

Name

CWID

Homework Assignment 1

**Due Date: Friday, Sept 13, 2019
(11:59 PM CDT)**

CS425 - Database Organization Results

Please leave this empty!

1.1	<input type="text"/>	1.2	<input type="text"/>	1.3	<input type="text"/>	1.4	<input type="text"/>	1.5	<input type="text"/>	1.6	<input type="text"/>
-----	----------------------	-----	----------------------	-----	----------------------	-----	----------------------	-----	----------------------	-----	----------------------

1.7	<input type="text"/>	1.8	<input type="text"/>	1.9	<input type="text"/>	1.10	<input type="text"/>	1.11	<input type="text"/>	1.12	<input type="text"/>	Sum	<input type="text"/>
-----	----------------------	-----	----------------------	-----	----------------------	------	----------------------	------	----------------------	------	----------------------	-----	----------------------

Instructions

- Try to answer all the questions using what you have learned in the class.
- Please write queries in *Relational Algebra*. Do not write SQL queries in this assignment!
- When writing a query, write the query in a way that it would work over all possible database instances (schema) and not just for the given example instance! In other words, queries can be created without the given example instance.

Consider the following database schema and example instance:

Athlete

<u>aid</u>	name	weight	age
01	Phil	250	39
02	Jay	240	46
03	Simeon	190	33
04	Jeremy	220	28
05	Jeff	195	25
...

Competition

<u>cid</u>	title	region	credit
MOL-US	Mr. Olympia	US	3
MOL-IL	Mr. Olympia	Illinois	2
MOL-CHI	Mr. Olympia	Chicago	1
MOL-NY	Mr. Olympia	New York	2
MOL-NYC	Mr. Olympia	New York City	1
ARC-US	Arnold Classic	US	3
ARC-IL	Arnold Classic	Illinois	2
ARC-CHI	Arnold Classic	Chicago	1
ARC-NY	Arnold Classic	New York	2
ARC-NYC	Arnold Classic	New York City	1
...

Result

<u>cid</u>	<u>aid</u>	rank	score
MOL-US	01	1st	20
MOL-US	02	2nd	10
MOL-US	04	3rd	5
MOL-IL	02	1st	20
MOL-IL	07	2nd	10
MOL-IL	09	3rd	5
ARC-US	01	1st	20
ARC-US	04	2nd	10
ARC-US	08	3rd	5
ARC-IL	02	1st	20
...

Subcompetition

<u>cid</u>	<u>sid</u>
MOL-US	MOL-IL
MOL-US	MOL-NY
MOL-IL	MOL-CHI
MOL-NY	MOL-NYC
...	...

Hints:

- This database is designed for storing the information of bodybuilding competitions. The data types of attributes *weight (lbs)*, *age*, *credit*, and *score* are defined as *Number*. The data types of the remaining attributes are defined as *String*. Attribute *title* refers to the title of the championship in the competition.
- Each underlined attribute is the primary key of a relation.
- The attributes *cid* and *aid* of relation *Result* is a foreign key to relations *Competition* and *Athlete*, respectively. Both attributes *cid* and *sid* are foreign keys (two different foreign keys) to relation *Competition* (which records the information of *direct sub-competitions*). For instance, 'MOL-CHI' is the direct sub-competition of 'MOL-IL' and 'MOL-IL' is the direct subcompetition of 'MOL-US'.
- Attribute *score* corresponds to attribute *rank* (e.g., an athlete receives score 20 if he is ranked as '1st' in the competition). There exists some redundancy in relation *Result* (please skip this in the queries).

Part 1.1 Relational Algebra (Total: 100 Points)

Question 1.1.1 (6 Points)

List the names of all the athletes who participate in the competition 'MOL-IL'.

Solution

$$\pi_{name}(\sigma_{cid='MOL-IL'}(Athelete \bowtie Result))$$

Question 1.1.2 (6 Points)

List the query result as “title, region, name, rank” where the athlete weight is greater than 230lbs.

Solution

$$\pi_{title,region,name,rank}(\sigma_{weight>230}(Competition \bowtie Result \bowtie Athlete))$$

Question 1.1.3 (8 Points)

List the athletes (result: aid, name) who participate in the direct subcompetitions of 'ARC-IL' and got a rank '2nd'. (Hint: only the direct subcompetition of 'ARC-IL')

Solution

$$\pi_{aid,name}(\sigma_{cid='ARC-IL' \wedge rank='2nd'}(Result \bowtie_{Result.cid=Subcompetition.sid} Subcompetition \bowtie Athlete))$$

Question 1.1.4 (10 Points)

List the athletes (result: aid, name) who have participated in all the competitions in the 'US'. (Hint: only national competitions with region 'US', not including subcompetitions)

Solution

$$\begin{aligned} E1 &\leftarrow \pi_{aid,name,cid}(Result \bowtie Athlete) \\ E2 &\leftarrow \pi_{cid}(\sigma_{region='US'} Result \bowtie Competition) \\ E &\leftarrow E1 \div E2 \end{aligned}$$

Question 1.1.5 (10 Points)

List the athletes (result: aid, name) whose score in 'MOL-US' is higher than his score in 'ARC-US'. (Hint: only count the athletes who have participated in both 'MOL-US' and 'ARC-US')

Solution

$$E1 \leftarrow \sigma_{cid='MOL-US'}(Result)$$
$$E2 \leftarrow \sigma_{cid='ARC-US'}(Result)$$
$$E \leftarrow \pi_{E1.aid, E1.name}(\sigma_{(E1.aid=E2.aid) \wedge (E1.score > E2.score)}(E1 \times E2))$$

Question 1.1.6 (8 Points)

List the athletes (result: aid, name) who never got a rank lower than '2nd' in any competition he has participated in. (Hint: rank lower than '2nd' is equivalent to "score less than 10")

Solution

$$\begin{aligned} S' &\leftarrow \pi_{aid}(Athlete) - \pi_{aid}(\sigma_{score < 10}(Result)) \\ S &\leftarrow \pi_{aid, name}(S' \bowtie Athlete) \end{aligned}$$

Question 1.1.7 (8 Points)

List all the 'US' competitions 'Jay' has not participated in (result: title, region). (Hint: only national competitions with region 'US', not including subcompetitions)

Solution

$$\pi_{title, region}(\sigma_{region='US'}(Competitions)) - \pi_{title, region}(\sigma_{name='Jay'}(Athlete \bowtie Result \bowtie Competition))$$

Question 1.1.8 (8 Points)

List the athletes and their weighted average scores (result: name, weighted) in all the competitions. (Hint: weighted = (score × credit)/total credits)

Solution

```
gs ← πResult.cid,aid,(score*credit) as gs (Result ⋈ Competition)
totalgs ←aid Gsum(gs) as tgs(gp)
totalcre ←aid Gsum(scores) as tcr(Result ⋈ Competition)
result ← πtotalgs.aid,(tgs/tcr) as weighted(totalgs ⋈ totalcre)
```

Question 1.1.9 (8 Points)

List all the competitions and the counts of their direct subcompetitions (result: cid, count) in the database. (Hint: if a competition does not have any direct subcompetition, return constant 0)

Solution

```
dirsub ← Result ⋈Result.cid=Subcompetition.cid Subcompetition
nodir ← πcid(Competition) − πcid(Subcompetition)
result ←cid Gcount(*)(dirsub) ∪ πcid,0(nodir)
```

Another typical solution is to use *left outer join*.

Question 1.1.10 (6 Points)

List all the competitions where the average age of the athletes in the competition is lower than 30 (result: cid, avg_age).

Solution

$$\begin{aligned} avgA &\leftarrow_{cid} \mathcal{G}_{avg(age) \text{ as } avg_age}(Result \bowtie Athlete) \\ result &\leftarrow \pi_{cid, avg_age}(\sigma_{avg_age < 30}(avgA)) \end{aligned}$$

Question 1.1.11 (10 Points)

List all the competitions and the name(s) of the athlete(s) with the highest score (result: cid, name) in each competition. (Hint: please do not use *max* to query the highest score)

Solution

$$\begin{aligned} nonhigh &\leftarrow \pi_{Result.cid, Result.aid}(\sigma_{(Result.score < x.score) \wedge Result.cid = x.cid}(Result \times \rho_x(Result))) \\ cidaid &\leftarrow \pi_{Result.cid, Result.aid}(\pi_{cid, aid}(Result) - nonhigh) \\ result &\leftarrow \pi_{title, name}(Competition \bowtie cidaid \bowtie Athlete) \end{aligned}$$

Question 1.1.12 (12 Points)

List the athletes (result: aid, name) who have participated in 'ARC-US' and all its subcompetitions. (Hint: all the subcompetitions of 'ACR-US' include not only the direct subcompetitions but also the subcompetitions of the subcompetition, ...)

Solution

$$\begin{aligned} R &:= \sigma_{cid='ARC-US'}(Subcompetition \cup \pi_{R.cid, Subcompetition.sid}(R \bowtie_{R.sid=Subcompetition.cid} Subcompetition)) \\ allsub &\leftarrow \rho_x(cid)((\pi_{sid} R) \cup 'ARC-US') \\ E1 &\leftarrow \pi_{aid, name, cid}(Result \bowtie Athlete) \\ result &\leftarrow E1 \div allsub \end{aligned}$$