

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Факультет радиофизики и компьютерных технологий
Кафедра интеллектуальных систем

Аннотация к дипломной работе

**Организация распределённых систем обработки
больших данных на основе Hadoop**

Масюк Виктория Романовна

Научный руководитель: старший преподаватель, Д.С. Штукатер

Минск, 2023

РЕФЕРАТ

Дипломная работа: 51 страница, 38 рисунков, 20 источников, 3 приложения

БОЛЬШИЕ ДАННЫЕ, РАСПРЕДЕЛЁННЫЕ СИСТЕМЫ, HADOOP, HADOOP DISTRIBUTED FILE SYSTEM (HDFS), MAPREDUCE

Объект исследования – распределённая система обработки данных.

Предмет исследования – особенности проектирования распределённой системы обработки данных на базе платформы Hadoop.

Цель работы – разработка распределенной системы для обработки больших данных на базе программной платформы Hadoop и тестирование отказоустойчивости файловой системы разработанной распределённой системы.

В работе рассматриваются методы и подходы к обработке больших данных. С учётом проведенного анализа развернут кластер, проведено тестирование его работоспособности и отказоустойчивости файловой системы этого кластера.

Программная реализация и тестирование выполнены на базе платформы Hadoop при использовании языка программирования Java и bash-скриптов. Результаты исследования могут быть полезны специалистам, занимающимся обработкой больших данных.

РЭФЕРАТ

Дыпломная праца: 51 старонка, 38 малюнкаў, 20 крыніц, 3 прыкладаньня

ВЯЛІКІЯ ДАНЫЯ, РАЗМЕРКАВАНЫЯ СІСТЭМЫ, HADOOP, HADOOP DISTRIBUTED FILE SYSTEM (HDFS), MAPREDUCE

Аб'ект даследавання – размеркаваная сістэма апрацоўкі даных.

Прадмет даследавання – асаблівасці праектавання размеркаванай сістэмы апрацоўкі даных на базе платформы Hadoop.

Мэта працы – распрацоўка размеркаванай сістэмы для апрацоўкі вялікіх даных на базе праграмнай платформы Hadoop і тэсціраванне адмоваўстойлівасці файлавай сістэмы распрацаванай размеркаванай сістэмы.

У працы разглядаюцца метады і падыходы да апрацоўкі вялікіх даных. З улікам праведзенага аналізу разгорнуты кластар, праведзена тэсціраванне яго працаздольнасці і адмоваўстойлівасці файлавай сістэмы гэтага кластара.

Праграмная рэалізацыя і тэсціраванне выкананы на базе платформы Hadoop пры выкарыстанні мовы праграмавання Java і bash-скрыптоў. Вынікі даследавання могуць быць карысныя адмыслоўцам, якія займаюцца апрацоўкай вялікіх даных.

ABSTRACT

Thesis: 51 pages, 38 images, 20 sources, 3 application

BIG DATA, DISTRIBUTED SYSTEMS, HADOOP, HADOOP
DISTRIBUTED FILE SYSTEM (HDFS), MAPREDUCE

The object of research – distributed data processing system.

The subject of research – design features of a distributed data processing system based on the Hadoop platform.

The purpose of research – development of a distributed system for processing big data based on the Hadoop software platform and testing the fault tolerance of the file system of the developed distributed system.

In this work, methods and approaches to processing big data are considered. Based on the analysis carried out, a cluster was deployed, its performance and fault tolerance of the file system of this cluster were tested.

Software implementation and testing were performed on the basis of the Hadoop platform using the Java programming language and bash scripts. The results of the study may be useful to specialists involved in the processing of big data.