International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org ||Volume 7 Issue 5 Ver V || May 2018 || PP 15-23

Challenges And Opportunities With Big Data, Analytics For A New Era Of Efficiency

Vanlalhruaia[1], Robert Laltanpuia[2] , R.Chawngsangpuii[3], Lalhmingliana[4]
[1]*Department of Computer Engineering, Mizoram University,Tanhril,796004,India*
[2]*Bachelor of Vocational, Mizoram University,Tanhril,796004,India*
[3]*Department of Information Technology, Mizoram University,Tanhril,796004,India*
[4]*Department of Computer Engineering, Mizoram University,Tanhril,796004,India*
[*]*Corresponding Author: Vanlalhruaia*

Abstract : *This paper provides the new technology and developments in big data for Government, public and private sectors. It outlines the architecture of Big data analysis and the working process of big data. It mentions the applications of Big Data Analysis for future generation. This analytics process of structured and unstructured data will provide a great benefits for company, universities as well as the government. It discussed about the challenges of big data, i.e 7 V's volume, velocity, variety, variability, value, visualization and veracity. In this challenging technology Big Data needs High-Performance Analytics software tools and applications for predictive analytics, data mining, text mining, forecasting and data optimization. Hadoop, NoSQL, Cassandra, HIVE, Mapreduce are used for storing a massivie. large volumes of data that a business has collected to determine which data is relevant and can be analyzed to drive better business decisions in the future. 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 to support a varieties of real time applications that include healthcare, security, market and business, sports, education system, gaming industry ,telecommunications and probably many others in future . Furthermore, challenges of big data,7 V's volume, velocity, variety, variability, value, visualization, veracity and cloud enabled big data with models and types are also described in this paper. It described the properties of Big Data , it mentioned the working of analyzing a Big data and various tools and requirements for analyzing Big data.*

Keywords : *Big data, big data analytics, big data platforms, data analysts, data mining.*

Date of Submission: 07-05-2018 Date of acceptance: 22-05-2018

I. Introduction

THIS big data analytics refers to the process of collecting, organizing and analyzing large sets of data (called big data) so as to discover patterns and other useful information. Big data analytics can help organizations to better understand the information contained within the data and will also help identify the data that is most important to the business and future business decisions. Big data analysts basically want the knowledge that comes from analyzing the data. Big Data is becoming one of the most talked about technology trends nowadays. Without big data the transaction can not be completed in every environment.The real challenge with the big organization is to get maximum out of the data already available and predict what kind of data to collect in the future. How to take the existing data and make it meaningful that it provides us accurate insight in the past data is one of the key discussion points in many of the executive meetings in organizations. With the explosion of the data the challenge has gone to the next level and now a Big Data is becoming the reality in many organizations.It helps in many way for the development of the social and economic environment in the technology.The use of Big Data[1] is becoming a crucial way for leading companies to outperform their peers. In most industries, established competitors and new entrants alike will leverage data-driven strategies to innovate, compete and capture value. Indeed, we found early examples of such use of data in every sector we examined. In healthcare, data pioneers are analyzing the health outcomes of pharmaceuticals when they were widely prescribed, and discovering benefits and risks that were not evident during necessarily more limited clinical trials. Other early adopters of Big Data are using data from sensors embedded in products from children's toys to industrial goods to determine how these products are actually used in the real world. Such knowledge then informs the creation of new service offerings and the design of future products.

Big Data will be helpful to create new opportunities and entirely new categories of companies, such as those that aggregate and analyse industry data. Many of these will be companies that sit in the middle of large information flows where data about products and services, buyers and suppliers, consumer preferences and intent can be captured and analysed[2]. Forward-thinking leaders across sectors should begin aggressively to build their organisations' Big Data capabilities.

The source and variety of big data involves new technologies that create , communicate,or are involved with data-generating activities, which produce different types/formats of data resources. What big data entails is structured and unstructured data that correspond to various

activities. Structured data entails data that is categorized and stored in a file according to a particular format description, where unstructured data is free-form text that takes on a number of types, such as those listed above.The cell phones of yesteryear have evolved into smartphones capable of texting,surfing,phoning,and playing a host of software-based

applications. Data resources can provide value to organizations from the information that can be extracted from them.This extraction process involves querying data resources for particular variables at particular levels of aggregation in a particular format and then initiating some type of analytic process.However,before conducting any of these activities,one essential task that underpins the information creation initiative involves the creation of a conceptual model.Inother words,whether you have terabytes of data or just a few thousand records,whether you are considering trends over the past few years or focusing on real-time data feeds, decision makers must determine what questions they are looking to answer with data and information.

In addition to the capabilities of Big Data, the real-time and high-frequency nature of the data are also important. For example, 'nowcasting,' the ability to estimate metrics such as consumer confidence, immediately, something which previously could only be done retrospectively, is becoming more extensively used, adding considerable power to prediction. Similarly, the high frequency of data allows users to test theories in near real-time and to a level never before possible.Due to this Big Data needs to analyzed to get sufficient result in the outcome.

II. Why The 3v's Are Not Sufficient To Describe Big Data

2.1 Big Data Requires High-Performance Analytics

When a large volume of data or cloud data stored in a storage is access by multiple-users,there may be a problem in accessing speed.To analyze a large volume of data, big data analytics is typically performed using specialized software tools and applications for predictive analytics, data mining, text mining, forecasting and data optimization. Collectively these processes are separate but highly integrated functions of high-performance analytics[3]. Using big data tools and software enables an organization to process extremely large volumes of data that a business has collected to determine which data is relevant and can be analyzed to drive better business decisions in the future.

2.2 The Challenges of Big Data Analytics

For most organizations, big data analysis is a challenge.It create and generate a new inovation for the technology.It changes the life in social,economic and political for the society which helps them in the promotion of the technology.It encourages the activies of all aspects. Consider the sheer volume of data and the different formats of the data (both structured and unstructured data) that is collected across the entire organization and the many different ways different types of data can be combined, contrasted and analyzed to find patterns and other useful business information.

The first challenge is in breaking down data silos to access all data an organization stores in different places and often in different systems. A second big data challenge is in creating platforms that can pull in unstructured[4] data as easily as structured data. This massive volume of data is typically so large that it's difficult to process using traditional database and software methods.

2.3 How Big Data Analytics is Used Today

All As the technology that helps an organization to break down data silos and analyze data improves, business can be transformed in all sorts of ways. The advances in analyzing big data allow researchers to decode human DNA in minutes, predict where terrorists plan to attack, determine which gene is mostly likely to be responsible for certain diseases and, of course, which ads you are most likely to respond.

Another example comes from one of the biggest mobile carriers in the world. France's Orange launched its Data for Development project by releasing subscriber data for customers in the Ivory Coast. Researchers accessed the data and sent Orange proposals for how the data could serve as the foundation for development projects to improve public health and safety. Proposed projects included one that showed how to improve public

safety by tracking cell phone data to map where people went after emergencies; another showed how to use cellular data for disease containment.

2.4 The Benefits of Big Data Analytics

Enterprises are increasingly looking to find actionable insights into their data. Many big data projects originate from the need to answer specific business questions. With the right big data analytics platforms in place, an enterprise can boost sales, increase efficiency, and improve operations, customer service and risk management.

About half of all respondents said they were applying big data analytics to improve customer retention, help with product development and gain a competitive advantage.

This same technology can fundamentally change the way government operates, breaking down hierarchies and silos, enabling preventive action, incorporating citizens into every aspect of governance and increasing overall efficiency. Data analytics offer us unprecedented opportunities to improve the effectiveness of government. The key to these opportunities is "big data," the ever-growing volume of information created and captured by the modern digitized world, from cloud-based systems to sensors to smart devices. New data-mining techniques allow governments to break through legacy-system barriers that seemed insurmountable only a couple of years ago. We soon will see new solutions in every area of government, from how public agencies hire, train and promote, to how performance is measured, how problems are identified and preempted, and how personalized services are delivered.These massive amounts of data will drive efficiency only when organized and analyzed in a manner that supports decision-making. Governments are just beginning to meaningfully incorporate data analytics into their operations.

III. Properties Of Big Data

Big data technologies process high-variety, high-volume and high-velocity to extract data value and ensure high-veracity of original data[5]. Volume, velocity and variety as the biggest challenges of data management. Big data definition having the following 7 V's (volume, velocity, variety, variability, value, visualization and veracity) properties[6] :

3.1 Volume:

Big data comes in one size: XXL (range 30-50 terabytes TBs) through enterprises .The available storage cannot handle structure and unstructured data; this is a big problem for enterprises.

90% of all data ever created, was created in the past 2 years. From now on, the amount of data[7] in the world will double every two years. By 2020, we will have 50 times the amount of data as that we had in 2011. The sheer volume[8] of the data is enormous and a very large contributor to the ever expanding digital universe is the Internet of Things with sensors all over the world in all devices creating data every second. The era of a trillion sensors is upon us.

If we look at airplanes they generate approximately 2,5 billion Terabyte of data each year from the sensors installed in the engines. Self-driving cars will generate 2 Petabyte of data every year. Also the agricultural industry generates massive amounts of data with sensors installed in tractors. Shell uses super-sensitive sensors to find additional oil in wells and if they install these sensors at all 10.000 wells they will collect approximately 10 Exabyte of data annually. That again is absolutely nothing if we compare it to the Square Kilometer Array Telescope that will generate 1 Exabyte of data per day.

In the past, the creation of so much data would have caused serious problems. Nowadays, with decreasing storage costs, better storage solutions like Hadoop and the algorithms to create meaning from all that data this is not a problem at all.

3.2 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. Underlying the volume numbers is an even larger trend, which is that 90 percent of extant data have been created in just the last two years. The speed at which data are generated, accumulated and analyzed is on a steep acceleration curve. As of next year, there will be 19 billion network connections globally feeding this velocity.

Although most data are warehoused before analysis, there is an increasing need for real-time processing of these enormous volumes, such as the 200 million emails, 300,000 tweets and 100 hours of Youtube videos that are passing by every minute of the day. Real-time processing reduces storage requirements while providing more responsive, accurate and profitable responses.

The Velocity[9] is the speed at which the data is created, stored, analysed and visualized. In the past, when batch processing was common practice, it was normal to receive an update from the database every night

or even every week. Computers and servers required substantial time to process the data and update the databases. In the big data era, data is created in real-time or near real-time. With the availability of Internet connected devices, wireless or wired, machines and devices can pass-on their data the moment it is created.

The speed at which data is created currently is almost unimaginable: Every minute we upload 100 hours of video on Youtube. In addition, every minute over 200 million emails are sent, around 20 million photos are viewed and 30.000 uploaded on Flickr, almost 300.000 tweets are sent and almost 2,5 million queries on Google are performed. The challenge organisations have is to cope with the enormous speed the data is created and used in real-time.

3.3 Variety:

Data can be structured, unstructured, semistructured or mix of three. It comes in many forms like logs files, tweets, images, videos, audio, text, PDF files, click streams etc. Variety describes different formats of data that do not lend themselves to storage in structured relational database systems. These include a long list of data such as documents, emails, social media text messages, video, still images, audio, graphs, and the output from all types of machine-generated data from sensors, devices, RFID tags, machine logs, cell phone GPS signals, DNA analysis devices, and more. This type of data is characterized as unstructured or semi-structured and has existed all along. In fact it's estimated by some studies to account for 90% or more of the data in organizations. Variety is also used to mean data from many different sources, both inside and outside of the company. What's changed is the realization that through analysis it can yield new and valuable insights not previously available.

There are two primary challenges here. First, storing and retrieving these data types quickly and cost efficiently. Second, during analysis, blending or aligning data types from different sources so that all types of data describing a single event can be extracted and analyzed together.

Then there is the interaction of variety with volume. Unstructured data is growing much more rapidly than structured data. Gartner estimates that unstructured data doubles every three months and offers the example that there are seven million web pages added each day.

In terms of opportunity, Variety is seen by business users as the major focus of new Big Data initiatives. Companies have been handling large volumes of data for many years and view that process as incremental and business and usual. But the new and unique opportunity to add unstructured data to the analytic mix is seen by many as a game changer.

Another challenge of Big Data processing goes beyond the massive volumes and increasing velocities of data but also in manipulating the enormous variety of these data. Taken as a whole, these data appear as an indecipherable mass without structure. Consisting of natural language, hashtags, geo-spatial data, multimedia, sensor events and so much more, the extraction of meaning from such diversity requires ever-increasing algorithmic and computational power. In the past, all data that was created was structured data, it neatly fitted in columns and rows but those days are over. Nowadays, 90% of the data that is generated by organisation is unstructured data. Data today comes in many different formats: structured data, semi-structured data, unstructured data and even complex structured data. The wide variety of data requires a different approach as well as different techniques to store all raw data.

There are many different types of data and each of those types of data require different types of analyses or different tools to use. Social media like Facebook posts or Tweets can give different insights, such as sentiment analysis on your brand, while sensory data will give you information about how a product is used and what the mistakes are.

3.4 Variability :

Big data is extremely variable.That is extremely difficult because words have different meanings an all depends on the context.Variability is often confused with variety. Say you have bakery that sells 10 different breads. That is variety. Now imagine you go to that bakery three days in a row and every day you buy the same type of bread but each day it tastes and smells different. That is variability.

Variability is thus very relevant in performing sentiment analyses. Variability means that the meaning is changing (rapidly). In (almost) the same tweets a word can have a totally different meaning. In order to perform a proper sentiment analyses, algorithms need to be able to understand the context and be able to decipher the exact meaning of a word in that context. This is still very difficult.

There are several potential meanings for Variability. Is the data consistent in terms of availability or interval of reporting? Does it accurately portray the event reported? When data contains many extreme values it presents a statistical problem to determine what to do with these 'outlier' values and whether they contain a new and important signal or are just noisy data.

Furthermore, the intrinsic meanings and interpretations of these conglomerations of raw data depends on its context. This is especially true with natural language processing. A single word may have multiple meanings. New meanings are created and old meanings discarded over time. Interpreting connotations is, for instance, essential to gauging and responding to social media buzz. The boundless variability of Big Data therefore presents a unique decoding challenge if one is to take advantage of its full value.

3.5 Veracity :

Veracity means "conformity with truth or fact".Data sources (even in same domain) are of different qualities with differences accuracy, coverage and timelineness.

Understanding what Big Data is telling you is one thing. However, it is useless if the data being analyzed are inaccurate or incomplete. This situation arises when data streams originate from diverse sources presenting a variety of formats with varying signal-to-noise ratios. By the time these data arrive at a Big Data analysis stage, they may be rife with accumulated errors that are difficult to sort out. It almost goes without saying that the veracity of the final analysis is degraded without first cleaning up the data it works with.

Having a lot of data in different volumes coming in at high speed is worthless if that data is incorrect. Incorrect data can cause a lot of problems for organisations as well as for consumers. Therefore, organisations need to ensure that the data is correct as well as the analyses performed on the data are correct. Especially in automated decision-making, where no human is involved anymore, you need to be sure that both the data and the analyses are correct. If you want your organisation to become information-centric, you should be able to trust that data as well as the analyses accountability.

3.6 Visualization :

A core task for any Big Data processing system is to transform the immense scale of it into something easily comprehended and actionable. For human consumption, one of the best methods for this is converting it into graphical formats. Spreadsheets and even three-dimensional visualizations are often not up to the task, however, due to the attributes of velocity and variety. There may be a multitude of spatial and temporal parameters and relationships between them to condense into visual forms.

This is the hard part of big data. Making all that vast amount of data comprehensible in a manner that is easy to understand and read. With the right analyses and visualizations, raw data can be put to use otherwise raw data remains essentially useless. Visualizations of course do not mean ordinary graphs or pie charts. They mean complex graphs that can include many variables of data while still remaining understandable and readable.

Visualizing might not be the most technological difficult part; it sure is the most challenging part. Telling a complex story in a graph is very difficult but also extremely crucial. Luckily there are more and more big data startups appearing that focus on this aspect and in the end, visualizations will make the difference. One of them is future this will be the direction to go, where visualizations help organisations answer questions they did not know to ask.

3.7 Value :

It refers to the processing of the data and produced it during analysis. Value[10] of data is not one time use and reused for future by combined with another data sets.

Although Value is frequently shown as the fourth leg of the Big Data stool, Value does not differentiate Big Data from not so big data. It is equally true of both big and little data that if we are making the effort to store and analyze it then it must be perceived to have value.

Big Data however is perceived as having incremental value to the organization and many users quote having found actionable relationships in Big Data stores that they could not find in small stores. Certainly it is true that if in the past we were storing data about groups of customers and are now storing data about each customer individually then the granularity of our findings is much finer and we approach that desired end-goal of offering each customer a personalization-of-one in their experience with us.

Another take on Value is that Big Data tends to have low value density, meaning that you have to store a lot of it to extract findings.This is likely true but since new Big Data storage and retrieval technologies are so much less expensive than previous, low value density should not be a hurdle that prevents us from searching for those valuable kernels.

Finally, there is at least one reviewer who goes to philosophical extremes quoting Sartre "existence precedes essence". By which he means that we may choose to store Big Data before even understanding exactly what use we have for it.We're not entirely sure about this. We still encourage business users to work backwards from the desired outcome before deciding exactly what Big Data to capture.

There are at least four additional characteristics that pop up in the literature from time to time. All of these share the same definitional problems of Value. That is they may be a descriptor of data but not uniquely of Big Data.

No one doubts that Big Data offers an enormous source of value to those who can deal with its scale and unlock the knowledge within. Not only does Big Data offer new, more effective methods of selling but also vital clues to new products to meet previously undetected market demands. Many industries utilize Big Data in the quest for cost reductions for their organizations and their customers. Those who offer the tools and machines to handle Big Data, its analysis and visualization also benefit hugely, albeit indirectly.

Although Volume, Velocity and Variety are intrinsic to Big Data itself, the other Vs of Variability, Veracity, Value and Visualzation are important attributes that reflect the gigantic complexity that Big Data presents to those who would process, analyze and benefit from it. All of them demand careful consideration, especially for enterprises not already on the Big Data bandwagon. These businesses may find that their current best practices related to data handling will require thorough revamping in order to stay ahead of the seven Vs. All that available data will create a lot of value for organisations, societies and consumers. Big data means big business and every industry will reap the benefits from big data. The properties of Big Data is shown in figure 1 clerary.

Of course, data in itself is not valuable at all. The value is in the analyses done on that data and how the data is turned

into information and eventually turning it into knowledge. The value is in how organisations will use that data and turn their organisation into an information-centric company that relies on insights derived from data analyses for their decision-making.

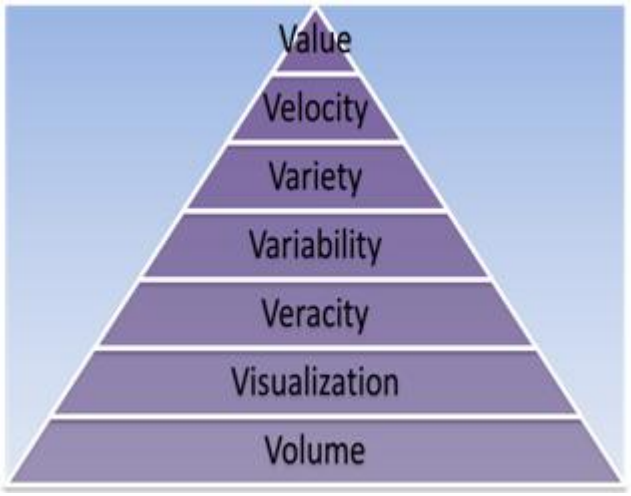

Fig.1.Properties of Big Data

IV. How Big Data Analysis Works

4.1 Data Preparation and Cleaning : The extracted data is subjected to statistical analysis for cleansing[11] and verification and indentifying outliers and applying appropriate treatment.

4.2 Transformation : Advanced algorithms or approaches are applied that extract information from a data sets both structured and unstructured and transform it into an understandable structure for further use.

4.3 Business Intelligence(BI) : Various Business Intelligence tools for analyzing structure data and tools, Hadoop being one of the most popular among them, for unstructured data are applied for predictive analysis.Thus,the architecture of Big Data is displayed in figure 2, for helping the analysis of a big data.

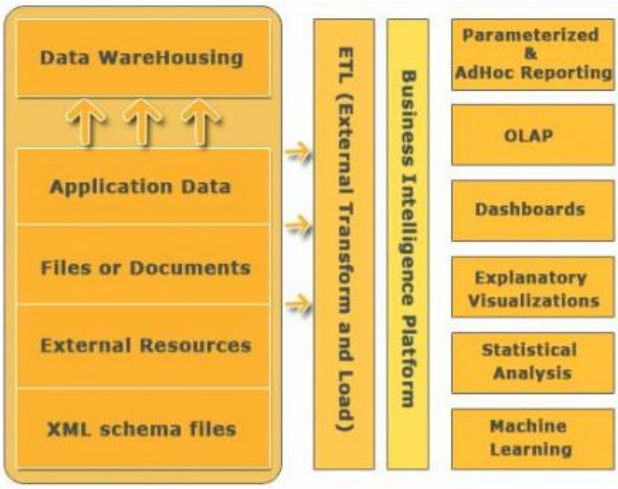

Fig.2. Architecture of Big Data

V. Tools For Analyzing Big Data

IT analytics tools are bringing big data processing techniques into the data center, for faster, more informed infrastructure management decisions. There's nothing quite like managing an environment where the concept of real time is measured by the infinitesimally small standards of electronics. In any one of those instants, things can go from smooth sailing to disaster. And, always, there is data to be moved and managed and data -- lots of it-- about how the data center itself is behaving. Into that challenging environment comes a new idea: applying new styles of processing and analysis borrowed from the world of big data technologies (e.g., Hadoop, NoSQL and Cassandra) and business analytics to help decision makers better master the challenges of IT management. With big data analytics offerings aimed at IT management, companies can combine real-time streaming data analysis along with terabytes of historical data correlation and analysis to detect patterns that can, in turn, help predict and prevent future outages or performance issues. Furthermore, they can leverage big data to understand usage patterns and geographical trends and gain insights about their heaviest users. They can also track and record Web activity and easily identify business impacts; accelerate profitable growth with insights into service utilization; and gather data across multiple systems to develop an IT services catalog.

There are five key approaches to analyzing big data and generating insight:

5.1 Discovery tools are useful throughout the information lifecycle for rapid, intuitive exploration and analysis of information from any combination of structured and unstructured sources. These tools permit analysis alongside traditional BI source systems. Because there is no need for up-front modeling, users can draw new insights, come to meaningful conclusions, and make informed decisions quickly.

5.2 BI tools are important for reporting, analysis and performance management, primarily with transactional data from data warehouses and production information systems. BI Tools provide comprehensive capabilities for business intelligence and performance management, including enterprise reporting, dashboards, ad-hoc analysis, scorecards, and what-if scenario analysis on an integrated, enterprise scale platform[12].

5.3 In-Database Analytics include a variety of techniques for finding patterns and relationships in your data.Because these techniques are applied directly within the database, you eliminate data movement to and from other analytical servers, which accelerates information cycle times and reduces total cost of ownership.

5.4 Hadoop is useful for pre-processing data to identity macro trends or find nuggets of information, such as outof-range values. It enables businesses to unlock potential value from new data using inexpensive commodity servers. Organizations primarily use Hadoop[13] as a precursor to advanced forms of analytics.

5.5 Decision Management includes predictive modeling, business rules, and self-learning to take informed action based on the current context. This type of analysis enables individual recommendations across multiple channels, maximizing the value of every customer interaction. Oracle Advanced Analytics scores can be integrated to operationalize complex predictive analytic models and create real-time decision processes.

All of these approaches have a role to play uncovering hidden relationships. Traditional data discovery tools like Oracle Endeca Information Discovery, BI tools like Oracle Exalytics, and decision management tools like Oracle Real Time Decisions are covered extensively in other white papers. In this paper, we mainly focus on the integrated use of Hadoop and In-Database Analytics to process and analyze a vast field of new data. Data is generated through interaction with social media, messaging, mobile applications, automated processes, sensors and computers[14]. Real time applications of big data in different industries like healthcare, network security, market and business, sports, education systems, gaming Industry, telecommunication.

VI. Big Data Analysis Requirements

In the previous section, Techniques for Analyzing Big Data, we discussed some of methods you can use to find meaning and discover hidden relationships in big data. Here are three significant requirements for conducting these inquiries in an expedient way:

- Minimize data movement
- Use existing skills
- Attend to data security

Minimizing data movement is all about conserving computing resources. In traditional analysis scenarios, data is brought to the computer, processed, and then sent to the next destination. For example, production data might be extracted from e-business systems, transformed into a relational data type, and loaded into an operational data store structured for reporting. But as the volume of data grows, this type of ETL architecture becomes increasingly less efficient. There's just too much data to move around. It makes more sense to store and process the data in the same place.With new data and new data sources comes the need to acquire new skills. Sometimes the existing skillset will determine where analysis can and should be done. When the requisite skills are lacking, a combination of training, hiring and new tools will address the problem. Since most organizations have more people who can analyze data using SQL than using MapReduce[15], it is important to be able to support both types of processing.Data security is essential for many corporate applications. Data warehouse users are accustomed not only to carefully defined metrics and dimensions and attributes, but also to a reliable set of administration policies and security controls. These rigorous processes are often lacking with unstructured data sources and open source analysis tools. Pay attention to the security and data governance requirements of each analysis project and make sure that the tools you are using can accommodate those requirements.

VII. Application Of Big Analysis In The Government : Few Scenarios

- Employment exchanges may help unemployed job seekers find jobs by combining their job qualifications, place of residence, age etc. with an analysis of available job opportunities.
- "Aadhaar " Cards in India:Based on the data transformed, it is possible to project the total number of people below poverty line,average economy of a person etc. Based on the findings government may take a decision to improve the life of poor people, take women and child care schemes to enhance woman population etc. It may be correlated with rural health statistical parameters to predict onset of a disease.
- Analysis of National health records for identifying the spread of epidemic or efficacy ofclinical trials, with respect to region, gender age etc.
- Data forensics of Logs or other data, logs data play a major role in understanding the customer demand and his usage of the applications.Based on the clicks that the user makes, the points through which he navigates provide us information that helps in enhancing the usability of the web pages and the application access.

VIII. 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.

Big data will change business, and business will change society. The hope, of course, is that the benefits will outweigh the drawbacks, but that is mostly a hope. The big-data world is still very new, and, as a society, we're not very good at handling all the data that we can collect now.

IX. Conclusions

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,7 V's volume, velocity, variety, variability, value, visualization, 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).

References

[1]. Advancing Discovery in Science and Engineering.Computing Community Consortium.Spring 2011.
[2]. Steve Lohr (2012) . The Age of Big Data.
[3]. Lavalle, S., Lesser, E., Shockley, R., Hopkins, M.S., & Kruschwitz, N. (2011) . Big Data , Analytics and the Path from Insights to Value . Sloan Management Review, 52(2) , 20-31.
[4]. Shirer, M., & Murray, P. (2012) . Will Be the Year of Mobile and Cloud Platform Wars as It Vendors Vie for Leadership While the Industry Redefines Itself, IDC Analyze.
[5]. Agneeswaran, V . S., Tonpay , P., & Tiwary , J. (2013) . Paradigms for realizing machine learning algorithms. Big Data , 1(4) , 207–214 .
[6]. Yuri Demchenko, Canh Ngo, Peter Membrey,(2013). Architecture Framework and Components for the Big Data Ecosystem .
[7]. Dewitt, D., & Gray, J. (1992) . Parallel Database Systems: The Future of High Performance Database Systems, 35(6), 85-98.
[8]. T. Niimura, T., Dhaliwal, M., & Ozawa,K.. (2001, July). Fuzzy regression models to represent electricity market data in deregulated power industry, 2556-2561.
[9]. Xu, Y., Kostamaa, P., & Gao, L. (2010) . Integrating Hadoop and Parallel Dbms : Proceedings of the 2010 ACM SIGMOD International Conference on Management of data, 969-974.
[10]. Matturdi,B., Zhou,X., Li ,S., & Lin, F.(2014, April).Big data security and privacy: A review, 135-145 .
[11]. Wang, J., Krishnan, S., Franklin, M.J., Goldberg, K.., Kraska, T., & Milo, T. (2014,June). A sample-and-clean framework for fast and accurate query processing on dirty data. In Proc. ACM SIGMOD Int. Conf. Manage. Data, 469-480 .
[12]. Amsterdam Yuri Demchenko,(2013). Defining the Big Data Architecture Framework (BDAF) : Outcome of the Brainstorming Session at the University of Amsterdam SNE Group, University of Amsterdam .
[13]. Sun, H., & Heller, P.,(2012). Oracle Information Architecture: An Architect's Guide to Big Data., Oracle, Redwood Shores.
[14]. NESSI White Paper, (2012,December). Big Data : A New World of Opportunities.
[15]. [15] Pavlo, A., Paulson, E., Rasin, A., Abadi, D.J., Dewitt, D.J., Madden, S., & Stonebraker, M.,(2009). A Comparison of Approaches to Large-Scale Data Analysis. Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data, 165-178.

Vanlalhruaia " Challenges And Opportunities With Big Data, Analytics For A New Era Of Efficiency "International Journal of Engineering Science Invention (IJESI), vol. 07, no. 05, 2018, pp 15-23