

ПСИХОЛОГИЯ ИНТЕРНЕТА

Big Data, или большие данные





Постоянное ускорение роста объема данных является неотъемлемым элементом современных реалий. Социальные сети, мобильные устройства, данные с измерительных устройств, бизнес-информация – это лишь несколько видов источников, способных генерировать гигантские массивы данных.

В настоящее время термин Big Data (большие данные) стал довольно распространенным. Далеко не все еще осознают то, насколько быстро и глубоко технологии обработки больших массивов данных меняют самые различные аспекты жизни общества. Перемены происходят во всех сферах, порождая новые проблемы и вызовы, в том числе и в сфере информационной безопасности, где на первом плане должны находиться такие важнейшие ее аспекты, как конфиденциальность, целостность, доступность и т. д.

По сути, определение термина лежит на поверхности:

Большие данные – это не просто много данных. Это неструктурированные данные, разного формата: комментарии всех пользователей, все видео, вся музыка, потоковая съемка видеонаблюдения, сообщения в мессенджерах и другое.

«Большие данные» означают управление очень большими объемами данных, а также их анализ. Если смотреть шире, то это информация, которая не поддается обработке классическими способами по причине огромных объемов. В 2010 году уже стали появляться первые продукты и решения, непосредственно связанные с обработкой больших данных. К 2011 году большинство крупнейших IT-компаний, включая IBM, Oracle, Microsoft и Hewlett-Packard, активно используют термин Big Data в своих деловых стратегиях. Постепенно аналитики рынка информационных технологий начинают активные исследования данной концепции.

В настоящее время этот термин приобрел значительную популярность и активно используется в самых различных сферах. Однако нельзя с уверенностью сказать, что Big Data – это какое-то принципиально новое явление. Напротив, большие источники данных существуют уже много лет. В маркетинге ими можно назвать базы данных по покупкам клиентов, кредитным историям, образу жизни и т. д. На протяжении многих лет аналитики использовали эти данные, чтобы помогать компаниям прогнозировать будущие потребности клиентов, оценивать риски, формировать потребительские предпочтения и т. д.

Сегодня ситуация изменилась в двух аспектах:

- Появились более сложные инструменты и методы для анализа и сопоставления различных наборов данных;
- Инструменты анализа дополнились множеством новых источников данных, что обусловлено повсеместным переходом на цифровые технологии, а также новыми методами сбора и измерения данных.

Исследователи прогнозируют, что технологии Big Data активнее всего будут использоваться в производстве, здравоохранении, торговле, госуправлении и в других сферах и отраслях.

Big Data – это не какой-либо определенный массив данных, а совокупность методов их обработки. Определяющей характеристикой для больших данных является не только их объем, но также и другие категории, характеризующие трудоемкие процессы обработки и анализа данных.

В качестве исходных данных для обработки могут выступать, например:

- Логи поведения интернет-пользователей.
- Интернет вещей.
- Социальные медиа.
- Метеорологические данные.
- Оцифрованные книги крупнейших библиотек.
- GPS-сигналы из транспортных средств.
- Информация о транзакциях клиентов банков.
- Данные о местонахождении абонентов мобильных сетей.
- Информация о покупках в крупных ритейл-сетях и т. д.

Со временем объемы данных и количество их источников непрерывно растет, а на этом фоне появляются новые и совершенствуются уже имеющиеся методы обработки информации.



Основные принципы Big Data:

- Горизонтальная масштабируемость – массивы данных могут быть огромными, и это значит, что система обработки больших данных должна динамично расширяться при увеличении их объемов.
- Отказоустойчивость – даже при сбое некоторых элементов оборудования, вся система должна оставаться работоспособной.
- Локальность данных. В больших распределенных системах данные обычно распределяются по значительному числу машин.

Однако по мере возможности и в целях экономии ресурсов данные часто обрабатываются на том же сервере, что и хранятся.

Для стабильной работы всех трех принципов и, соответственно, высокой эффективности хранения и обработки больших данных необходимы новые прорывные технологии, такие как, например, блокчейн.

Для чего нужны большие данные?

Сфера применения Big Data постоянно расширяется:

- Большие данные можно использовать в медицине. Так, устанавливать диагноз пациенту можно не только опираясь на данные анализа истории болезни, но также принимая во внимание опыт других врачей, сведения об экологической ситуации района проживания больного и многие другие факторы.
- Технологии Big Data могут использоваться для организации движения беспилотного транспорта.
- Обработывая большие массивы данных, можно распознавать лица на фото- и видеоматериалах.
- Технологии Big Data могут быть использованы ритейлерами – торговые компании имеют возможность активно использовать массивы данных из социальных сетей для эффективной настройки своих рекламных кампаний, которые могут быть максимально ориентированы под тот или иной потребительский сегмент.
- Данная технология активно используется при организации предвыборных кампаний, в том числе для анализа политических предпочтений в обществе.
- Использование технологий Big Data актуально для решений класса гарантирования доходов, которые включают в себя инструменты обнаружения несоответствий и углубленного анализа данных, позволяющие своевременно выявить вероятные потери либо искажение информации, способные привести к снижению финансовых результатов.
- Телекоммуникационные провайдеры агрегируют большие данные, в том числе о геолокации. В свою очередь, эта информация представляет коммерческий интерес для рекламных агентств, которые могут использовать ее для показа таргетированной и локальной рекламы, а также для ритейлеров и банков.
- Большие данные играют важную роль при решении открыть торговую точку в определенной локации на основе данных о наличии мощного целевого потока людей.

Таким образом, наиболее очевидное практическое применение технологии Big Data лежит в сфере маркетинга. Благодаря развитию интернета и распространению всевозможных коммуникационных устройств поведенческие данные становятся доступными в режиме реального времени.

Технологии больших данных могут также эффективно использоваться в финансах, для социологических исследований и во многих других сферах. Эксперты утверждают, что все эти возможности использования больших данных являются лишь видимой частью айсберга, поскольку в гораздо больших объемах эти технологии используются в разведке и контрразведке, в военном деле, а также во всем том, что принято называть информационными войнами.



В общих чертах последовательность работы с Big Data состоит из сбора данных, структурирования полученной информации с помощью отчетов и дашбордов, а также последующего формулирования рекомендаций к действию.

Рассмотрим вкратце возможности использования технологий Big Data в маркетинге. Как известно, для маркетолога информация – главный инструмент при прогнозировании и составлении стратегии. Анализ больших данных давно и успешно применяется для определения целевой аудитории, интересов, спроса и активности потребителей. А также, в частности, позволяет выводить рекламу только тем потребителям, которые заинтересованы в товаре или услуге.

Помимо высокой стоимости, одним из главных факторов, тормозящих внедрение Big Data в различные сферы, является проблема выбора обрабатываемых данных, то есть определения того, какие данные необходимо извлекать, хранить и анализировать, а какие – не принимать во внимание.

Еще одна проблема Big Data носит этический характер. Другими словами, возникает закономерный вопрос: можно ли подобный сбор данных (особенно без ведома пользователя) считать нарушением границ частной жизни? Не секрет, что информация, сохраняемая в поисковых системах Google и Яндекс, позволяет IT-гигантам постоянно дорабатывать свои сервисы, делать их удобными для пользователей и создавать новые интерактивные приложения. Для этого поисковики собирают пользовательские данные об активности пользователей в интернете, IP-адреса, данные о геолокации, интересах и онлайн-покупках, личные данные, почтовые сообщения и т. д. Все это позволяет демонстрировать контекстную рекламу в соответствии с поведением пользователя в интернете. При этом обычно согласия пользователей на это никто не спрашивает, а возможности выбора, какие сведения о себе предоставлять, не дает. То есть по умолчанию в Big Data собирается все, что затем будет храниться на серверах данных сайтов.

Из этого вытекает следующая важная проблема, касающаяся обеспечения безопасности хранения и использования данных. Например, безопасна ли та или иная аналитическая платформа, которой потребители в автоматическом режиме передают свои данные? Кроме того, многие представители бизнеса отмечают дефицит высококвалифицированных аналитиков и маркетологов, способных эффективно оперировать большими объемами данных и решать с их помощью конкретные бизнес-задачи.

Несмотря на все сложности с внедрением Big Data, бизнес намерен увеличивать вложения в это направление. По данным исследования Gartner, лидерами инвестирующих в Big Data отраслей являются медиа, ритейл, телеком, банковский сектор и сервисные компании.

Применение Big Data в маркетинге позволяет бизнесменам:

- Лучше узнавать своих потребителей, привлекать аналогичную аудиторию в интернете.
- Оценивать степень удовлетворенности клиентов.
- Понимать, соответствует ли предлагаемый сервис ожиданиям и потребностям.
- Находить и внедрять новые способы, увеличивающие доверие клиентов.
- Создавать проекты, пользующиеся спросом и т. д.

Например, сервис Google.trends может указать маркетологу прогноз сезонной активности спроса на конкретный продукт, колебания и географию кликов. Если сопоставить эти сведения со статистическими данными, собираемыми соответствующим плагином на собственном сайте, то можно составить план по распределению рекламного бюджета с указанием месяца, региона и других параметров.

Существует множество комбинаций программного и аппаратного обеспечения, которые позволяют создавать эффективные решения Big Data для различных бизнес-дисциплин: от социальных медиа и мобильных приложений до интеллектуального анализа и визуализации коммерческих данных.

Важное достоинство Big Data – это совместимость новых инструментов с широко используемыми в бизнесе базами данных, что особенно важно при работе с кросс-дисциплинарными проектами, например, такими, как организация мультимедийных продаж и поддержки покупателей.



Последовательность работы с Big Data состоит из сбора данных, структурирования полученной информации с помощью отчетов и дашбордов, создания инсайтов и контекстов, а также формулирования рекомендаций к действию.

Так как работа с Big Data подразумевает большие затраты на сбор данных, результат обработки которых заранее неизвестен, основной задачей является четкое понимание, для чего нужны данные, а не то, как много их есть в наличии. В этом случае сбор данных превращается в процесс получения исключительно нужной для решения конкретных задач информации.

Например, у телекоммуникационных провайдеров агрегируется огромное количество данных, в том числе о геолокации, которые постоянно пополняются. Эта информация может представлять коммерческий интерес для рекламных агентств, которые могут использовать ее для показа таргетированной и локальной рекламы, а также для ритейлеров и банков.

Есть пример измерения эффективности рекламы на outdoor-щитах в Лондоне. Сейчас охват подобной рекламы можно измерить, лишь поставив возле рекламных конструкций людей со специальным устройством, подсчитывающим прохожих. По сравнению с таким видом измерения эффективности рекламы, у мобильного оператора куда больше возможностей – он точно знает местонахождение своих абонентов, ему известны их демографические характеристики, пол, возраст, семейное положение и т. д.

На основе таких данных в будущем открывается перспектива менять содержание рекламного сообщения, используя предпочтения конкретного человека, проходящего мимо рекламного щита. Если данные показывают, что проходящий мимо человек много путешествует, то ему можно будет показать рекламу курорта. Организаторы футбольного матча могут оценить количество болельщиков только, когда те придут на матч. Но если бы они имели возможность запросить у оператора сотовой связи информацию, где посетители находились за час, день или месяц до матча, то это дало бы организаторам возможность планировать места для размещения рекламы следующих матчей.

Другой пример – как банки могут использовать Big Data для предотвращения мошенничества. Если клиент заявляет об утере карты, а при совершении покупки с ее помощью банк видит в режиме реального времени месторасположение телефона клиента в зоне покупки, где происходит транзакция, банк может проверить информацию по заявлению клиента, не пытался ли он обмануть его. Либо противоположная ситуация, когда клиент совершает покупку в магазине, банк видит, что карта, по которой происходит транзакция, и телефон клиента находятся в одном месте, банк может сделать вывод, что картой пользуется ее владелец. Благодаря подобным преимуществам Big Data расширяются границы, которыми наделены традиционные хранилища данных.

Для успешного принятия решения о внедрении решений Big Data компании необходимо рассчитать инвестиционный кейс, и это вызывает большие трудности из-за множества неизвестных составляющих. Парадоксом аналитики в подобных случаях становится прогнозирование будущего на основе прошлого, данные о котором зачастую отсутствуют. В этом случае важным фактором является четкое планирование своих первоначальных действий.

Во-первых, необходимо определить одну конкретную задачу бизнеса, для решения которой будут использоваться технологии Big Data. Эта задача станет стержнем определения верности выбранной концепции. Нужно сосредоточиться на сборе данных, связанных именно с этой задачей, а в ходе проверки концепции вы сможете использовать различные инструменты, процессы и методы управления, которые позволят принимать более обоснованные решения в будущем.

Во-вторых, маловероятно, что компания без навыков и опыта аналитики данных сможет успешно реализовать проект Big Data. Необходимые знания всегда вытекают из предыдущего опыта аналитики, что является основным фактором, влияющим на качество работы с данными. Важную роль играет культура использования данных, так как часто анализ информации открывает суровую правду о бизнесе, и чтобы принять эту правду и работать с ней, необходимы выработанные методы работы с данными.

В-третьих, ценность технологий Big Data заключается в предоставлении инсайтов. Хорошие аналитики остаются дефицитом на рынке. Ими принято называть специалистов, имеющих глубокое понимание коммерческого смысла данных и знающих, как правильно их применять.



Анализ данных является средством для достижения целей бизнеса, и чтобы понять ценность Big Data, требуется соответствующая модель поведения и понимание своих действий. В этом случае большие данные дадут массу полезной информации о потребителях, на основе которой можно принять полезные для бизнеса решения.

Несмотря на то, что казахстанский рынок Big Data только начинает формироваться, отдельные проекты в этой области уже реализуются достаточно успешно.

Если говорить об иностранных успешных кейсах, то в этом плане интересен опыт применения технологий Big Data в компании Dunkin`Donuts, использующей данные в режиме реального времени для продажи продукции. Цифровые дисплеи в магазинах отображают предложения, сменяющие друг друга каждую минуту, в зависимости от времени суток и наличия продукции. По кассовым чекам компания получает данные, какие именно предложения получили наибольший отклик у покупателей. Данный подход обработки данных позволил увеличить прибыль и оборачиваемость товаров на складе.

Как показывает опыт внедрения Big Data-проектов, эта область призвана успешно решать современные бизнес-задачи. При этом важным фактором достижения коммерческих целей при работе с большими данными является выбор правильной стратегии, которая включает в себя аналитику, выявляющую запросы потребителей, а также использование инновационных технологий в области Big Data.

В широком понимании информационная безопасность представляет собой защищенность информации и поддерживающей инфраструктуры от случайных или преднамеренных негативных воздействий естественного или искусственного характера.

В области информационной безопасности Big Data сталкивается со следующими вызовами:

- Проблемы защиты данных и обеспечения их целостности.
- Риск постороннего вмешательства и утечки конфиденциальной информации.
- Ненадлежащее хранение конфиденциальной информации.
- Риск потери информации, например, вследствие чьих-либо злонамеренных действий.
- Риск нецелевого использования персональных данных третьими лицами и т. д.

Информация – ценный актив, а это значит, что на первом плане должен стоять вопрос обеспечения основных аспектов информационной безопасности. Для того чтобы выстоять в конкурентной борьбе, компании должны идти в ногу со временем, а это значит, что им нельзя игнорировать те потенциальные возможности и преимущества, которые заключают в себе инструменты Big Data.