## 1. Развернуть кластер

Либо развернуть образ локально, например
https://hortonworks.com/downloads/#sandbox
Либо использовать облако, например
https://www.ibm.com/cloud/

Можно использовать другие образы/облака. Требование - нужен MapReduce. В дальнейшем в курсе понадобятся Hive и Spark.

## 2. Реализовать MapReduce задачу

Посчитать word co-occurrence в предложениях.
Вывести топ самых часто встречающихся пар слов в предложениях.
Использовать любой язык программирования (не обязательно Python)

Данные: http://www.umich.edu/~umfandsf/other/ebooks/alice30.txt
Данные разбить на 3 файла.
Результат написать в 3 файла (использовать 3 редьюсера)
MapReduce задача только считает статистику совместного встречания.
Сортировку и выбор самых частых слов сделать после.
Дополнительно посчитать количество пар (тоже после выполнения задачи)

Прислать код, топ 200 пар (понятно, что будет много знаков препинаний, надо поставить такой порог, чтобы в топ вошли осмысленные пары), количество пар.