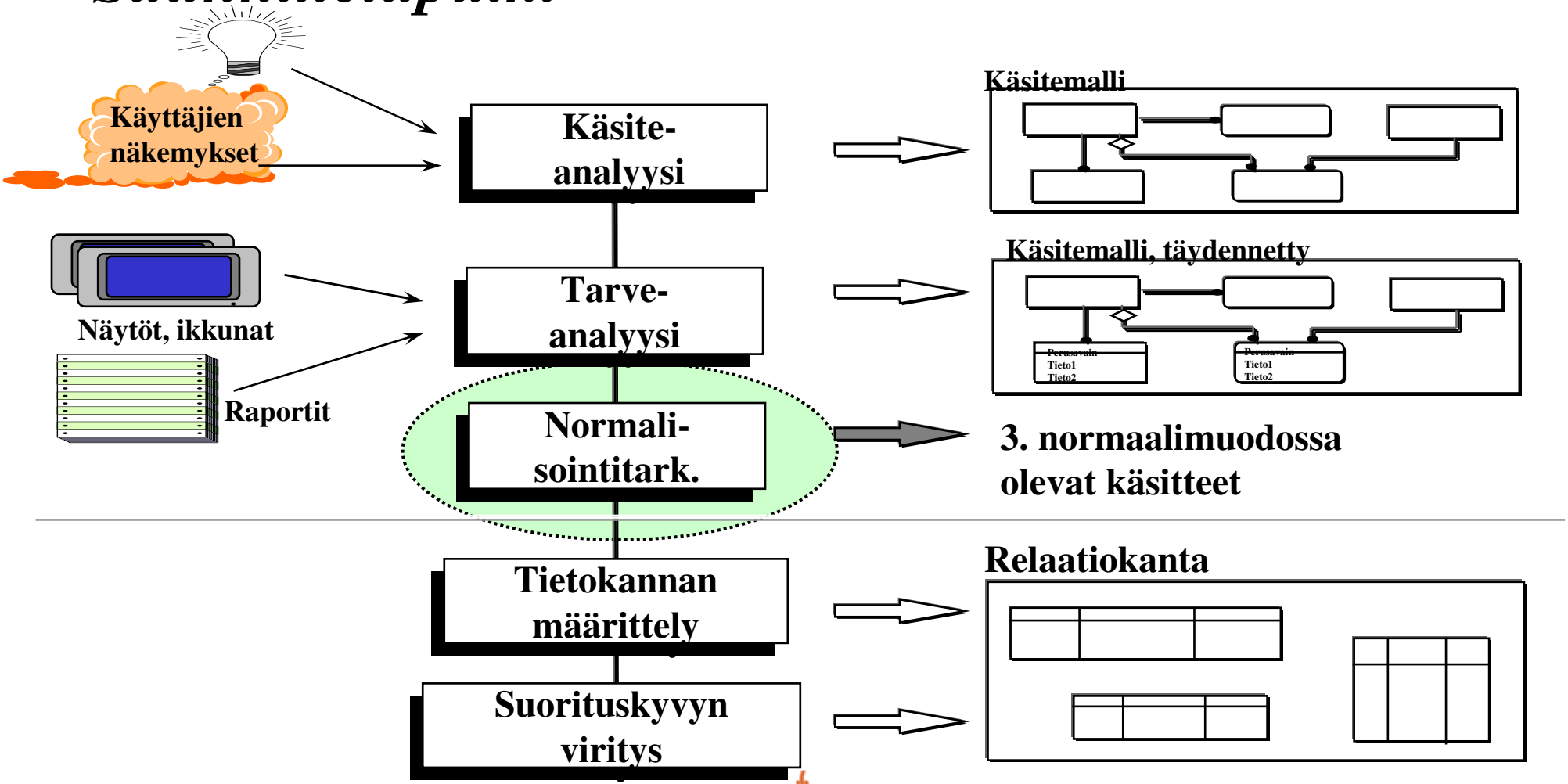
IIO30100 Tietokantojen suunnittelu

Normalisointi

kirjan Hovi, Huotari, Lahdenmäki: Tietokantojen suunnittelu & indeksointi, Docendo (2003, 2005) luku 5

© Jouni Huotari & Ari Hovi

Suunnitteluputki



JYVÄSKYLÄN AMMATTIKORKEAKOULU

Normalisoinnin tarkoitus

- Normalisointi (normalisation) on menetelmä, jonka avulla tietorakenteita voidaan jalostaa ”parempaan” tallennusmuotoon
- Parempi tarkoittaa tässä yhteydessä rakennetta,
  - jossa on tietojen toistaminen (redundanssi) minimoitu
  - joka on tehokas päivitysten kannalta
  - joka on helpompi pitää yhdenmukaisena, sillä tiedot tarvitsee päivittää vain yhteen paikkaan
  - joka on muutosjoustava

Normalisointivaiheet

1. Normaalimuoto

Erota toistuvat ryhmät ja moniarvoiset sarakkeet omaksi käsitteekseen

2. Normaalimuoto

Jokaisen ei-avaintiedon tulee olla riippuvainen koko perusavaimesta

3. Normaalimuoto

Poista sisäiset (ei-avaimeen kohdistuvat) riippuvuudet

1. Normaalimuoto

HENKILÖ

Htun	Nimi	Palkka	Pvm_1	Palkka_1	Pvm_2	Palkka_2

1. Normaalimuoto :

Poista toistuvat ryhmät ja moniarvoiset sarakkeet

HENKILÖ

Htun	Nimi	Palkka

PALKKAHISTORIA

Htun	Pvm	Palkka

JYVÄSKYLÄN AMMATTIKORKEAKOULU

1. Normaalimuoto (versio 2)

HENKILÖ

Htun	Nimi	Palkat	Pvmt

1. Normaalimuoto :

Poista toistuvat ryhmät ja moniarvoiset sarakkeet

HENKILÖ

Htun	Nimi	Palkka

PALKKAHISTORIA

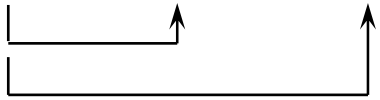
Htun	Pvm	Palkka

JYVÄSKYLÄN AMMATTIKORKEAKOULU

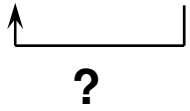
Funktionaalinen riippuvuus

Sarake B on funktionaalisesti riippuvainen A:sta, jos A:ta kohti on *korkeintaan* yksi B:n arvo kunakin ajanhetkenä.

HENKILÖ (Htun, HenkNimi, Osoite)

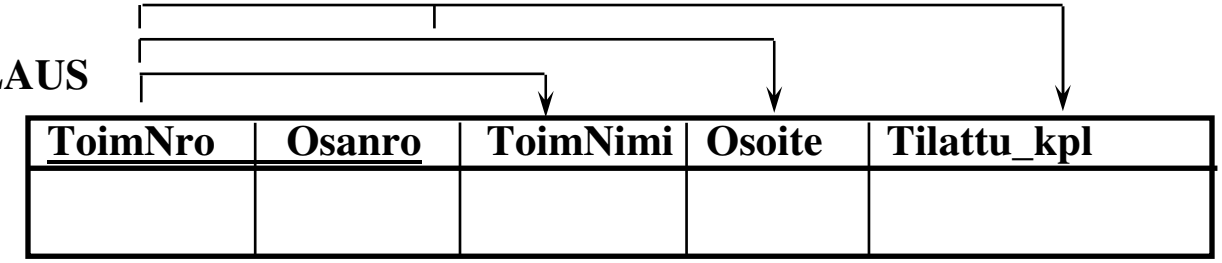


HENKILÖ (Htun, HenkNimi, Osoite)



JYVÄSKYLÄN AMMATTIKORKEAKOULU

2. Normaalimuoto



Mitä mieltä olet sarakkeiden nimistä? Ymmärrätkö mitä ne tarkoittavat?

2. Normaalimuoto:

Jos avain koostuu useasta sarakkeesta, on *jokaisen* muun sarakkeen oltava riippuvainen *koko* avaimesta.

TOIMITTAJA

ToimNro	Nimi	Osoite

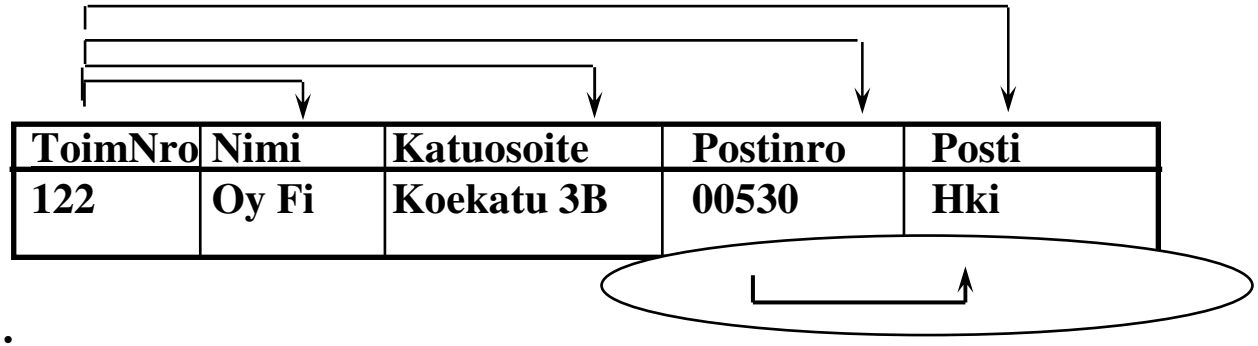
TILAUS

ToimNro	Osanro	Tilattu_kpl

- Mitä jos näin ei tehdä?
  - Ei voida tallettaa toimittajatietoja etukäteen
  - Vanhoja tilauksia poistettaessa tietyn toimittajan tiedot voivat häipyä kokonaan
  - Toimittajan osoitteen muutos tulee moneen paikkaan

3. Normaalimuoto

TOIMITTAJA



ToimNro	Nimi	Katuosoite	Postinro	Posti
122	Oy Fi	Koekatu 3B	00530	Hki

3. Normaalimuoto:

Kaikkien sarakkeiden tulee olla riippuvia *vain* perusavaimesta.

TOIMITTAJA

ToimNro	Nimi	Katuosoite	Postinro
122	Oy Fi	Koekatu 3B	00530

POSTI

Postinro	Postitoimip
00530	Hki

- Mitä jos näin ei tehdä?
  - Postitoimipaikkoja ei voida tallettaa ennen toimittajan talletusta
  - Postitoimipaikan nimen muutos tulee moneen paikkaan
- Lisähyötynä käyttäjätyytyväisyys ja virheiden ehkäisy:
  - Käyttöliittymä voidaan rakentaa siten, että käyttäjä valitsee listasta postinumeron => postitoimipaikka saadaan automaattisesti

3NM ja Käsitteet

- Suunnitteluputkessa normalisoidaan **käsitteitä**
  - Normalisointi voi tarvittaessa myöhemmin kohdistua tauluihin, tietueisiin jne.
  - Käsitemallin normalisointi tarkistetaan käsite käsitteeltä
  - Melko työläs operaatio, mutta vaivannäkö yleensä kannattaa
- Yhteenvetona (NM = normaalimuoto): kaikkien ei-avaintietojen tulee olla funktionaalisesti riippuvia
  - perusavaimesta (1. NM),
  - koko perusavaimesta (2. NM) ja
  - vain perusavaimesta (3. NM)

Tehtävä: normalisointi

- Tarkista tekemäsi käsitemalli (Demox Oy tai Formula-esimerkki)
  - missä normaalimuodossa malli on?
  - olisiko järkevää normalisoida mallia pitemmälle?
  - mitä etuja tai haittoja pitkälle normalisoidusta ratkaisusta on?
- Vinkki: usein tehdään nk. ideaalimalli, jossa normalisointi on ulotettu ”kaikkiin” tilanteisiin ja sen jälkeen tehdään perustellut rajaukset siitä, mikä tulee olemaan toteutettava ratkaisu (ts. palataan lähelle toista normaalimuotoa)
- Lisätehtävä: Tee kirjan (Hovi et. al. 2003) harjoitukset 9 ja 10 (sijaitsevat myös R5:ssä). Tutki asiaa myös webistä.

Denormalisointi

- Normalisoidun rakenteen yhdistäminen, esim. 3. normaalimuodosta 2. normaalimuotoon
- Ks. Kirjan luku 5 (s. 95)

Lisätietoa tietokannan normalisoinnista

- http://www.datamodel.org/NormalizationRules.html
- http://en.wikipedia.org/wiki/Database_normalization
- http://databases.about.com/od/specificproducts/a/normalization.htm
- http://www.databasejournal.com/sqletc/article.php/1428511
- http://dev.mysql.com/tech-resources/articles/intro-to-normalization.html
- http://www.utexas.edu/its/windows/database/datamodeling/rm/rm7.html