Big Data: A Review

Mrs. Jadhav Jayshree M. & Mrs. Kulkarni Chandraprabha V

ABSTRACT

The growing need in today's era poses a big challenge over the management of data which is produced from diverse sectors. Not alone this, but also to extract the useful and considerable amount of information from the data which is available in our surroundings. This paper focuses on the discussions of big data, the everyday challenges that are faced and solutions to overcome the same. The first part provides an introduction on the big data growing tremendously in industry on a day-to-day basis, second part includes the concepts involved in big data, third part relates to the daily challenges and the last section sees the efficient ways which are used for the management of big data.

London Journals Press

Big Data: A Review

Mrs. Jadhav Jayshree M.[α] & Mrs. Kulkarni Chandraprabha V.[σ]

ABSTRACT

The growing need in today's era poses a big challenge over the management of data which is produced from diverse sectors. Not alone this, but also to extract the useful and considerable amount of information from the data which is available in our surroundings. This paper focuses on the discussions of big data, the everyday challenges that are faced and solutions to overcome the same. The first part provides an introduction on the big data growing tremendously in industry on a day-to-day basis, second part includes the concepts involved in big data, third part relates to the daily challenges and the last section sees the efficient ways which are used for the management of big data.

Keywords: big data, challenges, big data analysis tools.

Author: Department Of Comp. Sci. & I.T. Rajarshi Shahu Mahavidyalaya (AUTONOMOUS), Latur Maharashtra.

I.   INTRODUCTION

Big data typically refers to the following types of data:

Traditional enterprise data: includes customer information from CRM systems, transactional ERP data, web store transactions, and general ledger data.

Machine-generated /sensor data – includes Call Detail Records ("CDR"), weblogs, smart meters, manufacturing sensors, equipment logs (often referred to as digital exhaust), trading systems data.

Social data: includes customer feedback streams, micro-blogging sites like Twitter, social media platforms like Facebook. In fact, there are four key characteristics that define big data:

1.1   Key Characteristics To Define Big Data

1. Volume: Machine-generated data is produced in much larger quantities than non-traditional data. For instance, a single jet engine can generate 10TB of data in 30 minutes. With more than 25,000 airline flights per day, the daily volume of just this single data source runs into the Peta-bytes. Smart meters and heavy industrial equipment like oil refineries and drilling rigs generate similar data volumes, compounding the problem.

2. Velocity: Social media data streams – while not as massive as machine-generated data – produce a large influx of opinions and relationships valuable to customer relationship management. Even at 140 characters per tweet, the high velocity (or frequency) of Twitter data ensures large volumes (over 8 TB per day).

3. Variety: Traditional data formats tend to be relatively well defined by a data schema and change slowly. In contrast, non-traditional data formats exhibit a dizzying rate of change. As new services are added, new sensors deployed, or new marketing executed, new data types are needed to capture the resultant information.

4. Value: The economic value of different data varies significantly. Typically there is good information hidden amongst a larger body of non-traditional data; the challenge is identifying what is valuable and then transforming and extracting that data for analysis. To make the most of big data, enterprises must evolve their IT infrastructures to handle these new high-volume, high-velocity, high-variety sources of data and

integrate them with the pre-existing enterprise data to be analyzed.

Table 1: Comparision Between Big Data and Small Data

Category	Big Data	Small Data
Data sources	Data generated outside the enterprise from non traditional data sources, Include: 1.Social media 2. Sensor data 3. Long data 4. Device data 5. Video ,Images ,etc.	Traditional enterprise data. Includes: 1.Enterprise Resource 2.Planning transactional data 3. Customer Relationship 4. Management(CRM) Systems 5. Web transactions 6. Financial data e.g. general ledger data
Volume	Terabytes Peta bytes Hexa bytes Zetta bytes	Gigabytes Terabytes
Velocity	Often real-time Requires immediate response	Batch or near real-time, Does not always require immediate response
Variety	Structured Un-structured Multi-structured	Structured Un-structured
Value	Complex, advanced, predictive business analysis and insights	Business intelligence , analysis and reporting

II. BIG DATA CHALLENGES

Data becomes big data when its volume, velocity, or variety exceeds the abilities of systems to ingest, store, analyze, and process it. Many organizations have the equipment and expertise to handle large quantities of structured data—but with the increasing volume and faster flows of data, they lack the ability to "mine" it and derive actionable intelligence in a timely way. Not only is the volume of this data growing too fast for traditional analytics, but the speed with which it arrives and the variety of data types necessitates new types of data processing and analytic solutions. However, big data doesn't always fit into neat tables of columns and rows. There are many new data types, both structured and unstructured, that can be processed to yield insight into a business or condition.

For example, data from twitter feeds, call detail reports, network data, video cameras, and equipment sensors often isn't stored in a data warehouse until you have pre-processed it to distill and summarize and perhaps to detect basic trends and associations. It is more cost effective to load the results into a warehouse for additional analysis. The idea is to "reduce" the data to the point that it can be put in a structured form. Then it can be meaningfully compared to the rest of your data, and scrutinized with traditional business intelligence tools.
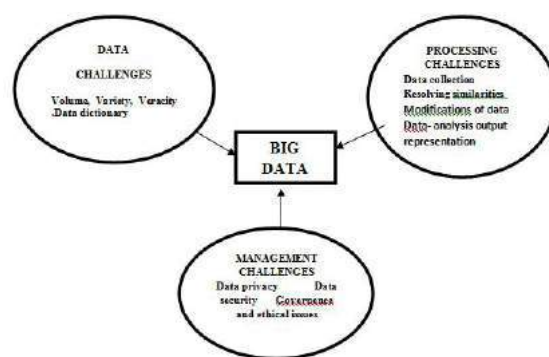

Figure 1: Big Data Challenges

There are several obstacles in the Big Data & Analytics process that need to be overcome in order to achieve success.

1. Data Quality: Data is coming from many disparate sources from all facts of the organization. In order to overcome this, a data warehouse is essential. However, when a data warehouse tries to combine inconsistent data from disparate sources, it encounters errors. Inconsistent data, duplicates, logic conflicts, and missing data all result in data quality challenges. Poor data quality results in

London Journal of Engineering Research

faulty reporting and analytics necessary for optimal decision making.

2. *Understanding Analytics:* The powerful analytics tools and reports available through integrated data will provide credit union leaders with the ability to make precise decisions that impact the future success of their organizations. When implementing a Big Data & Analytics solution, analytics and reporting will have to be taken into design considerations. In order to do this, the business user will need to know exactly what analysis will be performed. Envisioning these reports will be difficult for someone that hasn't yet utilized a Big Data & Analytics solution and is unaware of its capabilities and limitations.

3. *Quality Assurance:* The end user of a Big Data & Analytics solution is using reporting and analytics to make the best decisions possible. Consequently, the data must be 100 percent accurate or a credit union leader will make ill-advised decisions that are detrimental to the future success of their business. This high reliance on data quality makes testing a high priority issue that will require a lot of resources to ensure the information provided is accurate. The credit union will have to develop all of the steps required to complete a successful Software Testing Life Cycle (STLC), which will be a costly and time intensive process.

4. *Performance:* Implementing a Big Data & Analytics solution is similar to building a car. A car must be carefully designed from the beginning to meet the purposes for which it is intended. Yet, there are options each buyer must consider to make the vehicle truly meet individual performance needs. A Big Data & Analytics solution must also be carefully designed to meet overall performance requirements. While the final product can be customized to fit the performance needs of the organization, the initial overall design must be carefully thought out to provide a stable foundation from which to start. Major customizations are extremely expensive.

II. BIG DATA ANALYSIS TOOL

There are five key approaches to analyzing big data and generating insight:

1. *Discovery tools:* are useful throughout the information lifecycle for rapid, intuitive exploration and analysis of information from any combination of structured and unstructured sources. These tools permit analysis along side traditional BI source systems. Because there is no need for up-front modeling, users can draw new insights, come to meaningful conclusions, and make informed decisions quickly.

2. *BI tools:* are important for reporting, analysis and performance management, primarily with transactional data-from data warehouses and production information systems. BI Tools provide comprehensive capabilities for business intelligence and performance management, including enterprise reporting, dashboards, ad-hoc analysis, scorecards, and what-if scenario analysis on an integrated, enterprise scale platform.

3. *Database Analytics:* include a variety of techniques for finding patterns and relationships in your data. Because these techniques are applied directly within the database, you eliminate data movement to and from other analytical servers, which accelerates information cycle times and reduces total cost of ownership.

4. *Hadoop:* is useful for pre-processing data to identity macro trends or find nuggets of information, such as out of-range values. It enables businesses to unlock potential value from new data using inexpensive commodity servers. Organizations primarily use Hadoop as a precursor to advanced forms of analytics.

5. *Decision Management:* Includes predictive modeling, business rules, and self-learning to take informed action based on the current context. This type of analysis enables individual recommendations across multiple channels, maximizing the value of every customer interaction. Oracle Advanced Analytics scores can be integrated to operationalize complex predictive analytic models and create real-time decision processes.

III. CONCLUSION

The better analysis of the large volumes of data that are becoming available, there is the potential for making faster advances in many scientific

disciplines and improving the profitability and success of many enterprises. For that reason many technical challenges described in this paper must be addressed before this potential can be realized fully. The challenges include not just the obvious issues of scale, but also heterogeneity, lack of structure, error-handling, privacy, timeliness, provenance, and visualization, at all stages of the analysis pipeline from data acquisition to result interpretation. These technical challenges are common across a large variety of application domains, and therefore not cost-effective to address in the context of one domain alone. These challenges will require transformative solutions. This paper support and encourages fundamental research towards addressing these technical challenges if we are to achieve the promised benefits of Big Data.

REFERENCES

1. A. Gandomi, and M. Haider, "Beyond the Hype: Big DataConcepts, Methods, and Analytics", International Journal of Information Management (35), 2015, pp. 137-144.

2. Big Data Analytics: A Literature Review Paper (PDF Download Available). Available from: https://www.researchgate.net/publication/264 555968_Big_Data_Analytics_A_Literature_R eview_Paper [accessed Mar 29, 2017].

3. The Search for Analysts to Make Sense of Big Data. Yuki Noguchi. National Public Radio, Nov. 30, 2011. http://www.npr.org/2011/ 11/30/142 893065/the-search-for-analysts-to-make-sense -of-big-data

4. Towards the Internet of Services, Community Research and Development Information Service, European Commission, Apr. 2014; http://cordis .europa.eu/fp7/ict/ssai/.

5. "The Emerging Big Returns on Big Data", A TCS 2013 Global Trend Study..

6. www.googlebooks.com/Big Data

7. B. Chae, C. Yang, D. Olson, and C. Sheu, "The impact of advanced analytics and data accuracy on operational performance: A contingent resource based theory (RBT) perspective", Decision Support Systems (59), 2014, pp. 119 126

8. Challenges and Opportunities with Big Data", A Community White Paper Developed by Leading Researchers Across United States, 2012.