International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 3 | ISSN : 2456-3307

Big Data Analytics: Modern Techniques and Technologies

P. Bastin Thiyagaraj, M.Joyni, C. Abirami
Department of Information Technology, St. Joseph's College (Autonomous), Trichy, TamilNadu, India

ABSTRACT

Big data is something that is used to analyze insights and lead to better decision. Whereas Big data analytics involves automation insights into a certain data sets and supposes the usage of queries and data aggregation procedure. This paper presents about big data analytics, modern techniques & technology. Here big data analytics explains methodology, definition of big data, process of big data and benefits of big data. This paper only focuses on process of big data analytics and also focusing about modern techniques & technology.

Keywords: Big Data, Analytics, Volume, Variety

I. INTRODUCTION

In recent years "big data" has become something of a buzzword in business, computer science, information studies, information systems, statistics, and many other fields. As technology continues to advance, we constantly generate an ever-increasing amount of data. This growth does not differentiate between individuals and businesses, private or public sectors, institutions of learning and commercial entities. It is nigh universal and therefore warrants further study. This review aims to provide a brief overview of big data, and how it is used in analytics. After reviewing the methodology of research used in creating the review the investigation of big data will begin by attempting to craft a satisfying definition of the term. Many of the relevant technologies and techniques used in big data analytics will be covered briefly and the benefits of big data analytics across various sectors will be explored. The review will also present several of the challenges and barriers faced by purveyors of big data analytic tools and attempt to determine if the results of the analytics offset the costs of overcoming these challenges sufficiently to make them a wise investment.

is easy to see how these characteristics line up with Russom's Vs.

II. PROCESS OF BIG DATA ANALYTICS

In the process of transforming data to yield actionable insights it includes many steps from Collecting to Decision Making. In this process data is collected which is in raw form. This is then classified and organized. This data is associated with metadata and then converted into meaningful information. For the analysis of this data it is aggregated and summarized. This data is stored which helps in making the decisions.

Figure 2. Process of big data

III. TECHNIQUES AND TECHNOLOGIES

Since big data is not only large, but also varied and fast growing many technologies and analytical

techniques are needed in order to attempt extracting relevant information. Many of these are topics large enough to support an entire review on them alone. As such, this report is not designed to provide an in-depth knowledge of all these tools. Rather, it gives a broad overview of some of the most commonly used techniques and technologies to help the reader better understand what tools big data analytics is based on.

TECHNIQUES:

There are a myriad of analytic techniques that could be employed when attacking a big data project. Which ones are used depends on the type of data being analyzed, the technology available to you, and the research questions you are trying to solve. Some of the tools that came up frequently in the reviewed material are summarized here

- Association rule learning: A way of finding relationships among variables. It is often used in data mining and according to Chen, Chiang, and Storey it lends support to recommender systems like those employed by Netflix and Amazon.
- Data mining:[3]Manyika et al. (2011) calls data mining "combining methods from statistics and machine learning with database management" in order to pinpoint patterns in large datasets. [5] Picciano (2012) lists it as one of the most important terms related to data-driven decision making and describes it as "searching or 'digging into' a data file for information to understand better a particular phenomenon."
- Cluster analysis: Cluster analysis is a type of data mining that divides a large group into smaller groups of similar objects "whose characteristics of similarity are not known in advance." and attempts to discover what the similarities are among the smaller groups, and if they are new groups, what caused these qualities.
- Crowdsourcing: Crowdsourcing collects data from a large group of people through an open call, usually via a Web2.0 tool. This tool is used more for collecting data than for analyzing it.
- Machine learning:

Traditionally computers only know what we tell them, but in machine learning, a subspecialty of computer science, we try to craft "algorithms that allow computers to evolve based on empirical data. A major focus of machine learning research is to automatically learn to recognize complex patterns and make intelligent decisions based on data.[4]Miller (2011/2012) gives the example of the U.S. Department of Homeland Security, which uses machine learning to identify patterns in cell phone and email traffic, as well as credit card purchases and other sources surrounding security threats. They use these patterns to try to identify future threats so they can handle them before they become large problems.

- Text analytics: A large portion of generated data is in text form. Emails, internet searches, web page content, corporate documents, etc. are all largely text based and can be good sources of information. Text analysis can be used to extract information from large amounts of textual data. This can be done to model topics, mine opinions, answer questions, and other goals. These are just a few of the many techniques used in big data analytics. For the interested reader, some additional analytical tools not discussed here include classification, data fusion, network analysis, optimization, predictive modeling, regression, special analysis, time series analysis, and others.

Technology:

As with the analytical techniques, there are several software products and available technologies to facilitate big data analytics. Some of the most common will be discussed here.

- EDWs: Enterprise data warehouses are databases used in data analysis. [6]Russom writes that for many businesses that are trying to start handling big data the big question is "Can the current or planned enterprise data warehouse (EDW) handle big data and advanced analytics without degrading performance of other workloads for reporting and online analytic processing?" Some institutions manage their analytic data in the EDW itself while

others use a separate platform, which helps relieve some of the stress on the server resulting from managing your data on the EDW.

- Visualization products: One of the difficulties with big data analytics is finding ways to visually represent results. Many new visualization products aim to fill this need, devising methods for representing data points numbering up into the millions. [6] Russom (2011) lists this field as one of those having the most potential, saying it is "poised for aggressive adoption." Beyond simple representation visualization can also help in the information search. [2] Hey, Tansley, and Tolle's collection The Fourth Paradigm (2009) discussing visualization in data-intensive science in which they explain that visualization products allow us to compare models and datasets and "enables quantitative and qualitative decision-making." Their article stresses scalability in visualization technologies and their ability to track provenance in real-time.
- MapReduce&Hadoop: MapReduce is a programming model used to handle a lot of data simultaneously and Hadoop is one of the more popular open-source implementations of that model. Szalay and Blakeley wrote an article [2] In Hey, Tansley, and Tolle's The Fourth Paradigm (2009) in which they discuss this particular software. They explain that the principles MapReduce uses are similar to the "distributed grouping and aggregation capabilities that have existed in parallel relational database systems for some time" but they are able to scale very well to accommodate for exceptionally large data sets. They go on to explain that Hadoop implements a "data-crawling strategy over massively scaled-out, share-nothing data partitions" where various nodes in the system are able to perform different parts of a query on different parts of the data simultaneously. This works very well for big data, but for smaller projects they remind their readers that this product isn't as effective "when a good index might provide better performance by orders of magnitudes."
- NoSQL databases:

These databases are designed specifically to deal with very large amounts of information that don't utilize a relational model. Theyscale very well and are often useful for tracking and analyzing real-time lists which grow quickly. These cover some of the more common technologies used in big data analytics. But not everyone will use all these techniques and technologies for every project. Anyone involved in big data analytics must evaluate their needs and choose the tools that are most appropriate for their company or organization. These needs change, not only from business to business, but also from sector to sector. Now that some of the tools and techniques have been examined the application of big data in various sectors can be more closely examined.

Benefits of Big Data Analytics:

We just saw that user organizations have adopted big data analytics in appreciable numbers. To determine the potential benefits that are driving the adoption, TDWI's survey asked: "Which of the following benefits would ensue if your organization implemented some form of big data analytics?" The most likely benefits are those most often selected by survey respondents, and the likelihood of a benefit declines as the list proceeds downward.

Advanced analytics is common, and big data analytics has a good presence.Big data analytics can benefit customer relations, business intelligence, and many analytic applications.We practice some form of advanced analytics, but not with big data40%.W e practice some form of advanced analytics, and we apply it to big data34%.We do not practice any form of advanced analytics, and we are not leveraging big data via analytics.23%Don't know3%

40% . We practice some form of advanced analytics, but not with big data.We practice some form of advanced analytics, 34% and we apply it to big data.We do not practice any form of advanced 23% analytics, and we are not leveraging big data via analytics.

Don't know 3%The State of Big Data Analytics Anything involving customers could benefit from big data analytics. Near the top of the list (in Figure 5), this includes better-targeted social-influencer marketing (61%), customer-base segmentation (41%), and recognition of sales and market opportunities (38%). Recent economic changes worldwide have changed consumer behaviors. Big data analytics can help develop definitions of churn and other customer behaviors (35%), as well as an understanding of consumer behavior from clickstreams (27%). Business intelligence in general can benefit from big data analytics. This could result in more numerous and accurate business insights (45%), an understanding of business change (30%), better planning and forecasting (29%), and the identification of root causes of cost (29%). Specific analytic applications are likely beneficiaries of big data analytics. For example, consider analytic applications for the detection of fraud (33%), the quantification of risks (30%), or market sentiment trending (30%). At the leading edge, big data analytics might help automate decisions for real-time business processes such as loan approvals or fraud detection (37%). Potential benefits entered by survey respondents selecting "other" include customer loyalty, service experience optimization, healthcare delivery optimization, and supplier performance based on cost and quality.

Science and Technology:

[2] In Hey, Tansley, and Tolle's (2009) collection of articles The Fourth Paradigm there is an article by Parastatidis that calls for a fourth paradigm in science – a new research methodology. Parastatidis goes on to outline this need by saying that current technologies commonly used in science and research are great for things like managing and indexing data, but they fall short when asked to "discover, acquire, organize, analyze, correlate, interpret, infer, and reason." He ends his article by mentioning the MapReduce computational pattern and calls on the research community to develop equivalent platforms for knowledge-related actions like reasoning, aggregation, inference, etc. Parastatidis is not the only voice calling for change in this field. [1] Chen, Chiang, and Storey (2012) tell that the National Science Foundation now has a requirement that every project must provide a data management plan. NSF has a 2012 BIGDATA program supported by US funding that aims to advance the core scientific and technological means of managing, analyzing, visualization, and extracting useful information from large, diverse, distributed and heterogeneous data sets … encourage the development of new data analytic tools and algorithms [and] facilitate scalable, accessible, and sustainable data infrastructure. The call includes new systems in the scientific community capable of handling big data appears to have been answered.

Higher Education Education is utilizing available technologies more and more each year. [5] Picciano (2012) tells us that one third of the higher education population in 2010 enrolled in at least one fully online course and many more enrolled in blended courses (a mix of online and face-to-face teaching). Since teachers and students are trending towards increased use of technology for instruction there are a lot of recorded data generated. In fact, Siemens and Long (2011) claim that big data and analytics are going to be the biggest factors in what is going to shape the future of higher education. [5] Picciano (2012) gives at least four areas within higher education that can benefit from big data analytics. Siemens and Long (2011) list nine. Among these are recruitment and admissions processing, financial planning, student performance monitoring, administrative decision making, donor tracking, providing help to at-risk students, understanding an institutions successes and challenges, recognizing the hard and soft value of faculty activities, and others. Perhaps the most interesting of these is student performance monitoring. [5] Picciano (2012) cites a fascinating case of a school in Arizona that uses an

analytics program to track student progress on the website of online courses they offer. They track all student activity: login/logout information, number of mouse clicks, number of page views, how long students viewed each page, student post content, etc.

IV. CONCLUSION

Today we see information overload almost everywhere. Big data analytics is trying to take advantage of the excess of information to use it productively. The benefits are many and varied, ranging from higher quality education to cutting-edge medical research, and while further research is needed for things like ensuring people's information is protected from exploitation, there are many exciting discoveries waiting to be uncovered through big data analytics.

V. REFERENCES

[1]. Chen, H., Chiang, R. H. L., &Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly, 36(4).

[2]. Hey, T., Tansley, S., & Tolle, K. (Eds.). (2009). The fourth paradigm data-intensive scientific discovery. Redmond, Wash.: Microsoft Research.

[3]. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity (pp. 1–143).The McKinsey Global http://www.mckinsey.com/insights/businesstechnology/bigdata_the_next_frontier_for_innovation.

[4]. Miller, K. (n.d.). Big Data Analytics in Biomedical Research. Biomedical Computation Review, (Winter 2011/2012), 14–21.

[5]. Picciano, A. G. (2012). The Evolution of Big Data and Learning Analytics in American Higher Education. Journal of Asynchronous Learning Networks, 16(3), 9– 20.

[6]. Russom, P. (2011). TDWI Best Practices Report: Big Data Analytics (Best Practices) (pp. 1–35). The Data Warehouse Institute (TDWI). Retrieved from

[7]. Siemens, G., & Long, P. (2011). Penetrating the Fog: Analytics in Learning and Education. EDUCAUSE Review, 46(5), 30–32,.