

Assignment 01 | 190287R

Big Data Analytics using Hadoop and Spark on Amazon EMR

1. Task 1: Amazon EMR Setup

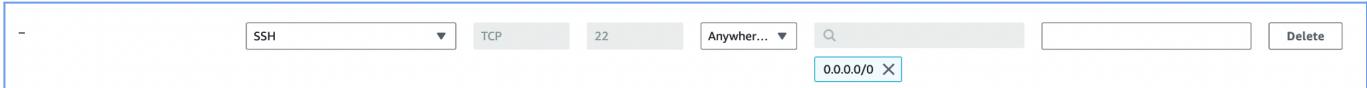
- Log into Amazon EMR and set up a new Spark cluster. Provide screenshots to show a successful cluster setup.

❖ Set up VPC:

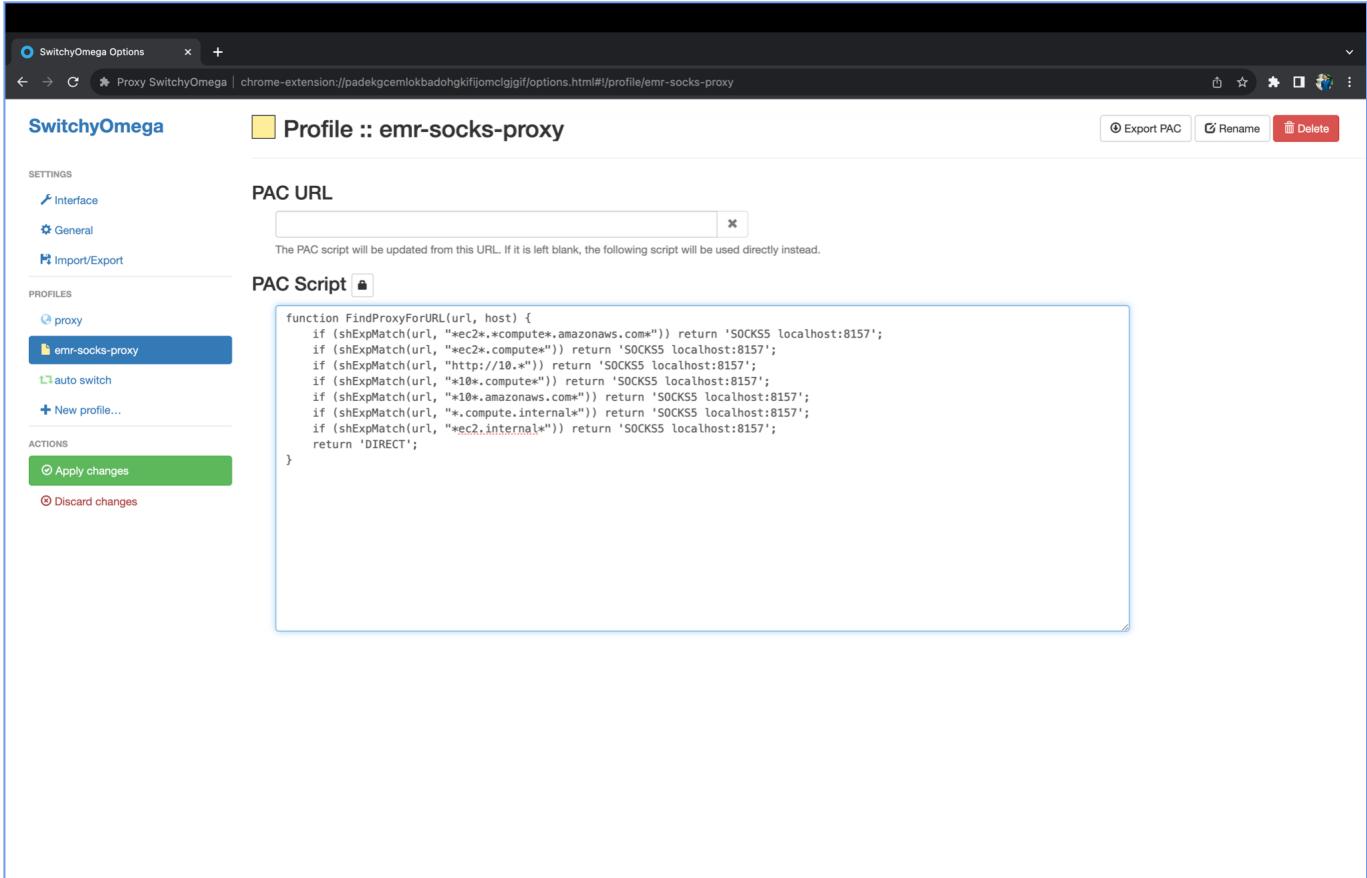
❖ Created key pair:

❖ Created EMR cluster with 1 primary and 2 core nodes (m4.large):

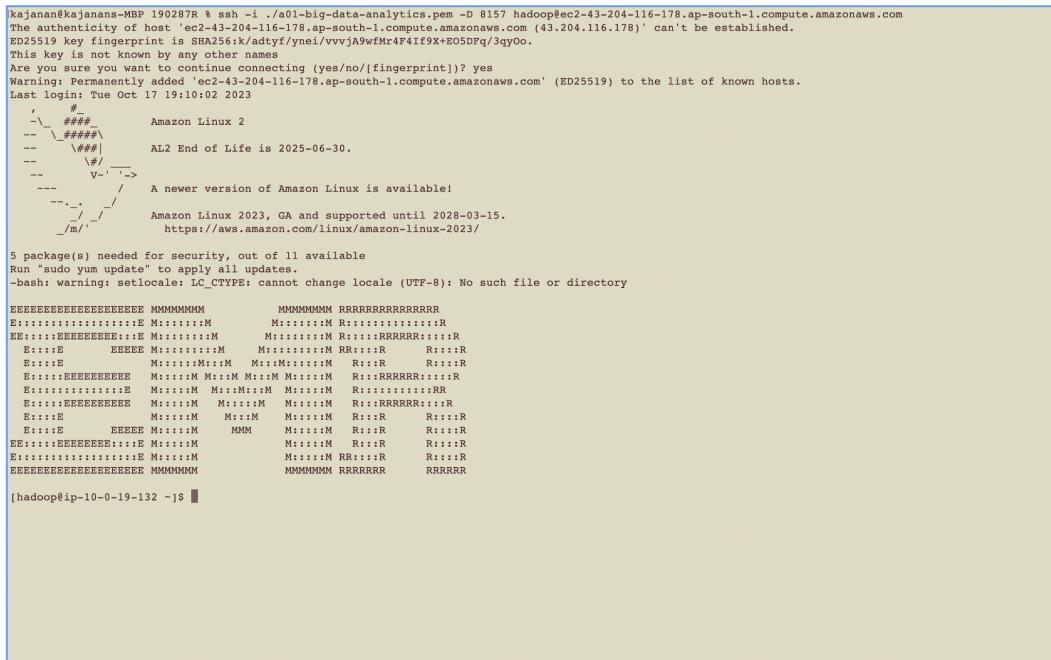
- ❖ Added inbound rule for security group of EMR-master:



❖ Set up proxy:



❖ SSH terminal:



❖ Resource manager UI:

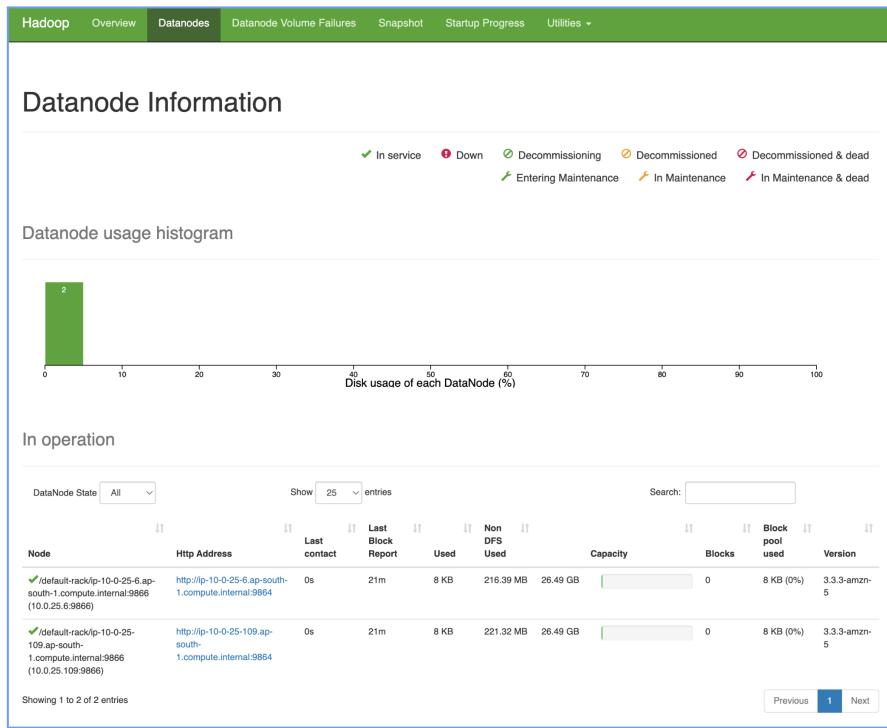
The screenshot shows the Hadoop Resource Manager UI with the title 'All Applications'. The left sidebar lists cluster metrics like 'Nodes' (0), 'Applications' (0), and a scheduler section. The main table has columns for 'ID', 'User', 'Name', 'Application Type', 'Application Tags', 'Queue', 'Application Priority', 'StartTime', 'LaunchTime', 'FinishTime', 'State', 'FinalStatus', 'Running Containers', 'Allocated CPU VCores', 'Allocated Memory MB', 'Allocated GPUs', 'Reserved CPU VCores', 'Reserved Memory MB', 'Reserved GPUs', '% of Queue', '% of Cluster', 'Progress', 'Tracking UI', and 'Blacklisted Nodes'. A search bar at the bottom right is empty.

❖ Spark history UI:

The screenshot shows the Apache Spark History Server UI with the title 'History Server' and version '3.4.1-amzn-0'. It displays an 'Event log directory' of 'hdfs:///var/log/spark/apps' and a note that it was last updated on '2023-10-18 00:46:33'. The client local time zone is listed as 'Asia/Colombo'. A bold message 'No completed applications found!' is displayed. Below it, a note asks if the correct logging directory is specified and whether the application ran to completion or stopped. A link to 'Show incomplete applications' is present.

❖ Node manager UI:

The screenshot shows the Hadoop Node Manager UI with the title 'Overview' for the IP 'ip-10-0-19-132.ap-south-1.compute.internal:8020'. It includes sections for 'Summary', 'NameNode Journal Status', 'NameNode Storage', and 'DFS Storage Types'. The 'Summary' section contains various metrics like 'Configured Capacity' (52.97 GB), 'DFS Used' (16 KB (0%)), and 'DataNodes usages%' (0.00% / 0.00% / 0.00%). The 'NameNode Journal Status' section shows a current transaction ID of 46 and a journal manager state of 'Active'. The 'NameNode Storage' section shows a storage directory of '/mnt/hamenode' with a type of 'IMAGE_AND_EDITS'. The 'DFS Storage Types' section shows a single entry for 'DISK' with a capacity of 52.87 GB and usage of 16 KB (0%).



2. Task 2: Importing and Managing Data in Amazon S3

- Import the provided data from the "voice_sample.csv" file and the "cell_centers.csv" file into your Amazon S3 bucket.

The screenshot shows the AWS S3 console for the bucket "a01-big-data-analytics". The bucket contains four objects: "data/", "logs/", "results/", and "scripts/". The "Actions" dropdown menu is open, showing options like Copy S3 URI, Copy URL, Download, Open, Delete, Create folder, and Upload. The "Upload" button is highlighted in orange. The table below shows the object details:

Name	Type	Last modified	Size	Storage class
data/	Folder	-	-	-
logs/	Folder	-	-	-
results/	Folder	-	-	-
scripts/	Folder	-	-	-

- List the files present in your Amazon S3 bucket. Provide screenshots of the command execution and the output.

- o aws s3 ls s3://a01-big-data-analytics/

```
[hadoop@ip-10-0-19-132 ~]$ aws s3 ls s3://a01-big-data-analytics/
                           PRE data/
                           PRE logs/
                           PRE results/
                           PRE scripts/
[hadoop@ip-10-0-19-132 ~]$ aws s3 ls s3://a01-big-data-analytics/data/
2023-10-16 16:12:06      58305 cell_centers.csv
2023-10-16 16:12:04  197669861 voice_sample.csv
[hadoop@ip-10-0-19-132 ~]$ aws s3 ls s3://a01-big-data-analytics/scripts/
2023-10-17 14:58:31      0
2023-10-17 18:57:12    1315 df_approach.py
2023-10-17 18:57:13    1628 sql_approach.py
[hadoop@ip-10-0-19-132 ~]$ aws s3 ls s3://a01-big-data-analytics/results/
2023-10-17 15:00:04      0
[hadoop@ip-10-0-19-132 ~]$ █
```

3. Task 3: Data Processing with Apache Spark via Amazon EMR

a. Spark DataFrame Approach

- i. Using an Amazon EMR notebook and Spark DataFrame API, extract the unique CALLER_IDs of users who have made at least one call every day. The calls must have been made from the Western Province. Include the Spark commands used and the output.

- **Initialization of Spark Session:** A Spark session was initiated to establish the Spark environment, providing the foundational framework for subsequent data processing steps.

 - `spark = SparkSession.builder.appName('a01-big-data-analytics-df-approach').getOrCreate()`

- **Data Import:** The two datasets, 'voice_sample.csv' and 'cell_centers.csv', were imported from Amazon S3 storage.

 - `spark.read.csv(s3_input_path, header=True, inferSchema=True)`

- **Date Column Transformation:** In the 'voice_sample' DataFrame, the 'CALL_DATE' column was derived from the 'CALL_TIME' column, facilitating date-based filtering.

 - `df1 = df1.withColumn('CALL_DATE', to_date('CALL_TIME', 'yyyyMMddHHmmss'))`

- **Location Filtering:** Subsequently, the 'cell_centers' DataFrame was filtered to include only records from the 'Western' province, employing the 'PROVINCE_NAME' column for the filter.

- **Data Join:** The filtered 'cell_centers' DataFrame was then joined with the 'voice_sample' DataFrame using the 'LOCATION_ID' column, enabling the association of calls with locations in the Western Province.

- **Data Grouping and Count:** With the data now integrated, the DataFrame was grouped by 'CALLER_ID', and the count of distinct 'CALL_DATE' values for each 'CALLER_ID' was calculated. This count represented the number of days calls were made by each user.

- **Filtering for Daily Callers:** Following this, the DataFrame was filtered to include only those rows where the 'NO_OF_DISTINCT_CALL_DATE' matched the total number of distinct

'CALL_DATE' values in the 'voice_sample' DataFrame. This ensured that only users who made calls every day were retained.

```
df3 = (
    df2
    .filter(df2['PROVINCE_NAME'] == 'Western')
    .join(df1, ['LOCATION_ID'])
    .groupBy('CALLER_ID').agg(countDistinct('CALL_DATE')).alias('NO_OF_DISTINCT_CALL_DATE'))
    .filter(col('NO_OF_DISTINCT_CALL_DATE') == df1.select('CALL_DATE').distinct().count())
    .select('CALLER_ID')
)
```

- **Result Storage:** The resulting DataFrame, containing unique CALLER_IDS, was saved to an Amazon S3 location for further analysis.

- df3.write.mode('overwrite').csv(s3_output_path, header=True)

- **Count Calculation and Storage:** Additionally, the count of unique CALLER_IDS was calculated and saved as a separate file in Amazon S3.

- count_df.write.mode('overwrite').text(s3_output_path)

- ❖ Added and ran the step in EMR cluster:

	Step ID	Status	Name	Log files	Creation time (UTC+05:30)	Start time (UTC+05:30)	Elapsed time
	s-091688826WQEZ1K03MEL	Completed	df-approach	controller syslog stderr stdout	18 October 2023, 01:02	18 October 2023, 01:02	3 minutes, 14...

❖ Results:

Amazon S3 > Buckets > a01-big-data-analytics > results/

results/ Copy S3 URI

Objects Properties

Objects
Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

C Copy S3 URI Copy URL Download Open Delete Actions Create folder Upload

Find objects by prefix

Name	Type	Last modified	Size	Storage class
df_approach_result_count/	Folder	-	-	-
df_approach_result_csv/	Folder	-	-	-

Amazon S3 > Buckets > a01-big-data-analytics > results/ > df_approach_result_count/

df_approach_result_count/ Copy S3 URI

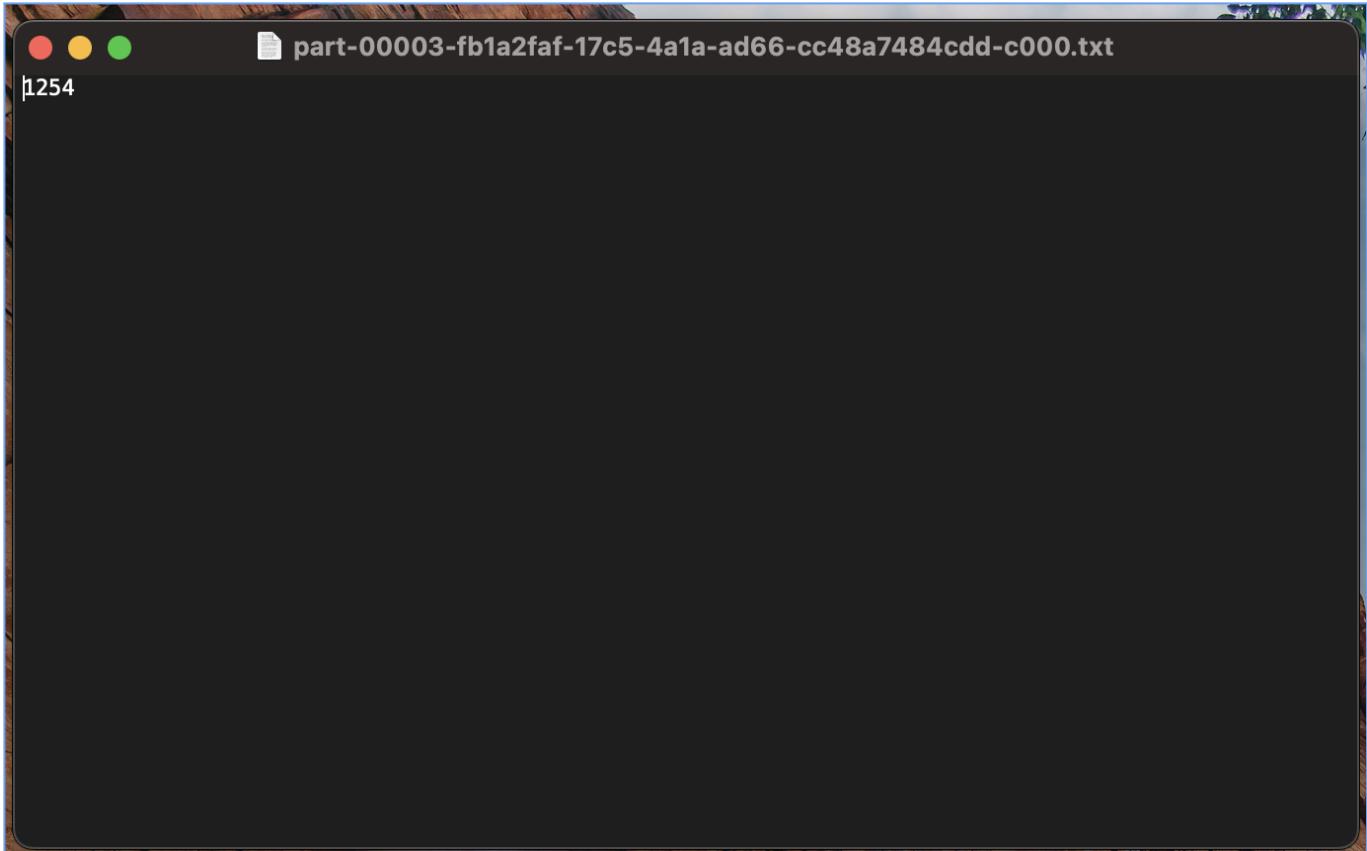
Objects Properties

Objects
Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

C Copy S3 URI Copy URL Download Open Delete Actions Create folder Upload

Find objects by prefix

Name	Type	Last modified	Size	Storage class
_SUCCESS	-	October 18, 2023, 01:06:05 (UTC+05:30)	0 B	Standard
part-00000-fb1a2faf-17c5-4a1a-ad66-cc48a7484cdd-c000.txt	txt	October 18, 2023, 01:06:05 (UTC+05:30)	0 B	Standard
part-00003-fb1a2faf-17c5-4a1a-ad66-cc48a7484cdd-c000.txt	txt	October 18, 2023, 01:06:05 (UTC+05:30)	5.0 B	Standard



Screenshot of the AWS S3 console showing the contents of the 'df_approach_result_csv' folder. The folder contains 26 objects, primarily CSV files, with one '_SUCCESS' file. The objects are listed by name, type (CSV), last modified date (October 18, 2023, at 01:05:38 UTC+05:30), size (e.g., 524.0 B, 474.0 B, 590.0 B, etc.), and storage class (Standard). The interface includes standard S3 actions like Copy, Delete, and Upload.

part-00000-c8d135cf-d037-43fc-ba8c-0f5bc9f04a9e-c000	
Caller ID	613942234 447293741 279783302 484898763 678252972 23690070 313081789 691036320 958666746 101668015 534165786 681290983 24495103 669700228 907364258 249137697 225900964 519552368 243102161 504155284 841809263 116177994 128690446 150390759 1007979574 389357873 805358769 458133565 710750981 5681719 203083573 838729027

b. Spark SQL Approach

- i. Repeat the same task as in 3.1.1 but use Spark SQL for data extraction. Include the Spark SQL commands used and the output.

- **Initialization of Spark Session:** Same as above.

 - ```
spark = SparkSession.builder.appName('a01-big-data-analytics-sql-approach').getOrCreate()
```

- **Data Import:** Same as above.

- **Temporary Views:** Both imported DataFrames were registered as temporary views with the names 'voice\_sample' and 'cell\_centers', respectively. This allows you to query them using Spark SQL.

- df1.createOrReplaceTempView("voice\_sample")
- df2.createOrReplaceTempView("cell\_centers")
- **Date Column Transformation:** A new column, 'CALL\_DATE', was derived in the 'voice\_sample' DataFrame using Spark SQL. The 'CALL\_TIME' column was converted into a date format ('yyyyMMddHHmmss' to date format) for later date-based filtering.
  - df1 = spark.sql('SELECT \*, TO\_DATE(CALL\_TIME, "yyyyMMddHHmmss") AS CALL\_DATE FROM voice\_sample')
- **Temporary View Update:** After the date transformation, the 'voice\_sample' DataFrame was updated and re-registered as a temporary view.
- **Data Extraction Using Spark SQL:** The core of the operation involved using Spark SQL to extract the unique CALLER\_IDs. The query selected 'CALLER\_ID' from the result of a subquery. The subquery involved a series of SQL operations, including:
  - Selecting 'LOCATION\_ID' from 'cell\_centers' where 'PROVINCE\_NAME' was 'Western', thereby narrowing down the locations in the Western Province.
  - Performing a LEFT JOIN between the filtered locations and the 'voice\_sample' DataFrame using the 'LOCATION\_ID' column to associate calls with locations.
  - Grouping the result by 'CALLER\_ID' and calculating the count of distinct 'CALL\_DATE' values for each 'CALLER\_ID'.
  - The outer query then filtered the results to only include rows where the 'NO\_OF\_DISTINCT\_CALL\_DATE' matched the total number of distinct 'CALL\_DATE' values in the 'voice\_sample' DataFrame. This ensured that only users who made calls every day were retained.

```
df3 = spark.sql(
 """
 SELECT CALLER_ID
 FROM (
 SELECT CALLER_ID, COUNT(DISTINCT CALL_DATE) AS NO_OF_DISTINCT_CALL_DATE
 FROM (
 SELECT LOCATION_ID
 FROM cell_centers
 WHERE PROVINCE_NAME = 'Western'
)
 LEFT JOIN voice_sample
 USING (LOCATION_ID)
 GROUP BY CALLER_ID
)
 WHERE NO_OF_DISTINCT_CALL_DATE = (SELECT COUNT(DISTINCT CALL_DATE) FROM voice_sample)
 """
)
```

- **Result Storage:** Similar to the DataFrame approach.
- **Count Calculation and Storage:** Similar to the DataFrame approach.

❖ Added and ran the step in EMR cluster:

**Add step Info**

**Step settings**

Type  
 Customised JAR  
Adds a step that enables you to write a customised script to process your data using the Java programming language.  
 Spark application  
Adds a step that submits work to the Spark framework on the cluster.  
 Shell script  
Troubleshoot your cluster.

Name

Deploy mode  
 Cluster mode  
Run your driver on a worker node.  
 Client mode  
Run your driver on the primary node as an external client.

Application location  
Path to a JAR with your application and dependencies (client deploy mode only supports a local path).

Spark-submit options - *optional*  
Specify other options for spark-submit.

Arguments - *optional* Info  
Specify optional arguments for your script.

**Step action**

Action if step fails  
The action to take when the step fails.  
 Continue  
Continues to the next step in the queue.  
 Cancel and wait  
Cancels any pending steps and returns the cluster to the waiting state.  
 Terminate cluster  
Shuts down the cluster.

|                          | Step ID                | Status    | Name         | Log files                             | Creation time (UTC+05:30) | Start time (UTC+05:30) | Elapsed time     |
|--------------------------|------------------------|-----------|--------------|---------------------------------------|---------------------------|------------------------|------------------|
| <input type="checkbox"/> | s-04188141GN04748QMXYO | Completed | sql-approach | controller   syslog   stderr   stdout | 18 October 2023, 01:34    | 18 October 2023, 01:34 | 3 minutes, 26... |

❖ Results:

Amazon S3 > Buckets > a01-big-data-analytics > results/

results/

**Objects (4)**  
Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

| <input type="checkbox"/> | Name                       | Type   | Last modified | Size | Storage class |
|--------------------------|----------------------------|--------|---------------|------|---------------|
| <input type="checkbox"/> | df_approach_result_count/  | Folder | -             | -    | -             |
| <input type="checkbox"/> | df_approach_result_csv/    | Folder | -             | -    | -             |
| <input type="checkbox"/> | sql_approach_result_count/ | Folder | -             | -    | -             |
| <input type="checkbox"/> | sql_approach_result_csv/   | Folder | -             | -    | -             |

Amazon S3 > Buckets > a01-big-data-analytics > results/ > sql\_approach\_result\_count/

sql\_approach\_result\_count/

**Objects**  
Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

| <input type="checkbox"/> | Name                                                      | Type | Last modified                          | Size  | Storage class |
|--------------------------|-----------------------------------------------------------|------|----------------------------------------|-------|---------------|
| <input type="checkbox"/> | _SUCCESS                                                  | -    | October 18, 2023, 01:38:23 (UTC+05:30) | 0 B   | Standard      |
| <input type="checkbox"/> | part-00000-f0d7f4dc-094b-4cce-840f-742816d7d9953-c000.txt | txt  | October 18, 2023, 01:38:23 (UTC+05:30) | 0 B   | Standard      |
| <input type="checkbox"/> | part-00003-f0d7f4dc-094b-4cce-840f-742816d7d9953-c000.txt | txt  | October 18, 2023, 01:38:23 (UTC+05:30) | 5.0 B | Standard      |



AWS Services Search [Option+S]

Amazon S3 > Buckets > a01-big-data-analytics > results/ > sql\_approach\_result\_csv/

sql\_approach\_result\_csv/

Copy S3 URI

Objects (26)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

| Name                                                    | Type | Last modified                          | Size    | Storage class |
|---------------------------------------------------------|------|----------------------------------------|---------|---------------|
| _SUCCESS                                                | -    | October 18, 2023, 01:37:29 (UTC+05:30) | 0 B     | Standard      |
| part-00001-d95fb4f-66f8-4c7c-932d-8596c75ca527-c000.csv | csv  | October 18, 2023, 01:37:27 (UTC+05:30) | 524.0 B | Standard      |
| part-00001-d95fb4f-66f8-4c7c-932d-8596c75ca527-c000.csv | csv  | October 18, 2023, 01:37:27 (UTC+05:30) | 474.0 B | Standard      |
| part-00002-d95fb4f-66f8-4c7c-932d-8596c75ca527-c000.csv | csv  | October 18, 2023, 01:37:27 (UTC+05:30) | 590.0 B | Standard      |
| part-00003-d95fb4f-66f8-4c7c-932d-8596c75ca527-c000.csv | csv  | October 18, 2023, 01:37:27 (UTC+05:30) | 461.0 B | Standard      |
| part-00004-d95fb4f-66f8-4c7c-932d-8596c75ca527-c000.csv | csv  | October 18, 2023, 01:37:27 (UTC+05:30) | 426.0 B | Standard      |
| part-00005-d95fb4f-66f8-4c7c-932d-8596c75ca527-c000.csv | csv  | October 18, 2023, 01:37:27 (UTC+05:30) | 425.0 B | Standard      |

CloudShell Feedback © 2023, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

part-00000-d95fb4f-66f8-4c7c-932d-8596c75ca527-c000

Sheet 1

View Zoom Add Category Pivot Table Insert Table Chart Text Shape Media Comment Collaborate Format Organize

Sheet Name Sheet 1

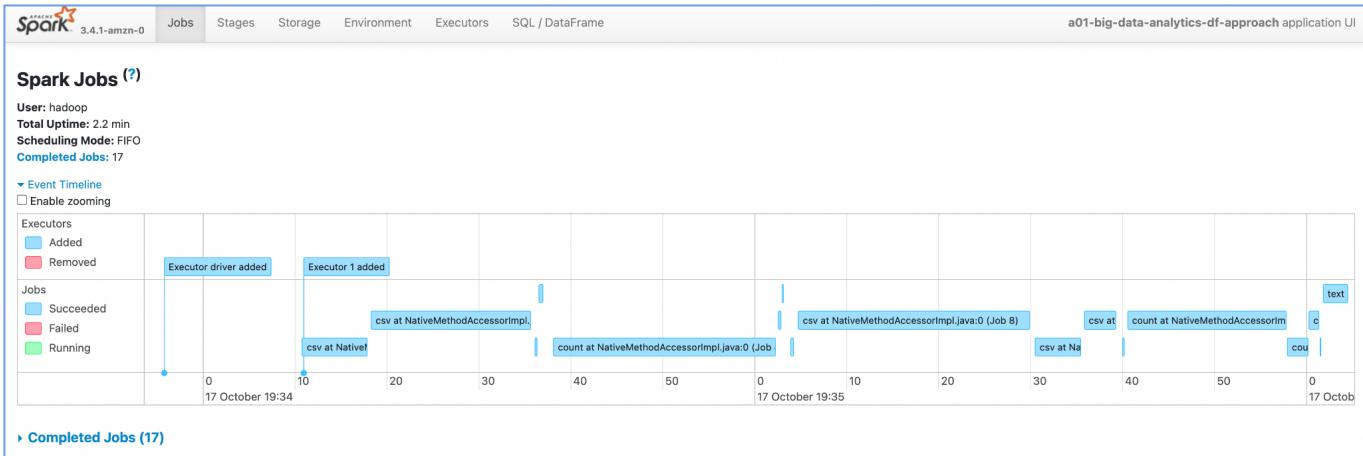
Background Duplicate Sheet Delete Sheet

| CALLER_ID  |
|------------|
| 613942234  |
| 447293741  |
| 276793302  |
| 484898763  |
| 678252972  |
| 23690070   |
| 313081789  |
| 691036320  |
| 958666746  |
| 101668015  |
| 534165786  |
| 681290963  |
| 24495103   |
| 66970228   |
| 907364258  |
| 249137697  |
| 22590094   |
| 519532368  |
| 243102161  |
| 504155264  |
| 128690446  |
| 116177994  |
| 841809263  |
| 150390759  |
| 1007979574 |
| 389357873  |
| 805358769  |
| 5651719    |
| 203083573  |
| 838729027  |
| 419490013  |
| 802830719  |

## c. Analysis and Comparison

- Provide screenshots of your Amazon EMR UI showing the execution of both jobs (DataFrame API and Spark SQL). The screenshots should show various stages of the job, tasks, and resource usage.

### ❖ DataFrame API:



**SQL / DataFrame**

Completed Queries: 6

Completed Queries (6)

| ID | Description                              | Submitted           | Duration | Job IDs              |
|----|------------------------------------------|---------------------|----------|----------------------|
| 5  | text at NativeMethodAccessorImpl.java:0  | 2023/10/17 19:36:01 | 3 s      | [16]                 |
| 4  | count at NativeMethodAccessorImpl.java:0 | 2023/10/17 19:35:39 | 22 s     | [11][12][13][14][15] |
| 3  | csv at NativeMethodAccessorImpl.java:0   | 2023/10/17 19:35:03 | 36 s     | [7][8][9][10]        |
| 2  | count at NativeMethodAccessorImpl.java:0 | 2023/10/17 19:34:37 | 25 s     | [4][5][6]            |
| 1  | csv at NativeMethodAccessorImpl.java:0   | 2023/10/17 19:34:35 | 0.3 s    | [2]                  |
| 0  | csv at NativeMethodAccessorImpl.java:0   | 2023/10/17 19:34:09 | 9 s      | [0]                  |

**SQL / DataFrame**

Completed Queries: 6

Completed Queries (6)

| ID | Description                              | Submitted           | Duration | Job IDs              |
|----|------------------------------------------|---------------------|----------|----------------------|
| 5  | text at NativeMethodAccessorImpl.java:0  | 2023/10/17 19:36:01 | 3 s      | [16]                 |
| 4  | count at NativeMethodAccessorImpl.java:0 | 2023/10/17 19:35:39 | 22 s     | [11][12][13][14][15] |
| 3  | csv at NativeMethodAccessorImpl.java:0   | 2023/10/17 19:35:03 | 36 s     | [7][8][9][10]        |
| 2  | count at NativeMethodAccessorImpl.java:0 | 2023/10/17 19:34:37 | 25 s     | [4][5][6]            |
| 1  | csv at NativeMethodAccessorImpl.java:0   | 2023/10/17 19:34:35 | 0.3 s    | [2]                  |
| 0  | csv at NativeMethodAccessorImpl.java:0   | 2023/10/17 19:34:09 | 9 s      | [0]                  |

| Jobs                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                                                  | Stages                 |                         | Storage         |                        | Environment             |                      | Executors             |                                 | SQL / DataFrame               |                            |              |               |                |                        |                         |            |                |                        |                         |                      |                       |                                 |                               |                            |               |          |           |                     |                 |                 |                |             |                     |       |              |               |      |                  |                     |              |              |                |             |                     |               |              |               |      |                  |     |      |                 |         |       |         |       |          |        |                 |                                                |               |       |               |               |               |               |               |                |               |          |   |   |   |   |   |   |   |                  |       |       |       |                  |   |                                                  |        |   |                 |                 |               |                     |                   |                 |                 |       |   |   |   |   |     |     |                |           |          |   |   |                  |  |  |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------|------------------------|-------------------------|-----------------|------------------------|-------------------------|----------------------|-----------------------|---------------------------------|-------------------------------|----------------------------|--------------|---------------|----------------|------------------------|-------------------------|------------|----------------|------------------------|-------------------------|----------------------|-----------------------|---------------------------------|-------------------------------|----------------------------|---------------|----------|-----------|---------------------|-----------------|-----------------|----------------|-------------|---------------------|-------|--------------|---------------|------|------------------|---------------------|--------------|--------------|----------------|-------------|---------------------|---------------|--------------|---------------|------|------------------|-----|------|-----------------|---------|-------|---------|-------|----------|--------|-----------------|------------------------------------------------|---------------|-------|---------------|---------------|---------------|---------------|---------------|----------------|---------------|----------|---|---|---|---|---|---|---|------------------|-------|-------|-------|------------------|---|--------------------------------------------------|--------|---|-----------------|-----------------|---------------|---------------------|-------------------|-----------------|-----------------|-------|---|---|---|---|-----|-----|----------------|-----------|----------|---|---|------------------|--|--|
| a01-big-data-analytics-df-approach application U                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                  |                        |                         |                 |                        |                         |                      |                       |                                 |                               |                            |              |               |                |                        |                         |            |                |                        |                         |                      |                       |                                 |                               |                            |               |          |           |                     |                 |                 |                |             |                     |       |              |               |      |                  |                     |              |              |                |             |                     |               |              |               |      |                  |     |      |                 |         |       |         |       |          |        |                 |                                                |               |       |               |               |               |               |               |                |               |          |   |   |   |   |   |   |   |                  |       |       |       |                  |   |                                                  |        |   |                 |                 |               |                     |                   |                 |                 |       |   |   |   |   |     |     |                |           |          |   |   |                  |  |  |
| Executors                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                                  |                        |                         |                 |                        |                         |                      |                       |                                 |                               |                            |              |               |                |                        |                         |            |                |                        |                         |                      |                       |                                 |                               |                            |               |          |           |                     |                 |                 |                |             |                     |       |              |               |      |                  |                     |              |              |                |             |                     |               |              |               |      |                  |     |      |                 |         |       |         |       |          |        |                 |                                                |               |       |               |               |               |               |               |                |               |          |   |   |   |   |   |   |   |                  |       |       |       |                  |   |                                                  |        |   |                 |                 |               |                     |                   |                 |                 |       |   |   |   |   |     |     |                |           |          |   |   |                  |  |  |
| Show Additional Metrics                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                  |                        |                         |                 |                        |                         |                      |                       |                                 |                               |                            |              |               |                |                        |                         |            |                |                        |                         |                      |                       |                                 |                               |                            |               |          |           |                     |                 |                 |                |             |                     |       |              |               |      |                  |                     |              |              |                |             |                     |               |              |               |      |                  |     |      |                 |         |       |         |       |          |        |                 |                                                |               |       |               |               |               |               |               |                |               |          |   |   |   |   |   |   |   |                  |       |       |       |                  |   |                                                  |        |   |                 |                 |               |                     |                   |                 |                 |       |   |   |   |   |     |     |                |           |          |   |   |                  |  |  |
| <strong>Summary</strong>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                  |                        |                         |                 |                        |                         |                      |                       |                                 |                               |                            |              |               |                |                        |                         |            |                |                        |                         |                      |                       |                                 |                               |                            |               |          |           |                     |                 |                 |                |             |                     |       |              |               |      |                  |                     |              |              |                |             |                     |               |              |               |      |                  |     |      |                 |         |       |         |       |          |        |                 |                                                |               |       |               |               |               |               |               |                |               |          |   |   |   |   |   |   |   |                  |       |       |       |                  |   |                                                  |        |   |                 |                 |               |                     |                   |                 |                 |       |   |   |   |   |     |     |                |           |          |   |   |                  |  |  |
| <table border="1"> <thead> <tr> <th>RDD Blocks</th><th>Storage Memory</th><th>On Heap Storage Memory</th><th>Off Heap Storage Memory</th><th>Disk Used</th><th>Cores</th><th>Active Tasks</th><th>Failed Tasks</th><th>Complete Tasks</th><th>Total Tasks</th><th>Task Time (GC Time)</th><th>Input</th><th>Shuffle Read</th><th>Shuffle Write</th><th>Excluded</th></tr> </thead> <tbody> <tr> <td>Active(2)</td><td>0</td><td>0.0 B / 3.1 GiB</td><td>0.0 B / 3.1 GiB</td><td>0.0 B / 0.0 B</td><td>0.0 B</td><td>4</td><td>0</td><td>0</td><td>128</td><td>128</td><td>8.7 min (11 s)</td><td>754.3 MiB</td><td>53.6 MiB</td><td>0</td></tr> <tr> <td>Dead(0)</td><td>0</td><td>0.0 B / 0.0 B</td><td>0.0 B / 0.0 B</td><td>0.0 B</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0.0 ms (0.0 ms)</td><td>0.0 B</td><td>0.0 B</td><td>0.0 B</td><td>0</td></tr> <tr> <td>Total(2)</td><td>0</td><td>0.0 B / 3.1 GiB</td><td>0.0 B / 3.1 GiB</td><td>0.0 B / 0.0 B</td><td>0.0 B</td><td>4</td><td>0</td><td>0</td><td>128</td><td>128</td><td>8.7 min (11 s)</td><td>754.3 MiB</td><td>53.6 MiB</td><td>0</td></tr> </tbody> </table>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                  |                        |                         |                 |                        |                         |                      |                       |                                 |                               |                            |              | RDD Blocks    | Storage Memory | On Heap Storage Memory | Off Heap Storage Memory | Disk Used  | Cores          | Active Tasks           | Failed Tasks            | Complete Tasks       | Total Tasks           | Task Time (GC Time)             | Input                         | Shuffle Read               | Shuffle Write | Excluded | Active(2) | 0                   | 0.0 B / 3.1 GiB | 0.0 B / 3.1 GiB | 0.0 B / 0.0 B  | 0.0 B       | 4                   | 0     | 0            | 128           | 128  | 8.7 min (11 s)   | 754.3 MiB           | 53.6 MiB     | 0            | Dead(0)        | 0           | 0.0 B / 0.0 B       | 0.0 B / 0.0 B | 0.0 B        | 0             | 0    | 0                | 0   | 0    | 0.0 ms (0.0 ms) | 0.0 B   | 0.0 B | 0.0 B   | 0     | Total(2) | 0      | 0.0 B / 3.1 GiB | 0.0 B / 3.1 GiB                                | 0.0 B / 0.0 B | 0.0 B | 4             | 0             | 0             | 128           | 128           | 8.7 min (11 s) | 754.3 MiB     | 53.6 MiB | 0 |   |   |   |   |   |   |                  |       |       |       |                  |   |                                                  |        |   |                 |                 |               |                     |                   |                 |                 |       |   |   |   |   |     |     |                |           |          |   |   |                  |  |  |
| RDD Blocks                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | Storage Memory                                   | On Heap Storage Memory | Off Heap Storage Memory | Disk Used       | Cores                  | Active Tasks            | Failed Tasks         | Complete Tasks        | Total Tasks                     | Task Time (GC Time)           | Input                      | Shuffle Read | Shuffle Write | Excluded       |                        |                         |            |                |                        |                         |                      |                       |                                 |                               |                            |               |          |           |                     |                 |                 |                |             |                     |       |              |               |      |                  |                     |              |              |                |             |                     |               |              |               |      |                  |     |      |                 |         |       |         |       |          |        |                 |                                                |               |       |               |               |               |               |               |                |               |          |   |   |   |   |   |   |   |                  |       |       |       |                  |   |                                                  |        |   |                 |                 |               |                     |                   |                 |                 |       |   |   |   |   |     |     |                |           |          |   |   |                  |  |  |
| Active(2)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | 0                                                | 0.0 B / 3.1 GiB        | 0.0 B / 3.1 GiB         | 0.0 B / 0.0 B   | 0.0 B                  | 4                       | 0                    | 0                     | 128                             | 128                           | 8.7 min (11 s)             | 754.3 MiB    | 53.6 MiB      | 0              |                        |                         |            |                |                        |                         |                      |                       |                                 |                               |                            |               |          |           |                     |                 |                 |                |             |                     |       |              |               |      |                  |                     |              |              |                |             |                     |               |              |               |      |                  |     |      |                 |         |       |         |       |          |        |                 |                                                |               |       |               |               |               |               |               |                |               |          |   |   |   |   |   |   |   |                  |       |       |       |                  |   |                                                  |        |   |                 |                 |               |                     |                   |                 |                 |       |   |   |   |   |     |     |                |           |          |   |   |                  |  |  |
| Dead(0)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | 0                                                | 0.0 B / 0.0 B          | 0.0 B / 0.0 B           | 0.0 B           | 0                      | 0                       | 0                    | 0                     | 0                               | 0.0 ms (0.0 ms)               | 0.0 B                      | 0.0 B        | 0.0 B         | 0              |                        |                         |            |                |                        |                         |                      |                       |                                 |                               |                            |               |          |           |                     |                 |                 |                |             |                     |       |              |               |      |                  |                     |              |              |                |             |                     |               |              |               |      |                  |     |      |                 |         |       |         |       |          |        |                 |                                                |               |       |               |               |               |               |               |                |               |          |   |   |   |   |   |   |   |                  |       |       |       |                  |   |                                                  |        |   |                 |                 |               |                     |                   |                 |                 |       |   |   |   |   |     |     |                |           |          |   |   |                  |  |  |
| Total(2)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 0                                                | 0.0 B / 3.1 GiB        | 0.0 B / 3.1 GiB         | 0.0 B / 0.0 B   | 0.0 B                  | 4                       | 0                    | 0                     | 128                             | 128                           | 8.7 min (11 s)             | 754.3 MiB    | 53.6 MiB      | 0              |                        |                         |            |                |                        |                         |                      |                       |                                 |                               |                            |               |          |           |                     |                 |                 |                |             |                     |       |              |               |      |                  |                     |              |              |                |             |                     |               |              |               |      |                  |     |      |                 |         |       |         |       |          |        |                 |                                                |               |       |               |               |               |               |               |                |               |          |   |   |   |   |   |   |   |                  |       |       |       |                  |   |                                                  |        |   |                 |                 |               |                     |                   |                 |                 |       |   |   |   |   |     |     |                |           |          |   |   |                  |  |  |
| <strong>Executors</strong>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                  |                        |                         |                 |                        |                         |                      |                       |                                 |                               |                            |              |               |                |                        |                         |            |                |                        |                         |                      |                       |                                 |                               |                            |               |          |           |                     |                 |                 |                |             |                     |       |              |               |      |                  |                     |              |              |                |             |                     |               |              |               |      |                  |     |      |                 |         |       |         |       |          |        |                 |                                                |               |       |               |               |               |               |               |                |               |          |   |   |   |   |   |   |   |                  |       |       |       |                  |   |                                                  |        |   |                 |                 |               |                     |                   |                 |                 |       |   |   |   |   |     |     |                |           |          |   |   |                  |  |  |
| Show 20 entries                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                  |                        |                         |                 |                        |                         |                      |                       |                                 |                               |                            |              |               |                |                        |                         |            |                |                        |                         |                      |                       |                                 |                               |                            |               |          |           |                     |                 |                 |                |             |                     |       |              |               |      |                  |                     |              |              |                |             |                     |               |              |               |      |                  |     |      |                 |         |       |         |       |          |        |                 |                                                |               |       |               |               |               |               |               |                |               |          |   |   |   |   |   |   |   |                  |       |       |       |                  |   |                                                  |        |   |                 |                 |               |                     |                   |                 |                 |       |   |   |   |   |     |     |                |           |          |   |   |                  |  |  |
| <table border="1"> <thead> <tr> <th rowspan="2">Executor ID</th><th rowspan="2">Address</th><th rowspan="2">Status</th><th rowspan="2">RDD Blocks</th><th rowspan="2">Storage Memory</th><th rowspan="2">On Heap Storage Memory</th><th rowspan="2">Off Heap Storage Memory</th><th rowspan="2">JVM Memory / OffHeap</th><th rowspan="2">Peak Memory / Offheap</th><th rowspan="2">Peak Execution Memory / OffHeap</th><th rowspan="2">Peak Storage Memory / OffHeap</th><th rowspan="2">Peak Pool Memory / OffHeap</th><th rowspan="2">Disk Used</th><th rowspan="2">Cores</th><th rowspan="2">Resources</th><th rowspan="2">Resource Profile Id</th><th rowspan="2">Active Tasks</th><th rowspan="2">Failed Tasks</th><th rowspan="2">Complete Tasks</th><th rowspan="2">Total Tasks</th><th rowspan="2">Task Time (GC Time)</th><th rowspan="2">Input</th><th rowspan="2">Shuffle Read</th><th rowspan="2">Shuffle Write</th><th rowspan="2">Logs</th><th>Exec Loss Reason</th></tr> <tr> <th>Peak</th><th>JVM</th><th>Peak</th><th>Peak</th><th>Peak</th><th>Peak</th><th>Disk</th><th>Used</th><th>Profile</th><th>Id</th><th>Time</th><th>(GC</th><th>Time</th><th>Input</th><th>Shuffle</th><th>Read</th><th>Shuffle</th><th>Write</th><th>Logs</th><th>Reason</th></tr> </thead> <tbody> <tr> <td>driver</td><td>ip-10-0-25-6.ap-south-1.compute.internal:33533</td><td>Active</td><td>0</td><td>0.0 B / 1 GiB</td><td>0.0 B / 1 GiB</td><td>0.0 B / 0.0 B</td><td>0.0 B</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>2.2 min (0.0 ms)</td><td>0.0 B</td><td>0.0 B</td><td>0.0 B</td><td>stdout<br/>stderr</td></tr> <tr> <td>1</td><td>ip-10-0-25-109.ap-south-1.compute.internal:44985</td><td>Active</td><td>0</td><td>0.0 B / 2.1 GiB</td><td>0.0 B / 2.1 GiB</td><td>0.0 B / 0.0 B</td><td>1.5 GiB / 132.4 MiB</td><td>261.9 MiB / 0.0 B</td><td>5.4 MiB / 0.0 B</td><td>19.3 kB / 0.0 B</td><td>0.0 B</td><td>4</td><td>0</td><td>0</td><td>0</td><td>128</td><td>128</td><td>6.5 min (11 s)</td><td>754.3 MiB</td><td>53.6 MiB</td><td>0</td><td>0</td><td>stdout<br/>stderr</td></tr> </tbody> </table> |                                                  |                        |                         |                 |                        |                         |                      |                       |                                 |                               |                            |              |               | Executor ID    | Address                | Status                  | RDD Blocks | Storage Memory | On Heap Storage Memory | Off Heap Storage Memory | JVM Memory / OffHeap | Peak Memory / Offheap | Peak Execution Memory / OffHeap | Peak Storage Memory / OffHeap | Peak Pool Memory / OffHeap | Disk Used     | Cores    | Resources | Resource Profile Id | Active Tasks    | Failed Tasks    | Complete Tasks | Total Tasks | Task Time (GC Time) | Input | Shuffle Read | Shuffle Write | Logs | Exec Loss Reason | Peak                | JVM          | Peak         | Peak           | Peak        | Peak                | Disk          | Used         | Profile       | Id   | Time             | (GC | Time | Input           | Shuffle | Read  | Shuffle | Write | Logs     | Reason | driver          | ip-10-0-25-6.ap-south-1.compute.internal:33533 | Active        | 0     | 0.0 B / 1 GiB | 0.0 B / 1 GiB | 0.0 B / 0.0 B  | 0.0 B / 0.0 B | 0.0 B    | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2.2 min (0.0 ms) | 0.0 B | 0.0 B | 0.0 B | stdout<br>stderr | 1 | ip-10-0-25-109.ap-south-1.compute.internal:44985 | Active | 0 | 0.0 B / 2.1 GiB | 0.0 B / 2.1 GiB | 0.0 B / 0.0 B | 1.5 GiB / 132.4 MiB | 261.9 MiB / 0.0 B | 5.4 MiB / 0.0 B | 19.3 kB / 0.0 B | 0.0 B | 4 | 0 | 0 | 0 | 128 | 128 | 6.5 min (11 s) | 754.3 MiB | 53.6 MiB | 0 | 0 | stdout<br>stderr |  |  |
| Executor ID                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | Address                                          | Status                 | RDD Blocks              | Storage Memory  | On Heap Storage Memory | Off Heap Storage Memory | JVM Memory / OffHeap | Peak Memory / Offheap | Peak Execution Memory / OffHeap | Peak Storage Memory / OffHeap | Peak Pool Memory / OffHeap | Disk Used    | Cores         |                |                        |                         |            |                |                        |                         |                      |                       |                                 |                               |                            |               |          |           |                     |                 |                 |                |             |                     |       |              |               |      | Resources        | Resource Profile Id | Active Tasks | Failed Tasks | Complete Tasks | Total Tasks | Task Time (GC Time) | Input         | Shuffle Read | Shuffle Write | Logs | Exec Loss Reason |     |      |                 |         |       |         |       |          |        |                 |                                                |               |       |               |               |               |               |               |                |               |          |   |   |   |   |   |   |   |                  |       |       |       |                  |   |                                                  |        |   |                 |                 |               |                     |                   |                 |                 |       |   |   |   |   |     |     |                |           |          |   |   |                  |  |  |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                  |                        |                         |                 |                        |                         |                      |                       |                                 |                               |                            |              |               | Peak           | JVM                    | Peak                    | Peak       | Peak           | Peak                   | Disk                    | Used                 | Profile               | Id                              | Time                          | (GC                        | Time          | Input    | Shuffle   | Read                | Shuffle         | Write           | Logs           | Reason      |                     |       |              |               |      |                  |                     |              |              |                |             |                     |               |              |               |      |                  |     |      |                 |         |       |         |       |          |        |                 |                                                |               |       |               |               |               |               |               |                |               |          |   |   |   |   |   |   |   |                  |       |       |       |                  |   |                                                  |        |   |                 |                 |               |                     |                   |                 |                 |       |   |   |   |   |     |     |                |           |          |   |   |                  |  |  |
| driver                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | ip-10-0-25-6.ap-south-1.compute.internal:33533   | Active                 | 0                       | 0.0 B / 1 GiB   | 0.0 B / 1 GiB          | 0.0 B / 0.0 B           | 0.0 B / 0.0 B        | 0.0 B / 0.0 B         | 0.0 B / 0.0 B                   | 0.0 B / 0.0 B                 | 0.0 B                      | 0            | 0             | 0              | 0                      | 0                       | 0          | 0              | 2.2 min (0.0 ms)       | 0.0 B                   | 0.0 B                | 0.0 B                 | stdout<br>stderr                |                               |                            |               |          |           |                     |                 |                 |                |             |                     |       |              |               |      |                  |                     |              |              |                |             |                     |               |              |               |      |                  |     |      |                 |         |       |         |       |          |        |                 |                                                |               |       |               |               |               |               |               |                |               |          |   |   |   |   |   |   |   |                  |       |       |       |                  |   |                                                  |        |   |                 |                 |               |                     |                   |                 |                 |       |   |   |   |   |     |     |                |           |          |   |   |                  |  |  |
| 1                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | ip-10-0-25-109.ap-south-1.compute.internal:44985 | Active                 | 0                       | 0.0 B / 2.1 GiB | 0.0 B / 2.1 GiB        | 0.0 B / 0.0 B           | 1.5 GiB / 132.4 MiB  | 261.9 MiB / 0.0 B     | 5.4 MiB / 0.0 B                 | 19.3 kB / 0.0 B               | 0.0 B                      | 4            | 0             | 0              | 0                      | 128                     | 128        | 6.5 min (11 s) | 754.3 MiB              | 53.6 MiB                | 0                    | 0                     | stdout<br>stderr                |                               |                            |               |          |           |                     |                 |                 |                |             |                     |       |              |               |      |                  |                     |              |              |                |             |                     |               |              |               |      |                  |     |      |                 |         |       |         |       |          |        |                 |                                                |               |       |               |               |               |               |               |                |               |          |   |   |   |   |   |   |   |                  |       |       |       |                  |   |                                                  |        |   |                 |                 |               |                     |                   |                 |                 |       |   |   |   |   |     |     |                |           |          |   |   |                  |  |  |
| Showing 1 to 2 of 2 entries                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                                  |                        |                         |                 |                        |                         |                      |                       |                                 |                               |                            |              |               |                |                        |                         |            |                |                        |                         |                      |                       |                                 |                               |                            |               |          |           |                     |                 |                 |                |             |                     |       |              |               |      |                  |                     |              |              |                |             |                     |               |              |               |      |                  |     |      |                 |         |       |         |       |          |        |                 |                                                |               |       |               |               |               |               |               |                |               |          |   |   |   |   |   |   |   |                  |       |       |       |                  |   |                                                  |        |   |                 |                 |               |                     |                   |                 |                 |       |   |   |   |   |     |     |                |           |          |   |   |                  |  |  |
| <span>Previous</span> <span>1</span> <span>Next</span>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                                                  |                        |                         |                 |                        |                         |                      |                       |                                 |                               |                            |              |               |                |                        |                         |            |                |                        |                         |                      |                       |                                 |                               |                            |               |          |           |                     |                 |                 |                |             |                     |       |              |               |      |                  |                     |              |              |                |             |                     |               |              |               |      |                  |     |      |                 |         |       |         |       |          |        |                 |                                                |               |       |               |               |               |               |               |                |               |          |   |   |   |   |   |   |   |                  |       |       |       |                  |   |                                                  |        |   |                 |                 |               |                     |                   |                 |                 |       |   |   |   |   |     |     |                |           |          |   |   |                  |  |  |

#### ❖ Spark SQL:

| Job Id (Job Group) •                      | Description                                                                                                           | Submitted           | Duration | Stages: Succeeded/Total | Tasks (for all stages): Succeeded/Total |
|-------------------------------------------|-----------------------------------------------------------------------------------------------------------------------|---------------------|----------|-------------------------|-----------------------------------------|
| 21                                        | text at NativeMethodAccessorImpl.java:0<br>text at NativeMethodAccessorImpl.java:0                                    | 2023/10/17 20:08:20 | 2 s      | 1/1                     | 4/4                                     |
| 20                                        | count at NativeMethodAccessorImpl.java:0<br>count at NativeMethodAccessorImpl.java:0                                  | 2023/10/17 20:08:19 | 66 ms    | 1/1 (5 skipped)         | 1/1 (81 skipped)                        |
| 19                                        | count at NativeMethodAccessorImpl.java:0<br>count at NativeMethodAccessorImpl.java:0                                  | 2023/10/17 20:08:18 | 1 s      | 1/1 (4 skipped)         | 25/25 (56 skipped)                      |
| 18 (bb2d61e0-4bf3-413d-8a8d-b71ba88a60d0) | Subquery (runid bb2d61e0-4bf3-413d-8a8d-b71ba88a60d0)<br>\$anonfun\$withThreadLocalCaptured\$1 at FutureTask.java:266 | 2023/10/17 20:08:18 | 42 ms    | 1/1 (2 skipped)         | 1/1 (5 skipped)                         |
| 17                                        | count at NativeMethodAccessorImpl.java:0<br>count at NativeMethodAccessorImpl.java:0                                  | 2023/10/17 20:08:18 | 70 ms    | 1/1 (1 skipped)         | 1/1 (4 skipped)                         |
| 16                                        | count at NativeMethodAccessorImpl.java:0<br>count at NativeMethodAccessorImpl.java:0                                  | 2023/10/17 20:07:56 | 22 s     | 1/1                     | 4/4                                     |
| 15                                        | count at NativeMethodAccessorImpl.java:0<br>count at NativeMethodAccessorImpl.java:0                                  | 2023/10/17 20:07:54 | 2 s      | 1/1 (3 skipped)         | 25/25 (31 skipped)                      |
| 14                                        | count at NativeMethodAccessorImpl.java:0<br>count at NativeMethodAccessorImpl.java:0                                  | 2023/10/17 20:07:49 | 5 s      | 1/1 (2 skipped)         | 26/26 (5 skipped)                       |
| 13                                        | count at NativeMethodAccessorImpl.java:0<br>count at NativeMethodAccessorImpl.java:0                                  | 2023/10/17 20:07:29 | 20 s     | 1/1                     | 4/4                                     |
| 12                                        | count at NativeMethodAccessorImpl.java:0<br>count at NativeMethodAccessorImpl.java:0                                  | 2023/10/17 20:07:29 | 0.3 s    | 1/1                     | 1/1                                     |
| 11                                        | csv at NativeMethodAccessorImpl.java:0<br>csv at NativeMethodAccessorImpl.java:0                                      | 2023/10/17 20:07:25 | 3 s      | 1/1 (4 skipped)         | 25/25 (56 skipped)                      |
| 10 (992898a1-01bf-4930-947a-dee7a40226d0) | Subquery (runid 992898a1-01bf-4930-947a-dee7a40226d0)<br>\$anonfun\$withThreadLocalCaptured\$1 at FutureTask.java:266 | 2023/10/17 20:07:24 | 0.2 s    | 1/1 (2 skipped)         | 1/1 (5 skipped)                         |
| 9                                         | csv at NativeMethodAccessorImpl.java:0<br>csv at NativeMethodAccessorImpl.java:0                                      | 2023/10/17 20:07:23 | 0.2 s    | 1/1 (1 skipped)         | 1/1 (4 skipped)                         |
| 8                                         | csv at NativeMethodAccessorImpl.java:0<br>csv at NativeMethodAccessorImpl.java:0                                      | 2023/10/17 20:07:06 | 17 s     | 1/1                     | 4/4                                     |
| 7                                         | csv at NativeMethodAccessorImpl.java:0<br>csv at NativeMethodAccessorImpl.java:0                                      | 2023/10/17 20:07:01 | 5 s      | 1/1 (3 skipped)         | 25/25 (31 skipped)                      |
| 6                                         | csv at NativeMethodAccessorImpl.java:0<br>csv at NativeMethodAccessorImpl.java:0                                      | 2023/10/17 20:06:48 | 12 s     | 1/1 (2 skipped)         | 26/26 (5 skipped)                       |
| 5                                         | csv at NativeMethodAccessorImpl.java:0<br>csv at NativeMethodAccessorImpl.java:0                                      | 2023/10/17 20:06:18 | 30 s     | 1/1                     | 4/4                                     |
| 4                                         | csv at NativeMethodAccessorImpl.java:0<br>csv at NativeMethodAccessorImpl.java:0                                      | 2023/10/17 20:06:18 | 2 s      | 1/1                     | 1/1                                     |
| 3                                         | csv at NativeMethodAccessorImpl.java:0<br>csv at NativeMethodAccessorImpl.java:0                                      | 2023/10/17 20:06:14 | 0.3 s    | 1/1                     | 1/1                                     |
| 2                                         | csv at NativeMethodAccessorImpl.java:0<br>csv at NativeMethodAccessorImpl.java:0                                      | 2023/10/17 20:06:14 | 0.3 s    | 1/1                     | 1/1                                     |
| 1                                         | csv at NativeMethodAccessorImpl.java:0<br>csv at NativeMethodAccessorImpl.java:0                                      | 2023/10/17 20:05:57 | 16 s     | 1/1                     | 4/4                                     |
| 0                                         | csv at NativeMethodAccessorImpl.java:0<br>csv at NativeMethodAccessorImpl.java:0                                      | 2023/10/17 20:05:50 | 6 s      | 1/1                     | 1/1                                     |

Page: 1 1 Pages. Jump to 1 . Show 100 items in a page. Go

1 Pages. Jump to  . Show  items in a page.

| Jobs                                               |   | Stages                                           |                 | Storage                |                         | Environment |       | Executors       |              | SQL / DataFrame        |             |                         |           |                                  |               |                                        |  |                                      |  |                                  |  |           |  |       |  |           |  |                     |  |              |  |                  |  |                |  |                |  |                     |  |           |  |              |  |               |  |        |  |                  |  |
|----------------------------------------------------|---|--------------------------------------------------|-----------------|------------------------|-------------------------|-------------|-------|-----------------|--------------|------------------------|-------------|-------------------------|-----------|----------------------------------|---------------|----------------------------------------|--|--------------------------------------|--|----------------------------------|--|-----------|--|-------|--|-----------|--|---------------------|--|--------------|--|------------------|--|----------------|--|----------------|--|---------------------|--|-----------|--|--------------|--|---------------|--|--------|--|------------------|--|
| a01-big-data-analytics-sql-approach application UI |   |                                                  |                 |                        |                         |             |       |                 |              |                        |             |                         |           |                                  |               |                                        |  |                                      |  |                                  |  |           |  |       |  |           |  |                     |  |              |  |                  |  |                |  |                |  |                     |  |           |  |              |  |               |  |        |  |                  |  |
| <b>Executors</b>                                   |   |                                                  |                 |                        |                         |             |       |                 |              |                        |             |                         |           |                                  |               |                                        |  |                                      |  |                                  |  |           |  |       |  |           |  |                     |  |              |  |                  |  |                |  |                |  |                     |  |           |  |              |  |               |  |        |  |                  |  |
| Show Additional Metrics                            |   |                                                  |                 |                        |                         |             |       |                 |              |                        |             |                         |           |                                  |               |                                        |  |                                      |  |                                  |  |           |  |       |  |           |  |                     |  |              |  |                  |  |                |  |                |  |                     |  |           |  |              |  |               |  |        |  |                  |  |
| <b>Summary</b>                                     |   |                                                  |                 |                        |                         |             |       |                 |              |                        |             |                         |           |                                  |               |                                        |  |                                      |  |                                  |  |           |  |       |  |           |  |                     |  |              |  |                  |  |                |  |                |  |                     |  |           |  |              |  |               |  |        |  |                  |  |
| ▲                                                  |   | RDD Blocks                                       | Storage Memory  | On Heap Storage Memory | Off Heap Storage Memory | Disk Used   | Cores | Active Tasks    | Failed Tasks | Complete Tasks         | Total Tasks | Task Time (GC Time)     | Input     | Shuffle Read                     | Shuffle Write | Excluded                               |  |                                      |  |                                  |  |           |  |       |  |           |  |                     |  |              |  |                  |  |                |  |                |  |                     |  |           |  |              |  |               |  |        |  |                  |  |
| Active(2)                                          | 0 | 0.0 B / 3.1 GiB                                  | 0.0 B / 3.1 GiB | 0.0 B / 0.0 B          | 0.0 B / 0.0 B           | 0.0 B       | 4     | 0               | 0            | 186                    | 186         | 12 min (10 s)           | 942.9 MiB | 196.4 MiB                        | 196.4 MiB     | 0                                      |  |                                      |  |                                  |  |           |  |       |  |           |  |                     |  |              |  |                  |  |                |  |                |  |                     |  |           |  |              |  |               |  |        |  |                  |  |
| Dead(0)                                            | 0 | 0.0 B / 0.0 B                                    | 0.0 B / 0.0 B   | 0.0 B / 0.0 B          | 0.0 B / 0.0 B           | 0.0 B       | 0     | 0               | 0            | 0                      | 0           | 0.0 ms (0.0 ms)         | 0.0 B     | 0.0 B                            | 0.0 B         | 0                                      |  |                                      |  |                                  |  |           |  |       |  |           |  |                     |  |              |  |                  |  |                |  |                |  |                     |  |           |  |              |  |               |  |        |  |                  |  |
| Total(2)                                           | 0 | 0.0 B / 3.1 GiB                                  | 0.0 B / 3.1 GiB | 0.0 B / 0.0 B          | 0.0 B / 0.0 B           | 0.0 B       | 4     | 0               | 0            | 186                    | 186         | 12 min (10 s)           | 942.9 MiB | 196.4 MiB                        | 196.4 MiB     | 0                                      |  |                                      |  |                                  |  |           |  |       |  |           |  |                     |  |              |  |                  |  |                |  |                |  |                     |  |           |  |              |  |               |  |        |  |                  |  |
| <b>Executors</b>                                   |   |                                                  |                 |                        |                         |             |       |                 |              |                        |             |                         |           |                                  |               |                                        |  |                                      |  |                                  |  |           |  |       |  |           |  |                     |  |              |  |                  |  |                |  |                |  |                     |  |           |  |              |  |               |  |        |  |                  |  |
| Show 20 entries                                    |   |                                                  |                 |                        |                         |             |       |                 |              |                        |             |                         |           |                                  |               |                                        |  |                                      |  |                                  |  |           |  |       |  |           |  |                     |  |              |  |                  |  |                |  |                |  |                     |  |           |  |              |  |               |  |        |  |                  |  |
| Search:                                            |   |                                                  |                 |                        |                         |             |       |                 |              |                        |             |                         |           |                                  |               |                                        |  |                                      |  |                                  |  |           |  |       |  |           |  |                     |  |              |  |                  |  |                |  |                |  |                     |  |           |  |              |  |               |  |        |  |                  |  |
|                                                    |   |                                                  |                 |                        |                         |             |       |                 |              |                        |             |                         |           |                                  |               |                                        |  |                                      |  |                                  |  |           |  |       |  |           |  |                     |  |              |  |                  |  |                |  |                |  |                     |  |           |  |              |  |               |  |        |  |                  |  |
| Executor ID                                        |   | Address                                          |                 | Status                 |                         | RDD Blocks  |       | Storage Memory  |              | On Heap Storage Memory |             | Off Heap Storage Memory |           | Peak JVM Memory OnHeap / OffHeap |               | Peak Execution Memory OnHeap / OffHeap |  | Peak Storage Memory OnHeap / OffHeap |  | Peak Pool Memory Direct / Mapped |  | Disk Used |  | Cores |  | Resources |  | Resource Profile Id |  | Active Tasks |  | Failed Tasks     |  | Complete Tasks |  | Total Tasks    |  | Task Time (GC Time) |  | Input     |  | Shuffle Read |  | Shuffle Write |  | Logs   |  | Exec Loss Reason |  |
| driver                                             |   | ip-10-0-25-6.ap-south-1.compute.internal:41621   |                 | Active                 |                         | 0           |       | 0.0 B / 1 GiB   |              | 0.0 B / 1 GiB          |             | 0.0 B / 0.0 B           |           | 0.0 B / 0.0 B                    |               | 0.0 B / 0.0 B                          |  | 0.0 B / 0.0 B                        |  | 0.0 B / 0.0 B                    |  | 0         |  | 0     |  | 0         |  | 0                   |  | 0            |  | 2.8 min (0.0 ms) |  | 0.0 B          |  | 0.0 B          |  | 0.0 B               |  | stdout    |  | stderr       |  |               |  |        |  |                  |  |
| 1                                                  |   | ip-10-0-25-109.ap-south-1.compute.internal:44569 |                 | Active                 |                         | 0           |       | 0.0 B / 2.1 GiB |              | 0.0 B / 2.1 GiB        |             | 0.0 B / 0.0 B           |           | 1.8 GiB / 1412.1 MiB             |               | 384 MiB / 0.0 B                        |  | 6.8 MiB / 0.0 B                      |  | 22.5 KiB / 0.0 B                 |  | 0.0 B     |  | 4     |  | 0         |  | 0                   |  | 0            |  | 186              |  | 186            |  | 8.9 min (10 s) |  | 942.9 MiB           |  | 196.4 MiB |  | 196.4 MiB    |  | stdout        |  | stderr |  |                  |  |

| Completed Stages (22) |                                                                                                                       |                              |          |                        |           |          |              |               |
|-----------------------|-----------------------------------------------------------------------------------------------------------------------|------------------------------|----------|------------------------|-----------|----------|--------------|---------------|
| Page: 1               |                                                                                                                       |                              |          |                        |           |          |              |               |
| Stage Id              | Description                                                                                                           | Submitted                    | Duration | Tasks: Succeeded/Total | Input     | Output   | Shuffle Read | Shuffle Write |
| 50                    | text at NativeMethodAccessorImpl.java:0                                                                               | +details 2023/10/17 20:08:20 | 2 s      | 4/4                    | 5.0 B     |          |              |               |
| 49                    | count at NativeMethodAccessorImpl.java:0                                                                              | +details 2023/10/17 20:08:19 | 42 ms    | 1/1                    |           | 1250.0 B |              |               |
| 43                    | count at NativeMethodAccessorImpl.java:0                                                                              | +details 2023/10/17 20:08:18 | 1 s      | 25/25                  |           | 11.3 MiB | 1250.0 B     |               |
| 38                    | Subquery (runId bb2d61e0-4bf3-413d-8a8d-b71ba88a60d0)<br>\$anonfun\$withThreadLocalCaptured\$1 at FutureTask.java:266 | +details 2023/10/17 20:08:18 | 22 ms    | 1/1                    |           | 50.0 B   |              |               |
| 35                    | count at NativeMethodAccessorImpl.java:0                                                                              | +details 2023/10/17 20:08:18 | 46 ms    | 1/1                    |           | 3.0 KIB  | 50.0 B       |               |
| 33                    | count at NativeMethodAccessorImpl.java:0                                                                              | +details 2023/10/17 20:07:56 | 22 s     | 4/4                    | 188.5 MiB |          | 3.0 KIB      |               |
| 32                    | count at NativeMethodAccessorImpl.java:0                                                                              | +details 2023/10/17 20:07:54 | 2 s      | 25/25                  |           | 22.9 MiB | 11.3 MiB     |               |
| 28                    | count at NativeMethodAccessorImpl.java:0                                                                              | +details 2023/10/17 20:07:49 | 5 s      | 26/26                  |           | 64.0 MiB | 22.9 MiB     |               |
| 25                    | count at NativeMethodAccessorImpl.java:0                                                                              | +details 2023/10/17 20:07:29 | 20 s     | 4/4                    | 188.5 MiB |          | 63.9 MiB     |               |
| 24                    | count at NativeMethodAccessorImpl.java:0                                                                              | +details 2023/10/17 20:07:29 | 0.2 s    | 1/1                    | 56.9 KIB  |          | 28.3 KIB     |               |
| 23                    | csv at NativeMethodAccessorImpl.java:0                                                                                | +details 2023/10/17 20:07:25 | 3 s      | 25/25                  |           | 12.4 KIB | 11.3 MiB     |               |
| 18                    | Subquery (runId 992898a1-01bf-4930-947a-dee7a40226d0)<br>\$anonfun\$withThreadLocalCaptured\$1 at FutureTask.java:266 | +details 2023/10/17 20:07:24 | 62 ms    | 1/1                    |           | 50.0 B   |              |               |
| 15                    | csv at NativeMethodAccessorImpl.java:0                                                                                | +details 2023/10/17 20:07:23 | 0.2 s    | 1/1                    |           | 3.0 KIB  | 50.0 B       |               |
| 13                    | csv at NativeMethodAccessorImpl.java:0                                                                                | +details 2023/10/17 20:07:06 | 17 s     | 4/4                    | 188.5 MiB |          | 3.0 KIB      |               |
| 12                    | csv at NativeMethodAccessorImpl.java:0                                                                                | +details 2023/10/17 20:07:01 | 5 s      | 25/25                  |           | 22.9 MiB | 11.3 MiB     |               |
| 8                     | csv at NativeMethodAccessorImpl.java:0                                                                                | +details 2023/10/17 20:06:48 | 12 s     | 26/26                  |           | 64.0 MiB | 22.9 MiB     |               |
| 5                     | csv at NativeMethodAccessorImpl.java:0                                                                                | +details 2023/10/17 20:06:18 | 29 s     | 4/4                    | 188.5 MiB |          | 63.9 MiB     |               |
| 4                     | csv at NativeMethodAccessorImpl.java:0                                                                                | +details 2023/10/17 20:06:18 | 2 s      | 1/1                    | 56.9 KIB  |          | 28.3 KIB     |               |
| 3                     | csv at NativeMethodAccessorImpl.java:0                                                                                | +details 2023/10/17 20:06:14 | 0.3 s    | 1/1                    | 56.9 KIB  |          |              |               |
| 2                     | csv at NativeMethodAccessorImpl.java:0                                                                                | +details 2023/10/17 20:06:14 | 0.2 s    | 1/1                    | 7.5 KIB   |          |              |               |
| 1                     | csv at NativeMethodAccessorImpl.java:0                                                                                | +details 2023/10/17 20:05:57 | 16 s     | 4/4                    | 188.5 MiB |          |              |               |
| 0                     | csv at NativeMethodAccessorImpl.java:0                                                                                | +details 2023/10/17 20:05:50 | 6 s      | 1/1                    | 1536.0 B  |          |              |               |

Page: 1

1 Pages. Jump to 1 . Show 100 items in a page. Go

ii. Compare the resource usage, execution time, and any other observable metrics between the DataFrame and SQL approaches in the EMR UI.

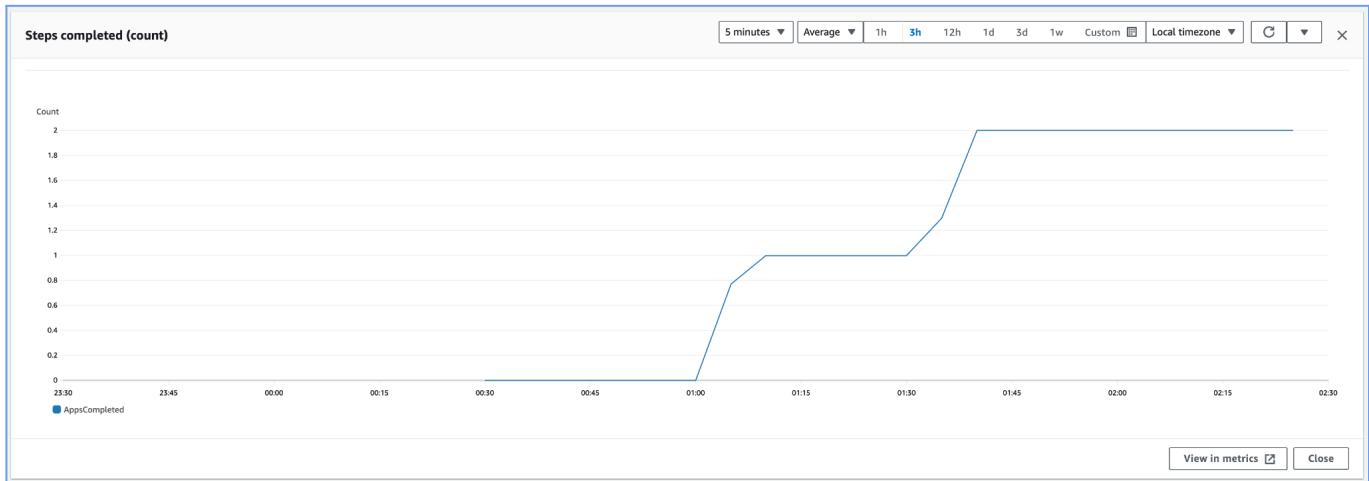
❖ Comparison using Spark history UI:

\*I have used green to indicate the best.

| Approach                       | DataFrame API             | SQL API                   |
|--------------------------------|---------------------------|---------------------------|
| Total uptime                   | 2.2 min                   | 2.8 min                   |
| Number of ran jobs             | 17                        | 22                        |
| Duration of the long run job   | 25 s                      | 30 s                      |
| Number of needed queries       | 6                         | 8                         |
| Query with longest duration    | Initial df => required df | Initial df => required df |
| Duration of the long run query | 36 s (ran by 4 jobs)      | 1.2 min (ran by 8 jobs)   |
| Completed stages               | 17                        | 22                        |
| Skipped stages                 | 12                        | 29                        |
| Duration of the long run stage | 25 s                      | 29 s                      |
| Tasks used in CSV write        | 25 Tasks                  | 25 Tasks                  |

Based on the Spark history UI information, we can say that the DataFrame approach is better than SQL approach.

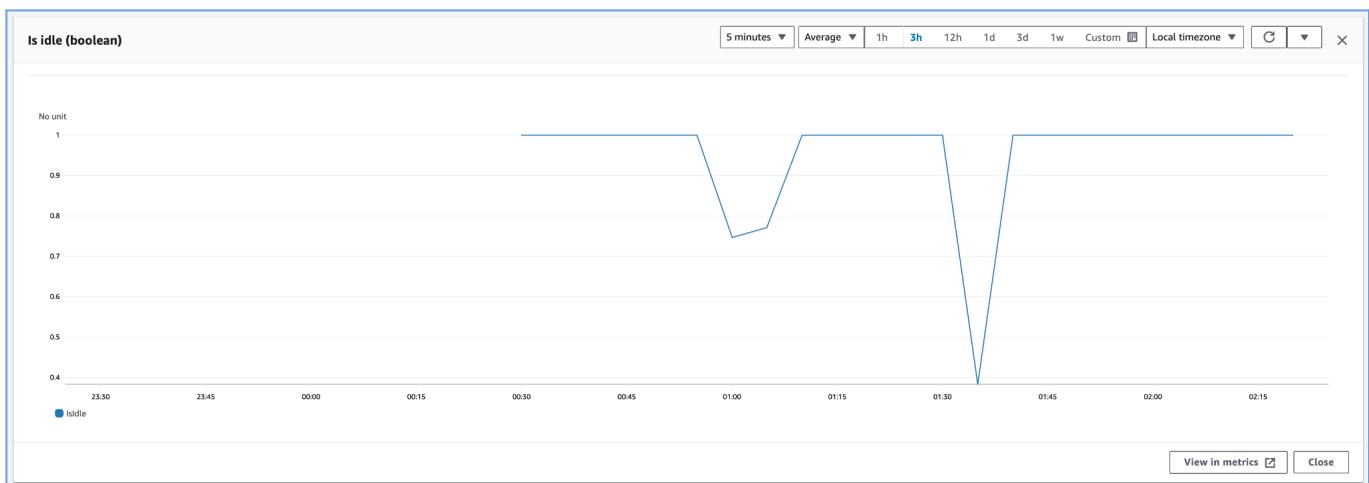
❖ Comparison using EMR monitoring information:



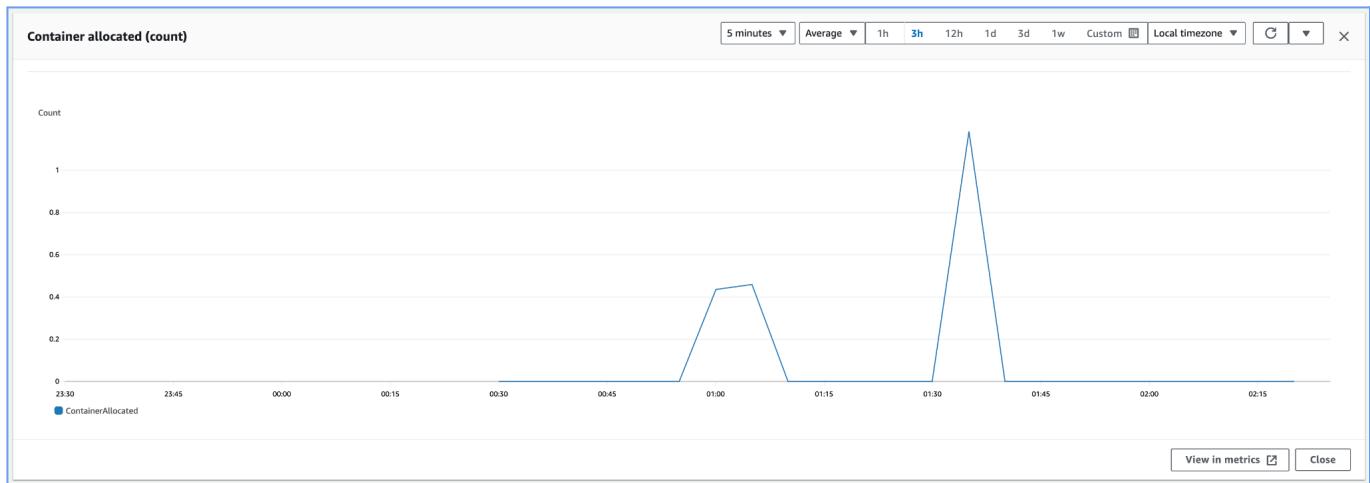
I have executed 2 steps one after the other. First, I ran the steps for the DataFrame approach. Therefore, the execution time of each approach:

1. DataFrame approach: Around 1:00 - 1:15 am
2. SQL approach: Around 1:30 - 1:45 am

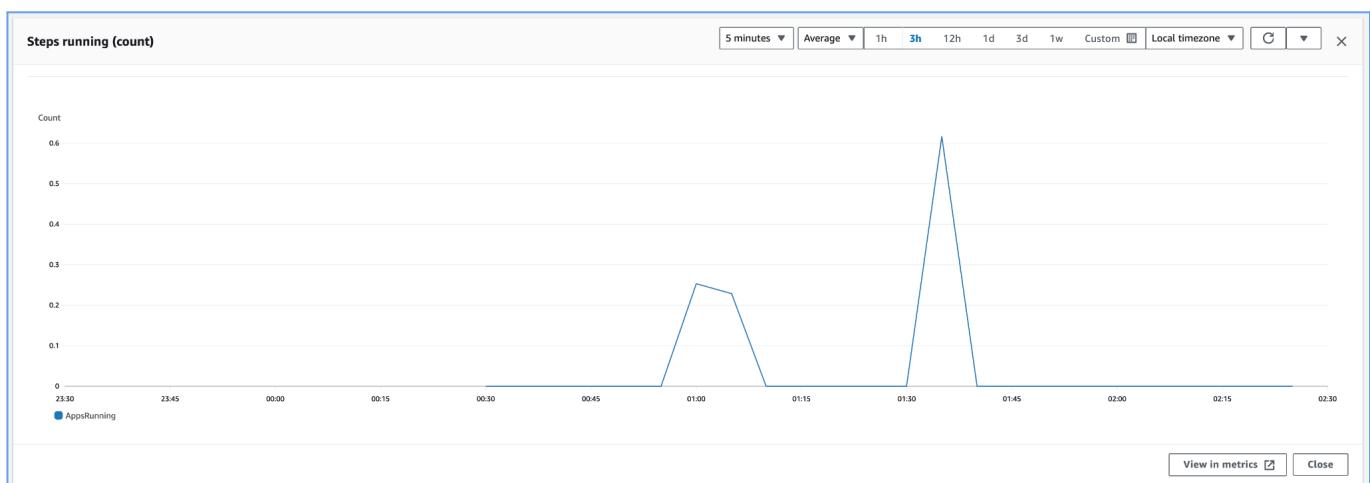
Let's compare both approaches based on the monitoring details. I have used green to indicate the best in each comparison.



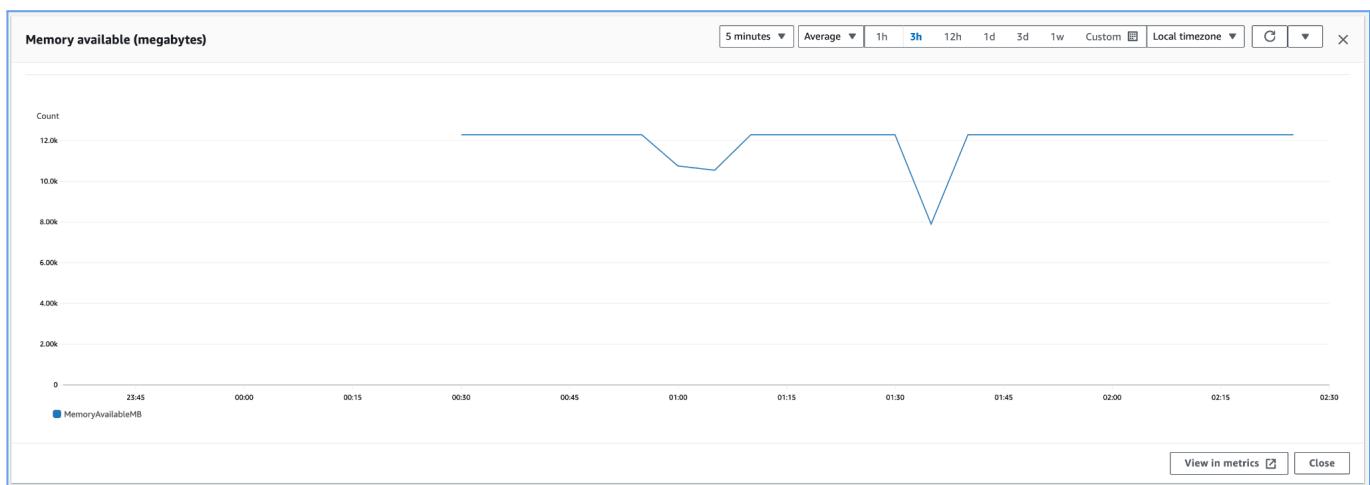
Boolean value of the EMR cluster is idle: **DataFrame approach > SQL approach**



The Number of containers were allocated: **DataFrame approach < SQL approach**



The Number of steps ran: **DataFrame approach < SQL approach**



Available memory of EMR cluster during the of step execution: **DataFrame approach > SQL approach**

Based on the EMR monitoring information also, we can say that the DataFrame approach is better than SQL approach. Because, to do the same task, the EMR cluster has run fewer steps in less memory during the DataFrame approach execution than that of SQL approach.