Quiz 2, CS344, Fall' 16, IITG

Hash Based Join

Input: relation1 file, relation 2 file, record size, page size, number of available pages, maximum number of hashing rounds

Example: relation1.txt, relation2.txt, 180, 130, 400, 5, 3

Number of records that fit per page are floor of page size/ record size.

relation1.txt

20

490

70

21

32

relation2.txt

20

70

21

199

7712

3201

42

3434

6789

1

Output:

Size of relation 1: 3 pages

Size of relation 2: 4 pages

Total number of available pages: 5

Number of buckets in hash table: 4

Hashing Round 1:

Reading relation1:

Tuple 1: 20 Mapped to bucket: 1

Tuple 2: 490 Mapped to bucket: 1

Page for bucket 1 full. Flushed to secondary storage.

Tuple3: 70 Mapped to bucket: 1

Tuple 4: 21 Mapped to bucket: 2

Tuple 5: 32 Mapped to bucket: 4

Done with relation1.

Created following files.

rel1.round1.bucket1: 2 pages

rel1.round1.bucket2: 1 page

rel1.round1.bucket4: 1 page

//create similar output for relation 2.

rel2.round1.bucket1: 3 pages

rel2.round1.bucket2: 1 page

rel2.round1.bucket3: 1 page

Bucket 1: Total size is 5 pages. Cannot perform in memory join.

Performing second round of hashing for round1.bucket1

Size of relation 1: 2 pages

Size of relation 2: 3 pages

Total number of available pages: 5

Number of buckets in hash table: 4

Reading relation1.round1.bucket1

Tuple 1: 20 Mapped to bucket: 1

Tuple 2: 490 Mapped to bucket: 2

Tuple 3: 70 Mapped to bucket: 3

.

.

.

.

.

.

.

Bucket 2: Total size is 2 pages

Total available pages 5.

Performing in memory join.

Matching pairs are

.

.

.

.

Bucket 3: No matching tuple from relation1. No further processing required

Bucket 4: No matching tuple from relation2. No further processing required