

BIG DATA: ПРОБЛЕМЫ И ТЕХНОЛОГИИ

А.Т. Абдыкаримова, магистр, старший преподаватель
Жетысуский государственный университета им. И. Жансугурова
(Казахстан, г. Талдыкорган)

DOI: 10.24411/2500-1000-2019-10859

Аннотация. В статье рассматривается понятие больших данных, история Big Data, описываются источники больших данных, дается понятие новым технологиям анализа больших данных, такие как: искусственный интеллект и Deep Learning, облачные хранилища, Блокчейн-технологии, Dark Data и программное обеспечение Statistica. В статье затрагиваются задачи и функции больших объемов данных, также раскрываются проблемы, которые возникают при работе с большими данными.

Ключевые слова: большие данные, технологии анализа данных, облачные хранилища, Blockchain, искусственный интеллект и Deep Learning, Statistica.

Понятие «большие данные» резко ворвались в нашу повседневную жизнь. Поток информации существовали всегда, но не так давно стали актуальны методы и технологии обработки больших объемов информации. Поэтому на данный момент

под Big Data понимают не только большой объем данных, но и технологии их обработки.

Новые технологии анализа больших данных приведены на рисунке 1 [1].

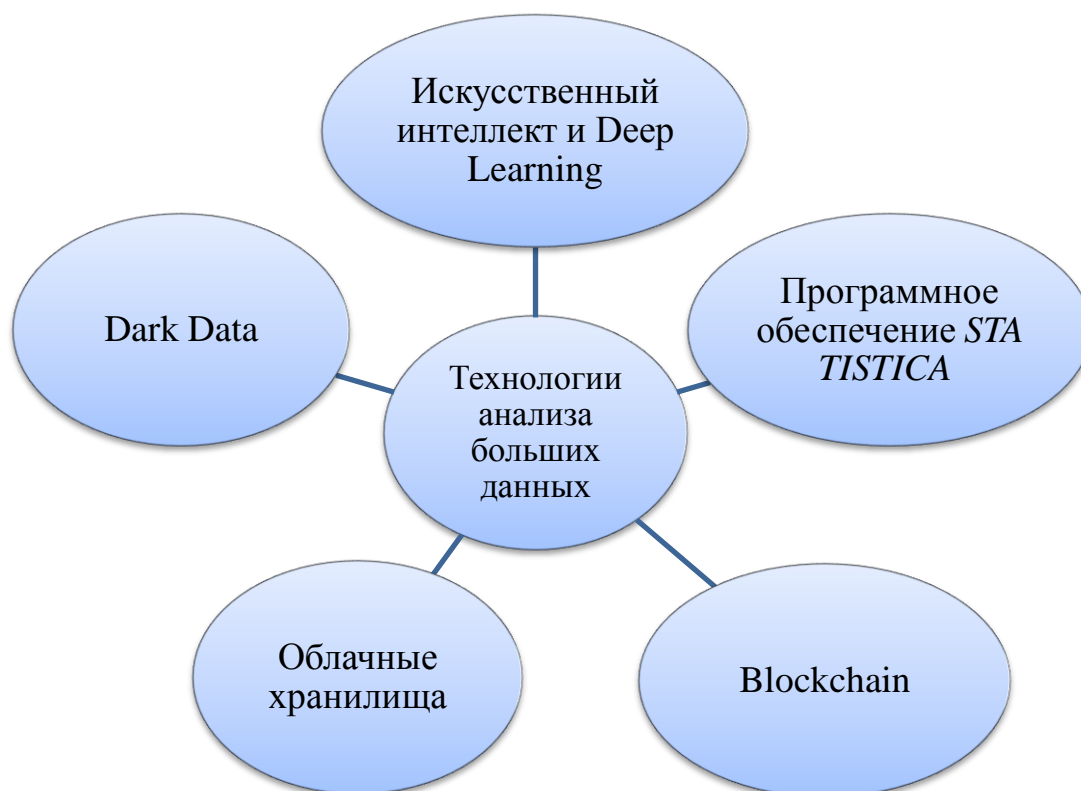


Рисунок. Технологии анализа больших данных

1. Искусственный интеллект и Deep Learning.

Deep Learning называется один из подходов к машинному обучению, который

позволяет предугадывать результаты по входным данным. Другими словами, автономная система самообучения, где используют данные, которые уже существуют

для обучения алгоритмов, чтобы найти образцы и в последующем применить их для прогнозирования новых данных.

2. Облачные хранилища.

Облачное хранилище данных, так называемая модель хранилища, где данные хранятся на многочисленных серверах, которые распределены в сети и предоставляются в пользование клиентам. Клиент же не знает о структуре «облака», и на каком из нескольких серверов хранятся данные, он имеет только доступ к аккаунту и управлением им.

3. Blockchain.

Блокчейн технология – это непрерывная последовательность блоков, которые содержат информацию и выстроены по определенным правилам. Это своего рода база данных, представляющая собой непрерывную цепь из блоков, хранящиеся на нескольких компьютерах. В данной базе-цепочке постоянно создаются новые блоки, которые содержат в себе группу упорядоченных записей (транзакций), которые накопились за последнее время.

4. Dark Data.

Dark Data – это вся нецифрованная информация, не играющая ключевой роли при ее использовании, но которая может послужить причиной для перехода на новый формат хранения сведений.

5. Программное обеспечение Statistica

Statistica – это программный продукт для статистической обработки данных, разработанный компанией StatSoft. Данный программный пакет реализует функции анализа данных, добычи данных, управления данными, также можно проводить визуализацию данных с использованием статистических методов. Программное обеспечение Statistica включает широкий набор различных аналитических процедур и методов: более ста типов графиков, разведочный анализ данных, описательные и внутригрупповые статистики, быстрые основные статистики и блоковые статистики, корреляции, интерактивный вероятностный калькулятор [2].

Как показывает практика, недостаточно знать и использовать технологии анализа больших данных. Есть ряд причин неудач проектов больших данных. В первую оче-

редь это дефицит специалистов, которые обладают знаниями и навыками в сфере анализа больших данных, управления данными и программирования. Ведь проект больших данных – это конкретная предметная область и недостаточный объем знаний его участников в данной прикладной сфере, а также отсутствие постановок задач анализа и критериев эффективности их решения, неосведомленность бизнес-аналитиков о потенциале технологий больших данных, методов и средств их анализа приводят к краху проекта. Также не правильное проектирование хранилища данных может привести к ошибкам в анализе данных. Ведь современный мир перешел не просто к большим данным, а к очень большим данным и из-за роста объема данных нарушается работоспособность хранилищ данных [3].

Главная проблема при работе с большими данными заключается не только в увеличении объема данных, но и в изменении характера данных.

С появлением больших данных многие предприятия осознали важность и необходимость работы с большим объемом структурированных и неструктурированных данных. Но для того чтобы внедрить эти процессы необходимы четкий план действий и правильно выбранные инструменты оптимизации процессов. Получить ожидаемый положительный результат от больших данных многие компании не могут, так как они используют унаследованные системы управления базами данных, а в них не хватает масштабируемости и функциональности. Ведь основной объем данных – это неструктурированная информация. Ее хранение и обработка на основе реляционных баз данных в привычных системах малоэффективна и реляционные системы управления базами данных не являются выходом для ряда ситуаций. И это привело к появлению целого семейства решений, так называемых NoSQL-системы.

Таким образом, предприятиям нужна стратегия, которая будет учитывать источники данных, также их жизненный

цикл, совместимость разных реляционных СУБД и масштабируемость хранения и конечно капиталовложения в IT-инфраструктуру предприятия, дополнительные ресурсы для сопровождения, необходимые для разработки новых приложений и сервисов.

Библиографический список

1. Абдыкаримова А. Т., «Технология больших данных», Наука и жизнь Казахстана, №2(78), 2019 г., с.223-226.

2. [Электронный ресурс]. – Режим доступа: <https://freshprogs.ru/programmy/7335-statsoft-statistica-v-100.html>

3. [Электронный ресурс]. – Режим доступа: <http://www.jetinfo.ru/stati/bolshie-dannye-bolshaya-problema>

BIG DATA: PROBLEMS AND TECHNOLOGIES

A.T. Abdykarimova, *master, senior lecturer*
Zhetysu state university named after I.Zhansugurov
(Taldykurgan, Kazakhstan)

Abstract. *The article covers the main principles and history of Big Data, describes various sources of Big Data and provides an introduction to new technologies of Big Data analysis such as artificial intelligence and Deep Learning, cloud storages, Blockchain technologies, Dark Data and Statistica software. Also the article discusses the tasks and functions of Big Data and the problems arising when working with Big Data.*

Keywords: *Big Data, data analysis technologies, cloud storages, Blockchain, artificial intelligence u Deep Learning, statistica.*