Analysis of Weather Data using Map Reduce

Jay D. Bodra

University of Texas at Arlington

Email: jay.bodra@mavs.uta.edu

Problem Statement

- Analyzing the weather data in order to find the stations with similar weather conditions for each month over the period of 5 year in Texas.
- Texas Weather Data to be used
- Attributes that can be considered:
  - Station Id: STN
  - Reading Date and Time: yearModa_hr
  - Weather Details: Hourly Temperature, Hourly Mean Wind Speed, Hourly Dew Point

Implementation: Specifications

- Weather patterns vary throughout the day.
  - For example: afternoon temperatures are higher than midnight temperatures.
- Two stations' weather would be similar only if their weather patterns are similar for every corresponding section of the day.
- Hence, the 4 sections of the day which we have considered are:
  - 1st section: 5 am – 11 am
  - 2nd section: 11 am – 5 pm
  - 3rd section: 5 pm – 11 pm
  - 4th section: 11 pm – 5 am

Implementation: MR Job 1

- A 3 attribute vector comprising of average temperature, average dew point and average wind speed calculated for a particular section of a day over a month is generated for each station.

Implementation: MR Job 2

- For each month a 12 attribute vector comprising of section wise average values of temperature, dew point and wind speed is generated for each station.

Sample Output

Key	Value
K1	63.1157|45.3379|8.7268|63.0115|45.3379|8.7268|63.1724|45.3379|8.7268|63.192|45.3379|8.7268
K2	78.0263|66.5433|7.3703|78.0521|66.5433|7.3703|77.9562|66.5433|7.3703|78.2049|66.5433|7.3703
K3	68.1174|51.8039|6.5092|68.1279|51.8039|6.5092|68.1162|51.8039|6.5092|68.1135|51.8039|6.5092
K4	75.0606|0.0|10.4278|74.9556|0.0|10.4278|75.0274|0.0|10.4277|75.0273|0.0|10.4277
K5	72.0442|59.2331|5.36|72.0997|59.2331|5.36|72.044|59.2331|5.36|72.038|59.2331|5.36

Thank You !!!