牛客大学高薪加成系列课

HyperLogLog

牛客大学

- 专业求职辅导 -

01 / HyperLogLog简介

1. HyperLogLog并不是一种新的数据类型，实际上它是字符串类型；
2. HyperLogLog是一个专门为了计算集合的基数而创建的概率算法，其优点在于它十分的节约内存空间；
3. HyperLogLog只需12KB的内存空间，就可以对2^{64}个元素进行计数，其标准误差仅为0.81%，结果是相当可信的。

1, 2, 3, 4, 5, 1, 2, 3, 1, 2, 3, 2, 3 -> 5

统计网站的独立访客（UV）：

示例：uv:20200101 -> 1.1.1.101, 1.1.1.102, 1.1.1.103, 1.1.1.102, 1.1.1.103, ...

说明：每当用户来访时，都通过HLL记录他的IP，可以统计出每个数据集的基数，也可以对多个数据集进行合并！

假设网站每天的UV约为1000万：

记录时长	HLL	Set
记录一个月的IP所需的内存	360KB	4.5GB
记录六个月的IP所需的内存	2.16MB	27GB
记录一整年的IP所需的内存	4.32MB	54GB

牛客大学
-专业求职辅导-

THANKS

关注【牛客大学】公众号
回复“牛客大学”获取更多求职资料