**Годишник на секция "Информатика"**
**Съюз на учените в България**
**ISSN 1313–6852**
**Том 10, 2019 / 2020, 81–105**

**Annual of "Informatics" Section**
**Union of Scientists in Bulgaria**
**ISSN 1313–6852**
**Volume 10, 2019 / 2020, 81–105**

# INDEX MATRICES AS A TOOL FOR DATA LAKEHOUSE MODELLING

**Veselina Bureva**

Laboratory of Intelligent Systems, "Prof. Dr. Assen Zlatarov" University,
1 "Prof. Yakimov" Blvd., Burgas 8010, Bulgaria
e-mail: vbureva@btu.bg

**Abstract:** The aim of the current research work is to present the possibilities for Big Data mathematical modeling using the apparatus of Index Matrices. The theory of index matrices is discussed. The capabilities of index matrix operations are investigated. An overview of data repositories and analytical tools are in the years is presented. In a series of papers index matrix operation over data warehouses are defined. Here, the authors present an overview of index matrix operations over data storing and processing processes.
**Keywords:** Big Data, Business Intelligence, Data Analytics, Data Lake, Data Lakehouse, Data Science, Data Warehouse, Index Matrices, Intuitionistic Fuzzy Sets, Operations.

## 1. Introduction

Index Matrix is a tool for mathematical modeling applied in different types of problems for decision making and optimization. Index matrix has indices on the rows and on the columns. The index matrix can be two-dimensional, three-dimensional or *n*-dimensional. The values of the index matrix can be real number, predicate, intuitionistic fuzzy pair. The definition of the index matrices are given in the Section 2. Series of papers for OLAP operations and data warehouse processes using index matrices are presented in Section 4. In Section 3, a brief review of the data storage development stages is presented. In the current research work the historic overview of Data Warehouses (DW) with OnLine Analytical Processing (OLAP) and their descendants is investigated. In the beginning, the enterprise data warehouses are introduced to provide storage for extracted, cleansed and transformed data. The workflow of the datasets from the data source to the data warehouse is executed by ETL (Extract, Transform, Load). Small parts of the topic data of the organization can be stored in data marts. The Analytical Processing (OLAP) is applied to analyze data and present it in different views of aggregation. The processes of business intelligence are actively supported. Data Warehouses are investigated in the years and three data warehouse approaches are proposed by Inmon, Kimball and Dan Linstedt. In the next stage the data lake is coined in 2011 to provide storage for capabilities for structured, semi-structured and unstructured data and Big Data analysis. Data Lakes are available in the cloud and it helps data scientists by providing access to the raw data. The modern data warehouses combine Data Warehouse and Data Lake. The vendors provide data lakes combining standard components with other features. In the 2020 Databricks defined Data Lakehouse which is optimized for Data Science and Machine Learning and multi-cloud. Data Lakehouse combines Data Lakeabd Data Warehouse in a single platform.

## 2. Short remarks on index matrices

The definition of index matrices is coined by Krassimir Atanassov [28, 42, 45, 46]. The concept of Index Matrix (IM) was discussed in series of papers [15, 18, 19, 22, 157, 162]. Initially they are applied mainly in the area of generalized nets - it is used for describing the transitions [16, 17, 23-27, 37, 48, 51, 77-79, 81-83, 132, 141, 155]. In the years different types of index matrix and their extensions are studied. Their operations, relations and operators are investigated [20, 21, 28, 32, 69-71, 160, 164, 166, 167]. Index matrices with elements index matrices are introduced in [18]. In [157] the extended index matrix (IM) in the case $n = 3$ is coined. Thereafter, in 2018, the $n$-dimensional extended index matrices ($n$-DEIM) are introduced [19]. Data warehouses and data lakehouse OLAP cubes are constructed from three or more dimensions. Therefore the notation of 3-IM has to be extended to $n$-dimensional extended index matrix ($n$-DEIM) to provide the capabilities for multidimensional modelling. Three dimensional intuitionistic fuzzy index matrices (IFIM) [28, 157, 162] are described [31, 44, 50]. They are helpful in intuitionistic fuzzy evaluations in the $n$-dimensional objects. Applications of index matrices [14, 30, 35, 36, 38, 39, 43, 47, 54, 55, 75, 80, 133, 134, 144-146, 159, 161, 163, 168, 169, 174] are developed in different fields of science. They will be explained in Section 4 of the current research work.

## 3. Short remarks on Data Warehouse (DW), Data Lake (DL) and Data Lakehouse (DLH)

Traditional DW are constructed as centralized data storage containing information from different data sources as: operational databases or files. In DW the input information is structured so it is optimized to store relational data. In the beginning, according to the Inmon approach, DW use star schema and construct data marts after the stage of data warehousing. Data mart contains appropriate type of information for exact needs of the department. DW can have many data marts with different type of information. Kimball improves the Inmon method with adding dimensionality and normalization of the dimensional tables (3NF – 3 normal form). In Kimball approach the data marts are constructed in the data warehouse before its processing. Thereafter Data Vault 1.0 and Data Vault 2.0 are introduced to provide storage for structured (the data is organized - RDBMS), semi-structures (the data is organized in non-fixed format - JSON) and unstructured data (unorganized data –audio, videos). The information for loading is extracted from relational databases, NoSQL databases, web logs, sensors and etc. The DW is extended to provide capabilities for Big Data technologies. Data Lake is introduced to store unstructured data for Big Data processing. The Big Data analysis allows processing techniques: batch processing and stream processing. These needs lead to the modern data warehouse which integrates Data Warehouse with Data Lake. Thereafter the data virtualization appears to enable the work with data stored in different data sources. The differences between DW, DL and DLH are visually explained by the imagepresented on the website of databricks data lakehouse [89].

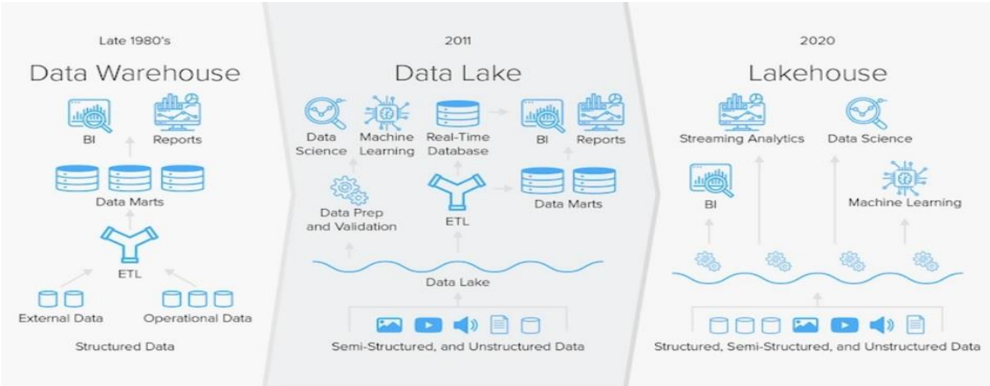



Fig. 1. Databricks Data Lakehouse

### 3.1. The beginning: short remarks on data warehouse

Data warehouses are constructed using three types of architecture depending on the tiers: single-tier architecture, two-tier architecture, three-tier architecture. Generally the three-tier architecture is used in the companies where the bottom tier contains the database server used to extract data from different data sources, the middle tier is OLAP server and the top tier is the client layer where the user interacts with data: data analysis, querying reporting and data mining. The data extraction, cleaning and transformation are usually processed using ETL (Extract, Transform, Load) services. ETL is an automated procedure which extracts data from various sources, transform it in appropriate form and load the datasets in data warehouse, data hub or data lake.

The DW contains integrated granular historical data. Three authors present approaches for data warehouse construction: Bill Inmon (CIF, DW2.0) [108-110], Ralph Kimball [113-114] (Kimball Group Method) and Dan Linstedt (Data Vault) [125]. In the next part of the section a brief overview of data warehouses is presented. The Inman and Kimbal models are considered to be the first methodologies while Dan Linstedt model introduces a method combining the previous two approaches (3 NF and star schema) and gives the flexibility to extend the storage including Big Data technologies and NoSQL Databases [3] to provide Modern Data Warehouse. Standard ETL process is executed in the staging area of the Inmon data warehouse approach: first the data is collected from different sources, second: the information is transformed and then the datasets are loaded into data warehouse. Depending on the need of the user different types of methods for extracting (full extraction, partial extraction without update notification, partial extraction with update notification), transforming (cleaning, filtering, transforming, joining, splitting, sorting, data validation) and loading (initial load, incremental load, full refresh) are used. The ELT process is applied in the staging area of Kimball data warehouse approach. The data is extracted, loaded into data marts and then the datasets are transformed for adding them to data warehouse [53].

The data warehouses are modelled using different types of schemas for data organization: star schema, snowflake schema and hybrid schema. The data schemes are constructed using fact tables and dimensional tables. Fact table contain measures for the OLAP cube and foreign keys of the dimensional tables. Dimensional tables present attributes of the hierarchy for appropriate object. Snowflake schema is an extension of star schema and uses additional dimensional tables to present the normalized data structure or Entity-Relationship (ER) diagram. Snowflake schema contains fully expanded hierarchies. Star schema contains one fact table in the center of the star and multiple dimensional tables.Star schema contains fully collapsed hierarchies. The dimensional tables can be denormalized. The queries execution and the cube processing are faster in the case of star schema. Galaxy schema or Fact Constellation Schema contains two fact table that share dimension tables between them. The data warehouses are implemented separately from the operational databases of the organizations [1, 96, 147].

Two types of system are frequently used for data warehousing: OLTP and OLAP**.**The OLTP (OnLine Transactional Processing) performs mainly read and write operation to database and are used for data source of data warehouse. The OLTP transactions are short, atomic, structured and repetitive. It selects a detailed, up-to-date data. The second, OLAP (OnLine Analytical Processing), presents processing of data by aggregating operations. The information is presented in detail or in aggregated view. The granularity presents the data in different level of summarization. OLAP is used to execute complex queries. The aim of the queries is to present the data in different views: detailed view, aggregated view, sub-view [74, 88, 137, 143, 170-172, 175, 178].

### 3.1.1. Inmon's data warehouse approach

The term Data Warehouse was introduced by Bill Inmon in 1990. According to Inman a data warehouse is "*a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process*". Bill Inmon is known as the "*Father of Data Warehousing*". He introduces the Corporate Information Factory (CIF) approach for data warehouse [108-110]. The normalized data model is used. It uses an Entity Relationship Model to model the data warehouse. The methodology of Bill Inmon is top-down approach. The data at the greatest level of detail is stored in the data warehouse. The data marts are extracted from data warehouse. The Inmon approach for data warehousing is presented in Fig. 2.

### 3.1.2. Kimball's data warehouse approach

Ralph Kimball is known as the "*Father of Business Intelligence*". He introduces a Dimensional Model to model the data warehouse [113-114]. Ralph Kimball approach is bottom-up design approach. The data marts are previously created to provide reporting and analytical capabilities for specific business processes and then they are integrated to create a data warehouse. The star schema is used. The data is segregated into two segments called 'Facts' and 'Dimensions'. Facts are the measurable data. Dimensions are the attributes which relates to these facts. The Ralph Kimball data warehousing approach is presented in Fig. 3.

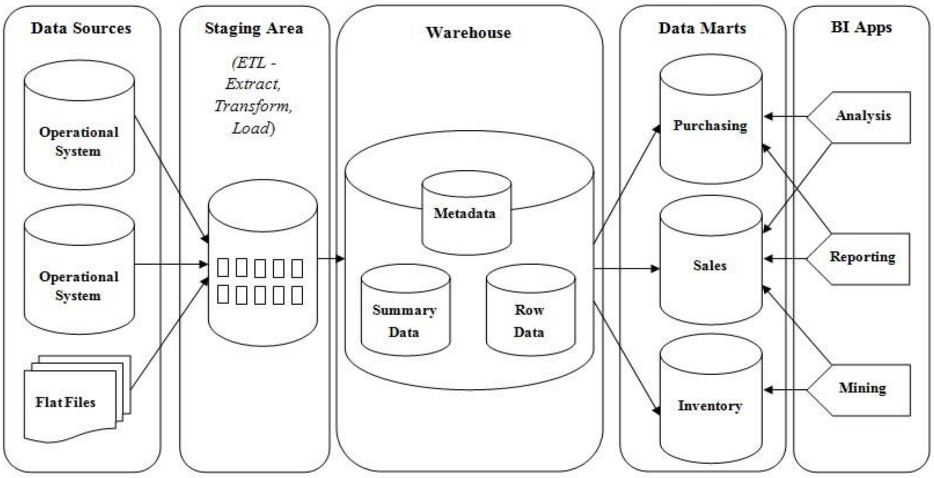



Fig. 2. Corporate Information Factory approach by Inmon

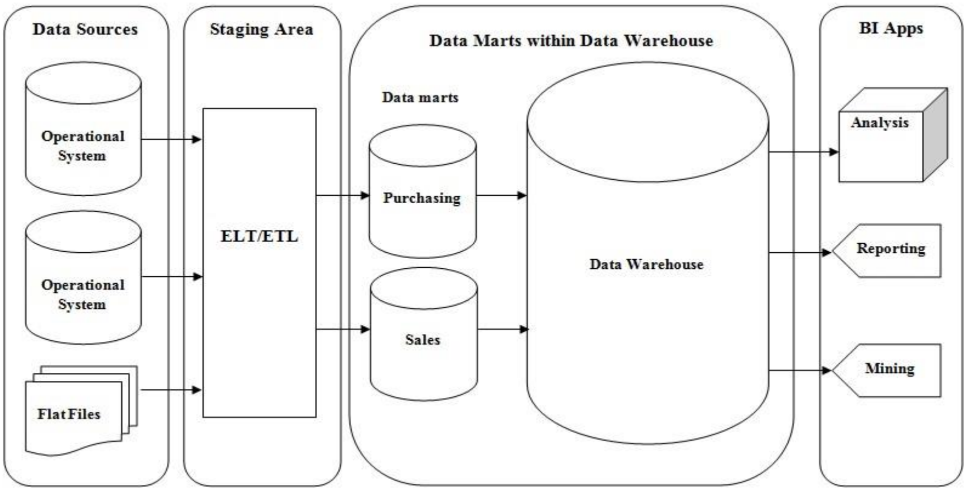



Fig. 3. Dimensional modeling approach by Kimball

### 3.1.3. Dan Linstedt data warehouse approach and Big Data technologies: Data Vault 1.0 and Data Vault 2.0

Dan Linstedt is the creator of a *Data Vault* model. Data Vault 1.0 is introduced by Linstedt in 1990. It combines Kimball and Inmon models.Dan Linstedt describes a resulting Data Vault 1.0 database as: "*A detail oriented, historical tracking and uniquely linked set of normalized tables that support one or more functional areas of business. It is a hybrid approach encompassing the best of breed between 3NF and Star Schemas. The design is flexible,* *scalable, consistent and adaptable to the needs of the enterprise.*" Data Vault 1.0 is mainly focused on the modeling: logical and physical data models for building the raw data warehouse. Data Vault 2.0 is proposed by Linstedt in 2013 as an extension of Data Vault 1.0. Data Vault 2.0 has the aim to improve performance and scalability by changing the model to interact with NoSQL and Big data systems [105]. It support relational databases and NoSQL databases. The replacement of the sequence number with hash key is considered for the main difference of Data Vault 1.0 and Data Vault 2.0. The Dan Linstedt data warehousing approach is presented in Fig. 4.

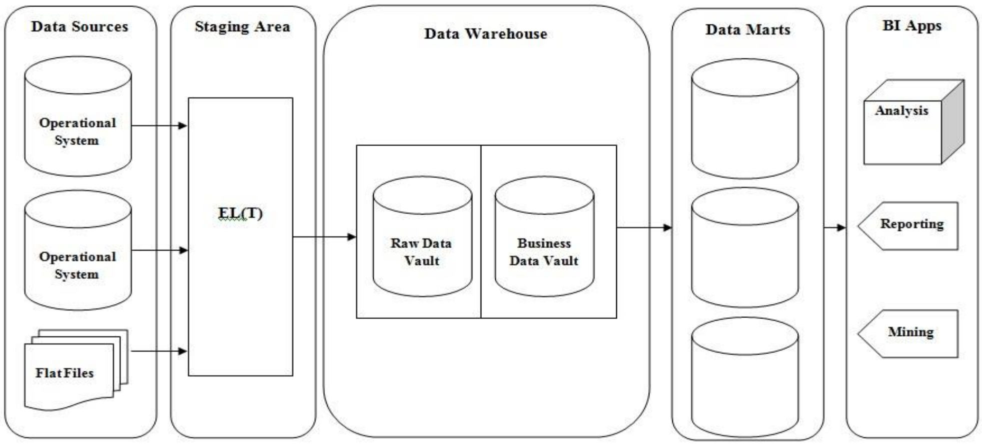



Fig. 4. Data Vault approach by Dan Linstedt

Raw data vault contains all raw data while Business data vault contains some preprocessed data. The Data Vault model is presented in detail in [125]. Here the authors do not discuss in detail the methodology of Data Vault architecture. The attention will be concentrated to the benefits whose the new architecture provides. The appearance of Data Vault 2.0 methodology enables parallel loading as much as possible which provides Big Data implementations. The new storages incorporate Hadoop, traditional data warehouse, data lakes receiving data from mobile devices, sensors, cloud and the Internet of Things. The Modern Data Warehouse supports variety of data sources, coexists with Data Lake, coexists with Hadoop [149], use Massively Parallel Processing (MPP) for operation on high volumes of data across distributed nodes. It supports data integration, data virtualization, in-memory structures, advanced analytics, flexible deployment and etc. The modernized data architecture as data lakehouse combines data lake and data warehouse as a single data management platform that serves data for both BI and data science. The concept of Data Lake and Data Lakehouse is discussed on the next section.

### 3.2. The era of Big Data

Data Vault is introduced due to the need to process large volumes of data in a structured and unstructured form. The Big Data methodology is presented around 1990 as a technology that has the aim to perform data storage, data analysis, search, sharing, transfer, visualization, querying, updating, information privacy of large amounts of data [75, 106, 111, 182]. In the same time Google introduced MapReduce as framework for processing data and the scalable distributed file system Google File System (GFS). MapReduce framework is considered as highly scalable according to the fact that it has a parallelized processing layer that distributes the data across multiple commodity machines [66, 76]. Google File System has good performance, availability, scalability and reliability. By 2004 Yahoo introduce an open source implementation of GFS as Nutch Distributed Filesystem (NDFS). NDFS file system allows large files to be set in the system. NDFS is fault tolerant and available. The development of these technologies leads to the introduction of Hadoop 1.0 which provides storage and processing framework. In 2008 Hadoop 2.0 is presented which provides new features [7]. YARN (Yet Another Resource Negotiator) is implemented to provide resource allocation according to the operations within the cluster. These are the three core elements of Hadoop ecosystem. Data lake interacting can be provided using the tools like Apache Pig [11], Apache Spark[12] and Apache Hive [9], Python. In addition to the technologies related to the Hadoop ecosystem Spark stands out for its determining role in the evolution of Big Data. Spark is a distributed data processing engine that can handle vast amount of data. Spark is evolution of Hadoop MapReduce providing advantages as working with in memory data, real-time processing,extension modules allowing Machine Learning, streaming and etc., working with different programming languages [112]. Depending of the company and of the provider of the Big Data implementation different architectures can be implemented. The interconnected systems provide the Ecosystem. Commercial vendors offered parallel database management systems for big data. The most popular big data ecosystems are Apache Hadoop [7], Apache Spark [12], Apache Flume [6], Microsoft Azure [129], Amazon Web Services (AWS) [2], Facebook ecosystem, LinkedIn ecosystem using Apache Kafka [10], Google ecosystem [100], Twitter ecosystem, Alibaba ecosystem. Big data architecture can include one or more of the following blocks (Fig. 5).

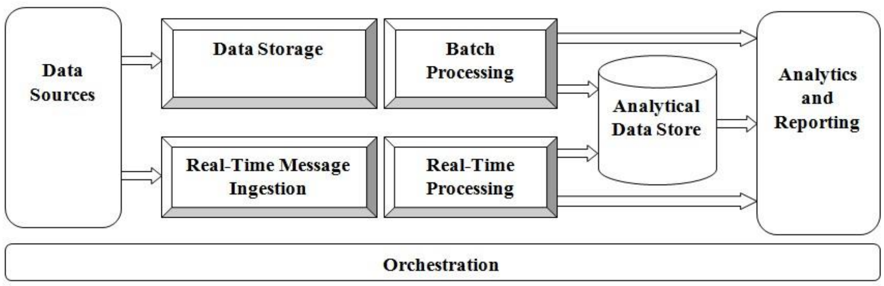



Fig. 5. Big Data architecture

### 3.3. Big data and Data Lake: Big Data Lake

The increasing popularity of the Big data provides rapid development of the Big Data ecosystem for data management. James Dixon, founder of Pentaho, present the term “Data

Lake" in 2010 as improved storage against to data mart, which is a smaller repository of interesting attributes derived from raw data. A data lake is a central location with schema-on-read architecture which stored a large amount of data in its native format. Data Lake storage can contain structured, semi-structured and unstructured/binary data taken from multiple sources and lacking a predefined schema. Structured data includes relational databases. Semi-structured data includes CSV, logs, XML, JSON files. Unstructured data is presented by emails, documents, PDFs. Binary data presents image,weblogs, audio and video files. Data lakes are flexible and adaptable to changes and can easily to be expanded through the scaling of its servers. Data Lakes stores the "valuable data". They provide to Data Scientists access to the raw information for their future investigations to consider if this data to be processed or not. A data lake can be established "on premises" or "in the cloud". In the first case the data lake is within an organization's data centers. In the second case Data Lake is established using cloud services from vendors such as Amazon [2], Microsoft [129], Google [100], Hortonworks [104], Oracle [131], Zaloni [179], Teradata [156], MongoDB [182]. Data lake tools and frameworks include Apache Spark [12], Apache Flink [5], Databriks [89], Amazon Web Services' Simple Storage Service [2], Microsoft Azure [129], Presto [138], Impetus [93]. The architecture of Enterprise Data Lake [103] is presented in the Fig.6. In the data acquisition layer the input data is collected and integrated. Thereafter the data is send to the processing layer for analytical layer building. It is possible to define the schema and structure for raw data if it is needed. Thereafter the data are preceded to the data consumption layer. Then the processes of predictive learning, business analytics, data science models and data visualizations are performed.

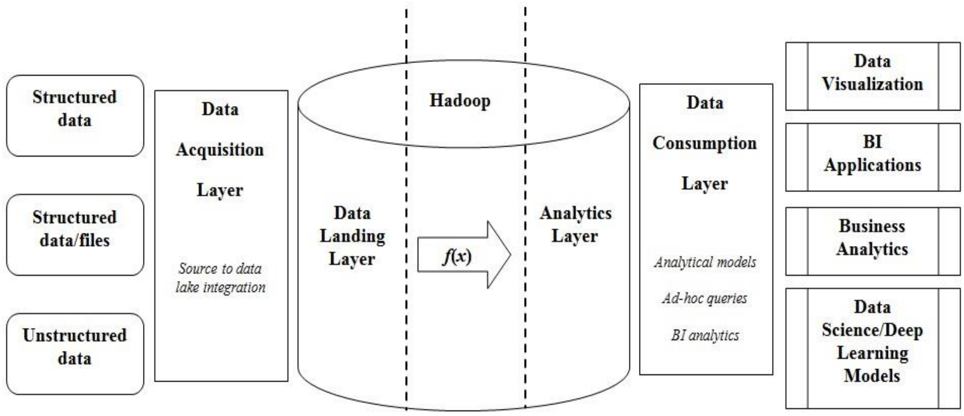



Fig. 6. Enterprise Data Lake Architecture

Let us notice the difference between Data Lake and Data Stamp.Data Swamp is represented as poorly organized Data Lake. Data swamp is a data lake providing little value: it provides poor management of data and it can be is inaccessible to its users. Data Lake for one organization can be a Data Swamp for another company. The Data Lakes become data swamps according to the lack of structure and governance. Data Lake establishes where large amount of different types of data can be stored estimated and analyzed. Data Lakes help Data Scientists to discover and analyze information performing minimal transformation over it with the aim to provide automated pattern identification [87, 124, 126].

## 3.4. Data Lakehouse: integrating Data Warehouse and Data Lake

Data Lakehouse is coined by Databriks in 2020 as a new data platform combining the data warehouse and data lake [13]. The storage contains the benefits of the two technologies and reduces its disadvantages.Data Lakes are great for data storage but data warehouses are organized. Data Warehouses were created to support Business Intelligence while the Data Lake was conceptualized to use data for data science and machine learning. Data Lakehouse is suitable for data analysts and for data scientists.In Data Lakehouse the redundancy of Data is removed. Data governance management is done easily. Data Lakehouse supports analysis of structured and unstructured data using tools for direct connection to BI tools as the open-source SQL query engine for Big Data exploration Apache Drill [4]. Data Lakehouse is based on open direct-access data formats, such as Apache Parquet [7]. Data Lakehouses have the schema support, executesdata reading and writing, end to end streaming, transaction support and high performance SQL. Data Lakehouse enable metadata layers for data lakes, new query engine designs and optimized access for data science and machine learning. The data flow of data lake is presented in the Fig.7.

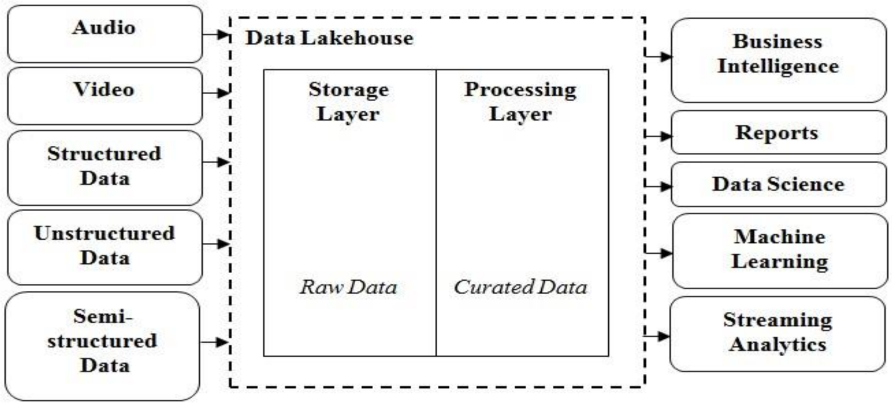
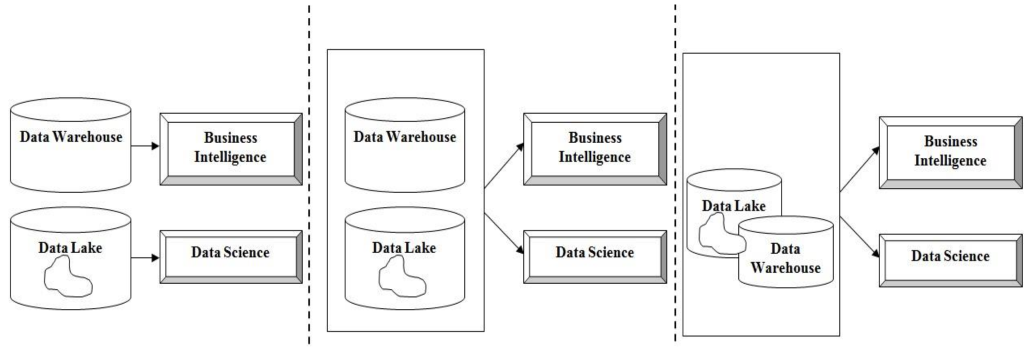



Fig.7. High-level Data LakeHouse data flow

Data Lakehouse tools as Google BigQuery [100], Apache Drill [4], Amazon Athena [2], Delta Lake [91] are available. BigQuery enables the Data Lakehouse concept by providing complete abstraction of the storage and computation layer.Apache Drill is the Schema-free distributed query engine. Athena is an interactive serverless query engine. Delta Lake is coined by the founders of Spark and Databricks. It is the open-source Data LakeHouse enables the ACID methodology on the Distributed storage. The notation of data warehouse, data lake and data lakehouse described up to now is presented graphically in Fig. 8 [184]. In the beginning data warehouses (DW) and data lakes (DL) are used separately for business intelligence tasks and data science processes. Thereafter the modernized data architecture is introduced to combine data lake and data warehouse as a single data management platform for both business intelligence and data science [148].

Data Lakehouse use data virtualization approach to merge data warehouse and data lake. Data virtualization is a logical data layer that integrates and manages all enterprise data to presents a real-time access to the users. It provides execution of distributed data management processing against multiple heterogeneous data sources. Data Virtualization executes data extract, transform and integrate virtually while ETL performs these tasks physically on data with transformation engines. Fig. 9 shows a brief history of data repositories [183].




a) Separate DW and DL b) Combined DW and DL c) Data Lakehouse

Fig. 8. Data Warehouse, Data Lake, Data Lakehouse

Data Lakehouse use data virtualization approach to merge data warehouse and data lake. Data virtualization is a logical data layer that integrates and manages all enterprise data to presents a real-time access to the users. It provides execution of distributed data management processing against multiple heterogeneous data sources. Data Virtualization executes data extract, transform and integrate virtually while ETL performs these tasks physically on data with transformation engines. Fig. 9 shows a brief history of data repositories [183].

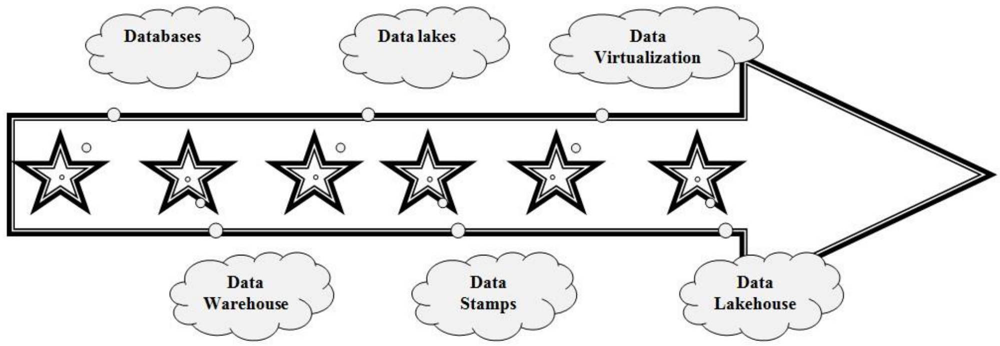



Fig. 9. Journey from databases to data lakehouse

# 4. Index matrices as tool for Data Lakehouse modelling

The authors present a summarized review of the research works related to index matrices as tool for mathematical modeling in the area of business intelligence processes, OLAP cube, data mining, Big Data computational frameworks and Data Science investigations.

## 4.1. Index matrices as tool for mathematical modelling of data lakehouse processes: Business intelligence and data science overview

Index matrix interpretations of the relational operations known as selection, projection and join are discussed in [54]. In the next papers the analytical processing of multidimensional data by OLAP cube operations are investigated. In the literature the OLAP operations are separated depending on different point of views. The main operations are the operations processing the content of an OLAP cube and operations between two cubes are based on the set operations. Operations processing the content of the OLAP cube can be separated into two subgroups: granular operations and extraction operations. Granular operations navigate in the levels of the hierarchies of the dimensions. Granularity is a hierarchisation of information at various levels detail (granularity levels). Roll-up groups and aggregate the data in the selected levels and drill-down do the opposite. Dice operation is a slice on more than two dimensions of a data cubes, i.e. it selects a sub-cube. Rotate operation gives a two-dimensional view of each side of the OLAP cube. The operations roll-up, drill-down, slice, dive and pivot are investigated by index matrix in [70, 164]. Data cube operation is extended roll-up (drill-down) operation [102, 101]. It groups and aggregates the data and adds the totals as the sum of the data by each row and each column. Slice operation performs projection according to a dimension of the cube, i.e. the result is a two-dimensional table [167]. The operations between OLAP cubes include the set operations and the join operation. The union operation combines the two input sets into one output set. The intersection operation returns the tuples that exist in both input sets. The difference operation returns tuples of the first set that do not exist in the second set. In series of papers the index matrix operation for OLAP procedures are defined. The join operator is a special case of Cartesian Product operator that is used to relate two cubes having one or more dimensions in common. The definition of index matrix operations performing Data Cube and Interset operations in an OLAP cube are investigated in [167]. Drill-across operation concatenates the measures of two OLAP cubes with the same dimensions [71]. New index matrix operations as multiplication, automatic reduction applied over data cube are introduced in [69, 20].

OLAP is powerful methodology used in different scientific fields to provide summarization and sub-views of the investigated datasets. According to the area of the interest the authors present its extensions. Here, we discuss one of the modifications of the OLAP cube: Graph cube. The authors consider that the graph representations of the OLAP cubes and their description are closely related to the index matrices developments. The index matrix interpretations of graphs operations over them are published [30, 133, 134, 144, 145]. The intuitionistic fuzzy representation of the tree is presented [146, 158].

Index matrices are successfully used in the area of decision making. Two improvements of the algorithm for job appointment are introduced in [55, 159]. Investigation on the index matrices as tool for assessment of human resources is discussed [161]. The managerial decision making by the index matrix operations is presented in [163]. The interpretation by index matrices of the transportation problem is explained [168]. Multiple scenarios of decision making are investigated by index matrices [169] and aggregation of expert value assessments are defined [174].

In the series of papers the neural networks processing is discussed. Index matrix interpretation of multilayer perceptron and index matrix interpretation of one type of extended neural networks are investigated in [35, 36].

## 4.2. Generalized Nets and InterCriteria analysis based on the index Matrices and intuitionistic fuzzy sets: Applications for Data Lakehouse knowledge discovery

The index matrices are used in Generalized nets (GN) for transition condition description. In the current section the GN models of the data mining processes [132, 150, 155] are presented. Data mining is the task performed in the end of the data warehousing process. The users extract stored data and analyze it. The analyses use methods from the area of artificial intelligence and machine learning. Thereafter, in the data lakehouse, the data science performs data mining processes in the parallel architecture. Generalized nets in artificial intelligence [16, 17, 83, 94, 118, 152] and expert system [37] are constructed. GN models of the techniques as association rule discovery [57, 62], decision trees [64], neural networks [119, 152], support vector machine [63], sequential pattern mining [68], genetic algorithms [141], cluster analysis [58-61, 67] and etc. are constructed. The GN models for hierarchal neural network [33], intuitionistic fuzzy feed forward neural network [34], multilayer neural network [121], early stopping method for parallel optimization of multilayer perceptron [120] are published. Implementation of GN models of feed forward networks is presented [99]. The GNs for biometric identification are constructed [56, 65, 98]. The data warehousing processes, operation and database deadlocks are investigated in [81-87, 116, 117]. Big Data MapReduce paradigm is modeled in [66].

InterCriteria analysis (ICA) is a new approach based on the index matrices and intuitionistic fuzzy sets [40]. ICA extensions are defined [49, 52, 72, 143, 165, 173]. Applications of the ICA are performed in the area of university rankings [122], economics, medicine [123], genetic algorithms [95, 135, 140], neural networks [151, 153, 154] and etc. using the software [107, 128].

## 5. Conclusion

In the current research work, the author presents a historical review of data repositories from data warehouse to data lakehouse. The brief summary of OLAP Analysis, Big Data methodology and Data Mining methods is presented. The published research works of index matrix interpretations in the area of OLAP, methods for decision making and neural networks are discussed. The OLAP operations by index matrices are investigated. GN models of data mining processes are presented. In the future research works the authors will continue to define data lakehouse processes by index matrix oerations. Data Science methods will be modelled. MapReduce index matrix interpretation will be presented.

## Acknowledgements

The authors are thankful for the support provided by the European Regional Development Fund and the Operational Program "Science and Education for Smart Growth" under contract UNITe No. BG05M2OP001-1.001-0004-C01 (2018-2023).

## References

[1] Alzebdi, M., Chountas, P., Atanassov, K.T., Enhancing DWH models with the utilisation of multiple hierarchical schemata. In: Proceeding of the 2010 IEEE International Conference on Systems Man and Cybernetics (SMC), Istanbul, 10-13 Oct. 2010, 488-492.

[2] Amazon Athena: interactive query service - AWS, Amazon Athena is an interactive query service, https://aws.amazon.com/athena.

[3] Apache Cassandra: Open Source NoSQL Database, https://cassandra.apache.org/.

[4] Apache Drill - Schema-free SQL for Hadoop, NoSQL and Cloud Storage, https://drill.apache.org/.

[5] Apache Flink — Stateful Computations over Data Streams, https://flink.apache.org/.

[6] Apache Flume: distributed, reliable, and available service for efficiently collecting, aggregating, and moving large amounts of log data, https://flume.apache.org/.

[7] Apache Hadoop, https://hadoop.apache.org/.

[8] Apache Hbase: Hadoop database, https://hbase.apache.org/.

[9] Apache Hive data warehouse software, https://hive.apache.org/.

[10] Apache Kafka: open-source distributed event streaming platform, https://kafka.apache.org/.

[11] Apache Pig: platform for analyzing large data sets, https://pig.apache.org/.

[12] Apache Spark: Unified Analytics Engine for Big Data, https://spark.apache.org/.

[13] Armbrust, M., Ghodsi, A., Xin, R., Zaharia, M., Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics.Conference on Innovative Data Systems Research - CIDR '21, Jan. 2021, Online, http://cidrdb.org/cidr2021/papers/cidr2021_paper17.pdf.

[14] Atanasov, K., Sotirov, S., Multylayer perceptron representation by index matrices with elements in a fixed interval. 2020 European Conference on Circuit Theory and Design (ECCTD), 2020, 1-3, doi: 10.1109/ECCTD49232.2020.9218374.

[15] Atanassov, K., Extended Interval Valued Intuitionistic Fuzzy Index Matrices. In: Atanassov K. et al. (eds) Uncertainty and Imprecision in Decision Making and Decision

Support: New Challenges, Solutions and Perspectives. IWIFSGN 2018. Advances in Intelligent Systems and Computing, vol. 1081, Springer, Cham., 2021, https://doi.org/10.1007/978-3-030-47024-1_1.

[16] Atanassov, K., Generalized Nets in Artificial Intelligence. Vol. 1: Generalized nets and Expert Systems, Ibid., 1998.

[17] Atanassov, K., Generalized Nets in Artificial Intelligence. Vol. 2: Generalized nets and Machine Learning, Ibid., 2000.

[18] Atanassov, K., Index Matrices with Elements Index Matrices. Proceedings of the Jangjeon Mathematical Society, vol. 21, No. 2, 2018, 221-228.

[19] Atanassov, K., *n*-Dimensional Extended Index Matrices. Advanced Studies in Contemporary Mathematics , vol. 21, No. 2, 2018, 245-259.

[20] Atanassov, K., Bureva, V., Four Operations over Extended Intuitionistic Fuzzy Index Matrices and Some of Their Applications. In: Dimov I., Fidanova S. (eds) Advances in High Performance Computing. HPC 2019. Studies in Computational Intelligence, vol. 902, 2021, Springer, Cham.

[21] Atanassov, K., Bureva, V., Operations over 3-Dimensional Extended Intuitionistic Fuzzy Index Matrices.Advanced Studies in Contemporary Mathematics, Vol. 30, No.3, 2020, 335–356.

[22] Atanassov, K., Extended temporal intuitionistic fuzzy index matrices. Notes on Intuitionistic Fuzzy Sets, Vol. 22, No. 5, 2016, 111–121.

[23] Atanassov K., G. Gluhchev, S. Hadjitodorov, A. Shannon, V. Vasilev, Generalized Nets in Image Processing and Pattern Recognition. Proceedings of the Sixth Int. Workshop on Generalized Nets, Sofia, 17 December 2005, 47–60.

[24] Atanassov, K., G. Gluhchev, S. Hadjitodorov, A. Shannon, V. Vassilev. Generalized Nets and Pattern Recognition. KvB Visual Concepts Pty Ltd, Monograph No. 6, Sydney, 2003.

[25] Atanassov K., Generalized Nets as a Tool for the Modelling of Data Mining Processes. In: Sgurev V., Yager R., Kacprzyk J., Jotsov V. (eds) Innovative Issues in Intelligent Systems. Studies in Computational Intelligence, vol 623., 2016, Springer, Cham. https://doi.org/10.1007/978-3-319-27267-2_6.

[26] Atanassov, K., Generalized Nets as Tools for Modelling of Intelligent Systems. Proceedings of the 9th InternationalWorkshop on Generalized Nets (K, Atanassov, A. Shannon, Eds.), Sofia, 4 July 2008, Vol. 1, 1–8.

[27] Atanassov, K., Gluhchev, G., Hadjitodorov, S., Shannon, A., Vasilev, V., Application of Generalized Nets in Biometrics. Comptes Rendus de l'Academie Bulgare des Sciences, Tome 56, No. 5, 2003, 13–18.

[28] Atanassov, K., Index Matrices: Towards an Augmented Matrix Calculus. Springer, Cham, 2014.

[29] Atanassov, K., On Generalized Nets Theory. “Prof. M. Drinov”, Academic Publishing House, Sofia, 2007.

[30] Atanassov, K., On intuitionistic fuzzy graphs and intuitionistic fuzzy relations. Pro-ceedings of the VI IFSA World Congress, Sao Paulo, Brazil, July 1995, Vol. 1, 551–554.

[31] Atanassov, K., On Intuitionistic Fuzzy Sets Theory. Springer, Berlin, 2012.

[32] Atanassov, K., Roeva O., Bureva V., Chesmedjiev P., Vassilev P., Atanassova V., Operators over 3-dimensional index matrices. Proceedings of International Conference on Big Data, Knowledge and Control Systems Engineering(BdKCSE'2019), 2019, Sofia, Bulgaria , 1–5.

[33] Atanassov, K., S. Sotirov, A. Shannon. Generalized Net Model of the Hierarchical Neural Networks. In: Proceedings of the 13th International Workshop on Generalized Nets, London, UK, 29 October 2012, 8–14.

[34] Atanassov, K., Sotirov, S.,Krawszak, M., Generalized Net Model of the Intuitionistic Fuzzy Feed Forward Neural Network. 13th International Conference on Intuitionistic Fuzzy Sets, Sofia, NIFS, Vol. 15, No 2, 2009, 18–23.

[35] Atanassov, K., Sotirov, S., Index matrix interpretation of the Multilayer perceptron. 2013 IEEE INISTA, 2013, 1-3, doi: 10.1109/INISTA.2013.6577637.

[36] Atanassov, K., Sotirov, S., Index matrix interpretation of one type of extended neural networks. International Journal of Reasoning-based Intelligent Systems (IJRIS), Vol. 6, No. 3/4, 2014, 93–97.

[37] Atanassov, K., Sotirova, E., Orozova, D., Generalized Net Model of an Expert System with Frame-type Data Bases. Proceedings of the Jangjeon Mathematical Society, Vol. 9, No. 1, 2006, 91–101.

[38] Atanassov, K., Szmidt, E., Kacprzyk, J., Bureva, V., Two examples for the use of 3-dimensionalintuitionistic fuzzy index matrices. Notes on Intuitionistic Fuzzy Sets, Vol. 20, No. 2, 2014, 52–59.

[39] Atanassov, K., Temporal intuitionistic fuzzy graphs.Notes on Intuitionistic Fuzzy Sets, Vol. 4, No. 4, 1998, 59–61.

[40] Atanassov, K., D. Mavrov, V. Atanassova. Intercriteria Decision Making: A New Approach for Multicriteria Decision Making, Based on Index Matrices and Intuitionistic Fuzzy Sets. Issues in Intuitionistic Fuzzy Sets and Generalized Nets, Vol. 11, 2014, 1–8.

[41] Atanassov, K., E. Szmidt, J. Kacprzyk, On intuitionistic fuzzy pairs. Notes on Intuitionistic Fuzzy Sets, Vol. 19, No. 3, 2013, 1–13.

[42] Atanassov, K., Generalized index matrices. Comptes rendus de l’Academie Bulgare desSciences, Vol. 40, No. 11, 1987, 15–18.

[43] Atanassov, K., Index matrix representation of the intuitionistic fuzzy graph.Fifth Scientific Session of the Mathematical Foundations of Artificial Intelligence Seminar, Sofia, Oct. 5, 1994, Preprint MRL-MFAIS-10-94, 36–41.

[44] Atanassov, K., Intuitionistic Fuzzy Sets. Springer, Heidelberg, 1999.

[45] Atanassov, K., On index matrices, Part 1: Standard cases. Advanced Studies in Contemporary Mathematics, Vol. 20, No. 2, 2010, 291–302.

[46] Atanassov, K., On index matrices, Part 2: Intuitionistic fuzzy case. Proceedings of the Jang-jeon Mathematical Society, Vol. 13, No. 2, 2010, 121–126.

[47] Atanassov, K., On index matrix interpretations of intuitionistic fuzzy graphs. Notes on Intuitionistic Fuzzy Sets, Vol. 8, No. 4, 2002, 73–78.

[48] Atanassov, K., Sotirova, S., Generalized Nets. Academic Publishing House "Prof. Marin Drinov", Sofia, 2017

[49] Atanassov, K., Atanassova, V.,Gluhchev, G.,InterCriteria Analysis: Ideas and problems, Notes on Intuitionistic Fuzzy Sets, Vol. 21, No. 1, 2015, 81–88.

[50] Atanassov, K., Intuitionistic Fuzzy Logics, Springer, Cham, 2017.

[51] Atanassov, K.,Generalized Nets and Intuitionistic Fuzziness in Data Mining. "Prof. Marin Drinov" Academic Publishing House, Sofia, 2020.

[52] Atanassova, V., Doukovska, L., Michalikova, A., Radeva, I., Intercriteria analysis: From pairs to triples. Notes on Intuitionistic Fuzzy Sets, Vol. 22, No. 5, 2016, 98–110.

[53] Blazic, G., Poscic, P., Jaksic, D., Data warehouse architecture classification. 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2017, 1491–1495.

[54] Bureva, V., Application of the Index Matrix Operations Over Relational Databases. Advanced Studies in Contemporary Mathematics, vol . 26, No. 4, 2016, 669–679.

[55] Bureva, V., Sotirova, E, Moskova, M, Szmidt, E., Solving the Problem of appointments using index matrices. Proceedings of the 13th International Workshop on Generalized Nets 29, 2012, 43–48.

[56] Bureva V., Sotirova E., Bozov H., "Generalized Net Model of Biometric Identification Process," 2018 20th International Symposium on Electrical Apparatus and Technologies (SIELA), 2018, 1-4, doi: 10.1109/SIELA.2018.8447104.

[57] Bureva, V., Sotirova, E., Generalized Net of the Process of Association Rules Discovery by Eclat Algorithm Using Weather Databases. 14th International Workshop on Generalized Nets, Burgas, 29–30 November 2013, 1–10.

[58] Bureva, V., Sotirova, E., Atanassov, K., Hierarchical Generalized Net Model of the Process of Clustering. Issues in Intuitionistic Fuzzy Sets and Generalized Nets, Volume 11, 2014, 73–80.

[59] Bureva V., Sotirova E., Atanassov K., Hierarchical Generalized Net Model of the Process of Selecting a Method for Clustering. Proceedings of the 15th International Workshop on Generalized Nets, 2014, 39–48.

[60] Bureva, V., E. Sotirova, S. Popov, D. Mavrov, V. Traneva, Generalized Net of Cluster Analysis Process Using STING: A Statistical Information Grid Approach to Spatial Data Mining. IN: Henning Christiansen, Hélène Jaudoin, Panagiotis Chountas, Troels Andreasen, Henrik Legind Larsen (Eds.): Flexible Query Answering Systems - 12th International Conference, FQAS 2017, London, UK, June 21-22, 2017. Proceedings. Lecture Notes in Computer Science, 10333, Springer, 2017, 239–248.

[61] Bureva, V., Sotirova, E., Popov, S., Mavrov, D., Traneva, V., Generalized Net of Cluster Analysis Process Using STING: A Statistical Information Grid Approach to Spatial Data Mining. Flexible Query Answering Systems, 2017, 239–248.

[62] Bureva, V., Generalized Net of the Process of Association Rules Mining by Frequent Pattern-Growth Method. Annual of Informatics Section Union of Scientists in Bulgaria, Vol. 6, 2013, 46–53 (in Bulgarian).

[63] Bureva, V., Puleva, M., Sotirova, E., Generalized Net of the Process of Support Vector Machines Classifier Construction. Annual of Informatics Section Union of Scientists in Bulgaria, Volume 7, 2014, 28–40.

[64] Bureva, V., Chountas, P., Atanassov, K., A Generalized Net Model of the Process of Decision Tree Construction. 13th Int. Workshop on Generalized Nets, London, 29 October 2012, 1–7.

[65] Bureva V., Yovcheva P., Sotirov S., Generalized Net Model of Fingerprint Recognition with Intuitionistic Fuzzy Evaluations. In: Kacprzyk J., Szmidt E., Zadrożny S., Atanassov K., Krawczak M. (eds) Advances in Fuzzy Logic and Technology 2017. EUSFLAT 2017, IWIFSGN 2017. Advances in Intelligent Systems and Computing, vol 641., 2018, Springer, Cham. https://doi.org/10.1007/978-3-319-66830-7_26, 286–294.

[66] Bureva, V., Popov, S., Sotirova, E., Atanassov, K., Generalized net of MapReduce computational model".In: "Uncertainty and Imprecision in Decision Making and Decision Support: Cross-Fertilization, New Models and Applications. IWIFSGN 2016.Advances in Intelligent Systems and Computing", Atanassov K. et al. (eds), Springer, Vol. 559, 2018, 305–315.

[67] Bureva, V., Popov, S., Traneva, V., Tranev, S., Generalized Net Model of Cluster Analysis using CLIQUE: Clustering in Quest, Biomedecine and Quality of Life – Youth in Science, Int. J. Bioautomation, vol. 23, No 2, 2019, 131–138.

[68] Bureva, V., Sotirova, E., Chountas, P., Generalized Net of the Process of Sequential Pattern Mining by Generalized Sequential Pattern Algorithm (GSP). Intelligent Systems'2014, Advances in Intelligent Systems and Computing (D. Filev et al., Eds), Vol. 323, Springer, Cham.

[69] Bureva, V., Traneva, V., Sotirova, E., Atanassov, K., Index Matrices amd OLAP- cube Part 5: Index Matrix Operations over OLAP-cube.Advanced Studies in Contemporary Mathematics, vol.30, No.1, 2020, 60–88.

[70] Bureva, V., Traneva, V., Sotirova, E., Atanassov, K., Index Matrices And OLAP-Cube Part 2: An presentation of the OLAP-analysis by Index Matrices.Advanced Studies in Contemporary Mathematics, Vol.27, No. 4, 2017, 647–672.

[71] Bureva, V., Traneva, V., Sotirova, E., Atanassov, K., Index matrices and olap-cube part 4: A presentation of the Olap "drill across" operation by index matrices. Advanced Studies in Contemporary Mathematics (Kyungshang), vol. 29, No 1, 2019, 109–123.

[72] Bureva V., Sotirova E., Atanassova V., Angelova N., Atanassov K., Intercriteria Analysis over Intuitionistic Fuzzy Data. In: Lirkov I., Margenov S. (eds) Large-Scale Scientific Computing. LSSC 2017. Lecture Notes in Computer Science, vol 10665. Springer, Cham. https://doi.org/10.1007/978-3-319-73441-5_352018, 333–340.

[73] Buyya, R., Calheiros, R., Dastjerdi, A., Big Data: Principles and Paradigms, Elsevier, 2016

[74] Chaudhuri, S., Dayal, U., An Overview of Data Warehousing and OLAP Technology.SIGMOD, Rec. 26 (1): 1997, 65–74.

[75] Chountas, P., Alzebdi, M., Shannon, A., Atanassov, K., On intuitionistic fuzzy trees. Notes on Intuitionistic Fuzzy Sets, Vol. 15, No. 2, 2009, 30–32.

[76] Chountas, P., Atanassov, K., Atanassova, V., Sotirova, E., Sotirov, S., Roeva, O., Big data, intuitionistic fuzzy sets and MapReduce operators. Notes on Intuitionistic Fuzzy Sets, Vol. 24, No. 2, 2018, 129–135.

[77] Chountas, P., Atanassov, K., Tasseva, V., Kolev, B., Generalized Net Model for Querying Operators over Multiple Intuitionistic Fuzzy Cubes. Developments in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics, Vol. 2, 2008, 21–30.

[78] Chountas, P., Atanassov, K., Sotirova, E., Bureva, V., Generalized Net Model of an Expert System Dealing with Temporal Hypothesis. Flexible Query Answering Systems 2015, Advances in Intelligent Systems and Computing, Vol. 400, 2016, 473–481.

[79] Chountas, P., Kolev, B., Tasseva, V., Atanassov, K., Generalized Net Model for Binary Operations over Intuitionistic Fuzzy OLAP Cubes. Proceedings of the Eighth Int. Workshop on Generalized Nets, Sofia, 26 June 2007, 66–72.

[80] Chountas, P., Shannon, A., Rangasamy, P., Atanassov, K., On Intuitionistic Fuzzy Trees and Their Index Matrix Interpretation, Fifth International Workshop on IFSs, Banska Bystrica, Slovakia, 19 Oct. 2009NIFS Vol. 15, No 4, 2009, 52–56.

[81] Chountas, P., Tasseva, V., Kolev, B., Atanassov, K., Generalized Net Models Representing Data Warehouse Operations. Annual of ”Informatics” Section Union of Scientists in Bulgaria, Vol. 1, 2008, 82–92.

[82] Chountas, P., Tasseva, V., Kolev, B., Atanassov K., Generalized Net Model of Selection Operation over Intuitionistic Fuzzy Cube. Issues in Intuitionistic Fuzzy Sets and Generalized Nets, Vol. 6, 2008, 86–95.

[83] Chountas, P., Kolev, B., Rogova, E., Tasseva, V., Atanassov, K., Generalized Nets in Artificial Intelligence. Volume 4: Generalized Nets, Uncertain Data and Knowledge Engineering, "Prof. Marin Drinov" Publishing house, 2007.

[84] Chountas, P., Petrounias, I.,. Precise enterprises and imprecise data. In: Barzdins, J. and Caplinskas, A. (ed.) Databases and information systems: Fourth International Baltic Workshop, Baltic DB&IS 2000, Vilnius, Lithuania, May 1-5, 2000, London, UK Kluwer Academic Publishers, 57–68.

[85] Chountas, P., Petrounias, I., Vasilakis, C., El-Darzi, E., Tseng, A., Kolev, B., Kodogiannis, V. and Georgiev, P., Temporality and intuitionistic fuzzy data warehouses. Notes in Intuitionistic Fuzzy Sets., vol. 10, No 4, 2004, 47–55.

[86] Chountas, P., Petrounias, I., Vasilakis, C., Tseng, A., El-Darzi, E., Atanassov, K.T. , Kodogiannis, V., On uncertainty and data-warehouse design. in: Yakhno, T. (ed.) Advances in Information Systems: Third International Conference, ADVIS 2004, Izmir, Turkey, October 20-22, 2004: proceedings Berlin, Germany Springer.

[87] Chountas, P., Rogova, E., Atanassov, K.T., Expressing hierarchical preferences in OLAP queries. In: Kacprzyk, J., Petry, F.E., Yazici, A. (ed.),Uncertainty approaches for spatial data modeling and processing. Berlin Heidelberg Springer, 2010.

[88] Ciferri, C.D., Ciferri, R.R., Gómez, L.I., Schneider, M., Vaisman, A.A., Zimányi, E., Cube Algebra: A Generic User-Centric Model and Query Language for OLAP Cubes. IJDWM, 9, 2013, 39–65.

[89] Databricks Lakehouse Platform, https://databricks.com/product/data-lakehouse.

[90] De Tré, G., Boeckling, T., Timmerman, Y., Zadrożny, S., Handling Veracity of Nominal Data in Big Data: A Multipolar Approach. In: Cuzzocrea A., Greco S., Larsen H., Saccà D., Andreasen T., Christiansen H. (eds) Flexible Query Answering Systems. FQAS 2019.Lecture Notes in Computer Science, vol. 11529. Springer, Cham., 2019.

[91] Delta Lake - Reliable Data Lakes at Scale, https://delta.io/.

[92] Doukovska, L., V. Atanassova, G. Shahpazov, F., Capkovic, InterCriteria Analysis Applied to Variuos EU Enterprises. In Proceedings of the Fifth International Symposium on Business Modeling and Software Design, Milan, Italy, 2015, 284–291.

[93] Impetus: Enabling a unified view of your business on cloud, https://www.impetus.com/.

[94] Fidanova, S., Atanassov,K., Marinov,P.,Generalized Nets in Artificial Intelligence. Vol. 5: Generalized Nets and Ant Colony Optimization. “Prof. M. Drinov” Academic Publishing House, Sofia, 2011.

[95] Fidanova, S., Roeva, O., Comparison of Different Metaheuristic Algorithms based on InterCriteria Analysis.Journal of Computational and Applied Mathematics, Available online 7 August 2017, https://doi.org/10.1016/j.cam.2017.07.028.

[96] Gallo, C., Perilli, M., Bonis, M., Data Warehouse Design and Management: Theory and Practice, 2010.

[97] Giebler, C., Gröger, C., Hoos, E., Schwarz, H., Mitschang B., Leveraging the Data Lake: Current State and Challenges. Proceedings of the 21st International Conference on Big Data Analytics and Knowledge Discovery(DaWaK), 2019.

[98] Gocheva, E., Sotirov, S., Modelling of the Verification by Iris Scanning by Generalized Nets. 9th International Workshop on Generalized Nets, Sofia, 4 July 2008, 9–13.

[99] Gocheva, P., Sotirov, S.,Gochev,V., Implementation of Generalized Nets Models of Feed-Forward Neural Networks. Issues in Intuitionistic Fuzzy Sets and Generalized Nets, Vol. 10, 2013, 125–135.

[100] Google BigQuery: Cloud Data Warehouse, https://cloud.google.com/bigquery.

[101] Gray J. et.al., Data Cube: A Relational Aggregation Operator Generalizing Group-by, Cross-Tab and Sub Totals.Data Mining and Knowledge Discovery Journal, Vol 1, No 1, 1997.

[102] Gray, J., Chaudhuri, S., Bosworth, A., Layman, A., Reichart, D., Venkatrao, M., Pellow, F., Pirahesh, H., Data Cube: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Totals. Data Mining and Knowledge Discovery, 1, 2004, 29–53.

[103] Gupta, S.,Giri, V., Practical Enterprise Data Lake Insights: Handle Data-Driven Challenges in an Enterprise Big Data Lake, Apress, 2018.

[104] Hortonworks Data Platform, https://www.cloudera.com/products/hdp.html.

[105] Harrison, G., Next Generation Databases: NoSQL, NewSQL, and Big Data, Apress, 2016.

[106] Hassanien, A., Azar, A., Snasael, V., Kacprzyk, J., Abawajy, J. (eds), Big Data in Complex Systems. Studies in Big Data, vol 9. Springer, Cham., 2015.

[107] Ikonomov, N., P. Vassilev, Roeva, O.,ICrAData – Software for InterCriteria Analysis.Int J Bioautomation, Vol. 22, No. 1, 2018, 1–10.

[108] Inmon, W., Building the Data Warehouse: Fourth Edition, John Wiley & Sons, 2002.

[109] Inmon, W., Linstedt, D., Data Architecture: A Primer For The Data Scientist: Big Data, Data Warehouse and Data Vault, Elsevier, 2015.

[110] Inmon, W., Strauss, D., Neushloss, G., DW 2.0: The Architecture for the Next Generation of Data Warehousing, Elsevier, 2008.

[111] Janev, V., Graux, D., Jabeen, H., Sallinger, E., Knowledge Graphs and Big Data Processing, Springer, 2020.

[112] Japkowicz, N., Stefanowski, J., Big Data Analysis: New Algorithms for a New Society, Springer, 2016.

[113] Kimball, R., Caserta, J., The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data, Wiley Publishing, 2004.

[114] Kimball, R., Reeves, L., Ross, M., Thornthwaite, W., The Data Warehouse Lifecycle Toolkit: Expert Methods For Designing Developing And Deploying Data Warehouses, John Wiley & Sons, 1998.

[115] Kimball, R., Ross, M., The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, Third Edition, John Wiley & Sons, 2013.

[116] Kolev, B., Intuitionistic Fuzzy Generalized Net Analysis of Periodic Deadlock Detection in Database Systems. Proceedings of IEEE Conference "Intelligent Systems", Vol. 3, 2002, 10–12.

[117] Kolev B., Chountas P., Petrounias I., Intuitionistic Fuzzy Generalized Net in Analysing Transaction Database Systems with Continuous Deadlock Detection. Notes on Intuitionistic Fuzzy Sets, Vol. 8, 2002, 95–100.

[118] Kolev, B., El-Darzi, E., Sotirova, E., Petrounias, I., Atanassov, K., Chountas, P., Kodogiannis, V., Generalized Nets in Artificial Intelligence. Volume 3: Generalized Nets, Relation Databases and Expert Systems, "Prof. Marin Drinov" Publishing house, 2006.

[119] Krawczak, M., Multilayer Neural Networks: A Generalized Net Perspective, Springer International Publishing, 2013.

[120] Krawczak, M., Sotirov S., Sotirova E., Generalized Net Model of the Early Stopping Method for Parallel Optimization of Multilayer Perceptron. Advances in Intelligent Systems and Computing, Springer International Publishing, 2017, 103–113.

[121] Krawczak, M., Sotirov, S.,Atanassov, K.,Multilayer Neural Network Modelling by Generalized Nets. Warsaw School of Information Technologies, 2010.

[122] Krawczak, M., Bureva V., Sotirova, E., Szmidt, E.,Application of the InterCriteria Decision Making Method to Universities Ranking. In: Novel Developments in Uncertainty Representation and Processing, Vol. 401,Advances in Intelligent Systems and Computing, 2016, Springer, 365–372.

[123] Krumova, S., Todinova, D. Mavrov, P. Marinov, V. Atanassova, K. Atanassov, S. Taneva, Intercriteria analysis of calorimetric data of blood serum proteome. Biochimica et Biophysica Acta – General Subjects, Vol. 1861, No. 2, 2017, 409–417.

[124] Laurent, A., Laurent, D., Madera, C., Data Lakes, Volume 1 and 2, Wiley, 2020

[125] Linstedt, D., Olschimke, M.,Building a Scalable Data Warehouse with Data Vault 2.0, 2015.

[126] Liu, R., Isah, H., Zulkernine, F., A Big Data Lake for Multilevel Streaming Analytics., 1st International Conference on Big Data Analytics and Practices (IBDAP), 2020, 1–6.

[127] Marz, N., Warren, J., Big Data: Principles and best practices of scalable real-time data systems, Manning publications, 2015.

[128] Mavrov, D., Software for InterCriteria Analysis: Implementation of the main algorithm, Notes on Intuitionistic Fuzzy Sets, Vol. 21, No. 2, 2015, 77–86.

[129] Microsoft Azure, https://azure.microsoft.com/en-us/.

[130] Microsoft Visual Studio, https://visualstudio.microsoft.com.

[131] Oracle: Integrated Cloud Applications and Platform Services, https://www.oracle.com/index.html.

[132] Orozova, D., Sotirova, E., Generalized Net Model of the Applying Data Mining Tools. Proceedings of the 10th International Workshop on Generalized Nets, Sofia, 5 Dec 2009, 22–26.

[133] Parvathi, R., Thilagavathi, S., Intuitionistic fuzzy directed hypergraphs. Advances in fuzzy Sets and System, 14 (1), 2013, 39–52.

[134] Parvathi, R., Shannon, A., Chountas, P., Atanassov, K., On intuitionistic fuzzy tree-interpretations by index matrices. Fifteenth Int. Conf. on IFSs, Burgas, 11-12 May 2011, NIFS, vol. 17, No2, 2011, 17–24.

[135] Pencheva, T., Angelova, M., Atanassova, V., Roeva, O., InterCriteria Analysis of Genetic Algorithm Parameters in Parameter Identification. Notes on Intuitionistic Fuzzy Sets, Vol. 21, No. 2, 2015, 99–110.

[136] Rogova, E., Chountas, P., On imprecision intuitionistic fuzzy sets & OLAP - the case for KNOLAP. in: Castillo, O., Patricia Melin, B., Pedryczev, W. and Kacprzyk, J. (ed.) Theoretical Advances and Applications of Fuzzy Logic and Soft Computing, Berlin and Heidelberg GmbH & Co. KG Springer, 11–22.

[137] Plattner, H. (2009), A common database approach for OLTP and OLAP using an in-memory column database, SIGMOD Conference.

[138] Presto: Distributed SQL Query Engine for Big Data, https://prestodb.io/.

[139] Ravat, F., Teste, O., Tournier, R., Zurfluh, G., Algebraic and graphic languages for OLAP manipulations. International Journal of Data Warehousing and Mining (IJDWM), vol. 4 , No 1, 2008.

[140] Roeva, O., Fidanova, S., Vassilev, P., Gepner, P., InterCriteria Analysis of a Model Parameters Identification using Genetic Algorithm, Annals of Computer Science and Information Systems, Vol. 5, 2015, 501–506.

[141] Roeva, O., Penchcva, T., Shannon, A., Atanassov, K., Generalized Nets in Artificial lntelligence.V ol. 7: Generalized nets and Genetic Algorithms. lbid., 2013.

[142] Roeva, O., Vassilev, P., Angelova, M., Su, J., Pencheva, T., Comparison of Different Algorithms for InterCriteria Relations Calculation. IEEE 8th International Conference on Intelligent Systems, 2016.

[143] Rogova, E., Chountas, P., On Imprecision Intuitionistic Fuzzy Sets & OLAP – The Case for KNOLAP, IFSA'07, Springer-Verlag, Theoretical Advances and Application of Fuzzy Logic and Soft Computing, 11–20.

[144] Shannon, A., Atanassov, K., A first step to a theory of the intuitionistic fuzzy graphs, Proc. of the First Workshop on Fuzzy Based Expert Systems(D. Lakov, Ed.), Sofia, Sept. 28-30, 1994, 59–61.

[145] Shannon, A., Atanassov, K., Intuitionistic fuzzy graphs from α-,β- and (α,β)-levels. Notes on Intuitionistic Fuzzy Sets, Vol. 1, No. 1, 1995, 32–35.

[146] Shannon, A., Rangasamy, P., Atanassov, K., Chountas, P., On intuitionistic fuzzy trees. Developments in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics, Vol. I: Foundations, 2010, 177–184.

[147] Singhal, A., An Overview of Data Warehouse, OLAP and Data Mining Technology. In: Data Warehousing and Data Mining Techniques for Cyber Security. Advances in Information Security, vol 31. Springer, Boston, MA.

[148] Smallcombe, M., Meet the Data Lakehouse, https://www.xplenty.com/blog/meet-the-data-lakehouse/, 9.05.2021.

[149] Song, J., Guo, C., Wang, Z., Zhang, Y., Yu, G., Pierson, J., HaoLap: A Hadoop based OLAP system for big data. Journal of Systems and Software, Vol. 102, 2015, 167–181.

[150] Sotirov, S., Sotirova, E., Krawczak, M., Application of Data Mining in Digital University: Multilayer Perceptron for Lecturers Evaluation with Intuitionistic Fuzzy Estimations. Issues in Intuitionistic Fuzzy Sets and Generalized Nets, Vol. 8, Warsaw, Poland, 2010, 102–107.

[151] Sotirov, S., Sotirova, E., Melin, P., Castilo, O., Atanassov, K., Modular Neural Network Preprocessing Procedure with Intuitionistic Fuzzy InterCriteria Analysis Method.In:-Flexible Query Answering Systems, Vol. 400, Advances in Intelligent Systems and Computing, 2015, Springer International Publishing, 2016, 175–186.

[152] Sotirov, S., Atanassov K., Generalized Nets in Artificial Intelligence. Vol. 6: Generalized nets and Supervised Neural Networks. lbid., 2012.

[153] Sotirov, S., Atanassova V., Sotirova E., Bureva V., Mavrov D., Application of the Intuitionistic Fuzzy InterCriteria Analysis Method to a Neural Network Preprocessing Procedure. 9th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT), 30.06-03.07.2015, Gijon, Spain, 1559–1564, doi:10.2991/ifsa-eusflat-15.2015.222.

[154] Sotirov, S., Opportunities for application of the intercriteria analysis method to neural network preprocessing procedures. Notes on Intuitionistic Fuzzy Sets, Vol. 21, 2015, No. 4, 143–152.

[155] Sotirova, E., Orozova, D., Generalized Net Model of the Phases of the Data Mining Process. Developments in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics. Vol. II: Applications, Warsaw, Poland, 2010, 247–260.

[156] Teradata, https://www.teradata.com/.

[157] Traneva, V., Atanassov, K., Parvathi, R., On 3-dimensional Extended Index Matrices. International Journal of Research, 1 , 2014, 64–68.

[158] Thamizhendhi, G., Parvathi, R., Intuitionistic Fuzzy Tree Center-Based Clustering Algorithm.International Journal of Soft Computing and Engineering (IJSCE), Vol. 6, Issue-1, March 2016.

[159] Traneva, V., Atanassova, V., Tranev, S., Index Matrices as a Decision-Making Tool for Job Appointment. Numerical Methods and Applications, Springer, 2019.

[160] Traneva, V., Sotirova, E., Bureva V., Atanassov, K., Aggregation operations over 3-dimensional extended index matrices. Advanced Studies in Contemporary Mathematics, 2015 (July), vol.25, No 3, 407–416.

[161] Traneva, V., Index Matrices in the assessment of human resources.Union of Scientists of Bulgaria, Sofia, 2019.

[162] Traneva, V., On 3-dimensional intuitionistic fuzzyindex matrices. 10th Int. Workshop on IFSs, Banska Bystrica, 29 Sept. 2014, Notes on Intuitionistic Fuzzy Sets, Vol. 20, No. 4, 2014, 59–64.

[163] Traneva, V., Tranev, S., Index Matrices as a Tool for Managerial Decision Making, Publishing house of the Union of Scientists in Bulgaria, 2017.

[164] Traneva, V., Bureva, V.,Sotirova, E.,Atanassov, K., Index Matrices And OLAP-Cube Part 1: Application of the Index Matrices to presentation of operations in OLAP-Cube, Advanced Studies in Contemporary Mathematics, Vol.27, No. 2, 2017, 253–278.

[165] Traneva, V., Tranev S., Szmidt E., Atanassov K., Three dimensional intercriteria analysis over intuitionistic fuzzy data. In: J. Kacprzyk et al. (eds.), Advances in Fuzzy Logic and Technology 2017, Advances in Intelligent Systems and Computing, Vol. 641, Springer International Publishing AG 2018, 442–449.

[166] Traneva, V., Tranev, S., Stoenchev, M. et al. Scaled aggregation operations over two- and three-dimensional index matrices. Soft Comput, 22, 2018, 5115–5120.

[167] Traneva, V., V. Bureva, E. Sotirova, K. Atanassov, Index matrices and OLAP-cube part 3: A presentation of the OLAP "InterCube Set" and “Data cube” operations by index matrices. Advanced Studies in Contemporary Mathematics (Kyungshang), 2018.

[168] Traneva, V., et al., Index Matrix Interpretations of A New Transportation-Type Problem. Comptes rendus de l'Académie bulgare des Sciences, Vol. 69, No. 10, 2016, 1275–1282.

[169] Tré, G., Zadrozny, S., Sotirov, S., Atanassov, K., Using index matrices for handling multiple scenarios in decision making, Proceedings of the 21st International Conference on Intuitionistic Fuzzy Sets. In Notes on Intuitionistic Fuzzy Sets 23(2)., 2017, 103–110.

[170] Vaisman, A.A., Zimányi, E., Graph OLAP. In: Sakr S., Zomaya A. (eds) Encyclopedia of Big Data Technologies, Springer, Cham, 2018.

[171] Vasilakis, C., El-Darzi, E., Chountas, P., Data cube modelling and OLAP algebra for hospital length of stay analysis. Craiova Medicala Journal: Proceedings of The 1st MEDINF International Conference on Medical Informatics & Engineering, 5 (3), 2003, 509–512.

[172] Vasilakis, C., El-Darzi, E., Chountas, P., An OLAP-enabled software environment for modelling patient flow. In: Proceedings of the 3rd International IEEE Conference on Intelligent Systems Los Alamitos, USA IEEE, 2006, 261–266.

[173] Vassilev, P., Todorova, L., Andonov, V., An auxiliary technique for InterCriteria Analysisvia a three dimensional index matrix.19th Int. Workshop on IFSs, Burgas, 4–6 June 2015, Notes on Intuitionistic Fuzzy Sets, Vol. 21, No. 2, 2015, 71–76.

[174] Vassilev, P., Ribagin S., Todorova, On an aggregation of expert value assignments using index matrices. Notes on Intuitionistic Fuzzy Sets, Vol. 23, No. 4, 2017, 75–78.

[175] Xuepeng, Y., Pedersen, T., Algebra-Based Optimization of XML-Extended OLAP Queries. COMAD, 2006.

[176] Yang, Q., Ge, M., Helfert, M., Analysis of Data Warehouse Architectures: Modeling and Classification. ICEIS, 2019.

[177] Yessad, L., Labiod, A., Comparative study of data warehouses modeling approaches: Inmon, Kimball and Data Vault, 2016 International Conference on System Reliability and Science (ICSRS), Paris, France, 2016, 95–99.

[178] Zadrożny, S., Kacprzyk, J., Bipolar queries: An aggregation operator focused perspective.Fuzzy Sets and Systems, Vol. 196, North-Holland, 69–81.

[179] Zaloni, https://www.zaloni.com/.

[180] Zhao, P., Li, X., Xin, D., Han, J., Graph cube: On warehousing and OLAP multidimensional networks. Proceedings of the ACM SIGMOD International Conference on Management of Data, 2011, 853–864.

[181] Zomaya, A., Sakr, S., Handbook of Big Data Technologies, Springer, 2017.

[182] Mongo, DB, https://www.mongodb.com/.

[183] Journey from Databases to Data Lakehouse, https://siddharth-mehta.medium.com/journey-from-databases-to-data-lakehouse-3e10e49b0d60.

[184] An Architect's View of the Data Lakehouse: Perplexity and Perspective, https://www.eckerson.com/articles/an-architect-s-view-of-the-data-lakehouse-perplexity-and-perspective.