

УДК 338.242.2

32. ИНТЕГРАЦИЯ BIG DATA И ОБЛАЧНЫХ ВЫЧИСЛЕНИЙ В БИЗНЕС-АНАЛИТИКУ (НА ПРИМЕРЕ ПРОГРАММНОЙ РЕАЛИЗАЦИИ)

Пармон М.В.¹, студент гр.073602, Бобрик М.А.¹, студент гр.073602

*Белорусский государственный университет информатики и радиоэлектроники¹
г. Минск, Республика Беларусь*

Ермакова Е. В. – канд. эконом. наук

Аннотация. Использование Big Data и облачных вычислений в сфере бизнес-аналитики широко распространено в современной экономике. При интеграции облачной среды выбор поставщиков облачных услуг, а также облачных сервисов, будет сильно влиять на успех бизнеса в повышении ценности собираемых данных, а умение правильно обработать эти большие объемы данных, помогает улучшить многие процессы в бизнес-аналитике. Обработка таких данных представляет собой работу с огромным количеством информации, иногда даже не сгруппированной или не отсортированной, что для работников усложняет задачу анализа этих данных и усложняет процесс принятия решения относительно рабочей ситуации.

Ключевые слова. Интеграция, Big Data, облачные вычисления, бизнес-аналитика, обработка больших объемов данных, хранение данных, параллельные вычисления, распределенные системы.

Сегодня у каждого предприятия существует информационное пространство, которое играет важную роль в его функционировании и развитии. Под информационным пространством предприятия понимается совокупность объектов, взаимодействующих друг с другом и с информацией в зависимости от иерархии обязанностей и уровня доступа к данным, а также технологий, обеспечивающих это взаимодействие. Основными компонентами информационного пространства являются информационные ресурсы, средства информационного взаимодействия и информационная инфраструктура. Оно позволяет эффективно управлять бизнес-процессами, принимать обоснованные решения на основе анализа данных, автоматизировать задачи и повышать производительность сотрудников.

На сегодняшний день информационное пространство предприятия содержит огромное количество данных, которые могут быть ценным активом при правильной обработке и анализе. Использование концепции Big Data и облачных вычислений помогает предприятиям разобраться в этих данных и извлечь ценную информацию для принятия стратегических решений. [1]

Интеграция аналитики Big Data и среды облачных вычислений стала областью внимания многих предприятий и организаций. Рост, который наблюдается в сфере Big Data, требует больше ресурсов для проведения аналитики. В бизнес-среде важно иметь возможность генерировать полезную информацию на основе собранных данных. При интеграции облачной среды выбор поставщиков облачных услуг, а также облачных сервисов, будет сильно влиять на успех бизнеса в повышении ценности собираемых данных. Основная цель предприятий и организаций — получить генерируемые данные в их окончательной форме, которая может повысить ценность их возможностей принятия решений. Возможность понять последствия перехода на облачные сервисы, проблемы безопасности и широкий спектр облачных сервисов на выбор будет иметь решающее значение для бизнеса. [2]

Big Data играют существенную роль в современной бизнес-аналитике. Они предоставляют компаниям огромные объемы информации, собранной из различных источников, таких как социальные сети, веб-сайты, мобильные приложения и датчики. Работа с Big Data позволяет бизнес-аналитикам получать глубокое понимание клиентских предпочтений, рыночных тенденций и эффективности бизнес-процессов.

Big Data можно использовать для измерения эффективности бизнеса, проблем, с которыми сталкиваются в конкретной отрасли, а также для создания возможностей роста для этого бизнеса. При нынешней динамике бизнес-среды использование облачных сервисов принесет организациям большую пользу в создании ценности для генерируемых данных. Однако интеграция облачной среды требует правильного выбора выбранных поставщиков и предлагаемых услуг. [3]

Применение Big Data в бизнес-аналитике сокращает издержки, повышает операционную эффективность, улучшает понимание клиентов, способствует разработке новых продуктов и услуг, обеспечивает более обоснованное принятие решений и снижает риски, включая предотвращение мошенничества. Это может привести к экономической эффективности компаний, улучшению доходов, снижению затрат и созданию конкурентных преимуществ.

Одним из основных преимуществ использования Big Data в бизнес-аналитике является возможность получать более полное и точное представление о клиентах. Анализ данных позволяет выявлять предпочтения и поведенческие паттерны клиентов, что помогает в создании

персонализированных продуктов и услуг, улучшении маркетинговых стратегий и увеличении уровня удовлетворенности клиентов.

Кроме того, Big Data позволяют проводить анализ рынка и конкурентов, что помогает бизнесу принимать более точные решения в отношении стратегии развития, ценообразования и позиционирования на рынке. Анализ данных также может помочь в выявлении новых возможностей и трендов, что способствует инновационному развитию бизнеса.

Облачные вычисления, известные как «революционная» технология, способны обрабатывать данные разных размеров и автоматически балансировать нагрузки с помощью определенных встроенных функций. Ежедневно генерируется около 2,5 квинтиллионов байт данных, и эти данные поступают из множества различных источников. Сообщения в социальных сетях, информация о климате, данные здравоохранения, такие как медицинские изображения и информация о пациентах, а также данные о транзакциях, собранные в банковской сфере, — все это некоторые примеры собираемых данных. Развитие больших данных и анализ этих данных оказываются полезным катализатором в обеспечении роста многих предприятий.

Обработка таких видов данных является необходимой для анализа их в последующей перспективе. Для этого можно использовать множество доступных ресурсов, которые явно показывают возможности обработки Big Data для последующего внедрения ее в бизнес-аналитику.

С помощью специализированных инструментов и технологий, таких как Apache Hadoop и машинное обучение, предсказательные модели могут быть разработаны для определения будущих тенденций, прогнозирования спроса и принятия решений в режиме реального времени. Большие данные демонстрируют свою ценность, помогая компаниям преодолевать вызовы и создавать конкурентные преимущества в динамичном и информационно насыщенном бизнес-окружении.

Hadoop предоставляет возможности распределенной обработки данных, которые позволяют эффективно обрабатывать Big Data. Он может обрабатывать различные формы данных. Он также обладает способностью к горизонтальному масштабированию путем добавления новых узлов кластера для обработки больших объемов данных.[4]

Распределенная обработка данных в Hadoop основана на принципе разбиения задач на более мелкие части и их параллельного выполнения на разных узлах кластера. Это позволяет обрабатывать большие объемы данных более эффективно и быстро.

Hadoop поддерживает обработку широкого спектра форматов данных, включая структурированные данные, такие как таблицы и реляционные базы данных, полуструктурированные данные, такие как XML и JSON, и неструктурированные данные, такие как текстовые файлы и мультимедийные данные.

Важнейшей особенностью Hadoop является горизонтальная масштабируемость: вместо того чтобы работать на одном мощном сервере, Hadoop использует кластеры из нескольких узлов, каждый из которых обладает собственными ресурсами (процессор, память и дисковое пространство). По мере необходимости в кластер можно добавлять новые узлы для увеличения мощности и вычислительной способности. Это позволяет Hadoop масштабироваться для удовлетворения потребностей в обработке больших объемов данных.[5]

Для хранения данных в распределенной среде Hadoop использует распределенную файловую систему Hadoop Distributed File System (HDFS). HDFS делит файлы на блоки фиксированного размера (по умолчанию 128 МБ) и реплицирует их на разные узлы кластера. Это обеспечивает отказоустойчивость, так как в случае выхода из строя одного узла данные всегда доступны на других репликах. HDFS обладает высокой пропускной способностью, так как данные читаются и записываются параллельно на разных узлах кластера (рисунок 1).

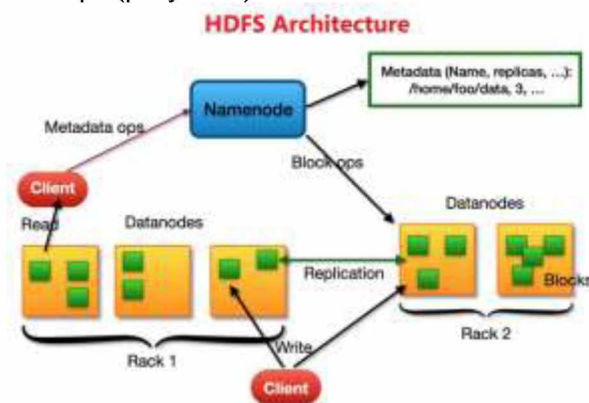


Рисунок 1 – Архитектура файловой системы Hadoop

Также существует Apache Spark. Он является мощным фреймворком для обработки и анализа данных, который может работать вместе с Hadoop. Он предоставляет API на различных языках программирования, таких как Scala, Java, Python и R, для разработки распределенных приложений обработки данных.

Apache HBase – это распределенная база данных, которая работает поверх Hadoop и предоставляет быстрый доступ к большим объемам структурированных данных. HBase поддерживает хранение и обработку структурированных данных с использованием модели ключ-значение. Он обладает высокой пропускной способностью и низкой задержкой при чтении и записи данных.[5]

Для работы с данными в режиме реального времени, используются системы потоковой обработки данных, такие как Apache Kafka и Apache Storm.

Apache Kafka – это распределенная система обмена сообщениями с открытым исходным кодом, предназначенная для публикации и подписки. Она обладает характеристиками, такими как постоянный обмен сообщениями, высокая пропускная способность, распределенность, поддержка нескольких клиентов и работа в реальном времени. Kafka позволяет обрабатывать большие объемы данных, обеспечивая стабильную производительность, и может быть интегрирована с другими системами аналитики. Ее использование позволяет сократить затраты на хранение данных, обеспечивает отказоустойчивость и обеспечивает оперативное принятие решений на основе данных в режиме реального времени.[6]

Kafka предоставляет механизм публикации и подписки на потоки данных в режиме реального времени. Он позволяет производителям данных (публикаторам) отправлять сообщения в темы, а потребителям (подписчикам) получать эти сообщения в режиме реального времени. На рисунке ниже показан типичный сценарий агрегирования и анализа Big Data, поддерживаемый системой обмена сообщениями Apache Kafka (рисунок 2)

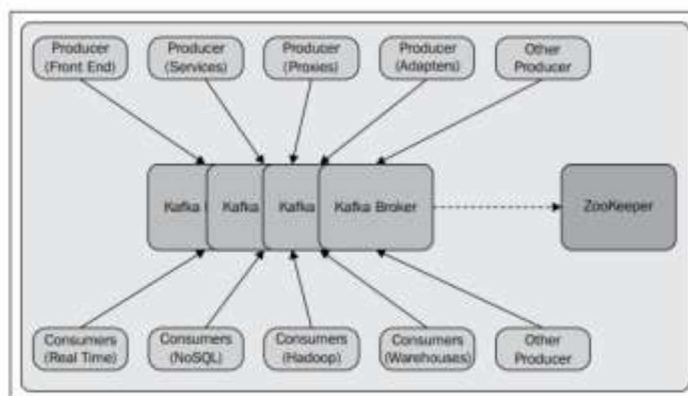


Рисунок 2 – Сценарий агрегирования и анализа Big Data

Apache Storm – это распределенная система обработки потоковых данных с открытым исходным кодом. Она предоставляет высокую производительность и отказоустойчивость при обработке реально-временных потоков данных. Apache Storm использует модель акторов и интегрируется с другими системами обработки данных, такими как Apache Kafka. Она находит применение в различных областях, включая реально-временную аналитику, обнаружение мошенничества и мониторинг сетевых данных (рисунок 3).[7]

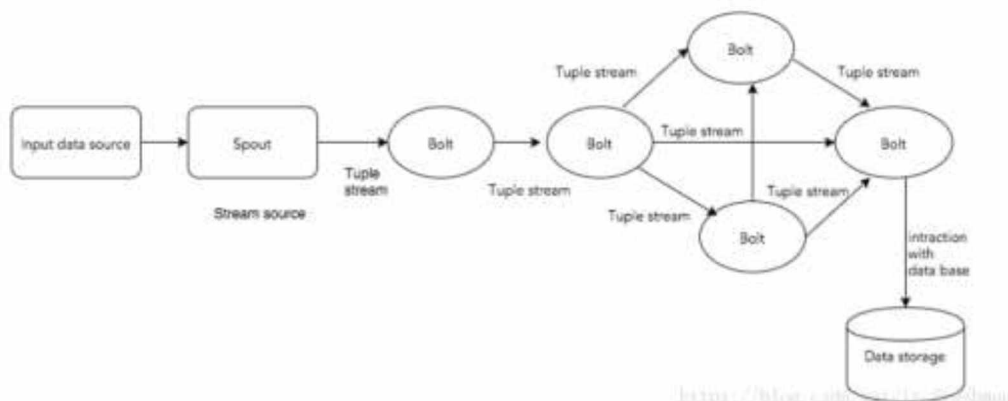


Рисунок 3 – Архитектура работы Apache Storm

Эти системы позволяют обрабатывать и анализировать данные по мере их поступления, что особенно полезно для мониторинга и анализа данных в реальном времени, например, для обнаружения аномалий или мгновенного реагирования на события.

Исходя из вышесказанного можно сделать вывод, что хранение данных является важнейшим компонентом при работе с большими объемами данных. Важно иметь безопасное место, в котором данные могут храниться и их можно легко извлечь при необходимости для анализа. Хранилище больших данных требует учета нескольких ключевых функций, включая масштабируемость, доступность, отказоустойчивость, согласованность и вторичное индексирование. В настоящее время платформа Hadoop широко используется многими исследователями для хранения больших данных, использует распределенную файловую систему Hadoop (HDFS) для хранения данных, в которой большие файлы разбиваются на более мелкие блоки, которые хранятся в разных узлах данных, а также представляет собой огромный выбор инструментов для их обработки.

Список использованных источников:

- Киселевский, О. С. Программное обеспечение операций электронного бизнеса : учебно-методическое пособие / О. С. Киселевский. – Минск : БГУИР, 2023. – 64 с. : ил.
- Rana M. E. et al. Integration of big data analytics and the cloud environment in harnessing valuable business insights //2021 International Conference on Data Analytics for Business and Industry (ICDABI). – IEEE, 2021. – С. 149-156.
- Струнин, Д. А. Бизнес-аналитика и большие данные / Д. А. Струнин. — Текст : непосредственный // Молодой ученый. — 2023. — № 32 (479). — С. 8-10. — URL: <https://moluch.ru/archive/479/105384>
- Russon P. et al. Big data analytics //TDWI best practices report, fourth quarter. – 2011. – Т. 19. – №. 4. – С. 1-34.
- O'Driscoll A., Daugelaite J., Sleator R. D. 'Big data', Hadoop and cloud computing in genomics //Journal of biomedical informatics. – 2013. – Т. 46. – №. 5. – С. 774-781.
- Garg N. Apache kafka. – Birmingham, UK : Packt Publishing, 2013. – С. 30-31.
- Iqbal M. H. et al. Big data analysis: Apache storm perspective //International journal of computer trends and technology. – 2015. – Т. 19. – №. 1. – С. 9-14.

UDC 338.242.2

INTEGRATION OF BIG DATA AND CLOUD COMPUTING INTO BUSINESS ANALYTICS (USING THE EXAMPLE OF A SOFTWARE IMPLEMENTATION)

Parmon M.V.¹, Bobrik M.A.¹

Belarusian State University of Informatics and Radioelectronics¹,
Minsk, Republic of Belarus

Ermakova E. V. - PhD in Economics

Annotation. The use of Big Data and cloud computing in the field of business intelligence is widespread in the modern economy. When integrating a cloud environment, the choice of cloud service providers, as well as cloud services, will greatly affect the success of a business in increasing the value of the data collected, and the ability to process these large amounts of data correctly helps to improve many processes in business analytics. Processing such data involves working with a huge amount of information, sometimes not even grouped or sorted, which will complicate the task of analyzing this data for an employee and complicate the process of making a decision regarding the work situation.

Keywords. integration, Big Data, cloud computing, business analytics, large data processing, data storage, parallel computing, distributed systems.