|  |  |
| --- | --- |
|  | |
| Name: Jason Tse |
| Course Date: 22/09/2025 |
|  |

**Table of contents**

Day 1: Task 1

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

|  |  |
| --- | --- |
| What is a primary key? | It is an identifier which uniquely identifies a particular record. |
| How does this differ from a secondary key? | They are not indexed. They may take longer to find than primary keys as primary keys are usually numbered. |
| How are primary and foreign keys related? | Foreign keys is a column or a set of columns in one table that links to he primary key of another table. |
| Provide a real-world example of a one-to-one relationship | An example of a one-to-one relationship would be citizens with passports. Each citizen can only hold one passport at a time. This is also the same with driving licenses. |
| Provide a real-world example of a one-to-many relationship | An example of a one-to-many relationship would be a parent with children. Each parent can have many children however all those children will relate to the one parent. Another example would be a customer can place multiple orders, but each order is linked to only one customer. |
| Provide a real-world example of a many-to-many relationship | An example of a many-to-many relationship is a student and their courses. A student can be enrolled in multiple courses, and a single course can have many students enrolled in it. Another example is customers and products. A customer might purchase many different products, and the same product could be purchased by many different customers. |

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 relational databases data is organized into tables (relations) of rows and columns, where each row is a record and each column is an attribute. They use a fixed, predefined schema that defines the structure of the data and the relationships between tables.  Non-relational (can also be known as NoSQL) databases handle unstructured data with flexible schemas, using various models like document, key-value, or graph, and are designed for horizontal scaling and high performance with diverse data types. Non-relational databases can be optimised for specific workloads, such as real-time analytics or high-throughput operations. |
| What type of data would benefit off the non-relational model?  Why? | Types of data that would benefit from this are unstructured or semi-structured, as well as data that requires high scalability, flexibility, and real-time performance. Examples of this include big data from social media and IoT devices, content management systems, real-time applications.  Non-relational databases support dynamic schemas, allowing for easy adaptation to new data types and evolving application requirements. They are designed for horizontal scaling, meaning they can handle increasing loads by adding more servers, which is crucial for large-scale applications. |

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 is a type of join operation in a relational database where a table is joined with itself. It allows you to combine rows from the same table based on a related condition.  For example, a table can have a primary key column, all employees in a company column, and managers that each employee in the company reports to column. |
| Right join | The SQL RIGHT JOIN statement joins two tables based on a common column. It selects records that have matching values in these columns and the remaining rows from the right table.  An example of using SQL RIGHT JOIN is when you want to retrieve all the records from one table (the right table) and only the matching records from a related table (the left table). |
| Full join | SQL FULL JOIN, also known as a FULL OUTER JOIN, is a type of SQL JOIN operation that retrieves all records from both tables, including matching and non-matching records.  SQL FULL JOIN can be used when you want to retrieve all records from both tables, regardless of whether there are matching records for example when merging data. When you need to combine data from two tables into a single result set while preserving all records. |
| Inner join | An INNER JOIN in SQL is a way to combine rows from two or more tables based on a matching condition in a common column, returning only the rows where the join condition is met in both tables.  You would use INNER JOIN to find shared records if you want to see only the users who have placed orders, you would join the Users table with the Orders table. |
| Cross join | The CROSS JOIN is used to generate a paired combination of each row of the first table with each row of the second table.  For example if there are multiple tables, the CROSS JOIN combines each row from the first table with every row in the second table, creating a new row for each combination. It repeats this process for each subsequent row in the first table and so on. |
| Left join | A SQL LEFT JOIN returns all rows from the left table and the matching rows from the right table. This type of join is useful when you want to include all records from one primary table, even if they don't have corresponding entries in a related table, such as showing all customers and any orders they've placed, or all products and their associated sales. |

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:*

* ***Understanding the Business Requirements****:*
* *What kind of data will the database need to store?*
* *Who will be the users of the database, and what will they need to accomplish?*
* ***Designing the Database Schema****:*
* *How would you structure the database tables to efficiently store inventory, sales, and customer information?*
* *What relationships between tables are necessary (e.g., how sales relate to inventory and customers)?*
* ***Implementing the Database****:*
* *What SQL commands would you use to create the database and its tables?*
* *Provide examples of SQL statements for creating tables and defining relationships between them.*
* ***Populating the Database****:*
* *How would you input initial data into the database? Give examples of SQL INSERT statements.*
* ***Maintaining the Database****:*
* *What measures would you take to ensure the database remains accurate and up to date?*
* *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 | The business will need data and tables about it's customers. The "Customer" table will have information such as Customer ID, customer name, email address, phone number. The "Inventory" table; item ID, item name, category, quantity, stock. "Sales" table; sales ID, customer ID, item ID, sales date. A loyalty program such as a club card for the store. The clubcard's table will include clubcard ID, customer ID, clubcard points, creation date, expiry date.  The shop manager will have full access to the data as they will need to keep track of the customers, sales, stock and inventory levels. The shop staff can use the data to record transactions and view and add loyalty points. The IT support will be able to maintain the data to ensure everything runs smoothly.  We can structure the database tables with a snowflake schema. With a snowflake schema we can create tables for each category: "Customer", "Purchases", "Inventory", Sales", "Clubcard". Using these tables we can create cardinalities such as "Purchase ID" in the Customer table to have a one-to-many relationship with Item ID". A one-to-one relationship can be formed by each customer having one clubcard.  Using MySQL, we can create tables with customer data. Here is an example of an SQL syntax of;  Creating a table:  CREATE TABLE Customers (  CustomerID,  First\_Name,  Last\_Name,  City);  Inserting data into a table:  INSERT INTO Customers (CustomerID, First\_Name, Last\_Name, City)  VALUES (559481, Nao, Saiki, Birmingham);  We have to keep our data updated consistently, For example whenever a customer purchases and item we need to know how much is being sold, which customer made the purchase, how much profit was made etc.    We need to check for errors and duplicate items in the data. The data we have should not have errors.  Our data needs be backed up just in case it gets lost or corrupted. We can have multiple copies stored in hard drives as Excel spreadsheets or Lucid charts for example.  To ensure accuracy, staff should update stock quantities after each sale, either automatically. Regular validation checks, such as verifying that stock counts match physical inventory, will also be needed.  The manager of the store will be able to edit all the data for example updating the prices on products, add new stock etc. Staff would only be able to make small additions such as recording sales. This is to make sure the data is secured and stored safely by the manager. Not everyone will have the ability to edit all the data.  For security, user access levels should be set (e.g., staff can record sales, but only the manager can delete records). Backups should be scheduled daily, with copies stored securely offsite or in the cloud to prevent data loss. Encryption and secure logins will help protect customer 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:**

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

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

* **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.

|  |
| --- |
|  |

* **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.

|  |
| --- |
|  |

* **"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.

|  |
| --- |
|  |

* **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.

|  |
| --- |
|  |

* **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.

|  |
| --- |
|  |

* **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.

|  |
| --- |
|  |

* **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.

|  |
| --- |
|  |

* **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.

|  |
| --- |
|  |

* **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.

|  |
| --- |
|  |

* **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.

|  |
| --- |
|  |

* **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.

|  |
| --- |
|  |

* **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.

|  |
| --- |
|  |

* **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.

|  |
| --- |
| Could not find this, SQL seemed to not work. |

* **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.

|  |
| --- |
|  |

* **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.

|  |
| --- |
|  |

* **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.

|  |
| --- |
|  |

* **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.

|  |
| --- |
|  |

* **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.

|  |
| --- |
|  |

* **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.

|  |
| --- |
|  |

* **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.**