

Distributed data processing method for extracting knowledge

V. Ababii, V. Sudacevschi, V. Lazu, S. Munteanu and V. Ungureanu

Department of Informatics and Systems Engineering, Technical University of Moldova, Chisinau, Republic of Moldova

e-mail: victor.ababii@calc.utm.md

Distributed computing [1] presents one of the most efficient data processing and storage methods. Examples of applications and development of distributed computing can be mentioned as Cloud Computing [2] and Big Data [3], which offer a set of services and computing applications, access to information and data storage. The research aims to develop a distributed data processing method for data located on a set of servers using mobile program codes that are teleported from Users (Agents) to the Server, where large amounts of data are stored (Big Data), to extract the necessary information or knowledge.

Let's define a set of Agents $\mathbf{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N\}$ which activates in the environment $\mathbf{E}(\mathbf{X}) \in \mathbf{R}^K$, where K is the dimension of space and $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K\}$ - the state of the activity environment. For each Agent \mathbf{a}_i is defined the strategy $\mathbf{S}_i = \bigcup_{j=1}^{J_i} \mathbf{O}_{i,j}, i = 1, N$, where

$\mathbf{O}_{i,j}, j = 1, J_i$ is the set of Objects (programs code) that solve the strategy \mathbf{S}_i , and $\bigcap_{i=1}^N \mathbf{S}_i \neq \emptyset$. In the space \mathbf{E} is also defined the set of Servers $\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_L\}$ that have sufficient computing resources and data storage $\mathbf{D} = \{\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_L\}$, where $\mathbf{d}_l = \mathbf{f}_l(\mathbf{X})$ is the quantity of data stored on the Server \mathbf{c}_l that determines the state of interest of services offered by server to the activity environment \mathbf{E} . Each Server $\mathbf{c}_l, l = 1, L$ is ready to host at necessity, Objects $\mathbf{O}_{i,j}, i = 1, \dots, N, j = 1, \dots, J_i$, to execute and to return the result (knowledge) to Agent which are transmitted (submitted) to the Object to execution.

The distributed data processing method based on the mobile program code will be used to solve the problem of searching in large amount of information (Cloud Storage Systems, Big Data Systems) for automatic extraction of new knowledge.

Bibliography

- [1] Ghosh, Sukumar, *Distributed systems: an algorithmic approach*, University of Iowa, Chapman & Hall/CRC, 390p., ISBN:1-58488-564-5.
- [2] Lewis, Grace, *Basics About Cloud Computing*, <http://www.sei.cmu.edu/library/abstracts/whitepapers/cloudcomputingbasics.cfm> (2010).
- [3] Mark A. Beyer and Douglas Laney, *The Importance of 'Big Data': A Definition*, Gartner, (2012).