

---

# Database Systems

Instructor: Hao-Hua Chu

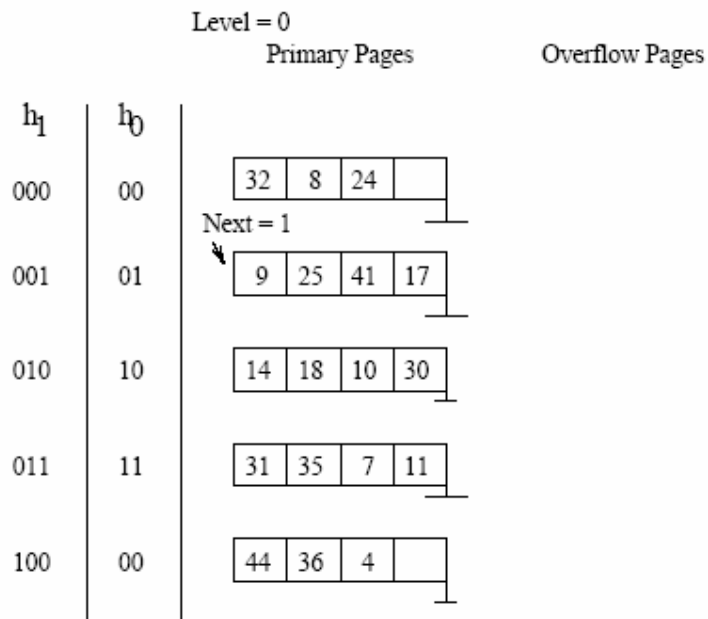
Fall Semester, 2004

---

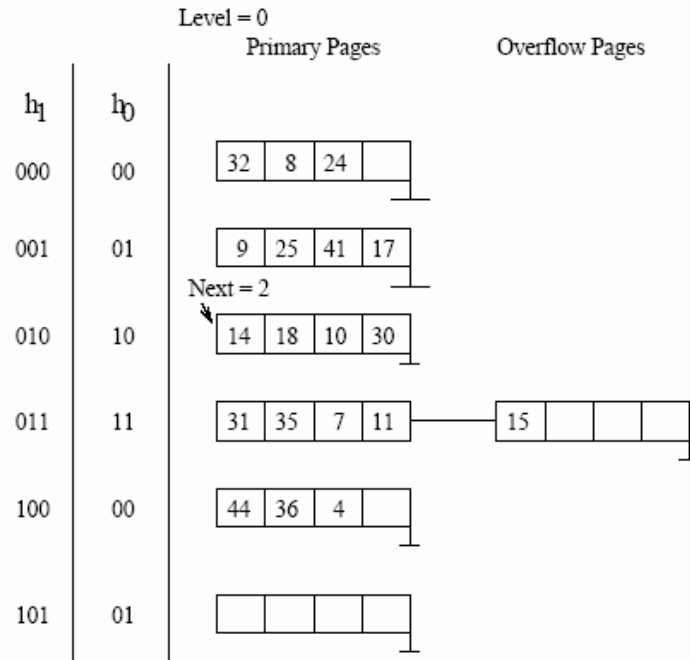
## Assignment 7 Answer: Extendible Hashing and Linear Hashing

### Answer 11.2

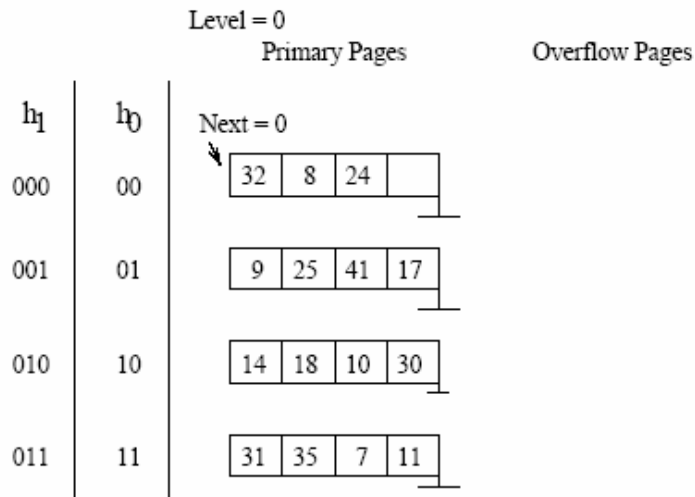
1. Nothing can be said about the last entry into the index: it can be any of the data entries in the index.
2. If the last item that was inserted had a hashcode  $h_0(\text{keyvalue}) = 00$  then it caused a split, otherwise, any value could have been inserted.
3. The last data entry which caused a split satisfies the condition  $h_0(\text{keyvalue}) = 00$ , as there are no overflow pages for any of the other buckets.
- 4.



5.



6.



7. The following constitutes the minimum list of entries to cause two overflow pages in the index :

63, 127, 255, 511, 1023

The first insertion causes a split and causes an update of *Next* to 2. The insertion of 1023 causes a subsequent split and *Next* is updated to 3 which points to this bucket.

This overflow chain will not be redistributed until three more insertions (a total of 8 entries) are made. In principle if we choose data entries with key values of the form  $2k + 3$  with sufficiently large  $k$ , we

can take the maximum number of entries that can be inserted to reduce the length of the overflow chain to be greater than any arbitrary number. This is so because the initial index has 31(binary 11111), 35(binary 10011), 7(binary 111) and 11(binary 1011). So by an appropriate choice of data entries as mentioned above we can make a split of this bucket cause just two values (7 and 31) to be redistributed to the new bucket. By choosing a sufficiently large  $k$  we can delay the reduction of the length of the overflow chain till any number of splits of this bucket.

**Answer 11.6:**

1. Let us take the data entries:

8, 16, 24, 32, 40, 48, 56, 64, 128, 7, 15, 31, 63, 127, 1, 10, 4

and the indexes shown in Fig 11.10 and Fig 11.11. Extendible hashing uses 9 pages including the directory page(assuming it spans just one page) and linear hashing uses 10 pages.

2. Consider the list of data entries

0, 4, 1, 5, 2, 6, 3, 7

and the usual hash functions for both and a page capacity of 4 records per page. Extendible hashing takes 4 data pages and also a directory page whereas linear hashing takes just 4 pages.

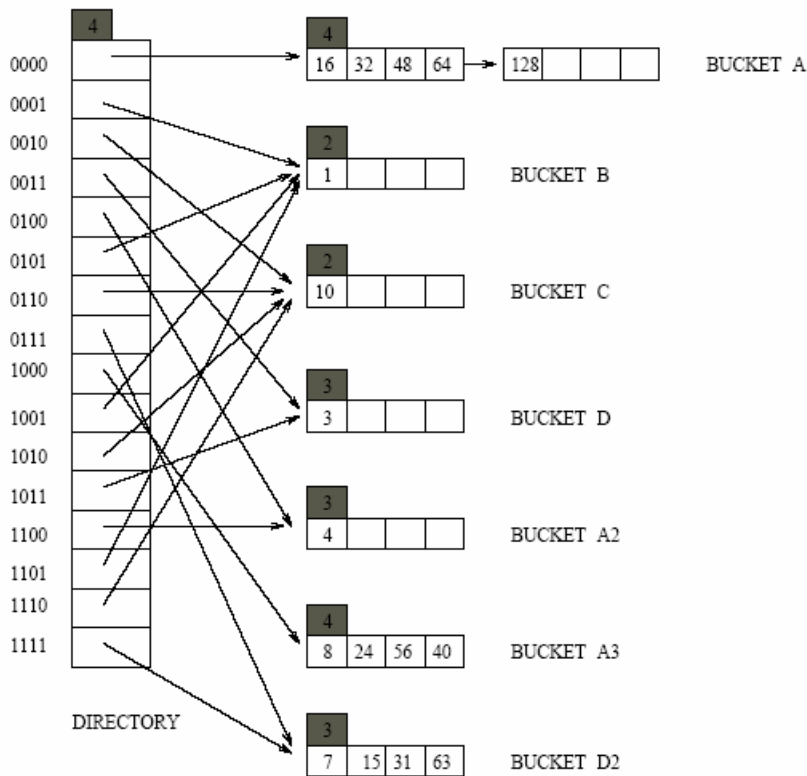


Figure 11.10

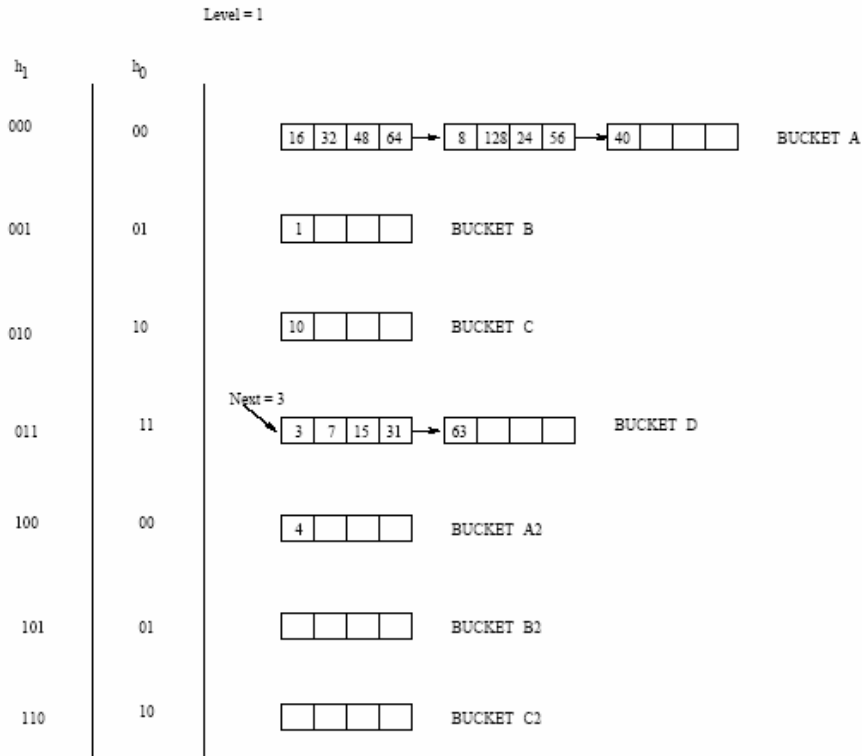


Figure 11.11

Consider Figure 11.14 in the R&G textbook, show a sequence of two record deletions that would lead to decrease of global depth to 2. Also draw the entire extendible hash table (directory, buckets, global/local depth, etc.) after the deletion sequence.

**Answer:** Remove 64 and 16.

