

Application of Big Data in Academic Libraries

Dr. Ganesh Ramdas Sanap

Librarian

MTES's Doshi Vakil Arts College and G. C. U. B. Sci. & Com. College, Goregaon, Mangaon, Raigad, Maharashtra
digilibtrend@gmail.com

Abstract: *Digital era produces vast amounts of data which leads to advances in data storage and analysis and the emergence of the concept of big data. Big data will be an important resource used in academic research and to make data driven decision making. This paper provides an overview about the big data concept and its application in libraries. This paper aims to create an awareness amongst librarians about big data, its applications and the role of librarian in big data scenario.*

Keywords: Big Data, Academic Libraries, Librarian's Role, Data Curation, etc

I. INTRODUCTION

Data is generated at an unprecedented level and at an incredible pace from everywhere such as governments, universities, Scientists, businesses & non-profit organizations. Big data affects the libraries directly in such a way that libraries can use the big data tools to analyze big data sets and tangentially as the faculties at the educational institutions can incorporate big data into their research. Big data is generated from various sources such as sensors, devices, videos, audios, network, log files, transactional applications, web and social media whose size or type is beyond the ability of traditional relational databases to capture, manage and process the data. Big data when analyzed allows users to make better and faster decisions.

‘As the library's role is to identify, evaluate, select, organize, and describe resources, as well as preserve, disseminate, use, and reuse resources and information, the library's role is getting more complex today than ever. This role for library, librarians, and information scientists as information providers finds them in a crossroad that forces them to choose if they will play a central role in the meta-fourth industrial revolution as central-information providers or if they will keep a profile as organizations that continue to provide “traditional” services to patrons. The dilemma is big, as big are the big data and the changes that will bring to our society’ (Garoufallou, 2021).

1.1 How Big is Big Data

Big data is loosely structured data which is often incomplete and inaccessible. It may consist of billions to trillions of records of millions of people all from different sources for e. g. web, sales, customer contact center, social media, mobile data etc. the starting of terabytes may be considered as starting of big data. It may be equivalent to petabyte or Exabyte.

Definition

“Big Data is the amount of data that cannot fit into the memory of a single computer system. With each passing day, Big data is growing bigger, is more difficult to make sense of, is being generated at a much faster rate and this trend is only going to intensify in our data-driven digital world” (<https://intellipaat.com>).

Ed Dumbill, the editor-in-chief of a journal devoted to the topic of big data, offered this broader, more conceptual definition:

“Big data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the strictures of your database architectures. To gain value from this data, you must choose an alternative way to process it.” (Dumbill, 2013)

‘Big data is defined as a term for data sets that are so large and complex that traditional data processing applications are inadequate to do analysis, capture, store, data curation, sharing, transfer and so on.’

1.2 Characteristics of Big Data: 5 Vs of Big Data

- Volume: The amount of data,
- Velocity: The speed of data in and out, and
- Variety: The range of data types and sources which include: unstructured text documents, picture, video, email, audio, stock ticker data, financial transactions, etc.
- Variability: At times, the data flow is highly inconsistent with periodic peaks which hamper the process of handling and managing data effectively.
- Complexity: As large volumes of data come from multiple sources, data management becomes a challenging task.

‘In fact, the data sets are so big and complex that it becomes very difficult and challenging to process them using traditional data processing applications. It is estimated that about 2.5 quintillion bytes of data are created every day.’ (Verma, 2022)

1.3 Importance of Big Data

The use of Big data is becoming very crucial and following are the importance of big data.

1. There is a significant amount of data which is yet not captured in digital form or made accessible or searchable through the network. Big data can unlock the value hidden in these forms of data by making information more transparent.
2. Big data allows narrower segmentation of customers and this helps in producing more precisely tailored products and services.
3. Sophisticated analytics would help to improve decision making.
4. Future generation products and services can be developed with the help of big data.

1.4 Big Data Technologies

Coursera (2022) has been described the big data technologies on their web page. According to him ‘Big data technologies can be categorized into four main types: data storage, data mining, data analytics, and data visualization. Each of these is associated with certain tools, and you’ll want to choose the right tool for your business needs depending on the type of big data technology required.

A. Data Storage

Big data technology that deals with data storage has the capability to fetch, store, and manage big data. It is made up of infrastructure that allows users to store the data so that it is convenient to access. Most data storage platforms are compatible with other programs. Two commonly used tools are Apache Hadoop and MongoDB.

B. Data Mining

Data mining extracts the useful patterns and trends from the raw data. Big data technologies such as Rapid miner and Presto can turn unstructured and structured data into usable information.

C. Data Analytics

In big data analytics, technologies are used to clean and transform data into information that can be used to drive business decisions. This next step (after data mining) is where users perform algorithms, models, and more using tools such as Apache Spark and Splunk.

D. Data Visualization

Finally, big data technologies can be used to create stunning visualizations from the data. In data-oriented roles, data visualization is a skill that is beneficial for presenting recommendations to stakeholders for business profitability and operations to tell an impactful story with a simple graph. The Tableau and Looker are popular tools in data visualization.



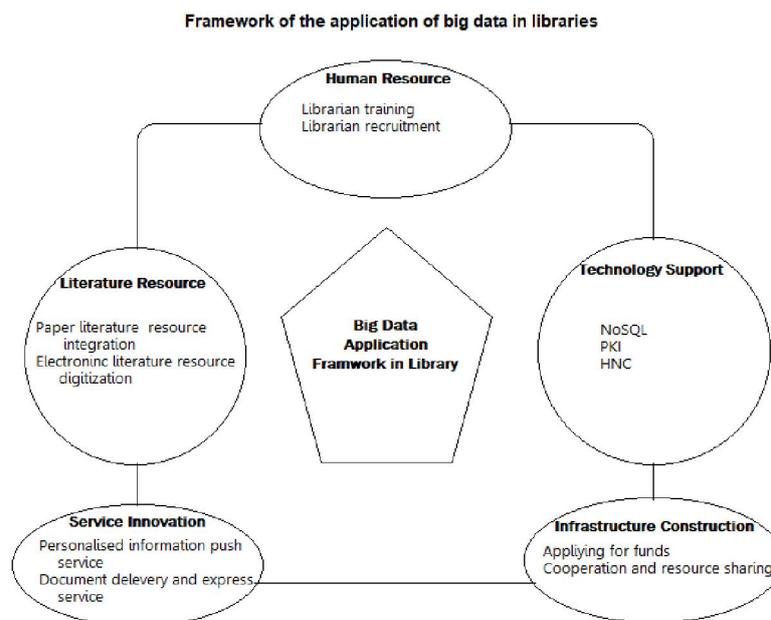
1.5 Benefits of Big Data in Different Sectors

Most of the industries today face challenges due to the large and complex amount of data that they deal. With the help of big data analytics, the following industries would benefit which is as follows:

- 1. Banking and securities: Big data could provide solutions in the areas such as card fraud detection, archival of audit trails, customer data transformation and so on.
2. Communication and media entertainment: Big data helps in creating quality content for the target audiences.
3. Health care providers: the technology involved in Big data allows faster and efficient identification of healthcare information which can help in tracking the spread of chronic diseases.
4. Education: it would help students to pick up courses according to their liking and even help in increasing the effectiveness of the teachers.
5. Government: government can use big data in energy exploration, financial market analysis and health related research and environmental protection.

1.6 Application of Big Data in Libraries

The application of big data in a library involves a complex framework as different modules interact and work together. Hussain (2022) presented framework of the application of big data in libraries which involves Human Resource, Technology support, Infrastructure Construction, service innovation and literature resources.



The unique skills of librarians are essential to make use of big data. Librarians' jobs have evolved from traditional setup into data mining, consultant, data management, librarian data services, data research scientist, architecture librarians, librarian data designs, coordinator, scientific data curation, librarian specialist/metadata, data curator, etc. Librarians should change their traditional role and learn the new skills needed in the 21st century. Large data sets can improve the quality of library services. Data resources of digital libraries can be used as big data by using techniques to bring innovation and introduce important digital changes. Big data provides insights into the utilization of resources, decision-making, and library user needs more accessible.

Academic libraries are already using high-level services to collect, evaluate and manage library resources. Changes in technology and new ways of communication are demanding the redefined academic library. A significant milestone was announced in 2010, which was the collaboration between the world's most extensive library – Library of Congress and Twitter. The purpose of the partnership was to analyze all the tweets on Twitter to retain and archive users' data. In 2012, the Harvard University library also published its metadata, including books, videos, audio recordings, manuscripts and other contents. Ensuring data privacy and security are also concerns when implementing big data analytics in library services (Bhat, 2018).

Big data analytics could be applied in the following areas of the libraries such as

- For superior search results: Data mining and text analytics on the past loan records and book bibliographies could enhance search results and recommendations.
- Demand Analysis: It would help in forecasting demand for new and existing titles.
- Planning Library Collection: The technology used would optimize to plan the category mix in the collection by taking into consideration the space and budget constraints.

Librarian's Roles in Big Data:

Jain, Priti (2016) has mentioned the various roles of librarian in big data, these are as follow:

- Data manager: Librarian is a well placed to deal with data management issues. Since they have expertise and experience with institutional repositories; they have a legacy of preserving knowledge for future generations, advocate open access and as such are advocates of the value of open data.
- Intermediaries: Librarians can act as intermediaries linking to faculties with peers in other disciplines to facilitate interdisciplinary research. They can create data literacy and awareness of data management issues and practices among students.
- Advocate of big data: Based on their vast experience with open data, librarians can foster the importance generation of researchers. Advocacy of big data is quite viable for librarians because by nature of their business they are always with their users.
- Big data curator: In this e-science and big data revolution age, libraries are expected to play key role in data curation, since librarians have the skills and most of the infrastructure to curate various types of data. Library user may or may not be knowledgeable in the areas of big data and the process of obtaining the data. Hence this is an excellent opportunity for librarians to assist library user on these pertinent issues.
- Thought leader on big data curation: Librarians have been doing a lot of work on data curation and we could be vital partners in terms of metadata standards, metadata creation, preservation and managing the whole information life cycle which librarians are really good at.

Facilitator in source of choice: every day the library users produces raw material in their quest for new insights into the past and new ways to explain culture, politics and philosophy. skilled librarian, who knows how to effectively search for not only books but primary sources of material across the world and who can understand, create and navigate catalog to accelerate a researcher's efforts.

E-research data manager: big data underpins e-research and it must be preserved for future re-use. Academic libraries, being research oriented and facilitating students and academia research initiatives are considered research libraries. They can play a significant role in research data management to accelerate their user's research activities.

II. CONCLUSION

The discussion in this article is focused on big data in libraries perspective. The capabilities of big data have captured the attention of the world of libraries. Library and information professionals can play a big role in the universe of big data as they have the skills, knowledge and service mentality to help universities, business and governments. With the help of these powerful analytics which big data technologies offers, librarians can look at the data in new ways thus adding value to different services and programs.

REFERENCES

- [1]. What is Big Data? . (2022). From Intellipaat.com: <https://intellipaat.com/blog/what-is-big-data/>
- [2]. Bhat, W. (2018). Long-term preservation of big data: prospects of current storage technologies in digital libraries. Library Hi-Tech. From (2018), "", Library Hi-Tech.
- [3]. Coursera. (2022, September 29). 4 Types of Big Data Technologies. From www.coursera.org: <https://www.coursera.org/articles/big-data-technologies>
- [4]. Dumbill, E. (2013). Making Sense of Big Data. Big Data, 1(1), 1-2. doi:<https://doi.org/10.1089/big.2012.1503>

- [5]. Garoufallou, E. &. (2021). Big Data: Opportunities and Challenges in Libraries: a Systematic Literature Review. *College & Research Libraries*, 82(3), 410. doi:<https://doi.org/10.5860/crl.82.3.410>
- [6]. Hussain, A. (2022, Feb). Impact of big data on library services. *Library High-Tech News*. doi:<https://doi.org/10.1108/LHTN-02-2022-0022>
- [7]. Huwe, T. (2014). "Big Data and the Library: A Natural Fit. *Computers in Libraries*, 34(2), 17-18.
- [8]. Laney, D. (2001). 3-D Data Management: Controlling Data Volume, Velocity and Variety. META Group Original Research Note.
- [9]. Priti, J. (2016). Big Data: Opportunities and Challenges for Librarians. *Future librarianship* (pp. 307-311). Jalgaon: Academic Book Publication.
- [10]. Verma, E. (2022). How Big is Big Data?: Vs of Big Data. From www.simplilearn.com: <https://www.simplilearn.com/how-big-is-big-data-rar335-article>