BCS HIGHER EDUCATION QUALIFICATIONS
Level 5 Diploma in IT

October 2010

EXAMINERS' REPORT

Database Systems

General

This report focuses on the questions Q3 and Q4. Table 1 shows the statistics based on the questions answered and marked by the examiner. About 87.46% candidates attempted Q3 and 73.39% attempted Q4, which means that they are most popular questions in this exam. Both have had the highest pass rates (63.29% for Q3 and 46.86% for Q4) among questions. For standard deviation results, both questions are not too low and not too high in comparison with other questions. Thus it may conclude that these two questions are fundamental in database technology and candidates were aware about this importance.

Overall, from the evidence of marking results, the data modelling is done well by most candidates, but the normalization is challenging to them. There are not many candidates could complete three normal forms. It follows that candidates are expected to improve the understanding in this area for BMS in their ability as the appreciation of these fundamental knowledge is important to the study in DBS.

TABLE 1: stats for all questions answered

Module:	Database Systems						
Level:	Diploma						
Session:	Oct-10						
Number Entered	393						
Number Sat	327						
	Q1	**Q2**	**Q3**	**Q4**	**Q5**	**Q6**	**Total**
Examiner (initials)							
Number Attempted	233	202	286	240	134	116	
% Attempted	71.25%	61.77%	87.46%	73.39%	40.98%	35.47%	
Number Accepted	233	200	286	239	133	112	
% Accepted	71.25%	61.16%	87.46%	73.09%	40.67%	34.25%	
Number Passed	96	78	181	112	53	38	141
% Passed	41.20%	39.00%	63.29%	46.86%	39.85%	33.93%	43.12%
Max Mark	23	24	22	23	22	22	74
Min Mark	1	0	0	0	0	1	1
Average Mark	8.83	8.18	10.64	9.50	8.59	7.74	33.74
Standard Deviation	4.96	5.21	4.59	5.34	5.93	5.70	16.70

Question 1

a) Discuss the main limitations of database systems that led to the evolution from first generation to the third generation of database systems.

(5 marks)

b) Discuss the major functionalities and standards provided in the web server (middle tier) that enables communication between the Internet and database systems?

(10 marks)

c) Compare and contrast the two-tier client server architecture for traditional database systems with the three-tier architecture. Why is the latter (i.e. the three tiers architecture) more appropriate for the Web database?

(10 marks)

Answer Pointers

a) The student could present different reasons that led to move from one generation of database to another generation. However for 5 marks one of the following reasons should be reported:
   - Although hierarchy & network data models provided good logical construct they faced with the problems of data dependency (any change in data structure requires changes in the programmes acting on that data and vice-versa), data redundancy, etc…
   - In 1970, relational data models came to solve the problem of data dependency. But later in 1980 and early 1990s the relational model faced with the problems of handling complex data types (1^{st} Normal Form).
   - This led to emerging object data model to deal with complex objects (such as video, sets, sound etc…) and also reduce difficulties in the development and the maintenance of large-scale database systems by introducing into database the OOP concepts such as encapsulation, inheritance, and polymorphism.

[5 Marks]

b) The three major functions that adapts the internet to database requirements are:
   - As it is an HTTP server, it processes the HTTP protocol i.e. receiving requests and generating responses in HTTP format,
   - It hosts scripting environment so developers can write code in languages like VBScript and JavaScript and execute the code on the Web Server
   - In database application, the Web Server create, read and delete views instances
   - Standards included HTML , CGI , XML and ODBC/JSP JDBC

[10 Marks]

c) The key points follow on from the answer to part b). The need for a third tier (often referred to as n-tier architecture) in a database driven web site is advantageous because of the separation of presentation (1^{st} tier), the application processing (2^{nd} tier), and the data management (3^{rd} tier) are logically separate processes. For example, an application that uses middleware to service data requests between a user and a database employs multi-tier architecture. 3-tier application architecture provides a model for developers to create a flexible and reusable application. By breaking up an application into tiers, developers only have to modify or add a specific layer, rather than have to rewrite the entire application over. There should be a presentation tier, a business or data access tier, and a data tier. 2 tier client server is less adaptable and depends on thick clients doing a lot of processing with high dependency on the operating system and the database connectivity (usually via ODBC/JDBC type protocols).

Examiners Comments
A popular question with a large variety of quality in answers produced. Most striking was the lack of understanding the development of database systems through its key stages from hierarchical to object oriented – a topic covered in most database texts.

Parta - The most popular answer was the first bullet point but data dependency was the key concept that was addressed. Many candidates also mentioned file based systems and this answer was incorrect as file based systems were not database systems – the only 1st generation systems were the hierarchical or network/CODASYL systems.

Part b - No real problems with most answers but many candidates used a lot of words instead of diagrams and example code which would explain concepts more convincingly. Examples of 3 tier architectures are essential to amplify points.

Part c - Again same comments as part b) generally too much writing and short on examples and diagrams to illustrate what would save many words.

Question 2

The following data structure and constraints exist for a magazine publishing company. The company publishes one regional magazine each in the states of Florida (FL), South Carolina (SC), Georgia (GA), and Tennessee (TN).

The company has 300.000 customers (subscribers) distributed throughout the four states. On the first of each month, an annual subscription INVOICE is printed and sent to all customers whose subscription is due for renewal. The INVOICE entity contains a REGION attribute to indicate the state (FL, SC, GA, TN) in which the customer resides.
CUSTOMER (CNUM, CNAME, ADDR, CITY, ST, ZIP, SUBSDATE)
INVOICE (INVNUM, REGION, CNUM, DATE, TOTAL)

The company's management is aware of the problems associated with centralised management and has decided that is time to decentralise the management of the subscription into its four regional subsidiaries. Each subscription site will handle its own customer and invoice data. The company's headquarters management wants to have access to customer and invoice data to generate annual reports and to issue ad hoc queries, such as:
List all current customers by region
List all new customers by region
Report all invoices by customer and by region

Given the above requirements answer the following questions:

a) What recommendations will you make regarding the type and characteristics of the required database system?

(3 marks)

b) What type of data fragmentation is needed for each table?

(3 marks)

c) What must be the criteria used to partition each database?

(7 marks)

d) Design the database fragments. Show an example with node names, location, fragment names, and attribute names.

(6 marks)

e) What type of distributed database operation or queries must be supported at each remote site?

(3 marks)

f) What type of distributed database operation or queries must be supported at the headquarters site?

(3 marks)

Answer Pointers

(a) A distributed database with the characteristics that data is located where there is a great demand for it for faster data access and data processing (localise the operations)

[3 Marks]

(b) Hybrid Fragmentation for both Customer and Invoice tables

[3 Marks]

(c) Since the HQ needs info by customers by region, new customers by region, invoices by customer and by region then Customer would be Derived VF using the criteria region of INVOICE such as CUTOMER-HQ(Cnum,Cname,SubDate) for the HQ and then CUSTOMER-R(Cnum,Addr,City,St, Zip) for each Region. Then CUSTOMER-R is again HF using the criteria region into CUSTOMER-FT, CUSTOMER-SC,CUSTOMER-GA and CUSTOMER-TN. Also the INVOICE would mixed fragmented such as INVOICE-HQ(INVNUM,Region) and INVOICE-R(INVNUM,CNUM,DATE,Total). Then INVOICE-R is horizontallu fragmented into INVOICE-FT,INVOICE-SC,INVOICE-GA, and INVOICE-TN.

[7 Marks]

(d) Student were expected to present examples based on the above fragmentations

[6 Marks]

(e) Calculation of the Total, Change of customer details e.g. address, city, etc., Date of invoice

[3 Marks]

(f) Customer list by region, new customer by region, invoice by customer by region, etc…

[3 Marks]

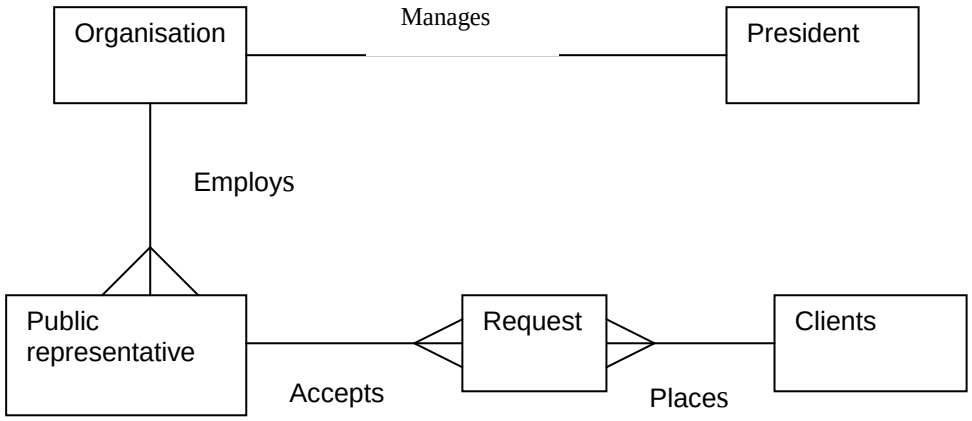
Examiners' Comments

A fairly popular question without any real concern in candidates answers apart from repeating the same answer to parts a) and b). For parts e) and f) many candidates supplied SQL expressions of the queries which gained the full marks. Candidates should were possible use SQL (even if it is not asked for) in answers to avoid any ambiguity. It is also useful to supply sample data.

Question 3

Use the concepts of entity integrity and referential integrity in relational database to answer the questions stated below.

An organisation's database system is shown in the data model below:



Work to do:

a) Explain what the data model represents. Explain the concepts of entity integrity and referential integrity, which may be used in the model.

(5 marks)

b) List the Primary key and foreign key(s) for each entity, stating the assumptions you made.

(10 marks)

c) For each relationship and in each direction explain your choice of optional/mandatory. State any assumptions made.

(10 marks)

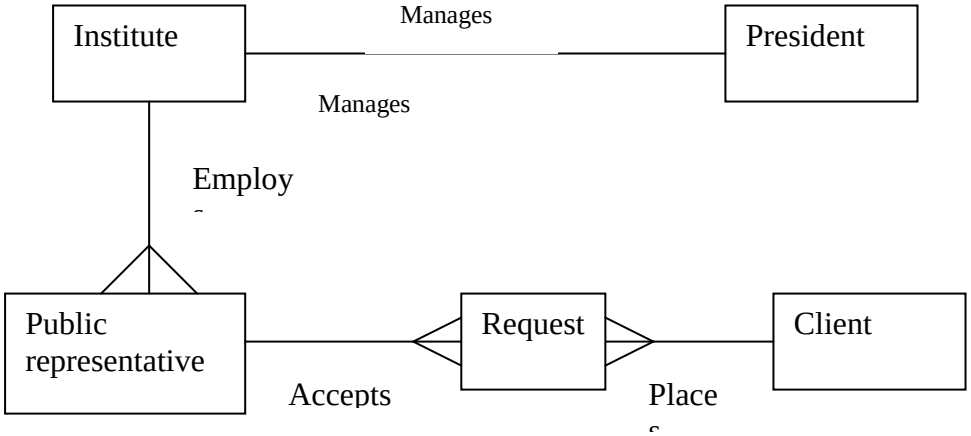
Answer Pointers

Question 3: 25 marks. 5 marks for Q3a, 10 marks for Q3b and Q3c.

A) This model shows that President manages the organisation; the institute employs public representative; PR person accepts requests that are placed by the clients of the organisation.
Explain the relationship correctly for 5 Marks

B) Institute: PK: InstituteID
President: PK: PresidentID; FK: InstituteID
Public representative: PK: PRID; FK: InstituteID
Requests: FK: PRID, ClientID.
Clients: PK: ClientID
Indicate PK correctly for 4 marks
Indicate FK correctly for 4 marks
Discussion for 2 Marks

C) Relationship suggestions:



Manages
Mandatory in both directions
Assuming one organisation must have one president and one president must be onto the organisation.

Employs — Optional in one direction, i.e. assuming one institute may have many public representatives
Mandatory in one direction, i.e. one public representative must be onto the organisation.

Accepts — Optional in one direction, i.e. assuming one public representative may accept many requests.
Mandatory in one direction, i.e. one request must be onto the PR person.

	Places	Optional in one direction, i.e. assuming one client may place many requests. Mandatory in one direction, i.e. one request must be from a client.
	Explain each relationship for each direction for 6 marks Explain assumption for each direction for 4 marks	
		[Total 25 marks]

Examiner's Comments

This is a very popular question as about 286 candidates with an attempt rate of 86.46%, average mark 10.64/25, standard deviation 4.59 which is the lowest among questions. About 63.29 % passed the question, which is the highest among questions. Above statistic number demonstrates that candidates are interested in answering this type of knowledge in the subject area, but about half number of candidates are not be able to answer Q3c correctly, indicating these areas needs further improvement.

Question 4

Normalization is key method used in relational database management systems. Use the concepts of normalization to answer the question stated below:

Analyse the following data stored in a typical university file to record information with regard to students, courses, modules and class times. Essentially, a student is enrolled onto a course and may take several modules as part of this course.

StudentID	StudentName	CourseID	CourseDuration	ModuleID	ModuleName	ClassRoom
C0123456	Mike	C101	3	M360	Database	CW3/19
				M301	WebTech	CW2/17
				M102	Software	CWG/02
C0212345	Anne	C102	4	M201	BusinessIT	CWG/01
				M203	MusicTech	CW5/18
				M102	Software	CWG/02
C0321234	Jack	C103	5	M301	WebTech	CW2/17
				M103	Network	CW4/04
				M360	Database	CW3/19
				M201	BusinessIT	CWG/01
C0432123	Helen	C104	4	M203	MusicTech	CW5/18
				M102	Software	CSG/02
				M360	Database	CW3/19
C0543212	Ben	C105	3	M102	Software	CWG/02
				M201	BusinessIT	CWG/01
				M203	MusicTech	CW5/18
				M301	WebTech	CW2/17

Work to do:

a) Represent the above an un-normalised relation and identify any candidate keys. From these candidates choose the primary key. State any assumptions you make.
(5 marks)

b) Normalise your un-normalised relation, showing the development of your design through the forms 1NF, 2NF and 3NF. At each of the three stages state any assumptions you make about the choice of primary key. Indicate the foreign keys for relevant entities.

(20 marks)

Answer Pointers

Question Q4: 25 marks. Q4a for 5 marks, Q4b for 20 marks

a) Unnormalised form: **Student - 0** (StudentID, StudentName, CourseID, CourseDuration, ModuleID, ModuleName, Lecturer)
Primary key for this data would be 'studentID'.
List unnormalised form for 2 marks
Indicate PK from candidate keys for 3 marks

b) Normalisation:

1NF - remove repeating groups, e.g. by separating out the repeating relation.
Assume that ModuleID is unique identifier for module item.

Student-1 **(StudentID, StudentName, CourseID, CourseDuration)**
Module - 1 **(StudentID, ModuleID, ModuleName, ClassRoom)**

2NF - remove partial dependencies. Split the second table to obtain the 2NF.

Student - 2 **(StudentID, StudentName, CourseID, CourseDuration)**
ModuleStudent - 2 **(StudentID, ModuleID)**
Module - 2 **(ModuleID, ModuleName, ClassRoom)**

3NF - remove transitive dependencies

Student - 3 **(StudentID, StudentName, CourseID)**
ModuleStudent - 3 **(StudentID, ModuleID)**
Module - 3 **(ModuleID, ModuleName, ClassRoom)**
Course - 3 **(CourseID, CourseDuration)**

Complete INF for 6 marks
Complete 2NF for 6 marks
Complete 3NF for 6 marks
Explanation for 2 marks

Examiner's comments

About 240 candidates with a rate of 73.39% attempted answering this question. But the pass rate 46.86% is relatively high in comparison with other questions. Again, it is one of the most popular questions in this exam. Most candidates are capable to PK, FK, candidate keys but are not be produce proper results for three normal forms. Finally it concludes that for most candidates normalization is still a challenging to them.

Question 5

a) Work out the overall function of the following SQL code samples:-

```
--- CODE SAMPLE 1
CREATE VIEW V1 AS SELECT name,
      amount_received,
      amount_sent,
      (amount_received - amount_sent) AS balance
FROM table_customers c
  JOIN accounts_table a
    ON a.customerid = c.customer_id

SELECT balance FROM V1 where CustomerType = 'Personal'

--- CODE SAMPLE 2
START TRANSACTION;
  UPDATE accounts_table SET balance=balance-200 WHERE
account_number=1234;
  UPDATE accounts_table SET balance=balance+200 WHERE
account_number=2345;
IF ERRORS=0 COMMIT;
IF ERRORS<>0 ROLLBACK;
```

(12 marks)

b) Describe the factors you would consider in deciding whether to use either a VIEW or a TABLE to express a database query.

(7 marks)

c) Describe the factors you would consider when deciding to use a either a single SQL statement or a database transaction to express a database update operation.

(6 marks)

Answer Pointers

Part a)
For 6 marks
SQL code 1
The function of this code is to CREATE a view that JOINS 2 tables to get the account balance of all personal type customers. Describe use of aliases and JOINS to generate the view and its realisation.
For 5 marks
This uses the concept of a transaction integrity to maintain consistency and integrity in concurrent environment

Part b
Views
Changing the data in a table alters the data shown in subsequent invocations of the view. But they can cause inefficiency
Views can provide advantages over tables though :

* Views can represent a subset of the data contained in a table
* Views can join and simplify multiple tables into a single virtual table
* Views can act as aggregated tables, where the database engine aggregates data (sum, average etc) and presents the calculated results as part of the data
* Views transparently partitioning the actual underlying table
* Views take very little space to store; the database contains only the definition of a view, not a copy of all the data it presents
* can provide extra security

*can provide abstraction, so database users can create abstraction by using views. In Views can make it easier to create lossless join decomposition.
Transactions Part b)
Must be applied in database updates that interleave other users access to the same data.
Must be short lived to avoid lock up and long duration transactions

Examiners' Comments

Most candidates could understand the transaction code and understand Views but many descriptions were misleading because the context to using the supplied code was missing. Clearly transactions have to apply the ACID properties of transactions and therefore the use of COMMIT and ROLLBACK and the scope of these operands was significant in the explanations. Similarly using Views in conjunction with transactions is significant because Views cannot be subject to ACID principles as Views are in effect virtual tables and therefore invalidate ACID principles. Very few candidates could make this connection and apparent flaw and therefore lost marks.

Question 6

Maintaining THE consistency between entities is essential for relational database management systems. The existence of Weak entity strongly affects on the integrity of a relational database system.

a) Explain what happens when a weak entity is subject to a cascade delete/update process.

(5 marks)

b) Illustrate an example of the use of cascaded delete/update process with an explanation.

(10 marks)

c) Explain the integrity rules which are applied in the cascaded delete/update process

(5 marks)

d) Give a definition and an alternative terminology for the term of weak entity

(5 marks)

Answer Pointers

a.

If one item has been deleted, all related items in relevant entities will be deleted in the database.

[5 marks]

b. Illustrate an example of the use of cascaded delete/update process

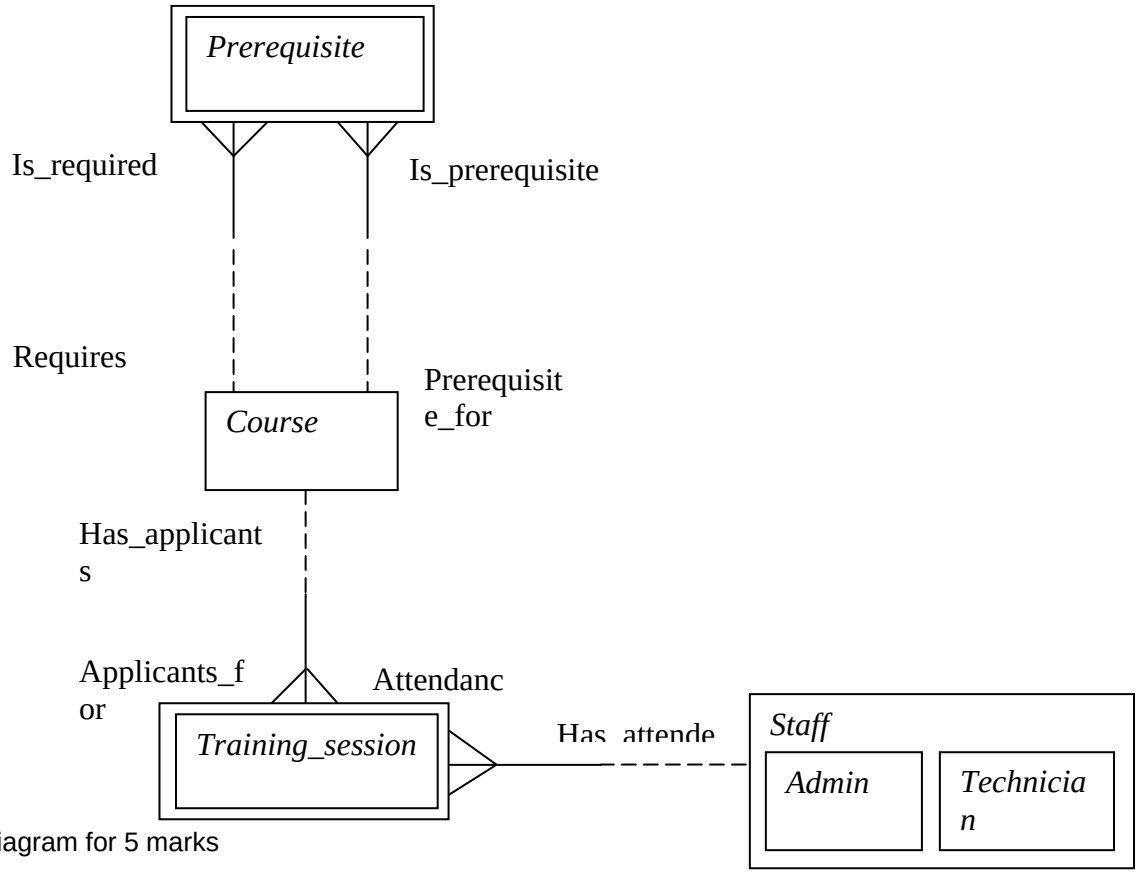


Diagram for 5 marks

Explanation:
Weak entities:
Prerequisite, which is associated with course, depends on the course entity;
Trianing_session, is associated with course entity, depends on the Course entity;
Admin and **Technician**, which are sub-entities of Staff, depend on the entity of Staff.
For cascade delete/update process, an attempt to delete/update a course should automatically delete/update all associated prerequisite records for that course. However, an attempt to delete/update a course which is a prerequisite for another course should fail as course entity is not dependent on the entity of prerequisite.
Explanation for 5 marks

c.

Referential integrity rules are applied in the cascaded delete process as one item has been deleted; other data of foreign keys in related entities are deleted as well.

d.

Definition: Weak entity is an entity attached to associated or parent entity.
Alternative term: A dependent on the other entity

Examiners' Comments
This was a popular question and generally well answered. The best answers illustrated answers with SQL code that defined the CASCADE operations at table definition time. However those candidates that answered this question poorly usually misread the question, omitted an accurate and poorly formulated diagram replaced instead with a lot of unnecessary description recalled from text books.