International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:10, No:10, 2016

Big Data: Concepts, Technologies and Applications in the Public Sector

A. Alexandru, C. A. Alexandru, D. Coardos, E. Tudora

***Abstract*—**Big Data (BD) is associated with a new generation of technologies and architectures which can harness the value of extremely large volumes of very varied data through real time processing and analysis. It involves changes in (1) data types, (2) accumulation speed, and (3) data volume. This paper presents the main concepts related to the BD paradigm, and introduces architectures and technologies for BD and BD sets. The integration of BD with the Hadoop Framework is also underlined. BD has attracted a lot of attention in the public sector due to the newly emerging technologies that allow the availability of network access. The volume of different types of data has exponentially increased. Some applications of BD in the public sector in Romania are briefly presented.

***Keywords*—**Big data, big data Analytics, Hadoop framework, cloud computing.

I. Introduction

In recent decades, the increasing importance of data to organisations has led to rapid changes in data collection and management. Traditional information management and data analysis methods ("analytics") are mainly intended to support internal decision processes. They operate with structured data types, existing mainly within the organization. Throughout the history of IT, each generation of organizational data processing and analysis methods acquired a new name.

With the launch of Web 2.0, a large amount of valuable business data started being generated beyond the organization by consumers and, generally, by web users. This data can be structured or unstructured, and can come from multiple sources such as social networks, products viewed in virtual stores, information read by sensors, GPS signals from mobile devices, IP addresses, cookies, bar codes, etc.

Some types of data, such as text and voice, have existed for a long time, but their volume is now exacerbated by the Internet and by other digital structures. This brings about a new era in existing technologies for data analysis. It is argued that the explosion in data volume is largely attributed to unstructured data, which partly comes from new sources through the passive behavior of the user (e.g. the case of online search terms or the user's location detected by mobile phone apps).

A. Alexandru is with National Institute for Research and Development in Informatics, Bucharest, 8-10 Averescu Ave, 011455, Romania (phone: +40-744-777-558; fax: +40-21=316-10-30; e-mail: adriana@ici.ro).

C.A. Alexandru is with the School of Informatics of the University of Edinburgh, 10 Crichton Street, Edinburgh EH8 9AB, United Kingdom (e-mail: calexan4@staffmail.ed.ac.uk).

D. Coardos is with National Institute for Research and Development in Informatics, 8-10 Averescu Ave, Bucharest, 011455, Romania (e-mail: coardos@ici.ro).

E. Tudora is with National Institute for Research and Development in Informatics, 8-10 Averescu Ave, Bucharest, 011455, Romania (e-mail: gilda@ici.ro).

II. The Concept of BD

A. BD Definition

In 2011, a report of the International Data Corporation has defined BD as "a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis" [1]. This definition highlights the most critical BD problem: that of uncovering value from data sets of huge dimensions, given the wide variety of the data types and the rapid generation of the data.

B. BD Characteristics

The characteristics of BD can be synthesized by 7 Vs [2]:

1. *Volume*: the increase in data volume in enterprise-type systems is caused by the amount of transactions and other traditional data types, as well as by new data types. Too much data becomes a storage problem, but also has a great impact on the complexity of data analysis;
2. *Velocity*: refers to both the speed with which data is produced and that with which it must be processed to meet demand. This involves data flows, the creation of structured records, as well as availability for access and delivery. The speed of data generation, processing and analysis is continuously increasing due to real-time generation processes, requests resulting from combining data flows with business processes, and decision-making processes. The velocity of the data processing must be high, while the processing capacity depends on the type of processing of the data flows;
3. *Variety*: converting large volumes of transactional information into decisions has always been a challenge for IT leaders, although in the past the types of generated or processed data were less diverse, simpler and usually structured. Currently, more information coming from new channels and emerging technologies - mainly from social media, the Internet of Things, mobile sources and online advertising - is available for analysis and generates semi-structured or unstructured data. This includes tabular data (databases), hierarchical data, documents, XML, emails, blogs, instant messaging, click streams, log files, data metering, images, audio, video, information about share rates (stock ticker), financial transactions etc.;
4. *Veracity*: refers to how reliable or questionable the data is.

The quality of BD is less controllable because it comes from different sources that cannot guarantee the quality of the content and the form of presentation. Experienced data analysts must assess the compliance, accuracy and truthfulness of the analysed data. This must be judged against the responsibilities of the initial data generator, the purpose of the data and the reactions of the receivers;

5. *Variability*: refers to how changing the meaning of the data is. This is found especially with natural language processing. Companies have to develop sophisticated programmes which can understand the context and decode the precise meaning of words;
6. *Visualization*: refers to how readable and accessible the data presentation is. Many spatial and temporal parameters and relationships between them have to be used in order to obtain something which is easily comprehensible and actionable;
7. *Value*: refers to the capacity of the data to bring new insights for creating knowledge.

C. BD Sets

BD systems can provide information both to public organizations and citizens. Some of the sources for BD are: paper documents, digital documents, points of access to the government Internet network, websites, social media and available operational systems.

The information provided by BD systems does not include information which is personal or restricted by control or confidentiality mechanisms.

According to the Gartner American consultancy firm [3], BD can be classified into five categories:

1. *Operational data*: information about customers, suppliers, partners and employees available based on transactions or from databases;
2. *Dark Data*: information gathered over time in archives, which cannot be clearly structured. This data includes emails, contracts, multimedia information;
3. *Commercial data*: information obtained from data aggregators (which read RSSs);
4. *Public data*: data belonging to public institutions (the Government, the ministries);
5. *Data from social media*: shows the activity of a user on a blog or on social networks. It is useful for determining trends, attitudes and preferences.

Once analysed, BD can help identify business trends, prevent disease and even combat crime.

To anticipate BD opportunities, companies from all areas of the industry collect and temporarily store a huge amount of operational, public, commercial or social data. For most industrial environments, and especially for the government, manufacturing and education sectors, combining these sources with "dark data" such as emails, multimedia and other enterprise content frequently represents an opportunity for transformation.

D. BD Analytics

Big Data Analytics (BDA) is a new approach in information management which provides a set of capabilities for revealing additional value from BD. It is defined as "the process of examining large amounts of data, from a variety of data sources and in different formats, to deliver insights that can enable decisions in real or near real time" [4]. BDA can be used to identify patterns, correlations and anomalies [4], [5].

BDA is a different concept from those of Data Warehouse (DW) or Business Intelligence (BI) systems.

Gartner defines a DW as "a storage architecture designed to hold data extracted from transaction systems, operational data stores and external sources. The warehouse then combines that data in an aggregate, summary form suitable for enterprise-wide data analysis and reporting for predefined business needs" [6].

BI is defined as "a set of methodologies, processes, architectures, and technologies that transform raw data into meaningful and useful information used to enable more effective strategic, tactical, and operational insights and decision-making" [7].

III. BD Architectures and Technologies

A. Architectures for BD Systems

The complexity of BD systems required the development of a specialized architecture. Nowadays, the most commonly used BD architecture is Hadoop. It has redefined data management because it processes large amounts of data, timely and at a low cost.

1. The Hadoop Framework

Traditional SQL database management systems are no longer suited to manage such large and complex data sets as in BD. When working with large volumes of data we need a solution that allows low cost storage, while also ensuring a good processing performance. One possible solution is the *Apache Hadoop* software framework.

Hadoop [8] is an open source project developed by Apache which can be used for the distributed processing of large data sets. It runs on multiple clusters using simple programming models. The design of the Hadoop framework ensured its scalability even when tasks are run on thousands of computers, each with its own processing and storage capability.

Since 2010, Hadoop has been widely adopted by organizations for the storage of large volumes of data and as a platform for data analysis. Hadoop is currently used by many companies for which the volume of data generated daily exceeds the storage and processing capacity of conventional systems. Adobe, AOL, Amazon.com, eBay, Facebook, Google, LinkedIn, Twitter, Yahoo are some of the companies using Hadoop.

Additional software packages can be installed on top of or alongside Hadoop, forming what is called the *Hadoop ecosystem*. They are designed to work together as an effective solution for the storage and processing of data. The Hadoop products which are integrated into most distributions are HDFS, MapReduce, HBase, Hive, Mahout, Oozie, Pig, Sqoop, Whirr, Zookeeper and Flume.

Fig. 1 illustrates the *Hadoop ecosystem*, and the relationship between its different components [9].

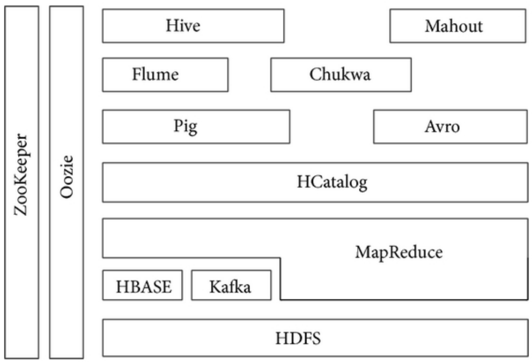

Fig. 1 The Hadoop Ecosystem [9]

The core of Apache Hadoop consists of two components: *a distributed file system (HDFS - Hadoop Distributed File System)* and *a framework for distributed processing (MapReduce)* [9]. Hadoop was designed to operate in a cluster architecture built on common server equipment.

Given the distributed storage, the location of the data is not known beforehand, being determined by Hadoop (HDFS). Each block of information is copied to multiple physical machines to avoid any problems caused by faulty hardware.

Unlike traditional systems, Apache Hadoop provides a limited set of functionalities for data processing (MapReduce), but has the ability to improve its performance and its storage capacity as it is installed on more physical machines. MapReduce processing divides the problem into sub-problems which can be solved independently (the *map* phase), in the manner of "divide et impera". Each of the sub-problems is executed as close to the data on which it must operate as possible. The results of the sub-problems are then combined according to needs (the *reduce* phase).

The core components build the foundation of four layers of the *Hadoop Ecosystem*, which make up a collection of additional software packages [9], [10]:

– *Data Storage Layer,* for storing data in a distributed file system. It consists of:
- HDFS: the main distributed storage;
- HBase: a NoSQL column-oriented distributed database based on the Google BigTable model which uses HDFS as storage media. It is used in Hadoop applications which require random read / write operations on very large data sets, or for applications which have many clients. HBase has three main components: a client library, a master server, and several region servers;

– *Data Processing Layer* where the scheduling, resource management and cluster management are calculated:
- YARN - a resource management platform which ensures security and data governance on different clusters;

– *Data Access Layer, where the request from the Management layer is sent to the Data Processing Layer.*
- Hive - a data storage platform (DW) used for querying and managing large data sets from distributed storage. Hive uses a SQL query language named HiveQL;
- Pig - a high level platform used for analysing large data sets, which has a specific language for describing data analysis programs. Pig allows parallel processing at run time. The Pig compiler produces MapReduce jobs. The Pig architecture generates a high-level scripting language (Pig Latin) and operates in real time on a platform which allows users to execute MapReduce on Hadoop. Pig offers more flexible data formats than Hive, providing its own data model. Pig has its own data type, *map*, which is used to represent semi-structured data;
- Mahout - a library for machine-learning algorithms and data mining, including classification and clustering algorithms. Many algorithms are written for compatibility with MapReduce such that they are scalable to large data sets;
- Avro - serializes the data, manages remote procedure calls and exchanges data from one program or language to another. Data is saved based on its own schema because this enables its use with scripting languages such as Pig;

– *Data Management Layer, which provides user access to the system through components like:*
- Oozie: a tool for workflow management and the coordination of MapReduce jobs. It enables the combination of several jobs into a logical workflow job;
- Chukwa: a framework for monitoring large distributed systems built on top of HDFS and MapReduce. It is used for collecting and processing data from distributed systems, which it then stores in Hadoop;
- Flume: a distributed service which enables the collection, aggregation and movement of large volumes of log data. Its architecture is based on data flows, which allows the development of analytic applications;
- Zookeeper: a coordination service for distributed applications. It maintains, configures and names large amounts of data. It also provides distributed synchronization and group services. Zookeeper contains master and slave nodes, and stores configuration information.

Companies use certain components of Hadoop, depending on the application.

2. The Integration of BD and Hadoop

There is no one BD architecture which can be used in all situations. Each processing layer of the architecture can make use of several solutions and techniques to create a robust environment. Each solution offers its own advantages and disadvantages for each particular workload.

BD integration raises issues due to the variety of data sources, the quality of the data to be integrated and data visualization [11].

The architecture of a BD integration ecosystem (see Fig. 2) includes the following components [11], [12]:

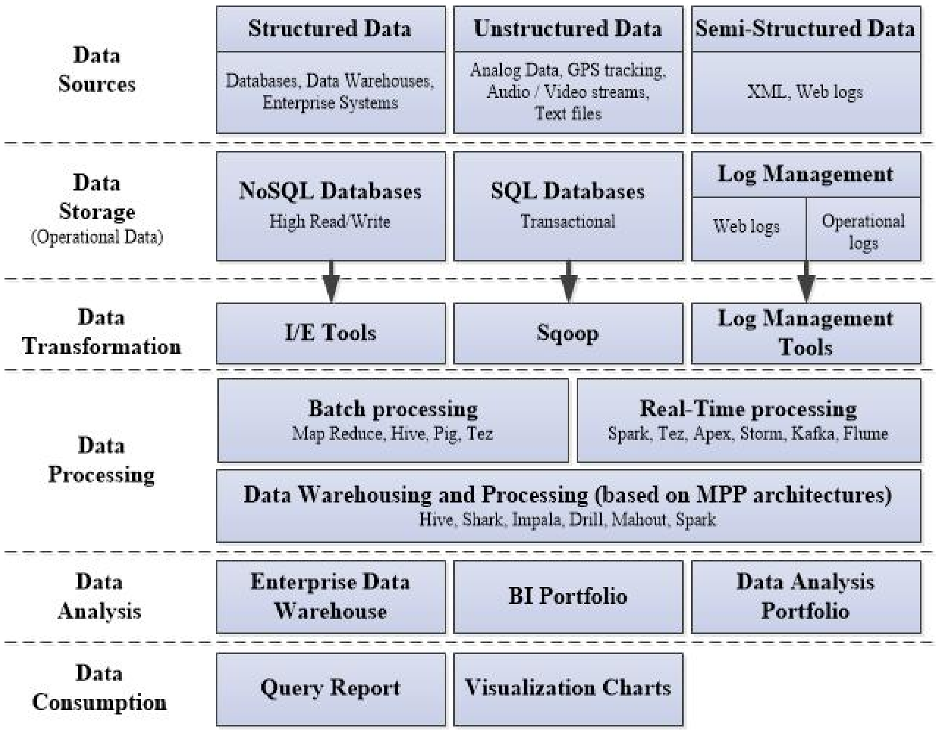

Fig. 2 The architecture of a BD integration ecosystem

1. *Data sources*: the emergence of tables which are stored in the Cloud and of mobile infrastructures has led to a significant increase in the size and complexity of data sets. The data integration ecosystems must thus include multiple strategies for the access and storage of a huge quantity of very varied data. The following classification can be made:
   - *Data sources for structured data* (e.g. databases, DWs, enterprise systems);
   - *Data sources for unstructured data (*e.g. analog data, GPS tracking, audio/video streams, text files);
   - *Data sources for semi-structured data* (e.g. XML, web logs);
2. *Data storage*: the data collected is stored in NoSQL/SQL databases, or Log Management systems for logs;
3. *Data Transformation:* in order to load data into the processing phase, it must first be transformed by using: import/export tools (SQL/NoSQL vendor specific tools), Sqoop (data source to Hadoop data transformation tool), Log management tools;
4. *Data Processing*: both structured and unstructured data are combined so that batch processing or real-time processing can be performed. Data Warehousing and Processing then generate usable data for data consumption.
5. *Data Analysis:* can be performed using:
   - *DWs*: ensure the necessary basic information. New functionality must be added for the better integration of unstructured data sources and for satisfying the level of performance required by analysis platforms. In order to perform strategic decisions, operational analysis has to be separated from deep analysis, which makes use of historical data. The data integration ecosystem must be able to separate operational information from the data sources used in developing long-term strategies. At the same time, it must allow a quick access to the most needed data;
   - *Portfolio for BI:* focuses on past results and performance. The data integration ecosystem will need to cope with the architectural challenges posed by the need for both self-service and mobile BI, and the need to integrate the BI and data analytics portfolios. This can lead to an increased demand for business information integration;
   - *Portfolio of data analytics*: the analysis activity in this portfolio must manage both company activity and data problems. The data integration ecosystem must provide access to structured and unstructured data for business analysis and allow reuse of previously conducted analyses to increase efficiency;
6. *Data Consumption*: the results of the data analysis have to be presented is a readable and accessible form to the final users**.** Query reports or visualisation charts can be used.

B. Technologies for BD

BD technologies are a rapidly developing field that deals with analysing and managing large volumes of data. They include both hardware and software systems that integrate, organize, manage, analyse and present data characterized by the 7 Vs of BD.

Technological innovations that have facilitated the emergence of BD can generally be structured into two categories: on one hand, storage technologies, especially fuelled by the development of Cloud Computing; on the other, processing technologies, including new databases suitable for unstructured data (Hadoop) and high-performance computing (MapReduce). These two innovations, supported by Google and Yahoo, have become the foundation for current BD processing. It is now possible to process large volumes of data of all types and at high speed, and much quicker than with previous technologies.

1. NoSQL

Traditional relational databases cannot handle the current challenges brought about by BD [13]. Lately, NoSQL (Not Only SQL) databases (non-relational databases) are increasingly popular for large scale data storage. They emerged from the need of companies like Google, Facebook or Twitter to manipulate huge amounts of data that traditional databases cannot handle.

NoSQL databases have been designed to store large volumes of data, usually partitioned across multiple servers. They offer flexible ways of working, support for easier data transfer, a simple API, and eventual data coherence. The main advantage of using NoSQL databases is that they allow efficient work with structured data, such as email, multimedia, text processors.

NoSQL databases, as a new generation of databases, are not relational, but distributed, open source, and are scalable horizontally. Another important feature of NoSQL systems is the "shared nothing" architecture in which each server node is independent and does not share memory or space.

NoSQL databases have a simpler structure and a different technology for data storage and retrieval than relational databases, which result in better performance for the real-time analysis of large volumes of data. NoSQL does not mean "no SQL" but rather "not only SQL" [14]. This new DBMS drops the structure, entity relationships, and most of the other principles of the ACID (Atomicity, Consistency, Isolation, Durability) model. Instead, it proposes a more flexible model names Basic Availability Soft State Eventual Consistency (BASE).

NoSQL databases do not use a specific schema [13]. Instead, data is stored using one of the following models:

- The key-value data model: very efficient and flexible, but not for self-describing data;
- The column data model: used for scattered data;
- The document data model: used for data XML deposits but inefficient;
- The graph data model: used for the explicit storage of data relationships, but inefficient in searches.

The NoSQL movement represents an attempt to exceed the limits of the relational model, and a step towards NewSQL, namely relational plus the NoSQL functionalities [15].

The most popular NoSQL databases at this time are: Cassandra, MongoDB, CouchDB, Redis, Riak, Membase, Neo4j and HBase. A larger list is provided in [16].

IV. BD Applications in Romania

Romania is developing its research into the types and complexity of the data used in the public administration, as well as its legal framework for the implementation of government BD systems [17].

Some examples of areas where BD projects are achievable in the public sector are: health (statistical analysis of cases, telemedicine etc.), e-Government, smart city, culture, e-Commerce, national security. Within these areas, BD could bring the greatest benefits in particular in:

- The improvement of health and health services;
- Customer-oriented advertising;
- The improvement and optimization of cities and countries;
- The performance optimization of devices and machines;
- The understanding and optimization of business processes;
- The increase of the political participation of citizens and the efficiency of the administrative system;
- The improvement of security systems and police efficiency.

A. Some Applications of BD in the Health Sector

Some BD based systems are briefly described below [18]:

1) *The SIUI Integrated EHR (Electronic Health Record) System*: an information system of the National Health Insurance House (CNAS), which integrates networked county health insurance houses (CJAS) and providers of health and pharmaceuticals, under the supervision of CNAS. It has been set into full operation since 2008;
2) *The SIVMED Integrated EHR System*: a complex IT-based solution for health units. It covers the main functional aspects of the hospital environment and provides solutions to both medical and administrative intra-hospital issues;
3) *The SIPE Electronic Prescription Information System*: a natural extension of the SIUI system, providing all of the functionalities necessary for the introduction of electronic prescribing for compensated and free prescriptions. It became operational on June 1, 2012;
4) *Armonia®sanita*: a hospital information system, available in the Cloud, with modules covering most workflows in a hospital, from patient scheduling, hospitalization and laboratory, pharmacy and even surgery management;
5) *Integrated Emergency Service System*: coordinates emergency services at a national level and improves the reaction of operational units. It helps manage 1,400,000 victims each month.

B. Some Applications of BD in the Government Sector

The following line graphs present the latest (2010-2014) e-Government Indicators for Romania (bottom line) as compared to the EU averages (top line) [19] (see Fig. 3 and Fig. 4).

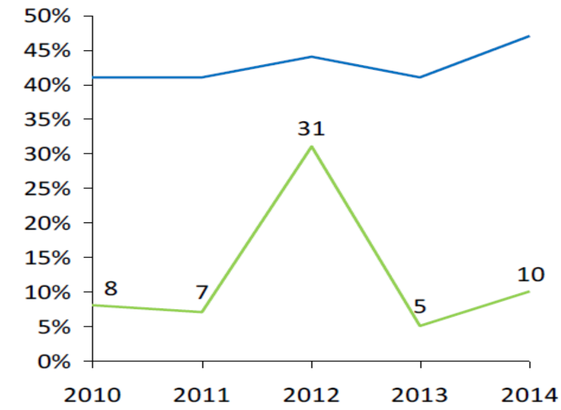

Fig. 3 Percentage of individuals using the internet for interacting with public authorities in Romania vs. EU [19]

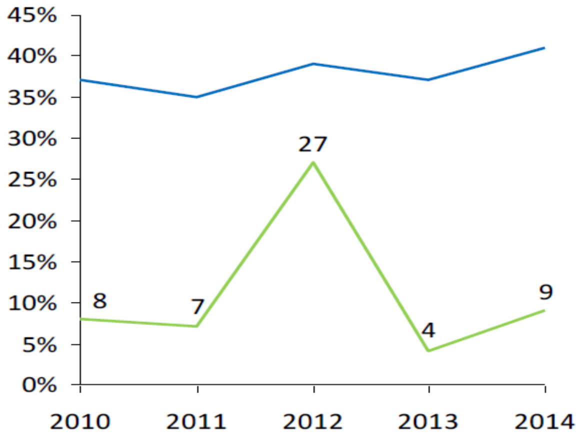

Fig. 4 Percentage of individuals using the internet for obtaining information from public authorities in Romania vs. EU [19]

On November 25, 2014, the Ministry for Information Society published the "National Strategy for the Digital Agenda for Romania 2014-2020" [17], which includes the strategy for implementing BD systems in Romania's public sector. This document is not only an instrument of convergence of Romania to the European Union in the field of ICT, but an opportunity to accelerate economic growth and sustainable development in Romania [19].

The following is a list of areas where the use of BD in the public sector is widespread and would bring the greatest benefits [20]:

1. *Electronic Public Procurement System (SEAP)*: a unified IT infrastructure which ensures the development of the procurement process in accordance with the law on public procurement through electronic auctions;
2. *Electronic Document Filing System (DEDOC)*: used for reporting individual budgets by means of electronic declarations and forms. It is in accordance with the legal provisions concerning the access to different sources of financing and the use of indicators of budgetary classification;
3. *Project for Modernizing Revenue Administration (RAMP)*: implemented by the National Agency for Fiscal Administration (ANAF) and generates prerequisites to achieve a higher level of revenue collection, with beneficial effects on the whole Romanian society;
4. *The Integrated Information System of the National Trade Register Office (NTRO)*: provides online services for the business community through a dedicated portal. It is used nationally by trade register offices attached to tribunals organized under the authority of NTRO;
5. *The Integrated System of the National House of Pensions (Orizont)*: manages the systems for public pensions and insurance in Romania, facilitating the management and access to information and benefits;
6. *National Cadastral and Estate Registry System (E-Terra)*: ensures the efficient management and electronic editing of land registers and the electronic management of cadastral coordinates.

III. Conclusions

BD is an opportunity to gain new insights into emerging types of data and content, in order to create more agile businesses and to answer to questions which were previously considered unanswerable.

BD is very attractive due to its ability to lead to better business results. It leads to radical changes to the operation of businesses, which shifts from using a model which was mostly based on the experience of the decider, to a model based on information which has real value to the business and to the organization itself.

The variety of data sources, the quality of the data to be integrated and data visualisation are some of the challenges for BD integration.

To evaluate the suitability of a technological solution based on BD, it is important to perform an examination of the types of business problems which need to be solved. Key factors for operational performance must also be considered. Examples of such key factors are scalability in relation to data volumes and the purpose of the analysis, the capability of the environment to grow together with the growth of the demand, and elasticity in satisfying the demand for mixed analytic task from a variety of businesses for different information consumers. The focus on key variables including data volumes, data variety, the need for performance and parallelization, is necessary.

Next, a strategic plan is developed for evaluating the different BD alternatives. Performance criteria can be used to select different suppliers. The clarification of success criteria allows the best estimation of value. The strategic plan must align BD technologies with existing infrastructures for BI and analytics.

References

[1] J. Gantz, D. Reinsel, "Extracting value from chaos", *IDC iView*, 2011, pp 1–12.
[2] E. Mcnulty, "Understanding Big Data: The Seven V's", *Dataconomy*, May 22, 2014, Retrieved from: http://dataconomy.com/seven-vs-big-data/.

[3] Gartner, "Big Data Strategy Components: Business Essentials", October 9, 2012, Retrieved from: https://www.gartner.com/doc/ 2191415/big-data-strategy-components-business
[4] Canada Inforoute, "Big Data Analytics in health", *White Paper, Full Report*, April 2013.
[5] A. Alexandru, D. Coardos, "BD in Tackling Energy Efficiency in Smart City", *Scientific Bulletin of the Electrical Engineering Faculty*, vol. 28, no. 4, pp. 14-20, 2014, Bibliotheca Publishing House, ISSN 1843-6188.
[6] Gartner, "IT glossary: big data" [webpage on the Internet]. Stamford, CT; 2012. Retrieved from: http://www.gartner.com/it-glossary/big-data.
[7] Frost & Sullivan White Paper, "Drowning in Big Data? Reducing Information Technology Complexities and Costs For Healthcare Organizations", 2012, Retrieved from http://www.emc.com/collateral/ analyst-reports/frost-sullivan-reducing-information-technology-complexities-ar.pdf.
[8] *Hadoop*, 2009, Retrived from: http://hadoop.apache.org/.
[9] N. Khan, I. Yaqoob, I. A. T. Hashem, et al., "Big Data: Survey, Technologies, Opportunities, and Challenges", *The Scientific World Journal*, vol. 2014, Article ID 712826, 18 pages, 2014. doi:10.1155/2014/712826
[10] Korneliusz. "Hadoop Ecosystem and Big Data", May 2, 2014. Retrieved from: https://blog.udemy.com/hadoop-ecosystem/
[11] V. Bodapati, "Data Integration Ecosystem for Big Data and Analytics", 2013. Retrieved from: http://smartdatacollective.com/raju-bodapati/103326/data-integration-ecosystem-big-data-and-analytics.
[12] V. Anuganti, "Typical “Big” Data Architecture", 2012. Retrieved from: http://venublog.com/2012/11/30/typical-big-data-architecture/
[13] P. Sadalage, “NoSQL Databases: An Overview”, Retrieved from: https://www.thoughtworks.com/insights/blog/nosql-databases-overview.
[14] D. Mc Creary, D., A. Kelly, *Making Sense of NoSQL: A guide for managers and the rest of us*, Manning, 2014, ISBN-13: 978-1617291074, ISBN-10: 1617291072.
[15] J. Piekos, “SQL vs, NoSQL vs. NewSQl: finding the right solution”, *Dataconomy* 2015, Retrieved from http://dataconomy.com/sql-vs-nosql-vs-newsql-finding-the-right-solution/.
[16] The Hadoop Ecosystem Table, Retrieved from: https://hadoopecosystemtable.github.io/
[17] National Strategy for the Digital Agenda for Romania 2014-2020, *Monitorul Oficial al României*, Partea I, No. 340 bis/19.V.2015, Retrived from: http://www.ancom.org.ro/uploads/links_files/Strategia_nationala_privind_Agenda_Digitala_pentru_Romania_2020.pdf
[18] A. Alexandru, D. Coardos, "Big Data in Health Care and Medical Applications in Romania", *Proceedings of AQTR 2016*, to be published.
[19] European Commission, *eGovernment in Romania*, January 2015, Edition 12.0.
[20] A. Alexandru, D. Coardoș, E. Tudora, "Using Big Data in the Government Sector", *Proceedings of IE 2016 Conference*, submitted for publication.

Adriana Alexandru (M’08) graduated the Faculty of Automatic Control, University Politehnica of Bucharest (1976), the Faculty of Mathematics, University of Bucharest (1982) and is Ph. D in Applied Informatics at University Politehnica of Bucharest (1998). She is Senior Researcher 1st degree and Head of research department at National Institute for Research and Development in Informatics, Bucharest and Professor at Valahia University of Târgoviște.

She coordinated several national projects in the field of e-Health, climate change, energy efficiency and renewable energies and was scientist in charge of 11 European projects in PECO, INCO COPERNICUS, SAVE, IST, EIE and Comenius European programmes. She is author of 3 books, co-author of 5 books and wrote over 100 articles published in Romania and abroad.

Prof. Alexandru is member of ISES, EHPA, VDI, SRIM, WASET, and Evaluation Board for RELANSIN, INFOSOC, CEEX, and PNII national research programmes and evaluator for FP V, EUREKA, and POC EU programmes.