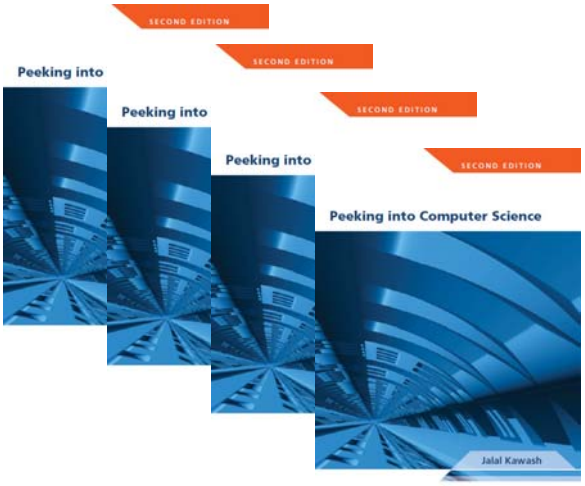
1 Databases & Data Modeling

Peeking into Computer Science

- Mandatory: Chapter 4 – Sections 4.1 to 4.3

Reading Assignment



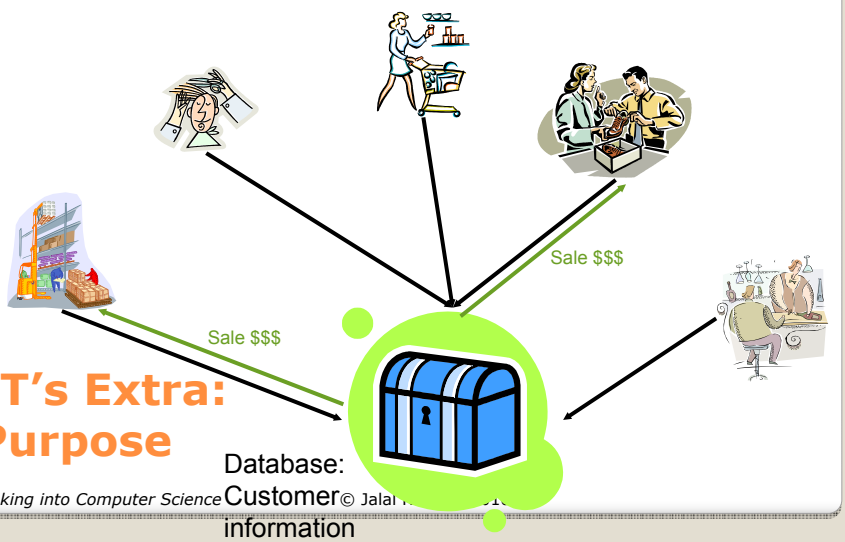
Databases

At the end of this section, you will be able to:

1. Describe what a database is
2. Draw the relationship between databases and mathematical relations
3. Describe what a database schema is

Objectives

- To store & retrieve information



JT's Extra: Purpose

- Why bother, why not use a simple file as an alternative?
  - E.g., tracking client information

MILES EDWARD O'BRIAN
DS9 Corp
Electrical engineering
2007 purchases: $10,0000,000
2006 purchases: $1,750,000

JAMIE SMYTHE
Cooperative services
Gasoline refining
2006 purchases: $5,000,0000
2005 purchases: $5,000,0000
2004 purchases: $5,000,0000
2003 purchases: $5,000,0000
2002 purchases: $5,000,0000

SCOTT BRUCE
Bryce Consulting
Investment analysis
2007 purchases: $500,000
2006 purchases: $1,500,000
2005 purchases: $2,500,000
2004 purchases: $500,000

ETC.

- If the list is short then a simple text file may suffice.
- As the list grows organizing and updating the information becomes more challenging (duplicates or inaccuracies?)
- Validity must be manually checked.
- Also searching the list according to specific criteria may become difficult .
  - e.g., Show all clients whose purchases in 2007 were between one and five million dollars
  - e.g., Show all clients that made in one year a purchase exceeding 10 million dollars.

JT's Extra: Why Bother?

- Organized collection of data
- Minimizes redundancy:
  - Wastes space and produce anomalies
- Makes it easier to access and modify data
- Examples: University and bank records
- Typically is a collection of *tables*

Database

EMPLOYEE

SIN	Fname	Lname	DOB	Gender	Salary	Number	Street	City	Pcode	Dnumber
171717171	Debra	Beacon	15-Aug-1961	Female	70000	15	Baron Hill	Calgary	T2X Y0Y	1
181817178	Sam	Field	17-Feb-1978	Male	40000	15	Kick Way	Calgary	Y2K K0K	1
123456789	Rajeet	Folk	30-Apr-1967	Male	78000	123	One Road	Toronto	H1H J9J	2
987654321	Marie	Band	12-Jan-1985	Female	53500	2828	Exit Close	Toronto	K8O O8K	2
666333999	Saleh	Dice	25-Mar-1970	Male	90400	66	Straight Way	Toronto	T4E T6B	3

DEPARTMENT

Dnumber	Dname	MGR_SIN	StartDate
1	IT	171717171	12-Feb-2008
2	Finance	123456789	1-Mar-2002
3	Marketing	666333999	1-Jan-2005

PROJECT

Pnumber	Pname	Location	Dnumber
1	Web Shopping	Calgary	1
2	Network Upgrade	Calgary	1
3	New Benefits	Toronto	2
4	Product XT345	Toronto	3

PROJ_EMP

SIN	Pnumber	Hours
171717171	1	15
171717171	2	20
171717171	4	5
181817178	1	30
181817178	2	10
123456789	3	40
666333999	4	40

Example Database

- Information is commonly stored in tables (relational database):

'Employees' table

SIN	LastName	FirstName	Address	City	Province
638666670	Cartland	Douglas	1109, 4944 Dalworth Dr	Silent Hill	Alberta
456789123	Cartman	Eric	456 Lynchview Road	Southpark	Alberta
670380456	Edgar	Maureen	300, Lockinvar Road	Calgary	Alberta
456889123	Flanders	Ned	60 Evergreen Terrace	Springfield	Alberta
413754621	Kennedy	Leon	808, 4900 Wildman Ave	Racoon City	Alberta
456438624	Lemoy	Leonard	55 Logic Way	Vulcan	Alberta
666666667	Mason	Harry	7 Luckstone Dr	Silent Hill	Alberta
666666666	Morris	Heather	7 Luckstone Dr	Silent Hill	Alberta
444638047	Redfield	Claire	653 Wildpark Place	Racoon City	Alberta
123115323	Simcox	Cole	311 Ocean View Drive	Vancouver	British C
456789124	Simpson	Homer	59 Evergreen Terrace	Springfield	Alberta
123456789	Smith	John	123 Peanut Lane	Calgary	Alberta
666666668	Sunderland	James	7 Heartbroken Ave	Silent Hill	Alberta
620451097	Williams	Amanda	25 Rodeo Drive	Edmonton	Alberta
666666669	Wolf	Claudia	66 Twisted View	Silent Hill	Alberta
371988812	Carswell	Mary	425 Remington Ave	Calgary	Alberta

JT's Extra: Storing Information In A Database

- Record: An example instance (row) of data within the table.

Records of the table (rows)

SIN	LastName	FirstName	Address	City	Province
638666670	Cartland	Douglas	1109, 4944 Dalworth Dr	Silent Hill	Alberta
456789123	Cartman	Eric	456 Lynchview Road	Southpark	Alberta
670380456	Edgar	Maureen	300, Lockinvar Road	Calgary	Alberta
456889123	Flanders	Ned	60 Evergreen Terrace	Springfield	Alberta
413754621	Kennedy	Leon	808, 4900 Wildman Ave	Racoon City	Alberta
456438624	Lemoy	Leonard	55 Logic Way	Vulcan	Alberta
666666667	Mason	Harry	7 Luckstone Dr	Silent Hill	Alberta
666666666	Morris	Heather	7 Luckstone Dr	Silent Hill	Alberta
444638047	Redfield	Claire	653 Wildpark Place	Racoon City	Alberta
123115323	Simcox	Cole	311 Ocean View Drive	Vancouver	British C
456789124	Simpson	Homer	59 Evergreen Terrace	Springfield	Alberta
123456789	Smith	John	123 Peanut Lane	Calgary	Alberta
666666668	Sunderland	James	7 Heartbroken Ave	Silent Hill	Alberta
620451097	Williams	Amanda	25 Rodeo Drive	Edmonton	Alberta
666666669	Wolf	Claudia	66 Twisted View	Silent Hill	Alberta
371988812	Carswell	Mary	425 Remington Ave	Calgary	Alberta

One record, 'Simpson, Homer'

JT's Extra: Storing Information In A Database (2)

- Field: are attributes used to describe each record in a table

Fields of the table (columns)

SIN	LastName	FirstName	Address	City	Province
638666670	Cartland	Douglas	1109, 4944 Dalhousie Dr	Silent Hill	Alberta
456789123	Cartman	Eric	456 Lynchview Road	Southpark	Alberta
670380456	Edgar	Maureen	300, Lockinvar Road	Calgary	Alberta
456889123	Flanders	Ned	60 Evergreen Terrace	Springfield	Alberta
413754621	Kennedy	Leon	808, 4900 Wildman Ave	Racoon City	Alberta
456438624	Lemoy	Leonard	55 Logic Way	Vulcan	Alberta
666666667	Mason	Harry	7 Luckstone Dr	Silent Hill	Alberta
666666666	Morris	Heather	7 Luckstone Dr	Silent Hill	Alberta
444638047	Redfield	Claire	653 Wildpark Place	Racoon City	Alberta
123115323	Simcox	Cole	311 Ocean View Drive	Vancouver	British C
456789124	Simpson	Homer	59 Evergreen Terrace	Springfield	Alberta
123456789	Smith	John	123 Peanut Lane	Calgary	Alberta
666666668	Sunderland	James	7 Heartbroken Ave	Silent Hill	Alberta
620451097	Williams	Amanda	25 Rodeo Drive	Edmonton	Alberta
666666669	Wolf	Claudia	66 Twisted View	Silent Hill	Alberta
371988812	Carswell	Mary	425 Remington Ave	Calgary	Alberta

'Address' field describes location

JT's Extra: Storing Information In A Database (3)

1. Create a unique and descriptive name.
2. Do not use words that convey physical characteristics or database terminology.
3. While names should be short avoid using acronyms and abbreviations unless they are well-known.
4. Do not use proper names or words that will restrict the type of data to be entered into the table.
5. Consider using the *plural* form of a name.
6. Avoid the use of spaces in names.

JT's Extra: Guidelines For Naming Tables

1. Create a unique and descriptive name.
2. Create a name that accurately, clearly and unambiguously identifies the characteristic that the field represents.
3. While names should be short avoid using acronyms and abbreviations unless they are well-known.
4. Use the *singular* form of a name.
5. Avoid the use of spaces in names.

JT's Extra: Guidelines For Naming Fields

PROJECT

Pnumber	Pname	Location	Dnumber
1	Web Shopping	Calgary	1
2	Network Upgrade	Calgary	1
3	New Benefits	Toronto	2
4	Product XT345	Toronto	3

$$\text{PROJECT} = \{(1, \textit{Web Shopping}, \textit{Calgary}, 1), (2, \textit{Backup}, \textit{Calgary}, 1), (3, \textit{New benefits}, \textit{Toronto}, 2), (4, \textit{XT345}, \textit{Toronto}, 3)\}.$$

Relations

- Let:
  - N be the set of natural numbers
  - M be the set of names
  - L be the set of locations
- Then:
  - PROJECT $\subseteq$ N x M x L

PROJECT

Pnumber	Pname	Location
1	Web Shopping	Calgary
2	Network Upgrade	Calgary
3	New Benefits	Toronto
4	Product XT345	Toronto

Relations

EMPLOYEE

SIN	Fname	Lname	DOB	Gender	Salary	Number	Street	City	Pcode	Dnumber

DEPARTMENT

Dnumber	Dname	MGR_SIN	StartDate

PROJECT

Pnumber	Pname	Location	Dnumber

PROJ_EMP

SIN	Pnumber	Hours

JT's Extra (database schema)

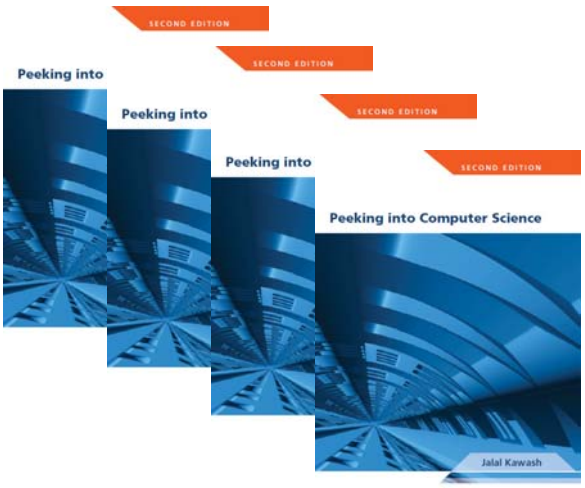
- Tables?
- Fields?
- Relationships?

Database Schema

	SIN	Last name	Given names	Sales
	111111111	Tam	James	$1

	SIN	Last name	Given names	Position
	222222222	Gtrezky	Wayne	Center

JT's Extra: Example Database Schema



Data Modeling

Entity-Relationship Model

At the end of this section, you will be able to:

1. Understand ER the model
   - Understand and differentiate between entity types, entities, relationship types, and relationships
   - Understand attributes and primary keys
2. Understand relationship type cardinality
3. Understand universal and existential participation in relationships
4. Use ER diagrams to design data models

Objectives

- Entity: an object that exists in the real world
  - Physically: book, car, student
  - Conceptually: job, route
- Entity-type: a class of entities
  - Employee
  - Project
  - Department

Entities

- Entities have attributes
  - Properties that describe entities
- An Employee can be described by:
  - SIN
  - Name
  - DOB
  - Gender
  - Address

Attributes

- They are used to graphically represent a database.
- An ERD shows:
  - Tables,
  - Fields of a table
  - Relationships between tables (more on this later).

JT's Extra: ERD's (Entity-relation diagrams)

EMPLOYEE
- SIN
- First name
- Last name
- DOB
- Gender
- Salary
- Number
- Street
- City
- Postal Code

PROJECT
- Number
- Name
- Location

DEPARTMENT
-Number
-Name

Entity Types in ER Diagrams

- Each table should typically have one field designated as the primary key:
  - The primary key must be unique (identifies one record from another)

Primary Key for table 'Employees' is the 'SIN' field

SIN	LastName	FirstName	Address	City	Province
638666670	Cartland	Douglas	1109, 4944 Dalworth Dr	Silent Hill	Alberta
456789123	Cartman	Eric	456 Lynchview Road	Southpark	Alberta
670380456	Edgar	Maureen	300, Lockinvar Road	Calgary	Alberta
456889123	Flanders	Ned	60 Evergreen Terrace	Springfield	Alberta
413754621	Kennedy	Leon	808, 4900 Wildman Ave	Racoon City	Alberta
456438624	Lemoy	Leonard	55 Logic Way	Vulcan	Alberta
666666667	Mason	Harry	7 Luckstone Dr	Silent Hill	Alberta
666666666	Morris	Heather	7 Luckstone Dr	Silent Hill	Alberta
444638047	Redfield	Claire	653 Wildpark Place	Racoon City	Alberta
123115323	Simcox	Cole	311 Ocean View Drive	Vancouver	British C
456789124	Simpson	Homer	59 Evergreen Terrace	Springfield	Alberta
123456789	Smith	John	123 Peanut Lane	Calgary	Alberta
666666668	Sunderland	James	7 Heartbroken Ave	Silent Hill	Alberta
620451097	Williams	Amanda	25 Rodeo Drive	Edmonton	Alberta
666666669	Wolf	Claudia	66 Twisted View	Silent Hill	Alberta
371988812	Carswell	Mary	425 Remington Ave	Calgary	Alberta

JT's Extra: Primary Key

- Primary Key: a collection of attributes the uniquely identify an entity
  - One attribute most of the time
- SIN for employee
- Student ID
- Underlined in ERD

Primary Keys

- A primary key must be unique to each record because it is the one thing that distinguishes them.
- If there is at least (or even exactly) one instance (however unlikely) where records can take on the same value for a field then that field cannot be a primary key. (When in doubt if this will ever be the case then verify with your users).
- If a single key field cannot be found then several fields can be combined into a composite key. (Each field is still a separate field but together they form a unique primary key for each record).

JT's Extra: Choosing A Primary Key

- If a unique primary key still cannot be found then 'invent' one.

JT's Extra: Choosing A Primary Key (2)

- One-to-one
- One-to-many (many-to-one)
- Many-to-many

Cardinality (JT: Multiplicity) of Relationships

1. One to one relationships
   - One entity participates in the relationship from the 'left' and one entity participates in the relationship from the 'right'.
   - Person : head
   - Worker : Social Insurance Number
   - This type of relationship is rare in databases
2. One to many relationships
   - On one side of the relationship one entity participates in the relationship while on the other side: zero or more entities may participate in the relationship.
   - Person : Hair
   - Department : Employee

JT's Extra: Cardinality

3. Many to many relationships
   - On each side of the relationship zero or more entities may participate in the relationship.
   - Students : Classes

JT's Extra: Cardinality (2)

3. Many to many relationships
   - This type of relationship is not directly implemented in databases:

Students table

StudentID	**StudentFirstName**	**StudentLast Name**	**StudentPhone**
123456	Jamie	Smyth	553-3992
123457	Stacey	Walls	790-3992
123458	Angel	Lam	551-4993

Classes table

ClassName	***ClassNumber***	***Lecture No***	**ClassDescription**
CPSC	*203*	*01*	Introduction to Problem…
CPSC	*231*	*01*	Introduction to Computer..
CPSC	*233*	*01*	Introduction to Computer..

JT's Extra: Cardinality (3)

3. Many to many relationships
   - Typically implemented as two one to many relationships in databases:

Students table

StudentID	**StudentFirstName**	**…**
123456	Jamie	
123457	Stacey	

Classes table

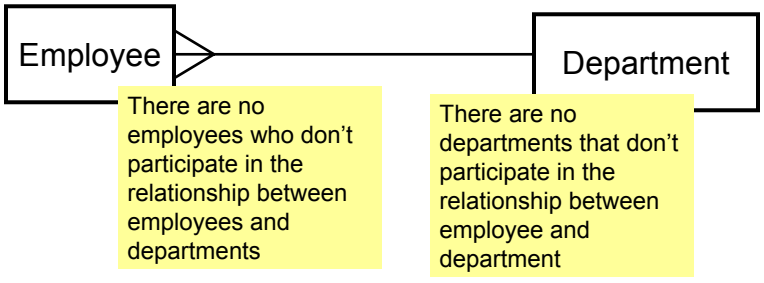
ClassName	**ClassNumber**	**…**
CPSC	203	
CPSC	231	

Registrations table (linking table)

StudentID	**ClassName**	**Class-Number**	**Lecture No**
123450	ENGL	201	01
123457	CPSC	203	01
123460	MATH	271	01

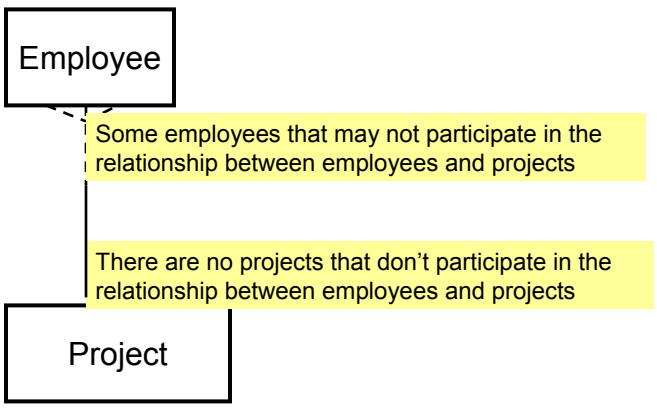
JT's Extra: Cardinality (4)

- Represented using a solid line.

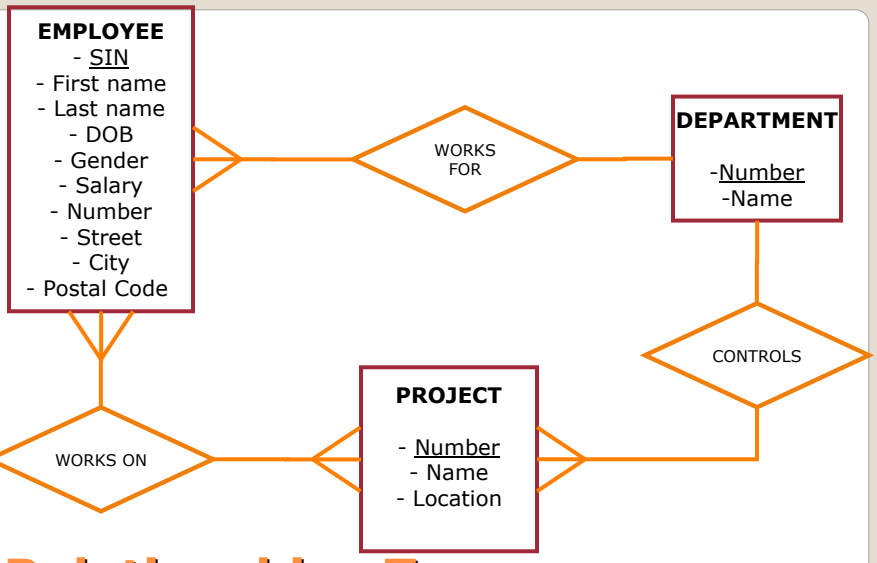


JT's Extra: Full Participation (ERD Representation)

- Represented using a dashed line.

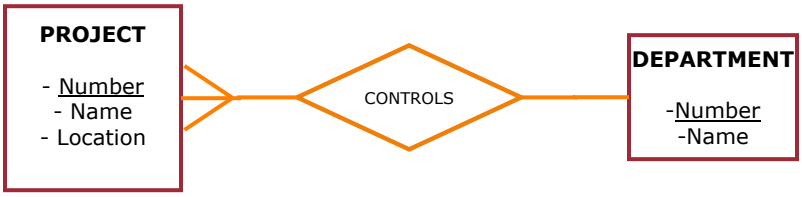


JT's Extra: Partial Participation (ERD Representation)



Relationships Types

- The degree of a relationship is the number of entity types it relates
- Controls is a binary relationship type

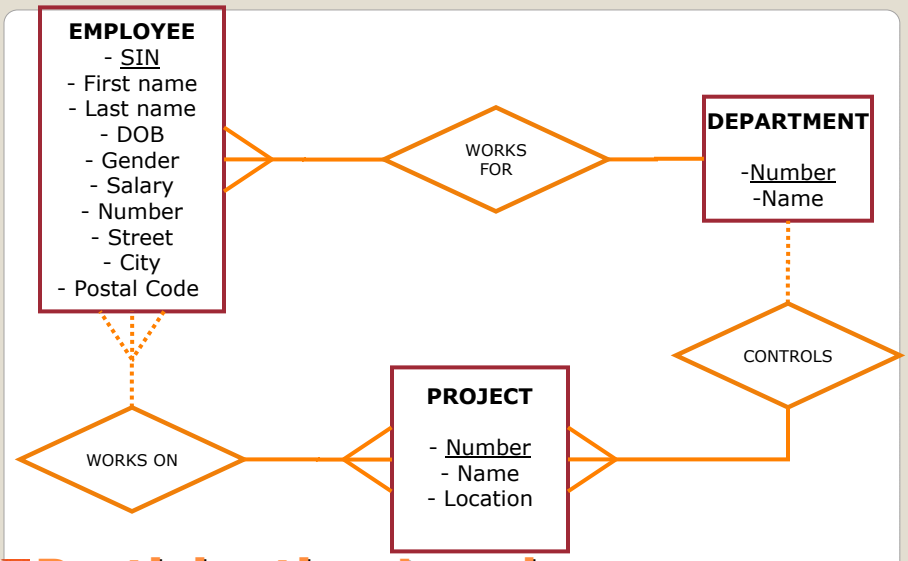


Relationship Degrees

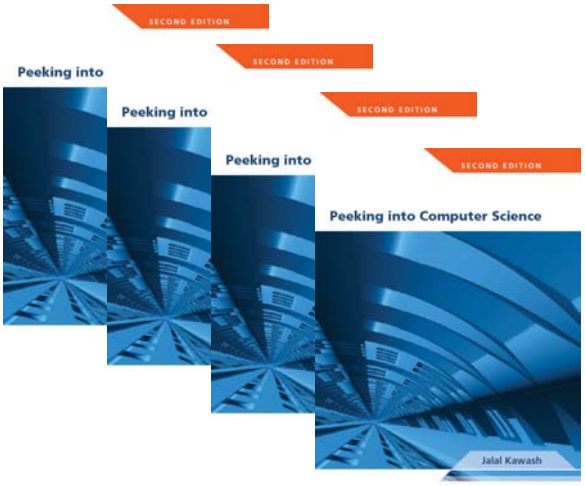
- Entity types participate
  - Fully (universal participation)
  - Partially (existential participation)

in relationship types

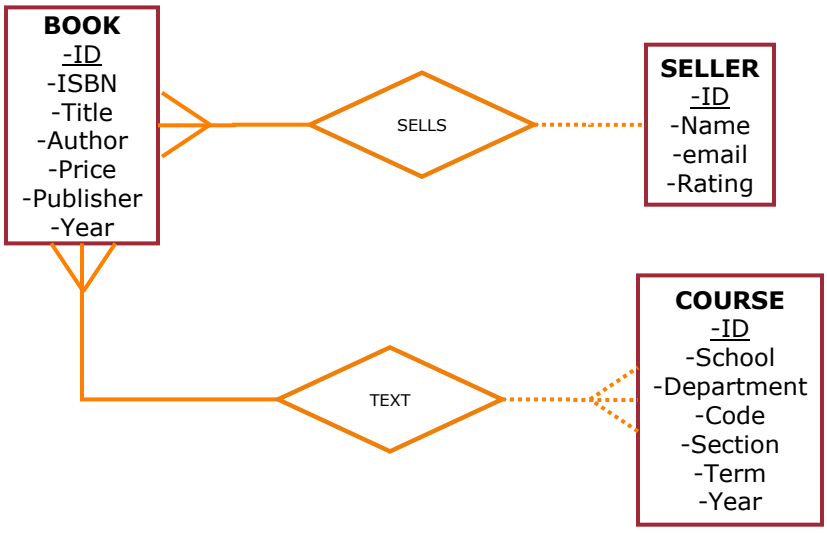
Participation Levels

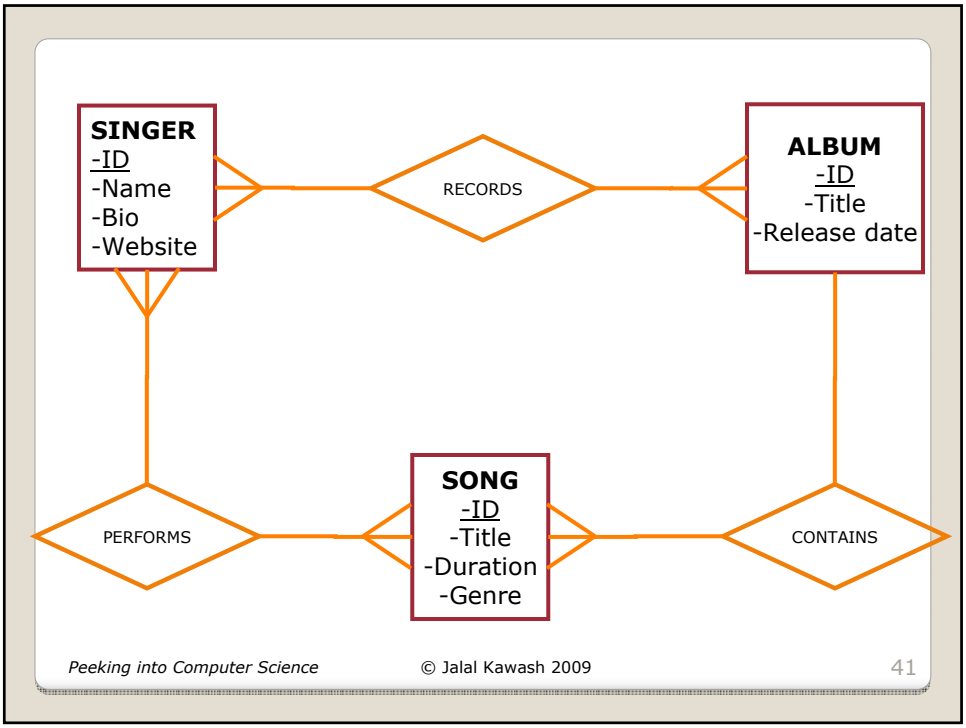


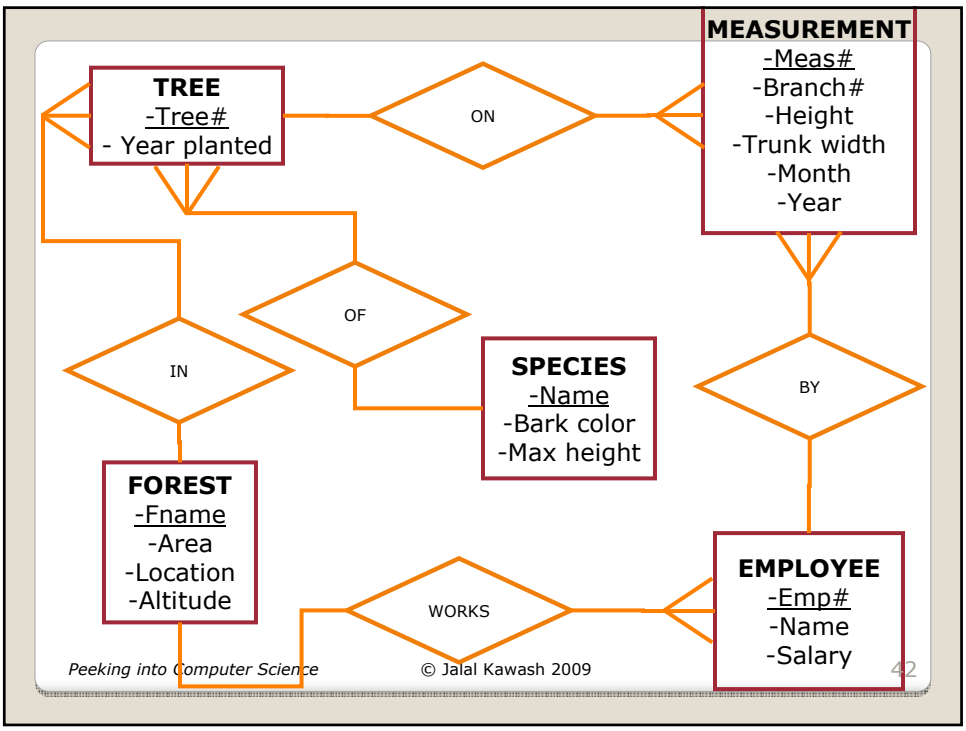
Participation Levels



ERD Examples




SINGER
-ID
-Name
-Bio
-Website
RECORDS
ALBUM
-ID
-Title
-Release date
SONG
-ID
-Title
-Duration
-Genre
PERFORMS
CONTAINS


MEASUREMENT
-Meas#
-Branch#
-Height
-Trunk width
-Month
-Year
TREE
-Tree#
- Year planted
ON
OF
IN
SPECIES
-Name
-Bark color
-Max height
BY
FOREST
-Fname
-Area
-Location
-Altitude
WORKS
EMPLOYEE
-Emp#
-Name
-Salary