

Большие данные – многие печали

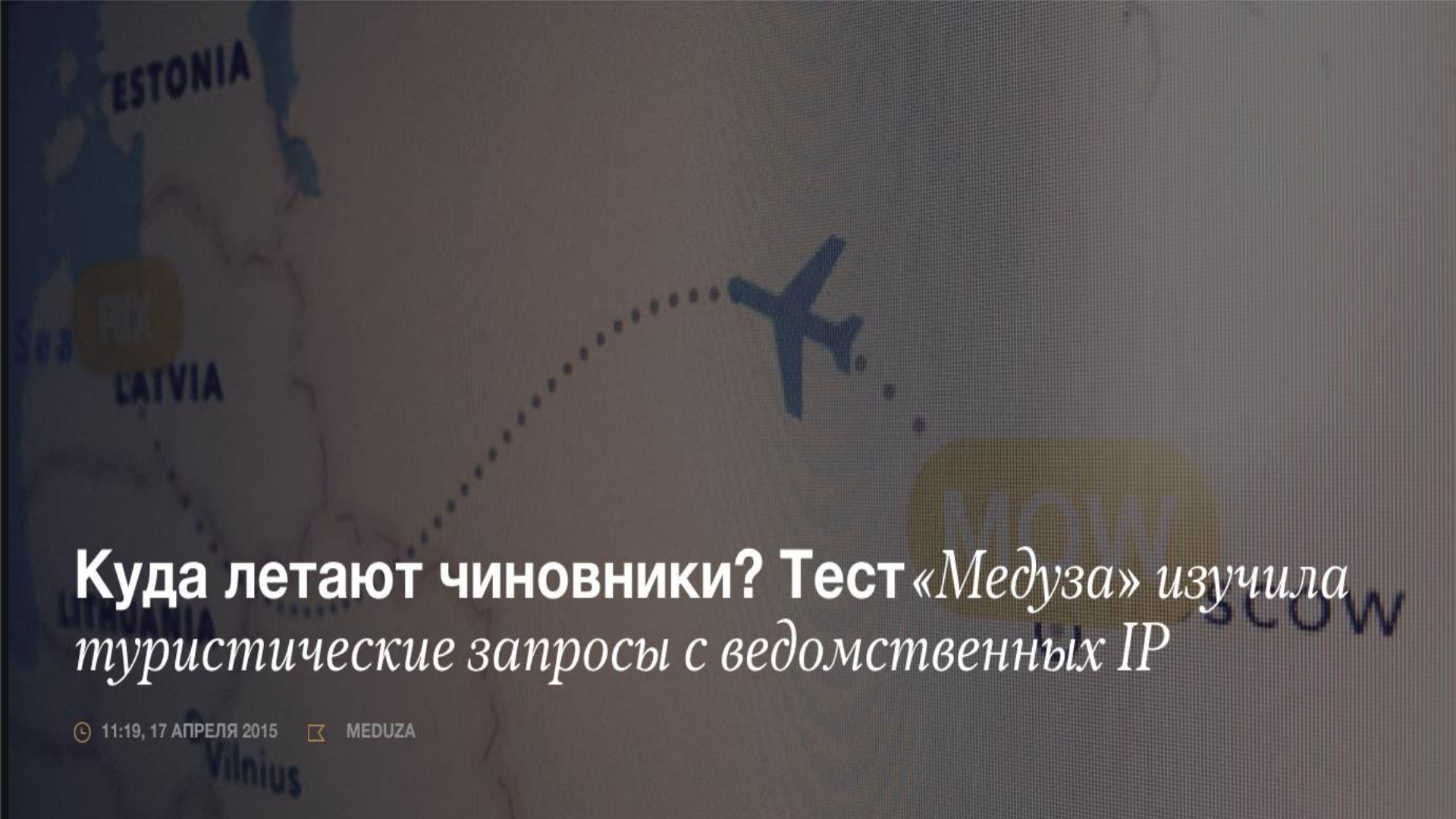
Ложечка дегтя в бочке преимуществ Big Data
для e-Commerce

Денис Безкоровайный, CISA, CISSP, CCSK
Вице-президент RISSPA

РИФ 2015
23 апреля 2015

Принципы сбора данных

- Де-идентификация (анонимизация)
- Агрегация



Куда летают чиновники? Тест «Медуза» изучила туристические запросы с ведомственных IP

🕒 11:19, 17 АПРЕЛЯ 2015 📄 MEDUZA

1

Куда больше всего тянет сотрудников Следственного комитета?

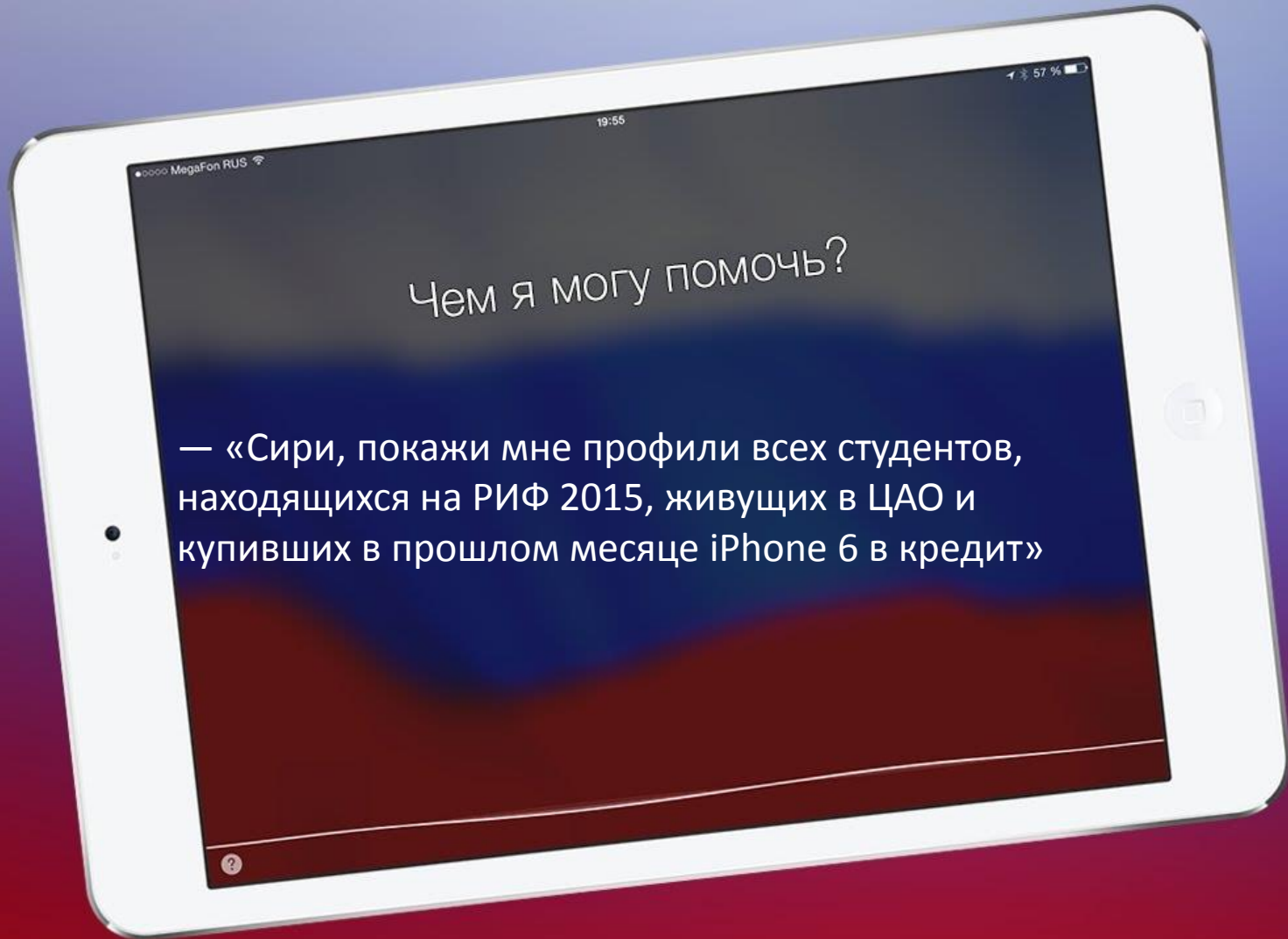


- В Сочи, конечно!
- Россия — для работы, а для отдыха — Шарм-эль-Шейх
- В Америке все плохо, но отдыхать уж больно приятно. Лос-Анджелес

Угроза приватности

Корреляция данных разных типов:
медицинские, финансовые,
демографические, геолокационные

Даже больше, чем персональные данные



Чем я могу помочь?

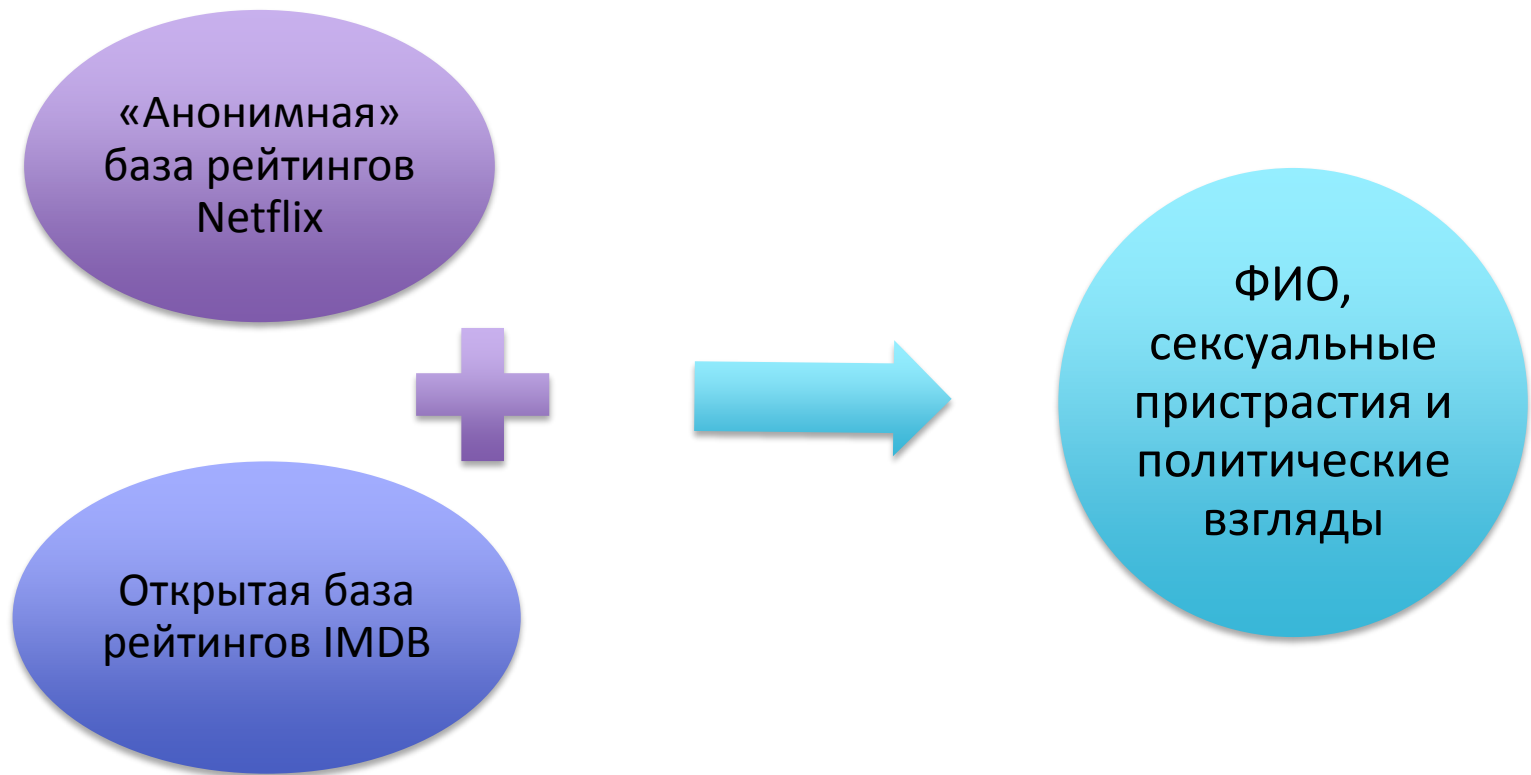
— «Сире, покажи мне профили всех студентов, находящихся на РИФ 2015, живущих в ЦАО и купивших в прошлом месяце iPhone 6 в кредит»

Риски персонализации

Отслеживание поведения потребителей и передача третьим сторонам **без авторизации** для таргетинга и других целей

«Я тебя по IP вычислю»

Де-анонимизация





Ценовая дискриминация



Пример дискриминации в онлайн-рекламе

Maryland

HOUSE OF DELEGATES

AISHA N. BRAVEBOY

Democrat, District

House Office Building

6 Bladen St., Annapolis

(410) 841-3707, (410) 841-3708

1-800-492-7122, cell: (410) 841-3709

e-mail: aisha.braveboy@maryland.gov

fax: (410) 841-3844

Ads by Google

[Aisha braveboy: Truth](#)

Arrests and Much More. Everything About Aisha braveboy
www.instantcheckmate.com/

[Sycle.net Patient Mngmt](#)

#1 Patient Management software in the hearing industry.
www.sycle.net/

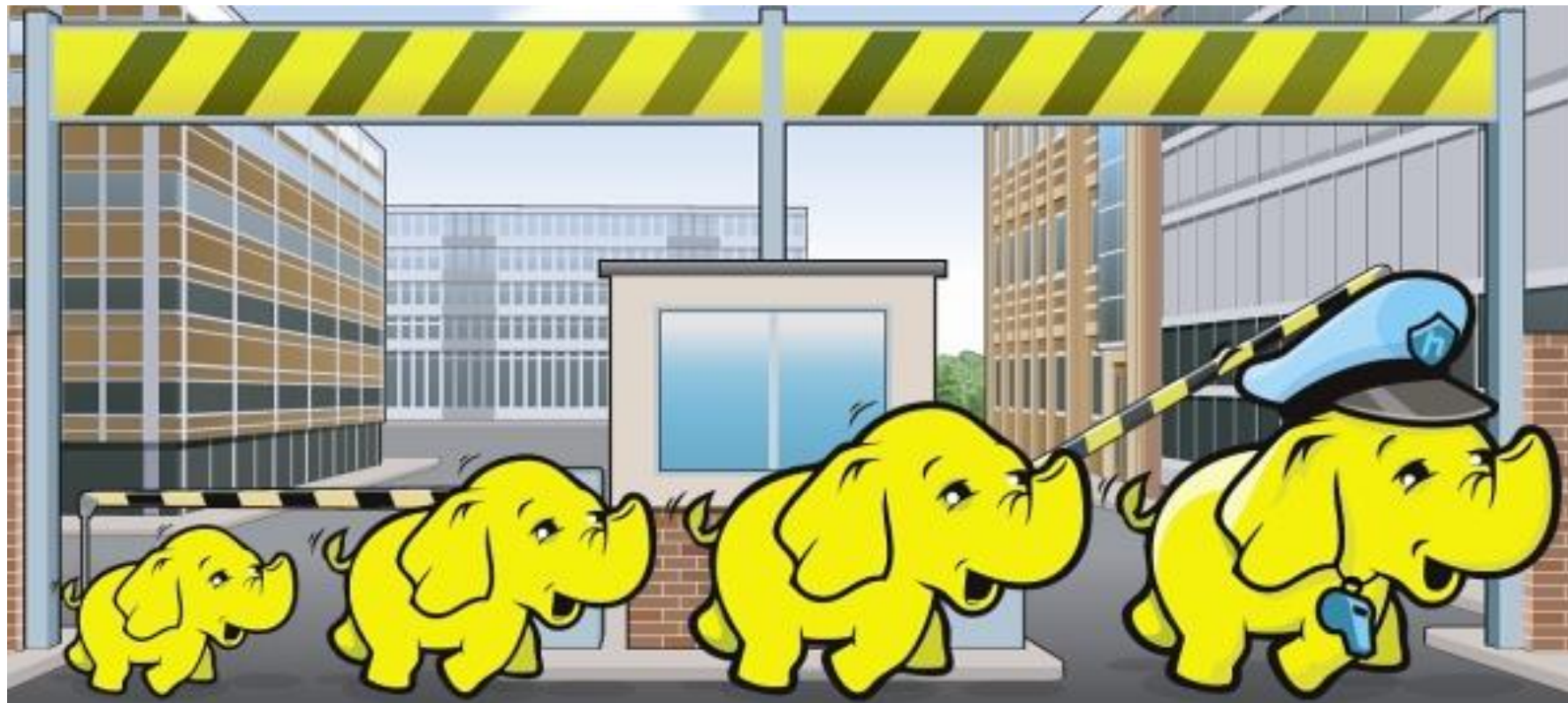
Регулирование

- Только защита «малых данных»
 - Закон о персональных данных
 - ФЗ 242
- Нет регулирования больших данных (пока)
- Что в мире?

Большие проблемы с информационной безопасностью

- Безопасность не была приоритетом
 - Вспомните TCP/IP (Hadoop)
- Часто изначальная задача – анализ
открытых данных
 - MapReduce

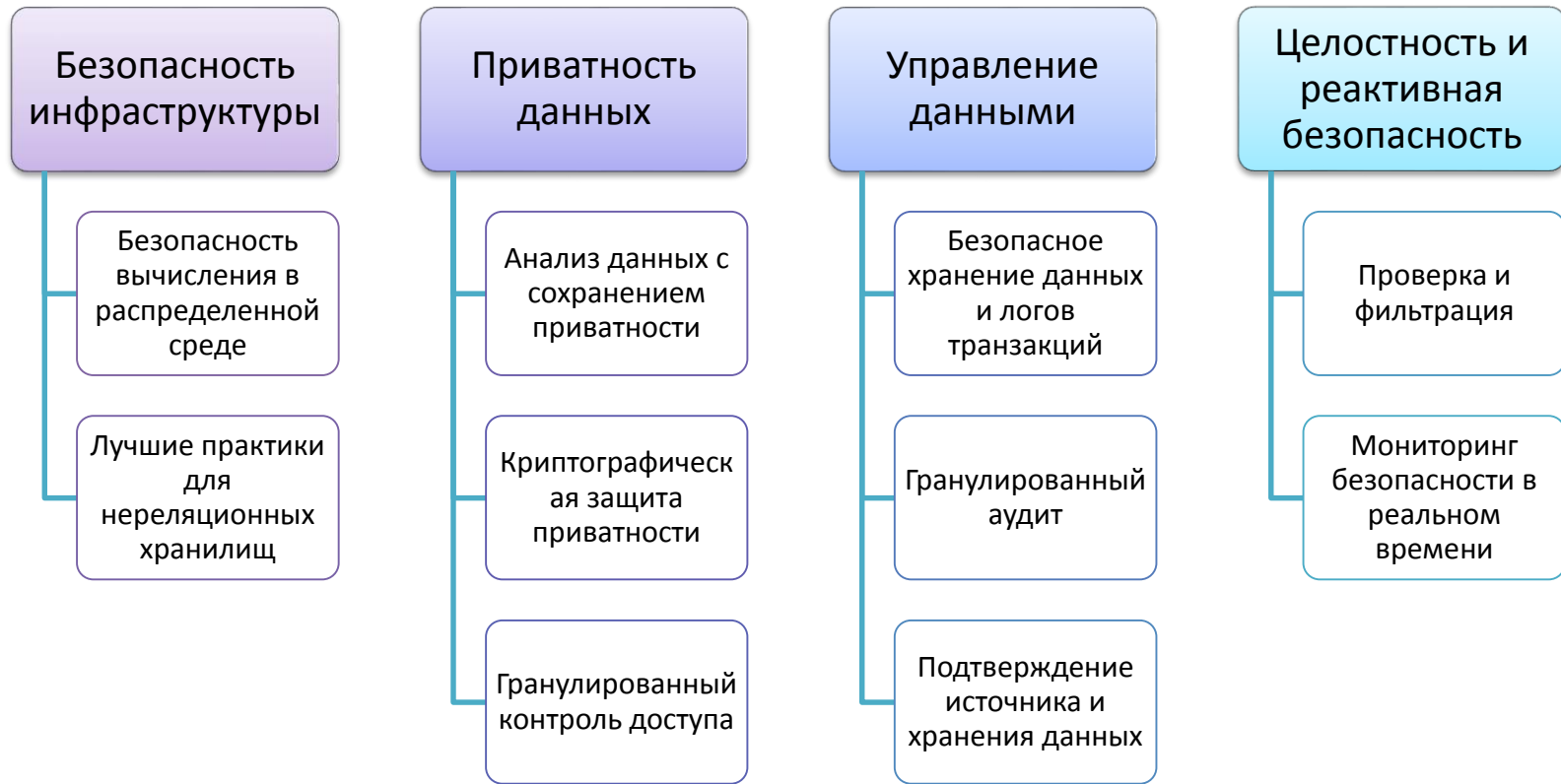
Защита инфраструктуры больших данных



Защита инфраструктуры больших данных

- Другая инфраструктура – другие защитные механизмы
- Расчет на сторонние компенсирующие меры (сложно и дорого)
- Множество хранилищ данных
- Недостаточно лучших практик

Топ 10 задач по обеспечению информационной безопасности больших данных (CSA)



Брокеры и агрегаторы данных – все непрозрачно

- Где хранит данные?
- Как обрабатывает?
- Кто имеет доступ к данным? С какой целью?
- Кому еще продаются данные ваших клиентов?

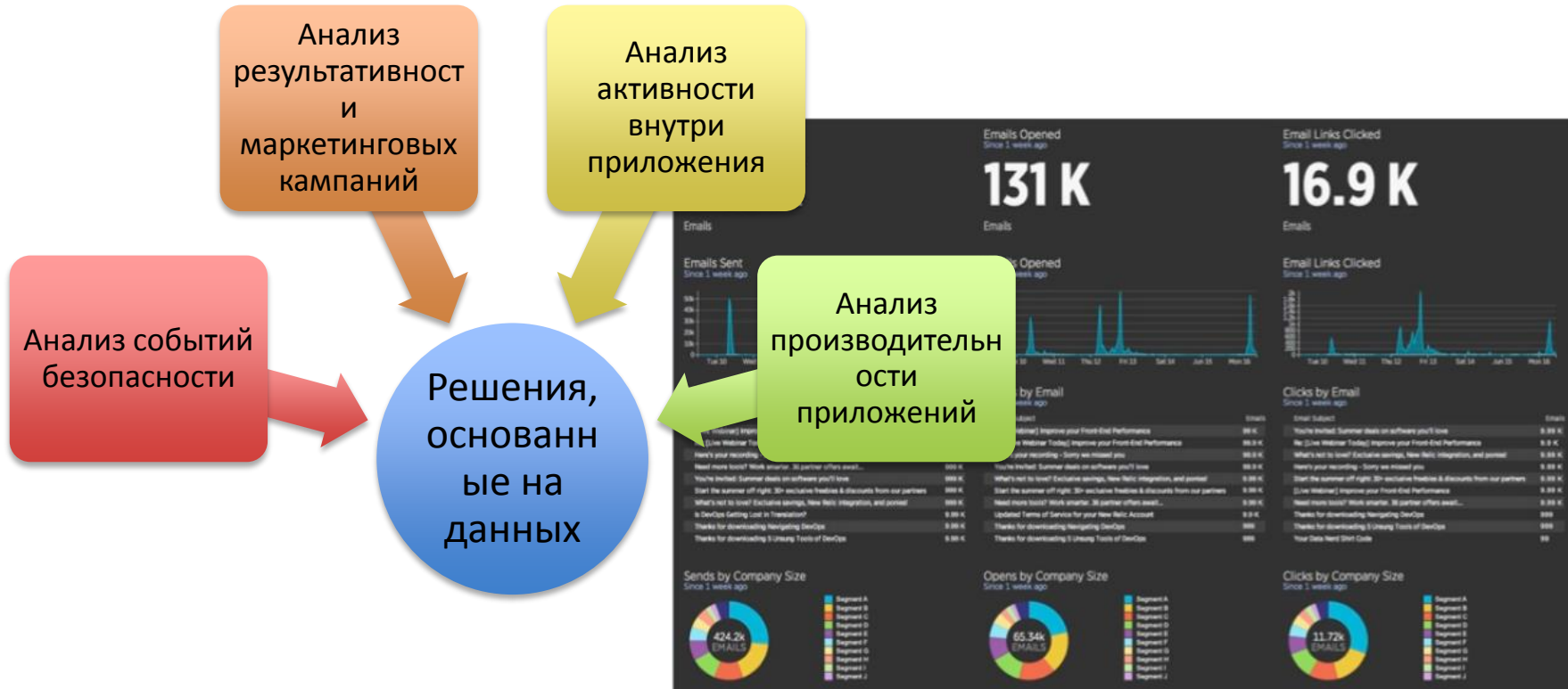
Обработка больших данных как сервис

- Самому или сервис?
- Выбор поставщика – вопросы безопасности и юрисдикции
- Какие данные собирает поставщик? Насколько эти данные анонимны?
- Утечки от брокеров – тень на поставщике данных (ответственность не аутсорсится)

Риски поставщиков и обработчиков больших данных

- Юридические
- Репутационные – потеря клиентов
 - Де-анонимизация, отнесение к когорте
 - Утечки данных (информационная безопасность)
 - Ценовая дискриминация
- Даже если вы используете внешний сервис!

Большие данные в E-commerce – это не только персонализация



Выводы

- Большие данные – новые риски для e-commerce
- Слабое регулирование (пока)
- Технологически защищать Big Data – сложнее
- Доверяете стороннему поставщику – кто он?
- Польза от больших данных не только в персонализации – множество путей увеличения прибыли

Вопросы?

Денис Безкоровайный

@bezkod

bezkod@risspa.ru