



Задача:

Разработать платформу управления данными, обеспечивающую :

- Обработку массивных потоков данных, содержащих информацию о поведении пользователей в сети;
- Оперативное подключения поставщиков данных и синхронизацию пользователей между поставщиками;
- Классификацию и структурирование потоков данных об аудитории в реальном времени;
- Получение отдельного профиля пользователя со временем отклика менее **10 мс** и пропускной способностью не менее **70 000** операций в секунду;
- Сегментирование аудитории (объединение пользователей по заданным условиям в одну группу).



Решение:

- Разработан сервер для загрузки «сырых» данных о поведении пользователей в Интернет;
- Обеспечено сохранение данных для последующей классификации и анализа (Apache Kafka, HDFS);
- Обеспечена классификация и структурирование сырых данных о поведении пользователей в сети;
- Разработаны Spark job для формирования сегментов, описанных с помощью предметно-ориентированного языка (разработана грамматика);
- Обеспечено сохранение и выдача профилей пользователей в базе данных «ключ-значение» (Aerospike).



Результаты:

- Возможность **монетизации** «сырых» данных о пользователях, предоставление сервиса **получения профиля** отдельного пользователя (для рынка RTB);
- Поставка оффлайн-**пользовательских сегментов** под конкретную рекламную кампанию клиента.

