BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **24**, No 1
Sofia • 2024 Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2024-0001

Integration Approaches for Heterogeneous Big Data: A Survey

Wafa' Za'al Alma'aitah[1], *Addy Quraan*[2], *Fatima N. AL-Aswadi*[3,4], *Rami S. Alkhawaldeh*[5], *Moutaz Alazab*[1], *Albara Awajan*[1]

[1]*Department of Intelligence Systems, Faculty of Artificial Intelligence, Al-Balqa Applied University, Al-Salt 19117, Jordan*
[2]*Department of Basic Sciences, Faculty of Science, The Hashemite University, Zarqa, Jordan*
[3]*Institute of Computer Science and Digital Innovation, UCSI University, 56000 Kuala Lumpur, Malaysia*
[4]*Faculty of Computer Science and Engineering, Hodeidah University, Al Hudaydah, Yemen*
[5]*Department of Computer Information Systems, The University of Jordan, Aqaba 77110, Jordan*
E-mails: wafaa_maitah@bau.edu.jo addy@hu.edu.jo Fatima.Nadeem@ucsiuniversity.edu.my fatima_aswadi@hoduniv.net.ye r.alkhawaldeh@ju.edu.jo m.alazab@bau.edu.jo a.awajan@bau.edu.jo

***Abstract*: *Modern organizations are currently wrestling with strenuous challenges relating to the management of heterogeneous big data, which combines data from various sources and varies in type, format, and content. The heterogeneity of the data makes it difficult to analyze and integrate. This paper presents big data warehousing and federation as viable approaches for handling big data complexity. It discusses their respective advantages and disadvantages as strategies for integrating, managing, and analyzing heterogeneous big data. Data integration is crucial for organizations to manipulate organizational data. Organizations have to weigh the benefits and drawbacks of both data integration approaches to identify the one that responds to their organizational needs and objectives. This paper aw well presents an adequate analysis of these two data integration approaches and identifies challenges associated with the selection of either approach. Thorough understanding and awareness of the merits and demits of these two approaches are crucial for practitioners, researchers, and decision-makers to select the approach that enables them to handle complex data, boost their decision-making process, and best align with their needs and expectations.*

***Keywords*: *Heterogeneous Big Data, Integration approaches, Big Data warehousing, Big Data federation.*

1. Introduction

The advancement of digital technology and the influx of data from various sources have made handling heterogeneous big data a vital task for organizations [1], which need to be able to merge and assess such data for deeper insight and effective decisions [1]. Additionally, traditional data management approaches are unable to handle heterogeneous data or deal with a wide range of data sources, formats, and quality [1]. Therefore, organizations need to employ a comprehensive approach drawing on advanced data management techniques [1]. This paper addresses big data federation and warehousing as integration approaches for organizations to adopt in managing heterogeneous big data. It also covers the specific management issues associated with the volume, velocity, and variety of big data. Data management experts define big data based on the 3Vs, namely, volume, velocity, and variety.

However, it can also be described by the 5Vs, which include the addition of Veracity (data reliability) and Value (insights). Organizations may efficiently handle and analyze heterogeneous data, obtain important insights, and enhance decision-making by recognizing these issues and potential solutions. The following issues are discussed in this paper: (a) two approaches for managing heterogeneity in big data, and (b) the difficulties related to researching big data heterogeneity. The integration approaches to manage big data heterogeneity are: (a) big data warehousing, and (b) big data federation. In the sequel, we mention the two approaches with DW and DF, respectively.

The remaining sections of this paper are structured as follows. Section 2 provides an overview of big data. The big data heterogeneity is presented in Section 3. Integration approaches are discussed in Section 4. A comparison between the DW and DF approaches, analyses, and findings are presented in Sections 5. Finally, Section 6 concludes the paper.

2. Overview of Big Data

Big Data refers to the enormous amount of organized and unstructured data that organizations are currently dealing with [2-5]. This enormous data volume is generated by numerous sources, including sensors [6, 7], e-Commerce transactions [8, 9], and social media [10]. According to [11-13], while technology develops, Big Data generation will only increase, necessitating the use of more advanced techniques for their storage, processing, and analysis. The primary big data characteristics that make them so distinct and difficult to manage include [14, 15] 5Vs.

- **Volume.** Organizations generate vast amounts of data, which continues to rise at an exponential rate [4, 16]. The growing usage of digital devices [17], the Internet of Things(IoT) [18], and cloud computing [19] are driving data growth. According to [20, 21], handling and storing such massive data volumes requires creative solutions, such as the use of distributed systems and cloud storage.

- **Velocity.** One important part of big data is how fast data are collected and handled [22]. In real-time situations, data are created at a rapid rate, and it is important to examine them quickly to get useful insights [23]. As [24, 25] report, it requires new ways to process data consisting of stream processing and real-time analytics to deal with the amount and speed of big data.

- **Variety.** Big data includes a wide range of data kinds, including structured, semi-structured, and unstructured data [26, 27]. Data displayed in a tabular fashion, as databases, are referred to as structured data [28]. Semi-structured data contain some structure but are not as well ordered as structured data [26, 29. Unstructured data are data that do not have a specific organization, such as photographs, videos, and text documents [30, 31]. Processing and analyzing such a wide set of data types necessitates the development of novel approaches and technologies, such as natural language processing and computer vision [32, 33]. It has been stated that while working with a range of data formats, there is always a possibility of receiving inaccurate results [34].
- **Veracity.** Big data quality and accuracy are key challenges. With such vast data amounts, it is difficult to ensure the data's reliability and consistency, making accurate decisions based on the data difficult [16, 35]. Modern data validation and cleansing processes, such as data standardization and profiling, are required to guarantee the integrity and reliability of data. Reference [36] proposed a solution to effectively handle authenticity concerns and significantly reduce the number of big data occurrences.
- **Value.** Big data has the possibility of giving useful insights and influencing corporate decisions, despite the limitations it faces [37]. Organizations can identify hidden patterns and relationships in data by studying them, as demonstrated by [38]. Organizations can find hidden patterns and connections by studying their data. According to [39], developing BDAC leads to better decision-making and outcomes. To derive meaningful insights from data, it is imperative to employ sophisticated data analytics methodologies, including predictive modeling and machine learning. Large datasets provide organizations with the opportunity to find complicated patterns and connections that would be challenging to recognize in smaller datasets [40]. This phenomenon has the potential to enhance decision-making processes, improve consumer experiences, and improve operational efficiency. It is necessary to implement specific tools and technologies, such as Hadoop, Spark, and NoSQL databases, to properly analyze massive data [41-43, 5]. These technologies are intended to deal with the volume, variety, velocity, and veracity of big data, while also providing quick and precise insights.

Table 1. Definition, advantage, and limitation of the key characteristics of Big Data

Metric	Definition	Advantage	Limitation
Volume	The amount of data	Captures scale of data	Does not consider the usefulness or quality of data
Velocity	The rate at which data is generated, captured, and processed	Captures the speed of data processing	Does not consider the usefulness or quality of data
Variety	The wide variety of data forms and origins	Captures the diversity of data	Does not consider the usefulness or quality of data
Veracity	The trustworthiness and accuracy of data	Captures the quality and reliability of data	Difficult to measure objectively
Value	The business value or impact of the data	Captures the usefulness of data	Difficult to measure objectively

Many studies have been dedicated to utilizing the capabilities of big data in various sectors. References [44-49] have made significant contributions to education. Similarly, significant progress has been made in health care by [50-52] using big data to improve patient care and medical research. Big data have served an important part in transforming industries, including finance [53-55], retail [56, 57], telecommunications [58, 59], and tourism [60, 59, 61]. The widespread usage of big data has demonstrated its revolutionary impact across a wide range of industries, motivating researchers to further investigate its enormous potential. The enormous dataset sizes are directly correlated with the increased likelihood of heterogeneity, as noted in [62]. This relationship is a very important aspect of big data. The definitions, advantages, and limitations of the main characteristics of big data outlined above are summarized in Table 1.

3. Big Data heterogeneity

In the world of data, the term “heterogeneity” encompasses the range of diverse and complex data types and sources. This poses challenges for organizations regarding the efficient handling and evaluation of such data [63, 1]. The data exponential rise has resulted in a significant increase in data generation, originating from many sources including structured, unstructured, and multimedia formats. Data integration from many sources with diverse formats, structures, and models is also a challenge when working with heterogeneous data, as has been pointed out by [64-66]. Despite of the challenges mentioned above, working with heterogeneous data offers several benefits such as gaining a comprehensive understanding of the problems at hand, and making improved decisions based on the insights obtained from analyzing such diverse datasets [67, 68]. Additionally, references [67, 69] assert that the effective management and analysis of heterogeneous data necessitates the acquisition of specialized skills, knowledge, and advanced technologies within the data management domain. It is worth noting that when managing numerous data types simultaneously, there is a possibility of obtaining inaccurate results [34]. Therefore, it becomes crucial to address any quality issues arising due to the dataset heterogeneity, which can make it more challenging to process [70].

Several studies have highlighted the importance of taking a comprehensive approach to unlock the full potential of diverse data sources. This involves integrating, preprocessing, analyzing, governing, and semantically integrating the data. The aim is to gain valuable insights that can lead to business success and a competitive advantage. The study of [25] highlights the growing importance and applicability of data stream research, along with the developments in tools, technologies, approaches, and strategies for big data stream analysis. Since data integration is essential to producing high-quality data, it is recognized as an essential component of the analytic process [71].

A broad variety of data types, such as social media, sensor, financial, healthcare, customer, supply chain, human resources, environmental, educational, transportation, and manufacturing data, are known as heterogeneous big data. To remain competitive in this industry, experts have stressed the significance of analyzing these big data sources and employing efficient tools and approaches. To ensure that high-quality data are generated, data integration is essential. It is worth noting that organizations are facing serious challenges in managing and analyzing big data. For organizations to overcome such challenges, they are forced to utilize professional and specific expertise, tools, and techniques in managing their organizational data. Table 2 below presents the various sources of big data heterogeneity.

Table 2. The heterogeneity of big data sources

Examples	References	Description	Kinds of data
Social media data	[72]	The data derived from social media sites, blogs, and forums	Audio, metadata, videos, pictures and text
Sensor data	[73-75]	The data derived from sensors, Internet of Things (IoT) devices and smart cities	Statistical, logical, and geographical
Financial data	[76, 77]	The data derived from banks, credit card firms, and financial institutions	Numerical, categorical, and temporal
Healthcare data	[78, 79]	Data gathered from hospitals, clinics, and healthcare facilities	Diagnostics, imaging, and patient histories
Customer data	[80-83]	The data produced from Customer Relationship Management (CRM) systems, websites, and loyalty programs	Transactional, behavioral, and demographic
Supply chain data	[84]	Data produced from warehouses, transportation systems, and factories	Logistical, financial, and operational
Human resources data	[85-87]	Data collected by human resources, payroll, and performance management systems	Employee files, payment information, and performance evaluations
Environmental data	[88-91]	The data obtained from weather stations, sensors, and satellite photography	Environmental, geological, and meteorological
Educational data	[92-95]	The data derived from educational institutions, online courses, and learning management systems	Student documentation, test scores, study tools
Transport data	[96-98]	Data derived from GPS units, transportation networks, and traffic sensors	Data on traffic patterns, locations of vehicles and routes
Manufacturing data	[99-103]	The data that is produced by manufacturing systems, sensors, and quality control systems	Data on production, quality, and maintenance

4. Integration approach

Data integration refers to the process of merging and combining data from sources and formats to create a unified and seamless view [104, 105]. Integrating data from programs, databases, and file systems can present challenges in this process [106]. One of the difficulties with integrating data is the vast amount of data coming from diverse sources with different structures that are constantly evolving [107, 108]. Key challenges in data integration include linking records, mapping schemas, and fusing data [109]. Additionally consolidating all the data into a repository improves accessibility, promotes understanding of the information, and safeguards against any potential loss. Moreover, experts in data management face tasks like extracting, combining, and exchanging information to create a comprehensive integrated view. As a result of these complexities, modern organizations are compelled to integrate sets of data, from sources and information systems into a centralized information system [110].

Data integration is a procedural mechanism that facilitates the use and accessibility of data by users inside an organization, all the while ensuring the preservation of its integrity and quality. Moreover, it facilitates the ongoing synchronization of alterations made to data stored in a single source with multiple additional sources [111]. To handle and exploit massive and complicated data, numerous organizations have inadequately integrated massive amounts of disorganized data into a central information system [109].

A notable initiative is a generic smart framework for big data integration [112], which seeks to enhance the quality of analytical outcomes by emphasizing data completeness and data veracity within the big data value chain process. A comprehensive review has been conducted in [113] for data integration approaches and techniques. This research revealed the increasing use of integration approaches in query processing between heterogeneous data sources and archives from heterogeneous data sources, demonstrating the importance of the current work. In the domain of industrial IoT applications, a real-time big data integration solution has been proposed in [114], which addresses data heterogeneity produced by IoT devices. The proposed solution manages data extraction, processing, and storage in diverse and heterogeneous repositories. Another notable contribution in [115] proposed a methodology that emphasizes system characteristics for data integration; however, update propagation control across various databases is not specified. The above reference proposes a solution that integrates data using two approaches: one based on a global data schema, where data are integrated across multiple databases using a unified schema, and the latter based on the concept of "peer" networks, where updates are propagated through the network of peers.

Furthermore, [116] emphasizes data integrity through the data migration procedure, introducing the Categorical Query Language (CQL) as an understandable language for data transfer and interaction with complicated schemas. Data streaming integration is not mentioned. The authors of [116] also emphasize the necessity for tools to merge heterogeneous datasets. A framework was proposed in [71], which enables data monitoring generated by IoT devices and sensors and their integration with historical data. The proposed approach is founded on the utilization of SQL and aims to enhance the accessibility and utilization of distributed data repositories that possess diverse data models. In addition, the framework enables users to enhance the data generated by Internet of Things (IoT) devices and sensors by seamlessly merging them with pre-existing historical datasets. Several studies have conducted surveys of the most up-to-date methods for data integration to deal with the problems caused by big data [117-121].

Open challenges, as pointed out in [122] include enabling high-velocity data to be processed in real-time with more advanced indexing techniques to make data more readily available for analysis. Reference [117] presents several data integration problems. The first important difficulty is schema heterogeneity, which occurs when various data sources represent the same topic using distinct schemas. The second challenge is data conflicts that can arise because of incomplete data, incorrect data, and out-of-date data. The approach of [123] developed a strategy that employs artificial intelligence technologies to automatically combine vast data volumes from many sources, either structured or unstructured. The suggested framework intends to analyze data based on their metadata to verify data similarity and feasible level of integration. Its structure is adaptable because it employs integration modules, which facilitate the maintenance, deployment, and integration of new data models as needed. To integrate heterogeneous big data, several approaches can be utilized, including data warehousing, data mapping, and data federation [124]. The data warehouse and data federation were the two dominant options [125].

4.1. Big Data warehouses

Big Data Warehouses (BDW) represent an advancement compared to Data Warehousing (DW). The big data era necessitates structures and innovative technologies to handle the overwhelming amount and diversity of data. Consequently, it has become increasingly important to complement DW to tackle the growing challenges posed by the volume, variety, velocity, and veracity of big data [126]. BDW is an adopted approach, for consolidating big data sources. It includes establishing a repository for data from multiple sources converted and loaded. This strategy is useful for combining data from several sources into a database that can be easily accessed and analyzed. The main reasons for switching to BDW are the traditional DW's significant limitations of high cost and limited efficiency in massive data storage and query analysis. To facilitate the integration of heterogeneous data from many sources, it is crucial to build a strong distributed data warehouse platform [127]. Because traditional DWs are better suited for organized and historical data analytics and have trouble expanding horizontally, it is quite challenging to ensure this [128]. Implementing a BDW system needs careful consideration of data modeling, complex mapping, and complex transformation procedures, making it a time-consuming and expensive procedure. The authors of [129, 130] have developed a customized BDW architecture designed to manage heterogeneous data for effective big data processing as a solution to these problems. BDWs can easier grow horizontally and can analyze data in real-time as shown in [131, 132]. Reference [133] suggested a drought data management architecture based on spatial-temporal BDW. To load data into Hadoop systems, Apache Flume was used in parallel to expedite data ingestion and increase overall system efficiency. Reference [134] proposed a BDW architecture for supporting big data analysis, the proposed architecture being capable of automatically or semi-automatically adapting to changes in requirements or data expansion. According to [126], the utilization of big data technologies and tools such as Hadoop, Apache Spark, Data Lake, and Delta Lake in a complimentary manner can effectively enhance and support existing DW systems. This integration not only enhances scalability but also contributes to cost reduction in the construction of conventional DW architectures. The concept of a data warehouse is a well-established and mature management paradigm that is supported by widely accepted methodologies. In contrast, the field of big data is still in a developmental stage, with several approaches attempting to handle certain aspects of the problem. However, a comprehensive and integrated solution for big data is yet to be fully realized. A summary of the comparison between BDW and traditional DW is shown in Table 3.

Table 3. Differences between big data warehouse and traditional data warehouse

Characteristics	BDW	Traditional DW
Data Types	It accepts as input data that is semi-structured, unstructured, or structured	It accepts only structured data as input
Architecture	The processing of big data is accomplished through the utilization of a distributed file system	The utilization of distributed file systems for processing is not employed
Tools	Massive amounts of data can be handled with Apache Hadoop	Database Management Systems (DBMS), Structured Query Language (SQL), Extract, Transform and Load (ETL) tools
Data Governance	Emerging; often less mature	Well-established
Data Sources	IoT devices, social media, logs, traditional databases	Traditional databases, operational systems
Complexity	Advanced analytics, machine learning, real-time analytics	Business intelligence, querying, reporting

4.2. Big Data federation

Big Data Federation (BDF) refers to the process of combining and analyzing data from sources spread out in locations to create a unified view, for advanced analytics and decision-making [135]. Unlike methods that involve consolidating data into one place, BDF aims to keep the data in its sources while still making it easily accessible for exploration and analysis across those sources. This approach is particularly useful when privacy, security, and control are concerns for data owners as it avoids the need for storage, which can be challenging or undesirable [136]. Data federation serves as a promising tool on its own or in conjunction with data integration. It effectively addresses the challenge of accessing diverse data sources by mapping them into a single schema like RDF(S)/OWL ontology or a relational schema. This unified schema allows queries such as SPARQL or SQL queries to run over it [135].

Moreover, modern data management systems often incorporate tools for federated query answering [137]. The main goal of federated query answering is to create a consistent way of accessing data, from sources without duplicating them in a central repository. This is achieved by using sub-queries that target the data sources within the federation and evaluating their results based on predefined rules. It is important to note that data federation across heterogeneous big data sources is an active field in both research and industry. However, there is still a need for a foundation and established principles for data federation systems [135].

In reference [138] a method is introduced to combine data from sources using Spark, Presto, and the OBDA framework. This allows queries in SPARQL, which utilizes ontology terms to consistently access types of data. This approach is commonly known as Ontology Based Data Access (OBDA) [138]. Additionally, reference [139] presents FEDSA, a data federation solution designed for query requirements, in law enforcement scenarios. It facilitates the gathering and exploration of information. The authors of [140] propose a framework that focuses on analyzing data in the context of the Internet of Things (IoT). This approach takes into account the network of IoT systems, each one with its unique data model.

4.3. Data warehousing versus data federation in heterogeneous Big Data

BDW and BDF are two unique techniques for dealing with massive amounts of data, each one with its own set of characteristics and advantages. BDW, which is designed to manage enormous datasets, excels at scalability, with horizontal and vertical extension options to handle data growth. Big data warehousing features real-time analytics which facilitates data-based decision making. Nevertheless, this process is complex and demanding as it requires the integration of data from various sources, huge data transformations, and modeling to ensure data integrity and robust governance, which entails significant efforts and additional implementation costs. On the other hand, big data federation reduces the need to transform data as it takes data from multiple sources to create a virtual data picture without physical integration. This cuts the cost of infrastructure because it is virtual and does not require storage.

Table 4. A comparison between BDW and BDF

Aspect	BDW	BDF
Scalability	Scales horizontally and vertically	Typically scales horizontally
Real-time analytics	Supports real-time data analysis	Suitable for batch and near real-time analysis
Data integration	Physically consolidates data	Creates virtual views of data
Data transformation	Requires significant transformation and modeling	Requires minimal data transformation
Data quality and governance	Emphasizes data quality and governance	Less emphasis on data quality and governance
Cost	Can be expensive due to infrastructure and resource requirements	Generally, more cost-efficient in terms of infrastructure
Complex queries	Handles complex queries efficiently	May face challenges with complex queries involving data from multiple sources

Table 5. Advantages and disadvantages of BDW and BDF

Approach	Description	Pros.	Constraints
Data warehousing	Data from many sources are processed and placed into a centralized repository	Provides a trustworthy data analysis and decision-making resource; Can expand horizontally with ease; Can perform real-time data analysis	Time-consuming and expensive to implement; Requires significant resources for data modeling and transformation
Data federation	Creating a representation of information obtained from different sources without physically combining the data	Less costly than data warehousing; minimal data modeling and transformation is required	Challenging when performing complex queries involving data from different sources

Despite its flexibility and reduced cost, it may experience some difficulty in executing queries that require pulling data from various sources. Organizations need to choose between these two data management strategies based on their respective goals, and budgetary limitations. Additionally, they have to consider the required type of Big Data, including infrastructure requirements, real-time analytics capabilities, scalability, integration, complexity, and implementation costs. A comparison between the two strategies, data warehousing, and data federation, against a set of relevant dimensions, is presented in Tables 4 and 5 [132, 141, 142].

5. Analysis and findings

The survey focuses on how organizations handle types of data and the challenges they face in the process. It highlights the importance of methods for integrating data with formats, structures, and content. Additionally, it discusses two approaches to managing data: data warehousing and data federation. Both approaches have their strengths and limitations. The goal of this survey is to provide an understanding of the complexities and challenges involved in handling Big Data for experts, researchers, practitioners, and decision-makers in the field of data management. This paper emphasizes that modern organizations must develop capabilities for managing Big Data to replace inefficient traditional methods. It also stresses the significance of utilizing up-to-date technology, tools, sound governance practices, and skilled professionals in managing big datasets. Furthermore, it sheds light on characteristics of Big Data including speed, volume, accuracy and trustworthiness (veracity) diversity (variety), and value.

The paper outlines challenges related to processing and analyzing datasets while highlighting their potential benefits such as enhancing decision-making processes discovering patterns/opportunities and making an impact, across various business sectors. Moreover, the research emphasizes the importance of having knowledge, skills, expertise, and the right tools to manage datasets. It presents approaches, to data integration that aim to combine types of data from multiple sources and formats to create a comprehensive and unified perspective. The document also sheds light on the challenges associated with data integration and consolidation such as mapping schemas, merging data, and linking records. It discusses two methods for managing data: BDW and BDF. The BDW approach involves integrating data from sources into a repository while employing modeling, mapping, and transformation techniques. Although this method requires time investment and additional expenses it enables scalability and real-time analytics. On the other hand, the BDF approach allows querying and analysis by creating a virtual view without physically relocating the original data. Additionally, it addresses concerns regarding privacy, security, and ownership related to storage infrastructure. Finally, when we compare BDW with BDF approaches we can see their advantages and drawbacks. Although BDW provides scalability and real-time analytics capabilities it requires effort in terms of data transformation and infrastructure development. On the other hand, the BDF option also requires effort in data transformation. Comes with minimal infrastructure costs. However, it may face challenges when dealing with data queries. Organizations make their decision on which approach to use based on factors such as resources, specific integration needs the level of complexity in data queries required, and considerations, for data security and privacy. By considering these factors organizations can make informed decisions to effectively manage data and improve the decision-making process.

6. Conclusion

This survey addresses the challenges and opportunities associated with managing heterogeneous big data in modern organizations. Heterogeneous big data consists of a combination of various data types, in different formats, from different sources, which poses challenges in analyzing and integrating such data. The paper suggests two approaches for heterogeneous big data integration: data warehousing and data federation. It also points out relevant data attributes such as velocity, volume, veracity, variety, and value and cites examples of the heterogeneity of data sources such as sensor data, social media, and healthcare information. It thoroughly analyzes the integration approaches mentioned earlier. This survey paper provides an investigation into the management of varied big data. It offers insights into the complexities and possibilities that come with handling heterogeneous big data. It is a resource for researchers, professionals, and decision-makers who wish to understand and make the most of the challenges and opportunities presented by managing heterogeneous big data, for informed decision-making and achieving business success.

References

1. Wang, L. Heterogeneous Data and Big Data Analytics. – Automatic Control and Information Sciences, Vol. **3**, 2017, No 1, pp. 8-15.
2. Gandomi, A., M. Haider. Beyond the Hype: Big Data Concepts, Methods, and Analytics. – International Journal of Information Management, Vol. **35**, 2015, No 2, pp. 137-144.
3. Baig, M. I., L. Shuib, E. Yadegaridehkordi. Big Data Adoption: State of the Art and Research Challenges. – Information Processing & Management, Vol. **56**, 2019, No 6, 102095.
4. Elgendy, N., A. Elragal. Big Data Analytics: A Literature Review Paper. – In: Proc. of 14th Industrial Conference Advances in Data Mining Applications and Theoretical Aspects (ICDM'14), St. Petersburg, Russia, 16-20 July 2014, Proceedings 14, 2014.
5. Venkatram, K., M. A. Geetha. Review on Big Data & Analytics – Concepts, Philosophy, Process and Applications. – Cybernetics and Information Technologies, Vol. **17**, 2017, No 2, pp. 3-27.
6. Hariri, R. H., E. M. Fredericks, K. M. Bowers. Uncertainty in Big Data Analytics: Survey, Opportunities, and Challenges. – Journal of Big Data, Vol. **6**, 2019, No 1, pp. 1-16.
7. Sun, J., W. Gan, Z. Chen, J. Li, P. S. Yu. Big Data Meets Metaverse: A Survey. – arXiv Preprint arXiv:2210.16282, 2022.
8. Zhao, X. A Study on e-Commerce Recommender System Based on Big Data. – In: Proc. of 4th IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA'19), 2019.
9. Yan, B., C. Wu, R. Yu, B. Yu, N. Shi, X. Zhou, Y. Yu. Big Data-Based e-Commerce Transaction Information Collection Method. – Complexity, Vol. **2021**, 2021, pp. 1-11.
10. Ghani, N. A., S. Hamid, I. A. T. Hashem, E. Ahmed. Social Media Big Data Analytics: A Survey. – Computers in Human Behavior, Vol. **101**, 2019, pp. 417-428.

11. N a e e m, M., T. J a m a l, J. D i a z-M a r t i n e z, S. A. B u t t, N. M o n t e s a n o, M. I. T a r i q, E. D e-l a-H o z-F r a n c o, E. D e-L a-H o z-V a l d i r i s. Trends and Future Perspective Challenges in Big Data. – In: Proc. of Advances in Intelligent Data Analysis and Applications: 2019, Arad, Romania, 2022.
12. P r a m a n i k, S., S. K. B a n d y o p a d h y a y. Analysis of Big Data. – In: Encyclopedia of Data Science and Machine Learning. IGI Global, 2023, pp. 97-115.
13. A l z o u b i, Y. I., A. A l j a a f r e h. Blockchain-Fog Computing Integration Applications: A Systematic Review. – Cybernetics and Information Technologies, Vol. **23**, 2023, No 1, pp. 3-37.
14. D a i, H.-N., H. W a n g, G. X u, J. W a n, M. I m r a n. Big Data Analytics for Manufacturing Internet of Things: Opportunities, Challenges and Enabling Technologies. – Enterprise Information Systems, Vol. **14**, 2020, No 9-10, pp. 1279-1303.
15. C h e n, H., R. H. C h i a n g, V. C. S t o r e y. Business Intelligence and Analytics: From Big Data to Big Impact. – MIS Quarterly, 2012, pp. 1165-1188.
16. G r o v e r, P., A. K. K a r. Big Data Analytics: A Review on Theoretical Contributions and Tools Used in Literature. – Global Journal of Flexible Systems Management, Vol. **18**, 2017, pp. 203-229.
17. S m i t h, G. J., L. B e n n e t t M o s e s, J. C h a n. The Challenges of Doing Criminology in the Big Data Era: Towards a Digital and Data-Driven Approach. – The British Journal of Criminology, Vol. **57**, 2017, No 2, pp. 259-274.
18. S e s t i n o, A., M. I. P r e t e, L. P i p e r, G. G u i d o. Internet of Things and Big Data as Enablers for Business Digitalization Strategies. – Technovation, Vol. **98**, 2020, 102173.
19. M o r a w i e c, P., A. S o ł t y s i k-P i o r u n k i e w i c z. Cloud Computing, Big Data, and Blockchain Technology Adoption in Erp Implementation Methodology. – Sustainability, Vol. **14**, No 7, 2022, 3714.
20. S a a d o o n, M., S. H. A. H a m i d, H. S o f i a n, H. H. A l t a r t u r i, Z. H. A z i z u l, N. N a s u h a. Fault Tolerance in Big Data Storage and Processing Systems: A Review on Challenges and Solutions. – Ain Shams Engineering Journal, Vol. **13**, 2022, No 2, 101538.
21. M e h m o o d, H., E. G i l m a n, M. C o r t e s, P. K o s t a k o s, A. B y r n e, K. V a l t a, S. T e k e s, J. R i e k k i. Implementing Big Data Lake for Heterogeneous Data Sources. – In: Proc. of 35th IEEE International Conference on Data Engineering Workshops (ICDEW'19), 2019.
22. Y o u n a s, M. Research Challenges of Big Data. – Service Oriented Computing and Applications, Vol. **13**, 2019, pp. 105-107.
23. T o m m a s i n i, R., P. B o n t e, F. S p i g a, E. D e l l a V a l l e. Taming Variety and Velocity. – In: Streaming Linked Data: From Vision to Practice. Springer, 2022, pp. 41-67.
24. M o h a m e d, A., M. K. N a j a f a b a d i, Y. B. W a h, E. A. K. Z a m a n, R. M a s k a t. The State of the Art and Taxonomy of Big Data Analytics: View from New Big Data Framework. – Artificial Intelligence Review, Vol. **53**, 2020, pp. 989-1037.
25. K o l a j o, T., O. D a r a m o l a, A. A d e b i y i. Big Data Stream Analysis: A Systematic Literature Review. – Journal of Big Data, Vol. **6**, 2019, No 1, 47.
26. A r o r a, Y., D. G o y a l. Review of Data Analysis Framework for Variety of Big Data. – In: Proc. of Emerging Trends in Expert Applications and Security (ICETEAS'18), 2019.
27. W a t s o n, H. J. Update Tutorial: Big Data Analytics: Concepts, Technology, and Applications. – Communications of the Association for Information Systems, Vol. **44**, No 1, 2019, 21.
28. L u, J., I. H o l u b o v á. Multi-Model Databases: A New Journey to Handle the Variety of Data. – ACM Computing Surveys (CSUR), Vol. **52**, 2019, No 3, pp. 1-38.
29. B a t i n i, C., A. R u l a, M. S c a n n a p i e c o, G. V i s c u s i. From Data Quality to Big Data Quality. – Journal of Database Management (JDM), Vol. **26**, 2015, No 1, pp. 60-82.
30. D a s, T. K., P. M. K u m a r. Big Data Analytics: A Framework for Unstructured Data Analysis. – International Journal of Engineering Science & Technology, Vol. **5**, 2013, No 1, 153.
31. A d n a n, K., R. A k b a r. Limitations of Information Extraction Methods and Techniques for Heterogeneous Unstructured Big Data. – International Journal of Engineering Business Management, Vol. **11**, 2019, 1847979019890771.
32. O u s s o u s, A., F.-Z. B e n j e l l o u n, A. A. L a h c e n, S. B e l f k i h. Big Data Technologies: A Survey. – Journal of King Saud University-Computer and Information Sciences, Vol. **30**, 2018, No 4, pp. 431-448.

33. S a n d h u, A. K. Big Data with Cloud Computing: Discussions and Challenges. – Big Data Mining and Analytics, Vol. **5**, 2021, No 1, pp. 32-40.
34. C a p p a, F., R. O r i a n i, E. P e r u f f o, I. M c C a r t h y. Big Data for Creating and Capturing Value in the Digitalized Environment: Unpacking the Effects of Volume, Variety, and Veracity on Firm Performance. – Journal of Product Innovation Management, Vol. **38**, 2021, No 1, pp. 49-67.
35. G h a s e m a g h a e i, M. Understanding the Impact of Big Data on Firm Performance: The Necessity of Conceptually Differentiating among Big Data Characteristics. – International Journal of Information Management, Vol. **57**, 2021, 102055.
36. S h u k l a, A. K., M. Y a d a v, S. K u m a r, P. K. M u h u r i. Veracity Handling and Instance Reduction in Big Data Using Interval Type-2 Fuzzy Sets. – Engineering Applications of Artificial Intelligence, Vol. **88**, 2020, 103315.
37. C a v a n i l l a s, J. M., E. C u r r y, W. W a h l s t e r. The Big Data Value Opportunity. – New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe, 2016, pp. 3-11.
38. G a n d o m i, A. H., F. C h e n, L. A b u a l i g a h. Machine Learning Technologies for Big Data Analytics. – MDPI, Electronics, Vol. **11**, 2022, Issue 3, 421.
39. M o n i n o, J.-L. Data Value, Big Data Analytics, and Decision-Making. – Journal of the Knowledge Economy, Vol. **12**, 2021, pp. 256-267.
40. G a n d o m i, A. H., F. C h e n, L. A b u a l i g a h. Big Data Analytics Using Artificial Intelligence. – MDPI, Electronics, Vol. **12**, 2023, Issue 1, 957.
41. C a s a d o, R., M. Y o u n a s. Emerging Trends and Technologies in Big Data Processing. – Concurrency and Computation: Practice and Experience, Vol. **27**, 2015, No 8, pp. 2078-2091.
42. L i p i c, T., K. S k a l a, E. A f g a n. Deciphering Big Data Stacks: An Overview of Big Data Tools. – Big Data Analytics: Challenges and Opportunities (BDAC-14), 2014.
43. R a o, T. R., P. M i t r a, R. B h a t t, A. G o s w a m i. The Big Data System, Components, Tools, and Technologies: A Survey. – Knowledge and Information Systems, Vol. **60**, 2019, pp. 1165-1245.
44. D r i g a s, A. S., P. L e l i o p o u l o s. The Use of Big Data in Education. – International Journal of Computer Science Issues (IJCSI), Vol. **11**, No 5, 2014, 58.
45. S i n, K., L. M u t h u. Application of Big Data in Education Data Mining and Learning Analytics – A Literature Review. – ICTACT Journal on Soft Computing, Vol. **5**, 2015, No 4, pp. 1035-1049.
46. K l a š n j a-M i l i ć e v i ć, A., M. I v a n o v i ć, Z. B u d i m a c. Data Science in Education: Big Data and Learning Analytics. – Computer Applications in Engineering Education, Vol. **25**, 2017, No 6, pp. 1066-1078.
47. B e e r k e n s, M. An Evolution of Performance Data in Higher Education Governance: A Path towards a "Big Data" Era? – Quality in Higher Education, Vol. **28**, 2022, No 1, pp. 29-49.
48. S h a h, T. H. Big Data Analytics in Higher Education. – Research Anthology on Big Data Analytics, Architectures, and Applications, 2022, pp. 1275-1293.
49. W a n g, L., R. C h e n, S. L i, Y. M a. Development Planning Path of Integration of Education, Production, and Education from the Perspective of Big Data Information Fusion. – In: Proc. of 3rd International Symposium on Computer Engineering and Intelligent Communications (ISCEIC'22), 2023.
50. G u o, C., J. C h e n. Big Data Analytics in Healthcare. – In: Knowledge Technology and Systems: Toward Establishing Knowledge Systems Science. Springer, 2023, pp. 27-70.
51. V e s o u l i s, Z. A., A. N. H u s a i n, F. S. C o l e. Improving Child Health through Big Data and Data Science. – Pediatric Research, Vol. **93**, No 2, 2023, pp. 342-349.
52. B e n z i d i a, S., O. B e n t a h a r, J. H u s s o n, N. M a k a o u i. Big Data Analytics Capability in Healthcare Operations and Supply Chain Management: The Role of Green Process Innovation. – Annals of Operations Research, 2023, pp. 1-25.
53. A l-D m o u r, H., N. S a a d, E. B a s h e e r A m i n, R. A l-D m o u r, A. A l-D m o u r. The Influence of the Practices of Big Data Analytics Applications on Bank Performance: Filed Study. – VINE Journal of Information and Knowledge Management Systems, Vol. **53**, 2023, No 1, pp. 119-141.

54. W a h y u d i, M., V. M e i l i n d a, A. K h o i r u n i s a. The Digital Economy's Use of Big Data. – International Transactions on Artificial Intelligence, Vol. **1**, 2022, No 1, pp. 62-70.
55. R a z z a q, A., X. Y a n g. Digital Finance and Green Growth in China: Appraising Inclusive Digital Finance Using Web Crawler Technology and Big Data. – Technological Forecasting and Social Change, Vol. **188**, 2023, 122262.
56. L u t f i, A., M. A l r a w a d, A. A l s y o u f, M. A. A l m a i a h et al. Drivers and Impact of Big Data Analytic Adoption in the Retail Industry: A Quantitative Investigation Applying Structural Equation Modeling. – Journal of Retailing and Consumer Services, Vol. **70**, 2023, 103129.
57. A r u l k u m a r, V., S. S r i d h a r, G. K a l p a n a, K. G u r u p r a k a s h. Real-Time Big Data Analytics for Improving Sales in the Retail Industry via the Use of Internet of Things Beacons. – In: Proc. of Expert Clouds and Applications (ICOECA'22), Springer, 2022, pp. 111-126.
58. K a s t o u n i, M. Z., A. A. L a h c e n. Big Data Analytics in Telecommunications: Governance, Architecture and Use Cases. – Journal of King Saud University-Computer and Information Sciences, Vol. **34**, 2022, No 6, pp. 2758-2770.
59. M o u m t z i d i s, I., M. K a m a r i o t o u, F. K i t s i o s. Digital Transformation Strategies Enabled by Internet of Things and Big Data Analytics: The Use-Case of Telecommunication Companies in Greece. – Information, Vol. **13**, 2022, No 4, 196.
60. L i, Q., J. S u n. Research on the Characteristics of Inbound Tourism in Yangtze River Delta Region Based on Big Data. – In: Proc. of International Conference on Cloud Computing, Big Data and Internet of Things (3CBIT'22), 2022.
61. B u i, V., A. R. A l a e i, H. Q. V u, G. L i, R. L a w. Revisiting Tourism Destination Image: A Holistic Measurement Framework Using Big Data. – Journal of Travel Research, Vol. **61**, 2022, No 6, pp. 1287-1307.
62. A l j o j o, N. Examining Heterogeneity Structured on a Large Data Volume with Minimal Incompleteness. – ARO-The Scientific Journal of Koya University, Vol. **9**, 2021, No 2, pp. 30-37.
63. A l m a'a i t a h, W. Z. a., A. Z. T a l i b, M. A. O s m a n. Opportunities and Challenges in Enhancing Access to Metadata of Cultural Heritage Collections: A Survey. – Artificial Intelligence Review, Vol. **53**, 2020, pp. 3621-3646.
64. K u m a r, G., S. B a s r i, A. A. I m a m, S. A. K h o w a j a, L. F. C a p r e t z, A. O. B a l o g u n. Data Harmonization for Heterogeneous Datasets: A Systematic Literature Review. – Applied Sciences, Vol. **11**, 2021, No 17, 8275.
65. A s p r i n o, L., E. D a g a, A. G a n g e m i, P. M u l h o l l a n d. Knowledge Graph Construction with a Façade: A Unified Method to Access Heterogeneous Data Sources on the Web. – ACM Transactions on Internet Technology, Vol. **23**, No 1, 2023, pp. 1-31.
66. M a h a r a n a, K., S. M o n d a l, B. N e m a d e. A Review: Data Pre-Processing and Data Augmentation Techniques. – Global Transitions Proceedings, Vol. **3**, 2022, No 1, pp. 91-99.
67. W a n i, M. A., S. J a b i n. Big Data: Issues, Challenges, and Techniques in Business Intelligence. – In: Proc. of Big Data Analytics: Proceedings of Computer Society of India, 2015, 2018.
68. P a s t o r i n o, R., C. D e V i t o, G. M i g l i a r a, K. G l o c k e r, I. B i n e n b a u m, W. R i c c i a r d i, S. B o c c i a. Benefits and Challenges of Big Data in Healthcare: An Overview of the European Initiatives. – European Journal of Public Health, Vol. **29**, 2019, No Supplement_3, pp. 23-27.
69. J a v e d, M. H. Challenges and Requirements of Heterogenous Research Data Management in Environmental Sciences: A Qualitative Study. Master's Thesis, University of Oulu, Information Processing Science, 2023.
70. W e n z, V., A. K e s p e r, G. T a e n t z e r. Clustering Heterogeneous Data Values for Data Quality Analysis. – ACM Journal of Data and Information Quality, 2023.
71. A m a r á, J., V. S t r ö e l e, R. B r a g a, M. D a n t a s, M. B a u e r. Integrating Heterogeneous Stream and Historical Data Sources Using SQL. – Journal of Information and Data Management, Vol. **13**, No 2, 2022.
72. K u m a r, A., S. R. S a n g w a n, A. N a y y a r. Multimedia Social Big Data: Mining. – Multimedia Big Data Computing for IoT Applications: Concepts, Paradigms and Solutions, 2020, pp. 289-321.

73. K r i s h n a m u r t h i, R., A. K u m a r, D. G o p i n a t h a n, A. N a y y a r, B. Q u r e s h i. An Overview of IoT Sensor Data Processing, Fusion, and Analysis Techniques. – Sensors, Vol. **20**, 2020, No 21, 6076.
74. B a l a k r i s h n a, S., M. T h i r u m a r a n, V. K. S o l a n k i. A Framework for IoT Sensor Data Acquisition and Analysis. – EAI Endorsed Transactions on Internet of Things, Vol. **4**, 2018, No 16, pp. e4-e4.
75. V a l d e o l m i l l o s, D., Y. M e z q u i t a, A. R. L u d e i r o. Sensing as a Service: An Architecture Proposal for Big Data Environments in Smart Cities. – In: Proc. of 0th International Symposium on Ambient Intelligence – Software and Applications, Vol. **1**, 2020.
76. G o l d s t e i n, I., C. S. S p a t t , M. Y e. Big Data in Finance. – The Review of Financial Studies, Vol. **34**, 2021, No 7, pp. 3213-3225.
77. C h e n g, X., S. L i u, X. S u n, Z. W a n g, H. Z h o u, Y. S h a o, H. S h e n. Combating Emerging Financial Risks in the Big Data Era: A Perspective Review. – Fundamental Research, Vol. **1**, 2021, No 5, pp. 595-606.
78. A g r a w a l, R., S. P r a b a k a r a n. Big Data in Digital Healthcare: Lessons Learnt and Recommendations for General Practice. – Heredity, Vol. **124**, 2020, No 4, pp. 525-534.
79. P i s e, A. A. An Introduction to Big Data Analytics Techniques in Healthcare. – In: S. K. Ohal, S. K. Pani, S Prasad, S. K. Mohapatra, Eds. Big Data Analytics and Machine Intelligence in Biomedical and Health Informatics: Concepts, Methodologies, Tools and Applications, 2022, pp. 1-20.
80. G o n č a r o v s, P. Data Analytics in Crm Processes: A Literature Review. – Information Technology and Management Science, Vol. **20**, 2017, No 1, pp. 103-108.
81. C i m i l u c a, K. Honors Thesis Marketing: Dr. Lutz Customer Relationship Management 1. Introduction of Customer Relationship Management. 2010.
82. K u n z, W., L. A k s o y, Y. B a r t, K. H e i n o n e n, S. K a b a d a y i, F. V. O r d e n e s, M. S i g a l a, D. D i a z, B. T h e o d o u l i d i s. Customer Engagement in a Big Data World. – Journal of Services Marketing, Vol. **31**, 2017, No 2, pp. 161-171.
83. B u t t l e, F., S. M a k l a n. Customer Relationship Management: Concepts and Technologies. – Routledge, 2019.
84. D a r v a z e h, S. S., I. R. V a n a n i, F. M. M u s o l u. Big Data Analytics and Its Applications in Supply Chain Management. – New Trends in the Use of Artificial Intelligence for the Industry, Vol. **4**, 2020, 175.
85. G a r c i a-A r r o y o, J., A. O s c a. Big Data Contributions to Human Resource Management: A Systematic Review. – The International Journal of Human Resource Management, Vol. **32**, 2021, No 20, pp. 4337-4362.
86. V e r m a, S., V. S i n g h, S. S. B h a t t a c h a r y y a. Do Big Data-Driven Hr Practices Improve Hr Service Quality and Innovation Competency of SMEs. – International Journal of Organizational Analysis, Vol. **29**, 2021, No 4, pp. 950-973.
87. X i a o, J. Iot Applications in the Innovation System of Enterprise Human Resource Performance Management Based on the Integration of Big Data. – Wireless Communications and Mobile Computing, Vol. **2022**, 2022.
88. C h e r g u i, N., M. T. K e c h a d i. Data Analytics for Crop Management: A Big Data View. – Journal of Big Data, Vol. **9**, 2022, No 1, pp. 1-37.
89. S u n, A. Y., B. R. S c a n l o n. How Can Big Data and Machine Learning Benefit Environment and Water Management: A Survey of Methods, Applications, and Future Directions. – Environmental Research Letters, Vol. **14**, 2019, No 7, 073001.
90. C h e r g u i, N., M.-T. K e c h a d i, M. M c D o n n e l l. The Impact of Data Analytics in Digital Agriculture: A Review. – In: Proc. of International Multi-Conference on Organization of Knowledge and Advanced Technologies (OKTA'20), 2020.
91. K s h e t r i, N., D. C. R o j a s T o r r e s, H. B e s a d a, M. A. M o r o s O c h o a. Big Data as a Tool to Monitor and Deter Environmental Offenders in the Global South: A Multiple Case Study. – Sustainability, Vol. **12**, 2020, No 24, 10436.
92. D a n i e l, B. K. Big Data in Higher Education: The Big Picture. – Big Data and Learning Analytics in Higher Education: Current Theory and Practice, 2017, pp. 19-28.
93. A n g, K. L.-M., F. L. G e, K. P. S e n g. Big Educational Data & Analytics: Survey, Architecture, and Challenges. – IEEE Access, Vol. **8**, 2020, pp. 116392-116414.

94. D a h d o u h, K., A. D a k k a k, L. O u g h d i r, F. M e s s a o u d i. Big Data for Online Learning Systems. – Education and Information Technologies, Vol. **23**, 2018, pp. 2783-2800.
95. C h e n, N.-S., C. Y i n, P. I s a i a s, J. P s o t k a. Educational Big Data: Extracting Meaning from Data for Smart Education. – Taylor & Francis, 2020, pp. 142-147.
96. Z h u, L., F. R. Y u, Y. W a n g, B. N i n g, T. T a n g. Big Data Analytics in Intelligent Transportation Systems: A Survey. – IEEE Transactions on Intelligent Transportation Systems, Vol. **20**, 2018, No 1, pp. 383-398.
97. Ka f f a s h, S., A. T. N g u y e n, J. Z h u. Big Data Algorithms and Applications in Intelligent Transportation System: A Review and Bibliometric Analysis. – International Journal of Production Economics, Vol. **231**, 2021, 107868.
98. A h m e d, S., Y. A b d e l-H a m i d, H. A. H e f n y. Traffic Flow Prediction Using Big Data and GIS: A Survey of Data Sources, Frameworks, Challenges, and Opportunities. – International Journal of Computing and Digital Systems, Vol. **13**, 2023, No 1, p. 1.
99. P u l i k o t t i l, T., L. A. E s t r a d a-J i m e n e z, J. J. P. A b a d í a, A. C a r r e r a-R i v e r a, A. T o r a y e v, H. U. R e h m a n, F. M o, S. N i k g h a d a m-H o j j a t i, J. B a r a t a. Big Data Life Cycle in Shop-Floor – Trends and Challenges. – IEEE Access, 2023.
100. W a n g, J., C. X u, J. Z h a n g, R. Z h o n g. Big Data Analytics for Intelligent Manufacturing Systems: A Review. – Journal of Manufacturing Systems, Vol. **62**, 2022, pp. 738-752.
101. X i a, C., Z. P a n, J. P o l d e n, H. L i, Y. X u, S. C h e n, Y. Z h a n g. A Review on Wire Arc Additive Manufacturing: Monitoring, Control and a Framework of Automated System. – Journal of Manufacturing Systems, Vol. **57**, 2020, pp. 31-45.
102. B e l h a d i, A., K. Z k i k, A. C h e r r a f i, M. Y. S h a'r i. Understanding Big Data Analytics for Manufacturing Processes: Insights from Literature Review and Multiple Case Studies. – Computers & Industrial Engineering, Vol. **137**, 2019, 106099.
103. P h u y a l, S., D. B i s t a, R. B i s t a. Challenges, Opportunities and Future Directions of Smart Manufacturing: A State of Art Review. – Sustainable Futures, Vol. **2**, 2020, 100023.
104. K a d a d i, A., R. A g r a w a l, C. N y a m f u l, R. A t i q. Challenges of Data Integration and Interoperability in Big Data. – In: Proc. of IEEE International Conference on Big Data (Big Data'14), 2014.
105. S i d d i q a, A., I. A. T. H a s h e m, I. Y a q o o b, M. M a r j a n i, S. S h a m s h i r b a n d, A. G a n i, F. N a s a r u d d i n. A Survey of Big Data Management: Taxonomy and State-of-the-Art. – Journal of Network and Computer Applications, Vol. **71**, 2016, pp. 151-166.
106. S a m u e l s e n, J., W. C h e n, B. W a s s o n. Integrating Multiple Data Sources for Learning Analytics – Review of Literature. – Research and Practice in Technology Enhanced Learning, Vol. **14**, 2019, pp. 1-20.
107. B i z e r, C., P. B o n c z, M. L. B r o d i e, O. E r l i n g. The Meaningful Use of Big Data: Four Perspectives – Four Challenges. – ACM Sigmod Record, Vol. **40**, 2012, No 4, pp. 56-60.
108. R o h, Y., G. H e o, S. E. W h a n g. A Survey on Data Collection for Machine Learning: A Big Data-Ai Integration Perspective. – IEEE Transactions on Knowledge and Data Engineering, Vol. **33**, 2019, No 4, pp. 1328-1347.
109. M h a m m e d i, S., N. G h e r a b i. Heterogeneous Integration of Big Data Using Semantic Web Technologies. – In: Intelligent Systems in Big Data, Semantic Web and Machine Learning, Springer, 2021, pp. 167-177.
110. M e r e l l i, I., H. P é r e z-S á n c h e z, S. G e s i n g, D. D'A g o s t i n o. Managing, Analysing, and Integrating Big Data in Medical Bioinformatics: Open Problems and Future Perspectives. – In: BioMed Research International. Vol. **2014**. 2014.
111. S i v a R a m a R a o, A. V., R. D h a n a L a k s h m i. A Survey on Challenges in Integrating Big Data. – In: Proc. of 2nd International Conference on Intelligent Computing and Applications (ICICA'15), 2017.
112. Y o u s f i, S., D. C h i a d m i, M. R h a n o u i. Smart Big Data Framework for Insight Discovery. – Journal of King Saud University – Computer and Information Sciences, Vol. **34**, 2022, No 10, pp. 9777-9792.
113. D e S o u z a C a m p o s, V. V., J. D. B r a n c h e r, F. P. F a r i a s, J. L. V. M. M i o n i, P. L. G. B r a h i m. Review and Comparison of Works on Heterogeneous Data and Semantic Analysis in Big Data. – Semina: Ciências Exatas e Tecnológicas, Vol. **42**, 2021, No 1, pp. 113-128.

114. C o s t a, F. S., S. M. N a s s a r, S. G u s m e r o l i, R. S c h u l t z, A. G. C o n c e i ç ã o, M. X a v i e r, F. H e s s e l, M. A. D a n t a s. Fasten Iiot: An Open Real-Time Platform for Vertical, Horizontal and End-to-End Integration. – Sensors, Vol. **20**, 2020, No 19, 5499.
115. A s a n o, Y., D.-F. H e r r, Y. I s h i h a r a, H. K a t o, K. N a k a n o, M. O n i z u k a, Y. S a s a k i. Flexible Framework for Data Integration and Update Propagation: System Aspect. – In: Proc. of IEEE International Conference on Big Data and Smart Computing (BigComp'19), 2019.
116. B r o w n, K. S., D. I. S p i v a k, R. W i s n e s k y. Categorical Data Integration for Computational Science. – Computational Materials Science, Vol. **164**, 2019, pp. 127-132.
117. N o a m a n, M., M. S. K h a n, M. F. A b r a r, S. A l i, A. A l v i, M. A. S a l e e m. Challenges in Integration of Heterogeneous Internet of Things. – In: Scientific Programming. Vol. **2022**. 2022.
118. D o n g, X. L., D. S r i v a s t a v a. Big Data Integration. – In: Proc. of 29th IEEE International Conference on Data Engineering (ICDE'13), 2013.
119. F a t h y, N., W. G a d, N. B a d r. A Unified Access to Heterogeneous Big Data through Ontology-Based Semantic Integration. – In: Proc. of 9th International Conference on Intelligent Computing and Information Systems (ICICIS'19), 2019.
120. A h a m e d, B., T. R a m k u m a r. Data Integration-Challenges, Techniques, and Future Directions: A Comprehensive Study. – Indian Journal of Science and Technology, Vol. **9**, 2016, No 44, pp. 1-9.
121. H e n d l e r, J. Data Integration for Heterogenous Datasets. – Big Data, Vol. **2**, 2014, No 4, pp. 205-215.
122. N a r g e s i a n, F., E. Z h u, R. J. M i l l e r, K. Q. P u, P. C. A r o c e n a. Data Lake Management: Challenges and Opportunities. – Proceedings of the VLDB Endowment, Vol. **12**, 2019, No 12, pp. 1986-1989.
123. S a e s, K. R. Abordagem Para Integração Automática de Dados Estruturados E Não Estruturados Em Um Contexto Big Data. – Universidade de São Paulo, 2018.
124. J a y a s h r e e, G., C. P r i y a. Data Integration with Xml Etl Processing. – In: Proc. of International Conference on Computer Science, Engineering and Applications (ICCSEA'20), 2020.
125. H u, H., Y. W e n, T.-S. C h u a, X. L i. Toward Scalable Systems for Big Data Analytics: A Technology Tutorial. – IEEE Access, Vol. **2**, 2014, pp. 652-687.
126. Z g u r o v s k y, M. Z., Y. P. Z a y c h e n k o. Big Data: Conceptual Analysis and Applications. Springer, 2020.
127. W u, Y., J. W u. Research on Decision Support System of Automobile Service Based on Distributed Data Warehouse. – In: Proc. of 2nd International Conference on Digital Society and Intelligent Systems (DSInS'22), 2023.
128. A r s h a d, M., M. N. B r o h i, T. R. S o o m r o, T. M. G h a z a l, H. M. A l z o u b i, M. A l s h u r i d e h. NoSQL: Future of Big Data Analytics Characteristics and Comparison with Rdbms. – In: The Effect of Information Technology on Business and Marketing Intelligence Systems, Springer, 2023, pp. 1927-1951.
129. S a d d a d, E., A. E l-B a s t a w i s s y, H. M. M o k h t a r, M. H a z m a n. Lake Data Warehouse Architecture for Big Data Solutions. – International Journal of Advanced Computer Science and Applications, Vol. **11**, No 8, 2020.
130. S i n d h u, C., N. P. H e g d e. Handling Complex Heterogeneous Healthcare Big Data. – International Journal of Computational Intelligence Research, Vol. **13**, 2017, No 5, pp. 1201-1227.
131. D i T r i a, F., E. L e f o n s, F. T a n g o r r a. Evaluation of Data Warehouse Design Methodologies in the Context of Big Data. – In: Proc. of 19th International Conference Big Data Analytics and Knowledge Discovery, France, 28-31 August 2017, Proceedings 19, 2017.
132. D h a o u a d i, A., K. B o u s s e l m i, M. M. G a m m o u d i, S. M o n n e t, S. H a m m o u d i. Data Warehousing Process Modeling from Classical Approaches to New Trends: Main Features and Comparisons. – Data, Vol. **7**, 2022, No 8, 113.
133. B a l t i, H., N. M e l l o u l i, A. B e n A b b e s, I. R. F a r a h, Y. S a n g, M. L a m o l l e. Enhancing Big Data Warehousing and Analytics for Spatio-Temporal Massive Data. – In: Track 2: Methodologies and Education, 2021.
134. S o l o d o v n i k o v a, D., L. N i e d r i t e. An Approach to Handle Big Data Warehouse Evolution. – arXiv preprint arXiv:1809.04284, 2018.

135. G u, Z., F. C o r c o g l i o n i t i, D. L a n t i, A. M o s c a, G. X i a o, J. X i o n g, D. C a l v a n e s e. A Systematic Overview of Data Federation Systems. – Semantic Web, No Preprint, 2022, pp. 1-59.
136. B o l o g a, A.-R., R. B o l o g a. A Perspective on the Benefits of Data Virtualization Technology. – Informatica Economica, Vol. **15**, 2011, No 4.
137. B o u a r a r, S., L. B e l l a t r e c h e, S. J e a n, M. B a r o n. Do Rule-Based Approaches Still Make Sense in Logical Data Warehouse Design? – In: Proc. of East European Conference on Advances in Databases and Information Systems, 2014.
138. M a m i, M. N., D. G r a u x, S. S c e r r i, H. J a b e e n, S. A u e r, J. L e h m a n n. Squerall: Virtual Ontology-Based Access to Heterogeneous and Large Data Sources. – In: Proc. of 18th International Semantic Web Conference (ISWC'19), Auckland, New Zealand, 26-30 October 2019, Proceedings, Part II 18, 2019.
139. L i, W., Z. F e n g, W. M a y e r, G. G r o s s m a n n, A. K. K a s h e f i, M. S t u m p t n e r. Fedsa: A Data Federation Platform for Law Enforcement Management. – In: Proc. of 22nd IEEE International Enterprise Distributed Object Computing Conference (EDOC'18), 2018.
140. K a l a m a r a s, I., N. K a k l a n i s, K. V o t i s, D. T z o v a r a s. Towards Big Data Analytics in Large-Scale Federations of Semantically Heterogeneous IoT Platforms. – In: Proc. of International Workshops Artificial Intelligence Applications and Innovations: AIAI 2018 IFIP WG 12.5, SEDSEAL, 5G-PINE, MHDW, and HEALTHIOT, Rhodes, Greece, 25-27 May 2018, Proceedings 14, 2018.
141. S a g g a r, S., C. B i t o n i, I. K h u r a n a, R. A l h a w a t. Data Warehouse with Big Data Technology for Higher Education. – Available at SSRN 4128707, 2022.
142. S h r e s t h a, S., S. B h a l l a. Survey on the Evolution of Models of Data Integration. – Int. J. Knowl. Based Comput. Syst, Vol. **8**, 2020, pp. 11-16.

Received: 02.10.2023; Second Version: 04.01.2023; Third Version: 18.02.2024; Accepted: 26.02.2024