name,age,city

Alice,30,New York

Bob,25,Los Angeles

Charlie,35,Chicago

**cut -d ',' -f 2-3 data.txt**

age,city

30,New York

25,Los Angeles

35,Chicago

**cut -d ',' -f 1,2 data.txt**

name,age

Alice,30

Bob,25

Charlie,35

拼接两个文件

paste -d ',' file1.txt file2.txt

a,1

b,2

c,3

**tr '[A-Z]' '[a-z]' < textfile**：将 textfile 中的所有大写字母转换为小写字母。

**tr -s ' ' ' ' < textfile**：将 textfile 中连续的多个空格压缩为一个空格。

**tr '\015' '\012' < windows\_textfile**：将 windows\_textfile 中的回车符转换为换行符，将Windows格式的文本文件转换为Unix格式。

**将所有小写字母转换为大写字母**

tr 'a-z' 'A-Z' < example.txt

将所有空格替换为下划线

tr ' ' '\_' < example.txt

uniq

-d：仅显示重复的行。

-u：仅显示不重复的行。

-c：统计行出现的次数。

-i：忽略大小写。

sort

-n：按数值排序。

-k：按特定字段排序。

-r：逆序排序。

-u：删除重复行。

-f：忽略大小写排序。

压缩和解压：gzip和gunzip

| **正则表达式** | **描述** | **示例** | **匹配结果** |
| --- | --- | --- | --- |

|  |  |  |  |
| --- | --- | --- | --- |
| . | 匹配任意单个字符（除换行符） | a.c | abc, a1c |

|  |  |  |  |
| --- | --- | --- | --- |
| ^ | 匹配字符串的开始 | ^abc | abc（在行首） |

|  |  |  |  |
| --- | --- | --- | --- |
| $ | 匹配字符串的结束 | abc$ | abc（在行尾） |

|  |  |  |  |
| --- | --- | --- | --- |
| \* | 匹配前一个字符零次或多次 | a\* | ``, a, aa |

|  |  |  |  |
| --- | --- | --- | --- |
| + | 匹配前一个字符一次或多次 | a+ | a, aa |

|  |  |  |  |
| --- | --- | --- | --- |
| ? | 匹配前一个字符零次或一次 | a? | ``, a |

|  |  |  |  |
| --- | --- | --- | --- |
| ` | ` | 逻辑或 | `a |

|  |  |  |  |
| --- | --- | --- | --- |
| \ | 转义字符 | \. | . |

|  |  |  |  |
| --- | --- | --- | --- |
| [abc] | 匹配 a、b 或 c 中的任意一个字符 | [abc] | a, b, c |

|  |  |  |  |
| --- | --- | --- | --- |
| [^abc] | 匹配除 a、b、c 之外的任意一个字符 | [^abc] | d, e |

|  |  |  |  |
| --- | --- | --- | --- |
| [a-z] | 匹配从 a 到 z 的任意一个字符 | [a-z] | a, b, ..., z |

|  |  |  |  |
| --- | --- | --- | --- |
| [0-9] | 匹配从 0 到 9 的任意一个数字 | [0-9] | 0, 1, ..., 9 |

|  |  |  |  |
| --- | --- | --- | --- |
| \d | 匹配任何数字 | \d | 0, 1, ..., 9 |

|  |  |  |  |
| --- | --- | --- | --- |
| \D | 匹配任何非数字字符 | \D | a, b, c |

|  |  |  |  |
| --- | --- | --- | --- |
| \w | 匹配任何字母、数字或下划线字符 | \w | a, 1, \_ |

|  |  |  |  |
| --- | --- | --- | --- |
| \W | 匹配任何非字母、数字或下划线字符 | \W | !, @, # |

|  |  |  |  |
| --- | --- | --- | --- |
| \s | 匹配任何空白字符 | \s | , \t |

|  |  |  |  |
| --- | --- | --- | --- |
| \S | 匹配任何非空白字符 | \S | a, 1, ! |

|  |  |  |  |
| --- | --- | --- | --- |
| {n} | 匹配前一个字符恰好 n 次 | a{3} | aaa |

|  |  |  |  |
| --- | --- | --- | --- |
| {n,} | 匹配前一个字符至少 n 次 | a{3,} | aaa, aaaa |

|  |  |  |  |
| --- | --- | --- | --- |
| {n,m} | 匹配前一个字符至少 n 次，至多 m 次 | a{3,5} | aaa, aaaa, aaaaa |

Player,Team,Machine,Password,salt,Score

Leon,BLUE,b,ff2889b3,related,5

有多少个队伍？（文件名arcade.csv）

tail -n +2 arcade.csv | cut -d, -f2 | sort | uniq | wc -l

有多少唯一的得分值？

tail -n +2 arcade.csv | cut -d, -f6 | sort | uniq | wc -l

列出所有重复的密码

tail -n +2 arcade.csv | cut -d, -f4 | sort | uniq -d

哪位玩家的得分最高？

tail -n +2 arcade.csv | sort -k6,6 -t, -n | tail -n1

哪个机器被最多玩家使用

tail -n +2 arcade.csv | cut -d, -f3 | sort | uniq -c | sort -n | tail -n1

#### 编写脚本**检查文件的第一行和最后一行是否相同。**

#!/bin/bash

if [[ -z $1 ]] || [[ ! -e $1 ]]

then

echo ERROR > /dev/stderr

exit 1

fi

first=$(head -n1 $1)

last=$(tail -n1 $1)

if [[ $first == $last ]]

then

echo "First and last lines are the same"

else

echo "First and last lines differ"

fi

ls /bin/g[^e]\*t

g[^e]\*t：g 开头，[^e] 表示第二个字母不是 e，\*t 表示以 t 结尾

编写脚本 sizes.sh，打印当前目录中每个文件的统计信息：脚本应打印每个文件的文件名和文件大小。打印所有文件的数据后，打印最大文件的大小和名称，以及所有文件大小的总和

#!/bin/bash

sum=0

max\_size=0

max\_name=''

for f in \*

do

if [[ -f $f ]]

then

size=$(wc -c $f | cut -f1 -d " ")

echo "$size $f"

if [[ $size -gt $max\_size ]]

then

max\_size=$size

max\_name=$f

fi

sum=$((sum + size))

fi

done

echo "Largest is $max\_size in $max\_name"

echo "Sum is $sum"

 for f in \*：遍历当前目录中的所有文件。

 if [[ -f $f ]]：检查当前项是否为文件。

 size=$(wc -c $f | cut -f1 -d " ")：使用 wc -c 获取文件大小，并用 cut 提取文件大小的数字部分。

 sum=$((sum + size))：累计文件大小。

 第一行打印一个 \*，第二行打印两个 \*，依此类推。

 行数由命令行参数 -r 指定。如果没有提供参数，则打印4行。

#!/bin/bash

if [[ $# -gt 1 && $1 -eq '-r' ]]

then

    size=$2

else

    size=4

fi

row=1

while [[ $row -le $size ]]

do

    col=1

    while [[ $col -le $row ]]

    do

        echo -n '\*'

        col=$((col+1))

    done

    echo

    row=$((row+1))

done

### 修改你的三角形脚本，按照以下规则打印字符**需求：**

### 打印最后一列时，打印 \

打印第一列时，打印 |

如果行号为奇数，打印 L

否则打印 \*

#!/bin/bash

if [[ $# -gt 1 && $1 -eq '-r' ]]

then

    size=$2

else

    size=4

fi

row=1

while [[ $row -le $size ]]

do

    col=1

    while [[ $col -le $row ]]

    do

        if [[ $col -eq $row ]]

        then

            echo -n '\'

        elif [[ $col -eq 1 ]]

        then

            echo -n '|'

        elif [[ $((row%2)) -eq 0 ]]

        then

            echo -n 'L'

        else

            echo -n '\*'

        fi

        col=$((col+1))

    done

    echo

    row=$((row+1))

done

rabbit”在Alice\_in\_Wonderland.txt中出现次数

grep rabbit Alice\_in\_Wonderland.txt | wc -l #5

grep Rabbit Alice\_in\_Wonderland.txt | wc -l#47

同时搜索

grep -i rabbit Alice\_in\_Wonderland.txt | wc -l # note this also matches RABBIT

#or

grep [Rr]abbit Alice\_in\_Wonderland.txt | wc -l

有多少行不包含“Caterpillar”或“caterpillar”一词？

grep -vic caterpillar Alice\_in\_Wonderland.txt

Alice\_in\_Wonderland.txt 中“Alice”一词出现了多少次？请注意，它有时会在同一行上出现多次。您可能想查看 grep -o

grep -o Alice Alice\_in\_Wonderland.txt | wc -l # 401

用grep命令解决arcade.csv的问题

打印 arcade.csv 中团队为绿色的所有行

grep GREEN arcade.csv

Molly的Score是多少

grep Molly arcade.csv | cut -d, -f6

打印 arcade.csv 中机器为 b 且分数以数字 4 开头的所有行

grep ,b, arcade.csv | grep ,4[0-9]\*$

您已收到一份大学入学数据列表 (australian-universities.csv)，但数据很混乱。编写一个 sed 脚本来执行以下操作

删除所有不包含单词 University 的行（不区分大小写，因此您还应该保留 UNIversity 等）

/university/I!d

删除名称字段后包含字母的所有行

/,.\*[a-zA-Z].\*/d

将所有句号 (.) 替换为逗号 (,)

s/[.]/,/g

删除所有尾随逗号

s/,$//

有mm/dd/yyyy 格式的生日列表，现在用awk打印日期为dd/mm/yyyy 格式

awk -F / '{print $2"/"$1"/"$3}' american\_dates.txt

Monash University,42339,22140,64479

得到第三行到第六行的数据

awk 'NR==3,NR==6{print $0}' australian-universities.csv

得到第三列和第二列的数据

awk '{print "third line: "$3,"second line: "$2}' australian-universities.csv

输入分隔符，默认为空格，空白字符，英文是field separator，变量名是FS

输出分隔符，output field separator，简称OFS

Entity Code Year GDP per capita, PPP (constant 2017 international $) Population (historical estimatesHomicide rate per 100,000 population - Both sexes - All ages Life expectancy - Sex: all - Age: at birth - Variant: estimates Cantril ladder score

Afghanistan AFG 2011 1961.0963 29249156 4.208668 61.4 4.25835

第二列和第三列

awk -F "\t" -v OFS="=====" '{print $2,$3}' sample2.tsv

第四行到第八行

awk -F "\t" -v OFS="=====" 'NR==4,NR==8{print $2,$3}' sample2.tsv



输出第二列，把行号和列号也输出出来

awk -F "\t" -v OFS="====" '{print NR,$2}' sample2.tsv

输出以university开头的行

awk -F "," '/^university/{print $0}' australian-universities.csv

打印文件 data.txt 中包含单词 "apple" 的行

awk '/apple/' data.txt

打印文件 data.txt中的每一行及其行号

awk '{print NR, $0}' data.txt

统计文件words.txt中每个单词的出现次数

awk '{count[$1]++} END {for (word in count) print word, count[word]}' words.txt

打印文件data.txt中第二个字段大于100的行

awk '$2 > 100' data.txt

将所有第二个字段包含 "error" 的行的第三个字段替换为 "fixed"

awk '$2 ~ /error/ {$3 = "fixed"} {print}' data.txt

合并第二和第三个字段为一个新字段，并按这个新字段的字母顺序排序后输出

awk '{new\_field = $2 $3; print new\_field, $0}' data.txt | sort -k1,1 | awk '{print $2, $3, $4}'

计算文件中每个唯一类别（第一列）的第二个字段的平均值

awk '{sum[$1] += $2; count[$1]++} END {for (i in sum) print i, sum[i]/count[i]}' data.txt

The Awk script gday.awk contains the following [2 marks]

BEGIN{

array[0] = "Hullo"

array[1] = "Hullo"

}

{for(i=0; i<=2; i++)

print array[i] " " $1

}

When the script is invoked with the command:

echo Fred | awk -f gday.awk which of the following happens:

A You see “Hullo Fred” twice

B You see “Hullo Fred” three times

C You see “Hullo Fred” twice, followed by “Fred”

D There is a runtime error because a variable has been referenced without having previously been assigned a value.

Ans:C

Q2. You have been given an Awk script to fix. The script aims to sum all the numbers in a file, i.e. both across the lines of the file and also down the file. Each line can have multiple values and can vary in length. Unfortunately, the script is not working as intended.

{

for(i=1; i< NF; i++)

sum = $i

}

END{

printf("sum = %d\n",sum)

}

When you run the script on a file of integers,

1 2 3

4 5 6

7 8 9 10

1. The sum was 9 (rather than 55). What caused that problem? **[4 Marks]**

Ans: sum = $i should be sum += $i (or sum = sum + $i)

1. When that error has been fixed, you rerun the code and are informed that the sum is 36. Still not right. What is the source of this problem? **[4 Marks]**

Ans: The loop test i < NF should be i <= NF (It is okay if they report the bugs in the wrong order)

Q3.

1. Write a Sed command that takes a file name containing spaces, and replaces each space with an underscore “\_”. **[2 Marks]**

Ans: s/ /\_/g

1. Write a Sed command that adds the string ‘.faa’ to a file name **[2 Marks]**

Ans: s/$/.faa/

1. Write a Sed command that adds “../data/” to the start of each file name. **[3 marks]**

Ans: s/^/..\/data\//

Q4. Write a runnable Bash script, do\_mvs <directory> which, given two directories, looks for ordinary files (i.e. not directories) in the first directory (and recursively through its subdirectories) , and moves each of the files to the second directory. For example, assume directory level1 contains file a and directory level2, and directory level2 contains file b:

level1

a

level2

b

then after do\_mvs(level1, other\_directory) the files a and b have been moved, other\_directory now has:

a

b

while level1 now looks like:

level 1

level 2

The program will first need to check that the files given as the argument exist and are directories. (Hint. You may wish to consider using the find command.) **[12 marks]**

Ans:

!/usr/bin/env bash

if [[ $# -eq 0 || ! -d $1 || ! -d $2 ]]

then

echo "usage: $0 <directory> <directory>" > /dev/stderr

exit 0

fi

find $1 -type f -exec mv '{}' $2 \;

Mark allocation

#!/ etc 2 marks

If statement(s) with 3 tests 4 marks

Sensible error message and exit 2

Sensible call to find 4 (don’t sweat the punctuation)

Q5.

Does the string rotor match regular Ed-style regular expression r..r **[2 Marks**

It Matches It Does Not Match

Does the string ababababc match regular Ed-style regular expression ^[ab]\*$ **[2 Marks]**

It Matches It Does Not Match

Does the string ababababc match regular Ed-style regular expression [ab]\*c[ab]\* **[2 Marks]**

It Matches It Does Not Match

Q6. Write one or more Ed style regular expressions that completely match all of the following lines. Use as few regular expressions as possible. Please note that the regular expression must involve the letters ‘a’, ‘b’, ‘c’, ‘d’ and ‘e’, but not ‘.’

a b c d

a a c d

a c d

a c e

1. What is/are the regular expression(s)  **[4 Marks]**

Ans: aa\*b\*c[de]

1. In words, explain how do your regular expressions can be used to match all four lines.  **[4 Marks]**

Ans: The first a is found in all lines, so is in the RE. a can then appear at least once more, hence a\*. b does follow in one line, but mostly not, so b\*. (Actually b? but that was not taught.) Then compulsory c then either d or e.

Q7. Sabine is using git to track her work. She has just updated a script called configure.sh. The old file looked like this:

**configure.sh**

#!/bin/bash  
  
if [[ $1 -gt 9000 ]]  
then  
a="stardust"  
else  
a="duchess"  
fi  
  
device\_code=$(grep "$a" codes.csv | cut -d, -f2)  
./activate.sh $device\_code $1

The updated file looks like this:

**configure.sh**

#!/bin/bash  
  
if [[ "$1" =~ ^[1-9][0-9]\*$ ]]  
then  
    power=$1  
else  
    echo "Usage: ./configure.sh <power> # power must be a positive integer"  
    exit 1  
fi  
  
if [[ $power -gt 9000 ]]  
then  
    device="stardust"  
else  
    device="duchess"  
fi  
  
device\_code=$(grep "$device" codes.csv | cut -d, -f2)  
./activate.sh $device\_code $power

a) What command (or commands) should Sabine use to record her changes in git? **[2 Marks]**

Ans: git add configure.sh (1 mark)

git commit (1 mark)

OR

git commit -a (2 marks)

OR

git add -A (1 mark)

git commit (1 mark)

b). Write a suitable git commit comment to describe Sabine's changes. **[3 Marks]**

Ans: something along the line of input validation/antibugging, improved code readability,

c). Git is a useful tool for group projects and individual projects. What is one advantage of git when used for anindividual project? **[2 Marks]**

Ans: Any of

 - git helps you understand past code changes

- git helps you know which version of your code is most recent

- git allows you to back up your code remotely (no need to mention github but no penalty for doing so)

- git allows you to easily revert mistakes/restore a working version

Q8. A computer science department noticed that there an unusual number of submissions for assignments on particular days. Luckily the department’s assignment submission system kept a log of submissions per day. You have been asked to write code to report the student IDs and the number of submissions for those unusual submissions on a particular day. The data you’ve been given is in a <tab> separated file with fields resembling:

*<date>* <TAB> *<time>* <TAB> *<ID>* <TAB> *<Assessment>*

You can ignore the first column as we know the data comes from a single day. Also, to keep things simple the time is just recorded as an integer, so 7:02 is 702, while 23:00 (11pm) is recorded as 2300.

For example,

today 900 james Ass2

today 2302 james Ass2

The input data is sorted by increasing time of the day. With that in mind, your code should look for 3 or more submissions for the assessment item

**Ass2** occurring after 11pm (**23:00**), and report the corresponding IDs and number of submissions. The list of IDs should be in alphabetical order. (Don’t worry about adding antibugging code.)

Hint: You can structure this as a single Awk script, very\_suby.awk, or an Awk script with a short Bash script very\_suby.sh **[10 Marks]**

Ans

# Expected format: Date Time ID Assessment

$2 >= 2300 {if($4 == "Ass2")

subs\_id[$3]++

}

END{

for(id in subs\_id)

if(subs\_id[id] >= 4)

printf("%s\t%s\n", id, subs\_id[id]) | "sort -k 1n"

}

Another (more likely) solution is to create an Awk script, but without the internal call sort, and place the call to the awk script in a Bash script, with the call to Awk followed by a call to sort.

Marks breakdown:

* 4 marks for gathering subs which are post 23:00 and for Ass2
* 3 for looping through to get those with at least 4 hits
* 3 for the final sort