# CSE 416

## Database Issues

1

## Database Preliminaries

- Recap important topics
- Recap terminology
- Use a good DB modelling tool (e.g., Workbench)
- You will implement the DB on a shared CS server (e.g., MySQL)



2

# An Entity

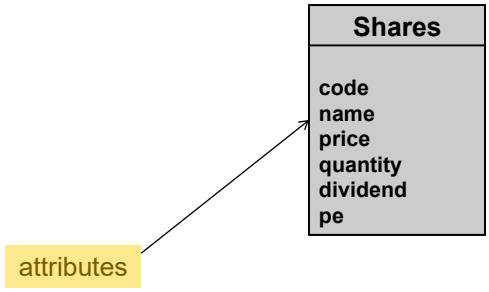
- Usually corresponds to something concrete in the domain of the application
- Represented by a rectangle
- An instance is a particular occurrence of an entity (corresponds to a row in a DB table)




3

# Attributes

- Also referred to as properties
- An attribute is a discrete data element that describes an entity
- Attribute names should be meaningful

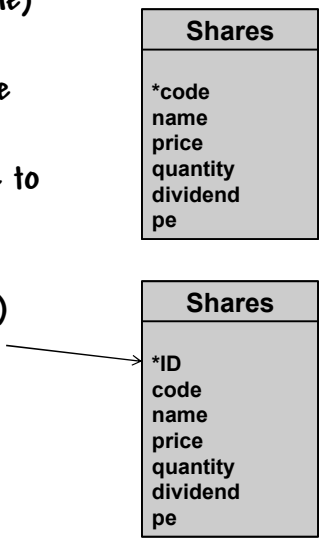





4

# Identifiers (Primary Keys)

- Every instance of an entity (think row of a table) must be uniquely identified
- An identifier (primary key) can be one or more attributes
- Better to use an identifier that does not relate to a domain attribute (guaranteed uniqueness)
- A leading asterisk denotes an identifier (sometimes, another notation is used, e.g., PK)






5

# DB Naming Conventions

- No universal standard
- Good to be consistent within a project
- Camel case used frequently



6

## CSE416 DB Naming Conventions

- Options
  - Camel case for table names (upper cc) and column names (lower cc)
  - All caps for table names with underscore (_) as a separator
- Table names - plural (unlike OO convention)
- Column names - singular
- Primary key field - ID
- Avoid acronyms and abbreviations except where well known (e.g., PI for Principal Investigator)



7

## Data Modeling

- A technique for modeling data
- We assume
  - RDM Model (Relational Data Model)
- The goal is to identify the structure of data to be stored in the database

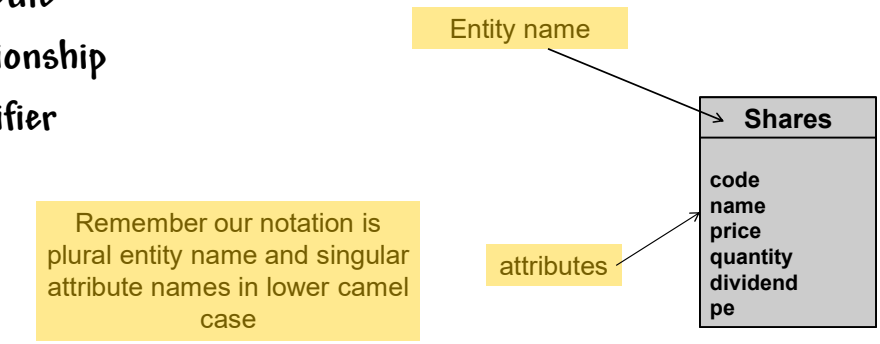
ER Model is applicable to non-relational DB, but we assume a relational implementation



8

## The Building Blocks

- Entity
- Attribute
- Relationship
- Identifier




Remember our notation is plural entity name and singular attribute names in lower camel case



9

## A Well-Formed Data Model

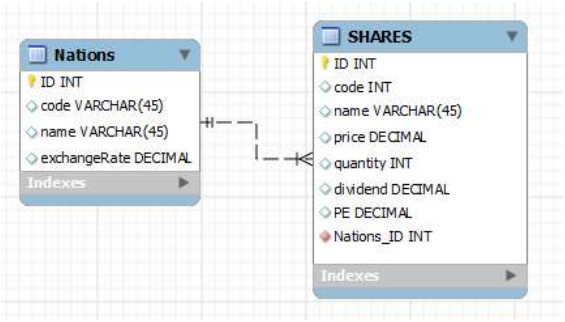
- Follow organization (e.g., your company) convention
- No ambiguity
  - All entities, attributes, relationships, and identifiers are defined
  - Names are meaningful to the client



10

# Relationships

- ERD and RDM show relationships between entities
  - 1-1
  - 1-many
  - Recursive
- ERD shows
  - Many-many
  - No foreign keys
- RDM usually shows
  - Associative entity (in-between table)
  - Foreign keys




Workbench uses more of a DB model style (not an ERD)



11

# 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

You might be asked to show your DB during your code review



* Wikipedia

12

## 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

Workbench uses more of a DB model style (not an ERD)



13

## 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

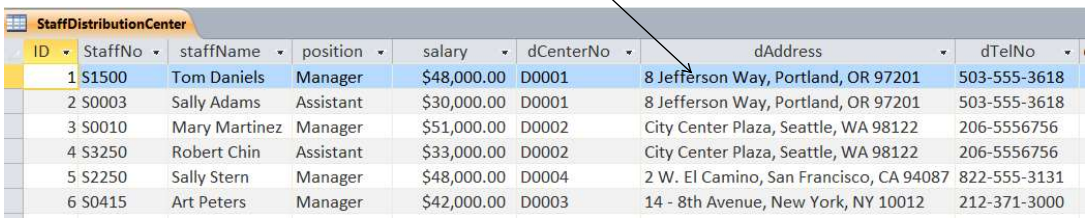
Problems associated with data redundancy are illustrated in the example on the following slides



14

# StaffDistributionCenters Table

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



StaffDistributionCenter

| ID | StaffNo | staffName | position | salary | dCenterNo | dAddress | dTelNo |
|---|---|---|---|---|---|---|---|
| 1 | S1500 | Tom Daniels | Manager | $48,000.00 | D0001 | 8 Jefferson Way, Portland, OR 97201 | 503-555-3618 |
| 2 | S0003 | Sally Adams | Assistant | $30,000.00 | D0001 | 8 Jefferson Way, Portland, OR 97201 | 503-555-3618 |
| 3 | S0010 | Mary Martinez | Manager | $51,000.00 | D0002 | City Center Plaza, Seattle, WA 98122 | 206-5556756 |
| 4 | S3250 | Robert Chin | Assistant | $33,000.00 | D0002 | City Center Plaza, Seattle, WA 98122 | 206-5556756 |
| 5 | S2250 | Sally Stern | Manager | $48,000.00 | D0004 | 2 W. El Camino, San Francisco, CA 94087 | 822-555-3131 |
| 6 | S0415 | Art Peters | Manager | $42,000.00 | D0003 | 14 - 8th Avenue, New York, NY 10012 | 212-371-3000 |



15

# 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



16

# Better Design

StaffDistributionCenter | Staff

| ID | staffNo | staffName | position | salary | dCenterID |
|---|---|---|---|---|---|
| 1 | S1500 | Tom Daniels | Manager | $48,000.00 | 1 |
| 2 | S0003 | Sally Adams | Assistant | $30,000.00 | 1 |
| 3 | S0010 | Mary Martinez | Manager | $51,000.00 | 2 |
| 4 | S3250 | Robert Chin | Assistant | $33,000.00 | 2 |
| 5 | S2250 | Sally Stern | Manager | $48,000.00 | 4 |
| 6 | S0415 | Art Peters | Manager | $42,000.00 | 3 |

Staff | DistributionCenters

| ID | dCenterNo | dAddress | dTelNo |
|---|---|---|---|
| 1 | D0001 | 8 Jefferson Way, Portland, OR 97201 | 503-555-3618 |
| 2 | D0002 | City Center Plaza, Seattle, WA 98122 | 206-5556756 |
| 3 | D0003 | 14 - 8th Avenue, New York, NY 10012 | 212-371-3000 |
| 4 | D0004 | 2 W. El Camino, San Francisco, CA 94087 | 822-555-3131 |



17

# First Normal Form (INF)

- 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 INF

Typical violation of INF

Customer

| Customer ID | First Name | Surname | Telephone Number |
|---|---|---|---|
| 123 | Robert | Ingram | 555-861-2025 |
| 456 | Jane | Wright | 555-403-1659<br>555-776-4100 |
| 789 | Maria | Fernandez | 555-808-9633 |

Resist the temptation to include repeated fields as CSV text



18

# Example - Table not INF

Staff | DistributionCenters

| ID | dCenterNo | dAddress | dTelNo |
|---|---|---|---|
| 1 | D0001 | 8 Jefferson Way, Portland, OR 97201 | 503-555-3618, 503-555-3619 |
| 2 | D0002 | City Center Plaza, Seattle, WA 98122 | 206-5556756 |
| 3 | D0003 | 14 - 8th Avenue, New York, NY 10012 | 212-371-3000 |
| 4 | D0004 | 2 W. El Camino, San Francisco, CA 94087 | 822-555-3131 |

Repeated field



19

# Converting to INF

DistributionCenters

| ID | dCenterNo | dAddress | dTelNo |
|---|---|---|---|
| 1 | D0001 | 8 Jefferson Way, Portland, OR 97201 | 503-555-3618, 503-555-3619 |
| 2 | D0002 | City Center Plaza, Seattle, WA 98122 | 206-5556756 |
| 3 | D0003 | 14 - 8th Avenue, New York, NY 10012 | 212-371-3000 |
| 4 | D0004 | 2 W. El Camino, San Francisco, CA 94087 | 822-555-3131 |

Replace a repeating group with a foreign key relationship

DistributionCenters1NF

| ID | dCenterNo | dAddress |
|---|---|---|
| 1 | D0001 | 8 Jefferson Way, Portland, OR 97201 |
| 2 | D0002 | City Center Plaza, Seattle, WA 98122 |
| 3 | D0003 | 14 - 8th Avenue, New York, NY 10012 |
| 4 | D0004 | 2 W. El Camino, San Francisco, CA 94087 |

DistribtionCentersTelephones

| ID | dCenterID | dTelNo |
|---|---|---|
| 1 | 1 | 503-555-3618 |
| 2 | 1 | 503-555-3619 |
| 3 | 2 | 206-5556756 |
| 4 | 3 | 212-371-3000 |
| 5 | 4 | 822-555-3131 |



20

# 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

Mainly applies to tables with multiple natural keys

| order | | | |
|---|---|---|---|
| itemno | customerid | quantity | customer-credit |
| 12 | 57 | 25 | OK |
| 34 | 679 | 3 | POOR |



21

# 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

Exchange rate is a fact about a nation

| stock | | |
|---|---|---|
| stock code | nation | exchange rate |
| MG | USA | 0.67 |
| IR | AUS | 0.46 |



22

# Example - not 3NF

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

Values in dAddress and dTelNo can be determined from dCenterNo

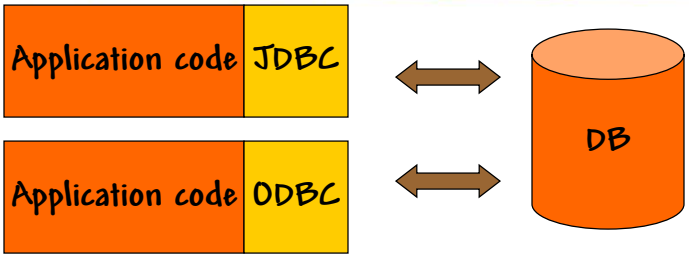
StaffDistributionCenter

| ID | StaffNo | staffName | position | salary | dCenterNo | dAddress | dTelNo |
|---|---|---|---|---|---|---|---|
| 1 | S1500 | Tom Daniels | Manager | $48,000.00 | D0001 | 8 Jefferson Way, Portland, OR 97201 | 503-555-3618 |
| 2 | S0003 | Sally Adams | Assistant | $30,000.00 | D0001 | 8 Jefferson Way, Portland, OR 97201 | 503-555-3618 |
| 3 | S0010 | Mary Martinez | Manager | $51,000.00 | D0002 | City Center Plaza, Seattle, WA 98122 | 206-5556756 |
| 4 | S3250 | Robert Chin | Assistant | $33,000.00 | D0002 | City Center Plaza, Seattle, WA 98122 | 206-5556756 |
| 5 | S2250 | Sally Stern | Manager | $48,000.00 | D0004 | 2 W. El Camino, San Francisco, CA 94087 | 822-555-3131 |
| 6 | S0415 | Art Peters | Manager | $42,000.00 | D0003 | 14 - 8th Avenue, New York, NY 10012 | 212-371-3000 |



23

# Interface Issues




- The application deals with objects, language specific data types, and higher level concepts
- The DB deals with relational tables and SQL
- Translation is usually performed to allow the two components to work together

Translation required DB code and application libraries



24