

Разработка инструмента для анализа социальных сетей

Выполнили:
Павлюченко Денис
Эдуардов Эдуард

Руководители:
Толстикова Никита
Чумак Анастасия

Проблема

Нет единого инструмента для решения задач
JetBrains Research Team по анализу
социальных сетей



Постановка задачи

Создать инструмент для анализа Twitter



Локальные задачи



План:

Трекинг числа подписчиков любых Twitter аккаунтов

Подсчет попарных пересечений подписчиков любых аккаунтов

Подсчет числа уникальных подписчиков для любых аккаунтов

Выявление «Influencers» для любого аккаунта

Выявление твитов с наибольшим числом ретвитов

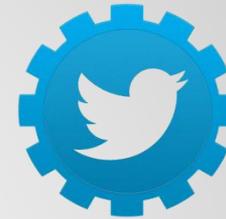
Подсчет среднего количества твитов и ретвитов за любое время

Сбор статистики по ссылкам и user mentions в твитах

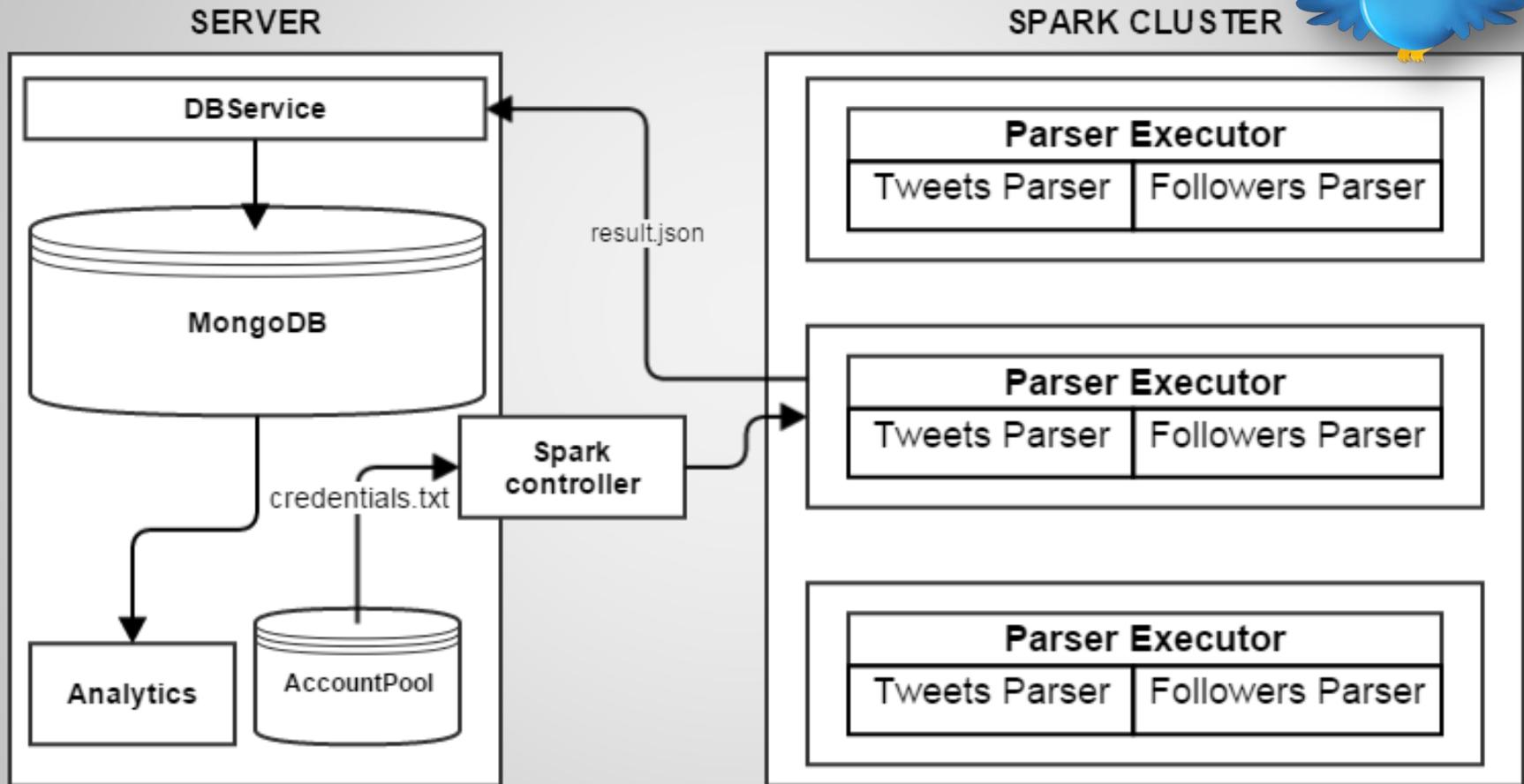
Экспорт списка подписчиков одного/нескольких Twitter-аккаунтов в формате .csv/.txt



Технологии



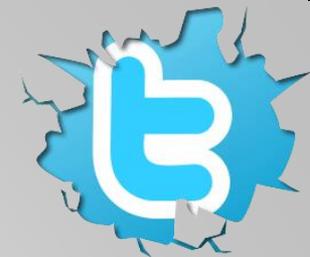
Архитектура системы



Проблемы в проекте

- ▶ Twitter REST API – жёсткие лимиты, неполнота данных
- ▶ Apache Spark – неинтуитивный сабмит приложений, проблемы с деплоем

Результаты



План:

Трекинг числа подписчиков любых Twitter аккаунтов



Подсчет попарных пересечений подписчиков любых аккаунтов



Подсчет числа уникальных подписчиков любых аккаунтов



Выявление «Influencers» для любого аккаунта



Выявление твитов с наибольшим числом ретвитов



Подсчет среднего количества твитов и ретвитов за любое время



Сбор статистики по ссылкам и user mentions в твитах



Экспорт списка подписчиков одного/нескольких Twitter-аккаунтов в формате .csv/.txt



Итоги работы



Реализовано:

- ▶ Распределенный сбор данных с Twitter
- ▶ Хранение данных о пользователях, твитах и ретвитах
- ▶ Подсчет различных агрегированных показателей

План на будущее:

- ▶ Web-интерфейс
- ▶ Визуализация данных
- ▶ Определение influencers
- ▶ Поддержка других социальных сетей



Разработка инструмента для анализа социальных сетей

Выполнили:
Павлюченко Денис
Эдуардов Эдуард

Руководители:
Толстиков Никита
Чумак Анастасия

Спасибо за внимание!