



The Evolution of Big Data

Rishabh Singh, Vibhor Jain, Rhythm Yadav, Ujjawal Jain

Inderprastha Engineering College, Ghaziabad, 201010, UP, India

Abstract: Big data is still an enigma to many people. It's a relatively new term that was only coined during the latter part of the last decade. While it is still an enigma to many people since its introduction, it's become increasingly clear over the years that what is big data and its important to so many different firms. There's a lot of history to study and analyse and this analysis shows how big data has improved and evolved in a very short span of time and also hints at the probable changes and modifications that would come in the near future.

In this paper we conduct an in-depth review of the topic evolution of big data and identify how the usage and scope of big data evolved and changed overtime.

Keywords: Big Data, Evolution, Research, History, Technology, Data, Information.

INTRODUCTION

Big data means data sets which are very large and complicated to deal with by conventional data-processing softwares. Data with several fields or rows offer big statistical power, while data with higher complexity may lead to a higher false discovery rate. The challenges to big data analysis include data storage, capturing, sharing, transfer data and visualization, updating, querying, information privacy, data analysis. Big data has three key concepts: *variety*, *volume* and *velocity*. Therefore, it often includes data that go beyond the capacity of conventional softwares to analyze within the required time.

The past of big data starts years before the present buzz around Big Data. Seventy years ago the first ever attempt to quantify the increase rate of data in the terms of the quantity of data was encountered which is popularly known as "information explosion".

In 1944, Fremont Rider, speculated that by 2040 Yale Library will possess "approximately 200,000,000 volumes, which will occupy over 6,000 miles of shelves which would need cataloging staff of at least six thousand persons."

In 1980, the sociologist Charles Tilly wrote "none of the big questions has actually yielded to the bludgeoning of the big-data people." where the term big-data was referenced for the first time in his article "The old-new social history and the new old social history".

The evolution of Big Data can approximately be divided in three main phases. Every phase has its unique capabilities and characteristics.

First phase of BD

Big Data, data analysis and Data analysis originate from the well-established domain of database management. It depends heavily on extraction, storage and optimization techniques which are similar to data that resides in Relational Database Management Systems (RDBMS).

Data warehousing and database management are considered the vital components of first phase of Big Data. It provides the base for data analytics as we know it today, using established techniques such as online analytical processing, database queries and standard reporting tools.

Second phase of BD

From the early 2000s, the Web began to offer data analysis opportunities and unique data collections. With the growth in web traffic and online stores, firms such as Yahoo, Amazon and eBay started to inspect customer behavior by examining IP-specific location data, click-rates and search logs which opened a door for a newly discovered world of possibilities. From a data analytics, data analysis and Big Data point of view, HTTP-based web traffic initiated a humongous growth in unstructured and semi-structured data. Along with the standard structured data types, organizations now needed to find new techniques and storage methods to deal with the new data types in order to examine and analyze them effectively. The advent of social media data greatly aggravated the need for technologies, tools and analysis techniques that were effective enough to extract the required and meaningful information out from the unstructured data.

Third phase of BD

Although web-based unstructured content is still the main focus for many organizations in data analysis, data analytics, and big data, the current possibilities to retrieve valuable information are emerging out of mobile devices.



Mobile devices provided the possibility to analyze and store data based on location with the help of GPS. With the technological advancement of these mobiles and devices, its possible to analyze physical behavior, track movement and also health-related data. This data provides new opportunities from health care, to city design and transportation. Concurrently, with the advent of sensor-based internet-enabled devices the data generation is increasing exponentially. Millions of thermostats, TVs, refrigerators and even wearables are now generating zettabytes of data every day. And the race to extract valuable information from these new sources of data has just begun.

LITERATURE REVIEW

Daphne R. Raban et al. studied the evolution of Big Data (BD) and Data Science (DS) literatures and the relationship among them is analysed using bibliometric indicators which help establish the course taken by publications in these areas of research. They observed a massive surge in the publications of big data along with a gradual increase in Data Science publications.

Sasa Batistic et al. reviews this body of research, using three bibliometric methods. First, he explain its intellectual foundations with the help of co-citation analysis. Then, he envisioned the past evolution of BDA and performance research and its substreams through algorithmic historiography. Lastly, it provides insights into the potential evolution of field via bibliographic coupling.

Deepak Gupta et al. discusses the evolution of BD which include a bibliometric study of academic and industry publications of the period 2000–2017 corresponding to big data, popular open-source frameworks of big data stream processing and prevalent research challenges that are vital to be addressed to comprehend and exploit the true power of big data.

Anthony G. Picciano examined the evolving world of big data and analytics in American higher education. Specifically, he looked at the basic nature of these concepts, gave basic definitions, considered feasible applications, and lastly identified concerns about their implementation and growth.

D.P. Acharjya explored in his paper that the impact of big data challenges, open research issues, and various tools associated with it. And provides a platform to explore big data at several stages.

RESULT

We have analysed the topic big data and its evolution, which has recently captured lot of buzz due to its perceived opportunities and benefits. Appropriately, the literature in the context was reviewed to provide scrutiny & analysis of the topic big data and its evolution which is being researched and its importance in decision making. We reckon that big data analysis is of great importance in the present time of data over-flow, and can provide unanticipated benefits and insights in decision making in several areas.

REFERENCES:

1. https://www.researchgate.net/publication/338954307_The_evolution_of_data_science_and_big_data_research_A_bibliometric_analysis
2. <https://www.bigdataframework.org/short-history-of-big-data>
3. <https://techvidvan.com/tutorials/big-data-history/>
4. https://en.wikipedia.org/wiki/Big_data
5. <https://files.eric.ed.gov/fulltext/EJ982669.pdf>
6. <https://journals.sagepub.com/doi/abs/10.1177/0165551518789880>
7. <https://www.qubole.com/blog/big-data-evolution>
8. https://thesai.org/Downloads/Volume7No2/Paper_67A_Survey_on_Big_Data_Analytics_Challenges.pdf
9. <https://onlinelibrary.wiley.com/doi/full/10.1111/1467-8551.12340>