Big Data Analytics Evaluation

[1] Mudassir Khan

Department of Computer Science College of Science & Arts Tanumah King Khalid University, Abha Saudi Arabia

Abstract: - In this age of technological and advanced world big data is prominent as a world new currency. The term big data is not a framework, language and Technology. Actually Big data is nothing but a problem statement. In the current era number of IOT enable devices is using data in huge amount. The data is coming from different datasets at an enormous amount. As data is increasing exponentially every year, the traditional system to store and process the data become incapable to handle it. The existing technologies are not capable to handle the big data. In this digital world, the data is generated automatically by the online interactions of big data applications. The Big data is used in the evaluation of emerging form of information. In the last two years data is growing at an enormous speed exponentially as compare to last twenty years. In this current era human life is totally dependable on IOT. This paper presents the overall changes in big data analytics evaluation growth in the recent years. The innovations in the technology and greater affordability of digital devices with internet made a new world of data are called big data. The data captured by enterprises such ad rise of IOT and multimedia has produce an overwhelming of data in either structured or un-structured format. It is a fact that data that is too big to process is also too big to transfer anywhere. So it is just an analytical program which needs to be moved (not the data) and this is only possible with cloud computing.

Index Terms - Big Data, Analytics, Issues & Challenges, Evaluation of data.

I. INTRODUCTION

The Big data is a data but in huge amount and complex, so that traditional or existing methods are not capable to process it. The volume and variety of data is increasing exponentially in the past two years. The data is getting popularity as data is coming from different sources like, social media, education, healthcare, multimedia, banking, agriculture and ecommerce more.

1-Big Data

In the recent years fast enhancement in technologies, storage and collective data capacity, the big data is increasing exponentially in all the fields. As we know that in the last two years data growth rate is very high about 90% of total data. The huge amount of datasets is also known as big data as big data come from different domains. The data is growing at an unprecedented rate. The data is getting significant growth from unknown domains or data sources. In the near future, expect that such networks will connect other entities such as software components, web based services, data sources and work flows. The peoples those are connected to the social media and other tools are producing a continuous data in huge amount that is deposited into a storage location of connected data. The business entities or individuals might conduct big data analytics on these connected data technologies like cloud. The connected data is the confluence, when social media and cloud are presented solution for big data analytics.

1.1 The Big data description with 7V's

The big data is called a huge data of different types because it comes from different data sources or domains. This data cannot be processed and store by the traditional or existing systems. Here data is described by its characteristics of big data 7V'S are as follows.

1.1.1 Volume.

The amount of data is adding from different data sources to the data warehouse (e.g in Bytes, KB, MB, GB, TB)

1.1.2 Variety

It divides the data in three categories:

- Structured data
- Semi Structured data
- Un-Structured data

1.1.3 Velocity

It describes the speed of data rate from different sources, that is, the speed of data production such as different social media sites.

1.1.4 Veracity

It represents the quality and accuracy of the data. It produces the unique data or extracting the valid data from the data sources

1.1.5 Value

The value of big data, i.e. it shows the importance of data after analysis.

1.1.6 Variability

Variability indicates to data or information whose meaning is constantly changing. This is particularly the case when gathering data relies on language processing.

1.1.7 Visualization

The way of presenting the data in a pattern that's readable and accessible- it is where visualization occurs. Visualizations can contain variety of variables and parameters.

1.2 Big data Issues
1.2.1 Storage wise

Getting ton's of data from different data sources but is very difficult to store it in the traditional system.

1.2.2 Processing Wise

Somehow we are able to manage the data but how to store and handle the data by using different hardware techniques or by server. When time comes to process the data, we are not capable to process and store it.

1.3 Problem with big data

1. The strong exponentially growing huge data sets.

2. Data sets generated in past two years are more than in previous history in total

3. By 2020 total digital data will grow to approximate 49 Zeta bytes.

1.4 Big data Challenges

The main challenge of big data is to explore the huge amount of data and extract the valid data for the future work or use.

Big data require the large computational infrastructure to ensure successful data processing and analysis.

II. BIG DATA ANALYTICS.

The Big data is considered as technical problem. Nowadays it has been become the business and future opportunities. Few years ago, in the early 2000, big data was considered as a serious problem by the organizations. As the data volume was growing at very fats rate, storage and CPU technologies swamp by the countless terabytes of data, at this point the Information Technology field faced a crisis on data capacity. As the time past the storage and CPU becomes powerful, intelligence and process the data at very fast speed as data is increasing. Today the numbers of organizations are exploring the big data to identify the valid information that was not possible before. Now all the business enterprises are producing data in huge amount that is based on their mass consumers. By using the analytics techniques, the organizations can identify current situation of running business is producing valid big data and can observe the consumer requirement on processed data. The Big data analytics is the approach of popular analytics technologies to verify big data sets.

The term "Big data" and "Big data analytics" are two technical entities. First, the big data is the huge amount of data with detailed information. Second, the analytics is a combination of different type of analytical tools such as statistical analysis, data mining, NPL and so on. If put them all together will get the big data analytics, which is currently used in BI. Based on this, the numbers of organizations are maintaining big data in their organization data warehouse (ODW). The DW should be auditable, well maintained and clean that can be used to generate business reports demand.

2.1 Benefits of big data analytics

The big data evaluation has been around for years, as most of the big companies understand that if they kept all the data that comes into their businesses. As a part of current business can be applied analytics on it and get accurate value from it. In the 1950's the term "big data" businesses were using normal analytics technique to open the insights trends. The most common benefit of big data analytics brings the evaluation of speed and efficiency. Few years ago a business would have collect the information, execute analytics and unearthed information that could be used for future decisions. Nowadays business can have the insights for advance decisions. The ability to work becomes faster and stay advanced- can give companies a competitive edge they did not have before.

2.2 Growth of analytical big data

The Modern world becomes exponentially digital, and data volumes increase, opportunities for utilizing data will grow: the amount of data that would be useful for analysis will rise from 32% to more than 45%. The overall worldwide market in big-data technology and services is supposed to increase at a compound annual growth rate of about 36% between 2014 and 2019, and worldwide revenue for big data and business analytics has been forecast to increase more than 60%. The main sectors for big data include manufacturing, banking and insurance, government, professional services, telecommunications, health, transport and retail.

2.3 Tools and techniques of big data

The Big data Analytics are structured, unstructured, formalized approaches to evaluating information. It

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

extended activities like arithmetic calculations, developing new information and documentation results. Formally it is a set of standardized tools. The tools serve as an investigation guide. It is useful in a situation where you became unfamiliar with the information.

As a result, the number of organizations that collect, process and analyze big data enhance to databases like Hadooop tools including:

• YARN: The cluster management technology and one of most common key feature in Hadoop (second generation).

• MapReduce: A programming environment that allows developers to create/write programs that can process massive amount of unstructured data in parallel across a distributed computer.

• Spark: An open source parallel processing environment that permit number of users to execute huge amount of data analytics applications across cluster systems.

• HBase: A column Oriented essential/utility data or information store built on the roof of the HDFS

• Hive: an open-source data warehouse cluster system for enquiring and analyzing large datasets stored in Hadoop files.

• Pig: an open-source technology that is capable to offers a high-level sturcture for the parallel programming of MapReduce jobs to be executed on Hadoop clusters.

III. BIG DATA ISSUES WITH EXISTING SYSTEM

As we know different high level programming languages like java, mainframe system and data warehouse tools (DWH, BI tools- RDBMS, MYSQL) are straining to store and process the data.

o Lack of proper goaled objectives,

o Lack of required skills

o The exponentially increased size of big data

o The undefined structure of the big data

o The difficulty of imposing data consistency

o Privacy

o Information/data management/integration,

o The Standard rights management,

o The data discovery tools

o Data veracity (How to deal with uncertainty, imprecision, missing details?),

o Data validation and verification,

o Technical challenges for big data analytics (While data is in running or processing state)

IV. EVALUATION OF BIG DATA

The popularity of big data comes in last two years. Actually big data is increasing exponentially in all the fields.
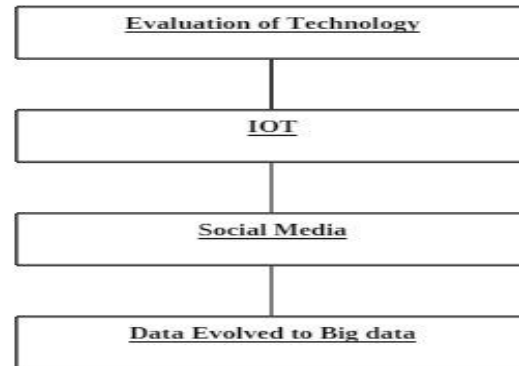

Figure-1 Evaluation of Big data

4.1 Big data monitoring and evaluation

The big data world is increasing rapidly and it becoming a complex hub of data. As a outcome of technology, the production rate of data is monitored in higher volumes from different unknown sources. Few digital data is harmful for public goods and for global enhancement. As new data analysis techniques are emerging, the monitoring and evaluation of data devices are not flexible, set of connected devices, or characterize that is enough to react for this change of rate. Data insurrection is required in the field of monitoring and evaluation, as different data analysis areas exist in this technological era.

4.2 Enhancement of Technology

The amount data generated by the Smartphone applications will not give any idea on every action. We can take the example of self driving car that records every minor obstacle, distance from obstacle and decide how to react with that. In this process huge amount of data generated to process. By 2020 the IOT devices will be increased by more than 50 billion.

V. CONCLUSION

Big Data helps the big organizations to enhance the exponentially increasing growth opportunities and entirely new categories of that organization that can combine and analyze industry data. These companies have variety of information about the items and services, buyers and suppliers, consumer's preferences that can be captured and analyzed.The Digital world is growing exponentially and become more complex in the capacity (terabyte to petabyte), variety (structured and un-structured), velocity (exponential growth) in nature. This describes as Big Data becomes a

global phenomenon. This is particularly considered to be a data collection that has collectively so large it cannot be managed directly or utilized using traditional data management tools: e.g., RDBMS (Relational database management system). To handle this problem, traditional RDBMS are specifically designed a rich set of alternative DBMS; such as Big data analytics tools and Search-based systems. This paper motivation is to provide - classification, characteristics and evaluation of Big Data Analytics. This report is deliberated to help users, mainly to the organizations to take an independent understanding of the pros and cons of various Analytics tools and techniques database approaches to supporting applications that process large amount of data.

REFRENCES

1. Codman EA. The classic: the registry of bone sarcomas as an example of the end-result idea in hospital organization. 1924. Clin Orthop Relat Res 2009; 467(11):2766–70.

2. Goldstein J. Private practice outcomes: validated outcomes data collection in private practice. Clin Orthop Relat Res 2010; 468(10):2640–5.

3-McGuire,T., Manyika, J. and Chui, M. (2012)"Why Big Data is the New Competitive advantage". Ivey Business Journal, Jul/Aug, Vol.76 Issue 4, pp. 1-4.

4- Y. Liu, and Y. Zhuang, "Research Model of Churn Prediction Based on Customer Segmentation and Misclassification Cost in the Context of Big Data", Journal of Computer& Communications, Vol. 3,2015,pp87-93

5- A. Tole, "Big data challenges" DatabaseSystem Journal, Vol.4 (3), 2013, PP31-40.

6- X. Wu and X.Zhu, "Data Mining with Big Data", IEEE transactions on Knowledge & Data Engineering, Vol 26(1), 2014, pp97-107

7- M.Musolesi "Big Mobile Data Mining: Good or Evil?" IEEE Internet Computing, 2014,pp2-5.

8- Mudassir Khan, "Design and Analysis of Security Aware Scheduling in Grid Computing Environment", International Journal of Computer Science and Information Technology Research (IJCSITR),Vol. 1, Issue 1, pp: (42-50), Month: October-December 2013

9- Afrati, F.N. & Ullman, J.D. (2011) Optimizing Multiway Joins in a Map-Reduce Environment. IEEE

Transactions on Knowledge and Data Engineering, 23(9), 1282-1298.

10- Jiang, D., Tung, A. & Chen, G. (2011) MAP-JOIN-REDUCE: Toward Scalable and Efficient Data Analysis on Large Clusters. IEEE Transactions on Knowledge and Data Engineering, 23(9), 1299-1311.

11- Kraska, T. (2013) Finding the Needle in the Big Data Systems Haystack. IEEE Internet Computing, 17(1), 84-86.

12- Lee, D., Kim J-S. & Maeng, S. (2013) A Large -scale incremental processing with MapReduce. Future Generation Computer System, 36, pp 66-79.

13- Bakshi, K. (2012) Considerations for Big Data: Architecture and Approach. IEEE Aerospace Conference, (pp.1-7). Big Sky, USA.

14- Mayer-Sch¨onberger V, Cukier K (2013) Big data: a revolution that will transform how we live, work, and think. Eamon Dolan/Houghton Mifflin Harcourt

15- S. Kaisler, F. Armour, J. A. Espinosa, and W. Money, "Big Data : Issues and Challenges Moving Forward," 46th Hawaii Int. Conf. Syst. Sci. IEEE, pp. 995–1004, 2013.

16- A. Katal, "Big Data : Issues , Challenges , Tools and Good Practices,"IEEE, pp. 404–409, 2013.

17- A. C. M. Webinar and W. H. Guide, "Simplifying Big Data … with Hadoop."

18- Internet Security Threat Report | Appendices," vol. 21, no. April,2016.

19- McAfee, Andrew and Brynjolfsson, Erik (2012), " Big Data: The Management Revolution" , Harvard Business Review, October Issue

20- M.Musolesi "Big Mobile Data Mining: Good or Evil?" IEEE Internet Computing, 2014,pp2-5.