

# Clustering plurirelazionale

Basi di dati II — 28 giugno 2016 — Compito A

## Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (i) memorizzazione separata delle due relazioni, entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

- (ii) memorizzazione in un cluster delle due relazioni pure entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità sempre  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

# Clustering plurirelazionale

- Vari termini, equivalenti (ma sovraccarichi)
  - cluster
  - cluster plurirelazionale
  - cluster multirelazionale
  - cluster eterogeneo
- Le ennuple delle due (o più) relazioni con gli stessi valori su un certo attributo (quello su cui è realizzato il cluster) sono memorizzate insieme

# Alcuni DBMS, conferma delle ambiguità

- Oracle (cluster multirelazionale – ciò di cui parliamo oggi)  
<https://docs.oracle.com/database/121/CNCPT/tablecls.htm#CNCPT88816>
- Postgres (cluster inteso come ordinamento, quindi un concetto diverso)  
<https://www.postgresql.org/docs/11/sql-cluster.html>

# To cluster or not to cluster (Shakespeare, Hamlet)

- Memorizzazione separata

R1	A: a1	C: c1	A: a2	C: c2	A: a3	C: c3
	A: a4	C: c4	A: a5	C: c5	A: a6	C: c6
	A: a7	C: c7	A: a8	C: c8	A: a9	C: c9
R1	A: a10	C: c10	A: a11	C: c11	A: a12	C: c12
	A: ...	C: ...	A: a...	C: c...	A: a...	C: c...
	A: a...	C: c...	A: a...	C: c...	A: a...	C: c...

# To cluster or not to cluster (Shakespeare, Hamlet)

- Memorizzazione separata

R1

A: a1	C: c1	A: a2	C: c2	A: a3	C: c3
A: a4	C: c4	A: a5	C: c5	A: a6	C: c6
A: a7	C: c7	A: a8	C: c8	A: a9	C: c9

R1

A: a10	C: c10	A: a11	C: c11	A: a12	C: c12
A: ...	C: ...	A: a...	C: c...	A: a...	C: c...
A: a...	C: c...	A: a...	C: c...	A: a...	C: c...

R2

A: a1	D: d1	E: e1	F: f1	A: ...	D: ...	E: ...	F: ...
A: ...	D: ...	E: ...	F: ...	A: ...	D: ...	E: ...	F: ...
A: ...	D: ...	E: ...	F: ...	A: ...	D: ...	E: ...	F: ...

R2

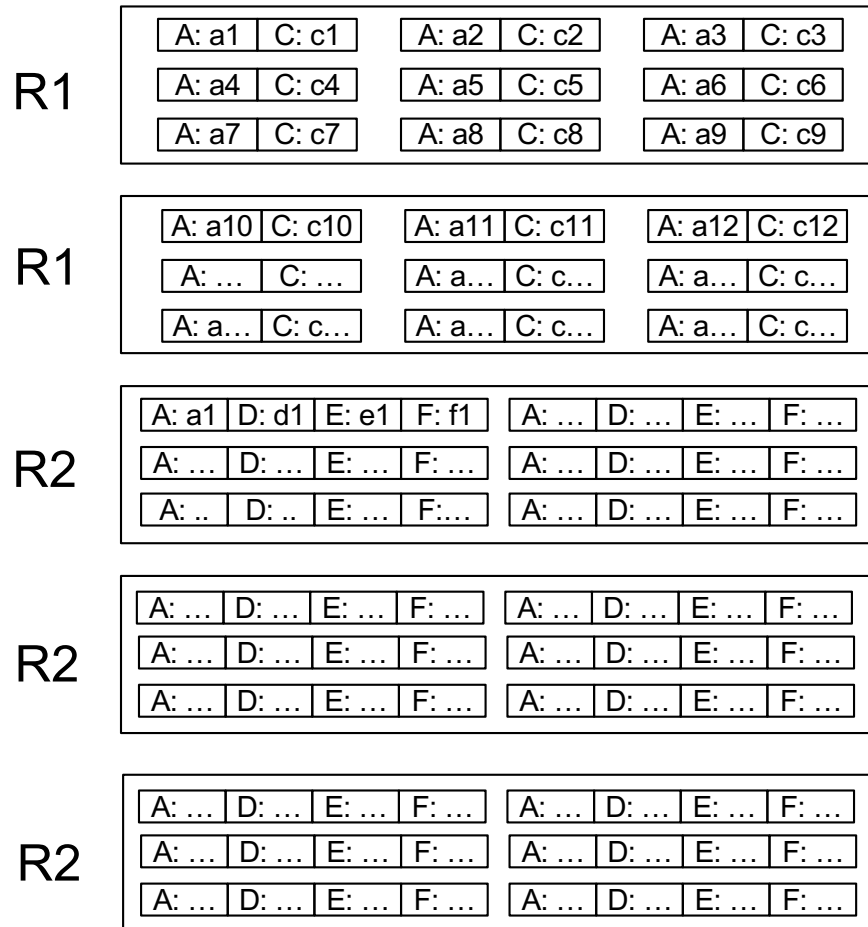
A: ...	D: ...	E: ...	F: ...	A: ...	D: ...	E: ...	F: ...
A: ...	D: ...	E: ...	F: ...	A: ...	D: ...	E: ...	F: ...
A: ...	D: ...	E: ...	F: ...	A: ...	D: ...	E: ...	F: ...

R2

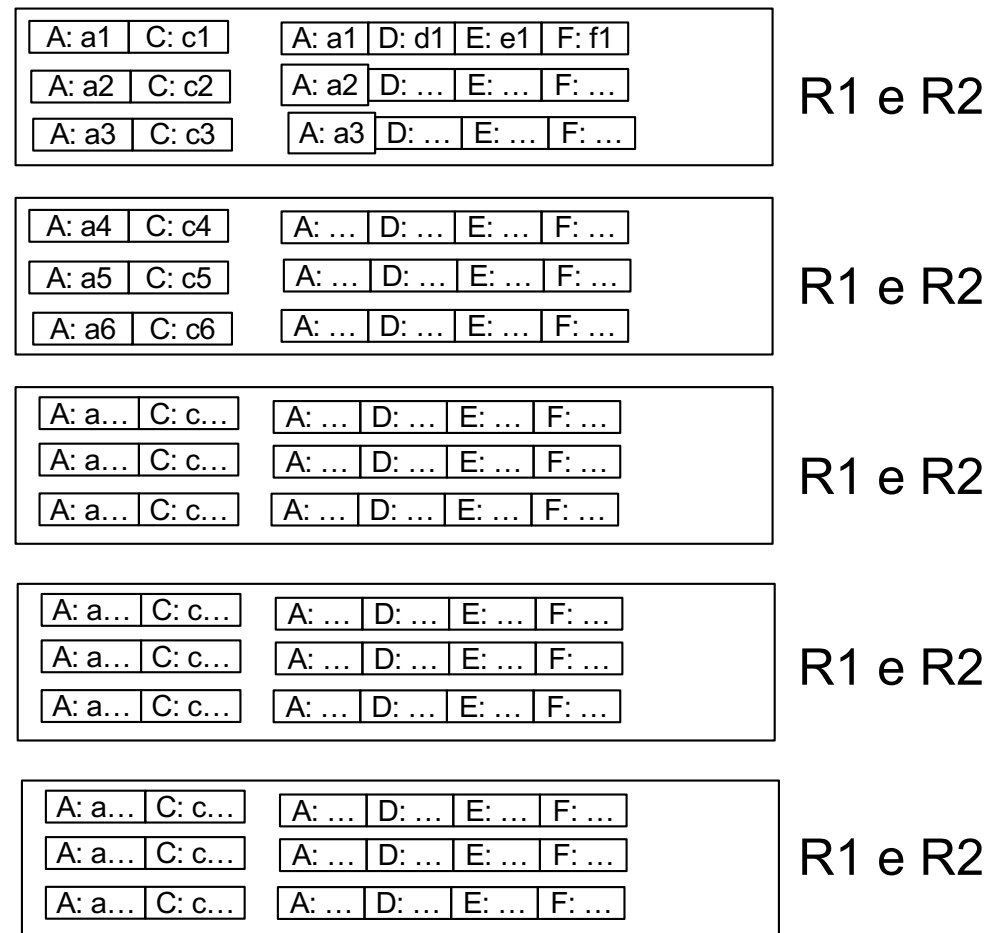
A: ...	D: ...	E: ...	F: ...	A: ...	D: ...	E: ...	F: ...
A: ...	D: ...	E: ...	F: ...	A: ...	D: ...	E: ...	F: ...
A: ...	D: ...	E: ...	F: ...	A: ...	D: ...	E: ...	F: ...

# To cluster or not to cluster (Shakespeare, Hamlet)

- Memorizzazione separata



- Memorizzazione in cluster



# Discussione

- Occupazione complessiva:
  - la stessa (nell'esempio, 5 blocchi e 5 blocchi)
- Costo delle operazioni
  - scansione di una relazione
  - join complessivo
  - selezione e join (ad esempio, una ennupla del join, supponendo che ci siano indici e troviamo subito le ennuple che contribuiscono ad essa)

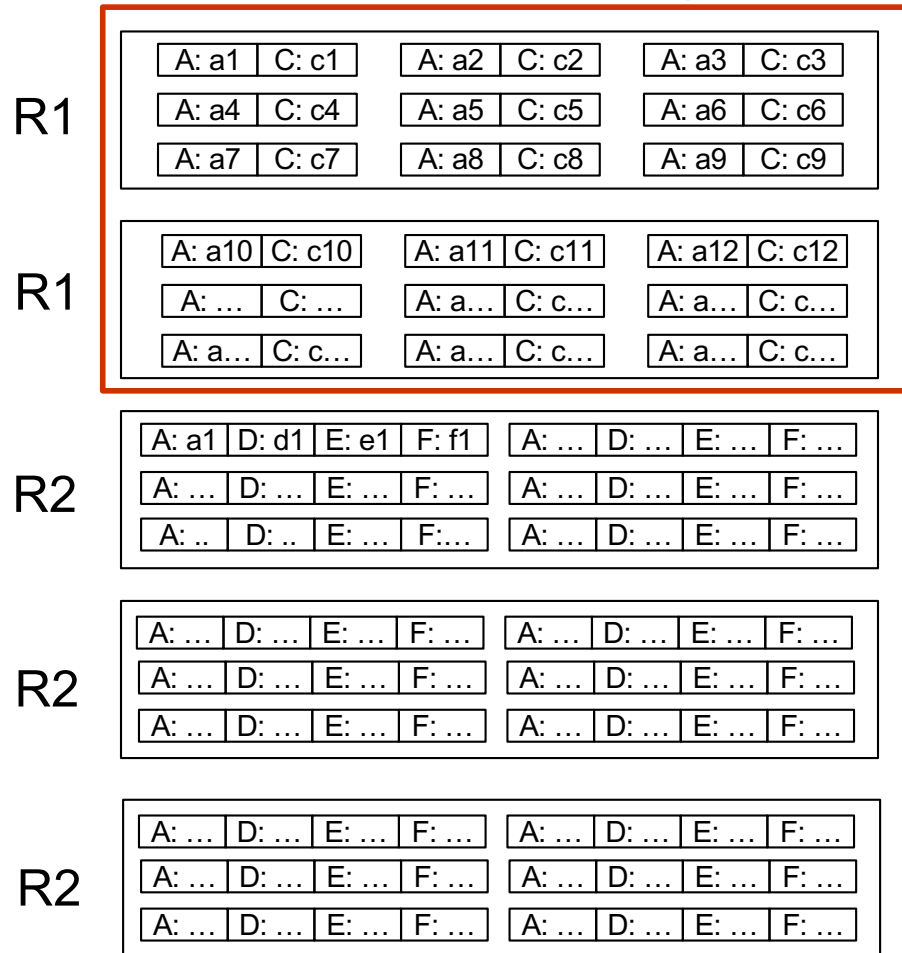
# Scansione di R1

```
SELECT *  
FROM R1
```



# To cluster or not to cluster: scansione

- Memorizzazione separata

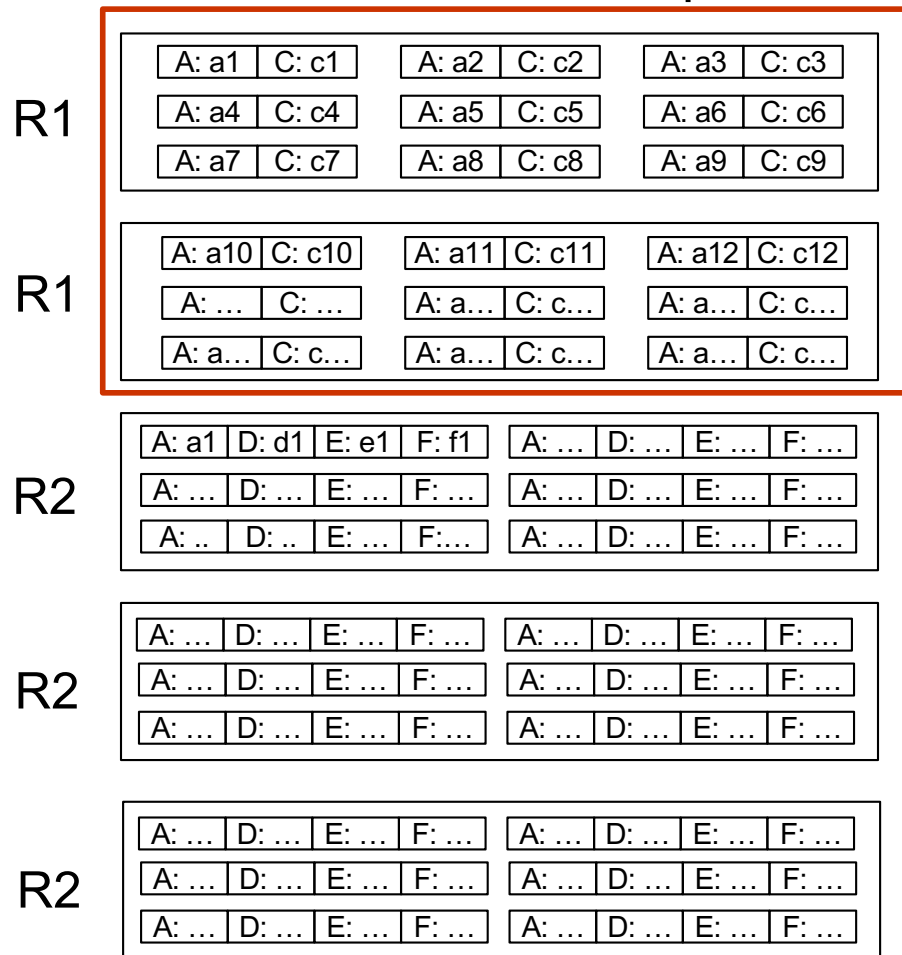


- Memorizzazione in cluster

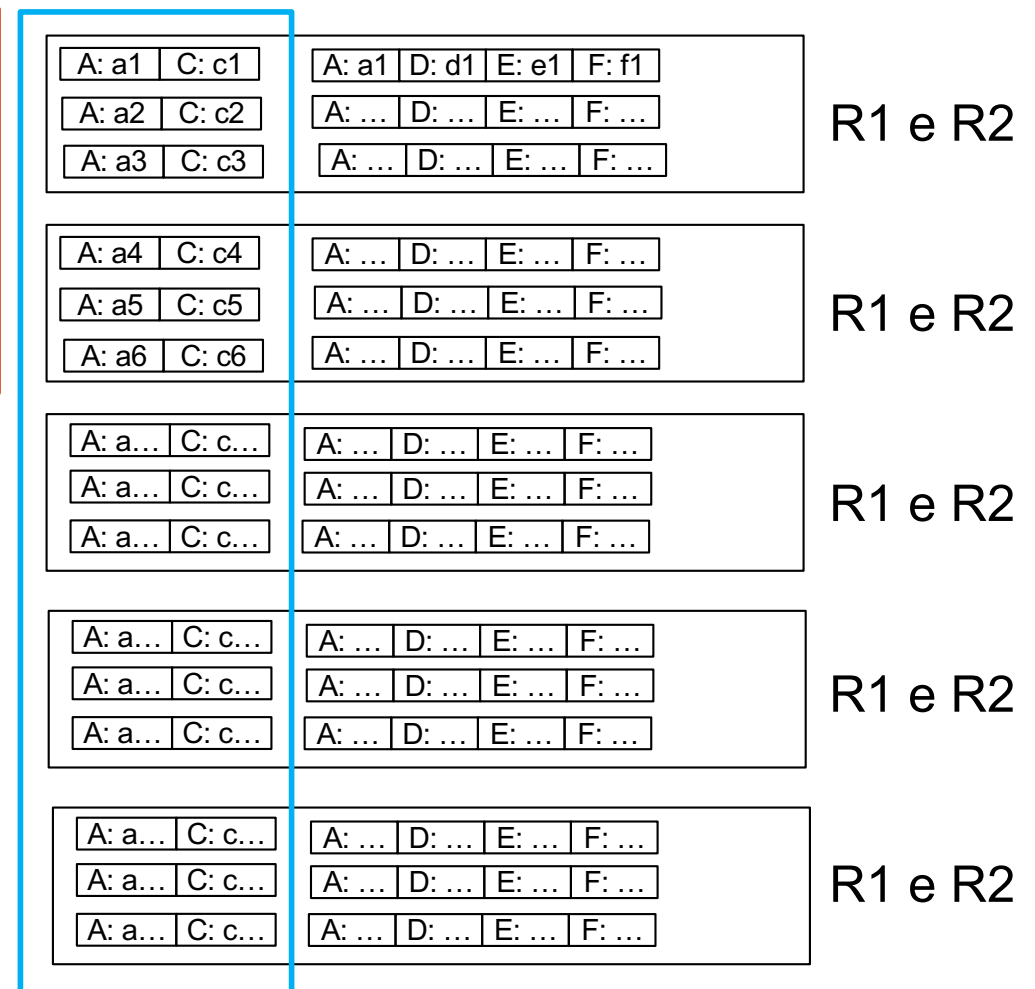


# To cluster or not to cluster: scansione

- Memorizzazione separata



- Memorizzazione in cluster



# To cluster or not to cluster: scansione

- Occupazione complessiva:
  - la stessa (nell'esempio, 5 blocchi e 5 blocchi)
- Costo delle operazioni
  - scansione di una relazione
    - memorizzazione separata: i soli blocchi della relazione
      - nell'esempio: 2
    - memorizzazione in cluster: i blocchi di entrambe
      - nell'esempio: 5

in sintesi

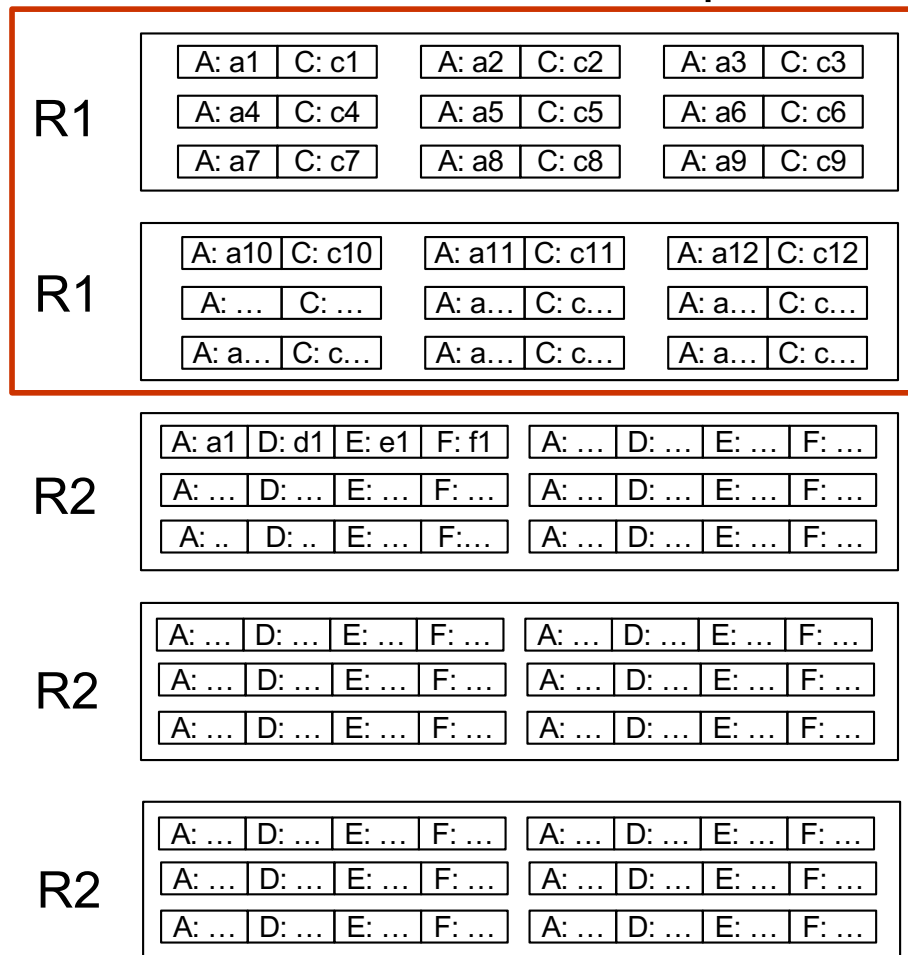
- preferibile la memorizzazione separata

# Join complessivo

```
SELECT *  
FROM R1 JOIN R2 ON R1.A = R2.A
```

# To cluster or not to cluster: join complessivo

- Memorizzazione separata

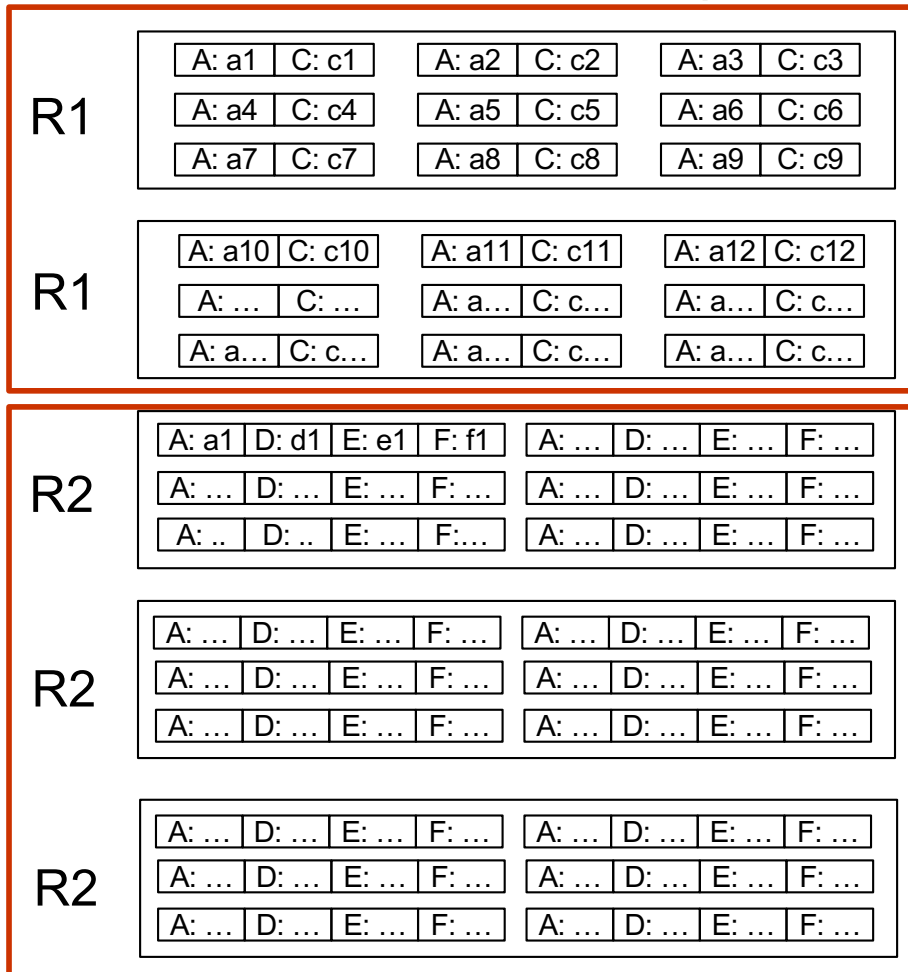


- Memorizzazione in cluster



# To cluster or not to cluster: join complessivo

- Memorizzazione separata

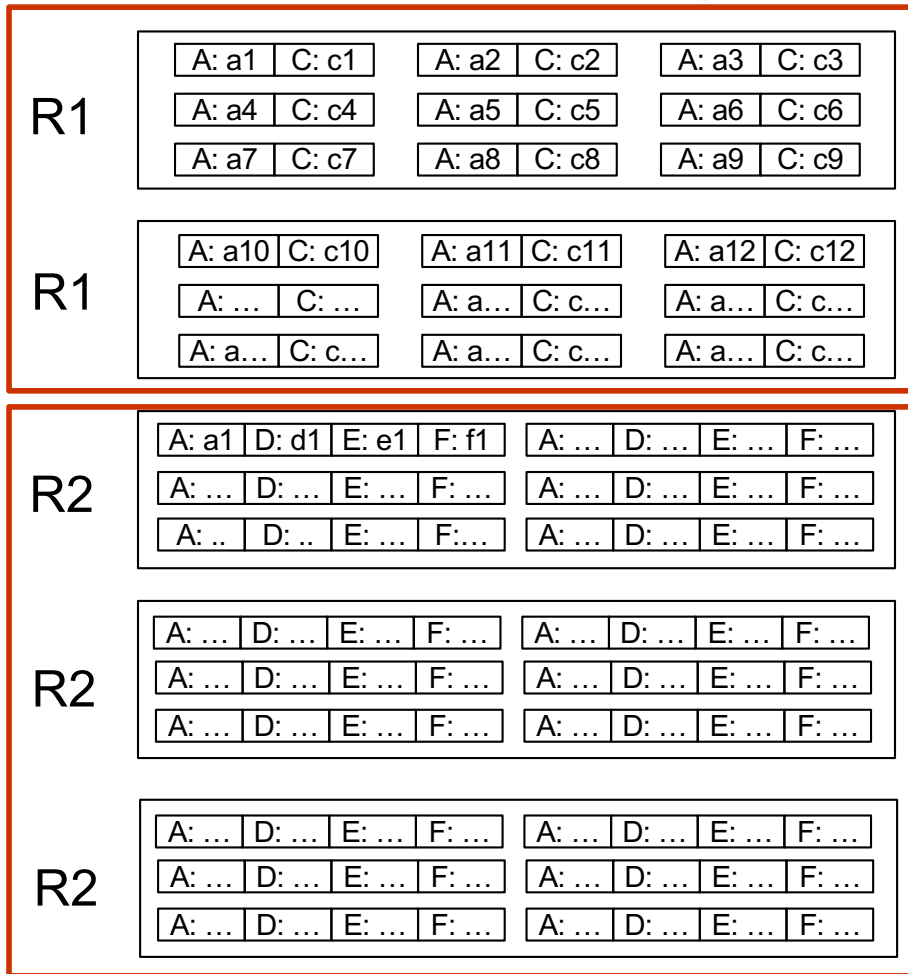


- Memorizzazione in cluster

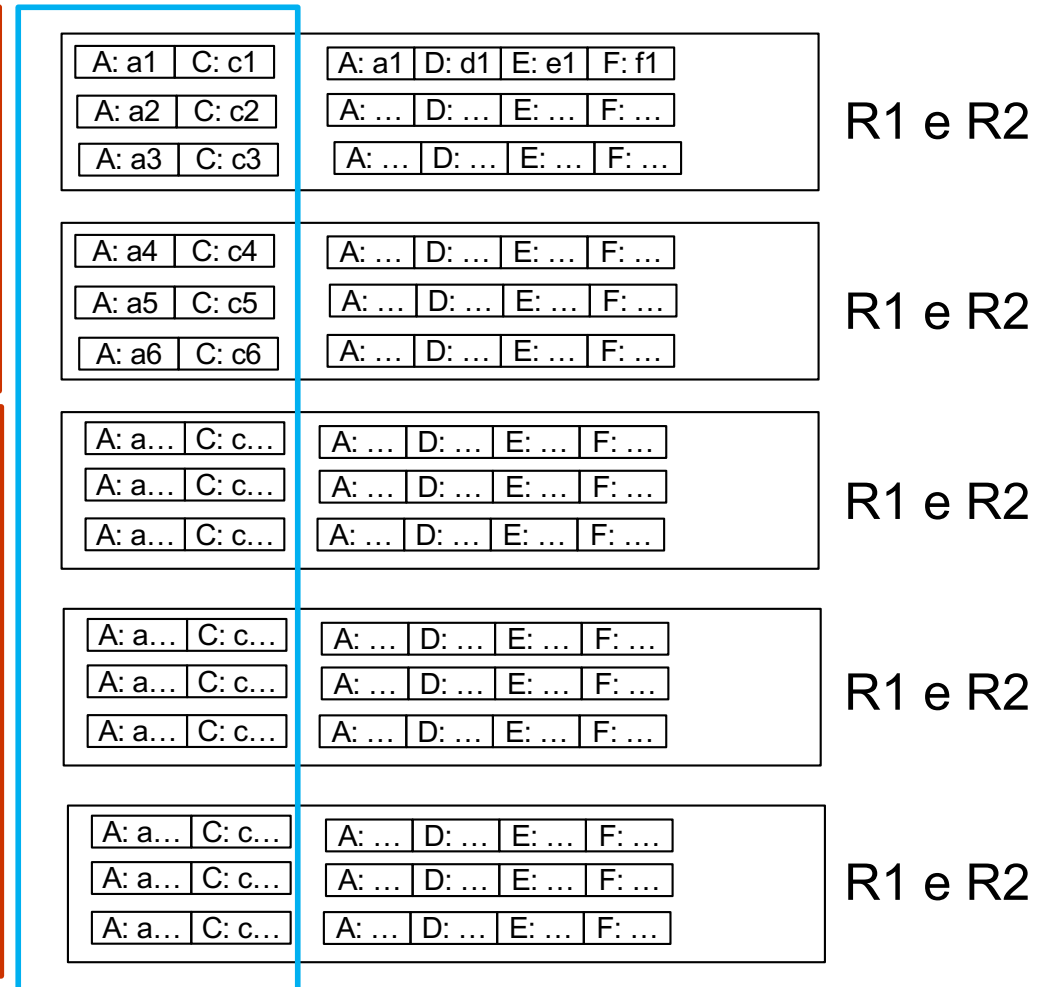


# To cluster or not to cluster: join complessivo

- Memorizzazione separata

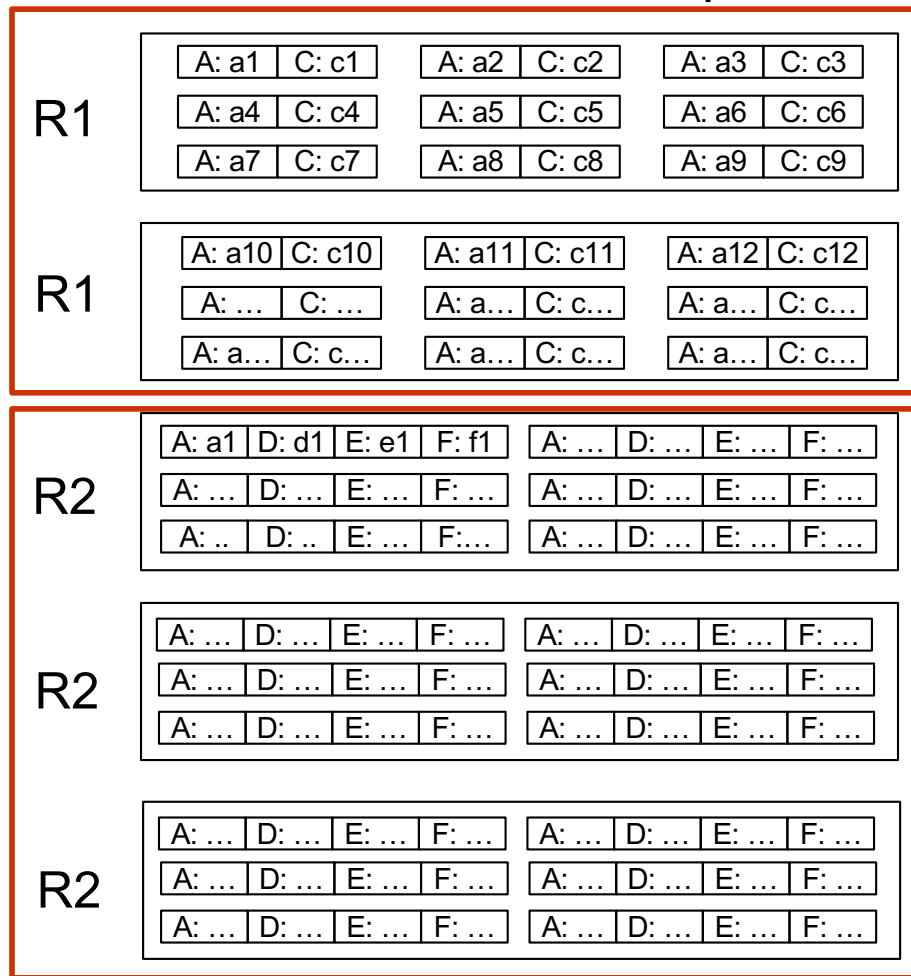


- Memorizzazione in cluster

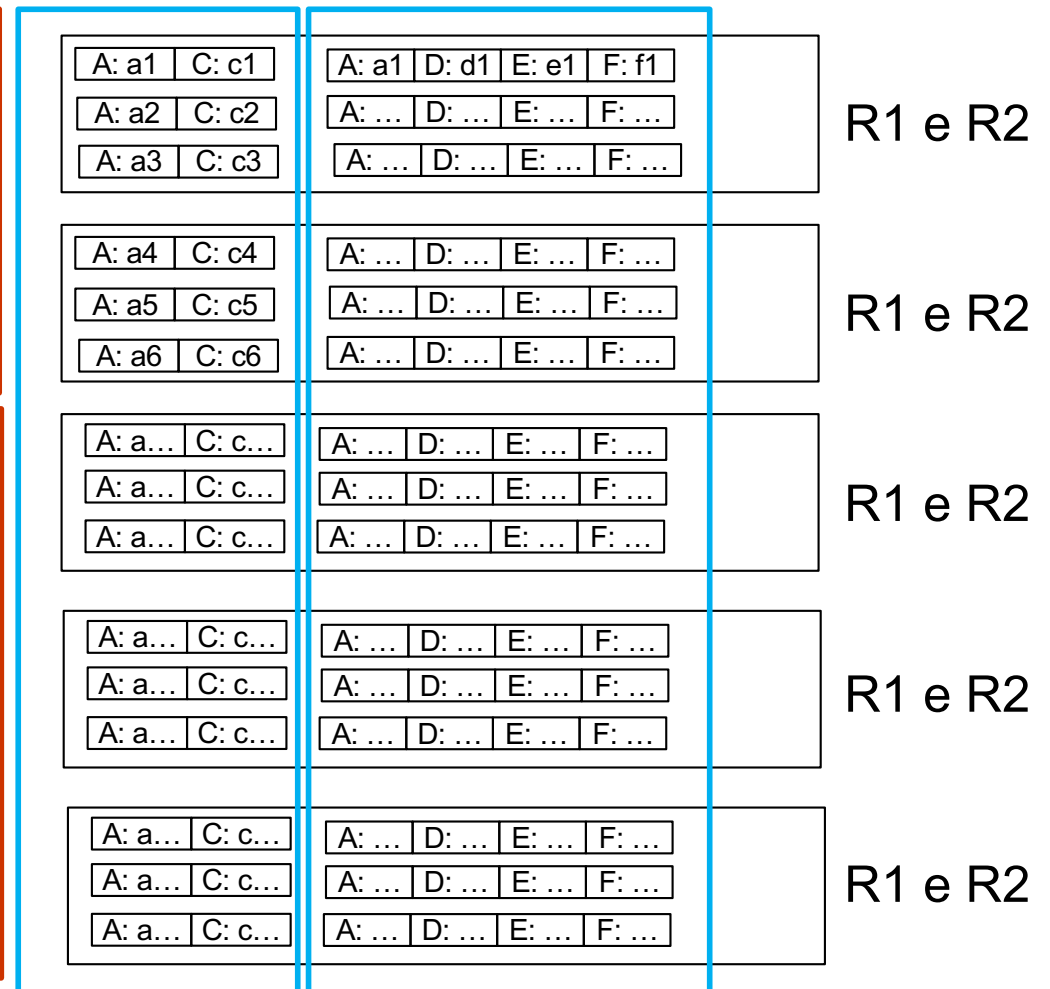


# To cluster or not to cluster: join complessivo

- Memorizzazione separata



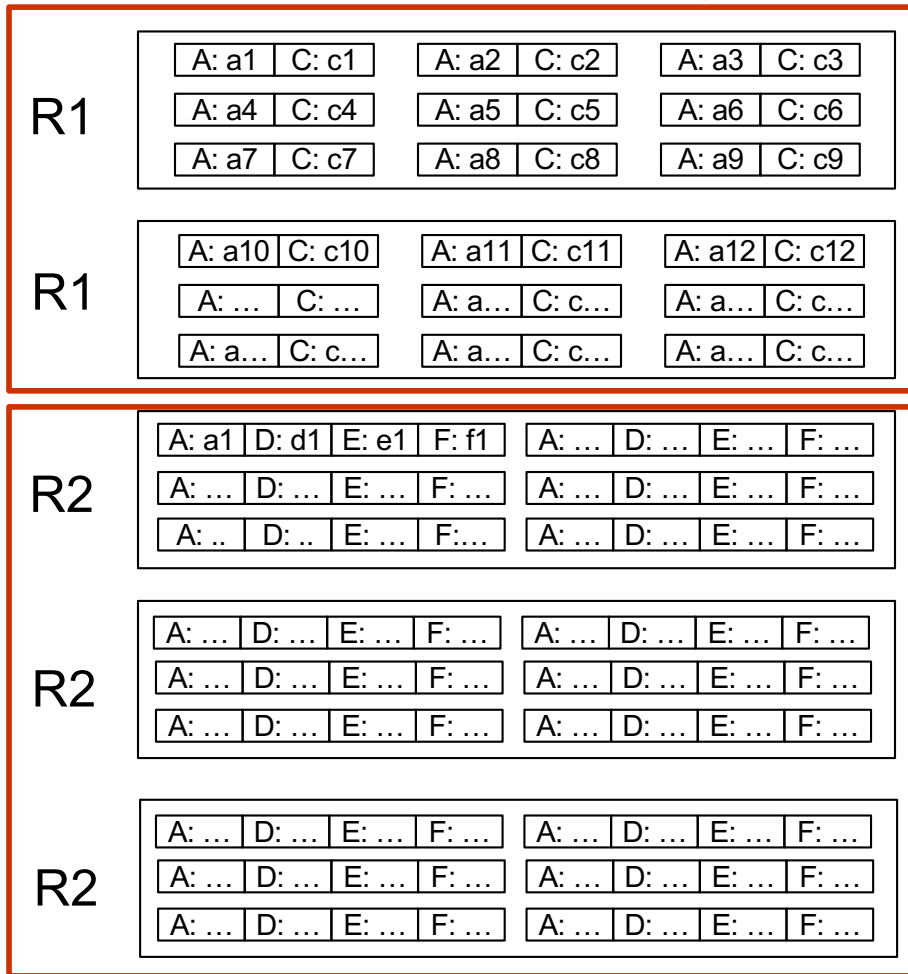
- Memorizzazione in cluster



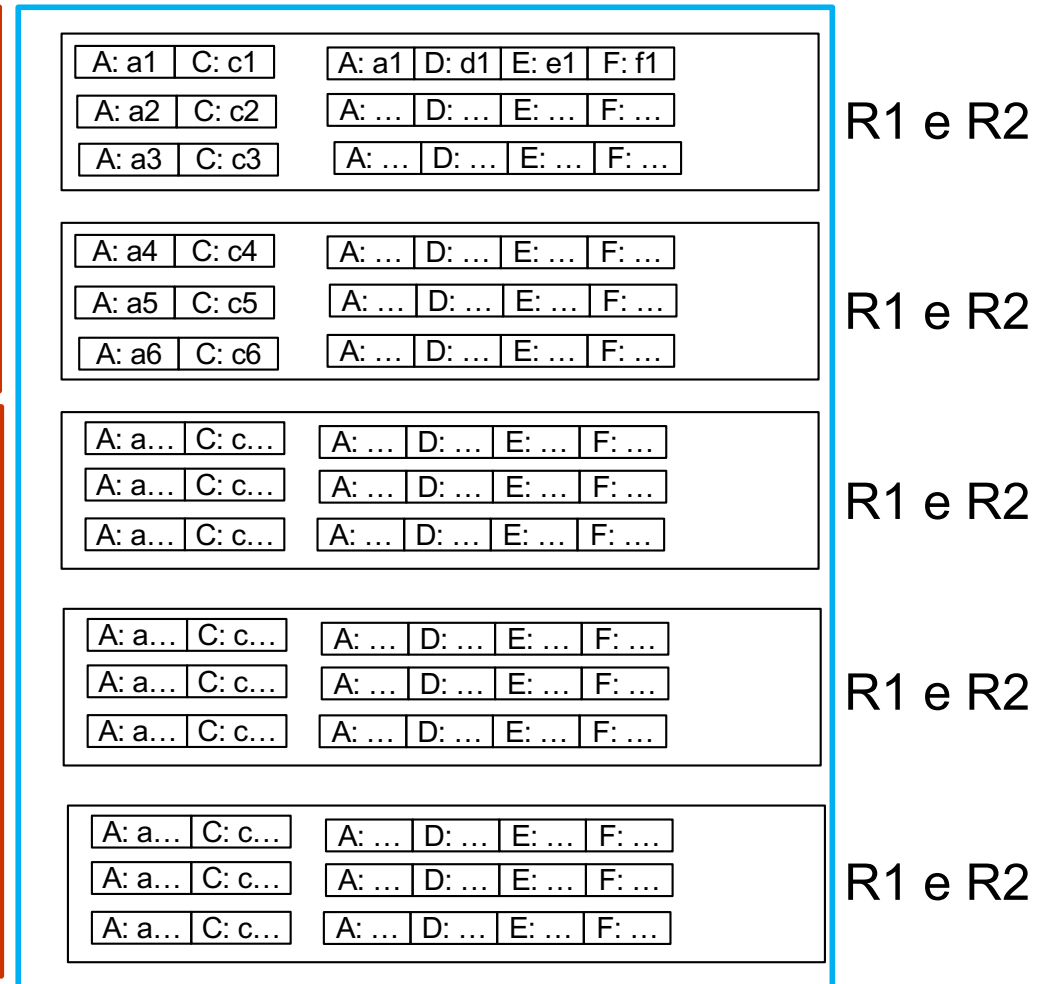


# To cluster or not to cluster: join complessivo

- Memorizzazione separata



- Memorizzazione in cluster



# To cluster or not to cluster: join complessivo

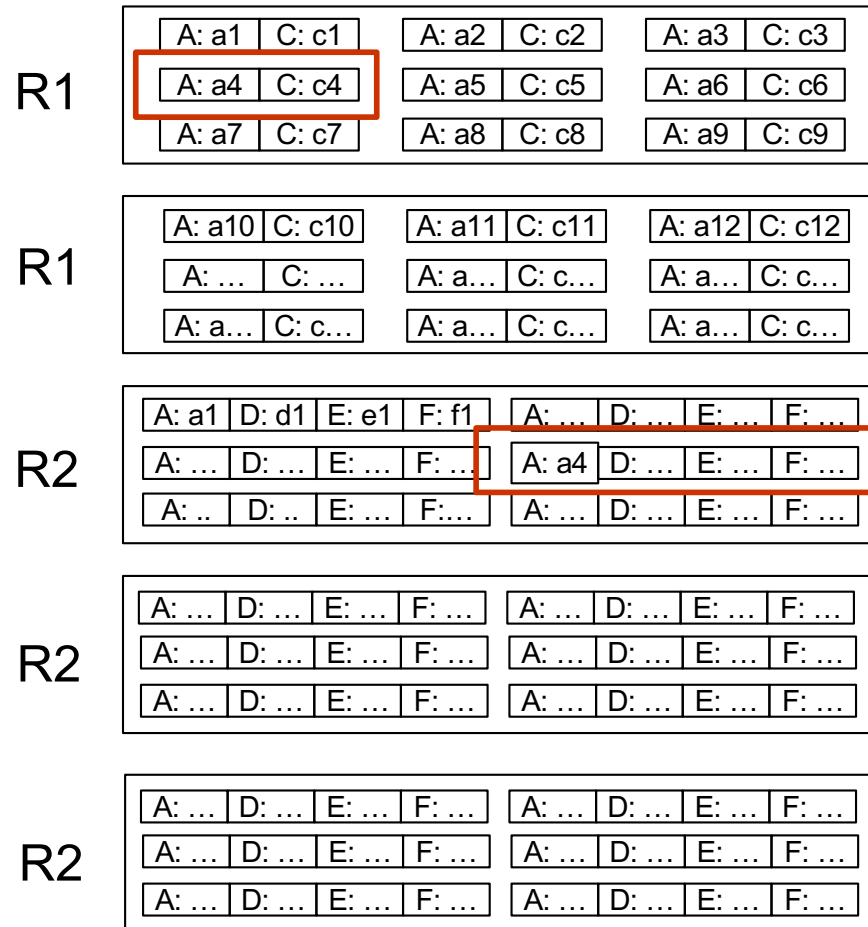
- Occupazione complessiva:
  - la stessa
- Costo delle operazioni
  - scansione di una relazione:
    - preferibile la memorizzazione separata
  - join complessivo
    - stesso costo per le due alternative (5 blocchi e 5 blocchi)

## To cluster or not to cluster una ennupla del join (o poche)

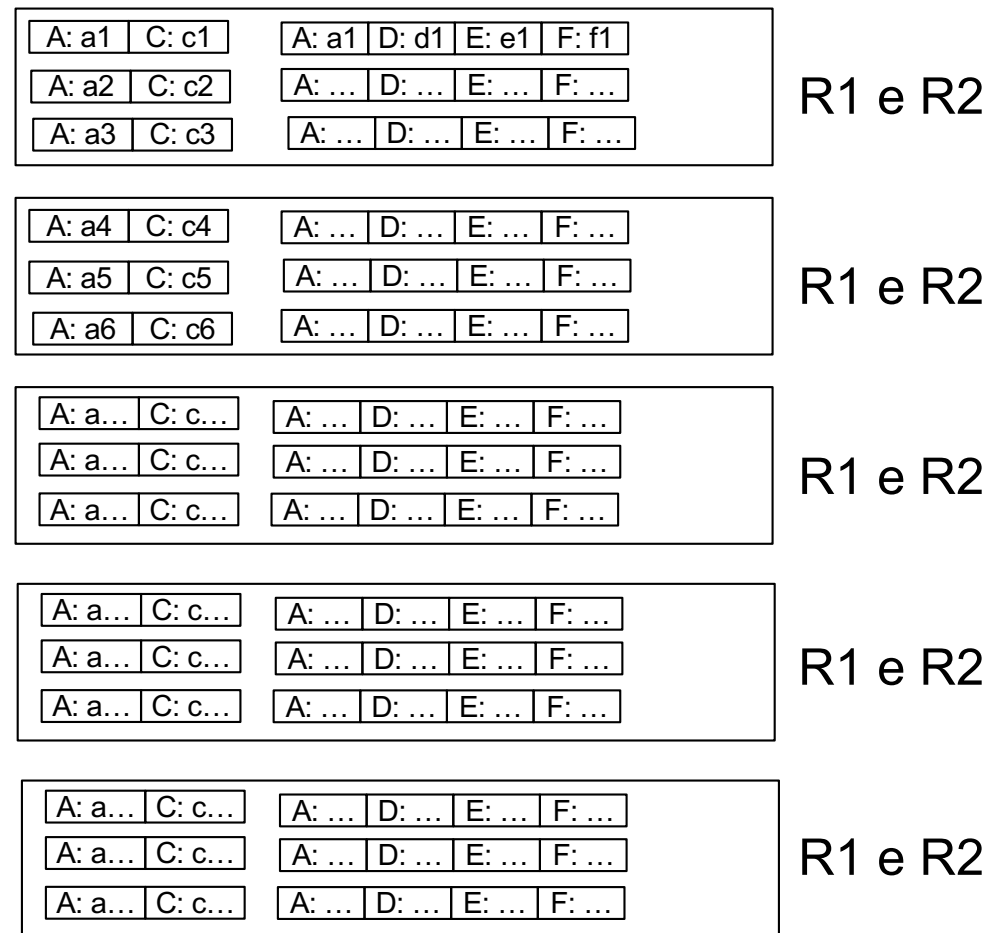
```
SELECT *  
FROM R1 JOIN R2 ON R1.A = R2.A  
WHERE R1.A = 'a4'
```

# To cluster or not to cluster una ennupla del join (o poche)

- Memorizzazione separata

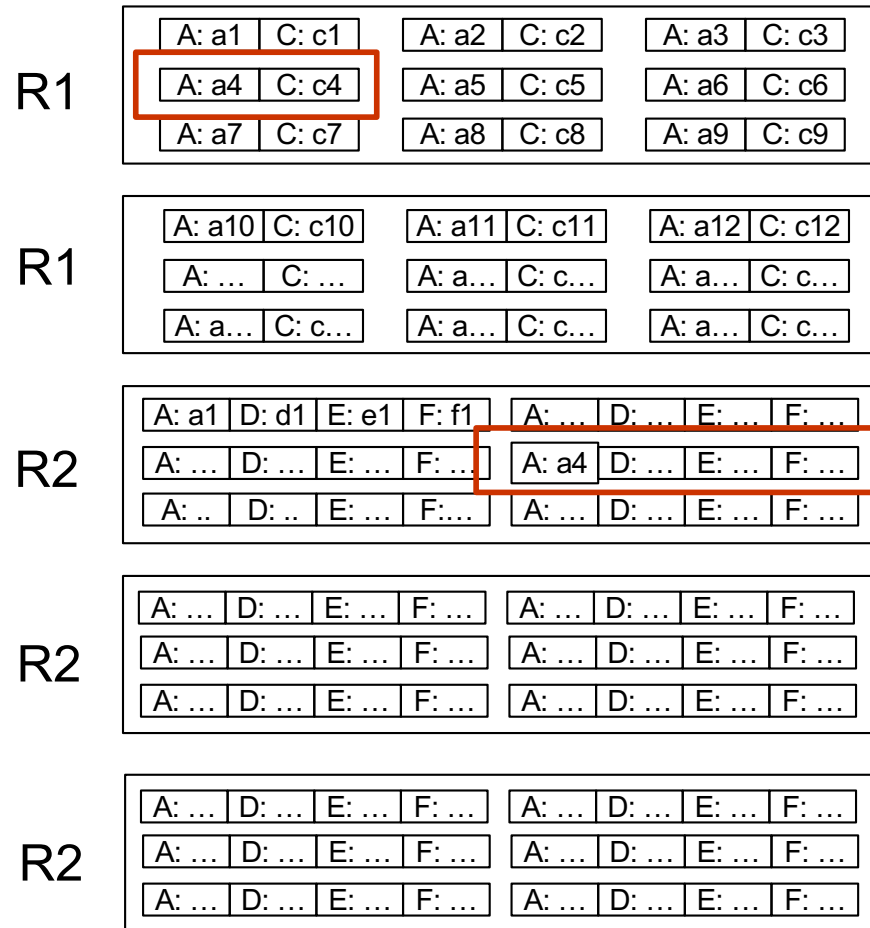


- Memorizzazione in cluster

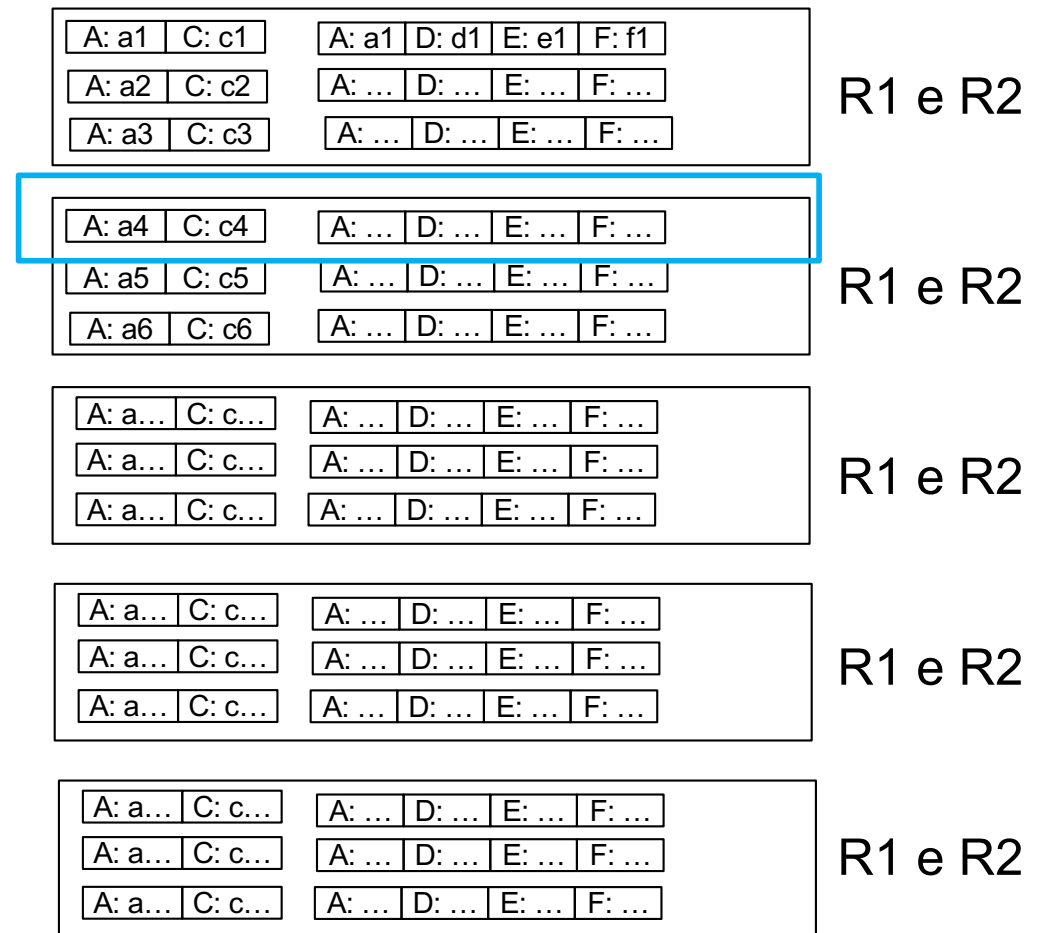


# To cluster or not to cluster una ennupla del join (o poche)

- Memorizzazione separata



- Memorizzazione in cluster



# To cluster or not to cluster: una ennupla del join (o poche)

- Occupazione complessiva:
  - la stessa
- Costo delle operazioni
  - scansione di una relazione:
    - preferibile la memorizzazione separata
  - join complessivo
    - stesso costo per le due alternative
  - una ennupla del join (o poche)
    - preferibile il cluster
      - trovo entrambe le componenti in un blocco anziché due

# To cluster or not to cluster: sintesi

- Occupazione complessiva:
  - la stessa
- Costo delle operazioni
  - scansione di una relazione:
    - preferibile la memorizzazione separata
  - join complessivo
    - stesso costo per le due alternative
  - una ennupla del join (o poche)
    - preferibile il cluster

# To cluster or not to cluster: sintesi

- Occupazione complessiva:
  - la stessa
- Costo delle operazioni
  - scansione di una relazione:
    - preferibile la memorizzazione separata
  - join complessivo
    - stesso costo per le due alternative
  - una enupla del join (o poche)
    - preferibile il cluster
- Per valutare, confrontiamo i costi complessivi, tenendo conto del "carico applicativo" (frequenze delle operazioni principali)

$$c1 \times f1 + c2 \times f2$$



# Torniamo all'esercizio

Basi di dati II — 28 giugno 2016 — Compito A

## Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (i) memorizzazione separata delle due relazioni, entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

- (ii) memorizzazione in un cluster delle due relazioni pure entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità sempre  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

# Scelta (i): memorizzazione separata

Basi di dati II — 28 giugno 2016 — Compito A

## Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (i) memorizzazione separata delle due relazioni, entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

# Scelta (i): memorizzazione separata

Basi di dati II — 28 giugno 2016 — Compito A

## Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (i) memorizzazione separata delle due relazioni, entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

$$c1 \times f1 + c2 \times f2$$

# Scelta (i): memorizzazione separata

Basi di dati II — 28 giugno 2016 — Compito A

## Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (i) memorizzazione separata delle due relazioni, entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

$$c1 \times f1 + c2 \times f2$$

$$f1 = 10.000$$

$$f2 = 1$$

# Scelta (i): memorizzazione separata

Basi di dati II — 28 giugno 2016 — Compito A

## Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (i) memorizzazione separata delle due relazioni, entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

$$c1 \times f1 + c2 \times f2$$

$$f1 = 10.000$$

$$f2 = 1$$

$$c1 =$$

# Scelta (i): memorizzazione separata

Basi di dati II — 28 giugno 2016 — Compito A

## Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (i) memorizzazione separata delle due relazioni, entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

$$c1 \times f1 + c2 \times f2$$

$$f1 = 10.000$$

$$f2 = 1$$

$$c1 = 3 + 3 = 6$$

# Scelta (i): memorizzazione separata

Basi di dati II — 28 giugno 2016 — Compito A

## Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (i) memorizzazione separata delle due relazioni, entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

$$c1 \times f1 + c2 \times f2$$

$$f1 = 10.000$$

$$f2 = 1$$

$$c1 = 6$$

# Scelta (i): memorizzazione separata

Basi di dati II — 28 giugno 2016 — Compito A

## Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (i) memorizzazione separata delle due relazioni, entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

$$c1 \times f1 + c2 \times f2$$

$$f1 = 10.000$$

$$f2 = 1$$

$$c1 = 6$$

$$c2 =$$



# Scelta (i): memorizzazione separata

Basi di dati II — 28 giugno 2016 — Compito A

## Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (i) memorizzazione separata delle due relazioni, entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

$$c1 \times f1 + c2 \times f2$$

$$f1 = 10.000$$

$$f2 = 1$$

$$c1 = 6$$

$$c2 =$$

numero di blocchi

# Scelta (i): memorizzazione separata

Basi di dati II — 28 giugno 2016 — Compito A

## Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (i) memorizzazione separata delle due relazioni, entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

$$c1 \times f1 + c2 \times f2$$

$$f1 = 10.000$$

$$f2 = 1$$

$$c1 = 6$$

$$c2 = \boxed{\text{□} / \text{□}}$$

numero di blocchi  
numero di ennuple

fattore di blocco

# Scelta (i): memorizzazione separata

Basi di dati II — 28 giugno 2016 — Compito A

## Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (i) memorizzazione separata delle due relazioni, entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

$$c1 \times f1 + c2 \times f2$$

$$f1 = 10.000$$

$$f2 = 1$$

$$c1 = 6$$

$$c2 = \boxed{N / \quad}$$

numero di blocchi  
numero di ennuple

fattore di blocco

# Scelta (i): memorizzazione separata

Basi di dati II — 28 giugno 2016 — Compito A

## Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (i) memorizzazione separata delle due relazioni, entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

$$c1 \times f1 + c2 \times f2$$

$$f1 = 10.000$$

$$f2 = 1$$

$$c1 = 6$$

$$c2 = \boxed{N / \left[ \frac{B}{L} \right]}$$

numero di blocchi  
numero di ennuple  
dimensione blocchi  
dimensione record  
fattore di blocco

# Scelta (i): memorizzazione separata

Basi di dati II — 28 giugno 2016 — Compito A

## Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (i) memorizzazione separata delle due relazioni, entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

$$c1 \times f1 + c2 \times f2$$

$$f1 = 10.000$$

$$f2 = 1$$

$$c1 = 6$$

$$c2 = \boxed{N / \left( B / \boxed{\phantom{0000}} \right)}$$

numero di blocchi  
numero di ennuple  
dimensione blocchi  
dimensione record  
fattore di blocco

# Scelta (i): memorizzazione separata

Basi di dati II — 28 giugno 2016 — Compito A

## Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (i) memorizzazione separata delle due relazioni, entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

$$c1 \times f1 + c2 \times f2$$

$$f1 = 10.000$$

$$f2 = 1$$

$$c1 = 6$$

$$c2 = \boxed{N / (B / 2 \times L)}$$

numero di blocchi  
numero di ennuple  
dimensione blocchi  
dimensione record  
fattore di blocco

# Scelta (i): memorizzazione separata

Basi di dati II — 28 giugno 2016 — Compito A

## Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (i) memorizzazione separata delle due relazioni, entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

$$c1 \times f1 + c2 \times f2$$

$$f1 = 10.000$$

$$f2 = 1$$

$$c1 = 6$$

$$c2 = N / (B / 2 \times L)$$



# Scelta (i): memorizzazione separata

Basi di dati II — 28 giugno 2016 — Compito A

## Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (i) memorizzazione separata delle due relazioni, entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

$$c1 \times f1 + c2 \times f2$$

$$f1 = 10.000$$

$$f2 = 1$$

$$c1 = 6$$

$$c2 = N / (B / 2 \times L) = 2.000.000 / (4.000/4) = 2.000$$



# Scelta (i): memorizzazione separata

Basi di dati II — 28 giugno 2016 — Compito A

## Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (i) memorizzazione separata delle due relazioni, entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

$$c1 \times f1 + c2 \times f2$$

$$f1 = 10.000$$

$$f2 = 1$$

$$c1 = 6$$

$$c2 = 2.000$$

# Scelta (i): memorizzazione separata

Basi di dati II — 28 giugno 2016 — Compito A

## Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (i) memorizzazione separata delle due relazioni, entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

$$c1 \times f1 + c2 \times f2 = 62.000$$

$$f1 = 10.000$$

$$f2 = 1$$

$$c1 = 6$$

$$c2 = 2.000$$

## Scelta (ii): memorizzazione in cluster

### Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (ii) memorizzazione in un cluster delle due relazioni pure entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità sempre  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

*scelta (i)*

$c_{\text{Tot}} = 62.000$

$f_1 = 10.000$

$f_2 = 1$

$c_1 = 6$

$c_2 = 2.000$

$c_1 \times f_1 + c_2 \times f_2$

*scelta (ii)*

$c_{\text{Tot}} =$

$f_1 = 10.000$

$f_2 = 1$

$c_1 =$

$c_2 =$

## Scelta (ii): memorizzazione in cluster

### Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

(ii) memorizzazione in un cluster delle due relazioni pure entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità sempre  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

*scelta (i)*

$c_{\text{Tot}} = 62.000$

$f_1 = 10.000$

$f_2 = 1$

$c_1 = 6$

$c_2 = 2.000$

$c_1 \times f_1 + c_2 \times f_2$

**o1: tutto in un blocco e un indice solo**

*scelta (ii)*

$c_{\text{Tot}} =$

$f_1 = 10.000$

$f_2 = 1$

$c_1 =$

$c_2 =$

## Scelta (ii): memorizzazione in cluster

### Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

(ii) memorizzazione in un cluster delle due relazioni pure entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità sempre  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

*scelta (i)*

$c_{\text{Tot}} = 62.000$

$f_1 = 10.000$

$f_2 = 1$

$c_1 = 6$

$c_2 = 2.000$

$c_1 \times f_1 + c_2 \times f_2$

**o1: tutto in un blocco e un indice solo**

*scelta (ii)*

$c_{\text{Tot}} =$

$f_1 = 10.000$

$f_2 = 1$

$c_1 = 3$

$c_2 =$

## Scelta (ii): memorizzazione in cluster

### Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (ii) memorizzazione in un cluster delle due relazioni pure entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità sempre  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

<i>scelta (i)</i>	$c1 \times f1 + c2 \times f2$	<i>scelta (ii)</i>
$c_{Tot} = 62.000$		$c_{Tot} =$
$f1 = 10.000$	$o2:$	$f1 = 10.000$
$f2 = 1$		$f2 = 1$
$c1 = 6$		$c1 = 3$
$c2 = 2.000$		$c2 =$

## Scelta (ii): memorizzazione in cluster

### Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (ii) memorizzazione in un cluster delle due relazioni pure entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità sempre  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

*scelta (i)*

$c_{\text{Tot}} = 62.000$

$f_1 = 10.000$

$f_2 = 1$

$c_1 = 6$

$c_2 = 2.000$

$c_1 \times f_1 + c_2 \times f_2$

$o_2$ : debbo accedere a tutti i blocchi (con le ennuple di entrambe le relazioni)

*scelta (ii)*

$c_{\text{Tot}} =$

$f_1 = 10.000$

$f_2 = 1$

$c_1 = 3$

$c_2 =$



## Scelta (ii): memorizzazione in cluster

### Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (ii) memorizzazione in un cluster delle due relazioni pure entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità sempre  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

<i>scelta (i)</i>	$c1 \times f1 + c2 \times f2$	<i>scelta (ii)</i>
$c_{Tot} = 62.000$		$c_{Tot} =$
$f1 = 10.000$	$o_2$ : debbo accedere a tutti i blocchi (con le ennuple di entrambe le relazioni)	$f1 = 10.000$
$f2 = 1$		$f2 = 1$
$c1 = 6$		$c1 = 3$
$c2 = 2.000$		$c2 =$
$c2 = N / (B / (2 + 4) \times L) = 6.000$		



## Scelta (ii): memorizzazione in cluster

### Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

(ii) memorizzazione in un cluster delle due relazioni pure entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità sempre  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

<i>scelta (i)</i>	$c1 \times f1 + c2 \times f2$	<i>scelta (ii)</i>
$c_{Tot} = 62.000$		$c_{Tot} =$
$f1 = 10.000$	$o_2$ : debbo accedere a tutti i blocchi (con le ennuple di entrambe le relazioni)	$f1 = 10.000$
$f2 = 1$		$f2 = 1$
$c1 = 6$		$c1 = 3$
$c2 = 2.000$		$c2 = 6.000$
$c2 = N / (B / (2 + 4) \times L) = 6.000$		

## Scelta (ii): memorizzazione in cluster

### Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (ii) memorizzazione in un cluster delle due relazioni pure entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità sempre  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

*scelta (i)*

$c_{\text{Tot}} = 62.000$

$f_1 = 10.000$

$f_2 = 1$

$c_1 = 6$

$c_2 = 2.000$

$c_1 \times f_1 + c_2 \times f_2$

*scelta (ii)*

$c_{\text{Tot}} =$

$f_1 = 10.000$

$f_2 = 1$

$c_1 = 3$

$c_2 = 6.000$

## Scelta (ii): memorizzazione in cluster

### Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

(ii) memorizzazione in un cluster delle due relazioni pure entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità sempre  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

*scelta (i)*

$$c_{\text{Tot}} = 62.000$$

$$f_1 = 10.000$$

$$f_2 = 1$$

$$c_1 = 6$$

$$c_2 = 2.000$$

$$c_1 \times f_1 + c_2 \times f_2$$

*scelta (ii)*

$$c_{\text{Tot}} = 36.000$$

$$f_1 = 10.000$$

$$f_2 = 1$$

$$c_1 = 3$$

$$c_2 = 6.000$$

## Scelta (ii): memorizzazione in cluster

### Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (ii) memorizzazione in un cluster delle due relazioni pure entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità sempre  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

*scelta (i)*

$$c_{\text{Tot}} = 62.000$$

$$f_1 = 10.000$$

$$f_2 = 1$$

$$c_1 = 6$$

$$c_2 = 2.000$$

$$c_1 \times f_1 + c_2 \times f_2$$

*scelta (ii)*

$$c_{\text{Tot}} = 36.000$$

$$f_1 = 10.000$$

$$f_2 = 1$$

$$c_1 = 3$$

$$c_2 = 6.000$$

## Scelta (ii): memorizzazione in cluster

### Domanda 5 (20%)

Si considerino due relazioni  $R_1(\underline{A}, C)$ ,  $R_2(\underline{A}, D, E, F)$ , in cui gli attributi hanno tutti la stessa dimensione  $L = 2\text{Byte}$ , molto più piccola della dimensione del blocco pari a  $B = 4000\text{ Byte}$ . Si supponga che le relazioni abbiano entrambe  $N = 2.000.000$  ennuple, con gli stessi valori su  $A$ , e che le operazioni più frequenti su di essa siano le seguenti:

- $o_1$  selezione di una ennupla del join di  $R_1$  e  $R_2$  (sulla base del valore di  $A$ ), con frequenza  $f_1 = 10.000$ ;
- $o_2$  scansione dell'intera relazione  $R_1$ , con frequenza  $f_2 = 1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (ii) memorizzazione in un cluster delle due relazioni pure entrambe ordinate su  $A$  e con indice primario su  $A$  con profondità sempre  $p = 4$ , con 2 livelli mediamente disponibili nel buffer.

*scelta (i)*

$$c_{\text{Tot}} = 62.000$$

$$f_1 = 10.000$$

$$f_2 = 1$$

$$c_1 = 6$$

$$c_2 = 2.000$$

$$c_1 \times f_1 + c_2 \times f_2$$

Conviene il cluster!

*scelta (ii)*

$$c_{\text{Tot}} = 36.000$$

$$f_1 = 10.000$$

$$f_2 = 1$$

$$c_1 = 3$$

$$c_2 = 6.000$$

## Seconda domanda

Ripetere la valutazione effettuata alla pagina precedente nel caso in cui le due frequenze sono invece  $f_1 = 1000$  e  $f_2 = 10$

*scelta (i)*

$c_{Tot} =$

$$f_1 = 1.000$$

$$f_2 = 10$$

$$c_1 = 6$$

$$c_2 = 2.000$$

$$c_1 \times f_1 + c_2 \times f_2$$

*scelta (ii)*

$c_{Tot} =$

$$f_1 = 1.000$$

$$f_2 = 10$$

$$c_1 = 3$$

$$c_2 = 6.000$$

## Seconda domanda

Ripetere la valutazione effettuata alla pagina precedente nel caso in cui le due frequenze sono invece  $f_1 = 1000$  e  $f_2 = 10$

*scelta (i)*

$$c_{\text{Tot}} = 26.000$$

$$f_1 = 1.000$$

$$f_2 = 10$$

$$c_1 = 6$$

$$c_2 = 2.000$$

$$c_1 \times f_1 + c_2 \times f_2$$

*scelta (ii)*

$$c_{\text{Tot}} =$$

$$f_1 = 1.000$$

$$f_2 = 10$$

$$c_1 = 3$$

$$c_2 = 6.000$$

## Seconda domanda

Ripetere la valutazione effettuata alla pagina precedente nel caso in cui le due frequenze sono invece  $f_1 = 1000$  e  $f_2 = 10$

*scelta (i)*

$$c_{\text{Tot}} = 26.000$$

$$f_1 = 1.000$$

$$f_2 = 10$$

$$c_1 = 6$$

$$c_2 = 2.000$$

$$c_1 \times f_1 + c_2 \times f_2$$

*scelta (ii)*

$$c_{\text{Tot}} = 63.000$$

$$f_1 = 1.000$$

$$f_2 = 10$$

$$c_1 = 3$$

$$c_2 = 6.000$$



## Seconda domanda

Ripetere la valutazione effettuata alla pagina precedente nel caso in cui le due frequenze sono invece  $f_1 = 1000$  e  $f_2 = 10$

*scelta (i)*

$$c_{\text{Tot}} = 26.000$$

$$f_1 = 1.000$$

$$f_2 = 10$$

$$c_1 = 6$$

$$c_2 = 2.000$$

$$c_1 \times f_1 + c_2 \times f_2$$

Il cluster non conviene!

*scelta (ii)*

$$c_{\text{Tot}} = 63.000$$

$$f_1 = 1.000$$

$$f_2 = 10$$

$$c_1 = 3$$

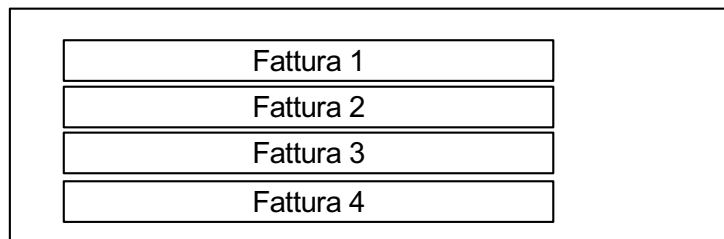
$$c_2 = 6.000$$

# Altro esempio di clustering

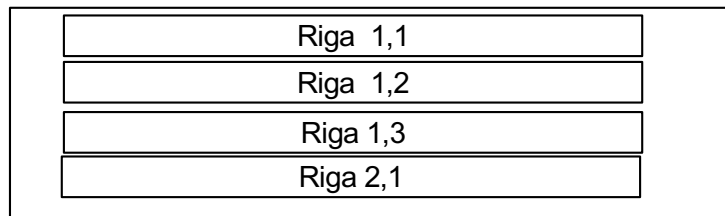
- Relationship uno a molti:

Fatture(Numero, Data, Cliente)

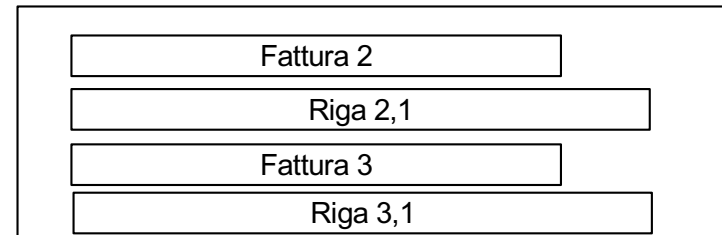
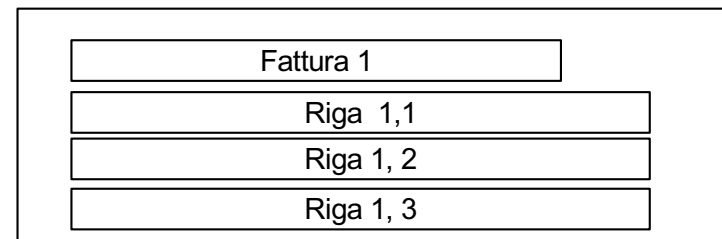
DettagliFatture(Numero, Riga, Quantità, Importo)



...



...



...