

УДК 004.75

ПРИМЕНЕНИЕ APACHE SPARK ДЛЯ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ В МОБИЛЬНЫХ СИСТЕМАХ



С.Н. Нестеренков

Декан факультета компьютерных систем и сетей БГУИР, кандидат технических наук, доцент
s.nesterenkov@bsuir.by



Д.В. Басак

Инженер-программист, ассистент кафедры ЭВМ БГУИР,
d.basak@bsuir.by



В.В. Куц

Студент 4 курса факультета компьютерных систем и сетей специальности ИиТП БГУИР
mazur.xenia94@gmail.com

С.Н. Нестеренков

Кандидат технических наук, доцент, декан факультета компьютерных систем и сетей Белорусского государственного университета информатики и радиоэлектроники, доцент кафедры программного обеспечения информационных технологий. Автор публикаций на тему машинного обучения, алгоритмов принятия решений, искусственных нейронных сетей и автоматизации.

Д.В. Басак

Окончил Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с исследованием проблем радиочастотной идентификации объектов, организацией учебного и научно-исследовательского процессов в техническом университете.

В.В. Куц

Обучается в БГУИР на 4 курсе специальности "Информатика и технологии программирования".

Аннотация. Выполнен краткий анализ роли больших данных в мобильной разработке, возможных способов решения новых сложных задач, поставленных традиционным системам хранения и обработки, и требований к эффективности программного обеспечения, используемого для работы с большими данными. Представлены основы архитектуры популярного фреймворка для разработки с использованием больших данных.

Ключевые слова: большие данные, мобильные приложения, spark.

Введение.

Развитие сферы мобильных приложений, вызванное увеличением количества владельцев смартфонов, стало причиной спроса на услуги компаний-разработчиков мобильных приложений со стороны владельцев бизнеса, нуждающихся в проектировании и разработке персонализированных и эффективных мобильных приложений, способных удовлетворить нужды пользователей.

В современном мире люди полагаются на помощь мобильных приложений при выполнении задач ежедневной рутины. Эти приложения собирают огромное количество данных, эффективная обработка которых требует уникальных средств управления данными. Здесь в дело вступает аналитика больших данных. Собранные данные используются компаниями для улучшения и поддержки приложений, индивидуализации рекламы, создания статистики. Надежное функционирование таких систем возможно при наличии достоверной и своевременной информации о состоянии устройств на объекте, телемеханики и связи.

Актуальность.

Огромные количества изменчивых и разнородных данных, создаваемых людьми и девайсами, нуждаются в развитых технологиях аналитики, хранения и сбора, способных предоставлять обработанную, отфильтрованную и проанализированную информацию, полезную для прогнозирования развития бизнеса и выработки стратегий развития.

Главным фактором в сфере больших данных является скорость, а инструменты этой области предоставляют бизнесу возможность получать данные в реальном времени и эффективно их

анализировать, не полагаясь на неточные или устаревшие наборы данных, что отвечает запросам современных клиентов.

Подробнее о «больших данных».

Исследовательская компания Gartner приводит следующее определение [3]: ««Большие данные» характеризуются объемом, разнообразием и скоростью, с которой структурированные и неструктурированные данные поступают по сетям передачи в процессоры и хранилища, наряду с процессами преобразования этих данных в ценную для бизнеса информацию».

Данное определение позволяет нам выделить четыре основные характеристики больших данных:

1. Объем. Увеличивающееся количество данных, производимых людьми и девайсами, вынуждает ИТ рынок формулировать и обеспечивать новые стандарты в отношении хранения, обработки и предоставления доступа.

2. Разнообразие. Используемые инструменты должны быть способны справиться с неоднородными, неструктурированными, сырыми данными.

3. Скорость. Важна скорость как сбора, так и анализа поступающих данных.

4. Ценность. Качественно обработанные данные могут позволить предсказать трудности или служить решением для уже существующих проблем [2].

Инструменты для обработки больших данных.

Сырые неструктурированные данные требуют качественной аналитики. Инструменты работы с большими данными позволяют разработчикам мобильных приложений собирать, упорядочивать и оценивать различные наборы данных для выявления предпочтений клиентов.

Самые популярные средства для разработки с использованием больших данных Cloudera, Hadoop, Tableau и Spark; они позволяют разработчикам создавать эффективные и современные приложения.

Подробнее об Apache Spark.

Spark предназначен для разработки приложений, нуждающихся в распределённой пакетной и потоковой обработке неструктурированных и слабоструктурированных данных. Основными компонентами архитектуры данного фреймворка выступают driver и executor.

Driver – компонент, отвечающий за:

1. отслеживание состояния обработки, генерацию задач и их перезапуск в случае отказов;
2. оркестровку вычислений – распределение задач среди вычислительных ресурсов с учетом использования ими данных;
3. контроль за вычислительными ресурсами, включая их выделение и возвращение менеджеру ресурсов.

Executor – компонент, ответственный за выполнение вычислений над данными на ресурсах соответствующего вычислительного узла. Его задачи:

1. контроль расхода вычислительных ресурсов;
2. хранение и предоставление доступа к данным, включая загрузку данных с диска или их десериализацию, для последующей обработки;
3. запуск задач и сохранение итогов работы[1], [4].

Заключение.

Компании-разработчики затрачивают большое количество ресурсов для создания интерактивных мобильных приложений, способных подстроиться под пользователя. Качественные приложения находят популярность из-за простоты использования, доступности интерфейса и постоянно растущего числа владельцев смартфонов, что позволяет им собирать огромные количества данных о предпочтениях и локации пользователей. Большие данные поставили перед традиционными системами хранения и обработки новые сложные задачи. Особую ценность имеет исследование возможностей упрощения анализа данных – любые улучшения или оптимизация, помноженные на огромный массив данных для обработки, могут сэкономить много времени. Владельцы программных продуктов, желающие оставаться конкурентноспособными, должны активно вкладываться в эффективные инструменты обработки собираемых данных.

Список литературы

- [1] Абдуллаев А.А., Мамадаев И.М., Червяков Е.В. Большие данные в мобильных приложениях. – Московский государственный технический университет им. Н.Э. Баумана, 2021. – 4 с.
- [2] Клеменков П.А., Кузнецов С.Д. Большие данные: современные подходы к хранению и обработке – 8 с.
- [3] Mark A. Beyer, Douglas Laney. The Importance of «Big Data»: A Definition, 21 June 2012.

[4] Бутаков Н.А., Петров М.В., Насонов Д. Обработка больших данных с Apache Spark. – Университет ИТМО, 2019. – 50 с.

[5] Jaroslav Pokorny. NoSQL databases: a step to database scalability in web environment. Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services, p. 278-283, ACM New York, NY, USA, 2011.

[6] Одиноченко, М. И. Использование облачных сервисов для решения задач, связанных с применением технологии BIG DATA / М. И. Одиноченко, С. Н. Нестеренков // BIG DATA and Advanced Analytics = BI DATA и анализ высокого уровня: сб. научных статей VIII Междунар. науч.-практ. конф. (Республика Беларусь, Минск, 11-12 мая 2022 года): / редкол. : В. А. Богущ [и др.]. – Минск : Бестпринт, 2022. – С. 252–256.

[7] Беляк, А. А. Анализ производительности технологии Hadoop / А. А. Беляк, С. Н. Нестеренков // BIG DATA and Advanced Analytics = BI DATA и анализ высокого уровня: сб. научных статей VII Междунар. науч.-практ. конф. (Республика Беларусь, Минск, 19-20 мая 2021 года): / редкол. : В. А. Богущ [и др.]. – Минск : Бестпринт, 2021. – С. 343–346.

APPLICATION OF APACHE SPARK TO PROCESSING BIG DATA IN MOBILE SYSTEMS

S.N. Nesterenkov

*Dean of the Faculty of Computer
Systems and Networks of BSUIR,
PhD, Associate Professor*

D.V. Basak

*Software engineer,
Assistant of the Department of
Electronic Computers of BSUIR*

V.V. Kuts

*4th year student of the Faculty of
Computer Systems and Networks of
BSUIR*

Faculty of computer systems and networks

Belarusian State University of computer science and Radio Electronics, Republic of Belarus

E-mail: mazur.xenia94@gmail.com

Abstract. A brief analysis of the role of big data in mobile development and the requirements for the efficiency of software used to work with big data is made. The basics of the architecture of the popular framework for development using big data are presented.

Keywords: big data, mobile applications, spark.