

# Visualizing the Progression of Real Time Data Using Web Technology

<sup>1</sup>Jannatul Fardous and <sup>2</sup>Shafiqul Islam

<sup>1</sup>Department of Computer science and Engineering, Institute of science &Technology, Dhanmondi, Dhaka, Bangladesh

<sup>2</sup>Department of Electrical Engineering, Institute of science &Technology, Dhanmondi, Dhaka, Bangladesh

Corresponding author email: koliiucja@gmail.com.

## Abstract

*In this digital age, the ability to explore the visual representation of massive data and detect meaningful patterns is very much important. Data visualization is a recent way of visual communication. It is being used to clarify the situation. It uses computer graphical effects to discover the patterns, trends, relationships out of datasets. This dissertation will examine how data visualization can be used for navigating and presenting large datasets. Recent years have seen rapid growth in Big Data methodologies throughout scientific research, business analytics, and online services This descriptive paper first familiar with data visualization and its related concepts, then it discuss about a propose system of Real time data visualization. This paper explores a project about it, and some explanations. The government, corporate, and industrial communities are faced with an increasing number of databases. These databases need not only to be managed, but also to be explored. It requires exploratory tools supporting the identification of elements such as patterns in data. Knowledge Discovery in databases is a relatively new research area that employs a variety of tools to explore and identify structure and patterns in these large databases. The data is mined for specific rules that are built incrementally and often steered by users with a specific set of goals in mind. This dissertation will present novel visualization systems designed for large datasets.*

**Keywords:** Real Time Data, Data visualization, web technology.

## 1. INTRODUCTION

Data visualization is the graphical representation of information. Representing large amounts of information in a visual form often allows us to see patterns that would otherwise difficult to understand, unconnected data sets. Data visualizations [1] [2] shows a way to find hidden patterns and correlations that can lead to important discoveries .Visualization is an increasingly important skill, and data visualizations are another way to develop the ability to process information visually. Visualization is to makes life easy and better .In order to improve information communication efficiency and clarity. Data visualization enables audience to absorb larger amount of data where traditional textual forms of data are slowly and time killing to understand. All in word data visualization is make our human life easy and better.

### 2.1 Data visualization techniques

Some commonly used data visualization techniques are -

- a).Charts: Bar Chart, pie Chart, Gantt chart etc.
- b).Graphs .Plots: Scatter plot, Scatter Plot (3D).
- c).Maps: Tree mapping.
- d).Image.

## 2. RELATED WORKS

In recent years L khufel et al, 2016 has been research into self-awarded Server monitoring system. Also this dissertation describe [3] the architectural concept of this working environment. Commercially

available server monitoring system such as “Hyperic HQ”, “Nagios”, “Zabbix”, SolarWinds” are capable to report large server frames [4][5]. These types of solutions are for huge and massive data collection and perform reporting. Our approach is to automate this analysis by identifying the affected servers.

### **2.1 Proposed systems comparison**

In previous systems they work in offline mode to prevent different attack scenario. In this system it works online based architecture, because individual server will communicate online to the monitoring server. There is no chance to attack whole system. On those systems every machine need to know their current health condition, which is unnecessary and time killing. But in this system it visualized the effected servers. So that anybody can understand easily which server is unhealthy and need to take action. There is no overload on host’s servers, because the full work is distributed to all the servers. As they perform generating health data and send the data to central machine. So system runs smoothly and this source code is very lightweight and working algorithm is simple. It avoids unnecessary loop and condition to make system faster and time efficient. The best thing of this system is that it is free and implementation method is easy. So the summery are-

1. System works online.
2. Visualized only the effected servers.
3. No need to go through every server.
4. All servers work individually. There is no overload on host’s server, because the full work is distributed to all the servers.

## **3. PROBLEM STATEMENT AND RESEARCH QUESTION**

### **3.1 Overall problem statement:**

What is real time data? What is data visualization? Manual work has to reduce.

How we can be benefitted with these processes? System need to be efficient and lightweight.

Proposed solution shouldn’t overload the system.

### **3.2 System side problem statement:**

How to gather all server utilization information? Writing a script to make an xml file containing all these information.

Server to server communication without password interaction. Sending the required information into central server. Make this whole processes simultaneous and automated.

### **3.3 Application side problem statement:**

How we can visualize the xml data into web? Insertion of xml data into MySQL database.

Previous data should be preserved in another data table. Every month’s data will be preserved in a separate table with the name of that month. Before insertion of data program should check whether the same data already exist into current data table. If exist then move those data to the table of that month.

Visualization program must should run at a certain time interval to ensure the visualization of current condition of the server. Data insertion must be continuous processes. This will provide the latest data into the database.

## **4. HYPOTHESIS**

First target of this system to build a website that will show the information in a visual way. Then i start building Figure 4 system website. First i have designed a database that contain all the information then build the pages that will pick the data and show the information. Now the toughest part was to create a program that will do all this job by itself. After completion of the program the system were successfully able to insert the data from dummy xml file. I thought my work is done. But system needs to make this work in real life. System has to test it into real machine. Then I decide to take this data Figure 1 from live machine. Real machine will generate the xml file and send the file to main server. The processing program will processes the data and system will be able to visualize the data of live machine. As of this

thought it deployed two Linux machine and wrote the script to generate xml file. This machine can transfer the file automatically to main server and our program is able to extract the data from the file and it successfully able to implement my thought.

**4.1 Real time data and Web Technology**

Real-time data (RTD) means the data that is being processed immediately after collection [6]. There should not be any delay in the providing the information. Web technology is the mechanisms to different computers to communicate and share resources. Mainly Web technologies mean the network of computers like Local Area Network (LAN), Metropolitan Area Network (MAN) or a Wide Area Network(WAN) [3].

**4.1.1 Applications of real time data**

- Stock market.
- Network management system.
- Aircraft management system.

**5. SERVER MONITORING SYSTEM**

Servers perform the heavy load operations almost all the time. As they are perform heavy duty, sometimes the resources of server get overloaded. At that time they need to do some maintenance work into server. So it need to monitor the server all the time but it is hard and painful job to do this. Because generally system have to login individually each server and perform health check. That takes long time. That will make disaster. When system is checking a healthy server which is unnecessary, at that time an unhealthy server may fall down. So it is not a proper way, all the manual processes are slow and cannot give proper solution. In this situation, that if system could make the servers tell about their health by themselves. This is our research topic to visualize the server health information live and at the real time. That makes the life easy and fast. So system makes decision fast and save our time. Servers perform the heavy load operations. Sometimes the resources of server get overloaded .So system take some actions are

1. Now need to monitor the server all the time.
2. Hard job to perform health check of individually server.
3. Take long time and may not able to keep an eye on all of them.
4. If we could make the servers tell about their health by themselves.
5. Visualize the server health information live and at the real time.

**5.1 Block Diagram**

All the servers in the network will generate their health status by themselves. One script named “health checkup” will create this information. The health information will be in xml format. This xml file then will be sent to central monitoring server at every 10 munities interval Figure 1. One script named “send xml” will do this job. Then the web application of central monitoring server will process these xml data and save them into database. At last from that database information will be visualized.



Figure 1 Block diagram of overall processes of this system

Here **Table 1** show the sample of data table of server and **table 2** shows visualize data table of those server.

**Table 1.A simple sample data table.**

Host	IP	Process	Process ID	CPU Usages
host2	192.100.100.90	process1	1203	45%
host3	192.100.100.95	process3	1203	40%
host3	192.100.100.95	process1	1203	96%
host1	192.100.100.85	process1	1203	50%
host4	192.100.100.96	process3	1203	50%
host1	192.100.100.85	process3	1203	85%
host4	192.100.100.96	process1	1203	79%
host4	192.100.100.96	process2	1203	90%
host3	192.100.100.95	process2	1203	78%
host2	192.100.100.90	process2	1203	96%
host2	192.100.100.90	process3	1203	77%
host1	192.100.100.85	process2	1203	75%

**Table 2 After Visualize data table.**

Host	IP	Process	Process ID	CPU Usages
host3	192.100.100.95	process1	1203	96%
host2	192.100.100.90	process2	1203	96%
host4	192.100.100.96	process2	1203	90%
host1	192.100.100.85	process3	1203	85%
host4	192.100.100.96	process1	1203	79%
host3	192.100.100.95	process2	1203	78%
host2	192.100.100.90	process3	1203	77%
host1	192.100.100.85	process2	1203	75%
host1	192.100.100.85	process1	1203	50%
host4	192.100.100.96	process3	1203	50%
host2	192.100.100.90	process1	1203	45%
host3	192.100.100.95	process3	1203	40%

### 5.1.1 Design and development

All the servers in the network will generate their health status by themselves. The health information will be in xml format. This xml file then will be sent to central monitoring server at every 10 munitities interval. Application will process these xml data and save them into database. From that database information will be visualized [7]

### 5.1.2 System Description

**System admin** will write a script to create the xml file and automate the xml generation processes. He will also ensure the automated machine to machine communication.

**Application developer** will develop the application to processes the xml file and visualize the information.

In the **client machine** this xml generation script will run every certain time interval. This machine will send this xml to monitoring server every 10 munitities interval.

**Central monitoring server** will processes the xml data file with the help of application and save them into database. From database it will then visualize the CPU, RAM and HDD information.

**Server owner, Owner of the hosted application, Owner of the hosted database And even server admin** can see the current condition of all servers in the network. Actually anybody can see the scenario if they are authorized to see.

## 6. RESULT AND ANALYSIS

There are four basic or fundamental things to do for implementing this project. They are:

1. XML generation into each and every server at a certain time interval
2. Machine to machine communication.
3. Process that xml data into the monitoring server.
4. Visualize that processed data.

### 6.1 Creating XML description

XML is a markup language much like HTML. XML was designed to describe data, not to display data . The tags are not predefined. We can define our own tags. XML is both human-readable and machine readable .There should be a starting and end tag for every element.

- We will collect CPU, RAM, and HDD utilization at a certain time interval.
- One script will do this job.
- XML file name will be the server unique IP address.
- Script will be called in every 10 minutes.

### 6.2 Gather utilization information

In Linux system we can gather the utilization by some commands. Following command will take the output of HDD utilization. Here is only one drive information will be collected. If we want we can add more command with drive to get their information. "df -h | grep drive". This command will just take the output. We need to dump it into a xml file. In the script we will write the code to dump that information into the xml file. The code of dumping into xml is:

```
<mount>
  <name>"$DRIVE_C_BOOT"</name>
<size>"$DRIVE_C_BOOT_PERC"</size>
<datetime>"$now"</datetime>
</mount>
```

The same processes go for CPU usages also. The code for that part is:

```
"top | head -12 | awk '{print $10}'"
<process>
<id>"${process_id[9]}"</id>
<name>"${process_name[9]}"</name>
<usages>"${process_pers[9]}"</usages>
<datetime>"$now"</datetime>
</process>
```

The same processes go for MEMORY usages also. The code for that part is:

```
"vmstat -s | head -10 | awk '{print $1}'"
<memory>
<Physical_memory>"${memory[0]}"</Physical_memory>
<virtual_memory>"${memory[7]}"</virtual_memory>
<free_memory>"${memory[4]}"</free_memory>
<used_memory>"${memory[1]}"</used_memory>
<datetime>"$now"</datetime>
</memory> .All these are for sample command. If we want to increase we can add more.
```

#### 6.2.1 Sending the XML file (Machine to Machine communication).

After generation of xml file into the machine now it's time to send the file to the monitoring server. To do this first it need to ensure that one machine can enter into another machine without authentication. Because when any machine wants to login another every time it asks for the password. And at that time human interaction will be necessary. But it want the system automated as it want to send the xml file every ten minutes interval. So first it need to make a password less communication. Now it will tell the machine what to do. First it will create a script named "send\_xml" into the location "/usr/local/bin". The content of the script will be:"rsync -azv -e ssh /home/jannat/ 192.168.56.3:/var/www/html/web/data "The command "rsync" means synchronize the data between the client machine and monitoring server. Where the client machines data location is "/home/jannat/" and monitoring server machines data location is"192.168.56.3:/var/www/html/web/data". The system has to make this script executable to run the script."chmod +x /usr/local/bin/send\_xml".Now it have to make the system automated. The system has to send the xml file every 10 minutes interval. There is a file in Linux system called "crontab" which is located to"/etc/crontab". We have to understand the architecture and mechanism .of that file. The above

picture is showing the format of crontab [7]. 1st star indicates the minutes of every hour. If system want to run any script in a certain minute of every hour it will set the minute into that place. now the system are going to use this star. 2nd star indicates the hour of every day. If system want to run any script in a certain hour of every day it will set the hour into that place 3rd star indicates the day of month. If system want to run any script in a certain day of every month it will set the day into that place. 4th star indicates the month of year. If it want to run any script in a certain month of every year it will set the month into that place. 5th star indicates the day of week[8].



Figure 3 Crontab structure.(command to be executed)

If system want to run any script in a certain day of every week it will set the month into that place. The content of system “crontab” file will be below. The script “/usr/local/bin/send\_xml” will run at every 10 minutes and synchronize Figure 3.the xml file.

```
00 * * * * root /usr/local/bin/send_xml 10 * * * * root /usr/local/bin/send_xml
```

### 6.2.2 Processes the XML file

By the script xml file has been reached to the monitoring server. Now system need to processes the xml data and save the data into database. There are three types of information is there. CPU utilization information, RAM utilization information, HDD utilization information. There are some conditions to save these information into database. The conditions are: First check whether if previous data exist in the database table. If exist then move it to another table and then insert the data. So that it doesn't create conflict with latest data. Previous data should be kept in monthly basis table. The algorithm of this program is below:

Take the IP of the server from xml file name. Check into the data table whether previous information exists .If exist then create a new table with same structure and name with the system date and time. Insert existing data to that table.

For every month there will be a new table. Each table contains one month's data .After insertion then deletes the data from current table.

Now it will insert the latest information into our latest table. Else doesn't exist previous data then just simply we will insert the information into current table.

## 7. VISUALIZE THE DATA

Monitoring server will processes these xml data. From these xml data CPU utilization information will be saved in CPU table in Database. Memory of RAM information will be saved in RAM table in Database. Hard drive information will be saved in HDD table in Database. From the database these information will be visualized.

### 7.1 CPU utilization

In this table we will get the info of server name, ip, process, and process id and processes utilization. If the processes are utilizing the CPU more than 80% then it will be red marked Figure 4. They are the prime concerned we should take care. If the Processes are utilizing the CPU more than 70% then it will be

yellow marked. They are secondary concern we should take care. Below 70% they will be green and they are safe. This window refreshes every minute and keeps the information updated.



Figure 4 Result of system review of CPU utilization.

### 7.1.2 RAM utilization

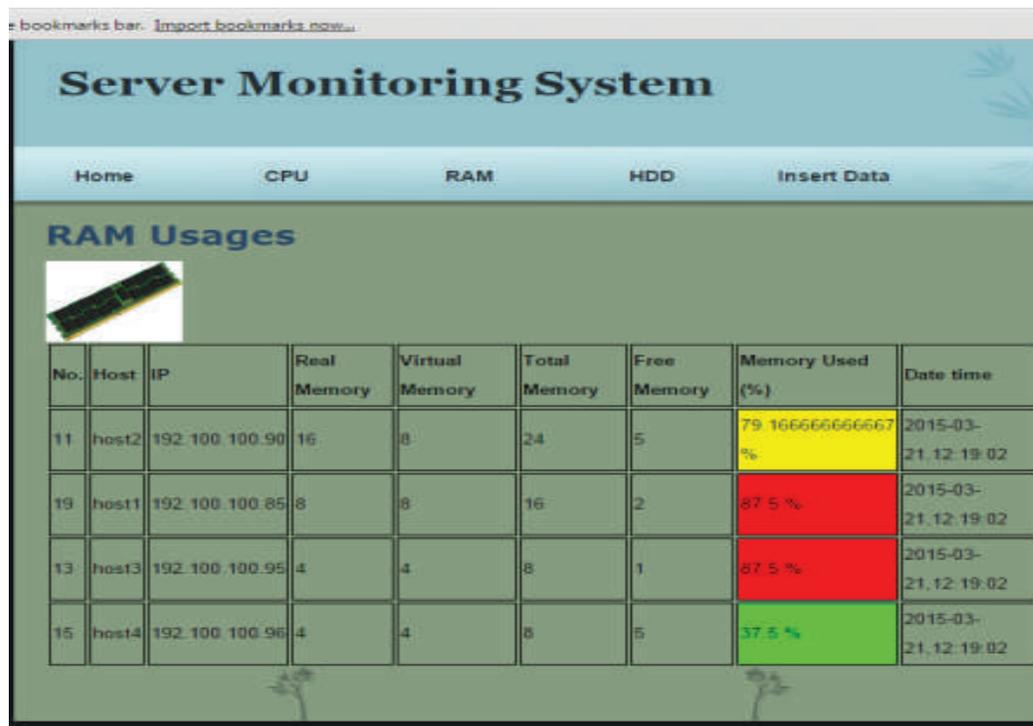


Figure 5 Results of system review of RAM utilization

If the RAM are utilizing more than 70% then it will be yellow marked. They are secondary concern we should take care. Below 70% they will be green and they are safe. This window refreshes every munities and keeps the information updated.

**7.1.3 Result of system monitoring HDD utilization**

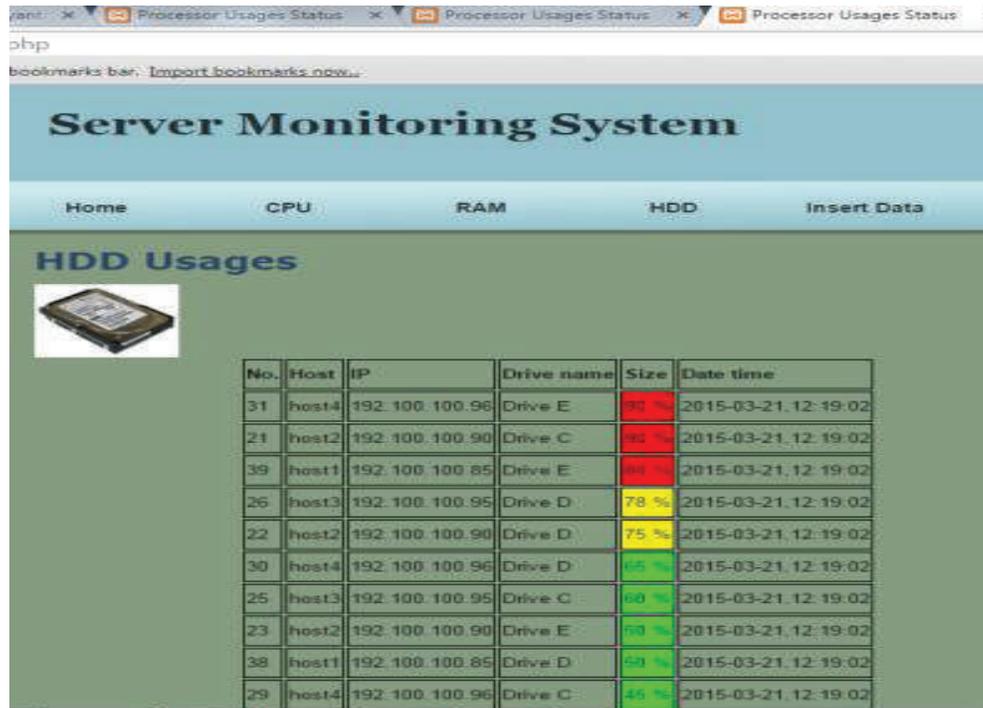


Figure 6 Result of system review of HDD utilization

In the figure 6 we see, we will get the info of server name, IP, Drive name, Drive usages. If the drives are being utilized more than 80% then it will be red marked Figure 6. They are the prime concerned we should take care. If the drives are being utilized more than 70% then it will be yellow marked. They are secondary concern we should take care. Below 70% they will be green and they are safe. This window refreshes every munities and keeps the information updated.

**8. TESTING AND EVALUATION**

After Visualization now system need to be tested. We are visualizing the data through color coding. We have set a certain level of threshold to take quick decision. From (80% to 100%) table cell color will be red and it is very much alarming. At the bellow level the threshold is (75% to 79%). The table cell color will be yellow and it is alarming. Because after sometimes it may reach to danger threshold. And below 75%is taken as a green color and they are safe. This percentage could be customized. We will do these threshold testing and besides we also do some system level and program level testing. We are using black box testing method for our testing purpose.

**Table 3 Test and evaluation.(monitoring approach)**

Test	No .of purpose of the test	Test (input)	Expected result of the test	Actual result of the test	Test corrective action
1	Check if Cpu utilization value is more than80%, then	95, 85, 90	The color of the table cell should be Red	Table cell color is red.	None taken.

	what the color of cell is.				
2	Check if Cpu utilization value is more than 75%, then what the color of cell is.	79, 75, 77	The color of the table cell should be yellow	Table cell color is Yellow.	None Taken
3	Check if Cpu utilization value is less than or equal 74%, then What the color of cell is.	65, 50, 40	The color of the table cell should be Green.	Table cell color is Green.	None Taken
4	Check if Memory utilization value is more than 80%, then what the color of cell is.	99, 84, 91	The color of the table cell should be Red	Table cell color is Red.	None Taken
5	Check if Memory utilization value is more than 75%, then what the color of cell is.	78, 73, 74	The color of the table cell should be Yellow.	Table cell color is Yellow.	None Taken
6	Check if Memory utilization value is less than or equal 74%, then what the color of Cell is.	20, 15, 45	The color of the table cell should be Green.	Table cell color is Green.	None Taken
	Check if Hard disk utilization value is more than 80%, then what the color of Cell is.	98, 85, 92	The color of the table cell should be Red	Table cell color is Red.	None Taken
8	Check if Hard disk utilization value is more than 75%, then what the color of Cell is.	77, 72, 76	The color of the table cell should be yellow	Table cell color is Yellow.	None taken
9	Check if Harddisk utilization Value is less than or equal 74%, then what the color of cell is.	10, 15, 25	The color of the table cell should be Green.	Table cell color is Green.	None Taken.
10	To check whether it create Xml file in the client machine.	Run the command "./script	It Should create xml file into /home/jannat/xml Location.	It creates xml file into home/jannat/xml Location.	None taken.
11	To check whether client machine can connect to the server machine.	Run the command "ssh192.168.56.101"	It should connect to the server machine.	But it didn't connect to the server machine	Need to copy ssh id to the server machine.
12	To check the transformation Of xml file to the server.	Run the command "rsync 192.168.56.102.xml"	It should send the xml file to server Specific location.	It did transfer the Xml file to the right location.	None taken

13	Checking the server data already exist inThe table.	Keep same data.	It should move the previous data into a separate table.	previous data moved into another table	None Taken.
14	Checking the server data Already exist in the table.	Keep same data.	It should move the Previous data into a separate table.	previous data moved into another table	None taken.
15	Keep the data into monthly Basis table.	Change data time of Server system.	Program should take current date and time and create a new table.	A new table was Created according to current date and time.	None taken.

### 9. CONCLUSION AND FUTURE SCOPE

Visualization is great and makes the work fast and easy. This system is applicable to any kind of monitoring system. Where data is available there visualization can also be done. The thing is how user wants to see it. System can set their customized threshold and make it workable. There are some servers those can highly impact business revenues if they fall down. So they need to maintain with higher priority. For these servers this system is not enough. Because if anybody want to monitor now system have to sit in front of computer monitor. In this situation my future plan is to make this system more dynamic. If these servers are impacted then the system will automatically send email to administrator and also give messages to mobile. So that he/she can take immediate action about this. Thus it will bring a full proof and automatic monitoring system. Some limitations of the proposed systems are a). System has tested this method in UNIX operating system b) systems are not tested in windows operating system.

### ACKNOWLEDGEMENT

First I wish to express my sincere thanks to Md. Mahmudul Hasan, Senior Lecturer, Department of CSE Daffodil International University, Dhaka. His endless patience, scholarly guidance, continual encouragement valuable advice, help me to complete this work. I would like to thank my friend Md .Illius for contributing the idea of real time data visualization also. I am extremely thankful number of course mate and colleagues in encouraging me to start the work and publish it.

### REFERENCES

- [1]P. Isenberg, T. Isenberg, T. Hesselmann, Bongshin Lee, U. von Zadow and A. Tang, "Data Visualization on Interactive Surfaces: A Research Agenda", *IEEE Computer Graphics and Applications*, vol. 33, no. 2, pp. 16-24, 2013.
- [2]B. Shneiderman, C. Plaisant and B. Hesse, "Improving Healthcare with Interactive Visualization", *Computer*, vol. 46, no. 5, pp. 58-66, 2013.
- [3] Ł. Kufel, "Tools for Distributed Systems Monitoring", *Foundations of Computing and Decision Sciences*, vol. 41, no. 4, pp. 237-260, 2016.
- [4] [Online]. Available: <http://www.monitis.com/blog/2011/02/22/11-top-server-management-monitoring-software/> (18 October 2015, 6:00 PM. [Accessed: 16- Oct- 2017].
- [5]"Permanant Automatic ssh login without password - FedoraForum.org", *Forums.fedoraforum.org*, 2017. [Online]. Available: <http://forums.fedoraforum.org/showthread.php?t=274268> (03 December 2015, 6:00 PM). [Accessed: 16- Oct- 2017].
- [6]Z. Liu, B. Jiang and J. Heer, "imMens: Real-time Visual Querying of Big Data", *Computer Graphics Forum*, vol. 32, no. 34, pp. 421-430, 2013.
- [7]. "Data visualization", *School of Computing* ,National ,University of Singapore
- [8]. [Online]. Available: <http://www.adminschoice.com/crontab-quick-reference> (14 December 2015, 8:00 PM). [Accessed: 16- Oct- 2017].