**IOSR Journal of Engineering (IOSRJEN)** **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Special Issue || Dec. 2019 || PP 56-59
International Conference on Management Practices, Innovations & Research 2019

# Big Data Analytics: An Overview

Mr. Y. R Rochlani, Mr. V. S. Gangwani, Mr. R. G. Anantwar
*CSE Department, H.V.P.M's COET, Amravati*
*CSE Department, H.V.P.M's COET, Amravati*
*CSE Department, H.V.P.M's COET, Amravati*

***Abstract:*** *In recent years, the web applications on internet have seen a lot of development in the field of Information Technology. These online applications and communication are generating the large size data continuously having different variety and with difficult structure data called as big data. In order to analyse this huge data Big Data Analytics is required. Big Data Analytics is the study of huge amount of stored data in order to extract complex and heavier patterns. The combination of high technology systems and mathematics together are capable of analysing all this information, hence providing great value of information for companies or government.*

*This paper describes introduction to Big Data Analytics, Architecture, need of Big Data Analytics and types of Big Data Analytics.*

***Keywords:*** *Big Data, Big Data Analytics.*

## I. Introduction

In today's world enormous amount of data is generated every minute on the web through smart phones, tablets, sensors spread all over our cities and bank cards. According to recent survey around 2.5 trillion bytes data is generated globally every day and stored by public administration and private companies. Besides, cities are having network of sensors that clusters all those technologies and mathematical developments dedicated to store, analyze and cross-reference all that information to try and find behavioural patterns. What can be done with all this information? This is where Big Data Analytics come into play.

Gartner Company believes that, Information or data will be the 21st century oil. In last 25 years, data has grown enormously in various fields with different types. According to the statistical report of International Data Corporation (IDC), in the year 2011, the overall data volume created in the world was 1.8ZB that was enhanced by nearly nine times within next five years [1]. Now with the inclusion of marketing, smart city, the results of disease control, prevention and business intelligence applications it can be effortlessly understand that big data plays a vital role everywhere in the universe [2].

**The Five V's in Big Data:**

According to industry the big data is articulated in following five V's:

***1) Volume (Data in Rest*)**: Organizations collect data from a variety of data sources, including commercial transactions, social media data and information from sensors or machine-to-machine data.

***2) Velocity (Data in Motion*)**: Data streams come in at unmatched speed and should be allocated with in an appropriate manner. Different kind of IoT sensors, RFID tags and smart metering are driving the necessity to deal with data flows in real time.

***3) Variety (Data in Many Forms*)**: Data comes in different kinds of formats such as structured, numeric data in traditional databases to unstructured text documents, email, video, audio, stock and financial transactions.

***4) Variability (Data in Highlight)*****:** Inconsistency of the data set can hamper processes to handle and manage it.

***5) Veracity (Data in Doubt)*****:** Refers to the messiness or trustworthiness of the data. The quality of captured data can vary greatly, affecting accurate analysis.

All major IT companies, including EMC, Microsoft, Google, Amazon, and Facebook, etc. already have started their big data projects. To extract patterns of information or data from big data, optimal processing power, analytics capabilities and skills are needed [5]. So, dealing the big data effectively requires generating the value against the volume, variety and veracity of big data [7].

## II. Big Data Analytics Architecture:

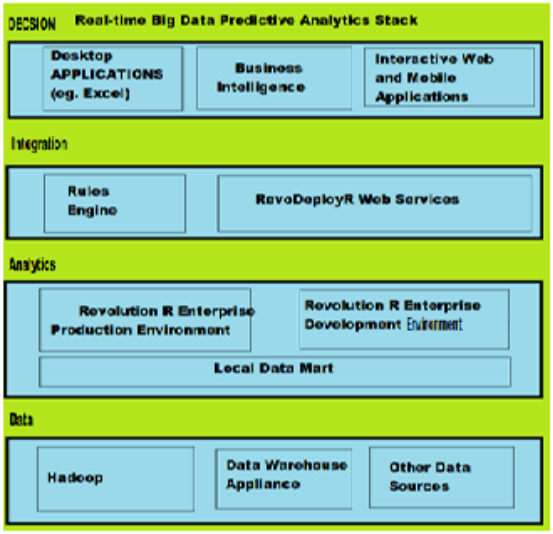


**Fig. 1 Layered Architecture of Big Data Analytics**

The implementation Layers are as follows:

***1) Data Layer***

This layer has Relational DBMS based structured, Semi-structured and unstructured based data. NoSQL databases are used to store the unstructured data. For example, MongoDB and Cassandra are the famous NoSQL databases. Fetching data from the web world, social media domain and data from IoT sensors are the examples to unstructured and semi-structured data. Software tools such as HBase, Hive, and HBase are also available at this layer. Hadoop and Map Reduce can also be used to support this layer.

***2) Analytics Layer***

Analytics layer has the environment to implement the dynamic data analytics and deploy the real time values. It has building models like producing and developing environment and has facility to modify the local data in regular interval. This process also improves the performance of the analytical engine.

***3) Integration Layer***

This layer integrates the end user applications and analytical engine. This includes usually a rules engine and an API for dynamic data analytics.

***4) Decision Layer***

This layer is where the end product is available in the market. This layer includes various applications of end user such as mobile applications, desktop applications, interactive web applications and business intelligence software. This is the layer where end users interact with the system. Each and every layer described here is associated with different sets of end users in real time and enables a important phase of real time data analytics implementation.

**Need of Big Data Analytics:**

Big data is basically defined by the huge volume of a data set. These data sets are generally huge – measuring tens of terabytes and even crossing the threshold of petabytes. The term big data was preceded by very large databases (VLDBs) which were managed using database management systems (DBMS). Today, big data falls under three categories of data sets – structured, unstructured and semi-structured.

- **Structured data sets** comprise of data which can be used in its original form to derive results. Examples include relational data such as employee salary records. Most modern computers and applications are programmed to generate structured data in preset formats to make it easier to process.

- **Unstructured data sets**, on the other hand, are without proper formatting and alignment. Examples include human texts, Google search result outputs, etc. These random collections of data sets require more processing power and time for conversion into structured data sets so that they can help in deriving tangible

results. Semi-Structured data sets are a combination of both structured and unstructured data. These data sets might have a proper structure and yet lack defining elements for sorting and processing. Examples include RFID and XML data.

- **Semi-Structured data sets** are a combination of both structured and unstructured data. These data sets might have a proper structure and yet lack defining elements for sorting and processing. Examples include RFID and XML data.

Big data processing requires a particular setup of physical and virtual machines to derive results. The processing is done simultaneously to achieve results as quickly as possible. These days, big data processing techniques also include Cloud Computing and Artificial Intelligence. These technologies help in reducing manual inputs and oversight by automating many processes and tasks.
The evolving nature of big data has made it difficult to give it a commonly accepted definition. Data sets are consigned the big data status based on technologies and tools required for their processing.

**Types of Big Data Analytics:**

- **Prescriptive analytics:** This type of analysis reveals what actions should be taken. This is the most valuable kind of analysis and usually results in rules and recommendations for next steps.

- **Predictive analytics:** An analysis of likely scenarios of what might happen. The deliverables are usually a predictive forecast.

The most commonly used technique; predictive analytics use models to forecast what might happen in specific scenarios. Examples of predictive analytics include next best offers, churn risk and renewal risk analysis.

- Forward looking
- Focused on non-discrete predictions of future states, relationship, and patterns
- Description of prediction result set probability distributions and likelihoods
- Model application
- Non-discrete forecasting (forecasts communicated in probability distributions)

- **Diagnostic analytics:** A look at past performance to determine what happened and why. The result of the analysis is often an analytic dashboard.

Data scientists turn to this technique when trying to determine why something happened. It is useful when researching leading churn indicators and usage trends amongst your most loyal customers. Examples of diagnostic analytics include churn reason analysis and customer health score analysis. Key points:

- Backward looking
- Focused on causal relationships and sequences
- Relative ranking of dimensions/variable based on inferred explanatory power)
- Target/dependent variable with independent variables/dimensions
- Includes both frequentist and Bayesian causal inferential analyses

- **Descriptive analytics:** What is happening now based on incoming data. To mine the analytics, you typically use a real-time dashboard and/or email reports. This technique is the most time-intensive and often produces the least value; however, it is useful for uncovering patterns within a certain segment of customers. Descriptive analytics provide insight into what has happened historically and will provide you with trends to dig into in more detail. Examples of descriptive analytics include summary statistics, clustering and association rules used in market basket analysis. Key points:

- Backward looking
- Focused on descriptions and comparisons
- Pattern detection and descriptions
- MECE (mutually exclusive and collectively exhaustive) categorization
- Category development based on similarities and differences (segmentation)

## III. Conclusion

In this paper, we studied the concept of Big Data and Big Data Analytics. Also we have seen layered architecture of Big Data Analytics and various types of Big Data Analytics including Prescriptive, Predictive and Descriptive Analysis. In this manner, we have seen importance of Big Data Analytics in online web applications in order to analyze huge amount of data.

## References


[1]. Shuhui Jiang, Xueming Qian, Tao Mei, Yun Fu, Personalized Travel Sequence recommendation on Multisource Big Social Media, 2016, IEEE Transactions on Big Data,Vol.2, Issue:1

[2]. Vallabh Dhoot, Shubham Gawande, Pooja Kanawade and Akanksha Lekhwani, Efficient Dimensionality Reduction for Big Data Using Clustering Technique, Imperial Journal of Interdisciplinary Research (IJIR), Vol-2, Issue-5, 2016, ISSN: 2454-1362

[3]. Cheikh Kacfah Emani, Nadine Cullot, Christophe Nicolle, Understandable Big Data: A survey, Mobile New

[4]. Applications 2014, 171-209

[5]. Kitchin R. The real-time city? Big data and smart urbanism. Geo J. 2014, 79(1), pp: 1–14.

[6]. Katrina Sin and Loganathan Muthu, Applications of big data in education data mining and learning analytics – A literature Review, ICTACT Journal on soft computing special issue on soft computing models for big data, July 2015, Vol:05, Iss: 04, pp: 1035-1049

[7]. Cheikh Kacfah Emani, Nadine Cullot, Christophe Nicolle, Understandable Big Data: A Survey , Computer Science Review, 2015, Vol: 17, pp: 71-80

[8]. K. Krishnan, Data warehousing in the age of big data, in: The Morgan Kaufmann Series on Business Intelligence, Elsevier Science, 2013.

[9]. H.V. Jagadish, D. Agarwal, P.Bernstein, Challenges and Opportunities in Big Data, The Community Research Association, 2015

[10]. K. Krishnan, Data warehousing in the age of big data, in the Morgan Kaufmano series on Business Intelligence, Elsevier Science, 2013.

[11]. Mike Barlow, Real-Time Big Data Analytics: Emerging Architecture, ISBN: 978-1-449-36421-2, 2013