

COP 4710: Database Systems Spring 2004

-Day 4 – January 14, 2004 –
Introduction to Data Modeling

Instructor : Mark Llewellyn
markl@cs.ucf.edu
CC1 211, 823-2790
<http://www.cs.ucf.edu/courses/cop4710/spr2004>

School of Electrical Engineering and Computer Science
University of Central Florida



Introduction to Data Modeling

- Semantic data models attempt to capture the “meaning” of a database. Practically, they provide an approach for conceptual data modeling.
- Over the years there have been several different semantic data models that have been proposed.
- By far the most common is the *entity-relationship data model*, most often referred to as simply the *E-R data model*.
- The E-R model is often used as a form of communication between database designers and the end users during the developmental stages of a database.



Introduction to Data Modeling (cont.)

- The E-R model contains an extensive set of modeling tools, some of which we will not be concerned with as our primary objective is to give you some insight into conceptual database design and not learning all of the ins and outs of the E-R model.
- Another conceptual modeling which is becoming more common is the *Object Definition Language* (ODL) which is an object-oriented approach to database design that is emerging as a standard for object-oriented database systems.



Introduction to Data Modeling (cont.)

- The database design process can be divided into six basic steps. Semantic data models are most relevant to only the first three of these steps.
1. *Requirements Analysis*: The first step in designing a database application is to understand what data is to be stored in the database, what applications must be built on top of it, and what operations are most frequent and subject to performance requirements. Often this is an informal process involving discussions with user groups and studying the current environment. Examining existing applications expected to be replaced or complemented by the database system.



Introduction to Data Modeling (cont.)

2. *Conceptual Database Design:* The information gathered in the requirements analysis step is used to develop a high-level description of the data to be stored in the database, along with the constraints that are known to hold on this data.
3. *Logical Database Design:* A DBMS must be selected to implement the database and to convert the conceptual database design into a database schema within the data model of the chosen DBMS.



Introduction to Data Modeling (cont.)

4. *Schema Refinement*: In this step the schemas developed in step 3 above are analyzed for potential problems. It is in this step that the database is *normalized*. Normalization of a database is based upon some elegant and powerful mathematical theory. We will discuss normalization later in the term.
5. *Physical Database Design*: At this stage in the design of a database, potential workloads and access patterns are simulated to identify potential weaknesses in the conceptual database. This will often cause the creation of additional indices and/or clustering relations. In critical situations, the entire conceptual model will need restructuring.

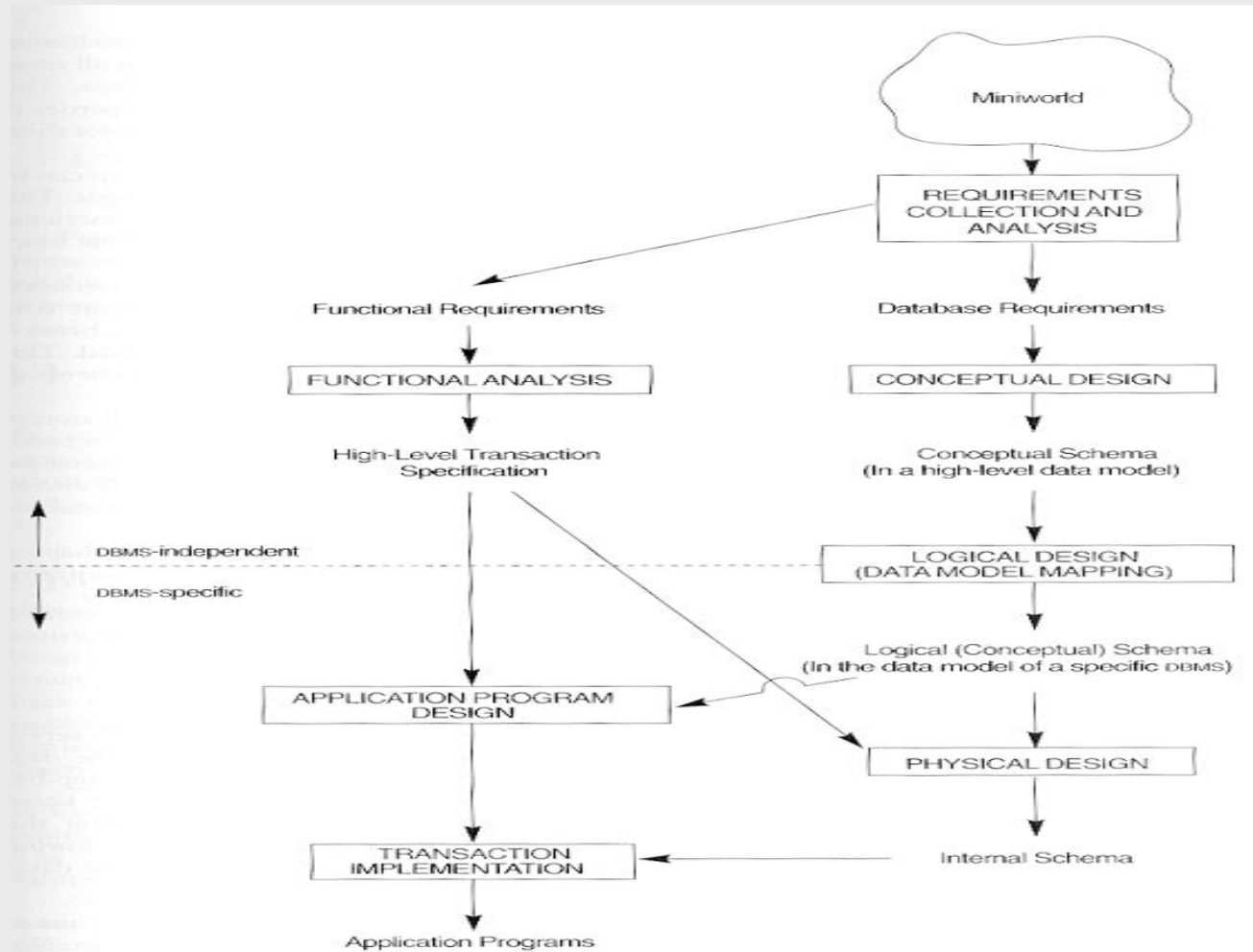


Introduction to Data Modeling (cont.)

6. *Security Design*: Different user groups are identified and their different roles are analyzed so that access patterns to the data can be defined.
- There is often a seventh step in this process with the last step being a *tuning phase*, during which the database is made operational (although it may be through a simulation) and further refinements are made as the system is “tweaked” to provide the expected environment.
- The illustration on the following page summarizes the main phases of database design.



Introduction to Data Modeling (cont.)



The Entity-Relationship Model

- The E-R model employs three basic notions: *entity sets*, *relationship sets*, and *attributes*.
- An *entity* is a “thing” or “object” in the real world that is distinguishable from all other objects. An entity may be either concrete, such as a person or a book, or it may be abstract, such as a bank loan, or a holiday, or a concept.
- An entity is represented by a set of *attributes*. Attributes are descriptive properties or characteristics possessed by an entity.
- An *entity set* is a set of entities of the same type that share the same attributes. For example, the set of all persons who are customers at a particular bank can be defined as the entity set *customers*.

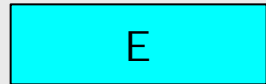


The Entity-Relationship Model (cont.)

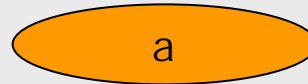
- Entity sets do not need to be disjoint. For example, we could define the entity set of all persons who work for a bank (*employee*) and the entity set of all persons who are customers of the bank (*customers*). A given person entity might be an *employee*, a *customer*, both, or neither.
- For each attribute, there is a permitted set of values, called the *domain* (sometimes called the *value set*), of that attribute. More formally, an attribute of an entity set is a function that maps from the entity set into a domain. Since an entity set may have several attributes, each entity in the set can be described by a set of <attribute, data-value> pairs, one for each attribute of the entity set.
- A database contains a collection of entity sets.



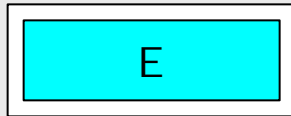
E-R Model Notation



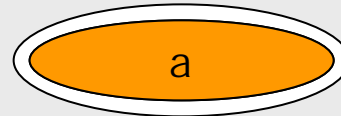
entity set



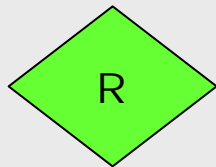
attribute



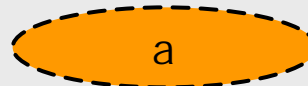
weak entity set



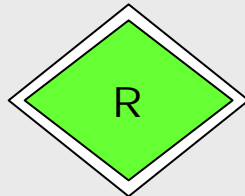
multi-valued attribute



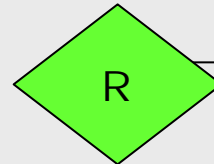
relationship



derived attribute



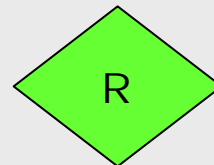
identifying relationship
for a weak entity set



total participation of
entity set in relationship



primary key



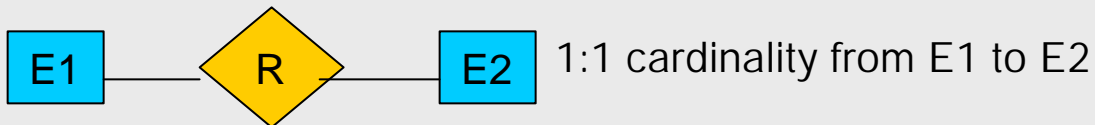
partial participation of
entity set in relationship



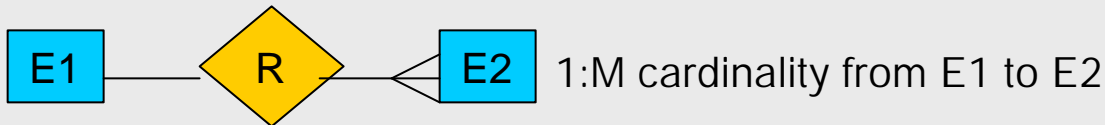
E-R Model Notation (cont.)



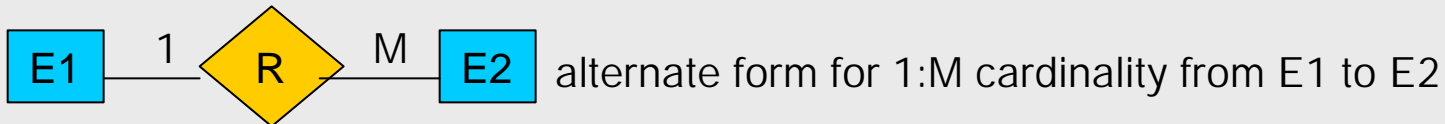
discriminating attribute of
a weak entity set



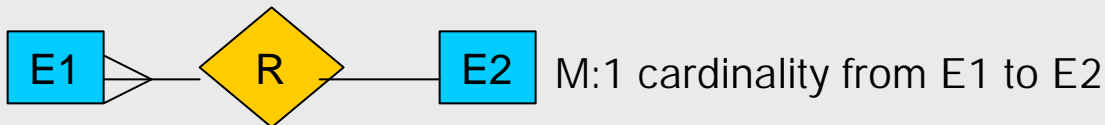
1:1 cardinality from E1 to E2



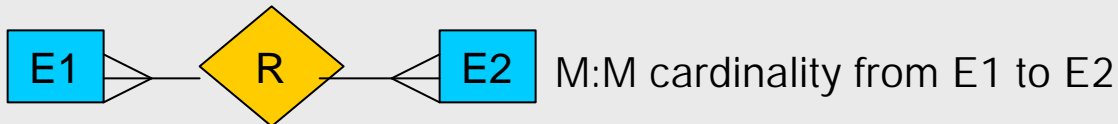
1:M cardinality from E1 to E2



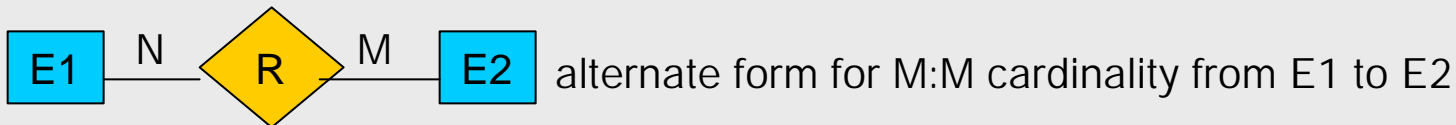
alternate form for 1:M cardinality from E1 to E2



M:1 cardinality from E1 to E2



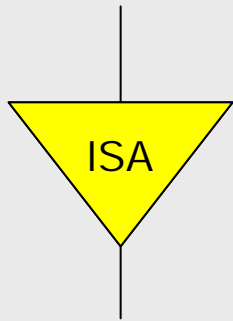
M:M cardinality from E1 to E2



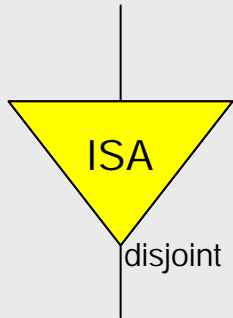
alternate form for M:M cardinality from E1 to E2



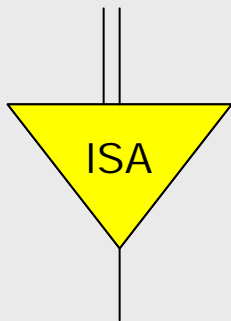
E-R Model Notation (cont.)



ISA (specialization or generalization)(partial participation)



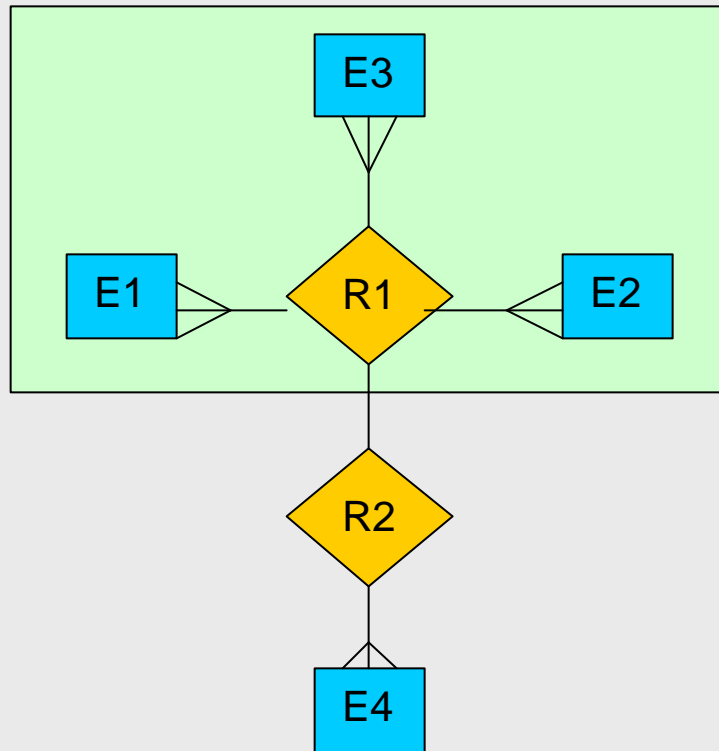
Disjoint ISA (specialization or generalization)



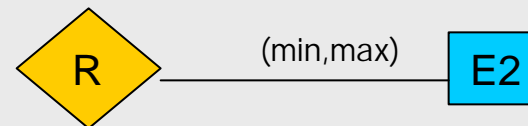
Total generalization



E-R Model Notation (cont.)



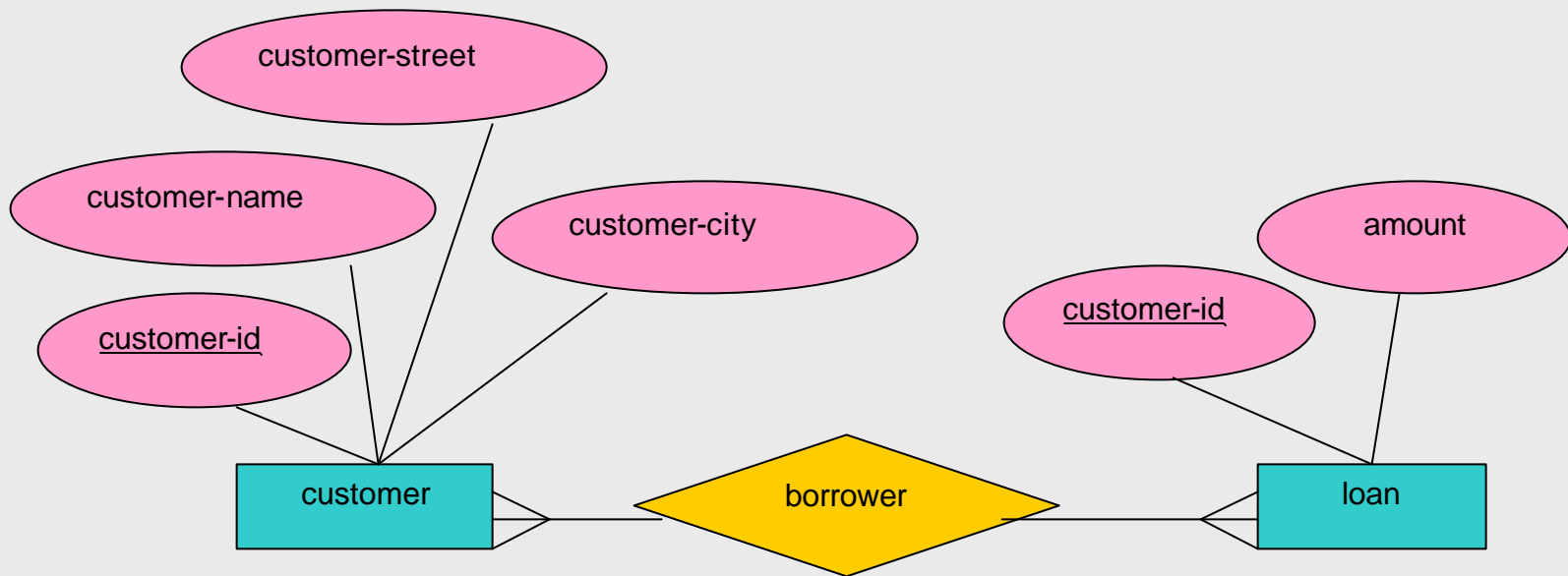
Aggregation: box drawn around relationship which is treated as an entity



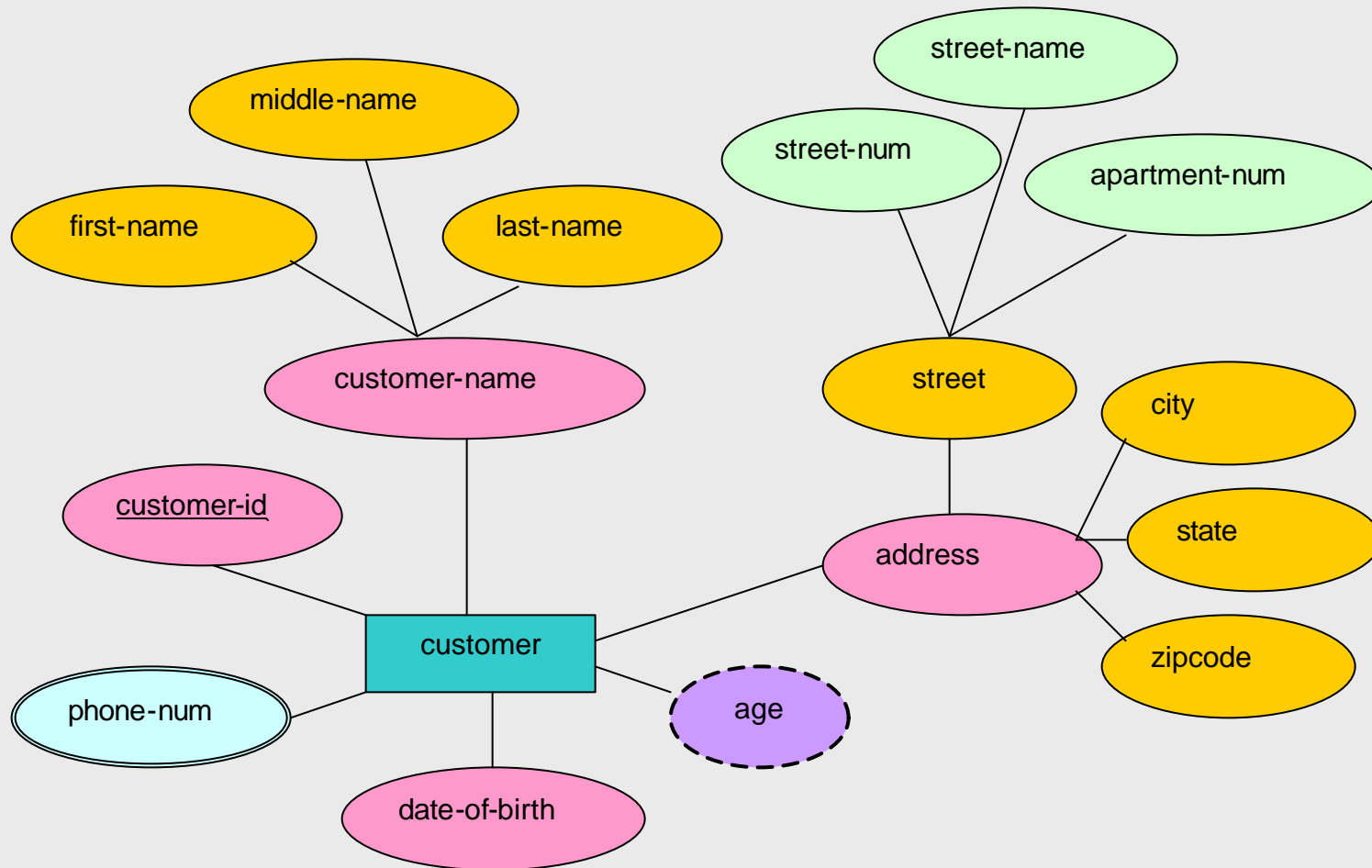
Structural constraint: (min,max) on the participation of an entity in a relationship



Example E-R Diagram (ERD)



Another Example ERD



Attributes in the E-R Model

- As used in the E-R model, an attribute can be characterized by the following attribute types:
- *Simple* or *Composite*: A simple attribute contains no subparts while a composite attribute will contain subparts. For example, consider the attribute *name*. If *name* represents a simple attribute then we must treat the first name, middle name, and last name as an atomic, indivisible attribute. On the other hand, if *name* represents a composite attribute then we have the option of dealing with the entire name as a whole or dealing only with one of the subparts. For example, we could look only at last names, something that we could not do with a simple attribute.



Attributes in the E-R Model (cont.)

- *Single-valued* or *Multi-valued*: A single-valued attribute may have at most one value at any particular time instance. A multiple-valued attribute may have several different values at any particular time instance.
 - For example, consider a particular course at UCF. At any given moment the number of students enrolled in that course is a single value, say 100, but not 100, 80, and 45! On the other hand, some attributes may contain different values at the same time instant. For example, consider an attribute of the entity set student which might be phone-number. At any given time instant a student may have several different phone numbers and thus a multi-valued attribute would be best to accurately model the student. It is also common to place lower and upper bounds on the number of different values that a multi-valued attribute may have at any given time.



Attributes in the E-R Model (cont.)

- *Derived*: This is an attribute whose value is derived (computed) from the values of other related attributes or entities.
 - For example, suppose that the bank customer entity set contains an attribute loans-held, which represents the number of loans a customer has from the bank. The value of this attribute can be computed for each customer by counting the number of loan entities associated with that customer.



Attributes in the E-R Model (cont.)

- *Null*: An attribute takes a null value when an entity does not have a value for it. Null values are usually special cases that can be handled in a number of different ways depending on the situation.
 - For example, it could be interpreted to mean that the attribute is “not applicable” to this entity, or it could mean that the entity has a value for this attribute but we don’t know what it is. We will see later in the term how different systems handle null values and the different interpretations that may be associated with this special value.



Relationships in the E-R Model

- A *relationship* is an association among several entities.
 - For example, we can define a relationship that associates you as a student in COP 4710. This relationship might specify that you are *enrolled* in this course.

A *relationship set* is a set of relationships of the same type.

More formally, it is a mathematical relation on $n \geq 2$ (possibly non distinct) entity sets.

If E_1, E_2, \dots, E_n are entity sets, then a relationship set R is a subset of:

$$\{(e_1, e_2, \dots, e_n) \mid e_1 \in E_1, e_2 \in E_2, \dots, e_n \in E_n\}$$

where (e_1, e_2, \dots, e_n) is the relationship.



Relationships in the E-R Model (cont.)

- The association between entity sets is referred to as **participation**; that is, the entity sets E_1, E_2, \dots, E_n participate in relationship R .
- A **relationship instance** in an E-R schema represents an association between named entities in the real world enterprise which is being modeled.
- A relationship may also have attributes which are called **descriptive attributes**. For example, considering the bank scenario again, suppose that we have a relationship set *depositor* with entity sets *customer* and *account*. We might want to associate with the *depositor* relationship set a descriptive attribute called *access-date* to indicate the most recent date that a customer accessed their account.



Constraints in the E-R Model

- As we have mentioned earlier, the values contained within a given database often have constraints placed upon them to ensure that they accurately model the real world enterprise captured in the database.
- The E-R model has the capability of modeling certain types of these constraints.
- We will focus on two types of constraints: **mapping cardinalities** and **participation constraints**, which are two of the more important types of constraints.

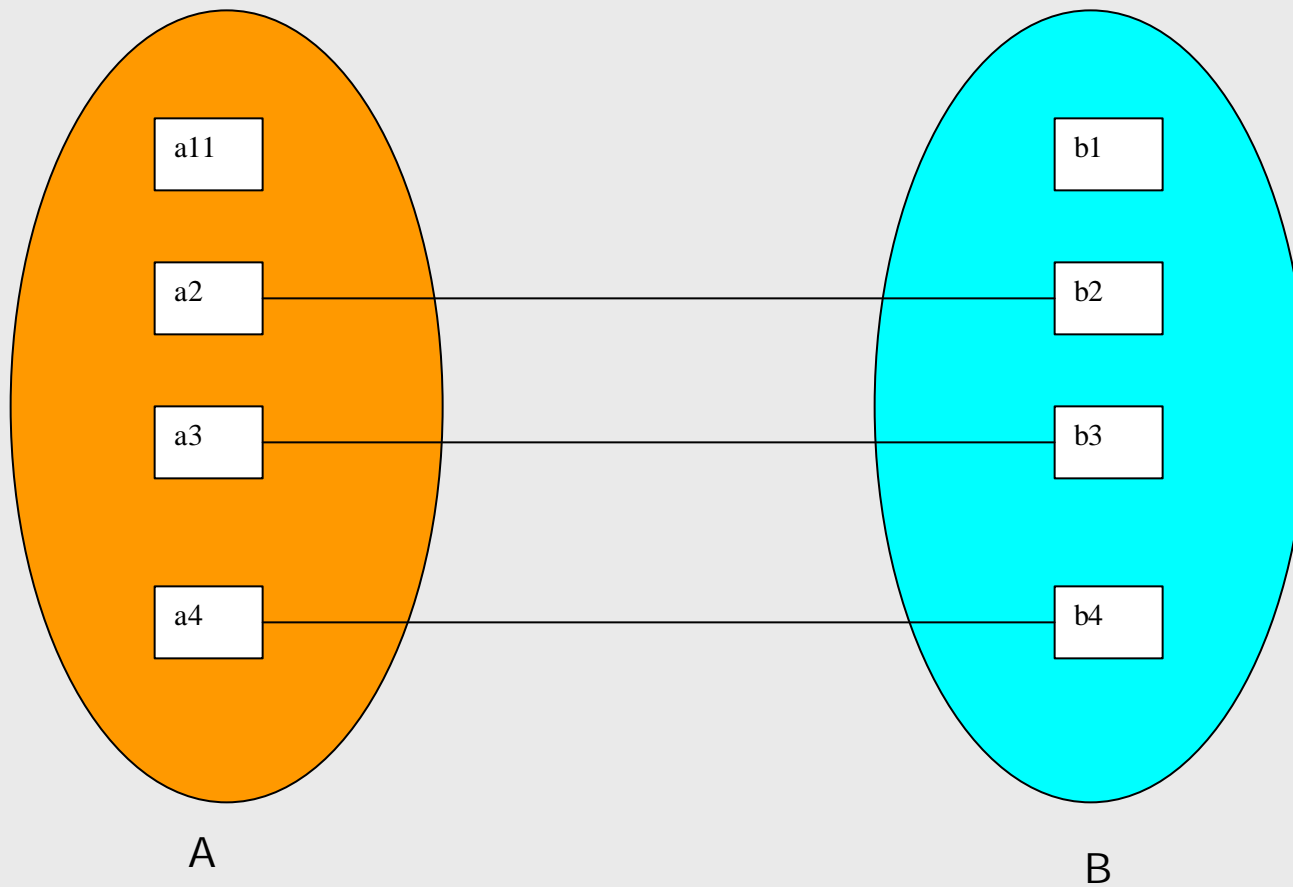


Constraints in the E-R Model (cont.)

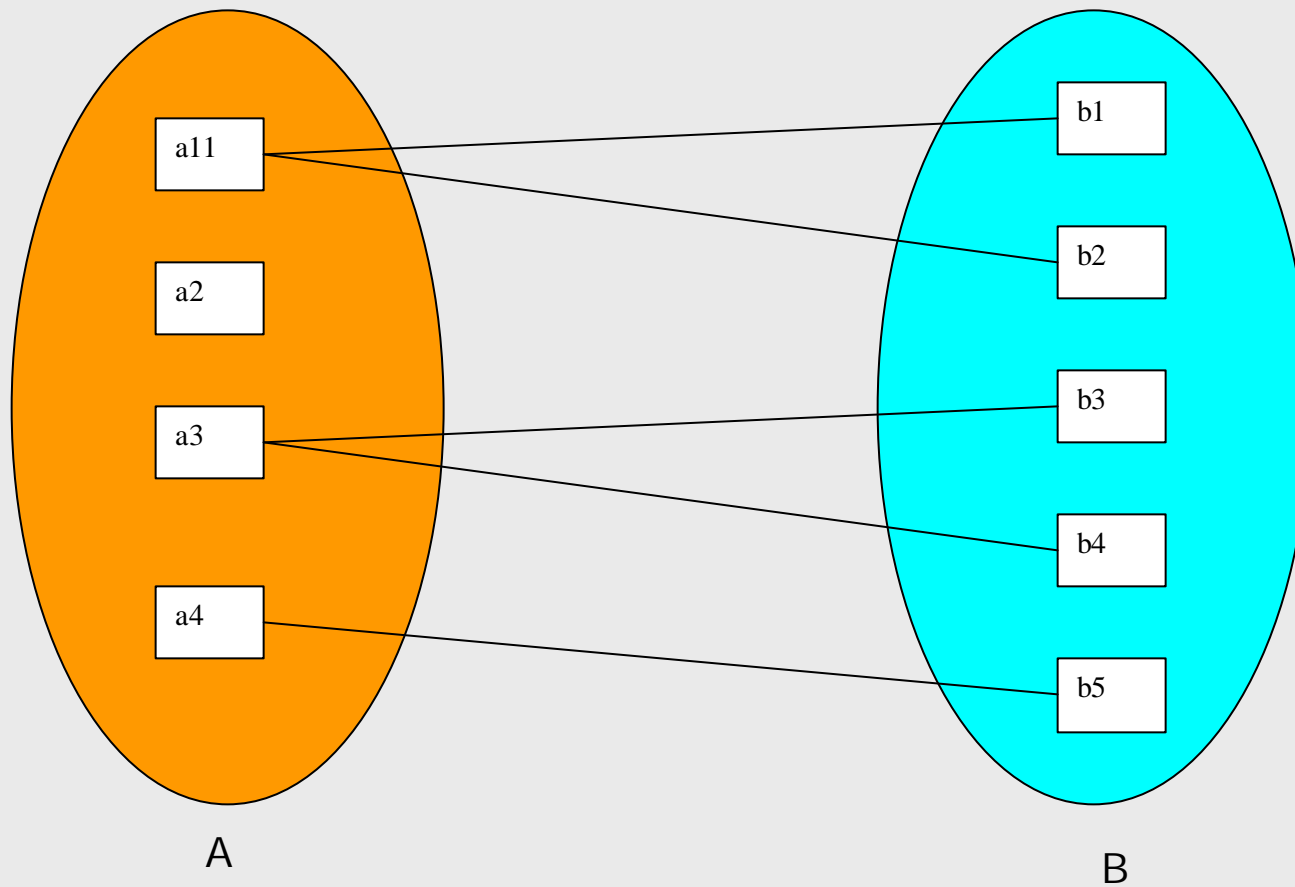
- **Mapping cardinalities** (also called *cardinality ratios*), express the number of entities to which another entity can be associated via a relationship set.
- Mapping cardinalities are most useful in describing binary relationships, although they can be helpful in describing relationship sets that involve more than two entity sets. We will focus only on binary relationships for now.
- For a binary relationship set R between entity sets A and B , the mapping cardinality must be one of the following:
 - (1:1) *one to one from A to B*
 - (1:M) *one to many from A to B*
 - (M:1) *many to 1 from A to B*
 - (M:M) *many to many from A to B*



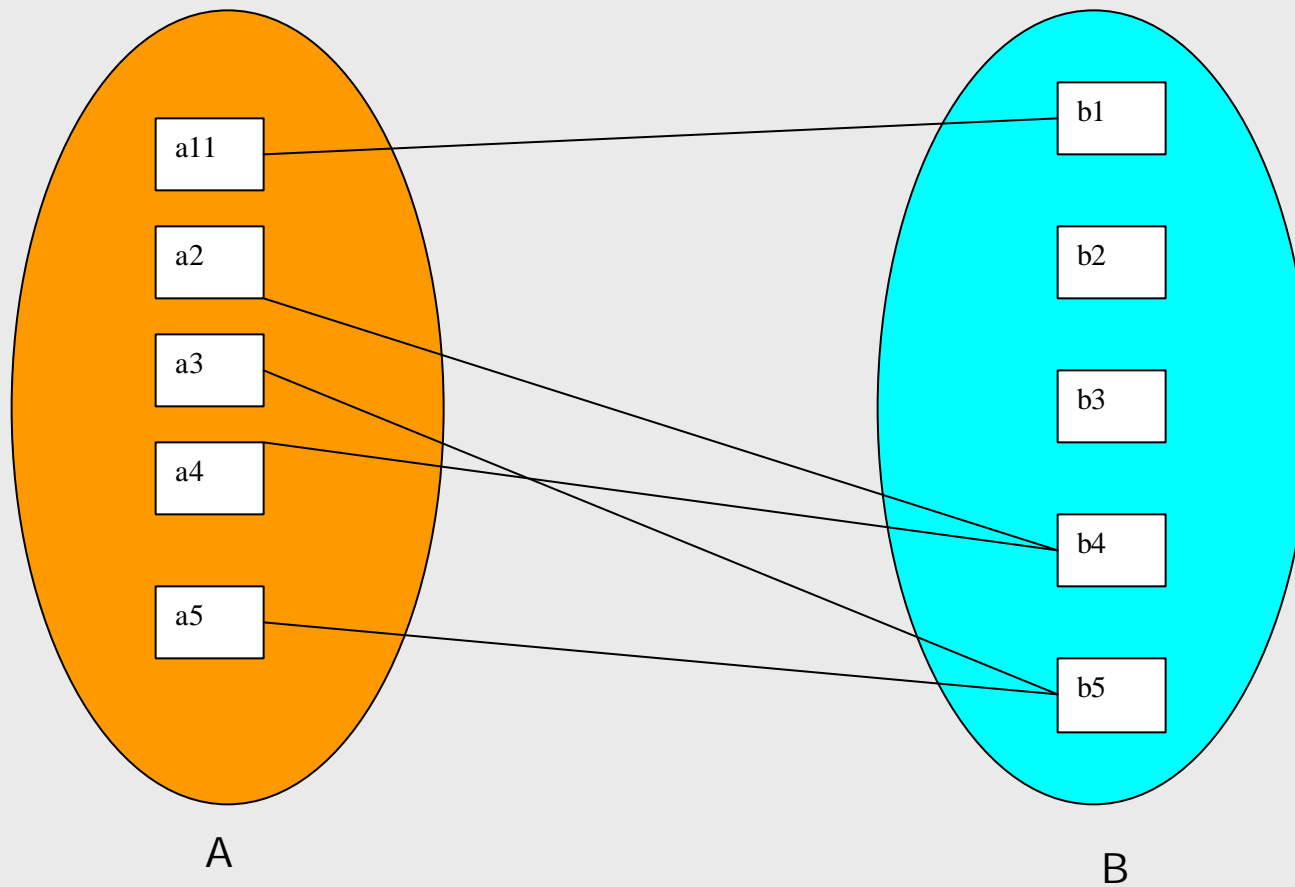
Mapping Cardinality: 1:1 from A to B



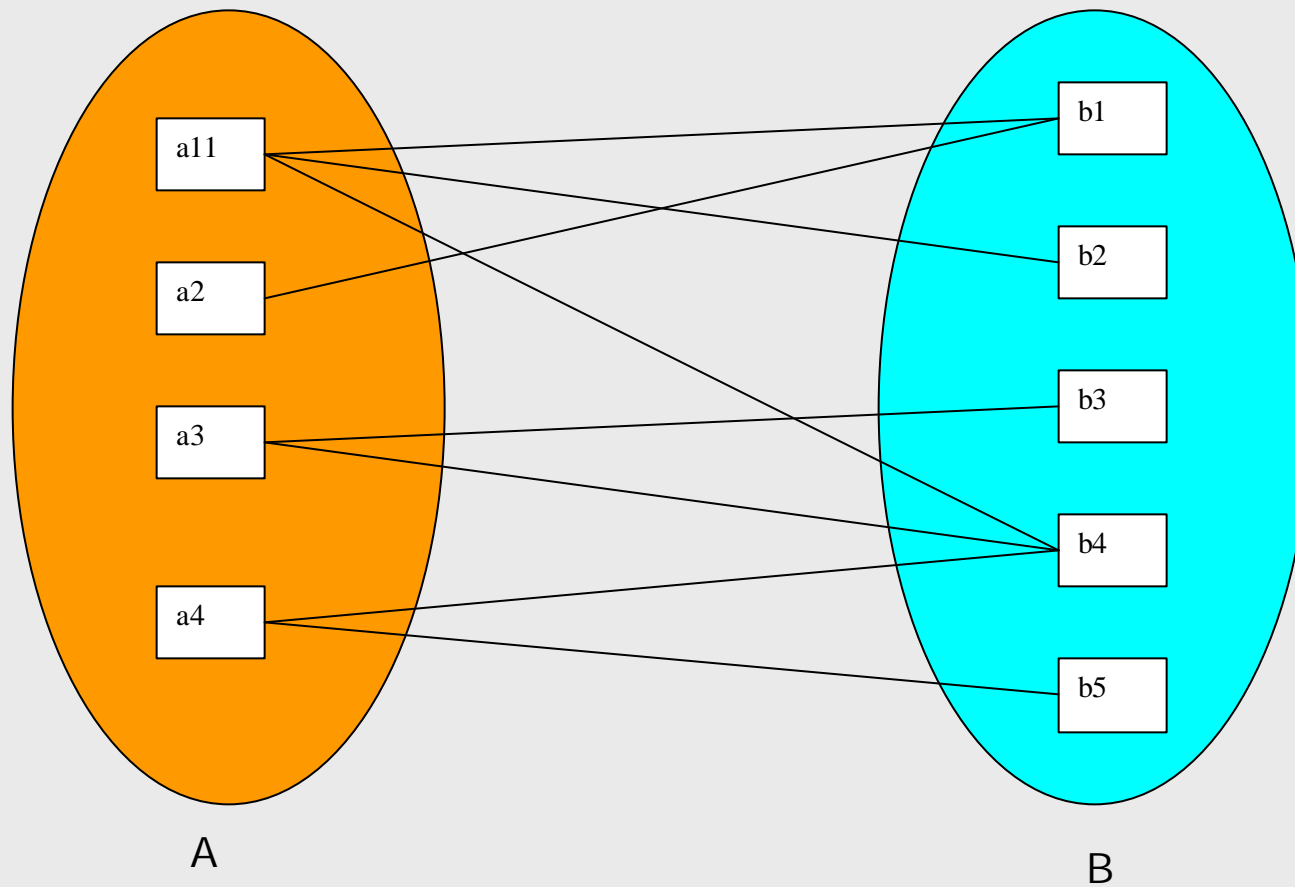
Mapping Cardinality: 1:M from A to B



Mapping Cardinality: M:1 from A to B



Mapping Cardinality: M:M from A to B



Participation Constraints in the E-R Model

- The participation of an entity set E in a relationship set R is said to be *total* if every entity in E participates in at least one relationship in R . If only some of the entities in E participate in a relationship in R , the participation of entity set E in relationship R is said to be *partial*.
- As examples, consider the banking example again. We would expect that every loan entity be related to at least one customer through a *borrower* relationship. Therefore the participation of *loan* in the relationship set *borrower* is total. In contrast, an individual can be a bank customer whether or not they have a loan with the bank. Thus, it is possible that only some of the customer entities will be related to a loan entity through the borrowers relationship. Therefore, the participation of the customer entity set in the borrower relationship is partial.



Keys of an Entity Set

- We must have some mechanism for specifying how entities within a given entity set are distinguished.
- Conceptually, individual entities are distinct; from a database perspective, however, the differences among them must be expressed in terms of their attributes. Therefore, the values of the attribute values of an entity must be such that they can *uniquely identify* the entity. In other words, no two entities in an entity set are allowed to have exactly the same value for all attributes.
- A *key* allows us to identify a set of attributes that suffice to distinguish entities from each other. Keys also help uniquely identify relationships, and thus distinguish relationships from one another.



Primary Keys, SuperKeys and Candidate Keys

- A *superkey* is a set of one or more attributes that, taken collectively, allow us to identify uniquely an entity in the entity set. Suppose that we have an entity set modeling the students in COP 4710. Suppose that we have the following schema for this entity set:

Students(SS#, name, address, age, major, minor, gpa, spring-sch)

- Among the attributes which we have associated with each student must be a set of attributes which will uniquely distinguish each student. Suppose that we define this set of attributes to be:

(SS#, name, major, minor)



Primary Keys, SuperKeys and Candidate Keys (cont.)

- This set of attributes (SS#, name, major, minor) defines a superkey for the entity set Students. Notice that the set of attributes (SS#, name) also defines a superkey for this entity set, because given this second set of attributes we can still uniquely distinguish each student in the set. The concept of a superkey is not a sufficient definition of a key because the superkey, as we can see from this example, may contain extraneous attributes.



Primary Keys, SuperKeys and Candidate Keys (cont.)

- If the set K is a superkey of entity set E , then so too is any superset of K . We are interested only in superkeys for which no proper subset of K is a superkey. Such a minimal superkey is called a *candidate key*.
- For a given entity set E it is possible that there may be several distinct sets of attributes which are candidate keys.
- Either there is only a single such set of attributes or there are several distinct sets from which only one is selected by the database designer and this set of attributes defines the *primary key* which is typically referred to simply as the *key* of the entity set.



Primary Keys, SuperKeys and Candidate Keys (cont.)

- A **key** (primary, candidate, and super) is a property of the entity set, rather than of the individual entities. Any two individual entities in the set are prohibited from having the same value on all attributes which comprise the key attributes at the same time. This constraint on the allowed values of an entity within the set is a *key constraint*.
- The database designer must use care in the selection of the set of attributes which comprise the key of an entity set to:
(1) be certain that the set of attributes guarantees the uniqueness property, and (2) be certain that the set of key attributes are never, or very rarely, changed.



Relationship Sets

- The primary key of an entity set allows us to distinguish among the various entities in the set. There must be a similar mechanism which allows us to distinguish among the various relationships in a relationship set.
- Let R be a relationship set involving entity sets E_1, E_2, \dots, E_n . Let K_i denote the set of attributes which comprise the primary key of entity set E_i . For now let's assume that
 - (1) all attributes names in all primary keys are unique, it will make the notation easier to understand and it really isn't a problem if the names aren't unique anyway, and
 - (2) each entity set participates only once in the relationship.
- Then the composition of the primary key for the relationship set depends on the set of attributes associated with the relationship set R in the following ways:



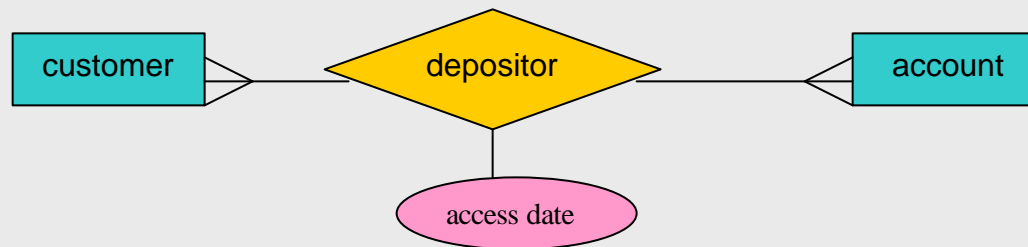
Relationship Sets (cont.)

- (a) If the relationship set R has no attributes associated with it, then the set of attributes: $K_1 \cup K_2 \cup \dots \cup K_n$ describes an individual relationship in set R .
- (b) If the relationship set R has attributes a_1, a_2, \dots, a_m associated with it, then the set of attributes: $K_1 \cup K_2 \cup \dots \cup K_n \cup \{ a_1, a_2, \dots, a_m \}$ describes an individual relationship in set R .
- In **both** of these cases, the set of attributes: $K_1 \cup K_2 \cup \dots \cup K_n$ forms a superkey for the relationship set.



Effect of Cardinality Constraints on Keys

- The structure of the primary key for the relationship set depends upon the mapping cardinality of the relationship set. Consider the following case:



- This E-R diagram represents a many to many cardinality for the relationship *depositor* with an attribute of *access date* associated with the relationship set with two entities *customer* and *account* participating in the relationship. The primary key of the relationship *depositor* will consist of the union of the primary keys of *customer* and *account*.



Effect of Cardinality Constraints on Keys

- To further clarify this situation consider for a moment the schemas of these two entity sets:

Customer (customer-id, customer-name, address, city)

Account (account-number, balance)

- A many-to-many relationship between these two sets means that it is possible for one customer to have several accounts and similarly for a given account to be held by several customers.
- To uniquely identify a relationship between two entities in *customers* and *accounts* will require the union of the primary keys in both entity sets.



Effect of Cardinality Constraints on Keys

(cont.)

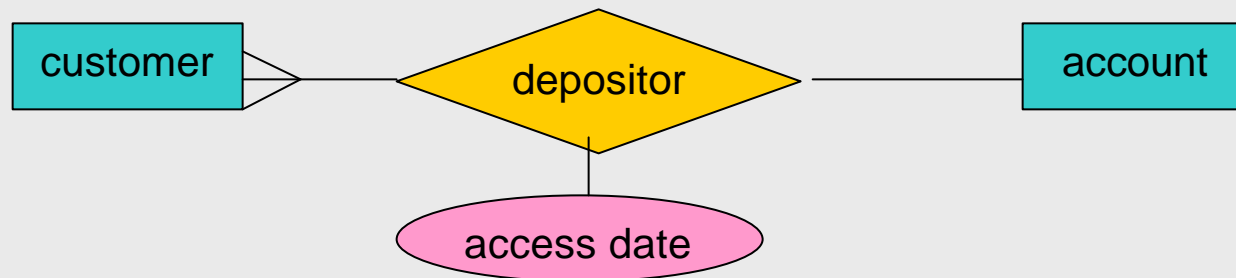
- In order to “see” the last deposit made to specific account number requires that we specify by whom the deposit was made since several account holders may have made deposits to the same account.
- The schema for the *depositor* relationship is then:

Depositor (customer-id, account-number, access-date)



Effect of Cardinality Constraints on Keys (cont.)

- Now consider the case when a customer is only allowed to have one account. This means that the *depositor* relationship is many-to-one from *customer* to *account* as shown in the following diagram.



- In this case the primary key of the *depositor* relationship is simply the primary key of the *customer* entity set. To clarify this, again look at the schemas of the entity sets:

Customer (customer-id, customer-name, address, city)

Account (account-number, balance)



Effect of Cardinality Constraints on Keys (cont.)

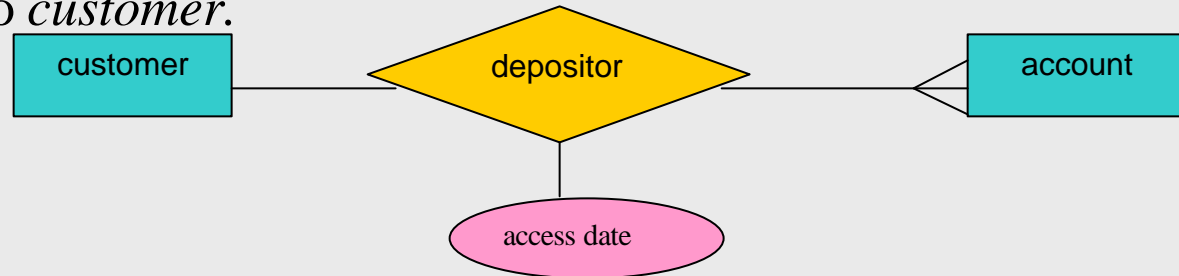
- A many-to-one relationship means that a given customer can have only a single account then the primary key of the *depositor* relationship is simply the primary key of the *customer* set since for a given customer they could only make a single most recent deposit since they only “own” one account, so specifying the account number is not necessary to identify a unique deposit by a given customer.
- The schema for the *depositor* relationship set is then:

Depositor (customer-id, access-date)



Effect of Cardinality Constraints on Keys (cont.)

- Now consider the case when the *depositor* relationship is many-to-one from *account* to *customer*.



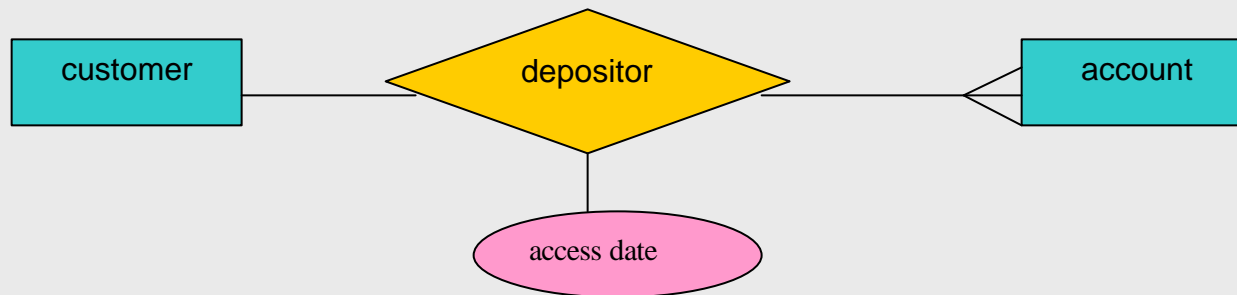
- A many-to-one relationship from *account* to *customer* means that each *account* is owned by at most one *customer* but each *customer* may have more than one *account*. In this situation the primary key of the *depositor* relationship is simply the primary key of the *account* entity set since there can be at most one most recent deposit to a given *account* because at most one *customer* could make the deposit. We do not need to uniquely identify which *customer* made the deposit in question because there could only be one.
- The schema for the *depositor* relationship is then:

Depositor (account-id, access-date)



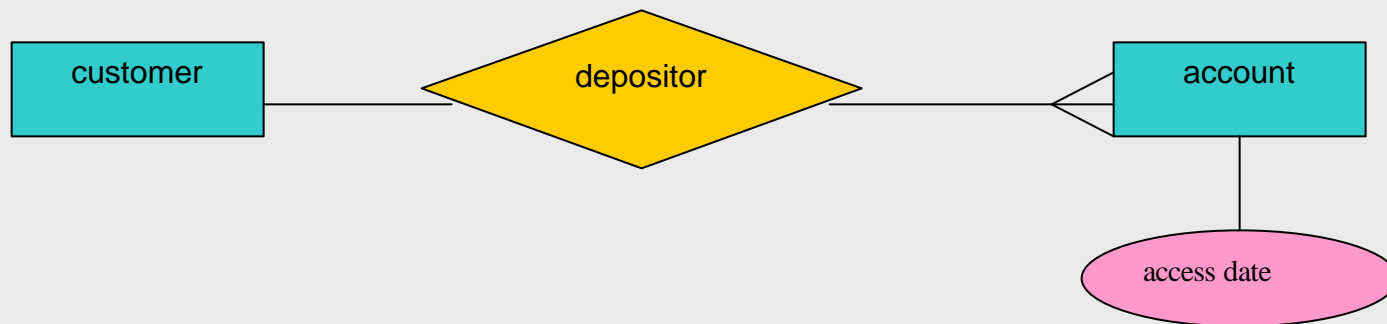
Placement of Relationship Attributes

- Just as the cardinality of a relationship set affects the set of attributes which comprise the primary key of the relationship set, so too does it affect the placement of the attributes.
- The attributes of a one-to-one or one-to-many relationship set can be associated with one of the participating entity sets, rather than with the relationship set itself. For example consider the following case:



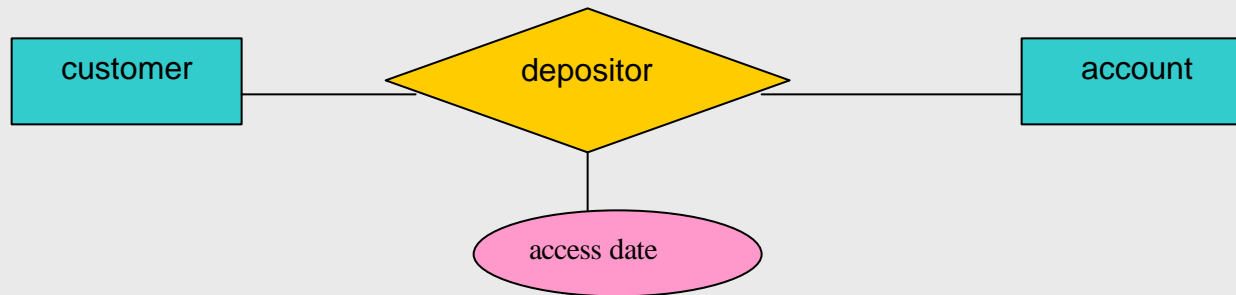
Placement of Relationship Attributes (cont.)

- The attribute *access-date* could be associated with the *account* set without loss of information. Since a given account can be owned by at most one customer it could have at most one access-date which could be stored in the *account*



Placement of Relationship Attributes (cont.)

Now consider the following case:

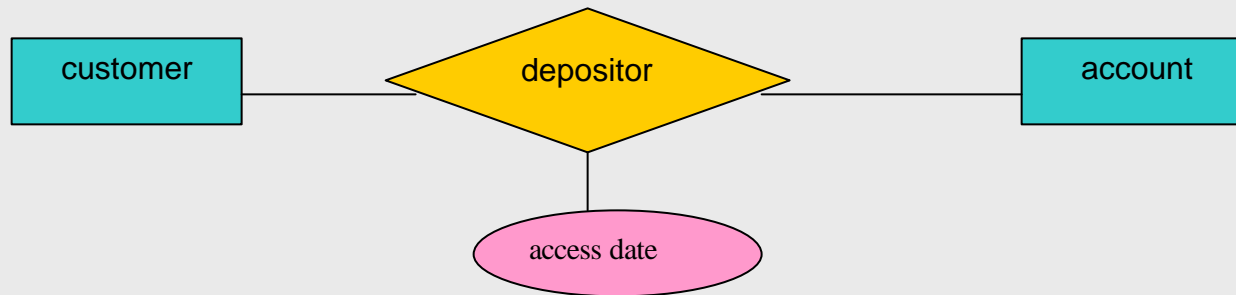


- The attribute *access-date* could be associated with either the *customer* set or the *account* set without loss of information. In this case a given account can be owned by at most one customer and a given customer can own at most one account. Therefore, if the *access-date* attribute is stored with the *customer* set then it must refer to the last access by this customer on the only account they can have. Similarly, if the *access-date* attribute is stored with the *account* set, then it must refer to the last access on this account by the only customer who owns this account.



Placement of Relationship Attributes (cont.)

Now consider the following case:

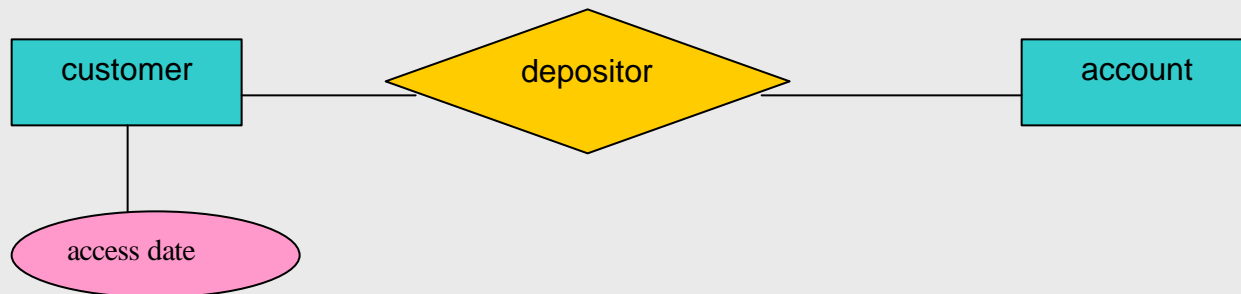
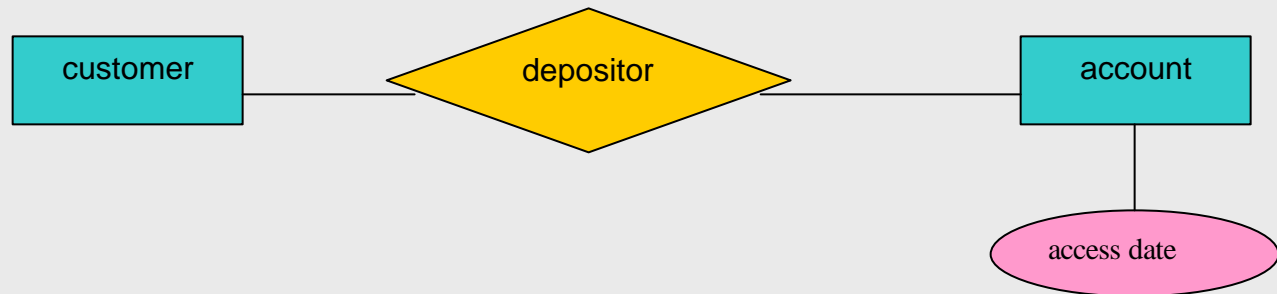


- The attribute *access-date* could be associated with either the *customer* set or the *account* set without loss of information. In this case a given account can be owned by at most one customer and a given customer can own at most one account. Therefore, if the *access-date* attribute is stored with the *customer* set then it must refer to the last access by this customer on the only account they can have. Similarly, if the *access-date* attribute is stored with the *account* set, then it must refer to the last access on this account by the only customer who owns this account.



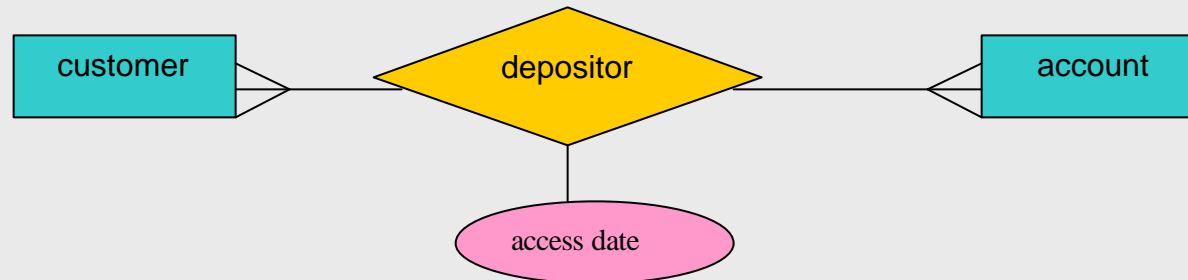
Placement of Relationship Attributes (cont.)

- Therefore, either diagram below would be a correct representation of this situation:



Placement of Relationship Attributes (cont.)

- When the relationship set has a cardinality constraint of many-to-many, the situation is much clearer. Consider the following situation:



- account may be owned by several customers, we see that associating the *access-date* attribute with either entity set will not properly model this situation without the loss of information. If we need to model the date that a specific customer last accessed a specific account the *access-date* attribute must be an attributed of the *depositor* relationship set, rather than one of the participating entities. For example, if *access-date* were an attribute of *account* we could not determine which customer made the last access to the account. If *access-date* were an attribute of *customer* we could not determine which account the customer last accessed.



Further Design Issues

- The notions of an entity set and a relationship set are not precise.
- It is possible to define a set of entities and the relationships among them in a number of different ways. We'll look briefly at some of these different approaches to the modeling of the data.
- To some extent this is where the “art” of database design becomes tricky. Sometimes several different design scenarios may all look equally plausible and even after refinement may still be suitable, sometimes not. Only a careful design will eliminate some of the problems we've discussed earlier.



Entity Sets vs. Attributes

- Consider the entity set: *Employee(emp-name, telephone-number, age)*
- It could easily be argued that a telephone is an entity in its own right with attributes of say, *telephone-number, location, manufacturer, serial-num*, and so on. If we take this point of view, then:

1. The *Employee* entity set must be redefined as:

*Employee (**emp-name**, age)*

2. Must create a new entity set:

*Telephone(**telephone-number**, location, manufacturer, serial-num,...)*

3. A relationship set must be created to denote the association between employees and the telephones that they have.

*Emp-Phone(**emp-name**,* *telephone-number**, age, location, manufacturer, serial-num)*



Entity Sets vs. Attributes (cont.)

- Now we must consider what is the main difference between these two definitions of an employee?
- Treating the telephone as an attribute *telephone-number* implies that employees have precisely one telephone number each. (Note that this must be true or otherwise the telephone-number attribute would need to be a part of the key for an employee and it isn't here – not considering multiple-valued attributes).
- Treating a telephone as an entity permits employees to have several phones (including zero) associated with them. However, we could easily make the *telephone-number* attribute be a multi-valued one to allow multiple phones per employee. So clearly, this is **not** the main difference in the two representations.



Entity Sets vs. Attributes (cont.)

- The main difference then is that treating a telephone as an entity better models a situation where one might want to keep additional information about a telephone, as we have indicated with our example above.
- If we used the original approach and wished to make the telephone an attribute of an employee and we wished to maintain this additional information about their phone, then the *Employee* entity set would look like:

Employee(emp-name, telephone-number, age, location, manufacturer,...)

- This is clearly not a good schema, for example, is the *age* attribute associated with the employee or the telephone? In this situation we are attempting to model two different entity sets inside a single entity set.



Entity Sets vs. Attributes (cont.)

- Conversely, it would not be appropriate to treat the attribute *emp-name* as an entity; it is difficult to argue that an employee name is an entity in its own right (in contrast to the telephone). Thus, it is entirely appropriate to have *emp-name* as an attribute of the *Employee* entity set.
- So, what constitutes an attribute and what constitutes an entity?
 - Unfortunately, there are no simple answers. The distinctions depend mainly upon the structure of the real-world scenario which is being modeled, and on the semantics associated with the attribute in question.



Entity Sets vs. Attributes (cont.)

- A common mistake is to use the primary key of an entity set as an attribute of another entity set, instead of using a relationship. For example, given our bank example again, it would not be appropriate to model *customer-id* as an attribute of *loan* even if each loan had only one customer associated to it. The relationship *borrower* is the correct way of representing the relationship between a loan and a customer, since it makes their connection explicit rather than implicit via an attribute.



Entity Sets vs. Relationship Sets

- It is not always clear whether an object is best expressed by an entity set or a relationship set.
- Consider the banking example. We have been modeling a loan as an entity. An alternative is to model a loan as a relationship between customers and say branches of the bank, with *loan-number* and *amount* as descriptive attributes. Each loan is then represented as a relationship between a customer and a branch.



Entity Sets vs. Relationship Sets (cont.)

- If every loan is owned by exactly one customer and is associated with exactly one branch, then it may be satisfactory to model the loan as a relationship.
- However, with this design we cannot represent in a convenient way the situation in which several customers jointly own a single loan.
 - To handle this type of situation, we would need to define a separate relationship for each holder of the joint loan.
 - Then we would replicate all of the values for the descriptive attributes *loan-number* and *amount* in each such relationship. Each such relationship must, of course, have the same value for the descriptive attributes.



Entity Sets vs. Relationship Sets (cont.)

- Two problems arise as a result of the replication:
 1. The data are stored in multiple locations (the very meaning of replication).
 2. Updates potentially leave the data in an inconsistent state, where the values in two different sets differ when they should be identical. We'll look at the complications that this replication causes as well as solution techniques (normalization theory) later in the course. Notice that the problem of replication is absent in our original version because *loan* is represented by an entity set in that case.
- One possible guideline in determining whether to use an entity set or a relationship set is to designate a relationship set to describe an action that occurs between entities. This approach can also be useful in deciding whether certain attributes may be more appropriately expressed as relationships.



Weak Entity Sets vs. Strong Entity Sets

- An entity set may not have sufficient attributes to form a primary key. Such an entity set is termed a *weak entity set*. An entity set that has a primary key is termed a *strong entity set*.
 - As an example, consider an entity set *payment*, which has three attributes: *payment-number*, *payment date*, and *payment amount*. Payment numbers are typically just sequential numbers, starting at 1 and are generated separately for each loan. Thus, although each *payment* entity is distinct, payments for different loans may share the same payment number, thus the set does not have a primary key and is a weak entity set.



Weak Entity Sets vs. Strong Entity Sets (cont.)

- For a weak entity set to be meaningful, it must be associated with another entity set which is called the *identifying* or *owner entity set*. Every weak entity must be associated with such an identifying entity set.
- The weak entity is said to be *existence dependent* on the identifying set. The identifying set is said to *own* the weak entity set that it identifies.
- The relationship associating the weak entity set with the identifying entity set is called the *identifying relationship*. The identifying relationship is many-to-one from the weak entity set to the identifying set and the participation of the weak entity set in the relationship is total.



Weak Entity Sets vs. Strong Entity Sets (cont.)

- Although a weak entity set does not have a primary key, we nevertheless need a means of distinguishing among all those entities in the weak entity set that depend upon one particular strong entity.
- The set of attributes of a weak entity that allows this distinction to be made is called the *discriminator* (sometimes also called the *partial key*). For example, the discriminator of the weak entity set *payment* from above is the attribute *payment-number*, since for each loan, a payment number uniquely identifies one single payment for that loan.



Weak Entity Sets vs. Strong Entity Sets (cont.)

- The primary key of a weak entity set is formed by the primary key of the identifying entity set, plus the weak entity set's discriminator.
 - For the case above, the primary key of the entity set *payment* would be: {*loan-number*, *payment-number*}, where *loan-number* would be the primary key of the identifying entity set *loan* and *payment-number* is the discriminator of the weak entity set *payment*.
- Within the E-R diagram, a weak entity set is represented by a rectangle with double lines and the identifying relationship for a weak entity set is represented by a diamond with double lines.



Weak Entity Sets vs. Strong Entity Sets (cont.)

Example of a weak entity set.

