

Калашніков В.І., Яременко Є.В., Україна, Харків

ЗАСТОСУВАННЯ МОДЕЛІ ПРОГРАМУВАННЯ MAPREDUCE ДЛЯ РОЗПАРАЛЕЛЕННЯ СТАТИСТИЧНОГО АНАЛІЗУ ВЕЛИКИХ ОБСЯГІВ ТЕКСТОВИХ ДАНИХ

Пропонується програма для статистичного аналізу великих обсягів текстових даних із застосуванням методів паралельної обробки. Реалізація побудована на платформі Apache Hadoop (зі зберіганням вхідних і вихідних текстових даних у розподіленій файловій системі HDFS (Hadoop Distributed File System), мова програмування – Java. Програма демонструє можливості паралельної обробки у платформі Apache Hadoop, а також знайомить із програмною моделлю MapReduce.

Калашніков В.І., Яременко Є.В., Україна, Харків

ИСПОЛЬЗОВАНИЕ МОДЕЛИ ПРОГРАММИРОВАНИЯ MAPREDUCE ДЛЯ РАСПАРАЛЛЕЛИВАНИЯ СТАТИСТИЧЕСКОГО АНАЛИЗА БОЛЬШИХ ОБЪЁМОВ ТЕКСТОВЫХ ДАННЫХ

Предлагается программа для статистического анализа больших объёмов текстовых данных с применением методов параллельной обработки. Реализация построена на платформе Apache Hadoop (с хранением входных и выходных текстовых данных в распределённой файловой системе HDFS (Hadoop Distributed File System), язык программирования – Java. Программа демонстрирует возможности параллельной обработки в платформе Apache Hadoop, а также знакомит с программной моделью MapReduce.

Kalashnikov V.I., Yaremenko. E.V., Ukraine, Kharkov

USING MAPREDUCE PROGRAMMING MODEL IN PARALLELIZING STATISTICAL ANALYSIS OF LARGE AMOUNTS OF TEXT DATA

Program for statistical analysis of large amounts using parallel computing methods is proposed. Implemented by means of Apache Hadoop platform (getting and storing data from/in HDFS (Hadoop Distributed File System) in Java programming language. Program demonstrate Apache Hadoop platform parallel computing facilities, familiarize with MapReduce programming model.