



# Big Data Analytics: Life cycle and its Tools

**Akhil Nashaboina**

Student of ACE Engineering College, Jawaharlal Nehru Technological University  
Department of Electrical and Electronics Engineering, Hyderabad, Telangana, India

**Abstract:** The big data analytics is basically the process of examining large data sets. This is done to extract meaningful insights from big data such as hidden patterns, market trends, and customer preferences. This will help organisations to make business decisions. In big data analytics the data can be driven by the life cycle which includes stages. And for examining the big data, various tools were developed to gain insights of what actually is going to happen in the future or done in the past.

**Keywords -** Data sets, Big data, Organisations, Life cycle, Data analytics

## I. INTRODUCTION

Big data are extremely large data sets that may be analysed computationally to reveal hidden patterns, hidden trends and associations, especially relating to human behaviour and interactions. As the name says a collection of data sets that is so large and complex, because it comes in different types it may be relational data, non-relational data, structured data or unstructured data. When all these types of different data get mixed up and when it is available for us, its very difficult to process using the existing database system tools. Because database system tools are usually for relational data, we can also get complex data but the processing is indeed expensive if we are using our traditional databases.

## II. LIFE CYCLE OF BIG DATA ANALYTICS

There are various stages how data is analysed from the beginning.

### Stage-1 Business case evaluation:

In business case evaluation the motive behind the analysis is identified. We need to understand why we are analysing, so that we know how to do it and what are the different parameters that have to be looked into. Once this is done it is clear for us and it becomes much easier for us to proceed with the rest.

### Stage-2 Identification of data:

Here, a broad variety of data sources are identified.

### Stage-3 Data filtering:

Once we get the required data we will have to see if the data we received is fit for analysis or not. All the data that we receive we'll have meaningful information some of it will surely just be corrupt data to remove this corrupt data we will pass this entire data through a filtering stage in this stage all the corrupt data will be removed.

### Stage-4 Data extraction:

Now we have the data - the corrupt data. We have to figure out which data will be compatible with the tool that we will be using for analysis. If we find data which is incompatible, we first extract it and then transform it to a compatible form depending on the tool that we use.

### Stage-5 Data aggregation:

In this stage all the data with the same fields will be integrated. This is known as aggregation stage.

### Stage-6 Data analysis:

Data analysis is the important stage in the lifecycle of big data analytics here in this step the entire process of evaluating our data using various analytical and statistical tools to discover meaningful information is done.

### Stage-7 Visualization of data:

The result of the data analysis stage is then graphically communicated using tools like tableau, power bi, click view.

### Stage-8 Final analysis result:

This analysis result will then be made available to different business stakeholders for various decision making.

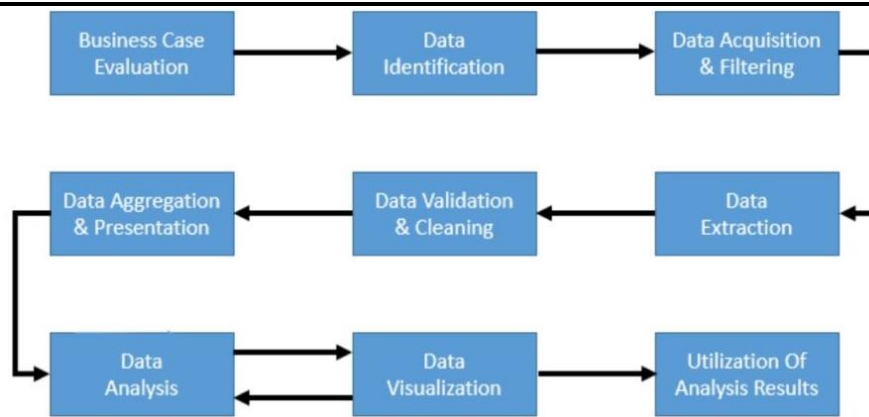


Fig-1 Life Cycle of Big Data Analytics

### III. TOOLS USED IN BIG DATA ANALYTICS

#### 1. Hadoop:

Apache Hadoop is used to store and process big data in a distributed and parallel fashion it allows us to process data very fast. Hadoop uses MapReduce big and high for analysing this big data. Hadoop is easily one of the most famous big data tools.



Fig-2 Hadoop

#### 2. MongoDB:

MongoDB is a cross-platform document-oriented database. It has the ability to deal with large amount of unstructured data. Processing of data which is unstructured and processing of datasets that change very frequently is done using MongoDB.



Fig-3 MongoDB

#### 3. Talend:

Talend provides software and services for data integration, data management and cloud storage. It specializes in big data integration. Talend open studio is a free open-source tool for processing data easily on a big data environment.



Fig-4 talend

#### 4. Cassandra:

Cassandra is used widely for an effective management of large amounts of data. It is similar to Hadoop in its feature of fault tolerance where data is automatically replicated to multiple nodes. Cassandra is preferred for real-time processing.



**cassandra**

Fig-5 Cassandra

#### 5. Spark:

Spark is used for data processing. This data processing engine is developed to process data faster than Hadoop MapReduce. This is done because spark does all the processing in the main memory of the data nodes and thus it prevents unnecessary input/output overheads with the disk. Whereas MapReduce is a disk based and hence spark proves to be faster than Hadoop MapReduce.



Fig-6 Spark

**6. Storm:**

Storm is a free big data computational system which is done in real time it is one of the easiest tools for big data analytics. It can be used with any programming language. This feature makes storm very simple to use.



Fig-7 Storm

**7. Kafka:**

Kafka is a distributed streaming platform which was developed by LinkedIn and later given to Apache software foundation. It is used to provide real-time analytics result. And it is also used for fault tolerant storage.



Fig-8 Kafka

**IV. BIG DATA APPLICATION DOMAINS****1. Ecommerce:**

Nearly 45% of the world is online and they create a lot of data every second. Big data can be used smartly in the field of e-commerce by predicting customer trend, forecasting demands, adjusting the price etc. Online retailers have the opportunity to create better shopping experience and generate higher sales.

**2. Marketing:**

Having big data doesn't automatically lead to better marketing strategy meaningful insights need to be derived from it in order to make right decisions by analysing big data we can have personalized marketing campaigns which can result in better and higher sales.

**3. Education:**

In the field of education depending on the market requirements new courses are developed. The market requirement needs to be analysed correctly with the respect to the scope of the course and accordingly a scope needs to be developed there is no point in developing a course which has no scope in the future hence to analyse the market requirement and to develop new courses we use big data analytics.

**4. Healthcare:**

There are a number of uses of big data analytics in the field of healthcare and one of it is to predict a patient's health issue that is with the help of their previous medical history, big data analytics can determine how likely they are to have a particular health issue in the future.

**5. Media and Entertainment:**

In the field of media and entertainment big data analytics is used to understand the demands or shows songs, movies etc to deliver personalised recommendation list as we saw with Netflix.

**6. Banking:**

Big data analytics is used in the field of banking for risk management. It is also used to analyse a customer's income and spend patterns and to help the bank predict if a particular customer is going to choose any of the bank offers such as loans, credit card schemes etc in this way the bank is able to identify the right customer who is interested in its offers.

**7. Telecommunications:**

Telecom companies have begun to embrace big data to gain profit. Big data analytics helps in analysing network traffic and called data records. It can also improve its service quality and improve its customer experience.

**8. Government:**

In government the big data analytics can be applied to analyse all the available data to understand crime patterns. Intelligence services can use predictive analytics to focus the crime which could be committed. In Durham the police department was able to reduce the crime rate using big data analytics with the help of data police could identify whom to target where to go when to patrol and how to investigate crimes. Big data analytics helped them to discover patterns of crime emerging in the area.



Fig-9 Big Data Applications

## V. CONCLUSION

Big data analytics refers to the complex process of analyzing big data for revealing information such as correlations, hidden patterns, market trends, and customer preferences. Big data analytics has the potential to transform the way healthcare provides advanced technologies to gain insights.

## REFERENCES

- [1] Francesco Corea, "An Introduction to Data", Springer Nature Switzerland AG publishers, 2019
- [2] S. Christian Al bright and Wayne L. Winton, "Business Analytics", Cengage Learning publishers, 6<sup>th</sup> edition, 2017
- [3] Wiley, "Big Data in Practice", Published by TJ International, 1<sup>st</sup> edition, 2016
- [4] Arshdeep Bahga and Vijay Madiseti, "Big Data Science & Analytics", Published by Arshdeep Bahga & Vijay Madiseti, 2016
- [5] Wiley, "Data Science & Big Data Analytics", Published by John Wiley & Sons, Inc, 2015

