UNIVERSITY OF CALIFORNIA
Department of EECS, Computer Science Division

CS186 Midterm I
Oct 5, 1999

NOTE: THIS IS THE EXAM THAT WAS GIVEN IN THE FALL 1999 SEMESTER. IT IS PROVIDED FOR YOUR STUDYING PURPOSES --- WE EXPECT THAT OUR EXAM WILL BE SIMILAR IN STYLE, BUT WE MAKE NO GUARANTEES IN THAT REGARD. USE THIS SIMPLY A STUDY AID.

Midterm Exam: Introduction to Database Systems

This exam has five problems. Each problem is made up of multiple questions. You should read through the exam quickly and plan your time-management accordingly. Before beginning to answer a question, be sure to read it carefully and to *answer all parts of every question!*

REFERENCE DATABASE . This is the Reference Database referred to in this exam.

There are four tables. SALESPERSON contains the names, ids, regions & quotas for the salespeople. Ids are unique, names are not, and there is only one salesperson per region. PRODUCTS contains the product names, product ids, and prices for the products. The product ids are unique.

SALESPERSON

Sname	Sid	Region	Quota
Frances	25	TX	$100
Bob	31	CA	$150
Frances	74	MA	$200
Mary	89	FL	$250

PRODUCTS

Pname	Pid	Pprice
disks	131	$100
pcs	152	$700
macs	831	$800
printers	255	$120
paper	221	$5

CUSTOMERS contains the customer names, customer ids, and regions for the customers (customer ids are unique), and ORDERS contains the customer id, the product id, and the product ordered per customer.

CUSTOMERS

Cname	Cid	Region
Bob	1	TX
Harry	2	TX
Lin	3	MA
Martha	4	FL
Lin	5	FL
Leyla	6	CA

ORDERS

Cid	Pid	Quantity
1	152	1
2	152	1
4	831	1
4	131	1
5	255	1
6	831	1

Name_____

Section You Usually Attend _____**Login**_____

1. Relational Model and ER Model (20 points).

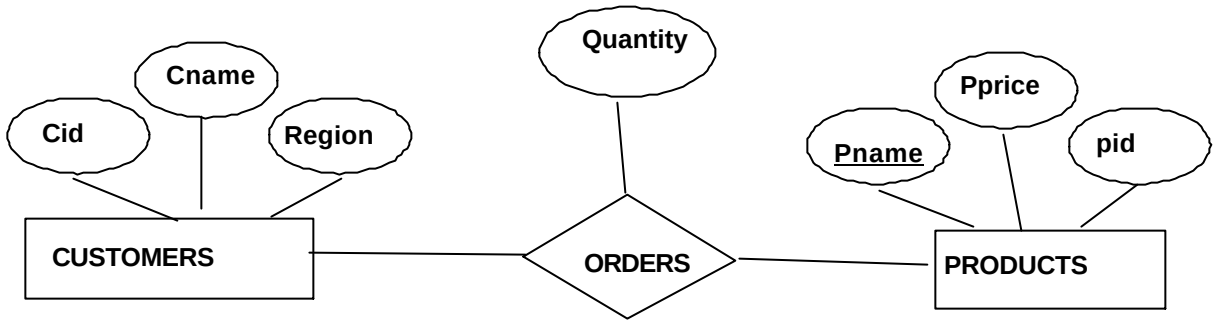
a. (2 points) Using the Reference Database as an example, give an example of data independence contrasting the use of a file system to store the information in the SALESPERSON table with the use of a relational database to store the same information.

b. (2 points) What is the cardinality of the table SALESPERSON in the Reference Database on page 1? What is the degree of the table SALESPERSON?

c. (4 points). Name the primary keys of each of the four tables in the Reference Database on Page 1.

d. (4 points) For each table in the Reference Database on Page 1, determine whether or not it has a foreign key, and if so, list what it is *and* what it references.

e. (8 points) Modify the following ER diagram to represent the Reference Database on Page 1, including the salesperson entities and relationships.



Name_____

2. Disks and Buffers (20 points)

a. [5 points] Consider a disk with a sector size of 512 bytes, 100 sectors per track.
Given a rotational speed of 7200 revolutions per second, what is the maximum rotational delay to the start of a sector?
Assuming that one track of data can be transferred per revolution, what is the transfer rate?

b. [5 points] In the organization of a data page, what is a slot directory? What is it used for? In the organization of a data record, what is a field offset? What is it used for?

For questions c & d, below, assume that in the CUSTOMERS table, records are much larger than shown in the table on the page – in fact, each record fits in exactly one page, so the CUSTOMERS table is 5 pages long. And also assume that there is room for only two buffers – that is, there are two buffer frames – in memory, and that these buffers start out empty.

c. [5 points] Is LRU or MRU a better strategy if we are reading the CUSTOMERS records in the order: Cid = 1, 1, 3, 4, 3, 1, 2. Explain your answer.

d. [5 points] Assume there are two processes that alternate references to the buffer pool. First process A references the buffer pool, then Process B, then A and so on. Process A keeps re-reading the record with Cid = 2, and Process B sequentially scans the entire CUSTOMERS table each time it gets a turn. Assume that the two processes must share the same two-buffer pool. Which algorithm (LRU, MRU or CLOCK) minimizes the number of reads? Explain why.

Name_____

3. **Indices** [20 points]

b. [5 points] In the Reference Database on Page 1, name one advantage of building a clustered index on the Sname field of the SALESPERSON table. Name one disadvantage. Under what circumstances would a nonclustered index on the Sname field be the best choice?

b. [5 points] In the PRODUCTS table in the reference database on page 1, **assume that there are actually many more records in the PRODUCTS table than are listed in the table shown** – assume that there are P pages with R records per page. Given an Extendible Hashing index on the Pid attribute, what is the number of I/Os required (worst case) to fetch all Pids between 131and 150, including 131 & 150? You should assume that bucket pages are internally organized like heap file pages. You may assume any alternative you wish for storing data entries in the index, and for buffering data in the system, but *you must state the alternative and the buffering assumptions,* and present the correct analysis *for those assumptions.*

c. [5 points] In the reference database on Page 1, assume that the CUSTOMERS records are much larger than shown in the table on the page – in fact, each record fits in exactly one page, so the CUSTOMERS table is 5 pages long.. Draw a B+ tree index on Cname, assuming that d = 2, using any of the alternatives discussed in text, but *explain which alternative you are using.*

d. [5 points] The DBA for the reference database on Page 1 has created a sparse index on the PRODUCTS table (again, assuming that the PRODUCTS table has many more records than are listed on Page 1). The sparse index is on Pname. Explain in terms of accesses to the table PRODUCTS the advantages of a sparse index on Pname. Explain the disadvantages.

Name_____

4. Relational Algebra (16 points)

A. (10 points; 2 points each) Show the rows resulting from the following queries, using the reference database.

1. $\sigma_{\text{Cname = "Lin"}}(\text{CUSTOMERS})$

2. $\pi_{\text{Cname, Pid}}(\text{CUSTOMERS} \bowtie_{\text{CUSTOMERS.Cid = ORDERS.Cid}} \text{ORDERS})$

3. $\pi_{\text{Sname}}((\text{SALESPERSON} \bowtie_{\text{SALESPERSON.Region = CUSTOMERS.Region}} \text{CUSTOMERS}) \bowtie_{\text{CUSTOMERS.Cid = ORDERS.Cid}} (\sigma_{\text{Pname = "Mac"}}(\text{PRODUCTS} \bowtie_{\text{PRODUCTS.Pid = ORDERS.Pid}} \text{ORDERS})))$

4. $\pi_{\text{sname}}\text{SALESPERSON} \cap \pi_{\text{Sname}}\text{CUSTOMERS}$

5. $\pi_{\text{sname}}\text{SALESPERSON} \cup \pi_{\text{Sname}}\text{CUSTOMERS}$

B. (6 points; 2 points each) State the relational algebra expression to find the following:

1. The names of the customers who live in the region TX.

2. The customer id's of customers who have ordered a pc.

3. The Sids of the Salespeople who have sold a printer.

Name_____

5. **SQL** [24 points]

Use the tables in the Reference Database, on Page 1, write the SQL queries to do the following.

a. [3 points] List the names of the customers who have bought more than one item.

e. [3 points] List the names, Pid and price of all the Products, whether or not the product has been ordered, but if it has been ordered display the Cids of the customer who ordered it.

f. [3 points] List the names and Cids of all customers, and, for those who have bought Products, the total amount of money they owe for their purchases.

g. [3 points] Find the Sids of the salespeople for the regions where there are the most customers.

Using the tables in the Reference Database on Page 1, write down the resulting rows for the following queries.

h. [2 points] SELECT DISTINCT Sname FROM SALESPERSON

i. [2 points] SELECT S.Sid, C.Cname from SALESPERSON S, CUSTOMERS C
WHERE S.Region = C.Region
AND C.Region = "MA"

j. [4 points] SELECT S.Sid, Tsales = Sum (P.Pprice)
from SALESPERSON S, PRODUCTS P, ORDERS O. CUSTOMERS C
WHERE S.Region = C.Region AND C.Cid = O.Cid and P.Pid = O.Pid
Group by S.Sid

k. [4 points] SELECT P.Pname from PRODUCTS P
WHERE NOT EXISTS
(SELECT O.Pid FROM ORDERS O
WHERE O.Pid = P.Pid)