424 Final review

- Exam
  - 200 pts
  - two hours
  - not comprehensive

Current grades:
- A-: 56.6 points (grades.cs, this is not %)
- B-: 50.1
- C-: 43.6

Questions:
- storage hierarchy
  - disk reads/writes
  - RAID (1,5)
  - differences w/ solid state drives
- buffer manager
  - LRU / MRU
  - pinning, force
- file organization
  - mapping of tables to files
    - sorted
      - choice of search/sort key important
    - hash-based
      - integral number (1?) buckets per page
- indexes
  - primary (#?) vs secondary (#?)
  - dense vs sparse
  - b+-trees
    - why good?
    - insert/delete
      - min \( \left\lceil \frac{n-1}{2} \right\rceil \) keys in leaf
      - min \( \left\lceil \frac{n}{2} \right\rceil \) pointers in non-leaf
    - splits on inserts
      - at least \( \left\lfloor \frac{n}{2} \right\rfloor \) keys in left leaf

As of 12/11/19, before the inclusion of P5.
• at least $\left\lceil \frac{n}{2} \right\rceil$ pointers in left leaf
  ■ fanout height
  ○ multi-level
  ○ hash
    ■ very fast on equality
    ■ really slow on ranges
• joins
  ○ nest-loop
  ○ block nested loop
  ○ merge join
  ○ hash join
• query selection costs
  ○ know the table (12.3)
  ○ now how to combine conjunctive / disjunctive
  ○ histograms
    ■ nothing after this in estimation
  ○ no optimization
  ○
• transactions
  ○ ACID
  ○ serializability
    ■ conflict
    ■ view
    ■ other
    ■ definitions
      • recoverability
      • cascading rollbacks
      • dirty reads
      • cascadeless schedules
    ■ concurrency control
      ■ locks
        ○ granularity
        ○ intention locks
        ○ two-phase
          ■ strict
          ■ rigorous
        ○ lock point
• deadlocks
  ○ finding
  ○ dealing with
  ○ preventing
  ○ starvation
• time-stamp based CC
  ○ validation / transaction aborts
  ○ thomas’ write rule
• snapshot isolation
  ○ write-ahead logging
  ▪ buffer manager
    ○ force vs no-force
    ○ steal vs no-steal
    ○ recover
      ▪ redo lists
      ▪ undo lists
      ▪ checkpointing
  ○ security
    ▪ SQL injection
    ▪ blacklist vs whitelisting
    ▪ prepared statements
    ▪ encryption
  ○ differential privacy
    ▪ inference
    ▪ k-anonymity
    ▪ output perturbation vs perturbed (statistical) db
• Distributed Systems
  ▪ CAP theorem
  ▪ eventual / causal consistency
  ▪ strict serializability