

《資料庫應用》

一、在大數據 (Big Data) 時代, NoSQL 資料庫已經是最常被使用的資料儲存解決方案, 而在各種 NoSQL 資料庫中, 文件式資料庫 (Document-oriented Database) 及鍵值式資料庫 (Key-Value-oriented Database) 為目前最常被使用的資料庫:

(一) 試說明文件式資料庫及鍵值式資料庫的資料儲存特性。(16 分)

(二) 若某一傳統關聯式資料庫有以下員工資料表, 試說明文件式資料庫儲存此資料表的情形。(7 分)

員工編號	員工姓名	性別	專長
001	林大明	男	網路、資料庫
021	張小可	女	網路、多媒體
031	許中天	男	資料庫

(三) 試說明鍵值式資料庫儲存上述資料表的情形。(7 分)

試題評析	本題考 NoSQL 資料庫, 不過出的算簡單, 同學只要知道文件式資料庫及鍵值式資料庫的概念應該都順利作答。
考點命中	《高點·高上資料庫總複習講義》, 唐箏編撰, 頁 10-17。

【擬答】

(一) 鍵值式資料庫是將一筆資料的結構簡化到只有一個 Key 值對應到一個 Value 值。每一筆資料之間沒有關聯性, 所以, 可以任意切割或調整, 也可以分散到不同的伺服器中建立副本。例如: 假設在以下資料表

員工編號	姓名	身分證字號
0001	王曉明	A123456789

可將 Primary key 加上欄位名稱形成新的 Key, 用 Key-Value 方式儲存, 會變成:

Key	Value
0001_姓名	王曉明
0001_身分證字號	A123456789

(二)

文件資料庫主要是用來儲存非結構性的文件, 對於一個給定物件資料, 文件資料庫將其所有資訊儲存在資料庫的單一文件中。如上方的例子, 可以存成一個 document

```
{
  員工編號: "0001"
  姓名: "王曉明",
  身分證字號: "A123456789",
}
```

```
{
  員工編號: "001"
  員工姓名: "林大明",
  性別: "男",
  專長: "網路、資料庫"
}
```

```
{
  員工編號: "021"
  員工姓名: "張小可",
  性別: "女",
```

【版權所有，重製必究！】

```

專長："網路、多媒體"
}
{
員工編號："031"
員工姓名："許中天",
  性別："男",
  專長："資料庫"
}

```

(三)

```

001_員工姓名      林大明
001_性別          男
001_專長          網路、資料庫
021_員工姓名      張小可
021_性別          女
021_專長          網路、多媒體
031_員工姓名      許中天
031_性別          男
031_專長          資料庫

```

二、資料庫從傳統關聯式 (Relational) 持續進展至最近的 NoSQL及NewSQL等新一代資料庫，試說明此三種資料庫的特性，並針對每一種資料庫列舉三個實例。(25 分)

試題評析	本題考了這幾年才出來的 NewSQL 資料庫，算非常新的題目。往後考試也可能會考類似的考題，同學要多注意 NewSQL 資料庫的發展。
考點命中	《高點·高上資料庫總複習講義》，唐箏編撰，頁 7-15。

【擬答】

(一)關聯資料庫是由關聯(Relation)的集合所組成的，關聯本身皆經正規化並且以共享值(shared value)的方式相互關聯，現實世界中的各種實體以及實體之間的各種聯繫均用關聯模型來表示。關聯資料庫使用標準資料查詢語言SQL。關聯資料庫需要有ACID特性，也就是在寫入或更新資料的過程中，要保證transaction是正確可靠的，所必須具備 1. Atomicity 不可分割性 2. Consistency 一致性 3.Isolation 獨立性 4.Durability 持久性。

例如：MySQL、PostgreSQL、Microsoft SQL Server

(二)NoSQL資料庫不使用SQL作為查詢語言。其資料儲存可以不需要固定的表格模式，可以儲存半結構化，或非結構化資料。NoSQL強調可以分散式儲存跟分散式計算。NoSQL並不要求ACID，更新速度有快有慢，只要最終資料能有一致性即可。

例如：MongoDB、BigTable、Hadoop HBase

(三)NewSQL是指新的可擴展/分散式運算資料庫的簡稱，這種新的DB不僅具有NoSQL大量資料的平行運算、儲存、管理能力，還能夠支持關聯資料庫的ACID和SQL等特性

例如：Google Spanner、CockroachDB、Apache Ignite

三、磁碟陣列RAID技術可用來提供容錯功能，以避免資料遺失，其做法可以分成很多種等級，試描述這些不同等級RAID的特性及優缺點。(20 分)

試題評析	本題老師把 RAID 全部列了出來。但是考試時，如果只寫常用的 Raid 0, Raid 1, Raid 5, Raid 6 寫清楚應該也可以獲得高分。
考點命中	《高點·高上資料庫講義》第三回，唐箏編撰，頁 107-109。

【擬答】

- (一)RAID 0 將兩個以上的磁碟並聯起來，成為一個大容量的磁碟。在存放資料時，同時對這些磁碟讀寫。
優點：因為平行處理，所以在所有的級別中，RAID 0 的速度是最快的。
缺點：RAID 0 不具備容錯能力，如果其中一個磁碟損壞，所有資料都會遺失，危險程度比單顆硬碟高。
- (二)RAID 1 是用磁碟相互作鏡像，其原理為在主硬碟上存放資料的同時也在鏡像硬碟上寫一樣的資料。當主硬碟損壞時，鏡像硬碟則代替主硬碟的工作。
優：只要一個磁碟正常即可維持運作，可靠性最高。
缺：無論用多少磁碟做RAID 1，僅算一個磁碟的容量，是所有RAID中磁碟利用率最低的一個級別。
- (三)RAID 2 這是RAID 0 的改良版，以漢明碼的方式將資料進行編碼後分割為獨立的位元，並將資料分別寫入硬碟中。因為在資料中加入錯誤修正碼，所以資料整體的容量會比原始資料大，只比RAID1 略好。
優：磁碟利用率比RAID 1 高。
缺：RAID 2 最少要三台磁碟機方能運作。資料利用率並沒有提升很多。很少使用。
- (四)RAID 3 採用Bit-interleaving 技術，它需要通過編碼再將資料位元分割後分別存在硬碟中，再將同位元檢查後單獨存在一個硬碟中，但由於資料內的位元分散在不同的硬碟上，因此就算要讀取一小段資料都可能需要所有的硬碟進行工作。
優：磁碟利用率比RAID 2 高。
缺：讀取少量資料效率不佳。很少使用。
- (五)RAID 4 它與RAID 3 不同的是它在分割時是以Block interleaving 的方式分別存在硬碟中，但每次的資料存取都必須從同位元檢查的那個硬碟中取出對應的同位元資料進行核對。
優：讀取少量資料效率比RAID 3 好。
缺：同位元檢查硬碟成為效能瓶頸、易壞。很少使用。
- (六)RAID 5 可以理解為是RAID 0 和RAID 1 的折衷方案，是把資料和相對應的奇偶校驗資訊儲存到組成RAID5 的各個磁碟上，並且奇偶校驗資訊和相對應的資料輪流儲存於不同的磁碟上。當RAID5 的一個磁碟資料發生損壞後，可以利用剩下的資料和相應的奇偶校驗資訊去恢復被損壞的資料。
優：RAID 5 具有和RAID 0 相近似的資料讀取速度，只是因為多了一個奇偶校驗資訊，寫入資料的速度相對單獨寫入一塊硬碟的速度略慢。磁碟空間利用率要比RAID 1 高，儲存成本相對較便宜。
缺：1.只能壞一顆硬碟。 2.計算量大，通常需要硬體幫忙。
- (七)RAID 6 與RAID 5 相比，RAID 6 增加第二個獨立的奇偶校驗資訊塊。兩個獨立的奇偶系統使用不同的演算法，資料的可靠性非常高。
優：任意兩塊磁碟同時失效時不會影響資料完整性。效能指比RAID 5 略差。
缺：1.必須具備四個以上的磁碟才能生效。2.計算量大，通常需要硬體幫忙。

四、某企業計畫開發一套資料庫應用系統，其需求如下所述：

- 1.需儲存供應其貨品的供應商相關資料，包含有供應商編號、名稱、地址、聯繫人姓名及電話號碼。
- 2.對於每次貨品供應，需建立一個帳單以儲存此次供應的日期、金額、付款期限以及有關說明。
- 3.每筆帳單的付款，可以在不同的時間以不同的方式支付（例如：現金、支票、信用卡），並記錄每次付款的日期及方式。
- 4.每個供應商可以多次供應貨品，唯每日最多只能供貨一次。

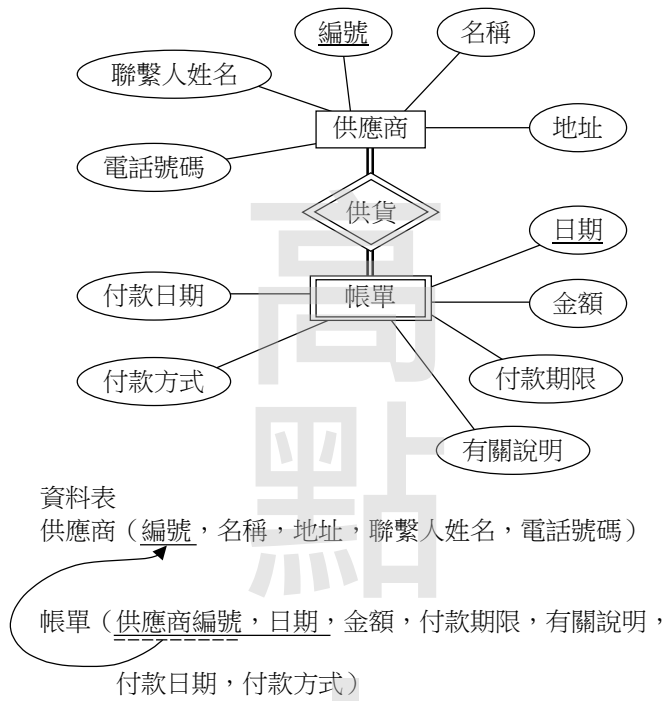
(一)試針對上述需求，描繪出實體關係圖。(15分)

(二)試依據實體關係圖，設計出此系統的資料表。(10分)

【版權所有，重製必究！】

試題評析	本題考實體關係圖的弱實體，同學應該能順利回答。
考點命中	《高點·高上資料庫講義》第一回，唐箏編撰，頁 29-35。

【擬答】



高上

【版權所有，重製必究！】