

Из предыдущей задачи появилось предположение, что в выводе средств замешаны дропы (<http://www.rusyaz.ru/is/ns/drop.html>).

Нам удалось получить файл с логами веб-сервера (data.csv) со следующими полями: timestamp, ip-адрес, логин, user agent. Необходимо найти самого активного дроппера.

Дропперы часто ходят в банк. Как правило, они работают с многих аккаунтов, но у них не меняется ip и user-agent. Предлагается сделать map-reduce по ключу ip+user agent и взять самый большой кластер.

Ключом к следующему заданию будет md5, вычисленная от строки user agent дроппера. Получившаяся md5 должна быть представлена в виде ASCII-строки и ее шестнадцатеричные цифры приведены к нижнему регистру.

К сожалению, CSV-файл, который мы получили из банка, оказался не совсем корректным, учитывайте это при выполнении задания.