

Lesson Plan 2/8

Thursday, February 8, 2018 2:01 PM

- HW Due Today!
 - o Questions, Clarification?
- HDSC: <https://eng.uber.com/dsw/>
- Data Representation Models
 - o Structure vs. Semantics
 - o Model vs. Schema
 - o Data Model:
 - Constructs
 - Constraints
 - Operations
 - o Example Data Models
 - o Data independence vs. Physical data independence
- Entity-Relationship (ER) and Relational Model
 - o ER Model:
 - Entities and their Attributes
 - Relationship between entities
 - Relationships also have attributes
 - o Types of relationships
 - o Relations as representation of E's and R's
- Keys
 - o Super Key
 - o Candidate Key
 - o Primary Key
 - o Foreign Key
 - Constraints
 - o Examples
- Tidy Data
 - o Nycflights13 is a tidy dataset
- SQL: operations over relations
 - o SFW
 - o Group-by summarize
 - o Subqueries



- Queries
- Baseball examples

Model vs. Schema

Entities: Flights

Attributes: Arr Delay: ~~Discrete~~ Discrete numeric

Dep Delay:

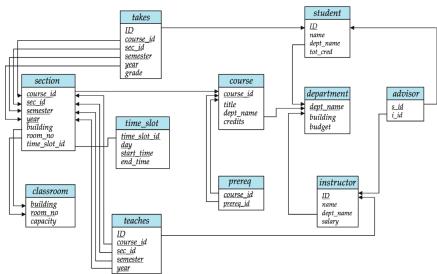
Origin: Categorical Order

Model:

- Constructs
- Constraints
- Operations

→ XML, JSON, Protocol Buffers, Avro

Entity - Relationship Model Relational Model

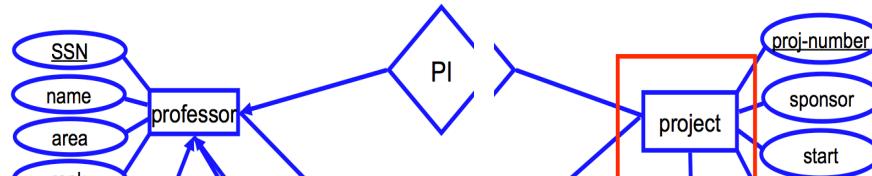


→ Physical Independence

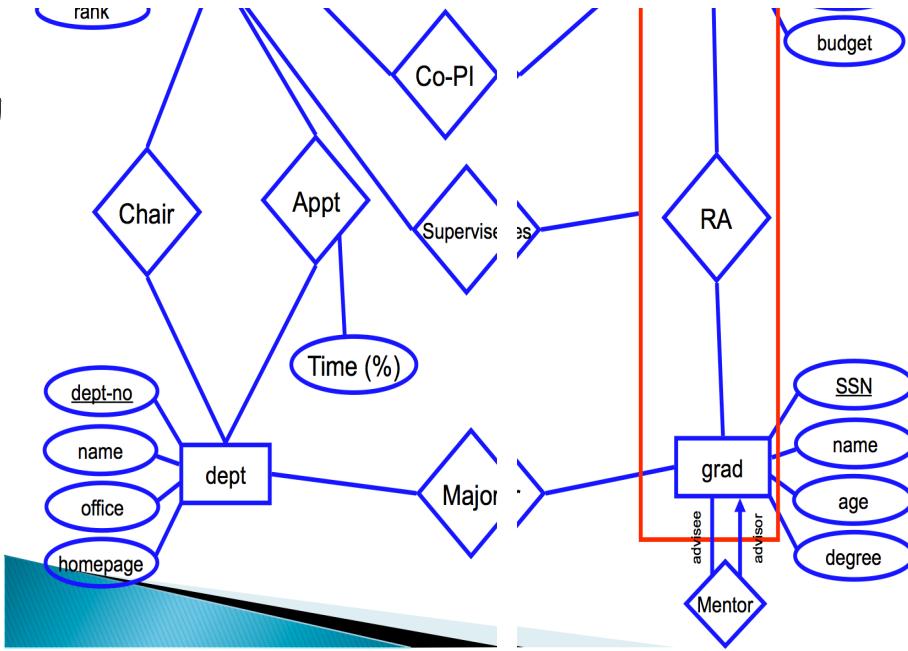
Entity - Relationship model

- Entities : Attributes
- Relationships : Attributes

D: entities



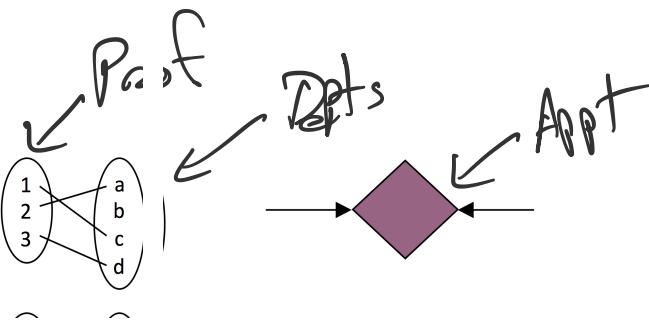
 : Relationships
 : Attributes



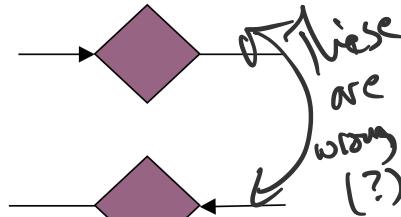
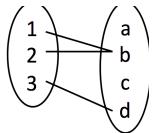
Professor (ssn, name, area, rank)

Chair (prof1, dept2)

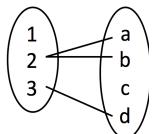
One-to-one:



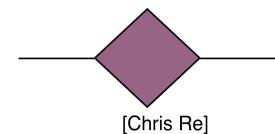
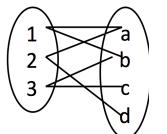
Many-to-one:



One-to-many:

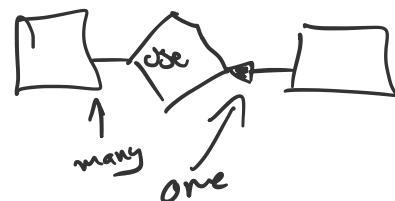


Many-to-many:



[Chris Re]

Flights planes



many - to - one

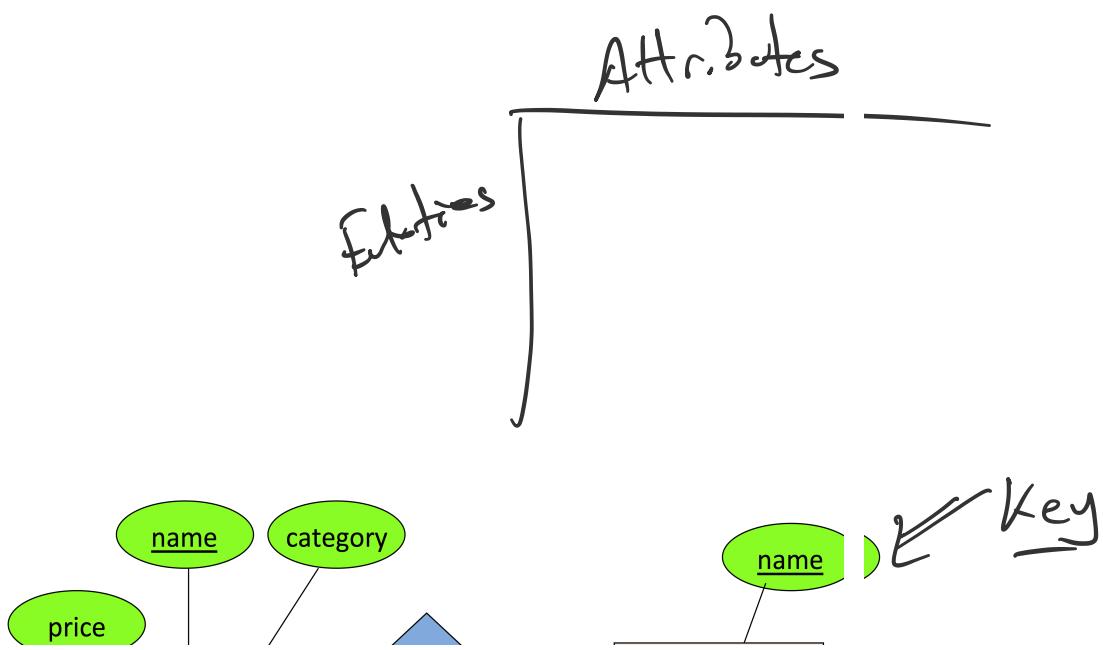


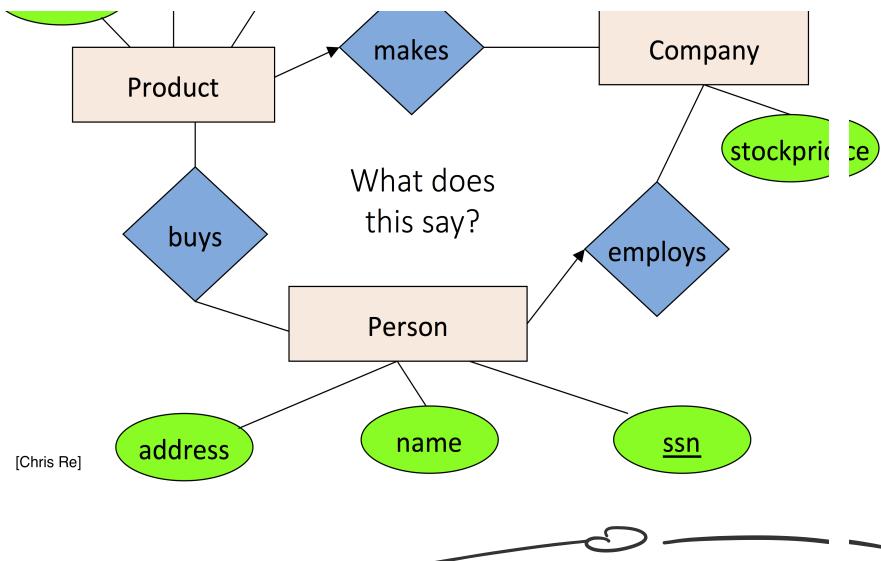
relations row

Relation: sets of tuples

- Entity: attribute values

Professor(123-456-7890, 'Joe', 'Assoc')





Entities: How entities are defined

→ Unique within a relation

→ Sets of attributes that define entities uniquely

Key

- Super Key: set of attributes which contain a subset of attributes that define entities uniquely
- Minimal Key: Super Key, that can't be minimized (can't remove any attribute)
- Candidate Key: Any minimal Key
- Primary Key: A single arbitrarily chosen candidate key

Why Keys?)

- Performance: Index entities

- Define relationships

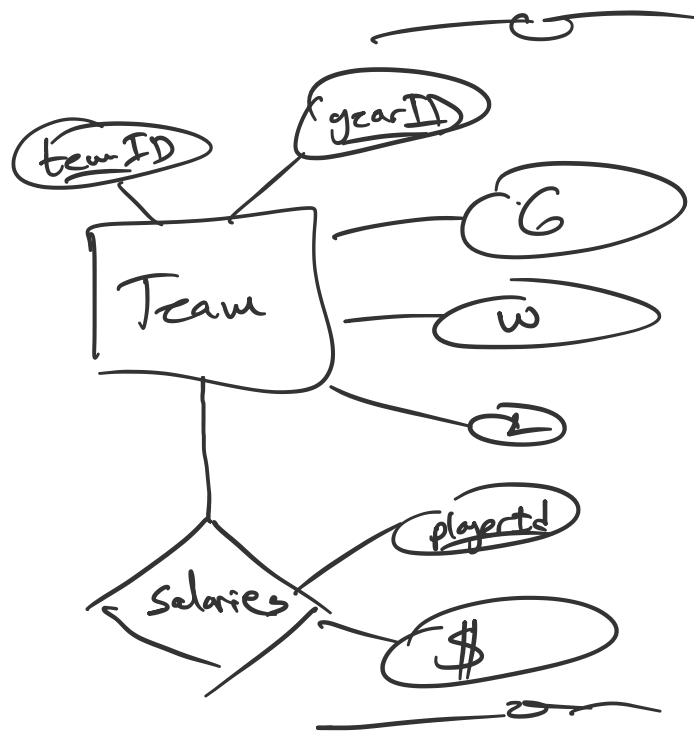
Relationship

$R : E_1 \times E_2$ (^{cross product}
_{of attr})

↑
primary key R
primary key

Foreign key: primary key + that appears in
another relation

Foreign key constraint: Primary must appear
in relationship



SQL: Structured Query Language

Declarative

Procedural df %>% Filter() or > 	Relations as arguments
---	------------------------

`tailor::coho
select() %>%
arrange()`

Describes the resulting
relation