CMSC 461 Final Exam Study Guide

Study Guide Key

Symbol	*Significance*
*	High likelihood it will be on the final
+	Expected to have deep knowledge of – can convey knowledge by working through an example problem

IMPORTANT: **Anything listed in this guide is a candidate for the final exam**. If it is not listed in this guide, it will not be on the exam. This key is meant to provide hints only. **If an item does not have an * next to it, that does NOT imply it will not be on the final.**

Assume any sub-item of an item inherits the symbols of that item.

Chapter 1

Purpose of Database Systems
Data abstraction levels
Data models
Database languages
Basic definition of a relational database

Chapter 2

Basic terms (relation, tuple, attribute schema, instance)
+Keys (superkey, candidate key, primary key, foreign key)
Relational operations

Chapter 6

+Relational algebra
- selection
- projection
- union
- set-difference
- Cartesian product
- rename
- set-intersection
- natural join
- outer join
- left and right outer join
- aggregation

Chapter 3, 4, 5

Definition of SQL
Data types

+How to define a schema
+Queries
    select
    from
    where
    natural join
    renaming
    correlation variables
    string function
    ordering
    set operations (union, intersection, except)
    aggregation (avg, min, max, sum,count)
    grouping
    joins (inner, outer, left and right outer, natural)
How to delete, update, insert
+Transactions (committing, rollback)
Integrity constraints
Understand the function of null
+Privileges and Authorization grant graph
+*Python (connecting, looping through resultset, creating statement, using transactions)

Chapter 7

Design process/phases
+Definition of ER model
    entity set
    relationship set
    attributes (simple, composite,single/multivalued,derived)
+Constraints
    cardinalities (1-1,1-n,n-1,n-m)
    participation
    keys
+Weak entities
Roles
Binary vs n-ary relationship sets
+Generalization

Chapter 8

+Features of good relational design
+*Functional dependency
    keys
    closure
    trivial vs non-trivial
+*Lossy vs. lossless decomposition
+Atomic domains
+First Normal Form
+*Third Normal Form

+*BCNF
+Armstrong's axioms
+*Closure of Attribute Sets
+*Canonical Cover
    extraneous attributes
+*Dependency Preservation

Chapter 10

Know difference between types of storage (hierarchy)
    primary, secondary, volatile, tertiary
Magnetic disk and flash
Performance measures
    seek time, rotational latency time, average latency time, data transfer rate, MTTF
How to optimize disk block access
RAID
+*File organization
    file
    block
    fixed length records vs. variable length
+Organization of records
    heap, sequential, hashing
Data Dictionary storage
    what is it and how it is used
    metadata
Database Buffer
    buffer manager
    replacement policies

Chapter 11

+Basic concepts of indexing and hashing
    ordered vs hash-index
    how to evaluate
+*Ordered indices
    clustering vs. non-clustering
    index-sequential files
    dense vs spare indices
    multilevel indices
    updating and deleting
    secondary indices
    indexing multiple keys
+*B+ trees
    advantages vs. disadvantages
    know how to query, insert, update and delete (assume a single search key)
    difference between leaf and non-leaf nodes
    how to split and how to coalesce
+*Static Hashing

- hash function
- organization
- buckets and overflows
- indices

+*Dynamic Hashing
- know how to query, update and delete
- difference between static and dynamic

Differences between ordered indexing and hashing

Chapter 12

What is query processing
- steps
- query execution plan
- measure of cost

Chapter 14

+*Definition
- ACID
- transaction management

Storage
- volatile, nonvolatile, stable

Atomicity and Durability
- transaction model

Isolation
- concurrency

+*Serializability
- difference between serial and non-serial transactions
- what is a schedule
- what is serializable
- conflict serializable
- what is a conflict
- conflict equivalent
- precedence graph
- serializability order
- topological sorting

+*Isolation and Atomicity
- recoverable schedules
    - partial
    - dependent
- cascadeless schedules
    - cascading rollback

Transaction Isolation Levels

Transactions as sql statements
- *phantom phenomenon

Chapter 15

+What is concurrency control, why is it needed?

+*Lock-based protocols

What is lock (shared vs exclusive)
requests, grants, compatibility function, wait
lock compatibility matrix
deadlock
starvation
two phase locking protocol
lock manager
graph based protocol (tree)
database graph
commit dependency

+*Deadlocks
how to prevent, detect, recover
wait-die, would-wait
lock timeouts
detection
wait for graph with and without cycles
recovery
victim
rollback

*Multiple granularity
intention lock modes (S,X,IS, IX, SIX)
multiple-granularity locking protocol

+*Timestamp-based protocol
how to create ts
W-TS vs R-TS
TS ordering protocols
Thomas Write Rule

*Validation based protocols
phases
validation test
optimistic concurrency control

*Multiversion schemes
multiversion concurrency control
multiversion ts
multiversion two phase locking

*Snapshot isolation
first committer wins vs first updater wins
write skew

Chapter 19

What does distributed systems offer and when should you use
How is it different than a parallel system
*Difference between homogeneous and heterogeneous
*Distributed data storage
replication vs fragmentation
Transparency

*Distributed transactions
    +manager vs coordinator
Two phase commit protocol
    how to use in a distributed system
+*Heterogeneous distributed databases
    multidatabase system
    unified view of database
    query processing (mediator)
    local vs global transactions
*Cloud-based databases
    definition
        what is cloud computing
        what is a virtual machine
    storage

*Review slides related to NoSQL databases, be prepared for questions related to NoSQL databases