

05 декабря 2017

Большие данные: панацея или PR?

В современном мире место технологий в бизнесе огромно. Не важно какие задачи вы перед собой ставите: повысить оборот, маржинальность, сократить затраты, привлечь новых клиентов — ответ на вопрос: «Как это сделать?» всегда подразумевает использование более совершенных технологий и инструментов при реализации ваших бизнес-идей.

Статья Константина Добротока, руководителя направления систем оптимизаций «КОРУС Консалтинг».

«Большие данные» в последнее время находятся на вершине технологической революции и их использование при принятии управленческих решений становится для компаний жизненной необходимостью. Итак, в чем суть этого нового явления и как нужно действовать, чтобы научиться извлекать выгоду из самого важного актива вашей компании — ваших данных.

Аналитики Gartner, Inc, в 2012 году дали следующее определение Большим данным: «Большие данные — это информационный актив, обладающий большим объемом, большой скоростью изменений и большим разнообразием, требующий экономически-эффективных и инновационных подходов для получения лучшего понимания и лучшего принятия решений». За прошедшие годы самые передовые международные компании проделали



большой путь для того, чтобы освоиться с новой реальностью и подходами. Уже в 2016 году опросы NewVantage Partners показали, что 62,5% топ-менеджеров компаний Fortune 1000 считают Большие данные основным направлением технологического развития.

Инвестиции в эту область растут от года к году, появилась и получает активное распространение новая должность - «директор по данным». Решения руководства всё больше основываются на данных, очень быстро растёт разнообразие источников информации для анализа, ученые по данным — самые востребованные специалисты на рынке, и за ними ведется настоящая охота...

Сейчас середина 2017 года. Оглянитесь вокруг. Видите ли вы в России большое количество компаний, которые совершили технологический прорыв за счет Big Data и уже начали опережать конкурентов? К сожалению, таких практически нет.

Давайте разберемся, почему. Попробуем понять, что мы делаем не так и что необходимо сделать, чтобы быстрее поймать волну и получить один из самых мощных в мире инструментов конкурентной борьбы. Самое важное, что нужно — это хорошо понимать само определение Больших данных. И тогда все встаёт на свои места, появляются новые мысли и идеи для реализации. Итак, определение состоит из двух частей и двадцати семи слов. И каждое слово здесь требует внимания. Первая часть описывает информационный актив.

Информационный актив



Перечислите все ваши корпоративные информационные системы — это и будет основа вашего информационного актива, ваши внутренние данные. Далее, возьмите данные о конкурентах, информацию из социальных сетей, геолокационные данные, результаты социологических опросов, метеосводки, новости из новостных лент и т.д. – и ваш актив будет обогащен внешними данными. И чем больше это обогащение, тем лучше.

Кажется, все просто и понятно. Но это не совсем так. При первой же попытке осознания своего информационного актива, перед нами, как грибы, вырастают большие проблемы, преодоление которых —отдельная большая задача. И решить ее мало кому удаётся.

В чем причина?

Проблема 1: Отсутствие культуры. Речь про корпоративную культуру бережного обращения данными (data-driven culture). Это краеугольный камень и источник большинства проблем. Задайтесь вопросом, насколько аккуратно и методично вы или ваши сотрудники вносят информацию в корпоративные системы, выполняя свои повседневные задачи? И ответ скорее всего вас сильно расстроит.

Без такой культуры бессмысленно планировать какие-либо инновации. Нужно, чтобы каждый сотрудник понимал: если он не корректно ввел одно (!) слово в одно поле в ERP-системе, то компания через какое-то время начнет терять прибыль и придет к катастрофе. Это чистая правда. За последнее время количество данных на планете растет в геометрической прогрессии и бережное отношение к данным — это уже не роскошь, а жизненная

необходимость. Кто не научится, уйдет с рынка. Вариантов других нет.

Вы должны начать процесс корректного ввода данных уже сейчас. Неважно, как именно он будет реализован, важно, чтобы он был инициирован со стороны руководства, как один из самых приоритетных для компании.

Допустим, вам удалось убедить всех, что корректность данных — это важно.

Все сотрудники аккуратно вносят данные, а тот, кто этого не делает – висит на стене позора или уволен. Далее встает вопрос: а что делать с историческими данными? Что делать с информацией, которой уже несколько лет и которая также нужна для принятия решений?

Ответ простой: ничего. Самой большой глупостью будет затея начать глобальную переделку текущих информационных систем, с тем, чтобы восполнить пробелы в исторических данных. Пока вы будете этим заниматься, вас обгонят и в два счета вытолкнут из конкурентной борьбы другие компании. Мы позже в этой статье расскажем о способе преодоления данной проблемы. Теперь посмотрим на внешние данные: все что мы можем получить из открытых источников.

Проблема 2: Нет понимания, что дает эффект. Эту проблему можно сформулировать также в виде вопроса: как понять, какие данные действительно важны, а какие нет? Вопрос этот сложный, и ответ на него далеко не очевиден.

Приведем простой пример: учет погоды в такой задаче в области Big Data, как прогнозирование спроса. На первый взгляд, все просто: будет жарко, значит, покупатель будет хотеть пить, значит, продажи воды вырастут. Прекрасно. Но что мы имеем на практике? Если вы жестко привяжете умный алгоритм

предсказания спроса на воду к прогнозу погоды, особенно на более длинных горизонтах планирования, то ошибки Гидрометцентра — тут же станут вашими собственными и будут напрямую влиять на ваши процессы планирования. Стоит хорошо подумать над целесообразностью такого союза с Гидрометцентром. Убытки он вам возвращать точно не будет.

Итак, перед какими-либо действиями правильно спросить себя: насколько действительно важен для бизнеса тот или иной внешний фактор? Отвечать на него нужно вдумчиво и взвешенно. И только после этого принимать решение об использовании новых внешних данных в своих расчетах.

Из своей практики мы видим, что большинство, на первый взгляд, очевидных и хорошо разрекламированных инноваций на самом деле имеют минимальный эффект и совершенно не стоят вложений. Здесь все, как в настоящей науке: все нужно проверять. Экспериментально доказывать перед промышленным применением. Как правильно построить этот процесс?

Тут мы плавно переходим ко второй части определения Gartner Больших данных: информационный актив требует экономически-эффективных и инновационных подходов для получения лучшего понимания и лучшего принятия решений. С пониманием этой части определения во всем мире очень плохо. Часто про эту часть вообще забывают! А ведь в ней содержится вся суть.

Лучшее принятие решений

Итак, финальная цель Больших данных одна: лучшее принятие решений руководством. Что значит лучшее? В современном мире лучше значит

быстрее и точнее предугадать ситуацию, которая быстро меняется. Кто не успел, тот проиграл, и хорошо, если при этом выжил.

Здесь очень важно понимать одну простую вещь: какие бы современные инновации ни внедрялись, в том числе и Большие данные, финальное принятие бизнес-решений всегда остается за руководством. И именно эти решения приносят результат. Ни один самый изощренный алгоритм никогда не даст правильных решений. Все, что могут сделать инновации, помочь вам. Дать инструмент. Хороший, лучший в мире, мощный инструмент, позволяющий человеку выйти на качественно иной уровень принятия решений. Не верьте мошенникам, которые будут говорить вам что-либо иное. Чудес не бывает.

Итак, суть Больших данных — дать вам возможность лучшего понимания и возможность лучшего принятия решений за счет применения экономически-эффективных, инновационных подходов при анализе вашего информационного актива.

На примере это выглядит вот так. Большие данные говорят: «Если покупатель X сегодня купил помидоры, то с вероятностью 78% через два дня он купит творог». И всё. Большого ждать от Больших данных нельзя. Далее руководитель должен внимательно отнестись к этому лучшему пониманию и применить эти знания на практике, приняв лучшее решение: персональная промо-акция покупателю через два дня: «приходи сегодня, и для тебя в этот день творог будет на 10% дешевле».

Бернард Марр из Forbes пишет: «Если Вы планируете проект Big Data, вы всегда упускаете один самый важный аспект. Это не должен быть проект в

области данных, это должен быть проект в области бизнеса. Если вы не знаете, на какие бизнес-вопросы вам нужен ответ и какая бизнес-проблема должна быть решена, вы не должны начинать проект».

Дэниэл Ньюмен также из Forbes указывает: «Большие данные — это не то, что вы знаете. Это то, что вы делаете с тем, что узнали. Это то, что все компании во всем мире должны усвоить». Очевидно, что если грамотно и своевременно применять новые знания, можно достичь невиданных ранее результатов в планировании промо-акций, ассортиментном планировании, ценообразовании, и др. И результаты эти помогут выйти на уровень, о котором ранее и мечтать не приходилось.

Но, как это обычно бывает, и здесь есть большие проблемы, которые нужно преодолевать. Если бы все было так просто, то в бизнесе шла бы революция за революцией. А этого не происходит. И вот почему.

Проблема 3: получить новое полезное знание из Больших данных — очень сложная задача с непредсказуемым результатом. Это не очень приятный факт. Это задачи для избранных людей, обладающих особым талантом и опытом. Еще более неприятным фактом является то, что таких людей практически нет. Называются они учеными по данным (Data Scientist). Есть известная характеристика Harvard Business Review: «Ученый по данным — самая сексуальная профессия XXI века». И ажиотаж в последнее время во всем мире вокруг этой профессии подтверждает эту характеристику.

Мы даём следующее определение: современный ученый по данным — это творческая профессия, требующая особого таланта искать и извлекать новые полезные знания из исторических и текущих данных компании. Причем

применение этих знаний руководством должно легко трансформироваться в конкретный бизнес-результат.

Сложности начинаются с самого начала: никто, включая самого ученого по данным, на старте не знает, где скрыто это новое знание, какие алгоритмы применить, какие данные нужно исследовать, как вообще подойти к корректному преобразованию бизнес-задачи в математическую и обратно и, самое важное, нужно ли это новое знание бизнесу. Также никто не застрахован от ситуации, когда у задачи может не быть решения в принципе.

Очень мало людей сейчас могут выполнять такую работу. Приведу пример из практики. На собеседованиях мы даем тест из семи практических задач на дом и одну неделю на решение. Причем у трех задач есть более-менее классическое решение, а вот у четырех задач готовых решений нет, и они не тривиальны. Бывает так, что задачи эти на текущий момент не до конца решены даже нашими собственными экспертами. Так вот: 90% кандидатов спустя неделю приходят с решенными первыми тремя задачами, по которым есть четкое понимание, что делать и статьи в интернете, описывающие схожую проблематику. И никто из этих 90% даже не делает попытки взять штурмом одну из четырех сложных задач. Вдумайтесь, 90% кандидатов — аналитиков и математиков с различным опытом (в том числе, и в бизнесе), успешно прошедших первичный отсев и допущенных к решению практических заданий, не имеют интереса к решению нетривиальных и непонятных задач. Они просто не подходят для этой профессии.

Итак, настоящий ученый по данным должен разбираться как в математике и программировании, так и в бизнес-области, уметь сам себе ставить задачи,

проводить математические исследования на Больших данных и уметь трансформировать найденные решения в такой вид, чтобы руководство компании смогло быстро эти знания применить и получить бизнес-эффект.

Причем дать ученому по данным исторические данные компании, в которых не совсем корректно велась информация — это нормальная задача. Да, он не будет от нее безмерно счастлив. Но его работа заключается в том, чтобы попытаться найти новые знания в любых данных. Единственный, но существенный минус — это то, что большого количества качественных открытий на плохих данных вы вряд ли увидите. Но какие-то результаты получить в любом случае можно и нужно. Допустим, вы нашли такого человека или обратились в специализированную компанию. И вот, наступил счастливый момент, когда эксперт садится работать. И тут же возникает еще одна проблема. В нашем списке она последняя.

4 проблема: без специальных ИТ-инструментов работать с Большими данными невозможно. И это проблема. Представьте себе гигабайты и терабайты внутренних и внешних данных, которые нужно исследовать. И делать это нужно быстро. Если эксперт будет тратить дни и недели на выполнение сортировок, преобразования форматов данных и т.д., то он выдаст вам результат через год, так как львиную долю времени будет тратить на ожидание выполнения тех или иных операций на сервере. А это неприемлемо. У этой проблемы есть решение: использовать специализированные инструменты. Это самые современные системы параллельных вычислений, без появления которых никакой области Больших данных не было бы в принципе.

McKinsey описывает это так: «У ученых по данным сейчас есть в распоряжении беспрецедентные компьютерные мощности и они способны придумать и применять алгоритмы колоссальной сложности». Но у таких систем есть один недостаток — они невероятно требовательны к ресурсам, что сразу же сказывается на себестоимости их использования. Представьте, что вам нужно закупить огромный дорогой сервер только для того, чтобы ученый по данным что-то разово посчитал. Что именно, когда и в каком объеме он будет считать в будущем никто не знает. Может быть сервера не хватит и нужен будет новый, а, может, мощности будут избыточны. А теперь представьте, что работать будут несколько специалистов одновременно.

А мы прекрасно помним из определения Gartner, что подходы к обработке Больших данных обязательно должны быть экономически-эффективными. И здесь нам на помощь приходит, пожалуй, один из главных участников современной технологической революции - облачные технологии.

Облачные технологии

Использовать для тяжелых вычислений облачные мощности — это, пожалуй, единственный современный и экономически эффективный способ работы с Большими данными. Причем речь не об использовании арендованного ЦОДа. Речь идет о полноценных инфраструктурных облачных платформах, которые позволяют выполнять параллельные вычисления на Больших данных, минимально затрачивая усилия на технические вопросы. То есть все должно быть максимально подготовлено для ученого по данным. Он должен иметь озеро данных (Data Lake), в которое по нажатию кнопки он заливает данные компании. И он должен иметь возможность применения языков



обработки данных (наиболее популярны сейчас R и Python), не думая об объеме этих данных. Облачная платформа должна предоставлять все эти возможности.

Ниже представлены данные Synergy Research Group за второй квартал 2017 года. По сути, мы имеем трех игроков на мировом рынке, которые в состоянии предоставить соответствующие инструменты Amazon, Microsoft, IBM.

Причем обратите внимание на рост доли Microsoft в облачном сегменте. Рост очень большой, и компания не собирается останавливаться. А Amazon был и остается самой технологичной компанией в мире, с большим отрывом от всех остальных. И у обеих компаний есть все, что необходимо для развития облачного направления.

Источник: Global CIO