

Praktisk om Postgres og filmdatabasen

Januar 2014

Alle som tar ett av databasekursene INF1300, INF3100, INF5100 får tilgang til postgres-serveren `dbpg-ifi-kurs.uio.no`. Innlogging fra unix-systemer til postgres med `psql`-grensesnittet gjøres slik fra linuxmaskiner på Ifi-nettet:

```
> psql -h dbpg-ifi-kurs -U brukernavn -d dbnavn
```

der `brukernavn` er ditt vanlige UiO-brukernavn. Det er opprettet en database (et arbeidsområde) for hver enkelt bruker, denne har samme navn som brukernavnet. Hvis du derfor gir kommandoen

```
> psql -h dbpg-ifi-kurs -U brukernavn -d brukernavn
```

eller

```
> psql -h dbpg-ifi-kurs -U brukernavn
```

får du tilgang på et område der du kan opprette ditt eget databaseskjema.

Når du har logget inn, responderer systemet med promptet `"=>"`. Du kan da bl.a. skrive inn SQL-kommandoer og postgres-kommandoer. Alle SQL-kommandoer må termineres med semikolon. Glemmer du det, får du promptet `"->"`, noe som betyr at `psql` tror at kommandoen fortsetter på neste linje.

I stedet for å skrive lange SQL-queries rett inn i `psql` kan man skrive dem inn i en tekstfil, f.eks. `filnavn.sql`, og så kjøre dem med kommandoen

```
=> \i filnavn.sql
```

Hvis du får feilmeldinger av typen `ERROR: character 0xc2bd of encoding "UTF8" has no equivalent in "LATIN9"`, så prøv kommandoen

```
=> \encoding UTF-8
```

`psql` avsluttes med kommandoen

```
=> \q
```

Dokumentasjon: Nåværende versjon av postgres er 9.2. Postgres-dokumentasjon finnes på

<http://www.postgresql.org/docs/9.2/interactive/>

For mer informasjon om `psql`, se

<http://www.postgresql.org/docs/9.2/interactive/app-psql.html>

Informasjon for dem som vil bruke postgres fra windows eller hjemmefra:

Det enkleste er å koble seg opp med `ssh` til Ifis login-cluster (`login.ifi.uio.no`) og kjøre `psql` derfra. Bruk f.eks. PuTTY. Det er også mulig å bruke VPN.

Det finnes Windows installers for dem som bruker Windows:
<http://www.postgresql.org/download/windows/> [Open URL]
Installasjonsprogrammet fra EnterpriseDB installerer bl.a. pgAdminIII og psql.

Filmdatabasen

Det er to måter å få tilgang til filmdatabasen, disse beskrives hver for seg under.

A. Tilgang til filmdatabasens arbeidsområde

Du får direkte tilgang til filmdatabasens arbeidsområde ved å logge inn på

```
> psql -h dbpg-ifi-kurs -U brukernavn -d fdb
```

eller med '\c fdb' når du er inne på eget arbeidsområde. (Dette fungerer også i omvendt rekkefølge; når du er inne på 'fdb' kommer du tilbake til eget arbeidsområde ved å gi kommandoen '\c brukernavn'.) Du har da lesetilgang til databasen og kan kjøre spørringer mot den. Eksempel på spørring: "Hvor mange kinofilmer inneholder databasen?"

```
fdb=> select count(*) as Ccount
fdb-> from filmitem
fdb-> where filmtyp = 'C';
  ccount
-----
   549782
(1 row)
```

```
fdb=>
```

Ulempen med å gjøre spørringene herfra, er at du ikke får lov til å definere egne hjelpetabeller eller views på dette området.

B. Tilgang til filmdatabasen via eget arbeidsområde

Fra eget arbeidsområde (> psql -h kurspg -U brukernavn -d brukernavn) kan du også kjøre spørringer mot filmdatabasen. Fordelen ved å gjøre dette fra eget arbeidsområde, er at du da kan opprette lokale tabeller og views.

For å få aksess til filmdatabasen fra eget arbeidsområde, må du sende spørringen som argument til en funksjon som heter `dblink`. Denne utfører et fjernkall mot filmdatabasen, og resultatet av kallet blir returnert til det lokale arbeidsområdet. Eksempel:

```
select *
from dblink('dbname=fdb host=dbpg-ifi-kurs.uio.no user=ifikurs_ro
password=ifikurs_ro',
$$select count(*) as Ccount
from filmitem
where filmtyp = 'C';$$)
as t1(ccnt integer);
```

Spørringen (i doble dollartegn - \$\$select count(*) as Ccount from filmitem where filmtype = 'C';\$\$) blir utført på fdb's arbeidsområde og kan benytte det som der er definert av indekser og andre strukturer som fremmer rask prosessering. Merk at du må definere et alias (her kalt `t1`, men aliasnavnet kan du velge fritt) og angi sammen med aliaset hvilke attributter og tilhørende datatyper som `dblink` er forventet å returnere (her: `ccnt` som er integer). Vær oppmerksom på at hvis resultatmengden er stor, så kan overføringen/streamingen av resultatmengden mellom fdb og det lokale arbeidsområdet, ta *svært* lang tid. Derfor er spørringer med `dblink` bare egnet hvis resultatmengden er forholdsvis liten.

Utsnitt av filmdatabasen

For at du skal kunne eksperimentere med filmdatabasen og spørringer mot den, har vi laget et utsnitt¹ av filmdatabasen som du får din egen versjon av på eget arbeidsområde. Tabellene i utsnittet har samme navn som i den fulle filmdatabasen. De inneholder essensielt alle de filmene der `language` i tabellen `filmlanguage` har verdien 'France'². Det er få restriksjoner på hva du har lov til å gjøre med tabeller på eget område, som å endre eller slette tabelldefinisjoner og endre innhold, det gjelder også dette utsnittet(!).

Det kan lønne seg å prøve ut tyngre spørringer mot utsnittet først, og deretter, når du er rimelig sikker på at en spørring er riktig satt opp, prøve spørringen mot den fulle versjonen av filmdatabasen - enten via `dblink` fra eget arbeidsområde eller direkte fra fdb-arbeidsområdet. Man kan i `psql` enkelt skifte mellom de to arbeidsområdene ved å gi kommandoen `\c fdb` eller `\c brukernavn` (avhengig av hvilket arbeidsområde man for tiden er på).

¹ Utsnittet er på rundt 8% av den fulle versjonen. Du kan anta at tidsforbruket til spørringene vil bli redusert tilsvarende.

² Hommage á Igor Rafienko, David Ranvig og Rune Aske