



Esercizio 2

E' data la query (con parametro)

```
SELECT * FROM TESI WHERE ARGOMENTO = X
```

sulla relazione:

```
TESI (COD, TITOLO, AUTORE, ANNO, ARGOMENTO)
```

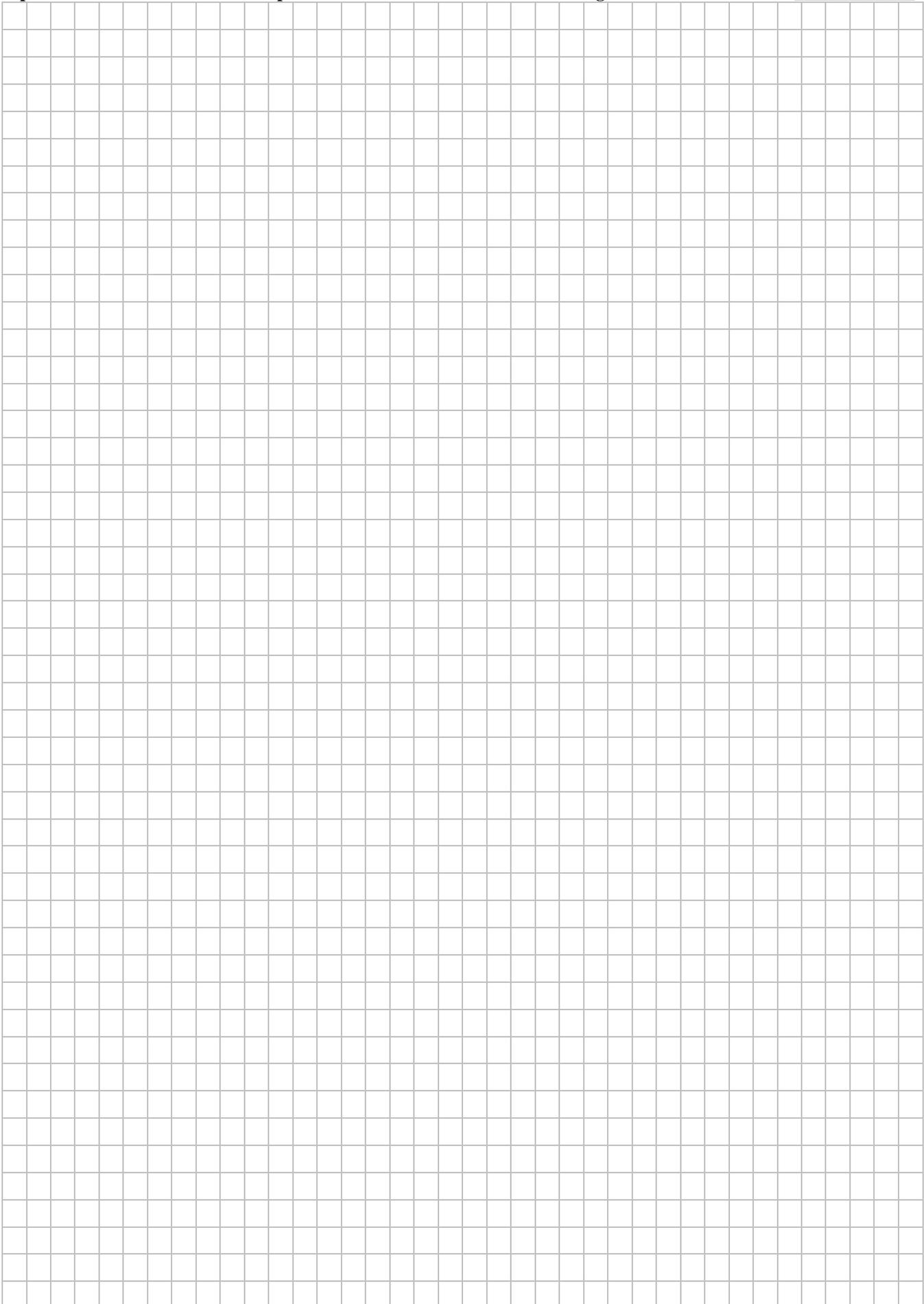
su cui è costruito un indice parziale unclustered su ARGOMENTO relativo a 1/20 degli argomenti presenti. Sapendo che in base al carico di lavoro stimato X appartiene all'insieme di valori indicizzati con probabilità dell'80% si calcoli *il numero di tuple attese* e *il costo atteso della query* noti i seguenti dati:

| | |
|---|--------------------------------|
| Dimensione di un nodo (indice e file dati) | D=1KB |
| Numero di record | NR=100000 |
| Numero valori di chiave distinti sul campo ARGOMENTO (si suppongano equamente distribuiti) | NK _{argomento} = 2000 |
| Lunghezza puntatore | len(p)= 4 byte |
| Lunghezza chiave del campo ARGOMENTO | len(tipo)= 10 byte |
| Lunghezza del record TESI | lentot= 50 byte |
| Ordine del B+-tree | g=11 |
| Utilizzazione (foglie e file dati) | u=ln 2 |

FORMULE UTILI

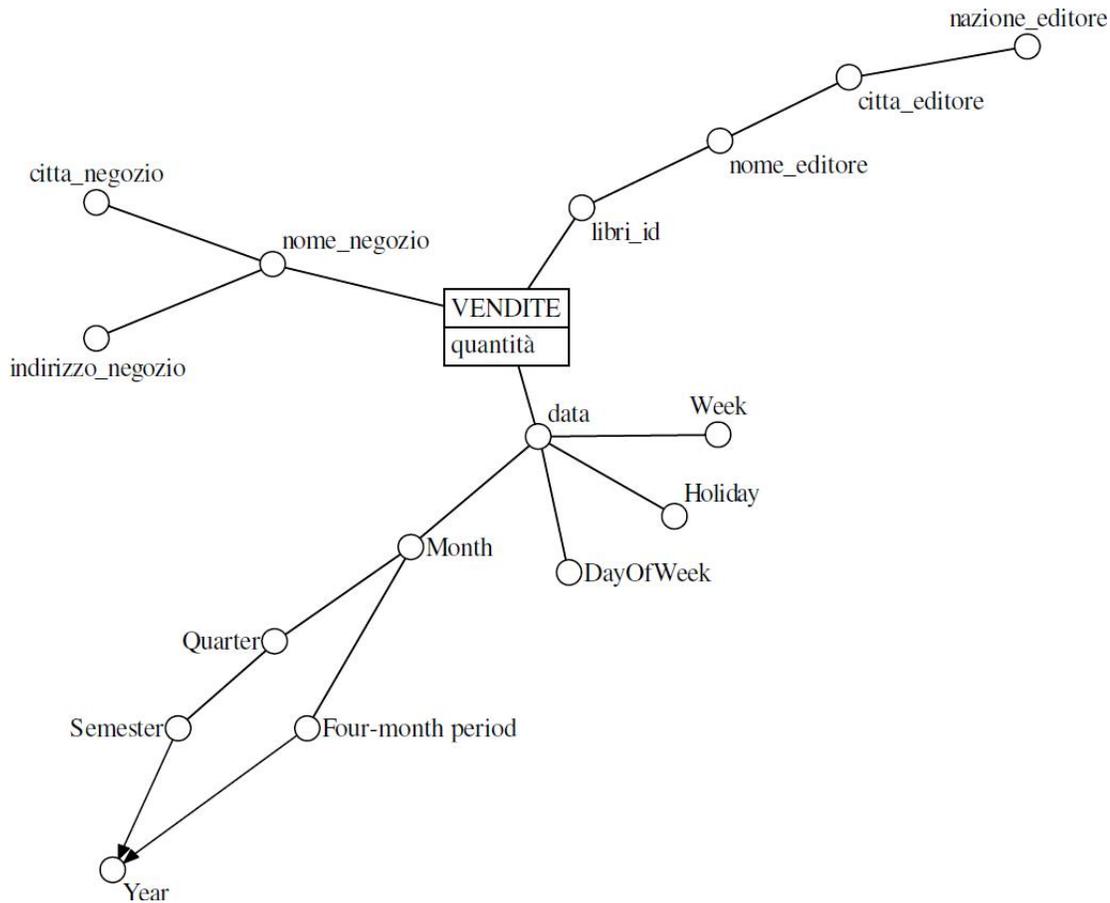
- **Dimensione di una relazione:** $NP = \lceil NR \times len(t) / (D \times u) \rceil$
- **Costo di ordinamento (Sort Merg Z vie):** $2 \times NP \times \lceil \log_z NP \rceil$
- **Accesso a tabelle con indice**
 - **Numero di foglie del B+-Tree:** $NL = \lceil (NK \cdot len(k) + NR \cdot len(p)) / (D \cdot u) \rceil$
 - **altezza del B+-Tree:** $\lceil \log_{2g+1} NN \rceil \leq h \leq 2 + \left\lceil \log_{g+1} \left(\frac{NN}{2} \right) \right\rceil$, $NN = \min (NL, NK)$
 - **Indice Clustered:** $h - 1 + \lceil EK / NK \cdot NL \rceil + \lceil EK / NK \cdot NP \rceil$
 - **Indice Unclustered:** $h - 1 + \lceil EK / NK \cdot NL \rceil + EK \cdot \Phi(NR / NK, NP)$
 - **Formula di Cardenas** $\Phi(ER, NP) = NP \times (1 - (1 - 1/NP)^{ER}) \leq \min\{ER, NP\}$
 -

Svolgimento





Dato il seguente schema di fatto:



Cubo: pubs

Dimensioni: Negozi, Data, Libri

Gerarchie: DataHierarchy, NegoziHierarchy, EditoriHierarchy

1. Spiegare in *linguaggio naturale* cosa visualizza la seguente query MDX:

```
SELECT NON EMPTY ([Negozi].[citta_negozio]) ON columns,  
NONEMPTYCROSSJOIN ({ [Data].[DataHierarchy].[Year].&[2008],  
[Data].[DataHierarchy].[Year].&[2009]}, { [Libri].[EditoriHierarchy].  
[citta_editore].&[Roma], [Libri].[EditoriHierarchy].  
[citta_editore].&[Milano]}) ON ROWS  
FROM [pubs]  
WHERE ([quantità])
```

2. Scrivere una *query MDX* che visualizzi le quantità vendute nell'anno 2008 per negozio e per nazione_editore

3. Scrivere una *query MDX* che visualizzi le quantità vendute nell'anno 2008 per mese e per città negozio (è vietato utilizzare l'attributo Data.anno, utilizzare la gerarchia)

Svolgimento

