ABSTRACT

Master dissertation: 65 pages, 19 figures, 7 tables, 1 application, 15 sources.

The relevance. Global computerization and progress in microelectronic and information technology causes enlargement in information quantity. Total amount of data in 2012 is more than 1,8 zetabytes(1,8 trillion GB). International Data Corporation researched that this amount is doubled every two years[9]. Ninety percent of world data was created in recent two years. Every day we create approximately 2,5 milliards GB. Saving of such amount of information needs a lot of resources and efforts. However, the biggest problem is not saving but processing. IDC prognoses that in 2020 only 35% of world data will be useful. So we need a mechanism to make data simplification for further analysis. One of such methods is cluster analysis.

Clasterization is used for data compression. We could cut down some data because objects inside each cluster is very similar so we could consider them like one object.

So automation and high quality cluster analysis in acceptable time without previous structuring is the task that is worthy of attention and research.

Dissertation is devoted of development of modification of known clasterization algorithms by using parallel calculation on graphical processor unit.

P**urpose and objectives of the study.** The goal is development of improved clasterization algorithms by using Nvidia Cuda parallel calculation to fasten them. To reach the goal it is needed to solve following tasks:

– Analyze modern methods to work with big data and clasterization;

– Analyze Nvidia CUDA technology for parallel computation on graphical processor units;

– Develop improved clasterization algorithms by using Nvidia Cuda parallel calculation;

– Develop application for data clasterization with implementation of chosen algorithms and their modifications;

– Conduct research on the effectiveness of the algorithms by testing.

The object of study is processes of data clasterization.

The subject of study is performance and effectiveness of cluster analysis.

Relationship with scientific programs, plans and themes. The work was carried out at the Department of Computer-Aided Management and Data Processing Systems of the National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute» within the framework of the theme «Methods and Technologies for High-Performance Computing and Processing of Big Data». State registration number 0117U000924.

Scientific novelty of the results is development and implementation of modification of known clasterization algorithms by using parallel calculation on Nvidia graphical processor unit.

Publications. Work results are published in conference abstracts of «INFORMATICS AND COMPUTER TECHNOLOGY – ICT-2018» and «INTELECTUAL SYSTEM OF DECISION-MAKING AND PROBLEMS OF COMPUTATIONAL INTELLIGENCE ».

CLASTER ANALYSIS, PARALLEL CALCULATIONS, GRAPHICAL PROCESSOR UNIT, NVIDIA CUDA