

АНОТАЦІЯ

У роботі розглянуто проблему актуальності підвищення ефективності використання ресурсів Hadoop кластера та їх розподілу між вхідними завданнями для можливості швидше та ефективніше працювати із великими об'ємами інформації.

Розглянуто алгоритми роботи існуючих систем управління ресурсами Hadoop кластера, їх переваги та недоліки. Вдосконалено алгоритми роботи найбільш популярних планувальників менеджера ресурсів. Реалізоване покращення надає можливість розділяти фізичні вузли кластера на логічні групи для використання різними підрозділами або завданнями з різними потребами (CPU, RAM) за рахунок надання гнучкого механізму маркування вузлів кластеру та планування на основі маркувань.

Ключові слова: YARN, маркування ресурсів, менеджер ресурсів, Hadoop, кластер, планувальник ресурсів.

Розмір пояснювальної записки – 87 аркушів, містить 22 ілюстрації, 24 таблиці, 6 додатків.

ABSTRACT

The paper considers the problem of increasing the efficiency of using Hadoop cluster's resources and their distribution between incoming tasks for the possibility of faster and more efficient work with large volumes of information.

The algorithms of existing Hadoop resource management systems, their advantages and disadvantages were considered. The most popular resource manager schedulers' algorithms were improved. The implemented improvement provides the ability to split physical cluster nodes into logical groups for use with different departments or tasks with different needs (CPU, RAM) by providing a flexible mechanism for marking cluster nodes and label based scheduling.

Keywords: YARN, resource labeling, resource manager, Hadoop, cluster, resource planner.

The size of the explanatory note is 87 sheets, it contains 22 illustrations, 24 tables, 4 annexes.