

- [13]. James Danziger Debora Dunkle. Methods of Training in the Workplace. Centre for Research on Information Technology and Organizations, School of Social Sciences University of California, Irvine CA 92697-4650
- [14]. The Forum Report.e-Learning: Adoption Rates and Barriers <http://www.internetime.com/itimegroup/FrmRep1.pdf>
- [15]. Ruth Clark. Six Principles of Effective e-Learning: What Works and Why. <http://www.e-Learningguild.com/pdf/2/091002DES-H.pdf>
- [16]. The advantages and challenges of e-learning implementation: The story of a developing nation
- [17]. Benjamin KWOFIE, Prof. Anders HENTEN

BIG DATA: THE NEXT BIG OPPORTUNITY FOR LIBRARIANS

Dr. Shantashree Sengupta | Librarian & Head | DLIS| Chandmal Tarachand Bora College | Shirur | Dist.Pune | Maharashtra | shantashrees@gmail.com

ABSTRACT

Big data includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a specific time period. Due to its widespread presence in recent times, librarians need to know the basics of Big Data and how it can impact academic research. The implications of big data are not only limited to research but to daily working of library professionals. The present paper details the concept of big data and how it can be beneficial in the field of librarianship.

Keywords: Big Data, Librarianship, Big Data Curation, Big Data Tools.

INTRODUCTION

Big data is a broad term used to describe a massive volume of both structured and unstructured data that is so large it is difficult to process using traditional database and software techniques. They are extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions.

Definition

Gartner Group Inc. (2001) defined Big Data as “high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization.”

An example of big data might be petabytes (1,024 terabytes) or exabytes (1,024 petabytes) of data consisting of billions to trillions of records of millions of people—all from different sources (e.g. Web, sales, customer contact center, social media, mobile data and so on). The data is typically loosely structured data that is often incomplete and inaccessible.

Characteristics of Big Data

Big data is characterized by three Vs: Volume, Velocity, and Variety.

- a) **The first V, volume:** The size of data sets is huge in big data compared to the regular data. The size depends on the specific discipline, but big data is loosely defined as data that cannot be stored or analyzed by conventional hardware and software. The size of the data determines the value and potential of the data under consideration, and whether it can actually be considered big data or not. Traditional software can handle megabyte and kilobyte sized data sets, while big data tools can handle terabyte and petabyte sized data sets.
- b) **The second V, velocity:** It covers the speed in which data is created. Velocity relates to the speed at which the data is generated and processed to meet the demands and the challenges that lie in the path of growth and development. It can be understood by imagining the striking speed in which messages, posts, tweets are created/ uploaded on social networking sites.
- c) **The third V, variety:** It makes big data sets more challenging to organize and analyze. This helps people who are associated with and analyze the data to effectively use the data to their advantage and thus uphold its importance. Big data sets can contain unstructured data such as email messages, photographs, postings on internet forums, and even phone transcripts.

Need of Big Data

The management and analyzing of big data sets initially was limited to the domain of big business, central governments and scientific and cultural communities engaged in higher education and research due to the huge

hardware and software costs involved in analyzing the big data. It was not possible for every profession to make use of big data sets and to deal with them. In due course of time, as a result of release of open source software tools, availability of cheaper hardware and software and lower costs of commodity servers, it became more easily available to business, academia, and local governments. Also, the ability to analyze big data in real time has changed. Early users of big data were born-digital firms that relied on analyzing large data sets to manage their success like Facebook, LinkedIn, Google, and Twitter.

Instead of calculating the amount of big data one has, it is very important to decide about the usage of that data. One can take data from any source and analyze it to find answers that enable-

- Cost reduction
- Time reduction
- New product development and optimize offering
- Smart decision making

Importance of Big Data for Librarians

The developments and challenges of 21st century have already influenced the Librarians to embrace ICT applications in daily library workings in order to provide effective and efficient services to the library users. With the onset of big data, it has now become important for the librarians to know how big datasets can be made more useful, visible and accessible by creating taxonomies, designing metadata schemes, and systematizing retrieval methods. Big data is creating a widespread presence in the world of information and is likely to impact on all kinds of libraries. Librarians depending on the type of library, in which they work, need to know the basics of big data and how it can affect the nature of work and how the big data can be used. It's not just increased amount of data but also improved tools to store, aggregate, combine and analyze the data. For eg. Corporate librarians need to know how companies hold big data, how such data mining provides a competitive advantage, and how students might need to tackle with big data sets in future employment. Librarians working in Scientific Research Libraries need to know how big data differs from other scientific data and the impact of emerging software and hardware used for its analysis. Humanities and Social Science librarians should know that big data is becoming common in their disciplines as well, and is no longer restricted to science and business only. Digital archivists, data curators, and other types of librarians are also asked to advise their faculty on the storage and accessibility of big data sets.

Library administration and management should examine what types of big data sets their library could be gathering and analyzing using big data tools. Library professionals should discuss what new they can measure with the help of the big data tools? Is the library dealing with some massive data set related to faculty research which previously was out of reach because of software and hardware constraints? Use of big data can also be done for developing library collection, tracking the use of library materials by its users etc. Big data can be used by the librarians in different ways for library administration, user satisfaction and retrieval of information in least possible time.

Computer Business Review (CBR, 2015) has published list of ten most popular Big Data tools. They are:

- 1) **Splice Machine:** This is a real-time SQL-on-Hadoop database which can help to derive real-time actionable insights, which is a clear benefit for those who are aiming for quick development. It is SQL 99 compliant with standard ANSI SQL and can scale from gigabytes to petabytes. As well as support for .NET, Java and Python, it also offers support for those written in JavaScript/AngularJS.
- 2) **MarkLogic:** MarkLogic is built to deal with heavy data loads and allow users to access it through real-time updates and alerts. It provides geographical data that is combined with content and location relevance along with data filtering tools. This tool is ideal for those looking at paid content search app development.
- 3) **Google charts:** This free tool comes with various capabilities for visualizing data from a website such as hierarchical tree maps or just simple charts. This tool is easily implemented by embedding JavaScript code on a website and allows you to sort, modify and filter data as well as the ability to connect to a database or pull data from a website. Offering support for popular languages and with the security of knowing that Google will likely keep on improving its offering, this is a good option for many standard developers.

- 4) **SAP in Memory:** SAP's HANA platform offers a number of advantages over some of the competition, such as the ability to integrate and analyze large workloads of data to be analyzed in real time. This is extremely beneficial for the developer who is looking for speed to market.
- 5) **Cambridge semantics:** Using the Anzo Software Suite, this open platform helps you to collect, integrate and analyze Big Data to help you build Unified Access solutions. The software has a data integration machine that streamlines data collection and assists with analytics. The key features include being able to combine data from multiple sources and customized dashboards to make analysis easy.
- 6) **MongoDB:** This is an open-source documental database that is ideal for developers who want to have precise control over the final results. MongoDB is also scalable and includes third party log tools such as Edda and Fluentd.
- 7) **Pentaho:** Pentaho joins data integration and business analytics for visualizing, analyzing and blending Big Data. The open and embeddable platform comes with extensive analytics capabilities with data mining and predictive analysis. The connectivity to any type of data source or source of data with native support for Hadoop, NoSQL and analytic databases. The data integration tools mean that users do not require coding in SQL or writing MapReduce Java functions.
- 8) **Talend:** Straight away, one of the key benefits of Talend's Open Studio is that it is open source, which means that improvements will keep on rolling out as the community tweaks the tool. Its tools include products for developing, testing and deploying data management and application integration products. Additionally the company manages the full lifecycle, even across enterprise boundaries.
- 9) **Tableau:** Tableau is one of the more well-known names in the data visualization sphere but it offers many tools for developers that are supported by an active community. Some of the key features of this software are its in-memory analytics database and advanced query language. API, XML, User Scripts, Python, and JavaScript are all supported and so are a number of browser extensions.
- 10) **Splunk:** Splunk specializes typically in harnessing machine data created from a number of different sources, such as websites, applications and sensors. The company also enables developers to write code using any technology platform, language or framework.

CONCLUSION

Big data is a buzzword and use and adoption of big data is beneficial and allows efficiencies in terms of cost, productivity, and innovation. Data must be processed with advanced tools to reveal meaningful information. Librarians have always been great at information organization and management. A librarian does not need to become a programmer but should try to know and have basic knowledge of how various software tools can transform the huge amount of data. Big data has given an opportunity to the librarians to play an important role in the big data universe because they have the skills, knowledge and service mentality to help all irrespective of their discipline.

REFERENCES

- [1]. Bell, Steven. (2013). Promise and Problems of Big Data from the Bell Tower. Retrieved November 30, 2015 from <http://lj.libraryjournal.com/2013/03/opinion/steven-bell/promise-and-problems-of-big-data-from-the-bell-tower/>
- [2]. Bieraugel, M. *Keeping up with...Big Data*. Retrieved November 10, 2015 from http://www.ala.org/acrl/publications/keeping_up_with/big_data
- [3]. *Big Data*. Retrieved on November 10, 2015 from https://en.wikipedia.org/wiki/Big_data
- [4]. Definition of Big Data. Retrieved November 10, 2015 from http://www.webopedia.com/TERM/B/big_data.html
- [5]. Ten of the most popular big data tools for developers. Retrieved November 20, 2015 from <http://www.cbronline.com/news/big-data/analytics/10-of-the-most-popular-big-data-tools-for-developers-4570483>
- [6]. What is Big Data? Retrieved November 20, 2015 from http://www.sas.com/en_us/insights/big-data/what-is-big-data.html