Database Qualifying Exam Reading List

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.

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 Gehrke, 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 the following (or equivalent) courses in preparation for the database qualifying exam.

- CS 564 Database Management Systems: Design and Implementation
- CS 764 Topics in Database Management Systems
- CS 784 Foundations of Data Management
- CS 774 (previous CS 784)

References

You are expected to be familiar with the coverage of the following topics in the text below (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 and manuscripts from the research literature, is listed below.

Data Models and Languages

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

  - Conjunctive Queries (Chapters 3, 4)
  - Chapter 6, Sections 6.2 and 6.4
  - Datalog (Chapter 12, Sections 12.1 - 12.3, Chapter 13, Section 13.1 - 13.3)

DBMS Architecture


Operating System Issues


File Organizations and Access Methods

Query Processing and Optimization


Concurrency Control and Recovery

- 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

Data Analysis and Decision Support


Data Extraction and Integration


- 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


Emerging Topics
