Data Modelling

Learning Outcomes

- Understand and be able to explain the basic concepts associated with Entity-Relationship(ER) model.
- Be able to use Entity–Relationship (ER) modelling in database design.
- Be able to build an ER model from a requirements specification.

Data model

- Recap: **Database** is a shared collection of logically related data (and a description of this data), designed to meet the information needs of an organization.
- How do we build a Database?
- A technique called **data modelling** helps you to understand the structure and meaning of data.
- A **data model** is a graphical description of the components of database.

Recap of Relational model

- E. F. Codd proposed relational data model in 1970.
- In the relational model, all data is logically structured within relations (tables).
- A **relation**, is a two-dimensional table arranged in columns and rows.
- A **relational database** is a collection of relations.
- One row of a table stores details of one case (instance) of an item. All the rows in a table store data about the same type of items.
- One column in the table contains the same type of data.

Recap of Relational model

- In a relational database, each row must be uniquely identified with **primary key**.

studentID	firstName	surname	age	programme
001	Mary	White	20	IoT
004	Tom	Hardy	19	Telecom
006	Mary	Bennet	20	E-commerce
032	John	Doe	21	Telecom
101	Ann	Martins	19	IoT

- The tables in a relational database are **connected** or **related** by means of the data in the tables.

Entity-relationship (E-R) modelling

- E-R modelling is a high-level *conceptual modelling technique for DB applications*
- Developed by Peter Chen and published in 1976 *
- Main concepts (building blocks)
  - Entity
  - Attributes
  - Relationship

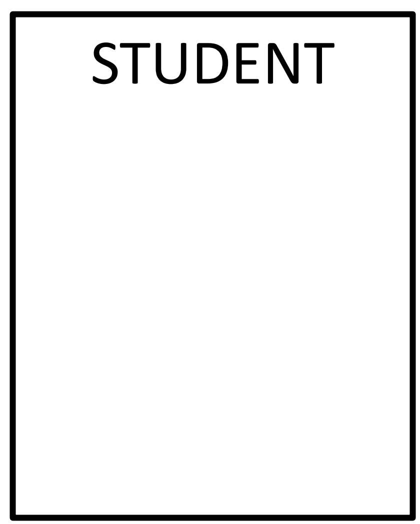
* Chen, Peter (March 1976). "The Entity-Relationship Model - Toward a Unified View of Data". *ACM Transactions on Database Systems* 1 (1): 9–36.

Entity

- Basic building block of a data model.
- An entity is a “thing” about which data should be stored.
- Group of objects with same properties, identified by enterprise as having an independent existence.
- Can be objects with a physical or conceptual existence
  - Physical existence: (e.g. student or a textbook)
  - Conceptual existence: (e.g. a module or an exam)

Entity

An entity is represented by a *rectangle*. The name of the entity is shown in singular form in uppercase in the top part of the rectangle.



Attributes

- An entity has characteristics or **attributes**.
- An attribute is a discrete element of data; it describes an entity.
- Attributes shown below the entity's name.
- Attribute names must be carefully selected so that they are self-explanatory and unique.
- A identifier (primary key) uniquely identifies an instance of an entity. Attribute(s) that are identifiers are labelled in the entity (here with a star).

STUDENT
* studentID
firstName
surName
age
programme

Exercise 1

Read the following description:

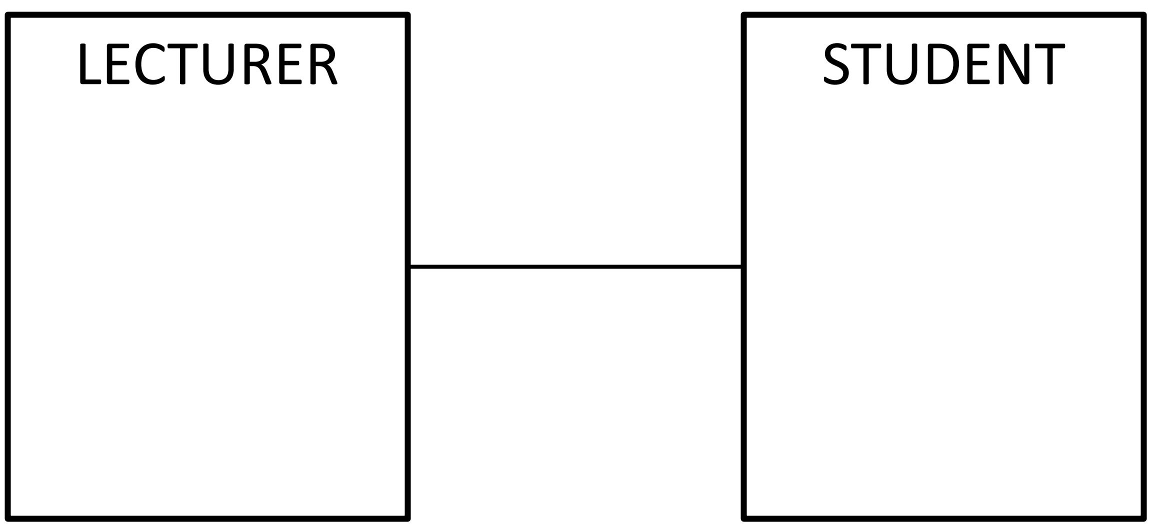
- An second hand bookshop want to keep a record of all the books in stock.
- Each book has a title, ISBN number, year of publication, price, condition of the book (like new, good condition, worn etc.)
- Design a single entity database using ER diagram. *Clearly* ***label everything***.

One possible solution

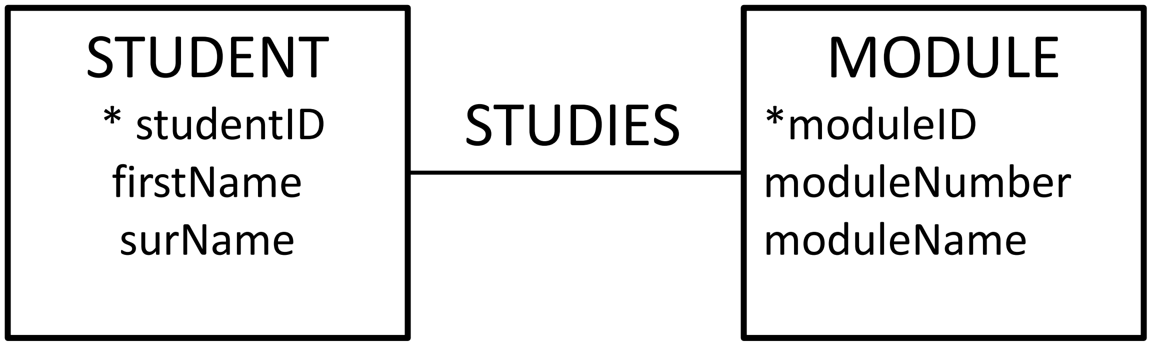
BOOK
title
*ISBN
yearPublished
price
condition

Relationship

- Entities are related to other entities.
- Relationship describes a linkage between two entities and is represented by an ***arc*** between them.



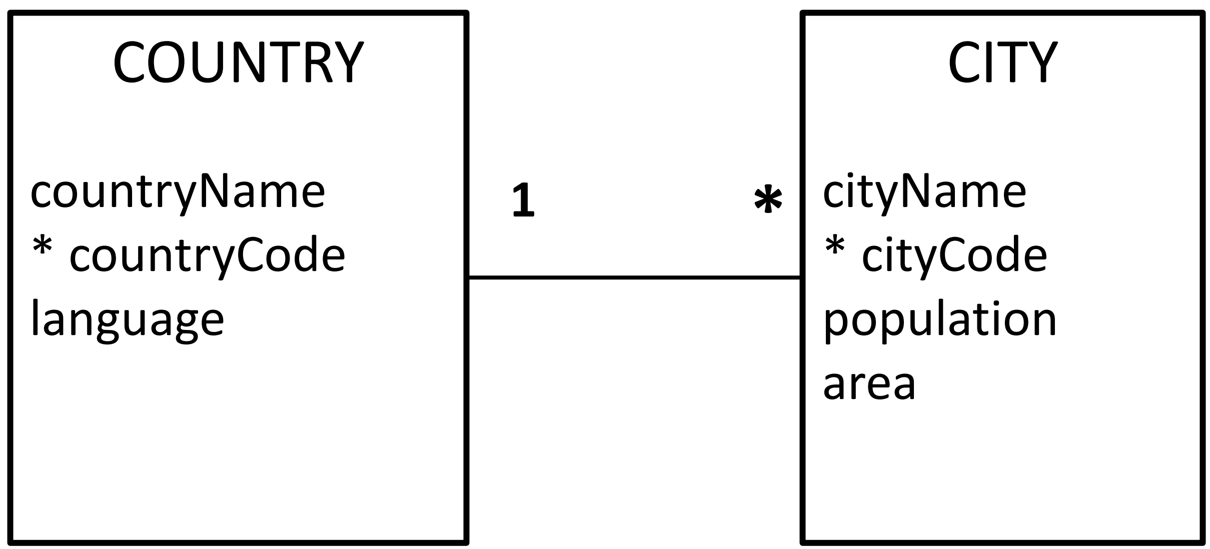
Test your understanding



- True or false:
  - Statement 1: STUDENT is an Entity
  - Statement 2: STUDIES is an Entity
  - Statement 3: moduleID is a primary key for module
  - Statement 4: moduleID is an attribute of MODULE

One-to-many (1:m) relationship

- Consider the database to record countries and cities. (On E-R diagram 1 and * shows 1 to many)



- This can be read as: “a country can have many cities, but a city belongs to only one country.”

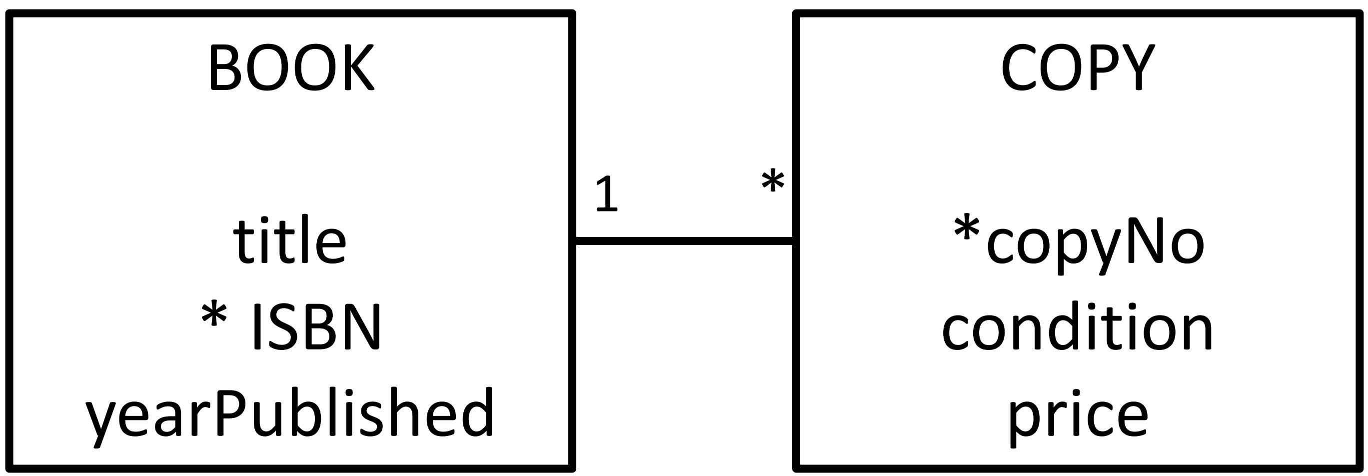
Exercise 2

The second hand bookshop now realised a problem in their earlier single entity database:

- one book may have *many* copies and each copy has its own condition, e.g. Harry Potter and the Philosopher's Stone has 3 copies, one nearly new condition and two worn condition; the nearly new condition copy has a higher price than the worn ones.

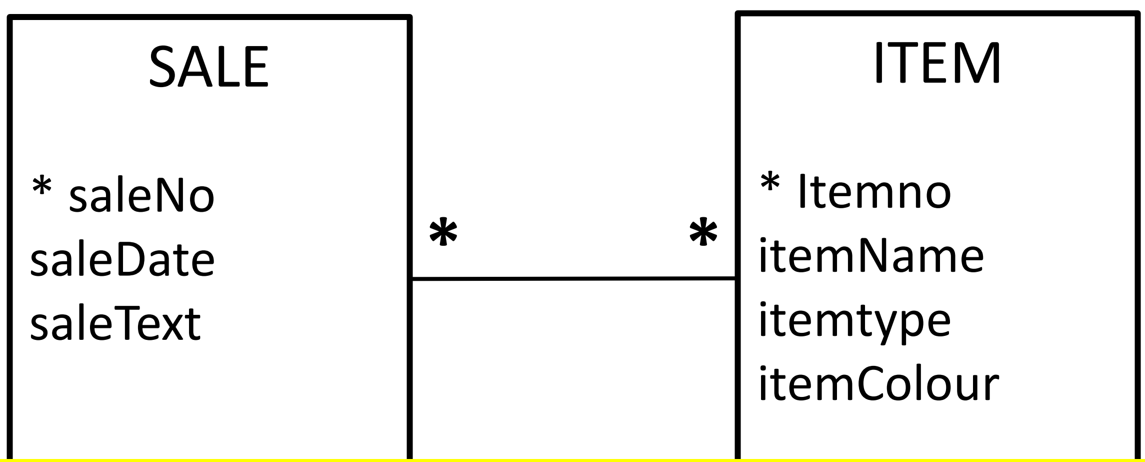
Redesign your database with 1:m relationship.

One possible solution



Many-to-Many (m:m) relationship

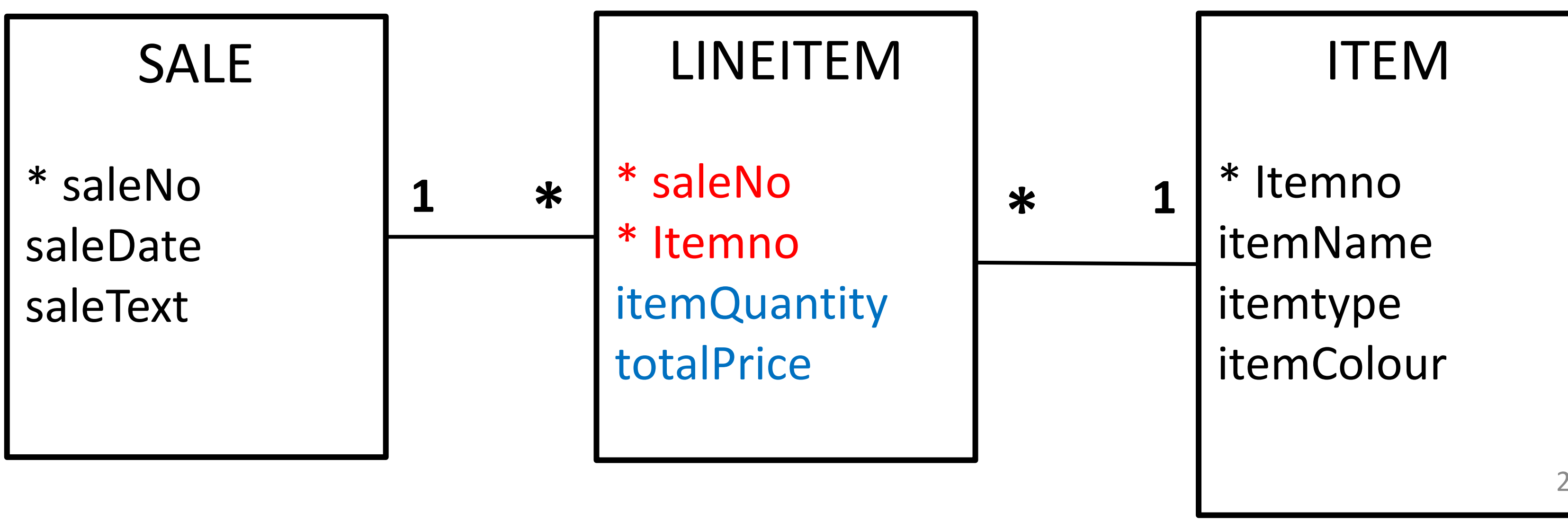
- Consider the case when items are sold. We can identify two entities: SALE and ITEM. A sale can contain many items, and an item can appear in many sales. (On ER diagram * and *)



But how do we record the information for the *m:m* relationship?

Many-to-Many (m:m) relationship

- Information missing from the relationship includes: quantity of an item being sold, total price etc.
- To store these information (attributes related to the m:m relationship), we create a third entity (**associative entity**)to link the entities through two 1:m relationships.



Relational keys

- **Candidate Key**
  - A set of attributes that uniquely identifies a tuple within a relation.
  - Uniqueness : In each tuple, candidate key uniquely identify that tuple.
  - Irreducibility: No proper subset of the candidate key has the uniqueness property.
- **Primary Key**
  - Candidate key selected to identify tuples uniquely within relation.
- **Foreign Key**
  - Attribute, or set of attributes, within one relation that matches candidate key of some (possibly same) relation.
- **Composite Key**
  - A candidate key that consists of two or more attributes.

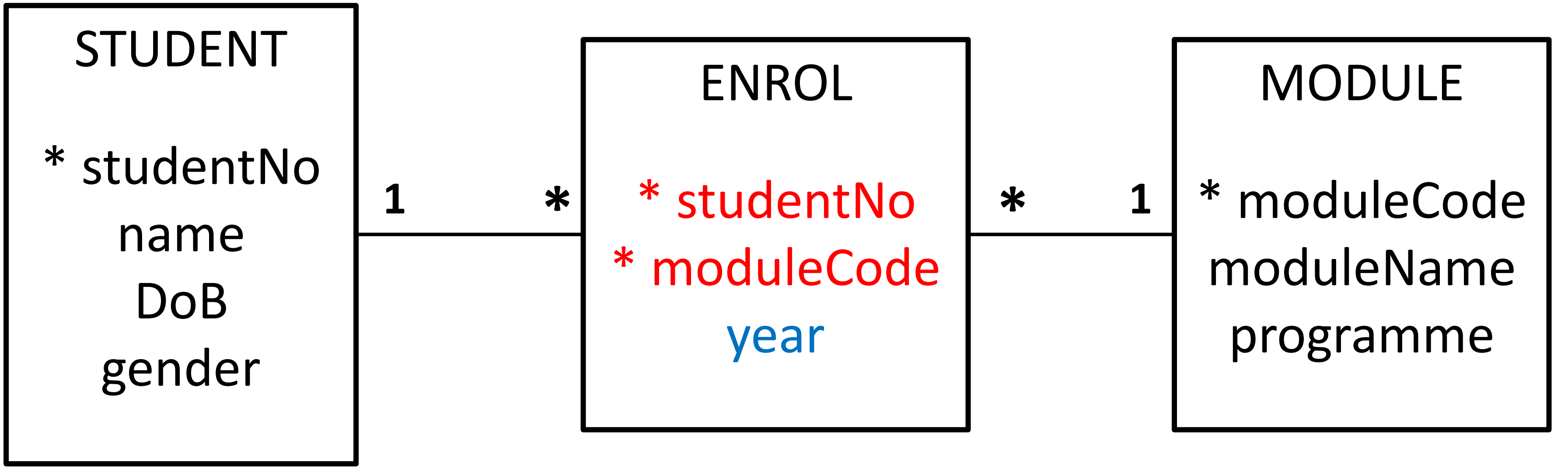
Exercise 3

Considering the following descriptions about students enrolling on modules:

- A student has various properties including student number, name, date of birth, gender.
- A module has various properties including module code, name, programme a module belongs to.
- A student can enrol on many modules and a module can be enrolled by many students. The database also need to record the year a student enrols on a module.

Draw the ER diagram for the above descriptions.

One possible solution

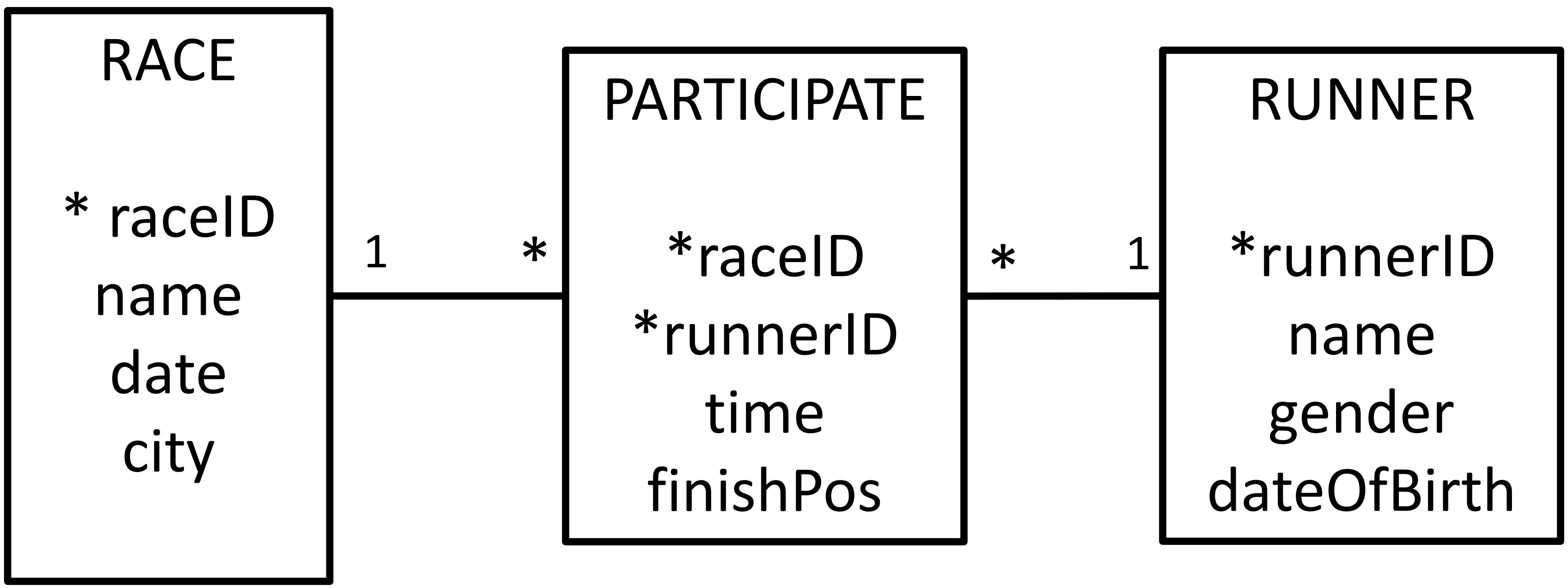


Exercise 4

Create an ER model for a sport magazine with the following descriptions:

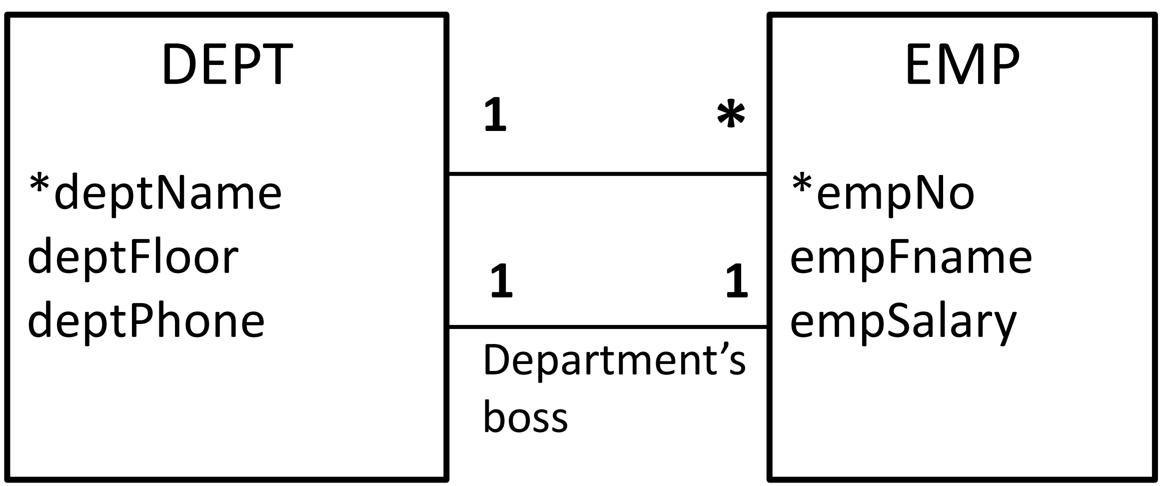
- The sport magazine regularly reports the performance of professional marathon runners. It has asked you to design a database to record the details of all major marathons (e.g., London, Beijing and Paris marathons).
- A professional marathon runner may compete in several races each year. A race will have many professional runners.
- For each race, the magazine reports a runner's time and finishing position and some personal details such as name, gender, and date of birth.

One possible solution



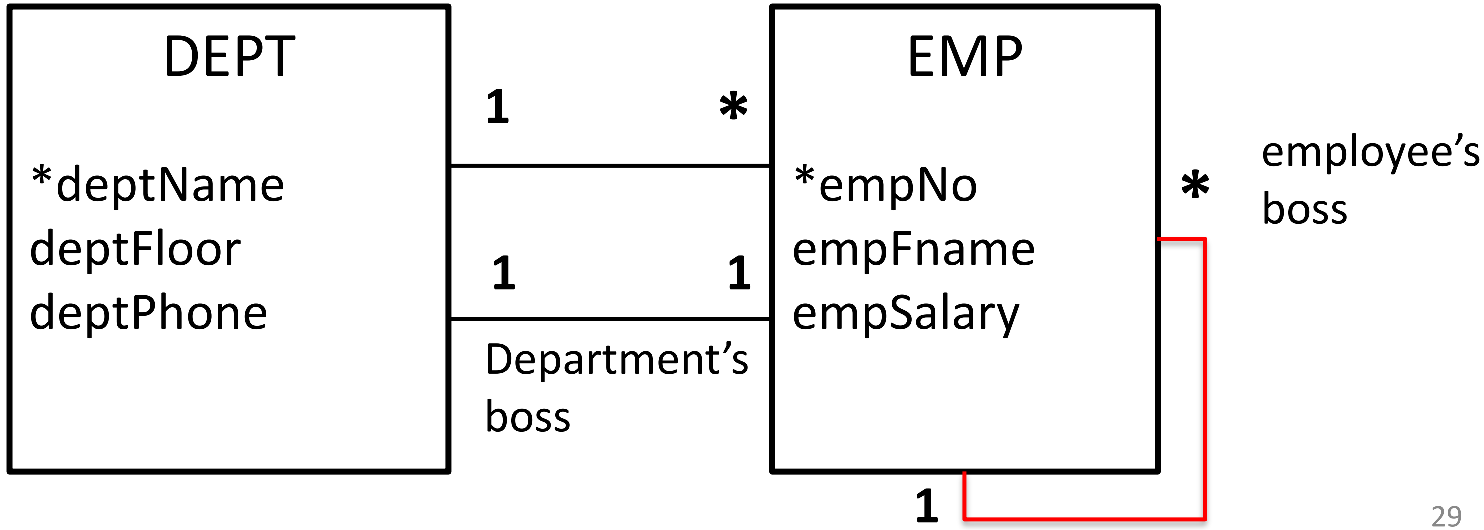
One-to-one relationship

- A department has one or more employees (EMP), and an employee belongs to one department (DEPT).
- A department has one boss, and a person is boss of only one department.
- Boss is a 1:1 relationship between DEPT and EMP.



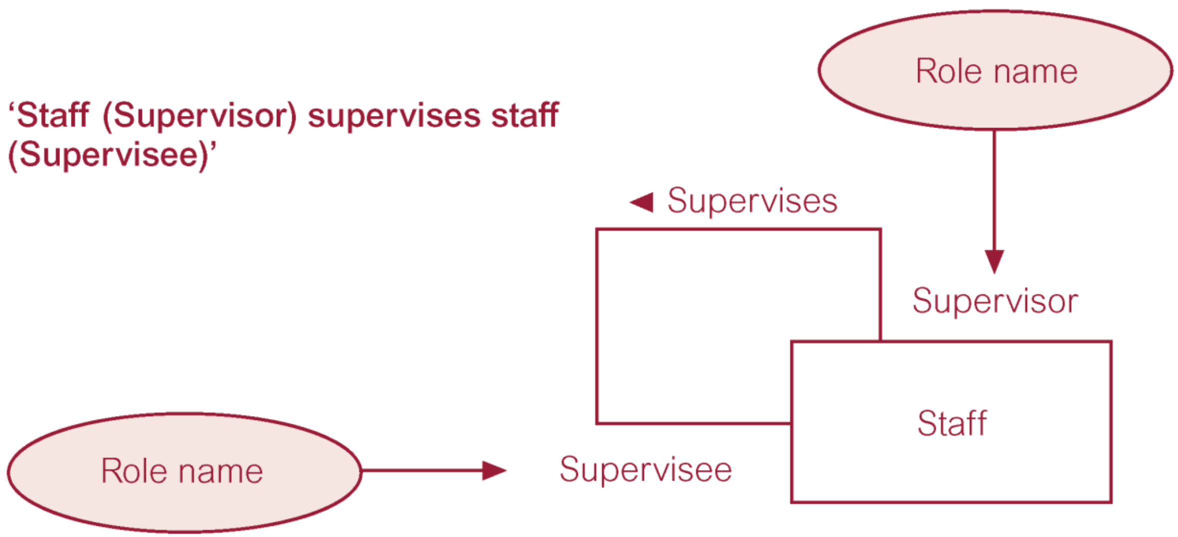
Recursive Relationships

- There is more to boss than just a department.
- People also have a boss. An employee can be boss to many other employees, and an employee has normally just one boss.
- The person-boss relationship is a recursive 1:m relationship.



Recursive Relationships

- Recursive Relationship
  - Relationship type where *same* entity type participates more than once in *different roles*.
- Relationships may be given role names to indicate purpose that each participating entity type plays in a relationship.



Structural Constraints

- Main type of constraint on relationships is called **multiplicity**.
- Multiplicity - number (or range) of possible occurrences of an entity type that may relate to a single occurrence of an associated entity type through a particular relationship.
- Represents policies (called *business rules*) established by user or company.

Structural Constraints

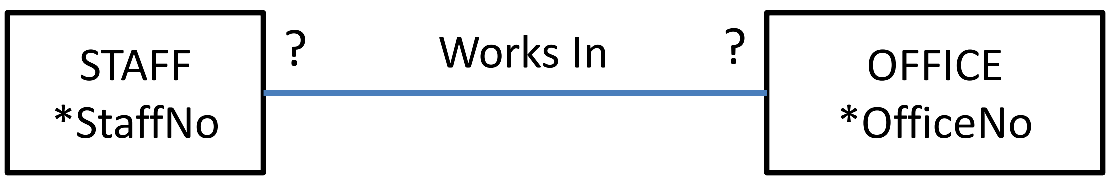
- The most common degree for relationships is binary.

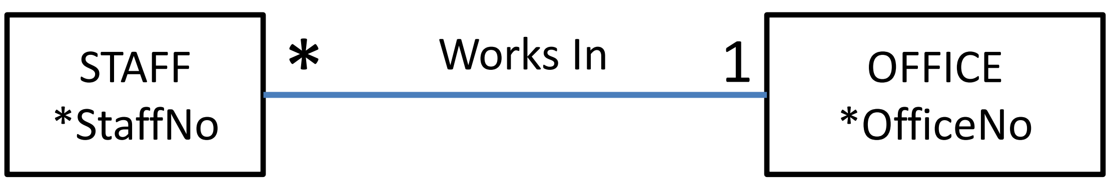
- Binary relationships are generally referred to as being:
  – one-to-one (1:1)
  – one-to-many (1:*)
  – many-to-many (*:*)

Exercise 5

- What is the multiplicity for the relationship:

Every member of staff has 1 (and only 1) office. Some offices empty, some have several members of staff. Replace the ? with the correct answer.





* means each office can have many staff

1 means each member of staff has exactly one office

Relationship Types

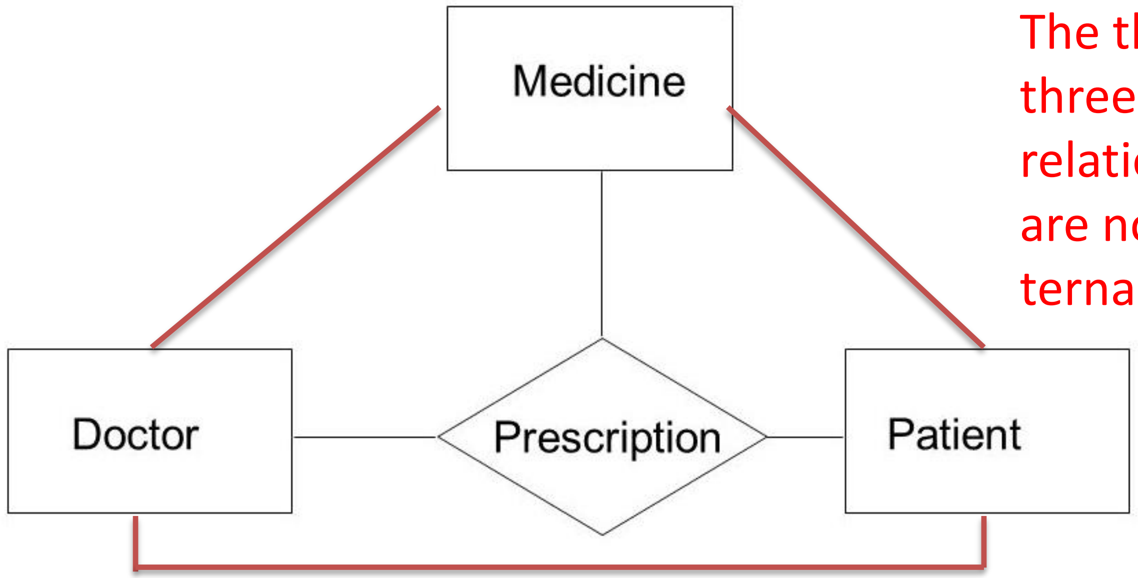
- Degree of a Relationship
  – Number of participating entities in a relationship.

- Degree of a Relationship:
  – two is binary (what we have seen so far are all binary relationships)
  – three is ternary
  – four is quaternary
  – n-ary is general n-entity relationship

Ternary relationship 1

- In a ternary relationship, <u>three</u> entities are ***simultaneously*** involved.

Ternary relationship of
“Doctor prescribes patients medicines”

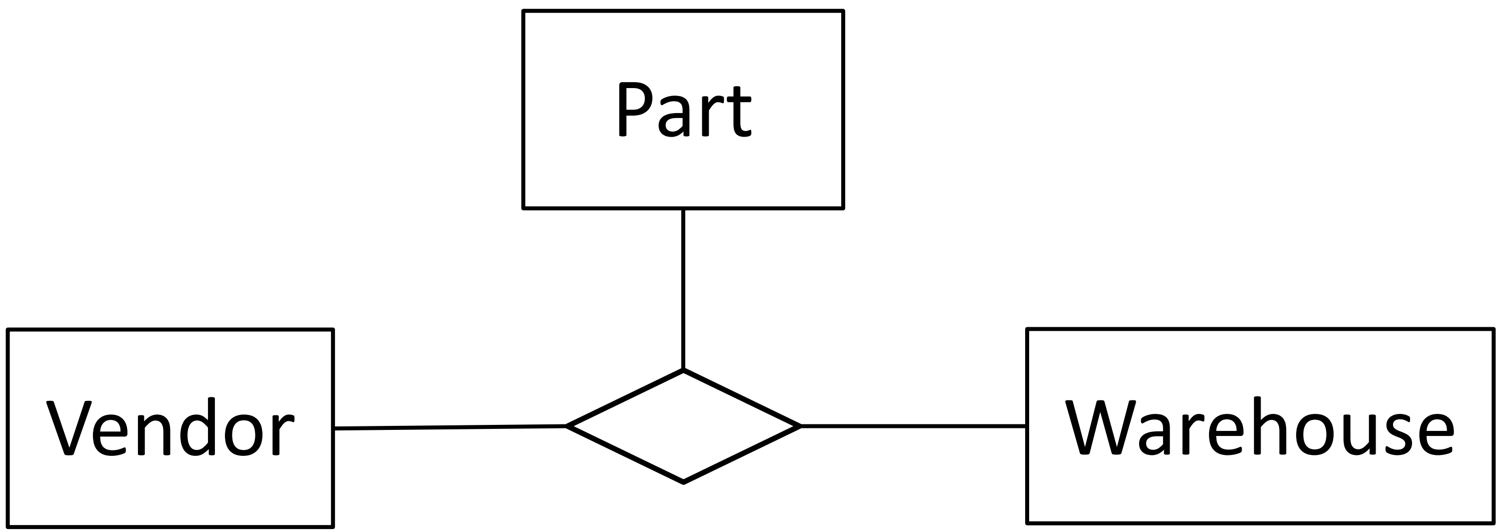


The three red lines are three binary relationships, and they are not the same as ternary relationship!

Note: a ternary relationship is ***not the same*** as three binary relationships!

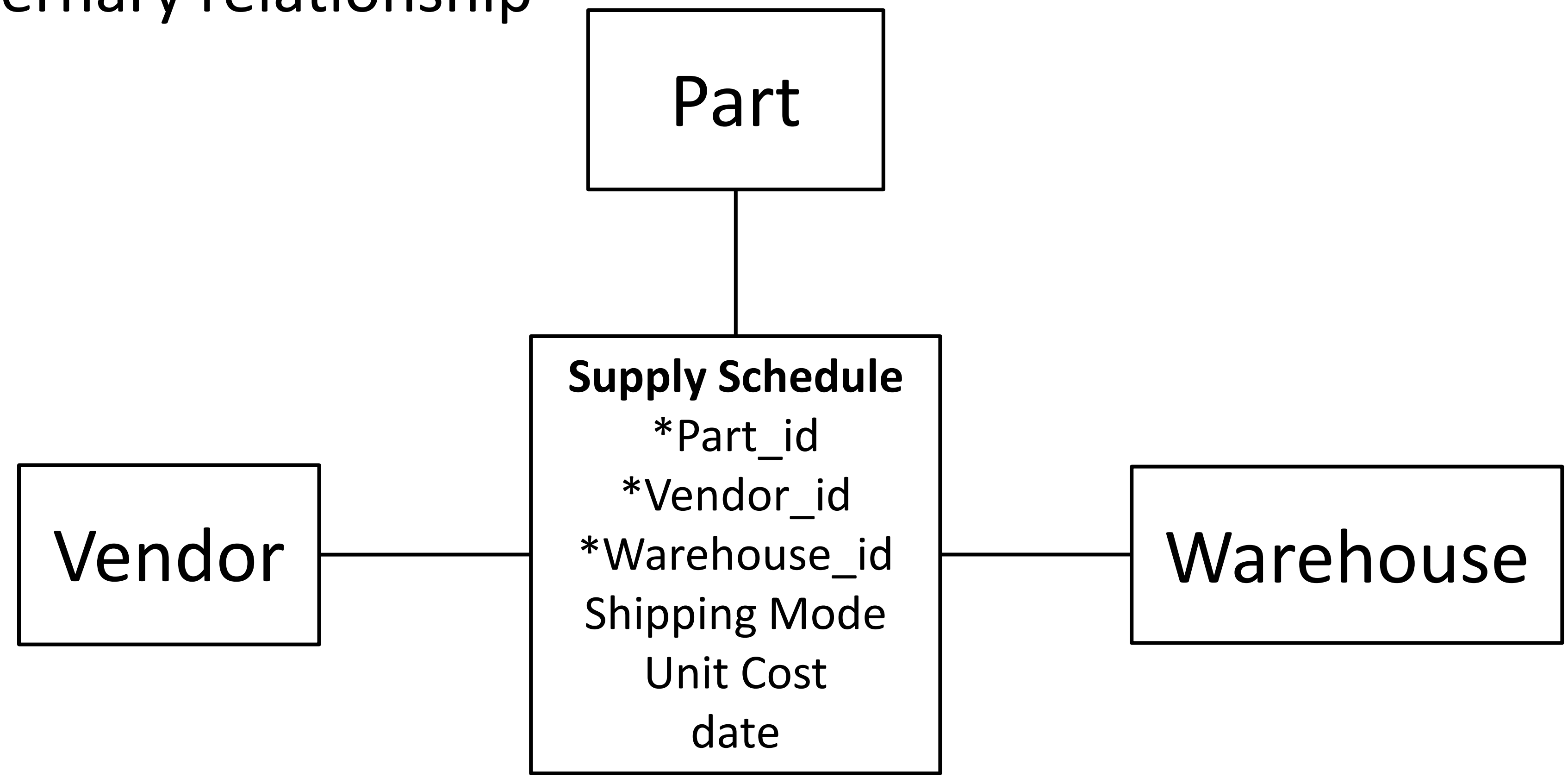
Ternary relationship 2

- Vendors can supply various parts to warehouses.
- Three entity types: Vendor, Part and Warehouse



Ternary Relationship 3

Attributes associated with the "*Supply Schedule*" ternary relationship



Note: a ternary relationship is ***not the same*** as three binary relationships! e.g. Unit cost cannot be properly associated with any one of the three possible binary relationships among the three entity types.

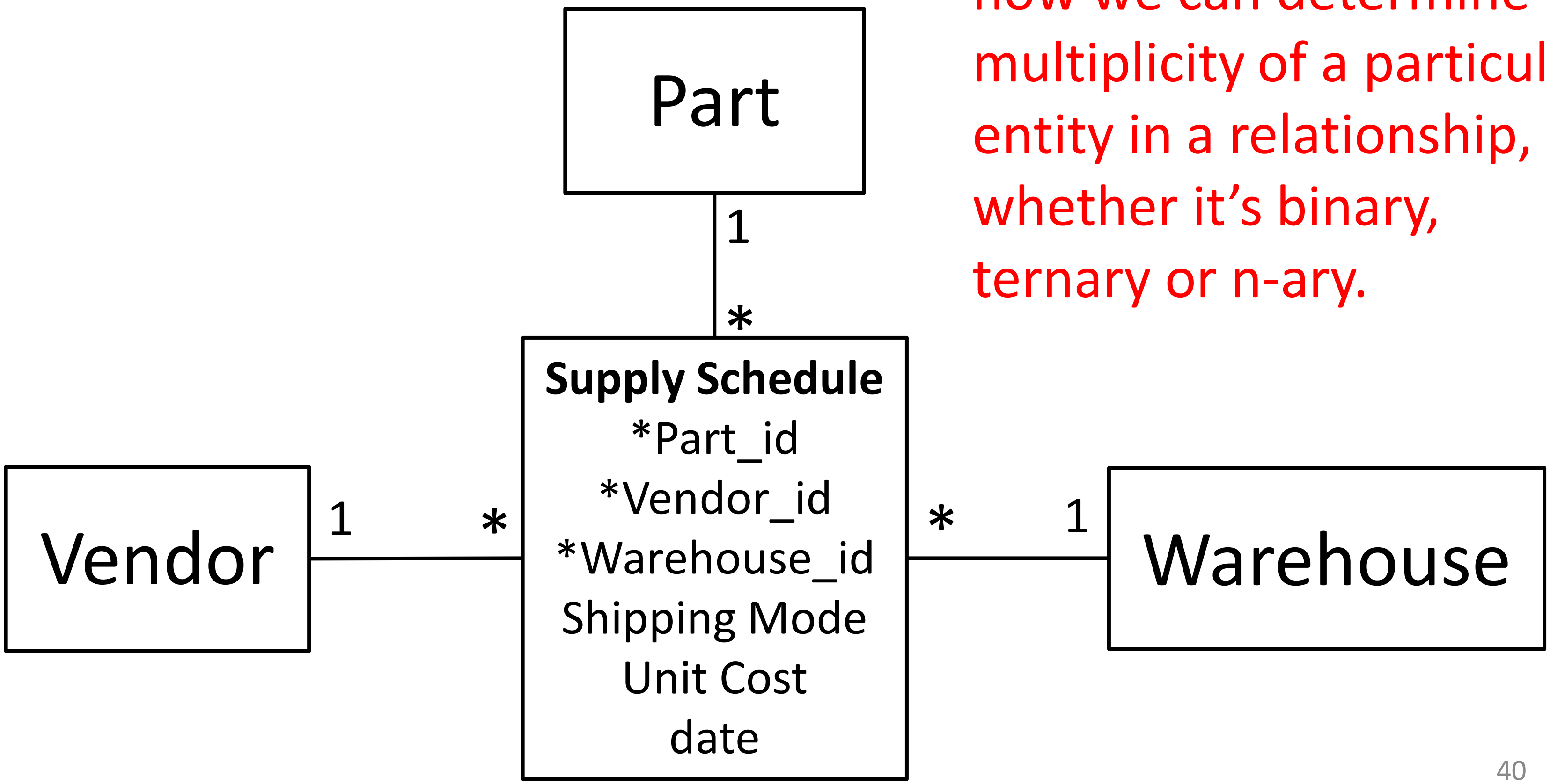
Ternary relationship 4

- Business rules
  - Each vendor supplies 0 or more parts to 0 or more warehouses.
  - Part can be supplied by 1 or more vendors to 1 or more warehouse.
  - Warehouse supplied with 0 or more parts from each of 0 or more vendors but must be supplied at least one part.

Ternary relationship 5

- Number (or range) of possible occurrences of an entity type in an *n*-ary relationship when other (*n*-1) values are fixed.

The above rule tells us how we can determine multiplicity of a particular entity in a relationship, whether it’s binary, ternary or n-ary.



Exercise 6

True or false:

1. A ternary relationship involves three entities.

   TRUE

2. A relationship with four entities is known as quaternary.

   TRUE

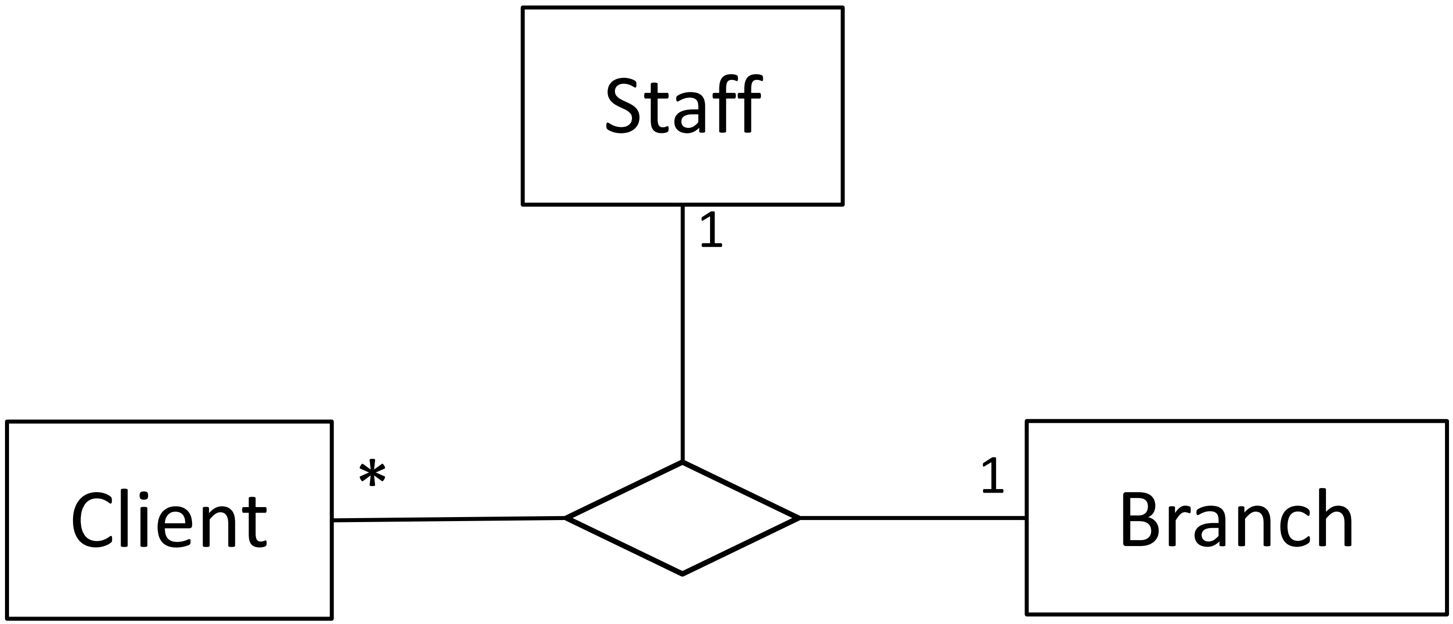
3. A ternary relationship can be simplified into three binary relationships

   FALSE – the relationship is between all three.

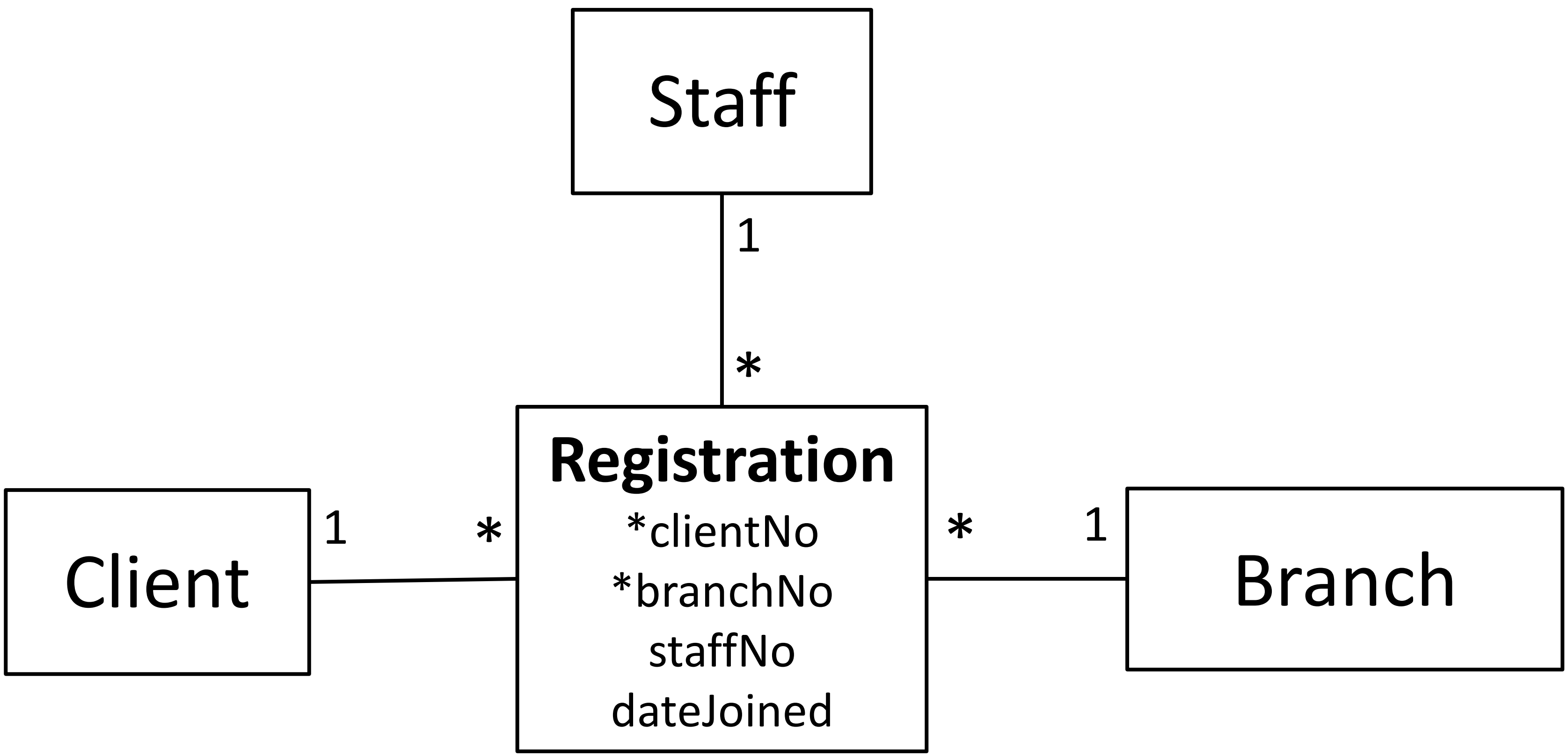
Exercise 7

- In DreamHome database, draw the ternary relationship of “a client is registered by a member of staff at a branch” and determine the multiplicity for the relationship.

One possible solution



One possible solution



What have we learned?

- Entity: (name, attributes, *primary key)
- Relationship
  - Degree: binary (2 entities), ternary (3 entities) …
  - In binary relationship:
    - 1:1, 1:m, m:m (create two 1:m relationships using associative entity)
  - Recursive relationships
- Multiplicity – defines the possible number of occurrences of each entity type in a relationship.

What have we learned?

- Relationship degree
  - Ternary relationship (three entities)
  - Not the same as three binary relationships
  - Quaternary is for four

Data modelling

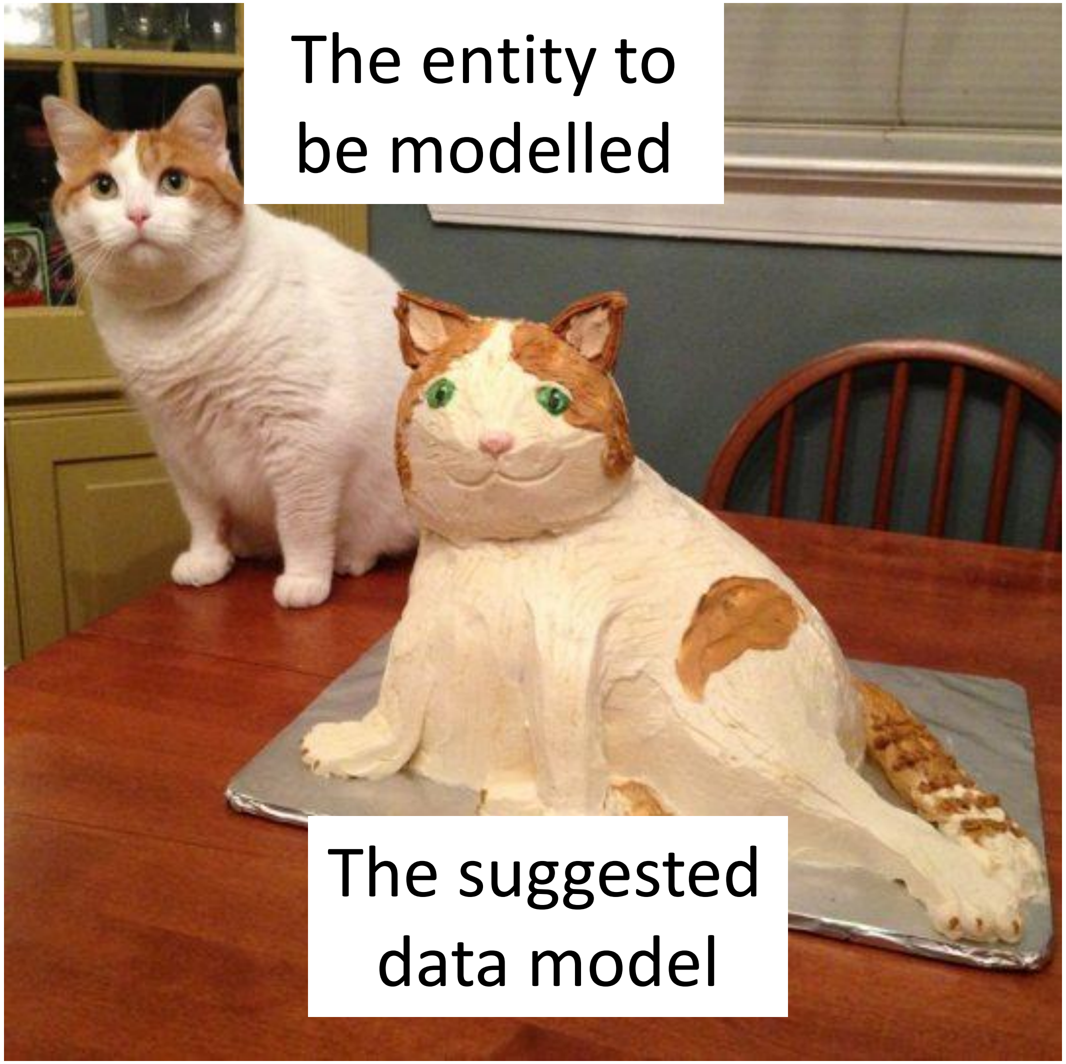
Data modelling

- A technique for modelling data
- A graphical representation of a database
- The goal is to identify the facts to be stored in a database
  - not concerned with how the data will be stored
  - not concerned with how the data will be processed
- Data modelling is a partnership between the client and designer
- Drawing a data model is an **iterative** process of trial and revision.

Data model quality

- The goal of a quality data model is to achieve:
  - A well-formed data model
  - A high fidelity image

Data model quality



Data model quality: A well-formed data model

- Construction rules obeyed
- No ambiguity
  - All entities, attributes, relationships, and identifiers are defined
  - All relationships are represented, using the correct notation
  - Relationships are labeled to avoid misunderstanding
  - All attribute names are meaningful and unique
  - Names are meaningful to the client

Data model quality: A high fidelity image

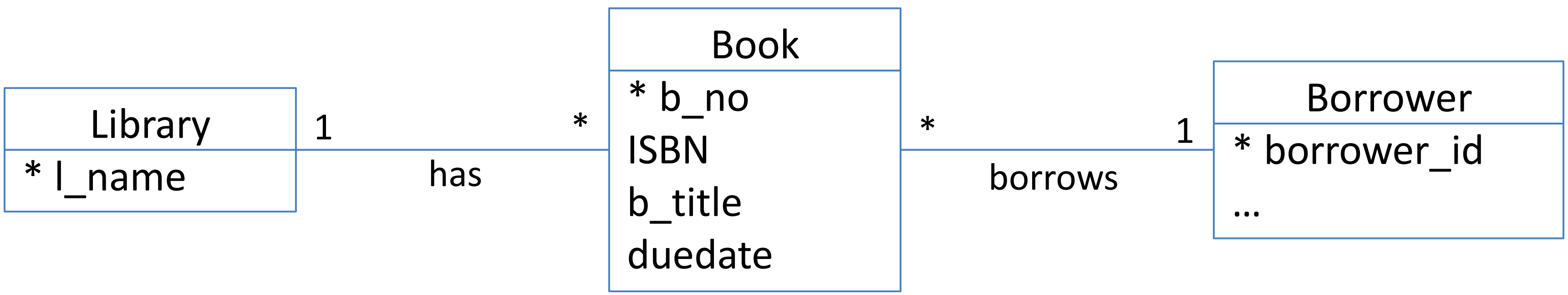
- Faithfully describes the world it is supposed to represent
- Relationships are of the correct degree
- Data model is complete, understandable, and accurate
- The data model makes sense to the client

Data model: Quality improvement

- Drawing a data model is an **iterative** process of trial and revision.
  - Is the level of detail correct?
  - Are all exceptions handled?
  - Is the model accurate?

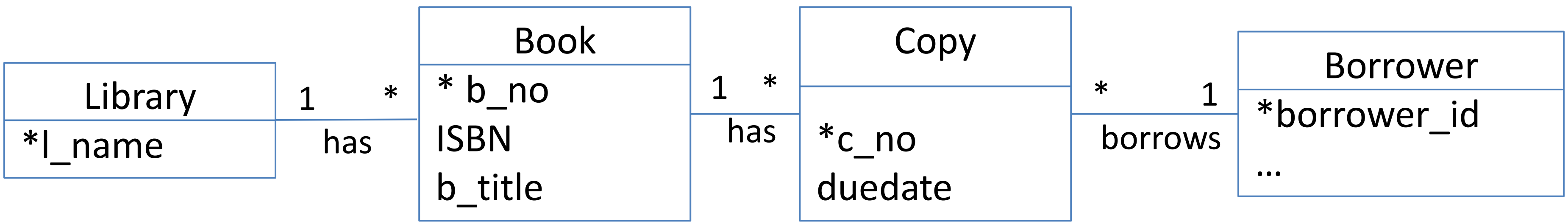
Data model Quality improvement: Library

- Consider the following data model for library.



- Is the above model an accurate model?
- What happens if the library has two copies of the book?
- Add an attribute to Book called *copy number*?
- ISBN vs b_no?

Revised library data model



Is this model an ***accurate*** representation of the data?

Hints on data modeling

- The model will expand and contract
- Invent identifiers(keys) where necessary
- Keys should have only one purpose – identification
- A data model does not imply ordering
- Create an attribute if ordering of instances is required
- An attribute’s meaning must be consistent

Hints on data modeling

- Single instance entities are OK
- Select names carefully
- Synonyms—different words have the same meaning
  - Get clients to settle on a common word or use views
- Homonyms—same word has different meanings
  - Clarify to avoid confusion
- Naming associative entities
  - Concatenate entity names if there is no obvious real world name

Hints on data modeling

- Uncover all exceptions
- Label relationships to avoid ambiguity
- Keep the data model well-formed and accurate

Making assumptions

- Data model is all about documenting rules and policies of an organisation.
- Database analyst should:
  - Identify and understand those rules that govern data
  - Represent those rules for information systems developers and users
  - Implement those rules in database technology
- Business rules can be gathered by *interviews* and *organisation documents*(policies, manuals, procedures etc)
- Sometimes a data analyst has to ask questions to clarify business rules.
- Occasionally a data analyst has to **make assumptions**.

Extra reading on business rules: Modern Database Management, Chapter 2, “Modelling the rules of the organization”

Making assumptions

- Different assumptions can result in different data models.
- Assumptions have to be sensible and reasonable.
- You must **clearly state the assumptions** if you make any – important for your coursework and exam!

Exercise 6

Create an ER model for the following descriptions:

- A cinema has multiple theatres. Each theatre has different seating capacity.
- Movies are shown through the day starting at 11am and finishing at 11am. A movie has its length, but each movie is given a two-hour time slot.
- One movie is never shown in more than one theatre at a time, but movies can be shifted among theatres because seating capacity varies.
- The cinema boss also want to store the data of how many people, classified by adults and children, attended each showing of a movie.
- Ticket prices vary by each show of movie. For example, X-men Apocalypse is £10 for everyone at 11am but is £15 at 9pm.

Clearly state the assumptions if you make any.

One possible solution