BCS THE CHARTERED INSTITUTE FOR IT

BCS HIGHER EDUCATION QUALIFICATIONS
BCS Level 6 Professional Graduate Diploma in IT

ADVANCED DATABASE MANAGEMENT SYSTEMS

EXAMINERS REPORT

September

Answer **any** THREE questions out of FIVE. All questions carry equal marks.

The marks given in brackets are **indicative** of the weight given to each part of the question.

Calculators are **NOT** allowed in this examination.

Section A

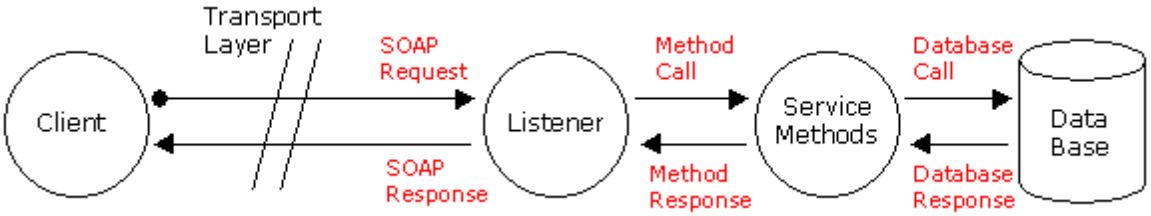
A1

EXAMINER'S GENERAL COMMENTS

Relatively few candidates attempted this question. The average mark was slightly below expected as the subject matter is frequently covered in this exam. There was a good range of marks and evidence of increasing knowledge of this emerging application of database technology to the World Wide Web.

Consider the following diagram (Fig A1) then answer the question parts a) and b) that follow.

Fig A1 Service Oriented architecture (SOA) supporting a client server database application



a) Briefly describe a database application that would benefit from using the above system architecture.

(5 Marks)

Answer Pointer

This requires an application that can benefit from a 'Service oriented' architecture typical of Cloud computing and most dynamic web based applications. Thus an e-commerce application could be applicable. Hence some description is required of the range of databases that might be connected to different services such as products held across different servers to meet a consumer purchase request. The ability to select different services and switch database providers from anywhere on the web in a secure fashion is a fundamental principle of this architecture.

EXAMINER'S COMMENTS

Candidates needed to outline in a sentence any application that deployed a centralised database server that could or actually did benefit from SOA. The role the database played in the application was not as important as the impact this architecture had. Some candidates misunderstood the term 'application' and related it to the software tools that are used to construct a SOA.

b) Describe the function and purpose of the above schematic diagram explaining how the various components support the database application that you have described in your answer to part (a). Include example code related to your application.

(10 Marks)

Answer Pointer

Different software systems often need to exchange data with each other, and a Web service is a method of communication that allows two software systems to exchange this data over the internet. The software system that requests data is called a service requester, whereas the software system that would process the request and provide the data is called a service provider.
Each component's role in supporting the chosen application could include source data and example transactions that convey information across the WWW and how this operates in practice. The main keywords SOAP, Client and Service define the specific architecture. Hence a description of a database application that needs the support of this architecture should be described. The architecture is loosely distributed. The web n-tier client server is based on the early ideas behind CORBA and now more recently founded on web services and data exchange by messages using XML or JSON as the transport mechanism.
Example generic code (with emphasis on XML) should illustrate examples of :

- How one system can request data from another system?
- Which specific parameters are needed in the data request?
- What would be the structure of the data produced? Normally, data is exchanged in XML files, and the structure of the XML file is validated against an .xsd file?

Simple/Indicative Code samples would include
XML data file
XML schema
SOAP message eg

```
<?xml version="1.0"?>
<soap:Envelope xmlns:soap="http://www.w3.org/soap-envelope">
  <soap:Header>
  </soap:Header>
  <soap:Body>
    <m:GetStockPrice xmlns:m="http://www.example.org/stock">
      <m:StockName>IBM</m:StockName>
    </m:GetStockPrice>
  </soap:Body>
</soap:Envelope>
```

EXAMINER'S COMMENTS

Candidates were able to demonstrate a fairly good understanding. The best answers applied the technology to the candidate's chosen application. The poorest answers were from candidates that tended to produce lengthy answers, recalling only generic information on the topic.

c) Discuss the main challenges that a service oriented architecture has on a database management system. **(10 Marks)**

Answer Pointer

Some discussion of the following is expected :

Managing metadata (data about data) in SOA-based environments can include many services that exchange messages to perform tasks. Depending on the design, a single application may generate millions of messages. Managing and providing information on how services interact can become complex. This becomes even more complicated when these services are delivered by different organisations within the company or even different companies (partners, suppliers, etc.).

Another challenge involves the lack of testing in SOA space. There are no sophisticated tools that provide testability of all headless services (including message and database services along with web services) in a typical architecture. Lack of horizontal trust requires that both producers and consumers need to test services on a continuous basis. Therefore, it is important to invest in a testing framework that would provide the visibility required to find the culprit in the architecture.

A further challenge relates to providing appropriate levels of security. Security models built into an application may no longer suffice when an application exposes its capabilities as services

that can be used by other applications. That is, application-managed security is not the right model for securing services.

EXAMINER'S COMMENTS

Most candidates could identify security and loss of loss of autonomy to some extent. But a lack of knowledge was demonstrated.

A2

EXAMINERS' GENERAL COMMENTS

The syllabus covers "Distributed relational systems and Data Replication". Candidates are expected at this level to show and apply concepts and techniques of data distribution to real world situations and to applications that encompass distributed databases. The results show the average mark is slightly above the pass mark and the range of marks was as expected.

Assume a journal library database holds the following Tables (with sample data) at a central site on a database server called (CTR)

JOURNAL

JournalID	JournalName
3215	Database Weekly
3216	Database Journal
3217	Oracle News
3218	ACM TODS

ARTICLE

ArticleID	ArticleTitle	AuthorID	JournalID
3215	ObjectOriented Analysis	23	3216
2409	Oracle indexing	18	3217
1398	DBA performance tools	23	3216
1289	Pioneers of Databases	23	3215
2554	Query optimisation	67	3216
1678	Daplex	18	3218
4561	Niam Fact model	18	3218

AUTHOR

AuthorID	AuthorFName	AuthorSname
23	Norman	Gray
18	Carlos	Santos
67	Modrich	Neuman

LOAN

ArticleID	BorrowerID	LoanDate	ReturnDate
2409	43	3/1/15	4/2/15
1398	43	3/1/15	24/1/15
1289	17	6/2/15	8/2/15
2554	26	1/2/15	12/2/15
2409	43	14/2/15	
2554	52	14/2/15	

BORROWER

BorrowerID	BorrowerFname	BorrowerLname	BorrowerTelNo
A52	Jane	Green	0156387562
43	Fred	Briggs	01985-86722
17	Henry	Dhura	01582-74238
26	Jonas	Smith	01933-632001

The above tables are currently held on a database at the single site called CTR.
Following reorganisation it is intended to distribute the journals held at the central database across 3 new branch libraries located at remote sites called GTR; UTC and TWE. The central library becomes the HQ (an administrative centre) meaning that it no longer keeps or loans out any journals itself. Instead journals are made available for loan to borrowers registered at any of the 3 new sites.

a) Briefly Describe in general terms THREE different approaches of achieving database distribution.

(6 Marks)

Answer Pointer

2 marks each
(i) Full transactional replication. The database should always be available to all sites and access to it should be fast even in the case of data communication failure. But part, or all, of the database is copies redundantly across sites
(ii) Fragmentation. There should be no redundancy in the allocation of data, i.e. only local data should be stored at a site. This might mean splitting tables horizontally or vertically.
(iii) Balanced or a mix of the above e.g. Merge Replication (updateable distributed transactions) where fragments are redundantly distributed as replicas over different sites.

EXAMINER'S COMMENTS
There was general satisfaction with the range of answers produced. Most candidates identified at least 2 options.

b) Describe in detail THREE different proposals for data distribution of the central database (CTR).
Hint: Show the distribution/replication of table fragments/partitions and explain any trade-off and pros/cons you think are relevant.

(12 Marks)

Answer Pointer

A reasonable set of proposals could be for 4 marks each:

Option1 :
Replication For example a master relation common to all sites e.g. the Loan relation is replicated.
Advantages: Availability and fast access, but there are update propagation issues and a lot
of storage is required.

Option 2:
The Journal and Article relations are horizontally partitioned into three fragments, which are stored one at the UTC branch, and one at the GTR and one at the TWE branch. The main CEN branch headquarters does not store any data. Fragment UTC = s (Branch="UTC") and
Fragment TWE =s (Branch="TWE"). Advantages: No redundancy of data, easier maintenance and security, efficient use of physical storage.
Disadvantages: Data becomes unavailable as a site or links to it fail.

Option 3:
The relation is horizontally fragmented and allocated to the UTC,TWE,GTR branches as above, but a copy of the whole relation is also allocated to the branch CEN Headquarters.

EXAMINER'S COMMENTS

Most candidates concentrated on physical fragmentation (options 2 and 3) the more theoretically sound approaches. Replication (option1) is a more practical proposal. There is evidence that candidates failed to demonstrate knowledge of replication techniques and might benefit from reading up on literature on how replication is supported by the major database vendors

c) Describe the criteria you would use to assess the effectiveness of your approach.

(7 Marks)

Answer Pointer

Criteria include handling of distributed transactions, software tools to facilitate replication, fragmentation and partitioning. Replication introduces higher maintenance overhead and includes redundancy and a possible loss of consistency, but the trade- off is higher availability.
Other criteria include integrity; security; access control; performance and locking

EXAMINER'S COMMENTS
A poorly answered question, with a large number of candidates failing to provide any significant analysis.

A3

EXAMINER'S GENERAL COMMENTS

This was a very popular question with some excellent attempts. A very high pass rate was achieved.

a) Given the following three linked tables:

```
Customers (custID, name, address)

Products (prodID, price)

Orders (orderID, custID*, prodID*, date)
```

And the following query:

```
SELECT Customers.name
FROM Customers, Orders, Products
WHERE Customers.custID = Orders.custID
AND  Orders.prodID = Products.prodID
AND Orders.date = '15-Jan-14'
AND Products.price > 100;
```

Draw a query tree that corresponds to the most efficient way of processing this query.

(11 Marks)

(b) Database security aims to minimise the loss of:

- Data Confidentiality
- Data Integrity
- Data Availability

(i) Describe each of the above concepts **(6 Marks)**

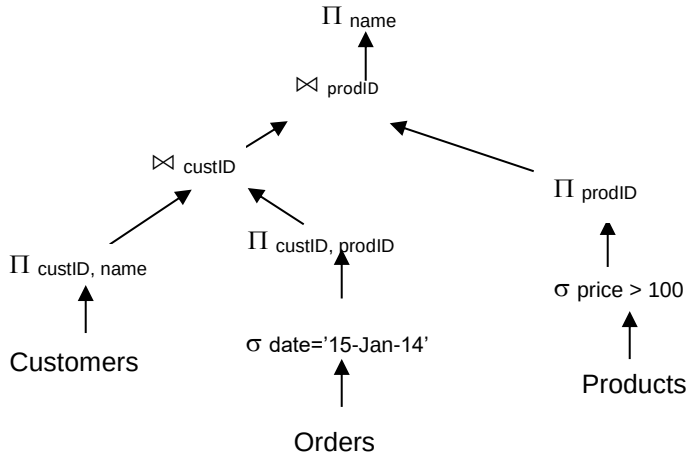
(ii) Explain how each of the following mechanisms contribute to providing security for a database:

- Authorisation
- Views
- Backup and Recovery
- Encryption

(8 Marks)

Answer Pointer

a) (11 marks: 1 for each operation (total of 8) + 3 for disposition of tables)



EXAMINER'S COMMENT:

Most candidates managed to correctly draw a sensible query tree. A few confused the symbols used for projection, selection and join.

(b)

(i) (2 marks each)

Confidentiality: The property that information is not made available or disclosed to unauthorised users or processes

Integrity: The property of maintaining and assuring the accuracy and consistency of data. This is usually enforced via integrity constraints.

Availability: the property of ensuring that data is accessible and usable upon demand by an authorised user.

EXAMINER'S COMMENTS

Most answers show a good understanding of database security concepts. The only exception was the confusion of authorisation (roles and privileges) with authentication (login and password).

Section B

B4

EXAMINER'S GENERAL COMMENTS
This was the most popular question and many candidates achieved very high marks.

(a) Using your own simple examples and/or diagrams, describe the potential data integrity and consistency problems that may occur in a *multi-user* database if *concurrency control* techniques are not fully implemented. You should comment specifically on:

(i) Transaction scheduling techniques between parallel database operations

(5 Marks)

(ii) The use of time-line diagrams to model two or more transactions accessing the same database

(5 Marks)

(iii) The distinction between, and issues caused by, accessing committed and uncommitted data

(5 Marks)

Answer Pointer

Marked holistically, but things to cover include: the lost update problem, dirty (uncommitted) read problem, non-repeatable reads & phantom reads. For each concept, a description of the problem is required along with a simple time-lapse example involving two transactions T_1 and T_2 is needed for full marks. Referral to the different types of schedule – serial and interleaved – and the impact on these problems by these concepts is also expected. Good diagrams gain bonus marks.

EXAMINER'S COMMENTS

This sub-question was generally well answered with the vast majority of candidates showing good appreciation of the major potential problems – lost updates, dirty reads etc. Most candidates then went on to provide well annotated time-line diagrams. Overall, a good question for the majority of those attempting it.

(b) In your own words, describe what is meant by the following terms:

(i) Two-Phase Locking

(3 Marks)

(ii) Serializability

(3 Marks)

You should elaborate on how the former ensures the latter. You should supply any suitable examples and/or diagrams that you deem appropriate to support your answer.

Answer Pointer

Two-phase locking should mention the concepts of a transaction having two distinct stages: a 'growing' phase where it acquires all locks necessary for it to complete its tasks (and cannot release any locks) and a 'shrinking' phase in which it systematically releases those locks and returns the data resources as it runs down (and is not allowed to acquire any new locks). In other words, 2PL ensures that a transaction must acquire a lock on a data item before doing any work on that data and once a transaction has finished with a lock and cannot grab more locks. Serializability is the idea that parallel transactions can execute concurrently - via interleaving (using a non-serial schedule) - yet without interfering with one another – so as give a final outcome on the database the same as if those transactions had been executed in a sequential (serial) manner. 2PL thus stops two competing transactions from colliding over the same data item(s) and thus violating the ACID principles.

EXAMINERS COMMENTS

This question was universally well answered with extensive and detailed responses, mostly with excellent diagrams clearly highlighting the growing/shrinking phases. There were many excellent answers and almost every candidate gained full marks.

(c) *Briefly* describe the various locking options available to a DBMS as part of the concurrency control function, paying particular attention to the granularity of the locking, the nature of the locks and the allowable operations and what happens when deadlock occurs. You should supply any examples or diagrams as you deem suitable (hint: think about the performance issues of different lock types too).

(4 Marks)

Answer Pointer

Marked holistically but good responses will cover locking as a pessimistic mechanism, row-level locking versus table-level locking, shared (read) locks versus exclusive (write) locks, the need for tracking all these locked data resources via a locking schedule and lock table, the impact on performance of over-locking (tying up large chunks of data that needs to be updated) or conversely having too many small locks and the associated overhead of managing the lock table. Lastly, candidates should state that deadlock is the situation whereby transactions have mutually locked the data needed by each other – thus leading to a temporary impasse.

EXAMINER'S COMMENTS

There were many excellent answers and almost every candidate gained full marks.

B5

EXAMINER'S GENERAL COMMENTS
An extremely popular question with a very high pass rate.

(a) Explain, using your own simple examples where appropriate, what each letter ('E', 'T' and 'L') of the term 'ETL' means with respect to data warehouses, taking care to highlight common problems or issues in each stage.

Two marks per letter.

(6 Marks)

Answer Pointer

*E (Extract) – To c*opy data from operational and other external sources like OLTP databases, spreadsheets, websites etc*. T (Transform) – To s*elect/reject, clean, re-format, order, (dis)aggregate data into a consistent structure and regular format. Issues could include what to do with missing (NULL) entries, standardizing on one format (for example 'M/F', 'Male/Female' for Sex), how many fields to use for the Address data or application of business rules. *L (Load) – To c*opy the selected and cleaned data into the data warehouse at regular intervals Specific examples gain bonus marks.

EXAMINER'S COMMENTS

Again, plenty of good ideas presented for this question.

(b) Describe *at least three* different methods or tools that an end-user may use to interact with a data warehouse, briefly highlighting the primary purpose of each (hint: how different *categories* – not specific products - of tools are used for different data extraction purposes).

Two marks each.

(6 Marks)

Answer Pointer

Major categories are: reporting tools (end-user desktop), query tools (SQL development and QBE interfaces), application development tools (for more advanced and regular tasks), EIS tools (decision-support), OLAP tools (multi-dimensional queries) and data mining tools (identification of new, unpredicted patterns and trends).

EXAMINER'S COMMENTS
The vast majority of candidates scored very heavily on this question. It was abundantly clear that they were very familiar with 'ETL'.

(c) Briefly describe the following data warehouse features:

(i) Summary Management **(3 Marks)**

(ii) Analytical Functions **(3 Marks)**

Answer Pointer

Summary management – DW queries are often seeking aggregated data, not the fine detail of individual rows, particularly aggregation via specific dimensions (month, product, region etc.) so the DBMS must support pre-computed summaries and aggregates to avoid run-time computation. *Analytical functions* – many BI and DW applications want to use SQL ranking and aggregate functions, cumulative aggregates or maybe the CUBE and ROLLUP operators for OLAP analysis.

EXAMINER'S COMMENTS

Not as well answered as the two previous sub-questions with some vague responses – relatively few got into details about specific functions or the concepts behind data aggregation. That said, those that did gave detailed answers.

(d) Using simple examples and/or diagrams of your own choosing, explain and demonstrate the similarities and differences between:

- Entity-Relationship Model s
- Star Schemas
- Snowflake Schemas

You should particularly address the roles of primary and foreign keys plus the role of normalized and de-normalized data on determining the number of dimensions in a given model.

(7 Marks)

Answer Pointer

This will be marked holistically with bonus marks for clear diagrams of the various models but the key points to be covered are: *dimension modeling* as a specialized example of an ER Model – both based on entities and relationships, the use of a *fact table* (with composite primary key) and a set of *dimension tables* (each with an atomic primary key) related to the fact table via foreign keys – thus producing a *star schema* (star join) model. The better candidates should then go on to discuss issues such as the fact table is much larger than the dimension tables, that the fact table works best when the 'facts' recorded are numeric (grades, prices, ages etc) thus allowing aggregated computations to be run leading to summarized data, that dimension tables tend to hold more descriptive data (names, addresses, identifiers etc), the use of de-normalized data to replicate attributes across multiple dimension tables (for example, storing address or contact data in several different dimension tables) thus avoiding additional joins and enhancing query performance. Finally, a few words on what a *snowflake schema* is (where dimensions can have their own dimensions) – caused by normalizing the original dimension table data down into two or more child dimension tables, all linked to the 'parent' dimension table via the familiar PK/FK technique. So, *star schemas* use de-normalized (repeated) data and *snowflake schemas* use normalized (minimized duplication) data.

(ii) (2 marks each)

- Authorisation: The granting of a right or privilege that enables a user to have legitimate access to a system or database object. This mechanism provides access control for the database by specifying what each database users can do/see.
- Views: a virtual relation that is the result of a query on the database. It provides a security mechanism by hiding parts of the database from certain users (user is given access to view instead of base table).
- Backup and Recovery: the process of periodically copying database to offline storage media. This will assist with the recovery of the database following a failure or security incident and contributes to preserving the availability of the database.
- Encryption: the encoding of the data by a special algorithm that renders the data unreadable by any program without the decryption key. This is used, in particular, to protect the confidentiality of sensitive data.

EXAMINER'S COMMENTS

This question was well answered by most candidates with many quality diagrams. A few did badly on this question, but many did extremely well – covering PK/FK, fact tables, dimension tables and normalization in appropriate fashion.