Dukkipati, Sai Kavya

UTA Id: 1000980778

CSE6331-CLOUD COMPUTING

**ASSIGNMENT-2**

DEVELOPMENT

* Hadoop version: 1.2.1
* Operating System: UBUNTU-12.04
* Java version: Java jdk-1.6.0\_32
* Java Hadoop development environment: Eclipse

Dataset:

The average temperatures for every year and then for every month and then for every season in the data set are calculated. The dataset used is a comma-separated one which contains ‘date’ in 2nd column and ‘temp’ in 3rd column.

1. Calculation of average temperature for years:

In the calculation of average temperature for year only year is extracted from the whole ‘date’ string and used as key in the mapper i.e. if date format is “yyyymmdd” , “yyyy” is used a key.

Changes in CPU time for (mappers, reducers):

(1, 1)- CPU time (ms)-3160

(2, 1)- CPU time (ms)-4240

(2, 2)- CPU time (ms)-4580

(10, 1)- CPU time (ms)-3320

(10, 2)- CPU time (ms)-13930

1. Calculation of average temperature for months:

In the calculation of average temperature for month only year and month are extracted from the whole ‘date’ string and used as key in the mapper i.e. if date format is “yyyymmdd” , “yyyymm” is used a key.

Changes in CPU time for (mappers, reducers):

(1, 1)- CPU time (ms)-2320

(2, 1)- CPU time (ms)-3070

(2, 2)- CPU time (ms)-3410

(10, 1)- CPU time (ms)-5950

(10, 2)- CPU time (ms)-9800

1. Calculation of average temperature for seasons:

In the calculation of average temperature for month only year and month are extracted from the whole ‘date’ string and used as key in the mapper i.e. if date format is “yyyymmdd” , “yyyymm” is used a key.

* 03<=mm<=05: Spring
* 06<=mm<=08: Summer
* 09<=mm<=11: Autumn
* mm= 12 or 01 or 02: Winter

Changes in CPU time for (mappers, reducers):

(1, 1)- CPU time (ms)-1870

(2, 1)- CPU time (ms)-5330

(2, 2)- CPU time (ms)-4820

(10, 1)- CPU time (ms)-5300

(10, 2)- CPU time (ms)-9790