

Применение технологий BigData: возможности и преимущества

Гюляра Мурадова

Азербайджанский Технический Университет
gularamu.aztu.edu.az

Аннотация – “Большие данные” для эффективного управления и экономического развития в научно-исследовательской деятельности, коммерческом секторе и государственном управлении. Прогнозирование, оценка моделей управления данными.

Ключевые слова - Bid Data, электронное правительство, прогнозирование

I. ВВЕДЕНИЕ

В современном мире, где информация часто обновляется и поступает из разных источников, для ее обработки приходится работать с огромными массивами данных.

Данные — это выраженные в разной форме сырые факты. В результате анализа обработанных данных человеком, появляется информация. Таким образом в результате анализа данные приобретают смысл и обеспечиваются потребительскими качествами. Данные теперь сложнее и разнообразнее они структурированные, не поддающиеся закону, структуре и квазиструктурированные, т.е. это документы в различных форматах с произвольными описаниями каждого документа. Применение новых подходов в управлении для быстро получаемых и меняющихся сведений и их анализ является актуальной задачей.

II. ЗАДАЧИ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ

Объем накопленных данных в из разных сфер деятельности различны. Согласно IDC (International Data Corporation), компании, занимающаяся анализом рынка Информационных Технологий (ИТ), телекоммуникаций и потребительской техники объем созданной информации до 2007 года соизмерим с доступными ресурсами по хранению, в следующих годах и по настоящее время информации произведено больше, чем средств для ее хранения. В 2007 году хранилось или передавалось 2,253 зеттабайта что в пять раз больше чем двадцать лет назад – 435 экзабайт. В 2013 году количество хранящейся информации в мире составило 1,2 зеттабайтов [1]. Обычные хранилища не справляются с колоссальным объемом больших данных. Например, пользователи социальной сети Facebook ежедневно регистрируют 2,7 млрд. комментариев. Такой гигантский объем данных не позволяют провести их должную сортировку, проанализировать и извлечь из них пользу. Сложность управления всеми этими данными связана и с тем, что они передаются со скоростью, нередко

многократно превышающей возможности обработки [2]. На сегодняшний день очень сложно, даже невозможно открыть публикацию в Интернете или реальном мире, не обращаясь к науке данных, аналитике, большой базе данных big data или к их другим комбинациям. Большие данные это набор данных, который слишком велик по объёму и сложен для того, чтобы его можно было обрабатывать. Использовать традиционные методы обработки данных дорого или даже в некоторых случаях невозможно и влечет за собой, как минимум затруднения в обеспечении необходимых темпов экономического роста. Огромное количество данных требуют новых подходов в управлении и анализе.

Для обработки нужны новые компьютерные алгоритмы, которые, могут приспосабливаться под задачу. Как следствие разработаны технологии Big Data, предлагающие эффективные решения и применяемые в настоящее время. Технологии Big Data, обрабатывая поступающие данные, помогают в режиме реального времени получать информацию. Big Data — это «технологии и архитектуры нового поколения для экономичного извлечения ценности из разно форматных данных большого объема путем их быстрого захвата, обработки и анализа». Много отраслей применяют эти технологии. На рис.1 показана зависимость индекса сбора данных в зависимости от сравнительного индекса ценности применения технологии Big Data [3].

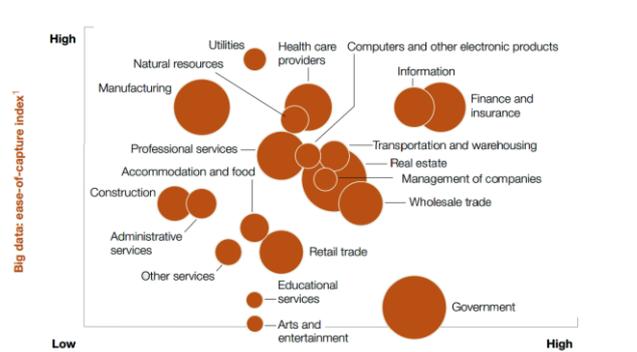


Рис.1 Источник: US Bureau of Labor Statistics

Радиус окружности показывает относительный объем индустрии. Перспективы внедрения и потенциал применения технологии Big Data в разных индустриях различны. Недвижимость, производство финансы и страхование, здравоохранение имеют больше радиусы окружности. Наиболее высокий индекс и радиус

окружности имеет применение Big Data в правительственных организациях.

Основными задачами обработки больших данных являются экспоненциальный рост данных и связанную с этим их непредсказуемость, а также средства для ее хранения. Эти задачи решают технологии Big Data.

III. АКТУАЛЬНОСТЬ ПРИМЕНЕНИЯ ТЕХНОЛОГИЙ BIG DATA

Мир больших данных радикально изменяет все: от бизнеса и естественных наук до здравоохранения, государственного управления, образования, экономики, гуманитарных наук и других аспектов жизни общества. Большие данные обеспечат себе большое виртуальное присутствие при незначительных материальных ресурсах, а также широко внедряют инновационные решения при небольших затратах [1].

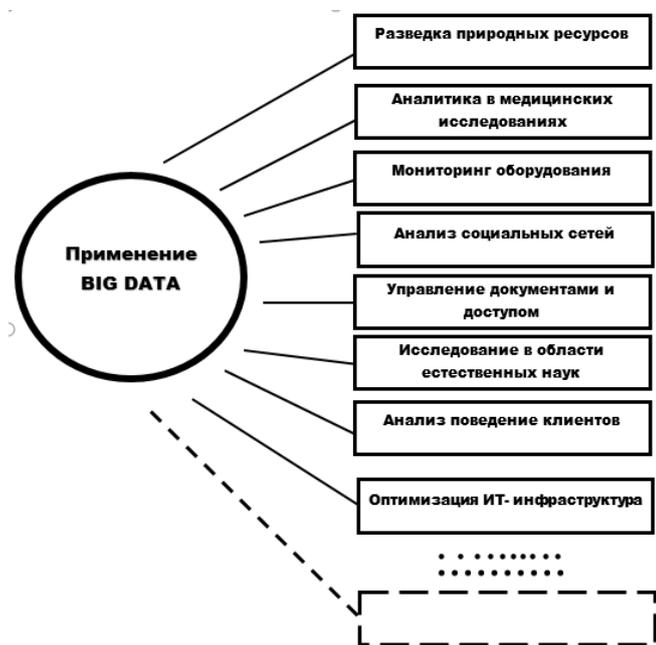


Рис.2 Примеры задач решаемых методами BigData

Невозможно перечислить все отрасли. Рассмотрим на примерах, как информация, собранная вместе и проанализированная с помощью технологий Big Data дает эффективность во многих отраслях сферы применения больших данных.

- Исследования в области естественных наук.
- Разведка природных ресурсов.

Гео-базы данных могут оказаться полезными в работе экологов, команд по ликвидации последствий стихийных бедствий и других специалистов. Разрабатывается картографические платформы, на которой специалисты визуализируют данные о численности популяций редких животных, чтобы защитить их от исчезновения. Для этого проекта, получившего название Earth Insights, в тропических лесах 16 стран мира были размещены

камеры и климатические датчики, собирающие данные о животных, растительности, температуре, атмосферных осадках, влажности и т.д. С помощью технологий Big data осуществляется управление тремя терабайтами информации, включая более 1,4 млн. фотографий и более трех млн. климатических показателей [4].

- Управление документами и доступом
- Мониторинг интеллектуальных счетчиков

С переводом госуслуг в электронный вид и созданием электронного правительства объем задействованных данных растет на порядки. Применение Big Data целесообразно, когда объемы данных составляет сотни терабайтов. Рейтинг развития электронного правительства ООН рассчитывается как интегральный показатель трех подындексов (или индексов-компонентов): развития электронных услуг, развития телекоммуникационной инфраструктуры и развития человеческого потенциала. В 2012 году администрация президента США объявила об Инициативе Исследования и Разработки Больших Данных (“Big Data Research and Development Initiative”), где изучались варианты использования больших данных для решения важных проблем, стоящих перед американским правительством, которая включала в себя 84 программы по управлению большими данными, которые впоследствии были внедрены в 6 департаментов [5]. Применительно к госуправлению - “электронному правительству” в Бразилии идет разработка стратегии управления «большими данными”. Мадрид будет использовать технологии big data для управления городской инфраструктурой. С их помощью городская администрация будет управлять работой с каждым сервис-провайдером и соответствующим образом оплачивать ее в зависимости от уровня услуг. Речь идет о подрядчиках администрации, которые следят за состоянием улиц, освещением, ирригацией, зелеными насаждениями, осуществляют уборку территории и вывоз, а также переработку мусора. В ходе проекта для специально выделенных инспекторов разработаны 300 ключевых показателей эффективности работы городских сервисов, на базе которых ежедневно будет осуществляться 1,5 тыс. различных проверок и замеров. Кроме того, город начнет использование инновационной технологической платформы под названием Madrid iNTeligente (MiNT) - Smarter Madrid. [6].

- Аналитика в медицинских исследованиях
Анализ данных электронных карт позволяет врачам видеть полную историю болезни конкретного пациента в случае обращения за медицинской помощью и оперативно принимать решения и таким образом на 20% может снизить здравоохранительная система США благодаря эффективной эксплуатации Big Data [7].
- Анализ поведения клиентов
25% Big data владеет финансовая индустрия. В среднем на каждую компанию 3,8 ПТБ данных приходится использование кредитных карточек. Это решение задач

залогах, кредитах, профилях клиентах, сбережения клиентов. Технологии Big Data увеличивают операционную рентабельность на 60% применяя прогнозирование трендов покупательского спроса : подготовка к резкому росту спроса на отдельные товары; оптимизация акций и цены; целевые маркетинговые кампании [7].

Приложение в поисковике Bing, помогает сэкономить авиапассажирам на покупке билета в среднем 50 долларов с каждого рейса, на основе прогноза изменения цен на авиарейсы [1].

- Оптимизация ИТ инфраструктуры

На 92% технологии BIG DATA могут уменьшить время обработки запроса пользователя, решая такие задачи как аналитика инфраструктуры, выявление мошенничества, улучшение качества связи, кредитование [7].

- Мониторинг оборудования

Аналитические технологии Big Data могут повысить точность распоряжения генераторами мощностей на 99%. В области энергетики применение технологий «больших данных» решает задачи ликвидации последствий отключений в реальном времени, с помощью: установки «умных» счётчиков и сенсоров; поиска новых оперативных каналов связи с клиентами. В Германии с помощью технологий Big Data рассчитываются спрос и предложение солнечной и ветряной энергии [7].

Концепция прогнозируемого производства, которая заключается в сокращении времени простоя до нуля и прозрачности процессов, требует огромного количества данных и современных инструментов прогнозирования для систематической переработки данных в ценную информацию. Концептуальная модель прогнозируемого производства начинается с получения таких сенсорных данных, как акустика, вибрация, давление, ток, напряжение и информация с датчиков. Большое количество сенсорных данных в дополнение к историческим составляют большие данные в области производства. Сгенерированные большие данные выступают в роли входного сырья для инструментов прогнозирования и развития превентивных стратегий (прогностика, здравоохранение) [5]. Анализ «больших данных» и их визуализация чрезвычайно важны для компании. Они помогают обнаружить ключевые показатели на ранней стадии, создать новую модель прогнозирования или вывести на рынок новый продукт.

Представление наиболее важной информации помогает принять обоснованные решения. Procter & Gamble создала около полусотни так называемых Business Spheres — офисов, где лица, принимающие ответственные решения, могут просматривать информацию в реальном времени на больших дисплеях. Плюс нововведения в том, что анализ «больших данных» происходит прямо на еженедельных обсуждениях, в реальном времени, создавая условия для экспериментов и мгновенной проверки гипотез. Применяя подобные методы работы, Procter & Gamble

насаждает новую, информационно ориентированную культуру работы. Знать, что произойдёт в будущем, а не анализировать, что было в прошлом, — вот стратегия «больших данных» [8].

Согласно прогнозам IDC (International Data Corporation) IDC компания, занимающаяся анализом рынка ИТ, телекоммуникаций и потребительской техники) совокупный среднегодовой темп роста (CAGR) компаний, специализирующихся в областях программного обеспечения, аппаратного обеспечения, сервисных услуг с 2010 по 2015 гг. составит 39,4%, или этот рынок будет расти примерно в семь раз быстрее, чем ИТ-рынок в целом [9].

Проблемы, вызванные экспоненциальным ростом данных, требуют особых технологий. В настоящее время для реализации технологии Big Data применяется Hadoop – это каркас с открытым исходным кодом, предназначенный для создания и запуска распределенных приложений. Кластер Hadoop состоит из многих машин, которые хранят и параллельно обрабатывают большие наборы данных. Клиентские компьютеры посылают в это вычислительное облако задания и получают результаты [10].

Большие данные помогут решению насущных глобальных проблем, таких как борьба с изменением климата, искоренение болезней, а также содействие эффективному управлению и экономическому развитию. Крупные вендоры создают платформы для анализа больших данных, стоимость которых зашкаливает за миллионы долларов, и нет сомнений, что каждый пиксель в любом уважающем себя интернет-проекте будет строиться с учетом больших данных не позднее, чем к 2020 году [1].

ЛИТЕРАТУРА

- [1] В.Майер-Шенбергер, К.Кукьер, “Большие данные”, Москва 2014, стр.9-109.
- [2] B. Gerhardt, K. Griffin, R. Klemann, “Unlocking Value in the Fragmented World of Big Data Analytics”, Cisco IBSG, 2011, p.3.
- [3] B. Brown, M. Chui, J. Manyika, “Are you ready for the era of ‘big data’?”, 2011, pp.6-8.
- [4] HP Earth Insights - a solution God might choose to monitor real life: <http://www.enterprisecioforum.com/en/blogs/enadhan/hp-earth-insights-solution-god-might-cho>
- [5] Статья Wikipedia "Big Data"(чвьсть 2) ECM-Journal.ru <http://ecm-journal.ru/docs/Statja-Wikipedia-Big-Data--2.aspx>
- [6] Madrid to implement environmental analytics project with IBM <http://enterpriseinnovation.net/article/madrid-implement-environmental-analytics-project-ibm-1399728302>
- [7] Apcon Mckinsey&company CNews Analytics 2013 <http://www.cnews.ru/reviews/?2013/12/10/552901>
- [8] Как в Procter & Gamble используют «большие данные» <http://www.computerra.ru/92102/procter-and-gamble-bigdata/>
- [9] IDC Big Data Market Sizing [https://www950.ibm.com/events/wwc/grp/grp037.nsf/vLookupPDFs/RICK%20-%20IDC_Calgary_Big_Data_Oil_andGas/\\$file/RICK%20%20IDC_Calgary_Big_Data_Oil_and-Gas.pdf](https://www950.ibm.com/events/wwc/grp/grp037.nsf/vLookupPDFs/RICK%20-%20IDC_Calgary_Big_Data_Oil_andGas/$file/RICK%20%20IDC_Calgary_Big_Data_Oil_and-Gas.pdf)
- [10] Ч.Лэм, “Hadoop в действии“, М.: ДМК Пресс, 2012, стр.23-25