



Столыпинский
вестник

Научная статья

Original article

УДК 330.47

**ПРИНЦИПЫ ТЕХНОЛОГИИ BIG DATA В ЭКОНОМИКЕ
СОВРЕМЕННОГО ПРЕДПРИЯТИЯ**
PRINCIPLES OF BIG DATA TECHNOLOGY IN THE ECONOMY OF A
MODERN ENTERPRISE

Андреева Софья Андреевна, студент, «Московский государственный технический университет им. Н. Э. Баумана» (05005, Москва, 2-я Бауманская ул., д.5, стр.1), andreevasa@student.bmstu.ru

Пушкин Александр Николаевич, студент, «Московский государственный технический университет им. Н. Э. Баумана» (05005, Москва, 2-я Бауманская ул., д.5, стр.1), pan20u718@student.bmstu.ru

Ядова Наталья Евгеньевна, к.э.н., доцент кафедры «Экономика и бизнес», «Московский государственный технический университет им. Н. Э. Баумана» (05005, Москва, 2-я Бауманская ул., д.5, стр.1), nyadova@bmstu.ru, ORCID: <https://orcid.org/0009-0000-4635-7150>

Andreeva Sofya Andreevna, student, Bauman Moscow State Technical University (Moscow, Russia), andreevasa@student.bmstu.ru

Pushkin Alexander Nikolaevich, student, Bauman Moscow State Technical University (Moscow, Russia), pan20u718@student.bmstu.ru

Yadova Natalia Evgenievna, PhD, professor of department “Economy and Business” Bauman Moscow State Technical University (Moscow, Russia), nyadova@bmstu.ru, ORCID: <https://orcid.org/0009-0000-4635-7150>

Аннотация. В статье дана характеристика технологии Big Data и принципы ее функционирования. В технологии используются как структурированные, так и неструктурированные данные. Big Data является основой для разработки и применения новых технологий, таких как искусственный интеллект, машинное обучение, технология распределенного обучения и обработки данных, технология интеллектуальной системы поддержки и принятия решений и технология финансовых аналитических систем. Рассмотрены принятые в научной литературе параметры технологии, на которых строится работа с большими данными. А также принципы работы: масштабирование, использование массивно-параллельной обработки данных, шардирование, защита и хранение данных. Раскрыта роль технологии в развитии различных отраслей экономики, в создании новых возможностей и повышении эффективности бизнеса. Показаны области применения и влияния технологии на предприятиях логистической отрасли, в частности – отслеживание поставок. Приведены результаты исследования использования технологии Big Data в одном из лидеров рынка логистических систем, в компании СДЭК. Проведен анализ современного состояния и перспектив развития технологии Big Data в РФ.

Annotation. The article describes the Big Data technology and the principles of its functioning. The technology uses both structured and unstructured data. Big Data is the basis for the development and application of new technologies such as artificial intelligence, machine learning, distributed learning and data processing technology, intelligent decision support and decision making technology and financial analytical systems technology. The parameters of the technology accepted in the scientific literature, on which work with big data is based, are considered. As well as the principles of operation: scaling, the use of massively parallel data processing, sharding, data protection and storage. The role of technology in the development of various sectors of the economy, in creating new opportunities and improving business efficiency is revealed. The areas of application and influence of technology in the enterprises of the logistics industry, in particular, supply tracking, are shown. The results of a study on the use of Big Data technology in one of the market leaders in logistics systems, in the SDEK company, are presented. The analysis of the current

state and prospects for the development of Big Data technology in the Russian Federation is carried out.

Ключевые слова: *Индустрия 4.0 (Четвертая промышленная революция), большие данные, наука о данных, СДЭК, анализ данных, масштабирование, шардирование, защита и хранение данных.*

Keywords: *Industry 4.0 (The Fourth Industrial Revolution), big data, data science, CDEK EXPRESS, data analysis, scaling, sharding, data protection and storage.*

XXI век ознаменовал наступление новой, цифровой эпохи. Каждой эпохе соответствуют особые политические и экономические события. Индустрия 4.0 (Четвертая промышленная революция) новый этап развития промышленности, который предполагает массовое внедрение информационных технологий, искусственного интеллекта, машинного обучения, а также различных аналитических систем во все сферы общества. Персональные компьютеры, глобальная сеть и цифровые системы уже стали неотъемлемой частью производственного процесса на подавляющем большинстве предприятий. Цифровые технологии позволяют увеличить эффективность производства и усовершенствуют взаимодействие с потребителем. Технология больших данных (Big Data) одна из наиболее характерных технологий индустрии 4.0. К концу первой четверти 21го века технология используется во многих естественных и гуманитарных науках и сферах жизнедеятельности современного общества. За последние два десятка лет (2023 г.) большие данные оказали существенное влияние на экономическую активность и уровень жизни человечества. Big Data является основой для разработки и применения новых технологий, таких как искусственный интеллект, машинное обучение, технология распределенного обучения и обработки данных, технология интеллектуальной системы поддержки и принятия решений и технология финансовых аналитических систем. Данные технологии способствуют появлению новых способов увеличения производительности предприятий. Благодаря аналитике большого объема потока данных, предприниматели на основе технологий больших данных могут экстраполировать конкретные экономические показатели, это позволяет

принимать своевременные эффективные решения. Такие возможности технологии положительно влияют на эффективность бизнес-процессов, способствуют улучшению качества продукции и оказываемых услуг. Big Data позволяет предприятиям анализировать поведение потребителей, предсказывать их предпочтения и потребности, проводить более персонализированные маркетинговые кампании. Таким образом технология открывает новые возможности для создания и практической реализации инновационных бизнес-моделей.

Big Data (с англ. большие данные) – это большие (несколько сотен гигабайт данных и более) наборы неструктурированных, полуструктурированных или структурированных данных, полученных из различных источников. Такими источниками выступают базы данных клиентов, медицинские записи, системы бизнес-транзакций, социальные сети, данные мобильных приложений и т.д.

По статистике, объём цифровой информации удваивается каждые восемнадцать месяцев. [12] Необходимость в систематизации и обработке этой информации является важным направлением в современной науке и в практических задачах. Большинство крупных предприятий в различных отраслях экономики и промышленности используют технологию Big Data для оптимизации производства и бизнес процессов, а также для исследования соответствующих рынков товаров и услуг, благодаря возможности получения и своевременной-поточковой обработки информации.

Big Data характеризуется 7 параметрами “7V”:

- объем (*англ. Volume*);
- скорость (*англ. Velocity*);
- ценность (*англ. Value*);
- разнообразность (*англ. Variety*);
- достоверность (*англ. Veracity*);
- переменчивость (*англ. Variability*);
- визуализация (*англ. Visualization*).

В практических задачах генерация данных различных форматов происходит с сравнительно большой скоростью, при больших объемах операций информация увеличивается нелинейно, технология должна обеспечивать надежный результат необходимый предпринимателю. Необходимо выделить, что для различных запросов или заданий данные могут кардинально отличаться, чтобы правильно анализировать информацию необходимо использовать различные методы технологии Big Data. Визуализация наиболее эффективна для восприятия человеком. Таблицы и диаграммы помогают упростить анализ данных и повысить результативность поставленных задач. Данные параметры “7V” необходимы для решения сложных заданий, прогнозов и правильных выводах о ситуации на рынке или производственной цепи.

На сегодняшний день (2023 г.) Big Data участвует практически во всех задачах по цифровизации бизнес-процессов. Big Data находит применение во многих сферах: в торговле, в медицине, в безопасности, в промышленности, в рекламе, в транспортировке, в сфере развлечений и услуг, в науке и так далее. Важно развивать цифровые сервисы, в основе которых лежат постоянно растущие объемы информации.

При обработке больших объемов операций наблюдается увеличение нагрузки на базу данных, что приводит к снижению производительности серверов и увеличению времени выполнения операций. Данное явление особенно критично в случае, когда обработка данных занимает больше времени, чем доступное время для выполнения этих операций. Этот эффект может быть объяснен рядом факторов, таких как увеличение количества запросов к базе данных, неравномерное распределение нагрузки между ресурсами базы данных, конкуренция за ресурсы обработки, ограниченная пропускная способность и задержки при передаче данных, а также проблемы с кешированием и индексированием данных.

В работе с большими объемами операций и данными, появляется необходимость в оптимизации баз данных, внедрении параллельной обработки и распределенных архитектур, а также в использовании эффективных алгоритмов и методов обработки данных, чтобы обрабатывать постоянно

увеличивающуюся информационную нагрузку и обеспечивать высокую производительность аналитической системы. Технология Big Data представляет собой совокупность ранее описанных процессов, закономерностей и терминов.

Принципы технологий Big Data в экономике предприятия заключаются в анализе и обработке больших объемов данных с целью получения ценной информации, которая может быть использована для принятия важных бизнес-решений.

Важны два принципа работы с Big Data: горизонтальная масштабируемость (возможность расширения системы обработки при увеличении объемов данных) и отказоустойчивость (возможность продолжения работы даже в случаях сбоев системы).

Масштабирование можно разделить на типы, перечисленные далее.

- Вертикальное масштабирование предполагает наращивание аппаратных ресурсов существующего сервера для возможности большей рабочей нагрузки. На практике это обычно самый простой способ масштабирования, но мощность оборудования нельзя увеличивать бесконечно.

- Горизонтальное масштабирование – это добавление дополнительных серверов для распределения рабочей нагрузки. Горизонтальное масштабирование сложнее в настройке и управлении, чем вертикальное.

- Использование массивно-параллельной обработки данных (MPP, Massively Parallel Processing), когда несколько серверов выполняют вычисления для параллельной обработки запросов. MPP-базы обычно горизонтально масштабируемы.

- Шардирование, что предполагает разделение большой базы данных на более мелкие управляемые фрагменты - шарды каждый из которых хранится на отдельном сервере.

Отказоустойчивость достигается в результате следующих факторов.

- Репликации одного и того же массива информации на несколько копий, каждая из которых размещается на отдельном кластере. При выходе из строя одного кластера данные можно получить из копии.

- В случае MPP-систем равнозначность узлов обработки и хранения данных. При отказе одного узла он легко заменяется другим.

В экономике предприятия объемы данных могут включать в себя информацию о продуктах, клиентах, операциях, финансовых показателях и многом другом. Сбор и хранение таких объемов данных становится все более дешевым и доступным, что создает возможности для более глубокого и детального анализа.

Однако, для работы с такими большими объемами данных, требуется методы эффективной обработки и анализа информации. В качестве этих методов применяются различные технологии Big Data, такие как параллельная обработка данных, распределенные системы хранения данных, машинное обучение и искусственный интеллект.

Параллельная обработка данных позволяет распределить задачи на несколько узлов обработки, что ускоряет время выполнения операций. Она подразумевает несколько этапов: распределение задач на части, дальнейшая их обработка отдельно и параллельно друг с другом и объединение результатов через агрегацию.

Распределенные системы хранения данных позволяют хранить и обрабатывать данные на нескольких узлах, обеспечивая более высокую пропускную способность и отказоустойчивость. Системы подразумевают собой репликацию одного и того же массива информации на несколько копий, каждая из которых размещается на отдельном кластере.

На рисунке 1 мы видим технологию MapReduce которую можно по праву назвать главной технологией Big Data, т.к. она изначально ориентирована на параллельные вычисления в распределенных кластерах. Эта модель была предложена компанией Google.

Данные разделяются на несколько частей (*англ. map*), и каждая из них обрабатывается независимо от остальных в одно и то же время, параллельно (*англ. shuffle*). После обработки происходит объединение результатов (*англ. reduce*).

Применение MapReduce позволяет существенно сократить время выполнения запросов к базе данных, что выводит возможности финансового анализа на качественно новый уровень.

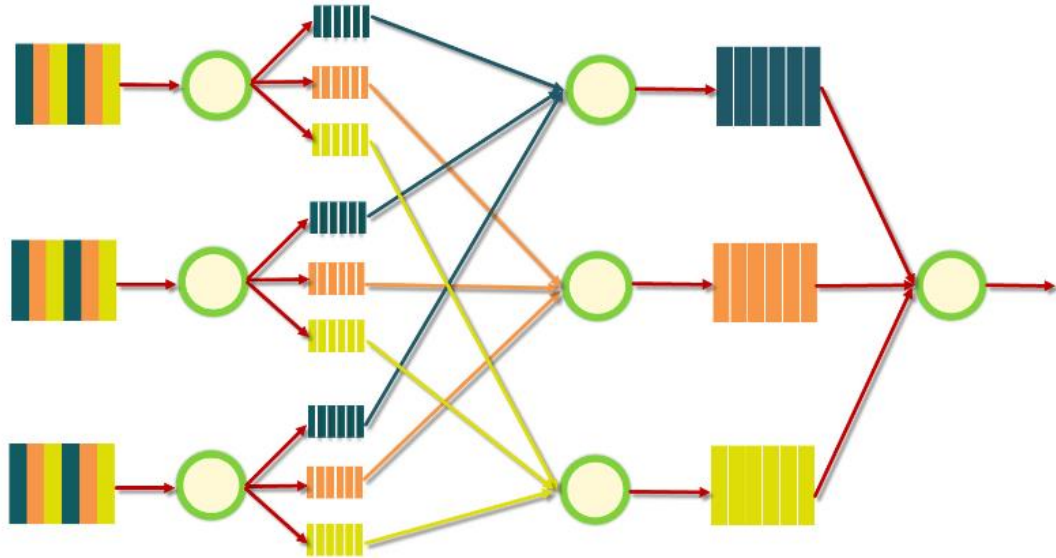


Рис. 1. Модель распределенной обработки данных, предложенная компанией Google [11]

Машинное обучение и искусственный интеллект используются для анализа данных и поиска закономерностей, что позволяет выявить скрытые связи и учесть влияние различных экономических явлений на бизнес-процессы.

Эффективное использование технологий Big Data в экономике предприятия может привести к ряду преимуществ, таких как повышение эффективности, снижение затрат, улучшение качества принимаемых решений, предсказание будущих тенденций, понимание потребностей клиентов и обеспечение конкурентных преимуществ на рынке.

При внедрении технологий Big Data необходимо учитывать различные факторы, такие как ресурсы, инфраструктура, безопасность данных, соблюдение законодательных требований, необходимость специалистов по работе с данными и этические нормы.

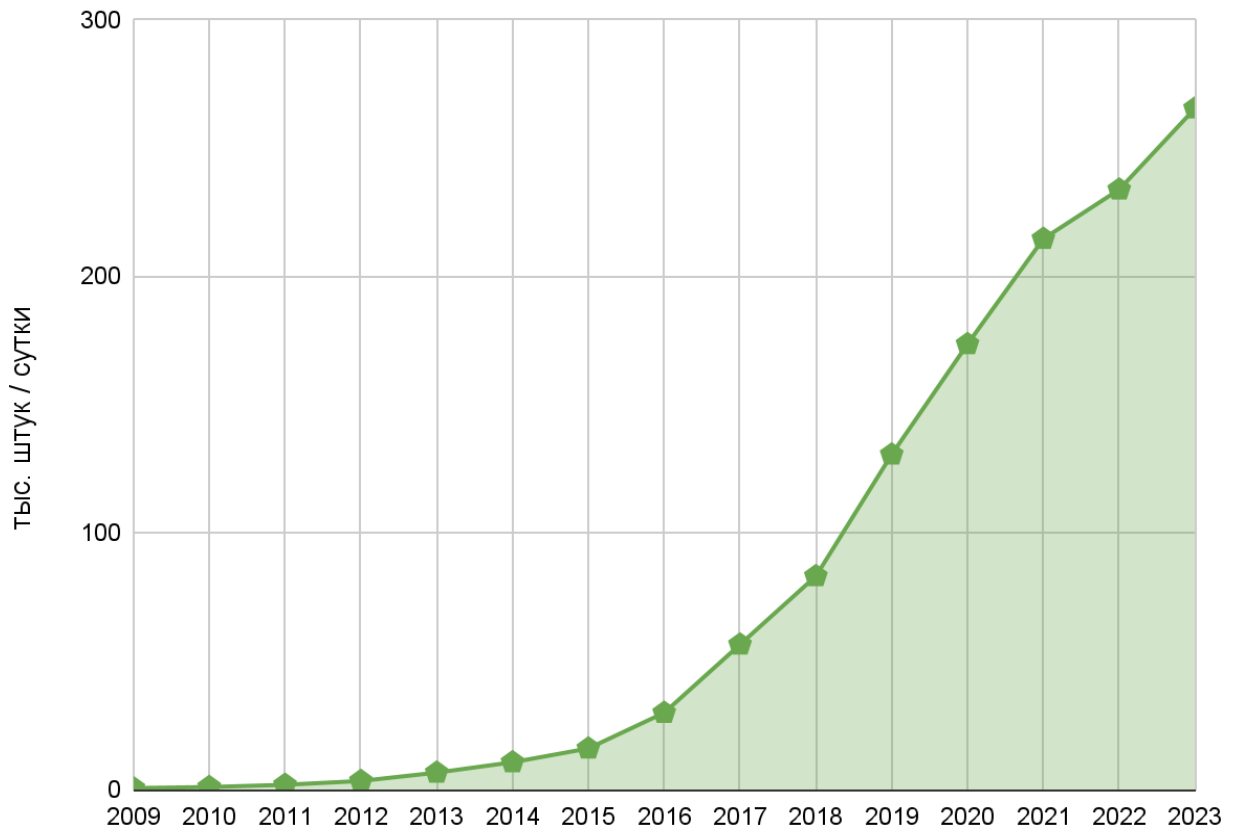
Принципы технологий Big Data в экономике предприятия позволяют использовать большие объемы данных в качестве ценного инструмента для принятия важных бизнес-решений и достижения конкурентных преимуществ.

Современный мир все больше ориентируется на эффективность доставки товаров и услуг. С развитием электронной коммерции и увеличением объемов интернет-торговли становится очевидной необходимость в развитии и совершенствовании логистических систем доставки. Большие данные помогают онлайн-ритейлерам управлять запасами в соответствии с задачами, специфичными для того или иного местоположения. Компании этого сектора используют большие данные для анализа личных и поведенческих данных клиентов и создания подробного профиля клиента.

Одной из ведущих на 2023 год компаний в области доставки отправок является «СДЭК» (или Служба доставки «Экспресс-курьер») — крупный российский логистический оператор. Основана в 2000 году и с тех пор активно развивается, предлагая широкий спектр услуг по доставке грузов и документов. [13]

Компания СДЭК выстроила свою сеть доставки по всей России и за ее пределами, включая страны СНГ и Европы. Главным преимуществом СДЭК является использование инновационных технологий, таких как автоматизированные системы сортировки, отслеживание грузов в режиме реального времени и гибкая система управления логистикой.

Использование Big Data в таких логистических компаниях как СДЭК в России началось примерно в середине 2010-х годов. По официальным данным сервис «PIM.Solutions», использующий принципы больших данных, появился в компании СДЭК в 2014 году. На графике 1 показано среднее значение отправок за сутки.



Граф 1 Динамика количестве отправлений (накладных) СДЭК [5,6]

В 2014 году среднее количество доставляемых грузов было в районе 10 тысяч в день, а в 2017 уже 60 тысяч в сутки, в настоящее время (2023) в России и в мире СДЭК доставляет 270 тысяч посылок каждый день.

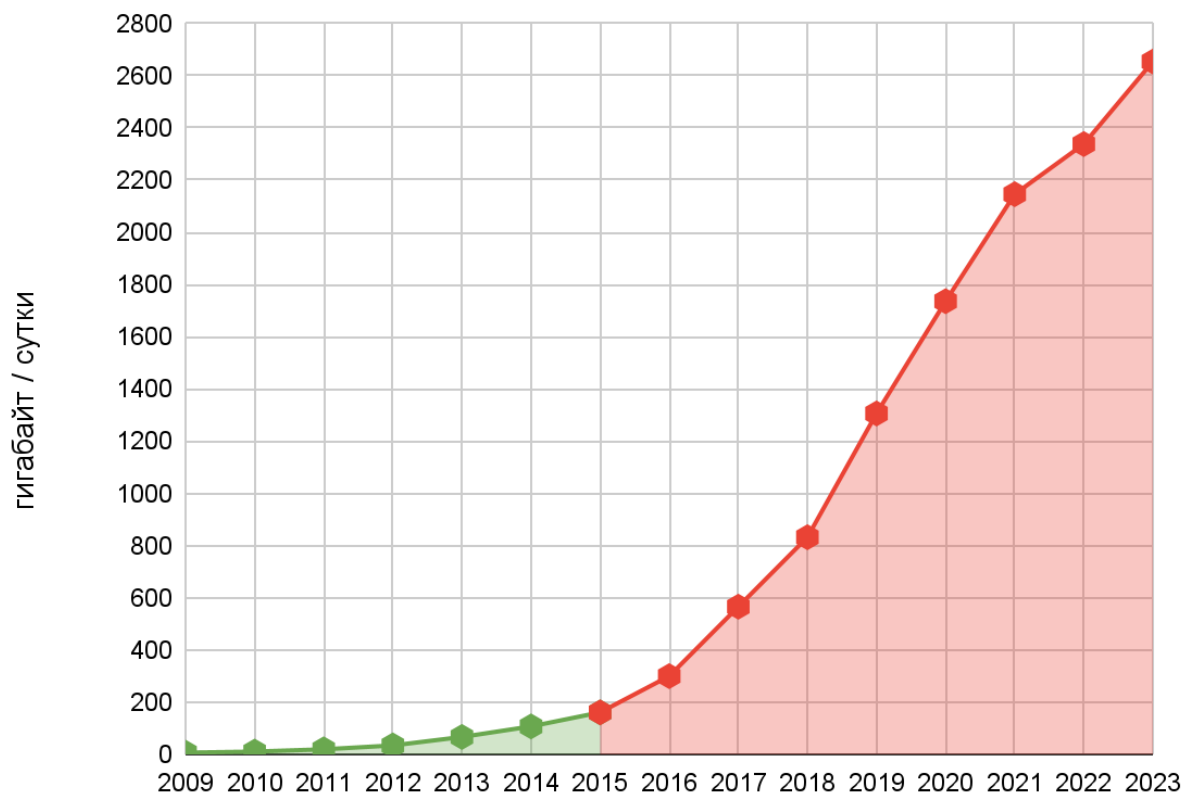
Для отправки одной посылки логистической компании необходимо обработать различные данные, включая информацию о грузе, отправителе, получателе, адресе доставки и другие детали. Размер этих данных может варьироваться в зависимости от конкретных требований компании и уровня детализации информации, но в среднем можно ожидать, что размер данных будет составлять несколько килобайт для каждой посылки. Когда количество посылок становится критичным, обычных технологий баз данных становится недостаточно.

В компании СДЭК в середине 2010-х годов технология Big Data позволила проводить экстраполяцию спроса, оптимизировать маршруты как на “микро” так и на “макро” уровне; проводить аналитику и документооборот, и что не мало

важно все эти параметры вычисляются в режиме реального времени со скоростью генерации данных. Именно в этих факторах заключаются основные преимущества технологии Биг дата.

В 2014 году СДЭК начинает внедрение технологий больших данных на основе продуктов от американской технологической компании «Teradata Corporation». Однако после политических событий весны 2022 года большинству Российских компаний как из банковского сегмента, так и логистического пришлось переходить на различные аналоги Teradata.

В данном исследовании, после консультирования со специалистами этой области, было выявлено, что в компании «СДЭК» необходимо обрабатывать 10 килобайт данных на доставку каждого отправления. На графике 2 отображено суточное количество обрабатываемых данных.

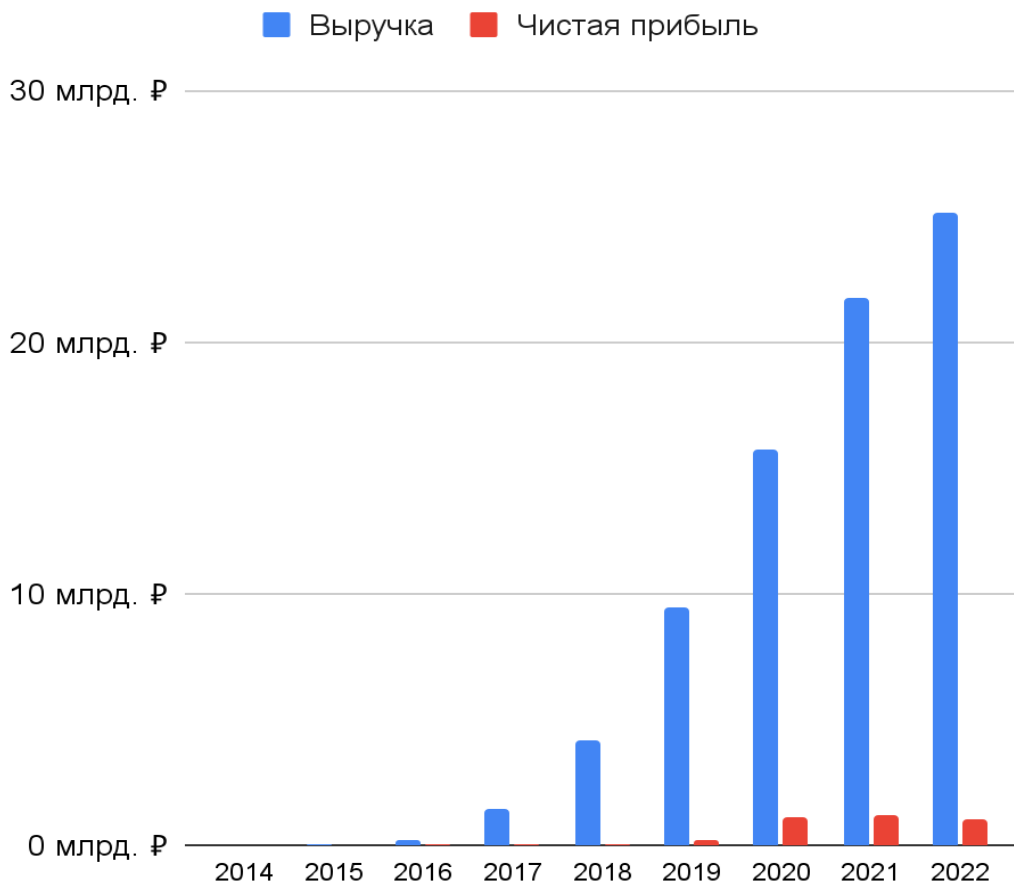


Граф. 2 Динамика количества обрабатываемой информации СДЭК

В 2015 году количество информации переступило отметку 150 гигабайт в сутки. Отправить такое количество посылок, не потеряв скорости и стоимость

отправления, стало возможно только благодаря эффективной работе технологии больших данных. В настоящее время (2023 г.) такие компании как СДЭК обрабатывают десятки терабайт в день. Представитель компании СДЭК Ксения Юркова в публикации на сайте компании «Сейчас на первый план выходит не столько наличие самих данных, сколько гибкость в их применении и правильная интерпретация. Именно последнее предоставляет шанс конвертировать инструменты аналитики из статике в бизнес-результат» [10].

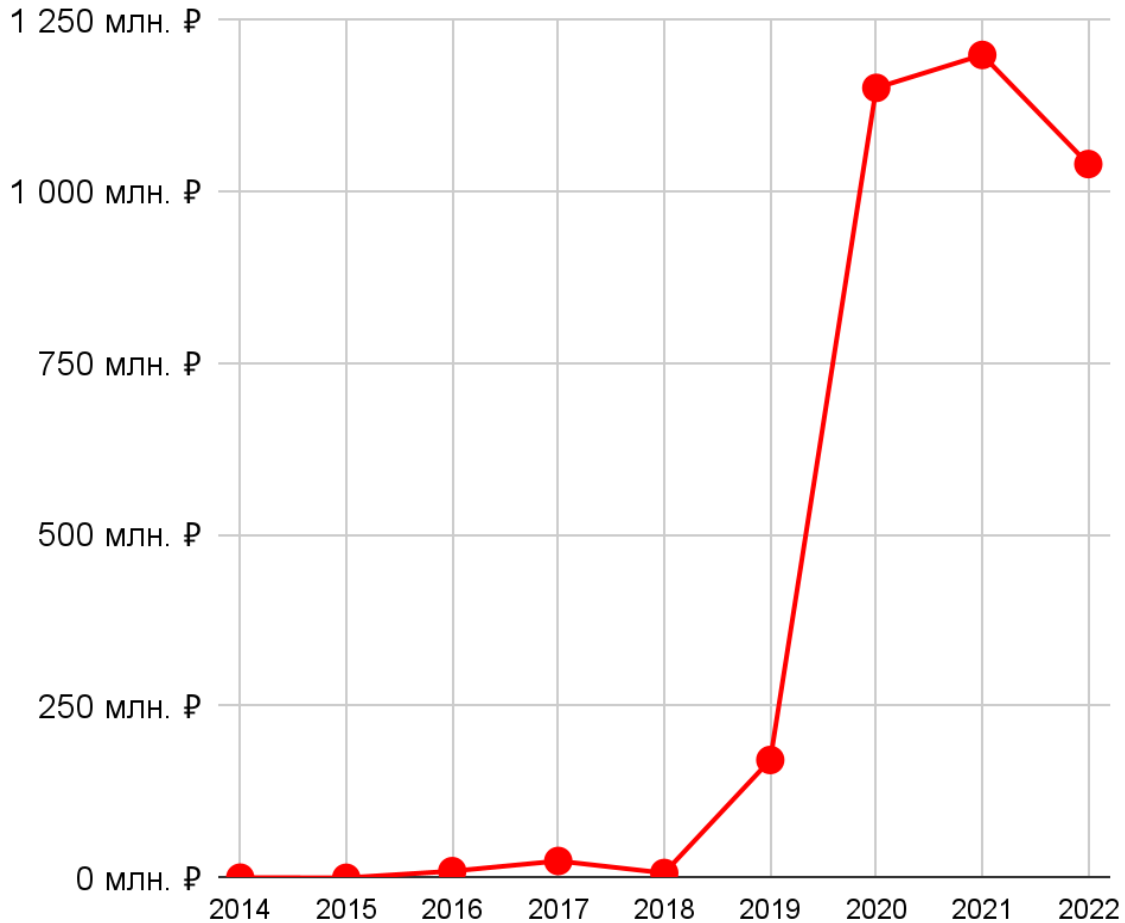
Активное экономическое развитие компания получила в период с 2014 по 2023. На основные показатели компании с течением времени влияли различные экономические, политические и технологические факторы. Инновационные технологии оказывают существенное влияние на бизнес-процессы, но основные показатели рентабельности (Граф. 3) в большей мере определены другими внешними экономическими факторами.



Граф.

3 Основные показатели рентабельности СДЭК [4]

График 4, полученный по данным из бухгалтерской отчетности ООО "СДЭК-Глобал", иллюстрирует чистую прибыль. Данный параметр позволяет оценить основную динамику экономического развития предприятия.

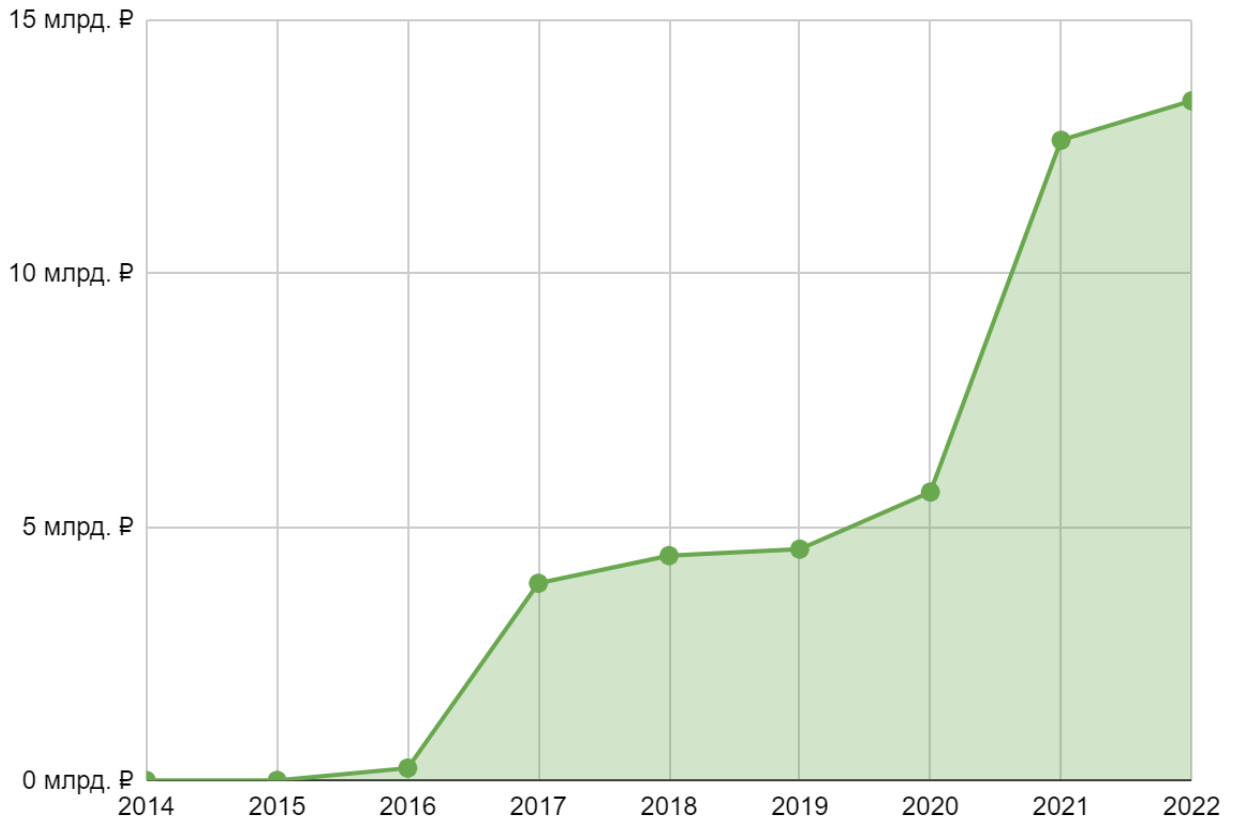


Граф.

4 Чистая прибыль СДЭК [4]

На основе аналитики бухгалтерской отчетности, оценки политической и экономической ситуации в России и времени внедрения технологии больших данных можно выявить некоторые закономерности. Сравнительный рост в 2017 году связан с внедрением новой технологии: она позволила улучшить эффективность уже существующих бизнес-процессов, что снизило расходы и напрямую отразилось на прибыли. Количество отправок и клиентов, увеличить которое сделала возможным технология Big Data после ее внедрения в 2014 году, дала ощутимые результаты в 2017 году. В 2017 резко возросла

положительная динамика количества отправок. На графике 5 показана капитализация компании СДЭК-глобал, резкий скачок капитализации приходится на 2017 год. Эффективное внедрение технологии больших данных оказало немалое влияние на величину этого скачка, а также на дальнейший рост капитализации.



Граф. 5 Динамика показателей капитала СДЭК [4]

В ближайшие годы большие данные будут столь же важны, как транзитные, энергетические и телекоммуникационные системы, с точки зрения общественного и финансового прогресса. Некоторые области, в том числе политика, экономика и общество, претерпят значительные трансформации и достигнут беспрецедентного прогресса в рамках больших данных, что глубоко повлияет на образ жизни людей, когнитивные и этические системы. Использование больших данных при анализе коммерческих экономических данных в будущем станет все более распространенным.

Большие объемы данных, собранные от различных источников, являются необходимым ресурсом для эффективной работы современного предприятия. Оптимизация производственных процессов основана на анализе и изучении этих данных с помощью огромного кластера математических и программных методов, что позволяет компаниям принимать обоснованные решения и достигать большей эффективности в различных условиях рынка.

Применение Big Data открывает новые возможности для оптимизации производственных процессов, повышения эффективности и принятия обоснованных решений. Предприятия, которые умеют правильно собирать, анализировать и использовать данные, могут получить значительные преимущества на рынке и улучшить свою конкурентоспособность.

В дальнейших исследованиях стоит обратить внимание на развитие новых методов обработки и анализа данных, а также на вопросы безопасности и этики использования Big Data. Важно провести дополнительные исследования, чтобы более глубоко понять потенциал и преимущества, которые могут быть получены с помощью Big Data в различных отраслях промышленности.

Литература

1. Алексеев К.А. Использование Big Data в современном бизнесе // Труды института системного программирования РАН. 2020. № 4. Том 32. С 7-20
2. Lynch C. Big data: how do your data grow? Nature.vol.455. № 7209. 2008. pp. 28-29
3. Прядка Р.И., Чуланова О.Л. Возможности и перспективы применения "Data Science" в управленческом консультировании // Материалы Афанасьевских чтений. 2021. №1(34). С. 45-51
4. Менщиков А.В., Перфильев В.Э., Федосенко М.Ю., Фабзиев И.Р. Основные проблемы использования больших данных в современных информационных системах // Столыпинский вестник. 2022. №1. С.316-327
5. Назаров А.Д., Плотников В.А. Data Science и обеспечение экономической безопасности в эпоху цифровой экономики // Известия СПбГЭУ. 2023. №4(142). С.151-155

6. "ООО СДЭК Глобал - отчетность и структура компании". Аудит-ИТ – UDL: https://www.audit-it.ru/buh_otchet/7722327689_ooo-sdek-global
7. “СДЭК демонстрирует уверенный рост по ряду показателей”. – UDL: <https://boomin.ru/publications/news/sdek-demonstriruet-uverennyj-rost-po-radu-pokazatelej/>
8. “Количество отправок СДЭК выросло в 2 раза за 2017 год”. – UDL: <https://logistics.ru/management/news/kolichestvo-otpravleniy-sdek-vyroslo-v-2-raza-za-2017-god>
9. "Зачем специалисту по Data Science нужна экономика". – Журнал Академии Яндекса. – UDL: <https://academy.yandex.ru/journal/zachem-spetsialistu-po-data-science-nuzhna-ekonomika>
10. Кельчевская Н.Р., Колясников М.С. Использование больших данных в стратегическом управлении знаниями компании, следующей трендам Индустрии 4.0 // Лидерство и менеджмент. 2020. №3. С. 405-422
11. “От Лямбда до Data Mesh: 7 архитектур данных для Big Data систем” – bigdataschool. – UDL: <https://bigdataschool.ru/blog/data-architectures-overview.html>
12. “СДЭК внедрил систему продвинутых отчетов сервиса PIM. Аналитика”. – 21 мая 2021 г. – UDL: <https://www.cdek.ru/ru/press/view/2021-05-21-sdek-vnedril-sistemu-prodvinutyh-otchetov-servisa-pimanalitika/>
13. Александр Петров @asash “Big Data от А до Я. Часть 1: Принципы работы с большими данными, парадигма MapReduce”. – 21 сен. 2015 г. – UDL: <https://habr.com/ru/articles/267361/>

Literature

1. Alekseev K.A. Using Big Data in contemporary business // Proceedings of the Institute of System Programming of the Russian Academy of Sciences. 2020. No. 4. Volume 32. From 7-20
2. Lynch C. Big data: how do your data grow? Nature.vol.455. № 7209. 2008. pp. 28-29

3. Pryadka R.I., Chulanova O.L. Possibilities and prospects of using Data Science in management consulting // Materials of the Afanasiev readings. 2021. №1 (34). P. 45-51
4. Menshchikov A.V., Perfiliev V.E., Fedosenko M.Yu., Fabziev I.R. The main problems of using big data in modern information systems // Stolypin Bulletin. 2022. No.1. pp.316-327
5. Nazarov A.D., Plotnikov V.A. Data Science and ensuring economic security in the era of the digital economy // Izvestiya SPbGEU. 2023. No.4(142). pp.151-155
6. "CDEK Global LLC - reporting and company structure". Audit-IT – UDL: https://www.audit-it.ru/buh_otchet/7722327689_ooo-sdek-global
7. "CDEK demonstrates steady growth in a number of indicators." – UDL: <https://boomin.ru/publications/news/sdek-demonstriruet-uverennyj-rost-po-radu-pokazatelej/>
8. "The number of CDEC shipments increased by 2 times in 2017." – UDL: <https://logistics.ru/management/news/kolichestvo-otpravleniy-sdek-vyroslo-v-2-raza-za-2017-god>
9. "Why a Data Science specialist needs economics." – Journal of the Yandex Academy. – UDL: <https://academy.yandex.ru/journal/zachem-spetsialistu-po-data-science-nuzhna-ekonomika>
10. Kelchevskaya N.R., Kolyasnikov M.S. The use of big data in strategic knowledge management of a company following Industry 4.0 trends // Leadership and Management. 2020. No.3. pp. 405-422
11. "From Lambda to Data Mesh: 7 data architectures for Big Data systems" – bigdataschool. – UDL: <https://bigdataschool.ru/blog/data-architectures-overview.html>
12. "SDEK has implemented a system of advanced reports of the PIM service. Analytics". – May 21, 2021 – UDL: <https://www.cdek.ru/ru/press/view/2021-05-21-sdek-vnedril-sistemu-prodvinytyh-otchetov-servisa-pimanalitika/>

13. Alexander Petrov @asash "Big Data from A to Z. Part 1: Principles of working with big data, the MapReduce paradigm". – September 21, 2015 – UDL: <https://habr.com/ru/articles/267361/>

© Андреева С.А., Пушкин А.Н., Ядова Н.Е., 2023 Научный сетевой журнал «Столтыпинский вестник» №12/2023.

Для цитирования: Андреева С.А., Пушкин А.Н., Ядова Н.Е. ПРИНЦИПЫ ТЕХНОЛОГИИ BIG DATA В ЭКОНОМИКЕ СОВРЕМЕННОГО ПРЕДПРИЯТИЯ//Научный сетевой журнал «Столтыпинский вестник» №12/2023.