

Storage and Indexing

CISC437/637, Lecture #11

Ben Carterette

Copyright © Ben Carterette

1

Physical Layer

- So far we have focused on high-level database layers
 - Conceptual: E-R model, E-R diagrams
 - Logical: relational model, SQL
- Ultimately the logical layer needs to be implemented in data structures and stored in files on disk
 - The **physical layer**
- Understanding how the files are organized is essential to using a DBMS effectively

Copyright © Ben Carterette

3

Physical Storage Media

- The actual physical objects that data is stored in
- It is convenient to classify media according to characteristics:
 - Speed of access
 - Cost per unit of storage
 - Reliability
 - Loss of data on power outage or crash
 - Failure of device
 - Volatility
 - Volatile storage loses contents when shut down
 - Non-volatile persists during shut down period

Copyright © Ben Carlsson

4

Physical Storage Media

- **Cache** – very fast; very expensive; volatile; managed by hardware/OS
- **Main memory** – fast; comparatively expensive; volatile
 - Usually too expensive to allow storage of entire DB
- **Flash memory** – fast read, slower write; roughly as costly as main memory; non-volatile

Copyright © Ben Carlsson

5

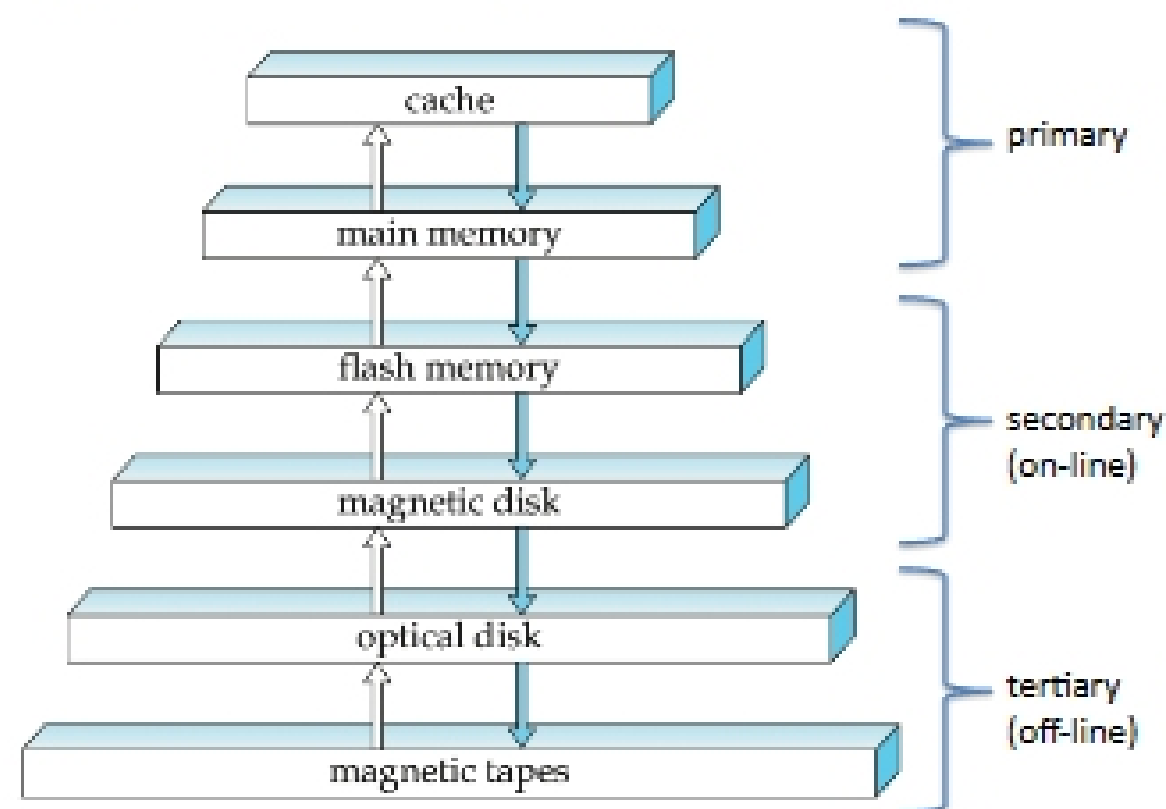
Physical Storage Media

- **Magnetic disk** – slow; cheap; non-volatile
 - This is where the full DB would be stored
 - DBMS must move data from disk to memory for access, and from memory to disk for storage
 - Doing this *efficiently* is the driving force behind storage decisions
- **Optical storage** – slow; cheap; non-volatile
 - CDs, DVDs
- **Tape** – very slow; very cheap; non-volatile
 - Only allows **sequential access**: data accessed in storage order from beginning
 - Compare to **direct access** offered by magnetic disks
 - Used for backups and archives

Copyright © Ben Carlsson

6

Physical Storage Hierarchy



Copyright © Ben Carlsson

7