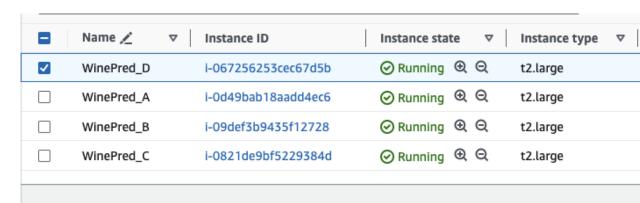
- a) Set up AWS EC2 Instances:
 - Launch 4 EC2 instances for model training and use 1 instance for the prediction application.



- b) Prepare the Environment:
 - Install Python and Apache Spark on all EC2 instances.
 - ii) Clone the repository to each instance or transfer the project files.
- c) Train the Model:
 - i) Execute the Spark application on the 4 instances to train the model using TrainingDataset.csv.

```
ubuntu@ip-_______:-/winePrediction$ ps

PID TTY TIME CMD

1621 pts/0 00:00:00 bash

2820 pts/0 00:00:15 java

4931 pts/0 00:00:00 ps
```

```
Last login: Mon Dec 4 01:01:49 2023 from 

""" sexport SPARK_HOME=/home/ubuntu/spark-3.5.0-bin-hadoop3

""" sexport PATH=$PATH:$SPARK_HOME/bin

""" scd $SPARK_HOME/sbin

""" starting org.apacne.spark.deploy.worker.Worker, logging to /home/ubuntu/spark-3.5.0-bin-hadoop3/logs/spark-ubunt

""" starting org.apacne.spark.deploy.worker.Worker-1-ip-1.2-2-2-1.out

""" spark-3.5.0-bin-hadoop3/sbin$

"" spark-3.5.0-bin-hadoop3/sbin$

""" spar
```

- d) Validate and Optimize the Model:
 - i) Use ValidationDataset.csv to validate and fine-tune the model's parameters.

- e) Build and Deploy the Docker Container:
 - i) Build the Docker container for the prediction application.
 - ii) Deploy the container to the prediction EC2 instance.
- f) Run the Prediction Application:
 - i) Execute the prediction application within Docker, using the test dataset.

