

AWK & sed homework (deadline:5/28)

請跑以下程式(print screen)並且說明此程式指令的功用於DOC中並上傳[PS. 特別注意以下跨號內黑體字]

(課程網站檔案下載:user.txt, consumer.txt, song.txt) (以下程式千萬不要複製貼上，字型不合)

awk -F": " '!a[\$2]++{print \$0}' user.txt (!a[\$0]++功能是甚麼? Google it!)

awk -F": " 'ARGIND==1{a[\$1]=\$0;next;}{if(\$1 in a){print a[\$1]"\t"\$0}}' user.txt consumer.txt (功能是甚麼?)

awk -F": " '{a[\$2]}END{asorti(a);for(i=1;i<=length(a);i++){print a[i]}}' user.txt (asorti功能是甚麼? Google it!)

sed -e '1,2d' song.txt (功能是甚麼?)

sed -e 's/6.30/7.30/g' song.txt >> song2.txt (s, g, >>功能是甚麼?)

sed 's/linux/Unix/' song.txt (沒有g有甚麼效果，請與原始檔案比較)

sed 's/linux/Unix/2' song.txt (加上2有甚麼效果，請與原始檔案比較)

sed 's/linux/Unix/g' song.txt (加上g有甚麼效果，請與前兩個比較)

sed -n 's/linux/Unix/p' song.txt (p有甚麼效果)

AWK & sed homework (deadline:5/28)

請跑以下程式(print screen)並且說明此程式指令的功用於DOC中並上傳

(課程網站檔案下載:info.txt) (以下程式千萬不要複製貼上，字型不合)

1. 首先將info.txt的檔案，將多餘的:去除，讓其只保留一個:在每個資料欄之間
2. 請利用awk將電話那一欄印出來
3. 請利用awk將姓名為Dan的人，印出其電話號碼
- 4.請利用awk將姓名開頭是J的人，印出其姓名以及電話號碼(用空白格開)
- 5.請利用awk將後面的三個欄位分別在數字前面加印\$字符號
6. 請利用sed將info.txt的檔案中John修改成 Joanthan
- 7.請利用sed將info.txt的檔案中刪除包含Lane的資料行

AWK & sed homework (deadline:5/28)

請跑以下程式(print screen)並且說明此程式指令的功用於DOC中並上傳 [PS. 特別注意以下跨號內黑體字]

(檔案下載: <https://www.kaggle.com/c/acquire-valued-shoppers-challenge/data?select=transactions.csv.gz>) , 不要解壓縮!! (以下程式千萬不要複製貼上, 字型不合)

由於檔案過大EC2硬碟不夠, Windows使用者請利用git bash或是CMder來實作

zcat 檔案路徑 | more (zcat, more功能是甚麼?)

zcat 檔案路徑 | wc -l (wc -l功能是甚麼?)

time zcat 檔案路徑 | wc -l (time功能是甚麼?)

zcat 檔案路徑 | head -n100 | tail -n50 > text.txt (head, tail, >功能是甚麼?)

zcat 檔案路徑 | awk -F", " '{if(NR==1)print\$0}' (awk -F, NR功能是甚麼?)

zcat 檔案路徑 | awk -F", " '{if(NR%1000==1)print\$0}' | more (%代表是甚麼?)

zcat 檔案路徑 | awk -F", " '{print \$3}' | more (awk \$3代表是甚麼?)

zcat 檔案路徑 | head -n10000000 | awk -F", " '\$10>100{print\$0}' | more

zcat 檔案路徑 | head -n10000000 | awk -F", " '\$7 ~/^2013-06-*/{print \$0}' | more (~代表是甚麼?)

zcat 檔案路徑 | awk -F", " '{print \$7}' | uniq -c | more (google “uniq” -c 功能是甚麼?)

zcat 檔案路徑 | awk -F", " 'NR>1 {print \$11}' | head -n10000000 | awk 'BEGIN{max=0} {if (\$1>max) max=\$1} END{print max}' (BEGIN,END, 功能是甚麼?)

Transactions.csv.gz(~3G)

id - A unique id representing a customer

chain - An integer representing a store chain

dept - An aggregate grouping of the Category (e.g. water)

category - The product category (e.g. sparkling water)

company - An id of the company that sells the item

brand - An id of the brand to which the item belongs

date - The date of purchase

productsize - The amount of the product purchase (e.g. 16 oz of water)

productmeasure - The units of the product purchase (e.g. ounces)

purchasequantity - The number of units purchased

purchaseamount - The dollar amount of the purchase

AWK & sed homework for MAC (deadline:5/28)

請利用MAC終端機，跑以下程式(**print screen**)並且說明此程式指令的功用於DOC中並上傳至Moodle[PS. 特別注意以下跨號內黑體字]

(檔案下載: <https://www.kaggle.com/c/acquire-valued-shoppers-challenge/data?select=transactions.csv.gz>)，不要解壓縮!! (以下程式千萬不要複製貼上，字型不合)

gunzip -c 檔案路徑 | more (**zcat**, **more**功能是什麼?)

gunzip -c 檔案路徑 | wc -l (**wc -l**功能是什麼?)

time gunzip -c 檔案路徑 | wc -l (**time**功能是什麼?)

gunzip -c 檔案路徑 | head -n100 | tail -n50 > text.txt (**head**, **tail**, **>**功能是什麼?)

gunzip -c 檔案路徑 | awk -F"," '{if(NR==1)print\$0}' (**awk -F**, **NR**功能是什麼?)

gunzip -c 檔案路徑 | awk -F"," '{if(NR%1000==1)print\$0}' | more (**%**代表是什麼?)

gunzip -c 檔案路徑 | awk -F"," '{print \$3}' | more (**awk \$3**代表是什麼?)

gunzip -c 檔案路徑 | head -n100000000 | awk -F"," '\$10>100{print\$0}' | more

gunzip -c 檔案路徑 | head -n100000000 | awk -F"," '\$7 ~/^2013-06-*/{print \$0}' | more (~代表是什麼?)

gunzip -c 檔案路徑 | awk -F"," '{print \$7}' | uniq -c | more (**google “uniq” -c** 功能是什麼?)

gunzip -c 檔案路徑 | awk -F"," 'NR>1 {print \$11}' | head -n10000000 | awk 'BEGIN{max=0} {if (\$1>max) max=\$1} END{print max}' (**BEGIN**,**END**,功能是什麼?)

Transactions.csv.gz(~3G)

id - A unique id representing a customer

chain - An integer representing a store chain

dept - An aggregate grouping of the Category (e.g. water)

category - The product category (e.g. sparkling water)

company - An id of the company that sells the item

brand - An id of the brand to which the item belongs

date - The date of purchase

productsize - The amount of the product purchase (e.g. 16 oz of water)

productmeasure - The units of the product purchase (e.g. ounces)

purchasequantity - The number of units purchased

purchaseamount - The dollar amount of the purchase