****

**University of Cape Town**

**Department of Information Systems**

**Introduction to Databases (INF2007) Test**

**-----------------------------------------------------------------------------------------------------------**

***Instructions to Candidates***

1. This test is **not** open book.
2. The test is out of **70 marks** and you have **1h30 Minutes.** Marks allocated to a question should be used as an indication of the level of detail expected in the answer.
3. The test has three sections. Answer all sections.
4. Write your **name and student number** on and answer questions in the provided test answer books. Each question should be answered on a new page.
5. At the end of the test you must hand-in this question paper, and any additional answer books used

**-------------------------------------------------------------------------------------------------------**

*For Marking Purposes*

|  |  |  |  |
| --- | --- | --- | --- |
| ***SECTION 1*** | ***SECTION 2*** | ***SECTION 3*** | ***Total*** |
| */20* | */10* | */40* | */70* |

|  |  |
| --- | --- |
| **Student Name:** | **Student Number:** |

***Fold the page over on this line and seal with the stickers supplied.***

**SECTION A MULTIPLE CHOICE [20]**

Answer the following question in your answer booklet by selecting the most appropriate answer for each question

1. Which statement about a file-based approach to data management is not correct?
   * 1. The file-based approach to data management causes the same information to be stored separately for different applications.
     2. In a file-based approach to data management, the data definitions are included in each application separately.
     3. In a file-based approach to data management, different applications could be using older and newer versions of the same data.
     4. In a file-based approach to data management, a change in the structure of a data file is easily handled because each application has its own data files.
2. Which statement about a database-based approach to data management is not correct?
   * 1. In a database approach to data management, all data are stored and managed centrally by a database management system
     2. In a database approach to data management, every application using the database has to explicitly write its own query and access procedures
     3. In a database approach to data management, applications are independent from the data and data definitions.
     4. In a database approach to data management, facilities for data querying and retrieval is provided.
3. A database management system is
   * 1. a combination of Database and Big Data used to define, create, use, and maintain a database
     2. a combination of DBMS and Volatile Data used to define, create, use, and maintain a database
     3. a software package used to define, create, use, and maintain a database
     4. a combination of a DBMS and a database used to define, create, use, and maintain a database
4. Which statement is not correct?
   * 1. In a conceptual data model, the data requirements from the business should be captured and modelled.
     2. A conceptual data model is implementation-dependent.
     3. A logical data model translates the conceptual data model to a specific implementation environment.
     4. Examples of implementations of logical data models are hierarchical, CODASYL, relational, or object-oriented models.
5. Which of the following is part of the query processor in the architecture of a DBMS?
   * 1. DDL compiler.
     2. DML compiler.
     3. Transaction manager.
     4. Security manager.
6. Which of the following concept specifies the various data items, their characteristics, and relationships, constraints, storage details, etc. and is specified during the database design?
   * 1. Database model.
     2. Catalog.
     3. Database state.
     4. None of the above
7. Which of the following DBMS types is not a classification based on a data model?
   * 1. Network DBMS.
     2. Hierarchical DBMS.
     3. Cloud DBMS.
     4. Object-relational DBMS.
8. Which statement is correct about the database state
   * 1. A database state, is specified during database design and is not expected to change frequently.
     2. A database state, comprises different data models, each describing the data from different perspectives.
     3. A database state, represents the data in the database at a particular moment
     4. The database state is stored in the catalog.
9. The external data model
   * 1. describes which data are stored where, in what format, which indexes are provided to speed up retrieval
     2. represents various subsets of the data items known as views, tailored toward the needs of specific applications or groups of users.
     3. is a translation or mapping of the conceptual data model toward a specific implementation environment
     4. is a communication instrument between the information architect and business user to make sure the data requirements are adequately captured and modelled
10. To ensure database transactions are processed in a reliable way, the DBMS must support the ACID properties. What does ACID stand for?
    * 1. Atomicity, Concurrency, Isolation, Durability
      2. Atomicity, Concurrence, Implementation, Durability
      3. Atomicity, Consistency, Implementation, Durability
      4. Atomicity, Consistency, Isolation, Durability
11. Which statement about DDL statements is correct
    * 1. DDL statements create data definitions that are stored in the catalog
      2. DDL statements are typically queried from a front-end tool, such as simple graphical user interface, or forms-based interface.
      3. DDL statements assist in parses the query into an internal representation format
      4. DDL statement provides a set of constructs to select, insert, update, and delete data
12. On-line transaction processing (OLTP) is
    * 1. A DBMS for point-of-sale (POS) application in a supermarket
      2. A DBMS for storage of Twitter tweets, Facebook
      3. A DBMS for storage for geographical information systems (GIS)
      4. A DBMS for smartphones, tablets, and other mobile devices.
13. Which statement(s) is/are correct?

* Statement A: An OLTP system is able to cope with real-time, simultaneous transactions that the database server is able to process in a large volume.
* Statement B: An OLAP system uses large amounts of operational data to run complex queries on and provide insights for tactical and strategic decision-making.
  + 1. Only A.
    2. Only B
    3. A and B.
    4. Neither A nor B.

1. Which statement(s) is/are correct?

* Statement A: The query parser optimizes and simplifies a query and then passes it on to the query executor.
* Statement B: In the DBMS architecture, the storage manager takes care of concurrency control.
  + 1. Only A
    2. Only B
    3. A and B
    4. Neither A nor B.

1. Which statement(s) is/are correct?

* Statement A: In a hierarchical DBMS, DML is declarative and set oriented with a query processor.
* Statement B: In a relational DBMS, there is data independence between the conceptual and internal data model.
  + 1. Only A
    2. Only B
    3. A and B
    4. Neither A nor B

1. Which statement about the ternary relationship is correct?
   * 1. In the case a ternary relationship type is represented as three binary relationship types, then semantics will get lost.
     2. A ternary relationship type can always be represented as three binary relationship types without loss of semantics.
     3. Three binary relationship types between three entity types can always be replaced by one ternary relationship type between the three participating entity types.
     4. A ternary relationship type cannot have attribute types
2. Which statements is correct?
   * 1. A weak entity type can only have one attribute type.
     2. A weak entity type is always existence-dependent.
     3. An existence-dependent entity type is always a weak entity type.
     4. An existence-dependent entity type always participates in a 1:1 relationship type.
3. Which of the following is not the limitation of the ER model
   * 1. ER model has temporal constraints, which are constraints spanning a particular time interval, cannot be modelled
     2. ER model cannot guarantee consistency across multiple relationship types.
     3. ER Model, cannot model domain, it is not possible to specify the set of values that can be assigned to an attribute type
     4. ER Model, has an existence-dependent limitation, where entity type always participates in a 1:1 relationship type.
4. Which of the following statements is correct?
   * 1. A foreign key of a relation A cannot refer to the primary key of the same relation A.
     2. A relation cannot have more than one foreign key.
     3. Every relation must have a foreign key.
     4. A foreign key can be NULL.
5. Consider the following relation R

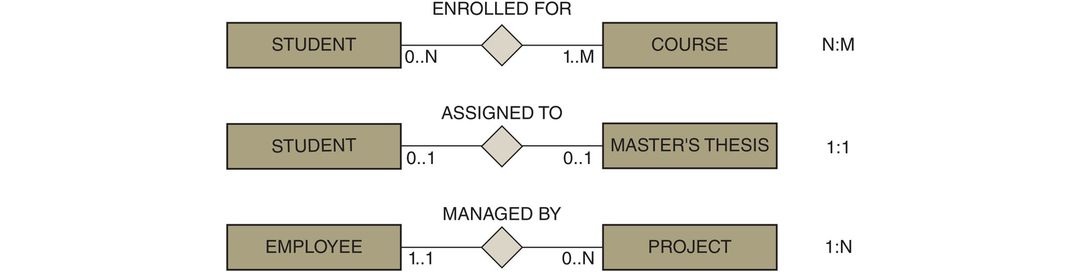
|  |  |  |  |
| --- | --- | --- | --- |
| **ORDER** | **DATE** | **CUSTOMER** | **ITEM\_NO** |
| O101  0100 | 3/3/2020  22/2/2020 | Salam | 6531  7890  7956 |
| O105 | 2/3/2020 | George | 3456 |
| O134 | 22/2/2020 | Obad | 4234  7956 |

* + 1. R may be in 2NF
    2. R may be in 3NF
    3. R may be in 1NF
    4. None of the above

**SECTION B: [10]**

State whether the following statements are true of false. Write the answer in your answer booklet.

1. Database technology is for traditional numeric and alphanumeric data only. F
2. Unstructured data, there are no finer-grain components in a file or series of characters that can be interpreted in a meaningful way by a DBMS or application. T
3. SQL is a DML language to retrieve, insert, delete, and modify data. It is stored in the catalog.F
4. The degree of a relationship type corresponds to the number of entity types participating in the relationship type. A unary relationship type has two participating entity types. F
5. The ER model can model both weak and existence-dependent entity types at once T
6. The following image is an example of a zero to many relationship type. F



1. A functional dependency X → Y between two sets of attribute types X and Y implies that a value of Y uniquely determines a value of X. F
2. A prime attribute type is an attribute type that can be either an alternative key or a primary key in a relation. T
3. 2NF ONLY applies to tables with composite primary keys T
4. A relation is in 3 NF if it satisfies 2 NF and no primary key of R is transitively dependent on a prime attribute type. F

**SECTION C: CASE STUDY [40]**

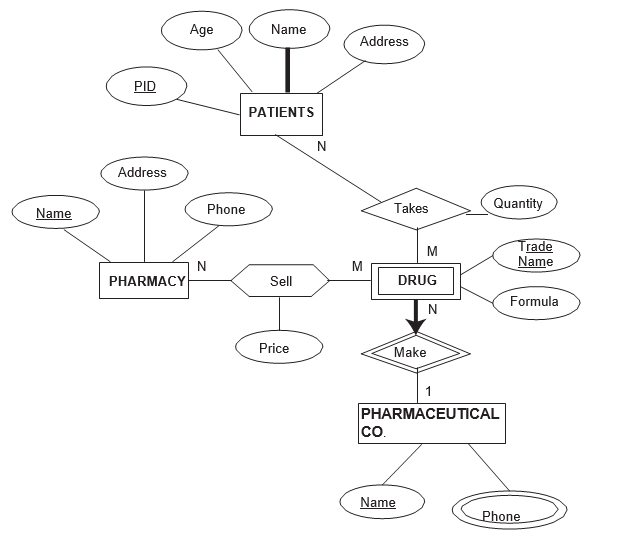
**Conceptual Data Modelling (20)**

1. The Prescriptions X chain of pharmacies has offered to give you a free lifetime supply of medicines if you design its database. Given the rising cost of health care, you agree. Here’s the information that you gather:

* Patients are identified by a Patient’s ID, and their names, addresses, and ages must be recorded.
* Doctors are identified by an ID . For each doctor, the name, speciality, and years of experience must be recorded.
* Each pharmaceutical company is identified by name and has a phone number.
* For each drug, the trade name and formula must be recorded. Each drug is sold by a given pharmaceutical company, and the trade name identifies a drug uniquely from among the products of that company. If a pharmaceutical company is deleted, you need not keep track of its products any longer.
* Each pharmacy has a name, address, and phone number.
* Every patient has a primary physician.
* Every doctor has at least one patient.
* Each pharmacy sells several drugs and has a price for each. A drug could be sold at several pharmacies, and the price could vary from one pharmacy to another.
* Doctors prescribe drugs for patients. A doctor could prescribe one or more drugs for several patients, and a patient could obtain prescriptions from several doctors. Each prescription has a date and a quantity associated with it. You can assume that if a doctor prescribes the same drug for the same patient more than once, only the last such prescription needs to be stored.
* Pharmaceutical companies have long-term contracts with pharmacies. A pharmaceutical company can contract with several pharmacies, and a pharmacy can contract with several pharmaceutical companies. For each contract, you have to store a start date, an end date, and the text of the contract.
* Pharmacies appoint a supervisor for each contract. There must be a supervisor for each contract, but the contract supervisor can change over the lifetime of the contract.

**Questions:**

1. The company hired a designer who produced an ER diagram, as seen in figure 1. The ER diagram did not capture all the required information. Modify and improve the ER diagram seen in figure 1 such that your new designed ER diagram captures all the above requirements. Identify any constraints that cannot be captured by the ER diagram. (8)



***Figure 1*: Pharmaceutical ER diagram version 1**

**SOLUTION**

Version 1 of the R diagram is missing

DOCTOR entity type 2

We want to know at least the speciality and the date of entry into the profession of each physician.

Primary-PHYSICIAN – relationship 1

Doctor and patients relationship

Each patient has to have one and only one primary physician(1..1). Each physician has at least one patient (1..M). Relationships physician: patient => (1:N).

Prescription relationship type 1

is a ternary relationship between (Drug,. Doctor and Patient). Instead of modeling only the fact that a patient takes certain drugs, model the fact that a patient takes certain drugs that are prescribed by a physician and the prescription date.

Contract relationship type – 2

Pharmaceutical companies have long-term contracts with pharmacies. A pharmaceutical company can contract with several pharmacies, and a pharmacy can have contracts with several pharmaceutical companies. (N:M)

Pharmacies appoint a supervisor for each contract (1:1). There must be a supervisor for each contract, but the contract supervisor can change over the lifetime of the contract. Add supervisor attribute type in the contrat relationship

Correct relationship constraints – 1

ER Model Constraints

ER model has temporal constraints, which are constraints spanning a particular time interval, cannot be modelled

* + - * + cannot model the fact that contract supervisor can change over the lifetime of the contract.

ER model cannot guarantee consistency across multiple relationship types.

ER Model, cannot model domain, it is not possible to specify the set of values that can be assigned to an attribute type: Domain constraints

A picture containing text, map

Description automatically generated

1. List the multi-valued attributes, composite attributes and Weak Entity types.
   1. Multi-Valued (3) (Any two of the following)
      1. Phone number- pharmacy
      2. Phone number – pharmaceutical company
      3. Speciality can be considered as a multivalued attribute
   2. Composite attributes – Address. Name (3)
      * + But if none is chosen assumptions must be made clear why none is considered as atomic attribute types.
   3. Weak entity types
      1. Drug (2) (8)
2. How would your design change if each drug must be sold at a fixed price by all pharmacies?
   1. If the drug is to be sold at a fixed price we can add the price attribute type to the Drug entity type and eliminate it from the Sell relationship type.

(4)

**Relational Model (Normalization) (20)**

1. Consider the Bank relational model below. BANK(StaffNo,Name,Position,BranchNo,BranchAddress,PhoneNo,StaffHours)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***StaffNo*** | ***Name*** | ***Position*** | ***BranchNo*** | ***BranchAddress*** | ***PhoneNo*** | ***Staffhours*** |
| 1001 | Paul Ryan | Manager | B001 | 8 Jefferson Way, Pinelands, Cape Town, 7405 | 021-555-3618, 021-555-2727, 021-555-6534 | 14 |
| 1002 | Vuke Josh | Supervisor | B002 | City Center Plaza, Gardens, Cape Town, 8801 | 021-555-6756, 021-555-8836 | 9 |
| 1003 | Jamie Shuttle | Manager | B003 | 14 – 8th Main Road, Claremont, Cape Town, 7708 | 021-371-3000 | 16 |
| 1004 | Sipho Seth | Assistant | B004 | 16 – 14th Atlas Road, Rondebosch, 7701 | 021-555-3131, 021-555-4112 | 16 |
| 1005 | Chad Zuma | Supervisor | B004 | 16 – 14th Atlas Road, Rondebosch, 7701 | 021-555-3131, 021-555-4112 | 7 |
| 1006 | Ziko Van | Manager | B002 | City Center Plaza, Gardens, Cape Town, 8801 | 021-555-6756, 021-555-8836 | 8 |
| 1009 | Pili Tatu | Manager | B004 | 16 – 14th Atlas Road, Rondebosch, 7701 | 021-555-3131, 021-555-4112 | 16 |

**Questions :**

1. Normalize the logical relational model from question 2 into a 1NF, 2NF and 3NF.
2. Clearly indicate each normalization level by showing the multi-valued attributes, Composite attribute, functional dependencies and transitive dependencies decomposition on their respective normal forms. .
3. Clearly indicate the primary–foreign key relationships and specify NOT NULL declarations where necessary. Show all the assumptions made.

(15)

**Solution**

Primary key are underlined : all primary keys must satisfy NOT NULL constraints

Foreign keys are in italic : Branch no must satisfy NOT NULL constraint everywhere it appears

**INF**

The first normal form is violated because phoneNo is a multivalued attribute type. The relation BANK can be brought in first normal form as follows:

BANK(StaffNo, BranchNo, Name,Position, BranchAddress,StaffHours,)

BRANCH\_PHONE(PhoneNo, BranchNo)

I noted that many of you considered PhoneNo to be an attribute of Staff relation, without stating your assumptions. Always remember to provide your assumptions when normalizing.

**2NF**

The relation BANK is not in second normal form because:

* BranchAdress is partially dependent on StaffNo and BranchNo. Because, it only depends on BranchNo.

BranchNo 🡪 BranchAdress

* Name and position is partially dependent on StaffNo and BranchNo. In fact, they only depend on StaffNo.

StaffNo🡪 Name

StaffNo🡪Position

* The only attribute type that is fully functional dependent on the composite key is StaffHours

StaffNo, BranchNo 🡪 StaffHours

* The relation BANK can be brought into second normal form as follows:

WORKS(*StaffNo BranchNo*,StaffHours)

BRANCH (BranchNo, BranchAddres)

BRANCH\_PHONE(PhoneNo , *BranchNo)*

STAFF (StaffNo, Name,Position)

**3NF**

**there is no transitive dependency will Same as 2NF –**

WORKS (StaffNoBranchNo,StaffHours)

BRANCH (BranchNo, BranchAddres)

BRANCH\_PHONE*(*, PhoneNo, *BranchNo*)

STAFF (StaffNo, Name,Position)

1. Explain the types of anomalies that might occur on this relational model when not normalized
   * 1. At least three types of anomaly may arise when working with an unnormalized relational model: an insertion anomaly, a deletion anomaly, and an update (modification) anomaly (0.5)

An explanation for each anomaly must follow.---

**Deletion Anomaly (1.5)**

If we delete a record from the BANK table that represents the last member of staff located at a branch, the details about that branch are also lost from the database.

For example, if we delete the record for staff Jamie Shuttle (1003) from the BANK table, the details relating to branch B003 are lost from the database. The design of the tables that separate the Staff and Branch table avoids this problem because branch records are stored separately from staff records and only the column branchNo relates the two tables. If we delete the record for staff Jamie Shuttle (1003) from the Staff table, the details on branch B003 in the Branch table remain unaffected.

**Update Anomaly (modification anomaly) (1.5)**

If we want to change the value of one of the columns of a particular branch in the BANK table, for example the telephone number for branch B004, we must update the records of all staff located at that branch (row 4,5 and 7). If this modification is not carried out on all the appropriate records of the BANK table, the database will become inconsistent. In this example, branch B004 would have different telephone numbers in different staff records. The above examples illustrate that the Staff and Branch tables have more desirable properties than the BANK table.

**Insertion anomaly (1.5)**

To insert details of a new branch that currently has no members of staff into the BANK table, it’s necessary to enter NULLs into the staff-related columns, such as staffNo. However, as staffNo is the primary key for the BANK table, attempting to enter nulls for staffNo violates entity integrity, and is not allowed. The design of the tables shown in Staff and Branch avoids this problem because new branch details are entered into the Branch table separately from the staff details. The details of staff ultimately located at a new branch can be entered into the Staff table at a later date.

In addition:

To insert the details of a new member of staff (staffNo, name and position) located at a given branch into the BANK table, we must also enter the correct details for the branch (branchNo, branchAddress and PhoneNo).

For example, to insert the details of a new member of staff at branch B004, we must enter the correct details of branch B004 so that the branch details are consistent with values for branch B004 in other records of the BANK table. The data shown in the BANK table is also shown in the Staff and Branch tables. These tables do have redundant data and do not suffer from this potential inconsistency, because for each staff member we only enter the appropriate branch number into the Staff table. In addition, the details of branch B004 are recorded only once in the database as a single record in the Branch table. (5)

Resulting table might look like this (Note \* Tables were not required \*)

|  |  |  |
| --- | --- | --- |
| ***StaffNo*** | ***Name*** | ***Position*** |
| 1001 | Paul Ryan | Manager |
| 1002 | Vuke Josh | Supervisor |
| 1003 | Jamie Shuttle | Manager |
| 1004 | Sipho Seth | Assistant |
| 1005 | Chad Zuma | Supervisor |
| 1006 | Ziko Van | Manager |
| 1009 | Pili Tatu | Manager |

|  |  |
| --- | --- |
| ***BranchNo*** | ***BranchAddress*** |
| B001 | 8 Jefferson Way, Pinelands, Cape Town, 7405 |
| B002 | City Center Plaza, Gardens, Cape Town, 8801 |
| B003 | 14 – 8th Main Road, Claremont, Cape Town, 7708 |
| B004 | 16 – 14th Atlas Road, Rondebosch, 7701 |

|  |  |
| --- | --- |
| **PhoneNo** | ***BranchNo*** |
| 021-555-3618 | B001 |
| 021-555-2727 | B001 |
| 021-555-6534 | B001 |
| 021-555-6756 | B002 |
| 021-555-8836 | B002 |
| 021-371-3000 | B003 |
| 021-555-3131 | B004 |
| 021-555-4112 | B004 |

|  |  |  |
| --- | --- | --- |
| ***StaffNo*** | ***BranchNo*** | ***Staffhours*** |
| 1001 | B001 | 14 |
| 1002 | B002 | 9 |
| 1003 | B003 | 16 |
| 1004 | B004 | 16 |
| 1005 | B004 | 7 |
| 1006 | B002 | 8 |
| 1009 | B004 | 16 |