

ПАРАЛЛЕЛЬНЫЕ АЛГОРИТМЫ ДЛЯ ОБРАБОТКИ BIGDATA

Big Data – это такой массив данных, который настолько велик и сложен, что традиционные методы обработки и анализа не могут справиться с ним. Это может включать данные из различных источников, таких как социальные сети, интернет, мобильные устройства, датчики и многое другое. Big Data требует использования специальных инструментов и методов для обработки и анализа, чтобы получить ценную информацию из этого огромного объема данных [1]. Обработка больших объемов данных становится все более актуальной проблемой – с каждым днем генерируются огромные объемы информации, которые необходимо хранить, обрабатывать и анализировать. Это касается различных сфер деятельности – от науки и медицины до бизнеса и социальных сетей.

Актуальность проблемы обусловлена несколькими факторами:

1. *Рост объема данных:* с развитием технологий и увеличением числа подключенных устройств объем генерируемых данных растет экспоненциально. В 2021 году было создано 64 зеттабайта данных, что в 10 раз больше, чем в 2015 году.

2. *Разнообразие данных.* Данные могут быть текстовыми, аудио, видео, геопространственными и т.д. Все это требует различных подходов к их обработке и анализу.

3. *Скорость генерации данных.* Данные генерируются с большой скоростью, что требует быстрых и эффективных методов обработки и анализа.

4. *Неструктурированные данные.* Многие данные не имеют четкой структуры, что усложняет их обработку и анализ.

5. *Безопасность данных.* Необходимо обеспечить безопасность и конфиденциальность данных, чтобы предотвратить их утечку или неправомерное использование.

С данными проблемами позволяют справиться методы параллельных вычислений. Они позволяют распределить нагрузку между несколькими процессорами или ядрами, что увеличивает производительность и ускоряет процесс обработки. Кроме того, параллельные алгоритмы позволяют использовать ресурсы системы более эффективно, так как каждый процессор или ядро работает над своей частью задачи, а не над всей задачей в целом.

Параллелизм – это свойство систем, при котором несколько вычислений выполняются одновременно, и при этом, возможно, взаимодействуют друг с другом. Существует два вида: физический и логический.

Физический параллелизм – это метод, при котором несколько задач выполняются одновременно без взаимодействия друг с другом. Этот вид параллелизма используется для ускорения процесса обработки данных, так как каждая задача выполняется на отдельном процессоре или ядре.

Логический параллелизм (конкуренция) – это метод, при котором задачи выполняются одновременно, но взаимодействуют друг с другом через общие ресурсы. Этот вид параллелизма позволяет более эффективно использовать ресурсы системы, так как задачи могут выполняться одновременно на одном процессоре или ядре, если они не используют одни и те же ресурсы [2].

В настоящий момент существует несколько реализация параллельных алгоритмов.

Алгоритм MapReduce является одним из самых популярных параллельных алгоритмов – был разработан для обработки больших объемов данных и разделения работы между несколькими компьютерами. Этот алгоритм состоит из двух основных этапов: Map и Reduce. На этапе Map происходит обработка данных, а на этапе Reduce – объединение результатов.

Параллельное программирование на основе потоков (Thread-based) позволяет выполнять несколько операций одновременно, используя несколько потоков. Это может увеличить производительность программы, но также может привести к снижению ее эффективности, если не использовать потоки правильно.

Параллельное программирование с использованием процессов (Process-based) работает путем разделения задачи на несколько частей и выполнения каждой части на отдельном процессе. Это позволяет использовать несколько процессоров или ядер для выполнения задачи, что может значительно ускорить процесс.

GPU-ускорение позволяет использовать графические процессоры для ускорения вычислений. Последние имеют большое количество ядер, которые могут выполнять параллельные вычисления. Этот метод особенно полезен для задач, связанных с обработкой изображений и видео [3].

Преимущества параллельных алгоритмов в обработке больших данных (Big Data) заключаются в следующем:

1. Увеличение производительности: параллельные алгоритмы позволяют выполнять несколько операций одновременно, что увеличивает общую скорость обработки данных.

2. Улучшение масштабируемости: параллельные системы могут легко масштабироваться для обработки больших объемов данных, добавляя больше вычислительных ресурсов по мере необходимости.

3. Улучшение эффективности использования ресурсов: параллельные алгоритмы могут использовать ресурсы более эффективно, распределяя нагрузку между несколькими процессорами или ядрами.

4. Возможность обработки сложных задач: параллельные алгоритмы могут использоваться для решения сложных задач, которые могут быть слишком большими или сложными для традиционных последовательных алгоритмов.

Кроме того, как и любое решение, параллельная обработка данных имеет следующие недостатки:

Недостатки параллельных алгоритмов при обработке больших данных включают:

1. Сложность координации и управления параллельно работающими процессами: Параллельные алгоритмы обычно требуют распределения данных и работы между несколькими процессорами или узлами. Это может привести к сложностям в координации и управлении этими процессами, особенно при наличии различных архитектур и платформ.

2. Проблемы с синхронизацией: часто возникает необходимость в синхронизации и обеспечении когерентности данных между различными процессами. Ошибки в синхронизации могут привести к ошибкам в алгоритме и снижению производительности.

3. Проблемы с масштабируемостью: в некоторых случаях, увеличение числа процессоров или узлов может не привести к пропорциональному увеличению производительности из-за проблем с коммуникацией, ограничением ресурсов или другими факторами.

4. Сложность разработки и отладки: разработка и отладка параллельных алгоритмов может быть сложной задачей, так как необходимо учитывать множество факторов, таких как распределение данных, синхронизация, масштабируемость и т.д.

Таким образом для дальнейших исследований за основу возьмем алгоритмы на основе модели MapReduce, а также специализированные инструменты и библиотеки, которые упрощают разработку и оптимизацию параллельных процессов. Особое внимание в дальнейшем будет уделено изменению (доработке) существующих алгоритмов с учетом особенностей обрабатываемых данных, например, уве-

личению уровня кластеризации. Заслуживающей внимания безусловно является использование алгоритмов обработки на основе модели перцептрона, что в может позволить уменьшить объём обрабатываемых данных с использованием алгоритмов их прогнозирования части данных, что позволит сократить время и ресурсы на чтения всего потока данных.

ЛИТЕРАТУРА

1. Ын, Анналин. Теоретический минимум по Big Data: все, что нужно знать о больших данных : пер. с англ. / Анналин Ын, Кеннет Су; [пер. с англ. А. В. Тимохин]. – Санкт-Петербург [и др.] : Питер: Прогресс книга, 2022. – 205 с.
2. Гергель, В.П. Теория и практика параллельных вычислений / В.П. Гергель. – Москва : Национальный Открытый Университет "ИНТУИТ", 2016. – 501 с.
3. Роби, Р. Параллельные и высокопроизводительные вычисления / Р. Роби, Д. Замора ; перевод с англ. А.В. Логунова. – Москва : ДМК Пресс, 2022. – 800 с.

УДК 004.05

С.А. Лукашевич, ст. преп.; А.Н. Купо, доц., канд. техн. наук;
Н.В. Лукашевич, оператор ЭВМ (ГГУ им. Ф. Скорины, г. Гомель)

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОНЛАЙН ПРЕПОДАВАНИИ

В Гомельском государственном университете имени Франциска Скорины постоянно проводится анализ применяющихся как в Республике Беларусь, так и за рубежом образовательных технологий (образовательных платформ) с использованием ИКТ.

В настоящее время существует несколько десятков программ (платформ, систем) для организации видеоконференцсвязи, и их количество продолжает увеличиваться. Все они обладают примерно одинаковым функционалом и многие даже практически не отличаются интерфейсом. Вот некоторые из них: Livestorm, WebinarJam, ProfiConf, GoToMeeting, Easymeeting, SuiteBox, Adobe Connect, Skype, Google Hangouts Meet, Zoom, join.me, BlueJeans, BigBlueButton, UberConference, Intermedia Unite, Mikogo, Vast Conference, MegaMeeting, Cisco Webex, Unified Meeting, ezTalks, WebRoom [1].

Кроме того, активно развивались уже ранее себя зарекомендовавшие (на протяжении 7-8 лет) специализированные платформы онлайн-обучения: iSpring, GetCourse, We Study, ZenClass, Etutorium,