Table of Contents

Need for database and DBMS

We live in a world in which information and data is a new fuel. There are many advancements in the field of data science and machine learning wherein you need to have historical data to build advanced machine learning algorithms for accurate predictions. Not only for machine learning purposes but also for conducting business analysis and predictive analysis using the data, is an important part of any industry to grow and align the business and marketing strategy in the line of maximum growing business.

You can take a look at any industry, whether it is FMCG, e-commerce, stock exchange, banking, edtech or any government schemes, that may involve collecting demographic data, issuing subsidies and Aadhar cards, and listing people as per the poverty benchmark. In each of the industries, you need to deal with data in terms of collection, processing or creating strategy.

So, in the world of highly advanced and data-driven analysis and technologies, it is extremely important to store and manage the data, which is also called database management. A database management system (DBMS) is a software that is designed to store, retrieve, define and manage data in a database and works on the top of the database. The DBMS software primarily functions as an interface between the end-user and the database, simultaneously managing the data, the database engine and the database schema in order to facilitate the organisation and manipulation of data.

OLTP vs OLAP

You will encounter mainly two types of processes in any database management system, which are online transactional processing (OLTP) and online analytical processing (OLAP). Online transaction processing systems are ones that support online transactions in databases on a real-time basis. Its primary objective is to process data, not analyse it. On the other hand, online analytical processing systems are those categories of software that provide analytical platforms to derive useful insights for business purposes.

Basic introduction to RDBMS

What if data is interlinked? A relational database management system (RDBMS) refers to the capabilities that are designed to enable the user to create, update and administer a relational database, which is characterised by its structuring of data into logically independent tables in which each table is related to the other tables in the database. A DBMS is capable of managing small amounts of data, whereas RDBMS can manage unlimited data in a distributed fashion.

The basic element under an RDBMS is entity relationship diagrams (ERDs). You can understand the concept of ERDs using a simple example. Suppose in an organisation, two databases exist–one is the employee table and the second is the department table. It is understood that there will be multiple employees in each department. So, there will be one-to-many relationships from the department table to the employee table. In this way, there can be one-to-one, many-to-one and one-to-many relationships between tables.

Online retailer business: toy replica case study

The online retailing business is quite famous currently. Here, you do not need to take orders by going to each customer. In online retailing, a company launches its website to sell the items in wholesale, and buyers can order the products that they require from that website only. In this case study, you are introduced to a company that manufactures toy replicas or models.

Suppose the name of this company is **Toylica**.

Suppose Toylica has captured a huge market globally and supplies toy replicas in almost all the countries. Wholesale buyers order products in their required quantity from the website only. As the company captured a large customer base and enhanced its workforce, it is essential for this company to maintain the database to navigate and store the employee-, customer- and product-related information.

The brand has a separate department for analysing its sales and building a machine learning model to predict future sales. Each customer's order is saved into the database using the OLTP process when they place their order. A team of data analysts and machine learning engineers perform OLAP tasks on the historic data for creating marketing strategies to enhance the business.

Suppose you are a data analyst working at this company. Here, you are only focusing on the SQL queries to get an idea of the company's sales. Now, let's understand the process and structure of the database that the company is using. The three sections in the database of the company are as follows:

- Employee section
- Customer section
- Product section

You can download the details of the attributes for each table from the file attached on the platform below.

1. **Employee section:** This section is divided into the following tables:
   a. **employees**: This table contains information about a company's employees. Each entry in this table is identified uniquely by the 'employeeNumber', which is the primary key of this table.
   b. **offices**: This contains information about the company's offices that are located worldwide. Each office is uniquely identified by the 'officeCode', which is the primary key of this table. Each employee must have only one office against their ID, in which they have to work.

2. **Customer section:** This section contains the following tables:
   a. **customers**: This table contains information about customers who buy the models from the company in bulk. They are retailers or shopkeepers who own firms or shops; they are not the final consumers. Each customer is uniquely identified by the 'customerNumber', which is the primary key of this table. A particular customer is either assigned an employee or not assigned any. The assigned employee is identified by the 'salesRepEmployeeNumber' column in this table. Multiple customers can be assigned to an employee.
   b. **orders**: This table contains information about the orders placed by the customers. Each order is uniquely identified by its 'orderNumber', which is the primary key of this table. Multiple orders can be placed by a particular customer, but each order should have a unique 'orderNumber'.

c. **orderdetails**: This table also contains details about the orders placed by the customers. However, both the 'orderNumber' and 'productCode' are part of the composite primary key here.

d. **payments**: This table contains information about the payments made by the customers after placing their orders. Here, you will notice a column named 'checkNumber', which refers to the unique number of the check through which the customer ('customerNumber') has made the payment. Under this table, 'checkNumber' is the primary key.

3. **Product section:** This section has the following tables:

a. **products**: The 'products' table contains information about the different products that are sold by the company. Each product is uniquely identified by the 'productCode', which is the primary key of this table.

b. **productlines**: This table is a broad categorisation of the type of toy models that are sold by the company, including vintage cars, motorcycles, trains and ships. Each product in the 'products' table must have a single product line.

Finally, you will have the following eight tables:

1. employees
2. offices
3. customers
4. orders
5. orderdetails
6. payments
7. products
8. productlines

Perform OLAP on online retailer business case study

You need to prepare the entity relational diagram (ERD) for this case study after understanding the descriptions of the above-mentioned tables.

You are provided with the data set. So, you do not need to create any table in the database. You are mainly required to write SQL queries to answer the questions.

You will find a blank MySQL workbench with inline comments attached to perform analysis on the data and to guide you to proceed in the case study.

You must code each step that has been mentioned in the commented workbench. The questions are related to each other; so, if you do not follow the steps diligently, then you will not be able to answer some of the questions that are given below.

You are provided with the MCQs that have ONLY one correct option and you need to follow all the instructions that have been given in the commented MySQL workbench.

Note: You need to run each step either by yourself or using the answer feedback that is given in the solution document, otherwise you will not be able to proceed in the further case study and won't be able to answer some of the questions.

Question 1

The first step of every RDBMS activity is to create a star schema for a particular data set. You have a total of eight tables as described earlier, and it is essential for you to create an entity relational diagram (ERD) in order to relate the various tables. Which of the following is the most appropriate relation among 'employees', 'customers', 'orders' and 'offices' based on the problem statement?

A. The 'customers' table has only a one-to-many/zero relation with the 'orders' and 'employees' tables. Also, the 'offices' table has only a one-to-many/zero relation with the 'employees' table.

B. The 'customers' table has only a one-to-many/zero relation with the 'orders' table. The 'employees' table has one/zero to many/zero relation with the 'customers' table. Also, the 'offices' table has only a one-to-many/zero relation with the 'employees' table.

C. The 'orders' table has only a one-to-many/zero relation with the 'customers' table. The 'employees' table has one/zero to many/zero relation with the 'customers' table. Also, the 'employees' table has only a one-to-many/zero relation with the 'offices' table.

D. The 'orders' table has only a one-to-many/zero relation with the 'customers' table. The 'employees' table has one/zero to many/zero relation with the 'customers' table. Also, the 'offices' table has only a one-to-many/zero relation with the 'employees' table.

Question 2

Which of the following table names has a self-referential relationship based on the descriptions of columns of each table given in the problem statement?

A. customers
B. orders
C. employees
D. Offices

Question 3

As you can see in the figure below, a one/zero-to-many/zero relation exists between the 'employees' and 'customers' tables. Identify the foreign key in the 'customers' table that relates this table with the 'employees' table.
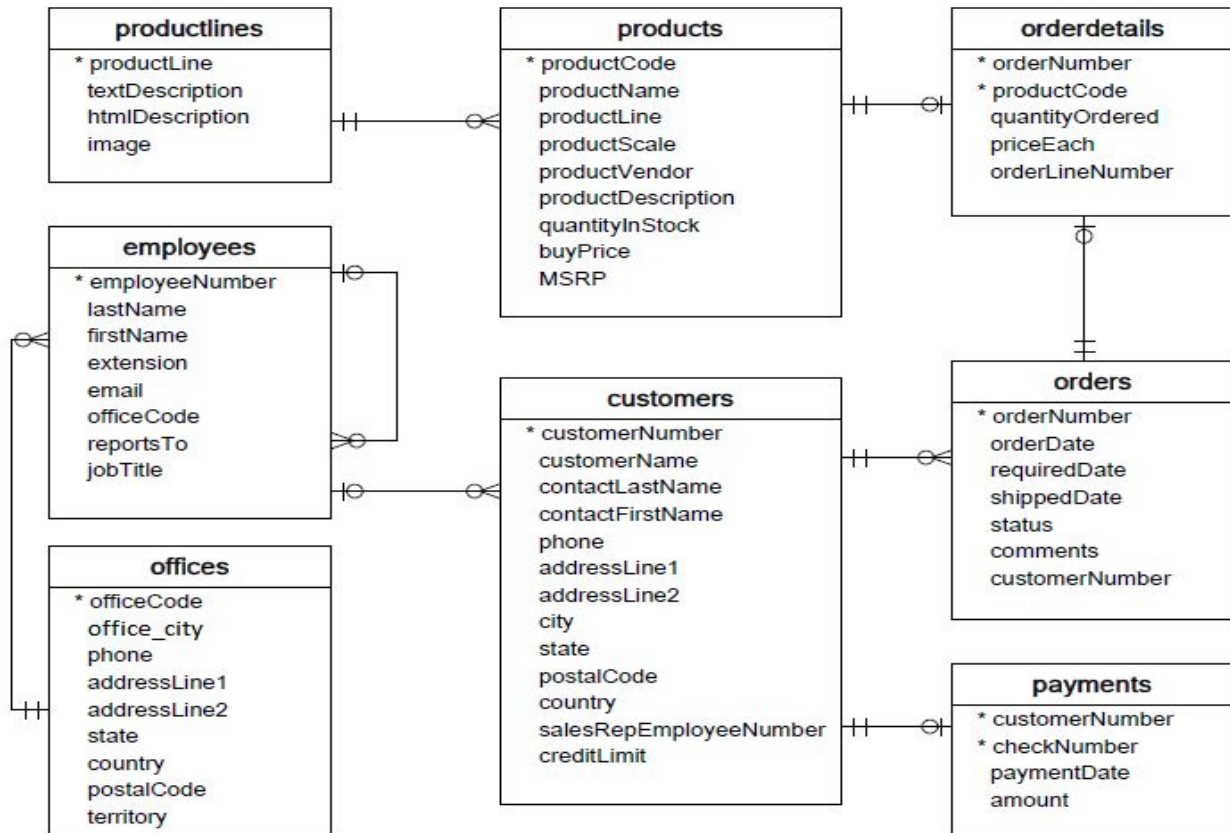
A. contactLastName

B. contactFirstname

C. country

D. salesRepEmployeeNumber

Question 4

Match the following actions with the appropriate SQL commands presented in the table given below.

a	Find all the employees from the 'employees' table	I	ALTER
b	Add information about a new employee	II	DROP
c	Change the structure of the employee table	III	SELECT
d	Remove the 'employees' table	IV	INSERT

    A. a-I, b-II, c-III, d-IV

    B. a-IV, b-III, c-II, d-I

    C. a-III, b-I, c-II, d-IV

    D. a-III, b-IV, c-I, d-II

Question 5

Suppose you want to create a table named 'payments' in SQL with the two primary keys 'customerNumber' and 'checkNumber', as mentioned in the ERD. Which of the following queries is the correct one to achieve this?

    A. CREATE TABLE `payments` (

        `customerNumber` int(11) NOT NULL,

        `checkNumber` varchar(12) NOT NULL,

        `paymentDate` datetime DEFAULT NULL,

        `amount` decimal(12,2) DEFAULT NULL,

        PRIMARY KEY (`customerNumber`,`checkNumber`)

        CONSTRAINT `Check` FOREIGN KEY (`customerNumber`) REFERENCES `customers` (`customerNumber`)

)

    B. CREATE TABLE `payments` (

`customerNumber` int(11) NOT NULL,

`checkNumber` varchar(12) NOT NULL,

`paymentDate` datetime DEFAULT NULL,

`amount` decimal(12,2) DEFAULT NULL,

PRIMARY KEY (`customerNumber'),

PRIMARY KEY ('checkNumber`)

)

C.   CREATE TABLE `payments` (

`customerNumber` int(11) NOT NULL,

`checkNumber` varchar(12) NOT NULL,

`paymentDate` datetime DEFAULT NULL,

`amount` decimal(12,2) DEFAULT NULL,

PRIMARY KEY (`checkNumber),

CONSTRAINT `Check` FOREIGN KEY (`customerNumber`) REFERENCES `customers` (`customerNumber`)

)

D.   CREATE TABLE `payments` (

`customerNumber` int(11) NOT NULL,

`checkNumber` varchar(12) NOT NULL,

`paymentDate` datetime DEFAULT NULL,

`amount` decimal(12,2) DEFAULT NULL,

PRIMARY KEY (`checkNumber),

)

Question 6

You need to enter the following entries in the 'customers' table to proceed in this assignment.

(495,'Diecast Collectables','Franco','Valarie','Boston','MA','51003','USA','1188',85100),

(496,'Kelly\'s Gift Shop','Snowden','Tony','Auckland','NULL','NULL','New Zealand','1612',110000)

Which of the following commands can be used to enter the two rows that are given above in the 'customers' table?

A. INSERT
B. UPDATE
C. ALTER
D. CREATE

Question 7

Suppose you need to change the 'employees' table by inserting the job title 'Sales Rep' in which the office code is equal to 4. Which of the following queries will you use to perform this operation?

A. modify employees

set jobTitle = 'Sales Rep'

where officeCode = 4;

B. update employees

set jobTitle = 'Sales Rep'

where officeCode = 4;

C. update employees

set jobTitle = 'Sales Rep'

where officeCode = '4';

D. modify employees

set jobTitle = 'Sales Rep'

where officeCode = '4';

Question 8

Delete the entries where the 'productLine' column has a value that equals to 'Boats' from the 'productlines' table. What is the number of rows in the 'productlines' table?

A. 8

B. 7

C. 6

D. 5

Question 9

Suppose you want to convert the data type of the 'quantityOrdered' column of the 'orderdetails' table from varchar into an integer. Which of the following code will you use to convert the varchar data type into an integer?

    A. alter table orderdetails

        change quantityOrdered int;

    B. alter table orderdetails modify column quantityOrdered int;

    C. alter table orderdetails

        change quantityOrdered from varchar to int;

    D. alter table orderdetails modify column quantityOrdered from varchar to int;

Question 10

You have tables named 'employees' and 'customers'. Code in the MySQL workbench that you have been provided and answer the following questions:

- Suppose you print the names of employees with the job title 'Sales Rep'. What is the first name of the employee that appears first in this query?

- What is the total number of employees from the 'employees' table? Alias it as 'Total_Employees'.

- How many customers are based in Australia? Alias it as 'Australia_Customers'.

    A. Leslie, 5, 23

    B. Jennings, 23, 5

    C. Leslie, 23, 5

    D. Jennings, 23, 5

Question 11

You have tables named 'products' and 'orders'. Code in the MySQL workbench that you have been provided and answer the following questions:

- Using the 'products' table, calculate the sum of quantity that is in stock for the product vendor 'Red Start Diecast' and the product line 'Vintage Cars'.
- What is the total number of orders that have not been shipped yet from the 'orders' table?

A. 22381, 303

B. 8073, 23

C. 22381, 23

D. 8073,303

Question 12

How many entries does the 'orderdetails' table contain for which the 'productCode' starts with S18 and the 'priceEach' is greater than 150?

A. 23

B. 24

C. 25

D. 26

Question 13

Which of the following are the top three countries (in order) that have the maximum number of customers?

A. Spain, Germany, France

B. USA, Germany, France

C. Germany, France, USA

D. Germany, France, Spain

Question 14

What is the average credit limit for Singapore from the 'customers' table?

    A. 67233

    B. 58600

    C. 58500

    D. 51700

Question 15

What is the total amount to be paid by the customer named 'Euro+ Shopping Channel'? You need to refer to the 'customers' and 'payments' tables to answer this question.

    A. 715738

    B. 715749

    C. 725750

    D. 725738

Question 16

Which month received the maximum aggregate payments from the customers along with the aggregated value of the payment of that month?

    A. November, 1551479

    B. December, 1645923

    C. January, 397887

    D. February, 503357

Question 17

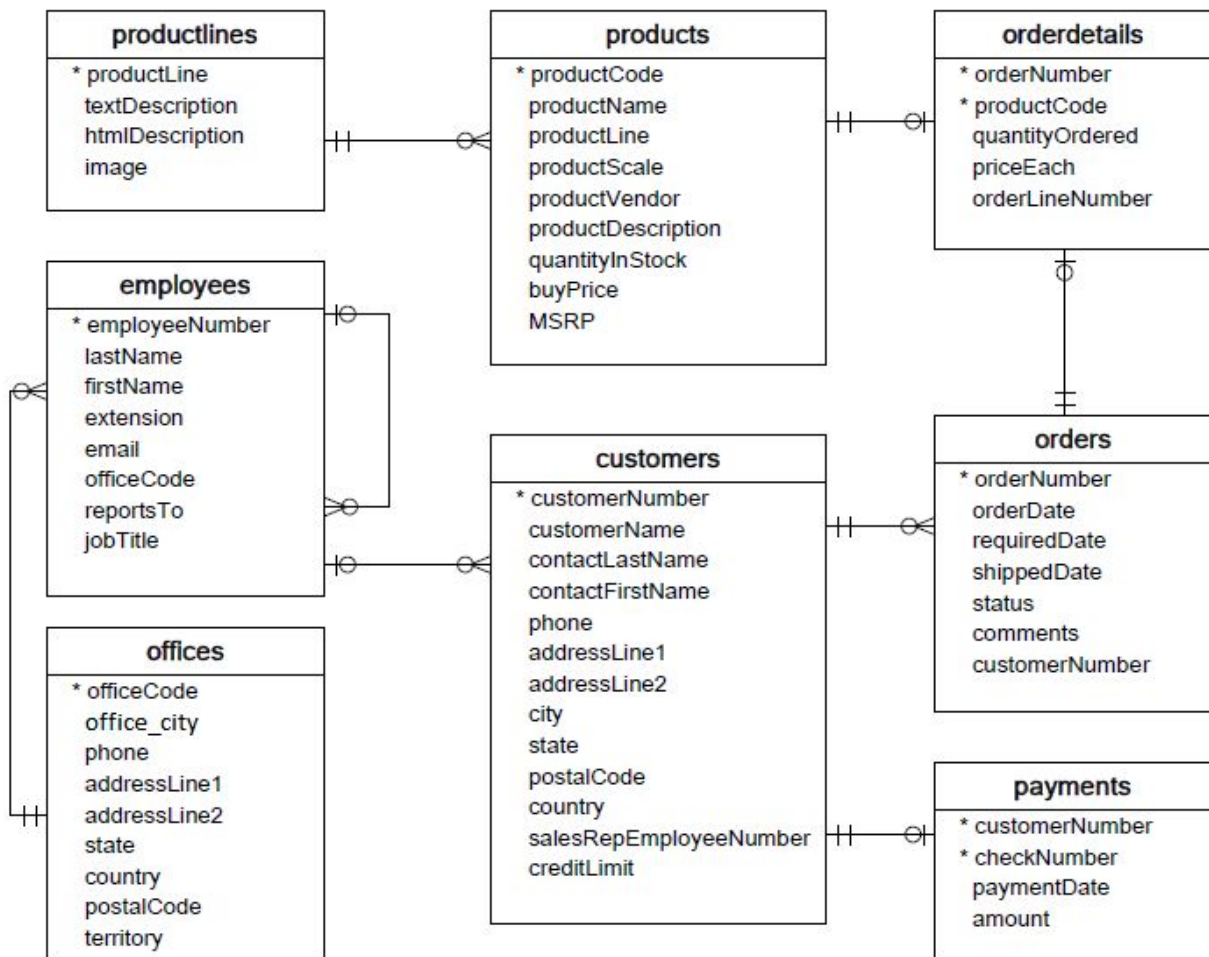What is the shipped date of the maximum quantity that was ordered for the product name '1968 Ford Mustang'?

Hint: Apply a nested query on the 'products', 'orderdetails' and 'orders' tables.

    A. 07-12-2003

B. 12-07-2003

C. 09-09-2003

D. 16-10-2003

Question 18

Multiple customers can be contacted by a single employee, as you can see in the schema given below. You need to use the 'inner join' clause on the 'employees' and 'customers' tables.

productlines	products	orderdetails
* productLine textDescription htmlDescription image	* productCode productName productLine productScale productVendor productDescription quantityInStock buyPrice MSRP	* orderNumber * productCode quantityOrdered priceEach orderLineNumber

employees
* employeeNumber lastName firstName extension email officeCode reportsTo jobTitle

customers
* customerNumber customerName contactLastName contactFirstName phone addressLine1 addressLine2 city state postalCode country salesRepEmployeeNumber creditLimit

orders
* orderNumber orderDate requiredDate shippedDate status comments customerNumber

offices
* officeCode office_city phone addressLine1 addressLine2 state country postalCode territory

payments
* customerNumber * checkNumber paymentDate amount

What is the average value of the credit limit that corresponds to the customers who have been contacted by the employees from the 'Tokyo' office after using the 'inner join' clause on these two tables? Hint: You need to apply the nested query as well.

A. 80900

B. 81900

C. 82900

D. 83900

Question 19

How will you find the name of the customer who paid the lowest amount for the product? You need to use the 'outer join' clause on the 'customers' and 'payments' tables and the 'sum' clause on the attribute 'amount' to answer this question.

A. Boards & Toy Co.

B. Auto-Moto Classics Inc.

C. Euro + Shopping Channel

D. Mini Gift Distributors Ltd.

Question 20

In which office does the employee with the job title 'VP Marketing' work?

A. Boston

B. Tokyo

C. San Francisco

D. Sydney

Question 21

What is the name of the customer who belongs to 'France' and has the maximum creditLimit among the customers in France?

A. Muscle Machine Inc

B. Handji Gifts & Co

C. Mini Wheels Co.

D. Mini Gifts Distributors Ltd.

Question 22

What will be the remaining stock of the product code that equals 'S18_1589' if it is sent to all the customers who have demanded it?

A. 8692

B. 9042

C. 350

D. None of the above

Question 23

What is the average amount paid by the customer 'Mini Gifts Distributors Ltd.'?

A. 63212

B. 64910

C. 65000

D. 65910