

МЕТОДИКИ АНАЛИЗА И ТЕХНОЛОГИИ BIG DATA В СОВРЕМЕННЫХ УСЛОВИЯХ

Семенченко А. А.

Витебский государственный технологический университет, г. Витебск

Научный руководитель: Краенкова К. И., ст. преподаватель

Big data (большие данные) – социально-экономический феномен, который связан с появлением новых технологических возможностей для анализа большого количества данных [1].

Когда говорят о Big Data, упоминают правила 3 «V» – три признака или свойства, которыми большие данные обладают: Volume – (объём) данные измеряются по величине физического объёма документов; Velocity – (скорость) данные регулярно обновляются, что требует их постоянной обработки; Variety – (разнообразие) данные могут иметь неоднородные форматы, быть неструктурированными или структурированными частично [2].

К основным характеристикам больших данных в современных условиях можно отнести:

- возможное полное автоматизированное генерирование без участия человека;
- соотношение только с новыми источниками данных;
- большие потоки информации, которые в ряде случаев не представляют собой ценность. Большая часть данных может быть вообще бесполезной.

Представленные характеристики показывают необходимость детального изучения инструментов и методов анализа в Big data.

Основные инструменты анализа можно разделить на 3 группы (рисунок 1).

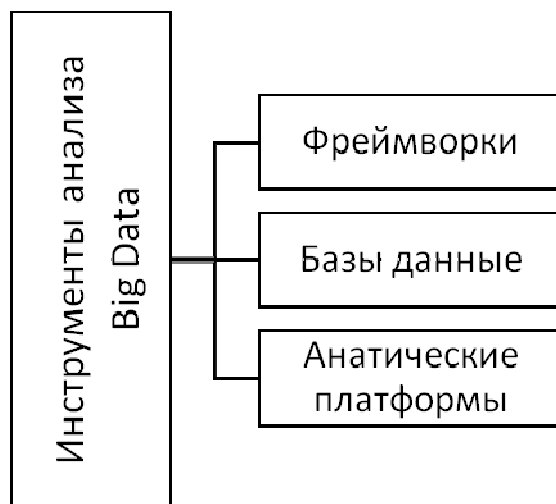


Рисунок 1 – Инструменты анализа Big Data

Составлено автором

Среди наиболее важных и значимых инструментов анализа больших данных выделяют Business Intelligence (BI-платформы).

Подробное изучение представленного вопроса позволило выделить 3 популярные BI-платформы – Tableau, Qlik, Power BI.

Tableau – это система интерактивной аналитики, позволяющая проводить глубокий и разносторонний анализ больших массивов информации и не требующая обучения бизнес-пользователей и дорогостоящего внедрения. Является одним из лидеров рынка платформ для бизнес-аналитики.

Qlik – BI-платформа, которая отличается широким ассоциативным поиском и возможностью обработки вычислений в оперативной памяти. Power BI – набор инструментов бизнес-аналитики от Microsoft для анализа и наглядного представления ценной информации [3].

Кроме оценки инструментов анализа, необходимо изучить основные методы и техники анализа больших данных.

Таблица 1 – Основные методы анализа больших данных

Data mining	Структурирование разнообразных сведений, поиск скрытых и неочевидных связей для приведения к единому знаменателю
Machine learning	Процесс машинного обучения, аналитика и прогнозирование на основе обработанной и структурированной информации
Deep learning	Позволил усовершенствовать нейронные сети до уровня ограниченного искусственного интеллекта. Система сама создаёт многоуровневые вычисления и делает выводы
Краудсорсинг	Позволяет получать данные одновременно из нескольких источников, причем количество последних практически не ограничено
Визуализация аналитических данных	Представление информации в виде рисунков, диаграмм, с использованием интерактивных возможностей и анимации для получения результатов, для использования в качестве исходных данных для дальнейшего анализа. Позволяет представить важные результаты в удобном для восприятия виде
Прогнозная аналитика	Специалисты в данной области стараются заранее предугадать и распланировать то, как будет вести себя подконтрольный объект, чтобы принять наиболее выгодное в этой ситуации решение
Сетевой анализ	Используется для исследования социальных сетей – после получения статистических данных анализируются созданные в сетке узлы, то есть взаимодействия между отдельными пользователями и их сообществами
Имитационное моделирование	Позволяет строить модели, описывающие процессы так, как они проходили бы в действительности

Составлено автором

Представленные методы анализа больших данных привели к росту популярности среди различных платформ, а также развитию новых технологий в работе с Big Data (таблица 2).

Таблица 2 – Технологии работы с Big Data

MapReduce	Модель распределённых параллельных вычислений в компьютерных кластерах, представленная компанией Google. Приложение разделяется на большое количество одинаковых элементарных заданий, выполняемых на узлах кластера и затем естественным образом сводимых в конечный результат
MySQL	Общий термин для различных нереляционных баз данных и хранилищ, не обозначает какую-либо одну конкретную технологию или продукт. Хорошо подходит для достаточно быстрых и однотипных запросов
Hadoop	Свободно распространяемый набор утилит, библиотек и фреймворк для разработки и выполнения распределённых программ, работающих на кластерах из сотен и тысяч узлов
R	Язык программирования для статистической обработки данных и работы с графикой
Аппаратные решения	Представляют собой аппаратно-программные комплексы, которые поставляются готовые к установке телекоммуникационные шкафы, содержащие кластер серверов и управляющее программное обеспечение для массово-параллельной обработки

Составлено автором

Таким образом, наличие большего количества данных приводит к развитию определённого вида анализа с использованием новых инструментов и методов.

Несмотря на большую популярность, у BI-платформ есть ряд недостатков. Актуальной проблемой является конфиденциальность Big Data. Это связано с тем, что большинство сер-

висов по обслуживанию клиентов переходят на онлайн-использование данных, очень легко стать очередной мишенью для киберпреступников. Также можно выделить проблему потери информации. Меры предосторожности требуют не ограничиваться простым однократным резервированием данных, а делать хотя бы 2-3 резервные копии хранилища.

Список литературы:

1. Корнев, М.С. История понятия «большие данные» (Big Data): словари, научная и деловая периодика, 2019 [Электронный ресурс] – URL: www.cyberleninka.ru
2. Технологии Big Data: как использовать большие данные, 2019 [Электронный ресурс] – URL: www.uplab.ru
3. Big Data – что такое системы больших данных? Развитие технологий Big Data, 2017 [Электронный ресурс] URL: www.promdevelop.ru

УДК 657

ОСНОВНЫЕ ПОДХОДЫ К АНАЛИЗУ ЭНЕРГОЭФФЕКТИВНОСТИ ПРОИЗВОДСТВА

Гуляева А. А.

Белорусский национальный технический университет, г. Минск

Научный руководитель: Лапченко Д. А., ст. преподаватель

К вопросам целесообразного потребления топливно-энергетических ресурсов в настоящее время подходят с большим вниманием. В данный момент крайне актуальны исследования, нацеленные на более доскональное изучение и анализ энергетического комплекса Беларуси и ее регионов; разработку и реализацию передовых энергосберегающих технологий и проектов; технико-экономический анализ эффективности их использования; стандартизацию и нормирование показателей потребления топливно-энергетических ресурсов (ТЭР) [1].

Реализация государственной политики по энергосбережению и повышению энергоэффективности в секторах экономики предполагает постоянный мониторинг их результативности. Энергоэффективность выражает степень эффективности (полноты) использования энергетического ресурса, подводимого к установке, его потребляющего, и при изучении понятия энергоэффективности необходимо делать различия между энергоустановками, которые производят энергию, потребляя энергетические ресурсы, и энергоустановками, которые потребляют энергию. Для количественного измерения энергоэффективности применяются различные показатели. Например, для тепловых электростанций используется такой показатель, как удельный расход топлива на отпущенную электроэнергию. Он применяется для сравнения экономичности, эффективности работы различных электростанций. Для электрических сетей энергоэффективность определяется величиной потерь электроэнергии в сетях, которая составляет в настоящее время примерно 11 % от отпущенной в сеть энергосистемы энергии. Для энергосистемы в целом может быть использован показатель удельного расхода топлива по всем электростанциям, относимый на полезно отпущенную потребителям электроэнергию. Для промышленных предприятий в качестве показателя энергоэффективности используется показатель удельного расхода энергии на производимую продукцию (показатель энергоемкости).

Система показателей энергоэффективности может быть представлена в виде четырехуровневой иерархии:

– верхний уровень, по экономике в целом: энергоемкость ВВП, (интегральный индекс энергоэффективности), энерго- и электропотребление на душу населения, сумма инвестиций в мероприятия по повышению энергоэффективности; показатель экономии ТЭР за счет внедрения энергосберегающих мероприятий, удельные и абсолютные выбросы вредных веществ от сжигания и переработки ТЭР;

– второй уровень, по основным секторам потребления энергии: показатели энергоэффективности секторов промышленности, транспорта, жилищно-коммунального сектора и других видов экономической деятельности (общие и частные показатели энергоемкости);