Тюрин Владислав

Нужно больше данных

2015

Нужно больше данных.

В публикации рассматривается понятие больших данных с точки зрения потребностей и возможностей бизнеса. Сфера внедрения технологий, основанных на больших данных, обширна. Но именно для бизнеса стоит насущная проблема их обработки исходя из экономической целесообразности и извлечения практической выгоды от больших и разнородных объемов данных. Сегодня эффективный развивающийся бизнес, который ориентирован на работу в плотной информационной среде, должен понимать преимущества и недостатки больших данных, разбираться в терминологии и в особенностях технологий, должен уметь формулировать задачи с ними связанные и использовать те предложения, которые существуют или появляются на рынке. Цель публикации дать представление о больших данных и аналитики, рационального их применения в бизнес среде.

Публикация осуществляется в рамках проекта Vlad’s Business Objects.
Сайт проекта: bizobj.ru

Содержание

Об очень больших данных

Обработка нарастающего объема разнообразных данных обуславливает мощность вычислительных устройств. А уверенный рост мощности электронно-вычислительных машин способствует увеличению объема и спектра обрабатываемых данных. Гонка производительности компьютеров и объема используемой информации привела к неизбежному появлению термина «большие данные» (big data), который ввел Клиффорд Линч в специальном номере журнала Nature на тему «Как могут повлиять на будущее науки технологии, открывающие возможности работы с большими объёмами данных?» (август 2008 года). Так как же могут повлиять большие объемы данных?

Рассмотрим понятие больших данных с точки зрения потребностей и возможностей бизнеса. Сфера применения технологий, основанных на больших данных, обширна. Но именно для бизнеса стоит насущная проблема их обработки исходя из экономической целесообразности и извлечения реальной практической выгоды от больших, очень больших и разнородных объемов данных. Для бизнеса характерна постановка и формулировка целей по сбору, обработке и анализу больших данных, не только как интересной исследовательской программы, но, в большей степени, как проекта для получения фактического конкурентного преимущества и экономической выгоды. Давайте посмотрим, что может бизнес найти в области больших данных. Какие перспективы у крупных компаний, которые настойчиво приступили к освоению громадных массивов данных для разнообразных исследований рынков и потребителей. Какую выгоду могут принести большие данные среднему и малому бизнесу и доступны ли они такой форме предпринимательства.

Прежде всего, уточним понятие «big data». Формулировки предлагаются разные, но их суть скорее сводится не к попытке установить физический объем, при котором данные становятся большими, а к определению некоторого массива данных превышающего «традиционный» нормальный объем обрабатываемых данных в рамках бизнеса. Поэтому остановимся на следующем варианте.

Большие данные *– разнообразные по своей сущности (содержанию и структуре) и существованию (форма записи, хранения и извлечения) данные в объемах значительно превышающих традиционные для бизнеса, эффективное использование которых обусловлено применением специальных знаний, методов, технологий и инструментов.*

Такое понимание больших данных основывается на открывающихся новых перспективах работы с данными, которые превосходят объемы, обычно используемые бизнесом. Становятся доступны огромные массивы внутренних и безграничные пространства открытых внешних данных. Для каждого бизнеса такие объемы были и остаются разными. Поэтому для конкретной компании будет свое выражение физического объема «больших данных» и оно всегда будет связано с необходимостью приложить существенно большие усилия для их использования.

Возникновение больших данных – это результат увеличения физического объема структурированных и слабо структурированных данных за счет роста технологий сбора и обмена информацией. В том числе в связи с появлением новых и интенсивным развитием существующих программно-аппаратных средств регистрации состояний, цифровой фиксации, взаимодействия пользователей и вычислительных устройств.

Не обошлось в популяризации технологий обработки больших объемов данных и без маркетинга. Определенный интерес был вызван новыми программными продуктами и технологиями, которые продвигаются на рынки, как способные обрабатывать громадные объемы неструктурированных данных. С одной стороны, их разработчики отвечают растущим потребностям бизнеса, с другой – заинтересовано демонстрируют потенциал больших данных и свои компетенции в этой области.

Сегодняшняя ситуация подводит к тому, что эффективный развивающийся бизнес, который ориентирован на работу в плотной информационной среде, должен понимать преимущества и недостатки больших данных, разбираться в терминологии и в особенностях технологий, должен уметь формулировать задачи с ними связанные и

применять те предложения, которые существуют или появляются на рынке. И всё это в условиях оптимального расходования своих ресурсов.

Сегодня не существует проблем в аппаратном и программном обеспечении процесса сбора, обработки и хранения больших данных. Существуют проблемы в понимании бизнесом определения и значения больших данных, в понимании их преимуществ и недостатков, в понимании источников данных и порядка их освоения.

При рассмотрении вопросов, связанных с большими данными, придется затронуть такие темы как бизнес-аналитика, бизнес-моделирование, структурирование бизнеса, кадровое обеспечение, экономика, маркетинг, информационные технологии и т.п. Некоторые вещи придется только упомянуть, поскольку если останавливаться на них подробно, то публикация превратиться в чрезвычайно «большие данные».

BIG MANAGED DATA

Большие данные – это большой потенциал для бизнеса. Их сбор, хранение, обработка, аналитика требуют внушительных усилий и расхода ресурсов. И это нужно для того, чтобы делать меньше ошибок, и чтобы минимизировать последствия сделанных ошибок. Иными словами, управление большими данными имеет главной целью принятие качественных решений.

А что значит управлять большими данными? Это весьма специфичный ресурс для бизнеса, нуждающийся в пристальном профессиональном внимании и последовательном изучении. Большие данные призваны решать серьезные проблемы и к ним предъявляются повышенные требования. Они должны быть управляемыми, т.е. понятны и готовы к обработке, анализу, форматированию, хранению, мониторингу, представлению. Чтобы стать *управляемыми* и полезными бизнесу, данные необходимо собирать, структурировать, классифицировать и сохранять.

Разберемся в природе больших данных и попробуем понять, что необходимо, для того чтобы иметь «большие управляемые данные».

Потоки данных

Большие данные обязаны своим появлением и укреплением позиций в мире бизнеса возросшему потоку цифровой информации. В значительной мере её избыток и неизбежность управлять таким потоком заставили задуматься о том, как это сделать наиболее рационально. С другой стороны, очевидные потребности компаний в получении дополнительной информации о рынках, потребителях, конкурентах, работниках, конъюнктуре приводят к поиску новых источников информации. Не стоит забывать, что появилось множество вариантов прямого сбора, в том числе регистрации широко спектра параметров и сведений, которые раньше отсутствовали. Взять хотя бы регистрирующие датчики в мобильных устройствах.

Расширяющиеся инструменты фиксации в различной цифровой и аналоговой форме – это тоже регистрация состояния, которое считывается, сохраняется и обрабатывается. Возросшее количество устройств сформировало сеть цифровых регистраторов, генерирующих гигантские объемы данных. Интенсивно развивающийся «интернет вещей» увеличит массивы обрабатываемых и хранимых данных. Не забудем и о том, что создают большие данные как непосредственно пользователи, так и цифровые устройства самостоятельно по заложенным алгоритмам. Наблюдая плотные потоки информации бизнес видит в них определённые преимущества и хочет ими воспользоваться на законных основаниях.

Феномен больших данных – это не столько результат увеличения некоторой информации в окружающем мире, сколько умение и способность собирать всё больше и больше информации из окружающего мира. Ещё «вчера» управляющий супермаркета вынужден был поставить работника на входе, чтобы считать посетителей для оценки популярности его заведения. А уже «сегодня» он снимает данные о движении посетителя по торговому залу и формирует оптимальный поток покупателей. Возможно ли будет «завтра» прогнозировать время визита конкретного потребителя и его покупки в «умном»

торговом пространстве с индивидуальными предложениями и сервисом? Но это не результат какого-то революционного изменения объемов информационного пространства – это результат повышения «плотности» собираемых данных.

Плотность регистрируемых данных *– объем данных собираемых об одном объекте (событии, процессе, явлении).* ***Повышение плотности данных*** *– это увеличение объема данных собираемых об одном объекте (событии, процессе, явлении).*

Особенность фиксации и регистрации цифровых данных – это дискретность. Каждый конкретный сбор данных – это отдельное событие, в определенной степени независимое от других подобных. Из-за этого набор собираемых данных в каждом событии сбора конечен. Как правило такой набор данных существует в рамках поставленной задачи и зависит от инструментов и методов регистрации. Допустимо говорить о *регистрации состояния* объекта (события, процесса, явления), как о формировании некоторого набора данных на заданный момент времени. Временные факторы сбора данных о состоянии имеют особое влияние, и они учитываются наравне со множеством других.

Нарастающий объем фиксируемых, собираемых, хранимых, обрабатываемых данных формирует их потоки. Производительность некоторых видов устройств, обильно генерирующих данные, создает проблемы массового обмена данными и их «упаковывания» в некие общие информационно-транспортные каналы или в производительные контейнеры хранения.

Весомая проблема потоков данных – их слабая структурность и связанность.

Структурирование данных

Не зависимо от того, являются ли данные для бизнеса большими или традиционными, они могут быть представлены (собраны и сохранены) в структурированном и не структурированном виде.

Для структурированных данных характерно их разделение на некоторые, сформированные по правилам, единицы и наборы данных, связанные или зависимые от других данных. В структурированных данных выделяются отдельные простые единицы данных, которые по определенным правилам составляют наборы данных. Для структурированных данных важно наличие разнообразных зависимостей.

Структурированные данные *– это данные в которых на основе их содержимого и формата выделены отдельные элементы данных, между которыми установлены заданные взаимные соответствия.*

Индексированные данные *– это данные для которых на основе их содержимого и формата выявлен набор указателей (индексов), которые позволяют осуществлять поиск целевых фрагментов данных.*

Прямая регистрация данных позволяет сформировать некоторый поток слабо структурированных или совсем неструктурированных данных. С потоком практически ничего толкового нельзя сделать и уж тем более провести какой-либо вразумительный анализ. Структурирование потребуется в любом случае, если поток необходимо каким-либо образом обработать. Это может быть простое индексирование или разметка, а может быть построение целой модели взаимосвязанных элементов данных и метаданных.

Без некоторого минимального уровня структурирования данных они превращаются в бесконечный поток без начала и конца и теряют всякий смысл. Красиво именуемые «озера данных» легко могут превратиться в сточные канавы, если в них сливать всё подряд и без разбора. Не стоит верить тем, кто убеждает в существовании необработанных массивов информации в «озерах». Уже при помещении данных в него, они как минимум оцениваются и индексируются по содержимому, источнику, свойствам. А как же иначе их потом изъять из озера? И чем более структурированы данные при помещении в хранилище, тем удобнее и полезнее они становятся для последующего извлечения, обработки и анализа. Отсутствие какой-либо структуры данных делает их практически бесполезными.

Озеро данных *– массив данных, хранимых в необработанном виде.*

Кстати метафора «озера данных» не совсем подходит к таким понятиям как «данные», «информация». Обратите внимание, что вода имеет очевидное свойство интенсивно смешиваться. Это категорически исключается в области данных: перемешивание на низком структурном уровне приводит к потере смысла.

Структурирование данных включает три базовых этапа.

Во-первых, в потоке данных выделяются некоторые минимально допустимые целостные единицы данных. Эти целостные единицы представляют собой значимую информацию и не должны утрачивать смысл в результате их вычленения в потоке. В зависимости от систем обработки и хранения информации целостными единицами данных могут быть: отдельные слова, характеристики, числовые данные, выражения, функции, объекты, документы, фрагменты, снимки, сегменты и т.п. Выделение в данных целостных единиц может сопровождаться их изменением для приведения к единому формату. Кроме того, при выделении единиц данных возможно добавление недостающих данных либо удаление избыточно-повторяющихся.

Во-вторых, выделенные в потоке целостные единицы данных упорядочиваются в контейнеры. Это могут быть таблицы, реляционные базы данных, иерархические структуры, наборы, списки, домены, классификаторы, библиотеки и т.п. Вводятся и соблюдаются правила упорядочивания единиц данных, что в какой-то степени позволяет проверить их корректность и полноту.

В-третьих, между единицами данных и между упорядоченными наборами единиц данных устанавливаются связи. При этом такие связи могут иметь дифференцированные атрибуты и назначение. Важно указать для дальнейшей обработки данных: что с чем, как и почему увязано. Так же, как и упорядочивание, определение связей в данных позволяет отследить их корректность и полноту.

Существуют различные методы и форматы структурирования данных. Их выбор зависит от конкретной ситуации и задач по хранению информации. Само содержание данных обуславливает такой выбор. Например, для хранения записей журнала о входящей корреспонденции разумно употребить таблицу или реляционную базу данных, а для хранения музыкальных альбомов – библиотеку аудио треков с индексацией по названию, исполнителю, длительности, году издания.

Структурирование данных сопряжено с распознаванием образов. В частности, это касается изображений и видеопотоков. Когда-то документы приходилось перепечатывать вручную или снимать фотокопии – сегодня на рынке предлагаются мощные средства распознавания текстов. Проблемы структурирования и индексации изображений, аудио и видео данных сегодня легко решаются, если у бизнеса есть в этом потребность.

Разные специалисты смотрят на структурирование данных и единицы данных по-разному. Для программиста – это типы и структуры языка программирования. Для разработчика реляционных баз данных – это типы полей таблиц, структуры и связи таблиц базы данных. Для аналитика – это типы количественных и качественных показателей, комплекты расчетных и аналитических сводных индексов (коэффициентов, функций).

Структурирование данных базируется на понятиях: единица данных, набор единиц данных, целостность данных и их зависимость.

Структурирование упорядочивает и связывает данные создавая читаемую и понятную основу, наделяя их дополнительными «данными о данных» – метаданными.

Хранение данных только в структурированном виде на практике малопригодно для решения актуальных задач. Может потребоваться сохранить данные в первоначальном виде как поток. А значит данные предпочтительно формировать по слоям:

- первый слой – это данные в первичном виде (как получены);
- второй слой – это структуры данных (в том числе индексы);
- третий слой – это аналитические данные (обработки, расчеты, коэффициенты);
- четвертый слой – это агрегированные данные и сводные показатели;
- пятый слой – это публичные презентационные данные, которые допускается представить широкой аудитории или руководству бизнеса.

Многослойное хранение данных удобно и обеспечивает эффективное их использование для различных целей как в настоящем, так и в будущем. Послойное хранение не имеет аппаратных и программных ограничений и существенно не влияет на издержки, связанные с информационной аналитикой.

Главная задача структурирования данных – это приведение их к такому виду, с которым результативно работать:

- обрабатывать,
- извлекать по запросам,
- формировать новое,
- обновлять,
- находить и исправлять ошибки,
- представлять в удобном виде,
- обнаруживать закономерности и проблемы,
- соотносить с целевыми объектами.

Работа с плохо структурированными данными сопряжена с трудностями, непредсказуемыми результатами и ошибками. Работа же с неструктурированными или слабо структурированными данными невозможна.

В подавляющем большинстве случаев человек имеет дело со структурированными данными, хотя и не фокусируется на этом, что удобно и правильно. Также, удобно и правильно передавать структурированные данные.

Есть структуры данных, которые мы даже не замечаем, потому что всегда имели и будем иметь с ними дело, потому что мы их знаем, мы их изучаем постоянно и постепенно обновляем в своей памяти, совершенствуем – это, например:

- тексты (строго структурированные по правилам лингвистики данные);
- изображения (образно-структурированные данные с различаемыми и узнаваемыми формами и цветами объектов).

Большие данные ещё больше нуждаются в структурировании. Это приоритетный способ разумно и эффективно их извлекать для обработки из мест хранения и обрабатывать без потери значимости. Без структурирования большие данные подвержены спутыванию, смешиванию, утрате смысла. Они теряются в огромных потоках информации и исчезают из поля зрения аналитиков. Когда мы говорим о возможности получать преимущества от использования больших данных, мы понимаем, что эти преимущества основаны на потенциале управления структурами данных.

Разные виды данных

Для управления большими данными потребуется разбираться в их сущности. Сопоставление данных, объединение, разделение, пересечение и иные способы обработки могут осуществляться только при понимании характеристик данных. Большие данные нуждаются в определении типа и вида, классификации, индексировании, ранжировании, стандартизации, в установлении приоритетов.

Классификация данных *– определение характерных свойств данных, которые позволяют разделять их на отдельные группы с учетом особенностей их сбора, обработки, (в том числе структурирования), хранения и извлечения.*

Само понятие «большие данные» наделяет информацию некоторым дополнительным смыслом. Например, традиционно для бизнеса было достаточным собрать и сохранить информацию о сделке в виде транзакционных сведений о покупке:

- какой товар куплен и сколько,
- по какой цене,
- какая скидка предоставлена,
- в какое время совершена покупка.

Теперь, учитывая понятие «больших данных», бизнес вынужден расширить традиционное представление о собираемых сведениях по сделке и увеличить плотность данных о регистрируемом событии, включив в данные о покупке:

- профиль клиента,
- предпочтения покупателя,

- передвижение покупателя по торговому залу,
- вопросы покупателя продавцам-консультантам,
- реакцию покупателя на обращение работников торговой точки,
- время, проведенное в торговом зале и в зоне кассового обслуживания,
- способ расчетов,
- инциденты, связанные с покупателем и работниками, которые с ним общались.

Соответственно, может оказаться полезным при структурировании и классификации выделять «традиционные» данные и «дополнительные данные». В том числе это позволит оценить эффект от использования больших данных.

По своей сути, когда для некоторых данных устанавливается определенная классификация – это введение дополнительных данных о других данных. Такие дополнительные данные не регистрируются и не собираются относительного некоторого субъекта, а формируются как характеристика других данных.

Метаданные *– данные о данных.*

Метаданные обеспечивают богатые возможности при обработке данных и их анализе. Как правило метаданные – это результат оценки обычных данных, наделяющий их дополнительными свойствами – что само по себе означает аналитическое («кабинетное») увеличение массивов больших данных.

Метаданные – это относительная категория: когда в отношении одних данных говориться, что они являются метаданными для других данных. Без такой относительности метаданные бессмысленны.

Несколько другая, но тоже относительная категория данных, это – контекстные данные.

Контекстные данные *– данные сопровождающие (окружающие) другие данные.*

Контекстными называются данные, которые показывают в какой ситуации были собраны, сохранены, обработаны, структурированы прямые данные. Нужны ли контекстные данные аналитику? Несомненно, они влияют на определение закономерностей, взаимных связей, формулирование выводов.

Чаще данные контекстно связываются через указание времени их регистрации (сбора, обработки). Контекст, как правило, зависит от последовательного изменения бизнеса и трансформации внешней среды и время - это основной элемент сопряжения прямых и контекстных данных. Так же, как и для метаданных, говорить о контекстных данных целесообразно в отношении каких-то других (прямых) данных. Часто данные составляют взаимный контекст друг другу.

Обработка данных неминуемо ведет к созданию новых массивов, которые уже не собраны от первоисточников и не являются результатом скрупулёзной регистрации, а представляют собой итоги расчетов, сортировки, фильтрации, слияния, пересечения. Разные уровни обработки данных порождают различные категории получаемых данных. За этим приходится следить и учитывать в дальнейшем. Исключить путаницу, специалистам помогает разделение на:

- первичные данные – это простые данные непосредственно полученные от источника информации;
- вторичные данные – это обработанные данные, которые определяют или иллюстрируют выводы, полученные на базе первичных данных;
- третичные данные – это обработанные и нормализованные данные, которые устанавливают и стандартизируют информацию (знания).

В целом же данные, в зависимости от степени их обработки, формируют некоторые слои в хранилище. Для последующего анализа требуется сохранить взаимные отношения между слоями, между первичными данными и полученными на их основе вторичными и третичными данными.

Классификацию данных можно осуществлять по:

- смысловому содержимому данных;
- плотности собираемых данных;
- размерности единиц данных;
- целям использования данных;

- источникам данных;
- способам обработки данных;
- форматам хранения данных;
- приоритету данных для бизнеса;
- отраслям данных;
- культурным особенностям данных.

Классифицируя данные принципиально понимать различие их источников. Данные объекта, это не то же самое, что данные события. Данные процесса, это совсем не то же самое, что данные атрибута. Всегда следует помнить, о чем и от какого источника получены ваши данные.

Почему большие данные нуждаются в классификации? Да потому что работать с большими данными сложно. Крайне сложно в больших данных уследить за отдельными элементами или наборами. Крайне сложно извлекать нужные, актуальные, релевантные данные для аналитики и принятия решения. Классификация – это инструмент управления большими данными и это тот инструмент, который разделив данные на группы, существенно повышает их управляемость.

Сбор и хранение данных

Сбор и хранение данных – это процесс получения данных, их преобразования, упаковывания и упорядочивания в форматы хранения, их внесение в некоторое место хранения в дополнение к уже имеющимся там.

Получение данных осуществляется двумя способами: непосредственно их сбором (регистрацией) от некоторого источника информации или получением от стороннего лица ранее собранных и обработанных массивов данных. Каждый из способов имеет свои преимущества и недостатки. Для примера достаточно указать, что сбор данных от источника занимает время и специальные инструменты, а приобретение данных у стороннего лица влечет финансовые затраты.

Традиционно, бизнес уделяет много внимания сбору транзакционных экономических и технологических данных. Для этого предназначены учетные системы, автоматизированные системы управления, корпоративные информационные системы. Данные полученные таким образом имеют весьма высокую степень релевантности и достоверности. В противовес транзакционным данным, покупка больших маркетинговых данных для соответствующего анализа сопряжена с финансовыми и репутационными рисками. Очевидно, что необходим тщательный выбор способов получения информации и её источников.

Источник информации *– это объект (процесс, событие, явление) от которого можно получить информацию о нем самом, либо о другом объекте (процессе, событии, явлении).*

Обычно источники делят на две большие группы: внутренние и внешние. Внутренние источники находятся внутри бизнеса, зависимы от него и в определенной степени подконтрольны. Внешние источники – вне бизнес-модели, относительно независимы от бизнеса и не всегда доступны. В наибольшей степени бизнес может влиять на сбор данных от внутренних источников и в наименьшей – от внешних. Иногда это оказывает решающее влияние на качество данных и мешает получению пригодной для дальнейшего оборота информации. Умение находить внешние и выявлять внутренние источники – это составляющая системы управления большими данными.

Получение данных сопряжено с некоторыми оценками их качества. Весьма полезно собирая наборы данных и помещая их на хранение, оценивать их с качественной стороны.

Показатели качества данных:

- актуальность данных – соответствие данных временному и смысловому контексту сбора, хранения и последующей их обработки;
- объективность и достоверность данных – насколько данные отражают реальную ситуацию и не противоречат ли действительности;

- полнота данных – в достаточном ли объеме собраны данные, для всестороннего анализа и не упущена ли какая-то их составляющая;

- релевантность данных – соответствие данных целям и задачам их сбора, хранения и последующей обработки;

- чистота данных – присутствие в данных избыточного их количества, дублирования, не нужных фрагментов и т.п.

- примитивность данных – являются ли данные объективно зарегистрированными или обработаны какими-либо инструментами;

- ценность данных – насколько важны данные для бизнеса в процессе сбора, хранения и последующей их обработки.

Приходится вводить простые и комплексные критерии качества данных. Но совершенно необходимо их вводить для оценки данных и их использования в последующих операциях обработки, хранения, извлечения. Именно по критериям качества существенно разняться внутренние и внешние источники. А во многих случаях оценка качества данных сопряжена с оценкой качества их источника.

Общепринятые методы сбора информации: наблюдение, считывание, тестирование, интервью, эксперимент, моделирование. В общей схеме сбора информации присутствуют такие элементы как источник информации, сборщик информации, фильтр (отбор) источников информации, метод сбора информации, информационный фильтр (критерии отбора данных). Организация сбора данных требует профессиональных знаний, опыта и предварительной подготовки с учетом специфики бизнеса.

Полученные данные для последующего использования необходимо сохранить. Структурирование имеет особое значение при сохранении данных в хранилищах (местах долговременного размещения данных). Отчетливо надо понимать, что в хранилища должны попадать управляемые данные. При этом технологии хранения строго устанавливают правила форматов и типов данных, которые допускается в них размещать. Конечно же любые правила можно обойти и искусно упаковать любую информацию в любое хранилище, по крайней мере в самом примитивном виде или представлении. Но такие нарушения окажут плохую службу в дальнейшем, когда возникнет потребность получить с хранения информацию, а обнаружить её и извлечь в целостности просто невозможно.

Во многом правила мест хранения, обусловленные технологиями и форматами, предъявляют существенные ограничения к структуре сохраняемых данных и к параметрам их индексации. Для решения ряда вопросов привлекается специалист по технологиям хранения данных. И может выйти так, что требования хранилища будут противоречить потребностям бизнеса. Эти противоречия придется рано или поздно устранять, изменяя потребности, корректирую поведение хранилища или перемещая данные в другое хранилище. Соответственно, приступая к работе с большими данными не стоит недооценивать вопрос выбора хранилища.

Задача хранилища – это выдать данные по требованию его клиента. Казалось бы, место хранения решает вопрос размещения данных. Но ведь по сути, мы сохраняем данные, чтобы потом их каким-то способом обработать. Извлечение данных из хранилища – это вторая и, пожалуй, наиболее важная сторона проблемы хранения данных. Зачастую клиент обращается в хранилище имея общее представление о запрашиваемой информации. Кроме того, получая какие-то данные, клиент должен понимать связаны ли с ними иные данные и какие. Извлечь некий набор конечных данных из места хранения – это половина дела. С целевыми данными могут быть соотнесены метаданные, контекстные данные или любые иные данные имеющиеся (или даже не имеющиеся) в хранилище. Да и сам процесс получения данных имеет ряд технологических особенностей, начиная от конкуренции за конкретные их фрагменты и кончая нарушением их целостности в хранилище. Возникает потребность в подготовке и проведении исследования данных в хранилищах.

Теперь, когда мы оценили сложность больших данных, разберемся, как и что с ними делать, чтобы существенно повысить их результативность для бизнеса.

BIG BUSINESS DATA

Использование больших данных в бизнесе начинается с понимания проблемы поиска эффективного управленческого решения в условиях дефицита информации. Затем, неминуемо, менеджмент приходит к необходимости четко сформулировать цели (задачи) и стратегию больших данных.

Постановка цели и выработка стратегии

Сочетание следующих трех факторов приводит к тому, что инициаторами перехода к аналитике на базе больших данных часто являются маркетинговые подразделения:

1) особенности и интенсивность маркетинга и технологий современных продаж, как «передовой» его составляющей,

2) характерные черты и доступность открытой информации о потребителях глобальной информационной сети,

3) производительность современных аппаратных и программных средств.

Проявляя положительную активность, маркетологи мотивируют руководство перейти от традиционных данных к более мощному, но более сложному и затратному анализу больших данных.

Задумываясь о таком инструменте как большие данные, менеджмент компании должен несколько отстраниться от темы маркетинга и оценить пользу новых технологий в рамках всей бизнес-модели. Ограничить большие данные исследованиями рынка – это низкоэффективный подход.

Прежде всего хорошо бы понимать суть и потенциал больших данных, оценить их преимущества и пользу для текущей бизнес-модели. Особенно полезен интегрированный переход к большим данным в рамках общей стратегии развития. Как и любая другая традиционная аналитика, аналитика больших данных связана со многими аспектами существования бизнеса. И сверх того, именно большая аналитика нуждается в системном подходе и внедрении.

Большая аналитика *– экономическая, финансовая, управленческая, маркетинговая и иная аналитика, основанная на больших данных и призванная подготовить принятие эффективных бизнес-решений.*

В реализации задач, поставленных перед большими данными, на разных этапах, будут принимать разные подразделения компании. Вероятнее, будут привлечены внешние консультанты, особенно на этапе перехода к аналитике больших данных.

Как и с любыми масштабными энергичными современными информационными технологиями, предварительно сложно оценить весь эффект и результат от их внедрения. Поэтому начинают с нескольких конкретных задач связывающих в первую очередь экономическую и маркетинговую составляющие бизнеса. В разных вариациях это могут быть отделы: маркетинга, аналитики, логистики, стратегического и оперативного планирования.

Важно сразу определить реалистичные и практические задачи для решения которых предполагается использовать большие данные. Первична постановка корректной цели по правилам стратегического планирования.

Аналитика больших данных – это проект и уместно проектное управление. Оптимальным на этапе внедрения аналитики больших данных считается формирование проектной команды. В последующем, функции распределяются в рамках расширения такого проекта по специалистам подразделений компании.

Но какими бы ни были организационные решения, от четкой постановки цели на уровне стратегии бизнеса не уйти. Вот вопросы, на которые руководство должно знать ответ до того, как приступит к финансированию проекта большой аналитики:

- Для чего собираются большие данные?
- Какой результат ожидается получить от больших данных?
- Сколько больших данных нужно и какие это данные?
- Как большие данные связаны со стратегией бизнеса?
- Готов ли бизнес к изменениям при использовании больших данных?

- Какие риски сопровождают большие данные?

При ответе на вопросы, не забывайте, что большие данные можно формировать во внутренней среде бизнеса. А значит в качестве первых задач для большой аналитики могут быть предъявлены не только маркетинговые, но и кадровые, финансовые, логистические, производственные (внутренние) проблемы.

Безусловно задачи, которые ставятся перед большими данными обязаны соответствовать стратегии бизнеса в целом. Противоречия не допустимы. Наличие обоснованных сомнений в необходимости аналитики больших данных является весомой причиной для отказа от неё. Как наиболее оптимальная и менее затратная альтернатива – постепенное наращивание плотности собираемых традиционных данных и простая экономическая и маркетинговая аналитика на их базе.

Переход к большим данным в любом случае ведет к существенным изменениям – первый раз на стадии внедрения технологий больших данных и второй раз после получения аналитических результатов, основанных на больших данных. Проект больших данных – это как на свой страх и риск передать на полное обследование собственный бизнес внешнему аудитору и ждать, что он рано или поздно вынесет свой суровый вердикт «здесь и здесь у вас плохо, а вот здесь – уже поздно что-то менять». Что делать бизнесу, если анализ показывает неизбежность кардинальных изменений? Ответ зависит от готовности топ-менеджмента к переменам. Конечно же, сегодня компании, особенно лидеры рынков, понимают насколько важно быть активным в изменениях и насколько важно вовремя осуществлять подстройку бизнес-модели под меняющуюся реальность.

Большим данным нужна своя стратегия. И пусть это не стратегия равная общей стратегии развития бизнеса, но это серьезная проектная стратегия для успешной реализации проекта и понятная база для работы проектной команды.

Вряд ли понадобиться создание отдельного подразделения, тем более на первых этапах работы. Однако руководству стоит задуматься о профессиональном отделе аналитики или о введении в управленческие структуры бизнеса позиций квалифицированных аналитиков по направлениям.

На этапе проработки проекта перехода к большой аналитике компания определяется со многими профессиональными, методическими, технологическими, кадровыми сторонами его реализации. От выбора методик, технологий, кадровых аппаратных и программных решений зависит успех проекта и его стоимость. В какой-то момент немаловажным станет вопрос построения хранилища.

Большие данные – это не столько большая информационная система, сколько процесс (и связанные с ним объекты и функционал). Этот процесс призван получить значительные и значимые объемы информации из внутренней и внешней среды с последующей их обработкой для выяснения объективной картины развития бизнеса и поиска его недостатков, особенностей, возможностей, угроз, преимуществ.

Вот о чём целесообразно подумать при переходе к большим данным:

1. Стратегия больших данных
    a. Основная цель и задача больших данных
        i. Зачем нужны большие данные бизнесу
        ii. Сколько данных нужно бизнесу
        iii. Кто поставляет данные
    b. Допустимые изменения бизнеса на основе больших данных (готовность к переменам)
        i. Ориентирование бизнеса на большую аналитику
        ii. Принятие решений на основе результатов большой аналитики
        iii. Глубина проникновения большой аналитики в бизнес-процессы компании
    c. Преимущества больших данных
        i. Рыночные (конкурентные) преимущества
        ii. Преимущества эффективных решений (управленческие преимущества)

iii. Финансово-экономические, производственные и логистические преимущества

2. Обеспечение больших данных
    a. Источники больших данных
        i. Внутренние источники
        ii. Внешние источники
        iii. Поставщики больших данных
    b. Инструменты больших данных
        i. Инструменты сбора (получения) больших данных
        ii. Инструменты обработки данных
        iii. Аналитические инструменты
    c. Технологии больших данных
        i. Хранилища больших данных
        ii. Аппаратные средства больших данных
        iii. ИТ-средства управления большими данными
3. Управление большими данными
    a. Проект больших данных
        i. Планирование проекта в рамках бизнеса
        ii. Подготовка команды проекта
        iii. Обеспечение проекта ресурсами
    b. Команда проекта больших данных
        i. ИТ-специалисты
        ii. Специалисты-аналитики
        iii. Руководство проекта
    c. Контроль и координация проекта
        i. Измеримые задачи и работы проекта
        ii. Привлечение сторонних консультантов и экспертов
        iii. Внешнее наблюдение (контроль и координация)

Всё выше описанное слишком масштабно для среднего и малого бизнеса со всеми проектными, информационными, аналитическими инструментами и технологиями. Так что же большие данные только для крупных корпораций? Конечно же нет. В больших данных есть большой потенциал и для субъектов среднего и малого предпринимательства. Имеются проблемы с развитием рынка больших данных для таких субъектов, которые вполне могли бы воспользоваться преимуществом больших данных на базе определенного вида аутсорсинга большой аналитики. Рынок услуг такого рода пока формируется и адекватных предложений недостаточно.

Уместным для малого/среднего бизнеса будет:

- постепенное повышение плотности собираемых традиционных данных,
- введение дополнительного аналитического функционала,
- назначение ответственного за работу с данными специалиста или координатора.

Конечно же многие задачи большой аналитики малому и среднему бизнесу решать лучше с внешними консультантами и профессионалами. Не исключена и кооперация по проблемным вопросам на основе имеющихся данных с другими субъектами рынка. Если же средний или малый бизнес является спутником (поставщиком, посредником) крупной компании, то очевидно получение существенного конкурентного преимущества через подключение к проекту больших данных такого партнера.

Если большие данные накапливаются с ними надо что-то делать

Основным мотивом для повышения уровня управления ресурсом является его избыток. Так излишки данных хорошо мотивируют их более эффективное использование. Действительно, если компании удалось накопить массивы информации, то очень хочется получить от них максимальный результат. Что толку, что они лежат цифровым грузом на серверах предприятия или в платных облачных сервисах. Кто-то их просто потеряет, а кто-то попытается извлечь выгоду. Данные накапливаются и с ними надо что-то делать.

Данные – это сильный ресурс:

1) текущие данные – это актуально для текущих решений – это тактика и оперативное планирование;

2) данные прошлых периодов – это понимание успешности и ошибочности выбранного пути и принятых ранее решений – это стратегия.

С другой стороны, во многих случаях бизнес упускает возможности самостоятельного сбора ценных данных, которые остаются вне наблюдения и регистрации. А некоторые могли бы помочь в принятии решений.

Но ведь есть и данные, которые бизнес собирает безуспешно – лишние и повторяющиеся, не достоверные и неполные.

Чтобы извлекать максимальную пользу из традиционных и больших данных, менеджмент должен понимать потенциал и особенности, а также уметь формировать цели и стратегию их использования. В сборе данных задействуются все направления: бухгалтерия, финансы, экономика, производство, маркетинг, логистика, ИТ, продажи, безопасность, кадры, планирование. Они же пользуются результатами сбора данных, их обработки и анализа. Есть разные уровни решений и все они могут быть поддержаны аналитикой больших данных.

Инструменты

Пожалуй, первое, о чем начинается разговор, когда необходимо обеспечить бизнесу работу с большими данными – это аппаратно-программный комплекс. Конечно же это важная сторона дела, но не первостепенная. Истратить финансы на высокопроизводительный сервер никогда не поздно. Тем более что предложений на рынке предостаточно и на самый изысканный вкус.

Хорошо бы начать с моделирования системы больших данных. Формализовать в понятном виде первичную и перспективную структуру, которая включит:

- источники данных,
- категории собираемых данных,
- уровни сохраняемых данных,
- логику обрабатываемых данных,
- результаты обработки данных.

Упростить данные до нескольких рабочих индикаторных показателей – преимущественная стратегия в отношении построения модели больших данных. Модель преобразования исходных данных в такие показатели формализуется в понятном виде для разных специалистов, участвующих в процессе.

Описанные потоки данных в рамках большой аналитики помогают в дальнейшей комплектации набора основных и дополнительных инструментов бизнеса.

Для моделирования потоков больших и традиционных данных используются специальные инструменты. Это позволяет подойти к проекту большой аналитики системно и профессионально.

Для работы с большими данными понадобятся инструменты, которые делятся на три группы:

1. Информационные инструменты – с помощью которых осуществляется непосредственный сбор, обработка и анализ данных. Информационные инструменты предназначены для работы с содержанием и структурой данных. К информационным инструментам относятся:

- методики сбора данных и технологии «добычи» данных (data mining – поиск и извлечение целостных данных и их структур из слабо структурированных или незнакомых данных);
- способы отбора источников информации;
- методы структурирования и индексирования данных;
- технологии обработки данных (их трансформации и приведения к целевому виду);
- методы анализа данных, в том числе математические и статистические;
- средства комплексной интеграции (де-интеграции) данных, их классификации.

2. Технологические инструменты (ИТ-инструменты) – с помощью которых осуществляется форматирование, хранение и представление данных на уровне программных и аппаратных решений. ИТ-инструменты ориентированы на работу с форматами и состояниями данных. Технологические инструменты это (в том числе):

- обеспечение сбора данных (регистраторы, сканеры, кодировщики, сенсоры, системы наблюдения, считыватели);
- серверы хранения данных (с функциями упаковывания данных в хранилище, поиска и извлечения данных);
- средства обработки и анализа данных;
- средства визуализации данных;
- технологии и протоколы обмена данными.

3. Организационные инструменты – это инструменты, с помощью которых осуществляется управление процессами большой аналитики. К ним относятся:

- методы проектного управления;
- методики и принципы командного подхода к реализации проекта больших данных;
- технологии моделирования потоков больших данных;
- средства применения результатов анализа больших данных.

Кроме того, все инструменты классифицируются по трем этапам работы с большими данными:

- сбор (регистрация, извлечение, получение) данных;
- обработка данных (структурирование, классификация, расчеты, моделирование, анализ);
- визуализация данных (панели индикаторов, мониторы данных, представление показателей и сводных индексов).

Цена и качество – два принципиальных критерия выбора инструмента. Значительные средства придется потратить на собственное программно-аппаратное обеспечение (ИТ-инструменты). Даже если ориентироваться на специальные сервисы больших данных затраты неизбежны (например, облачные сервисы). Это тот минимум, в который придется вложиться, хотя он не решающий для успешного использования больших данных.

Ряд информационных инструментов невозможно напрямую купить как продукт. Некоторые из них основаны на знаниях и опыте и приходят в бизнес только с соответствующим специалистом. Придется потрудится кадровым службам, чтобы найти профессионала или постепенно обучить кого-то из сотрудников.

Эффективное применение приобретённых или самостоятельно разработанных инструментов больших данных формирует то самое «волшебное» конкурентное преимущество на рынке. Не следует забывать об оценке экономической эффективности инструментов, поскольку результат от больших данных не должен превышать понесенные в связи с этим затраты. Такая оценка является экспертной из-за того, что не для всех инструментов определяется эффективность и полезный срок эксплуатации.

Один из способов относительного снижения затрат на большие данные – это совместное использование разных средств, программ, аппаратных комплексов несколькими подразделениями, их задействование в разных направлениях деятельности бизнеса. Особенно это удобно на начальных стадиях развития проекта, пока он ещё не набрал достаточные темпы и вполне способен уместиться на мощностях имеющихся вычислительных устройств.

Ещё из известных и набирающих популярность способов снижения издержек на ИТ-инфраструктуру – это облачные сервисы. Они позволяют нескольким бизнесам совместно, за адекватную плату, пользоваться технологическими инструментами.

Как и многое в сфере больших данных, выбор и использование инструментов для работы с ними – это зона ответственности высоко квалифицированных специалистов.

Команда

Встраивание аналитики больших данных в деятельность бизнеса на регулярной и профессиональной основе – это даже не отдельный функционал, а целое стратегическое направление. Без ответственных работников внедрение в жизнь больших планов в отношении больших данных скорее всего кончится неудачей.

В той или иной степени, с большими данными работают разные специалисты компании. К данным из внутренних источников имеют отношение буквально все работники.

Тем не менее на этапе перехода к большим данным и в процессе их использования есть ключевые функциональные роли, которые принципиальны для проекта.

Заказчик

Кто-то должен внутри бизнеса, находясь вне команды, поставить общую цель и сформулировать серию рабочих задач. Необходимо определить, как большие данные интегрируются в бизнес-модель и как измениться бизнес-модель после такой интеграции. Заказчик не обязан разбираться в деталях больших данных, но должен понимать зачем они бизнесу, какой результат они дают и как бизнес изменяется под их воздействием.

Заказчиком может быть некоторая группа специалистов и менеджеров, но имеющих полномочия по подготовке решений высокого уровня в отношении бизнес-модели компании.

Как правило, заказчик регулирует общие подходы к реализации проекта больших данных и выступает спонсором проекта.

Руководитель проекта

Возглавляет команду проекта больших данных – руководитель проекта. В его обязанности входит общая организация работ по проекту, в том числе: детализация целей, задач и планов проекта, оперативное планирование и контроль этапов проекта, планирование ресурсов и времени специалистов проекта.

Руководитель непосредственно отвечает перед заказчиком о ходе реализации проекта больших данных. Руководитель может совмещать свои функции с любыми функциями других ролей. Удачным был бы выбор, в качестве руководителя, профессионала с высоким уровнем подготовки и опыта в сфере экономического или маркетингового анализа данных обладающего также знаниями в области ИТ-технологий. Конечно же руководитель проекта подбирается из числа тех, кто способен возглавить проект как управленец.

Специалист по стратегическому планированию

Развитием проекта больших данных, в том числе укрупненным его планированием, может заниматься руководитель проекта. Для полноценного управления задачами проекта, с соблюдением намеченных заказчиком параметров и получаемых в рамках проекта результатов, привлекается специалист по стратегическому планированию. В его функционал включается планирование и координация проекта больших данных и бизнес-модели. Этот специалист должен иметь полномочия и права готовить решения высокого уровня по изменению стратегии развития компании и проекта развития больших данных, осуществляя при этом постоянную оценку результативности последнего.

Специалист по стратегическому планированию отслеживает текущий ход проекта по использованию больших данных, учитывает особенности бизнес-процессов и бизнес-объектов компании и имеет право предлагать решения по их максимально эффективному интегрированию.

Выделив специалиста по стратегическому планированию в отдельную роль команды проекта больших данных, бизнес существенно повышает ответственность проекта и снижает риск потери эффективности больших данных.

Функционал стратегического планирования можно разделить между заказчиком и руководителем проекта. Но это не самая лучшая идея, потому что у них просто не хватит время на повседневную работу в этом направлении.

Аналитик проекта

Одна из важнейших составляющих в проекте больших данных - аналитика. От работы аналитиков зависит конечный результат. Можно собрать идеальные громадные массивы красиво-структурированных данных и поместить их на великолепный суперсервер, но, если аналитик ничего не скажет полезного для бизнеса после того, как замучает хранилище запросами – проект провалиться. Аналитик в какой-то степени защищен от неудачи, если изначально разработана качественная модель потоков больших данных с выходными параметрами и показателями. Но от уровня профессионализма аналитика зависит очень многое, особенно когда на строгий суд заказчика понадобиться представить хоть что-нибудь впечатляющее и объяснить куда потрачены дефицитные финансовые ресурсы.

Аналитики проекта больших данных должны обладать профессиональными знаниями и умениями в области сбора и обработки данных, в области анализа экономических, финансовых, статистических и производственных данных. Фактически вся ключевая смысловая работа с данными ложиться на аналитиков проекта.

Выделим несколько специализаций:

- аналитик бизнес-модели (зона ответственности: понимание бизнес-модели, анализ бизнес-модели на основе традиционных и больших данных, формулировка и обоснование изменений в бизнес-модель, подготовка решений по бизнес-модели, увязка бизнес-модели и больших данных, требования к данным);
- аналитик структур данных (зона ответственности: понимание структур данных и их связь с элементами бизнес-модели, контроль и корректировка смысловой целостности данных и метаданных, изменение структур и классификации данных, контроль качества данных, сервис данных);
- аналитик рисков (зона ответственности: оценка потенциальных угроз данным и контроль информационных рисков, контроль достоверности данных и их источников, контроль рисков принятия решений на основе больших данных, вероятностная оценка прогнозных моделей).

Аналитик – это уникальный специалист, для каждого конкретного бизнеса. Он обладает компетенциями исключительными для понимания бизнес-модели. Уровень информации, к которой он имеет доступ – это фактически уровень топ-менеджмента соответствующего направления. По доступу к информации, осведомленности и пониманию особенностей бизнеса он ближе к руководству, чем к экспертной категории сотрудников. А иногда, аналитик объективней и реалистичней, чем само руководство. От результатов его работы зависит общее понимание экономической, финансовой, производственной, маркетинговой ситуации в которой оказался бизнес сегодня, что к этому привело и как он поведет себя в будущем.

Опыт работы профессионального аналитика весьма ценен для любого бизнеса, и он уникален также, как уникальна каждая отстроенная бизнес-модель. С другой стороны, аналитик должен иметь относительную независимость и не должен быть заинтересован своими расчетами и выводами подтверждать свою же правоту. Хороший аналитик сам заинтересован находить свои ошибки и исправлять их.

Не следует путать аналитика со статистиком или математиком. Он обычно понимает и умеет формулировать математические модели определенного класса, умеет применять статистические инструменты для обработки данных. В большей же степени он должен разбираться в том, как данные увязаны с бизнес-процессами и бизнес-объектами. Аналитику необходимо уметь разбираться в том, что означают собранные и обработанные данные сточки зрения экономических, производственных и рыночных процессов. Математические исследования и статистические доказательства – это не зона ответственности аналитика, это его инструментарий.

Аналитик – это и эксперт, и исследователь, и исполнитель, и дизайнер данных. Но аналитик не в состоянии заменить, например, «классного» менеджера по продажам. Это значит, что никакой глубокий, традиционный или большой анализ данных не наладит производственный или логистический процесс, не улучшит привлекательность и качество продукта, не гарантирует устойчивое финансовое положение. Аналитика лишь в состоянии показать, что идет не так в бизнесе, что заменить в бизнес-модели, на что обратить внимание.

ИТ-администратор проекта

Функции администраторов проекта составляют важные роли с точки зрения обеспечения ИТ-инфраструктуры проекта больших данных. По большому счету, специалисты, вовлекаемые в работу с большими данными со стороны подразделений ответственных за информационные технологии, решает общие вопросы бесперебойной работы программно-аппаратной инфраструктуры. Требования к ИТ с позиции больших данных имеют отличия по емкости, скорости и безопасности.

Можно говорить о следующих ключевых ИТ-администраторах проекта:

- администратор хранилища данных (зона ответственности: принятие данных в хранилище, проверка структуры данных, контроль размещения данных, исправление формата данных, формулирование и контроль запросов к хранилищу данных, контроль извлекаемых данных, сервис хранилища);
- администратор структур данных (зона ответственности: контроль и исправление структуры данных, классификация данных, контроль и получение метаданных, контроль и корректировка смысловой целостности данных, мониторинг качества данных);
- администратор системы защиты (зона ответственности: обеспечение защищенных соединений, контроль качества связи, защита конфиденциальности данных, управление учетными данными пользователей).

Очевидно, что одним из напрашивающихся способов сокращения команды проекта и издержек на такую команду – это ИТ-администратор в одном лице, выполняющий все упомянутые и сопутствующие им работы. Такой подход рекомендуется для старта проекта, но не для регулярной работы с большими данными.

Программист

Задача программиста разрабатывать программные средства обработки данных и автоматизировать работу с ними. Программист вовлекаемый в команду проекта больших данных должен иметь профессиональные знания и навыки не только в сфере объектно-ориентированного, функционального программирования и разработки алгоритмов, но и в сфере обработки крупных объемов информации.

Вопросы автоматизации больших данных бизнеса действительно серьезны для развития проекта. Большие объемы и задачи требуют существенного – в разы – сокращения времени на выполнение рутинных, типовых и повторяющихся операций. При этом следует понимать, что даже автоматическое выполнение операций компьютером сравнимое по затратам времени с выполнением той же операцией работником вручную (или в полуавтоматическом режиме) предпочтительней. Специалистов из команды больших данных надо освобождать от неквалифицированного труда. В этой связи, важен вопрос построения пользовательских интерфейсов программных продуктов для работы с большими данными. К ним несколько особых требований: простота, наглядность, логичность, системность, интуитивность и наличие подсказок. Громоздкие и сложные интерфейсы сведут на нет мощный функционал кода.

Супервайзер

Если у заказчика нет возможности компетентно и регулярно следить за ходом проекта больших данных, то имеет смысл ввести около-проектную позицию супервайзера.

Для объективной оценки работы команды проекта в целом и по отдельным задачам нужен относительно независимый контроль. А для того, чтобы избежать неожиданного провала проекта или временных задержек в реализации конкретных работ

по разным управляемым причинам, организуется постоянный, но не навязчивый контроль.

Желательно, чтобы супервайзер взаимодействовал с заказчиком, но не подчинялся ему. Супервайзером может быть внешний консультант понимающий суть и задачи проекта. Скорее всего, внешний консультант примет участие в проекте больших данных с самого его начала.

Эксперт

Команде проекта понадобиться участие различных экспертов. Если участие экспертов будет длительным, то их придется включать в команду проекта. Наверное, излишне говорить, что эксперты должны быть профессионалами в вопросах, которые помогают решать. Приветствуется привлечение независимых внешних консультантов.

Команда проекта больших данных по численному и качественному составу формируется в зависимости от сложности и амбициозности поставленных целей. Если поручено в сжатые сроки обеспечить внедрение большой аналитики в компании, то команда будет достаточно внушительна. Учитывая же практическую сторону вопроса и набирающий обороты рынок инструментов и специалистов, работающих с большими данными, начинать проект лучше с командой до 5 – 7 человек.

Команда, работающая с большими данными обязана постоянно совершенствоваться и развиваться, повышая свой профессиональный уровень и осваивая все более сложные и производительные инструменты. Такая команда становится экспертным центром бизнеса в области больших данных и аналитики. А являясь своеобразным центром компетенции, команда проекта больших данных вовлекает и обучает потребителей больших данных внутри бизнеса. Общение команды проекта с работниками компании на разных управленческих уровнях помогает в повышении качества отдельных элементов большой аналитики и её пользы не только для принятия стратегических решений, но и для повседневной работы по таким направлениям как экономика, кадры, финансы, логистика, операционный менеджмент, маркетинг, продажи, коммуникации, производство, качество, безопасность, гарантийный и послепродажный сервис.

Если бизнес стремиться извлечь максимальную выгоду от использования больших данных, то поддержание высокого профессионального уровня команды – одна из его базовых стратегических задач. Неминуемо возникнет проблема сохранения успешной команды проекта больших данных, а также проблема её расширения или трансформации во что-то большее. Утрата одного профессионала, особенно владеющего сложными инструментами, может привести к существенной потери производительности команды в целом. А если специалист был ключевым – то и к закрытию проекта больших данных.

Риски

Любой менеджмент несет в себе определенные риски из-за возможности принять неверное решение при информационных ограничениях. Именно для повышения эффективности принимаемых решений и снижения рисков неправильных решений компании обращаются к большим данным. Но ведь большим данным тоже сопутствуют риски. Оценим некоторые из них.

Риск конфиденциальности

Потеря контроля над данными и их передача в руки конкурентов может нанести серьёзный экономический и репутационный ущерб. Разглашение конфиденциальных данных в СМИ или в сети Интернет тоже нежелательны для бизнеса, даже если это не представляет явного коммерческого интереса для кого-то из игроков рынка.

Снизить опасность разглашения данных призвана система обеспечения безопасности.

В связи с риском конфиденциальности стоит отметить особый статус сервисов хранения и обработки данных, которые предоставляются сторонними компаниями

(«облака сторонних лиц»). Указанный риск здесь выше и непосредственно неподконтролен. Остается доверять порядочности таких поставщиков услуг и включать в контракты условия о компенсации разглашения данных третьим лицам.

Риск потери данных

Существенным риском для больших данных является их утрата (частичная или полная). Причины могут быть различны: от активности злоумышленников, до чрезвычайной ситуации. Единственный способ защитится – это резервирование данных. Очевидно однократное резервирование. Если оценка риска велика и сильно влияет на бизнес, то рекомендованы двукратное и трехкратное резервирование.

Одним из способов снижения рисков потери данных из-за ошибочных действий специалистов и пользователей – это предоставление рабочих копий данных (реплики полные или по запросам).

Риск переполнения хранилища

Неоптимальная система сбора и хранения больших данных в конечном итоге приведет к переполнению хранилища и утрате вновь получаемых данных при отсутствии места для физического их размещения. Особенность такой утраты данных – это потеря более актуальных «свежих» данных поступающих после события полного заполнения свободных объемов хранилища. Помогает тщательное планирование получения данных, умение оценивать их объемы и формировать хранилища, которые имеют адекватные емкости носителей для хранения.

Риск снижения эффективности больших данных

Четкость структуры собираемых и обрабатываемых данных, их управляемость и качество направлены на то, чтобы исключить снижение результативности работы с большими данными по мере разрастания их объемов. Помещение данных в хранилище должно быть управляемым и контролируемым. Даже если переполнение хранилища и не грозит, то сохранять в нем «всякие» данные не самый удачный вариант. Попробуйте разберитесь в них потом. Очевидно, что приходится затрачивать много времени и вычислительных ресурсов в хранилище с плохой структурой данных, с низким уровнем индексирования и классификации данных, с неясными типами и минимальными метаданными для поиска информации.

Для устранения риска снижения эффективности больших данных четко формулируются принципы упаковывания данных в хранилище и их структурирования. Сомнительные данные рекомендуется размещать обособлено.

Риск формирования неэффективного набора данных

Совокупность больших данных решает вполне конкретные цели и задачи, стоящие перед бизнесом. Бесконтрольный сбор (получение) и хранение данных могут привести к тому, что данные будут большими, четкими, удобными, но бесполезными по содержанию. Они могут быть неполными и не представлять полноценно фактическую сторону дела. На базе таких данных аналитики и менеджеры не смогут принять сколь-либо значимое решение.

Данные, а тем более большие, контролируются не только по форме, но и по содержанию, чтобы минимизировать риск формирования информационного набора неэффективного в целом или для решения отдельных поставленных задач. Допустимо рассматривать этот риск как несоответствие больших данных и бизнес-модели.

Риск ошибок больших данных

Несколько или одна примитивнейшая ошибка легко испортят долгую, кропотливую работу. Большие данные – не исключение. А учитывая, что объемы больших данных могут достигать огромных размеров – ошибки весьма вероятны как в содержании и структуре данных, так и в инструментах работы с данными.

Для снижения риска ошибок больших данных необходимо:

- проводить периодические ревизии данных (автоматизированные и выборочные);
- контролировать ключевые параметры данных;
- вести журнал выявленных ошибок и их устранения;
- разрабатывать инструменты и алгоритмы устранения или нивелирования ошибок и некорректных состояний данных;
- оценивать результативность инструментов;
- проводить независимую оценку и экспертизу;
- применять специальные средства тестирования данных и инструментов, разрабатываемых самостоятельно;
- использовать инструменты последовательно, подконтрольно и пошагово с постоянным контролем обрабатываемых данных в целом или по выборкам.

Риск ошибок бизнес-модели

В отличии от риска ошибок больших данных, риск ошибок бизнес-модели гораздо более серьезен и менее очевиден. Действительно утверждать об ошибке, допущенной в проектировании или понимании бизнес-модели может квалифицированный и опытный менеджер, знающий и понимающий бизнес. Ошибка или особенность бизнес-модели? В какой-то степени для ответа на этот вопрос и используются большие данные.

Риск экономической нецелесообразности

Не исключено, что аналитики не найдут ответы на проблемные вопросы бизнеса, обработав доступный им объем больших данных. Замена аналитиков, реформирование модели потоков больших данных, реструктуризация данных исправят как-то ситуацию в будущем. Однако затраты на проект произведены, а результат отсутствует.

Полностью избавиться от риска экономической нецелесообразности больших данных нельзя. Но минимизировать – реально, применяя:

- корректную постановку целей и задач проекта,
- стратегическое планирование проекта и его окружения,
- стратегию интегрирования больших данных в бизнес-модель,
- формирование профессиональной команды проекта,
- полноценное обеспечение проекта ресурсами,
- эффективное управление проектом,
- контроль за ходом проекта.

Риск внешнего консультанта

Большие данные – это сложный ресурс для бизнеса. Весьма вероятным является привлечение внешнего консультанта. Но это обуславливает соответствующий риск.

Внешней консультант помогает бизнесу, но остается вне поля его прямого воздействия. Хорошо прописанный контракт не спасает от разногласий и потери взаимного понимания. Полная передача на аутсорсинг работы с большими данными сторонней организации или заказ системы управления большими данными «под ключ» – это не очень разумный способ истратить денежные средства. Если бизнесу требуются большие данные, он обязан сам управлять ими. Конечно же, для среднего и малого бизнеса лучше искать разумный компромисс между внешним консультированием и собственными силами.

Бизнес всегда понимает и будет понимать о себе больше чем любой внешний консультант, который имеет также одну неприятную особенность – он в любой момент готов уйти и забрать с собой бесценные знания и опыт.

Риск неготовности к переменам

Может так оказаться, что большие данные и аналитика будут противоречить внутренней культуре компании и сложившемуся стилю руководства. Отсутствие в таком случае готовности к переменам сделает большие данные бесполезными. Придется от них отказаться, чтобы не тратить лишние средства, или оставить их в суррогатном виде для

создания видимости «информационно-инновационного современного развивающегося бизнеса».

Перед запуском проекта больших данных оцените готовность бизнеса к переменам, чтобы исключить или минимизировать риск их культурной несовместимости.

Риск мошенничества

Когда приходится сталкиваться с внешними консультантами или при создании команды проекта больших данных, существует вероятность столкнуться с банальным мошенничеством.

Особенно велик риск мошенничества при покупке больших данных «оптом и в розницу» или при подключении платных сервисов сбора и обработки больших данных. Проверить достоверность внешних данных или эффективность алгоритмов их обработки крайне сложно. Необходимо быть высоко квалифицированным и опытным специалистом, чтобы выявить подделанные или скомпрометированные данные. В самом деле, ну как для терабайтного массива цифровых данных провести полноценную экспертизу, да и сколько она будет стоить.

У мошенников много вариантов для формирования данных. Их можно специальным образом сгенерировать или имитировать, скрывая это за красивым фасадом «сверхчувствительного» алгоритма и «сверхумного» регистратора.

Качественные данные стоят не дешево и высок риск мошенничества с ними, поэтому и подходить к их покупке следует осторожно.

BIG DATA ILLUSION

Действительно ли большие данные – это объективная насущная проблема для бизнеса? Может быть это лишь красивый маркетинговый ход разработчиков мощных компьютеров и продуктов по хранению и обработке цифровых данных. Может быть это лишь привлекательная реклама консультантов по исследованию рынков и поведенческих моделей клиентов. А может это всего лишь модный тренд в сфере тотального наблюдения за субъектами рынка и прогнозирования их реакций. Возможно и нет никаких «больших» данных, а есть большая иллюзия о том, что удастся каким-то образом собрать огромный массив цифровой информации, обработать его неким волшебным образом и получить ответы на все вопросы, волнующие бизнесмена. Подводим итоги в последней части публикации и пытаемся определить основные ошибки, связанные с большими данных, а также перспективы развития рынка больших данных.

Иллюзия больших данных

Превалирующая иллюзия больших данных происходит от их названия. Кажется, имея big data, бизнес решает вопросы наивысшего порядка. И чем больше накопить данных, тем эффективней будут решаться всё более сложные вопросы.

Большие данные – это по сути ресурс аналитика. Это ресурс для людей, осуществляющих исследования и подготовку принятия решения. И как любой ресурс, большие данные без умения, знаний и технологий их использования не работают. Кто-то называет такое умение «добычей данных» (data mining) – по аналогии с добычей полезных ископаемых, делая акцент на глубоком проникновении и трудоемкости. Кто-то называет такое умение «интеллектом бизнеса» (business intelligent) – показывая насколько важным является «умственная» составляющая в этом процессе. Кому-то понравиться название «большая аналитика». Но известно, что даже наличие ресурса в большом количестве не означает его успешное и эффективное использование. Иногда избыточный объем ресурса позволяет строить бизнес-модель, не на глубокой его переработке в определенный набор продуктов, а на простом упаковывании и реализации в сыром виде. Зачем искать дополнительные варианты, когда, можно не прикладывая чрезмерных усилий, просто сбывать необработанный ресурс.

Большие данные, как информационная категория, имеют одну особенность в отличии от материальных ресурсов. Для их применения необходим высокий уровень

организации бизнес-объектов и бизнес-процессов компании. Без такого уровня подготовки, без наличия определенной квалификации у бизнеса, покупка (или сбор) больших данных будет отличаться низкой эффективностью. Настолько низкой, что не оправдает вложенные в них средства.

Зачем бизнесу тратить средства на большие данные, если не создан бизнес-слой управления (принятия решений) на основе аналитики? Абсолютно верно – незачем. К этому в той или иной степени приходят те компании, которые начали использование больших данных без принятия в контур управления аналитических технологий, техник принятия подготовленных аналитически решений и которые, по большому счету, не готовы к переменам. Такие субъекты рынка рано или поздно откажутся от больших данных. Особенно вопрос остро встанет при повышенной конкуренции за финансовые ресурсы внутри бизнеса.

Сегодня рынок больших данных сосредоточился на информационных технологиях. Это понятно и приятно, что развиваются инструменты работы с большими данными. Но интенсивный рост информационных сетей и совершенствование информационных технологий снимает барьеры по вычислительным мощностям. Это заставит передовые амбициозные бизнесы пересмотреть своё нынешнее увлечение и сместить акцент в сторону новых эффективных методик, инструментов, технологий менеджмента, базирующихся на знаниях и обучении.

Собственно, когда презентуют большие данные, часто речь идет о возможностях их хранения, транспортировки и обработки. Поисковые технологии гигантов сети Интернет яркий пример того, что бизнесу дают большие данные. Алгоритмы поиска – это мощнейшая обработка гигантских растущих объемов информации. Они постоянно находятся в процессе оптимизации, повышения производительности индексирования и структурирования информации. Но ведь за поисковыми технологиями в сети стоят не только большие данные. За ними стоят команды аналитиков, которые владеют высокотехнологичными знаниями в предметных областях. Поэтому разумное использование больших данных – это построение команды анализа данных, но никак не исключительное выстраивание серверов, облаков, систем добычи данных, машинного обучения и т.п.

Стоит заметить, что не очень показательно определение «добыча данных». Оно рисует несколько упрощенную ситуацию: есть «бесценные залежи» разнородных и перемешанных данных, а профессионал (или инструмент) берет и «раскапывает» в этих данных именно те, которые при «проникновенном» взгляде менеджера открывают ему глаза на всё происходящее и его вдруг осеняет праведная мысль о скрытых резервах бизнес-модели. Чудес не бывает и в больших данных тоже. Чтобы добыть ценную информацию из некоторого хранилища, её нужно туда сначала положить, потом извлечь, обработать и визуализировать. Очевидно, что акцент не корректно смещать на извлечение информации из хранилища, оставляя вне фокуса такие вещи как сбор (получение) данных, структурирование данных, упаковывание данных в хранилище, проверку качества данных, организацию процесса анализа данных, проблемы принятия решений на основе анализа больших данных и многое другое. Кроме того, даже для несложного data mining не помешает корректная постановка цели. Без грамотной постановки цели может выйти всё что угодно, а не осмысленный результат. Пусть эта цель выражена в виде гипотезы или вопросов, в виде проблемной ситуации или числовых показателей. Любые данные имеют контекст и метаданные, которые существенно ограничивают их использование в определенных ситуациях. Если условие контекста для задачи не задано, аналитик не в состоянии принять решение о соответствии данных поставленной задачи.

Не смотря на старания бизнеса сократить время от снятия информации о его состоянии до принятия решения об изменении такого состояния, существуют объективные причины непреодолимого временного лага. Задержка между принятием решения и изменением состояния бизнеса в соответствии с принятым решением может быть весьма существенной. Процессы и объекты перестраиваются, изменяется взаимодействие, корректируется поведения работников, подстраивается окружение. Поэтому любые данные и даже большие данные – это всегда данные о прошлом. Но руководство хочет

принимать на их базе решения для будущего. Здесь главное не переоценить возможности больших данных и аналитики.

Одно из заблуждений в отношении больших данных – это то, что они преимущественно внешние по отношению к бизнесу. Считается, что большие данные – это данные о клиентах (их поведении), данные о конкурентах, данные о разных факторах существования бизнеса (политические, социальные, культурные), данные о рынках и потребительских тенденциях, данные об активности других бизнесов. Частично это так. Но большие данные для бизнеса от внешних источников увязываются с данными о внутреннем состоянии, причем увязываются строго и контекстно. Это крайне необходимо, чтобы совместно оценивать самочувствие бизнес-модели и внешней среды. Внутренние данные также могут быть большими и весомыми для большой эффективной аналитики. Ведь ответ на вопрос что делать для исправления ситуации могут дать исключительно внутренние данные.

Ещё одна иллюзия которая способна помешать бизнесу – это то, что результативная аналитика основана только на больших данных. Существует реальная возможность и опыт, помноженный на талант некоторых аналитиков, предлагать решения в рамках традиционных объёмов внутренних данных, особенно когда речь идет о явных проблемах в бизнес-модели.

Отрицать огромное значение сбора и анализа больших данных для развития бизнеса невозможно. Особенно важны большие данные для распределённого и информационно-активного бизнеса. Пожалуй, большие данные – единственный эффективный инструмент быть в курсе всех дел для крупных корпораций и объединений с разветвленной сетью бизнес-единиц. Средний и малый бизнес также с учетом некоторых особенностей может оказаться в выигрыше от больших данных, особенно в кооперации с крупными компаниями и сообществами. Но нельзя подменять большими данными решение насущных проблем. Лучше рассматривать их как направление, которое поддерживает центральную стратегию бизнеса и позволяет быть в курсе произошедшего, происходящего и частично прогнозировать развитие ситуации в будущем. Но если у бизнеса нет вразумительной стратегии и если бизнес-модель видится примитивно и запутанно, то никакие большие данные не в состоянии помочь даже пассивному развитию. Некоторые менеджеры, понимая для себя отсутствие потребности в больших данных и не готовности к переменам, которые они сулят, не пытаются инициировать работу с ними – это тоже пример разумного поведения.

Как бы мы ни старались, большие данные не способны решить все проблемы. Никак нельзя с помощью большой аналитики построить эффективную бизнес-модель. Но они все-таки смогут помочь оптимизировать её в рамках выбранной стратегии.

«Волшебство» больших данных, которое несколько остается в стороне от общего внимания заключается в очевидном и обоснованном способе *размышлять о бизнесе* и искать пути его улучшения. Действительно проект больших данных улучшает бизнес и не столько из-за ценности каких-то массивов информации, а вследствие того, что менеджмент начинает смотреть на свою бизнес-модель с критической точки зрения, в том числе основываясь на некоторых информационных показателях и индикаторах. Если руководство вплотную интересуется большой аналитикой, то ему хочется понимать больше о своей компании и это – начало оптимизации бизнеса. Вместо больших данных можно выбрать иное средство развития бизнеса, например, маркетинговые исследования, статистические расчеты, экономико-математическое моделирование. Результат получится различным, но работа, нацеленная на «понимание» бизнес-модели, будет начата и несомненно даст положительный эффект. Если конечно она выполняется объективно, разумно, профессионально и с учетом воздействующих факторов.

Некоторые компании накопили ресурс – данные, а другие разработали мощные программные и аппаратные ИТ-решения. Этот ресурс и эти решения они постараются под тем или иным «маркетинговым соусом подать к столу бизнесменов» и заработать «хорошие чаевые». В ход пойдет активный сбыт и изощренный маркетинг, вежливые консультанты и веселые клиентские мероприятия богато «приправленные» красивым брендом и впечатляющей терминологией. Они будут говорить о построении надежнейших

систем обработки абсолютно не структурированных данных, о великолепных алгоритмах построения многоуровневых графов информации, о быстродействующих выборках на искусственном интеллекте, о самообучающихся нейросетевых механизмах. Не верьте на слово. Просите разъяснений, пояснений, демонстраций, документацию, независимые экспертные заключения, отзывы клиентов, нагрузочных тестов, пробного бесплатного периода.

Посудите сами, даже сам бренд «большие данные» выглядит выигрышно. Во-первых, в названии есть слово «большой», а значит это что-то хорошее, положительное, выгодное, впечатляющее, убедительное, ценное. Во-вторых, слово «данные», как бы указывает на что-то правильное, интеллектуальное, инновационное, эффективное, упорядоченное. Не поддавайтесь иллюзиям – большие данные не должны отрываться от реальности.

Не делайте это с большими данными

Категорически не рекомендуется делать следующие вещи с большими данными или с их помощью.

1. Не навязывайте товары и услуги, пусть и на основе анализа больших данных. Навязывать товары и услуги потребителю нельзя. Агрессивный сбыт характерен для низкокачественной или сомнительной продукции. Если вы абсолютно уверены и уверенность эта подкрепляется превосходной аналитикой больших данных, что конкретному потребителю нужен ваш товар (услуга) – не навязывайтесь. По крайней мере, сделайте ещё один шаг и обратитесь снова к большим данным – выясните как ваш потенциальный потребитель может быть проинформирован о конкурентных преимуществах вашего товара (услуги). Ни к чему хорошему не приведет назойливость типовых штампованных рекламных фраз. Зачем вообще тогда тратиться на большую аналитику, если вы и так решили добить вашего потребителя своей настойчивостью.

2. Не персонализируйте товар или услугу. Персонализация продуктов позволяет говорить об индивидуальном подходе и идеальном качестве товара или услуги для конкретного клиента (не путать с настройкой и подгонкой под размер). Большие данные помогут максимальным образом сформировать продукт под большую часть клиентов. Но с помощью больших данных не стоит излишне персонализировать продукт, тем более, что такая «персонализация» выглядит сомнительной. Клиент не особо доволен, когда вы предлагаете ему якобы созданный для него товар или услугу, если видит – они результат массового производства.

Обратите внимание, что попытка безусловно персонализировать коммерческое предложение – это индикатор проблем с качеством или другими потребительскими свойствами товара или услуги

3. Не «впадайте» в зависимость от активности на сайте. Собрать неограниченные объемы больших данных можно с помощью траффика и активности на корпоративном сайте (или, например, на официальной странице в социальной сети). Для этого много бесплатных инструментов. Но они будут говорить исключительно об удобстве интерфейса и не более. Строить далеко идущие выводы о том, что товар не нравиться потребителю используя мониторинг веб-страниц некорректно. Вы просто могли не так или не в тех цветах разместить информацию о товаре. Весьма полезно оптимизировать интерфейс на основе больших данных по активности в сети, но не более.

4. Не надо зависеть от индивидуальных предпочтений пользователя. Продукт не должен зависеть от индивидуальных предпочтений или особенностей поведения пользователя, даже если вы точно знаете, чего он хочет, проанализировав о нем все большие данные. Не получиться. Стандартизацию никто не отменял. Но никто не отменял и настройку пользователем продукта под свои предпочтения.

5. Не собирайте все данные обо всем. Самые мощные и емкие серверы данных – это не причина для тотального сбора и накопления чрезмерно больших данных. Накопление данных не является самоцелью для бизнеса, тем более, что на это приходится тратить ресурсы и порой не малые. Собирать данные «на всякий случай» тоже ошибочно,

если вы не в состоянии хотя бы примерно объяснить такой «всякий случай» с позиции бизнес-модели.

6. Не ждите, что большие данные сформулируют бизнес-модель. Большие данные не способны описать или формализовать её. Они на это не способны. Их задача в том, чтобы оценить существующий бизнес и скорректировать его. Как правило, большие данные собираются у тех источников, которые не выражают модель бизнеса или выражают её сильно опосредовано.

7. Не подменяйте будущее прогнозами на базе больших данных. Хочется заглянуть в будущее и понять, как поведет себя бизнес или рынок, потребитель или конкурент через неделю, через месяц, через год. Аналитики пытаются на основе обработки данных и расчетов строить прогнозные модели. Но зависимость от таких моделей, пусть весьма совершенных и изощрённых, приведут менеджмент к пассивности. Действительно, зачем предпринимать усилия в области продаж, если аналитик спрогнозировал их рост в связи с хорошими темпами роста рынка. А затем, чтобы не оказаться перед неожиданным схлопыванием доли рынка или целевых его сегментов. Нельзя с помощью больших данных узнать будущее, но можно понять куда направить усилия в будущем, чтобы сохранить конкурентные преимущества или уверенно нарастить их.

8. Не оставляйте без проверки большие данные. Они нуждаются в контроле. Особенно это касается приобретаемых на стороне массивов информации. Они покупаются при предоставлении убедительного доказательства подлинности и с учетом полного набора характеристик качества и обстоятельств сбора.

Непроверенные и фальсифицированные большие данные могут стать проблемой и даже навредить бизнесу. Кроме того, не исключен преднамеренный вброс фальсифицированных данных. Алгоритмы сбора и обработки больших данных могут допускать ошибки и поддаваться на скрытые их провокации. И подобные ошибки сложно обнаружить в больших и плохо структурированных данных.

9. Не надейтесь, что большие данные спасут бизнес или решат проблемы. Они предоставляют способ выяснить насколько сложна текущая ситуация и что к ней привело. Это лишь способ повысить эффективность бизнеса, но в комплекте с рациональным менеджментом. Как правило, нужен сильный и ответственный менеджер, чтобы понять и принять решения подготовленные большой аналитикой.

10. Большие данные могут навредить бизнесу, если попадут в руки конкурентов, поэтому не отдавайте контроль над большими данными. Они требуют пристального внимания и защиты. Допускается привлекать специалистов, экспертов, консультантов, внешние сервисы. Но не допускается полностью передавать на сторону управление большими данными или передавать сторонним подрядчикам стратегические и ключевые функции больших данных, так же как не допускается оставлять данные без надлежащей защиты.

11. Не подменяйте большими данными разумное поведение. Самое главное, чего не делайте с большими данными – это подменять ими другие не менее важные функции профессионального менеджмента, в том числе:

- экспертное принятие решений;
- стратегическое и тактическое планирование;
- исследование рынков и выявление потребностей клиентов;
- взаимодействие с потребителями;
- менеджмент качества;
- построение эффективной бизнес-команды;
- объективность и критичность в отношении результатов работы;
- бизнес-моделирование;
- и конечно же, здравый смысл.

Развитие больших данных

Большие данные дают большое конкурентное преимущество. И это утверждение верно, если их сбор, обработка и анализ сопровождаются соответствующей грамотной стратегией и готовностью бизнеса к переменам.

Сегодня большие данные доступны крупным и информационно-обеспеченным компаниям. Но завтра доступ к ним, с помощью тех или иных инструментов, получит средний и малый бизнес. Информационное развитие инфраструктуры бизнеса идёт впечатляющими темпами. И не корректно утверждать, что большие данные пригодны для массовых рынков сбыта и масштабных производств. Даже узкоспециализированные предприятия и сервисные компании способны получить от больших данных серьезный потенциал для развития.

Учитывая усиление специализации по отдельным компетенциям, связанным с большими данными, уместно говорить, что их будущее – это сервисный разделяемый функционал сторонних консультантов, а не тотальные процессы внутри одной компании. Оптимальной является модель, когда при общей внутренней стратегии управления большими данными, бизнес поручает внешним профессионалам узкие направления работ. Например, сбор конкретных данных и представление их заказчику в установленном виде. Вот примерно также как сейчас средства веб-аналитики, слабо понимая миссию и стратегию сайта, не интересуясь досконально его владельцами, авторами и выгодоприобретателями, собирают и предоставляют данные о количестве и качестве посетителей.

У больших данных есть все шансы со временем превратиться в индустрию большой аналитики. Сегодня в приоритете ИТ-инфраструктура сбора и обработки громадных массивов информации. Но по мере дальнейшего их совершенствования, всё очевидней и острее встают вопросы грамотной обработки данных и генерирования на их основе объективных и релевантных решений. Большая аналитика много сложней, чем просто большие данные. Но только при достаточном уровне проникновения последних в бизнес-среду и в профессиональные компетенции менеджмента, сформируется четкая потребность в анализе высокого уровня.

Большая аналитика должна быть обеспечена серьезными и удобными инструментами, как программными, так и непосредственно аналитическими. Очевидно, что увеличится потребность в квалифицированных кадрах. Но если для информационно-технологического развития сервисов больших данных можно привлекать подготовленных специалистов ИТ-сферы, то для большой аналитики потребуются специально подготовленные профессионалы. Они совмещают, в определенном ключе, знания и опыт информационных технологий со знаниями и опытом предметных областей. Не смотря на серьезные подвижки в области машинной обработки информации, до настоящего времени и в перспективе, обойтись без человека не получиться. Крайне необходимы специалисты, которые будут интенсивно исследовать данные и которые смогут формулировать задачи понятные с точки зрения алгоритма анализа. Поиск и устранение ошибок в данных – очевидная и насущная проблема, которую решают такие профессионалы. А вот что им предложит ИТ-рынок в качестве инструментария – посмотрим.

Несомненно, большие данные сформируют различные рынки: от тех на которых продают данные лотами разного объёма и качества, до тех на которых предоставляются высокотехнологичные сервисы с машинным временем суперкомпьютеров.

Переход к сбору и обработке информации в объемах, превышающих традиционные, может стать хорошим поводом для специализированного или широкого реинжиниринга бизнес-процессов (и вовлеченных в них бизнес-объектов). При этом придётся признать приоритет за моделью интегрирования больших данных в бизнес-модель по всей структуре и всем направлениям.

Не стоит забывать об этической стороне вопроса больших данных. Обезличить данные о клиентах, потребителях и других субъектах рынка нереально – это лишит их ценности. Но большие данные в том или ином виде, а особенно поступившие от внутренних источников, содержат персональные или атрибутивные личностные данные. Требуется серьезная защита от несанкционированного распространения больших данных. Да и клиенты не довольны, когда ведется тотальный сбор информации о них. Увещевания, что делается это для их же пользы не помогает. Никто не хочет, чтобы за ним следили. И если он не найдет в себе достаточной мотивации передавать кому-то данные о себе, то и

не станет этого делать. Нормальное и разумное поведение, которое приходится учитывать.

На законодательном уровне будут вводится всё более жесткие ограничения. И совершенно понятно, что эти ограничения будут вступать в конфликт с информационными потребностями бизнеса. Во что превратятся big data – в эффективный инструмент повышения комфорта жизни или в наводящую ужас слежку за всем и вся, удастся ли сформировать «вселенную полезной информации» или будет выстроена «бездна темных материалов».

Заглянуть в будущее. Получить ответы на разные вопросы касающиеся развития бизнеса, рынков, предпочтений клиентов, конъюнктурных факторов. Так или иначе профессионалы пытаются смоделировать ситуацию в будущем и извлечь адекватную информацию о возможностях и угрозах. Большие данные – один из инструментов для этого, весьма удобный и результативный. Но всё слишком непросто с предсказаниями и предсказателями, как в плане реализуемости прогнозов, так и в социально-этическом аспекте. Не случайно проблемами предвидения задаются не только экономисты, математики, физики, но и философы, писатели, для которых будущее – это не практическая выгода, а сложная и хрупкая сеть человеческих достоинств и пороков полная разумных сомнений и моральных противоречий. Повесть «Minority Report» («Особое мнение») известного писателя-фантаста Филипа Дика предлагает нам подумать на эту тему и решить в какой мере можно доверять и слепо следовать убедительным аналитическим прогнозам. Есть ли в таких ситуациях место для личного мнения и решения. И принимать ли во внимание особое мнение одного профессионала, пусть и сильно отличающееся от мнений большинства других уважаемых экспертов.

P.S. BAD DATA

В настоящей публикации много терминов из мира больших данных. Хочется в конце добавить ещё одно понятие.

Big-bad-data-driven (сокращенно: bbdd) – «плохо работающий на основе больших данных» – это нечто неправильно сделанное с использованием больших данных.

В качестве примера bbdd-сервиса приведу «как бы» таргетированную рекламу в сети Интернет. Казалось бы, для бизнеса, который пытается представить себя через глобальную сеть, есть очевидные преимущества персонализированного предложения (рекламы) клиенту. Но они сведены на нет в результате нарушения элементарных правил больших данных. Почему подобные bbdd-сервисы пытаются на базе неполноценного набора информации из прошлого, а порой из весьма далекого прошлого, предложить какой-то товар или услугу сегодня. Таргетирование делается как-то неумело и посредственно. Зачастую само рекламное сообщение предстает в низком качестве.

Все недостатки и ошибки обработки больших данных мы наблюдаем в изящно подбираемых для нас рекламных ссылках в браузерах, социальных сетях, условно-бесплатных сервисах. Если клиент активно искал когда-то и что-то в сети – это не значит, что он обязательно хочет это купить и уж совсем это не значит, что он исключительно это и хочет купить. Сколько раз можно показывать одну рекламу об одном и том же.

Гораздо менее раздражительно и более продуктивно показывать случайную рекламу. Так называемый рекламный шум – он не так раздражает и навязывается, но бывает весьма результативным. К сожалению, у простой не нацеленной рекламы ниже цена реализации. У неё нет «конкурентного преимущества», которое расхваливают агенты по продажам сетей «умной» рекламы. Персонализируемая реклама даже на основе сверхбольших данных никогда не вычислит будущие потребности кого-то и уж тем более не раскроет его истинные потребности. Да кто вообще знает свои потребности.