**awk 文字分析工具**

●常用在對文字和資料進行分析處理

●檔案逐行的讀入

**awk Script**

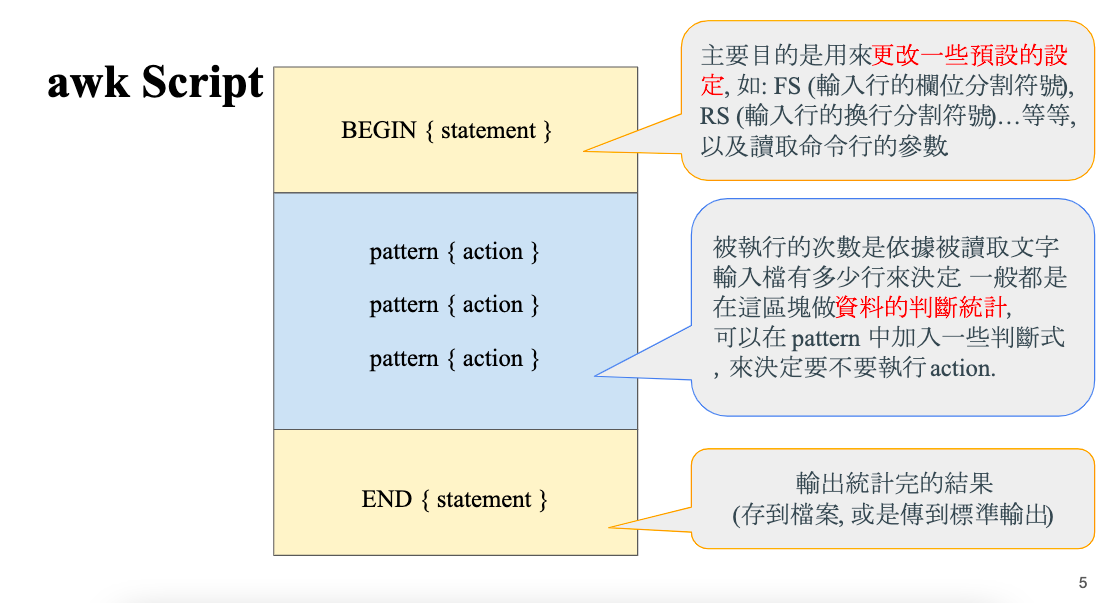
*awk ‘BEGIN{ print “start” } pattern{ commands } END{ print “end” }’ filename4*

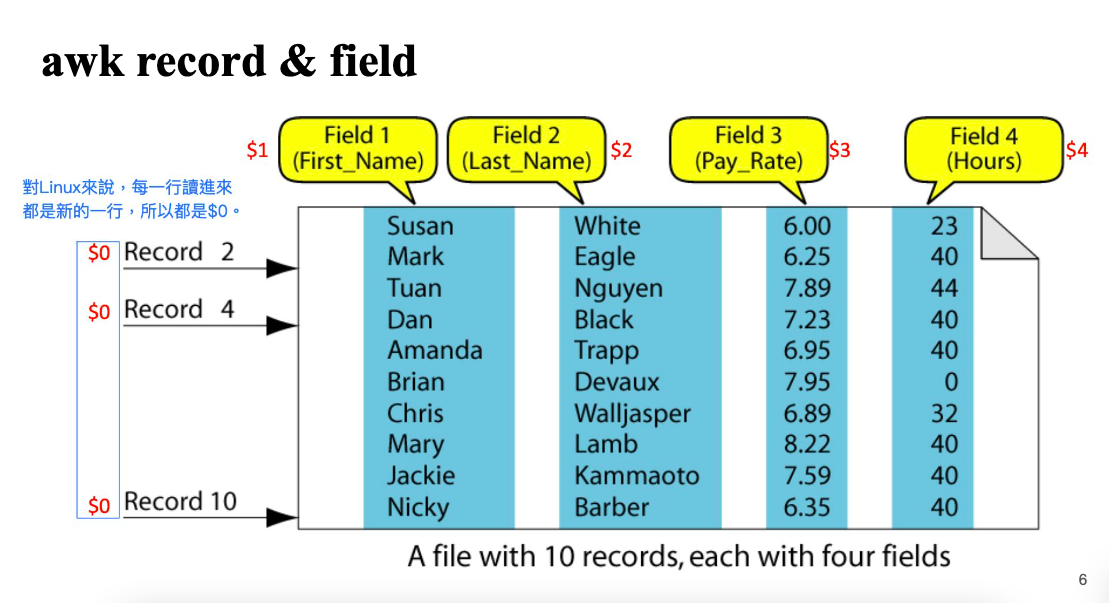
awk Script工作原理:

○第一步執行BEGIN 語句

○第二步從檔案或標準輸入讀取一行，然後再執行pattern語句，逐行掃描檔案到檔案全部被讀取

○第三步執行END語句





**awk 語法**

*awk [options] ‘scripts’ var=value filename7*

常用引數

○-F 指定分隔符(可是字串或正規表達法)，預設是空白(space)

○-f 從指令碼檔案(awk script)中讀取awk命令

○-v var=value賦值變數，將外部變數遞給awkawk [options] ‘scripts’ var=value filename

**awk print record & field**

*awk -F “\tab” ‘{print $0}’ fruit.txt*

*awk ‘{print $1, $2, $4}’ fruit.txt*

$0：entire line

$1：column 1

$2：column 2

awk 參數

➔$0 # 當前record（列、橫行）

➔$1~$n # 當前record的第 N 個欄位

➔FS # 輸入field直欄分隔符（-F相同作用）預設空格

➔RS # 輸入record（列、橫行）分割符，預設換行符

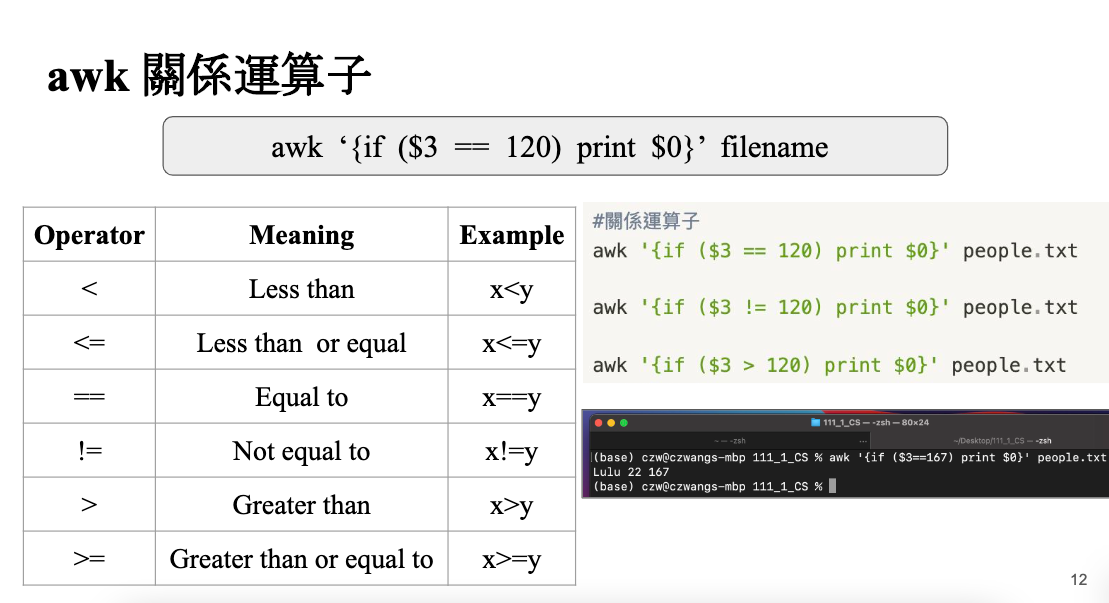
➔NF # 欄位個數

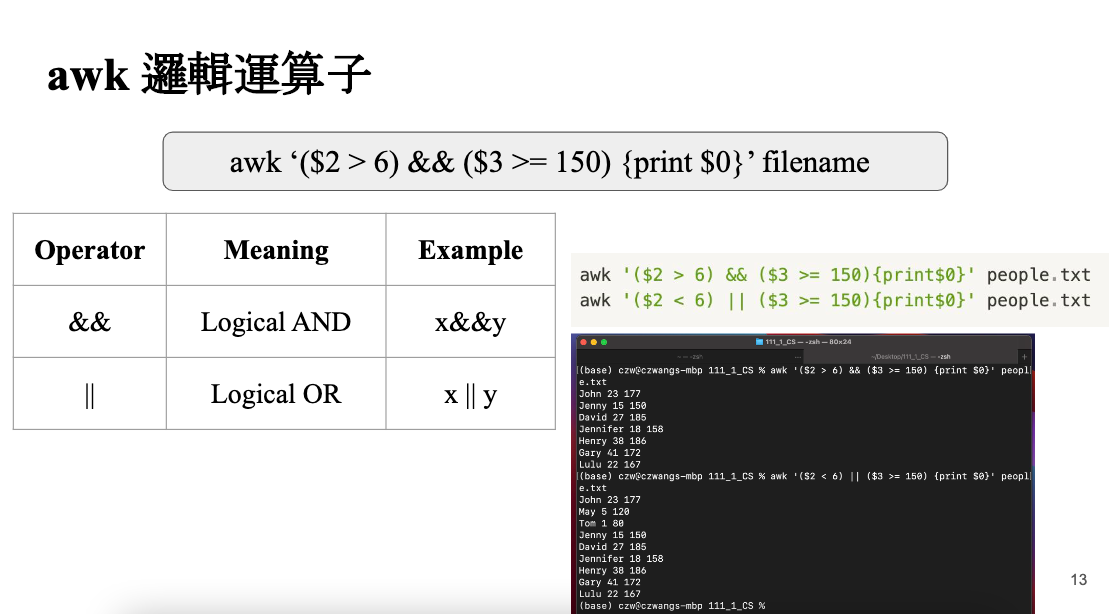
➔NR # record 數，就是行號，預設從 1 開始

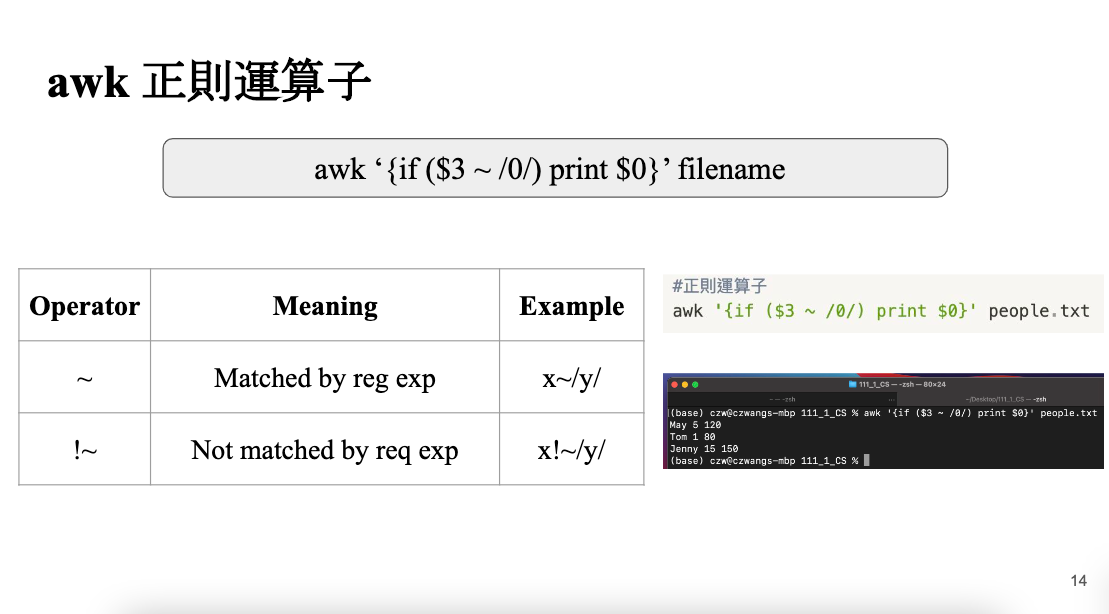
➔OFS # 輸出欄位分隔符，預設空格

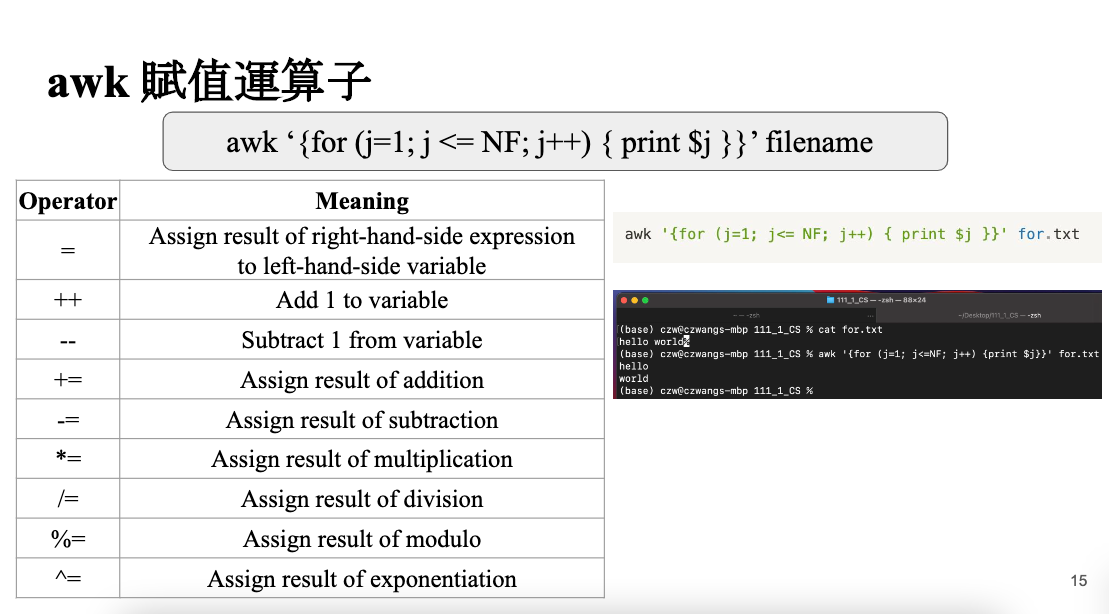
➔ORS # 輸出record分割符，預設換行符











**sed 文字分析工具**

●「stream editor 」的縮寫，顧名思義是進行串流 (stream) 的編輯

●字串取代、複製、刪除的功能

●自動化的修改文字檔

**Sed指令**

*sed [option] ‘[n1,n2] [command] / [pattern] / [replacement] / [flag]’ file.txt*

◆option：以 - 符號開頭的功能，如 -n、-r、-i，可省略

◆n1,n2：代表開始行數和結尾行數，不輸入代表每行都會執行command

◆command：進行的動作，如s, a, c, d, i

◆pattern：給command使用的參數

◆replacement：當使用s指令時會使用

◆flag：當使用s指令時會使用

**Sed語法 – 常用選項**

*sed [-nefi] ‘[n1,n2] [command] / [pattern] / [replacement] / [flag]’ file.txt*

option:

◆-n: 沉默模式，只有經過sed處理的那行才會被印出

◆-e : 直接在命令模式編輯，預設

◆-f : 直接將sed動作寫在一個檔案內，-f file會直接執行file裡面的動作

◆-i : 修改檔案

**Sed語法 – 常用指令**

*sed [-nefi] ‘[n1,n2] [command] / [pattern] / [replacement] / [flag]’ file.txt*

*command:*

◆a: 新增，在下一行插入字串，未指定行數的話，則是在每一行之後插入字串◆c: 取代，取代指定行數的內容

◆d : 刪除，後面通常不接任何東西

◆i : 插入，在指定行數的前一行插入字串

◆p : 印出，只印出受影響的行，常搭配-n使用

◆s : 搜尋、取代，最常用的指令，可搭配正規表達式使用，將搜尋的內容進行取代

**Sed語法 – 常用旗幟**

*sed [-nefi] ‘[n1,n2] [command] / [pattern] / [replacement] / [flag]’ file.txt*

flag:

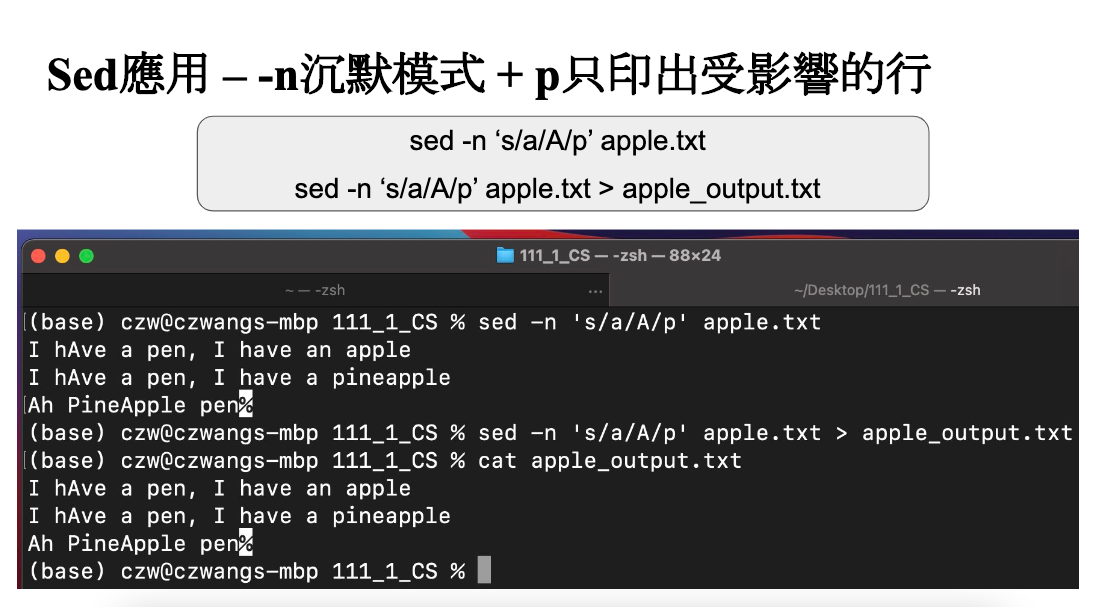
◆g: 全部取代

◆[0-9]: 每行第幾次出現

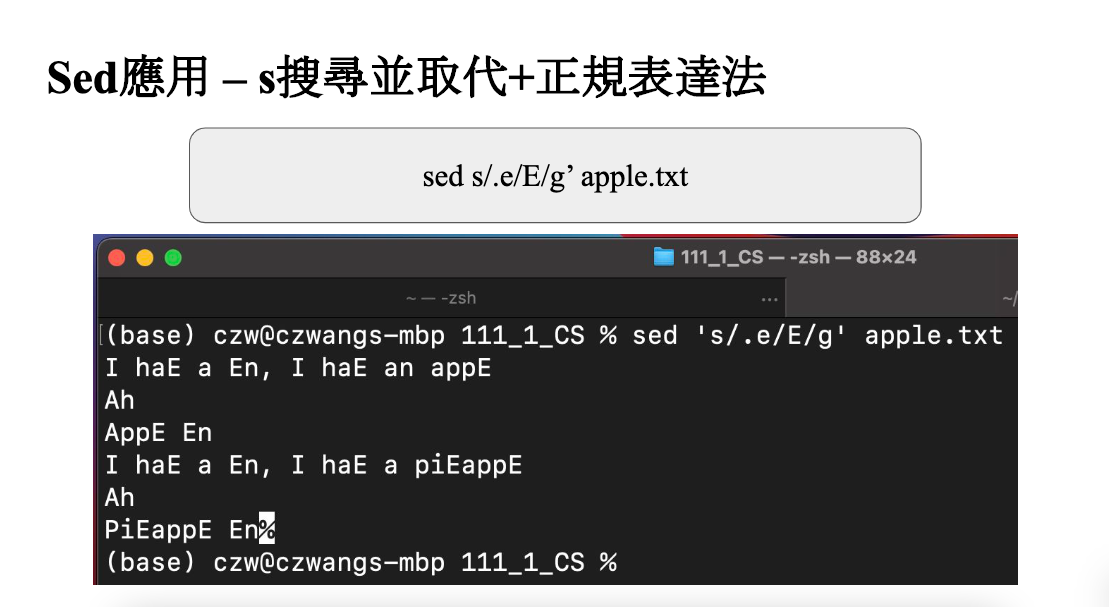
◆I : 忽略大小寫

◆w : 把符合的結果寫入檔案









**Grep, Awk, Sed比較**

➔grep：文字搜尋工具

◆可用正規表達法，找出匹配的內容

➔sed ：是一種線上編輯器

◆它一次處理一行內容，可搭配正規表達法

◆主要用來自動編輯一個或多個檔案，簡化對檔案的反覆操作

◆用於行間的內容操作，如新增、刪除、修改、查詢、替換

➔awk：文字分析工具

◆逐行的讀入，可搭配正規表達法

◆主要用在對文字和資料進行分析處理，以空格為預設分隔符號

◆同時也是程式語言

◆用於處理有欄位規則的行內內容，並支援格式化輸出