

Одиннадцатая независимая научно-практическая конференция «Разработка ПО 2015»

22 - 24 октября, Москва



От истории к векторам

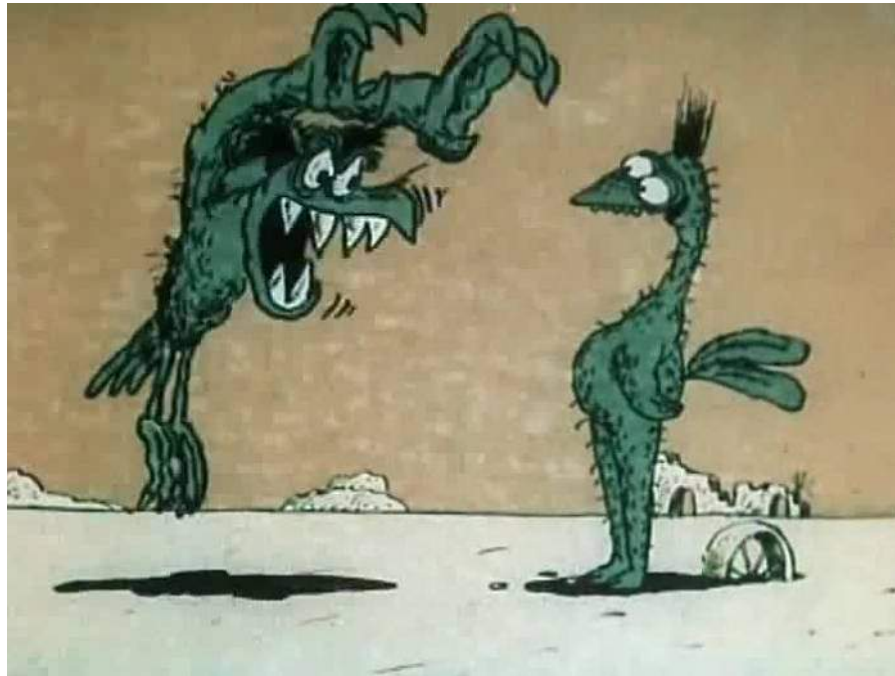
Хранилище данных для системы скоринга

Евгений Виноградов

Яндекс.Деньги

Про что это?

- Скоринг в реальном времени



Про что это?



Зачем это нужно

- Кредиты



Зачем это нужно

- Кредиты
- Безопасность



Зачем это нужно

- Кредиты
- Безопасность
- Аутентификация



Зачем это нужно

- Кредиты
- Безопасность
- Аутентификация
- Страхование

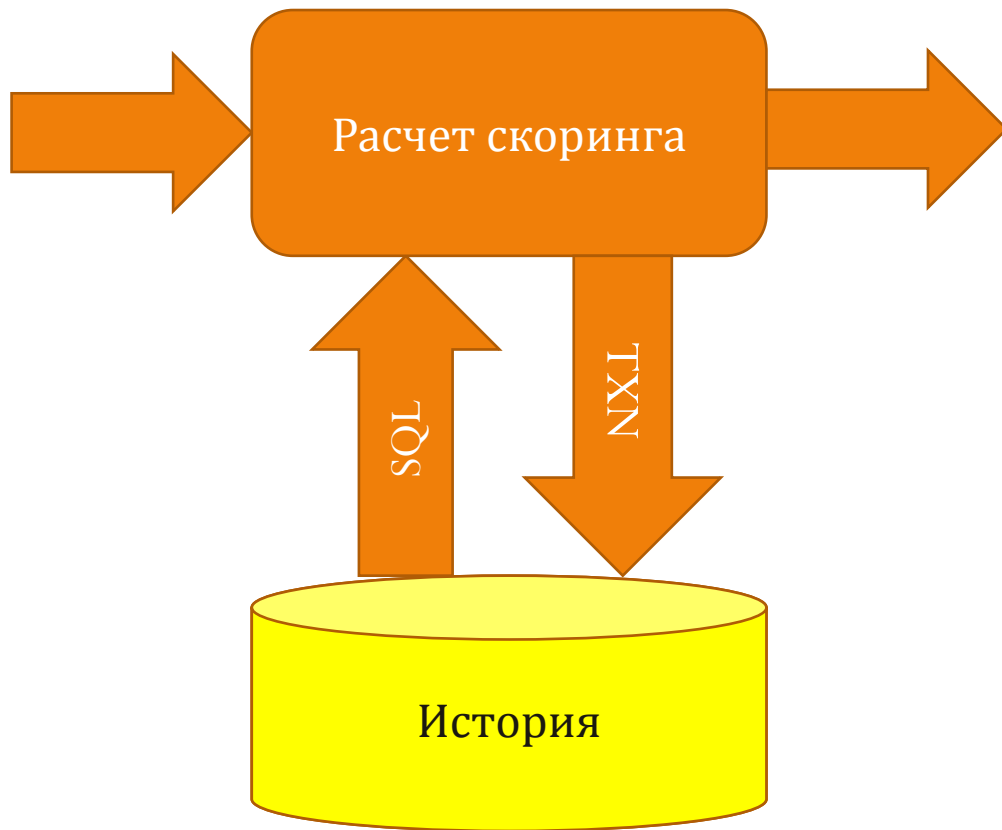
Зачем это нужно нам?



Скоринг в нашем случае

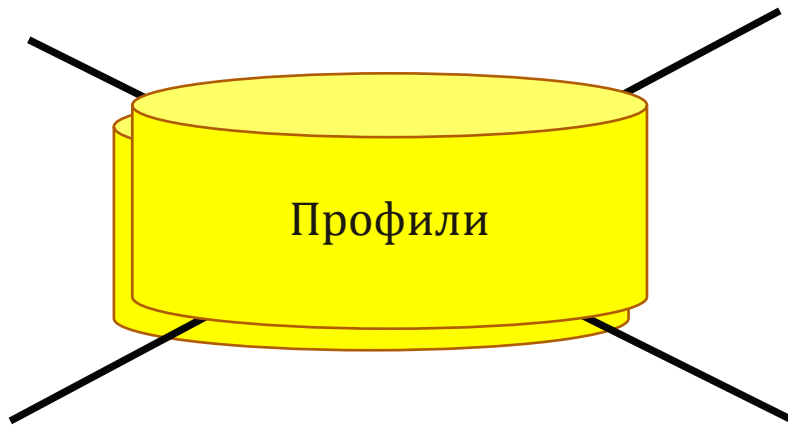
- Классификация
 - Статические профили
 - Динамические профили

С чего начиналось?



За что боремся?

- Скорость обработки
- Скорость модификации структуры профиля
- Масштабируемость
- Объем данных



Синхронно

Обработка

=

скоринг текущей транзакции

+

пересчет профиля

Однотипно

- Профиль = тип + ключ + вектор
- Поиск – только по типу и ключу
- Вектор = остальные атрибуты профиля

Однотипно

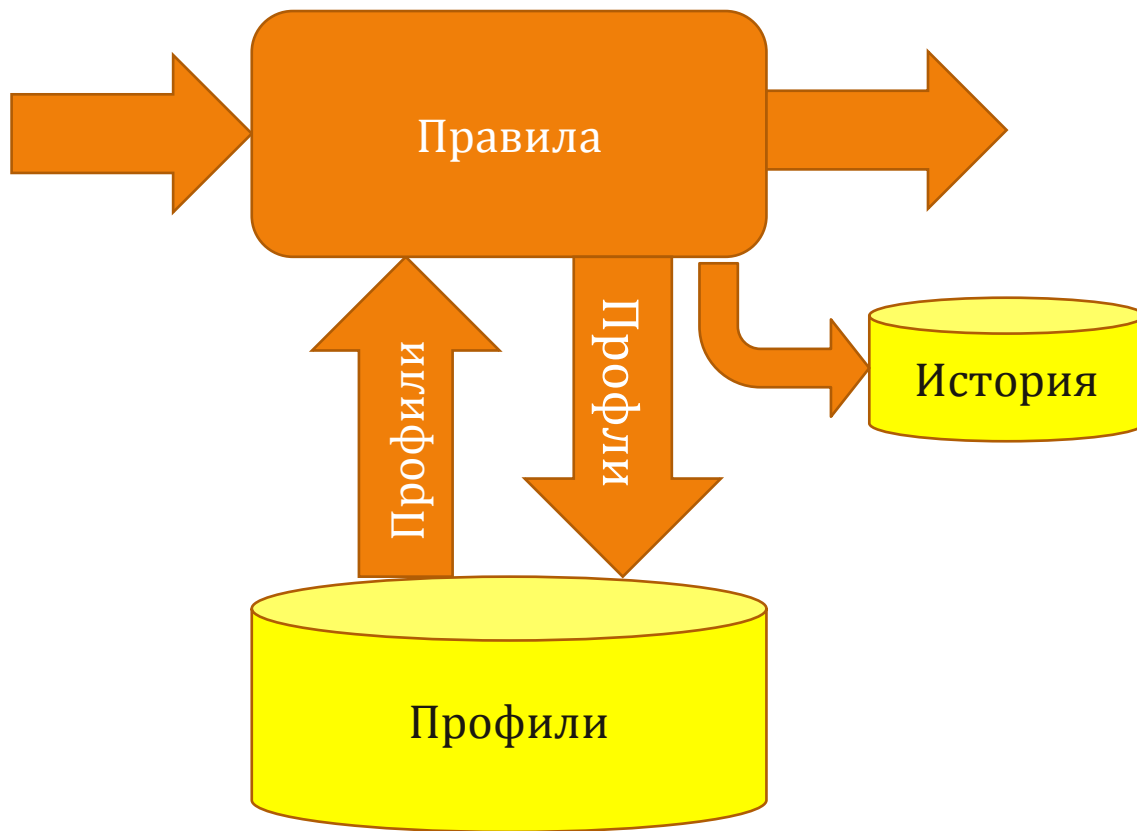
```
<PhoneProfile xmlns:i="http://www.w3.org/2001/XMLSchema-instance">
  <Payments>
    <Item>
      <Number>50.0</Number>
      <Time>2015-09-11T13:41:54.1648194+03:00</Time>
      <UpdateDate>2015-09-11</UpdateDate>
      <Value>41001111223333</Value>
    </Item>
  </Payments>
  <Bindings>
    <Item>
      <Value>41001111223333</Value>
    </Item>
  </Bindings>
</PhoneProfile>
```

|

Система правил

- Левая часть – вычисления
- Правая - модификации
 - Обновление профиля
 - Обновление скоринга
- Если профиля нет – он создается

Как оно работает?

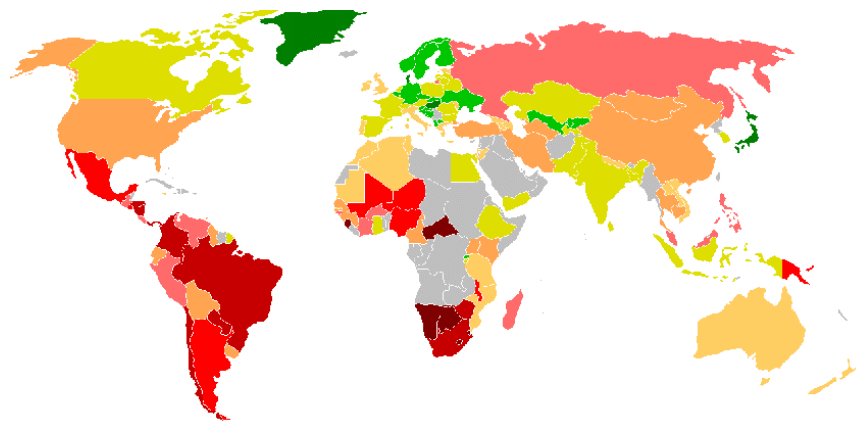


Ключи

- Натуральные ключи
 - Счет
 - Телефон
 - Адрес+счет

Ключи

- Шардинг по ключам
 - География



Особенности

- Шардинг по ключам
 - География
 - Математика
 - ...

Минимизируем объем

- Читаем и пишем только профили
- Нет или сложная развертка

Минимизируем объем

- Срок жизни профиля



Минимизируем объем

- Срок жизни профиля
- Старт с нуля



Минимизируем объем

- Внешнее хранилище
- Обогащение
 - Через транзакции
 - Отдельный тип
- Верификация

«Компрессия правил»

- Корреляция
- Опять же, внешнее хранилище
- Упрощение правил, минимизация вычислений
- Только экспериментально

Спасибо за внимание!

Вопросы?

- [linkedin.com/in/evinogradov](https://www.linkedin.com/in/evinogradov)
- jonny@yamoney.ru