduce/./hadoop-datajoin-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-examples-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./commons-net-3.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-sls.jar:/usr/lib/hadoop-mapreduce/./json-io-2.5.1.jar:/usr/lib/hadoop-mapreduce/./jersey-guice-1.9.jar:/usr/lib/hadoop-mapreduce/./hadoop-rumen-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./snappy-java-1.1.7.3.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-jobclient-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop-mapreduce/./leveldbjni-all-1.8.jar:/usr/lib/hadoop-mapreduce/./hadoop-azure-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./api-asn1-api-1.0.0-M20.jar:/usr/lib/hadoop-mapreduce/./jackson-databind-2.6.7.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-shuffle-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./gson-2.2.4.jar:/usr/lib/hadoop-mapreduce/./nimbus-jose-jwt-4.41.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-archives-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jsr305-3.0.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-ant.jar:/usr/lib/hadoop-mapreduce/./hadoop-sls-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-api.jar:/usr/lib/hadoop-mapreduce/./hadoop-datajoin.jar:/usr/lib/hadoop-mapreduce/./metrics-core-3.0.1.jar:/usr/lib/hadoop-mapreduce/./azure-data-lake-store-sdk-2.2.3.jar:/usr/lib/hadoop-mapreduce/./hadoop-archive-logs-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./guice-3.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-resourceestimator-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./java-util-1.9.0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-tests-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-nodemanager.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-client-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-timeline-pluginstorage-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-tests.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-applications-unmanaged-am-launcher.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-router-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-resourcemanager.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-applicationhistoryservice

e-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-resourcemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-applicationhistoryservice.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-applications-distributedshell-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-applications-unmanaged-am-launcher-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-registry-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-common.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-sharedcachemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-client.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-web-proxy.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-sharedcachemanager.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-web-proxy-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-router.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-registry.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-api-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-timeline-pluginstorage.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-common.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-applications-distributedshell.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-api.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-nodemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/lib/audience-annotations-0.5.0.jar:/usr/lib/hadoop-yarn/lib/javax.inject-1.jar:/usr/lib/hadoop-yarn/lib/jackson-jaxrs-1.9.13.jar:/usr/lib/hadoop-yarn/lib/commons-collections-3.2.2.jar:/usr/lib/hadoop-yarn/lib/spotbugs-annotations-3.1.9.jar:/usr/lib/hadoop-yarn/lib/stax2-api-3.1.4.jar:/usr/lib/hadoop-yarn/lib/curator-client-2.7.1.jar:/usr/lib/hadoop-yarn/lib/commons-lang3-3.4.jar:/usr/lib/hadoop-yarn/lib/jcip-annotations-1.0-1.jar:/usr/lib/hadoop-yarn/lib/jetty-util-6.1.26-emr.jar:/usr/lib/hadoop-yarn/lib/jaxb-api-2.2.2.jar:/usr/lib/hadoop-yarn/lib/jersey-client-1.9.jar:/usr/lib/hadoop-yarn/lib/asm-3.2.jar:/usr/lib/hadoop-yarn/lib/jettison-1.1.jar:/usr/lib/hadoop-yarn/lib/paranamer-2.3.jar:/usr/lib/hadoop-yarn/lib/java-xmlbuilder-0.4.jar:/usr/lib/hadoop-yarn/lib/jaxb-impl-2.2.3-1.jar:/usr/lib/hadoop-yarn/lib/avro-1.7.7.jar:/usr/lib/hadoop-yarn/lib/ehcache-3.3.1.jar:/usr/lib/hadoop-yarn/lib/aopalliance-1.0.jar:/usr/lib/hadoop-yarn/lib/curator-framework-2.7.1.jar:/usr/lib/hadoop-yarn/lib/jersey-json-1.9.jar:/usr/lib/hadoop-yarn/lib/curator-recipes-2.7.1.jar:/usr/lib/hadoop-yarn/lib/log4j-1.2.17.jar:/usr/lib/hadoop-yarn/lib/apacheds-kerberos-codec-2.0.0-M15.jar:/usr/lib/hadoop-yarn/lib/jets3t-0.9.0.jar:/usr/lib/hadoop-yarn/lib/jackson-xc-1.9.13.jar:/usr/lib/hadoop-yarn/lib/httpcore-4.4.11.jar:/usr/lib/hadoop-yarn/lib/guava-11.0.2.jar:/usr/lib/hadoop-yarn/lib/jsp-api-2.1.jar:/usr/lib/hadoop-yarn/lib/HikariCP-java7-2.4.12.jar:/usr/lib/hadoop-yarn/lib/xmlenc-0.52.jar:/usr/lib/hadoop-yarn/lib/servlet-api-2.5.jar:/usr/lib/hadoop-yarn/lib/commons-codec-1.4.jar:/usr/lib/hadoop-yarn/lib/api-util-1.0.0-M20.jar:/usr/lib/hadoop-yarn/lib/commons-math3-3.1.1.jar:/usr/lib/hadoop-yarn/lib/guice-servlet-3.0.jar:/usr/lib/hadoop-yarn/lib/activation-1.1.jar:/usr/lib/hadoop-yarn/lib/zookeeper-3.4.14.jar:/usr/lib/hadoop-yarn/lib/stax-api-1.0-2.jar:/usr/lib/hadoop-yarn/lib/commons-configuration-1.6.jar:/usr/lib/hadoop-yarn/lib/jersey-core-1.9.jar:/usr/lib/hadoop-yarn/lib/fst-2.50.jar:/usr/lib/hadoop-yarn/lib/commons-beanutils-1.9.4.jar:/usr/lib/hadoop-yarn/lib/json-smart-1.3.1.jar:/usr/lib/hadoop-yarn/lib/commons-logging-1.1.3.jar:/usr/lib/hadoop-yarn/lib/jetty-sslengine-6.1.26-emr.jar:/usr/lib/hadoop-yarn/lib/commons-compress-1.19.jar:/usr/lib/hadoop-yarn/lib/commons-io-2.4.jar:/usr/lib/hadoop-yarn/lib/jackson-core-asl-1.9.13.jar:/usr/lib/hadoop-yarn/lib/httpclient-4.5.9.jar:/usr/lib/hadoop-yarn/lib/netty-3.10.6.Final.jar:/usr/lib/hadoop-yarn/lib/commons-lang-2.6.jar:/usr/lib/hadoop-yarn/lib/jetty-6.1.26-emr.jar:/usr/lib/hadoop-yarn/lib/jackson-mapper-asl-1.9.13.jar:/usr/lib/hadoop-yarn/lib/geronimo-jcache\_1.0\_spec-1.0-alpha-1.jar:/usr/lib/hadoop-yarn/lib/msql-jdbc-6.2.1.jre7.jar:/usr/lib/hadoop-yarn/lib/commons-digester-1.8.jar:/usr/lib/hadoop-yarn/lib/protobuf-java-2.5.0.jar:/usr/lib/hadoop-yarn/lib/jsch-0.1.54.jar:/usr/lib/hadoop-yarn/lib/jersey-server-1.9.jar:/usr/lib/hadoop-yarn/lib/commons-cli-1.2.jar:/usr/lib/hadoop-yarn/lib/apacheds-i18n-2.0.0-M15.jar:/usr/lib/hadoop-yarn/lib/woodstox-core-5.0.3.jar:/usr/lib/hadoop-yarn/lib/htrace-core4-4.1.0-incubating.jar:/usr/lib/hadoop-yarn/lib/commons-net-3.1.jar:/usr/lib/hadoop-yarn/lib/json-io-2.5.1.jar:/usr/lib/hadoop-yarn/lib/jersey-guice-1.9.jar:/usr/lib/hadoop-yarn/lib/snappy-java-1.1.7.3.jar:/usr/lib/hadoop-yarn/lib/leveldbjni-all-1.8.jar:/usr/lib/hadoop-yarn/lib



```

ib/api-asn1-api-1.0.0-M20.jar:/usr/lib/hadoop-yarn/lib/gson-2.2.4.jar:/usr/lib/hadoo
p-yarn/lib/nimbus-jose-jwt-4.41.1.jar:/usr/lib/hadoop-yarn/lib/jsr305-3.0.0.jar:/us
r/lib/hadoop-yarn/lib/metrics-core-3.0.1.jar:/usr/lib/hadoop-yarn/lib/guice-3.0.ja
r:/usr/lib/hadoop-yarn/lib/java-util-1.9.0.jar:/usr/lib/hadoop/etc/hadoop/nm-config/
log4j.properties:/usr/lib/hadoop-yarn/./timelineservice/hadoop-yarn-server-timeline
service-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./timelineservice/hadoop-yarn-server
-timelineservice-hbase-coprocessor-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./timelin
eservice/hadoop-yarn-server-timelineservice-hbase-client-2.10.0-amzn-0.jar:/usr/lib/
hadoop-yarn/./timelineservice/hadoop-yarn-server-timelineservice-hbase-common-2.10.
0-amzn-0.jar:/usr/lib/hadoop-yarn/./timelineservice/lib/jsr311-api-1.1.1.jar:/usr/l
ib/hadoop-yarn/./timelineservice/lib/hbase-common-1.2.6.jar:/usr/lib/hadoop-yar
n/./timelineservice/lib/jcodings-1.0.8.jar:/usr/lib/hadoop-yarn/./timelineservice/
lib/metrics-core-2.2.0.jar:/usr/lib/hadoop-yarn/./timelineservice/lib/commons-csv-
1.0.jar:/usr/lib/hadoop-yarn/./timelineservice/lib/jackson-core-2.6.7.jar:/usr/lib/
hadoop-yarn/./timelineservice/lib/hbase-client-1.2.6.jar:/usr/lib/hadoop-yarn/./ti
melineservice/lib/joni-2.1.2.jar:/usr/lib/hadoop-yarn/./timelineservice/lib/htrace-
core-3.1.0-incubating.jar:/usr/lib/hadoop-yarn/./timelineservice/lib/hbase-protocol
-1.2.6.jar:/usr/lib/hadoop-yarn/./timelineservice/lib/netty-all-4.0.23.Final.jar:/u
sr/lib/hadoop-yarn/./timelineservice/lib/hbase-annotations-1.2.6.jar
STARTUP_MSG:   build = git@aws157git.com:/pkg/Aws157BigTop -r d1e860a34cc1aea3d600c5
7c5c0270ea41579e8c; compiled by 'ec2-user' on 2020-09-19T02:05Z
STARTUP_MSG:   java = 1.8.0_312
*****/
24/04/28 21:20:47 INFO nodemanager.NodeManager: registered UNIX signal handlers for
[TERM, HUP, INT]
24/04/28 21:20:47 INFO datanode.DataNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting DataNode
STARTUP_MSG:   host = algo-1/10.0.222.169
STARTUP_MSG:   args = []
STARTUP_MSG:   version = 2.10.0-amzn-0
STARTUP_MSG:   classpath = /usr/lib/hadoop/etc/hadoop:/usr/lib/hadoop/lib/audience-a
nnotations-0.5.0.jar:/usr/lib/hadoop/lib/jackson-jaxrs-1.9.13.jar:/usr/lib/hadoop/li
b/commons-collections-3.2.2.jar:/usr/lib/hadoop/lib/mockito-all-1.8.5.jar:/usr/lib/h
adoop/lib/spotbugs-annotations-3.1.9.jar:/usr/lib/hadoop/lib/stax2-api-3.1.4.jar:/us
r/lib/hadoop/lib/curator-client-2.7.1.jar:/usr/lib/hadoop/lib/commons-lang3-3.4.ja
r:/usr/lib/hadoop/lib/jcip-annotations-1.0-1.jar:/usr/lib/hadoop/lib/jetty-util-6.1.
26-emr.jar:/usr/lib/hadoop/lib/jaxb-api-2.2.2.jar:/usr/lib/hadoop/lib/asm-3.2.jar:/u
sr/lib/hadoop/lib/jettison-1.1.jar:/usr/lib/hadoop/lib/paranamer-2.3.jar:/usr/lib/ha
doo/lib/java-xmlbuilder-0.4.jar:/usr/lib/hadoop/lib/jaxb-impl-2.2.3-1.jar:/usr/lib/
hadoop/lib/avro-1.7.7.jar:/usr/lib/hadoop/lib/curator-framework-2.7.1.jar:/usr/lib/h
adoop/lib/jersey-json-1.9.jar:/usr/lib/hadoop/lib/curator-recipes-2.7.1.jar:/usr/li
b/hadoop/lib/log4j-1.2.17.jar:/usr/lib/hadoop/lib/apacheds-kerberos-codec-2.0.0-M15.
jar:/usr/lib/hadoop/lib/jets3t-0.9.0.jar:/usr/lib/hadoop/lib/jackson-xc-1.9.13.jar:/
usr/lib/hadoop/lib/httpcore-4.4.11.jar:/usr/lib/hadoop/lib/guava-11.0.2.jar:/usr/li
b/hadoop/lib/jsp-api-2.1.jar:/usr/lib/hadoop/lib/xmlenc-0.52.jar:/usr/lib/hadoop/li
b/servlet-api-2.5.jar:/usr/lib/hadoop/lib/commons-codec-1.4.jar:/usr/lib/hadoop/lib/
slf4j-api-1.7.25.jar:/usr/lib/hadoop/lib/api-util-1.0.0-M20.jar:/usr/lib/hadoop/lib/
commons-math3-3.1.1.jar:/usr/lib/hadoop/lib/activation-1.1.jar:/usr/lib/hadoop/lib/z
ookeeper-3.4.14.jar:/usr/lib/hadoop/lib/stax-api-1.0-2.jar:/usr/lib/hadoop/lib/commo
ns-configuration-1.6.jar:/usr/lib/hadoop/lib/jersey-core-1.9.jar:/usr/lib/hadoop/li
b/commons-beanutils-1.9.4.jar:/usr/lib/hadoop/lib/hamcrest-core-1.3.jar:/usr/lib/had
oop/lib/json-smart-1.3.1.jar:/usr/lib/hadoop/lib/commons-logging-1.1.3.jar:/usr/lib/
hadoop/lib/jetty-sslengine-6.1.26-emr.jar:/usr/lib/hadoop/lib/commons-compress-1.19.
jar:/usr/lib/hadoop/lib/commons-io-2.4.jar:/usr/lib/hadoop/lib/jackson-core-asl-1.9.
13.jar:/usr/lib/hadoop/lib/httpclient-4.5.9.jar:/usr/lib/hadoop/lib/netty-3.10.6.Fin

```

al.jar:/usr/lib/hadoop/lib/commons-lang-2.6.jar:/usr/lib/hadoop/lib/jetty-6.1.26-emr.jar:/usr/lib/hadoop/lib/jackson-mapper-asl-1.9.13.jar:/usr/lib/hadoop/lib/commons-digester-1.8.jar:/usr/lib/hadoop/lib/protobuf-java-2.5.0.jar:/usr/lib/hadoop/lib/jsc-h-0.1.54.jar:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.25.jar:/usr/lib/hadoop/lib/jersey-server-1.9.jar:/usr/lib/hadoop/lib/commons-cli-1.2.jar:/usr/lib/hadoop/lib/apacheds-i18n-2.0.0-M15.jar:/usr/lib/hadoop/lib/woodstox-core-5.0.3.jar:/usr/lib/hadoop/lib/htrace-core4-4.1.0-incubating.jar:/usr/lib/hadoop/lib/commons-net-3.1.jar:/usr/lib/hadoop/lib/snappy-java-1.1.7.3.jar:/usr/lib/hadoop/lib/api-asn1-api-1.0.0-M20.jar:/usr/lib/hadoop/lib/gson-2.2.4.jar:/usr/lib/hadoop/lib/nimbus-jose-jwt-4.41.1.jar:/usr/lib/hadoop/lib/jsr305-3.0.0.jar:/usr/lib/hadoop/lib/junit-4.11.jar:/usr/lib/hadoop/p//hadoop-openstack-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-azure-datalake.jar:/usr/lib/hadoop/p//hadoop-distcp-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-aliyun.jar:/usr/lib/hadoop/p//hadoop-ant-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-azure.jar:/usr/lib/hadoop/p//hadoop-nfs-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-auth.jar:/usr/lib/hadoop/p//hadoop-yarn-common-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-aliyun-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-yarn-server-resourcemanager.jar:/usr/lib/hadoop/p//hadoop-aws.jar:/usr/lib/hadoop/p//hadoop-azure-datalake-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-rumen.jar:/usr/lib/hadoop/p//hadoop-extras.jar:/usr/lib/hadoop/p//hadoop-yarn-server-applicationhistoryservice-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-annotations-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-archives.jar:/usr/lib/hadoop/p//hadoop-gridmix-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-auth-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-yarn-server-resourcemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-yarn-server-applicationhistoryservice.jar:/usr/lib/hadoop/p//hadoop-annotations.jar:/usr/lib/hadoop/p//hadoop-yarn-registry-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-extras-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-yarn-common.jar:/usr/lib/hadoop/p//hadoop-gridmix.jar:/usr/lib/hadoop/p//hadoop-streaming-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-distcp.jar:/usr/lib/hadoop/p//hadoop-yarn-server-common-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-yarn-server-web-proxy.jar:/usr/lib/hadoop/p//hadoop-aws-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-common-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop/p//hadoop-resourceestimator.jar:/usr/lib/hadoop/p//hadoop-openstack.jar:/usr/lib/hadoop/p//hadoop-yarn-server-web-proxy-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-archive-logs.jar:/usr/lib/hadoop/p//hadoop-nfs.jar:/usr/lib/hadoop/p//hadoop-yarn-registry.jar:/usr/lib/hadoop/p//hadoop-common-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-streaming.jar:/usr/lib/hadoop/p//hadoop-yarn-api-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-yarn-server-common.jar:/usr/lib/hadoop/p//hadoop-datajoin-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-common.jar:/usr/lib/hadoop/p//hadoop-sls.jar:/usr/lib/hadoop/p//hadoop-rumen-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-azure-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-archives-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-ant.jar:/usr/lib/hadoop/p//hadoop-sls-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-yarn-api.jar:/usr/lib/hadoop/p//hadoop-datajoin.jar:/usr/lib/hadoop/p//hadoop-archive-logs-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-resourceestimator-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/p//usr/lib/hadoop-hdfs/lib/jetty-util-6.1.26-emr.jar:/usr/lib/hadoop-hdfs/lib/asm-3.2.jar:/usr/lib/hadoop-hdfs/lib/okio-1.6.0.jar:/usr/lib/hadoop-hdfs/lib/commons-daemon-1.0.13.jar:/usr/lib/hadoop-hdfs/lib/log4j-1.2.17.jar:/usr/lib/hadoop-hdfs/lib/guava-11.0.2.jar:/usr/lib/hadoop-hdfs/lib/xmlenc-0.52.jar:/usr/lib/hadoop-hdfs/lib/xercesImpl-2.12.0.jar:/usr/lib/hadoop-hdfs/lib/servlet-api-2.5.jar:/usr/lib/hadoop-hdfs/lib/commons-codec-1.4.jar:/usr/lib/hadoop-hdfs/lib/xml-apis-1.4.01.jar:/usr/lib/hadoop-hdfs/lib/jersey-core-1.9.jar:/usr/lib/hadoop-hdfs/lib/jackson-core-2.6.7.jar:/usr/lib/hadoop-hdfs/lib/commons-logging-1.1.3.jar:/usr/lib/hadoop-hdfs/lib/commons-io-2.4.jar:/usr/lib/hadoop-hdfs/lib/jackson-core-asl-1.9.13.jar:/usr/lib/hadoop-hdfs/lib/netty-3.10.6.Final.jar:/usr/lib/hadoop-hdfs/lib/commons-lang-2.6.jar:/usr/lib/hadoop-hdfs/lib/jetty-6.1.26-emr.jar:/usr/lib/hadoop-hdfs/lib/jackson-annotations-2.6.7.jar:/usr/lib/hadoop-hdfs/lib/jackson-mapper-asl-1.9.13.jar:/usr/lib/hadoop-hdfs/lib/protobuf-java-2.5.0.jar:/usr/lib/hadoop-hdfs/lib/jersey-server-1.9.jar:/usr/lib/hadoop-hdfs/lib/commons-cli-1.2.jar:/usr/lib/hadoop-hdfs/lib/htrace-core4-4.1.0-in

cubating.jar:/usr/lib/hadoop-hdfs/lib/leveldbjni-all-1.8.jar:/usr/lib/hadoop-hdfs/lib/jackson-databind-2.6.7.jar:/usr/lib/hadoop-hdfs/lib/okhttp-2.7.5.jar:/usr/lib/hadoop-hdfs/lib/jsr305-3.0.0.jar:/usr/lib/hadoop-hdfs/lib/netty-all-4.0.23.Final.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-client-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-native-client.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-rbf-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-rbf-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-client-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-native-client-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-nfs.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-client.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-nfs-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-rbf.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-native-client-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop-yarn/lib/audience-annotations-0.5.0.jar:/usr/lib/hadoop-yarn/lib/javax.inject-1.jar:/usr/lib/hadoop-yarn/lib/jackson-jaxrs-1.9.13.jar:/usr/lib/hadoop-yarn/lib/commons-collections-3.2.2.jar:/usr/lib/hadoop-yarn/lib/spotbugs-annotations-3.1.9.jar:/usr/lib/hadoop-yarn/lib/stax2-api-3.1.4.jar:/usr/lib/hadoop-yarn/lib/curator-client-2.7.1.jar:/usr/lib/hadoop-yarn/lib/commons-lang3-3.4.jar:/usr/lib/hadoop-yarn/lib/jcip-annotations-1.0-1.jar:/usr/lib/hadoop-yarn/lib/jetty-util-6.1.26-emr.jar:/usr/lib/hadoop-yarn/lib/jaxb-api-2.2.2.jar:/usr/lib/hadoop-yarn/lib/jersey-client-1.9.jar:/usr/lib/hadoop-yarn/lib/asm-3.2.jar:/usr/lib/hadoop-yarn/lib/jettison-1.1.jar:/usr/lib/hadoop-yarn/lib/paranamer-2.3.jar:/usr/lib/hadoop-yarn/lib/java-xmlbuilder-0.4.jar:/usr/lib/hadoop-yarn/lib/jaxb-impl-2.2.3-1.jar:/usr/lib/hadoop-yarn/lib/avro-1.7.7.jar:/usr/lib/hadoop-yarn/lib/ehcache-3.3.1.jar:/usr/lib/hadoop-yarn/lib/aopalliance-1.0.jar:/usr/lib/hadoop-yarn/lib/curator-framework-2.7.1.jar:/usr/lib/hadoop-yarn/lib/jersey-json-1.9.jar:/usr/lib/hadoop-yarn/lib/curator-recipes-2.7.1.jar:/usr/lib/hadoop-yarn/lib/log4j-1.2.17.jar:/usr/lib/hadoop-yarn/lib/apacheds-kerberos-codec-2.0.0-M15.jar:/usr/lib/hadoop-yarn/lib/jets3t-0.9.0.jar:/usr/lib/hadoop-yarn/lib/jackson-xc-1.9.13.jar:/usr/lib/hadoop-yarn/lib/httpcore-4.4.11.jar:/usr/lib/hadoop-yarn/lib/guava-11.0.2.jar:/usr/lib/hadoop-yarn/lib/jsp-api-2.1.jar:/usr/lib/hadoop-yarn/lib/HikariCP-java7-2.4.12.jar:/usr/lib/hadoop-yarn/lib/xmlenc-0.52.jar:/usr/lib/hadoop-yarn/lib/servlet-api-2.5.jar:/usr/lib/hadoop-yarn/lib/commons-codec-1.4.jar:/usr/lib/hadoop-yarn/lib/api-util-1.0.0-M20.jar:/usr/lib/hadoop-yarn/lib/commons-math3-3.1.1.jar:/usr/lib/hadoop-yarn/lib/guice-servlet-3.0.jar:/usr/lib/hadoop-yarn/lib/activation-1.1.jar:/usr/lib/hadoop-yarn/lib/zookeeper-3.4.14.jar:/usr/lib/hadoop-yarn/lib/stax-api-1.0-2.jar:/usr/lib/hadoop-yarn/lib/commons-configuration-1.6.jar:/usr/lib/hadoop-yarn/lib/jersey-core-1.9.jar:/usr/lib/hadoop-yarn/lib/fst-2.50.jar:/usr/lib/hadoop-yarn/lib/commons-beanutils-1.9.4.jar:/usr/lib/hadoop-yarn/lib/json-smart-1.3.1.jar:/usr/lib/hadoop-yarn/lib/commons-logging-1.1.3.jar:/usr/lib/hadoop-yarn/lib/jetty-sslengine-6.1.26-emr.jar:/usr/lib/hadoop-yarn/lib/commons-compress-1.19.jar:/usr/lib/hadoop-yarn/lib/commons-io-2.4.jar:/usr/lib/hadoop-yarn/lib/jackson-core-asl-1.9.13.jar:/usr/lib/hadoop-yarn/lib/httpclient-4.5.9.jar:/usr/lib/hadoop-yarn/lib/netty-3.10.6.Final.jar:/usr/lib/hadoop-yarn/lib/commons-lang-2.6.jar:/usr/lib/hadoop-yarn/lib/jetty-6.1.26-emr.jar:/usr/lib/hadoop-yarn/lib/jackson-mapper-asl-1.9.13.jar:/usr/lib/hadoop-yarn/lib/geronimo-jcache\_1.0\_spec-1.0-alpha-1.jar:/usr/lib/hadoop-yarn/lib/mssql-jdbc-6.2.1.jre7.jar:/usr/lib/hadoop-yarn/lib/commons-digester-1.8.jar:/usr/lib/hadoop-yarn/lib/protobuf-java-2.5.0.jar:/usr/lib/hadoop-yarn/lib/jsch-0.1.54.jar:/usr/lib/hadoop-yarn/lib/jersey-server-1.9.jar:/usr/lib/hadoop-yarn/lib/commons-cli-1.2.jar:/usr/lib/hadoop-yarn/lib/apacheds-i18n-2.0.0-M15.jar:/usr/lib/hadoop-yarn/lib/woodstox-core-5.0.3.jar:/usr/lib/hadoop-yarn/lib/htrace-core4-4.1.0-incubating.jar:/usr/lib/hadoop-yarn/lib/commons-net-3.1.jar:/usr/lib/hadoop-yarn/lib/json-io-2.5.1.jar:/usr/lib/hadoop-yarn/lib/jersey-guice-1.9.jar:/usr/lib/hadoop-yarn/lib/snappy-java-1.1.7.3.jar:/usr/lib/hadoop-yarn/lib/leveldbjni-all-1.8.jar:/usr/lib/hadoop-yarn/lib/api-asn1-api-1.0.0-M20.jar:/usr/lib/hadoop-yarn/lib/gson-2.2.4.jar:/usr/lib/hadoop-yarn/lib/nimbus-jose-jwt-4.41.1.jar:/usr/lib/hadoop-yarn/lib/jsr305-3.0.0.jar:/usr/lib/hadoop-yarn/lib/metrics-core-3.0.1.jar:/usr/lib/hadoop-yarn/lib/guice-3.0.jar:/usr/lib/hadoop-yarn/lib/java-uti

1-1.9.0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-tests-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-nodemanager.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-client-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-timeline-pluginstorage-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-tests.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-applications-unmanaged-am-launcher.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-router-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-resourcemanager.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-applicationhistoryservice-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-resourcemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-applicationhistoryservice.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-applications-distributedshell-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-applications-unmanaged-am-launcher-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-registry-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-common.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-sharedcachemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-client.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-web-proxy.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-sharedcachemanager.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-web-proxy-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-router.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-registry.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-api-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-timeline-pluginstorage.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-common.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-applications-distributedshell.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-api.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-nodemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/lib/javax.inject-1.jar:/usr/lib/hadoop-mapreduce/lib/asm-3.2.jar:/usr/lib/hadoop-mapreduce/lib/paranamer-2.3.jar:/usr/lib/hadoop-mapreduce/lib/avro-1.7.7.jar:/usr/lib/hadoop-mapreduce/lib/aopalliance-1.0.jar:/usr/lib/hadoop-mapreduce/lib/log4j-1.2.17.jar:/usr/lib/hadoop-mapreduce/lib/guice-servlet-3.0.jar:/usr/lib/hadoop-mapreduce/lib/jersey-core-1.9.jar:/usr/lib/hadoop-mapreduce/lib/hamcrest-core-1.3.jar:/usr/lib/hadoop-mapreduce/lib/commons-compress-1.19.jar:/usr/lib/hadoop-mapreduce/lib/commons-io-2.4.jar:/usr/lib/hadoop-mapreduce/lib/jackson-core-asl-1.9.13.jar:/usr/lib/hadoop-mapreduce/lib/netty-3.10.6.Final.jar:/usr/lib/hadoop-mapreduce/lib/jackson-mapper-asl-1.9.13.jar:/usr/lib/hadoop-mapreduce/lib/protobuf-java-2.5.0.jar:/usr/lib/hadoop-mapreduce/lib/jersey-server-1.9.jar:/usr/lib/hadoop-mapreduce/lib/jersey-guice-1.9.jar:/usr/lib/hadoop-mapreduce/lib/snappy-java-1.1.7.3.jar:/usr/lib/hadoop-mapreduce/lib/leveldbjni-all-1.8.jar:/usr/lib/hadoop-mapreduce/lib/junit-4.11.jar:/usr/lib/hadoop-mapreduce/lib/guice-3.0.jar:/usr/lib/hadoop-mapreduce/./audience-annotations-0.5.0.jar:/usr/lib/hadoop-mapreduce/./javax.inject-1.jar:/usr/lib/hadoop-mapreduce/./hadoop-openstack-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-azure-datalake.jar:/usr/lib/hadoop-mapreduce/./hadoop-distcp-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jackson-jaxrs-1.9.13.jar:/usr/lib/hadoop-mapreduce/./commons-collections-3.2.2.jar:/usr/lib/hadoop-mapreduce/./hadoop-aliyun.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-hs-plugins-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-ant-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./azure-storage-5.4.0.jar:/usr/lib/hadoop-mapreduce/./spotbugs-annotations-3.1.9.jar:/usr/lib/hadoop-mapreduce/./stax2-api-3.1.4.jar:/usr/lib/hadoop-mapreduce/./hadoop-azure.jar:/usr/lib/hadoop-mapreduce/./curator-client-2.7.1.jar:/usr/lib/hadoop-mapreduce/./commons-lang3-3.4.jar:/usr/lib/hadoop-mapreduce/./jcip-annotations-1.0-1.jar:/usr/lib/hadoop-mapreduce/./jetty-util-6.1.26-emr.jar:/usr/lib/hadoop-mapreduce/./jaxb-api-2.2.2.jar:/usr/lib/hadoop-mapreduce/./jersey-client-1.9.jar:/usr/lib/hadoop-mapreduce/./azure-keyvault-core-0.8.0.jar:/usr/lib/hadoop-mapreduce/./asm-3.2.jar:/usr/lib/hadoop-mapreduce/./jettison-1.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-auth.jar:/usr/lib/hadoop-mapreduce/./ojslgo-43.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-common.jar:/usr/lib/hadoop-mapreduce/./paranamer-2.3.jar:/usr/lib/hadoop-mapreduce/./java-xmlbuilder-0.4.jar:/usr/lib/hadoop-mapreduce/./hadoop-aliyun-2.10.0-amzn-

0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-resourcemanager.jar:/usr/lib/hadoop-mapreduce/./jxb-impl-2.2.3-1.jar:/usr/lib/hadoop-mapreduce/./avro-1.7.7.jar:/usr/lib/hadoop-mapreduce/./aliyun-java-sdk-sts-3.0.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-aws.jar:/usr/lib/hadoop-mapreduce/./hadoop-azure-datalake-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-rumen.jar:/usr/lib/hadoop-mapreduce/./ehcache-3.3.1.jar:/usr/lib/hadoop-mapreduce/./aopalliance-1.0.jar:/usr/lib/hadoop-mapreduce/./aliyun-java-sdk-ram-3.0.0.jar:/usr/lib/hadoop-mapreduce/./curator-framework-2.7.1.jar:/usr/lib/hadoop-mapreduce/./jersey-json-1.9.jar:/usr/lib/hadoop-mapreduce/./hadoop-extras.jar:/usr/lib/hadoop-mapreduce/./curator-recipes-2.7.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-app.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-applicationhistoryservice-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./log4j-1.2.17.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-jobclient.jar:/usr/lib/hadoop-mapreduce/./apacheds-kerberos-codec-2.0.0-M15.jar:/usr/lib/hadoop-mapreduce/./jets3t-0.9.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-archives.jar:/usr/lib/hadoop-mapreduce/./jackson-xc-1.9.13.jar:/usr/lib/hadoop-mapreduce/./hadoop-gridmix-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./httpcore-4.4.11.jar:/usr/lib/hadoop-mapreduce/./guava-11.0.2.jar:/usr/lib/hadoop-mapreduce/./hadoop-auth-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jsp-api-2.1.jar:/usr/lib/hadoop-mapreduce/./aliyun-sdk-oss-3.4.1.jar:/usr/lib/hadoop-mapreduce/./HikariCP-java7-2.4.12.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-examples.jar:/usr/lib/hadoop-mapreduce/./xmlenc-0.52.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-resourcemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-applicationhistoryservice.jar:/usr/lib/hadoop-mapreduce/./servlet-api-2.5.jar:/usr/lib/hadoop-mapreduce/./commons-codec-1.4.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-jobclient-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./api-util-1.0.0-M20.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-registry-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-extras-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./commons-math3-3.1.1.jar:/usr/lib/hadoop-mapreduce/./guice-servlet-3.0.jar:/usr/lib/hadoop-mapreduce/./activation-1.1.jar:/usr/lib/hadoop-mapreduce/./aws-java-sdk-bundle-1.11.852.jar:/usr/lib/hadoop-mapreduce/./zookeeper-3.4.14.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-core.jar:/usr/lib/hadoop-mapreduce/./stax-api-1.0-2.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-common.jar:/usr/lib/hadoop-mapreduce/./commons-configuration-1.6.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-app-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jersey-core-1.9.jar:/usr/lib/hadoop-mapreduce/./fst-2.50.jar:/usr/lib/hadoop-mapreduce/./commons-beanutils-1.9.4.jar:/usr/lib/hadoop-mapreduce/./json-smart-1.3.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-gridmix.jar:/usr/lib/hadoop-mapreduce/./jackson-core-2.6.7.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-hs-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./commons-logging-1.1.3.jar:/usr/lib/hadoop-mapreduce/./hadoop-streaming-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-distcp.jar:/usr/lib/hadoop-mapreduce/./jetty-sslengine-6.1.26-emr.jar:/usr/lib/hadoop-mapreduce/./commons-compress-1.19.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-web-proxy.jar:/usr/lib/hadoop-mapreduce/./hadoop-aws-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./commons-io-2.4.jar:/usr/lib/hadoop-mapreduce/./hadoop-resourceestimator.jar:/usr/lib/hadoop-mapreduce/./jackson-core-asl-1.9.13.jar:/usr/lib/hadoop-mapreduce/./httpclient-4.5.9.jar:/usr/lib/hadoop-mapreduce/./netty-3.10.6.Final.jar:/usr/lib/hadoop-mapreduce/./commons-lang-2.6.jar:/usr/lib/hadoop-mapreduce/./hadoop-openstack.jar:/usr/lib/hadoop-mapreduce/./jetty-6.1.26-emr.jar:/usr/lib/hadoop-mapreduce/./jackson-annotations-2.6.7.jar:/usr/lib/hadoop-mapreduce/./jackson-mapper-asl-1.9.13.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-web-proxy-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./aliyun-java-sdk-ecs-4.2.0.jar:/usr/lib/hadoop-mapreduce/./geronimo-jcache\_1.0\_spec-1.0-alpha-1.jar:/usr/lib/hadoop-mapreduce/./hadoop-archive-logs.jar:/usr/lib/hadoop-mapreduce/./mssql-jdbc-6.2.1.jre7.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-registry.jar:/usr/lib/hadoop-mapreduce/./commons-digester-1.8.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-hs-plugins.jar:/usr

```

r/lib/hadoop-mapreduce/./protobuf-java-2.5.0.jar:/usr/lib/hadoop-mapreduce/./jsch-
0.1.54.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-hs.jar:/usr/lib/hadoo
p-mapreduce/./hadoop-mapreduce-client-core-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapre
duce/./jdom-1.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-common-2.1
0.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jersey-server-1.9.jar:/usr/lib/hadoop-ma
preduce/./commons-cli-1.2.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-
shuffle.jar:/usr/lib/hadoop-mapreduce/./hadoop-streaming.jar:/usr/lib/hadoop-mapred
uce/./hadoop-yarn-api-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./commons-httpcl
ient-3.1.jar:/usr/lib/hadoop-mapreduce/./aliyun-java-sdk-core-3.4.0.jar:/usr/lib/ha
doo-mapreduce/./hadoop-yarn-server-common.jar:/usr/lib/hadoop-mapreduce/./apached
s-i18n-2.0.0-M15.jar:/usr/lib/hadoop-mapreduce/./woodstox-core-5.0.3.jar:/usr/lib/h
adoop-mapreduce/./htrace-core4-4.1.0-incubating.jar:/usr/lib/hadoop-mapreduce/./ha
doo-datajoin-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-exampl
es-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./commons-net-3.1.jar:/usr/lib/hadoo
p-mapreduce/./hadoop-sls.jar:/usr/lib/hadoop-mapreduce/./json-io-2.5.1.jar:/usr/li
b/hadoop-mapreduce/./jersey-guice-1.9.jar:/usr/lib/hadoop-mapreduce/./hadoop-rumen
-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./snappy-java-1.1.7.3.jar:/usr/lib/had
oop-mapreduce/./hadoop-mapreduce-client-jobclient-2.10.0-amzn-0-tests.jar:/usr/lib/
hadoop-mapreduce/./leveldbjni-all-1.8.jar:/usr/lib/hadoop-mapreduce/./hadoop-azure
-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./api-asn1-api-1.0.0-M20.jar:/usr/lib/
hadoop-mapreduce/./jackson-databind-2.6.7.jar:/usr/lib/hadoop-mapreduce/./hadoop-m
apreduce-client-shuffle-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./gson-2.2.4.ja
r:/usr/lib/hadoop-mapreduce/./nimbus-jose-jwt-4.41.1.jar:/usr/lib/hadoop-mapreduc
e/./hadoop-archives-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jsr305-3.0.0.ja
r:/usr/lib/hadoop-mapreduce/./hadoop-ant.jar:/usr/lib/hadoop-mapreduce/./hadoop-sl
s-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-api.jar:/usr/lib/hadoo
p-mapreduce/./hadoop-datajoin.jar:/usr/lib/hadoop-mapreduce/./metrics-core-3.0.1.ja
r:/usr/lib/hadoop-mapreduce/./azure-data-lake-store-sdk-2.2.3.jar:/usr/lib/hadoop-m
apreduce/./hadoop-archive-logs-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./guice
-3.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-resourceestimator-2.10.0-amzn-0.jar:/us
r/lib/hadoop-mapreduce/./java-util-1.9.0.jar

```

```

STARTUP_MSG: build = git@aws157git.com:/pkg/Aws157BigTop -r d1e860a34cc1aea3d600c5
7c5c0270ea41579e8c; compiled by 'ec2-user' on 2020-09-19T02:05Z

```

```

STARTUP_MSG: java = 1.8.0_312

```

```

*****/

```

```

24/04/28 21:20:47 INFO datanode.DataNode: registered UNIX signal handlers for [TERM,
HUP, INT]

```

```

24/04/28 21:20:47 INFO namenode.NameNode: STARTUP_MSG:

```

```

/*****

```

```

STARTUP_MSG: Starting NameNode

```

```

STARTUP_MSG: host = algo-1/10.0.222.169

```

```

STARTUP_MSG: args = []

```

```

STARTUP_MSG: version = 2.10.0-amzn-0

```

```

STARTUP_MSG: classpath = /usr/lib/hadoop/etc/hadoop:/usr/lib/hadoop/lib/audience-a
nnotations-0.5.0.jar:/usr/lib/hadoop/lib/jackson-jaxrs-1.9.13.jar:/usr/lib/hadoop/li
b/commons-collections-3.2.2.jar:/usr/lib/hadoop/lib/mockito-all-1.8.5.jar:/usr/lib/h
adoop/lib/spotbugs-annotations-3.1.9.jar:/usr/lib/hadoop/lib/stax2-api-3.1.4.jar:/us
r/lib/hadoop/lib/curator-client-2.7.1.jar:/usr/lib/hadoop/lib/commons-lang3-3.4.ja
r:/usr/lib/hadoop/lib/jcip-annotations-1.0-1.jar:/usr/lib/hadoop/lib/jetty-util-6.1.
26-emr.jar:/usr/lib/hadoop/lib/jaxb-api-2.2.2.jar:/usr/lib/hadoop/lib/asm-3.2.jar:/u
sr/lib/hadoop/lib/jettison-1.1.jar:/usr/lib/hadoop/lib/paranamer-2.3.jar:/usr/lib/ha
doo/lib/java-xmlbuilder-0.4.jar:/usr/lib/hadoop/lib/jaxb-impl-2.2.3-1.jar:/usr/lib/
hadoop/lib/avro-1.7.7.jar:/usr/lib/hadoop/lib/curator-framework-2.7.1.jar:/usr/lib/h
adoop/lib/jersey-json-1.9.jar:/usr/lib/hadoop/lib/curator-recipes-2.7.1.jar:/usr/li
b/hadoop/lib/log4j-1.2.17.jar:/usr/lib/hadoop/lib/apacheds-kerberos-codec-2.0.0-M15.
jar:/usr/lib/hadoop/lib/jets3t-0.9.0.jar:/usr/lib/hadoop/lib/jackson-xc-1.9.13.jar:/

```

usr/lib/hadoop/lib/httpcore-4.4.11.jar:/usr/lib/hadoop/lib/guava-11.0.2.jar:/usr/lib/hadoop/lib/jsp-api-2.1.jar:/usr/lib/hadoop/lib/xmlenc-0.52.jar:/usr/lib/hadoop/lib/servlet-api-2.5.jar:/usr/lib/hadoop/lib/commons-codec-1.4.jar:/usr/lib/hadoop/lib/slf4j-api-1.7.25.jar:/usr/lib/hadoop/lib/api-util-1.0.0-M20.jar:/usr/lib/hadoop/lib/commons-math3-3.1.1.jar:/usr/lib/hadoop/lib/activation-1.1.jar:/usr/lib/hadoop/lib/zookeeper-3.4.14.jar:/usr/lib/hadoop/lib/stax-api-1.0-2.jar:/usr/lib/hadoop/lib/commons-configuration-1.6.jar:/usr/lib/hadoop/lib/jersey-core-1.9.jar:/usr/lib/hadoop/lib/commons-beanutils-1.9.4.jar:/usr/lib/hadoop/lib/hamcrest-core-1.3.jar:/usr/lib/hadoop/lib/json-smart-1.3.1.jar:/usr/lib/hadoop/lib/commons-logging-1.1.3.jar:/usr/lib/hadoop/lib/jetty-sslengine-6.1.26-emr.jar:/usr/lib/hadoop/lib/commons-compress-1.19.jar:/usr/lib/hadoop/lib/commons-io-2.4.jar:/usr/lib/hadoop/lib/jackson-core-asl-1.9.13.jar:/usr/lib/hadoop/lib/httpclient-4.5.9.jar:/usr/lib/hadoop/lib/netty-3.10.6.Final.jar:/usr/lib/hadoop/lib/commons-lang-2.6.jar:/usr/lib/hadoop/lib/jetty-6.1.26-emr.jar:/usr/lib/hadoop/lib/jackson-mapper-asl-1.9.13.jar:/usr/lib/hadoop/lib/commons-digester-1.8.jar:/usr/lib/hadoop/lib/protobuf-java-2.5.0.jar:/usr/lib/hadoop/lib/jsc-h-0.1.54.jar:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.25.jar:/usr/lib/hadoop/lib/jersey-server-1.9.jar:/usr/lib/hadoop/lib/commons-cli-1.2.jar:/usr/lib/hadoop/lib/apacheds-i18n-2.0.0-M15.jar:/usr/lib/hadoop/lib/woodstox-core-5.0.3.jar:/usr/lib/hadoop/lib/htrace-core4-4.1.0-incubating.jar:/usr/lib/hadoop/lib/commons-net-3.1.jar:/usr/lib/hadoop/lib/snappy-java-1.1.7.3.jar:/usr/lib/hadoop/lib/api-asn1-api-1.0.0-M20.jar:/usr/lib/hadoop/lib/gson-2.2.4.jar:/usr/lib/hadoop/lib/nimbus-jose-jwt-4.41.1.jar:/usr/lib/hadoop/lib/jsr305-3.0.0.jar:/usr/lib/hadoop/lib/junit-4.11.jar:/usr/lib/hadoop/p//hadoop-openstack-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-azure-datalake.jar:/usr/lib/hadoop/p//hadoop-distcp-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-aliyun.jar:/usr/lib/hadoop/p//hadoop-ant-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-azure.jar:/usr/lib/hadoop/p//hadoop-nfs-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-auth.jar:/usr/lib/hadoop/p//hadoop-yarn-common-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-aliyun-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-yarn-server-resourcemanager.jar:/usr/lib/hadoop/p//hadoop-aws.jar:/usr/lib/hadoop/p//hadoop-azure-datalake-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-rumen.jar:/usr/lib/hadoop/p//hadoop-extras.jar:/usr/lib/hadoop/p//hadoop-yarn-server-applicationhistoryservice-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-annotations-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-archives.jar:/usr/lib/hadoop/p//hadoop-gridmix-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-auth-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-yarn-server-resourcemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-yarn-server-applicationhistoryservice.jar:/usr/lib/hadoop/p//hadoop-annotations.jar:/usr/lib/hadoop/p//hadoop-yarn-registry-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-extras-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-yarn-common.jar:/usr/lib/hadoop/p//hadoop-gridmix.jar:/usr/lib/hadoop/p//hadoop-streaming-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-distcp.jar:/usr/lib/hadoop/p//hadoop-yarn-server-common-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-yarn-server-web-proxy.jar:/usr/lib/hadoop/p//hadoop-aws-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-common-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop/p//hadoop-resourceestimator.jar:/usr/lib/hadoop/p//hadoop-openstack.jar:/usr/lib/hadoop/p//hadoop-yarn-server-web-proxy-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-archive-logs.jar:/usr/lib/hadoop/p//hadoop-nfs.jar:/usr/lib/hadoop/p//hadoop-yarn-registry.jar:/usr/lib/hadoop/p//hadoop-common-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-streaming.jar:/usr/lib/hadoop/p//hadoop-yarn-api-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-yarn-server-common.jar:/usr/lib/hadoop/p//hadoop-datajoin-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-common.jar:/usr/lib/hadoop/p//hadoop-sls.jar:/usr/lib/hadoop/p//hadoop-rumen-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-azure-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-archives-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-ant.jar:/usr/lib/hadoop/p//hadoop-sls-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-yarn-api.jar:/usr/lib/hadoop/p//hadoop-datajoin.jar:/usr/lib/hadoop/p//hadoop-archive-logs-2.10.0-amzn-0.jar:/usr/lib/hadoop/p//hadoop-resourceestimator-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/p//usr/lib/hadoop-hdfs/lib/jetty-util-6.1.26-emr.jar:/usr/lib/hadoop-hdfs/lib/asm-3.2.jar:/usr/lib/hadoop-hdfs/lib/okio-1.6.0.jar:/usr/lib/hadoop-hdfs/lib/commons

-daemon-1.0.13.jar:/usr/lib/hadoop-hdfs/lib/log4j-1.2.17.jar:/usr/lib/hadoop-hdfs/lib/guava-11.0.2.jar:/usr/lib/hadoop-hdfs/lib/xmlenc-0.52.jar:/usr/lib/hadoop-hdfs/lib/xercesImpl-2.12.0.jar:/usr/lib/hadoop-hdfs/lib/servlet-api-2.5.jar:/usr/lib/hadoop-hdfs/lib/commons-codec-1.4.jar:/usr/lib/hadoop-hdfs/lib/xml-apis-1.4.01.jar:/usr/lib/hadoop-hdfs/lib/jersey-core-1.9.jar:/usr/lib/hadoop-hdfs/lib/jackson-core-2.6.7.jar:/usr/lib/hadoop-hdfs/lib/commons-logging-1.1.3.jar:/usr/lib/hadoop-hdfs/lib/commons-io-2.4.jar:/usr/lib/hadoop-hdfs/lib/jackson-core-asl-1.9.13.jar:/usr/lib/hadoop-hdfs/lib/netty-3.10.6.Final.jar:/usr/lib/hadoop-hdfs/lib/commons-lang-2.6.jar:/usr/lib/hadoop-hdfs/lib/jetty-6.1.26-emr.jar:/usr/lib/hadoop-hdfs/lib/jackson-annotations-2.6.7.jar:/usr/lib/hadoop-hdfs/lib/jackson-mapper-asl-1.9.13.jar:/usr/lib/hadoop-hdfs/lib/protobuf-java-2.5.0.jar:/usr/lib/hadoop-hdfs/lib/jersey-server-1.9.jar:/usr/lib/hadoop-hdfs/lib/commons-cli-1.2.jar:/usr/lib/hadoop-hdfs/lib/htrace-core4-4.1.0-incubating.jar:/usr/lib/hadoop-hdfs/lib/leveldbjni-all-1.8.jar:/usr/lib/hadoop-hdfs/lib/jackson-databind-2.6.7.jar:/usr/lib/hadoop-hdfs/lib/okhttp-2.7.5.jar:/usr/lib/hadoop-hdfs/lib/jsr305-3.0.0.jar:/usr/lib/hadoop-hdfs/lib/netty-all-4.0.23.Final.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-client-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-native-client.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-rbf-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-rbf-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-client-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-native-client-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-nfs.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-client.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-nfs-2.10.0-amzn-0.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-rbf.jar:/usr/lib/hadoop-hdfs/./hadoop-hdfs-native-client-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop-yarn/lib/audience-annotations-0.5.0.jar:/usr/lib/hadoop-yarn/lib/javax.inject-1.jar:/usr/lib/hadoop-yarn/lib/jackson-jaxrs-1.9.13.jar:/usr/lib/hadoop-yarn/lib/commons-collections-3.2.2.jar:/usr/lib/hadoop-yarn/lib/spotbugs-annotations-3.1.9.jar:/usr/lib/hadoop-yarn/lib/stax2-api-3.1.4.jar:/usr/lib/hadoop-yarn/lib/curator-client-2.7.1.jar:/usr/lib/hadoop-yarn/lib/commons-lang3-3.4.jar:/usr/lib/hadoop-yarn/lib/jcip-annotations-1.0-1.jar:/usr/lib/hadoop-yarn/lib/jetty-util-6.1.26-emr.jar:/usr/lib/hadoop-yarn/lib/jaxb-api-2.2.2.jar:/usr/lib/hadoop-yarn/lib/jersey-client-1.9.jar:/usr/lib/hadoop-yarn/lib/asm-3.2.jar:/usr/lib/hadoop-yarn/lib/jettison-1.1.jar:/usr/lib/hadoop-yarn/lib/paranamer-2.3.jar:/usr/lib/hadoop-yarn/lib/java-xmlbuilder-0.4.jar:/usr/lib/hadoop-yarn/lib/jaxb-impl-2.2.3-1.jar:/usr/lib/hadoop-yarn/lib/avro-1.7.7.jar:/usr/lib/hadoop-yarn/lib/ehcache-3.3.1.jar:/usr/lib/hadoop-yarn/lib/aopalliance-1.0.jar:/usr/lib/hadoop-yarn/lib/curator-framework-2.7.1.jar:/usr/lib/hadoop-yarn/lib/jersey-json-1.9.jar:/usr/lib/hadoop-yarn/lib/curator-recipes-2.7.1.jar:/usr/lib/hadoop-yarn/lib/log4j-1.2.17.jar:/usr/lib/hadoop-yarn/lib/apacheds-kerberos-codec-2.0.0-M15.jar:/usr/lib/hadoop-yarn/lib/jets3t-0.9.0.jar:/usr/lib/hadoop-yarn/lib/jackson-xc-1.9.13.jar:/usr/lib/hadoop-yarn/lib/httpcore-4.4.11.jar:/usr/lib/hadoop-yarn/lib/guava-11.0.2.jar:/usr/lib/hadoop-yarn/lib/jsp-api-2.1.jar:/usr/lib/hadoop-yarn/lib/HikariCP-java7-2.4.12.jar:/usr/lib/hadoop-yarn/lib/xmlenc-0.52.jar:/usr/lib/hadoop-yarn/lib/servlet-api-2.5.jar:/usr/lib/hadoop-yarn/lib/commons-codec-1.4.jar:/usr/lib/hadoop-yarn/lib/api-util-1.0.0-M20.jar:/usr/lib/hadoop-yarn/lib/commons-math3-3.1.1.jar:/usr/lib/hadoop-yarn/lib/guice-servlet-3.0.jar:/usr/lib/hadoop-yarn/lib/activation-1.1.jar:/usr/lib/hadoop-yarn/lib/zookeeper-3.4.14.jar:/usr/lib/hadoop-yarn/lib/stax-api-1.0-2.jar:/usr/lib/hadoop-yarn/lib/commons-configuration-1.6.jar:/usr/lib/hadoop-yarn/lib/jersey-core-1.9.jar:/usr/lib/hadoop-yarn/lib/fst-2.50.jar:/usr/lib/hadoop-yarn/lib/commons-beanutils-1.9.4.jar:/usr/lib/hadoop-yarn/lib/json-smart-1.3.1.jar:/usr/lib/hadoop-yarn/lib/commons-logging-1.1.3.jar:/usr/lib/hadoop-yarn/lib/jetty-sslengine-6.1.26-emr.jar:/usr/lib/hadoop-yarn/lib/commons-compress-1.19.jar:/usr/lib/hadoop-yarn/lib/commons-io-2.4.jar:/usr/lib/hadoop-yarn/lib/jackson-core-asl-1.9.13.jar:/usr/lib/hadoop-yarn/lib/httpclient-4.5.9.jar:/usr/lib/hadoop-yarn/lib/netty-3.10.6.Final.jar:/usr/lib/hadoop-yarn/lib/commons-lang-2.6.jar:/usr/lib/hadoop-yarn/lib/jetty-6.1.26-emr.jar:/usr/lib/hadoop-yarn/lib/jackson-mapper-asl-1.9.13.jar:/usr/lib/hadoop-yarn/lib/geronimo-jcache



\_1.0\_spec-1.0-alpha-1.jar:/usr/lib/hadoop-yarn/lib/mssql-jdbc-6.2.1.jre7.jar:/usr/lib/hadoop-yarn/lib/commons-digester-1.8.jar:/usr/lib/hadoop-yarn/lib/protobuf-java-2.5.0.jar:/usr/lib/hadoop-yarn/lib/jsch-0.1.54.jar:/usr/lib/hadoop-yarn/lib/jersey-server-1.9.jar:/usr/lib/hadoop-yarn/lib/commons-cli-1.2.jar:/usr/lib/hadoop-yarn/lib/acheds-i18n-2.0.0-M15.jar:/usr/lib/hadoop-yarn/lib/woodstox-core-5.0.3.jar:/usr/lib/hadoop-yarn/lib/htrace-core4-4.1.0-incubating.jar:/usr/lib/hadoop-yarn/lib/commons-net-3.1.jar:/usr/lib/hadoop-yarn/lib/json-io-2.5.1.jar:/usr/lib/hadoop-yarn/lib/jersey-guice-1.9.jar:/usr/lib/hadoop-yarn/lib/snappy-java-1.1.7.3.jar:/usr/lib/hadoop-yarn/lib/leveldbjni-all-1.8.jar:/usr/lib/hadoop-yarn/lib/api-asn1-api-1.0.0-M20.jar:/usr/lib/hadoop-yarn/lib/gson-2.2.4.jar:/usr/lib/hadoop-yarn/lib/nimbus-jose-jwt-4.41.1.jar:/usr/lib/hadoop-yarn/lib/jsr305-3.0.0.jar:/usr/lib/hadoop-yarn/lib/metrics-core-3.0.1.jar:/usr/lib/hadoop-yarn/lib/guice-3.0.jar:/usr/lib/hadoop-yarn/lib/java-util-1.9.0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-tests-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-nodemanager.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-client-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-timeline-pluginstorage-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-tests.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-applications-unmanaged-am-launcher.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-router-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-resourcemanager.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-applicationhistoryservice-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-resourcemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-applicationhistoryservice.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-applications-distributedshell-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-applications-unmanaged-am-launcher-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-registry-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-common.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-sharedcachemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-client.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-web-proxy.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-sharedcachemanager.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-web-proxy-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-router.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-registry.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-api-2.10.0-amzn-0.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-timeline-pluginstorage.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-common.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-applications-distributedshell.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-api.jar:/usr/lib/hadoop-yarn/./hadoop-yarn-server-nodemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/lib/javax.inject-1.jar:/usr/lib/hadoop-mapreduce/lib/asm-3.2.jar:/usr/lib/hadoop-mapreduce/lib/paranamer-2.3.jar:/usr/lib/hadoop-mapreduce/lib/avro-1.7.7.jar:/usr/lib/hadoop-mapreduce/lib/aopalliance-1.0.jar:/usr/lib/hadoop-mapreduce/lib/log4j-1.2.17.jar:/usr/lib/hadoop-mapreduce/lib/guice-servlet-3.0.jar:/usr/lib/hadoop-mapreduce/lib/jersey-core-1.9.jar:/usr/lib/hadoop-mapreduce/lib/hamcrest-core-1.3.jar:/usr/lib/hadoop-mapreduce/lib/commons-compress-1.19.jar:/usr/lib/hadoop-mapreduce/lib/commons-io-2.4.jar:/usr/lib/hadoop-mapreduce/lib/jackson-core-asl-1.9.13.jar:/usr/lib/hadoop-mapreduce/lib/netty-3.10.6.Final.jar:/usr/lib/hadoop-mapreduce/lib/jackson-mapper-asl-1.9.13.jar:/usr/lib/hadoop-mapreduce/lib/protobuf-java-2.5.0.jar:/usr/lib/hadoop-mapreduce/lib/jersey-server-1.9.jar:/usr/lib/hadoop-mapreduce/lib/jersey-guice-1.9.jar:/usr/lib/hadoop-mapreduce/lib/snappy-java-1.1.7.3.jar:/usr/lib/hadoop-mapreduce/lib/leveldbjni-all-1.8.jar:/usr/lib/hadoop-mapreduce/lib/junit-4.11.jar:/usr/lib/hadoop-mapreduce/lib/guice-3.0.jar:/usr/lib/hadoop-mapreduce/./audience-annotations-0.5.0.jar:/usr/lib/hadoop-mapreduce/./javax.inject-1.jar:/usr/lib/hadoop-mapreduce/./hadoop-openstack-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-azure-datalake.jar:/usr/lib/hadoop-mapreduce/./hadoop-distcp-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jackson-jaxrs-1.9.13.jar:/usr/lib/hadoop-mapreduce/./commons-collections-3.2.2.jar:/usr/lib/hadoop-mapreduce/./hadoop-aliyun.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-hs-plugins-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-ant-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./azure-storage-5.4.

0.jar:/usr/lib/hadoop-mapreduce/./spotbugs-annotations-3.1.9.jar:/usr/lib/hadoop-mapreduce/./stax2-api-3.1.4.jar:/usr/lib/hadoop-mapreduce/./hadoop-azure.jar:/usr/lib/hadoop-mapreduce/./curator-client-2.7.1.jar:/usr/lib/hadoop-mapreduce/./commons-lang3-3.4.jar:/usr/lib/hadoop-mapreduce/./jcip-annotations-1.0-1.jar:/usr/lib/hadoop-mapreduce/./jetty-util-6.1.26-emr.jar:/usr/lib/hadoop-mapreduce/./jaxb-api-2.2.2.jar:/usr/lib/hadoop-mapreduce/./jersey-client-1.9.jar:/usr/lib/hadoop-mapreduce/./azure-keyvault-core-0.8.0.jar:/usr/lib/hadoop-mapreduce/./asm-3.2.jar:/usr/lib/hadoop-mapreduce/./jettison-1.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-auth.jar:/usr/lib/hadoop-mapreduce/./ojslgo-43.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-common.jar:/usr/lib/hadoop-mapreduce/./paranamer-2.3.jar:/usr/lib/hadoop-mapreduce/./java-xmlbuilder-0.4.jar:/usr/lib/hadoop-mapreduce/./hadoop-aliyun-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-resourcemanager.jar:/usr/lib/hadoop-mapreduce/./jaxb-impl-2.2.3-1.jar:/usr/lib/hadoop-mapreduce/./avro-1.7.7.jar:/usr/lib/hadoop-mapreduce/./aliyun-java-sdk-sts-3.0.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-aws.jar:/usr/lib/hadoop-mapreduce/./hadoop-azure-datalake-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-rumen.jar:/usr/lib/hadoop-mapreduce/./ehcache-3.3.1.jar:/usr/lib/hadoop-mapreduce/./aopalliance-1.0.jar:/usr/lib/hadoop-mapreduce/./aliyun-java-sdk-ram-3.0.0.jar:/usr/lib/hadoop-mapreduce/./curator-framework-2.7.1.jar:/usr/lib/hadoop-mapreduce/./jersey-json-1.9.jar:/usr/lib/hadoop-mapreduce/./hadoop-extras.jar:/usr/lib/hadoop-mapreduce/./curator-recipes-2.7.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-app.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-applicationhistoryservice-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./log4j-1.2.17.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-jobclient.jar:/usr/lib/hadoop-mapreduce/./apacheds-kerberos-codec-2.0.0-M15.jar:/usr/lib/hadoop-mapreduce/./jets3t-0.9.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-archives.jar:/usr/lib/hadoop-mapreduce/./jackson-xc-1.9.13.jar:/usr/lib/hadoop-mapreduce/./hadoop-gridmix-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./httpcore-4.4.11.jar:/usr/lib/hadoop-mapreduce/./guava-11.0.2.jar:/usr/lib/hadoop-mapreduce/./hadoop-auth-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jsp-api-2.1.jar:/usr/lib/hadoop-mapreduce/./aliyun-sdk-oss-3.4.1.jar:/usr/lib/hadoop-mapreduce/./HikariCP-java7-2.4.12.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-examples.jar:/usr/lib/hadoop-mapreduce/./xmlenc-0.52.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-resourcemanager-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-applicationhistoryservice.jar:/usr/lib/hadoop-mapreduce/./servlet-api-2.5.jar:/usr/lib/hadoop-mapreduce/./commons-codec-1.4.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-jobclient-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./api-util-1.0.0-M20.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-registry-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-extras-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./commons-math3-3.1.1.jar:/usr/lib/hadoop-mapreduce/./guice-servlet-3.0.jar:/usr/lib/hadoop-mapreduce/./activation-1.1.jar:/usr/lib/hadoop-mapreduce/./aws-java-sdk-bundle-1.11.852.jar:/usr/lib/hadoop-mapreduce/./zookeeper-3.4.14.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-core.jar:/usr/lib/hadoop-mapreduce/./stax-api-1.0-2.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-common.jar:/usr/lib/hadoop-mapreduce/./commons-configuration-1.6.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-app-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jersey-core-1.9.jar:/usr/lib/hadoop-mapreduce/./fst-2.50.jar:/usr/lib/hadoop-mapreduce/./commons-beanutils-1.9.4.jar:/usr/lib/hadoop-mapreduce/./json-smart-1.3.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-gridmix.jar:/usr/lib/hadoop-mapreduce/./jackson-core-2.6.7.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-hs-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./commons-logging-1.1.3.jar:/usr/lib/hadoop-mapreduce/./hadoop-streaming-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-distcp.jar:/usr/lib/hadoop-mapreduce/./jetty-sslengine-6.1.26-emr.jar:/usr/lib/hadoop-mapreduce/./commons-compress-1.19.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-web-proxy.jar:/usr/lib/hadoop-mapreduce/./hadoop-aws-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce

```

e/./commons-io-2.4.jar:/usr/lib/hadoop-mapreduce/./hadoop-resourceestimator.jar:/usr/lib/hadoop-mapreduce/./jackson-core-asl-1.9.13.jar:/usr/lib/hadoop-mapreduce/./httpclient-4.5.9.jar:/usr/lib/hadoop-mapreduce/./netty-3.10.6.Final.jar:/usr/lib/hadoop-mapreduce/./commons-lang-2.6.jar:/usr/lib/hadoop-mapreduce/./hadoop-openstack.jar:/usr/lib/hadoop-mapreduce/./jetty-6.1.26-emr.jar:/usr/lib/hadoop-mapreduce/./jackson-annotations-2.6.7.jar:/usr/lib/hadoop-mapreduce/./jackson-mapper-asl-1.9.13.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-web-proxy-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./aliyun-java-sdk-ecs-4.2.0.jar:/usr/lib/hadoop-mapreduce/./geronimo-jcache-1.0_spec-1.0-alpha-1.jar:/usr/lib/hadoop-mapreduce/./hadoop-archive-logs.jar:/usr/lib/hadoop-mapreduce/./mssql-jdbc-6.2.1.jre7.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-registry.jar:/usr/lib/hadoop-mapreduce/./commons-digester-1.8.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-hs-plugins.jar:/usr/lib/hadoop-mapreduce/./protobuf-java-2.5.0.jar:/usr/lib/hadoop-mapreduce/./jsch-0.1.54.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-hs.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-core-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jdom-1.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-common-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jersey-server-1.9.jar:/usr/lib/hadoop-mapreduce/./commons-cli-1.2.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-shuffle.jar:/usr/lib/hadoop-mapreduce/./hadoop-streaming.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-api-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./commons-httpclient-3.1.jar:/usr/lib/hadoop-mapreduce/./aliyun-java-sdk-core-3.4.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-server-common.jar:/usr/lib/hadoop-mapreduce/./apache-s-i18n-2.0.0-M15.jar:/usr/lib/hadoop-mapreduce/./woodstox-core-5.0.3.jar:/usr/lib/hadoop-mapreduce/./htrace-core4-4.1.0-incubating.jar:/usr/lib/hadoop-mapreduce/./hadoop-datajoin-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-examples-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./commons-net-3.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-sls.jar:/usr/lib/hadoop-mapreduce/./json-io-2.5.1.jar:/usr/lib/hadoop-mapreduce/./jersey-guice-1.9.jar:/usr/lib/hadoop-mapreduce/./hadoop-rumen-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./snappy-java-1.1.7.3.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-jobclient-2.10.0-amzn-0-tests.jar:/usr/lib/hadoop-mapreduce/./leveldbjni-all-1.8.jar:/usr/lib/hadoop-mapreduce/./hadoop-azure-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./api-asn1-api-1.0.0-M20.jar:/usr/lib/hadoop-mapreduce/./jackson-databind-2.6.7.jar:/usr/lib/hadoop-mapreduce/./hadoop-mapreduce-client-shuffle-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./gson-2.2.4.jar:/usr/lib/hadoop-mapreduce/./nimbus-jose-jwt-4.41.1.jar:/usr/lib/hadoop-mapreduce/./hadoop-archives-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./jsr305-3.0.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-ant.jar:/usr/lib/hadoop-mapreduce/./hadoop-sls-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./hadoop-yarn-api.jar:/usr/lib/hadoop-mapreduce/./hadoop-datajoin.jar:/usr/lib/hadoop-mapreduce/./metrics-core-3.0.1.jar:/usr/lib/hadoop-mapreduce/./azure-data-lake-store-sdk-2.2.3.jar:/usr/lib/hadoop-mapreduce/./hadoop-archive-logs-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./guice-3.0.jar:/usr/lib/hadoop-mapreduce/./hadoop-resourceestimator-2.10.0-amzn-0.jar:/usr/lib/hadoop-mapreduce/./java-util-1.9.0.jar

```

```

STARTUP_MSG: build = git@aws157git.com:/pkg/Aws157BigTop -r d1e860a34cc1aea3d600c57c5c0270ea41579e8c; compiled by 'ec2-user' on 2020-09-19T02:05Z

```

```

STARTUP_MSG: java = 1.8.0_312

```

```

*****/

```

```

24/04/28 21:20:47 INFO namenode.NameNode: registered UNIX signal handlers for [TERM, HUP, INT]

```

```

24/04/28 21:20:47 INFO namenode.NameNode: createNameNode []

```

```

24/04/28 21:20:48 INFO impl.MetricsConfig: loaded properties from hadoop-metrics2.properties

```

```

24/04/28 21:20:48 INFO conf.Configuration: found resource core-site.xml at file:/etc/hadoop/conf.empty/core-site.xml

```

```

24/04/28 21:20:48 INFO security.Groups: clearing userToGroupsMap cache

```

```

24/04/28 21:20:48 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 1

```

```
0 second(s).
24/04/28 21:20:48 INFO impl.MetricsSystemImpl: NameNode metrics system started
24/04/28 21:20:48 INFO conf.Configuration: resource-types.xml not found
24/04/28 21:20:48 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
24/04/28 21:20:48 INFO namenode.NameNode: fs.defaultFS is hdfs://10.0.222.169/
24/04/28 21:20:48 INFO resource.ResourceUtils: Adding resource type - name = memory-
mb, units = Mi, type = COUNTABLE
24/04/28 21:20:48 INFO resource.ResourceUtils: Adding resource type - name = vcores,
units = , type = COUNTABLE
24/04/28 21:20:48 INFO conf.Configuration: found resource yarn-site.xml at file:/et
c/hadoop/conf.empty/yarn-site.xml
24/04/28 21:20:48 INFO event.AsyncDispatcher: Registering class org.apache.hadoop.ya
rn.server.resourcemanager.RMFatalEventType for class org.apache.hadoop.yarn.server.r
esourcemanager.ResourceManager$RMFatalEventDispatcher
24/04/28 21:20:48 INFO nodemanager.NodeManager: Node Manager health check script is
not available or doesn't have execute permission, so not starting the node health sc
ript runner.
24/04/28 21:20:48 INFO security.NMTOKENSecretManagerInRM: NMTOKENKeyRollingInterval:
86400000ms and NMTOKENKeyActivationDelay: 900000ms
24/04/28 21:20:48 INFO security.RMContainerTokenSecretManager: ContainerTokenKeyRoll
ingInterval: 86400000ms and ContainerTokenKeyActivationDelay: 900000ms
24/04/28 21:20:48 INFO event.AsyncDispatcher: Registering class org.apache.hadoop.ya
rn.server.nodemanager.containermanager.container.ContainerEventType for class org.ap
ache.hadoop.yarn.server.nodemanager.containermanager.ContainerManagerImpl$ContainerE
ventDispatcher
24/04/28 21:20:48 INFO event.AsyncDispatcher: Registering class org.apache.hadoop.ya
rn.server.nodemanager.containermanager.application.ApplicationEventType for class or
g.apache.hadoop.yarn.server.nodemanager.containermanager.ContainerManagerImpl$Applic
ationEventDispatcher
24/04/28 21:20:48 INFO event.AsyncDispatcher: Registering class org.apache.hadoop.ya
rn.server.nodemanager.containermanager.localizer.event.LocalizationEventType for cla
ss org.apache.hadoop.yarn.server.nodemanager.containermanager.ContainerManagerImpl$L
ocalizationEventHandlerWrapper
24/04/28 21:20:48 INFO event.AsyncDispatcher: Registering class org.apache.hadoop.ya
rn.server.nodemanager.containermanager.AuxServicesEventType for class org.apache.had
oop.yarn.server.nodemanager.containermanager.AuxServices
24/04/28 21:20:48 INFO event.AsyncDispatcher: Registering class org.apache.hadoop.ya
rn.server.nodemanager.containermanager.monitor.ContainersMonitorEventType for class
org.apache.hadoop.yarn.server.nodemanager.containermanager.monitor.ContainersMonitor
Impl
24/04/28 21:20:48 INFO event.AsyncDispatcher: Registering class org.apache.hadoop.ya
rn.server.nodemanager.containermanager.launcher.ContainersLauncherEventType for clas
s org.apache.hadoop.yarn.server.nodemanager.containermanager.launcher.ContainersLaun
cher
24/04/28 21:20:48 INFO checker.ThrottledAsyncChecker: Scheduling a check for [DISK]f
ile:/opt/amazon/hadoop/hdfs/datanode/
24/04/28 21:20:48 INFO event.AsyncDispatcher: Registering class org.apache.hadoop.ya
rn.server.nodemanager.containermanager.scheduler.ContainerSchedulerEventType for cla
ss org.apache.hadoop.yarn.server.nodemanager.containermanager.scheduler.ContainerSch
eduler
24/04/28 21:20:48 INFO security.AMRMTOKENSecretManager: AMRMTOKENKeyRollingInterval:
86400000ms and AMRMTOKENKeyActivationDelay: 900000 ms
24/04/28 21:20:48 INFO event.AsyncDispatcher: Registering class org.apache.hadoop.ya
rn.server.nodemanager.ContainerManagerEventType for class org.apache.hadoop.yarn.ser
ver.nodemanager.containermanager.ContainerManagerImpl
24/04/28 21:20:48 INFO event.AsyncDispatcher: Registering class org.apache.hadoop.ya
```

```
rn.server.nodemanager.NodeManagerEventType for class org.apache.hadoop.yarn.server.n
odemanager.NodeManager
24/04/28 21:20:48 INFO util.JvmPauseMonitor: Starting JVM pause monitor
24/04/28 21:20:48 INFO event.AsyncDispatcher: Registering class org.apache.hadoop.ya
rn.server.resourcemanager.recovery.RMStateStoreEventType for class org.apache.hadoo
p.yarn.server.resourcemanager.recovery.RMStateStore$ForwardingEventHandler
24/04/28 21:20:48 INFO event.AsyncDispatcher: Registering class org.apache.hadoop.ya
rn.server.resourcemanager.NodesListManagerEventType for class org.apache.hadoop.yar
n.server.resourcemanager.NodesListManager
24/04/28 21:20:48 INFO resourcemanager.ResourceManager: Using Scheduler: org.apache.
hadoop.yarn.server.resourcemanager.scheduler.capacity.CapacityScheduler
24/04/28 21:20:48 INFO hdfs.DFSUtil: Starting Web-server for hdfs at: http://0.0.0.
0:50070
24/04/28 21:20:48 INFO impl.MetricsConfig: loaded properties from hadoop-metrics2.pr
operties
24/04/28 21:20:48 INFO event.AsyncDispatcher: Registering class org.apache.hadoop.ya
rn.server.resourcemanager.scheduler.event.SchedulerEventType for class org.apache.ha
dooop.yarn.event.EventDispatcher
24/04/28 21:20:48 INFO event.AsyncDispatcher: Registering class org.apache.hadoop.ya
rn.server.resourcemanager.rmapp.RMAppEventType for class org.apache.hadoop.yarn.serv
er.resourcemanager.ResourceManager$ApplicationEventDispatcher
24/04/28 21:20:48 INFO event.AsyncDispatcher: Registering class org.apache.hadoop.ya
rn.server.resourcemanager.rmapp.attempt.RMAppAttemptEventType for class org.apache.h
adoop.yarn.server.resourcemanager.ResourceManager$ApplicationAttemptEventDispatcher
24/04/28 21:20:48 INFO event.AsyncDispatcher: Registering class org.apache.hadoop.ya
rn.server.resourcemanager.rmnode.RMNodeEventType for class org.apache.hadoop.yarn.se
rver.resourcemanager.ResourceManager$NodeEventDispatcher
24/04/28 21:20:48 INFO impl.MetricsConfig: loaded properties from hadoop-metrics2.pr
operties
24/04/28 21:20:48 INFO mortbay.log: Logging to org.slf4j.impl.Log4jLoggerAdapter(or
g.mortbay.log) via org.mortbay.log.Slf4jLog
24/04/28 21:20:48 INFO server.AuthenticationFilter: Unable to initialize FileSignerS
ecretProvider, falling back to use random secrets.
24/04/28 21:20:48 INFO impl.MetricsConfig: loaded properties from hadoop-metrics2.pr
operties
24/04/28 21:20:48 INFO http.HttpRequestLog: Http request log for http.requests.namen
ode is not defined
24/04/28 21:20:48 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 1
0 second(s).
24/04/28 21:20:48 INFO impl.MetricsSystemImpl: NodeManager metrics system started
24/04/28 21:20:48 INFO http.HttpServer2: Added global filter 'safety' (class=org.apa
che.hadoop.http.HttpServer2$QuotingInputFilter)
24/04/28 21:20:48 INFO http.HttpServer2: Added filter static_user_filter (class=org.
apache.hadoop.http.lib.StaticUserWebFilter$StaticUserFilter) to context hdfs
24/04/28 21:20:48 INFO http.HttpServer2: Added filter static_user_filter (class=org.
apache.hadoop.http.lib.StaticUserWebFilter$StaticUserFilter) to context logs
24/04/28 21:20:48 INFO http.HttpServer2: Added filter static_user_filter (class=org.
apache.hadoop.http.lib.StaticUserWebFilter$StaticUserFilter) to context static
24/04/28 21:20:48 INFO nodemanager.DirectoryCollection: Disk Validator: yarn.nodeman
ager.disk-validator is loaded.
24/04/28 21:20:48 INFO nodemanager.DirectoryCollection: Disk Validator: yarn.nodeman
ager.disk-validator is loaded.
24/04/28 21:20:48 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 1
0 second(s).
24/04/28 21:20:48 INFO impl.MetricsSystemImpl: ResourceManager metrics system starte
d
```

24/04/28 21:20:48 INFO security.YarnAuthorizationProvider: org.apache.hadoop.yarn.security.ConfiguredYarnAuthorizer is instantiated.

24/04/28 21:20:48 INFO event.AsyncDispatcher: Registering class org.apache.hadoop.yarn.server.resourcemanager.RMAppManagerEventType for class org.apache.hadoop.yarn.server.resourcemanager.RMAppManager

24/04/28 21:20:48 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).

24/04/28 21:20:48 INFO impl.MetricsSystemImpl: DataNode metrics system started

24/04/28 21:20:48 INFO event.AsyncDispatcher: Registering class org.apache.hadoop.yarn.server.resourcemanager.amlauncher.AMLauncherEventType for class org.apache.hadoop.yarn.server.resourcemanager.amlauncher.ApplicationMasterLauncher

24/04/28 21:20:48 INFO nodemanager.NodeResourceMonitorImpl: Using ResourceCalculatorPlugin : org.apache.hadoop.yarn.util.ResourceCalculatorPlugin@4a83a74a

24/04/28 21:20:48 INFO resourcemanager.RMNMIInfo: Registered RMNMIInfo MBean

24/04/28 21:20:48 INFO monitor.RMAppLifetimeMonitor: Application lifetime monitor interval set to 3000 ms.

24/04/28 21:20:48 INFO event.AsyncDispatcher: Registering class org.apache.hadoop.yarn.server.nodemanager.containermanager.loghandler.event.LogHandlerEventType for class org.apache.hadoop.yarn.server.nodemanager.containermanager.loghandler.NonAggregatingLogHandler

24/04/28 21:20:48 INFO event.AsyncDispatcher: Registering class org.apache.hadoop.yarn.server.nodemanager.containermanager.localizer.sharedcache.SharedCacheUploadEventType for class org.apache.hadoop.yarn.server.nodemanager.containermanager.localizer.sharedcache.SharedCacheUploadService

24/04/28 21:20:48 INFO containermanager.ContainerManagerImpl: AMRMPProxyService is disabled

24/04/28 21:20:48 INFO localizer.ResourceLocalizationService: per directory file limit = 8192

24/04/28 21:20:48 INFO util.HostsFileReader: Refreshing hosts (include/exclude) list

24/04/28 21:20:48 INFO conf.Configuration: found resource capacity-scheduler.xml at file:/etc/hadoop/conf.empty/capacity-scheduler.xml

24/04/28 21:20:48 INFO scheduler.AbstractYarnScheduler: Minimum allocation = <memory:1, vCores:1>

24/04/28 21:20:48 INFO scheduler.AbstractYarnScheduler: Maximum allocation = <memory:31784, vCores:8>

24/04/28 21:20:48 INFO http.HttpServer2: Added filter 'org.apache.hadoop.hdfs.web.AuthFilter' (class=org.apache.hadoop.hdfs.web.AuthFilter)

24/04/28 21:20:48 INFO http.HttpServer2: addJerseyResourcePackage: packageName=org.apache.hadoop.hdfs.server.namenode.web.resources;org.apache.hadoop.hdfs.web.resources, pathSpec=/webhdfs/v1/\*

24/04/28 21:20:48 INFO resource.ResourceUtils: Adding resource type - name = memory-mb, units = Mi, type = COUNTABLE

24/04/28 21:20:48 INFO resource.ResourceUtils: Adding resource type - name = vcores, units = , type = COUNTABLE

24/04/28 21:20:48 INFO capacity.CapacitySchedulerConfiguration: max alloc mb per queue for root is undefined

24/04/28 21:20:48 INFO capacity.CapacitySchedulerConfiguration: max alloc vcore per queue for root is undefined

24/04/28 21:20:48 INFO http.HttpServer2: Jetty bound to port 50070

24/04/28 21:20:48 INFO mortbay.log: jetty-6.1.26-emr

24/04/28 21:20:48 INFO capacity.ParentQueue: root, capacity=1.0, absoluteCapacity=1.0, maxCapacity=1.0, absoluteMaxCapacity=1.0, state=RUNNING, acls=ADMINISTER\_QUEUE:\*SUBMIT\_APP:\*, labels=\*, reservationsContinueLooking=true, orderingPolicy=utilization, priority=0

24/04/28 21:20:48 INFO capacity.ParentQueue: Initialized parent-queue root name=root, fullname=root

```

24/04/28 21:20:48 INFO resource.ResourceUtils: Adding resource type - name = memory-
mb, units = Mi, type = COUNTABLE
24/04/28 21:20:48 INFO resource.ResourceUtils: Adding resource type - name = vcores,
units = , type = COUNTABLE
24/04/28 21:20:48 INFO capacity.CapacitySchedulerConfiguration: max alloc mb per que
ue for root.default is undefined
24/04/28 21:20:48 INFO capacity.CapacitySchedulerConfiguration: max alloc vcore per
queue for root.default is undefined
24/04/28 21:20:48 INFO capacity.LeafQueue: Initializing default
capacity = 1.0 [= (float) configuredCapacity / 100 ]
absoluteCapacity = 1.0 [= parentAbsoluteCapacity * capacity ]
maxCapacity = 1.0 [= configuredMaxCapacity ]
absoluteMaxCapacity = 1.0 [= 1.0 maximumCapacity undefined, (parentAbsoluteMaxCapaci
ty * maximumCapacity) / 100 otherwise ]
userLimit = 100 [= configuredUserLimit ]
userLimitFactor = 1.0 [= configuredUserLimitFactor ]
maxApplications = 10000 [= configuredMaximumSystemApplicationsPerQueue or (int)(conf
iguredMaximumSystemApplications * absoluteCapacity)]
maxApplicationsPerUser = 10000 [= (int)(maxApplications * (userLimit / 100.0f) * use
rLimitFactor) ]
usedCapacity = 0.0 [= usedResourcesMemory / (clusterResourceMemory * absoluteCapacit
y)]
absoluteUsedCapacity = 0.0 [= usedResourcesMemory / clusterResourceMemory]
maxAMResourcePerQueuePercent = 0.1 [= configuredMaximumAMResourcePercent ]
minimumAllocationFactor = 0.9999685 [= (float)(maximumAllocationMemory - minimumAllo
cationMemory) / maximumAllocationMemory ]
maximumAllocation = <memory:31784, vCores:8> [= configuredMaxAllocation ]
numContainers = 0 [= currentNumContainers ]
state = RUNNING [= configuredState ]
acls = ADMINISTER_QUEUE:*SUBMIT_APP:* [= configuredAcls ]
nodeLocalityDelay = 40
rackLocalityAdditionalDelay = -1
labels=*,
reservationsContinueLooking = true
preemptionDisabled = true
defaultAppPriorityPerQueue = 0
priority = 0
maxLifetime = -1 seconds
defaultLifetime = -1 seconds
24/04/28 21:20:48 INFO capacity.CapacitySchedulerQueueManager: Initialized queue: de
fault: capacity=1.0, absoluteCapacity=1.0, usedResources=<memory:0, vCores:0>, usedC
apacity=0.0, absoluteUsedCapacity=0.0, numApps=0, numContainers=0
24/04/28 21:20:48 INFO capacity.CapacitySchedulerQueueManager: Initialized queue: ro
ot: numChildQueue= 1, capacity=1.0, absoluteCapacity=1.0, usedResources=<memory:0, v
Cores:0>usedCapacity=0.0, numApps=0, numContainers=0
24/04/28 21:20:48 INFO localizer.ResourceLocalizationService: Disk Validator: yarn.n
odemanager.disk-validator is loaded.
24/04/28 21:20:48 INFO capacity.CapacitySchedulerQueueManager: Initialized root queu
e root: numChildQueue= 1, capacity=1.0, absoluteCapacity=1.0, usedResources=<memory:
0, vCores:0>usedCapacity=0.0, numApps=0, numContainers=0
24/04/28 21:20:48 INFO placement.UserGroupMappingPlacementRule: Initialized queue ma
ppings, override: false
24/04/28 21:20:48 INFO capacity.WorkflowPriorityMappingsManager: Initialized workflo
w priority mappings, override: false
24/04/28 21:20:48 INFO capacity.CapacityScheduler: Initialized CapacityScheduler wit
h calculator=class org.apache.hadoop.yarn.util.resource.DefaultResourceCalculator, m

```

```
inimumAllocation=<<memory:1, vCores:1>>, maximumAllocation=<<memory:31784, vCores:8>>, asynchronousScheduling=false, asyncScheduleInterval=5ms
24/04/28 21:20:48 INFO event.AsyncDispatcher: Registering class org.apache.hadoop.yarn.server.nodemanager.containermanager.localizer.event.LocalizerEventType for class org.apache.hadoop.yarn.server.nodemanager.containermanager.localizer.ResourceLocalizationService$LocalizerTracker
24/04/28 21:20:48 INFO conf.Configuration: dynamic-resources.xml not found
24/04/28 21:20:48 INFO resourcemanager.AMSProcessingChain: Initializing AMS Processing chain. Root Processor=[org.apache.hadoop.yarn.server.resourcemanager.DefaultAMSPProcessor].
24/04/28 21:20:48 INFO monitor.ContainersMonitorImpl: Using ResourceCalculatorPlugin : org.apache.hadoop.yarn.util.ResourceCalculatorPlugin@710c2b53
24/04/28 21:20:48 INFO monitor.ContainersMonitorImpl: Using ResourceCalculatorProcessTree : null
24/04/28 21:20:48 INFO monitor.ContainersMonitorImpl: Physical memory check enabled: true
24/04/28 21:20:48 INFO monitor.ContainersMonitorImpl: Virtual memory check enabled: true
24/04/28 21:20:48 INFO monitor.ContainersMonitorImpl: ContainersMonitor enabled: true
24/04/28 21:20:48 INFO resourcemanager.ResourceManager: TimelineServicePublisher is not configured
24/04/28 21:20:48 WARN monitor.ContainersMonitorImpl: NodeManager configured with 31.0 G physical memory allocated to containers, which is more than 80% of the total physical memory available (30.6 G). Thrashing might happen.
24/04/28 21:20:48 INFO containermanager.ContainerManagerImpl: Not a recoverable state store. Nothing to recover.
24/04/28 21:20:48 INFO conf.Configuration: resource-types.xml not found
24/04/28 21:20:48 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
24/04/28 21:20:48 INFO resource.ResourceUtils: Adding resource type - name = memory-mb, units = Mi, type = COUNTABLE
24/04/28 21:20:48 INFO resource.ResourceUtils: Adding resource type - name = vcores, units = , type = COUNTABLE
24/04/28 21:20:49 INFO conf.Configuration: node-resources.xml not found
24/04/28 21:20:49 INFO mortbay.log: Logging to org.slf4j.impl.Log4jLoggerAdapter(org.mortbay.log) via org.mortbay.log.Slf4jLog
24/04/28 21:20:49 INFO resource.ResourceUtils: Unable to find 'node-resources.xml'.
24/04/28 21:20:49 INFO resource.ResourceUtils: Adding resource type - name = memory-mb, units = Mi, type = COUNTABLE
24/04/28 21:20:49 INFO resource.ResourceUtils: Adding resource type - name = vcores, units = , type = COUNTABLE
24/04/28 21:20:49 INFO nodemanager.NodeStatusUpdaterImpl: Nodemanager resources is set to: <memory:31784, vCores:8>
24/04/28 21:20:49 INFO server.AuthenticationFilter: Unable to initialize FileSignerSecretProvider, falling back to use random secrets.
24/04/28 21:20:49 INFO nodemanager.NodeStatusUpdaterImpl: Initialized nodemanager with : physical-memory=31784 virtual-memory=158920 virtual-cores=8
24/04/28 21:20:49 INFO http.HttpRequestLog: Http request log for http.requests.resourcemanager is not defined
24/04/28 21:20:49 INFO http.HttpServer2: Added global filter 'safety' (class=org.apache.hadoop.http.HttpServer2$QuotingInputFilter)
24/04/28 21:20:49 INFO http.HttpServer2: Added filter RMAuthenticationFilter (class=org.apache.hadoop.yarn.server.security.http.RMAuthenticationFilter) to context cluster
24/04/28 21:20:49 INFO http.HttpServer2: Added filter RMAuthenticationFilter (class=org.apache.hadoop.yarn.server.security.http.RMAuthenticationFilter) to context logs
```



24/04/28 21:20:49 INFO http.HttpServer2: Added filter RMAuthenticationFilter (class=org.apache.hadoop.yarn.server.security.http.RMAuthenticationFilter) to context static

24/04/28 21:20:49 INFO http.HttpServer2: Added filter static\_user\_filter (class=org.apache.hadoop.http.lib.StaticUserWebFilter\$StaticUserFilter) to context cluster

24/04/28 21:20:49 INFO http.HttpServer2: Added filter static\_user\_filter (class=org.apache.hadoop.http.lib.StaticUserWebFilter\$StaticUserFilter) to context logs

24/04/28 21:20:49 INFO http.HttpServer2: Added filter static\_user\_filter (class=org.apache.hadoop.http.lib.StaticUserWebFilter\$StaticUserFilter) to context static

24/04/28 21:20:49 INFO http.HttpServer2: adding path spec: /cluster/\*

24/04/28 21:20:49 INFO http.HttpServer2: adding path spec: /ws/\*

24/04/28 21:20:49 INFO common.Util: dfs.datanode.fileio.profiling.sampling.percentage set to 0. Disabling file IO profiling

24/04/28 21:20:49 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue queueCapacity: 2000 scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler

24/04/28 21:20:49 INFO datanode.BlockScanner: Initialized block scanner with targetBytesPerSec 1048576

24/04/28 21:20:49 INFO datanode.DataNode: Configured hostname is algo-1

24/04/28 21:20:49 INFO common.Util: dfs.datanode.fileio.profiling.sampling.percentage set to 0. Disabling file IO profiling

24/04/28 21:20:49 WARN conf.Configuration: No unit for dfs.datanode.outliers.report.interval(1800000) assuming MILLISECONDS

24/04/28 21:20:49 INFO datanode.DataNode: Starting DataNode with maxLockedMemory = 0

24/04/28 21:20:49 INFO ipc.Server: Starting Socket Reader #1 for port 0

24/04/28 21:20:49 INFO datanode.DataNode: Opened streaming server at /0.0.0.0:50010

24/04/28 21:20:49 INFO datanode.DataNode: Balancing bandwidth is 10485760 bytes/s

24/04/28 21:20:49 INFO datanode.DataNode: Number threads for balancing is 50

24/04/28 21:20:49 INFO mortbay.log: Logging to org.slf4j.impl.Log4jLoggerAdapter(org.mortbay.log) via org.mortbay.log.Slf4jLog

24/04/28 21:20:49 INFO mortbay.log: Started HttpServer2\$SelectChannelConnectorWithSafeStartup@0.0.0.0:50070

24/04/28 21:20:49 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ContainerManagementProtocolPB to the server

24/04/28 21:20:49 INFO server.AuthenticationFilter: Unable to initialize FileSignerSecretProvider, falling back to use random secrets.

24/04/28 21:20:49 INFO ipc.Server: IPC Server Responder: starting

24/04/28 21:20:49 INFO http.HttpRequestLog: Http request log for http.requests.datanode is not defined

24/04/28 21:20:49 INFO http.HttpServer2: Added global filter 'safety' (class=org.apache.hadoop.http.HttpServer2\$QuotingInputFilter)

24/04/28 21:20:49 INFO ipc.Server: IPC Server listener on 0: starting

24/04/28 21:20:49 INFO http.HttpServer2: Added filter static\_user\_filter (class=org.apache.hadoop.http.lib.StaticUserWebFilter\$StaticUserFilter) to context datanode

24/04/28 21:20:49 INFO http.HttpServer2: Added filter static\_user\_filter (class=org.apache.hadoop.http.lib.StaticUserWebFilter\$StaticUserFilter) to context logs

24/04/28 21:20:49 INFO http.HttpServer2: Added filter static\_user\_filter (class=org.apache.hadoop.http.lib.StaticUserWebFilter\$StaticUserFilter) to context static

24/04/28 21:20:49 INFO http.HttpServer2: Jetty bound to port 46759

24/04/28 21:20:49 INFO mortbay.log: jetty-6.1.26-emr

24/04/28 21:20:49 INFO security.NMContainerTokenSecretManager: Updating node address : algo-1:42607

24/04/28 21:20:49 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue queueCapacity: 500 scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler

24/04/28 21:20:49 INFO ipc.Server: Starting Socket Reader #1 for port 8040

```
24/04/28 21:20:49 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.
yarn.server.nodemanager.api.LocalizationProtocolPB to the server
24/04/28 21:20:49 INFO ipc.Server: IPC Server Responder: starting
24/04/28 21:20:49 INFO ipc.Server: IPC Server listener on 8040: starting
24/04/28 21:20:49 INFO localizer.ResourceLocalizationService: Localizer started on p
ort 8040
24/04/28 21:20:49 INFO webapp.WebApps: Registered webapp guice modules
24/04/28 21:20:49 INFO containermanager.ContainerManagerImpl: ContainerManager start
ed at /10.0.222.169:42607
24/04/28 21:20:49 INFO containermanager.ContainerManagerImpl: ContainerManager bound
to algo-1/10.0.222.169:0
24/04/28 21:20:49 INFO webapp.WebServer: Instantiating NMWebApp at algo-1:8042
24/04/28 21:20:49 WARN namenode.FSNamesystem: Only one image storage directory (dfs.
namenode.name.dir) configured. Beware of data loss due to lack of redundant storage
directories!
24/04/28 21:20:49 WARN namenode.FSNamesystem: Only one namespace edits storage direc
tory (dfs.namenode.edits.dir) configured. Beware of data loss due to lack of redunda
nt storage directories!
24/04/28 21:20:49 INFO http.HttpServer2: Jetty bound to port 8088
24/04/28 21:20:49 INFO mortbay.log: jetty-6.1.26-emr
24/04/28 21:20:49 INFO namenode.FSEditLog: Edit logging is async:true
24/04/28 21:20:49 INFO mortbay.log: Logging to org.slf4j.impl.Log4jLoggerAdapter(or
g.mortbay.log) via org.mortbay.log.Slf4jLog
24/04/28 21:20:49 INFO namenode.FSNamesystem: KeyProvider: null
24/04/28 21:20:49 INFO namenode.FSNamesystem: fsLock is fair: true
24/04/28 21:20:49 INFO namenode.FSNamesystem: Detailed lock hold time metrics enable
d: false
24/04/28 21:20:49 INFO namenode.FSNamesystem: fsOwner = root (auth:SIMPL
E)
24/04/28 21:20:49 INFO namenode.FSNamesystem: supergroup = supergroup
24/04/28 21:20:49 INFO namenode.FSNamesystem: isPermissionEnabled = true
24/04/28 21:20:49 INFO namenode.FSNamesystem: HA Enabled: false
24/04/28 21:20:49 INFO common.Util: dfs.datanode.fileio.profiling.sampling.percentag
e set to 0. Disabling file IO profiling
24/04/28 21:20:49 INFO mortbay.log: Started HttpServer2$SelectChannelConnectorWithSa
feStartup@localhost:46759
24/04/28 21:20:49 INFO mortbay.log: Extract jar:file:/usr/lib/hadoop/hadoop-yarn-com
mon-2.10.0-amzn-0.jar!/webapps/cluster to work/Jetty_10_0_222_169_8088_cluster____.b
pwzpj6/webapp
24/04/28 21:20:49 INFO blockmanagement.DatanodeManager: dfs.block.invalidate.limit:
configured=1000, counted=60, effected=1000
24/04/28 21:20:49 INFO blockmanagement.DatanodeManager: dfs.namenode.datanode.regist
ration.ip-hostname-check=true
24/04/28 21:20:49 INFO blockmanagement.BlockManager: dfs.namenode.startup.delay.bloc
k.deletion.sec is set to 000:00:00:00.000
24/04/28 21:20:49 INFO blockmanagement.BlockManager: The block deletion will start a
round 2024 Apr 28 21:20:49
24/04/28 21:20:49 INFO util.GSet: Computing capacity for map BlocksMap
24/04/28 21:20:49 INFO util.GSet: VM type = 64-bit
24/04/28 21:20:49 INFO util.GSet: 2.0% max memory 889 MB = 17.8 MB
24/04/28 21:20:49 INFO util.GSet: capacity = 2^21 = 2097152 entries
24/04/28 21:20:49 INFO blockmanagement.BlockManager: dfs.block.access.token.enable=f
alse
24/04/28 21:20:49 WARN conf.Configuration: No unit for dfs.heartbeat.interval(3) ass
uming SECONDS
24/04/28 21:20:49 WARN conf.Configuration: No unit for dfs.namenode.safemode.extensi
```

```

on(30000) assuming MILLISECOND
24/04/28 21:20:49 INFO blockmanagement.BlockManagerSafeMode: dfs.namenode.safemode.threshold-pct = 0.9990000128746033
24/04/28 21:20:49 INFO blockmanagement.BlockManagerSafeMode: dfs.namenode.safemode.min.datanodes = 0
24/04/28 21:20:49 INFO blockmanagement.BlockManagerSafeMode: dfs.namenode.safemode.extension = 30000
24/04/28 21:20:49 INFO blockmanagement.BlockManager: defaultReplication = 3
24/04/28 21:20:49 INFO blockmanagement.BlockManager: maxReplication = 512
24/04/28 21:20:49 INFO blockmanagement.BlockManager: minReplication = 1
24/04/28 21:20:49 INFO blockmanagement.BlockManager: maxReplicationStreams = 2
24/04/28 21:20:49 INFO blockmanagement.BlockManager: replicationRecheckInterval = 3000
24/04/28 21:20:49 INFO blockmanagement.BlockManager: encryptDataTransfer = false
24/04/28 21:20:49 INFO blockmanagement.BlockManager: maxNumBlocksToLog = 1000
24/04/28 21:20:49 INFO namenode.FSNamesystem: Append Enabled: true
24/04/28 21:20:49 INFO server.AuthenticationFilter: Unable to initialize FileSignerSecretProvider, falling back to use random secrets.
24/04/28 21:20:49 INFO namenode.FSDirectory: GLOBAL serial map: bits=24 maxEntries=16777215
24/04/28 21:20:49 INFO http.HttpRequestLog: Http request log for http.requests.nodemanager is not defined
24/04/28 21:20:49 INFO util.GSet: Computing capacity for map INodeMap
24/04/28 21:20:49 INFO util.GSet: VM type = 64-bit
24/04/28 21:20:49 INFO util.GSet: 1.0% max memory 889 MB = 8.9 MB
24/04/28 21:20:49 INFO util.GSet: capacity = 2^20 = 1048576 entries
24/04/28 21:20:49 INFO namenode.FSDirectory: ACLs enabled? false
24/04/28 21:20:49 INFO namenode.FSDirectory: XAttrs enabled? true
24/04/28 21:20:49 INFO namenode.NameNode: Caching file names occurring more than 10 times
24/04/28 21:20:49 INFO http.HttpServer2: Added global filter 'safety' (class=org.apache.hadoop.http.HttpServer2$QuotingInputFilter)
24/04/28 21:20:49 INFO snapshot.SnapshotManager: Loaded config captureOpenFiles: false skipCaptureAccessTimeOnlyChange: false
24/04/28 21:20:49 INFO http.HttpServer2: Added filter static_user_filter (class=org.apache.hadoop.http.lib.StaticUserWebFilter$StaticUserFilter) to context node
24/04/28 21:20:49 INFO util.GSet: Computing capacity for map cachedBlocks
24/04/28 21:20:49 INFO util.GSet: VM type = 64-bit
24/04/28 21:20:49 INFO util.GSet: 0.25% max memory 889 MB = 2.2 MB
24/04/28 21:20:49 INFO util.GSet: capacity = 2^18 = 262144 entries
24/04/28 21:20:49 INFO http.HttpServer2: Added filter static_user_filter (class=org.apache.hadoop.http.lib.StaticUserWebFilter$StaticUserFilter) to context logs
24/04/28 21:20:49 INFO http.HttpServer2: Added filter static_user_filter (class=org.apache.hadoop.http.lib.StaticUserWebFilter$StaticUserFilter) to context static
24/04/28 21:20:49 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
24/04/28 21:20:49 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
24/04/28 21:20:49 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
24/04/28 21:20:49 INFO http.HttpServer2: Added filter authentication (class=org.apache.hadoop.security.AuthenticationWithProxyUserFilter) to context node
24/04/28 21:20:49 INFO http.HttpServer2: Added filter authentication (class=org.apac

```

```

he.hadoop.security.AuthenticationWithProxyUserFilter) to context logs
24/04/28 21:20:49 INFO http.HttpServer2: Added filter authentication (class=org.apac
he.hadoop.security.AuthenticationWithProxyUserFilter) to context static
24/04/28 21:20:49 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
24/04/28 21:20:49 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total hea
p and retry cache entry expiry time is 600000 millis
24/04/28 21:20:49 INFO util.GSet: Computing capacity for map NameNodeRetryCache
24/04/28 21:20:49 INFO util.GSet: VM type          = 64-bit
24/04/28 21:20:49 INFO util.GSet: 0.029999999329447746% max memory 889 MB = 273.1 KB
24/04/28 21:20:49 INFO util.GSet: capacity        = 2^15 = 32768 entries
24/04/28 21:20:49 INFO http.HttpServer2: adding path spec: /node/*
24/04/28 21:20:49 INFO http.HttpServer2: adding path spec: /ws/*
24/04/28 21:20:49 INFO common.Storage: Lock on /opt/amazon/hadoop/hdfs/namenode/in_u
se.lock acquired by nodename 102@algo-1
24/04/28 21:20:49 INFO namenode.FileJournalManager: Recovering unfinalized segments
in /opt/amazon/hadoop/hdfs/namenode/current
24/04/28 21:20:49 INFO namenode.FSImage: No edit log streams selected.
24/04/28 21:20:49 INFO namenode.FSImage: Planning to load image: FSImageFile(file=/o
pt/amazon/hadoop/hdfs/namenode/current/fsimage_0000000000000000000, cpktTxId=0000000
0000000000000)
24/04/28 21:20:49 INFO namenode.FSImageFormatPBINode: Loading 1 INodes.
24/04/28 21:20:49 INFO web.DatanodeHttpServer: Listening HTTP traffic on /0.0.0.0:50
075
24/04/28 21:20:49 INFO util.JvmPauseMonitor: Starting JVM pause monitor
24/04/28 21:20:49 INFO datanode.DataNode: dnUserName = root
24/04/28 21:20:49 INFO datanode.DataNode: supergroup = supergroup
24/04/28 21:20:49 INFO namenode.FSImageFormatPBINode: Successfully loaded 1 inodes
24/04/28 21:20:49 INFO namenode.FSImageFormatProtobuf: Loaded FSImage in 0 seconds.
24/04/28 21:20:49 INFO namenode.FSImage: Loaded image for txid 0 from /opt/amazon/ha
doo/hdfs/namenode/current/fsimage_00000000000000000000
24/04/28 21:20:49 INFO namenode.FSNamesystem: Need to save fs image? false (staleIma
ge=false, haEnabled=false, isRollingUpgrade=false)
24/04/28 21:20:49 INFO namenode.FSEditLog: Starting log segment at 1
24/04/28 21:20:49 INFO ipc.CallQueueManager: Using callQueue: class java.util.concur
rent.LinkedBlockingQueue queueCapacity: 1000 scheduler: class org.apache.hadoop.ipc.
DefaultRpcScheduler
24/04/28 21:20:50 INFO ipc.Server: Starting Socket Reader #1 for port 50020
24/04/28 21:20:50 INFO delegation.AbstractDelegationTokenSecretManager: Updating the
current master key for generating delegation tokens
24/04/28 21:20:50 INFO datanode.DataNode: Opened IPC server at /0.0.0.0:50020
24/04/28 21:20:50 INFO delegation.AbstractDelegationTokenSecretManager: Starting exp
ired delegation token remover thread, tokenRemoverScanInterval=60 min(s)
24/04/28 21:20:50 INFO delegation.AbstractDelegationTokenSecretManager: Updating the
current master key for generating delegation tokens
24/04/28 21:20:50 INFO datanode.DataNode: Refresh request received for nameservices:
null
24/04/28 21:20:50 INFO datanode.DataNode: Starting BPOfferServices for nameservices:
<default>
24/04/28 21:20:50 INFO namenode.NameCache: initialized with 0 entries 0 lookups
24/04/28 21:20:50 INFO namenode.FSNamesystem: Finished loading FSImage in 261 msec
24/04/28 21:20:50 INFO datanode.DataNode: Block pool <registering> (Datanode Uuid un
assigned) service to algo-1/10.0.222.169:8020 starting to offer service
24/04/28 21:20:50 INFO ipc.Server: IPC Server Responder: starting
Apr 28, 2024 9:20:50 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFac
tory register
INFO: Registering org.apache.hadoop.yarn.server.resourcemanager.webapp.JAXBContextRe

```

```
solver as a provider class
Apr 28, 2024 9:20:50 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory register
INFO: Registering org.apache.hadoop.yarn.server.resourcemanager.webapp.RMWebServices as a root resource class
Apr 28, 2024 9:20:50 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory register
INFO: Registering org.apache.hadoop.yarn.webapp.GenericExceptionHandler as a provider class
Apr 28, 2024 9:20:50 PM com.sun.jersey.server.impl.application.WebApplicationImpl _initiate
INFO: Initiating Jersey application, version 'Jersey: 1.9 09/02/2011 11:17 AM'
24/04/28 21:20:50 INFO ipc.Server: IPC Server listener on 50020: starting
24/04/28 21:20:50 INFO webapp.WebApps: Registered webapp guice modules
Apr 28, 2024 9:20:50 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.resourcemanager.webapp.JAXBContextResolver to GuiceManagedComponentProvider with the scope "Singleton"
24/04/28 21:20:50 INFO http.HttpServer2: Jetty bound to port 8042
24/04/28 21:20:50 INFO mortbay.log: jetty-6.1.26-emr
24/04/28 21:20:50 INFO namenode.NameNode: RPC server is binding to algo-1:8020
24/04/28 21:20:50 INFO namenode.NameNode: Enable NameNode state context:false
24/04/28 21:20:50 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue queueCapacity: 1000 scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler
24/04/28 21:20:50 INFO ipc.Server: Starting Socket Reader #1 for port 8020
24/04/28 21:20:50 INFO mortbay.log: Extract jar:file:/usr/lib/hadoop/hadoop-yarn-common-2.10.0-amzn-0.jar!/webapps/node to work/Jetty_algo.1_8042_node____.afclh/webapp
24/04/28 21:20:50 INFO namenode.NameNode: Clients are to use algo-1:8020 to access this namenode/service.
24/04/28 21:20:50 INFO namenode.FSNamesystem: Registered FSNamesystemState MBean
24/04/28 21:20:50 INFO namenode.LeaseManager: Number of blocks under construction: 0
24/04/28 21:20:50 INFO blockmanagement.BlockManager: initializing replication queues
24/04/28 21:20:50 INFO hdfs.StateChange: STATE* Leaving safe mode after 0 secs
24/04/28 21:20:50 INFO hdfs.StateChange: STATE* Network topology has 0 racks and 0 datanodes
24/04/28 21:20:50 INFO hdfs.StateChange: STATE* UnderReplicatedBlocks has 0 blocks
24/04/28 21:20:50 INFO blockmanagement.BlockManager: Total number of blocks = 0
24/04/28 21:20:50 INFO blockmanagement.BlockManager: Number of invalid blocks = 0
24/04/28 21:20:50 INFO blockmanagement.BlockManager: Number of under-replicated blocks = 0
24/04/28 21:20:50 INFO blockmanagement.BlockManager: Number of over-replicated blocks = 0
24/04/28 21:20:50 INFO blockmanagement.BlockManager: Number of blocks being written = 0
24/04/28 21:20:50 INFO hdfs.StateChange: STATE* Replication Queue initialization scan for invalid, over- and under-replicated blocks completed in 22 msec
24/04/28 21:20:50 INFO ipc.Server: IPC Server Responder: starting
24/04/28 21:20:50 INFO ipc.Server: IPC Server listener on 8020: starting
24/04/28 21:20:50 INFO namenode.NameNode: NameNode RPC up at: algo-1/10.0.222.169:8020
24/04/28 21:20:50 INFO namenode.FSNamesystem: Starting services required for active state
24/04/28 21:20:50 INFO namenode.FSDirectory: Initializing quota with 4 thread(s)
```

24/04/28 21:20:50 INFO namenode.FSDirectory: Quota initialization completed in 6 milliseconds  
name space=1  
storage space=0  
storage types=RAM\_DISK=0, SSD=0, DISK=0, ARCHIVE=0  
24/04/28 21:20:50 INFO blockmanagement.CacheReplicationMonitor: Starting CacheReplicationMonitor with interval 30000 milliseconds  
Apr 28, 2024 9:20:50 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider  
INFO: Binding org.apache.hadoop.yarn.webapp.GenericExceptionHandler to GuiceManagedComponentProvider with the scope "Singleton"  
Apr 28, 2024 9:20:50 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory register  
INFO: Registering org.apache.hadoop.yarn.server.nodemanager.webapp.NMWebServices as a root resource class  
Apr 28, 2024 9:20:50 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory register  
INFO: Registering org.apache.hadoop.yarn.webapp.GenericExceptionHandler as a provider class  
Apr 28, 2024 9:20:50 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory register  
INFO: Registering org.apache.hadoop.yarn.server.nodemanager.webapp.JAXBContextResolver as a provider class  
Apr 28, 2024 9:20:50 PM com.sun.jersey.server.impl.application.WebApplicationImpl \_initiate  
INFO: Initiating Jersey application, version 'Jersey: 1.9 09/02/2011 11:17 AM'  
Apr 28, 2024 9:20:50 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider  
INFO: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.JAXBContextResolver to GuiceManagedComponentProvider with the scope "Singleton"  
Apr 28, 2024 9:20:50 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider  
INFO: Binding org.apache.hadoop.yarn.server.resourcemanager.webapp.RMWebServices to GuiceManagedComponentProvider with the scope "Singleton"  
Apr 28, 2024 9:20:50 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider  
INFO: Binding org.apache.hadoop.yarn.webapp.GenericExceptionHandler to GuiceManagedComponentProvider with the scope "Singleton"  
24/04/28 21:20:50 INFO mortbay.log: Started HttpServer2\$SelectChannelConnectorWithSafeStartup@10.0.222.169:8088  
24/04/28 21:20:50 INFO webapp.WebApps: Web app cluster started at 8088  
24/04/28 21:20:51 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue queueCapacity: 100 scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler  
24/04/28 21:20:51 INFO ipc.Server: Starting Socket Reader #1 for port 8033  
24/04/28 21:20:51 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.server.api.ResourceManagerAdministrationProtocolPB to the server  
24/04/28 21:20:51 INFO ipc.Server: IPC Server Responder: starting  
24/04/28 21:20:51 INFO ipc.Server: IPC Server listener on 8033: starting  
24/04/28 21:20:51 INFO resourcemanager.ResourceManager: Transitioning to active state  
24/04/28 21:20:51 INFO recovery.RMStateStore: Updating AMRMTOKEN  
24/04/28 21:20:51 INFO security.RMContainerTokenSecretManager: Rolling master-key for container-tokens  
24/04/28 21:20:51 INFO security.NMTOKENSecretManagerInRM: Rolling master-key for nm-tokens

24/04/28 21:20:51 INFO delegation.AbstractDelegationTokenSecretManager: Updating the current master key for generating delegation tokens

24/04/28 21:20:51 INFO security.RMDelegationTokenSecretManager: storing master key with keyID 1

24/04/28 21:20:51 INFO recovery.RMStateStore: Storing RMDTMasterKey.

Apr 28, 2024 9:20:51 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider

INFO: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.NMWebServices to GuiceManagedComponentProvider with the scope "Singleton"

24/04/28 21:20:51 INFO delegation.AbstractDelegationTokenSecretManager: Starting expired delegation token remover thread, tokenRemoverScanInterval=60 min(s)

24/04/28 21:20:51 INFO delegation.AbstractDelegationTokenSecretManager: Updating the current master key for generating delegation tokens

24/04/28 21:20:51 INFO security.RMDelegationTokenSecretManager: storing master key with keyID 2

24/04/28 21:20:51 INFO recovery.RMStateStore: Storing RMDTMasterKey.

24/04/28 21:20:51 INFO event.AsyncDispatcher: Registering class org.apache.hadoop.yarn.nodelabels.event.NodeLabelsStoreEventType for class org.apache.hadoop.yarn.nodelabels.CommonNodeLabelsManager\$ForwardingEventHandler

24/04/28 21:20:51 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue queueCapacity: 5000 scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler

24/04/28 21:20:51 INFO ipc.Server: Starting Socket Reader #1 for port 8031

24/04/28 21:20:51 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.server.api.ResourceTrackerPB to the server

24/04/28 21:20:51 INFO ipc.Server: IPC Server Responder: starting

24/04/28 21:20:51 INFO ipc.Server: IPC Server listener on 8031: starting

24/04/28 21:20:51 INFO ipc.Client: Retrying connect to server: algo-1/10.0.222.169:8020. Already tried 0 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECONDS)

24/04/28 21:20:51 INFO mortbay.log: Started HttpServer2\$SelectChannelConnectorWithSafeStartup@algo-1:8042

24/04/28 21:20:51 INFO webapp.WebApps: Web app node started at 8042

24/04/28 21:20:51 INFO nodemanager.NodeStatusUpdaterImpl: Node ID assigned is : algo-1:42607

24/04/28 21:20:51 INFO util.JvmPauseMonitor: Starting JVM pause monitor

24/04/28 21:20:51 INFO client.RMProxy: Connecting to ResourceManager at /10.0.222.169:8031

24/04/28 21:20:51 INFO util.JvmPauseMonitor: Starting JVM pause monitor

24/04/28 21:20:51 INFO nodemanager.NodeStatusUpdaterImpl: Sending out 0 NM container statuses: []

24/04/28 21:20:51 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue queueCapacity: 5000 scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler

24/04/28 21:20:51 INFO nodemanager.NodeStatusUpdaterImpl: Registering with RM using containers :[]

24/04/28 21:20:51 INFO ipc.Server: Starting Socket Reader #1 for port 8030

24/04/28 21:20:51 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationMasterProtocolPB to the server

24/04/28 21:20:51 INFO ipc.Server: IPC Server Responder: starting

24/04/28 21:20:51 INFO ipc.Server: IPC Server listener on 8030: starting

24/04/28 21:20:51 INFO datanode.DataNode: Acknowledging ACTIVE Namenode during handshakeBlock pool <registering> (Datanode Uuid unassigned) service to algo-1/10.0.222.169:8020

24/04/28 21:20:51 INFO common.Storage: Using 1 threads to upgrade data directories (dfs.datanode.parallel.volumes.load.threads.num=1, dataDirs=1)

```
24/04/28 21:20:51 INFO common.Storage: Lock on /opt/amazon/hadoop/hdfs/datanode/in_use.lock acquired by nodename 103@algo-1
24/04/28 21:20:51 INFO common.Storage: Storage directory /opt/amazon/hadoop/hdfs/datanode is not formatted for namespace 861840093. Formatting...
24/04/28 21:20:51 INFO common.Storage: Generated new storageID DS-7bb77e45-c9ec-428c-ac2f-c4e9031acad5 for directory /opt/amazon/hadoop/hdfs/datanode
24/04/28 21:20:51 INFO common.Storage: Analyzing storage directories for bpid BP-393350832-10.0.222.169-1714339246784
24/04/28 21:20:51 INFO common.Storage: Locking is disabled for /opt/amazon/hadoop/hdfs/datanode/current/BP-393350832-10.0.222.169-1714339246784
24/04/28 21:20:51 INFO common.Storage: Block pool storage directory /opt/amazon/hadoop/hdfs/datanode/current/BP-393350832-10.0.222.169-1714339246784 is not formatted for BP-393350832-10.0.222.169-1714339246784. Formatting ...
24/04/28 21:20:51 INFO common.Storage: Formatting block pool BP-393350832-10.0.222.169-1714339246784 directory /opt/amazon/hadoop/hdfs/datanode/current/BP-393350832-10.0.222.169-1714339246784/current
04-28 21:20 smspark-submit INFO      cluster is up
04-28 21:20 smspark-submit INFO      transitioning from status BOOTSTRAPPING to WAITING
04-28 21:20 smspark-submit INFO      starting executor logs watcher
04-28 21:20 smspark-submit INFO      start log event log publisher
Starting executor logs watcher on log_dir: /var/log/yarn
04-28 21:20 sagemaker-spark-event-logs-publisher INFO      Spark event log not enabled.
04-28 21:20 smspark-submit INFO      Waiting for hosts to bootstrap: ['algo-1']
24/04/28 21:20:51 INFO datanode.DataNode: Setting up storage: nsid=861840093;bpid=BP-393350832-10.0.222.169-1714339246784;lv=-57;nsInfo=lv=-63;cid=CID-fa2135ef-5d1b-4954-be9a-b407f6b3a20a;nsid=861840093;c=1714339246784;bpid=BP-393350832-10.0.222.169-1714339246784;dnuid=null
04-28 21:20 smspark-submit INFO      Received host statuses: dict_items([('algo-1', StatusMessage(status='WAITING', timestamp='2024-04-28T21:20:51.558014'))])
24/04/28 21:20:51 INFO datanode.DataNode: Generated and persisted new Datanode UUID 1e229d40-75ae-4892-8b3b-2ec66f1972fb
24/04/28 21:20:51 INFO impl.FsDatasetImpl: Added new volume: DS-7bb77e45-c9ec-428c-ac2f-c4e9031acad5
24/04/28 21:20:51 INFO impl.FsDatasetImpl: Added volume - /opt/amazon/hadoop/hdfs/datanode/current, StorageType: DISK
24/04/28 21:20:51 INFO impl.FsDatasetImpl: Registered FSDatasetState MBean
24/04/28 21:20:51 INFO checker.ThrottledAsyncChecker: Scheduling a check for /opt/amazon/hadoop/hdfs/datanode/current
24/04/28 21:20:51 INFO checker.DatasetVolumeChecker: Scheduled health check for volume /opt/amazon/hadoop/hdfs/datanode/current
24/04/28 21:20:51 INFO impl.FsDatasetImpl: Adding block pool BP-393350832-10.0.222.169-1714339246784
24/04/28 21:20:51 INFO impl.FsDatasetImpl: Scanning block pool BP-393350832-10.0.222.169-1714339246784 on volume /opt/amazon/hadoop/hdfs/datanode/current...
24/04/28 21:20:51 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue queueCapacity: 5000 scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler
24/04/28 21:20:51 INFO ipc.Server: Starting Socket Reader #1 for port 8032
24/04/28 21:20:51 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationClientProtocolPB to the server
24/04/28 21:20:51 INFO ipc.Server: IPC Server Responder: starting
24/04/28 21:20:51 INFO resourcemanager.ResourceTrackerService: NodeManager from node algo-1(cmPort: 42607 httpPort: 8042) registered with capability: <memory:31784, vCores:8>, assigned nodeId algo-1:42607
```



24/04/28 21:20:51 INFO rmnode.RMNodeImpl: algo-1:42607 Node Transitioned from NEW to RUNNING

24/04/28 21:20:51 INFO ipc.Server: IPC Server listener on 8032: starting

24/04/28 21:20:51 INFO capacity.CapacityScheduler: Added node algo-1:42607 clusterResource: <memory:31784, vCores:8>

24/04/28 21:20:51 INFO security.NMContainerTokenSecretManager: Rolling master-key for container-tokens, got key with id -1024705998

24/04/28 21:20:51 INFO security.NMTokenSecretManagerInNM: Rolling master-key for container-tokens, got key with id -429962112

24/04/28 21:20:51 INFO nodemanager.NodeStatusUpdaterImpl: Registered with ResourceManager as algo-1:42607 with total resource of <memory:31784, vCores:8>

24/04/28 21:20:51 INFO impl.FsDatasetImpl: Time taken to scan block pool BP-393350832-10.0.222.169-1714339246784 on /opt/amazon/hadoop/hdfs/datanode/current: 75ms

24/04/28 21:20:51 INFO impl.FsDatasetImpl: Total time to scan all replicas for block pool BP-393350832-10.0.222.169-1714339246784: 81ms

24/04/28 21:20:51 INFO impl.FsDatasetImpl: Adding replicas to map for block pool BP-393350832-10.0.222.169-1714339246784 on volume /opt/amazon/hadoop/hdfs/datanode/current...

24/04/28 21:20:51 INFO impl.BlockPoolSlice: Replica Cache file: /opt/amazon/hadoop/hdfs/datanode/current/BP-393350832-10.0.222.169-1714339246784/current/replicas does not exist

24/04/28 21:20:51 INFO impl.FsDatasetImpl: Time to add replicas to map for block pool BP-393350832-10.0.222.169-1714339246784 on volume /opt/amazon/hadoop/hdfs/datanode/current: 6ms

24/04/28 21:20:51 INFO impl.FsDatasetImpl: Total time to add all replicas to map for block pool BP-393350832-10.0.222.169-1714339246784: 12ms

24/04/28 21:20:51 INFO datanode.VolumeScanner: Now scanning bpid BP-393350832-10.0.222.169-1714339246784 on volume /opt/amazon/hadoop/hdfs/datanode

24/04/28 21:20:51 INFO datanode.VolumeScanner: VolumeScanner(/opt/amazon/hadoop/hdfs/datanode, DS-7bb77e45-c9ec-428c-ac2f-c4e9031acad5): finished scanning block pool BP-393350832-10.0.222.169-1714339246784

24/04/28 21:20:51 INFO datanode.DirectoryScanner: Periodic Directory Tree Verification scan starting at 4/28/24 10:11 PM with interval of 2160000ms

24/04/28 21:20:51 INFO datanode.DataNode: Block pool BP-393350832-10.0.222.169-1714339246784 (Datanode Uuid 1e229d40-75ae-4892-8b3b-2ec66f1972fb) service to algo-1/10.0.222.169:8020 beginning handshake with NN

24/04/28 21:20:51 INFO datanode.VolumeScanner: VolumeScanner(/opt/amazon/hadoop/hdfs/datanode, DS-7bb77e45-c9ec-428c-ac2f-c4e9031acad5): no suitable block pools found to scan. Waiting 181439969 ms.

24/04/28 21:20:51 INFO resourcemanager.ResourceManager: Transitioned to active state

24/04/28 21:20:51 INFO hdfs.StateChange: BLOCK\* registerDatanode: from DatanodeRegistration(10.0.222.169:50010, datanodeUuid=1e229d40-75ae-4892-8b3b-2ec66f1972fb, infoPort=50075, infoSecurePort=0, ipcPort=50020, storageInfo=lv=-57;cid=CID-fa2135ef-5d1b-4954-be9a-b407f6b3a20a;nsid=861840093;c=1714339246784) storage 1e229d40-75ae-4892-8b3b-2ec66f1972fb

24/04/28 21:20:51 INFO net.NetworkTopology: Adding a new node: /default-rack/10.0.222.169:50010

24/04/28 21:20:51 INFO blockmanagement.BlockReportLeaseManager: Registered DN 1e229d40-75ae-4892-8b3b-2ec66f1972fb (10.0.222.169:50010).

24/04/28 21:20:51 INFO datanode.DataNode: Block pool BP-393350832-10.0.222.169-1714339246784 (Datanode Uuid 1e229d40-75ae-4892-8b3b-2ec66f1972fb) service to algo-1/10.0.222.169:8020 successfully registered with NN

24/04/28 21:20:51 INFO datanode.DataNode: For namenode algo-1/10.0.222.169:8020 using BLOCKREPORT\_INTERVAL of 2160000msec CACHEREPORT\_INTERVAL of 10000msec Initial delay: 0msec; heartbeatInterval=3000

24/04/28 21:20:51 INFO blockmanagement.DatanodeDescriptor: Adding new storage ID DS-

```

7bb77e45-c9ec-428c-ac2f-c4e9031acad5 for DN 10.0.222.169:50010
24/04/28 21:20:51 INFO BlockStateChange: BLOCK* processReport 0x32554a57dc11ce2a: Pr
ocessing first storage report for DS-7bb77e45-c9ec-428c-ac2f-c4e9031acad5 from datan
ode 1e229d40-75ae-4892-8b3b-2ec66f1972fb
24/04/28 21:20:51 INFO BlockStateChange: BLOCK* processReport 0x32554a57dc11ce2a: fr
om storage DS-7bb77e45-c9ec-428c-ac2f-c4e9031acad5 node DatanodeRegistration(10.0.22
2.169:50010, datanodeUuid=1e229d40-75ae-4892-8b3b-2ec66f1972fb, infoPort=50075, info
SecurePort=0, ipcPort=50020, storageInfo=lv=-57;cid=CID-fa2135ef-5d1b-4954-be9a-b407
f6b3a20a;nsid=861840093;c=1714339246784), blocks: 0, hasStaleStorage: false, process
ing time: 2 msecs, invalidatedBlocks: 0
24/04/28 21:20:51 INFO datanode.DataNode: Successfully sent block report 0x32554a57d
c11ce2a, containing 1 storage report(s), of which we sent 1. The reports had 0 tota
l blocks and used 1 RPC(s). This took 6 msec to generate and 49 msecs for RPC and NN
processing. Got back one command: FinalizeCommand/5.
24/04/28 21:20:51 INFO datanode.DataNode: Got finalize command for block pool BP-393
350832-10.0.222.169-1714339246784
WARNING: Running pip install with root privileges is generally not a good idea. Try
`python3 -m pip install --user` instead.
Collecting pydeequ==0.1.5
  Downloading pydeequ-0.1.5-py3-none-any.whl (34 kB)
Installing collected packages: pydeequ
Successfully installed pydeequ-0.1.5
WARNING: Running pip install with root privileges is generally not a good idea. Try
`python3 -m pip install --user` instead.
Collecting pandas==1.1.4
  Downloading pandas-1.1.4-cp37-cp37m-manylinux1_x86_64.whl (9.5 MB)
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.7/si
te-packages (from pandas==1.1.4) (2.8.2)
Requirement already satisfied: numpy>=1.15.4 in /usr/local/lib64/python3.7/site-pack
ages (from pandas==1.1.4) (1.21.4)
Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.7/site-package
s (from pandas==1.1.4) (2021.3)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/site-packages (f
rom python-dateutil>=2.7.3->pandas==1.1.4) (1.16.0)
Installing collected packages: pandas
  Attempting uninstall: pandas
    Found existing installation: pandas 1.3.4
    Uninstalling pandas-1.3.4:
      Successfully uninstalled pandas-1.3.4
ERROR: After October 2020 you may experience errors when installing or updating pack
ages. This is because pip will change the way that it resolves dependency conflicts.
We recommend you use --use-feature=2020-resolver to test your packages with the new
resolver before it becomes the default.
pydeequ 0.1.5 requires pyspark==2.4.7, which is not installed.
Successfully installed pandas-1.1.4
s3a://sagemaker-us-east-1-211125778552/amazon-reviews-pds/tsv/
s3a://sagemaker-us-east-1-211125778552/amazon-reviews-spark-analyzer-2024-04-28-21-1
5-58/output
24/04/28 21:21:00 INFO spark.SparkContext: Running Spark version 2.4.6-amzn-0
24/04/28 21:21:00 INFO spark.SparkContext: Submitted application: PySparkAmazonRevie
wsAnalyzer
24/04/28 21:21:00 INFO spark.SecurityManager: Changing view acls to: root
24/04/28 21:21:00 INFO spark.SecurityManager: Changing modify acls to: root
24/04/28 21:21:00 INFO spark.SecurityManager: Changing view acls groups to:
24/04/28 21:21:00 INFO spark.SecurityManager: Changing modify acls groups to:
24/04/28 21:21:00 INFO spark.SecurityManager: SecurityManager: authentication disabl

```

```
ed; ui acls disabled; users with view permissions: Set(root); groups with view perm
issions: Set(); users with modify permissions: Set(root); groups with modify permis
sions: Set()
24/04/28 21:21:00 INFO util.Utils: Successfully started service 'sparkDriver' on por
t 37411.
24/04/28 21:21:00 INFO spark.SparkEnv: Registering MapOutputTracker
24/04/28 21:21:00 INFO spark.SparkEnv: Registering BlockManagerMaster
24/04/28 21:21:00 INFO storage.BlockManagerMasterEndpoint: Using org.apache.spark.st
orage.DefaultTopologyMapper for getting topology information
24/04/28 21:21:00 INFO storage.BlockManagerMasterEndpoint: BlockManagerMasterEndpoi
nt up
24/04/28 21:21:00 INFO storage.DiskBlockManager: Created local directory at /tmp/blo
ckmgr-2bc25380-7e99-4177-b0f7-7ce7f51d6b7d
24/04/28 21:21:00 INFO memory.MemoryStore: MemoryStore started with capacity 1008.9
MB
24/04/28 21:21:00 INFO spark.SparkEnv: Registering OutputCommitCoordinator
24/04/28 21:21:00 INFO util.log: Logging initialized @8996ms
24/04/28 21:21:01 INFO server.Server: jetty-9.3.z-SNAPSHOT, build timestamp: unknow
n, git hash: unknown
24/04/28 21:21:01 INFO server.Server: Started @9100ms
24/04/28 21:21:01 INFO server.AbstractConnector: Started ServerConnector@589e4ff2{HT
TP/1.1,[http/1.1]}{0.0.0.0:4040}
24/04/28 21:21:01 INFO util.Utils: Successfully started service 'SparkUI' on port 40
40.
24/04/28 21:21:01 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@2103411b{/jobs,null,AVAILABLE,@Spark}
24/04/28 21:21:01 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@70e63973{/jobs/json,null,AVAILABLE,@Spark}
24/04/28 21:21:01 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@3751b89b{/jobs/job,null,AVAILABLE,@Spark}
24/04/28 21:21:01 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@12ab3de{/jobs/job/json,null,AVAILABLE,@Spark}
24/04/28 21:21:01 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@40462f29{/stages,null,AVAILABLE,@Spark}
24/04/28 21:21:01 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@7a69c2da{/stages/json,null,AVAILABLE,@Spark}
24/04/28 21:21:01 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@5432fa35{/stages/stage,null,AVAILABLE,@Spark}
24/04/28 21:21:01 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@45d72c91{/stages/stage/json,null,AVAILABLE,@Spark}
24/04/28 21:21:01 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@6f7954c{/stages/pool,null,AVAILABLE,@Spark}
24/04/28 21:21:01 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@168c99cb{/stages/pool/json,null,AVAILABLE,@Spark}
24/04/28 21:21:01 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@5f8736f{/storage,null,AVAILABLE,@Spark}
24/04/28 21:21:01 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@79b764e5{/storage/json,null,AVAILABLE,@Spark}
24/04/28 21:21:01 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@43d0d36a{/storage/rdd,null,AVAILABLE,@Spark}
24/04/28 21:21:01 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@3160481a{/storage/rdd/json,null,AVAILABLE,@Spark}
24/04/28 21:21:01 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@77c52652{/environment,null,AVAILABLE,@Spark}
24/04/28 21:21:01 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@3517e403{/environment/json,null,AVAILABLE,@Spark}
```

```
24/04/28 21:21:01 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@70193a48{/executors,null,AVAILABLE,@Spark}
24/04/28 21:21:01 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@430a9cc9{/executors/json,null,AVAILABLE,@Spark}
24/04/28 21:21:01 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@1d23c54c{/executors/threadDump,null,AVAILABLE,@Spark}
24/04/28 21:21:01 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@171e0d17{/executors/threadDump/json,null,AVAILABLE,@Spark}
24/04/28 21:21:01 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@2eecbec6{/static,null,AVAILABLE,@Spark}
24/04/28 21:21:01 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@1598ff87{/,null,AVAILABLE,@Spark}
24/04/28 21:21:01 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@2e441a36{/api,null,AVAILABLE,@Spark}
24/04/28 21:21:01 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@138d3d09{/jobs/job/kill,null,AVAILABLE,@Spark}
24/04/28 21:21:01 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@1cc90872{/stages/stage/kill,null,AVAILABLE,@Spark}
24/04/28 21:21:01 INFO ui.SparkUI: Bound SparkUI to 0.0.0.0, and started at http://1
0.0.222.169:4040
24/04/28 21:21:01 INFO client.RMProxy: Connecting to ResourceManager at /10.0.222.16
9:8032
24/04/28 21:21:01 INFO yarn.Client: Requesting a new application from cluster with 1
NodeManagers
24/04/28 21:21:02 INFO resourcemanager.ClientRMService: Allocated new applicationId:
1
24/04/28 21:21:02 INFO conf.Configuration: resource-types.xml not found
24/04/28 21:21:02 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
24/04/28 21:21:02 INFO resource.ResourceUtils: Adding resource type - name = memory-
mb, units = Mi, type = COUNTABLE
24/04/28 21:21:02 INFO resource.ResourceUtils: Adding resource type - name = vcores,
units = , type = COUNTABLE
24/04/28 21:21:02 INFO yarn.Client: Verifying our application has not requested more
than the maximum memory capability of the cluster (31784 MB per container)
24/04/28 21:21:02 INFO yarn.Client: Will allocate AM container, with 896 MB memory i
ncluding 384 MB overhead
24/04/28 21:21:02 INFO yarn.Client: Setting up container launch context for our AM
24/04/28 21:21:02 INFO yarn.Client: Setting up the launch environment for our AM con
tainer
24/04/28 21:21:02 INFO yarn.Client: Preparing resources for our AM container
24/04/28 21:21:02 WARN yarn.Client: Neither spark.yarn.jars nor spark.yarn.archive i
s set, falling back to uploading libraries under SPARK_HOME.
24/04/28 21:21:05 INFO yarn.Client: Uploading resource file:/tmp/spark-ce0bcf76-b803
-4c9c-87b9-a475e13b18e4/___spark_libs___2046587094614679198.zip -> hdfs://10.0.222.16
9/user/root/.sparkStaging/application_1714339251228_0001/___spark_libs___2046587094614
679198.zip
24/04/28 21:21:07 INFO hdfs.StateChange: BLOCK* allocate blk_1073741825_1001, replic
as=10.0.222.169:50010 for /user/root/.sparkStaging/application_1714339251228_0001/___
spark_libs___2046587094614679198.zip
24/04/28 21:21:07 INFO datanode.DataNode: Receiving BP-393350832-10.0.222.169-171433
9246784:blk_1073741825_1001 src: /10.0.222.169:40978 dest: /10.0.222.169:50010
24/04/28 21:21:08 INFO DataNode.clienttrace: src: /10.0.222.169:40978, dest: /10.0.2
22.169:50010, bytes: 134217728, op: HDFS_WRITE, cliID: DFSClient_NONMAPREDUCE_-88721
1942_20, offset: 0, srvID: 1e229d40-75ae-4892-8b3b-2ec66f1972fb, blockid: BP-3933508
32-10.0.222.169-1714339246784:blk_1073741825_1001, duration(ns): 251952505
24/04/28 21:21:08 INFO datanode.DataNode: PacketResponder: BP-393350832-10.0.222.169
```

-1714339246784:blk\_1073741825\_1001, type=LAST\_IN\_PIPELINES terminating  
24/04/28 21:21:08 INFO hdfs.StateChange: BLOCK\* allocate blk\_1073741826\_1002, replicas=10.0.222.169:50010 for /user/root/.sparkStaging/application\_1714339251228\_0001/\_\_\_spark\_libs\_\_2046587094614679198.zip  
24/04/28 21:21:08 INFO datanode.DataNode: Receiving BP-393350832-10.0.222.169-1714339246784:blk\_1073741826\_1002 src: /10.0.222.169:40994 dest: /10.0.222.169:50010  
24/04/28 21:21:08 INFO DataNode.clienttrace: src: /10.0.222.169:40994, dest: /10.0.222.169:50010, bytes: 134217728, op: HDFS\_WRITE, cliID: DFSCClient\_NONMAPREDUCE\_-887211942\_20, offset: 0, srvID: 1e229d40-75ae-4892-8b3b-2ec66f1972fb, blockid: BP-393350832-10.0.222.169-1714339246784:blk\_1073741826\_1002, duration(ns): 214314802  
24/04/28 21:21:08 INFO datanode.DataNode: PacketResponder: BP-393350832-10.0.222.169-1714339246784:blk\_1073741826\_1002, type=LAST\_IN\_PIPELINES terminating  
24/04/28 21:21:08 INFO hdfs.StateChange: BLOCK\* allocate blk\_1073741827\_1003, replicas=10.0.222.169:50010 for /user/root/.sparkStaging/application\_1714339251228\_0001/\_\_\_spark\_libs\_\_2046587094614679198.zip  
24/04/28 21:21:08 INFO datanode.DataNode: Receiving BP-393350832-10.0.222.169-1714339246784:blk\_1073741827\_1003 src: /10.0.222.169:41000 dest: /10.0.222.169:50010  
24/04/28 21:21:08 INFO DataNode.clienttrace: src: /10.0.222.169:41000, dest: /10.0.222.169:50010, bytes: 134217728, op: HDFS\_WRITE, cliID: DFSCClient\_NONMAPREDUCE\_-887211942\_20, offset: 0, srvID: 1e229d40-75ae-4892-8b3b-2ec66f1972fb, blockid: BP-393350832-10.0.222.169-1714339246784:blk\_1073741827\_1003, duration(ns): 173879985  
24/04/28 21:21:08 INFO datanode.DataNode: PacketResponder: BP-393350832-10.0.222.169-1714339246784:blk\_1073741827\_1003, type=LAST\_IN\_PIPELINES terminating  
24/04/28 21:21:08 INFO hdfs.StateChange: BLOCK\* allocate blk\_1073741828\_1004, replicas=10.0.222.169:50010 for /user/root/.sparkStaging/application\_1714339251228\_0001/\_\_\_spark\_libs\_\_2046587094614679198.zip  
24/04/28 21:21:08 INFO datanode.DataNode: Receiving BP-393350832-10.0.222.169-1714339246784:blk\_1073741828\_1004 src: /10.0.222.169:41008 dest: /10.0.222.169:50010  
24/04/28 21:21:08 INFO DataNode.clienttrace: src: /10.0.222.169:41008, dest: /10.0.222.169:50010, bytes: 13532432, op: HDFS\_WRITE, cliID: DFSCClient\_NONMAPREDUCE\_-887211942\_20, offset: 0, srvID: 1e229d40-75ae-4892-8b3b-2ec66f1972fb, blockid: BP-393350832-10.0.222.169-1714339246784:blk\_1073741828\_1004, duration(ns): 16851589  
24/04/28 21:21:08 INFO datanode.DataNode: PacketResponder: BP-393350832-10.0.222.169-1714339246784:blk\_1073741828\_1004, type=LAST\_IN\_PIPELINES terminating  
24/04/28 21:21:08 INFO hdfs.StateChange: DIR\* completeFile: /user/root/.sparkStaging/application\_1714339251228\_0001/\_\_\_spark\_libs\_\_2046587094614679198.zip is closed by DFSCClient\_NONMAPREDUCE\_-887211942\_20  
24/04/28 21:21:08 INFO yarn.Client: Uploading resource file:/opt/ml/processing/input/jars/deequ-1.0.3-rc2.jar -> hdfs://10.0.222.169/user/root/.sparkStaging/application\_1714339251228\_0001/deequ-1.0.3-rc2.jar  
24/04/28 21:21:08 INFO hdfs.StateChange: BLOCK\* allocate blk\_1073741829\_1005, replicas=10.0.222.169:50010 for /user/root/.sparkStaging/application\_1714339251228\_0001/deequ-1.0.3-rc2.jar  
24/04/28 21:21:08 INFO datanode.DataNode: Receiving BP-393350832-10.0.222.169-1714339246784:blk\_1073741829\_1005 src: /10.0.222.169:41018 dest: /10.0.222.169:50010  
24/04/28 21:21:08 INFO DataNode.clienttrace: src: /10.0.222.169:41018, dest: /10.0.222.169:50010, bytes: 1714634, op: HDFS\_WRITE, cliID: DFSCClient\_NONMAPREDUCE\_-887211942\_20, offset: 0, srvID: 1e229d40-75ae-4892-8b3b-2ec66f1972fb, blockid: BP-393350832-10.0.222.169-1714339246784:blk\_1073741829\_1005, duration(ns): 2578805  
24/04/28 21:21:08 INFO datanode.DataNode: PacketResponder: BP-393350832-10.0.222.169-1714339246784:blk\_1073741829\_1005, type=LAST\_IN\_PIPELINES terminating  
24/04/28 21:21:08 INFO hdfs.StateChange: DIR\* completeFile: /user/root/.sparkStaging/application\_1714339251228\_0001/deequ-1.0.3-rc2.jar is closed by DFSCClient\_NONMAPREDUCE\_-887211942\_20  
24/04/28 21:21:08 INFO yarn.Client: Uploading resource file:/usr/lib/spark/python/lib/pyspark.zip -> hdfs://10.0.222.169/user/root/.sparkStaging/application\_17143392512

28\_0001/pyspark.zip  
24/04/28 21:21:08 INFO hdfs.StateChange: BLOCK\* allocate blk\_1073741830\_1006, replic  
as=10.0.222.169:50010 for /user/root/.sparkStaging/application\_1714339251228\_0001/py  
spark.zip  
24/04/28 21:21:08 INFO datanode.DataNode: Receiving BP-393350832-10.0.222.169-171433  
9246784:blk\_1073741830\_1006 src: /10.0.222.169:41020 dest: /10.0.222.169:50010  
24/04/28 21:21:08 INFO DataNode.clienttrace: src: /10.0.222.169:41020, dest: /10.0.2  
22.169:50010, bytes: 596339, op: HDFS\_WRITE, cliID: DFSCClient\_NONMAPREDUCE\_-88721194  
2\_20, offset: 0, srvID: 1e229d40-75ae-4892-8b3b-2ec66f1972fb, blockid: BP-393350832-  
10.0.222.169-1714339246784:blk\_1073741830\_1006, duration(ns): 1439996  
24/04/28 21:21:08 INFO datanode.DataNode: PacketResponder: BP-393350832-10.0.222.169  
-1714339246784:blk\_1073741830\_1006, type=LAST\_IN\_PIPELINES terminating  
24/04/28 21:21:08 INFO namenode.FSNamesystem: BLOCK\* blk\_1073741830\_1006 is COMMITTE  
D but not COMPLETE(numNodes= 0 < minimum = 1) in file /user/root/.sparkStaging/appl  
ication\_1714339251228\_0001/pyspark.zip  
24/04/28 21:21:09 INFO hdfs.StateChange: DIR\* completeFile: /user/root/.sparkStagin  
g/application\_1714339251228\_0001/pyspark.zip is closed by DFSCClient\_NONMAPREDUCE\_-88  
7211942\_20  
24/04/28 21:21:09 INFO yarn.Client: Uploading resource file:/usr/lib/spark/python/li  
b/py4j-0.10.7-src.zip -> hdfs://10.0.222.169/user/root/.sparkStaging/application\_171  
4339251228\_0001/py4j-0.10.7-src.zip  
24/04/28 21:21:09 INFO hdfs.StateChange: BLOCK\* allocate blk\_1073741831\_1007, replic  
as=10.0.222.169:50010 for /user/root/.sparkStaging/application\_1714339251228\_0001/py  
4j-0.10.7-src.zip  
24/04/28 21:21:09 INFO datanode.DataNode: Receiving BP-393350832-10.0.222.169-171433  
9246784:blk\_1073741831\_1007 src: /10.0.222.169:41022 dest: /10.0.222.169:50010  
24/04/28 21:21:09 INFO DataNode.clienttrace: src: /10.0.222.169:41022, dest: /10.0.2  
22.169:50010, bytes: 42437, op: HDFS\_WRITE, cliID: DFSCClient\_NONMAPREDUCE\_-887211942  
\_20, offset: 0, srvID: 1e229d40-75ae-4892-8b3b-2ec66f1972fb, blockid: BP-393350832-1  
0.0.222.169-1714339246784:blk\_1073741831\_1007, duration(ns): 1451759  
24/04/28 21:21:09 INFO datanode.DataNode: PacketResponder: BP-393350832-10.0.222.169  
-1714339246784:blk\_1073741831\_1007, type=LAST\_IN\_PIPELINES terminating  
24/04/28 21:21:09 INFO namenode.FSNamesystem: BLOCK\* blk\_1073741831\_1007 is COMMITTE  
D but not COMPLETE(numNodes= 0 < minimum = 1) in file /user/root/.sparkStaging/appl  
ication\_1714339251228\_0001/py4j-0.10.7-src.zip  
24/04/28 21:21:09 INFO hdfs.StateChange: DIR\* completeFile: /user/root/.sparkStagin  
g/application\_1714339251228\_0001/py4j-0.10.7-src.zip is closed by DFSCClient\_NONMAPRE  
DUCE\_-887211942\_20  
24/04/28 21:21:09 INFO yarn.Client: Uploading resource file:/tmp/spark-ce0bcf76-b803  
-4c9c-87b9-a475e13b18e4/\_\_spark\_conf\_\_4444323185622660411.zip -> hdfs://10.0.222.16  
9/user/root/.sparkStaging/application\_1714339251228\_0001/\_\_spark\_conf\_\_.zip  
24/04/28 21:21:09 INFO hdfs.StateChange: BLOCK\* allocate blk\_1073741832\_1008, replic  
as=10.0.222.169:50010 for /user/root/.sparkStaging/application\_1714339251228\_0001/\_\_  
spark\_conf\_\_.zip  
24/04/28 21:21:09 INFO datanode.DataNode: Receiving BP-393350832-10.0.222.169-171433  
9246784:blk\_1073741832\_1008 src: /10.0.222.169:41038 dest: /10.0.222.169:50010  
24/04/28 21:21:09 INFO DataNode.clienttrace: src: /10.0.222.169:41038, dest: /10.0.2  
22.169:50010, bytes: 245497, op: HDFS\_WRITE, cliID: DFSCClient\_NONMAPREDUCE\_-88721194  
2\_20, offset: 0, srvID: 1e229d40-75ae-4892-8b3b-2ec66f1972fb, blockid: BP-393350832-  
10.0.222.169-1714339246784:blk\_1073741832\_1008, duration(ns): 1448045  
24/04/28 21:21:09 INFO datanode.DataNode: PacketResponder: BP-393350832-10.0.222.169  
-1714339246784:blk\_1073741832\_1008, type=LAST\_IN\_PIPELINES terminating  
24/04/28 21:21:09 INFO namenode.FSNamesystem: BLOCK\* blk\_1073741832\_1008 is COMMITTE  
D but not COMPLETE(numNodes= 0 < minimum = 1) in file /user/root/.sparkStaging/appl  
ication\_1714339251228\_0001/\_\_spark\_conf\_\_.zip  
24/04/28 21:21:09 INFO hdfs.StateChange: DIR\* completeFile: /user/root/.sparkStagin

g/application\_1714339251228\_0001/\_\_\_spark\_conf\_\_\_zip is closed by DFSCClient\_NONMAPRED  
UCE\_-887211942\_20

24/04/28 21:21:09 INFO spark.SecurityManager: Changing view acls to: root  
24/04/28 21:21:09 INFO spark.SecurityManager: Changing modify acls to: root  
24/04/28 21:21:09 INFO spark.SecurityManager: Changing view acls groups to:  
24/04/28 21:21:09 INFO spark.SecurityManager: Changing modify acls groups to:  
24/04/28 21:21:09 INFO spark.SecurityManager: SecurityManager: authentication disabled;  
ui acls disabled; users with view permissions: Set(root); groups with view permissions:  
Set(); users with modify permissions: Set(root); groups with modify permissions: Set()  
24/04/28 21:21:11 INFO yarn.Client: Submitting application application\_1714339251228\_0001  
to ResourceManager  
24/04/28 21:21:11 INFO capacity.CapacityScheduler: Application 'application\_1714339251228\_0001'  
is submitted without priority hence considering default queue/cluster priority: 0  
24/04/28 21:21:11 INFO capacity.CapacityScheduler: Priority '0' is acceptable in queue :  
default for application: application\_1714339251228\_0001  
24/04/28 21:21:11 WARN rmapp.RMAppImpl: The specific max attempts: 0 for application: 1  
is invalid, because it is out of the range [1, 1]. Use the global max attempts instead.  
24/04/28 21:21:11 INFO resourcemanager.ClientRMService: Application with id 1 submitted by  
user root  
24/04/28 21:21:11 INFO rmapp.RMAppImpl: Storing application with id application\_1714339251228\_0001  
24/04/28 21:21:11 INFO resourcemanager.RMAuditLogger: USER=root#011IP=10.0.222.169#011  
OPERATION=Submit Application Request#011TARGET=ClientRMService#011RESULT=SUCCESS#011  
APPID=application\_1714339251228\_0001#011QUEUENAME=default  
24/04/28 21:21:11 INFO recovery.RMStateStore: Storing info for app: application\_1714339251228\_0001  
24/04/28 21:21:11 INFO rmapp.RMAppImpl: application\_1714339251228\_0001 State change from  
NEW to NEW\_SAVING on event = START  
24/04/28 21:21:11 INFO rmapp.RMAppImpl: application\_1714339251228\_0001 State change from  
NEW\_SAVING to SUBMITTED on event = APP\_NEW\_SAVED  
24/04/28 21:21:11 INFO capacity.ParentQueue: Application added - appId: application\_1714339251228\_0001  
user: root leaf-queue of parent: root #applications: 1  
24/04/28 21:21:11 INFO capacity.CapacityScheduler: Accepted application application\_1714339251228\_0001  
from user: root, in queue: default  
24/04/28 21:21:11 INFO rmapp.RMAppImpl: application\_1714339251228\_0001 State change from  
SUBMITTED to ACCEPTED on event = APP\_ACCEPTED  
24/04/28 21:21:11 INFO resourcemanager.ApplicationMasterService: Registering app attempt :  
appattempt\_1714339251228\_0001\_000001  
24/04/28 21:21:11 INFO attempt.RMAppAttemptImpl: appattempt\_1714339251228\_0001\_000001  
State change from NEW to SUBMITTED on event = START  
24/04/28 21:21:11 INFO capacity.LeafQueue: Application application\_1714339251228\_0001  
from user: root activated in queue: default  
24/04/28 21:21:11 INFO capacity.LeafQueue: Application added - appId: application\_1714339251228\_0001  
user: root, leaf-queue: default #user-pending-applications: 0 #user-active-applications: 1  
#queue-pending-applications: 0 #queue-active-applications: 1  
24/04/28 21:21:11 INFO capacity.CapacityScheduler: Added Application Attempt appattempt\_1714339251228\_0001\_000001  
to scheduler from user root in queue default  
24/04/28 21:21:11 INFO attempt.RMAppAttemptImpl: appattempt\_1714339251228\_0001\_000001  
State change from SUBMITTED to SCHEDULED on event = ATTEMPT\_ADDED  
24/04/28 21:21:11 INFO impl.YarnClientImpl: Submitted application application\_1714339251228\_0001  
24/04/28 21:21:11 INFO cluster.SchedulerExtensionServices: Starting Yarn extension services  
with app application\_1714339251228\_0001 and attemptId None

24/04/28 21:21:11 INFO allocator.AbstractContainerAllocator: assignedContainer application attempt=appattempt\_1714339251228\_0001\_000001 container=null queue=default clusterResource=<memory:31784, vCores:8> type=OFF\_SWITCH requestedPartition=

24/04/28 21:21:11 INFO capacity.ParentQueue: assignedContainer queue=root usedCapacity=0.0 absoluteUsedCapacity=0.0 used=<memory:0, vCores:0> cluster=<memory:31784, vCores:8>

24/04/28 21:21:11 INFO rmcontainer.RMContainerImpl: container\_1714339251228\_0001\_01\_000001 Container Transitioned from NEW to ALLOCATED

24/04/28 21:21:11 INFO resourcemanager.RMAuditLogger: USER=root#011OPERATION=AM Allocated Container#011TARGET=SchedulerApp#011RESULT=SUCCESS#011APPID=application\_1714339251228\_0001#011CONTAINERID=container\_1714339251228\_0001\_01\_000001#011RESOURCE=<memory:896, vCores:1>#011QUEUENAME=default

24/04/28 21:21:11 INFO security.NMTokenSecretManagerInRM: Sending NMToken for nodeId : algo-1:42607 for container : container\_1714339251228\_0001\_01\_000001

24/04/28 21:21:11 INFO rmcontainer.RMContainerImpl: container\_1714339251228\_0001\_01\_000001 Container Transitioned from ALLOCATED to ACQUIRED

24/04/28 21:21:11 INFO security.NMTokenSecretManagerInRM: Clear node set for appattempt\_1714339251228\_0001\_000001

24/04/28 21:21:11 INFO capacity.ParentQueue: assignedContainer queue=root usedCapacity=0.028190285 absoluteUsedCapacity=0.028190285 used=<memory:896, vCores:1> cluster=<memory:31784, vCores:8>

24/04/28 21:21:11 INFO capacity.CapacityScheduler: Allocation proposal accepted

24/04/28 21:21:11 INFO attempt.RMAppAttemptImpl: Storing attempt: AppId: application\_1714339251228\_0001 AttemptId: appattempt\_1714339251228\_0001\_000001 MasterContainer: Container: [ContainerId: container\_1714339251228\_0001\_01\_000001, AllocationRequestId: 0, Version: 0, NodeId: algo-1:42607, NodeHttpAddress: algo-1:8042, Resource: <memory:896, vCores:1>, Priority: 0, Token: Token { kind: ContainerToken, service: 10.0.222.169:42607 }, ExecutionType: GUARANTEED, ]

24/04/28 21:21:11 INFO attempt.RMAppAttemptImpl: appattempt\_1714339251228\_0001\_000001 State change from SCHEDULED to ALLOCATED\_SAVING on event = CONTAINER\_ALLOCATED

24/04/28 21:21:11 INFO attempt.RMAppAttemptImpl: appattempt\_1714339251228\_0001\_000001 State change from ALLOCATED\_SAVING to ALLOCATED on event = ATTEMPT\_NEW\_SAVED

24/04/28 21:21:11 INFO amlauncher.AMLauncher: Launching masterappattempt\_1714339251228\_0001\_000001

24/04/28 21:21:11 INFO amlauncher.AMLauncher: Setting up container Container: [ContainerId: container\_1714339251228\_0001\_01\_000001, AllocationRequestId: 0, Version: 0, NodeId: algo-1:42607, NodeHttpAddress: algo-1:8042, Resource: <memory:896, vCores:1>, Priority: 0, Token: Token { kind: ContainerToken, service: 10.0.222.169:42607 }, ExecutionType: GUARANTEED, ] for AM appattempt\_1714339251228\_0001\_000001

24/04/28 21:21:11 INFO security.AMRMTokenSecretManager: Create AMRMToken for ApplicationAttempt: appattempt\_1714339251228\_0001\_000001

24/04/28 21:21:11 INFO security.AMRMTokenSecretManager: Creating password for appattempt\_1714339251228\_0001\_000001

24/04/28 21:21:12 INFO ipc.Server: Auth successful for appattempt\_1714339251228\_0001\_000001 (auth:SIMPLE)

24/04/28 21:21:12 INFO containermanager.ContainerManagerImpl: Start request for container\_1714339251228\_0001\_01\_000001 by user root

24/04/28 21:21:12 INFO containermanager.ContainerManagerImpl: Creating a new application reference for app application\_1714339251228\_0001

24/04/28 21:21:12 INFO application.ApplicationImpl: Application application\_1714339251228\_0001 transitioned from NEW to INITING

24/04/28 21:21:12 INFO application.ApplicationImpl: Adding container\_1714339251228\_0001\_01\_000001 to application application\_1714339251228\_0001

24/04/28 21:21:12 INFO nodemanager.NMAuditLogger: USER=root#011IP=10.0.222.169#011OPERATION=Start Container Request#011TARGET=ContainerManagerImpl#011RESULT=SUCCESS#011APPID=application\_1714339251228\_0001#011CONTAINERID=container\_1714339251228\_0001\_01\_0



```
00001
24/04/28 21:21:12 INFO application.ApplicationImpl: Application application_17143392
51228_0001 transitioned from INITING to RUNNING
24/04/28 21:21:12 INFO container.ContainerImpl: Container container_1714339251228_00
01_01_000001 transitioned from NEW to LOCALIZING
24/04/28 21:21:12 INFO containermanager.AuxServices: Got event CONTAINER_INIT for ap
pId application_1714339251228_0001
24/04/28 21:21:12 INFO amlauncher.AMLauncher: Done launching container Container: [C
ontainerId: container_1714339251228_0001_01_000001, AllocationRequestId: 0, Version:
0, NodeId: algo-1:42607, NodeHttpAddress: algo-1:8042, Resource: <memory:896, vCore
s:1>, Priority: 0, Token: Token { kind: ContainerToken, service: 10.0.222.169:42607
}, ExecutionType: GUARANTEED, ] for AM appattempt_1714339251228_0001_000001
24/04/28 21:21:12 INFO attempt.RMAppAttemptImpl: appattempt_1714339251228_0001_00000
1 State change from ALLOCATED to LAUNCHED on event = LAUNCHED
24/04/28 21:21:12 INFO rmapp.RMAppImpl: update the launch time for applicationId: ap
plication_1714339251228_0001, attemptId: appattempt_1714339251228_0001_000001launchT
ime: 1714339272256
24/04/28 21:21:12 INFO recovery.RMStateStore: Updating info for app: application_171
4339251228_0001
24/04/28 21:21:12 INFO localizer.ResourceLocalizationService: Created localizer for
container_1714339251228_0001_01_000001
24/04/28 21:21:12 INFO localizer.ResourceLocalizationService: Writing credentials to
the nmPrivate file /tmp/hadoop-root/nm-local-dir/nmPrivate/container_1714339251228_0
001_01_000001.tokens
24/04/28 21:21:12 INFO nodemanager.DefaultContainerExecutor: Initializing user root
24/04/28 21:21:12 INFO nodemanager.DefaultContainerExecutor: Copying from /tmp/hadoo
p-root/nm-local-dir/nmPrivate/container_1714339251228_0001_01_000001.tokens to /tmp/
hadoop-root/nm-local-dir/usercache/root/appcache/application_1714339251228_0001/cont
ainer_1714339251228_0001_01_000001.tokens
24/04/28 21:21:12 INFO nodemanager.DefaultContainerExecutor: Localizer CWD set to /t
mp/hadoop-root/nm-local-dir/usercache/root/appcache/application_1714339251228_0001 =
file:/tmp/hadoop-root/nm-local-dir/usercache/root/appcache/application_1714339251228
_0001
24/04/28 21:21:12 INFO localizer.ContainerLocalizer: Disk Validator: yarn.nodemange
r.disk-validator is loaded.
24/04/28 21:21:12 INFO yarn.Client: Application report for application_1714339251228
_0001 (state: ACCEPTED)
24/04/28 21:21:12 INFO yarn.Client:
#011 client token: N/A
#011 diagnostics: AM container is launched, waiting for AM container to Register wit
h RM
#011 ApplicationMaster host: N/A
#011 ApplicationMaster RPC port: -1
#011 queue: default
#011 start time: 1714339271340
#011 final status: UNDEFINED
#011 tracking URL: http://algo-1:8088/proxy/application_1714339251228_0001/
#011 user: root
24/04/28 21:21:12 INFO rmcontainer.RMContainerImpl: container_1714339251228_0001_01_
000001 Container Transitioned from ACQUIRED to RUNNING
24/04/28 21:21:13 INFO yarn.Client: Application report for application_1714339251228
_0001 (state: ACCEPTED)
24/04/28 21:21:14 INFO container.ContainerImpl: Container container_1714339251228_00
01_01_000001 transitioned from LOCALIZING to SCHEDULED
24/04/28 21:21:14 INFO scheduler.ContainerScheduler: Starting container [container_1
714339251228_0001_01_000001]
```

```
24/04/28 21:21:14 INFO container.ContainerImpl: Container container_1714339251228_0001_01_000001 transitioned from SCHEDULED to RUNNING
24/04/28 21:21:14 INFO monitor.ContainersMonitorImpl: Starting resource-monitoring for container_1714339251228_0001_01_000001
24/04/28 21:21:14 INFO nodemanager.DefaultContainerExecutor: launchContainer: [bash, /tmp/hadoop-root/nm-local-dir/usercache/root/appcache/application_1714339251228_0001/container_1714339251228_0001_01_000001/default_container_executor.sh]
Handling create event for file: /var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/prelaunch.out
Handling create event for file: /var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/prelaunch.err
Handling create event for file: /var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stdout
Handling create event for file: /var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr
24/04/28 21:21:14 INFO yarn.Client: Application report for application_1714339251228_0001 (state: ACCEPTED)
24/04/28 21:21:15 INFO yarn.Client: Application report for application_1714339251228_0001 (state: ACCEPTED)
24/04/28 21:21:16 INFO yarn.Client: Application report for application_1714339251228_0001 (state: ACCEPTED)
24/04/28 21:21:16 INFO monitor.ContainersMonitorImpl: container_1714339251228_0001_01_000001's ip = 10.0.222.169, and hostname = algo-1
24/04/28 21:21:16 INFO monitor.ContainersMonitorImpl: Skipping monitoring container container_1714339251228_0001_01_000001 since CPU usage is not yet available.
24/04/28 21:21:17 INFO ipc.Server: Auth successful for appattempt_1714339251228_0001_000001 (auth:SIMPLE)
24/04/28 21:21:17 INFO resourcemanager.DefaultAMSPProcessor: AM registration appattempt_1714339251228_0001_000001
24/04/28 21:21:17 INFO resourcemanager.RMAuditLogger: USER=root#011IP=10.0.222.169#011OPERATION=Register App Master#011TARGET=ApplicationMasterService#011RESULT=SUCCESS#011APPID=application_1714339251228_0001#011APPATTEMPTID=appattempt_1714339251228_0001_000001
24/04/28 21:21:17 INFO attempt.RMAppAttemptImpl: appattempt_1714339251228_0001_000001 State change from LAUNCHED to RUNNING on event = REGISTERED
24/04/28 21:21:17 INFO rmapp.RMAppImpl: application_1714339251228_0001 State change from ACCEPTED to RUNNING on event = ATTEMPT_REGISTERED
24/04/28 21:21:17 INFO yarn.Client: Application report for application_1714339251228_0001 (state: RUNNING)
24/04/28 21:21:17 INFO yarn.Client:
#011 client token: N/A
#011 diagnostics: N/A
#011 ApplicationMaster host: 10.0.222.169
#011 ApplicationMaster RPC port: -1
#011 queue: default
#011 start time: 1714339271340
#011 final status: UNDEFINED
#011 tracking URL: http://algo-1:8088/proxy/application_1714339251228_0001/
#011 user: root
24/04/28 21:21:17 INFO cluster.YarnClientSchedulerBackend: Application application_1714339251228_0001 has started running.
24/04/28 21:21:17 INFO util.Utills: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 36237.
24/04/28 21:21:17 INFO netty.NettyBlockTransferService: Server created on 10.0.222.169:36237
24/04/28 21:21:17 INFO storage.BlockManager: Using org.apache.spark.storage.RandomBl
```

```
ockReplicationPolicy for block replication policy
24/04/28 21:21:17 INFO storage.BlockManagerMaster: Registering BlockManager BlockMan
agerId(driver, 10.0.222.169, 36237, None)
24/04/28 21:21:17 INFO storage.BlockManagerMasterEndpoint: Registering block manager
10.0.222.169:36237 with 1008.9 MB RAM, BlockManagerId(driver, 10.0.222.169, 36237, N
one)
24/04/28 21:21:17 INFO storage.BlockManagerMaster: Registered BlockManager BlockMana
gerId(driver, 10.0.222.169, 36237, None)
24/04/28 21:21:17 INFO storage.BlockManager: Initialized BlockManager: BlockManagerI
d(driver, 10.0.222.169, 36237, None)
24/04/28 21:21:17 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@58243bb5{/metrics/json,null,AVAILABLE,@Spark}
24/04/28 21:21:17 INFO cluster.YarnClientSchedulerBackend: SchedulerBackend is ready
for scheduling beginning after reached minRegisteredResourcesRatio: 0.0
24/04/28 21:21:17 INFO cluster.YarnClientSchedulerBackend: Add WebUI Filter. org.apa
che.hadoop.yarn.server.webproxy.amfilter.AmIpFilter, Map(PROXY_HOSTS -> algo-1, PROX
Y_URI_BASES -> http://algo-1:8088/proxy/application_1714339251228_0001), /proxy/appl
ication_1714339251228_0001
24/04/28 21:21:17 INFO internal.SharedState: Setting hive.metastore.warehouse.dir
('null') to the value of spark.sql.warehouse.dir ('file:/usr/lib/spark/spark-warehou
se').
24/04/28 21:21:17 INFO internal.SharedState: Warehouse path is 'file:/usr/lib/spark/
spark-warehouse'.
24/04/28 21:21:17 INFO ui.JettyUtils: Adding filter org.apache.hadoop.yarn.server.we
bproxy.amfilter.AmIpFilter to /SQL.
24/04/28 21:21:17 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@654d6f69{/SQL,null,AVAILABLE,@Spark}
24/04/28 21:21:17 INFO ui.JettyUtils: Adding filter org.apache.hadoop.yarn.server.we
bproxy.amfilter.AmIpFilter to /SQL/json.
24/04/28 21:21:17 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@c4973d6{/SQL/json,null,AVAILABLE,@Spark}
24/04/28 21:21:17 INFO ui.JettyUtils: Adding filter org.apache.hadoop.yarn.server.we
bproxy.amfilter.AmIpFilter to /SQL/execution.
24/04/28 21:21:17 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@6b97b3b4{/SQL/execution,null,AVAILABLE,@Spark}
24/04/28 21:21:17 INFO ui.JettyUtils: Adding filter org.apache.hadoop.yarn.server.we
bproxy.amfilter.AmIpFilter to /SQL/execution/json.
24/04/28 21:21:17 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@53568261{/SQL/execution/json,null,AVAILABLE,@Spark}
24/04/28 21:21:17 INFO ui.JettyUtils: Adding filter org.apache.hadoop.yarn.server.we
bproxy.amfilter.AmIpFilter to /static/sql.
24/04/28 21:21:17 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler
@1968531f{/static/sql,null,AVAILABLE,@Spark}
24/04/28 21:21:18 INFO cluster.YarnSchedulerBackend$YarnSchedulerEndpoint: Applicati
onMaster registered as NettyRpcEndpointRef(spark-client://YarnAM)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000001/stderr] 24/04/28 21:21:15 INFO util.SignalUtils: Registered signal handler
for TERM
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000001/stderr] 24/04/28 21:21:15 INFO util.SignalUtils: Registered signal handler
for HUP
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000001/stderr] 24/04/28 21:21:15 INFO util.SignalUtils: Registered signal handler
for INT
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000001/stderr] 24/04/28 21:21:15 INFO spark.SecurityManager: Changing view acls t
```

```

o: root
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] 24/04/28 21:21:15 INFO spark.SecurityManager: Changing modify acls to: root
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] 24/04/28 21:21:15 INFO spark.SecurityManager: Changing view acls groups to:
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] 24/04/28 21:21:15 INFO spark.SecurityManager: Changing modify acls groups to:
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] 24/04/28 21:21:15 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(root); groups with view permissions: Set(); users with modify permissions: Set(root); groups with modify permissions: Set()
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] 24/04/28 21:21:16 INFO yarn.ApplicationMaster: Preparing Local resources
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] 24/04/28 21:21:17 INFO yarn.ApplicationMaster: ApplicationAttemptId: appattempt_1714339251228_0001_000001
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] 24/04/28 21:21:17 INFO client.RMProxy: Connecting to ResourceManager at /10.0.222.169:8030
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] 24/04/28 21:21:17 INFO yarn.YarnRMClient: Registering the ApplicationMaster
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] 24/04/28 21:21:17 INFO client.TransportClientFactory: Successfully created connection to /10.0.222.169:37411 after 108 ms (0 ms spent in bootstraps)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] 24/04/28 21:21:17 INFO yarn.ApplicationMaster:
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] =====
=====
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] YARN executor launch context:
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] env:
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] CLASSPATH -> /usr/lib/hadoop-lzo/lib/*:/usr/lib/hadoop/hadoop-aws.jar:/usr/share/aws/aws-java-sdk/*:/usr/share/aws/emr/emrfs/conf:/usr/share/aws/emr/emrfs/lib/*:/usr/share/aws/emr/emrfs/auxlib/*:/usr/share/aws/emr/goodies/lib/emr-spark-goodies.jar:/usr/share/aws/emr/security/conf:/usr/share/aws/emr/security/lib/*:/usr/share/aws/hmclient/lib/aws-glue-datacatalog-spark-client.jar:/usr/share/java/Hive-JSON-Serde/hive-openx-serde.jar:/usr/share/aws/sagemaker-spark-sdk/lib/sagemaker-spark-sdk.jar:/usr/share/aws/emr/s3select/lib/emr-s3-select-spark-connector.jar<CPS>{{PWD}}<CPS>{{PWD}}/__spark_conf__<CPS>{{PWD}}/__spark_libs__/*<CPS>$HADOOP_CONF_DIR<CPS>$HADOOP_COMMON_HOME/share/hadoop/common/*<CPS>$HADOOP_COMMON_HOME/share/hadoop/common/lib/*<CPS>$HADOOP_HDFS_HOME/share/hadoop/hdfs/*<CPS>$HADOOP_HDFS_HOME/share/hadoop/hdfs/lib/*<CPS>$HADOOP_YARN_HOME/share/hadoop/yarn/*<CPS>$HADOOP_YARN_HOME/share/hadoop/yarn/lib/*<CPS>$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/*<CPS>$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/lib/*<CPS>{{PWD}}/__spark_conf__/_hadoop_conf__
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] SPARK_YARN_STAGING_DIR -> hdfs://10.0.222.169/user/root/.sparkStaging/application_1714339251228_0001

```

```

[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] SPARK_NO_DAEMONIZE -> TRUE
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] SPARK_USER -> root
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] SPARK_MASTER_HOST -> 10.0.222.169
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] SPARK_HOME -> /usr/lib/spark
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] PYTHONPATH -> {{PWD}}/pyspark.zip<CPS>{{PWD}}/py4j-0.10.7-src.zip
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr]
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] command:
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] LD_LIBRARY_PATH="/usr/lib/hadoop/lib/native:/usr/lib/hadoop-lzo/lib/native:$LD_LIBRARY_PATH" \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] {{JAVA_HOME}}/bin/java \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] -server \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] -Xmx26847m \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] '-verbose:gc' \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] '-XX:OnOutOfMemoryError=kill -9 %p' \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] '-XX:+PrintGCDetails' \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] '-XX:+PrintGCDateStamps' \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] '-XX:+UseParallelGC' \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] '-XX:InitiatingHeapOccupancyPercent=70' \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] '-XX:ConcGCThreads=2' \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] '-XX:ParallelGCThreads=6' \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] -Djava.io.tmpdir={{PWD}}/tmp \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] '-Dspark.driver.port=37411' \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] -Dspark.yarn.app.container.log.dir=<LOG_DIR> \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] org.apache.spark.executor.CoarseGrainedExecutorBackend \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] --driver-url \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] spark://CoarseGrainedScheduler@10.0.222.169:37411 \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] --executor-id \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] <executorId> \

```

```

[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr]      --hostname \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr]      <hostname> \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr]      --cores \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr]      8 \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr]      --app-id \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr]      application_1714339251228_0001 \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr]      --user-class-path \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr]      file:$PWD/___app___jar \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr]      --user-class-path \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr]      file:$PWD/deequ-1.0.3-rc2.jar \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr]      1><LOG_DIR>/stdout \
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr]      2><LOG_DIR>/stderr
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr]
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr]      resources:
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr]      __spark_conf__ -> resource { scheme: "hdfs" host: "10.0.222.16
9" port: -1 file: "/user/root/.sparkStaging/application_1714339251228_0001/___spark_c
onf__.zip" } size: 245497 timestamp: 1714339269971 type: ARCHIVE visibility: PRIVATE
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr]      pyspark.zip -> resource { scheme: "hdfs" host: "10.0.222.169"
port: -1 file: "/user/root/.sparkStaging/application_1714339251228_0001/pyspark.zip"
} size: 596339 timestamp: 1714339269007 type: FILE visibility: PRIVATE
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr]      deequ-1.0.3-rc2.jar -> resource { scheme: "hdfs" host: "10.0.2
22.169" port: -1 file: "/user/root/.sparkStaging/application_1714339251228_0001/deeq
u-1.0.3-rc2.jar" } size: 1714634 timestamp: 1714339268586 type: FILE visibility: PRI
VATE
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr]      __spark_libs__ -> resource { scheme: "hdfs" host: "10.0.222.16
9" port: -1 file: "/user/root/.sparkStaging/application_1714339251228_0001/___spark_l
ibs__2046587094614679198.zip" } size: 416185616 timestamp: 1714339268476 type: ARCHI
VE visibility: PRIVATE
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr]      py4j-0.10.7-src.zip -> resource { scheme: "hdfs" host: "10.0.2
22.169" port: -1 file: "/user/root/.sparkStaging/application_1714339251228_0001/py4j
-0.10.7-src.zip" } size: 42437 timestamp: 1714339269427 type: FILE visibility: PRIVA
TE
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr]
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] =====
=====

```

```
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000001/stderr] 24/04/28 21:21:17 INFO conf.Configuration: resource-types.xml not found
24/04/28 21:21:18 INFO state.StateStoreCoordinatorRef: Registered StateStoreCoordinator endpoint
24/04/28 21:21:18 INFO allocator.AbstractContainerAllocator: assignedContainer application attempt=appattempt_1714339251228_0001_000001 container=null queue=default clusterResource=<memory:31784, vCores:8> type=OFF_SWITCH requestedPartition=
24/04/28 21:21:18 INFO capacity.ParentQueue: assignedContainer queue=root usedCapacity=0.028190285 absoluteUsedCapacity=0.028190285 used=<memory:896, vCores:1> cluster=<memory:31784, vCores:8>
24/04/28 21:21:18 INFO rmcontainer.RMContainerImpl: container_1714339251228_0001_01_000002 Container Transitioned from NEW to ALLOCATED
24/04/28 21:21:18 INFO resourcemanager.RMAuditLogger: USER=root#011OPERATION=AM Allocated Container#011TARGET=SchedulerApp#011RESULT=SUCCESS#011APPID=application_1714339251228_0001#011CONTAINERID=container_1714339251228_0001_01_000002#011RESOURCE=<memory:29531, vCores:1>#011QUEUENAME=default
24/04/28 21:21:18 INFO capacity.ParentQueue: assignedContainer queue=root usedCapacity=0.95730555 absoluteUsedCapacity=0.95730555 used=<memory:30427, vCores:2> cluster=<memory:31784, vCores:8>
24/04/28 21:21:18 INFO capacity.CapacityScheduler: Allocation proposal accepted
24/04/28 21:21:19 INFO security.NMTokenSecretManagerInRM: Sending NMToken for nodeId : algo-1:42607 for container : container_1714339251228_0001_01_000002
24/04/28 21:21:19 INFO rmcontainer.RMContainerImpl: container_1714339251228_0001_01_000002 Container Transitioned from ALLOCATED to ACQUIRED
24/04/28 21:21:19 INFO ipc.Server: Auth successful for appattempt_1714339251228_0001_000001 (auth:SIMPLE)
24/04/28 21:21:19 INFO containermanager.ContainerManagerImpl: Start request for container_1714339251228_0001_01_000002 by user root
24/04/28 21:21:19 INFO nodemanager.NMAuditLogger: USER=root#011IP=10.0.222.169#011OPERATION=Start Container Request#011TARGET=ContainerManagerImpl#011RESULT=SUCCESS#011APPID=application_1714339251228_0001#011CONTAINERID=container_1714339251228_0001_01_000002
24/04/28 21:21:19 INFO application.ApplicationImpl: Adding container_1714339251228_0001_01_000002 to application application_1714339251228_0001
24/04/28 21:21:19 INFO container.ContainerImpl: Container container_1714339251228_0001_01_000002 transitioned from NEW to LOCALIZING
24/04/28 21:21:19 INFO containermanager.AuxServices: Got event CONTAINER_INIT for appId application_1714339251228_0001
24/04/28 21:21:19 INFO container.ContainerImpl: Container container_1714339251228_0001_01_000002 transitioned from LOCALIZING to SCHEDULED
24/04/28 21:21:19 INFO scheduler.ContainerScheduler: Starting container [container_1714339251228_0001_01_000002]
24/04/28 21:21:19 INFO container.ContainerImpl: Container container_1714339251228_0001_01_000002 transitioned from SCHEDULED to RUNNING
24/04/28 21:21:19 INFO monitor.ContainersMonitorImpl: Starting resource-monitoring for container_1714339251228_0001_01_000002
24/04/28 21:21:19 INFO nodemanager.DefaultContainerExecutor: launchContainer: [bash, /tmp/hadoop-root/nm-local-dir/usercache/root/appcache/application_1714339251228_0001/container_1714339251228_0001_01_000002/default_container_executor.sh]
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/prelaunch.out
Handling create event for file: /var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/prelaunch.out
Handling create event for file: /var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/prelaunch.err
Handling create event for file: /var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/prelaunch.err]
```

```
1/container_1714339251228_0001_01_000002/stdout
Handling create event for file: /var/log/yarn/userlogs/application_1714339251228_000
1/container_1714339251228_0001_01_000002/stderr
24/04/28 21:21:19 INFO monitor.ContainersMonitorImpl: container_1714339251228_0001_0
1_000002's ip = 10.0.222.169, and hostname = algo-1
24/04/28 21:21:19 INFO monitor.ContainersMonitorImpl: Skipping monitoring container
container_1714339251228_0001_01_000002 since CPU usage is not yet available.
24/04/28 21:21:19 INFO Configuration.deprecation: fs.s3a.server-side-encryption-key
is deprecated. Instead, use fs.s3a.server-side-encryption.key
24/04/28 21:21:19 INFO rmcontainer.RMContainerImpl: container_1714339251228_0001_01_
000002 Container Transitioned from ACQUIRED to RUNNING
24/04/28 21:21:20 INFO datasources.InMemoryFileIndex: It took 126 ms to list leaf fi
les for 1 paths.
24/04/28 21:21:21 INFO datasources.FileSourceStrategy: Pruning directories with:
24/04/28 21:21:21 INFO datasources.FileSourceStrategy: Post-Scan Filters:
24/04/28 21:21:21 INFO datasources.FileSourceStrategy: Output Data Schema: struct<re
view_id: string, star_rating: int, helpful_votes: int, total_votes: int ... 2 more f
ields>
24/04/28 21:21:21 INFO execution.FileSourceScanExec: Pushed Filters:
24/04/28 21:21:21 INFO execution.FileSourceScanExec: Pushed Filters:
24/04/28 21:21:21 WARN util.Utills: Truncated the string representation of a plan sin
ce it was too large. This behavior can be adjusted by setting 'spark.debug.maxToStri
ngFields' in SparkEnv.conf.
24/04/28 21:21:22 INFO scheduler.AppSchedulingInfo: checking for deactivate of appli
cation :application_1714339251228_0001
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:20 INFO executor.CoarseGrainedExecutorBackend: Star
ted daemon with process name: 1136@algo-1
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:20 INFO util.SignalUtils: Registered signal handler
for TERM
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:20 INFO util.SignalUtils: Registered signal handler
for HUP
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:20 INFO util.SignalUtils: Registered signal handler
for INT
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:20 INFO spark.SecurityManager: Changing view acs to:
root
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:20 INFO spark.SecurityManager: Changing modify acs
to: root
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:20 INFO spark.SecurityManager: Changing view acs g
roups to:
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:20 INFO spark.SecurityManager: Changing modify acs
groups to:
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:20 INFO spark.SecurityManager: SecurityManager: aut
hentication disabled; ui acs disabled; users with view permissions: Set(root); gro
ups with view permissions: Set(); users with modify permissions: Set(root); groups
with modify permissions: Set()
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:21 INFO client.TransportClientFactory: Successfully
```



```
created connection to /10.0.222.169:37411 after 129 ms (0 ms spent in bootstraps)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:21 INFO spark.SecurityManager: Changing view acls to: root
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:21 INFO spark.SecurityManager: Changing modify acls to: root
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:21 INFO spark.SecurityManager: Changing view acls groups to:
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:21 INFO spark.SecurityManager: Changing modify acls groups to:
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:21 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(root); groups with view permissions: Set(); users with modify permissions: Set(root); groups with modify permissions: Set()
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:21 INFO client.TransportClientFactory: Successfully created connection to /10.0.222.169:37411 after 2 ms (0 ms spent in bootstraps)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:21 INFO storage.DiskBlockManager: Created local directory at /tmp/hadoop-root/nm-local-dir/usercache/root/appcache/application_1714339251228_0001/blockmgr-02b6f9c4-4534-4a1a-81b5-a8d5ccdb3b5c
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:21 INFO memory.MemoryStore: MemoryStore started with capacity 13.8 GB
24/04/28 21:21:22 INFO cluster.YarnSchedulerBackend$YarnDriverEndpoint: Registered executor NettyRpcEndpointRef(spark-client://Executor) (10.0.222.169:41176) with ID 1
24/04/28 21:21:22 INFO codegen.CodeGenerator: Code generated in 349.732377 ms
24/04/28 21:21:22 INFO storage.BlockManagerMasterEndpoint: Registering block manager algo-1:40893 with 13.8 GB RAM, BlockManagerId(1, algo-1, 40893, None)
24/04/28 21:21:22 INFO codegen.CodeGenerator: Code generated in 16.701036 ms
24/04/28 21:21:22 INFO memory.MemoryStore: Block broadcast_0 stored as values in memory (estimated size 303.8 KB, free 1008.6 MB)
24/04/28 21:21:22 INFO memory.MemoryStore: Block broadcast_0_piece0 stored as bytes in memory (estimated size 27.6 KB, free 1008.6 MB)
24/04/28 21:21:22 INFO storage.BlockManagerInfo: Added broadcast_0_piece0 in memory on 10.0.222.169:36237 (size: 27.6 KB, free: 1008.9 MB)
24/04/28 21:21:22 INFO spark.SparkContext: Created broadcast 0 from collect at AnalysisRunner.scala:323
24/04/28 21:21:23 INFO execution.FileSourceScanExec: Planning scan with bin packing, max size: 4447362 bytes, open cost is considered as scanning 4194304 bytes, number of split files: 3, prefetch: false
24/04/28 21:21:23 INFO execution.FileSourceScanExec: relation: None, fileSplitsInPartitionHistogram: ArrayBuffer((1 fileSplits,3))
24/04/28 21:21:23 INFO scheduler.DAGScheduler: Registering RDD 3 (collect at AnalysisRunner.scala:323) as input to shuffle 0
24/04/28 21:21:23 INFO scheduler.DAGScheduler: Got map stage job 0 (collect at AnalysisRunner.scala:323) with 3 output partitions
24/04/28 21:21:23 INFO scheduler.DAGScheduler: Final stage: ShuffleMapStage 0 (collect at AnalysisRunner.scala:323)
24/04/28 21:21:23 INFO scheduler.DAGScheduler: Parents of final stage: List()
24/04/28 21:21:23 INFO scheduler.DAGScheduler: Missing parents: List()
24/04/28 21:21:23 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 0 (MapPart
```

```

itionsRDD[3] at collect at AnalysisRunner.scala:323), which has no missing parents
24/04/28 21:21:23 INFO memory.MemoryStore: Block broadcast_1 stored as values in mem
ory (estimated size 32.9 KB, free 1008.5 MB)
24/04/28 21:21:23 INFO memory.MemoryStore: Block broadcast_1_piece0 stored as bytes
in memory (estimated size 14.2 KB, free 1008.5 MB)
24/04/28 21:21:23 INFO storage.BlockManagerInfo: Added broadcast_1_piece0 in memory
on 10.0.222.169:36237 (size: 14.2 KB, free: 1008.9 MB)
24/04/28 21:21:23 INFO spark.SparkContext: Created broadcast 1 from broadcast at DAG
Scheduler.scala:1203
24/04/28 21:21:23 INFO scheduler.DAGScheduler: Submitting 3 missing tasks from Shuff
leMapStage 0 (MapPartitionsRDD[3] at collect at AnalysisRunner.scala:323) (first 15
tasks are for partitions Vector(0, 1, 2))
24/04/28 21:21:23 INFO cluster.YarnScheduler: Adding task set 0.0 with 3 tasks
24/04/28 21:21:23 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 0.0 (TID
0, algo-1, executor 1, partition 0, PROCESS_LOCAL, 8328 bytes)
24/04/28 21:21:23 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 0.0 (TID
1, algo-1, executor 1, partition 1, PROCESS_LOCAL, 8325 bytes)
24/04/28 21:21:23 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 0.0 (TID
2, algo-1, executor 1, partition 2, PROCESS_LOCAL, 8318 bytes)
24/04/28 21:21:23 INFO storage.BlockManagerInfo: Added broadcast_1_piece0 in memory
on algo-1:40893 (size: 14.2 KB, free: 13.8 GB)
24/04/28 21:21:24 INFO storage.BlockManagerInfo: Added broadcast_0_piece0 in memory
on algo-1:40893 (size: 27.6 KB, free: 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:22 INFO executor.CoarseGrainedExecutorBackend: Conn
ecting to driver: spark://CoarseGrainedScheduler@10.0.222.169:37411
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:22 INFO executor.CoarseGrainedExecutorBackend: Succ
essfully registered with driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:22 INFO executor.Executor: Starting executor ID 1 o
n host algo-1
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:22 INFO util.Utils: Successfully started service 'o
rg.apache.spark.network.netty.NettyBlockTransferService' on port 40893.
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:22 INFO netty.NettyBlockTransferService: Server cre
ated on algo-1:40893
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:22 INFO storage.BlockManager: Using org.apache.spar
k.storage.RandomBlockReplicationPolicy for block replication policy
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:22 INFO storage.BlockManagerMaster: Registering Blo
ckManager BlockManagerId(1, algo-1, 40893, None)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:22 INFO storage.BlockManagerMaster: Registered Bloc
kManager BlockManagerId(1, algo-1, 40893, None)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:22 INFO storage.BlockManager: Initialized BlockMana
ger: BlockManagerId(1, algo-1, 40893, None)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:23 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:23 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 1

```

```
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:23 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 2
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:23 INFO executor.Executor: Running task 0.0 in stage 0.0 (TID 0)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:23 INFO executor.Executor: Running task 1.0 in stage 0.0 (TID 1)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:23 INFO executor.Executor: Running task 2.0 in stage 0.0 (TID 2)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:23 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 1
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:23 INFO client.TransportClientFactory: Successfully created connection to /10.0.222.169:36237 after 2 ms (0 ms spent in bootstraps)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:23 INFO memory.MemoryStore: Block broadcast_1_piece0 stored as bytes in memory (estimated size 14.2 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:23 INFO broadcast.TorrentBroadcast: Reading broadcast variable 1 took 166 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:23 INFO memory.MemoryStore: Block broadcast_1 stored as values in memory (estimated size 32.9 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:24 INFO Configuration.deprecation: fs.s3a.server-side-encryption-key is deprecated. Instead, use fs.s3a.server-side-encryption.key
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:24 INFO executor.CoarseGrainedExecutorBackend: eagerFSInit: Eagerly initialized FileSystem at s3://does/not/exist in 2079 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:24 INFO codegen.CodeGenerator: Code generated in 244.489388 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:24 INFO datasources.FileScanRDD: TID: 2 - Reading current file: path: s3a://sagemaker-us-east-1-211125778552/amazon-reviews-pds/tsv/amazon_reviews_us_Gift_Card_v1_00.tsv.gz, range: 0-12134676, partition values: [empty row], isDataPresent: false
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:24 INFO datasources.FileScanRDD: TID: 0 - Reading current file: path: s3a://sagemaker-us-east-1-211125778552/amazon-reviews-pds/tsv/amazon_reviews_us_Digital_Video_Games_v1_00.tsv.gz, range: 0-27442648, partition values: [empty row], isDataPresent: false
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:24 INFO datasources.FileScanRDD: TID: 1 - Reading current file: path: s3a://sagemaker-us-east-1-211125778552/amazon-reviews-pds/tsv/amazon_reviews_us_Digital_Software_v1_00.tsv.gz, range: 0-18997559, partition values: [empty row], isDataPresent: false
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:24 INFO codegen.CodeGenerator: Code generated in 17.075007 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:24 INFO broadcast.TorrentBroadcast: Started reading
```

```
broadcast variable 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:24 INFO memory.MemoryStore: Block broadcast_0_piece
0 stored as bytes in memory (estimated size 27.6 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:24 INFO broadcast.TorrentBroadcast: Reading broadca
st variable 0 took 9 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:24 INFO memory.MemoryStore: Block broadcast_0 store
d as values in memory (estimated size 398.1 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:25 INFO zlib.ZlibFactory: Successfully loaded & ini
tialized native-zlib library
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:25 INFO compress.CodecPool: Got brand-new decompres
sor [.gz]
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:25 INFO compress.CodecPool: Got brand-new decompres
sor [.gz]
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:25 INFO compress.CodecPool: Got brand-new decompres
sor [.gz]
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:25 INFO codegen.CodeGenerator: Code generated in 1
7.692858 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:25 INFO codegen.CodeGenerator: Code generated in 7
3.691454 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:25 INFO codegen.CodeGenerator: Code generated in 8.
174423 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:25 INFO codegen.CodeGenerator: Code generated in 1
5.002028 ms
24/04/28 21:21:27 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 0.0 (TID
2) in 3842 ms on algo-1 (executor 1) (1/3)
24/04/28 21:21:27 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 0.0 (TID
1) in 3856 ms on algo-1 (executor 1) (2/3)
24/04/28 21:21:27 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 0.0 (TID
0) in 4276 ms on algo-1 (executor 1) (3/3)
24/04/28 21:21:27 INFO cluster.YarnScheduler: Removed TaskSet 0.0, whose tasks have
all completed, from pool
24/04/28 21:21:27 INFO scheduler.DAGScheduler: ShuffleMapStage 0 (collect at AnalysisRunner.scala:323) finished in 4.354 s
24/04/28 21:21:27 INFO scheduler.DAGScheduler: looking for newly runnable stages
24/04/28 21:21:27 INFO scheduler.DAGScheduler: running: Set()
24/04/28 21:21:27 INFO scheduler.DAGScheduler: waiting: Set()
24/04/28 21:21:27 INFO scheduler.DAGScheduler: failed: Set()
24/04/28 21:21:27 INFO adaptive.CoalesceShufflePartitions: advisoryTargetPostShuffle
InputSize: 67108864, targetPostShuffleInputSize 72.
24/04/28 21:21:27 INFO spark.SparkContext: Starting job: collect at AnalysisRunner.s
cala:323
24/04/28 21:21:27 INFO scheduler.DAGScheduler: Got job 1 (collect at AnalysisRunner.
scala:323) with 1 output partitions
24/04/28 21:21:27 INFO scheduler.DAGScheduler: Final stage: ResultStage 2 (collect a
t AnalysisRunner.scala:323)
```

```
24/04/28 21:21:27 INFO scheduler.DAGScheduler: Parents of final stage: List(ShuffleMapStage 1)
24/04/28 21:21:27 INFO scheduler.DAGScheduler: Missing parents: List()
24/04/28 21:21:27 INFO scheduler.DAGScheduler: Submitting ResultStage 2 (MapPartitionsRDD[6] at collect at AnalysisRunner.scala:323), which has no missing parents
24/04/28 21:21:27 INFO spark.ContextCleaner: Cleaned accumulator 47
24/04/28 21:21:27 INFO spark.ContextCleaner: Cleaned accumulator 39
24/04/28 21:21:27 INFO spark.ContextCleaner: Cleaned accumulator 31
24/04/28 21:21:27 INFO spark.ContextCleaner: Cleaned accumulator 46
24/04/28 21:21:27 INFO spark.ContextCleaner: Cleaned accumulator 36
24/04/28 21:21:27 INFO spark.ContextCleaner: Cleaned accumulator 52
24/04/28 21:21:27 INFO spark.ContextCleaner: Cleaned accumulator 29
24/04/28 21:21:27 INFO spark.ContextCleaner: Cleaned accumulator 32
24/04/28 21:21:27 INFO spark.ContextCleaner: Cleaned accumulator 48
24/04/28 21:21:27 INFO spark.ContextCleaner: Cleaned accumulator 37
24/04/28 21:21:27 INFO spark.ContextCleaner: Cleaned accumulator 51
24/04/28 21:21:27 INFO spark.ContextCleaner: Cleaned accumulator 43
24/04/28 21:21:27 INFO spark.ContextCleaner: Cleaned accumulator 33
24/04/28 21:21:27 INFO memory.MemoryStore: Block broadcast_2 stored as values in memory (estimated size 37.8 KB, free 1008.5 MB)
24/04/28 21:21:27 INFO memory.MemoryStore: Block broadcast_2_piece0 stored as bytes in memory (estimated size 15.9 KB, free 1008.5 MB)
24/04/28 21:21:27 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on 10.0.222.169:36237 (size: 15.9 KB, free: 1008.8 MB)
24/04/28 21:21:27 INFO spark.SparkContext: Created broadcast 2 from broadcast at DAGScheduler.scala:1203
24/04/28 21:21:27 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 2 (MapPartitionsRDD[6] at collect at AnalysisRunner.scala:323) (first 15 tasks are for partitions Vector(0))
24/04/28 21:21:27 INFO cluster.YarnScheduler: Adding task set 2.0 with 1 tasks
24/04/28 21:21:27 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 2.0 (TID 3, algo-1, executor 1, partition 0, NODE_LOCAL, 7778 bytes)
24/04/28 21:21:27 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on algo-1:40893 (size: 15.9 KB, free: 13.8 GB)
24/04/28 21:21:27 INFO storage.BlockManagerInfo: Removed broadcast_1_piece0 on 10.0.222.169:36237 in memory (size: 14.2 KB, free: 1008.9 MB)
24/04/28 21:21:27 INFO storage.BlockManagerInfo: Removed broadcast_1_piece0 on algo-1:40893 in memory (size: 14.2 KB, free: 13.8 GB)
24/04/28 21:21:27 INFO spark.ContextCleaner: Cleaned accumulator 49
24/04/28 21:21:27 INFO spark.ContextCleaner: Cleaned accumulator 45
24/04/28 21:21:27 INFO spark.ContextCleaner: Cleaned accumulator 50
24/04/28 21:21:27 INFO spark.ContextCleaner: Cleaned accumulator 30
24/04/28 21:21:27 INFO spark.ContextCleaner: Cleaned accumulator 44
24/04/28 21:21:27 INFO spark.ContextCleaner: Cleaned accumulator 35
24/04/28 21:21:27 INFO spark.ContextCleaner: Cleaned accumulator 41
24/04/28 21:21:27 INFO spark.ContextCleaner: Cleaned accumulator 34
24/04/28 21:21:27 INFO spark.ContextCleaner: Cleaned accumulator 42
24/04/28 21:21:27 INFO spark.ContextCleaner: Cleaned accumulator 53
24/04/28 21:21:27 INFO spark.ContextCleaner: Cleaned accumulator 38
24/04/28 21:21:27 INFO spark.ContextCleaner: Cleaned accumulator 40
24/04/28 21:21:27 INFO spark.MapOutputTrackerMasterEndpoint: Asked to send map output locations for shuffle 0 to 10.0.222.169:41176
24/04/28 21:21:28 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 2.0 (TID 3) in 258 ms on algo-1 (executor 1) (1/1)
24/04/28 21:21:28 INFO cluster.YarnScheduler: Removed TaskSet 2.0, whose tasks have all completed, from pool
```

```

24/04/28 21:21:28 INFO scheduler.DAGScheduler: ResultStage 2 (collect at AnalysisRun
ner.scala:323) finished in 0.272 s
24/04/28 21:21:28 INFO scheduler.DAGScheduler: Job 1 finished: collect at AnalysisRu
nner.scala:323, took 0.304407 s
24/04/28 21:21:28 INFO codegen.CodeGenerator: Code generated in 26.061336 ms
24/04/28 21:21:28 INFO codegen.CodeGenerator: Code generated in 13.79313 ms
24/04/28 21:21:28 INFO codegen.CodeGenerator: Code generated in 12.204502 ms
24/04/28 21:21:28 INFO spark.ContextCleaner: Cleaned accumulator 76
24/04/28 21:21:28 INFO spark.ContextCleaner: Cleaned accumulator 63
24/04/28 21:21:28 INFO spark.ContextCleaner: Cleaned accumulator 64
24/04/28 21:21:28 INFO spark.ContextCleaner: Cleaned accumulator 77
24/04/28 21:21:28 INFO spark.ContextCleaner: Cleaned accumulator 69
24/04/28 21:21:28 INFO storage.BlockManagerInfo: Removed broadcast_2_piece0 on 10.0.
222.169:36237 in memory (size: 15.9 KB, free: 1008.9 MB)
24/04/28 21:21:28 INFO storage.BlockManagerInfo: Removed broadcast_2_piece0 on algo-
1:40893 in memory (size: 15.9 KB, free: 13.8 GB)
24/04/28 21:21:28 INFO spark.ContextCleaner: Cleaned accumulator 74
24/04/28 21:21:28 INFO spark.ContextCleaner: Cleaned accumulator 55
24/04/28 21:21:28 INFO spark.ContextCleaner: Cleaned accumulator 71
24/04/28 21:21:28 INFO spark.ContextCleaner: Cleaned accumulator 61
24/04/28 21:21:28 INFO spark.ContextCleaner: Cleaned accumulator 73
24/04/28 21:21:28 INFO spark.ContextCleaner: Cleaned accumulator 66
24/04/28 21:21:28 INFO spark.ContextCleaner: Cleaned accumulator 58
24/04/28 21:21:28 INFO spark.ContextCleaner: Cleaned accumulator 67
24/04/28 21:21:28 INFO spark.ContextCleaner: Cleaned accumulator 56
24/04/28 21:21:28 INFO spark.ContextCleaner: Cleaned accumulator 60
24/04/28 21:21:28 INFO spark.ContextCleaner: Cleaned accumulator 59
24/04/28 21:21:28 INFO spark.ContextCleaner: Cleaned accumulator 75
24/04/28 21:21:28 INFO spark.ContextCleaner: Cleaned accumulator 65
24/04/28 21:21:28 INFO spark.ContextCleaner: Cleaned accumulator 70
24/04/28 21:21:28 INFO spark.ContextCleaner: Cleaned accumulator 68
24/04/28 21:21:28 INFO spark.ContextCleaner: Cleaned accumulator 62
24/04/28 21:21:28 INFO spark.ContextCleaner: Cleaned accumulator 72
24/04/28 21:21:28 INFO spark.ContextCleaner: Cleaned accumulator 57
24/04/28 21:21:28 INFO spark.ContextCleaner: Cleaned accumulator 78
24/04/28 21:21:28 INFO spark.ContextCleaner: Cleaned accumulator 54
[/var/log/yarn/userl+-----+-----+-----+-----+-----+-----+-----+-----+-----+
-----+
|entity      |instance                                |name                                |value                                |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|Column      |review_id                              |Completeness                        |1.0                                  |
|Column      |review_id                              |ApproxCountDistinct                |381704.0                             |
|Mutlicolumn |total_votes,star_rating                |Correlation                        |-0.08605234879757222                |
|Dataset     |*                                      |Size                               |396601.0                             |
|Column      |star_rating                            |Mean                               |4.102493437989314                    |
|Column      |top_star_rating                        |Compliance                         |0.7658931772738848                  |
|Mutlicolumn |total_votes,helpful_votes              |Correlation                        |0.9857511477962928                  |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
24/04/28 21:21:28 INFO output.FileOutputCommitter: File Output Committer Algorithm v
ersion is 2
24/04/28 21:21:28 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup
_temporary folders under output directory:false, ignore cleanup failures: false
24/04/28 21:21:28 INFO output.DirectFileOutputCommitter: Direct Write: DISABLED
24/04/28 21:21:28 INFO datasources.SQLConfCommitterProvider: Using output committer
class org.apache.hadoop.mapreduce.lib.output.DirectFileOutputCommitter
24/04/28 21:21:29 INFO codegen.CodeGenerator: Code generated in 13.322838 ms

```

```
24/04/28 21:21:29 INFO scheduler.DAGScheduler: Registering RDD 9 (save at NativeMethodAccessorImpl.java:0) as input to shuffle 1
24/04/28 21:21:29 INFO scheduler.DAGScheduler: Got map stage job 2 (save at NativeMethodAccessorImpl.java:0) with 7 output partitions
24/04/28 21:21:29 INFO scheduler.DAGScheduler: Final stage: ShuffleMapStage 3 (save at NativeMethodAccessorImpl.java:0)
24/04/28 21:21:29 INFO scheduler.DAGScheduler: Parents of final stage: List()
24/04/28 21:21:29 INFO scheduler.DAGScheduler: Missing parents: List()
24/04/28 21:21:29 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 3 (MapPartitionsRDD[9] at save at NativeMethodAccessorImpl.java:0), which has no missing parents
24/04/28 21:21:29 INFO memory.MemoryStore: Block broadcast_3 stored as values in memory (estimated size 5.4 KB, free 1008.6 MB)
24/04/28 21:21:29 INFO memory.MemoryStore: Block broadcast_3_piece0 stored as bytes in memory (estimated size 3.3 KB, free 1008.6 MB)
24/04/28 21:21:29 INFO storage.BlockManagerInfo: Added broadcast_3_piece0 in memory on 10.0.222.169:36237 (size: 3.3 KB, free: 1008.9 MB)
24/04/28 21:21:29 INFO spark.SparkContext: Created broadcast 3 from broadcast at DAGScheduler.scala:1203
24/04/28 21:21:29 INFO scheduler.DAGScheduler: Submitting 7 missing tasks from ShuffleMapStage 3 (MapPartitionsRDD[9] at save at NativeMethodAccessorImpl.java:0) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6))
24/04/28 21:21:29 INFO cluster.YarnScheduler: Adding task set 3.0 with 7 tasks
24/04/28 21:21:29 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 3.0 (TID 4, algo-1, executor 1, partition 0, PROCESS_LOCAL, 8108 bytes)
24/04/28 21:21:29 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 3.0 (TID 5, algo-1, executor 1, partition 1, PROCESS_LOCAL, 8116 bytes)
24/04/28 21:21:29 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 3.0 (TID 6, algo-1, executor 1, partition 2, PROCESS_LOCAL, 8124 bytes)
24/04/28 21:21:29 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 3.0 (TID 7, algo-1, executor 1, partition 3, PROCESS_LOCAL, 8092 bytes)
24/04/28 21:21:29 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 3.0 (TID 8, algo-1, executor 1, partition 4, PROCESS_LOCAL, 8100 bytes)
24/04/28 21:21:29 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 3.0 (TID 9, algo-1, executor 1, partition 5, PROCESS_LOCAL, 8108 bytes)
24/04/28 21:21:29 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 3.0 (TID 10, algo-1, executor 1, partition 6, PROCESS_LOCAL, 8132 bytes)
24/04/28 21:21:29 INFO storage.BlockManagerInfo: Added broadcast_3_piece0 in memory on algo-1:40893 (size: 3.3 KB, free: 13.8 GB)
24/04/28 21:21:29 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 3.0 (TID 4) in 38 ms on algo-1 (executor 1) (1/7)
24/04/28 21:21:29 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 3.0 (TID 9) in 36 ms on algo-1 (executor 1) (2/7)
24/04/28 21:21:29 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 3.0 (TID 10) in 37 ms on algo-1 (executor 1) (3/7)
24/04/28 21:21:29 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 3.0 (TID 8) in 39 ms on algo-1 (executor 1) (4/7)
24/04/28 21:21:29 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 3.0 (TID 7) in 39 ms on algo-1 (executor 1) (5/7)
24/04/28 21:21:29 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 3.0 (TID 5) in 41 ms on algo-1 (executor 1) (6/7)
24/04/28 21:21:29 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 3.0 (TID 6) in 41 ms on algo-1 (executor 1) (7/7)
24/04/28 21:21:29 INFO cluster.YarnScheduler: Removed TaskSet 3.0, whose tasks have all completed, from pool
24/04/28 21:21:29 INFO scheduler.DAGScheduler: ShuffleMapStage 3 (save at NativeMethod
```

```

odAccessorImpl.java:0) finished in 0.054 s
24/04/28 21:21:29 INFO scheduler.DAGScheduler: looking for newly runnable stages
24/04/28 21:21:29 INFO scheduler.DAGScheduler: running: Set()
24/04/28 21:21:29 INFO scheduler.DAGScheduler: waiting: Set()
24/04/28 21:21:29 INFO scheduler.DAGScheduler: failed: Set()
24/04/28 21:21:29 INFO spark.SparkContext: Starting job: save at NativeMethodAccesso
rImpl.java:0
24/04/28 21:21:29 INFO scheduler.DAGScheduler: Got job 3 (save at NativeMethodAccess
orImpl.java:0) with 1 output partitions
24/04/28 21:21:29 INFO scheduler.DAGScheduler: Final stage: ResultStage 5 (save at N
ativeMethodAccessorImpl.java:0)
24/04/28 21:21:29 INFO scheduler.DAGScheduler: Parents of final stage: List(ShuffleM
apStage 4)
24/04/28 21:21:29 INFO scheduler.DAGScheduler: Missing parents: List()
24/04/28 21:21:29 INFO scheduler.DAGScheduler: Submitting ResultStage 5 (ShuffledRow
RDD[10] at save at NativeMethodAccessorImpl.java:0), which has no missing parents
24/04/28 21:21:29 INFO memory.MemoryStore: Block broadcast_4 stored as values in mem
ory (estimated size 167.5 KB, free 1008.4 MB)
24/04/28 21:21:29 INFO memory.MemoryStore: Block broadcast_4_piece0 stored as bytes
in memory (estimated size 60.3 KB, free 1008.3 MB)
24/04/28 21:21:29 INFO storage.BlockManagerInfo: Added broadcast_4_piece0 in memory
on 10.0.222.169:36237 (size: 60.3 KB, free: 1008.8 MB)
24/04/28 21:21:29 INFO spark.SparkContext: Created broadcast 4 from broadcast at DAG
Scheduler.scala:1203
24/04/28 21:21:29 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Resul
tStage 5 (ShuffledRowRDD[10] at save at NativeMethodAccessorImpl.java:0) (first 15 t
asks are for partitions Vector(0))
24/04/28 21:21:29 INFO cluster.YarnScheduler: Adding task set 5.0 with 1 tasks
24/04/28 21:21:29 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 5.0 (TID
11, algo-1, executor 1, partition 0, NODE_LOCAL, 7778 bytes)
24/04/28 21:21:29 INFO storage.BlockManagerInfo: Added broadcast_4_piece0 in memory
on algo-1:40893 (size: 60.3 KB, free: 13.8 GB)
24/04/28 21:21:29 INFO spark.MapOutputTrackerMasterEndpoint: Asked to send map outpu
t locations for shuffle 1 to 10.0.222.169:41176
ogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 2
4/04/28 21:21:25 INFO codegen.CodeGenerator: Code generated in 43.363727 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:27 INFO executor.Executor: Finished task 2.0 in sta
ge 0.0 (TID 2). 2428 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:27 INFO executor.Executor: Finished task 1.0 in sta
ge 0.0 (TID 1). 2385 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:27 INFO executor.Executor: Finished task 0.0 in sta
ge 0.0 (TID 0). 2385 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:27 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 3
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:27 INFO executor.Executor: Running task 0.0 in stag
e 2.0 (TID 3)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:27 INFO spark.MapOutputTrackerWorker: Updating epoc
h to 1 and clearing cache
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:27 INFO broadcast.TorrentBroadcast: Started reading

```



```
broadcast variable 2
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:27 INFO memory.MemoryStore: Block broadcast_2_piece0 stored as bytes in memory (estimated size 15.9 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:27 INFO broadcast.TorrentBroadcast: Reading broadcast variable 2 took 9 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:27 INFO memory.MemoryStore: Block broadcast_2 stored as values in memory (estimated size 37.8 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:27 INFO spark.MapOutputTrackerWorker: Don't have map outputs for shuffle 0, fetching them
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:27 INFO spark.MapOutputTrackerWorker: Doing the fetch; tracker endpoint = NettyRpcEndpointRef(spark://MapOutputTracker@10.0.222.169:37411)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:27 INFO spark.MapOutputTrackerWorker: Got the output locations
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:28 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:28 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 11 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:28 INFO codegen.CodeGenerator: Code generated in 34.150405 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:28 INFO codegen.CodeGenerator: Code generated in 10.762936 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:28 INFO codegen.CodeGenerator: Code generated in 13.00211 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:28 INFO executor.Executor: Finished task 0.0 in stage 2.0 (TID 3). 3669 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 4
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO executor.Executor: Running task 0.0 in stage 3.0 (TID 4)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 5
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO executor.Executor: Running task 1.0 in stage 3.0 (TID 5)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 6
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO executor.Executor: Running task 2.0 in stage 3.0 (TID 6)
```

```
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 7
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 8
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO executor.Executor: Running task 3.0 in stage 3.0 (TID 7)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO executor.Executor: Running task 4.0 in stage 3.0 (TID 8)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 9
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 3
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 10
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO executor.Executor: Running task 5.0 in stage 3.0 (TID 9)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO executor.Executor: Running task 6.0 in stage 3.0 (TID 10)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO memory.MemoryStore: Block broadcast_3_piece0 stored as bytes in memory (estimated size 3.3 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO broadcast.TorrentBroadcast: Reading broadcast variable 3 took 10 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO memory.MemoryStore: Block broadcast_3 stored as values in memory (estimated size 5.4 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO executor.Executor: Finished task 0.0 in stage 3.0 (TID 4). 1384 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO executor.Executor: Finished task 5.0 in stage 3.0 (TID 9). 1384 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO executor.Executor: Finished task 6.0 in stage 3.0 (TID 10). 1384 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO executor.Executor: Finished task 4.0 in stage 3.0 (TID 8). 1384 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO executor.Executor: Finished task 3.0 in stage 3.0 (TID 7). 1427 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO executor.Executor: Finished task 2.0 in stage 3.0 (TID 6). 1384 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO executor.Executor: Finished task 1.0 in stage 3.0 (TID 5). 1384 bytes result sent to driver
```

```

ge 3.0 (TID 5). 1384 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 11
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO executor.Executor: Running task 0.0 in stage 5.0 (TID 11)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO spark.MapOutputTrackerWorker: Updating epoch to 2 and clearing cache
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 4
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO memory.MemoryStore: Block broadcast_4_piece 0 stored as bytes in memory (estimated size 60.3 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO broadcast.TorrentBroadcast: Reading broadcast variable 4 took 7 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO memory.MemoryStore: Block broadcast_4 stored as values in memory (estimated size 167.5 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO spark.MapOutputTrackerWorker: Don't have map outputs for shuffle 1, fetching them
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO spark.MapOutputTrackerWorker: Doing the fetch; tracker endpoint = NettyRpcEndpointRef(spark://MapOutputTracker@10.0.222.169:37411)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO spark.MapOutputTrackerWorker: Got the output locations
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO storage.ShuffleBlockFetcherIterator: Getting 7 non-empty blocks including 7 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:29 INFO output.DirectFileOutputCommitter: Direct Write: DISABLED
24/04/28 21:21:30 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 5.0 (TID 11) in 1180 ms on algo-1 (executor 1) (1/1)
24/04/28 21:21:30 INFO cluster.YarnScheduler: Removed TaskSet 5.0, whose tasks have all completed, from pool
24/04/28 21:21:30 INFO scheduler.DAGScheduler: ResultStage 5 (save at NativeMethodAccessorImpl.java:0) finished in 1.205 s
24/04/28 21:21:30 INFO scheduler.DAGScheduler: Job 3 finished: save at NativeMethodAccessorImpl.java:0, took 1.207326 s

```

```
24/04/28 21:21:31 INFO datasources.FileFormatWriter: Write Job 5ebce304-ef8e-4c5a-82
52-56b2a974031b committed.
24/04/28 21:21:31 INFO datasources.FileFormatWriter: Finished processing stats for w
rite job 5ebce304-ef8e-4c5a-8252-56b2a974031b.
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:29 INFO datasources.SQLConfCommitterProvider: Using
output committer clasPython Callback server started!
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 107
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 89
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 118
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 102
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 131
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 91
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 110
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 125
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 83
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 121
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 88
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned shuffle 1
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 104
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 127
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 136
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 106
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 99
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 79
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 92
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 80
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 122
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 113
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 116
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 111
24/04/28 21:21:31 INFO storage.BlockManagerInfo: Removed broadcast_3_piece0 on 10.0.
222.169:36237 in memory (size: 3.3 KB, free: 1008.8 MB)
24/04/28 21:21:31 INFO storage.BlockManagerInfo: Removed broadcast_3_piece0 on algo-
1:40893 in memory (size: 3.3 KB, free: 13.8 GB)
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 109
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 96
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 87
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 134
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 82
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 128
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 123
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 101
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 93
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 105
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 86
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 117
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 97
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 94
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 130
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 108
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 100
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 126
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 114
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 115
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 133
```

```
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 120
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 137
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 135
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 90
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 84
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 81
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 124
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 138
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 98
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 85
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 95
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 103
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 119
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 112
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 129
24/04/28 21:21:31 INFO storage.BlockManagerInfo: Removed broadcast_4_piece0 on 10.0.
222.169:36237 in memory (size: 60.3 KB, free: 1008.9 MB)
24/04/28 21:21:31 INFO storage.BlockManagerInfo: Removed broadcast_4_piece0 on algo-
1:40893 in memory (size: 60.3 KB, free: 13.8 GB)
24/04/28 21:21:31 INFO spark.ContextCleaner: Cleaned accumulator 132
24/04/28 21:21:31 INFO datasources.FileSourceStrategy: Pruning directories with:
24/04/28 21:21:31 INFO datasources.FileSourceStrategy: Post-Scan Filters:
24/04/28 21:21:31 INFO datasources.FileSourceStrategy: Output Data Schema: struct<ma
rketplace: string, review_id: string, star_rating: int ... 1 more fields>
24/04/28 21:21:31 INFO execution.FileSourceScanExec: Pushed Filters:
24/04/28 21:21:31 INFO execution.FileSourceScanExec: Pushed Filters:
24/04/28 21:21:31 INFO codegen.CodeGenerator: Code generated in 8.675859 ms
24/04/28 21:21:31 INFO codegen.CodeGenerator: Code generated in 20.069239 ms
24/04/28 21:21:31 INFO memory.MemoryStore: Block broadcast_5 stored as values in mem
ory (estimated size 303.8 KB, free 1008.3 MB)
24/04/28 21:21:31 INFO memory.MemoryStore: Block broadcast_5_piece0 stored as bytes
in memory (estimated size 27.6 KB, free 1008.3 MB)
24/04/28 21:21:31 INFO storage.BlockManagerInfo: Added broadcast_5_piece0 in memory
on 10.0.222.169:36237 (size: 27.6 KB, free: 1008.8 MB)
24/04/28 21:21:31 INFO spark.SparkContext: Created broadcast 5 from collect at Analy
sisRunner.scala:323
24/04/28 21:21:31 INFO execution.FileSourceScanExec: Planning scan with bin packing,
max size: 4447362 bytes, open cost is considered as scanning 4194304 bytes, number o
f split files: 3, prefetch: false
24/04/28 21:21:31 INFO execution.FileSourceScanExec: relation: None, fileSplitsInPar
titionHistogram: ArrayBuffer((1 fileSplits,3))
24/04/28 21:21:31 INFO scheduler.DAGScheduler: Registering RDD 15 (collect at Analys
isRunner.scala:323) as input to shuffle 2
24/04/28 21:21:31 INFO scheduler.DAGScheduler: Got map stage job 4 (collect at Analy
sisRunner.scala:323) with 3 output partitions
24/04/28 21:21:31 INFO scheduler.DAGScheduler: Final stage: ShuffleMapStage 6 (colle
ct at AnalysisRunner.scala:323)
24/04/28 21:21:31 INFO scheduler.DAGScheduler: Parents of final stage: List()
24/04/28 21:21:31 INFO scheduler.DAGScheduler: Missing parents: List()
24/04/28 21:21:31 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 6 (MapPart
itionsRDD[15] at collect at AnalysisRunner.scala:323), which has no missing parents
24/04/28 21:21:31 INFO memory.MemoryStore: Block broadcast_6 stored as values in mem
ory (estimated size 20.7 KB, free 1008.2 MB)
24/04/28 21:21:31 INFO memory.MemoryStore: Block broadcast_6_piece0 stored as bytes
in memory (estimated size 9.5 KB, free 1008.2 MB)
24/04/28 21:21:31 INFO storage.BlockManagerInfo: Added broadcast_6_piece0 in memory
```

on 10.0.222.169:36237 (size: 9.5 KB, free: 1008.8 MB)  
24/04/28 21:21:31 INFO spark.SparkContext: Created broadcast 6 from broadcast at DAG Scheduler.scala:1203  
24/04/28 21:21:31 INFO scheduler.DAGScheduler: Submitting 3 missing tasks from ShuffleMapStage 6 (MapPartitionsRDD[15] at collect at AnalysisRunner.scala:323) (first 15 tasks are for partitions Vector(0, 1, 2))  
24/04/28 21:21:31 INFO cluster.YarnScheduler: Adding task set 6.0 with 3 tasks  
24/04/28 21:21:31 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 6.0 (TID 12, algo-1, executor 1, partition 0, PROCESS\_LOCAL, 8328 bytes)  
24/04/28 21:21:31 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 6.0 (TID 13, algo-1, executor 1, partition 1, PROCESS\_LOCAL, 8325 bytes)  
24/04/28 21:21:31 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 6.0 (TID 14, algo-1, executor 1, partition 2, PROCESS\_LOCAL, 8318 bytes)  
24/04/28 21:21:31 INFO storage.BlockManagerInfo: Added broadcast\_6\_piece0 in memory on algo-1:40893 (size: 9.5 KB, free: 13.8 GB)  
24/04/28 21:21:31 INFO storage.BlockManagerInfo: Added broadcast\_5\_piece0 in memory on algo-1:40893 (size: 27.6 KB, free: 13.8 GB)  
s org.apache.hadoop.mapreduce.lib.output.DirectFileOutputCommitter  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:30 INFO output.FileOutputCommitter: Saved output of task 'attempt\_20240428212129\_0005\_m\_000000\_11' to s3a://sagemaker-us-east-1-211125778552/amazon-reviews-spark-analyzer-2024-04-28-21-15-58/output/dataset-metrics  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:30 INFO mapred.SparkHadoopMapRedUtil: attempt\_20240428212129\_0005\_m\_000000\_11: Committed  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:30 INFO executor.Executor: Finished task 0.0 in stage 5.0 (TID 11). 2477 bytes result sent to driver  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:31 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 12  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:31 INFO executor.Executor: Running task 0.0 in stage 6.0 (TID 12)  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:31 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 6  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:31 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 13  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:31 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 14  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:31 INFO executor.Executor: Running task 1.0 in stage 6.0 (TID 13)  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:31 INFO executor.Executor: Running task 2.0 in stage 6.0 (TID 14)  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:31 INFO memory.MemoryStore: Block broadcast\_6\_piece0 stored as bytes in memory (estimated size 9.5 KB, free 13.8 GB)  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:31 INFO broadcast.TorrentBroadcast: Reading broadcast variable 6 took 24 ms  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_

```
01_000002/stderr] 24/04/28 21:21:31 INFO memory.MemoryStore: Block broadcast_6 stored as values in memory (estimated size 20.7 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:31 INFO codegen.CodeGenerator: Code generated in 21.585775 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:31 INFO datasources.FileScanRDD: TID: 13 - Reading current file: path: s3a://sagemaker-us-east-1-211125778552/amazon-reviews-pds/tsv/amazon_reviews_us_Digital_Software_v1_00.tsv.gz, range: 0-18997559, partition values: [empty row], isDataPresent: false
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:31 INFO datasources.FileScanRDD: TID: 14 - Reading current file: path: s3a://sagemaker-us-east-1-211125778552/amazon-reviews-pds/tsv/amazon_reviews_us_Gift_Card_v1_00.tsv.gz, range: 0-12134676, partition values: [empty row], isDataPresent: false
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:31 INFO datasources.FileScanRDD: TID: 12 - Reading current file: path: s3a://sagemaker-us-east-1-211125778552/amazon-reviews-pds/tsv/amazon_reviews_us_Digital_Video_Games_v1_00.tsv.gz, range: 0-27442648, partition values: [empty row], isDataPresent: false
24/04/28 21:21:32 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 6.0 (TID 14) in 1044 ms on algo-1 (executor 1) (1/3)
24/04/28 21:21:32 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 6.0 (TID 13) in 1242 ms on algo-1 (executor 1) (2/3)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stdout] 2024-04-28T21:21:20.108+0000: [GC (Allocation Failure) [PSYoungGen: 120832K->6766K(140800K)] 120832K->6774K(463360K), 0.0058423 secs] [Times: user=0.02 sys=0.00, real=0.01 secs]
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stdout] 2024-04-28T21:21:20.544+0000: [GC (Allocation Failure) [PSYoungGen: 127598K->6174K(140800K)] 127606K->6190K(463360K), 0.0047884 secs] [Times: user=0.02 sys=0.00, real=0.00 secs]
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stdout] 2024-04-28T21:21:21.054+0000: [GC (Allocation Failure) [PSYoungGen: 127006K->6580K(140800K)] 127022K->6604K(463360K), 0.0073651 secs] [Times: user=0.04 sys=0.00, real=0.00 secs]
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stdout] 2024-04-28T21:21:21.427+0000: [GC (Allocation Failure) [PSYoungGen: 126161K->8031K(261632K)] 126185K->8063K(584192K), 0.0107788 secs] [Times: user=0.02 sys=0.02, real=0.01 secs]
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stdout] 2024-04-28T21:21:21.565+0000: [GC (Metadata GC Threshold) [PSYoungGen: 65455K->19938K(261632K)] 65487K->22790K(584192K), 0.0161732 secs] [Times: user=0.06 sys=0.01, real=0.01 secs]
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stdout] 2024-04-28T21:21:21.581+0000: [Full GC (Metadata GC Threshold) [PSYoungGen: 19938K->0K(261632K)] [ParOldGen: 2851K->22657K(181248K)] 22790K->22657K(442880K), [Metaspace: 20958K->20958K(1067008K)], 0.0381492 secs] [Times: user=0.15 sys=0.00, real=0.04 secs]
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stdout] 2024-04-28T21:21:22.458+0000: [GC (Allocation Failure) [PSYoungGen: 241664K->18912K(429056K)] 264321K->47785K(610304K), 0.0253269 secs] [Times: user=0.12 sys=0.01, real=0.03 secs]
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stdout] 2024-04-28T21:21:23.229+0000: [GC (Metadata GC Threshold) [PSYoungGen: 224321K->26087K(495104K)] 253194K->55287K(676352K), 0.0181260 secs] [Times: use
```

```

r=0.09 sys=0.01, real=0.02 secs]
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stdout] 2024-04-28T21:21:23.247+0000: [Full GC (Metadata GC Threshold) [PSYoungGen: 26087K->0K(495104K)] [ParOldGen: 29199K->49659K(266752K)] 55287K->49659K(761856K), [Metaspace: 34894K->34894K(1079296K)], 0.0346183 secs] [Times: user=0.12 sys=0.02, real=0.03 secs]
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stdout] 2024-04-28T21:21:24.951+0000: [GC (Allocation Failure) [PSYoungGen: 468992K->31213K(712704K)] 518651K->83317K(979456K), 0.0228511 secs] [Times: user=0.07 sys=0.02, real=0.03 secs]
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stdout] 2024-04-28T21:21:26.447+0000: [GC (Allocation Failure) [PSYoungGen: 712685K->29472K(718848K)] 764789K->540328K(1249792K), 0.1281464 secs] [Times: user=0.48 sys=0.28, real=0.13 secs]
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stdout] 2024-04-28T21:21:26.575+0000: [Full GC (Ergonomics) [PSYoungGen: 29472K->0K(718848K)] [ParOldGen: 510856K->269652K(715264K)] 540328K->269652K(1434112K), [Metaspace: 54999K->54929K(1097728K)], 0.0989945 secs] [Times: user=0.33 sys=0.03, real=0.10 secs]
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stdout] 2024-04-28T21:21:27.993+0000: [GC (Allocation Failure) [PSYoungGen: 681472K->1735K(924672K)] 951124K->271395K(1639936K), 0.0030559 secs] [Times: user=0.01 sys=0.00, real=0.00 secs]
24/04/28 21:21:33 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 6.0 (TID 12) in 1546 ms on algo-1 (executor 1) (3/3)
24/04/28 21:21:33 INFO cluster.YarnScheduler: Removed TaskSet 6.0, whose tasks have all completed, from pool
24/04/28 21:21:33 INFO scheduler.DAGScheduler: ShuffleMapStage 6 (collect at AnalysisRunner.scala:323) finished in 1.554 s
24/04/28 21:21:33 INFO scheduler.DAGScheduler: looking for newly runnable stages
24/04/28 21:21:33 INFO scheduler.DAGScheduler: running: Set()
24/04/28 21:21:33 INFO scheduler.DAGScheduler: waiting: Set()
24/04/28 21:21:33 INFO scheduler.DAGScheduler: failed: Set()
24/04/28 21:21:33 INFO adaptive.CoalesceShufflePartitions: advisoryTargetPostShuffleInputSize: 67108864, targetPostShuffleInputSize 16.
24/04/28 21:21:33 INFO codegen.CodeGenerator: Code generated in 16.283234 ms
24/04/28 21:21:33 INFO spark.SparkContext: Starting job: collect at AnalysisRunner.scala:323
24/04/28 21:21:33 INFO scheduler.DAGScheduler: Got job 5 (collect at AnalysisRunner.scala:323) with 1 output partitions
24/04/28 21:21:33 INFO scheduler.DAGScheduler: Final stage: ResultStage 8 (collect at AnalysisRunner.scala:323)
24/04/28 21:21:33 INFO scheduler.DAGScheduler: Parents of final stage: List(ShuffleMapStage 7)
24/04/28 21:21:33 INFO scheduler.DAGScheduler: Missing parents: List()
24/04/28 21:21:33 INFO scheduler.DAGScheduler: Submitting ResultStage 8 (MapPartitionsRDD[18] at collect at AnalysisRunner.scala:323), which has no missing parents
24/04/28 21:21:33 INFO memory.MemoryStore: Block broadcast_7 stored as values in memory (estimated size 13.6 KB, free 1008.2 MB)
24/04/28 21:21:33 INFO memory.MemoryStore: Block broadcast_7_piece0 stored as bytes in memory (estimated size 6.0 KB, free 1008.2 MB)
24/04/28 21:21:33 INFO storage.BlockManagerInfo: Added broadcast_7_piece0 in memory on 10.0.222.169:36237 (size: 6.0 KB, free: 1008.8 MB)
24/04/28 21:21:33 INFO spark.SparkContext: Created broadcast 7 from broadcast at DAGScheduler.scala:1203
24/04/28 21:21:33 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 8

```



```
tStage 8 (MapPartitionsRDD[18] at collect at AnalysisRunner.scala:323) (first 15 tasks are for partitions Vector(0))
24/04/28 21:21:33 INFO cluster.YarnScheduler: Adding task set 8.0 with 1 tasks
24/04/28 21:21:33 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 8.0 (TID 15, algo-1, executor 1, partition 0, NODE_LOCAL, 7778 bytes)
24/04/28 21:21:33 INFO storage.BlockManagerInfo: Added broadcast_7_piece0 in memory on algo-1:40893 (size: 6.0 KB, free: 13.8 GB)
24/04/28 21:21:33 INFO spark.MapOutputTrackerMasterEndpoint: Asked to send map output locations for shuffle 2 to 10.0.222.169:41176
24/04/28 21:21:33 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 8.0 (TID 15) in 41 ms on algo-1 (executor 1) (1/1)
24/04/28 21:21:33 INFO cluster.YarnScheduler: Removed TaskSet 8.0, whose tasks have all completed, from pool
24/04/28 21:21:33 INFO scheduler.DAGScheduler: ResultStage 8 (collect at AnalysisRunner.scala:323) finished in 0.047 s
24/04/28 21:21:33 INFO scheduler.DAGScheduler: Job 5 finished: collect at AnalysisRunner.scala:323, took 0.049915 s
24/04/28 21:21:33 INFO datasources.FileSourceStrategy: Pruning directories with:
24/04/28 21:21:33 INFO datasources.FileSourceStrategy: Post-Scan Filters: isNotNull(review_id#2)
24/04/28 21:21:33 INFO datasources.FileSourceStrategy: Output Data Schema: struct<review_id: string>
24/04/28 21:21:33 INFO execution.FileSourceScanExec: Pushed Filters: IsNotNull(review_id)
24/04/28 21:21:33 INFO execution.FileSourceScanExec: Pushed Filters: IsNotNull(none)
24/04/28 21:21:33 INFO codegen.CodeGenerator: Code generated in 13.432782 ms
24/04/28 21:21:33 INFO memory.MemoryStore: Block broadcast_8 stored as values in memory (estimated size 303.8 KB, free 1007.9 MB)
24/04/28 21:21:33 INFO memory.MemoryStore: Block broadcast_8_piece0 stored as bytes in memory (estimated size 27.6 KB, free 1007.9 MB)
24/04/28 21:21:33 INFO storage.BlockManagerInfo: Added broadcast_8_piece0 in memory on 10.0.222.169:36237 (size: 27.6 KB, free: 1008.8 MB)
24/04/28 21:21:33 INFO spark.SparkContext: Created broadcast 8 from count at GroupingAnalyzers.scala:80
24/04/28 21:21:33 INFO execution.FileSourceScanExec: Planning scan with bin packing, max size: 4447362 bytes, open cost is considered as scanning 4194304 bytes, number of split files: 3, prefetch: false
24/04/28 21:21:33 INFO execution.FileSourceScanExec: relation: None, fileSplitsInPartitionHistogram: ArrayBuffer((1 fileSplits,3))
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 178
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 143
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 173
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 176
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 153
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 214
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 175
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 209
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 199
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 180
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 163
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 151
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 215
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 190
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 170
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 149
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 195
```

```
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 177
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 140
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 164
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 185
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 187
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 182
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 157
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 201
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 145
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 166
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 174
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 194
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 217
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 144
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 213
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 161
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 158
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 200
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 179
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 172
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 204
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 141
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 205
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 169
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 139
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 171
24/04/28 21:21:33 INFO scheduler.DAGScheduler: Registering RDD 21 (count at GroupingAnalyzers.scala:80) as input to shuffle 3
24/04/28 21:21:33 INFO scheduler.DAGScheduler: Got map stage job 6 (count at GroupingAnalyzers.scala:80) with 3 output partitions
24/04/28 21:21:33 INFO scheduler.DAGScheduler: Final stage: ShuffleMapStage 9 (count at GroupingAnalyzers.scala:80)
24/04/28 21:21:33 INFO scheduler.DAGScheduler: Parents of final stage: List()
24/04/28 21:21:33 INFO scheduler.DAGScheduler: Missing parents: List()
24/04/28 21:21:33 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 9 (MapPartitionsRDD[21] at count at GroupingAnalyzers.scala:80), which has no missing parents
24/04/28 21:21:33 INFO storage.BlockManagerInfo: Removed broadcast_7_piece0 on 10.0.222.169:36237 in memory (size: 6.0 KB, free: 1008.8 MB)
24/04/28 21:21:33 INFO storage.BlockManagerInfo: Removed broadcast_7_piece0 on algo-1:40893 in memory (size: 6.0 KB, free: 13.8 GB)
24/04/28 21:21:33 INFO memory.MemoryStore: Block broadcast_9 stored as values in memory (estimated size 14.1 KB, free 1007.9 MB)
24/04/28 21:21:33 INFO memory.MemoryStore: Block broadcast_9_piece0 stored as bytes in memory (estimated size 7.4 KB, free 1007.9 MB)
24/04/28 21:21:33 INFO storage.BlockManagerInfo: Added broadcast_9_piece0 in memory on 10.0.222.169:36237 (size: 7.4 KB, free: 1008.8 MB)
24/04/28 21:21:33 INFO spark.SparkContext: Created broadcast 9 from broadcast at DAGScheduler.scala:1203
24/04/28 21:21:33 INFO scheduler.DAGScheduler: Submitting 3 missing tasks from ShuffleMapStage 9 (MapPartitionsRDD[21] at count at GroupingAnalyzers.scala:80) (first 15 tasks are for partitions Vector(0, 1, 2))
24/04/28 21:21:33 INFO cluster.YarnScheduler: Adding task set 9.0 with 3 tasks
24/04/28 21:21:33 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 9.0 (TID 16, algo-1, executor 1, partition 0, PROCESS_LOCAL, 8328 bytes)
24/04/28 21:21:33 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 9.0 (TID 17, algo-1, executor 1, partition 1, PROCESS_LOCAL, 8325 bytes)
```

```
24/04/28 21:21:33 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 9.0 (TID
18, algo-1, executor 1, partition 2, PROCESS_LOCAL, 8318 bytes)
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 212
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 196
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 142
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 203
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 202
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 168
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 147
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 184
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 154
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 216
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 162
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 148
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 181
24/04/28 21:21:33 INFO storage.BlockManagerInfo: Removed broadcast_6_piece0 on 10.0.
222.169:36237 in memory (size: 9.5 KB, free: 1008.8 MB)
24/04/28 21:21:33 INFO storage.BlockManagerInfo: Removed broadcast_6_piece0 on algo-
1:40893 in memory (size: 9.5 KB, free: 13.8 GB)
24/04/28 21:21:33 INFO storage.BlockManagerInfo: Added broadcast_9_piece0 in memory
on algo-1:40893 (size: 7.4 KB, free: 13.8 GB)
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 207
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 208
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 167
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 188
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 211
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 146
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 183
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 191
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 198
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned shuffle 2
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 160
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 155
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 197
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 186
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 210
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 206
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 150
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 159
24/04/28 21:21:33 INFO storage.BlockManagerInfo: Removed broadcast_5_piece0 on 10.0.
222.169:36237 in memory (size: 27.6 KB, free: 1008.8 MB)
24/04/28 21:21:33 INFO storage.BlockManagerInfo: Removed broadcast_5_piece0 on algo-
1:40893 in memory (size: 27.6 KB, free: 13.8 GB)
24/04/28 21:21:33 INFO storage.BlockManagerInfo: Added broadcast_8_piece0 in memory
on algo-1:40893 (size: 27.6 KB, free: 13.8 GB)
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 156
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 193
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 192
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 165
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 152
24/04/28 21:21:33 INFO spark.ContextCleaner: Cleaned accumulator 189
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stdout] 2024-04-28T21:21:32.725+0000: [GC (Allocation Failure) [PSYoungGe
n: 924359K->4626K(9661444339251228_0001/container_1714339251228_0001_01_000002/stderr
r] 24/04/28 21:21:31 INFO codegen.CodeGenerator: Code generated in 10.397632 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
```

```
01_000002/stderr] 24/04/28 21:21:31 INFO broadcast.TorrentBroadcast: Started reading
broadcast variable 5
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:31 INFO memory.MemoryStore: Block broadcast_5_piece
0 stored as bytes in memory (estimated size 27.6 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:31 INFO broadcast.TorrentBroadcast: Reading broadca
st variable 5 took 14 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:31 INFO memory.MemoryStore: Block broadcast_5 store
d as values in memory (estimated size 398.1 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:32 INFO executor.Executor: Finished task 2.0 in sta
ge 6.0 (TID 14). 1740 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:32 INFO executor.Executor: Finished task 1.0 in sta
ge 6.0 (TID 13). 1783 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:33 INFO executor.Executor: Finished task 0.0 in sta
ge 6.0 (TID 12). 1783 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:33 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 15
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:33 INFO executor.Executor: Running task 0.0 in stag
e 8.0 (TID 15)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:33 INFO spark.MapOutputTrackerWorker: Updating epoc
h to 3 and clearing cache
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:33 INFO broadcast.TorrentBroadcast: Started reading
broadcast variable 7
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:33 INFO memory.MemoryStore: Block broadcast_7_piece
0 stored as bytes in memory (estimated size 6.0 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:33 INFO broadcast.TorrentBroadcast: Reading broadca
st variable 7 took 11 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:33 INFO memory.MemoryStore: Block broadcast_7 store
d as values in memory (estimated size 13.6 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:33 INFO spark.MapOutputTrackerWorker: Don't have ma
p outputs for shuffle 2, fetching them
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:33 INFO spark.MapOutputTrackerWorker: Doing the fet
ch; tracker endpoint = NettyRpcEndpointRef(spark://MapOutputTracker@10.0.222.169:374
11)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:33 INFO spark.MapOutputTrackerWorker: Got the outpu
t locations
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:33 INFO storage.ShuffleBlockFetcherIterator: Gettin
g 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:33 INFO storage.ShuffleBlockFetcherIterator: Starte
```

```
d 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:33 INFO codegen.CodeGenerator: Code generated in 13.53789 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:33 INFO executor.Executor: Finished task 0.0 in stage 8.0 (TID 15). 1955 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:33 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 16
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:33 INFO executor.Executor: Running task 0.0 in stage 9.0 (TID 16)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:33 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 9
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:33 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 17
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:33 INFO executor.Executor: Running task 1.0 in stage 9.0 (TID 17)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:33 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 18
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:33 INFO executor.Executor: Running task 2.0 in stage 9.0 (TID 18)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:33 INFO memory.MemoryStore: Block broadcast_9_piece 0 stored as bytes in memory (estimated size 7.4 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:33 INFO broadcast.TorrentBroadcast: Reading broadcast variable 9 took 19 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:33 INFO memory.MemoryStore: Block broadcast_9 stored as values in memory (estimated size 14.1 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:33 INFO codegen.CodeGenerator: Code generated in 12.938693 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:33 INFO datasources.FileScanRDD: TID: 17 - Reading current file: path: s3a://sagemaker-us-east-1-211125778552/amazon-reviews-pds/tsv/amazon_reviews_us_Digital_Software_v1_00.tsv.gz, range: 0-18997559, partition values: [empty row], isDataPresent: false
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:33 INFO datasources.FileScanRDD: TID: 16 - Reading current file: path: s3a://sagemaker-us-east-1-211125778552/amazon-reviews-pds/tsv/amazon_reviews_us_Digital_Video_Games_v1_00.tsv.gz, range: 0-27442648, partition values: [empty row], isDataPresent: false
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:33 INFO datasources.FileScanRDD: TID: 18 - Reading current file: path: s3a://sagemaker-us-east-1-211125778552/amazon-reviews-pds/tsv/amazon_reviews_us_Gift_Card_v1_00.tsv.gz, range: 0-12134676, partition values: [empty row], isDataPresent: false
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:33 INFO datasources.FileScanRDD: TID: 18 - Reading current file: path: s3a://sagemaker-us-east-1-211125778552/amazon-reviews-pds/tsv/amazon_reviews_us_Gift_Card_v1_00.tsv.gz, range: 0-12134676, partition values: [empty row], isDataPresent: false
```

```
01_000002/stderr] 24/04/28 21:21:33 INFO codegen.CodeGenerator: Code generated in 9.175071 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:33 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 8
24/04/28 21:21:34 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 9.0 (TID 18) in 842 ms on algo-1 (executor 1) (1/3)
24/04/28 21:21:34 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 9.0 (TID 17) in 997 ms on algo-1 (executor 1) (2/3)
24/04/28 21:21:34 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 9.0 (TID 16) in 1349 ms on algo-1 (executor 1) (3/3)
24/04/28 21:21:34 INFO cluster.YarnScheduler: Removed TaskSet 9.0, whose tasks have all completed, from pool
24/04/28 21:21:34 INFO scheduler.DAGScheduler: ShuffleMapStage 9 (count at GroupingAnalyzers.scala:80) finished in 1.358 s
24/04/28 21:21:34 INFO scheduler.DAGScheduler: looking for newly runnable stages
24/04/28 21:21:34 INFO scheduler.DAGScheduler: running: Set()
24/04/28 21:21:34 INFO scheduler.DAGScheduler: waiting: Set()
24/04/28 21:21:34 INFO scheduler.DAGScheduler: failed: Set()
24/04/28 21:21:34 INFO adaptive.CoalesceShufflePartitions: advisoryTargetPostShuffleInputSize: 67108864, targetPostShuffleInputSize 16.
24/04/28 21:21:34 INFO codegen.CodeGenerator: Code generated in 10.388184 ms
24/04/28 21:21:34 INFO spark.SparkContext: Starting job: count at GroupingAnalyzers.scala:80
24/04/28 21:21:34 INFO scheduler.DAGScheduler: Got job 7 (count at GroupingAnalyzer.s.scala:80) with 1 output partitions
24/04/28 21:21:34 INFO scheduler.DAGScheduler: Final stage: ResultStage 11 (count at GroupingAnalyzers.scala:80)
24/04/28 21:21:34 INFO scheduler.DAGScheduler: Parents of final stage: List(ShuffleMapStage 10)
24/04/28 21:21:34 INFO scheduler.DAGScheduler: Missing parents: List()
24/04/28 21:21:34 INFO scheduler.DAGScheduler: Submitting ResultStage 11 (MapPartitionsRDD[24] at count at GroupingAnalyzers.scala:80), which has no missing parents
24/04/28 21:21:34 INFO memory.MemoryStore: Block broadcast_10 stored as values in memory (estimated size 7.6 KB, free 1008.2 MB)
24/04/28 21:21:34 INFO memory.MemoryStore: Block broadcast_10_piece0 stored as bytes in memory (estimated size 4.2 KB, free 1008.2 MB)
24/04/28 21:21:34 INFO storage.BlockManagerInfo: Added broadcast_10_piece0 in memory on 10.0.222.169:36237 (size: 4.2 KB, free: 1008.8 MB)
24/04/28 21:21:34 INFO spark.SparkContext: Created broadcast 10 from broadcast at DAGScheduler.scala:1203
24/04/28 21:21:34 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 11 (MapPartitionsRDD[24] at count at GroupingAnalyzers.scala:80) (first 15 tasks are for partitions Vector(0))
24/04/28 21:21:34 INFO cluster.YarnScheduler: Adding task set 11.0 with 1 tasks
24/04/28 21:21:34 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 11.0 (TID 19, algo-1, executor 1, partition 0, NODE_LOCAL, 7778 bytes)
24/04/28 21:21:34 INFO storage.BlockManagerInfo: Added broadcast_10_piece0 in memory on algo-1:40893 (size: 4.2 KB, free: 13.8 GB)
24/04/28 21:21:34 INFO spark.MapOutputTrackerMasterEndpoint: Asked to send map output locations for shuffle 3 to 10.0.222.169:41176
24/04/28 21:21:34 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 11.0 (TID 19) in 28 ms on algo-1 (executor 1) (1/1)
24/04/28 21:21:34 INFO cluster.YarnScheduler: Removed TaskSet 11.0, whose tasks have all completed, from pool
24/04/28 21:21:34 INFO scheduler.DAGScheduler: ResultStage 11 (count at GroupingAnal
```

```
yzers.scala:80) finished in 0.033 s
24/04/28 21:21:34 INFO scheduler.DAGScheduler: Job 7 finished: count at GroupingAnal
yzers.scala:80, took 0.035500 s
24/04/28 21:21:35 INFO datasources.FileSourceStrategy: Pruning directories with:
24/04/28 21:21:35 INFO datasources.FileSourceStrategy: Post-Scan Filters: isNotNull
(review_id#2)
24/04/28 21:21:35 INFO datasources.FileSourceStrategy: Output Data Schema: struct<re
view_id: string>
24/04/28 21:21:35 INFO execution.FileSourceScanExec: Pushed Filters: IsNotNull(revie
w_id)
24/04/28 21:21:35 INFO execution.FileSourceScanExec: Pushed Filters: IsNotNull(none)
24/04/28 21:21:35 INFO codegen.CodeGenerator: Code generated in 6.127687 ms
24/04/28 21:21:35 INFO codegen.CodeGenerator: Code generated in 41.710775 ms
24/04/28 21:21:35 INFO memory.MemoryStore: Block broadcast_11 stored as values in me
memory (estimated size 303.8 KB, free 1007.9 MB)
24/04/28 21:21:35 INFO memory.MemoryStore: Block broadcast_11_piece0 stored as bytes
in memory (estimated size 27.6 KB, free 1007.9 MB)
24/04/28 21:21:35 INFO storage.BlockManagerInfo: Added broadcast_11_piece0 in memory
on 10.0.222.169:36237 (size: 27.6 KB, free: 1008.8 MB)
24/04/28 21:21:35 INFO spark.SparkContext: Created broadcast 11 from collect at Anal
ysisRunner.scala:523
24/04/28 21:21:35 INFO execution.FileSourceScanExec: Planning scan with bin packing,
max size: 4447362 bytes, open cost is considered as scanning 4194304 bytes, number o
f split files: 3, prefetch: false
24/04/28 21:21:35 INFO execution.FileSourceScanExec: relation: None, fileSplitsInPar
titionHistogram: ArrayBuffer((1 fileSplits,3))
24/04/28 21:21:35 INFO scheduler.DAGScheduler: Registering RDD 27 (collect at Analys
isRunner.scala:523) as input to shuffle 4
24/04/28 21:21:35 INFO scheduler.DAGScheduler: Got map stage job 8 (collect at Analy
sisRunner.scala:523) with 3 output partitions
24/04/28 21:21:35 INFO scheduler.DAGScheduler: Final stage: ShuffleMapStage 12 (coll
ect at AnalysisRunner.scala:523)
24/04/28 21:21:35 INFO scheduler.DAGScheduler: Parents of final stage: List()
24/04/28 21:21:35 INFO scheduler.DAGScheduler: Missing parents: List()
24/04/28 21:21:35 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 12 (MapPar
titionsRDD[27] at collect at AnalysisRunner.scala:523), which has no missing parents
24/04/28 21:21:35 INFO memory.MemoryStore: Block broadcast_12 stored as values in me
memory (estimated size 27.7 KB, free 1007.9 MB)
24/04/28 21:21:35 INFO memory.MemoryStore: Block broadcast_12_piece0 stored as bytes
in memory (estimated size 13.0 KB, free 1007.9 MB)
24/04/28 21:21:35 INFO storage.BlockManagerInfo: Added broadcast_12_piece0 in memory
on 10.0.222.169:36237 (size: 13.0 KB, free: 1008.8 MB)
24/04/28 21:21:35 INFO spark.SparkContext: Created broadcast 12 from broadcast at DA
GScheduler.scala:1203
24/04/28 21:21:35 INFO scheduler.DAGScheduler: Submitting 3 missing tasks from Shuff
leMapStage 12 (MapPartitionsRDD[27] at collect at AnalysisRunner.scala:523) (first 1
5 tasks are for partitions Vector(0, 1, 2))
24/04/28 21:21:35 INFO cluster.YarnScheduler: Adding task set 12.0 with 3 tasks
24/04/28 21:21:35 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 12.0 (TI
D 20, algo-1, executor 1, partition 0, PROCESS_LOCAL, 8328 bytes)
24/04/28 21:21:35 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 12.0 (TI
D 21, algo-1, executor 1, partition 1, PROCESS_LOCAL, 8325 bytes)
24/04/28 21:21:35 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 12.0 (TI
D 22, algo-1, executor 1, partition 2, PROCESS_LOCAL, 8318 bytes)
24/04/28 21:21:35 INFO storage.BlockManagerInfo: Added broadcast_12_piece0 in memory
on algo-1:40893 (size: 13.0 KB, free: 13.8 GB)
```

```
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:33 INFO memory.MemoryStore: Block broadcast_8_piece0 stored as bytes in memory (estimated size 27.6 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:33 INFO broadcast.TorrentBroadcast: Reading broadcast variable 8 took 15 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:33 INFO memory.MemoryStore: Block broadcast_8 stored as values in memory (estimated size 398.1 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:34 INFO executor.Executor: Finished task 2.0 in stage 9.0 (TID 18). 1851 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:34 INFO executor.Executor: Finished task 1.0 in stage 9.0 (TID 17). 1808 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:34 INFO executor.Executor: Finished task 0.0 in stage 9.0 (TID 16). 1851 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:34 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 19
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:34 INFO executor.Executor: Running task 0.0 in stage 11.0 (TID 19)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:34 INFO spark.MapOutputTrackerWorker: Updating epoch to 4 and clearing cache
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:34 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 10
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:34 INFO memory.MemoryStore: Block broadcast_10_piece0 stored as bytes in memory (estimated size 4.2 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:34 INFO broadcast.TorrentBroadcast: Reading broadcast variable 10 took 6 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:34 INFO memory.MemoryStore: Block broadcast_10 stored as values in memory (estimated size 7.6 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:34 INFO spark.MapOutputTrackerWorker: Don't have map outputs for shuffle 3, fetching them
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:34 INFO spark.MapOutputTrackerWorker: Doing the fetch; tracker endpoint = NettyRpcEndpointRef(spark://MapOutputTracker@10.0.222.169:37411)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:34 INFO spark.MapOutputTrackerWorker: Got the output locations
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:34 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:34 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
```



```
01_000002/stderr] 24/04/28 21:21:34 INFO codegen.CodeGenerator: Code generated in 9.459436 ms
24/04/28 21:21:35 INFO storage.BlockManagerInfo: Added broadcast_11_piece0 in memory on algo-1:40893 (size: 27.6 KB, free: 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:34 INFO executor.Executor: Finished task 0.0 in stage 11.0 (TID 19). 1937 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:35 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 20
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:35 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 21
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:35 INFO executor.Executor: Running task 0.0 in stage 12.0 (TID 20)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:35 INFO executor.Executor: Running task 1.0 in stage 12.0 (TID 21)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:35 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 22
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:35 INFO executor.Executor: Running task 2.0 in stage 12.0 (TID 22)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:35 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 12
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:35 INFO memory.MemoryStore: Block broadcast_12_piece0 stored as bytes in memory (estimated size 13.0 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:35 INFO broadcast.TorrentBroadcast: Reading broadcast variable 12 took 7 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:35 INFO memory.MemoryStore: Block broadcast_12 stored as values in memory (estimated size 27.7 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:35 INFO codegen.CodeGenerator: Code generated in 48.487013 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:35 INFO codegen.CodeGenerator: Code generated in 14.215755 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:35 INFO codegen.CodeGenerator: Code generated in 5.088343 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:35 INFO codegen.CodeGenerator: Code generated in 6.980986 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:35 INFO datasources.FileScanRDD: TID: 20 - Reading current file: path: s3a://sagemaker-us-east-1-211125778552/amazon-reviews-pds/tsv/amazon_reviews_us_Digital_Video_Games_v1_00.tsv.gz, range: 0-27442648, partition values: [empty row], isDataPresent: false
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:35 INFO broadcast.TorrentBroadcast: Started reading
```

```

broadcast variable 11
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:35 INFO datasources.FileScanRDD: TID: 22 - Reading current file: path: s3a://sagemaker-us-east-1-211125778552/amazon-reviews-pds/tsv/amazon_reviews_us_Gift_Card_v1_00.tsv.gz, range: 0-12134676, partition values: [empty row], isDataPresent: false
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:35 INFO datasources.FileScanRDD: TID: 21 - Reading current file: path: s3a://sagemaker-us-east-1-211125778552/amazon-reviews-pds/tsv/amazon_reviews_us_Digital_Software_v1_00.tsv.gz, range: 0-18997559, partition values: [empty row], isDataPresent: false
24/04/28 21:21:36 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 12.0 (TID 21) in 1526 ms on algo-1 (executor 1) (1/3)
24/04/28 21:21:36 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 12.0 (TID 22) in 1536 ms on algo-1 (executor 1) (2/3)
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 12.0 (TID 20) in 1877 ms on algo-1 (executor 1) (3/3)
24/04/28 21:21:37 INFO cluster.YarnScheduler: Removed TaskSet 12.0, whose tasks have all completed, from pool
24/04/28 21:21:37 INFO scheduler.DAGScheduler: ShuffleMapStage 12 (collect at AnalysisRunner.scala:523) finished in 1.886 s
24/04/28 21:21:37 INFO scheduler.DAGScheduler: looking for newly runnable stages
24/04/28 21:21:37 INFO scheduler.DAGScheduler: running: Set()
24/04/28 21:21:37 INFO scheduler.DAGScheduler: waiting: Set()
24/04/28 21:21:37 INFO scheduler.DAGScheduler: failed: Set()
24/04/28 21:21:37 INFO adaptive.CoalesceShufflePartitions: advisoryTargetPostShuffleInputSize: 67108864, targetPostShuffleInputSize 732517.
24/04/28 21:21:37 INFO codegen.CodeGenerator: Code generated in 24.452843 ms
24/04/28 21:21:37 INFO scheduler.DAGScheduler: Registering RDD 30 (collect at AnalysisRunner.scala:523) as input to shuffle 5
24/04/28 21:21:37 INFO scheduler.DAGScheduler: Got map stage job 9 (collect at AnalysisRunner.scala:523) with 26 output partitions
24/04/28 21:21:37 INFO scheduler.DAGScheduler: Final stage: ShuffleMapStage 14 (collect at AnalysisRunner.scala:523)
24/04/28 21:21:37 INFO scheduler.DAGScheduler: Parents of final stage: List(ShuffleMapStage 13)
24/04/28 21:21:37 INFO scheduler.DAGScheduler: Missing parents: List()
24/04/28 21:21:37 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 14 (MapPartitionsRDD[30] at collect at AnalysisRunner.scala:523), which has no missing parents
24/04/28 21:21:37 INFO memory.MemoryStore: Block broadcast_13 stored as values in memory (estimated size 30.4 KB, free 1007.8 MB)
24/04/28 21:21:37 INFO memory.MemoryStore: Block broadcast_13_piece0 stored as bytes in memory (estimated size 14.1 KB, free 1007.8 MB)
24/04/28 21:21:37 INFO storage.BlockManagerInfo: Added broadcast_13_piece0 in memory on 10.0.222.169:36237 (size: 14.1 KB, free: 1008.8 MB)
24/04/28 21:21:37 INFO spark.SparkContext: Created broadcast 13 from broadcast at DAGScheduler.scala:1203
24/04/28 21:21:37 INFO scheduler.DAGScheduler: Submitting 26 missing tasks from ShuffleMapStage 14 (MapPartitionsRDD[30] at collect at AnalysisRunner.scala:523) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14))
24/04/28 21:21:37 INFO cluster.YarnScheduler: Adding task set 14.0 with 26 tasks
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 14.0 (TID 23, algo-1, executor 1, partition 0, PROCESS_LOCAL, 7767 bytes)
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 14.0 (TID 24, algo-1, executor 1, partition 1, PROCESS_LOCAL, 7767 bytes)

```

```
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 14.0 (TID 25, algo-1, executor 1, partition 2, PROCESS_LOCAL, 7767 bytes)
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 14.0 (TID 26, algo-1, executor 1, partition 3, PROCESS_LOCAL, 7767 bytes)
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 14.0 (TID 27, algo-1, executor 1, partition 4, PROCESS_LOCAL, 7767 bytes)
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 14.0 (TID 28, algo-1, executor 1, partition 5, PROCESS_LOCAL, 7767 bytes)
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 14.0 (TID 29, algo-1, executor 1, partition 6, PROCESS_LOCAL, 7767 bytes)
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 14.0 (TID 30, algo-1, executor 1, partition 7, PROCESS_LOCAL, 7767 bytes)
24/04/28 21:21:37 INFO storage.BlockManagerInfo: Added broadcast_13_piece0 in memory on algo-1:40893 (size: 14.1 KB, free: 13.8 GB)
24/04/28 21:21:37 INFO spark.MapOutputTrackerMasterEndpoint: Asked to send map output locations for shuffle 4 to 10.0.222.169:41176
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:35 INFO memory.MemoryStore: Block broadcast_11_piece0 stored as bytes in memory (estimated size 27.6 KB, free 13.6 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:35 INFO broadcast.TorrentBroadcast: Reading broadcast variable 11 took 8 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:35 INFO memory.MemoryStore: Block broadcast_11 stored as values in memory (estimated size 398.1 KB, free 13.6 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:36 INFO executor.Executor: Finished task 1.0 in stage 12.0 (TID 21). 4470 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:36 INFO executor.Executor: Finished task 2.0 in stage 12.0 (TID 22). 4470 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 0.0 in stage 12.0 (TID 20). 4470 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 23
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 24
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 0.0 in stage 14.0 (TID 23)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 1.0 in stage 14.0 (TID 24)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 25
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 2.0 in stage 14.0 (TID 25)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO spark.MapOutputTrackerWorker: Updating epoch to 5 and clearing cache
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
```

```
01_000002/stderr] 24/04/28 21:21:37 INFO broadcast.TorrentBroadcast: Started reading
broadcast variable 13
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 26
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 3.0 in stag
e 14.0 (TID 26)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 27
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 4.0 in stag
e 14.0 (TID 27)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 28
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 29
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 30
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 6.0 in stag
e 14.0 (TID 29)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 7.0 in stag
e 14.0 (TID 30)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO memory.MemoryStore: Block broadcast_13_piec
e0 stored as bytes in memory (estimated size 14.1 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 5.0 in stag
e 14.0 (TID 28)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO broadcast.TorrentBroadcast: Reading broadca
st variable 13 took 11 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO memory.MemoryStore: Block broadcast_13 stor
ed as values in memory (estimated size 30.4 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO spark.MapOutputTrackerWorker: Don't have ma
p outputs for shuffle 4, fetching them
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO spark.MapOutputTrackerWorker: Doing the fet
ch; tracker endpoint = NettyRpcEndpointRef(spark://MapOutputTracker@10.0.222.169:374
11)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO spark.MapOutputTrackerWorker: Don't have ma
p outputs for shuffle 4, fetching them
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO spark.MapOutputTrackerWorker: Don't have ma
p outputs for shuffle 4, fetching them
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO spark.MapOutputTrackerWorker: Don't have ma
```

p outputs for shuffle 4, fetching them  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:37 INFO spark.MapOutputTrackerWorker: Got the output locations  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:37 INFO spark.MapOutputTrackerWorker: Don't have map outputs for shuffle 4, fetching them  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:37 INFO spark.MapOutputTrackerWorker: Don't have map outputs for shuffle 4, fetching them  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 6 ms  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 4 ms  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 14.0 (TID 31, algo-1, executor 1, partition 8, PROCESS\_LOCAL, 7767 bytes)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 14.0 (TID 27) in 152 ms on algo-1 (executor 1) (1/26)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 14.0 (TID 32, algo-1, executor 1, partition 9, PROCESS\_LOCAL, 7767 bytes)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 14.0 (TID 24) in 162 ms on algo-1 (executor 1) (2/26)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 14.0 (TID 33, algo-1, executor 1, partition 10, PROCESS\_LOCAL, 7767 bytes)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 14.0 (TID 26) in 171 ms on algo-1 (executor 1) (3/26)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 14.0 (TID 34, algo-1, executor 1, partition 11, PROCESS\_LOCAL, 7767 bytes)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 14.0 (TID 25) in 185 ms on algo-1 (executor 1) (4/26)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 12.0 in stage 14.0 (TID 35, algo-1, executor 1, partition 12, PROCESS\_LOCAL, 7767 bytes)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 13.0 in stage 14.0 (TID 36, algo-1, executor 1, partition 13, PROCESS\_LOCAL, 7767 bytes)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 14.0 (TID 29) in 218 ms on algo-1 (executor 1) (5/26)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 14.0 (TID 23) in 219 ms on algo-1 (executor 1) (6/26)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 14.0 in stage 14.0 (TID 37, algo-1, executor 1, partition 14, PROCESS\_LOCAL, 7767 bytes)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 14.0 (TID 28) in 221 ms on algo-1 (executor 1) (7/26)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 15.0 in stage 14.0 (TID 38, algo-1, executor 1, partition 15, PROCESS\_LOCAL, 7767 bytes)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 14.0 (TID 30) in 221 ms on algo-1 (executor 1) (8/26)

D 30) in 229 ms on algo-1 (executor 1) (8/26)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 16.0 in stage 14.0 (TID 39, algo-1, executor 1, partition 16, PROCESS\_LOCAL, 7767 bytes)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 8.0 in stage 14.0 (TID 31) in 101 ms on algo-1 (executor 1) (9/26)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 17.0 in stage 14.0 (TID 40, algo-1, executor 1, partition 17, PROCESS\_LOCAL, 7767 bytes)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 12.0 in stage 14.0 (TID 35) in 56 ms on algo-1 (executor 1) (10/26)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 18.0 in stage 14.0 (TID 41, algo-1, executor 1, partition 18, PROCESS\_LOCAL, 7767 bytes)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 19.0 in stage 14.0 (TID 42, algo-1, executor 1, partition 19, PROCESS\_LOCAL, 7767 bytes)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 20.0 in stage 14.0 (TID 43, algo-1, executor 1, partition 20, PROCESS\_LOCAL, 7767 bytes)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 9.0 in stage 14.0 (TID 32) in 121 ms on algo-1 (executor 1) (11/26)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 11.0 in stage 14.0 (TID 34) in 97 ms on algo-1 (executor 1) (12/26)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 10.0 in stage 14.0 (TID 33) in 115 ms on algo-1 (executor 1) (13/26)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 21.0 in stage 14.0 (TID 44, algo-1, executor 1, partition 21, PROCESS\_LOCAL, 7767 bytes)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 13.0 in stage 14.0 (TID 36) in 85 ms on algo-1 (executor 1) (14/26)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 22.0 in stage 14.0 (TID 45, algo-1, executor 1, partition 22, PROCESS\_LOCAL, 7767 bytes)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 14.0 in stage 14.0 (TID 37) in 97 ms on algo-1 (executor 1) (15/26)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 23.0 in stage 14.0 (TID 46, algo-1, executor 1, partition 23, PROCESS\_LOCAL, 7767 bytes)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 15.0 in stage 14.0 (TID 38) in 91 ms on algo-1 (executor 1) (16/26)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 24.0 in stage 14.0 (TID 47, algo-1, executor 1, partition 24, PROCESS\_LOCAL, 7767 bytes)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 16.0 in stage 14.0 (TID 39) in 96 ms on algo-1 (executor 1) (17/26)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 25.0 in stage 14.0 (TID 48, algo-1, executor 1, partition 25, PROCESS\_LOCAL, 7767 bytes)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 19.0 in stage 14.0 (TID 42) in 67 ms on algo-1 (executor 1) (18/26)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 20.0 in stage 14.0 (TID 43) in 89 ms on algo-1 (executor 1) (19/26)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 21.0 in stage 14.0 (TID 44) in 70 ms on algo-1 (executor 1) (20/26)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 17.0 in stage 14.0 (TID 40) in 105 ms on algo-1 (executor 1) (21/26)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 25.0 in stage 14.0 (TID 48) in 31 ms on algo-1 (executor 1) (22/26)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 18.0 in stage 14.0 (TID 41) in 99 ms on algo-1 (executor 1) (23/26)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 24.0 in stage 14.0 (TID 47) in 36 ms on algo-1 (executor 1) (24/26)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 23.0 in stage 14.0 (TID 46) in 56 ms on algo-1 (executor 1) (25/26)  
24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 22.0 in stage 14.0 (T

ID 45) in 64 ms on algo-1 (executor 1) (26/26)

24/04/28 21:21:37 INFO cluster.YarnScheduler: Removed TaskSet 14.0, whose tasks have all completed, from pool

24/04/28 21:21:37 INFO scheduler.DAGScheduler: ShuffleMapStage 14 (collect at AnalysisRunner.scala:523) finished in 0.387 s

24/04/28 21:21:37 INFO scheduler.DAGScheduler: looking for newly runnable stages

24/04/28 21:21:37 INFO scheduler.DAGScheduler: running: Set()

24/04/28 21:21:37 INFO scheduler.DAGScheduler: waiting: Set()

24/04/28 21:21:37 INFO scheduler.DAGScheduler: failed: Set()

24/04/28 21:21:37 INFO adaptive.CoalesceShufflePartitions: advisoryTargetPostShuffleInputSize: 67108864, targetPostShuffleInputSize 22.

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 416

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 294

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 233

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 354

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 245

24/04/28 21:21:37 INFO codegen.CodeGenerator: Code generated in 26.816474 ms

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 410

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 399

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 247

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 231

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 222

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 408

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 332

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 260

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 232

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 255

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 297

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 246

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 411

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 269

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 270

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 264

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 393

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 362

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 290

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 236

24/04/28 21:21:37 INFO storage.BlockManagerInfo: Removed broadcast\_10\_piece0 on 10.0.222.169:36237 in memory (size: 4.2 KB, free: 1008.8 MB)

24/04/28 21:21:37 INFO storage.BlockManagerInfo: Removed broadcast\_10\_piece0 on algo-1:40893 in memory (size: 4.2 KB, free: 13.8 GB)

24/04/28 21:21:37 INFO spark.SparkContext: Starting job: collect at AnalysisRunner.scala:523

24/04/28 21:21:37 INFO scheduler.DAGScheduler: Got job 10 (collect at AnalysisRunner.scala:523) with 1 output partitions

24/04/28 21:21:37 INFO scheduler.DAGScheduler: Final stage: ResultStage 17 (collect at AnalysisRunner.scala:523)

24/04/28 21:21:37 INFO scheduler.DAGScheduler: Parents of final stage: List(ShuffleMapStage 16)

24/04/28 21:21:37 INFO scheduler.DAGScheduler: Missing parents: List()

24/04/28 21:21:37 INFO scheduler.DAGScheduler: Submitting ResultStage 17 (MapPartitionsRDD[33] at collect at AnalysisRunner.scala:523), which has no missing parents

24/04/28 21:21:37 INFO memory.MemoryStore: Block broadcast\_14 stored as values in memory (estimated size 8.7 KB, free 1007.8 MB)

24/04/28 21:21:37 INFO memory.MemoryStore: Block broadcast\_14\_piece0 stored as bytes in memory (estimated size 4.7 KB, free 1007.8 MB)

24/04/28 21:21:37 INFO storage.BlockManagerInfo: Added broadcast\_14\_piece0 in memory on 10.0.222.169:36237 (size: 4.7 KB, free: 1008.8 MB)

24/04/28 21:21:37 INFO spark.SparkContext: Created broadcast 14 from broadcast at DAGScheduler.scala:1203

24/04/28 21:21:37 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 17 (MapPartitionsRDD[33] at collect at AnalysisRunner.scala:523) (first 15 tasks are for partitions Vector(0))

24/04/28 21:21:37 INFO cluster.YarnScheduler: Adding task set 17.0 with 1 tasks

24/04/28 21:21:37 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 17.0 (TID 49, algo-1, executor 1, partition 0, NODE\_LOCAL, 7778 bytes)

24/04/28 21:21:37 INFO storage.BlockManagerInfo: Added broadcast\_14\_piece0 in memory on algo-1:40893 (size: 4.7 KB, free: 13.8 GB)

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 239

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 262

24/04/28 21:21:37 INFO spark.MapOutputTrackerMasterEndpoint: Asked to send map output locations for shuffle 5 to 10.0.222.169:41176

24/04/28 21:21:37 INFO storage.BlockManagerInfo: Removed broadcast\_13\_piece0 on algo-1:40893 in memory (size: 14.1 KB, free: 13.8 GB)

24/04/28 21:21:37 INFO storage.BlockManagerInfo: Removed broadcast\_13\_piece0 on 10.0.222.169:36237 in memory (size: 14.1 KB, free: 1008.8 MB)

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 361

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 252

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 271

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 355

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 291

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 403

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 366

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 248

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 333

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 219

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 241

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 220

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 370

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 229

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 285

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 234

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 404

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 396

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 369

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 253

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 413

24/04/28 21:21:37 INFO storage.BlockManagerInfo: Removed broadcast\_8\_piece0 on 10.0.222.169:36237 in memory (size: 27.6 KB, free: 1008.8 MB)

24/04/28 21:21:37 INFO storage.BlockManagerInfo: Removed broadcast\_8\_piece0 on algo-1:40893 in memory (size: 27.6 KB, free: 13.8 GB)

24/04/28 21:21:37 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 17.0 (TID 49) in 45 ms on algo-1 (executor 1) (1/1)

24/04/28 21:21:37 INFO cluster.YarnScheduler: Removed TaskSet 17.0, whose tasks have all completed, from pool

24/04/28 21:21:37 INFO scheduler.DAGScheduler: ResultStage 17 (collect at AnalysisRunner.scala:523) finished in 0.049 s

24/04/28 21:21:37 INFO scheduler.DAGScheduler: Job 10 finished: collect at AnalysisRunner.scala:523, took 0.054421 s

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 218

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 273

24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 265



```
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 256
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 276
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 251
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 334
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 357
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 223
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 295
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 409
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 360
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 339
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 267
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 266
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 250
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 330
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 352
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 228
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 272
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 275
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 286
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 351
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 338
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 296
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 359
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 224
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 282
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 288
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 261
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 375
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 367
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 394
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 259
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 258
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 401
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 289
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 348
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 412
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 287
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 225
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 358
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 300
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 283
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 284
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 415
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 400
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 240
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 280
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 242
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 353
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 347
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 397
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 235
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 274
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 278
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 349
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned shuffle 3
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 363
```

```
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 405
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 372
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 230
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 331
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 335
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 279
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 277
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 371
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 337
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 244
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 336
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 263
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 395
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 227
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 292
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 243
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 365
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 293
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 350
24/04/28 21:21:37 INFO storage.BlockManagerInfo: Removed broadcast_12_piece0 on 10.
0.222.169:36237 in memory (size: 13.0 KB, free: 1008.8 MB)
24/04/28 21:21:37 INFO storage.BlockManagerInfo: Removed broadcast_12_piece0 on algo
-1:40893 in memory (size: 13.0 KB, free: 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Gettin
g 3 non-empty blocks including 3 local blocVerification Run Status: Success
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 368
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 257
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 417
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 329
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 254
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 281
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 356
24/04/28 21:21:37 INFO storage.BlockManagerInfo: Removed broadcast_9_piece0 on 10.0.
222.169:36237 in memory (size: 7.4 KB, free: 1008.8 MB)
24/04/28 21:21:37 INFO storage.BlockManagerInfo: Removed broadcast_9_piece0 on algo-
1:40893 in memory (size: 7.4 KB, free: 13.8 GB)
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 249
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 364
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 268
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 346
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 406
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 237
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 238
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 398
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 221
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 226
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 402
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 414
24/04/28 21:21:37 INFO spark.ContextCleaner: Cleaned accumulator 407
24/04/28 21:21:37 INFO codegen.CodeGenerator: Code generated in 29.563977 ms
24/04/28 21:21:37 INFO codegen.CodeGenerator: Code generated in 8.248387 ms
24/04/28 21:21:37 INFO codegen.CodeGenerator: Code generated in 6.995667 ms
+-----+-----+-----+-----+
-----
-----+-----+-----+-----+
```

check	check_level	check_status	constraint
constraint_status	constraint_message		
Review Check	Error	Success	SizeConstraint(Size(None))
Success			
Review Check	Error	Success	MinimumConstraint(Minimum(star_rating,None))
Success			
Review Check	Error	Success	MaximumConstraint(Maximum(star_rating,None))
Success			
Review Check	Error	Success	CompletenessConstraint(Completeness(review_id,None))
Success			
Review Check	Error	Success	UniquenessConstraint(Uniqueness(List(review_id),None))
Success			
Review Check	Error	Success	CompletenessConstraint(Completeness(marketplace,None))
Success			
Review Check	Error	Success	ComplianceConstraint(Compliance(marketplace contained in US,UK,DE,JP,FR,`marketplace` IS NULL OR `marketplace` IN ('US','UK','DE','JP','FR'),None))
Success			

24/04/28 21:21:37 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2

24/04/28 21:21:37 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup \_temporary folders under output directory:false, ignore cleanup failures: false

24/04/28 21:21:37 INFO output.DirectFileOutputCommitter: Direct Write: DISABLED

24/04/28 21:21:37 INFO datasources.SQLConfCommitterProvider: Using output committer class org.apache.hadoop.mapreduce.lib.output.DirectFileOutputCommitter

ks and 0 remote blocks

[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 5 ms

[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks

[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 6 ms

[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 6 ms

[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks

[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 11 ms

[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks

[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 11 ms

```
01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Gettin
g 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Starte
d 0 remote fetches in 12 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Starte
d 0 remote fetches in 10 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO codegen.CodeGenerator: Code generated in 2
0.482514 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 4.0 in sta
ge 14.0 (TID 27). 3745 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 31
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 8.0 in stag
e 14.0 (TID 31)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 1.0 in sta
ge 14.0 (TID 24). 3788 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 3.0 in sta
ge 14.0 (TID 26). 3745 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Gettin
g 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Starte
d 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 32
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 9.0 in stag
e 14.0 (TID 32)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 33
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 10.0 in sta
ge 14.0 (TID 33)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Gettin
g 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Starte
d 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 2.0 in sta
ge 14.0 (TID 25). 3745 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Gettin
g 3 non-empty blocks including 3 local blocks and 0 remote blocks
```

```
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 34
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 11.0 in stage 14.0 (TID 34)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 0.0 in stage 14.0 (TID 23). 3745 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 6.0 in stage 14.0 (TID 29). 3745 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 5.0 in stage 14.0 (TID 28). 3745 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 35
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 12.0 in stage 14.0 (TID 35)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 36
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 13.0 in stage 14.0 (TID 36)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 37
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 7.0 in stage 14.0 (TID 30). 3745 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 14.0 in stage 14.0 (TID 37)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 8.0 in stage 14.0 (TID 31). 3788 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 4 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Getting
```

```
g 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 38
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 39
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 15.0 in stage 14.0 (TID 38)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 16.0 in stage 14.0 (TID 39)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 4 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 12.0 in stage 14.0 (TID 35). 3745 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 9.0 in stage 14.0 (TID 32). 3745 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 11.0 in stage 14.0 (TID 34). 3745 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 12 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 10.0 in stage 14.0 (TID 33). 3745 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 40
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 41
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 18.0 in stage 14.0 (TID 41)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
```

```
01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 17.0 in stage 14.0 (TID 40)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 42
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 43
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 19.0 in stage 14.0 (TID 42)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 20.0 in stage 14.0 (TID 43)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 13.0 in stage 14.0 (TID 36). 3745 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 44
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 21.0 in stage 14.0 (TID 44)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 14.0 in stage 14.0 (TID 37). 3745 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
```

```
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 45
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 22.0 in stage 14.0 (TID 45)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 15.0 in stage 14.0 (TID 38). 3745 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 46
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 16.0 in stage 14.0 (TID 39). 3745 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 23.0 in stage 14.0 (TID 46)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 47
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 24.0 in stage 14.0 (TID 47)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 19.0 in stage 14.0 (TID 42). 3745 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 48
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 20.0 in stage 14.0 (TID 43). 3745 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 25.0 in stage 14.0 (TID 48)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 21.0 in st
```



```
age 14.0 (TID 44). 3745 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 18.0 in stage 14.0 (TID 41). 3745 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 25.0 in stage 14.0 (TID 48). 3745 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 17.0 in stage 14.0 (TID 40). 3745 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 23.0 in stage 14.0 (TID 46). 3745 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 24.0 in stage 14.0 (TID 47). 3745 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 22.0 in stage 14.0 (TID 45). 3745 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 49
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Running task 0.0 in stage 17.0 (TID 49)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO spark.MapOutputTrackerWorker: Updating epoch to 6 and clearing cache
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 14
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO memory.MemoryStore: Block broadcast_14_piece0 stored as bytes in memory (estimated size 4.7 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO broadcast.TorrentBroadcast: Reading broadcast variable 14 took 8 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO memory.MemoryStore: Block broadcast_14 stored as values in memory (estimated size 8.7 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO spark.MapOutputTrackerWorker: Don't have map outputs for shuffle 5, fetching them
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO spark.MapOutputTrackerWorker: Doing the fetch; tracker endpoint = NettyRpcEndpointRef(spark://MapOutputTracker@10.0.222.169:37411)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:37 INFO spark.MapOutputTrackerWorker: Got the output locations
```

```
24/04/28 21:21:38 INFO scheduler.DAGScheduler: Registering RDD 36 (save at NativeMethodAccessorImpl.java:0) as input to shuffle 6
24/04/28 21:21:38 INFO scheduler.DAGScheduler: Got map stage job 11 (save at NativeMethodAccessorImpl.java:0) with 7 output partitions
24/04/28 21:21:38 INFO scheduler.DAGScheduler: Final stage: ShuffleMapStage 18 (save at NativeMethodAccessorImpl.java:0)
24/04/28 21:21:38 INFO scheduler.DAGScheduler: Parents of final stage: List()
24/04/28 21:21:38 INFO scheduler.DAGScheduler: Missing parents: List()
24/04/28 21:21:38 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 18 (MapPartitionsRDD[36] at save at NativeMethodAccessorImpl.java:0), which has no missing parents
24/04/28 21:21:38 INFO memory.MemoryStore: Block broadcast_15 stored as values in memory (estimated size 5.5 KB, free 1008.2 MB)
24/04/28 21:21:38 INFO memory.MemoryStore: Block broadcast_15_piece0 stored as bytes in memory (estimated size 3.3 KB, free 1008.2 MB)
24/04/28 21:21:38 INFO storage.BlockManagerInfo: Added broadcast_15_piece0 in memory on 10.0.222.169:36237 (size: 3.3 KB, free: 1008.8 MB)
24/04/28 21:21:38 INFO spark.SparkContext: Created broadcast 15 from broadcast at DAGScheduler.scala:1203
24/04/28 21:21:38 INFO scheduler.DAGScheduler: Submitting 7 missing tasks from ShuffleMapStage 18 (MapPartitionsRDD[36] at save at NativeMethodAccessorImpl.java:0) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6))
24/04/28 21:21:38 INFO cluster.YarnScheduler: Adding task set 18.0 with 7 tasks
24/04/28 21:21:38 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 18.0 (TID 50, algo-1, executor 1, partition 0, PROCESS_LOCAL, 8156 bytes)
24/04/28 21:21:38 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 18.0 (TID 51, algo-1, executor 1, partition 1, PROCESS_LOCAL, 8172 bytes)
24/04/28 21:21:38 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 18.0 (TID 52, algo-1, executor 1, partition 2, PROCESS_LOCAL, 8172 bytes)
24/04/28 21:21:38 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 18.0 (TID 53, algo-1, executor 1, partition 3, PROCESS_LOCAL, 8180 bytes)
24/04/28 21:21:38 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 18.0 (TID 54, algo-1, executor 1, partition 4, PROCESS_LOCAL, 8180 bytes)
24/04/28 21:21:38 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 18.0 (TID 55, algo-1, executor 1, partition 5, PROCESS_LOCAL, 8180 bytes)
24/04/28 21:21:38 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 18.0 (TID 56, algo-1, executor 1, partition 6, PROCESS_LOCAL, 8279 bytes)
24/04/28 21:21:38 INFO storage.BlockManagerInfo: Added broadcast_15_piece0 in memory on algo-1:40893 (size: 3.3 KB, free: 13.8 GB)
24/04/28 21:21:38 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 18.0 (TID 55) in 32 ms on algo-1 (executor 1) (1/7)
24/04/28 21:21:38 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 18.0 (TID 51) in 36 ms on algo-1 (executor 1) (2/7)
24/04/28 21:21:38 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 18.0 (TID 54) in 36 ms on algo-1 (executor 1) (3/7)
24/04/28 21:21:38 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 18.0 (TID 52) in 37 ms on algo-1 (executor 1) (4/7)
24/04/28 21:21:38 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 18.0 (TID 56) in 38 ms on algo-1 (executor 1) (5/7)
24/04/28 21:21:38 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 18.0 (TID 50) in 39 ms on algo-1 (executor 1) (6/7)
24/04/28 21:21:38 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 18.0 (TID 53) in 38 ms on algo-1 (executor 1) (7/7)
24/04/28 21:21:38 INFO cluster.YarnScheduler: Removed TaskSet 18.0, whose tasks have all completed, from pool
24/04/28 21:21:38 INFO scheduler.DAGScheduler: ShuffleMapStage 18 (save at NativeMet
```

```

hodAccessorImpl.java:0) finished in 0.046 s
24/04/28 21:21:38 INFO scheduler.DAGScheduler: looking for newly runnable stages
24/04/28 21:21:38 INFO scheduler.DAGScheduler: running: Set()
24/04/28 21:21:38 INFO scheduler.DAGScheduler: waiting: Set()
24/04/28 21:21:38 INFO scheduler.DAGScheduler: failed: Set()
24/04/28 21:21:38 INFO spark.SparkContext: Starting job: save at NativeMethodAccesso
rImpl.java:0
24/04/28 21:21:38 INFO scheduler.DAGScheduler: Got job 12 (save at NativeMethodAcces
sorImpl.java:0) with 1 output partitions
24/04/28 21:21:38 INFO scheduler.DAGScheduler: Final stage: ResultStage 20 (save at
NativeMethodAccessorImpl.java:0)
24/04/28 21:21:38 INFO scheduler.DAGScheduler: Parents of final stage: List(ShuffleM
apStage 19)
24/04/28 21:21:38 INFO scheduler.DAGScheduler: Missing parents: List()
24/04/28 21:21:38 INFO scheduler.DAGScheduler: Submitting ResultStage 20 (ShuffledRo
wRDD[37] at save at NativeMethodAccessorImpl.java:0), which has no missing parents
24/04/28 21:21:38 INFO memory.MemoryStore: Block broadcast_16 stored as values in me
mory (estimated size 167.6 KB, free 1008.1 MB)
24/04/28 21:21:38 INFO memory.MemoryStore: Block broadcast_16_piece0 stored as bytes
in memory (estimated size 60.4 KB, free 1008.0 MB)
24/04/28 21:21:38 INFO storage.BlockManagerInfo: Added broadcast_16_piece0 in memory
on 10.0.222.169:36237 (size: 60.4 KB, free: 1008.8 MB)
24/04/28 21:21:38 INFO spark.SparkContext: Created broadcast 16 from broadcast at DA
GScheduler.scala:1203
24/04/28 21:21:38 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Resul
tStage 20 (ShuffledRowRDD[37] at save at NativeMethodAccessorImpl.java:0) (first 15
tasks are for partitions Vector())
24/04/28 21:21:38 INFO cluster.YarnScheduler: Adding task set 20.0 with 1 tasks
24/04/28 21:21:38 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 20.0 (TI
D 57, algo-1, executor 1, partition 0, NODE_LOCAL, 7778 bytes)
24/04/28 21:21:38 INFO storage.BlockManagerInfo: Added broadcast_16_piece0 in memory
on algo-1:40893 (size: 60.4 KB, free: 13.8 GB)
24/04/28 21:21:38 INFO spark.MapOutputTrackerMasterEndpoint: Asked to send map outpu
t locations for shuffle 6 to 10.0.222.169:41176
24/04/28 21:21:39 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 20.0 (TI
D 57) in 849 ms on algo-1 (executor 1) (1/1)
24/04/28 21:21:39 INFO cluster.YarnScheduler: Removed TaskSet 20.0, whose tasks have
all completed, from pool
24/04/28 21:21:39 INFO scheduler.DAGScheduler: ResultStage 20 (save at NativeMethodA
ccessorImpl.java:0) finished in 0.876 s
24/04/28 21:21:39 INFO scheduler.DAGScheduler: Job 12 finished: save at NativeMethod
AccessorImpl.java:0, took 0.880282 s
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Gettin
g 26 non-empty blocks including 26 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO storage.ShuffleBlockFetcherIterator: Starte
d 0 remote fetches in 1 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO codegen.CodeGenerator: Code generated in 1
1.986798 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:37 INFO executor.Executor: Finished task 0.0 in sta
ge 17.0 (TID 49). 1931 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:38 INFO executor.CoarseGrainedExecutorBackend: Got

```

```
assigned task 50
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO executor.Executor: Running task 0.0 in stage 18.0 (TID 50)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 51
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 52
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO executor.Executor: Running task 1.0 in stage 18.0 (TID 51)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO executor.Executor: Running task 2.0 in stage 18.0 (TID 52)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 53
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 54
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO executor.Executor: Running task 3.0 in stage 18.0 (TID 53)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO executor.Executor: Running task 4.0 in stage 18.0 (TID 54)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 15
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 55
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 56
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO executor.Executor: Running task 6.0 in stage 18.0 (TID 56)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO executor.Executor: Running task 5.0 in stage 18.0 (TID 55)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO memory.MemoryStore: Block broadcast_15_piece0 stored as bytes in memory (estimated size 3.3 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO broadcast.TorrentBroadcast: Reading broadcast variable 15 took 22 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO memory.MemoryStore: Block broadcast_15 stored as values in memory (estimated size 5.5 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO executor.Executor: Finished task 5.0 in stage 18.0 (TID 55). 1384 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
```

```
01_000002/stderr] 24/04/28 21:21:38 INFO executor.Executor: Finished task 4.0 in stage 18.0 (TID 54). 1384 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO executor.Executor: Finished task 1.0 in stage 18.0 (TID 51). 1384 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO executor.Executor: Finished task 2.0 in stage 18.0 (TID 52). 1384 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO executor.Executor: Finished task 3.0 in stage 18.0 (TID 53). 1384 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO executor.Executor: Finished task 0.0 in stage 18.0 (TID 50). 1384 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO executor.Executor: Finished task 6.0 in stage 18.0 (TID 56). 1340 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 57
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO executor.Executor: Running task 0.0 in stage 20.0 (TID 57)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO spark.MapOutputTrackerWorker: Updating epoch to 7 and clearing cache
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 16
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO memory.MemoryStore: Block broadcast_16_piece0 stored as bytes in memory (estimated size 60.4 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO broadcast.TorrentBroadcast: Reading broadcast variable 16 took 10 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO memory.MemoryStore: Block broadcast_16 stored as values in memory (estimated size 167.6 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO spark.MapOutputTrackerWorker: Don't have map outputs for shuffle 6, fetching them
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO spark.MapOutputTrackerWorker: Doing the fetch; tracker endpoint = NettyRpcEndpointRef(spark://MapOutputTracker@10.0.222.169:37411)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO spark.MapOutputTrackerWorker: Got the output locations
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO storage.ShuffleBlockFetcherIterator: Getting 7 non-empty blocks including 7 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
24/04/28 21:21:39 INFO datasources.FileFormatWriter: Write Job 98e47c95-25ca-448d-bb6d-62b84081a6c2 committed.
```

```
24/04/28 21:21:39 INFO datasources.FileFormatWriter: Finished processing stats for write job 98e47c95-25ca-448d-bb6d-62b84081a6c2.
```

```
[/var/log/yarn/userlogs/applica+-----+-----+-----+-----+-----
```

-----+-----+

entity	instance	name	value
Column	review_id	Completeness	1.0
Column	review_id	Uniqueness	1.0
Dataset	*	Size	396601.0
Column	star_rating	Maximum	5.0
Column	star_rating	Minimum	1.0
Column	marketplace contained in US,UK,DE,JP,FR	Compliance	1.0
Column	marketplace	Completeness	1.0

```
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 506
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 455
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 313
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 373
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 490
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 459
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 441
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 303
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 321
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 342
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 498
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 298
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 493
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 327
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 443
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 420
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 462
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 505
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 432
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 500
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 471
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 485
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 423
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 344
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 453
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 419
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 384
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 318
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 299
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 389
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 317
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 454
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 343
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 324
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 440
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 380
24/04/28 21:21:39 INFO storage.BlockManagerInfo: Removed broadcast_11_piece0 on algo
-1:40893 in memory (size: 27.6 KB, free: 13.8 GB)
24/04/28 21:21:39 INFO storage.BlockManagerInfo: Removed broadcast_11_piece0 on 10.
0.222.169:36237 in memory (size: 27.6 KB, free: 1008.8 MB)
24/04/28 21:21:39 INFO storage.BlockManagerInfo: Removed broadcast_16_piece0 on 10.
0.222.169:36237 in memory (size: 60.4 KB, free: 1008.9 MB)
```

24/04/28 21:21:39 INFO storage.BlockManagerInfo: Removed broadcast\_16\_piece0 on algo-1:40893 in memory (size: 60.4 KB, free: 13.8 GB)

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 451

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 469

24/04/28 21:21:39 INFO storage.BlockManagerInfo: Removed broadcast\_14\_piece0 on 10.0.222.169:36237 in memory (size: 4.7 KB, free: 1008.9 MB)

24/04/28 21:21:39 INFO storage.BlockManagerInfo: Removed broadcast\_14\_piece0 on algo-1:40893 in memory (size: 4.7 KB, free: 13.8 GB)

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 437

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 479

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 478

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 326

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 487

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 509

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 434

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 320

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 495

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 481

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 467

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 464

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 429

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 308

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 436

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 508

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 483

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 430

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 302

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 458

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 319

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 456

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 463

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 501

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 307

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 433

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 391

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 496

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 390

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 438

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 312

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 473

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 323

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 499

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 494

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 435

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 427

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 460

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 465

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 382

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 466

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 426

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 480

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 449

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 446

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 421

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 315

24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 477

```
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 304
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 491
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 470
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 385
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 448
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 474
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 345
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 316
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 502
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 482
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 425
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 422
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 445
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 507
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned shuffle 5
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 468
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 309
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 424
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 497
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 489
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 311
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 418
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned shuffle 6
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 378
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 452
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 484
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 431
24/04/28 21:21:39 INFO storage.BlockManagerInfo: Removed broadcast_15_piece0 on 10.
0.222.169:36237 in memory (size: 3.3 KB, free: 1008.9 MB)
24/04/28 21:21:39 INFO storage.BlockManagerInfo: Removed broadcast_15_piece0 on algo
-1:40893 in memory (size: 3.3 KB, free: 13.8 GB)
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 325
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 341
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 486
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 386
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 488
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 450
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 461
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 457
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 442
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 340
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 492
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 379
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 328
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 428
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 381
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 439
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 376
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 383
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned shuffle 4
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 388
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 392
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 476
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 387
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 314
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 472
```



```
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 306
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 377
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 374
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 310
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 504
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 503
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 447
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 322
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 301
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 444
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 475
24/04/28 21:21:39 INFO spark.ContextCleaner: Cleaned accumulator 305
24/04/28 21:21:39 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
24/04/28 21:21:39 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
24/04/28 21:21:39 INFO output.DirectFileOutputCommitter: Direct Write: DISABLED
24/04/28 21:21:39 INFO datasources.SQLConfCommitterProvider: Using output committer class org.apache.hadoop.mapreduce.lib.output.DirectFileOutputCommitter
24/04/28 21:21:40 INFO scheduler.DAGScheduler: Registering RDD 42 (save at NativeMethodAccessorImpl.java:0) as input to shuffle 7
24/04/28 21:21:40 INFO scheduler.DAGScheduler: Got map stage job 13 (save at NativeMethodAccessorImpl.java:0) with 7 output partitions
24/04/28 21:21:40 INFO scheduler.DAGScheduler: Final stage: ShuffleMapStage 21 (save at NativeMethodAccessorImpl.java:0)
24/04/28 21:21:40 INFO scheduler.DAGScheduler: Parents of final stage: List()
24/04/28 21:21:40 INFO scheduler.DAGScheduler: Missing parents: List()
24/04/28 21:21:40 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 21 (MapPartitionsRDD[42] at save at NativeMethodAccessorImpl.java:0), which has no missing parents
24/04/28 21:21:40 INFO memory.MemoryStore: Block broadcast_17 stored as values in memory (estimated size 5.4 KB, free 1008.6 MB)
24/04/28 21:21:40 INFO memory.MemoryStore: Block broadcast_17_piece0 stored as bytes in memory (estimated size 3.3 KB, free 1008.6 MB)
24/04/28 21:21:40 INFO storage.BlockManagerInfo: Added broadcast_17_piece0 in memory on 10.0.222.169:36237 (size: 3.3 KB, free: 1008.9 MB)
24/04/28 21:21:40 INFO spark.SparkContext: Created broadcast 17 from broadcast at DAGScheduler.scala:1203
24/04/28 21:21:40 INFO scheduler.DAGScheduler: Submitting 7 missing tasks from ShuffleMapStage 21 (MapPartitionsRDD[42] at save at NativeMethodAccessorImpl.java:0) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6))
24/04/28 21:21:40 INFO cluster.YarnScheduler: Adding task set 21.0 with 7 tasks
24/04/28 21:21:40 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 21.0 (TID 58, algo-1, executor 1, partition 0, PROCESS_LOCAL, 8108 bytes)
24/04/28 21:21:40 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 21.0 (TID 59, algo-1, executor 1, partition 1, PROCESS_LOCAL, 8108 bytes)
24/04/28 21:21:40 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 21.0 (TID 60, algo-1, executor 1, partition 2, PROCESS_LOCAL, 8092 bytes)
24/04/28 21:21:40 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 21.0 (TID 61, algo-1, executor 1, partition 3, PROCESS_LOCAL, 8100 bytes)
24/04/28 21:21:40 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 21.0 (TID 62, algo-1, executor 1, partition 4, PROCESS_LOCAL, 8100 bytes)
24/04/28 21:21:40 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 21.0 (TID 63, algo-1, executor 1, partition 5, PROCESS_LOCAL, 8132 bytes)
24/04/28 21:21:40 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 21.0 (TID 64, algo-1, executor 1, partition 6, PROCESS_LOCAL, 8108 bytes)
```

```

24/04/28 21:21:40 INFO storage.BlockManagerInfo: Added broadcast_17_piece0 in memory
on algo-1:40893 (size: 3.3 KB, free: 13.8 GB)
24/04/28 21:21:40 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 21.0 (TI
D 64) in 17 ms on algo-1 (executor 1) (1/7)
24/04/28 21:21:40 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 21.0 (TI
D 60) in 20 ms on algo-1 (executor 1) (2/7)
24/04/28 21:21:40 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 21.0 (TI
D 58) in 21 ms on algo-1 (executor 1) (3/7)
24/04/28 21:21:40 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 21.0 (TI
D 61) in 20 ms on algo-1 (executor 1) (4/7)
24/04/28 21:21:40 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 21.0 (TI
D 59) in 30 ms on algo-1 (executor 1) (5/7)
24/04/28 21:21:40 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 21.0 (TI
D 62) in 29 ms on algo-1 (executor 1) (6/7)
24/04/28 21:21:40 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 21.0 (TI
D 63) in 29 ms on algo-1 (executor 1) (7/7)
24/04/28 21:21:40 INFO cluster.YarnScheduler: Removed TaskSet 21.0, whose tasks have
all completed, from pool
24/04/28 21:21:40 INFO scheduler.DAGScheduler: ShuffleMapStage 21 (save at NativeMet
hodAccessorImpl.java:0) finished in 0.036 s
24/04/28 21:21:40 INFO scheduler.DAGScheduler: looking for newly runnable stages
24/04/28 21:21:40 INFO scheduler.DAGScheduler: running: Set()
24/04/28 21:21:40 INFO scheduler.DAGScheduler: waiting: Set()
24/04/28 21:21:40 INFO scheduler.DAGScheduler: failed: Set()
24/04/28 21:21:40 INFO spark.SparkContext: Starting job: save at NativeMethodAccesso
rImpl.java:0
24/04/28 21:21:40 INFO scheduler.DAGScheduler: Got job 14 (save at NativeMethodAcces
sorImpl.java:0) with 1 output partitions
24/04/28 21:21:40 INFO scheduler.DAGScheduler: Final stage: ResultStage 23 (save at
NativeMethodAccessorImpl.java:0)
24/04/28 21:21:40 INFO scheduler.DAGScheduler: Parents of final stage: List(ShuffleM
apStage 22)
24/04/28 21:21:40 INFO scheduler.DAGScheduler: Missing parents: List()
24/04/28 21:21:40 INFO scheduler.DAGScheduler: Submitting ResultStage 23 (ShuffledRo
wRDD[43] at save at NativeMethodAccessorImpl.java:0), which has no missing parents
24/04/28 21:21:40 INFO memory.MemoryStore: Block broadcast_18 stored as values in me
memory (estimated size 167.5 KB, free 1008.4 MB)
24/04/28 21:21:40 INFO memory.MemoryStore: Block broadcast_18_piece0 stored as bytes
in memory (estimated size 60.3 KB, free 1008.3 MB)
24/04/28 21:21:40 INFO storage.BlockManagerInfo: Added broadcast_18_piece0 in memory
on 10.0.222.169:36237 (size: 60.3 KB, free: 1008.8 MB)
24/04/28 21:21:40 INFO spark.SparkContext: Created broadcast 18 from broadcast at DA
GScheduler.scala:1203
24/04/28 21:21:40 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Resul
tStage 23 (ShuffledRowRDD[43] at save at NativeMethodAccessorImpl.java:0) (first 15
tasks are for partitions Vector(0))
24/04/28 21:21:40 INFO cluster.YarnScheduler: Adding task set 23.0 with 1 tasks
24/04/28 21:21:40 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 23.0 (TI
D 65, algo-1, executor 1, partition 0, NODE_LOCAL, 7778 bytes)
24/04/28 21:21:40 INFO storage.BlockManagerInfo: Added broadcast_18_piece0 in memory
on algo-1:40893 (size: 60.3 KB, free: 13.8 GB)
24/04/28 21:21:40 INFO spark.MapOutputTrackerMasterEndpoint: Asked to send map outpu
t locations for shuffle 7 to 10.0.222.169:41176
tion_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:2
1:38 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_

```

```
01_000002/stderr] 24/04/28 21:21:38 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO output.DirectFileOutputCommitter: Direct Write: DISABLED
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:38 INFO datasources.SQLConfCommitterProvider: Using output committer class org.apache.hadoop.mapreduce.lib.output.DirectFileOutputCommitter
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:39 INFO output.FileOutputCommitter: Saved output of task 'attempt_20240428212138_0020_m_000000_57' to s3a://sagemaker-us-east-1-211125778552/amazon-reviews-spark-analyzer-2024-04-28-21-15-58/output/constraint-checks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:39 INFO mapred.SparkHadoopMapRedUtil: attempt_20240428212138_0020_m_000000_57: Committed
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:39 INFO executor.Executor: Finished task 0.0 in stage 20.0 (TID 57). 2477 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 58
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO executor.Executor: Running task 0.0 in stage 21.0 (TID 58)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 59
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 60
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO executor.Executor: Running task 1.0 in stage 21.0 (TID 59)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 61
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 62
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 63
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 64
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO executor.Executor: Running task 5.0 in stage 21.0 (TID 63)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO executor.Executor: Running task 6.0 in stage 21.0 (TID 64)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 17
```

```
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO executor.Executor: Running task 2.0 in stage 21.0 (TID 60)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO executor.Executor: Running task 4.0 in stage 21.0 (TID 62)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO executor.Executor: Running task 3.0 in stage 21.0 (TID 61)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO memory.MemoryStore: Block broadcast_17_piece0 stored as bytes in memory (estimated size 3.3 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO broadcast.TorrentBroadcast: Reading broadcast variable 17 took 8 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO memory.MemoryStore: Block broadcast_17 stored as values in memory (estimated size 5.4 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO executor.Executor: Finished task 6.0 in stage 21.0 (TID 64). 1384 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO executor.Executor: Finished task 0.0 in stage 21.0 (TID 58). 1384 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO executor.Executor: Finished task 3.0 in stage 21.0 (TID 61). 1384 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO executor.Executor: Finished task 2.0 in stage 21.0 (TID 60). 1384 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO executor.Executor: Finished task 5.0 in stage 21.0 (TID 63). 1384 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO executor.Executor: Finished task 1.0 in stage 21.0 (TID 59). 1341 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO executor.Executor: Finished task 4.0 in stage 21.0 (TID 62). 1384 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 65
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO executor.Executor: Running task 0.0 in stage 23.0 (TID 65)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO spark.MapOutputTrackerWorker: Updating epoch to 8 and clearing cache
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 18
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO memory.MemoryStore: Block broadcast_18_piece0 stored as bytes in memory (estimated size 60.3 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO broadcast.TorrentBroadcast: Reading broadcast
```

```
st variable 18 took 6 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO memory.MemoryStore: Block broadcast_18 stored as values in memory (estimated size 167.5 KB, free 13.8 GB)
24/04/28 21:21:41 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 23.0 (TID 65) in 1028 ms on algo-1 (executor 1) (1/1)
24/04/28 21:21:41 INFO cluster.YarnScheduler: Removed TaskSet 23.0, whose tasks have all completed, from pool
24/04/28 21:21:41 INFO scheduler.DAGScheduler: ResultStage 23 (save at NativeMethodAccessorImpl.java:0) finished in 1.047 s
24/04/28 21:21:41 INFO scheduler.DAGScheduler: Job 14 finished: save at NativeMethodAccessorImpl.java:0, took 1.049390 s
24/04/28 21:21:41 INFO datasources.FileFormatWriter: Write Job 0a656fe2-0cfc-4386-be62-217e10cb80d3 committed.
24/04/28 21:21:41 INFO datasources.FileFormatWriter: Finished processing stats for write job 0a656fe2-0cfc-4386-be62-217e10cb80d3.
24/04/28 21:21:41 INFO datasources.FileSourceStrategy: Pruning directories with:
24/04/28 21:21:41 INFO datasources.FileSourceStrategy: Post-Scan Filters:
24/04/28 21:21:41 INFO datasources.FileSourceStrategy: Output Data Schema: struct<marketplace: string, customer_id: string, review_id: string, product_id: string, product_parent: string ... 13 more fields>
24/04/28 21:21:41 INFO execution.FileSourceScanExec: Pushed Filters:
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 513
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 547
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 558
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 538
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 522
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 563
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 529
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 559
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 567
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 523
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 535
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned shuffle 7
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 521
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 568
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 555
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 539
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 519
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 545
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 532
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 531
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 548
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 536
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 554
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 530
24/04/28 21:21:41 INFO execution.FileSourceScanExec: Pushed Filters:
24/04/28 21:21:41 INFO storage.BlockManagerInfo: Removed broadcast_18_piece0 on 10.0.222.169:36237 in memory (size: 60.3 KB, free: 1008.9 MB)
24/04/28 21:21:41 INFO storage.BlockManagerInfo: Removed broadcast_18_piece0 on algo-1:40893 in memory (size: 60.3 KB, free: 13.8 GB)
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 550
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 510
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 528
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 526
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 543
```

```
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 562
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 515
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 542
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 520
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 544
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 518
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 566
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 541
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 546
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 533
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 514
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 534
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 557
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 552
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 564
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 553
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 517
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 537
24/04/28 21:21:41 INFO storage.BlockManagerInfo: Removed broadcast_17_piece0 on 10.0.222.169:36237 in memory (size: 3.3 KB, free: 1008.9 MB)
24/04/28 21:21:41 INFO storage.BlockManagerInfo: Removed broadcast_17_piece0 on algo-1:40893 in memory (size: 3.3 KB, free: 13.8 GB)
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 527
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 561
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 549
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 525
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 516
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 556
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 524
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 560
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 511
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 540
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 551
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 565
24/04/28 21:21:41 INFO spark.ContextCleaner: Cleaned accumulator 512
24/04/28 21:21:41 INFO codegen.CodeGenerator: Code generated in 18.221847 ms
24/04/28 21:21:41 INFO memory.MemoryStore: Block broadcast_19 stored as values in memory (estimated size 303.8 KB, free 1008.3 MB)
24/04/28 21:21:42 INFO memory.MemoryStore: Block broadcast_19_piece0 stored as bytes in memory (estimated size 27.6 KB, free 1008.3 MB)
24/04/28 21:21:42 INFO storage.BlockManagerInfo: Added broadcast_19_piece0 in memory on 10.0.222.169:36237 (size: 27.6 KB, free: 1008.8 MB)
24/04/28 21:21:42 INFO spark.SparkContext: Created broadcast 19 from collect at AnalysisRunner.scala:323
24/04/28 21:21:42 INFO execution.FileSourceScanExec: Planning scan with bin packing, max size: 4447362 bytes, open cost is considered as scanning 4194304 bytes, number of split files: 3, prefetch: false
24/04/28 21:21:42 INFO execution.FileSourceScanExec: relation: None, fileSplitsInPartitionHistogram: ArrayBuffer((1 fileSplits,3))
24/04/28 21:21:42 INFO scheduler.DAGScheduler: Registering RDD 49 (collect at AnalysisRunner.scala:323) as input to shuffle 8
24/04/28 21:21:42 INFO scheduler.DAGScheduler: Got map stage job 15 (collect at AnalysisRunner.scala:323) with 3 output partitions
24/04/28 21:21:42 INFO scheduler.DAGScheduler: Final stage: ShuffleMapStage 24 (collect at AnalysisRunner.scala:323)
24/04/28 21:21:42 INFO scheduler.DAGScheduler: Parents of final stage: List()
```

```
24/04/28 21:21:42 INFO scheduler.DAGScheduler: Missing parents: List()
24/04/28 21:21:42 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 24 (MapPartitionsRDD[49] at collect at AnalysisRunner.scala:323), which has no missing parents
24/04/28 21:21:42 INFO memory.MemoryStore: Block broadcast_20 stored as values in memory (estimated size 144.1 KB, free 1008.1 MB)
24/04/28 21:21:42 INFO memory.MemoryStore: Block broadcast_20_piece0 stored as bytes in memory (estimated size 46.4 KB, free 1008.1 MB)
24/04/28 21:21:42 INFO storage.BlockManagerInfo: Added broadcast_20_piece0 in memory on 10.0.222.169:36237 (size: 46.4 KB, free: 1008.8 MB)
24/04/28 21:21:42 INFO spark.SparkContext: Created broadcast 20 from broadcast at DAGScheduler.scala:1203
24/04/28 21:21:42 INFO scheduler.DAGScheduler: Submitting 3 missing tasks from ShuffleMapStage 24 (MapPartitionsRDD[49] at collect at AnalysisRunner.scala:323) (first 15 tasks are for partitions Vector(0, 1, 2))
24/04/28 21:21:42 INFO cluster.YarnScheduler: Adding task set 24.0 with 3 tasks
24/04/28 21:21:42 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 24.0 (TID 66, algo-1, executor 1, partition 0, PROCESS_LOCAL, 8328 bytes)
24/04/28 21:21:42 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 24.0 (TID 67, algo-1, executor 1, partition 1, PROCESS_LOCAL, 8325 bytes)
24/04/28 21:21:42 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 24.0 (TID 68, algo-1, executor 1, partition 2, PROCESS_LOCAL, 8318 bytes)
24/04/28 21:21:42 INFO storage.BlockManagerInfo: Added broadcast_20_piece0 in memory on algo-1:40893 (size: 46.4 KB, free: 13.8 GB)
24/04/28 21:21:42 INFO storage.BlockManagerInfo: Added broadcast_19_piece0 in memory on algo-1:40893 (size: 27.6 KB, free: 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO spark.MapOutputTrackerWorker: Don't have map outputs for shuffle 7, fetching them
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO spark.MapOutputTrackerWorker: Doing the fetch; tracker endpoint = NettyRpcEndpointRef(spark://MapOutputTracker@10.0.222.169:37411)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO spark.MapOutputTrackerWorker: Got the output locations
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO storage.ShuffleBlockFetcherIterator: Getting 7 non-empty blocks including 7 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO output.DirectFileOutputCommitter: Direct Write: DISABLED
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:40 INFO datasources.SQLConfCommitterProvider: Using output committer class org.apache.hadoop.mapreduce.lib.output.DirectFileOutputCommitter
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
```

```
01_000002/stderr] 24/04/28 21:21:41 INFO output.FileOutputCommitter: Saved output of
task 'attempt_20240428212140_0023_m_000000_65' to s3a://sagemaker-us-east-1-21112577
8552/amazon-reviews-spark-analyzer-2024-04-28-21-15-58/output/success-metrics
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:41 INFO mapred.SparkHadoopMapRedUtil: attempt_20240
428212140_0023_m_000000_65: Committed
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:41 INFO executor.Executor: Finished task 0.0 in sta
ge 23.0 (TID 65). 2434 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:42 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 66
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:42 INFO executor.Executor: Running task 0.0 in stag
e 24.0 (TID 66)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:42 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 67
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:42 INFO executor.Executor: Running task 1.0 in stag
e 24.0 (TID 67)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:42 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 68
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:42 INFO broadcast.TorrentBroadcast: Started reading
broadcast variable 20
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:42 INFO executor.Executor: Running task 2.0 in stag
e 24.0 (TID 68)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:42 INFO memory.MemoryStore: Block broadcast_20_piec
e0 stored as bytes in memory (estimated size 46.4 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:42 INFO broadcast.TorrentBroadcast: Reading broadca
st variable 20 took 5 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:42 INFO memory.MemoryStore: Block broadcast_20 stor
ed as values in memory (estimated size 144.1 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:42 INFO datasources.FileScanRDD: TID: 68 - Reading
current file: path: s3a://sagemaker-us-east-1-211125778552/amazon-reviews-pds/tsv/am
azon_reviews_us_Gift_Card_v1_00.tsv.gz, range: 0-12134676, partition values: [empty
row], isDataPresent: false
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:42 INFO datasources.FileScanRDD: TID: 66 - Reading
current file: path: s3a://sagemaker-us-east-1-211125778552/amazon-reviews-pds/tsv/am
azon_reviews_us_Digital_Video_Games_v1_00.tsv.gz, range: 0-27442648, partition value
s: [empty row], isDataPresent: false
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:42 INFO datasources.FileScanRDD: TID: 67 - Reading
current file: path: s3a://sagemaker-us-east-1-211125778552/amazon-reviews-pds/tsv/am
azon_reviews_us_Digital_Software_v1_00.tsv.gz, range: 0-18997559, partition values:
[empty row], isDataPresent: false
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:42 INFO codegen.CodeGenerator: Code generated in 1
```



```
2.147405 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:42 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 19
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:42 INFO memory.MemoryStore: Block broadcast_19_piece0 stored as bytes in memory (estimated size 27.6 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:42 INFO broadcast.TorrentBroadcast: Reading broadcast variable 19 took 6 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:42 INFO memory.MemoryStore: Block broadcast_19 stored as values in memory (estimated size 398.1 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:42 INFO codegen.CodeGenerator: Code generated in 10.152936 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:42 INFO codegen.CodeGenerator: Code generated in 15.661792 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:42 INFO codegen.CodeGenerator: Code generated in 9.284136 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:42 INFO codegen.CodeGenerator: Code generated in 171.593542 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:42 INFO codegen.CodeGenerator: Code generated in 6.514605 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:42 INFO codegen.CodeGenerator: Code generated in 5.055424 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:42 INFO codegen.CodeGenerator: Code generated in 5.028027 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:42 INFO codegen.CodeGenerator: Code generated in 4.995647 ms
24/04/28 21:21:45 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 24.0 (TID 67) in 3461 ms on algo-1 (executor 1) (1/3)
24/04/28 21:21:45 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 24.0 (TID 68) in 3821 ms on algo-1 (executor 1) (2/3)
24/04/28 21:21:46 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 24.0 (TID 66) in 4501 ms on algo-1 (executor 1) (3/3)
24/04/28 21:21:46 INFO cluster.YarnScheduler: Removed TaskSet 24.0, whose tasks have all completed, from pool
24/04/28 21:21:46 INFO scheduler.DAGScheduler: ShuffleMapStage 24 (collect at AnalysisRunner.scala:323) finished in 4.512 s
24/04/28 21:21:46 INFO scheduler.DAGScheduler: looking for newly runnable stages
24/04/28 21:21:46 INFO scheduler.DAGScheduler: running: Set()
24/04/28 21:21:46 INFO scheduler.DAGScheduler: waiting: Set()
24/04/28 21:21:46 INFO scheduler.DAGScheduler: failed: Set()
24/04/28 21:21:46 INFO adaptive.CoalesceShufflePartitions: advisoryTargetPostShuffleInputSize: 67108864, targetPostShuffleInputSize 849.
24/04/28 21:21:46 INFO spark.SparkContext: Starting job: collect at AnalysisRunner.scala:323
24/04/28 21:21:46 INFO scheduler.DAGScheduler: Got job 16 (collect at AnalysisRunner
```

```
r.scala:323) with 1 output partitions
24/04/28 21:21:46 INFO scheduler.DAGScheduler: Final stage: ResultStage 26 (collect
at AnalysisRunner.scala:323)
24/04/28 21:21:46 INFO scheduler.DAGScheduler: Parents of final stage: List(ShuffleM
apStage 25)
24/04/28 21:21:46 INFO scheduler.DAGScheduler: Missing parents: List()
24/04/28 21:21:46 INFO scheduler.DAGScheduler: Submitting ResultStage 26 (MapPartiti
onsRDD[52] at collect at AnalysisRunner.scala:323), which has no missing parents
24/04/28 21:21:46 INFO memory.MemoryStore: Block broadcast_21 stored as values in me
memory (estimated size 175.4 KB, free 1007.9 MB)
24/04/28 21:21:46 INFO memory.MemoryStore: Block broadcast_21_piece0 stored as bytes
in memory (estimated size 57.4 KB, free 1007.8 MB)
24/04/28 21:21:46 INFO storage.BlockManagerInfo: Added broadcast_21_piece0 in memory
on 10.0.222.169:36237 (size: 57.4 KB, free: 1008.7 MB)
24/04/28 21:21:46 INFO spark.SparkContext: Created broadcast 21 from broadcast at DA
GScheduler.scala:1203
24/04/28 21:21:46 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Resul
tStage 26 (MapPartitionsRDD[52] at collect at AnalysisRunner.scala:323) (first 15 ta
sks are for partitions Vector(0))
24/04/28 21:21:46 INFO cluster.YarnScheduler: Adding task set 26.0 with 1 tasks
24/04/28 21:21:46 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 26.0 (TI
D 69, algo-1, executor 1, partition 0, NODE_LOCAL, 7778 bytes)
24/04/28 21:21:46 INFO storage.BlockManagerInfo: Added broadcast_21_piece0 in memory
on algo-1:40893 (size: 57.4 KB, free: 13.8 GB)
24/04/28 21:21:46 INFO spark.MapOutputTrackerMasterEndpoint: Asked to send map outpu
t locations for shuffle 8 to 10.0.222.169:41176
24/04/28 21:21:46 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 26.0 (TI
D 69) in 275 ms on algo-1 (executor 1) (1/1)
24/04/28 21:21:46 INFO cluster.YarnScheduler: Removed TaskSet 26.0, whose tasks have
all completed, from pool
24/04/28 21:21:46 INFO scheduler.DAGScheduler: ResultStage 26 (collect at AnalysisRu
nner.scala:323) finished in 0.284 s
24/04/28 21:21:46 INFO scheduler.DAGScheduler: Job 16 finished: collect at AnalysisR
unner.scala:323, took 0.285980 s
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 611
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 646
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 638
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 607
24/04/28 21:21:46 INFO storage.BlockManagerInfo: Removed broadcast_21_piece0 on 10.
0.222.169:36237 in memory (size: 57.4 KB, free: 1008.8 MB)
24/04/28 21:21:46 INFO storage.BlockManagerInfo: Removed broadcast_21_piece0 on algo
-1:40893 in memory (size: 57.4 KB, free: 13.8 GB)
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 596
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 637
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 582
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 571
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 605
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 593
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 586
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 641
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 627
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 578
24/04/28 21:21:46 INFO storage.BlockManagerInfo: Removed broadcast_20_piece0 on 10.
0.222.169:36237 in memory (size: 46.4 KB, free: 1008.8 MB)
24/04/28 21:21:46 INFO storage.BlockManagerInfo: Removed broadcast_20_piece0 on algo
-1:40893 in memory (size: 46.4 KB, free: 13.8 GB)
```

24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 615  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 634  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 606  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 608  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 623  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 622  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 630  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 618  
24/04/28 21:21:46 INFO storage.BlockManagerInfo: Removed broadcast\_19\_piece0 on 10.0.222.169:36237 in memory (size: 27.6 KB, free: 1008.9 MB)  
24/04/28 21:21:46 INFO storage.BlockManagerInfo: Removed broadcast\_19\_piece0 on algo-1:40893 in memory (size: 27.6 KB, free: 13.8 GB)  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 572  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 632  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 612  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 602  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 601  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 614  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 592  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 573  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 642  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 639  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 577  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 583  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 610  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 625  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 569  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 590  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 588  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 603  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 644  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 599  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 629  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 604  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 624  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned shuffle 8  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 597  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 626  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 643  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 581  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 631  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 645  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 640  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 595  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 619  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 617  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 598  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 575  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 635  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 585  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 589  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 620  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 591  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 633  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 587  
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 616

```
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 576
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 628
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 613
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 600
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 570
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 579
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 621
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 584
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 609
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 580
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 594
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 636
24/04/28 21:21:46 INFO spark.ContextCleaner: Cleaned accumulator 574
24/04/28 21:21:47 INFO datasources.FileSourceStrategy: Pruning directories with:
24/04/28 21:21:47 INFO datasources.FileSourceStrategy: Post-Scan Filters:
24/04/28 21:21:47 INFO datasources.FileSourceStrategy: Output Data Schema: struct<customer_id: string, product_parent: string, star_rating: int, helpful_votes: int, total_votes: int ... 3 more fields>
24/04/28 21:21:47 INFO execution.FileSourceScanExec: Pushed Filters:
24/04/28 21:21:47 INFO execution.FileSourceScanExec: Pushed Filters:
24/04/28 21:21:47 INFO codegen.CodeGenerator: Code generated in 25.941838 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:42 INFO codegen.CodeGenerator: Code generated in 4.677641 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:42 INFO codegen.CodeGenerator: Code generated in 6.255669 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:42 INFO codegen.CodeGenerator: Code generated in 6.267663 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:42 INFO codegen.CodeGenerator: Code generated in 6.359029 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:42 INFO codegen.CodeGenerator: Code generated in 6.389966 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:42 INFO codegen.CodeGenerator: Code generated in 7.071582 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:42 INFO codegen.CodeGenerator: Code generated in 6.16827 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:42 INFO codegen.CodeGenerator: Code generated in 5.996477 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:45 INFO executor.Executor: Finished task 1.0 in stage 24.0 (TID 67). 2386 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:45 INFO executor.Executor: Finished task 2.0 in stage 24.0 (TID 68). 2386 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:46 INFO executor.Executor: Finished task 0.0 in stage 24.0 (TID 66). 2386 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:46 INFO executor.CoarseGrainedExecutorBackend: Got
```

assigned task 69  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:46 INFO executor.Executor: Running task 0.0 in stage 26.0 (TID 69)  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:46 INFO spark.MapOutputTrackerWorker: Updating epoch to 9 and clearing cache  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:46 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 21  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:46 INFO memory.MemoryStore: Block broadcast\_21\_piece0 stored as bytes in memory (estimated size 57.4 KB, free 13.8 GB)  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:46 INFO broadcast.TorrentBroadcast: Reading broadcast variable 21 took 5 ms  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:46 INFO memory.MemoryStore: Block broadcast\_21 stored as values in memory (estimated size 175.4 KB, free 13.8 GB)  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:46 INFO spark.MapOutputTrackerWorker: Don't have map outputs for shuffle 8, fetching them  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:46 INFO spark.MapOutputTrackerWorker: Doing the fetch; tracker endpoint = NettyRpcEndpointRef(spark://MapOutputTracker@10.0.222.169:37411)  
[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:46 INFO spark.MapOutputTrackerWorker: Got the output locations  
24/04/28 21:21:47 INFO codegen.CodeGenerator: Code generated in 42.98724 ms  
24/04/28 21:21:47 INFO memory.MemoryStore: Block broadcast\_22 stored as values in memory (estimated size 303.8 KB, free 1008.3 MB)  
24/04/28 21:21:47 INFO memory.MemoryStore: Block broadcast\_22\_piece0 stored as bytes in memory (estimated size 27.6 KB, free 1008.3 MB)  
24/04/28 21:21:47 INFO storage.BlockManagerInfo: Added broadcast\_22\_piece0 in memory on 10.0.222.169:36237 (size: 27.6 KB, free: 1008.8 MB)  
24/04/28 21:21:47 INFO spark.SparkContext: Created broadcast 22 from collect at AnalysisRunner.scala:323  
24/04/28 21:21:47 INFO execution.FileSourceScanExec: Planning scan with bin packing, max size: 4447362 bytes, open cost is considered as scanning 4194304 bytes, number of split files: 3, prefetch: false  
24/04/28 21:21:47 INFO execution.FileSourceScanExec: relation: None, fileSplitsInPartitionHistogram: ArrayBuffer((1 fileSplits,3))  
24/04/28 21:21:47 INFO scheduler.DAGScheduler: Registering RDD 55 (collect at AnalysisRunner.scala:323) as input to shuffle 9  
24/04/28 21:21:47 INFO scheduler.DAGScheduler: Got map stage job 17 (collect at AnalysisRunner.scala:323) with 3 output partitions  
24/04/28 21:21:47 INFO scheduler.DAGScheduler: Final stage: ShuffleMapStage 27 (collect at AnalysisRunner.scala:323)  
24/04/28 21:21:47 INFO scheduler.DAGScheduler: Parents of final stage: List()  
24/04/28 21:21:47 INFO scheduler.DAGScheduler: Missing parents: List()  
24/04/28 21:21:47 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 27 (MapPartitionsRDD[55] at collect at AnalysisRunner.scala:323), which has no missing parents  
24/04/28 21:21:47 INFO memory.MemoryStore: Block broadcast\_23 stored as values in memory (estimated size 53.6 KB, free 1008.2 MB)  
24/04/28 21:21:47 INFO memory.MemoryStore: Block broadcast\_23\_piece0 stored as bytes

```
in memory (estimated size 19.4 KB, free 1008.2 MB)
24/04/28 21:21:47 INFO storage.BlockManagerInfo: Added broadcast_23_piece0 in memory
on 10.0.222.169:36237 (size: 19.4 KB, free: 1008.8 MB)
24/04/28 21:21:47 INFO spark.SparkContext: Created broadcast 23 from broadcast at DA
GScheduler.scala:1203
24/04/28 21:21:47 INFO scheduler.DAGScheduler: Submitting 3 missing tasks from Shuff
leMapStage 27 (MapPartitionsRDD[55] at collect at AnalysisRunner.scala:323) (first 1
5 tasks are for partitions Vector(0, 1, 2))
24/04/28 21:21:47 INFO cluster.YarnScheduler: Adding task set 27.0 with 3 tasks
24/04/28 21:21:47 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 27.0 (TI
D 70, algo-1, executor 1, partition 0, PROCESS_LOCAL, 8328 bytes)
24/04/28 21:21:47 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 27.0 (TI
D 71, algo-1, executor 1, partition 1, PROCESS_LOCAL, 8325 bytes)
24/04/28 21:21:47 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 27.0 (TI
D 72, algo-1, executor 1, partition 2, PROCESS_LOCAL, 8318 bytes)
24/04/28 21:21:47 INFO storage.BlockManagerInfo: Added broadcast_23_piece0 in memory
on algo-1:40893 (size: 19.4 KB, free: 13.8 GB)
24/04/28 21:21:47 INFO storage.BlockManagerInfo: Added broadcast_22_piece0 in memory
on algo-1:40893 (size: 27.6 KB, free: 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:46 INFO storage.ShuffleBlockFetcherIterator: Gettin
g 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:46 INFO storage.ShuffleBlockFetcherIterator: Starte
d 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:46 INFO codegen.CodeGenerator: Code generated in 1
7.681652 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:46 INFO codegen.CodeGenerator: Code generated in 9.
214103 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:46 INFO codegen.CodeGenerator: Code generated in 1
4.649536 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:46 INFO executor.Executor: Finished task 0.0 in sta
ge 26.0 (TID 69). 7724 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:47 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 70
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:47 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 71
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:47 INFO executor.Executor: Running task 0.0 in stag
e 27.0 (TID 70)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:47 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 72
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:47 INFO executor.Executor: Running task 1.0 in stag
e 27.0 (TID 71)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:47 INFO broadcast.TorrentBroadcast: Started reading
broadcast variable 23
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
```

```
01_000002/stderr] 24/04/28 21:21:47 INFO executor.Executor: Running task 2.0 in stage 27.0 (TID 72)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:47 INFO memory.MemoryStore: Block broadcast_23_piece0 stored as bytes in memory (estimated size 19.4 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:47 INFO broadcast.TorrentBroadcast: Reading broadcast variable 23 took 6 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:47 INFO memory.MemoryStore: Block broadcast_23 stored as values in memory (estimated size 53.6 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:47 INFO codegen.CodeGenerator: Code generated in 33.143924 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:47 INFO datasources.FileScanRDD: TID: 71 - Reading current file: path: s3a://sagemaker-us-east-1-211125778552/amazon-reviews-pds/tsv/amazon_reviews_us_Digital_Software_v1_00.tsv.gz, range: 0-18997559, partition values: [empty row], isDataPresent: false
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:47 INFO datasources.FileScanRDD: TID: 70 - Reading current file: path: s3a://sagemaker-us-east-1-211125778552/amazon-reviews-pds/tsv/amazon_reviews_us_Digital_Video_Games_v1_00.tsv.gz, range: 0-27442648, partition values: [empty row], isDataPresent: false
24/04/28 21:21:48 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 27.0 (TID 72) in 979 ms on algo-1 (executor 1) (1/3)
24/04/28 21:21:48 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 27.0 (TID 71) in 1116 ms on algo-1 (executor 1) (2/3)
24/04/28 21:21:48 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 27.0 (TID 70) in 1442 ms on algo-1 (executor 1) (3/3)
24/04/28 21:21:48 INFO cluster.YarnScheduler: Removed TaskSet 27.0, whose tasks have all completed, from pool
24/04/28 21:21:48 INFO scheduler.DAGScheduler: ShuffleMapStage 27 (collect at AnalysisRunner.scala:323) finished in 1.448 s
24/04/28 21:21:48 INFO scheduler.DAGScheduler: looking for newly runnable stages
24/04/28 21:21:48 INFO scheduler.DAGScheduler: running: Set()
24/04/28 21:21:48 INFO scheduler.DAGScheduler: waiting: Set()
24/04/28 21:21:48 INFO scheduler.DAGScheduler: failed: Set()
24/04/28 21:21:48 INFO adaptive.CoalesceShufflePartitions: advisoryTargetPostShuffleInputSize: 67108864, targetPostShuffleInputSize 37.
24/04/28 21:21:48 INFO spark.ContextCleaner: Cleaned accumulator 672
24/04/28 21:21:48 INFO spark.ContextCleaner: Cleaned accumulator 676
24/04/28 21:21:48 INFO spark.ContextCleaner: Cleaned accumulator 678
24/04/28 21:21:48 INFO spark.ContextCleaner: Cleaned accumulator 674
24/04/28 21:21:48 INFO spark.ContextCleaner: Cleaned accumulator 670
24/04/28 21:21:48 INFO storage.BlockManagerInfo: Removed broadcast_23_piece0 on 10.0.222.169:36237 in memory (size: 19.4 KB, free: 1008.8 MB)
24/04/28 21:21:48 INFO storage.BlockManagerInfo: Removed broadcast_23_piece0 on algo-1:40893 in memory (size: 19.4 KB, free: 13.8 GB)
24/04/28 21:21:48 INFO codegen.CodeGenerator: Code generated in 36.717297 ms
24/04/28 21:21:48 INFO spark.ContextCleaner: Cleaned accumulator 679
24/04/28 21:21:48 INFO spark.ContextCleaner: Cleaned accumulator 686
24/04/28 21:21:48 INFO spark.ContextCleaner: Cleaned accumulator 689
24/04/28 21:21:48 INFO spark.ContextCleaner: Cleaned accumulator 692
24/04/28 21:21:48 INFO spark.ContextCleaner: Cleaned accumulator 691
24/04/28 21:21:48 INFO spark.ContextCleaner: Cleaned accumulator 680
```

```
24/04/28 21:21:48 INFO spark.ContextCleaner: Cleaned accumulator 694
24/04/28 21:21:48 INFO spark.ContextCleaner: Cleaned accumulator 685
24/04/28 21:21:48 INFO spark.ContextCleaner: Cleaned accumulator 682
24/04/28 21:21:48 INFO spark.ContextCleaner: Cleaned accumulator 687
24/04/28 21:21:48 INFO spark.ContextCleaner: Cleaned accumulator 690
24/04/28 21:21:48 INFO spark.ContextCleaner: Cleaned accumulator 675
24/04/28 21:21:48 INFO spark.ContextCleaner: Cleaned accumulator 693
24/04/28 21:21:48 INFO spark.ContextCleaner: Cleaned accumulator 671
24/04/28 21:21:48 INFO spark.ContextCleaner: Cleaned accumulator 684
24/04/28 21:21:48 INFO spark.ContextCleaner: Cleaned accumulator 683
24/04/28 21:21:48 INFO spark.ContextCleaner: Cleaned accumulator 673
24/04/28 21:21:48 INFO spark.ContextCleaner: Cleaned accumulator 681
24/04/28 21:21:48 INFO spark.ContextCleaner: Cleaned accumulator 677
24/04/28 21:21:48 INFO spark.ContextCleaner: Cleaned accumulator 688
24/04/28 21:21:48 INFO spark.SparkContext: Starting job: collect at AnalysisRunner.s
cala:323
24/04/28 21:21:48 INFO scheduler.DAGScheduler: Got job 18 (collect at AnalysisRunne
r.scala:323) with 1 output partitions
24/04/28 21:21:48 INFO scheduler.DAGScheduler: Final stage: ResultStage 29 (collect
at AnalysisRunner.scala:323)
24/04/28 21:21:48 INFO scheduler.DAGScheduler: Parents of final stage: List(ShuffleM
apStage 28)
24/04/28 21:21:48 INFO scheduler.DAGScheduler: Missing parents: List()
24/04/28 21:21:48 INFO scheduler.DAGScheduler: Submitting ResultStage 29 (MapPartiti
onsRDD[58] at collect at AnalysisRunner.scala:323), which has no missing parents
24/04/28 21:21:48 INFO memory.MemoryStore: Block broadcast_24 stored as values in me
memory (estimated size 50.7 KB, free 1008.2 MB)
24/04/28 21:21:48 INFO memory.MemoryStore: Block broadcast_24_piece0 stored as bytes
in memory (estimated size 16.2 KB, free 1008.2 MB)
24/04/28 21:21:48 INFO storage.BlockManagerInfo: Added broadcast_24_piece0 in memory
on 10.0.222.169:36237 (size: 16.2 KB, free: 1008.8 MB)
24/04/28 21:21:48 INFO spark.SparkContext: Created broadcast 24 from broadcast at DA
GScheduler.scala:1203
24/04/28 21:21:48 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Resul
tStage 29 (MapPartitionsRDD[58] at collect at AnalysisRunner.scala:323) (first 15 ta
sks are for partitions Vector(0))
24/04/28 21:21:48 INFO cluster.YarnScheduler: Adding task set 29.0 with 1 tasks
24/04/28 21:21:48 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 29.0 (TI
D 73, algo-1, executor 1, partition 0, NODE_LOCAL, 7778 bytes)
24/04/28 21:21:48 INFO storage.BlockManagerInfo: Added broadcast_24_piece0 in memory
on algo-1:40893 (size: 16.2 KB, free: 13.8 GB)
24/04/28 21:21:48 INFO spark.MapOutputTrackerMasterEndpoint: Asked to send map outpu
t locations for shuffle 9 to 10.0.222.169:41176
24/04/28 21:21:48 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 29.0 (TI
D 73) in 53 ms on algo-1 (executor 1) (1/1)
24/04/28 21:21:48 INFO cluster.YarnScheduler: Removed TaskSet 29.0, whose tasks have
all completed, from pool
24/04/28 21:21:48 INFO scheduler.DAGScheduler: ResultStage 29 (collect at AnalysisRu
nner.scala:323) finished in 0.058 s
24/04/28 21:21:48 INFO scheduler.DAGScheduler: Job 18 finished: collect at AnalysisR
unner.scala:323, took 0.060010 s
24/04/28 21:21:48 INFO datasources.FileSourceStrategy: Pruning directories with:
24/04/28 21:21:48 INFO datasources.FileSourceStrategy: Post-Scan Filters:
24/04/28 21:21:48 INFO datasources.FileSourceStrategy: Output Data Schema: struct<ma
rketplace: string, customer_id: string, review_id: string, product_id: string, produ
ct_parent: string ... 13 more fields>
```



```
24/04/28 21:21:48 INFO execution.FileSourceScanExec: Pushed Filters:
24/04/28 21:21:48 INFO memory.MemoryStore: Block broadcast_25 stored as values in me
memory (estimated size 303.8 KB, free 1007.9 MB)
24/04/28 21:21:48 INFO memory.MemoryStore: Block broadcast_25_piece0 stored as bytes
in memory (estimated size 27.6 KB, free 1007.9 MB)
24/04/28 21:21:48 INFO storage.BlockManagerInfo: Added broadcast_25_piece0 in memory
on 10.0.222.169:36237 (size: 27.6 KB, free: 1008.8 MB)
24/04/28 21:21:48 INFO spark.SparkContext: Created broadcast 25 from rdd at ColumnPr
ofiler.scala:591
24/04/28 21:21:48 INFO execution.FileSourceScanExec: Planning scan with bin packing,
max size: 4447362 bytes, open cost is considered as scanning 4194304 bytes, number o
f split files: 3, prefetch: false
24/04/28 21:21:48 INFO execution.FileSourceScanExec: relation: None, fileSplitsInPar
titionHistogram: ArrayBuffer((1 fileSplits,3))
24/04/28 21:21:48 INFO spark.SparkContext: Starting job: countByKey at ColumnProfile
r.scala:605
24/04/28 21:21:48 INFO scheduler.DAGScheduler: Registering RDD 65 (countByKey at Col
umnProfiler.scala:605) as input to shuffle 10
24/04/28 21:21:48 INFO scheduler.DAGScheduler: Got job 19 (countByKey at ColumnProfi
ler.scala:605) with 16 output partitions
24/04/28 21:21:48 INFO scheduler.DAGScheduler: Final stage: ResultStage 31 (countByK
ey at ColumnProfiler.scala:605)
24/04/28 21:21:48 INFO scheduler.DAGScheduler: Parents of final stage: List(ShuffleM
apStage 30)
24/04/28 21:21:48 INFO scheduler.DAGScheduler: Missing parents: List(ShuffleMapStage
30)
24/04/28 21:21:48 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 30 (MapPar
titionsRDD[65] at countByKey at ColumnProfiler.scala:605), which has no missing pare
nts
24/04/28 21:21:48 INFO memory.MemoryStore: Block broadcast_26 stored as values in me
memory (estimated size 21.2 KB, free 1007.8 MB)
24/04/28 21:21:48 INFO memory.MemoryStore: Block broadcast_26_piece0 stored as bytes
in memory (estimated size 10.5 KB, free 1007.8 MB)
24/04/28 21:21:48 INFO storage.BlockManagerInfo: Added broadcast_26_piece0 in memory
on 10.0.222.169:36237 (size: 10.5 KB, free: 1008.8 MB)
24/04/28 21:21:48 INFO spark.SparkContext: Created broadcast 26 from broadcast at DA
GScheduler.scala:1203
24/04/28 21:21:48 INFO scheduler.DAGScheduler: Submitting 3 missing tasks from Shuff
leMapStage 30 (MapPartitionsRDD[65] at countByKey at ColumnProfiler.scala:605) (firs
t 15 tasks are for partitions Vector(0, 1, 2))
24/04/28 21:21:48 INFO cluster.YarnScheduler: Adding task set 30.0 with 3 tasks
24/04/28 21:21:48 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 30.0 (TI
D 74, algo-1, executor 1, partition 0, PROCESS_LOCAL, 8328 bytes)
24/04/28 21:21:48 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 30.0 (TI
D 75, algo-1, executor 1, partition 1, PROCESS_LOCAL, 8325 bytes)
24/04/28 21:21:48 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 30.0 (TI
D 76, algo-1, executor 1, partition 2, PROCESS_LOCAL, 8318 bytes)
24/04/28 21:21:48 INFO storage.BlockManagerInfo: Added broadcast_26_piece0 in memory
on algo-1:40893 (size: 10.5 KB, free: 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:47 INFO datasources.FileScanRDD: TID: 72 - Reading
current file: path: s3a://sagemaker-us-east-1-211125778552/amazon-reviews-pds/tsv/am
azon_reviews_us_Gift_Card_v1_00.tsv.gz, range: 0-12134676, partition values: [empty
row], isDataPresent: false
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:47 INFO codegen.CodeGenerator: Code generated in 7.
```

563376 ms  
[//log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:47 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 22  
[//log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:47 INFO memory.MemoryStore: Block broadcast\_22\_piece0 stored as bytes in memory (estimated size 27.6 KB, free 13.8 GB)  
[//log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:47 INFO broadcast.TorrentBroadcast: Reading broadcast variable 22 took 6 ms  
[//log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:47 INFO memory.MemoryStore: Block broadcast\_22 stored as values in memory (estimated size 398.1 KB, free 13.8 GB)  
[//log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:48 INFO executor.Executor: Finished task 2.0 in stage 27.0 (TID 72). 1740 bytes result sent to driver  
[//log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:48 INFO executor.Executor: Finished task 1.0 in stage 27.0 (TID 71). 1740 bytes result sent to driver  
[//log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:48 INFO executor.Executor: Finished task 0.0 in stage 27.0 (TID 70). 1740 bytes result sent to driver  
[//log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:48 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 73  
[//log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:48 INFO executor.Executor: Running task 0.0 in stage 29.0 (TID 73)  
[//log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:48 INFO spark.MapOutputTrackerWorker: Updating epoch to 10 and clearing cache  
[//log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:48 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 24  
[//log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:48 INFO memory.MemoryStore: Block broadcast\_24\_piece0 stored as bytes in memory (estimated size 16.2 KB, free 13.8 GB)  
[//log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:48 INFO broadcast.TorrentBroadcast: Reading broadcast variable 24 took 5 ms  
[//log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:48 INFO memory.MemoryStore: Block broadcast\_24 stored as values in memory (estimated size 50.7 KB, free 13.8 GB)  
[//log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:48 INFO spark.MapOutputTrackerWorker: Don't have map outputs for shuffle 9, fetching them  
[//log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:48 INFO spark.MapOutputTrackerWorker: Doing the fetch; tracker endpoint = NettyRpcEndpointRef(spark://MapOutputTracker@10.0.222.169:37411)  
[//log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:48 INFO spark.MapOutputTrackerWorker: Got the output locations  
24/04/28 21:21:50 INFO storage.BlockManagerInfo: Added broadcast\_25\_piece0 in memory on algo-1:40893 (size: 27.6 KB, free: 13.8 GB)  
[//log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_

```
01_000002/stderr] 24/04/28 21:21:48 INFO storage.ShuffleBlockFetcherIterator: Gettin
g 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:48 INFO storage.ShuffleBlockFetcherIterator: Starte
d 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:48 INFO codegen.CodeGenerator: Code generated in 3
1.198592 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:48 INFO executor.Executor: Finished task 0.0 in sta
ge 29.0 (TID 73). 2197 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:48 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 74
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:48 INFO executor.Executor: Running task 0.0 in stag
e 30.0 (TID 74)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:48 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 75
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:48 INFO executor.Executor: Running task 1.0 in stag
e 30.0 (TID 75)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:48 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 76
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:48 INFO broadcast.TorrentBroadcast: Started reading
broadcast variable 26
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:48 INFO executor.Executor: Running task 2.0 in stag
e 30.0 (TID 76)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:48 INFO memory.MemoryStore: Block broadcast_26_piec
e0 stored as bytes in memory (estimated size 10.5 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:48 INFO broadcast.TorrentBroadcast: Reading broadca
st variable 26 took 5 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:48 INFO memory.MemoryStore: Block broadcast_26 stor
ed as values in memory (estimated size 21.2 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:50 INFO codegen.CodeGenerator: Code generated in 6.
853977 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:50 INFO datasources.FileScanRDD: TID: 75 - Reading
current file: path: s3a://sagemaker-us-east-1-211125778552/amazon-reviews-pds/tsv/am
azon_reviews_us_Digital_Software_v1_00.tsv.gz, range: 0-18997559, partition values:
[empty row], isDataPresent: false
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:50 INFO datasources.FileScanRDD: TID: 76 - Reading
current file: path: s3a://sagemaker-us-east-1-211125778552/amazon-reviews-pds/tsv/am
azon_reviews_us_Gift_Card_v1_00.tsv.gz, range: 0-12134676, partition values: [empty
row], isDataPresent: false
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:50 INFO datasources.FileScanRDD: TID: 74 - Reading
```

current file: path: s3a://sagemaker-us-east-1-211125778552/amazon-reviews-pds/tsv/amazon\_reviews\_us\_Digital\_Video\_Games\_v1\_00.tsv.gz, range: 0-27442648, partition value s: [empty row], isDataPresent: false

24/04/28 21:21:52 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 30.0 (TID 76) in 3258 ms on algo-1 (executor 1) (1/3)

24/04/28 21:21:52 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 30.0 (TID 75) in 3259 ms on algo-1 (executor 1) (2/3)

24/04/28 21:21:52 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 30.0 (TID 74) in 3844 ms on algo-1 (executor 1) (3/3)

24/04/28 21:21:52 INFO cluster.YarnScheduler: Removed TaskSet 30.0, whose tasks have all completed, from pool

24/04/28 21:21:52 INFO scheduler.DAGScheduler: ShuffleMapStage 30 (countByKey at ColumnProfiler.scala:605) finished in 3.862 s

24/04/28 21:21:52 INFO scheduler.DAGScheduler: looking for newly runnable stages

24/04/28 21:21:52 INFO scheduler.DAGScheduler: running: Set()

24/04/28 21:21:52 INFO scheduler.DAGScheduler: waiting: Set(ResultStage 31)

24/04/28 21:21:52 INFO scheduler.DAGScheduler: failed: Set()

24/04/28 21:21:52 INFO scheduler.DAGScheduler: Submitting ResultStage 31 (ShuffledRDD[66] at countByKey at ColumnProfiler.scala:605), which has no missing parents

24/04/28 21:21:52 INFO memory.MemoryStore: Block broadcast\_27 stored as values in memory (estimated size 3.1 KB, free 1007.8 MB)

24/04/28 21:21:52 INFO memory.MemoryStore: Block broadcast\_27\_piece0 stored as bytes in memory (estimated size 1922.0 B, free 1007.8 MB)

24/04/28 21:21:52 INFO storage.BlockManagerInfo: Added broadcast\_27\_piece0 in memory on 10.0.222.169:36237 (size: 1922.0 B, free: 1008.8 MB)

24/04/28 21:21:52 INFO spark.SparkContext: Created broadcast 27 from broadcast at DAGScheduler.scala:1203

24/04/28 21:21:52 INFO scheduler.DAGScheduler: Submitting 16 missing tasks from ResultStage 31 (ShuffledRDD[66] at countByKey at ColumnProfiler.scala:605) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14))

24/04/28 21:21:52 INFO cluster.YarnScheduler: Adding task set 31.0 with 16 tasks

24/04/28 21:21:52 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 31.0 (TID 77, algo-1, executor 1, partition 2, NODE\_LOCAL, 7673 bytes)

24/04/28 21:21:52 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 31.0 (TID 78, algo-1, executor 1, partition 5, NODE\_LOCAL, 7673 bytes)

24/04/28 21:21:52 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 31.0 (TID 79, algo-1, executor 1, partition 7, NODE\_LOCAL, 7673 bytes)

24/04/28 21:21:52 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 31.0 (TID 80, algo-1, executor 1, partition 8, NODE\_LOCAL, 7673 bytes)

24/04/28 21:21:52 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 31.0 (TID 81, algo-1, executor 1, partition 9, NODE\_LOCAL, 7673 bytes)

24/04/28 21:21:52 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 31.0 (TID 82, algo-1, executor 1, partition 10, NODE\_LOCAL, 7673 bytes)

24/04/28 21:21:52 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 31.0 (TID 83, algo-1, executor 1, partition 11, NODE\_LOCAL, 7673 bytes)

24/04/28 21:21:52 INFO scheduler.TaskSetManager: Starting task 14.0 in stage 31.0 (TID 84, algo-1, executor 1, partition 14, NODE\_LOCAL, 7673 bytes)

24/04/28 21:21:52 INFO storage.BlockManagerInfo: Added broadcast\_27\_piece0 in memory on algo-1:40893 (size: 1922.0 B, free: 13.8 GB)

24/04/28 21:21:52 INFO spark.MapOutputTrackerMasterEndpoint: Asked to send map output locations for shuffle 10 to 10.0.222.169:41176

24/04/28 21:21:52 INFO scheduler.TaskSetManager: Starting task 15.0 in stage 31.0 (TID 85, algo-1, executor 1, partition 15, NODE\_LOCAL, 7673 bytes)

24/04/28 21:21:52 INFO scheduler.TaskSetManager: Finished task 14.0 in stage 31.0 (TID 84) in 35 ms on algo-1 (executor 1) (1/16)

24/04/28 21:21:52 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 31.0 (TID 86, algo-1, executor 1, partition 0, NODE\_LOCAL, 7673 bytes)

D 86, algo-1, executor 1, partition 0, PROCESS\_LOCAL, 7673 bytes)  
24/04/28 21:21:52 INFO scheduler.TaskSetManager: Finished task 8.0 in stage 31.0 (TID 80) in 37 ms on algo-1 (executor 1) (2/16)  
24/04/28 21:21:52 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 31.0 (TID 87, algo-1, executor 1, partition 1, PROCESS\_LOCAL, 7673 bytes)  
24/04/28 21:21:52 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 31.0 (TID 88, algo-1, executor 1, partition 3, PROCESS\_LOCAL, 7673 bytes)  
24/04/28 21:21:52 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 31.0 (TID 78) in 39 ms on algo-1 (executor 1) (3/16)  
24/04/28 21:21:52 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 31.0 (TID 77) in 40 ms on algo-1 (executor 1) (4/16)  
24/04/28 21:21:52 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 31.0 (TID 89, algo-1, executor 1, partition 4, PROCESS\_LOCAL, 7673 bytes)  
24/04/28 21:21:52 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 31.0 (TID 90, algo-1, executor 1, partition 6, PROCESS\_LOCAL, 7673 bytes)  
24/04/28 21:21:52 INFO scheduler.TaskSetManager: Starting task 12.0 in stage 31.0 (TID 91, algo-1, executor 1, partition 12, PROCESS\_LOCAL, 7673 bytes)  
24/04/28 21:21:52 INFO scheduler.TaskSetManager: Starting task 13.0 in stage 31.0 (TID 92, algo-1, executor 1, partition 13, PROCESS\_LOCAL, 7673 bytes)  
24/04/28 21:21:52 INFO scheduler.TaskSetManager: Finished task 11.0 in stage 31.0 (TID 83) in 41 ms on algo-1 (executor 1) (5/16)  
24/04/28 21:21:52 INFO scheduler.TaskSetManager: Finished task 10.0 in stage 31.0 (TID 82) in 42 ms on algo-1 (executor 1) (6/16)  
24/04/28 21:21:52 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 31.0 (TID 79) in 42 ms on algo-1 (executor 1) (7/16)  
24/04/28 21:21:52 INFO scheduler.TaskSetManager: Finished task 9.0 in stage 31.0 (TID 81) in 42 ms on algo-1 (executor 1) (8/16)  
24/04/28 21:21:52 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 31.0 (TID 86) in 8 ms on algo-1 (executor 1) (9/16)  
24/04/28 21:21:52 INFO scheduler.TaskSetManager: Finished task 15.0 in stage 31.0 (TID 85) in 10 ms on algo-1 (executor 1) (10/16)  
24/04/28 21:21:52 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 31.0 (TID 89) in 6 ms on algo-1 (executor 1) (11/16)  
24/04/28 21:21:52 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 31.0 (TID 88) in 9 ms on algo-1 (executor 1) (12/16)  
24/04/28 21:21:52 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 31.0 (TID 90) in 9 ms on algo-1 (executor 1) (13/16)  
24/04/28 21:21:52 INFO scheduler.TaskSetManager: Finished task 13.0 in stage 31.0 (TID 92) in 9 ms on algo-1 (executor 1) (14/16)  
24/04/28 21:21:52 INFO scheduler.TaskSetManager: Finished task 12.0 in stage 31.0 (TID 91) in 9 ms on algo-1 (executor 1) (15/16)  
24/04/28 21:21:52 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 31.0 (TID 87) in 13 ms on algo-1 (executor 1) (16/16)  
24/04/28 21:21:52 INFO cluster.YarnScheduler: Removed TaskSet 31.0, whose tasks have all completed, from pool  
24/04/28 21:21:52 INFO scheduler.DAGScheduler: ResultStage 31 (countByKey at ColumnProfiler.scala:605) finished in 0.056 s  
24/04/28 21:21:52 INFO scheduler.DAGScheduler: Job 19 finished: countByKey at ColumnProfiler.scala:605, took 3.923498 s  
24/04/28 21:21:53 INFO spark.SparkContext: Starting job: runJob at PythonRDD.scala:153  
24/04/28 21:21:53 INFO scheduler.DAGScheduler: Got job 20 (runJob at PythonRDD.scala:153) with 1 output partitions  
24/04/28 21:21:53 INFO scheduler.DAGScheduler: Final stage: ResultStage 32 (runJob at PythonRDD.scala:153)  
24/04/28 21:21:53 INFO scheduler.DAGScheduler: Parents of final stage: List()

```
24/04/28 21:21:53 INFO scheduler.DAGScheduler: Missing parents: List()
24/04/28 21:21:53 INFO scheduler.DAGScheduler: Submitting ResultStage 32 (PythonRDD
[68] at RDD at PythonRDD.scala:53), which has no missing parents
24/04/28 21:21:53 INFO memory.MemoryStore: Block broadcast_28 stored as values in me
mory (estimated size 5.0 KB, free 1007.8 MB)
24/04/28 21:21:53 INFO memory.MemoryStore: Block broadcast_28_piece0 stored as bytes
in memory (estimated size 3.5 KB, free 1007.8 MB)
24/04/28 21:21:53 INFO storage.BlockManagerInfo: Added broadcast_28_piece0 in memory
on 10.0.222.169:36237 (size: 3.5 KB, free: 1008.8 MB)
24/04/28 21:21:53 INFO spark.SparkContext: Created broadcast 28 from broadcast at DA
GScheduler.scala:1203
24/04/28 21:21:53 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Resul
tStage 32 (PythonRDD[68] at RDD at PythonRDD.scala:53) (first 15 tasks are for parti
tions Vector(0))
24/04/28 21:21:53 INFO cluster.YarnScheduler: Adding task set 32.0 with 1 tasks
24/04/28 21:21:53 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 32.0 (TI
D 93, algo-1, executor 1, partition 0, PROCESS_LOCAL, 8303 bytes)
24/04/28 21:21:53 INFO storage.BlockManagerInfo: Added broadcast_28_piece0 in memory
on algo-1:40893 (size: 3.5 KB, free: 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:50 INFO broadcast.TorrentBroadcast: Started reading
broadcast variable 25
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:50 INFO memory.MemoryStore: Block broadcast_25_piec
e0 stored as bytes in memory (estimated size 27.6 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:50 INFO broadcast.TorrentBroadcast: Reading broadca
st variable 25 took 7 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:50 INFO memory.MemoryStore: Block broadcast_25 stor
ed as values in memory (estimated size 398.1 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Finished task 2.0 in sta
ge 30.0 (TID 76). 1919 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Finished task 1.0 in sta
ge 30.0 (TID 75). 1919 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Finished task 0.0 in sta
ge 30.0 (TID 74). 1919 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 77
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Running task 2.0 in stag
e 31.0 (TID 77)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 78
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Running task 5.0 in stag
e 31.0 (TID 78)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 79
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
```

```
01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Running task 7.0 in stage 31.0 (TID 79)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 80
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 81
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 82
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Running task 8.0 in stage 31.0 (TID 80)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Running task 9.0 in stage 31.0 (TID 81)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 83
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Running task 11.0 in stage 31.0 (TID 83)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Running task 10.0 in stage 31.0 (TID 82)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 84
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Running task 14.0 in stage 31.0 (TID 84)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO spark.MapOutputTrackerWorker: Updating epoch to 11 and clearing cache
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 27
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO memory.MemoryStore: Block broadcast_27_piece0 stored as bytes in memory (estimated size 1922.0 B, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO broadcast.TorrentBroadcast: Reading broadcast variable 27 took 5 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO memory.MemoryStore: Block broadcast_27 stored as values in memory (estimated size 3.1 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO spark.MapOutputTrackerWorker: Don't have map outputs for shuffle 10, fetching them
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO spark.MapOutputTrackerWorker: Don't have map outputs for shuffle 10, fetching them
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO spark.MapOutputTrackerWorker: Doing the fetch; tracker endpoint = NettyRpcEndpointRef(spark://MapOutputTracker@10.0.222.169:374
```

```
11)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO spark.MapOutputTrackerWorker: Don't have map outputs for shuffle 10, fetching them
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO spark.MapOutputTrackerWorker: Don't have map outputs for shuffle 10, fetching them
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO spark.MapOutputTrackerWorker: Don't have map outputs for shuffle 10, fetching them
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO spark.MapOutputTrackerWorker: Don't have map outputs for shuffle 10, fetching them
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO spark.MapOutputTrackerWorker: Don't have map outputs for shuffle 10, fetching them
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO spark.MapOutputTrackerWorker: Don't have map outputs for shuffle 10, fetching them
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO spark.MapOutputTrackerWorker: Got the output locations
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Getting 1 non-empty blocks including 1 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
```



```
01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Starte
d 0 remote fetches in 5 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Gettin
g 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Gettin
g 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Starte
d 0 remote fetches in 7 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Starte
d 0 remote fetches in 7 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Finished task 2.0 in sta
ge 31.0 (TID 77). 1301 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Finished task 8.0 in sta
ge 31.0 (TID 80). 1362 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Finished task 14.0 in st
age 31.0 (TID 84). 1307 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Finished task 5.0 in sta
ge 31.0 (TID 78). 1321 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Finished task 10.0 in st
age 31.0 (TID 82). 1301 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Finished task 7.0 in sta
ge 31.0 (TID 79). 1371 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Finished task 11.0 in st
age 31.0 (TID 83). 1301 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Finished task 9.0 in sta
ge 31.0 (TID 81). 1347 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 85
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 86
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Running task 15.0 in sta
ge 31.0 (TID 85)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Running task 0.0 in stag
e 31.0 (TID 86)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 87
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Running task 1.0 in stag
e 31.0 (TID 87)
```

```
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks including 0 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Getting 3 non-empty blocks including 3 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 88
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 89
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Running task 4.0 in stage 31.0 (TID 89)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Running task 3.0 in stage 31.0 (TID 88)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Finished task 0.0 in stage 31.0 (TID 86). 1134 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks including 0 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Finished task 15.0 in stage 31.0 (TID 85). 1301 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Finished task 4.0 in stage 31.0 (TID 89). 1091 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks including 0 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 90
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Running task 6.0 in stage 31.0 (TID 90)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 91
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Running task 12.0 in sta
```

```
ge 31.0 (TID 91)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 92
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Finished task 3.0 in stage 31.0 (TID 88). 1134 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Running task 13.0 in stage 31.0 (TID 92)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks including 0 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks including 0 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks including 0 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Finished task 6.0 in stage 31.0 (TID 90). 1091 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks including 0 local blocks and 0 remote blocks
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Finished task 13.0 in stage 31.0 (TID 92). 1091 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Finished task 12.0 in stage 31.0 (TID 91). 1134 bytes result sent to driver
24/04/28 21:21:53 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 32.0 (TID 93) in 540 ms on algo-1 (executor 1) (1/1)
24/04/28 21:21:53 INFO cluster.YarnScheduler: Removed TaskSet 32.0, whose tasks have all completed, from pool
24/04/28 21:21:53 INFO python.PythonAccumulatorV2: Connected to AccumulatorServer at host: 127.0.0.1 port: 37069
24/04/28 21:21:53 INFO scheduler.DAGScheduler: ResultStage 32 (runJob at PythonRDD.scala:153) finished in 0.549 s
24/04/28 21:21:53 INFO scheduler.DAGScheduler: Job 20 finished: runJob at PythonRDD.scala:153, took 0.552135 s
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 712
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 761
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 751
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 663
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 715
24/04/28 21:21:53 INFO storage.BlockManagerInfo: Removed broadcast_22_piece0 on 10.
```

```
0.222.169:36237 in memory (size: 27.6 KB, free: 1008.8 MB)
24/04/28 21:21:53 INFO storage.BlockManagerInfo: Removed broadcast_22_piece0 on algo
-1:40893 in memory (size: 27.6 KB, free: 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000/usr/lib/spark/python/lib/pyspark.zip/pyspark/sql/session.py:366: UserWarning:
Using RDD of dict to inferSchema is deprecated. Use pyspark.sql.Row instead
  warnings.warn("Using RDD of dict to inferSchema is deprecated. ")
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 719
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 754
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 717
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 786
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 704
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 711
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 733
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 794
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 731
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 805
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 755
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 752
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 785
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 765
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 700
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 703
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 705
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 654
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 718
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 720
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 768
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 699
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 753
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 791
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 777
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 781
24/04/28 21:21:53 INFO storage.BlockManagerInfo: Removed broadcast_27_piece0 on 10.
0.222.169:36237 in memory (size: 1922.0 B, free: 1008.8 MB)
24/04/28 21:21:53 INFO storage.BlockManagerInfo: Removed broadcast_27_piece0 on algo
-1:40893 in memory (size: 1922.0 B, free: 13.8 GB)
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 650
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 798
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 762
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 664
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 739
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 659
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 661
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 747
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 707
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 796
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 757
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned shuffle 10
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 724
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 787
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 763
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 772
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 716
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 774
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 736
```

```
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 710
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned shuffle 9
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 738
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 709
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 706
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 660
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 799
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 775
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 766
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 771
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 804
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 793
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 764
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 776
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 789
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 667
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 746
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 708
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 655
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 653
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 698
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 669
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 769
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 696
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 701
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 759
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 695
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 725
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 652
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 745
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 778
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 741
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 651
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 649
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 788
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 665
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 801
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 760
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 656
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 749
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 742
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 668
24/04/28 21:21:53 INFO storage.BlockManagerInfo: Removed broadcast_26_piece0 on 10.
0.222.169:36237 in memory (size: 10.5 KB, free: 1008.8 MB)
24/04/28 21:21:53 INFO storage.BlockManagerInfo: Removed broadcast_26_piece0 on algo
-1:40893 in memory (size: 10.5 KB, free: 13.8 GB)
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 721
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 662
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 744
24/04/28 21:21:53 INFO storage.BlockManagerInfo: Removed broadcast_24_piece0 on 10.
0.222.169:36237 in memory (size: 16.2 KB, free: 1008.8 MB)
24/04/28 21:21:53 INFO storage.BlockManagerInfo: Removed broadcast_24_piece0 on algo
-1:40893 in memory (size: 16.2 KB, free: 13.8 GB)
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 802
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 748
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 648
```

```
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 770
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 714
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 803
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 743
24/04/28 21:21:53 INFO storage.BlockManagerInfo: Removed broadcast_28_piece0 on 10.0.222.169:36237 in memory (size: 3.5 KB, free: 1008.8 MB)
24/04/28 21:21:53 INFO storage.BlockManagerInfo: Removed broadcast_28_piece0 on algo-1:40893 in memory (size: 3.5 KB, free: 13.8 GB)
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 780
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 773
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 795
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 767
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 740
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 790
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 666
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 750
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 722
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 647
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 697
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 758
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 735
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 737
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 779
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 783
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 734
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 657
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 658
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 800
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 756
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 792
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 723
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 732
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 797
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 782
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 702
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 784
24/04/28 21:21:53 INFO spark.ContextCleaner: Cleaned accumulator 713
24/04/28 21:21:53 INFO codegen.CodeGenerator: Code generated in 7.152389 ms
24/04/28 21:21:53 INFO spark.SparkContext: Starting job: showString at NativeMethodAccessorImpl.java:0
24/04/28 21:21:53 INFO scheduler.DAGScheduler: Got job 21 (showString at NativeMethodAccessorImpl.java:0) with 1 output partitions
24/04/28 21:21:53 INFO scheduler.DAGScheduler: Final stage: ResultStage 33 (showString at NativeMethodAccessorImpl.java:0)
24/04/28 21:21:53 INFO scheduler.DAGScheduler: Parents of final stage: List()
24/04/28 21:21:53 INFO scheduler.DAGScheduler: Missing parents: List()
24/04/28 21:21:53 INFO scheduler.DAGScheduler: Submitting ResultStage 33 (MapPartitionsRDD[75] at showString at NativeMethodAccessorImpl.java:0), which has no missing parents
24/04/28 21:21:53 INFO memory.MemoryStore: Block broadcast_29 stored as values in memory (estimated size 14.1 KB, free 1008.2 MB)
24/04/28 21:21:53 INFO memory.MemoryStore: Block broadcast_29_piece0 stored as bytes in memory (estimated size 8.2 KB, free 1008.2 MB)
24/04/28 21:21:53 INFO storage.BlockManagerInfo: Added broadcast_29_piece0 in memory on 10.0.222.169:36237 (size: 8.2 KB, free: 1008.8 MB)
24/04/28 21:21:53 INFO spark.SparkContext: Created broadcast 29 from broadcast at DA
```

```
GScheduler.scala:1203
24/04/28 21:21:53 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 33 (MapPartitionsRDD[75] at showString at NativeMethodAccessorImpl.java:0) (first 15 tasks are for partitions Vector(0))
24/04/28 21:21:53 INFO cluster.YarnScheduler: Adding task set 33.0 with 1 tasks
24/04/28 21:21:53 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 33.0 (TID 94, algo-1, executor 1, partition 0, PROCESS_LOCAL, 8303 bytes)
24/04/28 21:21:53 INFO storage.BlockManagerInfo: Added broadcast_29_piece0 in memory on algo-1:40893 (size: 8.2 KB, free: 13.8 GB)
24/04/28 21:21:53 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 33.0 (TID 94) in 67 ms on algo-1 (executor 1) (1/1)
24/04/28 21:21:53 INFO cluster.YarnScheduler: Removed TaskSet 33.0, whose tasks have all completed, from pool
24/04/28 21:21:53 INFO scheduler.DAGScheduler: ResultStage 33 (showString at NativeMethodAccessorImpl.java:0) finished in 0.077 s
24/04/28 21:21:53 INFO scheduler.DAGScheduler: Job 21 finished: showString at NativeMethodAccessorImpl.java:0, took 0.079026 s
24/04/28 21:21:53 INFO spark.SparkContext: Starting job: showString at NativeMethodAccessorImpl.java:0
24/04/28 21:21:53 INFO scheduler.DAGScheduler: Got job 22 (showString at NativeMethodAccessorImpl.java:0) with 4 output partitions
24/04/28 21:21:53 INFO scheduler.DAGScheduler: Final stage: ResultStage 34 (showString at NativeMethodAccessorImpl.java:0)
24/04/28 21:21:53 INFO scheduler.DAGScheduler: Parents of final stage: List()
24/04/28 21:21:53 INFO scheduler.DAGScheduler: Missing parents: List()
24/04/28 21:21:53 INFO scheduler.DAGScheduler: Submitting ResultStage 34 (MapPartitionsRDD[75] at showString at NativeMethodAccessorImpl.java:0), which has no missing parents
24/04/28 21:21:53 INFO memory.MemoryStore: Block broadcast_30 stored as values in memory (estimated size 14.1 KB, free 1008.2 MB)
24/04/28 21:21:53 INFO memory.MemoryStore: Block broadcast_30_piece0 stored as bytes in memory (estimated size 8.2 KB, free 1008.2 MB)
24/04/28 21:21:53 INFO storage.BlockManagerInfo: Added broadcast_30_piece0 in memory on 10.0.222.169:36237 (size: 8.2 KB, free: 1008.8 MB)
24/04/28 21:21:53 INFO spark.SparkContext: Created broadcast 30 from broadcast at DAGScheduler.scala:1203
24/04/28 21:21:53 INFO scheduler.DAGScheduler: Submitting 4 missing tasks from ResultStage 34 (MapPartitionsRDD[75] at showString at NativeMethodAccessorImpl.java:0) (first 15 tasks are for partitions Vector(1, 2, 3, 4))
24/04/28 21:21:53 INFO cluster.YarnScheduler: Adding task set 34.0 with 4 tasks
24/04/28 21:21:53 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 34.0 (TID 95, algo-1, executor 1, partition 1, PROCESS_LOCAL, 8870 bytes)
24/04/28 21:21:53 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 34.0 (TID 96, algo-1, executor 1, partition 2, PROCESS_LOCAL, 8879 bytes)
24/04/28 21:21:53 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 34.0 (TID 97, algo-1, executor 1, partition 3, PROCESS_LOCAL, 8767 bytes)
24/04/28 21:21:53 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 34.0 (TID 98, algo-1, executor 1, partition 4, PROCESS_LOCAL, 8383 bytes)
24/04/28 21:21:53 INFO storage.BlockManagerInfo: Added broadcast_30_piece0 in memory on algo-1:40893 (size: 8.2 KB, free: 13.8 GB)
24/04/28 21:21:53 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 34.0 (TID 97) in 31 ms on algo-1 (executor 1) (1/4)
24/04/28 21:21:53 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 34.0 (TID 95) in 37 ms on algo-1 (executor 1) (2/4)
24/04/28 21:21:53 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 34.0 (TID 98) in 36 ms on algo-1 (executor 1) (3/4)
```

24/04/28 21:21:53 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 34.0 (TID 96) in 58 ms on algo-1 (executor 1) (4/4)

24/04/28 21:21:53 INFO cluster.YarnScheduler: Removed TaskSet 34.0, whose tasks have all completed, from pool

24/04/28 21:21:53 INFO scheduler.DAGScheduler: ResultStage 34 (showString at NativeMethodAccessorImpl.java:0) finished in 0.062 s

24/04/28 21:21:53 INFO scheduler.DAGScheduler: Job 22 finished: showString at NativeMethodAccessorImpl.java:0, took 0.065087 s

24/04/28 21:21:53 INFO spark.SparkContext: Starting job: showString at NativeMethodAccessorImpl.java:0

24/04/28 21:21:53 INFO scheduler.DAGScheduler: Got job 23 (showString at NativeMethodAccessorImpl.java:0) with 11 output partitions

24/04/28 21:21:53 INFO scheduler.DAGScheduler: Final stage: ResultStage 35 (showString at NativeMethodAccessorImpl.java:0)

24/04/28 21:21:53 INFO scheduler.DAGScheduler: Parents of final stage: List()

24/04/28 21:21:53 INFO scheduler.DAGScheduler: Missing parents: List()

24/04/28 21:21:53 INFO scheduler.DAGScheduler: Submitting ResultStage 35 (MapPartitionsRDD[75] at showString at NativeMethodAccessorImpl.java:0), which has no missing parents

24/04/28 21:21:53 INFO memory.MemoryStore: Block broadcast\_31 stored as values in memory (estimated size 14.1 KB, free 1008.2 MB)

24/04/28 21:21:53 INFO memory.MemoryStore: Block broadcast\_31\_piece0 stored as bytes in memory (estimated size 8.2 KB, free 1008.2 MB)

24/04/28 21:21:53 INFO storage.BlockManagerInfo: Added broadcast\_31\_piece0 in memory on 10.0.222.169:36237 (size: 8.2 KB, free: 1008.8 MB)

24/04/28 21:21:53 INFO spark.SparkContext: Created broadcast 31 from broadcast at DAGScheduler.scala:1203

24/04/28 21:21:53 INFO scheduler.DAGScheduler: Submitting 11 missing tasks from ResultStage 35 (MapPartitionsRDD[75] at showString at NativeMethodAccessorImpl.java:0) (first 15 tasks are for partitions Vector(5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15))

24/04/28 21:21:53 INFO cluster.YarnScheduler: Adding task set 35.0 with 11 tasks

24/04/28 21:21:53 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 35.0 (TID 99, algo-1, executor 1, partition 5, PROCESS\_LOCAL, 8821 bytes)

24/04/28 21:21:53 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 35.0 (TID 100, algo-1, executor 1, partition 6, PROCESS\_LOCAL, 8783 bytes)

24/04/28 21:21:53 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 35.0 (TID 101, algo-1, executor 1, partition 7, PROCESS\_LOCAL, 8755 bytes)

24/04/28 21:21:53 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 35.0 (TID 102, algo-1, executor 1, partition 8, PROCESS\_LOCAL, 8373 bytes)

24/04/28 21:21:53 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 35.0 (TID 103, algo-1, executor 1, partition 9, PROCESS\_LOCAL, 9070 bytes)

24/04/28 21:21:53 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 35.0 (TID 104, algo-1, executor 1, partition 10, PROCESS\_LOCAL, 9247 bytes)

24/04/28 21:21:53 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 35.0 (TID 105, algo-1, executor 1, partition 11, PROCESS\_LOCAL, 8907 bytes)

24/04/28 21:21:53 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 35.0 (TID 106, algo-1, executor 1, partition 12, PROCESS\_LOCAL, 8508 bytes)

24/04/28 21:21:53 INFO storage.BlockManagerInfo: Added broadcast\_31\_piece0 in memory on algo-1:40893 (size: 8.2 KB, free: 13.8 GB)

24/04/28 21:21:53 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 35.0 (TID 107, algo-1, executor 1, partition 13, PROCESS\_LOCAL, 8758 bytes)

24/04/28 21:21:53 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 35.0 (TID 106) in 33 ms on algo-1 (executor 1) (1/11)

24/04/28 21:21:53 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 35.0 (TID 108, algo-1, executor 1, partition 14, PROCESS\_LOCAL, 8825 bytes)

24/04/28 21:21:53 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 35.0 (TID 99) in 33 ms on algo-1 (executor 1) (1/11)



```
D 99) in 36 ms on algo-1 (executor 1) (2/11)
24/04/28 21:21:53 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 35.0 (T
ID 109, algo-1, executor 1, partition 15, PROCESS_LOCAL, 8895 bytes)
24/04/28 21:21:53 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 35.0 (TI
D 102) in 44 ms on algo-1 (executor 1) (3/11)
24/04/28 21:21:53 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 35.0 (TI
D 100) in 46 ms on algo-1 (executor 1) (4/11)
24/04/28 21:21:53 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 35.0 (TI
D 103) in 61 ms on algo-1 (executor 1) (5/11)
24/04/28 21:21:53 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 35.0 (TI
D 105) in 73 ms on algo-1 (executor 1) (6/11)
24/04/28 21:21:53 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 35.0 (TI
D 101) in 75 ms on algo-1 (executor 1) (7/11)
24/04/28 21:21:53 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 35.0 (TI
D 104) in 79 ms on algo-1 (executor 1) (8/11)
24/04/28 21:21:53 INFO scheduler.TaskSetManager: Finished task 9.0 in stage 35.0 (TI
D 108) in 63 ms on algo-1 (executor 1) (9/11)
24/04/28 21:21:54 INFO scheduler.TaskSetManager: Finished task 10.0 in stage 35.0 (T
ID 109) in 55 ms on algo-1 (executor 1) (10/11)
24/04/28 21:21:54 INFO scheduler.TaskSetManager: Finished task 8.0 in stage 35.0 (TI
D 107) in 67 ms on algo-1 (executor 1) (11/11)
24/04/28 21:21:54 INFO cluster.YarnScheduler: Removed TaskSet 35.0, whose tasks have
all completed, from pool
24/04/28 21:21:54 INFO scheduler.DAGScheduler: ResultStage 35 (showString at NativeM
ethodAccessorImpl.java:0) finished in 0.107 s
24/04/28 21:21:54 INFO scheduler.DAGScheduler: Job 23 finished: showString at Native
MethodAccessorImpl.java:0, took 0.109140 s

+-----+-----+-----+
|code_for_constraint                                     |
|column_name      |constraint_name          |description
|current_value                                |
|rule_description
|suggesting_rule    |
+-----+-----+-----+
|isComplete("review_id")
|review_id         |CompletenessConstraint(Completeness(review_id,None))
|Completeness: 1.0           |'review_id' is not null
|If a column is complete in the sample, we suggest a NOT NULL constraint
|CompleteIfCompleteRule()   |
|.isUnique("review id")
```

```

|review_id      |UniquenessConstraint(Uniqueness(List(review_id),None))
|ApproxDistinctness: 0.9624383196209793|'review_id' is unique
|If the ratio of approximate num distinct values in a column is close to the number
of records (within the error of the HLL sketch), we suggest a UNIQUE constraint|Uniq
ueIfApproximatelyUniqueRule() |
|.isComplete("customer_id")
|customer_id    |CompletenessConstraint(Completeness(customer_id,None))
|Completeness: 1.0      |'customer_id' is not null
|If a column is complete in the sample, we suggest a NOT NULL constraint
|CompleteIfCompleteRule() |
|.isNonNegative("customer_id")
|customer_id    |ComplianceConstraint(Compliance('customer_id' has no negative valu
es,customer_id >= 0,None))
|Minimum: 10229.0      |'customer_id' has no negative values
|If we see only non-negative numbers in a column, we suggest a corresponding constrai
nt                                     |NonN
egativeNumbersRule() |
|.hasDataType("customer_id", ConstrainableDataTypes.Integral)
|customer_id    |AnalysisBasedConstraint(DataType(customer_id,None),<function1>,Som
e(<function1>),None)
|DataType: Integral      |'customer_id' has type Integral
|If we detect a non-string type, we suggest a type constraint
|RetainTypeRule() |
|.isComplete("review_date")
|review_date    |CompletenessConstraint(Completeness(review_date,None))
|Completeness: 1.0      |'review_date' is not null
|If a column is complete in the sample, we suggest a NOT NULL constraint
|CompleteIfCompleteRule() |
|.isComplete("helpful_votes")
|helpful_votes  |CompletenessConstraint(Completeness(helpful_votes,None))
|Completeness: 1.0      |'helpful_votes' is not null
|If a column is complete in the sample, we suggest a NOT NULL constraint
|CompleteIfCompleteRule() |
|.isNonNegative("helpful_votes")
|helpful_votes  |ComplianceConstraint(Compliance('helpful_votes' has no negative va
lues,helpful_votes >= 0,None))
|Minimum: 0.0      |'helpful_votes' has no negative values
|If we see only non-negative numbers in a column, we suggest a corresponding constrai
nt                                     |NonN
egativeNumbersRule() |
|.isComplete("star_rating")
|star_rating    |CompletenessConstraint(Completeness(star_rating,None))
|Completeness: 1.0      |'star_rating' is not null
|If a column is complete in the sample, we suggest a NOT NULL constraint
|CompleteIfCompleteRule() |
|.isNonNegative("star_rating")
|star_rating    |ComplianceConstraint(Compliance('star_rating' has no negative valu
es,star_rating >= 0,None))
|Minimum: 1.0      |'star_rating' has no negative values
|If we see only non-negative numbers in a column, we suggest a corresponding constrai
nt                                     |NonN
egativeNumbersRule() |
|.isComplete("product_title")
|product_title  |CompletenessConstraint(Completeness(product_title,None))
|Completeness: 1.0      |'product_title' is not null
|If a column is complete in the sample, we suggest a NOT NULL constraint

```

```

|CompleteIfCompleteRule()          |
|.isComplete("review_headline")
|review_headline |CompletenessConstraint(Completeness(review_headline,None))
|Completeness: 1.0                      |'review_headline' is not null
|If a column is complete in the sample, we suggest a NOT NULL constraint
|CompleteIfCompleteRule()          |
|.isComplete("product_id")
|product_id      |CompletenessConstraint(Completeness(product_id,None))
|Completeness: 1.0                      |'product_id' is not null
|If a column is complete in the sample, we suggest a NOT NULL constraint
|CompleteIfCompleteRule()          |
|.isComplete("total_votes")
|total_votes     |CompletenessConstraint(Completeness(total_votes,None))
|Completeness: 1.0                      |'total_votes' is not null
|If a column is complete in the sample, we suggest a NOT NULL constraint
|CompleteIfCompleteRule()          |
|.isNonNegative("total_votes")
|total_votes     |ComplianceConstraint(Compliance('total_votes' has no negative values, total_votes >= 0, None))
|Minimum: 0.0                      |'total_votes' has no negative values
|If we see only non-negative numbers in a column, we suggest a corresponding constraint
|NonNegativeNumbersRule()          |
|.isContainedIn("product_category", ["Gift Card", "Digital_Video_Games", "Digital_Software"])
|product_category |ComplianceConstraint(Compliance('product_category' has value range 'Gift Card', 'Digital_Video_Games', 'Digital_Software', 'product_category' IN ('Gift Card', 'Digital_Video_Games', 'Digital_Software'), None))
|Compliance: 1
|'product_category' has value range 'Gift Card', 'Digital_Video_Games', 'Digital_Software'
|If we see a categorical range for a column, we suggest an IS IN (...) constraint
|CategoricalRangeRule()          |
|.isComplete("product_category")
|product_category |CompletenessConstraint(Completeness(product_category, None))
|Completeness: 1.0                      |'product_category' is not null
|If a column is complete in the sample, we suggest a NOT NULL constraint
|CompleteIfCompleteRule()          |
|.isContainedIn("product_category", ["Gift Card", "Digital_Video_Games", "Digital_Software"], lambda x: x >= 0.99, "It should be above 0.99!")
|product_category |ComplianceConstraint(Compliance('product_category' has value range 'Gift Card', 'Digital_Video_Games', 'Digital_Software' for at least 99.0% of values, 'product_category' IN ('Gift Card', 'Digital_Video_Games', 'Digital_Software'), None))
|Compliance: 0.999999999
|'product_category' has value range 'Gift Card', 'Digital_Video_Games', 'Digital_Software' for at least 99.0% of values
|If we see a categorical range for most values in a column, we suggest an IS IN (...) constraint that should hold for most values
|FractionalCategoricalRangeRule(0.9) |
|.isComplete("product_parent")
|product_parent   |CompletenessConstraint(Completeness(product_parent, None))
|Completeness: 1.0                      |'product_parent' is not null
|If a column is complete in the sample, we suggest a NOT NULL constraint
|CompleteIfCompleteRule()          |
|.isNonNegative("product_parent")
|product_parent   |ComplianceConstraint(Compliance('product_parent' has no negative values, product_parent >= 0, None))
|Minimum: 209709.0                      |'product_parent' has no negative values
|If we see only non-negative numbers in a column, we suggest a corresponding constraint

```

```

int | NonN
negativeNumbersRule() |
+-----+-----+-----+
+-----+-----+-----+
+-----+-----+-----+
+-----+-----+-----+
+-----+-----+-----+
+-----+-----+-----+
+-----+-----+-----+
+-----+-----+-----+
+-----+-----+-----+
+-----+-----+-----+
+-----+-----+-----+

```

only showing top 20 rows

24/04/28 21:21:54 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2

24/04/28 21:21:54 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup \_temporary folders under output directory:false, ignore cleanup failures: false

24/04/28 21:21:54 INFO output.DirectFileOutputCommitter: Direct Write: DISABLED

24/04/28 21:21:54 INFO datasources.SQLConfCommitterProvider: Using output committer class org.apache.hadoop.mapreduce.lib.output.DirectFileOutputCommitter

002/stderr] 24/04/28 21:21:52 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 4 ms

[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:52 INFO executor.Executor: Finished task 1.0 in stage 31.0 (TID 87). 1134 bytes result sent to driver

[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:53 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 93

[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Running task 0.0 in stage 32.0 (TID 93)

[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:53 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 28

[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:53 INFO memory.MemoryStore: Block broadcast\_28\_piece0 stored as bytes in memory (estimated size 3.5 KB, free 13.8 GB)

[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:53 INFO broadcast.TorrentBroadcast: Reading broadcast variable 28 took 5 ms

[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:53 INFO memory.MemoryStore: Block broadcast\_28 stored as values in memory (estimated size 5.0 KB, free 13.8 GB)

[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:53 INFO python.PythonRunner: Times: total = 497, boot = 433, init = 64, finish = 0

[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Finished task 0.0 in stage 32.0 (TID 93). 1842 bytes result sent to driver

[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:53 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 94

[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Running task 0.0 in stage 33.0 (TID 94)

[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_01\_000002/stderr] 24/04/28 21:21:53 INFO broadcast.TorrentBroadcast: Started reading

```
broadcast variable 29
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO memory.MemoryStore: Block broadcast_29_piece0 stored as bytes in memory (estimated size 8.2 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO broadcast.TorrentBroadcast: Reading broadcast variable 29 took 7 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO memory.MemoryStore: Block broadcast_29 stored as values in memory (estimated size 14.1 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO codegen.CodeGenerator: Code generated in 5.177118 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO python.PythonRunner: Times: total = 6, boot = 3, init = 3, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Finished task 0.0 in stage 33.0 (TID 94). 2033 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 95
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 96
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Running task 1.0 in stage 34.0 (TID 96)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 97
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Running task 2.0 in stage 34.0 (TID 97)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 98
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Running task 3.0 in stage 34.0 (TID 98)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 30
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Running task 0.0 in stage 34.0 (TID 95)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO memory.MemoryStore: Block broadcast_30_piece0 stored as bytes in memory (estimated size 8.2 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO broadcast.TorrentBroadcast: Reading broadcast variable 30 took 7 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO memory.MemoryStore: Block broadcast_30 stored as values in memory (estimated size 14.1 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
```

```
01_000002/stderr] 24/04/28 21:21:53 INFO python.PythonRunner: Times: total = 7, boot
= 3, init = 4, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Finished task 2.0 in sta
ge 34.0 (TID 97). 2118 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:53 INFO python.PythonRunner: Times: total = 21, boo
t = 9, init = 12, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:53 INFO python.PythonRunner: Times: total = 19, boo
t = 6, init = 13, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Finished task 0.0 in sta
ge 34.0 (TID 95). 2356 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Finished task 3.0 in sta
ge 34.0 (TID 98). 2076 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:53 INFO python.PythonRunner: Times: total = 42, boo
t = -52, init = 94, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Finished task 1.0 in sta
ge 34.0 (TID 96). 2301 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:53 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 99
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Running task 0.0 in stag
e 35.0 (TID 99)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:53 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 100
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Running task 1.0 in stag
e 35.0 (TID 100)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:53 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 101
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:53 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 102
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Running task 3.0 in stag
e 35.0 (TID 102)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:53 INFO broadcast.TorrentBroadcast: Started reading
broadcast variable 31
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Running task 2.0 in stag
e 35.0 (TID 101)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:53 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 103
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
01_000002/stderr] 24/04/28 21:21:53 INFO executor.CoarseGrainedExecutorBackend: Got
assigned task 104
```

```
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Running task 4.0 in stage 35.0 (TID 103)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 105
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 106
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Running task 6.0 in stage 35.0 (TID 105)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Running task 7.0 in stage 35.0 (TID 106)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Running task 5.0 in stage 35.0 (TID 104)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO memory.MemoryStore: Block broadcast_31_piece0 stored as bytes in memory (estimated size 8.2 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO broadcast.TorrentBroadcast: Reading broadcast variable 31 took 10 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO memory.MemoryStore: Block broadcast_31 stored as values in memory (estimated size 14.1 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO python.PythonRunner: Times: total = 7, boot = 4, init = 3, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO python.PythonRunner: Times: total = 11, boot = 6, init = 5, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO python.PythonRunner: Times: total = 13, boot = 9, init = 4, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Finished task 0.0 in stage 35.0 (TID 99). 2240 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Finished task 7.0 in stage 35.0 (TID 106). 2217 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Finished task 3.0 in stage 35.0 (TID 102). 2083 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO python.PythonRunner: Times: total = 22, boot = 13, init = 9, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Finished task 1.0 in stage 35.0 (TID 100). 2139 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 107
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.CoarseGrainedExecutorBackend: Got
```

```
assigned task 108
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Running task 8.0 in stage 35.0 (TID 107)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Running task 9.0 in stage 35.0 (TID 108)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 109
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Running task 10.0 in stage 35.0 (TID 109)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO python.PythonRunner: Times: total = 45, boot = -32, init = 77, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Finished task 4.0 in stage 35.0 (TID 103). 2333 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO python.PythonRunner: Times: total = 49, boot = -33, init = 82, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO python.PythonRunner: Times: total = 49, boot = -18, init = 67, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Finished task 6.0 in stage 35.0 (TID 105). 2305 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Finished task 2.0 in stage 35.0 (TID 101). 2127 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO python.PythonRunner: Times: total = 57, boot = -46, init = 103, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Finished task 5.0 in stage 35.0 (TID 104). 2478 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO python.PythonRunner: Times: total = 47, boot = -5, init = 52, finish = 0
24/04/28 21:21:54 INFO scheduler.DAGScheduler: Registering RDD 77 (save at NativeMethodAccessorImpl.java:0) as input to shuffle 11
24/04/28 21:21:54 INFO scheduler.DAGScheduler: Got map stage job 24 (save at NativeMethodAccessorImpl.java:0) with 16 output partitions
24/04/28 21:21:54 INFO scheduler.DAGScheduler: Final stage: ShuffleMapStage 36 (save at NativeMethodAccessorImpl.java:0)
24/04/28 21:21:54 INFO scheduler.DAGScheduler: Parents of final stage: List()
24/04/28 21:21:54 INFO scheduler.DAGScheduler: Missing parents: List()
24/04/28 21:21:54 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 36 (MapPartitionsRDD[77] at save at NativeMethodAccessorImpl.java:0), which has no missing parents
24/04/28 21:21:54 INFO memory.MemoryStore: Block broadcast_32 stored as values in memory (estimated size 15.8 KB, free 1008.2 MB)
24/04/28 21:21:54 INFO memory.MemoryStore: Block broadcast_32_piece0 stored as bytes in memory (estimated size 9.3 KB, free 1008.2 MB)
24/04/28 21:21:54 INFO storage.BlockManagerInfo: Added broadcast_32_piece0 in memory
```



on 10.0.222.169:36237 (size: 9.3 KB, free: 1008.8 MB)

24/04/28 21:21:54 INFO spark.SparkContext: Created broadcast 32 from broadcast at DAGScheduler.scala:1203

24/04/28 21:21:54 INFO scheduler.DAGScheduler: Submitting 16 missing tasks from ShuffleMapStage 36 (MapPartitionsRDD[77] at save at NativeMethodAccessorImpl.java:0) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14))

24/04/28 21:21:54 INFO cluster.YarnScheduler: Adding task set 36.0 with 16 tasks

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 36.0 (TID 110, algo-1, executor 1, partition 0, PROCESS\_LOCAL, 8292 bytes)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 36.0 (TID 111, algo-1, executor 1, partition 1, PROCESS\_LOCAL, 8859 bytes)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 36.0 (TID 112, algo-1, executor 1, partition 2, PROCESS\_LOCAL, 8868 bytes)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 36.0 (TID 113, algo-1, executor 1, partition 3, PROCESS\_LOCAL, 8756 bytes)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 36.0 (TID 114, algo-1, executor 1, partition 4, PROCESS\_LOCAL, 8372 bytes)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 36.0 (TID 115, algo-1, executor 1, partition 5, PROCESS\_LOCAL, 8810 bytes)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 36.0 (TID 116, algo-1, executor 1, partition 6, PROCESS\_LOCAL, 8772 bytes)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 36.0 (TID 117, algo-1, executor 1, partition 7, PROCESS\_LOCAL, 8744 bytes)

24/04/28 21:21:54 INFO storage.BlockManagerInfo: Added broadcast\_32\_piece0 in memory on algo-1:40893 (size: 9.3 KB, free: 13.8 GB)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 36.0 (TID 118, algo-1, executor 1, partition 8, PROCESS\_LOCAL, 8362 bytes)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 36.0 (TID 119, algo-1, executor 1, partition 9, PROCESS\_LOCAL, 9059 bytes)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 36.0 (TID 117) in 64 ms on algo-1 (executor 1) (1/16)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 36.0 (TID 120, algo-1, executor 1, partition 10, PROCESS\_LOCAL, 9236 bytes)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 36.0 (TID 112) in 66 ms on algo-1 (executor 1) (2/16)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 36.0 (TID 111) in 66 ms on algo-1 (executor 1) (3/16)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 36.0 (TID 121, algo-1, executor 1, partition 11, PROCESS\_LOCAL, 8896 bytes)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 36.0 (TID 113) in 70 ms on algo-1 (executor 1) (4/16)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Starting task 12.0 in stage 36.0 (TID 122, algo-1, executor 1, partition 12, PROCESS\_LOCAL, 8497 bytes)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 36.0 (TID 115) in 73 ms on algo-1 (executor 1) (5/16)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Starting task 13.0 in stage 36.0 (TID 123, algo-1, executor 1, partition 13, PROCESS\_LOCAL, 8747 bytes)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 36.0 (TID 110) in 81 ms on algo-1 (executor 1) (6/16)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Starting task 14.0 in stage 36.0 (TID 124, algo-1, executor 1, partition 14, PROCESS\_LOCAL, 8814 bytes)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 36.0 (TID 114) in 84 ms on algo-1 (executor 1) (7/16)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Starting task 15.0 in stage 36.0 (TID 125, algo-1, executor 1, partition 15, PROCESS\_LOCAL, 8884 bytes)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 36.0 (TID 116) in 84 ms on algo-1 (executor 1) (8/16)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Finished task 8.0 in stage 36.0 (TID 118) in 51 ms on algo-1 (executor 1) (9/16)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Finished task 10.0 in stage 36.0 (TID 120) in 58 ms on algo-1 (executor 1) (10/16)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Finished task 9.0 in stage 36.0 (TID 119) in 62 ms on algo-1 (executor 1) (11/16)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Finished task 11.0 in stage 36.0 (TID 121) in 57 ms on algo-1 (executor 1) (12/16)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Finished task 13.0 in stage 36.0 (TID 123) in 50 ms on algo-1 (executor 1) (13/16)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Finished task 12.0 in stage 36.0 (TID 122) in 60 ms on algo-1 (executor 1) (14/16)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Finished task 15.0 in stage 36.0 (TID 125) in 53 ms on algo-1 (executor 1) (15/16)

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Finished task 14.0 in stage 36.0 (TID 124) in 58 ms on algo-1 (executor 1) (16/16)

24/04/28 21:21:54 INFO cluster.YarnScheduler: Removed TaskSet 36.0, whose tasks have all completed, from pool

24/04/28 21:21:54 INFO scheduler.DAGScheduler: ShuffleMapStage 36 (save at NativeMethodAccessorImpl.java:0) finished in 0.148 s

24/04/28 21:21:54 INFO scheduler.DAGScheduler: looking for newly runnable stages

24/04/28 21:21:54 INFO scheduler.DAGScheduler: running: Set()

24/04/28 21:21:54 INFO scheduler.DAGScheduler: waiting: Set()

24/04/28 21:21:54 INFO scheduler.DAGScheduler: failed: Set()

24/04/28 21:21:54 INFO spark.SparkContext: Starting job: save at NativeMethodAccessorImpl.java:0

24/04/28 21:21:54 INFO scheduler.DAGScheduler: Got job 25 (save at NativeMethodAccessorImpl.java:0) with 1 output partitions

24/04/28 21:21:54 INFO scheduler.DAGScheduler: Final stage: ResultStage 38 (save at NativeMethodAccessorImpl.java:0)

24/04/28 21:21:54 INFO scheduler.DAGScheduler: Parents of final stage: List(ShuffleMapStage 37)

24/04/28 21:21:54 INFO scheduler.DAGScheduler: Missing parents: List()

24/04/28 21:21:54 INFO scheduler.DAGScheduler: Submitting ResultStage 38 (ShuffledRowRDD[78] at save at NativeMethodAccessorImpl.java:0), which has no missing parents

24/04/28 21:21:54 INFO memory.MemoryStore: Block broadcast\_33 stored as values in memory (estimated size 167.8 KB, free 1008.0 MB)

24/04/28 21:21:54 INFO memory.MemoryStore: Block broadcast\_33\_piece0 stored as bytes in memory (estimated size 60.5 KB, free 1007.9 MB)

24/04/28 21:21:54 INFO storage.BlockManagerInfo: Added broadcast\_33\_piece0 in memory on 10.0.222.169:36237 (size: 60.5 KB, free: 1008.8 MB)

24/04/28 21:21:54 INFO spark.SparkContext: Created broadcast 33 from broadcast at DAGScheduler.scala:1203

24/04/28 21:21:54 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 38 (ShuffledRowRDD[78] at save at NativeMethodAccessorImpl.java:0) (first 15 tasks are for partitions Vector())

24/04/28 21:21:54 INFO cluster.YarnScheduler: Adding task set 38.0 with 1 tasks

24/04/28 21:21:54 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 38.0 (TID 126, algo-1, executor 1, partition 0, NODE\_LOCAL, 7778 bytes)

24/04/28 21:21:54 INFO storage.BlockManagerInfo: Added broadcast\_33\_piece0 in memory on algo-1:40893 (size: 60.5 KB, free: 13.8 GB)

24/04/28 21:21:54 INFO spark.MapOutputTrackerMasterEndpoint: Asked to send map output locations for shuffle 11 to 10.0.222.169:41176

[/var/log/yarn/userlogs/application\_1714339251228\_0001/container\_1714339251228\_0001\_

```
01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Finished task 9.0 in stage 35.0 (TID 108). 2273 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO python.PythonRunner: Times: total = 46, boot = -8, init = 54, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO executor.Executor: Finished task 10.0 in stage 35.0 (TID 109). 2286 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:53 INFO python.PythonRunner: Times: total = 47, boot = 2, init = 45, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Finished task 8.0 in stage 35.0 (TID 107). 2263 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 110
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Running task 0.0 in stage 36.0 (TID 110)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 111
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Running task 1.0 in stage 36.0 (TID 111)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 112
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 113
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Running task 3.0 in stage 36.0 (TID 113)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Running task 2.0 in stage 36.0 (TID 112)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 32
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 114
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 115
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 116
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Running task 5.0 in stage 36.0 (TID 115)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 117
```

```
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Running task 7.0 in stage 36.0 (TID 117)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Running task 4.0 in stage 36.0 (TID 114)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Running task 6.0 in stage 36.0 (TID 116)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO memory.MemoryStore: Block broadcast_32_piece0 stored as bytes in memory (estimated size 9.3 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO broadcast.TorrentBroadcast: Reading broadcast variable 32 took 6 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO memory.MemoryStore: Block broadcast_32 stored as values in memory (estimated size 15.8 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO python.PythonRunner: Times: total = 47, boot = -538, init = 585, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO python.PythonRunner: Times: total = 46, boot = -512, init = 558, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO python.PythonRunner: Times: total = 45, boot = -488, init = 533, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO python.PythonRunner: Times: total = 47, boot = -523, init = 570, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Finished task 1.0 in stage 36.0 (TID 111). 2046 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Finished task 7.0 in stage 36.0 (TID 117). 2046 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Finished task 2.0 in stage 36.0 (TID 112). 2046 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Finished task 3.0 in stage 36.0 (TID 113). 2046 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 118
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO python.PythonRunner: Times: total = 50, boot = -490, init = 540, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Running task 8.0 in stage 36.0 (TID 118)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO python.PythonRunner: Times: total = 52, boot = -489, init = 541, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Finished task 5.0 in sta
```

```
ge 36.0 (TID 115). 2046 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 119
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Running task 9.0 in stage 36.0 (TID 119)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO python.PythonRunner: Times: total = 57, boot = -502, init = 559, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO python.PythonRunner: Times: total = 59, boot = -499, init = 558, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 120
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 121
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Running task 10.0 in stage 36.0 (TID 120)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 122
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Running task 11.0 in stage 36.0 (TID 121)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Running task 12.0 in stage 36.0 (TID 122)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Finished task 0.0 in stage 36.0 (TID 110). 2046 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 123
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Running task 13.0 in stage 36.0 (TID 123)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Finished task 4.0 in stage 36.0 (TID 114). 2046 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Finished task 6.0 in stage 36.0 (TID 116). 2046 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 124
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Running task 14.0 in stage 36.0 (TID 124)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 125
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_
```

```
01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Running task 15.0 in stage 36.0 (TID 125)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO python.PythonRunner: Times: total = 42, boot = 1, init = 41, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Finished task 8.0 in stage 36.0 (TID 118). 2089 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO python.PythonRunner: Times: total = 43, boot = 4, init = 39, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Finished task 10.0 in stage 36.0 (TID 120). 2046 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO python.PythonRunner: Times: total = 47, boot = -7, init = 54, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO python.PythonRunner: Times: total = 50, boot = -2, init = 52, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Finished task 11.0 in stage 36.0 (TID 121). 2046 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Finished task 9.0 in stage 36.0 (TID 119). 2046 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO python.PythonRunner: Times: total = 48, boot = -5, init = 53, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO python.PythonRunner: Times: total = 44, boot = -5, init = 49, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Finished task 12.0 in stage 36.0 (TID 122). 2046 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Finished task 13.0 in stage 36.0 (TID 123). 2046 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO python.PythonRunner: Times: total = 44, boot = -4, init = 48, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Finished task 15.0 in stage 36.0 (TID 125). 2046 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO python.PythonRunner: Times: total = 47, boot = -4, init = 51, finish = 0
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Finished task 14.0 in stage 36.0 (TID 124). 2046 bytes result sent to driver
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.CoarseGrainedExecutorBackend: Got assigned task 126
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO executor.Executor: Running task 0.0 in stage 38.0 (TID 126)
```

```
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO spark.MapOutputTrackerWorker: Updating epoch to 12 and clearing cache
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO broadcast.TorrentBroadcast: Started reading broadcast variable 33
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO memory.MemoryStore: Block broadcast_33_piece0 stored as bytes in memory (estimated size 60.5 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO broadcast.TorrentBroadcast: Reading broadcast variable 33 took 5 ms
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO memory.MemoryStore: Block broadcast_33 stored as values in memory (estimated size 167.8 KB, free 13.8 GB)
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO spark.MapOutputTrackerWorker: Don't have map outputs for shuffle 11, fetching them
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:21:54 INFO spark.MapOutputTrackerWorker: Doing the fetch; tracker endpoint = NettyRpcEndpointRef(spark://MapOutputTracker@10.0.222.169:37411)
24/04/28 21:21:55 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 38.0 (TID 126) in 973 ms on algo-1 (executor 1) (1/1)
24/04/28 21:21:55 INFO cluster.YarnScheduler: Removed TaskSet 38.0, whose tasks have all completed, from pool
24/04/28 21:21:55 INFO scheduler.DAGScheduler: ResultStage 38 (save at NativeMethodAccessorImpl.java:0) finished in 0.988 s
24/04/28 21:21:55 INFO scheduler.DAGScheduler: Job 25 finished: save at NativeMethodAccessorImpl.java:0, took 0.990049 s
24/04/28 21:21:56 INFO datasources.FileFormatWriter: Write Job ea6b1c08-1f04-4158-9fbf-0c719b0c71b1 committed.
24/04/28 21:21:56 INFO datasources.FileFormatWriter: Finished processing stats for write job ea6b1c08-1f04-4158-9fbf-0c719b0c71b1.
24/04/28 21:23:04 INFO monitor.ContainersMonitorImpl: Skipping monitoring container container_1714339251228_0001_01_000002 since CPU usage is not yet available.
```

WARNING:sagemaker:Job ended with status 'Stopped' rather than 'Completed'. This could mean the job timed out or stopped early for some other reason: Consider checking whether it completed as you expect.

```
[/var/log/yarn/userlogs/application_1714339251228_0001/container_1714339251228_0001_01_000002/stderr] 24/04/28 21:23:04 INFO monitor.ContainersMonitorImpl: Skipping monitoring container container_1714339251228_0001_01_000002 since CPU usage is not yet available.
```

***Please Wait Until the ^^ Processing Job ^^ Completes Above.***

## Inspect the Processed Output

These are the quality checks on our dataset.

*The next cells will not work properly until the job completes above.*

```
In [ ]: !aws s3 ls --recursive $s3_output_analyze_data/

2024-04-28 21:21:40          0 amazon-reviews-spark-analyzer-2024-04-28-21-15-58/output/constraint-checks/_SUCCESS
2024-04-28 21:21:39        773 amazon-reviews-spark-analyzer-2024-04-28-21-15-58/output/constraint-checks/part-00000-708fbe20-5bba-4123-bc3d-4612f33e12b7-c000.csv
2024-04-28 21:21:56          0 amazon-reviews-spark-analyzer-2024-04-28-21-15-58/output/constraint-suggestions/_SUCCESS
2024-04-28 21:21:56       8615 amazon-reviews-spark-analyzer-2024-04-28-21-15-58/output/constraint-suggestions/part-00000-2e954ba6-bad9-42ef-bc37-117f6c64df64-c000.csv
2024-04-28 21:21:31          0 amazon-reviews-spark-analyzer-2024-04-28-21-15-58/output/dataset-metrics/_SUCCESS
2024-04-28 21:21:31        364 amazon-reviews-spark-analyzer-2024-04-28-21-15-58/output/dataset-metrics/part-00000-9cbd597f-fa02-47ea-9fbb-786f966a1d8d-c000.csv
2024-04-28 21:21:42          0 amazon-reviews-spark-analyzer-2024-04-28-21-15-58/output/success-metrics/_SUCCESS
2024-04-28 21:21:41        277 amazon-reviews-spark-analyzer-2024-04-28-21-15-58/output/success-metrics/part-00000-73acd998-6c74-47c8-9e3b-7d91c5214dee-c000.csv
```

## Copy the Output from S3 to Local

- dataset-metrics/
- constraint-checks/
- success-metrics/
- constraint-suggestions/

```
In [ ]: !aws s3 cp --recursive $s3_output_analyze_data ./amazon-reviews-spark-analyzer/ --e

download: s3://sagemaker-us-east-1-211125778552/amazon-reviews-spark-analyzer-2024-04-28-21-15-58/output/constraint-checks/part-00000-708fbe20-5bba-4123-bc3d-4612f33e12b7-c000.csv to amazon-reviews-spark-analyzer/constraint-checks/part-00000-708fbe20-5bba-4123-bc3d-4612f33e12b7-c000.csv
download: s3://sagemaker-us-east-1-211125778552/amazon-reviews-spark-analyzer-2024-04-28-21-15-58/output/dataset-metrics/part-00000-9cbd597f-fa02-47ea-9fbb-786f966a1d8d-c000.csv to amazon-reviews-spark-analyzer/dataset-metrics/part-00000-9cbd597f-fa02-47ea-9fbb-786f966a1d8d-c000.csv
download: s3://sagemaker-us-east-1-211125778552/amazon-reviews-spark-analyzer-2024-04-28-21-15-58/output/success-metrics/part-00000-73acd998-6c74-47c8-9e3b-7d91c5214dee-c000.csv to amazon-reviews-spark-analyzer/success-metrics/part-00000-73acd998-6c74-47c8-9e3b-7d91c5214dee-c000.csv
download: s3://sagemaker-us-east-1-211125778552/amazon-reviews-spark-analyzer-2024-04-28-21-15-58/output/constraint-suggestions/part-00000-2e954ba6-bad9-42ef-bc37-117f6c64df64-c000.csv to amazon-reviews-spark-analyzer/constraint-suggestions/part-00000-2e954ba6-bad9-42ef-bc37-117f6c64df64-c000.csv
```

## Analyze Constraint Checks



```
In [ ]: import glob
import pandas as pd
import os

def load_dataset(path, sep, header):
    data = pd.concat(
        [pd.read_csv(f, sep=sep, header=header) for f in glob.glob("{}/*.csv".format(path))]
    )

    return data
```

```
In [ ]: df_constraint_checks = load_dataset(path="./amazon-reviews-spark-analyzer/constraint_checks.csv", sep=";", header=0)
df_constraint_checks[["check", "constraint", "constraint_status", "constraint_message"]]
```

Out[ ]:

	check	constraint	constraint_status	constraint_message
0	Review Check	SizeConstraint(Size(None))	Success	
1	Review Check	MinimumConstraint(Minimum(star_rating,None))	Success	
2	Review Check	MaximumConstraint(Maximum(star_rating,None))	Success	
3	Review Check	CompletenessConstraint(Completeness(review_id,...	Success	
4	Review Check	UniquenessConstraint(Uniqueness(List(review_id,...	Success	
5	Review Check	CompletenessConstraint(Completeness(marketplace...	Success	
6	Review Check	ComplianceConstraint(Compliance(marketplace co...	Success	

## Analyze Dataset Metrics

```
In [ ]: df_dataset_metrics = load_dataset(path="./amazon-reviews-spark-analyzer/dataset-metrics.csv", sep=";", header=0)
df_dataset_metrics
```

Out[ ]:

	entity	instance	name	value
0	Column	review_id	Completeness	1.000000
1	Column	review_id	ApproxCountDistinct	381704.000000
2	Mutlicolumn	total_votes,star_rating	Correlation	-0.086052
3	Dataset	*	Size	396601.000000
4	Column	star_rating	Mean	4.102493
5	Column	top star_rating	Compliance	0.765893
6	Mutlicolumn	total_votes,helpful_votes	Correlation	0.985751

Analyze Success Metrics

In [ ]:

```
df_success_metrics = load_dataset(path="./amazon-reviews-spark-analyzer/success-met
df_success_metrics
```

Out[ ]:

	entity	instance	name	value
0	Column	review_id	Completeness	1.0
1	Column	review_id	Uniqueness	1.0
2	Dataset	*	Size	396601.0
3	Column	star_rating	Maximum	5.0
4	Column	star_rating	Minimum	1.0
5	Column	marketplace contained in US,UK,DE,JP,FR	Compliance	1.0
6	Column	marketplace	Completeness	1.0

Analyze Constraint Suggestions

In [ ]:

```
# df_constraint_suggestions = load_dataset(path='./amazon-reviews-spark-analyzer/co
# pd.set_option('max_colwidth', 999)

# df_constraint_suggestions = df_constraint_suggestions[['_1', '_2', '_3']].dropna(co
# df_constraint_suggestions.columns=['column_name', 'description', 'code']
# df_constraint_suggestions
```

Release Resources

In [ ]:

```
%html
```

```
<p><b>Shutting down your kernel for this notebook to release resources.</b></p>  
<button class="sm-command-button" data-commandlinker-command="kernelmenu:shutdown"  
  
<script>  
  try {  
    els = document.getElementsByClassName("sm-command-button");  
    els[0].click();  
  }  
  catch(err) {  
    // NoOp  
  }  
</script>
```

**Shutting down your kernel for this notebook to release resources.**

```
In [ ]: %%javascript  
  
  try {  
    Jupyter.notebook.save_checkpoint();  
    Jupyter.notebook.session.delete();  
  }  
  catch(err) {  
    // NoOp  
  }
```

```
In [ ]: pip install nbconvert
```

Requirement already satisfied: nbconvert in /opt/conda/lib/python3.10/site-packages (7.16.3)

Requirement already satisfied: beautifulsoup4 in /opt/conda/lib/python3.10/site-packages (from nbconvert) (4.12.3)

Requirement already satisfied: bleach!=5.0.0 in /opt/conda/lib/python3.10/site-packages (from nbconvert) (6.1.0)

Requirement already satisfied: defusedxml in /opt/conda/lib/python3.10/site-packages (from nbconvert) (0.7.1)

Requirement already satisfied: Jinja2>=3.0 in /opt/conda/lib/python3.10/site-packages (from nbconvert) (3.1.3)

Requirement already satisfied: jupyter-core>=4.7 in /opt/conda/lib/python3.10/site-packages (from nbconvert) (5.7.2)

Requirement already satisfied: jupyterlab-pygments in /opt/conda/lib/python3.10/site-packages (from nbconvert) (0.3.0)

Requirement already satisfied: markupsafe>=2.0 in /opt/conda/lib/python3.10/site-packages (from nbconvert) (2.1.5)

Requirement already satisfied: mistune<4,>=2.0.3 in /opt/conda/lib/python3.10/site-packages (from nbconvert) (3.0.2)

Requirement already satisfied: nbclient>=0.5.0 in /opt/conda/lib/python3.10/site-packages (from nbconvert) (0.10.0)

Requirement already satisfied: nbformat>=5.7 in /opt/conda/lib/python3.10/site-packages (from nbconvert) (5.10.3)

Requirement already satisfied: packaging in /opt/conda/lib/python3.10/site-packages (from nbconvert) (23.2)

Requirement already satisfied: pandocfilters>=1.4.1 in /opt/conda/lib/python3.10/site-packages (from nbconvert) (1.5.0)

Requirement already satisfied: pygments>=2.4.1 in /opt/conda/lib/python3.10/site-packages (from nbconvert) (2.17.2)

Requirement already satisfied: tinycss2 in /opt/conda/lib/python3.10/site-packages (from nbconvert) (1.2.1)

Requirement already satisfied: traitlets>=5.1 in /opt/conda/lib/python3.10/site-packages (from nbconvert) (5.14.2)

Requirement already satisfied: six>=1.9.0 in /opt/conda/lib/python3.10/site-packages (from bleach!=5.0.0->nbconvert) (1.16.0)

Requirement already satisfied: webencodings in /opt/conda/lib/python3.10/site-packages (from bleach!=5.0.0->nbconvert) (0.5.1)

Requirement already satisfied: platformdirs>=2.5 in /opt/conda/lib/python3.10/site-packages (from jupyter-core>=4.7->nbconvert) (4.2.0)

Requirement already satisfied: jupyter-client>=6.1.12 in /opt/conda/lib/python3.10/site-packages (from nbclient>=0.5.0->nbconvert) (8.6.1)

Requirement already satisfied: fastjsonschema in /opt/conda/lib/python3.10/site-packages (from nbformat>=5.7->nbconvert) (2.19.1)

Requirement already satisfied: jsonschema>=2.6 in /opt/conda/lib/python3.10/site-packages (from nbformat>=5.7->nbconvert) (4.17.3)

Requirement already satisfied: soupsieve>1.2 in /opt/conda/lib/python3.10/site-packages (from beautifulsoup4->nbconvert) (2.5)

Requirement already satisfied: attrs>=17.4.0 in /opt/conda/lib/python3.10/site-packages (from jsonschema>=2.6->nbformat>=5.7->nbconvert) (23.2.0)

Requirement already satisfied: pyparsing!=0.17.0,!=0.17.1,!=0.17.2,>=0.14.0 in /opt/conda/lib/python3.10/site-packages (from jsonschema>=2.6->nbformat>=5.7->nbconvert) (0.20.0)

Requirement already satisfied: python-dateutil>=2.8.2 in /opt/conda/lib/python3.10/site-packages (from jupyter-client>=6.1.12->nbclient>=0.5.0->nbconvert) (2.9.0)

Requirement already satisfied: pyzmq>=23.0 in /opt/conda/lib/python3.10/site-packages (from jupyter-client>=6.1.12->nbclient>=0.5.0->nbconvert) (25.1.2)

Requirement already satisfied: tornado>=6.2 in /opt/conda/lib/python3.10/site-packages

```
es (from jupyter-client>=6.1.12->nbclient>=0.5.0->nbconvert) (6.4)
WARNING: Error parsing requirements for fsspec: [Errno 2] No such file or directory:
'/opt/conda/lib/python3.10/site-packages/fsspec-2023.6.0.dist-info/METADATA'
Note: you may need to restart the kernel to use updated packages.
```