

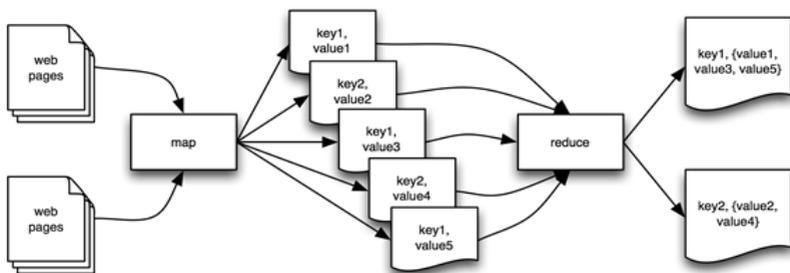
Информационный ресурс на базе MapReduce для управления данными

Шорин Р.И.

Белорусский национальный технический университет

В настоящее время практически не существует информационных ресурсов, которые могли бы полностью удовлетворить потребности в организации и управлении данными в образовательном пространстве, обеспечивая быстрый доступ к огромному количеству научных работ, статей и докладов. Одним из методов распределенной обработки больших данных для получения воспринимаемых человеком результатов, эффективных в условиях непрерывного роста данных, является метод MapReduce.

Разработанный информационный ресурс реализует модель MapReduce для выборки необходимой (удовлетворяющей некоторым условиям) и актуальной информации среди множества кластеров образовательного пространства. Структура модели MapReduce представлена на рисунке.



Структура модели распределенной обработки данных MapReduce

Информационный ресурс был реализован с помощью технологий Ruby on Rails, MySQL, Apache Hadoop и языка программирования Ruby.

Тестирование разработанного информационного ресурса показало, что метод MapReduce эффективно справляется с выборкой конкретного материала среди множества разнообразных данных. Была достигнута высокая производительность ресурса за счет имеющейся в MapReduce возможности использовать произвольное количество кластеров для распределенных вычислений.

Научный руководитель доцент Ковалева И.В.