Database Management System – 45 (Hashing Techniques – Internal Hashing)

Ajay James
Asst. Prof in CSE
Government Engineering College Thrissur

AJ-GEC THRISSUR

Outline

- Hashing
- Internal hashing
- Hashing functions
- Collision Resolution

Hashing

- File organization technique - Hash file
- Provides very fast access to records
- Search condition must be an equality condition on a single field, called the ***hash field***
- Hash field is a key field of the file - ***hash key***
- Idea behind hashing is to provide a function ***h***, called a hash function or randomizing function, which is applied to the hash field value of a record and yields the address of the disk block in which the record is stored
- For most records, we need only a single-block access to retrieve that record.

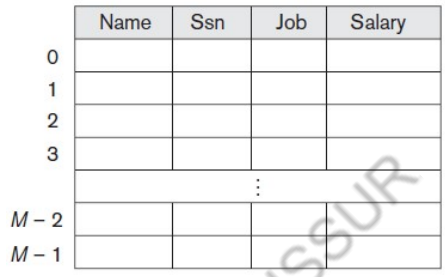
AJCEC THRISSUR

Internal Hashing

- Internal search structure within a program whenever a group of records is accessed exclusively by using the value of one field
- Implemented as a **hash table** through the use of an array of records

Internal Hashing

- Array index range is from 0 to M – 1
- Choose a hash function that transforms the hash field value into an integer between 0 and M − 1
- Common hash function is the h(K) = K **mod** M



AJ-CEC THRISSUR

Other hashing functions

- Folding
  - Applying an arithmetic function such as addition or a logical function such as exclusive or to different portions of the hash field
  - Example - with an address space from 0 to 999 to store 1,000 keys, a 6-digit key 235469 may be folded and stored at the address: (235+964) mod 1000 = 199)
- Picking some digits of the hash field value
  - Example - third, fifth, and eighth digits
  - Social Security number 301-67-8923 a hash value of 172

Collision

- Hashing functions do not guarantee that distinct values will hash to distinct addresses
- Social Security number 30**1**-6**7**-89**2**3 a hash value of **172**
- Social Security number 22**1**-8**7**-11**2**5 a hash value of **172**
- Collision
- Hash field value for inserted record hashes to address already containing a different record

AJCEC THRISSUR

Collision resolution

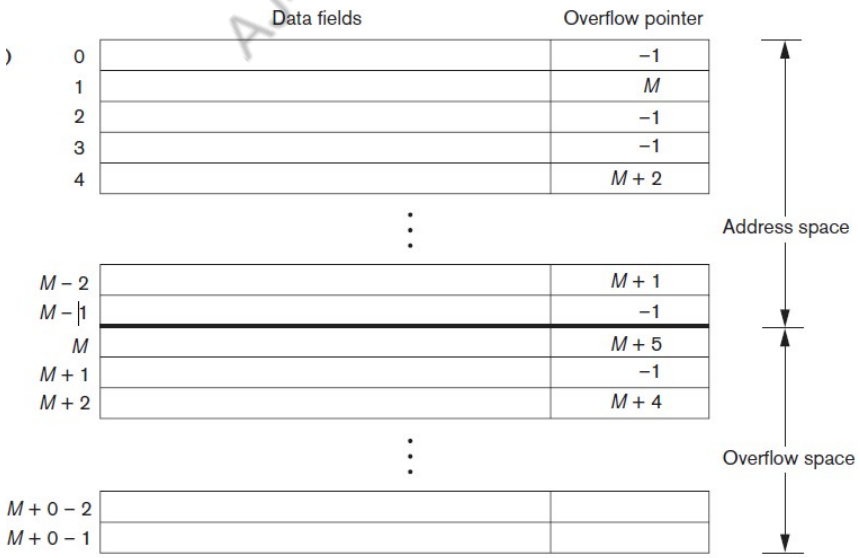
- Process of finding another position
  - Open addressing
  - Chaining
  - Multiple hashing
- Open addressing
  - Proceeding from the occupied position specified by the hash address, the program checks the subsequent positions in order until an unused (empty) position is found

Collision resolution

- Chaining
  - Various overflow locations are kept
  - A pointer field is added to each record location
  - Collision is resolved by placing the new record in an unused overflow location
  - Set the pointer of the occupied hash address location to the address of that overflow location

A-SEC THRISSUR

Collision resolution - Chaining



Collision resolution

- Multiple hashing
  - 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
  - Series of hash functions are used in the same order for retrieval

AJ-GEC THRISSUR

Reference

- Elmasri R. and S. Navathe, Database Systems: Models, Languages, Design and Application Programming, Pearson Education 6th edition and 7th edition

Thank you

AJ-GEC THRISSUR