IN2090 – Innleveringsoppgave 2
SQL 1

Publisert dato: 26.09.2024.

Innleveringsfrist: 11.10.2024 23:59.

Filer som skal leveres: 1: `innlevering2.sql`.

Antall forsøk: 2.

Les gjennom hele teksten nøye før du begynner å løse oppgavene. Oppgaven skal leveres som en ren tekstfil med filendelsen `.sql`. Bruk SQL-kommentarer for ren tekst og kommentarer.

Etter at romvesener har landet på Jorden (de fra forrige innleveringsoppgave) ønsker nå Departementet for Hemmelige, Utenomjordiske og/eller Farefulle Fenomener (HUFF) å finne ut hvor disse romvesnene kan ha kommet fra, altså hvilket solsystem og hvilken planet. De ønsker også å undersøke eventuelle muligheter for liv på andre planeter. Etter din formidable innsats i å hjelpe HUFF med ER-modellering, i forbindelse med romvesen-besøket tidligere, ønsker de nok en gang hjelp fra deg til dette!

Denne gangen trenger de hjelp til SQL og å hente ut informasjon fra et stort datasett de har fått samlet sammen. Databasen de har beskriver solsystemer, deres planeter, samt hvilke stoffer/molekyler som er oppdaget på disse planetene[1].

Databasen består av tre tabeller:

```
Stjerne(navn, [avstand], [masse])

Planet(navn, [masse], [oppdaget], [stjerne])

Materie(planet, molekyl)
```

[1]Dataene er ekte data om virkelige planeter og er hentet fra: `http://exoplanet.eu/catalog/`

Understrekede attributter angir primærnøklene. Attributter i klammeparenteser er attributter som ikke nødvendigvis har data, dvs. de kan inneholde `NULL`.

Tabellen `stjerne` beskriver stjerner, som identifiseres med deres `navn`, og har kolonnen `avstand` som inneholder stjernens avstand fra solen i parsecs, samt kolonnen `masse` som oppgir stjernens masse relativt til solen (hvor 1 betyr lik som solen, 2 betyr dobbelt så massiv som solen, 0.1 betyr en tiendedel av solens masse, osv.).

Tabellen `planet` beskriver planeter, som også identifiseres med dens `navn`, og inneholder også planetens `masse` relativt til Jupiter (den største planeten i vårt eget solsystem), et årstall som representerer når planeten ble `oppdaget`, samt en referanse til `stjernen` den går i bane rundt (altså er `planet(stjerne)` en fremmednøkkel som refererer til `stjerne(navn)`).

Tabellen `materie` beskriver hvilke stoffer som er oppdaget på hvilke planeter. Her er `planet` en fremmednøkkel som peker på `planet(navn)` og `molekyl` inneholder et molekyl som er oppdaget på planeten i `planet`-kolonnen. Merk at én planet kan ha mange molekyler, og kan derfor forekomme flere ganger i tabellen.

Oppgave 1

Oppgave 1 går bare ut på å kjøre et SQL-script som oppretter databasen du skal bruke i resten av oppgavene. For de resterende oppgavene skal du levere SQL-spørringene/kommandoene som løser oppgavene. Når du blir bedt om et antall, skal spørringen regne ut dette antallet.

Opprett en database på ditt eget database-område ved å følge en av de to alternativene under. I kommandoene må du bytte ut `brukernavn` med ditt UiO-brukernavn. I enkelte av kommandoene vil du bli bedt om å oppgi ditt UiO-passord.

Du trenger ikke levere det du har gjort i denne oppgaven, men oppgaven må gjøres før du kan gjøre de neste oppgavene. Om du har problemer er det lurt å sjekke PostgreSQL-guiden[2].

Med psql

For å bruke `psql` gjør du følgende:

a) last ned filen `planeter.sql` fra

[2] `https://www.uio.no/studier/emner/matnat/ifi/IN2090/h24/programvare/postgres.pdf`

`https://www.uio.no/studier/emner/matnat/ifi/IN2090/h24/innleveringer/planeter.sql`

til en mappe på ditt eget område på IFI (altså enten på en fysisk IFI-maskin, eller etter fjerninnlogging). Du kan eventuelt laste den ned via kommandolinjen ved å kjøre `wget <url>`, altså:

```
wget https://www.uio.no/studier/emner/matnat/ifi/IN2090/h24/innleveringer/planeter.sql
```

b) logge inn (fra samme mappe som filen over) på PostgreSQL med

```
psql -h dbpg-ifi-kurs03 -U brukernavn -d brukernavn
```

c) kjøre kommandoen `\i planeter.sql`

Med Edbit

For å bruke Edbit gjør du følgende:

a) last ned filen `planeter.sql` fra

`https://www.uio.no/studier/emner/matnat/ifi/IN2090/h24/innleveringer/planeter.sql`

til en mappe på din egen laptop, og last ned Edbit[3] til samme mappe

b) kjør Edbit ved å kjøre kommandoen under (fra samme mappe som filene over er lagt i)

```
java -jar edbit.jar -h dbpg-ifi-kurs03 -U brukernavn -d brukernavn \
     --sshHost login.ifi.uio.no --sshUser brukernavn
```

(Du vil nå bli bedt om ditt UiO-passord to ganger, én gang for fjerninnloggingen, og én gang for innlogging til databasesystemet.)

c) kjøre kommandoen `\i planeter.sql` fra Edbit

Etter dette er du innlogget i databasesystemet og har opprettet databasen beskrevet over. Du kan nå kjøre SQL-spørringer over databasen.

[3] `https://www.uio.no/studier/emner/matnat/ifi/IN2090/h24/programvare/edbit.jar`

Oppgave 2

Før HUFF er komfortable med å bruke dataene i databasen ønsker de å kontrollere korektheten av dataene ved å hente ut litt informasjon om noen solsystemer de allerede kjenner godt fra andre kilder. De ønsker også å vite hvor innholdsrik databasen er.

De ber deg derfor skrive SQL-spørringer som finner ut:

a) Navn på alle planeter som går i bane rundt stjernen med navn `Proxima Centauri`.

b) Hvilke årstall (uten duplikater) det ble oppdaget planeter i bane rundt stjernene med navn `TRAPPIST-1` og `Kepler-154`.

c) Antall planeter det er som ikke har oppgitt en masse (altså hvor masse er `NULL`).

d) Navn og masse på alle planeter oppdaget i 2020 med masse høyere enn gjennomsnittet av massen til alle planeter.

e) Antall år mellom første og siste oppdagede planet i `planet`-tabellen.

Oppgave 3

Basert på resultatet av spørringene dine er HUFF fornøyde med dataene og ønsker å gå videre med dem. De ønsker nå å begynne søket etter planeten som romvesnene som besøkte Jorden kom fra. De har ulike modeller som baserer seg på ulik informasjon, og din jobb blir å finne lister med planeter som passer hver modell. (Merk: Resultatene fra spørringene du skriver trenger ikke ha overlappende svar. Dersom det ikke er noen planet som passer alle modellene, kan det tyde på at romvesnene reiste fra en annen planet enn den de evolusjonært er tilpasset, altså at Jorden ikke er den første planeten de besøker.)

a) Den ene modellen baserer seg på anatomien til romvesnene: Biologer fra HUFF har nemlig funnet ut, basert på tykkelsen på bena og størrelsen på musklene, at de må komme fra en planet med masse mellom 3 og 10. Videre vet de at romvesnene inneholder vann (altså H2O), så planeten må inneholde dette molekylet. Skriv derfor en spørring som finner navn på alle planeter som passer denne informasjonen.

b) Den andre modellen baserer seg på romskipet og dens motor[4]: HUFFs ingeniører har nemlig funnet ut at romskipet høyst kan ha reist en avstand lik stjernens masse ganger `12`, altså må planeten gå i bane rundt en stjerne som har en avstand mindre enn `12` ganger dens masse. Videre har ingeniørene funnet ut at motoren bruker hydrogen som drivstoff, så planeten de reiser fra må ha molekyler som inneholder hydrogen (altså molekyler som inneholder bokstaven `'H'`). Skriv derfor en SQL-spørring som finner navn på planeter som passer informasjonen over.

c) Den tredje modellen er også laget av HUFFs ingeniører: Et alternativ til å bruke en massiv stjerne er å bruke to massive planeter for akselerasjon. Romvesnene kan ha reist fra en av to planeter fra samme solsystem, begge med masse større enn `10`, gitt at avstanden til deres solsystem er mindre enn `50`. Skriv en SQL-spørring som finner navn på alle planeter som passer med informasjonen over.

Oppgave 4

Nils, en av de ansatte på HUFF, ble svært imponert over hva du får til med SQL, og forsøkte seg derfor på å skrive en spørring selv. Nils ønsket å finne ut når planeter som går i bane rundt en stjerne med avstand større enn `8000` ble oppdaget, og skrev derfor denne spørringen:

```
SELECT oppdaget
FROM planet NATURAL JOIN stjerne
WHERE avstand > 8000;
```

Forklar Nils hvorfor spørringen hans ikke finner det han er ute etter og hvorfor den ikke gir noen rader i resultatet.

Oppgave 5

Datasettet brukt så langt inneholder kun stjerner og planeter utenfor vårt solsystem. HUFF ønsker nå at vår egen sol og egen planet (altså Jorda) skal legges inn i databasen, slik at disse kan bli med i fremtidige analyser og sammenlikninger. Skriv derfor SQL-kommandoer som:

a) Setter inn *Sola* i `stjerne`-tabellen, med `navn` lik `Sola`, `avstand` lik `0` og `masse` lik `1`.

b) Setter inn *Jorda* i `planet`-tabellen, med navn lik `Jorda`, `masse` lik `0.003146`, `oppdaget` lik `NULL`, og `stjerne` lik `Sola`.

[4]For romreise bruker man ofte et stort objekts gravitasjon for akselerasjon.

Oppgave 6

Nå som liv utenfor Jorden er bekreftet av de besøkende romvesnene, ønsker HUFF å utvide søket etter andre livsformer i rommet. De ønsker nå å lagre observasjoner av mulige tegn til liv i en egen tabell med navn `observasjon`, og ønsker hjelp av deg til å lage denne tabellen. Hver observasjon skal identifiseres med en egen `observasjons_id` som er et heltall. Hver observasjon skal også ha et tidspunkt med dato (altså et `timestamp`) knyttet til seg, samt en referanse til planeten observasjonen ble gjort på. Videre ønsker de å kunne lagre en kommentar til observasjonen som da bare er ren tekst (men en observasjon trenger ikke ha noen slik kommentar). Skriv en SQL-kommando som lager denne tabellen.

Levering

Innleverinsoppgaver leveres i Devilry[5]. Sørg for at du er registrert i systemet ved å logge inn og se at `innlevering2` er tilgjengelig som en oppgave for IN2090. *Sjekk dette før du begynner å løse oppgavene!*. Dersom du ikke er registrert, send en mail til `leifhka@ifi.uio.no`.

Etter innlevering vil du få tilbakemelding på innleveringen din, i likhet med en oblig. Selvom innleveringsoppgaven ikke er obligatorisk, er den et svært nyttig læringstilbud og en viktig del av undervisningen i kurset. Fokuser på å lære mest mulig og les tilbakemeldingene du får fra retterne nøye. Merk også at det ikke blir publisert løsningsforslag på innleveringsoppgavene, så å levere dem inn og deretter få tilbakemelding er eneste måte å få kjennskap til hvor god egen besvarelse er.

Du kan levere så mange ganger du vil, men det er kun den siste leveringen du vil få tilbakemelding på. Dersom du ønsker å levere et andre forsøk med tilhørende tilbakemelding, vennligst oppgi dette som en kommentar når du leverer, slik at retterne kan åpne for et andre forsøk etter å ha gitt tilbakemelding.

Lykke til!

[5] `https://devilry.ifi.uio.no/`