COMP-352 Tutorial #10

INDEXING

- So far we have seen multiple ways to organize our data: Arrays, Binary Search Trees, Heaps,…etc.

- Some have a concept of position [index] and others are positioned relativity [linked].

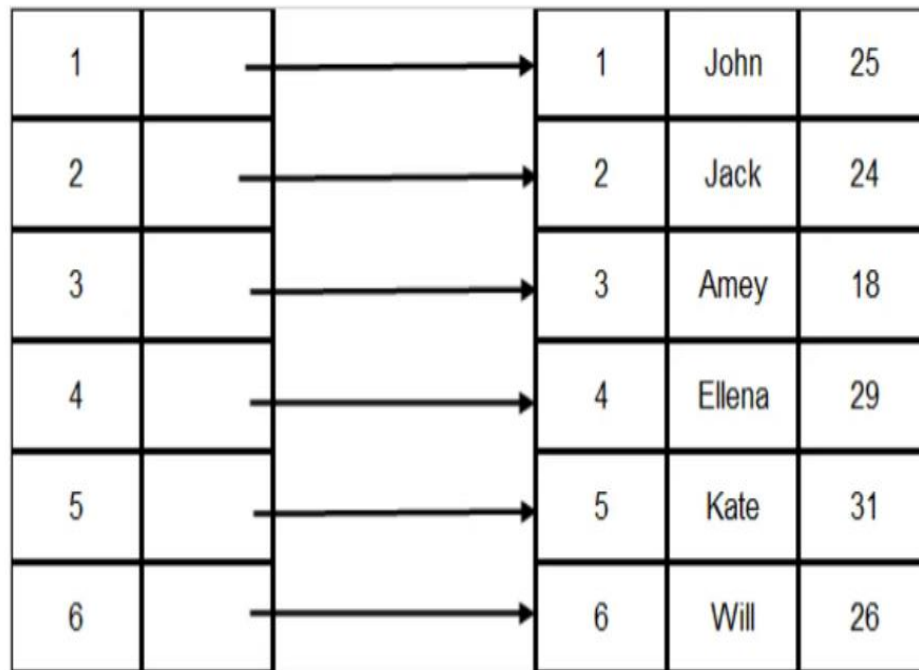**Indexing:** A data structure technique used to efficiently access data based on an key.

- This **key** is usually an index which is a data structure used to locate data.

- **E.G.,** an index could be based on an **ID** column in a database table.

- This technique/data structure offers us a constant access time **O(1)**

- Used in optimizing performance of databases.

Types of Indexing:

- **Primary Index:** Ordered column with two fields: index and a pointer to the data.

- **Secondary Index:** A two-level mapping technique to reduce the mapping size of 1st-level

- **Clustering Index:** An indexing technique used where multiple instance share a key
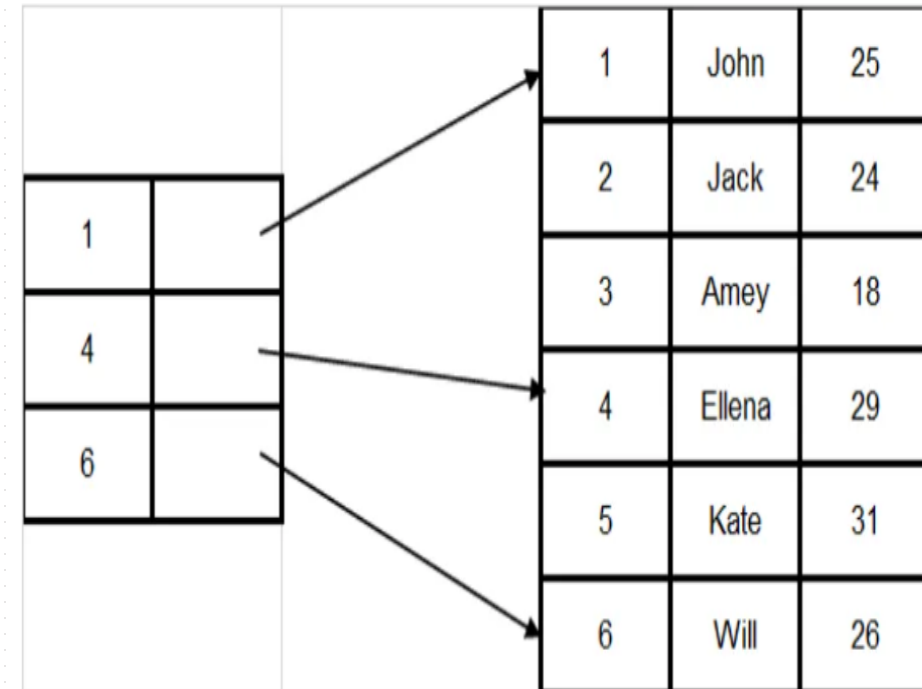
Primary Index

Dense Index
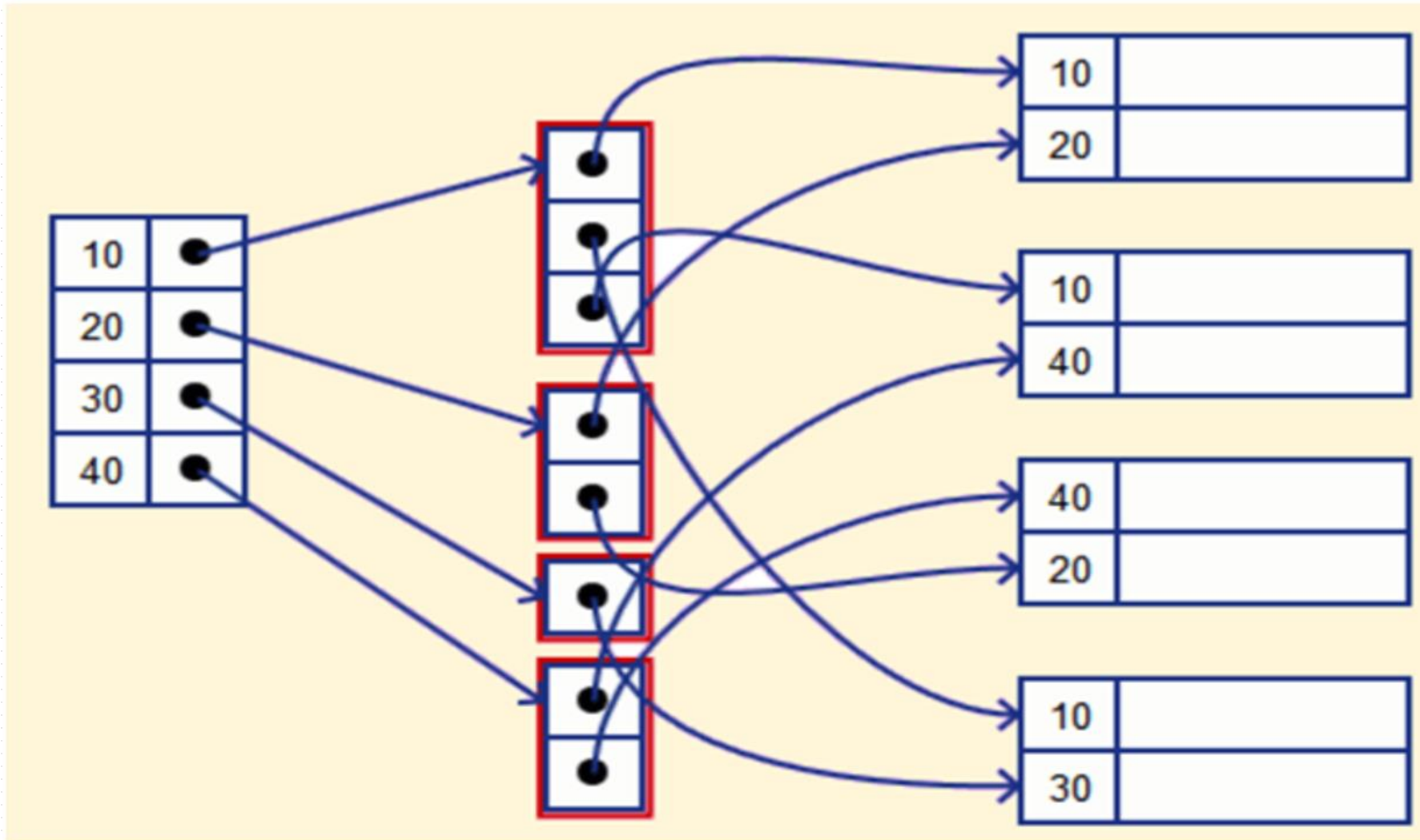

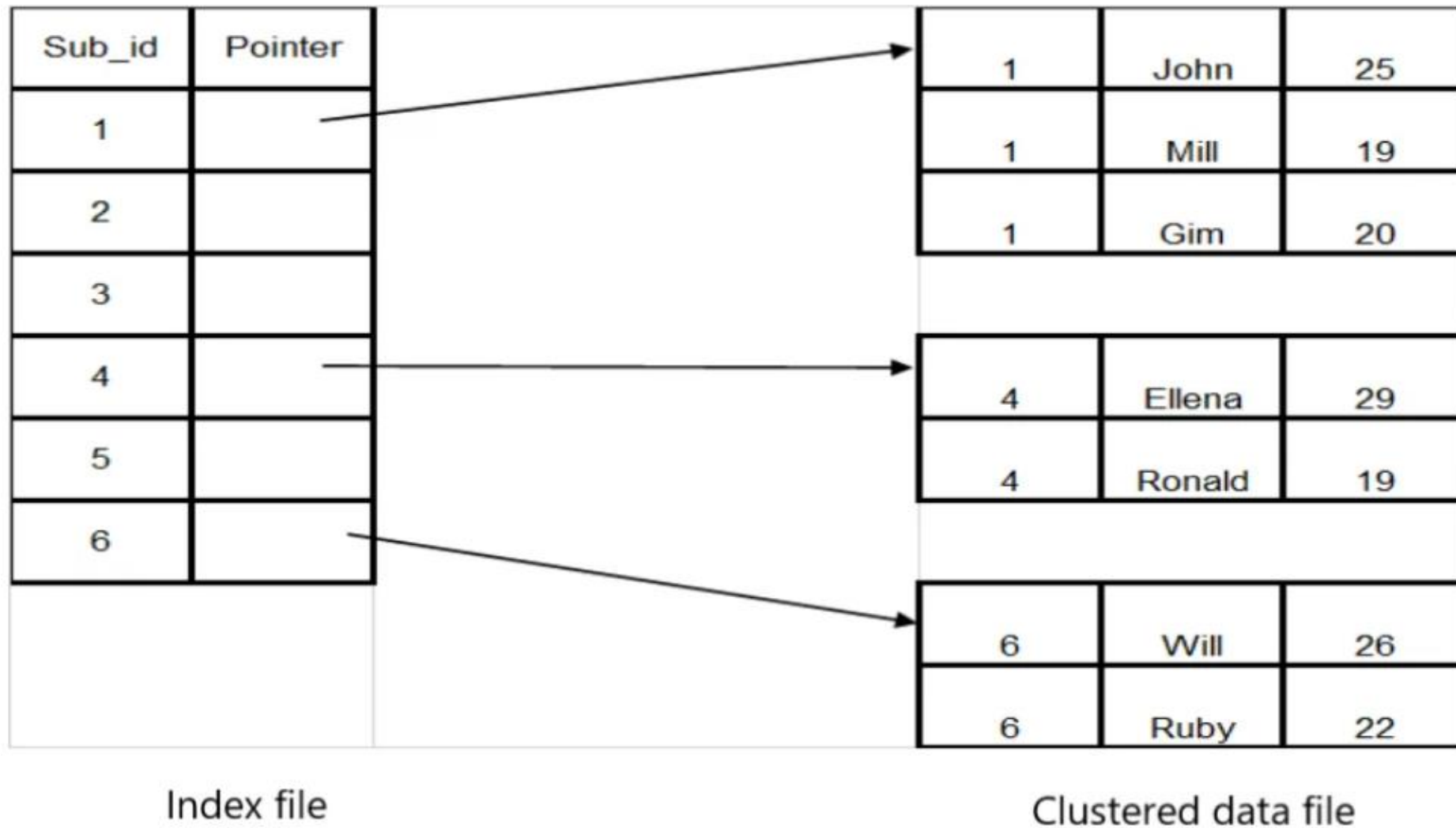Index record   Data block

Sparse Index

Index record   Data block

Secondary Index

Clustered Index

Index file

Clustered data file

5

Hashing

Hashing: A technique used to search the location of data instance without using an index.

Hash Function: A key-value mapping function that maps a hashed key to value.

• Based on a attribute of given data, we can generate special keys

• This can introduce duplicate keys, how to fix this issue?

> • **Probing:** A chain-hash function used to find the next available position
> • **Sequential Linked-list:** At a position there could exist a linked list for the duplicates.

Types of Probing: Linear Probing, Stepped Probing, Pseudo-Random Probing, and Quadratic Probing

Data Structures: Hash Table, Hash Maps, Dictionaries…etc.

EXERCISE: Convert Roman Numerals to Integers.

Symbol Table

I	V	X	L
1	5	10	50

Examples:

- "III" -> 3
- "LVIII" -> 58

Instructions:

Given a Roman numeral, convert it to an integer. The numeral will be valid and within the range [1, 89].

Constraints:

- 1 <= s.length <= 8
- s contains only ('I', 'V', 'X', 'L')
- Input Roman numerals are guaranteed to be valid in the range [1, 89].

THANK YOU