Data Management, Warehousing, And Analytics

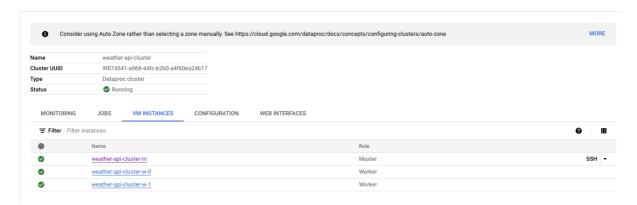
Lab5: Big Data, Hadoop & Apache Spark (Summer 2023)

Submitted by: Arihant Dugar (B00917961)

GitLab repo: https://git.cs.dal.ca/dugar/csci5408_s23_b00917961_arihant_dugar/-/tree/main/Lab5

Lab Exercise:

1. Setup Apache Spark On GCP Instance:



2. Write a Java program to filter the response data where the daily "feels_like" temperature for the next 5-days is greater than 15°C during the "day" time. Exclude the current, minutely, and hourly fields.

The code can be found in the repository - https://git.cs.dal.ca/dugar/csci5408 s23 b00917961 arihant dugar/- tree/main/Lab5

3. Save the filtered data into a new file – "summer_weather.json"

The file summer_weather.json is also attached for reference to view the filtered data. Below is the output of the code that ran on the spark instance.

Note:

Everything is uploaded to the Gitlab repo as well for reference.