Indexes

What is Index?

Book has been arranged via categories, subjects.

Same categories will be stored in the same area.

Book is data

Catalogue is the index

What is Index?

Phone number has been arranged in Alphabets, groups (work, friends, …)

Phone number is the data

The Alphabets and groups are the index.

What is Index?

- Auxiliary data
- Properly organised (data structure)
- To facilitate data search

Indexes

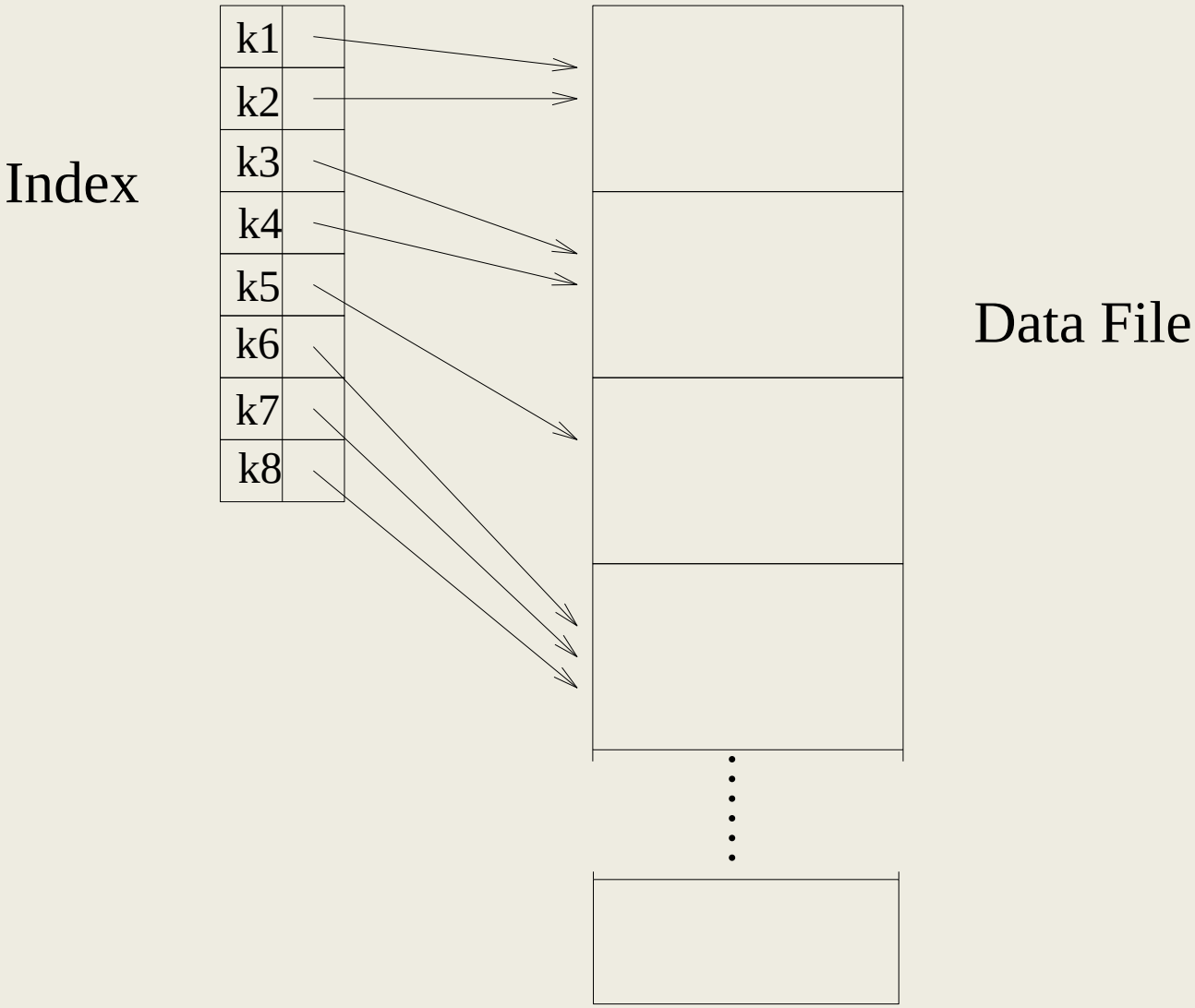
Word indexes in a book:

INDEXES

aardvark 25,36	lion 18
bat 12	llama 17,21,22
cat 1,5,12	sloth 18
dog 3	tiger 18
elephant . . 17	wombat . . . 27
emu 28	zebra 19

- A table of key values, where each entry gives places where key is used.
- Aim: efficient access to records via key.

Indexing Structure



Indexing Structure

- Index is collection of data entries k∗.
  - Each data entry k∗ contains enough information to retrieve (one or more) records with search key value k.

- Indexing:
  - How are data entries organized in order to support efficient retrieval of data entries with a given search key value?
  - Exactly what is stored as a data entry?

Alternatives for Data Entries in an Index

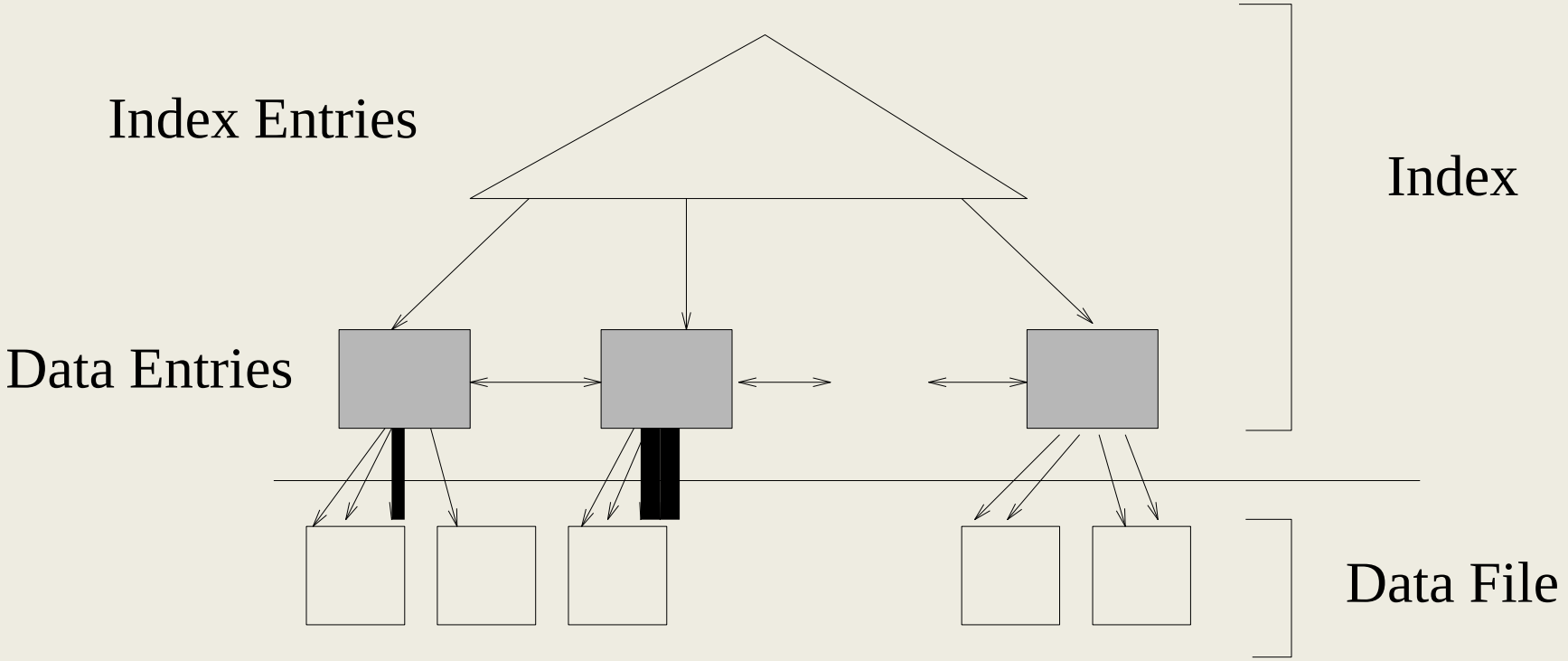
- A data entry k∗ is an actual data record (with search key value k).

- A data entry is (k, rid) pair (rid is the record id of a data record with search key value k).
  - E.g. Example: (Adams, page 12), (Adams, page 100)

- A data entry is a (k, rid − list) pair (rid − list is the list of record ids of data records with search key value k).
  - E.g. (Adams, page 12, page 100)

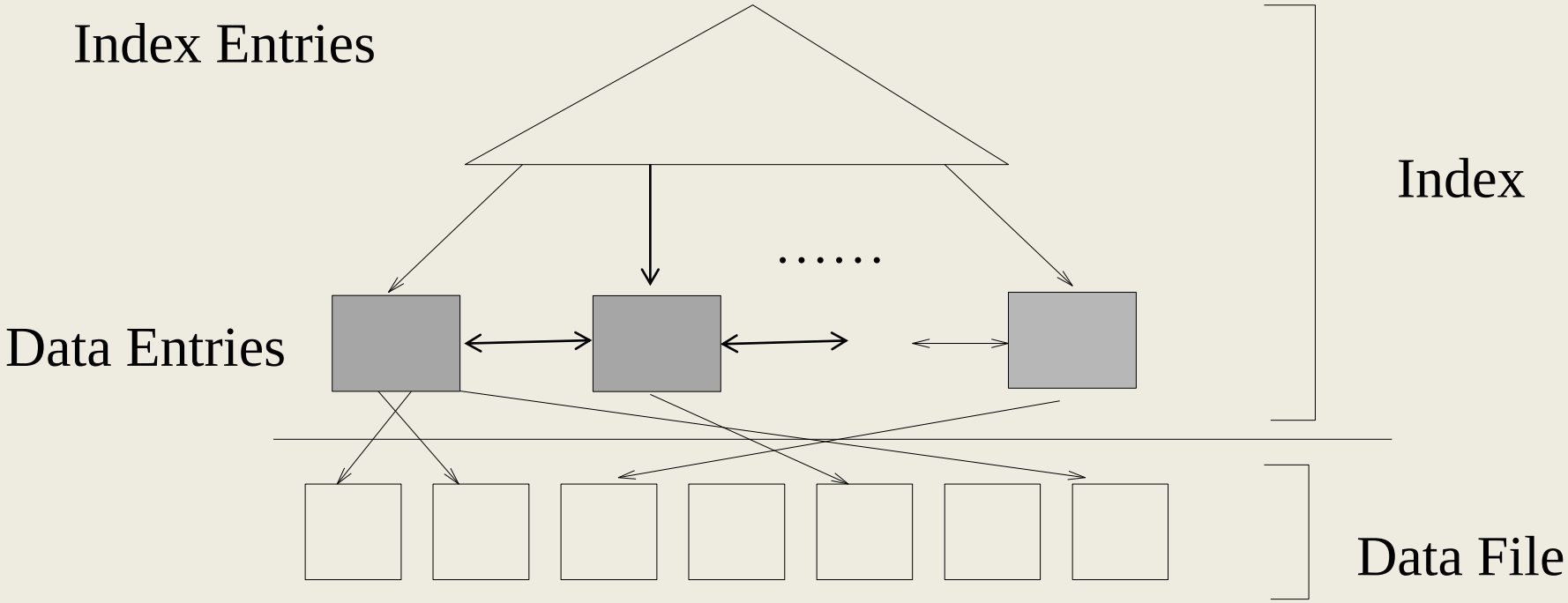
Clustered Index

- Clustered: a file is organized of data records is the same as or close to the ordering of data entries in some index.
- Typically, the search key of file is the same as the search key of index.



Unclustered Index

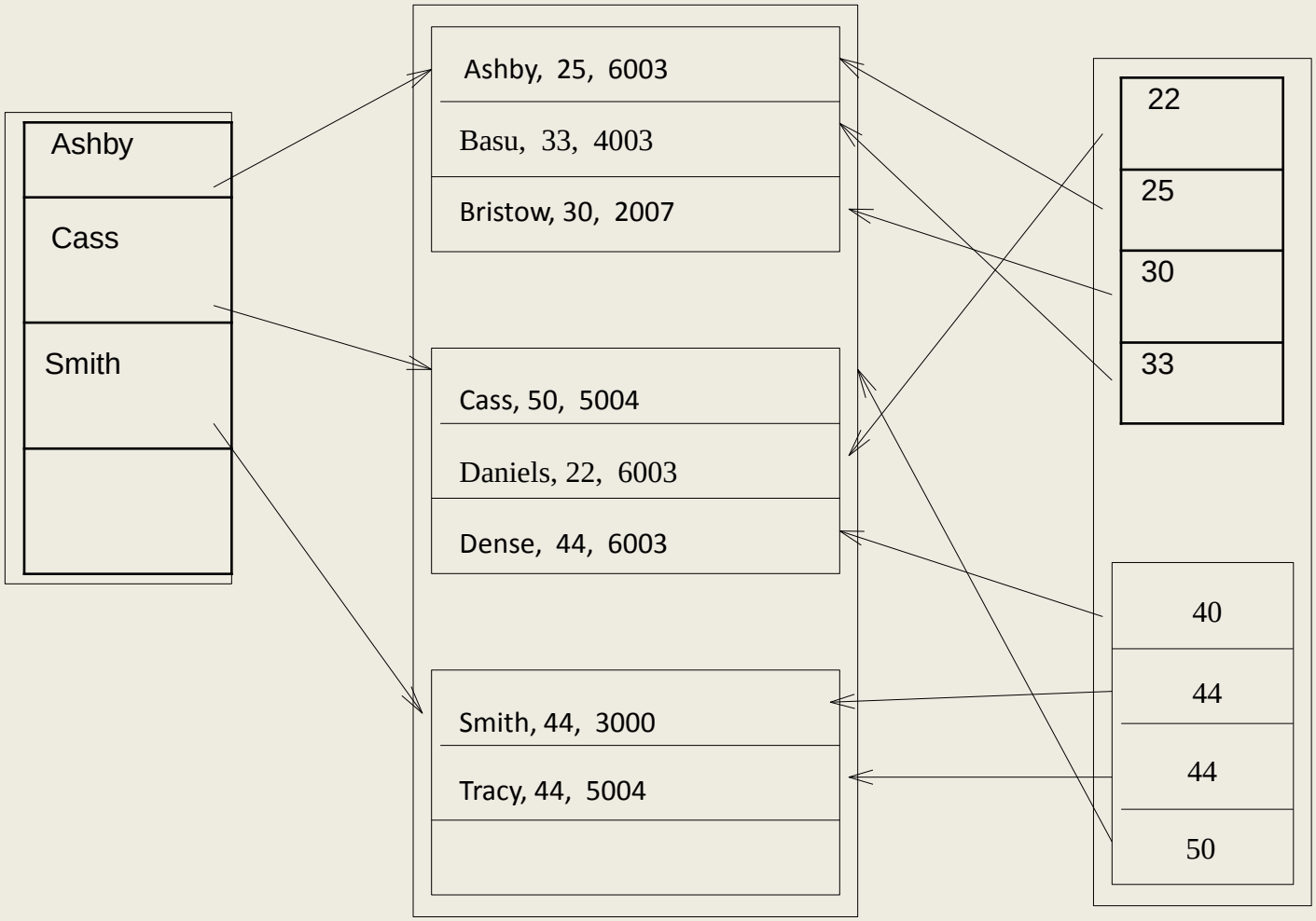
- Clustered indexes are relatively expensive to maintain.
- A data file can be clustered on at most one search key.



Dense VS Sparse Indexes

- Dense Index and Sparse Index
  - Dense Index contains (**at least**) one data entry for every search key value.
  - Sparse Index may note and one search key can points to a set of data entries

Q: Can we build a sparse index that is not clustered?



Sparse Index VS Dense Index

Primary and Secondary Indexes

- Primary: Indexing fields include **primary key**.
  - The file is physically ordered according to the primary key.
  - There can be at most one primary index for a table

- Secondary: otherwise.
  - Index specified on attribute fields which are non-ordering fields of the file (i.e., file is not ordered based on the field)
  - A file can have several secondary indexes

Learning Outcomes

- General understanding of the use of indexes

- The difference between clustered / unclustered indexes, dense / sparse indexes, primary / secondary indexes