

Больше чем данные

текст: Алексей Комов

Технологии анализа данных больших объемов, имеющие звучное наименование Big Data, единогласно признаются аналитиками одним из ключевых драйверов развития ИТ-рынка. В России рынок решений класса Big Data все еще находится на стадии зарождения, однако к руководителям отечественного бизнеса постепенно приходит понимание, что цифровые данные — это не просто информация, а актив, имеющий ценность и способный генерировать прибыль.



Исследовательская компания IDC прогнозирует, что к 2020 году мировой рынок Big Data вырастет более чем в полтора раза — со \$130,1 млрд в нынешнем году до \$203 млрд. При этом основными потребителями решений для анализа больших данных, по оценке аналитиков, станут банки и страховые компании, промышленные производства, телекоммуникационный бизнес,

а также индустрия профессиональных услуг и госсектор.

Как и многие другие отраслевые аналитики, эксперты IDC считают Big Data одним из наи-

более активно развивающихся направлений в сфере информационных технологий, по темпам роста опережающим сам ИТ-рынок. Однако для России Big Data



Денис Афанасьев,
генеральный директор компании CleverDATA
(ГК «Ланит»):

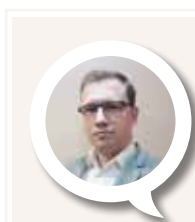
«Под термином BigData скрывается не какая-либо определенная технология, а концепция работы с данными и их применения в процессе принятия бизнес-решений».

остается новым и еще не распробованным явлением, в отношении к которому пока преобладает сдержанный скепсис.

А есть ли тренд?

Сам по себе термин Big Data, или большие данные, до сих пор трактуется многими неверно и вызывает множество споров и разногласий. Как правило, ИТ-компании подразумевают под этим термином технологии и решения для сбора, хранения, обработки и анализа большого объема информации, с которой затруднительно работать посредством традиционных подходов и решений. Однако некоторые представители бизнеса нередко пеняют на то, что под видом инновации, получившей название Big Data, им предлагаются решения и подходы к обработке данных, которые существуют и применяются уже очень давно.

По словам Мирослава Шестернина, менеджера по продуктам Orange Business Services, это распространенное заблуждение вызвано неразберихой в терминологии. Дело в том, что Big Data часто путают с решениями для бизнес-аналитики (Business Intelligence), чья цель — интерпретация большого количества данных с акцентом на ключевые факторы эффективности, а также моделирование исхода различных вариантов действий и рекомендаций, способствующих принятию управленческих решений. Отсюда и пошел миф, что никаких больших данных вообще не существует и их придумали маркетологи. «Big Data — это более



Мирослав Шестернин,
менеджер по продуктам компании Orange Business Services:

«Существующие модели Big Data позволяют оценивать актуальность данных по определенным дополнительным признакам, обнаруживать fraud и отсекаать махинации с информацией».



Андрей Орлов,
сертифицированный технический специалист по большим данным и управлению информацией компании IBM в России и СНГ:

«Управление данными и информацией — это гармоничное сочетание людей, процессов и технологий, позволяющее организации эффективно использовать данные в качестве актива предприятия».

углубленный научный подход к обработке и анализу данных, выявлению закономерностей путем машинного обучения и построения математических моделей», — поясняет Мирослав Шестернин. То есть под этим термином подразумеваются не сами данные, а методы их обработки, которые позволяют систематизировать огромные объемы структурированной и неструктурированной информации и представлять их в понятном виде, удобном для восприятия людей.

«Раньше термин Big Data использовался для описания любого объема данных и любых задач. Легко можно проследить по кривой Gartner движение данного тренда: он появился в 2011 году и пропал в 2015-м, точнее его заменил новый тренд — «машинное обучение». Люди поняли, что за Big Data скрываются именно подходы обработки и алгоритмы

исследования данных, которые описаны не одно десятилетие назад. Просто раньше технологии не позволяли решать подобные задачи, а теперь время изменилось и мы можем делать анализ на петабайтах информации», — комментирует Андрей Орлов, сертифицированный технический специалист по большим данным и управлению информацией компании IBM в России и СНГ.

По словам Максима Андрианова, руководителя отдела продаж управления бизнес-решений и заказной разработки компании Softline, предтечи Big Data в самом деле появились достаточно давно — ими являются так называемые нереляционные базы данных (NoSQL). Однако с тех пор объемы информации радикально увеличились: сейчас массивы данных растут как на дрожжах благодаря M2M-коммуникациям практически



Дмитрий Паршин,
директор центра разработки ПО компании Artezio
(ГК «Ланит»):

«На данный момент мало кто из бизнес-среды четко представляет, что такое Big Data, и тем более как эти технологии могут применяться в конкретной бизнес-области».



Татьяна Бочарникова,
глава представительства компании NetApp в России
и СНГ:

«Мы видим высокий интерес к большим данным, но проблема в том, что еще не все руководители понимают возможности различных технологий».

во всех отраслях — начиная от промышленности и заканчивая финансовой сферой. В частности, некоторые страховые компании предлагают автолюбителям снизить стоимость КАСКО, основываясь на манере вождения и на соблюдении водителем правил дорожного движения. Для этого в автомобиль устанавливается специальное устройство, фиксирующее скорость перемещения, количество перестроений, динамику разгонов и торможений, а также другие параметры, свидетельствующие, насколько аккуратен и дисциплинирован водитель и насколько целесообразно предоставить ему скидку.

«Конечно, во многом те методы и подходы, которые принято называть термином Big Data, являются реализацией давно существовавших математических абстракций. Революционна в данном случае именно их тех-

нологическая реализация, выражающаяся в применении давно известной математики на сверхбольших объемах данных», — дополняет Александр Смирнов, Hadoop-евангелист компании Teradata. К тому же, по его словам, есть множество задач, для которых не существовало технологических решений до появления Big Data. Классическим примером такого рода задач может служить индексированный полнотекстовый поиск.

По мнению Дениса Афанасьева, генерального директора компании CleverDATA (ГК «Ланит»), в последнее время технологии обработки информации сделали шаг вперед, благодаря чему бизнес получил возможность анализировать большие объемы данных за меньшее количество денег. Кроме того, объем информации, доступной для обработки, активно растет, чему во многом способствует все бо-

лее глубокое проникновение интернет-технологий в различные сферы жизни. Денис Афанасьев подчеркивает, что Big Data проходит стандартный жизненный цикл развития любой новой технологии. И если пару лет назад можно было говорить о том, что в этой области гораздо больше теории и маркетинговых обещаний, то теперь они постепенно уступают место практике и конкретным примерам применения.

Мало практики

Принято считать, что эффективное управление данными лежит в основе цифровой трансформации любого бизнеса и является первоочередным этапом этого процесса. По оценке компании IBM, многие представители российского бизнеса пытаются управлять своими данными — выстраивают соответствующие бизнес-процессы, внедряют специализированные ИТ-системы. Но на практике лишь единицы могут похвастаться правильно выстроенным подходом. «Российский рынок несколько отстает от мирового в части работы с данными. Особенно это касается представителей реального сектора экономики, в том числе добывающих компаний и предприятий топливно-энергетического комплекса. У них чаще всего отлично налажено цифровое управление предприятием, почти у всех внедрены ERP-системы, но при этом именно работе с данными и их анализу уделяется удручающе мало внимания», — сетует Александр Смирнов (Teradata).

Эксперты сходятся во мнении, что рынок решений Big Data

в России только зарождается и поэтому примеров их реального использования пока очень мало. Перед ИТ-сообществом стоит задача наглядно продемонстрировать заказчикам преимущества использования решений Big Data для извлечения максимальной пользы из данных для бизнеса. «По моему опыту, абсолютно все наши клиенты могут получить преимущества от применения технологий Big Data, но до реальных запросов дело пока доходит чуть более чем у половины заказчиков. Скорее всего, не все клиенты понимают, какие преимущества они могут получить, используя подобные решения, но мы над этим работаем», — отмечает Мирослав Шестернин (Orange Business Services).

Больше интереса

Несмотря на малое количество реальных кейсов, интерес к Big Data со стороны представителей российского бизнеса констатируют все ИТ-компании. Правда, по словам Дмитрия Паршина, директора центра разработки ПО компании Artezio (ГК «Ланит»), интерес этот в основном теоретический: отечественные компании ищут и анализируют примеры успешных зарубежных внедрений Big Data. «Аналогичные решения на практике реализуются очень медленно, что свидетельствует о смутном понимании российскими компаниями того, как работать с большими данными, а также о нежелании брать на себя инвестиционные риски. В такой ситуации нужно агрегировать лучший мировой опыт применения Big Data в раз-

личных областях и вырабатывать типовые решения для наиболее массовых бизнес-сегментов», — убежден Дмитрий Паршин.

Андрей Орлов (IBM) соглашается с тем, что за последние несколько лет интерес к Big Data существенно возрос, но теперь клиенты обращают внимание на то, как с помощью подобных решений извлечь выгоду. Очевидно, что для таких компаний хранение и управление данными являются ключевыми элементами ИТ-стратегий.

«В отраслях, традиционно ориентированных на работу с данными, таких как банкинг, телекоммуникации и ретейл, интерес к Big Data вылился в масштабные промышленные внедрения. Во многих других отраслях пока дальше пилотных проектов и архитектурных концепций дело не идет. При этом отдельные отрасли, в частности производство, добывающая промышленность и топливно-энергетический комплекс, только находятся на подступах к Big Data по при-

чине многолетней ориентации ИТ-стратегий на иные задачи. Также есть ряд отраслей, которые во всем мире являются едва ли не драйверами развития Big Data, а в нашей стране за счет хронического недофинансирования чаще всего даже и не думают начинать работы в этой области: речь идет обо всем, что связано с медицинской, биологией и клиническими исследованиями», — уточняет Александр Смирнов (Teradata).

Этап накопления

Прежде чем озаботиться обработкой и анализом больших данных, необходимо их собрать. По оценке Дмитрия Паршина (Artezio), российские компании пока находятся именно на стадии накопления данных. «Сбор качественных данных — важная часть комплексного решения Big Data. В ходе беседы с клиентом мы стараемся понять основную суть бизнес-задачи и порекомендовать ему возможные источники данных, которые могут пригодиться как на первых этапах проекта, так и в будущем. Часто бывает, что практически все данные находятся внутри контура клиента, и нужно правильно выстроить процесс интеграции данных, чтобы в дальнейшем облегчить работу исследователей данных», — делится опытом Андрей Орлов (IBM). По его словам, для интеграции данных из внутренних источников специалисты IBM обычно используют решение IBM InfoSphere Information Server, а в случае с внешними источниками неструктурированных и потоковых данных в качестве инструмента для аналитики мо-



Максим Андрианов,
руководитель отдела продаж
управления бизнес-решений
и заказной разработки компа-
нии Softline:

«Для большинства клиентов Big Data — это будущее».

жет применяться решение IBM Streams. Также в сборе данных заказчикам помогают разнообразные облачные сервисы, доступные на платформе Bluemix.net.

«Важно понимать, что Big Data позволяет использовать данные из самых разных источников. Например, мы помогаем клиентам собирать информацию, которая генерируется посредством M2M-технологий, мобильных приложений, контактных центров и всевозможных автоматизированных сервисов. К тому же при реализации проектов Big Data мы практикуем интеграцию с репозиториями заказчиков. Это могут быть базы данных о сделках, клиентах или производстве: как оператор связи, мы предоставляем специальный защищенный канал и помогаем обезличить данные, чтобы применять современные методы извлечения важной информации из максимального количества источников в полном соответствии с требованиями законодательства», — отмечает Мирослав Шестернин (Orange Business Services).

Лучше — больше

Без конкретной практической подоплеки информацию трудно оценить и превратить в актив компании, какими бы эффективными ни были решения для ее хранения, обработки и анализа. Однако Александр Смирнов (Teradata) рекомендует компаниям отталкиваться от выбранной бизнес-стратегии: если у заказчика есть понимание стоящих перед ним задач и того, какие данные необходимы для их решения,

то организация их сбора и анализа — это уже дело техники.

По словам Максима Андринова (Softline), с транзакционными компаниями все просто: у них имеется ряд метрик, влияющих на эффективность бизнеса. Чем больше данных, связанных с этими метриками, компании соберут и чем качественнее получатся корреляции с бизнес-результатом, тем больший позитивный эффект для их деятельности даст применение Big Data. В реальном секторе все намного сложнее, поскольку корреляции неочевидны. «Приходится прибегать к дата-майнингу, а это в большей степени эвристический, в чем-то даже шаманский метод. В одном банке с помощью анализа данных проводилось исследование возвратов кредита по кредитным картам. Выявилась корреляция с наличием домашнего животного в семье, что совершенно неочевидно. А другой банк выяснил, что на возврат кредитов влияет удаленность места жительства заемщика от офиса банка, в котором он брал кредит: есть определенное расстояние от офиса банка, после которого вероятность возврата кредита снижается в два раза», — рассказал Максим Андрианов.

Дмитрий Паршин (Artezio) убежден, что сбор данных во всех компаниях всегда можно трансформировать в сбор больших данных. Для этого нужно лишь снять ограничение на хранение только существенных или важных сведений. «Обычно компании не хотят тратить лишнее на хранение данных большого объема, поскольку не понимают, что может дать их обработка.

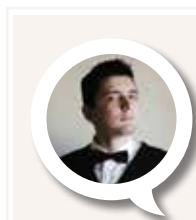
Технологии Big Data позволяют существенно снизить эти затраты и реализовать тотальный сбор данных с необходимой эффективностью. Особенно выгодно использовать гибридные облачные технологии, когда критичные данные остаются внутри предприятия, а существенные и гибкие ресурсы для обработки и визуализации развертываются в облаке», — поясняет Дмитрий Паршин. По его словам, аппаратные платформы для Big Data с самого начала были ориентированы не на дорогие высокопроизводительные кластеры, а на объединенные в сети частные компьютеры. Это также позволяет осуществлять приобщение к Big Data со сравнительно умеренным уровнем затрат. Но в любом случае в рамках каждого решения необходимо разрабатывать конкретную стратегию для требуемой предметной области, зависящую скорее и от нее самой, и от бизнес-процессов компании, нежели от объема информации.

В чем профит?

Что же может дать бизнесу хранение и анализ больших данных? По словам Дениса Афанасьева (CleverDATA), ценность данных проявляется прежде всего в их способности отвечать на вопросы бизнеса: кому из клиентов какой продукт или услугу лучше предложить, с кем и как лучше коммуницировать, в какое время и в каком месте это лучше делать, и так далее. «Конечно, лучше сохранять данные. Ведь сейчас мы не знаем, что нам понадобится спросить завтра, и именно поэтому стремимся собрать

и хранить все имеющиеся у компании сведения, а также идентифицировать их источники», — объясняет Денис Афанасьев.

Татьяна Бочарникова, глава представительства компании NetApp в России и СНГ, подчеркивает, что информация является главным активом бизнеса уже на протяжении довольно долгого времени. Но значимость цифровых данных резко возросла с распространением и популярностью интернет-доступа, появлением мобильных устройств, а также с тем, что уже давно знакомые человечеству бытовые приборы и записывающие устройства стали генерировать цифровые данные в огромных объемах. «Все это создает большие возможности для получения маркетинговой информации, повышения скорости и качества предоставления услуг, быстрой разработки новых продуктов, обеспечения безопасности и защиты информации. Но вместе с тем хранение и обработка данных требует довольно значительных затрат. Поэтому важно добиться, чтобы цифровые данные работали на бизнес и повышали его конкурентоспособность, а не являлись просто строкой в расходах», — резюмирует Татьяна Бочарникова. В качестве одного из примеров конструктивного подхода к управлению информацией Татьяна Бочарникова приводит сеть универмагов “John Lewis” в Великобритании, которая создала специальный департамент по работе с большими данными. Эти данные используются для подготовки стратегии развития бизнеса. В настоящее время “John Lewis” соби-



Александр Смирнов,
Hadoop-евангелист компании Teradata:

«В этом году у нас было несколько аналитических проектов для крупных ретейлеров, а также ряд архитектурных и аналитических проектов для банков и государственных организаций».



Андрей Тамбовский,
директор по технологиям «ФОРС Дистрибуция»:

Согласно данным IDC, среднегодовой рост мирового рынка технологий Big Data составляет 31,7% и, согласно прогнозам, к 2018 году достигнет \$41,5 млрд. Что касается российских реалий, то интерес к новым технологиям растет, однако на пути к практическому применению очень часто встает вопрос о целесообразности и получении конкретных результатов. При всех достоинствах систем Big Data получение преимуществ от их использования зависит прежде всего от двух факторов: умения сформулировать бизнес-задачу, или data science, и скорости внедрения решения.

рает информацию по продажам от 350 собственных магазинов каждого из 500 тыс. продуктовых наименований непосредственно в точках продаж. Эти сведения используются для того, чтобы улучшать обслуживание клиентов.

По словам Дениса Афанасьева (CleverDATA), компании, уже накопившие значительный объем статистической информации о клиентах, в частности операторы связи, получили возможность монетизировать эти данные, предоставляя другим компаниям всевозможные аналитические сервисы, основанные на подходах Big Data. Такие сервисы компенсируют отсутствие массива данных у большинства российских компаний. Например, на российском рынке

предлагаются сервисы на базе данных из социальных сетей и альтернативных источников. Распространены маркетинговые сервисы, всевозможные рекомендательные системы, сервисы клиентской аналитики и HR-scoring (цифровое ранжирование кандидатов), пользующиеся большим спросом. «Потребность в данных порождает интерес к классу систем типа “биржа данных”. Это совсем новый, только зарождающийся класс систем, который появляется благодаря тенденции к обогащению данных. С помощью таких систем взаимодействие между компаниями становится более легким, легитимным, прозрачным, регулируемым и безопасным», — подчеркивает Денис Афанасьев. ☒