лучших библиотек компьютерного зрения.

Делая выводы, можно сказать, что наиболее дефинитивной является библиотека OpenCV.

ЛИТЕРАТУРА

1. Bradski G.R. Learning OpenCV/ Dr. Gary Rost Bradski, Adrian Kaehler — Изд.: O'Reilly Media, Inc., 2008. — 556 с.

2. Laganière R. OpenCV 2 Computer Vision Application Programming Cookbook/ R.Laganière — Изд.: Packt Publishing, 2011. — 298 с.

3. D.L. Baggio Mastering OpenCV with Practical Computer Vision Projects/ Baggio D. L., Emami S., Escrivá D.M. — Изд.: P

4. LTI-lib Homepage [Electronic resource]. – Mode of access: http://ltilib.sourceforge.net/doc/homepage/index.shtml – Date of access: 15.01.2018.

5. The VXL Homepage documentation [Electronic resource]. – Mode of access: http://vxl.sourceforge.net/#docs – Date of access: 15.01.2018.

UDC 004.27

Mohamed Ahmad El Seblani, PhD student;
N. Ghilyak, PhD, associated prof.;
N. Savos, student
(BSTU, Minsk)

MODIFICATION OF THE DISTRIBUTED COMPUTING MODEL MAPREDUCE

Big Data is the amount of data just beyond technologies capability to store, manage and process efficiently.

Organizations have access to a wealth of information. They can't get value out of it because most of it is sitting in its most raw form or in a semistructured or unstructured form. As one of the most "hyped" terms in the market today, there is no consensus as to how to define big data.

It is true that data analysis and in many cases is an advanced analytics. This also forms the basis for the most used definition of big data, the three V'S:

− velocity: Large amounts of data from transactions with high refresh rate resulting in data streams coming at great speed and the time to act on the basis of these data streams will often be very short;

− variety: Data come from different sources. data can come in various format such as transaction and log data, structured data , semistructured data , unstructured data, images, video streams, and more;

– volume: Large amounts of data , from datasets with sizes of terabytes to zettabyte.

Many enterprises are investing in these big data technologies in order to derive valuable business insights from their stores of structured and unstructured data.

We consider the method MapReduce of distributing the problem to several nodes. Each node processes the data stored on this node. Consists of the phases created by the tugboat MapReduce. Between the Map and the Decrement is Shuffle and sort. Enter a key / value pair. The Map task takes a set of data and converts it into another set of data, where individual elements are broken down into tuples (key-value pairs). The Reduce task takes the output from the Map as an input and combines those data tuples (key-value pairs) into a smaller set of tuples.

The idea appeared in the analysis of MapReduce. And what if we reduce the number of steps, there by gaining in time?

In the case of our script, we skip the first two steps (Splitting and Mapping), and immediately go to Shuffling and Reducing. From our point of view, this will allow us to win a little bit in one of the "three V" Big-Dates - in speed, in comparison with MapReduce.

In the syntax of the Python language on which our script is based, the Map function is called Dict, and the Reduce-Map function.

Another important advantage of our script is the language in which it is written. Python has a huge number of standard and additional libraries (modules) that make our script adaptable for any tasks.

Including work with different types of databases. PostgreSQL, SQLServer, MySQL, SQLite3, Reddis and even many unpopular ones - work with almost all of them is supported by Python.

The aim of the further work with big data is the development of algorithm for processing large data scoring model.

REFERENCES

1. Konstantin B. Optimizations in computing the Duquenne–Guigues basis of implications / B. Konstantin // Annals of Mathematics and Artificial Intelligence. 2014. Vol. 70. No. 1-2. P. 5-24. doi

2. Obiedkov S. Modeling ceteris paribus preferences in formal concept analysis, in: Formal Concept Analysis / S.Obiedkov //Ed. by P. Cellier, F. Distel, B. Ganter. Vol. 7880. Berlin, Heidelberg : Springer, 2013. P. 188-202.