

ABSTRACT

The thesis is presented in 73 pages. It contains 2 appendixes and bibliography of 15 references. 13 figures and 5 table are given in the thesis.

Topic relevance. Today, people depend heavily on computers. This technical devices help in the calculations, processing and storage of important data. Since data is constantly supplemented and increasing in volume, it is important that computer programs performing data processing as soon as possible. For storing and processing large amounts of data gained popularity Big Data. But it is important to research and identify it is the fastest, most convenient platform components and components that will help to minimize costs. It is important to examine the method of minimizing the processing time search platform Hadoop. This method makes it possible to investigate the behavior of input data and very large volume processing speed on the platform. There are currently no Hadoop-friendly interface for the user.

Thesis connection to scientific programs, plans, and topics. The thesis was prepared according to the scientific research plan of the Applied Mathematics Department of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute.”

Research goal and objectives. The aim of the thesis is to develop a method to minimize processing time search platform HADOOP.

To accomplish this goal, the following objectives were reached:

- organize existing methods to minimize processing time;
- develop a model to minimize processing time;
- develop methods for solving the problem of minimizing the processing time;
- conduct a pilot study using data obtained.

Object of research is Hadoop is a platform for handling Big Data; querying of objects Hadoop ecosystem, such as Cloudera Impala, Apache Drill, Apache Hive, Apache Tez, Apache Spark, Cloudera Hue; various types such as parquet, snappy, gzip, bzip2.

Subject of research is a method of minimizing processing time demand for Hadoop platform to speed up the result of the query using Big Data.

Methods of research. To solve the task, the following methods were used: techniques to minimize the time (to develop models of treatment); forecasting methods (for the development of methods for solving the problem of forecasting the speed components); methods of the theory of algorithms and programming (for software implementation of the developed algorithms).

Scientific contribution consists of the following:

- first time the task of minimizing data platform Hadoop, which, unlike the existing ones, involves relying on the weather select the fastest component for processing the necessary data;
- improved forecasting methods, which, unlike existing data account for ambiguity, allowing to provide better and more accurate prediction;
- improved user interface platform, which differ from the present because it has intuitive controls.

Practical value of obtained results. The methods that can be used while minimizing processing time data. The methods, mathematical and software to minimize time data simplify software and time, contribute to quick fixes problems.

Approbation of the thesis results. Basic ideas and results of the research were presented at:

- Kulyk R.Y. research and comparative analysis of the Hadoop platform components R.Y. Kulyk // "Modern trends of development of science" (g Kharkov, May 26-27, 2017 g). - Kherson: Publishing House "Helvetyka" in 2017.
- Kulyk R.Y. Methodological aspects of processing time forecast demand for Hadoop platform R.Y. Kulyk // "Modern trends of development of science" (g Kharkov, May 26-27, 2017 g). - Kherson: Publishing House "Helvetyka" in 2017.

Keywords: Big Data, Hadoop, database, Cloudera Impala, Apache Drill, Apache Hive, Apache Tez, Apache Spark, Cloudera Hue.