**ETL Data Pipeline using EMR Cluster**

* We first created key-value pair in EC2. This led us to create EMR cluster in AWS.
* Then we went on o create SE bucket in order to add a file and save it in folder.
* We connected EMR master node using SSH. Then using Tableau, we visualized data.
* The data is Sales data. We connected Hive to Tableau using ODBC driver.
* The observations that were made are as follows.
  + Sales are maximum in Sub-Saharan Africa followed by Europe.
  + Cuba has maximum sales in the list of countries.
  + Then we analyzed the data on quarterly basis. In 2013-Q2, the company made highest sales.
  + Maximum number of orders were made by Czech Republic.

Chart, line chart

Description automatically generatedChart

Description automatically generatedMap

Description automatically generatedTable

Description automatically generatedBar chart

Description automatically generatedChart, treemap chart

Description automatically generatedChart

Description automatically generated with low confidenceA picture containing graphical user interface

Description automatically generated