

# Инструкция к занятию "Введение в Scala"

## Цели задания

Написать распределенное приложение для чтения JSON-файлов.

## Создание проекта

1. Для начала выполнения задания необходимы следующие инструменты:
  - a. [sbt](#)
  - b. [IntelliJ IDEA](#)
2. Для создания проекта введите в консоли  
`sbt new holdenk/sparkProjectTemplate.g8`
3. В настройках можете указать
  - a. `name` - `json_reader_{ваша фамилия}`
  - b. остальные параметры оставьте по-умолчанию (просто нажимайте Enter)
4. Импортируйте проект в IntelliJ IDEA  
File > New > Project from existing sources

## Задание

Напишите приложение, которое читает json-файл с помощью Spark **RDD API** ([https://storage.googleapis.com/otus\\_sample\\_data/winemag-data.json.tgz](https://storage.googleapis.com/otus_sample_data/winemag-data.json.tgz)), конвертирует его содержимое в case class'ы и распечатывает их в stdout.  
Расположение файла передается первым и единственным аргументом.

## Сборка и запуск приложения

1. Главный класс приложения должен называться `JsonReader`
2. Собрать приложение можно с помощью команды `sbt assembly`
3. Для запуска приложения через spark нужно скачать Spark (версия 2.3.0, scala 2.11)
4. Приложение запускается командой  
`/path/to/spark/bin/spark-submit --master local[*] --class com.example.JsonReader /path/to/assembly-jar {path/to/winemag.json}`

## Подсказки

1. Для чтения файла можно использовать  
`spark.sparkContext.textFile("README.md")`
2. Для десериализации JSON можно использовать библиотеку json4s
  - a. <https://eax.me/scala-json/> - статья с примерами
  - b. <https://github.com/json4s/json4s> - сама библиотека
  - c. `val decodedUser = parse(json).extract[User]`  
где User - это ваш case class

## Критерии оценки

1. Результат необходимо прислать в виде ссылки на git-репозиторий.
2. Если приложение запускается и распечатывает кейс-классы с данными из файла - 4 балла
3. Если задание сдано до 27.09 включительно - 1 балл