Applications of big Data: Current Status and Future Scope

[1]Sabia, [2]Sheetal Kalra

Department of Computer Science & Engineering
Guru Nanak Dev University Regional Campus, Jalandhar, India
Email: [1]sabiajal@gmail.com, [2]sheetal.kalra@gmail.com

Abstract : Big data is a main buzz phrase and new curve for IT today. Big data is driven data with high velocity, volume, variety, veracity and value. It comes from different sources like mobile devices, internet, social media, sensors, geospatial devices and other machine-generated data. Traditional data processing and analysis of structured data using RDBMS and data warehousing no longer satisfy the challenges of Big Data. Due to the high velocity and volume of big data, the effective option is to store the big data in cloud, because it has capability to store and process massive amount of big data. Cloud computing also offers the big data implementation in small and medium sized businesses. This paper presents various real time applications of big data that include healthcare, networking security, market and business, sports, education system, gaming and telecommunications.

Keywords: Big Data, Cloud Computing

I. INTRODUCTION

Devices and people are constantly generating data. In IT industry big data is a largest buzz phase. The data has increased day by day from last twenty years; some facts about data are 2 million searching queries on Google, 277,000 tweets, 100 million emails, and 350 GB data processing on facebook every minute [19]. Big data is a new opportunities for enterprise to extract huge volume of data real time and rational and nonrational data types. New technologies and personal communication producing the big data trends, the global internet population grew by 6.5percent from 2010 to 2011 [1]. In 2013, estimates reached 4 Zettabyte of data generated worldwide [5]. Big data describes any voluminous structured (array, files, records, table, tree), unstructured can be textual (PowerPoint, Word document, Email messages, instant messages) and non textual (JPEG, MP3 audio files, flash video files) and semi-structured (weblogs, social media feeds) [17]. In 2014, estimated worldwide data at a staggering 7ZB [13]. Today 2 billion people are connected together and generating massive amount of data every second and IDC study found that, by 2020 data volumes are expected to increase 50 times. 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 [15]. Cloud computing provides small to medium sized business to implement big data technologies and reduce processing cost and hardware cost. Data is generated through interaction with social media, messaging, mobile applications, automated processes, sensors and computers [2]. Real time applications of big data in different industries like healthcare, network security, market and business, sports, education systems, gaming Industry, telecommunication. Big data technologies process high-variety, high-volume and high-velocity to extract data value and ensure high-veracity of original data [6]. Volume, velocity and variety as the biggest challenges of data management. Big data definition having the following 5 V's (volume, velocity, variety, veracity and value) properties [3]:

1. Volume: Big data comes in one size: XXL (range 30-50 terabytes TBs) through enterprises [4]. The available storage cannot handle structure and unstructured data; this is a big problem for enterprises.
2. Velocity: Velocity defines the speed of data that enters the enterprise and then analyzed to increase the profit of business before the value of the information lost.
3. Variety: Data can be structured, unstructured, semi structured or mix of three [10]. It comes in many forms like logs files, tweets, images, videos, audio, text, PDF files, click streams etc.
4. Veracity: Veracity means "conformity with truth or fact". Data sources (even in same domain) are of different qualities with differences accuracy, coverage and timelineness.
5. Value: It refers to the processing of the data and produced it during analysis. Value of data is not one time use and reused for future by combined with another data sets.

The amount of data increased day-by-day that comes from various channels (Table 1).

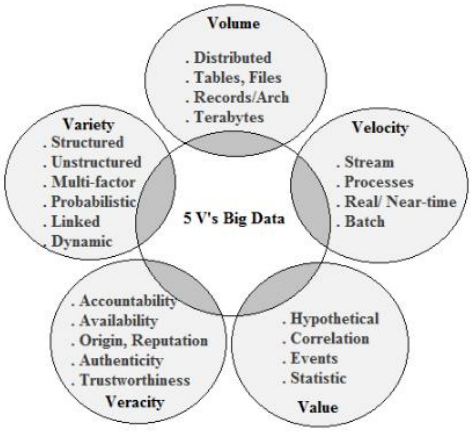

Fig. 1 Properties of Big Data

Table 1. Information explosion

Content Type	Quantity		Comments
Internet	20	Exabyte's (10^{18})	1Exabyte=1,000,000 terabytes
Blogs	70	Million (10^6)	36,718 listed on technocratic
YouTube Visitors	375	Million (10^6)	As of December 2009
Facebook Members	500	Million (10^6)	40% of online hours, top 10 properties
Social content Creators	600	Million (10^6)	People (33% of internet users)
Social Members	2.1	Billion (10^9)	Memberships-top 115 social sites
Live Posts	2.1	Billion (10^9)	Forums ,discussion boards
Tweets	20	Billion (10^9)	50 million user accounts
Web Pages	1.5	Trillion (10^{12})	Plus dark web
Formal Periodicals	10s	Thousands (10^3)	Newspapers, other publications

This paper is organised as follows. In section II cloud enabled big data have been described along with models and types of the paper. In section III various big data applications has been discussed. Future Scope has been discussed in section IV for direction to emerging researchers and section V gives the conclusion of the paper.

II. CLOUD ENABALED BIG DATA

The cloud consists of terrestrial servers across the internet that collectively manages store and process data. Cloud computing is use of resources (software and hardware) that are delivered as a service over the internet or other network. Cloud computing is a paradigm of service oriented computing. Three most popular cloud computing models include: Platform as a Service (PaaS), Infrastructure as a Service (IaaS), and Software as a Service (SaaS) [8].

1. SaaS: It provides the complete application to a cloud end user. It is accessed through a service oriented architectures and web portal based on web service technologies. The services seen on an application layer as an extension of ASP (Application service provider), in which application is maintained, run and supported by a service vendor. Examples of this are Gmail, Hot mail and online banking.
2. PaaS: It is an environment for provisioning and developing cloud applications. Benefits for using PaaS include the ability to upgrade or change and minimize expenses and streamlined version development and the main risk of PaaS is centralization requires different/new security measure. The popular PaaS is google app Engine.
3. IaaS: Infrastructure layer access the IT resources services (data storage resources, computing resources and communications channel) combined under the IaaS. Physical resources are abstracted by visualization; they can share several operating systems and end user environments on the virtual resources (CPU, RAM).

The cloud combines servers into large computing pool and divide the single server into multiple virtual machines. There are three types of cloud exist [7].

1. Public cloud: Public cloud computing model in which services (storage and applications) are use over the internet. It may be based on a pay-per-usage mode [16].
2. Private cloud: Private cloud is internal data center not available publically but operates within a firewall [16]. There are two variations of private cloud: on-premise private cloud and extremely hosted private cloud. An info-tech survey 76% IT decision-makers will focus on private cloud.
3. Hybrid cloud: It is a mix of public and private cloud. Hybrid architecture requires both on-premise resources and off-site server based cloud infrastructure.

III. APPLICATIONS OF BIG DATA

Big data applications solve and analyze real world problems using Hadoop and associated tools. Internet users and machine-to-machine connections are causing the data growth. Real time areas are defined following in which big data is used:

A. Big data in healthcare

Healthcare practices and policies differ tremendously around the world, there are three objectives regarding healthcare system [10]. The first objective is to improve the patient experience (including quality and satisfaction). Second, improving overall population health and reducing the cost of health care and third is traditional methods have fallen short to manage

healthcare and create modern technology to analyze large quantities of information. It is time consuming for clinical staff to Collecting massive amounts of data in healthcare. High-performance analytics are new technologies making easier to turn massive amounts of data into relevant and critical insights used to provide better care. Analytics helps to predict negative intervene and reactions. Unstructured data can be captured through text mining from patient records. It means information can be collecting without causing additional work for clinicians. Transparent, information can thus improve encourage and quality innovation. As information becomes increasingly available, comparable and transparent, patients will also be empowered and more involved in their own treatment through online health applications, which can integrate patient information with their health records and make it available to clinicians. A massive amount of data collected from different sources provides the best practices for today, and will help healthcare providers identify trends so they can achieve better results to improve medical facilities all around the world.

B. Network Security

Big data is changing the landscape of security technologies. The tremendous role of big data can be seen in network monitoring, forensics and SIEM [11]. Big data can also create a world where maintaining control over the revelation of our personal information is challenged constantly. Present analytical techniques don't work well at large scales and end up producing false positives that their efficacy is undermined and enterprises move to cloud architectures and gather much more data, the problem is becoming worse. Big data analytics is an effective solution for processing of large scale information as security is major concern in enterprises. Fraud detection is uses for big data analytics. Phone and credit card companies have conducted large-scale fraud detection for decades. Mainly big data tools are particularly suited to become fundamental for forensics and ATP.

C. Market and business

Big Data is the biggest game-changing opportunity for sales and marketing, since 20 years ago the Internet went main stream, because of the unprecedented array of insights into customer needs and behaviours it makes possible [12]. But many executives who agree that this is true aren't sure how to make the most of it and they also find themselves faced with overwhelming amounts of data and rapidly changing customer behaviours, organizational complexity and increased competitive pressures. According to Gartner, 50% internet connection between Internet of things (IoT) devices and number reached over 15 billion in 2011 and 30 billion by 2020[18]. Some companies are succeeding at turning that Big Data promise into reality. Those that use Big Data and analytics effectively show profitability and productivity rates that are 5–6% higher than those of their peers. The companies that succeed aren't the ones who have the most data, but the ones who use it best. Marketing of big data provides a strategic road map for executives who want to clear the chaos and start driving competitive advantage and top line growth. Using real-world examples additional downloadable resources, non-technical language, and a healthy dose of humour will help you discover the remedy offered by data-driven marketing. Big data reveals customers' behaviour and proven ways to elevate customer experiences. These insights to insure your business's success.

D. Sports

Sport, in business, an increasing volume of information is being collected and captured. Technological advances will fuel exponential growth in this area for the foreseeable future, as athletes are continuously monitored by tools as diverse as sports daily saliva, GPS systems and heart rate monitors tests. These statistics and many more like them are high performance in Big Data. These numbers there is a massive amount of potential insight and intelligence for trainers, administrators, coaches, athletes, sports medics and players. Statistics can be analyzed and collected to better understand what are the critical factors for optimum performance and success, in all facets of elite sport. Injury prevention, competition, Preparation, and rehabilitation can all benefit by applying this approach. Recruitment, Scouting and retention can also be enhanced by these powerful principles. Keeping an eye on various information a coach or a manager can easily and quickly understand which athletes and players need additional support, training, and guidance. Areas for reasons for success and improvement will be understood more clearly. Used consistently this is a powerful measure of progress and performance.

E. Education Systems

By using big data analytics in field of education systems, remarkable results can be seen [9]. Data on students online behaviour can provide educators with important insights, such as if a student requires more attention, the class understanding of a topic is not clear, or if the course has to be modified. Students are required to answer accompanying questions as they go through the set of online content before class. By tracking the number of students that have completed the online module, the time taken and accuracy of their answers, a lecturer can be better informed of the profile of his students and modify the lesson plan accordingly. The analysis of data also clarify about the interest of student looking at time spent in online textbook, online lectures, notes etc. As result instructor can guide choosing the future path effectively.

F. Gaming industry

The amount of data that video game players are generating on a daily basis is growing rapidaitely. Video game developers are using variety of IT techniques such as Hadoop to keep up the massive amount of gaming

data that's generated every day. People are playing video game and generated lot of data in separate areas: game data, player data and session data. In order to improve their game development, game experience, studios are turning to commercial Hadoop distributions such as MapR to analyze, collect and process data from these massive data streams. Armed with this valuable insight from big data, video game publishers are now able to enhance game player engagement and increase player retention by analyzing gamers' social behaviour, activity and tracking players' statistics, calculating rewards, quickly generating leader boards, changing game play and mechanics and delivering virtual prizes, so that experienced players will continue to play the game. By using advanced analytics to uncover rich player insights, developers can now focus on creating meaningful gaming experiences for their customers.

G. Telecommunication Industry

Telecommunications companies have unique advantage in marketplace by controlling the communication infrastructure. Today big challenges for telecommunication are volume, variety and complexity. Current data systems based on batch processing and traditional relation technology, they process big data in real time. Telcos combine ETL and traditional relational databases with big data technologies on a single platform [14]. Telcos technology parses, transforms and integrates the vast amount of data generated by location sensors, IPv6 devices, clickstream, CDRs, 4G networks and machine to machine monitors' information. Telcos parse and transforms from multiple formats and sources including unstructured mobile, media, web and machine monitor provide data. Telcos masking, managing and identifying sensitive data for regulatory compliance. Cloud data integration helps to control over off-premise data managed in the cloud.

IV. FUTURE SCOPE

The new applications are generating vast amount of data in structured and unstructured form. Big data is able to process and store that data and probably in more amounts in near future. Hopefully, Hadoop will get better. New technologies and tools that have ability to record, monitor measure and combine all kinds of data around us, are going to be introduced soon. We will need new technologies and tools for anonymzing data, analysis, tracking and auditing information, sharing and managing, our own personal data in future. So many aspects of life health, education, telecommunication, marketing, sports and business etc that manages big data world need to be polished in future.

V. CONCLUSION

The ability to analyze and store massive amount of structured, unstructured and semi-structure data promises ongoing opportunities for academic institutes, businesses and government organizations. However, a common horizontal big data analytics platform is necessary to support these varieties of real time applications that include healthcare, security, market and business, sports, education system, gaming industry, telecommunications and probably many others in future. The applications have been discussed in this paper. Furthermore, challenges of big data, 5 V's volume, velocity, variety, value, veracity and cloud enabled big data with models and types are also described in this paper. The main goal of our paper is to make a survey of various big data applications that are use in IT industries or organisation to store massive amount of data using technologies (Hadoop, HIVE, NoSQL, Mapreduce and HPCC).

REFRENCES

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

[2] "Big Data A New World of Opportunities", NESSI White Paper, December 2012.

[3] Yuri Demchenko, Canh Ngo, Peter Membrey, "Architecture Framework and Components for the Big Data Ecosystem", UNIVERSITEIT VAN AMSTERDAM System and Network Engineering ,2013.

[4] Yuri Demchenko "The Big Data Architecture Framework (BDAF)",Outcome of the Brainstorming Session at the University of Amsterdam, 2013

[5] "Applying big-data technologies to network architecture" E R I CSSONREVIEW 2012 BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES MAY 2014 Executive Office of the President

[6] Amsterdam Yuri Demchenko, "Defining the Big Data Architecture Framework (BDAF)" Outcome of the Brainstorming Session at the University of Amsterdam SNE Group, University of Amsterdam, 2013

[7] "Big data using cloud computing", Bernice M. Purcell Holy Family University Journal of Technology

[8] Divyakant Agrawal Sudipto Das Amr El Abbadi "Big Data and Cloud Computing: Current State and Future Opportunities"

[9] Anthony G. Picciano, "The Evolution of Big Data and Learning Analytics in American Higher Education" Journal of Asynchronous Learning Networks

[10] "Big Data is the Future of Healthcare Bill Hamilton" cognizant 20-20 insights 2012

[11] Sam curry "big data fuels intelligence –driven security" RSA Security brief 2013

[12] Harikesh S. Nair, Sanjog Misra, "Big Data and Marketing Analytics in Gaming: Combining Empirical Models and Field Experimentation"

[13] "Big Data: What It Is and Why You Should Care" IDC, 2011

[14] Big Data for the Telecommunications Industry Minimize Data Cost, Maximize Data Value Ericsson Traffic and Market Data Report November 2011

[15] Divyakant Agrawal Sudipto Das Amr El Abbadi "Big Data and Cloud Computing: Current State and Future"

[16] Bernice M. Purcell, "Big data using cloud computing" Holy Family University Journal of Technology Research

[17] 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

[18] Paulo Goes, Harry Jiannan Wang, Qiang Wei, "Guest Editorial: Business applications of Web of Things Decision Support Systems" Elsevier 2013

[19] Shilpa, Manjit Kaur, "BIG Data and Methodology-A review" International Journal of Advanced Research in Computer Science and Software Engineering

◈◈◈