**BIG DATA ANALYTICS**

Term paper

For

Master of Science

Information and Communication Technology

Nittala Srireshmi

University of Denver University College

August 2, 2017

Faculty: Austin Klahn

Director: Tom Tierney

Dean: Michael McGuire

**Abstract**

In our present world, we are bombarded with data from all possible directions. How the data can be calibrated and used in a constructive way plays the major importance. Many enterprises across the globe have embarked on a path to digitally transform their organization, ranging from introduction of data processing at every step of existing business practices and all the way to driving the enterprise towards monetisation of its data assets.

This paper examines how big data works in various ways across several industries. The paper will also highlight ways in which several organizations are already influenced by big data practices. More significantly, this paper highlights how Big Data is making a difference in the banking sector by facilitating the digitization of traditional processes and by continually improving customer experiences.

**Table of Contents**

1.Introduction....................................................................................................................pg1-3

2. Key categories of Big Data Tools....................................................................................pg4-5

3. Benefits of applying big data trends to current industries.............................................pg6-7

4. Big Data in Financial Industry.......................................................................................pg8-10

5. Challenges and Other emerging factors in Big Data..................................................pg10-12

6. Future trends in Big Data...........................................................................................pg13-14

7. Conclusion.......................................................................................................................pg15

8. Bibilography....................................................................................................................pg16

**Introduction**

Big Data has a 40-year-old history and it refers to extremely large datasets that cannot be captured, stored, managed, or analysed using traditional databases. The definition of Big Data has been evolving over the years. Big Data today not only refers to the data itself, but also a set of technologies to manage the data. The technologies help solve complex problems and unlock value from the data to make it more economical. Recent Big Data advancements also refer to the use of predictive analysis and user behaviour analytics. The following graph depicts the evolution of worldwide interest around the term Big Data.

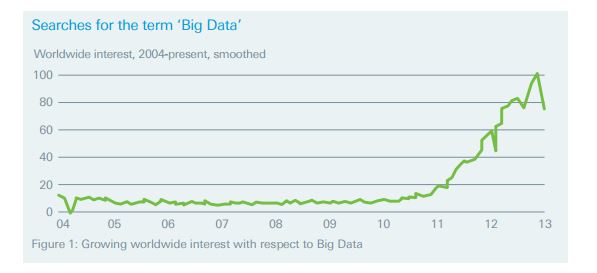


Figure 1: Growing worldwide interest with respect to Big Data

In 2001, according to a Gartner research report, challenges and opportunities related to data are three dimensional and can be categorized into Volume, Velocity and Variety. Industries today continue to use the 3V’s model for describing Big Data. In 2012, Gartner defined Big Data as ‘high volume, high velocity, and high variety information assets that demand cost effective, innovative forms of information processing that enable enhanced insight, decision making and process automation’.

The concept of Big data existed in many forms and tools that are often built by corporations with a special need. Commercial vendors historically offered parallel database management systems for big data as early as 1990s. For people new to the Big Data domain, it is important to know what the general understanding of Big Data is and clear out any misconceptions about this term. Here is a bar graph that tells us how people interpret the term Big Data.

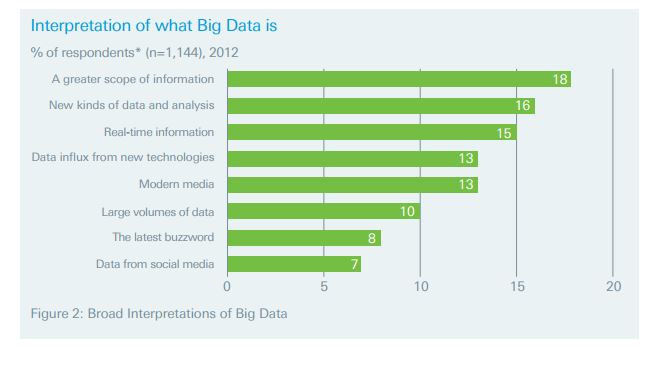


Figure 2: Broad Interpretations of Big Data

Successful business holders always consider data to be the fourth factor of production, as essential to business as land, labour, and capital. According to researches, the extensive use of Big Data these days has improved the performance of businesses by an average of 26% and the impact is estimated to grow by 42% in the coming five years.

In spite of the privacy concerns and organizational resistance, investments in Big Data continue to gain momentum worldwide. In the year 2014, the investments on Big Data are estimated to have reached 35billion U.S dollars. Technical experts believe that this rate is expected to grow up to 80billion USD by the end of 2020. Companies that consider data as their priority for their marketing and sales decisions, tend to increase their ROI (return on investment) by 15-25%. Smarter business decisions and increased revenue will continue to drive organizations make firm investments in big data. Big data has a potential to create more than 5million jobs by 2020, which shows the opportunities for analysts, computer scientists, mathematicians and other data job seekers.

**Key Categories of Big Data Tools**

Today, there exist a wide variety of tools that can be used to clean, store, process, and analyse big data. For easier understanding, the tools available can be categorized into Infrastructure, Data storage, Data processing & Management, and Data Analytics.

**Infrastructure:**

The most important feature of a big data infrastructure is the flexibility and scalability to manage at least terabytes of data. Generally, in a scenario where an organization has to deal with vast amounts of data, cloud-based services become a natural choice. Microsoft Azure, Amazon Web Services, Google Cloud Platform, and GoGrid are some of the existing cloud-based service providers that have already proved the benefits of the big data technologies.

**Data Storage:**

Traditional Legacy systems have storage methods that are sub-optimal due to its price and scalability restrictions. Today, new methods of storage, especially NoSQL and Distributed File Systems represent the changed hypothesis in storage era. Hadoop File System also popularly known as HDFS, is the most widely used storage system for Big Data. Over the years, numerous processing tools have been built on top of HDFS for faster data access and processing. Other familiar Big Data storage players include MongoDB and Cassandra.

**Data Processing & Management:**

It is not sufficient to store data overtime in large file systems like HDFS. What matters more is the ability to clean and quickly process unstructured data into readable format for data analysis. Fast Big Data processing frameworks like Spark, Storm, and Flink allow developers to write algorithms on top of terabytes of data available in storage systems. For easier management and monitoring of open source Big Data tools, organizations like Cloudera, Horton Works, and MapR provide enterprise and community versions.

**Data Analytics:**

Data Analytics can be defined as the ability to perform statistical analysis on Big Data, and to provide business intelligence which helps executives make important decisions regarding the business. Predictive analytics is another term closely associated with Big Data analytics. Being able to predict the future customer purchases based on past purchase behaviour is an example of predictive analysis that can be performed. Examples of data analytic tools in market include R, Spark, H20, and Splunk. Below is the graphical representation of some of the other popular categories of Big Data analysis tools.

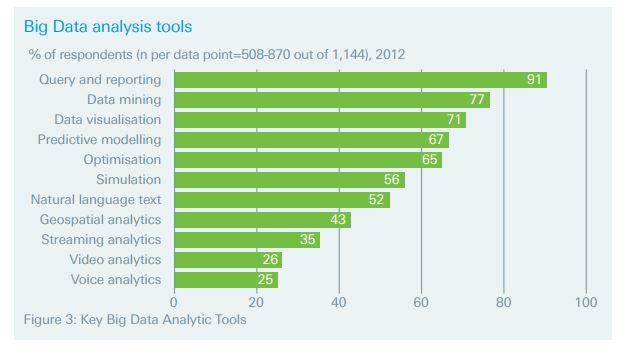


Figure 3: Key Big Data Analytic Tools

**Benefits of applying Big Data trends to current trends**

When companies add Big Data insights to their business, it benefits them in multiple ways. For example, a company operating retail business with an ecommerce platform can make use of big data to understand its site traffic, number of visits and activities such as pages viewed, purchasing history, to analyse its customers to improve their products and services. Tech giants like Google, Amazon and Facebook are facilitating smaller businesses to collect, store, and analyse big data. It has become one of the incredible trends in Information Technology history. Here are some of the industries in which big data has an impact.

**Retail industry**

To enhance the accuracy of forecasting and anticipation of changes in demand, many retail industries are already using Big Data analytics. *Brook Brothers* stand as a good example, as they introduced business analytics into their business which was developed by SAS. This transformation has helped in improving their stock management. Using analytics to forecast its global stock has enabled store managers to make better decisions about their stock levels and pricing. As a result, the number of times stores were out of stock when customers came in to purchase an item was reduced, and stock decreased by 27%.

**Engineering and Manufacturing**

In this field, big data can be used to predict maintenance problems, improve manufacturing quality and to manage costs. In car manufacturing and racing, the real time usage of SAP’s Hana data technology helps in analysing speed as well as to detect flex, vibration load, and many other measures that impact the machine performance. This real time analysis of car sensor data is compared with the historical data of previous performances to predict the performance based on newer features. This helps the teams to develop better performances and to make important corrections.

**Health care**

Currently in health care, big data techniques are widely used for predictive analysis towards critical events and for effective cost management. Recent improvements by health researchers include using big data analytics to improve quality of life in multiple sclerosis patients. Another example is the Human Genome project where genetic sequences available in really large files were processed using Big Data tools. IBM is also using their big data technology to capture and analyze the data from medical monitors, patient’s clinical manifest, and cancer origin illness.

**Transportation**

Using technology from SAS, Frankfurt airport’s air traffic controllers receive early warnings of storms. Managers can an access the overview of all key performance indicators near them including the average time for luggage delivery, delays, and other airport security levels. Some US truck companies uses data analytics and collects the data elements from numerous trucking systems like sensor data for fuel usage, engine operation and geospatial data etc. This helps in saving millions of dollars per year

**Government**

Major governments like UK and US have also accepted the usage of big data in to their systems. UK government was suggested by technical experts to raise funds on Big Data for fraud detection, which can save 2 billion pounds for them every year. In March, the US government and six federal agencies initiated their big data with an investment of 200million USD. It was so called as one of the most important public investments in to technology and this aims to the great access and understanding of public needs from the collection of huge volumes of data.

**Big Data in Financial Industry**

Financial industry is one of the important data oriented industries, which primarily needs big data. According to researches, it is proven that at the end of 2012 financial and security organizations were estimated to be managing 3.7 petabytes of data per firm. It is tedious to work with traditional databases with the immense amounts of data. It is evident that many large financial firms are already reaching their saturation limits in their legacy systems and are now seeking analytics systems and ERP frameworks.

Big data has a broad variety of data types that provide financial institutions the ability to tolerate and accept digital and physical channel interactions like customer data, geographical data, and graphical representations. Furthermore, most of the financial firm’s data is generally stored in an unstructured format. For example if we take insurance claim systems, the information stored will be in text format, but the ability of combining the text document and extracted information to structured data can make a big difference to provide better experience to customer and to improve their fraud detection abilities.

**Customer Understanding and Marketing analytics**

Now a days, customers maintain temporary relationships with many organizations simultaneously. So banks no longer have a detailed view of their customer preferences, their banking choices, buying and spending patterns. Facilities like PayPal, Amazon, Walmart and other peer-to-peer services have created a path to extract customer patterns to some extent and this has also disintermediated banks. Gaining more understanding on customer interests can help the banks to create richer predictive models.

In addition to utilizing just the internal resources, banks should invest on skilled employees and build the big data infrastructure to extract new and existing customer information from social networks and other important sources like call records, emails. American Express is an ideal example which has implemented this analysis approach. In 2012, American Express partnered with Fourth Square to offer promotions based on data collected from social networking sites, and provided their customers better experience both online and offline. Big data technologies facilitate the capability to integrate and augment unstructured data and operational data both inside and outside the organization.

**Corporate and Retail Banking Use cases**

* One of the main helpful big data use cases for banks is information aggregation. Few years ago, a reputed US retail bank faced a challenge of aggregating information about a single customer, which involves account checking, wealth management, and mortgage status of that same individual that was stored in different management systems. In similar situations, a centralized data infrastructure can provide banks with required details to leverage the analytics, where they can aggregate different types of data and provide better services to customers.
* Reputed bank Wells Fargo has recently made a huge step by developing a multi-channel data drive, which helps in deploying an analytics solution that integrates data from online and offline channels to represent the unified view of consumer. As a result, Wells Fargo has an observation on behavioural patterns of every bank user, and also pays special attention on transactions like changes in details of account and change in payment frequencies.
* Another important use case of big data for banks is reduction of credit card fraud risk. Ability to detect early risks in the accounts can save a lot of overheads and other losses. Real time risk assessment also allows proactive analysis to identify fraudulent behaviour before the financial damage occurs.
* Recently, Bank of America (BoA) used Big Data analytics to identify commercial customers who were migrating to smaller banks. This has helped BoA to solve their abrupt customer migration problem by offering promotions to customers.

Other use cases that banks are currently in the process of implementing are Risk Analysis, Customer Progress, Optimization, Predictive analysis, Matching Algorithms, Abnormality Identification through AI, and Normalization.

**Challenges and Other Emerging Factors in Big Data**

So far, we have discussed about the advantages and privileges of big data. Technical organizations have already accepted big data as a true asset and also embraced it enthusiastically. On the other hand, big data also has its own challenges and uncertainties as it involves handling of complex data types responsibly. Few of the challenges of big data include:

**Data Quality and size**

Big data has a capability to manipulate complex and unstructured data. However, not all data sets are clean and structured. Organizations should make sure that unnecessary data is filtered out before providing inputs for data analysis. Failure to do so will lead to waste of storage cost, processing time and storage space. According to the book *‘The Signal and the Noise’* by *Nate Silver* the quantity data in the world is increasing by 2.5quintillion bytes per day, out of which only one eight of the information is considered useful. In the name of Data Analytics, organizations are gathering tons of unused data, which is noise. Data noise is on the rise which is a direct result of the big data phenomenon, so it is definitely a drawback.

**Understanding the concepts**

Businesses are still in the process of learning the practices of big data. The powerful data analytics tools need observation and expertise to make proper decisions. When implementing big data into business, companies should know their priorities towards data and the reports that are produced should be considered in the next implementation cycle. Businesses should also gain understanding about which tools and techniques to use. Finally, business users should have the training to envision the reports which is another limitation.

**User Acceptance**

When it comes to customer base, generally consumers will not show any interest in changing their routine transactional mode. When a business adopts big data and analytics to their data, they need to change some specifications which may affect the customer base and will lead to customer dissatisfaction.

**Security**

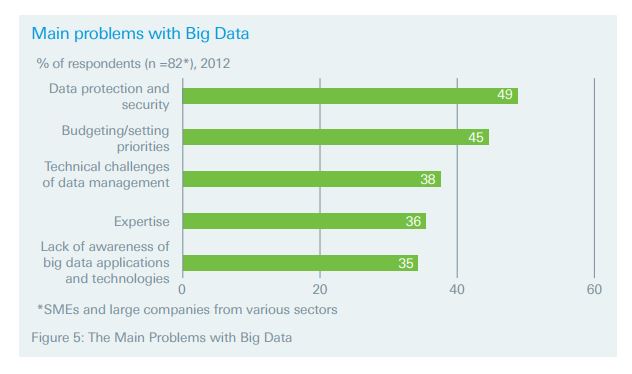
Enabling security around in-house as well as cloud-based big data solutions requires expensive expertise. The business needs to have a thorough understanding of the security pros and cons with in enterprise and community versions of big data technologies.

**Cost**

Cost varies immediately depending on the depth and size of data which can be a costly affair. There is a huge misconception in today’s technological industry that implementing big-data analytics can save money for the organization. The hidden truth is even if you have an expert team to develop an in-house solution, the effort to implement a solution within the organization is even more costly.

**Interpretation Issues**

Another challenge to face regarding big data is, sometimes the big data outputs are wrong and misleading. Many companies who are not so familiar with this technology overload chunks of data and this results in wrong predictions and misleading outputs. Some of the potential examples of big data pitfalls are wrong prediction of health care incidents. Therefore, more funds are needed to train internal employees in big data field so that these kinds of unpleasant incidents and error predictions can be avoided. Here is another brief representation of big data dependencies and problems.

Figure 4: Problems with Big Data

**Future Trends in Big Data**

In this digital era, big data has already created a technological history in terms of data processing. Technological experts claim that big data will continue to evolve and spread across all industries. Let us emphasize on how the big data hypothesis will grow in the upcoming years.

**IOT (Internet of Things)**

As people today are more internet driven and technological dependent, this kind of network helps big data to power grid every possible data source that is useful to public. The huge amounts of data streams that they collect and exchange will be available on internet. This constant whipping of data needs real-time data analytics to make useful decisions and improve the way services are built. IBM believes that IoT will grow up to 35% in the next five years. At the same time, many business analytics companies also expect that the increase in IoT based products will create big data an economic value of 4 trillion USD and 11trillion USD by 2030.

**Cognitive Computing**

Advanced Technologies like AI (Artificial Intelligence), Machine learning comes under the category of cognitive computing and they can create an easier platform for people to gain knowledge. Recent IBM invention Watson that has created a buzz by defeating two jeopardy champions and this was done because of huge big data analyses of medical data combined with AI, also known as cognitive computing. This proves us that cognitive computing is going to be in lime light soon.

**Increase in Robust visualizations**

One of those important considerations that make big data go to mainstream is the accuracy in its predictions. In the future, big data will be able to visualize more precise outcomes to organizations and public in the form of graphical representations.

**More companies are going to hire ‘Chief Data Officer’**

Growing of big data technology also results in creating more opportunities for chief data officers and big data scientists in coming years.

**Increase in Algorithm markets**

Forbes predicted that, in coming years businesses will start purchasing algorithms rather than program them and add their own data. Existing services like Algorithmia, Data Xu can be expected to multiply rapidly.

**Big Data in cloud**

In coming years, Improving Hadoop technologies will allow data and analytics to be done in the cloud easily. This would be a cheaper solution and also this is a burden free option because business don not have to manage the infrastructure physically all the time.

**Data will be everyone’s product**

Technical experts say that data will become a product with value to buy, sell or lose. Around this data new ways, business models and companies will be raised.

**Conclusion**

Finally, 63% of banking organizations believe that analysing and managing big data plays a significant role. However, this evidently shows that big data is going to be an important aspect of standardizing any business. The key point to successful usage of big data is the capability of the organization to collect and process relevant and quality data, and to introduce this data output into their business practices in right time. Instead of opting for big data based on its trend, companies should follow more pragmatic ways to understand and invest in big data. To understand and adopt big data into businesses and to avoid hidden pitfalls in big data, companies should educate themselves, explore opportunities through Research and Development practices, engage the business by emphasizing on value created, and finally execute the top revenue generating big data initiatives.

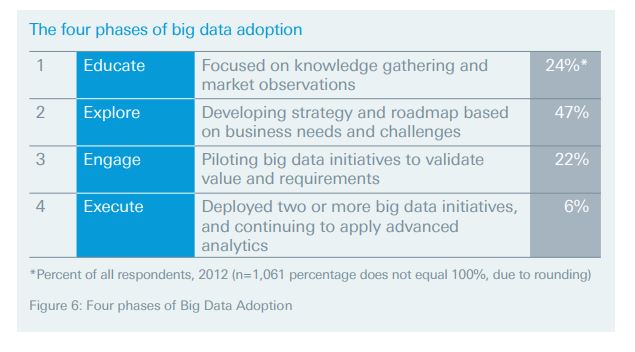


Figure 5: Four phases of big data adoption

**Bibliography**

Brown, Carol V, Daniel W. DeHayes, Jeffrey A. Hoffer, E. Wainright Martin, and William C. Perkins. 2012. Managing information technology. 7th edition. New Jersy: Pearson Prentice Hall.

Chen, Hsinchun, Roger HL Chiang, and Veda C. Storey. 2012. Business intelligence and analytics: from big data to big impact. MIS Quarterly 36, no. 5 (December): 1165-1188.

Clifton, Brian. 2012. Advanced web metrics with Google Analytics. 3rd edition. John Wiley & Sons, Inc..

Pallant, Julie. 2010. A step by step guide to data analysis using SPSS. Open University Press.

McCue, Colleen.2006. Data Mining and Predictive Analytics in Public Safety and Security. IT Professional 8,no. 4(July-August): 12-18.

Google Public Policy Blog.

<http://googlepublicpolicy.blogspot.com/2010/07/bringing-federal-it-into-cloud.html> (Accessed July 29, 2017)

IBM public blog.

https://www.ibm.com/analytics/us/en/technology/?lnk=mpr\_bua&lnk2=learn

(Accessed June 12-August 1, 2017)

Internet World Stats. 2010. “Internet usage statistics—the big picture.” Internet World Stats Web site, www.internetworldstats.com/ stats.htm (March).

“Big Data: Strategic Risks and Opportunities Looking Beyond the Technology Issues “ A White Paper by Richard Anderson and Daniel Roberts. N.p., n.d. Web. 19 Aug. 2014.

“Avoiding the Big Data traps.” Research-live.com – news, analysis and jobs for market researchers. N.p., n.d. Web. 19 Aug. 2014.