Большие данные: вчера, сегодня, завтра
Сергей Лихарев,
IBM Big Data Solutions
At the World Economic Forum last month in Davos, Switzerland, Big Data was a marquee topic. A report by the forum, “Big Data, Big Impact,” declared data a new class of economic asset, like currency or gold. Companies are being inundated with data—from information on customer-buying habits to supply-chain efficiency. But many managers struggle to make sense of the numbers.

“Increasingly, businesses are applying analytics to social media such as Facebook and Twitter, as well as to product review websites, to try to understand where customers are, what makes them tick and what they want,” says Deepak Advani, who heads IBM’s predictive analytics group.

“Big Data has arrived at Seton Health Care Family, fortunately accompanied by an analytics tool that will help deal with the complexity of more than two million patient contacts a year…”

“Companies are being inundated with data—from information on customer-buying habits to supply-chain efficiency. But many managers struggle to make sense of the numbers.”

“Data is the new oil.”

Clive Humby

“…now Watson is being put to work digesting millions of pages of research, incorporating the best clinical practices and monitoring the outcomes to assist physicians in treating cancer patients.”

The Oscar Senti-meter — a tool developed by the L.A. Times, IBM and the USC Annenberg Innovation Lab — analyzes opinions about the Academy Awards race shared in millions of public messages on Twitter.”

“Companies are being inundated with data—from information on customer-buying habits to supply-chain efficiency. But many managers struggle to make sense of the numbers.”

“Data is the new oil.”

Clive Humby

“The New York Times

“Increasingly, businesses are applying analytics to social media such as Facebook and Twitter, as well as to product review websites, to try to understand where customers are, what makes them tick and what they want,” says Deepak Advani, who heads IBM’s predictive analytics group.”

“Big Data has arrived at Seton Health Care Family, fortunately accompanied by an analytics tool that will help deal with the complexity of more than two million patient contacts a year…”

“Companies are being inundated with data—from information on customer-buying habits to supply-chain efficiency. But many managers struggle to make sense of the numbers.”

“Data is the new oil.”

Clive Humby

“The New York Times

“Increasingly, businesses are applying analytics to social media such as Facebook and Twitter, as well as to product review websites, to try to understand where customers are, what makes them tick and what they want,” says Deepak Advani, who heads IBM’s predictive analytics group.”

“Big Data has arrived at Seton Health Care Family, fortunately accompanied by an analytics tool that will help deal with the complexity of more than two million patient contacts a year…”

“Companies are being inundated with data—from information on customer-buying habits to supply-chain efficiency. But many managers struggle to make sense of the numbers.”

“Data is the new oil.”

Clive Humby

“The New York Times

“Increasingly, businesses are applying analytics to social media such as Facebook and Twitter, as well as to product review websites, to try to understand where customers are, what makes them tick and what they want,” says Deepak Advani, who heads IBM’s predictive analytics group.”

“Big Data has arrived at Seton Health Care Family, fortunately accompanied by an analytics tool that will help deal with the complexity of more than two million patient contacts a year…”

“Companies are being inundated with data—from information on customer-buying habits to supply-chain efficiency. But many managers struggle to make sense of the numbers.”

“Data is the new oil.”

Clive Humby

“The New York Times

“Increasingly, businesses are applying analytics to social media such as Facebook and Twitter, as well as to product review websites, to try to understand where customers are, what makes them tick and what they want,” says Deepak Advani, who heads IBM’s predictive analytics group.”

“Big Data has arrived at Seton Health Care Family, fortunately accompanied by an analytics tool that will help deal with the complexity of more than two million patient contacts a year…”

“Companies are being inundated with data—from information on customer-buying habits to supply-chain efficiency. But many managers struggle to make sense of the numbers.”

“Data is the new oil.”

Clive Humby

“The New York Times

“Increasingly, businesses are applying analytics to social media such as Facebook and Twitter, as well as to product review websites, to try to understand where customers are, what makes them tick and what they want,” says Deepak Advani, who heads IBM’s predictive analytics group.”

“Big Data has arrived at Seton Health Care Family, fortunately accompanied by an analytics tool that will help deal with the complexity of more than two million patient contacts a year…”

“Companies are being inundated with data—from information on customer-buying habits to supply-chain efficiency. But many managers struggle to make sense of the numbers.”

“Data is the new oil.”

Clive Humby

“The New York Times

“Increasingly, businesses are applying analytics to social media such as Facebook and Twitter, as well as to product review websites, to try to understand where customers are, what makes them tick and what they want,” says Deepak Advani, who heads IBM’s predictive analytics group.”

“Big Data has arrived at Seton Health Care Family, fortunately accompanied by an analytics tool that will help deal with the complexity of more than two million patient contacts a year…”

“Companies are being inundated with data—from information on customer-buying habits to supply-chain efficiency. But many managers struggle to make sense of the numbers.”

“Data is the new oil.”

Clive Humby

“The New York Times

“Increasingly, businesses are applying analytics to social media such as Facebook and Twitter, as well as to product review websites, to try to understand where customers are, what makes them tick and what they want,” says Deepak Advani, who heads IBM’s predictive analytics group.”

“Big Data has arrived at Seton Health Care Family, fortunately accompanied by an analytics tool that will help deal with the complexity of more than two million patient contacts a year…”

“Companies are being inundated with data—from information on customer-buying habits to supply-chain efficiency. But many managers struggle to make sense of the numbers.”

“Data is the new oil.”

Clive Humby

“The New York Times

“Increasingly, businesses are applying analytics to social media such as Facebook and Twitter, as well as to product review websites, to try to understand where customers are, what makes them tick and what they want,” says Deepak Advani, who heads IBM’s predictive analytics group.”

“Big Data has arrived at Seton Health Care Family, fortunately accompanied by an analytics tool that will help deal with the complexity of more than two million patient contacts a year…”

“Companies are being inundated with data—from information on customer-buying habits to supply-chain efficiency. But many managers struggle to make sense of the numbers.”

“Data is the new oil.”

Clive Humby

“The New York Times

“Increasingly, businesses are applying analytics to social media such as Facebook and Twitter, as well as to product review websites, to try to understand where customers are, what makes them tick and what they want,” says Deepak Advani, who heads IBM’s predictive analytics group.”

“Big Data has arrived at Seton Health Care Family, fortunately accompanied by an analytics tool that will help deal with the complexity of more than two million patient contacts a year…”

“Companies are being inundated with data—from information on customer-buying habits to supply-chain efficiency. But many managers struggle to make sense of the numbers.”

“Data is the new oil.”

Clive Humby

“The New York Times

“Increasingly, businesses are applying analytics to social media such as Facebook and Twitter, as well as to product review websites, to try to understand where customers are, what makes them tick and what they want,” says Deepak Advani, who heads IBM’s predictive analytics group.”

“Big Data has arrived at Seton Health Care Family, fortunately accompanied by an analytics tool that will help deal with the complexity of more than two million patient contacts a year…”

“Companies are being inundated with data—from information on customer-buying habits to supply-chain efficiency. But many managers struggle to make sense of the numbers.”

“Data is the new oil.”

Clive Humby

“The New York Times

“Increasingly, businesses are applying analytics to social media such as Facebook and Twitter, as well as to product review websites, to try to understand where customers are, what makes them tick and what they want,” says Deepak Advani, who heads IBM’s predictive analytics group.”

“Big Data has arrived at Seton Health Care Family, fortunately accompanied by an analytics tool that will help deal with the complexity of more than two million patient contacts a year…”

“Companies are being inundated with data—from information on customer-buying habits to supply-chain efficiency. But many managers struggle to make sense of the numbers.”

“Data is the new oil.”

Clive Humby

“The New York Times

“Increasingly, businesses are applying analytics to social media such as Facebook and Twitter, as well as to product review websites, to try to understand where customers are, what makes them tick and what they want,” says Deepak Advani, who heads IBM’s predictive analytics group.”

“Big Data has arrived at Seton Health Care Family, Fortunately accompanied by an analytics tool that will help deal with the complexity of more than two million patient contacts a year…”

“Companies are being inundated with data—from information on customer-buying habits to supply-chain efficiency. But many managers struggle to make sense of the numbers.”

“Data is the new oil.”

Clive Humby

“The New York Times

“Increasingly, businesses are applying analytics to social media such as Facebook and Twitter, as well as to product review websites, to try to understand where customers are, what makes them tick and what they want,” says Deepak Advani, who heads IBM’s predictive analytics group.”

“Big Data has arrived at Seton Health Care Family, Fortunately accompanied by an analytics tool that will help deal with the complexity of more than two million patient contacts a year…”

“Companies are being inundated with data—from information on customer-buying habits to supply-chain efficiency. But many managers struggle to make sense of the numbers.”

“Data is the new oil.”

Clive Humby

“The New York Times

“Increasingly, businesses are applying analytics to social media such as Facebook and Twitter, as well as to product review websites, to try to understand where customers are, what makes them tick and what they want,” says Deepak Advani, who heads IBM’s predictive analytics group.”
без анализа

Big Data
просто куча данных

Миф: Big Data – только большие базы данных; просто больше чем были раньше

Миф: Big Data означает Hadoop.. и все

Миф: Big Data означает ‘выбросьте и замените’ все что было раньше

Миф: NoSQL означает НЕТ SQL, никогда, забудьте про SQL

Миф: Big Data означает неструктурированные данные и для анализа настроений
In 2005 there were 1.3 billion RFID tags in circulation…
Сенсоры и датчики в современной среде генерируют ОГРОМНЫЕ объемы данных с МАШИННЫМИ СКОРОСТЯМИ…

1 МИЛИАРД строк кода КАЖДЫЙ двигатель - 10 ТВ каждые 30 минут!
6,000,000 пользователей Twitter создавали 300,000 сообщений в день.

500,000,000 пользователей Twitter создавали 400,000,000 сообщений в день.

83x

1333x

Ожидайте большего от ваших систем хранения.
Автоматическое обогащение данных
<table>
<thead>
<tr>
<th>Google</th>
<th>Google прекратил сообщать как много данных они хранят в 2010 (SEC filing): в то время это было 100 PBs</th>
</tr>
</thead>
<tbody>
<tr>
<td>YouTube</td>
<td>YouTube – порядки измеряются в Exabyte</td>
</tr>
<tr>
<td></td>
<td>• 72+ ч видео загружаются на YouTube каждую минуту</td>
</tr>
<tr>
<td></td>
<td>• YouTube второй по использованию поисковый движок после Google</td>
</tr>
<tr>
<td></td>
<td>• Последние данные 768+ PBs, 3-4 года назад: точно больше Exabyte сейчас</td>
</tr>
<tr>
<td>Facebook</td>
<td>Facebook перевалил за миллиард пользователей в августе 2012</td>
</tr>
<tr>
<td></td>
<td>• Население планеты стало больше 7B в прошлом году: 1/6\textsuperscript{th} – в Facebook</td>
</tr>
<tr>
<td></td>
<td>• 35% мировых фотографий по оценкам в Facebook</td>
</tr>
<tr>
<td>Twitter</td>
<td>Twitter - около 124 млрд tweets в год, в среднем 4500 в сек</td>
</tr>
<tr>
<td>Обмен сообщениями в мире</td>
<td>Обмен сообщениями в мире 193,000 смс/сек</td>
</tr>
<tr>
<td>Звонки в США: 2.2 трилиона минут в год; 19 мин/день/человека</td>
<td>Звонки в США: 2.2 трилиона минут в год; 19 мин/день/человека</td>
</tr>
<tr>
<td></td>
<td>• В несжатом виде это год YouTube</td>
</tr>
</tbody>
</table>
Характеристики больших данных

Растущий Объем

Растущая Скорость

Увеличивающееся Разнообразие

1 из 3 руководителей не доверяет информации, которую он использует для принятия решений

30 Billion RFID sensors and counting

80% данных не структурировано
Дилемма больших данных

• Процентное соотношение данных, которые компания может анализировать уменьшается по отношению с доступным данным
  
  • Проще говоря, как организация, мы становимся “более наивными” со временем

• Мы не знаем что мы могли бы знать….
Большие данные – горячая тема, потому что технологии сделали возможным анализ ВСЕХ доступных данных.

Эффективно с точки зрения затрат управлять и анализировать все доступные данные,

в их первозданном виде – структурированные, неструктурированные, потоковые.
Почему большие данные не появились раньше: разве у нас не было инструментов?
Взаимное дополнение подходов

Традиционная аналитика
Структура и повторяемость

Гипотеза → Вопрос → Ответ

Старт с гипотезы
Проверка по выбранным данным

Analyzed Information

Big Data аналитика
Исследование

Данные → Исследование

Данные определяют путь
Исследуем все, выявляем связи

All Information

Анализ после сохранения...

Анализ на лету...
Применение анализа больших данных

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

**Энергетика**
- Влияние погody 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
Отдельные системы отвечают на вопросы, по одному

“Что мы можем еще ей продать?”
“Как состояние поставок может повлиять на бизнес?”
“Какие материалы лучше послать?”
“Что мне лучше знать перед звонком с предложением продлить поддержку?”

“Что происходит с клиентом СЕГОДНЯ?”
“Как нам улучшить взаимодействие?”

Как нам получить больше таких клиентов?

...Но! 360º взгляд дает ответы в одном месте

Смещение данных из разных систем дает лучшее понимание — не просто факты

Ожидайте большего от ваших систем хранения
Радиологисты перегружены медицинскими снимками.
Усталость глаз, ошибки в диагнозах.
Помощь ценному ресурсу.
Телерадиология.
Ситуационный анализ в реальном времени

- Захватить данные с датчиков о погоде
- Проанализировать предполагаемую траекторию урагана
- Оценить влияние
- Дать рекомендации и уведомления
- Рассчитать социальную помощь и затраты на ремонт
- Провести симуляцию ущерба с вариантами ветра, осадков +++
Прогноз траектории в реальном времени

Динамически обновляемый профиль риска

Сопоставление погодных угроз и профилей риска для расчёта потенциального ущерба

Прогноз траектории в реальном времени
Ключевые компоненты портфеля IBM Big Data

**BIG DATA PLATFORM**
- Systems Management
- Application Development
- Discovery
- Hadoop System
- Stream Computing
- Data Warehouse
- Information Integration & Governance

**InfoSphere Data Explorer**
Поиск, навигация, визуализация всех данных

**Accelerators**
Аналитические функции
Аналитические приложения

**InfoSphere BigInsights**
Возможности Hadoop для предприятий

**InfoSphere Streams**
Анализ потоковых данных

**PureData for Analytics & InfoSphere Warehouse**
Глубокий анализ структурированных данных

**Information Integration and Governance**
Управление качеством и жизненным циклом информации

Ожидайте большего от ваших систем хранения
Appliance для Big Data

For Hadoop
Optimized system to accelerate big data analytics and online archive with appliance simplicity

For Analytics
Optimized system delivering data services for analytics & reporting

For Operational Analytics
Optimized system delivering data services for operational analytics

For Transactions
Optimized system delivering data services for transactions
THINK BIG
Thank You