CHAPTER 19

Recovery System

A computer system, like any other device, is subject to failure from a variety of causes: disk crash, power outage, software error, a fire in the machine room, even sabotage. In any failure, information may be lost. Therefore, the database system must take actions in advance to ensure that the atomicity and durability properties of transactions, introduced in Chapter 17, are preserved. An integral part of a database system is a **recovery scheme** that can restore the database to the consistent state that existed before the failure.

The recovery scheme must also support **high availability**, that is, the database should be usable for a very high percentage of time. To support high availability in the face of machine failure (as also planned machine shutdowns for hardware/software upgrades and maintenance), the recovery scheme must support the ability to keep a backup copy of the database synchronized with the current contents of the primary copy of the database. If the machine with the primary copy fails, transaction processing can continue on the backup copy.

Bibliographical Notes

Tutorial and survey papers on various recovery techniques for database systems include [Gray (1978)], [Lindsay et al. (1980)], and [Verhofstad (1978)]. The concepts of fuzzy checkpointing and fuzzy dumps are described in [Lindsay et al. (1980)].

Remote backup algorithms for disaster recovery are presented in [Lyon (1988)], [King et al. (1991)] and [Polyzois and Garcia-Molina (1994)].

The state of the art in recovery methods is best illustrated by the ARIES recovery method, described in [Mohan et al. (1992)] and [Mohan (1990)]. [Mohan and Levine (1992)] presents ARIES IM, an extension of ARIES to optimize B^+-tree concurrency control and recovery using logical undo logging. ARIES and its variants are used in several database products, including IBM DB2 and Microsoft SQL Server. Recovery in Oracle is described in [Lahiri et al. (2001)].

Specialized recovery techniques for index structures are described in [Mohan and Levine (1992)] and [Mohan (1993)]; [Mohan and Narang (1994)] describes recovery techniques for client–server architectures, while [Mohan and Narang (1992)] describes recovery techniques for parallel-database architectures. A generalized version of the theory of serializability, with short-duration lower-level locks during operations, combined with longer-duration higher-level locks, is described by [Weikum (1991)]. In Section 19.8.3, we saw the requirement that an operation should acquire all lower-level locks that may be needed for the logical undo of the operation. This requirement can be relaxed by performing all physical undo operations first, before performing any logical undo operations. A generalized version of this idea, called multilevel recovery, presented in [Weikum et al. (1990)], allows multiple levels of logical operations, with level-by-level undo passes during recovery.

[Faerber et al. (2017)] provide an overview of main-memory databases, including recovery techniques. [Diaconu et al. (2013)] provide an overview of the Hekaton main-memory engine which is part of SQL Server, including its recovery algorithms.

Bibliography

[Bayer et al. (1978)] R. Bayer, R. M. Graham, and G. Seegmuller, editors, *Operating Systems: An Advanced Course*, volume 60 of *Lecture Notes in Computer Science*, Springer Verlag (1978).

[Diaconu et al. (2013)] C. Diaconu, C. Freedman, E. Ismert, P.-A. Larson, P. Mittal, R. Stonecipher, N. Verma, and M. Zwilling, “Hekaton: SQL Server’s Memory-optimized OLTP Engine”, In *Proc. of the ACM SIGMOD Conf. on Management of Data* (2013), pages 1243–1254.

[Faerber et al. (2017)] F. Faerber, A. Kemper, P.-A. Larson, J. Levandoski, T. Neumann, and A. Pavlo, “Main Memory Database Systems”, *Foundations and Trends in Databases*, Volume 8, Number 1-2 (2017), pages 1–130.

[Gray (1978)] J. Gray. “Notes on Data Base Operating System”, In *[Bayer et al. (1978)]*, pages 393–481. Springer Verlag (1978).

[King et al. (1991)] R. P. King, N. Halim, H. Garcia-Molina, and C. Polyzois, “Management of a Remote Backup Copy for Disaster Recovery”, *ACM Transactions on Database Systems*, Volume 16, Number 2 (1991), pages 338–368.

[Lahiri et al. (2001)] T. Lahiri, A. Ganesh, R. Weiss, and A. Joshi, “Fast-Start: Quick Fault Recovery in Oracle”, In *Proc. of the ACM SIGMOD Conf. on Management of Data* (2001), pages 593–598.

[Lindsay et al. (1980)] B. G. Lindsay, P. G. Selinger, C. Galtieri, J. N. Gray, R. A. Lorie, T. G. Price, G. R. Putzolu, I. L. Traiger, and B. W. Wade. “Notes on Distributed Databases”, In Draffen and Poole, editors, *Distributed Data Bases*, pages 247–284. Cambridge University Press (1980).

[Lyon (1988)] J. Lyon, “Design considerations in replicated database systems for disaster protection”, In *IEEE Compcon* (1988).

[Mohan (1990)] C. Mohan, "Commit-LSN: A Novel and Simple Method for Reducing Locking and Latching in Transaction Processing Systems", In *Proc. of the International Conf. on Very Large Databases* (1990), pages 406–418.

[Mohan (1993)] C. Mohan, "IBM's Relational Database Products: Features and Technologies", In *Proc. of the ACM SIGMOD Conf. on Management of Data* (1993), pages 445–448.

[Mohan and Levine (1992)] C. Mohan and F. Levine, "ARIES/IM:An Efficient and High-Concurrency Index Management Method Using Write-Ahead Logging", In *Proc. of the ACM SIGMOD Conf. on Management of Data* (1992), pages 371–380.

[Mohan and Narang (1992)] C. Mohan and I. Narang, "Efficient Locking and Caching of Data in the Multisystem Shared Disks Transaction Environment", In *Proc. of the International Conf. on Extending Database Technology* (1992), pages 453–468.

[Mohan and Narang (1994)] C. Mohan and I. Narang, "ARIES/CSA: A Method for Database Recovery in Client-Server Architectures", In *Proc. of the ACM SIGMOD Conf. on Management of Data* (1994), pages 55–66.

[Mohan et al. (1992)] C. Mohan, D. Haderle, B. Lindsay, H. Pirahesh, and P. Schwarz, "ARIES: A Transaction Recovery Method Supporting Fine-Granularity Locking and Partial Rollbacks Using Write-Ahead Logging", *ACM Transactions on Database Systems*, Volume 17, Number 1 (1992), pages 94–162.

[Polyzois and Garcia-Molina (1994)] C. Polyzois and H. Garcia-Molina, "Evaluation of Remote Backup Algorithms for Transaction-Processing Systems", *ACM Transactions on Database Systems*, Volume 19, Number 3 (1994), pages 423–449.

[Verhofstad (1978)] J. S. M. Verhofstad, "Recovery Techniques for Database Systems", *ACM Computing Surveys*, Volume 10, Number 2 (1978), pages 167–195.

[Weikum (1991)] G. Weikum, "Principles and Realization Strategies of Multilevel Transaction Management", *ACM Transactions on Database Systems*, Volume 16, Number 1 (1991), pages 132–180.

[Weikum et al. (1990)] G. Weikum, C. Hasse, P. Broessler, and P. Muth, "Multi-Level Recovery", In *Proc. of the ACM SIGMOD Conf. on Management of Data* (1990), pages 109–123.

Credits

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