BIG DATA ANALYTICS

Mansi Gupta[1], Mrs. Indu Khatri[2]

[2]HOD, [1,2]Department of Computer Science Engineering, BhagwanMahavir College Of Engineering And Management, Sonipat, India

ABSTRACT: This study is about the attempts which are made to evolve a platform traditional technologies based notion of big data. If it can be analyzed to reach at a conclusionthen it is of worth. As big data comprises of structures, semi-structured and unstructured data, we canfind all the methods which can analyze the text, audio and videos as well in the study. Before thinking about the big data analytics the challenges imposed by the big data needs to be kept in mind.

I. INTRODUCTION

Means for analyzing data sets and drawing conclusions about them are provided by data analytics technologies and techniqueswhich is helpful for the organizations to make well informed business decisions. Business intelligence queries answer basic questions about business operations and performance.

It is a form of advanced analytics where complex applications with elements like predictive models, statistical algorithms etc.areinvolved.

## II. THE IMPORTANCE OF BIG DATA ANALYTICS

Big data analytics offers various business benefits, including:

- New revenue opportunities
- More effective marketing
- Better customer service
- Improved operational efficiency
- Competitive advantages over rivals

Big data analysts, data scientists, predictive modelers, statisticians and other analytics professionals for analyzing growing volumes of structured transaction data etc.are enabled by these applications.

III. DATA ANALYTICS

Data analytics is the source of big data analytics, but this study is about big data analytics so we will study only the output result evaluation of data mining are also studied.
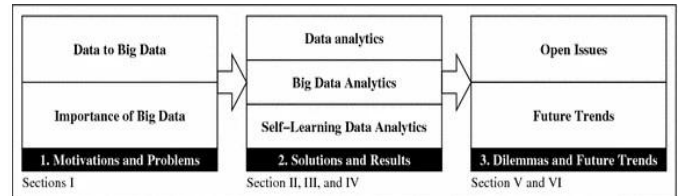
For solving the data mining problems,two major goals are:
1.Cohesion which says that the distance between each data and the centroid of its cluster should be as small as possible
2.Coupling which says that the distance between data belonging to different clusters should be as large as possible.

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i} D(x_{ij} - c_i), SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i} D(x_{ij} - c_i),$$
(1)

$$c_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}, c_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij},$$
(2)

$$D(p_i, p_j) = (\sum_{l=1}^{d}|p_{il}, p_{jl}|^2)^{1/2}, D(p_i, p_j) = (\sum_{l=1}^{d}|p_{il}, p_{jl}|^2)^{1/2},$$
(3)

$$ACC = \text{Number of cases correctly classified.}$$
(4)

With the confusion matrix at hand:

$$p = \frac{TP}{TP+FP}, p = \frac{TP}{TP+FP},$$
(5)

$$r = \frac{TP}{TP+FN}. r = \frac{TP}{TP+FN}.$$
(6)

$$F = \frac{2pr}{p+r}. F = \frac{2pr}{p+r}.$$
(7)

The computation cost and response time are another two well-known measurements for this evaluation.

IV. BIG DATA ANALYTICS TECHNOLOGIES AND TOOLS

In traditional data warehouses that are based on relational databases oriented to structured data sets the unstructured and semi-structured data types typically don't fit well.

Hence many of the organizations that collect, process and analyze big data turned to NoSQL databases, as well as Hadoop and its companion data analytics tools, including:

- YARN
- MapReduce
- Spark
- HBase
- Hive

- Kafka
- Pig
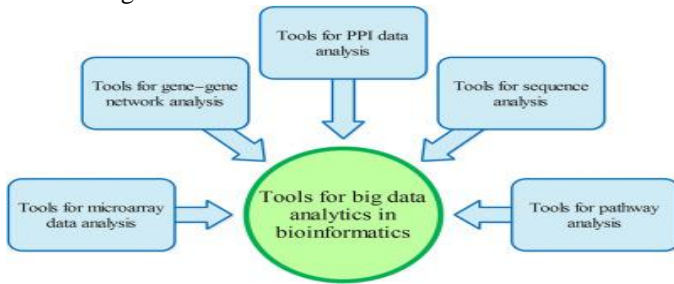
WORKING

Hadoop clusters and NoSQL systems are used primarily in some cases as landing pads and staging areas for data before loading it into a data warehouse or analytical database for analysis.

Concept of a Hadoop data lake which is served as the primary repository for incoming streams of raw data has been adopted by big data analytics.

Once the data is ready, for advanced analytics processes included tools for:

- data mining
- predictive analytics
- machine learning
- deep learning

V. BIG DATA ANALYTICS USES AND CHALLENGES

Data from both internal systems and external sources are often included by its applications. In addition,as the users look to perform real-time analytics on data fed into Hadoop systems through stream processing enginesthe streaming analytics applications are now becoming common in big data environments.

It has become increasingly beneficial in supply chain analytics. Specifically, big supply chain analytics expands datasets for increased analysis that goes beyond the traditional internal data found on enterprise resource planning and supply chain management systems.

VI. EMERGENCE AND GROWTH OF BIG DATA ANALYTICS

This term was first used to refer to increasing data volumes. The notion of big data was expanded to also include increases in the variety of data being generated by organizations and the velocity at which that data was being created and updated. The three factors volume, velocity and variety they became the 3Vs of big data.
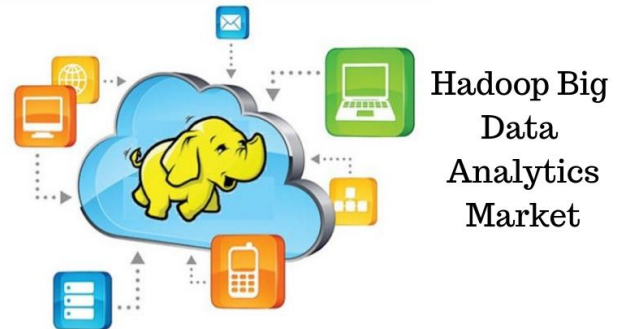
Initially, as the shape was taken by Hadoop ecosystem and it had started to mature, big data applications which was primarily the province of large internet and e-commerce companies such as Yahoo, Google and Facebook, as well as analytics and marketing services providers. Though, big data analytics has increasingly been embraced by retailers, financial services firms, etc.

VII. CONCLUSION

In this paper, the analysis framework refers to the whole system, from raw data gathering, data reformat, data analysis, all the way to knowledge representation.The learner typically represented the classification function which will create the classifier to help us classify the unknown input data and hence use it for various application. Hence the topic is vast, this is just a minor study about big data analytics.

REFERENCES

[1] Ding C, He X. K-means clustering via principal component analysis.
[2] Kollios G, Gunopulos D, Koudas N, Berchtold S. Efficient biased sampling for approximate clustering and outlier detection in large data sets.
[3] Fisher D, DeLine R, Czerwinski M, Drucker S. Interactions with big data analytics.
[4] Laney D. 3D data management: controlling data volume, velocity, and variety, META Group, Tech. Rep. 2001.
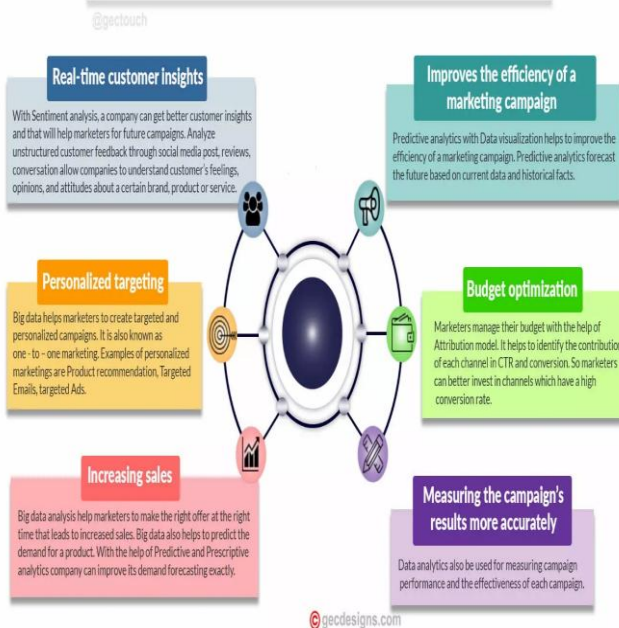
[5] vanRijmenam M. Why the 3v's are not sufficient to describe big data, BigDataStartups, Tech. Rep. 2013.

[6] Borne K. Top 10 big data challenges a serious look at 10 big data v's, Tech. Rep. 2014.

[7] Press G. $16.1 billion big data market: 2014 predictions from IDC and IIA, Forbes, Tech. Rep. 2013

[8] Big data and analytics—an IDC four pillar research area, IDC, Tech. Rep. 2013.

[9] Taft DK. Big data market to reach $46.34 billion by 2018, EWEEK, Tech. Rep. 2013.