

Big Data – как выпустить джинна из бутылки?

Big Data – let the genie out of the bottle



Александр Гольдштейн

Доцент кафедры ИКС СПбГУТ им проф. Бонч-Бруевича, директор НТЦ "Аргус", к.т.н.

Alexander Goldstein

Associate professor at SPbSUT, director of RTC ARGUS, PhD



Максим Скоринов

Преподаватель кафедры ИКС СПбГУТ, руководитель направления НТЦ "Аргус"

Maxim Skorinov

Lecturer at SPbSUT, product manager at RTC ARGUS



Михаил Феноменов

Аспирант кафедры ИКС СПбГУТ им проф. Бонч-Бруевича, технический директор НТЦ "Аргус"

Michail Fenomenov

Postgraduate at SPbSUT, technical director at RTC ARGUS

Концепция больших данных несет с собой не только большие потенциальные возможности, но и большие реальные трудности. Среди них и организация базовых принципов управления данными внутри компании, и изменение стратегических принципов постановки задач по анализу информации, и подбор сотрудников, которые лучше других будут соответствовать новым задачам. Как же избежать этих трудностей?

Conception of Big Data promises not only a lot of potential opportunities, but also a lot of real troubles. Among them there are organization of data manage-

Ключевые слова:

большие данные, обработка данных, подбор персонала, ошибочные представления

Keywords:

Big Data, data processing, recruitment, misconceptions

ment within the company, change of principles of setting targets for information analysis, careful selection of staff to meet new challenges. How can these difficulties be avoided?

Практически все IT-концепции и подходы, придуманные и воплощенные в жизнь за последние десятилетия, проходили и продолжают проходить общие стадии в своем развитии, которые очень хорошо иллюстрирует Gartner Hype Cycle: завышенные ожидания – разочарование – последовательное улучшение и выход на уровень реальной продуктивности. Конечно, не все идеи пережили вторую стадию, но те, кто смог пройти через переходный период, – по-прежнему с нами. По оценке IT-рынка от Gartner за 2014 г., концепция Big Data как раз перешагнула через пик "завышенных ожиданий" ("Наконец-то! Еще немного, и у нас будет магическая кнопка "Сделать все хорошо!", к тому же, она еще и нажиматься будет сама!") и передвинулась на вторую стадию. Почему так произошло и насколько радужными остаются перспективы? Давайте разберемся.

Отрицательная мифология

Наступление стадии разочарования в любой концепции практически неизбежно. Каждое новшество на пути своего развития обрастает различными мифами и легендами. Что же именно ожидалось первоначально от Big Data и почему эти ожидания не оправдываются?

Миф № 1. Big Data = анализ данных. Можно собрать в общее хранилище любые доступные нам данные, выполнить автоматически над ними какие-то магические пассы и заклинания – и на выходе получить готовый набор логически взаимосвязанных фактов, выводов и предложений по необходимым действиям.

К сожалению, Big Data – не философский камень, позволяющий превращать любой мусор в золото. Без изначальной постановки цели, исходного понимания, на какой именно вопрос требуется получить ответ, компания окажется в ситуации, что даже обладая "золотыми" данными, которые будут заложены в Big Data-инструменты, в результате обработки получатся "мусорные" знания, применить

которые окажется совершенно невозможно.

Миф № 2. Big Data = легкие деньги. Мы потратим гораздо меньше средств (например, по сравнению с внедрением технологий хранения и обработки данных, требующихся для классического Business Intelligence) на развертывание Big Data-решения, применим распространенные open-source-технологии и в результате получим очень быструю отдачу в виде новых продуктов и идей, которые нам подскажет выполненный анализ данных.

Эта идея тоже оказалась в определенной степени ошибочной. Несмотря на действительно более низкую удельную стоимость владения технологиями Big Data (особенно в области непосредственного хранения данных), в целом, практически половина компаний, решившихся на внедрение этих технологий, столкнулась со сложностями в самих принципах измерения уровня возврата инвестиций. Да, в качестве результата можно получить информацию о ранее неизвестных корреляциях в поведении клиентов, увидеть перспективную бизнес-идею, но эти результаты не приносят финансовой выгоды сами по себе, для их реализации и внедрения будут требоваться отдельные инвестиции. В результате бизнес приходит к пониманию, что на самом деле вложения в Big Data представляют собой, за редким исключением, очень длинные деньги, что отпугивает вчерашних адептов.

Миф № 3. Big Data – это Plug & Play. Самой концепции уже 8 лет, методы и технологии за прошедшее время уже устоялись, рынок предлагает значительный диапазон конкретных решений, мы всегда можем выбрать что-то, что нам точно подойдет, и без осложнений внедрить это у себя.

Технически – это действительно так. Но чтобы приносить пользу, инструмент должен не просто быть внедрен – он должен работать. А для работы будут требоваться данные. Для которых нужно описать модель, адаптировав ее под выбранное решение. И в среднем четверть компаний (согласно отчетам IDC), использующих у себя Big Data-решения, сталкиваются со сложностью разработки такой модели, а также с трудностями принципиального определения – какие именно данные требуются фиксировать и хранить, а какие можно игнорировать. А кроме самих данных инструменту нужны пользователи, которыми должны выступать аналитики. По данным того же отчета IDC, в котором перечис-

ляются причины, тормозящие повсеместное проникновение Big Data-концепции, – примерно 10% компаний испытывают проблемы с подбором соответствующего персонала. А еще 10% компаний подтверждают, что испытывают трудности в том, чтобы из всего многообразия, существующего на рынке, выбрать оптимальную технологию, наиболее точно подходящую для решения интересующих задач.

Миф № 4. Big Data – тот же BI, только больше. У нас уже есть продукт класса BI, а поскольку Big Data-решение – это тоже аналитический продукт, значит, у нас не будет никаких сложностей, чтобы его применить с пользой.

Наверное, это самый устойчивый миф. Есть даже поговорка: "BI-продукт есть уже почти у всех, а поскольку разработчикам ПО не хочется терять прибыли, было придумано новое название, чтобы то же самое продать второй раз". Однако в действительности цели и задачи BI и Big Data во многом сильно отличаются. BI – в первую очередь система, направленная на поддержку принятия решений, работающая в четком очерченных, задаваемых пользователем рамках, работающая на ретроспективных данных и позволяющая найти четкие доказательства, "почему было именно так". Big Data-концепция нацелена на поиск и предложения новых идей, открытия потенциальных продуктов, основанных на информации, автоматизируя области, которые оказываются чрезмерно ресурсоемкими для проведения анализа классическими способами, и основанная в своей идеологии фактически на "живых" онлайн-данных. В результате следования последнему мифу компания может оказаться просто не готовой к получению выгоды от использования Big Data.

Слагаемые успеха

Стратегия

Как ни банально это прозвучит, но для того, чтобы отправиться в успешное плавание по океану больших данных, необходимо четко представлять конечную точку маршрута, ту цель, достижение которой для компании оправдывает применение Big Data-концепции. Распыляя усилия сразу на все доступные области, собирая и пуская в ход все виды данных, которыми располагает или до которых может дотянуться, обрабатывая внешние источники, компания, тоже можно натолкнуться на неожиданное озарение, кото-

рое принесет прибыль, но это будет случайным исключением. Технологии обработки больших данных уже сейчас дают большой толчок совершенно различным сферам бизнеса, таким как:

- медицина (удаленный и персональный мониторинг, соответствие назначенных методов лечения и фактических результатов и т.д.);
- автомобильная промышленность (сбор и анализ информации с сотен датчиков и электронных систем, устанавливаемых на современных автомобилях, взаимодействие с государственными службами, с распределенными сетями для обеспечения дополнительной безопасности водителей и т.п.);
- телекоммуникации (анализ причин и перераспределение нагрузки в сетях передачи данных, оценка и контроль лояльности абонентов, формирование таргетированных предложений, основанных на определении интересов конкретного пользователя, и т.д.) и многих, многих других.

Разумеется, нельзя забывать о том, что внедрение технологий больших данных повлечет изменения и внутри отдельно взятой организации, перестройку ее сложившихся бизнес-процессов:

- для подразделений маркетинга (расширение базы для анализа, появление ранее недоступной информации, которая может помочь объяснить причины изменений в настроениях клиентов и предложить новые способы воздействия на них);
- для продавцов (использование данных со смартфонов и уже упомянутых выше датчиков с автомобилями, которые помогут определить эффективность конкретного продавца);
- для складских и логистических подразделений (комбинирование результатов GPS- и RFID-мониторинга, позволяющее более точно планировать перемещение оборудования, распределение его по складам и точкам продаж);
- отделов по работе с персоналом (определение лояльности сотрудников на основе анализа данных, получаемых из социальных сетей) и так далее.

И, конечно, одним из возможных результатов внедрения Big Data-концепции может стать открытие совершенно новых рынков, ранее недоступных для отрасли. Например, телекоммуникационные компании, которые буквально "сидят" на огромных объемах различных данных, генерируемых

самими абонентами, могут использовать получаемую в BSS/OSS-комплексе информацию не только с целью оптимизации внутренних бизнес-процессов, но и для формирования внешних предложений, превращаясь в "фабрику данных". Информация о типах звонков (локальные/международные), о местонахождении (в пределах страны/в роуминге), о частоте и качестве оплаты услуг формирует вполне законченный портрет абонента, который будет интересен не только самому телеком-оператору, но и внешним компаниям, также взаимодействующим с этим абонентом. И надо отметить, что в текущем году два мобильных оператора из большой тройки реализовали пилотные проекты по сотрудничеству с рядом банков, предлагая им обезличенный скоринговый балл по запрашиваемым абонентам, который в дальнейшем используется при оценке платежеспособности в процессе одобрения кредитов.

Таким образом, перспективы применения Big Data-концепции по-прежнему остаются поистине безграничными. Важно лишь выбрать наиболее выгодное направление для приложения усилий.

IT-ландшафт

Следующим по важности после определения направления применения Big Data-технологий является модернизация IT-ландшафта компании. Концепция больших данных для эффективного своего использования требует наличия сильной IT-инфраструктуры. И здесь речь идет не о том, чтобы компания просто располагала мощными серверами и сверхпрофессиональными сотрудниками. Мы говорим о том, что успешное внедрение Big Data зависит от качественной архитектуры IT. Ни для кого не секрет, что в крупных компаниях, например, телекоммуникационного сектора, заня-

тых в последние годы регулярными процессами слияния и поглощения, накопилось значительное количество дублирующих друг друга информационных систем, а связь между различными сегментами информационного поля может просто отсутствовать в силу разобщенности различных бизнес-направлений. Скажем, абонент, потребляющий услуги и мобильной, и фиксированной связи от одного провайдера услуг, может быть заведен одновременно в несколько различных учетных систем. В этой ситуации разработка непротиворечивой модели данных, которая будет в дальнейшем использоваться и в процессе обработки больших данных, может оказаться не просто сложной, а практически невыполнимой задачей.

Поэтому для того, чтобы Big Data-решения действительно показали свою эффективность, они должны быть размещены в прозрачном информационном поле. Соответственно, выравнивание IT-структуры, создание корпоративной шины данных, выделенных справочных классификаторов, настройка и оптимизация сквозных бизнес-процессов – все эти сопутствующие проекты должны быть предусмотрены и реализованы для того, чтобы работа с большими данными в действительности принесла пользу. Конечно, это увеличивает трудоемкость перехода на эти технологии, но в любом случае здесь есть как минимум одна хорошая новость – даже если впоследствии компания в силу каких-либо причин разочаруется в идеологии Big Data, порядок в IT-архитектуре уже будет наведен, и положительный эффект от этого будет распространяться на все области деятельности организации.

Люди

И последним, но от этого не менее важным слагаемым возможного успеха являются люди. Причем как те, кто

должен будет провести подготовительную работу, описанную в предыдущем разделе, организовав и успешно завершив проекты по трансформации IT-ландшафта, так и те, кто будет заниматься непосредственно работой с Big Data-инструментами, анализируя данные в поисках новых открытий. И если, возвращаясь к отчету IDC, компании, внедряющие у себя Big Data-решения, в трети случаев сетуют на то, что крайне сложно найти сотрудников с хорошими навыками управления проектами, то со сложностями в поиске людей, которые будут исполнять роль аналитиков данных, сталкивается уже 2/3 из опрошенных.

Так кто же сможет стать хорошим аналитиком данных? Какими навыками должен обладать такой сотрудник?

Программирование – умение писать код и понимание архитектуры стека Big Data-технологий. По-прежнему довольно редко можно найти источники данных, информацию из которых можно напрямую передавать на анализ. Нужно выполнить значительный объем подготовительных работ для того, чтобы полученные данные принесли пользу.

Научная деятельность – умение формировать решения на основе доказательств, стремление открывать что-то новое. Без таких навыков сотрудник, занятый обработкой результатов, полученных в ходе применения Big Data-инструментов, может просто "пройти мимо" чего-то действительно ценного.

Опыт ведения переговоров – навык убеждения оппонентов и взаимодействия с окружающими. Чаще всего недостаточно просто найти какую-то свежую идею, самое сложное – убедить остальных в том, что эта идея стоящая и действительно способна принести пользу компании.

Аналитика – опыт статистического и визуального анализа, знания в сфере машинного обучения и анализе

неструктурированных данных (текст, изображения, видео и т.д.). Инструменты, предлагаемые для работы с большими данными, не являются общепотребительными, необходимы практически навыки использования таких же или схожих методов анализа.

Экспертиза в бизнесе – знание, как работает бизнес, на котором специализируется компания, понимание того, как компания обеспечивает себе прибыль. Подобный профильный опыт позволит сосредоточиться на анализе информации, которая сможет оказать максимальное влияние на области, наиболее специфичные и важные для конкретного бизнеса, тем самым увеличив отдачу от самого проекта использования больших данных.

Разумеется, вероятность обнаружить аналитика, который полностью удовлетворял бы всем перечисленным характеристикам, внутри собственной компании или на открытом рынке бесконечно мала. К тому же, в настоящий момент по-прежнему отсутствуют выделенные направления в высших учебных заведениях, которые выпускали бы всесторонне подготовленных специалистов в области анализа данных. Тем не менее, в ряде университетов начали уделять этой области более

пристальное внимание и вводят в образовательную программу специфические дисциплины. И, конечно, в качестве стандартной практики многие ведущие разработчики ПО, предлагающие собственные продукты для работы с большими данными (такие как IBM, Oracle, Google), организуют отдельные курсы, позволяющие слушателям изучить особенности как общей методологии Big Data, так и особенности применения конкретных технологий.

Поэтому в настоящий момент наиболее правильным сценарием будет подбирать команды единомышленников, где все участники смогут дополнять друг друга различными навыками. И здесь самым важным для достижения успеха является правильное управление. Когда менеджмент компании не просто заинтересован в достижении быстрых результатов, но полностью понимает всю глубину и проблематику Big Data-концепции и готов регулярно и своевременно оказывать помощь в решении различных возникающих организационных задач.

Интерес сохраняется

И все же, несмотря на все разочарования и встреченные трудности,

интерес к теме Big Data и ощущение ее потенциальной выгоды для бизнеса сохраняется. И пусть на данный момент рынок Big Data-проектов в России еще совсем невелик и в основном находится на стадии пилотов, в количественном и денежном выражении этот рынок растет, и мы думаем, что темпы роста в перспективе будут только увеличиваться. Потому что данных меньше не становится. ■

Литература

1. Davenport, Thomas H. Big Data at work: dispelling the myths, uncovering the opportunities. – Harvard Business School Publishing Corporation, Boston, Massachusetts, 2014. – 230 p.
2. Майер-Шенбергер В. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим / Виктор Майер-Шенбергер, Кеннет Кукьер; пер. с англ. Инны Гайдюк. – М.: Манн, Иванов и Фербер, 2014. С. 240.
3. Rich R. Feel the heat: Big Data lifts off [Electronic Resource] / TMForum. – Electronic Data. – Morristown, NJ, 2015. – 36 p. [online] Доступ через: <https://www.tmforum.org/resources/research-and-analysis/feel-the-heat-big-data-lifts-off/>.

Ваше мнение и вопросы по статье
присылайте по адресу
tss@Groteck.ru