THE BCS PROFESSIONAL EXAMINATION
Professional Graduate Diploma

April 2002

EXAMINERS' REPORT

Advanced Database Management Systems

Question 1

"SQL is very far from being the perfect relational language". (C. J. Date 2000)

With reference to the way SQL supports the relational model, justify this statement. (25 marks)

Answer Pointers

To answer this question fully candidates must define what "supporting the relational model" means. A good answer would cover:

- Support for relations as the only data storage structure
- Support for primary keys with the stipulation that no part of a primary key may hold a null value (entity integrity).
- Support for foreign keys with the stipulation that a foreign key value must match a primary key in the database.
- Full support for all relational algebra operations.

- SQL does not support relations fully. In a relation every row must be unique. In SQL duplicate rows are permitted. SQL supports tables but not relations.

- SQL supports the concept of primary keys but does not enforce their use (duplicate rows problem). Where primary keys are nominated then entity integrity is supported.

- SQL now has full support for referential integrity although this was absent in earlier versions.

- Whilst all relational algebra operations can be achieved in SQL there are some that cannot be achieved easily. Natural join is a good example. Many implementations of SQL fail to implement parts of the standard that provide full coverage of relational algebra. Relational assignment is not in the standard

Question 2

Many server-side scripting languages (e.g. JSP, PHP and ASP), which are used to construct Web-enabled database applications, allow developers to open and close "sessions". Explain the importance of this facility and compare and contrast two mechanisms that are used to identify sessions. (25 marks)

Answer Pointers

Candidates should be aware that whilst HTTP is a stateless protocol, most database access is stateful.

Most databases require that users identify themselves to the database management system. In a stateless protocol every database action would require user authentication and this would clearly be inefficient.

Where transactions are required clearly each HTTP request must be related to the transaction it applies to. This is achieved via a session. At the start of transaction the code that is communicating with the database begins a session (usually identified with a session number). State information is then recorded for each session. At the end of a transaction the session is closed and the state information is discarded. Users must somehow transmit the session number to the server.

Two mechanisms for implementing sessions are cookies and URL re-writing. Cookies are directly supported by HTTP. In the case of databases a per-session cookie is normally used. The cookie is issued to the client when the session is created and returned in each HTTP request. The server can request the cookie at any time. There are concerns over privacy breaches caused by the use of cookies and at present the EU is considering regulating their use. In URL re-writing every page returned to the client is produced dynamically and each URL in the page has the session number appended to it. When these URLs are requested the server is able to determine which session the request applies to. This mechanism is much more cumbersome and less transparent to the user but it will work for those users who have turned cookie acceptance of in their browsers.

Question 3

"The SQL 3 standards activity will deliver an SQL which incorporates object-relational features." (Jackson 1999)

Discuss the features in the SQL 3 (now SQL/99) standard that achieve this. (25 marks)

Answer Pointers

The key issues in object-relational databases are:

- support for base type extension in an SQL context;
- support for complex objects in an SQL context;
- support for inheritance in an SQL context;
- support for a production rule system.

SQL/99 includes support for generated types, distinct types and structured types.

Structured types can be regarded as complex objects that is objects which are built out of variables of other types.

Structured types may be created via inheritance and SQL/99 supports subtables and supertables.

SQL/99 supports the CREATE TRIGGER statement which is a limited form of production rule.

Question 4

a) In query processing, one of the most expensive operations is the join. Various alternative implementation strategies are possible for the join operation. Discuss three alternative strategies and illustrate the difference that each strategy makes to the number of disk accesses required.

Assume the following relations in your illustration (primary keys are underlined).

r1 (A, B, C)
r2 (C, E)

r1 has 40,000 tuples, and 40 tuples fit into one physical block
r2 has 25,000 tuples, and 25 tuples fit into one physical block (18 marks)

b) Explain why some relational algebra expressions are more efficient than others. (7 marks)

Answer Pointers

a) Strategies

Possible strategies that can be discussed are :

- Nested Loop
- Indexed Nested Loop
- Block-oriented Nested Loop
- Hash-join
- Merge-join

For each of the three strategies selected for discussion:

Up to 3 marks for knowledge and understanding of strategy algorithm
Up to 2 marks for appropriate calculation of block reads

Up to 3 marks for lucid comparison of alternative strategies.

(18 marks)

b) Equivalence of Relational Algebra Expressions

Candidates should state and discuss the following rules for efficiency

- Perform selects early
- Perform projects early
- Use the commutivity and associativity rules of joins to reduce the size of intermediate relations

Up to 2 marks for coverage of each of the three operations (select, join, project), plus 1 mark for lucid discussion

Examiner's Comments

This was not a very popular question and also not handled particularly well by those candidates who selected it. Some candidates answered only part of the question and thus lost marks. Some candidates answered the question wrongly in that they discussed a complete query plan (i.e. select -> project -> join) when only a consideration of join strategies was required. Another common mistake was assuming a distributed infrastructure and working out a strategy based on moving data between sites. This also was not required

Question 5
Distributed database is a contradiction in terms. Discuss. (25 marks)

Answer Pointers
A candidate's argument may be along the following lines:

Distributed database may seem a contradiction in terms as the database is no longer a physically co-located collection of files. However conceptually the physically distributed collection of files is still seen as a single database.

To show this, the candidate may describe and discuss the well-known objectives and/or desirable features of distributed database:

- Local Autonomy
- Reliability
- Efficiency and flexibility
- Capability for incremental growth

- Location transparency
- Fragmentation transparency
- Replication Transparency
- Distributed Query Processing
- Distributed Transaction Processing
- Hardware and Systems Independence

Up to 4 marks for understanding shown of why possible contradiction of terms may or may not be seen.

Up to 5 marks for description of motivation for distributed database (first group of objectives/desirable features in the above list)

Up to 16 marks for understanding of issues involved with the second group of objectives above.

Location/Fragmentation/Replication Transparency	6 marks
Query/Transaction Processing	8 marks
Hardware/System Independence	2 marks

Examiner's Comments
This was a popular question and candidates mostly answered well. Few candidates got a fail mark on this question. In general marks were lost by leaving out one of the areas, providing sketchy coverage or straying off-track.