

Implementing Big Data for Intelligent Business Decisions

Dr. V. B. Aggarwal | Deepshikha Aggarwal

1(Jagan Institute of Management Studies, Delhi, India, vbaggarwal@jimsindia.org)

2(Jagan Institute of Management Studies, Delhi, India, deepshikha.aggarwal@jimsindia.org)

Abstract - BIG DATA is a buzz word these days both in the IT sector and the business world. The IT industry is busy selling the idea of big data to every corporation. We have conducted this research to understand what actually is big data and how important it is for the companies to use it in their business operations. Big data is not a technical term; it's used differently by different sectors. Any large collection of data can be called a big data with obviously a few characteristics that make it suitable to be used for decision making in businesses. Thus big data is more about decision making and business intelligence, hence it's not just about large volumes of data but the realization of greater business intelligence by storing, processing, and analyzing data that was previously ignored due to the limitations of traditional data management technologies. Data has always been there but until the last decade, it was never used in the way it's done now. Big data is about data analytics. Big data is an opportunity to find insights in new and emerging types of data and content, to make the businesses more agile, and to answer questions that were previously considered beyond reach. The potential of Big Data is in its ability to solve business problems and provide new business opportunities.

KEYWORDS: Big Data, Data Management, Big Data Analytics, Business Decision Making, Database Systems, Aids to Human Mind Management

1. INTRODUCTION TO BIG DATA

Big data is a collection of data sets so large and complex that it becomes difficult to process using on-hand state-of-art database management tools [9]. The challenges include capture, storage, search, sharing, analysis, and visualization. Put another way, big data is the realization of greater business intelligence by storing, processing, and analyzing data that was previously ignored due to the limitations of traditional data management technologies. This is the age of data explosion. 2.5 quintillion (2500 million billion) bytes of data are generated every day where a quintillion is 10¹⁸ bytes of data. Data comes from many quarters such as social media sites, emails, sms, sensors, digital photos, business transactions, location-based data, etc. Key enablers for the growth of "Big Data" are [5]:

- Increase of storage capacities
- Increase of processing power
- Availability of data

The characteristics of Big Data [11]:

Volume: A large heterogeneous volume of data is obviously the key characteristic of Big data. Many factors contribute to the increase in data volume. This data can be the transaction-based data generated through the day to day business transactions stored through the years. E-Commerce sites like Flipkart, Amazon, etc. are another source of Big Data as the number of online transactions is increasing day by day. Unstructured data streaming in from social media such as the Facebook page or the company website is a major contributor to big data. Increasing amounts of sensor and machine-to-machine data is being collected. In the past, excessive data volume was a storage issue. But with decreasing storage costs, other issues emerge, including how to determine relevance within large data volumes and how to use analytics to create value from relevant data.

Velocity: Quickly moving data as we are dealing with information accumulation in real world. Data is streaming in at unprecedented speed and must be dealt with in a timely manner. RFID tags, sensors and smart metering are

driving the need to deal with torrents of data in near-real time. Reacting quickly enough to deal with data velocity is a challenge for most organizations.

Variety: Data is not being collected in standard formats but we have structured data, unstructured data, images, videos taken through CCTV at sensitive locations like airports, etc. Data today comes in all types of formats. Structured, numeric data in traditional databases. Information created from line-of-business applications. Unstructured text documents, email, video, audio, stock ticker data and financial transactions. Managing, merging and governing different varieties of data is an area of concern for organizations.

Veracity: Trust and integrity is a challenge and is as important for big data just as for traditional relational Databases. However, it is necessary to connect and correlate relationships, hierarchies and multiple data linkages or the data can quickly spiral out of control.

Web 2.0 is all about data whether we talk about Facebook, twitter, LinkedIn, Instagram, its more and more data. All the Facebook posts or the tweets on twitter including the pictures and videos, everything is getting stored and making up the Big Data. More sources are generating more types of Data (Structured and unstructured Data)

Big data is the opportunity to extract insight from an immense volume, variety and velocity of data, in context, beyond what was previously possible. Massive volume, variety and velocity are defining characteristics of Big Data. It is obvious that a Big Data Problem will have, Big Data Volume. This Volume can start in the Terabytes and quickly move to the 100's of Petabytes. New Storage solution is needed to be used to have this type of volume. Velocity is obvious as well because no organization wants their answers slower! We hear demands for new insights or analytics ranging from; "we need it in 4 hours not 4 weeks" to "response must be real time, that is sub-second response." The third "V", Variety, is the least understood, and could be the most profound [3]. Big Data is derived mostly from sources not analyzed or not used before. Data

is not derived from classical transaction systems which lend themselves to structured models.

Data is exploding [1] in volume, variety and velocity as both structured and unstructured information will continue to grow at astronomical rates. This creates a tremendous opportunity for organizations to make timely decisions and achieve business goals. However, at the same time, organizations are struggling to gain deeper insights from this data. Business leaders continue to make decisions without access to the trusted information they need. CEOs understand that they need to do a better job in capturing and understanding information

2. HOW IS BIG DATA GENERATED

The amount of data which large organizations have collected is horrendously large, mind boggling and extremely difficult to comprehend by human mind. Just imagine the size the data, text, attachments and images stored on google servers as millions of emails are sent across the globe on daily basis. There is a great wealth of useful information waiting up to be exploited from there. Once the data can be analyzed in a meaningful and efficient manner, number of useful business decisions can be made. This is where big data analytics comes in to aid the human mind management. This is just one example. Imagine the videos and audios stored on YouTube, geographical information stored in maps, etc. On the other hand large banks and insurance companies have got huge depositories of data about customers, transactions, product information over the years. Formats of storage have been changing and it is not possible to use the traditional data management tools to handle the data as the data has become heterogeneous and unstructured in a major way. So big data analytics comes to the rescue of businesses and helps them to exploit the true potential of data collected for business decision making and strategy formulation.

Big data is generated by businesses themselves and by various other sources. Transaction data is the existing data generated on a day to day basis through all the business transactions happening in an organization and is the main source of big data. The main use of transactional data is to answer various queries and generate reports. In an RDBMS used for transactions, every time you update/delete/insert and select (queries), this information is logged. The business user can analyze this data for business purposes. Another source of data is "log data." which is generated in companies with large warehouses of product lines such as big stores. Walmart is probably the largest company using RFIDs (radio frequency ID technology) to allow it to track everything it has to deal with around the world. Another example of data collected through RFIDs is the metro rail where the passenger tokens are scanned to generate log data. This is sensor data that can be helpful to analyze use and inefficiencies [2].

Most traditional data was structured, or neatly organized in databases. Then the world went digital and the internet came along which led to unstructured data generated by all our digital interactions, from email to online shopping, text messages to tweets, Facebook updates to YouTube videos. Social Media (Facebook, Twitter, etc.) are large contributors to big data. Twitter generates approximately 8 TB of data everyday where Facebook generates around 10

TB. As the number of mobile phones grows globally, so does the volume of data they generate from call logs, texts, emails, social media updates, photos, videos, and location services. The number of gadgets recording and transmitting data, from smartphones to intelligent fridges, industrial sensors to CCTV cameras, has proliferated globally, leading to an explosion in the volume of data. These data sets are now so large and complex that we need new tools and approaches to make the most of them. Since all this data is the unstructured data, it cannot be managed by the traditional systems for data management and needs big data analytics to get analyzed and used in business decision making.

3. BIG DATA ANALYTICS

Big data is more than simply a matter of size; it is an opportunity to find insights in new and emerging types of data and content, to make the businesses more agile, and to answer questions that were previously considered beyond reach. Imagine if one could analyze the 12B TB of tweets being created each day to figure out what people are saying about their products, figure out who the key influencers are within the target demographics. Being able to mine this data to identify new market opportunities. What if hospitals could take the thousands of sensor readings collected every hour per patients in ICUs to identify subtle indications that the patient is becoming unwell, days earlier that is allowed by traditional techniques. Imagine if one could make risk decisions, such as whether or not someone qualifies for a mortgage, in minutes, by analyzing many sources of data, including real-time transactional data, while the client is still on the phone or in the office [4].

As these new sources of data continue to grow in volume, variety and velocity, so too does the potential of this data to revolutionize the decision-making processes in every industry. Aspects of the way in which users want to interact with their data:

Totality: Users have an increased desire to process and analyze all available data.

Exploration: Users apply analytic approaches where the schema is defined in response to the nature of the query.

Frequency: Users have a desire to increase the rate of analysis in order to generate more accurate and timely business intelligence.

Dependency: Users' need to balance investment in existing technologies and skills with the adoption of new techniques.

The focus is on supporting redundancy, distributed architectures, and parallel processing. Most data that is in a Big Data Platform is unstructured, or has part of it unstructured. It is not just that classic data warehouse platforms cannot store and access those volumes, it is that to cost and labor associated with storing that data on those platforms is prohibitively expensive and difficult to deploy. Answers that businesses are demanding have to be based on increasingly sophisticated analytics and the rate of response demanded can be an order of magnitude much faster.

In order to capitalize on this opportunity, enterprises must be able to analyze all types of data – relational and non-relational. Texts, sensor data, audio, video,

transactional. Sometimes, getting an edge over the competition can mean identifying a trend, problem or opportunity, seconds, or even microseconds before someone else. More and more of the data being produced today, has a very short life. Organizations must be able to analyze this data in real-time if they are to be able to find insights in this data. Big data analytics is THE opportunity to extract insight from an immense volume, variety and velocity of data, in context, beyond what was previously possible.

4. BIG DATA ANALYTICS IMPLEMENTATION

The Big Data analytics approach complements the traditional approach of business data management. The traditional approach business users determine what questions to ask and IT structures the data to answer that question. This is well suited to many common business processes, such as monitoring sales by geography, product or channel; extract insight from customer surveys; cost and profitability analyses. In the Big Data approach IT delivers a platform that consolidates all sources of information and enables creative discovery. Then the business users use the platform to explore data for idea and questions to ask. Most of the time, the data are raw data.

One of the most popular tools for Big Data analysis is HADOOP. Hadoop is a distributed file system and data processing engine that is designed to handle extremely high volumes of data in any structure [8]. Hadoop can handle all types of data from disparate systems: structured, unstructured, log files, pictures, audio files, communications records, email – just about anything one can think of, regardless of its native format. Even when different types of data have been stored in unrelated systems, it can all be dumped into the Hadoop cluster with no prior need for a schema. In other words, users don't need to know how they intend to query their data before they store it; Hadoop allows the users to frame queries as per their business requirements at any point of time. It makes all the data usable, not just what's stored in databases, Hadoop allows the formation of relationships that were hidden before and reveal answers that have always been just out of reach. Businesses can start making more decisions based on actual data instead of depending on the perceptions of managers and look at complete data sets, not just samples [12].

Hadoop has two components:

The Hadoop distributed file system (HDFS), which supports data in structured relational form, in unstructured form, and in any form in between these two forms. Thus Hadoop allows the users to store files bigger than what can be stored on one particular node or server. So huge large files and large number of files can be stored using Hadoop.

The second characteristic of Hadoop is its ability to process data, or provide a framework for processing the data. This is called MapReduce. The MapReduce programming paradigm is used for managing applications on multiple distributed servers. Rather than the conventional method of data processing where the data is moved over a network to be processed by software, MapReduce uses a smarter approach tailor made for big data sets. Moving data over a network can be very slow, especially for huge data sets, so rather than moving the data to the software, MapReduce moves the processing software to the data.

Many people think big data is all about the Hadoop technology but it's a lot more than that. One of the key requirements is to understand and navigate the heterogeneous sources of big data to discover data in place. New technologies have emerged that discover, index, search, and navigate diverse sources of big data. Data warehouses also manage big data but the applications are limited to the huge volumes of structured data. It requires massive parallel processing data warehouses and purpose-built appliances for big data analytics in data warehouses. Big data is in motion and the streaming data represents an entirely different big data problem – the ability to quickly analyze and act upon data while it's still moving. Big data analytics opens a world of possibilities – from processing volumes of data that were just not practical to store, to detecting insight and responding quickly. As much of the world's big data is unstructured and in textual content, for example the data generated from emails, text messages and social media posts, text analytics is a critical component to analyze and derive meaning from text. And finally the integration and governance technology – ETL, data quality, security, and lifecycle management establishes the veracity of big data, and is critical in determining whether information is trusted because whenever data analytics is happening, the results depend upon the accuracy of the information provided to the system.

5. BIG DATA FOR MEDIUM & SMALL BUSINESSES

Regardless of the size of a company, big data can be put to use for business because big data is not just about big volumes but it is the opportunity to learn from a company's existing data in order to make smarter business decisions. And as large, enterprise companies are taking advantage of the opportunity that big data analytics offers, medium and small businesses can do the same [4].

For a medium and small business, the ability to succeed is driven by the ability to understand the market, business forecast and ability to provide a superior service. Big data is changing the business landscape by helping the businesses to enhance product quality, improve marketing operations and further customer relationships.

It's not the size of the data that matters, but how it is used. Many medium and small business owners, think that they do not have sizeable data as compared to larger companies. But it's time to realize that the size of the data is less important than the integrity or ability to act upon it. Companies, whether medium and small or large, are collecting consumer data at an unprecedented rate, and the ability to integrate all of the different sources of data and shape it in a way that allows business leaders to make informed decisions is essentially at the heart of Big Data. Big Data benefits businesses in many ways, from supply chain logistics to financial forecasting. For the most part, however, Big Data is essentially a means to leverage whatever data sources are available to create a big-picture view of customer behaviors and preferences to personalize experiences,

If a business organization can gain a holistic view of their customer by connecting all the different data capture points, they have the opportunity to vastly improve both the immediate user experience and the long-term

perception of their brand. In the age of the empowered consumer, the execution of Big Data strategies will prove to be one of the largest differentiators for everything from medium and small business to Fortune 500 companies, particularly within retail and media industries.

Many medium and small business owners are confused by what Big Data is, and many also don't understand its purpose. A lot has been said and done in big data in big organizations but it's time to divert the attention towards medium and small businesses as well Big Data is just a buzz word for the businesses who want to capture as much information about a person/ community/ group as they possibly can and figure out how to make money off of that information. In retail, for example, they might want to know if you open emails, how often you purchase in-store versus online, how often you browse the website before purchasing, what you're posting on your social media accounts, and if you open promotional emails. To effectively mine and analyze Big Data — and get valuable results — what companies need is a strategy. Companies don't have to invest millions and gather a lot of data to be successful with Big Data, but they just have to start with a very concrete and focused question they're trying to solve, think through what data they need to solve it, and go get that data. For medium and small businesses, this means figuring out where data can be found and what problems this data can solve. medium and small businesses should just focus on all the different data that they have on customers and use this to answer the business problems identified already [6].

6. HOW TO APPROACH BIG DATA

Data that is collected without a particular set of questions in mind is unlikely to be useful. It should start with asking the right questions. For example, consider a company with a customer loyalty program that was originally designed to be a rewards program, not a strategic data collection initiative. The data that it provides likely answers many key questions. These include: Which customers responded to a particular promotion? What other items did they buy when they shopped? What was their total spending? Did the promotion affect the frequency of their visits?

Asking the right questions can also help businesses use Big Data to avoid costly mistakes. One example is a grocery store whose product data indicates that certain products, like organic foods, etc., are losers because of low sales. However, when such product data is integrated with customer data, it actually reveals that these "loser" products attract many of those customers who also buy many of the most profitable products. Thus to get the most out of big data, businesses require an integrated data strategy for collecting and analyzing data.

The following key tools for Big data analytics are provided by different vendors [13]:

□ Discovery tools that 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; users can draw new insights, come to meaningful conclusions, and make informed decisions quickly.

□ Business Intelligence tools 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.

□ In-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.

□ Decision Management includes predictive modelling, 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.

7. HOW BIG DATA WORKS

Big Data can give businesses extremely valuable information, but it does not do anything on its own nor does it magically unveil the secrets to success. Businesses need proper analysis of the big data and put it to work for their benefit. Big Data confusion starts with the misconception that it is some sort of technology which it certainly is not. It's actually about using the ever-increasing amount of data to create value for your customers and your business. The technologies and analytical techniques used to create such value can be simple or complex, but knowing the right approach to Big Data can make all the difference in making the process less baffling. The secret is to first identify the problem that needs to be solved. Big Data, however, won't provide business altering insights automatically but it requires a person to frame the question, identify the data that might be able to answer the question, and interpret the results to choose an action. This is how big data is to be implemented for businesses and only in this way the businesses can explore the true potential of the big data analytics.

Instead of worrying about technology, businesses have to concentrate on how it applies to and creates value for them. If you are a medium and small business with very few products, as opposed to over 200 million products like Amazon, you might not benefit from a Product Recommendation Engine. But understanding how customers interact with your website might be critical. It all depends on your business model.

There are five broad ways in which using big data can create value.

(i) Big data can unlock significant value by making information transparent and usable at much higher frequency.

(ii) As organizations create and store more transactional data in digital form, they can collect more accurate and detailed performance information on everything from product inventories to sick days, and therefore exposes variability and boost performance. Leading companies are using data collection and analysis to conduct controlled experiments to make better management decisions; others

are using data for basic low-frequency forecasting to high-frequency nowcasting to adjust their business levels just in time.

(iii) Big data allows ever-narrower segmentation of customers and therefore much more precisely tailored products or services.

(iv) Sophisticated analytics can substantially improve decision-making.

(v) Big data can be used to improve the development of the next generation of products and services. For instance, manufacturers are using data obtained from sensors embedded in products to create innovative after-sales service offerings such as proactive maintenance where preventive measures take place before a failure occurs or is even noticed [4].

The following links can be accessed to read some of the success stories of Big data analytics for businesses:

http://www.csc.com/big_data/success_stories

<ftp://ftp.software.ibm.com/software/data/sw-library/big-data/ibm-big-data-success.pdf>

<http://www.forbes.com/sites/mckinsey/2012/12/03/big-data-advanced-analytics-success-stories-from-the-front-lines/>

<http://www.informationweek.com/big-data/big-data-analytics/big-data-success-3-companies-share-secrets/d/d-id/1111815?>

<http://searchcio.techtarget.com/opinion/Ten-big-data-case-studies-in-a-nutshell>

8. CONCLUSION

Every day, we create 2.5 quintillion bytes of data and 90% of the data in the world today has been created in the last two years alone. This data comes from everywhere: sensors used to gather climate information, emails going around the world, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few. This data is big data. The real issue is not that we are acquiring large amounts of data. It's what we do with the data that counts. The hopeful vision is that organizations will be able to take data from any source, harness relevant data and analyze it to find answers that enable 1) cost reductions, 2) time reductions, 3) new product development and optimized offerings, and 4) smarter business decision making. Big data is not just a matter of size but it is an opportunity to find insights in new and emerging types of data and content gathered from different sources, to make the businesses more agile, and to answer questions that were previously considered beyond the reach. Till few years back, there was no practical way to harvest the opportunity to utilize the huge volumes of data but now various platforms for big data analytics have opened the door to a world of possibilities. Today's business leaders expect that big data analytics will enable more accurate business insights, allow them to better recognize sales and market opportunities and boost their company's social media marketing capabilities. With data growing so rapidly and the rise of unstructured data accounting for 90% of the data today, the time has come for enterprises to re-evaluate their approach to data storage, management and analytics. Legacy systems will remain necessary for specific day to day business transactions, and complement the use of Big data analytics for attaining the

long term goals of the organizations. But organizations in every industry are trying to take advantage of Big data analytics by developing systems and processes that can combine traditional structured data with semi-structured and unstructured sources of information. When properly captured and analyzed, big data can provide unique insights into market trends, customer satisfaction, buying patterns, product development, maintenance cycles and many other business issues, lowering costs, and enabling more targeted business decisions. To obtain value from big data analytics, organizations need a cohesive set of solutions for capturing, processing, and analyzing the data. The true power of Big data analytics will be harnessed when organizations are able to utilize the power of information systems, social media and Internet for effective decision making.

REFERENCES

- [1] <http://www.emc.com/collateral/analyst-reports/ar-the-economist-data-data-everywhere.pdf>
- [2] http://www.snia.org/sites/default/files/2/ABDS2012/Tutorials/RobPeglar_Introduction_AnalyticsBigData_Hadoop.pdf
- [3] http://www.greenplum.com/sites/default/files/SAS_BigData_final.pdf
- [4] http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation
- [5] http://www.planet-data.eu/sites/default/files/presentations/Big_Data_Tutorial_part4.pdf
- [6] <http://www.infosys.com/cloud/resource-center/Documents/big-data-spectrum.pdf>
- [7] <http://www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf>
- [8] Hadoop: The Definitive Guide (Paperback) by Tom White
- [9] Big Data: A Revolution That Will Transform How We Live, Work, and Think by Viktor Mayer-Schönberger
- [10] Big Data (ebook) by Nathan Marz
- [11] http://www.sas.com/en_us/insights/big-data/what-is-big-data.html
- [12] http://www.cloudera.com/content/cloudera/en/about_hadoop-and-big-data.html
- [13] <http://www.oracle.com/technetwork/database/options/advanced-analytics/bigdataanalyticswp0aa-1930891.pdf>