

CS 4700/6700 Programming Project

Simulation of Extendible Hashing.

1. The hash value of the hash-key attribute of each record is given as a randomly generated integer value.
2. Each data bucket (block) can store 40 records and is dynamically allocated. Each directory entry stores a pointer to a data bucket.
3. Generate up to 10^4 hash values (to simulate up to 10^4 records), and for every 10^2 hash values, display followings:

$$\text{utilization of the directory} = \frac{\text{number of data buckets}}{\text{number of entries in the directory}}$$

$$\text{utilization of the data buckets} = \frac{\text{number of total records}}{\text{number of records that can be stored in all the data buckets}}$$

Note: You can use any programming language, and you can plot the outputs if you want.

Reference:

1. Section 16.8.3 of the text book.
2. R. Fagin, J. Nievergelt, N. Pippenger, and H. R. Strong, Extendible Hashing: A fast access method for dynamic files, *ACM Trans. on Database Systems*, 4(3), 1979, pp. 315-344.