|  |  |
| --- | --- |
| **УДК 330.341.1** |  |
|  | **О.А. Кляус, П.Т. Кляус***ГОУВПО «Донецкий Национальный**Технический Университет»,**Донецк, Донецкая Народная Республика****O.A. Klyaus, P.T.Klyaus****Donetsk National Technical University,**Donetsk, Donetsk People's Republic* |
|  |  |
| ПРИМЕНЕНИЕ ТЕХНОЛОГИЙ BIG DATA В СОВРЕМЕННЫХ УСЛОВИЯХ |
|  |
| **USE OF BIG DATA TECHNOLOGIES IN MODERN CONDITIONS** |
|  |

*Аннотация. В статье проводится анализ основных подходов и методов обработки огромных объемов данных, применения технологий Big Data в различных сферах деятельности, исследуются тенденции их развития.*

*Ключевые слова: Big Data, технологии, цифровая экономика*

*Abstract. The article analyzes the main approaches and methods of processing huge volumes of data, the use of Big Data technologies in various fields of activity, studies the trends of their development.*

*Keywords: Big Data, technology,* *digital economy*

**Постановка проблемы.** Для своевременного принятия решения как на макро-, так и на микроуровнях необходима быстрая и качественная обработка информации. Но учитывая, что в современном мире наблюдается рост объемов информации по экспоненциальному закону, соответственно, объем генерируемой информации настолько увеличился, что хранить и обрабатывать ее традиционными способами стало очень сложно, к тому же данные, обработанные традиционными способами, появляются, как правило, с опозданием, что влияет на качество принимаемых решений [1].

С развитием цифровизации повышается спрос на инструменты обработки данных. Многие компании осознали, что цифровая трансформация — это переход к управлению, основанному на анализе данных. По прогнозам экспертов, особое внимание будет уделено интеграции разрозненных данных, обеспечении их безопасности и доступности.

**Цель статьи** *–*проанализировать возможности применения технологий обработки «больших данных» в организациях различного профиля деятельности, определить основные направления развития, эффективность применения в современных условиях.

**Анализ предыдущих исследований и публикаций.** Возможности применения, тенденции и перспективы развития концепции Big Data в прикладных целях в организациях различного профиля деятельности широко исследуются как российскими, так и иностранными специалистами.

Так, целесообразность и перспективность использования глобальных технологий Big Data в управлении экономическими системами рассматриваются докторами экономических наук С.А. Измалковой и Т.А. Головиной. Авторами проводится анализ возможностей повышения конкурентоспособности хозяйствующих субъектов при использовании технологий Big Data в маркетинговых стратегиях [2].

А.А. Паскова рассматривает практические аспекты использования технологий Big Data в автоматизации технологических и бизнес-процессов и их внедрение в процесс автоматизации на примере эффективного применения этих технологий в отечественной и зарубежной промышленности [3].

Коллектив авторов З.Т. Магеррамов, В.Г. Абдуллаев и А.З. Магеррамова [4], проведя обзор развития, характеристики и применения технологий Big Data, показывают возможности применения технологии Big Data на промышленных предприятиях на примере металлургического производства, а И.А. Халидов и К.Н. Миловидов [5] – в нефтегазовых компаниях.

В банковской сфере возможности работы с новейшими технологиями управления данными и их обработки с помощью системы Big Data с целью повышения конкурентоспособности кредитных организаций при привлечении новых клиентов рассмотрены В.В. Кузнецовой [6].

Возможности использования технологии Big Data при проведении маркетинговой аналитики в различных сферах бизнеса рассматривались в работах В.А. Бабурина и М.Е. Яненко, С.В. Мальцевой и В.В. Лазарева, В.В. Иванова [7].

Вместе с тем следует отметить, что вопросы использования технологии Big Data в управленческой деятельности предприятий различных секторов экономики пока недостаточно разработаны. Исследования современного состояния технологий Big Data и факторов, влияющих на их развитие и применение, а также выявление препятствий и проблем, тормозящих использование новых технологий, призваны обеспечить решение многих экономических проблем.

**Основные результаты исследования.** Использование Big Data в современном мире связано с тем, что появились новые технологические возможности для анализа и обработки огромного количества данных.

Существуют разные варианты авторства и даты появления термина Big Data.

Наиболее распространенное мнение в русскоязычных источниках: дата рождения термина «Big Data» – 3 сентября 2008 года, когда старейший британский научный журнал Nature выпустил специальный номер, полностью посвященный вопросу о влиянии технологий работы с большими данными на будущее науки, и подводивший итоги предыдущим дискуссиям о роли данных в науке вообще и, в частности, в электронной науке (e-science) [8].

По мнению же Сэмюэла Грингарда*,* короткое и внятное определение больших данных дал Дуг Лейни, ныне аналитик в компании Gartner, еще в 2001 г.: что большие данные включают три основных компонента – объем, скорость и разнообразие [9].

Не существует и строгого определения термина «Big Data». Согласно отчету McKinsey Global Institute «Большие данные: новый рубеж для инноваций, конкуренции и производительности», термин «Big Data» относится к наборам данных, размер которых превосходит возможности типичных баз данных (БД) по занесению, хранению, управлению и анализу информации [10].

Еще одну, достаточно краткую формулировку этого термина дала консалтинговая компания Forrester: «Большие данные объединяют техники и технологии, которые извлекают смысл из данных на экстремальном пределе практичности».

Одно из очень распространенных определений в русскоязычных источниках: технологии Big Data – это серия подходов, инструментов и методов обработки структурированных и неструктурированных огромных объемов и значительного многообразия данных, применяемые для получения воспринимаемых человеком результатов, эффективных в условиях непрерывного прироста, распределения информации по многочисленным узлам вычислительной сети [11].

Под термином «Big Data» («Большие данные») скрывается огромный объем разнообразной и сложной информации, обработка которой требует новых архитектуры, методов, алгоритмов и средств анализа.

При этом возникает несколько проблем в обработке таких данных:

* объем информации;
* формат представления: значительная часть представлена в неструктурированном формате (видеозаписи, геопространственные данные, веб-журналы, текстовые документы);
* установление взаимосвязи между данными при использовании множества разнообразных хранилищ.

Поэтому для получения на их основе значимых выводов необходимы специальные инструменты [12].

В настоящее время основными источниками Big Data являются:

* показания считывающих устройств: технических источников (Интернет вещей, в том числе промышленный Интернет вещей), метеорологических приборов, датчиков связи, устройств аудио и видео регистрации, измерительных комплексов, приборов учета потребления и других;
* социальные источники: социальные сети, виртуальные миры, специализированные форумы, профессиональные соцсети, блоги, фотохостинги, сайты отзывов, сайты знакомств и прочее;
* розничная торговля: информация о совершенных транзакциях, сведения из товарных чеков, из дисконтных карт и карт лояльности покупателей, из RFID-меток и прочая;
* здравоохранение: сведения о поставленных диагнозах и предложенных методиках лечения, восприимчивости пациентов к медицинским препаратам, оценке эффективности этих препаратов и проч., что находит отражение в медицинских картах, результатах лабораторных исследований и так далее;
* корпоративная информация – архивы, транзакции, базы данных.

В отечественной научной литературе часто Big Data упоминаются в контексте противопоставления данным, полученным традиционными методами. Рассмотрим разницу в подходах при обработке данных (рис. 1).

Редактирование и сортировка данных перед обработкой

Постепенный анализ небольших пакетов данных

Старт с гипотезы и ее тестирования относительно данных

Данные собираются, обрабатываются, хранятся и лишь затем анализируются

Данные обрабатываются в их исходном виде

Обработка сразу всего массива доступных данных

Поиск корреляций по всем данным до получения искомой информации

Анализ и обработка больших данных в реальном времени, по мере поступления

Рис. 1. Сравнение подходов при обработке данных

Существуют определенные требования, необходимые для корректного функционирования системы больших данных:

* горизонтальная масштабируемость – любая система, обрабатывающая большие данные, должна быть расширяемой;
* отказоустойчивость – необходимое условие при большом количестве машин, которые неизбежно будут выходить из строя;
* локальность данных – так как данные распределены по большому количеству вычислительных узлов, то, если они физически находятся на одном сервере, а обрабатываются на другом, расходы на передачу данных могут стать неоправданно большими. Поэтому обработку данных желательно проводить на той же машине, на которой они хранятся.

Для Big Data в качестве главных характеристик, подчеркивающих сложность задачи обработки и анализа, выделяют (рис. 2):

Рис. 2. Основные характеристики Big Data

Технологии Big Data могут быть полезны при решении множества задач:

* принятие обоснованных управленческих решений;
* оптимизация портфеля инвестиций;
* прогнозирование рыночных ситуаций;
* продвижение товаров;
* оптимизация продаж;
* повышение эффективности логистики и т.д.

Что очень важно, наблюдается тенденция постоянного роста как объемов глобальной сферы данных, так и скорости их обработки.

Согласно прогнозу, содержащемуся в аналитическом отчете о тенденциях и динамике изменений в сфере данных, подготовленном по результатам совместного исследования International Data Corporation (IDC) и Seagate Technology (2019 г.), в ближайшие семь лет глобальная сфера данных вырастет более чем в пять раз. Общий объем новых данных при этом увеличится до 175 зеттабайт (зеттабайт – 1021 (секстиллион) байт) по сравнению с 33 зеттабайтами в 2018 году [13, 14].

С одной стороны, в последние годы неуклонно растет объем информации на промышленных предприятиях. Происходит это за счет данных, полученных с датчиков, измерительных и «умных» устройств (автономное устройство, подключенное к Интернету, которое может отслеживаться и/или управляться удаленно). Самыми перспективными устройствами считаются датчики, которые могут передавать данные в режиме реального времени. С помощью таких датчиков все устройства на предприятии могут быть объединены в сеть, а технологии Big Data позволят обрабатывать поступающую с них информацию и проводить необходимые мероприятия в автоматическом режиме. Таким образом, автоматизация технологических и бизнес-процессов с использованием технологии Big Data повысит скорость реакции бизнеса на внешние и внутренние вызовы.

Аналитики предсказывают, что через шесть лет более половины всех накопленных данных, 90 Збайт, будут генерировать многочисленные и повсеместно распространённые устройства Интернета вещей.

Эксперты отмечают, что важность данных как экономического актива стала очевидной по итогам 2018 г. уже почти для половины (43%) исследованных компаний из региона ЕМЕА. Предсказан постепенный рост доли общего количества данных, приходящихся на коммерческие структуры, с 42% в 2015 г. до 58% в 2025-м. Такая динамика потребует от предприятий и государственных органов проведения модернизации инфраструктуры с учетом растущей базы пользователей.

С другой стороны, на рост сферы Big Data в ближайшие годы повлияет увеличение количества новых пользователей интернета, которые будут потреблять, создавать и публиковать информацию. Доля мирового населения, взаимодействующего с данными, вырастет к 2025 году до 75 процентов, а в Европе, Африке и на Ближнем Востоке эта цифра, несомненно, будет выше (рис. 3) [13, 14].



Рис. 3. Динамика числа взаимодействий людей с данными, чел./д.

Структура накапливаемых человечеством данных претерпевает серьёзные изменения. По оценке IDC, в 2015 г. 30% всей цифровой информации приходилось на визуальные данные неразвлекательного характера (включая научные, производственные, потоки с камер видеонаблюдения и т. п.), 54% — на развлекательный видеоряд, 9% — на невизуальную деловую информацию самого разного рода, 2% генерировалось элементами Интернета вещей, 5% занимал оцифрованный голос, прежде всего, передаваемый по сотовым сетям (рис. 4).

В 2025 г. ситуация обещает стать принципиально иной: падение объема визуальных данных неразвлекательного (до 22%) и развлекательного характера (до 36%), голосовых сообщений (до 2%) и рост объема деловой (до 25%) и связанными с ИВ (до 15%) информации (рис. 5).

Источник данных – IDC и Seagate Technology [13, 14]

Рис. 4. Структура накапливаемых человечеством данных в 2015 г.

Источник данных – IDC и Seagate Technology

Рис. 5. Накапливаемые человечеством данные в 2015 и 2025 гг.

Во всем мире использование аналитики больших данных значительно выросло за последние 5 лет – 50% компаний в 2019 г. по сравнению с 17% в 2015 г. Наиболее активно аналитика Big Data используется в США: более 55% компаний из самых разных сфер деятельности работают с этой технологией, немного ниже показатели в Европе и Азии — около 53%.

Активнее всего Big Data используется компаниями, работающими в сфере телекоммуникаций, финансовых услуг, анализа потребительского поведения. Затем идут компании, специализирующиеся на технологиях в здравоохранении.

При этом обычно не только на крупных, но и на средних предприятиях вводится должность директора по управлению данными (CDO – Chef Data Officer). К сожалению, для наших предприятий, даже крупных, такая должность все еще остается большой редкостью, хотя в последние годы российские компании значительно активизировались.

По данным экспертов IDC и Hitachi Vantara, которые провели в 2019 г. масштабные исследования «Аналитика больших данных как инструмент бизнес-инноваций» [15], 78% российских компаний отмечают значительный рост объема данных, более 80% российских компаний (из числа опрошенных) имеют стратегию цифровой трансформации.

В Центральной и Восточной Европе рост объемов рынка решений по аналитике больших данных довольно активный, ежегодно этот показатель увеличивается на 11% и прогнозируется, что к 2022 году он достигнет $5,4 млрд.

По мнению аналитиков IDC, во многом такое бурное развитие рынка обуславливается ростом этой сферы в России, являющейся крупнейшим региональным рынком: в 2018 году выручка от продажи решений по аналитике Big Data в РФ составила 40% от совокупного объема инвестиций в эти технологии всего региона.

Основное направление, в которое инвестируют российские компании – это анализ транзакционных данных, то есть данных, относящихся к бизнес-операциям, при этом больше всего на обработку Big Data тратят компании со стороны банковского и государственного секторов, телекоммуникационной индустрии и промышленности.

Эксперты говорят о двух подходах в работе с технологиями больших данных:

первый – когда приобретают не технологию, а уже готовый продукт со встроенной технологией Big Data, чаще всего это облачные решения;

второй – создание решения на базе технологий Big Data внутри компании самостоятельно или с привлечением внешних экспертов. Второй подход активно применяют банковский и страховой секторы, телекоммуникационные, производственные компании, ритейл.

При выборе решения для анализа данных компании, в первую очередь, руководствуются вопросами стоимости, простоты использования и интеграции с существующими бизнес-приложениями (рис. 6).



Источник данных – IDC

Рис.6 Требования к решениям для работы с неструктурированными данными

В РФ действует национальный проект «Цифровая экономика», в рамках которого внедряются технологии Big Data. Планируется до конца 2024 года увеличить внутренние затраты на развитие цифровой экономики (по доле в ВВП) не менее чем в три раза по сравнению с 2017 годом, с упором на использование программных продуктов российской разработки.

На прошедшем в марте московском форуме «Big Data 2019», собравшем более 400 представителей предприятий и организаций из различных отраслей, было отмечено, что проекты в области больших данных набирают обороты. Все больше промышленных предприятий внедряют технологии больших данных.

По прогнозам Ассоциации участников рынка Big Data, к 2024 году объем рынка больших данных в России может составить от 500 до 800 млрд рублей.

По оценкам специалистов, направление Big Data наиболее активно используется в компаниях, накопивших большие объемы разноплановой информации, как структурированной, так и неструктурированной: телекоммуникации, интернет-коммерция, финансовая сфера, ритейл.

Но и предприятия в других сферах деятельности все активнее используют или планируют использовать новую технологию, например, для нефтегазового сектора сбор и обработка больших данных важны для геологоразведки, управления фондом недр. В корпорации «Газпром» эти технологии начали применять для снижения потерь газа на инфраструктурных объектах.

ПАО «Газпром нефть» использует технологии Big Data для выявления причин сбоя автоматического перезапуска насосов после аварийного отключения электропитания. Для этого используются записи с контроллеров систем управления почти 1500 скважин.

«Газпромнефть – смазочные материалы» планирует запустить в работу в 2020 году проект, использующий технологии Big Data, по прогнозированию спроса. При этом планируется сокращение простоев на 10-15% при сроке окупаемости – 6-9 месяцев.

В ПАО «Сургутнефтегаз» для ведения бизнеса в режиме реального времени внедрена платформа SAP HANA, позволяющая автоматизировать учет продукции и осуществлять расчет скользящих цен онлайн, что значительно увеличило эффективность бизнес-процессов.

Для оптимизации технологического процесса производства стали на Магнитогорском металлургическом комбинате был внедрен сервис от Yandex Data Factory – «Снайпер», позволяющий обрабатывать параметры плавки и выдавать соответствующие рекомендации.

На Новолипецком металлургическом комбинате реализуется проект с использованием больших данных и искусственного интеллекта, решающий как задачи оптимизации работы электродуговой печи, так и прогнозирование ремонтов фурм в металлургических печах.

Согласно открытым источникам, в конце марта 2020 г. банк «ВТБ» и «Ростелеком» зарегистрировали совместное предприятие для работы с большими данными. Совместная компания будет заниматься разработкой продуктов на основе обработки больших данных с использованием машинного обучения. Первые проекты планируется запустить до конца года.

По мнению экспертов, в ближайшее время в перечень сфер, где востребованы технологии Big Data, войдут транспортная отрасль, энергетика, пищевая промышленность.

ПАО «Россети» планируют апробировать возможности больших данных в энергетическом комплексе РФ.

В РЖД разработана стратегия цифровой трансформации, охватывающая период до 2025 г., в рамках которой предполагается создание специализированных цифровых платформ для управления перевозочным процессом, пассажирскими и грузовыми перевозками.

Немецкая компания Siemens участвует в цифровизации авиационной промышленности РФ, предусматривающей создание единого цифрового киберфизического пространства с применением технологий промышленного Интернета Вещей.

В связи с тем, что госорганами уже накоплены обширные массивы данных, аналитики прогнозируют в ближайшем времени, для повышения эффективности процессов государственного управления, внедрение в госсекторе значительного количества программных решений с использованием технологии больших данных. Примером может служить ФНС России, являющийся обладателем одного из самых больших активов финансовых данных.

Применение Big Data имеет и ряд проблем, сдерживающих развитие рынка Big Data в России [16]:

* высокая стоимость решений и отсутствие быстрых результатов;
* дефицит специалистов: отсутствуют центры компетенций, которые бы занимались их массовой подготовкой;
* некачественные данные: на основе разрозненных или недостоверных данных эффективно решать аналитические задачи невозможно;
* ограниченный выбор решений, реально способных эффективно работать с большими объемами неструктурированных данных;
* опасения заказчиков в области конфиденциальности данных.

Кроме того, по мнению аналитиков, активный рост внедрения Big Data сдерживается тем, что российские компании реже, чем в Европе, Америке или странах Азии, используют облачные сервисы, руководствуясь соображениями безопасности данных, а в ряде случаев и требованиями законодательства.

**Выводы.** Оптимизация и автоматизация существующих бизнес-процессов с целью повышения эффективности деятельности компании – основная цель цифровизации, при этом в основе изменения бизнес-процессов лежит аналитика данных. Деятельность современных компаний сопровождается появлением значительного количества данных, часто с критически важной и ценной информацией о внешней и внутренней работе. Источниками внутренних данных могут выступать не только работа систем типа ERP, CRM, HR и тому подобных, но и внутренние процессы компании: оцифрованные документы, архивы, почтовые переписки. С помощью технологий Big Data предприятия могут анализировать огромные массивы такой информации и выявлять полезные закономерности, что позволит получить конкурентные преимущества.

Технологии Big Data становятся всё более востребованными на российских предприятиях, однако при этом требуют внесения изменений как в деятельность отдельных структурных единиц, так и компании в целом, и сопровождаются значительными финансовыми вложениями.

|  |
| --- |
|  |
| **Список литературы** |
|  | Филлипс, Т. Управление на основе данных. Как интерпретировать цифры и принимать качественные решения в бизнесе. – М. : Манн, Иванов и Фербер, 2017. |
|  | Измалкова С.А. Использование глобальных технологий «BIG DATA» в управлении экономическими системами / С.А. Измалкова, Т.А. Головина // Известия Тульского государственного университета. Экономические и юридические науки. – 2015.– Т. 4. № 1.– С. 151-158. |
|  | Паскова, А.А. Технологии BIG DATA в автоматизации технологических и бизнес-процессов // Научное обозрение. Технические науки. – 2018. – № 4. – С. 23-27. |
|  | Магеррамов З.Т. Big Data: проблемы, методы анализа, алгоритмы / З.Т. Магеррамов, В.Г. Абдуллаев, А.З. Магеррамова // Радиоэлектроника и информатика. – 2017. – № 3. – С. 42-52 |
|  | Халидов, И.А. Большие данные и цифровые месторождения в российских нефтегазовых компаниях / И.А. Халидов, К.Н. Миловидов. // Микроэкономика.– 2018.–№ 5. – С. 82-88.  |
|  | [Кузнецова,](https://www.elibrary.ru/author_items.asp?authorid=881491) В.В. Использование технологии БИГ ДАТА в банках / В.В. [Кузнецова,](https://www.elibrary.ru/author_items.asp?authorid=881491) М.И. Старостенко // Финансовая экономика. – 2019. – №12. – С.57-59 |
|  | Иванов, В.В. [Управление большими данными в международных контрактных и расчетных операциях](https://www.elibrary.ru/item.asp?id=36648332) / В.В. Иванов, Ю.К. Саркисьянц// Российский внешнеэкономический вестник. – 2018.– [№ 12](https://www.elibrary.ru/contents.asp?id=36648322&selid=36648332). – С. 98-114. |
|  | Черняк Л. «Большие данные» – новая теория и практика // Открытые системы. – 2011. – № 10. |
|  | Грингард, С. Интернет вещей. Будущее уже здесь / Сэмюэл Грингард – М.:«Альпина Паблишер», 2017 –188 с. |
|  | Big data: The next frontier for innovation, competition, and productivity [Электронный ресурс]. - Режим доступа: <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation> . – Загл. с экрана. |
|  | Иванов, П.Д. Технологии Big Data и их применение на современном промышленном предприятии./ П.Д. Иванов, В.Ж. Вампилова // Инженерный журнал: наука и инновации. – 2014. – вып. 8. |
|  | Тиндал, С. Большие данные: все, что вам необходимо знать. / Сьюзен Тиндал // PC Week/RE. – 2012. – № 25 (810). |
|  | Сфера данных в Европе, Африке и на Ближнем Востоке: быстрый рост и перенос на периферию [Электронный ресурс]. – Режим доступа : https://www.seagate.com/files/ www-content/our-story/trends/files/data-age-emea-ru-regional-idc.pdf. – Загл. с экрана. |
|  | The Digitization of the World From Edge to Core [Электронный ресурс]. – Режим доступа: https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf . – Загл. с экрана. |
|  | Аналитика данных как ключ к бизнес-стратегии компании [Электронный ресурс]. – Режим доступа :https://www.computerworld.ru/articles/Analitika-dannyh-kak-klyuch-k-biznes-strategii-kompanii (https://www.hitachivantara.com/ru/go/idc-bda/?ecid=ms\_ru\_bd\_ ru\_geidcbda2) . – Загл. с экрана. |
|  | Большие данные (Big Data) в России [Электронный ресурс]. – Режим доступа : [http://www.tadviser.ru/index.php/%D0%A1%D1%82%D0%B0%D1%82%D1%8C%D1%8F:%D0%91%D0%BE%D0%BB%D1%8C%D1%88%D0%B8%D0%B5\_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D0%B5\_(Big\_Data)\_%D0%B2\_%D0%A0%D0%BE%D1%81%D1%81%D0%B8%D0%B8](http://www.tadviser.ru/index.php/%D0%A1%D1%82%D0%B0%D1%82%D1%8C%D1%8F%3A%D0%91%D0%BE%D0%BB%D1%8C%D1%88%D0%B8%D0%B5_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D0%B5_%28Big_Data%29_%D0%B2_%D0%A0%D0%BE%D1%81%D1%81%D0%B8%D0%B8). – Загл. с экрана. |