THE IMPLICATIONS OF BIG DATA ANALYTICS AS NEW EDUCATION POLICY AND GOALS

[1]Veerla.NagamalleswaraRao, [2]T.Naga Prasad Rao
[1]Lecturer, Department of Computer Science, A.G&S.G Siddhartha College of Arts &Science,
[2]Lecturer, Department of Computer Science, A.G&S.G Siddhartha College of Arts &Science, Vuyyuru, AP, India, PIN:521165

ABSTRACT:

Big data exists in a wide variety of data-intensive areas such as atmospheric science, genome research, astronomical studies, and network traffic monitor. Big data exists in a wide variety of data-intensive areas such as atmospheric science, genome research, astronomical studies, and network traffic monitor. Collecting, analyzing and understanding Big Data is becoming a differentiated strategy today but it will become a fact of life tomorrow. A lot of work is done on big data. Now it is time to know its impact on welfare of the society. Therefore, the aim of the book is to understand real use cases, benefits, advantages, impact and future challenges

INTRODUCTION:

Social media has brought about a revolution and dictated a paradigm shift in the operational strategies of firms globally. It has resulted in collection of massive data from a variety of social media channels, necessitating use of this data for business intelligence purposes. Despite its importance, little research exists on the implications of the use of Big Data analytics for business intelligence purposes. This study fills this gap in knowledge by examining the role and implication of Big Data analytics on business intelligence for the data collected from Social media channels in India.

2.1 Key Characteristics of Big data

- **Volume(Scale):** The quantity of data produced today by different sources such as Internet usage, social networks, mobile devices, sensors, embedded systems, and enterprise IT is rapidly growing greater than anything ever seen. Data volume is increasing exponentially, which is a challenge of Big data.
- **Variety (Complexity**): It is the nature of data that exists within big data. This includes different data formats, data semantics and data structures types.
- **Velocity (Speed):** It is also about the rate of changes, about linking data sets that are coming with different speeds and about bursts of activities, rather than habitual steady tempo. It is important to realize that events in data arise out of the available data.
- **Harness of data:** As there is a huge amount of unformatted data generated, one should even know how to harness the data and process it for better decision making. These are the different ways in which the data can be processed.
- **Operational Database:** (OLTP) Online Transaction Processing (DBMSs). Data Warehousing: (OLAP) Online Analytical Processing. Stream Computing: Due to uninterrupted networks data is arriving in continuous stream 24/7 a day. This data need to be captured and processed using RealTime Analytics Processing (RTAP).

Big data isn't just a process for storing petabytes or exabytes of data in a data warehouse, (Wiley, et al, 2013). It's about the ability to make better decisions and take meaningful actions at the right time. The biggest challenge for most E-Commerce businesses is to collect, store and organize data from multiple data sources. Big data analytics refers to the strategy of analyzing large volumes of data, or big data.

This big data is gathered from a wide variety of sources, including social networks, videos, digital images, sensors, and sales transaction records. The aim in analyzing all this data is to uncover patterns and connections that might otherwise be invisible, and that might provide valuable insights about the users who created it.

2.2 Big Data Technology Some the resent technology emerged in Big data are:

- **Hadoop:** An open source (free) software framework for processing huge datasets on certain kinds of problems on a distributed system. Its development was inspired by Google's MapReduce and Google File System. It was originally developed at Yahoo! and is now managed as a project of the Apache Software Foundation.
- **HBase**: An open source (free), distributed, no relational database modeled on Google's Big Table. It was originally developed by power set and is now managed as a project of the Apache Software foundation as part of the Hadoop.
- **MapReduce:** A software framework introduced by Google for processing huge datasets on certain kinds of problems on a distributed system. Also implemented in Hadoop.
- **Cassandra:** An open source (free) database management system designed to handle huge amounts of data on a distributed system. This system was originally developed at Facebook and is now managed as a project of the Apache Software foundation.
- **Extract, transform, and load:** (ETL)Software tools used to extract data from outside sources, transform them to fit operational needs, and load them into a database or data warehouse.
- **Cloud computing:** Cloud computing as a computing paradigm in which highly scalable computing resources, often configured as a distributed system, are provided as a service through a network.
- **Data warehouse:** Specialized database optimized for reporting, often used for storing large amounts of structured data. Data is uploaded using ETL (extract, transform, and load) tools from operational data stores, and reports are often generated using business intelligence tools.
- **Data mart:** Subset of a data warehouse, used to provide data to users usually through business intelligence tools.
- **Google File System:** Proprietary distributed file system developed by Google; part ofthe inspiration for Hadoop.

3. Proposed Work

3.1 Impact of big data on business:

3.1.1. Data will become an asset to every business.

Even the smallest businesses generate data these days. If the business has a website, a social media presence, accepts credit cards etc., even a one-person shop has data it can collect on its customers, its user experience, web traffic, and more. This means companies of all sizes need a strategy for big data and a plan of how to collect, uses, and protect it. This also means that savvy businesses will start to offer data services to even very small companies.

It also means that businesses and industries that never thought big data would be "for them" might be scrambling to catch up. Let me just make this as plain as possible: If you own or operate a business, and you have questions about how to improve that business, you have data, your data is an asset, and it can be used to improve your business.

3.1.2. Big data will enable companies to collect better market and customer intelligence.

Like it or not, the companies you do business with know a lot about you and the quantity and diversity of what they know about you is increasing every year. Every company (from car manufactures who will monitor our driving to tennis racket manufacturers that know how often and how well we play) will get much better insights into what customers want, what they will use, what channels they use to buy, and so on.

The other half of this equation is that companies will need to be proactive about creating and maintaining their privacy policies and all the systems and security needed to protect that user data. As we've seen with the recent backlash against spotify to a lesser extent Microsoft 10, most people will allow companies to gather this data, but they want transparency around what's being collected and why and they want the ability to opt-out.

3.1.3. It will improve internal efficiency and operations

From using sensors to track machine performance, to optimizing delivery routes, to better tracking employee performance and even recruiting top talent, big data has the potential to improve internal efficiency and operations for almost any type of business and in many different departments.

Companies can use sensors to track shipments and machine performance, but also employee performance. Companies have started using sensors to track employee movements, stress, health, and even who they converse with and the tone of voice they use.

In addition, if data can successfully be used to quantify what makes a good CEO, it can be used to improve the HR and hiring process at any level. Data is breaking away from the IT department and becoming an integral part of every department in a company.

3.1.4. And data will allow companies to improve the customer experience and build big data into their product offering.

In the best of all possible worlds, companies will use the data they collect to improve their products and the customer experience. John Deere is an excellent example of a company that is not only using data to benefit its customers, but also as a new product offering.

All new John Deere tractors are equipped with sensors that can help the company understand how the equipment is being used, and predict and diagnose breakdowns. But they've also put the sensors to work for the farmers, offering access to data about when to plant, where, the best patterns for reaping, and more. It's become an entirely new revenue stream for an old company. As we invite more connected things into our lives from smart thermostats to Apple Watches and fitness trackers there will be more and more data, analysis, and insights that companies can sell back to consumers. These are just the top four impacts I predict big data will have on businesses of all types in the near future. Is your business ready to capitalize on these opportunities?

3.1.5 Challenges for in bigdata analytics in small business:

3.1.5.1. Get on Board

Big data is a mega-trend every bit as powerful as the alphabet, the printing press, or the Internet itself. It will inevitably transform the competitive landscape, and it will happen faster than you think. It took only 2.5 years to sell 100 million iPads, so how long will it take before 100 million people are wearing an Apple iWatch or Google Glass? And how long before your industry feels the effects of the shift from console transactions to programmatic transactions?

3.1.5.2. We're in an Arms Race Powered by Data

To remain competitive you need to be relentless in your pursuit of data. Is there data you can collect about your prospects, customers, products or employees that'll give you an edge over the competition? And just as important, do you have a real-time framework to turn that data into action, so you're able to programmatically increase the probability of good outcomes for your business, while decreasing bad outcomes?

3.1.5.3. Partner with Vendors Who Understand Predictive

While some companies have the resources to collect big data and build their own proprietary applications, it is likely that best-of-breed vendors have better end-to-end solutions. We've already seen this for A/B Testing, Search Optimization, Display Retargeting, Lead Scoring, and Product Adoption. In each of these categories, companies can ride the big data wave without having to hire a single data scientist.

We're living in an exciting time, especially for those who can embrace being awash in data. We need to find ways to profit from it rather than drown in it. By eliminating the deadweight in your business, you can shift energy to more productive initiatives and blow past the competition. Being a data-driven business means more quickly anticipating customer needs, and finding new ways to meet them so you can guarantee the most profitable outcomes.

3.1.6 Recent Trends in big data analytics

3.1.6.1. Big data analytics in the cloud

Hadoop, a framework and set of tools for processing very large data sets, was originally designed to work on clusters of physical machines. That has changed. "Now an increasing number of technologies are available for processing data in the cloud. Examples include Amazon's Redshift hosted BI data warehouse, Google's BigQuery data analytics service, IBM's Bluemix cloud platform and Amazon's Kinesis data processing service. Smarter Remarketer, a provider of SaaS-based retail analytics, segmentation and marketing services, recently moved from an in-house Hadoop and MongoDB database infrastructure to the Amazon Redshift, a cloud-based data warehouse. The Indianapolis-based company collects online and brick-and-mortar retail sales and customer demographic data, as well as real-time behavioral data and then analyzes that information to help retailers create targeted messaging to elicit a desired response on the part of shoppers, in some cases in real time.

3.1.6.2. Hadoop: The new enterprise data operating system

Distributed analytic frameworks, such as MapReduce, are evolving into distributed resource managers that are gradually turning Hadoop into a general-purpose data operating system. With these systems you can perform many different data manipulations and analytics operations by plugging them into Hadoop as the distributed file storage system.

What does this mean for the enterprise? As SQL, MapReduce, in-memory, stream processing, graph analytics and other types of workloads are able to run on Hadoop with adequate performance, more businesses will use Hadoop as an enterprise data hub. The ability to run many different kinds of [queries and data operations] against data in Hadoop will make it a low-cost, general-purpose place to put data that you want to be able to analyze.

3.1.6.3. Big data lakes

Traditional database theory dictates that you design the data set before entering any data. A data lake, also called an enterprise data lake or enterprise data hub, turns that model on its head, says Chris Curran, principal and chief technologist in PricewaterhouseCoopers' U.S. advisory practice. It says we'll take these data sources and dump them all into a big Hadoop repository, and we won't try to design a data model beforehand. Instead, it provides tools for people to analyze the data, along with a high-level definition of what data exists in the lake. People build the views into the data as they go along. It's a very incremental, organic model for building a large-scale database,. On the downside, the people who use it must be highly skilled.

3.1.6.4. More predictive analytics

With big data, analysts have not only more data to work with, but also the processing power to handle large numbers of records with many attributes. Traditional machine learning uses statistical analysis based on a sample of a total data set. You now have the ability to do very large numbers of records and very large numbers of attributes per record and that increases predictability. The combination of big data and compute power also lets analysts explore new behavioral data throughout the day, such as websites visited or location.

3.1.6.5. SQL on Hadoop: Faster, better

If you're a smart coder and mathematician, you can drop data in and do an analysis on anything in Hadoop. That's the promise — and the problem, says Mark Beyer, an analyst at Gartner. "I need someone to put it into a format and language structure that I'm familiar with," he says. That's where SQL for Hadoop products come in, although any familiar language could work, says Beyer. Tools that support SQL-like querying let business users who already understand SQL apply similar techniques to that data. SQL on Hadoop "opens the door to Hadoop in the enterprise," Hopkins says, because businesses don't need to make an investment in high-end data scientists and business analysts who can write scripts using Java, JavaScript and Python — something Hadoop users have traditionally needed to do.

3.1.6.7 More, better NoSQL

Alternatives to traditional SQL-based relational databases, called NoSQL (short for "Not Only SQL") databases, are rapidly gaining popularity as tools for use in specific kinds of analytic applications, and that momentum will continue to grow, says Curran. He estimates that there are 15 to 20 open-source NoSQL databases out there, each with its own specialization. For example, a NoSQL product with graph database capability, such as ArangoDB, offers a faster, more direct way to analyze the network of relationships between customers or salespeople than does a relational database.

3.1.6.8. Deep learning

Deep learning, a set of machine-learning techniques based on neural networking, is still evolving but shows great potential for solving business problems. Deep learning . . . enables computers to recognize items of interest in large quantities of unstructured and binary data, and to deduce relationships without needing specific models or programming instructions.

In one example, a deep learning algorithm that examined data from Wikipedia learned on its own that California and Texas are both states in the U.S. It doesn't have to be modeled to understand the concept of a state and country, and that's a big difference between older machine learning and emerging deep learning methods.

Big data will do things with lots of diverse and unstructured text using advanced analytic techniques like deep learning to help in ways that we only now are beginning to understand. For example, it could be used to recognize many different kinds of data, such as the shapes, colors and objects in a video or even the presence of a cat within images, as a neural network built by Google famously did in 2012. This notion of cognitive engagement, advanced analytics and the things it implies are an important future trend.

3.1.6.9. In-memory analytics

The use of in-memory databases to speed up analytic processing is increasingly popular and highly beneficial in the right setting, says Beyer. In fact, many businesses are already leveraging hybrid transaction/analytical processing (HTAP) allowing transactions and analytic processing to reside in the same in-memory database. But there's a lot of hype around HTAP, and businesses have been overusing it, Beyer says. For systems where the user needs to see the same data in the same way many times during the day and there's no significant change in the data in-memory is a waste of money.

REFERENCE:

1. Ahmad W. and Quadri B.S.M.K. Big Data promises value: Is hardware technology taken onboard?. Industrial Management & Data Systems. 115(9), 2015.
2. Assunção M.D., Calheiros R.N., Bianchi S., Netto M.A.S. and Buyya R. Big Data computing and clouds: Trends and future directions. Journal of Parallel and Distributed Computing. 79–80 (2015),
3. Avram (Olaru) M.G. Advantages and challenges of adopting cloud computing from an enterprise perspective. Procedia Technology. 12(2014), 529 – 534.
4. Davenport T.H. How strategists use 'big data' to support internal business decisions, discovery and production. Strategy & Leadership. 42(4), (2014), 45 – 50.
5. Erevelles S., Fukawa N. and Swayne L. Big Data consumer analytics and the transformation of marketing. Journal of Business Research. 69(2), (2016), 897-904.
6. García-Valls M., Biug T. and Lu C. Challenges in real-time virtualization and predictable cloud computing. Journal of Systems Architecture. 60(2014), 726–740.
7. Gandomi A. and Haider M. Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management. 35(2015), 137–144.
8. http://www.computerworld.com/article/2690856/big-data/8-big-trends-in-big-data-analytics.html
9. Ajit Roy, "Impact of Big Data Analytics on Business, Economy, Health Care and Society.Paradox,https://www.amazon.in/Analytics-Business-Economy-Society-Paradox-Analytics-ebook/dp/B015QWL352