IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, June 2022
Impact Factor: 6.252

Importance of Big Data Analytics

Akshay Rajamony Asari
Student, Department of MCA
Late Bhausaheb Hiray S S Trust's Hiray Institute of Computer Application, Mumbai, India

Abstract: *Big Data Analytics is a topic fraught with both positive and negative potential. Big Data is defined not just by the amount of information involved but also its variety and complexity, as well as the speed with which it must be analyzed or delivered. The amount of data being produced is already incredibly great, and current developments suggest that this rate will only increase in the near future. Improved service should result as companies better understand their customers, but it is also possible that this data will create privacy problems. Thus, Big Data is important not only to students who hope to gain employment using these techniques and those who plan to use it for legitimate research, but also for everyone who will be living and working in the 21st Century.*

Keywords: Big Data Analytics

I. INTRODUCTION

Big data is a blanket term for the non-traditional strategies and technologies needed to gather, organize, process, and gather insights from large datasets. While the problem of working with data that exceeds the computing power or storage of a single computer is not new, the pervasiveness, scale, and value of this type of computing has greatly expanded in recent years.

The volume of data that one has to deal has exploded to unimaginable levels in the past decade, and at the same time, the price of data storage has systematically reduced. Private companies and research institutions capture terabytes of data about their users' interactions, business, social media, and also sensors from devices such as mobile phones and automobiles. The challenge of this era is to make sense of this sea of data.

1.1 Overview of Big Data Analytics

The volume of data that one has to deal has exploded to unimaginable levels in the past decade, and at the same time, the price of data storage has systematically reduced. Private companies and research institutions capture terabytes of data about their users' interactions, business, social media, and also sensors from devices such as mobile phones and automobiles. The challenge of this era is to make sense of this sea of data. This is where big data analytics comes into picture.

Big Data Analytics largely involves collecting data from different sources, munge it in a way that it becomes available to be consumed by analysts and finally deliver data products useful to the organization business.

The process of converting large amounts of unstructured raw data, retrieved from different sources to a data product useful for organizations forms the core of Big Data Analytics.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, June 2022
Impact Factor: 6.252

1.2 What is Big Data

Big data analytics is the process of examining large and varied data sets -- i.e., big data -- to uncover hidden patterns, unknown correlations, market trends, customer preferences and other useful information that can help organizations make more-informed business decisions.

An exact definition of "big data" is difficult to nail down because projects, vendors, practitioners, and business professionals use it quite differently. With that in mind, generally speaking, big data is: large datasets the category of computing strategies and technologies that are used to handle large datasets

In this context, "large dataset" means a dataset too large to reasonably process or store with traditional tooling or on a single computer. This means that the common scale of big datasets is constantly shifting and may vary significantly from organization to organization.

1.3 Applications of Big Data Analytics

Big data has increased the demand of information management specialists so much sothat SoftwareAG, Oracle Corporation, IBM, Microsoft, SAP, EMC, HP and Dell have spent more than $15 billion on software firms specializing in data management and analytics.

A. Government

The use and adoption of big data within governmental processes allows efficiencies in terms of cost, productivity, and innovation, but does not come without its flaws. Data analysis often requires multiple parts of government (central and local) to work in collaboration and create new and innovative processes to deliver the desired outcome. CRVS (Civil Registration and Vital Statistics) collects all certificates status from birth to death. CRVS is a source of big data for governments.

B. Manufacturing

Big data provides an infrastructure for transparency in manufacturing industry, which is the ability to unravel uncertainties such as inconsistent component performance and availability. Predictive manufacturing as an applicable approach toward near-zero downtime and transparency requires vast amount of data and advanced prediction tools for a systematic process of data into useful information.

C. Healthcare

Big data analytics has helped healthcare improve by providing personalized medicine and prescriptive analytics, clinical risk intervention and predictive analytics, waste and care variability reduction, automated external and internal reporting of patient data, standardized medical terms and patient registries and fragmented point solutions. Some areas of improvement are more aspirational than actually implemented.

D. Media

To understand how the media utilizes big data, it is first necessary to provide some context into the mechanism used for media process. Media and Advertising approach big data as many actionable points of information about millions of individuals.

E. Banking

Large amounts of information will be streaming in into banks, managing all this data and getting proper insights would be possible only with big data analytics. This is important to understand customers and boost their satisfaction, and also to minimize risk and fraud.

1.4 History of Big Data Analytics

1 C 18,000 BCE • Humans use tally sticks to record data for the first time. These are used to track trading activity and record inventory.

2 C 2400 BCE • The abacus is developed, and the first libraries are built in Babylonia.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, June 2022
Impact Factor: 6.252

3 300 BCE – 48 AD • The Library of Alexandria is the world's largest data storage center – until it is destroyed by the Romans.
4 100 AD – 200 AD • The Antikythera Mechanism – the first mechanical computer – is developed in Greece
5 1663 • John Graunt conducts the first recorded statistical-analysis experiments in an attempt to curb the spread of the bubonic plague in Europe
6 1865 • The term "business intelligence" is used by Richard Millar Devens in his Encyclopaedia of Commercial and Business Anecdotes
7 1881 • Herman Hollerith creates the Hollerith Tabulating Machine which uses punch cards to vastly reduce the workload of the US Census.
8 1926 • Nikola Tesla predicts that in the future, a man will be able to access and analyze vast amounts of data using a device small enough to fit in his pocket.
9 1928 • Fritz Pfleumer creates a method of storing data magnetically, which forms basis of modern digital data storage technology.
10 1944 • Fremont Rider speculates that Yale Library will contain 200 million books stored on 6,000 miles of shelves, by 2040.
11 1958 • Hans Peter Luhn defines Business Intelligence as "the ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal."
12 1965 • The US Government plans the world's first data center to store 742 million tax returns and 175 million sets of fingerprints on magnetic tape.
13 1970 • Relational Database model developed by IBM mathematician Edgar F Codd. The Hierarchal file system allows records to be accessed using a simple index system. This means anyone can use databases, not just computer scientists.
14 1976 • Material Requirements Planning (MRP) systems are commonly used in business. Computer and data storage is used for everyday routine tasks.
15 1989 • Early use of term Big Data in magazine article by fiction author Erik Larson – commenting on advertisers' use of data to target customers.
16 1991 • The birth of the internet. Anyone can now go online and upload their own data, or analyze data uploaded by other people.
17 1996 • The price of digital storage falls to the point where it is more cost-effective than paper.
18 1997 • Google launch their search engine which will quickly become the most popular in the world. • Michael Lesk estimates the digital universe is increasing tenfold in size every year.
19 1999 • First use of the term Big Data in an academic paper – Visually Exploring Gigabyte Datasets in Realtime (ACM) • First use of term Internet of Things, in a business presentation by Kevin Ashton to Procter and Gamble.
20 2001 • Three "Vs" of Big Data – Volume, Velocity, Variety – defined by Doug Laney
21 2005 • Hadoop – an open source Big Data framework now developed by Apache – is developed. • The birth of "Web 2.0 – the user- generated web".
22 2008 • Globally 9.57 zettabytes (9.57 trillion gigabytes) of information is processed by the world's CPUs. • An estimated 14.7 exabytes of new information is produced this year.

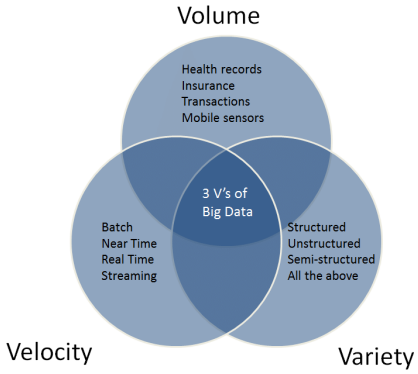
1.5 Characteristic of Big Data Analytics

The basic requirements for working with big data are the same as the requirements for working with datasets of any size. However, the massive scale, the speed of ingesting and processing, and the characteristics of the data that must be dealt with at each stage of the process present significant new challenges when designing solutions.

Volume

The sheer scale of the information processed helps define big data systems. These datasets can be orders of magnitude larger than traditional datasets, which demands more thought at each stage of the processing and storage life cycle.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, June 2022
Impact Factor: 6.252



Velocity

Another way in which big data differs significantly from other data systems is the speed that information moves through the system. Data is frequently flowing into the system from multiple sources and is often expected to be processed in real time

Variety

Big data problems are often unique because of the wide range of both the sources being processed and their relative quality.

Veracity

The variety of sources and the complexity of the processing can lead to challenges in evaluating the quality of the data (and consequently, the quality of the resulting analysis)

Variability

Variation in the data leads to wide variation in quality. Additional resources may be needed to identify, process, or filter low quality data to make it more useful.

Value

The ultimate challenge of big data is delivering value. Sometimes, the systems and processes in place are complex enough that using the data and extracting actual value can become difficult.

1.6 Why is Big Data Important

The importance of big data does not revolve around how much data a company has but how a company utilizes the collected data. Every company uses data in its own way; the more efficiently a company uses its data, the more potential it has to grow. The company can take data from any source and analyze it to find answers which will enable:

- **Cost Savings :** Some tools of Big Data l1ike Hadoop and Cloud-Based Analytics can bring cost advantages to business when large amounts of data are to be stored and these tools also help in identifying more efficient ways of doing business.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, June 2022
Impact Factor: 6.252

- **Time Reductions:** The high speed of tools like Hadoop and in-memory analytics can easily identify new sources of data which helps businesses analyzing data immediately and make quick decisions based on the learnings.

- **New Product Development :** By knowing the trends of customer needs and satisfaction through analytics you can create products according to the wants of customers.
- **Understand the Market Conditions:** By analyzing big data you can get a better understanding of current market conditions. For example, by analyzing customers' purchasing behaviors, a company can find out the products that are sold the most and produce products according to this trend. By this, it can get ahead of its competitors.
- **Control Online Reputation:** big data tool can do sentiment analysis. Therefore, you can get feedback about who is saying what about your company. If you want to monitor and improve the online presence of your business, then, big data tools can help in all this.

1.7 How it Works and Key Technologies

Data Management.

Data needs to be high quality and well-governed before it can be reliably analyzed. With data constantly flowing in and out of an organization, it's important to establish repeatable processes to build and maintain standards for data quality. Once data is reliable, organizations should establish a master data management program that gets the entire enterprise on the same page.

Data Mining

Data mining technology helps you examine large amounts of data to discover patterns in the data – and this information can be used for further analysis to help answer complex business questions. With data mining software, you can sift through all the chaotic and repetitive noise in data, pinpoint what's relevant, use that information to assess likely outcomes, and then accelerate the pace of making informed decisions.

Hadoop.

This open source software framework can store large amounts of data and run applications on clusters of commodity hardware. It has become a key technology to doing business due to the constant increase of data volumes and varieties, and its distributed computing model processes big data fast. An additional benefit is that Hadoop's open source framework is free and uses commodity hardware to store large quantities of data.

Predictive analytics.

Predictive analytics technology uses data, statistical algorithms and machine-learning techniques to identify the likelihood of future outcomes based on historical data. It's all about providing a best assessment on what will happen in the future, so organizations can feel more confident that they're making the best possible business decision. Some of the most common applications of predictive analytics include fraud detection, risk, operations and marketing.

Text mining.

With text mining technology, you can analyze text data from the web, comment fields, books and other text- based sources to uncover insights you hadn't noticed before. Text mining uses machine learning or natural language processing technology to comb through documents – emails, blogs, Twitter feeds, surveys, competitive intelligence and more – to help you analyze large amounts of information and discover new topics and term relationships.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, June 2022
Impact Factor: 6.252

II. BIG DATA SOLUTION

Case Studies

India

Big data analysis was tried out for the BJP to win the Indian General Election 2014.

The Indian government utilizes numerous techniques to ascertain how the Indian electorate is responding to government action, as well as ideas for policy augmentation.

Technology

eBay.com uses two data warehouses at 7.5 petabytes and 40PB as well as a 40PB Hadoop cluster for search, consumer recommendations, and merchandising.

- Amazon.com handles millions of back-end operations every day, as well as queries from more than half a million third- party sellers. The core technology that keeps Amazon running is Linux-based and as of 2005 they had the world's three largest Linux databases, with capacities of 7.8 TB, 18.5 TB, and 24.7 TB.
- Facebook handles 50 billion photos from its user base.
- Google was handling roughly 100 billion searches per month as of August 2012.

BIBLIOGRAPHY

[1]. www.tutorialspoint.com
[2]. www.webopedia.ieee.org
[3]. www.Wikipedia.com
[4]. www.techopedia.com
[5]. www.agiratech.com