Week 04
Relational Algebra

Labs

Relational Algebra symbols

Symbol	RA Term	SQL mapping
π, Π	Projection	SELECT
σ, Σ	Selection	FROM and WHERE
⋈	Join	JOIN
ρ, R	Rename	AS

Relational Algebra

$\pi_{attributesToProject}\sigma_{attributesToFilter/Select}$(Relation)

- Please note, that while we show some solutions, there are others that are also correct, and still others that are wrong
- It is important to ask the lab demonstrator if you are unsure about something you suggested (don’t just try to learn the same answer we offer!)

Book Schema

Oops: error in diagram!
Author table has primary key which is aname;
Birthdate should not be underlined too

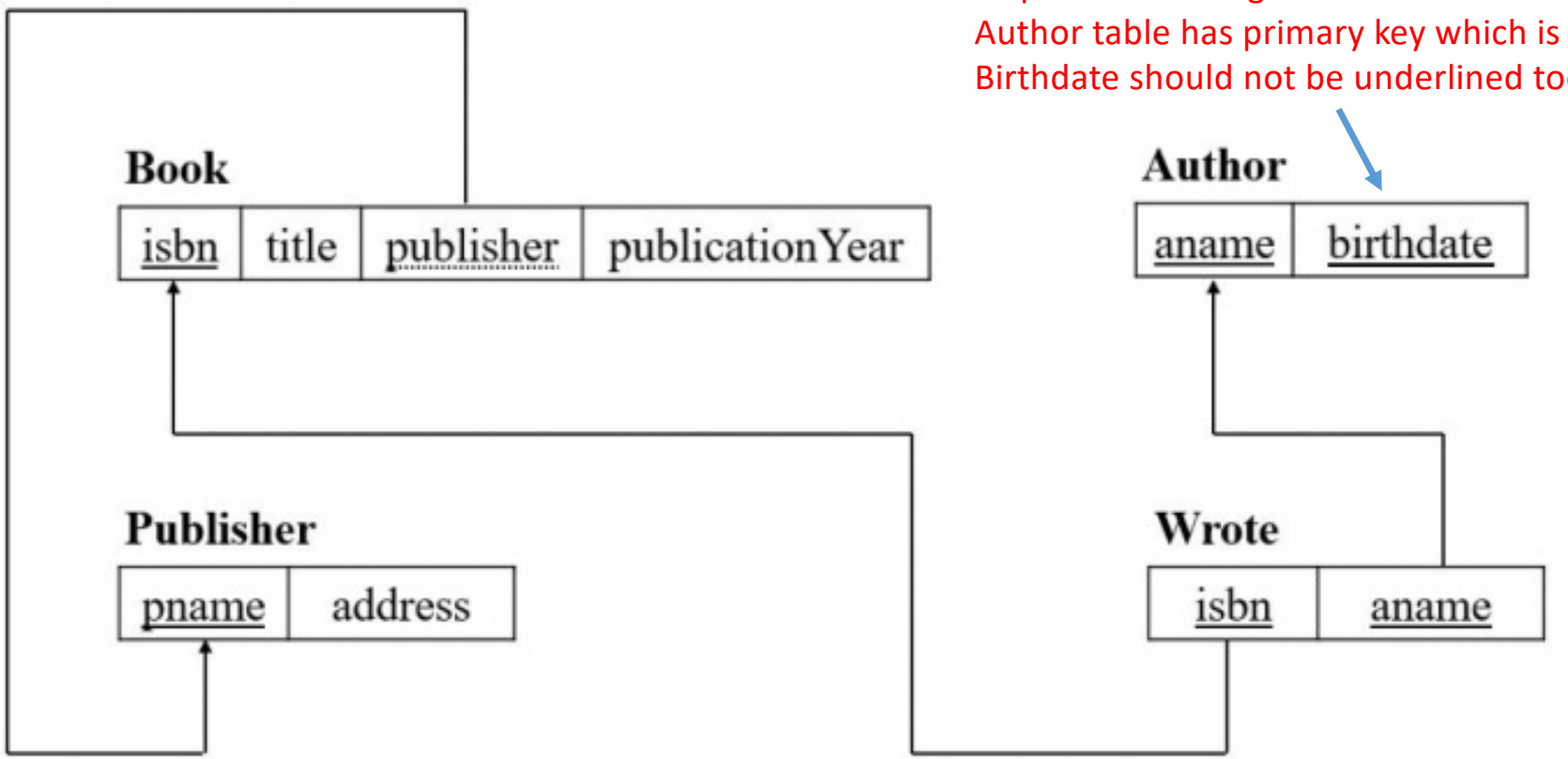


Figure 1: Schema Diagram

$\pi_{attributesToProject}\sigma_{attributesToFilter/Select}$(Relation)

$\pi_{title,publicationYear}$ (Book)

Find title, publication year of all Books

$\pi_{pname}\sigma_{address='New\ York'}$(Publisher)

Find names of Publishers with address New York.

$\pi_{aname}\sigma_{title='A\ First\ Discource\ in\ Database\ Systems'}$(Book⋈Wrote)

Find all names of authors who wrote the book 'A First Discourse in Database Systems'.

Why do we need the condition, instead of natural join?

$\pi_{address}\sigma_{title='Database' \vee title='Data\ Management'}$(Publisher $\bowtie_{pname\ =\ publisher}$ Book)

Find the address of Publishers who published a book with the title 'Database' or 'Data Management'

A (f)

Error 1 – no pname

Fix error 1, but then Error 2 – no title field in the result of projection, so selection won't work

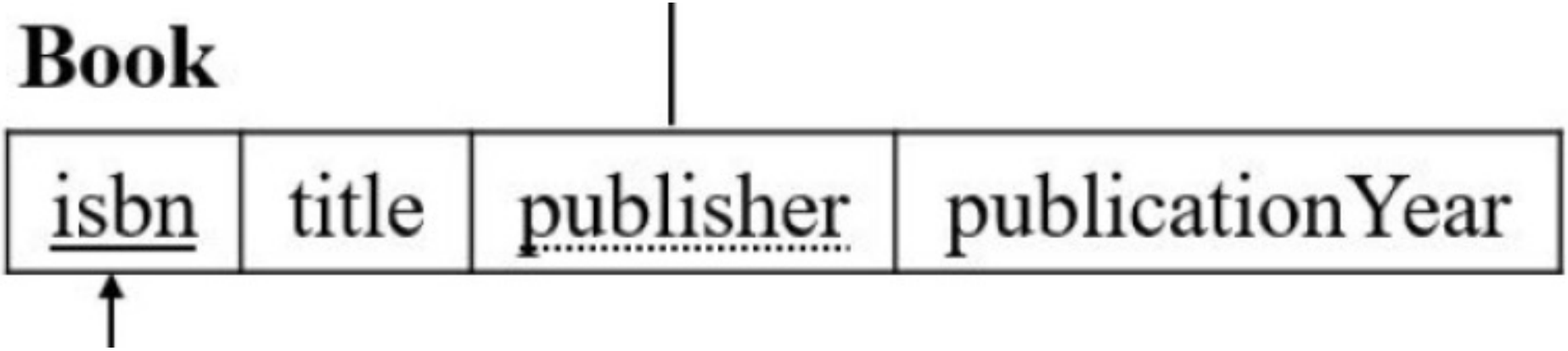
$\sigma_{title='New\ York'}(\pi_{pname}$ (Book))

$(\pi_{publisher}$ (Book))

$\sigma_{title='New\ York'}(\pi_{publisher}$ (Book))

publisher

$\sigma_{title='New\ York'}(\pi_{publisher,title}$ (Book))



B(a)

(a) Find the book titles published by Acme Publishers

$\pi_{title}\sigma_{publisher='ACME'}$(Book)

B(b)

(b) Find all authors (give their name) of the book with ISBN 0444455551

$\pi_{aname}\sigma_{isbn=044455551}$(Wrote)

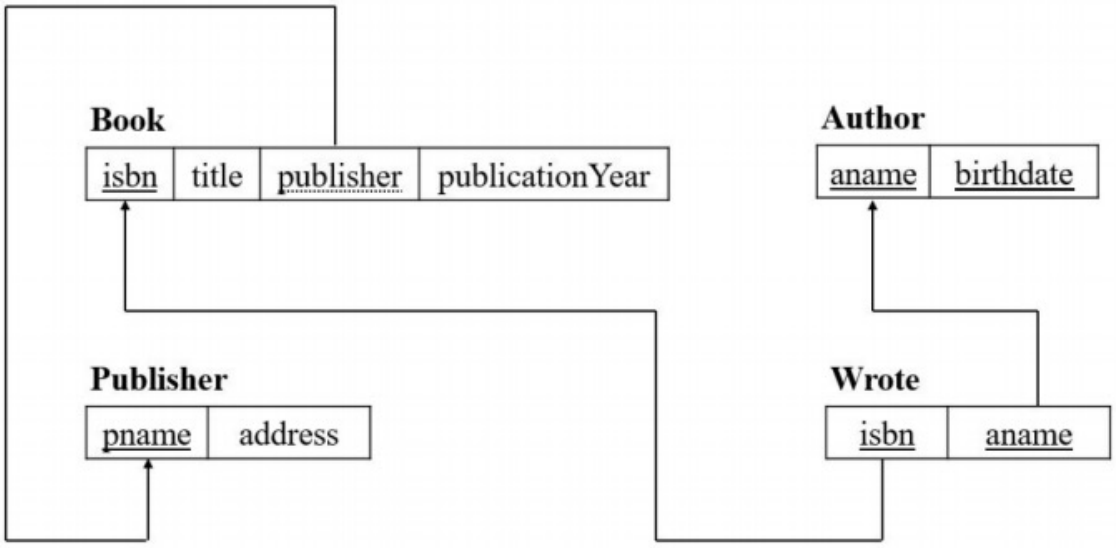


Figure 1: Schema Diagram

B(c)

(c) Find the authors (by name) who published at least one book with Acme Publishers

$\pi_{aname}\sigma_{publisher='ACME'}$(Book ⋈Wrote)

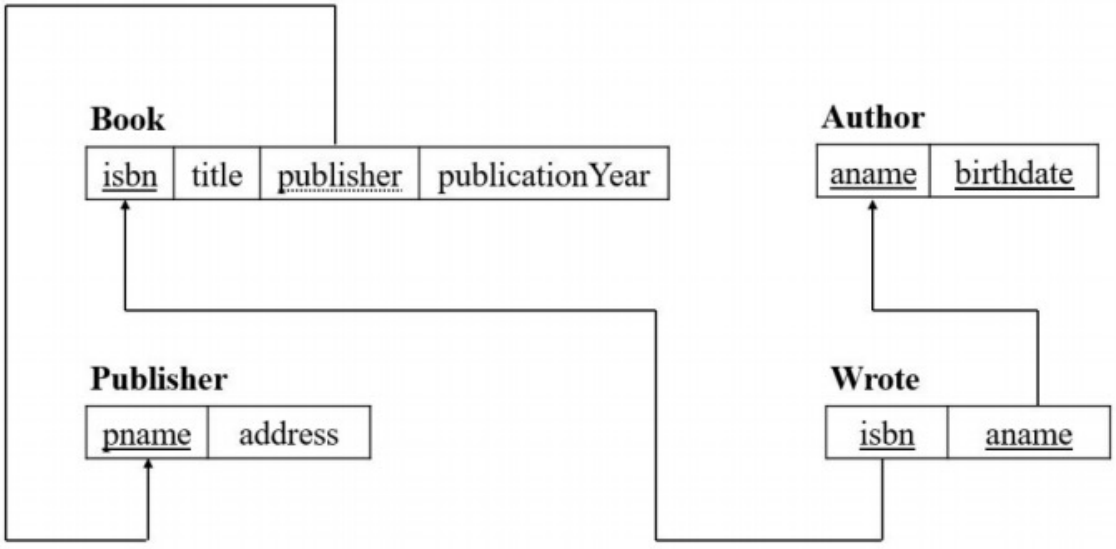


Figure 1: Schema Diagram

B(d)

(d) Find all authors (name) who have never published a book with Acme Publishers

=> Find all authors(name) – Find all author(name) who have published at least one book with ACME publishers

π_{aname}(Author) - $\pi_{aname}\sigma_{publisher='ACME'}$(Book ⋈Wrote)

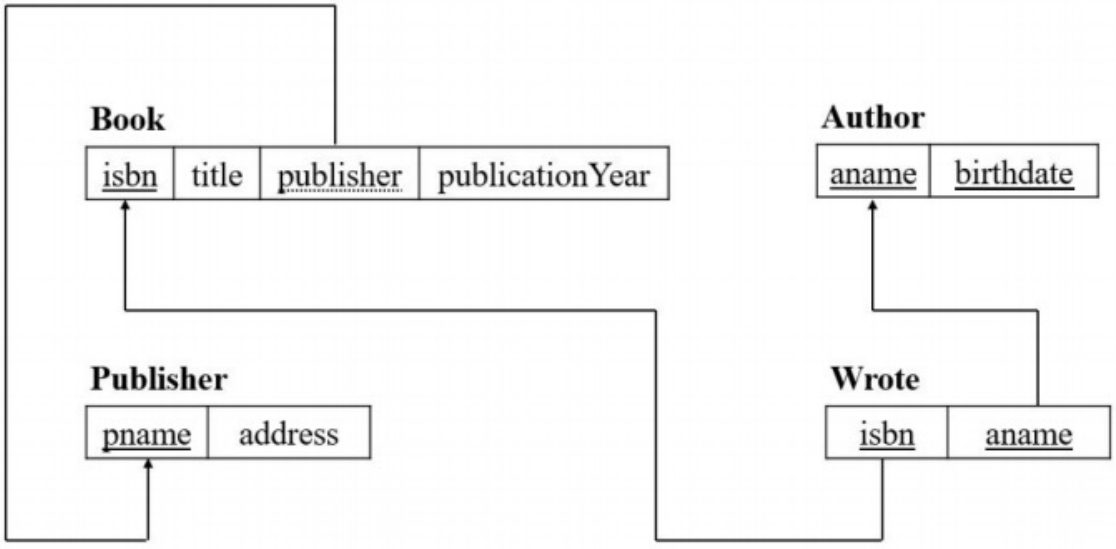


Figure 1: Schema Diagram

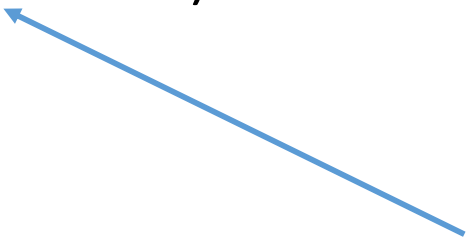
How to approach writing a SQL Query

1. Where can I find my data?
2. Do I need to combine any tables and on which attributes should I join them?
3. Which columns should I retrieve?
4. Do I need to apply any filters?
5. Do I need to apply any ordering?
6. Do I need to use set-based thinking?
7. Do I need to use a subquery?

C(a)

- How many books were published by each publisher

```
SELECT pname, COUNT(DISTINCT isbn)
FROM Book
GROUP BY pname
```



DISTINCT is not actually needed here, because Book already has isbn as primary key

C(b)

- For each author, show their name and the total number of books which they wrote and were published after 1975

```
SELECT Wrote.aname, COUNT(DISTINCT Wrote.isbn)
FROM Book, Wrote
WHERE Book.isbn = Wrote.isbn
        AND Book.publicationYear > 1975
GROUP BY Wrote.aname
```

C(c)

- Which book(s) were published most recently
- => Find books whosepublication year is the largest publication year among all books

```
SELECT b.isbn
FROM Book b
WHERE b.publicationYear = (SELECT MAX(publicationYear)
                           FROM Book
                          )
```

C(d)

- Find the authors who wrote a book published in 1995 and also wrote a book published in 2015
  - Use a set operation
- => (authors who wrote a book published in 1995) intersect (authors who wrote a book published in 2015)

(SELECT aname FROM Book JOIN Wrote WHERE publicationYear = 1995)

INTERSECT

(SELECT aname FROM Book JOIN Wrote WHERE publicationYear = 2015)

C(d)

- Find the authors who wrote a book published in 1995 and also wrote a book published in 2015
  - Use a subquery
- => (authors who wrote a book published in 1995 and are among the authors who wrote a book published in 1995)

```
(SELECT aname FROM Book JOIN Wrote
WHERE publicationYear = 1995
        AND aname IN (SELECT aname
                      FROM Book JOIN Wrote
                      WHERE publicationYear = 2015)
```

C(d)

- Find the authors who wrote a book published in 1995 and also wrote a book published in 2015
  - Use a simple query
- => (author who appears in two rows of Book JOIN Wrote result, one for a book published in 1995 and another for a book published in 2015)

```
(SELECT r1.aname
FROM (Book JOIN Wrote) r1, (Book JOIN Wrote) r2
WHERE r1.aname = r2.aname
        AND r1.publicationYear = 1995
        AND r2.publicationYear = 2015
```

C(e)

- Find any author who wrote every book which has “Database” in its title

=>Find author where number of books they wrote with Database in title, equals total number of books with Database in title

```
SELECT aname
FROM Book JOIN Wrote
WHERE title LIKE ‘%Database%’
GROUP BY aname
HAVING COUNT(DISTINCT isbn) = (SELECT COUNT(DISTINCT isbn)
                               FROM Book
                               WHERE title LIKE ‘%Database%’)
```