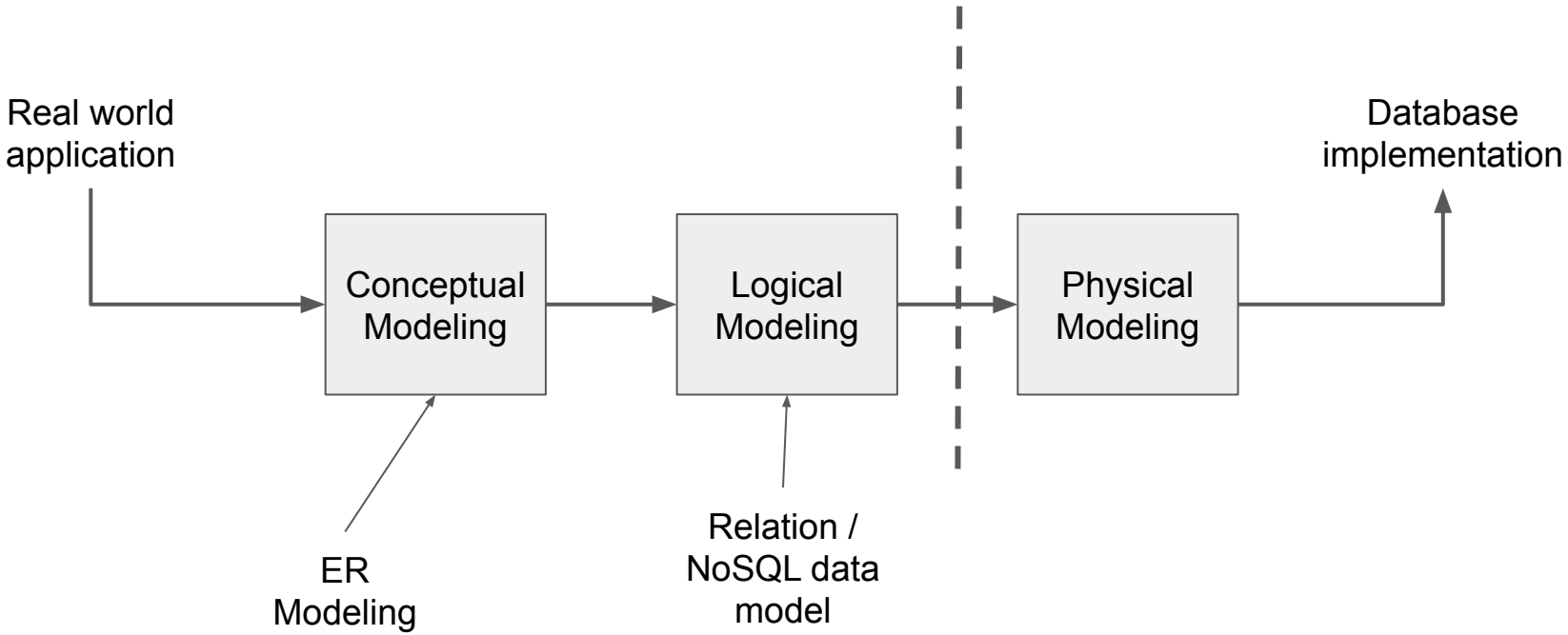
Database and Big Data Systems

Lab 2

Today

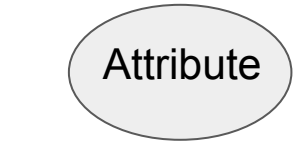
- Recap
- ER modeling
- Relational model
- Relational algebra

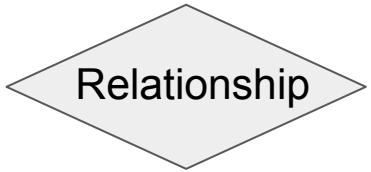
Recap


Real world application
Conceptual Modeling
Logical Modeling
Physical Modeling
Database implementation
ER Modeling
Relation / NoSQL data model

ER Modeling

- Describe data to other users
  - Conceptual
- A graphical diagram
- Entity set: a collection of similar objects
  - Object = entity
  - Attribute: property of an entity
- Entities in the same entity set have the same set of attributes
- Relationship: connection between entity sets

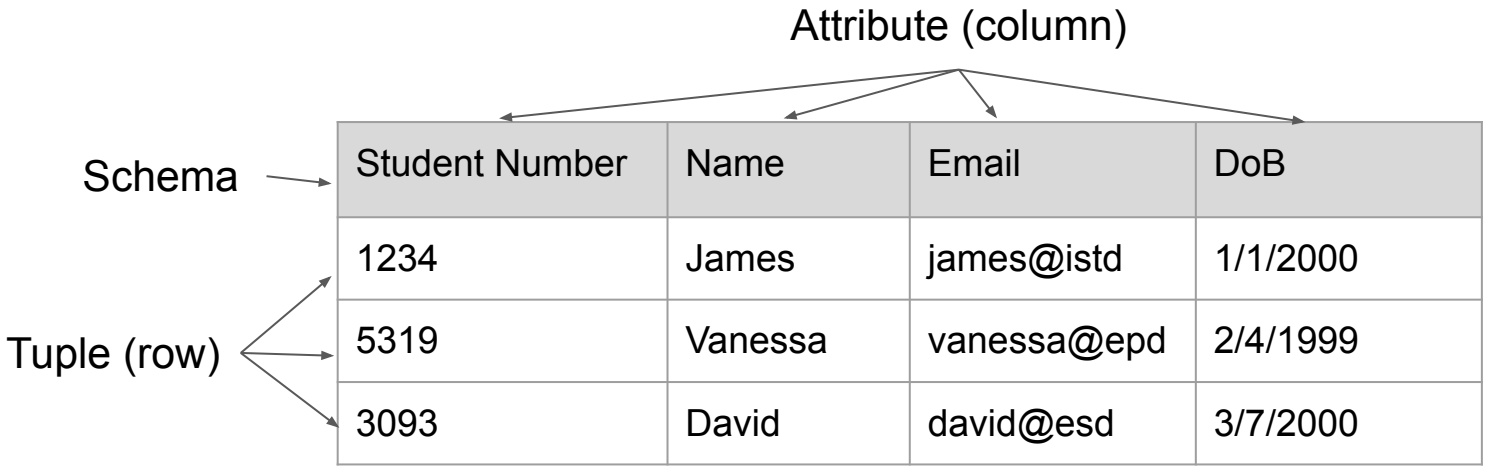




Data Model

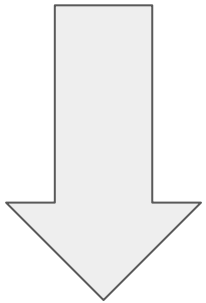
- Describe data to the database
  - Structure of the data
  - What we can do with the data
- Relational model:
  - Relation: unordered set representing relationship of attributes
  - Relation = {tuples}

Relation: Student



Student Number	Name	Email	DoB
1234	James	james@istd	1/1/2000
5319	Vanessa	vanessa@epd	2/4/1999
3093	David	david@esd	3/7/2000

ER Diagram



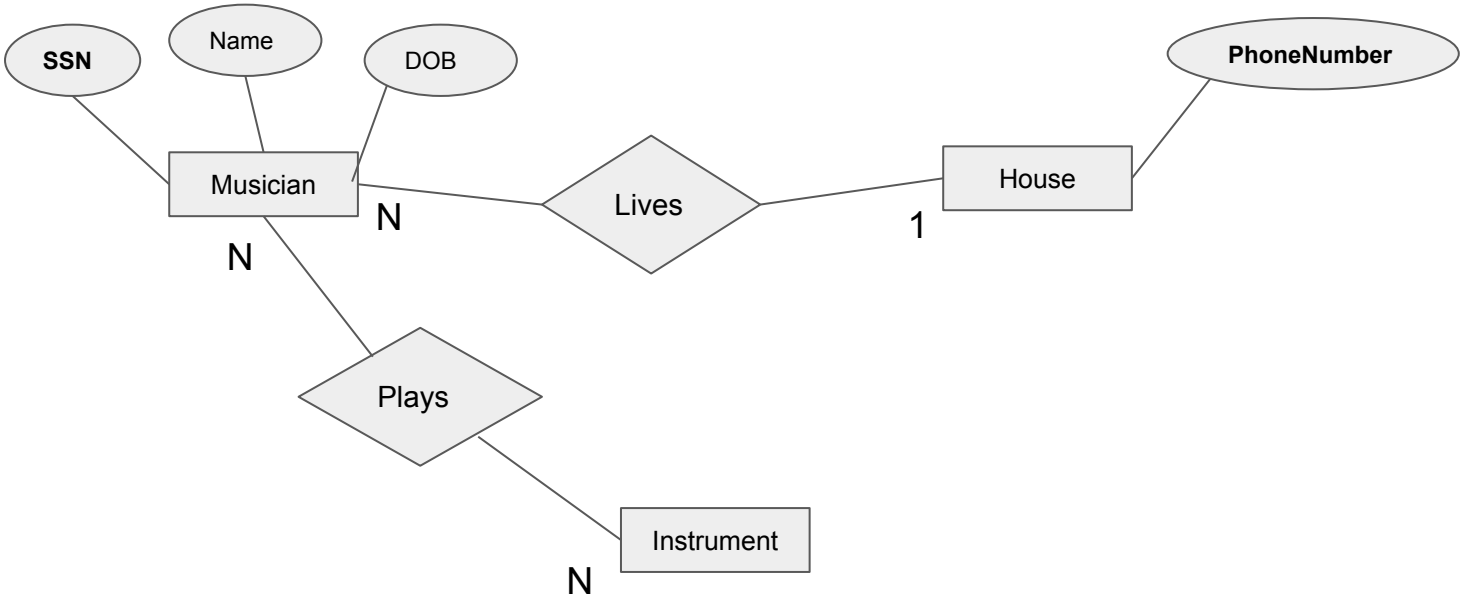
Rule 1: Entity set → Relation
Rule 2: Relationship → Relation
Rule 3: Merge relation

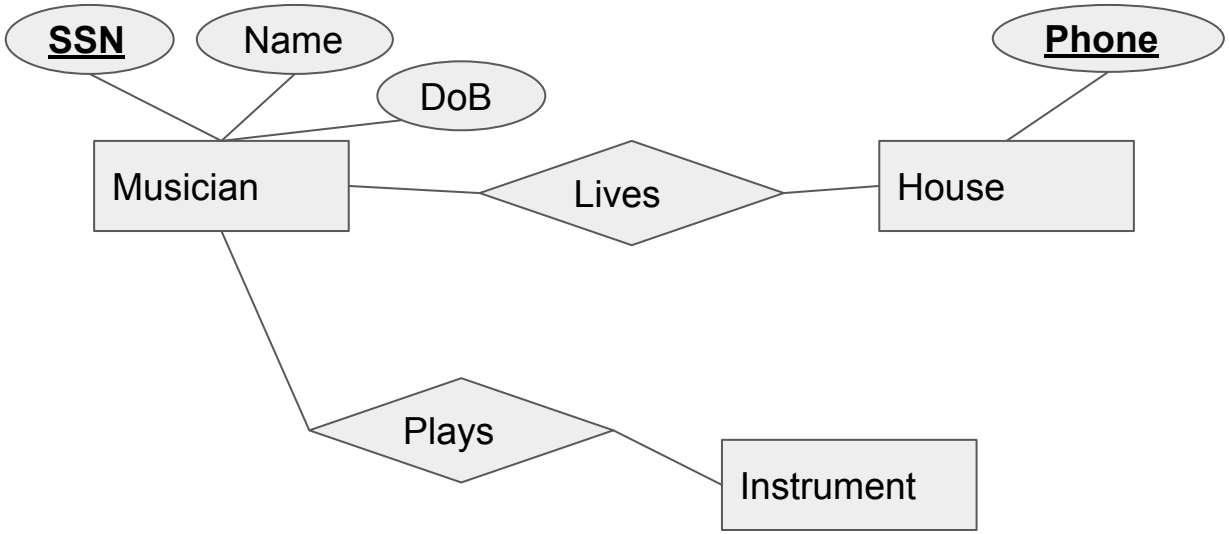
Relational Data Model

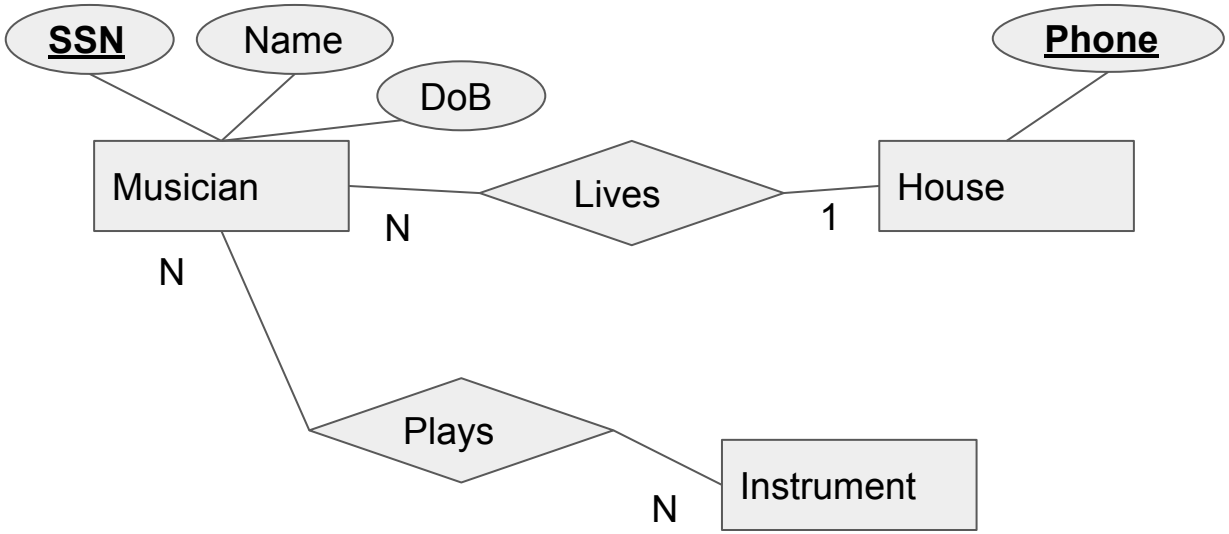
Exercise 1 - ER Diagram

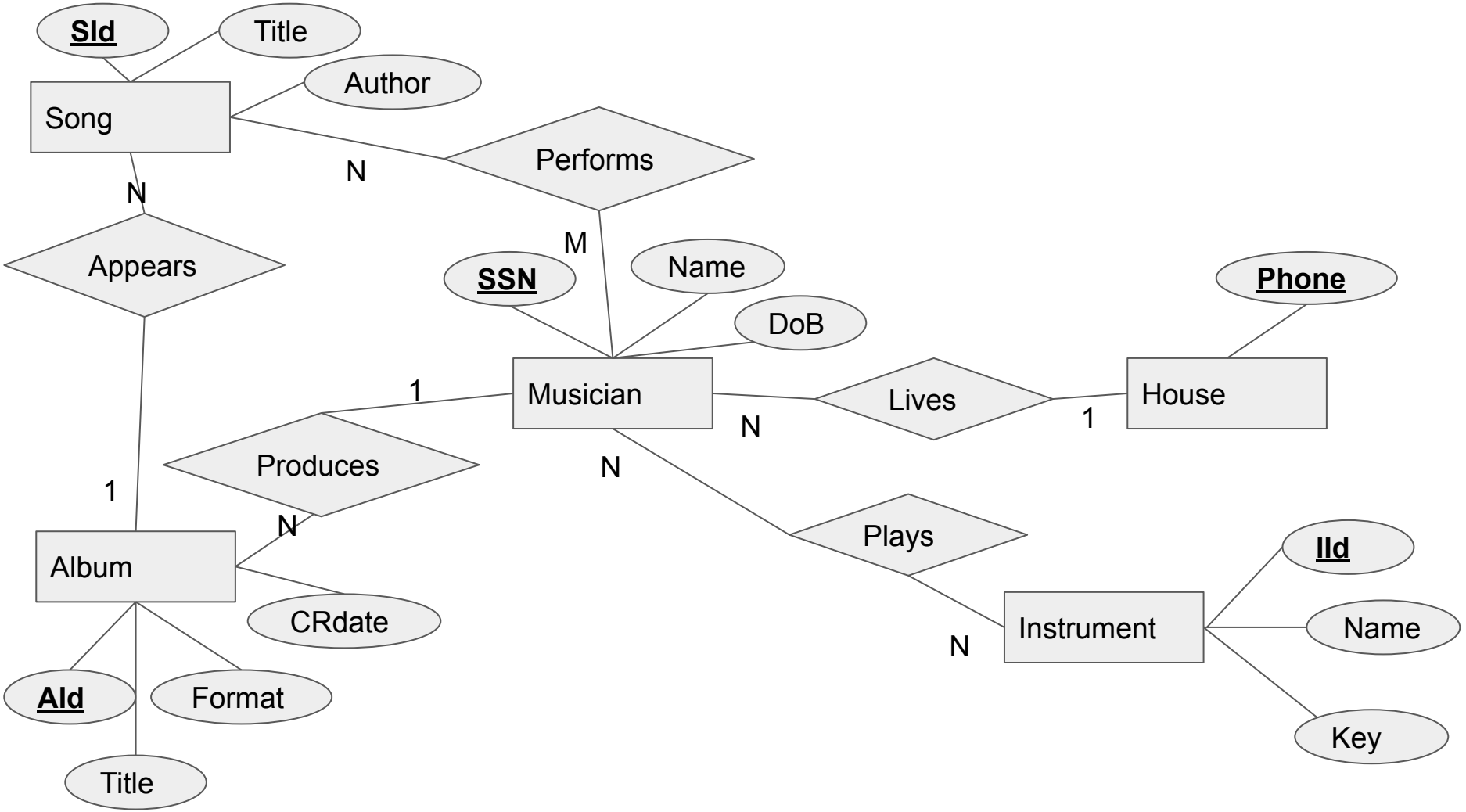
A new recording studio needs to maintain the following information:

- Each musician that records at Studify has an SSN, a name, and date of birth. Poorly paid musicians often live together at the same house which also has a phone.
- Each instrument used for recording at Studify has a unique identification number, a name (e.g., guitar, synthesizer, flute) and a musical key (e.g., C, B-flat, E-flat).
- Each album recorded has a unique identification number, a title, a copyright date, a format (e.g., CD or MC).
- Each song recorded has a unique identification number, a title, an author.
- Each musician may play several instruments, and a given instrument may be played by several musicians.
- Each album has a number of songs on it, but no song may appear on more than one album.
- Each song is performed by multiple musicians, and a musician may perform a number of songs.
- Each album has exactly one musician who acts as its producer. A musician may produce several albums.


SSN
Name
DOB
PhoneNumber
Musician
Lives
House
N
1
N
Plays
Instrument
N


SSN
Name
DoB
Phone
Musician
Lives
House
Plays
Instrument

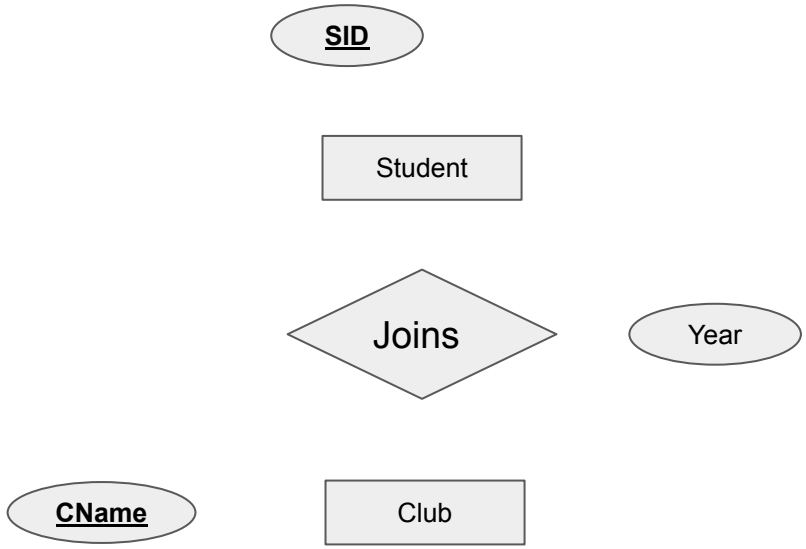

SSN
Name
DoB
Phone
Musician
Lives
House
N
1
N
Plays
Instrument
N

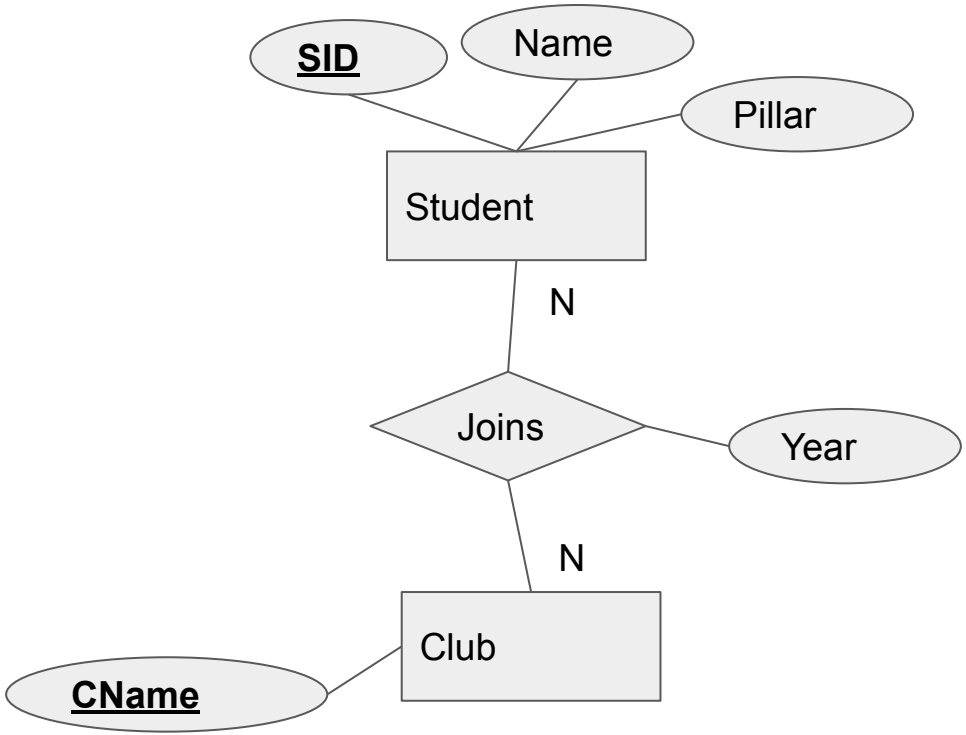

SId
Title
Author
Song
Performs
N
N
M
Appears
SSN
Name
DoB
Phone
Musician
Lives
House
1
N
1
Produces
N
1
N
Plays
Album
CRdate
IId
Instrument
Name
N
AId
Format
Key
Title

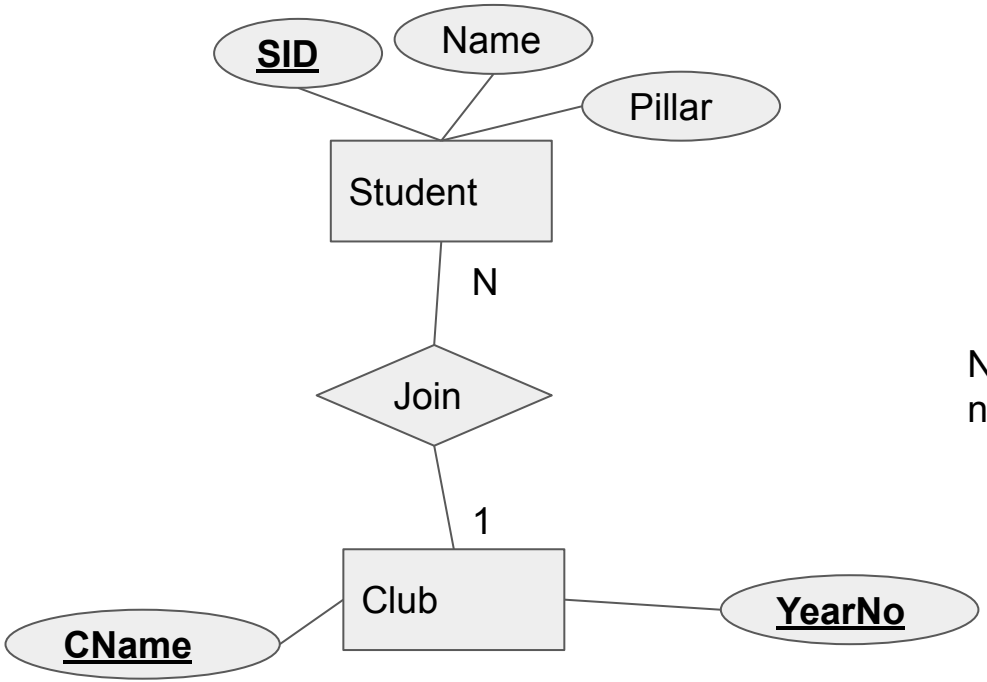
Exercise 2 - ER Diagram

You want to manage student clubs at the university. But you need to adhere to the following constraints:

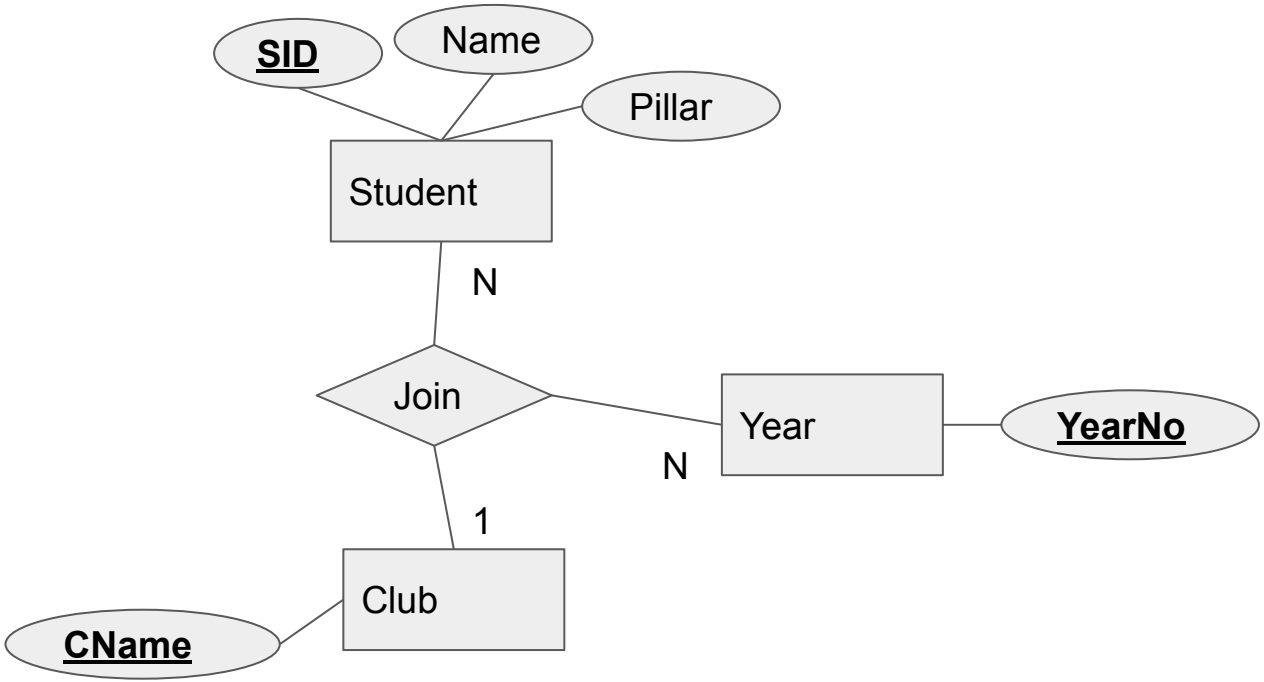
- Each student has a unique student identification number, a name, and pillar.
- Each club has a unique name
- Every year, students can join a club. One student may belong to different clubs in different year, but may belong to at most one club in any given year.
- Each club may travel to different cities. Every city has a name, and belongs to a country
- Each club may travel to different cities in different year, but only to one city in any given year.

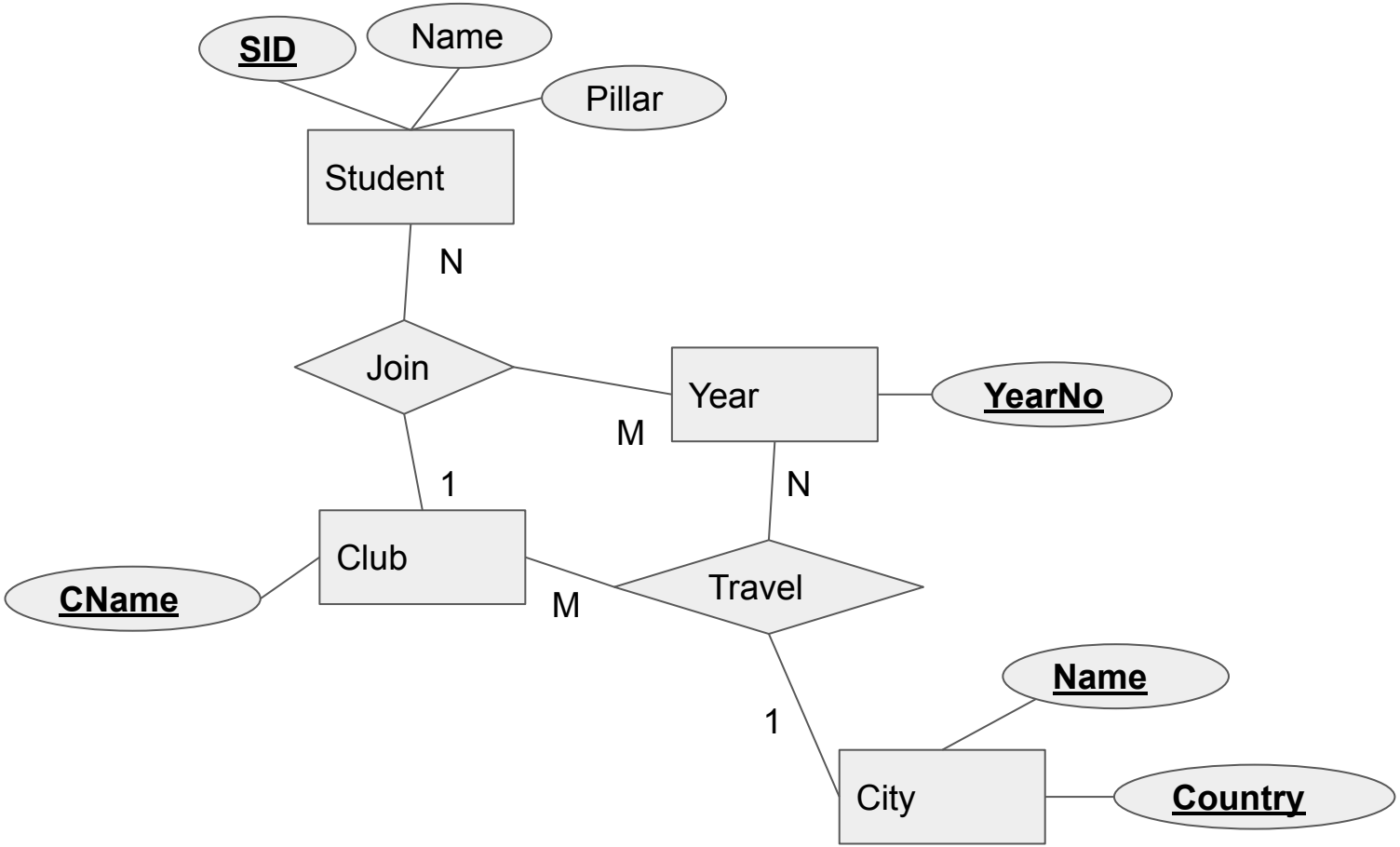

SID
Student
Joins
Year
CName
Club


SID
Name
Pillar
Student
N
Joins
Year
N
Club
CName



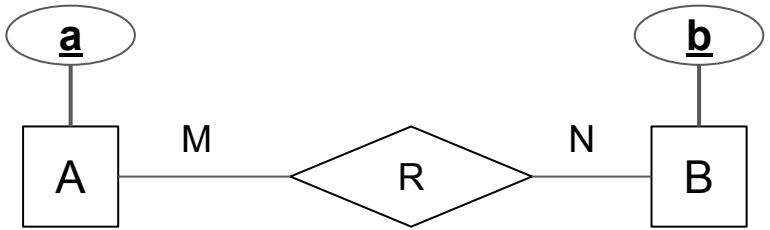
Note: this won't work as the uniqueness of CName is not enforced.


SID
Name
Pillar
Student
N
Join
Year
YearNo
N
1
Club
CName


SID
Name
Pillar
Student
N
Join
Year
YearNo
M
1
N
Club
CName
Travel
M
1
Name
City
Country

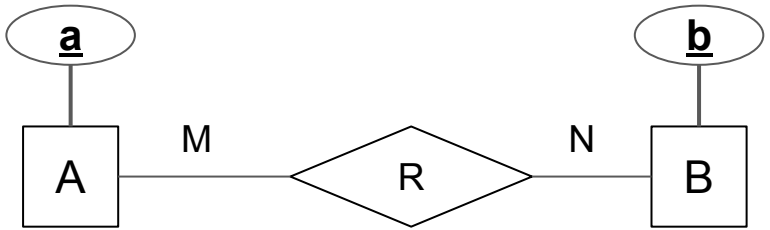
Exercise 3 - Relational Model

Translate this to relations



Exercise 3 - Relational Model

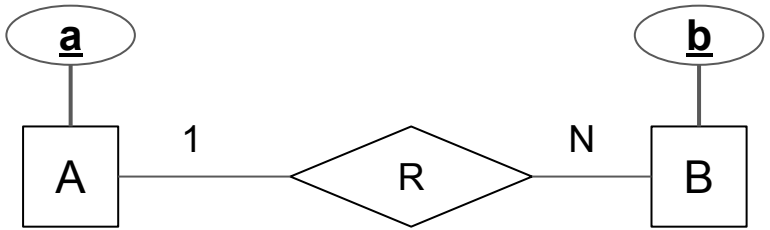
Translate this to relations



A(**a**)
B(**b**)
R(**a**,**b**)

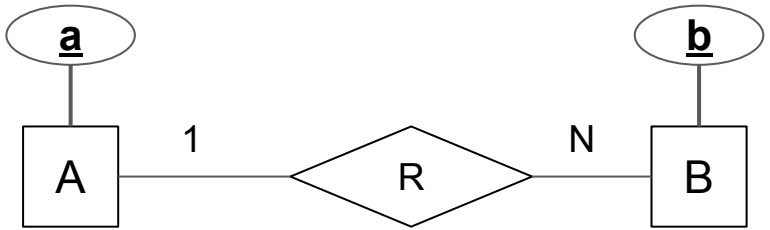
Exercise 4 - Relational Model

Translate this to relations



Exercise 4 - Relational Model

Translate this to relations



A(**a**)
B(**b**)

R(a,**b**)

Exercise 5 - Relational Algebra

Given the following table which records the results of running competition at the Olympic

Run(**<u>Name, Distance,</u>** Time)

Give an expression in relational algebra that finds all runners who:

- Take part in 100m category.
- Only take part in 100m category

$$\pi_{Name}(\sigma_{Distance=100m}(Run))$$

$$\pi_{Name}(\sigma_{Distance!=100m}(Run))$$

Exercise 5 - Relational Algebra

Given the following table which records the results of running competition at the Olympic

Run(**<u>Name, Distance,</u>** Time)

Give an expression in relational algebra that finds all runners who:

- Take part in 100m category.
- Only take part in 100m category

$$\pi_{Name}(\sigma_{Distance=100m}(Run))$$

$$\pi_{Name}(\sigma_{Distance=100m}(Run)) - \pi_{Name}(\sigma_{Distance!=100m}(Run))$$

Exercise 6 - Relational Algebra

Given the following relations modeling a library. A book in a library may have multiple copies.

Reader (ReaderID, FirstName, LastName)
Book (ISBN, Title, Author, PublicationDate, PublisherName)
Publisher (PublisherName, PublisherCity)
Copy(ISBN, CopyID, ShelfLocation)
Loan (ReaderID, ISBN, CopyID, ReturnDate)

Find names of readers who borrow more than 10 copies of a book.

$$R' = \sigma_{C>10}({}_{\text{ReaderID, ISBN}}\gamma_{\text{Count(CopyID) as C}}(\text{Loan}))$$

R'(ReaderID, ISBN, C)

Exercise 6 - Relational Algebra

Given the following relations modeling a library. A book in a library may have multiple copies.

Reader (ReaderID, FirstName, LastName)
Book (ISBN, Title, Author, PublicationDate, PublisherName)
Publisher (PublisherName, PublisherCity)
Copy(ISBN, CopyID, ShelfLocation)
Loan (ReaderID, ISBN, CopyID, ReturnDate)

Find names of readers who borrow more than 10 copies of a book.

$$\pi_{\text{FirstName, LastName}}(\text{Reader} \bowtie \sigma_{C>10}({}_{\text{ReaderID, ISBN}}\gamma_{\text{Count(CopyID) as C}}(\text{Loan}))$$

Exercise 7 - Relational Algebra

Given the following relations modeling a library. A book in a library may have multiple copies.

Reader (ReaderID, FirstName, LastName)
Book (ISBN, Title, Author, PublicationDate, PublisherName)
Publisher (PublisherName, PublisherCity)
Copy(ISBN, CopyID, ShelfLocation)
Loan (ReaderID, ISBN, CopyID, ReturnDate)

Find title and author of book from publisher in London or New York

$$\pi_{\text{Title, Author}}(\text{Book} \bowtie \sigma_{\text{PubliserCity=London OR PublisherCity=NewYork}}(\text{Publisher}))$$

Exercise 7 - Relational Algebra

Given the following relations modeling a library. A book in a library may have multiple copies.

Reader (ReaderID, FirstName, LastName)
Book (ISBN, Title, Author, PublicationDate, PublisherName)
Publisher (PublisherName, PublisherCity)
Copy(ISBN, CopyID, ShelfLocation)
Loan (ReaderID, ISBN, CopyID, ReturnDate)

Find title and author of book from publisher in London or New York

$$\pi_{\text{Title, Author}}(\sigma_{\text{PubliserCity=London OR PublisherCity=NewYork}}(\text{Book} \bowtie \text{Publisher}))$$

Exercise 8 - Relational Algebra

Given the following relations:

Reader (ReaderID, FirstName, LastName)
Book (ISBN, Title, Author, PublicationDate, PublisherName)
Publisher (PublisherName, PublisherCity)
Loan (ReaderID, ISBN, Copy, ReturnDate)

Find title and author of books that Anh Dinh borrowed

$$\pi_{\text{Title,Author}}(\text{Book} \bowtie (\sigma_{\text{FirstName= Anh AND LastName=Dinh}}(\text{Reader}) \bowtie \text{Loan})))$$

Exercise 9 - Relational Algebra

Given the following relations modeling a library. A book in a library may have multiple copies.

Reader (ReaderID, FirstName, LastName)
Book (ISBN, Title, Author, PublicationDate, PublisherName)
Publisher (PublisherName, PublisherCity)
Copy(ISBN, CopyID, ShelfLocation)
Loan (ReaderID, ISBN, CopyID, ReturnDate)

Find name of users who borrowed at least two different books

$$\pi_{\text{FirstName, LastName}}(\text{Reader} \bowtie \sigma_{\text{Loan.ISBN != Loan1.ISBN}}(\text{Loan} \bowtie_{\text{Loan.ReaderID = Loan1.ReaderID}} \rho_{\text{Loan1}}(\text{Loan})))$$

Reader	ISBN	Reader	ISBN
R1	I1	R1	I1
R1	I1	R1	I2
R1	I2	R1	I1
R1	I2	R1	I2
R2	I1	R2	I1

Reader	ISBN
R1	I1
R1	I2
R2	I1

Exercise 9 - Relational Algebra

Given the following relations modeling a library. A book in a library may have multiple copies.

Reader (ReaderID, FirstName, LastName)
Book (ISBN, Title, Author, PublicationDate, PublisherName)
Publisher (PublisherName, PublisherCity)
Copy(ISBN, CopyID, ShelfLocation)
Loan (ReaderID, ISBN, CopyID, ReturnDate)

Find name of users who borrowed at least two different books

$$\pi_{\text{Reader.FirstName, Reader.LastName}}(\text{Reader} \bowtie (\sigma_{\text{User\_Count}>1} ({}_{\text{ReaderID}}\gamma_{\text{Count-distinct(ISBN) as User\_Count}} (\text{Loan})))$$