B649 Cloud Computing

Project1 Report

Shashi Shankar, Chaitra Thipanna

* Transformation of data during the computations, i.e. data type of key, value

Data Pre-processing: Created a file by parsing the original input file and grouped them in 20.

Intermediate Computations: Computed the partial values for sum, min, max and sum of squares for the above data sets and wrote them into a CompositeWritable object. This object serves as input to reducers. Where after shuffle and sort the combined global values are calculated.

* The data structure used to transfer between Map and Reduce phases

Format of input in Map phase: <LongWritable, Text> as a <key, value> pair

Format of input from Reduce phase: <Text, CompositeWritable> as a key value pair

Output of the Reducer: <Text, DoubleWritable>.

* How the data flow happens through disk and memory during the computation

The pre-processed input file is stored in the input directory on the local disk. This input file is then copied to the HDFS. Then it is read line by line and the values are then stored in a double array in the memory.

Attached: screenshot of output

