International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 07 | July -2020 ISSN: 2582-3930

Big Data applications and challenges in India

AMAN KUMAR[1]

PG Scholar, Dept. of MCA, DSCE

Bangalore, India

SAMITHA KHAIYUM[2]

Assistant Professor, Dept. of MCA, DSCE

Bangalore, India

Abstract

Large volumes of data are generated due to the social media and other platforms. Most of the data are multimedia and are unstructured. The exponential growth of volume and quality of structured and unstructured type of datarefers to big data concepts. Managing and harnessing the benefits of big data is a real challenge. The various aspects that needs to be considered are the security and access control. The paper includes application of bigdata, opportunities it provides and challenges of managing large quantity of information for different types of applications.

Keywords: Big Data, Digital India, Governance, GFS, Analytics

1. Introduction

Big data refers to large amount of information collected and stored in either structured or unstructured form. The categories includes business transactions, photos, activity logs, data from social media, medical, banks, e-Governance services and others. If thisis effectively stored and analysed, will facilitate in knowledge visualisation of trends in business for numerous enterprises. The data from various sources would help the organizations to analyse the requirements of their customers and predict their preferences and demands.

The Digital Indiaprogram is introduced with the vision to remodel India into a digital society and data economy. Its vision is specifically targeted around these key areas:Infrastructure as a utility, Governance and services on demand and digital management of voters.

Fig-1: Unstructured Data

Big Data is characterised broadly speaking by five attributes: Volume, Velocity, Variety,Values &Complexity or Veracity.

Volume refers to the amount of data. It grows exponentially over an amount of time and might reach big. If we have a tendency to take all the info generated, constant quantity of information is probably going to be generated each minute within the returning days. Data tools use distributed systems in order tostore and analyse knowledge over databases that are are geographically scattered.

Velocity means the data that moves through the various sources like on-line systems,social media, internet click, and other mediums.

Variety means the data from several sources. Data are structured, semi-structured and unstructured (e-mails, text messages, etc. Most of the data generated through sources like Social Media are unstructured (text, images, video, voice, etc.). Big Data technology facilitates integration of various kinds of data like messages, social media conversations, photos and video or voice recordings.

Veracity implies that information should be ready to traverse through multiple data centres, the cloud and geographical zones. Handling, managing, analysing such huge amount of data is very complicated.

Value refers to measure of tangible or intangible advantages that an organisation could gain from data. Having access to data isn't any sensible unless we are able to flip it into worth value.

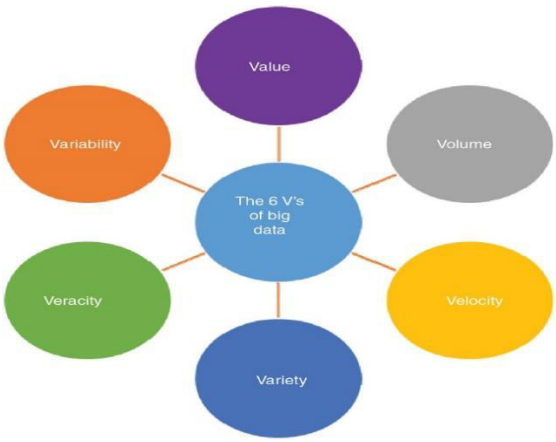


Fig-2: Big Data Characteristics

2. Body

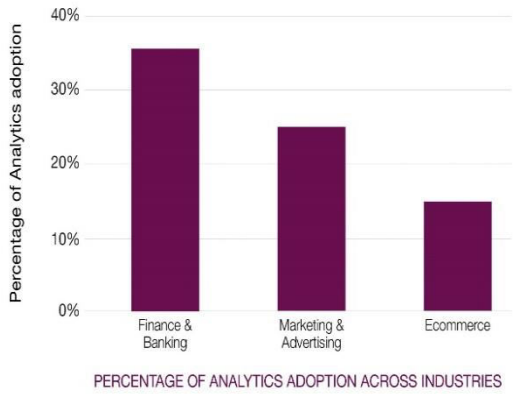
I. Review

About 90% of today's World data has been created in the last two years. Huge volumes of data are being generated everyday, unlocking various new possibilities for different stakeholders in the global ecosystem.

In India, there have been initiatives such as Digital India, Aadhaar, 100 smart cities missionand MyGov being implemented to enable participation of citizens in view of good governance. The data which is generated from these initiatives, they are combined with other sources which then leads to a boom in the generation of digitized data and also the subsequent implementation of analytics in the country.

The scale of adoption has been huge and it is robustly progressed to become atool for every business across various industries in the country. The domestic analytics market in India is estimated to be $3.03 Billion in 2020 and is also projected as seven times growth in the next seven years, according to a recent study. Big Data is important for India, with its vast pool of IT talent and the rapidly upgrading industries, which would help it become a model nation for enabling successful data-based decisions.

For 'Digital India' vision, the Indian Government has undertaken data-intensive initiatives and various projects in the public sectors. This ranges from agriculture, trade, and transport services to banking, policy-making, and others. The 2017 'Project Insight' stands as a success story of leveraging data mining techniques and its subsequent analysis to catch tax evaders and procure information on black money. By utilising Big Data and AI, the government was able to flag various accounts with sizable black money deposits and also to find patterns and trends to flag other bank accounts.



PERCENTAGE OF ANALYTICS ADOPTION ACROSS INDUSTRIES

II. Applications

Government

Big Data has proved to be very useful within the government sector. It had played an important role in not just the re-election campaign of Barack Obama but also the BJP and its allies to win Indian election 2014. Various techniques were employed by the Indian Government to seek out how the Indian electorate is responding to the actions taken and concepts for policy growth introduced by the govt.

Fraud Detection

One of the foremost compelling samples of Big Data application is in Fraud Detection, especially within the businesses which involve any claims or transaction processing. In most of the cases, fraud is discovered long after it's done, which suggests the damage is already done. Analysing requests and transactions in real time, detecting anomalous behaviour or identifying large-scale patterns across transactions through Big Data applications can definitely convince be a game changer in Fraud Detection.

Agriculture

Sensor data to optimise crop efficiency is employed by a biotechnological firm. This is often wont to measure how plants react to changes in various conditions by planting test crops and running simulations. The data about changes in attributes like temperature, water level, growth, gene sequencing and output of plant is being recorded. Such data helps in discovering the optimal condition for the precise gene types.

Marketing

To perform the research, one among the tools used was "a system to analyse facial expressions revealing viewers' feelings". Similarly, other markets also are using face recognition software to understand how well their advertising have succeeded or failed at generating interest in their product/service.

Healthcare

Usually, the health industry had lagged when it came to the utilization of Big Data. One among the foremost critical reasons behind it's the resistance to vary providers being familiar with making treatment decision autonomously, using their clinical judgement, rather than counting on protocols supported Big Data.

The organisations like single hospital (not a multi-speciality hospital) or pharmaceutical companies lack procedure for incorporating data and communicating findings. The organisations are switching to Big Data applications for analysing the info to get more action-oriented results.

Government Initiatives & Interventions

NITI Aayog is working on a plan in collaboration with the private players to develop the 'National Data & Analytics Platform', this will act as a single source of data for citizens, policymakers, and researchers, etc.

Also Ministry of Statistics and Programme Implementation has proposed to establish a 'National Data Warehouse on Official Statistics'. This would leverage technology to implement the use big data analytical tools for improving the quality of macro-economic statistics.

Use of Direct Benefit Transfer in MGNREGA & Aadhaar for authentication& availing welfare scheme benefits helps in the filtering of duplicate beneficiaries.

Ministry of Agriculture has signed a deal with the ISRO for the use of satellites for mapping of agricultural assets.

Smart City Mission, Digital India and digital economy proposals like BHIM app, are some importantgovernment initiatives that are using Big Data for achieving better governance in the country.[2]

III. Challenges

Big data has created a revolution in the field of IT, but a key question, about localization, remains unanswered that should the data of Indian consumers be stored in India only or not. Local storage of data would ensure the availability of information for legal reasons.

One of the challenges is invasion of privacy. Big Data analytics leads to multiple problems which are data privacy and net neutrality.

The other challenge is the data security. There have been several incidents of leakage of Aadhaar data which points to theneed for the government to increase the security and safety of virtual data it collects from its citizens.

There are various technical challenges as well. Big data involves several inherent limitations like, inefficient infrastructure for data collection & management of such humongous data, storage problems, computational problems, streaming issues and scalability.

There are various challenges related to governance of data. The government needs to adopt a consistent dynamic approach for effective policymaking using Big Data. There is a potential of revealing sensitive personal information.

Outsourcing of data analysis mainly increases the chances of security risks because the information like customers' earnings, savings and insurance policies are shared. There is always a concern about these data ending up in the wrong hands, and that is what discourages customers from sharing their personal information in return for customised offers

IV. Recommendations

The large scale and variety of data is to be absorbed and integrated in such a way that it is accessible globally. As the data is to be incorporated from diverse sources into an analytical platform, maintenance of the concurrency and consistency of data is to be handled properly. It is required to enable high-performance analytics so that used adequately by all different enterprises. Advance analytics helps to find out possible solution for the various challenges and to opportunities, new business facts that no one in the enterprise knew before. The various technology and tools organizations can use are predictive analytics, discovery analytics, statistical analytics, machine learning, data visualization and natural language processing.

Streaming algorithm can be used as an alternative model for dealing with large volumes of data with limited computational and storage resources. Learning algorithm based on linear classifier and multiclass classifiers can be used to design logistic regression model to make more accurate parameters. Multiclass classifiers based model can also be designed to do fraud detections in UIDAI scheme. The Map Reduce derives from providing an abstraction that allows

developers to harness power of large clusters but abstraction manage complexity by hiding details and presenting required features to users.

The Ethical issues related to data privacy need to get addressed in the current scenario.

- Data protection law should incorporate with some of the best practices followed over the World.
- It must strengthen the cybersecurity to safely utilize the large pool of virtually available data.

Infrastructural Requirement is one of the important aspect. To effectively analyse the large volumes of available data, well-equipped data centers are needed to be set up. It is essential to segregate the required relevant data from the irrelevant data pool.

3. Conclusion

This is the era ofBig Data. By doing better analysis of the massive volumes of data there's the potential for creating quicker advances in several scientific disciplines and rising the gain and success of the many enterprises.There are several challenges to be addressed. The various challenges are not just the apparent problems with scale, but also lack of structure, privacy, timeliness, and visual image, in the least stages of the analysis from information acquisition to results. Challenges are common across various application domains,and not cost-efficient to be handled.To achieve benefits of Big Data the technical challenges needs to be addressed. The Big Data technologies will further help conceive novel business models.

Big Data is a powerful tool that makes things easier in various fields. Apart from various mentioned sectors, Big Data application would also help in making a positive contribution in various fields.

References

[1].https://www.drishtiias.com/to-the-points/paper3/big-data-1

[2].https://globalstatement2015.wordpress.com/2015/08/12/big-data-in-india-the-governmental-intervention/

[3].https://inc42.com/resources/the-elemental-role-of-data-analytics-in-the-making-of-a-truly-digital-india/

[4].https://abhipedia.abhimanu.com/Article/IAS/NDEwOAEEQQVVEEQQVV/Scope-of-Big-data-in-Governance-Governance-Issues-IAS

[5].https://www.allerin.com/blog/growth-of-big-data-industry-in-india

[6].https://www.itproportal.com/features/how-big-data-will-impact-different-sectors/