International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 01 | Jan -2017 www.irjet.net p-ISSN: 2395-0072

A Review on data visualisation tools Used for Big Data

Bibhudutta Jena
School of Computer Engineering, KIIT University
bibhuduttajena728@gmail.com

Abstract-*Data visualization is an enactment of presenting the outcomes generated from analysis process of big data. On the basis of complexity of the data being analysed and the aspects from which it is analyzed, visuals can vary in terms of their dimensions such as one/two/multi dimensional. Now-a-days different class of tools are available in the market for data visualizing process. Some of them can be available on the open source platform which can be accessed and used with providing any cost. The paper aims to provide the notion of data visualization and need to visualize data in big data analytics. It also gives a brief idea about different tools used in data visualization to present the analyzed results.*

Key words*:* Big data, Data visualisation, Isoline, Isosurface image

1. INTRODUCTION

Data visualization is the study of presenting data in visual manner. Due to the enhancement of digital technologies, the width of manifold has enlarged manifold. Visuals in the form of graphs, images diagrams have completely escalated the internet. Visualization can be considered as an excellent medium to analyze and share information [1]. Instructional designers concentrate on scientific visualization to build learning content more interesting and easy to understand. Visual images assist to convey a huge amount of information to the brain of human beings. Similarly Visual analytics integrates the power of computational intelligence and computational power of modern computers. Data visualization also helps in detecting problems perceiving trends and outliers and also point out absorbing in a large dataset [2].

The rest of the paper has been summed up in the following way. The section 2 describes different techniques used for visual data representation. Similarly section 3 describes types of data visualization. Section 4 describes visualization of big data and gives a basic idea about different tools used in big data visualization. Finally section 5 concludes the paper.

2. TECHNIQUES IN VISUAL DATA REPRESENTATION

Data can be represented in multiple forms which consist of simple line diagrams, bar graphs, tables, matrices etc. Some techniques are generally used for a visual presentation of data are as follows.

Isoline – It is basically a 2D data representation of a curved line that generally transfers constantly on the surface of the graph, the plotting of line generally drawn on the basis of data arrangement instead of data visualization [3].

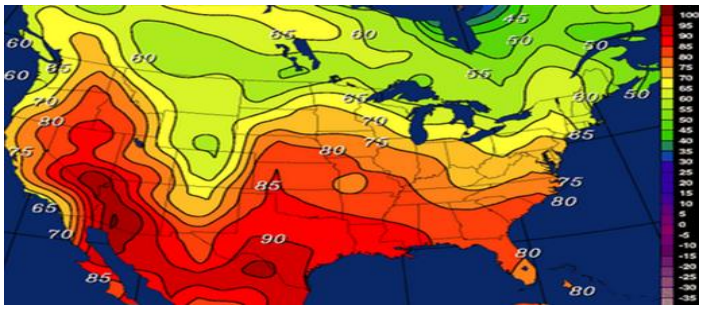

Fig 1. Isoline image [3]

Isosurface – It is a 3D representation of an isoline. Isosurfaces are designed to present points that are bound by a constant value in a volume of space i.e in a domain that covers 3D space [4].

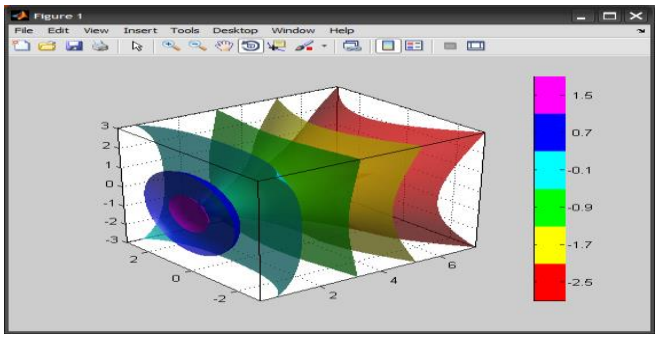

Fig2. Isosurface figure drawn in MATLAB

DVR (Direct Volume Rendering) In this method we can obtain a 2D projection for a 3D data set and it is very clear and transparent in visualization process [5].

Streamline – It is a field that is generated from the description of velocity vector field of the data flow. It is now-a-days widely used in data visualization process [6].

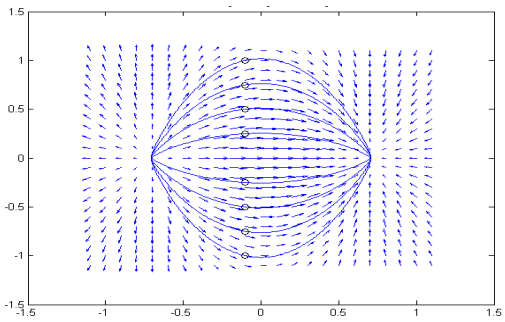

Fig 3. Streamline snapshot

Map: It is generally used to represent the location of different areas of a country. It is generally drawn on a plain surface. Google maps is generally widely used for data visualization. Now –a-days it is widely used for finding the location in different domains of country [7], [8].

Fig 4. Google map location of my residence area

Venn Diagram : Generally it is used to represent the logical relations that is existing in between finite collection of sets and Venn diagram generally are of many types i.e intersection of two sets and union of two sets etc [9].

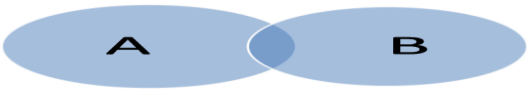

Fig 5. Intersection Venn Diagram

Ordinogram : It is generally used to perform the analysis operation of various sets of multivariate objects which are generally used in different domain. Simple two-dimensional graph is an example of ordinogram.

In the above section the paper has described number of ways by which the data can be visualized which are in the form of 1D,2D or in 3D, hence in the below section the paper has described different types of data visualization.

3. TYPES OF DATA VISUALIZATION

Data visualization can be done in different ways that the paper has described in the above in this section various data visualization techniques will be discussed.

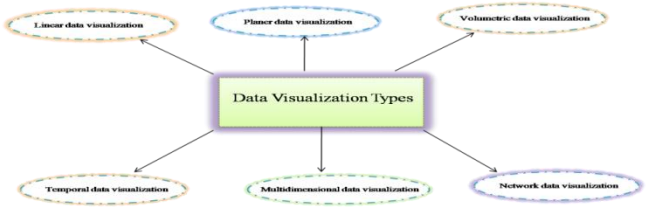

Fig 6.Types of data visualization

Linear Data Visualization- In this visualization technique data always represented in list format. Basically we can't consider it as a visualization technique rather than it is consider as a data organization technique. Hence in this process no tool is used to visualize the data. It is also called as 1D data visualization [10].

Planar Data Visualization – In this type of visualization data generally take in the form of images or charts over a plane surface. The best example of this type of data visualization is Cartogram. Some tools used to build planar data visualization are Geocommons, Poly maps etc.

Volumetric Data Visualization- In this approach the presentation of data generally involves with three dimensions to present the simulations, surface and volume rendering and commonly used scientific studies. Basic tools used for it are AC3D, True-space etc.

Temporal Data Visualization- In some approach visualizations are time dependent in nature so to visualize on the analyses of time the temporal data visualization is used which consist of Gantt chart, Time

series and sanky diagram etc. Now-a-days it is widely used to visualize the real time data.

Multidimensional data visualization- In this approach numerous data are generally used to represent the data. Generally pie charts, histograms, bar charts etc are generally used to multidimensional data visualization. Google charts tool is used to create multidimensional data visualization.

Network Data Visualization- This approach is generally used to represent the relations that are too complex in the form of hierarchies. Some of the basic tools used for network data visualization are hive plot, Pajek, Gephi, NodeXL etc.

However various type of data visualization techniques are present for data visualization out of these some of the techniques are described in the above.

4. BIG DATA VISUALIZATION

Now-a-days every organisation is struggling to monitor and process the huge amount of data generated in every second from multiple sources. So Data visualisation is a major process to reduce the turn-around time consumed during interpreting of Big data. Traditional visualization techniques are not sufficient to store or interpret the information that big data holds. Just for the example the traditional visualization techniques are not able to interpret videos, audios and complex sentences and beside to this during the interpretion process volume, velocity characteristics also a biggest challenge. Big data is highly dynamic in function hence existing visualization tools are not able to generate quality results. The paper has described some tools which are used for visualisation process of big data [11].

Excel- It is a new tool that is generally utilized for big data analysis and it helps to track and visualize data referring better insights..In the below figure [7] a snap shot of an excel sheet has been provided.

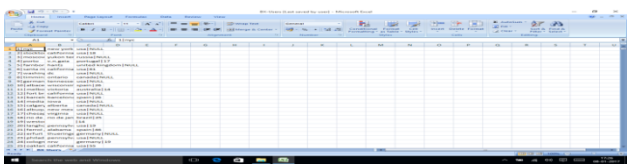

Fig 7. Excel sheet

Arc- This visualisation tool has developed by Digg and generally it is shown the results swaddled around a sphere. In the below figure [8] an arc diagram has been shown and it is basically used to find out the relationship among the users.

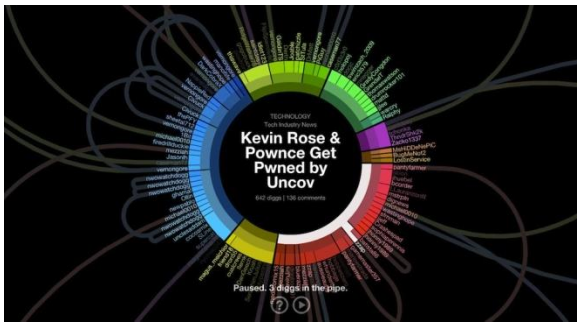

Fig 8.Arc Visual

Google Charts API- This tool allows the user to construct dynamic charts to be introduced in a web page. The chart which is obtained from the data and formatting parameters transferred through a hyper text transfer protocol and after that converted in to a PNG image. The below figure [9] describes the example of Google charts API.

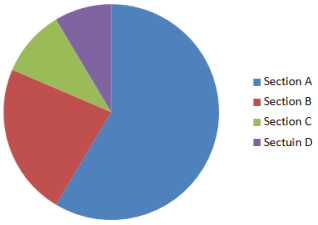

Fig 9. Charts obtained from Google charts API

TwittEarth- This tool is used to presenting live tweets from the entire world over a 3D globe. It is attempt to enhance social media visualisation and yield a global image mapping in tweets. The below figure [10] provides an example of Twittearth visual.

Fig 10. Twittearth Visual

Although there are no of tools are present for visualizing the big data and provide the analysis result but the paper has summed up some of them and discussed in the above section.

5. CONCLUSION

Big data analytics requires the implementation of improvised and advanced tools and methods. Due to economic and infrastructural constraints every organisations are not able to buy all the domains required for analyzing the data. Therefore to fulfil the requirement of advanced tools and technologies organizations are used open source data visualization tools for analyse the big data [12]. These tools guarantee to provide customer-centric solutions with some additional features such as real time accessibility to information, adaptive application domains etc. These tools are more secure and useful for providing the accurate results after analysis of big data. This paper discusses about different data visualization types and different data visualisation tools used for representing the analyzed results of big data.

REFERENCES

[1] D. Beaver, S. Kumar, H. C. Li, J. Sobel and P. Vajgel, “Finding a needle in Haystack: facebook's photo storage”, Proceedings of the 9th USENIX conference on Operating systems design and implementation, Vancouver, BC, Canada, (2010) October 4–6, pp. 1-8.

[2] J. Cui, T. S. Li and H. X. Lan, “Design and development of the mass data storage platform based on Hadoop”, Journal of Computer Research and Development, vol. 49, (2012), pp. 12-18.

[3] F. Mu, W. Xue, J. W. Shu and W. M. Zheng, “Metadata management mechanism based on route directory”, Journal of Tsinghua University (Sci & Tech), vol. 49, no. 8, (2009), pp. 1229-1232.

[4] L. N. Song, H. D. Dai and Y. Ren, “A learning method of hot-spot extent in multi-tiered storage medium based on huge data storage file system”, Journal of Computer Research and Development, vol. 49, (2012), pp. 6-11.

[5] Bibhudutta Jena"A review on Big Data Challenges and Techniques Used to Overcome from it" International Journal of Science and Research (IJSR), Volume 6 Issue 1, January 2017, 201 – 204.

[6] Bibhudutta Jena"A review on Joining Approaches in Hadoop Framework and Skewness Associate to it " Vol. 2 - Issue 6 (Nov - Dec 2016), International Journal of Engineering and Techniques (IJET) , ISSN: 2395 – 1303.

[7] Richard K. Lomotey and Ralph Deters , " Terms Mining in Document-Based NoSQL: Response to Unstructured Data " , 2014 IEEE International Congress on Big Data.

[8] S. Ghemawat, H. Gobioff and S. T. Leung, “The Google file system”, Proceedings of the 19th ACM symposium on Operating systems principles, Bolton Landing, NY, USA, (2003) October 19–22, pp. 29- 43.

[9] S. Yu, X. L. Gui, R. W. Huang and W. Zhuang, “Improving the storage efficiency of small files in cloud storage, Journal of Xi'An Jiao Tong University, vol. 45, no. 6, (2011), pp. 59-63.

[10] Muhammad H. Alizai, Georg Kunz, Olaf Landsiedel, Klaus Wehrle, Power to a first class metric in network simulations, in: Proceedings of the Workshop on Energy Aware Systems and Methods, February, 2010.

[11] Carson K. Leung, Hao Zhang, "Management of Distributed Big Data for Social Networks ",2016 16th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing.

[12] Rodney S. Tucker, Jayant Baliga, Robert Ayre, Kerry Hinton, Wayne V. Sorin, Energy consumption in IP networks, in: Proceeding of Optical Communication, 2008, pp.1.

AUTHOR

Bibhudutta Jena, is a CSI Accredited Student. Currently pursuing M. Tech (Computer Science and Engineering) at the School of Computer Engineering, KIIT University, Bhubaneswar. His area of interest Data Analytics, Data mining etc. He can be reached at bibhuduttajena728@gmail.com