

Data Modeling and Database Design

Data Modeling

A *data model* is a blueprint for organizing information.

Three kinds of components –

- *structure* – what kinds of things exist and how they relate to each other
- *constraints* – rules limiting which states of the data are considered valid
- *operations* – how you are allowed to interact with and manipulate the data

Not every model will have all three components.

Data Modeling

Data modeling for databases typically proceeds through three levels.

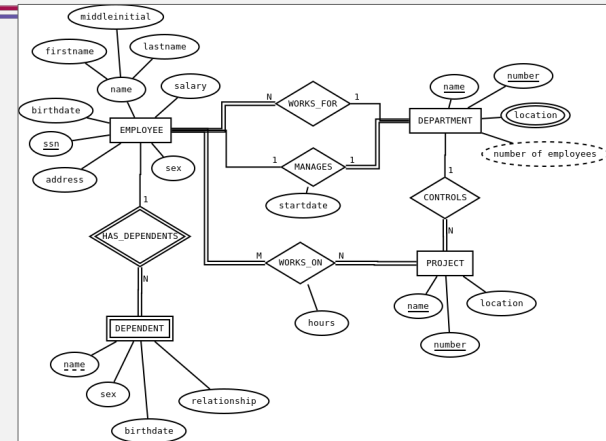
- the *conceptual data model* has high level concepts to identify what the things are and how they relate to each other
 - concepts are close to those in the problem domain
 - implementation independent
 - we will study the entity-relationship (ER) and enhanced entity-relationship (EER) models
- the *logical data model* has concepts representing the structures available in the database
 - structure of the data is translated into something implementable in the database
 - interactions with the database (queries) can be thought of at this level of abstraction
 - we will study the relational model
- the *physical data model* deals with how things will actually be built in the database
 - implementation details
 - we will use MySQL

Why Data Modeling Matters

- a communication tool between everyone involved
 - domain experts, developers, DBAs
- a way to manage complexity
 - reason at the right level of abstraction for the task

Goals of Good Design

- correctness
 - faithfully captures the real-world requirements
- minimal redundancy
 - avoids storing the same information in more than one place
- appropriate use of structures
 - modeling constructs are used for what they are meant to represent
- clarity
 - understandable and maintainable by someone other than its author

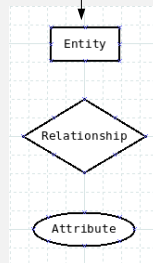


Entity-Relationship Model

Core concepts and notation:

- an *entity* represents some abstract object
 - like a book, borrower, or library branch
- an *entity set* represents a collection of the same kind of entity
- a *relationship* represents a connection between two or more entity sets
- an *attribute* is a property of an entity set or relationship
 - like a title, card number, or branch name

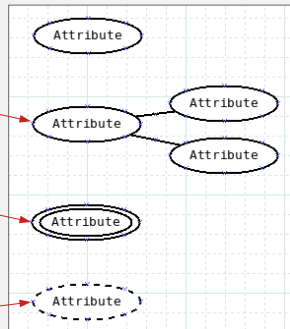
usually referred to as an entity, but really is an entity set



Which of the following are entities?	Which of the following are relationships?	Which of the following are attributes?
☐ EMPLOYEE	☐ ssn	☐ EMPLOYEE
☐ CONTROLS	☐ number of employees	☐ DEPARTMENT
☐ number of employees	☐ birthdate	☐ DEPENDENT
☐ hours	☐ DEPARTMENT	☐ relationship
☐ ssn	☐ EMPLOYEE	☐ name
☐ relationship	☐ DEPENDENT	☐ number of employees
☐ name	☐ hours	☐ location
☐ DEPARTMENT	☐ name	☐ WORKS_FOR
☐ DEPENDENT	☐ WORKS_FOR	☐ birthdate
☐ location	☐ relationship	☐ hours
☐ birthdate	☐ CONTROLS	☐ CONTROLS
☐ WORKS_FOR	☐ location	☐ ssn

Types of Attributes

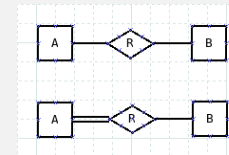
- simple or composite
 - is the attribute atomic, or can it be divided into meaningful subparts?
- single-valued or multivalued
 - can an entity have more than one value for the attribute at a time?
- stored or derived
 - can the attribute's value be determined from other information in the database?



Participation

Participation constraints dictate the minimum number of relationship instances a given entity participates in.

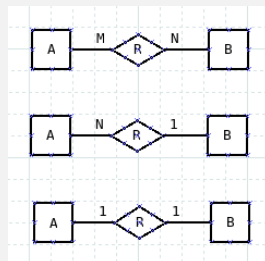
- normally this can be 0 (“may”)
- *total participation* requires ≥ 1 (“must”)



Cardinality

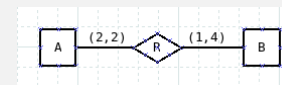
Cardinality constraints dictate the maximum number of relationship instances a given entity participates in.

- M:N (many-many)
 - one entity of A can participate in R with any number of entities of B, vice versa
- N:1 (many-one)
 - one entity of A can participate in R with at most one entity of B
 - one entity of B can participate in R with any number of entities from A
- 1:1 (one-one)
 - one entity of A can participate in R with at most one entity of B, and vice versa



Participation and Cardinality

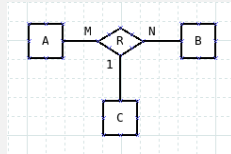
If you need more than may/must and one/many, an alternate notation allows explicit specification of lower and upper bounds.



- an entity of A participates in at least 1 and at most 4 relationship instances with an entity of B
- an entity of B participates in exactly 2 relationship instances with an entity of A

Multiway Relationships

Relationships can involve more than two entity sets.



a particular pair of entities from A and B can participate in R with at most one entity from C

a particular pair of entities from A, C or from B, C can participate in R with any number of entities from the third entity set

Note that not all cardinality constraints involving the various entity sets can be represented with this notation.

- e.g. can't express that one entity from A can participate in at most one relationship instance