

УДК 004.93

Alexey Kolodka¹, Gennadii Snizhnoi², Valentin Pogosov²

¹ consultant “Easy Online Solutions“ LTD

² professor NU “Zaporizhzhia Polytechnic”

INNOVATION CLUSTERS FOR PROCESSING AND STORING DATA

Clusters can have different hardware and virtualized solutions with various operating systems and custom software configurations solving a wide range of tasks [1]. The level of cluster performance is highly dependent on the software used for its build. The proposed structure of a highly loaded cluster is using only free software and widely distributed hardware components.

The cluster structure includes networking and balancing systems, a compute module, a database, and data storage subsystems. The networking systems (LAN, VRRP) provide connection of all components within the cluster and its connection with the world via the Internet. The balancing system is represented by two nodes based on Keepalived software connected to each other in Linux Virtual Server (LVS) at the 4th level of the OSI network model to provide a high level of performance and the VRRP monitoring protocol to ensure fault tolerance in case of failure of one of the balancing nodes. The computing system is represented by nodes with Nginx software to provide the ability of information transfer at level 7 of the OSI network model to compute blocks. The database subsystem is represented by multiple nodes configured with the MySQL Galera Cluster software solution. The data storage system is represented by multiple nodes configured as a software object storage network with a CEPH software solution. While the testing were used: Arista 7050T-64 switch, 64 ports with a speed of 10 Gbit/s; 2 balancing nodes with configuration: Dual Intel (R) Xeon® CPU E5-2620 v2, 12 cores, 24 threads; 32G RAM; CentOS Version 7 x86_64; 10 compute nodes with

configuration: Dual Intel (R) Xeon® CPU E5-2690 v2, 20 cores, 40 threads; 64G RAM

; CentOS Version 7 x86_64.

The test results indicate a high level of performance and redundancy of the deployed cluster. Due to the usage of the same configuration of nodes in each of the systems, cluster configuration, internal parameters and tasks can be variously changed and easily automated.

REFERENCES

1. Shakhovska, N. Big data processing technologies in distributed information / N. Shakhovska, N. Boyko, Y. Zasoba, E. Benova // Procedia Computer Science. – 2019. – No. 160. – P. 561–566.