

# HyperLogLog

---

用来统计一组数据中的基数的，比如说可以统计某个系统中一段时间内总共有多少个人登录不计算他登录的次数

## 基数

- 基数时数据集去重后元素个数
- HyperLogLog 是用来做基数统计的，运用了LogLog的算法

**{1, 3, 5, 7, 5, 7, 8}**

**基数集：{1, 3, 5, 7, 8}**

**基数：5**

**{1, 1, 1, 1, 1, 7, 1}**

**基数集：{1, 7}**

**基数：2**

## 基础操作

## HyperLogLog类型的基本操作

- 添加数据

```
pfadd key element [element ...]
```

- 统计数据

```
pfcount key [key ...]
```

- 合并数据

```
pfmerge destkey sourcekey [sourcekey...]
```

```
127.0.0.1:6379> flushdb
OK
127.0.0.1:6379> pfadd hll 001
(integer) 1
127.0.0.1:6379> pfadd hll 001
(integer) 0
127.0.0.1:6379> pfadd hll 001
(integer) 0
127.0.0.1:6379> pfadd hll 001
(integer) 0
127.0.0.1:6379> pfadd hll 001
(integer) 0
127.0.0.1:6379> pfadd hll 001
(integer) 0
127.0.0.1:6379> pfadd hll 002
(integer) 1
127.0.0.1:6379> pfadd hll 002
(integer) 0
127.0.0.1:6379> pfcount hll
(integer) 2
127.0.0.1:6379> █
```

## 相关说明

- 用于进行基数统计，不是集合，不保存数据，只记录数量而不是具体数据
- 核心是基数估算算法，最终数值存在一定误差
- 误差范围：基数估计的结果是一个带有0.81%标准错误的近似值
- 耗空间极小，每个hyperloglog key占用了12k的内存用于标记基数
- pfadd命令不是一次性分配12k的内存，会随着基数的增加内存逐渐增大
- pfmerge命令合并后占用的存储空间为12k，无论合并前数据量多少

