CS430 Homework 1

Brendan Nguyen

Due: Wednesday, Feb. 15, 2023

Question 1 (30 points)

Consider a database schema with three relations:

- *Books* (`bid`: int, `bname`: string, `author`: string, `pubyear`: int, `pubcompany`: string)
- *Students* (`sid`: int, `sname`: string, `age`: real, `state`: string)
- *Reads* (`sid`: int, `bid`: int, `year`: int)

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. The following relations instances are also given:

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	alien
103	other book	joyce	1920	penguin

Reads

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

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

a) $\sigma_{author='joyce'}Books$

$\sigma_{author='joyce'}Books$

bid	bname	author	pubyear	pubcompany
102	ulysses	joyce	1920	simon
103	other book	joyce	1920	penguin

b) $\pi_{author,pubyear}(\sigma_{author='joyce'}Books)$

$\pi_{author,pubyear}(\sigma_{author='joyce'}Books)$

author	pubyear
ulysses	joyce
other book	joyce

c) $(\sigma_{sname='mary'}Students) \bowtie Reads$

$(\sigma_{sname='mary'}Students) \bowtie Reads$

sid	sname	age	state	bid	year
20	mary	21	MA	101	2020
20	mary	21	VT	102	2021

d) $(\sigma_{state='MA'}Students) \times (\sigma_{year=2020}Reads)$

$(\sigma_{state='MA'}Students) \times (\sigma_{year=2020}Reads)$

(sid)	sname	age	state	(sid)	bid	year
20	mary	21	MA	20	101	2020
20	mary	21	MA	30	103	2020
30	joe	21	MA	20	101	2020
30	joe	21	MA	30	103	2020
60	linda	23	MA	20	101	2020
60	linda	23	MA	30	103	2020

e) $\rho(A(bname \rightarrow name), \sigma_{(pubcompany='simon')\vee(pubcompany='alien')}Books)$

$\rho(A(bname \rightarrow name), \sigma_{(pubcompany='simon')\vee(pubcompany='alien')}Books)$

bid	name	author	pubyear	pubcompany
102	ulysses	joyce	1920	simon
101	lord of the rings	tolkien	1954	alien

f) [CS630 only] $Students \bowtie Reads \bowtie (\sigma_{pubyear<1950}Books)$

[CS630 only] $Students \bowtie Reads \bowtie (\sigma_{pubyear<1950}Books)$

sid	sname	age	state	bid	year	bname	author	pubyear	pubcompany
30	joe	21	MA	103	2020	other book	joyce	1920	penguin
40	mary	21	VT	102	2021	ulysses	joyce	1920	simon

Question 2 (40 points)

Consider a database schema with three relations:

- *Actors* (aid: int, aname: string, age: real, city: string, state: string)
- *Playsin* (aid: int, mid: int, character: string)
- *Movies* (mid: int, mname: string, year: int, 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 is the name of the character played by the actor with aid when playing in mid.

Write relational algebra queries for the following queries:

a) Find the information about movies produced by 'WB' or 'Universal' studios

$\sigma_{(studio='WB')\vee(studio='Universal')}Movies$

b) Find the names of actors who are older than 25 and are from state VT.

$\pi_{aname}(\sigma_{(age>25)\wedge(state='VT')}Actors)$

c) Find the names and ages of the actors who played only in movies only in 2015.

$\pi_{aname,age}(Actors \bowtie Playsin \bowtie (\sigma_{year=2015}Movies))$

d) Find the names, age and city of actors who are from Boston MA and played some movies produced by 'Universal' studio.

$\pi_{aname,age,city}((\sigma_{(city='Boston')\wedge(state='MA')}Actors) \bowtie Playsin \bowtie (\sigma_{studio='Universal'}Movies))$

e) Find the name and age of the actors who played in movies both in 2012 and 2018.

$\pi_{aname,age}((\sigma_{year=2012}Movies \bowtie Playsin) \cap (\sigma_{year=2018}Movies \bowtie Playsin)) \bowtie Actors$

f) Find the names of the actors older than 30 who played in a movie produced by 'WB' studio in 2018.

$\pi_{aname}((\sigma_{age>30}Actors) \bowtie (Playsin \bowtie (\sigma_{(studio='WB')\wedge(year=2018)}Movies)))$

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.

$\pi_{aname,age,mname}(Actors \bowtie Playsin \bowtie Movies)$

h) Find the names and ages of actors from MA who played as character 'Batman'.

$\pi_{aname,age}(\sigma_{(city='MA')\wedge(character='Batman')}(Actors \bowtie Playsin))$

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: int, bname: string, author: string, pubyear: int, pubcompany: string)
- *Students* (sid: int, sname: string, age: real, state: string)
- *Reads* (sid: int, bid: int, year: int)

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.

Write the relational algebra expressions for the following queries:

a) Find the information about the youngest students.

$\rho(S1, \textit{Students})$
$\rho(S2, \textit{Students})$
$\rho(\textit{Temp}(1 \to \textit{fsid}, 2 \to \textit{fsname}, 3 \to \textit{fage}, 4 \to \textit{fstate}), S1 \bowtie_{\textit{S1.age} > \textit{S2.age}} S2)$
$\rho(\textit{TempLeft}, \pi_{\textit{fsid,fsname,fage,fstate}} \textit{Temp})$
$\textit{Students} - \textit{TempLeft}$

b) Find the information about the books that are either published in 2010 or 2020.

$\sigma_{(\textit{year=2010}) \wedge (\textit{year=2020})} \textit{Books}$

c) Find the names, pub year and pub company of the oldest books.

$\rho(B1, \textit{Books})$
$\rho(B2, \textit{Books})$
$\rho(\textit{Temp}(1 \to \textit{fbid}, 2 \to \textit{fbname}, 3 \to \textit{fauthor}, 4 \to \textit{fpubyear}, 5 \to \textit{fpubcompany}), B1 \bowtie_{\textit{B1.pubyear} > \textit{B2.pubyear}} B2)$
$\rho(\textit{TempLeft}, \pi_{\textit{fbname,fpubyear,fpubcompany}} \textit{Temp})$
$\textit{Books} - \textit{TempLeft}$

d) Find the names of the students from MA who read some books both in 2015 and 2018.

$\pi_{\textit{sname}}(((\sigma_{\textit{year=2015}} \textit{Reads}) \bowtie (\sigma_{\textit{state='MA'}} \textit{Students})) \cap ((\sigma_{\textit{year=2018}} \textit{Reads}) \bowtie (\sigma_{\textit{state='MA'}} \textit{Students})))$

e) Find the names of the books that were read by all students.

$\pi_{\textit{bname}}((\pi_{\textit{bid,sid}} \textit{Reads} / \pi_{\textit{sid}} \textit{Students}) \bowtie \textit{Books})$

f) [CS630 only] Find the names, pubyear and pubcompany of the books that were read by all students from MA.