**Harjoitus 1 - Datatyypit**

Big Data -ympäristöt

Tapani Alastalo

**Sisältö**

No table of contents entries found.

# Vaihe 1

## Mitä tauluja SQL-tiedoston rakenne sisältää ja mikä on taulujen rakenne?

Tiedosto sisältää taulut DEPT, EMPS ja SDEPT.

DEPT koostuu merkkijonoista (varchar) DEPT, MANAGER ja DUTY, sekä decimaaliarvosta (decimal) BUDGET.

EMPS koostuu useammista merkkijonoista (char ja varchar) ja decimaaliarvoista (decimal).

SDEPT koostuu merkkijonoista (varchar) DEPT ja MANAGER, sekä decimaaliarvosta (decimal) CODE.

## Sisältääkö tiedosto myös tauluihin tallennettavaa dataa?

Kyllä. DEPT sisältää kymmeniä esimerkki syötteitä. EMPS sisältää 500 esimerkki syötettä ja SDEPT kolme esimerkki syötettä.

## Jos sisältää, niin anna pari esimerkkiriviä niistä tauluista, joihin dataa tallennetaan.

DEPT:

('A','SMITH A','ACCOUNTING',100000.00), ('B','JONES B','INF SYSTEMS',120000.00)

EMPT:

(1,'A','KOO','SARA','234 WEST ST','BANFF','AB','T0R0R0','421-6923',6.00,0,100.00),(2,'B','MARSH','JOHN','456 EAST AVE','BANFF','AB','T0R0R0','963-2176',NULL,1,75.00)

SDEPT:

('A','SMITH A',1),('B','JONES B',2)

# Vaihe 2

## Vertaile avaamaasi tiedostoa rakenteen osalta sekä tekstieditorissa että Excel/CSV Viewer -ohjelmistossa. Miten tiedoston rakenne näyttäytyy tekstieditorissa ja kuinka Excel/CSV Viewer -ohjelmistot sen esittävät?

Tekstieditorissa tiedoston rakenne on rivitettyä pilkulla eroteltua tekstiä, minkä esimerkiksi Excel osaa jäsennellä taulukoksi, missä pilkku toimii erottimena rivin tekstin jakamisena sarakkeisiiin.

## Millaista dataa tiedosto sisältää ja mikä on mahdollisesti ollut sen alkuperäinen käyttötarkoitus?

Tiedosto sisältää myynnissä olleiden asuntojen tietoja, kuten osoite, kaupunki, postinumero, osavaltio, makuuhuoneiden lukumäärä, hinta, myyntipäivä ja koordinaatit.

Datan käyttötarkoitus on mahdollisesti ollut myytyjen asuntojen hinnoittelu esimerkiksi kaupungeittain. Lisäksi on voitu vertailla makuuhuoneiden lukumäärän vaikutusta hintakehitykseen eri alueilla.

# Vaihe 3

## Vertaa tätä tiedostoa edellisen tehtävän CSV-tiedostoon. Miten rakenne poikkeaa?

## Millaista dataa tiedostoon on tallennettu?

## Anna muutaman rivin esimerkki siitä millaiselta aiemmassa tehtävässä esitetty CSV-data näyttäisi XML-muodossa.

# Vaihe 4

## Millainen datarakenne on kahteen aiempaan semirakenteelliseen datatyyppiin nähden?

## Millaista dataa tiedostoon on tallennettu?

## Luo esimerkki JSON-tiedosto, johon on tallennettu kaksi riviä toisen vaiheen CSV-tiedostosta ja kolme riviä kolmannen vaiheen XML-tiedostosta.

# Vaihe 5

## Millaista dataa kuva sisältää?

## Onko siitä havaittavissa minkäänlaista rakenteellista piirrettä?