

Laboratorio di Basi di Dati

Matricola:

Appello del 13/07/2016 (100 minuti) Nome e cognome:

Service: si-oracle-11.csr.unibo.it

Login: esame____ Password: _____

FILA

A

1) Un sistema informativo per la ricerca scientifica deve gestire le statistiche relative alla produttività dei ricercatori. Nel DB vengono memorizzate informazioni relativi agli articoli pubblicati e alle citazioni presenti in ogni articolo.

ARTICOLO (IdArt, Titolo, Rivista, vol, num, pagg)

RICERCATORE (IdRic, Nome, Cognome, email, afferenza)

AUTORE (IdArt: Articolo, IdRic: Ricercatore)

CITA (IdArticolo: Articolo, Posizione, ArticoloCitato: Articolo)

Si vuole calcolare l'H-index di un ricercatore. L'H-index è un indice bibliometrico per quantificare la prolificità e l'impatto del lavoro degli scienziati che misura il numero di citazioni ricevute rapportate agli articoli pubblicati.

Secondo la definizione, uno scienziato ha un indice n se almeno n lavori tra quelli che ha pubblicato sono stati citati almeno n volte ciascuno.

Il calcolo dell'indice viene eseguito in base alla distribuzione delle citazioni che le pubblicazioni di un ricercatore ricevono.

In altre parole, uno studioso con un indice pari a 10 ha pubblicato 10 lavori citati almeno 10 volte ciascuno (i suoi restati lavori sono citati meno di 10 volte). Ad esempio nel caso seguente di un autore che ha pubblicato 6 lavori con il numero di citazioni di seguito riportato l'H-index è 4:

pubblicazione A, citazioni 12

pubblicazione B, citazioni 8

pubblicazione C, citazioni 5

pubblicazione D, citazioni 5

pubblicazione E, citazioni 1

pubblicazione F, citazioni 1

a) Si definisca la base di dati su ORACLE.

b) Si scriva la funzione $HIndex(vIdRis)$ che:

- calcola il numero di citazioni ricevute da ciascuna articolo di cui $vIdRis$ è autore

- restituisce e stampa a video l'H-index del ricercatore considerato ($vIdRis$)

ATTENZIONE: Nel caso in cui i campi Nome, Cognome e Login non siano compilati in modo leggibile il compito non sarà corretto

2) Utilizzando il database TPCD, si disegni l'albero di esecuzione proposto da ORACLE e si calcoli il costo di accesso della seguente query.

```
select P_BRAND, count(*)
from TPCD.PART, TPCD.ORDERS, TPCD.LINEITEM
where P_PARTKEY=L_PARTKEY and L_ORDERKEY=O_ORDERKEY
      and O_ORDERPRIORITY='5-LOW'
group by P_BRAND;
```

Si facciano le seguenti assunzioni e si estraggano dal DB eventuali dati mancanti:

$D = 4096$ byte $\text{len}(P) = \text{len}(K) = 4$ byte $NB = 101$ $u = 0.69$

Si assumo inoltre che ORACLE non applichi proiezioni sui risultati intermedi e che non esegua operazioni in pipeline.

