数据提取练习

1_1. tr命令对文件重命名，内容的替换操作

- 解答步骤

> 先把字符串保存到一个变量num中： num = '1 2 3 4 5 6 7 9 a v 你好 . /8 ' ；
>
> 用tr命令替换， tr-c替换所有不属于第一个字符集的内容， 只留下数字 “[:digit:]”, 除了数字以外的内容都先替换成‘+’

`echo ${num} | tr -c "[:digit:]" +`

> tr -s + 压缩+

`echo ${num} | tr -c "[:digit:]" + | tr -s "+"`

> 用字符串拼接解决最后一个+问题

```
1  echo `echo ${num} | tr -c "[:digit:]" + | tr -s "+"`0 | ab
```

使用管道符号 得出结果

```
yali@ip-172-31-2-107 ~ % num='1 2 3 4 5 6 7 9 a v 你好 . /8 ~'                [1]
yali@ip-172-31-2-107 ~ % echo ${num}                                          [0]
1 2 3 4 5 6 7 9 a v 你好 . /8 ~
yali@ip-172-31-2-107 ~ % num='1 2 3 4 5 6 7 9 a v 你好 . /8'                  [0]
yali@ip-172-31-2-107 ~ % echo ${num}                                          [0]
1 2 3 4 5 6 7 9 a v 你好 . /8
yali@ip-172-31-2-107 ~ % echo ${num} | tr -c "[:digit:]" +                    [0]
1+2+3+4+5+6+7+9+++++++++++++++8+%
yali@ip-172-31-2-107 ~ % echo ${num} | tr -c "[:digit:]" + | tr -s "+"        [0]
1+2+3+4+5+6+7+9+8+%
yali@ip-172-31-2-107 ~ % echo 'echo ${num} | tr -c "[:digit:]" + | tr -s "+"'0
echo ${num} | tr -c "[:digit:]" + | tr -s "+"0
yali@ip-172-31-2-107 ~ % echo ${num} | tr -c "[:digit:]" + | tr -s "+"        [0]
1+2+3+4+5+6+7+9+8+%
yali@ip-172-31-2-107 ~ % echo `echo ${num} | tr -c "[:digit:]" + | tr -s "+"`0
1+2+3+4+5+6+7+9+8+0
yali@ip-172-31-2-107 ~ % echo `echo ${num} | tr -c "[:digit:]" + | tr -s "+"`0 |
 bc
45
```

！！！中间echo转化时，注意一下是反单引号

另解：

> echo -n 表示没有换行符号（echo打印的时候自动加了换行符号，printf没有， 二则的区别）

`echo -n ${num} | tr -c "[:digit:]" +`

bc 处理计算时要以换行符号结束，所以再用一次echo

```
1  echo `echo -n ${num} | tr -c "[:digit:]" + | tr -s "+"` | bc
```

```
yali@ip-172-31-2-107 ~ % echo -n ${num} | tr -c "[:digit:]" +                [0]
1+2+3+4+5+6+7+9+++++++++++++++8%
yali@ip-172-31-2-107 ~ % echo -n ${num} | tr -c "[:digit:]" + | tr -s "+"    [0]
1+2+3+4+5+6+7+9+8%
yali@ip-172-31-2-107 ~ % echo -n ${num} | tr -c "[:digit:]" + | tr -s "+" | bc
(standard_in) 1: syntax error
yali@ip-172-31-2-107 ~ % echo `echo -n ${num} | tr -c "[:digit:]" + | tr -s "+"`
 | bc
45
yali@ip-172-31-2-107 ~ %                                                     [0]
```

1_2. 请将该**文**件中所有**大**写字母转换为**小**写

```
yali@ip-172-31-2-107 ~ % echo "ABCefg" | tr "[:upper:]" "[:lower:]"          [0]
abcefg
yali@ip-172-31-2-107 ~ % echo "ABCefg" | tr A-Z a-z                         [0]
abcefg
```

用tr替换， 2种方法

2 找到**PATH** 变量中的最后一个路径

a: 一般方法； b: 利用转义符\n； c：应用反转

```
yali@ip-172-31-2-107 ~ % echo ${PATH}                                        [0]
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/loc
al/games:/snap/bin
yali@ip-172-31-2-107 ~ % echo ${PATH} | cut -d: -f 9                        [0]
/snap/bin
yali@ip-172-31-2-107 ~ % echo ${PATH} | tr : "\n"                           [0]
/usr/local/sbin
/usr/local/bin
/usr/sbin
/usr/bin
/sbin
/bin
/usr/games
/usr/local/games
/snap/bin
yali@ip-172-31-2-107 ~ % echo ${PATH} | tr : "\n" | tail -1                 [0]
/snap/bin
yali@ip-172-31-2-107 ~ % echo ${PATH} | rev | cut -d: -f1                   [0]
nib/pans/
yali@ip-172-31-2-107 ~ % echo ${PATH} | rev | cut -d: -f1 | rev              [0]
/snap/bin
```

3. 使**用last** 命令,输出所有的登录**用**户名及登录次数，按登录次数由多及少排序

> last | cut -d " " -f1 (所有的登录用户)

```
yali@ip-172-31-2-107 ~ % last | cut -d " " -f1
yali
yali
yali
yali
Admin
Admin
Admin
yali
Admin
Admin
ubuntu
yali
Admin
Admin
yali
yali
yali
yali
yali
ubuntu
ubuntu
reboot

wtmp
```

> last | cut -d " " -f1 | grep -v "^$" | grep -v wtmp 去除无效的列

```
yali@ip-172-31-2-107 ~ % last | cut -d " " -f1 | grep -v "^$" | grep -v wtmp
yali
yali
yali
yali
Admin
Admin
Admin
yali
Admin
Admin
ubuntu
yali
Admin
Admin
yali
yali
yali
yali
yali
ubuntu
ubuntu
reboot
```

> 统计登录次数先排序last | cut -d " " -f1 | grep -v "^$" | grep -v wtmp | sort

> （先不用uniq去重的机制：找上下临近的值，临近值重复才可以实现去重）

```
yali@ip-172-31-2-107 ~ % last | cut -d " " -f1 | grep -v "^$" | grep -v wtmp | sort
Admin
Admin
Admin
Admin
Admin
Admin
Admin
reboot
ubuntu
ubuntu
ubuntu
yali
yali
yali
yali
yali
yali
yali
yali
yali
yali
yali
yali
```

> last | cut -d " " -f1 | grep -v "^$" | grep -v wtmp | sort | uniq -c 去重并统计次数

```
yali@ip-172-31-2-107 ~ % last | cut -d " " -f1 | grep -v "^$" | grep -v wtmp | sort | uniq -c
      7 Admin
      1 reboot
      3 ubuntu
     11 yali
```

> sort 按字母排序， sort -n 按数字从小到大排序，sort -nr 反向排序

```
yali@ip-172-31-2-107 ~ % last | cut -d " " -f1 | grep -v "^$" | grep -v wtmp | sort | uniq -c | s
ort
      1 reboot
      3 ubuntu
      7 Admin
     11 yali
yali@ip-172-31-2-107 ~ % last | cut -d " " -f1 | grep -v "^$" | grep -v wtmp | sort | uniq -c | s
ort -n
      1 reboot
      3 ubuntu
      7 Admin
     11 yali
yali@ip-172-31-2-107 ~ % last | cut -d " " -f1 | grep -v "^$" | grep -v wtmp | sort | uniq -c | s
ort -nr
     11 yali
      7 Admin
      3 ubuntu
      1 reboot
```

4. 在云主机上查找系统登录用户的总人次

> 先剔除无效的行，再统计 last | grep -v "^$" | grep -v wtmp | wc -l

```
yali@ip-172-31-2-107 ~ % last | grep -v "^$" | grep -v wtmp | wc -l
22
```

5. 将/etc/passwd 中的内容按照用户名排

cat /etc/passwd | sort -t : -k 1 (拿出用户名， 以：作为分隔符， 按第一列排序)

```
yali@ip-172-31-2-107 ~ % cat /etc/passwd | sort -t : -k 1                    [0]
Admin:x:1002:1002::/home/Admin:/usr/bin/zsh
_apt:x:104:65534::/nonexistent:/usr/sbin/nologin
backup:x:34:34:backup:/var/backups:/usr/sbin/nologin
bin:x:2:2:bin:/bin:/usr/sbin/nologin
daemon:x:1:1:daemon:/usr/sbin:/usr/sbin/nologin
dnsmasq:x:107:65534:dnsmasq,,,:/var/lib/misc:/usr/sbin/nologin
games:x:5:60:games:/usr/games:/usr/sbin/nologin
gnats:x:41:41:Gnats Bug-Reporting System (admin):/var/lib/gnats:/usr/sbin/nologin
irc:x:39:39:ircd:/var/run/ircd:/usr/sbin/nologin
landscape:x:108:112::/var/lib/landscape:/usr/sbin/nologin
list:x:38:38:Mailing List Manager:/var/list:/usr/sbin/nologin
lp:x:7:7:lp:/var/spool/lpd:/usr/sbin/nologin
lxd:x:105:65534::/var/lib/lxd/:/bin/false
mail:x:8:8:mail:/var/mail:/usr/sbin/nologin
man:x:6:12:man:/var/cache/man:/usr/sbin/nologin
messagebus:x:103:107::/nonexistent:/usr/sbin/nologin
news:x:9:9:news:/var/spool/news:/usr/sbin/nologin
nobody:x:65534:65534:nobody:/nonexistent:/usr/sbin/nologin
pollinate:x:110:1::/var/cache/pollinate:/bin/false
proxy:x:13:13:proxy:/bin:/usr/sbin/nologin
root:x:0:0:root:/root:/bin/bash
sshd:x:109:65534::/run/sshd:/usr/sbin/nologin
sync:x:4:65534:sync:/bin:/bin/sync
sys:x:3:3:sys:/dev:/usr/sbin/nologin
syslog:x:102:106::/home/syslog:/usr/sbin/nologin
systemd-network:x:100:102:systemd Network Management,,,:/run/systemd/netif:/usr/sbin/nologin
systemd-resolve:x:101:103:systemd Resolver,,,:/run/systemd/resolve:/usr/sbin/nologin
ubuntu:x:1000:1000:Ubuntu:/home/ubuntu:/bin/bash
uucp:x:10:10:uucp:/var/spool/uucp:/usr/sbin/nologin
uuidd:x:106:110::/run/uuidd:/usr/sbin/nologin
www-data:x:33:33:www-data:/var/www:/usr/sbin/nologin
yali:x:1001:1001:,,,:/home/yali:/usr/bin/zsh
```

6. 将/etc/passwd 中的内容按uid 排序

> cat /etc/passwd | sort -t : -k 3 -n(uid在第三列， -n按数字排列)

```
yali@ip-172-31-2-107 ~ % cat /etc/passwd | sort -t : -k 3 -n                    [0]
root:x:0:0:root:/root:/bin/bash
daemon:x:1:1:daemon:/usr/sbin:/usr/sbin/nologin
bin:x:2:2:bin:/bin:/usr/sbin/nologin
sys:x:3:3:sys:/dev:/usr/sbin/nologin
sync:x:4:65534:sync:/bin:/bin/sync
games:x:5:60:games:/usr/games:/usr/sbin/nologin
man:x:6:12:man:/var/cache/man:/usr/sbin/nologin
lp:x:7:7:lp:/var/spool/lpd:/usr/sbin/nologin
mail:x:8:8:mail:/var/mail:/usr/sbin/nologin
news:x:9:9:news:/var/spool/news:/usr/sbin/nologin
uucp:x:10:10:uucp:/var/spool/uucp:/usr/sbin/nologin
proxy:x:13:13:proxy:/bin:/usr/sbin/nologin
www-data:x:33:33:www-data:/var/www:/usr/sbin/nologin
backup:x:34:34:backup:/var/backups:/usr/sbin/nologin
list:x:38:38:Mailing List Manager:/var/list:/usr/sbin/nologin
irc:x:39:39:ircd:/var/run/ircd:/usr/sbin/nologin
gnats:x:41:41:Gnats Bug-Reporting System (admin):/var/lib/gnats:/usr/sbin/nologin
systemd-network:x:100:102:systemd Network Management,,,:/run/systemd/netif:/usr/sbin/nologin
systemd-resolve:x:101:103:systemd Resolver,,,:/run/systemd/resolve:/usr/sbin/nologin
syslog:x:102:106::/home/syslog:/usr/sbin/nologin
messagebus:x:103:107::/nonexistent:/usr/sbin/nologin
_apt:x:104:65534::/nonexistent:/usr/sbin/nologin
lxd:x:105:65534::/var/lib/lxd/:/bin/false
uuidd:x:106:110::/run/uuidd:/usr/sbin/nologin
dnsmasq:x:107:65534:dnsmasq,,,:/var/lib/misc:/usr/sbin/nologin
landscape:x:108:112::/var/lib/landscape:/usr/sbin/nologin
sshd:x:109:65534::/run/sshd:/usr/sbin/nologin
pollinate:x:110:1::/var/cache/pollinate:/bin/false
ubuntu:x:1000:1000:Ubuntu:/home/ubuntu:/bin/bash
yali:x:1001:1001:,,,:/home/yali:/usr/bin/zsh
Admin:x:1002:1002::/home/Admin:/usr/bin/zsh
nobody:x:65534:65534:nobody:/nonexistent:/usr/sbin/nologin
```

7. 将本地的/etc 目录下的文件及目录，每十条保存到一个文件中

> 把内容先存放在 ls_etc.txt 文件中放在home路径下ls > ~/ls_etc.txt
>
> 每10条切割一下，放在practice文件夹下split -l 10 ls_etc.txt practice/ls_etc_

```
yali@ip-172-31-2-107 ~ % split -l 10 ls_etc.txt practice/ls_etc_
yali@ip-172-31-2-107 ~ % cd practice/
yali@ip-172-31-2-107 practice % ls
aa  ac  ae  ag  ai  ak  am  ao  aq  as        ls_etc_ab  ls_etc_ad  ls_etc_af  ls_etc_ah  ls_etc_aj  ls_etc_al  ls_etc_an  ls_etc_ap  ls_etc_ar
ab  ad  af  ah  aj  al  an  ap  ar  ls_etc_aa  ls_etc_ac  ls_etc_ae  ls_etc_ag  ls_etc_ai  ls_etc_ak  ls_etc_am  ls_etc_ao  ls_etc_aq  ls_etc_as
yali@ip-172-31-2-107 practice % ls -la ls*
-rw-rw-r-- 1 yali yali  85 Jun 27 20:47 ls_etc_aa
-rw-rw-r-- 1 yali yali 159 Jun 27 20:47 ls_etc_ab
-rw-rw-r-- 1 yali yali 113 Jun 27 20:47 ls_etc_ac
-rw-rw-r-- 1 yali yali 109 Jun 27 20:47 ls_etc_ad
-rw-rw-r-- 1 yali yali  83 Jun 27 20:47 ls_etc_ae
-rw-rw-r-- 1 yali yali 104 Jun 27 20:47 ls_etc_af
-rw-rw-r-- 1 yali yali  90 Jun 27 20:47 ls_etc_ag
-rw-rw-r-- 1 yali yali 104 Jun 27 20:47 ls_etc_ah
-rw-rw-r-- 1 yali yali 109 Jun 27 20:47 ls_etc_ai
-rw-rw-r-- 1 yali yali 105 Jun 27 20:47 ls_etc_aj
-rw-rw-r-- 1 yali yali  99 Jun 27 20:47 ls_etc_ak
-rw-rw-r-- 1 yali yali  93 Jun 27 20:47 ls_etc_al
-rw-rw-r-- 1 yali yali 106 Jun 27 20:47 ls_etc_am
-rw-rw-r-- 1 yali yali  64 Jun 27 20:47 ls_etc_an
-rw-rw-r-- 1 yali yali  87 Jun 27 20:47 ls_etc_ao
-rw-rw-r-- 1 yali yali  62 Jun 27 20:47 ls_etc_ap
-rw-rw-r-- 1 yali yali  86 Jun 27 20:47 ls_etc_aq
-rw-rw-r-- 1 yali yali  97 Jun 27 20:47 ls_etc_ar
-rw-rw-r-- 1 yali yali  22 Jun 27 20:47 ls_etc_as
yali@ip-172-31-2-107 practice % cat ls_etc_aa
NetworkManager
X11
acpi
adduser.conf
alternatives
apm
apparmor
apparmor.d
apport
apt
```

8. 将/etc/passwd 中存放的第10到20个用户，输出uid， gid 和groups

> 取出10-20个用户cat /etc/passwd | head -20 | tail -10

> 用户名分割取第一列 cut -d: -f1
>
> 参数代换xargs -n 1 id ； 每次读一个参数给id

```
yali@ip-172-31-2-107 practice % cat /etc/passwd | head -20 | tail -10
uucp:x:10:10:uucp:/var/spool/uucp:/usr/sbin/nologin
proxy:x:13:13:proxy:/bin:/usr/sbin/nologin
www-data:x:33:33:www-data:/var/www:/usr/sbin/nologin
backup:x:34:34:backup:/var/backups:/usr/sbin/nologin
list:x:38:38:Mailing List Manager:/var/list:/usr/sbin/nologin
irc:x:39:39:ircd:/var/run/ircd:/usr/sbin/nologin
gnats:x:41:41:Gnats Bug-Reporting System (admin):/var/lib/gnats:/usr/sbin/nologin
nobody:x:65534:65534:nobody:/nonexistent:/usr/sbin/nologin
systemd-network:x:100:102:systemd Network Management,,,:/run/systemd/netif:/usr/sbin/nologin
systemd-resolve:x:101:103:systemd Resolver,,,:/run/systemd/resolve:/usr/sbin/nologin
yali@ip-172-31-2-107 practice % cat /etc/passwd | head -20 | tail -10 | cut -d: -f1
uucp
proxy
www-data
backup
list
irc
gnats
nobody
systemd-network
systemd-resolve
yali@ip-172-31-2-107 practice % cat /etc/passwd | head -20 | tail -10 | cut -d: -f1 | xargs -n 1 id
uid=10(uucp) gid=10(uucp) groups=10(uucp)
uid=13(proxy) gid=13(proxy) groups=13(proxy)
uid=33(www-data) gid=33(www-data) groups=33(www-data)
uid=34(backup) gid=34(backup) groups=34(backup)
uid=38(list) gid=38(list) groups=38(list)
uid=39(irc) gid=39(irc) groups=39(irc)
uid=41(gnats) gid=41(gnats) groups=41(gnats)
uid=65534(nobody) gid=65534(nogroup) groups=65534(nogroup)
uid=100(systemd-network) gid=102(systemd-network) groups=102(systemd-network)
uid=101(systemd-resolve) gid=103(systemd-resolve) groups=103(systemd-resolve)
```

9.将按照用户名查看/etc/passwd 中的用户，读到'sync' 用户时结束

> cat /etc/passwd | sort -t : -k 1 <<sync(错误)
>
> << 是在前面命令从标准输入读入时作为结束字符的，但是前面命令只有标准输出完全没有从标准输入读入数据。要用xargs参数代换-esync

```
yali@ip-172-31-2-107 ~ % cat /etc/passwd | cut -d : -f 1 | xargs  -n 1 -egames id          [0]
uid=0(root) gid=0(root) groups=0(root)
uid=1(daemon) gid=1(daemon) groups=1(daemon)
uid=2(bin) gid=2(bin) groups=2(bin)
uid=3(sys) gid=3(sys) groups=3(sys)
uid=4(sync) gid=65534(nogroup) groups=65534(nogroup)
yali@ip-172-31-2-107 ~ % cat /etc/passwd | cut -d : -f 1 | xargs -p -n 1 -egames id         [0]
id root ?...y
uid=0(root) gid=0(root) groups=0(root)
id daemon ?...y
uid=1(daemon) gid=1(daemon) groups=1(daemon)
id bin ?...y
uid=2(bin) gid=2(bin) groups=2(bin)
id sys ?...y
uid=3(sys) gid=3(sys) groups=3(sys)
id sync ?...y
uid=4(sync) gid=65534(nogroup) groups=65534(nogroup)
```

> cat /etc/passwd | cut -d : -f 1 | xargs -n 1 -esync id
>
> cat /etc/passwd | cut -d : -f 1 | xargs -p -n 1 -esync id
>
> 图例中展示的是到games

10. 词频统计（cat 重定向到a.txt 一直到xxx结束）

cat >> a.txt << xxx nihao hello hello 你好 nihao hello ls cd world pwd xxx

> tr " " "\n" 把空格替换成换行

```
yali@ip-172-31-2-107 ~ % cat a.txt
nihao hello hello 你好
nihao
hello

ls

cd
world
pwd
yali@ip-172-31-2-107 ~ % cat a.txt | tr " " "\n"
nihao
hello
hello
你好
nihao
hello

ls

cd
world
pwd
```

> grep -v "^$" 去除空行
>
> sort | uniq -c 先排序再去重统计
>
> sort -n 从小到大排序
>
> sort -nr 逆向排序

```
yali@ip-172-31-2-107 ~ % cat a.txt | tr " " "\n" | grep -v "^$" | sort | uniq -c | sort -n
      1 cd
      1 ls
      1 pwd
      1 world
      1 你好
      2 nihao
      3 hello
yali@ip-172-31-2-107 ~ % cat a.txt | tr " " "\n" | grep -v "^$" | sort | uniq -c | sort -nr
      3 hello
      2 nihao
      1 你好
      1 world
      1 pwd
      1 ls
      1 cd
```