THIS PAPER IS NOT TO BE REMOVED FROM THE EXAMINATION HALLS

UNIVERSITY OF LONDON

CO2209 ZA

(291 0209 ZA)

BSc Examination

COMPUTING AND INFORMATION SYSTEMS AND CREATIVE COMPUTING

Database Systems

Tuesday 15 May 2012 : 10.00 – 1.00 pm

Duration: 3 hours

There are 5 questions in this paper. Candidates should answer **4** questions. All questions carry equal marks, and full marks can be obtained for complete answers to a total of **4** questions. The marks for each part of a question are indicated at the end of the part in [.] brackets.

Only your first 4 answers, in the order that they appear in your answer book, will be marked.

There are 100 marks available on this paper.

Calculators may be used. Electronic calculators must not be programmed prior to the examination. Calculators which display graphics, text or algebraic equations are not allowed. The make and type of machine must be stated clearly on the front cover of the answer book

Question 1 ER and relational Models

The following, an extract from a relation for a (fictitious) classical music shop, describes CDs that they stock.

CD code (ISRC)	Title	Year	Price	Track	Composer	Nationality	Work	Date of composition	Movement	Key Signature
01142223	Great Symphonies	2003	£4.99	1	Beethoven	German	Symphony No 9	1827	i	D maj
01142223	Great Symphonies	2003	£4.99	2	Beethoven	German	Symphony No 9	1827	ii	D maj
01142224	Great Symphonies	2004	£4.99	1	Brahms	Austrian	Symphony No 1	1854	i	D maj
02342376	Acis & Galatea	2002	£25.99	1	Handel	German	Acis & Galatea	1723	Overture	D maj
11124283	Beethoven Complete Works	1999	£55.99	1	Beethoven	German	Symphony No 9	1827	i	D maj
...	...	...	...	...	...	...	...	...	...	...

CDs have a unique ID value (the ISRC), a title, date and price. Each track corresponds to a movement of a classical work. Each work has a date and a single composer – whose nationality is stored – and each of its movements has a key signature. Work names are unique to composers, whilst a given movement name occurs once only in a given work.

(a) i. What sort of anomaly is risked in modifying this table in the event of CD 01142223 being increased in price? Why? [2]

ii. This table is susceptible to other anomalies with regard to update operations. Give **two**, in each case naming the anomaly and giving an example. [4]

(b) i. A friend thinks that this table is in 1NF. What do they mean? Are they correct? [2]

ii. Identify the functional dependencies in this table. [4]

iii. Give an example of a transitive dependency from your answers. [1]

iv. Revise this schema to be in BNCF, giving new tables and specifying their keys. [6]

v. Provide the SQL table definitions for your revised schema including relevant constraints. [6]

Question 2 Constraints and Data Protection

A bank database contains the following tables:

```
CREATE TABLE Customers (
  Customer_ID     CHAR(15),
  First_Names     VARCHAR(30),
  Last_Name       VARCHAR(30),
  Year_of_Birth   INT,
  Address         VARCHAR(250),
  PRIMARY KEY (Customer_ID));

CREATE TABLE Accounts (
  Account_Number   INT,
  Customer_ID      CHAR(15),
  Account_Type     CHAR(10),
  Balance          INT,
  PRIMARY KEY (Account_Number),
  FOREIGN KEY (Customer_ID) REFERENCES Customers);
```

(a) What would you add to ensure that:

i. Deleting a customer's record resulted in all their account records being removed? [2]

ii. Changing a customer's ID in Accounts was not possible (because it is a foreign key) once it was registered with an account? [2]

(b) Construct an SQL query that would return the list of all customers who had *at least one* account overdrawn (i.e. with a balance below 0). [3]

(c) A customer transfers money between two accounts. The procedure consists of: a) checking the balance of the first account; b) debiting the money from the first account; c) crediting the second account. The first step in this procedure is to ensure that the transfer does not take the first account beyond its credit limit.

i. What is a transaction? [1]

ii. Explain the lost update problem, using the above example as an illustration. How is this normally avoided? [5]

iii. What is a deadlock? What strategies are there for recognizing and resolving deadlocks? [4]

(d) George Claw (user id: `gcla0121`) is a new administrator in the bank. He is allowed to view customers' details, along with the account numbers and types for their accounts. He is not allowed to update details, nor to view any account balances.

i. Provide the necessary SQL to establish these abilities (based on the tables given at the beginning of the question). [4]

ii. How would you modify the SQL statement to allow George to allocate the same privileges to others? [1]

iii. George is fired for dishonesty. What rule would remove his privileges *and those he had allocated*? [3]

Question 3 Distributed databases & other advanced topics

A chain of department stores is considering developing a distributed database.

(a) Give **three** objectives of a distributed database and, in each case, explain why it offers an advantage. [6]

(b) What extra factors should the query optimiser take into consideration for a distributed database? [2]

(c) How is recovery control made more complicated in a database distribution? [2]

(d) One employee suggests some other database models. For each model, give **one** advantage and **one** disadvantage:

i. a deductive database model; [2]

ii. an object-oriented database model. [2]

(e) Currently, the company system is split into multiple separate resources. What is a two-phase commit? Why would it be necessary in this case? [2]

(f) The ANSI/SPARC architecture for DBMS specifies three levels of abstraction.

i. Name and briefly describe each level. [6]

ii. Do distributed databases conform to this architecture? Explain your answer. [3]

Question 4 SQL

A museum has a card catalogue of exhibits that they hope to develop into a database. Each exhibit has a unique catalogue number, name, year and country of manufacture, and item type (e.g. 'vase'). It also records year of acquisition and the identity of the collector (the person who found it).

```
CREATE TABLE Exhibits (
  Catalogue_No  INT,
  Name          VARCHAR(40),
  Year_Made     INT,
  Country       VARCHAR(50),
  Type          VARCHAR(20),
  Year_Acquired INT,
  Collector_ID  INT,
  PRIMARY_KEY   ???
);
```

(a) What should be entered as the primary key (instead of ???)? [1]

(b) The museum only houses artefacts dating from the years 1066-1799.

i. How would you specify a constraint to ensure that all entries come from these dates? [3]

ii. Is the form you used in (b)i. above a domain integrity constraint, a column constraint, a base table integrity constraint or a database integrity constraint? Explain your answer. [2]

(c) A separate table exists for collectors, storing their first name(s), surnames, birth year and death year, along with their ID.

i. Give the SQL definition for this table, including any appropriate keys. [3]

ii. Clearly, an exhibit cannot have been acquired in a year during which the collector was not alive. Specify an appropriate constraint to check for this. [4]

(d) i. What is a view? [1]

ii. Define a view for listing exhibit name, year made, country along with the first name and surname of the collector for all exhibits of type 'Gargling bowl'. [3]

iii. Define a query using the view from (d)ii. above, which presents, for each country the total number of gargling bowls collected. [3]

(e) i. Construct a ***relational algebra*** statement for the query "Get the names and countries of all exhibits found by Henry Jones". [2]

ii. Would you expect the optimiser to perform the join or the restriction components of this query first? Why? [3]

Question 5 Database design

A public zoo needs a database to keep track of animal type, individual animals and designated keepers. Each enclosure in the zoo holds at least one species of animal, with no animal type appearing in more than one enclosure. Animal types are uniquely identified by two pieces of information – genus and species – but non-unique classification information – phylum, class, order and family – are also stored. Assume that phylum names are unique, but include many genera (genuses) and species, and similarly that orders are unique, but contain many phyla, classes are unique and contain many orders, and that families are unique and contain many classes.

Sometimes – usually only for larger animal types – an individual will be stored on the database, uniquely identified by name and species, but also with details of weight, sex and date of birth.

For keepers, the zoo needs to keep track of role (head keeper, keeper or trainee keeper), with each role having an associated salary. Each keeper will be associated with at least one animal type, and each animal type must have at least one keeper. Assume that keepers have unique names.

(a) Draw an Entity/Relationship diagram expressing this information. [9]

(b) What is a fan trap? Is there one in your diagram? If there is, where is it? [3]

(c) What are the functional dependencies for these relations? [4]

(d) Indicating all primary keys, design a relational schema normalized to at least BCNF.

Include as sample data a polar bear called Vicks. He was born on December 6th 2010 and now weighs 80Kg. The genus/species classification of polar bears is Ursus maritimus, and the family, order class and phylum are ursidae, carnivora, mammalia and chordata respectively. Polar bears are kept in enclosure E22, and looked after by head keeper Anja Peeters and trainee Nils Ritter. The two keepers are paid 50,000€ and 20,000€ respectively. [9]