

# Midterm 1, Review Sheet, V2.0

## Database Systems

CS386D  
Professor Daniel Miranker

### Time, Place, Format:

The midterm will be held in class, Monday May 2, 6:00 – 7:15pm. The exam will have an unusual format. The first hour, 6:00 – 7:00, it will be a closed-book exam. (calculators ok, but not phones serving as calculators). At 7:00 it will be announced that you have permission to open your books, and the [same] exam paper may be addressed per an open book exam.

### Material:

This review sheet is intended only as a study guide concerning the breadth of the exam. You are expected to know all the terminology presented as covered in class, the texts and the required supplemental reading. To be clear, individual terms and topics in this document are indicative of the breadth, *not* a comprehensive syllabus for the exam.

### Reading:

Assumed background knowledge: Ch. 2.1-2.4, 4.1-4.5, 5.1,5.2 Ch. 6,  
As indicated on the lecture and reading schedule.

### Topics:

1. Basic Relational Database Concepts
  - a. Schema(s)
  - b. Associative Access and Declarative (logic) Programming as the way databases are thought about.
  - c. Keys
    - i. Candidate key
    - ii. Primary key
    - iii. Foreign key
    - iv. Search/index key
    - v. \*\* per preperation for the exam, explore just how many uses of the word key, or *<adjective>* key, have gone by. Get to a point that you are certain you've identified all of them, and can answer, per advice in lecture, "what is the important concept that deserved to be named". Don't forget, its ok to ask your mother, just don't tell me you asked your mother.
  - d. Basic Organization
    - i) Data/index files
    - ii) transaction log
    - iii) SQL Engine
    - iv) Role of RDBMS in a three-tier architecture
5. Disks and Data
  - a. Physical properties of disk drives.
  - b. Sequential access vs. random access
    - the supporting details of HDDs and SSDs such that, despite very different technologies, one can replace an HDD with an SSD, and be successful without any change in software.
  - ~~b. Two-phase multiway merge sort~~
  - ~~c. RAID~~
6. Indexing
  - a. Methods – "The 3 B's"
    - i. B+-trees
    - iii. Bit-vector index methods
    - iii. Bloom filters
  - b. Secondary Indexes

- i. Applicability
- ii. Clustering
- iii. Effectiveness
  - Measures of effectiveness
  - Parameters,  $B(R)$ ,  $T(R)$ ,  $V(R, attr)$

## 8. Transactions

- a. ACID properties
  - a. A...
  - C...
  - I...
  - D...
- b. No deeper than a log, on stable store exists in a relational database to support transactions. But its existence and role is noted. (later in the semester we will cover algorithms that speak precisely to what and when is written to the log.)

## 9. Storage Management in Contemporary, Cloud-native Databases

- b. Machine organizations. e.g. shared-nothing
- c. Data Partitioning methods
  - round-robin
  - hashing
  - range based
- d. Application and use of hash partitioning in Cloud databases
  - Basically the scope of the very last homework problem
- e. Semi-join reduction
  - core idea -
  - application to reduce communication in a simple distributed join.
  - approximate implementation using Bloom filters.
  - (this topic was introduced at this early, and shallow level, in order to press home the point that you must be comfortable with the relational algebra, and the relational algebra and this very early material is foundational and you have in fact seen enough to have at least some visibility in to big database systems. Thus, my advise as you prepare is simply to be completely comfortable with the individual ideas and their interaction, and not worry about complex problem solving. Much more intricate application and integration is coming later in the semester.)