

A decorative graphic on the left side of the slide, consisting of a network of white lines and small circles on a teal background, resembling a circuit board or a neural network.

# WEEK 3

ADDING RECORDS TO THE HARD DRIVE USING A HEAP 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 a heap organization
  - Given adding, modifying and deleting records, determine which operations are efficient and which operations are costly
  - Determine when a heap 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

# CONSIDER HOW “EXPENSIVE” EACH FILE ORGANIZATION WOULD BE:

- **A heap (or a pile or unordered file)** → each time you get a new record, just add it to then end.
- Consider inserting the following items (25, 12, 89, 64, 1, 13, 19) into the disk below:

25	12	89	64	1	13	19	
----	----	----	----	---	----	----	--

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

*insert would be cheapest*

*modify*

*deleting is the most expensive: no not leave a gap, you have to move all the follow records one position forward.*

- **Good in the following situations:**

- Data is being bulk loaded into the relation.
- Relation is only a few pages long (searches don't have lots of data to go through)
- When every tuple has to be retrieved every time the relation is used
- When the relation has an additional access structure such as an indexed key

- **Bad in the following situations:**

- Search (linear search, on average  $b/2$  where  $b$  is the number of blocks that must be brought into main memory)
- Deletions, same as searches and it leaves open blocks (wasted storage) or require reorganization occasionally to clean up.
- Sorting

# EXAMPLE

**QUESTION:** Find the average search time to find a record if you use a heap organization for the following scenario:

- $r = 100,000$  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) =  $2048/500 = \underline{4}$  records per block (fill in the blank)
- # of blocks needed is  $\underline{100,000/4} = \underline{25,000}$  blocks
- Linear Search: on average  $b/2 = \underline{25,000}/2 = \underline{12,500}$  block accesses

spreads out

But early on

100