CHAPTER 13

Data Storage Structures

In Chapter 12 we studied the characteristics of physical storage media, focusing on magnetic disks and SSDs, and saw how to build fast and reliable storage systems using multiple disks in a RAID structure. In this chapter, we focus on the organization of data stored on the underlying storage media, and how data are accessed.

Bibliographical Notes

[Hennessy et al. (2017)] is a popular textbook on computer architecture, which includes coverage of hardware aspects of translation look-aside buffers, caches, and memory-management units.

The storage structure of specific database systems, such as IBM DB2, Oracle, Microsoft SQL Server, and PostgreSQL are documented in their respective system manuals, which are available online.

Algorithms for buffer management in database systems, along with a performance evaluation, were presented by [Chou and Dewitt (1985)]. Buffer management in operating systems is discussed in most operating-system texts, including in [Silberschatz et al. (2018)].

[Abadi et al. (2008)] presents a comparison of column-oriented and row-oriented storage, including issues related to query processing and optimization.

Sybase IQ, developed in the mid 1990s, was the first commercially successful column-oriented database, designed for analytics. MonetDB and C-Store were column-oriented databases developed as academic research projects. The Vertica column-oriented database is a commercial database that grew out of C-Store, while VectorWise is a commercial database that grew out of MonetDB. As its name suggests, VectorWise supports vector processing of data, and as a result supports very high processing rates for many analytical queries. [Stonebraker et al. (2005)] describe C-Store, while [Idreos et al. (2012)] give an overview of the MonetDB project and [Zukowski et al. (2012)] describes Vectorwise.

The ORC and Parquet columnar file formats were developed to support compressed storage of data for big-data applications that run on the Apache Hadoop platform.

With the rapid increase in CPU speeds, cache memory located along with the CPU has become much faster than main memory. Although database systems do not control what data are kept in cache, there is an increasing motivation to organize data in memory and write programs in such a way that cache utilization is maximized. Work in this area includes [Rao and Ross (2000)], [Ailamaki et al. (2001)], [Zhou and Ross (2004)], [Garcia and Korth (2005)], and [Cieslewicz et al. (2009)].

Buffering data in mobile systems is discussed in [Imielinski and Badrinath (1994)], [Imielinski and Korth (1996)].

SAP HANA is an in-memory database that uses a compressed column-oriented representation for data that is stored in memory; [Lee et al. (2013)] provide an overview of HANA. HyPer [Kemper et al. (2012)] is a main-memory database that supports column-oriented storage, but with the ability to store multiple columns together for more efficient access, to efficiently support both transaction processing and analytical query processing.

Bibliography

[Abadi et al. (2008)] D. J. Abadi, S. Madden, and N. Hachem, "Column-Stores vs. Row-Stores: How Different Are They Really?", In *Proc. of the ACM SIGMOD Conf. on Management of Data* (2008), pages 967–980.

[Ailamaki et al. (2001)] A. Ailamaki, D. J. DeWitt, M. D. Hill, and M. Skounakis, "Weaving Relations for Cache Performance", In *Proc. of the International Conf. on Very Large Databases* (2001), pages 169–180.

[Chou and Dewitt (1985)] H. T. Chou and D. J. Dewitt, "An Evaluation of Buffer Management Strategies for Relational Database Systems", In *Proc. of the International Conf. on Very Large Databases* (1985), pages 127–141.

[Cieslewicz et al. (2009)] J. Cieslewicz, W. Mee, and K. A. Ross, "Cache-Conscious Buffering for Database Operators with State", In *Proc. Fifth International Workshop on Data Management on New Hardware (DaMoN 2009)* (2009), pages 43–51.

[Garcia and Korth (2005)] P. Garcia and H. F. Korth, "Multithreaded Architectures and the Sort Benchmark", In *Proc. of the First International Workshop on Data Management on Modern Hardware (DaMoN)* (2005).

[Hennessy et al. (2017)] J. L. Hennessy, D. A. Patterson, and D. Goldberg, *Computer Architecture: A Quantitative Approach*, 6th edition, Morgan Kaufmann (2017).

[Idreos et al. (2012)] S. Idreos, F. Groffen, N. Nes, S. Manegold, K. S. Mullender, and M. L. Kersten, "MonetDB: Two Decades of Research in Column-oriented Database Architectures", *IEEE Data Engineering Bulletin*, Volume 35, Number 1 (2012), pages 40–45.

[Imielinski and Badrinath (1994)] T. Imielinski and B. R. Badrinath, "Mobile Computing — Solutions and Challenges", *Communications of the ACM*, Volume 37, Number 10 (1994), pages 18–28.

[Imielinski and Korth (1996)] T. Imielinski and H. F. Korth, editors, *Mobile Computing*, Kluwer Academic Publishers (1996).

[Kemper et al. (2012)] A. Kemper, T. Neumann, F. Funke, V. Leis, and H. Mühe, "HyPer: Adapting Columnar Main-Memory Data Management for Transaction AND Query Processing", *IEEE Data Engineering Bulletin*, Volume 35, Number 1 (2012), pages 46–51.

[Lee et al. (2013)] J. Lee, M. Muehle, N. May, F. Faerber, V. Sikka, H. Plattner, J. Krüger, and M. Grund, "High-Performance Transaction Processing in SAP HANA", *IEEE Data Engineering Bulletin*, Volume 36, Number 2 (2013), pages 28–33.

[Rao and Ross (2000)] J. Rao and K. A. Ross, "Making B+-Trees Cache Conscious in Main Memory", In *Proc. of the ACM SIGMOD Conf. on Management of Data* (2000), pages 475–486.

[Silberschatz et al. (2018)] A. Silberschatz, P. B. Galvin, and G. Gagne, *Operating System Concepts*, 10th edition, John Wiley and Sons (2018).

[Stonebraker et al. (2005)] M. Stonebraker, D. J. Abadi, A. Batkin, X. Chen, M. Cherniack, M. Ferreira, E. Lau, A. Lin, S. Madden, E. J. O'Neil, P. E. O'Neil, A. Rasin, N. Tran, and S. B. Zdonik, "C-Store: A Column-oriented DBMS", In *Proc. of the International Conf. on Very Large Databases* (2005), pages 553–564.

[Zhou and Ross (2004)] J. Zhou and K. A. Ross, "Buffering Database Operations for Enhanced Instruction Cache Performance", In *Proc. of the ACM SIGMOD Conf. on Management of Data* (2004), pages 191–202.

[Zukowski et al. (2012)] M. Zukowski, M. van de Wiel, and P. A. Boncz, "Vectorwise: A Vectorized Analytical DBMS", In *Proc. of the International Conf. on Data Engineering* (2012), pages 1349–1350.