

Database Qualifying Exam Reading List, Fall 2007

INFORMATION

The Database Management Systems (DBMS) qualifying exam is intended to cover a wide range of database systems literature, testing the students' preparedness and ability for pursuing serious research in the database area. As such, students who do not intend to pursue database-related Ph.D. topics are advised not to attempt taking this exam. The exam will be graded with the expectation that each examinee will have a fairly deep understanding of the basic issues plus a broad knowledge of recent and current database research work.

The exam is intended to cover a set of fundamental topics in databases. Students should have a solid grasp of the database systems area at the advanced undergraduate / introductory graduate level, as can be obtained by studying Ramakrishnan and Gerhke, or a comparable textbook, and in addition, they should have a deep understanding of many of the advanced topics discussed in the research literature. The following list of topics and papers indicates the material that all students will be expected to be familiar with. Students should take CS 564, CS 764, and CS 784 in preparation for the database qualifying exam.

REFERENCES

As mentioned above, either of the following textbook (among others) is a good introductory source of database management system material.

Ramakrishnan, R., and Gehrke, J., *Database Management Systems*, 3rd Edition, McGraw-Hill, 2003.

In particular, you are expected to be familiar with the coverage of the following topics at the level of this text: file organizations, database design, data models and languages, data mining, decision support, data warehousing, deductive databases, information retrieval, object-oriented and object-relational databases, query processing, transaction management, view maintenance, security and integrity, and XML. Further required reading, in the form of papers from the research literature, is listed below.

Case Studies

Chamberlin, D., et al., "A History and Evaluation of System R", *Communications of the ACM* 24(10), 1981.

Stonebraker, M., et al., "The Design and Implementation of INGRES", *ACM Transactions on Database Systems* 1(3), 1976.

File Organizations and Access Methods

Comer, D., "The Ubiquitous B-Tree", *ACM Computing Surveys* 11(2), June 1979.

Nievergelt, J., et al., "The Grid File: An Adaptable, Symmetric Multikey File Structure", *ACM Transactions on Database Systems* 9(1), March 1984.

Guttman, A., "R-Trees: A Dynamic Index Structure for Spatial Searching", *Proceedings of the ACM SIGMOD International Conference on the Management of Data*, 1984.

Brin, S. and Page, L., "The Anatomy of a Large-Scale Hypertextual Web Search Engine", *Proceedings of Computer Networks and ISDN Systems*, 1998.

Data Models and Languages

Stonebraker, M., "Inclusion of New Types In Relational Data Base Systems", *Proceedings of the International Conference on Data Engineering*, 1986.

Stonebraker, M. and L. Rowe, "The Design of Postgres," *Proceedings of the ACM-SIGMOD International Conference on Management of Data*, 1986.

Lamb, C., Landis, G., Orenstein, J., and Weinreb, D., "The ObjectStore Database System", *Communications of the ACM*, 34(10), 1991.

Blakely, J.A., Larson, P-A., and Tompa, F.W., "Efficiently Updating Materialized Views", *Proceedings of the ACM-SIGMOD International Conference on Management of Data*, 1986.

Zaniolo, C., "The Database Language GEM", *Proceedings of the ACM-SIGMOD International Conference on Management of Data*, 1983.

Chamberlin, D., "XQuery: An XML Query Language", *IBM Systems Journal*, Vol 41, No. 4, 2002.

Query Processing and Optimization

Shapiro, L. D., "Join Processing in Database Systems with Large Main Memories", *ACM Transactions on Database Systems* 11(3), 1986.

Selinger, P., et al., "Access Path Selection in a Relational Database Management System", *Proceedings of the ACM-SIGMOD International Conference on Management of Data*, 1979.

Concurrency Control and Recovery

Gray, J., et al., "Granularity of Locks and Degrees of Consistency in a Shared Data Base", *Proceedings of the IFIP Working Conference on Modeling of Data Base Management Systems*, 1979.

Kung, H., and Robinson, J., "On Optimistic Methods for Concurrency Control", *ACM Transactions on Database Systems* 6(2), June 1981.

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

Lehman, P. and Yao, S., "Efficient Locking for Concurrent Operations on B-Trees", *ACM Transactions on Database Systems*, 6(4): 650-670.

Distributed and Parallel Databases

Lohman, G., and Mackert, L., "R* Optimizer Validation and Performance Evaluation for Distributed Queries", *Proceedings of the 12th International Conference on Very Large Data Bases*, 1986.

Mohan, C., Lindsay, B., and Obermarck R., "Transaction Management in the R* Distributed Database Management System", *ACM Transactions on Database Systems* 11 (4), 1986.

DeWitt, D., et al, “The GAMMA Database Machine Project,” *IEEE Transactions on Knowledge and Data Engineering* 2(1), 1990.

Gray, J., et al., “The Dangers of Replication and a Solution”, *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 1996.

Operating System Issues

Chou, H., and DeWitt, D., “An Evaluation of Buffer Management Strategies for Relational Database Systems”, *Proceedings of the 11th International Conference on Very Large Data Bases*, 1985.

Stonebraker, M., “Operating System Support for Database Management”, *Communications of the ACM* 24(7), 1981.

Database Theory

Aho, A., et al, “Equivalences Among Relational Expressions”, *SIAM Journal on Computing* 8(2), 1979.

Date, C. D., and Fagin, R., “Simple Conditions for Guaranteeing Higher Normal Forms in Relational Databases”, *ACM Transactions on Database Systems* 17(3), 1992.

Aho, A., and Ullman, J., “Universality of Data Retrieval Languages”, *Proceedings of the 6th ACM Symposium on Principles of Programming Languages*, 1979.

Bernstein, P.A., Hadzilacos, V., and Goodman, N., “Concurrency Control and Recovery in Database Systems”, *Addison-Wesley*, 1987; can be freely downloaded from Bernstein’s webpage. (Chapters 1 and 2)

Temporal Databases

Snodgrass, R., and Ahn, I., “A Taxonomy of Time in Databases”, *Proceedings of the ACM-SIGMOD International Conference on Management of Data*, 1985.

Decision Support

O’Neil, P., and D. Quass, “Improved Query Performance with Variant Indexes.” *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 1997.

Gray, J., et al., “Data Cube” A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Totals.” In *Data Mining and Knowledge Discovery*, 1(1): 29—53.

Data Mining

Agrawal, R., and R. Srikant, “Fast Algorithms for Mining Association Rules.” In *Proceedings of the 20th International Conference on Very Large Data Bases*, 487—499.

Zhang, T., R. Ramakrishnan, and M. Livny, “BIRCH: A Clustering Algorithm for Large Multidimensional Datasets.” *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 1996.

Schema Integration

Rahm, E., and P. Bernstein, "A Survey of Approaches to Automatic Schema Matching." *The VLDB Journal*, 10(4):334-350, December 2001.

DBMS and Search Engines

Brewer, Eric, "Combining Systems and Databases: A Search Engine Retrospective." *Readings in Database Systems*, Fourth Edition. Joseph M. Hellerstein and Michael Stonebraker eds, MIT Press, Cambridge, MA, 2005.

Information Retrieval

Singhal, A., "Modern Information Retrieval: a Brief Overview." *IEEE Data Engineering Bulletin*, 24(4), 35-43, 2001.