REVIEW PAPER ON
BIG DATA: CHALLENGES AND APPLICATIONS

Priya Parhate[1], Gaurav Ghogle[2], Jyoti Bhange[3], Ashwini Ingle[4]

[1]Professor, Dept. of Computer Science & Engineering, STC, SPRT, Shegaon, Maharashtra, India
[2]Diploma Student, Dept. of CSE, STC, SPRT, Shegaon, Maharashtra, India
[3]Diploma Student, Dept. of CSE, STC, SPRT, Shegaon, Maharashtra, India
[4]Diploma Student, Dept. of CSE, STC, SPRT, Shegaon, Maharashtra, India

---***---

Abstract - *Big data is a new curve for IT today and a main buzz phrase. Big data is driven data with high velocity, variety, volume, value and veracity. It comes from different sources like internet, sensors, mobile devices, social media, machine-generated data and other geospatial devices. Traditional data processing and analysis of structured data using relational database management systems (RDBMS) and data warehousing no longer satisfy the challenges of Big Data. Due to the big data's high volume and velocity, the effective solution is to store the big data in cloud, because it has capability to store and process huge amount of big data. Cloud computing also offers the implementation of big data in small sized and medium sized businesses. This paper presents big data's various real time applications that include healthcare, networking security, sports, market and business, gaming, education system, and telecommunications. The interest in big data has been ever-increasing due to its rapid growth and it covers diverse areas of applications. Hence, there seems to be a need for an analytical review of recent developments in the big data technology. This paper provides a comprehensive review of the big data state of the art, conceptual explorations, challenges and applications. In addition to that, several future directions for big data research are highlighted [1].*

1. INTRODUCTION

People and devices are constantly generating data. In the IT industry, big data is a main largest buzz phase. The data has increased day by day from last 20 years; some facts about data are 100 million emails, 2 million searching queries on Google, 277000 tweets, 350 GB data processing on Facebook per minute [2]. Big data is a new opportunities for enterprise to extract huge volume of data rational, non-rational, and real time data types. Personal communication and new technologies producing the big data trends, the global internet population from 2010 to 2011 have been grew by 6.5% [3]. In 2013, estimates reached 4 ZB of data generated worldwide. Big data describes any voluminous structured (files, array, records, tree, table), unstructured can be textual (Word document, PowerPoint, instant messages, Email messages) and non-textual (MP3 audio files, JPEG, flash video files) and semi-structured (weblogs, social media feeds) [4]. In 2014, estimated worldwide data at a staggering is 7ZB. Today 2 billion people are connected together and generating massive amount of data every second and IDC study found that, data volumes are expected to increase 50 times by 2020. Cloud computing enabled with big data by features such as pay-per-use, elasticity, low time to market, transfer of risks and low upfront investment.

2. FIVE V'S OF BIG DATA

The properties associated with big data are Volume, Variety, Velocity, Variability and Value [1].

2.1 Volume

The volume of big data is constantly increasing day by day. The data accumulated through sensor networks and social websites rising from petabytes to Zetabytes.

2.2 Variety

Data produced are from different categories, consists of standard, unstructured, semi structured and raw data which are very difficult to be handled by traditional systems.

2.3 Velocity

It indicates the speed at which the data generated and become historical. Big data is able to handle the outgoing and incoming data rapidly.

2.4 Variability

Variability describes the amount of variance used in summaries kept within the data bank and refers how they are closely clustered or spread out within the data set.

2.5 Value

All e-commerce systems and enterprises are enthusiastic in improving the relationship with customers by providing value added services. For that, study on customer trends and attitudes in the market are to be analyzed. Users can also query the data store to find out business trends and accordingly they can change their master plan or strategies. By making big data open to all, it creates functional analysis

transparency. Supporting real time decisions and experimental analysis in different locations, datasets can do admirable things for enterprises.

3. BIG DATA CHALLENGES

3.1 Volume

The challenge is how to deal with the size of Big Data.

3.2 Variety

It means combining Multiple Data Sets. The challenge is how to handle multiplicity of sources, types, and formats.

3.3 Velocity

One of the important challenges is how to react to the flood of information in the time required by the application.

3.4 Veracity

There are several challenges:
- How can we cope with imprecision, uncertainty, missing values, misstatements or untruths?
- How good is the data? How broad is the coverage?
- How fine is the sampling resolution? How periodically are the readings?
- How well understood are the sampling biases?
- Is there data available, at all?

3.5 Data Discovery

This is a huge challenge to find out high-quality data from the vast collections of data that are out there on the Web.

3.6 Relevance and Quality

The challenge is determining the quality of data sets and relevance to particular issues.

3.7 Data Comprehensiveness

Are there areas without coverage? What are the implications?

3.8 Management Challenges

The main management challenges are: Data privacy, Governance, Ethical, Security.

4. BIG DATA CHALLENGES, APPLICATIONS AND PROVIDERS

4.1 Banking and Securities

4.1.1 Industry-Specific big data challenges:
The challenges in this industry include: securities fraud early warning, tick analytics, enterprise credit risk reporting, trade visibility, archival of audit trails, and customer data transformation, social analytics for trading, IT operations analytics, and IT policy compliance analytics, among others.

4.1.2 Applications of big data in the securities and banking industry:
The Securities Exchange Commission (SEC) is using big data to check the progress of financial market activity. They are currently using natural language processor and network analytics to catch illegal trading activity in the financial markets. Big banks, Retail traders, hedge funds use big data for trade analytics used in high frequency trading, pre-trade sentiment measurement, decision-support analytics, Predictive Analytics etc.
This industry also heavily relies on big data for risk analytics including; demand enterprise risk management, anti-money laundering, KYC (Know Your Customer), and fraud mitigation.

4.1.3 Big Data providers specific to this industry include:
Panoptic on Software, 1010data, Stream base Systems, Nice Actimize and Quartet FS.

4.2 Media, Communication, and Entertainment

4.2.1 Industry-Specific big data challenges:
Since consumers expect rich media on-demand in different formats and in a variety of devices, some big data challenges in the media, communications, and entertainment industry include:
- Analyzing, collecting, and utilizing consumer insights
- Social media content and Leveraging mobile
- Understanding patterns of real-time, media content usage

4.2.2 Applications of big data in the Media, Communications, and Entertainment industry:
Organizations in this industry simultaneously analyze their consumer data along with behavioral data to create detailed consumer profiles that can be used to:
- Create content for audiences of different targets
- Recommend on demand content
- Content performance measurement

Spottily, an on-demand music service uses Hadoop big data analytics, to collect data from its billions of users worldwide and then uses the analyzed data to give informed music recommendations to individual user. Amazon Prime, which is driven to provide a great customer experience by offering, music, video, and books in a one-stop shop also heavily, utilizes big data.

4.2.3 Big Data Providers in this industry include:
Splunk, Infochimps, Pervasive Software, and Visible Measures.

4.3 Healthcare

4.3.1 Industry-Specific challenges:
The healthcare field has access to huge amounts of data but has been plagued by failures in utilizing the data to curb the cost of rising healthcare and by inefficient systems that restrain faster and better healthcare benefits across the board. This is mainly because of unavailability of electronic data, inadequate, or unusable. Additionally, the healthcare databases that hold health-related information have made it tough to link data that can show patterns useful in the medical sector. Other challenges related to big data include: the exclusion of patients from the decision making process and the usage of data from different readily available sensors.

4.3.2 Applications of big data in the healthcare sector:
Some hospitals are using data collected from mobile application, from millions of patients, to allow doctors to use evidence-based medicine as opposed to administering several medical or lab tests to all patients those go to the hospital or clinic. A battery of tests can be efficient but they can also be usually ineffective and expensive also. Free public health data and Google Maps have been used by the University of Florida to create visual data that allows for fastest identification and efficient analysis of healthcare information, used to track the spread of chronic disease. Obamacare has also utilized big data in various ways.

4.3.3 Big Data Providers in this industry include:
Humedica, Recombinant Data, Explorys and Cerner

4.4 Education

4.4.1 Industry-Specific big data challenges:
From a practical point of view, institutions and staff have to learn the new data management and analysis tools. And from a technical point of view, a major challenge in the education industry is to incorporate big data from several suppliers and sources and to utilize it on platforms that were not designed for the varying data. On the technical side, there are challenges to integrate data from several sources, on different platforms and from different suppliers that were not designed to work with one another. Politically, an issue of personal data protection and privacy associated with big data used for educational purpose is a challenge.

4.4.2 Applications of big data in Education Sector:
Big data is used quite significantly in higher education. For example, the Australian Tasmania University with over 26000 students has deployed a Learning and Management Tracking System which tracks among other things, when a student logs onto the system, how much time is he or she spent on different pages in the system, as well as the overall progress of a student over time. Big data is also used to measure effectiveness of teachers to ensure a good experience for both students and teachers. Performance of teacher can be fine-tuned and measured against student numbers, subject matter, student aspirations, student demographics, behavioral classification and several other variables. On a governmental level, the Educational Technology Office in the United State Department of Education is using big data to develop analytics to help course correct students those are going away from the correct direction/path while using online big data courses. Click patterns are also used to detect boredom.

4.4.3 Big Data Providers in this industry include:
Carnegie Learning and Knewton and MyFit/ Naviance

4.5 Government

4.5.1 Industry-Specific challenges:
In government the challenges are the integration and interoperability of big data across several different government departments and affiliated organizations.

4.5.2 Applications of big data in Government:
Big data has a very wide range of applications in public services including: energy exploration, fraud detection, financial market analysis, health related research and environmental protection. Some more examples are as follows:
Big data is being used in the analysis of large amounts of social disability claims, made to the Social Security Administration (SSA), that come in the form of unstructured data. The analytics are used to process medical information rapidly and efficiently for faster decision making and to detect fraudulent or suspicious claims.
The FDA (Food and Drug Administration) is using big data to study and detect patterns of food-related illnesses and diseases. This allows for faster response which has led to faster treatment and less death.
The Homeland Security Department uses big data for several different use cases. Big data is analyzed from

several different agencies of government and is used to protect the country.

4.5.3 Big Data Providers in this industry include:
Socrata, HP, and Digital Reasoning

4.6 Insurance

4.6.1 Industry-Specific challenges:
In a survey conducted by Market force challenges identified by professionals in the insurance industry include underutilization of data gathered by loss adjusters and a hunger for better insight. Lack of personalized pricing, lack of personalized services and the lack of targeted services to new segments and to specific market segments are some of the main challenges.

4.6.2 Applications of big data in the insurance industry:
Big data has been used in the insurance industry to provide customer insights for transparent and simpler products, by predicting and analyzing customer behavior through data derived from social media, CCTV footage and GPS-enabled devices. It also allows for better customer retention from insurance companies. When it comes to management of claims, predictive analytics from big data has been used to offer faster service since huge amounts of data can be analyzed especially in the underwriting stage. Fraud detection has also been enhanced. Through huge amounts of data from digital channels and social media, real-time monitoring of claims throughout the claims cycle has been used to provide insights.

4.6.3 Big Data Providers in this industry include:
Octo Telematics, Sprint, Qualcomm, The Climate Corp.

4.7 Retail and Whole sale trade

4.7.1 Industry-Specific challenges:
From traditional brick and wholesalers and mortar retailers to current day e-commerce traders, the industry has collected a lot of data over time. This data derived from loyalty cards of customer, POS scanners, RFID etc. is not being used enough to improve customer experiences on the whole. Any improvements and changes made in it, have been quite slow.

4.7.2 Applications of big data in the Retail and Wholesale industry:
Big data from customer loyalty data, store inventory, POS, local demographics data continues to be collected by retail and wholesale stores.
In New York's Big Show retail trade conference in 2014, companies like Cisco, Microsoft and IBM pitched the need for the retail industry to big data utilization for analytics and for other uses including:
- Reduced fraud
- Optimized staffing through data from shopping patterns, local events, and so on
- Timely analysis of inventory

Social media use also has a lot of potential usage and continues to be slowly but surely adopted especially by brick and mortar stores. Social media is used for customer prospecting, promotion of products, customer retention, and more.

4.7.3 Big Data Providers in this industry include:
First Retail, First Insight, Infor, Epicor and Vistex

4.8 Transportation

4.8.1 Industry-Specific challenges:
In recent times, massive amount of data from location-based social networks and high speed data from telecoms have affected travel behavior. Unfortunately, research to understand travel behavior has not progressed as quickly. In most places, transport demand models are based on poorly understood new social media structures.

4.8.2 Applications of big data in the transportation industry:
Some applications of big data in transportation by governments, private organizations and individuals include:
- **Government:** traffic control, intelligent transport systems, , route planning, congestion management (by predicting traffic conditions)
- **Private sector:** revenue management, logistics, technological enhancements and for competitive advantage by optimizing freight movement and consolidating shipments
- **Individual:** for travel arrangements in tourism, route planning to save on fuel and time etc.

4.8.3 Big Data Providers in this industry include:
Qualcomm and Manhattan Associates

REFERENCES

Research Papers:
[1] S. Justin Samuel1, Koundinya RVP2, Kotha Sashidhar3 and C.R. Bharathi4, "A SURVEY ON BIG DATA AND ITS RESEARCH CHALLENGES", ARPN, May 2015.
[2] Shilpa, Manjit Kaur, "BIG Data and Methodology-A review", International Journal of Advanced Research in Computer Science and Software Engineering.
[3] "Introduction to Big Data: Infrastructure and Networking", White Paper Juniper Networks 2012.
[4] Jianwu Wang, Daniel Crawl, Ilkay Altintas, Weizhong Li, " Big Data Applications using Workflows for Data Parallel

Computing" University of California, San Diego Digital Object Indentifier, IEEE 2014.

[5] "Introduction to Big Data: Infrastructure and Networking", White Paper Juniper Networks 2012.

[6] Roberto V. Zicari, "Big Data: Challenges and Opportunities".

Web Sites:

[7] https://www.simplilearn.com/big-data-applications-in-industries-article

BIOGRAPHIES

Prof. Priya Parhate.
Computer Science & Engineering Department,
Siddhivinayak Technical Campus,
School of Polytechnic & Research Technology, Shegaon
e-mail: priya_parhate@rediffmail.com

Mr. Gaurav Ghogle.
Studying in Third Year,
CSE Deptt,
STC, SPRT,
Shegaon
e-mail: ghoglegaurav29@gmail.com

Miss. Jyoti Bhange.
Studying in Third Year,
CSE Deptt,
STC, SPRT,
Shegaon
e-mail: jyotibhange11@gmail.com

Miss. Ashwini Ingle.
Studying in Third Year,
CSE Deptt,
STC, SPRT,
Shegaon
e-mail: aaashwiniingle1996@gmail.com