

# **DATU-BASE ERLAZIONALEN NORMALIZAZIOA**

# Indizea

1. Diseinuaren egokitasuna. Zer da diseinu on bat?
2. Mendekotasun-funtzionalak
3. Forma normalak eta normalizazioa

# 1. Diseinuaren egokitasuna

- Diseinu kontzeptuala egin dugu. Egokia ote?
- Diseinuaren “egokitasuna” eta “ontasuna” neurtu behar dugu diseinuaren kalitatea ebaluatzeko
- HELBURUA: erlazio-eskemetako atributu taldekatzeen multzo bat beste bat baino hobea zergatik den neurtzea

# 1. Zer da diseinu on bat?

*Zergatik honakoa EZ DA diseinu ona?*

SALMENTAK(SukKod, EnpIzen, SukTfnoa, IFK, Unitateak, ProKod, Salneurria, Banatzailea, Sum)

*Zergatik da hau hobe?*

SALMENTAK(SukKod, EnpIzen, ProKod, Unitateak)

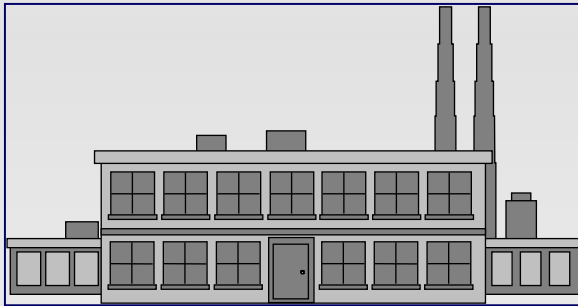
SUKURTSALA(SukKod, EnpIzen, SukTfnoa)

ENPRESA(EnpIzen, IFK)

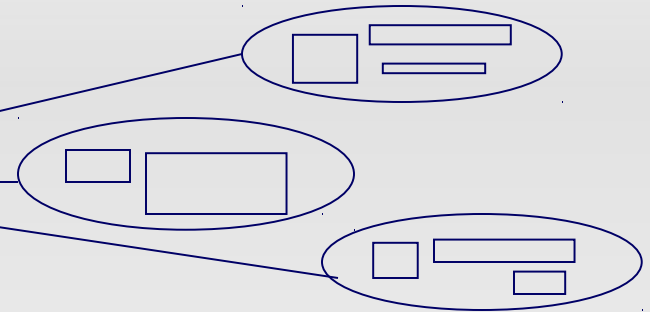
PRODUKTUA(ProKod, Salneurria, Banatzailea, Sum)

# Normalizazioa

Domeinu bat



Hainbat diseinu



- Normalizazioa

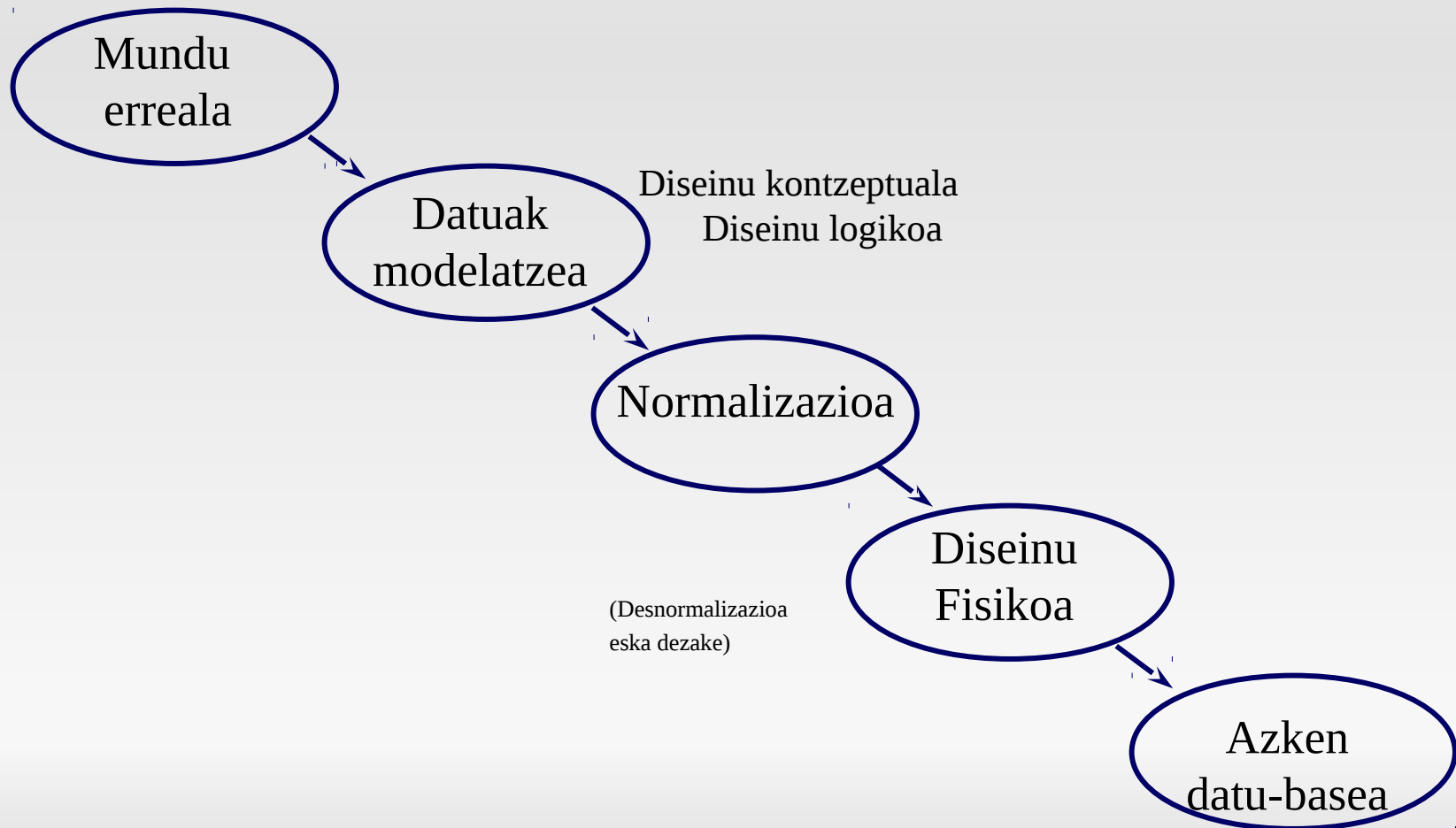
**Arazo** kopurua  
metodologia.

**minimizatzen**

duen

diseinu

# Testuingurua



# Normalizazioaren helburuak

- Diseinu logikoa hobetzea eta balidatzea
  - Datuen erredundantzia ekiditea
  - Ongi egituraturako erlazioak lortzea
- 
- Ideia intuitiboa: Taula bakoitzak kontzeptu bakarra lantzen duela ziurtatzea

# Arazoak diseinuan

- Akatsak
  - txertatzean
  - ezabatzean
  - eguneratzean
- NULL balio gehiegi
- Tupla aizunen (faltsuen) agerpena



# Aldatze-akatsak

<u>SukKod</u>	<u>EnpIzen</u>	<u>SukTfnoa</u>	<u>IFK</u>	<u>Uni</u>	<u>ProKod</u>	<u>Sal</u>	<u>Ban</u>
suk1	corte_inglés	193	ci	12	#prod1	123	agro
suk2	corte_inglés	294	ci	18	#prod1	123	agro
suk2	corte_inglés	294	ci	7	#prod2	827	telna
suk2	corte_inglés	294	ci	86	#prod3	123	agro
suk2	corte_inglés	294	ci	45	#prod6	123	agro
suk2	corte_inglés	294	ci	98	#prod2	827	telna
suk1	eroski	185	ab	103	#prod2	827	telna
suk1	eroski	185	ab	40	#prod5	308	xcon
suk2	eroski	200	ab	27	#prod7	126	agro

**EGUNERATZEA:** *#prod1*-en salneurria 200-era igo da

➡ 2 tupla aldatu behar dira. Erredundantzia! **Akatsa!**

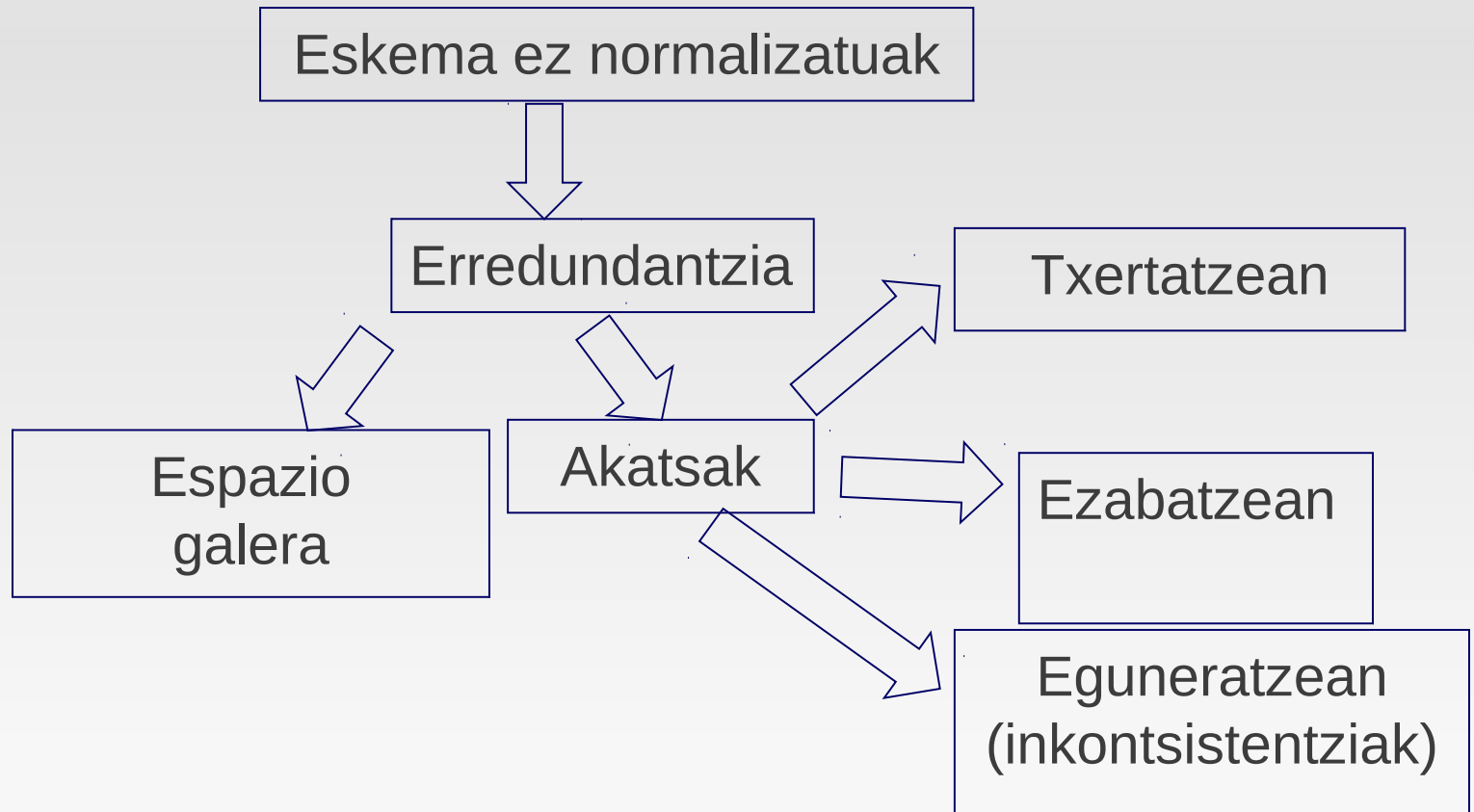
**TXERTATU:** *agro* banatzaileak *#prod9* produktu berria banatzen du

➡ ezin dugu txertatu, ez baitakigu gako osoa **Akatsa!**

**EZABATU:** *eroski*-ko *suk2* sukurtsala itxi egingo da

➡ *#prod7* produktuari buruz genuen informazio galdua! **Akatsa!**

# Motibazioa

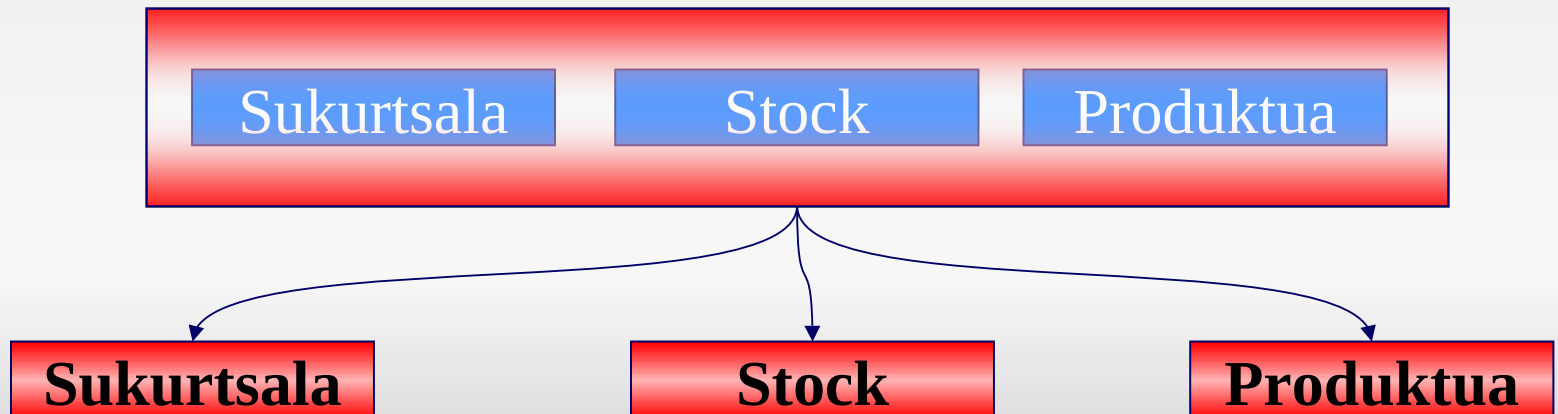


# Zergatik akatsak?

- *Arrazoia*
  - Erlazio batek entitate mota bat baino gehiago du
- *Soluzioa*
  - Erlazioak deskonposatzea, modu honetan entitateak “isolatuaz”, beti ere, informazioa eta mendekotasunak babestuz (mantenduaz)

# Jarraipena

<u>SukKod</u>	<u>Enplzen</u>	<u>SukTfnoa</u>	<u>IFK</u>	<u>Uni</u>	<u>ProKod</u>	<u>Sal</u>	<u>Ban</u>
suk1	corte_inglés	193	ci	12	#prod1	123	agro
suk2	corte_inglés	294	ci	18	#prod1	123	agro
suk2	corte_inglés	294	ci	7	#prod2	827	telna
suk2	corte_inglés	294	ci	86	#prod3	123	agro
suk2	corte_inglés	294	ci	45	#prod6	123	agro
suk2	corte_inglés	294	ci	98	#prod2	827	telna
suk1	eroski	185	ab	103	#prod2	827	telna
suk1	eroski	185	ab	40	#prod5	308	xcon
suk2	eroski	200	ab	27	#prod7	126	agro



# Diseinu onerako gida-lerroak

1. **gida-lerroa**: Erlazio-eskema erraz azaltzeko moduan definitu behar da

Taula berean entitate bat baino gehiago adieraztea ekidin

Bestela, anbiguitasun semantikoa

Adib. LANGILE-PROIEKTUA ez egokia, entitate ezberdinetako atributuen nahasketa

# Diseinu onerako gida-lerroak

2. **gida-lerroa:** Diseinatu oinarri-erlazioen eskemak, txertatze-, ezabatze- edo eguneratze-akatsik gabe, hau da aldatze-akatsik gabe

Eraginkortasunagatik akatsen arriskua onartzen bada, argi adierazi eta arrazoitu

# Adibide gehiago

## Aldatze-akatsak: Txertatze-akatsa

LANG\_SAIL

<u>NAN</u>	Izena	JaioData	Helbidea	Szenb	Sizena	ZuzNAN
34	Pepe	1960-1-1	Matia, 23-Donostia	1	LSI	30
30	Aitor	1961-2-1	Matia, 23-Donostia	1	LSI	30
20	Miren	1963-6-8	Urdaibai, 2-Bilbo	3	KZAA	14
15	Ana	1960-1-1	Alameda, 54-Bilbo	2	KAT	15
17	Asier	1950-8-1	Heriz, 3-Gazteiz	3	KZAA	14
18	Gorka	1945-7-7	Gaztelu, 6-Iruña	4	LSI	10
16	Ainhoa	1945-5-1	Nagusia, 45-Bilbo	1	LSI	30
14	Leire	1956-8-8	Berria, 21-Iruña	3	KZAA	14
10	Aitor	1950-9-9	Goenkale, 20-Iruña	4	LSI	10

*Suposatu ez direla SAILA eta LANGILEA bakoitza bere aldetik existitzen*

Bi mota:

- (45, 'Josu', '1970-1-1', 'Ona-1, Iruña') langilea 2 kodea duen sailean sartu => Saileko balioak ongi sartu besteeekin koherenteak izan daitezzen
- 5, KZAA saila sartu (langilerik gabea) => Langilean null => Arazoa! NAN-ek ezin du null izan. Adib, SAILA taulan sartzeko arazorik ez.

# Aldatze-akatsak: Ezabatze-akatsa

## LANG\_SAIL

<u>NAN</u>	Izena	JaioData	Helbidea	Szenb	Sizena	ZuzNAN
34	Pepe	1960-1-1	Matia, 23-Donostia	1	LSI	30
30	Aitor	1961-2-1	Matia, 23-Donostia	1	LSI	30
20	Miren	1963-6-8	Urdaibai, 2-Bilbo	3	KZAA	14
15	Ana	1960-1-1	Alameda, 54-Bilbo	2	KAT	15
17	Asier	1950-8-1	Heriz, 3-Gazteiz	3	KZAA	14
18	Gorka	1945-7-7	Gaztelu, 6-Iruña	4	LSI	10
16	Ainhoa	1945-5-1	Nagusia, 45-Bilbo	1	LSI	30
14	Leire	1956-8-8	Berria, 21-Iruña	3	KZAA	14
10	Aitor	1950-9-9	Goenkale, 20-Iruña	4	LSI	10

*Suposatu ez direla SAILA eta LANGILEA bakoitza bere aldetik existitzen*

- LANG\_SAILetik tupla bat ezabatzen badugu, eta sail batentzat lan egiten duen azken langilea bada, sail horri buruzko informazioa DBtik ezabatuta!



# Aldatze-akatsak: Eguneratze-akatsa

## LANG\_SAIL

<u>NAN</u>	Izena	JaioData	Helbidea	Szenb	Sizena	ZuzNAN
34	Pepe	1960-1-1	Matia, 23-Donostia	1	LSI	30
30	Aitor	1961-2-1	Matia, 23-Donostia	1	LSI	30
20	Miren	1963-6-8	Urdaibai, 2-Bilbo	3	KZAA	14
15	Ana	1960-1-1	Alameda, 54-Bilbo	2	KAT	15
17	Asier	1950-8-1	Heriz, 3-Gazteiz	3	KZAA	14
18	Gorka	1945-7-7	Gaztelu, 6-Iruña	4	LSI	10
16	Ainhoa	1945-5-1	Nagusia, 45-Bilbo	1	LSI	30
14	Leire	1956-8-8	Berria, 21-Iruña	3	KZAA	14
10	Aitor	1950-9-9	Goenkale, 20-Iruña	4	LSI	10

*Suposatu ez direla SAILA eta LANGILEA bakoitza bere aldetik existitzen*

- 1. saileko zuzendariaren NAN zenbakia aldatzen badugu => sail horretako langile guztiei zuzendariaren NAN aldatu!

# Diseinu onerako gida-lerro informalak

3. **gida-lerroa**: Ahal dela, saihestu *null* balioak sarri izan ditzaketen atributuak oinarri-erlazioetan kokatzea

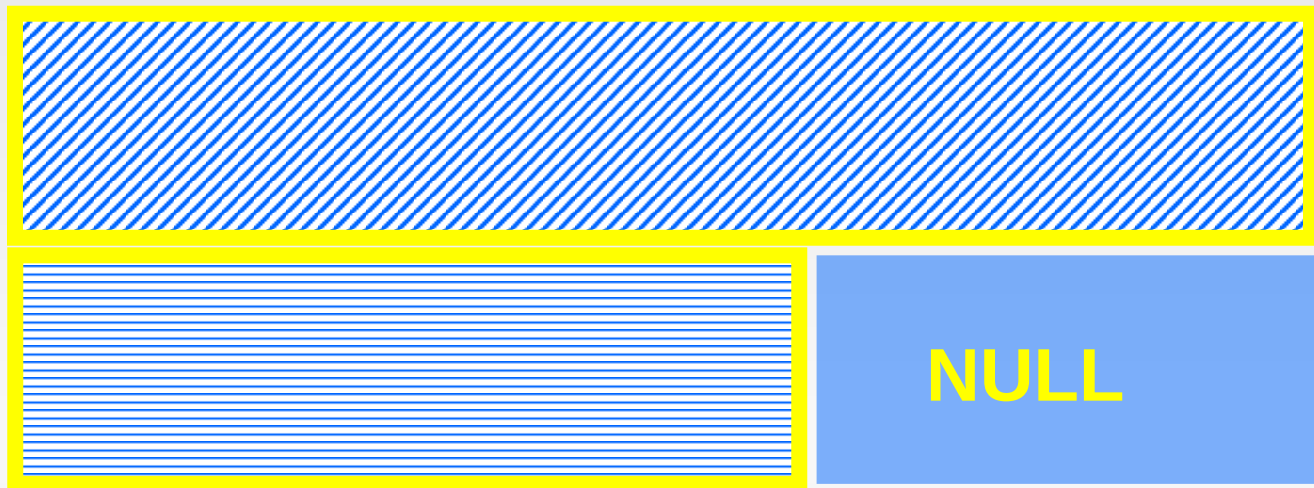
*Null* balio asko:

- Espazioa alferrik galduta
- Atributuaren esanahia ulertzea zaila
- KONBINATU (barne, kanpo) zaila
- COUNT, SUM eragiketak
- Null-en interpretazioak

Saihestu ezin badira, ziurtatu ohiz kanpoko kasuetan soilik aplikatzen direla

# NULL balioak

- Atributu asko ez badira erabiltzen erlazioko tupla guztietan, tupla horietan *null* pila izango dugu.
- Biltegiatze-espazioaren galera sor dezake



# NULL balioak

- Gainera, NULL balioek ondorengo esanahiak izan ditzakete:
  - Ez aplikagarria (N/A)
  - Ezezaguna (Unknown)
  - Ezagutzen da baina falta da

Esanahi ezberdinak nahas daitezke

# Tupla aizunak

- Konbinazio (join) baten emaitzan, “**espero ez diren**” tuplak agertzen dira

Iraklzen	Ikaslelzen
FBD	Alaitz
TBD	Alaitz
TBD	Juan

Iraklzen	Gela
TBD	318
TBD	121
TBD	125
FBD	330
FBD	325
ALG	102
ALG	330

Iraklzen	Ikaslelzen	Gela
FBD	Alaitz	330
FBD	Alaitz	325
TBD	Alaitz	318
TBD	Alaitz	121
TBD	Alaitz	125
TBD	Juan	318
TBD	Juan	121
TBD	Juan	125
ALG	Juan	102
ALG	Juan	330

Tupla aizunek baliozkoa ez den informazioa errepresentatzen dute  
(kasu honetan natural join)

# Diseinu onerako gida-lerro informalak

4. **gida-lerroa:** Erlazioak diseinatu, oinarrizko gakoak eta gako arrotzak berdintasun-baldintzen bidez bildu/konbinatu ahal izateko moduan. Horrela, tupla *aizunik (faltsurik)* ez dela sortuko ziurtatuko dugu.

# Adibide gehiago

## Tupla Aizunak (faltsuak)

LANGILEA

<u>NAN</u>	Izena	Sizena
34	Pepe	LSI
30	Aitor	LSI
20	Miren	KZAA
17	Asier	KZAA
18	Gorka	LSI
16	Ainhoa	LSI
14	Leire	KZAA
10	Aitor	LSI

SAILA

<u>Szenb</u>	Sizena
1	LSI
3	KZAA
4	LSI

LANGILEA  
NATURAL JOIN  
SAILA

Arrazoa: Sizena ez  
OG, ez GA

*Tupla aizuna*

Ikus Lang\_sail taula

<u>NAN</u>	Izena	Sizena	Szenb
34	Pepe	LSI	1
*	34	Pepe	LSI
	30	Aitor	LSI
*	30	Aitor	LSI
	20	Miren	KZAA
	17	Asier	KZAA
*	18	Gorka	LSI
	18	Gorka	LSI
	16	Ainhoa	LSI
*	16	Ainhoa	LSI
	14	Leire	KZAA
*	10	Jon	LSI
	10	Jon	LSI

4<sup>3</sup>

# Laburpena. Arazoak

- Tresna gehigarrik gabe detekta daitezkeenak
  - Erlazio batean txertatzeak eta eguneratzeak egitean lan erredundantea sorraraz dezaketen akatsak, eta ezabaketak egitean nahi gabe informazioa galtzea eragin dezaketenak
  - *Null*-balioen ondorioz gordetze-espazioa alferrik galtzea eta agregazio-eragiketak eta konbinazioak egiteko zailtasunak
  - Behar bezala erlazionatuta ez dauden erlazioekin konbinazioak egitean datu baliogabeak eta aizunak sortzea



# Baina ...

- Nola deskonposatu modu egokian erlazioak?
  - Nola detektatu entitate zuzenak?
- *Bada ...*
  - Mendekotasunetan oinarrituta. **Mendekotasun funtzional** bat datu-baseko bi atributu multzoren arteko murriztapen bat da

## 2. Mendekotasun funtzionala

- Datu-baseko bi atributu multzoren arteko murriztapena da
- Demagun datu-base erlazionalaren eskemak  $n$  atributu dituela,  $A_1, A_2, \dots, A_n$
- Datu-base osoa (eskema unibertsal bakarra):  
 $R = \{A_1, A_2, \dots, A_n\}$

## 2. Mendekotasun funtzionala. Definizioa

- Izan bitez  $X$  eta  $Y$  atributu multzoak eta  $R$  eskemaren azpimultzoak direnak.  $Y$ -k  $X$ rekiko mendekotasun funtzionala duela esango dugu ( $X \rightarrow Y$ ), baldin eta edozein  $r(R)$  estentsiorako  $r$ -ko  $t1$  eta  $t2$  tupla guztientzat zera betetzen bada:

- $t1.X = t2.X \quad \Rightarrow \quad t1.Y = t2.Y$

- Eragiketak erabiliz:

$$\text{PROY}_Y (\text{HAUTESPENA}_{x=\text{'balioa'}} (r)) = \text{tupla bakarra}$$

- $X$ ren balioa ezagututa, modu unibokoan identifikatzen da  $Y$ ren balioa
- Determinatzailea: mendekotasunaren ezker aldea (determinatzailea  $\rightarrow$  atributua)

## 2. Mendekotasun funtzionala. Definizioa

*Y atributu-multzoa X atributu-multzoaren mendeko funtzionala da  
Xtik Yrako mendekotasun funtzionala dago  
X atributu-multzoak Y atributu-multzoa funtzionalki zehazten du*

- Mendekotasun funtzionala atributuen esanahia edo semantikaren ezaugarria da
  - Eskemaren gainean definitzen da
  - Hedapen guztiek betetzen dute

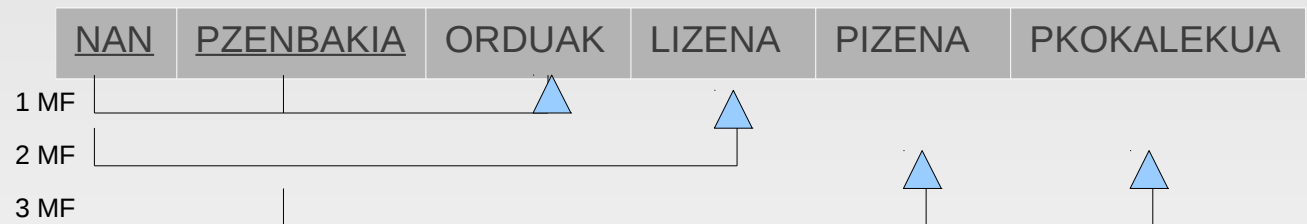
# Mendekotasun funtzionala. Adibidea

- Adibidea (informala):

$\{\text{PROBINTZIA, MATRIKULA\_ZENBAKIA}\} \rightarrow \text{NAN}$

Aplikagarria EAEko pertsona heldu guztientzat

- LANG\_PROI



- $\{\text{NAN, PZENBAKIA}\} \rightarrow \text{ORDUAK}$

$\text{NAN} \rightarrow \text{LIZENA}$

$\text{PZENBAKIA} \rightarrow \{\text{PIZENA, PKOKALEKUA}\}$

- 1 MF: NAN eta PZENBAKIA konbinazioak modu unibokoan identifikatzen du langileak asteko sartzen duen ordu-kopurua
- 2 MF: Langilearen NAN balioak modu unibokoan zehazten du langilearen izena
- 3MF?

# Mendekotasun funtzionala. Adibidea

- Ez dira gertatzen:
  - Irakasgaia  $\rightarrow$  Irakaslea
  - (Irakasgaia, taldea)  $\rightarrow$  Irakaslea
  - (Irakasgaia, gaia)  $\rightarrow$  Irakaslea
- Gertatzen dira
  - (Irakasgaia, taldea, gaia)  $\rightarrow$  Irakaslea
- *Mendekotasunak egiazta al daitezke, taularen estentsio bat begiratuaz?*
  - *EZ! domeinuaren ezaugarri bat dira, integritate-murritzapenak bezala*

# Mendekotasun funtzionalak

- $X$   $R$ -ko gako hautagaia bada
  - $X \rightarrow Y$  ( $R$ -ko edozein  $Y$  azpimultzotarakoa)
- $X \rightarrow Y \not\Rightarrow Y \rightarrow X$ 
  - $Na \rightarrow Tfnoa$  versus  $Tfnoa \rightarrow Na$

# MFak inferitzen

- Mendekotasun-funtzionalak semantikoki begi-bistakoak dira
- MF batzuk *inferitu* edo deduzitu daitezke
- Adib:
  - Sail bakoitzak kudeatzaile bakarra du, eta saila SAIL\_ZKak modu unibokoan zehazten du (SAIL\_ZK → KUD\_NAN)
  - Kudeatzaileak KUD\_TELEFONOA zenbaki bakarra du (KUD\_NAN → KUD\_TELEFONOA)
- Orduan,

SAIL\_ZK → KUD\_TELEFONOA



# Armstrong-en axiomak

## MFtarako Inferentzia-erregelak

MF batzuetatik beste batzuk inferi daitezke, erregela hauek erabiliz:

- Bedi **R** eta bere atributu-multzoak **A**, **B**, **C**

**bihurkorrena:** *baldin*  $B \subseteq A$  *orduan*  $A \rightarrow B$

- $\text{izena} \subseteq \{\text{izena}, \text{tutore\_id}\}$  *orduan*  $\{\text{izena}, \text{tutore\_id}\} \rightarrow \{\text{izena}\}$

**handitzearena:** *baldin*  $A \rightarrow B$  *orduan*  $AC \rightarrow BC$

- $\{\text{tutore\_id}\} \rightarrow \{\text{espezializazioa}\}$  *orduan*  $\{\text{tutore\_id}, \text{izena}\} \rightarrow \{\text{espezializazioa}, \text{izena}\}$

**iragankorrarena:** *baldin*  $A \rightarrow B$  *eta*  $B \rightarrow C$  *orduan*  $A \rightarrow C$

- *Baldin*  $\{\text{tutore\_id}\} \rightarrow \{\text{espezializazioa}\}$  *eta*  $\{\text{espezializazioa}\} \rightarrow \{\text{lab}\}$  *orduan*  $\{\text{tutore\_id}\} \rightarrow \{\text{lab}\}$

# Armstrong-en inferentzia erregelak

- Armstrong-ek (1974) frogatu zuen, inferentzia-erregelak (bihurkorrena, handitzearena, iragankorrena) zuzenak eta osoak direla.
  - **Zuzenak:** R erlazio baten gainean zehaztutako F mendekotasun funtzioen multzo bat izanik, erregelak erabiliz F-tik inferi daitekeen edozein mendekotasun, F-ren mendekotasunak betetzen dituen R-ren egoera guztietan beteko da.
  - **Osoak:** Behin eta berriz aplikatzen baditugu erregelak, gehiago inferitzea posible ez den arte, Ftik inferi daitezkeen mendekotasun posible guztien multzo osoa lortuko dugu.

# MFtarako Inferentzia-erregelak

Bedi **R** eta bere atributu-multzoak **A**, **B**, **C**

## *Sinplifikazio-erregelak*

bildurarena  $A \rightarrow B, A \rightarrow C \Rightarrow A \rightarrow BC$

deskonposaketarena  $A \rightarrow BC$  eta  $B \subseteq BC$   
 $\Rightarrow A \rightarrow B$  eta  $A \rightarrow C$

sasi-iragankorrarena  $A \rightarrow B, CB \rightarrow D \Rightarrow CA \rightarrow D$

# Mendekotasun funtzionalak

- Datu-baseen diseinatzaileek lehenengo atributuen semantikatik abiatuta lor daitekeen mendekotasun funtzionalen  $F$  multzoa zehazten dute. Gero, erregelak erabiltzen dira gainerako mendekotasun funtzionalak lortzeko.

# Mendekotasun funtzionalen multzoaren itxidura (I)

- Bedi  $F$  mendekotasun funtzionalen multzoa,  **$F$ -ren itxidura** ( $F^+$ ) multzotik eratortzen/inferitzen diren mendekotasunen multzoa da (Armstrong-en erregelak aplikatuta)

$$(A \rightarrow B, B \rightarrow C) \Rightarrow (A \rightarrow C)$$

- Bitez  $E$  eta  $F$  mendekotasun funtzionalen bi multzo, **baliokideak** dira, baldin eta  $E^+ = F^+$

# Itxidura. Adibidea

- $F = \{a \rightarrow b, b \rightarrow c, ca \rightarrow d, db \rightarrow e, aej \rightarrow k\}$
- $a^+ = \{a, b, c, d, e\}$
- $ad^+ = \{a, b, c, d, e\}$
- $j^+ = \{j\}$
- $aj^+ = \{a, b, c, d, e, j, k\}$

# Atributu-multzo baten itxidura

- Bedi  $X$  atributu-multzoa,  **$X$ -ren itxidura  $F$ -n oinarrituta** ( $X^+$ ),  $F$  mendekot. funtzionaletan oinarrituta,  $X$ -ren mendekoak diren atributuen multzoa

```
itxidura := {X};                                // erregela bihurkorra
loop
    lag := itxidura;
    for  $F$ -ko  $Y \rightarrow Z$  mendekotasun funtzional bakoitzeko loop
        if  $Y \subseteq \text{itxidura}$  then  $\text{itxidura} := \text{itxidura} \cup \{Z\}$ ; end if;
    end loop;                                    // erregela iragankorra
while not (lag = itxidura)
```

# Itxidura. Beste modu batera esanda...

- Nola egin?
  - Fn mendekotasun funtzionalen batean ezker aldean agertzen diren Xren atributu multzo guztiak zehaztu.
  - Gero, Xren mendeko diren atributu guztiak lortu. Honi  $X^+$  deitzen zaio.



# Atributu-multzo baten itxidura: Adibidea

$$R = \{A, B, C, D, E, G\}$$

$$F = \{ \{A, B\} \rightarrow \{C\}, \{C\} \rightarrow \{A\}, \{B, C\} \rightarrow \{D\}, \{A, C, D\} \rightarrow \{B\}, \{D\} \rightarrow \{E, G\}, \\ \{B, E\} \rightarrow \{C\}, \{C, G\} \rightarrow \{B, D\}, \{C, E\} \rightarrow \{A, G\} \}$$

$$X = \{B, D\}$$

$$X^{(0)} = \{B, D\}$$

- $\{D\} \rightarrow \{E, G\},$

$$X^{(1)} = \{B, D, E, G\},$$

- $\{B, E\} \rightarrow \{C\}$

$$X^{(2)} = \{B, C, D, E, G\},$$

- $\{C, E\} \rightarrow \{A, G\}$

$$X^{(3)} = \{A, B, C, D, E, G\}$$

$$X^{(4)} = X^{(3)}$$

# Itxidura. Beste adibide bat.

**$Na \rightarrow Izena$**

**$IrakKod \rightarrow \{IrakIzen, kred\}$**

**$\{Na, IrakKod\} \rightarrow Gela$**

F

**$\{Na\}^+ = \{Na, Izena\}$**

**$\{IrakKod\}^+ = \{IrakKod, IrakIzen, kred\}$**

**$\{Na, IrakKod\}^+ = \{Na, IrakKod, Izena, IrakIzen, kred, Gela\}$**

# Mendekotasun funtzionalak eta gakoak

Adibidea: Ikaslea(id, izena, tutore\_id, espezializazioa)

- Zera idatziko dugu:

$\{id\} \rightarrow \{izena, tutore\_id, espezializazioa\}$

- Id-k atributu guztiak funtzionalki zehazten ditu (tupla osoa)
- Ikaslea erlazioko bi tuplek id bera badute, atributu guztietan balio berak eduki behar dituzte
- Bestela esanda, atributu bera izan behar dute (eredu erlazionalean ez dira errepikatutako tuplek onartzen)

## Gakoak eta bertako atributuak definitzea

### Supergakoa, Gako hautagaia, Oinarrizko gakoa, Atributu lehena

- A atributu-multzoa R erlazioaren **supergakoa** da:

$$A \rightarrow R$$
$$[ \text{edo } A^+ = R ]$$

- A atributu-multzoa R erlazioaren **gako hautagaia** da:

$$A \text{ supergakoa da, eta}$$
$$Z \subset A \text{ guztietarako } (A-Z) \not\rightarrow R$$

- A atributu-multzoa R-ren **oinarrizko gakoa** da:

gako hautagaietako bat

- Z atributua R-ren atributu **lehena** da, R-ren gako hautagaien batean azaltzen bada

# Gakoak bilatzen

Adibidea: Izan bedi  $R(A,B,C,D)$  erlazio eskema, honako mendekotasun funtzionalekin  $\{A\} \rightarrow \{C\}$  eta  $\{B\} \rightarrow \{D\}$ .

$\{A,B\}$  gako hautagaia al da?

$\{A,B\}$  gako hautagaia izateko, zera bete behar du

- Atributu guztiak funtzionalki zehaztu behar ditu (supergakoa)
- Minimoa izan behar du

$\{A,B\}$  supergakoa da honakoagatik:

- $\{A\} \rightarrow \{C\} \Rightarrow \{A,B\} \rightarrow \{A,B,C\}$  (*AB handitzearena*)
- $\{B\} \rightarrow \{D\} \Rightarrow \{A,B,C\} \rightarrow \{A,B,C,D\}$  (*A,B,C handitzearena*)
- $\{A,B\} \rightarrow \{A,B,C,D\}$  lortzen dugu (*iragankorra*)

Beste *modu batera*, *atributuen itxidura erabiliz*:

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

$\{A,B\}$  minimoa da, ez  $\{A\}$  bere aldetik, ezta  $\{B\}$  bere aldetik ere, ez dira gako hautagaiak

# Supergakoa lortzen. Adibidea

- Bisita (gaixoa, ospitalea, sendagilea)
- $F = \{ (gaixoa, ospitalea) \rightarrow sendagilea ; sendagilea \rightarrow ospitalea \}$
- $gaixoa^+ = \{gaixoa\}$
- $ospitalea^+ = \{ospitalea\}$
- $sendagilea^+ = \{sendagilea, ospitalea\}$
- $(ospitalea, sendagilea)^+ = \{ospitalea, sendagilea\}$
- $(gaixoa, ospitalea)^+ = \{gaixoa, ospitalea, sendagilea\}$  supergakoa
- $(gaixoa, sendagilea)^+ = \{gaixoa, sendagilea, ospitalea\}$  supergakoa
- $(gaixoa, sendagilea, ospitalea)^+ = \{gaixoa, sendagilea, ospitalea\}$ 
  - BAI da supergakoa baina ez da gakoa, ez baita minimoa

# Gakoak bilatzen

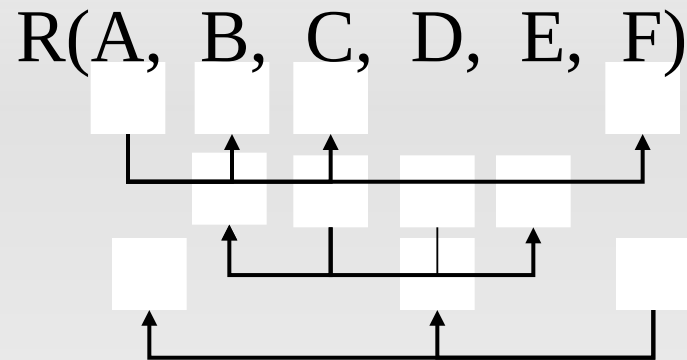
- Informazio gehigarria:
- [http://www.youtube.com/watch?v=s1DNVWKeQ\\_w](http://www.youtube.com/watch?v=s1DNVWKeQ_w)

# Supergakoa, Gako hautagaia, Oinarrizko gakoa, Atributu lehenak. Adibidea

$A \rightarrow BCF$

$CD \rightarrow BE$

$F \rightarrow AD$



Supergakoak:

Gako hautagaiak:

Oinarrizko gakoa:

Atributu lehenak:

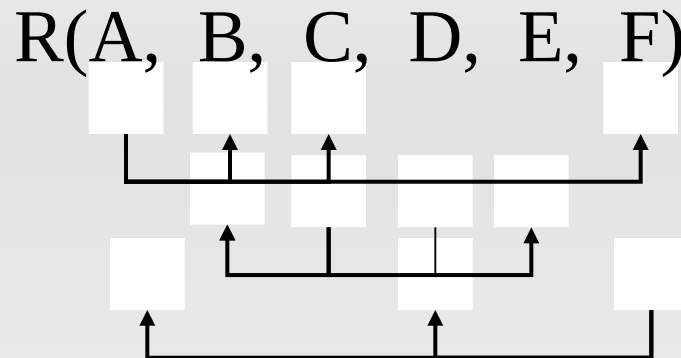


# Supergakoa, Gako hautagaia, Oinarrizko gakoa, Atributu lehenak. Adibidea

$A \rightarrow BCF$

$CD \rightarrow BE$

$F \rightarrow AD$



Supergakoak:  $CDA, CDF, CDAF$ , etab.

Gako hautagaiak:  $A, F$

Oinarrizko gakoa:  $A$  edo  $F$

Atributu lehenak:  $A$  eta  $F$

# MF multzoen baliokidetasuna

F-k G estaltzen du, Gko MF guztiak  $F^+$ -n ere badaude  
F eta G **baliokideak** dira, *haien itxidurak baliokideak badira*,  $F^+ = G^+$

Adibidez bi multzo hauek baliokideak dira:

1.  $\{XY \rightarrow Z, X \rightarrow Y\}$
2.  $\{X \rightarrow Z, X \rightarrow Y\}$

- Baliokidetasunak zera esan nahi du: Fko MF guztiak Gtik inferi daitezkeela, eta Gko MF guztiak Ftik inferi daitezkeela.
- Beste modu batera: Fk G estaltzen du eta Gk F estaltzen du

# Estaldura Minimoa. Motibazioa

- Izan bedi  $F$ , mendekotasun funtzionalen multzo bat. Helburua, beste  $F_c$  multzo bat topatzea da,  $F$ ren baliokidea dena, baina zeinetan egiaztatu beharreko mendekotasun kopurua minimoa den

$F$  multzoa  $G$ ren *estaldura minimoa* da,  $F$  minimoa bada eta  $G$ ren baliokidea

- Adib.

$\{X \rightarrow Z, X \rightarrow Y\}$  estaldura minimoa da ondorengoarentzako

$$\{XY \rightarrow Z, X \rightarrow Y\}$$

# Estaldura Minimoa

- Mendekotasun funtzionalen multzo batek estaldura minimo batzuk izan ditzake (beti, gutxienez bat topa daiteke)

# Estaldura Minimoa. Definizioa

MF minimoa da baldin eta,

- Fko mendekotasun guztiek atributu bakarra dute eskuin aldean
  - motibazioa: MF guztiek forma estandar bera izan dezaten nahi dugu, algoritmoek prozesa dezaten. Deskonposizioaren erregelaren bitartez MF guztiak forma estandarrera bihur daitezke
- Ezin dugu  $F_{tik}$  inolako mendekotasunik kendu eta Fren baliokidea den mendekotasun multzoa izaten jarraitu
  - motibazioa: mendekotasun erredundanteak kentzea
- Ezin dugu inolako  $X \rightarrow A$  mendekotasunik ordeztu  $Y \subseteq X$  izanik  $Y \rightarrow A$  batekin eta Fren baliokidea den MF multzo bat edukitzen jarraitu
  - motibazioa: atributu erredundanteak kentzea

# Ariketa

- Ftik abiatuta, lortu MF multzo minimoa
  - 1º) Forma kanonikoa

F

$Na \rightarrow Izena$ $\{Na, IrakKod\} \rightarrow Gela$ $IrakKod \rightarrow \{Iraklzen, kred\}$
--

$IrakKod \rightarrow \{Iraklzen, kred\}$



deskonposizioa

$IrakKod \rightarrow Iraklzen$   
 $IrakKod \rightarrow kred$

# Ariketa

- Ftik abiatuta, lortu MF multzo minimoa
  - 1º) Forma kanonikoa
  - 2º) Ezin da inolako eskuineko alderik kendu

**Na → Izena**

**{Na, IrakKod} → Gela**

**IrakKod → Iraklzen**

**IrakKod → kred**

# Ariketa

- Ftik abiatuta, lortu MF multzo minimoa
  - 1º) Forma kanonikoa
  - 2º) Ezin da inolako eskuineko alderik kendu
  - 3º) Ezin da inolako ezkerreko alderik laburtu

**Na → Izena**

**{Na, IrakKod} → Gela**

**IrakKod → Iraklzen**

**IrakKod → kred**



# Adibidea

- $F = \{A \rightarrow BC, A \rightarrow C, B \rightarrow C, AB \rightarrow E\}$
- Estaldura minimoa lortzea
  - 1.Deskonposizioaren erregela erabiliz eskuin aldean atributu bat baino gehiago duten MFak banatu egiten dira:  $A \rightarrow BC$ , honakoetan banatzen da  $A \rightarrow B$  eta  $A \rightarrow C$
  2. $A \rightarrow C$  erredundantea da.  $A \rightarrow B$  eta  $B \rightarrow C$  erregelatan iragankortasuna aplikatuz lor daiteke
  3. $AB \rightarrow E$  erregela  $A \rightarrow E$  moduan geratzen da B erredundante baita ( $A \rightarrow B$  betetzen da)
- Beraz, Frentzako estaldura minimoa honakoa da:
  - $A \rightarrow B$
  - $B \rightarrow C$
  - $A \rightarrow E$

# Mendekotasun trantsitiboak

IKASLEA(EspZbkia, IkasleIzen, Fakultatea, Dekanoa, ...)

$\text{EspZbkia} \rightarrow \text{Fakultatea}$

$\text{Fakultatea} \not\rightarrow \text{EspZbkia}$

$\text{Fakultatea} \rightarrow \text{Dekanoa}$

Fakultatean bitartez, Dekanoa EspZbkiaren mendeko da

- Izan bitez **A**, **B**, **C** atributu multzoak (R eskemakoak). **A** iragankortasuna erabiliz **C** ren mendekoa da **R**pean baldin eta

$A \rightarrow B$

$B \not\rightarrow A$  (ez da betetzen)

$B \rightarrow C$

# Eskema baten deskonposizioa

- **R(A, F) erlazio-eskema**

A, atributu-multzoa, F mendekotasun funtzionalen multzoa

R-n akatsak sortzeko arriskua badago, deskonposatu

$\{R_i(A_i, F_i)\}$  non  $i=1\dots n$

- Deskonposizioa

## **R-ren baliokidea**

Atributuen kontserbazioa:  $A = A_1 \cup A_2 \cup \dots \cup A_n$ , hau da A-ko atributu guztiak  $A_n$ -ren batean daude

Mendekotasun funtzionalen kontserbazioa

Informazioaren kontserbazioa

## **R baino hobe**

Erredundantzia gutxiago (Akats gutxiago)

# Mendekotasun funtzionalen kontserbazioa (I)

R-ko mendekotasun funtzionalen multzoa,  $F$

$R_i$ -ko mendekotasun funtzionalen multzoa,  $F_i$

$\{R_1, R_2, \dots, R_n\}$  multzoa **mendekotasun funtzionalak kontserbatzen** dituen **deskonposizioa**, baldin eta

$$F^+ = (\cup F_i)^+$$

$$F \subset (\cup F_i)^+$$

Oharra:  $R_i$  deskonposizioaren ondorengo erlazioa da.

# Informazioaren kontserbazioa (I)

- $\{R_1, R_2, \dots, R_n\}$  multzoa informazioaren galerarik gabeko deskonposizioa da,  $R$ -ren edozein  $r$  hedapenerako betetzen bada

$$r = r_1 |X| r_2 |X| r_3 |X| \dots |X| r_n$$

non  $r_i$ ,  $R_i$ -ren atributuak hartuz,  $r$ -ren proiektzioa den

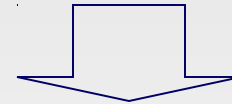
Hau da, tupla faltsurik/aizunik sortzen ez denean

# R-ren galerarik gabeko deskonposizioa

Azter ezazu honako deskonposizioa

$AB \rightarrow D$   
 $AC \rightarrow E$   
 $ABD \rightarrow F$

$R(\underline{A}, \underline{B}, \underline{C}, D, E, F)$



$R1(\underline{A}, \underline{B}, D, F)$

$R2(\underline{A}, \underline{C}, E)$

Ez da MFrik galtzen baina bai ABC hirukoteak

# Informazioaren kontserbazioa (II)

- R-ko mendekotasun funtzionalen multzoa,  $F$   $\{R1, R2\}$  informazioa kontserbatzen duen R-ren deskonposizioa da,  $F^+$ en gutxienez hauetako bat azaltzen bada

$R1 \cap R2 \rightarrow R1$ , hots  $R1 \cap R2$  atributua  $R1$ en supergakoa da

$R1 \cap R2 \rightarrow R2$ , hots  $R1 \cap R2$  atributua  $R2$ ren supergakoa da

# Jarraipena

- Izan bitez:
  - $F, R$ ko mendekotasun funtzionalen multzoa
  - $F_i, R_i$  barruko mendekotasunak (beraz, egiaztapenak ez du JOINik behar)
- Deskonposaketa batek **mendekotasunak mantentzen** ditu baldin eta:

$$F^+ = E^+ \text{ non } E = \cup F_i$$



# Adibidea

$R = \{(liburua, argitaletx., herri), \{liburua \rightarrow argitaletx., argitaletx. \rightarrow herri\}\}$

Proiektzioa?

Atributu  
guztiak?

Tupla aizunik gabe

Mendekotasunik  
galdu gabe

$R[liburua, herri]$   
 $R[argitaletx., herri]$

Bai

Bai

Ez. Ez dira  
betetzen  
herri komunean

Ez  
 $liburua \rightarrow argitaletx.$   
Galtzen da

$R[liburua, argitaletx.]$   
 $R[liburua, herri]$

Bai

Bai

Bai.  
 $Liburua \rightarrow$   
 $argitaletx.$   
Betetzen da

Ez.  
 $argitaletx. \rightarrow herri$   
Galtzen da

$R[liburua, argitaletx.]$   
 $R[argitaletx., herri]$

Bai

Bai

Bai.  
 $argitaletx. \rightarrow herri$   
Betetzen da

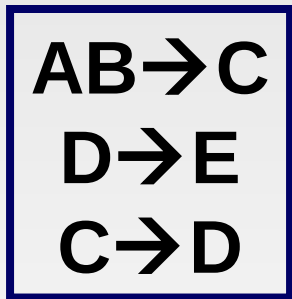
Bai



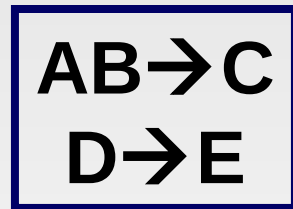
# Deskonposizioa mendekotasunak mantenduz

- Adibidea 1
  - $R1(ABC)$  eta  $R2(DE)$

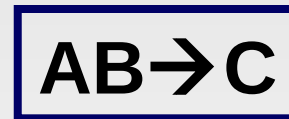
$F$



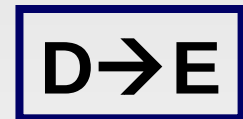
$E = \cup F_i$



$F_1$



$F_2$



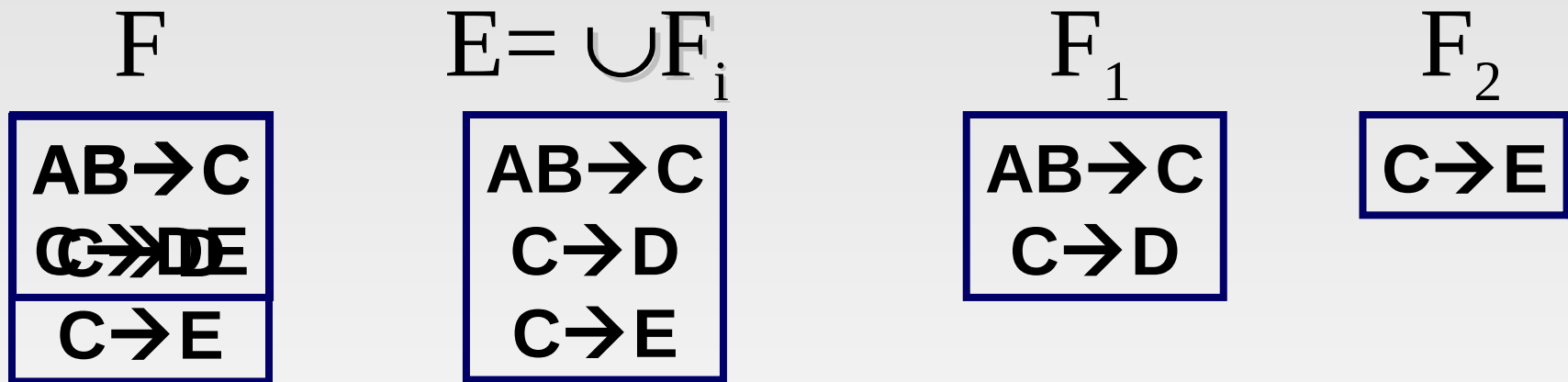
$C \rightarrow D$ , E-tik?

**EZ**

Ez dira mendekotasunak mantentzen

# Deskonposizioa mendekotasunak mantenduz

- Adibidea 2
  - $R1(ABCD)$  eta  $R2(CE)$



**E eta F berdina dira**

**BAI mantentzen dira mendekotasunak**

# Deskonposizioa mendekotasunak mantenduz

- Adibidea 3
  - $R1(ABC)$  eta  $R2(DE)$

$F$

$AB \rightarrow C$
$CD \rightarrow E$

$E = \cup F_i$

$AB \rightarrow C$

$F_1$

$AB \rightarrow C$
--------------------

$F_2$

--

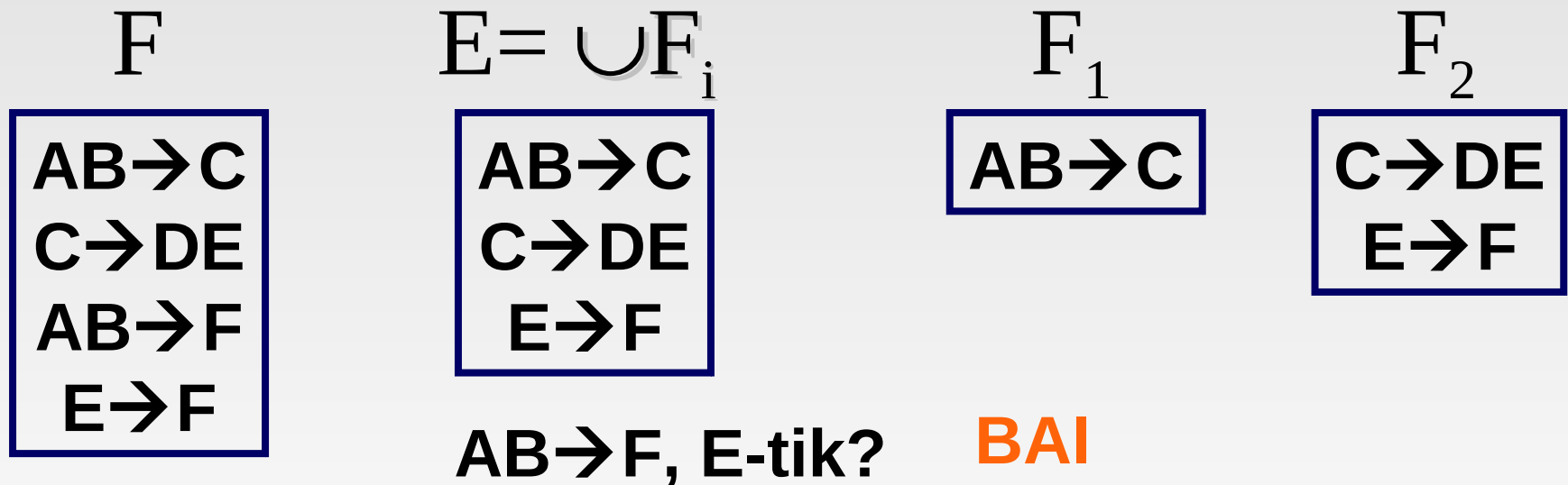
$CD \rightarrow E$  , E-tik?

**EZ**

Ez dira mendekotasunak mantentzen

# Deskonposizioa mendekotasunak mantenduz

- Adibidea 4
  - $R1(ABC)$  eta  $R2(CDEF)$



BAI mantentzen dira mendekotasunak

# Mendekotasun funtzionalen galera. Ondorioaz jabetzea

$A \rightarrow B$   
 $R \quad B \rightarrow C$

A	B	C
a1	b1	c1
a2	b1	c1
a3	b3	c1

**Deskonposizioa**

$B \rightarrow C$  mendek. funtzionala galdu da

R1		R2	
A	B	A	C
a1	b1	a1	c1
a2	b1	a2	c1
a3	b3	a3	c1



R1-en eta R2-n (a4, b3) eta (a4, c4) tuplak sartzea, jatorrizko R taulan nola adieraziko litzateke?

(a4, b3, c4) tupla sartuta. Baina hori ezinezkoa da aurretik (a3, b3, c1) badagoelako. Kontrol hori  $B \rightarrow C$  mendekotasunak darama. Deskonposaketaren ondoriozko egoran ez da kontrol hori eramaten, mendekotasuna desagertu delako.

# 3. Oinarritzko gakoetan oinarritutako Forma Normalak

- Normalizazioa: Codd (1972). Erlazio-eskema bati hainbat test aplikatzen dizkio, **forma normal** jakin bat betetzen ote duen jakiteko
- Def: Datuen **normalizazioa** erlazio-eskemak haien MFetan eta oinarritzko gakoetan oinarrituta, analizatzeko prozesua da. Kalitate handiko diseinuak lortzen dira
- Def2: Normalizazioa: erlazio “txarrak” deskonposatzeko prozesua, erlazioaren atributuak erlazio txikiagotan banatuaz
- Helburua: erredundantzia eta txertatze-, ezabatze eta aldatze-akatsak minimizatzea

# Jarraipena.

- Izan bedi  $R(A, D)$  erlazio eskema bat,  $A$  atributu multzo bat eta  $D$ , mendekotausnen multzo bat
- $R_n$  akatsak gertatzen badira,
  - $\{R_i(A_i, D_i)\}$  azpieskemetan banatzen da non  $i=1\dots n$ .
- Deskonposizioak honelakoa izan behar du

## ➤ **R-ren baliokidea**

- Atributuen kontserbazioa:  $\cup A_i = A$
- Edukiaren kontserbazioa:  $r_1 * r_2 * r_3 * \dots * r_n = r$
- Mendekotasunaren kontserbazioa:  $E^+ = D^+$  non  $E = \cup D_i$   
(BCFNan izan ezik)

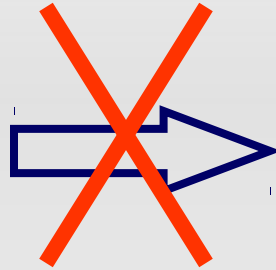
## ➤ **R baino hobe**

- Erredundantzia txikiagoa (akats kopuru txikiagoa)



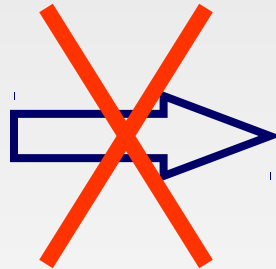
# Normalizazio prozesua (2)

**Erlazio  
Normalizatua**



**Diseinu  
ona**

**Diseinu  
ona**



**FN altuenera  
iritsi**

**Errendimentua  
kontuan hartu!**

# Idea intuitibo batzuk ...

<u>SukKod</u>	<u>Enplzen</u>	SukTfno	IFK	Uni	<u>ProKod</u>	Sal	Ban
suk1	corte_inglés	193	ci	12	#prod1	123	agro
suk2	corte_inglés	294	ci	18	#prod1	123	agro
suk2	corte_inglés	294	ci	7	#prod2	827	telna
suk1	eroski	185	ab	103	#prod2	827	telna

## Eguneratze-akatsa:

- #prod1 produktuaren salneurria eguneratzeak, tupla batzuk aldatzea dakar

## Arrazoa:

- *KodPro-k Sal zehazten du: KodPro  $\rightarrow$  Sal Ban*
- *KodPro ez da supergakoa, beraz, tupla batzuk egon daitezke salneurri berarekin*

## Ondorioa: erredundantzia

- Produktu baten salneurria hainbat tupletan errepikatzen da

# Entitateak ezagutzen...

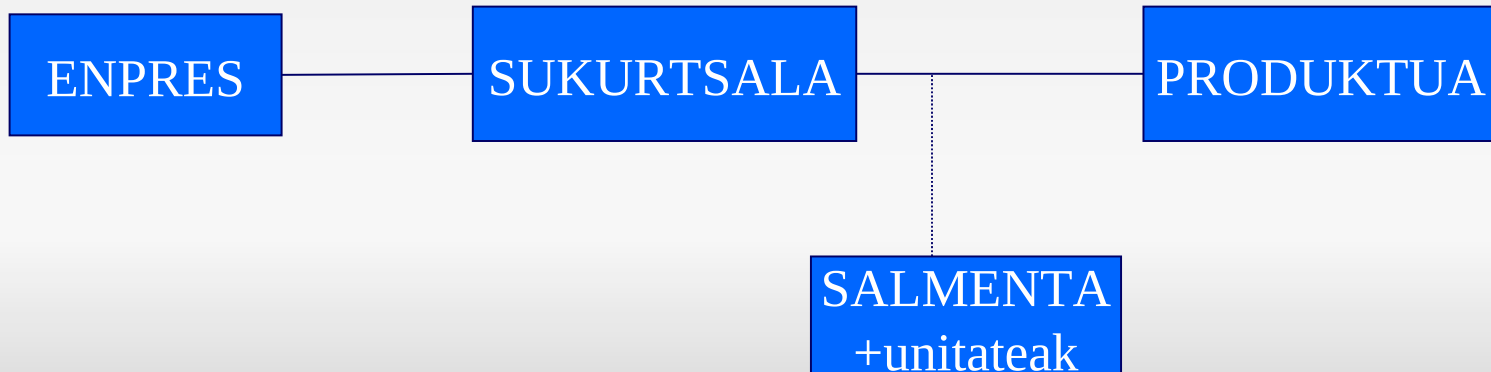
*SALMENTAK[ProKod, Salneurria, Banatz.] → PRODUKTU entitatea*

*SALMENTAK[SukKod, EnpIzen, SukTfnoa] → SUKURTSALA entitatea*

*SALMENTAK[EnpIzen, IFK] → ENPRESA entitatea*

*SALMENTAK(SukKod, EnpIzen, Unitateak, ProKod) → SALMENTA entitatea*

Diseinu on batek entitateak ezagutzen ditu!!  
Normalizazioa, beraz, egiaztatze prozesua litzateke...



# Normalizazio prozesua

- Prozesu bat da zeinetan eskema bati hainbat proba egiten zaizkion forma normal zehatz batean ote dagoen ezagutzeko.
- Forma normalak
  - Erlazio bat **1FN** moduan dago baldin eta soilik baldin atributu guztiak atomikoak dituen
  - Erlazio bat **2FN** moduan dago baldin eta soilik baldin **1FN**n dago eta
    - **Lehena ez den** atributu oro, gako hautagaiekiko guztiz mendekoa den
  - Erlazio bat **3FN** moduan dago baldin eta soilik baldin **2FN**n dago eta
    - Inolako **lehena ez den** atributuk ez duen **mendekotasun iragankorrik gako hautagaietatik**

# Forma Normalak

- Forma normalak:
  - **1FN, 2FN, 3FN (Codd)**
  - **3FN sakonago = BCFN (Boyce and Codd)**
  - 4FN, 5FN
  - Normalean 3FN, BCFN edo 4FN arte
- Propietateak:
  - *Galerarik gabeko konbinazioaren propietatea (ezinbestekoa):* deskonposatu ondoren erlazio-eskemetan ez da tupla aizunik egongo
  - *Mendekotasunak kontserbatzearen propietatea (komenigarria):* MF bakoitza deskonposizioaren bidez lorturiko banakako erlazioen batean ordezkatuta dago

# Normalizazioa

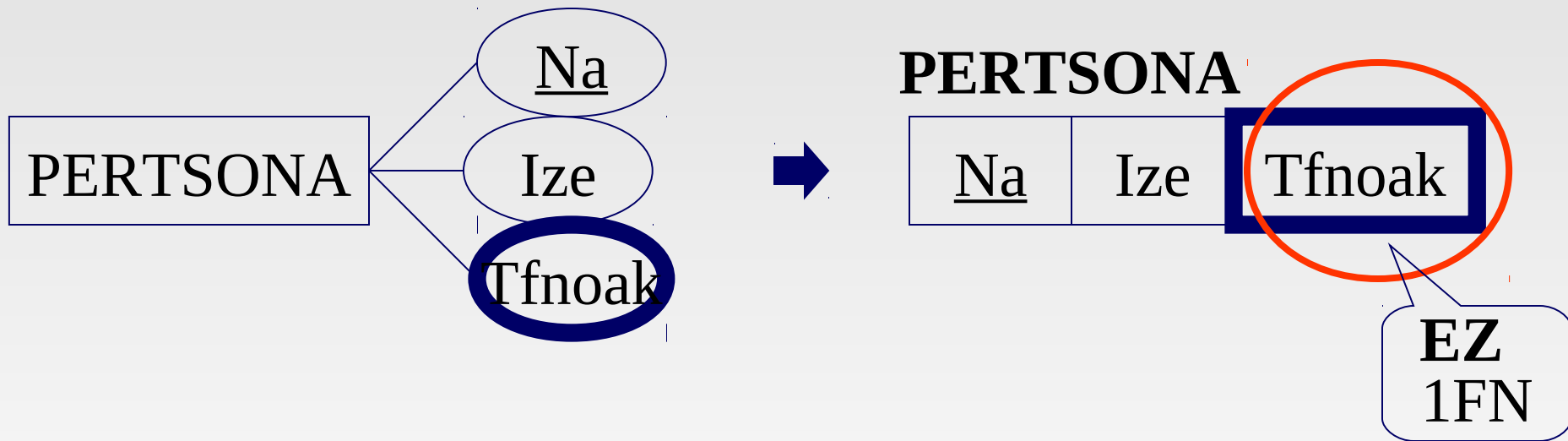
Erredundantzia gutxitzea helburu duen prozesua  
Hots, aldatze eragiketak exekutatzearan akatsak  
sortzeko arriskua gutxitzea

Lehen Forma Normala  
Bigarren Forma Normala  
Hirugarren Forma Normala  
Boyce-Codd-en Forma Normala  
Laugarren Forma Normala

# Lehen Forma Normala (1FN)

- Baldintza: atributuek balio atomikoak izan behar dituzte
  - Tupla bakoitzerako, atributu batek balio bat:
    - Ez dago multzo errepikakorrik
    - Ez dago atributu konposaturik
  - Eredu erlazionaletik datorren murriztapena

# Lehen Forma Normala (1FN)



**SOLUZIOA:**

**PERTSONA1**

<u>Na</u>	Ize
-----------	-----

**PERTSONA2**

<u>Na</u>	<u>Tfnoa</u>
-----------	--------------



# Lehen Forma Normala (1FN)

Atributuen balioak atomikoak izan behar dira  
(sinpleak eta banaezinak)

*Atributu  
balioanitza*

*~~SAILA(Szenb, Sizena, zuzenNAN, STokiak)~~*

<b>Szenb</b>	<b>Sizena</b>	<b>zuzenNAN</b>	<b>STokiak</b>
1	LSI	30000000	{Donostia, Eibar, Gazteiz}
2	KAT	15000000	{Donostia}
3	KZAA	14000000	{Leioa, Donostia}

# Lehen Forma Normala. Nola lortu?

1. *Stokiak* atributua kendu eta kokatu *SAILEN\_TOKIAK* izeneko beste erlazio batean, *Szenb* oinarritzko gakoarekin batera. Erlazio berriaren OG {*Szenb*, *Stokia*} da. 1FN ez den erlazioa bi 1FNtan deskonposatzen da.
2. Hedatu gakoa, *SAILAren* kokaleku bakoitzerako tupla bat. OG {*Szenb*, *Stokiak*} da. Desabantaila: erredundantzia!

Adib:

<b>Szenb</b>	<b>Sizena</b>	<b>zuzenNAN</b>	<b>STokiak</b>
1	LSI	30000000	Donostia
1	LSI	30000000	Eibar
1	LSI	30000000	Gazteiz
2	KAT	15000000	Donostia
3	KZAA	14000000	Leioa
3	KZAA	14000000	Donostia

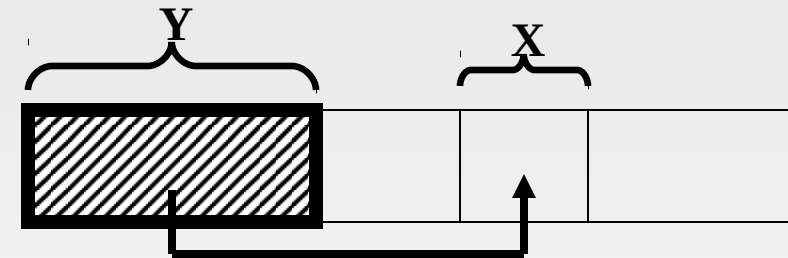
# Lehen Forma Normala. Nola lortu?

3. “*Gehienezko balio kopurua ezagutzen bada*”, adibidez sail batek gehienez 3 leku, ordeztu hiru atributu atomikorekin: STOKIA1, STOKIA2 eta STOKIA3. Desabantaila: *null* balioak sartzen dira. Kontsulta zailagoak.
- Aurreko hiru irtenbideetatik ***lehena*** hoberena

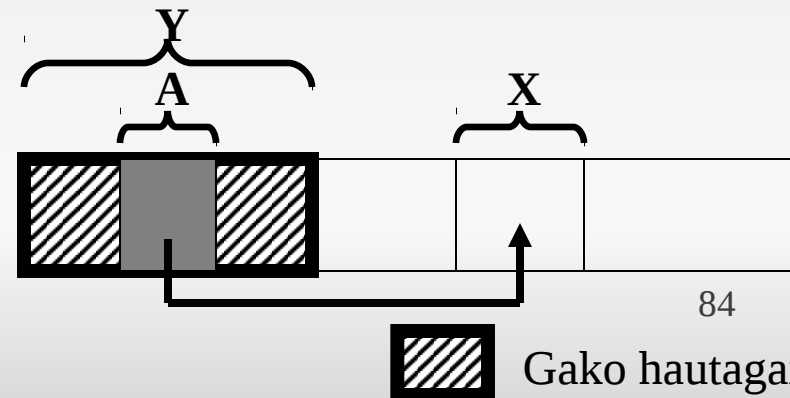
# Bigarren Forma Normala (2FN)

- Erlazioa **2FN**an dago
  - 1FNan badago, eta
  - Ez-lehenak diren atributu **guztiak gako hautagai guztien mendeko osoak** badira

$X$ ,  $Y$ ren *mendeko osoa* da, baldin  $Y \rightarrow X$  eta  $(Y-A) \not\rightarrow X$   
( $Y$ tik  $A$  atributu bat kenduta, mendekotasuna galdu egiten da)



$X$  atributua  $Y$ ren *mendeko partziala* da, baldin  $Y \rightarrow X$  eta  $(Y-A) \rightarrow X$   
( $Y$ tik  $A$  atributu bat kenduta, mendekotasuna mantendu egiten da)



# Bigarren Forma Normala (2FN)

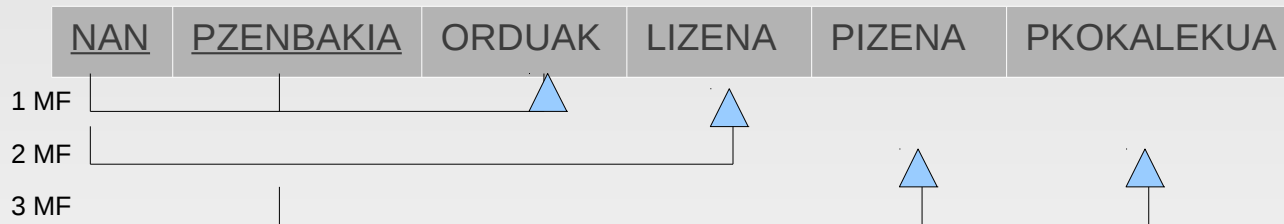
- Mendekotasun osoa vs partziala. Adibideak:
  - $\{NAN, PZENBAKIA\} \rightarrow ORDUAK$  mendekotasun osoa  
NAN kenduta MF galdu  
PZENBAKIA kenduta MF galdu  
(ez  $NAN \rightarrow ORDUAK$ , ez  $PZENBAKIA \rightarrow ORDUAK$  ez dira betetzen)
  - $\{NAN, PZENBKIA\} \rightarrow LIZENA$  mendekotasun partziala  
 $NAN \rightarrow LIZENA$  betetzen da

# Bigarren Forma Normala (2FN)

- Nola lortu:
  - Mendekotasun partzialak taula berrietara eramaten dira
  - Hasierako taulan gakoa eta gakotik guztiz mendekoak diren atributuak geratzen dira

# Bigarren Forma Normala

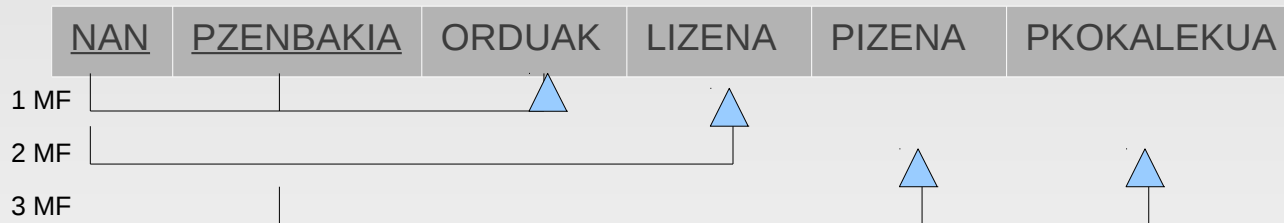
- 2FNrako proba: ezker aldeko atributuak oinarritzko gakoaren barruan dituzten mendekotasun funtzionalak aztertu behar dira. OGk atributu bakarra badu, ez egin proba
- Adib: LAN\_PROI 1FN bai, 2FN ez!



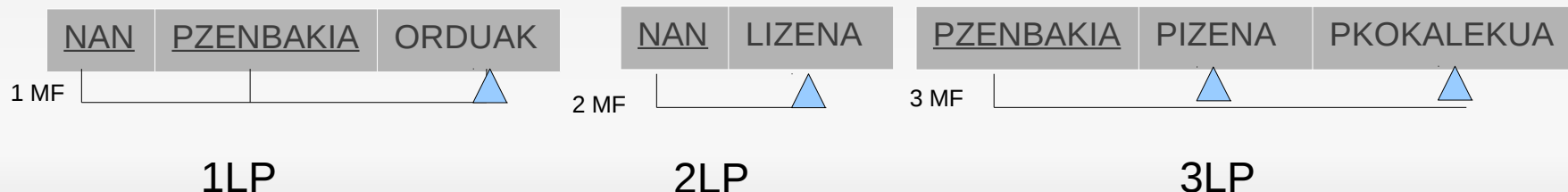
- 2FN ez:
  - 2MF eta 3MFren ondorioz, LIZENA, PIZENA eta PKOKALEKUAK LANG\_PROIko {NAN, PZENBAKIA} oinarritzko gako OSOAREKIN partzialki mendekoak dira → 2FN hautsi!
  - Adib. NAN → LIZENA, PIZENA → PZENBAKIA, PZENBAKIA → PKOKALEKUA partzialak

# Bigarren Forma Normala. Soluzioa

- Bigarren normalizazioa: Hainbat 2FN erlazio bihurtu. Erlazio horietan, atributu ez-lehenak mendeotasun funtzional osoa duten oinarrizko gakoaren zatiarekin bakarrik daude lotuta
- Adibidea: LANG\_PROI, 1LP, 2LP eta 3LP erlazioetan **Deskonposatu**

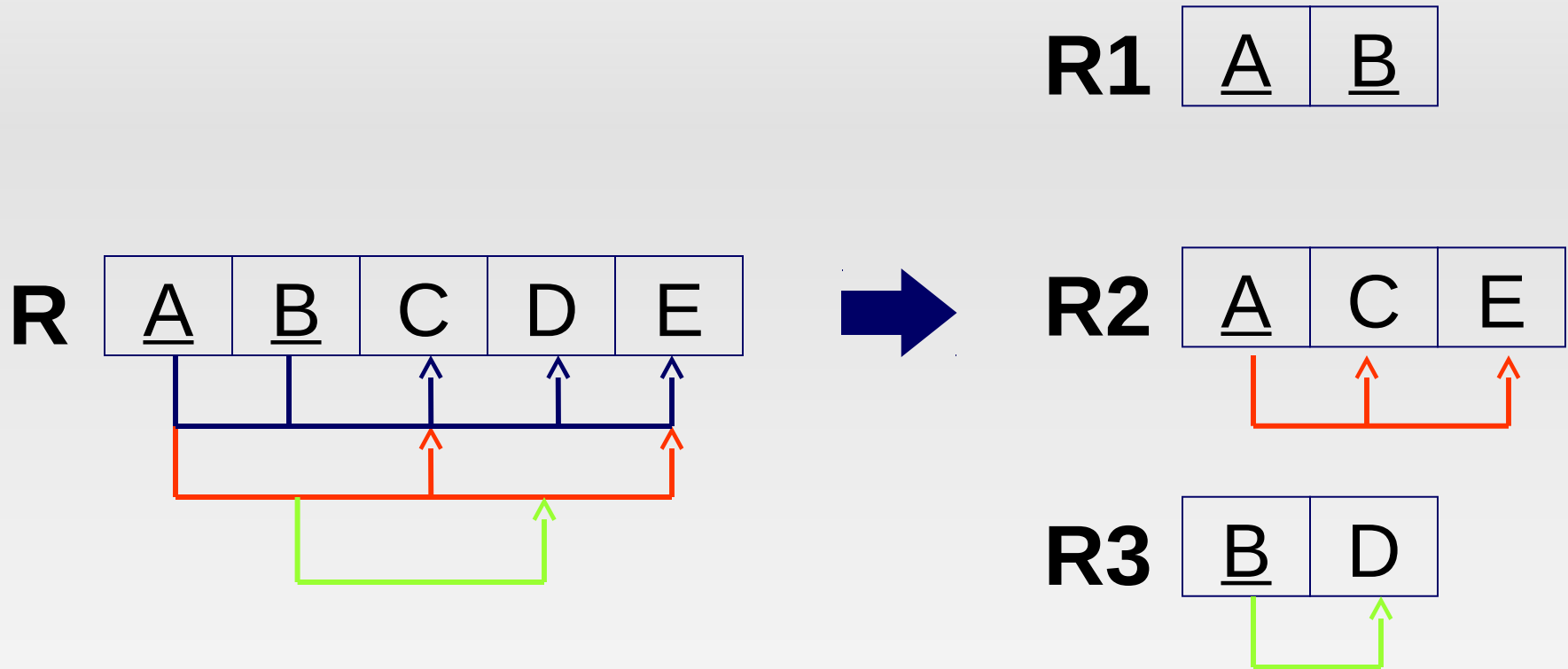


- 2FN NORMALIZAZIOA (ez-lehen guztiek MF OSOA. **Ezin dute MF partzialik izan!**):





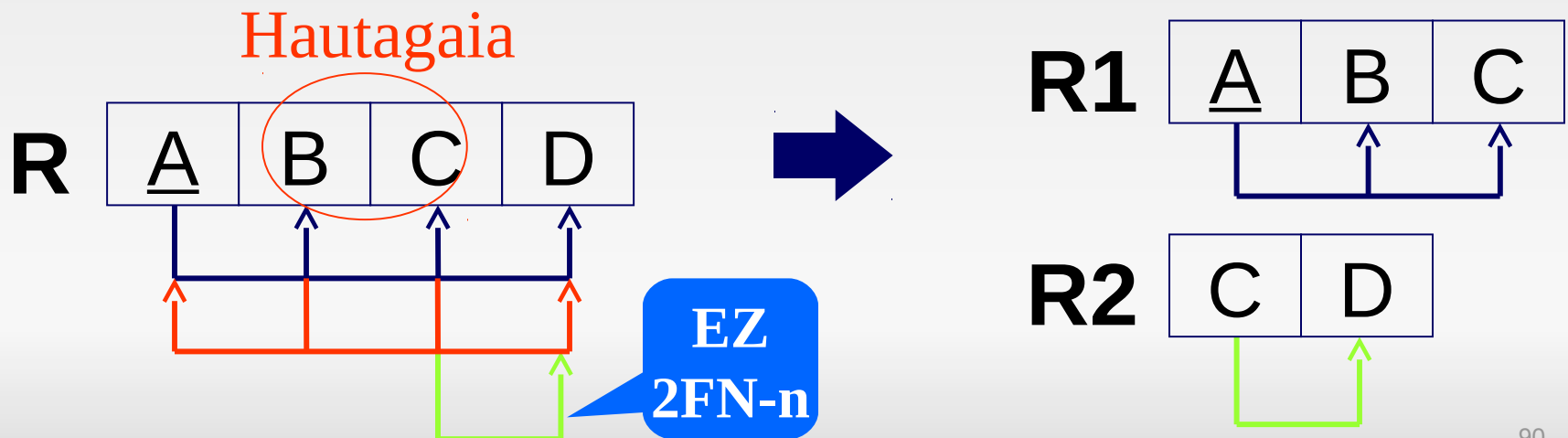
# 2FN: adibidea



# 2FN gako hautagaiaz

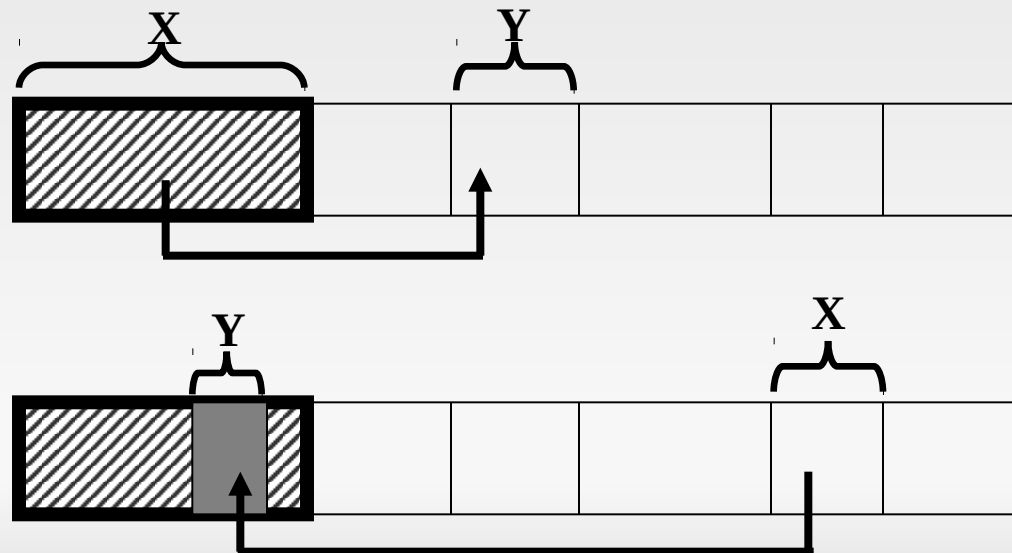
- Eta, **mendekotasun partzialak** badaude baina **gako hautagai** batekin, **oinarrizko gako** batekin beharrean?

Modu berean ebazten da



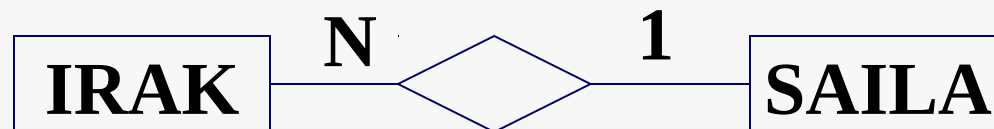
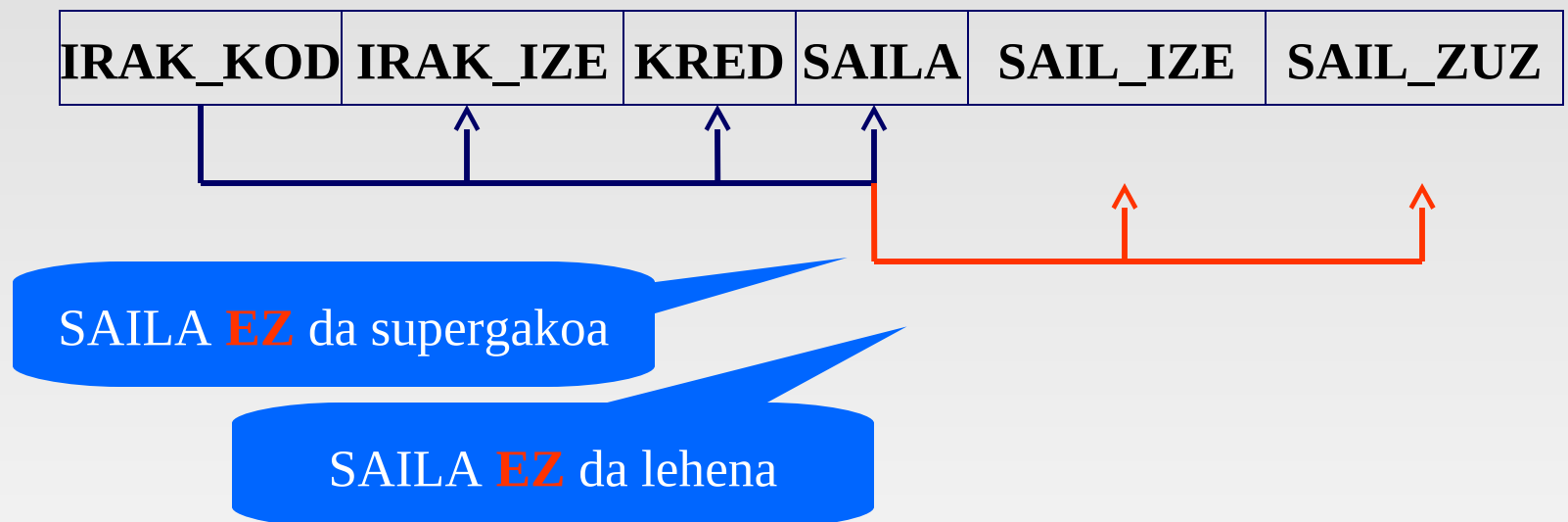
# Hirugarren Forma Normala (3FN)

- Erlazioa 3FNan dago
- 2FNan badago, eta
- $X \rightarrow Y$  mendeotasun guztietan **X supergakoa** bada *edo* **Y atributu lehena** bada



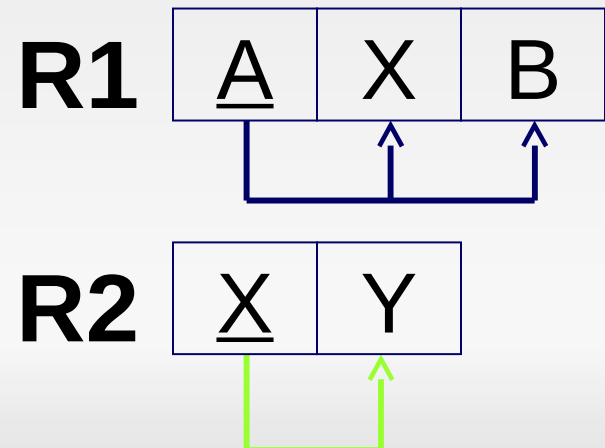
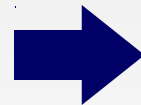
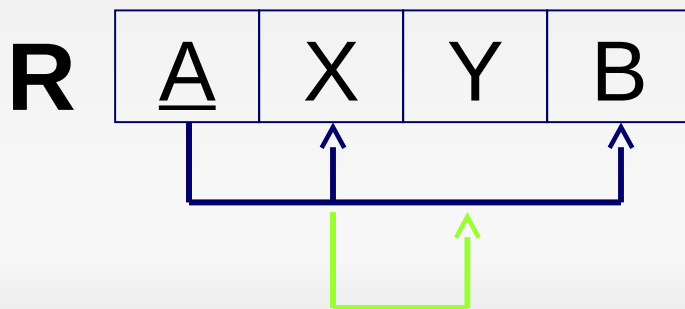
# 3FN: adibidea

Ez dago 3FNn



# 3FN: Soluzioa

- $R(A, X, Y, B)$  non  $X \rightarrow Y$  mendekotasun funtzionalak ez duen 3FNa betetzen
- Sortu  $R'$  erlazioa  $X^+$  duena eta  $X$  gakotzat duena
- $R$ -tik  $Y$  ezabatu

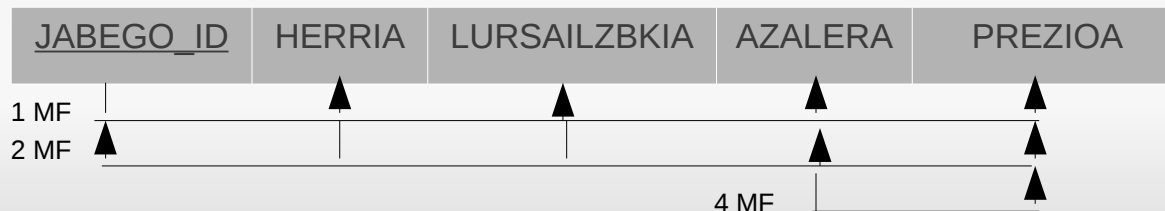


# Normalizazioa. Adibidea.

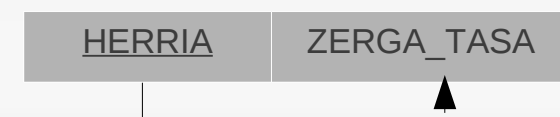


- Gako hautagaiak: JABEGO\_ID eta {HERRIA, LURSAILZBKIA}, beraz, 1MF eta 2MF betetzen dira
- 3MF: zerga-tasa finkoa da herri jakin batean (Ez da aldatzen herri bereko lursail batetik bestera)
- 4MF: lursail baten prezioa bere azaleraren araberakoa da (berdin dio eskualdea)
- 2 forma normala hausten al da? BAI
  - ZERGA\_TASA atributuak {HERRIA, LURSAILZBKIA} gako hautagaiarekiko mendekotasun partziala => 3MF ez dago 2FNn
  - Konponketa:

LURSAILAK1

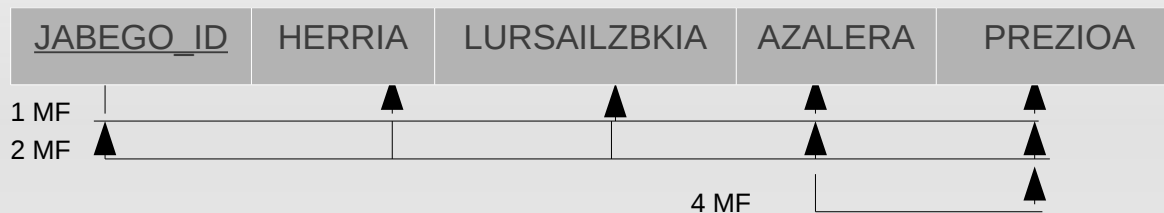


LURSAILAK2

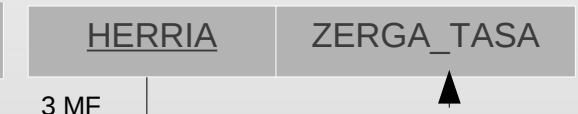


# Normalizazioa. Adibidea.

LURSAILAK1

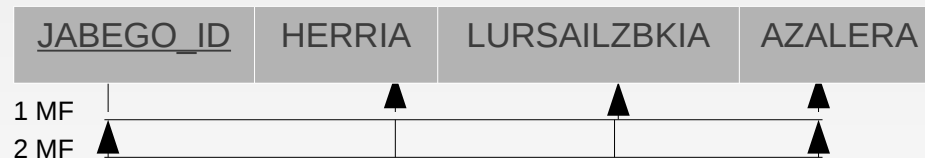


LURSAILAK2

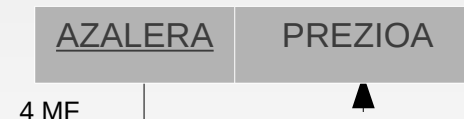


- 3 forma normala hausten da? BAI
  - LURSAILAK1 erlazioko 4MFko AZALERA ez da supergakoa eta PREZIOA ez da atributu lehena
  - Konponketa: Deskonposaketa
    - 3FN hausten duen PREZIOA LURSAILAK1etik kendu eta eta LURSAILAK2n ipini AZALERArekin batera

LURSAILAK1A



LURSAILAK1B



# Forma normalak modu informalean

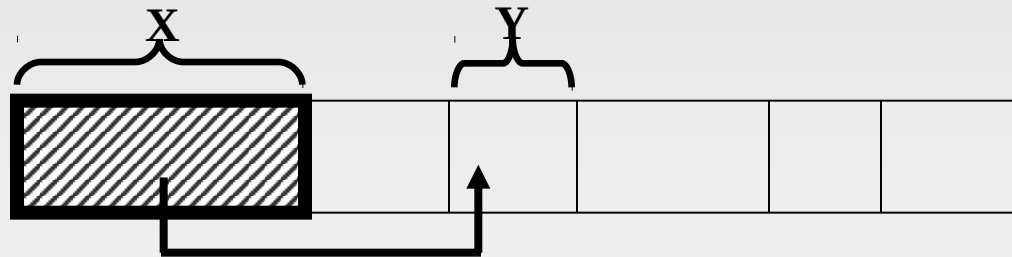
- 1. forma normala
  - Atributu guztiak oinarritzko gakoaren mendeko dira
- 2. forma normala
  - Atributu guztiak oinarritzko gako **osoaren** mendeko dira
- 3. forma normala
  - Atributu guztiak soilik gakoaren mendeko dira



# Boyce-Codd-en Forma Normala

Erlazioa **BCFN**an dago

$X \rightarrow Y$  mendekotasun guztietan **X supergakoa** bada



BCFNan dauden erlazioak 3FNan ere badaude

3FNan dauden batzuk BCFNan ez egotea gerta daiteke  
(adibidez 3FNan daude Y lehena delako)

# BCFN betearazteko prozesua

Aurrekoaren berdina

Baina batzuetan, mendekotasunaren bat galtzea gerta daiteke.

Erabakia diseinatzailearen ardura

- 3FNn utzi: erredundantzia onartu eta kontrolatu behar
  - Irakurketa eragiketak lehenesten dira
- BCFNra pasa: galdutako mendekotasuna esplizituki kontrolatu behar
  - Aldaketa eragiketak lehenesten dira

# Mendekotasunen galera. Adibidea

- $R=(A, B, C)$ :
  - $F = \{AB \rightarrow C, C \rightarrow B\}$
- Gako hautagaiak:  $(A, B), (A, C)$
- Ez dago BCFNan  $(C \rightarrow B)$
- Deskonposatzean  $AB \rightarrow C$  galduko da

# Hori dela eta 3FN-ren baliagarritasuna...

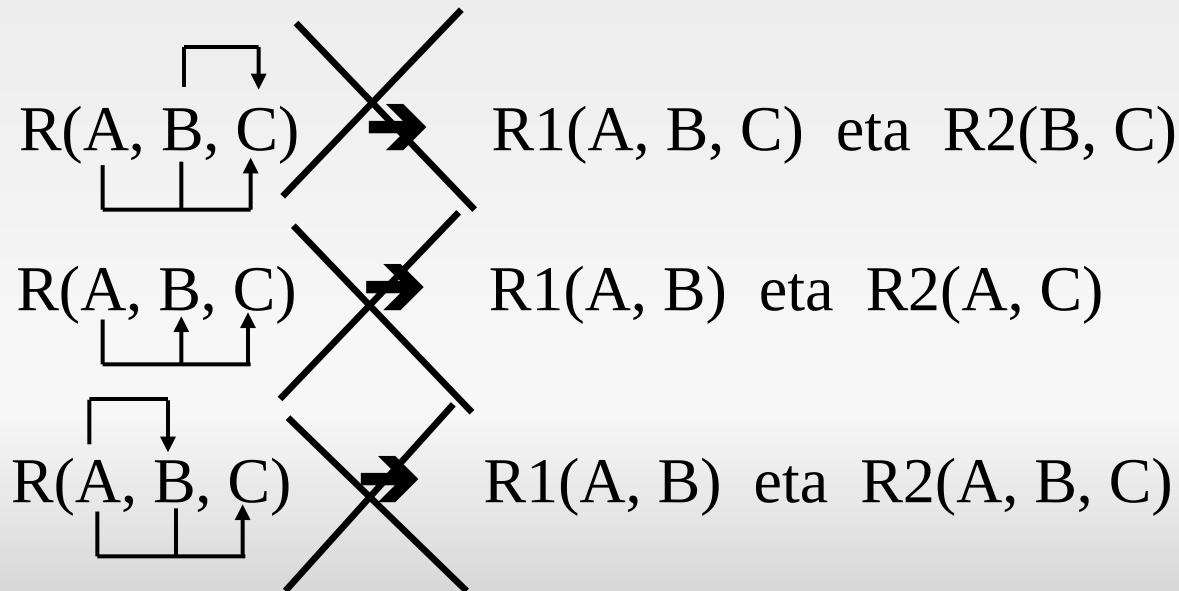
- Beti lor daiteke deskonposaketa 3FNan:
  - Galerarik gabe
  - Mendekotasun funtzionalak mantenduaz

**BCFN:** atributu guztiak soilik gakotik guztiz  
dependente dira

**3FN:** ez-gakoak diren atributu guztiak soilik gakotik  
guztiz dependente dira

# Normalizazioa: ondorioak ateraz ...

- Mendekotasunak, oinarritzko gakoaren mekanismoaren bitartez betearaziko dira
  - DBKSren ardura
- Normalizazioa ez da mendekotasun funtzionalak taula banatan jartzea



# Balio anitzetako mendekotasunak

- $A \rightarrow \rightarrow B$
- Mendekotasun funtzionalaren orokortzea
  - B-ko atributuek hartzen dituzten balioak R-A-B ko atributuek hartzen dituztenetatik independente dira
  - Baldin  $A \rightarrow \rightarrow B$ , orduan  $A \rightarrow \rightarrow R - B$
- $A \rightarrow \rightarrow B$  balio anitzetako mendekotasun arrunta da ondorengo edozein 2 kasuetan:
  - $A \cup B = R$
  - $B \subseteq A$

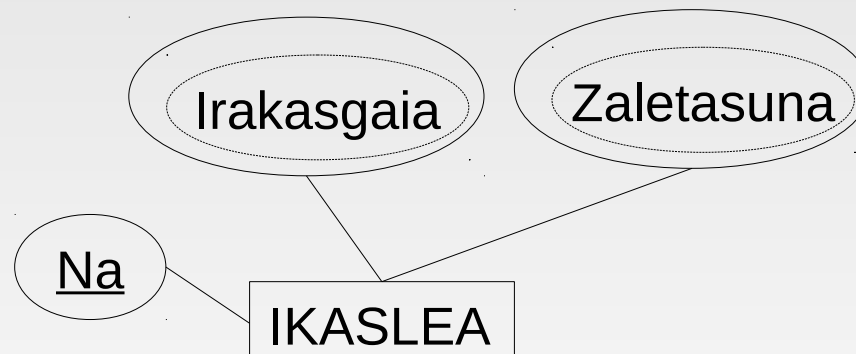
Balio anitzeko mendekotasunen determinatzaileek gako izan behar dute

- Beste modu batera esanda:
  - BCFNan dago eta balio anitzeko mendekotasun guztiak, mendekotasun funtzionalak dira
  - Hau da, mendekotasun guztiak gakoek zehazten dituzte

$4FN \Rightarrow FNBC \Rightarrow 3FN \Rightarrow 2FN \Rightarrow 1FN$

# 4FN adibidea

$\text{Ikaslea} = (\underline{\text{Na}}, \underline{\text{Irakasgaia}}, \underline{\text{Zaletasuna}});$





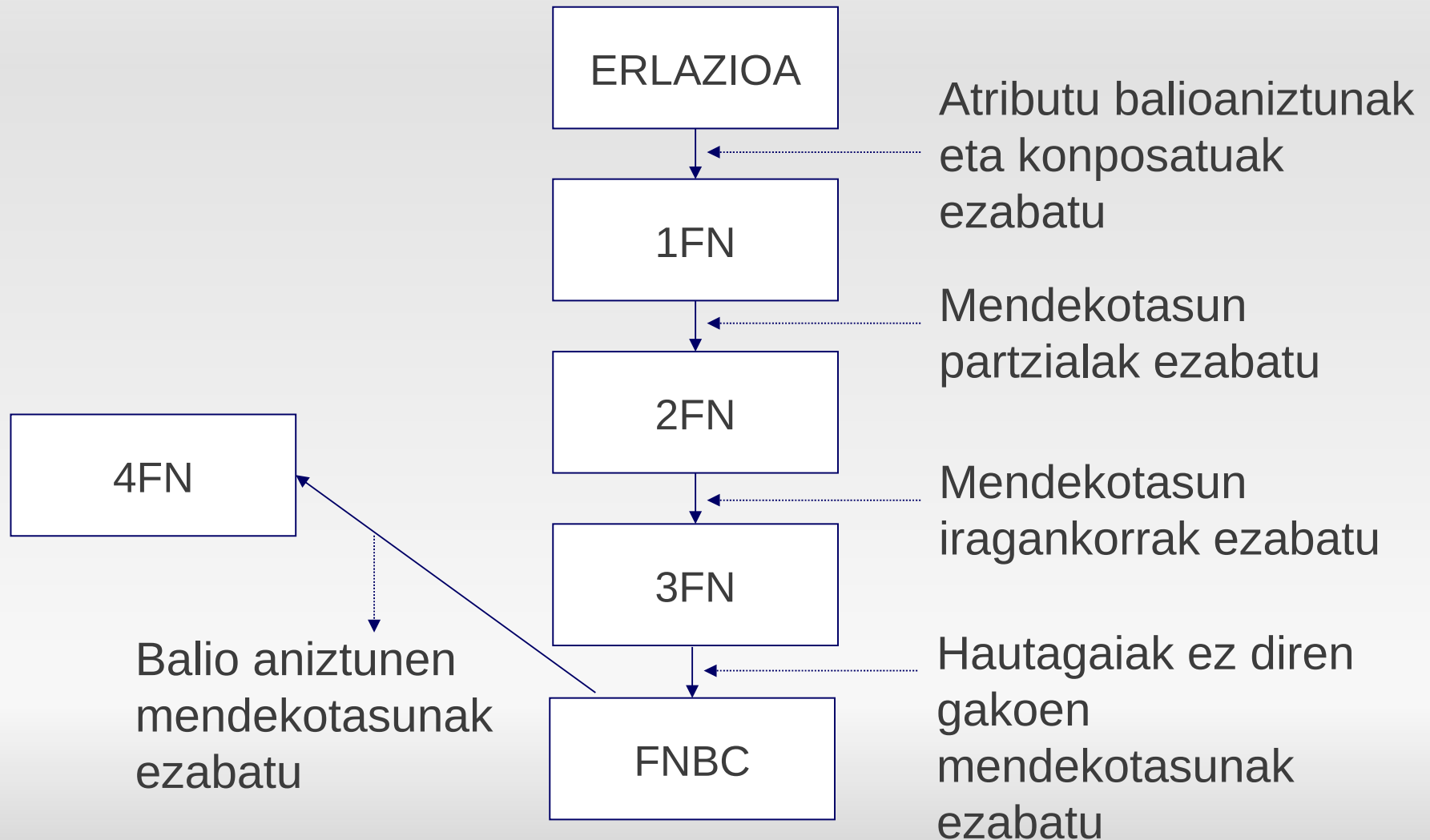
# 4FN adibidea

$\text{Ikaslea} = (\underline{\text{Na}}, \underline{\text{Irakasgaia}}, \underline{\text{Zaletasuna}});$

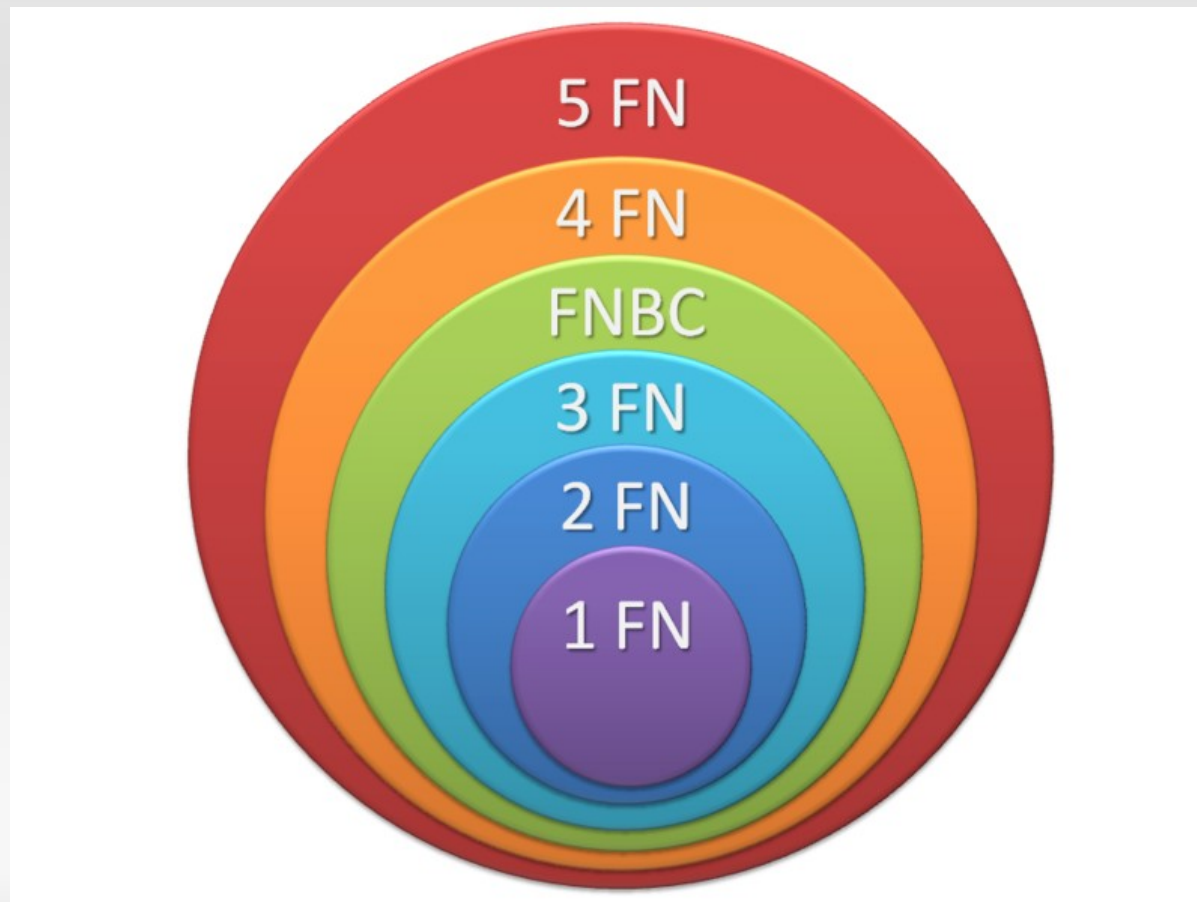
- $\text{Na} \rightarrow \rightarrow \text{Irakasgaia}$
  - $\text{Na} \rightarrow \rightarrow \text{Zaletasuna}$
- } Irakasgaiak eta zaletasunak independenteak dira
- Beraz, deskonposatu egin behar dira:

$\text{Irakasgaia}(\underline{\text{Na}}, \underline{\text{Irakasgaia}});$   
 $\text{Zaletasuna}(\underline{\text{Na}}, \underline{\text{Zaletasuna}})$

# Normalizazio prozesuaren laburpena



# Forma Normalen arteko erlazioa



1FN, 2FN → Arrazoi historikoengatik hartzen dira kontuan

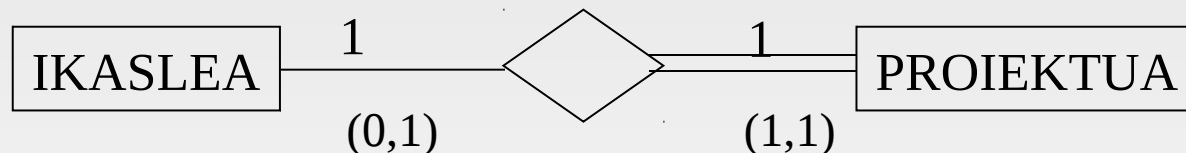
# DesNormalizazioa

- Maiz exekutatzen diren galderen eraginkortasuna hobetzearen, erredundantzia sartzeko prozesua
- Forma normal baxuagoak onartzen ditugu
- Erredundantziarekin *join* egitea ekiditen dugu
- Motak:
  - 1:1 erlazioa duten taulen konbinaketa
  - 1:N erlazioetan gakoarenak ez diren atributuen bikoizketa
  - M:N motako erlazioetan atributuen bikoizketa
  - Atributu balioaniztunak

# DesNormalizazioa.1:1 erlazioa duten taulen konbinaketa

Baldintzak:

- Taulak batera askotan atzitzen dira
- Taulak bereiztuta gutxitan atzitzen dira
- Egon daitezkeen *null* balioen kopurua txikia



IKASLEA			PROIEKTUA		
<u>NAN</u>	izena	...	Pkodea	izena	...
10000000	Jon Lasa	...	LSI-1	Datu-baseak eta internet	...
15000000	Miren Alkorta	...	LSI-2	Internet teknologiak	...
16000000	Aitor Ansa	...	KAT-1	XINU Sistema Eragilea	
12000000	Ane Landa	...	null	null	

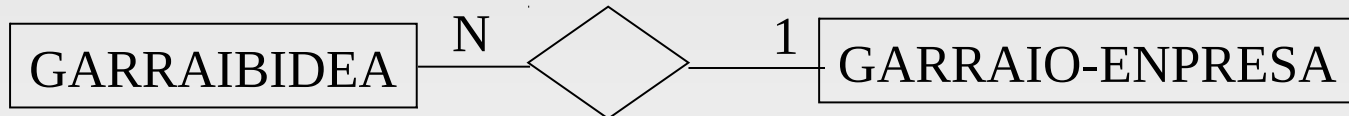
# DesNormalizazioa. 1:N erlazioetan gakoarenak ez diren atributuen bikoizketa

## Baldintzak:

Bikoiztutako atributuak oso gutxitan aldatzen dira

Bikoiztutako atributuak garestiak diren join-etan maiz azaltzen dira

Taula batean egindako aldaketek, bestean ez lukete eguneraketa asko eragingo



GARRAIBIDEA				
<u>Gkodea</u>	mota	...	GEIFZ	izena
G1	kamioia	...	10000	San José
G2	itsasontzia	...	20000	Guipuzcoana
G3	kamioia	...	30000	TDN
G4	hegazkina	...	40000	Alvarez
G7	kamioia	...	10000	San José
G8	kamioia	...	10000	San José

GARRAIO-ENPRESA		
<u>IFZ</u>	izena	...
10000	San José	...
20000	Guipuzcoana	...
30000	TDN	...
40000	Alvarez	...

# DesNormalizazioa. M:N motako erlazioetan atributuen bikoizketa

## Baldintzak

- Bikoiztutako atributuak oso gutxitan aldatzen dira
- Bikoiztutako atributuak garestiak diren *join*-etan maiz azaltzen dira
- Taula batean egindako aldaketek, bestean ez lukete eguneraketa asko eragingo



ENPRESA-SAILA			
<u>TEIFZ</u>	<u>Skodea</u>	izena	Sizena
17000	S1	Tratam1	Karburoak
17000	S2	Tratam1	Erradiaktiboak
17000	S4	Tratam1	Gasak
27000	S2	Tratam2	Erradiaktiboak
37000	S1	Tratam3	Karburoak
47000	S3	Tratam4	Azidoak
57000	S2	Tratam5	Erradiaktiboak

# DesNormalizazioa. Atributu balioaniztunak

## Baldintzak

- Balioen kopuru maximoa ezaguna eta txikia
- Balioen kopuru maximoa finkoa da

*SAILA(Szenb, Sizen, zuzenNAN, STokia1, STokia2, STokia3)*

<b>Szenb</b>	<b>Sizen</b>	<b>ZuzenNAN</b>	<b>STokia1</b>	<b>STokia2</b>	<b>STokia3</b>
1	LSI	30000000	Donostia	Eibar	Gazteiz
2	KAT	15000000	Donostia	null	null
3	KZAA	14000000	Leioa	Donostia	null