Questions to answer
Memory, disks, buffers
What are the three key components of disk latency?

A. Reading time, buffering time, replacement delay
B. Rotating time, stop time, transfer time
C. Seek time, rotational delay, transfer time
D. Warming up time, search time, transfer time
What is an advantage of magnetic disks over RAM? Check all that apply.

A. Multiple storage surfaces (platters)
B. Larger capacity
C. Faster random access
D. Lower price
E. Persistence of data
What is a dirty buffer page?

A. The page with pin count > 0
B. The page that contains data that changed since it was read from disk
C. The page that contains incorrect data
D. The page that got corrupted in memory
Which of the following classes of buffer pages are written to disk if we need to free some buffer space?

A. Pages with pin count zero, and dirty flag zero (0 0)
B. Pages with pin count greater than zero, and dirty flag one (>=1 1)
C. Pages with pin count zero, and dirty flag one (0 1)
D. Pages with pin count greater than zero, and dirty flag zero (>=1 0)
External-memory sorting
How big a relation (in bytes) that we can sort in 2 passes

M = 5 GB = 5,000,000,000 Bytes = 5 \times 10^9 \text{Bytes}

B = 10 KB

<table>
<thead>
<tr>
<th>Capacity</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 KB</td>
<td>10^3</td>
</tr>
<tr>
<td>1 MB</td>
<td>10^6</td>
</tr>
<tr>
<td>1 GB</td>
<td>10^9</td>
</tr>
<tr>
<td>1 TB</td>
<td>10^{12}</td>
</tr>
<tr>
<td>1 PB</td>
<td>10^{15}</td>
</tr>
</tbody>
</table>

A. 250 TB
B. 2.5 TB
C. 25 TB
D. 2.5 PB

\(B\) - block size in bytes.
\(M\) - main memory in bytes.
What is the state after the next step of 2PMMS?

- **Phase 2.**
- **On disk:**
  - Sub-list 1: 20 25 27 29 30
  - Sub-list 2: 18 23 35 45 65
  - Sub-list 3: 21 22 29 34 39
- **Main Memory (4 buffers)**
  - Input Buffer1: 20 25
  - Input Buffer2: 18 23
  - Input Buffer3: 21 22
  - Output Buffer:
- **Sorted list:**

A. Output buffer: 20, sorted list: empty

B. Output buffer: 18, sorted list: empty

C. Output buffer: 20, sorted list: 18
What software and hardware methods we can use to improve performance of multi-way sort, given that the available memory is constrained and cannot be increased? Check all that applies

- Double buffering
- Cylindrification
- Replacing hard disks with tapes
- Multiple disks
- Stronger CPU
Disk files
The record header may contain a directory of field offsets. What problems does it solve? Check all that apply.

- Minimizing an overall space occupied by the record
- Efficient access to the beginning of the field data
- Defining the order of fields within each record
- Efficient representation of nulls
B-trees
An index on a search key $K$ can be created even if the data file is not sorted by $K$. Such an index can be dense or sparse.

- False – only dense
In order to maintain the pre-defined capacity range, internal nodes of B*-tree must be joined or split. The insertion may cause the splitting of internal nodes.

• True
The deletion of a key from a B*-tree may result in the following tree modifications (check all that apply):

- The structure of the tree remains unchanged
- A key from one sibling is transferred to another sibling
- A parent and a child merge into a single node
- Two siblings merge into a single node