



链滴

# redis 实现搜索热词统计 “当日 top10” 和 “当周 top10” 两种方案的抉择

作者: [JellyfishMIX](#)

原文链接: <https://ld246.com/article/1591149175459>

来源网站: 链滴

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

# 核心问题

一个项目中，遇到了搜索热词统计的需求，我使用了redis的五大数据类型之一Sorted Set实现。目前两项数据需要统计：“当日搜索热词top10”和“当周搜索热词top10”。

关于这两项数据的统计方法，目前想到了两种实现方法：

1. 两个redis的Sorted Set实现，一个Sorted Set A统计当天，0点top10记录进mysql，Sorted Set 清零。一个Sorted Set B统计当周，每周日top10记录进mysql，Sorted Set B清零。
2. 只使用一个Sorted Set记录当天搜索热词，0点top10记录进mysql，Sorted Set清零。到周日时会有7 \* 10行记录。把这7 \* 10行遍历，每次便利都记录进Sorted Set，全部遍历结束后，再从Sorted Set中取出top10记录进mysql的周热词统计表中。

Sorted Set是redis的数据结构，方法1会占用两份内存，一份当天的，一份当周的。方法2会提高系的复杂度，并且在统计周表时，可能会出现短时间内大量的计算（当然可以使用定时任务放到凌晨进行）。

请问我该选择哪种解决办法比较好呢？

# 硬件与用户量

我在写的是一个学生项目，ECS是1核2GB内存的学生机（别歧视呀哈哈，穷学生☹️ weat\_smile）。如果当天的搜索热词和当周的搜索热词都使用redis内存记录，我怕内存会爆炸。

目前生产环境的ECS内存用了1GB，保守还剩900MB，我怕redis消耗两份内存会炸。目前是一个小序的后端，快要上线了，预计初期用户数量不多。关键词长度限制最大8

天马行空想一下，1MB = 1048576字节，按两个字节存一个字算，1MB能存1048576/2/8 = 65,536个不重复的搜索关键词，看起来好像可以存很多。我没熟练使用过redis，不知道这样天马行空地计算是否真的可以呢🤔thinking

# 问题中涉及的相关知识（如已经了解问题，这部分可忽略）

一个项目中，遇到了搜索热词统计的需求。我使用了redis的五大数据类型之一Sorted Set实现。

Redis 有序集合(sorted set)

Redis 有序集合和集合一样也是string类型元素的集合,且不允许重复的成员。

不同的是每个元素都会关联一个double类型的分数。redis正是通过分数来为集合中的成员进行从小到大的排序。

有序集合的成员是唯一的,但分数(score)却可以重复。

集合是通过哈希表实现的，所以添加，删除，查找的复杂度都是O(1)。集合中最大的成员数为  $2^{32} - 1$  (4294967295, 每个集合可存储40多亿个成员)。

ZSET:  Size: 8 TTL: -1

row ▼	value	score
1	后端	1
2	少林	1
3	前端	2
4	产品	3
5	足球	3
6	工具	4
7	中国	6
8	面试	11

如上图，redis的Sorted Set自带排序功能。

操作方法也比较简单，在本项目中，核心是两个方法：

zincrby命令，对于一个Sorted Set，存在的就把分数加x (x可自行设定)，不存在就创建一个分数为1成员。

zrevrange，查询集合中指定顺序的值。返回有序的集合中，score大的在前面。