International Journal of Advanced Information Science and Technology (IJAIST) ISSN: 2319:2682
Vol.5, No.7, July 2016 DOI:10.15693/ijaist/2016.v5i7.88-92

Big Data Visualization and its Tools: A Literature Review

Kanika[1], Mansi Gupta[2], Sangeeta Bhagat[3]

[123]*Department of Computer Science Engineering, I K Gujral Punjab Technical University, Kapurthala (India)*

Abstract—In recent years, a large amount of data has been generated by government, science and business. The size of data collected over the web i.e. from social media and mobile is even greater. Thus, visualization is very effective for presenting essential information in vast amounts of data. While Apache Hadoop and other technologies are emerging to support back-end concerns such as storage and processing, on the other hand visualization-based data discovery tools focus on the front end of big data—on helping businesses explore the data more easily and understand it more fully. In this paper, the introduction about this new technology, its principles, key features of visualization based data discovery tools and some visualization methods used by them is presented. This paper concludes with the Good Big data visualization practices to be followed and comparison of some popular big data visualization tools.

Index terms *-Analytics; Data Visualization (DV).*

I. INTRODUCTION

As we've entered in an era where data are continuously driven for a variety of purposes it is important to make timely decisions based on available data which is crucial to cyber, medical, business success, national security, and disaster management. Big data are the collection of large amounts of structured, semi-structured and unstructured data. Big data means large amount of data, such large that is difficult to predict, collect, store, analyze, visualize, manage and model the data. Thus, data visualization helps to manage this massive amount of data which is called big data.

II. DATA VISUALIZATION

Big data visualization is deriving real meaning from big data. It is both art and science. Its purpose is insight, not pictures. It is not about aesthetics. Visualization of information takes the advantage of the vast, and often misspent, capacity of the human eye to detect information from pictures and illustrations. Data visualization shifts the load from numerical reasoning to visual reasoning. Getting information from pictures is far more time-saving than looking through text and numbers – that's why many decision makers would rather have information presented to them in graphical form, as opposed to a written or textual form. Its primary objectives are:

- Communicating knowledge clearly and effectively.
- Displaying data to understand cause and effect.

III. KEY FEATURES OF VISUALIZATION BASED DATA DISCOVERY TOOLS

Visualization-based data discovery tools allow business users to mash up disparate data sources to create custom analytical views with flexibility and ease of use that simply didn't exist before. End users can view the graphics on the same gadgets, or on even smaller mobile devices such as tablets or in smartphones.

Key Features of These Visualization-based Data Discovery Tools [1].

- Enable real-time data analysis.
- Support real-time formation of dynamic, interactive presentations and reports.
- Allow users to interact with data.
- Hold data in-memory, where it is accessible to multiple users.
- Allow users to collaborate and share securely.

A. Addressing the Three Vs [1]

These Visualization-based data discovery tools take the challenges presented by the "three Vs" of big data and turn them into opportunities for growth.

Volume [1]

Visualization based-data discovery tools are designed to work with an immense no of datasets, so business can turn their attention from simply managing the deluge of data to gaining rich insights. These tools enable business to derive meaning from large and growing volumes of data.

Variety [1]

Visualization-based data discovery tools are designed to mash up, or combine, as many data sources as needed. That means businesses can derive more meaning from structured data, semi structured and unstructured data sources such as social media and sensor data. Using interactive bubble charts, 3-D data landscapes, tree maps, box plots, heat maps, word clouds, and many other types of graphics, businesses can view, interact, and interpret with complex data from a multitude of sources.

Velocity [1]

With visualization-based data discovery tools, organization can replace batch processing with real-time processing of ceaselessly updated input streams. The tools also support the democratization of data discovery, so more people can access real-time data sources such as click streams, and analyze and view the data without having to wait for reports.

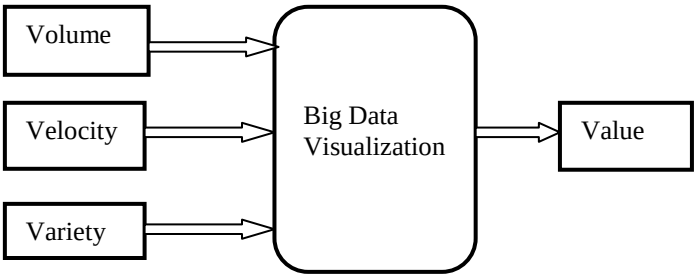

Figure 1. The 4th V of Big Data

When businesses address the three Vs in parallel, they achieve the fourth V: Value (Figure 1).

Value
User can run certain queries against the data stored and thus can deduct important results from the filtered data obtained and can also rank it according to the dimensions they require. These reports help the people to find the business trends according to which they can change their strategies.
Visualization-based data discovery tools don't just enable users to create attractive infographics and heatmaps but also create business value by enabling more workers to gain more insights from more data. Instead of waiting weeks or months for static reports, employees can visualize and analyze real-time data on their own. They can also collaborate with co-workers using interactive graphics which are online to generate new ideas and identify previously unseen trends.

IV. Five Best Practices and Principles of Big Data Visualization

There are five principles of big data visualization which are as following:

Principle 1: Context is king [2]

- The context in which big data is visually placed impacts the knowledge that can be communicated.
  Mapping with D3.js
  a) (D3 – Data Driven Document is a java script library for visualizing data. It helps bring data to life using SVG, HTML and CSS. It makes the data interactive through the use of transformations and transitions (Zooming and Panning). It builds data visualization framework.)
  b) D3.js includes routines for handling geographic information.
- Make the right comparisons for the context.
- Many problems are multivariate (i.e. multiple variables) which needs to be recognized in data visualization.

Principle 2: Visualizations must match data [2]

- The knowledge interpreted through visualizations must match underlying data.
- Follow convention in modelling your data and axis.
- The objective of data visualization is to communicate information to the viewer, misleading by deception or confusion (even accidently) will not serve your purpose

.
Principle 3: Escape flatland when useful [2]
Although devices present data in two dimensions, this desolate flatland can be escaped. Using three dimensions to show to convey information.

Principle 4: Show your work [2]
Aggregating whole details can reveal the knowledge present in data.

Principle 5: Insights from Hay Bales [2]

- Layering and parallelizing data visualization can reveal insights but be careful not to form haystacks.
- Layering data on a common X or Y axis maximizes visualization of coincidence and anomalies and best use for time series data.
- Parallelizing the data is as powerful as layering data to show significant differences between multiple data sets.

V. Visualization Methods

There is a fairly large amount of data visualization tools that offer different possibilities. It can be classified as a character of data to be visualized by the tool; visualization techniques and the samples as the data can be submitted; interoperability with visual imagery and techniques for better data analysis[3].

Visualization tools are able to work with:

- univariate data – one dimensional arrays, time series, etc.;
- two-dimensional data – point two-dimensional graphs, geographical coordinates, etc.;
- multidimensional data – financial indicators, results of experiments, etc.;
- texts and hypertexts – newspaper articles, web documents, etc.;
- hierarchical and links – the structure subordination in the organization, e-mails of people, documents and hyperlinks, etc.;
- algorithms and programs – information flows, debug operations, etc.

Processed and Visualized Data

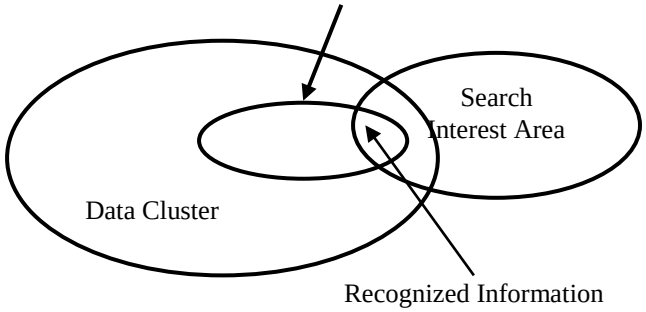

Figure 2. Human Perception Capability Issue

However, different visualization methods are currently used to represent various types of data [6]. Obviously, the number of forms is only limited by human imagination. The main requirement is clarity and ease of the analysis of the represented data. Visualization techniques can be both elementary (line graphs, charts, bar charts, etc.) and complex (based on the mathematical apparatus). Furthermore, the visualization can be used as a combination of various methods. However, visualized representation of data is abstract and extremely limited by ones perception capabilities and requests (see Figure 2). Types of visualization techniques are classified as following:

- 2D/3D standard figure [5] – bars, line graphs, etc. The main drawback of this type is the complexity of the acceptable visualization for complicated data structures;

- Geometric transformations [6] – scatter diagram of data, parallel coordinates, etc. This type is aimed to the multi-dimensional data sets transformation in order to display it in the Cartesian and non-Cartesian geometric spaces. This class includes methods of mathematical statistics unit;

- Display icons [7] – ruled shapes (needle icons) and star (star icons). Basically, this type is displaying the values of elements of multidimensional data in properties of images. Such images may be: human faces, arrows, stars, etc. Images can be grouped together for a holistic analysis. The result of the visualization is a texture pattern, which is different corresponding to specific characteristics of data;
  Methods focused on the pixels [8] – recursive tem-plates and cyclic segments. The main idea is to display the values in each dimension into the coloured pixel and to merge some of them according to specific measurements. Since one pixel is used to display a single value, therefore visualization of large amounts of data (over one million values) can be reachable with this methodology;

- Hierarchical images [9] – tree maps and overlay measurements. These type methods are intended to be used with the hierarchical structured data.

VI. BIG DATA VISUALIZATION TOOLS

There are many big data visualization tools are available and some are the most flexible, comprehensive, sophisticated visualization tools which are capable of handling big data. Some tools are Open-Source applications that can be used in conjunction with one another or with your existing design applications, using HTML5, JSON, Python, SVG, JavaScript or drag and drop functionality with no programming knowledge required at all like d3.js, Plymaps, Nodebox, Flot, Inkscape, Leaflet and many more.[11].There is not a single best visualization product. For example, Tibco's Spotfire has best web client and analytical functionality. Qlikview is the best visualization product for interactive drill-down capabilities. On the other hand, The Microsoft BI platform provides better price-performance ratio and good as a backend for Data Visualization (especially with release of SQL Server 2012). Tableau has the best ability to interact with OLAP cubes etc. Thus, comparison of some of Data Visualization platforms such as Tibco's Spotfire, Qlikview, Tableau Software and the Microsoft BI platform (PowerPivot, Excel 2010, SQL Server with its VertiPaq, SSAS, SSRS and SSIS) on the basis of technical and business criteria is as following [12].

Business Criteria	**Qlikview**	**Spotfire**	**MS BI Stack**	**Tableau**	**Comment Scalability, Price, Speed**
Scalability	Limited by RAM	Unlimited	Good	Very Good	Need the expert in scalable SaaS
Implementation Speed	High	Good	Average	Good	Qlikview is fastest to implement
Pricing	Above Average	High	Average	High	Microsoft is the price leader
Licensing/ Support Cost	High	High	Average	High	Smart Client is the best way to save
Enterprise Readiness	Good for SMB	Excellent	Excellent	Good for SMB	Partners are the key to SMB market
Long – Term Viability	1 Product	Good	Excellent	Average	Microsoft are 35+ years in business
Mindshare	Growing fast	Analytics Market	3rd attempt to win BI	Growing fast	Qlikview is a Data Visualization Leader and successful IPO.

Table 1. Comparison of four Big Data Visualization Tools on the basis of technical criteria [12]

Table 2. Comparison of four Big Data Visualization Tools on the basis of business criteria [12]

Technical Criteria	Qlikview	Spotfire	MS BI Stack	Tableau	Comment Analytics, Drilldown, UI
Clients for End Users	Limited by RAM	Unlimited	Good	Very Good	Need the expert in scalable SaaS
Interactive Visualization	High	Good	Average	Good	Qlikview is fastest to implement
Data Integration	Above Average	High	Average	High	Microsoft is the price leader
Visual Drill-Down	High	High	Average	High	Smart Client is the best way to save
Dashboard Support	Good for SMB	Excellent	Excellent	Good for SMB	Partners are the key to SMB market
Integration with GIS	1 Product	Good	Excellent	Average	Microsoft are 35+ years in business
Modelling and Analytics	Growing fast	Analytics Market	3rd attempt to win BI	Growing fast	Qlikview is a Data Visualization Leader and successful IPO.
UI & Set of Visual Controls	Best	Very Good	Good	Very Good	Need for UI expert to integrate DV components
Development Environment	Scripting, Rich API	Rich API, S+	Excellent	Average	Tableau requires less consulting than competitors
64-bit In-Memory Columnar DB	Excellent	Very Good	Very Good	In – Memory Data Engine	64-bit RAM allows huge datasets in memory
Summary - Best for:	***DV, Drilldown***	***Visual Analytics***	***Backend for DV***	***Visual OLAP***	***Good Visualization requires a customization***

VII. CONCLUSION

This paper described the new concept of big data visualization, its importance, its principles and the existing tools. To adapt this new technology, many challenges and issues exist which need to be brought up right in the beginning before it is too late. These challenges and issues will help the business organizations which are moving towards this technology for increasing the value of the business to consider them right in the beginning and to find the ways to counter them. Thus, it can be concluded that the data visualization methodology may be improved by placing the most essential data in the most recognizable area of the human visual field.

REFERENCES

[1] Intel IT Center White Paper Big Data Visualization. *Turning Big Data Into Big Insights The Rise of Visualization-based Data Discovery Tools*, 15th March2013, http://www.intel.com.

[2] Rami Sayar, Data Visualization in Practice, http://www.slideshare.net/ramisayar/fitc-data-visualization-in-practice-42809637.

[3] Ekaterina Olshannikova, AleksandrOmetov, YevgeniKoucheryavy. *Towards Big Data Visualization forAugmented Reality.*

[4] M. Friendly, Milestones in the history of thematic cartography, statistical graphics, and data visualization, August 2009.

[5] M. Tory, A. E. Kirkpatrick, M. S. Atkins, and T. Moller, *Visualization task performance with 2d, 3d, and combination displays,* IEEE Trans-actions onVisualization and Computer Graphics, vol. 12, January 2006.

[6] S. R. M. Oliveria and O. R. Zaiane, *Geometric datatransformation for privacy preserving clustering*, DEPARTMENT OF COMPUTING SCIENCE, 2003.

[7] C. G. Healey and J. T. Enns, *Large dataset at a glance:Combining textures and colors in scientific visualization*, IEEE Transactions on Visualization and Computer Graphics, vol. 5, no. 2, and 1999.

[8] D. A. Keim, *Designing pixel-oriented visualizationtechniques: Theory and applications*, IEEE Transactionson visualization and computer graphics, vol. 6, no. 1, 2000.

[9] M. Kamel and A. Camphilho, *Hierarchic imageclassification visual-ization*, ICIAR, 2013.

[10] Deepa Gupta, Sameera Siddiqui. BIG DATA IMPLEMENTATION AND VISUALIZATION.

[11] Andi Lurie. 39 Data Visualization Tools for Big Data, 13th February, 2014, http://blog.profitbricks.com/39-data-visualization-tools-for-big-data.

[12] Andrei Pandre, "Comparison of DV platforms",https://apandre.wordpress.com/tools/comparison.

Author Profile

Kanika received the B.Tech. degree in computer science engineering from CTIEMT, Shahpur Jalandhar. Currently doing M.Tech in computer science engineering from I. K. Gujral PTU Main Campus. Her research interest includes big data analytics, machine learning, data mining.

Mansi Gupta is an assistant professor in I. K. Gujral PTU Main Campus. Her research interest includes wireless sensor network.

Sangeeta Bhagat received the B.Tech. degree in computer science engineering from CTIEMT, Shahpur Jalandhar. Currently doing M.Tech in computer science engineering from I. K. Gujral PTU Main Campus. Her research interest includes big data analytics, data mining.