SGG 4653
Advance Database System

Database vs. GeoDatabase

Database Management System

- A software that allows the creation of database and manipulation and management of data in database.

- DBMS perform the tasks through commands (aka queries) written in Structured Query Language (SQL).

Structured Query Language

- A database language that allow user to:
  - Create database and relation structures
  - Perform basic data management tasks, such as insertion, modification and deletion of data from relations
  - Perform simple and complex queries

- 2 major components:
  - Data definition language (DDL)
  - Data manipulation language (DML)

SQL: DDL

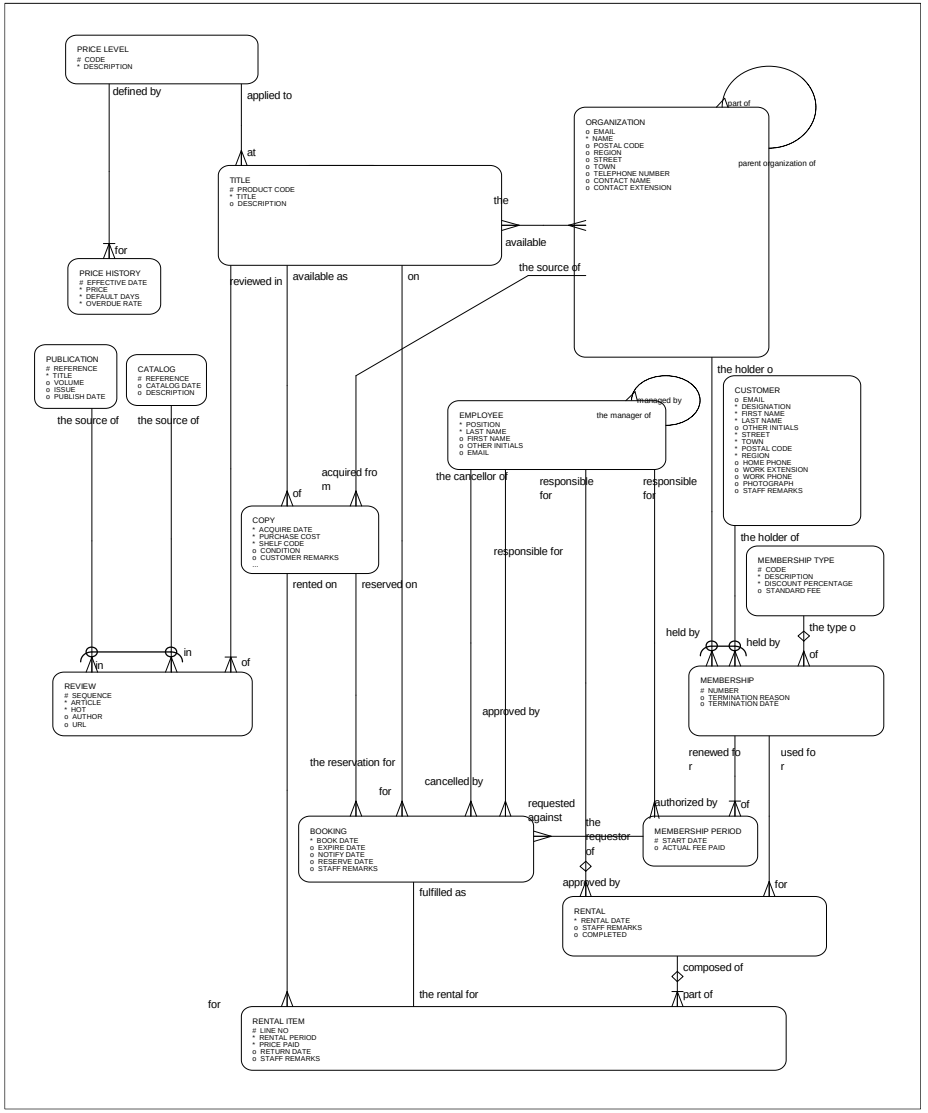
- Define database structure and controlling access to data
  - CREATE TABLE – to create table or database
  - ALTER TABLE – to modify structure of table that has been created
  - DROP TABLE – to delete table that has been created

SQL: DML

- Adding data into tables, to update data in tables, to delete data from tables and to display records in tables.
  - INSERT INTO
  - UPDATE SET
  - DELETE FROM
  - SELECT ... FROM
  - WHERE .... IN
  - WHERE .... LIKE
  - WHERE ..... IS NULL
  - ORDER BY
  - COUNT, SUM, AVG, MIN, MAX
  - GROUP BY
  - HAVING

Entity Relationship Modeling

- Models business, not implementation
- Is a well-established technique
- Has a robust syntax
- Results in easy-to-read diagrams…
- …although they may look rather complex at first sight



Entity Relationship Modeling

Goals:

- Capture *all* required information
- Information appears *only* once
- Model: *no* information that is derivable from other information already modeled
- Information is in a predictable, logical place

ER Modeling Notations

- UML notation
- Chen notation
- Crow's Feet notation

Concepts of the ER Model

- Entity types
- Relationship types
- Attributes

Relationship Types

- Relationship type
  - Set of meaningful associations among entity types.
  - Express how entities are mutually *related*
  - Always exist between *two* entities (or one entity *twice*)

- Relationship occurrence
  - Uniquely identifiable association, which includes one occurrence from each participating entity type.

Relationship Examples

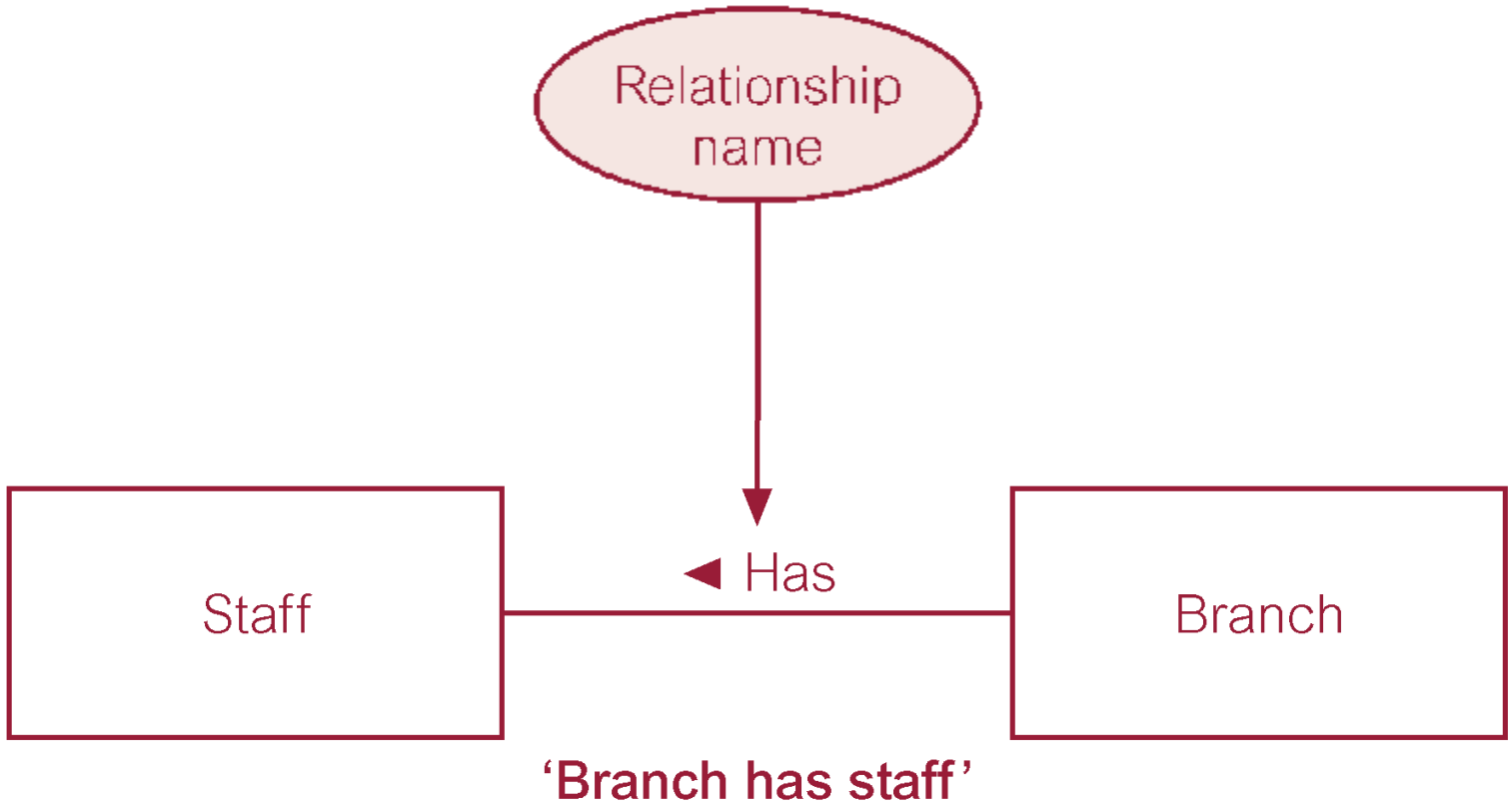
- BRANCH has STAFF
  STAFF belongs to BRANCH

- EMPLOYEES *have* JOBS
  JOBS *are held by* EMPLOYEES

- PEOPLE *make* TICKET RESERVATIONS
  TICKET RESERVATIONS *are made by* PEOPLE

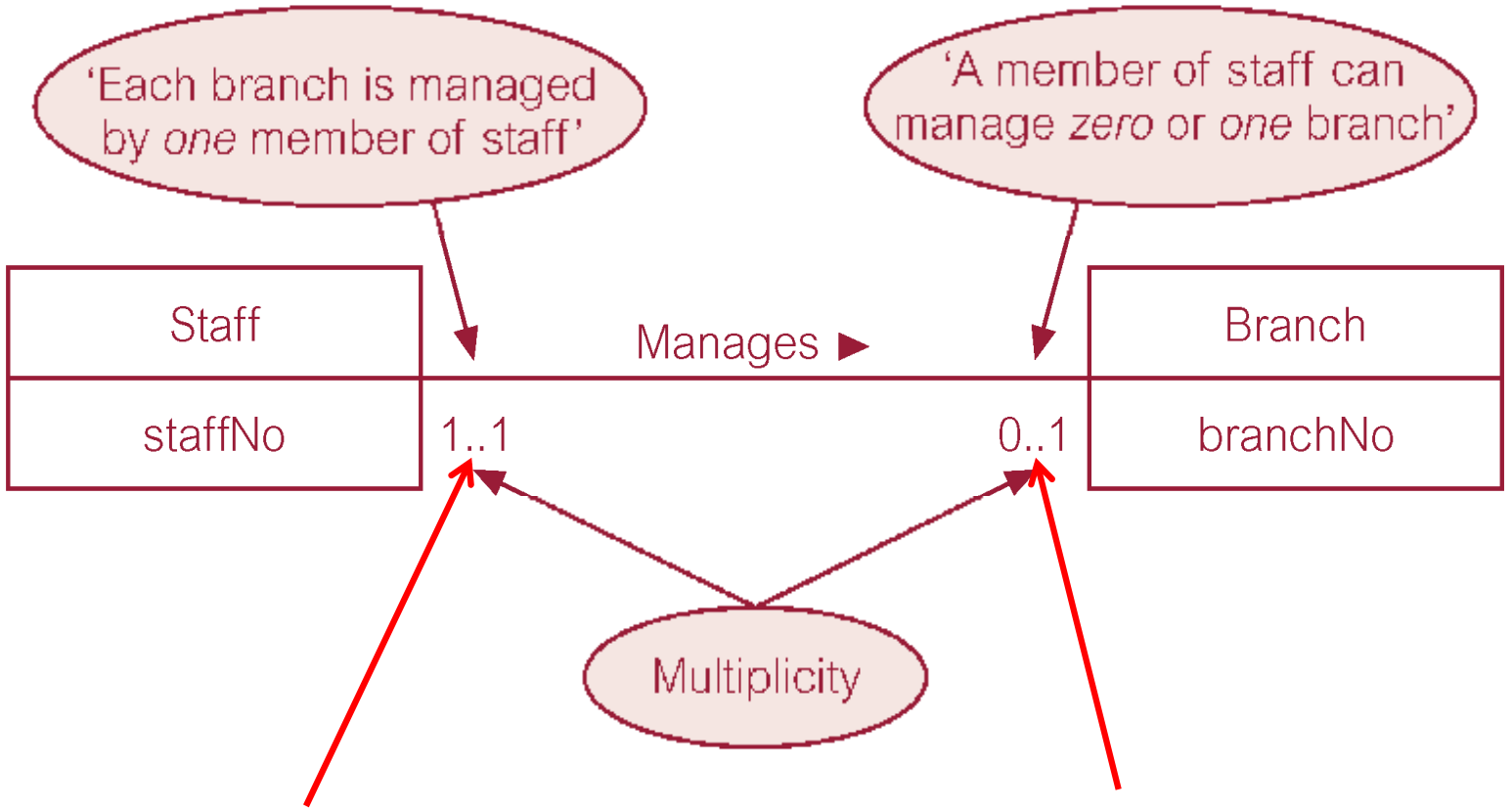
ERD: Branch *Has* Staff Relationship

UML NOTATION:



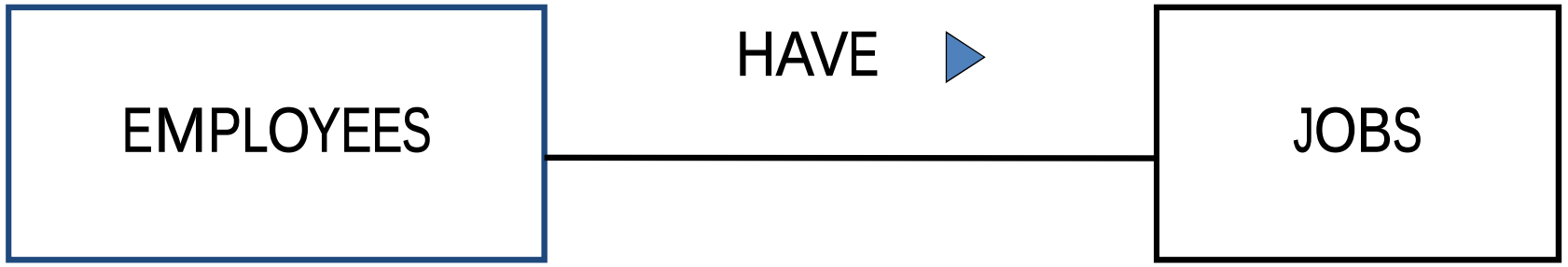
'Branch has staff'

ERD: Staff Manages Branch

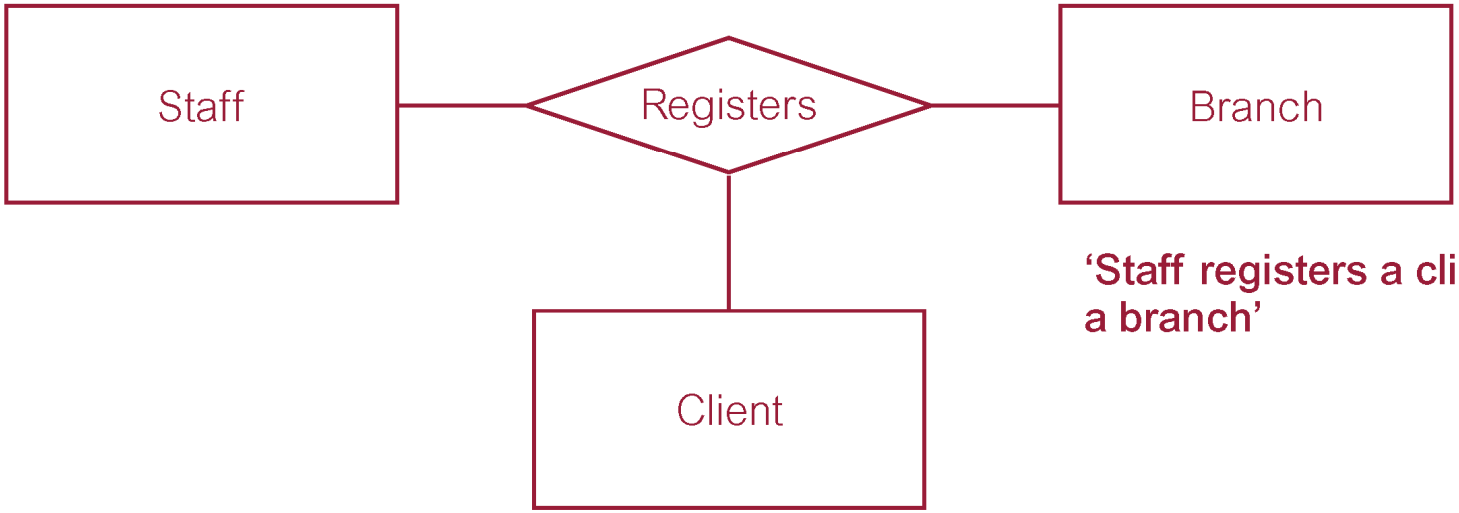


Manages relationship is a one-to-one (1:1)relationship **à** *shown through the maximum range value on the multiplicities at both ends of relationship*

ERD: Employees Have Jobs

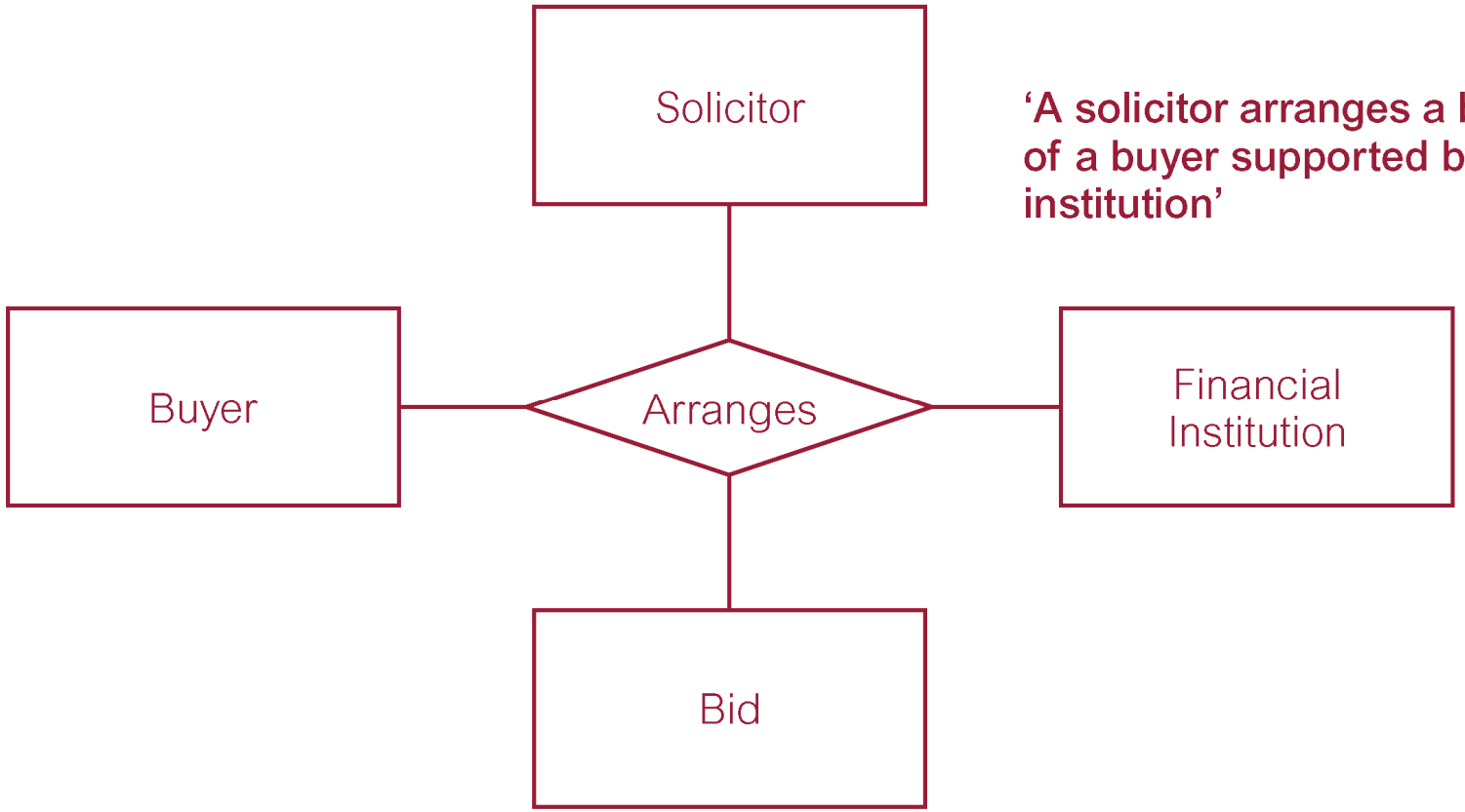


Ternary Relationship called *Registers*



'Staff registers a client at a branch'

Quaternary Relationship called *Arranges*

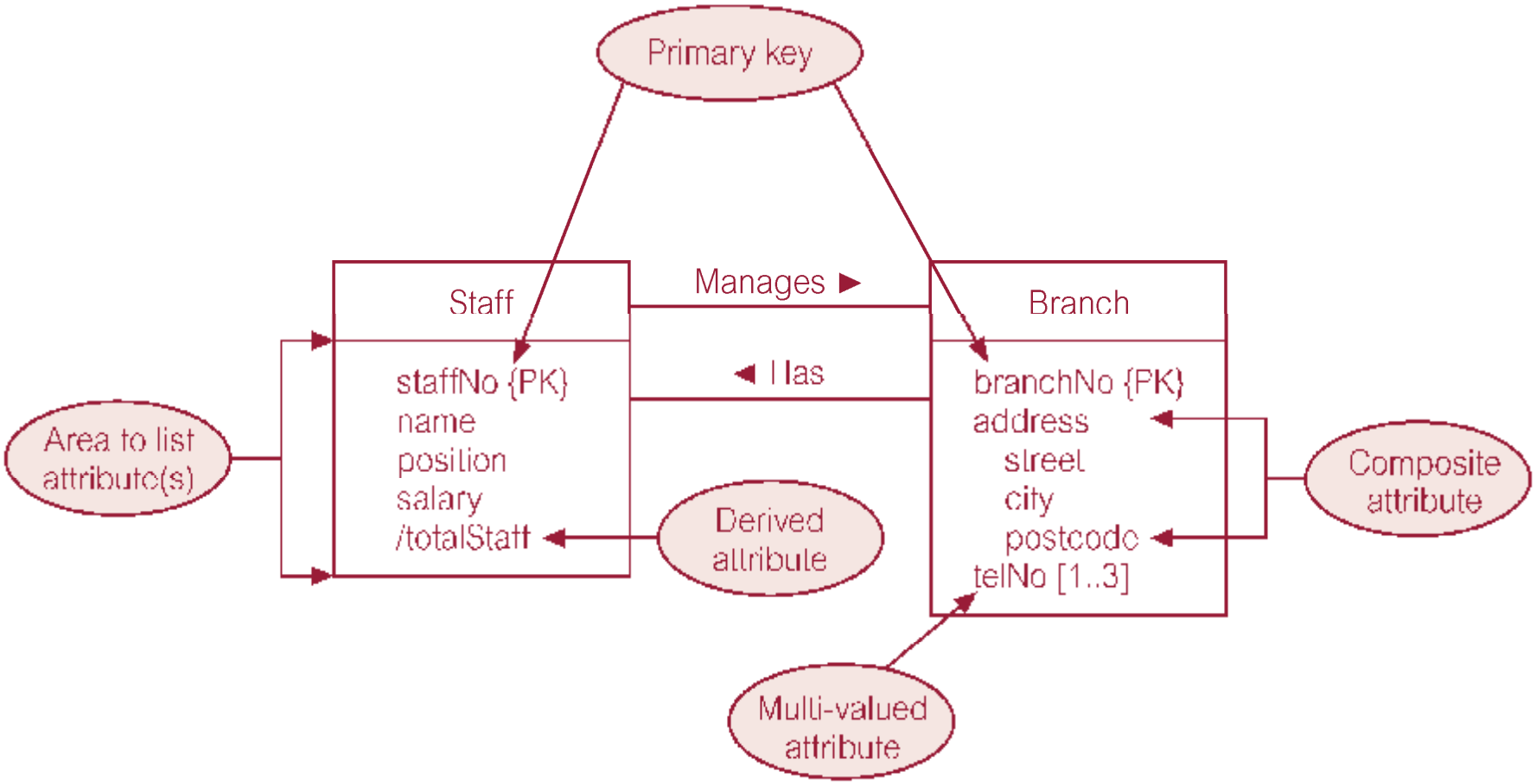


'A solicitor arranges a bid on behalf of a buyer supported by a financial institution'

Attribute Examples

Entity	Attribute
EMPLOYEE	Family Name, Age, Shoe Size, Town of Residence, Email, ...
CAR	Model, Weight, Catalog Price, ...
ORDER	Order Date, Ship Date, ...
JOB	Title, Description, ...
TRANSACTION	Amount, Transaction Date, ...
EMPLOYMENT CONTRACT	Start Date, Salary, ...

ER Diagram of Staff and Branch Entities and their Attributes



Database Design (DBLC)

Conceptual

- Process of constructing a model of information used in an enterprise, independent of all physical considerations.
- Entity Relationship Model; Data dictionary

Logical

- **Process of constructing a model of information used in an enterprise based on a specific data model (e.g. relational), but independent of a particular DBMS and other physical considerations.**
- Normalized database (relational) schemas

Physical

- **Process of producing a description of the implementation of the database on secondary storage; it describes the base relations, file organizations, and indexes design used to achieve efficient access to the data, and any associated integrity constraints and security measures.**

Normalization

- A technique for producing a set of relations with desirable properties, given the data requirements of an enterprise.
- The characteristics of a suitable set of relations include the following:
  - the minimal number of attributes necessary to support the data requirements of the enterprise.
  - attributes with a close logical relationship (describes as functional dependency) are found in the same relation.
  - minimal redundancy with each attribute represented only once with the important exception of attributes that form all or part of foreign keys, which are essential for the joining of related relations.

Benefit of Normalization

- Minimize data redundancies in a database, thus will reduce storage space required to store the data
- Reduce data anomalies
- Easy for user to access data from a database
- Easy for user to maintain data

Data Redundancy

Staff

staffNo	sName	position	salary	branchNo
S0001	Majid Ahmad	Manager	25000	B005
S0002	Aleesha Tan	Assistant	14000	B003
S0003	Raj Kumar	Supervisor	17000	B003
S0004	Peter Lim	Assistant	7300	B007
S0005	Syed Adhadi	Manager	29000	B003
S0006	Imran Hakeem	Assistant	8100	B005

Branch

branchNo	bAddress
B003	10, Jalan Skudai
B005	5, Taman Perling 2
B007	222, Jalan Petaling

Figure 1

vs.

StaffBranch

staffNo	sName	position	salary	branchNo	bAddress
S0001	Majid Ahmad	Manager	25000	B005	5, Taman Perling 2
S0002	Aleesha Tan	Assistant	14000	B003	10, Jalan Skudai
S0003	Raj Kumar	Supervisor	17000	B003	10, Jalan Skudai
S0004	Peter Lim	Assistant	7300	B007	222, Jalan Petaling
S0005	Syed Adhadi	Manager	29000	B003	10, Jalan Skudai
S0006	Imran Hakeem	Assistant	8100	B005	5, Taman Perling 2

Figure 2

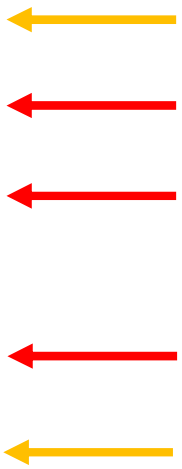
- Information on staff in relation Staff
- Information on branch in relation Branch

- Both information on staff and branch are in relation StaffBranch

Update Anomalies

StaffBranch

staffNo	sName	position	salary	branchNo	bAddress
S0001	Majid Ahmad	Manager	25000	B005	5, Taman Perling 2
S0002	Aleesha Tan	Assistant	14000	B003	10, Jalan Skudai
S0003	Raj Kumar	Supervisor	17000	B003	10, Jalan Skudai
S0004	Peter Lim	Assistant	7300	B007	222, Jalan Petaling
S0005	Syed Adhadi	Manager	29000	B003	10, Jalan Skudai
S0006	Imran Hakeem	Assistant	8100	B005	5, Taman Perling 2



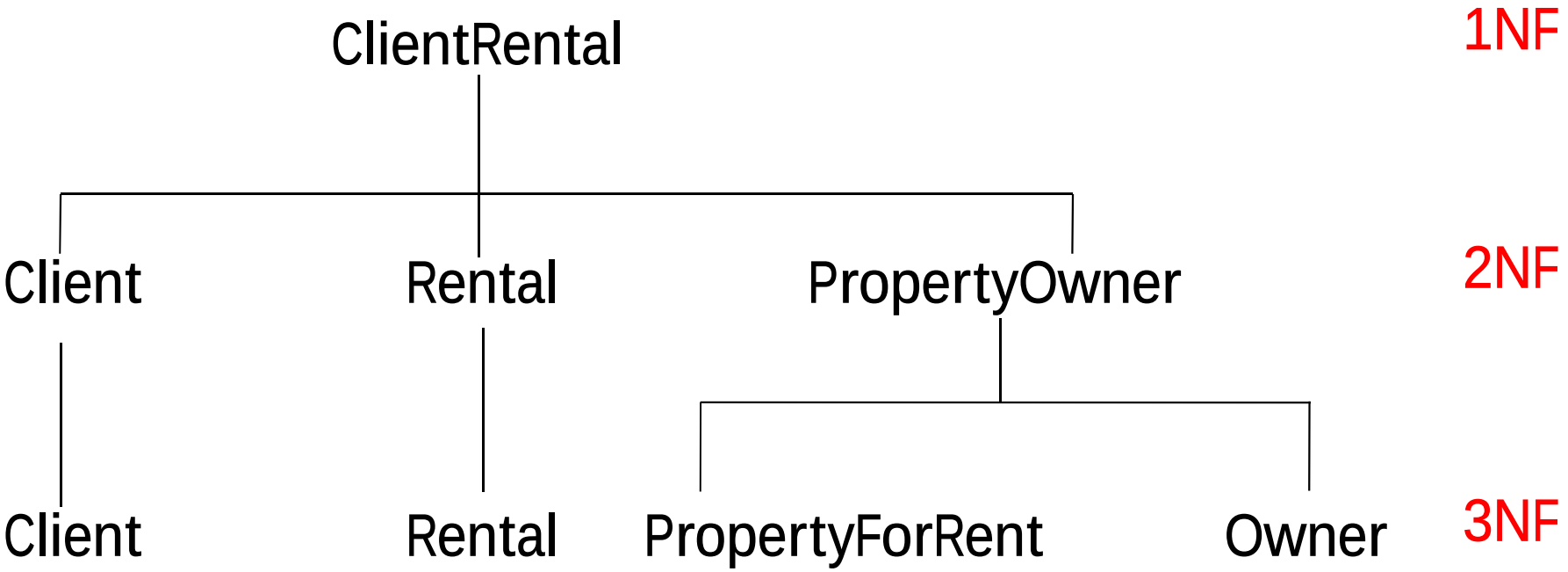
Unwanted data redundancy may cause update anomalies

Look at tuples in row 2, 3 and 5. The details of a branch (B003) are repeated for every member of staff located at that branch

à data redundancy, hence defeat one of the purpose of building a database.

Category of update anomalies : insertion, deletion and modification

Example: 1NF à 3NF



Geodatabase

- Geographic objects (spatial) + database
- Store:
  - Attributes
  - Topological information (feature class)
  - Spatial geometry (point, line, polygon)
  - Spatial Reference System (Coordinate System, Datum)