**Visualization of Big Data**

**Dr. Deepti Gaur, PhD. Ritu Taneja Associate Professor (CSE & IT) M.tech Computer Science Institute of Technology And Management Institute of Technology And Management University , Gurgaon University,Gurgaon** [**deeptigaur@itmindia.edu**](mailto:deeptigaur@itmindia.edu) **ritu.cse.91@gmail.com**

**ABSTRACT**

Big Data as the name signifies large amount of data which is increasing day by day exponentially. To process this data and manage it, lots of time and memory is needed. With the increase in memory, several issues regarding processing speed and space concerns makes it difficult to manage it along with its presentation to make it easy to grasp things. A technique to communicate via images or diagram to present large amount of data or knowledge is termed as Visualization. This paper describes us about the different visualization methods and some of the tools associated. Visualization is essentially the way of representing Big Data in brevity such that it will be easy to learn and understand it by human in short span of time. We will be discussing about different data visual analytics. After performing visualization on data, it will be easily understood by human for learning purpose as well as understanding it to use for further work in business. Thus helps in growth and development of business as well as capital of country.

**GENERAL TERMS**

Big data; Visualization; Application of Visualization

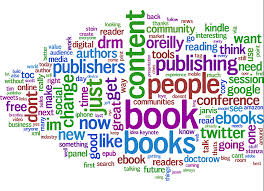
**INTRODUCTION**

Data from any area of search like from corporate world , web data, industries, or any other organization has rich amount of knowledge and information regarding its employees, customers ,products, policies, monthly turn over and much more. As the data is becoming vaster it has a problem of storing and data warehouse still there. Taking out value from Big Data is a big concern needs best output with less cost incurred.

It’s importance is getting so much attention in today’s world as a latest and trending topic in just few years. Various top companies like IBM, HP have not only started working on Big Data but also made successful growth in this field. Many Eyes is one of the examples of IBM project relating to Big Data.

As said correctly by IBM “a picture is worth millions of cells”.

In this real world of Big Data, it has become a great need to visualize the data using which thousand lines of data can be concise in the form of pictures like graphs, charts, tables etc. There are various tools that can be used to visualize the data but some are commercial and some are for public use. Commercial visualization tools are paid and provided to only those who are authorized to use it by filling their particulars on respective sites before downloading it and paying the amount related to it. And for public tools with some trial versions available, there is a need to register for it to use and then download it to run and apply visual effects on it.



**Figure 1 Visualizing Large Amount Of Data**

The two main purpose of data visualization are Data Exploration and Communicating Information.

1. To discover unpredicted relationships involving different data sets of huge databases is known as Data Exploration
2. After discovering a relationship visualization can be used to communicate that relation to others.

### Information is beautiful. The use of visual representations to relocate knowledge between at least two persons by using [computer](http://en.wikipedia.org/wiki/Computer) and manually-generated graphic representations through [sketches](http://en.wikipedia.org/wiki/Sketch_(drawing)), [diagrams](http://en.wikipedia.org/wiki/Diagram), [images](http://en.wikipedia.org/wiki/Image), [objects](http://en.wikipedia.org/wiki/Object_(image_processing)), interactive visualizations, information visualization applications, and imaginary visualizations as in [stories](http://en.wikipedia.org/wiki/Narrative) are of great concern. The process to identify, analyze, store, and retrieve massive quantities of data is central to successful visualization.

In this paper we will discuss the 5 applications of visualization and the tools that can be used for each type. We will also visualize data using different tools and learn how to deal with various data formats.

**II APPLICATIONS OF VISUALIZATION**

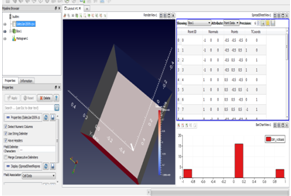
In this section we will discuss different applications of visualization:-

1. **Scientific Visualization**

* Researchers find it easy to figure out trends and understand the characteristics of large amount of datasets from their research on it. Being in the world of high technology it has also lead to visualize 3D along with 2D data sets.
* Scientific visualization has various applications, natural sciences, mathematics, formal sciences, geography and ecology.
* Various tools that falls into it are paraview, Interactive data language (IDL), OpenGL.

Figure 2 depicts visualization of a box using Paraview in which different values been assigned to the box with respect to orientation,axis and different filters been applied Data Analysis, Material Analysis, Quadrature points and much more.

It shows 3D graphics of a box in which a box with different dimensions along with some transitions applied on it . Its data is shown in right and below data analysis is done using histogram.



**Figure 2 Visualization of 3D box using ParaView**

### Educational Visualization

### Visualization for educational purpose can be done by breaking down information in palatable portions and provide context around specific themes, thus serving readers value key concepts more rapidly.

### User can upload their own data for that it needs to be uploaded on the tool that is required with the help of which data model can be prepared in the form of table. This data is only the metadata consisting of core points of the data uploaded. Finally visualization is done on the basis of that metadata.

### The various tools for educational visualization can be Many Eyes, InfoGram, creately, Tag Crowd and many more.

### Most tools are quick in use, fast to share, and easy on eye along with drag and drop facility

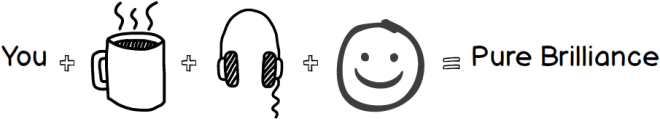
### TagCrowd is a network application for visualizing word frequencies in any content by creating word cloud, text cloud or tag cloud.



### Figure 3 TagCrowd to represent visualization

### Infographics Visualization

* A study of visual representation of numeric or non numeric abstract data for better understanding and analyzing of data to human.
* It is basically a research area for data without any correspondences such as symbolic, tabular, networked, or textual information sources through the visualization techniques.
* Various tools that fall into it are Visual.ly, Prefuse, D3, Timeline.js, Balsamiq.



**Figure 4 Balsamiq to represent data with graphics**

1. **Software Visualization**

### Software though being invisible and intangible, is kept when into files can only be used due to its unknown physical size or shape. Thus there is a need to understand software structures.

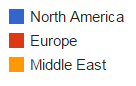
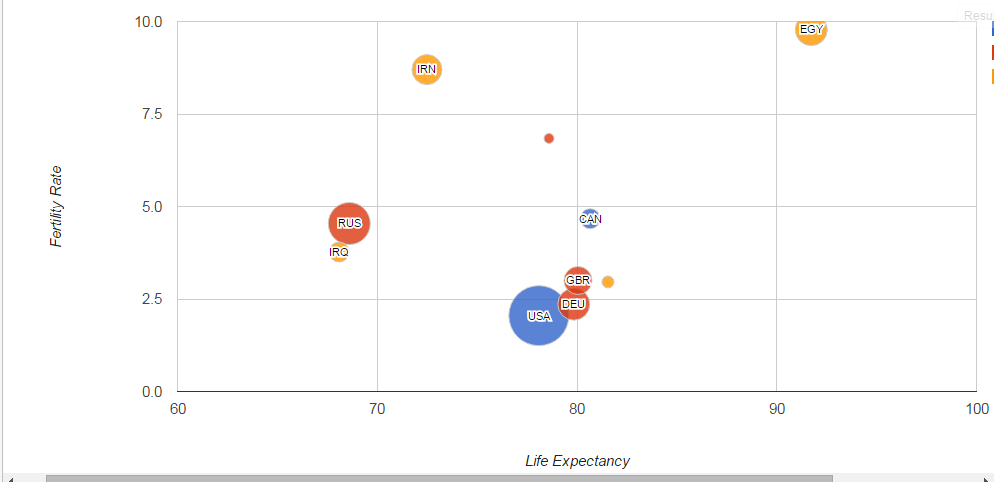
### It can be used as tool to explore and analyze software system information through the architecture of its source code or its run-time behavior and software structures using its visualization in 2D or 3D.

### Software structure can be described using Directed graphs, Run time behavior based on program or traces can reveal bugs and performance anomalies or by using graphical representation of data structures and source code can be visualized using printers only.

### Various tools that falls into it are Google chart tools, GraphViz, Flare

### Some of these tools like Google Chart tools use HTML5/SVG codes to develop charts of different types like line, area, bubble, calendar charts and many more.

### Figure 5 depicts visualization using Google Chart tool showing correlation between different countries with respect to life expectancy, fertility rate of some countries in year 2010 by coding it online and getting its result in the form of chart.



**Figure 5 Correlation between Life Expectancy, Fertility Rate of some countries (2010) using Google Charts.**

**Table 1. Description of various tools of Visualization**

|  |  |
| --- | --- |
| **Tools** | **Description** |
| **ParaView** | 1. ParaView is an open-source for making quick visualization results to examine the data using qualitative and quantitative techniques.   ParaView supports distributed memory computational models to process large data sets ranging from supercomputers to laptops.   1. It runs on distributed and shared memory, parallel as well as single processor systems and has been successfully tested on different operating system. |
| **QlikView** | 1. Provides the right information available and easily accessible user interfaces that are clean, simple, and straightforward. 2. Helps users make time-saving and accurate decisions. |
| **TagCrowd** | 1. It’s a web application for visualizing word frequencies in any text by creating a word cloud, text cloud or tag cloud. 2. For [visual analysis of qualitative data](http://onlinelibrary.wiley.com/doi/10.1111/j.1475-4762.2010.00952.x/full). 3. Acts as blog tool for [search engine optimization (SEO)](http://www.associatedcontent.com/article/2641221/seo_tips_using_tagcrowd_as_an_seo_tool.html). |
| **Cytoscape** | 1. It imports a network file in a variety of file formats: SIF, NNF, GML,XGMML, SBML, BioPAX, PSI-MI Level 1 and 2.5, Delimited text and Excel Workbook. 2. To analyze public network data and to find interactions of set of genes. |
| **Circos** | 1. Circos visualizes data in a [circular layout](http://circos.ca/intro/circular_approach). 2. It is designed for visualizing [genomics](http://blogs.nature.com/ofschemesandmemes/2014/02/20/beautiful-science-picturing-data-inspiring-insight-at-the-british-library) to [mathematical art](http://www.theguardian.com/science/alexs-adventures-in-numberland/gallery/2014/mar/14/pi-day-pi-transformed-into-incredible-art-in-pictures). |
| **Microsoft Power BI** | 1. To make it flexible approach ,it has drag and drop feature. 2. Visual analytics can be prepared at fingertips with perceptive report authoring |
| **Tableau** | 1. Creates amazing interactive visuals and publish them quickly, without the help of programmers or IT. 2. It has drag and drop feature you can easily create stories or reports using just mouse and a little imagination. |
| **Google Charts** | 1. To make [the](https://developers.google.com/chart/interactive/docs/customizing_charts) website look presentable charts are customized. Highly interactive and expose [events](https://developers.google.com/chart/interactive/docs/events) to create complex [dashboards](https://developers.google.com/chart/interactive/docs/gallery/controls) or other experiences [integrated with webpage](https://developers.google.com/chart/interactive/docs/examples#full_html_page_example). 2. Provides cross browser compatibility and platform independency charts are turned in using HTML5/SVG technology to iPhones, iPads and Android. |

**III RELATED WORK**

Big Data has been the part of various organizations that need to be managed and understood for its further use [1].Most of the things which we do in the area of computer graphics and animation to depict abstract data falls into category of visualization. It helps to grasp the importance of raw data using knowledge and information to compute the semantic relationship of data for its efficient processing [2].

The need to collect data from different sources and using tools to visualize the data is a great work to be done and requires some special programming skills along with programming languages. Hence experts or specialists can only work on those tools [3].

Taking out information from data and finding abstract knowledge from it much fast in 2D or 3D space such that complex and large data is summarized [4].

The process of building a framework to acquire the data sets with the purpose of understanding its meaning and to show information status like its visual effects is known as information visualization [5].

To visualize various heterogeneous information sources, preprocessing needs to be done to display domain information. First dropping the unstructured information and taking out useful information from it and obtaining a well defined structure for it. Then making visualizing sets to control abstraction sets and its frame of reference.[6]

Thus, Big data analysis is a way to adapt new things easily, improves teaching and learning process and for decision making.[7]

**IV. CONCLUSION AND FUTURE WORK**

In this work, we have explored the various visualization application areas and the tools that are associated with each along with description of few of the tools. Each tool has its advantage of using it to apply for different data formats. This approach tells a story to focus on information that is important. Further work can be done to make a single platform for using various formats of data in a single tool.

**V. REFERENCES**

1. Ekaterina Olshannikova , Aleksandr Ometov , “Towards Big Data Visualization for Augmented Reality” IEEE 16th Conference on Business Informatics 978-1-4799-5779-8/14 $31.00 © 2014 IEEE DOI 10.1109/CBI.2014.42 (2014)

2. Jinsong Zhang, Yan Chen, Taoying Li , “Opportunities of Innovation under Challenges of Big Data” 10th International Conference on Fuzzy Systems and Knowledge Discovery 978-1-4673-5253-6/13/$31.00 ©2013 IEEE

3. Fernanda B. Viégas, Martin Wattenberg, Frank van Ham, Jesse Kriss, Matt McKeon, “Many Eyes: A Site for Visualization at Internet Scale” IEEE transactions on visualization and computer graphics, vol. 13, no. 6, November/December (2007)

4. M. Ghanbari , “Visualization Overview”, IEEE Member, Department of Computer Science Alabama A&M University 39th Southeastern Symposium on System Theory Mercer University Macon, GA, 31207, March 4-6, 2007

5. Kanae MATSUI, Masato YAMANOUCHI, Hideki SUNAHARA, “A Proposal of Framework for Information Visualization in Developing of Web Application” IEEE/IPSJ International Symposium on Applications and the Internet DOI 10.1109/SAINT.2011.85 (2011)

# 6. Matthias Kreuseler , Norma López , Heidrun Schumann , “ A scalable framework for information visualization” IEEE Symposium on Information Visualization 0-7695-0804-9/00 $10.00(2000)

7. B.Tulasi, “Significance of Big Data and Analytics in Higher Education” , International Journal of Computer Applications (0975 – 8887) Volume 68– No.14, April 2013.