

BIG DATA АНАЛИТИКА И ЕЕ ПРИМЕНЕНИЕ



В.С. Дроздов

Ассистент кафедры инженерной психологии и эргономики,
аспирант БГУИР, магистр
технических наук

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь
E-mail: fxtraid@tut.by

Abstract. Big data provides the opportunity to discover phenomena and laws, which until now was beyond our understanding; linking various aspects of the system to help you better understand their behavior; to assist in the description of complex phenomena and processes; coordinate descriptions of complex systems and enables you to build models that predict their dynamic behavior. These capabilities are deeply relevant to many research fields such as weather forecasting and climate, studies of the brain; determining the state of the global economy; the assessment of performance in agriculture; demographic projections

Big Data – это группа технологий и методов производительной обработки очень больших объемов данных, в том числе неструктурированных, в распределенных информационных системах, обеспечивающих организацию качественно новой полезной информации. Технологии Big Data предоставляют услуги, позволяющие раскрыть потенциал мегамассивов данных за счет выявления скрытых закономерностей и фактов. Под очень большими наборами данных подразумеваются данные объемом от терабайт до сотен петабайт. Рассмотрим важность использования Big Data в различных сферах жизни и производства.

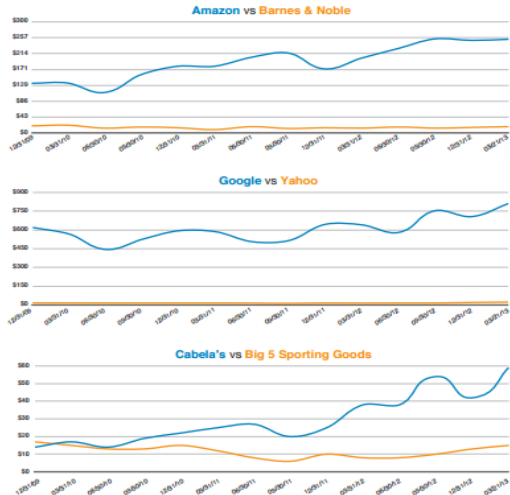
Признаки, характеризующие Большие Данные

Признаки	Пояснение
Объем (volume)	Представляет собой большой объем информации, который трудоемко обрабатывать и хранить традиционными способами
Скорость (velocity)	Данный признак указывает на скорость обработки данных, что в последнее время более востребовано
Многообразие (variety)	Возможность одновременной обработки структурированной и неструктурированной разноформатной информации
Достоверность данных (veracity)	Все большее значение пользователи стали придавать значимость достоверности имеющихся данных
Ценность накопленной информации (value)	Большие данные должны быть полезны компании и приносить особую ценность для нее

Почему Big Data аналитика так важна?

1. Производительность от управления данными

Компании, которые используют Big Data аналитику (синие линии) более продуктивны, чем те компании, которые этого не делают



2. Результаты использования Big Data аналитики

Компании из всех отраслей промышленности используют большие данные для:

- Увеличения выручки
- Сокращения затрат
- Повышения продуктивности

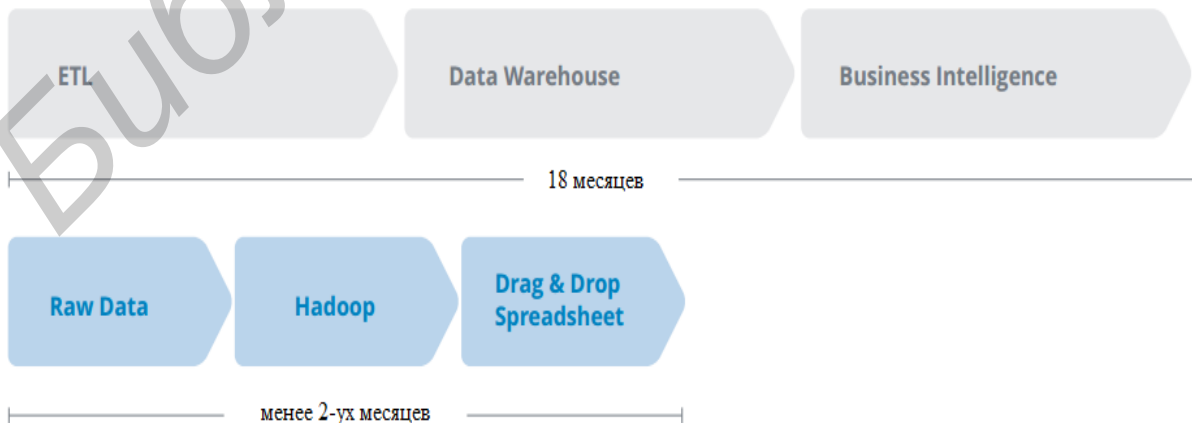


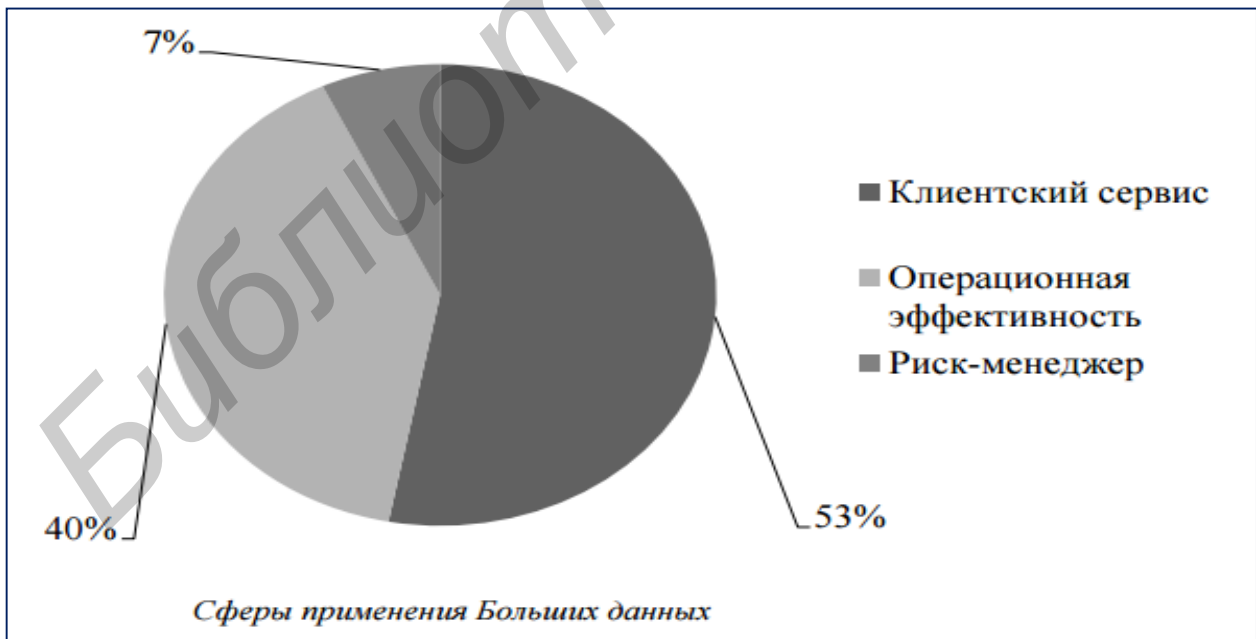
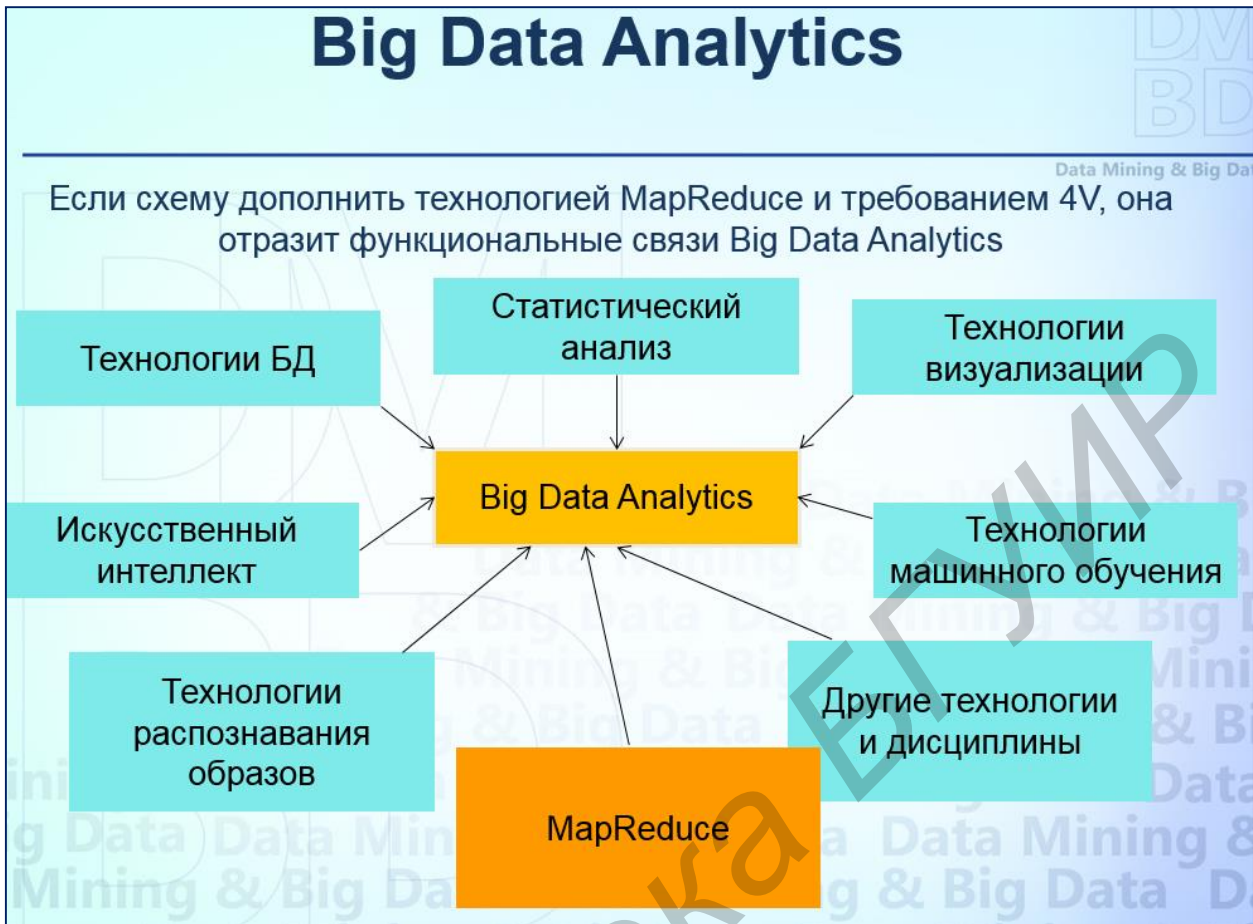
* Powered by Datameer

Что такое Big Data аналитика и что делает ее такой сильной?

Проблематика

До hadoop у нас было ограниченное пространство для хранения данных а также вычислительные ресурсы, что привело к длительной аналитической обработке данных(см.ниже). Во-первых, обработка проходит через длительный процесс (часто известный как ETL), чтобы получить каждый новый источник данных, готовый быть сохраненным. После того, как данные готовы, они помещаются в базу данных или хранилище данных, и в статическую модель данных. Проблема с этим подходом заключается в том, что она разрабатывает сегодня модели данных которые содержат вчерашние данные, и вы должны надеяться, что они будут актуальны для завтра.





Косвенные направления генерирования сегмента «Big Data»	
Направления	Описание генерирования мультипликативных эффектов
Разработка новых видов продуктов и услуг	Это осуществляется за счет формирования более полного представления о потребностях, предпочтениях, пожеланиях покупателей, а также о них самих. В этом плане по расчетам каждое вновь созданное рабочее место в сфере «Big Data» влечет за собой возникновение 3 новых рабочих за пределами ИКТ-сектора.
Повышение эффективности хозяйственной деятельности субъектов в уже существующих отраслях	Приведем несколько примеров, например: - производитель продуктов питания Nestle за счет ИКТ-оптимизации и систематизации баз данных по своим 550 тыс. поставщикам уменьшил операционные расходы на 1 млрд. долл.; - мобильному оператору Cablecom (Швейцария) удалось снизить уровень оттока абонентов с 20 до 5%, а розничной сети Royal Shakespeare Company - увеличить число посетителей на 70% посредством развернутого анализа данных о своих клиентах.

Эффекты применения технологии «Big Data» в различных сферах	
Сфера применения	Результативность применения технологии «Big Data»
Добыча полезных ископаемых	Плоскость применения помимо геологоразведки затрагивает непосредственно сам процесс добычи полезных ископаемых. Таким образом, на одном месторождении достигается сокращение операционных расходов на 10-25% и рост уровня производительности на 5%.
Государственное управление	Результируется в сокращении бюджетных расходов администрирования на 15-20%; повышение уровня собираемости налогов на 10% и увеличении эффективности государственных закупок на 30%. Также предполагается наличие действующего и масштабного по географии и спектру охвата сфер жизнедеятельности институтов электронного правительства.
Здравоохранение	Точкой опоры в данном случае выступает повседневная клиническая практика, где технологии «Big Data» могут быть использованы для существенного повышения эффективности и качества медицинского обслуживания при параллельном сокращении затрачиваемых на эти цели из государственного бюджета сумм посредством.
Наука	Пути применения их по конкретным направлениям исследований проиллюстрированы также как на примере сферы здравоохранения: прогностическое моделирование при разработке новых видов лекарственных средств, организация клинических испытаний на основе статистических данных, анализ тенденций заболеваемости, анализ данных клинических исследований.

Большие данные – большой успех

Согласно результатам исследования большие данные представляют значительную ценность для пользователей, реализовавших как минимум один проект.

Подавляющее большинство (92%) пользователей отмечают, что они удовлетворены бизнес-результатами. Кроме того, 94% заявили, что активное использование больших данных полностью удовлетворяет их потребности. Крупные компании в большей степени ощущают на себе чрезвычайно высокую значимость больших данных для реализации их цифровой стратегии (см. рис. 1).

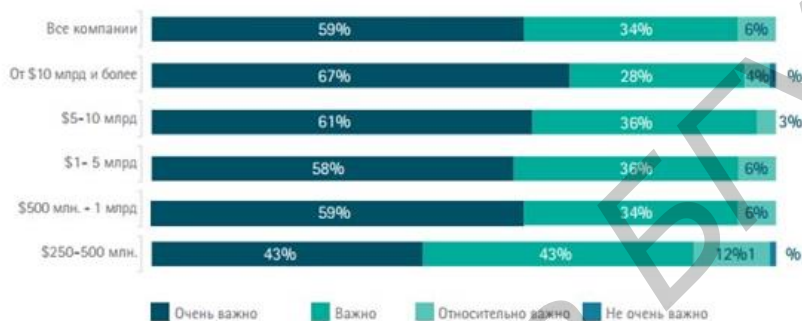
В то же время, возможно, еще многие организации пока находятся вне игры, но те, кто начал внедрять у себя большие данные, уже смогли увидеть собственными глазами практическую пользу и ощутить их ценность.

В компаниях осознают всю важность больших данных для широкого спектра стратегических корпоративных целей, начиная от поиска новых источников дохода и выхода на новые рынки и заканчивая повышением качества обслуживания клиентов и эффективности предприятия в целом.

Один из крупнейших в мире контейнерных грузоперевозчиков также входит в список крупнейших пользователей больших данных, ежегодно расходуя на хранение и обработку 16 петабайт данных, собранных со всех бизнес-подразделений, \$1 миллиард.¹

Важность больших данных

Насколько важны большие данные для вашей организации?



¹ Источник: <http://smartdatacollective.com/bigdatastartups/201286/why-ups-spendsover-1-billion-big-data-annually>

Применение анализа больших данных

Финансы

- Решения по рискам
- Анализ мнения клиентов
- Борьба с отмыванием денег



Энергетика

- Влияние погоды на генерацию энергии
- Анализ данных от умных счетчиков

Транспорт

- Влияние погоды и трафика на доставку и потребление топлива



ИТ

- Анализ логов от разных транзакционных систем

Колл центр

- Анализ расшифровок разговоров для понимания поведения клиентов

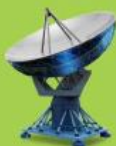


E Commerce

- Анализ поведения и покупательских моделей

Телко

- Анализ операций и сбоев сети



Интеграция каналов взаимодействия

- Моделирование поведения клиентов

Объемы Больших данных

	Google прекратил сообщать как много данных они хранят в 2010 (SEC filing): в то время это было 100 PBs
	YouTube – порядки измеряются в Exabyte •72+ ч видео загружаются на YouTube каждую минуту •YouTube второй по использованию поисковый движок после Google •Последние данные 768+ PBs, 3-4 года назад: точно больше Exabyte сейчас
	Facebook перевалил за миллиард пользователей в августе 2012 • Население планеты стало больше 7B в прошлом году: 1/6 th – в Facebook •35% мировых фотографий по оценкам в Facebook
	Twitter - около 124 млрд tweets в год, в среднем 4500 в сек
	Обмен сообщениями в мире 193,000 смс/сек
	Звонки в США: 2.2 триллиона минут в год; 19 мин/день/человека

Заключение. Обоснованы необходимость использования и перспективность применения технологий Big Data. Приведены результаты исследований применения технологий Big Data. Исследованы современное состояние и тенденции развития технологий Big Data.