**题目：列出访问次数最多的10个IP，以及对应的次数**

分析：根据要处理的是文件还是文件夹，设计了两种命令：

命令一：处理单个文件

命令二：处理文件夹中文件名中含有“accesslog”的所有文件

**命令一：**

要查询的文件是/dataset/access\_log/accesslog.txt

文件片段如下：

tcp 0 0 127.0.0.1:8652 127.0.0.1:40193 TIME\_WAIT

tcp 0 0 127.0.0.1:8652 127.0.0.1:40192 TIME\_WAIT

tcp 0 0 127.0.0.1:8652 127.0.0.1:40196 TIME\_WAIT

tcp 0 0 127.0.0.1:8652 127.0.0.1:40199 TIME\_WAIT

tcp 0 0 127.0.0.1:8652 127.0.0.1:40201 TIME\_WAIT

tcp 0 0 127.0.0.1:8652 127.0.0.1:40204 TIME\_WAIT

tcp 0 0 127.0.0.1:8652 127.0.0.1:40207 TIME\_WAIT

tcp 0 0 127.0.0.1:8652 127.0.0.1:40210 TIME\_WAIT

**命令：**

cat /home/sean/dataset/access\_log/accesslog.txt | awk '{printf("%s\n",$5)}'| cut -d : -f 1 | sort | uniq -c | sort -n -r | head -n +10 | sed 's/^[\t]\*//g' | awk '{printf("Ip地址:%s 访问次数:%s\n",$2,$1)}' | tee /home/sean/dataset/result

**命令注释：**

cat /home/sean/dataset/accesslog.txt #打开文件

awk '{printf("%s\n",$5)}' #匹配文件的第五域，例如文件中的127.0.0.1:40207

cut -d : -f 1 #根据“：”分隔符，截取第一域，也就是127.0.0.1

sort #将所有ip地址字典序排序

uniq –c # 统计每个ip地址出现的次数

sort -n –r # 将IP地址按出现次数从大到小进行排序

head -n +10 # 取出前10位，也就是top10

sed 's/^[\t]\* #去掉每一行的行首的空格

awk '{printf("Ip地址:%s 访问次数:%s\n",$2,$1)}' #按“Ip地址:127.0.0.1 访问次数:24”将结果输出

tee /home/sean/dataset/result #将统计的结果写入文件/home/sean/dateset/result中

**命令二：**

要查询的文件是/dataset/access\_log文件夹下的文件名中含有“accesslog”的所有文件

文件夹如下：







**命令：**

find /home/sean/dataset/access\_log -name "\*accesslog\*" -exec cat {} \; | awk '{printf("%s\n",$5)}'| cut -d : -f 1 | sort | uniq -c | sort -n -r | head -n +10 | sed 's/^[\t]\*//g' | awk '{printf("Ip地址:%s 访 问次数:%s\n",$2,$1)}' | tee /home/sean/dataset/result

**命令注释：**

find /home/sean/dataset/access\_log -name "\*accesslog\*" -exec cat {} \; #找到文件夹中所有文件名中含有accesslog的文件，并查看内容

awk '{printf("%s\n",$5)}' #匹配文件的第五域，例如文件中的127.0.0.1:40207

cut -d : -f 1 #根据“：”分隔符，截取第一域，也就是127.0.0.1

sort #将所有ip地址字典序排序

uniq –c # 统计每个ip地址出现的次数

sort -n –r # 将IP地址按出现次数从大到小进行排序

head -n +10 # 取出前10位，也就是top10

sed 's/^[\t]\* #去掉每一行的行首的空格

awk '{printf("Ip地址:%s 访问次数:%s\n",$2,$1)}' #按“Ip地址:127.0.0.1 访问次数:24”将结果输出

tee /home/sean/dataset/result #将统计的结果写入文件/home/sean/dateset/result中