



ЧТО ТАКОЕ BIG DATA (БОЛЬШИЕ ДАННЫЕ) В МАРКЕТИНГЕ: ПРОБЛЕМЫ, АЛГОРИТМЫ, МЕТОДЫ АНАЛИЗА

Гараджаева Сульгун Атаевна,

Старший преподаватель Туркменского государственного университета имени Махтумкули
г. Ашхабад Туркменистан

Оразгелдиев Абдымухаммет Ходжамбердиевич

Преподаватель института Телекоммуникаций и информатики г. Ашхабад Туркменистан

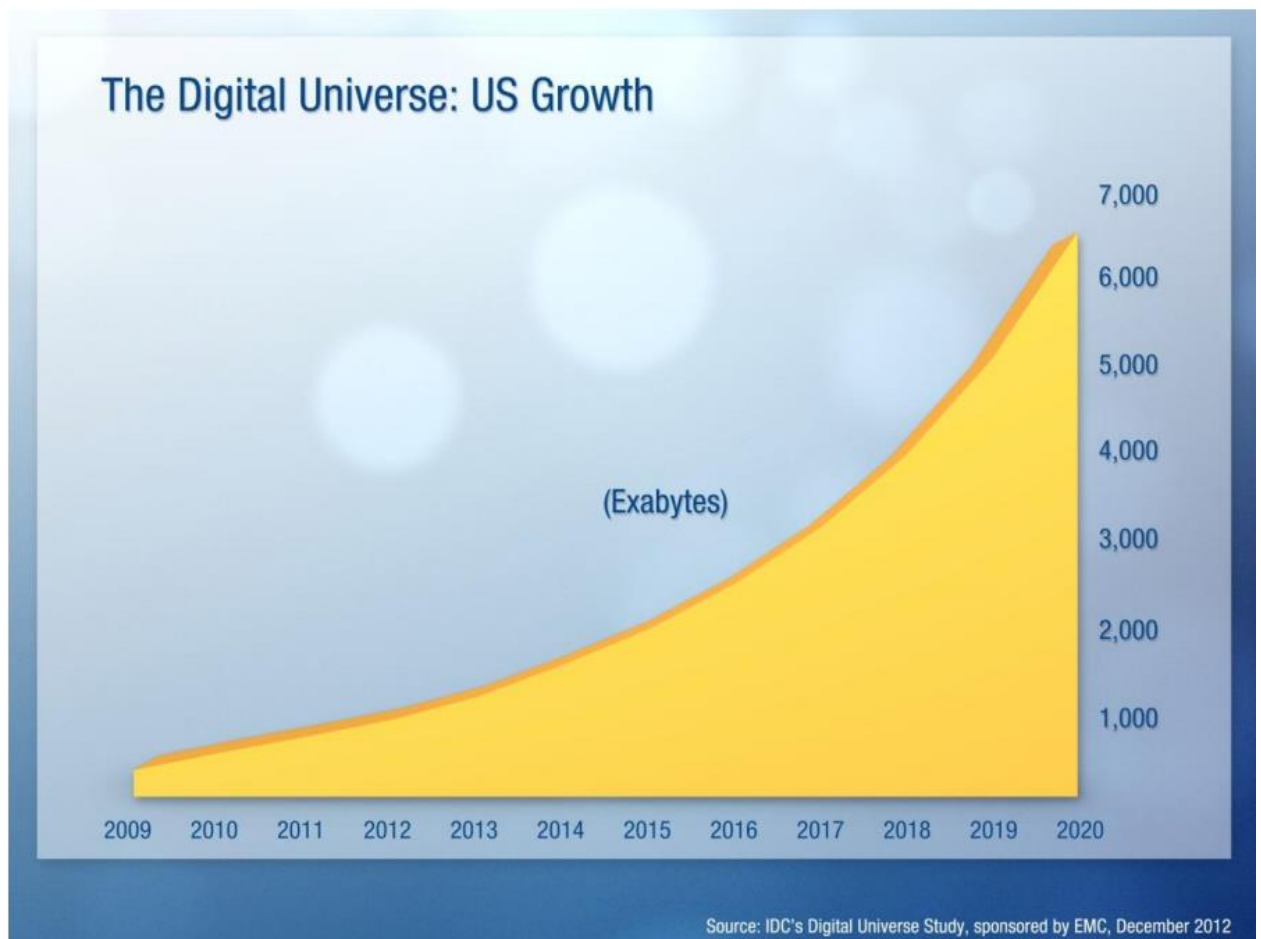
Большие данные (Big Data) — довольно распространенное понятие в IT и интернет-маркетинге. По сути, определение термина лежит на поверхности: «большие данные» означает управление и анализ очень больших объемов данных. Если смотреть шире, то это информация, которая не поддается обработке классическими способами по причине больших объемов.

Big Data — что это такое?

Цифровые технологии присутствуют во всех областях жизни человека. Объем записываемых в мировые хранилища данных каждую секунду растет, а это означает, что такими же темпами должны изменяться условия хранения информации и появляться новые возможности для наращивания ее объема.

Эксперты в области IT высказывают мнение, что расширение Big Data и ускорение темпа роста стало объективной реальностью. Каждую секунду гигантские объемы контента генерируют такие источники, как социальные сети, информационные сайты, файлообменники — и это лишь сотая часть поставщиков.

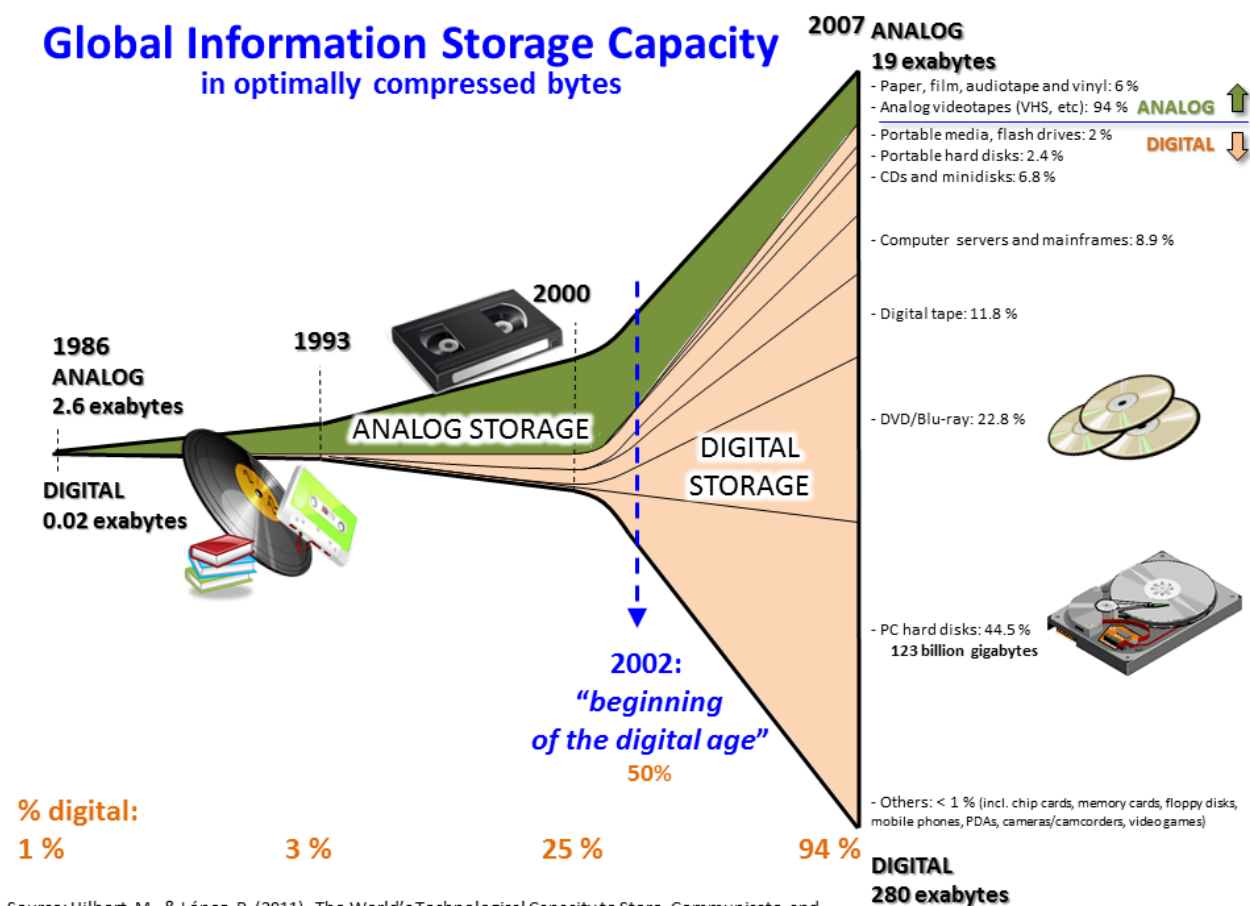
Согласно исследованию IDC Digital Universe, в ближайшие пять лет объем данных на планете вырастет до 40 зеттабайтов, то есть к 2020 году на каждого живущего на Земле человека будет приходиться по 5200 Гб.



Известно, что основной поток информации генерируют не люди. Источником служат роботы, находящиеся в постоянном взаимодействии друг с другом. Это приборы для мониторинга, сенсоры, системы наблюдения, операционные системы персональных устройств, смартфоны, интеллектуальные системы, датчики и прочее. Все они задают бешеный темп роста объема данных, что приводит к появлению потребности наращивать количество рабочих серверов (и реальных, и виртуальных) — как следствие, расширять и внедрять новые data-центры.

По сути, большие данные — довольно условное и относительное понятие. Самое распространенное его определение — это набор информации, по объему превосходящей жесткий диск одного персонального устройства и не поддающейся обработке классическими инструментами, применяемыми для меньших объемов.

Global Information Storage Capacity in optimally compressed bytes



Source: Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60–65. <http://www.martinhilbert.net/WorldInfoCapacity.html>

Технология Big Data — что это? Обобщенно говоря, технологию обработки больших данных можно свести к трем основным направлениям, решающим три типа задач:

- Хранение и перевод поступающей информации в гигабайты, терабайты и зеттабайты для их хранения, обработки и практического применения.
- Структурирование разрозненного контента: текстов, фотографий, видео, аудио и всех иных видов данных.
- Анализ Big Data и внедрение различных способов обработки неструктурированной информации, создание различных аналитических отчетов.

В сущности, применение Big Data подразумевает все направления работы с огромным объемом самой разрозненной информации, постоянно обновляемой и разбросанной по разным источникам. Цель предельно проста — максимальная эффективность работы, внедрение новых продуктов и рост конкурентоспособности.

Проблема Big Data

Проблемы системы Big Data можно свести к трем основным группам: объем, скорость обработки, неструктурированность. Это три V — Volume, Velocity и Variety.

Хранение больших объемов информации требует специальных условий, и это вопрос пространства и возможностей. Скорость связана не только с возможным замедлением и «торможением», вызываемом старыми методами обработок, это еще и вопрос интерактивности: чем быстрее процесс, тем больше отдача, тем продуктивнее результат.

Проблема неоднородности и неструктурированности возникает по причине разрозненности источников, форматов и качества. Чтобы объединить данные и эффективно их обрабатывать, требуется не только работа по приведению их в пригодный для работы вид, но и определенные аналитические инструменты (системы).

Но это еще не все. Существует проблема предела «величины» данных. Ее трудно установить, а значит трудно предугадать, какие технологии и сколько финансовых вливаний потребуются для дальнейших разработок. Однако для конкретных объемов данных (терабайт, к примеру) уже применяются действующие инструменты обработки, которые к тому же и активно развиваются.

Существует проблема, связанная с отсутствием четких принципов работы с таким объемом данных. Неоднородность потоков только усугубляет ситуацию. Каким образом подходить к их применимости, чтобы вынести из них что-то ценное? Здесь требуется разработка такого направления, как новые методы анализа Big Data, чтобы этот поток стал полезным источником информации. Возможно, согласно утверждениям представителей университетов США (Нью-Йоркского, Вашингтонского и Калифорнийского), сегодня пришло время ввести и развивать новую дисциплину — науку о Big Data.

Собственно, это и является главной причиной отсрочки внедрения в компании проектов Big Data (если не брать во внимание еще один фактор — довольно высокую стоимость).

Подбор данных для обработки и алгоритм анализа может стать не меньшей проблемой, так как отсутствует понимание, какие данные следует собирать и хранить, а какие можно игнорировать. Становится очевидной еще одна «болевая точка» отрасли — нехватка профессиональных специалистов, которым можно было бы доверить глубокий анализ, создание отчетов для решения бизнес-задач и как следствие извлечение прибыли (возврат инвестиций) из Big Data.

Еще одна проблема Big Data носит этический характер. А именно: чем сбор данных (особенно без ведома пользователя) отличается от нарушения границ частной жизни?

Так, информация, сохраняемая в поисковых системах Google и Яндекс, позволяет им постоянно дорабатывать свои сервисы, делать их удобными для пользователей и создавать новые интерактивные программы.

Поисковики записывают каждый клик пользователя в Интернете, им известен его IP-адрес, геолокация, интересы, онлайн-покупки, личные данные, почтовые сообщения и прочее, что, к примеру, позволяет демонстрировать контекстную рекламу в соответствии с поведением пользователя в Интернете. При этом согласия на это не спрашивается, а возможности выбора, какие сведения о себе предоставлять, не дается. То есть по умолчанию в Big Data собирается все, что затем будет храниться на серверах данных сайтов.

Здесь можно затронуть другую проблему — обеспечение безопасности хранения и использования данных. Например, сведения о возможных покупателях и их история переходов на сайтах интернет-магазинов однозначно применимы для решения многих бизнес-задач. Но безопасна ли аналитическая платформа, которой потребители в автоматическом режиме (просто потому, что зашли на сайт) передают свои данные, — это вызывает множество споров. Современную вирусную активность и хакерские атаки не сдерживают даже супер-защищенные серверы правительственных спецслужб.

История больших данных

Сами по себе алгоритмы Big Data возникли при внедрении первых высокопроизводительных серверов (мэйнфреймов), обладающих достаточными ресурсами для оперативной обработки информации и пригодных для компьютерных вычислений и для дальнейшего анализа..

Сам термин Big Data впервые был озвучен в 2008 году на страницах спецвыпуска журнала Nature в статье главного редактора Клиффорда Линча. Этот номер издания был посвящен взрывному росту глобальных объемов данных и их роли в науке.

Специалисты утверждают, что большими данными допустимо называть любые потоки информации объемом более 100 Гб в сутки.

Однако в последние 2-3 года ученые отмечают, что термин Big Data стал лишком популяризирован, его употребляют практически везде, где упоминаются потоки данных, и как следствие он стал восприниматься слишком обобщенно и размыто. Виной тому не совсем сведущие журналисты и малоопытные предприниматели, которые попусту злоупотребляют данным понятием. По мнению западных экспертов, термин давно дискредитировал себя и пришло время от него отказаться.

Сегодня мировое сообщество вновь заговорило о больших данных. Причины — в неизменном росте объемов информации и отсутствии какой-либо структуры в ней. Предпринимателей и ученых волнуют вопросы качественной интерпретации данных, разработки инструментов для работы с ними и развитие технологий хранения. Этому способствует внедрение и активное использованию облачных моделей хранения и вычислений.

Big Data в маркетинге

Информация — это главный аспект успешного прогнозирования роста и составления маркетинговой стратегии в умелых руках маркетолога. Анализ больших данных давно и успешно применяется для определения: целевой аудитории, интересов, спроса, активности потребителей. Таким образом, Big Data является точнейшим инструментом маркетолога для предсказания будущего компании.

К примеру, анализ больших данных позволяет выводить рекламу (на основе известной модели RTB-аукциона — Real Time Bidding) только тем потребителям, которые заинтересованы в товаре или услуге.

Применение Big Data в маркетинге позволяет бизнесменам:

- лучше узнавать своих потребителей, привлекать аналогичную аудиторию в Интернете;
- оценивать уровень удовлетворенности клиентов;
- понимать, соответствует ли предлагаемый сервис ожиданиям и потребностям;
- находить и внедрять новые способы, увеличивающие доверие клиентов;
- создавать проекты, пользующиеся спросом.

Например, сервис Google.trends очень точно укажет маркетологу прогноз сезонной активности спроса на конкретный продукт, колебания и географию кликов. Достаточно сопоставить эти сведения со статистическими данными собственного сайта и можно составить качественный план по распределению рекламного бюджета с указанием месяца и региона.

Заключения

Сегодня, в пик высоких технологий и огромных потоков информации, у компаний появилось гораздо больше возможностей для достижения превосходных показателей в ведении бизнеса благодаря использованию Big Data.

Высоких вам конверсий!