

Использование технологий параллельного программирования для сортировки данных большого объёма

Михелёв Владимир Владимирович

Белгородский государственный технологический университет им. В.Г. Шухова (Белгород), Россия
e-mail: keeper121@ya.ru

Научный руководитель - канд. техн. наук, профессор. Синюк В. Г.

В настоящее время во всех современных информационных системах данные обычно сортируются с целью облегчения последующей их обработки: поиска, добавления или исключения объектов [3]. Однако прогресс не стоит на месте, организации создают огромные объёмы данных, называемые «большие данные» («big data» - англ.).

В сущности, понятие больших данных подразумевает работу с информацией огромного объема и разнообразного состава, весьма часто обновляемой и находящейся в разных источниках в целях увеличения эффективности работы, создания новых продуктов и повышения конкурентоспособности. В связи с этим, проблема ускорения процесса выполнения сортировки стоит достаточно остро, поэтому нами были разработаны улучшенные алгоритмы сортировок с использованием технологии параллельного программирования [1, 2] и проведено их исследование и разработаны рекомендации по их использованию..

В работе рассмотрены различные алгоритмы сортировок [1, 3] и приведен сравнительный анализ последовательных и параллельных алгоритмов их реализации.

Выявлено, что пузырьковая сортировка является достаточно медленным алгоритмом сортировки, однако возможно его значительно ускорить за счет распараллеливания. Для этого можно применить метод нечетной перестановки.

Алгоритм сортировки выбором – один из классических методов упорядочивания элементов последовательности. Суть алгоритма заключается в последовательном нахождении минимального значения в выборке и перестановки найденного значения в начало выборки, таким образом, получаем отсортированную последовательность. Можно распараллелить поиск минимального значения в выборке, это значительно повысит эффективность данного алгоритма.

Быстрая сортировка является существенно улучшенным вариантом алгоритма сортировки с помощью прямого обмена, известного своей низкой эффективностью. Принципиальное отличие этого алгоритма состоит в том, что в первую очередь производятся перестановки на наибольшем возможном расстоянии и после каждого прохода элементы делятся на две независимые группы. Известно, что с ростом объема данных эффективность сортировок уменьшается. Этот факт и взят за основу параллельного алгоритма быстрой сортировки.

Полученные результаты вычислительного эксперимента показывают, что распараллеливание алгоритмов сортировки действительно приводит к увеличению их эффективности при совместном использовании технологий параллельно программирования MPI и OpenMP. Эффективность использования технологий параллельного программирования особенно высока для сортировки данных большого

объема.

Список литературы

1. Д. Кнут. Искусство программирования. Том 1. Основные алгоритмы. М., Изд-во Вильямс, 2010.
2. В. Гергель: Современные языки и технологии параллельного программирования. М., Изд-во МГУ, 2012.
3. Синюк, В. Г. Структуры и алгоритмы обработки данных: лабораторный практикум: учебное пособие / В.Г. Синюк, Ю.Д. Рязанов. - Белгород: Изд-во БГТУ, 2009. – 196 с.