# WEEK 3

## ADDING RECORDS TO THE HARD DRIVE USING AN ORDERED ORGANIZATION

CS3319

# STUDENT OBJECTIVES

- Upon completion of this video, you should be able to:
  - Explain how the records are added to the disk when using an ordered or sequential organization
  - Given adding, modifying and deleting records, determine which operations are efficient and which operations are costly
  - Determine when an ordered organization is appropriate
  - Given a number of records, record size and block size, figure out the average number of searches needed to find a record and the worst case scenario for searching for a given record

# ORDERED ORGANIZATION

- **An Ordered File or Sequential File→**

    - before adding any records, pick a field that you want to order the records in the file. If the field guarantees uniqueness, it is called a key field.

    - each time you get a new record, make sure you add the record in the correct location on the disk

# Ordered on disk based on the field *SSN*

**Block 1**

| SSN | FirstName | LastName | BirthDate | Salary | JobCode |
|-----|-----------|----------|-----------|--------|---------|
| 10001 | Homer | Simpson | Aug 12, 1967 | 40,000 | SO |
| 10002 | Lisa | Flanders | Feb 18, 1990 | 86,333 | TE |
| ... | | | | | |
| 10008 | Marge | Griffin | Dec 24, 1959 | 122,000 | BK |

**Block 2**

| 10011 | Bill | Ozzy | Jan 12, 1967 | 49,000 | SS |
|-------|------|------|--------------|--------|-----|
| 10013 | Doug | Reid | Feb 11, 1993 | 186,333 | SO |
| ... | | | | | |
| 10022 | Sylvia | Vancise | March 24, 1979 | 54,000 | BK |

**Block 3**

| 10033 | Harvey | Keitel | May 12, 1977 | 47,000 | TE |
|-------|--------|--------|--------------|--------|-----|
| 10034 | Brian | Griffin | Feb 11, 1993 | 85,333 | BK |
| ... | | | | | |
| 10039 | Sue | Becky | Aug 24, 1989 | 84,000 | BK |

**Block n**

| 20911 | Hanan | Almattar | Jan 12, 1967 | 144,000 | SS |
|-------|-------|----------|--------------|---------|-----|
| 20919 | Ari | Said | Jun 11, 1993 | 286,333 | SO |
| ... | | | | | |
| 20954 | Angus | Carmichael | Jun 24, 1959 | 66,000 | BK |

- **QUESTION: Using an ordered organization , of the 3 operations (insert, modify, delete), which ones would be cheap (fast and easy), which ones would be expensive?**

- Can use a binary search when trying to find a record based on the sorting field

- Offers NO advantage when trying to find records based on the values in the non ordering fields.

# ORDERED FILES CONTINUED…

- Physically order the records on the disk using one of the fields (usually the key field but not always)
- **Advantages:**
  - Searching on the key field can use a binary search
  - No sorting required
- **Disadvantages**
  - Inserts and Delete can be expensive
  - Binary search on a disk can be very expensive
- Rarely used unless there is also a primary index

# EXAMPLE

**QUESTION: Find the** **worst** **case search time to find a record if you are search for a field that the records are sorted on by on the disk:**

- r = 100,000 records stored on a disk with block size B = 2048 bytes.

- Each record (R) is a fixed size of R = 500 bytes.

- Blocking Factor (bfr) = 2048/500 = __4__ records per block (fill in the blank)

- # of blocks needed is __100,000/4__ = __25,000__ blocks

- Binary Search: **Worst Case:** $\log_2 b$ = $\underline{\log_2 25000}$ = $\underline{14.6\ (15)}$ **block accesses**

# AVERAGE ACCESS TIMES:

- The second below is if you are looking for a field that is NOT the sorting field:

**Table 17.2** Average Access Times for a File of $b$ Blocks under Basic File Organizations

| Type of Organization | Access/Search Method | Average Blocks to Access a Specific Record |
|---|---|---|
| Heap (unordered) | Sequential scan (linear search) | $b/2$ |
| Ordered | Sequential scan | $b/2$ |
| Ordered | Binary search | $\log_2 b$ |