

## Praktisk om Postgres og filmdatabasen

Februar 2010

Alle som tar et av databasekursene INF1300, INF3100, INF5100 får tildelt et eget passord mot postgres-serveren `kurspg`. Innlogging fra unix-systemer til postgres med `psql`-grensesnittet gjøres slik når man er på Ifi-nettet:

```
> psql -h kurspg -U brukernavn -d dbnavn
```

der `brukernavn` er ditt vanlige 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 kurspg -U brukernavn -d brukernavn
```

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

**For førstegangsbrukere:** Når dere har kommet så langt som prompten `"=>"`, dvs. når dere har logget inn på postgres, så skift passord med en gang. Det gjøres slik:

```
=> alter role brukernavn with encrypted password 'nyttpassord';
```

Ikke glem semikolonet (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
```

Merk at slike `psql`-spesifikke kommandoer ikke skal avsluttes av semikolon.<sup>1</sup>

`psql` avsluttes med kommandoen

```
=> \q
```

**Dokumentasjon:** Nåværende versjon av postgres er 8.4. Postgres-dokumentasjon finnes på <http://www.postgresql.org/docs/8.4/interactive/>

For mer informasjon om `psql`, se

<http://www.postgresql.org/docs/8.4/interactive/app-psql.html>

**Informasjon for dem som vil bruke kurspg fra windows:** Bruk en X-server, f.eks. X-win32, og kjør `psql` via denne.

**Informasjon for dem som vil bruke kurspg 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`, den finnes på Ifi-dvden.

---

<sup>1</sup> I Postgres 8.4 går det bra å avslutte også `psql`-kommandoer med semikolon.

## **Filmdatabasen**

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

### **A. Tilgang til filmdatabasens arbeidsområde**

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

```
> psql -h kurspg -U brukernavn -d fdb
```

Dere 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 dere 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 dere også kjøre spørringer mot filmdatabasen. Fordelen ved å gjøre dette fra eget arbeidsområde, er at dere da kan opprette lokale tabeller og views.

For å få aksess til filmdatabasen fra eget arbeidsområde, må dere 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 user=fdb password=fdb',
           $$select count(*) as Ccount
              from filmitem
              where filmtyp = 'C';$$)
as t1(ccnt integer);
```

Spørringen (i doble dollartegn - \$\$select count(\*) as Ccount from filmitem where filmtyp = 'C';\$\$) blir utført på fdbs arbeidsområde og kan benytte det som der er definert av indekser og andre strukturer som fremmer rask prosessering. Merk at dere må definere et alias (her kalt `t1`, men aliasnavnet kan dere 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 dere skal kunne eksperimentere med filmdatabasen og spørringer mot den, har vi i tillegg laget et utsnitt<sup>2</sup> av filmdatabasen som dere får deres 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'`<sup>3</sup>. Det er få restriksjoner på hva dere har lov til å gjøre med tabeller på eget område, som å endre tabelldefinisjoner og innhold, det gjelder også dette utsnittet.

Det kan lønne seg å prøve ut tyngre spørringer mot utsnittet først, og deretter, når dere er rimelig sikre 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 du for tiden er på).

## **Referanser**

[1] The Internet Movie Database, <http://www.imdb.com/>

---

<sup>2</sup> Utsnittet er på rundt 8% av den fulle versjonen. Dere kan rimeligvis anta at tidsforbruket til spørringene vil bli redusert minst like mye som dette.

<sup>3</sup> Hommage á Igor Rafienko, David Ranvig og Rune Aske