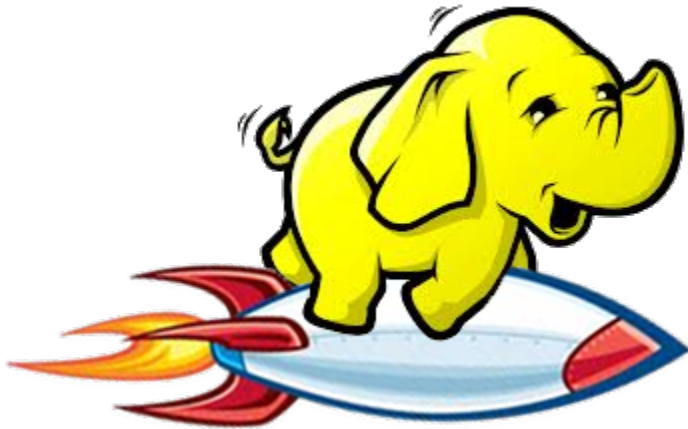


Собрать групповую статистику из лог-файла с помощью MapReduce, предварительно развернув Hadoop кластер из 5 машин с помощью Cloudera Manager



Задача:

1. (уже умеете, но придется проделать еще раз) Развернуть работающий Hadoop-кластер с помощью бесплатного продукта Cloudera Manager.
2. По файлу `advert.log` - <http://newprolab.com/bigdata/advert.log.bz2> - (ip, userid, country, bannerid, payout) размером 2Гб получить ответ (country, totalpayout) в котором для каждой страны должна стоять сумма payout по этой стране. Ответ должен располагаться в папке `/users/adverts`.
3. После проверки автоматическим скриптом, вам нужно прислать на почту `npl@digitaloctober.com` читаемый код: mapper, reducer и строчку, которой вы запустили это в shell.

Внимание! На каждую попытку дается 2 часа, этого должно хватить вам с запасом! К 3-м попыткам на выполнение основного задания лабораторной №1 мы добавляем еще одну попытку на выполнение супер-ачивки.

В результате вы сможете проверить правильность выполнения лабораторной работы с помощью автоматического скрипта автопроверки прямо в вашем личном кабинете.

Дедлайн: Вторник 17 марта 19:00

Жесткий дедлайн (минус 30%): Четверг 19 марта 19:00

Никаких подсказок - это суперачивка!

Примечание: файл `advert.log` был сгенерирован специально для вас. Пожалуйста, не воспринимайте его, как файл реальных данных.