

Unit 1

1. What is the objective to study DBMS?

:- The objective of a database management system is to facilitate the creation of data structures and relieve the programmer of the problems of setting up complicated files. Database management systems have developed from a concept of the data base as something distinct from the programs accessing it.

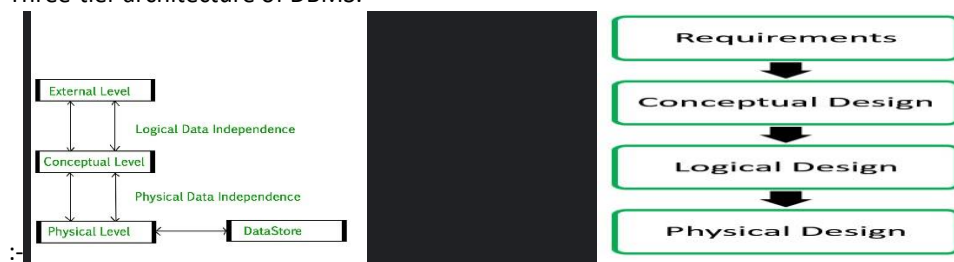
2. The difference between file system and database system.

Basis	File System	DBMS
Structure	The file system is software that manages and organizes the files in a storage medium within a computer.	DBMS is software for managing the database.
Data Redundancy	Redundant data can be present in a file system.	In DBMS there is no redundant data.
Backup and Recovery	It doesn't provide backup and recovery of data if it is lost.	It provides backup and recovery of data even if it is lost.
Query processing	There is no efficient query processing in the file system.	Efficient query processing is there in DBMS.
Consistency	There is less data consistency in the file system.	There is more data consistency because of the process of normalization.
Complexity	It is less complex as compared to DBMS.	It has more complexity in handling as compared to the file system.
Security Constraints	File systems provide less security in comparison to DBMS.	DBMS has more security mechanisms as compared to file systems.
Cost	It is less expensive than DBMS.	It has a comparatively higher cost than a file system.

3. Give examples DBMS?

- :- MySQL. ...
- Oracle Database. ...
- MongoDB. ...
- IBM Db2 DBMS. ...
- Amazon RDS. ...
- PostgreSQL. ...

4. Three-tier architecture of DBMS.



1. **Physical Level:** At the physical level, the information about the location of database objects in the data store is kept. Various users of DBMS are unaware of the locations of these objects. In simple terms, physical level of a database describes how the data is being stored in secondary storage devices like disks and tapes and also gives insights on additional storage details.
2. **Conceptual Level:** At conceptual level, data is represented in the form of various database tables. For Example, STUDENT database may contain STUDENT and COURSE tables which will be visible to users but users are unaware of their storage. Also referred as logical schema, it describes what kind of data is to be stored in the database.
3. **External Level:** An external level specifies a view of the data in terms of conceptual level tables. Each external level view is used to cater to the needs of a particular category of users. For Example, FACULTY of a university is interested in looking course details of students, STUDENTS are interested in looking at all details related to academics, accounts,

courses and hostel details as well. So, different views can be generated for different users. The main focus of external level is data abstraction.

Advantages of DBMS

DBMS helps in efficient organization of data in database which has following advantages over typical file system:

- **Minimized redundancy and data inconsistency:** Data is normalized in DBMS to minimize the redundancy which helps in keeping data consistent. For Example, student information can be kept at one place in DBMS and accessed by different users. This minimized redundancy is due to primary key and foreign keys
- **Simplified Data Access:** A user need only name of the relation not exact location to access data, so the process is very simple.
- **Multiple data views:** Different views of same data can be created to cater the needs of different users. For Example, faculty salary information can be hidden from student view of data but shown in admin view.
- **Data Security:** Only authorized users are allowed to access the data in DBMS. Also, data can be encrypted by DBMS which makes it secure.

5. Two-tier architecture of DBMS

:- the client is on the first tier. The database server and web application server reside on the same server machine, which is the second tier.

6. One-tier architecture of DBMS

:-, the data is directly provided to the user and that user can directly use the database through the computer

7. Write are the Full form of DDL, DML and DCL?

:- DDL – Data Definition Language. DQL – Data Query Language. DML – Data Manipulation Language. DCL – Data Control Language.

8. Explain DDL, DML, DCL and TCL.

:- DDL (Data Definition Language):

DDL or Data Definition Language actually consists of the SQL commands that can be used to define the database schema. It simply deals with descriptions of the database schema and is used to create and modify the structure of database objects in the database. DDL is a set of SQL commands used to create, modify, and delete database structures but not data. These commands are normally not used by a general user, who should be accessing the database via an application.

List of DDL commands:

- **CREATE:** This command is used to create the database or its objects (like table, index, function, views, store procedure, and triggers).
- **DROP:** This command is used to delete objects from the database.
- **ALTER:** This is used to alter the structure of the database.
- **TRUNCATE:** This is used to remove all records from a table, including all spaces allocated for the records are removed.
- **COMMENT:** This is used to add comments to the data dictionary.
- **RENAME:** This is used to rename an object existing in the database.

DML(Data Manipulation Language):

The SQL commands that deals with the manipulation of data present in the database belong to DML or Data Manipulation Language and this includes most of the SQL statements. It is the component of the SQL statement that controls access to data and to the database. Basically, DCL statements are grouped with DML statements.

List of DML commands:

INSERT : It is used to insert data into a table.

UPDATE: It is used to update existing data within a table.

DELETE : It is used to delete records from a database table.

LOCK: Table control concurrency.

CALL: Call a PL/SQL or JAVA subprogram.

TCL (Transaction Control Language):

Transactions group a set of tasks into a single execution unit. Each transaction begins with a specific task and ends when all the tasks in the group successfully complete. If any of the tasks fail, the transaction fails. Therefore, a transaction has only two results: success or failure. You can explore more about transactions here. Hence, the following TCL commands are used to control the execution of a transaction:

COMMIT: Commits a Transaction.

ROLLBACK: Rolls back a transaction in case of any error occurs.

SAVEPOINT: Sets a save point within a transaction.

SET TRANSACTION: Specifies characteristics for the transaction.

DCL (Data Control Language):

DCL includes commands such as GRANT and REVOKE which mainly deal with the rights, permissions, and other controls of the database system.

List of DCL commands:

GRANT: This command gives users access privileges to the database.

REVOKE: This command withdraws the user's access privileges given by using the GRANT command.

Unit 2

9. Write SQL query to get the Student name whose salary is between 1000 and 2000.

**:- SELECT * FROM students
WHERE salary BETWEEN 1000 and 3000;**

10. Write SQL query to get the Students Name and address from Student table.

:- SELECT S_NAME, S_ADDRESS, FROM STUDENTS ;

11. Write SQL query to get the Students Name who have got marks greater than 90 from Student table Explain Union and Intersection in DBMS.

:- SELECT StudID, StudName ,Marks FROM STUDENT_MARKS

WHERE MARKS >80 ORDER BY RIGHT(StudName,3), StudID ASC;

12. Explain join in DBMS.

:- In DBMS, a join statement is mainly used to combine two tables based on a specified common field between them. If we talk in terms of Relational algebra, it is the cartesian product of two tables followed by the selection operation. Thus, we can execute the product and selection process on two tables using a single join statement. We can use either 'on' or 'using' clause in MySQL to apply predicates to the join queries.

A Join can be broadly divided into two types:

1. **Inner Join :-** Inner Join is a join that can be used to return all the values that have matching values in both the tables. Inner Join can be depicted using the below diagram.

The inner join can be further divided into the following types:

Equi Join :- Equi Join is an inner join that uses the equivalence condition for fetching the values of two tables.

Natural Join :- Natural Join is an inner join that returns the values of the two tables on the basis of a common attribute that has the same name and domain. It does not use any comparison operator. It also removes the duplicate attribute from the results.

Eg :- Select * from employee Natural Join department;

2. **Outer Join :-** Outer Join is a join that can be used to return the records in both the tables whether it has matching records in both the tables or not.

The outer join can be further divided into three types

Left-Outer Join :- The Left-Outer Join is an outer join that returns all the values of the left table, and the values of the right table that has matching values in the left table. If there is no matching result in the right table, it will return null values in that field

Right-Outer Join :- The Right-Outer Join is an outer join that returns all the values of the right table, and the values of the left table that has matching values in the right table

Full-Outer Join :- The Full-Outer join contains all the values of both the tables whether they have matching values in them or not

13. Explain Cartesian Product in DBMS with example.

:- The final product of the sets will be a collection of all ordered pairs obtained by the product of the two non-empty sets. In an ordered pair, two elements are taken from each of the two sets. Example: Cartesian product of two sets $A = \{2,3\}$ and $B = \{x,y\}$ will be: $A \times B = \{(2,x), (2,y), (3,x), (3,y)\}$.

CROSS PRODUCT is a binary set operation means, at a time we can apply the operation on two relations. But the two relations on which we are performing the operations do not have the same type of tuples, which means Union compatibility (or Type compatibility) of the two relations is not necessary.

Notation:

$A \times S$

where A and S are the relations,

the symbol 'X' is used to denote the CROSS PRODUCT operator.

Example:-

STUDENTS X DETAILS

SNO	FNAME	LNAME
1	Albert	Singh
2	Nora	Fatehi

ROLLNO	AGE
5	18
9	21

SNO	FNAME	LNAME	ROLLNO	AGE
1	Albert	Singh	5	18
1	Albert	Singh	9	21
2	Nora	Fatehi	5	18
2	Nora	Fatehi	9	21

We can observe that the number of tuples in STUDENT relation is 2, and the number of tuples in DETAIL is 2. So the number of tuples in the resulting relation on performing CROSS PRODUCT is $2 \times 2 = 4$.

Important points on CARTESIAN PRODUCT(CROSS PRODUCT) Operation:

- The cardinality (number of tuples) of resulting relation from a Cross Product operation is equal to the number of attributes(say m) in the first relation multiplied by the number of attributes in the second relation(say n).
- Cardinality = $m \times n$
- The Cross Product of two relation $A(R_1, R_2, R_3, \dots, R_p)$ with degree p, and $B(S_1, S_2, S_3, \dots, S_n)$ with degree n, is a relation $C(R_1, R_2, R_3, \dots, R_p, S_1, S_2, S_3, \dots, S_n)$ with degree $p + n$ attributes.
- Degree = $p + n$
- In SQL, CARTESIAN PRODUCT(CROSS PRODUCT) can be applied using CROSS JOIN.
- In general, we don't use cartesian Product unnecessarily, which means without proper meaning we don't use Cartesian Product. Generally, we use Cartesian Product followed by a Selection operation and comparison.

14. Explain Select and Project.

:- [YHA CLICK KRKE PDH LO BHAI](#)

Unit

3

1. What is partial functional dependency?
2. What is transitive functional dependency?
3. What is full functional dependency?
4. Explain 1NF, 2NF, 3NF, BCNF, 4NF with Example.
5. What is the difference between 3NF and BCNF.
6. $R(C, S, Z)$
7. $CS \rightarrow Z$
 $Z \rightarrow C$
8. Find the list of candidate key.
9. Find Normal Form

Emp(ename, did, empid, salary)

ename, did \rightarrow salary did, empid \rightarrow salary ename

\rightarrow empid empid \rightarrow ename Find the Normal

Form

$R(A, B, C, D, E)$

$A \rightarrow BC$

$BC \rightarrow E$

$E \rightarrow DA$

Unit 4

1. What is concurrent schedule?
2. What is deadlock?
3. What do you mean by serializable schedule.
4. What is serial schedule?
5. What do you mean by ACID property?
6. check the given schedule is serializable or not. Consider the following schedule for transaction T1, T2, T3

T1	T2	T3
r(x)		
	r(y)	
		r(y)
	w(y)	
w(x)		
		w(x)
	r(x)	

	$w(x)$	
--	--------	--

Unit 5

1. What is temporary update problem?

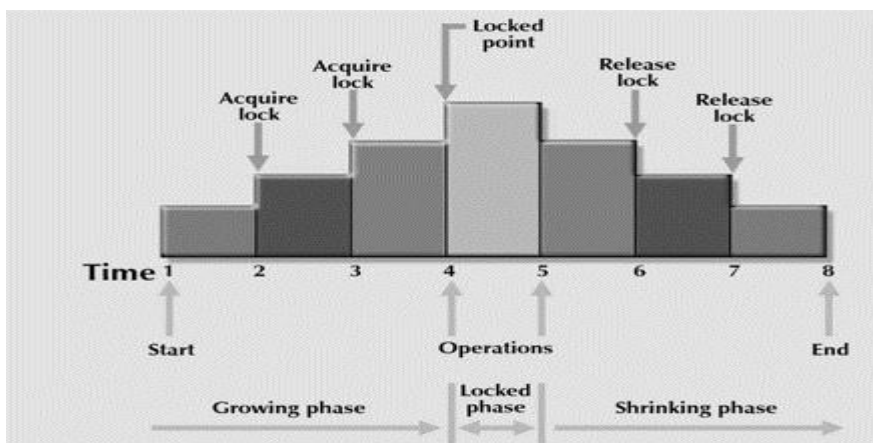
:- Temporary update or dirty read problem occurs when one transaction updates an item and fails. But the updated item is used by another transaction before the item is changed or reverted back to its last value

2. Explain Two Phase Locking Protocol .

:- Every transaction will lock and unlock the data item in two different phases.

- **Growing Phase** – All the locks are issued in this phase. No locks are released, after all changes to data-items are committed and then the second phase (shrinking phase) starts.
- **Shrinking phase** – No locks are issued in this phase, all the changes to data-items are noted (stored) and then locks are released.

The 2PL locking protocol is represented diagrammatically as follows –



In the growing phase transaction reaches a point where all the locks it may need has been acquired. This point is called LOCK POINT.

After the lock point has been reached, the transaction enters a shrinking phase.

Types

Two phase locking is of two types –

Strict two phase locking protocol

A transaction can release a shared lock after the lock point, but it cannot release any exclusive lock until the transaction commits. This protocol creates a cascade less schedule.

Cascading schedule: In this schedule one transaction is dependent on another transaction. So if one has to rollback then the other has to rollback.

Rigorous two phase locking protocol

A transaction cannot release any lock either shared or exclusive until it commits.

The 2PL protocol guarantees serializability, but cannot guarantee that deadlock will not happen.

Example

Let T1 and T2 are two transactions.

$T1=A+B$ and $T2=B+A$

T1	T2
Lock-X(A)	Lock-X(B)
Read A;	Read B;
Lock-X(B)	Lock-X(A)

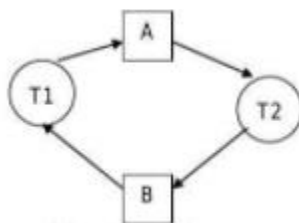
Here,

Lock-X(B) : Cannot execute Lock-X(B) since B is locked by T2.

Lock-X(A) : Cannot execute Lock-X(A) since A is locked by T1.

In the above situation T1 waits for B and T2 waits for A. The waiting time never ends. Both the transaction cannot proceed further at least any one releases the lock voluntarily. This situation is called deadlock.

The wait for graph is as follows –



Wait for graph: It is used in the deadlock detection method, creating a node for each transaction, creating an edge T_i to T_j , if T_i is waiting to lock an item locked by T_j . A cycle in WFG indicates a deadlock has occurred. WFG is created at regular intervals.

3. What is the objective to do concurrency control?

:- Concurrency control is provided in a database to:

- (i) enforce isolation among transactions.
- (ii) preserve database consistency through consistency preserving execution of transactions.
- (iii) resolve read-write and write-read conflicts

- Concurrency control manages the transactions simultaneously without letting them interfere with each other.
- The main objective of concurrency control is to allow many users perform different operations at the same time.
- Using more than one transaction concurrently improves the performance of system.
- If we are not able to perform the operations concurrently, then there can be serious problems such as loss of data integrity and consistency.
- Concurrency control increases the throughput because of handling multiple transactions simultaneously.
- It reduces waiting time of transaction.

4. Explain Dirty read problem.

:- **SHASHANK SINGH**

The dirty read problem occurs when one transaction updates an item of the database, and somehow the transaction fails, and before the data gets rollback, the updated database item is accessed by another transaction. There comes the Read-Write Conflict between both transactions.

A dirty read can cause duplicate rows to be returned where none exist. Alternatively, a dirty read can cause no rows to be returned when one (or more) actually exists. In some cases, dirty reads can return data that was never in the database at all (e.g., rolled back before committed)

Dirty data refers to data that contains erroneous information. It may also be used when referring to data that is in memory and not yet loaded into a database. The complete removal of dirty data from a source is

impractical or virtually impossible. Dirty Reads A dirty read occurs when a transaction reads data that has not yet been committed. For example, suppose transaction 1 updates a row. Transaction 2 reads the updated row before transaction 1 commits the update.

5. What is concurrency control?

:- Concurrency control concept comes under the Transaction in database management system (DBMS). It is a procedure in DBMS which helps us for the management of two simultaneous processes to execute without conflicts between each other, these conflicts occur in multi user systems.

Concurrency can simply be said to be executing multiple transactions at a time. It is required to increase time efficiency. If many transactions try to access the same data, then inconsistency arises. Concurrency control required to maintain consistency data.

For example, if we take ATM machines and do not use concurrency, multiple persons cannot draw money at a time in different places. This is where we need concurrency.

Advantages

- Waiting time will be decreased.
- Response time will decrease.
- Resource utilization will increase.
- System performance & Efficiency is increased.

Control concurrency

The simultaneous execution of transactions over shared databases can create several data integrity and consistency problems.

For example, if too many people are logging in the ATM machines, serial updates and synchronization in the bank servers should happen whenever the transaction is done, if not it gives wrong information and wrong data in the database.

Main problems in using Concurrency

Updates will be lost – One transaction does some changes and another transaction deletes that change. One transaction nullifies the updates of another transaction.

Uncommitted Dependency or dirty read problem – On variable has updated in one transaction, at the same time another transaction has started and deleted the value of the variable there the variable is not getting updated or committed that has been done on the first transaction this gives us false values or the previous values of the variables this is a major problem.

Inconsistent retrievals – One transaction is updating multiple different variables, another transaction is in a process to update those variables, and the problem occurs is inconsistency of the same variable in different instances.

Concurrency control techniques

Locking

Lock guaranties exclusive use of data items to a current transaction. It first accesses the data items by acquiring a lock, after completion of the transaction it releases the lock.

Types of Locks

Shared Lock [Transaction can read only the data item values]

Exclusive Lock [Used for both read and write data item values]

Time Stamping

Time stamp is a unique identifier created by DBMS that indicates relative starting time of a transaction. Whatever transaction we are doing it stores the starting time of the transaction and denotes a specific time.

Optimistic

It is based on the assumption that conflict is rare and it is more efficient to allow transactions to proceed without imposing delays to ensure serializability.

6. What is Lost Update problem?

:- Consider the case where two users are about to update the same row/document in some data store. For example, let user A retrieve some row first. After that, assume that user B retrieves the same row; however, B writes his update immediately, and in particular before A writes his update.

In the lost update problem, an update done to a data item by a transaction is lost as it is overwritten by the update done by another transaction.

7. Explain validation-Based Protocol.

:- Validation Based Protocol is also called Optimistic Concurrency Control Technique. This protocol is used in DBMS (Database Management System) for avoiding concurrency in transactions. It is called optimistic because of the assumption it makes, i.e. very less interference occurs, therefore, there is no need for checking while the transaction is executed.

In this technique, no checking is done while the transaction is been executed. Until the transaction end is reached updates in the transaction are not applied directly to the database. All updates are applied to local copies of data items kept for the transaction. At the end of transaction execution, while execution of the transaction, a validation phase checks whether any of transaction updates violate serializability. If there is no violation of serializability the transaction is committed and the database is updated; or else, the transaction is updated and then restarted.

Optimistic Concurrency Control is a three-phase protocol. The three phases for validation based protocol:

Read Phase:

Values of committed data items from the database can be read by a transaction. Updates are only applied to local data versions.

Validation Phase:

Checking is performed to make sure that there is no violation of serializability when the transaction updates are applied to the database.

Write Phase:

On the success of the validation phase, the transaction updates are applied to the database, otherwise, the updates are discarded and the transaction is slowed down.

The idea behind optimistic concurrency is to do all the checks at once; hence transaction execution proceeds with a minimum of overhead until the validation phase is reached. If there is not much interference among transactions most of them will have successful validation, otherwise, results will be discarded and restarted later. These circumstances are not much favourable for optimization techniques, since, the assumption of less interference is not satisfied.

In order to perform the Validation test, each transaction should go through the various phases as described above. Then, we must know about the following three time-stamps that we assigned to transaction T_i , to check its validity:

1. Start(T_i): It is the time when T_i started its execution.
2. Validation(T_i): It is the time when T_i just finished its read phase and begin its validation phase.
3. Finish(T_i): the time when T_i end it's all writing operations in the database under write-phase.

Two more terms that we need to know are:

1. Write_set: of a transaction contains all the write operations that T_i performs.
2. Read_set: of a transaction contains all the read operations that T_i performs.

In the Validation phase for transaction T_i the protocol inspect that T_i doesn't overlap or intervene with any other transactions currently in their validation phase or in committed. The validation phase for T_i checks that for all transaction T_j one of the following below conditions must hold to being validated or pass validation phase:

1. Finish(T_j) < Starts(T_i), since T_j finishes its execution means completes its write-phase before T_i started its execution(read-phase). Then the serializability indeed maintained.
2. T_i begins its write phase after T_j completes its write phase, and the read_set of T_i should be disjoint with write_set of T_j .
3. T_j completes its read phase before T_i completes its read phase and both read_set and write_set of T_i are disjoint with the write_set of T_j

Unit 6

1. What do you mean by research?

- Research databases are organized collections of computerized information or data such as periodical articles, books, graphics and multimedia that can be searched to retrieve information

2. Explain Quantitative data Analysis

- Data analysis is defined as a process of cleaning, transforming, and modeling data to discover useful information for business decision-making. The purpose of Data Analysis is to extract useful information from data and taking the decision based upon the data analysis.

In this Data Science Tutorial, you will learn:

If your business is not growing, then you have to look back and acknowledge your mistakes and make a plan again without repeating those mistakes. And even if your business is growing, then you have to look forward to making the business to grow more. All you need to do is analyze your business data and business processes.

- **Data Analysis Tools**
- **Data Analysis Tools**
- **Data Analysis Tools**

Data analysis tools make it easier for users to process and manipulate data, analyze the relationships and correlations between data sets, and it also helps to identify patterns and trends for interpretation. Here is a complete list of tools used for data analysis in research.

Types of Data Analysis: Techniques and Methods

There are several types of Data Analysis techniques that exist based on business and technology. However, the major Data Analysis methods are:

- Text Analysis
- Statistical Analysis
- Diagnostic Analysis
- Predictive Analysis
- Prescriptive Analysis
- Text Analysis

Text Analysis is also referred to as Data Mining. It is one of the methods of data analysis to discover a pattern in large data sets using databases or data mining tools. It used to transform raw data into business information. Business Intelligence tools are present in the market which is used to take strategic business decisions. Overall it offers a way to extract and examine data and deriving patterns and finally interpretation of the data.

Statistical Analysis

Statistical Analysis shows “What happen?” by using past data in the form of dashboards. Statistical Analysis includes collection, Analysis, interpretation, presentation, and modeling of data. It analyses a set of data or a sample of data. There are two categories of this type of Analysis – Descriptive Analysis and Inferential Analysis.

Descriptive Analysis

analyses complete data or a sample of summarized numerical data. It shows mean and deviation for continuous data whereas percentage and frequency for categorical data.

Inferential Analysis

analyses sample from complete data. In this type of Analysis, you can find different conclusions from the same data by selecting different samples.

Diagnostic Analysis

Diagnostic Analysis shows “Why did it happen?” by finding the cause from the insight found in Statistical Analysis. This Analysis is useful to identify behavior patterns of data. If a new problem arrives in your business process, then you can look into this Analysis to find similar patterns of that problem. And it may have chances to use similar prescriptions for the new problems.

Predictive Analysis

Predictive Analysis shows “what is likely to happen” by using previous data. The simplest data analysis example is like if last year I bought two dresses based on my savings and if this year my salary is increasing double then I can buy four dresses.

Prescriptive Analysis

Prescriptive Analysis combines the insight from all previous Analysis to determine which action to take in a current problem or decision. Most data-driven companies are utilizing Prescriptive Analysis because predictive and descriptive Analysis are not enough to improve data performance. Based on current situations and problems, they analyze the data and make decisions.

Data Analysis Process

The Data Analysis Process is nothing but gathering information by using a proper application or tool which allows you to explore the data and find a pattern in it. Based on that information and data, you can make decisions, or you can get ultimate conclusions.

Data Analysis consists of the following phases:

- Data Requirement Gathering
- Data Collection
- Data Cleaning
- Data Analysis
- Data Interpretation
- Data Visualization

Data Requirement Gathering

First of all, you have to think about why do you want to do this data analysis? All you need to find out the purpose or aim of doing the Analysis of data. You have to decide which type of data analysis you wanted to do! In this phase, you have to decide what to analyze and how to measure it, you have to understand why you are investigating and what measures you have to use to do this Analysis.

Data Collection

After requirement gathering, you will get a clear idea about what things you have to measure and what should be your findings. Now it's time to collect your data based on requirements. Once you collect your data, remember that the collected data must be processed or organized for Analysis. As you collected data from various sources, you must have to keep a log with a collection date and source of the data.

Data Cleaning

Now whatever data is collected may not be useful or irrelevant to your aim of Analysis, hence it should be cleaned. The data which is collected may contain duplicate records, white spaces or errors. The data should be cleaned and error free. This phase must be done before Analysis because based on data cleaning, your output of Analysis will be closer to your expected outcome.

Data Analysis

Once the data is collected, cleaned, and processed, it is ready for Analysis. As you manipulate data, you may find you have the exact information you need, or you might need to collect more data. During this phase, you can use data analysis tools and software which will help you to understand, interpret, and derive conclusions based on the requirements.

Data Interpretation

After analyzing your data, it's finally time to interpret your results. You can choose the way to express or communicate your data analysis either you can use simply in words or maybe a table or chart. Then use the results of your data analysis process to decide your best course of action.

Data Visualization

Data visualization is very common in your day to day life; they often appear in the form of charts and graphs. In other words, data shown graphically so that it will be easier for the human brain to understand and process it. Data visualization often used to discover unknown facts and trends. By observing relationships and comparing datasets, you can find a way to find out meaningful information.

3. Explain Quantitative data Analysis
:- same question dal, diya inhone
4. Importance of research in social aspect.