



Big Data Analytics: Prospects, Challenges, and Applications

Omer Abdullah ¹, Mohd Yousuf ²

¹ CSE, Lords Institute of Engineering and Technology, Hyderabad, India

² CSE, ISL Engineering College, Hyderabad, India

Abstract:

In the era of the fourth industrial revolution (Industry 4.0), big data has major impact on businesses, since the revolution of networks, platforms, people and digital technology have changed the determinants of firms' innovation and competitiveness. An ongoing huge hype for big data has been gained from academics and professionals, since big data analytics leads to valuable knowledge and pro motion of innovative activity of enterprises and organizations, transforming economies in local, national and international level. In that context, data science is defined as the collection of fundamental principles that promote information and knowledge gaining from data. The techniques and applications that are used help to analyze critical data to support organizations in understanding their environment and in taking better decisions on time. Nowadays, the tremendous increase of data through the Internet of Things (continuous increase of connected devices, sensors and smartphones) has contributed to the rise of a "data-driven" era, where big data analytics are used in every sector (agriculture, health, energy and infrastructure, economics and insurance, sports, food and transportation) and every world economy. The growing expansion of available data is a recognized trend worldwide, while valuable knowledge arising from the information come from data analysis processes. In that context, the bulk of organizations are collecting, storing and analyzing data for strategic business decisions leading to valuable knowledge. The ability to manage, analyze and act on data ("data-driven decision systems") is very important to organizations and is characterized as a significant asset. A huge repository of terabytes of data is generated each day from modern information systems and digital technologies such as Internet of Things and cloud computing. Analysis of these massive data requires a lot of efforts at multiple levels to extract knowledge for decision making. Therefore, big data analysis is a current area of research and development. The basic objective of this paper is to explore the potential impact of big data Prospects, Challenges, and Applications .

Keyword :

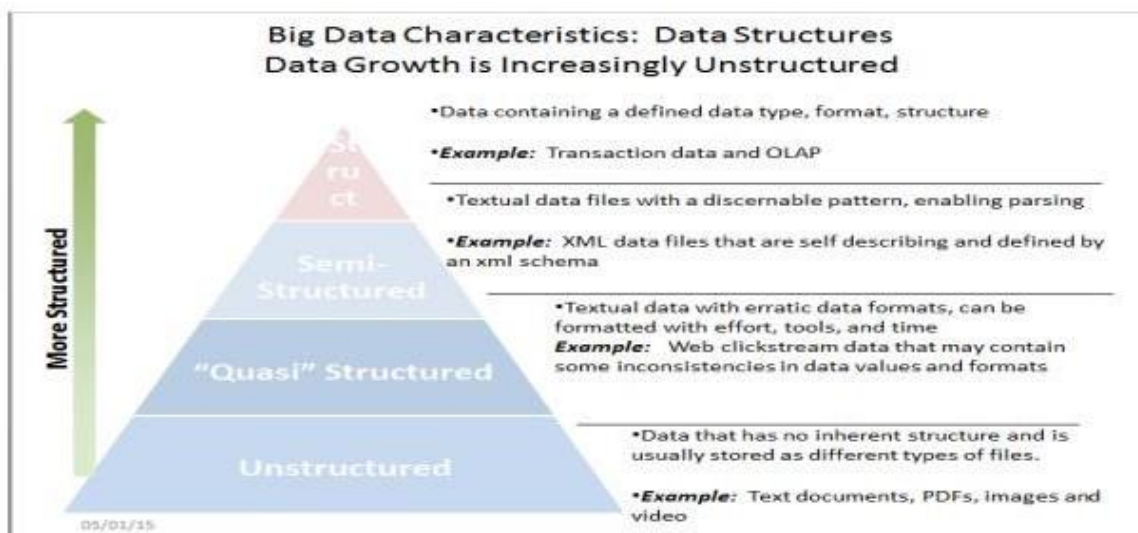
Big data · Big data analytics · Performance · Enterprises Knowledge management · Internet of things (IoT)

1 Introduction :

There are various ways of defining the big data. In online big data is defined as a group of large volume of datasets that is not possible process using usual computing methods. More than 90% of word's data is generated in the last 5 years and day by day data is growing very rapidly due to the arrival of novel technologies, devices and communication ways for example social sites, sensor networks and hand held devices. Today the amount of data that is creating in one day through social networking sites is equal to the data created since beginning of time till 2006. Every day more than 2.5 quintillion bytes of information is created and it is still increasing very fast. Can you imagine if this amount of data is written on disks, it may fill an whole cricket field? So there is a need to process this huge amount of data to get knowledge which can be useful for analyzing the customers and taking the remedial actions, healthcare, whether report

creation etc.

Big data is still in fantasy, it will grow enormously in near future with the use of IOT devices and home automations. The Application of big data analytics will continuously grow in personal and professional use. As the data is growing the companies have started to invest in analytics infra and management of information. The big data analytics has become their integral part for the rapid growth of the company. It is known to all of us that opportunities always come with



challenges and difficulties. The companies must be able to resolve the challenges to utilize the full potential of big data analytics. Nowadays, in the 4th Industrial revolution era, organizations and governments focus on the development of capabilities that provide knowledge extracted from large and complex data sets, commonly known as “big data”. Big data is a buzz word in the last years in the business and economics fields, since it plays an essential role in economic activity and has strengthened its role in creating economic value by enabling new ways to spur innovation and productivity growth. Hence, the ability of management, analysis and acting is significant under the context of knowledge-based capital (KBC) that is associated with digital information, innovative capacity and economic aspects.

Fig: Nature of Data

Characteristics of Data

Accuracy	Is the information correct in every detail?
Completeness	How comprehensive is the information?
Reliability	Does the information contradict other trusted resources?
Relevance	Do you really need this information?
Timeliness	How up- to-date is information? Can it be used for real-time reporting?

“THE FIVE V’s”: volume, variety, velocity, value and veracity .

- Volume:** Volume indicates more data; it is granular in nature and unique. Big data requires processing high volumes of low density unstructured Hadoop data- that is, data of unknown value such as Twitter data feeds, click streams on web page, mobile app and many more. It is the task of big data to convert such Hadoop data into valuable information. For some organizations it might be tens of terabytes for other it might be hundreds of petabytes.
- Velocity:** The fast at which data is received and acted upon. The highest velocity data normally streams directly into memory. Some IOT applications have health and safety ramifications that require real time evaluation and action. Operationally mobile phone users have large user populations, increased network traffic and

expectations for immediate response.

3. **Variety:** Unstructured and semi structured data types such as audio and video. These types of data require different types of analysis or different type tools to use in order to process it. Unstructured data has many of the same requirements as structured data such as summarization, lineage, auditability and privacy. Complexity arises when data source changes without notice.

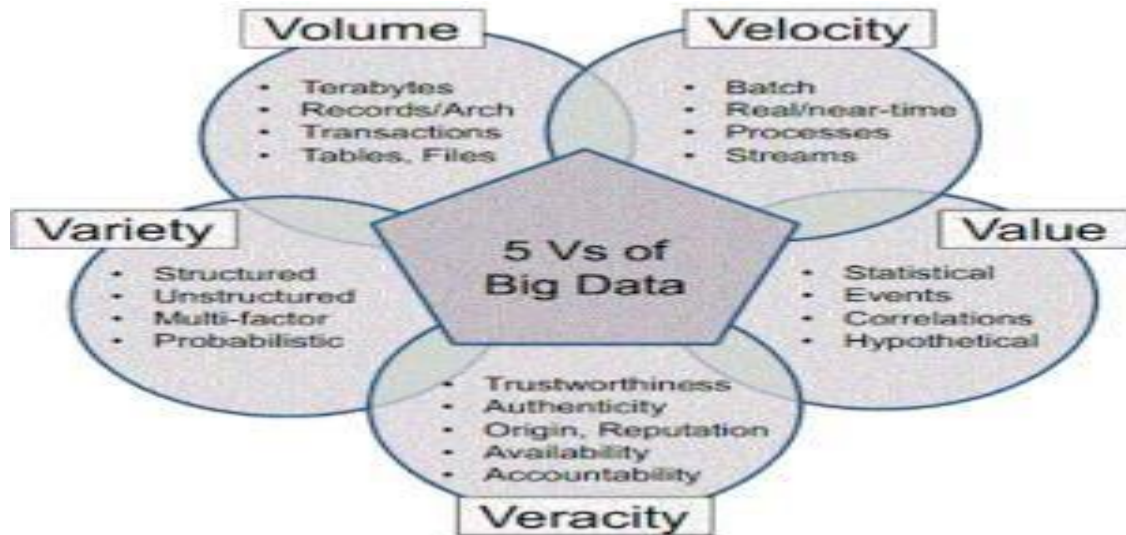


Fig : V's of Big Data

4. **Value:** There are a range of quantitative and investigative techniques to derive value from data. Big data involves in extracting valuable information from mass data to predict future trends and make decisions.
5. **Veracity:** It refers to messiness and trustworthiness of data. With many forms of Big data, quality and accuracy are less controllable (Twitter hash tags, abbreviations etc.) Big data analytic technologies will help to work with this type of data. Volumes often make up with lack of quality and accuracy.

Big Data Important :

1. **Cost Savings:** Some tools of Big Data like Hadoop and Cloud-Based Analytics can bring cost advantages to business when large amounts of data are to be stored and these tools also help in identifying more efficient ways of doing business.

2. **Time Reductions:** The high speed of tools like Hadoop and in-memory analytics can easily identify new sources of data which helps businesses analyzing data immediately and make quick decisions based on the learning.

3. **Understand the market conditions:** By analyzing big data you can get a better understanding of current market conditions. For example, by analyzing customers' purchasing behaviors, a company can find out the products that are sold the most and produce products according to this trend. By this, it can get ahead of its competitors.

4. **Control online reputation:** Big data tools can do sentiment analysis. Therefore, you can get feedback about who is saying what about your company. If you want to monitor and improve the online presence of your business, then, big data tools can help in all this.

5. **Using Big Data Analytics to Boost Customer Acquisition and Retention :** The customer is the most important asset any business depends on. There is no single business that can claim success without first having to establish a solid customer base. However, even with a customer base, a business cannot afford to disregard the high competition it faces. If a business is slow to learn what customers are looking for, then it is very easy to begin offering poor quality products. In the end, loss of clientele will result, and this creates an adverse overall effect on business success. The use of big data allows businesses to observe various customer related patterns and trends. Observing customer behavior is important to trigger loyalty.

6. **Using Big Data Analytics to Solve Advertisers Problem and Offer Marketing Insights:** Big data analytics can

help change all business operations. This includes the ability to match customer expectation, changing company's product line and of course ensuring that the marketing campaigns are powerful.

7. Big Data Analytics: As a Driver of Innovations and Product Development Another huge advantage of big data is the ability to help companies innovate and redevelop their products.

2-Big Data Analytics Prospects :

Analytics in decision making procedure is not something new, since business analytics appeared as early as in the mid1950s—Analytics 1.0 era—with the advent of tools that were able to generate and capture larger amounts of data in enterprises data warehouses and discover patterns more quickly than human minds with business intelligence tools. In that first era, managers gained a data-based comprehension going beyond intuition in decision making. Until mid-2000s, the rapid growth of data generation and the arrival of big data have signaled a new era— Analytics 2.0—where enterprises have the opportunity to leverage that data with new more powerful tools. The need of new innovative technologies appeared and enterprises moved quickly to acquire the necessary capabilities and knowledge for gaining insights from big data, with the major difference between eras being in skills required for data analysis.

In the next era, analytics is an integral part of enterprises supporting decision making and enterprises move to creation of analytics-based products/services. Moving ahead, the next era—Analytics 3.0 or “data economy”—is characterized by the tremendous increase of data generation coming from the growth of Internet of Things (IoT) with 8.4 billions connected devices in 2017 globally and 20.4 billion by 2020.

The most recent era—Analytics 4.0—includes cognitive technologies including machine learning, where actions and decision making are shifted to augmentation with dynamic machine automation.

In the current era of analytics, the emerging new technologies will increase the generation of data, thus enterprises and organizations have to face up technical challenges in order to have access to more and better data. The worldwide revenues of big data and business analytics (BDA) will be more than \$203 billion in 2020 and banking, manufacturing, government and professional services will be the top industries in BDA investments according to International Data Corporation (IDC).

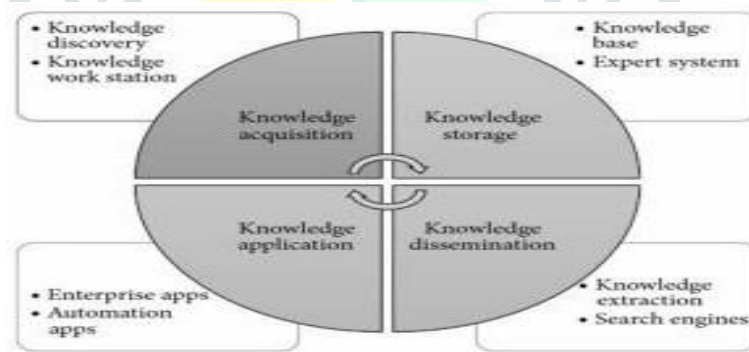


Fig: IoT Knowledge Exploration System

3-

CHALLENGES OF BIG DATA ANALYTICS :

With the advancement of technology like Internet of Things, Sensor Networks, Social Media , Access of web in handheld devices has brought us in a new era of information that is structured to unstructured in nature called as Big data. Sorting and extracting the knowledge from such data is acknowledged by business organizations. But test with this large data is that traditional techniques can not process such large data. New models and systems are used to process such large data. Hadoop is an ecosystem framework that is used for this purpose.

1. Data Challenges:

It means set of challenges associated with features of data. It includes amount of data, pace of data, diversity of data, data heterogeneity, visualization and value.

➤ **Storage and management of unstructured data:** The big data experts have acknowledged management

of data as a big challenge in year 2020. management of data is a process of organizing, storing and managing the data collected by any organization. Another data challenge is huge arrival of unstructured data that sound and videos etc.

- **AI and ML Challenge:** Artificial Intelligence and Machine learning is also a big challenge in near future since these technologies mostly depends on data management. Both these technologies requires quality data to work efficiently and generating great results.
- **Internet of Things:** As the IOT is growing and more and more devices are being connected using sensors. These networked devices are going to create huge data. This data will generate the best results only if it is efficiently stored organized and managed.

2. Process Challenges:

The challenges that occurs during processing of data. It includes data acquirement and storing, mining and cleaning, aggregation and integration of data, analysis and data interpretation.

- **How to deal with data growth:** The heterogeneous data in huge volume is coming from different sources with very fast speed that needs to be processed immediately to generate useful insights with in time frame.. So dealing with such data is a big challenge.
- **Integration of different datasets:** Big data is collected in real-time from various sources in different formats and quality. So challenge is to store in scalable data warehouses and integrating this data after removing inconsistencies.
- **Validation of data:** Data is being created, captured, changed, processed and analyzed at very extraordinary level. So quality of data is not up to the level so poor data quality can include huge cost for analysis. Therefore validation of data quality is very essential to improve the data quality and reduction of processing cost.
- **Processing of data having complex structure:** The data may be heterogeneous type that is unstructured to structured. So difficult to process.

3. Management Challenges:

It deals with the security, governance, sharing of data , operational expenses and data rights.

- **Maintaining Security, integrity and privacy:** The collected data from different sources and organizations uses various tools to process such data so data should be kept secure and integrity, privacy of data must be maintained during processing.
- **Capture:** How to quickly capture the high volume of data created from various sources. Data is growing faster speed than writing capacity of disks.
- **Storages of exponentially increasing data:** Data streams are coming very fast and huge. Where to store it. It is difficult to store the whole data. Data created in previous 2 years is more that total data created in history. Till 2025, it is expected that 463 exabytes of data will be generated every day worldwide which is equal to 212765957 DVDs .The solution for it is Hadoop Distributed File System (HDFS).
- **Sharing:** It is a challenge to share the data that is huge in volume. It will require huge bandwidth.
- **Analysis:** In big data most of the time we do not know the type of dealing data. So analysis is difficult.

4. New Technology to Learn:

There are a lot of technology to learn in depth and also changing very fast. Right and skilled people are still missing and very few technology master experts are available today. Solution for the major challenge of storage of big data may be the hybrid processing that is some data is processed at cloud and some on premises. Data lakes can be used as a cheap storage for data optimized algorithms can be used to reduce huge computing power.To manage the diverse, complex and unreliable data must be cleaned using some powerful existing tools and for sharing of data map reduce can be a solution.

5. Need For Synchronization Across Disparate Data Sources:

As data sets are becoming bigger and more diverse, there is a big challenge to incorporate them into an analytical platform. If this is overlooked, it will create gaps and lead to wrong messages and insights.

6. Acute Shortage Of Professionals Who Understand Big Data Analysis :

The analysis of data is important to make this voluminous amount of data being produced in every minute, useful. With the exponential rise of data, a huge demand for big data scientists and Big Data analysts has been created in the market. It is important for business organizations to hire a data scientist having skills that are varied as the job of a data scientist is multidisciplinary. Another major challenge faced by businesses is the shortage of professionals who understand Big Data analysis. There is a sharp shortage of data scientists in comparison to the massive amount of data being produced.

7. Getting Meaningful Insights Through The Use Of Big Data Analytics:

It is imperative for business organizations to gain important insights from Big Data analytics, and also it is important that only the relevant department has access to this information. A big challenge faced by the companies in the Big Data analytics is mending this wide gap in an effective manner.

8. Getting Voluminous Data Into The Big Data Platform :

It is hardly surprising that data is growing with every passing day. This simply indicates that business organizations need to handle a large amount of data on daily basis. The amount and variety of data available these days can overwhelm any data engineer and that is why it is considered vital to make data accessibility easy and convenient for brand owners and managers.

9. Uncertainty Of Data Management Landscape:

With the rise of Big Data, new technologies and companies are being developed every day. However, a big challenge faced by the companies in the Big Data analytics is to find out which technology will be best suited to them without the introduction of new problems and potential risks

10. Security And Privacy Of Data :

Once business enterprises discover how to use Big Data, it brings them a wide range of possibilities and opportunities. However, it also involves the potential risks associated with big data when it comes to the privacy and the security of the data. The Big Data tools used for analysis and storage utilizes the data disparate sources. This eventually leads to a high risk of exposure of the data, making it vulnerable. Thus, the rise of voluminous amount of data increases privacy and security concerns.

ANALYTIC PROCESS AND TOOLS

There are 6 analytic processes:

1. Deployment :

Here we need to: – plan the deployment and monitoring and maintenance, – we need to produce a final report and review the project. – In this phase, • we deploy the results of the analysis. • This is also known as reviewing the project.

2. Business Understanding:

– The very first step consists of business understanding. – Whenever any requirement occurs, firstly we need to determine the business objective, – assess the situation, – determine data mining goals and then – produce the project plan as per the requirement. • Business objectives are defined in this phase

3. Data Exploration:

• The second step consists of Data understanding. – For the further process, we need to gather initial data, describe and explore the data and verify data quality to ensure it contains the data we require. – Data collected from the various sources is described in terms of its application and the need for the project in this phase. – This is also known as data exploration. • This is necessary to verify the quality of data collected.

4. Data Preparation:

• From the data collected in the last step, – we need to select data as per the need, clean it, construct it to get useful information and – then integrate it all. • Finally, we need to format the data to get the appropriate data. •

Data is selected, cleaned, and integrated into the format finalized for the analysis in this phase

5. Data Modeling :

- we need to – select a modeling technique, generate test design, build a model and assess the model built.
- The data model is build to – analyze relationships between various selected objects in the data, – test cases are built for assessing the model and model is tested and implemented on the data in this phase.



Benefits of Using Big Data Analytics:

- Identifying the root causes of failures and issues in real time
- Fully understanding the potential of data-driven marketing
- Generating customer offers based on their buying habits
- Improving customer engagement and increasing customer loyalty
- Reevaluating risk portfolios quickly
- Personalizing the customer experience
- Adding value to online and offline customer interactions
- Big Data has to deal with large and complex datasets that can be structured, Semi-structured, or unstructured and will typically not fit into memory to be Processed.

4- Big Data Analytics Applications :

Nowadays, as the growing generation of available data is a recognized trend across enterprises, countries and market segments, the majority of enterprises regardless industry is collecting, storing and analyzing data in order to capture value. Digital economy through the tremendous use of internet and digital services has transformed almost all the industry sectors, including agriculture and manufacturing, to more service-centered. There are many and different sectors, like e-commerce politics, science & technology, health, government services etc.



Some examples of the ways BDA are exploited showing the significance of analytics in various themes :

Marketing	Market basket analysis	Recommendation systems	Customer Intelligence	Retention modeling	Customer churn prediction
Processes	Supply chain analytics	Demand and supply forecasting	Business Processes analytics	HR analytics	
Government	Fraud detection	Terrorism Detection	Tax avoidance	Cost reduction	Social security
Risk Management	Credit risk modeling	Market risk modeling	Fraud detection		
Web and Social media	Web analytics	Social media analytics	Multivariate testing		

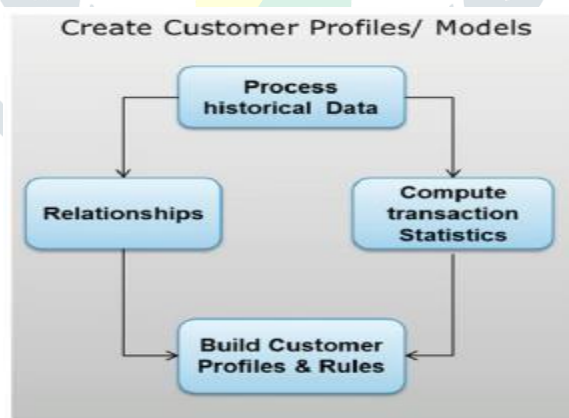
1 Applications in Banking and Financial Industries:

Massive amounts of data are being generated by the banking and financial industries through their various service offerings such as checking/savings accounts, mobile banking, credit and debit cards, loans, insurance, and investment services. Most of these data are structured data. Also, most of these organizations have set up their presence online for better serviceability and marketing through which lots of data are collected.



2 Fraud Detection :

Various surveys and studies indicate that banking and financial services industry is the victim of the most of the fraud cases among various industries.



3 Customer Churn :

It is well known that the customer churn is a big headache for the all the telecom service providers. Customers leaving the existing service provider and signing up with a competitor cause revenue/profit/loss. It is a costly affair to acquire new customers with new promotions and has an effect of increased marketing costs which in turn has the effect on profitability.

4 Applications in Telecommunications :

With the expansion of telecommunications services across the globe, the telecommunications industry is trying to penetrate various markets with diverse service offerings in voice, video, and data. With the development of new

technologies and services across multiple countries, the market is growing rapidly and has become highly competitive between various service providers.

5 Applications in Manufacturing :

Manufacturing companies have become highly competitive across the world with the margins of doing business going down every day. The manufactures are always on the lookout for optimizing costs in running factories thereby increasing the margins.

5-Conclusions :

The growth of Internet with the beginning of Web 2.0 era enabled companies getting access to big amounts of data easier and cheaper, while the opportunities for external data collection have even increased with the appearance of the Web 3.0.

This chapter gives basics of big data analytics and talks about tool and technologies like hadoop, map reduce, Hbase,Hive etc. for processing the huge volume of heterogeneous data in real-time in a efficient way to extract the desired insights.

This chapter highlighted the past as well as present challenges of big data analytics and also proposed some solutions for the same. After that it has given the existing areas where efforts can be made by the researchers for their improvements in near future.

Big data has become a new era for every economy, industry and organizations. Through the scrutiny of huge volume of data that are becoming available, there is the possibility for making faster advances in various technical areas. Big data analysis is becoming essential for automatic discovering of intelligence for the occurrence of patterns and hidden rules. Big data analysis makes it easier for companies in decision making, predicting and identifying the new opportunities. In this paper we have discussed about the issues and challenges of big data and also big data analysis tools which help researchers for extracting useful knowledge out of big data.

REFERENCES:

6-

01. Lodefalk, M.: Servicification of manufacturing—evidence from Sweden. *Int. J. Econom. Bus. Res.* 6(1), 87 (2013)
02. Analytics 3.0: Harvard Business Review, 01 Dec 2013. <https://hbr.org/2013/12/analytics-30>. Accessed 21 Jun2017
03. Davenport, T.: *Big Data at Work: Dispelling the Myths*. Harvard Business Review Press, Uncovering the Opportunities (2014)
04. Chen, H., Chiang, R., Storey, V.C.: Business intelligence and analytics: from big data to big impact. *Miss. Q.* 36(4), 1165–1188 (2012)
05. Sivarajah, U., Kamal, M.M., Irani, Z., Weerakkody, V.: Critical analysis of big data challenges and analytical methods. *J. Bus. Res.* 70, 263–286 (2017)
06. M. K.Kakhani, S. Kakhani and S. R.Biradar, Research issues in big data analytics, *International Journal of Application or Innovation in Engineering & Management*, 2(8) (2015), pp.228-232
07. A. Gandomi and M. Haider, Beyond the hype: Big data concepts, meth- ods, and analytics, *International Journal of Information Management*, 35(2) (2015), pp.137-144.
08. Bhadani, A., Jothimani, D., “ Big data: Challenges, opportunities and realities” in the journal of Effective Big Data Management and Opportunities for Implementation (pp. 1-24), Pennsylvania, USA, IGI Global,2017.
09. Mohd Vasim Ahamad, Misbahul Haque,“Insight Into Big Data Analytics:Challenges, Recent Trends and Future Prospects.
10. Apache Spark Features <http://spark.apache.org/>
11. Big data Analytics Tools<https://www.qubole.com/big-data-analytics/>