3. Database management

1. Databases and tables
2. Searching and summarizing data
3. Database design
4. DB management in business environments

Inquiry

• What would ordinary users benefit from knowing about databases?
• What are the accepted principles for organizing data?
3. Database management

Topic objective

Describe the querying and design of databases in business environments

Essential and priority objectives

3.0a Recall basic database concepts*  
3.1 Create a database table*†  
3.2a Design a database query*  
3.2b Explain basic database concepts**  
3.3 Design a relational database*
1. Databases and tables

- How do you organize your paper data?
- What’s a *table*?

Subtopic objective

3.1 Create a database table*†
3. Database management

Why study this?

• Consistency and non-redundancy of data require more than a list format
• Certain information is easiest to manage in tables
• End users retrieve information from these tables using queries
• Today DB experts and users also design tables
• Use of general-purpose database management systems (DBMSs) is growing

Database management

• Supports retrievable arrangement of data
• Examples: library catalog, student records, many business records
• Want to separate the database management software from the data design, enabling data design on site
• Likewise query design is possible on site
• Effective database and query design requires analytical thinking: breaking a problem apart
# Tables

- *Table:* stores data about instances of entities
- A table has *attributes* (columns) and *instances* (rows after first)
- One instance of an entity is called a *tuple*, *record*, or *row*
- Example: A *student* has an *ID*, a *name*, and a *Year-of-graduation*
- Simple DB management can be done with MS Word tables, Excel spreadsheets

---

# Tables

- Store information for lookup
- Database tables (relations) include *metadata* (headers), may be empty
- Entities have *attributes* (fields), which have names, types, values
- *Example:* student table (*name, ID, email, major*)
- Row representing instance of entity is a *tuple* (record)
- *Sets* of tuples are unordered, by definition
- *Primary key* is unique identifier of a tuple
A relation is a set of tuples

<table>
<thead>
<tr>
<th>DB term</th>
<th>Math term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table, class, entity</td>
<td>relation</td>
</tr>
<tr>
<td>Record, row, object</td>
<td>tuple</td>
</tr>
<tr>
<td>Column, attribute</td>
<td></td>
</tr>
</tbody>
</table>

**Students**

<table>
<thead>
<tr>
<th>stu-ID</th>
<th>name</th>
<th>major</th>
</tr>
</thead>
<tbody>
<tr>
<td>1234</td>
<td>shaw,d</td>
<td>bus</td>
</tr>
<tr>
<td>4312</td>
<td>smith,d</td>
<td>psych</td>
</tr>
<tr>
<td>5678</td>
<td>cox,t</td>
<td>cs</td>
</tr>
</tbody>
</table>

Database views

- **View**: A logical (vs. physical) table, created by an operation on a table
- **Views** are a way to select data for use
- Some views are a window into a table
- **Selection**: specification of a set of rows chosen by some criterion (e.g., all contacts with salary over $55,000)
- **Projection**: specification of a set of columns (e.g., name and salary)
Other operations on tables

- **Union**: Combines two compatible tables
  (*Example*: all students in either of two sections of a course)
- **Difference** (*Example*: Students who are *not* CS majors)
- **Product**: Creates combined table
  (*Example*: students \( \times \) courses), used to create smaller *join* relation
- **Join**: A *product*, restricted to matching values for common fields

---

**Example table: List of professional contacts**

- Example file: *contacts.xlsx*
- Schema: (*Name*, *ID*, *Salary*, *Title*, *Company*)

<table>
<thead>
<tr>
<th>Name</th>
<th>ID</th>
<th>Salary</th>
<th>Title</th>
<th>Company</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smith</td>
<td>2345</td>
<td>70000</td>
<td>Trainer</td>
<td>Xyz corp</td>
</tr>
</tbody>
</table>

---
2. Searching and summarizing data

• How do you look up a telephone number?
• Could excel store a database?
• How do you size up a company you might work for or buy from?

Subtopic objectives

3.2a Design a database query*
3.2b Explain basic database concepts**
Accessing data

- **Query** (filter): generate selected view
- **Subtotal**: sum of numeric values in a field of a view
- **Sorting**: any table may be sorted on any attribute
- **Pivot tables**: show correlations between independent and dependent variables
- **Correlation** establishes patterns in data

Queries

- **Query**: specification of a view
- **Query languages**
  - Query by example (enter value in a field)
  - Structured Query Language (SQL)
- **Examples**:
  - Select all student records where major is Business Administration
  - Select all employees where salary > 25000 and salary < 50000
  - Select name, salary from employees
Querying (searching) in Excel

- In Excel, *filters* enable selection of rows of a table according to values of attributes
- In Excel, select column headings; then *Data / Filter* places tabs at tops of columns
- *Example*: with *listings.xls*, to select all home listings with more than 2 bedrooms:
  - Click triangle beside “BDRM”
  - Choose “Custom”; “Is greater than”; “2”
- Multiple conditions may be specified using “AND” and “OR” operators

Sorting in Excel and MS Word

- Columns of Excel spreadsheets and of tables, or lists of paragraphs, in Word documents may be sorted
- Sorts may be *ascending* or *descending*
- *Word*: use “A→Z” button
- *Excel*: use *Data* tab
- Sorting enables faster searching
Data correlations

- We look for patterns to understand our environment
- Example: Does success correlate with effort? Luck? IQ? Family wealth?
- The corre function in Excel supports finding correlations among sets of data stored in spreadsheet columns
- Example: What is the effect of location and acreage on home price?

Subtotals

- In Excel, Data / Subtotals permits automatic insertion of subtotal information in a view; e.g., total salary for EMC employees
- Use of subtotals and pivot tables (see next two slides) enables higher-level data analysis than use of cell data alone
Summarizing data

- A *pivot table* in Excel enables computation of functions, e.g., *sum*, on all instances of a given field of all records that match by two criteria
- *Example*: total salaries for all TJX vice presidents

Pivot tables

- The columns used at the edges of the pivot table are the independent variables
- The column used in the body of the table is the dependent variable
- The pivot table shows how the dependent variable is affected by the values of the independent variables
3. Database design

• How do we learn to understand the world?
• What’s a *thing*?
• A *category* of things?
• Can the world be represented by mathematical *sets*?

<table>
<thead>
<tr>
<th>Employee</th>
</tr>
</thead>
<tbody>
<tr>
<td>name</td>
</tr>
<tr>
<td>ID</td>
</tr>
<tr>
<td>salary</td>
</tr>
<tr>
<td>display()</td>
</tr>
<tr>
<td>input()</td>
</tr>
<tr>
<td>calc-paycheck()</td>
</tr>
</tbody>
</table>

Subtopic objective

3.3 Design a relational database*
Database integrity and design

• An entity (aka class) is a category of instances (objects) – things, people, places, transactions
• Integrity principle: No two records may represent the same real-world object
• This principle is enforced using primary keys and using a method called entity-relationship design
• A more general integrity principle is that any fact may be represented only once in a database

Entity-relationship design

• The database design process may include consideration of relationships among entities stored as tables
• Example: Where Student and Course are entities, the real-world relationship “Student registers for Course” may guide design
• In that case, a course-registration table may be designed to store records whose attributes might include Student ID, Course ID, and Date
Entity example: Course registration

- A student has an ID, a name, and a Year-of-grad
- Primary key is ID, created to be unique
- A course section has a course-ID, section-ID, name, and instructor
- Entity-relationship diagram:

```
 Student ----> Registers for ----> Course section
```

- A Registrations table could embody the relationship “registers for” between the Students and Course-sections entities

Joining tables

- Join: A view that combines data from two or more tables where records have coinciding field values (e.g. course number)
- Example:
  - A course table lists course numbers, course titles
  - A course-registration table lists student IDs and course numbers of the courses the student registered for
  - A transcript will access both tables
**Join views**

- Suppose a business maintains tables of *customers, products for sale,* and *transaction details* (instances of orders of a single item)
- Each transaction detail record includes customer ID and transaction detail identification, e.g., date, item #, qty., price
- To show all data about a single customer order, a *join* is performed, yielding the customer information plus all detail data for that transaction (i.e., a bill of sale)

**Information integrity in DBs**

- *Rule:* A *fact* should appear only *once* in a database
- *Example:* Non-duplication assured by use of *primary key* prevents anomalies such as deleting only one record and leaving others for the same supposedly deleted entity
- *Example:* A student’s address should appear with the student’s record, not in every record of a student registering for a course. What if address changes?
3. Database management

Integrity enforcement guidelines

- Dependencies are reflected in design.
- *Example:* If Student name in a course registration depends on student ID, the course registration record contains the course ID but not the student name.
- Many-to-one and many-to-many relationships are implemented using *key* references to participants in relationship.

4. Database management in business environments

- How do businesses store information?
- How much data do businesses need?
- Who does it belong to?
### Subtopic objective

3.4a Group and summarize data to show patterns†

3.4b Explain applications of databases

---

### DB management systems

- Excel works for simple tables but is not a full-featured DBMS
- A *database management system*, such as Access or Oracle, is software designed for a wide range of database operations
- *Web-based database access* uses *middleware* to make database information available to the server to support responses to client requests
- Data for one DB may be stored at distributed locations
Other database tools

- *Web pages* may access and even update database files for customer service
- *Forms* enable easy entry of data into a relation
- *Reports* format DB information for distribution
- *Graphical design tools* enable study of linked relationships among tables

Database file formats

- *Specialized* (e.g., email address books)
- *Tables* in word-processor or spreadsheet documents
- General-purpose DBMS software (Access, Oracle, Ingres)
- XML (Extended Markup Language), which uses HTML-like tags to denote field names, as *semantic* information
Data analysis

- **Decision support systems** attempt to predict future sales, profits, costs, based on current data and alternative scenarios.
- **Data mining** uses statistical methods to find unseen relationships in data.
- **Data warehouses** collect data from multiple sources for analysis.
- **Data analysis** is used to profile individuals for market and other purposes.

Data warehouses

- **Data warehouse**: large-scale archival storage of all or a business’s data.
- **Sources**: internal, external, sources, customers, web visitors.
- **Scope**: enterprise.
- **Data marts** organize access to data by department.
IT-enabled business practices

- Business Process Reengineering
- Customer Relationship Management (CRM)
- Enterprise Application Integration
- Electronic Data Interchange (EDI)
- Enterprise Resource Planning
- Just-in-Time Manufacturing
- Manufacturing Resource Planning
- Total Quality Management (TQM)

ISs for management of business

- *Office support systems*: day-to-day task and communication support; e.g., MS Office
  - *Transaction processing systems*: e.g., course registrations, purchases
    - All use databases
  - *Management information systems*: detail, summary, and exception reports for managers
  - *Decision support systems*: work with models and knowledge bases to provide basis for decisions
  - Enterprise resource planning systems tie together multiple data sources and business processes
Information systems and organizations

- **Organization**: People working together to accomplish a mission
- **Mission**: an ongoing set of services accessed at many points
- Many organizations are structured hierarchies
- **Management planning**: strategic, tactical, operational
- **Information system**: A system that uses computers and operates within an organization
- Networks are essential for communication

Components of an IS

- **Transaction processing system**
  - Transaction: an atomic interaction
  - Many transactions today occur online
- **Management information system** (MIS): groups and summarizes data
- **Decision support system**: enables managers to design models of situations, explore alternatives
- **Expert system**: rule or knowledge bases enabling recommendations or diagnosis
Data mining and Big Data

- Gathers and analyzes very large amounts of data
- **Chief goal:** to analyze customer behavior and attributes
- Analysis and gathering methods:
  - Classification (e.g., good/bad risk)
  - Estimation (e.g., residence patterns of low-risk persons)
  - Association rules (e.g., products bought together)
  - Clustering, description, visualization

Recent developments

- **Big Data** is the collection of massive information about individual transactions
- **Medical records** are becoming electronic and web based
- **Privacy issues** are raised by these developments
- **Cloud storage** means that business data is stored on web servers not owned by the businesses
**Metadata in database management**

- Database designs
- Queries

**Users co-create their computing environments**

- *Database queries* are small, reusable programs
- *DB tables* are larger reusable computing environments
C. Breuning. Handout on using database features of Excel.
