Unit -7 Security & Advance DBMS Concepts

Database Security

Security of databases refers to the array of controls, tools, and procedures designed to ensure and safeguard confidentiality, integrity, and accessibility. Security for databases must cover and safeguard the following aspects:

- The database containing data.
- Database management systems (DBMS)
- Any applications that are associated with it.
- Physical database servers or the database server virtual, and the hardware that runs it.
- The infrastructure for computing or network that is used to connect to the database.

Why Database Security is Important?

According to the definition, a data breach refers to a breach of data integrity in databases. The amount of damage an incident like a data breach can cause our business is contingent on various consequences or elements.

- **Stolen Intellectual Property**: If our trade secrets, inventions, or proprietary methods are stolen, it could destroy our competitive edge, making it hard to recover or stay ahead in the industry.
- **Brand Reputation Damage**: Customers and partners may lose trust in us and stop doing business with us if they feel their data isn't secure with our company.
- **Business Disruption**: Some companies can't operate until a data breach is resolved, leading to potential business shutdowns.
- **Fines and Penalties**: Non-compliance with regulations like GDPR, HIPAA, or PCI DSS can result in severe fines, sometimes reaching millions of dollars per violation.
- **Cost of Fixing the Breach**: Companies must pay for investigations, system repairs, customer notifications, and crisis management, adding to the financial burden after a breach.

When it comes to IT security, especially in computer systems, user's access to certain resources and information must be regulated. Of particular importance in managing information in a system is the ability to decide who can have access to what. Of all the models, the two most common models are the

- ✓ Discretionary Access Control (DAC)
- ✓ Mandatory Access Control (MAC)

There are differences in features, advantages, and drawbacks of each model. This article will compare and contrast DAC and MAC, as well as identify their characteristics, applications as well as their impact on system security.

Discretionary Access Control (DAC)

DAC is identity-based access control. DAC mechanisms will be controlled by user identification such as username and password. DAC is discretionary because the owners can transfer objects or any authenticated information to other users. In simple words, the owner can determine the access privileges.

Examples:

1. File Sharing in Personal Computers:

In Windows or macOS, the owner of a file or folder can decide who gets access to it. For instance:

- ✓ A user can right-click on a folder, go to *Properties* > *Sharing*, and allow specific users to read, write, or execute.
- ✓ The user may choose to give read-only access to one person and full control to another.

2. Collaboration Tools - Google Drive or Dropbox:

- ✓ The file owner decides who can view, comment, or edit the file by explicitly sharing it with specific email addresses or making it accessible to "anyone with the link."
- ✓ The control is in the owner's hands, making it discretionary.

3. Small Business Networks:

- ✓ In a small office setting, the administrator creates shared folders and gives team members varying levels of access based on trust or role. The owner can later revoke or transfer permissions to others.

Mandatory Access Control (MAC)

The operating system in MAC will provide access to the user based on their identities and data. To gain access, the user has to submit their personal information. It is very secure because the rules and restrictions are imposed by the admin and will be strictly followed. MAC settings and policy management will be established in a secure network and are limited to system administrators.

Examples

1. **Government Systems and Classified Information:**
    - Systems used in the defense sector, such as those for handling classified documents (Top Secret, Secret, Confidential), use MAC.
    - For example:

- ✓ Only users with the appropriate security clearance and role-based permissions can access certain files.
- ✓ Even if a person accidentally gains access to a computer with classified data, they cannot open restricted files due to enforced policies.

2. **Banking Systems- Core Banking Applications:**

- Employees in a bank have access based on roles (teller, manager, IT admin).
- A teller can only view customer account details for transactions, while a manager may approve loans or higher-value withdrawals.
- Access rules are predefined and cannot be modified by individual users, ensuring compliance.

3. **Healthcare Management Systems-Electronic Health Records (EHR):**

- Only doctors with the appropriate access level can view a patient's medical history, while nurses may have limited access (e.g., vitals or treatment logs).
- These rules are implemented by administrators, not end users, and align with legal standards like HIPAA.

4. **Military Communication Systems:**

- Secure networks for military communication apply MAC policies, ensuring that only authorized personnel can view or send certain communications based on rank or mission clearance.

Differences Between DAC and MAC

SNo	DAC	MAC
1.	DAC stands for Discretionary Access control.	MAC stands for Mandatory Access Control.
2.	DAC is easier to implement.	MAC is difficult to implement.
3.	DAC is less secure to use.	MAC is more secure to use.
4.	In DAC, the owner can determine the access and privileges and can restrict the resources based on the identity of the users.	In MAC, the system only determines the access and the resources will be restricted based on the clearance of the subjects.
5.	DAC has extra labor-intensive properties.	MAC has no labor-intensive property.
6.	Users will be provided access based on their identity and not using levels.	Users will be restricted based on their power and level of hierarchy.

7.	DAC has high flexibility with no rules and regulations.	MAC is not flexible as it contains lots of strict rules and regulations.
8.	DAC has complete trust in users.	MAC has trust only in administrators.
9.	Decisions will be based only on user ID and ownership.	Decisions will be based on objects and tasks, and they can have their own ids.
10.	Information flow is impossible to control.	Information flow can be easily controlled.
11.	DAC is supported by commercial DBMSs.	MAC is not supported by commercial DBMSs.
12.	DAC can be applied in all domains.	MAC can be applied in the military, government, and intelligence.
13.	DAC is vulnerable to trojan horses.	MAC prevents virus flow from a higher level to a lower level.

Introduction to Data Mining

What is data mining?

Data mining is the process of sorting through large data sets to identify patterns and relationships that can help solve business problems through data analysis. Data mining techniques and tools help enterprises to predict future trends and make more informed business decisions.

Data mining is a key part of data analytics and one of the core disciplines in data science, which uses advanced analytics techniques to find useful information in data sets. At a more granular level, data mining is a step in the knowledge discovery in databases (KDD) process, a data science methodology for gathering, processing and analyzing data. Data mining and KDD are sometimes referred to interchangeably, but they're more commonly seen as distinct things.

The process of data mining relies on the effective implementation of data collection, warehousing and processing. Data mining can be used to describe a target data set, predict outcomes, detect fraud or security issues, learn more about a user base, or detect bottlenecks and dependencies. It can also be performed automatically or semi automatically.

Data mining is more useful today due to the growth of big data and data warehousing. Data specialists who use data mining must have coding and programming language experience, as well as statistical knowledge to clean, process and interpret data.

Why is data mining important?

Data mining is an essential part of analytics that helps organizations make better decisions. It provides valuable insights for both historical data analysis and real-time data processing. These insights are used in business intelligence (BI) and advanced analytics.

Data mining helps businesses in various ways, such as improving marketing, advertising, sales, and customer support. It is also useful for managing operations like manufacturing, supply chain management (SCM), finance, and human resources (HR). Additionally, it plays a key role in detecting fraud, managing risks, and planning cybersecurity.

Beyond businesses, data mining is important in fields like healthcare, government, scientific research, mathematics, and sports. It enables smarter strategies and efficient management across all these areas.

The data mining process: How does data mining work?

Data mining is often done by data scientists and skilled analytics professionals, but business analysts and others with data knowledge can also perform it. Modern tools, like machine learning and artificial intelligence (AI), have made data mining easier and more automated, enabling the analysis of large data sets like customer records, transaction logs, and sensor data.

The data mining process generally has four main stages:

1. **Data Gathering:**

   Collect relevant data from various sources, such as databases, data warehouses, or data lakes. External data can also be included. The data is often moved to a central repository, like a data lake, for analysis.

2. **Data Preparation:**

   Clean and prepare the data for analysis. This involves exploring the data, fixing errors, handling duplicates or missing values, and transforming it into a consistent format. Sometimes, raw data is analyzed directly for specific applications.

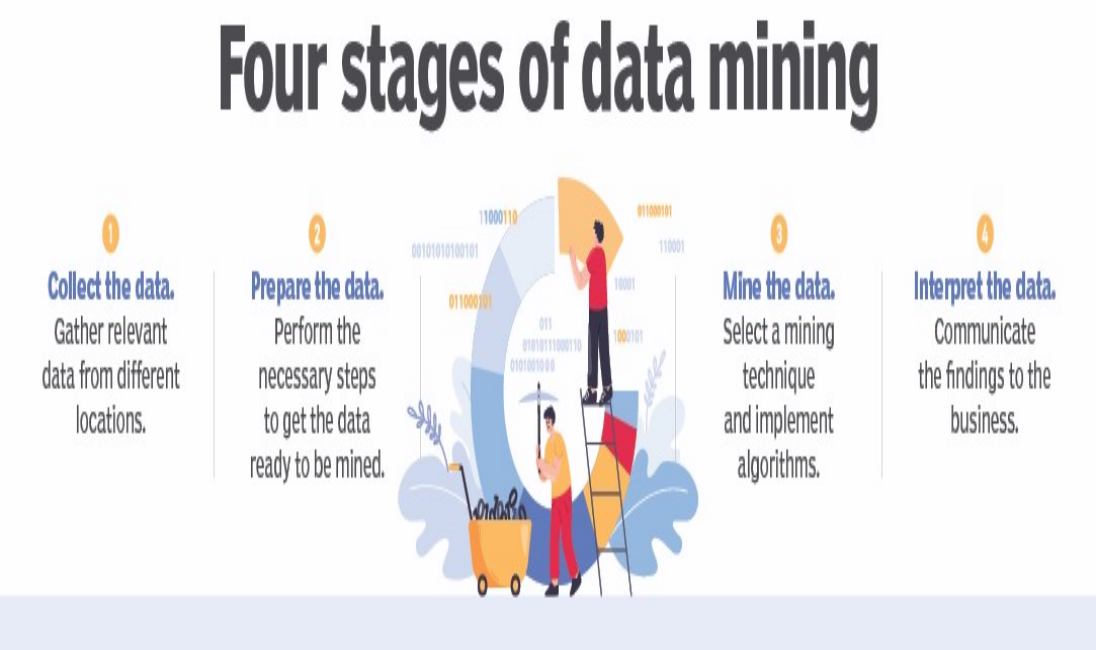
3. **Data Mining:**

   Select the right data mining techniques and algorithms to analyze relationships, patterns, and trends in the data. For machine learning, algorithms are first trained on sample data before being applied to the full dataset.

4. Data Analysis and Interpretation:

Use the mining results to build analytical models for decision-making. Findings are shared with business teams through visualizations and storytelling to ensure the insights are actionable.

This process helps organizations make better decisions and uncover valuable patterns in their data.



Introduction to NoSQL

NoSQL is a type of database management system (DBMS) that is designed to handle and store large volumes of unstructured and semi-structured data.

Unlike traditional relational databases that use tables with pre-defined schemas to store data, NoSQL databases use flexible data models that can adapt to changes in data structures and are capable of scaling horizontally to handle growing amounts of data.

The term NoSQL originally referred to “non-SQL” or “non-relational” databases, but the term has since evolved to mean “not only SQL,” as NoSQL databases have expanded to include a wide range of different database architectures and data models.

NoSQL databases are generally classified into four main categories:

1. **Document databases:** These databases store data as semi-structured documents, such as JSON or XML, and can be queried using document-oriented query languages.
2. **Key-value stores:** These databases store data as key-value pairs, and are optimized for simple and fast read/write operations.

3. **Column-family stores:** These databases store data as column families, which are sets of columns that are treated as a single entity. They are optimized for fast and efficient querying of large amounts of data.
4. **Graph databases:** These databases store data as nodes and edges, and are designed to handle complex relationships between data.

Data Encryption

Data Encryption is an important part of preserving data integrity, and confidentiality, its importance cannot be overestimated. Almost the whole thing on the internet has been encrypted at some point. In this article, we will discuss data encryption and its importance.

What is Data Encryption?

Data Encryption is a method of preserving data confidentiality by transforming it into ciphertext, which can only be decoded using a unique decryption key produced at the time of the encryption or before it. The conversion of plaintext into ciphertext is known as encryption.

Key Objective of Encryption Data

- **Confidentiality:** Encryption ensures that only authorized parties can get access to data and recognize the information.
- **Data Integrity:** Encryption can also provide data integrity by making sure that the encrypted data remains unchanged during transmission. Any unauthorized changes to the encrypted information will render it undecipherable or will fail integrity checks.
- **Authentication:** Encryption may be used as part of authentication mechanisms to verify the identification of the communication party.
- **Non-Repudiation:** Through encryption, events can make sure that they cannot deny their involvement in growing or sending a selected piece of data.

Importance of Data Encryption

The significance of encryption cannot be overstated in any way. Even though your data is stored in a standard infrastructure, it is still possible for it to be hacked. There's always the chance that data will be compromised, but with data encryption, your information will be much more secure. Consider it this way for a moment. If your data is stored in a secure system, encrypting it before sending it out will keep it safe. Sanctioned systems do not provide the same level of protection.

So, how do you think this would play out in real life?

Suppose the user has access to sensitive information while at work. The user may put the information on a portable disc and move it anywhere they choose without any encryption. If the encryptions are set in place ahead of time, the user can still copy the information, but the data

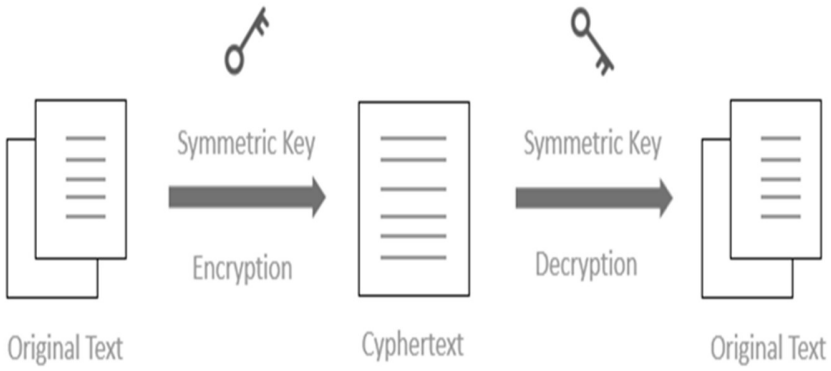
will be unintelligible when they try to see it someplace else. These are the benefits of data encryption that demonstrate its genuine value.

Types of Data Encryption

There are multiple encryption techniques, each of which have been developed with various security requirements in mind. Symmetric and Asymmetric encryption are the two types of data encryption.

1. **Symmetric Key Encryption**

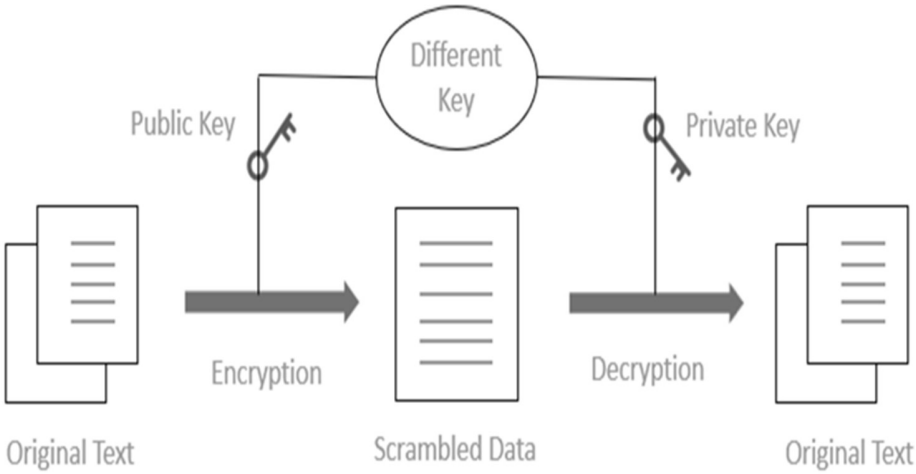
There are a few strategies used in cryptography algorithms. For encryption and decryption processes, some algorithms employ a unique key. In such operations, the unique key must be secured since the system or person who knows the key has complete authentication to decode the message for reading. This approach is known as "symmetric encryption" in the field of network encryption.



Symmetric Encryption

2. **Asymmetric Key Encryption**

Some cryptography methods employ one key for data encryption and another key for data decryption. As a result, anyone who has access to such a public communication will be unable to decode or read it. This type of cryptography, known as "public-key" encryption, is used in the majority of internet security protocols. The term "asymmetric encryption" is used to describe this type of encryption.



Asymmetric Encryption

How does Encryption work?

When data or information is shared over internet, it passes via a number of global network devices that are a component of the public internet. Data that is transmitted via the open internet leads to the risk of being stolen or hacked by hackers. Users can install particular hardware or software to guarantee the safe transfer of data or information in order to avoid this. In network security, these operations are referred to as encryption. The process of transforming plaintext into ciphertext, is called encryption. A cryptographic key, or a collection of agreed-upon mathematical values, is used during the encryption process. The data is decrypted by the recipient using the key, restoring readable plaintext. Because brute force assaults, which involve testing random numbers until the right combination is known, are less likely to be used by third parties to decrypt data, encryption is more safe the more difficult the cryptographic key. Passwords are also secured by encryption. Hackers cannot decipher your password because password encryption techniques jumble it up.

States of Data Encryption

- **Data encryption in transit:** Information that is actively moving from one point to another, such as via the internet or over a private network, is referred to as data in transit. Data is deemed less safe when in transit due to the weaknesses of transfer techniques.
- **Encryption of data at rest:** Data encryption at rest decreases the risk of data breach caused by lost or stolen devices, inadvertent password sharing, or accidental permission granting by increasing the time it takes to access information and providing the time required to discover data loss, ransomware attacks, remotely erased data, or changed credentials.

How the Data Encryption takes place?

Assume a person possesses a box containing a few documents. The individual looks after the box and secures it with a lock. The individual sends this box of paperwork to his or her pal after a few days. The key is also kept by a buddy. This signifies that both the sender and the recipient have the same key. The buddy has now been given permission to open the box and see the document. The encryption method is the same as we mentioned in the sample. Encryption is performed on digital communications, though. This technological procedure is designed to prevent a third party from deciphering the signal's secret content. Consumers conduct transactions for goods purchases over the internet. There are millions of web services that can help various trained employees do their responsibilities. Furthermore, to utilize these services that demand personal information, most websites require substantial identification. One of the most common ways, known as "encryption," is to keep such information safe and secure.

Plain text *CipherText*

Encryption Process

The security of networks is intimately related to encryption. Encryption is useful for concealing data, information, and things that are incomprehensible to a normal human. Because both encryption and decryption are effective ways of cryptography, which is a scientific procedure for performing secure communication, the encrypted information may be transformed back to its original condition following the decryption process. There are a variety of algorithms for data encryption and decryption. However, "keys" can also be utilized to obtain high-level data security.

Uses of Data Encryption

- Using digital signatures, Encryption is used to prove the integrity and authenticity of the information. Digital-rights management and copy protection both require encryption.
- Encryption can be used to erase data. But since data recovery tools can sometimes recover deleted data, if you encrypt the data first and then throw away the key, the only thing anyone can recover is the ciphertext, not the original data.
- Data Migration is used when transferring data over a network to ensure that no one else on the network can read it.

- VPNs (Virtual Private Networks) uses encryption, and you should encrypt everything you store in the cloud. This can encrypt the entire hard drive as well as voice calls.

Advantages of Data Encryption

1. Data encryption keeps information distinct from the security of the device on which it is stored. Encryption provides security by allowing administrators to store and send data via insecure channels.
2. If the password or key is lost, the user will be unable to open the encrypted file. Using simpler keys in data encryption, on the other hand, makes the data insecure, and anybody may access it at any time.
3. Encryption improves the security of our information.

Disadvantages of Data Encryption

1. If the password or key is lost, the user will be unable to open the encrypted file. Using simpler keys in data encryption, on the other hand, makes the data insecure, and anybody may access it at any time.
2. Data encryption is a valuable data security approach that necessitates a lot of resources, such as data processing, time consumption, and the use of numerous encryption and decryption algorithms. As a result, it is a somewhat costly approach.
3. Data protection solutions might be difficult to utilize when the user layers them for contemporary systems and applications. This might have a negative influence on the device's normal operations.
4. If a company fails to realize any of the restrictions imposed by encryption techniques, it is possible to set arbitrary expectations and requirements that might undermine data encryption protection.

Audit trail

What is an audit trail in DBMS?

An audit trail is a chronological record of all database transactions, including insertions, updates, and deletions. It captures both the old and new values of modified data, as well as metadata such as the user or application responsible for the change, the date and time of the change, and the type of change (e.g., insert, update, delete).

An audit trail can be used to track and monitor database activity, identify and troubleshoot issues, and ensure data integrity and security. For example, if a user accidentally deletes important data from the database, the audit trail can be used to identify the responsible user and restore the deleted

data. Similarly, if data is corrupted or modified in an unauthorized manner, the audit trail can help to identify the cause and take corrective action.

Types of Audit Trails

In a database management system (DBMS), an audit trail is a record of changes made to the database. There are several types of audit trails that can be used to track changes in a DBMS. The three main types of audit trails are internal, external, and IRS (Internal Revenue Service) audit trails.

- **Internal audit trails** − These audit trails are used by organizations to track changes made to their own databases. They are typically used to ensure data integrity, detect and correct errors, and meet regulatory requirements.
  Example − A company might use an internal audit trail to track changes made to its financial records or customer database.
- **External audit trails** − These audit trails are used by external organizations or auditors to review the data in a database. They are often used to verify the accuracy and reliability of the data for regulatory or compliance purposes.
  Example − An external auditor might use an external audit trail to review the financial records of a company for compliance with accounting standards.
- **IRS audit trails** − These audit trails are used by the Internal Revenue Service (IRS) to track changes made to tax records. They are used to ensure the accuracy and integrity of tax information and to detect and prevent tax fraud.
  Example − The IRS might use an IRS audit trail to track changes made to an individual's tax records, such as changes to income or deductions.

Some other important types of audit trails and their examples are mentioned below.

- **Log-based audit trails** − These audit trails use a log file to record changes made to the database. The log file contains information about each change, such as the time the change was made, the user who made the change, and the type of change (e.g., insert, update, delete).
  Example − In a financial database, a log-based audit trail might be used to track changes to account balances or transactions.

- **Trigger-based audit trails** − These audit trails use triggers, which are special types of database objects that are activated when a specific event occurs (e.g., a row is inserted or updated). Triggers can be used to record changes made to the database in an audit table.
  Example − In a healthcare database, a trigger-based audit trail might be used to track changes to patient records, such as changes to medication lists or vital signs.
- **Version-based audit trails** − These audit trails use versioning to track changes made to the database. Each time a change is made to a row in the database, a new version of the row is created with the updated data. The old version of the row is retained, allowing you to view the history of changes made to the row.
  Example − In a project management database, a version-based audit trail might be used to track changes to project tasks, such as changes to due dates or completion status.
- **Shadow tables** − These are tables that are used to store copies of rows as they are updated in the main table. The shadow table contains both the old and new versions of the row, allowing you to see the history of changes made to the row.
  Example − In a customer relationship management (CRM) database, a shadow table might be used to track changes to customer profiles, such as changes to contact information or purchasing history.

Statistical Databases in DBMS

A **Statistical Database (SDB)** is a specialized type of database used to store data that supports statistical analysis. It is designed to provide aggregate data (such as averages, sums, and counts) while often restricting access to individual records to maintain privacy and confidentiality.

Statistical databases are particularly useful in fields like market research, census data analysis, healthcare, and scientific research, where analyzing trends and patterns is more important than accessing detailed individual data.

Types of Statistical Databases

1. **Pure Statistical Databases:**
   Contain only statistical data, such as pre-computed aggregates, for direct use in statistical analysis.
2. **Mixed Statistical Databases:**
   Combine raw data with statistical data, allowing access to individual records but with strict controls to ensure confidentiality.

Components of a Statistical Database

1. **Data Tables:**
   Store the raw or aggregated data.
2. **Metadata:**
   Information about the structure and meaning of the data (e.g., variable names, units of measurement).
3. **Statistical Query Interface:**
   Tools and interfaces to perform statistical operations like summing, averaging, or finding correlations.

Examples of Statistical Database Usage

1. **Census Bureau Data:**

   Provides aggregate population statistics without revealing individual responses.
2. **Healthcare Databases:**

   Analyze patient trends and outcomes while keeping personal health information confidential.
3. **Economic Surveys:**

   Aggregate data on consumer spending, employment rates, and market trends.

Challenges in Statistical Databases

1. **Inference Attacks:**

   Even with restricted access, users might deduce individual data by combining multiple statistical queries.

   Example: Repeatedly querying for average income by filtering subsets of data.
2. **Performance Issues:**

   Aggregating large datasets can be computationally expensive, especially with complex queries.
3. **Balancing Privacy and Usability:**

   Ensuring privacy without overly restricting the utility of the database can be challenging.

Techniques to Ensure Privacy

1. **Data Perturbation:**

   Add noise to the data or results to prevent the exact identification of individual records.
2. **Query Restriction:**

   Limit the number and type of queries a user can execute.

3. **Data Aggregation:**
   Only provide pre-aggregated data, removing access to raw records.

Statistical databases in DBMS are vital tools for analyzing large datasets while ensuring data privacy and confidentiality. They are widely used in research, government, and healthcare applications where statistical insights are more valuable than individual-level data. While they come with challenges like inference attacks and performance optimization, advancements in database technology and privacy-preserving techniques continue to improve their efficiency and security.