《065\_基于HyperLogLog的网站UV统计程序》

hyperloglog，数据结构+概率算法，组合而成的，去重统计，近似数

1023468个用户来访问过你

1011325，算出来这样的一个近似值，对于很多大数据统计类的程序，精准也没太大的必要，CEO层面，100万的日活

如果基于set来计数，太耗费内存了，基于HyperLogLog算法来计数，是近似计数，有0.8%的误差，但是误差不会太大，可以给出一个相对准确的近似计数，而且就占用12kb的内存，不需要存一个set来统计uv

pfadd key 一大堆item，可以对数据进行计数，如果计算过这个元素，就返回0，没计算过就返回1，他也是可以去重的

pfcount key，可以获取计数结果

这个其实是很有用的，其实一些网站数据分析，什么pv、uv之类的，没必要过于精准，直接就是用redis的hyperloglog计数就挺好的