Data Storage and Indexing:

Alternative File Organizations

Ira A. Fulton Schools of Engineering
Arizona State University

Objectives

Objective

Recognize major data storage layouts

Alternative File Organizations

Heap Files

- Random order
- Suitable when typical access is a file scan retrieving all records.

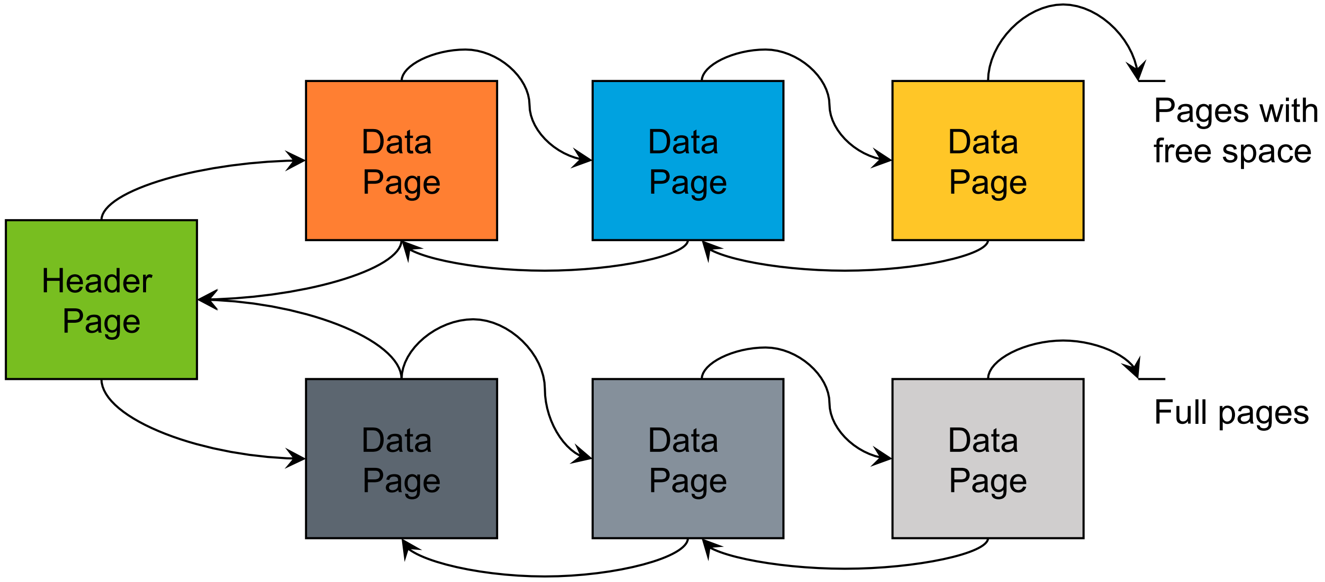
Sorted Files

- Best if records must be retrieved in some order, or only a `range’ of records is needed.

Indexes

- Data structures to organize records via trees or hashing.

Basic File Organization: Heap Files



The Cost Model

Cost Measure

Number of page accesses

Reasoning

Page access cost is usually the dominant cost of database operations

An accurate model is too complex for analyzing algorithms

Heap File Advantages/ Disadvantages

Advantages

Efficient

- for bulk loading data.
- for relatively small relations as indexing overheads are avoided.
- when queries that need to fetch large proportion of stored records.

Disadvantages

Not Efficient

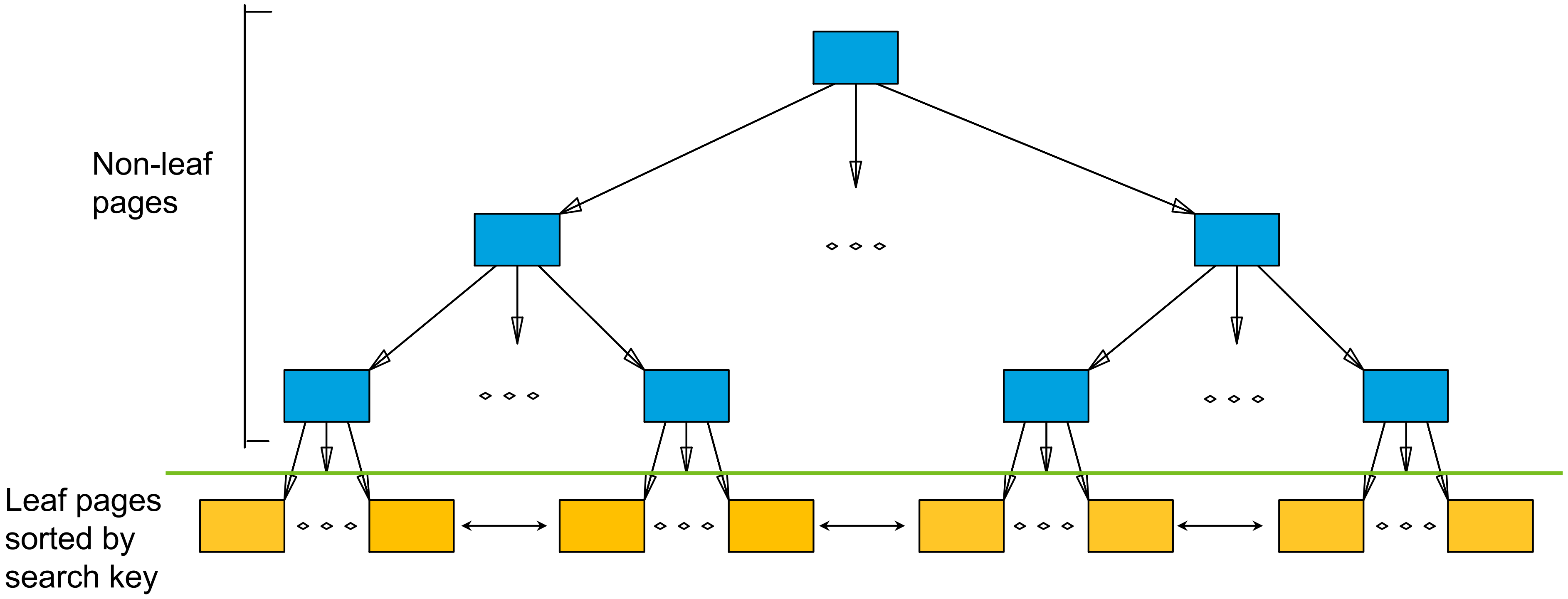
- for selective queries.
- for sorting, may be time-consuming.

Indexes

File Index

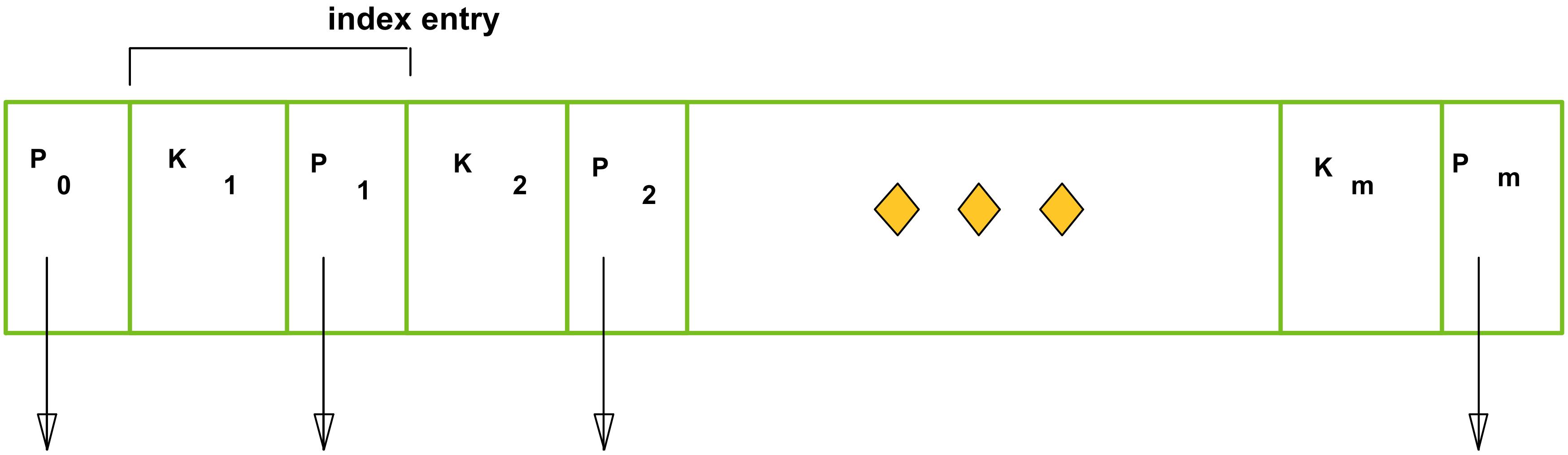
- Speeds up selections on the search key fields
- Any subset of the fields of a relation can be the search key for an index on the relation.
- An index contains a collection of data entries, and supports efficient retrieval of all data entries k* with a given key value k.

B+ Tree Indexes

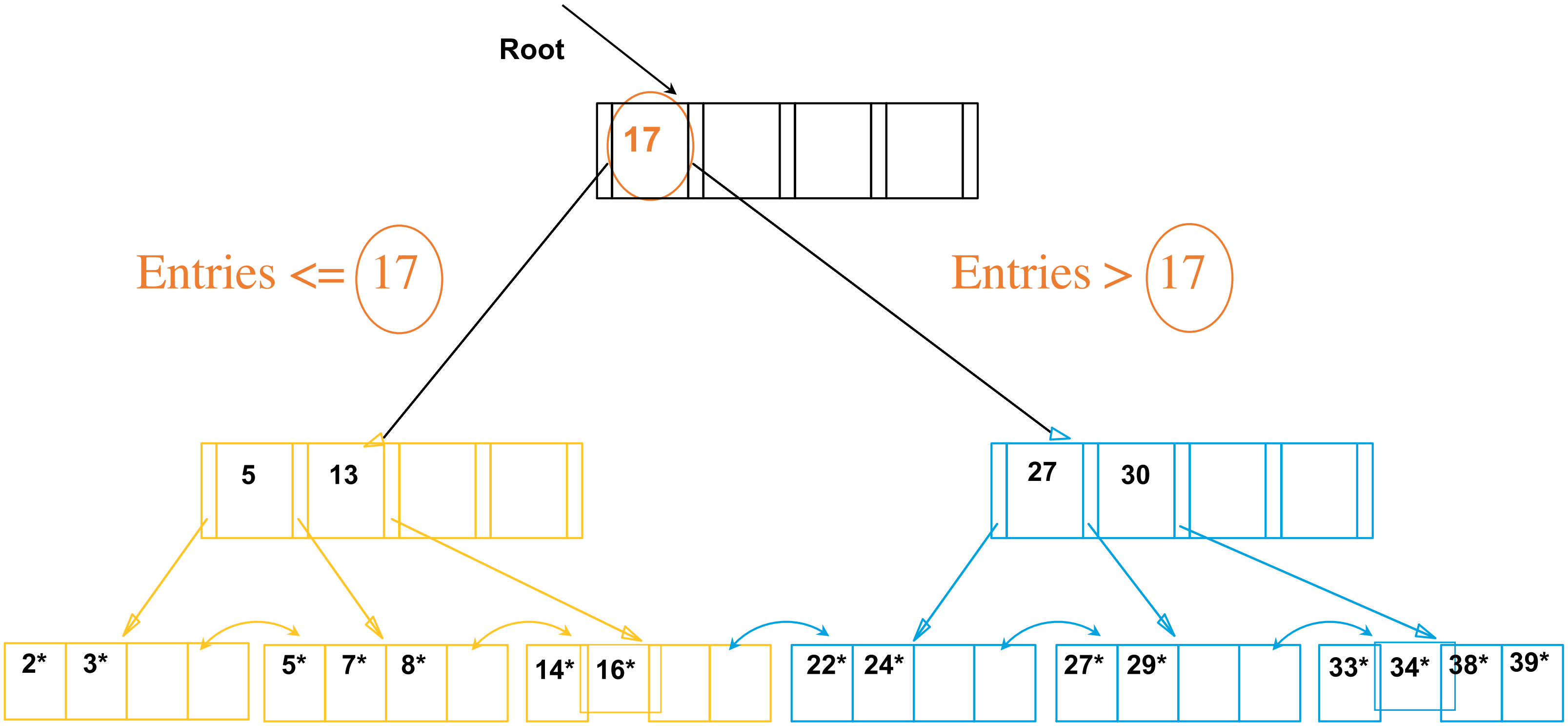


B+ Tree Indexes

Non-leaf pages have *index entries;* only used to direct searches:



Example B+ Tree


Root
17
Entries <= 17
Entries > 17
5
13
27
30
2*
3*
5*
7*
8*
14*
16*
22*
24*
27*
29*
33*
34*
38*
39*