**CSE DBMS 5331 Project – 3 Report**

**Team No. 2**

**Hemakumar Gokulakannan(1001415255)**

**Overall Status**

The Project has been completed and all the methods that were to be implemented have been completed. Methods that have been implemented in this are Map, Reduce, Sampler and Partitioner.

1. **Map():**

This method was used to extract Temperature fields from the the **.g** filesprovided to us. Once they were extracted we set the Temperature as **Key** and the entire row with all the values, including the Temperature, as Value to the Partitioner.

1. **Partitioner():**

In this function we first split the Temperature according to different ranges as specified in the code. Based on the temperature value and the ranges the temperature values are sent to their corresponding partition. After this they are sent to the reducer.

1. **Sampler():** This function we used a random sampler which filtered Temperature data based on Sampling value, which was .01 in our case. We also set a limit on maximum records which was 10,000.
2. **Reducer():**

Reducer takes the input from each partition and writes the output to corresponding output file in a sorted order.

**Performance Measure**

|  |  |  |  |
| --- | --- | --- | --- |
| **System Configuration : Linux OS, 8 GB RAM.** | | | |
| **Task** | **# of Mappers** | **# of Reducers** | **Time Taken(in Mins)** |
| Mapper+Partitioner+Reducer | 1 | 2 | 2.5 |
| Sampler+Mapper+Reducer | 1 | 2 | 5.5 |

**File Description**

* TempSort.java – This file contains all the function as described in the above sections.

**Logical Errors and How they were Handled**

1. Hadoop Setup:

The setup must be done very carefully. The files that need to be edited, like core-site.xml, yarn-site.xml, mapred-site.xml and hdfs-site.xml. One mistake in editing these can result in improper configuration.

1. Formatting and Starting Order:

You must know the exact order in which you need to format the your Hadoop Node. Improper start and stop of the node can cause it to behave aberrantly. You must first execute hdfs namenode –format, then start-dfs.sh and lastly start-yarn.sh. If you need to reformat the node at anytime make sure there is no Job running.

1. DoubleWritable Not a class error.

We go this error while running the Mapper and Reducer. Sometime the data which we were sending from Mapper was not what the Reducer was expecting. We corrected this error by making the data type same(DoubleWriteable) at both the Mapper output and Reducer Input end.