# 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 column (or set of columns) in a database table that uniquely identifies each record. No two rows in the table have the same primary key value. This maintains data integrity and makes retrieving information efficient. |
| **How does this differ from a secondary key?** | A candidate key is not chosen as the primary key, or a field used for indexing/searching, however not chosen as a primary key. |
| **How are primary and foreign keys related?** | A foreign key is a column in one table that refers to the primary key in another table. |
| **Provide a real-world example of a one-to-one relationship** | Take passports for example, Each person has exactly one passport.  Each passport belongs to exactly one person. |
| **Provide a real-world example of a one-to-many relationship** | One customer can place many orders.  Each order belongs to exactly one customer. |
| **Provide a real-world example of a many-to-many relationship** | One student can enrol on many courses.  One course can have many students. |

Day 1 Task 2

|  |
| --- |
| What is the difference between a relational and non-relational database? |
| What type of data would benefit off the non-relational model?  Why? |

Relational database stores structured data in tables with rows and columns, its linked by keys, and uses SQL for queries. A non-relational database (NoSQL) stores data in flexible formats like documents, key-value pairs, or graphs, making it better for unstructured or large-scale data.

Unstructured or large scale data, such as graphs and large documents.

NoSQL databases store data in a highly flexible way.

# 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** | A self-join occurs when a table is joined to itself, typically by using table aliases to distinguish between the two instances of the same table within the query. This approach allows for the comparison or association of rows within a single dataset.  Self-joins are particularly useful in scenarios involving hierarchical data, such as when representing relationships between employees and their managers in an organisation. They can also be applied to identify relationships among entries within a dataset, for example, customers residing in the same city. Additionally, self-joins are employed to compare rows for tasks such as detecting duplicate entries or analysing sequential data within the table. |  |
| **Right join** | Retrieves all records from the right table, along with any corresponding records from the left table that meet the join criteria.  Lists all orders even without customer records. |  |
| **Full join** | Returns every row from both tables, matching where possible and filling in gaps for unmatched rows—essentially combining left and right joins.  Afull join is used when you want to combine two tables and keep all records from both, even if they don’t have matching values. On values are recorded as Null values. |  |
| **Inner join** | An inner join returns only those records where a matching value exists in both tables. This means that if a row from one table does not have a corresponding match in the other table, it will not appear in the results. Only the rows where there is a direct match between the two tables are included in the output, ensuring that the final dataset consists solely of related records from both sources.  An inner join requires related data fields between two tables. Typically, these are: Primary Keys & Foreign Keys. |  |
| **Cross join** | A cross join is a type of join operation that generates the Cartesian product of two tables. This means that every row from the first table is paired with every row from the second table, resulting in a combination of all possible row pairs between the two tables.  Unlike other types of joins, a cross join does not require any specific condition or clause to combine the tables. However, it is possible to apply a filter later on to narrow down the results if needed. The cross join is often used when you need to evaluate all possible combinations between two datasets. |  |
| **Left join** | Left Join A left join is a method used to combine data from two tables, focusing primarily on the left table. It ensures that all rows from the left table are included in the result set. If there are matching rows in the right table, those values are also returned alongside the left table’s data. However, if there is no corresponding match in the right table, the columns from the right table will display NULL values. This approach allows for a comprehensive view of all entries from the left table, regardless of whether they have associated records in the right table. |  |

# 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** | A retail database, when designed properly, serves as the backbone for efficient business management, ensuring accuracy, consistency, and security of essential operational data. Within this framework, retailers can systematically organise and access information concerning suppliers, inventory, customers, purchases, and sales. This document presents a concise overview of such a schema, outlining table structures, relational logic, and the security mechanisms required to maintain a robust data environment.  Establishing a dedicated database environment is fundamental. The foundational steps include creating a new database and activating it for further operations:  CREATE DATABASE RetailDB;  USE RetailDB;  At the heart of an effective retail database are distinct yet interconnected tables:   * Suppliers: Uniquely identifies each supplier, facilitating management and avoiding duplication. * Inventory: Associates each product with a supplier, recording details such as category, price, and available stock. * Sales\_Transactions: Documents each sale, linking it to customers and transaction specifics. * Sales\_Items: Itemises products sold in each transaction, noting quantity and price. * Customer: Compiles essential customer data, including loyalty points for engagement programmes. * Purchase\_Transactions: Logs purchases and updates stock levels to ensure inventory accuracy.   Keys and constraints are critical for relational integrity:   * Primary keys ensure each record is unique. * Foreign keys enforce valid cross-table references, preventing data inconsistencies.   The following examples demonstrate table creation and relational enforcement:  CREATE TABLE Suppliers (  Supplier\_ID INT PRIMARY KEY,  Supplier\_Name VARCHAR(100) NOT NULL,  Email VARCHAR(100),  Contact\_Number VARCHAR(20)  );  CREATE TABLE Inventory (  Product\_ID INT PRIMARY KEY,  Supplier\_ID INT,  Product\_Name VARCHAR(100) NOT NULL,  Product\_Category VARCHAR(50),  Selling\_Price DECIMAL(10,2),  Stock\_Quantity INT,  Stock\_Count INT,  FOREIGN KEY (Supplier\_ID) REFERENCES Suppliers(Supplier\_ID)  );  CREATE TABLE Sales\_Transactions (  Sales\_ID INT PRIMARY KEY,  Loyalty\_Card\_Number VARCHAR(20),  Date DATE,  Amount DECIMAL(10,2),  Discount DECIMAL(5,2),  Quantity INT  );  CREATE TABLE Sales\_Items (  Sales\_Item\_ID INT PRIMARY KEY,  Sales\_ID INT,  Product\_ID INT,  Quantity INT,  Selling\_Price DECIMAL(10,2),  FOREIGN KEY (Sales\_ID) REFERENCES Sales\_Transactions(Sales\_ID),  FOREIGN KEY (Product\_ID) REFERENCES Inventory(Product\_ID)  );  CREATE TABLE Customer (  Loyalty\_Card\_Number VARCHAR(20) PRIMARY KEY,  First\_Name VARCHAR(50),  Last\_Name VARCHAR(50),  Email VARCHAR(100),  Contact\_Number VARCHAR(20),  Loyalty\_Points INT  );  Sample data insertion serves to validate table relationships:  INSERT INTO Suppliers VALUES (1, 'Global Foods Ltd', 'contact@globalfoods.com', '+44-123-456-789');  INSERT INTO Inventory VALUES (101, 1, 'Organic Rice', 'Food', 2.50, 500, 500);  INSERT INTO Customer VALUES ('LC001', 'Alice', 'Johnson', 'alice.johnson@email.com', '+44-111-222-333', 120);  INSERT INTO Sales\_Transactions VALUES (5001, 'LC001', '2025-12-03', 90.00, 5.00, 3);  INSERT INTO Sales\_Items VALUES (7001, 5001, 101, 2, 2.50);  Valid supplier references are enforced through foreign keys, as demonstrated in:  FOREIGN KEY (Supplier\_ID) REFERENCES Suppliers(Supplier\_ID)   * Suppliers support multiple products, enabling a diverse inventory. * Purchases trigger inventory updates for accurate stock counts. * Sales transactions link customers to detailed, itemised purchase records. * Sales facilitate stock management and customer engagement via loyalty programmes and reporting.   Sales records, inclusive of customer and product details, may be retrieved with the following query:  SELECT S.Sale\_ID, C.Customer\_Name, P.Product\_Name, S.Quantity, S.Price  FROM Sales S  JOIN Customers C ON S.Customer\_ID = C.Customer\_ID  JOIN Products P ON S.Product\_ID = P.Product\_ID;  To maintain database security and reliability, it is essential to implement regular backups, role-based access controls, encryption and audits. Keys, constraints, normalisation, triggers, and procedures uphold data integrity, while routine validation minimises errors and ensures consistent performance.  This schema enables retailers to efficiently manage all core business data, ensuring accuracy, integrity, and comprehensive reporting for both operational and analytical needs. By adhering to these practices, the database remains secure, reliable, and optimally functional.  A diagram of a company  AI-generated content may be incorrect.  A diagram of a company  AI-generated content may be incorrect.  The schematic diagrams above show the original conception and follow up of the database showing enhanced relationships to serve the retail business. |

# 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  AI-generated content may be incorrect. |

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  AI-generated content may be incorrect. |

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  AI-generated content may be incorrect. |

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  AI-generated content may be incorrect. |

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  AI-generated content may be incorrect. |

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  AI-generated content may be incorrect. |

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  AI-generated content may be incorrect. |

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  AI-generated content may be incorrect. |

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  AI-generated content may be incorrect. |

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  AI-generated content may be incorrect. |

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.

|  |
| --- |
| A screenshot of a computer  AI-generated content may be incorrect. |

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  AI-generated content may be incorrect. |

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 |

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  AI-generated content may be incorrect. |

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  AI-generated content may be incorrect. |

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  AI-generated content may be incorrect. |

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