

МОДЕЛИ И ПРОГРАММНО-АЛГОРИТМИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ОБРАБОТКИ БОЛЬШИХ МАССИВОВ ДАННЫХ В ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНОЙ СРЕДЕ

Д. А. КРОТОВ

НАУЧНЫЕ РУКОВОДИТЕЛИ – Ю. А. СКУДНЯКОВ, КАНДИДАТ ТЕХНИЧЕСКИХ НАУК, ДОЦЕНТ;

Н. Н. ГУРСКИЙ, КАНДИДАТ ТЕХНИЧЕСКИХ НАУК, ДОЦЕНТ

Разработаны математические модели и программно-алгоритмическое обеспечение обработки больших массивов данных в информационно-вычислительной среде.

Ключевые слова: математическая модель, программно-алгоритмическое обеспечение, большие массивы данных, информационно-вычислительная среда.

Цель работы состоит в исследовании, разработке и применении математического и программно-алгоритмического обеспечения для обработки больших массивов данных в информационно-вычислительной среде.

При выполнении работы проведен анализ существующих проблем и подходов при обработке больших массивов данных в информационно-вычислительной среде. Большие данные предполагают нечто большее, чем просто анализ больших объемов информации. Проблема заключается не в том, что создается большое количество данных, а в том, что в основном эти данные представлены в неструктурированном формате. Это могут быть как видеофайлы, так и обычные текстовые документы, таблицы или же данные о местоположении. Все эти данные хранятся в различных местах. Кроме того, все данные, которые необходимо проанализировать, постоянно обновляются, что не позволяет применить к ним традиционные методы анализа информации. В результате организации имеют доступ к большому количеству данных, но не имеют нужных инструментов, которые могли бы их правильно обработать.

В работе рассмотрен процесс, а также способы загрузки данных в базу данных. Проведен обзор различных форматов хранения табличной информации и их различия. Проанализированы различные подходы к загрузке, среди них выявлен способ, который наиболее удовлетворяет задаче работы. Рассмотрена выгрузка данных и частные задачи, возникающие в ходе загрузки данных. С учетом поставленных требований к средству загрузки и проанализированной информации были выявлены возможности для улучшения метода обработки больших массивов данных. Разработано решение для автоматизации выбранного способа загрузки данных.

Разработанный редактор позволяет перенести часть накладных расходов, связанных с проверкой целостности данных, с процесса загрузки на процесс обработки / модификации данных, а также выполнять загрузку данных за меньшее количество действий со стороны оператора. Таким образом, используя полученный метод, можно добиться прироста в скорости загрузки больших массивов данных, если найти удачное средство для загрузки с возможностью отключения проверок целостности. Кроме того, за счет упрощения загрузки снижается количество этапов, на которых можно допустить ошибку. Приведен пример использования проектного решения.

В результате проведения исследований разработаны математические модели и программно-алгоритмическое обеспечение обработки больших массивов данных в информационно-вычислительной среде. Эффективность данной разработки подтверждается ее высокими показателями качества и практическими результатами использования в процессе обработки больших массивов данных в информационно-вычислительной среде.

Применение полученных результатов позволяет в значительной мере автоматизировать и ускорить процесс обработки больших массивов данных, что в свою очередь способствует повышению качества и расширению сферы использования больших данных.