

# СБОР, ХРАНЕНИЕ И ПОТОКОВЫЙ АНАЛИЗ ДАННЫХ СЕНСОРОВ ANDROID СМАРТФОНОВ

Мануилов Г., 13641/2  
Лесик Д., 13641/2

# МОТИВАЦИЯ

- Поведенческая биометрия
- Задача сбора датасета для создания алгоритма
  - Была решена с помощью RethinkDB и мобильных драйверов, собрано ~28 Гб данных
- Возможность переиспользования инфраструктуры

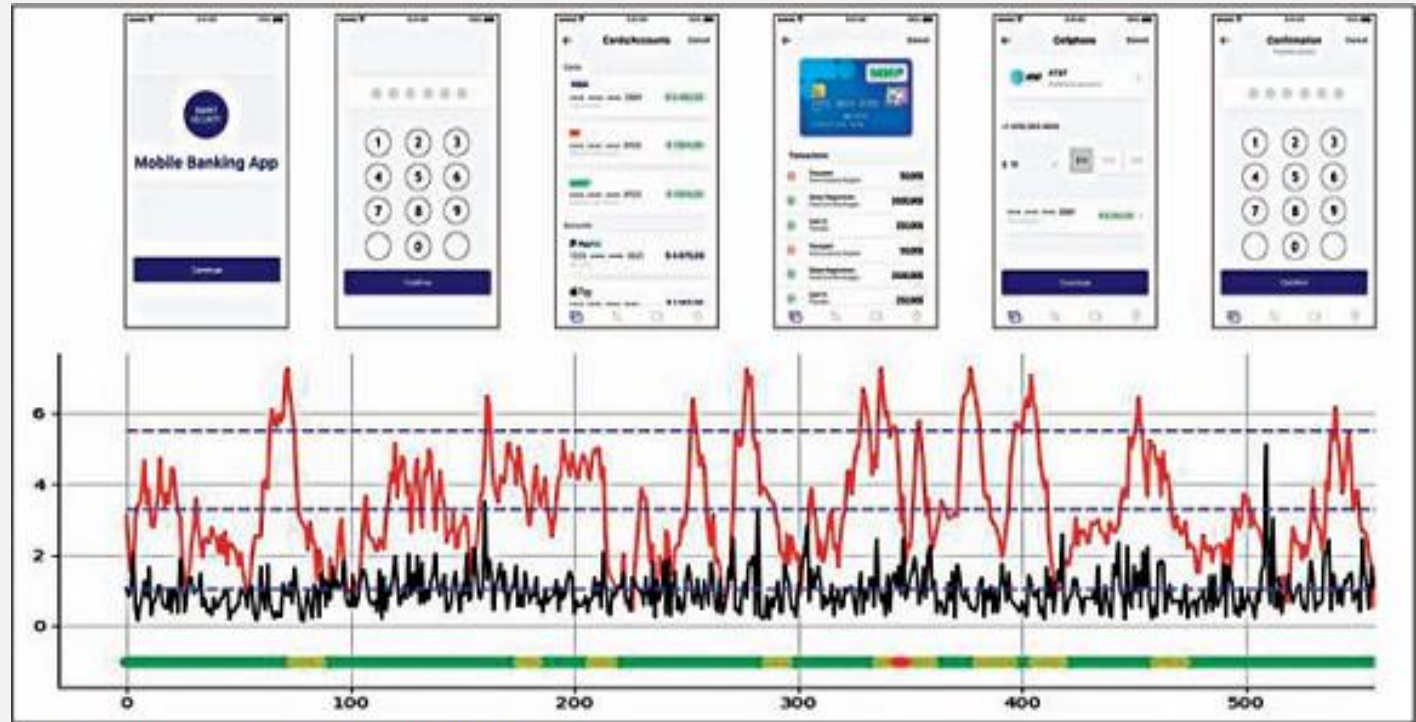


Рис. 7. Непрерывная аутентификация

## ЗАДАЧИ РАБОТЫ

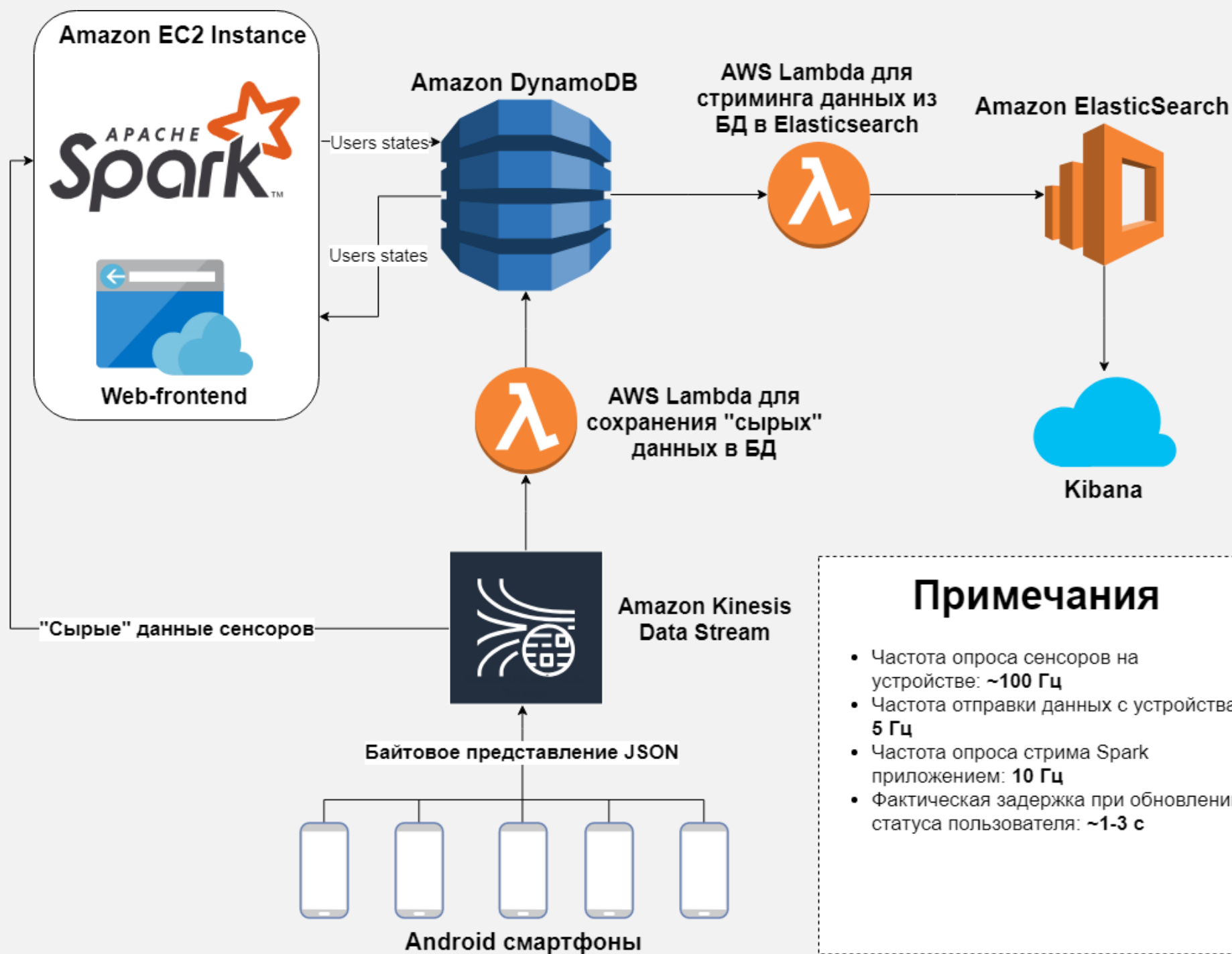
- Непрерывный потоковый сбор и сохранение данных сенсоров Android смартфонов
- Непрерывный анализ характера действий пользователя (идет, стоит, звонит и т.п.) в реальном времени
- Непрерывное формирование статистики по полученным данным
- *Непрерывная идентификация и аутентификация пользователя*

## РАЗРАБОТАННОЕ ПО

- ПО
  - Android приложение для сбора и отправки данных сенсоров
  - Android приложение для отладки модуля аналитики
  - Spark приложение для потоковой обработки данных
  - Веб-фронтенд для отображения состояний пользователей
  - Лямбда функции **AWS** для связывания компонентов системы
- Технологии и языки программирования
  - Amazon Kinesis Data Streams,
  - Apache Spark
  - AWS Lambda
  - Amazon DynamoDB,
  - Amazon Elasticsearch + Kibana
  - Java, Scala, Python

## НЕКОТОРЫЕ ТЕХНИЧЕСКИЕ АСПЕКТЫ

- Amazon Kinesis Data Streams: резюме
- AWS Lambda: резюме, сравнение с Azure Functions
- Spark и MapReduce в нашей задаче
- AWS Java SDK: наш личный рейтинг – 2/10

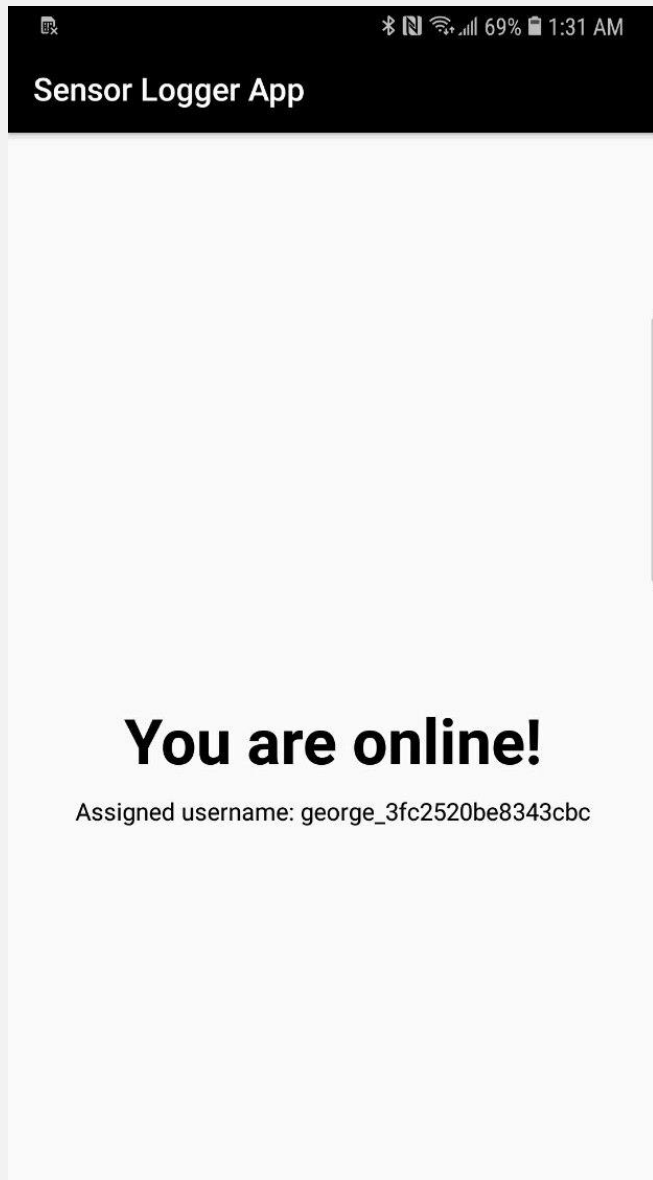


## Примечания

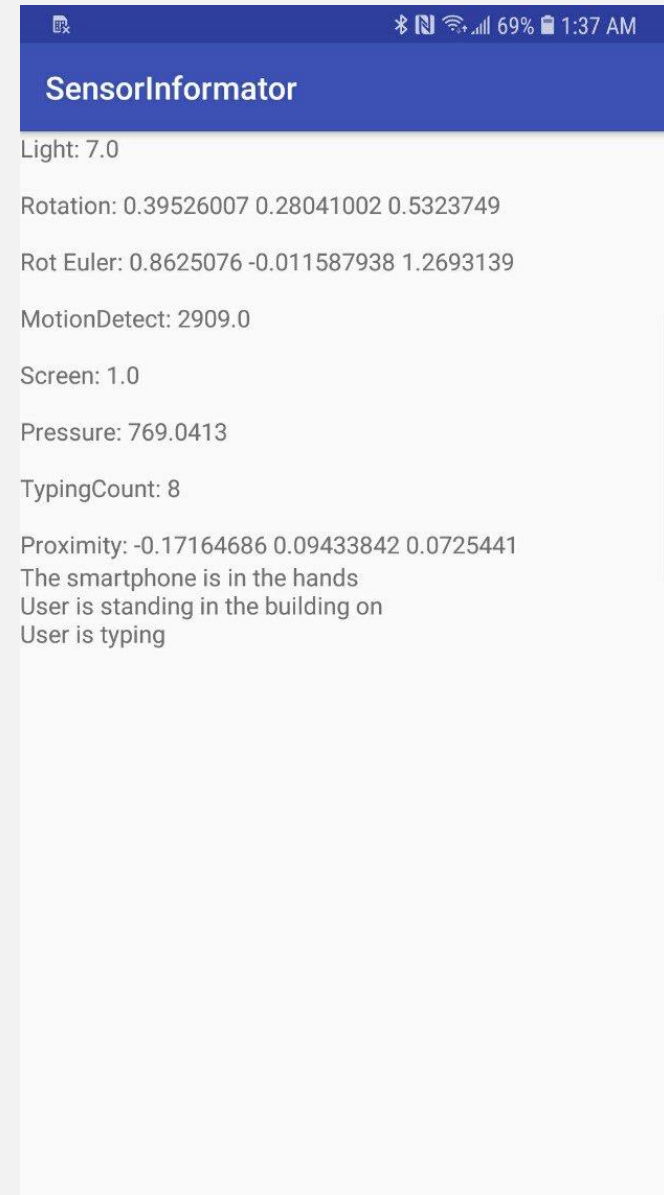
- Частота опроса сенсоров на устройстве: **~100 Гц**
- Частота отправки данных с устройства: **5 Гц**
- Частота опроса стрима Spark приложением: **10 Гц**
- Фактическая задержка при обновлении статуса пользователя: **~1-3 с**

# СБОР ДАННЫХ

## Sensor Logger App

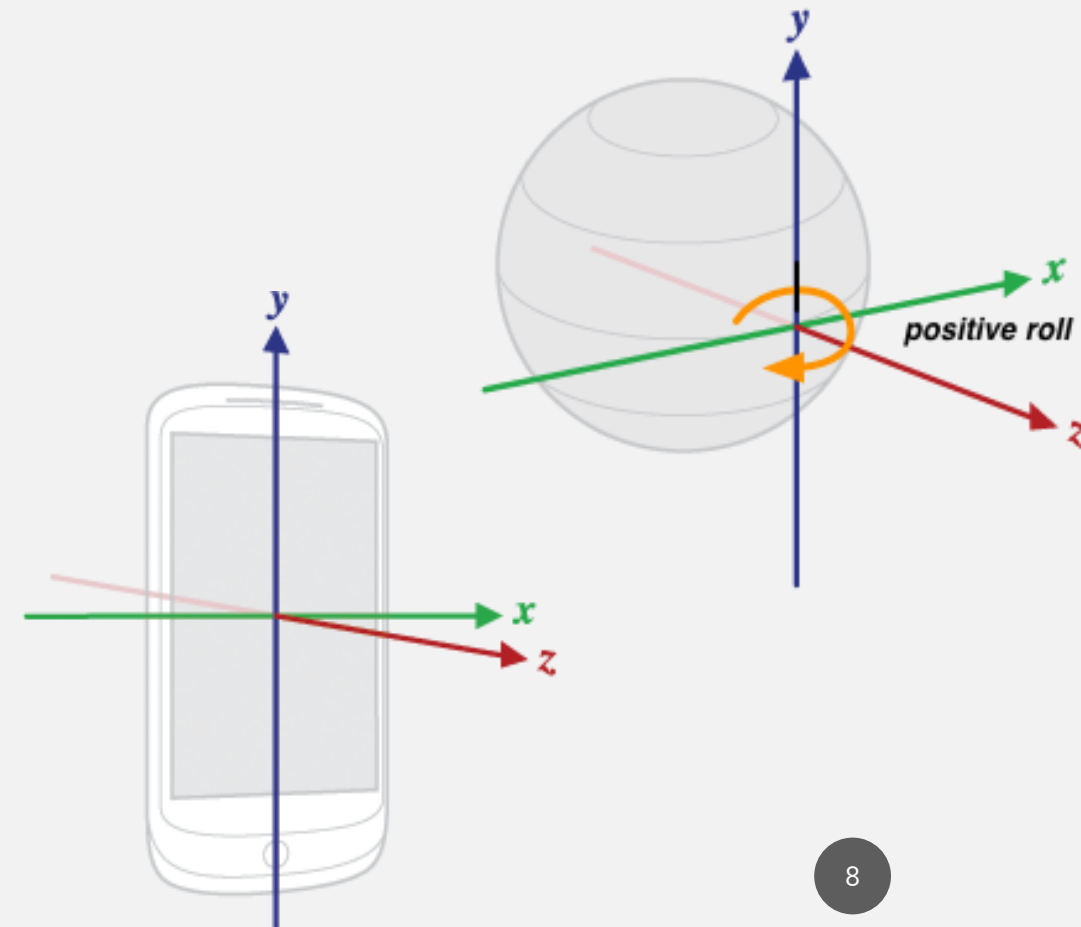


## Sensor Debugger App



# АНАЛИЗ: ДАННЫЕ

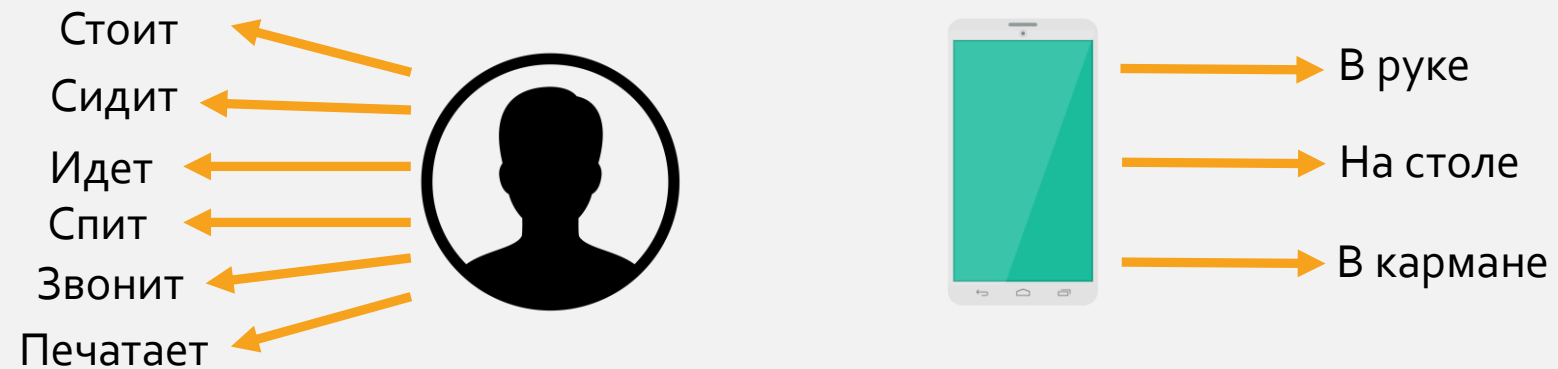
Сенсор	Значение	Тип
Акселерометр	Ускорение по трём осям, $\text{м/с}^2$	<code>float[3]</code>
Гироскоп	Скорость вращения устройства по трём осям, рад/с	<code>float[3]</code>
Барометр	Давление, мбар	<code>float[1]</code>
Датчик магнитного поля	Показатели магнитного поля по трём осям, мТл	<code>float[3]</code>
Датчик вектора поворота (виртуальный)	Положение устройства в пространстве в виде угла относительно оси	<code>float[4]</code>





# АНАЛИЗ: ИССЛЕДОВАНИЕ

- Выделение категорий



- Обработка сырых данных:









$$\begin{bmatrix} \phi \\ \theta \\ \psi \end{bmatrix} = \begin{bmatrix} \text{atan2}(2(q_0 q_1 + q_2 q_3), 1 - 2(q_1^2 + q_2^2)) \\ \text{asin}(2(q_0 q_2 - q_3 q_1)) \\ \text{atan2}(2(q_0 q_3 + q_1 q_2), 1 - 2(q_2^2 + q_3^2)) \end{bmatrix}$$

- Выделение признаков: разметка + литература

## АНАЛИЗ: ПОЧЕМУ НЕ ML?

- Версии:
  - **A.** Высокая сложность выделения признаков из-за малой частоты поступления данных
  - **B.** Не самая подходящая для данного метода архитектура
  - **C.** Недостаток времени
  - **D.** Всё вышеперечисленное

# ВИЗУАЛИЗАЦИЯ

 Не защищено | 18.223.136.131:5000

MATRIX: Sensor analyzer

For statistics please visit [Kibana](#)  
Wanna join The Matrix? Download an [APK](#)

---

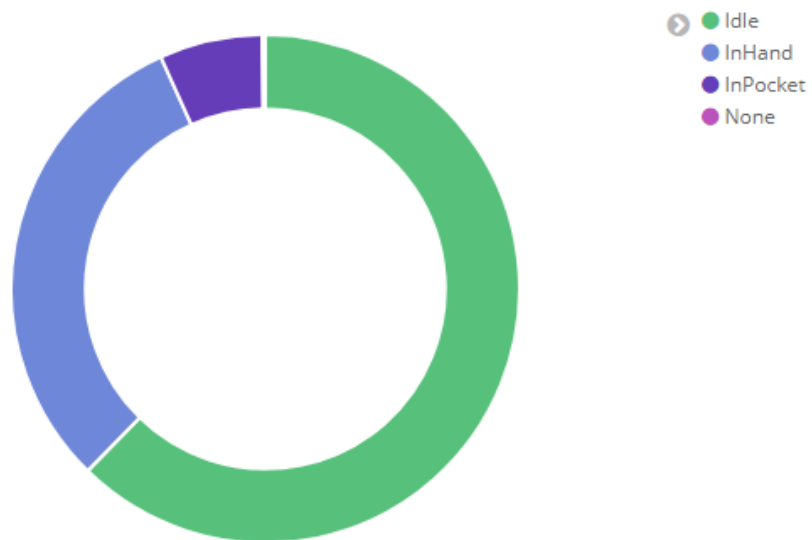
## Users online

**george** (*3fc2520be8343cbc*), user's state: **Away**, device state: **Idle**

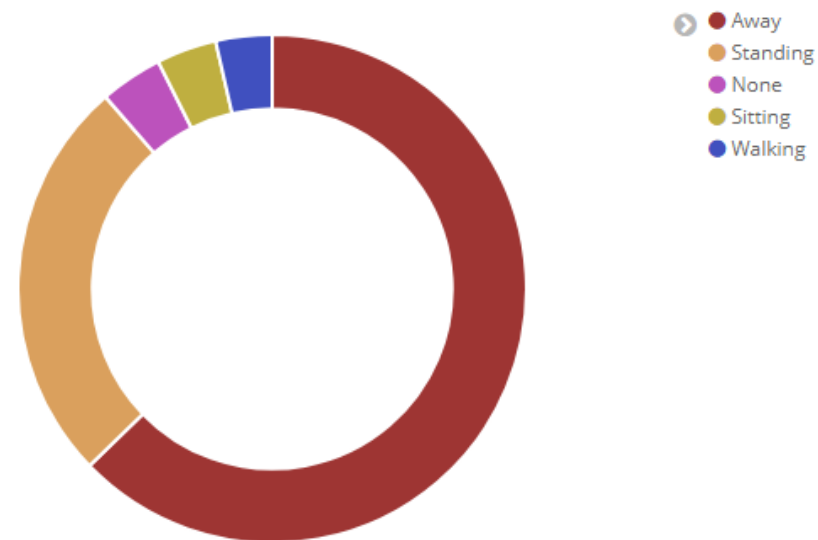
**varvara** (*3fd065bc208cdf92*), user's state: **Sitting**, device state: **InHand**

# ВИЗУАЛИЗАЦИЯ

Device states



Users states



Raw sensor data metrics

**918,520**  
Total sensor samples processed

**225,576,362**  
Total bytes processed

User states metrics

**61,995**  
Total states stored

**12**  
Total users tracked

## НЕМНОГО СТАТИСТИКИ

- По состоянию на утро **21.12.2018**
  - Обработано и сохранено **260 Мб** сырых данных сенсоров (до очистки БД 20.12.2018 – порядка **2 Гб**)
  - Сохранено **5 Мб** пользовательских состояний
  - Обработано **1,071,578** событий сенсоров
  - Сгенерировано **62,009** состояний пользователей

## ЗАКЛЮЧЕНИЕ

- Была разработана масштабируемая система для непрерывного сбора и анализа данных сенсоров Android смартфонов
- Был получен опыт работы с инфраструктурой Amazon Web Services
- Был получен опыт работы с Apache Spark

## ССЫЛКИ

- <http://18.223.136.131:5000>
- <https://github.com/dev0x13/sensor-analysis>

**СПАСИБО ЗА ВНИМАНИЕ!**