APPLICATIONS OF BIG DATA ANALYTICS

Kaveesh Nayak

Research Scholar

Punjabi University,

Patiala, Punjab

Abstract : Every day we see a huge amount of data being generated in the form of social media uploads, classifieds, government records and many other forms. We call this data as "Big Data". It is called "Big Data" because of its three characteristics "Volume, Veracity, Velocity". We know that most of the social media updates are in multimedia format, classifieds are uploaded in digitized document formats and there are so many other sources which generate a huge amount of data. So most of this data is unstructured, semi structured or quasi structured. "Big Data Analytics" is the process of analysing the Big Data to extract useful information from it. In this paper, we try to highlight the various applications of Big Data Analytics.

***IndexTerms* - Big Data, Big Data Applications, Big Data in Healthcare, Social Media.**

I. INTRODUCTION

Most of the data we see now days is generated in huge amount and that too at a high speed. Every day approximately 2.5 quintillion bytes of data is generated both in structured as well as unstructured forms. With the help of this information we can predict how massive can data sets be in coming years. This huge amount of data that we would be dealing with revolves around the concept of "Big Data".

Big Data refers to the collection of data that is too massive that traditional data management techniques cannot handle. This data can come from sources like sensors, click streams, multimedia devices and many others. The rate at which such a huge data is generated requires not only an efficient processing but also requires a quick method to access it. In Real-Time systems timely operation is of utmost importance. So if a huge amount of data is to be handled by Real Time applications not only should it be handled efficiently but quickly also.

There is no benefit in collecting and storing a huge amount of data if no analysis is possible on it. To perform analysis the terms "Big Data" and "Analytics" were combined thus forming the term "Big Data Analytics". With Big Data Analytics one can analyze huge volumes of data that is beyond the scope of conventional analytics. Even the most complex of problems can be efficiently solved by Big Data Analytics. Techniques like predictive analysis, text mining and forecasting helps in discovering new patterns and finding more than one solution to problems. Big Data Analytics can help discover new business opportunities, prospective customers and can help in understanding sales seasonality etc. Apart from business analysis Big Data Analytics has many other applications. Some of these applications have been described in this paper.

II. SOME COMMON TECHNIQUES USED IN BIG DATA ANALYTICS

A. Data Fusion

Combining and integrating data from several sources is known as Data Fusion. Data fusion is done with the aim of creating new insights. Data fusion is widely used in geospatial domain as it includes huge data sets from various sources which need to be combined to create new insights.

B. Association Rule Learning

Association Rule Learning refers to the method for discovering relation between objects (variables) that most of the times seem to be unrelated. This technique is mostly used by point-of-sale (POS) systems to discover regularities in data that includes different items. Various rules are used in this method to find the regularities. Now days Association Rule Mining is also being used in many application areas like Web usage mining and Bioinformatics. Some of the popular algorithms used for generating association rules are: Apriori algorithm, FP- growth algorithm and Eclat algorithm.

C. Decision Tree Learning

This technique makes use of the decision trees to extract the knowledge out of Big Data. One strategy of applying decision tree learning to Big Data could be to break large data set into a number of partitions then perform Decision Tree Learning on each of them. After doing learning on each of the partition, next task is to combine them to generate a single solution. This techniques is helpful in classification and regression.

D. Genetic Algorithms

These algorithms work on the principle of evolution including mechanisms like inheritance, mutation and natural selection to solve a problem. With data changing everyday Genetic Algorithms help in doing adaptive analysis.

E. Sentiment Analysis

Sentiment Analysis helps in determining the sentiments of and individual regarding a topic. It is done mostly with the help of Natural Language Processing and Text Analysis. Both structured and unstructured data is stored using Big Data techniques and then a learning and reporting engine is designed which gives out the result as sentiments of different individuals. The result is often converted into a visual form like bar charts, graphs etc.

F. Regression Analysis

This technique involves finding out the relationship between dependent and independent variables. This technique helps in prediction and forecasting. Probabilistic methods like Bayesian method is used for Regression Analysis.

G. Machine Learning

It deals with designing algorithms that can learn and make predictions about data. It works well with both structured and unstructured data. This technique is widely used in applications like content recommendation; spam filtering and computer vision.

III. APPLICATION OF BIG DATA ANALYTICS

A. Healthcare

Around 25,000 petabytes of healthcare data is expected to be generated by 2020. Medical and health care apps and remote health monitoring devices are getting popular day by day. For example consider an application software keeping track of water consumption of an individual in a day. Consider this software to be used by millions of people around the globe. Huge amount of data generated by this software if properly analyzed using Big Data Analytics, can help in making some useful predictions about diseases related to deficiency of water. Most of the data generated in clinics and hospitals is unstructured data: hospital admission and discharge notes, paper prescriptions, notes that are handwritten by doctors and nurses, MRI, CT and other images. Real-time data analytics can prove to be useful in ICUs in hospitals. Big Data Analytics can help in analyzing disease patterns and can help track outbreaks. Analysis of large data sets can help match the right medicine to the right patient at the right time. "Google Flu Trends" is a useful web service by Google to gather data from an individual and predicts the likeliness of flu in an individual. "Nextbio" is another application software used in healthcare based on Big Data Analytics. It uses Big Data technology to make correlations between data points from public domain with private clinical data sets. "Predixion Software" used in health care domain makes use of Big Data Analytics, Data Mining and Machine Learning to find out and explain patterns in datasets of hospital. This software helps in reducing readmissions in hospitals and in prevention of disease like diabetes. There are many other application softwares used in healthcare which make use of Big Data Analytics; "Practice Fusion" and "iBlueButton" are some of them.

B. Financial Sector

Big Data Analytics can be used in financial sector to generate new insights. Industries generate a large chunk of data. These industries can earn higher profit by analyzing the data generated by them.

Some useful functions that can be performed in financial sector with the help of Big Data Analytics are:

(i) Analysis of Customer Segmentation: Segmentation of customers in a business is done with the aim of maximizing the value of each customer to the business. This is done by dividing the customer base in an organization. With the help of Big Data Analytics customer segmentation can be done from the data generated by sources like call centre interactions and website activity.

(ii) Detection of Fraud: All transactions taking place in an organization can be protected by using Big Data Analytics. Any fraudulent activity that can lead to loss in business can be tracked in real-time.

(iii) Recommendation of Investments and Stocks: With Big Data Analytics a huge amount of unstructured data can be used to identify the best companies to invest in. Stock market traders and investment portfolio managers can analyze this Big Data to compare among many companies and then deciding which one to invest in. Price list changes, product reviews, company news contributes to most of the data. Mathematical models are used to check beneficial stocks.

(iv) Credit Risk Assessment: Big Data Analytics can help in analyzing customer credit score accurately. This helps in answering the question how secure it is to lend the money to this individual.

(v) Customer Retention: Customers' requirements can be found out by tracking their data regarding behavior on social networks, website activities. With this tracking of data better customer services can be provided and this data analysis can also help in customer retention. "Open Text Actuate Big Data Analysis" is an application software mainly built for business and financial applications. It does not require any coding skills and can be used by business persons without an IT support.

C. Education

Big Data Analysis can provide immediate feedback to students and teachers about academic performance. With such analysis teachers can discover patterns regarding student outcomes such as dropping out, needing extra help. Teachers can use data from all the schools in a state and can analyze which subject requires more attention and then teachers try applying new methods to improve performance. After this analysis feedback is given by students which again helps in checking effectiveness of the new method applied. Pre-test and Post-test assessment can be done by teachers easily using analysis of big data. Different methods implemented by students to solve problems can be checked by teachers using analysis and teachers can identify the creative students.

"Tableau.com" provides data analysis application software for analysis in educational sector. It includes a dashboard for both teachers and students; for students to check their performance and compare it, and for teachers to find patterns among students' performance.

D. Government Sector

We know that organizations use Big Data Analytics to make predictions and generate new insights. Same way even the governments have an enormous amount of data (PAN Card details, Voter IDs etc.) which if analyzed properly can help them in framing new policies and implementing them effectively. In USA Data Analytics played a major role in presidential elections. President Barack Obama analyzed various activities of people and they started receiving emails regarding policies to be framed, campaigning information and information regarding the facilities they would get after the elections.

Big Data Analytics can be used by governments to minimize tax fraud and social security fraud by citizens. Irregular behavior in patterns while analyzing the data. For example checking if someone who has applied for social security is really allowed to receive it or he is just for applying for the sake of it.

With the help of Big Data Analytics government can monitor malicious activities that can cause threats like terrorist attacks and hijacks. For example PRISM, a surveillance program operated by United States National Security Agency used data mining and analysis to track terrorist and other criminal activities.

E. Telecom Domain

Telecom sector generates a large amount of data in the form of "Call Detail Record" everyday and this data needs efficient processing. Call Detail Record is known as the data that provides details of a telephone call or other communication transactions. These CDRs are analyzed by telecom companies and this analysis needs to be done in real time to prevent dropped calls and poor sound quality.

Machine Learning implemented with Big Data Analytics can aid in providing optimal schedule for maintenance of the backbone (Transmission towers and their related connections) and can be useful for setting up new towers. Data Analysis also helps in allocating right bandwidth at right time.

F. Social Media

The most common use of Big Data Analytics is to observe opinions of people regarding a subject or a thing. Due to the boom of Social Media, lots of status updates and tweets are uploaded every day. With the emergence of "hashtags" it is easy to predict the mood of an individual, how he or she feels about a certain thing. If concentrated on the online activities of individuals, a pool of data can be obtained out of which relevant information can be extracted.

Being a critical part of the web social media is of interest to almost all applications worldwide and it is widely used for research purposes also. Businesses use social media data as a rich pool of information which if tapped can help them in product design, management of customer relations and marketing. Similarly politicians and governments use social media as a good medium to monitor the effectiveness of the new policies launched by them and can help them gather public support. Terrorist activities can be tracked on social media by security and intelligence department using social media analytics. Application software like "Google Analytics" is used for social media analytics.

G. Network Security

Network monitoring and security can be done by big data analytics. As we know network security is of primary concern in every enterprise, so network administrators can use Big Data Analytics to quickly detect and remove threats with the aim of securing the network and their customers' data.

H. Sports

With athletes being continuously monitored, we can expect a huge amount of data to be generated. Drug testing kits, digital health bands, smart watches and all such devices used by sportspersons and athletes generate data which if analyzed can provide them with the measures that can be taken to improve performance. Factors like preparation, injury prevention and rehabilitation can be improved by the approach of Big Data Analytics.

I. Gaming Industry

We see nowadays that video gaming has come a long way since its emergence, with gaming consoles implementing artificial intelligence and gaming on smart phones has increased a lot. With this revolution in gaming industry, it is natural that data will also increase which leads to large game data, player data, and session data. This huge data being generated is analyzed by gaming studios to enhance gaming experience, send notification to gamers and generate score boards that increases player retention.

IV. CHALLENGES IN BIG DATA ANALYTICS

A. Integration of Data: Big Data consists of data from various sources which includes structured as well as unstructured data. The challenge is to integrate or combine this data quickly and efficiently.

B. Scaling Issues: Data volume is increasing day by day but computer resources and CPU speeds are same as before which is a bottleneck.

C. Privacy: Data from social networking websites is continuously being analyzed since the advent of Social Media Analytics. This poses as a threat to privacy of individuals.

V. CONCLUSION

A. Integration of Data: Big Data consists of data from various sources which includes structured as well as unstructured data. The challenge is to integrate or combine this data quickly and efficiently.

B. Scaling Issues: Data volume is increasing day by day but computer resources and CPU speeds are same as before which is a bottleneck.

REFERENCES

[1] Ibm.com, 'IBM Big Data – What is Big Data – United States. [Online]. Available: http://www.ibm.com/big-data/us/en/. [Accessed: 08- Aug- 2015]

[2] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Richard Dobbs, C. Roxburgh and A. Byers, 'Big data: The next frontier for innovation, competition, and productivity', *Mckinsey.com*, 2011. [Online]. Available: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation. [Accessed: 10- Aug- 2015].

[3] P. Duggal and S. Paul, 'Big Data Analysis: Challenges and Solutions', *International Conference on Cloud, Big Data and Trust 2013*, pp. 269-276, 2013.

[4] S. Kar, 'Big Data Techniques for Analyzing Large Data Sets - Infographic | CloudTimes', *Cloudtimes.org*, 2013. [Online]. Available: http://cloudtimes.org/2013/06/13/big-data-techniques-for-analyzing-large-data-sets-infographic/. [Accessed: 11- Aug- 2015].

[5] V. Upadhyay and I. Shaikh, 'Big Data Analytics', *IJARCSSE*, vol. 5, no. 6, pp. 1076-1082, 2015.

[6] A. Katal, M. Wazid and R. Goudar, 'Big data: Issues, challenges, tools and Good practices', *2013 Sixth International Conference on Contemporary Computing (IC3)*, pp. 404-409, 2013.

[7] S. Madden, 'From Databases to Big Data', *IEEE Internet Computing*, no. 3, pp. 4-6, 2012.

[8] R. Rea, 'Real-time Big Data Analytics for Telecommunications | The Big Data Hub', *Ibmbigdatahub.com*, 2013. [Online]. Available: http://www.ibmbigdatahub.com/blog/real-time-big-data-analytics-telecommunications. [Accessed: 17- Aug- 2015].

[9] D. Klobucher, 'Sentiment Analysis with Big Data - SAP News Center', *SAP News Center*, 2013. [Online]. Available: http://news.sap.com/sentiment-analysis-with-big-data/. [Accessed: 17- Aug- 2015].

[10] W. Raghupati and V. Raghupati, 'Big data analytics in healthcare: promise and potential', *Health Information Science and Systems*, vol. 2, 2014.

[11] H. Parmar, 'Applications of Big Data in Government Sector', in *International Conference on Big Data Management & Cloud Computing*, Bhubaneswar, 2013, pp. 29-31.