

**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 table which indicates unique row in table. |
| How does this differ from a secondary key? | Secondary key is a alternative key, which usually linked to primary key and contains additional info |
| How are primary and foreign keys related? | To identify and manage data records and they both are indexed to improve query perfomance. Usually have shared goals, with different hierarchy |
| Provide a real-world example of a one-to-one relationship | Person and a national insurance number. Normally, IN available only per person at a time. |
| Provide a real-world example of a one-to-many relationship | Costumer and orders. A costumer may have either one or multiple orders at a time. But orders cannot have several costumers. |
| Provide a real-world example of a many-to-many relationship | Products and categories where product might belong to several categories at a time and categories contains lot of different products. |

# 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? | Structure, data organisation, the way of querying and the purpose of usage. Where in the relational one its a structed data with already existed ways to apply it. And non relational databases are flexible and can be stored in json files. |
| What type of data would benefit off the non-relational model?  Why? | A data with lot of changes with quick velocity, unstructured data, big data with unpredictable tendency such as twitter etc.  It has wider variety of structure, gives an opportunity to apply changes without disturbing schema, give access to complex relationships wich works better with graph 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 | Where a table join with itself. Very good practice for comparing different column’s data records. |
| Right join | It takes relevant info from first table and join it, but it differs from the left join by a fully provided info. For example, if there’s no matched information with right table, it’ll bring row with null as result. |
| Full join | It returns all the record which mentioned in table 1 and 2 in full |
| Inner join | Can return only data which matches in both tables |
| Cross join | This join works better with WHERE, otherwise all the information in both tables would lose it’s meaning. Works as a math multiplying function 2\*2=4. Doesn’t rely much on relatioships between tables. |
| Left join | It is including all data from both tables (eg. If it’s null at only few columns, it’ll return null). |

# 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 | We essentially need to understand data and being able to identify the area it will be stored. For different companies, it may vary. Database should support tracking stock (numbers of items available) and purchases customers are pending. It is very useful for all parts of a company: eg. Sales team need to update inventory or managers, who’ll need to have sales reports. It will include product name, category and quantity. From customer’s side it might include any personal information, any sales details or details of the product.  For that we’d create 3 tables with costumer’s info, sales and product details. In product table we’ll include name, price, quantity in stock and category. In costumers table would be name, phone number, address and loyalty program. Sales will have date of sale, price, seller info, quantity and total price.  Relationships are necessary: a foreign key in Sales (e.g., customer\_id) relates to Customers, and another (e.g., product\_id) ties to Products, reflecting how sales connect customers and inventory. This normalization reduces redundancy and ensures data integrity.  By command CREATE DATABASE we created Retail\_db, the we started USE retail\_db. Then we create table Product with rows: product\_id, product\_title, then category, price and quantity in stock.  After it worked, we need to create a Customers table with customer\_id, name, phone number, address and discount.  Finally, we created table Sales and connected it to Costumers and Products tables via Foreign key.  We need to use command INSERT INTO and provide all ids we used for each table with relevant information. |

# 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)** [**here**](https://justit831-my.sharepoint.com/:u:/g/personal/danpe_justit_co_uk/Ef6vAaaYVi5FhHqKGxqnn60B9g2khoYekEIO3Y7J00UcJQ?e=pv9NNE)
2. **Follow each step to create your database** [**here**](https://justit831-my.sharepoint.com/:b:/g/personal/danpe_justit_co_uk/EdeCKl2Sas1Hl7u9amDy0fIB9jGVCKxSR0u2-lFOvS5rXw?e=xKv1U7)

**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.

|  |
| --- |
| Syntax is :  SELECT COUNT(\*) AS total\_cities  FROM city  WHERE countrycode = ‘USA’; |

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.

|  |
| --- |
| SELECT Code, LifeExpectancy  FROM country  ORDER BY LifeExpectancy DESC  LIMIT 1; |

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.

|  |
| --- |
| SELECT name, countrycode  FFOM city  WHERE name LIKE ‘%New%’; |

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.

|  |
| --- |
| SELECT name, population  FROM city  ORDER BY Population DESC  LIMIT 10; |

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.

|  |
| --- |
| SELECT name, population  FROM city  WHERE population > 2000000; |

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.

|  |
| --- |
| SELECT name, countrycode  FFOM city  WHERE name LIKE ‘%Be%’; |

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.

|  |
| --- |
| SELECT name, countrycode  FROM city  WHERE population BETWEEN 500000 AND 1000000 |

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.

|  |
| --- |
| SELECT name  FROM city  ORDER BY name ASC; |

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.

|  |
| --- |
| SELECT name, countrycode  FROM city  ORDER BY population DESC  LIMIT 1; |

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.

|  |
| --- |
| SELECT name, COUNT(\*) as Frequency  FROM city  GROUP BY name  ORDER BY name ASC; |

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.

|  |
| --- |
| SELECT name, countrycode  FROM city  ORDER BY population ASC  LIMIT 1; |

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.

|  |
| --- |
| SELECT name  FROM country  ORDER BY population DESC  LIMIT 1; |

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.

|  |
| --- |
| SELECT capital as capital\_name  FROM country  WHERE name = ‘SPain’; |

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.

|  |
| --- |
| SELECT name AS EUROPE  FROM country  WHERE continent = 'Europe'; |

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.

|  |
| --- |
| SELECT CountryCode, AVG(Population) AS Average\_Population  FROM City  GROUP BY countrycode; |

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.

|  |
| --- |
| SELECT Name, Population, SurfaceArea, (Population / SurfaceArea) AS PopulationDensity  FROM country  WHERE (Population / SurfaceArea) < 10  ORDER BY PopulationDensity ASC;    Frankly speaking, here it was hard and been done with a help of AI... |

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.**