THE BCS PROFESSIONAL EXAMINATIONS
Diploma

October 2005

EXAMINERS' REPORT

Database Systems

Question 1

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.

a) Data Integrity
b) Data Security
c) Data Redundancy
d) Data Maintenance
e) Data Consistency **(5 x 5 marks)**

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.

Examiner's Comments

It was pleasing to note that, almost without exception, this question was very well answered – with most students giving 'textbook' responses to the issues raised in the question. There was obvious dependency/redundancy between some of the topics but most students managed to express their views clearly and with minimum repetition. All round, a successfully answered question (and certainly the most popular one – with only a handful of students not attempting it).

Question 2

2. Using your own SQL code examples and any appropriate diagrams, discuss and explain all the relevant concepts and constructs that must be addressed and implemented by a database developer or administrator when satisfying the following database issues:

a) Security constraints (on users) **(15 marks)**
b) Integrity constraints (on data) **(10 marks)**

Answer Pointers

This was marked holistically but students will be expected to cover the following topics. Good examples and diagrams will gain extra credit.

Security

- Access control – passwords (environment and database)
- Views – to restrict data visibility
- Privileges – system-level and object-level (GRANT & REVOKE)
- Roles – as a privilege-management tool
- Use of data dictionary and meta-data
- Profiles – to limit access to database resources
- SQL examples

Integrity

- Domain Integrity
    - Creation of user-defined domains
    - Use of NULL / NOT NULL
    - Use of CHECK constraints
    - Use of default values
    - SQL examples
- Entity Integrity
    - Selection of primary keys
    - Uniqueness criteria
    - No duplicates
    - No NULLs
    - SQL examples
- Referential Integrity
    - Selection of foreign keys
    - Allowance of NULLs
    - Presence of referenced value in the referenced table
    - ON DELETE CASCADE
    - SQL examples

Examiner's Comments

Again, this was generally very well answered (although by no means as popular). The vast majority of students who attempted it covered the core issues very well – although many failed to capitalize on their undoubted knowledge by not providing any/enough good actual examples – as requested by the question. All that said, it was clear that most students have a very good grasp of the key database themes raised in this question.

Overall Thoughts (Q 1 & 2)
The general standard of written English was very good and most students expressed complex technical points clearly and with conviction. The quality of most answers and many of the provided SQL examples leaves me in no doubt that, on the basis of these two questions, database knowledge is alive and well at these BCS centres.

Question 3

3. Explain the purpose of the ANSI/SPARC 3-level architecture for a database management system, and describe the functions of each level giving examples to illustrate your arguments. **(25 marks)**

Answer Pointers
A major aim of the database is to provide the user with an abstract view of the data
Hiding certain details of how data is stored and manipulated. Abstraction is used as the starting point for the design of a database for a given organisation. Different users may have different views of the data held in the database

To achieve abstraction and the variety of views required for the design of database systems, the ANSI/SPARC 3-levels architecture is provided in most available commercial DBMS. The three levels are:

External Level
This level describes that part of the database that is relevant to a particular user,
A users' view of the database. The view might be a subset of the data stored in the database or a virtual data different real views of the data stored in the database e.g. the date as (day, month, year) or as (month, day, year) virtual view such as the age that is calculated from the date of birth stored in the database.

Conceptual Level or Schema
This level describes what data is stored in database and relationships among the data.
It is Community view of the database. It contains the logical structure of the entire database e.g. entity, attribute and constraint information. It does not contain any storage-dependent details. It supports each external view i.e. any viewed data must be contained/derived from the conceptual level

Internal Level or Schema
This level describes how the data is stored in the database i.e. physical representation of the database on the computer. It covers the data structure and the file organisation
It interfaces the operating system access methods. It is used for accessing and retrieving data records. It is concerned with storage space allocation, record description for storage, etc...

Examiner's Comments
Most of the students have answered this question quite well especially the last sub question the explanation of each level (external, conceptual and external) of the **ANSI/SPARC 3-level architecture.** However, very few have explained why the need for this architecture. The examiner suggests that candidates should 'read around' topics such as this and also go through real example to appreciate the need for such architecture that realised abstraction and reduced complexity in dealing with database applications.

Question 4

4. Consider the following relations (primary keys are underlined):
AUTHOR (AName, name, address, speciality)
PUBLISHER (PName, name, Location)
BOOK (Title, AName, PName)
Where
AUTHOR contains author details and **AName** is the primary key.
PUBLISHER contains publisher details and **PName** is the primary key. BOOK contains details of the book and its primary key is title **(Title)**.

Write relational algebra statements for each of the following queries:

a) What are all titles published by *Pitman*?
b) What is the speciality of all authors publishing a book with *MIT Press*?
c) What is the location of the publisher of the book '*A guide to DB2*'?
d) Get the names of all publishers publishing a book by *Smith* and a book by *Jones*?
e) What are the addresses of all authors publishing a book with all the publishers who located in *Paris*? **(5 x 5 marks)**

Answer Pointers

The student could either use relational algebra notation such as Π, σ etc... or English words Project, select, Join, …

(a) What are all titles published by *Pitman*?

Project(Select BOOK (PName='Pitman') (Title)
OR
Π_{title} ($\sigma_{PName='Pitman'}$ BOOK)

(b) What are the specialisms of all authors publishing a book with *MIT Press*?

Project (Join (Πt(Select BOOK($\sigma_{PName= \text{ 'MIT Press'}}$)(AName), AUTHOR:[AName=AName]) (Specialism)
OR
$\Pi_{Specialism}$, Π_{AName} ($\sigma_{PName='MIT\ Press'}$ BOOK) $\bowtie$ BOOK. $\sigma_{AName=AUTHOR.AName}$ ($\Pi_{AName,\ Specialism}$ AUTHOR))

(c) What is the location of the publisher of the book '*A guide to DB2*'?

Π (Join(Π($\sigma_{(Title= 'A Guide to DB2')}$ BOOK) (PName)), PUBLISHER[PName=PName] (Location)

(d) Get the names of all publishers publishing a book by *Smith* and a book by *Jones*?

Divide (Π BOOK (ANname, PName), D:[AName | AName] where D(AName) is a relation containing two tuples-values Smith and Jones.

(e) What are the addresses of all authors publishing a book with all the publisher located in *Paris*?

Project (Join (Select (Join (Project BOOK (AName, PName)), PUBLISHER : [PName=PName]) (Location= 'Paris')), AUTHOR: [AName=AName]) (AName, Address)

Examiner's Comments

Most of the students have answered correctly the simple queries (a) and (b). A large number of these students have answered their questions in SQL rather than relational algebra expressions. However, a small number of students have answered correctly the complex queries (c), (d) and (e). The examiner suggests that candidates should have the opportunity to practice more with complex queries and be able to write the answer in both SQL and relational algebra.

Question 5

5. Read the following scenario then attempt the task that follows.

Scenario

You are a database consultant and you have been given a contract to build a corporate database for a large manufacturing company. The company has many departments and each department has a database to support its own data processing requirements. The local databases that run in these departments are a mixture of flat file databases and relational databases.

You are required to merge all the local departmental databases into one centralised database. A centralised approach is seen as a way of controlling the duplication of data and resources. It is also seen as an opportunity for the company's directors to develop an improved management information system so they can analyse company wide information such as resource planning and comparative performance of departments. A phased approach to development is recommended to minimise the loss of service to the existing systems.

Task

Produce a report for the technical director of the manufacturing company that will include a description of the phases of development and outline any risks associated with each phase. **(25 marks)**

Answer Pointers

The phases may consist of:

System definition (revised to include User Views)
Requirements collection and analysis
Define/Elicit Business Rules
Conceptual Database design (merge Views)
DBMS selection (optional)
Application design (reverse engineering existing applications)
Prototyping (optional)
Implementation
Data conversion and loading
Testing
Tuning and Operational maintenance

Candidates should not simply regurgitate bookwork but carefully apply their answers to the scenario. For example it is important that MIS represents accurate interpretation of Business Rules an example. Candidates should explain the significance of VIEWS expressed in department databases and the preservation of these Views (a risk is that these are not captured and hence some functionality is lost). Other risks include parallel working alongside existing and target databases – when should the switch be made or do we keep 2 copies synchronised? Is there another risk of who owns what information now that the database is centralised it might de-motivate those users who cherished working on the old local databases. Security and integrity issues?

Examiner's Comments

This was not a popular question but it certainly produced a wide range of answers. Unfortunately many answers were simply memory recalls from text books and not pertinent to the answer. Also slightly worrying is that some answers did not reflect the modern trends of incorporating databases into business intelligence or the incorporation of business rules. This means apart from reading old text books some candidates could not relate to any experience they have of IT businesses and database support. The strong point of many candidates' answers was the realisation of phased development and it was pleasing to see diagrams reflecting a plan of development even though the detail of how the phases unravelled was missing.

Question 6

6. Read the following discourse then answer the questions that follow.

Discourse: Health Centre Application

There are many doctors assigned to treat patients at a health centre. Patients must be registered with an associated doctor before they can book an appointment. However a patient when attending an appointment may not always see their own doctor, instead they may see another doctor working at the health centre.

The doctor sees the patient and he/she then makes a diagnosis of the illness/ailment. Medicines (if required) to treat the illness/ailment are recorded by the doctor on a form called a prescription. There may be many medicines recorded on a prescription and there may be many prescriptions for a patient if they have many illnesses/ailments. The patient is given prescriptions so that they can collect/buy the medicines from a local drug store or pharmacist. The doctor also records the details of the prescription this includes the medicine name, the category and the dose (amount taken and frequency) and other instructions if applicable (e.g. avoid alcohol). Repeat prescriptions (where a prescription extends over a period of time) are usually sent to the patient by post.

Medicines are classified according to their use, e.g. flu remedies, skin complaint remedies. Some medicines may fit into more than one category, e.g. penicillin.

The following query (QUERY1) needs to be supported:

List all the prescriptions made out to patient 'X' that were prescribed by doctor 'Y' involving medicines of category = 'Penicillin'. List the patient ailment/illness, the medicine, the date, the dose for each prescription dispensed and indicate whether it is a repeat or single prescription. Also list the date and time of the original appointment from which the prescription was dispensed.

a) Produce a logical data model for the above discourse. Your model should represent:

- *i)* The entities.
- *ii)* The relationships.
- *iii)* Multiplicity of the relationships, i.e. cardinality and optionality.
- *iv)* Draft tables required (including the primary and foreign keys).

Justify any modelling decisions and state any assumptions you have made. **(18 marks)**

b) Show with the aid of a diagram the access paths to the tables required to execute QUERY1. Justify the benefits of using access paths during database design. **(7 marks)**

Answer Pointers

Part a)

The ER model is shown at the end of this report.

The model below and corresponding Tables reflects the accuracy of modelling decisions expected as well as the consistency of mapping the ER model to the relational model. Referential integrity is also important and the marks are compensated for excellence in some areas against deficiencies in other areas of the modelling.

D R A F T T A B L E S

Patient (Patient_no, name, registered_doctor)

Doctor (Doctor_no, name)

Prescription (Doctor_no, Patient_no , consultation_date

Prescription_item (Doctor_no, Patient_no , consultation_date,
Drug_code. Dosage)

Drug (Drug_code, Drug_name)

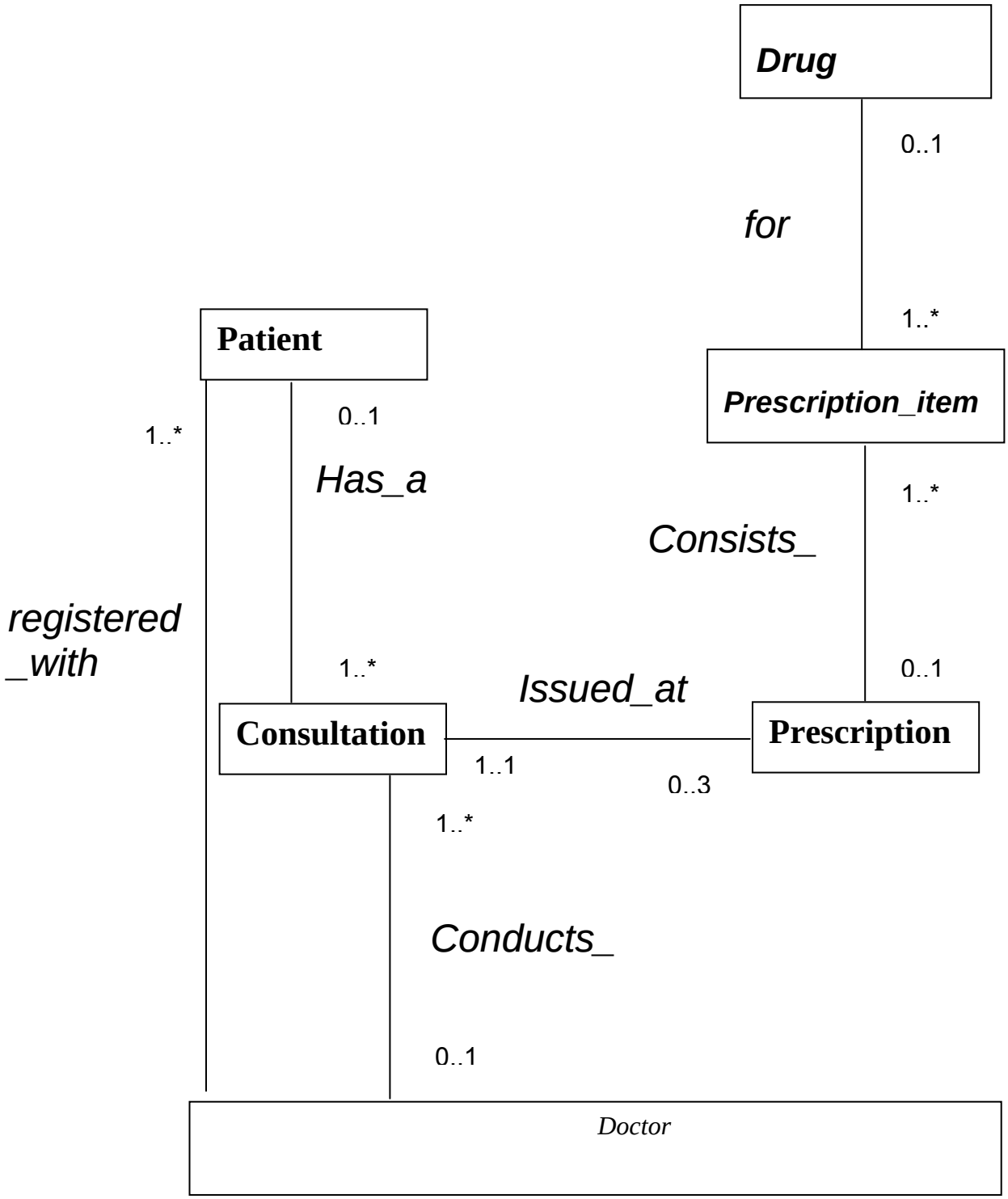
Consultation (Patient_no , Doctor_no, consultation_date)

Part b)
The access paths depend on what model is produced and may indicate a connection trap and/or the need for indexes. Occasionally it may be the case that a relationship is missing for example to maintain integrity following an insert.
3 marks for the access paths (arrows added to ER model above)
4 marks for justification

The ER model is shown at the end of this report.

Examiner's Comments
Candidates are getting much better at data modelling every year. This is perhaps candidates are better prepared as they now expect some modelling question. Data modelling is a key skill and therefore the presence of data modelling is likely to form some part of exam paper. However when candidates misinterpret a scenario this causes most problems with the allocation of marks. Unless the candidate produces a justification or an assumption behind a modelling decision the examiner has to mark the answer down. This happens in a number of cases (but not all) as it appears candidates have simply guessed or remembered another scenario. Also candidates need to be clear of any assumptions and state the precise modelling notation they have used. In Part b the simple access paths can be drawn diagrammatically simply by redrawing the ER model and annotating the paths on the model. Some candidates obviously wasted a lot of time drawing individual access paths or worse still writing SQL code to prove data could be related this way. The latter was not marked down but it would have involved too much work for little reward.

E R Model


Drug
0..1
for
1..*
Patient
Prescription_item
0..1
1..*
Has_a
1..*
Consists_
registered
_with
1..*
0..1
Issued_at
Consultation
Prescription
1..1
0..3
1..*
Conducts_
0..1
Doctor