

**Basi di dati II**  
**Prof. Stefano Paraboschi**  
*II prova - 22/12/2015*

- A. Si ha una base di dati per un'azienda che si occupa di logistica, con il trasporto mediante camion di merci. Vi sono un certo numero di centri di distribuzione, ciascuno responsabile di un'area. La merce viene raccolta nel centro di distribuzione associato al luogo di origine della spedizione, viene trasferita (se necessario) al centro di distribuzione della destinazione finale e quindi consegnata al destinatario.

CENTRODISTRIBUZIONE(CodCD,Nome,Indirizzo)  
TRATTA(CodTratta,CodCDDOrig,CodCDDDest)  
RECUPERO(CodRec,CodCD,CapDisp,TempoInizio,TempoFine)  
TRASPORTO(CodTrasporto,CodTratta,CapDisp,TempoInizio,TempoFine)  
CONSEGNA(CodCons,CodCD,CapDisp,TempoInizio,TempoFine)  
COLLO(CodCollo,IndOrig,CodCDDOrig,IndDest,CodCDDDest,TempoCreazione,Deadline,FlagOnTime)  
ITERCOLLO(CodCollo,CodPasso,TempoAssegnamento)

Il sistema assume che i colli vengano inseriti nella tabella e un sistema di trigger deve andare a gestire la creazione di tuple in ITERCOLLO, dove l'attributo CodPasso descrive i codici dei recuperi dalla sorgente, degli eventuali trasporti tra centri, e della consegna finale. In ogni caso, un collo può essere gestito se rimane abbastanza capacità e se i tempi dei vari passi sono compatibili; ovvero, il TempoFine di un passo deve precedere il TempoInizio del successivo; il TempoInizio del recupero deve essere successivo al TempoCreazione del collo. L'attributo FlagOnTime deve essere impostato a TRUE se il TempoFine della consegna precede il valore di Deadline.

Si considerino questi due scenari. Nel primo scenario, si assuma che il trasporto tra centri possa solo essere diretto. Nel secondo scenario, i trigger devono verificare se esiste un percorso composto da due tratte che riesca a recapitare il collo in un tempo inferiore.

(Punto aggiuntivo) Descrivere, senza scrittura di codice, come si potrebbe gestire un numero di trasporti arbitrario.

- B. Un sistema distribuito è composto da un DBMS  $N_1$  che gestisce la tabella STUDENTE(Matr,Cognome,Nome,DataNascita,Indirizzo) e un DBMS  $N_2$  che gestisce la tabella ESAME(Matr,CodCorso,Data,Voto). La tabella STUDENTE occupa 10MB e contiene 10K tuple. La tabella ESAME occupa 100MB e contiene 100K tuple. La struttura primaria è sequenziale entry-sequenced per entrambe le tabelle. Si consideri la seguente query eseguita sul server  $N_2$ :

```
select * from Studente S join Esame E on S.Matr = E.Matr  
where Cognome = "Rossi" and Data > 1/1/2014 and CodCorso = "DB2"
```

La base di dati ha 100 studenti con cognome "Rossi", 10K esami nel 2014, 100 esami di DB2 in totale, di cui 10 nel 2014. Si considerino due scenari. Nel primo, vi è un solo indice secondario sull'attributo Matr di STUDENTE (albero B+ a 3 livelli). Nel secondo, vi sono anche indici costruiti su CodCorso e Data (entrambi alberi B+ a 3 livelli). Stimare il tempo di esecuzione della query visto da  $N_2$  assumendo blocchi da 1KB, tempo di accesso del disco di 10 ms, banda di rete pari a 100KB/s e ritardo di trasmissione tra i 2 server pari a 10ms.

- C. <!ELEMENT GaraATappe (Nome, Anno, Squadra+, Tappa+)>  
<!ELEMENT Squadra (Nome, MediaTempiTotale, PosizioneClassifica, Corridore+) >  
<!ELEMENT Corridore (Nome, DataNasc, LuogoNasc, TempoTotale, PosizioneClassifica) >  
<!ATTLIST Corridore Numero ID #REQUIRED >  
<!ELEMENT Tappa (Data, Nome, Partenza, Arrivo, TempoLimite, TempoVincitore, Performance+) >  
<!ELEMENT Performance Tempo >  
<!ATTLIST Performance CorridoreNumero IDREF #REQUIRED >

Lo schema descrive i risultati di gare ciclistiche. Gli elementi non specificati contengono PCDATA.

1. Estrarre in XQuery le squadre con almeno 3 ciclisti diversi che hanno vinto una tappa nella stessa gara a tappe.
  2. Estrarre in XQuery le squadre in cui nessun ciclista ha vinto una tappa nel Giro d'Italia del 2015.
  3. Estrarre in XQuery le squadre per cui il valore di MediaTempiTotale non è corretto, in quanto non presenta la media dei tempi di tutti i ciclisti della squadra.
- D. Progettare uno schema multidimensionale per i dati relativi alle gare ciclistiche descritti nell'Esercizio C.