

**Data Technician**

|  |
| --- |
|  |

|  |
| --- |
| Name: |
| Course Date: |
|  |

**Table of contents**

[Day 1: Task 1 3](#_Toc106466110)

[Day 1: Task 2 3](#_Toc1673105856)

[Day 3: Task 1 4](#_Toc1682646377)

[Day 4: Task 1: Written 6](#_Toc135095808)

[Day 4: Task 2: SQL Practical 9](#_Toc1228298484)

[Course Notes 19](#_Toc631049525)

[Additional Information 19](#_Toc1612679949)

# Day 1: Task 1

Please research and complete the below questions relating to key concepts of databases.

|  |  |
| --- | --- |
| What is a primary key? | A primary key is a field/column where it’s values uniquely identifies each row. |
| How does this differ from a secondary key? | A primary key is a mandatory requirement by databases and secondary are not required but can aid to speed up searches. |
| How are primary and foreign keys related? | A primary key is native to a table and identifies the records in that table, however, when that same primary key is used in a different table, it becomes a foreign key in the new table and therefore creates a relationship with the first table. Primary keys are unique to one table but that same table can have multiple foreign keys. |
| Provide a real-world example of a one-to-one relationship | A citizen has only one driving license. In the digital world, ideally, an email should only have one user. |
| Provide a real-world example of a one-to-many relationship | One digital subscription product can have many subscribers, theoretically unlimited number of subscribers. |
| Provide a real-world example of a many-to-many relationship | A train can have many journeys and many passengers, in a day this may be easy to manage, but over time, navigating through a train’s relationships will be difficult. |

# Day 1: Task 2

Please research and complete the below questions relating to key concepts of databases.

|  |  |
| --- | --- |
| What is the difference between a relational and non-relational database? | In an RDBMS, there must be simple or complex relationships with data stored in tables, while non-RDBMS doesn’t not need to have relationships in-between the data set. |
| What type of data would benefit off the non-relational model?  Why? | The store of digital files: videos, pictures, application files. These records do not need to be linked with other sets of data and typically the users of such media want flexibility in how they organise the data. Whereas an RDBMS must have relationships between the data, using an RDBMS will only complicate the users of media files needs. |

# Day 3: Task 1

Please research the below ‘JOIN’ types, explain what they are and provide an example of the types of data it would be used on.

|  |  |
| --- | --- |
| Self-join | This allows a table to be joined to itself, which enables comparisons between different rows in the same table, but it must use aliases.  It is particularly useful to find relationship within one table like comparing customers and the shopping habits. Also, useful to understand hierarchical relationships in a data table. |
| Right join | This joins all rows/records from the right table with matching records from the left table.  Similar to Left join below, this can be useful to find a customer that has met a condition like those that have made an order, have done the same action. |
| Full join | This joins all rows from both tables regardless of matching or non-matching as well as joining NULL values, if they are there. Useful to see when you need to see all data regardless of matching records. |
| Inner join | This combines two tables based on a common column and a condition. It is useful for finding, say a group of customers that meet a certain condition like having placed an order or having the same status of service or product request. |
| Cross join | This join combines every row from one table with all the rows another table in every combination possible. Essentially, similar to a multiplication matrix.  Apart from creating test data for exploratory analysis, cross join can be used in a case you need to see products that have different dimensions like size. |
| Left join | This join returns all values from the left table with matching records from the right table and returns null values if there aren’t any matching records.  This can be useful when you need to see all customers and compare the ones that have met conditions versus others that haven’t. Like if customers have made orders or not. |

# Day 4: Task 1: Written

In your groups, discuss and complete the below activity. You can either nominate one writer or split the elements between you. Everyone however must have the completed work below:

*Imagine you have been hired by a small retail business that wants to streamline its operations by creating a new database system. This database will be used to manage inventory, sales, and customer information. The business is a small corner shop that sells a range of groceries and domestic products. It might help to picture your local convenience store and think of what they sell. They also have a loyalty program, which you will need to consider when deciding what tables to create.*

*Write a 500-word essay explaining the steps you would take to set up and create this database. Your essay should cover the following points:*

1. ***Understanding the Business Requirements****:*
   1. *What kind of data will the database need to store?*
   2. *Who will be the users of the database, and what will they need to accomplish?*
2. ***Designing the Database Schema****:*
   1. *How would you structure the database tables to efficiently store inventory, sales, and customer information?*
   2. *What relationships between tables are necessary (e.g., how sales relate to inventory and customers)?*
3. ***Implementing the Database****:*
   1. *What SQL commands would you use to create the database and its tables?*
   2. *Provide examples of SQL statements for creating tables and defining relationships between them.*
4. ***Populating the Database****:*
   1. *How would you input initial data into the database? Give examples of SQL INSERT statements.*
5. ***Maintaining the Database****:*
   1. *What measures would you take to ensure the database remains accurate and up to date?*
   2. *How would you handle backups and data security?*

*Your essay should include specific examples of SQL commands and explain why each step is necessary for creating a functional and efficient database for the retail business.*

|  |  |
| --- | --- |
| Please write your 500-word essay here | Implementing a New Database System Part 1 In our group we were “hired” by the business, **Running Big Savers** to streamline their database system to improve their operations.  After deliberating, we concluded to have simplify our database in the first implementation with inventory data, sales data, and customer data.  The business sells a few items, five essentials; milk, bread, eggs, detergent, and cake. We decided not to track details of their suppliers in this stage of capturing data.  We determined the database would hold products, customers, sales, and employee data. Tracking loyalty points was very important to the owners of **Running Big Savers.**  Therefore, we decided to create triggers embedded in the database that would count each pound(currency) of sales for a customer as one (1) loyalty point. This can be found in a DEMLIMITER statement below courtesy of Rodrigo, who handled all database and sql tasks. Part 2SCHEMA   Customers   * CustomerID: Primary Key * Name: full name * Email * LoyaltyPoints   Products   * ProductID * Name * StockQuantity * Price   Sales   * SalesID * SalesDate * CustomerID * ProductID   Employee   * EmployeeID * Name * Role * Permissions   The owners decided to set the prices of the items at.  Milk   * Price: £1.50 * Quantity: 100   Bread   * Price: £1.00 * Quantity: 200   Detergent   * Price: £5.99 * Quantity: 500   Egg   * Price: £2.00 * Quantity: 300   Cake   * Price: £20.00 * Quantity: 15   Find A Schema diagram of the database below.  A screenshot of a computer  Description automatically generated    The database will be implemented using mysql and MySQL Workbench as visual tool to create and manage the database. Part 3 and 4Code to Create the Database. CREATE DATABASE RunningBigSavers; use RunningBigSavers; show tables; Code to Create the Tables. CREATE TABLE Products (     ProductID INT PRIMARY KEY AUTO\_INCREMENT,     Name VARCHAR(255) NOT NULL,     StockQuantity INT NOT NULL,     Price DECIMAL(5,2) NOT NULL );  CREATE TABLE Customers (     CustomerID INT PRIMARY KEY AUTO\_INCREMENT,     Name VARCHAR(255) NOT NULL,     Email VARCHAR(255) UNIQUE NOT NULL,     LoyaltyPoints INT DEFAULT 0 );  CREATE TABLE Sales (     SaleID INT PRIMARY KEY AUTO\_INCREMENT,     SaleDate DATE NOT NULL,     CustomerID INT,     ProductID INT,     Quantity INT NOT NULL,     TotalPrice DECIMAL(5,2) NOT NULL,     FOREIGN KEY (CustomerID) REFERENCES Customers(CustomerID),     FOREIGN KEY (ProductID) REFERENCES Products(ProductID) );  CREATE TABLE Employees (     EmployeeID INT PRIMARY KEY AUTO\_INCREMENT,     Name VARCHAR(255) NOT NULL,     Role VARCHAR(255) NOT NULL,     Permissions TEXT NOT NULL ); Loyalty Points Database Algorithm DELIMITER $$   CREATE TRIGGER UpdateLoyaltyPoints   AFTER INSERT ON Sales   FOR EACH ROW   BEGIN       UPDATE Customers       SET LoyaltyPoints = LoyaltyPoints + NEW.TotalPrice       WHERE CustomerID = NEW.CustomerID;   END$$   DELIMITER; Part 5Roles of Employees in Running Big Savers  1. Shop Manager  * Track inventory levels -needs to track stock and sales. * Analyse sales trends to make informed decisions. * Check customer loyalty points for rewards and discounts.  1. Shop Staff  * Record transactions at the point of sale. * Check stock availability and update inventory. * Check customer loyalty points for rewards and discounts.  1. IT Support Team  * Support and secure the database to ensure smooth operation. * Perform backups to prevent data loss. * Optimize database performance.    Code to Populate Products Table INSERT INTO Products (Name, StockQuantity, Price) VALUES ('Bread', 20, 1.00); Example Code to Populate Customers Table INSERT INTO Customers (Name, Email, LoyaltyPoints) VALUES ('Jane Doe', 'jane.doe@example.com', 50);   Maintaining the DatabaseRegular Updates  * The inventory will be updated after a sale: Whenever a product is sold, the inventory must be adjusted to reflect the new stock levels. * We need a **trigger** to calculate new stock and automatically reduce the stock levels (quantity). * Updating the records by staff should make this activity happen.  Updating Customer Loyalty Points:  * If a store has a loyalty program, customers should earn points after every purchase. The database must update these points accordingly.  Monitoring Data Quality  * Checking for Duplicate Customers * Ensuring Product Information is Complete and correct. * This is responsibility IT support to check that this automatic “trigger” is functioning. * Maintaining high data quality improves business efficiency and prevents errors that could impact customers or sales.    Backing Up the Database  * Creating Automatic Backups * By maintaining regular backups, the business can recover any lost data quickly and continue operations without disruption. * There needs to be redundancy in the backups. Cloud back up can be done daily in the morning and an on-site backup storage can be in the manager’s office.    Data Security  * Restricting Access to Sensitive Data: Only managers should have the ability to change product prices, while cashiers should only be able to process sales transactions. * IT Support will have full access but authority level approval should be in place to control the change of any sensitive data such as price, staff information, customer information and more.    Securing Customer Information:   * Customer data, including email addresses and phone numbers, should be protected using Data protection act and GDPR as a policy. * Only authorized personnel should be able to access or update this information. * And these authorised personnel should be the manager and the data owners.    Stored procedures Stored procedures will be created to create the above rules in the database to ensure data quality and security.   * To Update Loyalty Points. * To Update Stock each time a sale is made. * To prevent a product being added with incomplete information. * To validate data entry with all tables; customers, products, and employees. * NB. Sales data is created by the trigger of a point of sales (POS) system. |

# Day 4: Task 2: SQL Practical

In your groups, work together to answer the below questions. It may be of benefit if one of you shares your screen with the group and as a team answer / take screen shots from there.

**Setting up the database:**

1. **Download world\_db(1)**
2. **Follow each step to create your database**

**For each question I would like to see both the syntax used and the output.**

1. **Count Cities in USA:** *Scenario:* You've been tasked with conducting a demographic analysis of cities in the United States. Your first step is to determine the total number of cities within the country to provide a baseline for further analysis.

|  |
| --- |
| A screenshot of a computer  Description automatically generated |

1. **Country with Highest Life Expectancy:** *Scenario:* As part of a global health initiative, you've been assigned to identify the country with the highest life expectancy. This information will be crucial for prioritising healthcare resources and interventions.

|  |
| --- |
| A screenshot of a computer  Description automatically generated |

1. **"New Year Promotion: Featuring Cities with 'New :** *Scenario:* In anticipation of the upcoming New Year, your travel agency is gearing up for a special promotion featuring cities with names including the word 'New'. You're tasked with swiftly compiling a list of all cities from around the world. This curated selection will be essential in creating promotional materials and enticing travellers with exciting destinations to kick off the New Year in style.

|  |
| --- |
| A screenshot of a computer  Description automatically generated |

1. **Display Columns with Limit (First 10 Rows):** *Scenario:* You're tasked with providing a brief overview of the most populous cities in the world. To keep the report concise, you're instructed to list only the first 10 cities by population from the database.

|  |
| --- |
| A screenshot of a computer  Description automatically generated |

1. **Cities with Population Larger than 2,000,000:** *Scenario:* A real estate developer is interested in cities with substantial population sizes for potential investment opportunities. You're tasked with identifying cities from the database with populations exceeding 2 million to focus their research efforts.

|  |
| --- |
| A screenshot of a computer  Description automatically generated |

1. **Cities Beginning with 'Be' Prefix:** *Scenario:* A travel blogger is planning a series of articles featuring cities with unique names. You're tasked with compiling a list of cities from the database that start with the prefix 'Be' to assist in the blogger's content creation process.

|  |
| --- |
| A screenshot of a computer  Description automatically generated |

1. **Cities with Population Between 500,000-1,000,000:** *Scenario:* An urban planning committee needs to identify mid-sized cities suitable for infrastructure development projects. You're tasked with identifying cities with populations ranging between 500,000 and 1 million to inform their decision-making process.

|  |
| --- |
| A screenshot of a computer  Description automatically generated |

1. **Display Cities Sorted by Name in Ascending Order:** *Scenario:* A geography teacher is preparing a lesson on alphabetical order using city names. You're tasked with providing a sorted list of cities from the database in ascending order by name to support the lesson plan.

|  |
| --- |
| A screenshot of a computer  Description automatically generated |

1. **Most Populated City:** *Scenario:* A real estate investment firm is interested in cities with significant population densities for potential development projects. You're tasked with identifying the most populated city from the database to guide their investment decisions and strategic planning.

|  |
| --- |
| A screenshot of a computer  Description automatically generated |

1. **City Name Frequency Analysis: Supporting Geography Education** *Scenario*: In a geography class, students are learning about the distribution of city names around the world. The teacher, in preparation for a lesson on city name frequencies, wants to provide students with a list of unique city names sorted alphabetically, along with their respective counts of occurrences in the database. You're tasked with this sorted list to support the geography teacher.

|  |
| --- |
| A screenshot of a computer  Description automatically generated |

1. **City with the Lowest Population:** *Scenario:* A census bureau is conducting an analysis of urban population distribution. You're tasked with identifying the city with the lowest population from the database to provide a comprehensive overview of demographic trends.

|  |
| --- |
| A screenshot of a computer  Description automatically generated |

1. **Country with Largest Population:** *Scenario:* A global economic research institute requires data on countries with the largest populations for a comprehensive analysis. You're tasked with identifying the country with the highest population from the database to provide valuable insights into demographic trends.

|  |
| --- |
| A screenshot of a computer  Description automatically generated |

1. **Capital of Spain:** *Scenario:* A travel agency is organising tours across Europe and needs accurate information on capital cities. You're tasked with identifying the capital of Spain from the database to ensure itinerary accuracy and provide travellers with essential destination information.

|  |
| --- |
| A screenshot of a computer  Description automatically generated |

1. **Country with Highest Life Expectancy:** *Scenario:* A healthcare foundation is conducting research on global health indicators. You're tasked with identifying the country with the highest life expectancy from the database to inform their efforts in improving healthcare systems and policies.

|  |
| --- |
| A screenshot of a computer  Description automatically generated  Andorra |

1. **Cities in Europe:** *Scenario:* A European cultural exchange program is seeking to connect students with cities across the continent. You're tasked with compiling a list of cities located in Europe from the database to facilitate program planning and student engagement.

|  |
| --- |
| A screenshot of a computer program  Description automatically generated |

1. **Average Population by Country:** *Scenario:* A demographic research team is conducting a comparative analysis of population distributions across countries. You're tasked with calculating the average population for each country from the database to provide valuable insights into global population trends.

|  |
| --- |
| A screenshot of a computer  Description automatically generated |

1. **Capital Cities Population Comparison:** *Scenario:* A statistical analysis firm is examining population distributions between capital cities worldwide. You're tasked with comparing the populations of capital cities from different countries to identify trends and patterns in urban demographics.

|  |
| --- |
| A screenshot of a computer  Description automatically generated |

1. **Countries with Low Population Density:** *Scenario:* An agricultural research institute is studying countries with low population densities for potential agricultural development projects. You're tasked with identifying countries with sparse populations from the database to support the institute's research efforts.

|  |
| --- |
| A screenshot of a computer  Description automatically generated |

1. **Cities with High GDP per Capita:** *Scenario:* An economic consulting firm is analysing cities with high GDP per capita for investment opportunities. You're tasked with identifying cities with above-average GDP per capita from the database to assist the firm in identifying potential investment destinations.

|  |
| --- |
| A screenshot of a computer  Description automatically generated |

1. **Display Columns with Limit (Rows 31-40):** *Scenario:* A market research firm requires detailed information on cities beyond the top rankings for a comprehensive analysis. You're tasked with providing data on cities ranked between 31st and 40th by population to ensure a thorough understanding of urban demographics.

|  |
| --- |
| A screenshot of a computer  Description automatically generated |

|  |
| --- |
| **Course Notes** |

It is recommended to take notes from the course, use the space below to do so, or use the revision guide shared with the class:

|  |
| --- |
| **Additional Information** |

We have included a range of additional links to further resources and information that you may find useful, these can be found within your revision guide.

**END OF WORKBOOK**

**Please check through your work thoroughly before submitting and update the table of contents if required.**

**Please send your completed work booklet to your trainer.**