

- d) This refers to VFX (Visual Effects), and it indicates the technique of synthesizing a real recorded subject matter by using computers. SFX (Special Effects) is similar or synonymous to this, but it refers to synthesis technique that does not use computers, and in other words, it is “special photographing.” In some cases, no distinction is made between VFX and SFX.

Q3-12 a)

Computer graphics

When a shape in computer graphics is drawn, technology used for making the step-like jagged appearance near edges less noticeable is anti-aliasing. An image is displayed by changing the color of pixels constituting the image. However, pixels can only display the color corresponding to the assigned numeric value, and therefore the number of colors that can be simultaneously displayed is only one at maximum. Therefore, a boundary of two colors has a step-like jagged appearance, and it is necessary to assign intermediate color to make it less noticeable. This process is called anti-aliasing. Therefore, a) is the correct answer.

- b) Clipping refers to specifying a portion of an image for restricting the processing range.
- c) Shading refers to assigning shade for giving a three-dimensional appearance.
- d) Morphing is creating an intermediate image from images before and after the change so that the image changes smoothly.

Q3-13 d)

Explanation of relational database

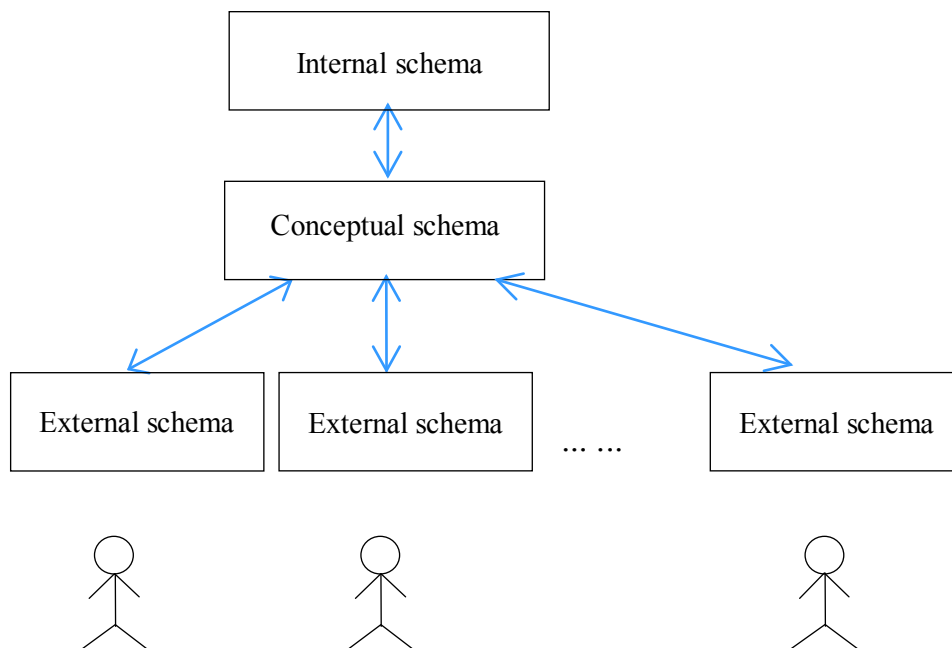
Database models are broadly divided into a relational model and structured model, and the structured model is further classified into a hierarchical model and network model. While the parent-child relationship in a hierarchical model is “1:n”, the parent-child relationship in a network model is “n:m” and it is also called the CODASYL model. Relational model represents the collection of data in several two-dimensional tables, and relation is represented by associating key items between these tables. Therefore, d) is the appropriate explanation.

- a) A hierarchical model database shows relation between data and the hierarchical structure.
- b) A list links the related data by using pointers, and rather than database, it is one of the methods of showing linked structures of data.
- c) An object oriented database (OODB) is formed by combining data and its operation procedure in the form of a database.

Q3-14 b)

Three-schema architecture of database

The three-layer schema architecture is composed of an external schema (a group of data items describing the required logical structure as seen from the viewpoint of an individual user), a conceptual schema (a group of data items describing the logical relation to be originally contained in the data), and an internal schema (a group of data items written in view of hardware, performance, recovery, and security). The relation between these three is shown in the figure below.



Logical data independence is achieved by separating the conceptual schema (which shows the logical relation of data) from the external schema (which shows the view desired by users). In specific terms, even when there is a change in the logical relation of data (addition or deletion of tables or data items), if this change is not related to the view of data desired by users, the schema is not affected by this change. Therefore, b) is the appropriate description.

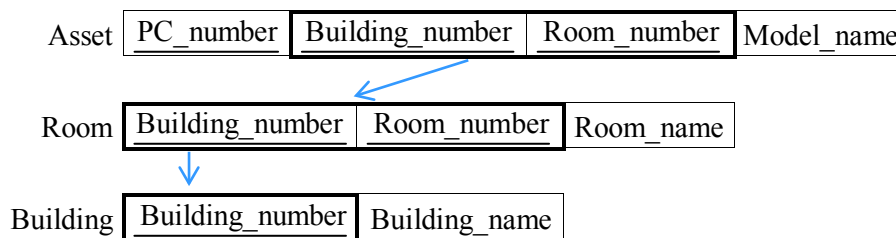
- a) As mentioned above, the three-layer schema architecture refers to external schema, conceptual schema, and internal schema.
- c) (Incorrect) Internal schema → (Correct) Conceptual schema
- d) (Incorrect) External schema → (Correct) Internal schema

Q3-15 b)

Sequence of data entry in table

This question is about referential constraints in a database. Referential constraint refers to the constraint that “when a value in a column of some table refers to another table, a column in the referenced table must also have that value.” In a relational database, the foreign key (an item that is used for association with another table, and that is the primary key of the referenced table) must maintain referential constraint with the primary key of the referenced table.

Here, first it is necessary to understand from the data items of a table how referential constraints are set. In the “Asset” table, PC_number, Building_number, and Room_number are the primary keys. It is required to check whether or not the combination of Building_number and Room_number actually exists in the “Room” table. Moreover, it is also required to check whether or not Building_number in the “Room” table exists in the “Building” table. Summarizing this gives the result shown below.



As the referenced data must be entered first, sequence of data entry in the tables is Building → Room → Assets. Therefore, b) is the appropriate answer.

Q3-16 c)

Primary key of a relational database

The field required for identifying each row is called the primary key, and in the same table no two rows have the same primary key. Therefore, c) is the appropriate answer. In other words, columns that have rows with the same primary value cannot be a primary key.

- Rows can be searched even if search conditions are specified for items other than the column identified as a primary key.
- Arithmetic operations can be performed on any column irrespective of whether it is a primary key or not.
- Multiple columns can be specified as a primary key. In this case, a primary key is also called a composite key.

Q3-17 c)

Third normal form

The outline of procedures until achieving the third normal form is given below.

- First normal form: Separate the repeated parts.
- Second normal form: Separate the attributes that are functionally dependent on the entire composite key (full functional dependency), and that are functionally dependent on only a part of the composite key.
- Third normal form: Other than key data, search the data that can become a key, and separate it.

Applying the tasks involved in this procedure to the answer choices, it is clear that the table in c) is in the third normal form.

- a) Employee_name is functionally dependent on Employee_number only, while Skill_name is functionally dependent on Skill_code only. Therefore, there is no relation between employee and skill.
- b) Employee_name is functionally dependent on Employee_number only, and just like a), conditions of the second normal form are not satisfied. In this situation, when one employee has multiple skills, the same combination of Employee_number and Employee_name is shown redundantly.
- d) There is only one item of Skill_years_of_experience for each employee, and years of experience for each skill cannot be recorded.

Q3-18 d)

Identifying the primary key from functional dependency

A primary key of a relation is a minimum set of attributes with which a tuple of a relation (called a row in SQL) can be uniquely identified. Minimum means that even if one attribute is removed from that set of attributes, it doesn't fulfill the role of a primary key. In other words, it indicates the minimal requirement.

- a) From the details of functional dependency of (1) through (7), it is clear that product number, product name, or quantity cannot be decided from order number only. Therefore, (Order_number) is not the primary key.
- b) Just like a), product number, product name, or quantity cannot be decided. Moreover, in the functional attribute (2), there is $\text{Order_number} \rightarrow \text{Customer_number}$. Therefore, Customer_number can be decided by Order_number, so it does not make sense to add Customer_number (excess or redundant). Although (Order_number, Customer_number) can uniquely identify a tuple, this combination is not minimum. It is clear that it is not the primary key.
- c) With these three attributes, all other attributes can be decided. However, just like b), as it includes Customer_number, it is not minimum, and it cannot be the primary key.
- d) According to the functional dependencies of (1) through (7), all attributes can be decided with (Order_number, Product_number) (identification and equivalent of tuple), and it is minimum. Therefore, it matches the definition of primary key, and d) is the correct answer. Moreover, by comparing with c), it should be kept in mind that the required quality of a primary key is "minimum."

Q3-19 a)

Operation of a relational database

In the operation of a relational database, retrieving the respective rows that satisfy the specified conditions from multiples tables, and then creating a new table (derived table) is called “Join”. Therefore, a) is appropriate.

- b) Project – Retrieving columns (not rows) that meet certain conditions from a table.
- c) Select – Retrieving rows (not columns) that meet certain conditions from a table.
- d) Insert – Inserting specific rows (not columns) in a table.

Q3-20 a)

Data manipulation not subject to referential constraints

Referential constraints are constraints concerning the integrity of reference relationship between tables in a relational database. In a relational database, for referring to an item in another table by using a foreign key, the referring item must exist in the referenced table.

In the three tables provided in the question, items referring to other tables as a foreign key (items indicated with a dotted underline) are Product_code and Customer_code in the Order table, and they refer to the Product table and the Customer table respectively.

- a) The “Customer” table is the referenced table, and new records can be added to this table without any constraints.
- b) The “Product” table is the referenced table, and when the records whose product codes exist in the “Order” table are deleted, constraints are applied.
- c) The “Product” table is the referenced table, and when product codes to be changed exist in the “Order” table, constraints are applied.
- d) As Product_code and Customer_code of the record to be newly added to the “Order” table must already be registered in the “Product” table and the “Customer” table respectively, constraints are applied.

Therefore, a) is the correct answer.

Q3-21 c)

Information that cannot be obtained from data

As one buyer may purchase multiple products, from “Daily sales data of sales agents,” “Daily variation in the number of buyers for each sales agent” cannot be obtained. Moreover, as the Date field is not present in “Product data bought by customers,” “Daily variation in the number of buyers for each sales agent” cannot be obtained. As both “Daily sales data of sales agents” and “Product data bought by customers” have Sales_agenc as a common field, it is possible to join the records in these tables where the Sales_agenc field has the same value. However, even after the daily sales data and product data bought by customers are joined, it is not possible to determine the daily number of buyers. Therefore, c) is the correct answer.

- a) “Daily variation in sales quantity for each product” can be obtained from Date, Product, and Sales_quantity in “Daily sales data of sales agents.”

- b) A single record can be created by joining the records in “Customer data” and “Product data bought by customers” that have the same value in the Customer field. Therefore, “Hot-selling products by gender” can be obtained from Gender, Product, and Sales_quantity.
- d) “Age distribution of the buyers for each sales agent” can be obtained from Customer, Date_of_birth, and Sales_agent.

Q3-22 b)

Description concerning SQL statements

Verifying the results of SQL statements a) through d) including set functions provided in the questions gives the following outcome.

- a) **AVG** is a function for calculating a mean value. Data in the first and the third record is applicable for Product_number NP200 specified in the selection condition, and calculating the mean value for this quantity gives $(3+1) \div 2 = 2$.
- b) **COUNT** is a function for calculating the number of applicable data records. As no specific selection condition is specified, all records are counted. Therefore, the outcome is 4.
- c) **MAX** is a function for calculating the maximum value. Calculating the maximum value of quantity gives 3.
- d) **SUM** is a function for calculating the total value. Data in the third and the fourth record is applicable for date 2006-10-11 specified in the selection condition. Calculating the total quantity for these records gives $1+2=3$.

Based on the above results, the maximum value is obtained in b).

Q3-23 b)

Correct SQL statement

When there is a **GROUP BY** clause in the inquiry, the selection list of the **SELECT** clause must be a column (arithmetic expression including that column) specified in the **GROUP BY** clause, or it must be a constant number or a set function. In other words, if there is a **GROUP BY** clause, only values representing a group can be specified in the **SELECT** clause. b) satisfies this condition, and it is the correct answer.

- a) Although set function is specified in the selection list of the **SELECT** clause, **GROUP BY** clause is not included in this inquiry. Therefore, this can be considered as one group. Set function of a value representing one (1) group is allowed, however, Order_date is not a value (one in this case) representing a group, and hence it cannot be allowed.
- c) There is an error here as the set function is nested.
- d) Since a set function is included in comparison predicate of the **WHERE** clause, there is an error. For specifying a set function, use the **HAVING** clause.

Q3-24 a)

Selection of SQL statement

As this “SQL statement is used to search for Department_code of departments that have less than five employees with the job duty of Programmer,” the retrieval conditions are as follows:

- (1) Employees with the job duty of programmer
- (2) Departments that have less than five employees falling under (1)

Normally, an SQL statement that joins the “Department” table and the “Employee” table by using Department_code can be considered, and then employees with a Job_duty of Programmer can be retrieved. However, only the “Employee” table is provided in the question statement. Therefore, as given in answer choices, let the “Employee” tables be S1 and S2, and S2 be the correlated subquery to be referred for one row of S1.

Here, “departments that have less than five employees with the job duty of Programmer” means that not only those departments that have one to four employees as programmer, departments with no employees as a programmer should also be considered.

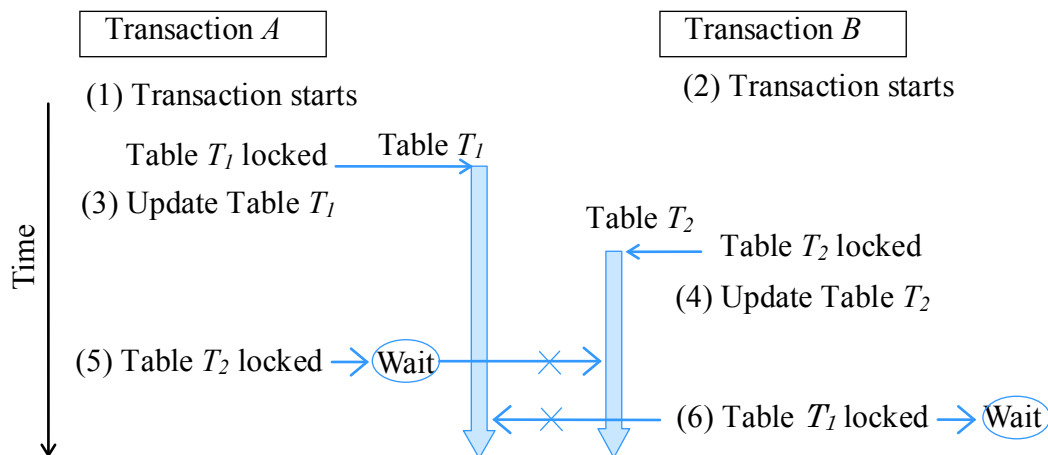
- a) S2 is referred to for one row of S1, and the number of rows of Employee_number that have the same Department_code and Job_duty of Programmer are tabulated with the COUNT function. As a result, the Department_code of rows that have less than five (including zero (0)) as an outcome is retrieved from S1. However, as a result of referring to S2, all rows that have the same Department_code are retrieved, and therefore the overlapping is removed with DISTINCT. Therefore, Department_code of departments that have less than five employees that have Job_duty of Programmer are searched for, and Therefore a) is the appropriate answer.
- b) With the same method as a), the number of rows where Job_duty is not Programmer is tabulated, and they are retrieved when the number of rows is over five. Having six or more employees who are not a programmer does not necessarily mean that there are less than five programmers.
- c) According to the conditional expressions in the GROUP BY clause and the HAVING COUNT clause, the table S1 is not narrowed down by job duty, and departments that have five or less and even one programmer are searched for under those conditions.
- d) Because the IN clause is used, S2 is referred to for one row of S1, and Department_code where Job_duty is programmer is retrieved, and in that if Department_code is the same as S1 then Department_code of S1 is retrieved. Therefore, departments that have no programmer are not searched for.

Q3-25 d)

Deadlock when a table is updated

Deadlock means when multiple transactions attempt to lock a shared resource, they mutually wait for the unlocking (releasing) of the resource locked by the other party, and they end up perpetually waiting. Key points in this question are: locking is done for each table just before update, and unlocking is done after the transaction is completed.

After the transactions start, Transaction *A* and Transaction *B* lock Table T_1 and Table T_2 respectively, and performs the update process ((3) and (4)). However, unlocking is done after the transaction is completed, and therefore the process continues with tables remaining locked. Prior to process (5), Transaction *A* attempts to lock Table T_2 . However, it is not able to lock as Table T_2 is being used by Transaction *B*, and Transaction *A* gets into the waiting mode (At this stage, Transaction *A* is in waiting mode, but deadlock hasn't occurred). Next, prior to process (6), Transaction *B* attempts to lock Table T_1 . However, it is not able to lock as Table T_1 is being used by Transaction *A*, and Transaction *B* gets into the waiting mode. At this stage (6), both Transaction *A* and Transaction *B* are in mutually waiting mode for resources, and deadlock occurs. Therefore, d) is the correct answer.



Q3-26 a)

Exclusive control of database

When a database record is being used by a transaction, exclusive control of a database is a function that restricts access to this database record by another transaction. In this exclusive control (lock), there is a “shared lock” and an “exclusive lock.” For example, when a record is used for reference purposes, it causes inconvenience if another transaction updates this record while it is being referred to. However, it does not cause any inconvenience even if it is simultaneously referenced by another transaction. In this case, a shared lock is used. In other words, although reference by multiple transactions at a time is permitted, update is not allowed. On the other hand, when a record is used for update purposes, it causes inconvenience if another transaction refers to this record during the process of update. An exclusive lock is used in such case. In an exclusive lock, the corresponding record is exclusively occupied, and even reference to the record is not allowed. Therefore, a) is the appropriate answer. Although shared locks may coexist, other combinations cannot coexist.

Q3-27 d)

Files for database recovery

A backup file is a file where the entire database of a certain point in time is saved. In the case of a media failure, once the media is recovered, by loading the contents of the latest backup file, it is possible to return to the point where the backup was taken. However, it does not return to the state when the failure occurred (down point). Log file (journal file) contains update history (information before update and information after update) for every data item, and therefore the information is recorded in chronological order. With this, the recovered database is updated to the point right before the down point. At that time, information after update is used. This process is called a rollforward process. Therefore, d) is the correct answer.

- a) Transaction file: This file contains update information (daily transaction information) of the master file.
- b) Master file: This file is maintained as a base file. Like a journal, this file contains the most basic information of operations.
- c) Rollback file: This file contains information before update for rollback. However, generally it is included in d) log file.

Q3-28 d)

Techniques of recovering transactions

For a database recovery process, a log is stored in the database management system. In this log, starting point of transaction, information before the data is updated, and information after update are recorded. Moreover, by setting the checkpoint and by generating the log, a transaction identifier (information for identifying the transaction) under execution is generated, and then changes made up to that point in time are written to the database. As the checkpoint is obtained when Transaction T is running, changes made up to obtaining the checkpoint are written on the database. Therefore, in order to restore the database to the point of right after the completion of Transaction T, it is necessary to re-run the process from the checkpoint recorded in the redo log until the completion of transaction, and then write the result to the database. Therefore, d) is the correct answer.

- a) A two phase locking (2PL) is divided into two phases, namely a phase where locks are acquired (growing phase) and a phase where locks are removed (shrinking phase). In the case that a transaction needs multiple locks, instead of locking and unlocking each time to maintain consistency, after the lock is applied, it is necessary to keep the lock until the writing is completed.
- b) Transaction scheduling decides the execution sequence of transactions and lock granularity so that deadlock or inconsistency of update does not occur. Two phase locking in (a) is also a type of transaction scheduling.
- c) Roll back is a recovery technique for the restoration of the database to the state prior to the transaction by using the undo log.

Q3-29 b)

Data mining

Data mining is a technique of uncovering semantic information such as underlying rules and relations in a large amount of historical data. For this, advanced mathematical techniques such as artificial intelligence and statistics are used, and it attempts to predict the future based on past data. As supporting tools for this, software programs equipped with a fast search function through a large amount of data and automatic rule detection algorithms are developed. Therefore, b) is the appropriate description.

- a) Although data mining is used as one of the fast search methods, it is not limited only to that.
- c) This describes data warehouse.
- d) This describes data mart which is a part (or subset) of data warehouse and is rearranged for each department.

Q3-30 a)

Transparency of a distributed database

Distributed database refers to distributing and keeping data in multiple locations, and it functions by mutual cooperation over communication lines. An application program that uses this database can use the entire distributed database as a single database without realizing the physical location of the data. Moreover, transparency of a distributed database means that data can be handled without knowing the storage location of data or other circumstances specific to the distributed database.

Databases reside on “multiple” servers and can be accessed “as if they are operating on one server.” This means that a) is the appropriate choice.

- b) An application program “needs not know” which server’s database should be accessed.
- c), d) A distributed database cannot exist “on one server.”

Q3-31 b)

Communication load between client and database server

An effective method for solving the problem of the communication load of SQL statements is to use a stored procedure function that involves storing typical processes of frequently used SQL statements in the database management system beforehand. Communication load between client and server can be reduced with this function. Therefore, b) is the correct answer.

- a) Redefinition of the index: Although it streamlines the database search, it does not result in the reduction of communication load of the SQL statements.
- c) Reorganization of the database: This is used when repetitive addition, deletion, and update of data lower the storage efficiency of a database. It is not a solution for reducing the communication load.
- d) Use of dynamic SQL: Dynamic SQL is one of the methods of embedded SQL. As SQL statements are still sent in this method, it is not an appropriate solution for reducing the communication load.

Q3-32 a)

Data buffering

As encoding speed and communication speed are represented in bits while size of audio data is represented in bytes, the conversion of everything in units of bytes gives the encoding speed of 8 Kbps and a communication speed of 6 Kbps.

The time required for encoding 1.2 Mbytes of audio data is $1.2 \text{ Mbytes} \div 8 \text{ Kbps} = 150 \text{ seconds}$. On the other hand, the time required for downloading is $1.2 \text{ Mbytes} \div 6 \text{ Kbps} = 200 \text{ seconds}$. Therefore, it is clear that the downloading time is 50 seconds longer than the encoding time. For playing back uninterruptedly, it is necessary to simultaneously complete downloading and encoding. For simultaneously completing both of these, 50 seconds of data buffering is required before the playback is started.

Therefore, a) is the correct answer.