Experiment - 13.

(Group B3)

- Title : Locate dataset (eg. sample - weather - txt) for working on weather data which reads the text input files and finds average for temperature, dew point and wind speed.

- Pre-requisite: Basics of java.

- Objective: Students must be able to perform operations on text data.

- THEORY :

(1) Scraping weather data -

- Wunderground has a 'personal' weather station for which fantastic historical weather data is available. - covering temperature, pressure, wind speed and direction and rainfall, all available on minute level.

- It has interactive graphs and downloadable .csv. datasets.

(2) Cleansing and data processing -

- We need to apply pre-processing techniques to clean the data and bring it to desired form.

(3). Data summarization and aggregation.

The data can then be aggregate to generate comprehensive series.

(4). Implementation -

Given a weather log file with entries formatted as -

Date, Temperature (°C), Dew Point (°C), Wind Speed (km/h).

- The goal is to compute Average temperature, average due point, average wind speed.

* Approach -

(a) Mapper class -

- Processes such as each line of weather log.
- Extracts temperature, dew point and wind speed values.
- Emits key value pairs for each metric.
  ("temperature", 15.5)
  ("dewpoint", 10.2)
  ("windspeed", 5.6).

(b) Reducer class -

- For each key (temperature, dewpoint, windspeed), receives a list of views values.
- Sums all values and counts the number of records.
- Computes and emits the average of each metric.

5] Why hadoop map reduce?

- Distributed processing - efficient for handling large datasets.
- Fault tolerance - automatically recovers from failures.
- Scalability - Suitable for petabyte scale data.
- Parallel execution - Runs mappers and reducers in parallel for performance gains.

6] Applications of weather data analysis -

- Predictive weather modelling and climate research.
- Agricultural planning and crop advisory.
- Air quality and environmental monitoring.
- Renewable energy forecasting (wind, solar).

* **Conclusion**: Hence we successfully performed operations on data using text file inputs.

* * * * * * *