

Code : BFV4DBS1	Tentamen: Databases 1			
Datum: 2020-01-24	Tijd: 11:30-13:00	School: ILST		
Lokaal: ZP11 D1.08	Klas: BFV2	Duur: 90'		
Naam docent(en) : PARN Naam docent collegiale consultatie: (WERD) Tijdens het tentamen te bereiken: secretariaat ILST: 050 – 595 45 69		Aantal pagina's: 3 (incl. voorblad)		
<table><tr><td>Hulpmiddelen: kladpapier BIN-pc in tentamenmodus, incl. mysql documentatie 1920Hdatabases.sql</td><td>Overig hulpmiddelen: [OverigeH]</td></tr></table>			Hulpmiddelen: kladpapier BIN-pc in tentamenmodus, incl. mysql documentatie 1920Hdatabases.sql	Overig hulpmiddelen: [OverigeH]
Hulpmiddelen: kladpapier BIN-pc in tentamenmodus, incl. mysql documentatie 1920Hdatabases.sql	Overig hulpmiddelen: [OverigeH]			
Opgave inleveren: nee Kladpapier inleveren: nee				
Bijzonderheden: De mysql documentatie wordt beschikbaar gesteld in de mysql client en op je desktop. Pak dit archief uit, daarna kun je de bestanden benaderen met een browser				

Hernoem het bestand `voornaam_achternaam.sql`. Vervang `voornaam` door je voornaam en `achternaam` door je achternaam. Eventuele tussenvoegsels laat je weg. Ik zelf zou bijvoorbeeld het bestand moeten hernoemen naar `arne_poortinga.sql`.

Vul vervolgens je naam, studentnummer en de computer waar je achter werkt in op de aangegeven plaats.

De antwoorden van vragen 1 en 2 zet je op de aangegeven plaatsen in dit bestand. Let er op dat dit sql-script (want dat is het tenslotte) geen foutmeldingen genereert! Test dit voordat je het inlevert!

Het antwoord op vraag 3 bestaat uit een `.dia` bestand waarvan de naam dezelfde regels volgt maar waarvan de extensie `.dia` is (bijvoorbeeld `arne_poortinga.dia`).

Aan het einde van de toets lever je beide bestanden in via het submascript `"submit_your_work"`. Geef commando `"submit_your_work help"` om te zien hoe dit werkt.

Je cijfer wordt als volgt berekend: $(\text{aantal behaalde punten} / \text{maximaal te behalen punten}) * 10$. Mocht je minder dan een 1 gescored hebben, dan wordt dit 'afgerond' tot een 1. Je hebt een voldoende als je een 5.5 of hoger gescored hebt.

Om tijdens het tentamen in te kunnen loggen op de databaseserver moet je gebruik maken van de volgende gegevens (staan ook in `inloggegevens.txt` op je Desktop):

- user: tentamen
- database: Tentamen
- host: localhost (meegeven met `-h`)
- wachtwoord: tenta556

vraag 1 (40 pnt):

- Laat het script `1920Hdatabases.sql` uitvoeren en geef als antwoorden *de query* die je zou gebruiken om de vragen te beantwoorden, dus *niet* het *resultaat* van de query!
- Gebruik (tenzij er expliciet iets anders gevraagd wordt) de 'join-syntax' om tabellen aan elkaar te koppelen; het maken van een carthesisch produkt gevolgd door het filteren van de overtollige tupels m.b.v. de where-clausule (de 'where-syntax') wordt fout gerekend.
- Overtollige joins worden fout gerekend.
- Voer de query uit in de mysql client, dat voorkomt slordigheidsfouten.
- Verzamel de antwoorden in je sql script zoals boven aangegeven.

1a (2 pnt): Wat is de 'x_coord' de cel met het id 14?

1b (4 pnt): welke gene_expressions bevatten expressedata van de levercellen? Gebruik de 'where-syntax' en geef alleen het id.

1c (4 pnt): welke gene_expressions bevatten expressedata van de levercellen? Gebruik de join-syntax en geef alleen het id.

1d (5 pnt): Wat is de hoogste id in de tabel genes?

1e (5 pnt): Geef de celtypen waarbij de laagste expressiewaarde voorkomt.

1f (5 pnt): Geef de gene_symbols van de genen waarvan expressies gemeten zijn waarvan het celtype onbekend is. Sorteer op gene_symbols.

g (5 pnt): Wat zijn de gene_symbols van de genen die in levercellen voorkomen? Sorteer in omgekeerde volgorde.

1h (10 pnt): Geef per gen de gemiddelde expressiewaarde van dat gen in gestimuleerde bloedcellen. Geef alleen de gene_symbol en de gemiddelde expressiewaarde gesorteerd op expressiewaarde van hoog naar laag.

vraag 2 (30 pnt):

Voor deze vragen kan het zijn dat je meer dan één query nodig hebt. Als dat zo is geef dan de queries in de goede volgorde.

2a (15 pnt): De genen met de symbols '>gi|124505655 MAL1P4.06b' en '>gi|124505651 MAL1P4.04' blijken hetzelfde te zijn. Verwijder daarom het gen met het symbol '>gi|124505655 MAL1P4.06b' en ken de gene_expressions daarvan toe aan '>gi|124505655 MAL1P4.04'

2b (5 pnt) Verwijder de kolom sample uit de tabel cells.

2c (10 pnt) Maak gene_symbol een verplicht veld met als default waarde 'N.N.'

vraag 3 (30 pnt):

Maak een ontwerp in dia voor een database die de gegevens van de onderstaande informatiebehoefte kan bevatten.

- Geef primary (PK) en foreign keys (FK) duidelijk aan en let op het gebruik van de correcte uiteindes van de verbindinglijnen (de meer-kant moet dus een kraaienpootje zijn).
- Geef duidelijk (als commentaar in je schema) aan naar welke velden in welke tabellen de foreign keys verwijzen
- Geef ook bij elk veld het sql-datatype aan en of ze leeg (null) mogen zijn.

Voor een onderzoek naar de effecten van mutaties moet een database gebouwd worden. Mutatie zijn bijvoorbeeld het vervangen van een base door een of meer andere basen, deleties en inserties. Van elke mutatie wordt bijgehouden wat de originele waarde van de base(n) is en waardoor dit vervangen wordt en op welke genen zij zich voorgedaan hebben. Van de genen waar de mutaties in plaatsgevonden hebben moet de unieke naam (ook bekend als gene symbol) opgeslagen worden. Van elke mutatie wordt bijgehouden op welke positie in een gen hij zich voorgedaan heeft en wat dan zijn effect was, als dit bekend is. Dit is een korte omschrijving. Een gen kan meerdere mutaties bevatten.

=====Einde=====