Chapter 4

Data Warehousing and Online Analytical Processing

4.1 Bibliographic Notes

There are a good number of introductory-level textbooks on data warehousing and OLAP technology, e.g., Kimball, Ross, Thorntwaite et al. [KRTM08], Imhoff, Galemmo and Geiger [IGG03], Imhoff, Galemmo and Geiger [IGG03], and Inmon [Inm96]. Chaudhuri and Dayal [CD97] provide an early overview of data warehousing and OLAP technology. A set of research papers on materialized views and data warehouse implementations were collected in Materialized Views: Techniques, Implementations, and Applications by Gupta and Mumick [GM99].

The history of decision support systems can be traced back to the 1960s. However, the proposal of the construction of large data warehouses for multidimensional data analysis is credited to Codd [CCS93] who coined the term OLAP for online analytical processing. The OLAP council was established in 1995. Widom [Wid95] identified several research problems in data warehousing. Kimball and Ross [KR02] provide an overview of the deficiencies of SQL regarding the ability to support comparisons that are common in the business world and present a good set of application cases that require data warehousing and OLAP technology. For an overview of OLAP systems versus statistical databases, see Shoshani [She97].

Gray et al. [GCB+97] proposed the data cube as a relational aggregation operator generalizing group-by, crosstabs, and subtotals. Harinarayan, Rajaraman, and Ullman [HRU96] proposed a greedy algorithm for the partial materialization of cuboids in the computation of a data cube. Data cube computation methods have been investigated by numerous studies, such as Sarawagi and
Stonebraker [SS94], Agarwal et al. [AAD+96], Zhao, Deshpande, and Naughton [ZDN97], Ross and Srivastava [RS97], Beyer and Ramakrishnan [BR99], Han, Pei, Dong, Wang [HPDW01], Xin, Han, Li and Wah [XHLW03]. These methods will be discussed in depth in Chapter 5. The concept of iceberg queries was first introduced in Fang, Shivakumar, Garcia-Molina, et al. [FSGM+98]. The use of join indices to speed up relational query processing was proposed by Valduriez [Val87]. O’Neil and Graefe [OG95] proposed a bitmapped join index method to speed up OLAP-based query processing. A discussion of the performance of bitmapping and other nontraditional index techniques is given in O’Neil and Quass [OQ97].

For work regarding the selection of materialized cuboids for efficient OLAP query processing, see e.g., Chaudhuri and Dayal [CD97], Harinarayan, Rajaraman, and Ullman [HRU96], Sristava et al. [SDJL96]. Methods for cube size estimation can be found in Deshpande et al. [DNR+97], Ross and Srivastava [RS97], and Beyer and Ramakrishnan [BR99]. Agrawal, Gupta, and Sarawagi [AGS97] proposed operations for modeling multidimensional databases. Methods for answering queries quickly by online aggregation are described in Hellerstein, Haas, and Wang [HHW97] and Hellerstein et al. [HAC+99]. Techniques for estimating the top $N$ queries are proposed in Carey and Kossman [CK98] and Donjerkovic and Ramakrishnan [DR99]. Further studies on intelligent OLAP and discovery-driven exploration of data cubes are presented in the Bibliographic Notes of Chapter 5.
Bibliography


