

**Chapter Six**  
**File Organization & Indexing**

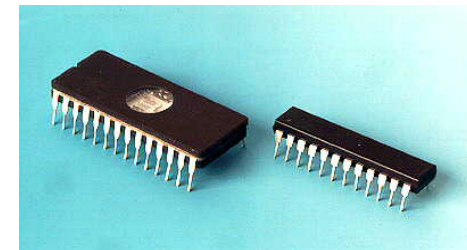


## File Organization

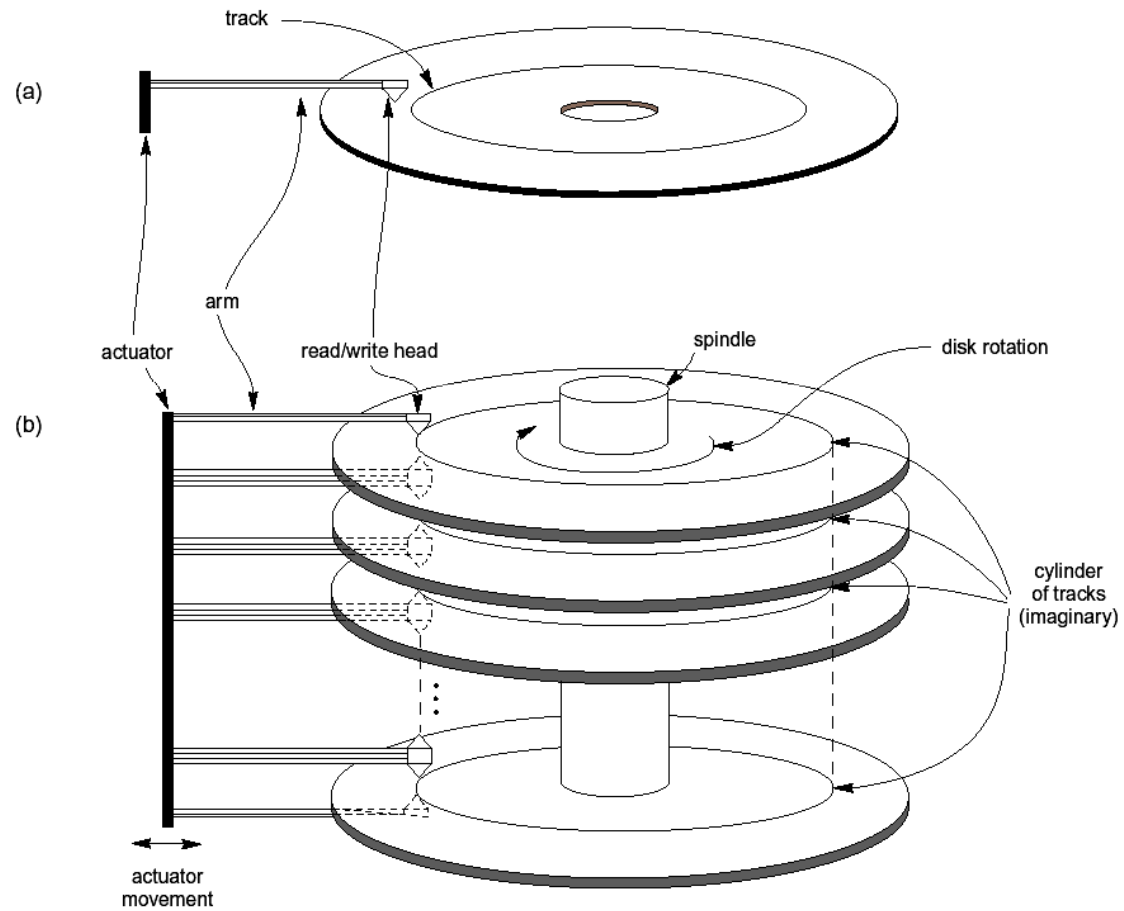
### Storage Hierarchy

- Primary
- Secondary

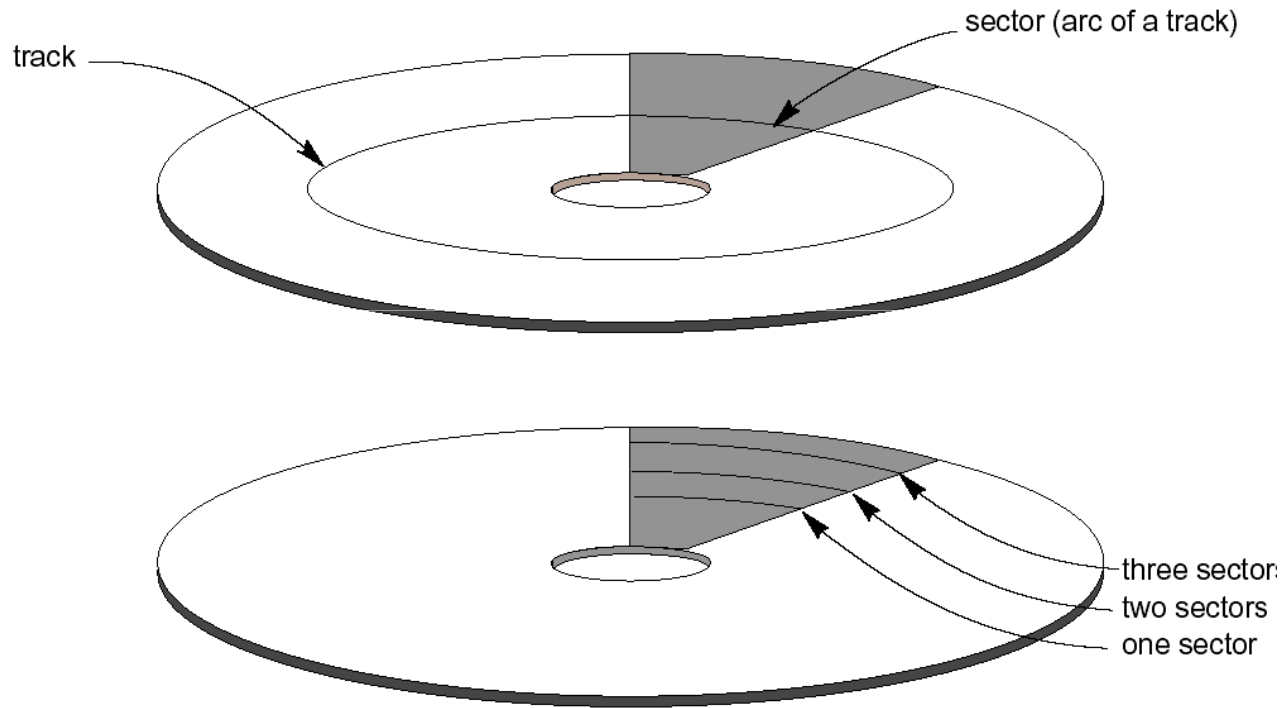
Factors: Capacity, Loss of Stored Data, Cost



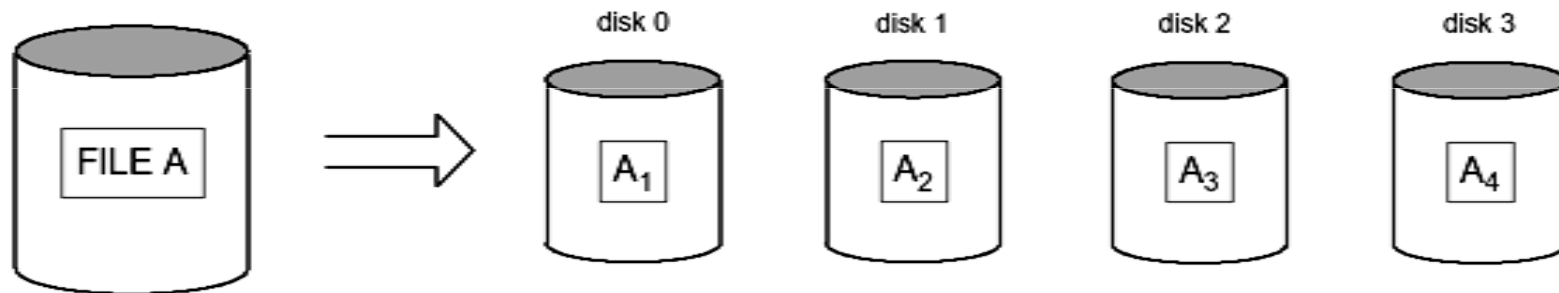
## Hard Disk Mechanism



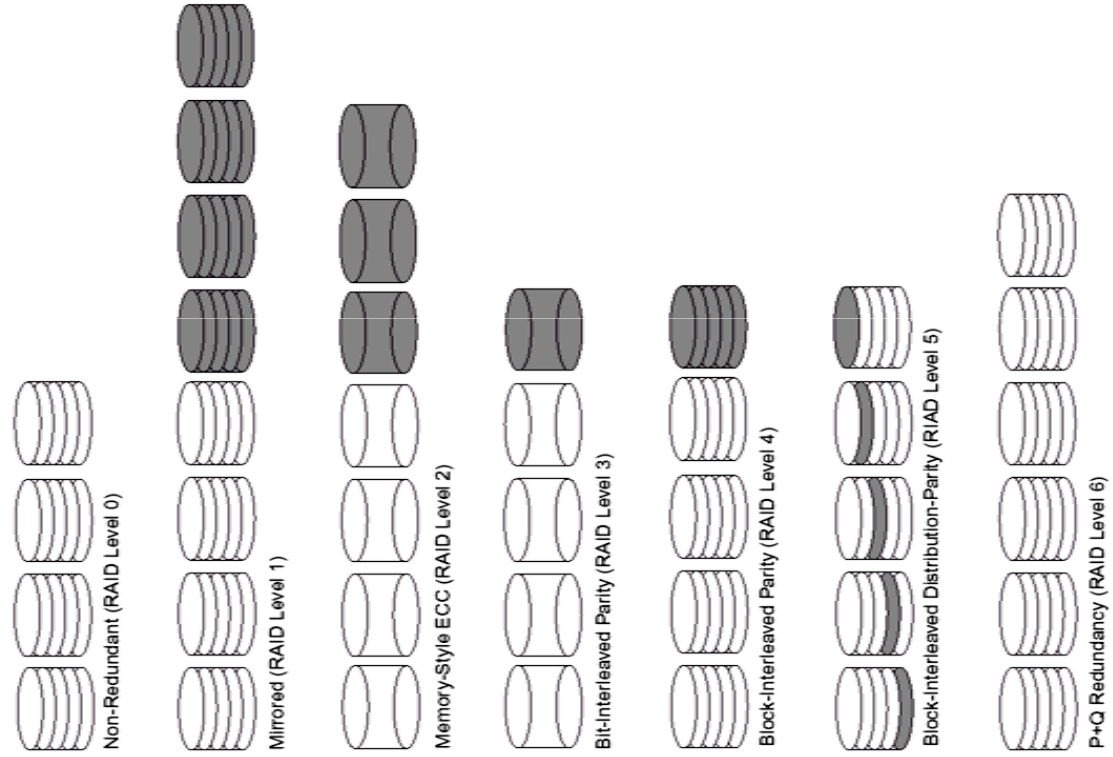
## Sectors and Tracks



Data Striping  
File A is striped across four disks



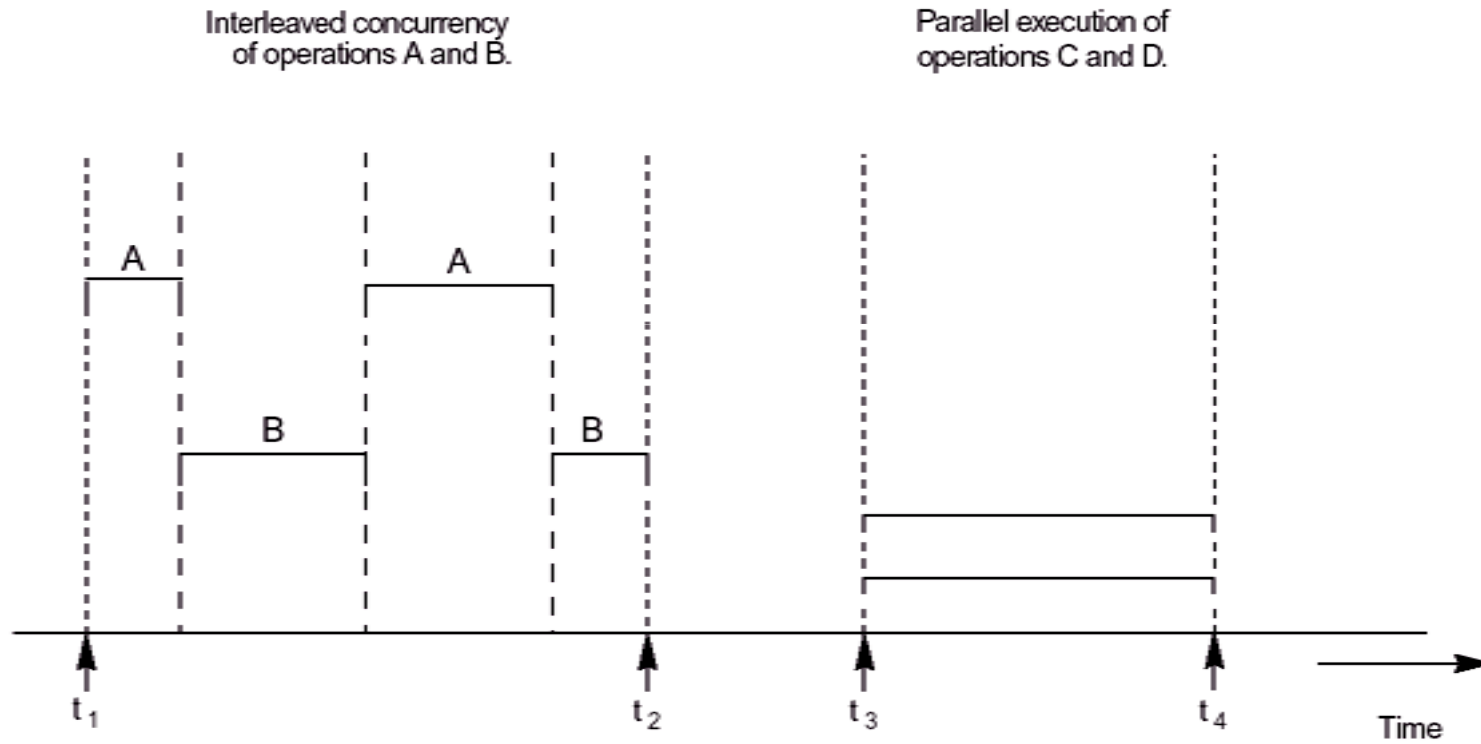
## RAID Levels



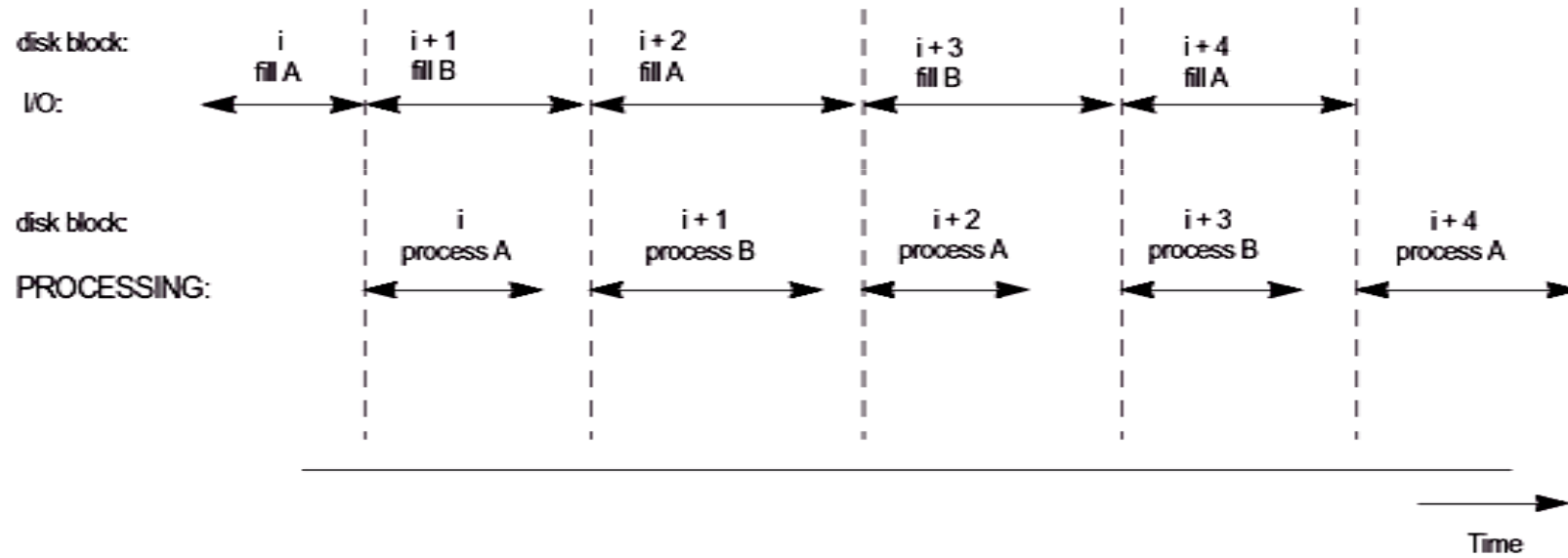
## Hardware Description of Disk Devices

- Units of data transfer (Block)
  - Hardware Address: surface, track, block
  - Access time: seek + latency + bulk transfer rate
  - Buffering of blocks: concurrently interleaved, concurrently simultaneous
-

## Interleaved concurrency versus parallel execution



Use of two buffers, A and B, for reading from disk



- n Data types: Int(4), long(8), real(4), Bool(1), date(4), fixed k character(kB)
- n Fixed length and variable length records
- n Spanned and unspanned records

bfr(blocking factor)=floor(block size/record size)  
in (records/blocks)

---

- n Linear Search: Number of blocks/2
  - n Binary Search:  $\log_2(\text{Number of Blocks})$
-

**Example:**

No. of records:30,000

Block size:1024B

File records (unspanned)

Bfr=10record/block

No. of blocks=3000 blocks

Binary search: 12 block

Linear search:1500 blocks



## Primary File Organization

### File Organizations:

#### 1.Heap file

#### Advantages

- i. Inserting a new record is very efficient
- ii. It is used to collect and store data records for future use

#### Disadvantages

- i. Searching is expensive
  - ii. In deletion a record, an extra unused space in the disk block is leaved
-

## 2. Primary index

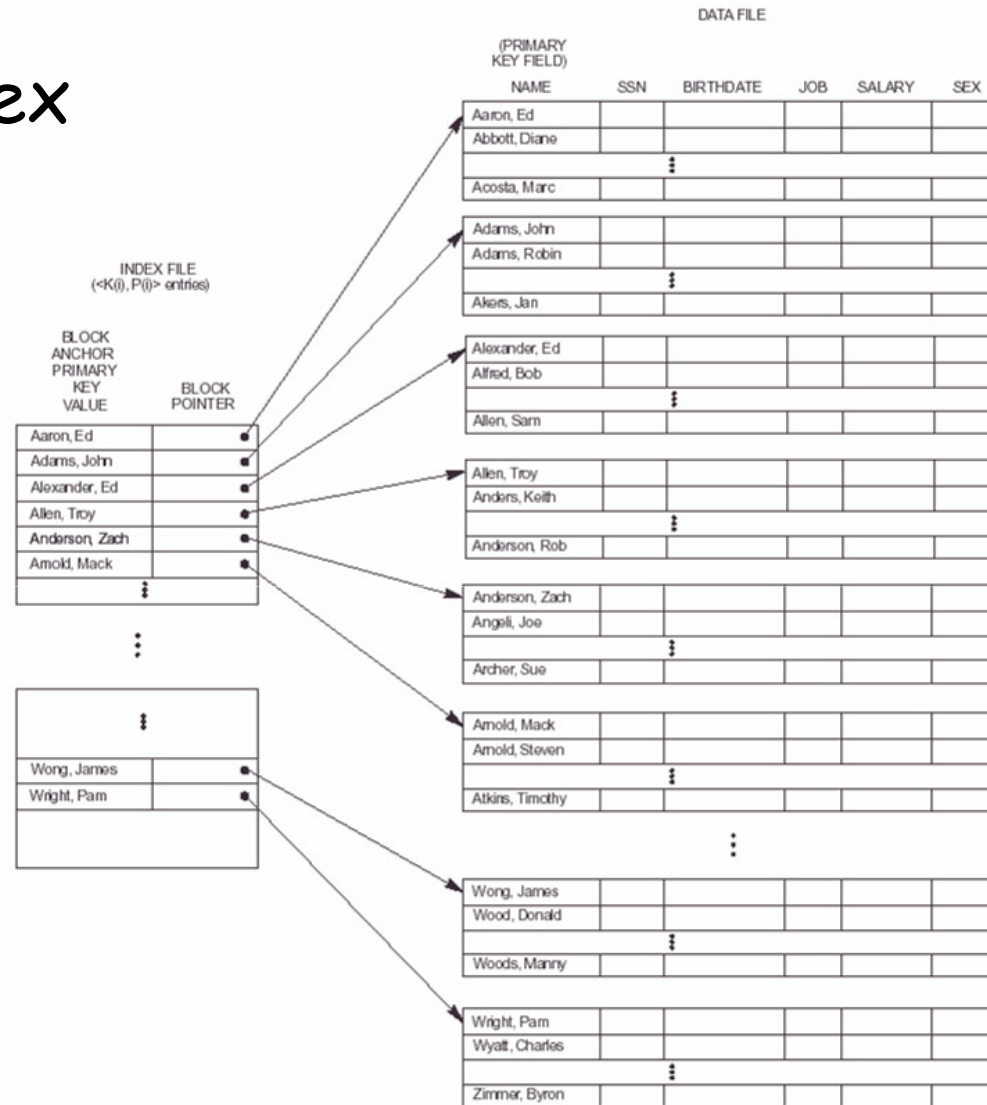
- n Defined on an ordered data file
  - n The data file is ordered on a **key field**
  - n One index entry for each block
-

# Ibrahim I Hamar

	NAME	SSN	BIRTHDATE	JOB	SALARY	SEX
block 1	Aaron, Ed					
	Abbott, Diane					
			⋮			
	Acosta, Marc					
block 2	Adams, John					
	Adams, Robin					
			⋮			
	Akers, Jan					
block 3	Alexander, Ed					
	Alfred, Bob					
			⋮			
	Allen, Sam					
block 4	Allen, Troy					
	Anders, Keith					
			⋮			
	Anderson, Rob					
block 5	Anderson, Zach					
	Angeli, Joe					
			⋮			
	Archer, Sue					
block 6	Arnold, Mack					
	Arnold, Steven					
			⋮			
	Atkins, Timothy					
			⋮			
block n-1	Wong, James					
	Wood, Donald					
			⋮			
	Woods, Manny					
block n	Wright, Pam					
	Wyatt, Charles					
			⋮			
	Zimmer, Byron					



# Ex. Primary index



**Example:**

No. of records:30,000

Block size:1024B

File records (unspanned), primary indexed

Bfr=10record/block

No. of blocks=3000 blocks

Binary search:?

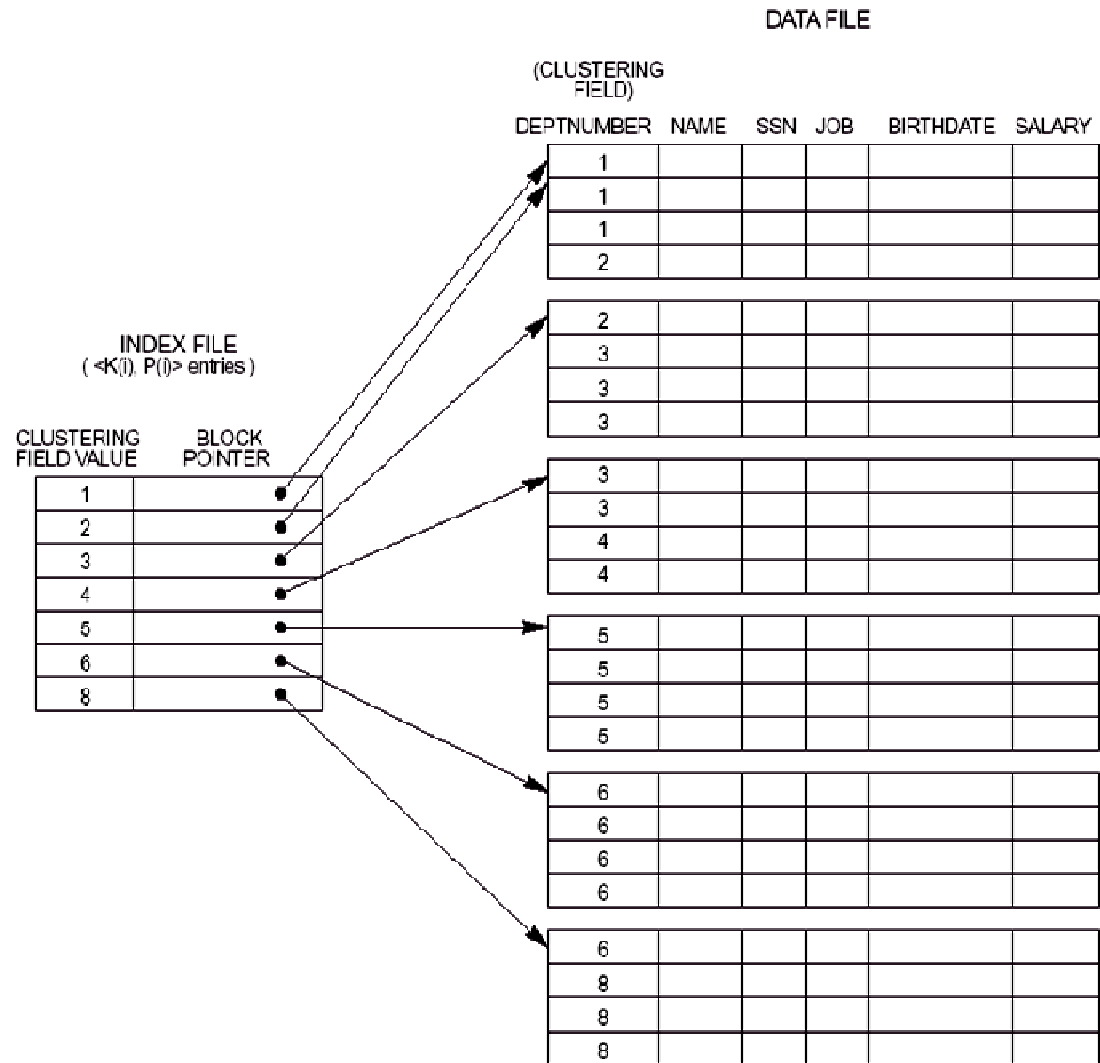
Linear search:?



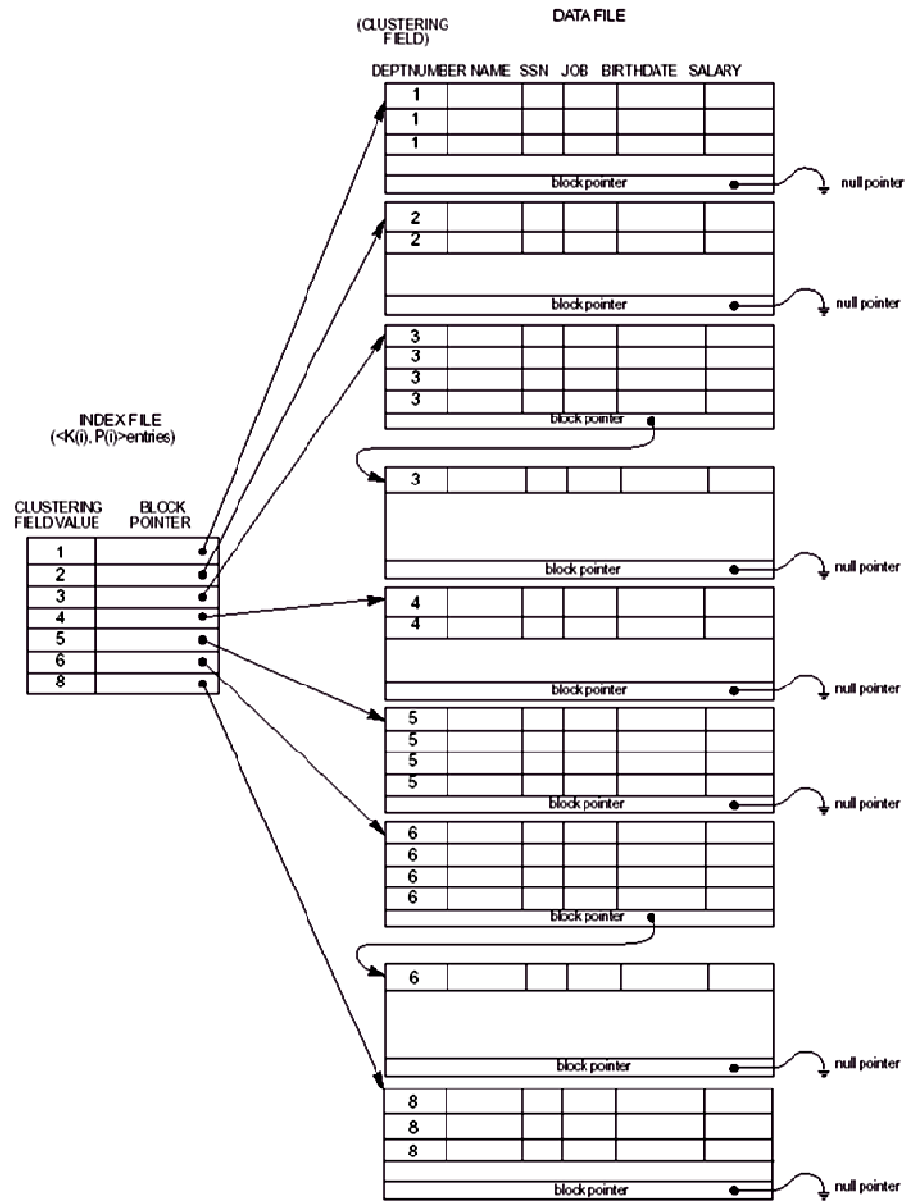
### 3. Clustering index

- n The data file is ordered on a non-key field including one index entry for each distinct value of the field
-

# Ex. Clusterina index ordered on a non-key field

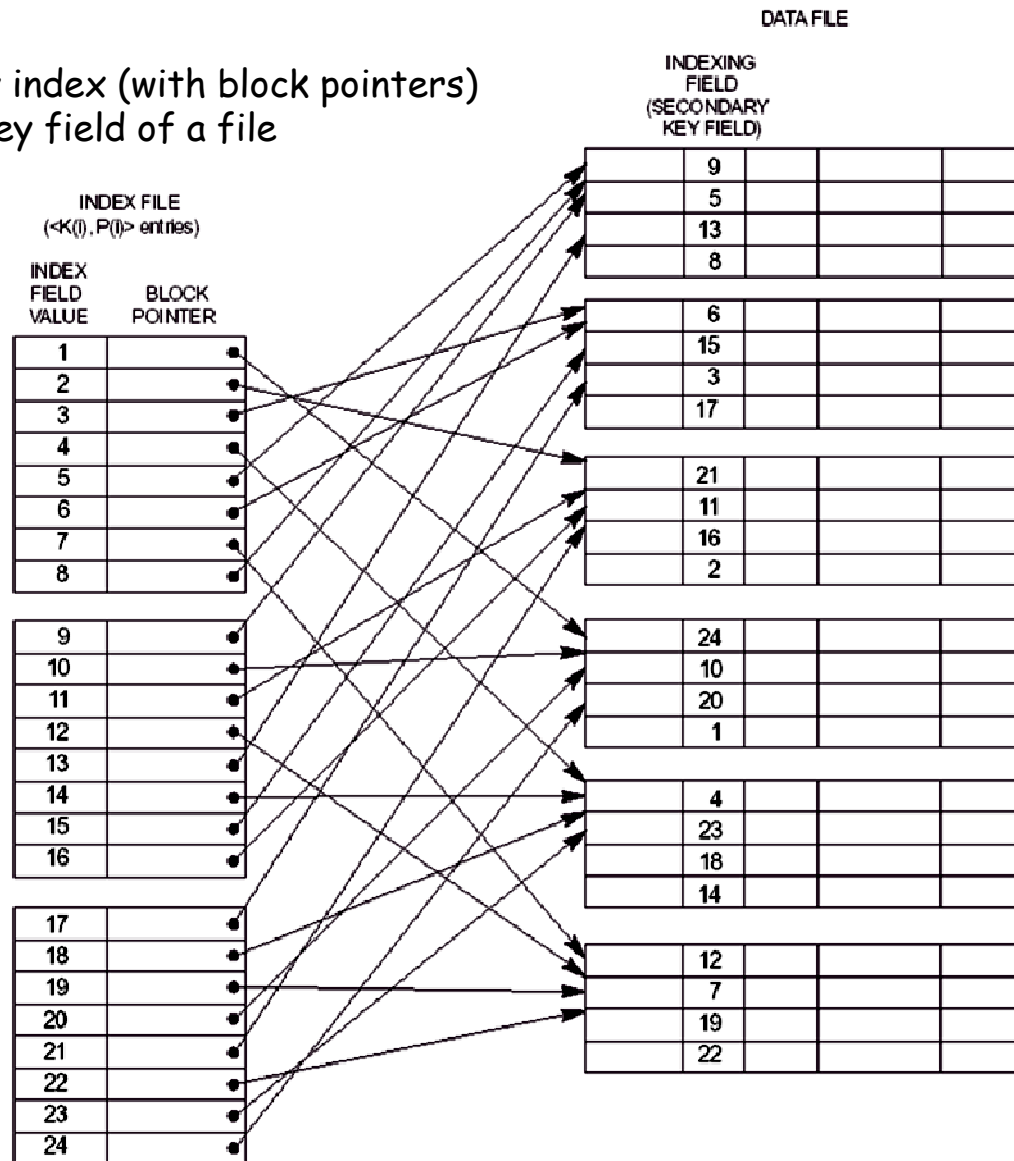


# Ibrahim I Hamarast

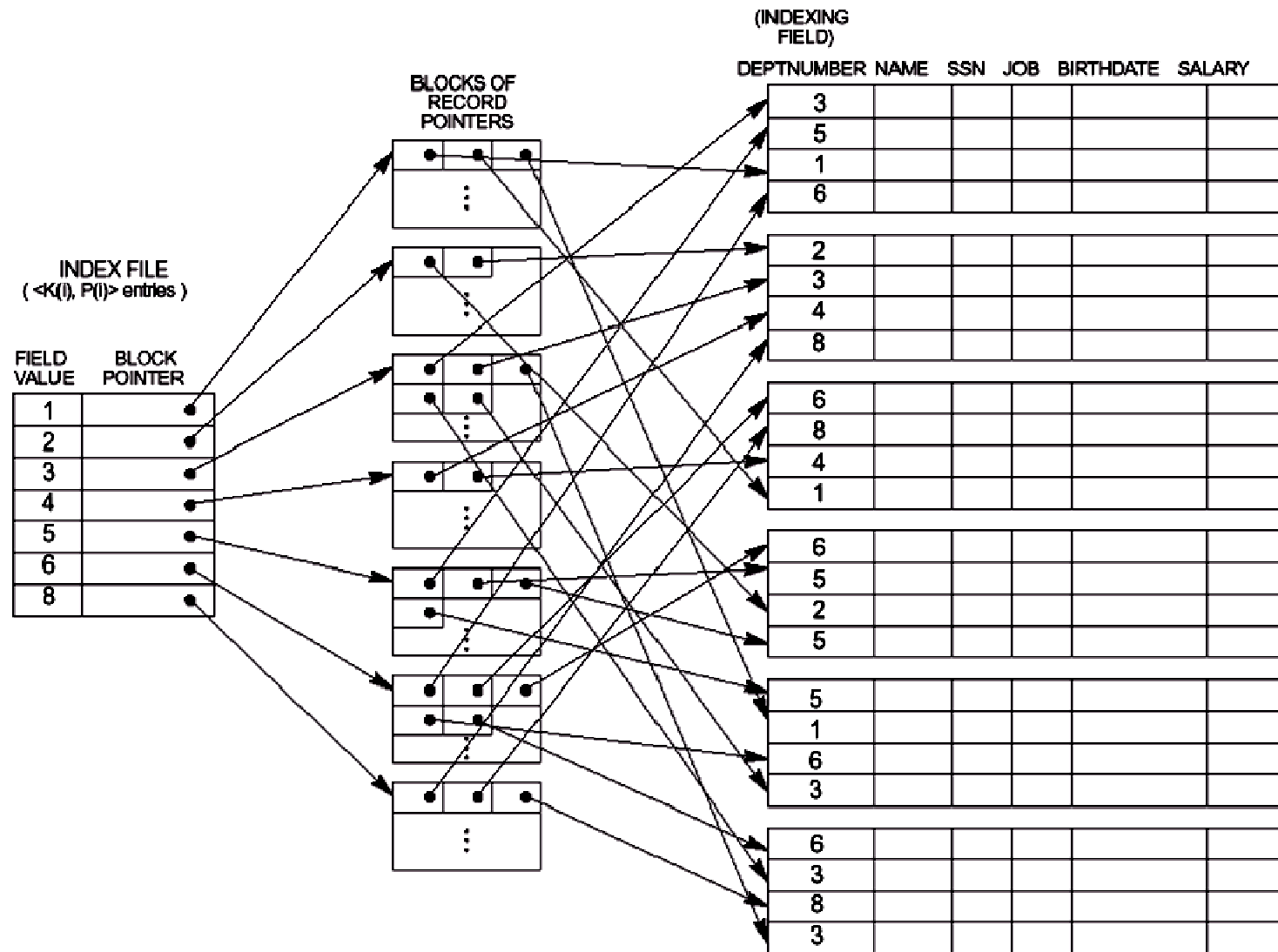


# Ibrahim I Hamar

A dense secondary index (with block pointers) on a nonordering key field of a file



Ibrahim I Hamarash. Database Systems. Ch-VI



A two-level primary index resembling ISAM (Indexed Sequential Access Method) organization.

