For the assignment, Students form a team of two and should pick one domain. The team should demonstrate (with example and source code) how the following concepts can be used to solve a particular problem in the selected domain. One from each team has to upload this document along with mapreduce source code(zipped) to cse portal at (<http://mycse.mahe.manipal.net/mod/assign/view.php?id=422>):

**Assignment 1:**

The data should be stored in MongoDb database and MapReduce should be used for computation.

Team Members

|  |  |
| --- | --- |
| Name | Registration No. |
| Yash Swaika | 120905638 |
| Sristi Mohanty | 120905244 |

**Selected Domain :**

Data for Red Wine Quality Testing.

**Explain what fields are present in csv file used to import data to mongodb:**

**Fixed acid, Volatile acidity, citric acid, residual sugar, free sulfur dioxide, total sulfur dioxide, sulphates, density, pH, alcohol, quality.**

The grouping and classification was done on residual sugar ingredient.

Command used to import data to mongodb:

Mongoimport -db test –c ass1 –type csv --file ass.csv --headerline

mongodb query:

db.createCollection(“ass1”);

db.ass1.find().pretty();

Command used to export data from mongodb:

The output is from the Map Reduce program

Explain what problem is solved using MapReduce:

The csv file is taken as the input and mapReduce is applied on it. The various attributes are stored in the collection as documents. The MapReduce programs groups and classifies the % residual sugar from red wine for quality testing.

Briefly explain what is done in mapper:

The residual sugar is taken as the key for the mapper and the % of that attribute as value. The output is the list of values of various % for residual sugar.

Briefly explain what is done in reducer:

Various ranges are set to check the interval and reducer is applied. The ranges are taken as the key and expected result is in the form of output from reducer. The reduced result is stored in output file.