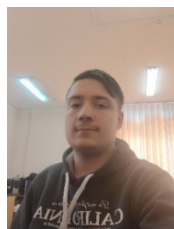


УДК 004.021:004.75

АНАЛИЗ BIG DATA В GAMEMOBILE



Г.А. Трофимук
Студент БГУИР
gleb.trofimuk@gmail.com



П.А. Архиреев
Студент БГУИР
arhir.13.07.2002@gmail.com



А.Н. Марков
Старший преподаватель,
магистр технических наук,
заместитель начальника Центра
информатизации и
инновационных разработок
БГУИР
a.n.markov@bsuir.by

П.А. Архиреев
Студент кафедры информатики БГУИР.

Г.А. Трофимук
Студент кафедры информатики БГУИР.

А.Н. Марков
Магистр технических наук, старший преподаватель кафедры ПИКС, заместитель начальника Центра информатизации и инновационных разработок Белорусского государственного университета информатики и радиоэлектроники.

Аннотация. Данная статья исследует причины возникновения больших данных в мобильных приложениях, их особенности и характеристики. Кроме того, для анализа больших данных мобильных приложений авторы рассматривают распространенный фреймворк Spark, который предлагает эффективный и экономичный способ хранения данных в больших объемах, включая слабоструктурированные и неструктурированные данные.

Ключевые слова: большие данные, Big Data, Hadoop, Spark, мобильные приложения.

Введение.

Современные информационные системы ориентированы на обработку огромных объемов данных, которые генерируются ежедневно [1].

В научных, промышленных и бизнес-отраслях необходимо уметь эффективно использовать инструменты для обработки доступных Big Data, чтобы получить ценные результаты.

Одним из наиболее активно развивающихся направлений в области информационных технологий, связанных с большими данными, являются мобильные технологии, которые объединяют Интернет вещей, ориентированный на устройства, и мобильные приложения, ориентированные на пользователя [2].

Актуальность работы с BIG DATA.

Big Data - это набор подходов, методов и инструментов для обработки значительных объемов данных с различной степенью структурированности.

Основная цель обработки Big Data заключается в оперативном и эффективном использовании всей доступной информации, которая постоянно растет и меняется в огромных объемах.

Обработка больших данных позволяет работать с объемом информации, который невозможно обработать стандартными аппаратными средствами и инструментами, такими как системы управления реляционными базами данных (БД) [3].

Основные задачи Big Data заключаются в хранении и обработке больших объемов информации, что требует использования специальных технологий и инструментов, предназначенных для работы со сверхбольшим объемом данных.

Со стороны разработчика, Big Data в мобильных играх - крайне важный инструмент для понимания поведения и предпочтений пользователей.

С помощью такого анализа разработчики могут получить ценную информацию о том, как игроки взаимодействуют с игрой, какие элементы им нравятся больше всего, какие сложности возникают и как улучшить игровой процесс.

Основные характеристики Big Data, собираемых в мобильных приложениях, включают в себя [4]:

1. Объем данных, который превышает возможности типичных БД, что требует новых подходов и инструментов для сбора, хранения, обработки и анализа данных.

2. Скорость обработки данных в режиме реального времени, что позволяет быстро реагировать на текущую информацию, ограниченную временным периодом приложения.

3. Многообразие источников данных, которые содержат различные типы и форматы данных, включая структурированные, полуструктурированные и неструктурированные данные, извлекаемые из умных устройств, сенсоров, мобильных приложений и социальных сетей. Это позволяет интегрировать большое количество источников данных для обработки и анализа.

4. Недостоверные данные могут затруднить их анализ, поэтому достоверность информации является одним из наиболее важных критериев для пользователей.

5. Большие данные имеют ценность благодаря своей способности улучшать процессы составления отчетности, организации бизнес-процессов, оптимизации расходов компаний и других сфер деятельности, где накопление информации играет важную роль.

Из-за большого количества разнородной информации, Big Data собранные в результате работы мобильного приложения, могут содержать как множество полезной информации, так и наоборот – мусора.

Без правильного анализа она не может быть отфильтрована в полной мере, ведь грамотный анализ больших данных в данной отрасли может помочь разработчикам создавать более привлекательные игры, лучше соответствующие потребностям пользователей и способствующие удержанию игроков в игре на больший период времени.

Именно массивное развитие мобильных устройств и приложений привело к возникновению больших объемов данных, которые сейчас пользователи получают в огромных количествах.

Для обработки такого объема данных, разработчикам мобильных приложений необходимо создавать сложные системы хранения, использующие различные модели данных [5].

В этом случае, для разработки приложений, обрабатывающих большие объемы данных, программисты прибегают к применению соответствующих архитектурных шаблонов проектирования.

Поскольку данные являются ценным активом, необходимо обеспечить надежную систему защиты, которая должна учитываться при любом выборе архитектуры. Кроме того, при разработке систем, предназначенных для системного анализа больших объемов данных в мобильных приложениях, необходимо учитывать ряд факторов [6]:

1. Для успешного анализа данных необходимо предварительно обработать их, поскольку исходная информация неупорядочена и требует очистки, объединения, изменения, перемешивания и других манипуляций для извлечения полезной информации.

2. В сфере обработки больших данных ключевым понятием является итерация.

Для проведения анализа и моделирования часто требуются многократные проходы по одной и той же информации в связи со статистическими процедурами и алгоритмами машинного

обучения. Однако, если при каждом обращении к диску фреймворку необходимо читать один и тот же набор сведений, это может привести к задержкам, которые замедляют всю работу.

3. Для полного использования возможностей Big Data необходимо преобразование хорошо функционирующей модели из регрессионных коэффициентов, которые хранятся в текстовом файле на устройстве разработчика, в информационные приложения, где модели становятся частью сервиса, работающего в эксплуатационном режиме и требующего периодической перестройки.

Конечно, при этом так же следует учитывать вопросы конфиденциальности данных и обеспечивать защищенность личной информации самих игроков, так как нет в мире системы, что невозможно взломать.

Однако, если выполнять все необходимые требования по безопасности, то анализ Big Data может стать мощным инструментом для улучшения мобильных игр и повышения качества пользовательского опыта.

Необходимо отметить, что на данное время уже есть отличные примеры грамотной реализации этой технологии:

- Компания King Digital Entertainment, известная своей игрой "Candy Crush Saga", использует анализ данных для мониторинга поведения своих пользователей.

Они собирают данные о том, как часто игроки проходят уровни и покупают игровые предметы, чтобы определить, насколько привлекательной является игра для пользователей.

Эта информация помогает компании разработать более интересные игры и удерживать пользователей в своих приложениях.

- Playtika - крупнейший разработчик социальных казино-игр. Компания использует анализ данных для определения того, какие типы игр работают лучше всего для различных групп пользователей. Они также анализируют данные о том, как игроки взаимодействуют с играми, чтобы улучшить пользовательский опыт и повысить удовлетворенность пользователями.

- Machine Zone - разработчик мобильных стратегических игр, использует анализ Big Data для создания целенаправленной рекламы и улучшения процесса монетизации.

Они используют данные о том, как пользователи играют и покупают игровые предметы, чтобы настроить рекламу и предложения, которые лучше соответствуют интересам и потребностям пользователей.

Эти примеры демонстрируют, что при умелом подходе к Big Data в мобильных играх, полученные после анализа данные - это не только хороший инструмент заработка, но и возможность совершенствовать свой продукт, опираясь на данные, которые предоставляют ваши же клиенты.

Принципы функционирования Apache Spark.

При разработке мобильных приложений, которые предназначены для практического использования, разработчики занимаются операционным анализом и превращают свои модели в сервисы, которые могут быть использованы для решения различных задач.

Такие приложения постоянно отслеживаются на производительность, контроль SLA и время доступности сервиса, а также совершенствуются различные характеристики.

А для этого необходим фреймворк, который обеспечивает удобство моделирования и годится для разработки промышленных систем разного типа. Apache Spark является одним из таких фреймворков.

Он обладает открытым исходным кодом и сочетает в себе элегантную модель написания кода поверх кластера машин с механизмом распределения программ. Он так же является частью экосистемы проектов Hadoop [7].

Apache Spark представляет собой распределенное кластерное приложение, которое используется для обработки данных. Оно также является универсальной платформой, которая поддерживает несколько режимов обработки данных, включая пакетную и потоковую обработку, а также обработку графов и SQL-запросы [1].

Кроме того, Spark предоставляет различные инструменты для этих целей. Эта платформа специально разработана для создания распределенных приложений, которые могут обрабатывать данные. В связи с этим она предлагает фиксированную master-slave архитектуру, основными компонентами которой являются:

1. Driver, который отвечает за управление вычислениями, в том числе распределение задач на вычислительные ресурсы с помощью executor'ов, учитывая информацию об их использовании. Он также генерирует задачи и перезапускает их в случае отказов, контролирует состояние обработки и управляет вычислительными ресурсами, выделяя и возвращая их менеджеру ресурсов.

2. Executor - компонент, который отвечает за выполнение вычислений на ресурсах того вычислительного узла, где располагается информация.

Он обеспечивает хранение информации и предоставление доступа к ней для выполнения обработки, включая десериализацию или загрузку данных с диска.

Он также контролирует потребление вычислительных ресурсов и запускает задачи, фиксируя их результаты.

Spark, подобно своему предшественнику MapReduce, обеспечивает отказоустойчивость и линейную масштабируемость, но расширяет их возможности в нескольких важных направлениях [6]:

- Выполнение универсального ориентированного ациклического графа, что позволяет передавать промежуточные результаты непосредственно на следующий этап конвейера.

- Широкий спектр преобразований, которые позволяют пользователям задавать вычисления естественным образом. Spark имеет потоковый API, который может определять конвейеры всего в нескольких строках кода.

- Обработка в оперативной памяти, что позволяет разработчикам взаимодействовать с любым местом в обрабатываемом конвейере в машинной памяти кластера. Это избавляет от необходимости повторного вычисления данных или чтения их с диска на следующих этапах и расширяет возможности использования механизмов для распределенной обработки.

- Spark - оптимальный выбор для высокоитеративных алгоритмов и приложений, которые быстро реагируют на запросы пользователей благодаря просмотру больших наборов данных в памяти. Он также характеризуется отличной интеграцией со многими инструментами из экосистемы Hadoop.

- Исходя из представленных данных, можно подытожить несколько преимуществ, что позволяют нам остановиться на выборе именно этого фреймворка для работы с Big Data, получаемых из мобильных приложений:

- Высокая скорость обработки данных: Apache Spark может обрабатывать данные в режиме реального времени, что позволяет быстро анализировать большие объемы данных в мобильных приложениях.

- Универсальность: Apache Spark поддерживает множество языков программирования, включая Java, Scala, Python и R, что делает его универсальным и легко интегрируемым в различные проекты.

- Распределенная обработка данных: Apache Spark использует распределенную обработку данных, что позволяет обрабатывать большие объемы данных на кластерах из множества узлов.

- Масштабируемость: Apache Spark легко масштабируется, что позволяет увеличивать объемы обрабатываемых данных и количество узлов в кластере.

- Разнообразие возможностей: Apache Spark предоставляет широкий набор библиотек и инструментов для анализа данных, включая машинное обучение, графовые алгоритмы, обработку потоков данных и многое другое.

- Открытый исходный код: Apache Spark является проектом с открытым исходным кодом, что делает его доступным и гибким для использования в различных проектах.

Заключение.

Работа с большими объемами данных в мобильных приложениях представляет ряд сложностей, таких как недостаточная инфраструктурная мощность, необходимость обеспечения должного уровня безопасности и организационные трудности при внедрении инновационных методов сбора данных из-за ограниченных ресурсов.

Применение фреймворка Apache Spark позволяет частично решить данные проблемы, улучшив отказоустойчивость и скорость обработки данных, а также эффективно выполнять итеративные алгоритмы благодаря кэшированию результатов в памяти и расширения технологий экосистемы Hadoop.

Список литературы

- [1] Бутаков Н.А., Петров М.В., Насонов Д. Обработка больших данных с Apache Spark. – СПб.: Университет ИТМО, 2019. – 50 с.
- [2] Комаров М., Кучерявый Е., Источники Больших Данных и современные способы хранения данных // Тезисы докладов конференции «Большие Данные в национальной экономике» / под ред. Дубовой Н.А. – М.: «Открытые системы», 2014. – С. 9-11.
- [3] Серебряков М.А. Исследование архитектур построения и алгоритмов работы систем хранения и анализа больших данных: выпуск. квалиф. работа: 09.04.03 / М.А. Серебряков; Тольятт. гос. ун-т. – Тольятти, 2019. – 97 с.
- [4] Кобзаренко Д.Н., Мустафаев А.Г. Учебное пособие дисциплины «Анализ больших данных» для направления подготовки 38.03.05 «Бизнес-информатика», профиль «Электронный бизнес». – Махачкала: ДГУНХ, 2019. – 107 с.
- [5] Бойкова М.А., Петрова С.Ю. Пример реализации интероперабельности больших данных для мобильных приложений // Вестник Новгородского государственного университета. – 2017. – № 6 (104). – С. 6-10.
- [6] Риза С., Лезерсон У., Оуэн Ш., Уиллис Д. Spark для профессионалов: современные паттерны обработки больших данных. – М.: Из-во Питер, 2017. – 272 с.
- [7] Стаселько И.Д., Сычев А.Ю., Протасов А.П., Позняков Т.Д., Алексеев Ю.И. Форматы файлов больших данных: хранение данных в экосистеме Hadoop // Big Data and Advanced Analytics. – 2020. – № 6-2. – С. 335-339.

BIG DATA ANALYSIS IN GAMEMOBILE

G.A. Trofimuk

Student of the BSUIR

P.A. Arkhireenko

Student of the BSUIR

A.N. Markov

*Senior lecturer of the department,
Deputy head of the Center for
Informatization and Innovative
Developments*

Department of Computer Science

Faculty of Computer Systems and Networks

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

E-mail: arhir.13.07.2002@gmail.com

Abstract. This article explores the reasons for the emergence of big data in mobile applications, their peculiarities, and characteristics. In addition, to analyze big data in mobile applications, the authors consider the widely used Spark framework, which offers an efficient and cost-effective way to store large volumes of data, including weakly structured and unstructured data.

Keywords: Big Data, Hadoop, Spark, Mobile applications.