1. (20 points) The following E-R diagram shows a database for the sale records from an online bookseller. If the information of a customer appears in this database, the customer must have ordered some book(s). It is similar that if the information of a book appears, that book must have been ordered by some customer(s).

a. Make some correction if there is anything wrong for the mapping cardinalities. (You can describe the error in the following space or directly make correction in the diagram):

b. Please use associated minimum and maxim cardinality to specify the number of times that the Books and Customers entities participate in the Orders relationship. If there is any total participation, please also specify it by using double lines in this diagram.

c. If all books can have both hardcover and paper-cover versions, which has different prices. Please extend this E-R diagram to model it.

d. If the bookseller allows each customer to pay his (her) balance month by month. Please add a payment entity and a necessary relationship to this diagram. Each payment has a payment number, payment-date and payment-amount. The payment numbers are associated to each customer and they are only distinct in the payment record for each customer.

Solutions:
2. (10 points) Convert the E-R diagram in Question 1 to a collection of relation tables. Please include the entities and relationships you added in the question 1.

Customer (C_id, name, address)
Papercovers(ISBN, price)
Hardcovers(ISBN, price)
Cust_Pays(C_id, payment_num, payment_amount, payment_date)
Orders(C_id, ISBN, Order_date, quantity)

3. (25 points) Consider the following relations

Employee(eid: Integer, name: String, salary: real)
Works(eid: Integer, dname: String, title: String)
Department(dname: String, location: String, managerid: Integer)

Note: A manager is also an employee.

a. Define the Department relation only in SQL. Please specify every department is guaranteed to have a manager. Please also specify both the primary key and foreign key(s) (if it has).

Answer:
Create table Department(
    dname char(20) not null,
    location char(20) not null,
    managerid integer not null,
    primary key dname,
    foreign key (managerid) references employee)

b. Write SQL and relation algebra expressions for the following Query:

“Find the name of the department, such that the average salary of its employees is the highest among all departments.”

Answer-SQL:
(One possible answer)
With avg-dep-salary(dname, avg-salary) as
    Select dname, avg(salary) 
    From Employee, Works 
    Where Employee.eid = Works.eid 
    Group by dname 
Select dname 
From avg-dep-salary 
Where avg-salary in (Select max(avg-salary) From avg-dep-salary)
(Another possible answer)
Select dname
From Works, Employee
Where Works.eid = Employee.eid
Group by dname
Having avg(salary) >= all(Select avg(salary)
    From Works, Employee
    Where Works.eid = Employee.eid
    Group by dname)

Answer-Algebra:
\[\text{temp1} \leftarrow \text{dname} \sigma_{\text{avg}}(\text{salary}) \text{as avg.salary} \text{from Works, Employee} \]
\[\text{temp2} \leftarrow \text{G}_{\text{max}}(\text{avg.salary}) \text{as max.avg.salary} \text{from temp1} \]
\[\Pi_{\text{dname}} (\sigma_{\text{temp1.avg.salary=\text{temp2.max.avg.salary}}} (\text{temp1} \times \text{temp2}))\]

4. (20 points) Consider the same relations in question 3 and the following datalog program.

\[\text{Query } (x, y) : \text{P} (x, y) \]
\[\text{P} (x, y) : \text{Employee} (x, y, z), z > 50000.00, \text{Works}(x, d, t), \text{Department}(d, l, m), \text{Employee}(m, \text{“John Smith”}, s)\]

Please write down the SQL and relational algebra expression for this query. (You can get half points if you only can write down this query by using natural language, i.e. English.)

Answer-Natural Language:
Find the eid and name for those employees who earn more than $50000 and work in the department where John Smith is the manager.

Answer-SQL
Select Employee.eid, Employee.name
From Employee, Works,
Where salary > 50000 and Employee.eid = Works.eid and
Works.dname in (Select dname From Department
    Where managerid in (Select eid from Employee
        Where name = “John Smith”))

Answer-Algebra
\[\text{temp} \leftarrow \Pi_{\text{dname}} (\sigma_{\text{Employee.name=}\text{“John Smith”} \cdot \text{Department.managerid=}\text{Employee.eid}} (\text{Employee} \bowtie \text{Department}))\]
\[\Pi_{\text{eid.name}} (\sigma_{\text{salary}>50000} (\text{Employee} \bowtie \text{Works} \bowtie \text{temp}))\]
5. (25 points) Consider a relation $R(A,B,C,D)$ and a set of functional dependencies that hold for $R$: \{AB $\rightarrow$ C, C $\rightarrow$ A, C $\rightarrow$ D\}.

a. Identify the candidate key(s) for $R$.

$AB$, $BC$

b. Show $R$ is not in BCNF

$C \rightarrow D$ is not trivial, and $C$ is not a superkey.

c. Give a lossless-join decomposition into BCNF of $R$.

Use the algorithm described in the book for lossless-join BCNF decomposition.

1. Consider $C \rightarrow D$, we can decompose $R$ to $R_1=(C,D)$ and $R_2=(A,B,C)$.

2. Then consider $R_2(A,B,C)$. $C \rightarrow A$ violates BCNF. We could decompose $R_2$ into $(C,A)$ and $(C,B)$.

So one solution is $(C,D)$, $(C,A)$ and $(C,B)$.

The other solution is $(A,C,D)$ and $(B,C)$. 