MIS 0855 Spring 2015 – Data Science

Day 32 – *Designing Data*

Min-Seok Pang

Management Information Systems
Fox School of Business, Temple University
minspang@temple.edu
Apr. 3rd, 2015

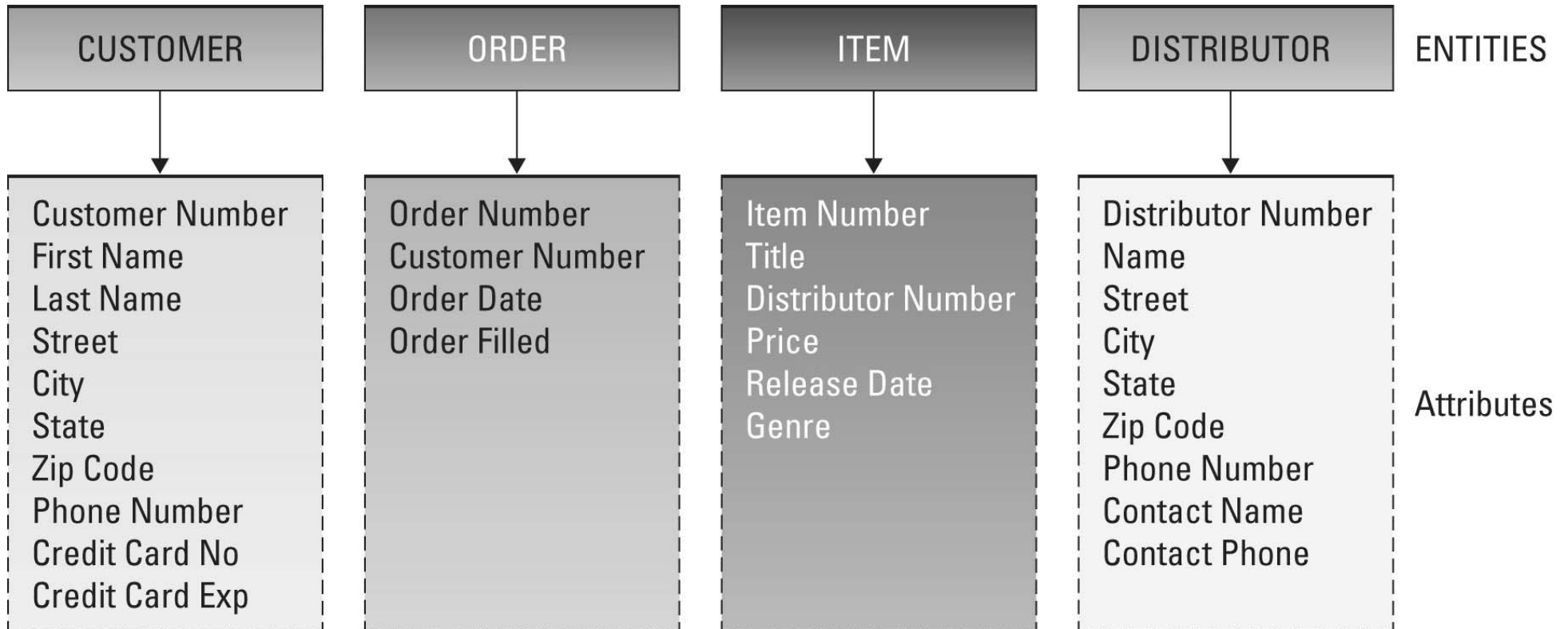Fox School of Business
TEMPLE UNIVERSITY

Relational Data Model

- <u>Data Model</u>: a formal way to express data relationships to a database

- <u>Relational Data Model</u>: a type of model that represents its information in the form of logically-related *2-D tables*

  - The most common, intuitive, *de facto* model

  - consists of Entities, Attributes, and Relationships

- Entity-Relationship Diagram (ERD)

Entities and Attributes (1/2)

● <u>Entity</u>, aka *table*, is an object or event about which information is stored.

   ▪ Object: e.g. Customer, Product, Employee, Factory

   ▪ Event: e.g. Order, Registration, Contract, Payment

   ▪ A company is usually not an entity unless a database stores information of *multiple* companies.

● <u>Attributes</u>, aka *fields* or *columns*, are characteristics or proper-ties of an entity.

   ▪ A customer (entity) can be described by *customer number, name, address, phone number* (attributes).
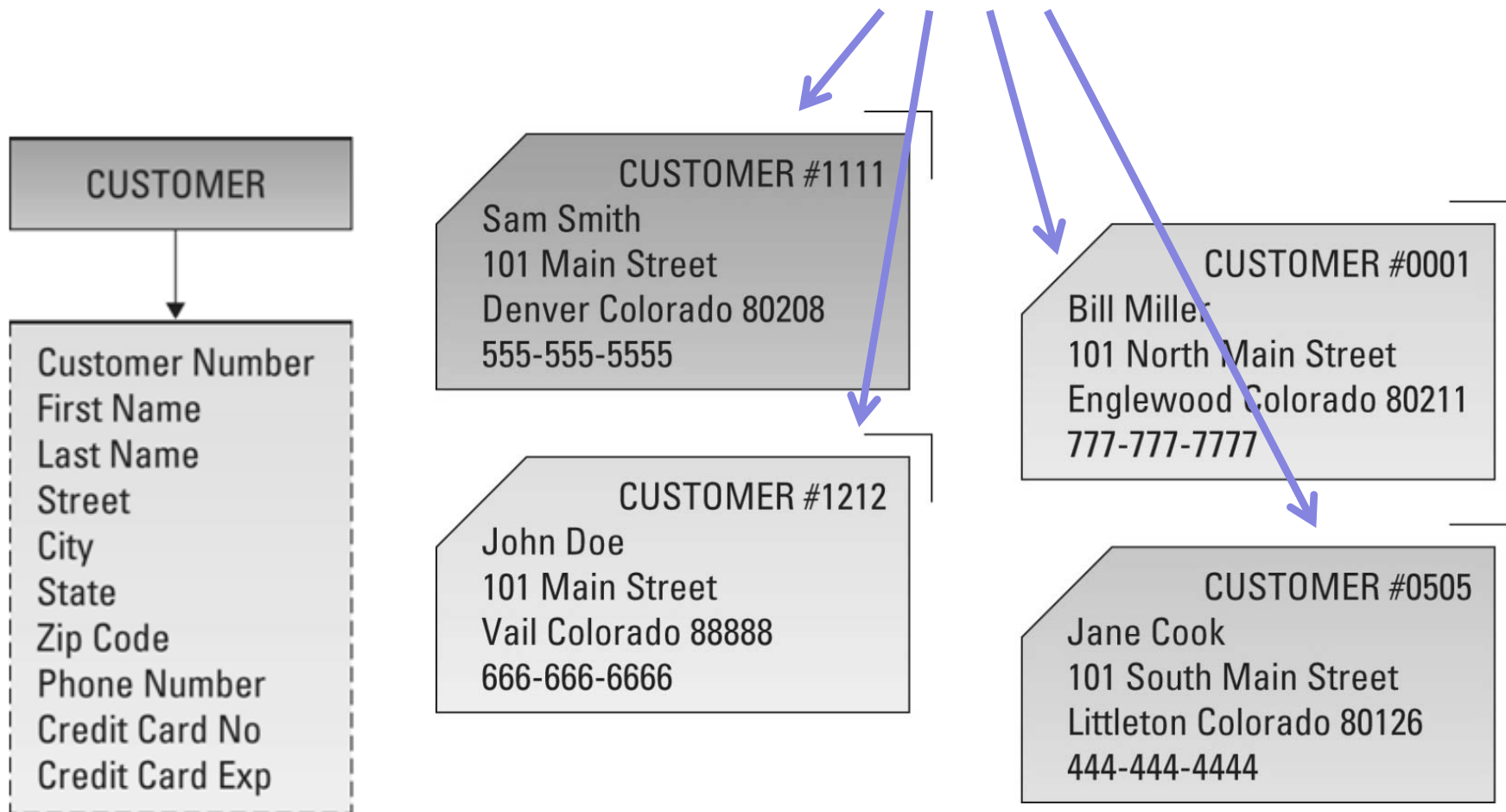
Entities and Attributes (2/2)

CUSTOMER	ORDER	ITEM	DISTRIBUTOR	ENTITIES
Customer Number First Name Last Name Street City State Zip Code Phone Number Credit Card No Credit Card Exp	Order Number Customer Number Order Date Order Filled	Item Number Title Distributor Number Price Release Date Genre	Distributor Number Name Street City State Zip Code Phone Number Contact Name Contact Phone	Attributes

● How to distinguish between Entity and Attribute?

▪ An Attribute is a <u>characteristic or property</u> of an Entity.

▪ An Attribute becomes a column of a Table.

Instances

● An entity consists of multiple <u>Instances</u>, aka records or rows.

CUSTOMER

| Customer Number |
| First Name |
| Last Name |
| Street |
| City |
| State |
| Zip Code |
| Phone Number |
| Credit Card No |
| Credit Card Exp |

CUSTOMER #1111
Sam Smith
101 Main Street
Denver Colorado 80208
555-555-5555

CUSTOMER #1212
John Doe
101 Main Street
Vail Colorado 88888
666-666-6666

CUSTOMER #0001
Bill Miller
101 North Main Street
Englewood Colorado 80211
777-777-7777

CUSTOMER #0505
Jane Cook
101 South Main Street
Littleton Colorado 80126
444-444-4444

Entity

Attribute

State	Region	Population	Code	Industry	Firms	Sales ($1,000)
Alabama	East South Central	4,627,851	20100	Groceries & other foods	7,011	8,220,149
Alabama	East South Central	4,627,851	20140	Packaged liquor	3,476	895,545
Alabama	East South Central	4,627,851	20160	Drugs & health aids	5,309	6,804,147
Alabama	East South Central	4,627,851	20180	Soaps & detergents	3,667	545,909
Alabama	East South Central	4,627,851	20190	Paper & towels	3,616	471,518
Alabama	East South Central	4,627,851	20200	Men's wear	2,618	811,448
Alabama	East South Central	4,627,851	20220	Women's wear	3,319	1,787,968
Alabama	East South Central	4,627,851	20240	Children's wear	2,020	536,742
Alabama	East South Central	4,627,851	20260	Footwear	2,829	638,494
Alabama	East South Central	4,627,851	20270	Sewing materials & supplies	425	63,190
Alabama	East South Central	4,627,851	20280	Curtains & table coverings	1,681	302,326
Alabama	East South Central	4,627,851	20300	Major household appliances	985	404,591
Alabama	East South Central	4,627,851	20310	Small electric appliances	1,409	129,511

Instance

Country	Continent	Year	Population	Current GDP	International Trade
Afghanistan	Asia	2000	23,898,198	$ 291.86	$ 209.47
Afghanistan	Asia	2001	23,997,412	$ 285.88	$ 201.65
Afghanistan	Asia	2002	25,268,405	$ 491.48	$ 304.59
Afghanistan	Asia	2003	27,060,359	$ 542.84	$ 521.14
Afghanistan	Asia	2004	28,513,677	$ 583.20	$ 519.10
Afghanistan	Asia	2005	29,928,987	$ 666.48	$ 595.65
Afghanistan	Asia	2006	31,056,997	$ 729.94	$ 592.52
Afghanistan	Asia	2007	31,889,923	$ 815.85	$ 606.68
Afghanistan	Asia	2008	24,506,839	$ 1,137.13	$ 788.99
Afghanistan	Asia	2009	25,390,062	$ 1,322.18	$ 809.86
Albania	Europe	2000	3,158,000	$ 3,101.12	$ 1,568.24

Entity Identifier (Primary Key)

● <u>Entity Identifier</u>, aka *primary key*, is an *attribute* that

- ■ ensures each instance has a *unique* value that distinguishes it from every other instance.

- ■ Every entity must have an entity identifier.

- ■ TUID #

- ■ Social Security #

- ■ Plate #

- ■ other examples?

CUSTOMER #1111
Sam Smith
101 Main Street
Denver Colorado 80208
555-555-5555

CUSTOMER #0001
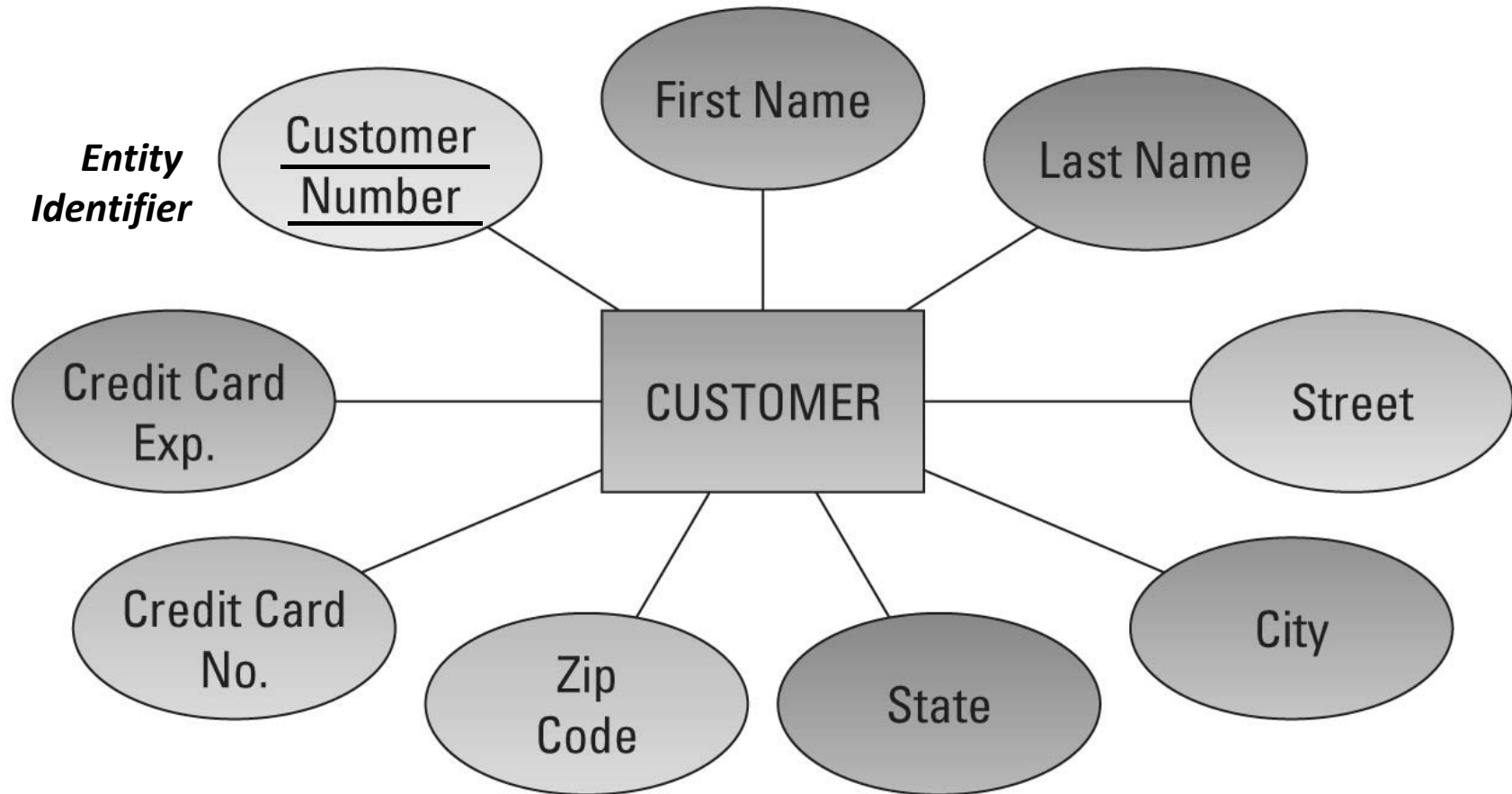Bill Miller
101 North Main Street
Englewood Colorado 80211
777-777-7777

CUSTOMER #1212
John Doe
101 Main Street
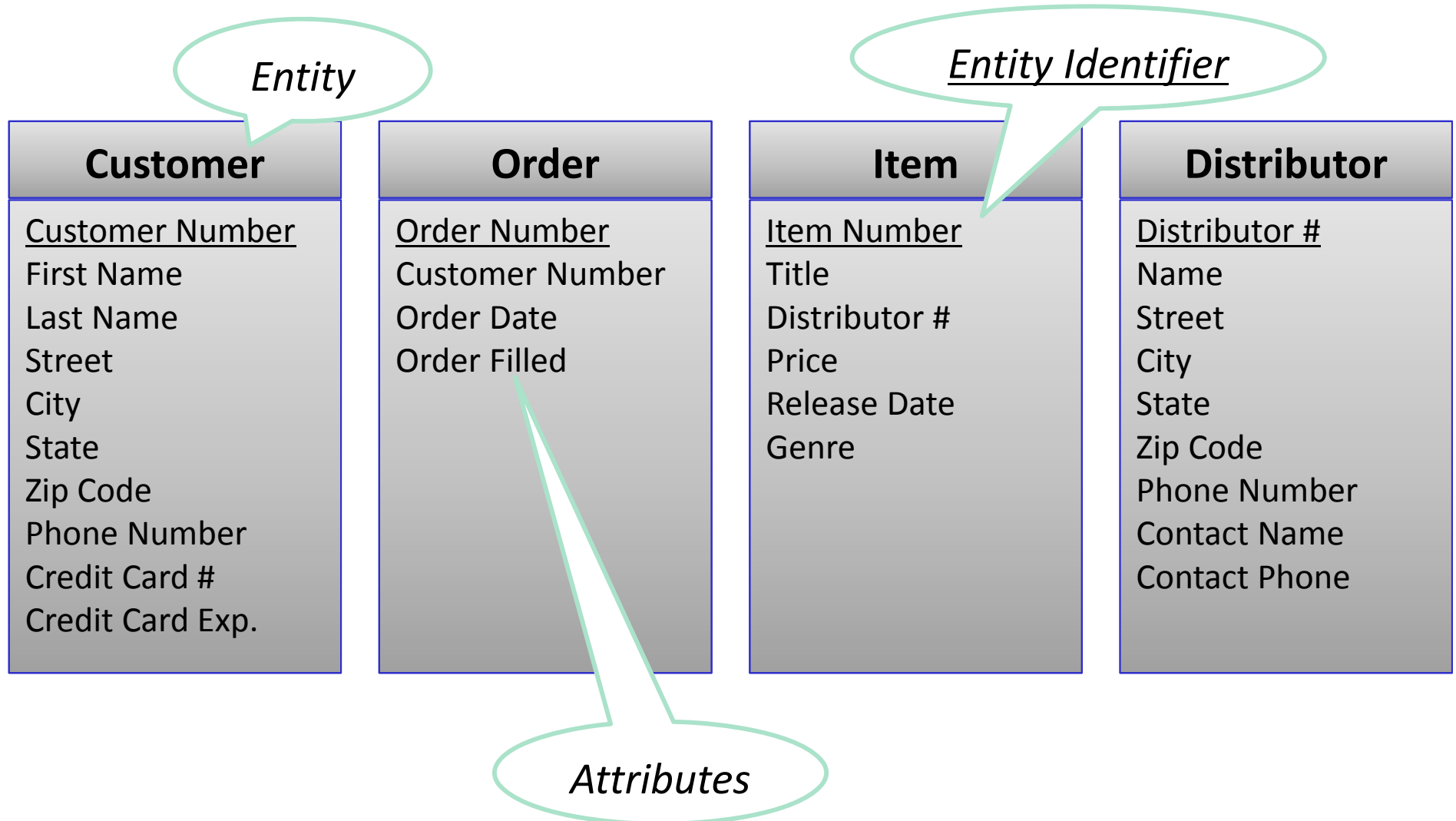Vail Colorado 88888
666-666-6666

CUSTOMER #0505
Jane Cook
101 South Main Street
Littleton Colorado 80126
444-444-4444

# How to draw Entities and Attributes (1/2)

Entity Identifier

Customer Number

First Name

Last Name

Credit Card Exp.

CUSTOMER

Street

Credit Card No.

Zip Code

State

City

How to draw Entities and Attributes (2/2)

Entity

Entity Identifier

Customer	Order	Item	Distributor
Customer Number	Order Number	Item Number	Distributor #
First Name	Customer Number	Title	Name
Last Name	Order Date	Distributor #	Street
Street	Order Filled	Price	City
City		Release Date	State
State		Genre	Zip Code
Zip Code			Phone Number
Phone Number			Contact Name
Credit Card #			Contact Phone
Credit Card Exp.			

Attributes

Example

SportTech Events puts on athletic events for local high school athletes. The company needs a database designed to keep track of the sponsor for the event and where the event is located.
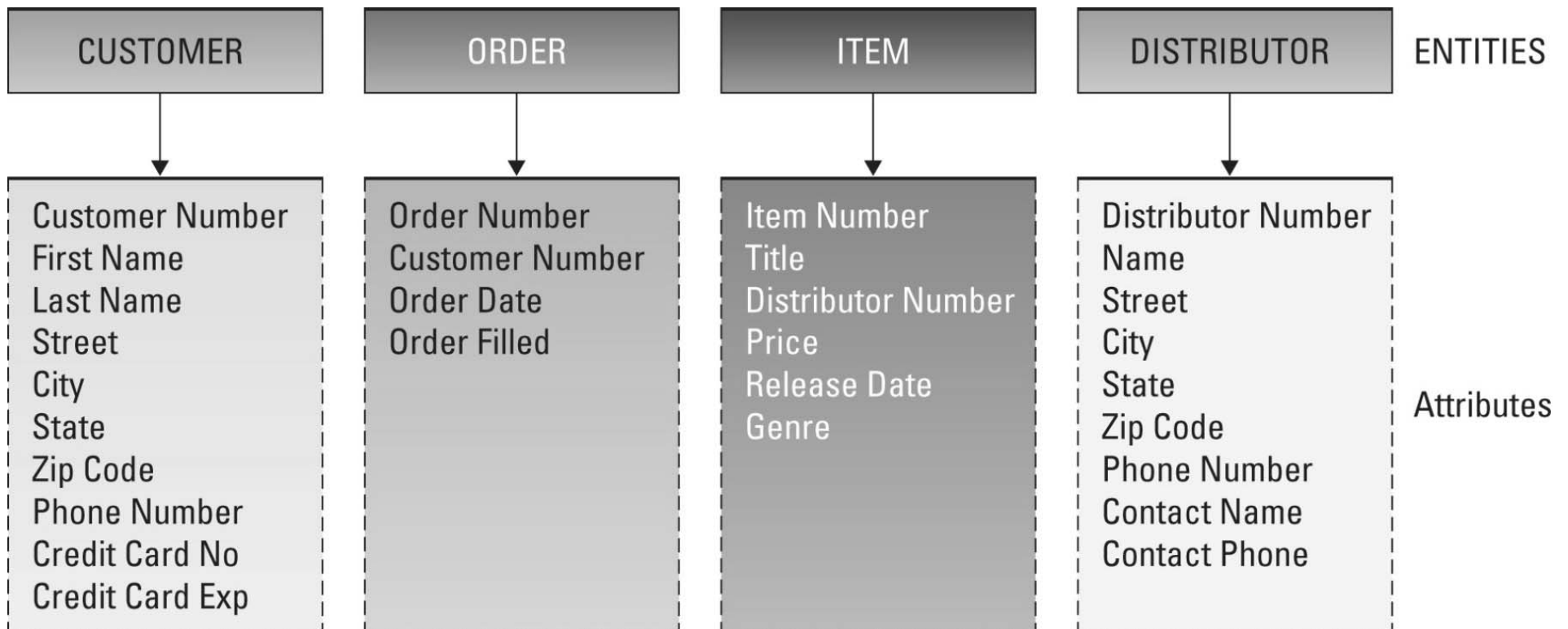
Each event needs a description, date, and cost. Separate costs are negotiated for each event. The company would also like to have a list of potential sponsors that includes each sponsor's contact information such as the name, phone number, and address.

Each event will have a single sponsor, but a particular sponsor may sponsor more than one event. Each location will need an ID, contact person, and phone number. A particular event will use only one location, but a location may be used for multiple events.
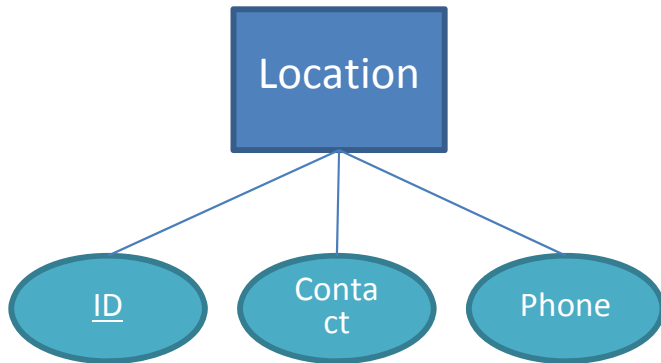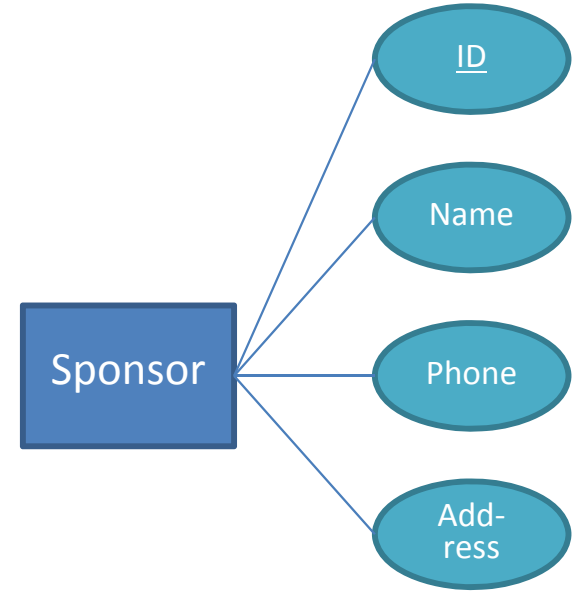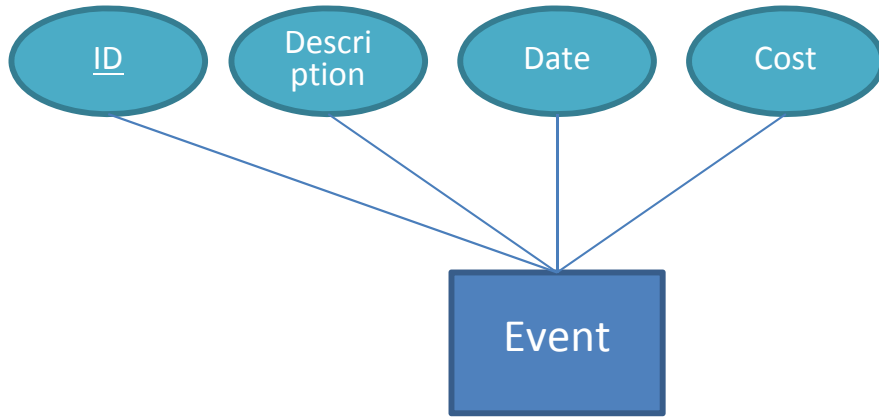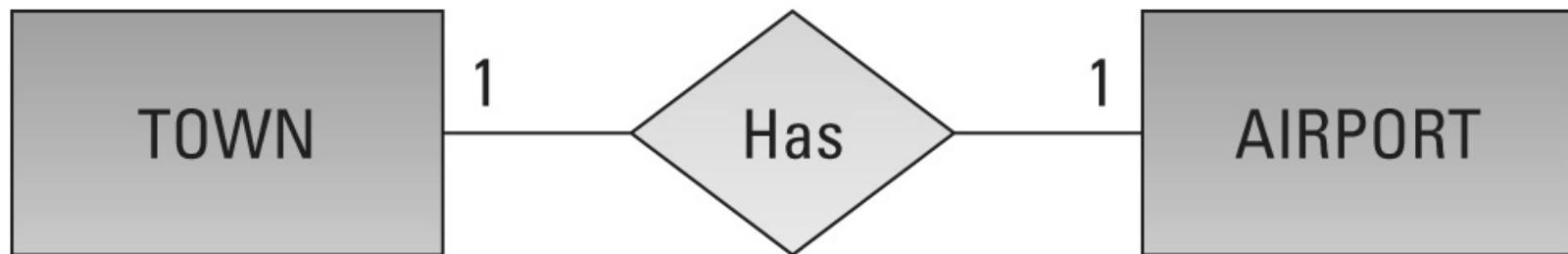
Relationship

- An entity should be related to one or more other entities.

- Any relationship between the entities below?

| CUSTOMER	ORDER	ITEM	DISTRIBUTOR	ENTITIES
Customer Number First Name Last Name Street City State Zip Code Phone Number Credit Card No Credit Card Exp	Order Number Customer Number Order Date Order Filled	Item Number Title Distributor Number Price Release Date Genre	Distributor Number Name Street City State Zip Code Phone Number Contact Name Contact Phone	Attributes

One-to-One (1:1) Relationship

● A relationship between two entities in which

▪ an instance of entity A can be related to <u>only one instance of entity B</u> and

▪ an instance of entity B can be related to <u>only one instance of entity A</u>.

Name of relationship usually with a *verb*

One-to-Many (1:M) Relationship

● A relationship between two entities, in which

▪ an instance of entity A can be related to <u>one or more instances of entity B</u> and

▪ an instance of entity B can be related to <u>only one instance of entity A</u>.
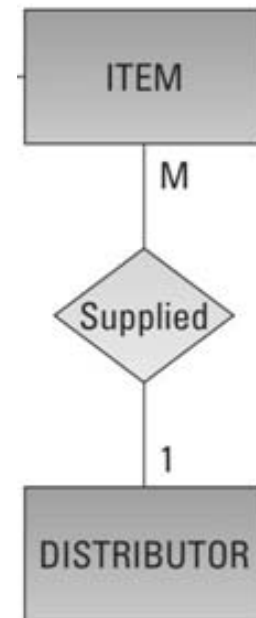
Many-to-Many (M:N) Relationship

● A relationship between two entities, in which

▪ an instance of entity A can be related to <u>one or more instances of entity B</u> and

▪ an instance of entity B can be related to <u>one or more instances of entity A</u>.
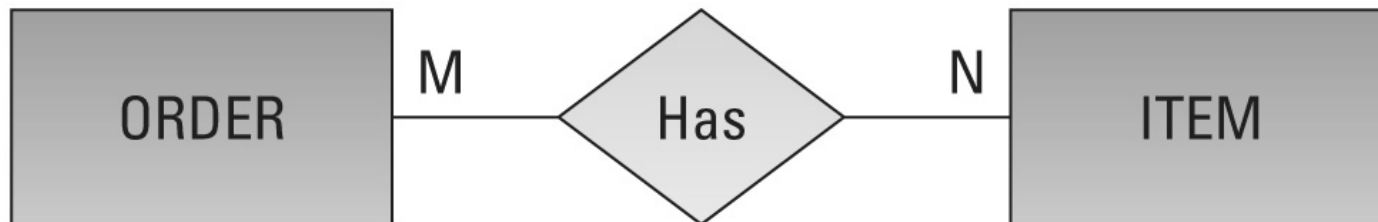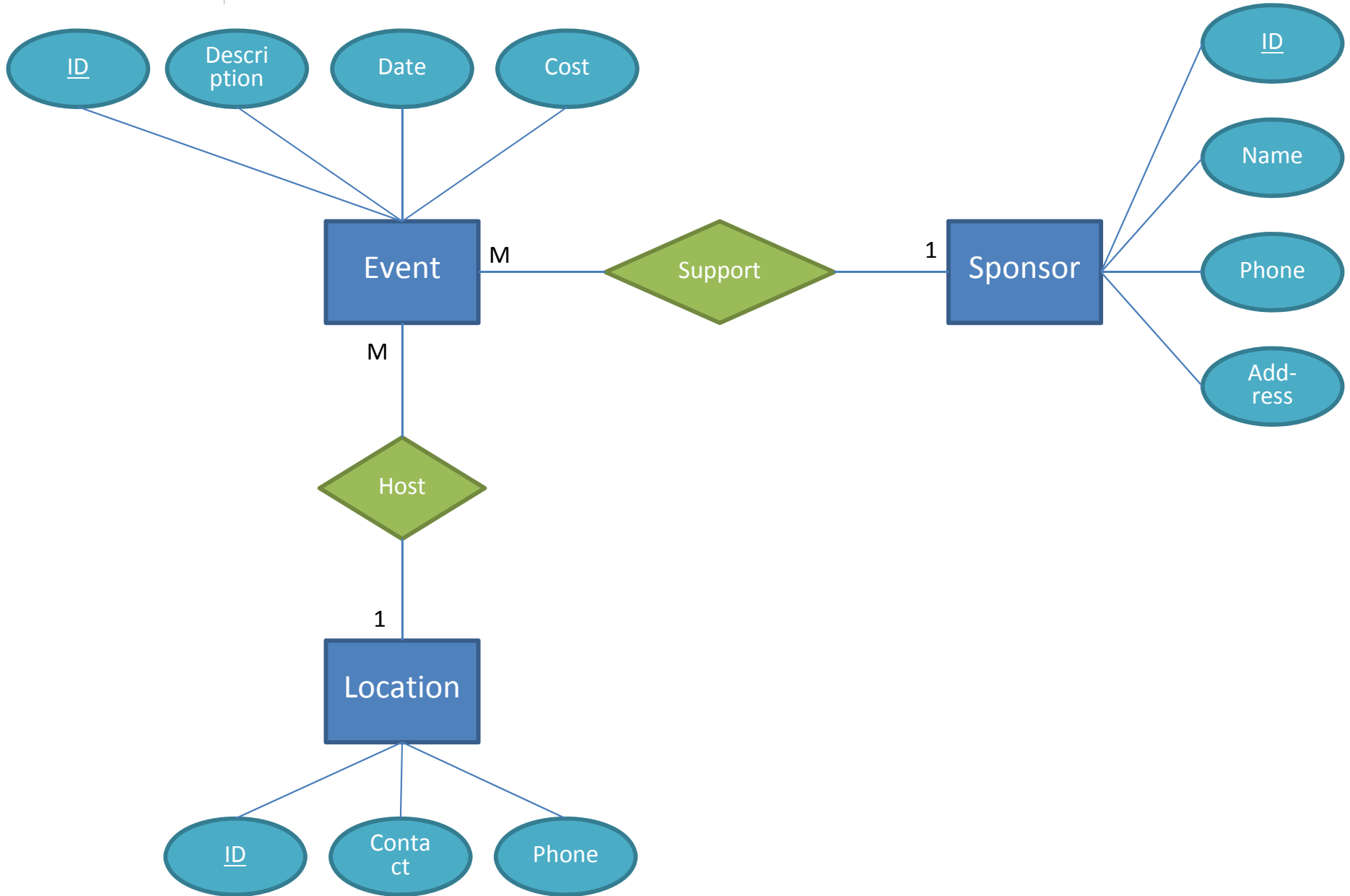
Relationship Depends on Business Rules

● There is no right answer in relationship types. It depends on business descriptions and rules.

● If a customer can place only one order, it's 1:1 relationship.

● If an order can have multiple customers, it's M:N.

Database Normalization (1/2)

EmpNo	Ename	DeptNo	DeptName	Date	Expenses
101	Abigail	10	Marketing	Jan 2014	$1,000
102	Bob	20	Purchasing	Jan 2014	$500
103	Carolyn	10	Marketing	Jan 2014	$1,500
101	Abigail	10	Marketing	Feb 2014	$250
102	Bob	20	Purchasing	Feb 2014	$1,000
104	Kevin	30	R&D	Feb 2014	$900
101	Abigail	10	Marketing	Mar 2014	$400
102	Bob	20	Purchasing	Mar 2014	$1,750
103	Carolyn	10	Marketing	Mar 2014	$2,000

Database Normalization (1/2)

EmpNo	Ename	DeptNo	DeptName	Date	Expenses
101	Abigail	10	Marketing	Jan 2014	$1,000
102	Bob	20	Purchasing	Jan 2014	$500
103	Carolyn	10	Marketing	Jan 2014	$1,500
101	Abigail	10	Marketing	Feb 2014	$250
102	Bob	20	Purchasing	Feb 2014	$1,000
104	Kevin	30	R&D	Feb 2014	$900
101	Abigail	10	Marketing	Mar 2014	$400
102	Bob	20	Purchasing	Mar 2014	$1,750
103	Carolyn	10	Marketing	Mar 2014	$2,000

● There are too many redundant data (Department name, employee names). What if there are millions of records?

● Normalization : a process to remove redundant data

Database Normalization (2/2)

DeptNo	DeptName
10	Marketing
20	Purchasing
30	R&D

EmpNo	Ename	DeptNo
101	Abigail	10
102	Bob	20
103	Carolyn	10
104	Kevin	30

EmpNo	Date	Expenses
101	Jan 2014	$1,000
102	Jan 2014	$500
103	Jan 2014	$1,500
101	Feb 2014	$250
102	Feb 2014	$1,000
104	Feb 2014	$900
101	Mar 2014	$400
102	Mar 2014	$1,750
103	Mar 2014	$2,000