链滴

Redis- 简单动态字符串

作者：SimpleBin

原文链接：https://ld246.com/article/1492744171693

来源网站：链滴

Redis没有直接使用C语言传统的字符串，而自己构建了一种简单动态字符串（Simple Dynamic Strin）的抽象类型，简称为SDS。并将SDS作为Redis的默认字符串表示。

例如：当我们在redis-client执行如下的命令时;

```
> redis 127.0.0.1:6379>set message 'Hello World'
>
> OK
```

那么Redis将会在数据库中创建一个新的键值对，其中：

- 键值对的键是一个字符串对象，对象底层的实现是一个保存着字符串“msg”的SDS。
- 键值对的值也是一个字符串对象，对象底层的实现是一个保存着字符串“Hello World”的SDS。

SDS的定义

每个sds.h/sdshdr结构表示一个SDS值：

```
struct sdshdr{

 int len; //记录buf中已使用字节的数量等于SDS所保存的字符串的长度

 int free; //记录buf中未使用字节的长度

 char []buf; //字节数据，保存字符串

}
```

获取字符串的长度

Redis获取字符串长度可以通过SDS中的len属性来获得，时间复杂度为O(1)。而C语言获取字符串需去遍历char数据来叠加获得，时间复杂度为O(n)。这就保证了在获取字符串长度时不会成为Redis的能瓶颈。

杜绝缓存区溢出

除了获取字符串长度复杂度高之外，C语言不记录自身字符串长度带来的另一个问题是容易造成缓存溢出（Buffer Overflow）。SDS属性len可以记录自身字符串的长度，因此当对SDS字符串进行拼接，SDS API会先去检测buf的空间，如果空间不足，SDS先会扩容buf的空间，之后进行拼接。反之，接去拼接。

减少修改字符串带来的内存重分配次数

- 空间预分配

空间预分配用于优化SDS字符串增长操作，当SDS的API对SDS的字符串进行修改，并且要对SDS进行间扩展的时候，程序不仅会为SDS分配修改所必须要的空间，还会为SDS分配额外未使用的空间。分公式主要有SDS中的len属性的大小决定。当len属性值小于1MB时，buf数组的实际长度为 len(修改的len值)*2+1byte，这时len和free相等，一个字节用于保存空字符。当len属性值大于1MB时，buf组的实际长度为 len(修改后的len值)+1MB+1byte，这时free的值为1MB。通过内存预分配策略，Reis可以减少连续执行字符串增长操作所需要的内存重分配次数。

- 惰性空间释放

惰性空间释放用于优化SDS字符串缩短操作：当SDS的API需要缩短SDS所保存的字符串时，程序并不即去使用内存重分配来回收字符串缩短后多出来的字节，而是使用free属性将这些字节的数量记录起并等待将来使用。

二进制安全

由于SDS是通过len属性来判断字符串是否结束的，而不是使用空字符串来判断字符串是否结束的，且SDS的API都是二进制安全的，所有的SDS API都会以处理二进制的方式来处理SDS存放在buf数组的数据，程序不会对数组中的数据进行任何的过滤，限制，假设，数据在写入和读出的时候是一致的。

兼容部分C字符串函数

虽然SDS API是二进制安全的，但它依然遵循C字符串以空字符结尾的惯例。

总结

C语言字符串	SDS
获取字符串长度的时间复杂度为O(N)	获取字符串
API是不安全的，可能会造成缓存区溢出	API是安
修改字符串N次必须需要执行N次内存重分配	修
只能保存文本数据	可以保存文本和二进制数据
可以使用所有的函数	可以使用部分函数

SDS API

函数	作用	时间复杂度
sdsnew	创建一个给定包含C字符串的SDS	
sdsempty	创建一个不包含任何内容的空 SDS	
sdsfree	释放给定的SDS	O(N),N
sdslen	返回SDS已使用的字节数	
sdsavail	返回SDS未使用的字节数	
sdsdup	创建一个给定SDS的副本（copy）	
sdsclear	清空给定的SDS字符串内容	
sdscat	将给定的C字符串拼接到SDS字符串的末尾	

sdscatsds	将给定的SDS字符串拼接到另一个SDS字符串的末尾
(N),N为拼接SDS字符串的长度	
sdscpy	将给定的C字符串复制到SDS中，覆盖原来SDS中的字符串
(N),N为被复制C字符串的长度	
sdsgrowzero	用空字符串将SDS扩展到指定的长度
(N),N被扩展的新增字节数	
sdsrange	保留SDS给定区域的数据，不在区域的数据覆盖或者清除
(N),N被保留的字节数	
sdstrim	接受一个SDS和C字符串的参数，从SDS中移除所有在C字
串中出现过的字符	O(N*N),N为给定C字符串的长度
sdscmp	对比两个SDS是否相等 O
N),N两个SDS中len较小的值	