

# Homework 2

Jacob Taylor Cassady

CECS 535: Introduction to Databases

February 19, 2020

The following database schema is given:

RESEARCHER(rid,name,institution,city,country)

PAPER(title,journal,volume,number,year)

AUTHOR(resid,title)

where rid is the identifier (primary key) of RESEARCHER, name is the researcher's name, institution is where the researcher works, and city and country the city and country where the institution is located; title is the paper identifier (primary key) of PAPER, journal is the journal where it was published (in volume volume and number number), and year is the year it appeared. Finally, in AUTHOR resid is a foreign key for RESEARCHER and title is a foreign key for PAPER. A researcher may write several papers, and papers may be jointly written by several researchers; this means that the key of Author is (resid,title).

- 
- 1 LIST THE NAMES OF AUTHORS OF ANY PAPER PUBLISHED IN JOURNAL "DATABASES" IN 2019 IN VOLUME 12.
- 

**SELECT** R.name

**FROM** RESEARCHER **AS** R, AUTHOR **AS** A, PAPER **AS** P

**WHERE** R.resid **and** P.title = A.title **and** P.journal = "Databases" **and** P.year = 2019 **and** P.volume=12;

- 
- 2 LIST THE NAMES OF AUTHORS WHO HAVE PUBLISHED A PAPER IN BOTH "NATURE" AND IN "SCIENCE" (JOURNALS).
- 

**SELECT** name

**FROM** (SELECT name, rid

**FROM** RESEARCHER **AS** R, AUTHOR **AS** A, PAPER **AS** P

**WHERE** R.resid = A.resid **and** P.title = A.title **and** R.journal = "Nature")

**INTERSECT**

(SELECT name, rid

**FROM** RESEARCHER **AS** R, AUTHOR **AS** A, PAPER **AS** P

**WHERE** R.resid = A.resid **and** P.title = A.title **and** R.journal = "Science");

- 
- 3 LIST THE NAMES OF AUTHORS WHO HAVE PUBLISHED TWO OR MORE PAPERS IN "NATURE" (JOURNAL).
- 

**SELECT** name

**FROM** RESEARCHER **AS** R, AUTHOR **AS** A

**WHERE** R.resid = A.resid **and** R.resid **IN** (SELECT A.resid

**FROM** PAPER **AS** P, AUTHOR **AS** A

**WHERE** P.title = A.title

**GROUP BY** A.resid

**HAVING** sum(CASE journal="Nature" 1 else 0) >= 2);

- 4 LIST THE NAMES OF AUTHORS WHO HAVE PUBLISHED A PAPER IN "NATURE" BUT NEVER IN "SCIENCE" (JOURNALS).
- 

```
SELECT name
FROM (SELECT name, resid
      FROM RESEARCHER AS R, AUTHOR AS A, PAPER AS P
      WHERE R.resid = A.resid and P.title = A.title and R.journal = "Nature")
EXCEPT
(SELECT name, resid
 FROM RESEARCHER AS R, AUTHOR AS A, PAPER AS P
 WHERE R.resid = A.resid and P.title = A.title and R.journal = "Science");
```

- 5 LIST THE NAMES OF AUTHORS WHO HAVE PUBLISHED A PAPER IN "NATURE" (JOURNAL) ALONE (I.E. WITHOUT CO-AUTHORS).
- 

```
SELECT name
FROM (SELECT name, resid
      FROM RESEARCHER AS R, AUTHOR AS A
      WHERE R.resid = A.resid and A.title IN (SELECT title
                                             FROM (SELECT title, COUNT(resid) AS aCount
                                                  FROM AUTHOR
                                                  GROUP BY title)
                                             WHERE aCount = 1;
```

- 6 LIST THE TITLE OF PAPERS WHERE ALL AUTHORS ARE FROM THE SAME INSTITUTION. NOTE: THIS INCLUDES SOLE AUTHORS TOO.
- 

```
SELECT P.title
FROM RESEARCHER AS R, AUTHOR AS A, PAPER AS P
WHERE R.resid = A.resid and P.title = A.title and A.institution = (SELECT
                                                                FROM RESEARCHER AS R1,
                                                                AUTHOR AS A1, PAPER AS P1
                                                                WHERE P.title = P1.title and
                                                                R1.resid = A1.resid);
```

- 7 FIND OUT THE AVERAGE NUMBER OF AUTHORS PER PAPER.
- 

```
SELECT A.title, AVG(resid)
FROM AUTHOR AS A, PAPER AS P
WHERE A.title = P.title
GROUP BY A.title
```

- 8 FIND OUT THE NUMBER OF PAPERS PER COUNTRY (COUNT EACH PAPER ONCE, REGARDLESS OF THE NUMBER OF AUTHORS).
- 

```
SELECT R.country, COUNT(DISTINCT P.title)
FROM AUTHOR AS A and PAPER AS P and RESEARCHER AS R
WHERE A.title = P.title and R.resid = P.resid
GROUP BY R.country;
```

- 9 LIST THE NUMBER OF PAPERS PER AUTHOR (IDENTIFIED BY AUTHOR ID) BUT COUNT EACH PAPER IN PROPORTION TO THE NUMBER OF AUTHORS. THAT IS, IF A PAPER HAS N AUTHORS, SPLIT THE CREDIT EQUALLY AMONG THEM ( $1/N$  UNITS OF CREDIT). EXAMPLE: IF RESEARCHER ULLMAN HAS 3 PAPERS, ONE ALONE, ONE WITH SOMEBODY ELSE, AND ONE WITH TWO OTHER CO-AUTHORS, HE GETS 1, .5 AND .3 UNITS OF CREDIT RESPECTIVELY, FOR A TOTAL OF 1.8 PAPERS.
- 

```
CREATE VIEW TitleResearcherCount(title, researchercount) AS  
SELECT A.title, COUNT(A.resid)  
FROM AUTHOR AS A  
GROUP BY A.title;
```

```
CREATE VIEW ResearcherTitleScore(rid, title, authorscore) AS  
SELECT R.rid, A.title, 1/COUNT(A.*)  
FROM RESEARCHER AS R, AUTHOR AS A, TitleResearcherCount AS TRC  
WHERE R.rid = A.resid and TRC.title = A.title  
GROUP BY R.rid;
```

```
SELECT R.rid, SUM(RTS.authorscore)  
FROM RESEARCHER AS R, ResearcherTitleScore AS RTS  
WHERE R.rid = RTS.rid  
GROUP BY R.rid;
```

- 10 LIST THE JOURNALS THAT PUBLISHED MORE PAPERS IN 2019 THAN IN 2018.
- 

```
SELECT P.journal, P.year  
FROM PAPER AS P  
WHERE P.year = 2019  
GROUP BY P.journal, P.year  
HAVING COUNT(DISTINCT P.title) > (SELECT 2018_COUNT  
                                FROM (SELECT COUNT(DISTINCT P1.title) AS 2018_COUNT,  
                                P1.journal, P1.year  
                                FROM PAPER AS P1  
                                WHERE P1.journal = P.journal and P1.year = 2018));
```