Solution to Midterm Exam (Fall 2018)
(60 minutes; closed-book and closed-notes)

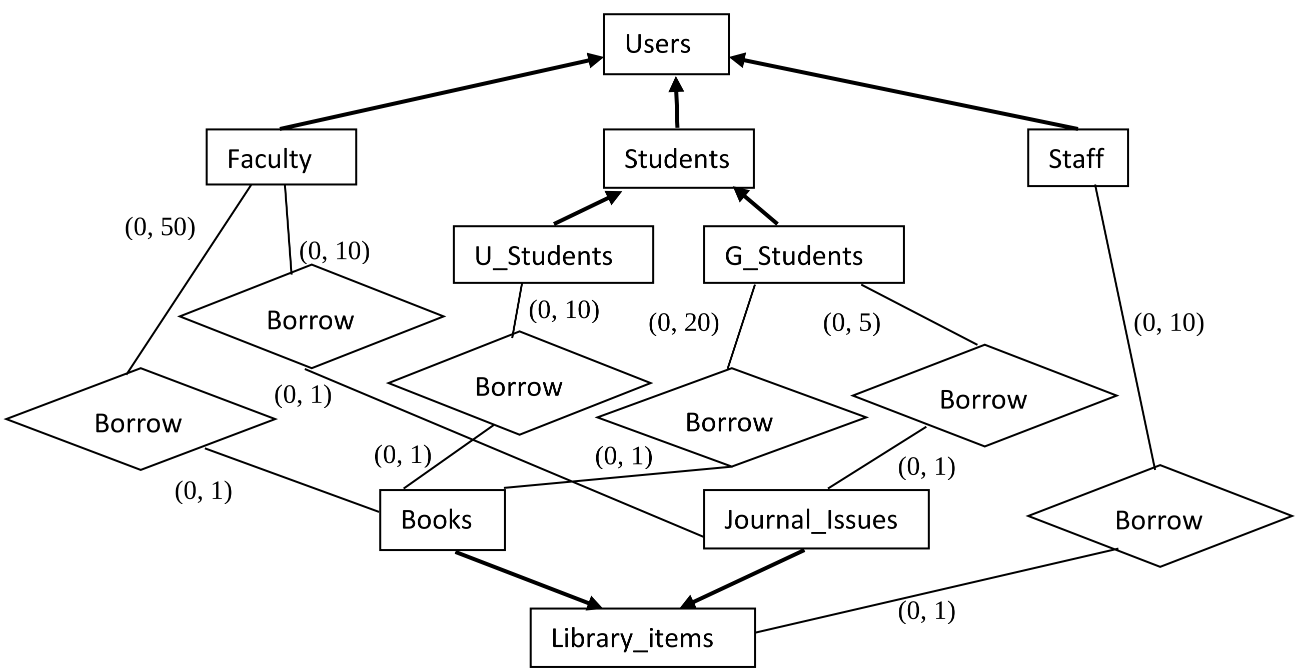
1. (8 points) Explain the 3-level database architecture. Specifically, what are the 3 levels (provide both the name of each level and a one-sentence description of what each level means) and why do we have them (very briefly)?

   Answer: The three levels are the conceptual schema in the middle, the internal schema at the bottom (closest to the physical data) and the external schemas at the top. The conceptual schema describes how the entire data in the database are conceptually organized, the internal schema describes how the data are stored as files, and the external schemas describe different portions of the database for different groups of users.

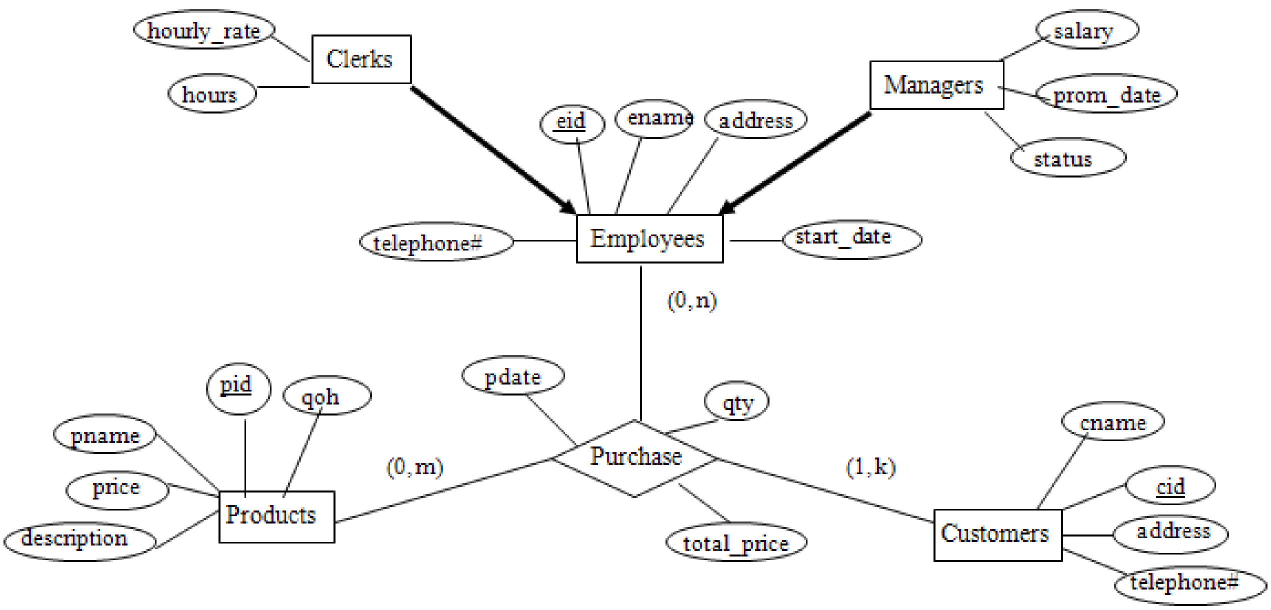
   The reason for having this architecture is to achieve a high degree of data independence (logical data independence and physical data independence) between the application programs and the data structure through the two mappings, one between the external schemas and the conceptual schema and the other between the conceptual schema and the internal schema.

2. (18 points) Consider a university library database. A user of the library is either a student, a faculty or a staff. The materials that can be borrowed are either books or issues of journals. Assume that each faculty can borrow up to 50 books and 10 issues of journals, each graduate student can borrow up to 20 books and 5 issues of journals, each undergraduate student can borrow up to 10 books but no journal issues, and each staff can borrow up to a total of 10 items, including both books and journal issues. Construct an E-R diagram for the database. (**No need to add attributes to entity sets)**.

Answer: It is acceptable if we do not have Students entity set. It is also acceptable if we do not have both Users and Students.



3. (16 points) Transform the following ER diagram into relations. Use Method 1 to transform the IS-A hierarchy in the diagram. Also, describe the key and foreign key(s) of each relation.



Answer: Products(pid, pname, price, description, qoh)
Customers(cid, cname, address, telephone#)
Employees(eid, ename, address, telephone#, start_date)
Clerks(eid, hours, hourly_rate)
Managers(eid, salary, status, prom_date)
Purchase(cid, pid, eid, pdate, qty, total price)

Clerks.eid, Managers.eid and Purchase.eid are foreign keys referencing Employees.eid
Purchase.cid is a foreign key referencing Customers.eid
Purchase.pid is a foreign key referencing Products.pid

4. (20 points) Answer **True** or **False** for each of the following independent questions based on relation schema R(A, B, C). No justification is needed.

(a) If AB together is a superkey of R, but neither A nor B is a superkey, then AB must be a candidate key. (True)

(b) If AB together is the primary key of R, then BC together cannot be a candidate key. (False)

(c) If ABC together is the primary key of R, then no tuple in R can have null value. (True)

(d) If R.C is a foreign key referencing S.D (where S is another relation schema), then the values under attribute D must be unique. (True)

(e) If R.C is a foreign key referencing S.D (where S is another relation schema), then the values under C must be unique. (False)

(f) If A → B is true (→ represents "functional determines", same below), then B cannot have more distinct values than A. (True)

(g) If B → AC is true, then B is a candidate key of R. (True)

(h) If there are no trivial functional dependencies on R, then R is in BCNF. (True)

(i) If A → C is true, then (AB, BC) is a lossless join decomposition of R. (False)

(j) If F = {A → B, C → B, AC → B}, then Fmin (the minimum cover of F) = F. (False)

5. (8 points) Prove the Decomposition Rule using Armstrong's Axioms (i.e., IR1-IR3).

Proof: This is to show that {X → YZ} |= {X → Y, X → Z}.

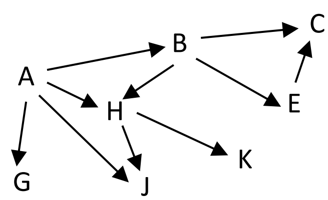
By (IR1): YZ → Y, YZ → Z;

By X → YZ (given), YZ → Y (above) and (IR3): X → Y;

By X → YZ (given), YZ → Z (above) and (IR3): X → Z.

6. Consider a relation schema R(A, B, C, E, G, H, J, K) and the functional dependencies F = {A → BGHJ, B → CEH, E → C, H → JK}.

a. (5 points) Find all candidate keys of R. Draw the dependency graph.
Answer: Dependency Graph:



It is clear Vni = {A} and A can determine all attributes. Therefore, A is the only candidate key.

b. (5 points) Is the schema in BCNF? You need to justify your answer.
Answer: It is not in BCNF. Many functional dependencies make R not in BCNF such as B → CEH, E → C and H → JK. These are all non-trivial FDs and the left hand side is not a superkey.

c. (5 points) Explain why this schema is not in 3NF.
Answer: Any of the FDs that make the schema not in BCNF also makes it not in 3NF as no attribute on the right of any of these FDs is a prime attribute. So we have non-prime non-trivially depends on a non-superkey.

d. (5 points) Find a minimal cover Fmin of F.
Answer: Fmin = {A → B, A → G, B → E, B → H, E → C, H → J, H → K}.

e. (5 points) Decompose R into 3NF schemas using Algorithm LLJ-DPD-3NF. Show the 3rd and the 4th steps of the algorithm (1st and 2nd steps are already done above).
Answer: Step 3: R1 = ABG, R2 = BEH, R3 = EC, R4 = HJK
Step 4: R1 contains the key A, no additional schema is generated.

f. (5 points) Identify the candidate key(s) and foreign key(s) in each schema obtained in step (e) above.
Answer: A is the candidate key for R1, B is the candidate key for R2, E is the candidate key for R3, and H is the candidate key for R4.
R1.B is a foreign key referencing R2.B, R2.E is a foreign key referencing R3.E, R2.H is a foreign key referencing R4.H.