H4 – DATABASE DEVELOPMENT
SOLUTIONS & MARKING SCHEME   JUNE 2013

PART A.

Answer A1.

Explain the term metadata in the context of a data dictionary and how a database management system makes use of it.

Data that is used to describe the objects in a database (tables, columns, users, privileges etc.). Used by DBMS to validate queries, check on user rights to access data etc.

(5 marks)

Answer A2.

End users in a database environment can be described as naïve or sophisticated. Explain how these different types may access the database

Naïve users probably unaware of the DBMS or the DB structure. They use simple commands/menu options/form filling to start DB operations
Sophisticated users familiar with the DB structure and facilities of the DBMS. They may use SQL or similar query languages or possibly write applications programs.

(5 marks)

Answer A3.

Briefly describe the meaning of the following aspects of SQL data definition commands

  i.   CREATE TABLE
  ii. DROP TABLE
  iii. NOT NULL
  iv. DEFAULT
  v. UNIQUE

i) Creates a new table in the users database schema
ii) Permanently deletes a table and its data
iii) Ensures that a column will not have null values
iv) Assigns a value to an attribute when a new row is added. The value can be overwritten by the user
v) Ensures that all values in a column are unique   **(5 marks)**

Answer A4.

Explain the difference between a primary key and a foreign key using a suitable example to illustrate your answer.

(5 marks)

Primary Key – attribute that is chosen to uniquely identify a row in a table (or tuple in a relation) - 1 mark

Foreign Key – attribute in a table that relates to a primary key in another table - 1 mark

Suitable answer (e.g. a Department number identifying a department in a table Primary Key (PK), and department number appearing in an employee table Foreign Key (FK) – 3 marks

Answer A5.

Give an example of a valid INSERT SQL statement. Give **TWO** reasons why a valid INSERT SQL statement might fail to insert data into a database? For each reason give an example.

(5 marks)

Any example that fits the standard definition
INSERT INTO Tablename[(column list)] VALUES (dataValueList) – 1 mark
Any two from (1 mark for reason, 1 mark for explanation)
  Data already present for UNIQUE attribute
  Breaks relationship rules (non-existent foreign key)
  User permissions may not allow inserts
  Other reasonable answer

Answer A6.

Describe the characteristics of a Distributed Database Management System.

The description should contain any five of the following (1 mark each up to a maximum of 5 marks)
  • *Collection of logically related shared data*
  • *Data split into fragments*
  • *Fragments may be replicated*
  • *Fragments/replicas allocated to sites*
  • *Sites linked by a communication network*
  • *Data at each site under the control of a DBMS*
  • *DBMS handles local applications autonomously*
  • *DBMS participates in global applications*

Answer A7.

Explain the difference between the hierarchical model and the network model

Hierarchical model represents data as a tree structure (1 mark) each parent record having zero (1 mark) or more children (1 mark), each child record only belonging to one parent (1 mark). By contrast, the network model allows a record to have more than one parent (1 mark).
(Up to 5 marks)

Answer A8.

Give a definition of the term "data mining"

The process of extracting valid (1 mark), previously unknown (1 mark) comprehensible (1 mark) and actionable (1 mark) information from large databases (1 mark) and using it to make critical business decisions (1 mark). **(Up to 5 marks)**

Total (40 marks)

PART B.

Answer B9.

The Three-Level ANSI-SPARC architecture is important in providing DBMS functionality. Describe in detail the three levels using diagrams where appropriate.

(8 marks)

External level – users' perception of the data (2 marks)

Internal level – where the DBMS and OS perceive the data is stored

(2 marks)

Conceptual – provides a mapping between external and internal levels

(2 marks)

Suitable diagram(s) (2 marks)

Explain why it is so important.

Importance (2 marks each)

- *Each user should be able to access the same data that is customised to their view*

- *Each user should be able to change this view without affecting other users*

- *Users should not have to deal directly with the details of the physical database storage*

- *DBA should be able to change the DB storage structures without affecting users' view*

- *Internal structure should be unaffected by changes to the physical storage*

- *DBA should be able to change the conceptual structure without affecting all users*

(12 marks)

(Total 20 marks)

Answer B10.

Dave Rave Promotions organises pop concerts in various countries. Part of the data that is stored to enable the company to keep track of who is performing and which company is supplying security for the event is shown below.

C#	CName	CDate	CVenue	S#	SName	SCont	A#	AName	ACont	Fee
C1	Rockit	1/6/13	Munich	S1	Bond	998123	A1	Locus	333456	4000
							A2	Freddie	665675	3000
C2	CoolTunes	4/8/13	Durban	S2	Mooney	888234	A1	Locus	333456	4500
							A3	Cheese	776345	2500
C3	Grunge	8/9/13	Istanbul	S1	Bond	998123	A1	Locus	333456	5000
							A2	Freddie	665675	4000
							A3	Cheese	776345	3500

C# - Concert number, a unique identifier

CName – Concert name

CDate – Concert date

CVenue – Concert venue

S# - Security firm number, a unique identifier

SName – Security firm name

SCont – Security firm contact

A# - Act number, a unique identifier

AName – Act name

ACont – Act contact

Fee – Fee to be paid to the act for that concert

Using this data explain the transformation into first, second and third normal form

(Total 20 marks)

Explanation of 1st normal form – remove repeating groups   **(2 marks)**

     CONCERT (C#, CName, CDate, CVenue, S#, SName, SCont)

     ACT (C#, A#, AName, Acont, Fee)   **(4 marks)**

Explanation of 2nd normal form – remove partial dependency   **(2 marks)**

     CONCERT (C#, CName, CDate, CVenue, S#, SName, SCont)

     CONCERT-ACT (C#, A#, Fee)

     ACT (C#, A#, AName, Acont)

                                                     (5 marks)

Explanation of 3rd normal form – remove partial dependency   **(2 marks)**

      CONCERT (C#, CName, CDate, CVenue, S#, SName, SCont)

      SECURITY (S#, SName, SCont)

      CONCERT-ACT (C#, A#, Fee)

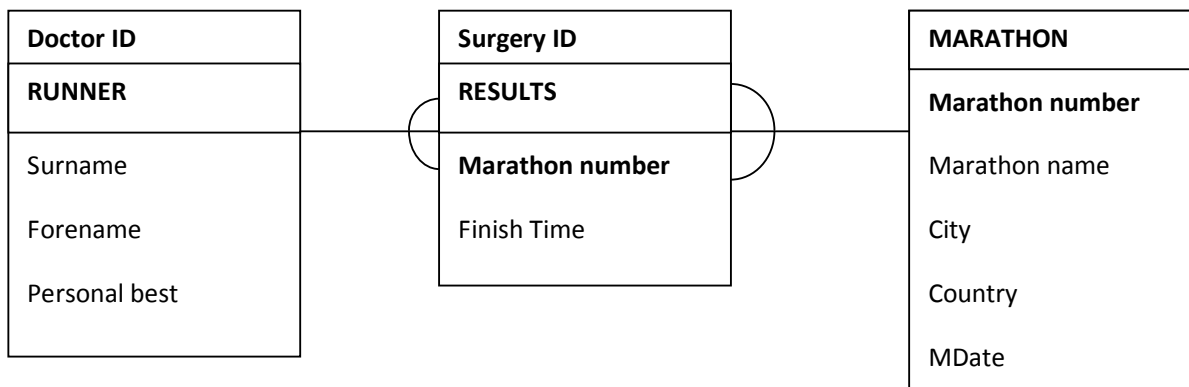   **ACT (C#, A#, AName, Acont)**

                                                       (5 marks)

Answer B11.

The following relationship model and table details indicate how data is stored for the recording of results in marathon events taking place throughout Africa, Use the data definition to answer questions B12a, B12b and B12c :-

Marathon Database Tables and Relationships

Doctor ID	Surgery ID	MARATHON
RUNNER	**RESULTS**	**Marathon number**
Surname	**Marathon number**	Marathon name
Forename	Finish Time	City
Personal best		Country
		MDate

Marathon					
MarathonNumber	MarathonName		**City**	**Country**	**MDate**
1	Marathon International de Marrakech		Marrakech	Morocco	27/01/2013
2	Deloitte Pretoria Marathon	Pretoria		South Africa	03/02/2013
3	Old Mutual Buffs Marathon		East London	South Africa	23/02/2013
4	Rwanda Challenge Marathon		Rwamgena	Rwanda	24/02/2013
5	Sahara Marathon		Tindouf	Algeria	25/02/2013
6	Kilimanjaro Marathon		Moshi	Tanzania	03/03/2013
7	Nedbank City of Cape Town Marathon		Cape Town	South Africa	22/09/2013

Runner			
VestNumber	**Surname**	**Forename**	**PersonalBest**
122	Mboya	Frederick	02:33:12
134	Jones	Andy	03:54:35
456	Ledyard	Eric	04:09:30
467	Giroux	Pierre	03:45:40
565	Tchana	Merryman	04:00:30
587	Varney	Regina	04:03:43
677	Kuwali	Donna	03:44:50
809	Smithies	Ron	03:22:44
888	Mgano	Charles	04:00:04
899	Thomas	Mary	03:09:00

Results		
MarathonNumber	**VestNumber**	**FinishTime**
1	122	02:43:10
1	456	04:39:30
2	122	03:03:10
2	134	03:35:55
2	467	03:35:45
3	122	03:13:10
3	134	04:05:25

Results		
MarathonNumber	**VestNumber**	**FinishTime**
3	587	04:22:43
4	587	04:33:33
5	888	04:00:54
7	122	02:55:10
7	134	04:15:05
7	456	04:49:30
7	888	04:10:54
7	899	03:09:00

Answer B11 a)

Write the SQLstatement required to display the Marathon Name, Date and City for all races taking place in South Africa giving:-

SouthAfricanMarathons		
MarathonName	**MDate**	**City**
Deloitte Pretoria Marathon	03/02/2013	Pretoria
Old Mutual Buffs Marathon	23/02/2013	East London
Nedbank City of Cape Town Marathon	22/09/2013	Cape Town

(4 marks)

Answer B11 b)

Write the SQL statement to give the number of finishers for each marathon giving –

Finishers	
MarathonName	**Finishers**
Deloitte Pretoria Marathon	3
Marathon International de Marrakech	2
Nedbank City of Cape Town Marathon	5
Old Mutual Buffs Marathon	3
Rwanda Challenge Marathon	1
Sahara Marathon	1

(5 marks)

Answer B11 c)

Write the SQL statement required to display the Marathon Date, Surname , Forename Personal Best and Finish Time for all runners with a finish time better than their previous personal best giving -

NewPB				
MDate	**Surname**	**Forename**	**PersonalBest**	**FinishTime**
03/02/2013	Jones	Andy	03:54:35	03:35:55
03/02/2013	Giroux	Pierre	03:45:40	03:35:45

(11 marks)

(Total 20 marks).

a) SELECT Marathon.MarathonName, Marathon.MDate, Marathon.City (1 mark)

FROM Marathon (1 mark)

WHERE (((Marathon.Country)="South Africa")); (2 marks)

b) SELECT Marathon.MarathonName, Count(Results.FinishTime) AS Finishers (2 marks)

FROM Marathon INNER JOIN Results ON Marathon.MarathonNumber = Results.MarathonNumber (2 marks)

GROUP BY Marathon.MarathonName; (1 mark)

c) SELECT Marathon.MDate, Runner.Surname, Runner.Forename, Runner.PersonalBest, Results.FinishTime (3 marks)

FROM Runner INNER JOIN (Marathon INNER JOIN Results ON Marathon.MarathonNumber = Results.MarathonNumber) ON Runner.VestNumber = Results.VestNumber (5 marks)

WHERE (((Results.FinishTime)<[Runner].[PersonalBest])); (3 marks)

Answer B12.

a) The processing of a query can be subdivided up into the SQL parsing phase, the execution phase and the fetching phase. Briefly describe the execution and fetching phases.

(4 marks)

b) Briefly describe FOUR of the steps in parsing a SQL query.

(8 marks)

c) Modern database management systems use in-built query optimisers to speed up queries. Describe two other aspects that might affect performance, giving appropriate examples.

(8 marks)

(Total 20 marks)

a) Execution Phase – all I/O operations indicated in the access plan are executed including acquiring of lock and fetching of data into DBMS data cache (2 marks)

Fetching Phase – All rows matching the conditions are retrieved, sorted, grouped and aggregated if required and returned to the client (2 marks)

b) Any four from (2 marks each)

- *Validate for syntax compliance*

- *Validate against data dictionary for correct table/column names*

- *Validate against data dictionary for access rights*

- *Analyse and decompose into atomic components*

- *Transform into an equivalent but more efficient SQL query*

- *Determine the most efficient execution plan*

c) Any two from (4 marks each)

- *Faster processor or more processors*

- *Larger size RAM*

- *Faster disk*

- *Better network connection*

© Institute for the Management of Information Systems – J2013.

- **Better use of indexes**

- **Other suitable examples**

Question B13.

A bank has its head office in Paris. It has branches all over Europe. Each customer account belongs to one branch only. Currently, the company has one database located in the head office. Customers at branches access this database via a communication network for whatever they need.

One of the relations in this centralised database system is the Account relation, where data about the accounts is kept. The attributes of the Account relation are: the account number (AccNo), account type (AccType), the customer's identification (Name), the branch where the account is held, the current balance for the account (balance) and an indicator which tells head office if the publicity (information on updates to regulations, interest rates etc.) for a particular account type for a customer has been sent out. An instance of the account relation is given below:

Acc No	Acc Type	Name	Branch	Balance	Publicity
191688	002	Jones	London	2000	yes
779865	005	Smith	Paris	600	no
158756	002	Green	Paris	8	yes
125467	007	White	Paris	10000	no
124678	005	Black	London	150	yes

a) The bank has decided to move to a distributed database system where each of the sites has its own database. Using the above data explain the concept of horizontal and vertical fragmentation.

(8 marks)

b) Discuss three reasons why replication might be used in a distributed database environment.

(6 marks)

c) What are the problems of introducing replication?

(6 marks)

a) Horizontal fragmentation – the division of a relation (table) into subsets of tuples (records) – 2 marks
   Example – 2 marks
   Vertical fragmentation – the division of a relation into attribute (column) subsets – 2 marks
   Example – 2 marks

b) Any two from(2 marks each)
   Improved reliability
   Improved availability
   Improved performance
   Supports mobile environments
   Other reasonable points

c) All replicas must be updated to ensure consistency. For synchronous replication If one replica is unavailable the transaction cannot complete.
   (3 marks)
   The extra messages to keep the replicas in step add a burden onto networks.   (3 marks)
   Or any other suitable argument

(Total 20 marks)