

# #W5L101: 0 кypce “Spark: from zero to hero”

---

1. Kypc “Spark: from zero to hero”

---

**2**



## 1. Курс "Spark: from zero to hero"

Привет!

Если ты это читаешь, значит ты уже знаком с тем, как хранятся данные в HDFS, какие ключевые принципы функциональной парадигмы MapReduce, где у нее узкие места и как распределенный кеш, Combiner, Comparator и Partitioner позволяют оптимизировать вычисления, минимизировать объем пересылаемых данных и объем оперативной памяти, потребляемый на Worker'ами. Я также ожидаю, что ты знаком с Hive, который строит план запроса и оптимизирует цепочку вычислений MapReduce. А еще ты умеешь прокидывать порты (ssh port forwarding), поскольку приходилось работать на кластере.

В рамках следующих трех учебных модулей мы разберем современный и высокоуровневый инструмент для обработки больших массивов информации, а именно - Spark.

В рамках первого учебного модуля по Spark мы изучим:

1. Архитектуру Spark-приложений и низкоуровневый интерфейс Spark RDD;
2. Разницу между преобразованиями и действиями, а также разницу между узкими и широкими преобразованиями;
3. Как кешировать датасеты, как работать с PairRDD и делать join'ы.

В обязательном порядке тебя будут ждать:

- практические задания без оценки;
- вопросы на понимание;
- несколько задач на программирование на оценку.

Почти все как обычно, но есть нюанс. На курсе учатся люди, которые проходили наш курс по промышленной разработке на Python и знакомы с немного другим форматом обучения, а именно live demo (возможно ты - один из них). Мы проанализировали ваши отзывы и комментарии. Во-первых, большое спасибо, что не поленились их написать. Во-вторых, мы услышали явный сигнал добавить таких live demo в курс по работе с большими данными.

Live demo - это формат, где преподаватели курса показывают в интерактивном режиме как работать с различными инструментами, в частности Big Data. В ближайших модулях таких интерактивных демо будет сильно больше. Продолжайте оставлять обратную связь и комментарии, будем все анализировать и обновлять учебные материалы под ваши запросы.



# BIGDATA TEAM

С уважением,  
Алексей Драль и команда курса

P.S. ссылки на обратную связь, если они прошли мимо тебя:

- [фидбек по учебному модулю](#)
- [фидбек по домашнему заданию](#)
- **(new)** [фидбек по прошедшей части курса "HDFS, Map Reduce, Hive"](#)