ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 3, Issue 5, May 2016

Data Visualization: A Handy Plug-In

[1]Ms. Lavanya Patil [2] Dr. Jagdeesh D. Pujari (Guide)
[1] M.Tech in Information Technology, [2] Head of Department
[1][2] Department of Information Science and Engineering,
SDM College of Engineering and Technology, Dharwad, India

Abstract: **---- The amount of data that is being generated in today's world is increasing rapidly. Communicating this data to the world is essential. In doing so, there is no wrong way. But the traditional way of communicating data through texts is mediocritic and requires more attention to details to understand the data. Data visualization does not belong to a particular field and is a part of every field and only through visualization one can enunciate the results in a better way.**

Keywords: **Big data, Centrifuge, Plug-in, Transformation principle**

I. INTRODUCTION

Data is being generated everyday in multiples of it being generated the previous day. Such a huge data needs to be processed properly and analyzed to make better decisions. The data cannot be thrown away or kept un-utilised. It provides better insight to the business giants and techies in taking decisions for the companies. Hence it is very essential to formulate such huge data to a specific format. The data could be generated from varies sources such as weather station, sensor data, data from banks, log analyses, stock market etc.

Example, consider the data from weather stations. The first question that arises is how do we consider this data as huge? The answer is, assume there are 10 weather stations in Delhi. The data coming from all the weather stations will contain the weather station ID, Humidity value, Day, Date etc. So data from all the stations makes big data to be collected. This data should and must be understood, analysed and interpreted well to make better decisions as shown in the figure 1.1.Below we can see the temperature and humidity for the months of January, February, March and April. From the image we can easily interpret that the temperature is maximum for the month of April. This is the data on the average calculations from all the weather stations

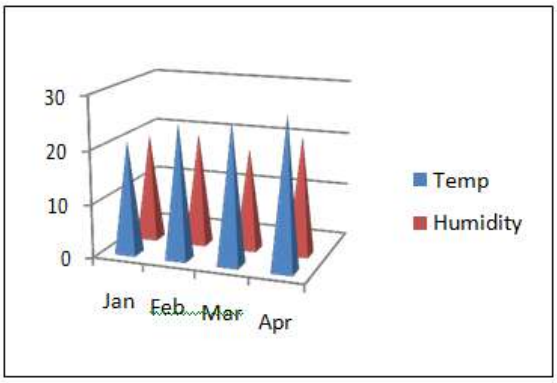

Fig 1.1: Weather station data

Visualization is nothing but presentation of data is one of the available ways. Even in visualization there are many algorithms and tools available. Some of the algorithms that provide visual display are merge sort, segment tree quick sort as shown in the images below:

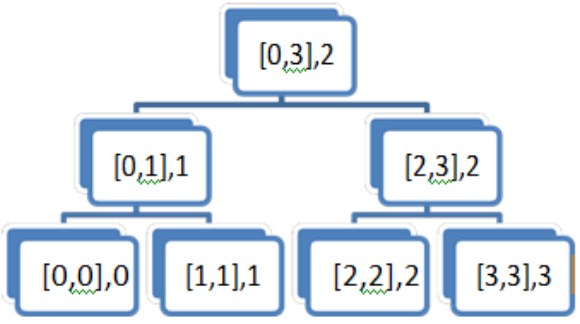

Fig 1.2: Segment tree algorithm visualization

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 3, Issue 5, May 2016

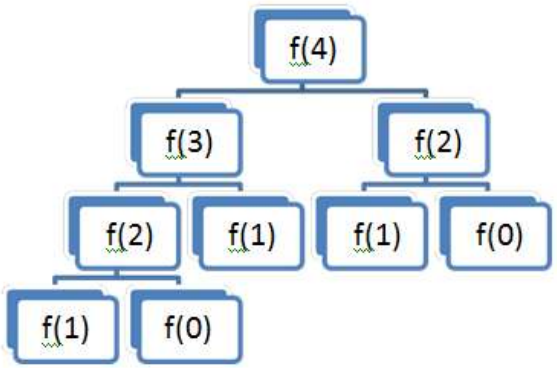

Fig 1.3: Visualization of recursion tree/DAG

And every data must satisfy the three V's that define big data i.e., Volume, Velocity and Variety.

- Volume: This gives the measure of the quantity of data. If the data is more than 1TB then it is said to be big. Similarly any data that does not fit into one's system is known as big data relatively. The data stored at APMC's or the data at the warehouse of a company can be considered to be big data.
- Velocity: The rate at which data is being generated and the rate at which it is being processed define if the data is big data.
- Variety: The rate, at which data is increasing, is emerging with different types of data. Variety implies to the data set nature, its complexity and applications. There is data of structured, unstructured and semi-structured type. And all these types have different characteristics.

II. IMPLEMENTATION OVERVIEW

There are many algorithms in today's world to visualize data and so are many approaches. To name a few:

- Network analysis: In the early days the data would be small but when internet happened the data grew and handling this data on the net was tedious yet important. This also happened as the number of sensors increased a long side. Hence keeping a track of the information flow among these sensors became important. There are various tools to analyze network flows like: *Centrifuge* which gives a suite of capabilities to the investigators to analyse data. Other one is *EgoNet* which is a program for collection and analysis of egocentric network data.
- Optimization: There are many classic optimization algorithms that allow huge data to be reduced to data of a size that can be analysed easily.
- Cluster analysis: As data increased it was necessary to group the data into smaller and similar groups.
- Along with the approaches there are many visualization tools available such as:

- Datawrapper: This tool was designed for the news institutions by a journalist. It basically allowed the journalists which happen to be non-technical people to create a chart on the data they collect in just a few steps.
- Chart JS: This is a chart library and allows the users to include the API in their code. After this the user can run the code and get the visualization. This tool is good but if one does not want to get technical with coding, this tool will not be helpful.
- D3: D3 stands for data driven documents.vD3 provides APIs that can be applied to DOM elements and transform the resulting HTML, SVG, or CSS documents. But again, this method may appeal to programmer's more than average users because it involves writing code to creating graphs.
- After knowing all these, we need to understand the basic principle of all the visualization algorithms. If we consider all the algorithms like merge sort or quick sort, all these reduce the huge data sets into smaller datasets and then generate results as shown in fig 2.1

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 3, Issue 5, May 2016

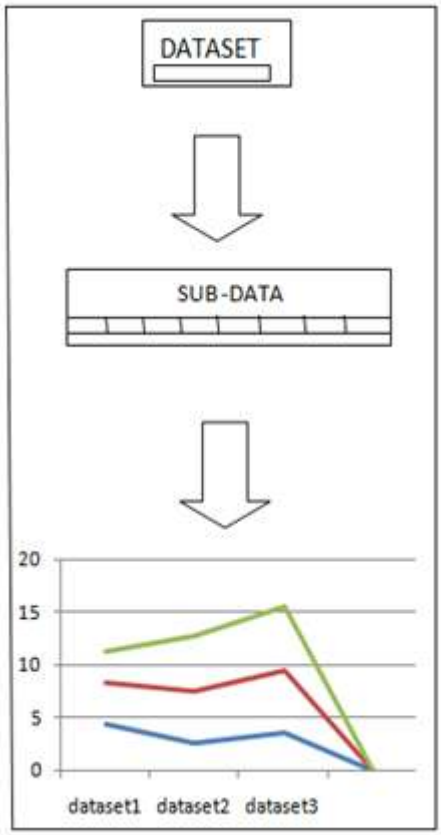

Fig 2.1: Transformation Principle

Hence the basic idea of transformation remains the same for each and every algorithm and the tool of mapping the data sets and then reducing them to the required format for further utilization.

2.1 Need for plug-in:

PLUG-IN are addition software's that allow the user system to generate results on a particular type of content without installing additional requirements. Hence in this paper we understand how we use visualization as a PLUG-IN and analyse data in different fields such as agriculture

2.2 Proposed system

Using the transformation principle, the plug-in that is being implemented works as shown below in the fig 2.2.

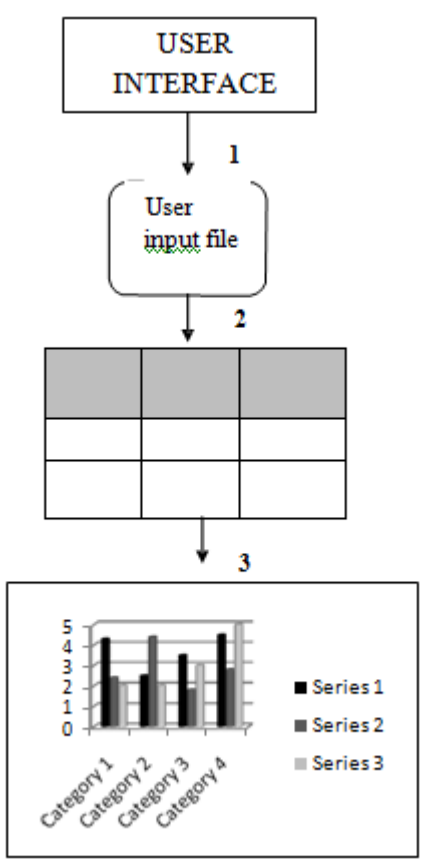

Fig 2.2: Visualization Plug-In

The visualization plug-in that we are implementing occurs in three steps:

Step 1: We provide a user interface; where in the user can provide a file which contains the data in the csv format. Later as a, the data from this file is processed and reduced into smaller data set.

Step 2: In this step, the reduced data set is stored into a csv file. This file acts as Meta data storage file.

Step 3: The data from the Meta data stored file is processed and results are generated. The results can be bar charts, pie charts and others such as time series.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 3, Issue 5, May 2016

2.3 Results

The images below show the results by using visualization as a PLUG-IN.

Fig 2.3: User file input to the PLUG-IN

In the fig 2.3, the user is requested to provide the file that contains the required data. The format of the user file should be .csv extension. The data from the user's file is given as an input to the visualization PLUG-IN and then reduced by using the filters. Later the filtered data is chosen to display the results.

In the fig 2.4, the results of the users data is displayed in the form of a bar chart.

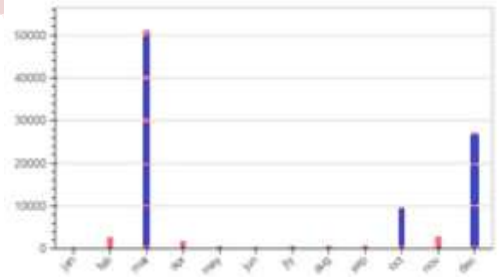

Fig 2.4: Bar chart for the user files data

III. GAPS IN STUDY

Business's today are growing at an ever-increasing rate. They need an efficient, easy deployable and handy data visualization tool.

Tableau8 which provided with better visualization as it allowed the web incorporation but still did lack in its speed, free form of calculation and data analytics. This was covered by Tableau9. This tool also provided with better colours for the background to differentiate the results.

Tableau9 is the recent visualization tool which came in

2015. This tool works on the driver technology. Hence every system would require the mandatory driver installation which is time consuming. This would reduce the time required in installing the software. This tool will reduce the complexity of using the visualization tool. It enables the users to use files of different data format in one tool which does not exist in the current available tools. These tools require the user to convert their files into the file format which the tool accepts and later would provide visualization. But now that extra work can be eliminated.

IV. VISUALIZATION SCOPE

i. H e a l t h care: The visualization data used in health care is generally called health care informatics.

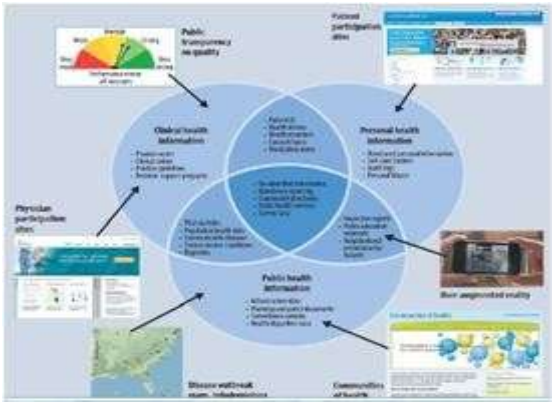

Fig 4.1: Patient details through visualization (Src: Improving health and healthcare with interactive visualization methods, IEEE, 2013)

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 3, Issue 5, May 2016

This has improved the health care revolutionarily. But the only drawback of using visualization in health care is that, not only using visualization is important but also to read the reports properly must be taught to the users

ii. Financial: Financial data can be termed as large and multi-dimensional. Due to this giving the user an up-to-date information access is essential. Financial data and visualization in work on the principle "See it, understand it, and pursue it".

iii. Communication: Providing visualization in the field of communication is a crucial task. In doing so, only the maximum and minimum value frequencies are marked on the chart and according to those values a pictorial representation is obtained.

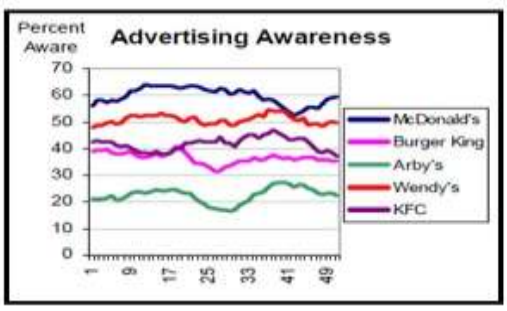

Fig 4.2: Communication with visualization (SRC: Visualization for Communication: The Importance of Aesthetic Sizzle" IEEE IV05, 2005)

iv. Transportation: The recent development in technology has made the interpretation of traffic and transportation easy. Denver Council of Government's (DRCOG) Transportation Improvement Program (TIP) uses various forms of charts, graphs to indicate the traffic flow.

Fig 4.3: Navigation through visualization. (Src: Guide for Transportation Agencies, 2003)

v. Security: Visualization is used in computer security and in other forms also. There are various applications of visualization in computer security such as correlating intrusion detection, forensic analysis, network defence analysis etc.

V. CONCLUSION

This paper covers the various concepts of big data such as the 3V's that describe the data. It also covers the approaches of visualization and the tools used. Even though there are numerous visualization tools available in the market, the basic principle of transformation is not hampered. Then we finally understand the need for a plug-in.

REFERENCES

[1] Urvashi Pathania and Aman Singh, "Visualization Tools of Data Structures Algorithms", IJARCSSE,2014

[2] Brian K. Sheffler, "The Design and Theory of Data Visualization Tools and Techniques", SANS Institute.

[3] Michelle A. Borkin, Chelsea S. Yeh, Madelaine Boyd, Peter Macko, Krzysztof Z. Gajos, Margo Seltzer, Hanspeter Pfister, "Evaluation of Filesystem Provenance Visualization Tools", IEEE,2013.

[4] Nishith Sharma (co-founder frrole) social intelligence, "The 14 best data visualization tools", 2015.

[5] Xiaoyan Bai, David White, David Sundaram titled — "Context Adaptive Visualization for Effective Business Intelligence", Proceedings of ICCT2013.

[6] John R. Goodall, "Introduction to Visualization for Computer Security".

[7] AASHTO Task Force on Environmental Design prepared a document, "Guide for Transportation Agencies", 2003.

[8] Richard Brath, Mike Peters, Robin Senior, "Visualization for Communication: The Importance of Aesthetic Sizzle" IEEE IV05,2005

[9] Yingjian Qi, Xinyan Yu, Guoliang Shi, Ying Li, titled ―Visualization in Media Big Data Analysis, at ICIS, June 2015, USA.

[10] Stephen Few, ―Data visualization past, present, and future, Perceptual Edge, 2007.

[11] W.Schroeder, K. Martin and B. Lorensen, The Visualization Toolkit, 2nd Edition. Pretice Hall PTR, 1998, ISBN 0-13-954694-4.

[12] Ben Shneiderman, Catherine Plaisant, and Bradford W. Hesse, "Improving health and healthcare with interactive visualization methods", IEEE Computer Special Issue on Challenges in Information Visualization, 2013.