**To Run Model Training on 4 Parallel EC2 Instances using EMR on AWS**

**Create cluster using following steps**

1. Login into AWS Account

2. Open EMR Service

3. Create cluster

a) Config cluster Name

b) Software configuration as Spark

c) Hardware Configuration Number of Instance e.g., 5

d) Security and access Pass EC2 Key pair

4. Click on the security group for master to change inbound rules for traffic.

5. Select ssh to get login

6. Copy program file to cluster from s3

a) Command: aws s3 cp s3://bucket\_name/file\_name

7. Run the program using spark-submit Model.py

8. The model will be saved to the exiting passed path of the bucket.

**To Run Model Prediction on single EC2 Instance without Docker**

Preprocess Configuration log:

a) Create EC2 instance

b) configure with spark

To Run the model:

1) Update the bucket\_name in the Prediction.py

2) Command: aws s3 cp s3://bucket\_name/Prediction.py

3) Run the program using spark-submit Prediction.py

**To Run Model Prediction on single EC2 Instance with Docker**

Preprocess Configuration log:

a) Ec2 instance is created

b) It has docker installed on it

1) Log into Ec2 instance

2) docker pull

docker pull shubham973979/final-project-2-sd858

3) Run this command from the current directory of Validation file:

docker run -it -v "$(pwd)":/file shubham973979/final-project-2-sd858

# Text Description automatically generated