

Basi di Dati - Prima prova parziale - 8/11/2022 - Compito C  
Domande

Rispondere alle seguenti domande su questo foglio. Usare il foglio protocollo per la brutta copia. Si ricorda che ritirandosi dalla prova si perde il bonus di 2 punti.

Tempo: 20 minuti

Quiz: (contrassegnare con una crocetta la lettera relativa alla risposta corretta - solo una risposta è corretta - non c'è penalizzazione per risposte sbagliate) – totale: 4 punti

1) Nell'ambito del modello relazionale, data una relazione, esiste sempre una superchiave?

- (a) in generale no
- (b) esiste una superchiave solo se non esistono chiavi
- (c) sì, sempre
- (d) sì, purché almeno un attributo ammetta valori nulli

2) Quale di queste sequenze di attività descrive meglio la realizzazione di un database?

- (a) si definisce prima lo schema logico, poi quello concettuale, poi quello esterno
- (b) si definisce prima lo schema concettuale, poi quello logico, poi quello esterno
- (c) si definisce prima lo schema logico, poi quello esterno, poi quello concettuale
- (d) si definisce prima lo schema esterno, poi quello logico, poi quello concettuale

3) A cosa serve un Data Definition Language (DDL)?

- (a) Per l'interrogazione e l'aggiornamento di istanze nella basi di dati
- (b) Per la definizione e la modifica di schemi (logici, esterni, fisici)
- (c) Per la definizione ma non per la modifica di schemi (logici, esterni, fisici)
- (d) Per l'interrogazione ma non per l'aggiornamento di istanze nella basi di dati

4) Nel livello fisico del DBMS, qual è il vantaggio dell'Organizzazione disordinata dei record su disco rispetto a quella ordinata?

- a) L'organizzazione disordinata è più efficiente per ricerche su range di valori
- b) L'organizzazione disordinata è più efficiente in caso di ricerche sul valore della chiave
- c) L'organizzazione disordinata non presenta vantaggi rispetto a quella ordinata
- d) L'organizzazione disordinata è più efficiente in caso di inserimenti, modifiche e cancellazioni

Domande - totale: 6 punti

5) Spiegare cosa si intende per indipendenza logica e fisica dei dati.

L'INDIPENDENZA LOGICA RENDE INDEPENDENTE IL LIVELLO ESTERNO DA QUELLA LOGICO. MODIFICHE AL LIVELLO ESTERNO E LOGICO NON AFFECTANO LA MODIFICA DEL LIVELLO FISICO.

L'INDIPENDENZA FISICA SIGNIFICA CHE IL LIVELLO LOGICO ED ESTERNO SONO INDEPENDENTI DAL LIVELLO FISICO. ILE MODIFICHE AL LIVELLO LOGICO CHE INTRODUCONO INNELLATO LO SCHEMA ESTERNO SONO TRASFERITE QUINTO ALE TERRITORIE A LIVELLO FISICO A MENO CHE NON INTRODUCANO L'IMPLEMENTAZIONE DEL APPLICATIVO. NON SONO RILEVANTI AI LIVELLI SUPERIORI.

6) Spiegare le analogie e le differenze tra la relazione matematica e la relazione del modello relazionale di una base di dati.

LA DIFFERENZA SOSTANZIALE È CHE IN UNA RELAZIONE MATEMATICA NICE ~~PERMETTE~~ UN SISTEMA POSIZIONALE IN CUI COORDINE DEI DOMINI CONTA MENTRE IN UN MODELLO RELAZIONALE NO., I RDBMS (RELATIONAL DATABASE MANAGEMENT SYSTEM) ~~ASSEGNA~~ UN NOME CHE RENDE OGNI DOMINIO UNIVOCO E QUINDI RICONOSCIBILE A PRESCINDERE DALLA SUA POSIZIONE.

7) Qual è la semantica del valore NULL? Cosa succede se in una selezione si applica una condizione (ad esempio "SELECT ... WHERE Stipendio > 50000") su un attributo (ad es. 'Stipendio') che ha valore NULL?

IL VALORE NULL SERVE A NON CREARE "PROBLEMI" SE DURANTE UNA RICERCA SE UN CAMPO È VUOTO NEL CASO PRESO IN ESAME LA SELEZIONE NON RESTITUISCE LE RIGHE IN CUI L'ATTRIBUTO È NULL.

Rispondere alle seguenti domande su questo foglio. Usare il foglio protocollo per la brutta copia.

Tempo: 25 minuti

1) Quiz: (contrassegnare con una crocetta la lettera relativa alla risposta corretta) – 2 punti

Considerare la relazione  $R(A, B, C)$ . Da quali tuple è composta la relazione  $SEL_{A>B}(R)$ ?

- (a) Le tuple di  $R$  in cui il valore di  $A$  è maggiore di quello di  $B$ , più quelle per cui  $A$  è NULL ma  $B$  è diverso da NULL.
- (b) Le tuple di  $R$  in cui il valore di  $A$  è maggiore di quello di  $B$ , più quelle per cui  $B$  è NULL ma  $A$  è diverso da NULL.
- (c) Le tuple di  $R$  in cui il valore di  $A$  è maggiore di quello di  $B$ , più quelle per cui  $A$  è NULL oppure  $B$  è NULL (o entrambi sono NULL).
- (d) Le tuple di  $R$  in cui il valore di  $A$  e di  $B$  è diverso da NULL e il valore di  $A$  è maggiore di quello di  $B$ .

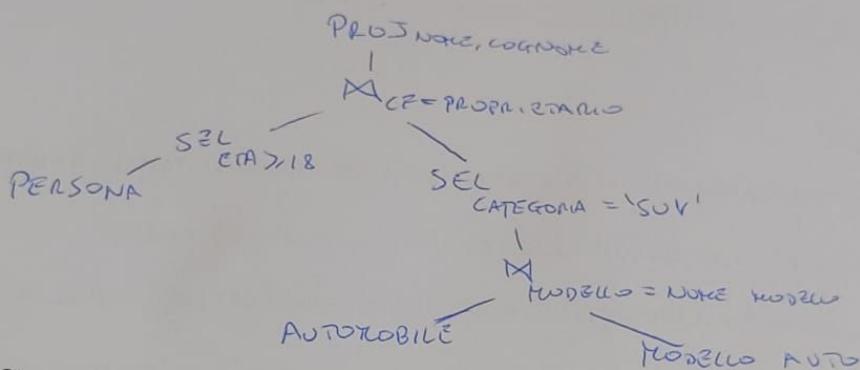
### Espressioni in algebra relazionale

Si consideri il seguente schema che gestisce un elenco di case costruttrici, automobili, e relativi proprietari. Per semplicità si assuma che nessun campo delle relazioni possa avere valore NULL.

- PERSONA(CodiceFiscale, Nome, Cognome, Età)
- CASACOSTRUTTRICE(Codice, NomeCasa, Nazione)
- MODELLOAUTO(CodiceCasa, NomeModello, Categoria)
  - con vincolo di integrità referenziale fra CodiceCasa e la relazione CASACOSTRUTTRICE
- AUTOMOBILE(Targa, Casa, Modello, Anno, Proprietario)
  - con vincolo di integrità referenziale tra l'attributo Casa e la relazione CASACOSTRUTTRICE
  - con vincolo di integrità referenziale fra gli attributi <Casa, Modello> e la relazione MODELLOAUTO
  - con vincolo di integrità referenziale fra Proprietario e la relazione PERSONA

Si scrivano le espressioni in algebra relazionale che traducono le seguenti interrogazioni:  
 2) Elenca la targa delle auto di categoria 'SUV'. - 3 punti  
 3) Elenca nome e cognome dei diciottenni che possiedono un'auto di categoria 'SUV'. Disegnare anche l'albero dell'espressione (expression tree). - 4 punti

Expression tree:

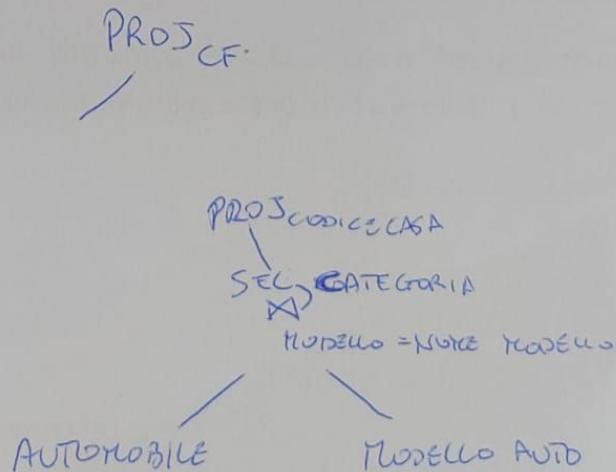


Espressione algebrica:

$$\text{PROJ}_{\text{NomeCognome}} \left( \left( \text{SEL}_{\text{ETÀ} > 18} \text{ PERSONA} \right) \text{JOIN}_{\text{CF} = \text{PROPRIETARIO}} \left( \text{AUTOMOBILE} \text{ JOIN}_{\text{Modello} = \text{Nome Modello}} \text{ Modello AUTO} \right) \right)$$

4) Elenca il codice fiscale delle persone che possiedono sia una 'BMW' che una 'KIA' (N.B. devono possederle entrambe, non solo una delle due). 'BMW' e 'KIA' sono nomi di case costruttrici (attributo NomeCasa della relazione CasaCostruttrice). Disegnare anche l'albero dell'espressione. - 3 punti

Expression tree:



Espressione algebrica:

Basi di Dati - Prima prova intermedia - 8/11/2022 - Compito D  
Livello fisico del DBMS

Rispondere alle seguenti domande su questo foglio. Usare il foglio protocollo per la  
brutta copia.

Tempo: 20 minuti

- 1) Spiegare brevemente il funzionamento di un indice Hash. Quali vantaggi ha un indice Hash rispetto ad un B Tree? In cosa consiste il problema delle collisioni e come si può risolvere? – 3 punti

UN INDICE HASH HA L'OBBIETTIVO DI ACCEDERE IN MODO DIRETTO A UN  
INSIEME DI RECORD UTILIZZANDO I VALORI DI UN CAMPO, QUESTO  
AVVIENE ATTRAVERSO UNA FUNZIONE HASH CHE HA UN COSTO COMPUTAZIONALE  
EDIO DI  $\log(n)$ . ANCHE DI CONTRO LE FUNZIONI HASH SUFFRANO PER PROBLEMI  
DI COLLISIONI OVVERO ESSENDO LA CARDINALITÀ DELL'INSIEME PEGLI INPUT  
PIÙ PROSSIMA ALLA CARDINALITÀ DELL'INSIEME DEI POSSIBILI VALORI DELLE  
CADE CHE ALCUNI OUTPUT DELLA FUNZIONE COINCIDANO PER INPUT DIVERSE.  
PER RISOLVERE IL PROBLEMA DELLE COLLISIONI SI PUÒ AUMENTARE LO SPAZIO DI RIDONDANZA  
REQUISITA UTILIZZANDO UNA FUNZIONE HASH.

delle  
2) Nella scrittura dei record su file system, un DBMS può usare una tecnica spanned oppure unspanned.  
Spiegare cosa si intende per tecnica spanned e unspanned, e discutere vantaggi e svantaggi di entrambe - 3 punti

LA TECNICA UNSPANNED CONSISTE NELLA SCRITURA DELL'INTERO RECORD SU UN BLOCCO. HA COME VANTAGGI LA FACILITÀ D'IMPLEMENTAZIONE E LA VELOCITÀ D'ESECUZIONE. DI CONTRO ITA CHE SI SPRECA SPAZIO. LA TECNICA SPANNED, INVECE, CERCA DI RILEMPIRE AL MASSIMO DELLA CAPACITÀ IL BLOCCO ~~ATTENDENDO~~, SE NECESSARIO, I RECORD. I VANTAGGI SONO UN MIGLIOR UTILIZZO DELLA MEMORIA, DI CONTRO È DIFFICILE DA IMPLEMENTARE.

3) In base a cosa si calcola il costo di una operazione su un DBMS? Nell'architettura di un BDMS, quale processo di ottimizzazione attua l'ottimizzatore delle query? Spiegare cosa si intende per euristica 'Push selections down' e 'Push projection down', e perché queste euristiche possono diminuire il costo di esecuzione di una query - 3 punti

IL COSTO DI UN'OPERAZIONE SU UN DBMS È DATO DAL NUMERO DI ACCESSI ALLA MEMORIA SECONDARIA. PER EURISTICA "PUSH SELECTION DOWN" E "PUSH PROJECTION DOWN" SI RIFERISCE AL PROCESSO DI OTTIMIZZAZIONE DELLE QUERY, OVVERO, ESEGUIRE IL PRIMA POSSIBILE LE SELECTION E Poi LE PROJECTION RENDE IL COSTO, IN TERMINI DI MEMORIA, PIÙ EFFICIENTE ~~PERCHÉ~~ PERCHÉ EVITA DI TRASFERIRSI INFORMAZIONI NON UTILI PER TUTTE LE OPERAZIONI FUTURE.