Fundamentals of Data Engineering

(COE848)

Sample Final Exam

This is a CLOSED BOOK exam. Textbooks, notes, laptops, calculators, personal digital assistants, cell phones, and Internet access are NOT allowed.

Please read each question carefully, and write your answers legibly in the space provided. You may do the questions in any order you wish, but please

USE YOUR TIME WISELY.

When you are finished, please hand in your exam paper and make sure you are **signed out**.

Good luck!

Student Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Student ID: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Score: \_\_\_\_\_ %

|  |  |  |
| --- | --- | --- |
| Question | Maximum Mark | Received |
| 1 | 15 |  |
| 2 | 5 |  |
| 3 | 16 |  |
| 4 | 14 |  |
| 5 | 15 |  |
| 6 | 15 |  |
| Total | 80 |  |

Question 1 (15 Marks): Explain the following in 2-3 sentences:

1. Transitive Dependency
2. Surrogate Key
3. 2NF

Question 2 (5 Marks): Within the context of database normalization, there is a rule called *pseudo-transitivity*, which is expressed as follows: if A🡪B, BC🡪D then AC🡪D.

Based on the functional dependency rules that you already learnt in class this term, show how *pseudo-transitivity* can be derived.

Question 3 (16 Marks): Consider the relation schema S (A, B, C, D) and the following functional dependencies on S:

A🡪BCD, B🡪C, CD🡪A

For each of the following short questions, be sure to briefly explain your answer.

1. S is not in BCNF, but is it in 3NF? Explain your answer.
2. Consider the decomposition of S into S1 (A, B, C) and S2 (B, C, D). Is this a valid decomposition into BCNF? Explain.
3. Consider the decomposition of S into S3 (A, B, D) and S4 (B, C). Is this a valid decomposition into BCNF? Explain.
4. Assume the decomposition in part c is implemented and the user is allowed to enter data into tables S3 and S4. Is it guaranteed that all the initial functional dependencies on S (A🡪BCD, B🡪C, CD🡪A) will be respected in S3 and S4?

Question 4 (14 Marks): Assume you are in charge of managing the program committee for an important conference. The following database stores information about papers submitted to the conference (Paper table), reviewers on the program committee (Reviewer table), and the assignment of reviewers to papers (Reviews table). Each reviewer on the program committee will have to review a set of papers. Each paper will be reviewed by some subset of reviewers.

Paper (pid, title)

Reviewer (rid, name)

Reviews (rid, pid)

a) Write a SQL query that finds all papers with fewer than three reviewers assigned to them. The output of the query should be a list of paper titles. The result should also include papers without any reviewers assigned to them.

b) Write a SQL query that finds the reviewers with the most papers assigned to them. There can be more than one such reviewer. The output of the query should be a list of reviewer names. A reviewer should be listed if no other reviewer has strictly more papers to review.

Question 5 (Mark 15): Refer to the following schema design and answer the queries using SQL queries:

Country (CName, Continent, GDP, Population)

Politician (Pname, Gender, Age, Office, Country, Region)

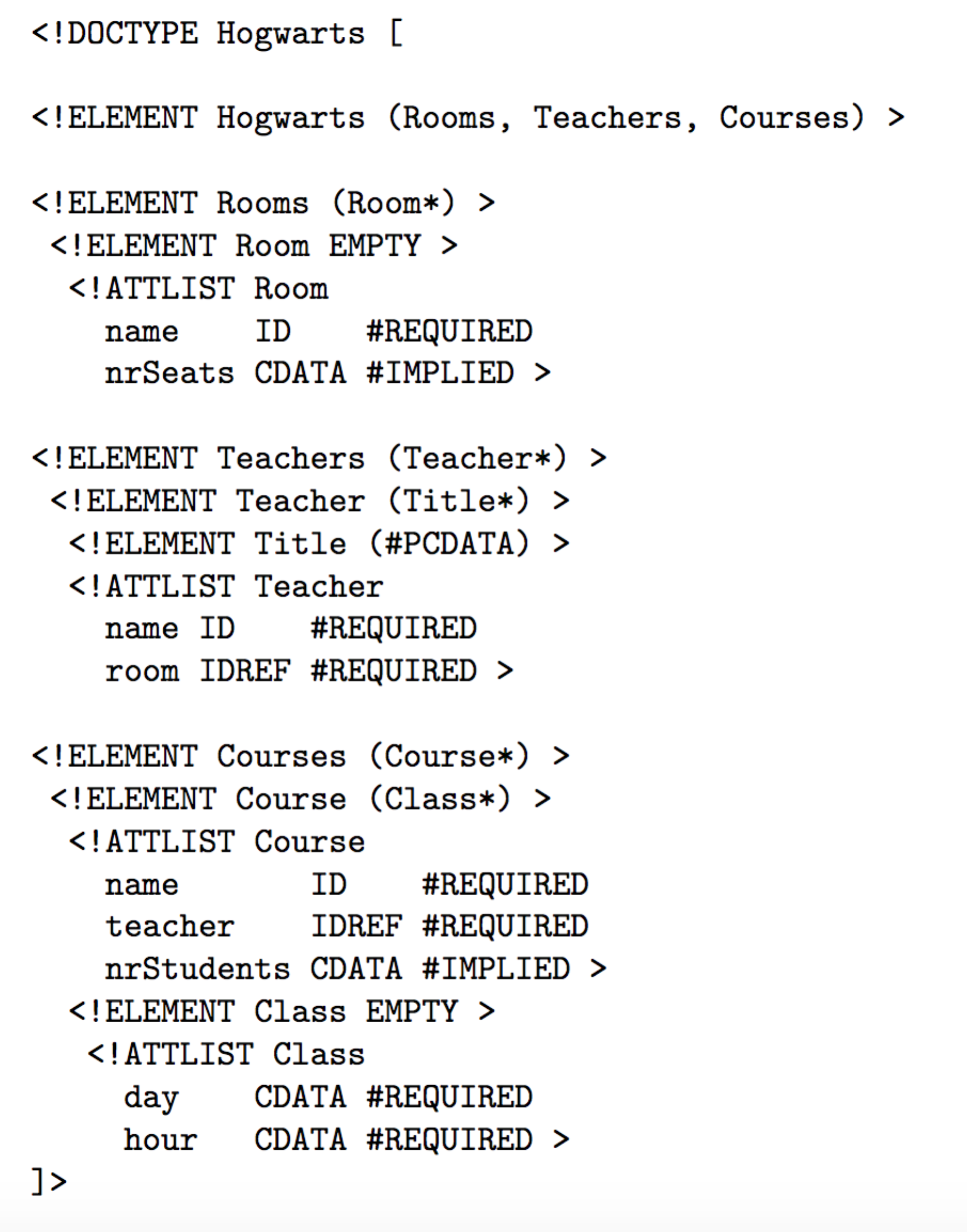
Head\_Of\_Government (Country, PName)

Alliance\_Membership (Country, Alliance)

a) List the countries in the database where all of the senators (office=senator) are women.

b) List the countries in the database that are in NATO (Alliance=Nato) but do not have a female head of government.

Question 6 (15 Marks): Consider the following DTD Document.



a) Write an XPath expression to find all courses that have at least 20 students

b) Write an XPath expression to list all professors at the school

c) Find all rooms that are used on Mondays