UDC 004.67

Denis S. Shibaev[1], Graduate student of the Department of Technical Cybernetics and Information Technology, E-mail: denscreamer@gmail.com, ORCID: 0000-0002-3260-5843
Vladimir V. Vyuzhuzhanin[2], Doctor of Technical Sciences, Professor, Head of the Department of Information Technologies, E-mail: vint532@yandex.ua, Scopus ID: 57193025809, ORCID: 0000-0002-6302-1832
Nikolay D. Rudnichenko[2], Candidate of Technical Sciences, Associate Professor at the Department of Information Technologies, nickolay.rud@gmail.com, ORCID: 0000-0002-7343-8076
Natalia O. Shibaeva[2], Candidate of Technical Sciences, Associate Professor, Department of Information Technologies, nati.shibaeva@gmail.com, ORCID: 0000-0002-7869-9953
Tatyana V. Otradskaya[3], Candidate of Technical Sciences, Director of Odessa Computer Technology College "Server", tv_61@ukr.net, ORCID: 0000-0002-5808-5647
[1]Odessa National Maritime University, st. Mechnikov, 34, Odessa, Ukraine, 65029
[2]Odessa National Polytechnic University, Avenue Shevchenko, 1, Odessa, Ukraine, 65044
[3]Odessa College of Computer Technology "Server", Polish Descent, 1, Odessa, Ukraine, 65026

DATA CONTROL IN THE DIAGNOSTICS AND FORECASTING THE STATE OF COMPLEX TECHNICAL SYSTEMS

Annotation. *The analysis of management methods Big Data is carried out. In order to obtain timely results of analyzing the state of complex technical systems on the basis of the list of parameters established by regulatory documentation that are of paramount (for critical components) and minor importance in diagnosing the state of components ensuring the operation of complex technical systems, it is necessary to develop a method for managing data with high speed and losslessly separate and transfer Big Data from IIS to relational and non-relational databases. A method is proposed that ensures the distribution of data coming from information-measuring systems to relational and non-relational databases in diagnosing and predicting the state of complex technical systems. The expediency of using the concept of Data Mining in SCADA systems to control Big Data is substantiated. Algorithms for transmission, distribution and analysis of data in an information-measuring system for diagnosing and predicting the state of complex technical systems have been developed. The scheme of data transmission in devices using the CAN bus. The proposed method for managing Big Data in diagnosing and predicting the state of complex technical systems is based on ensuring the dynamic distribution of data in an information-measuring system with regard to the requirements imposed on the structure of the local-computer network. The method is based on the application of the principles of the construction of software-configured networks, allowing to manage the network by using the results of the analysis of data flows passing through the node-based network devices. A software implementation of a data distribution system in a local network has been developed using the principle of analyzing network packets as they arrive at the switching nodes of an information-measuring system. The system of program logic of data distribution from information-measuring systems transmitted over local networks or via CAN bus has been developed. From the conducted research, it follows that the best performance of the data separation process according to the developed method is achieved with distributed execution of computational processes by the developed program in four continuous modes, and the data separation process in non-relational bases for all experiments performed is faster than for relational data. The use of the developed Big Data management method with data distribution in relational and non-relational databases provides an increase in the speed of the information-measuring system in diagnosing and predicting the state of complex technical systems. Allows you to predict the technical condition of critical components of the systems during their short-term in an emergency condition, as well as to carry out a long-term forecast of the technical condition of the entire complex technical system. The use of software distributors of transmitted information provides an increase in the speed of the information-measuring system in diagnosing and predicting the state of complex technical systems, thereby ensuring timely assessment of the state of the critical components of complex technical systems whose failure affects the operation of the systems.*
Keywords: *data control; Big Data; DataMining; SCADA systems; relational and non-relational databases; information measuring system; data distribution; complex technical systems; diagnostics; forecasting the state of technical systems*

Introduction

The implementation and expansion of the information systems in complex technical systems (CTS) usage provides the level of quality and an assessment of their effectiveness increasing in various operational modes.

The efficiency of the operation of the CTS largely depends on their reliability. Reliable functioning of CTS interconnected components depends on their safety, survivability, controllability, etc. [1]. During the operation period, the CTS technical state is subject to a change, which means that timely analysis of such systems operation reliability is necessary [2-4].

This leads to the complication of information systems for the CTS and raising the volume of processed Big Data in real time, coming from information-measuring systems (IMS) to relational and non-relational databases. Increased requirements for speed of data transfer to databases, in which incoming Big Data should be stored for the purpose of their high-speed processing in order to timely determine the CTS components failures risk, thus ensuring reliable CTS operation [5-8].

It is known that the relational database model provides atomicity, consistency, isolation, and reliability when manipulating data.

However, this hinders the maintenance of high availability and speed in the case of the Big Data distribution across several services. To ensure high performance in a distributed way of storing information, non-relational databases are used.

Thus, the technologies used at all stages of receiving, processing, storing and transmitting information should fully ensure the speed and high quality management of the distribution of data received from IMS to relational and non-relational databases in diagnosing and predicting the CTS state and thereby ensure reliable operation of such systems [9-13].

Problem statement

Many data management structures for monitoring the CTS state have a scheme that corresponds to the SCADA automatic control and information collection system [14-16].

The system is designed to develop or provide real-time systems for collecting, processing, displaying and archiving information about the diagnostic object.

SCADA has requirements, the fulfillment of which also affects the quality of CTS functioning diagnosed, namely:

– system reliability (technological and functional);

– security management;

– accuracy of data processing and presentation;

– system expansion simplicity.

One of the problems arising from the use of data management systems is the problem of analyzing the information used after transferring it to the IMS central processing unit.

This is due to IMS high loads when analyzing the current state of the CTS and the complexity of building the IMS hardware.

To do this, SCADA systems use the methods of processed data synchronization using dispatch control.

Another problem is the need to reduce time and computational costs when Big Data searching and transferring to various distributed scalable systems in order to analyze them.

This problem can be solved by developing and applying methods and means of data distribution, taking into account their characteristics (data types, variability, accuracy, etc.) and interrelations.

When transferring Big Data, their loss may occur due to incorrect distribution of data, errors in the database management system operation, problems of network equipment, software failures affecting the operation of the automatic control system and information collection of SCADA. To overcome the above problems using SCADA-architecture, dispatchers are involved to control most of the solving data management tasks processes.

However, in this case there are no guarantees of the transmitted data loss, which is unacceptable when assessing the state of the CTS responsible for security and survivability of technical systems for special purposes.

In order to obtain timely results of the CTS state analysis based on the list of parameters established by the regulatory documentation that are of primary (for critical components) and of secondary importance in diagnosing the components state ensuring the CTS functioning, it is necessary to develop a data management method that allows high-speed and lossless separation and transfer Big Data from IMS to relational and non-relational databases.

Recent research and publications analysis

Improving the quality of the CTS state assessing requires the development of IMS information subsystems, in particular, ensuring the collection, storage and processing of Big Data.

Big Data analysis methods are diverse and include mathematical statistics, algorithmic classification, clustering, and so on. Each of the methods describes the patterns of data changes with varying degrees of accuracy. However, the practical application of such methods significantly complicates the concept of diagnosing the CTS condition, since additional information is required related to the specifics of the methods used [17-20]

Many modern software systems focused on data management for assessing and predicting the state of the CTS are unable to interact with Big Data in IMS.

This is due to the use of equipment for transmitting, storing and processing information that is unable to perform complex analytical operations in real time with changing data sets. A methodology capable of solving multi-class problems of searching and interpreting results is the Data Mining [21-22]. The main feature of the methodology is to find patterns in the data.

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

The use of such solutions for data management allows us to build a complete information system for analyzing Big Data in order to assess the state of the CTS.

The aim of the work

The aim of this paper is to improve the speed and reliability of Big Data transmission based on data management by programmatically distributing them during transmission over CMS local networks to relational and non-relational databases.

Research methods

The methods of technical analysis of hardware systems, building complex network architectures, designing specialized software solutions are used.

Main research material

One of the tasks aimed at improving the quality of assessments of the CTS state diagnosed is the management of Big Data in IMS integrated into functional modules.

The work of the modules is based on the use of a central data storage system.

As data transmission lines, local computer networks (LANs) [23-25] and specialized onboard information buses IMS, installed in transport are used IMS work with LAN is based on the use of switches and routers, as well as data conversion devices. The working network is integrated with telemetric, analytical and diagnostic systems that provide the resulting conclusion of the IMS state assessment. However, this doesn’t guarantee high accuracy of the analysis of the CTS state and can’t provide high speed data analysis in real time. This is due to the large amounts of diagnostic information stored in the central repository.

In addition, the structural complexity of the CTS can be so great that the amount of incoming information from the IMS can exceed the capabilities of the data management system in diagnosing the CTS state. In this case, the diagnostics of the CTS state is performed remotely in specialized centers for evaluating the results of modular IMS to which data is transmitted via global networks or directly connected to the IMS local network. However, the implementation of this approach based on classical methods of organizing network switching is not very effective and requires modification

To overcome this problem, a method for managing streaming data is proposed by distributing it in real time between relational and non-relational databases.

The essence of the method consists in combining the concept of software-configured networks (SCN) and Data Mining methods in order to identify, with high speed, the most important diagnostic information for assessing and predicting the CTS state.

The principles of operation of the SCN is based on flexible network management by applying the results of the data streams passing through the node network devices analysis. The use of transmitted information software distributors is intended to improve the speed of the system for diagnosing the CTS condition, as well as to carry out a primary analysis of the CTS critical elements state, the failure of which can affect the entire system performance. The structure of IMS uses numerous sensors and converting amplifying devices, which are combined into measuring-control devices (MCD) modules with their own interface (Fig. 1).

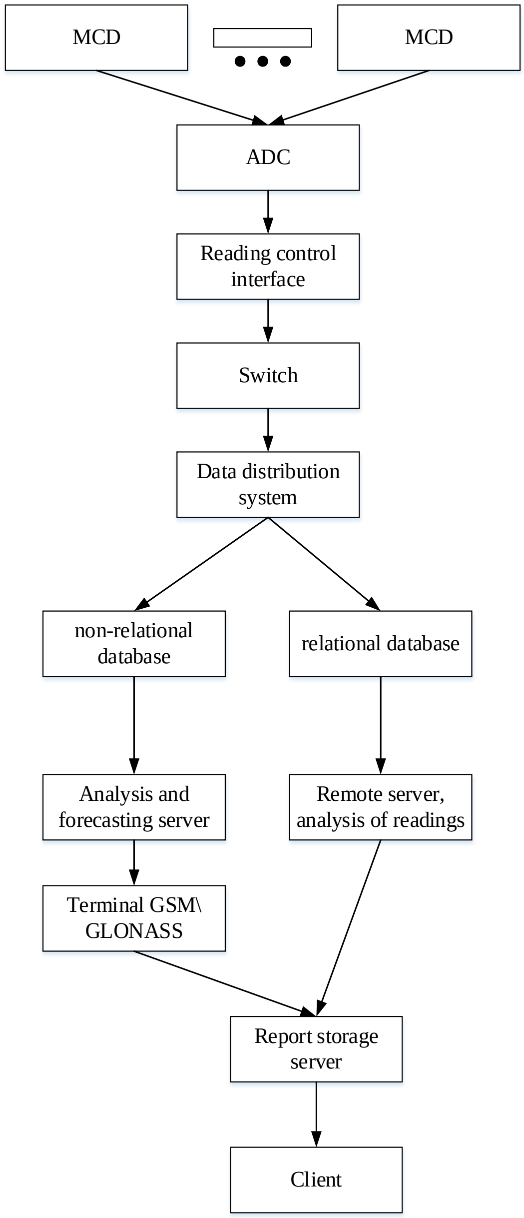


Fig. 1. Data distribution scheme in information and measurement systems

The monitoring control module can be either an on-board computer or plug-in devices for decoding signal readings, depending on the type of CTS. Each module forms a data packet and transmits it to the network for further analysis. The distribution of data is performed at the level of network switches based on the use of the developed additional software.

To implement a software solution, it is recommended to use switches that can analyze network data and distribute information across relational and non-relational databases in the used hardware.

The basis of the developed method, which provides a dynamic distribution of testimony from the IMS, is based on the requirements for the organizational device of a common LAN. The method uses the principles of real-time network traffic analysis, as well as the use of vertical data analysis methods for individual network packets whose data is analyzed.

The direction of the entire data flow analysis requires data transfer only at the transport level with the applicable connection-oriented network type.

Therefore, it is not possible to use such an approach in diagnosing and predicting the CTS state due to multi-structured equipment operating at different levels of the OSI model.

In this regard, the optimal solution for intercepting network traffic in IMS CTS is the direction of vertical analysis, which includes a set of known methods:

– surface packet analysis (SPI);
– average packet analysis (MPI);
– deep packet analysis (DPI).

Each of these methods is an effective solution and differs only in the amount of information read from each packet and the possibilities of using the information received.

In the developed method of the diagnosed information distribution, the MPI method is used, which is the most preferable for vertical analysis of data packets. To implement the MPI method, a separate functional module was developed that allows using such a solution in any LAN segment.

The developed software parses the packet headers to the transport layer of the OSI model and a small part of the packet to match the parsed part with the pre-created list of matches (parse list). Depending on the information received and the comparison made with the parse list, a solution is formed for the final storage of information in one of the storage systems.

Priorities in the distribution of data are established according to the regulatory documentation list of parameters that are of primary (for critical components) and of secondary importance used in diagnosing the components state that ensure the functioning of the CTS in different modes of operation.

The data distribution process is performed when a network packet is modified on a LAN. To provide it we need to add the address of the end node of data storage corresponding to the required data storage model.

The network packet structure is shown in Fig. 2, the content of which is read by means of network analysis. Packet modification is performed by changing the destination ip-address.

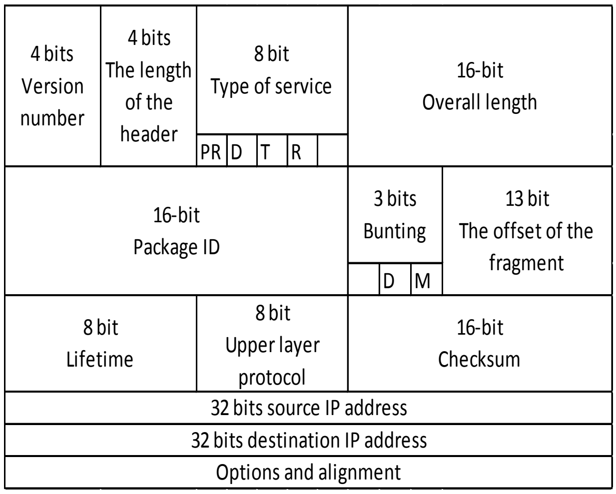


Fig. 2. Network package structure

To build a short-term prediction of the CTS state it is necessary to use data sets stored in non-relational storage. This will provide the ability to perform a quick prediction of the CTS state with great accuracy with low computational resources.

An important factor affecting the data distribution system is their consistency, since IMS CTS may consist of various hardware components.

The same applies to IMS, which, depending on the components used, may have different hardware solutions.

In this regard, the work of the data distribution module may differ significantly for each individual case.

To solve this problem, it is necessary to use a unified system for the subsequent transfer of a network packet to the storage system. This is implemented in the developed network packet collector, which provides a single diagnostic data structure.

The result of selecting data and recording them in a non-relational database is shown in Fig. 3.

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

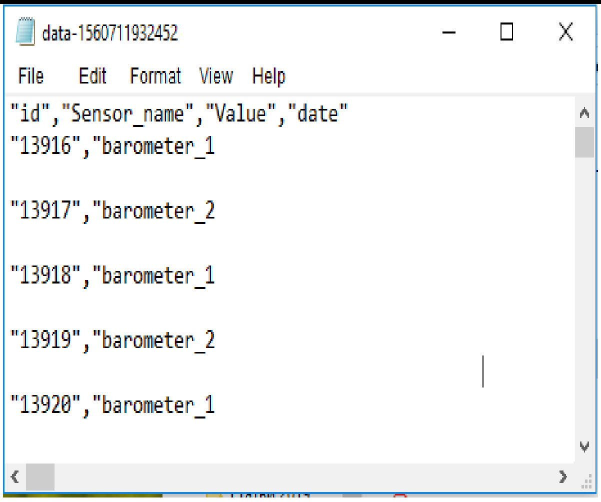


Fig. 3. Data representation in a non-relational database

Data distribution system software development

The implementation of the developed data distribution method for diagnosing the CTS state includes a software design stage with its further development. UML notation language was chosen as a tool for designing software systems.

The developed software solution has client-server architecture. The server part is located in the IMS CTS, which ensures the continuous operation of the system for collecting data on the current system states. The client part can be remote from the object of diagnosis and work using data transfer protocols. This solution provides the ability to simultaneously remotely diagnose the condition of a large number of CTS.

The developed software solution has a modular architecture and can be expanded by adding new functionality. Since the developed system is organized according to the SCADA principle, the central server receives data from remote CTS servers in real time. This allows us to provide a constant update of the data and use them for more accurate prediction of the CTS state.

Communication with remote servers IMS CTS is carried out using satellite or telephone data lines. This solution contributes to a permanent connection to the central SCADA server regardless of the current location of the CTS. In order to increase the stability of the software, the client part runs on a central SCADA server, and users connect in terminal mode. This ensures data integrity and the ability to simultaneously use all data stored on servers.

An important factor is the provision of the possibility of data processing centers remote location, thereby forming the possibility of building territorial expert centers. Scheme SCADA architecture using miscellaneous data storage is shown in Fig. 4.

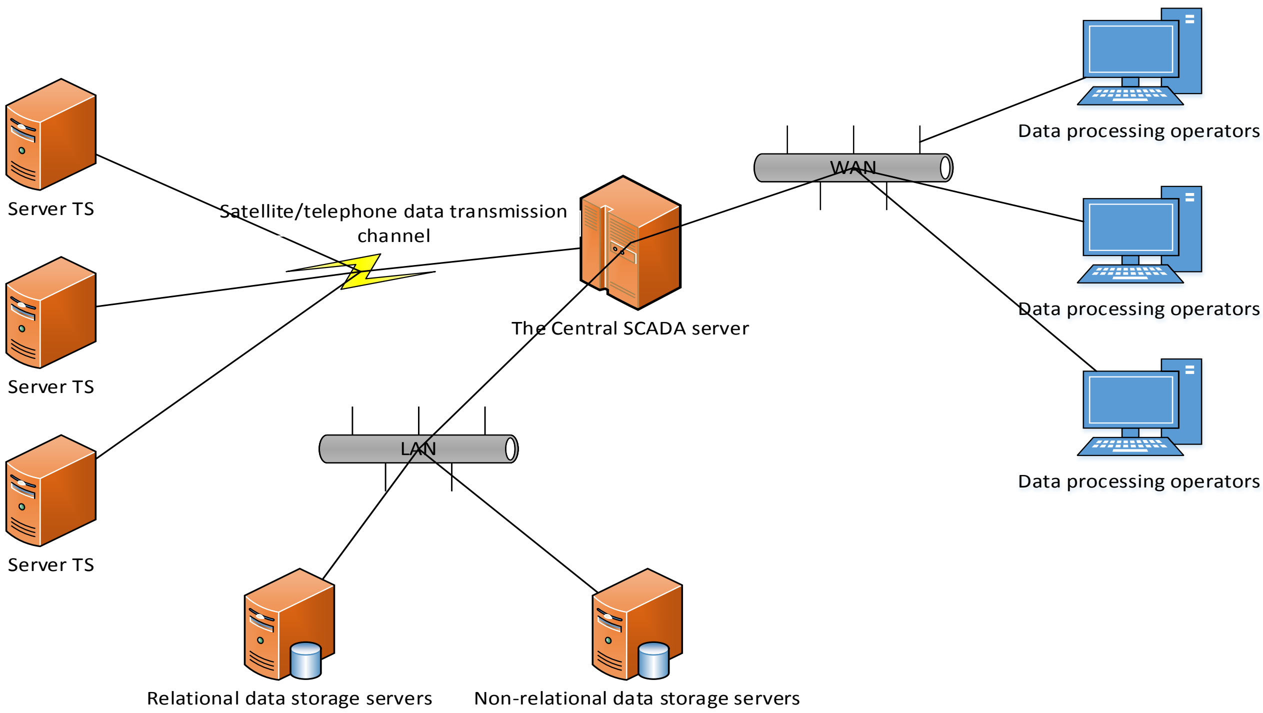


Fig. 4. Network architecture of SCADA-system scheme using multi-structure data storage

The initial stage of the task formalization is the development of a use case diagram (Fig. 5), reflecting the functionality of the software system.

The availability of functionality for the operators of the client part of the software system allows us to define the logic of the prediction module and to develop a software part, explaining the order of actions needed to be performed at the stage of building the prediction of the CTS state. For this, the CTS state diagram is used (Fig. 6).

The use of the developed project documentation and the established features of the software components made it possible to implement the development of software for managing data in diagnosing and predicting the CTS state.

The implementation of the client part is done using the C ++ development language and the QT cross-platform framework (Fig. 6).

PostgresSqL is used as a storage system for relational data, MongoDB is used for non relational data. The server side is also implemented by means of C++, which provides high-speed processing and transfer of data.

The class diagram in Fig. 7 shows the classes “ParserData”, “AsioListener”, “Listener”, collecting data from hardware CTS control devices, extracting a header from a network packet and performing the distribution according to a pre-defined rule based on the comparison of a key field with a parse list.

The parse list is stored with the system as an xml file, an example of which is shown in Fig. 8.

The basis of the rule is the entry in the non-relational database of critical subsystems (components) for analyzing and forecasting their condition, assessing the risk of failures.

The data necessary for diagnosing and predicting the state of auxiliary subsystems (components) that do not critically affect the work of the CTS are entered into the relational database.

The selection of data takes place according to parse list, after which, in order to the rule, the data is distributed among relational and non-relational databases. The distribution takes into account the device priority, which is selected from the priority field and the database is the field of the parse list file.

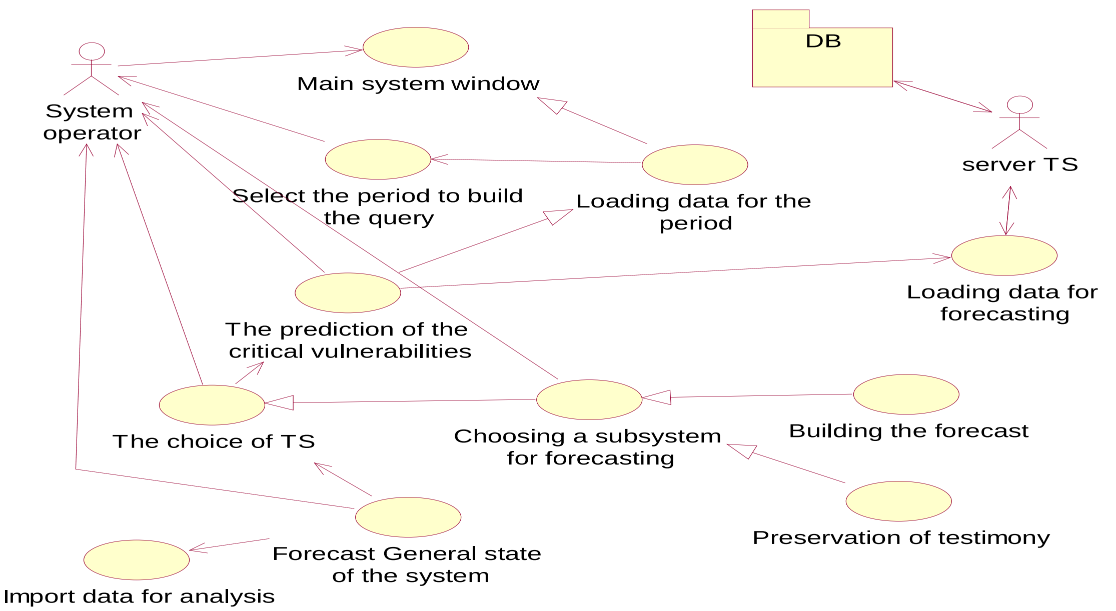


Fig. 5. Software use case diagram for predicting the CTS state

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

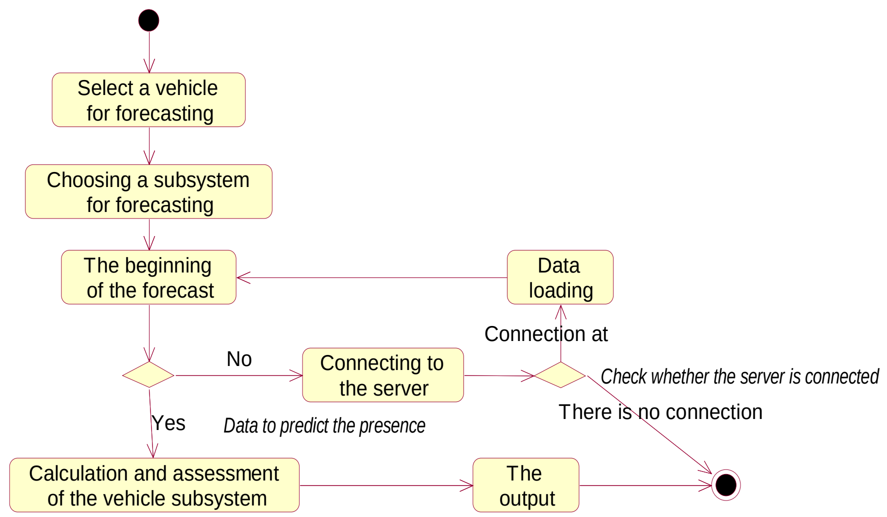


Fig. 6. CTS State Diagram

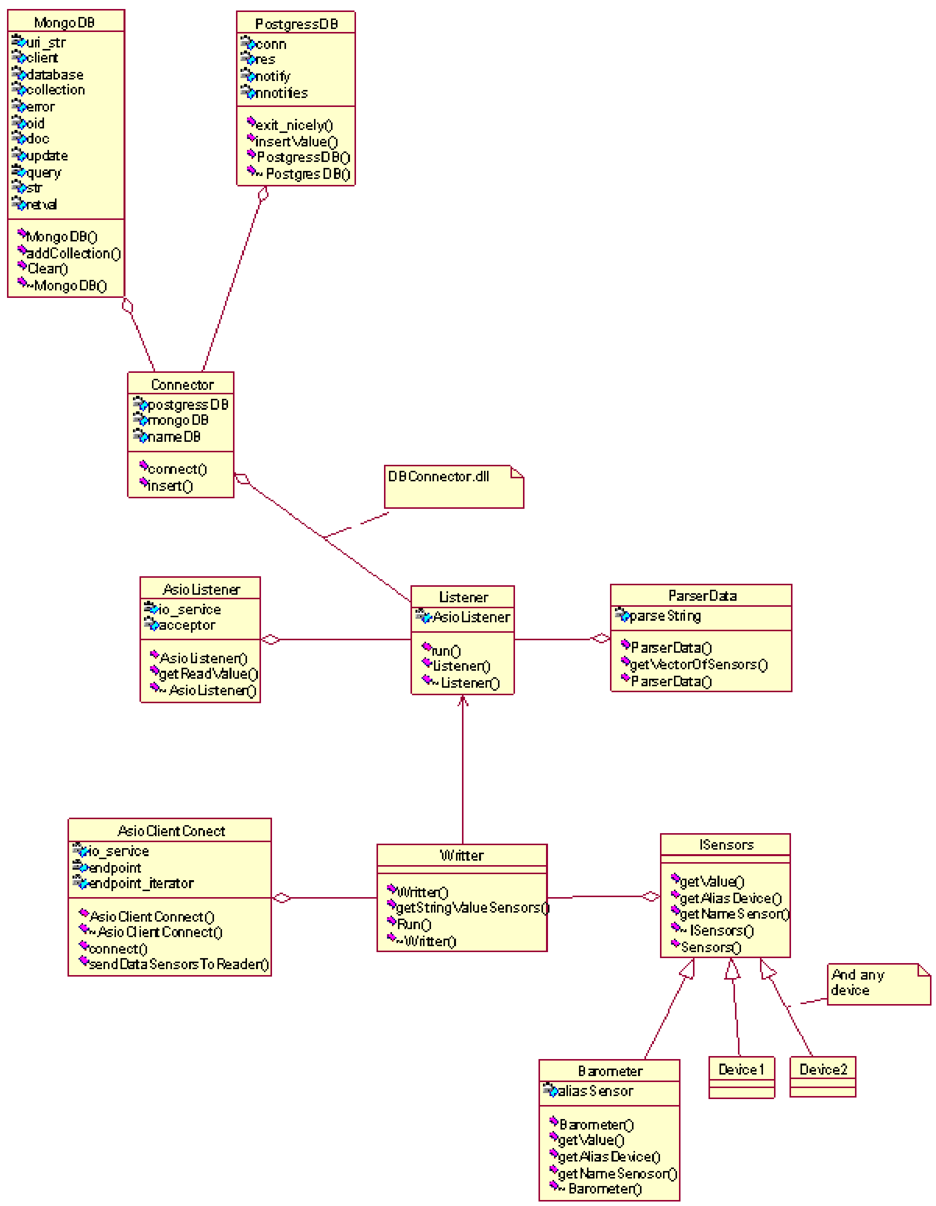


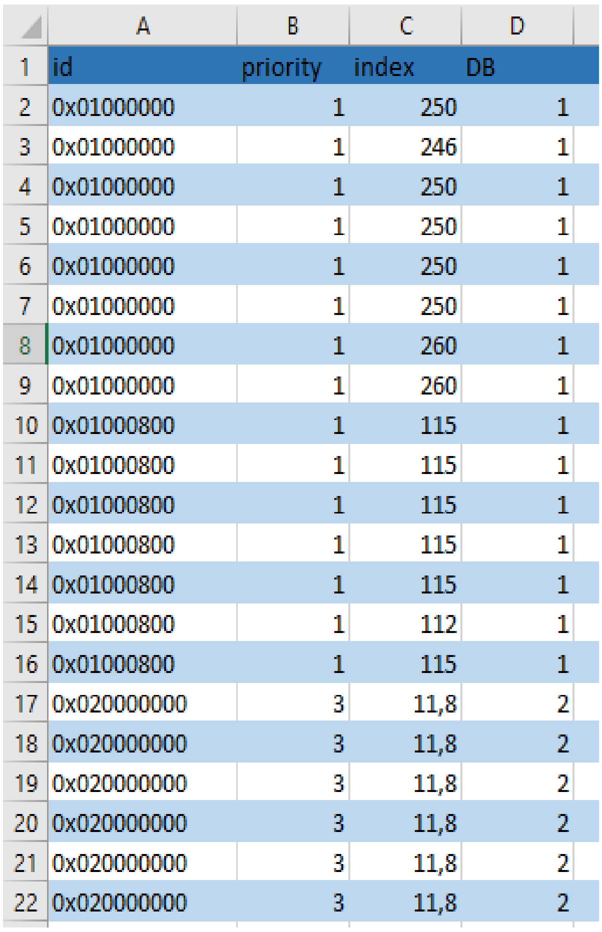
Fig. 7. Class diagram for the data distribution software module

Research of the data distribution and processing

For a numerical evaluation of the data separation processes speed, the system carried out profiling of the created software application work by means of the development environment on a test bench: Intel Core i5 7200u central processor, DDR4 16 GB RAM. 10 computational experiments were performed on the separation of various volumes data, on the basis of which the time expenditures were calculated when executing the program in 1, 2 and 4 streams (T1, T2 and T4, respectively).

The research results are summarized in Table 1. The results are graphically shown in Fig. 9 and Fig. 10.

From the conducted research, it follows that the best performance of the data separation process according to the developed method is achieved with distributed execution of computational processes by the developed program in four continuous modes, and the data separation process in non-relational bases for all the experiments performed is faster than for data in relational databases (on average by 8-10 %).



	A	B	C	D
1	id	priority	index	DB
2	0x01000000	1	250	1
3	0x01000000	1	246	1
4	0x01000000	1	250	1
5	0x01000000	1	250	1
6	0x01000000	1	250	1
7	0x01000000	1	250	1
8	0x01000000	1	260	1
9	0x01000000	1	260	1
10	0x01000800	1	115	1
11	0x01000800	1	115	1
12	0x01000800	1	115	1
13	0x01000800	1	115	1
14	0x01000800	1	115	1
15	0x01000800	1	112	1
16	0x01000800	1	115	1
17	0x020000000	3	11,8	2
18	0x020000000	3	11,8	2
19	0x020000000	3	11,8	2
20	0x020000000	3	11,8	2
21	0x020000000	3	11,8	2
22	0x020000000	3	11,8	2

Fig. 8. Parse list structure

Table 1. Software profiling results

No.	Data size, GB	Relational data T1, c	T2, c	T3, c	Non-relational data T1, c	T2, c	T3, c
1	100	211	198	186	193	182	179
2	150	326	298	269	297	273	254
3	200	442	401	364	401	375	352
4	250	511	499	398	478	458	441
5	300	621	553	501	576	532	493
6	500	1050	876	751	987	898	851
7	600	1284	1171	1034	1011	976	912
8	750	1610	1478	1344	1576	1543	1521
9	900	1905	1811	1745	1794	1752	1711
10	1000	2256	2021	1882	1874	1672	1511

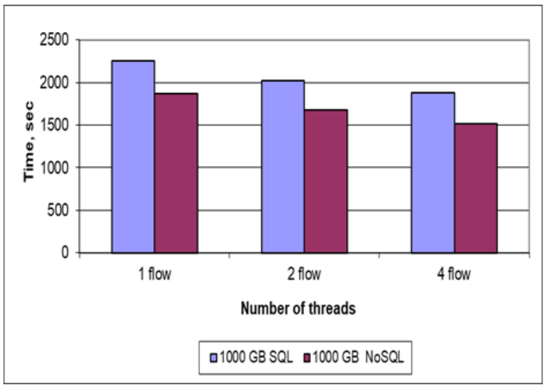


Fig. 9. Results of software profiling on 1000 GB

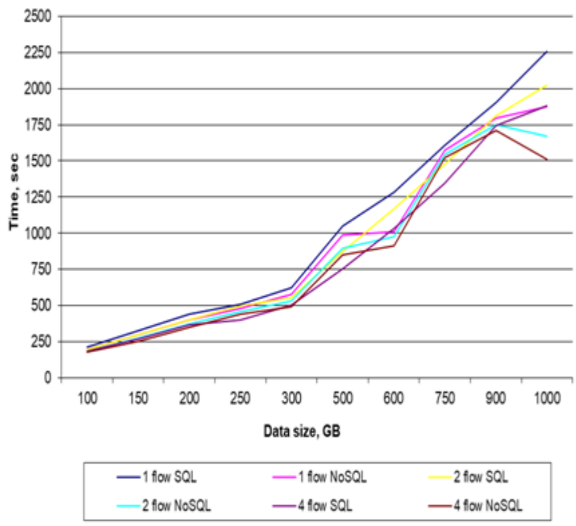


Fig. 10. Dependencies of the different volumes data separation duration

This result is due to the lack of the need to form links between individual data blocks and the possibility of separating computational operations into different cores with caching of reading and writing processes.

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

In order to assess the state of the components of the CTS from the data from the relevant databases, the ship power plant (SPP), which is either a functional complex of subsystems or the relationship of these complexes, was selected as the object being researched for diagnosing the CTS condition. SPP is a particularly important element of the vessel and affects the performance of all its other elements.

Therefore, the operation of the SPP from the standpoint of reliability must be considered in mutual connection with all its elements.

The operational readiness of the SPP is crucial for the efficient use of the vessel as a means of transport [26-31].

The CTS includes such subsystems:

– fire system (PS);

– control of the main engine (RUGD);

– compressed air (SSV);

– control of propulsion and steering complex (DRK);

– boiler installation (KU);

– power station (PS);

– fire protection system (FPS);

– main engine (GD);

– remote automated control (DAU) of the main engine;

– ballast-drainage (BOS);

– power transmission from the main engine to the propulsion unit (PM);

– emergency drive DRC (AP);

– propulsion and steering complex (DRK);

– control system propulsion and steering complex (SUDRK);

– sanitary water treatment (SSVP).

The structural scheme of the SPP and the interconnection of the subsystems are presented in Fig. 11.

All subsystems can be represented as interrelated components.

Each of the subsystems affects the subsystems of the next level.

All the data received from the sensors of the parameters of the subsystems comes in a software module.

The developed software module provides for the use of the Bayesian trust network, which allows us to determine the probability of the each CTS subsystem components failure.

When analyzing data, their sample is formed for a specific scenario of the CTS subsystems components risk failure.

The data is then interpreted and processed using knowledge acquisition, decision support and rule management.

As a result, they are replenished with new data from the knowledge, data and rule base, which are then fed to the block for analyzing the decisions taken.

In the study of the selected SPP, the risk of subsystem failures was determined during the operation of the ship's power plant 20000 hours.

From the research results it follows that with an increase in the element failure risk at the inlet of the CTS from 0.09 to 0.2, the values of the failures risk all affiliated and lower subsystems of the SPP increase in accordance with the data given in Table. 2

Table 2. Subsystem failure risk SPP

№	The name of the element	Risk of failure on input element 0,09	Risk of failure on input element 0,2
1	PS	0,07	0,16
2	SSV	0,08	0,17
3	RUGD	0,07	0,16
4	SSVP	0,07	0,14
5	DAU	0,07	0,13
6	SE	0,08	0,15
7	GD	0,08	0,14
8	BOS	0,07	0,12
9	AP	0,07	0,12
10	SUDRK	0,07	0,13
11	KU	0,07	0,13
12	PM	0,06	0,1
13	DRK	0,08	0,13

Technical risk is a comprehensive indicator of the CTS components reliability, namely the combination of the hazards occurrence probabilities of a certain class and damages from undesirable events due to imperfection, violation of the rules for technical systems operation
[26-31].

The risk of failure is defined as the product of the CTS component failure probability the damage caused by the failure of the component.

When using the developed IMS software, which provides real-time programmatic distribution and transfer of Big Data to a relational and non-relational database, the load on the local computing server that processes data used to determine the risk of failures of CTS components is reduced by 17.4 - 18.5% compared to the standard use of a computing server with a relational data store.

There was no load on the switch performing the data distribution function in the relational and non-relational databases.

Predicting the use of non-relational storage is more efficient due to the use of a simpler data retrieval system in the storage array and not tracking the constant hierarchy, as is the case in structured arrays.

A positive feature that ensures the operation of the proposed data distribution system is the integration of the software-configured networks concept and Data Mining methods in order to identify the most important diagnostic information for assessing and building a prediction of the CTS state.

This approach allows us to increase the accuracy of estimating the CTS states when selecting data from their critical components. The generated data sets are used in their analysis with an assessment of the CTS components failures risk in order to predict their technical condition.

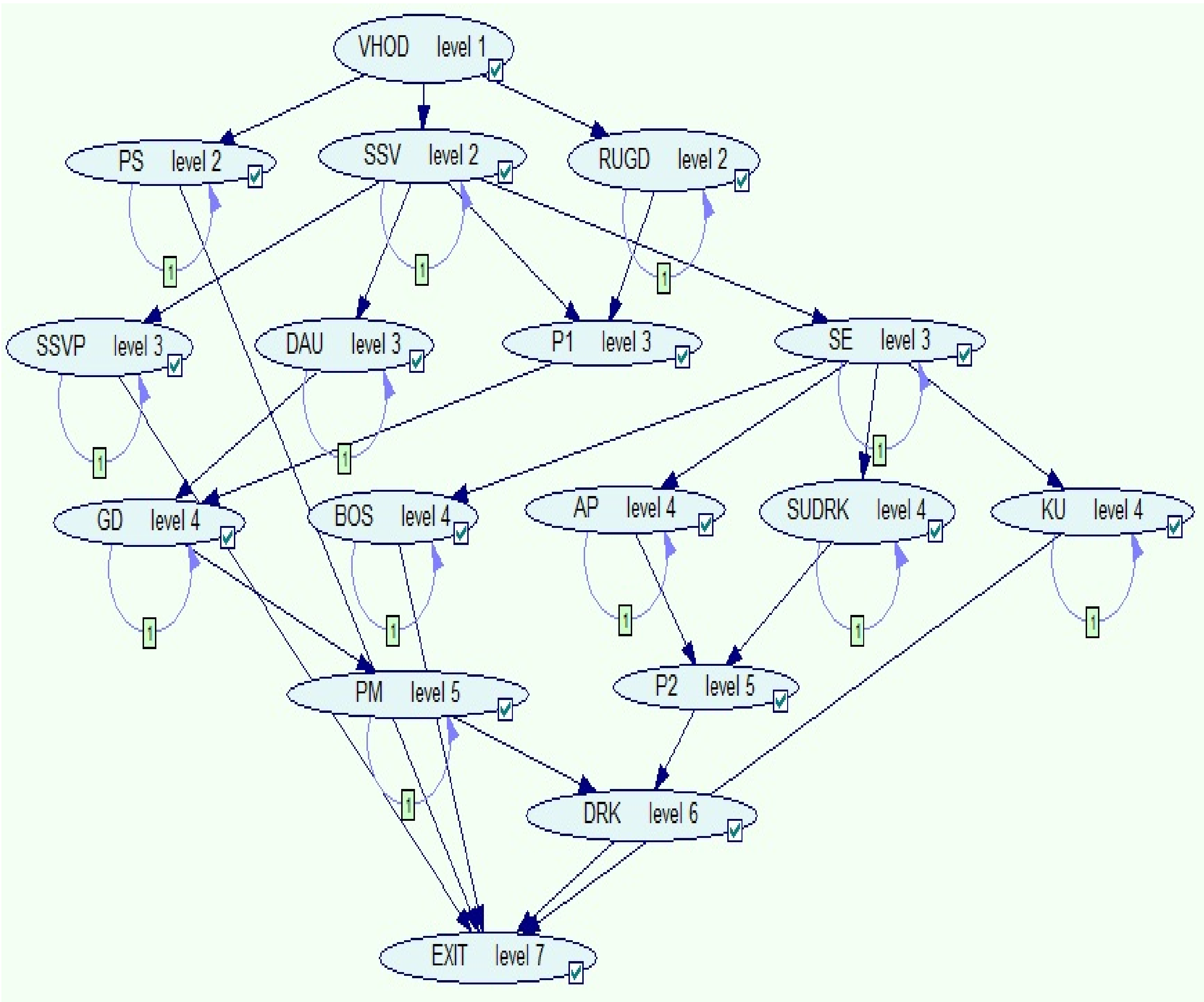


Fig 11. The structure of the relationship subsystems SPP

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

Conclusion

Due to research results it follows that the highest performance of the programmatic data separation process according to the developed method is achieved with distributed execution of computational processes by the developed program in four continuous modes, and the data separation process in non-relational bases for all the experiments performed is faster than for relational data (an average of 8-10 %).

The use of the developed software IMS, providing software distribution and transfer of Big Data in real time to a relational and non-relational database, reduces by 17.4 - 18.5% the load on the local computing server that processes data used to determine the CTS components failures risk compared to the standard use of a computing server with a relational data store.

The use of the developed Big Data management method with data distribution in relational and non-relational databases allows using its results to assess and predict the state of the critically vulnerable CTS components.

References

1. Vychuzhanin, V., & Rudnichenko, N. (2019). "Metody informatsionnykh tekhnologiy v diagnostike sostoyaniya slozhnykh tekhnicheskikh system: мonografiya". [Information technology methods in diagnosing the state of complex technical systems: мonograph], *Ekologiya*, 178 p. [in Russian].

2. Tsyganov, A., Kuzovkin, A., & Shchukin, B. (2010). "Upravleniye dannymi". [Data Control], Moscow, Russian Federation, *Academia*, 256 p. [in Russian].

3. Bigus, G., Daniyev, Yu., Bystrova, N., & Galkin, D. (2014). "Diagnostika tekhnicheskikh ustroystv". [Diagnostics of technical devices], 615 p. [in Russian].

4. Pankratova, N. (2008). "Sistemnyy analiz v dinamike diagnostirovaniya slozhnykh tekhnicheskikh system". [System analysis in the dynamics of diagnosing complex technical systems], *Systematic analysis and information technology*, Ukraine, No.1, pp. 33-49 [in Russian].

5. Vychuzhanin, V. V. (2018). "Raspredelennyy programmnyy kompleks na baze freymvorka APACHE SPARK dlya obrabotki potokovykh BIG DATA ot slozhnykh tekhnicheskikh system". [Distributed software complex based on the APACHE SPARK framework for processing streaming BIG DATA from complex technical systems], *Informatics and mathematical methods in modeling*, Ukraine, Vol. 8, No. 2, pp. 146-154 [in Russian]. Doi: 10.15276/imms.v8. no2.146.

6. Vychuzhanin, V., & Rudnichenko, N. (2014). "Assessment of risks structurally and functionally complex technical systems", *Eastern-European Journal of Enterprise Technologies*, Ukraine, Vol. 1, No. 2, pp.18-22. Doi: 10.15587/1729-4061.2014.19846.

7. Rudnichenko, N., & Vychuzhanin, V. (2013). "Informatsionnaya kognitivnaya model tekhnologicheskoy vzaimozavisimosti slozhnykh tekhnicheskikh system", [Informational cognitive model of technological interdependence of complex technical systems], *Computer Science and Mathematical Methods in Modeling*, Ukraine, No. 3, pp. 240-247 [in Russian].

8. Boyko, V., & Vychuzhanin, V. (2012). "Model otsenki zhivuchesti sudovykh tekhnicheskikh system". [Model for assessing the survivability of ship technical systems], Bulletin of Mykolaiv Shipbuilding University, Ukraine, No. 3, pp. 62-67 [in Russian].

9. Karpenkov, S. (2009). "Sovremennyye sredstva informatsionnykh tekhnologiy". [Modern means of information technology], Moscow, Russian Federation, *Knorus*, 400 p. [in Russian].

10. Arasu, A., Ganti, V., & Kaushik, R. (2006). "Efficient exact set-similarity joins", *Proceedings of the 32nd international conference on Very large data bases*, VLDB '06, VLDB Endowment, pp. 918-929.

11. Kudryavtsev, K., & Korotkov, A. (2012). "Methods of increasing the speed of information search in databases", *LAP Lambert Academic Publishing*, 84 p.

12. Kristofer, D., Ragkhavan, P., & Shyuttse, K. (2011). "Vvedeniye v informatsionnyy poisk". [Introduction to information retrieval], Moscow, Russian Federation, OOO I.D. Williams, 528 p. [in Russian].

13. Tsukert, A. (2001). "Problemy i perspektivy informatsionnogo poiska". [Problems and prospects of information retrieval], News of the Taganrog State University of Radio Engineering, Russian Federation, Vol. 21, No. 3, pp. 194-201 [in Russian].

14. Andreyev, E., Kutsevich, N., & Sinenko, O. (2004). "SCADA-sistemy: vzglyad iznutri" [SCADA-systems: a view from the inside], Moscow, Russian Federation, *RTSoft*, 176 p. [in Russian].

15. Proshin, D., & Gur'yanov, L. (2010). "Problemy vybora instrumental'nykh sredstv postroyeniya SCADA-sistem". [Problems of choice of SCADA-systems building tools], *Informatization*

and Control Systems in Industry, Russian Federation, No. 1(25), pp. 21-25 [in Russian].

16. Yefimov, I., & Soluyanov, D. (2010). "SCADA-sistema TraceMode" [SCADA-system TraceMode], Ulyanovsk, Russian Federation, *UlSTU*, 158 p. [in Russian].

17. Ibrahim, A., Ibrar, Y., & Nor, B. (2015). "The rise of "big data" on cloud computing, Review and open research issues", *Information Systems*, No. 47, pp. 98-115. Doi: 10.1016/j.is.2014.07.006.

18. Wu, X., Zhu, X., Wu, GQ, & Ding, W. (2014). "Data mining with big data", IEEETrans Knowl Data Eng., No. 26 (1), pp. 97-107. Doi: 10.1109/tkde.2013.109.

19. Steven, F. (2014). "Predictive Analytics, Data Mining and Big Data. Myths, Misconceptions and Methods". Basingstoke: Palgrave Macmillan, 15 p. Doi: 10.1057/9781137379283.

20. Huai, Y., Lee, R., Zhang, S., Xia, C. H., & Zhang, X. (2011). "DOT: a matrix model for analyzing, optimizing and deploying software for big data analytics in distributed systems". *In: Proceedings of the ACM Symposium on Cloud Computing*, pp. 4-14 Doi: 10.1145/2038916.2038920.

21. Adibi, J., & Faloutsos, C. (2018). "KDD-2002 Workshop Report". Fractals and Self-similarity in Data Mining: Issue and Approaches URL: http://www.sigkdd.org/sites/default/ files/issues/4-2-2002-12/adibi.pdf (accessed: 16.12.2018).

22. Barbara, D. (2010). "Fractal Mining – Self Similarity-based Clustering and its Applications", Data Mining and Knowledge Discovery Handbook, pp. 573-589. Doi: 10.1007/978-0-387-09823-4_28.

23. David, L. (2016). "CISA Certified Information Systems Auditor Study Guide", 632 p. Doi: 10.1002/9781119419211.

24. Filimonov, P., & Ivanov, M. (2015). "Sovremennyye podkhody k klassifikatsii trafika fizicheskikh kanalov seti Internet". [Modern approaches to the classification of traffic of physical channels of the Internet], *Proceedings of the 18th International Conference "Distributed Computer and Communication Networks: Control, Computing, Communication"* (DCCN-2015), October 19-22, Russian Federation, pp. 466-474 [in Russian].

25. Risso, F., Baldi, M., Morandi, O., Baldini, A., & Monclus, P. (2008) "Lightweight, payload-based traffic classification: An experimental evaluation", In Proc. IEEE ICC, pp. 5869-5875. Doi: 10.1109/icc.2008.1097

26. Shibayeva, N., Shibayev, D., Vychuzhanin, V., & Rudnichenko, N. (2017). "Optimizatsiya otbora i analiza informatsii v raznostrukturnykh khranilishchakh dannykh". [Optimization of the selection and analysis of information in multi-structured data warehouses], *Informatics and mathematical methods in modeling,* Russian Federation, No. 4. pp. 318-324 [in Russian].

27. Andersen, B. (2011). "A Diagnostic System for Remote Real-Time Monitoring of Marine Diesel-Electric Propulsion Systems", Leipzig, 45 p.

28. Krarowski, R. (2014). "Diagnosis modern systems of marine diesel engine", *Journal of KONES Powertrain and Transport*, pp. 191-198. Doi: 10.5604/12314005.1133203.

29. Sørensen, A. (2013). "Marine Control Systems Propulsion and Motion Control of Ships and Ocean Structures" / Asgeir J. Sørensen, 526 p.

30. Vyuzhuchanin, V., & Rudnichenko, N. (2014). "Technical risks of complex complexes of functionally interconnected structural components of ship power plants", Odesky National Nautical University, Ukraine, No. 2 (40), pp. 68-77 [in Russian].

31. Changben, J., Brian, L., & David, R. (2002). "Ship Hull and Machinery Optimization using Physics Based Design Software", *Marine Technology*, Vol. 39, No. 2, pp. 109-117.

Received 15.05.2019

УДК 004.67

[1]**Шібаев, Денис Сергійович**, аспірат кафедри технічної кібернетики та інформаційних технологій,
E-mail: denscreamer@gmail.com, ORCID: 0000-0002-3260-5843
[2]**Вичужанін, Володимир Вікторович**, д-р техніч. наук, професор, завідувач кафедри інформаційних технологій, E-mail: vint532@yandex.ua, Scopus ID: 57193025809, ORCID: 0000-0002-6302-1832
[2]**Рудніченко, Микола Дмитрович**, канд. техніч. наук, доцент кафедри інформаційних технологій, E-mail: nickolay.rud@gmail.com, ORCID: 0000-0002-7343-8076
[2]**Шібаева, Наталя Олегівна**, канд. техніч. наук, доцент кафедри інформаційних технологій, E-mail: nati.shibaeva@gmail.com, ORCID: 0000-0002-7869-9953
[3]**Отрадская, Тетяна Василівна**, канд. техніч. наук, директор Одеського коледжу комп'ютерних технологій «Сервер», E-mail: tv_61@ukr.net, ORCID: 0000-0002-5808-5647
[1]Одеський національний морський університет, вул. Мечникова, 34, Одеса, Україна, 65029
[2]Одеський національний політехнічний університет, проспект Шевченка,1, Одеса, Україна, 65044
[3]Одеський коледж комп'ютерних технологій «Сервер», Польський спуск, 1, Одеса, Україна, 65026

УПРАВЛІННЯ ДАНИМИ ПРИ ДІАГНОСТИЦІ І ПРОГНОЗУВАННІ СТАНУ СКЛАДНИХ ТЕХНІЧНИХ СИСТЕМ

Анотація. *Проведено аналіз методів управління Big Data. З метою отримання своєчасних результатів аналізу стану складних технічних систем на основі встановленого згідно з нормативною документацією переліку параметрів, що мають першорядне (для критично важливих компонентів) і другорядне значення при діагностуванні стану компонентів, що забезпечують функціонування складних технічних систем необхідно розробити метод управління даними, що дозволяє з високою швидкодією і без втрат розділяти і передавати Big Data від ІВС в реляційні і нереляційні бази даних. Запропоновано метод, що забезпечує розподіл даних, що надходять з інформаційно-вимірювальних систем в реляційні і нереляційні бази даних при діагностиці та прогнозуванні стану складних технічних систем. Обґрунтовано доцільність використання концепції Data Mining в SCADA системах для управління Big Data. Розроблено алгоритми передачі, розподілу та аналізу даних в інформаційно-вимірювальної системи при діагностиці та прогнозуванні стану складних технічних систем. Розроблено схему передачі даних в пристроях, що використовують шину CAN. В основу запропонованого методу управління Big Data при діагностиці та прогнозуванні стану складних технічних систем покладено забезпечення динамічного розподілу даних в інформаційно-вимірювальної системи з урахуванням вимог, що пред'являються до використовуваної структурі локально-обчислювальної мережі. Метод ґрунтується на застосуванні принципів побудови програмно-конфігуруються мереж, що дозволяють управляти мережею за рахунок використання результатів аналізу потоків даних, що проходять через вузлові мережеві пристрої. Розроблена програмна реалізації системи розподілу даних в локальній мережі з використанням принципу аналізу мережевих пакетів при їх надходженні на комутаційні вузли інформаційно-вимірювальної системи. Розроблено систему програмної логіки розподілу даних з інформаційно-вимірювальних систем, що передаються по локальних мережах або по комутаційної шині CAN. З проведених досліджень випливає, що найбільша продуктивність процесу поділу даних по розробленим методом досягається при розподіленому виконанні обчислювальних процесів розробленої програми в чотирьох потоковому режимі, причому процес поділу даних в нереляційні бази для всіх проведених експериментів виконується швидше, ніж для даних в реляційні бази. Застосування розробленого методу управління Big Data з розподілом даних в реляційні і нереляційні бази даних забезпечує підвищення швидкодії інформаційно-вимірювальної системи при діагностиці та прогнозуванні стану складних технічних систем. Дозволяє прогнозувати технічний стан критично-вразливих компонентів систем при їх короткочасному знаходженні в аварійному стані, а також здійснювати довгостроковий прогноз оцінки технічного стану всієї складної технічної системи. Використання програмних розподільників переданої інформації забезпечує підвищення швидкодії інформаційно-вимірювальної системи при діагностиці та прогнозуванні стану складних технічних систем, тим самим забезпечуючи своєчасну оцінку стану критично важливих компонентів складних технічних систем, вихід з ладу яких впливає на працездатність систем.*

Ключові слова: *управління даними; Big Data; DataMining; SCADA-системи; реляційні і нереляційні бази даних; інформаційно-вимірювальна система; розподіл даних; складні технічні системи; діагностика; прогнозування стану технічних систем*

УДК 004.67

[1]**Шибаев, Денис Сергеевич,** аспират кафедры технической кибернетики и информационных технологий, E-mail: denscreamer@gmail.com, ORCID: 0000-0002-3260-5843
[2]**Вычужанин, Владимир Викторович** д-р технич. наук, профессор, заведующий кафедрой информационных технологий, E-mail: vint532@yandex.ua, Scopus ID: 57193025809, ORCID: 0000-0002-6302-1832
[2]**Рудниченко, Николай Дмитриевич,** канд. технич. наук, доцент кафедры информационных технологий, E-mail: nickolay.rud@gmail.com, ORCID: 0000-0002-7343-8076
[2]**Шибаева, Наталья Олеговна,** канд. технич. наук, доцент кафедры информационных технологий, E-mail: nati.shibaeva@gmail.com, ORCID: 0000-0002-7869-9953
[3]**Отрадская, Татьяна Васильевна**, канд. технич. наук, директор Одесского колледжа компьютерных технологий «Сервер», E-mail: tv_61@ukr.net, ORCID: 0000-0002-5808-5647
[1] Одесский национальный морской университет, ул. Мечникова, 34, Одесса, Украина, 65029
[2] Одесский национальный политехнический университет, проспект Шевченко, 1, Одесса, Украина, 65044
[3] Одесский колледж компьютерных технологий «Сервер», Польский спуск, 1, Одесса, Украина, 65026

УПРАВЛЕНИЕ ДАННЫМИ ПРИ ДИАГНОСТИКЕ И ПРОГНОЗИРОВАНИИ СОСТОЯНИЯ СЛОЖНЫХ ТЕХНИЧЕСКИХ СИСТЕМ

Аннотация. *Проведен анализ методов управления Big Data. В целях получения своевременных результатов анализа состояния сложных технических систем на основе установленного по нормативной документации перечня параметров, имеющих первостепенное (для критически важных компонентов) и второстепенное значение при диагностировании состояния компонентов, обеспечивающих функционирование сложных технических систем необходимо разработать метод управления данными, позволяющий с высоким быстродействием и без потерь разделять и передавать Big Data от ИИС в реляционные и нереляционные базы данных. Предложен метод, обеспечивающий распределение данных, поступающих с информационно-измерительных систем в реляционные и нереляционные базы данных при диагностике и прогнозировании состояния сложных технических систем. Обоснована целесообразность использования концепции Data Mining в SCADA системах для управления Big Data. Разработаны алгоритмы передачи, распределения и анализа данных в информационно-измерительной системе при диагностике и прогнозировании состояния сложных технических систем. Разработана схема передачи данных в устройствах, использующих шину CAN. В основу предложенного метода управления Big Data при диагностике и прогнозировании состояния сложных технических систем положено обеспечение динамического распределения данных в информационно-измерительной системе с учетом требований, предъявляемых к используемой структуре локально-вычислительной сети. Метод основывается на применении принципов построения программно-конфигурируемых сетей, позволяющих управлять сетью за счет использования результатов анализа потоков данных, проходящих через узловые сетевые устройства. Разработана программная реализации системы распределения данных в локальной сети с использованием принципа анализа сетевых пакетов при их поступлении на коммутационные узлы информационно-измерительной системы. Разработана система программной логики распределения данных с информационно-измерительных систем, передаваемых по локальным сетям либо по коммутационной шине CAN. Из проведенных исследований следует, что наибольшая производительность процесса разделения данных по разработанному методу достигается при распределенном выполнении вычислительных процессов разработанной программой в четырех поточном режиме, причем процесс разделения данных в нереляционные базы для всех проведенных экспериментов выполняется быстрее, чем для данных в реляционные базы. Применение разработанного метода управления Big Data с распределением данных в реляционные и нереляционные базы данных обеспечивает повышение быстродействия информационно-измерительной системы при диагностике и прогнозировании состояния* *сложных технических систем. Позволяет прогнозировать техническое состояние критически-уязвимых компонентов систем при их кратковременном нахождении в аварийном состоянии, а также осуществлять долгосрочный прогноз оценки технического состояния всей сложной технической системы. Использование программных распределителей передаваемой информации обеспечивает повышение быстродействия информационно-измерительной системы при диагностике и прогнозировании состояния сложных технических систем, тем самым обеспечивая своевременную оценку состояния критически важных компонентов сложных технических систем, выход из строя которых влияет на работоспособность систем.*

Ключевые слова: *управление данными; Big Data; DataMining; SCADA-системы; реляционные и нереляционные базы данных; информационно-измерительная система; распределение данных; сложные технические системы; диагностика; прогнозирование состояния технических систем*

Denis Sergeyevich Shibaev
Graduate Student of the Department of Technical Cybernetics and Information Technology
Scientific interests:
Information and technological support of functional and topological structures of distributed systems according to the survivability criteria of ship systems

Vladimir Viktorovich Vyuzhuzhanin
Doctor of Technical Sciences, Professor, Head of the Department of Information Technologies
Scientific interests:
Information and technological support of functional and topological structures of distributed systems

Nikolay Dmitrievich Rudnichenko
Candidate of Technical Sciences, Associate Professor at the Department of Information Technologies
Scientific interests:
Methods and tools for intellectual analysis of large amounts of data
Information and technological support of functional and topological structures of distributed systems according to the survivability criteria of ship systems
Development of software systems for monitoring the technical condition of complex systems

Natalia Olegovna Shibaeva
Candidate of Technical Sciences, Associate Professor, Department of Information Technologies
Scientific interests:
Information and technological support of functional and topological structures of distributed systems according to the survivability criteria of ship systems.
Development of software for monitoring the technical condition of complex systems.

Tatyana Vasilevna Otradskaya
Candidate of Technical Sciences, Director of Odessa Computer Technology College “Server”
Scientific interests:
Information and technological support of functional and topological structures of distributed systems according to the survivability criteria of ship systems.
Development of software for monitoring the technical condition of complex systems.

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)