

**Basi di dati II, primo modulo (ord.270)**  
**Tecnologia delle basi di dati (ord.509)**

**18 febbraio 2010**

**Tempo a disposizione: un'ora e quindici minuti.**

**Domanda 1 (20%)**

Commentare brevemente la seguente affermazione: “le tecniche per il controllo di concorrenza basate su 2PL e su timestamp si basano su condizioni sufficienti ma non necessarie per la view-serializzabilità.” In particolare, chiarire se essa è vera o falsa, spiegare perché e motivare le ragioni pratiche per le quali si utilizzano tali tecniche anziché la view-serializzabilità stessa.

**Domanda 2 (30%)**

Si considerino un sistema con blocchi di dimensione  $B = 1000$  byte e puntatori ai blocchi di  $P = 2$  byte e una relazione  $T(\underline{A}, B, C, D)$  di cardinalità pari circa a  $R = 6.000.000$ , con ennuple di  $L = 50$  byte e campo chiave  $A$  di  $L_A = 5$  byte e campo  $B$  di  $L_B = 3$  byte. Valutare i pro e i contro relativamente alla presenza di un indice secondario sulla chiave  $A$  e di un altro, pure secondario, su  $B$ , in presenza del seguente carico applicativo:

1. inserimento di una nuova ennupla (con verifica del soddisfacimento del vincolo di chiave), con frequenza  $f_1 = 300$
2. ricerca di una ennupla sulla base del valore della chiave  $A$ , con frequenza  $f_2 = 100$
3. ricerca di ennuple sulla base del valore di  $B$ , con frequenza  $f_3 = 500$
4. elenco di tutte le ennuple, ordinato secondo il valore di  $B$ , con frequenza  $f_4 = 10$

Ragionare in termini di costo degli accessi a memoria secondaria, assumendo disponibilità di buffer che permettano di mantenere stabilmente in memoria due livelli per ciascun indice e considerando che la relazione possa essere memorizzata in forma contigua (assumendo un rapporto 100:1 fra tempo di posizionamento della testina e tempo di lettura). Trascurare le problematiche relative alla concorrenza.

**Domanda 3 (15%)**

Si consideri un B-tree con nodi intermedi che contengono due chiavi e tre puntatori e foglie con due chiavi. Mostrare un possibile contenuto della struttura a seguito di inserimenti delle chiavi nel seguente ordine (a partire dall'albero vuoto): 2, 17, 7, 15, 20, 19, 16, 18, 5, 1. Mostrare anche i passi salienti che portano a tale contenuto.

**Domanda 4 (20%)**

Si supponga di dover realizzare un data mart relativo ad esami universitari. Fra i vari dati di interesse vi sono i voti riportati negli esami. Spiegare quali differenze possono sussistere nel data mart fra il caso in cui interessa ragionare sul voto medio e quello in cui interessa ragionare sulle distribuzioni per fasce di voto (cioè, ad esempio, quanti voti fra 18 e 20, quanti fra i 21 e 23 e così via). Spiegare anche come procedere se interessano entrambi gli aspetti.

**Domanda 5 (15%)**

Spiegare perché, scrivendo un programma che accede a due basi di dati diverse utilizzando JDBC, non è possibile garantire il commit a due fasi. Indicare quali funzionalità aggiuntive sono necessarie per i server locali che vogliono realizzare tale servizio e come si può procedere con un sistema che non disponga di tali funzionalità.