

# Homework 1

CS430 and CS630

100 points

Due date: **Wed, July 26, 2023 before 10am**

**Instructions (please read carefully):** The homework needs to be submitted in a **single PDF file** called **H1\_<yourStudentID>\_<FirstName>.PDF** (e.g. if your student id is 12345678 and your name is Joe Doe, the file will be named H1\_12345678\_JoeDoe.PDF). **At the beginning of the PDF please also type your name and your student id.** The file must be sent to [cristina.maier@umb.edu](mailto:cristina.maier@umb.edu) from your umb email address. **Subject of the email must be exactly [CS430][HW1] or [CS630][HW1] .**

Up to three submissions will be accepted, given that they are received before the due date. If more submissions, only the last one will be considered and graded. If more than three submissions, the third one will be considered.

Homework **must be typed** (not handwritten). A handwritten homework won't be accepted. No submission after the due date will be accepted.

**The exercises starting with [CS630 only] are only for CS630 students** (will not be graded for CS430 students). **All the other exercises are for both CS430 and CS630 students.**

## Question 1 (30 points)

Consider a database schema with three relations:

Books(bid:integer, bname:string, author:string, pubyear:integer, pubcompany:string)

Students(sid:integer, sname:string, age:real, state:string)

Reads(sid:integer, bid:integer, year:integer)

(The primary keys are underlined in each relation. A book is uniquely identified by bid. A student is uniquely identified by sid. If a student reads a book, a record will be present in the Reads relation, with that sid and bid and the year the book was read.)

and the following relations instances:

Students

sid	sname	age	state
20	mary	21	MA
10	anne	20	NY
30	joe	21	MA
40	mary	21	VT
60	linda	23	MA

Books

bid	bname	author	pubyear	pubcompany
102	ulysses	joyce	1920	simon
101	lord of the rings	tolkien	1954	allen
103	other book	joyce	1920	penguin

Reads

sid	bid	year
20	101	2020
20	102	2021
30	103	2020

Notes:

- For CS430 students, each problem (a through e) carries 6 points possible.
- For CS630 students, each problem (a through f) carries 5 points possible.

Using the relation instances from above, show the resulted relation for each of the following relational algebra expressions:

- a)  $\sigma_{author='joyce'} Books$
- b)  $\pi_{author, pubyear}(\sigma_{author='joyce'} Books)$
- c)  $(\sigma_{sname='mary'} Students) \bowtie Reads$
- d)  $(\sigma_{state='MA'} Students) \times (\sigma_{year=2020} Reads)$
- e)  $\rho(A(bname \rightarrow name), \sigma_{(pubcompany='simon') \vee (pubcompany='allen')} Books)$
- f) [CS630 only]  $Students \bowtie Reads \bowtie (\sigma_{pubyear < 1950} Books)$

**Note:** for each resulted relation, you must show both the header with the column names, in the expected order, as well as the records.

## Question 2 (40 points)

Consider a database schema with three relations:

Actors (aid:integer, aname:string, age:real, city:string, state:string)

Playsin(aid:integer, mid:integer, character:string)

Movies(mid:integer, mname:string, year:integer, studio:string)

Primary keys are underlined in each relation. Attribute aid uniquely identifies an actor in Actors relation. An actor has an id (aid), a name (attr. aname), an age (attr. age), and a city and state (attributes city and state). Attribute mid uniquely identifies a movie in relation Movies. A movie has an id (mid), a name (attr. mname), a year (attr. year) and a studio that produced it (attr. studio). Relation Playsin contains information about actors who played in movies. Attribute character = the character name of the actor with aid when playing in mid.

Notes:

- For CS430 students, each problem (a through h) carries 5 points possible.
- For CS630 students, each problem (a through j) carries 4 points possible.

Write relational algebra queries for the following queries:

- a) Find the information about movies produced by 'WB' or 'Universal' studios.
- b) Find the names of actors who are older than 25 and are from state VT.
- c) Find the names and ages of the actors who played only in movies only in 2015.
- d) Find the names, age and city of actors who are from Boston MA and played some movies produced by 'Universal' studio.
- e) Find the name and age of the actors who played in movies both in 2012 and 2018.

- f) Find the names of the actors older than 30 who played in a movie produced by 'WB' studio in 2018.
- g) Find the information about actors and movies they played in. The result should contain the name and age of actors and the name of movies.
- h) Find the names and ages of actors from MA who played as character 'Batman'.
- i) [CS630 only] Find the name and age of actors who played in movies produced by 'Paramount' (in any year) and never played in any movie produced by 'WB' in year 2020.
- j) [CS630 only] Find names of movies in which actors from both MA and NY states played.

### Question 3 (30 points)

Consider a database schema with three relations:

Books(bid:integer, bname:string, author:string, pubyear:integer, pubcompany:string)

Students(sid:integer, sname:string, age:real, state:string)

Reads(sid:integer, bid:integer, year:integer)

Primary keys are underlined in each relation. A book is uniquely identified by bid. A book has an id (bid), a name (bname), one author (attribute author), a publication year (pubyear), and a publishing company (pubcompany). A student is uniquely identified by sid. A student has an id (sid), a name (attr. sname), age (attr. age) and a state (attr. state). If a student reads a book, a record will be present in the Reads relation, with that sid and bid and the year the book was read.

Notes:

- For CS430 students, each problem (a through e) carries 6 points possible.
- For CS630 students, each problem (a through f) carries 5 points possible.

Write the relational algebra expressions for the following queries:

- a) Find the information about the youngest students.
- b) Find the information about the books that are either published in 2010 or 2020.
- c) Find the names, pub year and pub company of the oldest books.
- d) Find the names of the students from MA who read some books both in 2015 and 2018.
- e) Find the names of the books that were read by all students.
- f) [CS630 only] Find the names, pubyear and pubcompany of the books that were read by all students from MA.