ISSN (ONLINE): 2395-695X
ISSN (PRINT): 2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.3, Special Issue.24, March 2017

SURVEY ON BIG DATA

THENMOZHI L, Research scholar, Asst prof, Dept of BCA, MGR COLLEGE. Hosur, Tamilnadu, India.
Dr. CHANDRA KALA N, Head of the department, Dept of CS ,SSM arts and science college, kumarapalayam, Tamilnadu, India.

ABSTRACT

In this paper, we discuss about the journey of big data. The main focus of this big data definitions, big data characteristics, Big data architecture, big data technologies which includes big data storage technologies, inside the technologies then discuss about the concepts of types and sources of big data, big data analytics, big data challenges, data base security framework, security and privacy challenges, opportunities for big data analytics, big data applications ,need of security in big data, opportunities and challenges on big data those topics will be covered in this survey.

KEYWORDS:
Analytics, Security. Hadoop, NoSQL.

INTRODUCTION

In order to analyze complex data and to identify patterns it is very important to securely store, manage and share large amounts of complex data. Big Data could play a pivotal role in some of the advanced technologies like data mining, speech recognition, cross lingual information retrieval.

Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualization, and information privacy. The term often refers simply to the use of predictive analytics or other certain advanced methods to extract value from data, and seldom to a particular size of data set. Accuracy in big data may lead to more confident decision making. And better decisions can mean greater operational efficiency, cost reductions and reduced risk.

Data sets grow in size in part because they are increasingly being gathered by cheap and numerous information-sensing mobile devices, aerial (remote sensing), software logs, cameras, microphones, radio-frequency identification (RFID) readers, and wireless sensor networks. The world's technological per-capita capacity to store information has roughly doubled every 40 months since the 1980s; as of 2012, every day 2.5 Exabyte's (2.5×1018) of data were created; The challenge for large enterprises is determining who should own big data initiatives that straddle the entire organization.

Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time. Big data "size" is a constantly moving target, as of 2012 ranging from a few dozen terabytes to many petabytes of data. Big data is a set of techniques and technologies that require new forms of integration to uncover large hidden values from large datasets that are diverse, complex, and of a massive scale. Software firms specializing in data management and analytics.

In 2010, this industry was worth more than $100 billion and was growing at almost 10 percent a year: about twice as fast as the software business as a whole. Developed economies make increasing use of data-intensive technologies. There are 4.6 billion mobile-phone subscriptions worldwide and between 1 billion and 2 billion people accessing the internet between 1990 and 2005 ,more than 1 billion people worldwide entered the middle class which means more and more people who gain money will become more literate which in turn leads to information growth.

The world's effective capacity to exchange information through telecommunication networks was 281 petabytes in 1986, 471 petabytes in 1993, 2.2 Exabyte's in 2000, 65 Exabyte's in 2007and it is predicted that the amount of traffic flowing over the internet will reach 667 Exabyte's annually by 2014. It is estimated that one third of the globally stored information is in the form of alphanumeric text and still image data, which is the format most useful for most big data applications. This also shows the potential of yet unused data (i.e. in the form of video and audio content).

DEFINITION

Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time. Big data "size" is a constantly moving target, as of 2012 ranging from a few dozen terabytes to many petabytes of data. Big data is a set of techniques and technologies that require new forms of integration to uncover large hidden values from large datasets that are diverse, complex, and of a massive scale.

In a 2001 research report and related lectures, META Group (now Gartner) analyst Doug Laney defined data growth challenges and opportunities as being three-dimensional, i.e. increasing volume (amount of data), velocity (speed of data in and out), and variety (range of data types and sources). Gartner, and now much of the industry, continue to use this "3Vs" model for describing big data.

In 2012, Gartner updated its definition as follows: "Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and

ISSN (ONLINE): 2395-695X
ISSN (PRINT): 2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.3, Special Issue.24, March 2017

process optimization." Additionally, a new V "Veracity" is added by some organizations to describe it.

If Gartner's definition (the 3Vs) is still widely used, the growing maturity of the concept fosters a more sound difference between big data and Business Intelligence, regarding data and their use:

- Business Intelligence uses descriptive statistics with data with high information density to measure things, detect trends etc.;

- Big data uses inductive statistics and concepts from nonlinear system identification to infer laws (regressions, nonlinear relationships, and causal effects) from large sets of data with low information density to reveal relationships, dependencies and perform predictions of outcomes and behaviors.

Wikipedia: "Big data is the term for a collection of data sets so large and complex that it becomes difficult to process using on hand database management tools or traditional data processing applications."

Horrigan (2013): "I view Big Data as non-sampled data, characterized by the creation of databases from electronic sources whose primary purpose is something other than statistical inference."

Rodriguez (2012): "For years, statisticians have been working with large volumes of data in fields as diverse as astronomy, bioinformatics, and data mining. Big Data is different because it is generated on a massive scale by countless online interactions among people, transactions between people and systems, and sensor-enabled machinery

A more recent, consensual definition states that "Big Data represents the Information assets characterized by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value".

The useful perspective is to characterize big data as having high volume, high velocity, and high variety—the three Vs [Russom, 2011]

-***High volume***—the amount or quantity of data
-***High velocity***—the rate at which data is created
-***High variety***—the different types of data

In short, "big data" means there is more of it, it comes more quickly, and comes in more forms.

Both of these perspectives are reflected in the following definition [Mills, Lucas, Irakliotis, Rappa, Carlson, and Perlowitz, 2012; Sicular, 2013]:

Big data is a term that is used to describe data that is high volume, high velocity, and/or high variety; requires new technologies and techniques to capture, store, and analyze it; and is used to enhance decision making, provide insight and discovery, and support and optimize processes.

It is important to understand that what is thought to be big data today won't seem so big in the future [Franks, 2012].

Many data sources are currently untapped—or at least underutilized. For example, every customer e-mail, customer-service chat, and social media comment may be captured, stored, and analyzed to better understand customers' sentiments.

Web browsing data may capture every mouse movement in order to better understand customers' shopping behaviors. Radio frequency identification (RFID) tags may be placed on every single piece of merchandise in order to assess the condition and location of every item.

CHARACTERISTICS

Big Data is the word used to describe massive volumes of structured and unstructured data that are so large that it is very difficult to process this data using traditional databases and software technologies. The term "Big Data [5]" is companies who had to query loosely structured very large distributed data. The three main terms that signify Big Data have the following properties:

a) **Volume:** Many factors contribute towards increasing Volume streaming data and data collected from sensors etc.
b) **Variety:** Today data comes in all types of formats emails, video, audio, transactions etc.
c) **Velocity:** This means how fast the data is being produced and how fast the data needs to be processed to meet the demand.

The other two dimensions that need to consider with respect to Big Data are Variability and Complexity.

d) **Variability:** Along with the Velocity, the data, flows can be highly inconsistent with periodic peaks.
e) **Complexity:** Complexity of the data also needs to be considered when the data is coming from multiple sources. The data must be linked, matched, cleansed and transformed into required formats before actual processing.

Factory work and Cyber-physical systems may have a 6C system:

1. Connection (sensor and networks),
2. Cloud (computing and data on demand),
3. Cyber (model and memory),
4. Content/context (meaning and orrelation),
5. Community (sharing and collaboration),
6. Customization (personalization and value).

ARCHITECTURE

In 2000, Seisint Inc. developed C++ based distributed file sharing framework for data storage and querying. Structured, semi-structured and/or unstructured data is stored and distributed across multiple servers. Querying of data is done by modified C++ called ECL which uses apply scheme on read method to create structure of stored data during time of query.

In 2004 LexisNexis acquired Seisint Inc. and 2008 acquired Choice Point, Inc. and their high speed parallel processing platform. The two platforms were merged into HPCC Systems and in 2011 was open sourced under Apache v2.0 License. Currently HPCC and Quant

ISSN (ONLINE): 2395-695X
ISSN (PRINT): 2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.3, Special Issue.24, March 2017

cast File System are the only publicly available platforms capable of analyzing multiple Exabyte's of data.

In 2004, Google published a paper on a process called Map Reduce that used such architecture. The Map Reduce framework provides a parallel processing model and associated implementation to process huge amounts of data. With Map Reduce, queries are split and distributed across parallel nodes and processed in parallel (the Map step).

The results are then gathered and delivered (the Reduce step). The framework was very successful, so others wanted to replicate the algorithm. Therefore, an implementation of the Map Reduce framework was adopted by an Apache open source project named Hadoop.

MIKE2.0 is an open approach to information management that acknowledges the need for revisions due to big data implications in an article titled "Big Data Solution Offering". The methodology addresses handling big data in terms of useful permutations of data sources, complexity in interrelationships, and difficulty in deleting (or modifying) individual records.

Recent studies show that the use of multiple layer architecture is an option for dealing with big data. The Distributed Parallel architecture distributes data across multiple processing units and parallel processing units provide data much faster, by improving processing speeds. This type of architecture inserts data into a parallel DBMS, which implements the use of Map Reduce and Hadoop frameworks. This type of framework looks to make the processing power transparent to the end user by using a front end application server.

Big Data Analytics for Manufacturing Applications can be based on a 5C architecture (connection, conversion, cyber, cognition, and configuration).

Big Data Lake - With the changing face of business and IT sector, capturing and storage of data has emerged into a sophisticated system. The big data lake allows an organization to shift its focus from centralized control to a shared model to respond to the changing dynamics of information management. This enables quick segregation of data into the data lake thereby reducing the overhead time.

BIG DATA STORAGE TECHNOLOGIES

The ability to store massive amounts of data is a necessity for business executives to use big data. Two major means of storing big data are clustered network-attached storage (NAS), also called scale-out NAS, and object-based storage systems (Sliwa, 2011). Without a change to data storage technology, executives will not be able to collect big data.

Scale-out NAS is built upon a traditional NAS system. NAS is a storage device that is based on a computer with no keyboard or mouse; this computer only serves as a device to retrieve data for users (White, 2011). To support the demands of big data, several NAS devices are connected, or clustered, and each NAS device can search through devices attached to the other NAS devices.

As indicated in Figure 1 (Appendix), each NAS is attached to several storage devices, which the NAS is able to search. In turn this "NAS pod" is connected by a switch to another "NAS pod" which does the same function. Because the pods are connected through the switch, both pods can be searched for data by any client. Clients may be directly connected on a local network, a VPN, or somewhere on the cloud attached through a network.

In object-based storage systems, users deal not with files but with sets of objects which are distributed over several devices (Wang, Brandt, Miller, & Long, 2004). Object-based storage systems provide high capacity and throughput as well as reliability and scalability, which are all needed for big data storage (Wang, Brandt, Miller, & Long, 2004). It is the layout of the objects themselves is what provides the efficiency of the storage and searching, rather than the configuration of the storage system as in scale-out NAS.

TECHNOLOGIES

Big data requires exceptional technologies to efficiently process large quantities of data within tolerable elapsed times. A 2011 McKinsey report suggests suitable technologies include A/B testing, crowd sourcing, data fusion and integration, genetic algorithms, machine learning, natural language processing, signal processing, simulation, time series analysis and visualization.

Multidimensional big data can also be represented as tensors, which can be more efficiently handled by tensor-based computation, such as multilinear subspace learning. Additional technologies being applied to big data include massively parallel-processing (MPP) databases, search-based applications, data mining, distributed file systems, distributed databases, cloud based infrastructure (applications, storage and computing resources) and the Internet.

Some but not all MPP relational databases have the ability to store and manage petabytes of data. Implicit is the ability to load, monitor, back up, and optimize the use of the large data tables in the RDBMS.

DARPA's Topological Data Analysis program seeks the fundamental structure of massive data sets and in 2008 the technology went public with the launch of a company called Ayasdi.

The practitioners of big data analytics processes are generally hostile to slower shared storage, preferring direct-attached storage (DAS) in its various forms from solid state drive (SSD) to high capacity SATA disk buried inside parallel processing nodes. The perception of shared storage architectures—Storage area network (SAN) and Network-attached storage (NAS) —is that they are relatively slow, complex, and expensive. These qualities

ISSN (ONLINE): 2395-695X
ISSN (PRINT): 2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.3, Special Issue.24, March 2017

are not consistent with big data analytics systems that thrive on system performance, commodity infrastructure, and low cost.

Real or near-real time information delivery is one of the defining characteristics of big data analytics. Latency is therefore avoided whenever and wherever possible. Data in memory is good—data on spinning disk at the other end of a FC SAN connection is not. The cost of a SAN at the scale needed for analytics applications is very much higher than other storage techniques.

There are advantages as well as disadvantages to shared storage in big data analytics, but big data analytics practitioners as of 2011 did not favour it.

INSIDE THE TECHNOLOGY

Big-data technologies are usually engineered from the bottom up with two things in mind: scale and availability. Consequently, most solutions are distributed in nature and introduce new programming models for working with large volumes of data. Because most of the legacy database models cannot be effectively used for big data, the current approach to ensuring availability and partitioning needs to be revised.

NoSQL

Like key-value storage of semi-structured data, NoSQL systems are designed with specific characteristics in mind, such as relaxed models of consistency. They run applications that tend to be read/write-intensive.

To take advantage of scaling capacity as new nodes are added to a network, many NoSQL databases are designed to expand horizontally and run on low-cost commodity hardware. NoSQL databases have far more relaxed or even nonexistent data-model restrictions. Such databases allow applications to store almost any kind of structure in a data element, and the responsibility for maintaining the logical data structure is transferred to the application.

Most NoSQL databases are key-value stores that hold schema-less collections of entities that do not necessarily share the same properties. The data consists of a string containing the key, and the actual data is considered to be the value in the key-value relationship. For example: document stores – containing complex data structures that are usually stored in JavaScript

Object Notation (JSON); and graph stores, or graph databases – an emerging NoSQL category using nodes (stand-alone objects or entities) and edges (lines used to connect nodes and properties) to represent and store information. Another current trend is to combine SQL and NoSQL in a non-conflicting manner that enables the technologies to work well together.

TYPES AND SOURCES OF BIG DATA

TYPES OF DATA

The value that can be derived from using big-data technologies depends on the use case and the data associated with it. Apart from volume and velocity, the value that can be gained from the variability of data tends to be overlooked. Put simply, the less structured the data, the greater the requirement to apply big-data technologies. Variability is typically categorized into three different data types:

Structured – data is well organized, there are several choices for abstract data types, and references such as relations, links and pointers are identifiable;

Unstructured – data may be incomplete and/or heterogeneous, and often originates from multiple sources. It is not organized in an identifiable way, and typically includes bitmap images or objects, text and other data types that are not part of a database; and

Semi-structured – some data is organized, containing tags or other markers to separate semantic elements,but unstructured data may also be present.

Executives need to be cognizant of the types of data they need to deal with. There are three main types of data, regardless of whether or not a company is using big data – unstructured data, structured data, and semistructured data. Unstructured data are data in the format in which they were collected; no formatting is used (Coronel, Morris, & Rob, 2013).

Some examples of unstructured data are PDF's, e-mails, and documents (Baltzan, 2012). Structured data are formatted to allow storage, use, and generation of information (Coronel, Morris, & Rob, 2013). Traditional transactional databases store structured data (Manyika et al., 2011). Semi-structured data have been processed to some extent (Coronel, Morris, & Rob, 2013). XML or HTML-tagged text are examples of semistructured data (Manyika et al., 2011).

Business executives with traditional database management systems need to broaden their data horizons to include collection, storage, and processing of unstructured and semi-structured data. Data collection of unstructured and semistructured data is done through several internet-based technologies. Chui, Löffler, and Roberts (2010) describe sensors providing big data as being part of the Internet of Things.

The Internet of Things is described as sensors and actuators that are embedded in physical objects that provide data through wired and wireless networks (Chui, Löffler, & Roberts, 2010). Some industries that are creating and using big data are those that have recently begun digitization of their data content; these industries include entertainment, healthcare, life sciences, video surveillance, transportation, logistics, retail, utilities, and telecommunications (Chui, Löffler, & Roberts, 2010). Devices generating data in these industries include IPTV cameras, GPS transceiver, RFID tag readers, smart meters, and cell phones (Chui, Löffler, & Roberts, 2010).

ISSN (ONLINE): 2395-695X
ISSN (PRINT): 2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.3, Special Issue.24, March 2017

BIG DATA ANALYTICS

Storing big data is only part of the picture. Special techniques are needed to analyze big data. Executives need to become familiar with the big data methodologies, adopt the technology appropriate for their business, and ensure that employees develop skill with the technology.

Data storage techniques differ depending on whether the data are unstructured or structured. Unstructured and semi structured data can be analyzed using software like Hadoop. Users analyzing structured big data can use software such as NoSQL, MongoDB, and TerraStore.

Hadoop is based on a programming paradigm called Map Reduce, as discussed in Google's 2004 paper on Hadoop (Eaton, Deroos, Deutsch, Lapis, & Zikopoulos, 2012). The name Map Reduce comes from the two distinct tasks that the Hadoop program will perform using key-value pairs when a query is made (Eaton, Deroos, Deutsch, Lapis, & Zikopoulos, 2012).

The mapping task is given a piece of data known as a key to search on, finds relevant values based on this key, and converts the key and values into another dataset query (Eaton, Deroos, Deutsch, Lapis, & Zikopoulos, 2012). The reducing task takes the final resultant output (the key and value combinations) from the mapping and reduces the output into a small dataset which answers the query (Eaton, Deroos, Deutsch, Lapis, & Zikopoulos, 2012).

Hadoop works well in a scale-out NAS environment. The mapping task will search all possible datasets for the data being queried. Due to the size of the environment, this will produce a huge dataset for the output. The reduce task will analyze the dataset output from mapping and check that only data the directly answers the query is returned.

For example, if the user queries the system for the highest sales amount for each of four sales people, the map task will search the system for all sales for the four sales people, and the reduce task will limit the output to the highest sales amount for each sales person. Researchers from Techaisle found that 73% of businesses in their study preferred using Hadoop because of its capability to process large volumes of big data (Business & Finance Week editors, 2012).

Due to the volume of data stored, structured data can also be considered big data depending upon how it is stored (scale-out NAS or object-based storage). There are several different software options commonly used to analyze structured big data.

NoSQL, which can mean either 'no SQL' or 'not only SQL,' is characterized by data that is Basically Available, Soft state, and Eventually consistent (BASE), rather than the traditional database data characteristics of Atomicity, Consistency, Isolation, and Durability (ACID) (Oracle, 2011). Data analyzed using NoSQL, therefore, is at times in a state of transition and may not be directly available; the data is in flux rather than set as in traditional database environments. MongoDb and TerraStore are both NoSQL-related products that are used for "document-oriented applications" such as storage and searching of whole invoices rather than the individual data fields from the invoice (Sasirekha, 2011).

BIG DATA CHALLENGES

Infrastructure Security

- Secure computations in distributed programming frameworks
- Security best practices for non-relational data stores Data Privacy
-Scalable privacy-preserving data mining and analytics
-Cryptographically enforced data centric security
-Granular access control

Data Management

-Secure data storage and transactions logs
-Granular audits
-Data provenance

Integrity & Reactive Security

-End-point input validation/filtering
-Real-time security monitoring

Big Data Security Framework

The following section provides the target security architecture framework for Big Data platform security.
The core components of the proposed Big Data Security Framework are the following:

1. Data Management
2. Identity & Access Management
3. Data Protection & Privacy
4. Network Security
5. Infrastructure Security & Integrity

TOP 10 SECURITY & PRIVACY CHALLENGES

The Cloud Security Alliance Big Data Security Working Group has compiled the following as the Top 10 security and privacy challenges to overcome in Big Data [4].

1. Secure computations in distributed programming frameworks
2. Security best practices for non-relational data stores
3. Secure data storage and transactions logs
4. End-point input validation/filtering
5. Real-time security monitoring
6. Scalable privacy-preserving data mining and analytics
7. Cryptographically enforced data centric security
8. Granular access control
9. Granular audits

ISSN (ONLINE): 2395-695X
ISSN (PRINT): 2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.3, Special Issue.24, March 2017

10. Data provenance

OPPORTUNITIES FOR BIG DATA ANALYTICS

Following are opportunities for big data analytics in different industries [Franks, 2012].

Automobile insurance [Rose,S., 2013] – pricing, client risk analysis, fraud detection, faster claims processing

Telecommunications [IBM, 2013] – analysis of patterns of services across social networks, profitability of customers' social networks, churn minimization

Manufacturing, distribution, and retail [CGT, 2012] – tracking shelf availability, assessing the impact of promotional displays, assess the effectiveness of promotional campaigns, inventory management, pricing, advanced clickstream analysis

Transportation and logistics [IBM, 2010] – real-time fleet management, RFID for asset tracking

Utilities [Oracle, 2013] – analysis of smart grid data to determine variable pricing models, smart meters to forecast energy demand, customized rate plans for customers

Gaming [Chulis, 2012] – game play analysis to provide feedback to game producers, opportunities for ingame offers

Law enforcement [Wyllie, 2013] – identifying people linked to known trouble groups, determining the location of individuals and groups

.

BIG DATA APPLICATIONS

The big data application refers to the large scale distributed applications which usually work with large data sets. Data exploration and analysis turned into a difficult problem in many sectors in the span of big data. With large and complex data, computation becomes difficult to be handled by the traditional data processing applications which triggers the development of big data applications.

Google's map reduce framework and apache Hadoop are the defacto software systems for big data applications, in which these applications generates a huge amount of intermediate data. Manufacturing and Bioinformatics are the two major areas of big data applications.

Big data provide an infrastructure for transparency in manufacturing industry, which has the ability to unravel uncertainties such as inconsistent component performance and availability. In these big data applications, a conceptual framework of predictive manufacturing begins with data acquisition where there is a possibility to acquire different types of sensory data such as pressure, vibration, acoustics, voltage, current, and controller data. The combination of sensory data and historical data constructs the big data in manufacturing. This generated big data from the above combination acts as the input into predictive tools and preventive strategies such as prognostics and health management.

Another important application for Hadoop is Bioinformatics which covers the next generation sequencing and other biological domains. Bioinformatics which requires a large scale data analysis, uses Hadoop.

Government

-United States of America
-India
-United Kingdom
International development
Manufacturing
Cyber-Physical Models

Media

-Internet of Things (IoT) Technology
-eBay.com
- Amazon.com
-Facebook
-GOOGLE

Private sector

-Retail
- Retail Banking
-Real Estate
-Science
-Science and Research

Health care

-Right living

-Right care

-Right provider

-Right value

-Right innovation

NEED OF SECURITY IN BIG DATA

For marketing and research, many of the businesses uses big data, but may not have the fundamental assets particularly from a security perspective. If a security breach occurs to big data, it would result in even more serious legal repercussions and reputational damage than at present.

In this new era, many companies are using the technology to store and analyze petabytes of data about their company, business and their customers. As a result, information classification becomes even more critical. For making big data secure, techniques such as encryption, logging, honeypot detection must be necessary.

In many organizations, the deployment of big data for fraud detection is very attractive and useful. The challenge of detecting and preventing advanced threats and malicious intruders must be solved using big data style analysis. These techniques help in detecting the threats in

the early stages using more sophisticated pattern analysis and analyzing multiple data sources.

Not only security but also data privacy challenges existing industries and federal organizations. With the increase in the use of big data in business, many companies are wrestling with privacy issues. Data privacy is a liability, thus companies must be on privacy defensive.

But unlike security, privacy should be considered as an asset, therefore it becomes a selling point for both customers and other stakeholders. There should be a balance between data privacy and national security.

OPPORTUNITIES AND CHALLENGES

Businesses

Benefits of big data use to business executives include enhanced data sharing through transparency, improved performance though analysis, augmented market segmentation, increased decision support through advanced analytics, and greater ability to innovate products, services and business models.

Business owners need to follow trends in big data carefully to make the decision that fits their businesses.

Healthcare

The healthcare system will have to change significantly for stakeholders to take advantage of big data. The old levers for capturing value—largely cost-reduction moves, such as unit price discounts based on contracting and negotiating leverage, or elimination of redundant treatments—do not take full advantage of the insights that big data provides and thus need to be supplemented or replaced with other measures related to the new value pathways.

Similarly, traditional medical-management techniques will no longer be adequate, since they pit payors and providers against each other, framing benefit plans in terms of what is and isn't covered, rather than what is and is not most effective. Finally, traditional fee-for-service payment structures must be replaced with new systems that base reimbursement on insights provided by big data—a move that is already well under way.

CONCLUSION

In this above given data's are analyzed based on the big data developments in future, we have to proposed some efficient data conventional mechanism to storing the large variety of data's .storing large volume of data having many mechanism when getting the accurate data's is an inefficient one to solve this inconvenience future research planed to provide the solution for data retrieval from large data sets with the efficient way of compressing and decompressing of large data's in big data.

REFERENCES:

[1] EMC Big Data 2020 Projects http://www.emc.com/leadership/digital-universe/iview/big-data-2020.htm

[2] NIST Special Publication 1500-1 *NIST Big Data Interoperability Framework: Volume 1, Definitions* http://bigdatawg.nist.gov/_uploadfiles/M0392_v1_3022325181.pdf.

[3] Securosis – Securing Big Data Security issues with Hadoop environments https://securosis.com/blog/securing-big-data-security-issues-with-hadoop-environments

[4] Top 10 Big Data Security and Privacy Challenges, Cloud Security Alliance, 2012 https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Big_Data_Top_Ten_v1.pdf

[5] A, Katal, Wazid M, and Goudar R.H. "Big data: Issues, challenges, tools and Good practices.". Noida:2013, pp. 404 – 409, 8-10 Aug. 2013.

[6]. Meredith A. Barrett,1,2* Olivier Humblet,1,2* Robert A. Hiatt,3 and Nancy E. Adler BIG DATA AND DISEASE PREVENTION :From Quantified Self to Quantified Communities.

[7].Waldrop M. Big data: Wikiomics. Nature 2008; 455:22.

[8]. Dumbill E, Liddy ED, Stanton J, et al. Educating the next generation of data scientists. Big Data 2013; 1:21–27.

[9] F.C.P, Muhtaroglu, Demir S, Obali M, and Girgin C. "Business model canvas perspective on big data applications." *Big Data, 2013 IEEE International Conference*, Silicon Valley, CA, Oct 6-9, 2013, pp. 32 - 37.

[10] Zhao, Yaxiong , and Jie Wu. "Dache: A data aware caching for big-data applications using the MapReduce framework." *INFOCOM, 2013 Proceedings IEEE,* Turin, Apr 14-19, 2013, pp. 35 - 39.

[11] Xu-bin, LI , JIANG Wen-rui, JIANG Yi, ZOU Quan "Hadoop Applications in Bioinformatics." *Open Cirrus Summit (OCS), 2012 Seventh*, Beijing, Jun 19-20, 2012, pp. 48 - 52.

[12].Venkata Narasimha Inukollu1 , Sailaja Arsi1 and Srinivasa Rao Ravur "SECURITY ISSUES ASSOCIATED WITH BIG DATA IN CLOUD COMPUTING" International Journal of Network Security & Its Applications (IJNSA), Vol.6, No.3, May 2014.

ISSN (ONLINE): 2395-695X
ISSN (PRINT): 2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.3, Special Issue.24, March 2017

[13] Bernice Purcell, **The emergence of "big data" technology and analytics ,** Journal of Technology Research .