THE BCS PROFESSIONAL EXAMINATION
BCS Level 6 Professional Graduate Diploma

April 2008

EXAMINERS' REPORT

Advanced Database Management Systems

General Comments

Table 1 shows the statistics based on questions answered, including where some students answered four questions. Averages are fairly consistent across questions, though, Q1 is a little on the low side. Standard deviation results are fairly low, except Q4 is relatively high. Questions 2, 3 and 4 were the most popular amongst candidates. The relative general questions (Q2 and Q3) were attempted by the largest number of candidates. Thus it concluded that candidates appear to prefer general questions on ADBMS issues and technologies rather than specific programming type questions on particular issues.

TABLE 1: statistics for all questions answered

	Q1	Q2	Q3	Q4	Q5	Total
Number Attempted	44	91	85	79	57	
% Attempted	37.29%	77.12%	72.03%	66.95%	48.31%	
Number Accepted	44	91	85	78	55	
% Accepted	37.29%	77.12%	72.03%	66.10%	46.61%	
Number Passed	33	81	58	52	29	95
% Passed	75.00%	89.01%	68.24%	66.67%	52.73%	80.51%
Max Mark	20	24	24	25	23	89.33
Min Mark	1	0	4	1	0	14.67
Average Mark	12.55	13.85	12.53	12.37	9.87	49.58
Standard Deviation	4.31	3.69	4.91	5.50	4.89	15.20

It is clear that the most popular questions are Q2 (77.12%) and Q3 (72.03%). Q1 (37.29%) and Q5 (48.31%) are less popular questions. For the pass rate, Q1 (75%) and Q2 (89.01%) reach the highest pass rate among all questions.

Question 1

a)   Traditionally Data Definition Language (DDL) is used to define the database structures and Data Manipulation Language (DML) is used to manipulate the data stored in a relational database. Explain:

  i)   What can "Object Definition Language (ODL)" and "Object Query Language (OQL)" be used for and their relationship with Relational Database Management System (RDBMS)

(5 marks)

  ii)   The main advantages of ODL and OQL in database management systems?

(5 marks)

b) A summer school maintains its database for the enrolment of students. The following ODL is a class definition for an object type, Student, instances of which represents students of the summer school.

```
class Student : Person_IF
(extent students
key student_ID)
{
attribute String<ccode_length>
student_ID;
relationship List<Enrolment>
request_enroll
inverse Enrolment::enrolled_by;
Enrolment request_enroll
(List<EnrolLine> details;)
raise (course_doesnot_exist);
}
```

Explain the meanings of each statement for the above ODL.

(10 marks)

c) The following OQL will partition property objects such that there are four partitions, respectively containing:
- properties located in Leeds with average price lower than £180,000;
- properties in Leeds with average price £ 180,000 or more;
- properties not in Leeds with average price lower than £180,000;
- properties not in Leeds with average price £ 180,000 or more.

A property object type includes 'location' and 'averageprice' attributes.

```
select *
from properties p
group by
    Which_city ? :
p.location = "Leeds", LessExpensive?: p.avergeprice < 180,000;
```

Explain the meaning of each statement for the above OQL.

(5 marks)

INDICATIVE SOLUTIONS

Question 1: 25 marks. 10 marks for 1a in which 5 marks for (i) and 5 marks for (ii); 10 marks for 1b; 5 marks for 1c.

a.
  i. Traditionally, RDBMSs support a Data Definition Language (DDL) and a Data Manipulation Language (DML). OODBMSs support Object Definition Language (ODL) and Object Query Language (OQL). ODL is used for defining interfaces to object type equivalent to the DDL of RDBMS. OQL is used for the retrieval and manipulation of objects equivalent to DML of RDBMS.
  Answer "what can ODL and OQL be used for" = **1 marks**
  Answer the relationship: ODL equivalent to DDL = **1 mark**
  Answer the relationship: OQL equivalent to DML = **1 mark**
  Answers are precise and complete with clarity = **2 marks**
  Total 5 marks
  ii. The main advantages of ODL and OQL include the following points:
  ODL and OQL can be used to bind other programming language, such C++, JAVA and Small Talk so that they are able to port object database systems from one programming environment to another and able to operate many object database systems together, even implemented by different languages, and potentially benefit to the access of global information resource through Internet. However, although OQL is relatively advanced in modern database management systems, it is not computationally complete and must be used in conjunction with ODL and other programming languages.
  Identify advantages and discuss limitations = **3 marks**
  Answers are precise and complete with clarity = **2 marks**
  Total 5 marks
b. The meanings of each provided ODL statement are:
  1. The class specified is for the Student object type (line 1)

      (1 mark)
  2. Student is a subtype of the object type Person_IF (line 1)

      (1 mark)
  3. The extend, i.e. the collection of instances, is called students(line 2)

      (1 mark)
  4. The attributes, student_ID, is a key, i.e. each instance of Student is uniquely distinguishable from other instances of Student by the value of student_ID (line 3)

      (2 mark)
  5. The state includes an attribute of type String<ccode_length>, called student_ID (line 5-6)

      (1 mark)
  6. The state includes a relationship between instances of Student and sets of instances of Enrolment (lines 7-9) - the relationship from Student to Enrolment to Student is called request_enroll (lines 8), and the inverse relationship from Enrolment to Student is called enrolled_by (line 9);

      (2 mark)
  7. The behaviour is an operation called request_enroll (lines 10-12), which takes an object of type List<EnrolLine> as its parameter, and returns an object of type Enrolment;

      (1 mark)
  8. The request_enroll operation raises an exception (line 12), course_doesnot_exit, in exceptional circumstances.

      (1 mark)

      Total 10 Marks
c. OQL specifies two partition attributes (which_city? and LessExpensive?), which are defined on functions (p.location = "Leeds", and p.averageprice <180,000). The partition attributes which_city? and LessExpensive? are of the types of the value returned by their respective partition functions, i.e. they are both Boolean. The objects which are grouped are property objects contained in the type's extent, properties.

Examiner's Comments

The minority of candidates (37.29%) answered Question 1 and made a reasonable attempt (average 12.55/25; standard deviation 4.31). "What can ODL and OQL be used for" has been answered well, but most students failed to mention the relationship with RDBMS. For Q1b it seems an easy part of this question as most students can explain the meanings of provided ODL statements. For Q1c, although most students have attempted, they did not link the concept of object database to the question. Overall, the question has achieved a reasonable good pass rate (75.00%).

Question 2

Database security is the mechanisms that protect database against threat.

i)   Breach of database security is a serious issue. List 5 typical situations where the database security has been breached; and for each of your case provide reason(s) of why the database security has been deemed as breached.

(7 marks)

ii)   For the following five specific problem areas (listed below): 1) give one example of the problem which has occurred in each area; 2) propose possible countermeasure(s) to rectify each problem that you have given in the five specific areas.
- Authorisation,
- Inappropriate access to data,
- Loss of or damage to data,
- Denial of service,
- Repudiation.

(12 marks)

iii)   What is the purpose of using the following Standard Query Language (SQL) syntax in database security control?

GRANT SELECT ON Student_tbl TO PUBLIC;
GRANT All On Student_tbl To former_student;

(6 marks)

INDICATIVE SOLUTIONS

Question 2: 25 marks. 7 marks for q2i, 12 marks for q2ii, 6 marks for q2iii.	
	a). The following situations can be considered that database security has been breached: Loss or destruction of data, or theft and fraud Modification of data Denial of service Errors in software Repudiation Loss of confidentiality (secrecy) Loss of privacy Loss of integrity Loss of availability Above points may be caused by unauthorised amendment, or copying of data; data corruption (as server); failure of security checks; unauthorised program alteration; and theft of programs. Each situation = **1 mark, 5 marks** for 5 or more situations Answers are precise and complete with clarity = **2 marks** **Total 7 marks**
b).	Examples in specific problem areas are: 1. Authorisation: Unauthorised access to server files,

Possible countermeasures: Encryption, firewalls, Privileges
2. Inappropriate access to data: Unauthorised reading/updating, introduction of viruses

Possible countermeasures: Privileges, views (problems with conflicting roles), Anti-virus software
3. Loss of or damage to data: Theft of disks, physical damage to equipment, data corruption

Possible countermeasures: Physical access restrictions, backup and recovery procedures, fault-tolerant hardware (e.g. RAID drives, protected power supply)
4. Denial of service: Spamming

Possible countermeasures: Monitoring traffic behaviour
5. Repudiation: Proof of customer's order or payment

Possible countermeasures: Trusted organisations; digital certification

Identify 5 problems in specific areas = **5 marks**

Propose possible countermeasure(s) for each area, and total 5 countermeasures = **5 marks**

Answers are precise and complete with clarity = **2 marks**

Total 12 marks

c). The statements GRANT SELECT privileges on the Students table hierarchy to everyone but restrict the privilege to make changes to the student data to users who are part of the former student role.

Explain privileges = **2 marks**

Explain restrict privilege = **2 marks**

Answers are precise and complete with clarity = **2 marks**

Total 6 marks

Examiner's Comments

This is the most popular question among candidates, about 77.12% candidates have attempted. Results have achieved the highest points in pass rate (89.01%) and average mark (13.85), and the lowest point in standard deviation (3.69) among the questions. It is clearly indicated that the impact of security issues must be addressed on advanced database management systems. In general there is no serious problem caused for this question.

Question 3

The increase in global communication through the internet results in a rise in the deployment of web services in distributed systems.

a) Describe the four key features of XML technologies (listed below) which are deployed for web services:
- Integration with the Web,
- Data can freely travel over theWeb,
- Strong Data independence,
- Simple Object Access Protocol (SOAP), Web Services Description Language (WSDL), and Universal Description Discovery and Integration (UDDI).

(13 marks)

b) Security for e-business over Internet is vital. Explain the meaning of the following concepts within the context of information security:
- Confidentiality,
- Authentication,
- Data Integrity.

(12 marks)

Answer Pointers

a) Some key features of XML technologies in web services are as follows:

XML integrates with standard Web protocols such as HTTP and FTP thus data defined by XML format can be delivered from server to server through HTTP or FTP. (2 marks)

Web services builds upon SOAP and HTTP so that the data can freely travel the web. (2 marks)

XML is strong in data independent so that data can be available to web connected platforms without considering to programmatic infrastructures. (2 marks)

The key technologies of Web services are XML based specifications: SOAP, UDDL and WSDL. SOAP is a protocol to be used to move information across the web; WSDL is a part of the web services framework. It uses XML syntax to describe the specifics of entering a web service; UDDL is a protocol to be used to register web services both privately and publicly. (2 marks)
Critical discussion = 2 marks
Total 10 marks

b. The basic requirements for e-business are: Confidentiality: information is not disclosed to unauthorised parties, such as individuals, groups, or processes.

Authentication: using login process to determine a message really comes from sender with proof, such as password.

Data integrity: the information arrives should be its original form without alternation after transaction process.

(1 mark each point)
Critical discussion = 2 marks

c. XML based security framework is:
XML Digital Signature,
XML Encryption/XML Decryption,
XML Key Management Systems
(3 marks)
Critical discussion = 2 marks

Examiners Comments

Part a) Many candidates seemed to have not kept informed and updated on latest technology Concerned with Web services and the Service Oriented Architectures that pervade the WWW. Many candidates still had fixed in their minds the concept of standard services supplied by dedicated unconnected web servers. A clue was the word 'distributed'. Candidates consequently struggled with the last part to the question a need to relate concepts in a) practice.

Question 4

Modern businesses are building **Data Warehouses** *or* **Data Marts** *and use OLAP (On Line Analytical Processing) - (either* **Relational OLAP** *or* **Multi-dimensional OLAP***) tools to access information regarding the organisation and its environment. However, building a data warehouse is not an easy task. The data population (also referred to as the Extract Transform-Load (***ETL***)) process is one of the most frequently cited reasons for delays in, or cancellation of, a data warehousing project.*

Explain, within the context of the above statement, the meaning of each term highlighted (in a bold font). Illustrate your answer with examples.

(25 marks)

Answer Pointers

(What follows is a summary answer – further exposition and discussion of these points expected, perhaps with reference to tools supported by a DBMS product such as SQL Server Analysis Services.)

Each term approximate to 5 marks

A **data warehouse** can be the main repository of an organization's historical data. It contains the raw material for management's decision support. The critical factor leading to the use of a data warehouse is that a data analyst can perform complex queries, analysis, forecasting and data mining, on the information without slowing down the operational systems.

A **data mart** (**DM**) is a specialized version of a data warehouse. Like data warehouses, data marts contain a snapshot of operational datathat helps analysts to strategize based on analyses of past trends and experiences. The key difference is that the creation of a data mart is predicated on a specific, predefined need for a certain grouping and configuration of select data. A data mart configuration emphasizes easy access to relevant information.

Operational database data re-modelled and re-loaded as multi-dimensional cubes in a separate data warehouse, which can support complex queries, analysis and perhaps also data mining.

On Line Analytical Processing – running complex analyses against cube models.

Extract, Transform, and Load (**ETL**) is a process in data warehousing that involves

- Extracting data from outside sources,
- transforming it to fit business needs (which can include quality levels), and ultimately
- loading it into the end target, i.e. the data warehouse.

ETL is important, as it is the way data actually gets loaded into the warehouse. The term ETL can in fact refer to a process that loads any database. ETL can also be used for the integration with legacy systems. Usually ETL implementations store an audit trail on positive and negative process runs. In almost all designs, this audit trail is not at the level of granularity which would allow to reproduce the ETL's result if the raw data were not available.

Examiners Comments

This question was popular and attracted a lot of recall from textbooks in a large number of cases. The best candidates could see the connection between the various terms/components and could discuss the process or work flow through a development of a data warehouse. For example the pivotal role of ETL. Very few candidates could cite real development experience of this process and consequently lost some value in their answers.

Question 5

a) State the four essential properties of a transaction that ensure data reads and writes do not compromise the data integrity of a database.

(4 marks)

b) Explain how these properties would be applied given the following Transactions that have been submitted to a DBMS engine concurrently from different user sessions: Session_1 and Session_2. Give examples to illustrate your answer.

(8 marks)

Session_1: user 1 issues a request to execute the following transaction

USE DATABASE ORDERS
BEGIN TRANSACTION
UPDATE Customers SET ContactName = 'Bill Smith'
WHERE CustomerID ='ABC'
SELECT ContactName FROM Customers WHERE CustomerID = 'ABC'
ROLLBACK TRANSACTION
SELECT ContactName FROM Customers WHERE CustomerID = 'ABC'

Session_2: user 2 issues a request to execute the following code

USE DATABASE ORDERS
BEGIN TRANSACTION
SELECT * FROM Customers WHERE CustomerID = 'ABC'
UPDATE Customers SET ContactName = 'Bill Smith'
WHERE CustomerID = 'ABC'

c) Explain the affect of each of the following statements in the above transactions.

SET TRANSACTION ISOLATION LEVEL SERIALIZABLE
SET TRANSACTION ISOLATION LEVEL READ COMMITTED

(6 marks)

d) Describe the range of tools and mechanisms that a Database Administrator of an e-commerce OLTP (On Line Analytical Processing) type of application can use to trade-off the need to achieve both high performance (via fast throughput of transactions) against maintaining data integrity.

(7 marks)

Answer Pointers

(a)
ACID = Atomic, consistent, isolated, durable.

(b)
In Session1 the whole transaction is treated as the smallest recoverable unit and is thus atomic (not possible to break down further or break out.
The Begin and End Trans ensures isolation form the effects of other transactions. The consistency is achieved by allowing Rollback of failed transactions so that the state of a process is known and hence consistent. Durable transactions are supported by Recovery beyond the normal Rollback this means that transaction can be serialized – ie can be run as a serialized schedule.   Examples from the code are plenty and should be referenced.

(c)
The transaction isolation levels define the type of locks acquired on read operations. Shared locks acquired for READ COMMITTED are generally row locks, although the row locks can be escalated to page or table locks if a significant number of the rows in a page or table are referenced by the read. If a row is modified by the transaction after it has been read, the transaction acquires an exclusive lock to protect that row, and the exclusive lock is retained until the transaction completes.

SERIALIZABLE   Specifies that: the statements contained within a transaction cannot read data that has been modified but not yet committed by other transactions and no further transactions can modify data that has been read by the current transaction until the current transaction completes.This blocks other transactions from updating or inserting any rows that would qualify for any of the statements executed by the current transaction. Consequently concurrency is lower with this option but isolation is the strictest.

(d)
The range of tools cover techniques such as de-normalisation and redundancy mechanisms such as Views Temporary tables. Other techniques to consider are database mirroring, fail over servers, distribution by federation ie scaling up.

Tools such as profiling or monitoring tools to check the effect of critical transactions; indexed views and other optimization techniques. Many of these tools have graphical representation of monitoring variables such as no of page locks, disk I/O, caching efficiency for example.

Examiners Comments

In parts a and b candidates seemed to have grasped the concept that isolation refers to WRITE operations on the database but many could not appreciate the need to relax READ operations level and relate these to part c).. The setting of an isolation levl can b e used to allow interleaving of transactions so that reads can be uncommited but a consistent state is maintained at the database level. Not sure if the majority of candidates realsed this fact judging by their answers to part c).

Part d) needed a justification of choice of tool or technique for a givn situation perhaps practical knowledge or experience would help give a fuller more reflective answer.