SECURITY IN COMPUTING, FIFTH EDITION

Chapter 7: Databases

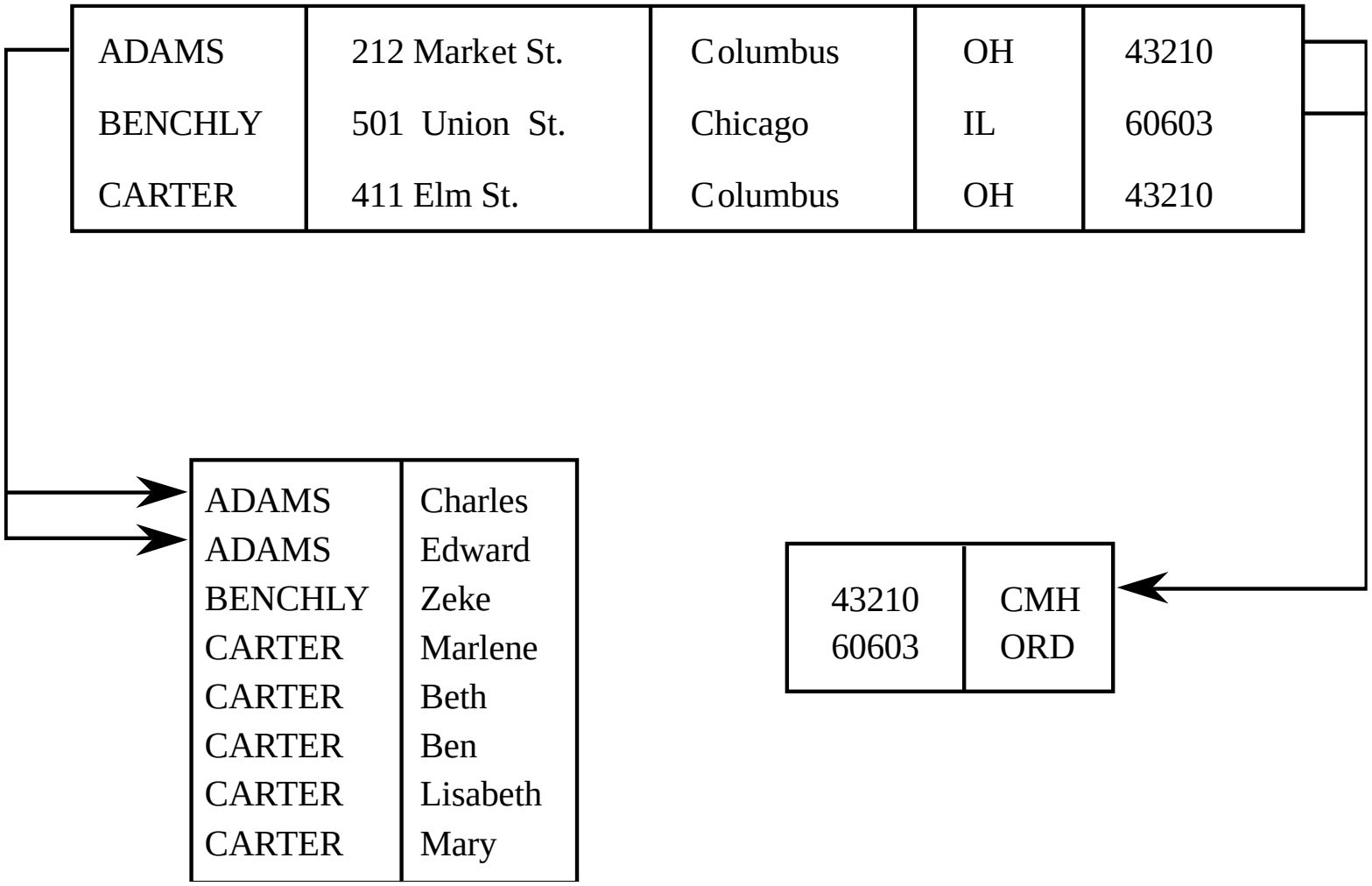
Objectives for Chapter 7

- Basic database terminology and concepts
- Security requirements for databases
- Implementing access controls in databases
- Protecting sensitive data
- Data mining and big data

Database Terms

- Database administrator
- Database management system (DBMS)
- Record
- Field/element
- Schema
- Subschema
- Attribute
- Relation

Database Example



Schema Example

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
BENCHLY	Zeke	501 Union St.	Chicago	IL	60603	ORD
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

Queries

- A query is a command that tells the database to retrieve, modify, add, or delete a field or record
- The most common database query language is SQL

Example SQL Query

- `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

Database Security Requirements

- Physical integrity
- Logical integrity
- Element integrity
- Auditability
- Access control
- User authentication
- Availability

Reliability and Integrity

- Reliability: in the context of databases, reliability is the ability to run for long periods without failing
- Database integrity: concern that the database as a whole is protected against damage
- Element integrity: concern that the value of a specific data element is written or changed only by authorized users
- Element accuracy: concern that only correct values are written into the elements of a database

Two-Phase 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
- 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

Other Database Security Concerns

- Error detection and correction codes to protect data integrity
- For recovery purposes, a database can maintain a change log, allowing it to repeat changes as necessary when recovering from failure
- Databases 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

Sensitive Data

- 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

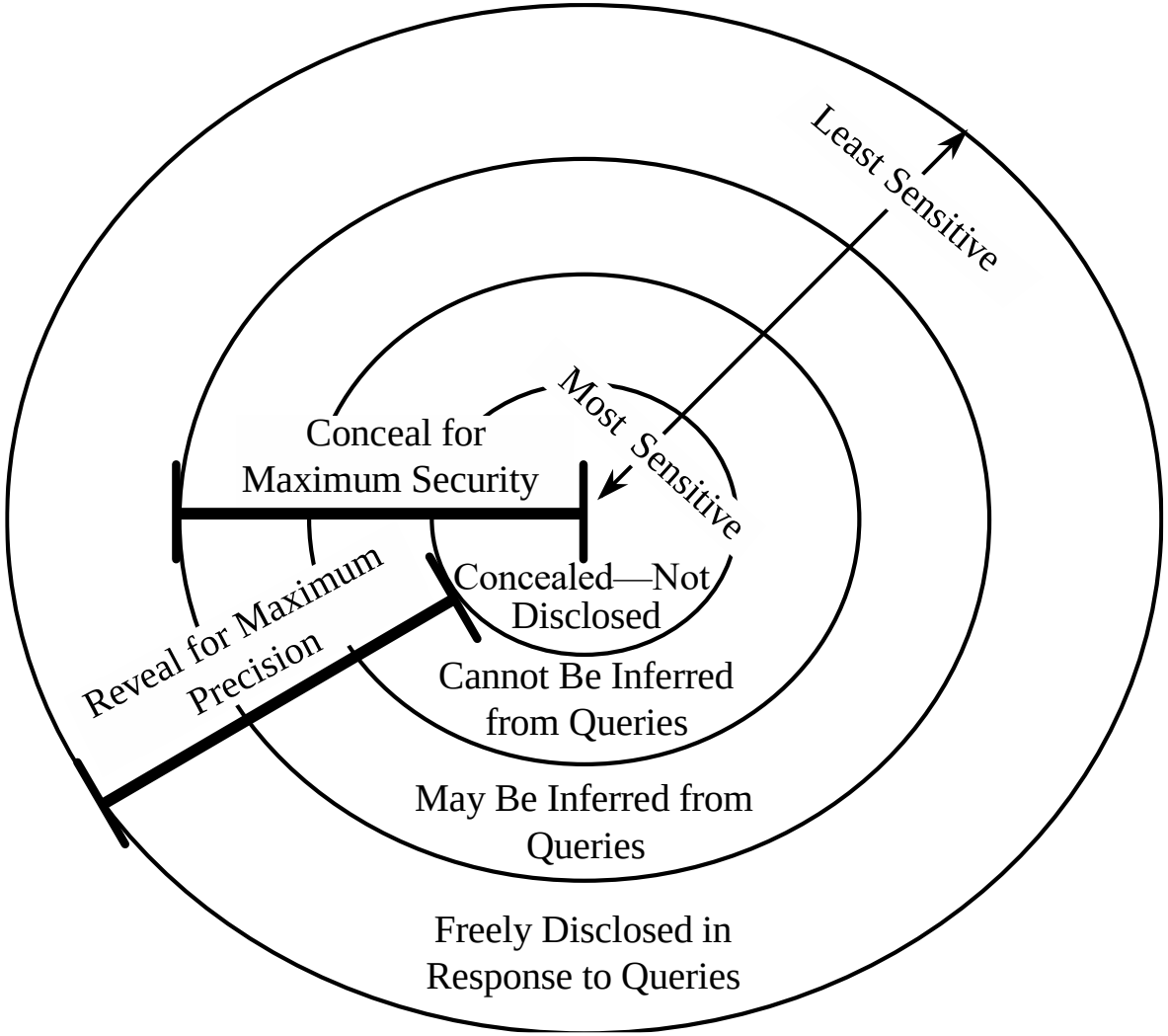
Types of Disclosures

- Exact data
- Bounds
- Negative result
- Existence
- Probable value
- Direct inference
- Inference by arithmetic
- Aggregation
- Hidden data attributes
  - File tags
  - Geotags

Preventing Disclosure

- Suppress obviously sensitive information
- Keep track of what each user knows based on past queries
- Disguise the data

Security vs. Precision



Suppression Techniques

- Limited response suppression
  - Eliminates certain low-frequency elements from being displayed
- Combined results
  - Ranges, rounding, sums, averages
- Random sample
- Blocking small sample sizes
- Random data perturbation
  - Randomly add or subtract a small error value to/from actual values
- Swapping
  - Randomly swapping values for individual records while keeping statistical results the same

Data Mining

- 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
- Preserving privacy
- Granular access control
- Secure data storage
- Transaction logs
- Real-time security monitoring

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