Question : Relational Algebra

Consider the relations of a database for the `2016 Olympics` as shown in Table 1. These relations record the athletes, events, and outcomes/results of the 2016 Olympic games.

athlete_id	country	name	age
A1	U.S.A.	Michael Phelps	31
A2	U.S.A.	Justin Gatlin	34
A3	U.S.A.	Ryan Lochte	32
A4	Canada	Andre De Grasse	21
A5	Jamaica	Usain Bolt	30
A6	France	Christophe Lemaitre	26
A7	Japan	Masato Sakai	24
A8	Japan	Naito Ehara	60
A9	GBR	Duncan Scott	35
A10	GBR	James Guy	32

(a) `Athletes` Table

event_id	name
E1	100m Sprint
E2	200m Sprint
E3	200m Butterfly
E4	4x200 Freestyle Relay

(b) `Events` Table

event_id	athlete_id	result
E1	A5	Gold
E1	A2	Silver
E1	A4	Bronze
E2	A5	Gold
E2	A4	Silver
E3	A1	Gold
E3	A7	Silver
E3	A9	Bronze
E4	A1	Gold
E4	A3	Gold
E4	A7	Silver
E4	A8	Silver
E4	A9	Bronze
E4	A10	Bronze

(c) `Event_Results` Table: the outcome of every event

Table 1: Relations of the `Olympic` database.

We have the following tables:

- **Athletes**: For every athlete, we record a unique **athlete_id**, the country they represent, their name, and their age.
- **Events**: This table lists all the events that are part of the 2016 Olympic games. Every event has a unique integer **event_id** and a name.
- **Event_Results**: Lists the outcomes of all events. Every outcome records the **event_id** of the event, the **athlete_id** of the athlete that won a medal in the event, (or was part of the team that won), and the standing of the athlete (i.e., gold, silver or bronze).

Given this database instance, answer the following questions:

(a) [**2 points**] Which of the following is the meaning of the expression:

$\sigma_{\texttt{age} \geq 25}(\texttt{Athletes})$

1. It lists the **athlete_id** and **name** of all athletes that are at least 25 years old.
2. It lists all attributes of all athletes whose **age** is greater than or equal to 25.
3. It lists all athlete **age**s that are greater than or equal to 25 .
4. None of the above. The real answer is ....

Solution: #2

(b) [**5 points**] We want to list the **name**s of the athletes that have won at-least one gold medal (eliminating duplicate names) . Which, if any, of the following expressions achieve that ? Mark **all** valid expressions.

1. $\pi_{\texttt{name}}\left(\sigma_{\texttt{result}='Gold'}(\texttt{Athletes} \bowtie \texttt{Event\_Results})\right)$
2. $\pi_{\texttt{name}}\left(\texttt{Athletes} \bowtie \sigma_{\texttt{result}='Gold'}(\texttt{Event\_Results})\right)$
3. $\pi_{\texttt{name}}\left(\sigma_{\texttt{result}='Gold'}\left(\texttt{Athletes} \bowtie \pi_{\texttt{athlete\_id,result}}(\texttt{Event\_Results})\right)\right)$
4. $\pi_{\texttt{name}}(\texttt{Athletes}) \bowtie \sigma_{\texttt{result}='Gold'}(\texttt{Event\_Results})$
5. $\pi_{\texttt{name}}(\texttt{Athletes}) - \pi_{\texttt{name}}\left(\texttt{Athletes} \bowtie \sigma_{\texttt{result}\neq'Gold'}(\texttt{Event\_Results})\right)$

Solution: #1, #2 , and #3
Grading info:
  - *+1 point for each expression correctly listed (or not listed)*

(c) For the following expression:

$$\sigma_{\texttt{age}<25}(\texttt{Athletes} \bowtie \texttt{Event\_Results})$$

i. [**0 points**] *Optional*: describe in English what the expression does

Solution: List **athlete_id**, **country**, **name**, **age**, **event_id**, and **result** of athletes younger than 25 that have won medals.

ii. **[1 point]** How many, and which are the columns (= attributes) in the answer?

Solution: Six columns: `athlete_id`, `event_id`, `country`, `name`, `age`, and `result`.

iii. **[3 points]** How many tuples are in the answer?

Solution: 4

iv. **[3 points]** List all the tuples in the answer, as a table.

Solution:

athlete_id	country	name	age	event_id	result
A4	Canada	Andre De Grasse	21	E1	Bronze
A4	Canada	Andre De Grasse	21	E2	Silver
A7	Japan	Masato Sakai	24	E3	Silver
A7	Japan	Masato Sakai	24	E4	Silver

Grading info:
- *ii. -1 for missing columns or having extra columns*
- *iii. -1 for being off by one on number of tuples; -3 for getting the wrong number of tuples by more than one*
- *iv. -1 for missing one tuple; -2 for performing wrong operation (e.g. × instead of ⋈) or getting some columns and tuples wrong; -3 for all tuples and columns wrong*

(d) For the following expression:

$$\pi_{\texttt{athlete\_id,event\_id}}\big(\texttt{Event\_Results}\big) \div \pi_{\texttt{event\_id}}\big(\sigma_{\texttt{athlete\_id='A5'}}\big(\texttt{Event\_Results}\big)\big)$$

i. **[0 points]** *Optional*: describe in English what the expression does

Solution: List the `athlete_id`s of all athletes that have won medals in every event that athlete with ID "A5" (i.e., Usain Bolt) has also won a medal in.

ii. **[2 points]** How many, and which are the columns (= attributes) in the answer?

Solution: One column: `athlete_id`.

iii. **[3 points]** How many tuples are in the answer?

Solution: 2

iv. **[3 points]** List all the tuples in the answer, as a table.

Solution:

athlete_id
A4
A5

Grading info:

(e) For the following expression:

$$\pi_{\text{A.athlete\_id}}\big(\rho_A(\text{Athletes})\big) - \pi_{\text{ER1.athlete\_id}}\big($$

$$\rho_{ER1}(\text{Event\_Results}) \bowtie_{ER1.\text{athlete\_id}=ER2.\text{athlete\_id} \wedge ER1.\text{result} \neq ER2.\text{result}} \rho_{ER2}(\text{Event\_Results})\big)$$

   i. **[0 points]** *Optional*: describe in English what the expression does

   Solution: Finds all the athletes (by athlete_id) that did not win two or more types of medals.

   ii. **[2 points]** How many, and which are the columns (= attributes) in the answer?

   Solution: One column: A.`athlete_id`.

   iii. **[3 points]** How many tuples are in the answer?

   Solution: 9

   iv. **[3 points]** List all the tuples in the answer, as a table.

   Solution:

A.athlete_id
A1
A2
A3
A5
A6
A7
A8
A9
A10