Course code   : **CSE2007**
Course title   : **Database Management System**
Module   : **6**
Topic   : **3**

Hashing Techniques

Objectives

This session will give the knowledge about

- Introduction to Hashing Techniques

- Internal Hashing

- External Hashing

Hashing Techniques

Hash function (randomizing function)

- Applied to hash field value of a record

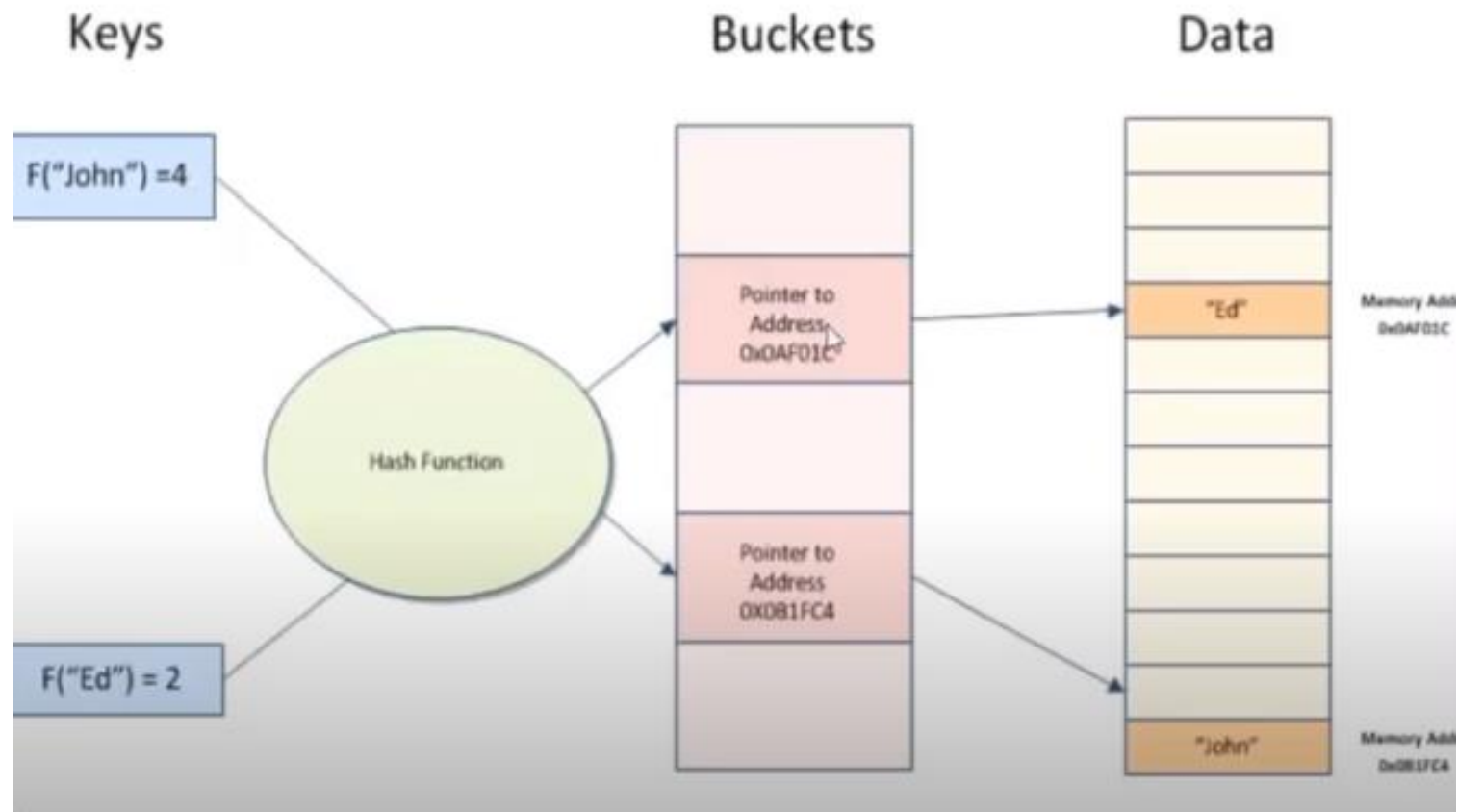- Yields address of the disk block of stored record

Organization called hash file

- Search condition is equality condition on the hash field

- Hash field typically key field

Hashing also internal search structure

- Used when group of records accessed exclusively by one field value

Hashing Techniques

Hashing Techniques

Hash file organization of account file, using branch_name as key

There are 10 buckets,

The binary representation of the ith character is assumed to be the integer i.

The hash function returns the sum of the binary representations of the characters modulo 10

E.g. h(Perryridge) = 5   h(Round Hill) = 3
h(Brighton) = 3

bucket 0		

bucket 1		

bucket 2		

bucket 3		
A-217	Brighton	750
A-305	Round Hill	350

bucket 4		
A-222	Redwood	700

bucket 5		
A-102	Perryridge	400
A-201	Perryridge	900
A-218	Perryridge	700

bucket 6		

bucket 7		
A-215	Mianus	700

bucket 8		
A-101	Downtown	500
A-110	Downtown	600

bucket 9		

Hashing Techniques

Internal hashing

- Implemented as a hash table through the use of an array of records.

- Hash table

	Name	Ssn	Job	Salary
0				
1				
2				
3				
⋮				
M − 2				
M − 1				

Internal hashing data structures. (a) Array of M positions for use in internal hashing.

Hashing Techniques

Collision

- Hash field value for inserted record hashes to address already containing a different record

Collision resolution

- Open addressing

  Open addressing. Proceeding from the occupied position specified by the hash address, the program checks the subsequent positions in order until an unused (empty) position is found
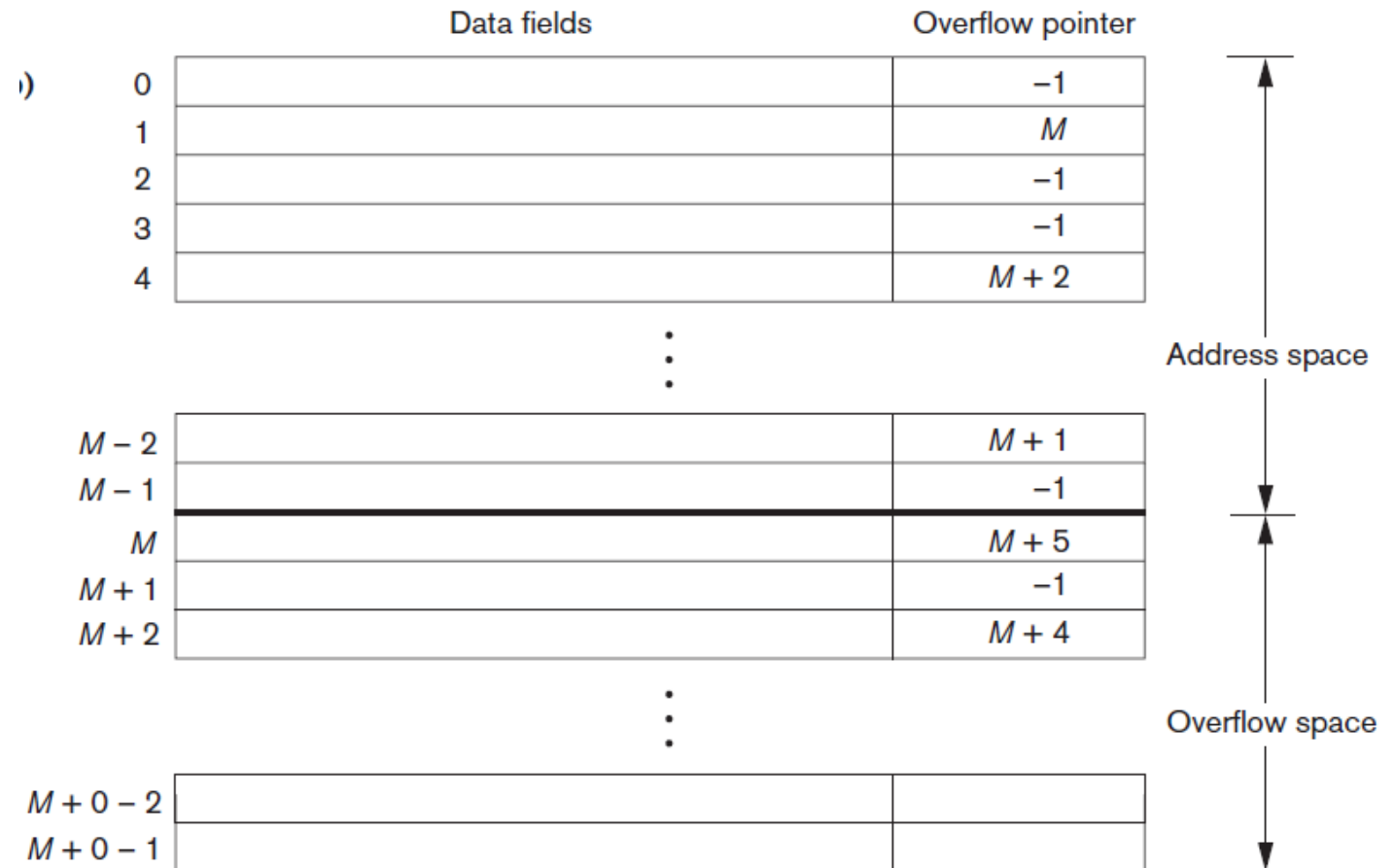
Hashing Techniques

Chaining

- Various overflow locations are kept.

- By extending the array with a number of overflow positions.

- Additionally, a pointer field is added to each record location.

- A collision is resolved by placing the new record in an unused overflow location and setting the pointer of the occupied hash address location to the address of that overflow location.

Hashing Techniques

Collision resolution by

Chaining

- null pointer = −1
- overflow pointer refers to position of next record in linked list

Hashing Techniques

Multiple hashing

- The program applies a second hash function if the first results in a collision.

- If another collision results, the program uses open addressing or applies a third hash function and then uses open addressing if necessary.

- Note that the series of hash functions are used in the same orderfor retrieval.

Hashing Techniques

External hashing for disk files

- Target address space made of buckets

- Bucket: one disk block or contiguous blocks
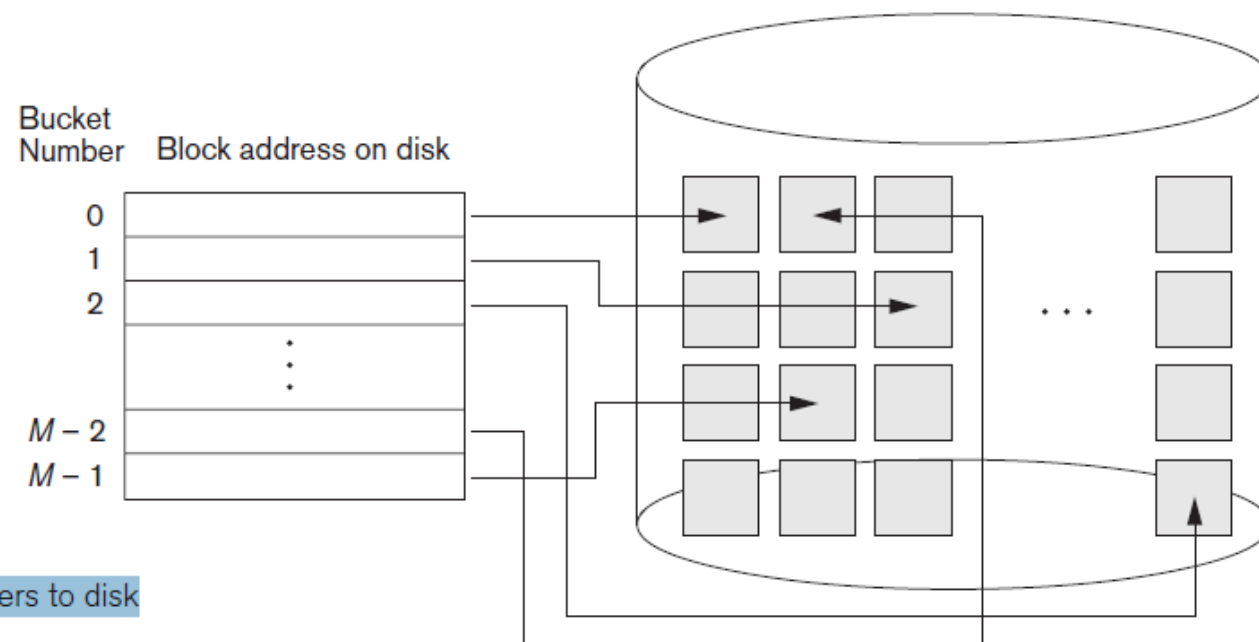
Figure 16.9
Matching bucket numbers to disk block addresses.

Hashing Techniques

Hashing function maps a key into relative bucket

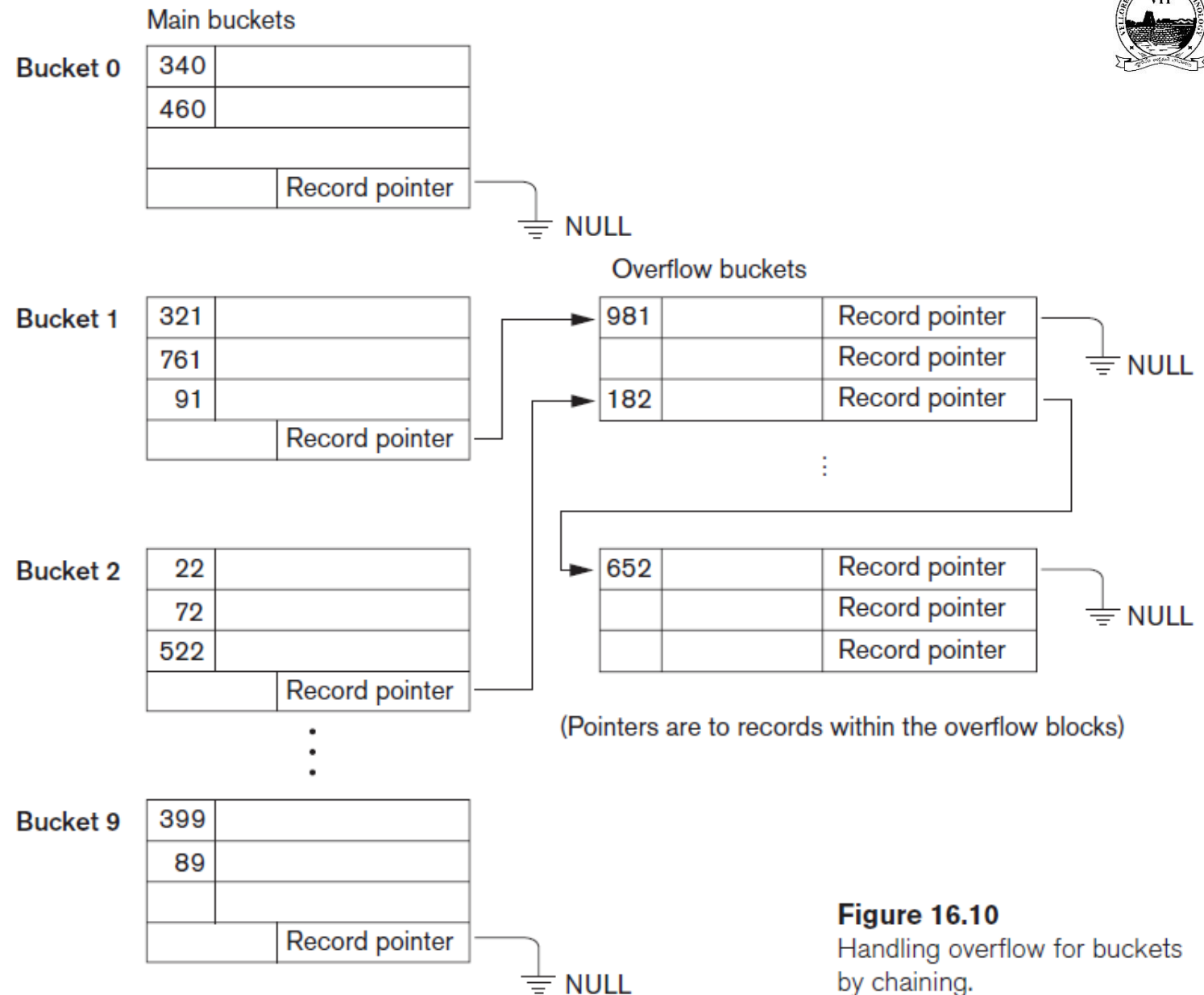- Table in file header converts bucket number to disk block address

Collision problem less severe with buckets

Static hashing

- Fixed number of buckets allocated

Hashing

Handling overflow for buckets by chaining.

Figure 16.10
Handling overflow for buckets by chaining.

Hashing Techniques

Hashing techniques that allow dynamic file expansion

- Extendible hashing
  - File performance does not degrade as file grows

- Dynamic hashing
  - Maintains tree-structured directory

- Linear hashing
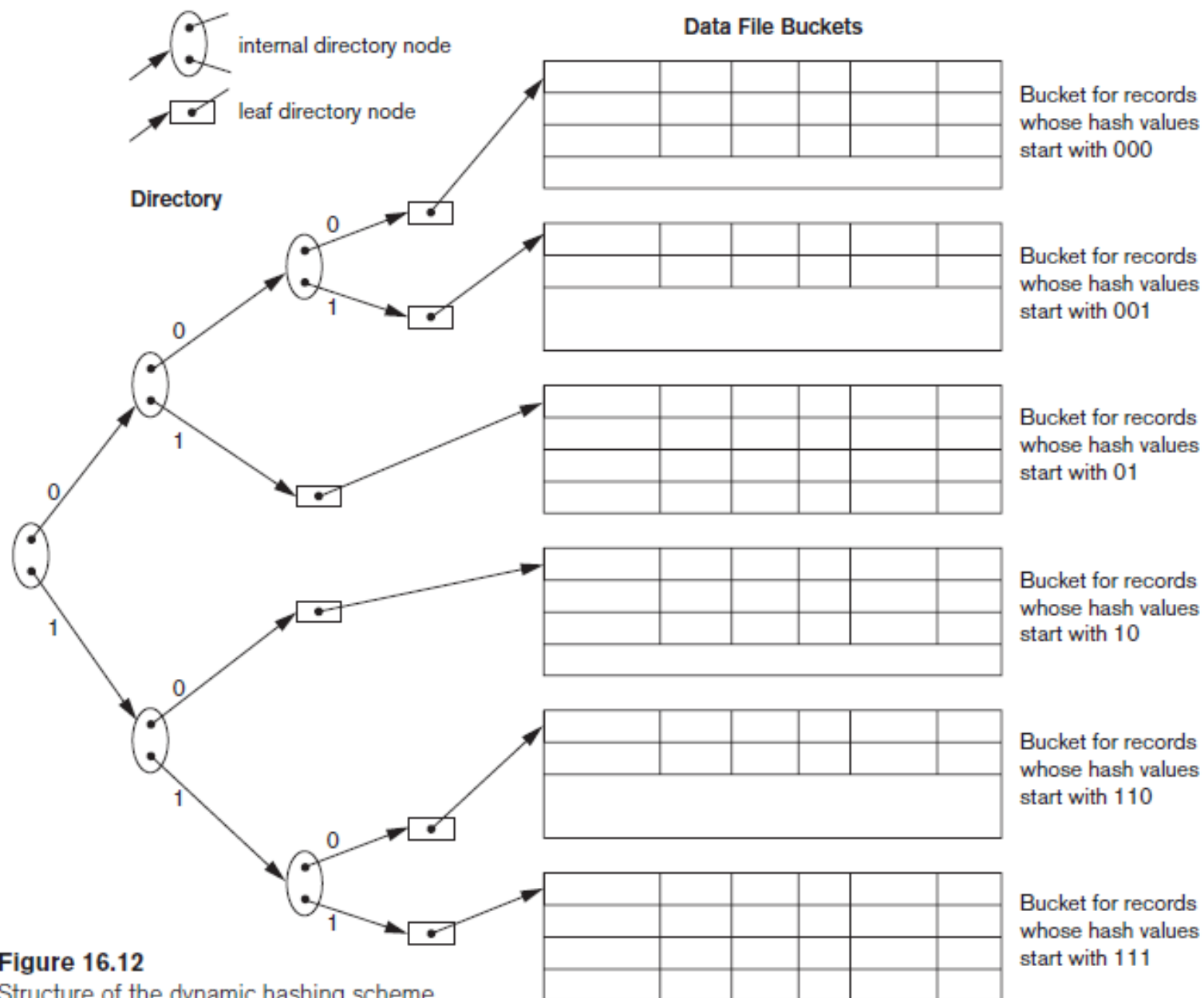 - Allows hash file to expand and shrink buckets without needing a directory

Hashing

Extendible hashing
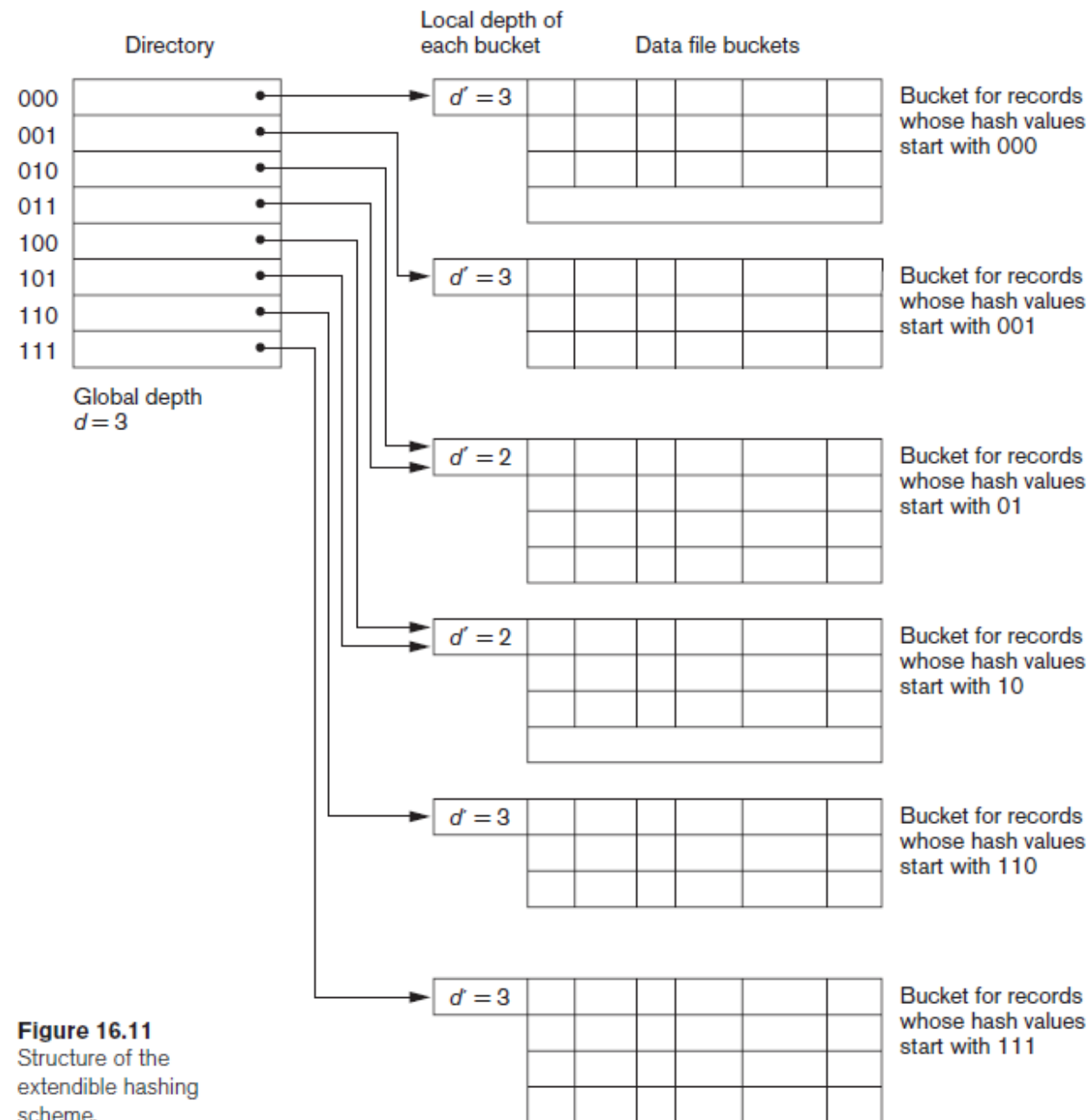
Figure 16.11
Structure of the extendible hashing scheme.

Hashing

Dynamic hashing

Figure 16.12
Structure of the dynamic hashing scheme.

Other Primary File Organizations

Files of mixed records

- Relationships implemented by logical field references
- Physical clustering

B-tree data structure

Column-based data storage

Summary

This session will give the knowledge about

- Introduction to Hashing Techniques

- Internal Hashing

- External Hashing