

kurtosis and others. On the base of the characteristics received as a result of modeling it is possible to draw a conclusion about financial stability of the bank, to estimate risk of the introduction at a zone of financial unreliability (a risk zone). For an introduction risk assessment in a zone of unreliability it is offered to use coefficient of risk of an exit of values of the capital below critical level.

The stochastic model of dynamics of the bank's capital should be used if there is information on nature of capital allocation of bank and its increments.

Otherwise, if it is possible to transform temporary ranks to a stationary type – it is recommended to use a simulation model of dynamics of the commercial bank's capital.

The offered models allow to solve actual problems of an assessment, the analysis and capital management of commercial bank, in particular, to research the size of the bank's capital in dynamics, to estimate risk of bankruptcy of bank, to develop the strategy of investment of means of bank to satisfy the level of acceptable risk.

REFERENCES

1. *Реннер А. Г., Ленерт А. Г.* Моделирование банковской паники на примере АКБ «Форштадт» (ЗАО) // «Экономика и современный менеджмент». 2013. № 30. С. 21-26.
2. *Ситникова А. Ю.* Формирование денежных потоков коммерческого банка как участника инвестиционного рынка // Модели. системы. сети в экономике, технике, природе и обществе. 2013. № 1 (5). С. 94-98.
3. *Проказа Т. В.* Дисконтирование денежных потоков коммерческого банка // Экономика промышленности. 2012. № 3-4. С. 81-86.
4. *Лученко К. Л.* Экономико-математическая модель денежных потоков банка при кредитовании малого предприятия // Научно-исследовательский финансовый институт. Финансовый журнал. 2011. № 2. С. 105-116.
5. *Сорокина М. Г.* Оценка эффективности увеличения собственного капитала коммерческого банка за счет внутренних источников // Экономические науки. 2011. № 20 (83). С. 168-171.
6. *Туренова Е. Л.* Исследование процесса изменения капитала при непрерывном расходовании средств, зависящем от их объема на активном счете // Вестник Томского государственного университета. 2006. № 290. С. 203-208.
7. *Ерофеев В. Т.* Уравнения с частными производными и математические модели в экономике М. : Книжный дом «ЛИБРОКОМ», 2013. 248 с.
8. *Волков И. К., Зуев С. М., Цветкова Г. М.* Случайные процессы. МГТУ им. Н.Э. Баумана, 2006. 448 с.
9. *Яркова О. Н., Реннер А. Г., Буреш А. И.* Моделирование инвестиционного портфеля страховой компании в статике и динамике: монография. Самара : Изд-во СамНЦ РАН, 2014. 207 с.

DEVELOPMENT OF TOOLS FOR PROCESSING BIG DATA

M. G. Tindova

*Saratov social-economic Institute of «Russian economic
University G.V. Plekhanov», Russia
E-mail: mtindova@mail.ru*

The author has described the creation process means processing of large data generated by the operation and service of vending machines. In the development processing, the author analyzed the process of gathering information from vending machines, identified problems in the storage and processing of information, and formulated the requirements to the developed means of processing big-data. The author showed the process of modeling, which identified the main classes of the system, their attributes, methods and relationships between them, chose software and modeling tools have made the terms of reference. On the basis of the project work we developed a tool for the development of the data processing system to service the vending machines.

In modern post-industrial society the main value is information. The relevance information, the speed and accuracy of processing are the Foundation of a successful business. In connection with the development of information technology, with the advent of affordable computing devices and high power software implementing optimal algorithms, processing of large data volumes has become affordable for most organizations. The analysis of large amounts of information allows to obtain non-obvious information about the analyzed object and find hidden connections. But often the amount of information that must be processed is very large and it is necessary to use special means of information processing.

Therefore, the aim of this work is to study the problems of processing large scale data and development of a system for processing big data. The object of study are data on the work vending machines.

Big-data is data of such volumes, the processing of which standard tools are too complicated and takes too much time. Most often it is unstructured or semi-structured data large volumes that must quickly process, resulting to the means of their processing requirements parallelism, scalability and fault tolerance [1].

In this is regard, it became necessary to use technologies for distributed computing that effectively address the following issues affecting the quality of results: first, the processing speed, which depends on the performance of the processor, speed write/read from disk, amount of RAM. To solve this problem is clustering units, i.e. parallel connection of several computers used as a single, unified computer resource [2].

Second is reliability, because the more machinery is used in the cluster, the greater the probability of failure of a node, data processing system, big data must correctly respond to the failure of a certain number of machines and such events should not stop the operation of the system.

By far the most popular software tools for processing big data are NoSQLбазы data, specialized frameworks (Hadoop, Spark), and a parallel DBMS. The most well-known hardware-software complexes for processing big data are the Aster MapReduce Appliance for Teradata Corporation, the Big Data Appliance, Oracle, Greenplum Appliance of EMC Corporation.

Feature of NoSQL technologies is the idea of unlimited scale and abandoning consistency in favor of throughput. The basic concept of parallel DBMS is a fragment parallelism, which implies the horizontal fragmentation of each table of the database on disks of the cluster system. In this case, one and the same query is executed concurrently by each agent over "their" fragments of the database tables, and then the resulting partial results are merged into the result table. Frameworks for distributed data

processing is divided into processing technology, streaming data and technology for batch processing. Stream processing involves the receipt of result in real time. Batch treatment is a treatment of historical data obtained over a certain period of time [3].

The purpose of this study, and based on the fact that the object is the data of vending machines, i.e. devices engaged in small retail trade in goods and services, payment and delivery which are enforced through technical devices that do not require direct participation of salesman [4], that the developed information processing system should be the basis of online monitoring devices. The service user should be provided information about the current state of his machines, a log of events, the number of bills in the bill acceptor, and other information. Thus, the task of processing the large number of vending machines is reduced to the problem of stream processing semi-structured data, which is a special case of the problem of processing big-data. To solve this problem it is necessary to develop scalable and fault-tolerant software tool that allows you to extend the functionality of the service if necessary.

The process developed product consists of a collection of data that is sent vending machines, the processing and storage of processing results in a format suitable for display in a web application, and analysis. Data collection should be allocated a separate server on which you installed a program that receives HTTP requests and writes them in text format in a distributed file system (HDFS). To ensure fault tolerance, scalability and adequate bandwidth, you need to deploy several of these data collectors and to distribute the load between them using a load balancer.

Data processing will be performed using a distributed system parallel computing that reads data from the directory into which the collector writes the data and puts the results of processing to the database from which data is read by the web application.

The process of the system is as follows: vending machines send requests to a service address. The load balancer receives requests and distributes them among multiple collectors. The data collector is a service that accepts requests, interprets them, checks in the database entry about the registration, making the request, and if the record exists, and stores the request in a distributed file system (HDFS). The data processing application continuously reads data from HDFS and begins processing them as soon as new data appear. The data is processed and stored in the database in order to read out their web application.

To develop the data collector we will use NodeJS platform based on JavaScript that adds the possibility of interaction with input devices and output through your API (written in C++), the ability to connect other external libraries written in different languages, providing challenges to them from JavaScript code.

To distribute the load between multiple servers, data collectors will be used Nginx is a HTTP server and reverse proxy server, mail proxy server, and TCP/UDP proxy server for General use.

To provide the service "Personal Cabinet" data processing vending machines must occur in real time. It is necessary to use the technology to produce streaming data processing. In this regard, the best solution is to use Apache Spark, which is a General-purpose framework for processing big data. With it can be treated as coming at the moment (streaming) and historical (batch) data.

For reliable distributed storage of the source data, i.e. data coming from the vending machines, will use a specialized distributed file system Hadoop Distributed File System (HDFS), which is a hierarchical file system can support nested directories.

Thus, for the organization of a distributed stream processing we use the Apache Spark to work on the basis of Apache Hadoop. In this scenario, use Spark uses YARN as a dispenser of resources and HDFS as a distributed file system. In the end, the subsystem of data processing based on Apache Hadoop, which is required for distributed data storage and cluster resource management, and Apache Spark, are used as a framework for data processing.

Another aspect of the development of processing big data is to develop data storage system for a web application, which should ensure the maximum availability of the data to be easily extensible and provide the ability to change existing data structures without loss of efficiency. So we will use MongoDB – document-oriented database management system (DBMS) open source, does not require the description of the table schema.

The result is a technical specification [5] to develop data processing system for monitoring service of vending machines can be represented in the following form:

- *The purpose and objectives of the system:* the system is designed for processing logs of vending machines in real time to provide information about the current and past status of machines, and calculating statistics regarding use of the machines.

- *Functional system requirements:* the system must contain the following functional subsystems: data acquisition subsystem, storage subsystem, and a subsystem of data processing.

- *Reliability requirements:* the system must ensure data processing 50,000 machines, the maximum waiting time is 8 minutes, and the system should ensure correct handling of emergency situations, caused by incorrect format or invalid values for the input data. In these cases, the system should automatically process incorrect data and continue the normal work.

- *Requirements on security of information:* security of information should be provided in case of following events: the failure of communication and network equipment; loss of connections in the Internet; failure or loss of power supply (including temporary), voltage surges in the mains; the failure of one or more computing machines.

- *Feature requirements subsystems of data collection:* data acquisition subsystem must implement the following functions: receiving queries from the system for attached devices; write queries in JSON format in HDFS; processing HTTP requests with the POST method in the body which contains the text version of the transaction in the format EVA-DTS.

- *Requirements to the functions of the storage subsystem:* the storage subsystem needs to provide continuous availability of data for the web application "my account" and reliable storage with replication.

- *Feature requirements data processing sub-systems:* the subsystem of data processing must implement the following functions: structuring data and bringing

them to the required format.

As a conclusion I want to note that the search for effective solutions to problems of processing of big-data is an important task and that development of systems handling such data should be considered as features of processed data and scenarios of their use.

REFERENCES

1. *Vendrov A. M.* The software Design of economic information systems. M. : "finances and statistics", 2006. 376 p.
2. *Argerich L.* Professional PHP programming, 2nd edition. SPb. : Symbol-Plus, 2003. 1048 p.
3. *Tindova M. G., Zhuravskaya K. G.* Development applications for maintenance loans // *Evraziyskiy Union of scientists*. 2015. № 8 (17). Part 1. P. 83-85.
4. *Tindova M. G.* the Use of modern information technologies in the development of instrumental means of assessment of natural resources // *Information security of regions*. 2013. № 1. P. 47.
5. *Orlov S. A.* Technology development programs. SPb.: Piter, 2004. 609 p.

MODELING MULTIVALUED DYNAMIC SERIES OF FINANCIAL INDEXES ON THE BASIS OF MINIMAX APPROXIMATION IN THE HAUSDORFF METRIC WITH THE CONSTRAINT

I. Yu. Vygodchikova

Saratov State University named after N. G. Chernyshevsky, Russia

E-mail: irinavigod@yandex.ru, vlgodchikovaiy@info.sgu.ru

The problems of modeling a time series by various methods were studied by many authors, including, D.A. Dickey, W.A. Fuller [1], C.A. Sims, J.H. Stock, M.W. Watson [2], Schumaker L. L. [3]. However, the question of mathematical modeling of the multivalued dynamic series regardless of the distribution law, with additional restrictions, little studied in the literature. This paper examines the model of approximation of the multivalued time series, represented as a series of ranges of numerical values of the indicator of financial market, with constraints to approximating function.

For approximation of the dynamic number of applicable optimization problem of minimizing the maximum of the Hausdorff distances between the ranges of the dynamic series and the values of the approximating function.

The purpose of this paper is to develop a methodology of mathematical modeling of multivalued dynamic series, in the presence of restrictions on the values of the approximating function, the development of rational algorithm.

Introduction. Will call by the multivalued dynamic series a sequence ordered in time ranges of numerical values of some object. In the literature was not considered the criterion of uniform approximation to be used for multivalued mappings using the Hausdorff distance [4], including the restrictions on the approximating function. One of the effective methods for the analysis of unambiguous time series of different nature is the criterion of uniform approximation by Chebyshev [5]. However, when considering the ranks of the specified ranges, there is a problem with the rationale for the selection of the point inside the range, which will allow you to achieve