

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 :

Consider using Auto Zone rather than selecting a zone manually. See <https://cloud.google.com/dataproc/docs/concepts/configuring-clusters/auto-zone> [MORE](#)

Name	weather-api-cluster
Cluster UUID	9f074541-a968-44fc-b2b0-a4f60ea24b17
Type	Dataproc cluster
Status	Running

MONITORINGJOBSVM INSTANCESCONFIGURATIONWEB INTERFACES

Filter instances

	Name	Role	
✓	weather-api-cluster-m	Master	SSH
✓	weather-api-cluster-w-0	Worker	
✓	weather-api-cluster-w-1	Worker	

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.

```
arihant0996@weather-api-cluster-m:~$ spark-submit --jars gson-2.2.2.jar,json-simple-1.1.1.jar --class Main weather.jar
arihant0996@weather-api-cluster-m:~$ ls
gson-2.2.2.jar  json-simple-1.1.1.jar  summer_weather.json  weather.jar  weather.json
arihant0996@weather-api-cluster-m:~$ cat summer_weather.json
{
  "timezone": "America/Halifax",
  "timezone_offset": -10800,
  "daily": [
    {
      "moonset": 1655920320,
      "sunrise": 1655886551,
      "temp": {
        "min": 9.44,
        "max": 17.74,
        "ave": 15.62,
        "night": 11.48,
        "day": 17.74,
        "morn": 11.27
      },
      "moon_phase": 0.8,
      "lunr": 9.48,
      "moonrise": 1655873940,
      "pressure": 1024,
      "clouds": 9,
      "feels_like": {
        "eve": 15.07,
        "night": 11.11,
        "day": 17.19,
        "morn": 10.8
      },
      "wind_gust": 6.89,
      "dt": 1655913600,
      "pop": 0,
      "wind_deg": 192,
      "dew_point": 9.67,
      "sunset": 1655942600,
      "weather": [
        {
          "icon": "01d",
          "description": "clear sky",
          "main": "Clear",
          "id": 800
        }
      ],
      "humidity": 62,
      "wind_speed": 4.75
    },
    {
      "moonset": 1656010740,
      "rain": 12.63,
      "sunrise": 1655972966,
```

Note:

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