СОВМЕСТНОЕ ИССЛЕДОВАНИЕ
VK CLOUD И ARENADATA

Технологии для работы с Big Data: готовность к использованию и основные барьеры

VK Cloud × ARENADATA

Цели исследования

Объем глобального рынка Big Data в 2022 году достигнет 274,3 млрд долларов. Этому способствует то, что 97,2% организаций инвестируют в решения для работы с большими данными.

Положительную динамику демонстрирует и российский рынок: по оценкам Ассоциации больших данных, к 2024 году он вырастет до 300 млрд рублей с примерно 30 млрд рублей в 2022 году.

Цель исследования VK Cloud и Arenadata — выявить подходы и планы российских компаний по работе с большими данными и определить, как они могут меняться в 2022 году.
В исследовании мы также сфокусировались на следующих вопросах:

- готовность к использованию технологий для работы с большими данными и основные барьеры внедрения Big Data-проектов;
- особенности внедрения Big Data-проектов: задачи, инструменты, результаты;
- проблемы и сложности при работе с большими данными;
- применение облачных технологий в работе с большими данными.

© VK Cloud, 2022 г.

Методология исследования

Исследование проводилось в два этапа:

1. **Опрос**. Мы разослали респондентам анкету из 13 вопросов. Результаты анкетирования позволили сформировать общее видение направлений по работе с Big Data-проектами на российском рынке.
2. **Глубинное интервью**. Провели глубинные интервью, в ходе которых получили подробные ответы на вопросы о планах работы с большими данными от CIO, CDO, CDTO крупных российских компаний.

© VK Cloud, 2022 г.

Респонденты

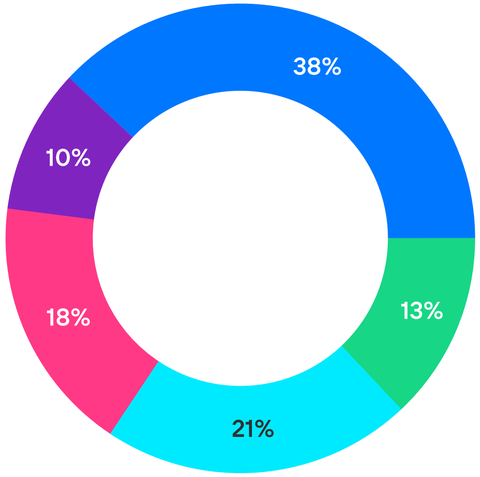
В исследовании приняли участие 150 человек.

Респондентами выступили главы ИТ-департаментов, CIO, CTO, CDTO, CDO крупных российских компаний.
Из 150 респондентов, принявших участие в исследовании, 17% — представители сферы FMCG (fast-moving consumer goods, товары повседневного спроса), 16% — розничной торговли и e-commerce, 15% — банковских и страховых организаций, 10% — машиностроения, 7% — организации топливно-энергетического комплекса, 6% — предприятия медицинской и фармацевтической отрасли.

© VK Cloud, 2022 г.

Как давно компании используют решения для работы с большими данными?

Как давно вы используете решения для работы с Big Data?



- Менее года
- 1-2 года
- 3-4 года
- Более 5 лет
- Не используем

Исследование демонстрирует, что решения для работы с большими данными уже используют 62% респондентов. При этом 10% компаний-респондентов работают с большими данными меньше 1 года. В течение последних двух лет такие проекты внедрены в компаниях 28% респондентов. Эта динамика может объясняться повсеместной цифровизацией предприятий в период пандемии, увеличением объема собираемых данных, стремлением повысить прибыль в условиях высококонкурентного бизнеса (например, в розничной торговле).

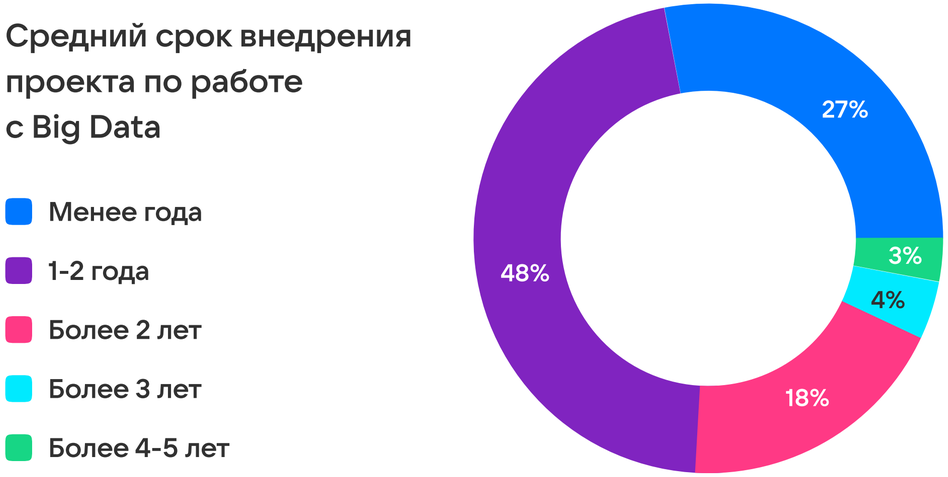
38% респондентов не работают с большими данными. Это связано со следующими факторами:

- отсутствие больших данных и потребности в их обработке;
- нехватка специалистов с нужным уровнем компетенций;
- разрозненность ИТ-инфраструктуры;
- проблема каталогизации данных;
- отсутствие механизмов интеграции результатов аналитики в производственные процессы и в бизнес-процессы;
- несформированная культура работы с данными.

© VK Cloud, 2022 г.

Как быстро внедряются проекты по работе с большими данными

Средний срок внедрения проекта по работе с Big Data



Внедрение проектов по работе с большими данными — сложный и длительный процесс, который требует комплексной подготовки данных (сбор, хранение, обработка), инфраструктуры, операционных процессов и специалистов.

В большинстве случаев на внедрение проекта нужно от 1 до 2 лет — так ответили 48% участников опроса. 27% респондентов заявили, что внедрили Big Data-решение менее чем за год. 18% компаний смогли реализовать проект более чем за 2 года. 7% респондентов отметили, что на запуск проекта ушло от 3 до 5 лет.

Срок внедрения зависит не от профиля компании, а от:

- масштаба проекта;
- степени подготовки инфраструктуры;
- компетенций команд и процессов разработки;
- готовности данных и их качества;
- культуры работы с данными;

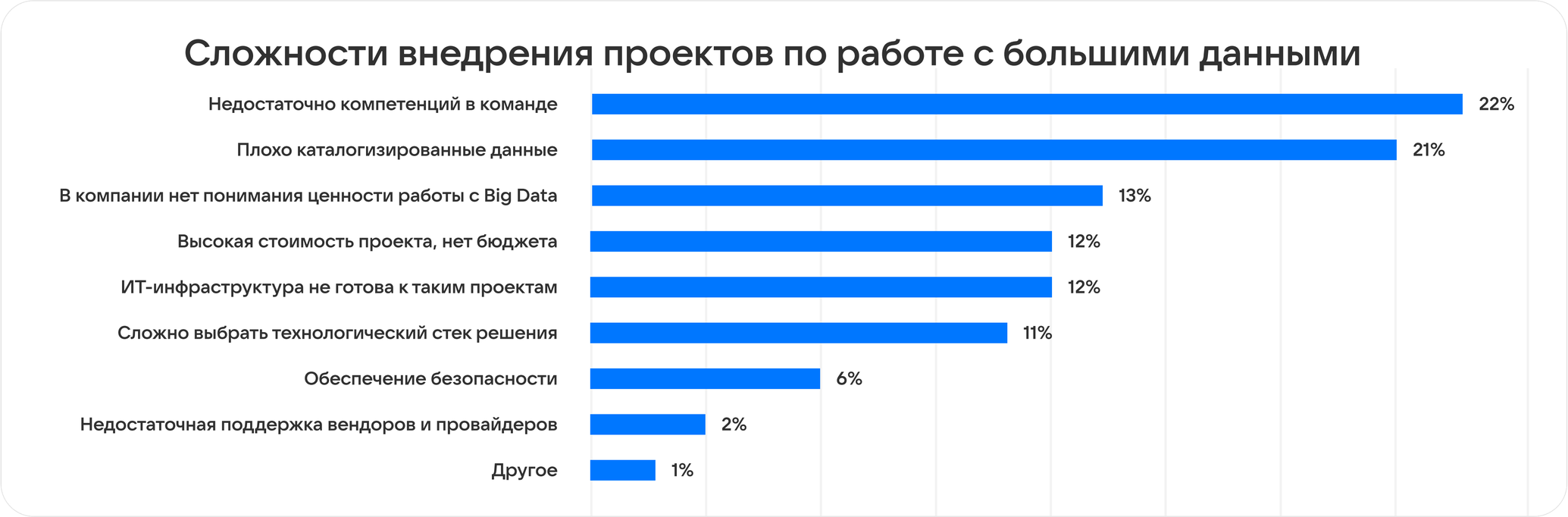
Узкое место на каждом из этих этапов может становиться причиной, по которой внедрение решения может затянуться.

При этом внедрение проекта не означает завершение работы над ним: система может требовать регулярной доработки аналитической модели. Принцип «внедрил и забыл» здесь не работает.

© VK Cloud, 2022 г.

Основные сложности внедрения проектов по работе с большими данными

По результатам опроса, среди главных препятствий при внедрении проектов для работы с большими данными — отсутствие компетенций в команде, низкое качество данных, отсутствие понимания ценности работы с большими данными, плохая интеграция компонентов ИТ-инфраструктуры и сложность выбора технологического стека решений.



© VK Cloud, 2022 г.

Недостаточно компетенций в команде

Чаще всего работать с большими данными не позволяет отсутствие компетенций в команде — об этом заявили 22% респондентов.

В части компаний эту проблему решают, привлекая внешних экспертов и агентства для определенных этапов проекта или его реализации «под ключ». В ближайшее время такой подход сохранится: на рынке дефицит специалистов по работе с большими данными (уровней middle и senior), поэтому формировать команды сложно и дорого.

Еще одна стратегия — обучение сотрудников или привлечение профильных экспертов в качестве консультантов. Такой подход предполагает учет некоторых факторов: на обучение сотрудников нужно время, а при работе консультантов важно контролировать, чтобы рекомендации внедрялись на практике, а не оставались на бумаге.

Ставку на построение Inhouse-центров компетенций в области больших данных делают преимущественно компании, где проекты с большими данными активно развиваются, а массивы данных стремительно растут.

Плохо каталогизированные данные

Респонденты признают, что низкое качество данных является проблемой: плохие данные снижают точность и объективность аналитики.

Среди способов, которые компании рассматривают для решения этой проблемы, инвентаризация собранных данных, ведение каталогов данных и устранение факторов получения некачественных данных.

Однако респонденты отмечают, что даже при выделенных подразделениях Data Quality, Data Governance проблема качества данных сохраняется. Это связано с отсутствием культуры работы с данными, а также с избыточным или недостаточным количеством их источников — не все данные информативны и требуют учета.

© VK Cloud, 2022 г.

Нет понимания ценности работы с Big Data

Исследование выявило, что важной проблемой в компаниях является отсутствие понимания ценности работы с большими данными — об этом говорят 13% респондентов.

Зачастую ИТ-департаментам приходится доказывать важность сбора и анализа больших данных и ценность интегрирования результатов аналитики в операционную деятельность. Это усложняет и замедляет внедрение проектов, особенно если сотрудники саботируют процесс и отказываются принимать решения, сделанные «машиной».

Опрос показал, что преодолеть это препятствие возможно, только если все бизнес-пользователи понимают важность качественного сбора данных и использования результатов аналитики в операционной деятельности. Как правило, это возможно в компаниях, где такая практика регламентирована, а инициаторами и заинтересованными сторонами в изменении культуры работы с данными выступают CDTO или CIO, CFO, CBDO, CMO.

Высокая стоимость проектов

Несмотря на появление доступных инструментов для сбора и хранения данных, а также на снижение стоимости оборудования и технологий для работы с большими данными, Big Data-проекты по-прежнему требуют значительных инвестиций в инфраструктуру и программное обеспечение, реорганизацию бизнес-процессов, усиление компетенций команд.

12% участников опроса отметили стоимость в качестве сдерживающего фактора на пути внедрения Big Data-проектов.

На такую позицию может влиять сложность оценки экономического эффекта от внедрения проектов для работы с большими данными: при высокой стоимости проектов компании не всегда видят возможности монетизации собранных данных.

© VK Cloud, 2022 г.

ИТ-инфраструктура не готова к таким проектам

Для работы с большими данными нужна надежная производительная ИТ-инфраструктура. Она должна охватывать все департаменты компании, позволять работать с большим количеством источников данных и легко масштабироваться. Такую инфраструктуру важно не просто создать — ее нужно регулярно модернизировать и поддерживать, что сопряжено с необходимостью постоянных инвестиций. Неготовность инфраструктуры является главным препятствием для 12% респондентов исследования.

Сложно выбрать технологический стек решения

Технологий для работы с большими данными много. Компаниям сложно выбрать оптимальные решения для своих задач. Усложняет выбор и тот факт, что проприетарные решения дорогие, а Open-Source-технологии предъявляют высокие требования к компетенции специалистов.
По мнению 11% респондентов, это затрудняет внедрение Big Data-проектов.

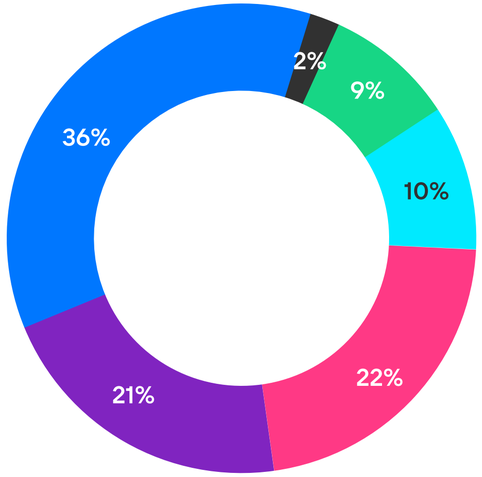
Примечание. Решить основные проблемы помогает миграция в облако. Развертывание проектов в облаках позволяет привлекать меньше инженеров с узкой специализацией, снижать затраты на инфраструктуру, быстро создавать пилотные проекты для проверки гипотез, получать доступ к удобным Big Data-инструментам.

© VK Cloud, 2022 г.

Статистика привлечения внешних специалистов

Привлекаете ли вы внешних специалистов для внедрения проектов

- Да, привлекаем интеграторов для разработки проекта «под ключ»
- Нет, развиваем компетенции внутри компании или берем новых сотрудников в штат
- Да, привлекаем агентства для решения локальных задач
- Нет, у команды есть все компетенции
- Да, работаем с консалтинговыми компаниями
- Другое



36% компаний привлекают интеграторов для внедрения проекта «под ключ», 22% пользуются услугами аналитических агентств, 9% работают с консалтинговыми компаниями.

Причина, по которой компании обращаются к внешней экспертизе, — нехватка кадров с компетенциями, необходимыми для работы с различными типами данных и технологий, а также потребность в экспертизе и консультационных услугах.

21% респондентов формируют экспертизу внутри компании. Ставку на Inhouse-центры ИТ-компетенций объясняют:

- ростом сложности проектов,
- быстрой сменой технологий,
- необходимостью глубокого погружения в логику бизнес-процессов компании,
- важностью внедрения результатов аналитики данных в операционные процессы.

Исследование показало, что внутреннюю экспертизу чаще развивают крупные компании со зрелым подходом к работе с большими данными.

© VK Cloud, 2022 г.

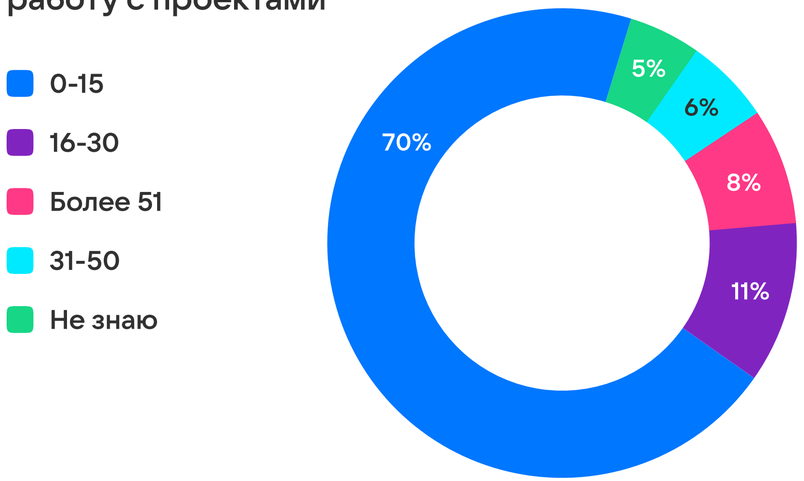
Сколько людей задействуют для работы с большими данными

В ходе глубинных интервью мы узнали, что российские компании по-разному подходят к формированию отделов по работе с Big Data. Прежде всего, на это влияют профиль и масштаб компании. От этого же зависит и количество людей, задействованных в работе с большими данными.

В небольшой компании это могут быть аналитик и представитель ИТ-подразделения: согласно ответам 70% респондентов, в команду по работе с большими данными входят не больше 15 человек.

Исключение — крупные компании, в которых за работу с большими данными могут отвечать несколько подразделений и специалистов с разными ролями: Data Science, Data Engineer, системный аналитик, архитектор, бизнес-аналитик, продакт-менеджер и другие.

Количество сотрудников, которые обеспечивают работу с проектами



© VK Cloud, 2022 г.

Что осложняет работу с Big Data-решениями

Анализ показал, что трудности при работе с Big Data во многом схожи со сложностями внедрения таких проектов. Например, 32% респондентов сталкиваются с нехваткой кадров, 21% — со сложностью выбора инструментов, 15% — с трудностями из-за необходимости капитальных вложений.

Участники исследования отмечают, что появление подобных сложностей часто связано с ошибками при стратегическом планировании: нередко проблемы, возникающие при внедрении проектов, решаются ситуативно, из-за чего они появляются снова уже при работе с большими данными.

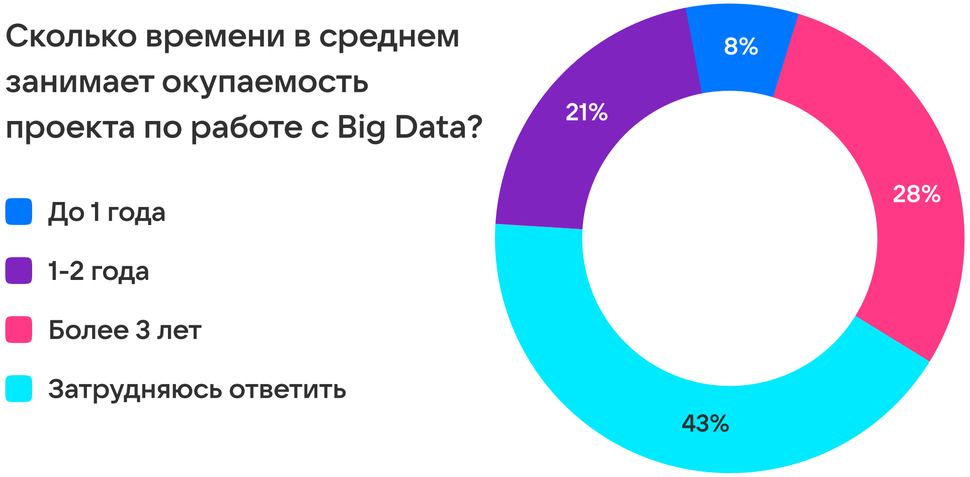
Это подтверждает важность анализа и аналитики всех возникающих проблем: компании, следующие этому принципу, сталкиваются с трудностями реже.

Трудности при работе с Big Data-проектами

© VK Cloud, 2022 г.

Статистика окупаемости проекта по работе с большими данными

Сколько времени в среднем занимает окупаемость проекта по работе с Big Data?



28% респондентов ответили, что окупаемость проектов занимает более 3 лет. 21% опрошенных заявили, что для этого нужно 1–2 года. 8% участников исследования отметили, что получают эффект от внедрения решений для работы с большими данными менее чем за 1 год.

Разные темпы окупаемости объясняются разным масштабом внедряемых проектов и разным эффектом от использования результатов анализа больших данных.

43% респондентов не смогли ответить на вопрос о сроках окупаемости Big Data-проектов. Это указывает на то, что компании не всегда видят ценность работы с большими данными и выгоду от использования результатов аналитики. Из-за этого компании часто не готовы инвестировать в разработку проектов по работе с Big Data.

Исследование показало, что оценить рентабельность инвестиций в большие данные компаниям часто мешает:

- сложность определения метрик успешности реализации проекта;
- отсутствие концепции монетизации результатов анализа больших данных;
- слабая интеграция результатов аналитики в бизнес-процессы;
- отсутствие культуры работы с данными в компании: сотрудники могут игнорировать выводы аналитики и принимать решения, основываясь на «традициях» и наработанном личном опыте («неверие бизнеса в выводы, сделанные машиной, а не человеком»).

© VK Cloud, 2022 г.

Полезными практиками могут стать:

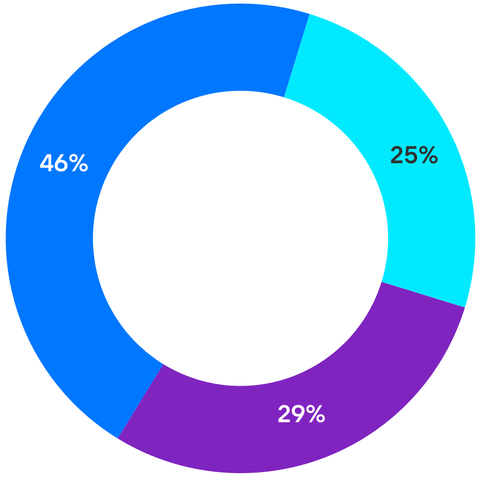
1. Запуск пилотных проектов или небольших проектов для оценки узких мест реализации решения и возможностей применения результатов аналитики.
2. Подбор технологий, которые позволят максимально эффективно развернуть проект с точки зрения затрат и получения результатов.

© VK Cloud, 2022 г.

Статистика использования облачных решений для работы с Big Data

Применение облачных решений для работы с большими данными

- Да, используем облачные решения
- Нет, планируем использовать облачные решения
- Нет, в планах нет



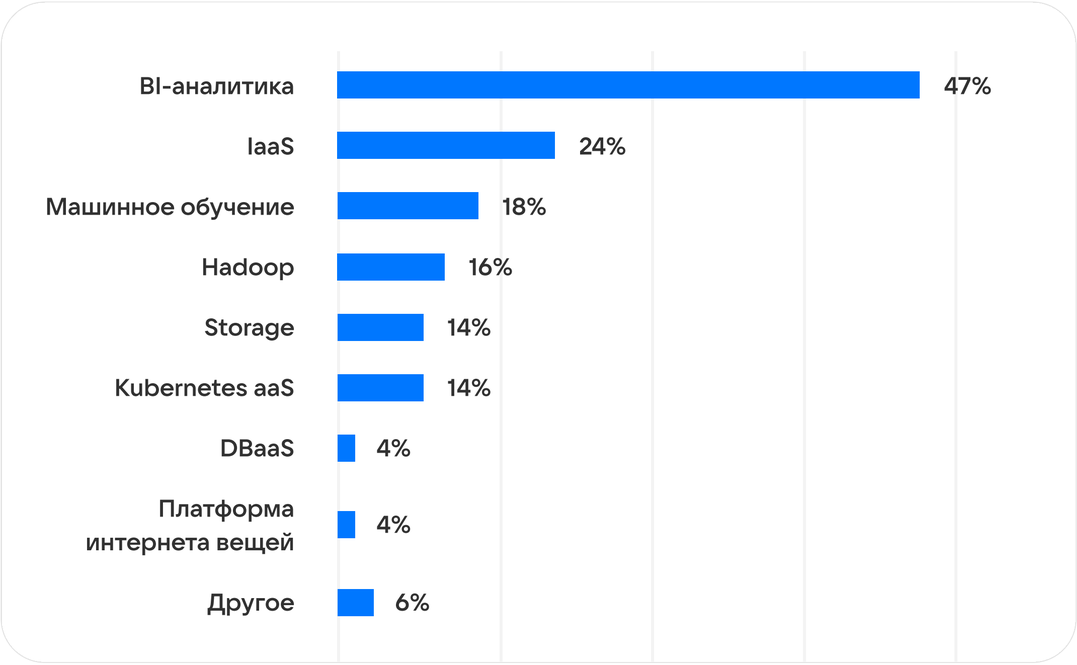
Облачные решения упрощают работу с большими данными: помогают сократить издержки, переложить часть задач по поддержке и настройке на облачного провайдера и работать с данными, привлекая меньше инженеров с узкой специализацией.

По данным нашего исследования, в России облачные решения для работы с Big Data-проектами уже используют 46% компаний, а 29% планируют начать работать с облаками в ближайшее время. Участники глубинных интервью отметили, что использовать преднастроенные сервисы в облаке выгодно: не нужно самостоятельно администрировать сервисы, обслуживать инфраструктуру, выполнять базовую настройку. Облако лучше и потому, что не каждая компания может позволить себе Inhouse-команду: кадров не хватает, а держать их в штате дорого.

При этом по-прежнему сдерживающим фактором респонденты считают требования информационной безопасности. Бизнес-подразделения видят преимущества работы с облаками, но миграцию в облако часто усложняют возражения со стороны служб информационной безопасности: они не доверяют облакам, даже если облака соответствуют государственным требованиям к защите персональных данных и безопасности (ФЗ-152, УЗ-1).

© VK Cloud, 2022 г.

Какие облачные решения в проектах по работе с большими данными используют чаще



Анализ ответов показал, что наиболее востребованным облачным решением являются инструменты для BI-аналитики: их применяют в компаниях 47% респондентов.

24% респондентов ответили, что работают в облаке с IaaS-решениями: виртуальными серверами и сетями, сервисами резервного копирования, средствами управления ресурсами. IaaS выбирают для экономии на инфраструктуре, использования современных технологий, снижения нагрузки на ИТ-отдел.

Для 18% опрошенных основным облачным решением являются инструменты для машинного обучения. Как правило, это крупные компании с большим бюджетом и объемом данных. Российские компании еще только приступают к использованию технологий машинного обучения в проектах — оценивают стоимость внедрения и рентабельность.

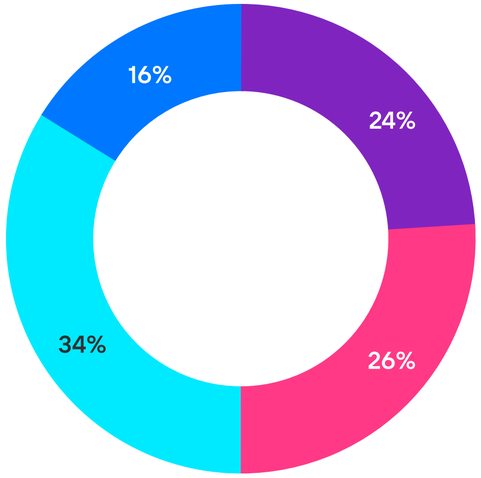
Инструменты для обработки и хранения данных, облачные базы данных и другие инструменты замыкают список. В России начинает формироваться тренд на применение этих решений — в перспективе они могут стать основными драйверами российского рынка облачных решений.

© VK Cloud, 2022 г.

Какие типы облачных решений используют российские компании

Типы облачных IaaS-решений, которые используют респонденты

- Облака у облачного провайдера
- Затрудняюсь ответить
- Виртуализированная инфраструктура On-premise и облака провайдера
- Виртуализированная инфраструктура On-premise



Исследование показало, что в компаниях 46% респондентов уже используют облачные технологии. При этом чаще всего (в 34% компаний) используют виртуализированную инфраструктуру.

26% участников исследования ответили, что используют виртуализированную инфраструктуру и облака провайдера. Комплексную миграцию респонденты объяснили необходимостью оптимизации операционных и бизнес-процессов.

Реже, в 16% случаев, российские компании используют только облака для сбора, хранения и обработки данных.

Примечание. Для всех этих задач можно использовать облако VK Cloud. На платформе доступен большой набор сервисов для сбора, хранения, анализа и визуализации больших данных.

© VK Cloud, 2022 г.

Основные проблемы миграции в облака

© VK Cloud, 2022 г.

Опыт компаний показал, что миграция в облако выгодна для бизнеса: повышает безопасность, упрощает соблюдение законов, ускоряет выход на прибыль.

Вместе с тем компании продолжают опасаться мигрировать в облако. Так, для 18% респондентов критическими стали проблемы обеспечения безопасности: нужно защитить данные, распределить права доступа, организовать безопасный канал сбора и получения данных из облака, настроить защиту от внешних угроз.

По мнению 16% опрошенных, главное препятствие — необходимость соблюдения 152-ФЗ при работе с персональными данными. Тенденция последних лет показывает, что проблема остается актуальной, однако почти все облачные провайдеры сертифицировали свою инфраструктуру в соответствии с ФЗ-152. Поэтому решение, которое размещено в облаках, это требование выполняет.

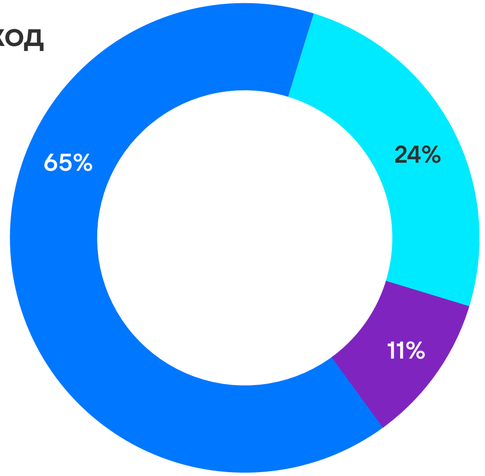
Реже мигрировать в облака мешает разрозненность инфраструктуры, недостаток бюджета и нехватка кадров — на эти проблемы указали по 13% респондентов. По мнению участников исследования, часто преодолеть эти преграды мешает отсутствие стратегического планирования внутри компании.

© VK Cloud, 2022 г.

Как изменится подход к затратам на инфраструктуру для Big Data-проектов в 2022 году

Планируете ли вы изменить подход к затратам на инфраструктуру для Big Data в 2022 году?

- Нет, не меняем
- Да, увеличим затраты
- Да, сократим затраты



Наше исследование показало, что российские компании, несмотря на возникающие трудности, готовы работать с большими данными, как и прежде. 65% респондентов заявили, что их компании сохранят бюджет на инфраструктуру для Big Data на запланированном уровне. Согласно анализу результатов глубинного интервью, это компании, которые давно внедрили Big Data-проекты, вышли на их окупаемость и эффективно используют результаты аналитики.

24% респондентов увеличат затраты. Это представители банков и финансовой индустрии, розничной торговли, FMCG, ТЭК. В ряде случаев затраты связаны с необходимостью смены вендора, изменения подхода к инфраструктуре и ее поставщика. Так, компании, работающие в стратегически важных для государства областях, инвестируют в перезапуск Big Data-проектов на основе отечественных платформ и облаков.

© VK Cloud, 2022 г.

Одновременно с этим 11% компаний планируют сократить затраты. По мнению респондентов, их ожидания от работы с большими данными не оправдались: инфраструктура и технологии для работы с данными «избыточные» и сложные для использования. По оценке специалистов VK Cloud и Arenadata, это может быть следствием неверного расчета рентабельности проекта и определения его целей в момент внедрения: ошибочно выбраны инфраструктура, инструменты, способы монетизации результатов аналитики.

Отдельная причина сокращения бюджетов — внешние факторы, которые стимулируют ряд компаний переходить «в режим экономии». Вместе с тем, если проекты уже реализованы, работают и «устраивают компанию», их не всегда выгодно сворачивать и проще поддерживать.

© VK Cloud, 2022 г.

Ключевые выводы

1. С Big Data-решениями работают уже 62% российских компаний. Многие обрабатывают данные уже более 3 лет.
   Это демонстрирует, что рынок Big Data в России сформирован, а подходы к работе устоялись.

2. В России сохраняется рост интереса к работе с большими данными: 28% компаний начали работать с Big Data в течение последних двух лет. При этом чаще всего внедрение Big Data-проектов занимает от 1 до 2 лет.

3. На всем цикле работы с большими данными — от внедрения проектов до миграции в облако — компании сталкиваются со схожими проблемами: недостатком финансирования, нехваткой кадров, неготовностью ИТ-инфраструктуры. Преодолеть эти трудности помогает в том числе тщательное планирование ресурсов и целей, а также грамотный подход к реализации Big Data-проектов.

4. Большинство компаний привлекают внешних специалистов и интеграторов для работы над проектами по созданию Big Data-решений.

5. Почти половина российских компаний работает с большими данными в облаках или планирует использовать облачные решения. Из облачных решений чаще используют инструменты для BI-аналитики и машинного обучения, сервисы для построения инфраструктуры и обработки данных.

6. Переход в облако — предпочтительный выбор российских компаний, которым необходимо оперативно решить задачи расширения инфраструктуры, дефицита компетенций и инструментов, чтобы обеспечить возможность масштабирования проектов или построения новых решений. Облачные технологии позволяют оперативно запускать пилотные проекты для оценки потенциала работы с Big Data для бизнеса. Это может быть хорошей практикой для поиска дополнительных бизнес-выгод в период нестабильности и неопределенности.

© VK Cloud, 2022 г.

7. Российский рынок Big Data-решений достаточно зрелый: есть технологический стек, выявлены типовые проблемы внедрения проектов, определены способы их решения. Это снижает порог входа в работу с технологиями для новых компаний: они могут использовать сформированный опыт, пробовать внедрять решения через Proof-of-concept или пресейл-проекты для оценки возможностей работы с Big Data и экономического эффекта от внедрения решений.

8. Несмотря на трудности и ограничения, российские компании сохраняют объем инвестиций в проекты по работе с большими данными. В ближайшее время компании будут нацелены прежде всего на поддержание и сохранение текущих Big Data-проектов.

© VK Cloud, 2022 г.