Big Data Analytics Opportunities and Challenges for the Smart Enterprise

Brahim Jabir, Falih Noureddine, and Khalid Rahmani

Abstract Today, the real challenge is examining large and varied data sets to get the value in order to uncover information including hidden patterns, unknown correlations, market trends, and customer preferences that can help organizations make informed business decisions. Big data analytics is a form of advanced analytics, which involves complex applications with elements such as predictive models, statistical algorithms, and what-if analysis powered by high-performance analytics systems, but the problem is the field still suffering from gaps and problems, and there is no a complete framework that can achieve the business objectives. In this chapter, we present a literature review, explain how big data analytics helps improving business objectives, and describe some analytics systems used in this goal.

Keywords Business analytics · Big data analytics · Smart enterprise

1 Introduction

Organizations handle and collect large volume of data and try to exploit it to achieve their business goals. So that makes data pass through several organizational, strategical, and procedural stages. Over the years, various applications and approaches [1] appeared and helped to collect and analyze the right data to make the right decisions.

Lately big data analytics comes as a new strategy of analyzing large amount of data, or big data [2]. This big data is gathered from a wide variety of sources, including social networks, videos, digital images, sensors, and sales transaction records, that might provide valuable insights. Through this insight, businesses may be able to gain an edge over their rivals and make superior business decisions. This

B. Jabir · F. Noureddine
LIMATI Laboratory, Polydisciplinary Faculty Mghila, BP 592, BeniMellal, Morocco

K. Rahmani (✉)
ERPTM Laboratory, Polydisciplinary Faculty Mghila, BP 592, BeniMellal, Morocco

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_70

chapter highlights the challenges and pressures that the healthcare systems face, identifies its opportunities, and discusses the important role of analytics and its framework.

2 Business Analytics

Business analytics is an advanced concept emerged from business intelligence. It is an approach combining different disciplines to extract value from data in order to make a valuable decision [3] via regular analysis with different plans and strategies. So, BA is used by organizations for the aim to develop existing processes, identify new opportunities, discover more product features, change and evolve new services and systems, better understand the behavior of customers, and expect problems before they happen. These disciplines are computer science, statistics, data management, decision science, and scientific research methods.

2.1 Existing Type of Analytics

To better explain this concept of analytics, it is necessary to understand it through its three types [4], which shorten its role and purposes:

2.1.1 Descriptive Analytics

Descriptive analytics is a type of analytics known as business reporting, provides an interpretation and extrapolation of historical data, and helps to understand the significant change in the enterprise [5]. Its main result is making the raw data understandable for the various components of the organization (employees, managers, investors, supplier, etc.), this type enables the company to answer the questions of "what happened" and/or "what happening" [6] like:

- How many stocks have been delivered last year?
- What is the medium sales volume for the last year?
- What is the kind of the products returned for last month?
- What are the best-selling products?
- How many customers purchased last month?
- How much paid for the general costs last year?

This analytics type uses many techniques and tools such as data mining and data aggregation to provide information, create a report of available data, and prepare it for further processing in order to provide insights and predictions, so that to help understand why and how some event happened and explain why some results occur, all while trying to improve employee engagement and productivity.

2.1.2 Predictive Analytics

Predictive analytics is a branch of analytics comes as a kind of analytic modeling and requires several statistical tools that can analyze current and historical events to provide insights and make predictions about unknown events about future [7].

Experts use this type to deploy future business planning to predict the problems before they happen, discover new services and more opportunities to reduce time, increase productivity, and minimize risks. Its principal outcome is to answer the question of "what will happen?" or/and "why will it happen?" Examples are:

- Who is the most likely employee to leave our organization?
- What is the risk of losing money on new project investment?
- What will be the revenue if sales service decreases by X percent?
- What will be the revenue in case of a boycott for an X time?
- What will happen if supplier prices grow by an X percent?
- What do we expect to pay for X services over the next year?

By answering these questions, the enterprise examines the results to detect new patterns and links to enhance their performance through its different business areas, operations, finance, and marketing.

2.1.3 Prescriptive Analytics

This analytic is the final step of the analytics process. It defines the actions to be taken to avoid future risk or to take full advantage of a promising trend. It uses also historical data and external information due to the nature of statistical algorithms to identify opportunities and identify the reasons behind failure or success. Prescriptive analytics uses sophisticated tools and technologies, like machine learning, business rules, and algorithms. It answers the question of "what I should do?" and/or "why should I do it?" [8]. Examples are:

- What is the alternative plan to maintain maximum profit if X employee leaves?
- How many products do we need to sell to maximize revenue?
- What is the best way to minimize costs and fees?

The answers obtained by this processing help the organization to set new criteria for success and failure in order to reconstruct the business with reliable predictions to develop efficiency and reduce costs.

Figure 1 shows the analysis levels of the business analytics with its three types, and it clarifies that descriptive analytics provide insights into the past. Predictive analytics helps to understand the future and prescriptive analytics to advise on possible outcomes.

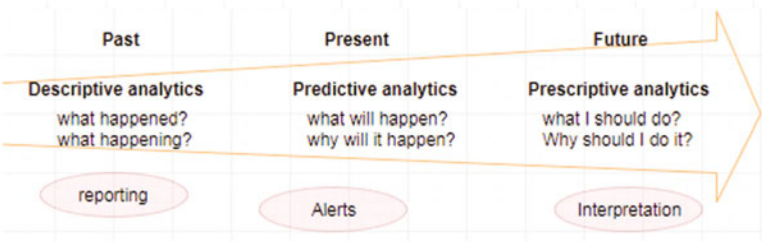

Fig. 1 Business analytics types view

3 Big Data Overview

3.1 What Is Big Data?

From all existed definitions for "big data," we can choose that big data is data that is "too big," "too hard," and "too fast" [9]. "Too big" indicates the large available quantities of data, known as "volume." "Too hard" relates to the different type and several format of data and also a variety of sources, which need some kind of analysis and some kind of tools to process, that is known as "variety." "Too fast" means that data need a high speed to process and that refers to "velocity." These three Vs are the big data dimensions and attributes (volume, velocity, variety) [10] (Fig. 2).

3.2 Defining Big Data Analytics

According to the big data research article [2], the big data analytics is the point where advanced analytics technique has been emerged and runs in the big data to process data for extracting value [11]. This analytics adoption drive potential benefits, and it is the key for the enterprises to exploit data-dependent capabilities and provide insight and direction to enhance and help making decision. Big data analytics unites different tools and disciplines using statistical algorithms, predictive models, data mining, and various other tools, in order to provide valuable information from a large amount of data (Fig. 3).

4 Smart Enterprise

Mellote affirms that smart enterprise is an organization use and is a kind of exploiting technologies to answer key challenges, and also to manage and extract meaning from the diversity and volume of data that is available to them. A second

Fig. 2 The three Vs of big data

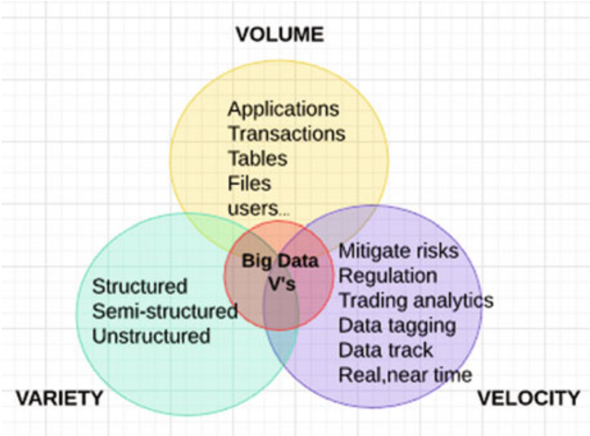

Fig. 3 Big data analytics for enterprise

group of definitions [12] emphasis that "smart enterprise" is an organization that offers an integration of all recent analytics technological advances to acquire, transfer, interpret, and analyze the information. A third definition focuses on the insight aspect and accord that smart enterprise is an organization that embeds analytics to transform information into insights and predictions and then into action. From all those definitions, we can conclude that smart enterprise is a new enterprise performance optimization strategy, enable some methods and approaches defined under a solution which is big data analytics and come to turn data into value in order to make a decision and affects all enterprise areas to enhance their competitiveness (Fig. 4).

4.1 Enterprise Big Data Sources

The available information varies significantly in its volume, its format, and the speed depending on the type of the company [13]. It is necessary that information combine data typically managed by the HR department, customer satisfaction, and operational data. This data expanded, linked, and analyzed with several tools to find what is really happening in the organization and discover what will happen and what should the organization do. Some examples of the types of data are described in Table 1.

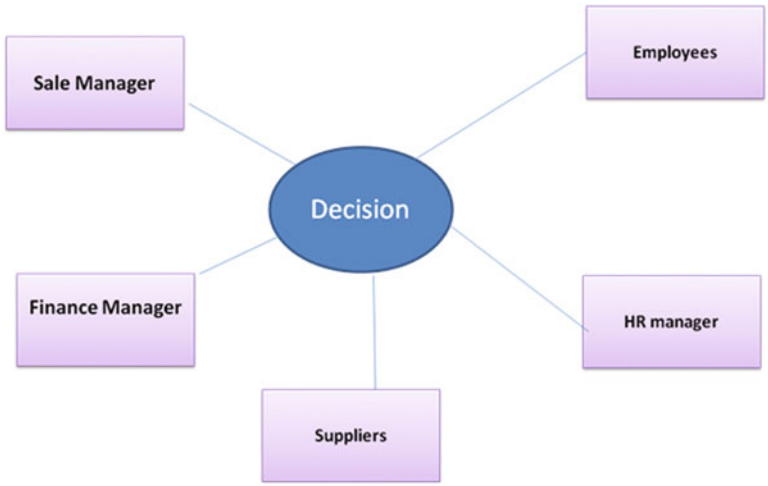

Fig. 4 Making decision in enterprise areas

5 Challenges and Opportunities of Big Data Analytics

Big data analytics has become essential as it helps in improving business, decision makings and providing the significant advantage over the competitors. This applies for organizations as well as professionals in the analytics field. For smart enterprises, who are skilled in big data analytics, there is a wide range of opportunities out there.

Decision-making is a major challenge [14] given the many changes that have occurred in all areas of the enterprise; to this end, big data analytics has been emerged as a set of strategies, tools, and methods to facilitate decision-making. It makes enterprise ensure the following benefits:

- Better targeted social influencer marketing.
- More numerous and accurate business insights.
- Segmentation of customer base.
- Recognition of sales and market opportunities.
- Automated decisions for real-time processes.
- Definitions of churn and other customer behaviors.
- Detection of fraud.
- Greater leverage and ROI for big data.
- Quantification of risks.
- Trending for market sentiments.
- Understanding of business change.
- Better planning and forecasting.
- Identification of root causes of cost.
- Understanding consumer behavior from click-streams.

Table 1 Several information sources used by enterprise

Information source	Description	Examples
HR database	Data collections contain information about employees, customers, products, etc., such as employee personal details, performance, diversity data, and promotion details	Database: Oracle, SAP, etc. Information: Age, gender, salary, department, performance rating, sickness absence, location, team, price, etc.
Employee attitude survey data	A range of information usually stored in survey programs and exported to files contains the attitude of employees and their engagement data, usually managed by providers organization	Job strain level, employee engagement, employee performance, satisfaction, perception of justice, stress level, etc.
Customer satisfaction survey data	Also stored in survey programs, provides information about customers' preferences, customers' experiences, customers' satisfaction, etc.	Customer rating Customer loyalty Preferences Satisfaction Purchases Likelihood of further business
Sales performance data	Information usually owned by the sales function, recording details of sales performance and revenues. It is a useful information that help to determine how the organization reached the business goal	Sales of month New purchases Revenue attained Best-selling Products' characteristics
Operational performance data	Information refers to the efficiency of the organization. It is about measuring the successful running of the business	Number of complaints resolved Number of calls dropped out Number of queries resolved Time consumed in some operations

- Manufacturing yield improvements.
- Discover new facts about their customers, markets, partners, costs, and operations.

Data generated through big data analytics sources can help companies better understand their performance than previous technologies [15], but there is no doubt that big data presents technical challenges due to its volume, variety, and velocity. Data volume alone is a showstopper for some organizations, and most of them face the very real risk of information overload generated by the different systems, the complexity of data, the lack of experts in this field, and the costs of systems, and there is not yet a complete framework that can offer a solution for this major problems and help enterprise for a sample transformation to be a smart enterprise that adopt analytics to get value from data in order to take the right decision (Fig. 5). So, finding a way to exploit the data at their disposal and leverage

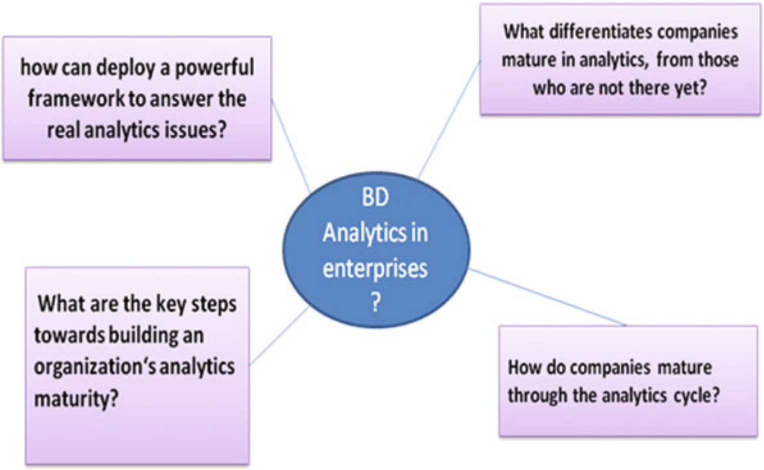

Fig. 5 Big data analytics issues in enterprise

them to improve business and organizational performance becomes a necessity [16], this can only be achieved when analytical tools and techniques are combined and integrated in a structured and rigorous framework, and this combination can identify new opportunities for improvement or suggest innovative ways to address old challenges...

5.1 Big Data Analytics Technique Solutions

In this section, we describe the current analytics solutions that can be a major key to create and deploy an efficient predictive management model. This analytic management model will produce predictions and insights and help HR managers make better decisions for the company. Of course, most of them are expensive but there is a lot of similarity between the packages, and the analysis generally produces the same results. So, all you need is analytic methods and skills, and it is possible to apply them in most other systems. We consolidate in Table 2 the most popular analytics system with some details:

5.2 Components of Big Data Analytics

The levels and layers of abstraction presented in the Fig. 6 show components of big data analytics, forms of stacks, and their integration with each other; in most of the cases, HDFS/Hadoop forms the core of most of the big data-centric applications, but that is not a generalized rule:

Table 2 Popular big data analytics systems with details

Analysis software system	Details
Apache zookeeper	Apache zookeeper [17] is a framework that can federate the communications between distributed systems. It functions by providing a memory space shared by all the instances of servers. These several machines connected to each other solve problems together and process large quantity of data together that accelerates the processing, offers real-time access, and handles system breakdown problems
Apache storm	Apache storm is an open-source project started with the idea of developing a stream processing system came in December 2010, by Nathan Marz and adopted by apache on 17 September 2014 [18], this system offer a real time distributed processing systems that can process the unbounded streams of data very fast than ever [19], it is easy to use and provide low latency with guaranteed data processing, it support communication over a JSON-based protocol. It offers services of filtering, aggregation, join, read/write to and from a several sources
KAFKA	The Kafka project used to build real time data pipelines and streaming applications [20], it is a distributed, partitioned and replicated service, it support parallel data loading provide by a various producers (Frontend web applications, services, adapters...), consumed by real-time consumers (filter and sift information in databases and trigger alert), near real time consumers (save data in any NoSQL system) and offline consumers (storing information in traditional data warehouse for offline analysis). This following diagram shows a typical big data analysis and aggregation scenario supported by the Apache Kafka system
SPARK	Apache spark is a powerful open source framework for smart data processing, developed by AmPlab, in 2009, adopted by apache in 2010 [21], makes sophisticated analysis and designed for speed and ease, it has several advantages over other technologies, it is an efficient solution for smart data analytics offers text and graphics visualization, and supports streaming requests
HADOOP	Apache Hadoop is an open source designed for the distributed processing of large amount data sets across clusters of computers using simple programming models [22], created to scale up from single servers to lots of machines, each offering local storage and computation. Rather than it can deliver high-availability, detect and handle failures at the application layer, it is composed from several components, that work together to process data: HDFS (the distributed file system layer that coordinates storage and replication across the cluster nodes) YARN (the cluster coordinating component of the Hadoop stack), MapReduce (the native batch processing engine for Hadoop)
CASSANDRA	Apache Cassandra is a powerful distributed database of the NoSql family designed to collect a large amount of data from multiple sources [23], data stored is automatically replicated to multiple physical instance which is nodes offering none downtime and offers also a high availability and scalability, architecture of this database constitutes from nodes uses peer-to-peer connection, also from clusters, data centers and a partitions

Original data	Storage systems	Task trackers	Higher level language	Security and management	Modeling	Loading analytic databases	Analytics applications	presentation
Files	Hadoop				ETL modeling tools	e.g., greenplum netezza	Merced,	Reports
external data bases	file system,e.g.,HDFS	MapReduce engine	Hive	Cascading,			ClickFox	dashbords
Surveys	NoSqlDB,e.g.,Hbase Cassandra		Pig	Kerberos				Alerts
business apps								
media								

Fig. 6 Big data analytics components

6 Proposed Big Data Analytics Framework

In socioeconomic world, the decision-making challenge turn around type and quality of data, people, business planning, business objectives, etc., and not only around technological side, that is why it is necessary to think about a comprehensible multidimensional framework that project on all the dimensions that can extract value from large amount of data and help to turn this value into action in order to take the right decision. We proposed a simple framework based on the existed ones that are relatively limited in scope, this framework presents a complete model that combines several dimensions to interact with each other (Fig. 7):

- Actors: Include software developers, personnel, customers, managers.
- Interaction models: Interface between user and system aspects.
- Computing infrastructure: Devices and the software required.
- Communication: Business workflow as a collaboration that requires a significant two-way communication.
- Enterprise content: Information on attitude of employees, personnel details and their performance, promotions details, customers information.
- Data sources: The source systems such data warehouses, supply chain systems, and other operational systems, surveys.
- Internal organizational policies, procedures, and culture: Purchase of hardware and software, data backups, etc.
- External forces: External rules, regulations, and pressures that place constraints or help on the deployment.
- Regular basis measuring and monitoring of the effects of information technology.

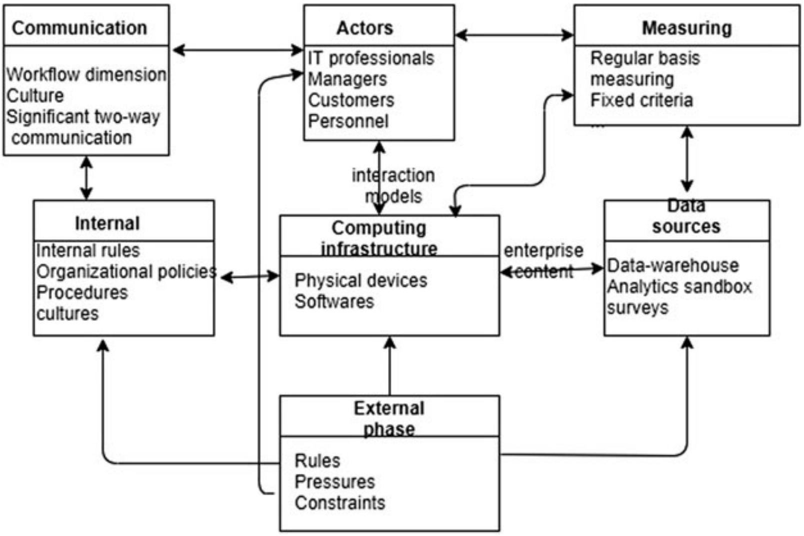

Fig. 7 Proposed analytics framework

6.1 Discussion

6.1.1 Benefits

The model offers the possibility of its hierarchical decomposition; in other words, the possibility of breaking down a complex process, system, or device into its components, this offers the possibility to study them and then integrate the results trying to understand the functioning of complete system. This model can offer the capability to feed management reports and dashboards with deep insight into past, current, and even future performance.

6.1.2 Barriers

There is always a gap between analytics development and the use of analytics within enterprises so the challenge is not to find a complete framework [24], but the analytics deployment is confronted by the following barriers:

- Lack of experienced people that can understand the analytical systems.
- Distrust of the information and gaps to extract correct data.
- Models are expensive and complex to deploy.
- Turning information and insights into decision requires an immense experience.

7 Conclusion

In this chapter, we described big data analytics approach for the company as a set of analysis of a large amount of data that comes to drive business planning and deploy the future business planning as well as a proposed framework that support big data analytics aspect.

Big data analytics generates potential benefits for the company. It is the major key behind the reaching of business goals. For this, our future contribution will be about concretizing this notion of big data analytics by a specific and original approach about scenario modeling, even predicting employee performance.

References

1. Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences, 1*(1), 83–98.
2. Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*(2), 137–144.
3. Kohavi, R., et al. (2002). Emerging trends in business analytics. *Communications of the ACM, 45*(8), 45–48.
4. Van Barneveld, A., et al. (2012). Analytics in higher education: Establishing a common language. *EDUCAUSE Learning Initiative, 1*(1), 1–11.
5. Banerjee, A., et al. (2013). Data analytics: Hyped up aspirations or true potential? *Vikalpa, 38*(4), 1–12.
6. Delen, D., & Demirkan, H. (2013). Data, information and analytics as services. *Decision Support Systems, 55*(1), 359–363.
7. Eckerson, W. W. (2007). Predictive analytics. Extending the value of your data warehousing investment. *TDWI Best Practices Report, 1*, 1–36.
8. Ransbotham, S., et al. (2015). Minding the analytics gap. *MIT Sloan Management Review, 56*(3), 63.
9. Madden, S. (2012). From databases to big data. *IEEE Internet Computing, 3*, 4–6.
10. Jothimani, D., et al. (2014). *A big data analytical framework for portfolio optimization*. Cornell University.
11. Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Harvard Business Press.
12. Springett, J. (2007). U.S. patent application no. 11/613,847.
13. Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. In *Collaboration Technologies and Systems (CTS), 2013 International Conference on* (p. 4).
14. Harrison, E. F. (1995). *The managerial decision-making process* (Vol. 4, pp. 1–39). Houghton Mifflin.
15. Marcu, O. C., et al. (2016). Spark versus flink: Understanding performance in big data analytics frameworks. In *CLUSTER computing (CLUSTER), 2016 IEEE international conference on* (pp. 433–442). IEEE.
16. Grigoroudis, E., et al. (2012). Strategic performance measurement in a healthcare organisation: A multiple criteria approach based on balanced scorecard. *Omega, 40*(1), 104–119.
17. Haloi, S. (2015). *Apache ZooKeeper essentials*. Packt Publishing.
18. Iqbal, M. H., & Soomro, T. R. (2015). Big data analysis: Apache storm perspective. *International Journal of Computer Trends and Technology, 19*(1), 9–14.

19. van der Veen, J. S., et al. (2015). Dynamically scaling apache storm for the analysis of streaming data. In *Big data computing service and applications (BigDataService), 2015 IEEE first international conference on* (pp. 154–161). IEEE.
20. Ranjan, R. (2014). Streaming big data processing in datacenter clouds. *IEEE Cloud Computing, 1*(1), 78–83.
21. Xin, R., et al. (2014). *Graysort on apache spark by databricks*. GraySortCompetition.
22. Zikopoulos, P., & Eaton, C. (2011). *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media.
23. Chebotko, A., et al. (2015). A big data modeling methodology for apache Cassandra. In *Big data (BigData congress), 2015 IEEE International Congress on* (pp. 238–245). IEEE.
24. Williams, I., & Bryan, S. (2007). Understanding the limited impact of economic evaluation in health care resource allocation: A conceptual framework. *Health Policy, 80*(1), 135–143.