**背景需求：**

有两张表，暂且定义为t1和t2

t1中有一个字段 client\_ip（我们暂且只关注这一个），string类型， 形式为192.168.1.100

t2中有三个字段  start\_ip,end\_ip,city\_id，其中start\_ip和end\_ip都是long类型，形式为16777216,16777471

注：start\_ip和end\_ip是client\_ip通过一个算法得到的。

现在要用MR写一段程序，把t1表中的client\_ip截取出来，然后与t2表中的start\_ip,end\_ip作比较，看其落在哪个区间内，然后返回对应区间的city\_id,并且增加到t1表的末尾

例如：

t2

|  |  |  |
| --- | --- | --- |
| start\_ip | end\_ip | city\_id |
| 16777216 | 16777471 | 010 |
| 16777472 | 16778239 | 020 |

t1表中有个client\_ip=192.168.1.100，经过算法计算，等于16777480，那么其落在16777472-16778239这个区间，则返回020，并增加到t1表末尾。

**本人的思路：**

第一个map程序，截取t1表中的client\_ip，并进行算法转换，得到和start\_ip,end\_ip相同类型的数据

第二个map程序，截取t2表中的start\_ip,end\_ip,city\_id三个字段

**然后怎么进行比较，怎么返回city\_id并添加到t1表的末尾，没有什么经验，求指点！！！！！！！！！！**

经过多方打听与查询：

方法一：

T1表 尾部加a是为了排序时在start\_ip的后面

client\_ip算法值 client\_ip

16777216a 192.168.1.100

16777217a 192.168.1.110

16777480a 192.168.1.102

T2表

|  |  |  |
| --- | --- | --- |
| start\_ip | end\_ip | city\_id |
| 16777216 | 16777471 | 010 |
| 16777472 | 16778239 | 020 |

T1 t2 按算法值和start\_ip作为key排序

Key value

16777216 010

16777216a 192.168.1.100

16777217a 192.168.1.110

16777472 020

16777480a 192.168.1.102

跨行输入map， 需要自己实现recordreader

一个record

16777216 010

16777216a 192.168.1.100

16777217a 192.168.1.110

另一个record

16777472 020

16777480a 192.168.1.102

…

把结果写到文件里，再写个mr把两个数据关联起来写回去：

record里面的数据取到16777216后面的区间

然后和record里所有的ip拼一遍

**但是这样做，数据量大的话，一定会爆内存的！！！！不可取！**