CHAPTER 24

Advanced Indexing Techniques

We studied the concept of indexing, as well as a number of different index structures in Chapter 14. While some index structures, such as B*-trees, were covered in detail, others such as hashing, write-optimized indices, bitmap indices, and spatial indices were only briefly outlined in Chapter 14. In this chapter we provide further details of these index structures. We provide detailed coverage of the LSM tree and its variants. We then provide a detailed description of bitmap indices. Next, we provide more detailed coverage of spatial indexing, covering quad trees and R-trees in more detail. Finally, we cover hashing, with detailed coverage of dynamic hashing techniques.

Bibliographical Notes

The Log-Structured Merge (LSM) tree is presented in [O’Neil et al. (1996)], while the Stepped Merge tree is presented in [Jagadish et al. (1997)]. [Vitter (2001)] provides an extensive survey of external-memory data structures and algorithms.

Bitmap indices, and variants called bit-sliced indices and projection indices, are described in [O’Neil and Quass (1997)]. They were first introduced in the IBM Model 204 file manager on the AS 400 platform. They provide very large speedups on certain types of queries, and are today implemented on most database systems. Research on bitmap indices includes [Wu and Buchmann (1998), Chan and Ioannidis (1998), Chan and Ioannidis (1999)], and [Johnson (1999)].

[Samet (2006)] provides a textbook coverage of spatial data structures. [Samet (1995)] provides an overview of the large amount of work on spatial index structures. An early description of the quad tree is provided by [Finkel and Bentley (1974)]. [Samet (1990)] and [Samet (1995)] describe numerous variants of quad trees. [Bentley (1975)] describes the k-d tree, and [Robinson (1981)] describes the k-d-B tree. The R-tree was originally presented in [Guttman (1984)]. Extensions of the R-tree are presented by [Sellis et al. (1987)], which describes the R+ tree, and [Beckmann et al. (1990)], which describes the R* tree. These structures provide better worst case com-
plexity guarantees for search than R-trees, but at a higher space cost. [Roussopoulos et al. (1995)] describe algorithms for nearest neighbor search on R-trees.

Discussions of the basic data structures in hashing can be found in [Cormen et al. (2009)]. [Knuth (1973)] analyzes a large number of different hashing techniques. Several dynamic hashing schemes exist. Extendable hashing was introduced by [Fagin et al. (1979)]. Linear hashing was introduced by [Litwin (1978)] and [Litwin (1980)]. A performance comparison with extendable hashing is given by [Rathi et al. (1990)]. An alternative given by [Ramakrishna and Larson (1989)] allows retrieval in a single disk access at the price of a high overhead for a small fraction of database modifications. Partitioned hashing is an extension of hashing to multiple attributes, and is covered in [Rivest (1976), Burkhard (1976)], and [Burkhard (1979)].

Bibliography


[Jagadish et al. (1997)] H. V. Jagadish, P. P. S. Narayan, S. Seshadri, S. Sudarshan, and R. Kanneganti, “Incremental Organization for Data Recording and Warehousing”, In Pro-
![Image content]

Further Reading


**Credits**

The photo of the sailboats in the beginning of the chapter is due to ©Pavel Nesvadba/Shutterstock.