

# ДЗ 11

Кондраев Дмитрий

Задание — напишите юнит-тест WordCount-приложения на Spark Streaming (10 баллов).

Исходники см. в репозитории <https://github.com/mariohuq/2022-polis-ml>. Команды выполняются из корневой папки репозитория:

```
$ git clone https://github.com/mariohuq/2022-polis-ml.git kondraev-ml
$ cd kondraev-ml
```

## 1. Запуск WordCount, читающего из TCP-сокета

1. Старт контейнера, созданного в прошлом ДЗ:

```
$ docker start -i zeppelin
```

2. Сборка и копирование WordCount<sup>1</sup>

```
$ cd 11-spark-streaming-test/WordCountStreaming/
$ sbt package
$ docker cp ./**/wordcountstreaming_2.11-0.1.0-SNAPSHOT.jar zeppelin:/home/hduser/
```

3. Запуск источника данных (TCP-сокеты)

```
$ docker exec -it zeppelin bash # запуск второго терминала для контейнера
hduser@localhost:~$ sudo apt-get install netcat
hduser@localhost:~$ nc -lkp 9999
```

4. Запуск обработчика

```
# в первом терминале
hduser@localhost:~$ spark-submit --class ok.ml.WordCount \
    wordcountstreaming_2.11-0.1.0-SNAPSHOT.jar
```

## 2. Запуск юнит теста

```
$ cd 11-spark-streaming-test/WordCountStreaming/
$ sbt test
```

```
[info] WordCountTest:
[info] - WordCount basic
[info] Run completed in 5 seconds, 978 milliseconds.
[info] Total number of tests run: 1
[info] Suites: completed 1, aborted 0
[info] Tests: succeeded 1, failed 0, canceled 0, ignored 0, pending 0
[info] All tests passed.
[success] Total time: 8 s, completed Dec 11, 2022 7:14:14 PM
```

---

<sup>1</sup>To use \*\* in bash command enable globstar shell option: \$ shopt -s globstar