BCS HIGHER EDUCATION QUALIFICATIONS
Level 5 Diploma in IT

April 2011

EXAMINERS' REPORT

Database Systems

Question 1

Consider the following scenario:
"A small Bank has two branches, one in Bristol and one in Bath, and has its head-quarters in London. Currently, it has a centralised database in its headquarters where it keeps data about its customers. Locals use consists of report generation for trend analysis. On the other hand applications at the two branches access this database via a communication network for whatever data they need. There is also a communication link between the two branches, which is currently used only when one of the main links to the London headquarters fails.

The only relation in this centralized database system is the Customer relation, where data about customer accounts are kept. The attributes of the Customer relation are, the account number (Acc_no), the customer's name (Cust_name), the branch where the account is kept (Branch) and its current balance (Balance). An instance of the Customer relation follows:

Customer relation

Acc_no	Cust_name	Branch	Balance
200	Jones	Bath	1000
324	Smith	Bristol	250
153	Gray	Bristol	38
426	Dorman	Bath	796
500	Green	Bristol	168
683	Roy	Bath	1500
252	Elmore	Bath	330

Due to heavy network traffic the bank's service to its customers is poor. The Bank is concerned and has asked you to investigate database distribution designs that will improve its service."

Propose three distribution designs one for each of the following requirements. Also justify each proposal and outline its advantages and disadvantages.

i)   The database should always be available to all sites and access to it should be fast even in the case of data communication link failures.

(7 marks)

ii)   There should be no redundancy in the allocation of data, i.e. only local data should be stored at a site.

(10 marks)

iii)   A reasonable compromise between requirements (i) and (ii). And also justify each proposal and outline its advantages and disadvantages.

(8 marks)

Note: There are a number of reasonable proposals that would satisfy the above requirements, especially for requirement (iii). You will be assessed on how well you argue your proposals, and not on whether you come up with a particular one.

Answer Pointers and Marking Scheme

A reasonable set of proposals could be:
(i)   The Customer Relation is replicated to all three sites. Advantages: Availability and fast access. Disadvantages: difficult to maintain, update propagation, a lot of storage is required.

[7 marks]

(ii)   The Customer relation is horizontally partitioned into two fragments, Fragment Bath = **s** (Branch="Bath") which will be stored at Bath branch and Fragment Bristol =**s** (Branch="Bristol") which will be stored at the Bristol one as it follows:

Bath Branch

200	Jones	Bath	1000
426	Dorman	Bath	796
683	Roy	Bath	1500
252		Elmore	330

Bristol Branch

324	Smith	Bristol	250
153	Gray	Bristol	38
500	Green	Bristol	168

NB: We can even remove the column Branch containing Bath and Bristol so the tables will be reduced to three columns only.

However, the London headquarters do not store any data. Advantages: No redundancy of data, easier maintenance and security, efficient use of physical storage. Disadvantages: Data becomes unavailable is a site or links to it fails.

[10 marks]

(iii)   The Customer Relation is horizontally fragmented and allocated to Bath and Bristol branches as above, but a copy of the whole Customer Relation is also allocated to the London Headquarters.

[8 Marks]

Examiner's Comments
A large number of students have selected this question. It is apparent that most students have been exposed to topic of distributed database concepts and some of them showed that they have practiced enough to grasp the concepts. The examiner suggests that candidates should practice more on distributed database applications to be able to design database applications with advanced requirements which are not met by central database systems.

Question 2

1. What is normalization? (5 marks)
2. Use the following provided table to complete first, second and third normal forms: (10 marks)

PONo	SupCode	SupName	SupAdd	DelDate	Partno	Desc	Qty	Unit	Price
P001	S123	Windwin	York	3/2/11	PT104	Door	5	10	80.00
					PT105	Bolt	10	10	2.00
P002 Etc.	S456	Flyaway	Manchester	3/2/01	PT105	Hinge	20	10	5.00

The attributes used in the form are:

PONo stands for purchase order number
SupCode stands for supply code
SupName stands for supplier name
SupAdd stands for supplier address
DelDate stands for delivery date
Partno stands for part number
Desc stands for product description
Qty stands for quantity
Unit stands for unit per part package
Price stands for price of the part(s)

3. Use completed third normal form to generate database model with proper constraints for the relationship. (10 marks)

Answer Pointers and Marking Schemes

1. What is normalization? (5 marks)
   Normalization is a process to organize the data in an effective way, e.g. remove redundancy, improve manipulation operations, such as insertion, deletion, updating and so on for relational database management systems.
   Explanation for 3 marks
   Correct example(s) for 2 marks

2. Use the provided table to complete first, second and third normal forms: (10 marks)

a) Purchase_order – 0 (PONo, SupCode, SupName, SupAdd, DelDate, Partno, Desc, Qty, Unit, Price)

b) INF
Remove repeating group
Purchase_order-1 (PONo, Supcode, SupName, SupAdd, DelDate)
PartsOnOrder-1 (PONo, Partno, Desc, Qty, Unit, Price)

2NF
Remove partial dependence

Purchase_order-2 (<u>PONo</u>, <u>Supcode</u>, DelDate)
Supplier-2 (<u>Supcode</u>, SupName, SupAdd)
Part_order-2 (<u>PONo</u>, <u>Partno</u>, Qty)
Part-2 (<u>Partno</u>, Desc, Unit, Price)

3NF
Remove indirect dependence, above relationship is already in 3NF.

Purchase_order-3 (<u>PONo</u>, <u>Supcode</u>, DelDate)
Supplier-3 (<u>Supcode</u>, SupName, SupAdd)
Part_order-3 (<u>PONo</u>, <u>Partno</u>, Qty)
Part-3 (<u>Partno</u>, Desc, Unit, Price)

INF   for 2 marks
2NF   for 2 marks
3NF   for 2 marks
Assumption stated for 4 marks

3. Use completed third normal form to generate database model with correct relationship between entities. (10 marks)
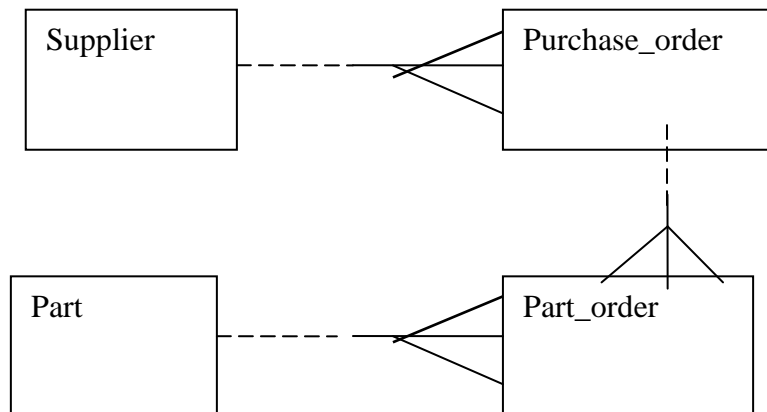
There four entities are involved in the model, i.e. Supplier, Purchase_order, Part and Part_order.

Relationships between:
1. Supplier and Purchase_order is one to many,
2. Part and Part_order is one to many
3. Purchase_order to Part_order is one to many

Correct entity names for 2 marks
Correct entity relationships for 2 marks
Correct link entity to remove many to many relationships for 2 marks
Explanations for 4 marks

Examiner's Comments
This is a very popular question as about 367 candidates with an attempt rate of 97.35%, average mark 8.92/25, standard deviation 4.02 which is lower than Q5. About 47.14 % passed the question. Above statistic number demonstrates that candidates are interested in answering this type of knowledge in the subject area, but less than half candidates are not be able to answer it correctly, indicating these areas needs further improvement.

Question 3
Read the discourse below and then use the information given in the discourse to expand your answers to the following questions.

    a) What is database modelling? Briefly describe its relevance in developing a database system. (5 marks)
    b) Explain entity types and Key attributes in a data model (10 marks)
    c) What are constraints on relationships of a data model (10 marks)

Answer Pointers and Marking Scheme
Part a)
Database modelling is the ability to interpret data driven scenarios or user specifications using an informal data analysis and functional analysis techniques. Data analysis is the key to changing from an interpretation to concrete ideas that form part of a data model. Functional analysis forms the changes or actions on the data model to confirm they can be supported as transactions.   2 marks for a sensible definition and 3 marks for its use and its profound effect on information systems as a systematic method in its own right

Part b)
Here actual modelling is expected and an understanding of one systematic method of data analysis and functional analysis. So using the discourse marks awarded for key entities being swimmers, venues, events, competition and heats and rounds. These entity types must be associated with attributes such as swimmer name, competition type, venue location. Assign attributes to entity types is worth 3 marks Entity types is worth 4 marks and allocation of key attributes is worth 2 marks. No diagram necessary but justification needed of why the application would store X in Y i.e. apply functional analysis   1 mark

Part c)
A partial data model diagram is needed to show the relationships and constraints. The marks are for a diagram illustrating the key differences in representation.
5 marks for accuracy and clarity and consistency with answer to part b)
The data model constraints include anything that forms part of the larger model so it is not required to produce a full complete model only carefully selected examples to illustrate
1 to many relationship with participation constraints for the 1 side (e.g. non obligatory) and obligatory for the many side.
3 marks
Example (2 marks) e.g. could be a swimmer may attend many events that takes place at a particular venue. The event must have a competitor and an event must have a venue but that particular swimmer does not need to compete in all events.

Examiner's Comments
A common problem with most answers was the inability or unwillingness to apply examples to the case study/discourse. Instead very generic answers were produced and hence many marks were lost. In previous questions on this topic candidates needed to apply a given case study – this question was no different apart from localising the data model construct from the full ERD. Overall knowledge of the theory/method is sound but more practice in data modelling is necessary to turn this into good answers and higher marks.

Question 4

Refer to Appendix A and the Ticket data (Figure A1).

    a) Express the ticket data given in Figure A1 as populated Tables using the information given in the discourse. (Do not show any working such as Normalisation or SQL code)

        (5 marks)

    b) Using one or two of the Tables you produced above in part a), write an SQL script (a logical sequence of SQL code) to implement the following tasks:-
        i)   Create the Tables without columns apart from the primary key (3 marks)
        ii)   Add columns to existing Tables   (5 marks)
        iii)   Add foreign keys to existing Tables.   (6 marks)

    c) Sketch out the design of an SQL script that populates all the tables you created in the tasks above in part b).

        (6 marks)

Answer Pointers and Marking Scheme

Candidates need to 'intuitively' or otherwise normalise the given data – however the process of normalisation is not marked. A rough test for normalisation should be whether 2NF or BCNF is achieved (for example BCNF every determinant is a candidate identifier). Tables imply Functional dependency diagrams so that each table will show whether this rule is upheld. Marks were awarded in terms of does the table have determinants and Candidate Identifiers as well as ordinary attributes? Do the tables exhibit uncontrolled redundancy? Simply recasting the entire data into one table is not acceptable. For example for 5 marks the examiner would look out for these principles being upheld given the data sample could indicate what attribute is FD on other attributes (or collective of attributes)

TICKET NO <-> CustomerNo,PerfNo,Seat,SeatCode

Hence TicketNo is equivalent to identified by them

these 4 attributes and uniq seatcode cannot be determined by seatno or customerno

TicketNO->ISSUE DATE
PerfNo->PERFDate/TIME
THEATRE Welldon
CustNo->CUSTOMER
**   ->CUST-ADD**

SEAT-> SEAT AREA

SEATCODE,PRODNo->
AMOUNT PAID

PRODUCTION No -> SPONSOR
** ->PERF**
->PRODUCTION COMPANY
-> PRODUCTION TITLE   --- note the need to create a unique ID

Using the above to get tables

Bookings

TicketNo	IssueDate	CustNo	PerfNo	SeatNo	SeatStatus
079231	13-Dec-2010	10032	8320	H3	C
309232	13-Dec-2010	10032	8320	H4	C
309998	15-Dec-2010	3424	869321	MM3	C
306298	15-Mar-2011	3424	9767	MM3	C
736228	15-Oct-2010	3420	9770	C11	C
079232	13-Dec-2010	10035	8320	H5	C

Performance Table

PerfNo	PerfDate	Production No	Theatre
8320	2-Jan-2011	1	Welldon
869321	15-Dec-2010	2	Byron
9767	15-Mar-2011	2	Welldon
9770	15-Oct-2010	3	Byron

Production Table

ProductionNo	ProductionTitle	Production Company	Sponsor
1	12th Night	RSC	Teesside Polymers
2	Chopped Carrot	Shaw-Taylor	The Vegan Society
3	12th Night	Shaw-Taylor	Gardeners World

Customer Table

CustomerNo	Customer Name	Customer Address
10032	R. Sayers	'Tess' Ilkley Moor
7243	P. Smith	'Homeblest', Preston Capes
10035	V. Singh	23 Belle Vue St, Odiham
3420	P.Smith	Dove Cottage Stratford

Seats Table

SeatNo	Seat Area	Seat Code	Theatre
MM3	Circle	D	Welldon
MM3	Stalls	D	Byron
H2	Front Stalls		Welldon
H3	V. Singh		Welldon
H4	P.Smith		Welldon

Part b)
i) This is straightforward SQL
CREATE TABLE seats (seatNo varchar(3) ......)   ;
Etc
Marks for correct syntax and data types
ii)
ALTER TABLE Seats ADD COLUMN Seattype varchar(4));
iii)
ALTER TABLE Seats
ADD CONSTRAINT Cdt_PK PRIMARY KEY
SeatCode

Part c)
The script is simply a set of statements that will CREATE TABLES , ADD CONSTRAINTS,
INSERT Rows
The order of CREATING tables with foreign key references is very important. The 'base' tables
that have Foreign key references to another table must be created first. Traditionally the script has
an ADD CONSTRAINT to existing tables to prevent referential integrity errors.
Finally a script that populates the tables in the reverse order that they are created is required
using the SQL syntax INSERT INTO table VALUES (col1, col2, col3 ....) preserving the primary
key integrity. Optionally a DROP table script should exist to delete the tables in the required order.

Examiner's Comments
The first part of the question seemed to be misinterpreted by many candidates simply because
they didn't read the full question. If they read the full question they would have realised the need
to produce normalised tables so that the tables were a suitable transformation of the un-
normalised ticket data. Candidates should realise that ticket data as it was presented could not be

used as a relational model of data. Unfortunately some candidates didn't realise this as thus got 0 marks. For the rest of the question most candidates could recall the SQL syntax for most tasks but because the tables were not well formed could not apply the code correctly. Marks were awarded for correct SQL syntax even though they were not applied to the tables in part a)

Question 5
The following tables (with row counts) belong to a large Alternative Healing Clinic.

Therapist(therapistID,therapistName, therapistPostCode, ...)	**100 rows**
Treatments(treatmentID, treatmentName,HourlyFee)	**209 rows**
TreatmentOfferedBy (treatmentId,therapistID)	**450 rows**
Appointments (appointmentID,therapistID,patientID,AppointmentDateTime)	**250000 rows**
Patients(patientID, patientFirstName, patientSurname, patientPostCode)	**50000 rows**

These tables are stored on a database server that supports web based client server architecture. Patients make appointments to see a therapist who treats a patients' ailment/illness by using alternative therapy (for example acupuncture).

a) Write an SQL statement that lists which treatments are offered by which therapists. The resulting table should contain 2 columns - the names of the treatments and a concatenation of the first and surnames of the therapist.

(3 marks)

b) Write an SQL statement that lists the details of appointments for treatment by "acupuncture" with a patient named Myfawy (firstname) Jones (surname) with a post code containing the string "TS1". The resulting table should contain 4 columns- patient Surname, postcode, appointment date and appointment time.

(5 marks)

c) Explain how you would extend the above query to include a count of the number of appointments satisfying the above query but instead for postcodes that match the string "TS".

(4 marks)

d) What indexes would you need to improve the performance of the queries you wrote in a) and b) above? Justify your answer.

(6 marks)

e) Explain, with reference to a DBMS of your choice, how you would measure the performance of the indexes you provided in part d above.

(7 marks)

Answer Pointers and Marking Scheme

Part a)
SELECT TreatmentName, TherapistFname+TherapistSname as TherapistName
FROM Treatments as tr , Therapists as th, treatmentsOfferedBy tob
WHERE tob.treatmentId = tr.treatmentID and tob.therapistID = th.therapistID
         3 marks, 1 mark for each row
Part b)
SELECT PatientSName, PatientPostCode, Appointmentdatetime
FROM   treatments as tr , patients as p, appointments a, treatmemntofferedby as tob
WHERE tob.treatmentID = tob.treatmentID and th.treatmentname = 'acupuncture'
AND tob.therapistID = a.therapistID and p.patientID = a.patientID and patientsname = 'Mcfawry'
5 marks, including 3 marks for correct join and use of aliases

Part c)
Extend the SQL using COUNT (appointmentID) and GROUP BY patientSName, patientFname
And alter the string to TS as a having clause
4 marks

Part d)
Indexes used on patient name maybe better to combine an index on fname and surname to improve performance
Index the treatmentname and use a non clustered index to apply sorting in sort merge of Join application 2 marks
Index the primary keys using appropriate fill factors as the primary keys should have identity functions to
autogenerate an index key. Clustered index for each table is therefore needed.
2marks + 2 marks for justification

Part e)
Many DBMS have DBA tools such as profiler query analyser and the candidates should give diagrams or commands to show how the impact of indexes can be measured ie throughput prior to index and projected this allows the actual query plan to be viewed and show the number of records accessed. The row numbers in the above tables should be shown in terms of volatility as a test of performance following updates etc.
3 marks for tools and description of use
2 marks for measures of performance and real time monitoring (statistics)
2 marks for clarity and application to the tables supplied

Examiner's Comments
Most answers to the SQL questions were satisfactory. There are still some weaknesses with JOIN conditions and a lack of understanding of COUNT and GROUP BY. For GROUP BY a corresponding matching list of attributes must exist in the SELECT/COUNT CLAUSE. For part e) the answers were very disappointing overall with evidence of lack of practical exposure to the tools available in a DBMS product. Any aspiring database practitioner has to know about these 'performance measuring' tools when they deliver a database to a customer.

Question 6
Refer to the discourse in Appendix A. Assume that at peak times the database server processes as many as 10 bookings made by different customers for seats at the same time.
a) What is the purpose of 'concurrency control' in a database system? Give examples of the concurrency control measures that would be required in the theatre booking application.

[8 marks]

b) What is meant by the terms READ COMMITTED and SERIALIZABLE in concurrency control?

[8 marks]

c) Suppose you were required to develop the 'registering customers as members of the Theatre' part of the Theatre bookings application. Explain the techniques needed to guarantee security of the data and the confidentiality of customer data.

[9 marks]

Answer Pointers and Marking Scheme
Part a)
Concurrency control is a mechanism by which transactions that are simultaneously submitted to the database engine are dealt with using ACID principles ie atomic, consistent, isolated and durable. This means updates are committed before another update that writes to the same table occurs. ACID principles attracted 2 marks definition 2 marks and example transactions applied 3 marks.

Part b)
READ COMMITTED AND SERIALIZABLE are settings for concurrency strategy that a DBA imposes the default is usually the former setting with the latter a strict setting for pessimistic concurrency control
Examples required of setting 3 marks 2 marks for consequences of each setting on ACID properties
Part c)
A broad question that examines knowledge of access control authorisation and permissions across the range of services and data objects. Expected some discussion on the granularity of access and the concept of role based authentication and experiences of using as DBMS to achieve this. Also expect a list of roles and permissions and what access control is provided for customers, casual users (web) admin and developers.

Ezaminer's Comments
Generally sound answers but most answers too dependent on bookwork. The best answers referenced the case study and applied knowledge to the problems of security and access control in particular.
In part b) the question could be answered (correctly) in two ways – one was the principles of setting a particular isolation level such as READ COMMITTED and SERIALIZABLE in terms of concurrency control, this is what most candidates interpreted. The other way was the use of the effect of applying these terms (READ COMMITTED and SERIALIZABLE) as keywords in SQL as a way of setting the transaction isolation level before any transactions (ie UPDATE, INSERT, DELETE) are invoked in a concurrent environment.

APPENDIX A: Theatre Booking Database
Discourse: The following data represents a snapshot of ticket data generated from a theatre bookings database. This database holds bookings made by customers for seats at a performance of a particular production taking place at a particular theatre. Bookings are made over the WWW by registered customers who have agreed for certain personal details (such as name, age and address) to be held in return for priority booking and information of forthcoming productions. Customers can either reserve a seat or pay for a seat in full at the time of booking. Reservations must be fully paid (and hence confirmed) within 7 days of the booking if the performance is more than 14 days ahead. Following a successful booking a 'printable' ticket is generated for the customer to print off and present at the theatre when they attend the performance or if they request a refund.
Seats are organised into areas with the most expensive in the circle with individual seats identified by a seat number (eg H3) and given a seat code, such as Reserved for sponsors (code R) ; reserved for Disabled customers (code D) otherwise the seat code is NULL. Seats are also given a seat status where 'C' indicates a booking for a seat has been confirmed and paid for otherwise the seat status is NULL if the seat is reserved.
When a performances is cancelled and customers are refunded the cost of their tickets in full subject to the customer producing a printed ticket (this acts as a proof of purchase). Alternatively customers can re-book for another performance of the same production on a different date. They must re-book or request a refund within 14 days after the date of the performance.

Fig A1 Printed Theatre Tickets

===============================	==========================	======Receipt========
TICKET NO **079231**	ISSUE DATE **13-Dec-2010**	TICKET NO **079231**
THEATRE **Welldon**	SEAT AREA **upper stalls**	
CUSTOMER **P. Smith**	SEAT H3	CUSTOMER No **10032**
CUST-ADDRESS **Dove Cottage**	PRICE £5.75	
Stratford	SEAT CODE **RC**	
PRODUCTION TITLE **12th Night**	PRODUCTION COMPANY **R.S.C.**	

PERFORMANCE DATE/am **02-Jan 2011 M**	SPONSOR **Teesside Polymers**	PERF_NO **8320**
===============================	=========================	======Receipt=======
TICKET NO **309232**	ISSUE DATE **13-Dec-2010**	TICKET NO **309232**
THEATRE **Welldon**	SEAT AREA **upper stalls**	
CUSTOMER **P. Smith**	SEAT **H4**	CUSTOMER No **10032**
CUST-ADDRESS **Dove Cottage Stratford**	PRICE **£12.75**	
	SEAT CODE **C**	
PRODUCTION TITLE **12ᵗʰ Night**	PRODUCTION COMPANY **R.S.C.**	
PERFORMANCE DATE/am **02-Jan 2011 M**	SPONSOR **Teesside Polymers**	PERF_NO **8320**
===============================	=========================	======Receipt=======
TICKET NO **309998**	ISSUE DATE **15-Dec-2010**	TICKET NO **309998**
THEATRE **Byron**	SEAT AREA **upper stalls**	
CUSTOMER **R. Sayers**	SEAT **MM3**	CUSTOMER No **3424**
CUST-ADDRESS '**Tess' Ilkley Moor**	PRICE **£13.75**	
	SEAT CODE **DC**	
PRODUCTION TITLE **Chopped Carrot**	PRODUCTION COMPANY **ShawTaylor**	
PERFORMANCE DATE/am **02-Jan 2011 M**	SPONSOR **The Vegan Society**	PERF_NO **869321**
===============================	======================	==================
===============================	=========================	======Receipt=======
TICKET NO **306298**	ISSUE DATE **13-Dec-2010**	TICKET NO **306298**
THEATRE **Welldon**	SEAT AREA **circle**	
CUSTOMER **P. Smith**	SEAT **MM3**	CUSTOMER No **7243**
CUST-ADDRESS '**Homeblest'**, **Preston Capes**	PRICE **£5.75**	
	SEAT CODE **C**	
PRODUCTION TITLE **12ᵗʰ Night**	PRODUCTION COMPANY **ShawTaylor**	
PERFORMANCE DATE/am **20-Jan 2011 P**	SPONSOR **Gardeners World**	PERF_NO **9770**
===============================	=========================	======Receipt=======
TICKET NO **079232**	ISSUE DATE **13-Dec-2010**	TICKET NO **079232**
THEATRE **Welldon**	SEAT AREA **upper stalls**	
CUSTOMER **V. Singh**	SEAT **H5**	CUSTOMER No **10035**
CUST-ADD **23 Belle Vue St, Odiham**	AMOUNT PAID **£5.75**	
	SEAT CODE **C**	
PRODUCTION TITLE **12ᵗʰ Night**	PRODUCTION COMPANY **R.S.C.**	
PERFORMANCE DATE/TIME **02-Jan 2011 M**	SPONSOR **Teesside Polymers**	PERF_NO **8320**
===============================	=========================	==================