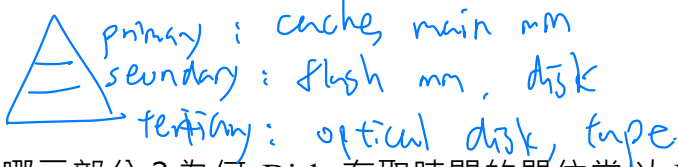

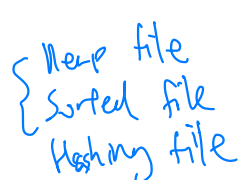


# DBMS 2024 Spring 期末考參考題綱

\* 這是題綱，不是題庫。僅供同學複習核檢用，考題與題型與題綱沒有直接關係。

1. 何謂 Storage Hierarchy ? 
2. Disk 存取資料的時間包括哪三部分？為何 Disk 存取時間的單位常以 Block I/O 次數來代表？  $\text{Disk Access Time} = \text{seek time} + \text{latency time} + \text{transfer time}$
3. 何謂 Disk Scheduling ? The method by which the OS decides the order in which disk I/O requests are served
4. 何謂 Fixed Length Records ? 何謂 Variable Length Records ? Variable Length Records 有哪幾種 ? Relational Data Model 有 Repeating Fields, Optional Fields, Mixed File 嗎 ? NoSQL 的 Document-Based Model 有 Repeating Fields, Optional Fields, Mixed File 嗎 ?
5. 何謂 Blocking Factor ?  $\frac{\#(\text{records})}{\text{disk block}}$  unspanned vs. spanned
6. 何謂 Spanned, Un-spanned Records ?  variable-length field, repeating field, optional field, mixed file
7. 何謂 Primary File Organization ? "data file"
8. Primary File Organization 有哪三種？各有何優缺點？ 
9. Sorting File 如何 Binary Search ?  $\rightarrow \text{Algo. (11 p.53)}$
10. Extendible Hashing 如何處理 Insertion ? Hashing Function 如何 Dynamically 隨著 Data Insertion 而變化 ?

- 
11. 已知 Primary File 的 Records 筆數、每筆 Record 所需 Byte 數，給定硬碟的 Block Size，其 Blocking Factor 是多少？此 Primary File 共有多少個 Blocks ? Linear Search Worst Case

$$BF = \frac{\#(\text{records})}{\text{block}} = \frac{C}{B}$$

$$N = \frac{A}{\frac{C}{B}} = \frac{AB}{C}$$

需要存取多少個 Blocks ? Binary Search Worst Case 需要存取多少個 Blocks ?

12. 何謂 Secondary File Organization ? "index file"

Does it have index entry to every search value?

(Q13) 13. Single Level Secondary File Organization 有哪四種 ? 請舉例畫出這四種架構圖。

14. 什麼情況建立的 Index 是 Primary Index ? Clustering Index ? Secondary Index ? 哪些 Index

是 Dense Index ? 哪些是 Sparse Index ?

	ordering file	otherwise
key	primary	secondary (key)
non-key	clustering	" (non-key)

15. 建立 Secondary File (Index File) 有何優缺點 ?

{ pros: look up performance  
cons: storage, maintenance

(Q16) 16. 已知硬碟 Block Size、Primary File 的 Records 筆數、每筆 Record 所需 Byte 數，若根據

其中一個 Non-Key Field 建立 Secondary Index，已知此 Non-Key Field 的 Byte 數及 Block

Pointer(Block Address) 的 Byte 數，此 Secondary Index File 的 Blocking Factor 是多少 ? 共

有多少個 Index Entries ? 此 Index 共需多少個 Blocks ?

$$BF = \frac{C}{F + P}$$

17. 何謂 B+Tree ? B+Tree 如何處理 Query ? 需要存取多少 Blocks ? B+Tree 如何處理

Insertion ?

18. 何謂 Range Query ? 何謂 Multi-attribute (Multi-dimensional) Query ?

(Q19) 19. 何謂 Bit Map Index ? Bit Map Index 如何處理 Query ?

20. 資料庫系統如何將一個 SQL Query Transform 成 Relational Algebra 的運算式 ?

21. Query Processor 有哪些方法可以處理 Selection Operation ? Projection Operation ? Join

Operation ? Union, Intersection Operation ? Aggregation Operation ?

22. 為何資料庫系統可以做 Query Optimization ? Query Optimization 有哪幾種 Approaches ?

23. Heuristic based Query Optimization 如何將 SQL Query 做 Query Optimization ? 包括哪些

步驟 ?

- nested-loop join  
- index-based "  
- sort-merge join  
- hash join

1. SELECT → 6 AND 6 AND...

2. move 6 down the tree

3. rearrange leaf nodes → the most restrictive operations go to the left of the tree

4. 8 + 6 = 14 5. move down to create 2

6. identify subtrees → 7 also.

24. 如果 Employee 資料表有 10,000 筆記錄，且根據薪水欄位 Salary 排序儲存在 2,000 disk blocks 中。假設已經根據部門代號欄位 Dno 建立 Secondary Index，且此 Secondary Index 是 2 層的 B<sup>+</sup>-tree。若這家公司共有 125 個部門，且假設每個部門人數都差不多。針對  $\sigma_{Dno=2}(Employee)$  這查詢，Cost-based Query Optimization 會如何處理？如果 Query Plan 是直接在 Primary File 做 Linear Search，其 Cost 是多少個 Block I/O？如果 Query Plan 是運用 Secondary Index，其 Cost 是多少個 Block I/O？

25. Query Optimization 如何有效地處理多個 Join Operation？

26. 何謂 Semantic Query Optimization？請舉例說明。

27. 資料庫系統的效能調校可由哪三個方向著手？

28. 調校資料庫系統的效能時，需要知道哪些統計資訊？

29. 何謂 Denormalization？為何要做 Denormalization？

30. 何謂 Vertical Partitioning？何謂 Horizontal Partitioning？何謂 Sharding？適用時機為何？

31. Tuning Query 有哪些常見的技巧？這些技巧背後的原因為何？為何要將有 OR 條件的 SQL query 轉換成 Union？

32. Tuning Index 必須考慮哪些事情？

33. 何謂 Transaction？何謂 Transaction Processing？

34. 為何資料庫系統需要 Concurrency？需要 Recovery？

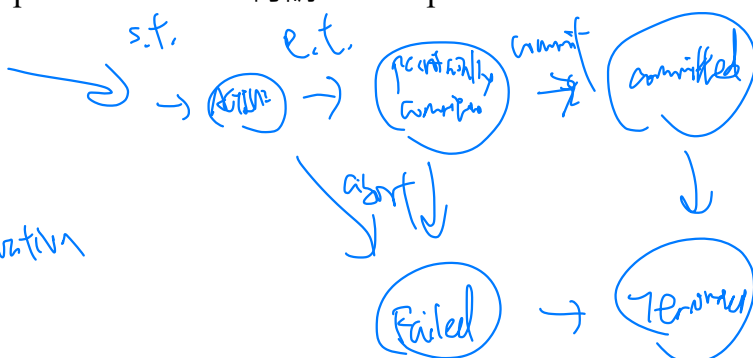
35. 兩個 Interleaved Transactions 會發生 Lost Update Problem。何謂 Lost Update Problem？

36. 資料庫系統的 Transaction States 為何？

37. 何謂 ACID？

Atomicity  
Consistency  
Isolation  
Durability

Preservation



38. 何謂 Conflict Operations ?

Conflict equivalent to serializable schedule  
the order of 2 conflict operations is the same in both schedules  
何謂 Serial Schedule ? 何謂 Serializable Schedule ? 何謂

Conflict Serializable Schedule ?

39. 如何判斷是否 Conflict Serializable Schedule ?

conflict serializable  $\Rightarrow$  acyclic in precedence graph

40. 多數資料庫系統如何做到 Serializable Schedule ?

2-phase locking

41. 相較於傳統 Relational Data Model, NoSQL 的特性為何 ?

semi structured, self-describing  
high performance, replication  
sharding

42. 何謂 CAP Theorem ? CAP Theorem 與 ACID 的關係為何 ?

C consistency  
A availability  
P partition tolerance