

# CS-A1150 Tietokannat

19.3.2019

## Oppimistavoitteet: tämän luennon jälkeen

- ▶ Osaat muuttaa huonon tietokantakaavion paremmaksi: Tiedät mitä tarkoitetaan sillä, että relaatio on Boyce-Codd-normaalimuodossa. Jos relaatio ei ole Boyce-Codd-normaalimuodossa, osaat jakaa sen uusiksi relaatioiksi, jotka ovat BCNF:ssä.
- ▶ Osaat selittää, mitä tarkoittavat moniarvoiset riippuvuudet ja 4. normaalimuoto.

## Kertausta: kaksi tapaa esittää tuotteita ja valmistajia

### Relaatio Products1

<i>number</i>	<i>prodName</i>	<i>description</i>	<i>price</i>	<i>manufID</i>	<i>manufName</i>	<i>phone</i>
T-33441	Galaxy A5	cellphone	250.0	S123	Samsung	020-7300
S-65221	Brasserie 24	pan	33.50	F542	Fiskars	020-43910
T-33442	NX 300 Smart	camera	399.0	S123	Samsung	020-7300
T-33455	Cyber-shot	camera	463.0	L711	Sony	020-6500
R-43118	Samsung LT 24	TV	169.0	S123	Samsung	020-7300
R-27113	Sony 32 KDL	TV	347.0	L711	Sony	020-6500

## Esimerkki jatkuu

### Relation Products

<i>number</i>	<i>prodName</i>	<i>description</i>	<i>price</i>	<i>manufID</i>
T-33441	Galaxy A5	cellphone	250.0	S123
S-65221	Brasserie 24	pan	33.50	F542
T-33442	NX 300 Smart	camera	399.0	S123
T-33455	Cyber-shot	camera	463.0	L711
R-43118	Samsung LT 24	TV	169.0	S123
R-27113	Sony 32 KDL	TV	347.0	L711

### Relation Manufactures

<i>ID</i>	<i>manufName</i>	<i>phone</i>
S123	Samsung	020-7300
L711	Sony	020-6500
F542	Fiskars	020-43910

## Kertausta: funktionaalinen riippuvuus (esimerkki)

- ▶ Tarkastellaan relaatiota  $R(A, B, C, D, E)$ . Oletetaan, että relaatiossa on voimassa funktionaalinen riippuvuus  $B C \rightarrow D$ .
- ▶ Jos relaatiossa kahdella monikolla on samat arvot sekä  $B$ - että  $C$ -attribuuteilla, niin niillä on sama arvo myös  $D$ -attribuutilla.
- ▶ Oletetaan, että relaation instanssissa on monikko

("15.9.2016", "112234", "CSE-A1200", 3, 5.0)

- ▶ Tällöin relaation instanssissa ei voi olla monikkoa

("10.11.2016", "112234", "CSE-A1200", 2, 5.0)

- ▶ Sen sijaan esimerkiksi monikot

("15.9.2016", "112234", "CSE-A1111", 2, 5.0)

("15.9.2016", "227533", "CSE-A1200", 5, 5.0)

("15.9.2016", "112234", "CSE-A1200", 3, 2.0)

ovat täysin mahdollisia, ellei ole muita funktionaalisia riippuvuuksia.

## Kertausta: Ositus/yhdistämissääntö

- ▶ Ositussääntö (engl. splitting rule): jos riippuvuuden oikealla puolella esiintyy useampi attribuutti, voidaan riippuvuus jakaa useammaksi riippuvuudeksi, esimerkiksi:

$$A B \rightarrow C D E$$

on ekvivalentti seuraavan funktionaalisten riippuvuuksien joukon kanssa:

$$A B \rightarrow C$$

$$A B \rightarrow D$$

$$A B \rightarrow E$$

- ▶ **Riippuvuuden vasenta puolta ei saa jakaa!**
- ▶ Yhdistämissääntö (engl. combining rule): jos usealla funktionaalisella riippuvuudella on täsmälleen sama vasen puoli, voidaan ne yhdistää yhdeksi riippuvuudeksi.

# Riippuvuuksien luokittelua

- ▶ Tarkastellaan riippuvuutta

$$A_1 A_2 \dots A_n \rightarrow B_1 B_2 \dots B_m$$

- ▶ Riippuvuus on *triviaali* (trivial), jos kaikki  $B$ :t ovat  $A$ :iden osajoukko. Esimerkiksi riippuvuus  $A B C \rightarrow A B$  on triviaali.
- ▶ Riippuvuus on epätriviaali (nontrivial), jos vähintään yksi  $B$  ei sisälly  $A$ :iden joukkoon. Esimerkiksi riippuvuus  $A B \rightarrow A B D$  on epätriviaali.
- ▶ Riippuvuus on täysin epätriviaali (completely nontrivial), jos mikään  $B$  ei sisälly  $A$ :iden joukkoon. Esimerkiksi riippuvuus  $A B \rightarrow D$  on täysin epätriviaali.
- ▶ Riippuvuuden oikealta puolelta voidaan aina poistaa sellaiset attribuutit, jotka esiintyvät myös sen vasemmalla puolella.

# Transitiivisuussääntö

- ▶ Tarkastellaan relaatiota  $R(A, B, C, D)$ . Jos on voimassa funktionaaliset riippuvuudet  $A \rightarrow B$  ja  $B \rightarrow C$ , niin silloin on voimassa myös riippuvuus  $A \rightarrow C$ .
- ▶ Todistus:
  - ▶ Olkoon  $R$ :n instanssissa kaksi monikkoa, joilla on samat arvot attribuutille  $A$ :  $(a, b_1, c_1, d_1)$  ja  $(a, b_2, c_2, d_2)$ .
  - ▶ Koska  $R$ :lle pätee  $A \rightarrow B$ , täytyy tällöin olla  $b_1 = b_2$ .
  - ▶ Koska nyt molemmilla monikoilla on sama arvo attribuutille  $B$  ja  $B \rightarrow C$ , täytyy myös monikoiden  $C$  arvot olla samat:  $c_1 = c_2$ .
  - ▶ On siis todistettu, että jos kahdella monikolla on sama arvo attribuutille  $A$ , niillä on myös sama arvo attribuutille  $C$  eli funktionaalinen riippuvuus  $A \rightarrow C$  pätee.



## Attribuuttien sulkeuma

- ▶ Olkoon annettu attribuuttijoukko  $\{A_1, A_2, \dots, A_n\}$  ja joukko funktionaalisia riippuvuuksia  $S$ .
- ▶ Mitkä kaikki attribuutit riippuvat tällöin funktionaalisesti joukosta  $\{A_1, A_2, \dots, A_n\}$ ?
- ▶ Näiden attribuuttien joukkoa kutsutaan joukon  $\{A_1, A_2, \dots, A_n\}$  sulkeumaksi riippuvuusjoukon  $S$  suhteen.
- ▶ Sulkeumaa merkitään

$$\{A_1, A_2, \dots, A_n\}^+$$

# Attribuuttien sulkeuman laskeminen, algoritmi

- ▶ Attribuuttien joukon  $\{A_1, A_2, \dots, A_n\}$  sulkeuman funktionaalisten riippuvuuksien joukon  $S$  suhteen voi laskea seuraavasti:
  1. Jaa tarvittaessa funktionaaliset riippuvuudet käyttämällä ositussääntöä niin, että jokaisen riippuvuuden oikealla puolella esiintyy vain yksi attribuutti.
  2. Olkoon  $X$  joukko, joka lopulta sisältää sulkeuman. Alusta  $X$ :ksi aluksi joukko  $\{A_1, A_2, \dots, A_n\}$ .
  3. Etsi jokin funktionaalinen riippuvuus

$$B_1 B_2 \dots B_m \rightarrow C$$

siten, että kaikki  $B_1, B_2, \dots, B_m$  sisältyvät joukkoon  $X$ , mutta  $C$  ei sisälly. Lisää  $C$  joukkoon  $X$ .

4. Toista kohtaa kolme, kunnes  $X$ :ää ei enää voi kasvattaa.
5. Joukko  $X$  on nyt  $\{A_1, A_2, \dots, A_n\}^+$ .

## Attribuuttien sulkeuman laskeminen, esimerkki

- ▶ Tarkastellaan relaatiota  $R(A, B, C, D, E, F)$ , jolla on riippuvuudet  $B C \rightarrow A D$ ,  $A B \rightarrow C$ ,  $D \rightarrow E$  ja  $C F \rightarrow B$ . Mikä on  $\{A, B\}^+$ ?
- ▶ *Ratkaisu:* Jaetaan ensin ensimmäinen riippuvuus kahdeksi riippuvuudeksi. Saadaan siis riippuvuudet:

$$B C \rightarrow A$$

$$B C \rightarrow D$$

$$A B \rightarrow C$$

$$D \rightarrow E$$

$$C F \rightarrow B$$

(jatkuu seuraavalla kalvolla)

## Attribuuttien sulkeuman laskeminen, ratkaisu jatkuu

- ▶ Aloitetaan asettamalla  $X = \{A, B\}$ .
- ▶ Kolmannen riippuvuuden perusteella voidaan lisätä  $X$ :ään  $C$ , siis  $X = \{A, B, C\}$ .
- ▶ Toisen riippuvuuden perusteella lisätään  $D$ , jolloin  $X = \{A, B, C, D\}$ .
- ▶ Nyt neljännen riippuvuuden perusteella voidaan lisätä  $E$ , joten  $X = \{A, B, C, D, E\}$ .
- ▶ Tämän jälkeen  $X$ :n kasvattaminen ei enää onnistu minkään riippuvuuden avulla.  $\{A, B\}^+$  on siis  $\{A, B, C, D, E\}$

## Miksi laskea attribuuttien sulkeuma?

- ▶ Attribuuttien sulkeuman perusteella voidaan päätellä, seuraako jokin riippuvuus  $A_1 A_2 \dots A_n \rightarrow B$  annetusta funktionaalisten riippuvuuskien joukosta  $S$ :
  - ▶ Lasketaan ensin  $\{A_1, A_2, \dots, A_n\}^+$ .
  - ▶ Jos  $B$  kuuluu sulkeumaan, niin  $A_1 A_2 \dots A_n \rightarrow B$  seuraa  $S$ :stä, jos  $B$  ei kuulu sulkeumaan, niin tarkasteltu riippuvuus ei seuraa joukosta  $S$ .
- ▶ Näin voidaan päätellä, seuraako jokin riippuvuuksien joukko toisesta riippuvuuksien joukosta, tai ovatko annetut kaksi riippuvuuksien joukkoa ekvivalentteja.
- ▶ Sulkeuman avulla voidaan myös päätellä, onko annettu attribuuttijoukko relaation yliavain. Lasketaan attribuuttijoukon sulkeuma. Jos relaation kaikki attribuutit kuuluvat sulkeumaan, attribuuttijoukko on relaation yliavain.

## Takaisin alun esimerkkiin erilaisista tietokantakaavioista

- ▶ Ensimmäisessä tavassa tiedot on tallennettu relaatioon  
`Products1(number, prodName, description, price, manufID,  
          manufName, phone)`
- ▶ Toisessa tavassa tiedot on jaettu kahteen relaatioon  
`Products(number, prodName, description, price, manufID)`  
`Manufacturers(ID, manufName, phone)`
- ▶ Etsitään yleinen tapa päästä ensimmäisestä tavasta toiseen.

## Relaation osittaminen

- ▶ Tarkoituksena on välttää anomalia osittamalla alkuperäinen relaatio  $R(A_1, A_2, \dots, A_n)$  kahdeksi uudeksi relaatioksi  $R_1(B_1, B_2, \dots, B_m)$  ja  $R_2(C_1, C_2, \dots, C_k)$  siten, että uudet relaatiot sisältävät saman informaation kuin vanha:
  1.  $\{A_1, A_2, \dots, A_n\} = \{B_1, B_2, \dots, B_m\} \cup \{C_1, C_2, \dots, C_k\}$
  2.  $R = R_1 \bowtie R_2$
- ▶ (Aivan kaikkea informaatiota relaation riippuvuuksista ei pystytä aina säilyttämään, tästä myöhemmin.)

## Esimerkki epäonnistuneesta osituksesta

- ▶ Relaatio

```
Products1(number, prodName, description, price, manufID,  
           manufName, phone)
```

on jaettu relaatioihin

```
P1(number, prodName, description, price)
```

```
P2(description, manufID, manufName, phone)
```

- ▶ Mikä on ongelma?



## Relaatio P1

<i>number</i>	<i>prodName</i>	<i>description</i>	<i>price</i>
T-33441	Galaxy A5	cellphone	250.0
S-65221	Brasserie 24	pan	33.50
T-33442	NX 300 Smart	camera	399.0
T-33455	Cyber-shot	camera	463.0
R-43118	Samsung LT 24	TV	169.0
R-27113	Sony 32 KDL	TV	347.0

## Relaatio P2

<i>description</i>	<i>manufID</i>	<i>manufName</i>	<i>phone</i>
cellphone	S123	Samsung	020-7300
pan	F542	Fiskars	020-43910
camera	S123	Samsung	020-7300
camera	L711	Sony	020-6500
TV	S123	Samsung	020-7300
TV	L711	Sony	020-6500

## Esimerkki epäonnistuneesta osituksesta, jatkoa

- ▶ Uudet relaatiot sisältävät kaikki samat attribuutit ja niiden arvot kuin alkuperäinen relaatio.
- ▶ Uusien relaatioiden perusteella ei kuitenkaan pystytä enää päättelemään, mikä valmistaja on valmistanut minkin tuotteen.
- ▶ Siis

$$P1 \bowtie P2 \neq \text{Products1}$$

toisin kuin vaadittiin.

## Boyce-Codd-normaali muoto

- ▶ Jos relaatio on Boyce-Codd-normaali muodossa (Boyce-Codd normal form, BCNF), siinä ei ole lainkaan funktionaalisista riippuvuuksista johtuvaa toisteista tietoa.
- ▶ Relaatio  $R$  on BCNF:ssä (Boyce-Codd-normaali muodossa) jos ja vain jos kaikille  $R$ :ssä voimassa oleville epätriviaaleille funktionaalisille riippuvuuksille  $A_1 A_2 \dots A_n \rightarrow B_1 B_2 \dots B_m$  pätee se, että  $\{A_1, A_2, \dots, A_n\}$  on relaation  $R$  yliavain.
- ▶ Esimerkki: aikaisemmin kuvattu relaatio

Products1(number, prodName, description, price, manufID,  
manufName, phone)

ei ole BCNF:ssä, koska relaatiolle pätee riippuvuus

$\text{manufID} \rightarrow \text{manufName phone}$

mutta riippuvuuden vasen puoli ei ole relaation yliavain, vaan relaation avain on  $\{\text{number}\}$

## Relaation osittaminen Boyce-Codd-normaaliin, rautalankaohjeet

- ▶ Laske annettujen epätriviaalien riippuvuuksien vasenten puolen sulkeumat.
- ▶ Jos yksikin sulkeumista ei sisällä relaation kaikkia attribuutteja (eli vasen puoli ei ole relaation yliavain), relaatio ei ole BCNF:ssä, ja se pitää osittaa.
- ▶ Valitse yksi niistä riippuvuuksista, jossa vasen puoli ei ole relaation yliavain ja ota tämän riippuvuuden vasemman puolen sulkeuma.
- ▶ Jaa ositettava relaatio kahdeksi uudeksi relaatioksi:
  - ▶ Ensimmäiseen tulee em. sulkeuma.
  - ▶ Toiseen tulee edellä valitun riippuvuuden vasen puoli ja lisäksi ne ositettavan relaation attribuutit, jotka eivät kuulu vasemman puolen sulkeumaan.
- ▶ Tutki, mitkä riippuvuudet ovat voimassa uusissa relaatioissa.
- ▶ Laske uusien relaatioiden riippuvuuksien vasempien puolien sulkeumat. Jos jokin sulkeuma ei sisällä uuden relaation kaikkia attribuutteja, uusi relaatio ei ole BCNF:ssä ja se pitää osittaa samalla tavalla.

## Relaation osittaminen BCNF:ään, esimerkki 1

- ▶ Tarkastellaan aikaisempaa esimerkkiä

Products1(number, prodName, description, price, manufID,  
manufName, phone)

- ▶ Relaatiolla on riippuvuudet

number → prodName description price manufID manufName  
phone

manufID → manufName phone

- ▶ Näistä ensimmäinen ei ole ongelma, koska siinä vasen puoli on relaation (yli)avain, mutta toinen riippuvuus rikkoo BCNF:n ehtoja.
- ▶ Algoritmin mukaan lasketaan  $\{\text{manufID}\}^+$ . Saadaan

$$\{\text{manufID}\}^+ = \{\text{manufID}, \text{manufName}, \text{phone}\}$$

## Esimerkki jatkuu

- ▶ Products1 jaetaan kahteen relaatioon, joista ensimmäisessä on sulkeumaan kuuluvat attribuutit ja toisessa riippuvuuden vasen puoli ja loput attribuutit:

Products2(manufID, manufName, phone)

Products3(manufID, number, prodName, description, price)

- ▶ Relaatio Products2 on nyt BCNF:ssä, sillä sillä on riippuvuus

manufID → manufName phone

mutta {manufID} on relaation Products2 avain.

- ▶ Relaatiolla Products3 on riippuvuus

number → prodName description price manufID

mutta {number} on relaation avain, joten myös se on BCNF:ssä.

- ▶ Näin alkuperäinen relaatio Products1 on ositettu kahdeksi relaatioksi Products2 ja Products3, jotka molemmat ovat BCNF:ssä.

## Relaation osittaminen Boyce-Codd-normaaliinmuotoon (täsmällinen algoritmi)

- ▶ Olkoon relaatio  $R_0$ , jolle pätee funktionaalisten riippuvuuksien joukko  $S_0$ . Jos  $R_0$  ei ole BCNF:ssä, se voidaan osittaa BCNF:ään seuraavalla algoritmilla:
  1. Aloita asettamalla  $R = R_0$  ja  $S = S_0$ . Tarkista, onko  $R$  BCNF:ssä. Jos on, mitään ei tarvitse tehdä. Palauta vastauksena  $\{R\}$ .
  2. Olkoon eräs BCNF:n estävä riippuvuus  $X \rightarrow Y$ . Laske  $X^+$ . Muodosta kaksi uutta relaatiota  $R_1$  ja  $R_2$  siten että,  $R_1$ :n relaatiokaavioon tulee kaikki  $X^+$ :ssä esiintyvät attribuutit ja  $R_2$ :n relaatiokaavioon tulee  $X$  ja lisäksi kaikki ne  $R_0$ :n attribuutit, jotka eivät esiinny  $X^+$ :ssä. (Huomaa, että  $X$  voi koostua usemmasta kuin yhdestä attribuutista.)
  3. Laske  $R_1$ :n ja  $R_2$ :n funktionaaliset riippuvuudet (selitetty myöhemmin); nimeä nämä riippuvuudet  $S_1$ :ksi ja  $S_2$ :ksi.
  4. Osita rekursiivisesti  $R_1$  ja  $R_2$  käyttämällä tätä samaa algoritmia. Palauta vastausten unioni koko ongelman ratkaisuna.

## Relaation osittaminen BCNF:ään, esimerkki 2

- ▶ Tarkastellaan relaatiota  $R(A, B, C, D)$ , jossa on voimassa riippuvuudet  $A \rightarrow B$  ja  $B \rightarrow C, D$ .
- ▶ Lasketaan riippuvuuksien vasenten puolten sulkeumat:

$$\{A\}^+ = \{A, B, C, D\}$$

$$\{B\}^+ = \{B, C, D\}$$

Toinen sulkeuma ei sisällä kaikkia  $R$ :n attribuutteja, joten  $B$  ei ole relaation yliavain eikä relaatio ole BCNF:ssä.

- ▶ Jaetaan relaatio kahdeksi uudeksi relaatioksi käyttämällä hyväksi jälkimmäistä sulkeumaa. Saadaan  $R_1(B, C, D)$  ( $B$ :n sulkeumaan kuuluvat attribuutit) ja  $R_2(A, B)$  ( $B$  ja loput  $R$ :n attribuutit).
- ▶ Tutkitaan, mitkä riippuvuudet ovat voimassa uusissa relaatioissa.



## Relaation osittaminen BCNF:ään, esimerkki 2 jatkuu

- ▶  $R_1$ :ssä on riippuvuus  $B \rightarrow C D$ . Lasketaan vasemman puolen sulkeuma  $R_1$ :n riippuvuuksien avulla:

$$\{B\}^+ = \{B, C, D\}$$

Koska sulkeuma sisältää kaikki  $R_1$ :n attribuutit,  $B$  on  $R_1$ :n yliavain ja  $R_1$  on BCNF:ssä.

- ▶  $R_2$ :ssä on riippuvuus  $A \rightarrow B$ . Lasketaan vasemman puolen sulkeuma  $R_2$ :n riippuvuuksien avulla:

$$\{A\}^+ = \{A, B\}$$

Koska sulkeuma sisältää kaikki  $R_2$ :n attribuutit,  $A$  on  $R_2$ :n yliavain ja  $R_2$  on BCNF:ssä.

- ▶ Molemmat uudet relaatiot ovat BCNF:ssä, joten ositus on valmis ja lopputulos on  $R_1(B, C, D)$  ja  $R_2(A, B)$

## Relaation osittaminen BCNF:ään, lisäesimerkki

- ▶ Tarkastellaan relaatiota  $R(A, B, C, D)$ , jossa on voimassa riippuvuudet  $A \rightarrow B C$  ja  $B C \rightarrow D$ .
- ▶ Lasketaan riippuvuuksien vasenten puolten sulkeumat:

$$\begin{aligned}\{A\}^+ &= \{A, B, C, D\} \\ \{B, C\}^+ &= \{B, C, D\}\end{aligned}$$

Toinen sulkeuma ei sisällä kaikkia  $R$ :n attribuutteja, joten  $\{B, C\}$  ei ole relaation yliavain eikä relaatio ole BCNF:ssä.

- ▶ Jaetaan relaatio kahdeksi uudeksi relaatioksi käyttämällä hyväksi jälkimmäistä sulkeumaa. Saadaan  $R_1(B, C, D)$  ( $\{B, C\}$ :n sulkeumaan kuuluvat attribuutit) ja  $R_2(A, B, C)$  ( $B, C$  ja ne  $R$ :n attribuutit, jotka eivät kuulu sulkeumaan).
- ▶ Tutkitaan, mitkä riippuvuudet ovat voimassa uusissa relaatioissa.

## Relaation osittaminen BCNF:ään, lisäesimerkki jatkuu

- ▶  $R_1$ :ssä on riippuvuus  $B \rightarrow C D$ . Lasketaan vasemman puolen sulkeuma  $R_1$ :n riippuvuuksien avulla:

$$\{B, C\}^+ = \{B, C, D\}$$

Koska sulkeuma sisältää kaikki  $R_1$ :n attribuutit,  $\{B, C\}$  on  $R_1$ :n yliavain ja  $R_1$  on BCNF:ssä.

- ▶  $R_2$ :ssä on riippuvuus  $A \rightarrow B C$ . Lasketaan vasemman puolen sulkeuma  $R_2$ :n riippuvuuksien avulla:

$$\{A\}^+ = \{A, B, C\}$$

Koska sulkeuma sisältää kaikki  $R_2$ :n attribuutit,  $A$  on  $R_2$ :n yliavain ja  $R_2$  on BCNF:ssä.

- ▶ Molemmat uudet relaatiot ovat BCNF:ssä, joten ositus on valmis ja lopputulos on  $R_1(B, C, D)$  ja  $R_2(A, B, C)$

## Funktionaalisten riippuvuuksien projisointi

- ▶ *Ongelma:* Kun relaatio ositetaan uusiksi relaatioiksi, niin mistä tiedetään, mitkä riippuvuudet ovat voimassa uusissa relaatioissa?
- ▶ Jos kaikki alkuperäisen relaation riippuvuuden attribuutit esiintyvät uudessa relaatioissa, on alkuperäinen riippuvuus voimassa myös uudessa relaatioissa.
- ▶ Lisäksi uudessa relaatioissa voi kuitenkin esiintyä myös sellaisia riippuvuuksia, joita ei mainittu alkuperäisen relaation riippuvuuksissa, mutta jotka syntyvät niiden "yhteisvaikutuksesta".
- ▶ Esimerkki: olkoon relaatioissa  $R(A, B, C, D)$  voimassa riippuvuudet  $A \rightarrow B$ ,  $B \rightarrow C$  ja  $C \rightarrow B D$ . Jos relaatio jaetaan relaatioiksi  $R_1(B, C, D)$  ja  $R_2(A, C)$ , niin  $R_2$ :ssa on voimassa riippuvuus  $A \rightarrow C$ , vaikka sitä ei mainittu alkuperäisten riippuvuuksien joukossa.

## Funktionaalisten riippuvuuksien projisointi, jatkoa

- ▶ Olkoon  $R$  relaatio, joka on ositettu relaatioksi  $S$  ja muiksi relaatioiksi, ja olkoon  $F$  relaatiossa  $R$  voimassa olevien riippuvuuksien joukko.
- ▶ Tarkastellaan jokaista attribuuttijoukkoa  $X$ , joka on  $S$ :n osajoukko. Jos seuraavat ehdot ovat voimassa
  1.  $B$  on  $S$ :n attribuutti.
  2.  $B$  sisältyy joukkoon  $X^+$  (riippuvuuksien  $F$  mukaan).
  3.  $B$  ei sisälly joukkoon  $X$niin riippuvuus  $X \rightarrow B$  on voimassa  $S$ :ssä.

## Esimerkki riippuvuuksien projisoinnista

- ▶ Tarkastellaan relaatiokaaviota  $R(A, B, C, D)$ . Olkoon siinä voimassa riippuvuudet  $A \rightarrow B$ ,  $B \rightarrow C$  ja  $C \rightarrow B D$ .
- ▶ Oletetaan, että  $R$ :n osituksen tuloksena on syntynyt relaatiokaavio  $R_2(A, C)$ . Mitkä riippuvuudet ovat voimassa  $R_2$ :ssä?
- ▶ Lasketaan jokaisen joukon  $\{A, C\}$  aidon osajoukon sulkeuma.
  1.  $\{A\}^+ = \{A, B, C, D\}$ . Koska  $B$  ja  $D$  eivät sisälly  $R_2$ :n attribuutteihin, riippuvuus  $A \rightarrow B D$  ei ole voimassa  $R_2$ :ssä. Sen sijaan  $C$  kuuluu  $R_2$ :n attribuutteihin, joten riippuvuus  $A \rightarrow C$  on voimassa  $R_2$ :ssä.
  2.  $\{C\}^+ = \{C, B, D\}$ , mutta se ei tuota uusia riippuvuuksia, koska  $B$  ja  $D$  eivät sisälly  $R_2$ :n attribuutteihin.
- ▶  $R_2$ :ssä on siis voimassa riippuvuus  $A \rightarrow C$ .

## Relaation osittaminen BCNF:ään, esimerkki 3

- ▶ Tarkastellaan relaatiota  $R(A, B, C, D, E)$ , jossa on voimassa riippuvuudet  $A \rightarrow B$ ,  $B \rightarrow D E$  ja  $C \rightarrow E$
- ▶ Lasketaan riippuvuuksien vasenten puolten sulkeumat:

$$\{A\}^+ = \{A, B, D, E\}$$

$$\{B\}^+ = \{B, D, E\}$$

$$\{C\}^+ = \{C, E\}$$

- ▶ Mikään sulkeumista ei sisällä kaikkia attribuutteja, joten mikään vasemmista puolista ei ole yliavain. Kaikki riippuvuudet rikkovat BCNF:ää.
- ▶ Valitaan rikkovista riippuvuuksista keksimmäinen (yhtä hyvin voitaisiin valita mikä tahansa) ja ositetaan relaatio sen avulla:  $R_1(B, D, E)$  ( $B$ :n sulkeuma) ja  $R_2(A, B, C)$  ( $B$  ja ne  $R$ :n attribuutit, jotka eivät kuulu  $B$ :n sulkeumaan).

## Relaation osittaminen BCNF:ään, esimerkki 3 jatkuu

- ▶  $R_1$ :ssä on voimassa riippuvuus  $B \rightarrow D E$ . Lasketaan sen vasemman puolen sulkeuma:

$$\{B\}^+ = \{B, D, E\}$$

Koska sulkeuma sisältää kaikki  $R_1$ :n attribuutit,  $B$  on relaation yliavain ja  $R_1$  on BCNF:ssä.

- ▶  $R_2$ :ssä on voimassa riippuvuus  $A \rightarrow B$ . Lasketaan sen vasemman puolen sulkeuma:

$$\{A\}^+ = \{A, B\}$$

Sulkeuma ei sisällä  $C$ :tä, joten  $A$  ei ole relaation yliavain eikä  $R_2$  ole vielä BCNF:ssä, vaan se pitää osittaa uudelleen.

- ▶ Ositetaan  $R_2(A, B, C)$ :  $R_3(A, B)$  ( $A$ :n sulkeuma) ja  $R_4(A, C)$  ( $A$  ja  $R_2$ :n loput attribuutit).



## Relaation osittaminen BCNF:ään, esimerkki 3

- ▶  $R_3$ :ssä on voimassa riippuvuus  $A \rightarrow B$ . Lasketaan sen vasemman puolen sulkeuma:

$$\{A\}^+ = \{A, B\}$$

Koska sulkeuma sisältää kaikki  $R_3$ :n attribuutit,  $A$  on relaation yliavain ja  $R_3$  on BCNF:ssä.

- ▶  $R_4$ :lla ei ole epätriviaaleja funktionaalisia riippuvuuksia, joten  $R_4$  on BCNF:ssä.
- ▶ Lopullinen ositus on siis  $R_1(B, D, E)$ ,  $R_3(A, B)$  ja  $R_4(A, C)$ .

## Huomaa osituksesta

- ▶ Usein osituksessa voidaan valita, missä järjestyksessä riippuvuuksia käytetään. Eri järjestykset voivat johtaa erilaisiin lopputuloksiin. Esimerkissä 3 ensimmäinen ositus oltaisiin yhtä hyvin voitu tehdä ensimmäisen tai kolmannen riippuvuuden perusteella.
- ▶ Jos ensimmäinen ositus olisi tehty ensimmäisen riippuvuuden perusteella, olisivat relaatiot olleet ensimmäisen osituksen jälkeen  $R_1(A, B, D, E)$  ( $A$ :n sulkeuma) ja  $R_2(A, C)$  ( $A$  ja ne  $R$ :n attribuutit, jotka eivät kuulu  $A$ :n sulkeumaan). Näistä  $R_1$  ei vielä ole BCNF:ssä, vaan se pitää osittaa uudelleen.
- ▶ Jos ensimmäinen ositus olisi tehty kolmannen riippuvuuden perusteella, olisivat relaatiot olleet ensimmäisen osituksen jälkeen  $R_1(C, E)$  ( $C$ :n sulkeuma) ja  $R_2(A, B, C, D)$  ( $C$  ja ne  $R$ :n attribuutit, jotka eivät kuulu  $C$ :n sulkeumaan). Näistä  $R_2$  ei vielä ole BCNF:ssä, vaan se pitää osittaa uudelleen.
- ▶ Tentissä ja harjoitustehtävissä riittää, että esittää vaihtoehtoisista ratkaisuista vain yhden.

## BCNF:n puutteita

- ▶ Jos relaatio on ositettu uusiksi relaatioiksi, jotka ovat BCNF:ssä, ei tieto kaikista alkuperäisessä relaatiossa olleista riippuvuuksista välttämättä säily siirryttäessä uusien relaatioiden riippuvuuksiin.
- ▶ Jos relaatio ositetaan BCNF:n sijasta *kolmanteen normaalimuotoon*, säilytetään tieto riippuvuuksista, mutta sen sijaan tietokantaan voi jäädä toisteista tietoa.
- ▶ Vaikka relaatio olisi BCNF:ssä, siinä voi olla toisteista tietoa, joka ei aiheudu funktionaalisista riippuvuuksista, vaan moniarvoisista riippuvuuksista. Ongelma voidaan ratkaista osittamalla relaatio *neljänteen normaalimuotoon*.

## Moniarvoiset riippuvuudet

- ▶ Esimerkki (oppikirjasta): tarkastellaan relaatiota *StarsIn1*, joka kuvaa sitä, missä elokuvassa kukin näyttelijä esiintyy. Relaatioon otetaan mukaan myös tieto näyttelijän osoitteesta. Relaatiokaavio on siis `StarsIn1(name, street, city, title, year)`,
- ▶ Oletetaan, että samalla näyttelijällä voi olla useita eri osoitteita samassa kaupungissa tai eri kaupungeissa.
- ▶ Relaatiolla ei ole lainkaan epätriviaaleja funktionaalisia riippuvuuksia, sillä mikään attribuuttien yhdistelmä ei määrää suoraan jonkin muun attribuutin arvoa. (Eri kaupungeissa sijaitsevilla asunnoillakin voi olla sama osoite, joten edes riippuvuus `name street` → `city` ei ole voimassa.)

## Moniarvoiset riippuvuudet, jatkuu

- ▶ Osa *StarsIn1*-relaation sisällöstä voisi näyttää seuraavalta:

Relaatio StarsIn1

<i>name</i>	<i>street</i>	<i>city</i>	<i>title</i>	<i>year</i>
Carrie Fischer	123 Maple St.	Hollywood	Star Wars	1977
Carrie Fischer	5 Locust Ln.	Malibu	Star Wars	1977
Carrie Fischer	123 Maple St.	Hollywood	Empire Strikes Back	1980
Carrie Fischer	5 Locust Ln.	Malibu	Empire Strikes Back	1980
Carrie Fischer	123 Maple St.	Hollywood	Return of the Jedi	1983
Carrie Fischer	5 Locust Ln.	Malibu	Return of the Jedi	1983

- ▶ Saman elokuvan tiedot on esitetty kahteen kertaan ja sama osoite kolmeen kertaan, vaikka relaatio on BCNF:ssä.
- ▶ Osoitteet ja elokuvat ovat toisistaan riippumattomia, mutta molemmat riippuvat näyttelijän nimestä, joten jokainen osoite pitää esittää jokaisen elokuvan yhteydessä. Tätä ilmiötä sanotaan *Moniarvoiseksi riippuvuudeksi* (multivalued dependency).

# Moniarvoisen riippuvuuden määritelmä

► Moniarvoinen riippuvuus

$$A_1 A_2 \dots A_n \twoheadrightarrow B_1 B_2 \dots B_m$$

on voimassa relaatiossa  $R$ , jos sen jokaiselle monikkoparille  $t$  ja  $u$ , joilla on  $A$ -attribuuttien osalta samat arvot, voidaan löytää  $R$ :stä monikko  $v$ , jolle pätee:

1.  $v$ :llä on samat arvot  $A$ -attribuuttien osalta kuin  $t$ :llä ja  $u$ :lla,
2.  $v$ :llä on samat arvot  $B$ -attribuuttien osalta kuin  $t$ :llä,
3.  $v$ :llä on samat arvot kuin  $u$ :llä niiden  $R$ :n attribuuttien osalta, jotka eivät sisälly  $A$ -attribuutteihin eikä  $B$ -attribuutteihin.

## Esimerkki moniarvoisesta riippuvuudesta

- ▶ Olkoon relaation  $R$  relaatiokaavio  $R(A, B, C)$  ja siinä voimassa riippuvuus  $A \twoheadrightarrow B$ . Jos relaation instanssissa on monikot  $(a, b_1, c_1)$  ja  $(a, b_2, c_2)$ , niin moniarvoisen riippuvuuden määritelmän mukaan siinä on myös monikko  $(a, b_1, c_2)$ .
- ▶ Kun moniarvoisen riippuvuuden määritelmässä vaihdetaan  $u$ :n ja  $t$ :n roolit, saadaan määritelmästä myös se, että  $R$ :n instanssissa pitää olla monikko  $(a, b_2, c_1)$ .
- ▶ Relaation instanssissa pitää siis olla sen kahden ensimmäisen monikon perusteella ainakin seuraavat monikot:

Relaatio  $R$

$A$	$B$	$C$
$a$	$b_1$	$c_1$
$a$	$b_2$	$c_2$
$a$	$b_1$	$c_2$
$a$	$b_2$	$c_1$

## Toinen esimerkki

- ▶ Tarkastellaan aikaisemmin esitettyä *StarsIn1*-relaatiota. Siinä on selvästi voimassa moniarvoinen riippuvuus

`name → street city`

- ▶ Näin esimerkiksi relaation instanssin ensimmäisen ja neljännen rivin perusteella

<i>name</i>	<i>street</i>	<i>city</i>	<i>title</i>	<i>year</i>
Carrie Fischer	123 Maple St.	Hollywood	Star Wars	1977
Carrie Fischer	5 Locust Ln.	Malibu	Empire Strikes Back	1980

relaatiossa täytyy olla myös rivit

<i>name</i>	<i>street</i>	<i>city</i>	<i>title</i>	<i>year</i>
Carrie Fischer	5 Locust Ln.	Malibu	Star Wars	1977
Carrie Fischer	123 Maple St.	Hollywood	Empire Strikes Back	1980



## Moniarvoiset riippuvuudet tällä kurssilla

- ▶ Tämän kurssin vaatimukseen kuuluu ymmärtää, mitä tarkoittaa moniarvoinen riippuvuus ja mitä tarkoittaa 4. normaalimuoto. Esimerkiksi ositusta 4. normaalimuotoon ei tarvitse osata.
- ▶ Seuraavilla kalvoilla on kuitenkin esitetty kurssivaatimuksia enemmän asioita niille, jotka haluavat (itse)opiskella lisää tai tarvitsevat näitä asioita työtehtävissä.

## Moniarvoisia riippuvuuksia koskevia määritelmiä

- ▶ Moniarvoinen riippuvuus

$$A_1 A_2 \dots A_n \rightarrow B_1 B_2 \dots B_m$$

on *triviaali*, jos sen oikealla puolella esiintyy vain sellaisia attribuutteja, jotka esiintyvät myös sen vasemmalla puolella, toisin sanoen

$$\{B_1, B_2, \dots, B_m\} \subseteq \{A_1, A_2, \dots, A_n\}$$

- ▶ Relaatiota  $R$  koskeva moniarvoinen riippuvuus

$$A_1 A_2 \dots A_n \rightarrow B_1 B_2 \dots B_m$$

on triviaali myös silloin, jos siinä esiintyvät kaikki  $R$ :n attribuutit eli  $R$ :n relaatiokaavio on  $R(A_1, A_2, \dots, A_n, B_1, B_2, \dots, B_m)$ .

- ▶ Trivaalit moniarvoiset riippuvuudet ovat voimassa jokaisessa relaatiossa.

# Moniarvoisia riippuvuuksia koskevia määritelmiä, jatkuu

- ▶ Relaatiota  $R$  koskeva moniarvoinen riippuvuus

$$A_1 A_2 \dots A_n \twoheadrightarrow B_1 B_2 \dots B_m$$

on *epätriviaali*, jos

1. vähintään yksi  $B$ -attribuuteista ei sisälly  $A$ -attribuutteihin ja
2.  $A$ -attribuuttien ja  $B$ -attribuuttien yhdiste ei sisällä kaikkia  $R$ :n attribuutteja.

## Moniarvoisia riippuvuuksia koskevia sääntöjä

- ▶ *Transitiivisuussääntö* (transitive rule): Jos on voimassa  $A_1 A_2 \dots A_n \rightarrow B_1 B_2 \dots B_m$  ja  $B_1 B_2 \dots B_m \rightarrow C_1 C_2 \dots C_k$ , niin silloin pätee myös

$$A_1 A_2 \dots A_n \rightarrow C_1 C_2 \dots C_k$$

Alimmasta riippuvuudesta voidaan poistaa oikealta puolelta kaikki ne  $C$ :t, jotka esiintyvät myös  $A$ :iden joukossa.

- ▶ Huomaa, että moniarvoisiin riippuvuuksiin **ei voi** soveltaa ositus- eikä yhdistämissääntöä.
- ▶ Jokainen funktionaalinen riippuvuus on myös moniarvoinen riippuvuus, eli jos

$$A_1 A_2 \dots A_n \rightarrow B_1 B_2 \dots B_m$$

on voimassa jollekin relaatiolle, niin myös

$$A_1 A_2 \dots A_n \rightarrow B_1 B_2 \dots B_m$$

on voimassa samalle relaatiolle.

## Moniarvoisia riippuvuuksia koskevia sääntöjä, jatkuu

- ▶ *Komplementointisääntö (complementation rule)*: Jos  $A_1 A_2 \dots A_n \twoheadrightarrow B_1 B_2 \dots B_m$  on voimassa relaatiolle  $R$ , niin tällöin relaatiolle  $R$  on voimassa myös

$$A_1 A_2 \dots A_n \twoheadrightarrow C_1 C_2 \dots C_k$$

missä  $C$ -attribuutit ovat kaikki ne  $R$ :n attribuutit, jotka eivät ole  $A$ -attribuutteja eikä  $B$ -attribuutteja.

- ▶ *Esimerkki*: Tarkastellaan relaatiota  
`StarsIn1(name, street, city, title, year)`,  
Aikaisemmin on todettu, että siinä on voimassa moniarvoinen riippuvuus

$$\text{name} \twoheadrightarrow \text{street city}$$

Komplementointisäännön perusteella siinä täytyy olla myös voimassa moniarvoinen riippuvuus

$$\text{name} \twoheadrightarrow \text{title year}$$

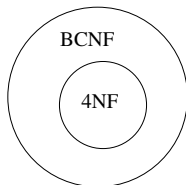
## Neljäs normaalimuoto

- ▶ Jos relaatio on *neljännessä normaalimuodossa* (fourth normal form, 4NF) siinä ei ole funktionaalisista riippuvuuksista eikä moniarvoisista riippuvuuksista johtuvaa toisteisuutta.
- ▶ Relaatio  $R$  on neljännessä normaalimuodossa, jos jokaiselle siinä voimassa olevalle epätriviaalille moniarvoiselle riippuvuudelle

$$A_1 A_2 \dots A_n \twoheadrightarrow B_1 B_2 \dots B_m$$

pätee se, että attribuuttijoukko  $\{A_1, A_2, \dots, A_n\}$  on relaation  $R$  yliavain.

- ▶ Jos relaatio on 4NF:ssä, se on aina myös BCNF:ssä. Päinvastainen ei välttämättä päde.



## Neljäs normaalimuoto, esimerkki

- ▶ Tarkastellaan aikaisempaa relaatiota

`StarsIn1(name, street, city, title, year),`

Aikaisemmin on todettu, että siinä on voimassa moniarvoinen riippuvuus

`name → street city`

- ▶ Attribuutti `name` ei kuitenkaan riitä relaation avaimeksi, koska se ei määrää funktionaalisesti kaikkia muita relaation attribuutteja.
- ▶ Tämän perusteella relaatio ei ole 4NF:ssä.

## Relaation osittaminen 4NF:ään

- ▶ Olkoon relaatio  $R_0$ , jolle pätee moniarvoisten riippuvuuksien joukko  $S_0$ . Jos  $R_0$  ei ole 4NF:ssä, se voidaan osittaa 4NF:ään samantapaisella algoritmilla kuin millä relaatio ositettiin BCNF:ään:
  1. Aloita asettamalla  $R = R_0$  ja  $S = S_0$ .
  2. Etsi jokin 4NF:n estävä epätriviaali riippuvuus  $A_1 A_2 \dots A_n \twoheadrightarrow B_1 B_2 \dots B_m$ , missä  $\{A_1, A_2, \dots, A_n\}$  ei ole relaation yliavain. (Huomaa, että myös funktionaaliset riippuvuudet ovat samalla moniarvoisia riippuvuuksia.)
  3. Muodosta kaksi uutta relaatiota  $R_1$  ja  $R_2$  siten, että  $R_1$ :n relaatiokaavioon tulee kaikki  $A$ - ja  $B$ -attribuutit ja  $R_2$ :n relaatiokaavioon kaikki  $A$ -attribuutit ja ne  $R$ :n attribuutit, jotka eivät kuulu  $A$ - tai  $B$ -attribuutteihin.
  4. Laske  $R_1$ :n ja  $R_2$ :n funktionaaliset ja moniarvoiset riippuvuudet (tästä lisää myöhemmin); nimeä nämä riippuvuudet  $S_1$ :ksi ja  $S_2$ :ksi.
  5. Osita rekursiivisesti  $R_1$  ja  $R_2$  käyttämällä tätä samaa algoritmia, jos  $R_1$  ja  $R_2$  eivät ole 4NF:ssä. Palauta vastausten unioni koko ongelman ratkaisuna.



# Osittettujen relaatioiden moniarvoisten riippuvuuksien laskeminen

- ▶ Mistä tiedetään, mitä moniarvoisia riippuvuuksia on osituksen tuloksena syntyneillä relaatioilla  $R_1$  ja  $R_2$ ?
- ▶ Tarkan ratkaisun saamiseen on suhteellisen työläs algoritmi, joka on esitetty oppikirjan (3. painos) kappaleessa 3.7.
- ▶ Useimmissa käytännön tilanteissa riittää laskea osittettujen relaatioiden funktionaaliset riippuvuudet aikaisemmin kuvatulla tavalla sekä ottaa lisäksi mukaan ne alkuperäisen relaation moniarvoiset riippuvuudet, joiden kaikki attribuutit esiintyvät ositetussa relaatioissa. Lisäksi on harkittava niiden moniarvoisten riippuvuuksien mukaanottamista, joiden attribuuteista osa esiintyy ositetussa relaatioissa.

## Esimerkki relaation osittamisesta 4NF:ään

- ▶ Tarkastellaan aikaisempaa relaatiota

StarsIn1(name, street, city, title, year)

Relaatiolla ei ole lainkaan epätriviaaleja funktionaalisia riippuvuuksia (mikään neljän tai vähemmän attribuutin joukko ei riitä avaimeksi), mutta sillä on voimassa moniarvoinen riippuvuus

$$\text{name} \twoheadrightarrow \text{street city}$$

missä name ei ole relaation yliavain.

- ▶ Ositetaan relaatio kahdeksi uudeksi relaatioksi. Ensimmäiseen tulee moniarvoisessa riippuvuudessa esiintyvät attributit:

StarsIn2(name, street, city)

ja toiseen tulee esitetyn moniarvoisen riippuvuuden vasemman puolen attribuutti sekä alkuperäisen relaation ne attributit, jotka eivät esiinny käsiteltävässä riippuvuudessa:

StarsIn3(name, title, year)

## Esimerkki relaation osittamisesta 4NF:ään, jatkoa

- ▶ Ositetuissa relaatioissa ei ole edelleenkään epätriviaaleja funktionaalisia riippuvuuksia.
- ▶ Relaatiossa *StarsIn2* on voimassa riippuvuus

`name → street city`

mutta se on triviaali, koska siinä esiintyy kaikki relaation attribuutit.

- ▶ Komplementointisäännön perusteella alkuperäisessä relaatiossa oli voimassa myös funktionaalinen riippuvuus

`name → title year`

Tämä riippuvuus siirtyy ositettuun relaatioon *StarsIn3*, koska kaikki riippuvuudessa esiintyvät attribuutit kuuluvat relaation *StarsIn3* relaatiokaavioon. Riippuvuus on kuitenkin triviaali, koska se sisältää kaikki *StarsIn3*:n attribuutit.

- ▶ Näin sekä *StarsIn2* että *StarsIn3* ovat 4NF:ssä, joten ositus on valmis.