

New York Institute of Technology

College of Engineering and Computing Sciences

Big Data Project Report

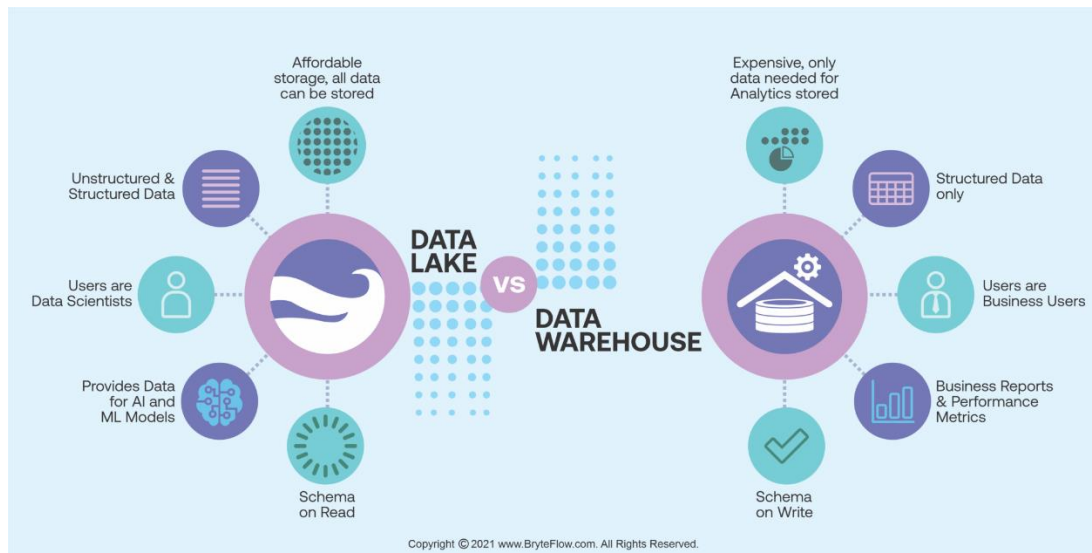
Professor: Liangwen Wu

GAJJELA HARIKA | TEJASWI VENUMULA | RAGHURAM BOBBA | NITHIN RAO DUGYALA

Aim: The objective of the project is to generate visual representations of an analysis on New York City parking tickets. This involves conducting diverse aggregations and showcasing a variety of statistics obtained from the dataset.

Data Warehouse vs Data Lake:

- **Data Warehouse:** A data warehouse is a centralized and organized storage system for a company's information. It acts like a comprehensive library, bringing together data from different parts of the business into one accessible place. This makes it easier for people in the company to analyze and understand the data, aiding in better decision-making.
- **Data Lake:** A data lake is like a vast, unstructured reservoir for storing all types of data—structured, semi-structured, and unstructured—without the need for predefined formats. It's a flexible storage space where businesses can collect and keep large amounts of raw data from various sources. Unlike a data warehouse, which organizes data neatly, a data lake allows for more flexibility and exploration, making it possible to analyze diverse and evolving datasets for insights and decision-making.
- A data warehouse and a data lake are both storage architectures designed for handling and analyzing large volumes of data,
- In our Project, we used Data Warehouse, as it specially designed for structured data, it also ensures data consistency, accuracy along with that it helps in performing complex queries and analysis on aggregated data.



Pictorial Representation of Data Lake vs Data Lake

Data Collection:

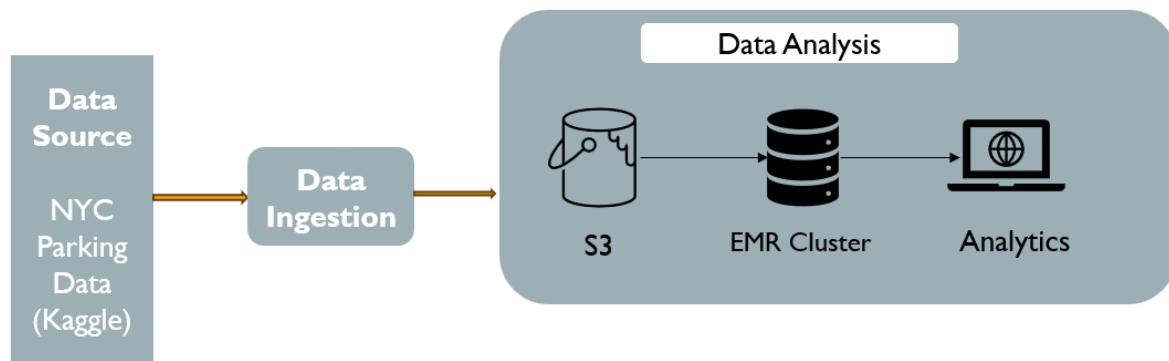
In this Project we have taken dataset from Kaggle, from the below link.

<https://www.kaggle.com/datasets/new-york-city/nyc-parking-tickets/data>

Data Pipelining of the Project:

A data pipeline is a set of processes and tools that move and transform data from one or multiple sources to a destination for analysis, it involves the extraction, transformation, and loading (ETL) of data, ensuring that it flows smoothly from source to destination.

In this Project, we commonly used Amazon Simple Storage Service (S3) and Amazon Elastic MapReduce (EMR) are two services offered by Amazon Web Services (AWS). Where S3 is Amazon Simple Storage Service (S3) and Amazon Elastic MapReduce (EMR) are two services offered by Amazon Web Services (AWS) and Amazon EMR (Elastic MapReduce) is a cloud-based big data platform that utilizes Apache Hadoop and Apache Spark to process and analyze large datasets. EMR clusters can be part of the data processing steps in a data pipeline. Data from S3 can be loaded into an EMR cluster for processing and analysis using tools like Apache Spark or Apache Hive. The results of these computations can then be stored back in S3 or sent to other destinations.



AWS Pipeline

Analysis of the Project:

S3 Setup:

- **Sign in to AWS Console:** Log in to AWS Management console, once logged in, find and click on “S3” under storage.
- **Create S3 bucket:** Click the "Create bucket" button > Choose a globally unique bucket name and select the region for the bucket > Configure other settings like versioning, logging, and tags>Click "Create bucket" to finish.
- **Configure bucket permissions:** Click on the newly created bucket > Navigate to the "Permissions" tab > Adjust the bucket policy and access control list (ACL) settings.
- **Document Upload:** Click on the "Upload" button > Click "Add files" or "Add folder" to select the document to upload > Configure optional settings like permissions, metadata, and storage class > Click "Upload" to upload the document.
- **Document URL:** Once the document is uploaded, select it in the S3 bucket > In the details pane, find the "Object URL" or "Link" and get the URL Path of the document.

EMR Cluster setup:

- **Sign in to AWS Console:** Log in to AWS Management console, once logged in,

find and click on "EMR" under the "Analytics" section in the AWS Management Console.

- **Creating a Cluster:** Click the "Create cluster" button > Choose the "Go to advanced options" to customize the cluster.
 - **Configure cluster settings:** Enter name for the cluster > Configure logging options.
 - **Choose Hardware and software:** Select the EMR release version > Choose the spark application or framework > Choose 1 primary node and 5 core nodes.
 - **Security Configuration:** Configure EC2 key pair, IAM role.
- Launch Cluster:** click on "Create cluster" to launch EMR cluster.

Data Processing:

- **EMR Cluster Creation:** Once the EMR Cluster is created, the cluster status will transition to “Waiting”.

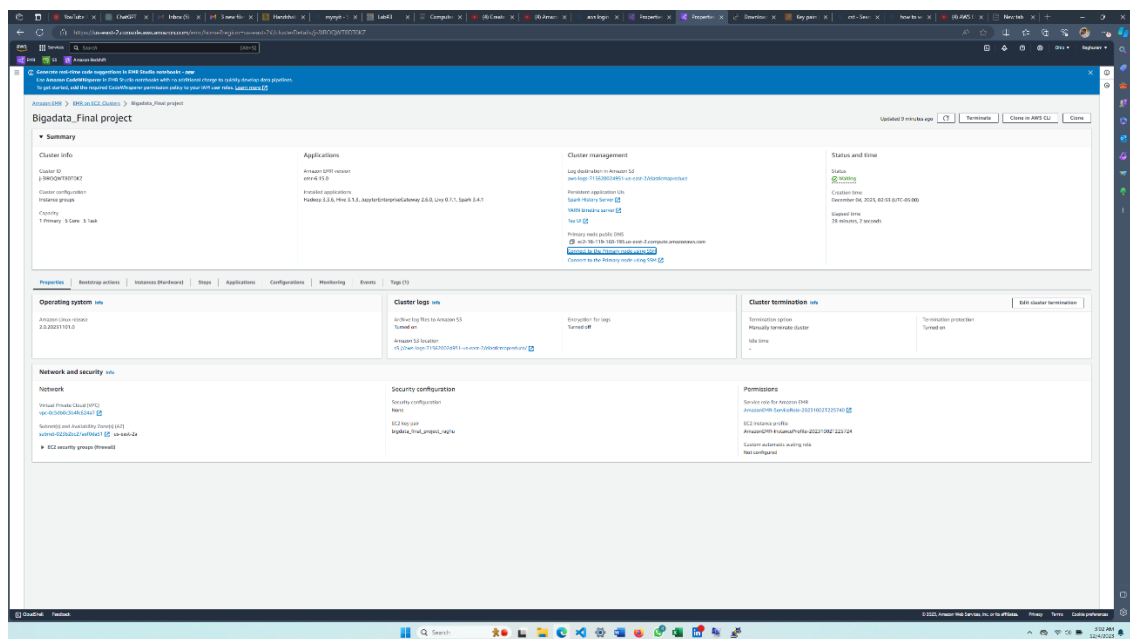


Figure 1

- **Connect to SSH:** Click on "Connect to Remote SSH" to access the remote Linux/Unix system hosted on your EMR cluster.

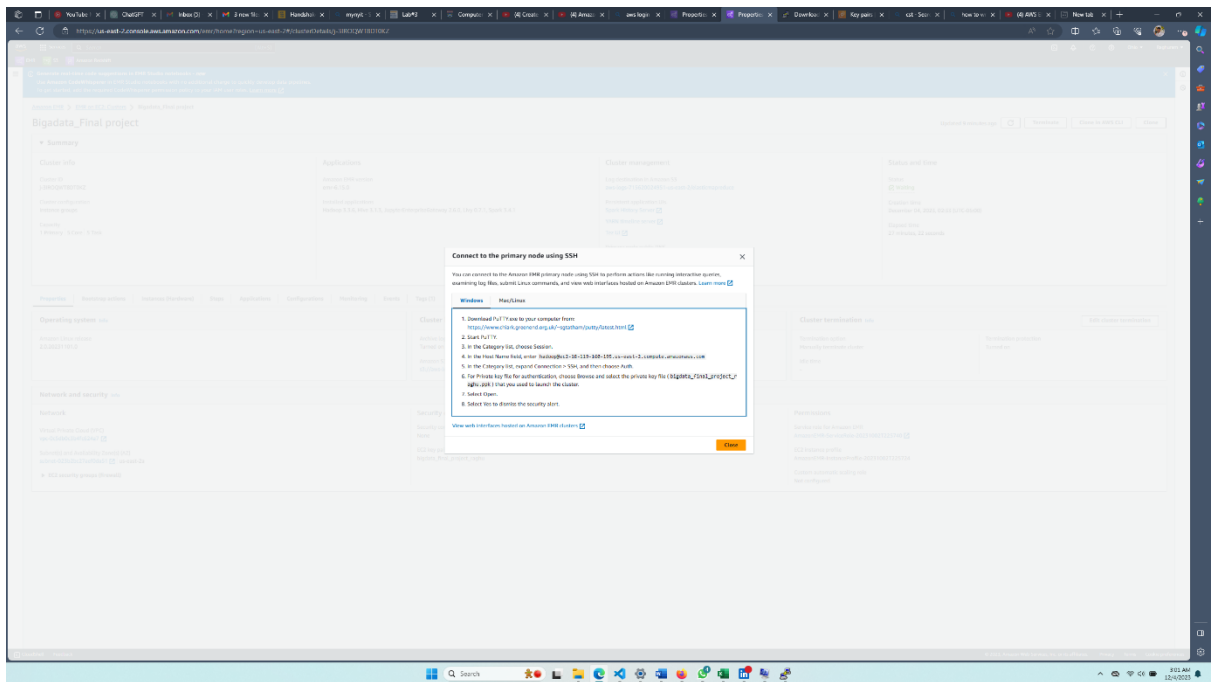


Figure 2

- **Establish Remote connection using Putty:** By using putty, we get access to connect to the remote server which helps to access the local system.



Figure 3

- **Access to the remote Linux/Unix server:** Upon successful connection, gain access to the remote Linux/Unix system from your local environment. This establishes a seamless bridge between your local system and the EMR cluster.

- **Python file creation:** Generate a Python file (e.g., filename.py) to develop the code for data processing on the EMR cluster.
- **Initiate the spark code:** Save the Python file and initiate the Spark job using the following command:
spark-submit filename.py.
- Once the spark job has been submitted, it calls python file and execute the code and displays the respective results.

```

23/12/04 20:35:12 INFO BlockManagerInfo: Removed broadcast_14_piece0 on ip-172-31-13-133.us-east-2.compute.internal:42233 in memory (size: 13.3 KiB, free: 911.9 MiB)
23/12/04 20:35:12 INFO BlockManagerInfo: Removed broadcast_14_piece0 on ip-172-31-2-185.us-east-2.compute.internal:34995 in memory (size: 13.3 KiB, free: 4.8 GiB)
23/12/04 20:35:12 INFO BlockManagerInfo: Removed broadcast_14_piece0 on ip-172-31-6-111.us-east-2.compute.internal:37253 in memory (size: 13.3 KiB, free: 4.8 GiB)
23/12/04 20:35:12 INFO BlockManagerInfo: Removed broadcast_14_piece0 on ip-172-31-4-160.us-east-2.compute.internal:39485 in memory (size: 13.3 KiB, free: 4.8 GiB)
23/12/04 20:35:12 INFO BlockManagerInfo: Removed broadcast_14_piece0 on ip-172-31-2-228.us-east-2.compute.internal:43419 in memory (size: 13.3 KiB, free: 4.8 GiB)
23/12/04 20:35:12 INFO BlockManagerInfo: Removed broadcast_14_piece0 on ip-172-31-14-29.us-east-2.compute.internal:44955 in memory (size: 13.3 KiB, free: 4.8 GiB)
23/12/04 20:35:12 INFO RepOutputTrackerMasterEndpoint: Asked to send map output locations for shuffle 6 to 172.31.14.29:47946
23/12/04 20:35:12 INFO BlockManagerInfo: Removed broadcast_16_piece0 on ip-172-31-2-228.us-east-2.compute.internal:43419 in memory (size: 19.1 KiB, free: 4.8 GiB)
23/12/04 20:35:12 INFO BlockManagerInfo: Removed broadcast_16_piece0 on ip-172-31-6-111.us-east-2.compute.internal:37253 in memory (size: 19.1 KiB, free: 4.8 GiB)
23/12/04 20:35:12 INFO BlockManagerInfo: Removed broadcast_16_piece0 on ip-172-31-2-185.us-east-2.compute.internal:34995 in memory (size: 19.1 KiB, free: 4.8 GiB)
23/12/04 20:35:12 INFO BlockManagerInfo: Removed broadcast_16_piece0 on ip-172-31-4-160.us-east-2.compute.internal:39485 in memory (size: 19.1 KiB, free: 4.8 GiB)
23/12/04 20:35:12 INFO BlockManagerInfo: Removed broadcast_16_piece0 on ip-172-31-13-133.us-east-2.compute.internal:43289 in memory (size: 19.1 KiB, free: 4.8 GiB)
23/12/04 20:35:12 INFO BlockManagerInfo: Removed broadcast_16_piece0 on ip-172-31-10-65.us-east-2.compute.internal:38007 in memory (size: 19.1 KiB, free: 4.8 GiB)
23/12/04 20:35:12 INFO BlockManagerInfo: Removed broadcast_16_piece0 on ip-172-31-5-103.us-east-2.compute.internal:45627 in memory (size: 19.1 KiB, free: 4.8 GiB)
23/12/04 20:35:12 INFO BlockManagerInfo: Removed broadcast_16_piece0 on ip-172-31-1-53.us-east-2.compute.internal:38173 in memory (size: 19.1 KiB, free: 4.8 GiB)
23/12/04 20:35:12 INFO BlockManagerInfo: Removed broadcast_16_piece0 on ip-172-31-8-64.us-east-2.compute.internal:46827 in memory (size: 19.1 KiB, free: 4.8 GiB)
23/12/04 20:35:12 INFO BlockManagerInfo: Removed broadcast_16_piece0 on ip-172-31-14-29.us-east-2.compute.internal:44955 in memory (size: 19.1 KiB, free: 4.8 GiB)
23/12/04 20:35:12 INFO BlockManagerInfo: Removed broadcast_16_piece0 on ip-172-31-13-133.us-east-2.compute.internal:42233 in memory (size: 19.1 KiB, free: 911.9 MiB)
23/12/04 20:35:12 INFO BlockManagerInfo: Removed broadcast_20_piece0 on ip-172-31-5-64.us-east-2.compute.internal:46827 in memory (size: 13.2 KiB, free: 4.8 GiB)
23/12/04 20:35:12 INFO BlockManagerInfo: Removed broadcast_20_piece0 on ip-172-31-6-111.us-east-2.compute.internal:37253 in memory (size: 19.1 KiB, free: 4.8 GiB)
23/12/04 20:35:12 INFO BlockManagerInfo: Removed broadcast_20_piece0 on ip-172-31-13-133.us-east-2.compute.internal:43289 in memory (size: 19.1 KiB, free: 4.8 GiB)
23/12/04 20:35:12 INFO BlockManagerInfo: Removed broadcast_20_piece0 on ip-172-31-10-65.us-east-2.compute.internal:38007 in memory (size: 19.1 KiB, free: 4.8 GiB)
23/12/04 20:35:12 INFO BlockManagerInfo: Removed broadcast_20_piece0 on ip-172-31-2-228.us-east-2.compute.internal:43419 in memory (size: 19.1 KiB, free: 4.8 GiB)
23/12/04 20:35:12 INFO BlockManagerInfo: Removed broadcast_20_piece0 on ip-172-31-14-29.us-east-2.compute.internal:44955 in memory (size: 19.1 KiB, free: 4.8 GiB)
23/12/04 20:35:12 INFO BlockManagerInfo: Removed broadcast_20_piece0 on ip-172-31-13-133.us-east-2.compute.internal:42233 in memory (size: 19.1 KiB, free: 911.9 MiB)
23/12/04 20:35:12 INFO TaskSetManager: Finished task 0.0 in stage 21.0 (TID 29) in 79 ms on ip-172-31-14-29.us-east-2.compute.internal (executor 9) (1/1)
23/12/04 20:35:12 INFO DAGScheduler: Removed TaskSet 21.0, whose tasks have all completed, from pool
23/12/04 20:35:12 INFO DAGScheduler: ResultStage 21 (showString at NativeMethodAccessorImpl.java:0) finished in 0.107 s
23/12/04 20:35:12 INFO DAGScheduler: Job 14 is finished. Cancelling potential speculative or zombie tasks for this job
23/12/04 20:35:12 INFO YarnScheduler: Killing all running tasks in stage 21: Stage finished.
23/12/04 20:35:12 INFO BlockManagerInfo: Removed broadcast_5_piece0 on ip-172-31-13-133.us-east-2.compute.internal:42233 in memory (size: 4.3 KiB, free: 911.9 MiB)
23/12/04 20:35:12 INFO DAGScheduler: Job 14 finished: showString at NativeMethodAccessorImpl.java:0, took 0.110119 s
=====
|Vehicle Body Type| count|
=====
SUM|161649|
4DSD|300389|
VAN|137456|
EXV| 64890|
SM| 41649|
2DSD|26842|
FIC| 21644|
REFG| 9778|
TRAC| 7022|
TAM| 5836|
UTIL| 5734|
4 DR| 5067|
BSE| 4177|
CONV| 4384|
null| 3993|
WAGO| 3624|
4D| 3418|
MCV| 2873|
TK| 2832|
TMS| 2463|
=====
only showing top 20 rows

```

Figure 4

```
23/12/04 20:41:12 INFO TaskSetManager: Finished task 37.0 in stage 22.0 (TID 339) in 841 ms on ip-172-31-10-63.us-east-2.compute.internal (executor 2) (34/40)
23/12/04 20:41:12 INFO TaskSetManager: Finished task 17.0 in stage 22.0 (TID 309) in 555 ms on ip-172-31-10-63.us-east-2.compute.internal (executor 2) (35/40)
23/12/04 20:41:12 INFO TaskSetManager: Finished task 27.0 in stage 22.0 (TID 319) in 579 ms on ip-172-31-10-63.us-east-2.compute.internal (executor 2) (36/40)
23/12/04 20:41:12 INFO TaskSetManager: Finished task 26.0 in stage 22.0 (TID 318) in 586 ms on ip-172-31-4-160.us-east-2.compute.internal (executor 8) (37/40)
23/12/04 20:41:12 INFO TaskSetManager: Finished task 7.0 in stage 22.0 (TID 259) in 595 ms on ip-172-31-10-63.us-east-2.compute.internal (executor 2) (38/40)
23/12/04 20:41:12 INFO TaskSetManager: Finished task 6.0 in stage 22.0 (TID 260) in 600 ms on ip-172-31-4-160.us-east-2.compute.internal (executor 8) (39/40)
23/12/04 20:41:12 INFO TaskSetManager: Finished task 16.0 in stage 22.0 (TID 308) in 663 ms on ip-172-31-4-160.us-east-2.compute.internal (executor 8) (40/40)
23/12/04 20:41:12 INFO YarnScheduler: Removed TaskSet 22.0, whose tasks have all completed, from pool
23/12/04 20:41:12 INFO DAGScheduler: ShuffleMapStage 22 (showString at NativeMethodAccessorImpl.java:0) finished in 0.675 s
23/12/04 20:41:12 INFO DAGScheduler: looking for newly runnable stages
23/12/04 20:41:12 INFO DAGScheduler: running: Set()
23/12/04 20:41:12 INFO DAGScheduler: waiting: Set()
23/12/04 20:41:12 INFO DAGScheduler: failed: Set()
23/12/04 20:41:12 INFO ShufflePartitionUtil: For shuffle(7), advisory target size: 67108864, actual target size: 1048576, minimum partition size: 1048576
23/12/04 20:41:12 INFO ShufflePartitionUtil: For shuffle(7), advisory target size: 67108864, actual target size: 1048576, minimum partition size: 1048576
23/12/04 20:41:12 INFO HashAggregateExec: spark.sql.codegen.aggregate.map.twolevel.enabled is set to true, but current version of codegen fast hashmap does not support this aggregate.
23/12/04 20:41:12 INFO SparkContext: Starting job: showString at NativeMethodAccessorImpl.java:0
23/12/04 20:41:12 INFO DAGScheduler: Got job 16 (showString at NativeMethodAccessorImpl.java:0) with 2 output partitions
23/12/04 20:41:12 INFO DAGScheduler: Final stage: ResultStage 24 (showString at NativeMethodAccessorImpl.java:0)
23/12/04 20:41:12 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 23)
23/12/04 20:41:12 INFO DAGScheduler: Missing parents: List()
23/12/04 20:41:12 INFO DAGScheduler: Submitting ResultStage 24 (MapPartitionsRDD[72] at showString at NativeMethodAccessorImpl.java:0), which has no missing parents
23/12/04 20:41:12 INFO MemoryStore: Block broadcast_26 stored as values in memory (estimated size 25.9 KiB, free 910.8 MiB)
23/12/04 20:41:12 INFO MemoryStore: Block broadcast_26_piece0 stored as bytes in memory (estimated size 13.2 KiB, free 910.7 MiB)
23/12/04 20:41:12 INFO BlockManagerInfo: Added broadcast_26_piece0 in memory on ip-172-31-13-133.us-east-2.compute.internal:40943 (size: 13.2 KiB, free: 912.1 MiB)
23/12/04 20:41:12 INFO DAGScheduler: Submitting 2 missing tasks from ResultStage 24 (MapPartitionsRDD[72] at showString at NativeMethodAccessorImpl.java:0) (first 15 tasks are for partitions Vector(0, 1))
23/12/04 20:41:12 INFO YarnScheduler: Adding task set 24.0 with 2 tasks resource profile 0
23/12/04 20:41:12 INFO TaskSetManager: Starting task 0.0 in stage 24.0 (TID 332) (ip-172-31-15-105.us-east-2.compute.internal, executor 5, partition 0, PROCESS_LOCAL, 7374 bytes)
23/12/04 20:41:12 INFO TaskSetManager: Starting task 1.0 in stage 24.0 (TID 333) (ip-172-31-14-29.us-east-2.compute.internal, executor 3, partition 1, PROCESS_LOCAL, 7374 bytes)
23/12/04 20:41:12 INFO DAGScheduler: ResultStage 24 (showString at NativeMethodAccessorImpl.java:0) finished in 0.480 s
23/12/04 20:41:12 INFO DAGScheduler: Added broadcast_26_piece0 in memory on ip-172-31-14-29.us-east-2.compute.internal:43321 (size: 13.2 KiB, free: 4.8 GiB)
23/12/04 20:41:12 INFO MapOutputTrackerMasterEndpoint: Asked to send map output locations for shuffle 7 to 172.31.15.105:43958
23/12/04 20:41:12 INFO MapOutputTrackerMasterEndpoint: Asked to send map output locations for shuffle 7 to 172.31.14.29:50974
23/12/04 20:41:12 INFO TaskSetManager: Finished task 0.0 in stage 24.0 (TID 332) in 229 ms on ip-172-31-15-105.us-east-2.compute.internal (executor 5) (1/2)
23/12/04 20:41:12 INFO TaskSetManager: Finished task 1.0 in stage 24.0 (TID 333) in 476 ms on ip-172-31-14-29.us-east-2.compute.internal (executor 3) (2/2)
23/12/04 20:41:12 INFO YarnScheduler: Removed TaskSet 24.0, whose tasks have all completed, from pool
23/12/04 20:41:12 INFO DAGScheduler: ResultStage 24 (showString at NativeMethodAccessorImpl.java:0) finished in 0.480 s
23/12/04 20:41:12 INFO DAGScheduler: Job 16 is finished. Cancelling potential speculative or zombie tasks for this job
23/12/04 20:41:12 INFO YarnScheduler: Killing all running tasks in stage 24: Stage finished
23/12/04 20:41:12 INFO DAGScheduler: Job 16 finished: showString at NativeMethodAccessorImpl.java:0, took 0.483915 s

-----
[Violation Time] count
-----
1140A| 24621
0836A| 24781
1136A| 25653
0804A| 24671
0840A| 24561
0536A| 24251
1145A| 23981
1142A| 23821
0840A| 23781
0958A| 23531
1138A| 23371
0810A| 23271
0812A| 22971
0843A| 22821
0453A| 22611
1139A| 22491
0804A| 22371
0838A| 22441
1150A| 22421
0937A| 22381
-----
only showing top 20 rows
```

Figure 5

```
hadoop@ip-172-31-13-133:~$
23/12/04 20:41:12 INFO DAGScheduler: Submitting ResultStage 21 (MapPartitionsRDD[64] at showString at NativeMethodAccessorImpl.java:0), which has no missing parents
23/12/04 20:41:12 INFO MemoryStore: Block broadcast_23 stored as values in memory (estimated size 28.9 KiB, free 910.1 MiB)
23/12/04 20:41:12 INFO BlockManagerInfo: Removed broadcast_9_piece0 on ip-172-31-15-105.us-east-2.compute.internal:40717 in memory (size: 42.9 KiB, free: 4.8 GiB)
23/12/04 20:41:12 INFO BlockManagerInfo: Removed broadcast_9_piece0 on ip-172-31-4-103.us-east-2.compute.internal:37857 in memory (size: 42.9 KiB, free: 4.8 GiB)
23/12/04 20:41:12 INFO BlockManagerInfo: Removed broadcast_9_piece0 on ip-172-31-6-111.us-east-2.compute.internal:43353 in memory (size: 42.9 KiB, free: 4.8 GiB)
23/12/04 20:41:12 INFO BlockManagerInfo: Removed broadcast_9_piece0 on ip-172-31-2-228.us-east-2.compute.internal:44663 in memory (size: 42.9 KiB, free: 4.8 GiB)
23/12/04 20:41:12 INFO BlockManagerInfo: Removed broadcast_9_piece0 on ip-172-31-4-160.us-east-2.compute.internal:33741 in memory (size: 42.9 KiB, free: 4.8 GiB)
23/12/04 20:41:12 INFO BlockManagerInfo: Removed broadcast_9_piece0 on ip-172-31-2-188.us-east-2.compute.internal:33259 in memory (size: 42.9 KiB, free: 4.8 GiB)
23/12/04 20:41:12 INFO MemoryStore: Block broadcast_23_piece0 stored as bytes in memory (estimated size 18.2 KiB, free 910.6 MiB)
23/12/04 20:41:12 INFO BlockManagerInfo: Added broadcast_23_piece0 in memory on ip-172-31-14-29.us-east-2.compute.internal:43321 in memory (size: 42.9 KiB, free: 4.8 GiB)
23/12/04 20:41:12 INFO BlockManagerInfo: Added broadcast_23_piece0 in memory on ip-172-31-1-53.us-east-2.compute.internal:39177 in memory (size: 42.9 KiB, free: 4.8 GiB)
23/12/04 20:41:12 INFO BlockManagerInfo: Added broadcast_9_piece0 on ip-172-31-5-64.us-east-2.compute.internal:42407 in memory (size: 42.9 KiB, free: 4.8 GiB)
23/12/04 20:41:12 INFO BlockManagerInfo: Added broadcast_23_piece0 in memory on ip-172-31-13-133.us-east-2.compute.internal:40943 (size: 13.2 KiB, free: 912.0 MiB)
23/12/04 20:41:12 INFO SparkContext: Created broadcast 23 from DAGScheduler.scala:11892
23/12/04 20:41:12 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 21 (MapPartitionsRDD[64] at showString at NativeMethodAccessorImpl.java:0) (first 15 tasks are for partitions Vector(0))
23/12/04 20:41:12 INFO YarnScheduler: Adding task set 21.0 with 1 tasks resource profile 0
23/12/04 20:41:12 INFO TaskSetManager: Starting task 0.0 in stage 21.0 (TID 294) (ip-172-31-2-188.us-east-2.compute.internal, executor 1, partition 0, PROCESS_LOCAL, 7374 bytes)
23/12/04 20:41:12 INFO BlockManagerInfo: Removed broadcast_9_piece0 on ip-172-31-13-133.us-east-2.compute.internal:40943 in memory (size: 130.3 KiB, free: 912.2 MiB)
23/12/04 20:41:12 INFO BlockManagerInfo: Removed broadcast_9_piece0 on ip-172-31-15-105.us-east-2.compute.internal:40717 in memory (size: 130.3 KiB, free: 4.8 GiB)
23/12/04 20:41:12 INFO MapOutputTrackerMasterEndpoint: Asked to send map output locations for shuffle 6 to 172.31.12.185:34250
23/12/04 20:41:12 INFO TaskSetManager: Finished task 0.0 in stage 21.0 (TID 294) in 42 ms on ip-172-31-2-188.us-east-2.compute.internal (executor 1) (1/1)
23/12/04 20:41:12 INFO YarnScheduler: Removed TaskSet 21.0, whose tasks have all completed, from pool
23/12/04 20:41:12 INFO DAGScheduler: ResultStage 21 (showString at NativeMethodAccessorImpl.java:0) finished in 0.071 s
23/12/04 20:41:12 INFO DAGScheduler: Job 14 is finished. Cancelling potential speculative or zombie tasks for this job
23/12/04 20:41:12 INFO YarnScheduler: Killing all running tasks in stage 21: Stage finished
23/12/04 20:41:12 INFO DAGScheduler: Job 14 finished: showString at NativeMethodAccessorImpl.java:0, took 0.079153 s

-----
[Vehicle Body Type] count
-----
SRW| 3414464
40SD| 3003989
VAN| 1374560
BLV| 648901
SDM| 416499
JSD| 248420
FCR| 216641
RZF| 97781
TRAC| 70221
TXN| 88361
UTL| 57341
4 DR| 56671
BUS| 44171
CONV| 43841
MUL| 39931
WAG| 36241
4D| 36181
NCT| 29721
TR| 28321
TRLE| 26631
-----
only showing top 20 rows
```

Figure 6

Monitor Hardware Instances:

- Navigate to the AWS Console to monitor hardware instances associated with the EMR cluster. This provides an overview of the infrastructure utilized during the data processing task.

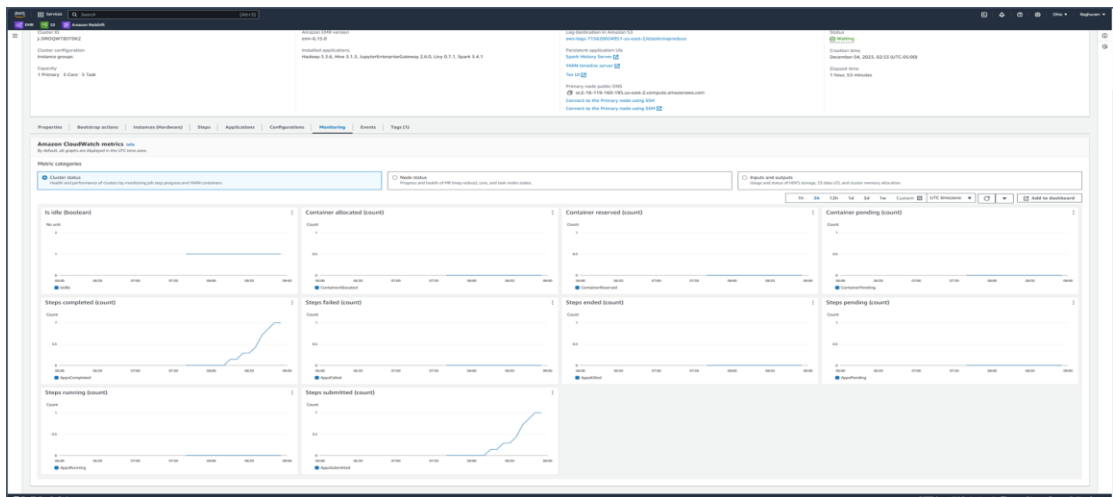


Figure 7

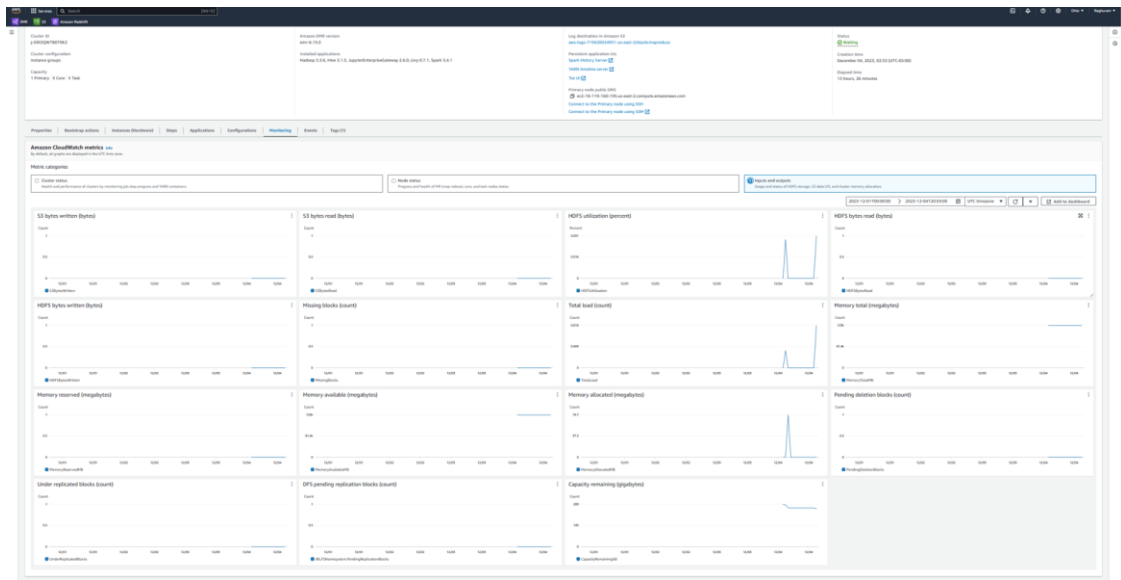


Figure 8

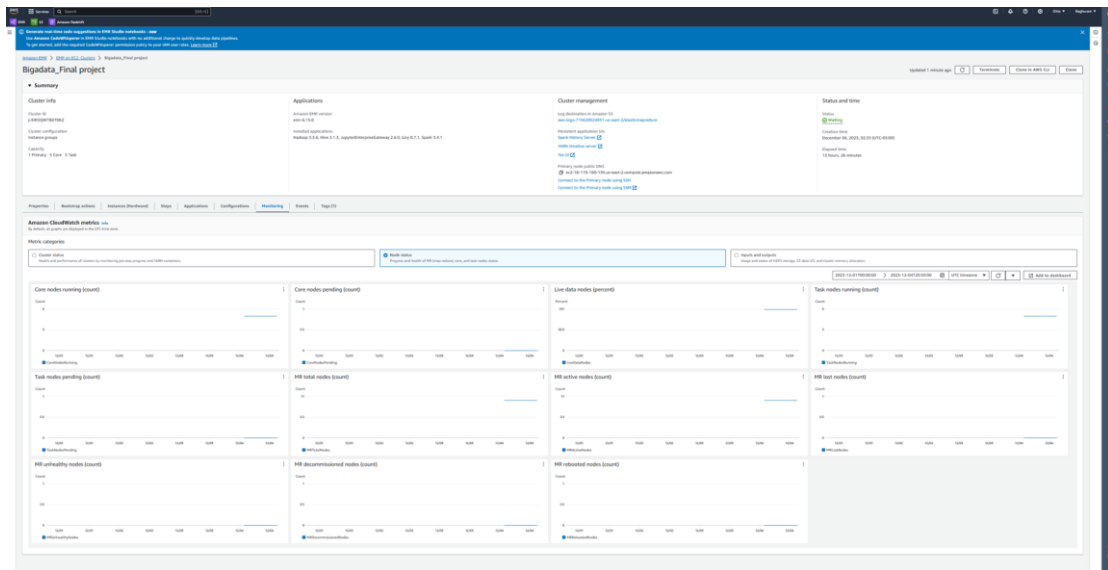


Figure 9

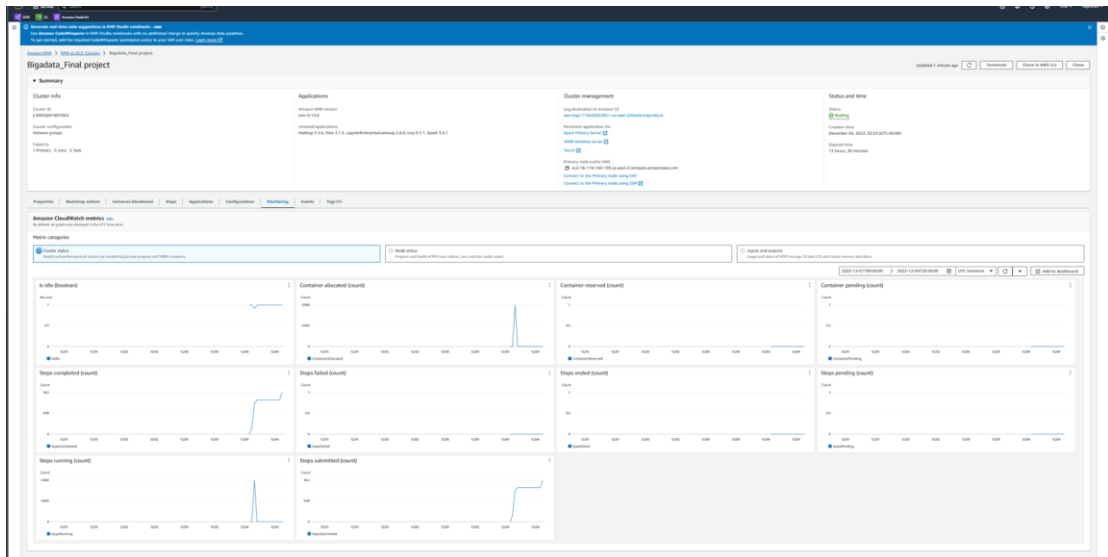


Figure 10

Conclusion:

This project successfully analyzed the NYC parking ticket data using AWS services. The initial findings offer valuable insights into parking trends and violations in the city. Further analysis can be conducted to gain deeper understanding and potentially inform policy decisions or improvements to parking infrastructure.