

**Tietojen tallennusrakenteet**

- Jokaisella tiedostolla on **otsake (header)**, joka sisältää tiedostoon liittyvää hallintatietoa
  - tiedot tiedostoon kuuluvista lohkoista esim. taulukkona, joka voi muodostua ketjutetuista osista
  - tietoja tietueiden rakenteesta
  - muuta hallintatietoa

1

**Tietojen tallennusrakenteet**

- Tiedoston sivut sisältävät ainakin:
  - sivuotsakkeen
    - sivun hallintaan tarvittavaa tietoa
  - tietuealueen
    - varsinaiset tietueet ja vapaa tila
  - tietuehakemiston
    - tietueiden sijaintitiedot sivulla

2

**Tietojen tallennusrakenteet**

- Sivun alussa sijaitseva **sivuotsake (page header)** sisältää
  - sivutunnisteen** (page identifier)
    - yleensä sivun järjestysnumero tiedostossa, on tiedostokuvaajan perusteella muunnettavissa fyysisiksi levyosoitteeksi
  - tietuehakemiston alkioiden lukumäärän
  - vapaan tilan hallintaan liittyvää tietoa
  - mahdollisesti sivun tyyppin
  - tiedon siitä mihin tiedostoon sivu kuuluu (sisäinen tunniste)
  - mahdollisesti linkkejä muille sivuille (esim. jatkosivun tunniste)**
  - käyttötietoa**

3

**Sivun rakenne tyypillisesti**

tietueiden tila ja tietuehakemisto kasvavat vastakkaisiin suuntiin

node id, file\_id, page\_no, record\_index, address

4

**Sivun rakenne**

- Tietueiden talletusalue alkaa heti sivuotsakkeen jälkeen
- Tietuehakemisto on taulukko, jossa on kunkin sivulle talletetun tietueen alkukohta
- Tietueen talletukseen käytetty tila ja tietuehakemistolle varattu tila kasvavat vastakkaisiin suuntiin.

5

**Sivun rakenne**

- Tietueen tunniste (record identifier, RID) muodostaa pari
  - sivutunniste
  - tietueen indeksi (tietuehakemistossa)
- Tuple identifier (TID) = record identifier

6

**Tietueen rakenne**

- Samassa tiedostossa olevat tietueet voivat olla keskenään
  - saman pituisia, eli kiinteäpituisia (fixed length)
  - vaihtuvapituisia (variable length)
    - erityyppisiä esim. kurssitietueita ja osallistujatietueita
    - saman tyyppisiä, mutta tietokenttien pituus tai määrä vaihtelee
    - tietueessa voi olla kiinteäpituisia tai vaihtuvapituisia kenttiä, yksikin vaihtuvapituinen kenttä tekee tietueesta vaihtuvapituisen

7

**Tietueen rakenne**

- Kiinteäpituisuus on jokaisessa tietueessa saman pituinen
  - smallint, integer, float, double, date ja timestamp tyyppiset arvot tallennetaan tyyppillisesti (ei kuitenkaan välttämättä) kiinteäpituisiin kenttiin – kentän pituus riippuu tietotyyppistä. Esimerkiksi kokonaisluvut voitaisiin tallentaa binäärisinä 4 tavun pituisina kenttinä
  - SQL:n tietotyyppi char määrittelee kiinteäpituisen merkkijonon. Sillekin kiinteäpituisuus soveltuisi

8

**Tietueen rakenne**

- Vaihtuvapituisia kenttiä käytetään tyyppillisesti tilanteissa, joissa tietoalkioiden arvot vaihtelevat merkittävästi pituudeltaan
  - Esimerkiksi vaihtuvapituiset merkkijonot
  - Yleensä vaihtuvapituisilla kentillä pyritään säästämään tilaa, mutta tilansäästöä voidaan saada muutenkin, esimerkiksi tiivistämällä tietueet

9

**Tietueen rakenne**

- Relatioon employee(name, ssn, salary, address) monikon ('Smith, John', '010263-189F', 17000, 'Park Avenue ...') esittäminen:
- kiinteänmittaisilla kentillä (alkukohta laskettavissa kun tiedetään kenttien pituudet):

20 tavua	11 tavua	4 tavua	30 tavua
Smith, John	010263-189F	17000	Park Avenue ...
0	20	34	38

10

**Tietueen rakenne**

- Vaihtuvanmittaiselle kentälle erilaisia esitystapoja:
- erotinmerkkien avulla

4	Smith, John	\$	010263-189F	\$	17000	\$	Park Avenue ...	\$
1	13	24	30					

• pituuskentän avulla

4	11	Smith, John	11	010263-189F	5	17000	25	Park Avenue ...
1	2	1	14	25	31			

kenttien lukumäärä


11

**Tietueen rakenne**

- siirtymätaulukon avulla


4	5	16	27	32	Smith, John	010263-189F	17000	Park Avenue ...
1	5	16	27	32				

12

 Tietueen rakenne


- Usein tietueissa on sekä kiinteän että vaihtuvan pituisia osia, tällöin
  - kiinteänmittaiset ovat yleensä tietueen alussa kiinteissä paikoissa
  - vaihtuvanmittaisen osan alkukohta samoin kenttien järjestys (vähintään) on kuvattava esim. sivun otsikkotietueessa

13

 Tietueen rakenne


- Oraclen tietuerakenteessa kaikki kentät ovat periaatteessa vaihtuvapituisia/valinnaisia
  - Kentät esiintyvät tietueessa siinä järjestyksessä, missä ne on esitelty 'create table' -lauseessa (myöhemmät lisäykset tietueen loppuun).
  - Kunkin kentän alussa on **pituus** 1-3 tavua
    - alle 250 pituisten pituus ilmaistaan yhdellä tavulla ja yli 250 kolmella tavulla, ensimmäisen pituustavun arvo 255 kertoo, että 2 seuraavaa sisältävät pituuden
  - Jos kenttä on tyhjä (null), niin pituus on 0.
  - Oracle tietueessa on lisäksi alussa otsakeosa (version 8.0 laskukaavojen mukaan 3 tavua)

14

 Tiedostorakenteet


- Tiedostojen tehokkuutta yhtä kyselyä kohti arvioidaan usein tarvittavien **levyhakujen määrällä**
  - kuten levykäsittelyn yhteydessä todettiin levyhakuja on kahden tyyppisiä
    - satunnaishakuja ja
    - peräkkäishakuja

15

 Tiedostorakenteet


- Kokonaistehokkuuteen vaikuttaa myös puskureiden määrä
- 'oikean levyhaun' tarpeen todennäköisyys vähenee puskureiden määrän kasvaessa, koska todennäköisyys sille, että tarvittava sivu on jo puskurissa kasvaa
- Levyjen ominaisuudet ja verkkoyhteyksien nopeus vaikuttavat kokonaistehokkuuteen.

16

 Tiedostorakenteet

- Levyhakujen määrä on yksinkertainen ja selkeä tehokkuusmitta:
  - ei riipu laitteistosta
  - riippuu tiedostorakenteesta
  - arvioinnissa voi käyttää keskiarvoja, maksimia, jne. [vrt. tietorakenteiden analyysi](#)

17

 Tiedostorakenteet

- Tiedostorakenteen tasolla vaikuttavia tekijöitä ovat mm.
  - **tietueiden järjestys**
  - **tietueiden sijainti suhteessa toisiinsa**
    - Tyypillisesti kunkin relaatiotietokannan taulun tiedot muodostavat oman tiedostonsa. Joissakin järjestelmissä, esimerkiksi Oraclessa, tietueita voidaan haluttaessa **ryvästää** (cluster) siten, että usein yhdessä tarvittavat tietueet (vaikkapa opiskelijatietue ja opiskelijan suoritus-tietueet) sijoitetaan lähemmäksi toisiaan samalle sivulle
  - **miten tietueita kytketään**
    - tiedoston sisällä
    - tiedostojen välillä
    - käytetäänkö suorita **osoitepohjaisia kytkentöjä** vai ko relaatiokantojen **avainperustaisia epäsuoria kytkentöjä**

18

**Tiedostorakenteet**

- Tiedostorakenteet tarjoavat tiedostojen käsittelyyn varsinaisten datatietueiden sijoittelun lisäksi erilaisia apurakenteita (indeksejä), joilla pyritään nopeuttamaan käsittelyä
- Relaatiotietokantojen avainperusteiset viiteavainkennät vaativat apurakenteita toimiakseen tehokkaasti
- Apurakenteet vaativat levytilaa, puskuritilaa, omia ylläpitoimia lisäysten ja muutosten yhteydessä
- Apurakenteiden käyttömahdollisuudet riippuvat datatietueille valitusta organisaatiosta

19

**Tiedostorakenteet**

perusrakenne = varsinaisten datatietueiden tiedosto

Apurakenteina indeksit tarjoavat tehokkaita saantipolkuja datatietueisiin

Indeksien ylläpito aiheuttaa kustannuksia

20

**Tiedostorakenteet**

- Tietokannan kokonaistehokkuuteen vaikuttaa myös erilaisten tietokantaoperaatioiden osuus ja jakautuma kokonaiskuormassa
  - paljonko on tietyn tyyppisiä kyselyjä, milloin
  - paljonko on lisäyksiä, muutoksia ja poistoja ja miten ne jakautuvat
- Operaatioiden tarpeet voivat olla ristiriitaisia
- Kuorma ei välttämättä pysy samanlaisena vaan muuttuu ajan myötä

21