Конференция Большие данные 2015
Москва, 1 октября 2015 г.

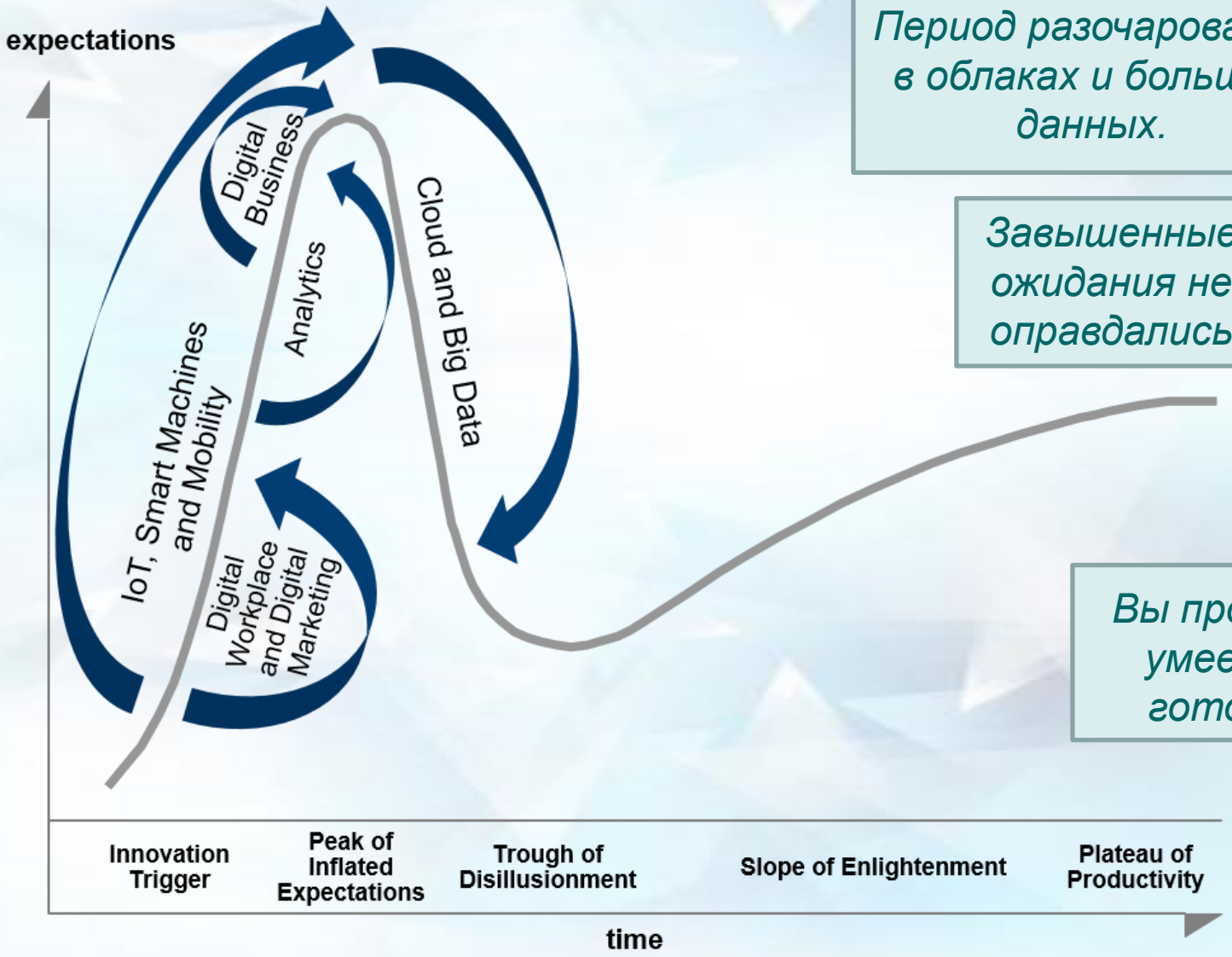
Чем могут помочь большие данные в сложных экономических условиях

Марина Аншина,

Председатель Комитета по стандартам
Российского Союза ИТ-директоров

CNews Conferences
агентство маркетинговых коммуникаций

Развитие технологий по Гартнер



Период разочарования в облаках и больших данных.

Завышенные ожидания не оправдались

Вы просто не умеете их готовить

Прогноз Гартнер

Мегатренд 4: Пик аналитики

- Настало время аналитики, особенно с учетом Интернета вещей, мобильности, получения доступа к данным о предпочтениях клиентов и безопасности.
- Стали доступны возможности доступа к новым областям данных и их обработки в реальном времени.
- В 2014 г. аналитика была в числе наиболее быстро развивающихся областей.
- В 2015 г. аналитические технологии и сервисы развиваются, но по оценке Гартнер находятся вблизи пика завышенных ожиданий.

Мегатренд 5: Большие данные и облака двигаются в сторону разочарования от завышенных ожиданий

- В 2014 большие данные прошли пик завышенных ожиданий. Однако информация и облачные технологии продолжат в 2015 г. оставаться критически важными для организаций. Некоторые наиболее быстро развивающиеся области ИТ в 2015 г. по прогнозу Гартнер относятся к большим данным.
- Однако по мере того, как большие данные и облака укореняются в бизнесе и ИТ, клиенты всё больше внимания обращают на то, как они используются в конкретной отрасли и удовлетворяют специфические потребности бизнеса и меньше внимания обращают на инновационную составляющую.

Gartner®

Gartner: инвестиции в обработку Больших Данных растут медленнее, чем раньше

- В большинстве компаний ожидают отдачи от инвестиций в подобные инициативы, 43% из тех, кто только планирует инвестиции и 38% из тех, кто их уже сделал, сообщают, что пока не знают, какой она будет. Это указывает на проблемы в определении ценности проектов Больших Данных, указывают аналитики.
- Три с лишним четверти компаний, участвовавших в проведенном аналитиками Gartner опросе, уже инвестируют в технологии обработки Больших Данных, или планируют заняться этим в ближайшие два года. Это лишь на 3% больше, чем в 2014 году, отмечают аналитики. Темп роста инвестиций снизился, но главное заключается в том, что обработка больших объемов данных перестает быть самостоятельной проблемой и становится частью обычной работы.
- Как и раньше, чаще всего главной целью внедрения этих технологий в компаниях называют улучшение обслуживания клиентов. Так ответило 64% опрошенных. По 47% опрошенных называют в числе целей повышение эффективности процессов и точности целевого маркетинга. Доля тех, кто ставит задачи повышения безопасности, выросла с прошлого года с 15% до 23%.
- Хотя в большинстве компаний ожидают положительного коэффициента доходности инвестиций, 43% из тех, кто только планирует инвестиции и 38% из тех, кто их уже сделал, сообщают, что пока не знают, каким он будет. Это указывает на проблемы в определении ценности проектов Больших Данных, указывают аналитики.
- В прошлом году 37% проектов начинались по предложениям руководителей информационных служб, а 25% — по инициативе бизнес-подразделений. В этом году доли почти сравнялись — 32% и 31%, соответственно.

Открытые системы, сентябрь, 2015

Пример проекта	Это большие данные?	Комментарий
Анализ большого объема неструктурированной информации из различных источников в режиме реального времени	V да	Классический пример
Анализ перемещения покупателей в магазинах по сигналам WiFi	V да	Этот тип проектов будет относиться к большим данным, если необходимо производить анализ в реальном времени и при возникновении определенных событий (например, покупатель подошел к стенду) предпринимать определенные действия или же на основе анализа данных за большой период делать решения об изменении расстановки товаров, соотносить траектории с типами клиентов и т.д.
Быстрое формирование финансовой отчетности высокопроизводительными средствами больших данных	? мнения разделились	В классическом понимании нет, хотя технологии больших данных могут быть использованы, если объемы и сложность расчетов и преобразований не позволяют традиционным технологиям удовлетворять требованиям по быстродействию
Анализ большой по объемам, но структурированной информации из внутренних источников	? мнения разделились	В классическом понимании нет, хотя в этом случает использование технологий больших данных оправдано, чтобы уменьшить стоимость хранения и обработки больших объемов информации и сократить время анализа или провести анализ по очень сложным алгоритмам (телеком)
Потоковая обработка однородных структурированных данных	? мнения разделились	Подобные задачи традиционно решались в производственных и биржевых системах до появления технологий больших данных. Однако существует целый ряд новых задач, когда нужно «на лету» объединить разнородную информацию и обеспечить принятие решений в реальном времени. Например, сверить биометрические данные в потоках людей с базами злоумышленников, данными о пересечении границ, авиабилетами и пр.
Обработка неструктурированных данных (текстовая и видеоаналитика)	? мнения разделились	Зачастую подобные решения относят к большим данным, хотя алгоритмы интеллектуальной обработки информации развиваются в течение десятков лет. Машинное обучение успешно применялось до появления термина «большие данные». В наши дни можно говорить об отнесении подобных решений к категории больших данных, если они реализуются на больших объемах информации в режиме времени близком к реальному
Обработка небольших объемов неструктурированной информации	X нет	Например, подключение к социальной сети с целью анализа информации о каком-то заранее известном пользователе
Терабайт данных, обрабатываемый на одном большом сервере	X нет	Размер не имеет значения – см. выше

Cnews, 2014

Что же делать с данными?

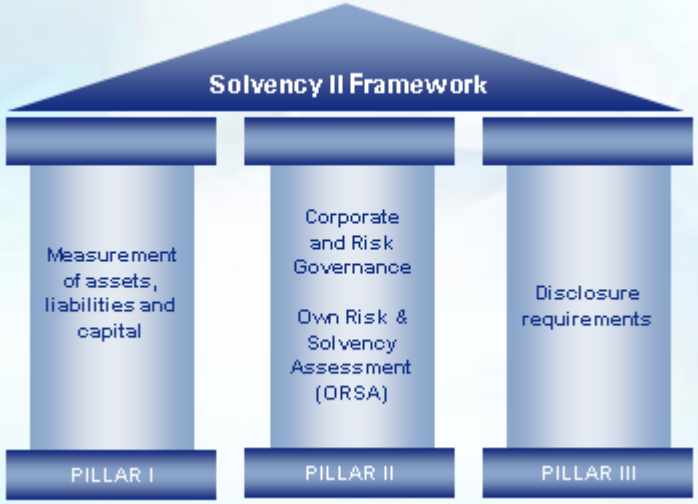
<u>90% данных в мире сформированы за последние 2 года</u>

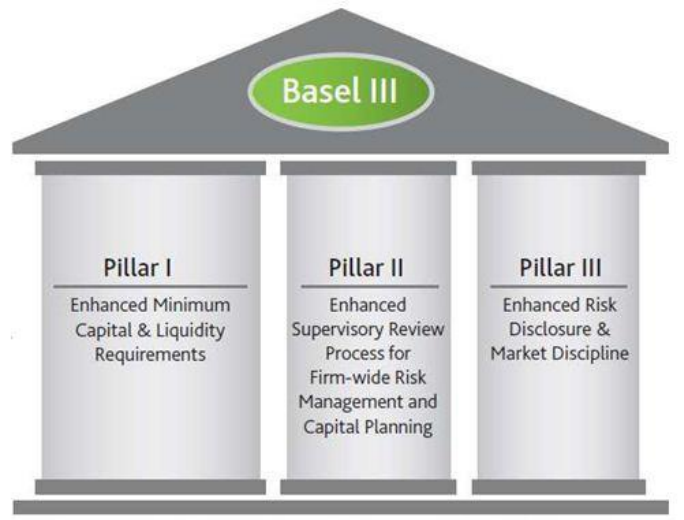
Количество данных на планете будет удваиваться каждые два года вплоть до 2020 года

- **ИТ-лидеры:** Инвестиции и программы связанные с выявлением и анализом нарушений нормального функционирования бизнеса. Обучение руководителей и лидеров маркетинга работе с аналитическими системами.
- **Инвесторы и предприниматели:** Используют аналитику для выявление и проектирования новых возможностей бизнеса.
- Большие данные и распределенные вычисления подогрели интерес и привели **инвестиции в аналитику**. Гугл например купил компанию DeepMind (аналитический стартап) за более чем $500 млн.
- IBM Watson, например, позволяет выявлять **рыночные закономерности** на основе рыночных данных, мировых событий, клиентской истории и доступных заказов.
- Большое количество **стартапов** в области бизнес-аналитики.
- Оксфорд прогнозирует, что **более половины рабочих мест в США** в ближайшем будущем будут связаны с бизнес-аналитикой.

Европейские нормативные требования

- Обработка больших объемов данных становится серьезной проблемой для многих компаний, и это повышает спрос на специалистов по управлению базами данных. Помимо роста объемов неструктурированных данных (например, сообщений в социальных сетях), спрос повышается из-за подготовки к введению в Европе новых нормативных требований — в том числе стандартов платежеспособности Solvency II для страховых компаний и стандартов капитала и ликвидности Basel III для банковского сектора.





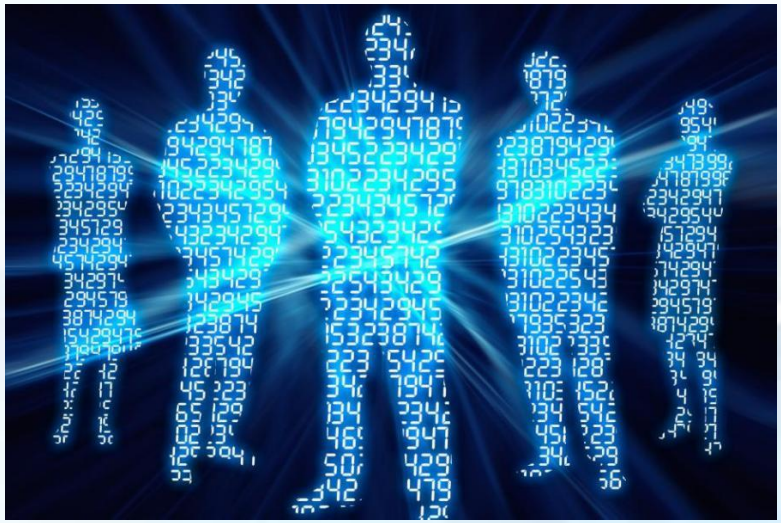
Традиционная автоматизация и облака

- Существующие бизнес-модели, несмотря на Интернет и электронный бизнес, начинают уходить от классической автоматизации. Цифровой бизнес и основные тренды (Nexus of Forces): облака, социальные сети, большие данные и мобильность привели к ускоренному развитию цифровых технологий.
- **Всё как сервис** — облака оказывают огромное влияние на развитие технологий.

Облака позволяют получить доступ к современным средствам обработки данных организациям СМБ

Гартнер больше не считает большие данные отдельным трендом!

- Обработка больших и неструктурированных массивов данных становится интегральной частью бизнеса большинства организаций и их вычислительных сред



О терминологии

- Большие данные, Smart Company, Digital Company, расширенная аналитика, продвинутая аналитика, перспективная аналитика
- Новый уровень извлечения и использования данных

Пик развития

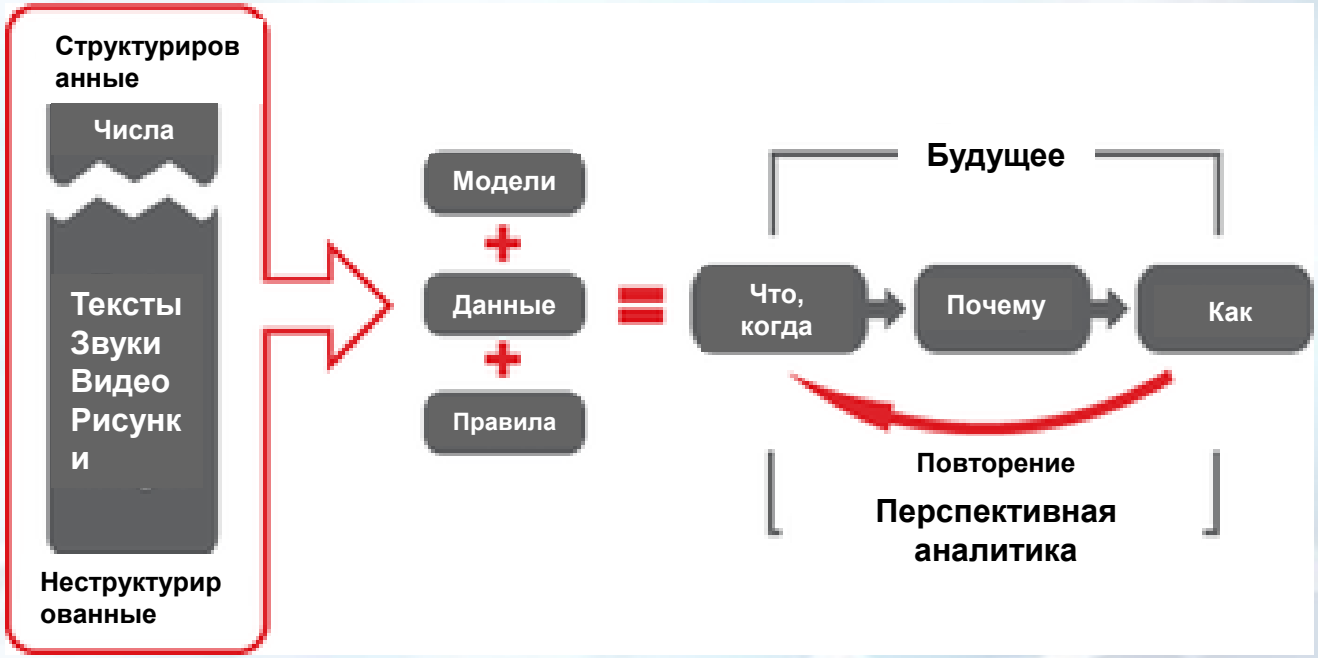
- Управление принятием решений
- Предиктивная аналитика
- Перспективная аналитика
- Обработка потока событий
- Геопространственная аналитика
- Бизнес – анализ, основанный на PaaS (baPaaS)
- Управляемое обнаружение данных (Governed Data Discovery)

Governed Data Discovery

- Обнаружение данных о конечном пользователе
- С соблюдением требований безопасности

Предиктивная и перспективная аналитика

Перспективная аналитика

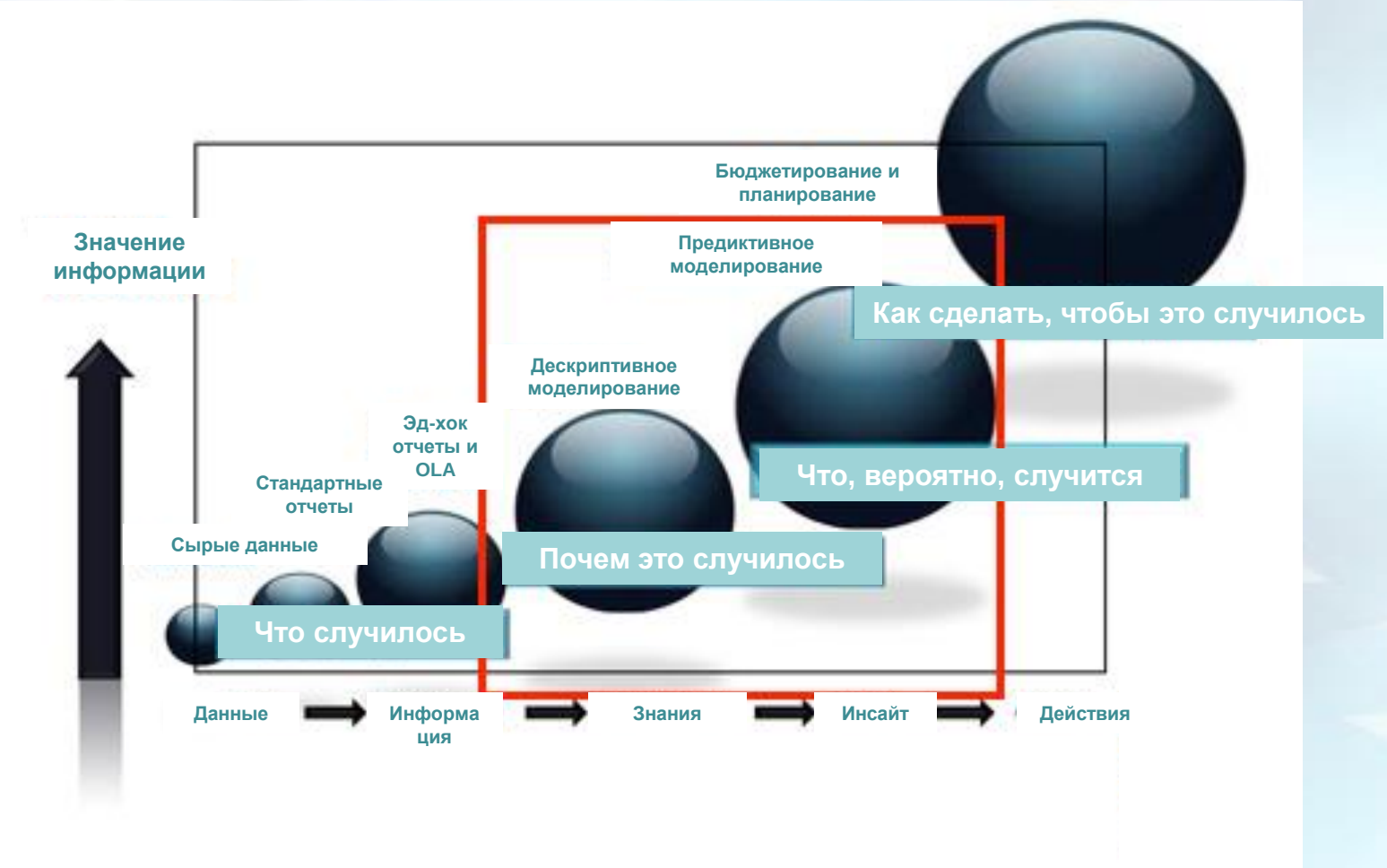


Перспективная аналитика

Спад от завышенных к реальным ожиданиям

- Хранилища данных
- Natural-Language Question Answering
- Большие данные
- Мобильная BI
- Сервисы бизнес-аналитики
- R
- Анализ текстов

Развитие аналитики

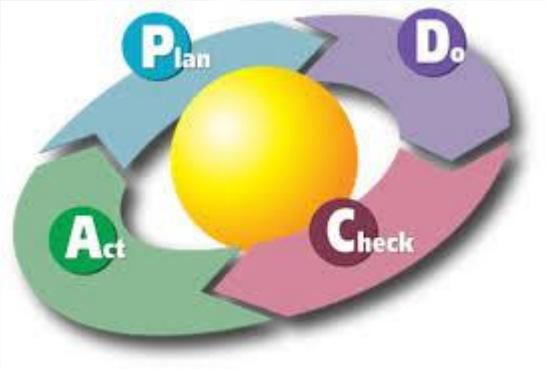


Как сделать первый шаг к продвинутой аналитике

- Выбрать бизнес-проблему, которая может обеспечить первую победу
- Использовать аутсорсинг и закупить приложение, если у вас ещё нет опыта продвинутой аналитики
- Выявить стейкхолдеров в организации, которым нужны результаты продвинутой аналитки
- Решить, стоит ли развивать компетенции и инструменты внутри организации

Цикл PDCA в применении к аналитике

- Выбери инструмент
- Установи и настрой
- Используй
- Оптимизируй



БОЛЬШИЕ ДАННЫЕ в банках

У БАНКОВСКОГО БИЗНЕСА ПОЯВИЛСЯ НОВЫЙ АКТИВ

Банки США сейчас накопили **1 экзабайт** данных

275 млрд записей mp3

Классические источники БОЛЬШИХ ДАННЫХ в банках

Как банки используют БОЛЬШИЕ ДАННЫЕ

Решение задач в финансовом секторе с помощью больших данных

Предотвращение оттока клиентов

В 2012 году **50%**

РЕШЕНИЕ

Определение оптимальной численности штата

66%

РЕШЕНИЕ

Понимание потребностей клиентов

РЕШЕНИЕ

Управление растущими расходами на безопасность

РЕШЕНИЕ

Идеи для развития продуктов

Кредитный скоринг

РЕШЕНИЕ

800 млрд записей,

ЗАЧЕМ BIG DATA

в автомобилестроении?

CNews Analytics

ОТРАСЛЬ

1,5 КИЛОМЕТРА проводов

50 блоков УПРАВЛЕНИЯ

в **20** РАЗ ВЫШЕ вычислительная мощность, чем У ПРОДВИНУТОГО ПК

СОВРЕМЕННЫЙ АВТОМОБИЛЬ:

ДАТЧИКИ В АВТОМОБИЛЕ:

- **НА РУЛЕ** – отслеживают температуру лица, а сенсоры на внутренней **БОКОВОЙ ПОВЕРХНОСТИ РУЛЯ** – температуру ладоней
- **СНАРУЖИ** – датчик **температуры** окружающей среды
- **ПОД КОЛОНКОЙ РУЛЕВОГО УПРАВЛЕНИЯ** – **сравнивает** общую температуру с **температурой воздуха** рядом с водителем
- **В РЕМНЕ БЕЗОПАСНОСТИ** – пьезоэлектрический датчик отслеживает **дыхание водителя**

480 ТБ данных собрано **в 2013 г.** мировым авторынком / **11,1 ПБ** прогноз к 2020 г.

ПРОИЗВОДСТВО

15 000 ЗАРЕГИСТРИРОВАННЫХ НЕИСПРАВНОСТЕЙ — **ДАННЫЕ С ТЕСТ-ДРАЙВОВ** прототипов автомобилей и сведения из последних **ОТЧЕТОВ АВТОМАСТЕРСКИХ** анализирует система IBM SPSS **ДО НАЧАЛА СЕРИЙНОГО ПРОИЗВОДСТВА**

500 СОТРУДНИКОВ из разных подразделений и дочерних организаций ИСПОЛЬЗУЮТ **250** АНАЛИТИЧЕСКИХ ПРОГРАММ

10 000 ИТ-СПЕЦИАЛИСТОВ

НОВАЯ СТРАТЕГИЯ — превращение в **ЦИФРОВУЮ ФАБРИКУ**

в 1,5 РАЗА СНИЗИТСЯ в США НОРМА **РАСХОДА ТОПЛИВА** для ЛЕГКОВЫХ АВТОМОБИЛЕЙ и МАЛОТОННАЖНЫХ ГРУЗОВИКОВ **к 2025 г.**

Аналитика Big Data поможет производителям достичь нового стандарта

$400 млн привлекла компания MOBILEYE на развитие собственных АВТОПИЛОТИРУЕМЫХ УСТРОЙСТВ

1 ГБ данных в секунду будут ГЕНЕРИРОВАТЬ СЕНСОРЫ для АВТОПИЛОТА

ВОЖДЕНИЕ

2 ПБ данных генерируются одним автомобилем за **600** ЧАСОВ В ГОД, которые в среднем проводит за рулем АМЕРИКАНЕЦ

АВТОМОБИЛИ НА ФОРМУЛЕ-1:

ТЕЛЕМЕТРИЯ НА F1 используется с конца 80-х гг.

130 СЕНСОРОВ у гоночного автомобиля на F1

ДАННЫЕ в реальном времени передаются с борта авто в автомастерскую

за **2 ЧАСА** гонки автомобиль дает столько же информации, сколько собрано в нескольких телефонных справочниках

Hertz. ТЕХНОЛОГИИ BIG DATA ПОМОГАЮТ HERTZ (компания по прокату автомобилей, представлена в 8,3 тыс. городов в 145 странах)

- **МОНИТОРИТЬ** блоги, социальные сети, комментарии на вебсайтах, почту и другие **ИСТОЧНИКИ ИНФОРМАЦИИ**,
- **ЧТОБЫ ОЦЕНИТЬ УРОВЕНЬ УДОВЛЕТВОРЕННОСТИ КЛИЕНТОВ**

ПРОГНОЗ

$14,5 МЛН ДОХОДА ПРИНЕСУТ СИСТЕМЫ ОБРАБОТКИ ДАННЫХ, ПЕРЕДАННЫХ С АВТОМОБИЛЕЙ, к 2020 г.

152 МЛН ВЕБ-МОБИЛЕЙ БУДЕТ В МИРЕ к 2020 г.

ИСТОЧНИКИ: Auto Rental, BigdataStartups, CNews Analytics, Crains Detroit Business, Datanami, Forbes, IBM, Technet, 2013-2014

при поддержке IBM

БОЛЬШИЕ ДАННЫЕ В РИТЕЙЛЕ

Подготовлено CNews Analytics

20% ПОКУПАТЕЛЕЙ крупного супермаркета генерируют **80%** ПОКУПОК

это

1% ВСЕХ ПРОДУКТОВ, выставленных на полках магазина в течение года

Большинство покупателей приобретают

КАТЕГОРИЙ ТОВАРА	в год	в квартал	в неделю
	260	80	13

26% ПОКУПАТЕЛЕЙ в непраздничные недели

33% ПОКУПАТЕЛЕЙ в праздничные недели

ПРИОБРЕТАЮТ РЕКЛАМИРУЕМЫЕ ТОВАРЫ

ОНЛАЙН-РОЗНИЦА РФ

В **18** РАЗ меньше китайской

Оборот российской онлайн-розницы В **4,5** РАЗА меньше оборота Amazon

ЧИСЛО ОНЛАЙН-МАГАЗИНОВ

>550 ТЫСЯЧ в Европе

<40 ТЫСЯЧ в России

3/4 СПЕЦИАЛИСТОВ РОЗНИЧНОЙ ТОРГОВЛИ считает, что большие данные необходимы для развития конкурентоспособной стратегии продвижения по разным каналам связи с покупателями

60% РИТЕЙЛЕРОВ не определились, нужно ли работать с большими данными

20% РИТЕЙЛЕРОВ уже используют большие данные

20% РИТЕЙЛЕРОВ посчитали работу с ними нецелесообразной

ПО СТАТИСТИКЕ hadoop ПОСЛЕ ВНЕДРЕНИЯ СИСТЕМЫ В РИТЕЙЛЕ:

производительность аналитики увеличивается до обработки **125** АЛГОРИТМОВ ЗА **20** МСЕК

прибыль растет НА **7-10%**

магазины, анализируя данные о поведении клиентов, ПРОГНОЗИРУЮТ СПРОС

amazon

240 млн клиентов

$75 млрд выручка в год

Анализ данных о покупателях и разработка предложений

Используют Hadoop

macy's

НА **383%** выросла стоимость бренда

Анализ поведения покупателей в интернете

BURGER KING

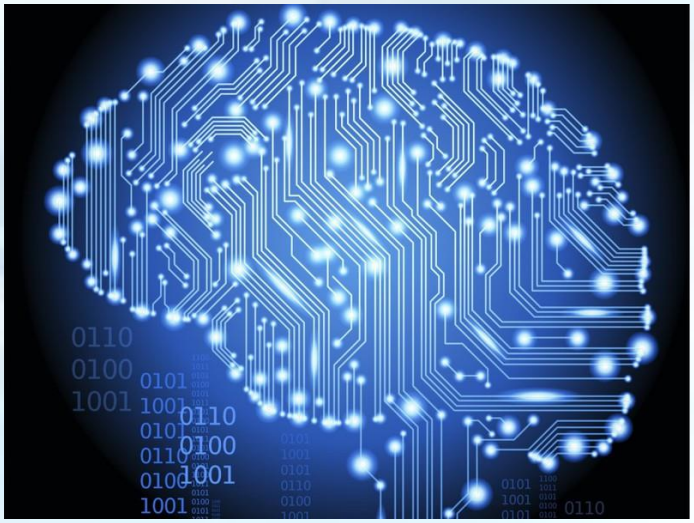
7 200 ЗАВЕДЕНИЙ планируют использовать ПО для анализа поведения покупателей

Анализ поведения покупателей позволит выдавать постоянным клиентам купоны на бесплатную еду, отправлять индивидуальные предложения по СМС, e-mail, а также приготовить гамбургер по индивидуальному рецепту сделавшего заказ с помощью мобильного приложения

Планирует внедрить Hadoop

Особенности поведения в кризис

- Гибкость
- Масштабируемость
- Точный расчёт
- Быстрые победы



Основные потребители информации

- Топ-менеджеры
- ЛПР
- Маркетологи и продавцы
- Аналитики

?

СПАСИБО
ЗА ВНИМАНИЕ

anshina@mail.ru