IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

BIG DATA ANALYTICS: TOOLS AND TECHNIQUES

Miss Divya S. Belekar
Assistant Professor in Department of Computer Science
KavikulguruKalidas Sanskrit University, Ramtek

Abstract : Big data analyticsis all about extracting valuable insights and patterns from large sets of data. It involves using various tools and methods to analyze and interpret this data. Some popular tools for big data analytics include Hadoop, Spark, and Apache flink. As for Techniques, there are different approaches like Data fusion and data integration, A/B testing, data mining , machine learning. These techniques help business make informed decision and discover hidden trends. This paper attempts to offer a broader definition of Big data analytics , it's tools and technique with there characteristics and application.

Keywords : Big data analytics, Tools, Techniques , Characteristics , Application, Process.

I. INTRODUCTION

Big data analytics is a field that focuses on extracting meaningful insights from large and complex datasets. It involves using various tools and techniques to analyze, process, and interpret this data.
When it comes to tools , popular ones include Hadoop, Spark , and Apache flink. These tools provide the infrastructure and frameworks to handle the immense volume, velocity, and variety of big data. They enable distributed computing, parallel processing, and efficient data storage, allowing organizations to tackle big data challenges effectively.

The techniques used in big data analytics such as descriptive analytics involve summarizing and visualizing data to gain a better understanding of past events and trends. Predictive analytics uses historical data to make prediction and forecast about future events. Prescriptive analytics goes beyond prediction and provide recommendations on what actions to take based on the data. Text analytics focuses on extracting insights from unstructured textual data.
By leveraging these tools and techniques, organizations can uncover valuable insights. Make data-driven decision, and gain a competitive advantage in today's data-driven world.

This paper documents the basic concept of big data analytics , there characteristics and application. The key contribution of this paper is that stud about the different tools and techniques used in big data processing.

This paper is organized is as follows. Paper begin with defining big data definitions given by different authors. Then it highlight to its characteristics and applications. Then it expands the discussion on various types of big data. Then it uncover the different tools and techniques used to process the big data.

II. BIG DATA ANALYTICS

Big data refers to the large, diverse sets of information that grow at ever-increasing rates. It encompasses the volume of information, the velocity or speed at which it is created and collected, and the variety or scope of the data points being covered. Big data often comes from data mining and arrives in multiple formats.

Big data analytics refers to the methods, tools, and applications used to collect, process, and derive insights from varied, high-volume, high-velocity data sets. These data sets may come from a variety of sources, such as web, mobile, email, social media, and networked smart devices. They often feature data that is generated at a high speed and varied in form, ranging from structured (database tables, Excel sheets) to semi-structured (XML files, webpages) to unstructured (images, audio files).
Traditional forms of data analysis software aren't equipped to support this level of complexity and scale, which is where the systems, tools, and applications designed specifically for big data analysis come into play.

Nowadays, the data that need to be analyzed are not just large, but they are composed of various data types, and even including streaming data [1]. Since big data has the unique features of "massive, high dimensional, heterogeneous, complex, unstructured, incomplete, noisy, and erroneous," which may change the statistical and data analysis approaches [2]. Although it seems that big data makes it possible for us to collect more data to find more useful information, the truth is that more data do not necessarily mean more useful information. It may contain more ambiguous or abnormal data. For instance, a user may have multiple accounts, or an account may be used by multiple users, which may degrade the accuracy of the mining results [3]. Therefore, several new issues for data analytics come up, such as privacy, security, storage, fault tolerance, and quality of data [4].

III. CHARACTERISTICS

Big data is data whose scale, distribution, diversity, and/or timeliness require the use of new technical architectures, analytics, and tools in order to enable insights that unlock new sources of business value. Three main features characterize big data: volume, variety, and velocity, or the three V's. The volume of the data is its size, and how enormous it is. Velocity refers to the rate with which data is changing, or how often it is created. Finally, variety includes the different formats and types of data, as well as the different kinds of uses and ways of analyzing the data [4]. Data volume is the primary attribute of big data. Big data can be quantified by size in TBs or PBs, as well as even the number of records, transactions, tables, or files. Additionally, one of the things that make big data really big is that it's coming from a greater variety of sources than ever before, including logs, clickstreams, and social media. Using these sources for analytics means that common structured data is now joined by unstructured data, such as text and human language, and semi-structured data, such as eXtensible Markup Language (XML) or Rich Site Summary (RSS) feeds. There's also data, which is hard to categorize since it comes from audio, video, and other devices. Furthermore, multi-dimensional data can be drawn from a data warehouse to add historic context to big data. Thus, with big data, variety is just as big as volume.

Moreover, big data can be described by its velocity or speed. This is basically the frequency of data generation or the frequency of data delivery. The leading edge of big data is streaming data, which is collected in real-time from the websites [5]. Some researchers and organizations have discussed the addition of a fourth V, or veracity. Veracity focuses on the quality of the data. This characterizes big data quality as good, bad, or undefined due to data inconsistency, incompleteness, ambiguity,latency, deception, and approximations [6]

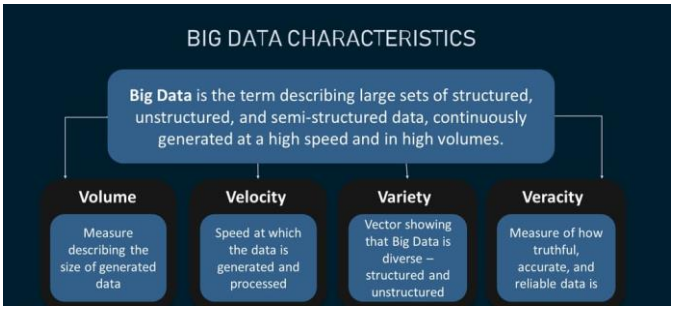

Figure 1 : Characteristics of Big Data

Again some researchers and organization have discover new 3-V's ,which is Variability , Value and Visualization. So now there are 7-V's we have found.

The Seven V's of Big Data Analytics are Volume, Velocity, Variety, Variability, Veracity, Value, and Visualization. This framework offers a model for working with large and complex data sets. Comprehending these dimensions is crucial for developing an effective big data strategy that enables the management, analysis, and extraction of valuable business insights from large datasets.[8]

3.1. Volume

In the current scenario, the amount of data that companies possess matters. For big data analytics, you will need to process higher volumes of structured and unstructured data. This data can be of indefinite value, such as Facebook and Instagram datasets, or data on numerous web or mobile applications. As per the market trends, the volume of data will upsurge considerably in the coming years, and there is a lot of room for extensive data analysis and pattern-finding.

3.2. Velocity

Velocity refers to the swiftness of data processing. A higher data processing rate is significant for any big data procedure's real-time evaluation and performance. More data will be accessible in the future, but the processing speed will be equally important for companies to benefit from big data analytics.

3.3. Variety

Variety refers to the diverse categories of big data. It is among the prime challenges the big data industry faces as it impacts productivity.

3.4. Veracity

Veracity denotes the precision of your data. It is essential as low veracity can negatively impact the accuracy of your big data analytics results.

3.5. Variability

Big data often contains noisy and incomplete data points, which can obscure valuable insights. Addressing this variability typically involves data cleaning and validation processes to ensure data quality.

3.6. Visualization

Visualization plays a vital role in data analytics, as it involves presenting the analyzed data in a visually comprehensible manner. When planning data visualization, it is essential to consider the end user and the decisions the visualizations aim to support. Well-executed data visualization facilitates swift and well-informed decision-making.[8]

3.7. Value

A successful big data analytics strategy must generate value. The insights derived from the analysis should provide meaningful guidance for improving operations, enhancing customer service, or creating other forms of value. An integral part of developing a big data analytics strategy is distinguishing between data that can contribute value and data that cannot.[8]

With the rising usage of big data, data comes in new data groups. Different data categories, like text, audio, and video, need extra pre-processing to back metadata and derive enhanced value.[7]

IV. APPLICATION

Big data has increased the demand of information management specialists so much so that Software AG, Oracle Corporation, IBM, Microsoft, SAP, EMC, HP and Dell have spent more than $15 billion on 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.[10]

The global big data analytics market size was valued at $307.51 billion in 2023 & is projected to grow from $348.21 billion in 2024 to $924.39 billion by 2032.

Application area of Big Data Analytics are Government ,International development , Manufacturing , Healthcare , Education and Media.[11]

V. TOOLS

As we're growing with the pace of technology, the demand to track data is increasing rapidly. Today, almost 2.5 quintillion bytes of data are generated globally and it's useless until that data is segregated in a proper structure. It has become crucial for businesses to maintain consistency in the business by collecting meaningful data from the market today and for that, all it takes is the right data analytic tool and a professional data analyst to segregate a huge amount of raw data by which then a company can make the right approach.[12]

Large numbers of tools are available to process big data. In this section, we discuss some current techniques for analyzing big data with emphasis on 10 important emerging tools. Most of the available tools concentrate on batch processing, stream processing,and interactive analysis. Most batch processing tools are based on the Apache Hadoop infrastructure . Stream data applications are mostly used for real time analytic. Some examples of large scale streaming platform are Strom. The interactive analysis process allow users to directly interact in real time for their own analysis.

5.1. APACHE Hadoop

It's a Java-based open-source platform that is being used to store and process big data. It is built on a cluster system that allows the system to process data efficiently and let the data run parallel. It can process both structured and unstructured data from one server to multiple computers. Hadoop also offers cross-platform support for its users. Today, it is the best big data analytic tool and is popularly used by many tech giants such as Amazon, Microsoft, IBM, etc.

Features of Apache Hadoop:

Free to use and offers an efficient storage solution for businesses. Offers quick access via HDFS (Hadoop Distributed File System). Highly flexible and can be easily implemented with MySQL, and JSON. Highly scalable as it can distribute a large amount of data in small segments. It works on small commodity hardware like JBOD or a bunch of disks.

5.2. Cassandra

APACHE Cassandra is an open-source NoSQL distributed database that is used to fetch large amounts of data. It's one of the most popular tools for data analytics and has been praised by many tech companies due to its high scalability and availability without compromising speed and performance. It is capable of delivering thousands of operations every second and can handle petabytes of resources with almost zero downtime. It was created by Facebook back in 2008 and was published publicly.

Features of APACHE Cassandra:

Data Storage Flexibility: It supports all forms of data i.e. structured, unstructured, semi-structured, and allows users to change as per their needs.

Data Distribution System: Easy to distribute data with the help of replicating data on multiple data centers.

Fast Processing: Cassandra has been designed to run on efficient commodity hardware and also offers fast storage and data processing.

Fault-tolerance: The moment, if any node fails, it will be replaced without any delay.

5.3. Qubole

It's an open-source big data tool that helps in fetching data in a value of chain using ad-hoc analysis in machine learning. Qubole is a data lake platform that offers end-to-end service with reduced time and effort which are required in moving data pipelines. It is capable of configuring multi-cloud services such as AWS, Azure, and Google Cloud. Besides, it also helps in lowering the cost of cloud computing by 50%.

Features of Qubole:

Supports ETL process: It allows companies to migrate data from multiple sources in one place.
Real-time Insight: It monitors user's systems and allows them to view real-time insights
Predictive Analysis: Qubole offers predictive analysis so that companies can take actions accordingly for targeting more acquisitions.
Advanced Security System: To protect users' data in the cloud, Qubole uses an advanced security system and also ensures to protect any future breaches. Besides, it also allows encrypting cloud data from any potential threat.

5.4. Xplenty

It is a data analytic tool for building a data pipeline by using minimal codes in it. It offers a wide range of solutions for sales, marketing, and support. With the help of its interactive graphical interface, it provides solutions for ETL, ELT, etc. The best part of using Xplenty is its low investment in hardware & software and its offers support via email, chat, telephonic and virtual meetings. Xplenty is a platform to process data for analytics over the cloud and segregates all the data together.

Features of Xplenty:

Rest API: A user can possibly do anything by implementing Rest API
Flexibility: Data can be sent, and pulled to databases, warehouses, and salesforce.
Data Security: It offers SSL/TSL encryption and the platform is capable of verifying algorithms and certificates regularly.
Deployment: It offers integration apps for both cloud & in-house and supports deployment to integrate apps over the cloud.

5.5. Spark

APACHE Spark is another framework that is used to process data and perform numerous tasks on a large scale. It is also used to process data via multiple computers with the help of distributing tools. It is widely used among data analysts as it offers easy-to-use APIs that provide easy data pulling methods and it is capable of handling multi-petabytes of data as well. Recently, Spark made a record of processing 100 terabytes of data in just 23 minutes which broke the previous world record of Hadoop (71 minutes). This is the reason why big tech giants are moving towards spark now and is highly suitable for ML and AI today.

Features of APACHE Spark:

Ease of use: It allows users to run in their preferred language. (JAVA, Python, etc.) Real-time Processing: Spark can handle real-time streaming via Spark Streaming Flexible: It can run on, Mesos, Kubernetes, or the cloud.

5.6. Mongo DB

Came in limelight in 2010, is a free, open-source platform and a document-oriented (NoSQL) database that is used to store a high volume of data. It uses collections and documents for storage and its document consists of key-value pairs which are considered a basic unit of Mongo DB. It is so popular among developers due to its availability for multi-programming languages such as Python, Jscript, and Ruby.

Features of Mongo DB:

Written in C++: It's a schema-less DB and can hold varieties of documents inside.
Simplifies Stack: With the help of mongo, a user can easily store files without any disturbance in the stack.
Master-Slave Replication: It can write/read data from the master and can be called back for backup.

5.7. Apache Storm

A storm is a robust, user-friendly tool used for data analytics, especially in small companies. The best part about the storm is that it has no language barrier (programming) in it and can support any of them. It was designed to handle a pool of large data in fault-tolerance and horizontally scalable methods. When we talk about real-time data processing, Storm leads the chart because of its distributed real-time big data processing system, due to which today many tech giants are using APACHE Storm in their system. Some of the most notable names are Twitter, Zendesk, NaviSite, etc.

Features of Storm:
Data Processing: Storm process the data even if the node gets disconnected
Highly Scalable: It keeps the momentum of performance even if the load increases
Fast: The speed of APACHE Storm is impeccable and can process up to 1 million messages of 100 bytes on a single node.

5.8. SAS

Today it is one of the best tools for creating statistical modeling used by data analysts. By using SAS, a data scientist can mine, manage, extract or update data in different variants from different sources. Statistical Analytical System or SAS allows a user to access the data in any format (SAS tables or Excel worksheets). Besides that it also offers a cloud platform for business analytics called SAS Viya and also to get a strong grip on AI & ML, they have introduced new tools and products.

Features of SAS:
Flexible Programming Language: It offers easy-to-learn syntax and has also vast libraries which make it suitable for non-programmers
Vast Data Format: It provides support for many programming languages which also include SQL and carries the ability to read data from any format.
Encryption: It provides end-to-end security with a feature called SAS/SECURE.

5.9. Data Pine

Datapine is an analytical used for BI and was founded back in 2012 (Berlin, Germany). In a short period of time, it has gained much popularity in a number of countries and it's mainly used for data extraction (for small-medium companies fetching data for close monitoring). With the help of its enhanced UI design, anyone can visit and check the data as per their requirement and offer in 4 different price brackets, starting from $249 per month. They do offer dashboards by functions, industry, and platform.

Features of Datapine:
Automation: To cut down the manual chase, datapine offers a wide array of AI assistant and BI tools.
Predictive Tool: datapine provides forecasting/predictive analytics by using historical and current data, it derives the future outcome.
Add on: It also offers intuitive widgets, visual analytics & discovery, ad hoc reporting, etc.

5.10. Rapid Miner

It's a fully automated visual workflow design tool used for data analytics. It's a no-code platform and users aren't required to code for segregating data. Today, it is being heavily used in many industries such as ed-tech, training, research, etc. Though it's an open-source platform but has a limitation of adding 10000 data rows and a single logical processor. With the help of Rapid Miner, one can easily deploy their ML models to the web or mobile (only when the user interface is ready to collect real-time figures).

Features of Rapid Miner:
Accessibility: It allows users to access 40+ types of files (SAS, ARFF, etc.) via URL
Storage: Users can access cloud storage facilities such as AWS and dropbox
Data validation: Rapid miner enables the visual display of multiple results in history for better evaluation.[13]

VI. TECHNIQUES

6.1. Types of big data analytics

There are four main types of big data analytics—descriptive, diagnostic, predictive, and prescriptive. Each serves a different purpose and offers varying levels of insight.

Collectively, they enable businesses to comprehensively understand their big data and make decisions to drive improved performance.

6.1.1. Descriptive analytics

This type focuses on summarizing historical data to tell youwhat's happened in the past. It uses aggregation, data mining, and visualization techniques to understand trends, patterns, and key performance indicators (KPIs).

Descriptive analytics helps you understand your current situation and make informed decisions based on historical information.

6.1.2. Diagnostic analytics

Diagnostic analytics goes beyond describing past events and aims to understand why they occurred. It separates data to identify the root causes of specific outcomes or issues.

By analyzing relationships and correlations within the data, diagnostic analytics helps you gain insights into factors influencing your results.

6.1.3. Predictive analytics

This type of analytics uses historical data and statistical algorithms to predict future events. It spots patterns and trends and forecasts what might happen next.

You can use predictive analytics to anticipate customer behavior, product demand, market trends, and more to plan and make strategic decisions proactively.

6.1.4. Prescriptive analytics

Prescriptive analytics builds on predictive analytics by recommending actions to optimize future outcomes. It considers various possible actions and their potential impact on the predicted event or outcome.

Prescriptive analytics help you make data-driven decisions by suggesting the best course of action based on your desired goals and any constraints.[14]

6.2. Six big data analysis techniques

Big data is characterised by the three V's: the major volume of data, the velocity at which it's processed, and the wide variety of data.[15] It's because of the second descriptor, velocity, that data analytics has expanded into the technological fields of machine learning and artificial intelligence.[16]Alongside the evolving computer-based analysis techniques data harnesses, analysis also relies on the traditional statistical methods.[17] Ultimately, how data analysis techniques function within an organisation is twofold; big data analysis is processed through the streaming of data as it emerges, and then performing batch analysis' of data as it builds – to look for behavioural patterns and trends.[18] As the generation of data increases, so will the various techniques that manage it. As data becomes more insightful in its speed, scale, and depth, the more it fuels innovation.

The world is driven by data, and it's being analysed every second, whether it's through your phone's Google Maps, your Netflix habits, or what you've reserved in your online shopping cart.

McKinsey's big data report identifies a range of big data techniques and technologies, that draw from various fields such as statistics, computer science, applied mathematics, and economics.[19] As these methods rely on diverse disciplines, the analytics tools can be applied to both big data and other smaller datasets:

6.2.1. A/B testing

This data analysis technique involves comparing a control group with a variety of test groups, in order to discern what treatments or changes will improve a given objective variable. McKinsey gives the example of analysing what copy, text, images, or layout will improve conversion rates on an e-commerce site.[20] Big data once again fits into this model as it can test huge numbers, however, it can only be achieved if the groups are of a big enough size to gain meaningful differences.

6.2.2. Data fusion and data integration

By combining a set of techniques that analyse and integrate data from multiple sources and solutions, the insights are more efficient and potentially more accurate than if developed through a single source of data.

6.2.3. Data mining

A common tool used within big data analytics, data mining extracts patterns from large data sets by combining methods from statistics and machine learning, within database management. An example would be when customer data is mined to determine which segments are most likely to react to an offer.

6.2.4. Machine learning

Well known within the field of artificial intelligence, machine learning is also used for data analysis. Emerging from computer science, it works with computer algorithms to produce assumptions based on data. It provides predictions that would be impossible for human analysts.

6.2.5. Natural language processing (NLP).

Known as a subspecialty of computer science, artificial intelligence, and linguistics, this data analysis tool uses algorithms to analyse human (natural) language.[21]

6.2.6. Statistics.

This technique works to collect, organise, and interpret data, within surveys and experiments.[22]

Other data analysis techniques include spatial analysis, predictive modelling, association rule learning, network analysis and many, many more. The technologies that process, manage, and analyse this data are of an entirely different and expansive field, that similarly evolves and develops over time. Techniques and technologies aside, any form or size of data is valuable. Managed accurately and effectively, it can reveal a host of business, product, and market insights. What does the future of data analysis look like? It's hard to say with the tremendous pace analytics and technology progresses, but undoubtedly data innovation is changing the face of business and society in its holistic entirety.

VII. PROCESSING

Big data analytics refers to collecting, processing, cleaning, and analyzing large datasets to help organizations operationalize their big data.

7.1. Collect Data

Data collection looks different for every organization. With today's technology, organizations can gather both structured and unstructured data from a variety of sources — from cloud storage to mobile applications to in-store IoT sensors and beyond. Some data will be stored in data warehouses where business intelligence tools and solutions can access it easily. Raw or unstructured data that is too diverse or complex for a warehouse may be assigned metadata and stored in a data lake.

7.2. Process Data

Once data is collected and stored, it must be organized properly to get accurate results on analytical queries, especially when it's large and unstructured. Available data is growing exponentially, making data

processing a challenge for organizations. One processing option is **batch processing**, which looks at large data blocks over time. Batch processing is useful when there is a longer turnaround time between collecting and analyzing data. **Stream processing** looks at small batches of data at once, shortening the delay time between collection and analysis for quicker decision-making. Stream processing is more complex and often more expensive.

7.3. Clean Data

Data big or small requires scrubbing to improve data quality and get stronger results; all data must be formatted correctly, and any duplicative or irrelevant data must be eliminated or accounted for. Dirty data can obscure and mislead, creating flawed insights.

7.4. Analyze Data

Getting big data into a usable state takes time. Once it's ready, advanced analytics processes can turn big data into big insights. Some of these big data analysis methods include:

- **Data mining** sorts through large datasets to identify patterns and relationships by identifying anomalies and creating data clusters.
- **Predictive analytics** uses an organization's historical data to make predictions about the future, identifying upcoming risks and opportunities.
- **Deep learning** imitates human learning patterns by using artificial intelligence and machine learning to layer algorithms and find patterns in the most complex and abstract data.[14]

VIII. IMPORTANCE

Big data analytics may look simple, but there are a large number of processes that are composed in it. We can think of big data as something that has huge volume, velocity, and variety. Big data analytics tools can make sense of the large volumes of data and convert them into valuable business insights.

The need for big data analytics comes from the fact that we are generating data at an extremely rapid pace, and every enterprise needs to make sense of this data. By the year 2020, we were generating a staggering 1.7 MB of data every second.

This indicates the importance of big data analytics for making sense of large volumes of data. It helps us organize, transform, and model the data based on the requirements of an enterprise. It also allows us to identify patterns and draw conclusions from the data.

The larger the size of the data, the bigger the problem. So, big data may be defined as the data where its size itself poses the problem, and newer ways of handling the same are needed. The analysis of data that is high in volume, velocity, and variety means that the traditional methods of working with data are not applicable here.

IX. CONCLUSION

Big data has been in limelight for the past few years and will continue to dominate the market in almost every sector for every market size. The demand for big data is booming at an enormous rate and ample tools are available in the market today, all you need is the right approach and choose the best data analytic tool as per the project's requirement.

Aside from techniques and technologies, data of any size or form is valuable. When managed correctly and effectively, it can reveal a wealth of information about a company's business, products, and markets. Where does data analysis go from here? With the rapid progress of analytics and technology, it is difficult to say, but it is evident that data innovation has revolutionized business and society as a whole.

REFERENCES

[1] Russom P. Big data analytics. TDWI: Tech. Rep ; 2011.
[2] Ma C, Zhang HH, Wang X. Machine learning for big data analytics in plants. Trends Plant Sci. 2014;19(12):798–808.
[3] Boyd D, Crawford K. Critical questions for big data. Inform Commun Soc. 2012;15(5):662–79.
[4] EMC: Data Science and Big Data Analytics. In: EMC Education Services, pp. 1–508 (2012)
[5] Russom, P.: Big Data Analytics. In: TDWI Best Practices Report, pp. 1–40 (2011)
[6] TechAmerica: Demystifying Big Data: A Practical Guide to Transforming the Business of Government. In: TechAmerica Reports, pp. 1–40 (2012)
[7] JanahaVivek **:**Assistant Marketing Manager @ Zuci Systems.-2016
[8] TrigynTechnologies ,The Seven V's of Big Data Analytics
[9] Technology / Big data analytics market : April 8 , 2024 : Report ID : FBI106179
[10] SrikanthGangadhara , Associate Professor, CSE Department, VITS Karimnagar,Telangana @2018 JETIR September 2018, Volume 5, Issue 9, (ISSN-2349-5162)
[11] (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 2, 216
[12] https://www.geeksforgeeks.org/10-most-popular-big-data-analytics-tools
[13] https://www.naukri.com/code360/library/techniques-for-big-data-analysis
[14] https://intellipaat.com/blog/big-data-analytics.
[15] 'Big data. Retrieved fromSearch Data Management.
[16] 'Big data. Retrieved fromSearch Data Management.
[17] Erl, T & Khattak, W & Buhler, P. (2016). 'Big data fundamentals: Concepts, drivers & techniques'. Retrieved from Safari Books Online
[18] Erl, T & Khattak, W & Buhler, P. (2016). 'Big data fundamentals: Concepts, drivers & techniques'. Retrieved from Safari Books Online
[19] Manyika, J, et al. (2011). 'Big data: The next frontier for innovation, competition, and productivity'. Retrieved from McKinsey.
[20] Manyika, J, et al. (2011). 'Big data: The next frontier for innovation, competition, and productivity'. Retrieved from McKinsey.
[21] Reavie, V. (Aug, 2018). 'Do you know the difference between data analytics and AI machine learning?'. Retrieved from Forbes.
[22] Manyika, J, et al. (2011). 'Big data: The next frontier for innovation, competition, and productivity'. Retrieved from McKinsey.