Bigdata and It's Applications

Centre of Excellence in IT PNG

October 2019

1 Introduction to BigData

There is no place where Big Data does not exist!

The curiosity about what is Big Data has been soaring in the past few years. Let me tell you some mind-boggling facts! Forbes reports that every minute, users watch 4.15 million YouTube videos, send 456,000 tweets on Twitter, post 46,740 photos on Instagram and there are 510,000 comments posted and 293,000 statuses updated on Facebook!

Just imagine the huge chunk of data that is produced with such activities. This constant creation of data using social media, business applications, telecom and various other domains is leading to the formation of Big Data.

Big Data refers to the large amounts of data which is pouring in from various data sources and has different formats. Big Data is much more than a collection of datasets with different formats, it is an important asset which can be used to obtain enumerable benefits.

The major sources of Big Data are social media sites, sensor networks, digital images/videos, cell phones, purchase transaction records, web logs, medical records, archives, military surveillance, eCommerce, complex scientific research and so on. All these information amounts to around some Quintillion bytes of data. By 2020, the data volumes will be around 40 Zettabytes which is equivalent to adding every single grain of sand on the planet multiplied by seventy-five.

2 Big Data Analytics

Big data analytics examines large and different types of data to uncover hidden patterns, correlations and other insights

Now that I have told you what is Big Data and how it's being generated exponentially, let me present to you a very interesting example of how Starbucks, one of the leading coffeehouse chain is making use of this Big Data.

I came across this article by Forbes which reported how Starbucks made use of Big Data to analyse the preferences of their customers to enhance and personalize their experience. They analysed their member's coffee buying habits along with their preferred drinks to what time of day they are usually ordering. So, even when people visit a "new" Starbucks location, that store's point-of-sale system is able to identify the customer through their smartphone and give the barista their preferred order. In addition, based on ordering preferences, their app will suggest new products that the customers might be interested in trying. This my friends is what we call Big Data Analytics.

Basically, Big Data Analytics is largely used by companies to facilitate their growth and development. This majorly involves applying various data mining algorithms on the given set of data, which will then aid them in better decision making.

2.1 Big Data Applications

These are some of the following domains where Big Data Applications has been revolutionized:

1. Entertainment: Netflix and Amazon use Big Data to make shows and movie recommendations to their users.
2. Insurance: Uses Big data to predict illness, accidents and price their products accordingly.
3. Driver-less Cars: Google's driver-less cars collect about one gigabyte of data per second. These experiments require more and more data for their successful execution.
4. Education: Opting for big data powered technology as a learning tool instead of traditional lecture methods, which enhanced the learning of students as well aided the teacher to track their performance better.
5. Automobile: Rolls Royce has embraced Big Data by fitting hundreds of sensors into its engines and propulsion systems, which record every tiny detail about their operation. The changes in data in real-time are reported to engineers who will decide the best course of action such as scheduling maintenance or dispatching engineering teams should the problem require it.
6. Government: A very interesting use of Big Data is in the field of politics to analyse patterns and influence election results. Cambridge Analytica Ltd. is one such organisation which completely drives on data to change audience behaviour and plays a major role in the electoral process.
7. Smarter Healthcare: Making use of the petabytes of patient's data, the organization can extract meaningful information and then build applications that can predict the patient's deteriorating condition in advance.
8. Telecom: Telecom sectors collects information, analyzes it and provide solutions to different problems. By using Big Data applications, telecom companies have been able to significantly reduce data packet loss, which occurs when networks are overloaded, and thus, providing a seamless connection to their customers.

9. Retail: Retail has some of the tightest margins, and is one of the greatest beneficiaries of big data. The beauty of using big data in retail is to understand consumer behavior. Amazon's recommendation engine provides suggestion based on the browsing history of the consumer.
10. Traffic control: Traffic congestion is a major challenge for many cities globally. Effective use of data and sensors will be key to managing traffic better as cities become increasingly densely populated.
11. Manufacturing: Analyzing big data in the manufacturing industry can reduce component defects, improve product quality, increase efficiency, and save time and money.
12. Search Quality: Every time we are extracting information from google, we are simultaneously generating data for it. Google stores this data and uses it to improve its search quality.
13. Finance: Banks and financial services firms use analytics to differentiate fraudulent interactions from legitimate business transactions. The analytics systems suggest immediate actions, such as blocking irregular transactions, which stops fraud before it occurs and improves profitability.

3 Big Data Applications

The primary goal of Big Data applications is to help companies make more informative business decisions by analyzing large volumes of data. It could include web server logs, Internet click stream data, social media content and activity reports, text from customer emails, mobile phone call details and machine data captured by multiple sensors

Organisations from different domain are investing in Big Data applications, for examining large data sets to uncover all hidden patterns, unknown correlations, market trends, customer preferences and other useful business information.

3.1 Healthcare

The level of data generated within healthcare systems is not trivial. Traditionally, the health care industry lagged in using Big Data, because of limited ability to standardize and consolidate data.

But now Big data analytics have improved healthcare by providing personalized medicine and prescriptive analytics. Researchers are mining the data to see what treatments are more effective for particular conditions, identify patterns related to drug side effects, and gains other important information that can help patients and reduce costs.

With the added adoption of mHealth, eHealth and wearable technologies the volume of data is increasing at an exponential rate. This includes electronic health record data, imaging data, patient generated data, sensor data, and other forms of data.

By mapping healthcare data with geographical data sets, it's possible to predict disease that will escalate in specific areas. Based of predictions, it's easier to strategize diagnostics and plan for stocking serums and vaccines.

3.2 Manufacturing

Predictive manufacturing provides near-zero downtime and transparency. It requires an enormous amount of data and advanced prediction tools for a systematic process of data into useful information.

3.3 Media & Entertainment

Various companies in the media and entertainment industry are facing new business models, for the way they – create, market and distribute their content. This is happening because of current consumer's search and the requirement of accessing content anywhere, any time, on any device.

Big Data provides actionable points of information about millions of individuals. Now, publishing environments are tailoring advertisements and content to appeal consumers. These insights are gathered through various data-mining activities. Big Data applications benefits media and entertainment industry by:

- Predicting what the audience wants
- Scheduling optimization
- Increasing acquisition and retention
- Ad targeting
- Content monetization and new product development

3.4 Internet of Things (IoT)

Data extracted from IoT devices provides a mapping of device inter-connectivity. Such mappings have been used by various companies and governments to increase efficiency. IoT is also increasingly adopted as a means of gathering sensory data, and this sensory data is used in medical and manufacturing contexts

3.5 Government

The use and adoption of Big Data within governmental processes allows efficiencies in terms of cost, productivity, and innovation. In government use cases, the same data sets are often applied across multiple applications & it requires multiple departments to work in collaboration. Since Government majorly acts in all the domains, thus it plays an important role in innovating Big Data applications in each and every domain. Let me address some of the major areas:

Cyber security & Intelligence The federal government launched a cyber security research and development plan that relies on the ability to analyze large data sets in order to improve the security of U.S. computer networks.

The National Geospatial-Intelligence Agency is creating a "Map of the World" that can gather and analyze data from a wide variety of sources such as satellite and social media data. It contains a variety of data from classified, unclassified, and top-secret networks.

Crime Prediction and Prevention Police departments can leverage advanced, real-time analytics to provide actionable intelligence that can be used to understand criminal behaviour, identify crime/incident patterns, and uncover location-based threats.

Pharmaceutical Drug Evaluation According to a McKinsey report, Big Data technologies could reduce research and development costs for pharmaceutical makers by 40*billionto*70 billion. The FDA and NIH use Big Data technologies to access large amounts of data to evaluate drugs and treatment.

Scientific Research The National Science Foundation has initiated a long-term plan to:

Implement new methods for deriving knowledge from data Develop new approaches to education Create a new infrastructure to "manage, curate, and serve data to communities".

Weather Forecasting The NOAA (National Oceanic and Atmospheric Administration) gathers data every minute of every day from land, sea, and space-based sensors. Daily NOAA uses Big Data to analyze and extract value from over 20 terabytes of data.

Tax Compliance Big Data Applications can be used by tax organizations to analyze both unstructured and structured data from a variety of sources in order to identify suspicious behavior and multiple identities. This would help in tax fraud identification.

Traffic Optimization Big Data helps in aggregating real-time traffic data gathered from road sensors, GPS devices and video cameras. The potential traffic problems in dense areas can be prevented by adjusting public transportation routes in real time.