

Seminarska naloga PB za leto 2016

Podatki

Podatki, na katerih boste delali, so pridobljeni s strani www.amazon.com s pajkom. Podatkovna baza vsebuje podatke o albumih, ki so bili na voljo poleti leta 2014. Tako imate za albume na voljo podatke o naslovu albuma, imenu založnika, imenu izvajalca, ceni, letu izdaje, številu medijev in tipu medija. Baza je narejena tako, da se je za prvi album izbral White album skupine The Beatles. Od tukaj naprej so pregledani vsi albumi, ki so dosegljivi preko strani glede na amazonova priporočila. Bolj točno podatki o »user who bought this also bought«. Podatki zajemajo vse strani, ki so od White albuma oddaljeni največ devet strani.

Sheme

Podatkovna baza sestoji iz naslednjih shem:

Album(AlbumID, AlbumName) – vsebuje imena albumov

Author(#AlbumID, #AuthorID) – za vsak album vsebuje ključ njenega izvajalca

AuthorNames(AuthorID, AuthorName) – vsebuje ključe in imena izvajalcev

Label(#AlbumID, #LabelID) – za vsak album vsebuje ključ njenega založnika

LabelNames(LabelID, Label) – vsebuje ključe in imena založnikov

Medium(#AlbumID, #MediumID) – za vsak album vsebuje ključ medija

MediumNames(MediumID, Medium) -- vsebuje ključe in imena tipov medija

Year(#AlbumID, Year) – za vsak album vsebuje letnico izdaje (lahko je NULL)

Price(#AlbumID, Price) – za vsak album vsebuje ceno albuma (lahko je NULL)

DiscNumber(#AlbumID, NumberOfDiscs) – za vsak album vsebuje število medijev (CD-jev) (lahko je NULL)

Recommendation(#AlbumID, #RecommendationID) – Vsebuje preslikavo Album-Album. Zapis (x,y) v tabeli pomeni, da je kupec, ki je kupil album s šifro x kupil tudi album s šifro y.

Priprava baze

Z učilnice prenesite datoteko albums.sql. Datoteka vsebuje skrpito za pripravo vseh prej omenjenih tabel. Ker je datoteka velika (100 Mb), jo ne boste mogli izvesti iz workbench. Za zagon skripte uporabite ukaz »mysql –h localhost –u pb –p vaje < albums.sql« v ukazni vrstici. Pozor! Ustvarjanje baze lahko traja nekaj minut. Tabele so dostopne tudi na FRI strežniku pod shemo seminarpb.

Oddaja naloge

Pri reševanju naloge striktno upoštevajte navodila. Oddajte tri datoteke, ki jih poimenujete s svojo vpisno številko, in sicer v naslednjem vrstnem redu:

1. Poročilo v .pdf obliki
2. SQL koda (rešitve vseh nalog)
3. Python koda

Pri delu na skupnem MariaDB strežniku si naredite lastne, enolično poimenovane kopije originalnih tabel, najbolj priporočljivo pa je delo na lastnih inštalacijah.

Naloge

1. Naloga (DML) (30%)

Naloge rešujte na originalni/nespremenjeni bazi. Se pravi preden rešite naloge 2,3 in 4! Vse naloge rešujte z eno samo poizvedbo. Rešitve z večimi poizvedbami se štejejo za delno pravilne.

Nad podanimi tabelami izvedite naslednja povpraševanja:

- a) Za vsak album izpišite njegov ID in ceno.
- b) Enako kot a), le da brez albumov, ki nimajo podane cene.
- c) Za vsak album izpišite njegov ID, število medijev ter tip medija. Razvrstite padajoče po razmerju cena/število medijev.
- d) Med katerimi leti obstajajo zapisi v bazi? Napišite eno poizvedbo, ki vam poda odgovor.
- e) Za vsakega avtorja izpišite njegov ID, ime ter število albumov, ki jih prodaja. Izpis uredite padajoče po številu albumov.
- f) Poiščite imena založnikov, ki dosegajo najboljšo ceno glede na povprečno ceno prodanega albuma iz njihove založbe. Rezultat izpišite padajoče po tej povprečni ceni in to samo za tiste založnike, ki prodajajo vsaj 10 albumov s **podano** ceno.
- g) Nekateri albumi imajo na koncu imena albuma podana še dodatna določila v oglatih oklepajih. Koliko je takih albumov?
- h) Napišite proceduro Priporocam(*isci*). Procedura vzame niz *isci* ter poišče vse albume s tem imenom. Za vsak album, ki vsebuje to ime, izpišite ID albuma, njegovo ime ter imena albumov, ki so v bazi shranjeni kot priporočila za ta iskani album.
- i) *Za izvajalca U2 poiščite druge izvajalce, ki jih sistem priporoča. Rezultat uredite padajoče po številu prodanih albumov posameznega izvajalca.
- j) Tabela Recommendation ima približno 1.7 milijona zapisov. Ali lahko iskanje po tej tabeli pohitríte? Ste pohitrili prejšno poizvedbo?

2. Naloga (DDL) (30%)

- a) Ustvarite pogled razsirjeni_izvajalci, ki bo vseboval ID izvajalca, ime izvajalca ter število njegovih albumov. (5%)
- b) Ustvarite tabelo z enakimi podatki kot v pogledu iz točke a). (5%)
- c) Ustvarite bazne prožilce, ki bodo tabelo iz točke b) pravilno posodobili v primeru, da se spremeni tabela Album. (20%)

3. Naloga (DDL) (15%)

Napišite skripto, ki bo ustvarila novo tabelo GroupedData, in bo vsebovala podatke o letnici albuma, ceni albuma, številu medijev in šifrah tipov medija. Nato odstranite tabele, ki jih ne potrebujete več ter nastavite manjkajoče omejitve primarnih in tujih ključev.

4. Naloga (ODBC – Python) (25%)

Kot ste opazili imajo nekatera imena albumov dodatna določila med oglatimi oklepaji.

- a) S Pythonom se povežite na bazo in izločite vsa možna dodatna določila. Pazite, med oklepaji je lahko samo eno dodatno določilo (npr.: [Live]) ali več določil ločenih z vejico (npr.: [Extra tracks, Import, Original recording remastered]).
- b) Ustvarite tabelo ExtraDataNames(ExtraID, ExtraName), kjer je ExtraName eno izmed dodatnih določil, ExtraID pa njen ključ. Tabelo napolnite s pravilnimi podatki!
- c) Ustvarite tabelo ExtraData(#AlbumID,# ExtraID), kjer je AlbumID tuji ključ, ki kaže na tabelo Album, ExtraID pa tuji ključ, ki kaže na tabelo ExtraDataNames. Napolnite tabelo!