# Study of Basic Concepts in the Context of Process Management and Strategies Based on Big Data

Bogdan Danov, Daniela Gotseva
Department of Computer Systems, Faculty of Computer Systems and Technologies
Technical University of Sofia
8 Kliment Ohridski blvd., 1000 Sofia, Bulgaria
b.danov@tu-sofia.bg, dgoceva@tu-sofia.bg

2024 12th International Scientific Conference on Computer Science (COMSCI) | 979-8-3503-9216-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/COMSCI63166.2024.10778493

*Abstract* – **Big Data is both an advantage and a challenge due to the explosive increase in the volume of large data sets. Big Data analysis derives knowledge to improve business processes and can be critical in decision making. However, for this purpose, the right strategies need to be implemented. This document provides an overview of the key concepts, issues and challenges in integrating big data with business processes.**

*Keywords* – **Big Data; Massive Datasets; Big Data Challenges; Business Processes; Process Model.**

## I. Introduction

The huge arrays of data, the diverse and complex sets of digital information, the volume of which is constantly increasing and exerting an ever-increasing influence in the modern world, are usually referred to by the term Big Data.

Usually, the arrays of information defined by this term are characterized by difficulties in storing, analyzing and visualizing for further processes or results [1], among their other characteristics are high dimensionality and large sample size [2].

The authorship of this term is attributed to two of the researchers at NASA's Ames Research Center, that's according to John Sprague, deputy chief information officer for technology, data and innovation at NASA headquarters [3]. In his words he said, they were describing a data set too large for a desktop computer to handle.

Another person could also be named as the author of the term - Doug Laney, from Gartner, who defined the three main characteristics of big data - the so-called. three V's.

Some authors call the emergence of big data a revolution, because due to the limited capacity of traditional technologies, companies cannot store all their archives for long periods of time, nor can they effectively manage huge data sets [4].

It is important to emphasize that Big Data is experiencing remarkable growth thanks to the rapid development of mobile data, cloud services, machine learning and the IoT industry.

For example, thanks to the constant accumulation and generation of information, the volume of data worldwide is expected to reach over 180 zettabytes by 2025. [5].

The volume and variety of available data is increasing at a rapid pace. The reason for this is the massive collection of information from various devices, including:

- mobile devices such as smartphones, tablets, etc.;
- Internet of things – cheap and numerous devices that collect and transmit data;
- generated software log files;
- fixed and mobile cameras;
- RFID readers;
- wireless sensor networks collecting data from various sensors.

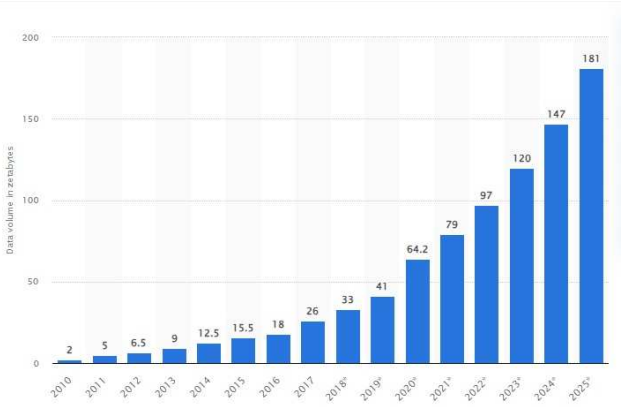


Fig. 1. Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025. Taken from [5]

All this leads to huge volumes of data that can be used for various purposes, such as:

- data analysis – such as discovering patterns and trends in data;
- • machine learning, involving the development of models that can predict future events;
- artificial intelligence – systems that could think and act independently.

## II. Characteristics of Big Data

The three key characteristics of Big Data – Volume, Variety and Velocity, also known as the three Vs of Big Data, as already mentioned were first introduced by Gartner in 2001.

Some authors [7,8] expand these three concepts to five (or more), adding to the definition two more characteristics - value and credibility of big data; The V's of big data:

- Volume - The main characteristic of big data is its enormous size. This refers to the massive amount of information generated continuously by social media, IoT devices, transaction records and more.
- Velocity - Nowadays, information is often created in or near real time. This also requires that the data be processed, accessed and analyzed at the same speed for it to be useful.

979-8-3503-9216-6/24/$31.00 ©2024 IEEE

- Variety - Big data is varied in nature. They come from multiple sources and can be in different formats:
  - Structured - organized in tables or databases, with a fixed structure;
  - Unstructured – unorganized, without a fixed structure;
  - Semi-structured – partly organized, with some elements of structure.

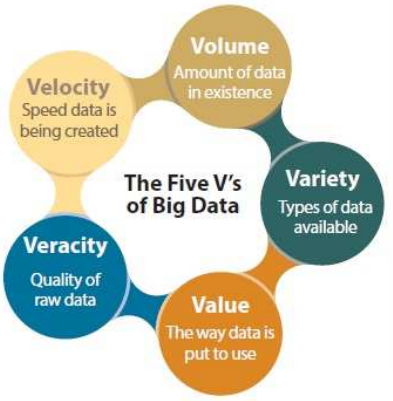


Fig. 2. The five "V's" of big data. Taken from [6]

- Veracity - Big data is not always good quality. They can be difficult to work with and analyze, contain errors and inaccuracies, and lack important data. High data reliability is essential to the success of many companies and organizations.
- Value - Big data is only valuable if it is useful for business. The goal is to turn vast amounts of data into meaningful information that can be used for strategic decisions, uncover new market trends, customer needs and potential sources of revenue. To generate new ideas for products, services and business models.

As some authors point out [9], big data analysis would definitely lead to valuable knowledge for many organizations and help to continuously improve business processes, which is a challenging task [10].

## III. Business Process Management

Looking at the BPM (Business Process Management) lifecycle, it is probably one of the best options to delve into its terminology and characteristics. Multiple approaches and methods to deal with different aspects of business processes are proposed by different researchers.

It is important to point out that Business Process Management (BPM) is not a one-time event, but is in fact a process in itself – a continuous one at that. Its scope includes design, analysis, monitoring and optimization of work processes in organizations.

It is divided into six phases [11] (Fig 3.):

**Phase 1 – process identification:**

- defining the scope: the aim is to clarify which business processes BPM covers;
- scoping: a precise definition of the specific process is needed;
- determining KPIs and Targets: identify key performance indicators (KPIs) and targets that will be measured to evaluate the effectiveness of the process.

**Phase 2 – discovery of the processes:**

- creating an "as-is" model: through a suitable modeling language (BPMN, UML, etc.);
- gathering information: through observation and analysis of documents;
- definition of roles and responsibilities: it is necessary to accurately distinguish the roles and competences of the participants in the process.

**Phase 3 – process analysis:**

- problem detection: for this purpose, qualitative and quantitative analysis are applied to detect problems;
- cause analysis: clarifying the specific causes of problems;
- • impact assessment: performed by evaluating the potential benefits of process optimization.

**Phase 4 – process redesign:**

- developing a model: it should represent the new state of the process;
- selection of redesign methods: the goal is to select those that are most appropriate;
- setting new KPIs and targets: Setting new KPIs and targets that reflect the desired improvement in performance.

**Phase 5 – implementation of the processes:**

- developing a plan: it should include all the points necessary to implement the new process;
- development of IT solutions: by implementing or designing IT systems that support the implementation of the new process;
- testing and validation: performing tests to establish the correct functioning of the new process.

**Phase 6 - monitoring and control of processes:**

- data collection: how the post-implementation process is performed;
- data analysis: to identify deviations and problems in the planned execution;
- • taking corrective action: making the necessary changes to the process to ensure that it continues to meet the objectives and deliver the desired results.

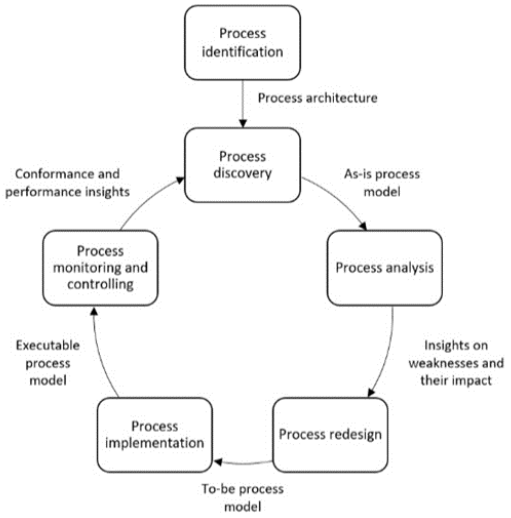


Fig. 3. The Business Process Management Lifecycle. Taken from [11]

It needs to be pointed out that the six phases of the BPM Lifecycle Model are not a linear process. It may be necessary to backtrack between phases until the business process is optimized.

## IV. OPTIMIZATION OF BUSINESS PROCESSES

The optimization of business processes aims to take new measures and decisions based on the obtained results in order to achieve real improvements. For this purpose, the processes undergo modeling and in-depth analysis. However, these actions should not be an end in themselves, as applied alone will not lead to a real result, for example, process modeling would be ineffective if it is not followed by an analysis to evaluate the results. The methods for performing modeling and analysis of business processes that some authors [12] propose are:

- modeling the process by visualizing the steps in it.
- identifying problems by finding inefficient processes;
- developing improvements by using process optimization analysis.

## V. PERFORMANCE MEASUREMENT

After the optimization of business processes, it is extremely important to measure the performance. Performance measurement is the process of quantifying the results of actions [13].

It should be pointed out here that for this purpose it is necessary to correctly select the right performance indicators in order to obtain maximum benefits. For this purpose a Performance Measurement System (PMS) can be used. PMS is a system that integrates activities at different levels. It consists of a set of performance indicators (PI) that indicate the degree of success with which the objectives have been achieved. Performance indicators can be: individual that determine the effectiveness of individual tasks or processes or be part of the PMS. Among the most important dimensions of performance indicators are quality, time, cost and flexibility.

## VI. BUSINESS PROCESSES AND BIG DATA ANALYSIS

The analysis of business processes (BPM), as already mentioned, aims to lead to their optimization. However, implementing Big Data into this analysis can make it even more effective [14].

Through the use of Big Data analysis, knowledge can be extracted to improve processes, leading to specific solutions and providing a competitive advantage.

It should be pointed out here that because they store structured data, relational database management systems (RDBMS) are unable to cope with the volume, speed and variety of Big Data. This is because such data is usually semi-structured or unstructured.

In addition, RDBMS are known to be expensive to scale, which makes them unsuitable for large volumes of data. On the other hand - NoSQL databases are more flexible and scalable than RDBMS, making them more suitable for Big Data.

## VII. BENEFITS OF INTEGRATION OF BIG DATA INTO PROCESS MANAGEMENT

Big Data solutions can offer a number of advantages in integration is process management. These advantages can be categorized into three main groups [15]: increased efficiency, better business intelligence and expanded capabilities.

Increased efficiency includes optimizing business processes through procurement, product development, manufacturing, distribution, marketing, pricing, merchandising, sales, human resource management. This in turn leads to more optimized and successful decision making that is based on data and analytics.

Better business information is characterized by transparency by providing access to data for stakeholders. What's more, this data is detailed as a result of being received in real time. Their analysis and segmentation could serve to personalize the actions to be taken, as well as to automate processes for better efficiency. This would also lead to the development of new models – products, services and business models.

Advanced capabilities can come both from recommendation systems through which offers can be personalized and from sentiment analysis – e.g. from social media. Similar capabilities also include risk modeling for financial operations and fraud detection. Several types of analysis can also be used – for churn, social graph, customer experience. Among the extended advantages of the implementation of Big Data solutions in process management can also be mentioned network monitoring to detect problems or the possibility of applying them in new research and development.

## VIII. STRATEGIES BASED ON BIG DATA

After integrating Big Data into the process, organizations must determine which strategy is best suited to their needs. The choice of strategy depends on the needs of the organization and the characteristics of the data [16].

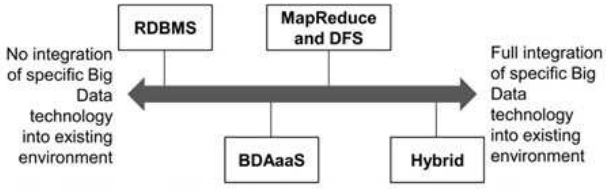


Fig. 4. Degree of infrastructure integration of the four presented Big Data strategies. Taken from [16]

One possible strategy would be to continue using traditional relational database management systems (RDBMS). An advantage of such a choice is that it is based on well-known technologies and no additional integration will be required. Such an approach would not be suitable for processing the huge amount of data typical of Big Data.

Another possible strategy would be to use Big Data Analytics as a Service (BDAaaS). Such a step means the use of cloud-based Big Data analysis tools, which may require

the use of additional software to connect the company's data to the cloud service.

Another strategy is to use MapReduce and Distributed File Systems (DFS), which however requires more significant revision and integration of entirely new software alongside existing applications.

It is also possible to take a hybrid approach, combining elements of the strategies listed so far. Such a measure would include, on the one hand, the integration of Big Data technologies, but with the use of existing technological solutions.

Of course, these four strategies for Big Data analysis with varying degrees of integration with an existing IT system are not the only options, and a number of variations and combinations are possible.

## IX. Conclusion

Choosing the right big data strategy can be critical. This problem will become more significant in view of the ever-increasing volume of information, which makes efficient data processing a challenge. This paper provides an overview of key concepts, issues and challenges in the context of big data-based process management and strategies.

## Acknowledgment

The work presented in the paper is supported by research grant 242ПД0034-09 financed by the Research and Development Sector of TU-Sofia.

## References

[1] S. Sagiroglu, D. Sinanc, “Big data: A review”, International Conference on Collaboration Technologies and Systems (CTS), San Diego, CA, USA, 2013, pp. 42-47

[2] J. Fan, F. Han, H. Liu, “Challenges of Big Data analysis”, National Science Review, vol. 1, 2, June 2014, pp. 293–314

[3] Small Steps, Giant Leaps: Episode 5, Big Data, available online at: https://www.nasa.gov/podcasts/small-steps-giant-leaps/small-steps-giant-leaps-episode-5-big-data/

[4] A. Oussous, F. Benjelloun, A. A. Lahcen, S. Belfkih, "Big Data technologies: A survey", Journal of King Saud University - Computer and Information Sciences, vol. 30, 4, 2018, pp. 431-448 [4] What is Big Data Analytics?, available online at: https://www.ibm.com/topics/big-data-analytics

[5] Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025 , available online at: https://www.statista.com/statistics/871513/worldwide-data-created

[6] Enhancing Active Transportation and Demand Management (ATDM) with Advanced and Emerging Technologies and Data Sources - Chapter 5. Design and Deployment Elements and Methods - FHWA Office of Operations, available online at:https://ops.fhwa.dot.gov/publications/fhwahop19010/ch5.htm

[7] Big Data Defined: Examples and Benefits, available online at: https://cloud.google.com/learn/what-is-big-data#section-3

[8] What is Big Data Analytics?, available online at: https://www.ibm.com/topics/big-data-analytics

[9] T. -M. Choi, H. K. Chan and X. Yue, "Recent Development in Big Data Analytics for Business Operations and Risk Management," in IEEE Transactions on Cybernetics, vol. 47, no. 1, pp. 81-92, Jan. 2017

[10] A. Vera-Baquero, R. Colomo-Palacios and O. Molloy, "Business Process Analytics Using a Big Data Approach," in IT Professional, vol. 15, no. 6, pp. 29-35, Nov.-Dec. 2013

[11] B.Wurm, T.Grisold, J.Mendling, J.vom Brocke, "Business Process Management and Routine Dynamics", In Handbook of Routine Dynamics, Cambridge University Press, 2020

[12] K. Vergidis, A. Tiwari, B. Majeed, "Business Process Analysis and Optimization: Beyond Reengineering," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 38, no. 1, pp. 69-82, Jan. 2008, doi: 10.1109/TSMCC.2007.905812

[13] L. M. Glavan, "Understanding Process Performance Measurement Systems", Business Systems Research, 2011, 2, 25-38, 10.2478/v10305-012-0014-0

[14] A. Hassani, S. A. Gahnouchi, "A framework for Business Process Data Management based on Big Data Approach", Procedia Computer Science, Volume 121, 2017, Pages 740-747, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2017.11.096

[15] A. C. L. Ziora, "The Role of Big Data Solutions in the Management of Organizations. Review of Selected Practical Examples", Procedia Computer Science, Volume 65, 2015, Pages 1006-1012, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2015.09.059.

[16] K. Ebner, T. Bühnen, N. Urbach, "Think Big with Big Data: Identifying Suitable Big Data Strategies in Corporate Environments," 2014, 47th Hawaii International Conference on System Sciences, Waikoloa, HI, USA, 2014, pp. 3748-3757, doi: 10.1109/HICSS.2014.466.