Normalization

- A theoretical foundation for the relational model
- Application of a series of rules that gradually improve the design
  - Minimize redundancy
  - Minimize dependency
- Objectives*
  - Free the collection of relations from undesirable insertion, update and deletion dependencies
  - Isolate data so that additions, deletions, and modifications of a field can be made in just one table and then propagated through the rest of the database

* Wikipedia
Functional Dependency

- A relationship between attributes in an entity
  - One or more attributes determine the value of another attribute
- An identifier functionally determines all the attributes of an entity
  - Shares.ID → stock code, firm name, stock price, stock quantity, stock dividend
  - If we know ID, we know the value of firm name, etc.
- Multivalued dependency
- Formulae
  - (stock dividend, stock price) → yield
  
  \[ \text{yield} = \left( \frac{\text{dividend}}{\text{price}} \right) \times 100 \]

Full Functional Dependency

- Yield is fully functionally dependent on stock dividend and stock price because both of these attributes are required to determine the value of yield
  - (stock dividend, stock price) → yield
- Determinant
  - An attribute that fully functionally determines another attribute
  - e.g., ID determines stock PE
Attribute Relationships

- One-to-one
  - Value of an attribute determines the value of another attribute and vice versa
  - $A \rightarrow B$ and $B \rightarrow A$
  - e.g.,
    - CH $\rightarrow$ Switzerland
    - Switzerland $\rightarrow$ CH

- One-to-many
  - A value of one attribute determines the value of another attribute but **not** vice versa
  - $A \rightarrow B$
  - e.g.,
    - country name $\rightarrow$ currency unit
    - currency unit not $\rightarrow$ country name
Normal Forms

- Based on rules about relationships among the columns of a table
- Removes data redundancies that can cause update anomalies
- A classification of relations
  - 1NF
  - 2NF
  - 3NF
  - BCNF
  - 4NF
  - 5NF

Usually, only the first 3 normal forms are applied to a DB.

Data Redundancy

- Major aim of relational database design is to group columns into tables to:
  1. minimize data redundancy and
  2. reduce file storage space required by implemented base tables
- Problems associated with data redundancy are illustrated by

Problems associated with data redundancy are illustrated in the example on the following slides.
StaffDistributionCenters Table

• Note the details of a distribution center are repeated for every employee (not normal form)

<table>
<thead>
<tr>
<th>ID</th>
<th>staffNo</th>
<th>staffName</th>
<th>position</th>
<th>salary</th>
<th>dCenterNo</th>
<th>dAddress</th>
<th>dTelNo</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$5500</td>
<td>Tom Daniels</td>
<td>Manager</td>
<td>$48,000</td>
<td>00000</td>
<td>8 Jefferson Way, Portland, OR 97201</td>
<td>503-555-3518</td>
</tr>
<tr>
<td>2</td>
<td>50003</td>
<td>Sally Adams</td>
<td>Assistant</td>
<td>$30,000</td>
<td>00000</td>
<td>8 Jefferson Way, Portland, OR 97201</td>
<td>503-555-3518</td>
</tr>
<tr>
<td>3</td>
<td>50001</td>
<td>Mary Martinez</td>
<td>Manager</td>
<td>$51,000</td>
<td>00002</td>
<td>City Center Plaza, Seattle, WA 98122</td>
<td>206-555-5510</td>
</tr>
<tr>
<td>4</td>
<td>51250</td>
<td>Robert Chin</td>
<td>Assistant</td>
<td>$35,000</td>
<td>00002</td>
<td>City Center Plaza, Seattle, WA 98122</td>
<td>206-555-5510</td>
</tr>
<tr>
<td>5</td>
<td>52250</td>
<td>Sally Stern</td>
<td>Manager</td>
<td>$48,000</td>
<td>00004</td>
<td>2 W. El Camino, San Francisco, CA 94157</td>
<td>415-555-3131</td>
</tr>
<tr>
<td>6</td>
<td>50415</td>
<td>Art Peters</td>
<td>Manager</td>
<td>$42,000</td>
<td>00003</td>
<td>14 - 38th Avenue, New York, NY 10012</td>
<td>212-371-3000</td>
</tr>
</tbody>
</table>

Update Anomalies

• Tables that contain redundant information may potentially suffer from update anomalies
• Types of update anomalies include:
  • Insertion – how do you insert details of a new distribution center that has no employees?
  • Deletion – when we delete the last employee in a distribution center, we lose the information about the distribution center
  • Modification – changes to a distribution center must be made for all records containing that distribution center
Better Design

First Normal Form (1NF)

- All rows must have the same number of columns
- Single valued attributes only

No universal agreement as to what would disqualify a table from being in 1NF

Resist the temptation to include repeated fields as CSV text
Example – Table not 1NF

Repeated field

Converting to 1NF

Replace a repeating group with a foreign key relationship
Second Normal Form (2NF)

- Violated when a non-key column is a fact about part of the primary key
- A column is not fully functionally dependent on the primary key
  - **customer-credit** in this case

<table>
<thead>
<tr>
<th>itemno</th>
<th>customerid</th>
<th>quantity</th>
<th>customer-credit</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>57</td>
<td>25</td>
<td>OK</td>
</tr>
<tr>
<td>34</td>
<td>679</td>
<td>3</td>
<td>POOR</td>
</tr>
</tbody>
</table>

Mainly applies to tables with multiple natural keys

---

Third Normal Form (3NF)

- Violated when a non-key column is a fact about another non-key column, restated as
  - A column is not fully functionally dependent on the primary key

<table>
<thead>
<tr>
<th>stock code</th>
<th>nation</th>
<th>exchange rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>MG</td>
<td>USA</td>
<td>0.67</td>
</tr>
<tr>
<td>IR</td>
<td>AUS</td>
<td>0.46</td>
</tr>
</tbody>
</table>
Example - **not 3NF**

Values in `dAddress` and `dTelNo` can be determined from `dCenterNo`.

Values in `staffNo`, `staffName`, `position`, and `salary` are determined from `ID`.

<table>
<thead>
<tr>
<th>ID</th>
<th>StaffNo</th>
<th>StaffName</th>
<th>Position</th>
<th>Salary</th>
<th>dCenterNo</th>
<th>dAddress</th>
<th>dTelNo</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>51500</td>
<td>Tom Daniels</td>
<td>Manager</td>
<td>$48,000.00</td>
<td>00001</td>
<td>8 Jefferson Way, Portland, OR 97201</td>
<td>503-555-1618</td>
</tr>
<tr>
<td>2</td>
<td>50003</td>
<td>Sally Adams</td>
<td>Assistant</td>
<td>$30,000.00</td>
<td>00001</td>
<td>8 Jefferson Way, Portland, OR 97201</td>
<td>503-555-3018</td>
</tr>
<tr>
<td>3</td>
<td>50010</td>
<td>Mary Martinez</td>
<td>Manager</td>
<td>$53,000.00</td>
<td>00002</td>
<td>City Center Plaza, Seattle, WA 98122</td>
<td>206-555-0575</td>
</tr>
<tr>
<td>4</td>
<td>53250</td>
<td>Robert Chin</td>
<td>Assistant</td>
<td>$33,000.00</td>
<td>00002</td>
<td>City Center Plaza, Seattle, WA 98122</td>
<td>206-555-0575</td>
</tr>
<tr>
<td>5</td>
<td>52250</td>
<td>Sally Stern</td>
<td>Manager</td>
<td>$40,000.00</td>
<td>00004</td>
<td>2 W. El Camino, San Francisco, CA 94087</td>
<td>822-555-3131</td>
</tr>
<tr>
<td>6</td>
<td>52415</td>
<td>Art Peters</td>
<td>Manager</td>
<td>$42,000.00</td>
<td>00003</td>
<td>14 - 8th Avenue, New York, NY 10012</td>
<td>212-373-3000</td>
</tr>
</tbody>
</table>