Basi di dati I — Prova di autovalutazione 2 novembre 2015
Soluzioni

Domanda 1 Si consideri una base di dati sulle relazioni

- $R_1(\underline{A}, B, C)$
- $R_2(\underline{D}, \underline{E}, F)$

Scrivere interrogazioni in SQL equivalenti alle seguenti espressioni dell'algebra relazionale:

1. $\pi_{BC}(\sigma_{C>10}(R_1))$

```
SELECT DISTINCT B, C
FROM R1
WHERE C > 10
```

2. $\pi_B(R_1 \bowtie_{C=D} \sigma_{F=2}(R_2))$

```
SELECT DISTINCT B
FROM R1 JOIN R2 ON C = D
WHERE F = 2
```

Domanda 2 Con riferimento alla base di dati nella domanda 1 scrivere espressioni dell'algebra relazionale equivalenti alle seguenti interrogazioni SQL

1.
```
SELECT DISTINCT A , B
FROM R1, R2
WHERE C = D AND E > 100
```

$$\pi_{AB}(R_1 \bowtie_{C=D} \sigma_{E>100}(R_2))$$

2.
```
SELECT DISTINCT A , B
FROM R1 X1
WHERE NOT EXISTS
(SELECT *
FROM R1 Y1, R2
WHERE Y1.C = D AND X1.A = Y1.A AND F>10)
```

$$\pi_{AB}(R_1) - \pi_{AB}(R_1 \bowtie_{C=D} \sigma_{F>10}(R_2))$$

Nota bene, se la chiave di R_1 fosse stata AB, allora l'interrogazione sarebbe sata più complessa:

$$\pi_{AB}(\pi_A(R_1) - \pi_A(R_1 \bowtie_{C=D} \sigma_{F>10}(R_2))) \bowtie_{A=A'} \rho_{A' \leftarrow A}(R_1))$$

Domanda 3 Ancora con riferimento alla base di dati nella domanda 1, indicare, per ciascuna delle seguenti interrogazioni, se la parola chiave `DISTINCT` è necessaria

1. l'interrogazione 1 nella domanda 2 *Risposta*: sì
2. l'interrogazione 2 nella domanda 2 *Risposta*: no
3.
```
SELECT DISTINCT A , B
FROM R1, R2
WHERE B = D AND C = E
```
Risposta: no
4.
```
SELECT DISTINCT B , C
FROM R1, R2
WHERE B = D AND C = E
```
Risposta: sì

Domanda 4 Considerare la base di dati relazionale definita per mezzo delle seguenti istruzioni:

```
create table studenti (matricola numeric not null primary key,
                       cognome char(20) not null,
                       nome char(20) not null,
                       età  numeric);
create table corsi (codice numeric not null primary key,
                       titolo char(20) not null,
                       CFU numeric not null);
create table esami (corso numeric not null references corsi(codice),
                       studente numeric not null references studenti(matricola),
                       data date not null,
                       voto numeric not null,
                       primary key (corso, studente));
```

Supponendo che le relative relazioni abbiano rispettivamente le cardinalità $S = 10.000$ (studenti), $C = 1.000$ (corsi) e $E = 40.000$ (esami), indicare le cardinalità minime e massime (in simboli e numeri) dei risultati delle seguenti interrogazioni:

	Min (simboli)	Max (simboli)	Min (valore)	Max (valore)
`SELECT matricola, codice FROM studenti, corsi`	$S \times C$	$S \times C$	10.000.000	10.000.000
`SELECT * FROM studenti, esami WHERE matricola = studente`	E	E	40.000	40.000
`SELECT matricola, codice FROM studenti, esami, corsi WHERE matricola = studente AND corso = codice`	E	E	40.000	40.000

Domanda 5 Con riferimento alla base di dati usata nella domanda precedente formulare le seguenti interrogazioni in algebra relazionale:

1. trovare matricole e cognomi degli studenti che hanno preso almeno un trenta

$$\pi_{matricola,cognome}(\text{studenti} \bowtie_{matricola=studente} \sigma_{voto=30}(\text{esami}))$$

2. trovare le matricole degli studenti che hanno sostenuto almeno due esami

$$\pi_{studente}(\sigma_{corso \neq corso'}(\text{esami} \bowtie_{studente=studente'} \rho_{X \leftarrow X'}(\text{esami})))$$

Domanda 6 Con riferimento alla base di dati usata nelle domande precedenti, formulare le seguenti interrogazioni in SQL

1. trovare codici e titoli di corsi nei cui esami è stato assegnato almeno un trenta

```
SELECT DISTINCT codice, titolo
FROM corsi JOIN esami ON codice = corso
WHERE voto = 30
```

2. trovare le coppie di studenti (mostrare le sole matricole) per i quali uno dei due ha riportato un voto più alto di quello riportato dall'altro in tutti gli esami superati da entrambi.

```
SELECT e1.studente, e2.studente
FROM esami e1, esami e2
WHERE e1.voto > e2.voto
AND e1.studente <> e2.studente
AND e1.corso = e2.corso
AND NOT EXISTS (                  SELECT *
                                  FROM esami e3, esami e4
                                  WHERE e3.corso = e4.corso
                                  AND e3.studente = e1.studente
                                  AND e4.studente = e2.studente
                                  AND e3.voto <= e4.voto )
```

3. trovare lo studente con la media più alta; mostrare i dati dello studente, la media in questione e il numero di esami superati

```
CREATE VIEW MediaVoti AS SELECT studente, AVG(voto) AS media, COUNT(*) AS numEsami
FROM esami
GROUP BY studente

SELECT studente, media, numEsami
FROM MediaVoti, studenti
WHERE studente = matricola
AND media = (SELECT MAX(media)
                  FROM MediaVoti)
```

oppure

```
SELECT matricola, cognome, nome, AVG(voto), COUNT(*)
FROM esami join studenti on studente = matricola
GROUP BY matricola, cognome, nome
HAVING AVG(voto) >= ALL
  (SELECT AVG(voto)
  FROM esami
  GROUP BY studente)
```

Domanda 7 Con riferimento al seguente schema di base di dati:

CITTÀ(Nome, Regione, Abitanti)
ATTRAVERSAMENTI(Città, Fiume)
FIUMI(Fiume, Lunghezza)

formulare, in algebra relazionale e in SQL, le seguenti interrogazioni:

1. visualizzare nome, regione e abitanti per le città che (i) hanno più di 50.000 abitanti e (ii) sono attraversate dal Po o dall'Adige;

$$\pi_{nome,regione,abitanti}(\sigma_{abitanti>50000 \wedge (fiume="Po" \vee fiume="Adige")}(\text{Citta} \bowtie_{nome=citta} \text{Attraversamenti}))$$

```
select distinct citta.*
from citta join attraversamenti on nome = citta
where abitanti > 50000
   and (fiume ='Po' or Fiume='Adige');
```

2. trovare le città che sono attraversate da (almeno) due fiumi, visualizzando il nome della città e quello del più lungo di tali fiumi (supponendo per semplicità che nessuna città sia attraversata da più di due fiumi)

$$\text{AttrFiume} := \text{Fiume} \bowtie_{fiume=f} (\rho_{f \leftarrow fiume}(\text{Attraversamenti}))$$

$$\pi_{citta,fiume}(\sigma_{lunghezza>lunghezza'}(\text{AttrFiume} \bowtie_{citta=citta'} \rho_{X \leftarrow X'}(\text{AttrFiume})))$$

```
create view attrfiume as
   select citta, f.fiume, lunghezza
   from attraversamenti a join fiumi f on a.fiume=f.fiume;

select distinct a1.citta, a1.fiume
from   attrfiume a1 join attrfiume a2 on a1.citta =a2.citta
where  a1.fiume <> a2.fiume and
a1.lunghezza >= a2.lunghezza;
```

3. stessa interrogazione precedente, ma senza l'ipotesi semplificativa

$$\pi_{citta,fiume}(\sigma_{lunghezza>lunghezza'}(\text{AttrFiume} \bowtie_{citta=citta'} \rho_{X \leftarrow X'}(\text{AttrFiume}))) -$$

$$\pi_{citta,fiume}(\sigma_{lunghezza<lunghezza'}(\text{AttrFiume} \bowtie_{citta=citta'} \rho_{X \leftarrow X'}(\text{AttrFiume})))$$

```
select distinct a1.citta, a1.fiume
from attrfiume a1 join attrfiume a2 on a1.citta = a2.citta
where a1.fiume <> a2.fiume
except
select distinct a1.citta, a1.fiume
from attrfiume a1 join attrfiume a2 on a1.citta = a2.citta
where a1.lunghezza < a2.lunghezza
```

oppure

```
select distinct a1.citta, a1.fiume
from attrfiume a1 join attrfiume a2 on a1.citta = a2.citta
where a1.fiume <> a2.fiume
  and not exists (select *
                  from attrfiume a2
                  where a1.citta = a2.citta
                    and a2.lunghezza > a1.lunghezza);
```