Normalisation

Business Data Management and Analytics

Contributions by Arthur Adamopoulos and Vince Bruno

Normalisation

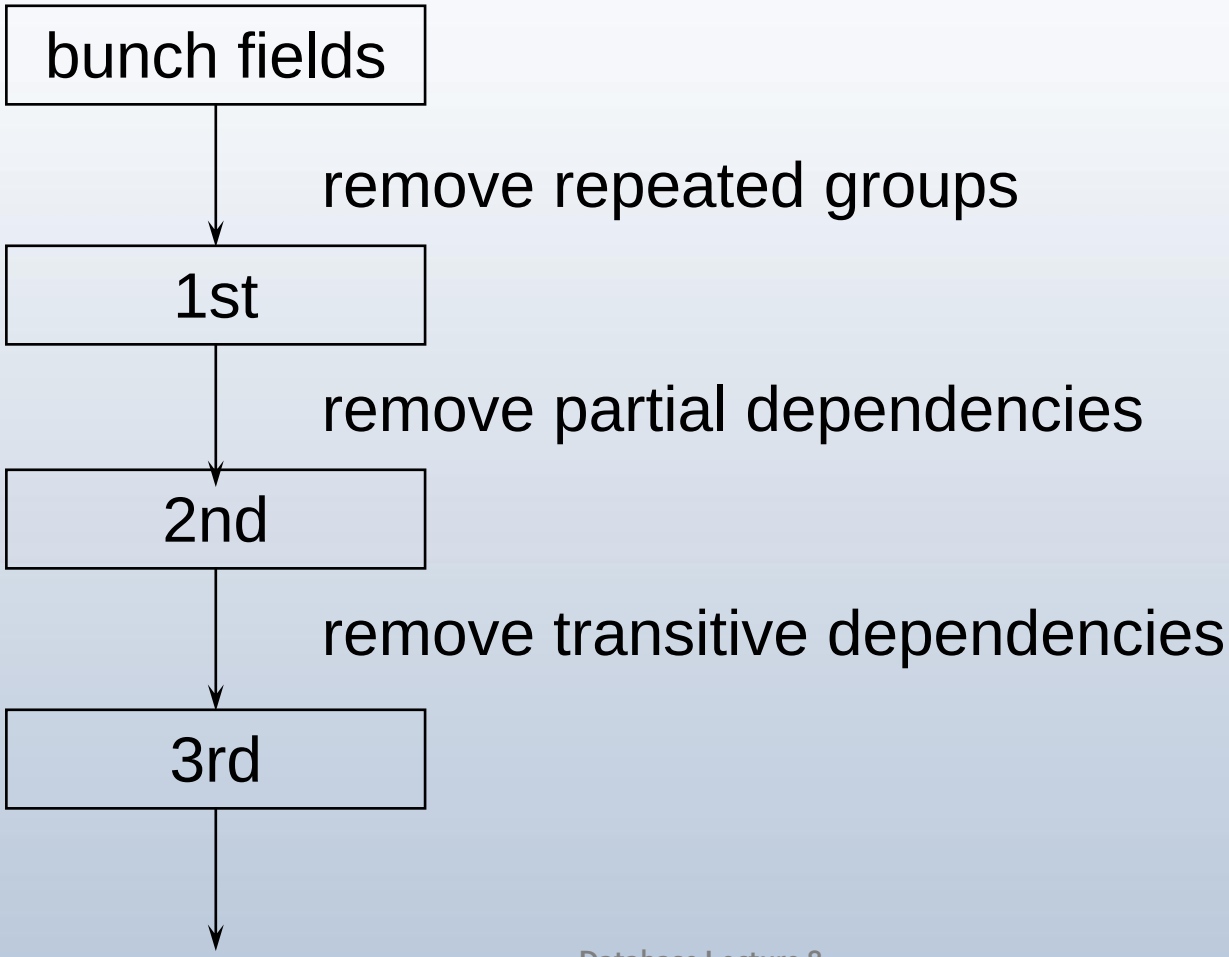
- Normalisation [none,1st,2nd,3rd forms]
- Functional Dependence
- Partial/Transitive dependencies
- Primary Keys
- Referential Integrity
- EERD

Normalisation - Steps

- Each stage is a normal form

- Normal forms relate by applying simple rules about dependencies.

Normalisation - steps

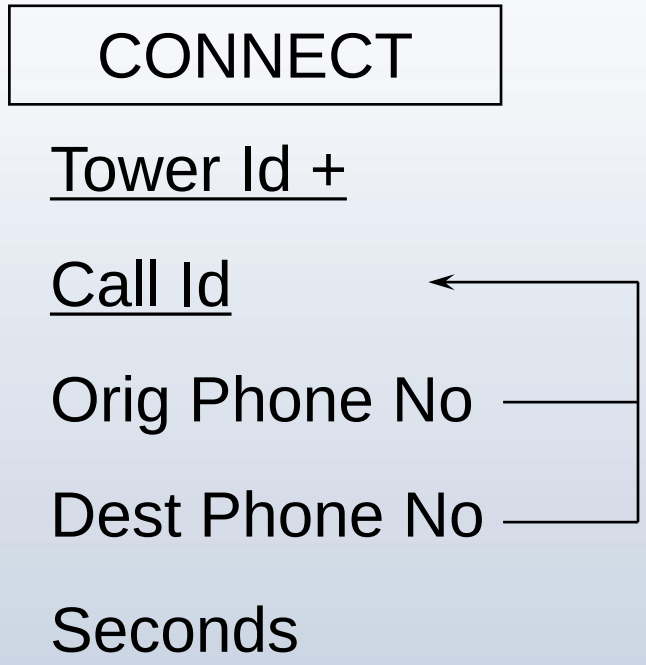


Functional Dependence

- A relationship between two attributes
- One field is dependant on another, the first field value would not come into existence unless the second field value does.
- IF A depends on B there is only 1 value for A for each value of B
- A only exists when B exists first.
- B ---> A

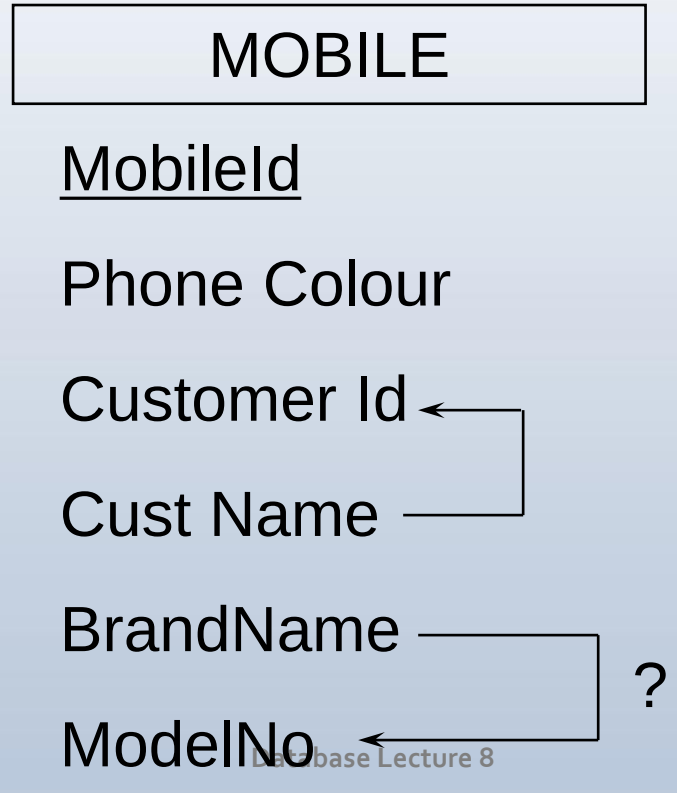
Partial Dependency

- A field that depends on part of the primary or candidate key.
- If tables do NOT have multi-part keys, NO partial dependency can exist.



Transitive Dependency

- A field that depends on a NON-Key field(s).



Zero Normal Formal

MOBILE

- Mobile Id
- PhoneNumber
- BrandName
- Cust Surname
- Cust Given
- Dob
- Joined
- Cancelled
- PlanName
- ConnectFee
- PeakFee

CALL

- Call Id
- Mobile Id
- Dest Phone No
- Call Date
- Call Time
- Call Duration
- Tower1
- Location 1
- Seconds 1
- Tower 2
- Location 2
- Seconds 2
- Tower 3
- Location 3
- Seconds 3

1st Normal Formal

MOBILE

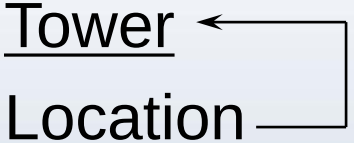
- Mobile Id
- PhoneNumber
- BrandName
- Cust Surname
- Cust Given
- Dob
- Joined
- Cancelled
- PlanName
- ConnectFee
- PeakFee

CALL

- Call Id
- Mobile Id
- Dest Phone No
- Call Date
- Call Time
- Call Duration

CONNECT

- Call Id
- Tower
- Location
- Seconds



PD

2nd Normal Formal

MOBILE

- Mobile Id
- PhoneNumber
- BrandName
- Cust Surname
- Cust Given
- Dob
- Joined
- Cancelled
- PlanName
- ConnectFee
- PeakFee

TD

TD

CALL

- Call Id
- Mobile Id
- Dest Phone No
- Call Date
- Call Time
- Call Duration

CONNECT

- Call Id
- Tower Id
- Seconds

TOWER

- Tower Id
- Location

3rd Normal Formal

MOBILE

- Mobile Id
- PhoneNumber
- BrandName
- Joined
- Cancelled
- Cust Id
- PlanName

CALL

- Call Id
- Mobile Id
- Dest Phone No
- Call Date
- Call Time
- Call Duration

CONNECT

- Call Id
- Tower Id
- Seconds

CUSTOMER

- Cust Id
- Cust Surname
- Cust Given
- Dob

PLAN

- PlanName
- ConnectFee
- PeakFee

TOWER

- Tower Id
- Location

Other Normal Forms

- Boyce-Codd Normal Form
  - remove remaining anomalies resulting from functional dependencies;
- Fourth Normal Form
  - remove anomalies that result from a multi-valued dependencies;
- Fifth Normal Form
  - designed to cope with dependency known as join dependency.

Keys

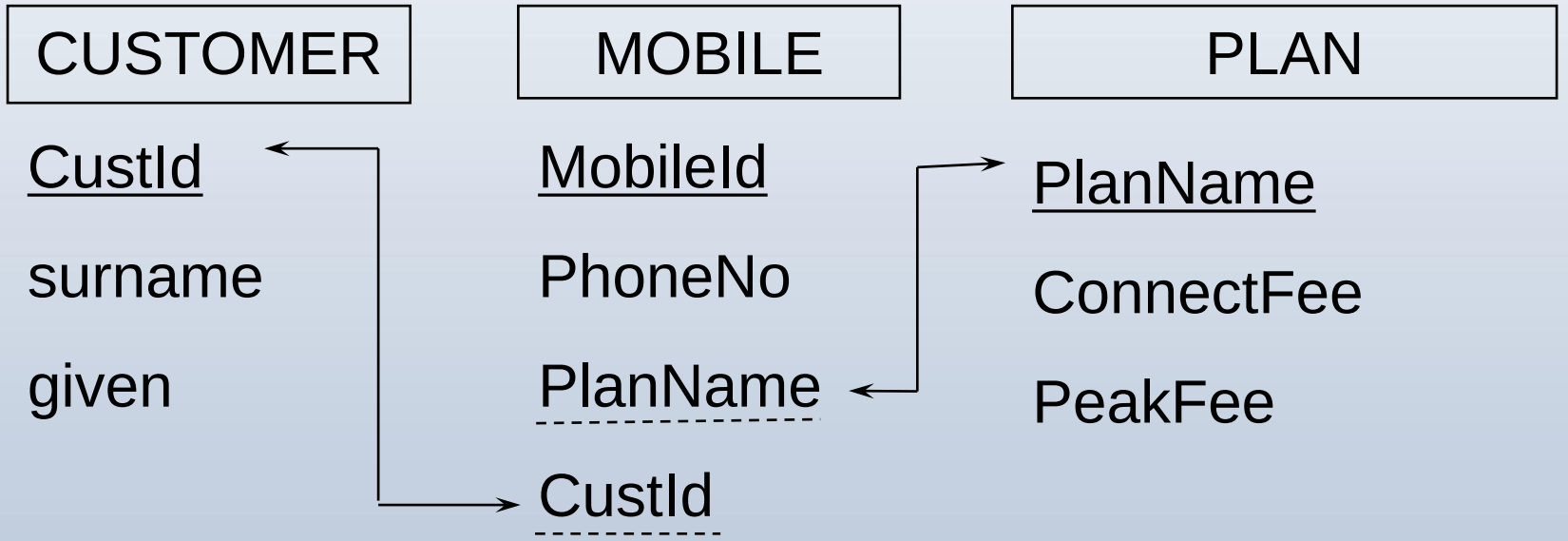
- Primary Key
  - Unique identifier (field or fields) of a table
  - properties of a primary key are:
    - Uniqueness
    - Availability
    - Stability
    - Minimality
- Candidate Key
  - A field or fields that could be a primary key
- Composite Key
  - a primary key containing more than one field.

Constraints - Domain

- input checking
  - type
  - length
  - formats
  - allowable values
  - min/max ranges
  - Optional/mandatory

Referential Integrity

- related to foreign keys only
- values in foreign key must exist in primary key of related file.

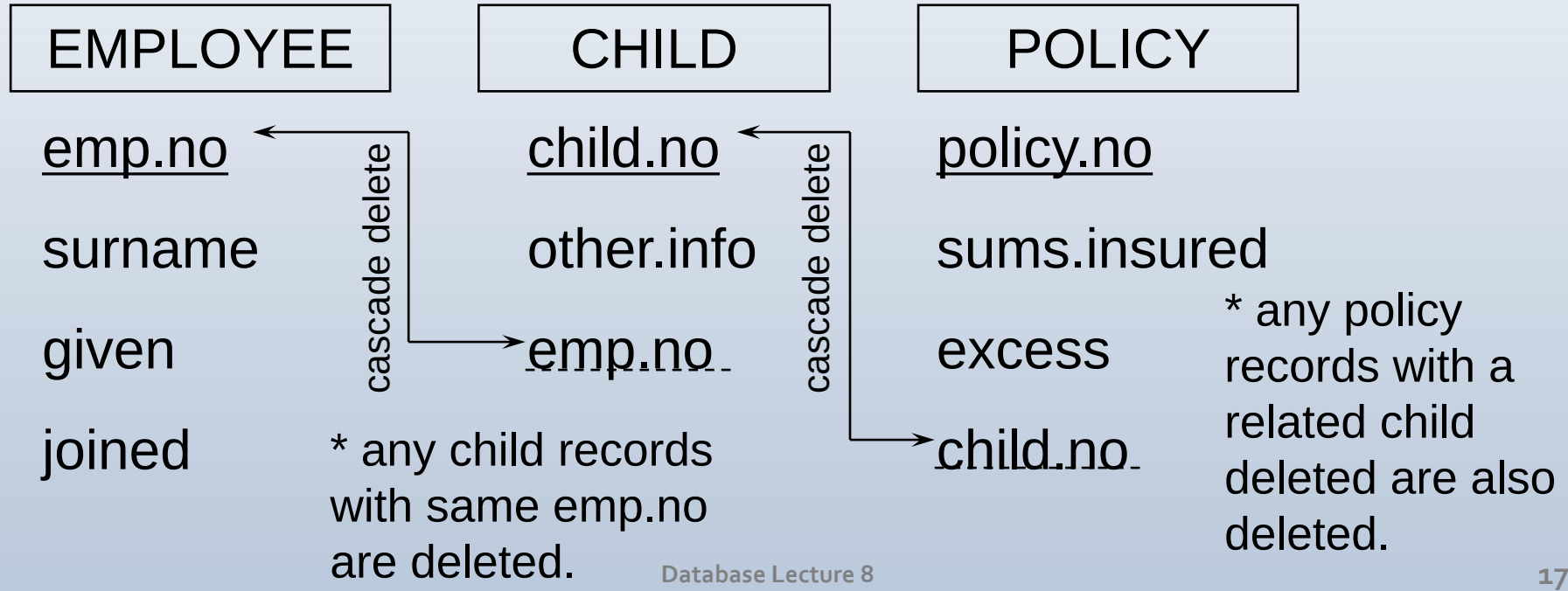


RI - Other Issues

- Insert/Update
  - value inserted/change in foreign key must already exist in primary key of other file.
- Delete - three options:
  - not allow;
  - null out the corresponding foreign key(s);
  - cascade delete - remove entire record and any related foreign keys.

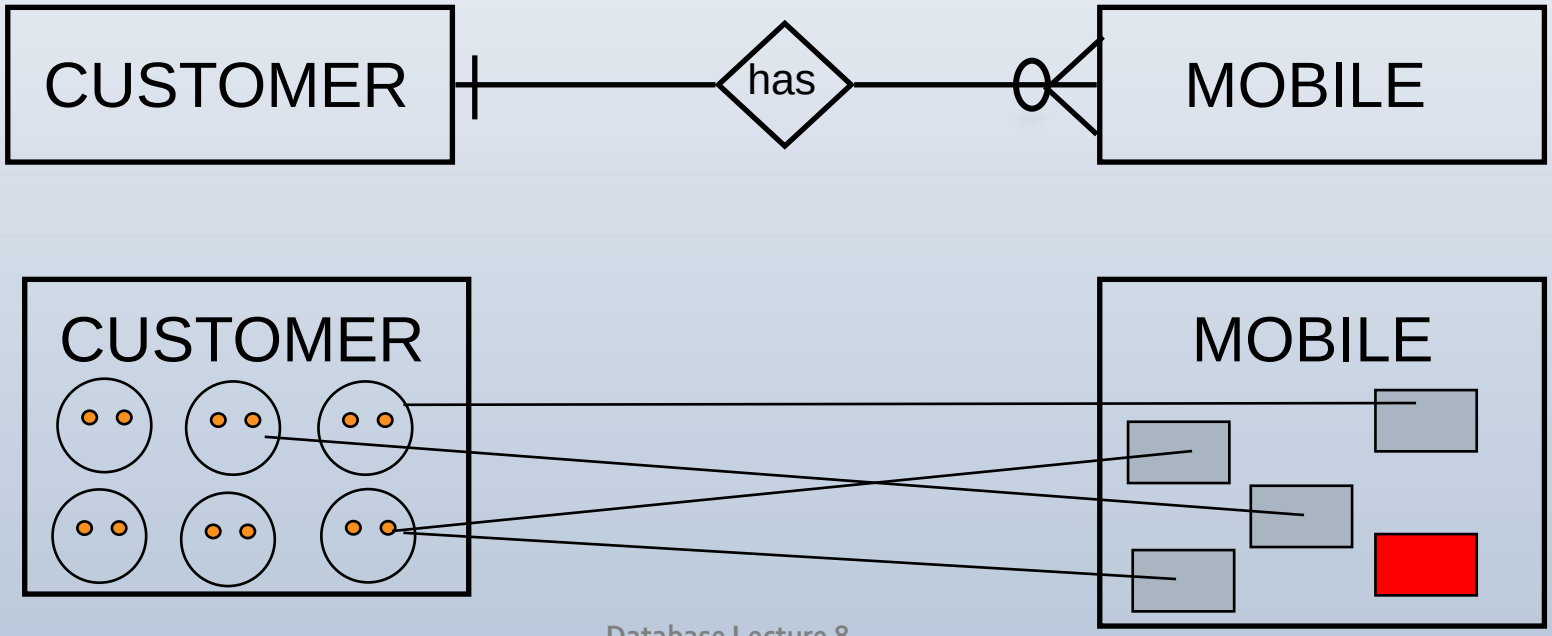
RI - Cascade Delete

- A delete is issued to delete an employee from the EMPLOYEE file.



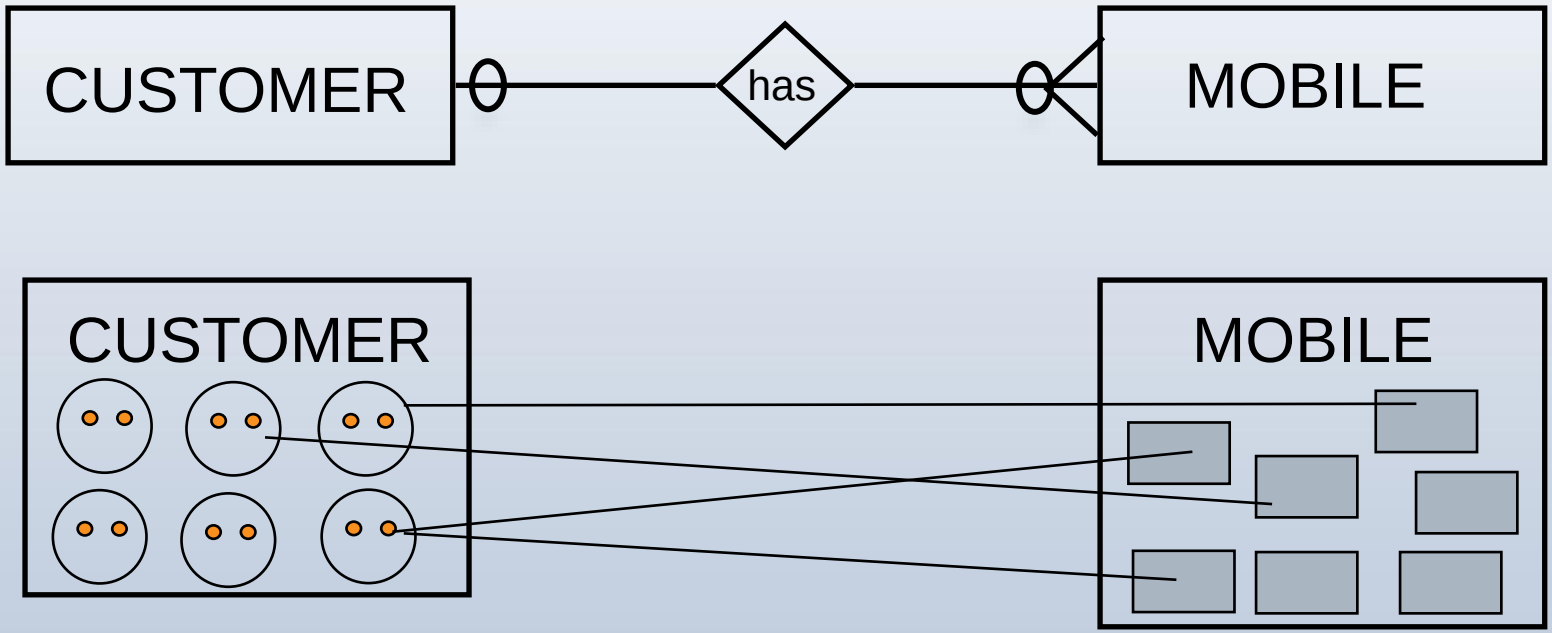
Relationships

- One other aspect to relationships, which can further define the allowed associations between entities is:
  - Mandatory – at least one association must exist



Relationships

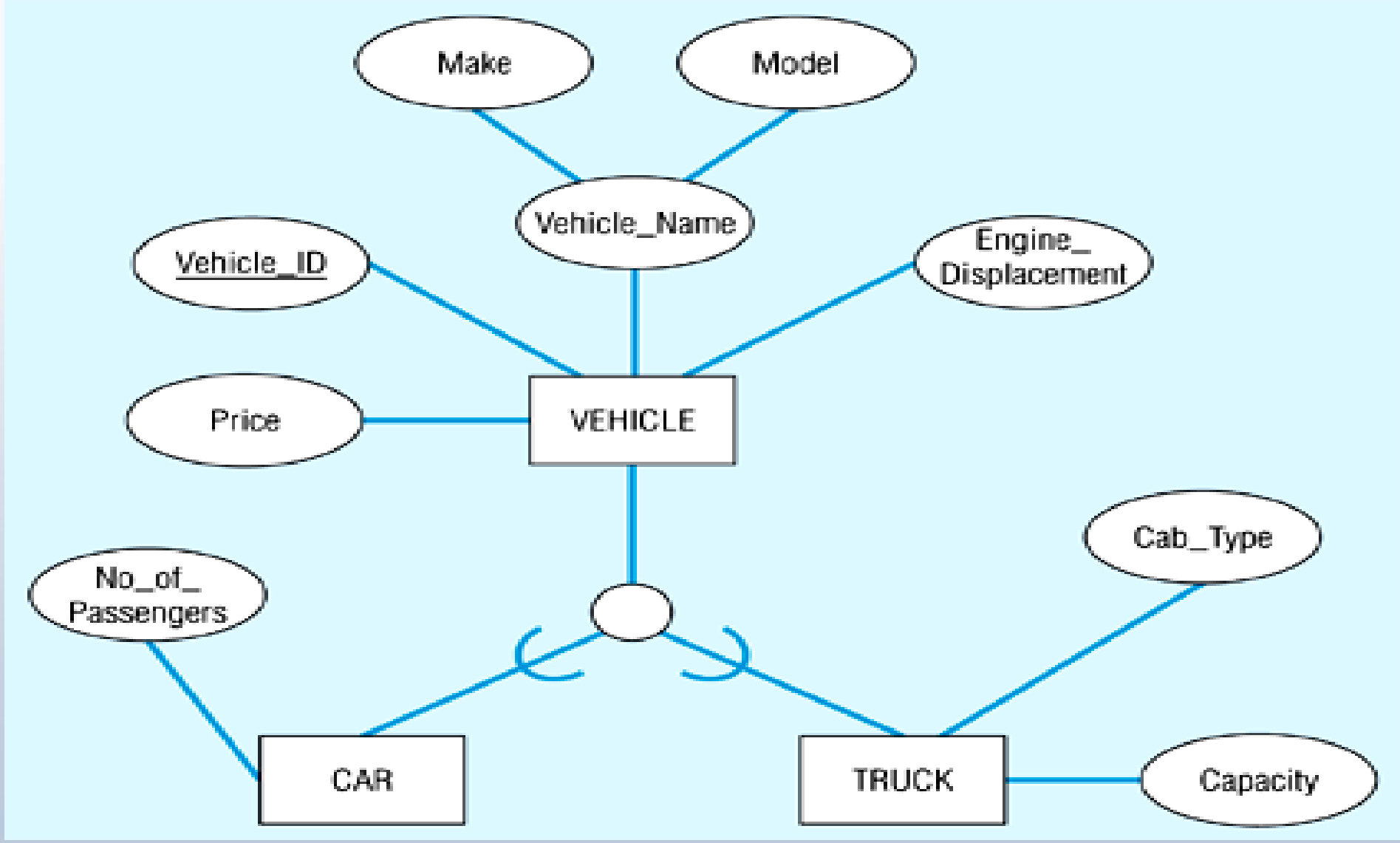
- Optional – zero or more association can exist



Enhanced ER Model (EER)

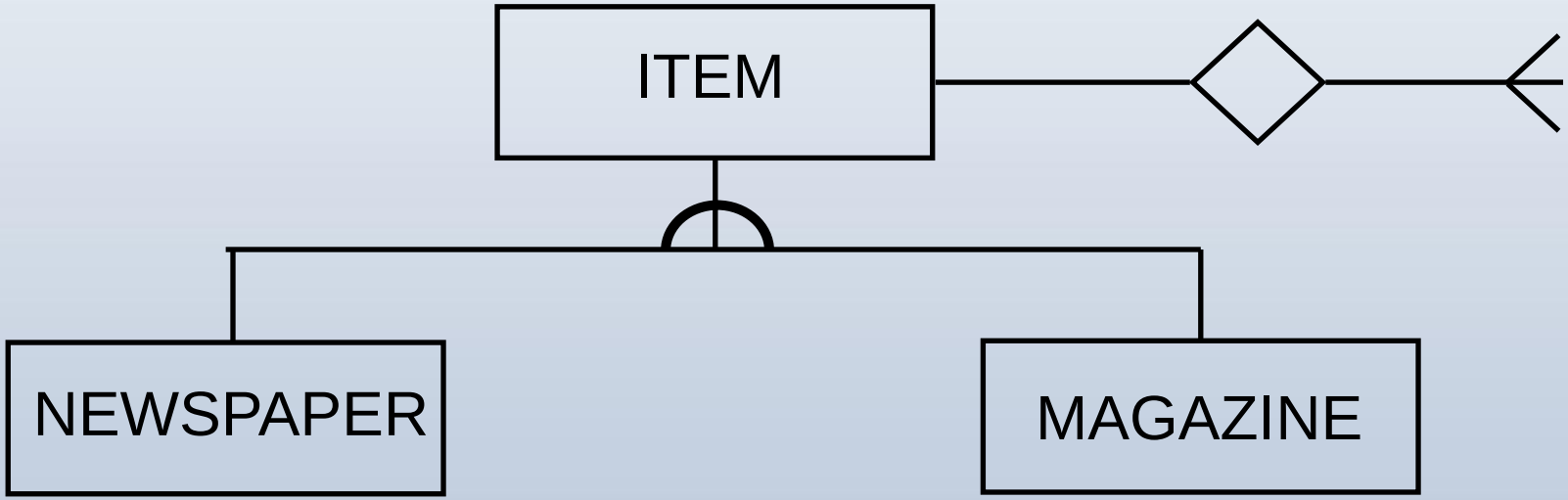
- Super and Sub types
- The ER model has been enhanced by various people to include inheritance.
- Concepts have simply been borrowed from the Object Oriented model but drawn differently.
- Both supertypes and subtypes can participate in relationships.
- Page 153 in the Hoffer (textbook) describes the various ways this is represented.

EER Model Super/Sub Types

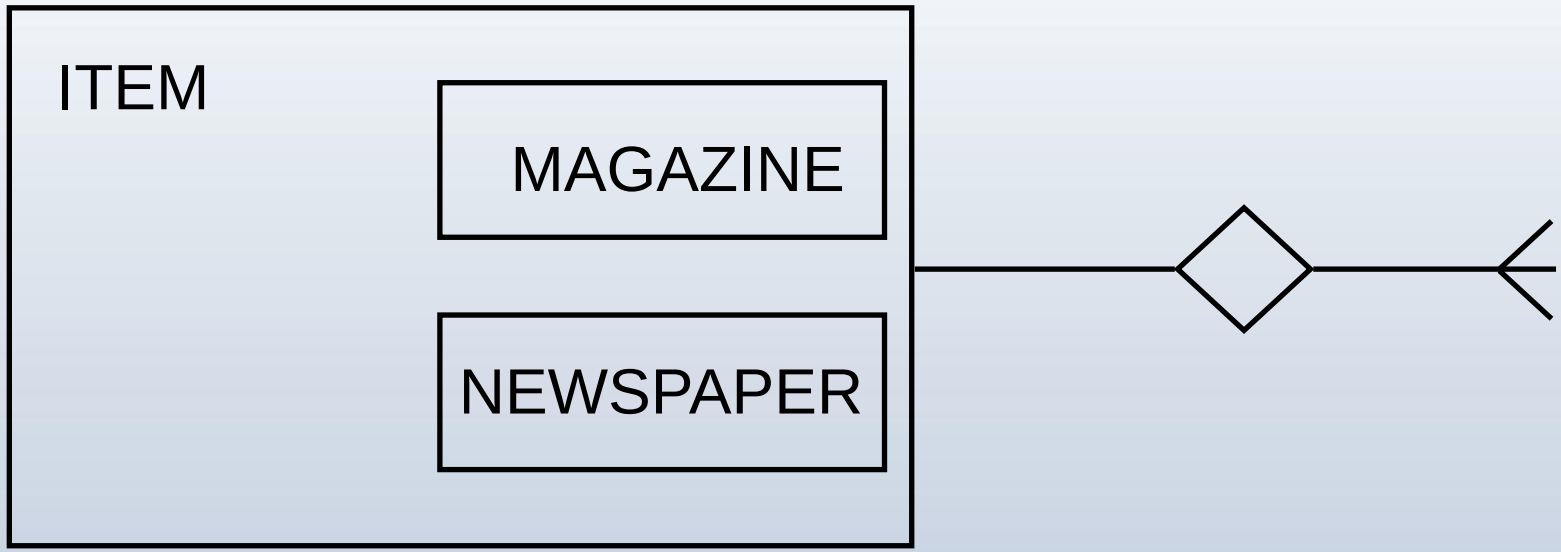


EER: Example PEOPLE

- An ITEM delivered by newsagent has a primary key of ITEM No and each of the sub-types have their own primary key.



EER: Example PEOPLE



EER Model Super/Sub Types

- Three options to translate super/sub types into a relational model
  - Keep all levels
  - Roll Up
  - Roll Down
- Hoffer (text) describes keep all level transformation, page 224-225

EER: Example PEOPLE

- Keep All Levels

ITEM	NEWSPAPER	MAGAZINE
item no+	newspaper no	magazine no
type	Field1	Field3
Name	field2	field4
frequency		
price		

EER: Example PEOPLE

- Roll up

ITEM

item no

type

Name

Frequency

Price

Field1-4

EER: Example PEOPLE

- Roll down

NEWSPAPER	MAGAZINE
newspaper no	magazine no
Name	Name
Frequency	Frequency
Price	Price
Field1	Field3
field2	field4

Exercise 1: Normalise

- What is wrong with the following set of tables, designed for the Kangaroo Holiday Park ? How would you redesign them ?

CABINS

Cabin No
PeopleCatered
Ensuite (Y/N)
Tourist_No
Kitchen (Y/N)
Date Arriving
Date Leaving
Type Of Shelter

TOURIST

Tourist No
Tourist Name
Tourist Address
Tourist Phone
Date Booked
Cabin_No1
Cabin_No2
Cabin_No3
Total Cost
Duration Of Stay

Exercise 2: Normalise

- What is wrong with the following set of tables, designed for a small antique book library? How would you redesign them?

BOOK

Catalog_No
Author
Title
Borrower_No
Borrower_Address
Date_Due

BORROWER

Borrower_No
Borrower_Name
Borrower_Address
Catalog_Numbers*
Fines_Owing
Fines_Paid
Date_Returned