**COMP-SCI 5570 Architecture of Database Systems RTT Project**

**Work Distribution Report**

**Project Group Name:** Team Elite

**Group Leader Name:** Sri Harsha Chennavajjala

**Group Members’ Names:** Teja Garidepally, Vinay Kumar Matoori, Pallavi Ramineni, Swetha Chandra Karroti, Uday Kiran Chowdary Mallineni, Raj Kiran Reddy Munnangi, Aishwarya Rao Venkata

**Project objective understanding and identification of primary way of working:**

**Involved members:** Everyone

We discussed among ourselves on the project objective and the functionalities that we need to develop.

We identified the different technologies and tools that we’ve to use in order to develop the required functionalities.

We also defined a way of working to interchange the ideas and to share the project status updates daily.

**Working with PlanetLab:**

**Involved members:** Everyone

After getting the PlanetLab account, we started exploring the website. We added the nodes to our slice. We ensured that the chosen nodes are geographically diversified. We chose the nodes such that the distance between the nodes will span from small to large distance. Later we create our public-private key pair and uploaded the public key to PlanetLab to get *ssh* access to the nodes.

Raj Kiran and Teja were monitored the uptime and downtimes of added nodes and added few more nodes to ensure the availability of more nodes at a given time.

**Client-Server application development:**

**Involved members:** Sri Harsha Chennavajjala and Teja Garidepally

We chose socket programming in C language to develop the client-server applications. First we developed the logic to serve a single client at a time. We tested this functionality on some nodes. Later Harsha used fork mechanism to serve multiple clients simultaneously. There by we removed the queueing delay. Harsha created the files with different sizes (32B, 1KB, 256KB, 512KB and 1MB) and a file (nodesList.txt) with hostnames of all the nodes in our slice. Harsha uploaded the client-server programs and other files to all the nodes using *scp* tool.

The main problem that we come across is load on a single server application. As we start all the client programs at a time, each server is being hit by all the clients at a time. This slowing down the file download process. So, in order to overcome this issue, Harsha implemented a logic such that each client will choose a server randomly. This improved the file transfer rates.

**Database maintenance:**

**Involved members:** Sri Harsha Chennavajjala

We used MySQL as our database server. Harsha created the database (IGOD) and all the required tables. Harsha created the *sql* scripts to load the data from flat files into the database.

**Shell scripts to start and stop the client server applications and to download the log files:**

**Involved members:** Sri Harsha Chennavajjala and Raj Kiran Chowdary

We’ve some shell scripts that run on the nodes and others which run on our local system. Harsha developed all the shell scripts to automate the process of starting and stopping the server/client programs on all nodes. Raj Kiran developed the shell scripts to download and merge all the log files.

**Network Tuning:**

**Involved members:** Sri Harsha Chennavajjala and Uday Kiran Chowdary Mallineni

We primarily identified two factors that affect the variance in delay.

1. Data block size that we are using to send each file. We tested the file collection delays with various block sizes and we got optimal results with data block size (Receive Window) of 256KB.
2. Time is affecting the delay variance. So we scheduled the file collection at different times in a day.

**File collection monitoring:**

**Involved members:** Sri Harsha Chennavajjala, Teja Garidepally and Raj Kiran Reddy Munnangi

We’ve constantly observed the file collection process. Some nodes are down for more than one week. Data transfer rates of some nodes are too slow. This causing the delay in downloading the files on other nodes. We identified such nodes and removed them from our list.

**Calculation of Distance Delay Ratio (DDR):**

**Involved members:** Vinay Kumar Matoori and Uday Kiran Chowdary Mallineni

We’ve created a separate view to store the DDR values for each node. Vinay and Uday created the required logic to calculate the DDR values for each node. They also created the logic to calculate the estimate distance using the DDR values and the given delay.

**Data Analysis:**

**Graphical representation of collected data statistics:**

**Involved members:** Sri Harsha Chennavajjala, Teja Garidepally, Raj Kiran Reddy Munnangi, Uday Kiran Chowdary Mallineni and Vinay Kumar Matoori

First, we analyzed the collected data samples and created varieties of sub sets of data to keenly analyze the data. We used Microsoft Excel and MATLAB to draw the graphs. First, we exported all the required data into a csv file and calculated the corresponding cumulative distribution values. Later we imported the data into MATLAB and drew the diagrams.

**Error Correction Techniques:**

**Involved members:** Aishwarya Rao Venkata, Pallavi Ramineni, and Swetha Chandra Karroti

Aishwarya, Pallavi and Swetha were worked on error correction techniques. They analyzed the Linear Regression techniques and Weighted Least Mean Square techniques in order to minimize the error in distance calculation. They used R language to graphically represent the observed and optimized errors.

**Documentation:**

**Involved members:** Everyone

Teja documented all the observations into our report. We’ve maintained a track record of weekly work done. We’ve created a document on our environment setup and way of application workflow.

Environment setup (Linux operating system, and MySQL installation) was done on Harsha’s system. All of our team members were connected to Harsha’s system using putty to perform the required operations.