

ФЕНОМЕН БОЛЬШИХ ДАННЫХ ДЛЯ РАЗВИТИЯ ОБЩЕСТВА И НАУКИ**Елисеева Дина Юрьевна**

старший преподаватель кафедры информационных систем, сетей и безопасности,
ФГБОУ ВО «Российский государственный социальный университет»

**Мнацаканян Ольга Леонидовна**

кандидат педагогических наук, доцент кафедры информационных систем, сетей и безопасности,
ФГБОУ ВО «Российский государственный социальный университет»

**Карягина Татьяна Васильевна**

кандидат технических наук, доцент кафедры информатики и прикладной математики,
ФГБОУ ВО «Российский государственный социальный университет»

**Юшков Андрей Сергеевич**

студент факультета информационных технологий, «Программная инженерия», бакалавр
ФГБОУ ВО «Российский государственный социальный университет»

Аннотация: Настоящая статья посвящена изучению феномена больших данных для развития общества и науки, а также изучению причины возникновения популярности больших данных.

Проведен анализ методов работы с большими данными. В статье рассмотрено понятие «Big Data» и его теоретическая сущность, а также ключевые причины возникновения данной технологии. Определены основные задачи Big Data в современной торговле. Выделены ключевые преимущества, получаемые торговыми предприятиями от использования и применения инструментов Big Data.

Ключевые слова: BigData, инновации, анализ, ИТ, информация, обработка, данные, систематизация, развитие.

Abstract: This article is devoted to the big data phenomenon study for the development of society and science, as well as to the study of the reasons for the popularity of big data.

The analysis of methods for working with big data. The article discusses the concept of «Big Data» and its theoretical essence, as well as the key causes of this technology. The main tasks of Big Data in modern trade are determined. Highlighted the key benefits that retailers get from using and using Big Data tools.

Keywords: BigData, innovation, analysis, IT, information, processing, data, systematization, development.

Введение. Одной из тенденций современного общества XXI века является существенный рост различного рода информации. Объем данных, который генерируется и собирается финансовыми институтами, социальными сетями, научно-исследовательскими центрами уже привычно измеряется петабайтами. Facebook, одна из крупнейших социальных сетей в мире, хранит в своих дата-центрах уже более 18 млрд. изображений, а NYSE – Нью-йоркская фондовая биржа, создает и реплицирует ежедневно около 1 Тб данных, Большой адронный коллайдер получает около 1 Пб данных в секунду.

Актуальность настоящего исследования заключается в том, что большие данные включают в себя больше, чем просто анализ огромного количества информации. Организации создают огромные объемы данных и основная проблема, что большинство из них представлено в формате, который не соответствует традиционному формату структурированной базы данных – видеокассетам, текстовым документам, веб-журналам, компьютерному коду или, например, геопространственные данные. Все это хранится в различных хранилищах, иногда даже вне организации. В результате корпорации могут иметь доступ к огромному количеству своих данных и могут не иметь необходимых инструментов для установления взаимосвязей между этими данными и для того, чтобы сделать из них существенные выводы. К этому добавляется факт, что данные обновляются все чаще и чаще, и вы попадаете в ситуацию, когда традиционные методы анализа информации не успевают за огромными объемами постоянно обновляемых данных, что в конечном итоге открывает путь для технологий больших данных [4].

Инструменты и методы анализа больших данных

С развитием технологии и увеличением количества данных, поступающих из ежедневных организаций, возникла необходимость в более быстрых и эффективных способах анализа таких данных. Имея груды данных на руках больше недостаточно, чтобы сделать эффективные решения в нужное время.

Такие наборы данных больше не могут быть легко проанализированы с помощью традиционного управления данными и методы анализа и инфраструктуры. Поэтому возникает необходимость в новых инструменты и методы, специализированные для анализа больших данных, а также необходимые архитектуры для хранения и управления такими данными. Соответственно, появление big data имеет влияние на все, от самих данных и их сбора, до обработки, до готовых извлеченных решений.

Следовательно, был предложен Big-Data, Analytics, and Decisions фреймворк, который включает инструменты и методы анализа больших данных в процесс принятия решений [1]. Платформа отображает различные инструменты хранения, управления и обработки больших данных, инструменты и методы аналитики, а также визуализацию и инструменты оценки для различных этапов процесса принятия решений [8]. Следовательно, изменения, связанные с

Big data analytics, отражаются в трех основных областях: big data хранение и архитектура, обработка данных и аналитика и, наконец, анализ больших данных, который может быть применен для обнаружения знаний и принятия обоснованных решений [3].

Хранение и управление большими данными

Одной из первых вещей, которыми организации должны управлять при работе с большими данными, является, где и как эти данные будут сохранены после приобретения. Традиционные методы структурированного хранения и извлечение данных включает реляционные базы данных, витрины данных и хранилища данных.

Данные загружаются в хранилище из операционных хранилищ данных с помощью извлечения, преобразования, загрузки (Extract, Transform, Load (ETL)) или извлечение, загрузка, преобразование (Extract, Load, Transform (ELT)), эти инструменты производят извлечение данных из внешних источников, преобразование данных в соответствии с оперативными потребностями и наконец, загрузку данных в базу данных или хранилище данных. Таким образом, данные очищаются, становятся преобразованными и каталогизированными перед тем, как стать доступным для интеллектуального анализа данных и онлайн аналитических функций [2].

Однако среда больших данных требует магнитного, гибкого, глубокого (Magnetic, Agile, Deep MAD) анализа, которые отличаются от аспектов традиционного корпоративного хранилища данных (Enterprise Data Warehouse EDW). Прежде всего, традиционные подходы EDW препятствуют включению новых источников данных, пока они не будут очищены и интегрированы. Из-за вездесущности данных в настоящее время большие среды данных должны быть магнитными, привлекая тем самым все источники данных, независимо от качества данных. Кроме того, учитывая растущее число данных источники, а также сложность анализа данных, хранение больших данных должны позволить аналитики легко производят и быстро адаптировать данные. Для этого требуется гибкая база данных. Логическое и физическое содержимое может адаптироваться синхронно с быстрой эволюцией данных. Поскольку текущий анализ данных использует сложные статистические методы, аналитики должны иметь возможность изучать огромные наборы данных для этого требуется, чтобы репозиторий данных был довольно большого объема [7].

Нереляционные базы данных, такие как не только SQL (NoSQL), были разработаны для хранения и управление неструктурированными или нереляционными данными. Цель баз данных NoSQL – массовое масштабирование, гибкость модели данных, упрощенная разработка приложений и развертывание. В отличие от реляционных баз данных, базы данных NoSQL разделяют управление данными и хранение данных. Такие базы данных скорее ориентированы на высокопроизводительное масштабируемое хранилище данных и позволяют записывать задачи управления данными в протокол прикладного уровня

вместо того, чтобы писать его на конкретных языках баз данных.

С другой стороны, базы данных в памяти управляют данными в памяти сервера, таким образом, исключение дискового ввода-вывода (ввода-вывода) и включение ответов в реальном времени из базы данных. Вместо использования механических дисков можно хранить основную базу данных в Кремниевой основной памяти. Это приводит к улучшению производительности на порядки и позволяет разрабатывать совершенно новые приложения.

Вывод.

В этом исследовании рассмотрели инновационную тему больших данных, которая в последнее время приобрела большой интерес из-за ее беспрецедентных возможностей и преимуществ. Мы живем в информационный век, большие объемы данных производятся ежедневно, и в них заложены внутренние детали и модели скрытых знаний, которые должны быть извлечены и использованы [5]. Следовательно, изучение методов обработки больших данных, может быть применено для изменения бизнеса и повышения эффективности принятия решений, путем применение передовых аналитических методов к большим данным, а также выявление скрытых идей и ценных знаний.

Было обнаружено, что аналитика больших данных может предоставить широкие горизонты возможностей в различных приложениях и областях, таких как интеллект клиентов, мошенничество обнаружение и управление цепочками поставок. Кроме того, изучение эффективной работы с большими данными может служить различным большим отраслям, такие как здравоохранение, розничная торговля, телекоммуникации, различное крупное производство продукции и т.д. [6].

Заключение.

Подводя итоги, это исследование предоставило людям и организациям примеры различных инструментов, методов и технологий изучения работы больших данных, которые можно применять. Это дает пользователям представление о необходимых технологиях, а также разработчикам представление о том, что они могут сделать, чтобы предоставить более совершенные решения для анализа больших данных в поддержку принятия решений.

Наконец, любая новая технология, при правильном применении может принести с собой несколько потенциальных преимуществ и инновации, не говоря уже о больших данных, что является замечательной областью с ярким будущим, при правильном подходе и изучении. Однако с большими данными очень сложно иметь дело. Это требует надлежащего хранения, управления, интеграции, объединения, очистки, обработки, и постоянный анализ поступающих данных.

Анализ больших данных имеет большое значение в эту эпоху переполнения данных и может предоставить непредвиденные идеи и преимущества для лиц, принимающих решения в различных областях.

При правильном использовании и применении, аналитика больших данных может обеспечить основу для достижений на научных, технологических и гуманитарных уровнях.

Список литературы

1. Бабурин В.А., Яненко М.Е. Технологии Big Data в сервисе: новые рынки, возможности и проблемы // Техничко-технологические проблемы сервиса (ТППС). 2014. № 1 (27). С. 100–105.
2. Блинов А.О., Рахманкулов И.Ш. Управление организациями с помощью когнитивных технологий // Вестник Уфимского государственного нефтяного технического университета (УГНТУ). Наука, образование, экономика. Серия: Экономика. 2017. № 4 (14). С. 33–39.
3. Быкова А.А., Морковкина Е.В. Отношенческий капитал как фактор повышения стоимости компании // Корпоративные финансы. 2014. № 4 (28). С. 19–36.
4. Волкова Ю.С. Большие Данные в современном мире // Концепт. 2016. Т. 11. С. 1171–1175.
5. Кузубов С.А. Развитие концептуальных подходов к измерению гудвилл с исторической перспективы // Корпоративные финансы. 2014. № 3 (31). С. 94–104.
6. Макеева Е.Ю., Аршавский И.В. Применение нейронных сетей и семантического анализа для прогнозирования банкротства // Корпоративные финансы. 2014. № 4 (32). С. 130–141.
7. Мальцева С.В., Лазарев В.В. Маркетинговая аналитика в сфере электронного бизнеса на основе больших данных // Информационные технологии в проектировании и производстве. 2015. С. 62–67.
8. Шайкин А.Н., Карягина Т.В. Принятие решений на основе парных сравнений в проектной деятельности студентов // Ученые записки Российского государственного социального университета. 2011. № 9. С.186–189.