MATHEMATICS 277

Time: 2 Hrs. **Final Examination** **Fall 2002-2003**

Family Name	
First Name	
Id. Num	

Write down your answers on the question sheet. You can use the reverse side for scratch work. **DO NOT CUT OFF ANY PAGE.**

Problem	Grade	Out Of
1		10
2		20
3		20
4		15
5		15
6		10
7		15
8		15
Total		**120**

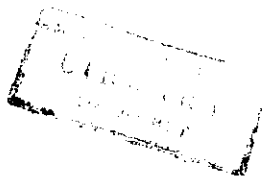
Problem 1 (10 Points)

The relational database design schema given below is for Zakaria movie theater.

MOVIE (TITLE, YEAR, STUDIONAME, PRODUCERC#, INCOLOR)
STAR (STARNAME, SEX, BIRTHYEAR)
STARS_IN (TITLE, STARNAME)
EXECUTIVE (CERT#, NAME, ADDRESS)
STUDIO (STUDIONAME, CITY, COUNTRY, PRESIDENTC#)

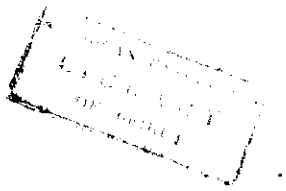
The database is supposed to keep track of the different movies, stars, studios that produce these movies, and the executives who either direct these movies, or are presidents of studios. A movie is assumed to be identified by its title, and for each movie in the database is kept its year of production, its director, the studio it was filmed at, and whether it is in color or not (INCOLOR being "yes" or "no"). A star is identified by the name, and the database maintains information about the gender and the year of birth. An executive here is either a studio president or a movie director or others. Every executive goes through a certification procedure run by some high authority. Once certified, the executive gets a certificate number, CERT#, which is assumed to identify the executive. For a studio, the database keeps track of its name, which identifies the studio, the city and country, and the presidents certificate number. The primary key for STARS_IN is the combination of its two attributes.

a. Give an appropriate ER schema diagram for the given relational schema, including cardinality and participation constraints.



Problem 2. (20 Points)

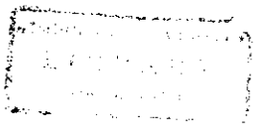
a. List the set of referential integrity constraints that are present in the given relational design.



b. Give appropriate SQL statements that create the 3 relations MOVIE, STAR, and STARS_IN as given. Include key and referential integrity constraints.

c. Based on the referential integrity constraints, what should a database management system do if a INSERT INTO STARS_IN statement is to be executed. Briefly explain.

d. Based on the referential integrity constraints, what should a database management system do if a DELETE STUDIO statement is to be executed. Briefly explain.



Problem 3.(20 Points)
This problem also refers to the relational design of Problem 1.

a. Suppose that there are no colored movies prior to year 1945, and Zakaria wants to enforce this "business rule" (integrity constraint) in the database. Suggest a scheme that can be implemented in SQL to enforce this constraint. Give the appropriate SQL statements.

b. Can such a business rule be expressed in the ER diagram that you gave earlier?

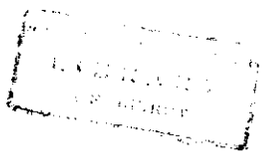
c. Suppose that every movie must have a female star and a male star. Can such a constraint be enforced in SQL ? How ? Give the appropriate SQL statements.

d. Can the constraint in (c) be modeled in the ER diagram? Give a briefly justified answer.

Problem 4. (15 Points)

With reference to the relational design of Problem1, write the following in SQL:

a. Find all titles of movies that were developed at Universal Studios during 1995.



b. Find all names and birth-years of those stars who have starred in all movies filmed at universal studios during the year 1995

c. Find the average age of the stars in the movie "Thunderstorm"

Problem 5. (15 Points)

a. Suppose that r is a relation on the schema $R(A, B, C)$. Write an SQL query to test whether the functional dependency $A \rightarrow C$ holds on r. Assume that no null values are present Briefly explain your strategy.

For the rest of the problem, consider the scheme
STUDIUO (STUDIONAME, CITY, COUNTRY, PRESIDENTC#)

b. Suppose that in the scheme for STUDIO above, the following functional dependency hold : CITY → COUNTRY. Is the scheme BCNF ? Why ?

c. Suggest a decomposition of the scheme that will result in a BCNF schema.

d. Is your decomposition lossless join ? Why ?

e. Is your decomposition dependency preserving ? Why ?

Problem 6. (10 Points)

Suppose that the data contained in the relation

MOVIE (TITLE, YEAR, STUDIONAME, PRODUCERC#, INCOLOR)

of the relational design of Problem 1 above is to be stored as follows: **all records corresponding to movies of the same year will be stored in a bucket**. (By a record here is meant a tuple of MOVIE) Suppose that there are at most 100 movies per year. Also assume the following storage requirements: 30-bytes for TITLE, 2-bytes for YEAR, 30-bytes for STUDIONAME, 4-bytes for PRODUCER#, and 2-byte for INCOLOR. Assume that the block size is 4096 bytes. No spanning is allowed (i.e. one record must be stored in one block, and cannot cross over two blocks). Answer the following showing your computations:

a. How many blocks of data at most would a bucket require?

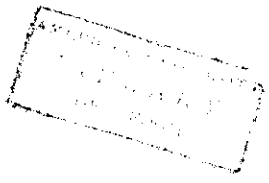
b. What will be the total number of blocks needed to store the relation MOVIE?

Problem 7. (15 Points)

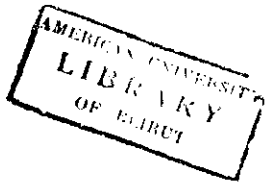
Suppose that a B^+-tree index of degree ***n*** with search-key field YEAR is to be built for the data buckets containing MOVIE as described above. Assume that a pointer to a block requires 6-bytes of storage. Answer the following, explaining and **showing your computations**:

a. What is the appropriate value for *n*, the degree of the B^+-tree?

b. What will the height of the B^+-tree be, and how many blocks will the tree occupy?



c. Explain how the index structure can be used to perform query (a) of Problem 4 above.



d. How many block transfers are needed to perform query (b) of Problem 4 above using the index structure?

Problem 8. (15 Points)

a. Suppose that extendable hashing is used to store a file that contains records with the following search-key values: 14, 15, 18, 28. Show the extendable hash structure, including the hash table, for this file if the hash function is $h(x) = x \bmod 16$ and buckets can hold 2 values each. (Write down the binary representations of the values $h(x)$.)

b. Ideally, the file should fit in 2 blocks (since there are 4 values). Is this happening? What is the percentage of actually used/reserved blocks in (a) to the ideal number of blocks?

c. What is your conclusion from your answers in (a) and (b) ?

d. Can the hashing scheme above useful for queries that search for records whose search key value is less than 25 ? Why ?