# Database Design and Development

# 06 August 2019

# Marking Scheme

This marking scheme has been prepared as a **guide only** to markers. This is not a set of model answers, or the exclusive answers to the questions, and there will frequently be alternative responses which will provide a valid answer. Markers are advised that, unless a question specifies that an answer be provided in a particular form, then an answer that is correct (factually or in practical terms) **must** be given the available marks.

If there is doubt as to the correctness of an answer, the relevant NCC Education materials should be the first authority.

**Throughout the marking, please credit any valid alternative point.**

**Where markers award half marks in any part of a question, they should ensure that the total mark recorded for the question is rounded up to a whole mark.**

**Answer ALL questions**

**Marks**

## Question 1

**a)** With the use of an example, explain the concept of a domain. **2**

***Award 1 mark for the explanation and 1 mark for an example.***

***Explanation***
***Domain - The set of valid values of an attribute.***

***Example***
***The valid values in the domain 'sex' would be 'Male' and 'Female'.***

***Note: Credit valid alternative answers.***

**b)** With the use of an example, explain the concept of functional dependency. **4**

***Functional dependency represents a relationship between attributes (1 mark) such that if the value of one attribute is known the value of the other attribute is known (1 mark). 1 mark for an example". It is not reversible (1 mark).***

**c)** Explain the purpose of normalisation. **2**

***Normalisation aims to overcome potential anomalies that can occur when data is replicated (1 mark). Normalisation aims to eliminate replication (1 mark).***
***Credit alternative answers.***

**d)** Explain the concept **and** purpose of de-normalisation. **2**

***De-normalisation is the process of introducing redundancy or duplication (1 mark). It is done to improve performance and/or scalability (1 mark).***

**Total 10 Marks**

**Marks**

## Question 2

Consider the following table which shows orders for an office supplies company. Note that the table is not normalised and contains duplications e.g. order 2100 appears twice.

| Order ID | Order Date | Product Code | Product Name | No of Products | Customer ID | Customer Name |
|---|---|---|---|---|---|---|
| 2100 | 01-JAN-2019 | P11 | Deluxe Desk | 1 | C789 | Sally Smith |
| 2100 | 01-JAN-2019 | C11 | Standard Desk | 5 | C789 | Sally Smith |
| 2611 | 03-JAN-2019 | P99 | Super Projector | 1 | C789 | Sally Smith |
| 1309 | 31-DEC-2019 | J99 | Standard Office Chair | 1 | C980 | Colin Clark |
| 1299 | 01-DEC-2019 | XP1 | Deluxe Office Chair | 1 | C876 | Humzah Khan |
| 1567 | 03-DEC-2019 | J99 | Standard Office Chair | 2 | C754 | Johnny West |

**a)** Explain what anomalies might arise because of the way this table is structured. **6**

***An insert anomaly is where a table is not properly normalized and errors inserting data results (1 mark). If we want to record a new product but no one has yet ordered it we cannot do so because we need an order id and this is likely part of the primary key and therefore cannot be null (1 mark).***

***Update anomaly (1 mark). Because if we need to update the name of a customer or product we have to do it in more than one place, that there is duplication of data (e.g. Sally Smith) (1 mark).***

***Deletion anomaly (1 mark). Because if an order that contains a product that doesn't appear on any other order is deleted then we lose the information about that product entirely (1 mark).***

**b)** Explain how this data can be restructured so that anomalies do not arise. **4**

***There should be a table for Orders (1 mark), a table for Products Sold for that Order (Order Line or Order Product) (1 mark) there should be tables for Customers (1 mark) and Products (1 mark )***

**Total 10 Marks**

**Marks**

**Question 3**

**a)** Define the term datatype **and** provide TWO (2) examples of datatypes available in SQL. **3**

***Award 1 mark for the definition and 1 mark for each example, up to a maximum of 3 marks.***

- ***A datatype defines the broad domain of valid values for a given attribute.***

- ***Examples of SQL data types are integer, char, varchar, date etc.***

***Note: Credit other valid definitions and examples.***

**b)** With the use of an example, explain what is meant by an outer join in SQL. **3**

***The outer join is used to retrieve rows from a table where there are unmatched values in the data set (1 mark). For example if we have Students and Courses tables: if the Course and Students tables are joined then Students who do not yet have a course will NOT be retrieved with a normal join. An Outer Join will remedy this (up to 2 marks for example).***

**c)** With the use of an example, explain what the Cartesian product is of two tables in SQL. **3**

***The Cartesian product is every row from one table along with every row from another table regardless of any link between them (1 mark). Example if we have Students and Courses tables: the Cartesian product would give a result set with every course matched with every student even if that student was not taking that course (up to 2 marks for example).***

**d)** Describe the entity integrity rule. **1**

***The rule states that if an attribute or group of attributes is specified as a Primary Key then it cannot contain a NULL value (1 mark).***

**Total 10 Marks**

**Marks**

## Question 4

**a)** Explain what is meant by derived data in a database system. **4**

***Derived data is derived from elsewhere in the database rather than being directly entered (1 mark). It could be a column whose value is derived from the value of one or more other columns in the database (1 mark). These might be columns within the same table or columns from one or more other tables (1 mark). Derived data might use aggregate functions (1 mark)***

***Other valid points should be credited.***

**b)** Give TWO (2) examples of situations where derived data might be used. **3**

***Possible suitable examples:***

- ***To display up-to-data totals of money (for example of sales or costs) derived from other columns.***
- ***To display a tally (count) of some other columns.***
- ***To set a non-numeric value of one item of data that is dependent on another such as the colour of one car part (door) being dependent on that of another (body).***

***Credit any other valid examples.***

**c)** With the use of an example, briefly explain what is meant by a redundant relationship in an ER model. **2**

***Explanation***

- ***A relationship is redundant if the same information can be found via some other relationships /A relationship that represents something that might be derived through one or more other relationships (1 mark)***

***Example (1 mark)***

- ***Examples include an Order Line having a Customer ID when the Customer ID is also on the parent Order entity.***

***Note: Credit alternative valid explanations and examples.***

**d)** What is the usual approach during logical design to redundant relationships in an ER diagram? **1**

***They are usually removed (1 mark)***

**Total 10 Marks**

**Marks**

## Question 5

**a)** Explain the terms: Data Definition Language (DDL) **and** Data Manipulation Language (DML) in SQL. Provide TWO (2) examples to illustrate the difference. **6**

***DDL is about creating data structures (1 mark) such as tables, columns, views etc. (1 mark for any example). It uses commands like CREATE TABLE (1 mark for this or similar example). DML is about performing operations on those structures (1 mark) such as queries and updates (1 mark) using commands like SELECT (1 mark).***

**b)** Explain what is meant by the term *tuple* in the relational model. **2**

***A tuple is an instance of an entity (1 mark). It corresponds to a row in a database table (1 mark).***

**c)** With the use of an example explain what is meant by the term *cardinality* in the relational model. **2**

***Cardinality refers to the number of instances of an entity that there can be with regard to another entity / accept: the number of rows in table (1 mark) +1 mark for example (such as one-to-one, one-to-many, many-to-many)***

**Total 10 Marks**

**Marks**

## Question 6

**a)** Give a definition of logical database design. **2**

***The data is designed in terms of the chosen model such as the relational model (1 mark) but independently of any DBMS (1 mark).***

**b)** Identify FOUR (4) of the activities undertaken during logical design. **4**

***Identify entities***
***Identify relationships***
***Identify and associate attributes with entities***
***Determine attribute domains***
***Determine candidate, primary and alternative key attributes***
***Move from entities to tables.***
***Document in a data dictionary***
***1 mark identifying each activity (up to 4 marks)***

**c)** Discuss why different levels of access are required for different users when accessing a database system. **4**

***It is important to recognise that not every user is the same (1 mark). Users will need to access different parts of a system in different ways depending on their job role (1 mark) within the organisation and/or the level of authority that they hold (1 mark). This is also good practice in terms of security to prevent unauthorised access (1 mark).***

**Total 10 Marks**

**Marks**

## Question 7

**a)** Outline the chief activities in the physical design stage of database development. **4**

***In physical design the database is designed with the target Database Management System (DBMS) in mind. (1 mark) Therefore it entails knowledge of the chosen DBMS whether it is Oracle, MySQL, SQL Server etc. (1 mark). Physical design entails designing the actual data structures as they will be implemented using the DBMS. (1 mark) Meta-data (data dictionaries) for structures such as tables, columns etc. will be defined.(1 mark)***

**b)** With the use of an example define business rules **and** the role they might have in a database. **4**

***Business rules are rules about the way an organisation goes about its operations (1 mark). Many such rules can be represented in the normal constraints that exist in a database for example referential integrity and domains (1 mark). However there may be a number of rules of a type that require more complex representation (1 mark). For example if we have an order in a sales system that can only have up to 10 items. (1 mark)***

**c)** Identify TWO (2) items of information that would need to be recorded for each column of a table at the physical design stage. **2**

- ***The domain of that column including the data-type, length and any additional constraints that apply to that column***
- ***A default value for the column***
- ***Whether the column is derived and if it is derived then how it is computed***
- ***Whether or not the column can contain null values***

***1 mark for any of the above points up to 2 marks.***

**Total 10 Marks**

**Marks**

**Question 8**

**a)** Identify **and** explain THREE (3) different types of indexes used in a database system. **6**

***The different types of indexes available are primary index, secondary index and clustering index. (1 mark each for identifying up to 3 marks)***
***A primary index is built around a key field with a unique value for every entry.***
***A secondary index is used for specifying additional key columns.***
***A clustering index is built around a field which can have many corresponding rows in the table.(+1 mark for each elaboration)***

**b)** Explain TWO (2) disadvantages of database indexes. **2**

***Disadvantages of indexes can be the overheads that sometimes accompany their use (1 mark).***
***Adding data or updating data on a table that is indexed will usually take longer as changes (1 mark)***

**c)** Identify TWO (2) types of columns that are likely to be indexed in a database system. **2**

***The columns used most often in joins (1 mark)***
***The columns that are used for ordering (1 mark)***

**Total 10 Marks**

**Marks**

**Question 9**

Top Plate is a ceramics company that manufactures crockery items such as cups and plates. They are based in Stoke-on-Trent in the United Kingdom. They have spread their business from its original base and are now a large multinational company that deals with the manufacture, import and export of their products. They operate in a number of Asian, South American and North American countries. They have sales branches in more than one country. Their manufacturing bases are in India and Brazil. From there, goods are shipped to the sales branches. Their management offices which oversee the whole business are still in Stoke-on-Trent.

**a)** Describe FIVE (5) advantages Top Plate might gain by implementing a distributed database. **5**

- ***Emulating organisational structure.***
- ***Greater control. Data is located where it is needed and so it is easier to control.***
- ***Improved availability. Locally stored data is less likely to have access disrupted by network problems or issues with the central database.***
- ***Greater reliability.***
- ***Better performance. where most data is retrieved from a local version (either replicated or fragmented) this could result in an increase in performance***
- ***Easier Growth. Where local sites have control over data relevant to them it is easier for them to plan and structure the growth of the database.***
- ***1 mark each for any FIVE of these points.***

***Alternative answers should be credited.***

**b)** Why might security be an issue in a distributed database? **2**

***Data travelling over possibly insecure networks (1 mark). Local copies of the data might exist in situations where the security is not as robust as the central site (1 mark).***

**c)** Identify THREE (3) implementation or performance factors that need to be considered when choosing a DBMS for a distributed database. **3**

**Marks**

***Any DBMS that needs to operate for a distributed database will be more complex than for a stand-alone database (1 mark).***

***More complex system catalogue. To keep track of potentially different structures in different sites and to record location of specific structures (1 mark).***

***Concurrency control. Needs to be more sophisticated for example to make sure an update that affects more than one site operates in such a way to keep the database consistent and maintain integrity (1 mark)***

***Query optimiser. Tracing paths through tables for queries becomes more complex when the tables way be fragmented across different sites (1 mark)***

***Finally the DBMS itself must be distributed which means that tuning and managing it becomes a more complex set of tasks (1 mark)***

***1 mark for each issue identified up to 3.***

***Alternative answers should be credited.***

**Total 10 Marks**

**Marks**

**Question 10**

**a)** Outline the key reasons that organisations are increasingly implementing data warehouse systems. **6**

***The two main reasons for a data warehouse are the need for business intelligence (1mark) and as a way of integrating data from different systems.(1 mark)***

***Additional marks (up to 4 can be given for relevant points such as those below)***

***Business intelligence is needed in a business environment that is increasingly competitive. Organisations compete with one another for shares of markets and customers. This can even be true of public sector and charitable organisations, for example colleges competing for students or different charities competing for donations. The 'classic' firm in economics (private sector proving goods or services) is always in competition with other firms providing the same or similar goods or services.***

***In order to effectively compete, organisations need to have knowledge of the market, customers, suppliers etc. This will enable them to make effective decisions.***

***As organisations have implemented IT solutions it has led to a proliferation of different systems. A data warehouse is a way of bringing these systems together in a way that provides an organisation with the sort of business intelligence they need.***

**b)** Identify **and** discuss TWO (2) types of tools that can be used for interaction with a data warehouse. . **4**

***Query tools: using a standard query language like SQL or query languages specific to the commercial tool.***
***OLAP: On-line analytical processing. Which allows multi-dimensional views of data.***
***Statistics: Statistical analysis tools.***
***Discovery tools. Used to uncover trends in database. What is known as 'data mining' whereby the tools discover trends that the users might not necessarily be aware of.***
***1 mark for identifying (up to 2 marks)***
***1 mark for detail (up to 2 marks).***

**Total 10 Marks**

# End of Paper

**Marks**

## Learning Outcomes matrix

| Question | Learning Outcomes assessed | Marker can differentiate between varying levels of achievement |
|---|---|---|
| 1 | 1,3 | Yes |
| 2 | 3 | Yes |
| 3 | 2,3 | Yes |
| 4 | 4 | Yes |
| 5 | 5 | Yes |
| 6 | 4 | Yes |
| 7 | 4 | Yes |
| 8 | 1,2 | Yes |
| 9 | 1 | Yes |
| 10 | 1 | Yes |

## Grade descriptors

| Learning Outcome | Pass | Merit | Distinction |
|---|---|---|---|
| Understand the enterprise application of database systems | Demonstrate adequate level of understanding | Demonstrate robust level of understanding | Demonstrate highly comprehensive level of understanding |
| Understand how to enhance the design of and further develop a database system | Demonstrate ability to perform the task | Demonstrate ability to perform the task consistently well | Demonstrate ability to perform the task to the highest standard |
| Be able to enhance a logical database design | Demonstrate ability to perform the task | Demonstrate ability to perform the task consistently well | Demonstrate ability to perform the task to the highest standard |
| Be able to develop a physical database design | Show adequate development | Show sound and appropriate development | Show innovative and highly appropriate development |
| Be able to enhance a database system using SQL | Demonstrate ability to perform the task | Demonstrate ability to perform the task consistently well | Demonstrate ability to perform the task to the highest standard |