

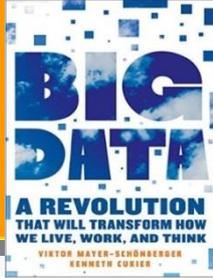
Больше данных, хороших и разных!



Большие данные 2018 (CNews)

Москва, 19 апреля 2018 года

Артемьев Валерий Иванович (Банк России)



Больше данных, хороших и разных!

Структурированные данные

- Сгенерированные машиной
 - Данные датчиков и IoT
 - Поток событий
 - Логи
 - Операционные данные
 - Отчётные данные
 - Реестровые данные
- Сгенерированные человеком
 - Данные ручного ввода
 - Click-данные
 - Данные игр

Неструктурированные данные

- Сгенерированные машиной
 - Данные в формате обмена
 - Изображения со спутника
 - Научные данные
 - Фото и видео
 - Данные радара/сонара
- Сгенерированные человеком
 - Тексты и офисные документы
 - Социальные медиа
 - Мобильные данные
 - Web-контент

Можно ли говорить о качестве данных неструктурированных данных

Big data – не только неструктурированная информация

Датификация и цифровизация – многообразие источников

Можно идти от ДАННЫХ, но главное решаемые ЗАДАЧИ!

Больше аналитики, хорошей и разной!



Бизнес-аналитика (Business Intelligence)

- Производственная отчётность (Reporting)
- Произвольные запросы (Ad hoc query)
- Деловая графика
- Многомерный анализ (OLAP)
- Информационные панели (Dashboarding)
- Оценочные карты (Scorecarding)
- Визуальное исследование данных (Data discovery)

Интеллектуальный анализ данных (Data mining)

- Машинное обучение
- Классификация
- Регрессионные модели
- Кластеризация
- Ассоциативные модели
- Нейронные сети
- Временные последовательно
- Поиск существенных атрибутов
- Модели исключений
- Выделение признаков

Продвинутая аналитика (Advanced analytics)

- Статистический анализ
- Анализ временных рядов
- Прогнозирование
- Моделирование
- Риск-анализ (Value at risk, ...)
- Глубинное обучение нейросетей
- Анализ текстов и полуструктурированной информации
- Распознавание образов
- Рекомендации
- Анализ тональности контента
- Продвинутая визуализация
- Предписывающая аналитика

Самообслуживание бизнес-аналитика.
Фокус на визуальное исследование

Упрощение работы бизнес-аналитика.
Фокус на предсказательную аналитику

Самообслуживание команды data scientist. Фокус на операционную, предсказательную и, возможно, предписывающую аналитику

Data scientist – и швец и жнец и на дуде игрец?



Data science – раздел информатики, изучающий методы и средства анализа, обработки и представления данных (включая Big data)

- ☑ **Аналитик?** – Несомненно!
- ☑ **Статистик и математик?** – Да
- ☑ **Лингвист?** – Самую малость
- ☑ **Спец по интеграции данных?** – Да
- ☑ **Спец по базам данных?** – Конечно!
- ☑ **Программист?** – Без этого никак!
- ☑ **Системщик?** – Приходится
- ☑ **Бизнес-аналитик?** – Хотелось бы

Методы и средства бизнес-аналитики,
data mining и продвинутой аналитики
Основы математики и математической
статистики

Азы лингвистики и особенности анализа
текстов (text mining)

Интеграция данных и обеспечение качества
данных (ELT, Data federation, DQA)

Управление базами данных SQL и NoSQL

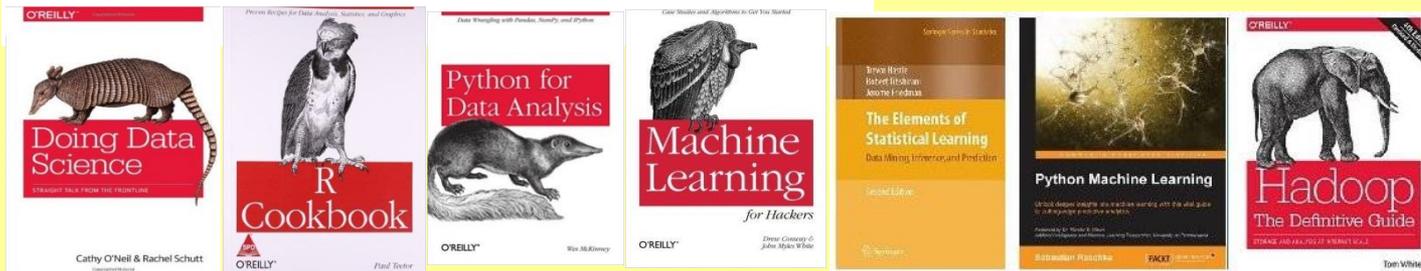
Среды программирования Python, R, ...

Распределённая обработка данных, потоков
и событий

Распределённая файловая система HDFS

Файловая система и команды Linux

Моделируемые бизнес-процессы/функции



Больше навыков,

хороших и разных!

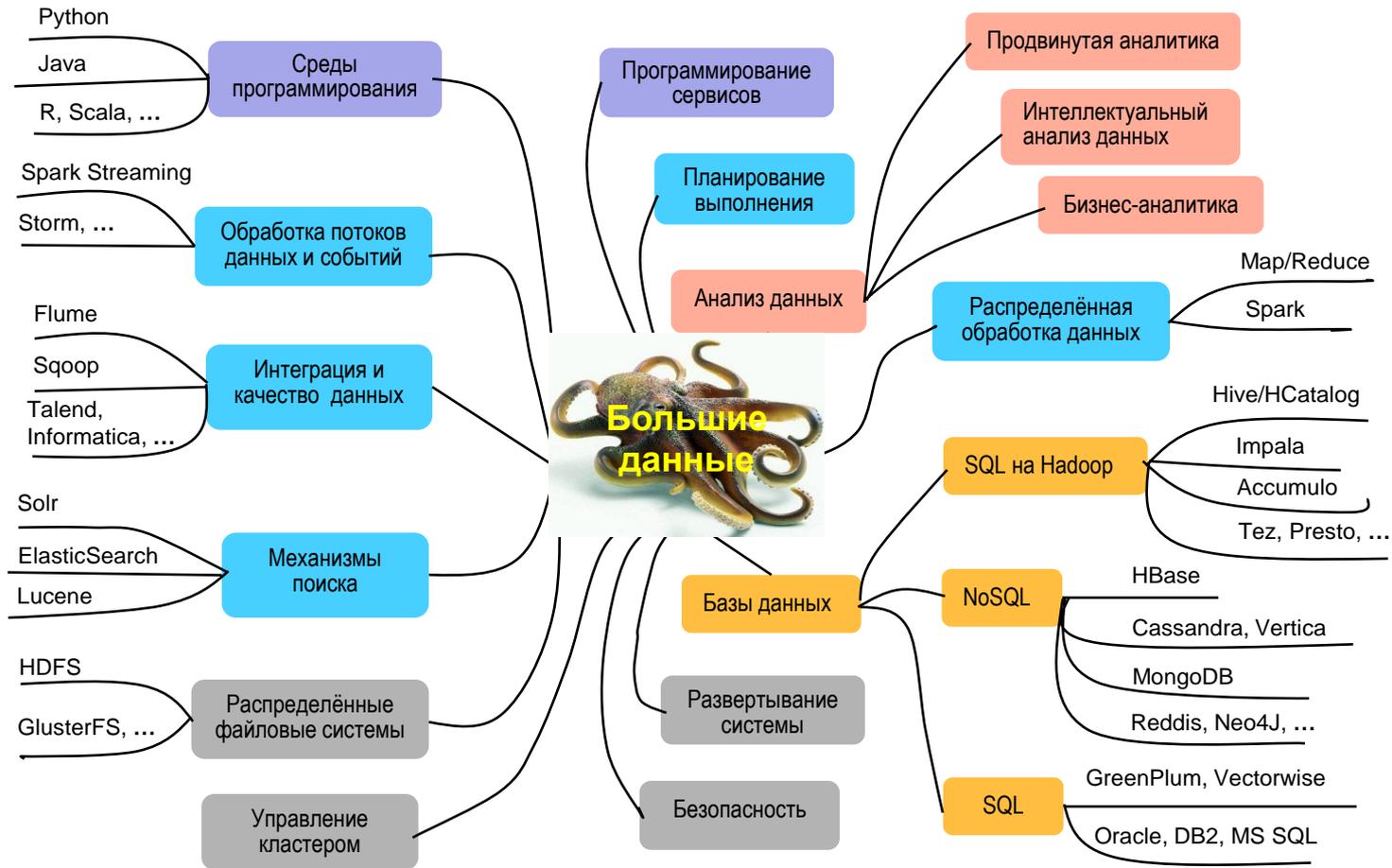


Data Science – командная игра. Важен специалист от бизнеса

Совмещение ролей – больше зарплата специалиста

Через лабораторию данных к DataOps=Agile+DevOpS+BigData

Больше инструментов, хороших и разных!



Преобладание OpenSource.
Экосистемы вокруг Hadoop/MR
и Spark. От batch к real time

**Влияние на традиционное
управление данными,
интеграцию и анализ данных**

Имеются решения в облаках.
Появились архитектурные
шаблоны Big data

Новые архитектурные стили



Несколько слов на посошок

- Бизнес–проекты, а не технологии ради технологий
- Не ищите большие данные, ищите и решайте серьёзные аналитические задачи
- Повышайте осведомлённость и компетенции по Data science и Big data
- Экспериментируйте, если можете себе это позволить
- Осторожно, аналитика – коварная штука
- Большие данные нужны далеко не всем
- Не хакеры–аналитики, а команда единомышленников (аналитики + бизнес + ИТ)
- Инструменты Data science и Big data со временем станут менее изощрёнными
- Big data станет пройденным этапом и исчезнет

Спасибо за внимание.

Ответы на вопросы

Департамент статистики и управления данных
Банка России

Валерий Иванович Артемьев
