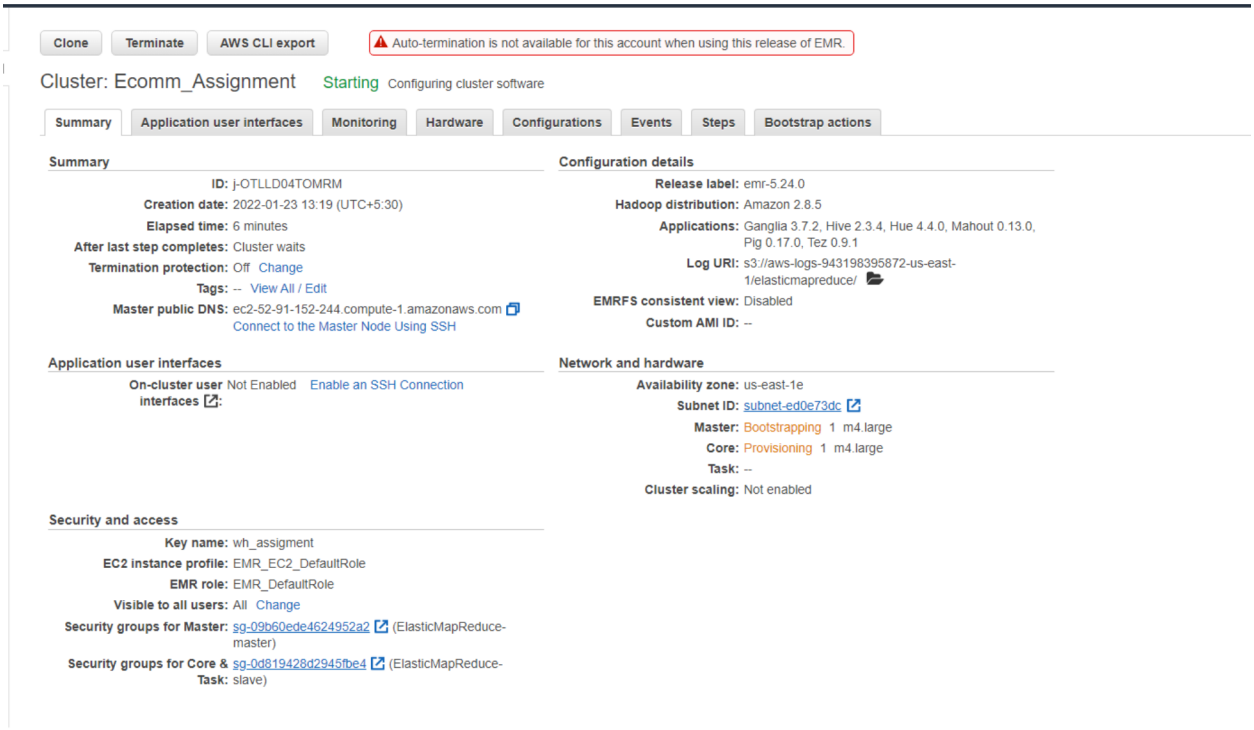
E-Commerce Assignment

Problem Statement:

With online sales gaining popularity, tech companies are exploring ways to improve their sales by analyzing customer behavior and deriving insights about product trends. Furthermore, the e-commerce websites make it easier for customers to find the products that they require without much scavenging. The role of big data engineers is among the most sought-after job profiles of this decade. Therefore, as part of this assignment, you, as a big data engineer, will be required to extract data and gather insights from a real-life data set of an e-commerce company

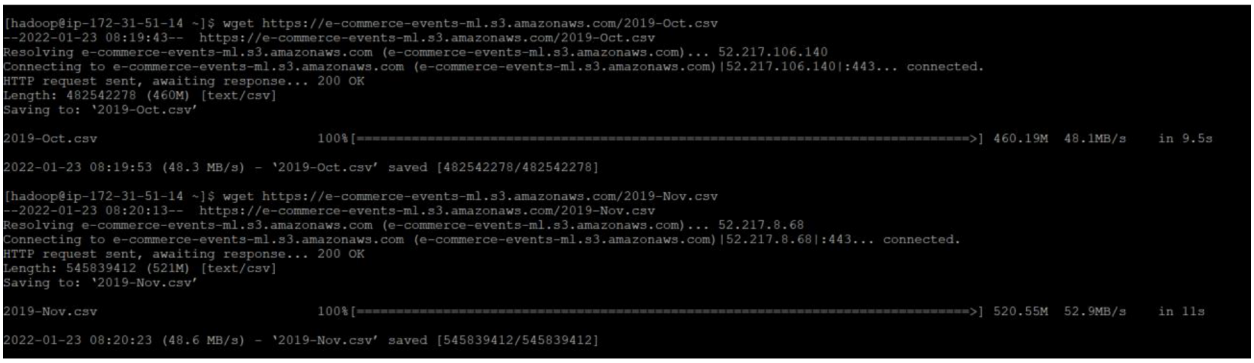
Stepwise Solution to the problem statement:

Step 1: An EMR cluster was created with the desired specifications:



Step 2: An EMR cluster was launched using Putty:

Step 3: Data from s3 bucket was imported to the HDFS using wget command



Step 4: logging into hive and creating database and tables for the analysis. Following types of tables were created:

Step 5: Loading data into the required tables