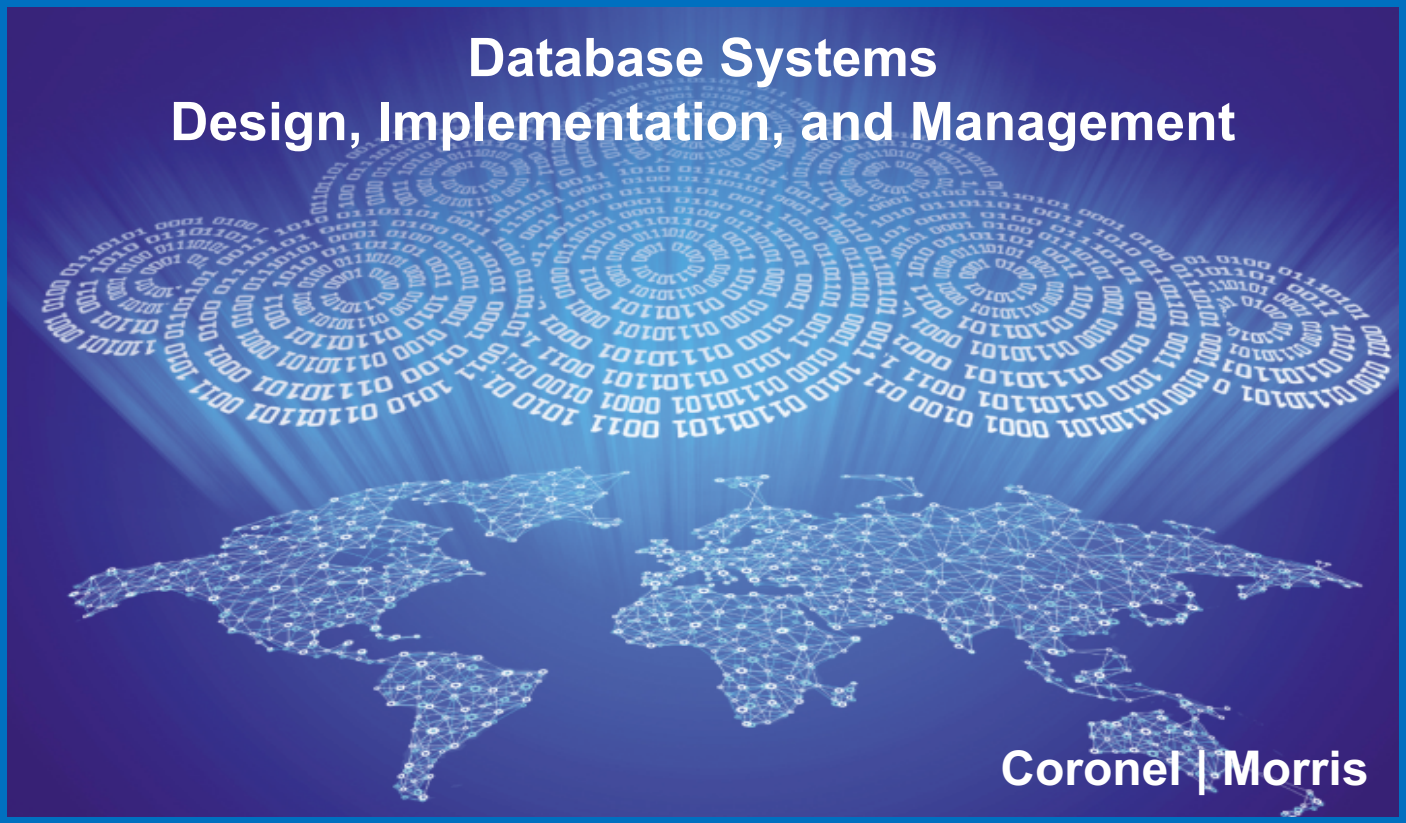
11e



Chapter 5
Advanced Data Modeling

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Learning Objectives

- In this chapter, you will learn:
  - About the extended entity relationship (EER) model
  - How entity clusters are used to represent multiple entities and relationships
  - The characteristics of good primary keys and how to select them
  - How to use flexible solutions for special data-modeling cases

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Extended Entity Relationship Model (EERM)

- Result of adding more semantic constructs to the original entity relationship (ER) model
- **EER diagram (EERD)**: Uses the EER model

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Entity Supertypes and Subtypes

- Most employees possess a wide rage of skills and special qualifications, data modeler must find a variety of group employees based on their characteristics.
  - A university could group employees as faculty, staff, and administrators.
  - A retail company could group employees as salaried and hourly.

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Entity Supertypes and Subtypes

- 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.

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Entity Supertypes and Subtypes

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	SEL/MEL/Instr	1	20-Oct-97
105	Williams	Gabriel	U	COM	SEL/MEL/Instr/CFI	2	08-Nov-97
106	Duzak	Mario		COM	SEL/MEL/Instr	2	05-Jan-04
107	Diante	Venite	L				02-Jul-97
108	Wiesenbach	Joni					18-Nov-95
109	Travis	Brett	T	COM	SEL/MEL/SES/Instr/CFII	1	14-Apr-01
110	Genkazi	Stan					01-Dec-03

Cengage Learning © 2015

- Many pilot characteristics are not shared by other employees.
- Pilots participated in some relationships that are unique to their qualifications. (employee flies airplane)

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

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 the usage
  - There must be different, identifiable kinds of the entity in the user's environment
  - The different kinds of instances should each have one or more attributes that are unique to that kind of instance

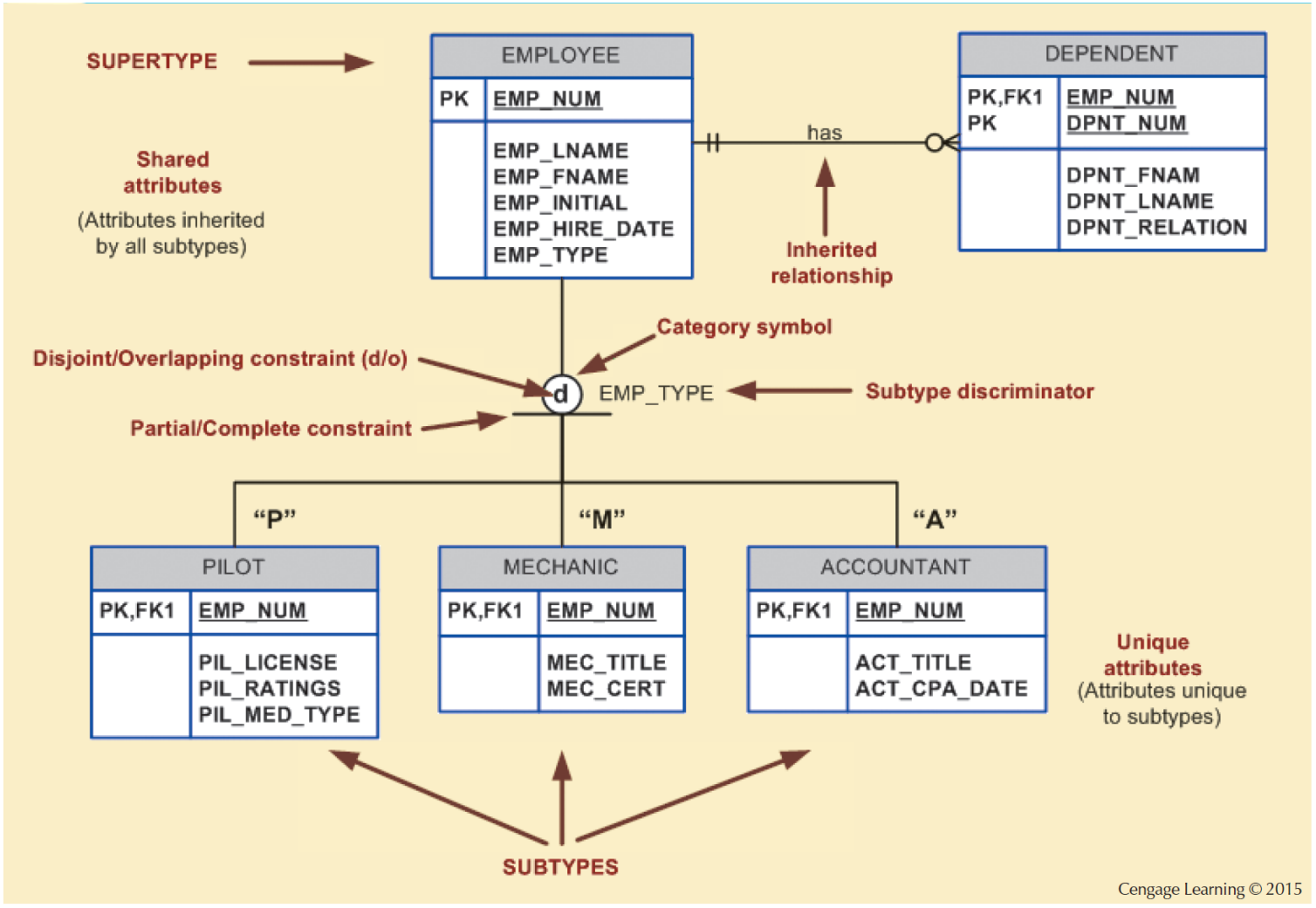
©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Specialization Hierarchy

- Entity supertypes and subtypes are organized in a specialization hierarchy, which depicts arrangement of higher-level entity supertypes and lower-level entity subtypes
- Relationships are described in terms of "is-a" relationships (a pilot is an employee)
- Subtype exists within the context of a supertype
- Every subtype has one supertype to which it is directly related
- Supertype can have many subtypes

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Figure 5.2 - Specialization Hierarchy



©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Specialization Hierarchy

- Provides the means to:
  - Support attribute inheritance
  - Define a special supertype attribute known as the subtype discriminator
  - Define disjoint/overlapping constraints and complete/partial constraints

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Inheritance

- Enables an entity subtype to inherit attributes and relationships of the supertype
- All entity subtypes inherit their primary key attribute from their supertype
- At the implementation level, supertype and its subtype(s) maintain a 1:1 relationship
- Entity subtypes inherit all relationships in which supertype entity participates
- Lower-level subtypes inherit all attributes and relationships from its upper-level supertypes

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Inheritance

Table name: EMPLOYEE

EMP_NUM	EMP_LNAME	EMP_FNAME	EMP_INITIAL	EMP_HIRE_DATE	EMP_TYPE
100	Kolmycz	Xavier	T	15-Mar-88	
101	Lewis	Marcos		25-Apr-89	P
102	Vandam	Jean		20-Dec-93	A
103	Jones	Victoria	R	28-Aug-03	
104	Lange	Edith		20-Oct-97	P
105	Williams	Gabriel	U	08-Nov-97	P
106	Duzak	Mario		05-Jan-04	P
107	Diante	Venite	L	02-Jul-97	M
108	Wiesenbach	Joni		18-Nov-95	M
109	Travis	Brett	T	14-Apr-01	P
110	Genkazi	Stan		01-Dec-03	A

Table name: PILOT

EMP_NUM	PIL_LICENSE	PIL_RATINGS	PIL_MED_TYPE
101	ATP	SEL/MEL/Instr/CFII	1
104	ATP	SEL/MEL/Instr	1
105	COM	SEL/MEL/Instr/CFI	2
106	COM	SEL/MEL/Instr	2
109	COM	SEL/MEL/SES/Instr/CFII	1

Cengage Learning © 2015

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Subtype Discriminator

- Attribute in the supertype entity that determines to which entity subtype the supertype occurrence is related (EMP_TYPE)
  - If the EMP_TYPE values is “A”, the supertype is related to the ACCOUNTANT subtype.
- Default comparison condition is the equality comparison

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

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
  - An employee(supertype) who is a pilot (subtype) can appear only in the PILOT subtype, not in any of the other subtypes.

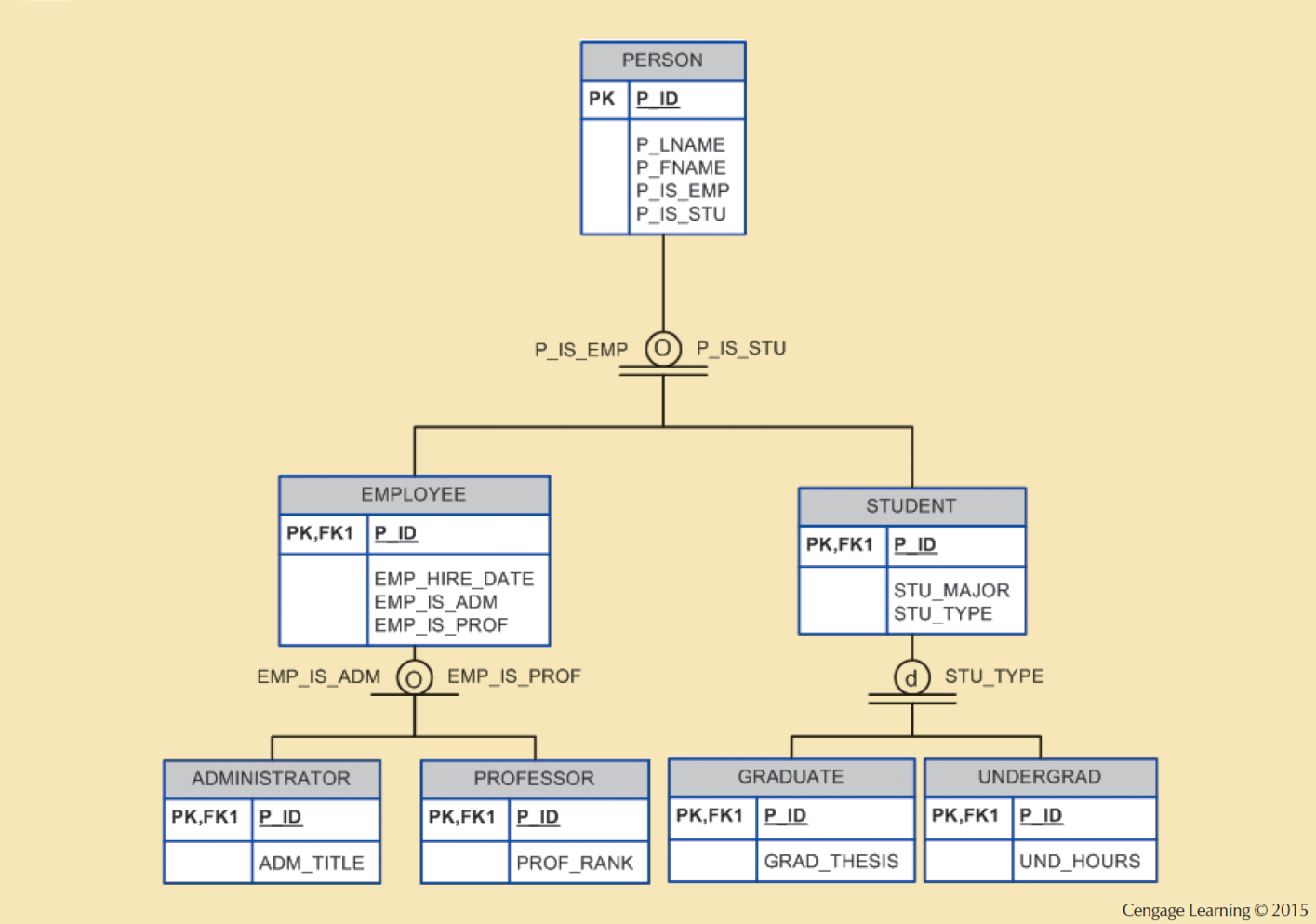
©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Disjoint and Overlapping Constraints

- **Overlapping subtypes**: Contain nonunique subsets of the supertype entity set
  - Implementation requires the use of one discriminator attribute for each subtype
    - In a university environment, a person may be an employee, a student, or both.
    - An employee may be a professor as well as an administrator.
    - An employee may also be a student, STUDENT and EMPLOYEE are overlapping subtypes.
    - PROFESSOR and ADMINISTRATOR are overlapping subtypes of the supertype EMPLOYEE.

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Figure 5.4 - Specialization Hierarchy with Overlapping Subtypes



©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Table 5.1 - Discriminator Attributes with Overlapping Subtypes

DISCRIMINATOR ATTRIBUTES		COMMENT
Professor	Administrator	
Y	N	The Employee is a member of the Professor subtype.
N	Y	The Employee is a member of the Administrator subtype.
Y	Y	The Employee is both a Professor and an Administrator.

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Completeness Constraint

- Specifies whether each supertype occurrence must also be a member of at least one subtype
- Types
  - **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 subtype.

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Table 5.2 - Specialization Hierarchy Constraint Scenarios

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

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

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

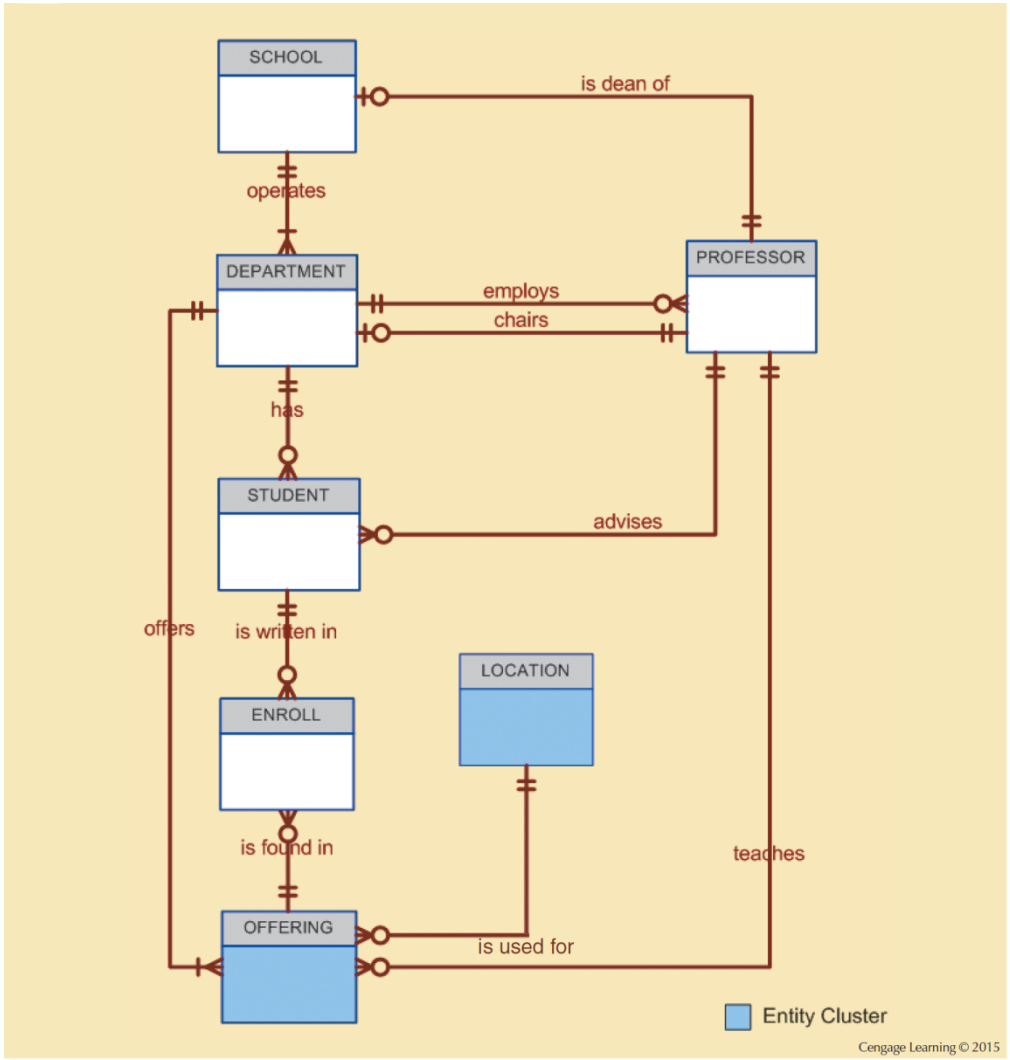
©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Entity Cluster

- Virtual entity type used to represent multiple entities and relationships in ERD
- Avoid the display of attributes (no key attributes) to eliminate complications that result when the inheritance rules change

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Figure 5.5 - Tiny College ERD Using Entity Clusters



- OFFERING, groups the COURSE and CLASS entities and relationships.
- Location, groups the ROOM and BUILDING entities and relationships.

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Primary Keys

- Single attribute or a combination of attributes, which uniquely identifies each entity instance
  - Guarantees entity integrity
  - Works with foreign keys to implement relationships

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Natural Keys or Natural Identifier

- Real-world identifier used to uniquely identify real-world objects
  - Familiar to end users and forms part of their day-to-day business vocabulary
  - Also known as natural identifier
  - Used as the primary key of the entity being modeled

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Desirable Primary Key Characteristics

- Non intelligent
- No change over time
- Preferably single-attribute
- Preferably numeric
- Security-compliant

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Desirable Primary Key Characteristics

PK CHARACTERISTIC	RATIONALE
Unique values	The PK must uniquely identify each entity instance. A primary key must be able to guarantee unique values. It cannot contain nulls.
Nonintelligent	The PK should not have embedded semantic meaning other than to uniquely identify each entity instance. An attribute with embedded semantic meaning is probably better used as a descriptive characteristic of the entity than as an identifier. For example, a student ID of 650973 would be preferred over *Smith, Martha L.* as a primary key identifier.
No change over time	If an attribute has semantic meaning, it might be subject to updates, which is why names do not make good primary keys. If *Vickie Smith* is the primary key, what happens if she changes her name when she gets married? If a primary key is subject to change, the foreign key values must be updated, thus adding to the database work load. Furthermore, changing a primary key value means that you are basically changing the identity of an entity. In short, the PK should be permanent and unchangeable.
Preferably single-attribute	A primary key should have the minimum number of attributes possible (irreducible). Single-attribute primary keys are desirable but not required. Single-attribute primary keys simplify the implementation of foreign keys. Having multiple-attribute primary keys can cause primary keys of related entities to grow through the possible addition of many attributes, thus adding to the database workload and making (application) coding more cumbersome.
Preferably numeric	Unique values can be better managed when they are numeric, because the database can use internal routines to implement a counter-style attribute that automatically increments values with the addition of each new row. In fact, most database systems include the ability to use special constructs, such as Autonumber in Microsoft Access, sequence in Oracle, or uniqueidentifier in MS SQL Server to support self-incrementing primary key attributes.
Security-compliant	The selected primary key must not be composed of any attribute(s) that might be considered a security risk or violation. For example, using a Social Security number as a PK in an EMPLOYEE table is not a good idea.

Cengage Learning © 2015
©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Use of Composite Primary Keys

- Identifiers of composite entities
  - Each primary key combination is allowed once in M:N relationship
  - STUDENT, ENROLL, CLASS.
- Identifiers of weak entities
  - Weak entity has a strong identifying relationship with the parent entity

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Use of Composite Primary Keys

- When used as identifiers of weak entities, represent a real-world object that is:
  - Existence-dependent on another real-world object
    - EMPLOYEE, DEPENDENT
  - Represented in the data model as two separate entities in a strong identifying relationship
    - The real-world invoice object is represented by two entities in a data model: INVOICE and LINE. LINE entity does not exist in the real world as an independent object, but as part of an INVOICE.

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Figure 5.6 - The M:N Relationship between STUDENT and CLASS

Database name: Ch05_Tinycollege

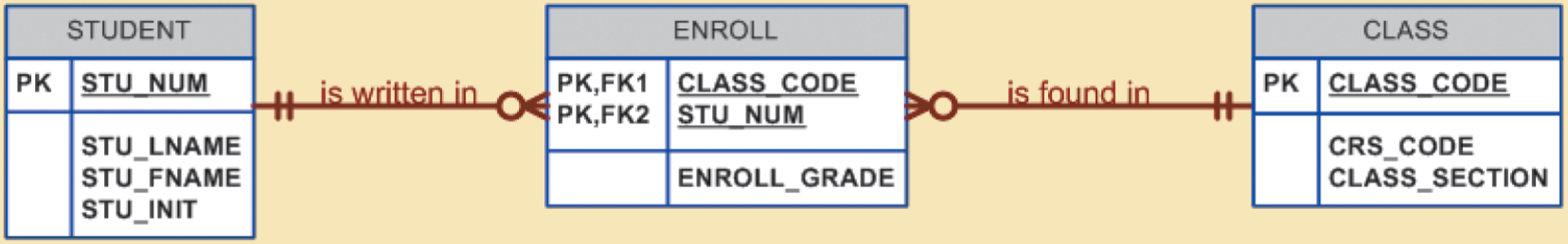


Table name: STUDENT
(first four fields)

STU_NUM	STU_LNAME	STU_FNAME	STU_INIT
321452	Bowser	William	C
324257	Smithson	Anne	K
324258	Brewer	Juliette	
324269	Oblonski	Walter	H
324273	Smith	John	D
324274	Katinga	Raphael	P
324291	Robertson	Gerald	T
324299	Smith	John	B

Table name: ENROLL

CLASS_CODE	STU_NUM	ENROLL_GRADE
10014	321452	C
10014	324257	B
10018	321452	A
10018	324257	B
10021	321452	C
10021	324257	C

Table name: CLASS
(first three fields)

CLASS_CODE	CRS_CODE	CLASS_SECTION
10012	ACCT-211	1
10013	ACCT-211	2
10014	ACCT-211	3
10015	ACCT-212	1
10016	ACCT-212	2
10017	CIS-220	1
10018	CIS-220	2
10019	CIS-220	3
10020	CIS-420	1
10021	QM-261	1
10022	QM-261	2
10023	QM-362	1
10024	QM-362	2
10025	MATH-243	1

Cengage Learning © 2015

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Surrogate Keys

- Primary key used to simplify the identification of entity instances are useful when:
  - There is no natural key
  - Selected candidate key has embedded semantic contents or is too long
- Require ensuring that the candidate key of entity in question performs properly
  - Use unique index and not null constraints

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Table 5.4 - Data Used to Keep Track of Events

DATE	TIME_START	TIME_END	ROOM	EVENT_NAME	PARTY_OF
6/17/2014	11:00AM	2:00PM	Allure	Burton Wedding	60
6/17/2014	11:00AM	2:00PM	Bonanza	Adams Office	12
6/17/2014	3:00PM	5:30PM	Allure	Smith Family	15
6/17/2014	3:30PM	5:30PM	Bonanza	Adams Office	12
6/18/2014	1:00PM	3:00PM	Bonanza	Boy Scouts	33
6/18/2014	11:00AM	2:00PM	Allure	March of Dimes	25
6/18/2014	11:00AM	12:30PM	Bonanza	Smith Family	12

Cengage Learning © 2015

- Primary key: (DATE,TIME_START,ROOM) or (DATE, ETIME_END, ROOM)?
- RESOURCE(**RSC_ID**, RSC_DESCRIPTION, RSC_TYPE, RSC_QTY, RSC_PRICE)
- COMPOSITE entity: EVNTRSC (DATE, TIME_START, ROOM, RSC_ID, QTY_USED)

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

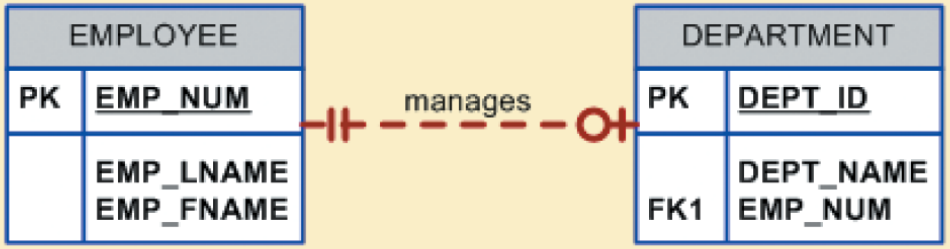
Design Case 1: Implementing 1:1 Relationships

- Foreign keys work with primary keys to properly implement relationships in relational model
- Rule
  - Put primary key of the parent entity on the dependent entity as foreign key
- Options for selecting and placing the foreign key:
  - Place a foreign key in both entities
  - Place a foreign key in one of the entities

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Figure 5.7 - The 1:1 Relationship between Department and Employee

A One-to-One (1:1) Relationship:
An EMPLOYEE manages zero or one DEPARTMENT;
each DEPARTMENT is managed by one EMPLOYEE.



Cengage Learning © 2015

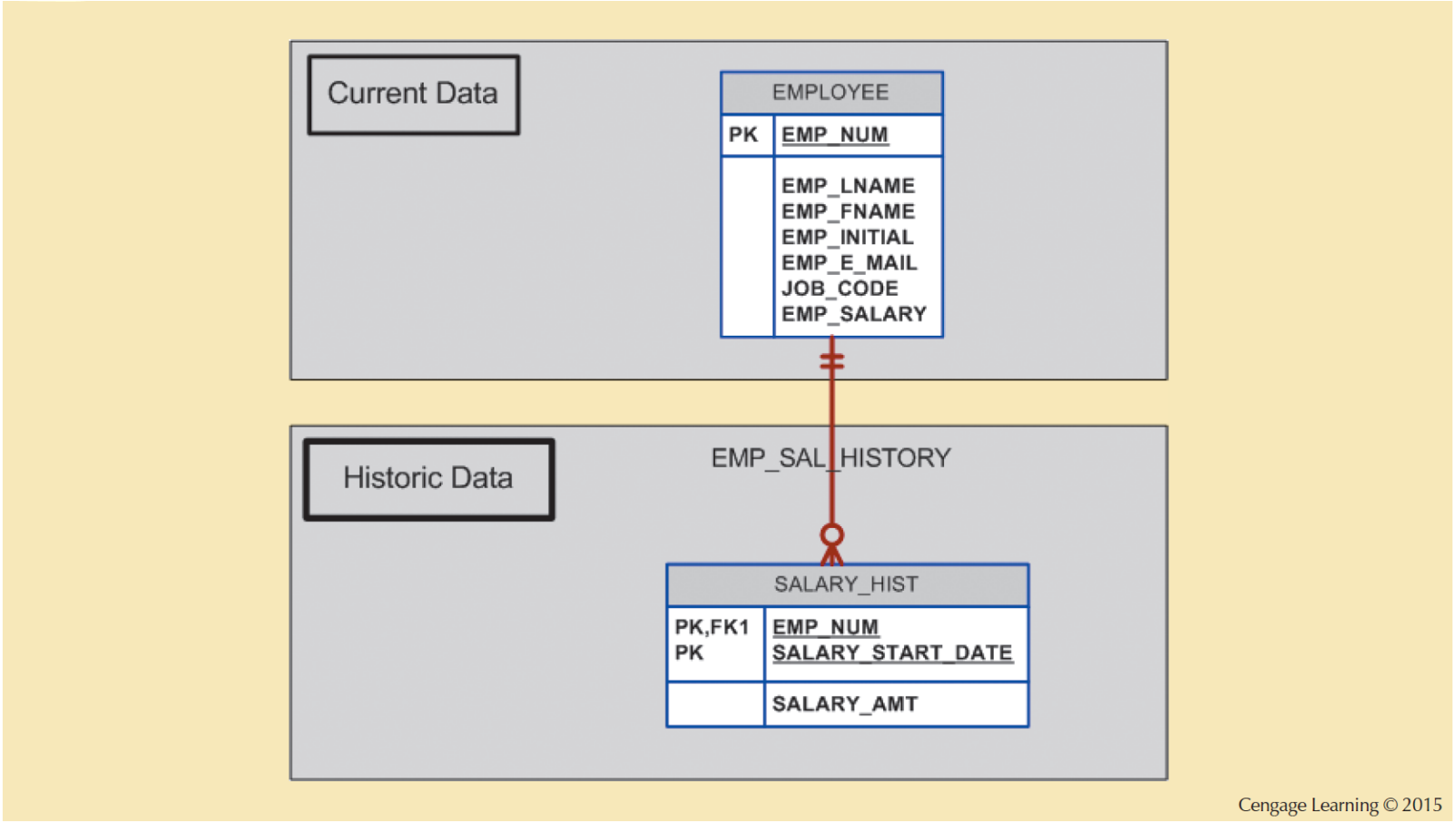
©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Design Case 2: Maintaining History of Time-Variant Data

- **Time-variant data**: Data whose values change over time and for which a history of the data changes must be 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 other pertinent attribute

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Figure 5.8 - Maintaining Salary History



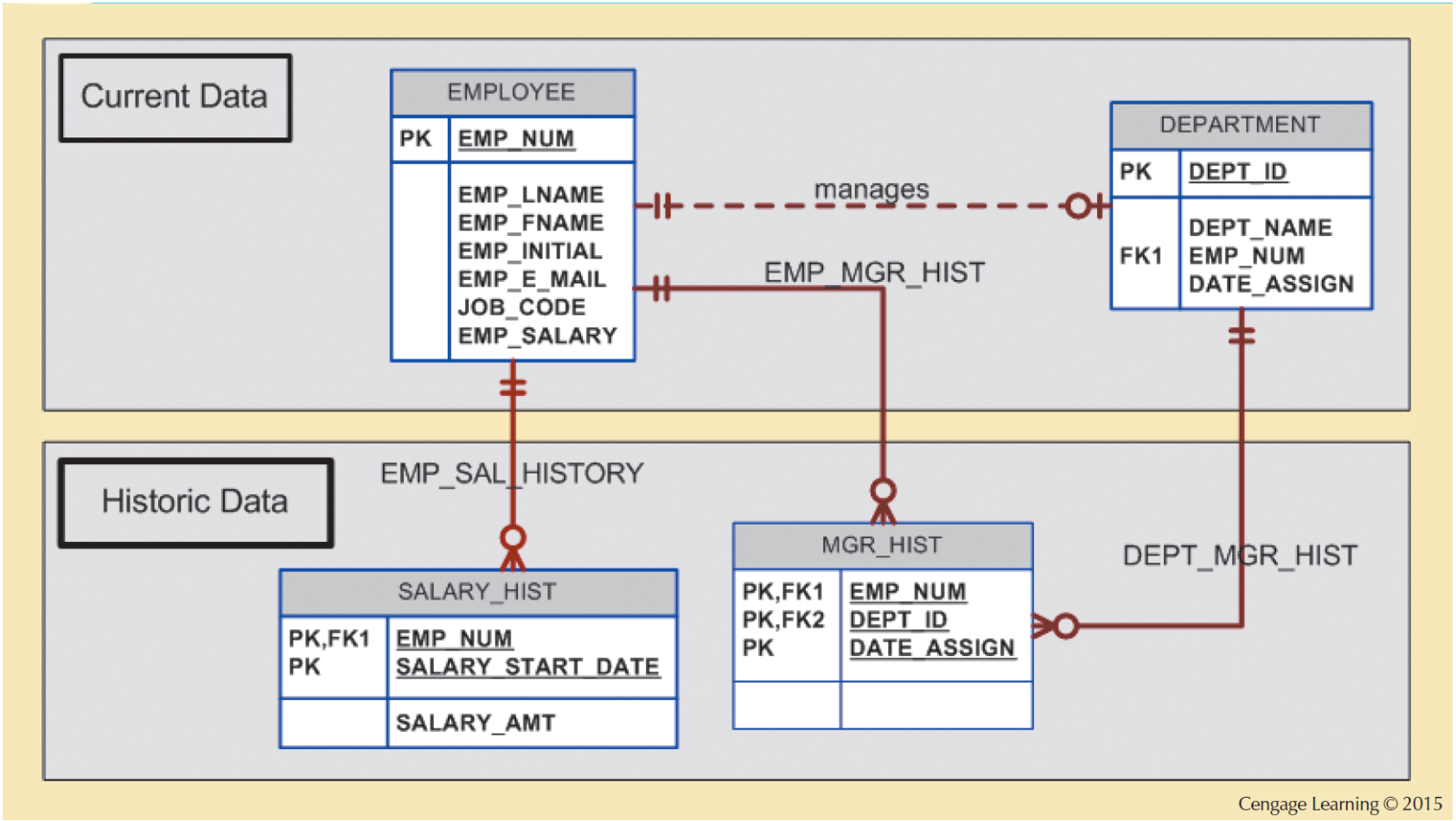
©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Design Case 2: Maintaining History of Time-Variant Data

- Data model includes data about the different departments in the organization and which employee manages each department.
- Each department is managed by only one employee and each employee can mange one department.
- 1:1 relationship would exist between EMPLOYEE and DEPARTMENT.
- You want to keep track of the history of all department mangers as well as the current manager.

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Figure 5.9 - Maintaining Manager History



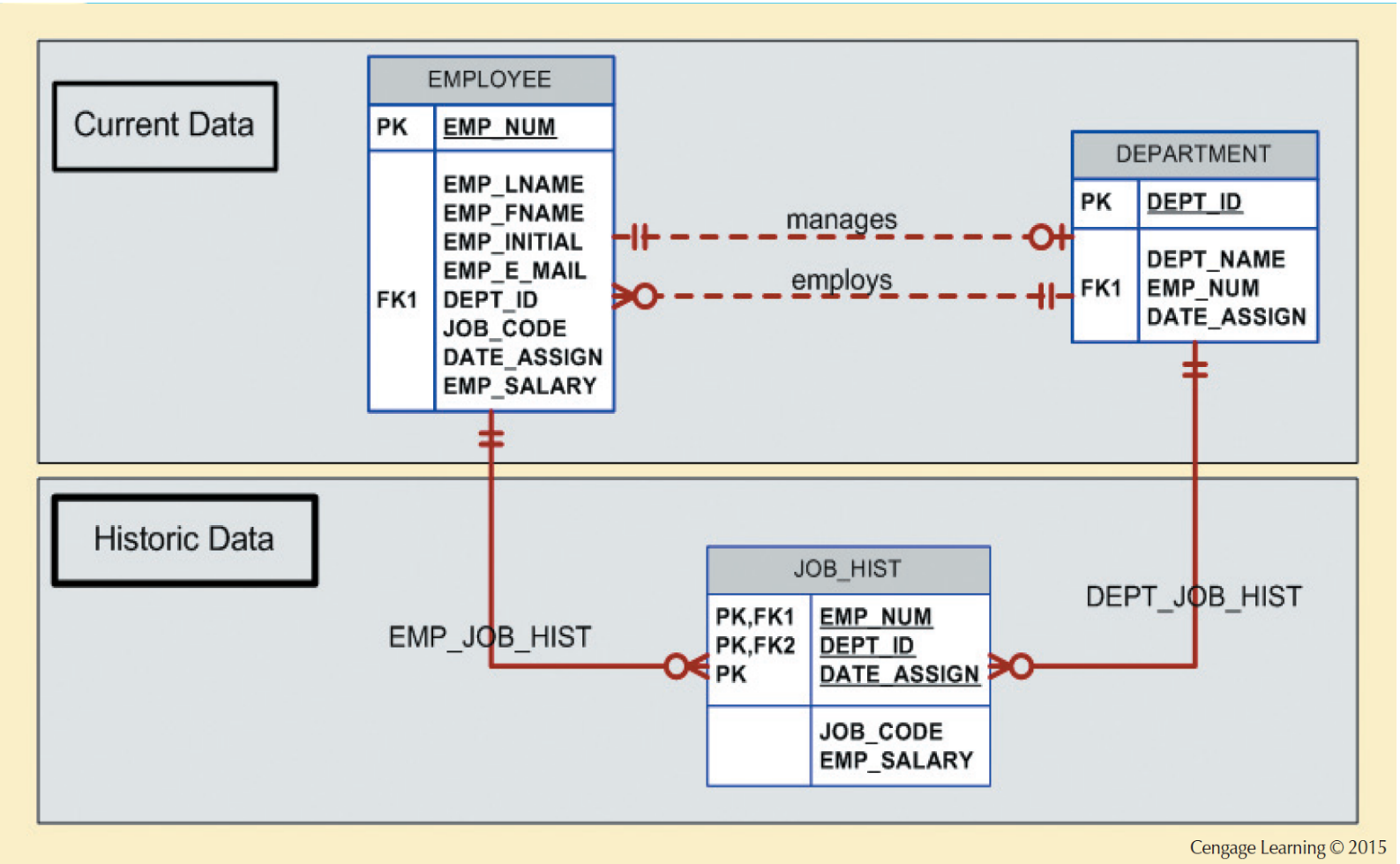
©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Design Case 2: Maintaining History of Time-Variant Data

- Now suppose you would like to keep track of the job history for each of the company's employees.
- You want to store the department, the job code, the date assigned, and the salary.

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Figure 5.10 - Maintaining Job History



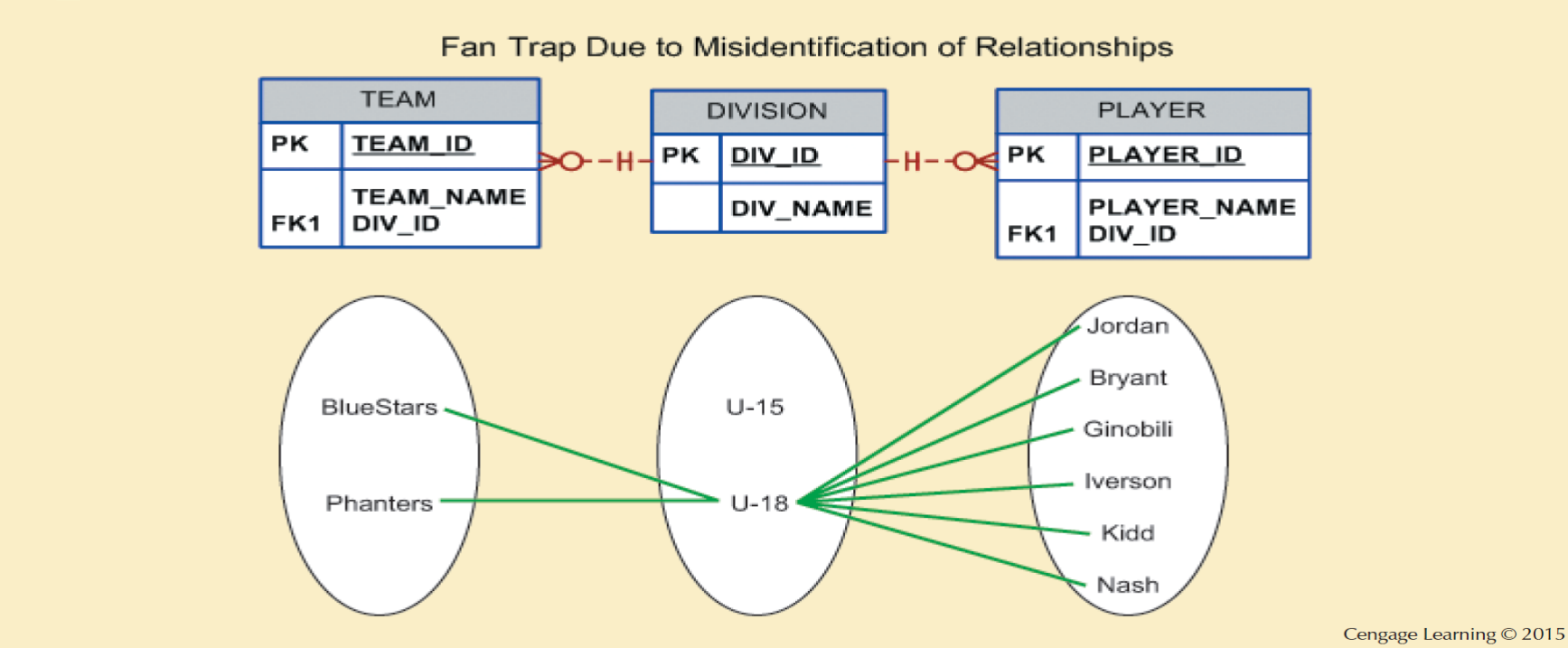
©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Design Case 3: Fan Traps

- Due to miscommunication or incomplete understanding of the business rules or processes, it is not uncommon to misidentify relationship among entities.
- **Design trap**: Occurs when a relationship is improperly or incompletely identified
  - Represented in a way not consistent with the real world
- **Fan trap**: Occurs when one entity is in two 1:M relationships to other entities
  - Produces an association among other entities not expressed in the model

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

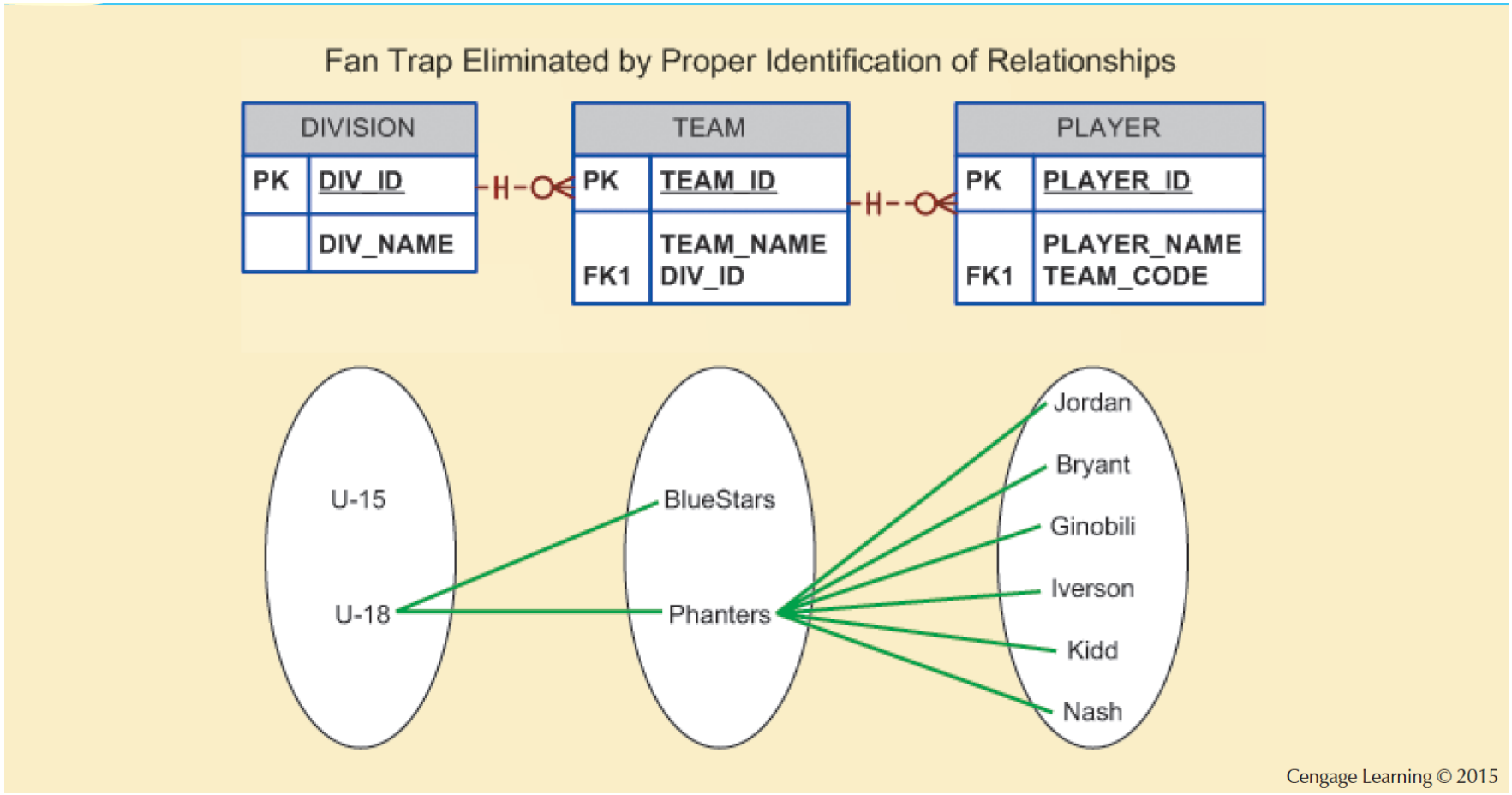
Figure 5.11 - Incorrect ERD with Fan Trap Problem



- There is no way to identify which players belong to which team.

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Figure 5.12 - Corrected ERD After Removal of the Fan Trap



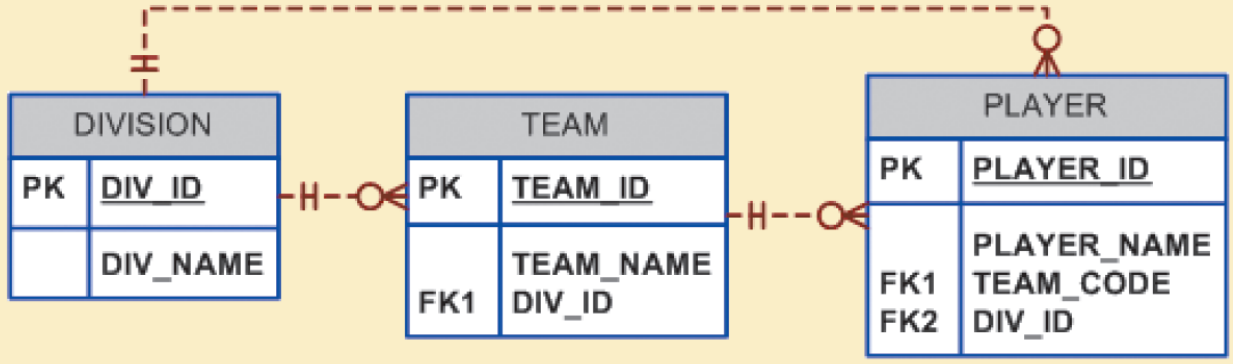
©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Design Case 4: Redundant Relationships

- Occur when there are multiple relationship paths between related entities
- Need to remain consistent across the model
- Help simplify the design

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Figure 5.13 - A Redundant Relationship



Cengage Learning © 2015

- There is a transitive 1:M relationship between DIVISION and PLAYER through the TEAM entity set.
- The relationship that connects DIVISION and PLAYER is redundant.
- The relationship could be safely deleted with out losing any information generation capabilities in the model.

©2015 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.