

23 декабря 2016

Обзор российского рынка Big Data

Российский рынок Big Data находится на начальной стадии развития. Основными потребителями таких технологий являются банки, телеком-операторы и крупные ритейлеры. Главными проблемами развития направления больших данных являются нехватка квалифицированных кадров и отсутствие достаточного опыта российских внедрений.

Состояние российского рынка Big Data

Российский рынок Big Data находится на начальной стадии развития и под этим термином часто понимаются традиционные BI-подходы. Основными потребителями технологий больших данных, так же, как и основными носителями больших объемов данных, являются компании в банковском секторе, телекоме и торговле. Для них анализ больших объемов данных, связанных с анализом платежеспособности клиентов, потребительским поведением и рыночной конъюнктурой является важнейшим инструментом для поддержания конкурентного преимущества.

В последние годы во всех компаниях из большой тройки мобильных операторов появились подразделения, специализирующиеся на работе с большими данными, причем они являются не просто информационными подразделениями для развития клиентских профайлов, они являются бизнес-единицами, которые призваны генерировать дополнительную прибыль.



«Именно в телекоме начали переводить большой контур систем на Hadoop. Большие массивы данных из биллинг-систем, CRM и других источников складываются в Hadoop, агрегируются и уже над этой информацией надстраивают BI, позволяющий понимать, где в данный момент находится абонент и каковы его потребности, чтобы предложить лучший сервис, сделать для каждого конкретного клиента наиболее привлекательной предложение, - рассказывает **Андрей Байбутов, директор по развитию бизнеса департамента BI ГК «Корус Консалтинг».**

«Ритейл тоже находится в числе первопроходцев рынка Big Data. Все больше компаний из этого сегмента создают отдельные подразделения по работе с данными, чтобы как можно глубже погружаться в строки чеков за 2+ лет и находить новые скрытые взаимосвязи», **добавляет Байбутов.**

Иван Вахмянин, генеральный директор компании Visiology, полагает, что первая шумиха уже прошла, и сейчас происходит накопление реального опыта, как у заказчиков, так и у исполнителей.

«Мы ожидаем в ближайшее время появления интересных кейсов, причем не только в «традиционных» для Big Data отраслях, таких как финансы, телеком и ритейл, но и в промышленности, логистике, строительстве и здравоохранении,» - добавляет Вахмянин.

Павел Адылин, исполнительный директор компании Artezio (группа компаний ЛАНИТ), считает, что потенциальным заказчиком Big Data проектов в России в ближайшее время должны выступить также и компании госсектора, т.к. они обладают огромными накопленными объемами данных, пригодными для анализа.

Константин Черноусов, заместитель генерального директора Vesolv, приводит пример реализованного проекта в госсекторе: «Например, Федеральная Налоговая Служба завершила первый проект с использованием Big Data по отслеживанию цепочки плательщиков НДС и пресечению мошенничеств по уводу НДС».

Что касается решений, предлагаемых разработчиками, то это либо международные коммерческие продукты от Oracle, SAP и подобных, либо решения на базе open source технологий. Отечественного софта для обработки большого объема данных практически нет, добавляет Черноусов.

Андрей Нугманов, партнер AT Consulting, директор блока BI, считает, что в секторе «нового BI» – анализа больших данных, процессинга событий и принятия решений в реальном времени – стек СПО активно теснит продукты традиционных вендоров. Оно развивается в свете обновленного видения функциональных требований к BI и технологически во многом догнало проприетарный стек.

«Открытый код, транспарентность развития, правовая чистота и доступность, гарантированная и не замкнутая на одного вендора поддержка, толерантность к оборудованию, высочайшая популярность СПО, в первую очередь среди молодых и перспективных специалистов, – все это становится причинами активного вытеснения и вымывания «старого» проприетарного стека из традиционных ниш,» - уверен Нугманов.

Вендоры пытаются если не оседлать волну, то хотя бы не быть погребенными ею. Кто-то открывает код и переходит на СПО-модель бизнеса, пытаясь возродить интерес у публики, а значит и у лидеров мнений в среде

покупателей, к своим традиционным продуктам. Другие активно интегрируются с крупными поставщиками сервисов по поддержке стека Hadoop, стараясь снизить стоимость владения своими традиционными продуктами за счет использования открытых возможностей Big Data и достичь синергетических эффектов гибридного решения.

«Клиент не всегда готов сразу платить за лицензии вендору и пытается самостоятельно протестировать технологию, понять степень ее применимости и набрать необходимую экспертизу для дальнейшей эксплуатации. Выбор СПО позволяет обеспечить быстрое внедрение интересующей функциональности без лицензионных платежей и – благодаря отсутствию закупочных процедур – в минимальные сроки. Каких-то серьезных препятствий в развитии данных технологий у клиентов мы не видим. И экспертиза присутствует на рынке, и как минимум есть внятный бизнес-кейс, обеспечивающий сокращение операционных затрат на хранение значительных объемов информации,» - говорит Нугманов

С точки зрения технологий, в AT Consulting наблюдают, что на первый план выходят решения использующие In-Memory Data Grid (IMDG).

«Hadoop позволяет собрать разнородную информацию и хранить. Теперь пришло время следующего шага – проводить сложные аналитические расчеты в онлайн-режиме. Классические MPP-платформы уже не могут обеспечить быстрой реакции по причине наличия операций чтения и записи на диски и специфики операционной среды. Также важен и вопрос стоимости таких технологий,»-- рассказывает партнер AT Consulting. - «Мы видим, что для серьезных аналитических задач все чаще применяются in-memory решения. Они обеспечивают возможность высокопроизводительного параллельного выполнения запросов на сильно

нагруженных аналитических системах для обслуживания тысяч пользователей в режиме высокой доступности.»-

Роман Баранов, руководитель направления бизнес-аналитики и хранилищ данных компании «Крок», отмечает важность понимания того, что сам термин Big Data с каждым годом становится все более размытым. Перечень технологий, которые можно отнести к этому понятию, становится все больше. Они уже являются обыденной реальностью большинства современных компаний. Кроме того, произнося «большие данные» многие уже давно имеют в виду не только сбор и хранение данных, но и аналитику, и интернет вещей, и многое другое.

Тренды российского и мирового рынка Big Data

Основная тенденция российского рынка Big Data — проникновение технологий больших данных в те области, в которых раньше их было сложно представить.

Если ранее огромное число сегментов, например, производство, не так активно уделяло внимание технологиям работы с большими данными, то сейчас возможность собрать информацию со всех датчиков и другого оборудования завода дает гигантские возможности.

*«Это позволит существенно оптимизировать работу на самом производстве, а также повысить эффективность планирования и конвертировать полученную информацию в деньги, которые теряются при отклонении от плана или не дозарабатываются с точки зрения утерянной прибыли,» - говорит **Андрей Байбутов, директор по развитию бизнеса***

департамента BI ГК «Корус Консалтинг».

«Если говорить о мировых трендах, то в первую очередь можно говорить о тренде переноса инфраструктуры Big Data в облако,» - считает Иван Вахмянин, генеральный директор компании Visiology. По мнению Константина Черноусова, заместителя генерального директора Vesolv, общий тренд заключается в том, что все хотят использовать Big Data, поскольку анализ больших данных повышает эффективность и конкурентоспособность компании. И одним из движимых фактов является, как ни странно, появляющиеся опасения от того, что конкурент начал извлекать выгоду, используя новую технологию.

«Это имеет смысл для многих компаний, так как серверные мощности для Big Data стоят очень дорого, а нужны не всегда на постоянной основе. Например, мы в Visiology большинство своих Big Data экспериментов проводим в облаке Amazon. Кроме того, облачные Big Data продукты часто сильно облегчают работу инженеров – порог входа большинства Big Data программных продуктов в плане развертывания очень высокий, а в облаке можно сразу получить уже сконфигурированный кластер,» - рассказывает Вахмянин.

Второй тренд, по его словам, - это потоковая (streaming) аналитика, которая позволяет анализировать поступающие данные в реальном масштабе времени. Эта возможность особенно важна для приложений, построенных поверх данных, собираемых с датчиков (IoT, IIoT).

Павел Адылин из Artezio, добавляет, что для мирового рынка характерно разделение направления Big Data, которое у нас пока понимают в общем

виде, на множество самостоятельных направлений, решающих более узкие специфические задачи.

Например, по его данным, можно выделить: программно-аппаратные средства обеспечения хранения больших данных, средства параллельной обработки данных, средства фильтрации данных и построения моделей, средства визуализации данных и их взаимосвязи, средства работы с изображениями, машинное обучение, интеллектуальные интерфейсы, автоматизация умственного труда.

С таким разделением также связано появление готовых отраслевых решений для малого и среднего бизнеса, работающих как автономные приложения, так и по моделям SaaS или BDaS (Big Data as Service).

Барьеры российского рынка Big Data

■ Нехватка специалистов

Одна из главных проблем рынка Big Data в России - трудности с поиском квалифицированных специалистов.

По мнению Ивана Вахмянина из Visiology, дефицит таких кадров наблюдается не только потому, что они должны обладать довольно сложным набором навыков и компетенций, но и потому что сегодня мало кто понимает, как их готовить, оценивать и правильно организовывать их работу.

Константин Черноусов, заместитель генерального директора Vesolv, рассказывает, что сейчас постепенно входит в обиход такая профессия как

Data Scientist. Она довольно редкая, но спрос на нее уже колоссальный: на одно резюме такого специалиста приходится порядка 50 запросов на работу. *«В России таких специалистов, которые расскажут руководству о возможностях анализа с использованием Big Data, посчитают бюджет и реализуют проект, мало, и увеличить их количество быстро не получится, поскольку отсутствуют не просто курсы, а даже материалы на русском языке,»--* отмечает Черноусов.

Андрей Тиунов, генеральный директор компании «БиАй Партнер» (ГК «Ай-Теко»), уточняет, что Data Scientist - это эксперты со стороны компании-заказчика, которые понимают тенденции своего рынка, отлично знают бизнес, находят возможности для его роста и умеют использовать потенциал данных, которыми владеют, для решения тех или иных задач. Именно они обладают ключевыми компетенциями по решениям Big Data, говорит Тиунов.

Андрей Байбутов, директор по развитию бизнеса департамента BI ГК «Корус Консалтинг», также считает, что хороших специалистов в области Big Data на рынке критически мало.

«Если вы сейчас выйдете на рынок ресурсов в поиске хорошего специалиста с опытом работы с Big Data, machine learning, IoT и т.д., вряд ли сходу сможете найти человека с опытом от двух до пяти лет работы, да ещё и с необходимым продуктовым портфелем. Поэтому многие компании стараются возвращать собственных специалистов под данные задачи,»-- поясняет эксперт.

Любовь Ведешина, руководитель практики бизнес-аналитики компании «Интерпроком», видит проблему в том, что в России пока не сформировалось экспертное сообщество аналитиков в области больших данных, не появились

компетентный заказчик и компетентный исполнитель.

«На стороне потенциальных заказчиков мы видим нехватку специалистов, которые одинаково хорошо разбирались бы как в отраслевой специфике, так и в подходах, инструментах и методах обработки больших данных. На стороне исполнительней уже появились эксперты в области больших данных мирового уровня, некоторые даже находятся в мировом топе. Но их единицы,»-- подчеркивает Ведешина.

Профессия аналитика в области больших данных пока еще не стала массовой. В вузах нет соответствующих программ подготовки, опять же потому что пока мало компетентных преподавателей. Корпорации отчасти компенсируют нехватку специалистов, предлагая собственные программы обучения. Например, ШАД (Школа анализа данных) от «Яндекса» и платные курсы в «Билайне».

«Однако этих курсов недостаточно, должно пройти еще какое-то время, чтобы число квалифицированных аналитиков больших данных изменило бы качество спроса и предложения на рынке Big Data,»-- полагает Любовь Ведешина.

Павел Адылин, исполнительный директор Artezio, добавляет, что в силу недостатка специалистов в области Big Data в России чаще всего не разделяются профессии Data Scientist, Data Analyst и Data Engineer. Если первый из них является создателем новых технологий извлечения информации из данных, алгоритмов машинного обучения, искусственного интеллекта, то последний является разработчиком комплексов программного или программно-аппаратного обеспечения для решения конкретных задач больших данных. Для подготовки этих различных

специалистов уже сейчас требуется внедрять различные методологические подходы, уверен Адылин.



Нехватка опыта внедрений

Основным сдерживающим фактором в развитии рынка Big Data в России ряд экспертов называют небольшое количество российских кейсов, на которые могли бы опираться как заказчики, так и интеграторы. В результате чего, Big Data проекты оказываются рискованными.

«Часто приходится слышать от клиентов – «гарантируйте нам, что внедрение Big Data аналитики принесет нам экономию N рублей», но таких гарантий дать невозможно, как минимум, до проведения эксплоративного анализа накопленных данных и построения и верификации первых моделей, что само по себе требует вложения ресурсов», - отмечает Иван Вахмянин, генеральный директор компании Visiology.

Об этом же говорит и Любовь Ведешина, руководитель практики бизнес-аналитики компании «Интерпроком». По её мнению, потенциальные потребители не понимают, какую выгоду для их компании и отрасли несут технологии применения больших данных. Заказчики сомневаются, что их инвестиции в технологии обработки и анализа больших данных окупятся.

Аналогичного мнения придерживается и Константин Черноусов из Vesolv. По его словам, сдерживает развитие российского рынка отсутствие знаний о возможных выгодах от использования Big Data.



Представитель «Корус Консалтинг» Андрей Байбутов также ссылается на опыт внедрений, как на барьер рынка:

«Я знаю только о единицах, максимум — паре десятков внедрений.

Большая часть проектов с большими данными делается зачастую на продуктах с открытым исходным кодом, опыта работы с которыми у российских специалистов тоже немного. В результате есть методологическая неготовность, что мешает понимать, как делать проекты, и технологическая — из-за отсутствия необходимых программных компетенций». «Западные кейсы мало на кого производят впечатление, потому что российские реалии довольно сильно отличаются. Поэтому на данном этапе рынок Big Data двигают компании, которые не боятся экспериментировать, вкладываться в исследовательские проекты, в расчете на те выгоды и конкурентные преимущества, которые Big Data может принести,» - добавляет Иван Вахмянин.



Проблемы качества данных

Важной проблемой для применения технологий Big Data в России является отсутствие практики накопления больших данных и низкое качество этих данных.

По мнению Любви Ведешиной из «Интерпроком», даже если потенциальный заказчик сформировал понимание своих выгод от анализа Big Data и нашел отраслевых экспертов-аналитиков в области больших данных, он сталкивается с проблемой качества и количества данных, которые у него накоплены. Как правило, данные, стихийно накопленные заказчиками,



находятся в состоянии, не пригодном для анализа и получения выгоды для компании, отмечает она.

Такую же проблему видит и Павел Адылин из Artezio. По его словам качество данных оставляет желать лучшего из-за наличия искажений (выбросов) и недостаточной глубины. Таким образом, требуется значительно расширять наборы данных для анализа, но для этого нет возможности, т.к. в связи с защитой персональных данных в нашей стране практически отсутствует рынок купли/продажи информации в виде бирж данных (Data Exchange).

«Возможно, накоплению данных могла бы помочь программа государственной поддержки открытых источников оцифрованных данных, например, доступ к первичным данным Росстата и т.д.,» – считает эксперт.

Материал опубликован на портале TAdviser, декабрь 2016

[Подробнее о системах бизнес-анализа в линейке «КОРУС Консалтинг»](#)