Database Qualifying Exam Reading List, Last Revision 2014

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 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 are encouraged to take CS 564, CS 764, and CS 784 in preparation for the database qualifying exam.

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

You are expected to be familiar with the coverage of the following topics in this text (or a comparable text book): file organizations, indexing (B-tree and hash), 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 and Architecture

- Chamberlin, D. D., Astrahan, M. M., Blasgen, M. W., Gray, J. N., King, W. F., Lindsay, B. G., Lorie, R., Mehl, J. W., Price, T. G., Putzolu, F., Selinger, P. G., Schkolnick, M., Slutz, D. R., Traiger, I. L., Wade, B. W. and Yost, R. A. "A History and Evaluation of System R", Communications of the ACM 24(10), 1981.

- Stonebraker, M., Wong E., Kreps P., and Held G., "The Design and Implementation of INGRES", ACM Transactions on Database Systems 1(3), 1976.

- Hellerstein J. M., Stonebraker M. and Hamilton, J. R., "Architecture of a Database System", Foundations and Trends in Databases 1(2), 2007.

File Organizations and Access Methods

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

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

Data Models and Languages

- J. Ullman. Database and Knowledge Base Systems, vol. I. Chapter 3 (Logic as a Data Model).

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

- 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.

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., Lorie, R. A., Pulzolu, G. R., Traiger, I. L., "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.

- Berenson, H., Bernstein, P. A., Gray, J., Melton, J., O'Neil, E. J., O'Neil, P. E., "A Critique of ANSI SQL Isolation Levels", *Proceedings of the ACM-SIGMOD International Conference on Management of Data,* 1995.

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

- 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.

- 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)

Distributed and Parallel Data Processing

- 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.

- Dean J. and Ghemawat S., "MapReduce: Simplified Data Processing on Large Clusters", Proceedings of the Symposium on Operating System Design and Implementation (OSDI) 2004.

Operating System Issues

- Chou, H., and DeWitt, D., "An Evaluation of Buffer Management Strategies for Relational Database Systems", *Proceedings of the International Conference on Very Large Data Bases (VLDB),* 1985.

- O'Neil, E. J., O'Neil, P. E., Weikum G., "The LRU-K Page Replacement Algorithm For Database Disk Buffering", *Proceedings of the ACM-SIGMOD International Conference on Management of Data,* 1993.

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

Database Theory

- S. Abiteboul, R. Hull, V. Vianu. Foundations of Databases. Chapter 6, Sections 6.2 (Global Optimizations), Chapter 12, Sections 12.1 – 12.3 (Datalog), Chapter 13, Section 13.1 (Datalog evaluation). ; Available for free from: http://webdam.inria.fr/Alice/

- 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.

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., Chaudhuri, S., Bosworth, A., Layman, A., Reichart D., Venkatrao, M., Pellow, F., and Pirahesh, H., "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., Ramakrishnan R., and Livny M., "BIRCH: A Clustering Algorithm for Large Multidimensional Datasets", *Proceedings of the ACM SIGMOD International Conference on Management of Data,* 1996

Data Integration

- Levy, Alon, "Logic-based Techniques in Data Integration", available on the Web at http://homes.cs.washington.edu/~alon/site/files/levy-di00.ps, read up to and include Section 5.1.

- Doan, A., Halevy, A., Ives, Z., "Principles of Data Integration", Chapter 1, Chapter 5, Chapter 4: read 4.1, 4.2.1 (only "Edit Distance"), 4.2.2 (only "Overlap", "Jaccard", and "TF/IDF"), 4.2.4, and 4.3 (only "Inverted Index" and "Size Filtering"). Chapter 7: up to and including 7.5.3. Chapter 9: read 9.1, 9.2, and 9.3.1. Chapters available from http://pages.cs.wisc.edu/~anhai/courses/dibook-chapters.

DBMS and Search Engines

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

- 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.

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

- Page, L. and Brin, S. and Motwani, R. and Winograd, T., "The Pagerank Citation Ranking: Bringing Order to the Web", Technical Report, 1999.