Reflections on the Effect of Massive Data on Digitalized Accounting Information

Traian Cristin Nicolae
„Ovidius" University of Constanta, Faculty of Economic Sciences, Romania
traian.nicolae.profesor@gmail.com

Abstract

The purpose of this work is the issue of new software and hardware applicable to the realization of practical accounting work. Accounting as an academic discipline and practice of records, with millennial existence, is in the process of reconstructing its founding instrumentation. The impact of new technologies - the use of Information Technology (IT) tools, the expansion of applications based on artificial intelligence, radically changes the way the accounting system is thought and developed. The transformations generated by the emergence of new technologies in the concrete record-keeping activities of the entities, the new opportunities created by the existing software and hardware tools or proposed to be implemented, constitute as many challenges for the science of accounting and the practical techniques of organizing the accounting of the entities..

Key words: accountings paradigms, new technology
J.E.L. classification: M41, F60

1. Introduction

Although the use of digital technologies is essential for all sectors of the economy, each field has its own particularities, causing different speeds of implementation. (www.hyperledger.org)

Technologies of large volumes of data - BigData (BD) and/or complex data analysis solutions are used in digital platforms that mediate the alternative transport of people, in the communications sector and by large retailers, but much less used in the banking sector, within price comparison platforms and in e-commerce.

The implementation of BigData solutions brings general benefits, valid for any sector in which they were implemented, such as: optimization of operational and technical processes, efficiency of operational costs and resources, but also for a series of activities specific to each field of activity. (www.hyperledger.org)

Digitization investments are substantial, involve significant resources and long implementation times.

The barriers/ risks in the implementation of BigData technologies are generated by the high costs related to the acquisition/development and implementation of solutions, the inherent complexity of the data, the lack of skills within the companies for data interpretation, the difficulty of hiring specialized personnel, ensuring security and preserving data confidentiality.

The use of BigData technologies can bring competitive advantages to companies, such as: adjusting, in real time, the price level to changes in the demand/supply ratio, thus ensuring the rebalancing of the demand and supply of accounting information. (www.hyperledger.org)

2. Theoretical background for the use of the IT techniques in accounting

BigData (BD) represents the technologies to process and analyze, cost-effectively, the large and varied volumes of available data, at the desired speed, with the aim of improving business operations and decision-making processes. (www.hyperledger.org)

The BD architecture represents the basic system used to capture, store and process, in bulk and in real time, large volumes of data, so that they can be analyzed, including through predictive analysis and machine learning (automatic learning) tools. (www.hyperledger.org)

BigData Analytics (BDA) is the process of analyzing large volumes of data.

Dynamic Machine Learning (ML) algorithms are algorithms that decide for themselves what information to use (from a specified set) and how to process it to achieve the result they are programmed to achieve. (www.hyperledger.org)

Business Intelligence (BI) defines a term that refers to a set of software applications and technologies designed to collect, manage and analyze information from different business management systems or other databases. BI involves activities such as: analysis of large volumes of data, online analytical processing, queries and reporting. (www.hyperledger.org)

The notion of BD appears in the digital world in the 2000s, with reference to a large set of data, at that time, impossible to manage and process with existing technical solutions. The exponential growth of information and communication technologies, the expansion of global Internet access in recent years have stimulated the emergence of the digital economy and allowed the development of business models based on the collection and processing of BigData. The term BigData refers to the tools, processes and procedures that enable the creation, processing, storage and management of large data sets. The explosion of data volume characteristic of the BigData phenomenon also comes from unstructured data, and the large volume of data that must be processed requires specific techniques, tools and architectures. BigData generates value by storing and processing large volumes of digital information (structured, semi-structured or unstructured), which can no longer be analyzed with traditional techniques. (www.hyperledger.org)

The amount of digital data generated by the evolution of the Internet, the emergence of IoT (Internet of Things) and better resolution of digital images worldwide is expanding exponentially. It is estimated that the volume of data will reach 44 ZB(zettabytes) in just 20 years after entering the 21st century. And every year, 4.5 ZB of new data can be generated: 1 zettabyte (ZB) = 1 billion terabytes (TB). (www.hyperledger.org)

Big Data is responsible for managing, managing and processing different types of data such as structured, semi-structured and unstructured. Big data technologies are cost-effective in terms of maintaining a large amount of data, they work on a distributed database system. One can save large amounts of data for a long time using BigData techniques. So it is easy to manage historical data and generate accurate reports. The data processing speed is very fast and therefore social media uses Big Data techniques. Data accuracy is a big advantage of Big Data. Big data technologies allow users to make effective decisions for their business based on current and historical data and error management, version control and customer experience are very effective in BigData.(Nicolae, 2010)

3. Research methodology. Evolution of IT techniques in the accounting record system - empirical aspects of research

The research methodology used in this article consists of empirically researching a data set available on the Internet and building models of accounting records.

Big Data influences the activity of companies, which are forced to reconsider their organization and business processes, taking into account the data they have and which could be transformed into a competitive advantage in the new information-based market. The decision-making process is moving from a model based primarily on the experience of the decision-maker to one based on information and often on the experience of the user. The use of Big Data offers a new perspective by improving the practices of analysis and predictive modeling, with a positive effect in the decision-making process in real time. (www.hyperledger.org)

Just as it helps companies increase their productivity, BD enables public administrations to improve their IT systems and public sector administration, helping global organizations analyze information for strategic planning. (www.hyperledger.org)

Like any new field, the BD concept comes with a series of challenges regarding the acceptance and implementation of these specific technologies. (www.hyperledger.org)

The insufficient understanding and acceptance at the organizational level of the BD phenomenon is a challenge at present. Without a clear understanding of the BD phenomenon: what BD is, the benefits it brings, which architecture lends itself best to the organization, the BD implementation project risks failure. The implementation of BD technologies represents a major change within the organization, a change that must be recognized and accepted at all its levels. (www.hyperledger.org)

The variety of existing BD technologies and architectures on the market makes it difficult to choose the most appropriate one for the organization to implement. Taking into account the permanent expansion of data volumes, the chosen solution must allow, without decreasing performance, the possibility of including new processing and storage capacities. (www.hyperledger.org)

Local implementation involves costs with new hardware, new employees – administrators/developers. Although most framework software systems are open source, there are costs associated with the development, configuration and maintenance of new programs. And the cloud storage solution involves, in addition to infrastructure access costs and costs with new employees, development, configuration and maintenance. For the implementation of BD, it is necessary to establish a team of data analysts and competent programmers (who filter and analyze the data), whose employment is difficult and expensive. (www.hyperledger.org)

BD technologies are never 100% accurate, data comes from various sources in a variety of formats, processing it and filtering useful information from the very large volume of raw data is difficult. It is also difficult to validate the veracity of data, and erroneous and/or redundant data can provide wrong or contradictory solutions. (www.hyperledger.org)

Data security and confidentiality involves adequate back-up mechanisms for large and varied volumes of data and a unified access control mechanism for multiple sources and types of data.

Currently, BD technologies are subject to national privacy and data protection legislation. The current legal system focuses on specific forms of data protection (personal, private data, statistical data), from the data collection stage and on the principle of purpose limitation (data cannot be used for purposes other than those for which they were collected). Instead, the logic of BD technologies is more data, more accurate results and is based on the general collection of data, combining them from several sources and their secondary use in other analyses. The most important stage of BD technologies, the analysis stage, in which the algorithms, data sources, classification and weighting criteria are established, is not yet regulated, and the automatic decision-making process is not prohibited. (www.hyperledger.org)

Given the rapid evolution of technologies and the long implementation period of a BD solution, there is a risk that, by the time the implementation period is completed, the technology will already be outdated.

4. Findings. Concise review of specific developments regarding the application of computer technologies (IT) in the accounting record system

Blockchain, translated from English chain of blocks, is a distributed database, which is shared between the nodes of a computer network. (IASB, 2022)

The digital market is the space in which goods, services and data are traded, using digital technologies and mainly includes electronic commerce, digital accounting and telecommunications. Trading in digital markets is done through digital, Internet-based channels. (IASB, 2022)

In the conditions of the increasingly rapid evolution of digital technologies, on which all sectors of the economy are currently based, increasing competitiveness and productivity crucially depend on the ability to effectively generate, amplify and exploit digital innovations. (www.hyperledger.org)

To support Europe's role in the global digital economy, in 2015 the Commission launched the Digital Single Market Strategy, based on the concept of an internal market, in which the free movement of goods, people, services and capital is ensured. Through the digital single market, the EU aims to remove national barriers to online transactions and give European companies the opportunity to operate in any other EU country, legally, safely, securely and cost-effectively. The basis of an effective single market is provided by common standards that ensure the interoperability of digital technologies, foster innovation and reduce barriers to entry to the digital single market.

The data market, a component of the digital market, is the market where digital data is traded in the form of products or services, following their raw processing. Maximizing the growth of the data market is one of the important objectives of the EU and of the digital single market strategy for Europe.

According to the study on the European data market published in June 2020, the size of the European data market reached a level of over 58 billion euros in 2019, an increase of approximately 5% compared to the previous year. The evolution of the European data market maintained an upward trend, during the period 2013-2019, with an average annual compound growth rate of 7.4%.(www.ziare.com)

The data market impact measures the overall effects of the data market on the EU economy and includes, as mentioned previously, the direct (equal to the size of the data market), indirect and induced effects of the data market on the economy.

In 2019, the impact of the data market on the EU economy was approximately €325 billion, up 7.7% on the previous year, representing the equivalent of 2.6% of EU GDP. In the period 2013-2019, the trend was an upward one, with an average annual compound growth rate of 8.2%.(www.ziare.com)

Europe has a dynamic, growing (digital) data ecosystem. The evolution of the data market highlights the constant investments in the digital economy, artificial intelligence and robotics. Thus, the data market and its impact on the European economy had a constant and significant growth in the period 2013-2019. The compound annual average growth rate, in the period 2013-2019, was 7.4% for the data market, respectively 8.2% for its impact on the European economy. The faster growth of impact highlights the positive multiplier effect of innovative technologies based on information and data. (www.ziare.com)

Digital platforms are a promising tool in accounting. Currently, more than one million businesses in the EU trade their products or services, digital or not, through the platforms.

The use of online platforms improves activity efficiency and competitiveness, increases consumer choice, offers new business opportunities for enterprises of all sizes, by generating alternative business models, based on new technologies and access to the global market.

Digital platforms, like any classic intermediary, connect two distinct groups: sellers/ buyers, accountants/ users of accounting data, teachers/ students. They cover a wide range of activities: financial accounting, financial reporting, management accounting, price comparators, application platforms, search engines, payment systems, collaborative economy. (IASB, 2022)

To meet the challenges of new data processing requirements, Big Data platforms are complex systems made up of a suite of components with a well-defined role in processing data volumes, as follows: the data collection component (structured, semi-structured, unstructured); the data quality component (cleaning, transformation and ETL loading); the data storage component (Data Lake, Data Warehouse); data analysis component; the component of exposing the result of data processing to end users. (www.hyperledger.org)

Using Big Data technologies, companies can collect more precise and detailed data, both structural and behavioral, from internal or external sources, to help them understand the needs of customers, their preferences, but also the phenomena that influence performance to achieve superior results. The volume of data is growing exponentially, the phenomenon affecting every company/institution. But it is not only the volume of data that is increasing, but also the number of data sources. (www.hyperledger.org)

Different approaches to data acquisition have emerged depending on their scope, developing a series of protocols and tools, including open source solutions, that support the data acquisition process. 90% of the data currently available globally was produced in the last 2 years. The source and nature of this data is diverse, the speed with which it is collected and exchanged is one in real time (through sensors that take data from the movement or through smartphone applications, tablets, etc.), and the variety of their formats poses data management problems (you can collect anything from emails to financial transactions). (www.hyperledger.org)

In order to have access to these huge volumes of data, they must meet several conditions, such as: availability, accessibility, persistence, interoperability.

The underlying assumption is that multiple volatile data sources generate information that must be captured, stored and analyzed by a big data processing platform. New information generated by the data source is passed to the data store using a data acquisition framework that implements a predefined protocol.

Several of the organizations that rely on big data processing have designed protocols specific to their purposes, which are not public and therefore cannot be described in this chapter. Open protocols are frequently used for data acquisition.

Advanced Message Queuing Protocol (AMQP) is an open protocol developed to meet the data acquisition requirements of large companies. (www.hyperledger.org)

The Java Message Service (JMS) protocol API. According to the specification, JMS provides a common way for Java programs to create, send, receive, and read messages in an enterprise messaging system. (www.hyperledger.org)

As for data acquisition software tools, many of them are well known and available on the Internet.

Some examples are shown below.

Many companies and applications use open-source application Storm to power a wide variety of production systems that process data, including: Groupon, The Weather Channel, fullcontact.com, and Twitter. (www.ziare.ro)

Kafka was originally developed by LinkedIn to monitor user engagement and define the relevance of their posts. (www.ziare.ro)

There are mainly four recognized techniques for acquiring data from the Internet.

Crawlers - are tools used to download many pages from the Internet. The main properties of crawlers are scalability and volume. They follow links on web pages on the Internet (or within a website) and download pages. They can be distributed across multiple devices to download tens of thousands of web pages. Among the most well-known technologies for crawling: Heritrix – from Open Internet; Archive Nutch – from Apache; Aspider - from Search Technologies etc. (www.ziare.ro)

Scrapers - are usually less scalable and require more programming work than crawlers and are generally used for extracting structured data from web pages based on presentation structure.

Browser Automation - fetches and renders the page like a web browser. Browser automation tools actually run JavaScript extracted from web pages and render HTML (and other data structures). These can then be combined with custom scripts to explore the results and download content that would otherwise be inaccessible. Among the most well-known browser automation tools: Stropi, PhantomJS, Selenium, WebDriver.io, Nightmare, etc. (www.ziare.ro)

API (Application Programming Interface) - APIs are made available by content providers to facilitate the integration of different software products. Examples: Thomson Reuters, LexisNexis, Bing, Factiva, NewsCred, etc. (www.ziare.ro)

These tools must meet requirements such as: being able to pick up frequent data changes to ensure that the data destination is in sync with the content source; store the data as it is retrieved; any operation on the data (normalization, cleaning, enhancement) is performed later in the loading/storage process; be able to obtain all necessary metadata from the content source; be able to retrieve files stored in the content source.

In conclusion, data acquisition is an important process and allows the downstream components of data processing to perform their functions correctly.

In the case of BD technologies, data storage is provided by a specially designed infrastructure for recording, managing and retrieving massive amounts of files and data objects. A BD platform is built for much greater scale, speed and performance than traditional enterprise data storage has allowed. It must allow large volumes of data to be stored and sorted in such a way that it can be easily accessed, used and processed by specific applications. Large data storage needs to be able to scale flexibly as needed.

A misconception about big data is that the term only refers to the size of the data set. While this is true in principle, the science behind big data goes deeper. The intent is to extract specific subsets of data from multiple large storage volumes. This data may be widely dispersed across different systems and may not have an obvious correlation. The goal is to unify the data in an intelligent way to enable rapid analysis.

From the point of view of big data storage, the analysis can be carried out in two dimensions closely interconnected and dependent on each other, respectively: hardware infrastructure and software infrastructure.

Data Warehouse (DW) are central repositories of structured data integrated from one or more disparate sources, used for reporting and analysis using Business Intelligence (BI) technologies. In fact, DW is a database optimized to analyze relational data from transactional systems and specific business applications. DW architecture was born in the 80s as an architectural pattern designed to support the flow of data from operational systems to decision support systems.

The trend of moving from structured to unstructured data makes traditional relational databases unsuitable for storage. This shortcoming of relational databases motivates the development of efficient distributed storage mechanisms.

Thus, NoSQL (not only SQL) technologies emerged that introduce flexible data models, horizontal scalability and unstructured data models. These databases aim to provide ease of scalability and handling of large volume data. NoSQL databases provide some level of transaction management, so they are suitable for social networks, email and other web-based applications. To improve data accessibility to its users, data is distributed and replicated across multiple sites using cloud infrastructures. (www.ziare.ro)

Combining the capabilities of the two technologies for structured and unstructured data, Data Lakes have been developed, which store relational data from business applications and non-relational data from mobile applications, IoT devices and social media. The data structure or schema is not defined when the data is captured. This means that data can be stored without careful design or the need to know what questions might need answers in the future. Different types of analytics such as SQL queries, big data analytics, full-text search, real-time analytics, and machine learning are used to discover diverse information. (www.ziare.ro)

Recently, organizations that have implemented DW systems have seen the advantages of Data Lakes and are developing them in the sense of implementing these new technologies that allow various query capabilities, data science use cases and advanced capabilities for discovering new information models. (www.ziare.ro)

Cloud computing is a modern concept in the IT field. Cloud-based solutions are the benchmark for managing large volumes of data, services and their associated analytics. (www.hyperledger.org)

There is a general tendency of companies to migrate to cloud services due to the following: reduced time required to implement new services, real-time cost control, scalability according to demand, ability to adapt the resources used.

Thus, cloud computing represents a viable alternative to on-premise corporate data centers, enabling the on-demand delivery of IT resources and applications from and over the Internet, with a pay-as-you-go pricing model. (www.hyperledger.org)

On-demand delivery involves provisioning resources to the customer only when they need it, without prior planning. The infrastructure and software resources are already available in the data centers of the cloud provider and are only made available to the customer when he requests them, immediately or in a very short time frame. (www.hyperledger.org)

IT resources are all or parts of an IT infrastructure or services. In the Cloud, they can be categorized as Infrastructure as a Service - abbreviated IaaS. Examples of cloud IT resources: virtual servers, virtual hard disks, network cards, network addresses (ips), Internet gateways, virtual load balancers, etc. (www.ziare.ro)

Cloud apps are software that you use and pay for as a service. The concept is called Software as a Service abbreviated SaaS. Examples of SaaS: document editing suites-Amazon WorkDocs, Google G Suite, Office 365, ERP-type applications, etc. Most software developers have come to offer a Cloud version, with a monthly payment, in parallel with the classic sales offer. (www.ziare.ro)

5. A case study of the accounting of set of transactions

Table no. 1 Table of transactions related to the case study

Transactions	Textual description of the transactions
01.01.2023	The Astra9 company purchases 2,000,000 Digital Euros on 01.01.2023, at the price of 6 monetary units (m.u.). (Nicolae, 2010).
December 31, 2023	On 31.12.2023 the re-estimation is done and it is found that the price of a Digital Euros has dropped to 5.5 m.u.
December 31, 2024	On 31.12.2024, the re-estimation is done again, the Digital Euros reaching the price of 5.7 mu
December 31, 2025	On 31.10.2025 the company sells the Digital Euros at the price of 6.2 m.u.

Source: Case study data proposed by the author

Note:
Debit = D
Credit = C

Accounting data
(Making entries by the author)

Table no. 2 *Calculations at 01.01.2023*

Calculations
Purchase cost: 2,000,000 Digital Euros (EuDi) x 6 um = 12,000,000 um

Source: Calculations made by the author

Table no. 3 *Purchase of Digital Euros (EuDi) at 01.01.2023*

Account - D	Account - C	Amount
Cash equivalents – EuDi	*Bank accounts*	12,000,000

Source: Calculations made by the author

Table no. 4 *Calculations at 31.12.2023*

Calculations
Cost of re-estimation = 2,000,000 *EuDi* x (6 – 5.5) = 1,000,000 um

Source: Calculations made by the author

Table no. 5 Re-estimation *of of Digital Euros (EuDi) at 31.12.2023*

Account - D	Account - C	Amount
Other financial expenses	*Cash equivalents – EuDi*	1,000,000

Source: Calculations made by the author

Table no. 6 *Calculations at 31.12.2024*

Calculations
Revenue from *EuDi* re-estimation = 2,000,000 *EuDi* x (5.7 – 5.5) = 400,000 mu

Source: Calculations made by the author

Table no. 7 Re-estimation *of of Digital Euros (EuDi) at 31.12.2024*

Account - D	Account - C	Amount

Source: Calculations made by the author

Table no. 8 *Calculations at 31.12.2025*

Calculations
Downloading the book value of Cash equivalents – *EuDi* = 2,000,000 *EuDi* x 5,7 = 11,400,000 mu
Financial income = 2,000,000 *EuDi* x (6.2 – 5.7) = 1.000,000 mu
Total amount collected = 2,000,000 *EuDi* x 6.2 = 12,400,000 mu

Source: Calculations made by the author

Table no. 9 Sale of of Digital Euros (EuDi) at 31.12.2025

Account - D	Account - C	Amount
Bank accounts	*Cash equivalents – EuDi*	11,400,000
Bank accounts	*Other financial revenues*	1.000,000

Source: Calculations made by the author

6. Conclusions

In my opinion, the application of new, innovative tools and techniques, which appeared as a result of the development of the informational capabilities of information technology (IT), provides the necessary foundation for a paradigm shift in the processing of accounting information, the development of new ways of thinking and practice in the activity financial accounting of the entities.

The widespread global use of information technology tools – smart phones, tablets, laptops, wireless accounting information transmission devices and 5G/6G, all of which open up new opportunities in processing large volumes of accounting data in real time.

Financial reporting - annual financial statements in the classic sense - are about to be transformed and incorporated into a system of integrated financial reporting. Integrated financial reporting allows multi-dimensional and multi-criteria treatment of financial-accounting information. It improves, dramatically reduces the time required to transmit accounting information generated by entities to users.

A little-studied aspect refers to the energy consumption generated by the use of smartphones, tablets, laptops, 5G/6G devices, etc., much higher consumption compared to the traditional paper-based processing of accounting data.

7. References

- IASB (International Accounting Standards Board), 2011. *International Financial Reporting Standard 13 Fair Value Measurement*. [online] Available at: http://eifrs.ifrs.org
- IASB, 2022. *International Financial Reporting Standards (IFRSs)*. [online] Available at: http://eifrs.ifrs.org.
- Nicolae, T., 2010. *Standarde contabile [Accounting standards]*. Constanta: Ex Ponto Publishing House.
- Nicolae, T.C., 2015, *Contabilitate financiară aprofundată: caiet de studiu individidual: învățământ la distanță(ID) [Advanced financial accounting: individual study book for distance learning (DL)]*. Constanta: Ovidius University Press.
- * * * *Law of the Fiscal Code of Romania.* [online] Available at: www.anaf.ro.
- www.arbc.ro
- www.ziare.ro
- www.bnr.ro
- www.hyperledger.org
- www.insse.ro