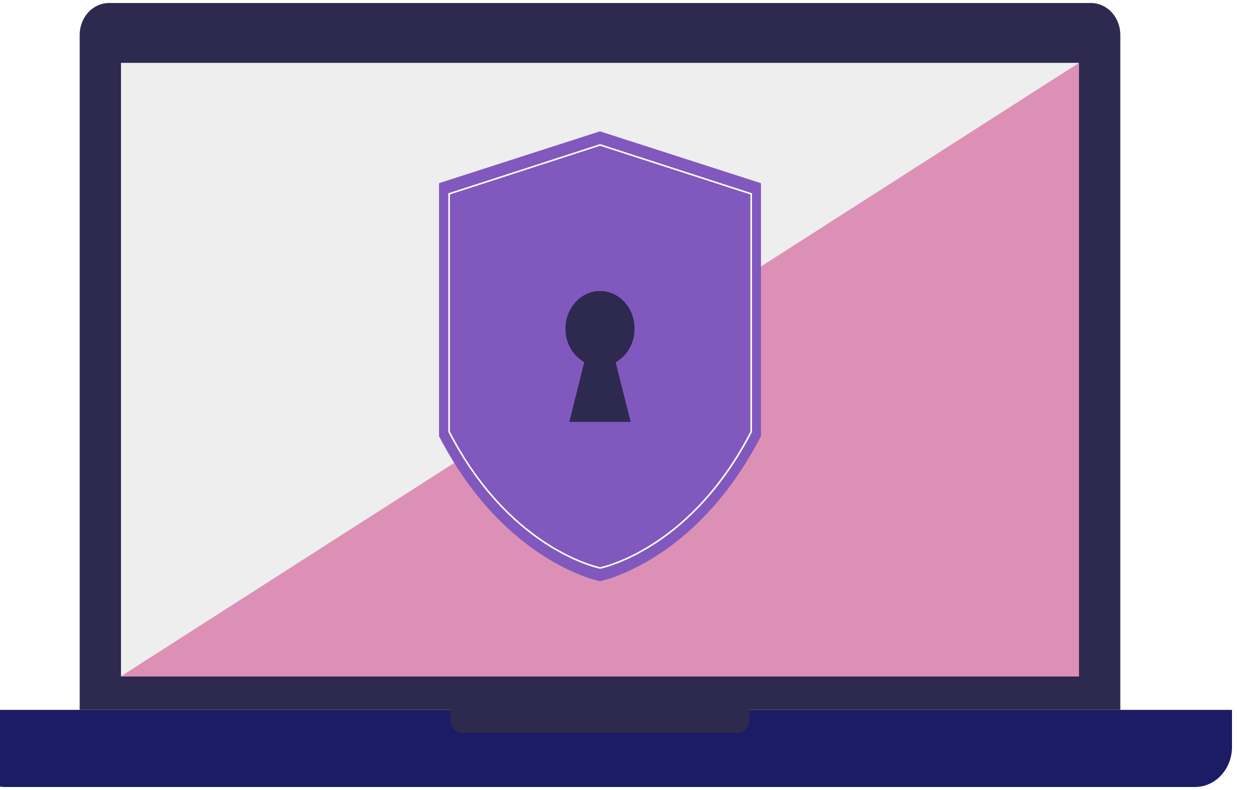
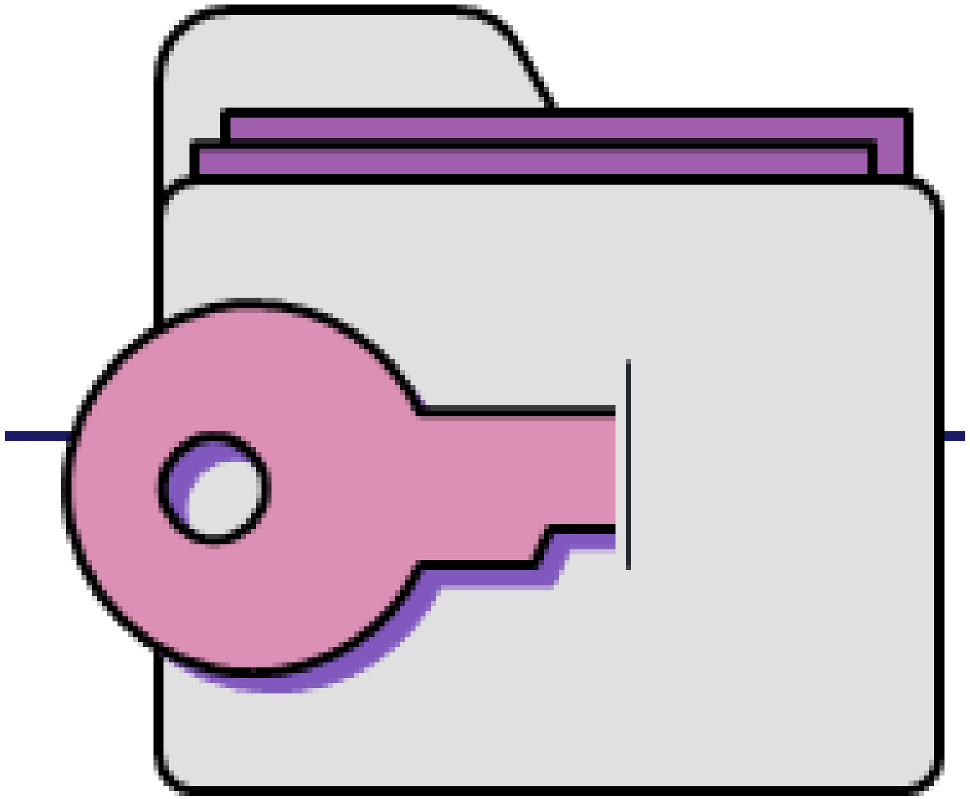


# Computer Security

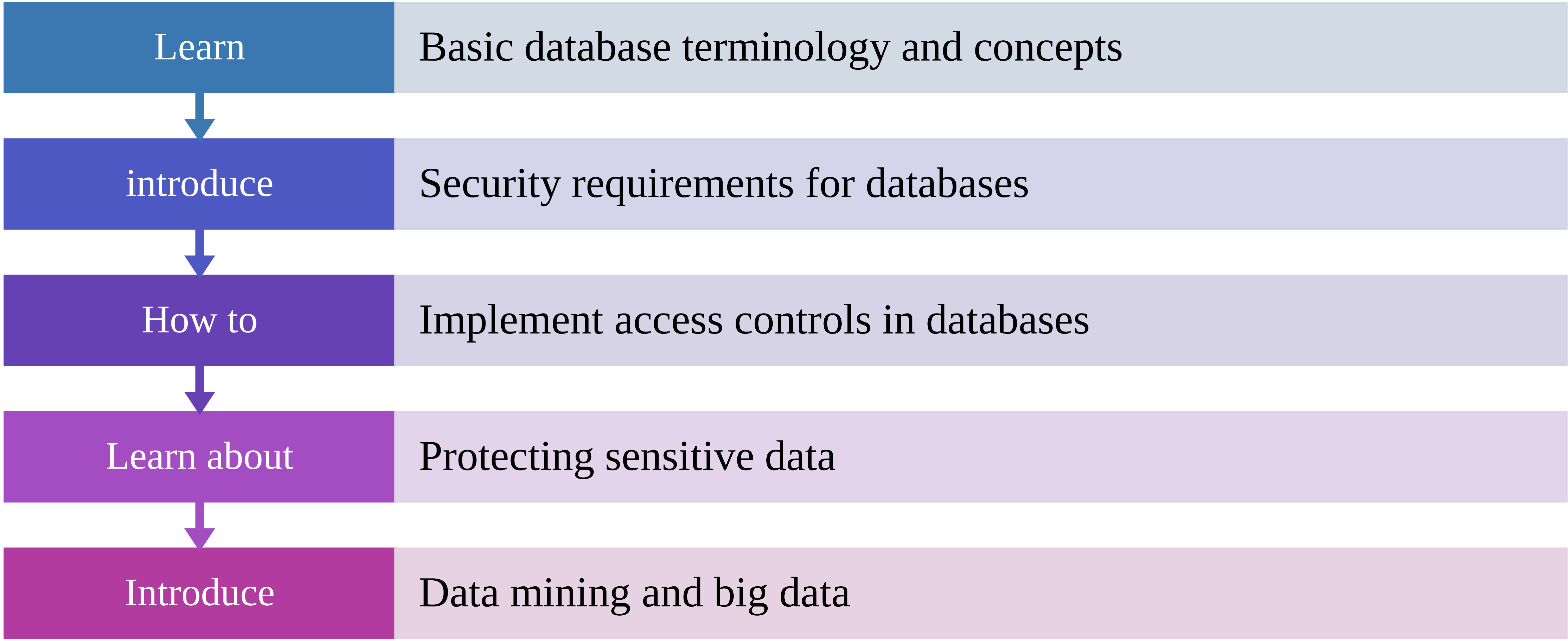
***CS433***



# Chapter 7

Databases

# Objectives



| | |
|---|---|
| Learn | Basic database terminology and concepts |
| introduce | Security requirements for databases |
| How to | Implement access controls in databases |
| Learn about | Protecting sensitive data |
| Introduce | Data mining and big data |

# Database Terms

**Database:** A collection of data and a set of rules that organize the data by specifying certain relationships among the data.

- ✓ **Logical format vs. Physical format**

**Database administrator:** Person who defines the rules that organize the data and controls who should have access to what parts of the data.

**Database management system:** The system through which users interact with the database (front end).

# Database Terms

***Components of Databases***

**Record:** One related group of data.

**Field/element:** Elementary data items that make up a record (e.g., name, address, city).

**Schema:** Logical structure of a database.

**Subschema:** The portion of a database a given user has access to.

**Attribute:** A column in a database.

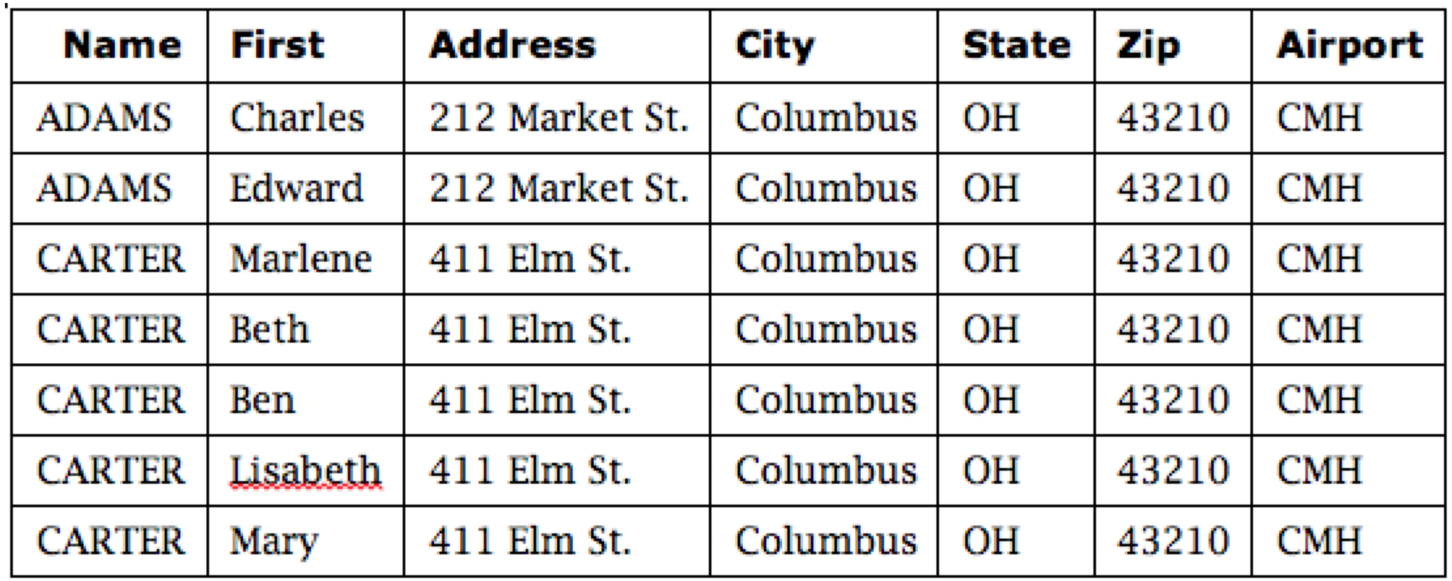
**Relation:** A set of database columns.

FROM *SECURITY IN COMPUTING, FIFTH EDITION*, BY CHARLES P. PFLEEGER, ET AL. (ISBN: 9780134085043).

# Database Terms

**Query:** A command that tells the database to retrieve, modify, add, or delete a field or record.

- ✓ Users extract data through use of queries
- ✓ The most common database query language is SQL
- ✓ Example. SELECT ZIP=‘43210’



| Name | First | Address | City | State | Zip | Airport |
|---|---|---|---|---|---|---|
| ADAMS | Charles | 212 Market St. | Columbus | OH | 43210 | CMH |
| ADAMS | Edward | 212 Market St. | Columbus | OH | 43210 | CMH |
| CARTER | Marlene | 411 Elm St. | Columbus | OH | 43210 | CMH |
| CARTER | Beth | 411 Elm St. | Columbus | OH | 43210 | CMH |
| CARTER | Ben | 411 Elm St. | Columbus | OH | 43210 | CMH |
| CARTER | Lisabeth | 411 Elm St. | Columbus | OH | 43210 | CMH |
| CARTER | Mary | 411 Elm St. | Columbus | OH | 43210 | CMH |

Subschema

# DB Security Requirements

## 1. Physical database integrity

- ✓ Immunity to physical catastrophe, such as power failures, media failure
  - physically securing hardware
  - regular backups

## 2. Logical database integrity

- ✓ The logical structure is preserved
  - a modification to the value of one field does not affect other fields

## 3. Element integrity

- ✓ The data of each element are accurate.

# DB Security Requirements

## 4. Auditability

- ✓ Tracking who or what has accessed/modified the db.

## 5. Access Control

- ✓ A user is allowed to access only authorized data.
- ✓ Different users can be restricted to different access modes.

## 6. User authentication

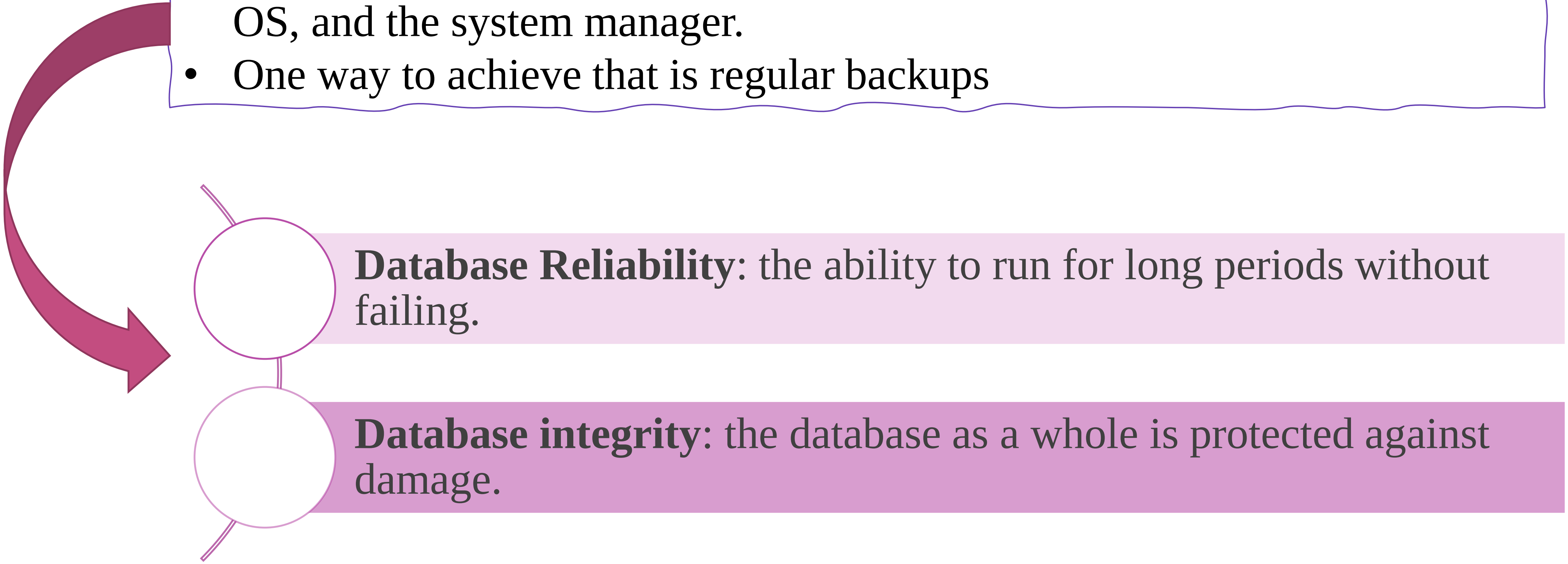
- ✓ Users are positively identified.

## 7. Availability

- ✓ Users can access the database in general and all the data for which they are authorized.

# DB Reliability & Integrity

- Assuring the *integrity of the db* is the responsibility of the DBMS, the OS, and the system manager.
- One way to achieve that is regular backups



**Database Reliability**: the ability to run for long periods without failing.

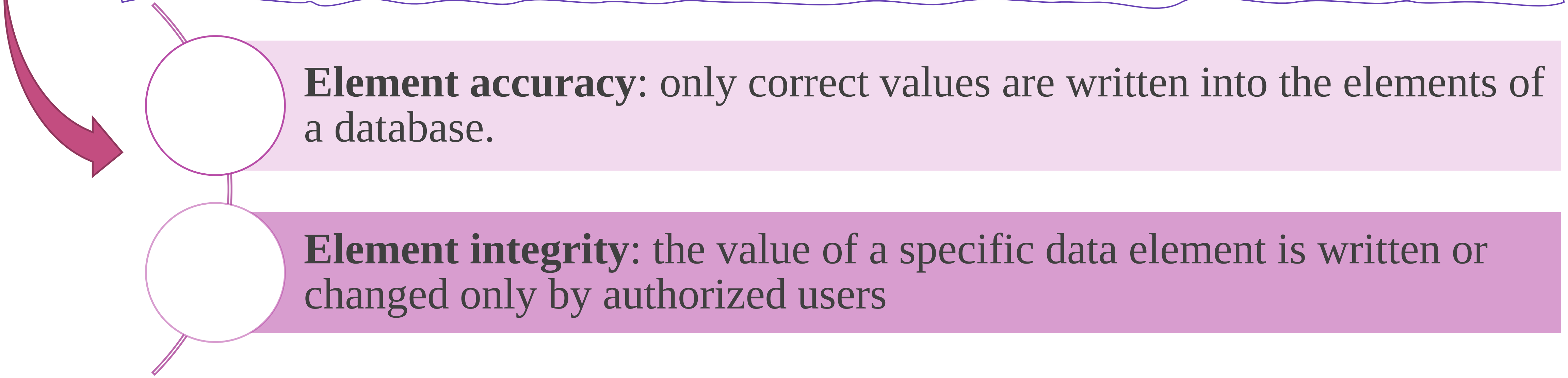
**Database integrity**: the database as a whole is protected against damage.

# DB Reliability & Integrity

Assuring the *integrity of data* integrity is the responsibility of the DBMS and authorized entities.

**Achieved through** field checks; access control and change logs.

DBMS implement their own access control at a level finer than what an OS handles.



**Element accuracy**: only correct values are written into the elements of a database.

**Element integrity**: the value of a specific data element is written or changed only by authorized users

# DB Reliability & Integrity

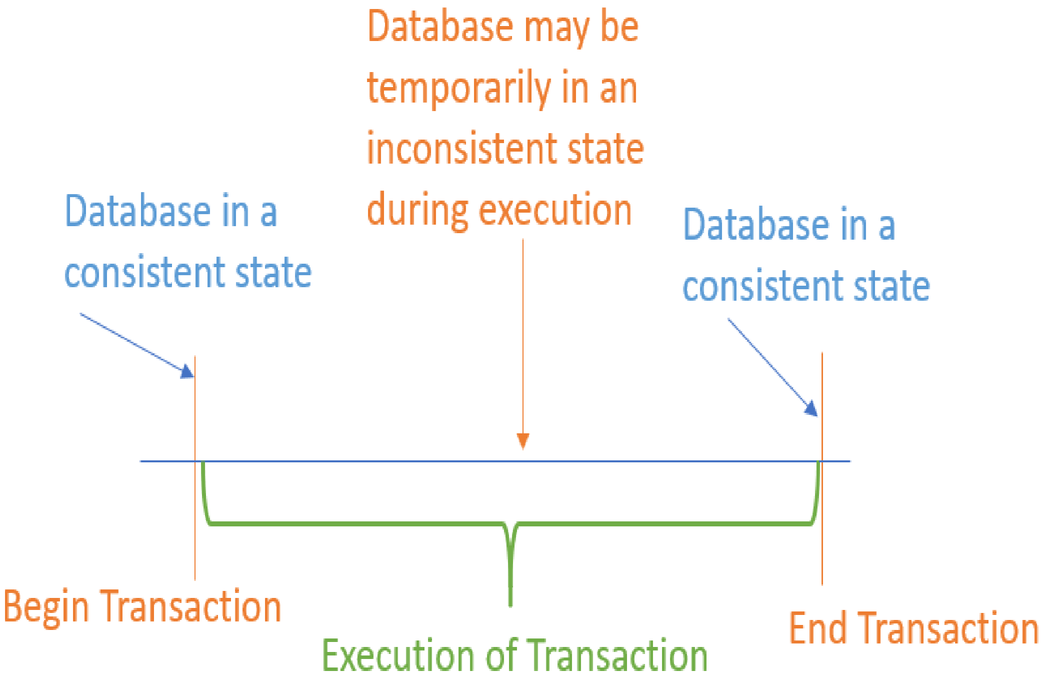
## Two-Phase Update

- ✓ A solution to DB system failure in the middle of an update.

### ***Phase 1: Intent***

- ✓ DBMS does everything it can, other than making changes to the database, to prepare for the update.
  - Collects records, opens files, locks out users, makes calculations.
- ✓ DBMS commits by writing a commit flag to the database.

If the system fails during execution of the first phase, no harm is done because all these steps can be restarted and repeated after the system resumes processing

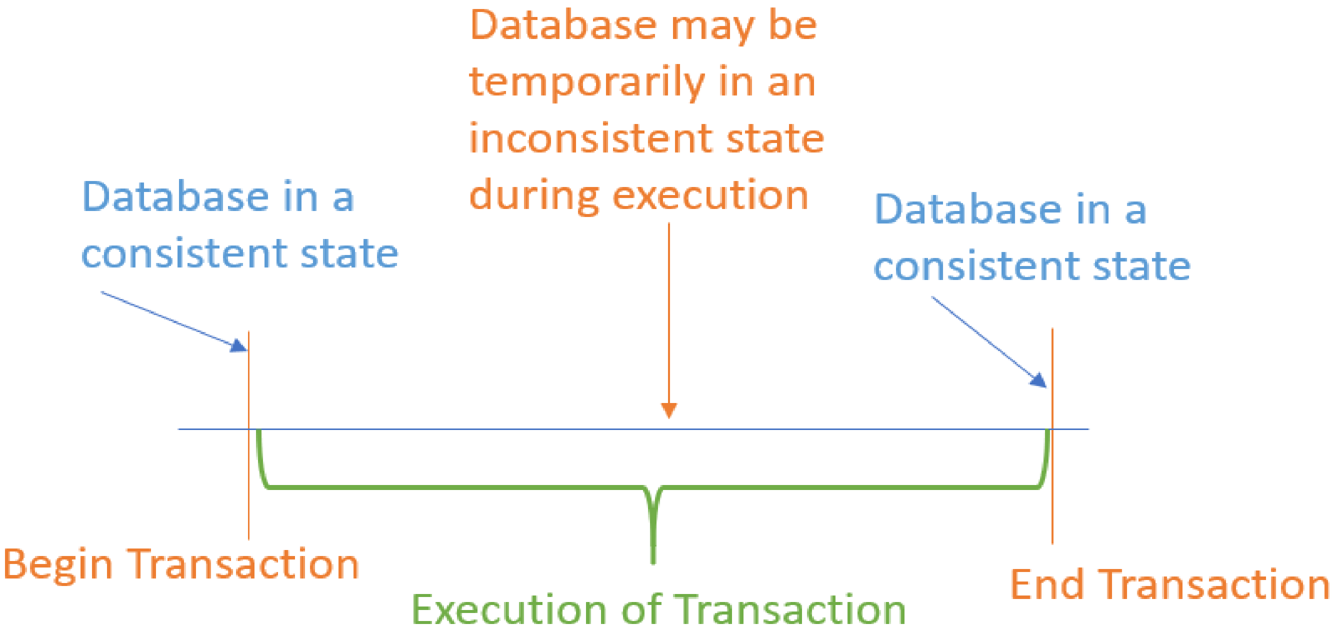

Database may be temporarily in an inconsistent state during execution
Database in a consistent state
Database in a consistent state
Begin Transaction
Execution of Transaction
End Transaction

# DB Reliability & Integrity

***Phase 2: Write***

- ✓ DBMS completes all write operations
- ✓ DBMS removes the commit flag
- ✓ If the DBMS fails during either phase 1 or phase 2, it can be restarted and repeat that phase without causing harm

If the system fails during the second phase, the database may contain incomplete data, but the system can repair these data by performing all activities of the second phase.


Database may be temporarily in an inconsistent state during execution
Database in a consistent state
Database in a consistent state
Begin Transaction
Execution of Transaction
End Transaction

# Other DB Security Concerns

- **DBMS serves multiple users at once**
  - Should maintain data consistency
    - DBMS use locks and atomic operations to maintain consistency
      - Writes are treated as atomic operations.
      - Records are locked during write so they cannot be read in a partially updated state.

- **Error detection and correction codes to protect data integrity**
  - Applied to single fields, records, or the entire database
    - point out the place of the error; or
    - indicate what the correct value should be

- **For recovery purposes, a database can maintain a change log**
  - In the event of a failure, the database is reloaded from a backup copy and all later changes are then applied from the audit log.

# Database Disclosure

**Sensitive Data** is data that should not be made public.

**Sensitive Data can be:**

- ✓ **Inherently sensitive**: Passwords, locations of weapons
- ✓ **From a sensitive source**: Confidential informant
- ✓ **Declared sensitive**: Classified document, name of an anonymous donor
- ✓ **Part of a sensitive attribute or record**: Salary attribute in an employment database
- ✓ **Sensitive in relation to previously disclosed information**: An encrypted file combined with the password to open it

# Types of DB Disclosures

***Sensitive data should not be disclosed or anything about it!***

**Exact data.** Sensitive data values are requested

```
SELECT Income FROM Payroll
WHERE Name = Omar Ali
```

**Bounds.** Revealing that a value falls between two values

```
SELECT Name FROM Payroll
WHERE Income BETWEEN 20000 AND
80000
```

**Existence.** Indicating that sensitive data, regardless of the actual value, is present

```
SELECT Aid FROM Payroll
```

# Types of DB Disclosures

**Negative result.** Use a query to determine a negative result

```
SELECT Name FROM Payroll
WHERE Qualification = PhD
```

**Probable value.** Determine the probability that a certain element has a certain value

```
COUNT Age = 42 & Gender = Male
& Qualification = Primary
COUNT Age = 42 & Gender = Male
& Qualification = Primary
& PreviousEmployer = ABC
```

**Direct inference.** Deriving sensitive data from nonsensitive data

```
SELECT Name FROM Payroll
WHERE (Income > 5000) OR
 (Gender = Male AND Gender = Female)
```

# Types of DB Disclosures

**Inference by arithmetic.** Inferring a final result based on one or more intermediate statistical results

```
SELECT sum(Income) FROM payroll
WHERE Name <> Omar Ali
SELECT sum(Income) FROM payroll
```

**Aggregation**. Building sensitive results from less sensitive inputs

Previous Knowledge: F and M have same qualification, same job,  and were hired the same date 10 years ago; ABC claims equal pay!

**Hidden data attributes.** Attributes that add meaning to the data

- File tags
- Geotags

```
COUNT Service = 10 & Gender = Male
& Qualification = Primary
& Income<2000
COUNT Service = 10 & Gender = Female &
Qualification = Primary
& Income<2000
```

# Preventing Disclosure

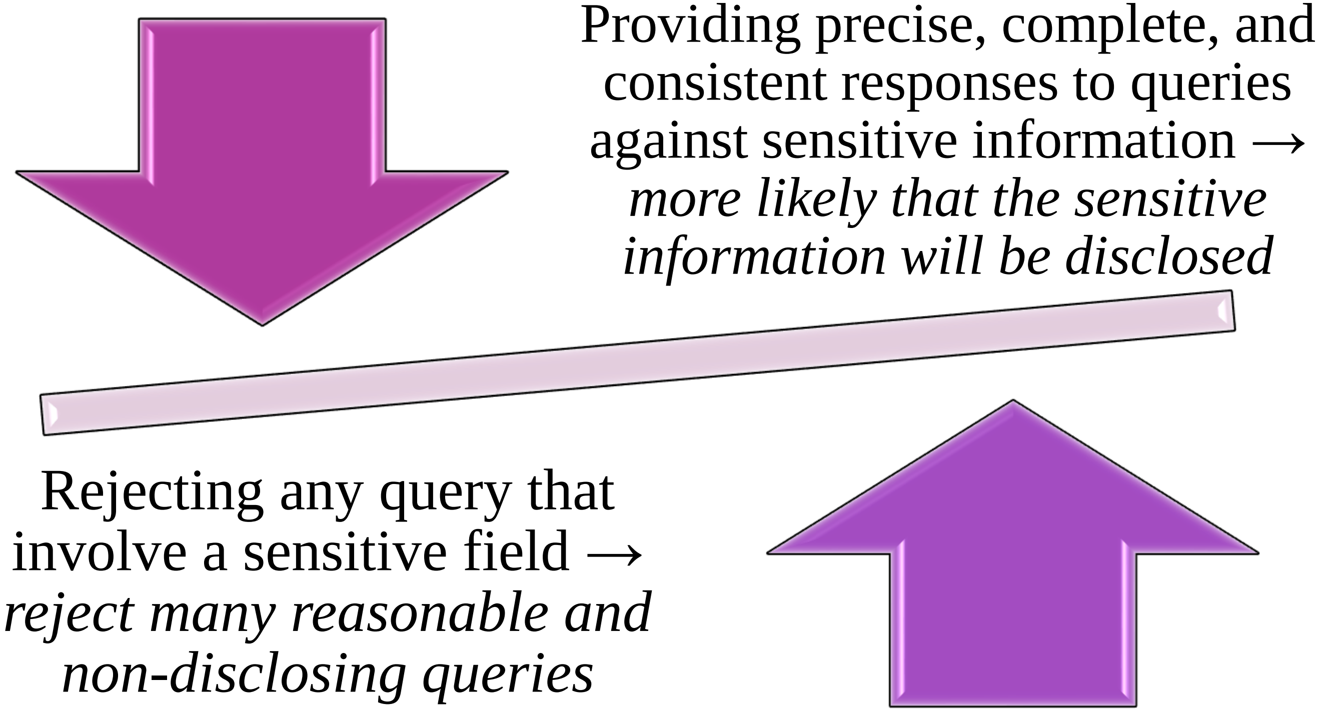
**Using the following approaches:**

1. Suppress obviously sensitive information
   - Block the release of sensitive data by limiting the queries accepted- easily applied.

2. Keep track of what each user knows
   - Based on past queries, limit data provided in response to a query- extremely costly.

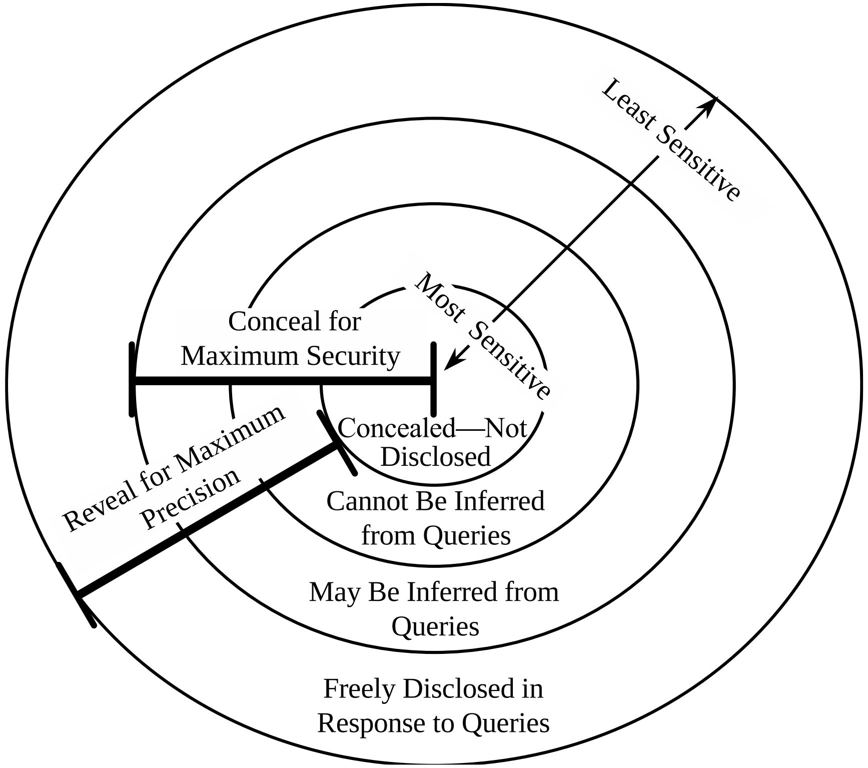
3. Disguise the data
   - Applies only to released data- provide slightly incorrect or possibly inconsistent results.

# Security vs. Precision



Providing precise, complete, and consistent responses to queries against sensitive information → *more likely that the sensitive information will be disclosed*

Rejecting any query that involve a sensitive field → *reject many reasonable and non-disclosing queries*


Least Sensitive
Most Sensitive
Conceal for Maximum Security
Reveal for Maximum Precision
Concealed—Not Disclosed
Cannot Be Inferred from Queries
May Be Inferred from Queries
Freely Disclosed in Response to Queries

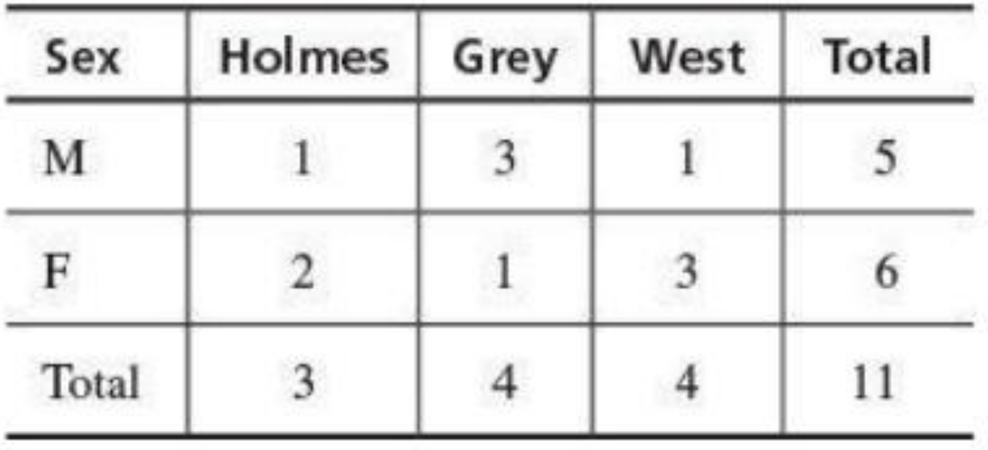
The ideal combination of security and precision allows maintaining perfect confidentiality with maximum precision. Disclosure prevention must be balanced against the database requirements, as the loss of precision and completeness may make the database unusable.

# Suppression Techniques

## Statistical Suppression

1. **Limited response suppression**
   - Eliminates certain low-frequency elements from being displayed.
   - It is not sufficient to delete them, however, if their values can also be inferred.

Example: The data in this table suggest that the cells with counts of 1 should be suppressed

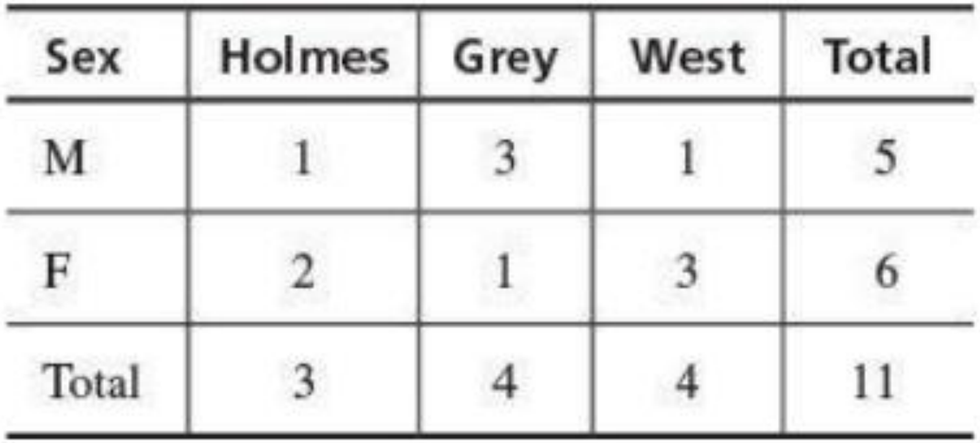


| Sex | Holmes | Grey | West | Total |
|---|---|---|---|---|
| M | 1 | 3 | 1 | 5 |
| F | 2 | 1 | 3 | 6 |
| Total | 3 | 4 | 4 | 11 |

| Sex | Holmes | Grey | West | Total |
|---|---|---|---|---|
| M | | | | 5 |
| F | | | | 6 |
| Total | 3 | 4 | 4 | 11 |

# Suppression Techniques

2. **Combined results**
   - Ranges, rounding, sums, averages.



| Sex | Holmes | Grey | West | Total |
|---|---|---|---|---|
| M | 1 | 3 | 1 | 5 |
| F | 2 | 1 | 3 | 6 |
| Total | 3 | 4 | 4 | 11 |

| Sex | | |
|---|---|---|
| M | | |
| F | | |

# Suppression Techniques

3. **Random sample**
   - Result is not derived from the whole database, instead computed on a random sample of the database.
     - The sample chosen is large enough to be valid.
     - The same sample set should be chosen for equivalent queries.

# Suppression Techniques

4. **Blocking small sample sizes**
   - Block results when a small number of people make up a large proportion of a category.
     - “n items over k percent”

5. **Random data perturbation**
   - Randomly add or subtract a small error value to/from actual values.

6. **Swapping**
   - Randomly swapping values for individual records while keeping statistical results the same.

# Data Mining

- ✓ Closely related to the concept of big data
  - Collection of massive amounts of data.
- ✓ Data mining uses statistics, machine learning, mathematical models, pattern recognition, and other techniques to discover patterns and relations on large datasets.
- ✓ The size and value of the datasets present an important security and privacy challenge, as the consequences of disclosure are naturally high.

# Data Mining Challenges

- ✓ **Correcting mistakes in data**
  - data is often moved to more databases before the original database can be corrected.
- ✓ **Preserving privacy**
  - inference works on big data just as it does in databases.
- ✓ **Granular access control**
  - uses unstructured datasets, flat, two-dimensional tables; access control is imposed at the file level.
- ✓ **Secure data storage**
  - data may be collected globally and through cloud providers.
- ✓ **Transaction logs**
  - tracking access is expensive, especially if accesses are numerous; detailed access auditing is uncommon for big data.
- ✓ **Real-time security monitoring**
  - real-time security monitoring is not intended for complex, shared, fluid network architectures.

# Summary

- ✓ Database security requirements include:
  - Physical integrity
  - Logical integrity
  - Element integrity
  - Auditability
  - Access control
  - User authentication
  - Availability
- ✓ There are many subtle ways for sensitive data to be inadvertently disclosed, and there is no single answer for prevention
- ✓ Data mining and big data have numerous open security and privacy challenges