

**Data Technician**

|  |
| --- |
|  |

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

**Table of contents**

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

[Day 1: Task 2 4](#_Toc1673105856)

[Day 3: Task 1 5](#_Toc1682646377)

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

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

[Course Notes 25](#_Toc631049525)

[Additional Information 26](#_Toc1612679949)

# Day 1: Task 1

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

|  |  |
| --- | --- |
| What is a primary key? | Primary key is a unique identifier of data records in a table within a database. It is used to provide identification for data and to prevent data duplication. |
| How does this differ from a secondary key? | Secondary key can have duplications and can be a different data type comparing to a primary key, it can be a part of the primary key, and is commonly used for easier data search within a database. |
| How are primary and foreign keys related? | Primary key is a unique data within one table, while foreign key is a primary key in a second table that was taken from the first table. Foreign key is the same primary key from the previous table, but is used in a different table to create a relationship. |
| Provide a real-world example of a one-to-one relationship | A persona can have only one national insurance number, or a residence permit card. |
| Provide a real-world example of a one-to-many relationship | A movie can have many reviews at once, or a person can possess many items at once. |
| Provide a real-world example of a many-to-many relationship | A student may have many teachers, while teachers may have many students to teach. |

# 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? | The two types of databases have their advantages and disadvantages and are used for different purposes, relational databases are commonly used to provide consistency where a clear structure is required, they always have a fixed schema and a table can define relationships between tables. Non-relational database, on the other hand, is usually used to store unstructured or semi-structured data, is horizontally scalable and has fast processing speed. Non-relational databases don’t handle relationships between databases. |
| What type of data would benefit off the non-relational model?  Why? | Unstructured and semi-structured data types would benefit from non-relational model for the model being horizontally scalable, meaning it can provide higher processing speed for the unstructured data, which is difficult to read for relational models. |

# 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 | Self-join is a method in databases that allows a table to be joined with itself, creating a way to explore relationships inside the same dataset. Unlike other types of joins where two or more different tables are combined, a self-join treats one table as if it were two by giving it temporary aliases, this allows rows to be compared to other rows within the same structure. Self-join makes hidden connections in data visible where items in the same dataset have some kind of dependency or hierarchy. By using this join, it becomes possible to show relationships that are not obvious at first look, and this makes a simple table into a more dynamic and complex source of information. An example of self-join could be having an employee table where a name of an employee could be alongside a manager’s ID, but they both are employees of a company, with self-join it is possible to match every employee to their manager and to show these pairs in a single result. |
| Right join | Right join focuses on keeping all rows from a table on the right side of a query, while only including rows of the left table if they match. Right join is useful when the main aim is to keep the data on the right ensuring no information is lost from the right side even if some connections are missing due to the join filling the left table values with null if there is no match. Integers, strings, and decimals — are commonly used because they clearly define the relationships between two tables and make the join meaningful. |
| Full join | Full join in SQL combines results of both left and right joins together, it keeps all rows from both of the tables even if there are no matches. If there is a match – the data is connected and shows together, in case of mismatch the unmatched sides is filled with the null value. Full join is commonly used to complete an overall picture of information from two databases ensuring nothing is lost on the both of the sides. Data types such as integers for IDs, descriptive attributes for titles or description, or other details with use of dates and decimal numbers, are commonly used in full join to provide link between tables, while text (string) gives some meaningful context. |
| Inner join | \ |
| Cross join | sss |
| Left join |  |

# 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 | 1. 1a. *What kind of data will the database need to store?*   The first steps that should be made before defining the scope of the project and implementing the solution is to define client’s requirements to define the project scope, any additional adjustments and the essential data to be stored. The business owner should provide information about their requirements and data if any exists to be immediately used in the database. Discussing the loyalty program will be useful too to clarify how the business owner wants to use it to gain profit, therefore which data should be stored.  The database will store the essential data about the customers, sales, items and inventory, and loyalty program data for the customers. The tables in the table will include necessary data about all of the listed data arrays above such as profits, item prices, customer details, etc.. The users of the database will be the IT staff who will manage and adjust the database for the business to improve efficiency and to populate the database, other team members such as sales assistants or shop managers can take part in data input for the database, but only with restrictions to keep the database safe from invalid data or any other cyber threat that may appear during the databases lifetime. |

# 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.

|  |
| --- |
|  |

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.

|  |
| --- |
|  |

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.

|  |
| --- |
|  |

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.

|  |
| --- |
|  |

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.

|  |
| --- |
|  |

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.

|  |
| --- |
|  |

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.

|  |
| --- |
|  |

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.

|  |
| --- |
|  |

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.

|  |
| --- |
|  |

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.

|  |
| --- |
|  |

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.

|  |
| --- |
|  |

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.

|  |
| --- |
|  |

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.

|  |
| --- |
|  |

1. **Country with Shortest 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.

|  |
| --- |
|  |

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.

|  |
| --- |
|  |

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.

|  |
| --- |
|  |

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.

|  |
| --- |
|  |

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.

|  |
| --- |
|  |

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.

|  |
| --- |
|  |

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.

|  |
| --- |
|  |

|  |
| --- |
| **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.**