CSE 462 Homework #4: Database Design

Name: _____ Date: March 28, 2011

***** Due on Due on April 11, 2011 at the beginning of class. *****

Instructions. This problem set is worth 300 points.

1. (100pts) Consider the ER model for an online store shown in figure 1 and provide short answers for the questions below. You must justify your answers based on the ER model.

1. Can there exist customers without any contact? (2 pts)

   Yes, Customer participation in Reaches is optional.

2. Can there exist a contact that is neither an e-mail account, nor a phone number, nor an address? (2 pts)

   No, Contact has total (disjoint) participation in the is-a relationship, so it must be one of e-mail, number, or address.

3. Can a customer have multiple e-mail accounts, phone numbers, and addresses? (2 pts)

   Yes, Customer participation in Reaches is one-to-many.

4. Can an order be placed by multiple customers? (2 pts)

   No, Order participation in Places is many-to-one.

5. Can an order be shipped to multiple addresses? (2 pts)

   No, Order participation in Ships is many-to-one.

6. Must the address to which an order is shipped be associated with the customer who places the order? (2 pts)

   No, there is no relationship that associates the customer (C1) of a shipping address with the customer (C2) who places an order.

7. Can an order have products from different sellers? (2 pts)

   Yes, an Order may have multiple OrderEntry entities, each of which associated with a different seller.

8. Can there exist orders without at least one order item? (2 pts)

   No, Order has total participation in PartOf.

9. Can there exist an order with multiple entries of the same product? (2 pts)

   Yes, but they must be sold by different sellers.

10. Are customers allowed to review products which they did not purchase? (2 pts)

    Yes, a Review associates a Product and a Customer, not a product purchased by the customer.

Construct a relational schema for this ER model. Break your work into phases:

1. Create the initial schema for the ER model by mapping entity sets and their attributes, relationships, and defining all foreign keys. (40pts)

Customer(cid, lastName, firstName, dob)

Contact(cid, label)

Email(cid, label, email)

Phone(cid, label, phone)

Address(cid, label, street, city, state, zip)

Product(sku, name)

UnitPrice(sku, date, price)

(UnitPrice.sku) references (Product.sku)

Seller(sid, name)

Order(oid, date, shipDate)

OrderEntry(oid, sid, sku, qty)

(OrderEntry.oid) references (Order.oid)

(OrderEntry.sid) references (Seller.sid)

(OrderEntry.sku) references (Product.sku)

Places(oid, cid)

(Places.oid) references (Order.oid)

(Places.cid) references (Customer.cid)

Ships(oid, cid, label, date, type)

(Ships.oid) references (Order.oid)

(Ships.cid, Ships.label) references (Address.cid, Address.label)

Reviews(cid, sku, date, stars, comments)

(Reviews.cid) references (Customer.cid)

(Reviews.sku) references (Product.sku)

Rates(cid, sid, date, stars, comments)

(Rates.cid) references (Customer.cid)

(Rates.sid) references (Seller.sid)

2. Create an optimized schema for the ER model by applying all possible optimizations to the schema obtained above. (20pts)

Customer(cid, lastName, firstName, dob)

Contact(cid, label)

Email(cid, label, email)

Phone(cid, label, phone)

Address(cid, label, street, city, state, zip)

Product(sku, name)

UnitPrice(sku, date, price)

(UnitPrice.sku) references (Product.sku)

Seller(sid, name)

Order(oid, date, shipDate, places_cid, ship_cid, ship_label, ship_date, ship_type)

(Order.places_cid) references (Customer.cid)

(Order.ships_cid, Order.ships_label) references (Address.cid, Address.label)

OrderEntry(oid, sid, sku, qty)

(OrderEntry.oid) references (Order.oid)

(OrderEntry.sid) references (Seller.sid)

(OrderEntry.sku) references (Product.sku)

Reviews(cid, sku, date, stars, comments)

(Reviews.cid) references (Customer.cid)

(Reviews.sku) references (Product.sku)

Rates(cid, sid, date, stars, comments)

(Rates.cid) references (Customer.cid)

(Rates.sid) references (Seller.sid)

3. Provide SQL DDL commands to create the optimized schema above in a relational database. Choose reasonable types for your attributes, include relevant `CHECK` constraints, and enforce total participation where necessary. (10pts)

straightforward

4. Entity sets *Order* and *Order Entry* have derived attributes. Create views `vOrder` and `vOrderEntry` that return: the attributes of the corresponding *tables* as well as the respective derived attribute (*subtotal* for `vOrderEntry` and *total* for `vOrder`). *Hint:* define view `vOrderEntry` first and use it in the definition of view `vOrder`. As usual, you may create any additional views to simplify your job. (10pts)

straightforward

2. (50pts) Answer problem 4.4.4 (p156).

a) Course is a weak entity set, number is its discriminator. There exists an identifying relationship between Course and Department, so that the department name and the course number uniquely identify a course.

b) Team is a weak entity set, name is its discriminator. There exists an identifying relationship between Team and League, so that the team name and the league name uniquely identify a team. Player is a weak entity set, number its discriminator. There exists an identifying relationship between Player and Team. Because Team is itself a weak entity set, the player number, the team name, and the league name must be used to uniquely identify a player.

3. Extra-Credit (50pts) Answer problem 4.2.5 (p147).

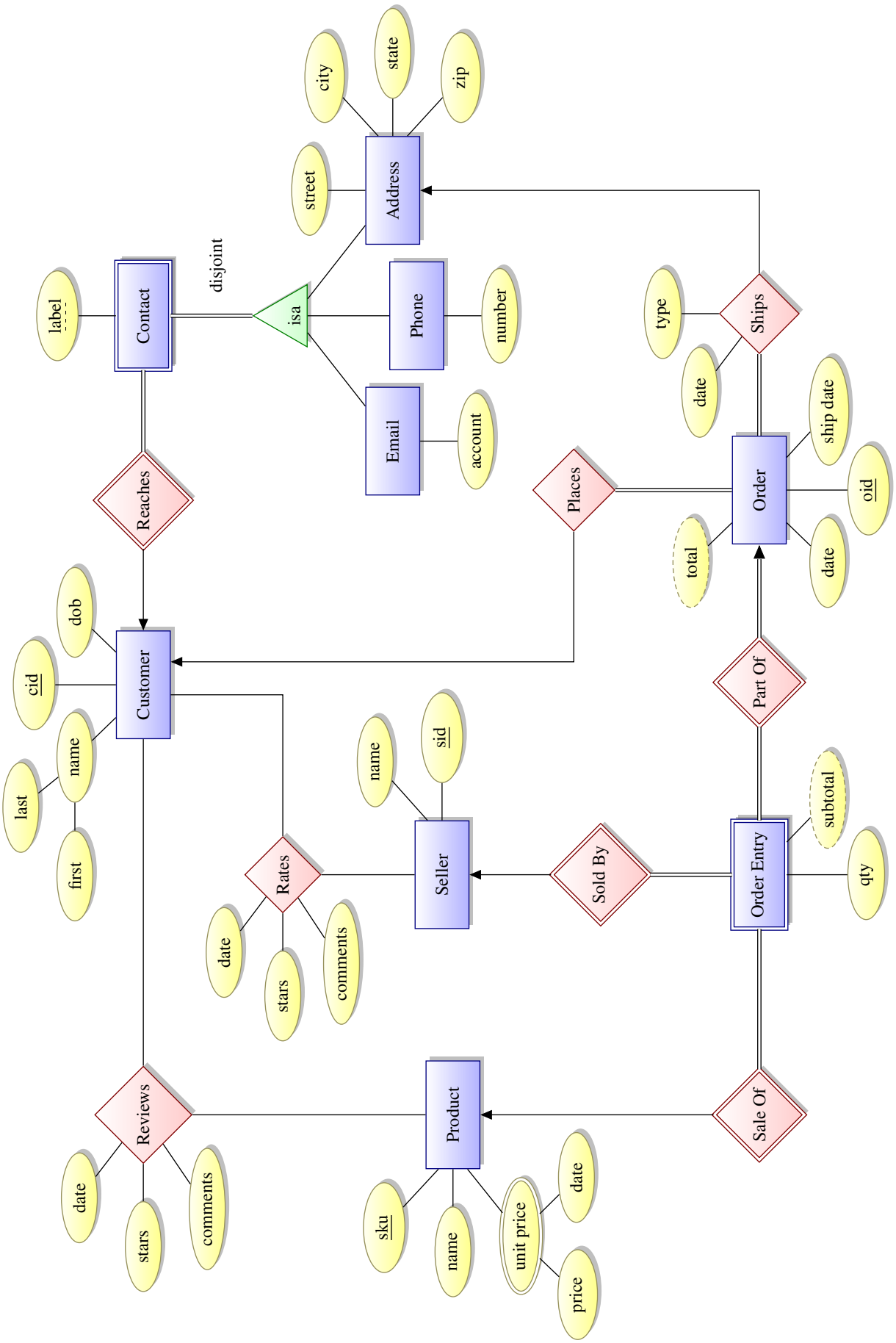

Figure 1: ER Model of an online store represented diagrammatically.

4. (100pts) Consider the relation schema $R(A, B, C, D, E, F)$, the set of FDs $\mathcal{F} = \{A \rightarrow BC, BD \rightarrow A, C \rightarrow AD, EF \rightarrow D\}$ that hold in R, and a decomposition $\{R_1(A, B, C, D), R_2(B, E, F)\}$ of R. Then, answer the questions below.

a) Compute the key(s) of R. Show *all necessary* attribute set closures to justify your answer. Then, show two distinct FDs, the first one only violates BCNF and the second one violates 3NF. (20pts)

b) Project the set of functional dependencies $\mathcal{F}$ (algorithm 3.12, p82) onto the attributes of R_1 and R_2 to obtain the sets of functional dependencies $\mathcal{F}_1$ that hold in R_1 and $\mathcal{F}_2$ that hold in R_2. (20pts)

c) Show that the decomposition of R into R_1 and R_2 is not dependency-preserving by providing instances of R_1 and R_2 and showing the dependency violation in the result of the natural join of R_1 and R_2. *Important:* the instances of R_1 and R_2 you provide must satisfy the projected FDs computed in (b). (20pts)

d) Using the chase, determine whether the decomposition has the lossless join property. You *must* show the initial tableau setup, the FD applied at each iteration, and the state of the tableau after each iteration (just as we did in class). Conclude by saying whether the decomposition has the lossless join property, based on the result of the chase. (20pts)

e) Decompose R into 3NF using the synthesis algorithm (algorithm 3.26, p103). Make sure you show each step of the algorithm. You may use the results obtained in (a)–(d) if those will help in the computation. (20pts)

5. (50pts) Provide relation instances to show that the following rules involving MVDs *do not hold*. You must apply a step-by-step chase for each item, just as we did in the classroom. Based on the result of the chase, provide the requested conterexample relation instances. (cf. problem 3.7.4, p120).

1. If $A \twoheadrightarrow BC$, then $A \twoheadrightarrow B$. (15 pts)
2. If $A \twoheadrightarrow B$, then $A \rightarrow B$. (20 pts)
3. If $AB \twoheadrightarrow C$, then $A \twoheadrightarrow C$. (15 pts)