W228/404C, W8900/404C

DUBLIN INSTITUTE OF TECHNOLOGY

DT228 BSc. (Honours) Degree in Computer Science

Year 4

DT8900/1 International Pre Masters for MSc in Computing

WINTER EXAMINATIONS 2015/2016

ADVANCED DATABASES [CMPU4003]

DR. PIERPAOLO DONDIO
DR. DEIRDRE LILLIS
MR. PAUL COLLINS

FRIDAY 15TH JANUARY 1:00 PM – 3:00 PM

TWO HOURS

ANSWER **FOUR** QUESTIONS OUT OF **FIVE**.

ALL QUESTIONS CARRY EQUAL MARKS.

Question 1. **[25 marks in total]**

a.

(i) Describe the concept of RAID. Describe the following RAID configuration discussing their storage efficiency and performance.

i. RAID 3
ii. RAID 1+0
iii. RAID 5

[6 marks]

(ii) Using RAID 3, can two writing operations be performed at the same time? What about RAID 5? Justify your answer.

[4 marks]

b.

Describe how indexes can make an inner join between two tables faster. Discuss the use of no indexes, an index only on one of the two joint tables and the use of two indexes, one for each table. Suppose you are joining on the indexed field. Show how using two indexes a merge-sort algorithm can be used to make the join faster compared to a nested loops or hash join solution.

[8 marks]

c.

Explain the CAP theorem for distributed databases, providing examples of one system that drop Consistency, a system that drop Availability and a system that drop Partition-tolerance

[7 marks]

Question 2. **[25 marks in total]**

a.

(i) What is a Bitmap Index in Oracle? When is it better to use it? When is it not?

[3 marks]

(ii) Suppose to have a million record and a filed "weekday" that can take only seven values (one for each day of the week). How big will be the bitmap index?

[4 marks]

b.

Describe the concept of dynamic hashing, providing an example. Discuss the advantages with respect to a static hashing scheme.

[7 marks]

c.

Insert in a (2,3) b-tree the following index values:

28, 61, 69, 30, 33, 55, 11, 78, 89, 99

[11 marks]

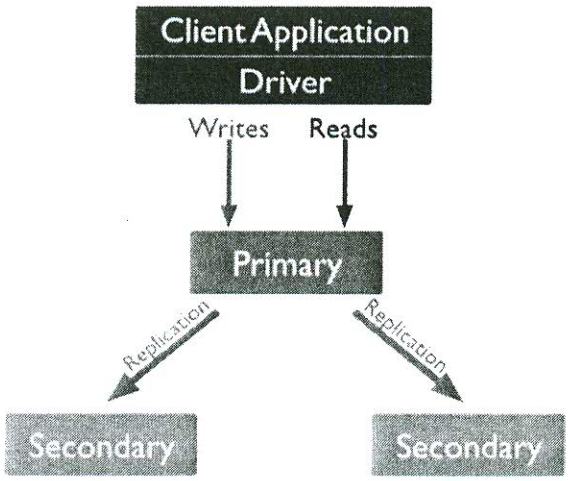
Question 3. **[25 marks in total]**

a. Describe the strengths of a Dimensional Model, including the advantages over an ER model

[9 marks]

b. Referring to the following figure, explain the concept of *replica set* in MongoDB. Discuss the functioning of a replica set under normal conditions and during a failover. What is the difference between sharding and replication in MongoDB?

[7 marks in total]



c. Describe the ACID properties of a relational database. [3]
Describe the BASE properties of a NOSQL database. [3]
Discuss the difference between the two approaches and for which application the BASE approach is more suitable [3]

[9 marks in total, 3 for each question]

Question 4. **[25 marks in total]**

a. Which are the 3 types of strategies used for managing changes in dimensions? Provide an example for each of them

[6 marks]

b. A bank requires designing of a data warehouse to record customer transactions. A transaction has two types: lodgement or withdrawal. Customers can have multiple accounts. There are three types of accounts: saving, premium and student. The following ER diagram is available (primary keys are underlined, FK = foreign keys):

ACCOUNT (aCode, OpeningDate, Balance,BranchID,AccountType, cCode (FK))
CUSTOMER (cCode, Name, Address, Phone, BDay, Gender, Salary)
TRANSACTION (Tcode, aCode (FK), Date, Type, amount_in_euro)
BRANCH (BrenchID, Name, Address,RegionID (FK), BankID (FK))
REGION (RegionID, RegionName, Country)
BANK (BCode, Name,)

(i) Produce a star schema for the above ER diagram. The diagram should support queries about the total amount of transaction (in euro) that have been done at different periods of time (down to daily reports) at each branch and by which type of transaction and by which type of account and customer. Justify your choices (grain, facts and dimensions)

[14 marks]

(j) ii. Write an SQL query to get the total amount (in euro) of all the transactions over students accounts for each branch in 2009

[5 marks]

Question 5. **[25 marks in total]**

a. Describe the concept of write concern in MongoDB. Discuss the difference between the following 4 write concern levels:
   (i) Unacknowledged
   (ii) Acknowledged
   (iii) Journaled
   (iv) Replica Acknowledged

[10 marks]

b. Explain the following dimensional model concepts:
   (i) Surrogate Key
   (ii) Business Key
   (iii) Staging Area
   (iv) Grain

[4 marks]

c. Consider the two following transactions:

(a)

T_1
read_item (X);
X:=X-N;
write_item (X);
read_item (Y);
Y:=Y+N;
write_item (Y);

(b)

T_2
read_item (X);
X:=X+M;
write_item (X);

Nothing is known about the time sequence of the above instructions. Because of concurrency, problems with data consistency might happen. Can you provide two examples of potential problem? (for instance, you might describe the lost updated problem or the temporary update problem).

[9 marks]