CHAPTER 21

Parallel and Distributed Storage

Parallelism is used to provide speedup, where queries are executed faster because more resources, such as processors and disks, are provided. Parallelism is also used to provide scaleup, where increasing workloads are handled without increased response time, via an increase in the degree of parallelism.

There are many techniques for data storage and indexing in parallel and distributed database systems.

Teradata was one of the first commercial shared-nothing parallel database systems designed for decision support systems, and it continues to have a large market share. Teradata supports partitioning and replication of data to deal with node failures. The Red Brick Warehouse was another early parallel database system designed for decision support (Red Brick was bought by Informix, and later IBM).

Bibliographical Notes

In the late 1970s and early 1980s, as the relational model gained reasonably sound footing, people recognized that relational operators are highly parallelizable and have good dataflow properties. Several research projects, including GAMMA ([DeWitt (1990)]), XPRS ([Stonebraker et al. (1988)]), and Volcano ([Graefe (1990)]) were launched to investigate the practicality of parallel storage of data and parallel execution of queries.

Information on the Google file system can be found in [Ghemawat et al. (2003)], while the Google Bigtable system is described in [Chang et al. (2008)]. The Yahoo! PNUTS system is described in [Cooper et al. (2008)], while Google Megastore and Google Spanner are described in [Baker et al. (2011)] and [Corbett et al. (2013)] respectively. Consistent hashing is described in [Karger et al. (1997)], while Dynamo, which is based on consistent hashing, is described in [DeCandia et al. (2007)].

Bibliography

[Baker et al. (2011)] J. Baker, C. Bond, J. C. Corbett, J. J. Furman, A. Khorlin, J. Larson, J.-M. Leon, Y. Li, A. Lloyd, and V. Yushprakh, "Megastore: Providing Scalable, Highly Available Storage for Interactive Services", In *Proceedings of the Conference on Innovative Data system Research (CIDR)* (2011), pages 223–234.

[Chang et al. (2008)] F. Chang, J. Dean, S. Ghemawat, W. C. Hsieh, D. A. Wallach, M. Burrows, T. Chandra, A. Fikes, and R. E. Gruber, "Bigtable: A Distributed Storage System for Structured Data", *ACM Trans. Comput. Syst.*, Volume 26, Number 2 (2008).

[Cooper et al. (2008)] B. F. Cooper, R. Ramakrishnan, U. Srivastava, A. Silberstein, P. Bohannon, H.-A. Jacobsen, N. Puz, D. Weaver, and R. Yerneni, "PNUTS: Yahoo!'s Hosted Data Serving Platform", *Proceedings of the VLDB Endowment*, Volume 1, Number 2 (2008), pages 1277–1288.

[Corbett et al. (2013)] J. C. Corbett et al., "Spanner: Google's Globally Distributed Database", *ACM Trans. on Computer Systems*, Volume 31, Number 3 (2013).

[DeCandia et al. (2007)] G. DeCandia, D. Hastorun, M. Jampani, G. Kakulapati, A. Lakshman, A. Pilchin, S. Sivasubramanian, P. Vosshall, and W. Vogels, "Dynamo: Amazons Highly Available Key-value Store", In *Proc. of the ACM Symposium on Operating System Principles* (2007), pages 205–220.

[DeWitt (1990)] D. DeWitt, "The Gamma Database Machine Project", *IEEE Transactions on Knowledge and Data Engineering*, Volume 2, Number 1 (1990), pages 44–62.

[Ghemawat et al. (2003)] S. Ghemawat, H. Gobioff, and S.-T. Leung, "The Google File System", *Proc. of the ACM Symposium on Operating System Principles* (2003).

[Graefe (1990)] G. Graefe, "Encapsulation of Parallelism in the Volcano Query Processing System", In *Proc. of the ACM SIGMOD Conf. on Management of Data* (1990), pages 102–111.

[Karger et al. (1997)] D. Karger, E. Lehman, T. Leighton, R. Panigrahy, M. Levine, and D. Lewin, "Consistent Hashing and Random Trees: Distributed Caching Protocols for Relieving Hot Spots on the World Wide Web", In *Proc. of the ACM Symposium on Theory of Computing* (1997), pages 654–663.

[Stonebraker et al. (1988)] M. Stonebraker, R. H. Katz, D. A. Patterson, and J. K. Ousterhout, "The Design of XPRS", In *Proc. of the International Conf. on Very Large Databases* (1988), pages 318–330.