1. Sql query : 非管理階層員工，dependents數 > 2；列出EmpID、名字、主管名字、Project數

🡪SELECT A.empId, a.empname, b.empname, pnumber FROM

employee as b, works\_on,

(SELECT empid FROM employee, dependent

WHERE employee.empid = dependent.empid

GROUP BY employee.empid

HAVING COUNT(employee.empid) > 2) AS sub,

(SELECT empid, empname, supervisorid from employee, department

WHERE employee.empid not in (department.managerid)) AS a

WHERE a.empid = b.empid AND a.supervisorid = b.empid

1. File system block 計算
2. 用Binary search要存取多少次blocks，線性搜尋平均幾次?

🡪 一個block存5個records(512/100 = 5)

🡪 800,000 / 5 = 16,000 個 blocks

Binary access time = log(16,000) = 14

Average access time = 16,000/2 = 8,000

1. primary index, key field = 10 bytes & block pointer = 10 bytes, 總共多少blocks，需存取幾次?

🡪1 entry = 10 + 10 = 20 bytes

→ 1 block = 512 / 20 = 25 entries

→ 16000 / 25 = 640；Log640 = 10次

→ 640 塊

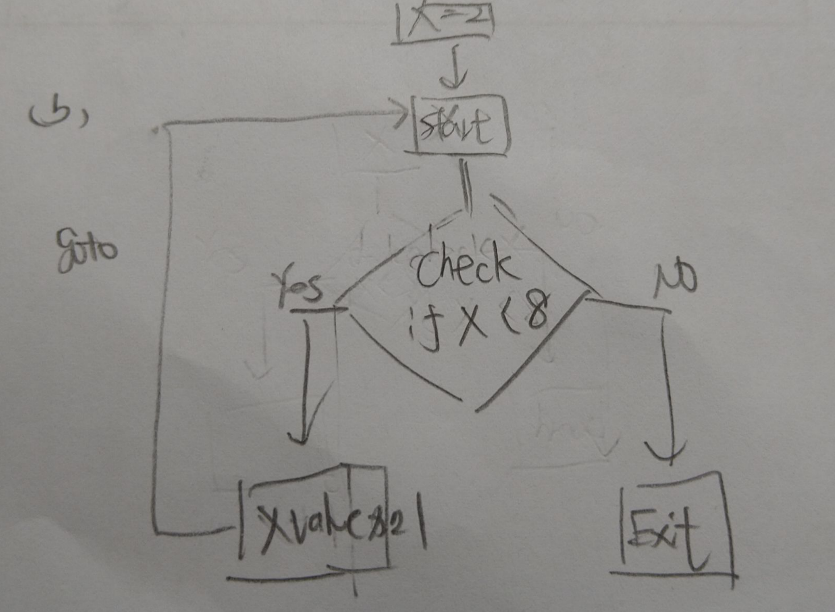
1. secondary index, key field = 10 bytes & block pointer = 10 bytes, 總共多少blocks，需存取幾次?

🡪80000/25 = 3200；log3200 = 12次

→ 3200 塊

1. multilevel index，entries數，需存取幾次

🡪3200/25 = 128個；log128 = 7 次

1. 單字配對(資料探勘 + DBMS)
   1. N
   2. K
   3. G
   4. I
   5. B
2. 為甚麼在高階程式語言GOTO不受歡迎
   1. 會使得程式可讀性變差不好維護，讓程式結構鬆散，增加錯誤的機會
   2. 
3. (a) 簡單敘述class與object

Class 🡪 用來定義物件的特徵，類似物件的設計圖、範本，內部包含該class的屬性與方法，以class 狗為例子，皆有特徵(屬性)有長短毛、毛色、大小，可以跑、跳(方法)

Object 🡪 為具現化class的實例(instance)，接續上述，吉娃娃和柴犬都是class狗的object，但其特徵內容會有不同的差異，例如:大小不同，但都是class狗所以可以跑或跳

(b) 為甚麼online db不被用來做資料探勘

資料探勘多數使用資料倉儲等已經不會再變動更改的過去歷史資料作分析，而線上資料庫的資料內容仍有變動的可能性不適合用來做資料探勘

1. LSTM, GAN AI
   1. Overfitting為何?如何避免?

🡪所訓練之模型與其內部參數太過吻合traing test，導致無法預測普遍之資料集，也就是variance小旦bias大即離實際目標遠的意思；可以減少模型的複雜度、增加樣本大小、正規化、dropout

* 1. 線性回歸與深度學習差異，與是否同意b的敘述

🡪 線性回歸為利用線性方程式在誤差最小之情況於不同參數間找到關係的模型；深度學習則是使用多層神經網路與激勵函數所搭建起的計算模型以進行預測；同意，深度學習仰賴資料大小，當資料過小的時候可能會有overfitting的問題，訓練難度上會比線性回歸高

* 1. Gan為何提出你的方法改善模型

🡪 GAN的方法訓練兩組模型，一組以正確資料訓練之模型負責鑑識稱為Discriminator network，另一組則負責產生資料稱為Generator network，不斷產生資料並讓鑑識模型負責判斷，生成模型根據結果修正模型內部參數，直到產出之資料能讓鑑識之模型認為是正確資料為止；可以訓練GAN生成假新聞，以提供負責辨識假新聞之模型樣本不足之問題，讓模型不會overfitting

* 1. Lstm為何

🡪為rnn衍生之可長期記憶的深度學習模型，具有四大元素 : input gate、forget gate, memory cell, output gate，因為有memory cell與forget gate，因此在計算時會考慮先前的結果並根據結果決定是否需要輸出或與要選擇遺忘哪些資料，可用於對於時間序列等問題，例如語音辨識

1. 評估模型之方法(F - Score & Confusion Matrix)
   1. Accuracy :
   2. False Positive Rate :
   3. F1 score :