

```

/// 第一节
//练习1
cd Desktop
touch data.txt
cd /
cd lib
du --max-depth=1 > ../home/datascience/Desktop/data.txt
cd ~
cd Desktop
sort -n data.txt -o data.txt
touch result.txt
head -n 10 data.txt > result.txt
/*

```

这里使用到了重定向，把结果写入文件，并进行排序
最后去前10个 重定向到结果文件即可

```

*/
// 练习2
cd Desktop
touch data2.txt
cd ..
du --max-length=1 > Desktop/data2.txt
cd Desktop/
sort -n data2.txt -o data2.txt
head -n 5 data2.txt > result2.txt

```

// 第二节

```

// 练习1
curl -OL
https://raw.githubusercontent.com/biddata/datascience/master/F14/lab1/data/
wc_day6_1_log.tar.bz2

tar -xjf wc_day6_1_log.tar.gz2
cut -d ' ' -f 7 wc_day6_1.log | sort | uniq -c | sort -n | tail - 5

```

```

// 练习2
cut -d ' ' -f 7,9 wc_day6_1.log | grep "^\\/.*\s^(200)]" | sort |
uniq -c | sort -n | tail -10

```

```

// 练习3
cut -d ' ' -f 7,9 wc_day6_1.log | grep "^\\/.*\s404" | wc -l
result: 1293
// 查找某一天404的页面数
cut -d ' ' -f 4,7,9 wc_day6_1.log | grep "^30\|Apr\|
1998.*\s\|.*\s404" | wc -l

```

// 第三节

// 练习1

/*
先梳理下思路，这里主要用到的是sed命令对数据进行替换

原始数据是这样的：

```

57 - - [30/Apr/1998:22:00:48 +0000] "GET /english/images/
lateb_new.gif HTTP/1.0" 200 1431

```

先去除两个 '-' sed 's/ - -//'

然后来去除 '[' sed 's/\[///'

然后去除 ' +0000]' sed 's/ +0000]//'

继续：去除第一个 ':' sed 's/:/ /'

然后： 去除两个'"(引号)' sed 's/\\"//g'

数据清理完毕,最后在每个空格插入';'

最后重定向到一个csv文件中即可

用一个命令写的话就是：

*/

```
cat wc_day6_1.log | sed 's/ - -//; s/\[///; s +0000]//; s/:/ /; s/ \"/g;' | sed 's/ /;/g' > result.csv
```

// 练习2

```
cat wc_day6_1.log | grep "^[0-9]+;[1-30]\/[Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec]\/[0-9]{2};[0-23]:[(0[0-9])|10-59]:[(0[0-9])|10-59];[GET|POST];.*\/1\[0-9];[0-9]+;[0-9]+"
```