

Integration of Big Data in Business Analytics

Autor: Antulio J. Lorenzo Montes de Oca

Mentor: Dr. Nelliud Torres

Department of Computer Science

Abstract

In a world where technology constantly evolves and the amount of data being created is exponentially increasing, data processing and analysis becomes crucial on decision and risk evaluation. Companies around the world can agree that fast processing, predictive analysis, and real-time data can give strategic advantage. Business Intelligence (BI) is an approach of information gathering and analysis to give executives a better understanding based on data. BI can be applied using concepts from statistics, artificial intelligence, and data mining, among other, focusing on achieving to help companies in decision- and risk- taking. Big Data is used to describe large sets of information that traditional systems, such as databases, cannot process and maintain reliably. These techniques make the processing of high volumes of data easier, which in turn may result in quick analysis. One of the benefits of integrating Big Data and Business Intelligence is the management of both structure and unstructured data from diverse and large collections.

Introduction

In every area of interest, such as academia, business or engineering, the more data that is gathered and processed, the better the accuracy in decision-making. Big Data has been a term of discussion since the early 1990s. The idea of analyzing huge amounts of data was mesmerizing. Even though the concept was conceived nearly three decades ago, it has gained general popularity in the last decade [1]. Business Intelligence (BI) allows a corporation's executives to acquire a better understanding of its customers, the market, supply and resources, and competitors in order to make effective strategic decisions. BI technologies provide historical, current and predictive views of business operations, such as reporting, online analytical processing, business performance management, competitive intelligence, benchmarking, and predictive analytics. Business Intelligence and Analytics have taken many forms and constantly apply new techniques and methods to achieve competitive advantage. Tools designed for a strategic advantage become valuable assets for any business. Integration of Big Data with business intelligence is the next level of how companies will succeed. Advances in technology have driven us to generate immense amount of data. Finding patterns, analysis and correlation of data can result in business success. It can be deployed on the effectiveness of a drug marketing campaign, how well a pharmaceutical is performing, or even measuring tendencies on how new human resource policies affect the yield of its employees [3]. The main challenge faced is to prepare the data for predictive analysis, as raw data cannot be directly utilized for analysis.

Business Intelligence and Analytics (BI&A)

The term Business Intelligence began to emerge on the late 1800s. Richard Millar Devens wrote a book called Cyclopædia of Commercial and Business Anecdotes, where he explains how a corrupt banker uses the information gathered on events of the era to make decisions in his business. Even though he is portrayed as a corrupt banker, the concepts of gathering data and doing the analysis set the basis of how helpful and beneficial BI is. The topic of having a system to perform decision-making intelligence began to increase rapidly as technology did great advances on computing. Decision-making for a company is one of the most important roles. Making tough, assertive, and accurate decisions is a difficult task if all the elements of judgement are not considered. As technology has evolved, companies produce extensive amounts of information in different formats. These records by themselves are useless without some form of manipulation. Business Intelligence addresses this need of gathering and displaying meaningful trends [3]. BI has been able to improve the success of organizations by providing better decision-making with the use of information, which regular reporting did not provide. Business intelligence, in theory, is the development of tools to aid and guide management in day-to-day decision-making. It can give insights in an optimal way of planning and make changes oriented to business needs. What BI does is summarize and simplify data in the most accurate way in which management can view every aspect of a company in a single system. The goal of BI is to ensure the data that management views will display the most current perspective as to insight trends [3]. Data analysts use one or more of the predictive analytics modeling tools to perform various Business intelligence analysis (figure 1). These trends can be performed by many techniques on the field, such as statistical analysis, machine learning, artificial intelligence, data mining, big data, and multidimensional analysis.

Advantages of BI&A

The main advantage of BI is to have a single source capable of viewing processed data from different systems, databases, and records.

This includes the following benefits:

- Making smart business decisions
- Developing better strategies
- Increasing operational efficiency
- Identifying and implementing cost reductions
- Establishing measurable goals
- Tracking progress

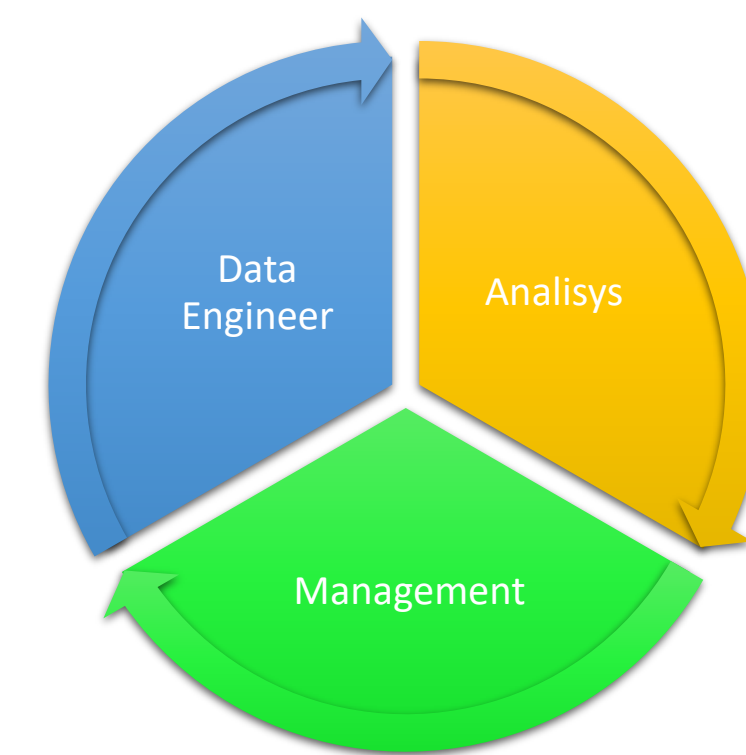


Figure 1
Iterations between BI and management

A knowledge dashboard should be capable of managing, processing, and displaying data from many sources [4]. Although big data analysis has been predominant in large companies and financial institutions, many SMEs have started using big data to transform their business operations. Enterprises in all levels develop strategies and goals to expand in the market. Making smart assertive business decisions is a critical aspect regardless of the market or how big an enterprise is. For better strategy development, an overall view all aspects of the business must be considered.

Disadvantages of BI

Even though the concept and theory of BI makes it a great asset to a company, the implementation of such technology can be rather daunting. This is due to the complex hardware and software architecture that must be managed to implement an efficient and capable system.

Many tools that aid in the decision-making process may be used in Business Intelligence. This broadens the term, and many tools do not necessarily meet a business's needs. Sometimes a business is discouraged to adopt these new systems due to:

- Security breaches
- Hardware initial cost
- Long lead times on implementation
- Difficult user adoption
- Time-consuming data cleaning

One of the most important aspects that could be a disadvantage of any system is the cost of maintenance and support during product life [4]. Although the market is flooded with big data analytic tools, all these tools are geared towards the provision of real-time, data-driven business solutions. As technology grows and evolves, new techniques and tools are available and an implemented product becomes obsolete or rather not supported. The migration or retrofit option on an architecture that has hundreds of thousands or even millions of transactions can be a difficult task

Big Data

Big Data is a new concept that allows large collections of raw data to be processed promptly [1]. BI tools can range from simple MS-Excel spreadsheets to querying and reporting software, OLAP as well as data mining tools. But it may be conclusively reported that business intelligence tools can be categorized into generalized or tools that function on normal data sets, of smaller size, in structured or semi-structured format and big data specific tools that utilize the big data, i.e., data large enough but of no explicit size and that too in any structural format be it structured, semi-structured, or unstructured.

How can a company or entity determine that big data is an approach that meets its needs? The answer is as simple as to evaluate the characteristics of the five Vs (figure 2).

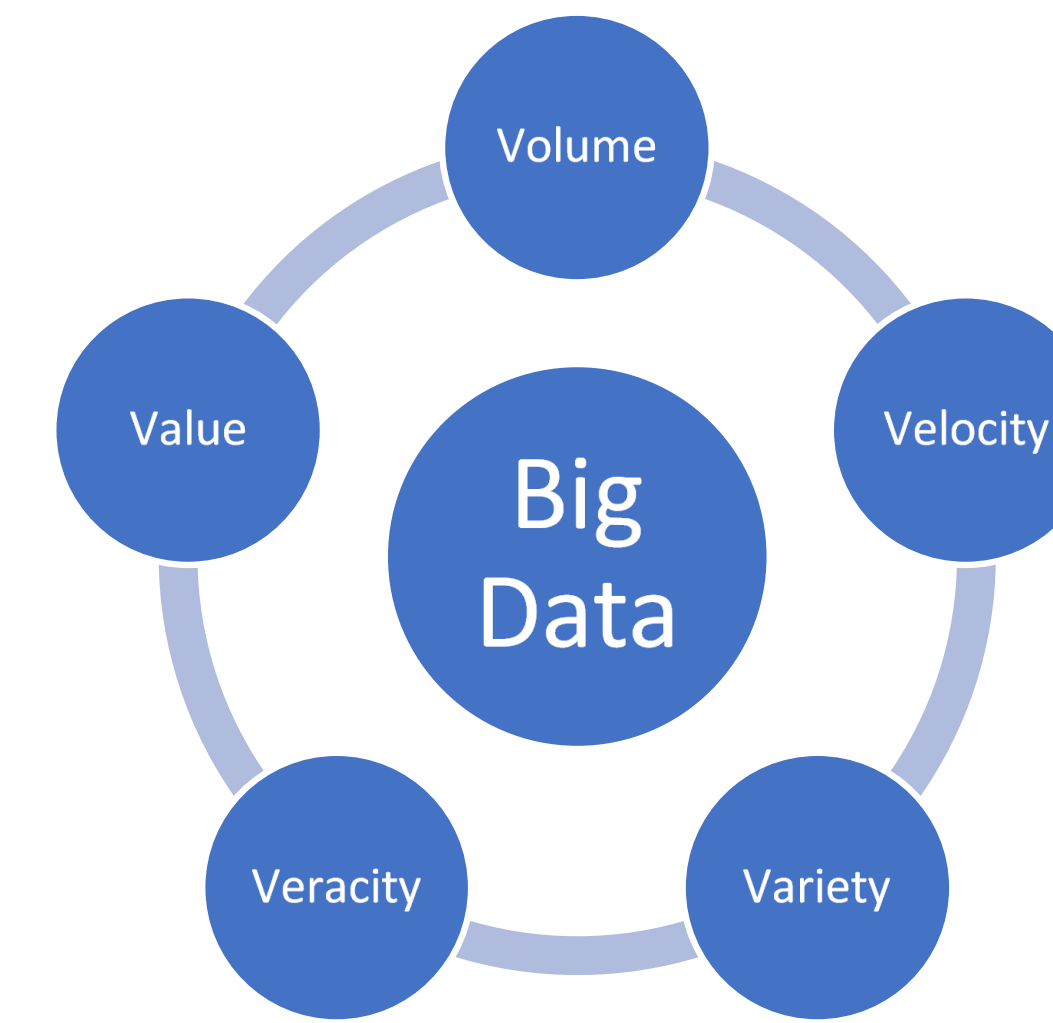


Figure 2
The five Vs of Big Data

- Volume: this term refers to the amount of data that is hard to be managed in a single system. Every system architecture is different and evaluation is needed to understand the performance of a single system.
- Velocity: it refers to data generation at high rates. As data increases exponentially, it affects the amount of data being handled.
- Variety: the most crucial and beneficial aspect of big data. Data can come from many sources and it can be structured, unstructured, or a combination of both.
- Veracity: how accurate the data is, as well as the accuracy in the processing of such data.
- Value: it depends on how useful are the system and the data provided. The value is dependent on the four preceding Vs.

One of the first steps after selecting the appropriate system to perform Big Data is to apply Extract Transform and Load (ETL). ETL is a principle to normalize data from different sources in order to have data in the same format. Multiple sources may have different conventions for the data. The data extracted to normalize into the system may have to apply conversion formulas or formats; for example, temperature data from two different sources, one measured in Celsius degrees and the other measured in Fahrenheit degrees; or it may be as simple as a yes/no answer one source renders un binary format and the other source, as a "Y/N" string.

Currently, the most adopted technology on the market is MapReduce models to distribute data and perform parallel processes. This model is a popular alternative to structure computation to ease the process in distributed machine. The first step is to Map that distribute the raw data in the machine; then the shuffle step does the allocation of the computation for the reduced step that allows to ease calculation in the most efficient method.

Advantages of Big Data

In every area of interest, such as academia, business or engineering, the more data that is gathered and processed, the better the accuracy in decision-making [1]. 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 a business more agile, and to answer questions that were previously considered beyond one's reach.

Some of the benefits of using Big Data includes:

- Real-time reporting and analysis
- Making data-driven decisions
- Establishing metric priorities based on current business health
- Redundancy of data
- System performance optimization

There are many benefits of Big Data to support businesses, including that most solutions on the market have the availability to create a dynamic dashboard that the user can manipulate at their convenience.

Big Data has the advantage to create a view of all the enterprise's aspects, to monitor its overall.

Tedious, time-consuming report generation could be done faster by implementing Big Data. Multidimensional analysis could be implemented to dig deeper into trends with minimal resource personnel making reports based on their department. The industry is opening new jobs as data scientists, data analysts, and data engineers, their role based solely on managing data. In legacy industry models, the area expert manages data reporting and management acquires and pieces different reports for decision-making.

Disadvantages of Big Data

Technology is great and can give a user advantage if appropriately adopted [5]. In spite of their pervasive presence in business, dashboards can be problematic to develop, design, and implement, and frequently fail to accomplish their objectives. The first problems that arise are security and stability.

Data integrity is a major drawback of having a system with immense amount of data. Inconsistencies in these systems can lead to poor management decisions that could negatively affect enterprise. Systems are dependent on how much of the data's integrity is compromised. If it's only a negligible amount of data, then this would mean low noise on a mathematical model, but as noise increases, accuracy decreases.

Vulnerabilities in the system can lead to security issues and harm a company. This may be performed by injecting false data to misguide the dashboards and hence lead management to make poor decisions. If such attacks are made, the recovery would be a difficult task due to the redundancy of immense amount of valid data processing paralleled with false information.

Big Data implementation needs initial expensive hardware for most companies. For a company to invest, the return of investment (ROI) must be on average less than four years. For big companies, it makes sense that the ROI will yield results in a short amount of time. For SMEs, the implementation cost could be prohibitive. Emerging solutions are cloud services based on consumer needs. The major drawback to small enterprises is to gather the knowledge necessary for the implementation. New systems can crash with the cultural change. Sometimes people get used to the tools they have always used, and even though a better system can be implemented, the business may phase a negative reaction to change.

Conclusion

The integration of Big Data helps business achieve better results and improve efficiency. BI tools, with the scope of their methodology, reveal a new way of evolution in business. SMEs that cannot react to this technology fast enough will be the most affected when, as time keeps passing, more enterprises with these powerful tools target the market and gain advantage. The world is moving into a digital era in which systems are generating massive amounts of data that can be correlated into meaningful insights of behavioral patterns. The ability to process and discover patterns can result the most beneficial to companies. Not long ago, the commercial market industry realized the power of gathering knowledge about its customer. This can be seen in loyalty program cards of stores around the country. Loyalty programs gather useful information about peak customer hours, best-selling products, and what products are bought together, among many other. With this information, sellers can offer an effectively targeted, personalized ad campaign. The integration of Big Data in Business Intelligence yields the most positive results in business health. As more items connect to generate data, clearer is the path in the need of a system that can store, manage, and process information fast and reliably. Data produced by sensors, machines, systems, cashiers, weather stations, phones, cars and many others is being stored. As processing algorithms and techniques such as Big Data become stronger, society will evolve intellectually, making discoveries. Careers will begin to merge and technology will begin to be used in areas such as psychology, politics, and social studies for better analysis.

References

- [1] P. T. Chung and S. H. Chung, "On data integration and data mining for developing business intelligence," in 2013 IEEE Long Island Systems, Applications and Technology Conference, Farmingdale, NY, 2013. [Online]. doi: 10.1109/CSNT.2013.6578235
- [2] P. Nagar, L. Atriwal, H. Mehra and S. Tayal, "Comparison of generalized and big data business intelligence tools," in 3rd International Conference on Computing for Sustainable Global Development, New Delhi, 2016, pp. 3585-3588.
- [3] P. Wazurkar, R. S. Bhadoria and D. Bajpai, "Predictive analytics in data science for business intelligence solutions," in 7th International Conference on Communication Systems and Network Technologies, Nagpur, 2017. [Online]. doi: 10.1109/CSNT.2017.70
- [4] K. SM and M. Belwal, "Performance dashboard: Cutting-edge business intelligence and data visualization," in International Conference On Smart Technologies For Smart Nation, Bangalore, 2017, pp. 1201-1207.
- [5] L. Rajabion, "Application and adoption of big data technologies in SMEs," in International Conference on Computational Science and Computational Intelligence, Las Vegas, 2018. [Online]. doi: 10.1109/CSCI46756.2018.00219