Database Systems

Instructors: Hao-Hua Chu
Winston Hsu
Fall Semester, 2007

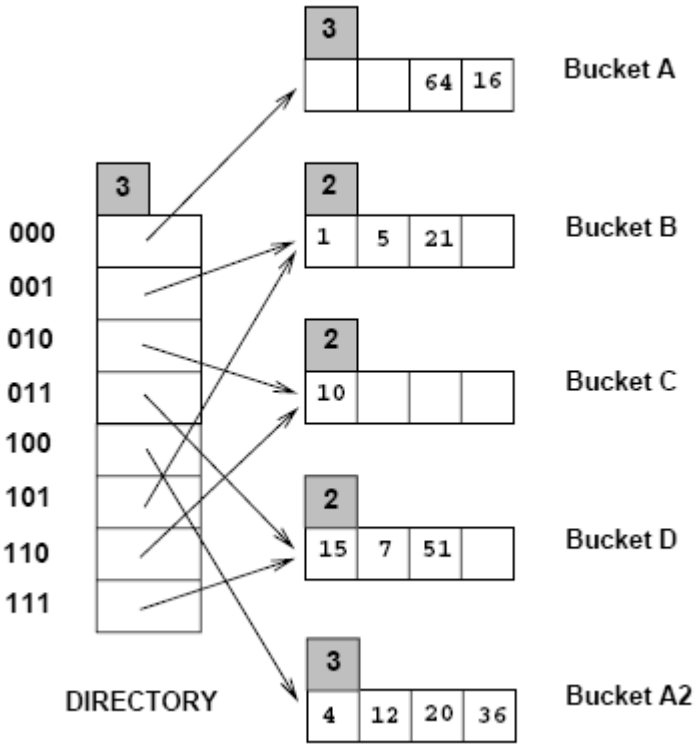
Assignment 5

Deadline: At the end of the class, Dec 17 (Mon), 2007
This is an individual assignment, that is, no group submissions are allowed.
Cheating Policy: If you are caught cheating, your grade is 0.
Late Policy: We will not accept any assignment submissions.

Questions

1. Consider the Extendible Hashing index shown in Figure 1.1. Answer the following questions about this index:

1. What can you say about the last entry that was inserted into the index?
2. What can you say about the last entry that was inserted into the index if you know that there have been no deletions from this index so far?
3. Suppose you are told that there have been no deletions from this index so far. What can you say about the last entry whose insertion into the index caused a split?
4. Show the index after inserting an entry with hash value 68.
5. Show the index after inserting entries with hash values 17 and 69.
6. Show the index after deleting the entry with hash value 21.
   (Assume that the full deletion algorithm is used.)
7. Show the index after deleting the entry with hash value 10.
   Is a merge triggered by this deletion? If not, explain why. (Assume that the full deletion algorithm is used.)



gure 1.1 Figure for Exercise 1

2. Consider the Linear Hashing index shown in Figure 1.2. Assume that we split whenever an overflow page is created. Answer the following questions about this index:

1. What can you say about the last entry that was inserted into the index?
2. What can you say about the last entry that was inserted into the index if you know that there have been no deletions from this index so far?
3. Suppose you know that there have been no deletions from this index so far. What can you say about the last entry whose insertion into the index caused a split?
4. Show the index after inserting an entry with hash value 4.
5. Show the index after inserting an entry with hash value 15.
6. Show the index after deleting the entries with hash values 36 and 44.

. (Assume that the full deletion algorithm is used.)

7. Find a list of entries whose insertion into the original index would lead to a bucket with two overflow pages. Use as few entries as possible to accomplish this. What is the maximum number of entries that can be inserted into this bucket before a split occurs that reduces the length of this overflow chain?

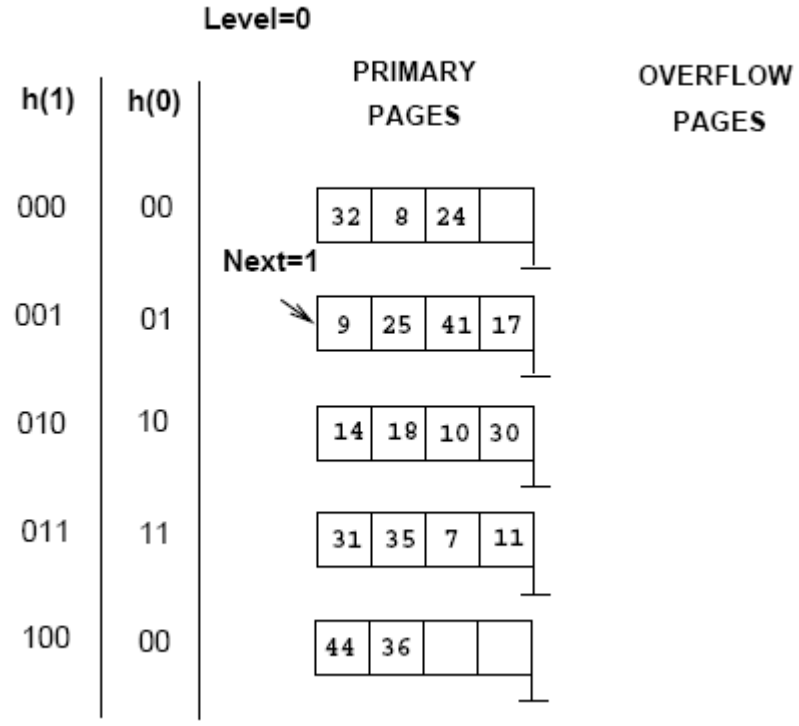


Figure 1.2 Figure for Exercise 2

3. Briefly answer the following questions:

1. Describe three techniques commonly used when developing algorithms for relational operators. Explain how these techniques can be used to design algorithms for the selection, projection, and join operators.
2. What is an access path? When does an index *match* an access path? What is a *primary conjunct*, and why is it important?
3. What information is stored in the system catalogs?
4. What are the benefits of making the system catalogs be relations?
5. What is the goal of query optimization? Why is optimization important?
6. Describe *pipelining* and its advantages.
7. Give an example query and plan in which pipelining *cannot* be used.
8. Describe the *iterator* interface and explain its advantages.
9. What role do statistics gathered from the database play in query optimization?
10. What were the important design decisions made in the System R optimizer?
11. Why do query optimizers consider only left-deep join trees? Give an example of a query and a plan that would not be considered because of this restriction.