BCS HIGHER EDUCATION QUALIFICATIONS
BCS level 5 Diploma in IT

April 2010

EXAMINERS' REPORT

Database Systems

Question 1

The following concepts are important parameters when judging the effectiveness of data storage solutions. For EACH concept, briefly compare and contrast how effective the database approach and file-based approach are, highlighting any particular strengths or weaknesses of either. **Five marks each**.

- Data Integrity
- Data Security
- Data Redundancy
- Data Maintenance
- Data Consistency

Answer Pointers

Data Integrity – The database approach is much stronger due to the presence of primary and foreign keys in addition to various domain enforcement techniques such as CHECK constraints and NOT NULL etc. All these prevent the accidental creation of invalid data in the tables. Conversely, simple data files have very weak integrity checks and such validity checking is reliant upon the application logic accessing these files.

Data Security – Data held in a database is under the control of a host DBMS which has very strong user security mechanisms including passwords, profiles, privileges and roles – all of which is captured as meta-data in the data dictionary. Data files have none of these inherent features and any security mechanisms must be provided by the attendant application programs.

Data Redundancy – Via the process of normalisation a good database design will minimize redundant data storage (without completely removing it). As there will be one central database for all associated application programs, only a single data source is required – thus minimizing data redundancy. With simple data files, each application will have its own set of data files, each possibly storing duplicate (and redundant) copies of the same data values. Data files encourage distributed, duplicated data storage whilst databases encourage centralised, minimal data storage.

Data Maintenance – If databases support centralized, minimal data storage and data files encourage distributed, duplicated data then it is obvious that data held in a databases will be easier to maintain (updates, deletions etc)– as there is only one copy.

Data Consistency – Related to the last point, the more copies of a given data item you hold, the more difficult it will be to ensure that all copies of that data item are consistent when updates or deletes are applied. Hence, data files are inferior to databases on this issue.

Examiners Comments

This question was generally answered fairly well and was very popular. However many candidates seemed to answer the same points in other parts of the question particularly Data integrity and data consistency repeating the same comments on the advantages of a having a centralised database versus a collection of files. Although this comment is valid it would be expected to advance the distinctive argument of a centralised database in terms of its integrity (central control of handling restoration of corrupt data) and consistency (programmatic handling of constraints). Those candidates who clearly distinguished these features gained the most marks and few if any marks were awarded for repeating some general comments that were recalled from text books.

Question 2

In web-based auction sites users submit bids and compete with other users bidding for the same product. Bidding continues for a specified period of time before the highest bidder (who becomes the buyer) secures the product. When the seller receives payment he/she posts the product to the buyer. A transaction is completed when the buyer receives the product. Since the seller is anonymous no correspondence from the buyer to the seller occurs, therefore a large amount of trust occurs. The integrity and honesty of sellers is recorded and this is made known to potential bidders.

a) Describe the requirements of a DBMS and database server needed to support the application outlined above.

(8 marks)

b) Explain the interaction between a database server and a web server in order to present data stored in a database on a web browser. Illustrate your answer with references to application data and program code applicable to a web-based auction site.

(10 marks)

c) Discuss the trade-offs of implementing the program logic and business rules on:-

i) The application/web server;
ii) The database server.

(7 marks)

Answer Pointers

(a) Marks allocated for issues pertinent to the application described:

- Protection/integrity of users: Access control and security
- Monitoring abuse of rules in the application
- Provide membership roles and establish rules/protocols
- Resource switching and connection time out
- Reliable with adequate backup control
- Extra marks for good examples **(8 marks)**

(b) Marks spread equally for:

- Connection strings and Data Source Names
- The concept of ODBC and object level drivers
- How ownership of database equates to connection properties
- Grants and permission set for anonymous web user (eg ASP.NET)
- OPEN/CLOSE connection to named data source
- ASP.NET users may mention web config and other options
- Mention the use of cursors or equivalent data streaming measures

(10 marks)

Marks for emphasising choices and dilemmas:

- Transaction integrity (eg DTC – distributed transaction coordinator) being built into DBMS.
- Portability of code (eg poorer in DBMS because of lack of standards)
- SQL model better for closeness in mirroring I/O actions
- Performance DBMS better due to direct access to query optimiser and server processes
- Power and generality of OO programming frameworks

(7 marks)

Examiners Comments

Most of the students who selected this question have answered it well especially sub question (b). However, few students have explained the sub question (a) i.e. requirements of a DBMS and database server needed to support the e-commerce applications or (c) the trade-offs of implementing the program logic and business rules. The examiner suggests that candidates should 'read around' topics and also go through real example to appreciate the need for such three tiers architecture for e-commerce database applications.

Question 3

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. Local 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 fail.

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

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 suffering. 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:

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)

Answer Pointers

Propose three distribution designs one for each of the following requirements:

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.
ii) There should be no redundancy in the allocation of data, i.e. only local data should be stored at a site.
iii) A reasonable compromise between requirements (i) and (ii).

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, which are stored one at the Bath branch, and one at the Bristol one. The London headquarters do not store any data. Fragment Bath = **s** (Branch="Bath") and Fragment Bristol =**s** (Branch="Bristol"). 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 fail.

(10 marks)

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

(8 marks)

Examiners' Guidance Notes

A large number of students have selected this question on distributed databases. Most of them have answered correctly this question. It is clear that suggestions by the examiners in previous report to expose student to distributed database have been taken on board and that is why students have achieved good marks in this questions.

Question 4

Refer to Appendix A (2BHired discourse)

Produce an ER data model for the discourse given in Appendix A using UML class modelling notation. An example of UML notation is given below in figure 4.1

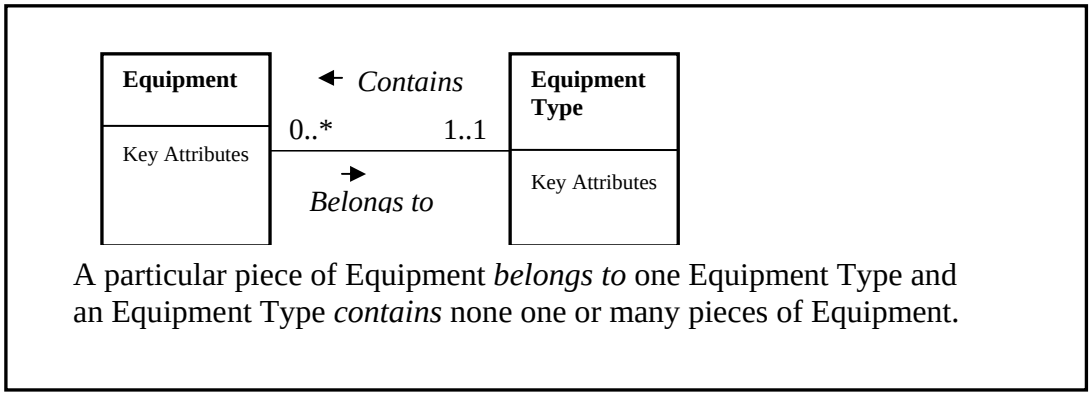
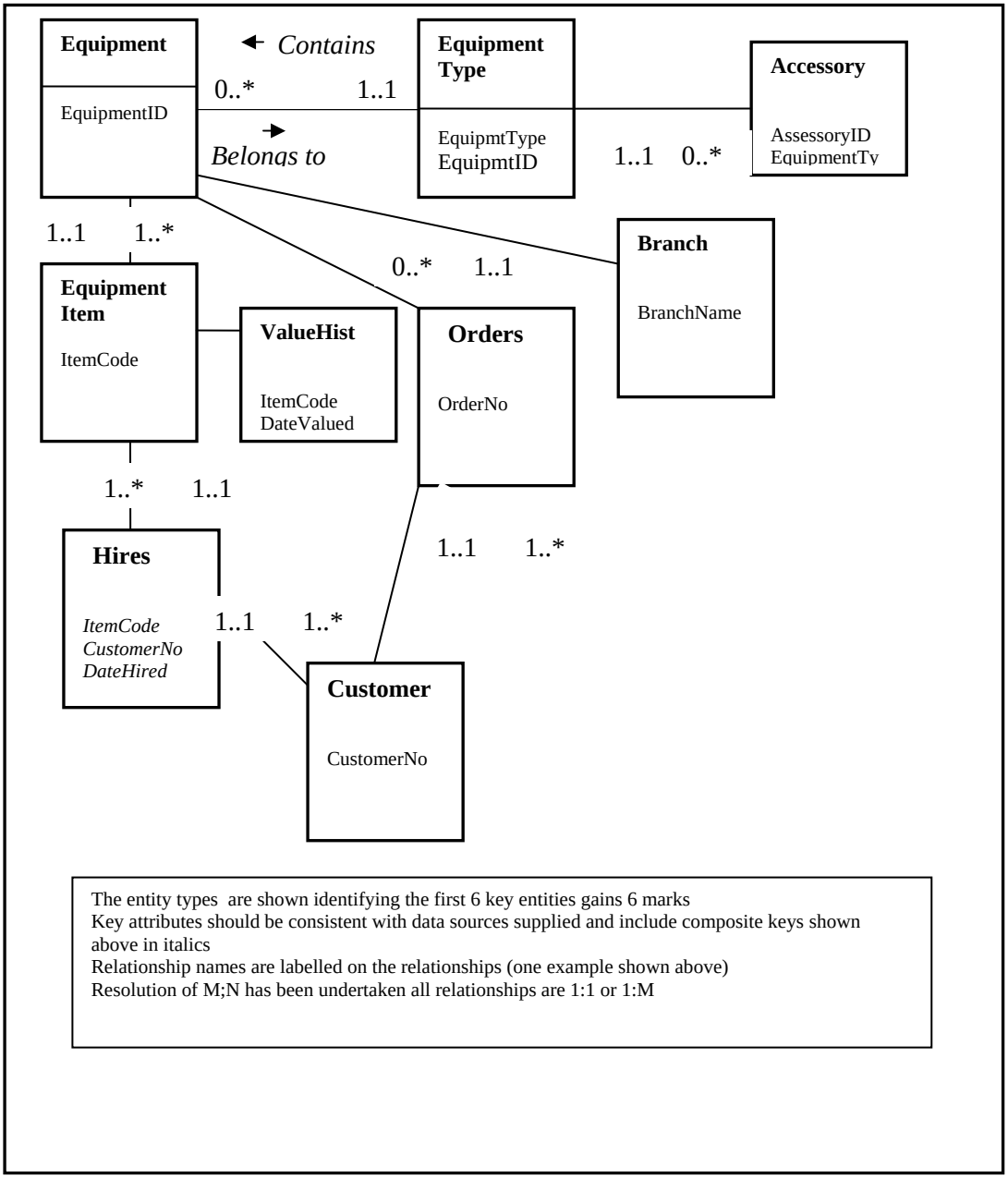


Figure 4.1: UML Class diagram notation

Marks are awarded on the ooverall accuracy, expressiveness and completeness of your ER model in representing:-

- Entity Types (Classes) **(6 marks)**
- Key Attributes **(5 marks)**
- Relationship names (associations) **(3 marks)**
- Relationship Constraints **(7 marks)**
- Resolution of many to many relationships. **(4 marks)**

Answer Pointers



The entity types are shown identifying the first 6 key entities gains 6 marks
Key attributes should be consistent with data sources supplied and include composite keys shown above in italics
Relationship names are labelled on the relationships (one example shown above)
Resolution of M;N has been undertaken all relationships are 1:1 or 1:M

Examiners Comments

This question was generally answered fairly poorly with very few candidates getting high marks. It was disappointing to see some candidates ignore the notation provided to represent an ERD in UML (figure 4.1). In Figure 4.1 the candidates were provided with an initial ERD to develop their answer. It was surprising to see so many poor attempts given this is a regular type of question in this paper.
The marks were broken down as shown above but there was some discretion given to candidates who developed alternative models that showed understanding of the functional requirements and backed up by assumptions that did not contradict the discourse.

Question 5

Refer to Appendix A.

a) Refer to Figure A1 in Appendix A

Table A1 represents data that is presented on a spreadsheet and is stored in a format called a comma separated value (csv) 'flat file'.

Describe the main disadvantages of processing data held in a flat file compared with processing data held in a database table

(9 marks)

b) Explain how EACH of the following problems associated with the existing 2BHIRED information system can be overcome using a database approach:-

- The record card system has created problems such as the same piece of equipment has been entered twice with different serial numbers.
- Store persons often bypass editing changes on the record card on the computer by writing those changes on the printed card.
- The process of recording a hire and issuing a hire agreement to a customer can be very slow at busy times. Customers can become very impatient with the delays this causes.
- Personnel at Head Office find it difficult to extract operational, marketing and management information. This is needed to develop the business and to give them a competitive edge.

(16 marks)

Answer Pointers

Part a) 3 main points are • Lack of data independence, i.e. have to know the physical structure and have to reflect logical changes (i.e. add a column) directly on physical structure • No meta definition, i.e. schema to impart logical conceptual design • Abstract notion of DDL and DML embedded in host language hence lack of standardisation

Part b) expand on the following pointers
i) Enforce Entity Integrity + example
ii) Views or Access control or DBMCL (example grant/deny create view v1 as …..)
iii) Automated constraint checking using defaults bypassing manual data entry
iv) Use OLAP (data analysis services) that are built on top of a db synchronised to -> data warehouse

Examiners Comments

The candidates answered this question (Q5) with an attempt rate of 75.27%, average mark 9.38/25, standard deviation 4.44 which is relatively low in comparison with Q6. About 50% passed the question. Above statistic number demonstrates that candidates are interested in answering this type of background knowledge in the subject area, but the basic knowledge in these areas needs further improvement.

Question 6

Refer to Appendix A

Draft a set of tables (using SQL code) that will support the following requirement:

> Occasionally a hiring cannot be met because an item of equipment is being serviced at the same time it was due to be hired out. The system needs to recognise also that there may be more than one identical item available or there may be similar items that could be offered if the requested item is not available.

Marks are awarded on the extent your Table design satisfies the above requirement supported by the following:-

Design is sound and Rules of normalisation have been applied	**(5 marks)**
Column names assigned to Tables	**(5 marks)**
Primary keys and Foreign Keys have been assigned	**(4 marks)**
Applied column constraints (domain, default, Null)	**(3 marks)**
Supplied consistent sample data	**(4 marks)**
SQL code is reasonably correct and is consistent	**(4 marks)**

State any assumptions you make.

Answer Pointers

Question Q6: 25 marks.

Marks are awarded on the extent your Table design satisfies the above requirement supported by the following:-

Design is sound and Rules of normalisation have been applied [5 marks]
Column names assigned to Tables [5 marks]
Primary keys and Foreign Keys have been assigned [4 marks]
Applied column constraints (domain, default, Null) [3 marks]
Supplied consistent sample data [4 marks]
SQL code is reasonably correct and is consistent [4 marks]
State any assumptions you make.*Underline PK and italisise for FK*

Equipment (<u>equipmentID</u>, EquipmentDescription ,** EquipmentCategory**)
***Item(<u>itemCode</u>, ….., statusCode,** equipmentID*
Hiring (<u>Hirecode</u>,** EquipmentIDRequested, itemCodeHired**,** CustAccountID**,
HireDate, ReturnDateDue)
Service (<u>serviceCode</u>,** itemCode**, MechanicName , ServiceDueDate,
ServiceActualdate)
CustomerAccount(<u>CustAccountID</u>, MemberStartDate, CompanyAddress)

Assumption is that Equipment Category is needed to associate similar types of equipment and that status codes check the status of an item at any date it is due to be hired out and thus would validate the integrity of a hire made on an item that is being serviced. It could be there is another item available.

SQL code example

CREATE TABLE Equipment (EquipmentID integer NOT NULL Primary key,
EquipmentDescription varchar(99) NOT NULL DEFAULT 'default',
EquipmentCategory varchar(59) NOT NULL CHECK EquipmentCategory IN ('A', 'B', 'C'));

CREATE TABLE Item (ItemCode integer NOT NULL Primary Key,
StatusCode char(3) NOT NULL DEFAULT ('SPA'), EquipmentID integer NOT NULL
Constraint FK_EquipmentID References Equipment(equipmentID);

Data types DateTime, Integer, varchar(N)
Sample data must exhibit the FK-PK references and data types and constraints

Examiners Comments

About 36.02% candidates attempted answering this question. But the pass rate is relatively high in comparison with Q5. Although there is a relatively small number of candidates choose to answer this question, it seems they are relatively confident in SQL coding rather than theoretic descriptions in comparison of Q5. Most candidates are capable to produce the SQL but do not provide explanation and assumptions for their answers. Most candidates lost their marks by missing database design and sample data. Finally it concludes that for most candidates SQL coding is managed reasonably well.