DATABASE CONCEPTS AND THE STRUCTURED QUERY LANGUAGE

Chapter-1
Relational Databases

Revision Notes

- Data, in the context of databases, refers to all the single items that are stored in a database, either individually or as a set.
- Database is a collection of interrelated data which is stored together to serve multiple applications.

  A database has the following properties :
  - It is a logically coherent collection of data with some inherent meaning. A random assortment of data can't correctly be called as database.
  - A database is a design, build properly with data for a specific purpose.
  - A database can be of any size and complexity.

Database Management System (DBMS)

- It is a collection of programs that enables users to create and maintain a database.
- The DBMS is a general purpose system that facilitates the process of defining, constructing, manipulating and sharing database among various users & application.

  e.g. : ORACLE, MS-ACCESS, FOXPRO, SQL, etc.
- Defining a database involves specifying a data type, structure and constraint of the data to be stored in the database.
- Sharing a database allows multiple users and programs to access the database simultaneously.
- The goal of a DBMS is to provide an environment that is both convenient and efficient to use in
  - retrieving Information from the database.
  - Storing information into the database.

  Databases are usually designed to manage large bodies of information. This involves:
  - definition of structures for information storage (data modelling).
  - provision of mechanisms for the manipulation of information (file and systems structure, query processing).
  - providing for the safety of information in the database (crash recovery and security).
  - concurrency control if the system is shared by users.

Components of Database Systems

- **Components of Database Systems**

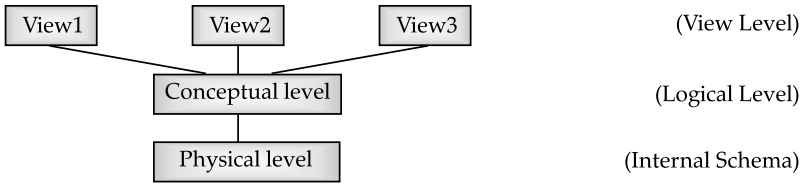
  (i) Data (ii) Software

  (iii) Hardware (iv) Users

  Data Abstraction

  Main purpose of a database system is to provide users with an abstract view of the system. The system hides certain details of how data is stored and created and maintained. All complexities are hidden from database users.

  Levels of Data abstraction : There are three levels of abstraction



1. **Physical Level**
   - How the data are stored :
     e.g. index, B-tree, hashing
   - Lowest level of abstraction :
     e.g. Data compression and encryption techniques
   - Complex low level structures
2. **Conceptual Level**
   - Next highest level of abstraction
   - Describes what data are stored.
   - Describes the relationships among data
   - Database administrator level
3. **View Level**
   - Highest level
   - Describes part of the database for a particular group of users.
   - Can be many different views of a database.
     e.g., tellers in a bank get a view of customer accounts, but not of payroll data.

Data Model

- Data model is a collection of conceptual tools for describing data, data relationship, data semantics and consistency constraints.

There are three different data models as:

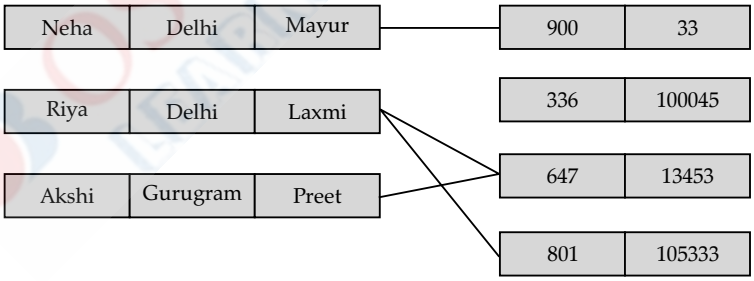
(i) Relational model (ii) Network model (iii) Hierarchical model

(i) Relational model : Data and relationship are represented by a collection of tables. Each table has multiple columns with unique names. *e.g.* customer, account

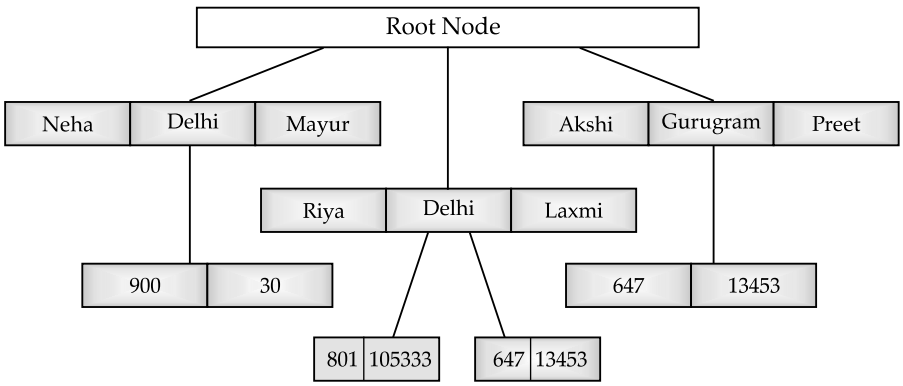
Customer			
Name	**City**	**Street**	**Number**
Neha	Delhi	Mayur	900
Riya	Delhi	Laxmi	647
Riya	Delhi	Laxmi	801
Akshi	Gurugram	Preet	647

Account	
Number	**Balance**
900	33
336	100045
647	13453
801	105333

(ii) Network Model : Data are represented by collections of records. Relationships among data are represented by links. Organization is that of an arbitrary graph.



(iii) Hierarchical Model : It is similar to the network model. Organisation of the records is as a collection of trees rather than arbitrary graphs.



Relational Model

- The relational model represents the database as a collection of relations. A relation is nothing but a table of values. Every row in the table represents a collection of related data values. These rows in the table denote a real-world entity or relationship.
- The table name and column name are helpful to interpret the meaning of values in each row. The data are represented as a set of relations. In the relational model, data are stored as table, however, the physical storage of the data is independent of the way the data are logically organized.

Relational Model Terminology :

- **Attribute :** In a Database Management System (DBMS) an attribute refers to a database component such as table columns. It also may refer to a database field. Attribute describes the instance in the row of database. ***E.g.*** Student, Roll no., Name, etc.
- **Relation :** It is sometimes used to refer to a table in a relational database but is more commonly used to describe the relationships that can be created between those tables in a relational database. Relations have three important properties a name, cardinality and degree. These are described as :

  Name : The first property of a relation is its name which is represented by the title or entity identifier.

  Cardinality : It refers to the number of rows (tuples) in relation that defines the uniqueness of data value contained in a column.

  Degree : It refers to the number of column (attributes) in each tuple.
- **Domain :** It is defined as the set of all unique values permitted for an attribute. For example; a domain of dates is the set of all possible valid dates, a domain of integer is all possible whole number, a domain of day of week is Monday, Tuesday ..... Sunday.
- **Tuples :** Rows of relations are generally termed as tuples.

Keys

- It is a data item that allows to uniquely identify individual occurrences of an entity type.
- An entity type usually has an attribute whose values are distinct for each individual entity in the entity set such a attributes is called key attribute.
- A key is normally correlated with one column in table and it might be associated with multiple tables.

(i) Primary Key

- The primary key of a relation can be said to be a minimal super key.
- The field or group of fields which forms the unique identifier for a table is called the table's primary key.
- The primary key uniquely identifies each record in the table and must never be the same for two records.

e.g. Emp-code can be primary key for entity set Employee. The primary key should be chosen such that its attributes are never or very rarely changed. For instance, the address field of a person should not be part of the primary key, since it is likely to change.

Emp-Code, on the other hand, is guaranteed to never change, till the employee is in the organisation.

(ii) Candidate Key

- There is only one primary key in a table. But there can be multiple candidate keys.
- A candidate key is an attribute or set of attributes that uniquely identifies a record.
- These attributes or combinations of attributes are called candidate keys.
- When there are more than one candidate keys, one of the candidate key is chosen to be a primary key. The remaining candidate keys are called alternate keys.

(iii) Foreign Key

- In a relation, column whose data values correspond to the values of a primary key column in another relation is called foreign key.
- In a relational database, the foreign key of a relation would be the primary key of another relation.

(iv) Super key

For an entity, it is a set of one or more attributes whose combined value uniquely identifies the entities in the entity set.

e.g. for an entity set Employees, the set of attributes (Emp-name, Address) can be considered to be a super key, if we assume that there are no two employees with the same name Emp-name and same address.

(v) Alternate key

If any one of the candidate keys among the different candidate keys available, is selected as primary key then remaining keys are called

❐❐❐

Chapter-2
Structured Query Language (SQL)

Revision Notes

- SQL provides commands for a variety of tasks including
  - **(i)** querying data.
  - **(ii)** inserting, updating and deleting rows in a table.
  - **(iii)** creating, altering, and dropping objects.
  - (iv) controlling access into the database and its objects.
- SQL (Structured Query Language) consists of commands, each command has an operational part and condition part. Operation is executed through a search in all relations defined in relational database.
- Required solution (list of data) is returned by the command. The data listed in solution satisfies all conditions given in that command.
- SQL is a comprehensive database language, it has statements for data definition, query and update. Hence, it is both DDL and DML.
- In addition, it has facilities for defining videos on the database for specifying security and authorization, for defining integrity constrains and for specifying transaction controls. It also has rules for embedding SQL statements into a general purpose programming language such as C or Pascal.
- Structured Query language is a query language that is used to access, create and operate relational databases.
- It was earlier known as sequel.
- It can be used by both casual users as well as skilled programmers.
- Types of SQL statements

  SQL statements can be divided into three major categories:
  (1) **Data Manipulation Language (DML):** These statements consist of queries that retrieve data from tables in a database and statements that change the data in the database. SQL statements under this category are SELECT, INSERT, UPDATE, DELETE, etc.
  (2) **Data Definition Language (DDL):** These statements define the structure of the database DDL consists of these statements that CREATE, ALTER and DROP database objects and statements that grant and revoke privileges and roles to users of the database. Statements under this category are CREATE, ALTER, DROP, GRANT, REVOKE, etc.
  (3) **Transaction Control Language (TCL):** These commands manage changes made by data manipulation.

Data Definition Language

- The DDL provides a set of definitions to specify the storage structure and access methods used by the database system.
- A data dictionary is a file that contains the metadata.
- Whenever data is read or modified in the database system, the data directory is consulted.
- **Creating database objects:** CREATE command

  Create CREATE command is used to create database objects.

  Syntax

  CREATE TABLE name

```
(Column_name 1 datatype    [column constraint],
Column –name 2 datatype    [column constraint],
⋮
Column_name n datatype     [column constraint]);
```

Data Manipulation Language

- A Data Manipulation language is a language that enables users to access or manipulate data as organized by the appropriate data model.
- DMLs are of two types:
  - Procedural DMLs require the user to specify what data is needed and how to get it.
  - Non- procedural DMLs require the user to specify what data is needed without specifying how to get it.

Data Types

- A data type defines the type of data that is stored in a column.

Commonly used data types in MySQL

Data type	Description
CHAR(n)	Specifies character type data of length n where n could be any value from 0 to 255. CHAR is of fixed length, means, declaring CHAR (10) implies to reserve spaces for 10 characters. If data does not have 10 characters (e.g., 'city' has four characters), MySQL fills the remaining 6 characters with spaces padded on the right.
VARCHAR(n)	Specifies character type data of length where n could be any value from 0 to 65535. But unlike CHAR, VARCHAR(n) is a variable-length data type. That is, declaring VARCHAR (30) means a maximum of 30 characters can be stored but the actual allocated tytes will depend on the length of entered string. So 'city' in VARCHAR (30) will occupy space needed to store 4 characters only.
INT	INT specifies an integer value. Each INT value occupies 4 bytes of storage. The range of unsigned values allowed in a 4 byte integer type are 0 to 4,294,967,295. For values larger than that, we have to use BIGINT, which occupies 8 bytes.
FLOAT	Holds numbers with decimal points. Each FLOAT value occupies 4 bytes.
DATE	The DATE type is used for dates in 'YYYY-MM-DD' format. YYYY is the 4 digit year, MM is the 2 digit month and DD is the 2 digit date. The supported range is '1000-01-01' To '9999-12-31'.

- **Constraints applied on a table:**
  - **Unique Constraint:** It ensures no two rows in the table have same value for this field. UNIQUE keyword is used to apply the unique constraint. The NOT NULL constraint should be applied to the field with UNIQUE constraint.
  - **Primary Key constraint:** declares a column as the primary key of the table. PRIMARY KEY keyword is used to declare a column as primary key and it should also be declared NOT NULL.
  - **Default constraint** specifies a default value using the DEFAULT clause.
  - **Check Constraint** limits values that can be inserted into a column of table. It can consist of
    - An expression using IN
    - A range of constant expressions specified using BETWEEN
    - A pattern specified using LIKE
    - Multiple conditions using OR, AND etc.
  - **Foreign key constraint** is used to set relationship between tables using REFERENCES keyword.

The following statement creates an EMPLOYEE table :

```
CREATE TABLE EMPLOYEE
(Emp_No INT(6) PRIMARY KEY,
Emp_Name char (30) NOT NULL,
DEPT        CHAR (20)
Salary      INT
);
```

- **Altering definition of table : Alter Command**

ALTER TABLE command is used to alter the definition of a table in the database.

Syntax

ALTER TABLE Table_name ADD column_name datatype size :

e.g ALTER TABLE employee ADD age INT;

Sometimes we need to change the datatype of a column. To do this, we use ALTER TABLE with MODIFY column command

Syntax:

ALTER TABLE Table –name MODIFY column name. datatype (size);

e.g ALTER TABLE Employee MODIFY salary INT (10);

- **INSERT Command**

INSERT command is used to add rows to a table.

Syntax

INSERT INTO Table_name VALUES (data_value1, data_value2);

The number and sequence of data values should match that of columns in the table, If the number of data values less then specify the column names into which data is being entered as illustrated.

INSERT INTO Table_name (column1, column2) VALUES (data_value1, data_value 2);

- To insert null values, NULL may be used,

  INSERT INTO EMP VALUES (1001,'sharma',NULL, 3000, NULL);

- **SELECT Command**

  SELECT command is used to retrieve the sub part of rows or columns from one or more tables.

  Syntax:

  SELECT column_name FROM Table_name;

Table:Employee

Emp_code	Emp_name	Designation	Salary	Joining-Date
1001	Rahul	Accountant	25000	2011-5-25
1002	Krishna	Clerk	20000	2010-6-19
1003	Akshat	Accountant	22000	2012-7-22
1004	Apoorvi	Clerk	18000	2013-3-17
1005	Nishant	Supervisor	24000	2016-4-23
1006	Sonam	Accountant	22000	2010-5-24
1007	Pihu	Manager	38000	2012-6-18

e.g SELECT Emp_Name FROM Employee;

Output

Emp_ Name
Rahul
Krishna
Akshat
Apoorvi
Nishant
Sonam
Pihu

If you want to display the details of all employee, asterisk * is used.

Syntax SELECT*FROM Table_name ;

e.g. SELECT* FROM Employee;

- **Using Distinct keyword**

  DISTINCT keyword is used to eliminate the duplicate rows from the result of SELECT statement.

  Syntax SELECT DISTINCT column_name FROM Table_name;

  e g SELECT DISTINCT Designation FROM Employee;

  Output

Designation
Clerk
Accountant
Supervisor
Manager

- **Using ALL Keyword**

  If you use All Keyword instead of DISTINCT, it will give all rows with duplicate from the result of a SELECT statement.

  Syntax SELECT ALL column_name From Table_name ;

  e.g. SELECT ALL Designation FROM Employee;

Output

Designation
Accountant
Clerk
Accountant
Clerk
Supervisor
Accountant
Manager

- **Using WHERE clause**

  WHERE clause in SELECT statement specifies the criteria for selection of rules to be returned. It gives the particular result based on some condition.

  Syntax SELECT column_name 1, column_name 2

  FROM Table_name WHERE condition;

  e.g. SELECT Emp_name Designation FROM Employee WHERE salary >22000;

  Output

Emp_ name	Designation
Rahul	Accountant
Sonam	Accountant
Pihu	Manager

- **Logical Operator**

  Logical operators are used with WHERE clause. These operators are following:

  - **OR operator**

    e.g. To list the employee's details having Emp_code as 1004 or 1006 from table Employee.

    SELECT Emp_code, Emp_name, Designation, salary FROM Employee WHERE (Emp_code =1004 OR Emp_code 1006);

  - **AND Operator**

    e.g. To list the employee's details having Emp_code as 1005 but with Salary less than 24000.

    SELECT Emp_code, Emp_name, Designation, Salary FROM Employee WHERE (Emp_code 1005 AND Salary < 24000);

  - **NOT operator**

    e.g. To list all employees details whose Emp_code are other than 1004.

    SELECT Emp_code, Emp_name, Designation, Salary FROM Employee Where (NOT Emp_code =1004).

  - **BETWEEN operator** This Operator defines the specified range of values that come to make condition true. In this, lower limit and upper limit are given as a range.

    e.g. SELECT Emp_code, Emp_Name , Designation FROM Employee

    WHERE Salary BETWEEN 21000 AND 25000;

    Output

Emp_code	Emp_Name	Designation
1001	Rahul	Accountant
1003	Akshat	Accountant
1005	Nishant	Supervisor
1006	Sonam,	Accountant

- **IN Operator**

  This operator allows you to easily test if the expression matches any value in the list of values. You can also use NOT IN to exclude the rows in your list.

  e.g. To list all the employee's with designation Clerk, Manager

  SELECT Emp_Code, Emp_name, Designation, Salary FROM Employees.

  WHERE Designation IN ('Clerk ','Manager);

Output

Emp_ code	Emp_Name	Designation	Salary
1002	Krishna	Clerk	20000
1004	Apoorvi	Clerk	18000
1007	Pihu	Manager	38000

But if you want to display the details of those employee whose designation NOT IN Accountant.

SELECT* FROM Employees

WHERE Designation NOT IN ('Accountant');

Emp_ code	Emp_Name	Designation	Salary	Joining Date
1002	Krishna	Clerk	20000	2010-6-19
1004	Apoorvi	Clerk	18000	2013-3-17
1005	Nishant	Supervisor	24000	2016-4-23
1007	Pihu	Manager	38000	2012-6-18

- **LIKE Operator**

  This operator is used in a WHERE clause to search for specified pattern in a column. Pattern are defined using two special wildcard characters.

  - percent (%) matches any substring
  - underscore (_) matches any one character
  - To illustrate pattern matching
  - "abc%" matches any string that start with 'abc'
  - "abc%" matches any string in which abc comes in mid
  - "_ _ _ _" matches any string of exactly four characters.
  - "_ _ _ _%" matches any string of at least 3 characters

  e.g. SELECT Emp_code, Emp_name FROM Employee WHERE Emp_Name Like "%oo%";

  Output

Emp_code	Emp _Name
1004	Appoorvi

- **ORDER BY Clause**

  This clause is used with SELECT statement for arranging retrieved data in sorted order. To sort the data in descending order DESC keyword is used while order by clause by default sorts the retrieved data in ascending order.

  Syntax SELECT column_name 1 , column_name 2
  FROM Table_name
  WHERE condition
  ORDER By column_name ;

  e.g SELECT Emp_Code, Emp_name, Salary
  FROM Employee
  WHERE Salary <25000
  ORDER By Emp_Name;

Output

Emp_code	Emp_Name	Salary
1003	Akshat	22000
1004	Apoorvi	18000
1002	Krishna	20000
1005	Nishant	24000
1006	Sonam	22000

- **DELETE Command**

  DELETE Command is used to remove the rows of the Table. It does not delete individual field from the table.

  Syntax DELETE FROM Table_Name where condition;

  e.g. DELETE FROM Employee WHERE Emp_code = 1004;

SELECT* FROM Employee;

Output

Emp_ code	Emp-Name	Designation	Salary	Joining Date
1001	Rahul	Accountant	25000	2011-5-25
1002	Krishna	Clerk	20000	2010-6-19
1003	Akshat	Accountant	22000	2012-7-22
1005	Nishant	Supervisor	24000	2016-4-23
1006	Sonam	Accountant	22000	2010-5-24
1007	Pihu	Manager	38000	2012-6-18

To remove the all rows from table use

DELETE FROM Table_name;

It will delete all rows but not a table structure.

- **DROP Command**

  DROP command is used to delete a whole database or just a table. It destroys the objects like an existing database, table, index, etc.

  Syntax DROP TABLE Table_Name;

  Or

  DROP DATABASE database_name ;

- **UPDATE Command**

  UPDATE Command is used to modify the existing records in a table. You can use WHERE clause with the UPDATE command to update the selected rows, otherwise all rows would be affected.

  Syntax

  UPDATE Table_name
  SET Column 1 = value 1 Column 2 = value .....,
  WHERE condition;

  e.g. To change the salary of all employees to 35000, use this

  UPDATE Employee
  Set Salary =35000;
  Where Salary < 24000;

  But if you want to change the salary to 35000 only for those Employees that have salary less then or equal to 20000, use this

  UPDATE Employee
  SET Salary = 35000
  WHERE Salary <=20000;

  To update multiple columns change the Salary and designation of those Employees who have Emp_code 1002

  UPDATE Employee
  SET Salary = 40000, Designation = Supervisor WHERE Emp_code 1002;

❐❐❐

Chapter-3
Introduction to the Emerging Trends

Revision Notes

- The Emerging trends (or Trending words) is a list of words that describe the topics which emerged faster than any others over the last 24 hours.
- Artificial intelligence is the simulation of human intelligence processes by machines, especially computer systems. Specific applications of AI include expert systems, natural language processing, speech recognition and machine vision.

- Machine learning is a method of data analysis that automates analytical model building. It is a branch of Artificial Intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.
- Natural Language Processing (NLP) refers to the branch of computer science—and more specifically, the branch of Artificial Intelligence or AI—concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.
- Augmented reality is an interactive experience of a real-world environment where the objects that reside in the real world are enhanced by computer-generated perceptual information, sometimes across multiple sensory modalities, including visual, auditory, haptic, somatosensory and olfactory.
- Virtual reality is a simulated experience that can be similar to or completely different from the real world. Applications of virtual reality include entertainment, education and business.
- Robotics is an interdisciplinary branch of computer science and engineering. Robotics involves design, construction, operation, and use of robots. The goal of robotics is to design machines that can help and assist humans.
- Big data is a field that treats ways to analyze, systematically extract information from, or otherwise deal with data sets that are too large or complex to be dealt with by traditional data-processing application software.
- Volume, velocity, variety, veracity and value are the five keys to making big data a huge business.
- The Internet of things describes physical objects that are embedded with sensors, processing ability, software, and other technologies, and that connect and exchange data with other devices and systems over the Internet or other communications networks.
- Top Internet-of-Things (IoT) Examples to Know:
  (i) Connected appliances.
  (ii) Smart home security systems.
  (iii) Autonomous farming equipment.
  (iv) Wearable health monitors.
  (v) Smart factory equipment.
  (vi) Wireless inventory trackers.
  (vii) Ultra-high speed wireless internet.
  (viii) Biometric cyber security scanners.
- A sensor is a device used to measure a property, such as pressure, position, temperature, or acceleration, and respond with feedback. This includes sensors in cars and sensors for home systems, sensors in blood pressure devices and sensors for robotics.
- A smart city uses information and communication technology (ICT) to improve operational efficiency, share information with the public and provide a better quality of government service and citizen welfare. Confident and progressive city plans. People able to live and work within the city, using its resources.
- Cloud computing is the on-demand availability of computer system resources, especially data storage and computing power, without direct active management by the user. Large clouds often have functions distributed over multiple locations, each location being a data center.
- Cloud computing refers to storing and accessing data and programs over the Internet instead of your computer's hard drive.
- Clouds are of two types
  (i) Public cloud
  (ii) Private cloud
- Public cloud refers to a common cloud service made available to multiple subscribers.
- Organizations or individuals that use the same cloud are called Cloud "tenants."
- A private cloud consists of computing resources exclusively owned by one business or organization.
- **Infrastructure as a Service (IaaS):** IaaS provides users access to raw computing resources such processing power, data storage capacity, and networking, in the context of a secure data center.
- **Platform as a Service (PaaS):** Geared toward software development teams, PaaS offerings provide computing and storage infrastructure and also a development platform layer, with components such as web servers, database management systems, and software development kits (SDKs) for various programming languages.
- **Software as a Service (SaaS):** SaaS providers offer application-level services tailored to a wide variety of business needs, such as customer relationship management (CRM), marketing automation, or business analytics.
- Grid computing is the use of widely distributed computer resources to reach a common goal. A computing grid can be thought of as a distributed system with non-interactive workloads that involve many files.
- Blockchain is a system of recording information in a way that makes it difficult or impossible to change, hack, or cheat the system. A blockchain is essentially a digital ledger of transactions that is duplicated and distributed across the entire network of computer systems on the blockchain.

❐❐❐