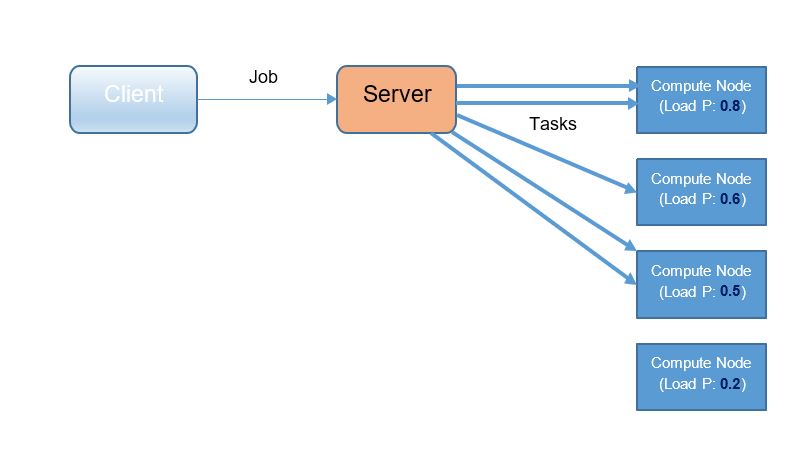
Design Document Anubhav Panda (panda047) Sanjit Dash(dash0030)  
Design of The PA1



**Over View**

In this project we have implemented a simple Map reduce Algorithm for Sentiment Analysis using Java and Thrift .

The Design of the PA1 can be divided into following sections   
1 Client  
2 Thrift files used for client server communication

3 Server

4 Thrift files used for server compute node communication

5 Compute Node

**Client**

The client reads the input file names from the input directory (./data/input\_dir) and stores it in a collection and it uses Thrift communication to send the data to the server and it uses the port number 9095 for the communication and the server should be hosted in the same machine as client as it is using local host. The following files are the part of the client section

Client.java

**Thrift files used for client server communication**

The following are the thrift files that are used for client server communication .It contains two methods   
ping to check there can be a communication between client and the server and process which is used to process the map reduce algorithm and it takes the argument which contains the list of input file names.

SentimentAnalysisServer.Thrift

**Server**

The server receives the job submitted by the client.it then selects the compute node for each of the file by and based on the configuration it also pass an parameter (based on which the scheduling policy is decided) and the file name (to be processed)on the compute node.

At First it stores all the files names in the queue.   
 The following process is applied for each file.

Remove the file from the queue.

Generate a random number and based on the random number assign a compute node to the given file.  
Generate anew thread and call the ComputeMap function of the compute node passing scheduling parameter and the file name .  
if ComputeMap returns false add the file to the queue.

Once all the files are processed(ie queue size=0) call the Computesort function of the compute node.   
The following file contains all the server code

SentimentAnalysisServerHandler.java  
  
  
**Thrift files used for server compute node communication**

The following are the thrift files that are used for server to compute node communication .It contains the following three methods   
bool ping()

The method is used to check the communication between server and compute node

bool ComputeMap(1:string fileName,2:double probability)

this method is used to Compute the map functionality of the map reduce algorithm.

bool Computesort()

This method is used to compute the sort operation of the map reduce algorithm.  
Files used for thrift communication  
SentimentAnalysiscomputeNode.thrift  
  
 **Compute Node**

The compute nodes execute tasks (either map or sort) sent to them by the server. The following functions are part of the compute node.

ComputeMap(String fileName,double probability)  
This function accepts two parameters and compute the Sentiment score for each file and store the data in the intermediate directories .  
based on the second parameter Probability it decides whether to use Random or Load-Balancing Scheduling policy.  
The method is implemented as follows.

If the Scheduling policy is for Load-Balancing  
first it generates a random number and based on that it compares it with load probability it checks whether it will process the files or not. if it rejects it ,it returns false.

- Load injecting  
it delays the computation by some time.it generates a random number and if its value is less than the load probability it delays the computation for 3 seconds.  
  
once it decides to process it the following steps are followed.  
it stores the positive and the negative words in the two hash sets and read the file and line by line and computes the sentiment score and store the result in a intermediate directory.  
  
Computesort()  
This function is use to sort the sentiment score in a particular order and display it.it uses a treemap  
to store the values of the score and corresponding filename in sorted order. Then it store the output file in the output folder.  
Files used for ComputenodeServer

SentimentAnalysiscomputeNodeHandler.java