

外部記憶體

磁碟 / RAID / SSD / 光學儲存 / 磁帶

摘要

本章討論記憶體階層底部的大容量外部儲存。首先介紹磁碟的物理結構、資料組織與存取時間模型 (尋道、旋轉延遲、傳輸); 接著詳述以冗餘換取效能與可靠度的 **RAID 0-6** 各級; 然後補充 Stallings 第 6 版之後的重大變革——固態硬碟 (**SSD**); 最後簡述光學儲存與磁帶。本章對應 Stallings 第 6 章並更新至現代。

目錄

1	磁碟	2
1.1	物理結構與資料組織	2
1.2	存取時間模型	2
2	RAID	2
3	固態硬碟 (SSD)	3
4	光學儲存與磁帶	3
5	本章重點回顧	3
6	複習問題	3

1 磁碟

1.1 物理結構與資料組織

磁碟是塗有磁性材料的圓形碟片 (platter), 讀寫頭懸浮於其上。資料組織:

- 磁軌 (track): 同心圓環; 相鄰磁軌間有間隙以防干擾。
- 磁區 (sector): 磁軌上的固定大小區段 (傳統 512 B, 現代 4 KiB), 是讀寫的最小單位。
- 磁柱 (cylinder): 多碟片機構中, 所有碟面上半徑相同的磁軌集合。
- 定速旋轉 (CAV) 下, 外圈磁軌線密度低; 現代磁碟採分區位元記錄 (multiple zone recording), 外圈磁軌容納更多磁區以提高容量。

1.2 存取時間模型

$$T_{access} = T_{seek} + T_{rot} + T_{transfer}$$

- 尋道時間 T_{seek} : 讀寫頭移到目標磁軌, 平均約 4–9 ms;
- 旋轉延遲 T_{rot} : 等待目標磁區轉到讀寫頭下, 平均為半圈時間。7200 RPM 時 = $0.5 \times 60 / 7200 = 4.17$ ms;
- 傳輸時間 $T_{transfer} = b / (rN)$: b 為要傳的位元組、 N 為每磁軌位元組、 r 為轉速。

循序 vs. 隨機存取的巨大差距

磁碟: 平均尋道 4 ms、7200 RPM、每磁軌 512 個 4 KiB 磁區。讀取連續的 2560 個磁區 (10 MiB): 第一磁區需 4 + 4.17 ms, 其後順序讀完 5 圈約 5×8.33 ms, 總計約 50 ms。若同樣 2560 個磁區隨機散布: 每個都要 4 + 4.17 + 0.016 ms, 總計約 **21** 秒——相差 400 倍。作業系統的排程與檔案配置 (第 8 章)、資料庫的存取模式設計, 皆源於此。

2 RAID

1988 年 Berkeley 提出 **RAID**(Redundant Array of Independent Disks): 以多顆磁碟組成邏輯上單一的大磁碟, 藉條帶化 (striping) 提升平行度、藉冗餘容忍磁碟失效。

級	技術	冗餘方式	特點
0	條帶化	無	純效能/容量; 任一碟壞即全毀
1	鏡射	完整複製	讀可雙路、寫需兩份; 成本 2 倍; 恢復最簡單
2	位元條帶 + 漢明碼	漢明碼	需同步主軸; 實務上未被採用
3	位元/位元組條帶	單一同位碟	高循序傳輸率; 每次 I/O 動用全部碟
4	區塊條帶	單一同位碟	同位碟成為寫入瓶頸
5	區塊條帶	分散式同位	同位輪流放各碟, 消除瓶頸; 最常用
6	區塊條帶	雙重分散同位 (P+Q)	容忍兩碟同時失效; 重建期間安全

重點觀念

RAID 5 的寫入代價: 更新一個資料區塊需「讀舊資料、讀舊同位、寫新資料、寫新同位」共 4 次 I/O(read-modify-write); 同位計算為 $P_{new} = P_{old} \oplus D_{old} \oplus D_{new}$ 。XOR 同位的原理與第 5 章的同位檢查一脈相承。

3 固態硬碟 (SSD)

SSD 以 NAND 快閃記憶體構成, 無機械部件:

- 效能: 無尋道與旋轉延遲, 隨機讀取比磁碟快 2–3 個數量級; NVMe 介面直連 PCIe, 消除 SATA 瓶頸。
- 寫入特性: 快閃「先抹除後寫入」且抹除以大區塊為單位, 因此控制器採寫入放大管理、磨損平均 (**wear leveling**) 與 **FTL**(快閃轉譯層) 把邏輯區塊映射到實體頁。
- 壽命: 每單元抹寫次數有限 (SLC>MLC>TLC>QLC); 以 TBW/DWPD 標示耐寫度。
- 階層定位: SSD 填補了 DRAM 與磁碟之間的空隙; 資料中心常見 DRAM–SSD–HDD–磁帶四級儲存。

4 光學儲存與磁帶

- **CD-ROM/CD-R/CD-RW**(650–700 MB): 單一螺旋軌、定線速度 (CLV); 雷射讀取坑紋反射差。
- **DVD**(4.7–17 GB): 更短波長雷射、更密坑紋、雙層雙面。
- **Blu-ray**(25–128 GB): 藍紫光雷射。光學媒體今日主要用於冷資料發行與典藏。
- 磁帶: 循序存取, 存取時間以秒計, 但每 TB 成本最低、離線保存最安全; LTO 世代持續演進 (LTO-9 原生 18 TB), 仍是備份/歸檔的主力。

5 本章重點回顧**本章要點**

- 磁碟存取時間 = 尋道 + 旋轉延遲 + 傳輸; 循序與隨機存取效能差距可達數百倍。
- RAID 以條帶化與冗餘平衡效能、容量與可靠度; RAID 5/6 為主流, RAID 0 無冗餘。
- SSD 消除機械延遲, 但需 FTL、磨損平均管理快閃特性。
- 磁帶以最低成本承擔冷資料; 儲存階層依「熱度」分級部署。

6 複習問題

1. 磁碟 10000 RPM、平均尋道 3 ms、每磁軌 1 MiB。求讀取一個 4 KiB 磁區的平均時間。
2. 分區位元記錄解決什麼問題?

3. RAID 1 與 RAID 5 在成本與寫入效能上如何取捨?
4. 說明 RAID 5 小量寫入的 4 次 I/O 代價;RAID 6 為何在大容量碟時代更受青睞?
5. SSD 為何需要磨損平均與 FTL?
6. 你要為「每天寫入一次、保存十年、極少讀取」的資料選儲存媒體, 選什麼? 為什麼?

參考資料

- [1] W. Stallings, *Computer Organization and Architecture*, 6th ed., Chapter 6.
- [2] D. Patterson, G. Gibson, R. Katz, “A Case for Redundant Arrays of Inexpensive Disks (RAID),” *SIGMOD* 1988.
- [3] NVM Express, Inc., *NVMe Specification*. <https://nvmexpress.org>