

26 февраля 2025

Динамика, перспективы и проблемы российского рынка больших данных (Часть 2)

Вторая часть обзора российского рынка больших данных. Первая часть [здесь](#).

ОБЛАСТИ ПРИМЕНЕНИЯ, ПРОБЛЕМЫ И ПЕРСПЕКТИВЫ

В упомянутом ранее отчете «Big data: выгодное вложение или дорогостоящий эксперимент?», со ссылкой на исследования TAdviser отмечается, что «Основными потребителями технологий big data являются банки, телеком-операторы и крупные ритейлеры. Главные проблемы развития направления больших данных — нехватка квалифицированных кадров, отсутствие достаточного опыта российских внедрений, а также высокая стоимость решений».

На вопрос «Согласны ли вы с этими утверждениями об основных потребителях технологий big data и проблемах развития направления больших данных?» коммерческий директор Группы Arenadata Андрей Жуков ответил так: «Развитие рынка больших данных сопровождается рядом



сложностей, которые сдерживают его потенциал. Многие компании сталкиваются с дефицитом квалифицированных кадров, имеющих практический опыт работы с технологиями аналитики данных, что замедляет процесс внедрения ИТ-продуктов и приводит к ошибкам на этапе реализации. Высокая стоимость решений также создаёт барьеры, особенно для среднего бизнеса, который не всегда может позволить себе дорогостоящие инфраструктуры и поддержку. Решение этих проблем требует комплексного подхода. Важнейший шаг — активное инвестирование в подготовку кадров через сотрудничество с вузами и реализацию внутренних образовательных программ. Использование готовых и зарекомендовавших себя решений вендоров сможет снизить порог входа для внедрения технологий обработки больших данных. Кроме того, комбинирование инфраструктуры on-prem с облачными платформами позволит компаниям использовать гибкие модели финансирования реализации data-проектов и быстрее адаптироваться к изменениям. При успешном решении этих задач рынок больших данных в России сохранит высокие темпы роста и будет способствовать цифровой трансформации бизнеса».

Руководитель продукта Artimate, компания «Пруфтек ИТ» Дмитрий Кошкин комментирует ситуацию так: «Однозначно основными потребителями технологий big data являются банки, телеком-операторы и крупные ритейлеры». Эти отрасли **наиболее активно внедряют аналитические инструменты, что обусловлено рядом факторов:**

- Высокая конкуренция и борьба за клиента. В телекоме наблюдается низкий барьер для смены оператора, а в банковском секторе — высокая

доступность карт и возможность мгновенного перевода зарплатных проектов.

- Сетевой эффект в ритейле. Крупные торговые сети конкурируют за покупателя в одних и тех же локациях, а значит, активно применяют инструменты big data для анализа потребительского поведения, прогнозирования спроса и персонализированного маркетинга.
- Огромные массивы данных. Во всех этих отраслях ежедневно генерируются колоссальные объемы информации, которые необходимо обрабатывать в режиме реального времени.

«Что касается ключевых проблем внедрения big data, — продолжил эксперт, — мы бы добавили еще один важный аспект — зрелость рынка и готовность к внедрению технологий. Например, в сегменте AIOps (Artificial Intelligence for IT Operations, методика управления ИТ-инфраструктурой с использованием технологий искусственного интеллекта), который напрямую связан с анализом больших данных в ИТ-инфраструктуре предприятий и организаций. В данном сегменте Северная Америка опережает Россию: там уже 60-70% крупных компаний используют подобные решения, в то время как в нашей стране этот рынок только формируется. Это говорит о том, что, помимо кадрового и финансового барьеров, компаниям еще предстоит преодолеть организационные и технологические вызовы, связанные с переходом к более продвинутым моделям работы с данными».

А вот мнение технического директора DIS Group Олега Гиацинтова: «Да, основными потребителями технологий big data по-прежнему являются финансы, ритейл и телекомы, но уже видно применение таких решений в

промышленности и государственном секторе. Применимость решений big data фактически не ограничена отраслями. Вопрос в наличии задач для аналитики. Что же касается кадров, то на рынке всегда будет недостаток специалистов нужных направлений. Просто надо их растить и развивать таким образом рынок».



Исторически финансовая сфера, телеком и ритейл были и остаются наиболее передовыми отраслями с точки зрения генерации, обработки и анализа большого объема данных. При этом мы видим, что крупные проекты начали появляться в таких сферах как фармацевтика или, например, пищевая промышленность, поскольку технологические процессы порождают достаточный объем данных, в том числе для онлайн-аналитики и ИИ. Также инструментами для сбора, хранения и анализа данных начинают активно пользоваться добывающие корпорации и DIY-компании.

На отечественном рынке достаточно много сильных игроков и зрелых продуктов. Однако, если говорить о визуализации данных, отчетности, ИИ и машинном обучении, зачастую внедрение или импортозамещение таких решений является более дорогостоящим, чем на зарубежных аналогах. Причина ясна: имеющиеся российские продукты часто дорабатываются под запросы конкретного заказчика, а не внедряются как унифицированное коробочное решение. В части обработки данных (хранилища, платформы данных) дело обстоит лучше — эти решения дольше на рынке страны и являются достаточно зрелыми

Александр Зенькович,

Руководитель направления ML и Big Data компании «ДАР»

(входит в ГК «КОРУС Консалтинг»)

О ХРАНИЛИЩАХ, ОЗЕРАХ И БОЛОТАХ ДАННЫХ

Для понимания эволюции технологий хранения и обработки больших данных напомним: концепция Data Warehouse (DWH, хранилища данных) была предложена в 1988 г. В 1990-е появились первые корпоративные хранилища данных (Enterprise Data Warehouses, EDW). Их появление было обусловлено тем, что компании столкнулись с проблемой данных, разбросанных в разных системах (CRM, ERP, бухгалтерия и так далее). Подход DWH позволил централизовать данные для отчетности и аналитики. При этом для работы использовались различные реляционные базы данных. В итоге создавалось высококачественное структурированное хранилище, в которое данные загружались через процессы ETL (Extract, Transform, Load), проходили очистку и оптимизировались для аналитики.

Однако в 2010-х годах крупные ИТ-компании столкнулись с тем, что традиционные DWH-хранилища не справляются с неструктурированными данными (логами, видео, сенсорными данными). Решением проблемы стал подход Data Lake («озера данных»), который позволял хранить любые данные (структурированные, неструктурированные, полуструктурированные) в сыром виде. При этом данные загружались в «озеро» и обрабатывались по мере необходимости. Данный подход позволил аналитикам использовать для анализа данных не только традиционные методы, но и технологии ИИ.

Со временем многие предприятия обнаружили, что их «озера данных», куда время от времени загружались сырые данные не очень высокого качества, стали превращаться в «болота данных». Из которых вытаскивать нужные данные порой приходилось, образно говоря, с таким же трудом, как «из болота тащить бегемота». Поэтому в 2017-2018 гг. была предложена гибридная архитектура Data Lakehouse («озеро-хранилище»), объединяющая преимущества подходов Data Lake (гибкость и дешевизна хранения больших объемов данных) и Data Warehouse (структурированность и высокая производительность запросов). По прозвучавшим недавно оценкам некоторых аналитиков (см., например, здесь), основными факторами всё продолжающегося роста популярности «озер-хранилищ» данных являются экономическая эффективность, возможность унифицированного доступа к данным и повышенная простота их использования.

В ЗАКЛЮЧЕНИЕ

Надо отметить, что «конструкций под ключ» для хранения и обработки больших данных никто не предлагает. Однако из четырехцветной «Тепловой карты импортозамещения ПО (рис. 3), содержащейся в недавно обнародованном 30-страничном отчете Альфа Банка «Импортозамещение в IT-секторе. Итоги и перспективы 2025+» следует, что с отечественными «строительными блоками» для практической реализации различных архитектур хранения данных дела обстоят относительно неплохо.

Во всяком случае, ситуация с наличием в Реестре отечественного ПО программных продуктов, функциональность которых соответствует таким классам как «Средства анализа данных» и «Средства обработки и визуализации массивов данных», складывается значительно лучше, чем, к примеру, с наличием в нём софта категорий «Промышленное ПО» и «Средства управления процессами организации». Это вселяет надежду, что российский рынок больших данных (если понимать под ним доходы отечественных разработчиков ПО и системных интеграторов) ждет достаточно светлое будущее. Однако всегда надо помнить, что от внезапного появления так называемых «черных лебедей» не застрахована ни одна высокотехнологичная компания.

