

Tieto - data

- n Digitaalisesti tallennettua informaatiota jostakin kohteesta
 - n Vapaamuotoinen tieto (unformatted)
 - n Esim. teksti, puhe, kuvat, ...
 - n Sisältö jäsentämätöntä
 - n Koneellinen käsittely vaikeaa, paitsi kokonaisuutena
 - n Jotain määrämuotoisuutta esiintyy, esim. merkkien esitystapa, kuvapisteiden koodaus
 - n Rakenteinen tieto (structured)
 - n Koneellinen käsittely helppoa
 - n Jäsenneiltyä
 - n Pääsy tiedon osiin
 - n Tällä kurssilla käsitellään rakenteista tietoa

1

Tietokanta

- n Tietokanta (database)
 - n jotakin käyttötarkoitusta varten laadittu kokoelma toisiinsa liittyviä säilytettäviä tietoja

2

Tietomallit

- n Tietomallilla (data model) tarkoitetaan tiedon rakenteen ja tiedolle suoritettavan käsittelyn määrittelevää käsitteistöä
- n Tietoa voidaan tarkastella eri näkökulmista - eri abstraktiotasoilla
- n Perinteinen jako
 - n Käsitetaso
 - n Rakennetaso (tai esitystaso)
 - n Talletustaso

3

Tietoabstraktiot

- n Abstraktiotasojen ajatuksena tiedon mallinnuksessa on jakaa tiedon kuvaaminen ja määrittely osiin ja täten yksinkertaistaa kuvauksia, kun mukana ei ole kuvauksen käyttötarkoituksen kannalta ylimääräistä informaatiota

4

Näkökulmat tietoon

- n Käsitetaso, kohdetaso (conceptual level, real word level):
 - n Mitä tietoja käsitellään
 - n millainen on se kohde, jota tiedoilla pitäisi kuvata
 - n asiat, joita pitäisi esittää
 - n mitä tietoelementtien arvot tarkoittavat
 - n Miten tiedot liittyvät yhteen
 - n Mitä sääntöjä tietoihin liittyy
 - n käsittelyohjelmistoista riippumaton kuvaus tiedoista

5

Näkökulmat tietoon

- n Rakennetaso, looginen taso, esitystaso (structural level, logical level):
 - n Minkälaisia käsiteltäviä rakenteita tiedot muodostavat - esim. eri ohjelmointikielten näkemykset tiedosta
 - n Millaisina rakenteina ohjelmoijat ja tietokannan suoraikäyttäjät näkevät tiedot

6

Näkökulmat tietoon

- n Talletustaso, fyysinen taso (**physical level**):
 - n Minkälaisina koneenläheisinä teknisinä rakenteina tiedot tallennetaan ja miten niitä voidaan käsitellä
 - n Tietueiden muoto, tiedon koodaustapa, salakirjoitus, tiivistystekniikka, ...
 - n Minkälaiset rakenteet tehostavat tiedon hakua
 - n Onko tiedot hajautettu vai ovatko kaikki keskitetysti samassa paikassa

7

Abstraktiotasot

The diagram illustrates three levels of abstraction for data storage:

- käsitetaso** (Conceptual level): A thought bubble containing the text "On tietoja autoista ja henkilöistä. Henkilöistä tallennetaan nimi. Autoista tallennetaan tieto omistajasta, joka on henkilö." (There is information about cars and people. From people, names are stored. From cars, information about the owner, who is a person, is stored.)
- rakennetaso** (Logical level): A thought bubble containing "Taulu AUTO(rekno, malli)" and "Taulu HENKILO(nimi,...)". (Table AUTO (license plate, model) and Table HENKILO (name,...))
- talletustaso** (Physical level): A thought bubble containing "Tiedosto Auto levyllä X1135 toteuttaa taulun auto" and "Indeksi AiX (auto, rekno)". (File Auto on disk X1135 implements the table auto and Index AiX (auto, license plate)).

8

Abstraktiotasot tiedon käsittelyssä

käsitetaso	tietosisällön suunnittelu
rakennetaso	kyselyt, ohjelmointi, rajapinnat
fyysinen taso	Tallennustapa, viritys, hajautus, tehokkuus

9

Mitä malleja on tarjolla?

käsitetaso	<ul style="list-style-type: none"> • Entity-Relationship mallit (ER) • Oliomallit (UML) • Semanttiset tietomallit
rakennetaso	<ul style="list-style-type: none"> • Relatiomalli • Oliomallit • XML • Hierarkkinen malli • Verkkomalli • Muita
fyysinen taso	<ul style="list-style-type: none"> • Toimittajakohitteiset käsitteet

10

Mallinnuksen kohteet

- n Monitasoinen tiedon mallinnus ei liity vain tietokantoihin.
 - n Esimerkiksi tietoliikenteessä voidaan tarkastella välitettävien viestien
 - n sisältöä käsitetasolla,
 - n muotoa rakennetasolla ja
 - n teknistä toteutusta yhtenä tai useampana talletustason abstraktiona.

11

Tietokanta vs. tiedosto

- n Edellä tietokanta määriteltiin yleisesti tietokokoelmana, jatkossa tarkastellaan ominaisuuksia, joita tietokannalle yleensä asetetaan verrattuna perinteiseen ohjelmointikielten **tiedosto** käsitteeseen.
- n Tarkastelun kohteena ovat rakenteiset tiedot

12

Tiedosto

- Ohjelmointikielissä apumuistiin tallennettuja tietoja käsitellään tiedostoina (file, data stream).
- Tiedostoja on rakenteisia (structured) ja vapaamuotoisia (text file).
- Rakenteinen tiedosto muodostuu tietueista (record). Tietue puolestaan jakautuu kenttiin (field)

13

Rakenteinen tietue

- Rakenteisessa tietueessa kenttien sisältöön päästään käsiksi kentän tunnuksen tai sijainnin perusteella
 - merkkipositioissa 10-15 sijaitseva tieto
 - kolmannessa kentässä sijaitseva tieto
 - kentän SYNTYMA-AIKA sisältö
- Alimmalla toiminta-abstraktion tasolla tiedon käsittely perustuu aina tiedon sijaintiin
- Kiinteämuotoisissa tietueissa kentän tunnuksen ja sijainnin välisen yhteyden tietää ohjelmointikielen kääntäjä, vaihtumuotoisissa yhteys ratkaistaan suoritusajaisesti tietueisiin sisältyvän **rakennetiedon** perusteella.

14

Rakenteinen tietue

Perinteinen:

XML:

```
<henkilotietue>  
<hnumero>2345</hnumero>  
<etunimi>Arttu</etunimi>  
<sukunimi>Aho</sukunimi>  
</henkilotietue>
```

Rakennetietona elementtien alku ja loppumerkit

15

Tietokanta - tietoriippumattomuus

- Perinteisissä ohjelmointikielissä käsiteltävien tiedostojen **tietuerakenne** määrittellään osana ohjelmaa.
 - miten taataan yhtenäisyys samoja tietoja käyttävien ohjelmien välillä ?
 - miten toisella kielellä kirjoitetun ohjelman pitäisi määrittellä tiedot ?

16

Tietokanta - tietoriippumattomuus

- helppoa, jos A ja B on laadittu samalla ohjelmointikielellä (include)

17

Tietokanta - tietoriippumattomuus

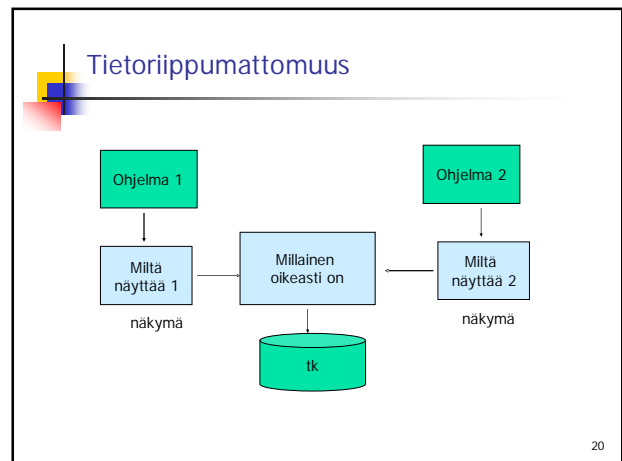
- Tietokannoissa tiedon kuvaus on irrallaan ohjelmista **tietokantakaaviona**

18

Tietokanta - tietoriippumattomuus

- n Kun jokin ohjelma tarvitsee tiedostoon lisätietoja (tietueisiin uusia kenttiä)
 - n perinteisiä tiedostoja käytettäessä kaikki muuttunutta tiedostoa käsittelevät ohjelmat on muutettava ja käännettävä uudelleen,
- n Ohjelman tietoriippumattomuus:
 - n Ohjelma täytyy kääntää uudelleen vain, jos sen itse tarvitsemien tietojen rakenne muuttuu
 - n Tietokannan kuvaukseen sisältyvät erilaiset näkymät tietoon mahdollistavat tietoriippumattomuuden
 - n Eri ohjelmilla voi olla erilainen käsitys rakenteesta

19



Tietokanta / samanaikainen käyttö

- n Perinteiset tiedostot
 - n yksi käyttäjä kerrallaan saa haltuunsa koko tiedoston
- n Tietokannat tyypillisesti
 - n useita samanaikaisia käyttäjiä
 - n käyttäjät häiritsevät toisiaan mahdollisimman vähän – muut käyttäjät eivät välttämättä näy käyttäjälle

21

Tietokannat / hakumahdollisuudet

- n Tiedosto
 - n peräkkäiskäsittely = käy läpi kaikki tietueet siinä järjestyksessä, jossa tietueet sijaitsevat tiedostossa
 - n haku osoitteen (esim. tietueen järjestysnumero) perusteella (*hae kymmenes tietue*)
- n Tietokanta
 - n kohteena annetun hakukriteerin täyttävät
 - n hakukriteeri perustuu tiedon sisältöön ei sijaintipaikkaan (*hae tietueet, joissa kentässä palkka oleva arvo <2000*)

22

Tietokanta / suojaus

- n Käyttöjärjestelmä
 - n suojaa tiedostoja ja hakemistoja
 - n luku-, kirjoitus-, suoritusoikeus koko tiedostolle
- n Tietokannat
 - n täsmäsuojaus, jopa kentätasolla
 - n käyttöhistoria operaatiotarkkuudella niin haluttaessa

23

Tietokanta / tapahtumakäsittely ja elvytys

- n Tiedosto
 - n käyttäjän tai sovellusohjelman on huolehdittava varmuuskopiointista, varmistuksista, operaatioiden peruutuksista, virheistä toipumisesta
- n Tietokanta
 - n takaa 'vahvistetun' operaation läpimenoa
 - n operaatiosarjan peruminen mahdollista vahvistukseen asti
 - n elpyminen virhetilanteista

24

Tietokanta / suuret dynaamiset tietomäärät

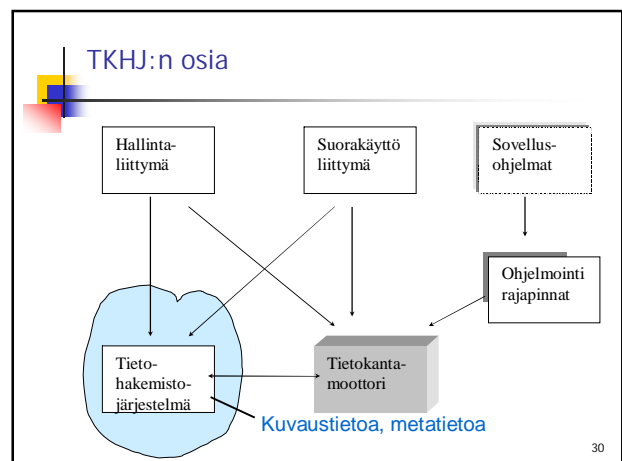
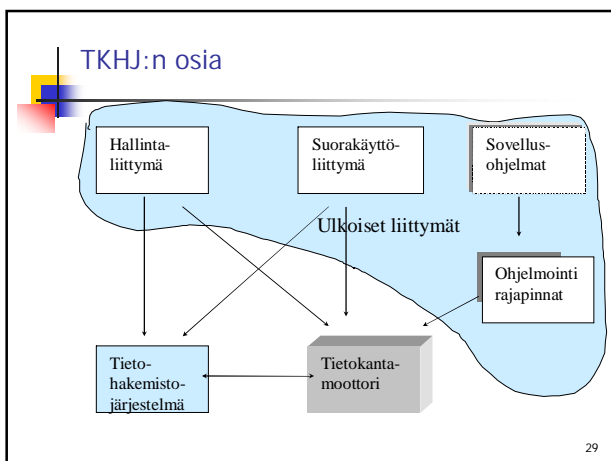
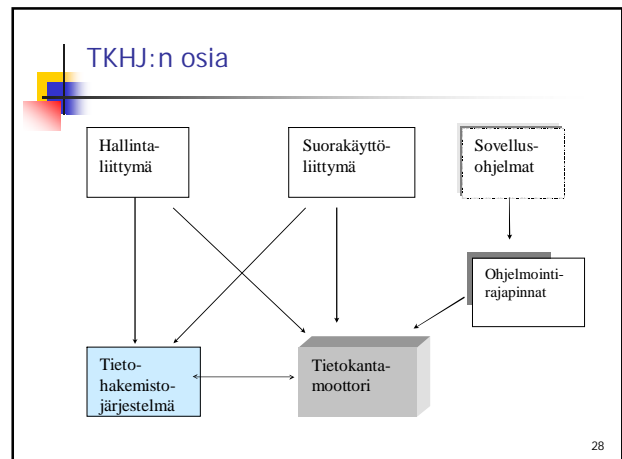
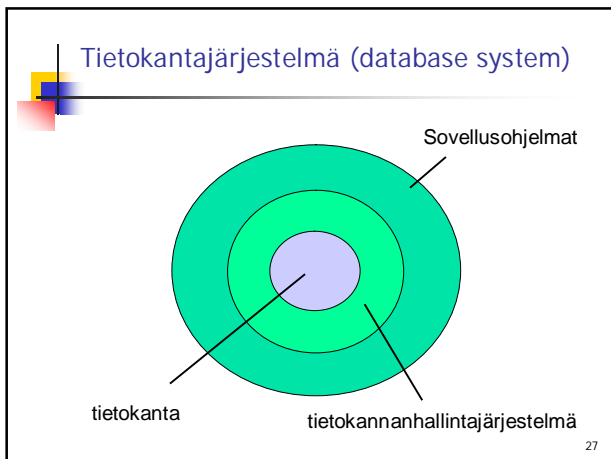
- n Kasvavat tietomäärät edellyttävät mukautuvia ja tehokkaita talletusrakenteita
 - n nopea haku annetun kriteerin perusteella
 - n tiedoston koon kasvaminen ei saa merkittävästi vaikuttaa tiedonsaannin aikaan
- n Rakenteiden viritys tehokkaammiksi ilman, että ohjelmia joudutaan muuttamaan

25

Tietokanta / tietojen väliset riippuvuudet

- n Tietokanta kuvaa reaali maailman ilmiöitä
- n Reaali maailmassa vallitsevien sääntöjen (esim. **Henkilöllä vain yksi puoliso kerrallaan**) on oltava voimassa myös reaali maailmaa kuvaavassa tietokannassa
 - n jos sääntöjä valvotaan sovellusohjelmissa, on valvonta sisällytettävä jokaiseen tietojä muokkaavaan ohjelmaan
 - n vaihtoehtona erillinen sovellusohjelmien ulkopuolinen valvontamekanismi (tietokanta valvoo)

26



Tietokantamoottori

- n Pääsyn valvonta (authorization control)
 - n tarkastaa käyttäjien oikeudet operaatioihin
- n Kyselyn optimoija (query optimizer)
 - n laatii toteutussuunnitelman halutuille tietokantaoperaatiolle – miten tehdään
- n Transaktion hallinta (transaction manager)
 - n valvoo samanaikaisia operaatioita
 - n varaa ja vapauttaa resursseja,
 - n huolehtii operaatioiden päättymisestä ja mahdollisista peruutuksista.

31

Tietokantamoottori

- n Eheyden valvonta (integrity control)
 - n valvoo, etteivät tietokantaan kohdistuvat muutokset riko tietokannalle määriteltyjä oikeellisuussääntöjä.
- n Suorittaja (command processor)
 - n ohjaa operaatioiden suoritusta
- n Puskurien hallinta (buffer management),
 - n vastaa keskusmuistin ja apumuistin välisestä tiedonsiirrosta
- n Hakumenetelmät (access methods)
 - n suorittaa tiedonhaut toteutussuunnitelman mukaisesti

32