

SIT103/SIT772 Data and Information Management

Week 3

Entity Relationship Diagram (ERD)

Sunil Aryal

Last Week



- Database design Conceptual and logical design
- Relational model Entity, Attribute, Relationships, and constraints
- Keys (composite, super, candidate, primary, natural, surrogate, foreign, secondary)
- Integrity Rules Entity and Referential Integrity
- Entity Relationship Diagram

Last Week's OnTrack Task



- 2.1P Database Modelling Tools
 - Basics of relational database modelling
 - To familiarise you with modelling tools
 - LucidChart and MS Visio

This task is due this Friday!

Questions?



Any questions/comments so far

Last week's content

OnTrack tasks

Anything in general about the unit

This week



- More on conceptual and logic design
 - Entity Relationship Diagram (ERD)
- Some case studies

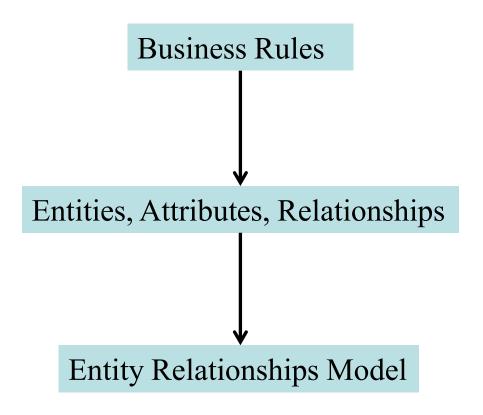
Entity Relationships Model (Recap)



- Entity: person, place, thing, or event about which data will be collected and stored
 - e.g., Student, Course, Product, Order, Transaction, etc.
- Attribute: characteristic of an entity
 - e.g., ID, Name, DoB, Address, etc.
- Relationship: association among entities
 - One-to-one (1:1 OR 1..1)
 - One-to-many (1:M OR 1..*)
 - Many-to-many (M:N OR *..*)
- Constraint: restriction placed on data
 - Ensures data integrity, e.g., Unique, Not NULL, etc.

Entity Relationships Model (2)





Attributes and domain



- Required attribute: must have a value (NOT NULL)
- Optional attribute: may be left empty (NULL is allowed)
- Domain: set of possible values for an attribute
 - States of Australia {ACT, NT, NSW, QLD, SA, TAS, VIC, WA}
 - Height, Weight decimal numbers
 - Age integer numbers

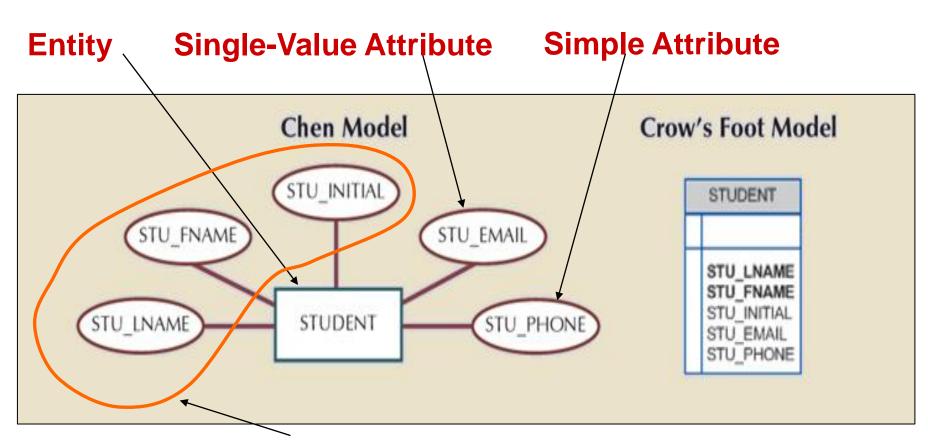
Types of Attributes



- Composite attribute <u>can</u> be further <u>subdivided</u>
- Simple attribute *cannot* be further *subdivided*
- Single-valued attribute can only have a <u>single value</u> at a particular instance of time *e.g.*, a person has one weight
- Multi-valued attribute can have <u>many values</u>, e.g., a person can have several <u>aliases</u>, or multiple <u>contact numbers</u> or several <u>skills</u>, or several <u>qualifications</u>, and so on

Types of Attributes (2)





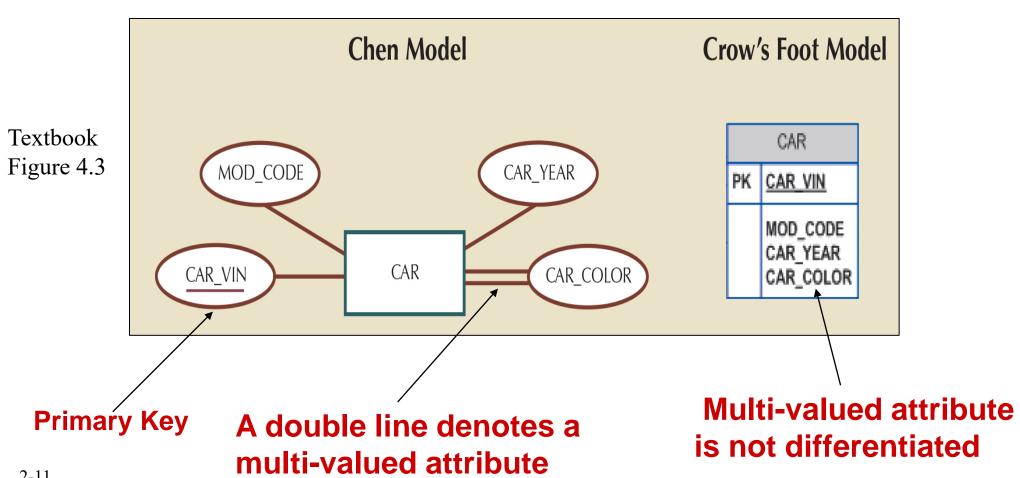
Composite Attribute STU_NAME can be subdivided into these three simple attributes

Textbook Figure 4.1

Multi-valued Attribute



Car's color – body color, roof color, and trim color



Multi-valued Attribute (2)

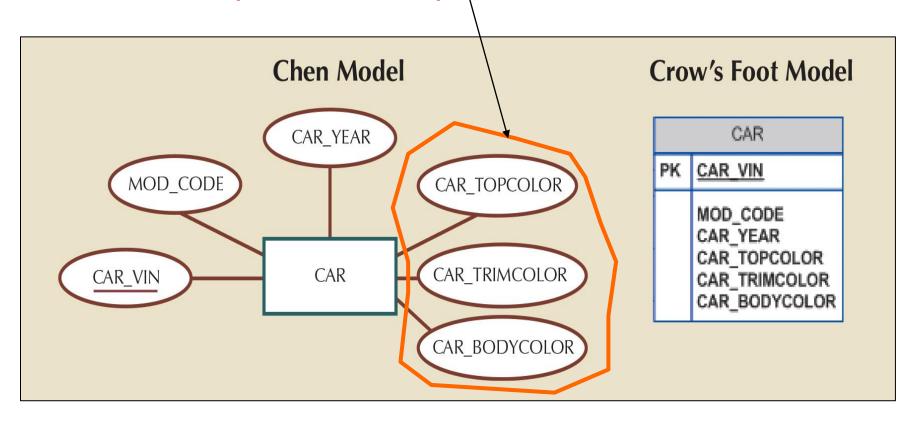


- Problematic we should avoid them
 - Remember, in relational model
 - "Each row and column intersection must represent a single data value"
- Two possible solutions
 - Create new attributes one for each of the original multivalued attribute's components
 - Create a new entity composed of original multi-valued attribute's components

Multi-valued Attribute (3)



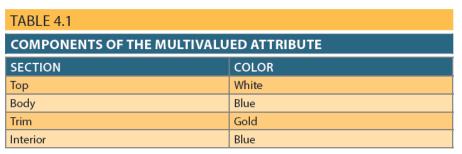
Splitting the multi-valued attribute (CAR_COLOR) into three attributes



Textbook Figure 4.4

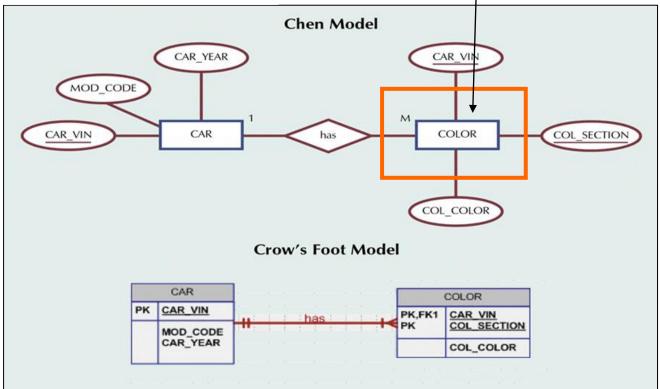
Multi-valued Attribute (4)





A new entity from a multi-valued attribute

Textbook Table 4.1



Textbook
Figure 4.5
(bottom figure)

13-14

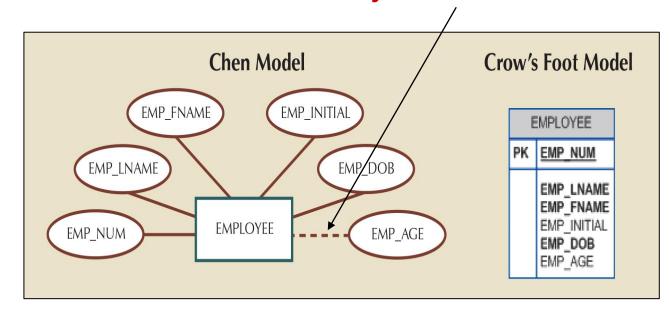
Derivable Attributes



- An attribute whose value may be calculated (derived) from other attributes
 - Need not be physically stored in the DB
 - Can be derived when needed

A dashed line also depicts a derived attribute, alternative symbol in Chen's notation

Textbook Figure 4.6



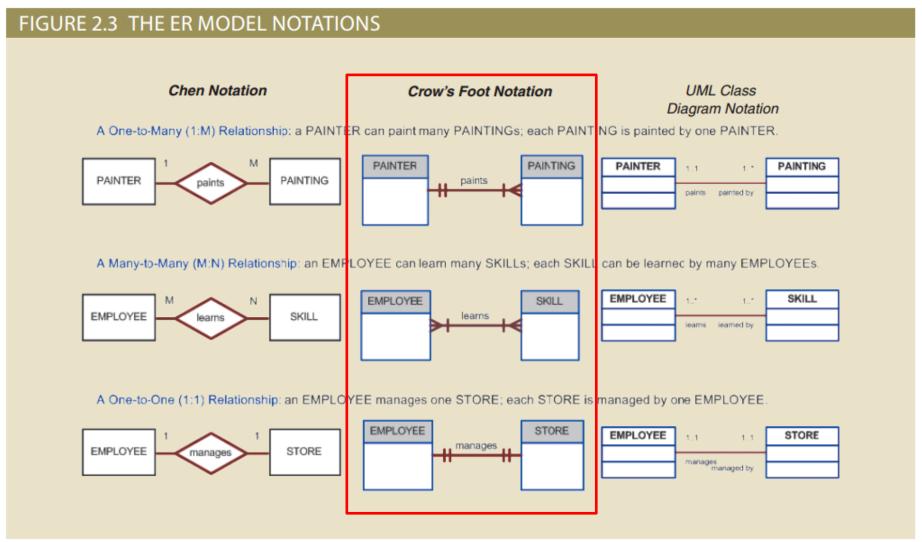
Should we store derived attributes?



Table 4.2	Stored	Not Stored
Advantage	Saves CPU processing cycles Saves data access time Data value is readily available Can be used to keep track of historical data	Saves storage space Computation always yields current value
Disadvantage	Requires constant maintenance to ensure derived value is current, especially if any values used in the calculation change	Uses CPU processing cycles Increases data access time Adds coding complexity to queries

ERD Notations (Recap)





ERD: Relationships



Mandatory vs Optional relationships

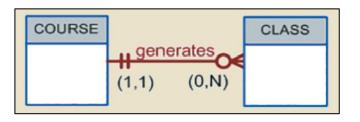
Textbook Table 4.3

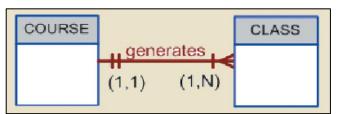
TABLE 4.3

CROW'S FOOT SYMBOLS

SYMBOL	CARDINALITY	COMMENT
O€	(0,N)	Zero or many; the "many" side is optional.
l€	(1,N)	One or many; the "many" side is mandatory.
II	(1,1)	One and only one; the "1" side is mandatory.
О	(0,1)	Zero or one; the "1" side is optional.

Textbook Figure 4.13





Textbook Figure 4.14

ERD: Relationship Strength



• Weak (non-identifying) relationship

- Primary key of the related entity does not contain a primary key component of the other entity
- The relationship is denoted using dashed line in the ERD

• Strong (identifying) relationships

- Primary key of the related entity contains a primary key component of the other entity
- A relationship that occurs when two entities are existence dependent
- The relationship is denoted using solid line in the ERD

Weak Relationship







Table name: COURSE

Database i	name:	Ch04_	TinyCollege
------------	-------	-------	-------------

CRS_CODE	DEPT_CODE	CRS_DESCRIPTION	CRS_CREDIT
ACCT-211	ACCT	Accounting I	3
ACCT-212	ACCT	Accounting II	3
CIS-220	CIS	Intro. to Microcomputing	3
CIS-420	CIS	Database Design and Implementation	4
MATH-243	MATH	Mathematics for Managers	3
QM-261	CIS	Intro. to Statistics	3
QM-362	CIS	Statistical Applications	4

Table name: CLASS

CLASS_CODE	CRS_CODE	CLASS_SECTION	CLASS_TIME	ROOM_CODE	PROF_NUM	
10012	ACCT-211	1	MVVF 8:00-8:50 a.m.	BUS311	105	
10013	ACCT-211	2	MVVF 9:00-9:50 a.m.	BUS200	105	
10014	ACCT-211	3	TTh 2:30-3:45 p.m.	BUS252	342	
10015	ACCT-212	1	MVVF 10:00-10:50 a.m.	BUS311	301	
10016	ACCT-212	2	Th 6:00-8:40 p.m.	BUS252	301	
10017	CIS-220	1	MVVF 9:00-9:50 a.m.	KLR209	228	
10018	CIS-220	2	MVVF 9:00-9:50 a.m.	KLR211	114	
10019	CIS-220	3	MVVF 10:00-10:50 a.m.	KLR209	228	
10020	CIS-420	1	vV 6:00-8:40 p.m.	KLR209	162	
10021	QM-261	1	MVVF 8:00-8:50 a.m.	KLR200	114	
10022	QM-261	2	TTh 1:00-2:15 p.m.	KLR200	114	
10023	QM-362	1	MVVF 11:00-11:50 a.m.	KLR200	162	
10024	QM-362	2	TTh 2:30-3:45 p.m.	KLR200	162	
10025	MATH-243	1	Th 6:00-8:40 p.m.	DRE155	325	

Textbook Figure 4.8

Strong Relationship



FIGURE 4.9 A STRONG (IDENTIFYING) RELATIONSHIP BETWEEN COURSE AND CLASS



Table name: COURSE

Database name: Ch04_TinyCollege_Alt

CRS_CODE	DEPT_CODE	CRS_DESCRIPTION	CRS_CREDIT
ACCT-211	ACCT	Accounting I	3
ACCT-212	ACCT	Accounting II	3
CIS-220	CIS	Intro. to Microcomputing	3
CIS-420	CIS	Database Design and Implementation	4
MATH-243	MATH	Mathematics for Managers	3
QM-261	CIS	Intro. to Statistics	3
QM-362	CIS	Statistical Applications	4

Table name: CLASS

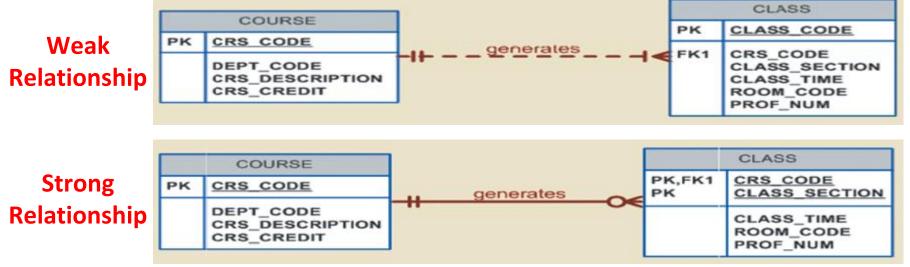
lable liame. CLASS						
CRS_CODE	CLASS_SECTION	CLASS_TIME	ROOM_CODE	PROF_NUM		
ACCT-211	1	MVVF 8:00-8:50 a.m.	BUS311	105		
ACCT-211	2	MVVF 9:00-9:50 a.m.	BUS200	105		
ACCT-211	3	TTh 2:30-3:45 p.m.	BUS252	342		
ACCT-212	1	MVVF 10:00-10:50 a.m.	BUS311	301		
ACCT-212	2	Th 6:00-8:40 p.m.	BUS252	301		
CIS-220	1	MVVF 9:00-9:50 a.m.	KLR209	228		
CIS-220	2	MVVF 9:00-9:50 a.m.	KLR211	114		
CIS-220	3	MVVF 10:00-10:50 a.m.	KLR209	228		
CIS-420	1	√V 6:00-8:40 p.m.	KLR209	162		
MATH-243	1	Th 6:00-8:40 p.m.	DRE155	325		
QM-261	1	MVVF 8:00-8:50 a.m.	KLR200	114		
QM-261	2	TTh 1:00-2:15 p.m.	KLR200	114		
QM-362	1	MVVF 11:00-11:50 a.m.	KLR200	162		
QM-362	2	TTh 2:30-3:45 p.m.	KLR200	162		

Textbook Figure 4.9

Strong vs Weak Relationships



Textbook Figure 4.8



• Entity and Referential Integrity

Textbook

- Primary key: NOT NULL

Figure 4.9

- Foreign key: Can be NULL

Relationship can be strong/weak depending on how primary keys are defined in entities

A class record can be created without course information – Weak relationship

A class record can't be created without course information – Strong relationship

Existence Dependence



• Existence dependence

- Entity exists in the database only when it is associated with another related entity occurrence

• Existence independence

- Entity exists apart from all of its related entities
- Referred to as a strong entity or regular entity

Weak Entities



Conditions of a weak entity

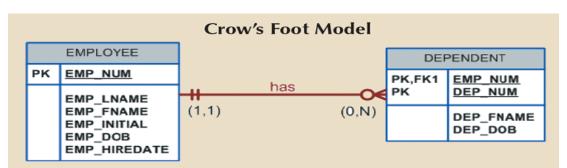
- Existence-dependent
- Has a primary key that is partially or totally derived from parent entity in the relationship
- Condition of strong relationships

Weak entity has a strong relationships with another entity

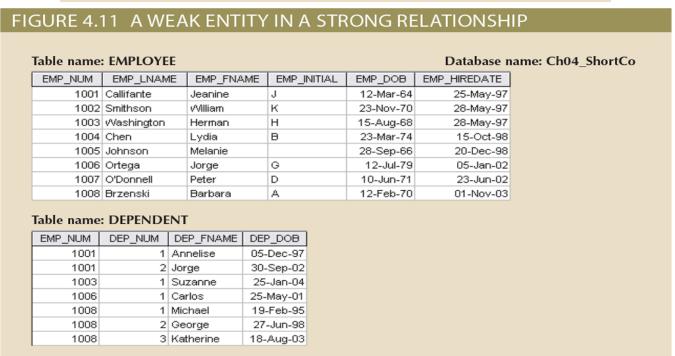
Weak Entities



A company insurance policy insures an employee and any dependents



Textbook Figure 4.10



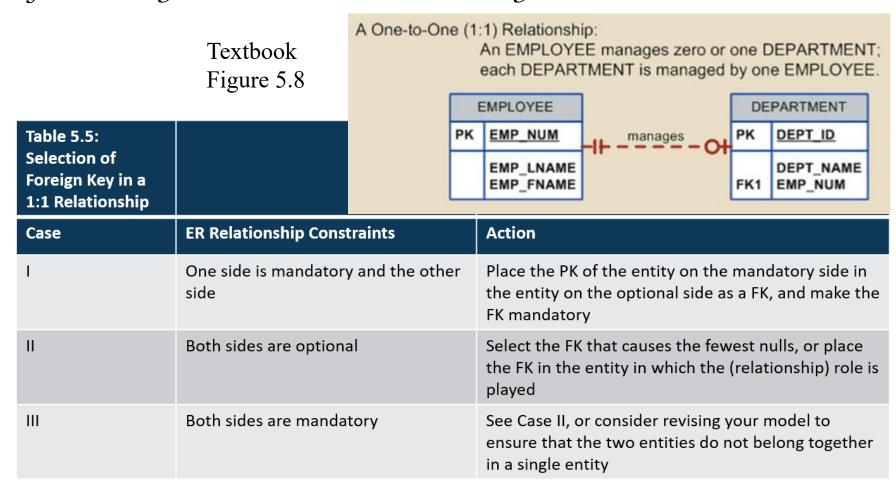
Textbook Figure 4.11

Implementing Relationships



1:1 Relationship:

PK of one entity as a FK in another entity



Implementing Relationships (2)



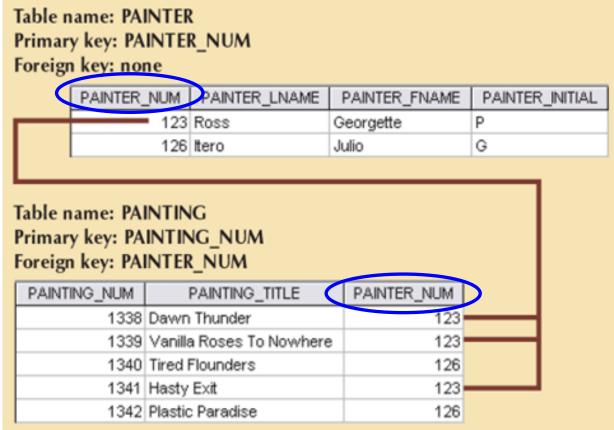
1:M Relationship:

PK of the "1" side in the table of the "M" side as a FK

Textbook Figure 3.17



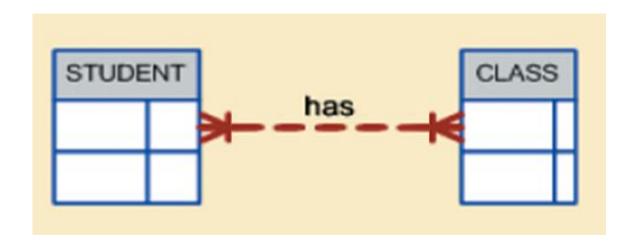
Textbook Figure 3.18



Implementing Relationships (3)



M:N Relationship:



Textbook Figure 3.23

Any idea how can we implement it?

Do you think it can be implemented just using PK/FK?

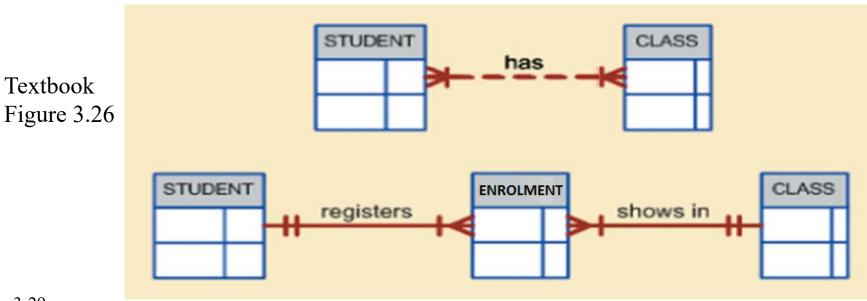
It is little bit tricky!

Implementing M:N Relationship



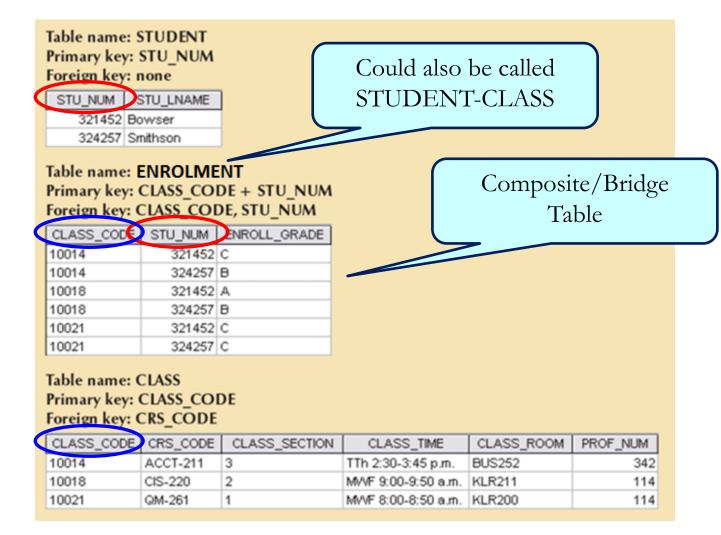
Replacing it with two 1:M relationships by creating a new composite table

- bridge table or associative entity
- Including foreign keys based on the primary keys of the 2 tables
- Assigning additional attributes as needed to the composite table



M:N Relationship Example



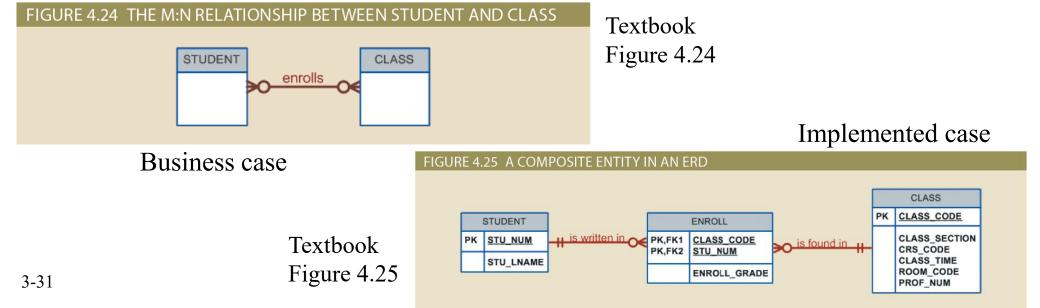


Textbook Figure 3.25

Associative Entity (Revisit)



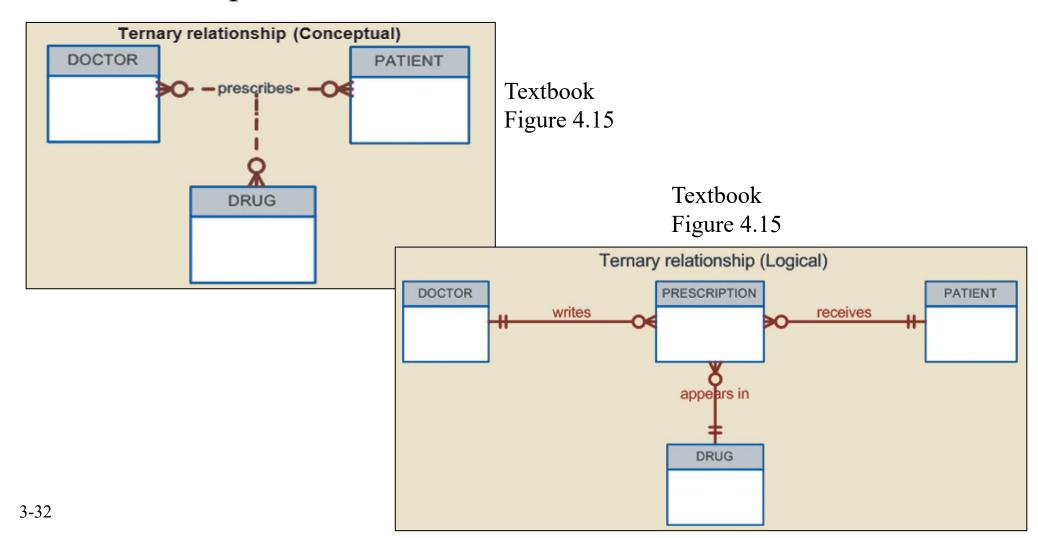
- Used to implement an M:N relationship between two or more entities
- Composed of the primary key attributes of each parent entity
- May also contain additional attributes that play no role in connective process



Associative Entity (Revisit) (2)



• DOCTORs prescribe DRUGs for PATIENTs.



Extended ER Model



- Advanced data modelling
- Result of adding more semantic constructs to ER model
 - modelling data requirements in complex real-world applications
 - Entity supertypes
 - Entity subtypes
 - Entity clustering

Extended ER Model (2)



- Consider a scenario, in which most employees possess a wide range of skills and special qualifications,
- Database designer must find a variety of ways to group employees based on their characteristics.
- For instance, a retail company could group employees as salaried and hourly,
- while a university could group employees as faculty, and admin staff
- The grouping of employees into various types provides two important benefits:
 - It avoids unnecessary NULLs in attributes when some employees have characteristics that are not shared by other employees.
- It enables a particular employee type to participate in relationships that are unique to that employee type.

Issues of NULLs



Textbook Figure 5.1

FIGURE 5.1 NULLS CREATED BY UNIQUE ATTRIBUTES

Database name: Ch05_AirCo

EMP_NUM	EMP_LNAME	EMP_FNAME	EMP_INITIAL	EMP_LICENSE	EMP_RATINGS	EMP_MED_TYPE	EMP_HIRE_DATE
100	Kolmycz	Xavier	T				15-Mar-88
101	Lewis	Marcos		ATP	SEL/MEL/Instr/CFII	1	25-Apr-89
102	Vandam	Jean					20-Dec-93
103	Jones	Victoria	R				28-Aug-03
104	Lange	Edith		ATP	SELMEL/Instr	1	20-Oct-97
105	v∕illiams	Gabriel	U	COM	SEL/MEL/Instr/CFI	2	08-Nov-97
106	Duzak	Mario		COM	SELMEL/Instr	2	05-Jan-04
107	Diante	Venite	L				02-Jul-97
108	√lesenbach	Joni					18-Nov-95
109	Travis	Brett	T	COM	SEL/MEL/SES/Instr/CFII	1	14-Apr-01
110	Genkazi	Stan			A		01-Dec-03

These attributes are applicable to certain types of Employees only

Entity Supertypes and Subtypes



Entity supertype

- Generic entity type related to one or more entity subtypes
- Contains common characteristics

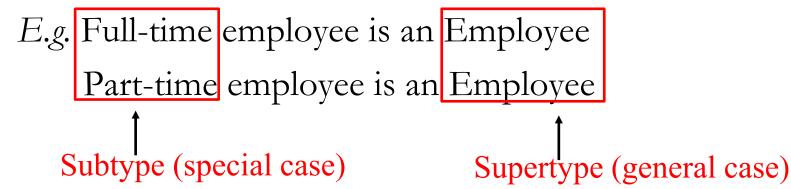
Entity subtype

- Contains unique characteristics of each entity subtype
- Criteria to determine usage
 - The different kinds of instances should each have one or more attributes that are unique to that kind of instance
 - Define a special Supertype attribute known as the Subtype discriminator.
 - Define disjoint or overlapping constraints and complete or partial constraints.

Specialization Hierarchy



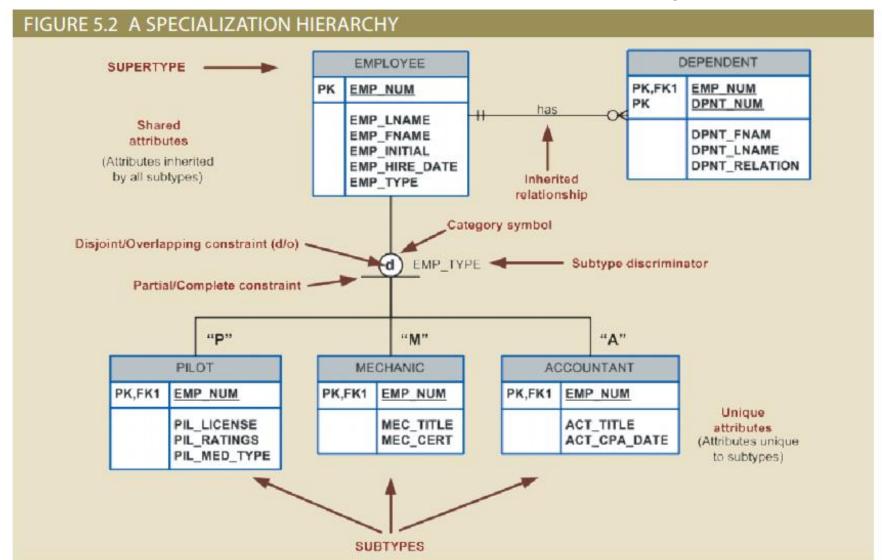
- Entity supertypes and subtypes are organized in a specialization hierarchy
 - Depicts arrangement of higher-level entity supertypes and lower-level entity subtypes
 - Relationships are described in terms of "is-a" relationships
 - Every subtype has one supertype to which it is directly related
 - Supertype can have many subtypes



Specialization Hierarchy (2)



Textbook Figure 5.2



Disjoint and Overlapping Constraints



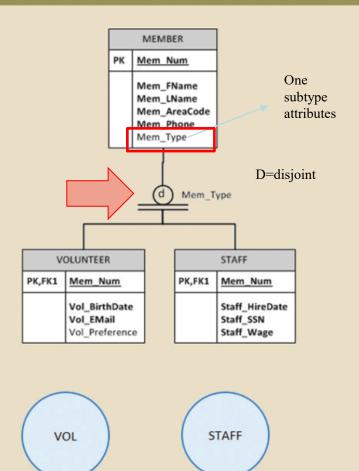
- **Disjoint subtypes**: contain a unique subset of the supertype entity set
 - Known as nonoverlapping subtypes
 - Implementation is based on the value of the subtype discriminator attribute in the supertype (in the parent entity)
- Overlapping subtypes: contain nonunique subsets of the supertype entity set
 - Implementation requires the use of one discriminator attribute for each subtype (in the parent entity)

Disjoint and Overlapping Constraints (2)

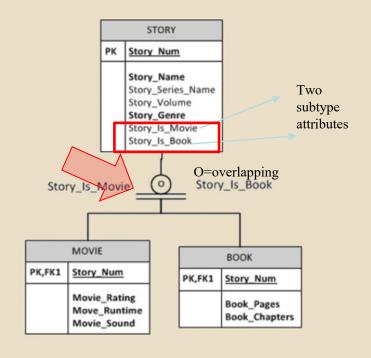
Textbook Figure 5.5

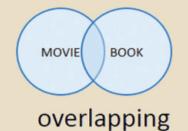


FIGURE 5.5 DISJOINT AND OVERLAPPING SUBTYPES









A STORY can be in MOVIE, BOOK, or both

Completeness constraints

Subtype sets are unique.



- Specifies whether each supertype occurrence must also be a member of at least one subtype
 - Partial completeness: not every supertype occurrence is a member of a subtype
 - Total completeness: every supertype occurrence must be a member of at least one subtypes

 Textbook

TABLE 5.2 Figure 5.2

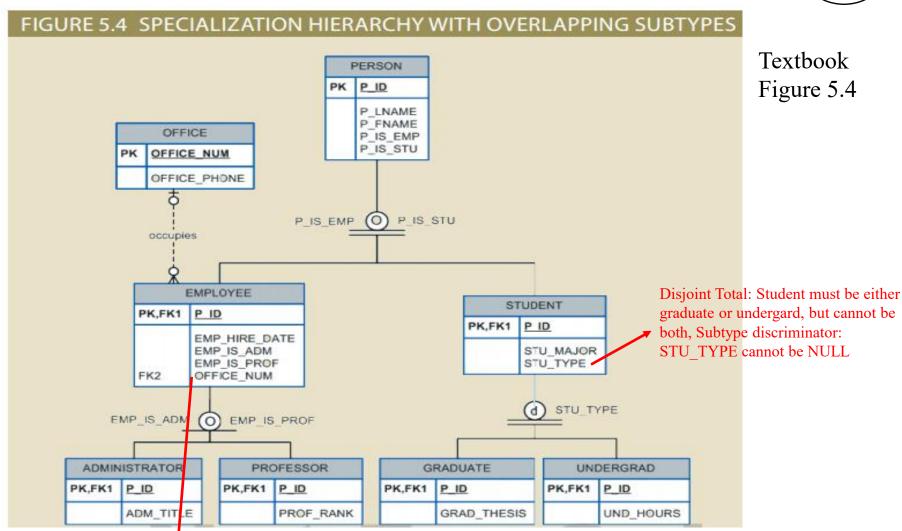
SPECIALIZATION HIERARCHY CONSTRAINT SCENARIOS

TYPE DISJOINT CONSTRAINT OVERLAPPING CONSTRAINT **Partial** Supertype has optional subtypes. Supertype has optional subtypes. Subtype discriminator can be null. Subtype discriminators can be null. Subtype sets are unique. Subtype sets are not unique. Total Every supertype occurrence is a member of only Every supertype occurrence is a member of at one subtype. least one subtype. Subtype discriminator cannot be null. Subtype discriminators cannot be null.

Subtype sets are not unique.

Levels of hierarchy





Overlapping partial: All employee may not be (either administrator or a professor) Subtype discriminator: EMP IS ADM,EMP IS PROFC for this can be NULL

Specialization and Generalization,



Specialization

- Top-down process
- Identifies lower-level, more specific entity subtypes from a higher-level entity supertype
- Based on grouping unique characteristics and relationships of the subtypes

Generalization

- Bottom-up process
- Identifies a higher-level, more generic entity supertype from lower-level entity subtypes
- Based on grouping common characteristics and relationships of the subtypes

Entity Clustering



"Virtual/Abstract" entity type used to represent multiple entities and relationships in ERD

- To simply complex ERD
 - in some problems you may have hundreds of entities
- Formed by combining multiple interrelated entities into a single, abstract entity object
- **General rule**: avoid the display of attributes to eliminate complications that result when the inheritance rules change
- Not an actual entity from the business rule
- Not implemented

Entity Clustering (2)

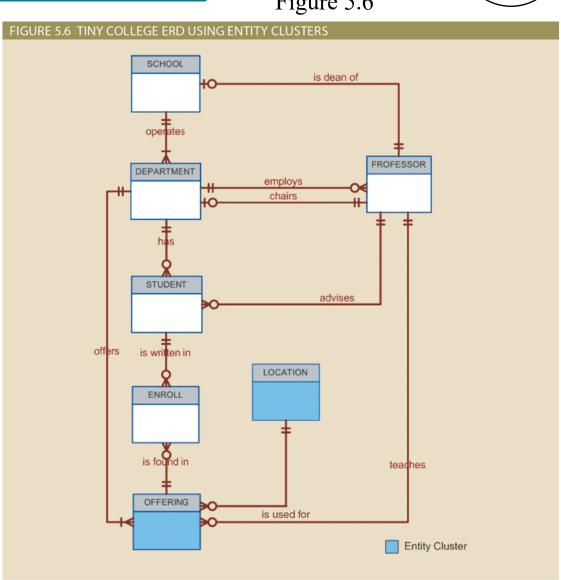
DEAKIN UNIVERSITY

Textbook Figure 5.6

OFFERING groups SEMESTER, COURSE, and CLASS

LOCATION groups
BUILDING and ROOM

Relationships of Entities within the Entity cluster is represented in a separate ERD



History of Time-variant data

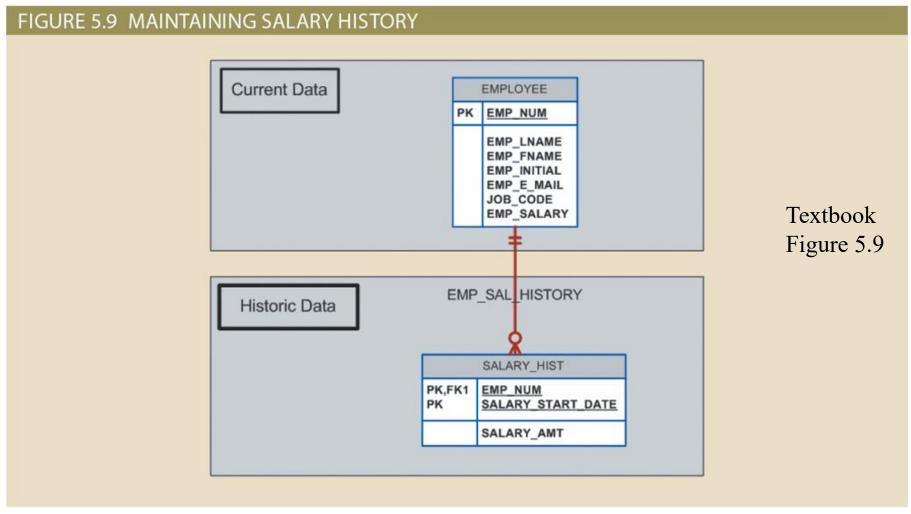


- Time-variant data: data whose values change over time and for which a <u>history of the data changes must be</u> retained
 - Requires creating a new entity in a 1:M relationship with the original entity
 - New entity contains the new value, date of the change, and any other pertinent attribute

E.g., tracking salary histories of employees

Maintaining data history

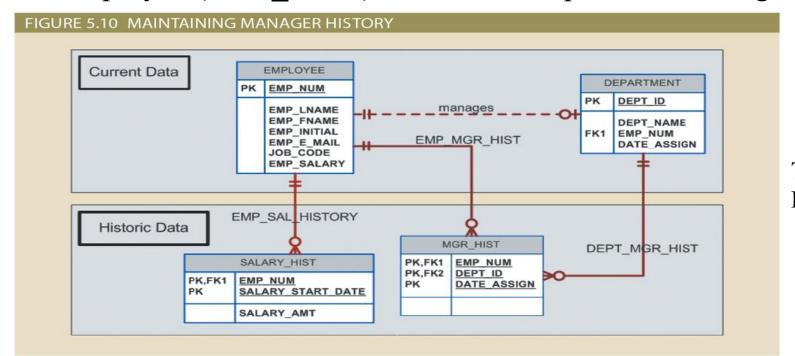




Maintaining data history (2)



- An employee could be the manager of many different departments over time,
- A department could have different managers over time.
- Because you are recording time-variant data, you must store the DATE_ASSIGN attribute in the MGR_HIST entity to provide the date that the employee (EMP_NUM) became the department manager.



Textbook Figure 5.10

ERD Exercise: Case 1



Draw an ER diagram for the following description

The HEG has twelve instructors and can handle up to thirty trainees per class. HEG offers five "advanced technology" courses, each of which may generate several classes. If a class has fewer than ten trainees in it, it will be cancelled. It is, therefore, possible for a course not to generate any classes during a session. Each class is taught by one instructor. Each instructor may teach up to two classes or may be assigned to do research only. Each trainee may take up to two classes per session.

ERD Exercise: Case 1 (2)



Determine Entities

The HEG has twelve instructors and can handle up to thirty trainees per class. HEG offers five "advanced technology" courses, each of which may generate several classes. If a class has fewer than ten trainees in it, it will be cancelled. It is, therefore, possible for a course not to generate any classes during a session. Each class is taught by one instructor. Each instructor may teach up to two classes or may be assigned to do research only. Each trainee may take up to two classes per session.

ERD Exercise: Case 1 (3)



Determine Relationships

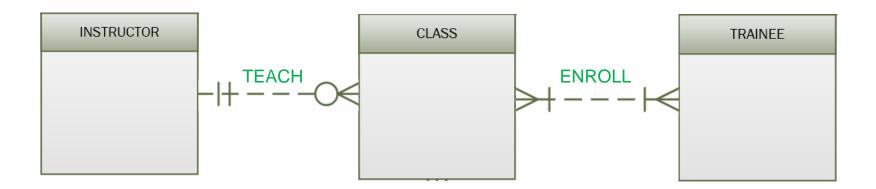
- One <u>instructor</u> → Many <u>class</u>es (teach)
- One <u>class</u>
 One <u>instructor</u> (taught by)
- Some instructors MAY NOT teach any classes
- One <u>course</u> → Many <u>class</u>es (*generate*)
- One <u>class</u> → One <u>course</u> (*generated by*)
- Some courses MAY NOT generate any classes
- One <u>trainee</u>
 Many <u>classes</u> (enroll)
- One <u>class</u>
 Many <u>trainees</u>
 (enrolled by)

ERD Exercise: Case 1 (4)



HEG has twelve <u>instructors</u> and can handle up to thirty <u>trainees</u> per <u>class</u>. ...

Also consider the business rules about: Instructor, Class, Trainee

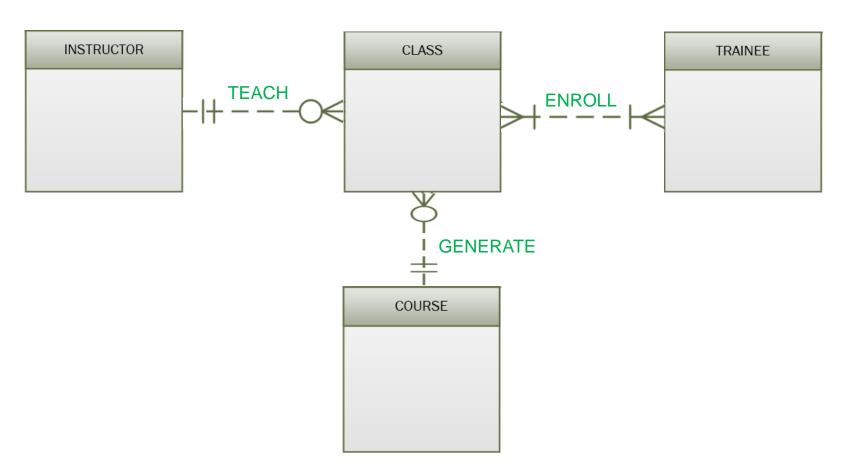


ERD Exercise: Case 1 (4)



... HEG offers five "advanced technology" <u>courses</u>, each of which may generate several <u>classes</u>. ...

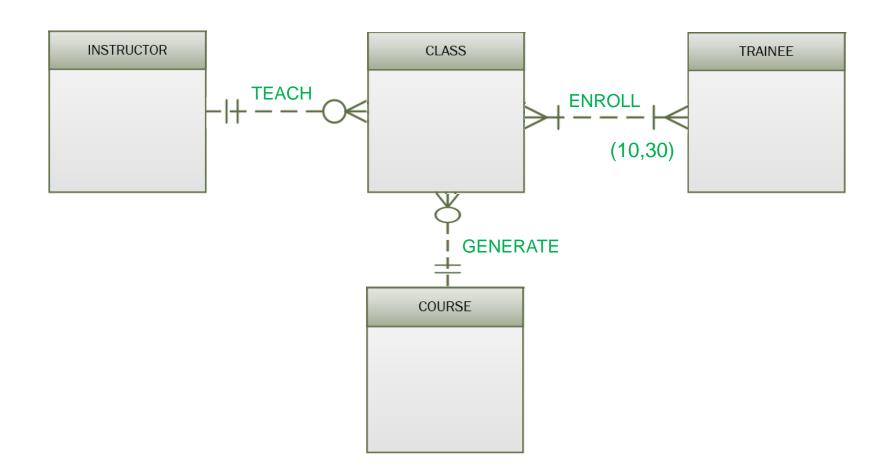
Also consider the business rules about: Class, Course



ERD Exercise: Case 1 (5)



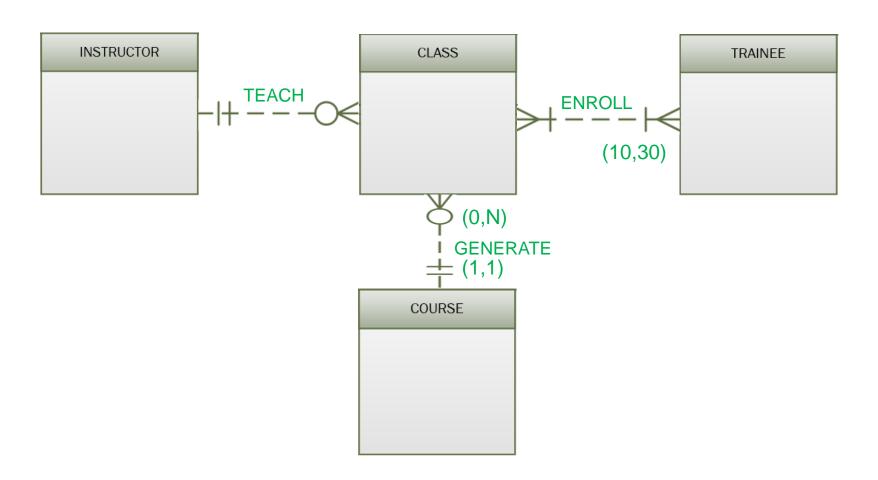
... twelve instructors can handle up to thirty trainees per class. ... If a class has fewer than ten trainees in it, it will be cancelled.



ERD Exercise: Case 1 (6)



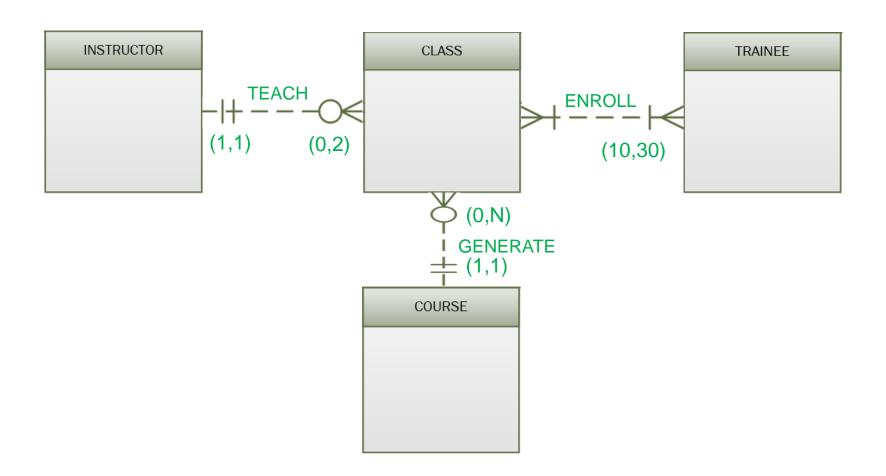
... each course may generate <u>several</u> classes. ... It is, therefore, possible for a course <u>not to generate any classes</u> during a session.



ERD Exercise: Case 1 (7)



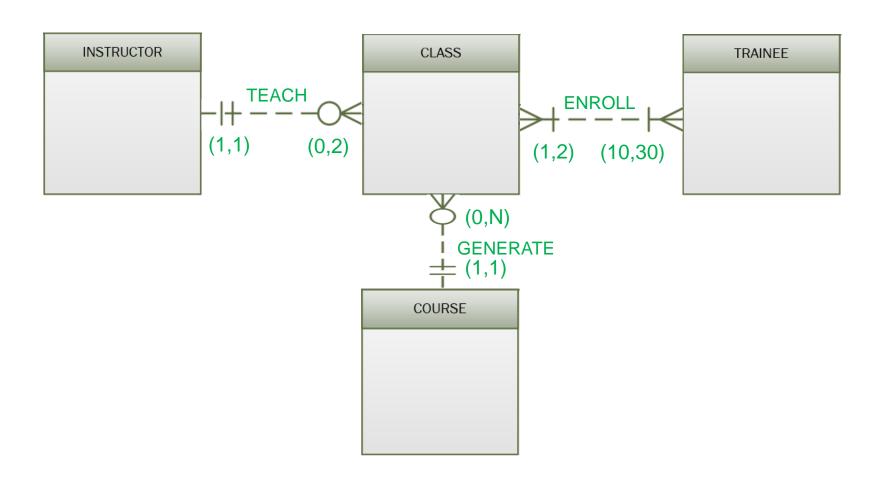
.... Each class is taught by **one instructor**. Each instructor may teach **up to two classes** or may be assigned to **do research only**.



ERD Exercise: Case 1 (8)



... Each trainee may take up to two classes per session.



ERD Exercise: Case 1 (3)



Note that we did not consider Attributes here

What next:

Attributes

This is a homework for you!

- Primary keys
- Maintaining relationships
 - using FK
 - using bridge/associative entity (M:N relationships)

ERD Exercise: Case 2



Draw an E-R diagram for the following description

The Jonesburgh County Basketball Conference (JCBC) is an amateur basketball association. Each city in the county has one team that represents it. Each team has a maximum of twelve players and a minimum of nine players. Each team also has up to three coaches (offensive, defensive and PT coaches). Each team plays two games (home and visitor) against each of the other teams during the season.

ERD Exercise: Case 2 (2)



Determine Entities

The Jonesburgh County Basketball Conference (JCBC) is an amateur basketball association. Each city in the county has one team that represents it. Each team has a maximum of twelve players and a minimum of nine players. Each team also has up to three coaches (offensive, defensive and PT coaches). Each team plays two games (home and visitor) against each of the other teams during the season.

ERD Exercise: Case 2 (3)



Determine Relationships

- One <u>city</u> →
- One team
- One team
- One player
- One team
- One coach
- One team
- One team

One team (posses)

One city (represent)

Many players (has)

One team (belong to)

Many coaches (employ)

One team (employed by)

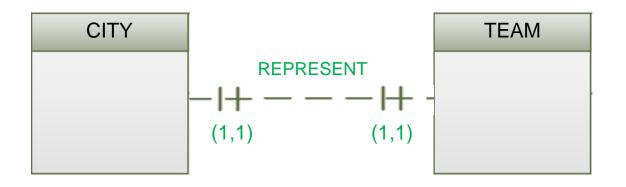
Many teams (play against)

Many teams (play against)

ERD Exercise: Case 2 (4)



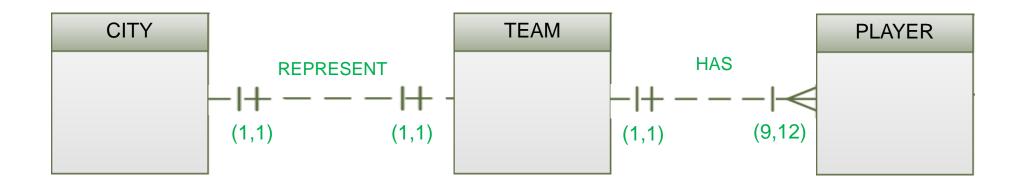
... Each <u>city</u> in the county has one <u>team</u> that represents it. ...



ERD Exercise: Case 2 (5)



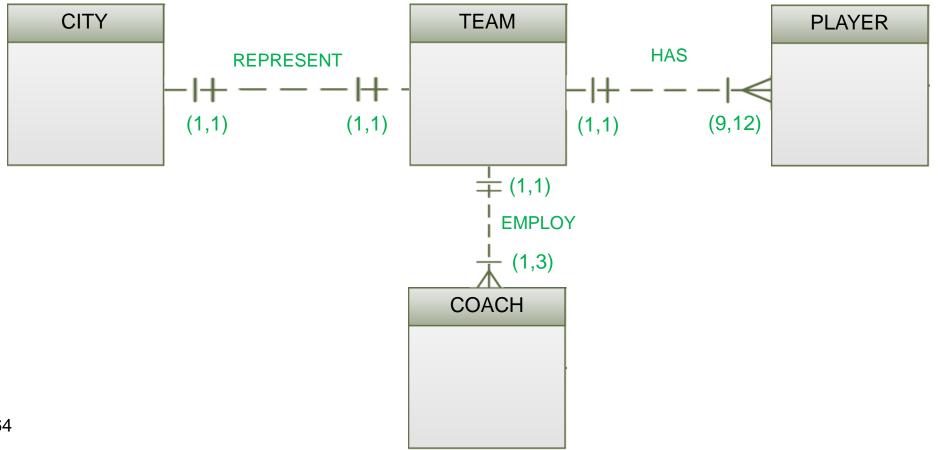
.... Each <u>team</u> has a minimum of nine <u>players</u> and a maximum of twelve players.....



ERD Exercise: Case 2 (6)



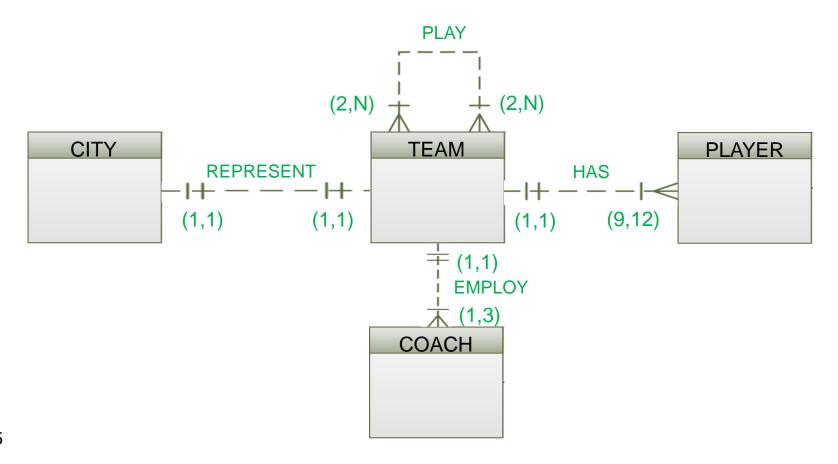
... Each **team** also has up to three **coaches** ...



ERD Exercise: Case 2 (7)



.... Each <u>team</u> plays two games (home and visitor) against each of the other <u>teams</u> during the season.

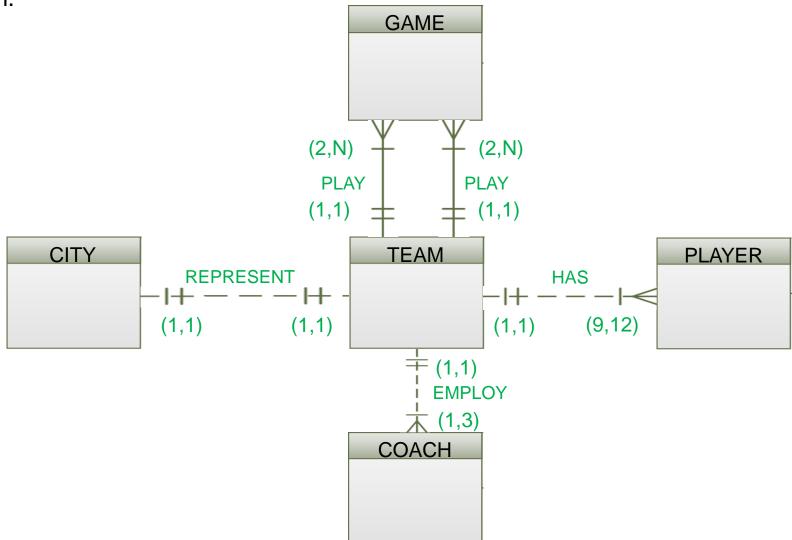


ERD Exercise: Case 2 (8)



... Each **team** plays two games (home and visitor) against each of the other **teams** during the

season.



Summary



- Types of Attributes
- Strengths of Relationships
- Types of Entities
- Implementing Relationships
- Extended/Advanced ERD concepts
- Modelling historical time-variant data
- Case studies

This Week's OnTrack Tasks



- 3.1P Modelling database for a given business scenario in terms of ERD
 - Using Lucid Chart or MS Visio
- 3.2HD Research Report and Presentation
 - On a topic of your interest related to database and/or data management
 - Select one from the given list or propose your own (discuss with us)
 - Due on Friday of Week 9 (16 Sept 2022)

Please check the task sheets and start working on them

Next Week



Normalisation

Thank you

See you next week

Any questions/comments?

Readings and References:



• Chapters 2-5 and 9

Database Systems: Design, Implementation, & Management 13TH EDITION, by Carlos Coronel, Steven Morris